RUZICKA Jan, RUZICKOVA Katerina (Eds.)

# GIS Ostrava 2012: Surface models for geosciences

**Ostrava, Czech Republic**


January 23rd - 25th 2012

**Auspices**

AGILE - Association of Geographic Information Laboratories for Europe
ISPRS - International Society for Photogrammetry and Remote Sensing
CAGI - Czech Association for Geoinformation
SAGI - Slovak Association for Geoinformatics
Prof. Ivo Vondrák, CSc., Rector of VSB – Technical University of Ostrava
Prof. Vladimir Slivka, CSc., dr.h.c., Dean FMG VŠB-TU Ostrava

**VŠB – Technical University of Ostrava**

## Editors

Jan Ruzicka
VŠB -Technical University of Ostrava
Faculty of Mining and Geology
Institute of Geoinformatics
17. listopadu 15, 708 33, Ostrava - Poruba, Czech Republic
jan.ruzicka@vsb.cz

Katerina Ruzickova
VŠB -Technical University of Ostrava
Faculty of Mining and Geology
Institute of Geoinformatics
17. listopadu 15, 708 33, Ostrava - Poruba, Czech Republic
katerina.ruzickova@vsb.cz

Production:

Institute of Geoinformatics
VŠB -Technical University of Ostrava
17. listopadu 15,
708 33, Ostrava - Poruba,
Czech Republic
jan.ruzicka@vsb.cz



ISBN 978-80-248-2667-7
ISSN 1213-2454

**GIS Ostrava 2012: Surface models for geosciences**

January 23rd - 25th 2012, Ostrava, Czech Republic

| | |
|---|---|
| Printed by: | VŠB -Technical University of Ostrava, 2012 |
| Bound by: | OFTIS Ostrava, Chelčického 693/7, 702 00 Ostrava, Czech Republic |
| Edition: | First edition, 2012 |
| Number of pages: | 311 |

## Preface

Surface models are frequently used for many applications in geosciences. Models with different scales, various data sources, alternative construction approaches and representations are searched to best fit for the given purpose. The common focus is a model quality adequate to satisfy our needs and efficiency requirements. The controlled quality is essential to manage uncertainty in our processing and decision making. Extensive field studies and research have been conducted to address these issues. The aim of the conference is to present and discuss new methods, issues and challenges encountered in all parts of a complex process of gradual development and utilization of digital surface models. The process covers data capture, storage, model creation, validation, manipulation, visualization and utilization. Each phase requires adequate methods and contains issues which may substantially decrease the value of the model. Surface models play important roles in wide spectrum of domains. The proceedings contains results from latest research in the area of surface models for geosceinces.

## Papers selection

Number of registered papers: 83
Number of papers published in this proceedings: 24
Number of papers published in journal Transactions in GIS: 9

Countries from which come authors of the papers published in this proceedings:
Republic of Serbia, Czech Republic, Germany, Slovakia, Iran, Republic of Ireland, Sweden, Malaysia, United States of America, Turkey, Poland, Slovenia and Australia

## Programme committee

Dr. Kateřina Růžičková (VŠB-Technical University of Ostrava, CZ) - chairman

Dr. Farès Belhadj (Université Paris, FR)

Assoc. prof. Lena Halounová (Czech Technical University in Prague, CZ)

Dr. Tomislav Hengl (ISRIC World Soil Information institute, Wageningen, NL)

Assoc. prof. Tomáš Hlásny (National Forest Centre Zvolen, SK)

Assoc. prof. Jaroslav Hofierka (University of Presov in Presov, SK)

Assoc. prof. Jiří Horák (VŠB-Technical University of Ostrava, CZ)

Assoc. prof. Josef Kohout (University of West Bohemia, CZ)

Assoc. prof. Jaromir Kolejka (MU Brno, CZ)

Prof. Gotffried Konecny (Hannover, DE)

Prof. Brian Lees (The University of New South, AUS)

Dr. Xiaoye Liu (University of Southern Queensland, AUS)

Assoc. prof. Helena Mitášová (NC State Univ., USA)

Assoc. prof. František Staněk (VŠB-Technical University of Ostrava, CZ)

Dr. Tomáš Vaníček (Czech Technical University in Prague, CZ)

Prof. Vít Voženílek (Palacky University in Olomouc, CZ)

Prof. J. P. Wilson (University of Southern California, USA)

Prof. Qiming Zhou (Hong Kong Baptist University, HK)

## Content

# A CONCEPTUAL HYDROGEOLOGICAL MODEL FOR THE OPEN PIT MINE "POLJE E" (KOLUBARA COAL BASIN, SERBIA)

Dragoljub BAJIC[1] and Dusan POLOMCIC[2]

[1,2]University of Belgrade, Faculty of mining and geology, Department of hydrogeology, Djusina 7, 11000, Belgrade, Republic of Serbia

*[1]osljane@orion.rs, [2]dupol2@gmail.com*

**Abstract**

The Kolubara Coal Basin ranks among the largest coal basins in the Republic of Serbia. Numerous exploration-exploitation fields: "Polje A", "Polje B", "Polje C, "Polje D", "Polje E", "Polje F", "Polje G", "Veliki Crljeni", "Šopić-Lazarevac", "Tamnava – Zapadno polje", "Tamnava – Istočno polje", "Radaljevo", "Zvizdar", "Ruklade" and "Trlić" have been delineated administratively in the Basin. The "Polje E" of the Kolubara Coal Basin is the exploration area dealt with in this paper.

The analysis and drawing of specific conclusions on the hydrogeological function of present lithological members at the exploration area, as well as realizing of their spatial relations are extremely significant while forming the conceptual hydrogeological model of an area. The main purpose of a conceptual hydrogeological model is drawing conclusions on the distribution of present aquifer types in the area, their mutual relations, as well as conditions of groundwater movement. As such, later it represents a basis for the superstructure - the development hydrodynamic model of groundwater regime.

Observed in a vertical profile, according to the hydrogeological function of rock masses, there have been separated overall seven layers in the open pit "Polje E", among which four are confining strata, the fourth layer is a water-bearing one, and two are combined water-bearing-confining strata. Both the separation of these layers and their contouring were carried out on the basis of conducted analyses of numerous geological and hydrogeological explorations.

A conceptual hydrogeological model of the "Polje E" was developed on the basis of data with coordinates of borehole positions and separated layer elevations and with carried out schematization, where by software support, the separation and survey of some lithological members, as well as the survey of unlimited number of geological and hydrogegological profiles are enabled.

**Keywords: a conceptual hydrogeological model, open pit mine, aquifer, coal basin**

## INTRODUCTION

The Kolubara Coal Basin covers an area of about 600 km$^2$. It is situated about 50 km south east of Belgrade (Fig. 1). It spreads in the middle and lower flow of the River Kolubara and its tributaries: the Tamnava on the left, and the Peštan on the right sides.

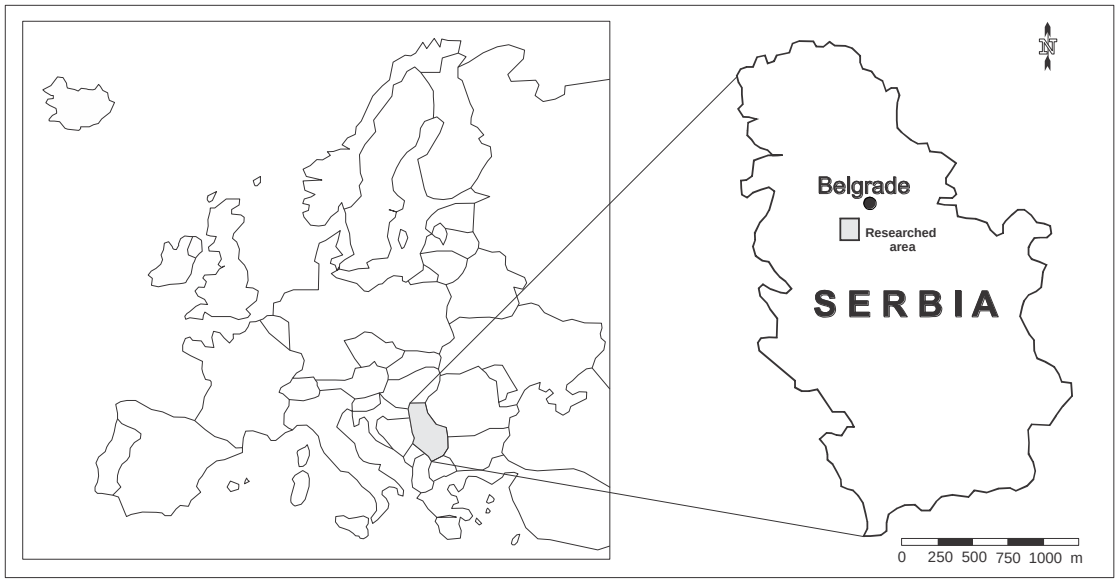



**Fig.1.** Geographic position of exploration area

Parts made of Palaeozoic and Mesozoic formations and vulcanite (south and south–east areas), as well as those made of Neogene sediments (north-east area) can be separated in the geological setting of the east Kolubara Basin (the “Polje E”). The geological map of the exploration area is presented in Fig. 2, while the geological profile of the terrain is presented in Fig. 3.

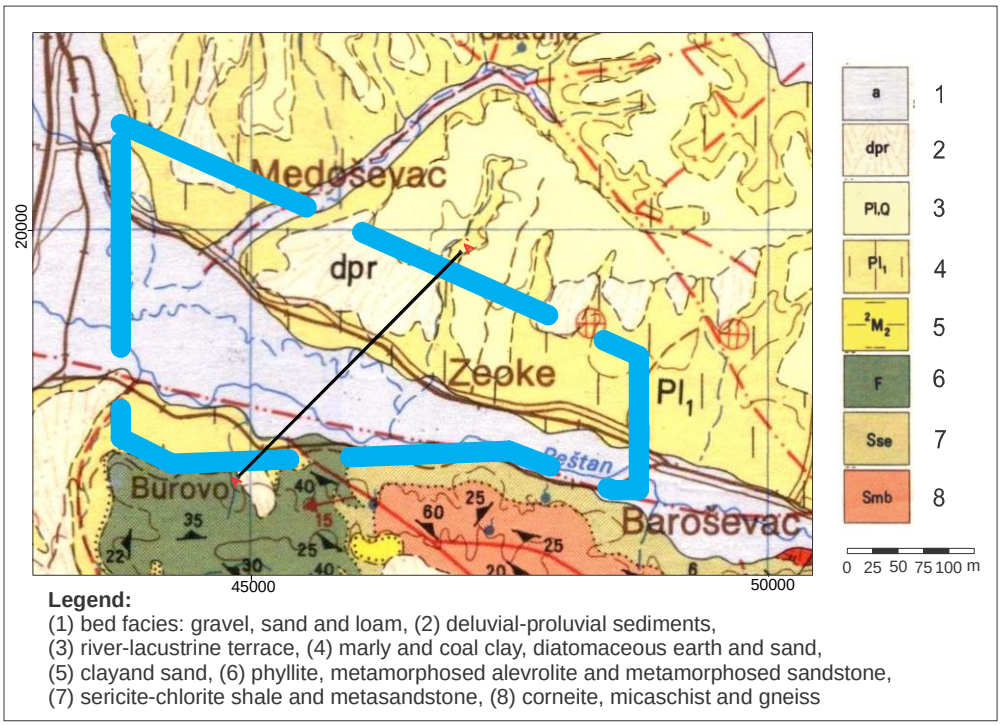



**Fig. 2.** Geological map of exploration area

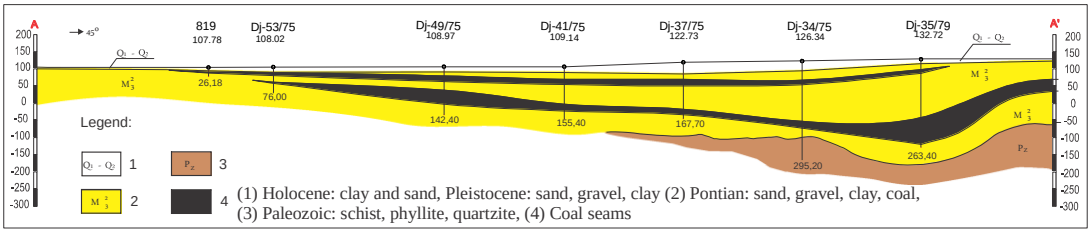



**Fig. 3.** Geological profil A-A'

The tectonic structure is clearly pronounced due to the impact of the Alpine-Hercynian orogeny. Palaeozoic formations were modified by movements of the Hercynian orogeny, and consequences of a volcanic activity and contact metamorphism are present as well.

Mesozoic formations were reshaped in younger phases of the Alpine orogeny. Complex breaking zones, particularly faults of E-W and NW-SE strike are distributed here, and in addition fault structures of WNW-ESE and NW-SE strike [1].

Miocene-Pliocene and Quaternary formations belong to the upper structural stage and are characterized by simple tectonic setting.

Neogene sediments are sub-horizontal in position, while these layers are bent even more steeply (to 25°) at some places as the consequence of post-Pontian tectonics when old faults were activated again [1].

The consequence of these movements, is that a coal series in east and south parts of the Basin has been disturbed, therefore coal lies in shallow synclines and anticlines of the WNW-ESE strike. The coal series was partly to completely eroded in limbs and apexes of anticlines during the Quaternary, while, in synclines, it has been completely preserved. Neogene sediments are covered by lacustrine - river alluvial deposits of Quaternary age on its complete surface [1]. The block diagram of the Neogene substratum, with the survey of the rupture texture, is presented in Fig. 4, while the block diagram of the current relief, with the survey of the rupture texture, is presented in Fig. 5.

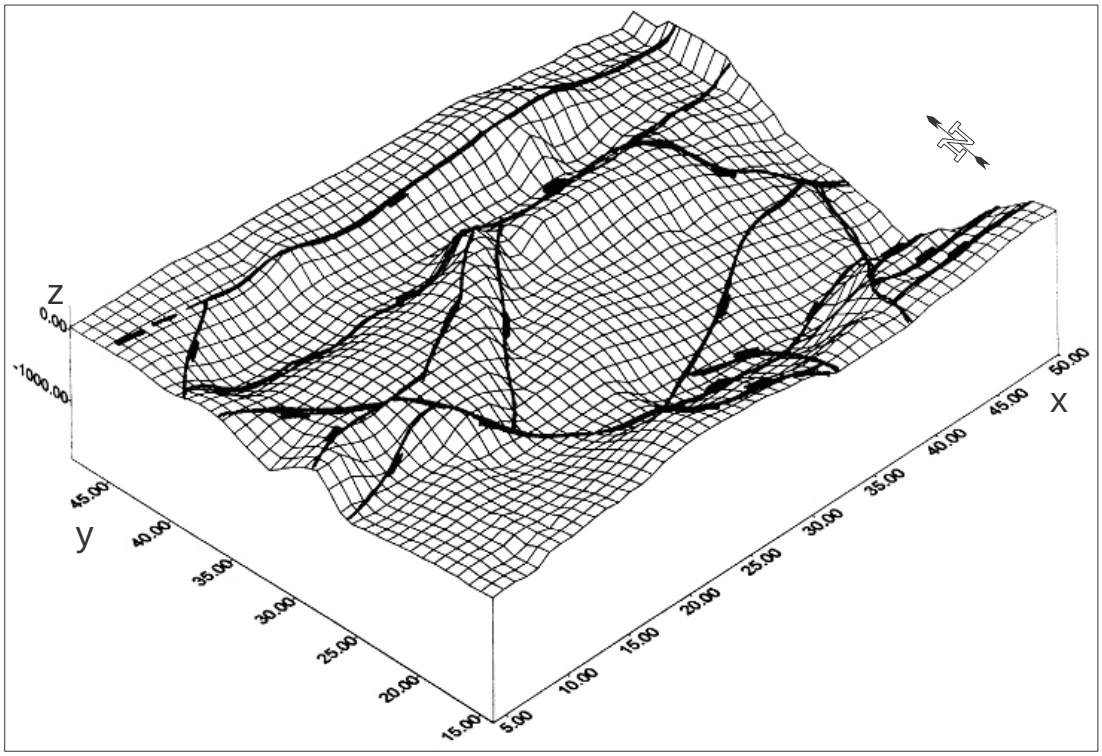



**Fig. 4.** Block diagram Neogen substratum with survey of rupture texture in exploration area

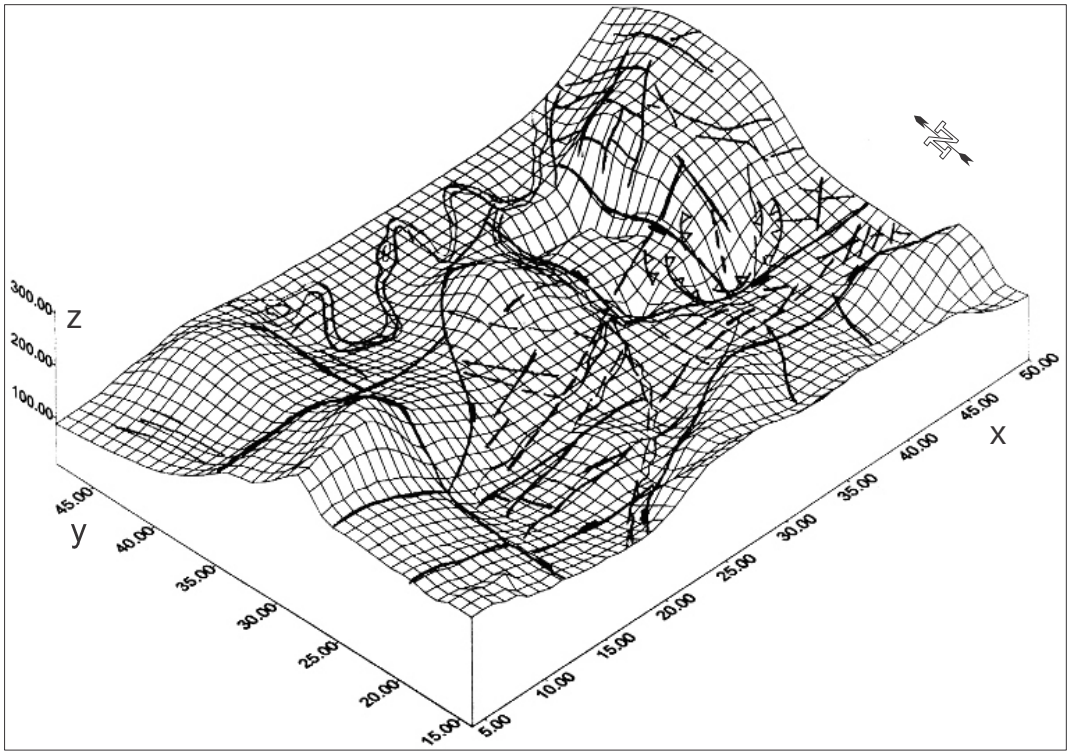



**Fig. 5.** Block diagram of current relief with survey of rupture texture in exploration area

## HYDROGEOLOGICAL CHARACTERISTICS OF OPEN PIT MINE “POLJE E”

In the wider area of the Kolubara Coal Basin, there have been formed three aquifers: overlying, interlayered, and substratum ones.

*The overlying aquifer* was formed in alluvial and Upper Pontian sand and gravel in the roof of the upper coal seam. In east part of the open pit mine “Polje E”, as well as in the neighbouring open pit mine “Polje C”, whose normal extension is the “Polje E”, there is also immediate contact of alluvial and interlayered aquifers. Taking into consideration the erosion of the main coal seam also occurred in the open pit “Polje C”, there is no possibility of water overflow from the overlying aquifer of the main coal seam into the alluvial aquifer [2]. The overlying aquifer is cut off in the north by mine works of the open pit mine “Polje D”, and in the south and east is related to the alluvion of the Peštan River. The aquifer is recharged by infiltration of precipitation and river water of the Peštan, while drainage is diffusive on bottom sides of gullies and brooks.

*The interlayered aquifer* is formed between two coal seams in so called interlayered sands. By its distribution, location, and thickness it is the most significant aquifer within the Kolubara Coal Basin. Within the open pit mine “Polje C” this aquifer represents an extension of the overlaying aquifer from the south part of the open pit mine “Polje D”, where the main coal seam represents a substratum aquifuge. The interlayered aquifer is recharged in the area of the unconfined aquifer where water is infiltrated by precipitation and in some zones of the “Polje C”, where the upper coal seam is eroded, thus in these parts the alluvion is in immediate contact with interlayered sands [2]. The drainage is carried out naturally via the alluvion of the Kolubara River, where hydraulic connection between the interlayered aquifer and the aquifer in the alluvion of the Kolubara River is realized, and artificially via water tapping facilities (wells).

*The substratum aquifer* was formed in Lower Pontian sands underneath the main coal seam. It is situated below the level of the erosion basis and the water level of the Turija and Peštan rivers. As the main coal seam lies as deep as the – 200 m elevation, this aquifer is under pressure as the main coal seam makes the roof of the substratum aquifer [3]. The aquifer is recharged in terrains beyond the “Polje E” by infiltration of precipitation in the area of eroded coal and uncovered substratum sand, in addition via water infiltration of the Turija river. The substratum aquifer is drainaged in the area of the Kolubara river alluvion, and the secondary drainage occurs through the substratum part of coal, through holes that remained after drilling in the area of the Junkovac shaft [3].

On the basis of terrain exploratory work and the conducted analysis of distributed lithological members and their hydrogeological function there have been separated oveall seven layers in he vertical profile on the open pit mine “Polje E”. Observed from the surface of the terrain, from the hydrogeological aspect, corresponding terrain layers are [4]:

- first confining stratum: overlying clayey sediments of Quaternary age,
- second combined water – bearing: alluvial gravel and sand in which overlying aquifer was formed,
- confining stratum: clayey sediments in south pat of terrain,
- third mainly confining stratum: upper coal seam changing into sandy sediments in south part of terrain,

- fourth water- bearing stratum: Pliocene sandy sediments of interlayered aquifer,
- fifth mostly confining stratum:main coal seam wedging out in south part where sand of substratum aquifer and decayed shale appear,
- sixth combined water-bearing: sandy sediments in north part of terrain where substratum aquifer was formed changing into,
- confining stratum: clayey sediments,
- seventh mainly confining stratum: shale with isolated parties of sandy layers in north part of terrain.

The contour determination of separated layers was carried out on the basis of data of numerous test holes (about 220) distributed on the whole exploration area. The location of boreholes within the coal deposit “Polje E” used for defining spatial characteristics of separated layers is presented in Fig. 6.

On the basis of data processing from grain size analyses of core samples from boreholes, as well as the data obtained by data processing of well test pumping, values of hydrogelogical parameters of the porous environment (porosity, hydraulic conductivity and specific storage) are given in Table 1 [4].

Table 1: Values of hydrogeological parameters

| **Lithostratigraphic unit** | **Hydraulic conductivity (m/s)** | **Specific storage (1/m)** | **Porosity (-)** |
|---|---|---|---|
| Quaternary clay | $1 \times 10^{-7}$ | $1 \times 10^{-3}$ | 0.44 |
| Alluvial sand and gravel | $1 \times 10^{-4}$ | $2.2 \times 10^{-5}$ | 0.25 |
| Upper coal seam | $1 \times 10^{-8}$ | $6 \times 10^{-5}$ | 0.1 |
| Pl sand of interlayered aquifer | $7 \times 10^{-6} – 5.6 \times 10^{-4}$ | $1.9 \times 10^{-4}$ | 0.34 |
| Main (lower) coal seam | $1 \times 10^{-8}$ | $6 \times 10^{-5}$ | 0.08 |
| Substratum fine-grained sand | $8.7 \times 10^{-7}$ | $1 \times 10^{-5}$ | 0.36 |
| Substratum clay | $1 \times 10^{-7}$ | $1 \times 10^{-3}$ | 0.42 |
| Decayed shale | $2.5 \times 10^{-6}$ | $6.1 \times 10^{-5}$ | 0.1 |

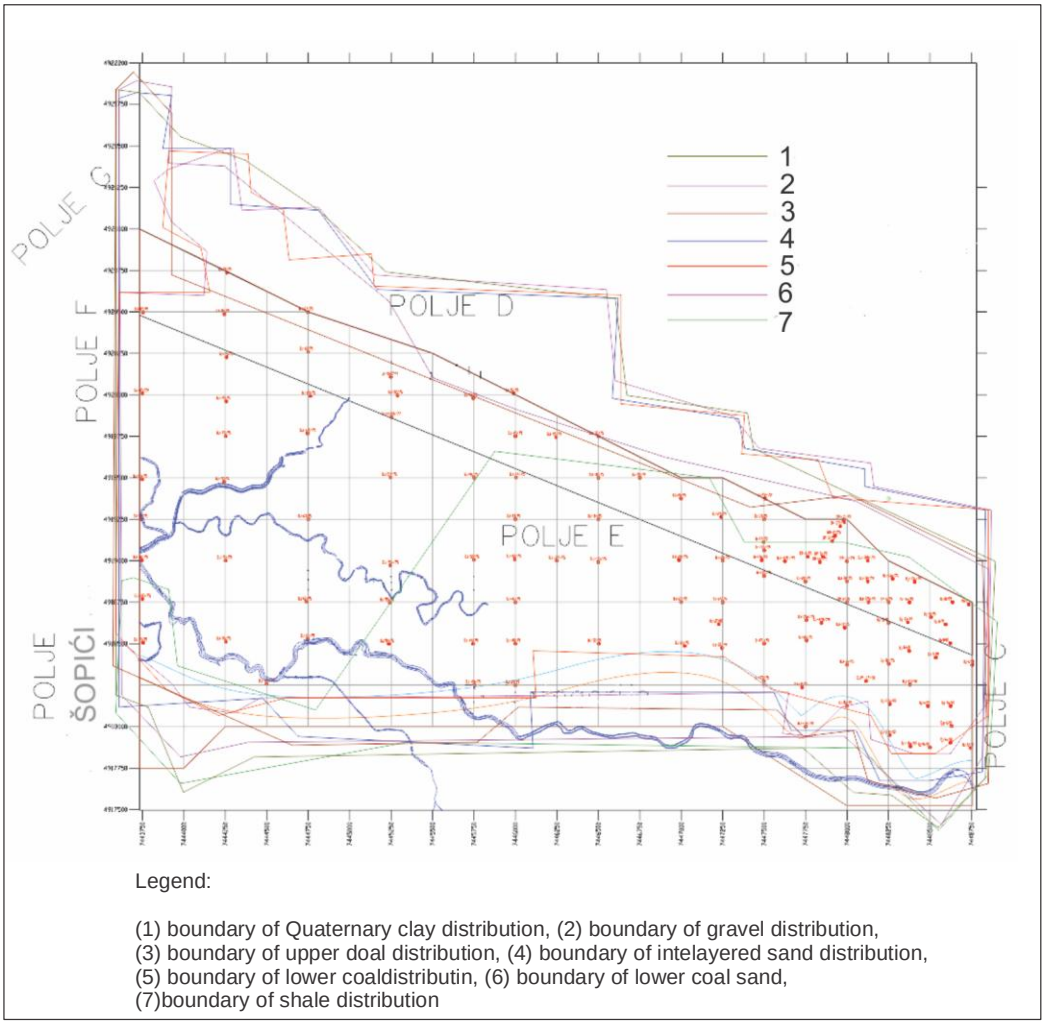



**Fig. 6.** Map of borehole location and contoures of all lithological members of “Polje E” coal deposit

## METHODOLOGY OF DESIGNING THE 3D CONCEPTUAL HYDROGEOLOGICAL MODEL

A hydrogeological system is defined by specific incoming elements, environmental characteristics, processes in the monitored environment and outgoing elements, whereby incoming and outgoing elements are interdependent. Characteristics of a hydrogeological system are its geometry, namely the spatial distribution of a water-bearing environment and its seepage characteristics [5]. Basic processes occurring here are the filtration of groundwater in a porous environment, and the matter and heat transfer by aquifer flow. Owing to the impact of these processes, there also occur changes in the system itself, reflecting in the fluctuation of piezometric level of aquifers, the alteration of qualitative properties of aquifer water and soil deformation of filtration.

On the basis of above mentioned it can be concluded that each hydrogeological system is characterized by the complexity of involved processes which cannot be interpreted directly

mathematically. Therefore some simplification of the system is required, a form convenient for mathematical description of processes occurring in it - so called schematization of a hydrogeological system. Such simplification results in the development of a hydrogeological model. Thus a calculating scheme is formed which will serve as a basis for the development of a hydrodynamic (mathematical) model of the studied aquifer regime.

In the process of hydrogeological model development, the following pieces of schematization of the hydrogeological system are used [6]:

1. The schematization of the flow area (aquifer geometry) by which the changes of thickness of the water-bearing environment reach average values, or the aquifer flow is divided into parts of the same thickness;

2. The schematization of seepage characteristics of the porous environment and the hydrodynamic state of the aquifer flow caried out owing to the heterogeneity of seepage characteristics of the environment, and by it, zones of the same seepage characteristics of the porous environment are separated, then in case that a hydrogeological system is characterised by a complex character of aquifer water, the aquifer flow is divided into parts with the same flow conditions (parts of the aquifer under pressure and parts of the aquifer with unconfined level). This schematization also embraces an analysis of conditions of groundwater movement, whereby the unsteady state of groundwater flow is reduced to the quasy-steady one, and spatial (three dimensional) groundwater movement is also reduced to more simple plane flow types.

3. The schematization of the aquifer regime on the basis of which specific regularities of both the occurrence and the distribution of factors affecting the groundwater regime are determined.

4. The schematization of groundwater balance elements, where the number of incoming and outgoing balance elements is reduced by determining of specific relations among those elements, by their uniting or knowingly neglecting some of balance elements supposed to be less significant.

5. The schematization of boundary conditions of the hydrogeological system which, generally speaking, represent water inflow and outflow to the model.

6. The schematization of initial conditions, wich represent values of piezometric levels in some initial time.

The code selected to develop the conceptual hydrogeological model was MODFLOW-2000; a modular, three-dimensional finite difference groundwater flow model developed by US Geological Survey [7]. The program used in this work is Groundwater Vistas 5.33b (Environmental Simulations International, Ltd.).

## 3D CONCEPTUAL HYDROGEOLOGICAL MODEL OF COAL DEPOSIT “POLJE E”

As stated earlier, there have been separated seven layers being of heterogeneous composition. Carried out schematizations enabled the development of the conceptual hydrogeological “Polje E” model. The developed conceptual hydrogeological model with the

software support enables the separation and survey of some lithological members and the survey of unlimited number of geological or hydrogeological profiles. Hydrogeological profiles of the open pit mine “Polje E” with the strike south-north and west-east are presented in Fig. 7 and Fig. 8, as an illustration.

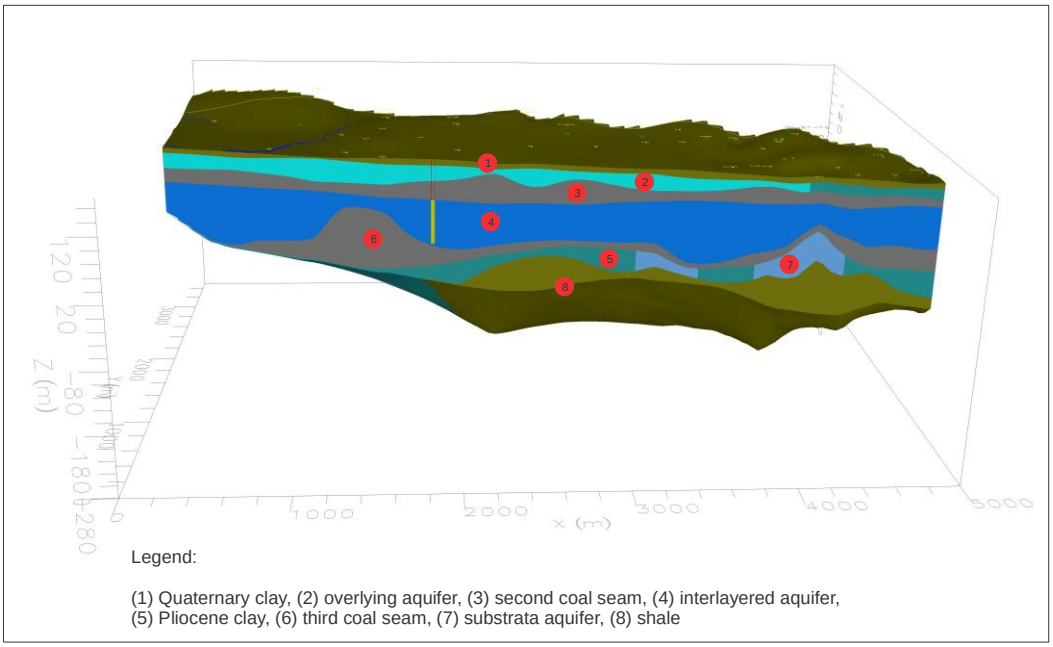



**Fig. 7.** Hydrogeological profile of strike south-north

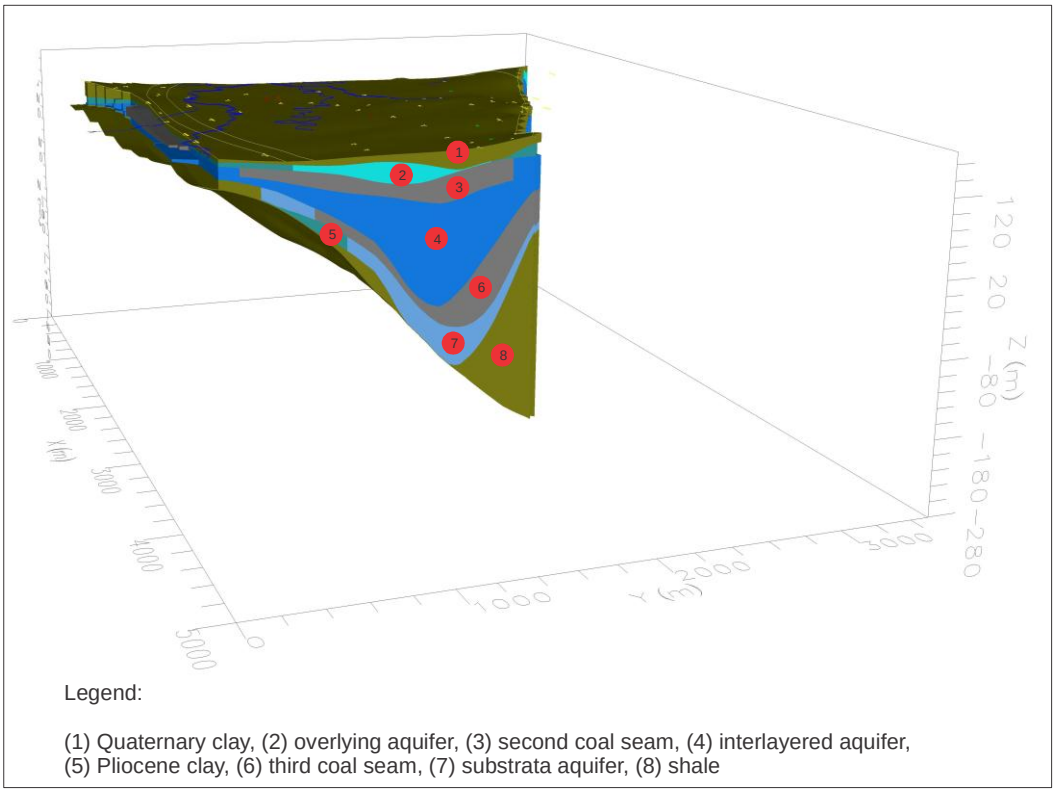



**Fig. 8.** Hydrogeological profile of strike west-east

The distribution of all seven layers can be stated in presented profiles whereby it is characteristic of the distribution of most layers (with the exception of the first and the fourth ones, observed from the surface of the terrain) that two and even three lithological members participate within one layer. Within the sixth combined water-bearing-confining stratum, there was formed a substratum aquifer in the north part of the terrain within sandy sediments that wedge out towards south. This layer occurs even further, whereby, laterally (towards south), it changes into clayey sediments spreading to the left bank of the Peštan where they wedge out and afterwards (in the furthest south) change into decayed shale located in the seventh layer. Thus, the sandy layer is represented as it exists in nature, namely in the form of lenses. Similarly, the upper coal seam exists within the third layer which in the north and the south of the terrain changes into sandy sediments. The main (third) coal seam occurs as the prevailing lithological member within the fifth layer. It wedges out in the south as well as, where the sand of substratum aquifer and decayed shale occur. Decayed shale also occurs within the seventh layer in the south and central part of the terrain (the open pit mine) with isolated parties of sandy layers in the north part of the open pit mine.

It has already been emphasised that three aquifers have been formed at the wider area of the Kolubara Basin: overlying, interlayered and substratum ones. Spatial positions of interlayered and stratum aquifers in the open pit mine “Polje E” are shown in Fig. 9 and Fig. 10, as an illustration of the stated schematization.

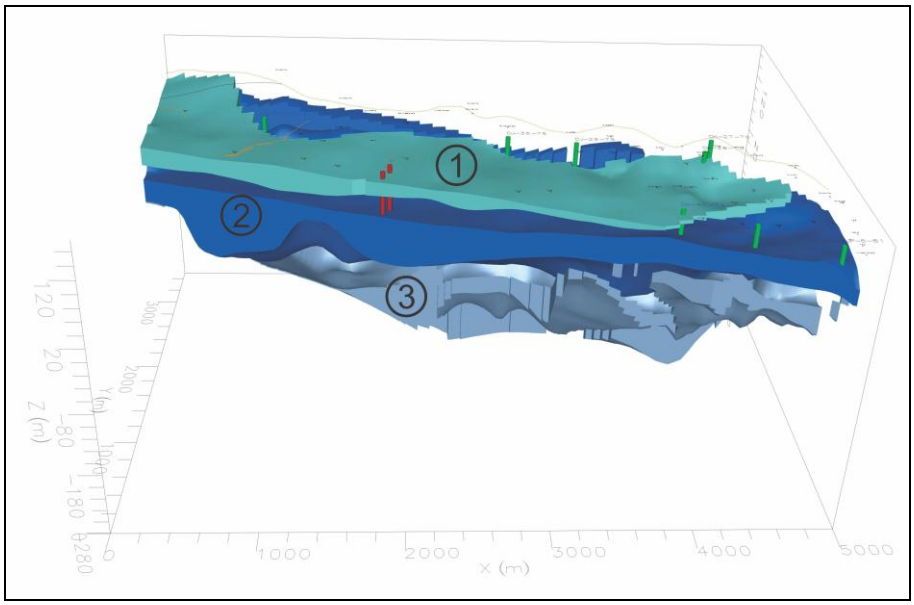



**Fig. 9.** Survey of distributed types of aquifers (cross-section west-east) - from surface of terrain: overlying (1), interleyered (2) and substrata ones (3)

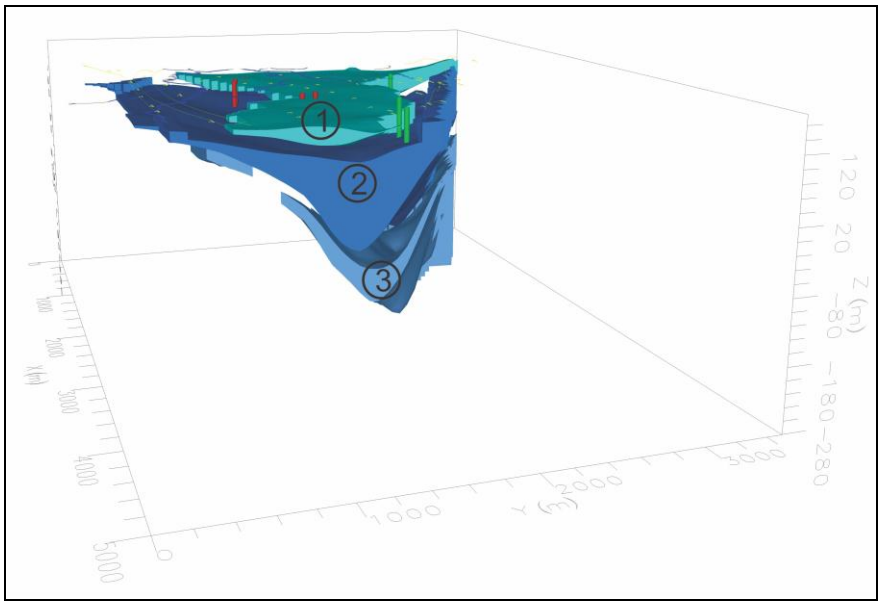



**Fig. 10.** Survey of represented types of aquifers (cross-section south-north) - from surface of terrain: overlying (1), interlayered (2) and substata ones (3)

## CONCLUSION

Natural, geological conditions of the formation of lithological members in the study area have conditioned the geological setting and the texture of sediments as well as hydrogeological relations of lithological members.

The spatial (3D) flow of groundwater is pronounced in contact zones of sandy-gravely sediments with the sand of the interlayed aquifer, and lower, the sand of the stratum aquifer, where lithological stratification in a vertical profile and uneven horizontal distribution of lithological members are observed. This has impacted the selection of basic model characteristics, thus a multilayer hydrogeological model with seven layers in a vertical profile has been developed.

The completed conceptual hydrogeological model represents a direct basis for the development of groundwater regime hydrodynamic model that should enable carrying out of the hydrodynamic analysis of the “Polje E” protection conditions from groundwater with the selection of optimal system of drainage facilities.

## REFERENCES

[1] Bogdanovic, V. and Ilic, Z., Polomcic, D. (2006) Preliminary programme with Justifiability study of “Polje E” open pit mine opening and developing. Technical documentation from Kolubara projekt, Lazarevac, Serbia. University of Belgrade, Faculty of Mining and Geology, Belgrade.

[2] Polomcic, D. (2008) Hydrodynamical model of the open pit “Polje C” (Kolubara Coal Basin, Serbia). IV International Conference “Coal 2008”, Belgrade, Serbia, 407-419.
ISBN 978-86-7352-193-0.

[3] Bogdanovic, V. and AndIlic, Z. (2007) Selection of limitation and opening of South Field open pit mine in Kolubara Coal Basin, Technical documentation from Kolubara projekt, Lazarevac, Serbia. University of Belgrade, Faculty of Mining and Geology, Belgrade.

[4] Polomcic, D. and Bajic, D. (2011) 3D Hydrodynamic model of open pit mine “Polje E” (Kolubara coal basin). V International Conference “Coal 2011”, Zlatibor, Serbia, 320-330.
ISBN 978-86-83497-17-1.

[5] Polomcic, D. (2002) Schematization types of hydrogeological system for needs of hydrodynamic model development. Proceedings of XIII Yugoslav symposium on hydrogeology and engineering geology. Herceg Novi, Serbia, 389-396.

[6] Polomcic, D. (2001) Shematization types of hydrodynamic research, opening and managing of groundwater sources in intergranular porous environment. University of Belgrade, Faculty of Mining and Geology, Belgrade.

[7] Harbaugh, A.W., Banta, E.R., Hill, M.C. and McDonald, M.G. (2000) MODFLOW-2000: The U.S. Geological Survey Modular Ground-Water Model - User Guide to Modularization Concepts and the Ground-Water Flow Process, U.S. Geological Survey Open-File Report 00-92, Reston, VA, USA, http://water.usgs.gov/nrp/gwsoftware/modflow2000/ofr00-92.pdf

# THE QUALITY OF THE BASE DEM AS A KEY FACTOR IN MODELING OF SLOPE AND SOLAR INSOLATION

Jaroslav BURIAN[1], Jana SVOBODOVA[1] and Vit VOZENILEK[1]

[1]Department of geoinformatics, Faculty of Science, Palacký University, Olomouc, tř. Svobody 26, 771 46, Olomouc, Czech Republic

*j.svobodova@upol.cz, jaroslav.burian@upol.cz, vit.vozenilek@upol.cz*

**Abstract**

The results of analysis of spatial conflicts are standard content of territorial planning documents for creating or modifying urban master plans. The paper presents the results of spatial modeling of selected functional spatial conflicts with the characteristics derived from the earth's surface (slope and insolation rate). These two characteristics can be, in terms of comfort and quality of life, considered as very important. The urban planners describe that higher slope affect bigger technical complexity of construction and subsequent maintenance of buildings and technical infrastructure. The insolation rate impacts especially the quality of the environment and has strong influence to the localization of the activities defined as the visual activity (according to standard “CSN 730580” [1]).

The main aim of this paper is not just finding the spatial conflicts in the model area. The aim is also to evaluate the differences between slope and insolation rate derived from digital terrain model (DEM) of different quality. The differences for concrete selected area are expressed by numerical quantification. Spatial descriptions (maps) show the qualitative changes of conflicts importance. The results of the analysis show that the quality of the base DEM affects the definition of areas with critical values of observed phenomena very strongly. For use as a content of territorial planning documents is creation of high quality DEM important point of urban planning processes.

**Keywords: slope, insolation, spatial conflict, digital terrain model**

## INTRODUCTION

There most human activities which take place in space, i.e. in a particular area, are to a great extent dependent on the nature of the Earth’s surface. The Earth's relief determines suitability of localization of these activities in a particular area or limits of utilization of an area for a given activity [2]. If those limits and restrictions are not observed, conflicts of various gravity may arise [3], [4], [5], [6], [7].

Problems of modelling spatial conflicts of functional utilization of an area bearing characteristics derived from the Earth’s surface are generally dealt with by urban planning. Significant conflicts always occur when there is a clash of functional areas for housing, sports and recreational activities with other characteristics, specifically gradients of slopes and rate of their solar insolation. To derive these characteristics of the Earth's surface in the environment of geographic information systems (GIS) a digital elevation model (DEM) is used, which substitutes the real Earth’s surface, e.g. [8], [9], [10]. To derive a continuous DEM from input data, (mostly discrete elevation points) it is possible to use various interpolation methods

[11], [12], results of which can, however, be outputs of different quality [13], [14], [15], [16], [17], [18]. Therefore the primary objective of this article is an evaluation of influences of DEM quality (from the point of view of the chosen method of interpolation) on results of modelling of the aforementioned spatial conflicts within the frame of urban planning.

Model area for the elaboration of the demonstrations has been placed in Olomouc region (Municipality with extended power Olomouc).

## ROLE OF EARTH'S SURFACE IN URBAN PLANNING

Shape and pattern of the Earth's surface can be described by its basic characteristics, the so-called morphometric variables. [19] have divided them into two basic groups. The first group consists of characteristics derived directly from DEM (of earth's surface) such as slope, aspect or curvature of relief, the second group represents characteristics describing natural processes taking place on the Earth's surface. However, while evaluating suitability of an area for utilization within the frame of urban planning, the most important aspects are gradient and aspect of slope.

Gradient of slope is a property of relief which, to a great extent, influences most human activities in any area. Flatlands and easy slopes were always sought-after and settled sooner than slopes with higher gradients. From the urbanistic point of view, the higher the gradient of slope the more technically demanding the construction and the subsequent maintenance of buildings and public utilities are. The most suitable building sites are flat or gently sloped (with a gradient up to 10%). More detailed limits are stated by [20] or by [21]. Consequences of ground shaping in steeper areas have to be dealt with already in the basic urbanistic concept which has to observe not only the aspects of construction and operating economy, balancing moving of earth, concept of local roads and engineering infrastructure but also a general influence it has on landscape [20]. Especially in protected localities, adjustments resulting in levelling terrain undulations in sloping areas can fundamentally change character of landscape and devastate its present values [22].

Gradient of slope together with its aspect implicate rate of solar insolation of relief (or angle of incidence of sunbeams) which, within the frame of urban planning, influences, above all, quality of living environment and localization of activities defined by the so-called visual activities. "It is necessary to preserve conditions of visual comfort during daylight irrespective of the amount of clouds (in indoor residential premises in compliance with their function)" [1]. When designing housing, the values defined by the ČSN 730580 (Czech National Standard) [1] are being observed; these include not only day lighting but also the values of max. hillshade angle of obstacles (i.e. by surrounding buildings) or terrain.

Suitable orientation towards cardinal points in connection with rate of solar insolation has, however, also a very practical effect. In recent years there has been a boom in passive or low energy house construction for which this factor is of vital importance. Most suitable locations for such constructions can be found on south-facing slopes where one can get from 10 to 30% more insolation than on a flatland. Yet a south-facing slope with unfavourable direction of prevailingly cold winds looses this advantage. Formally designed street lines which sometimes do not allow for a sufficient spacing of buildings also decrease the guarantee of a

sufficient amount of insolation of dwelling rooms for the whole life duration of a building. This can increase heating requirements by tens of percents [23].

The general problems of situating buildings are dealt with in more detail e.g. by Jablonská et al. (2010) who analyses suitable orientation of buildings in relation not only to cardinal points but also to surrounding buildings and roads.

## MODELLING OF SPATIAL CONFLICTS IN AN AREA

### Elaboration of Background DEM and Derivation of Morphometric Variables

As mentioned above, the objective of this article is an evaluation of influences of DEM quality, from the point of view of the chosen method of interpolation, on results of modelling of spatial conflicts. At the beginning of the modelling process it was therefore necessary to obtain two DEMs of different quality which would be subjected to subsequent analyses and would allow us to monitor changes in the results (Fig. 1). When using identical input data for the DEMs which are being created, the one which achieves the best results when evaluated by non-spatial indicators of metric accuracy is designated as the high-quality DEM. The DEM which achieves the lowest score is then designated as the so-called low-quality DEM. Therefore the selection of the high-quality and the low-quality DEM is conditioned by the elaboration of a set of DEMs which enter the process of accuracy evaluation.

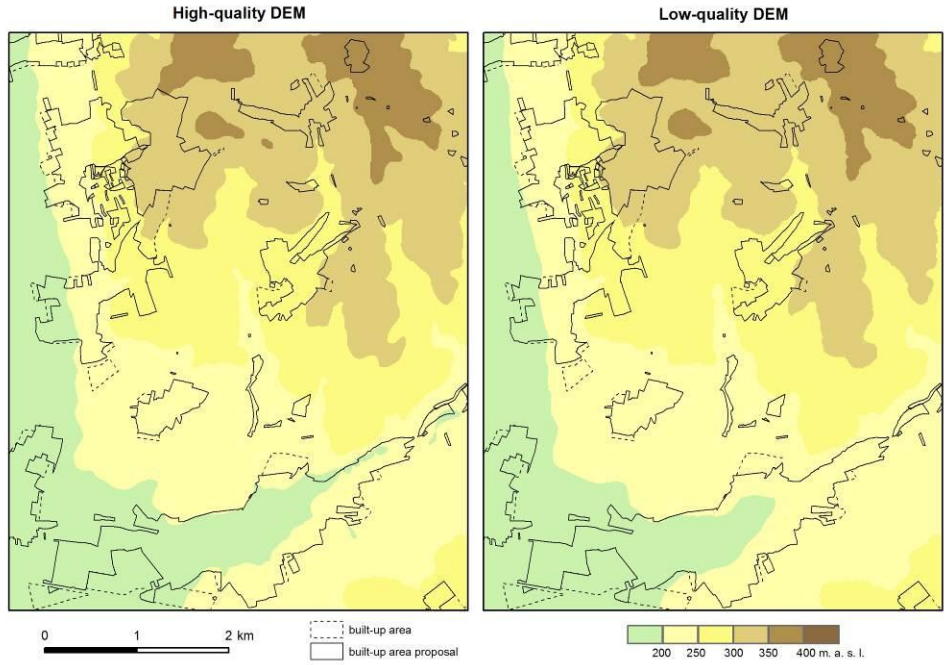



**Fig. 1** Vertical situation of the small part of Olomouc region

A number of publications have already been devoted to the topic of evaluation of DEM accuracy (i.e. [13], [14], [15], [16], [17], [24]). The choice of the most suitable interpolation method and configuration of its variables for the specific types of reliefs were dealt with e.g. by

[25], [26] or by [18]. These publications imply that the method suitable for an elaboration of a digital elevation model (from the DLM 25 contours) of the type of relief defined as flat highland is the interpolation method of kriging employing for instance exponential or spherical models of theoretical semivariogram which stress the influence of the closest entry points values on the estimated values; it is also possible to use the tension spline method with a lower value of the weight variable (e.g. 0,1), which results in a high tension of the resulting surface. According to the aforementioned recommendations and a subsequent evaluation of the DEM's metric accuracy, the method of kriging using the exponential model of theoretical semivariogram and the variable n = 20 points has been employed to elaborate the high-quality DEM of the model area. The low-quality DEM of the given area has been created using the IDW method with the values of power = 2 and n = 2. The setup was also based on the results of the aforementioned publications. The contours from the DLM 25 data set with the equidistance of 5 m were used to create the digital elevation models. The size of pixel has been set to 10 m. Since the selected area has a sufficient density of input data the calculation of non-spatial indicators was sufficient to verify the accuracy of both the models (table 1).

**Table 1.** Values of non-spatial variables of metric accuracy of high- and low-quality DEM

| DEM | Interpolating method (settings) | RMSE | AE | H |
|---|---|---|---|---|
| High-quality | kriging (exponential model) | 1,51 | 1251,39 | 0,44 |
| Low-quality | IDW (power = 2) | 2,50 | 2343,50 | 0,45 |

Accuracy of the quality evaluation of the created DEMs was confirmed also by their spatial visualization using shaded relief (Fig. 2) or by the visualization of the derived morphometric variables – gradient and aspect of slopes (Fig. 3 and Fig. 4).

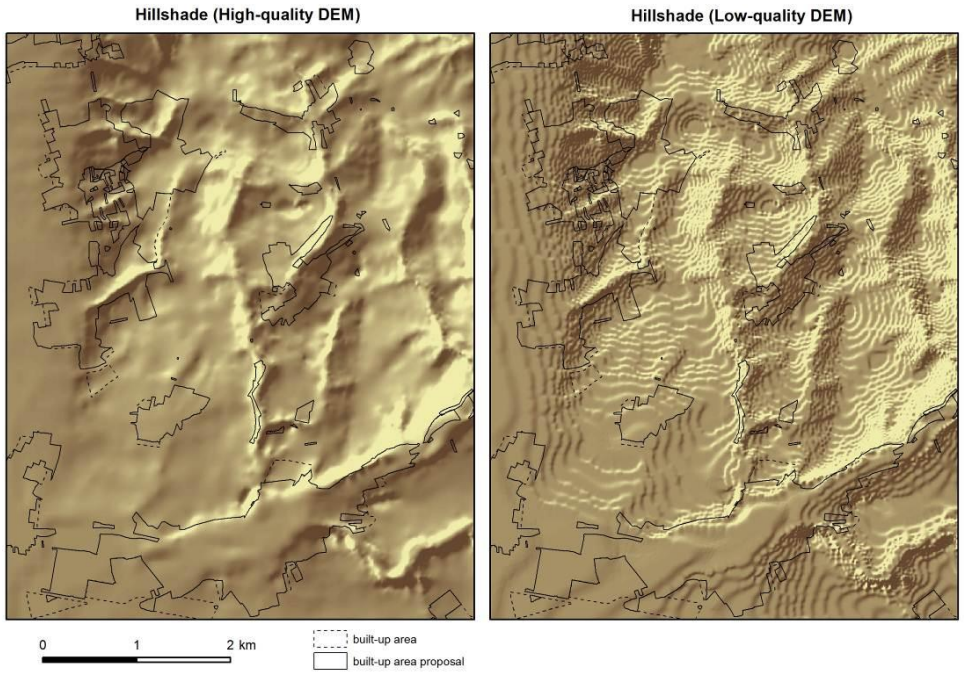



Fig. 2 Shaded relief derived from the HQ DEM and from the LQ DEM The surface of the HQ DEM is smooth whereas in the case of the LQ DEM small terraces occurred on the slopes.

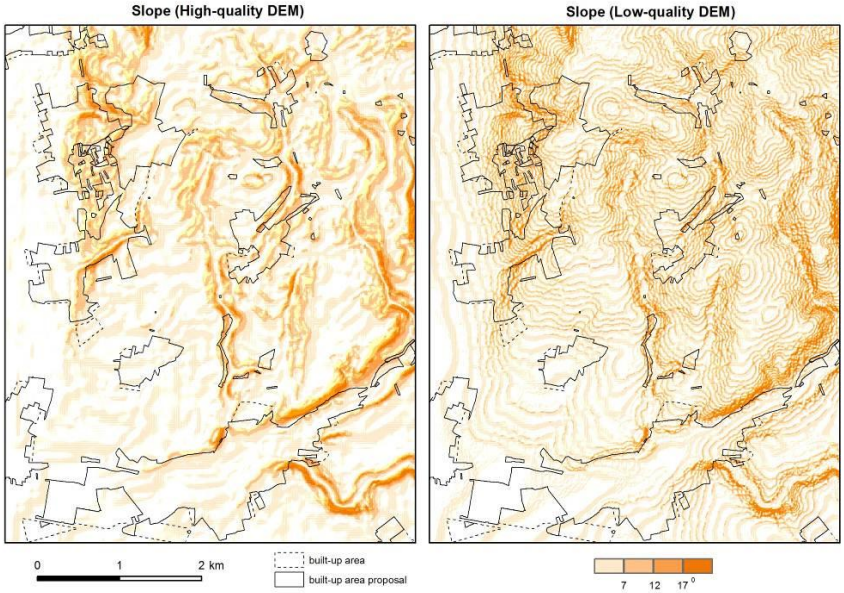



**Fig. 3** Gradient of slope derived from the HQ DEM traces the seamless flow of the background DEM and the individual categories of gradient form compact areas. The behavior of the gradient of slope on the LQ DEM traces the existing terraces. The values of gradient of slopes are therefore underestimated in the flat parts of the terraces and overestimated in their slanting parts.

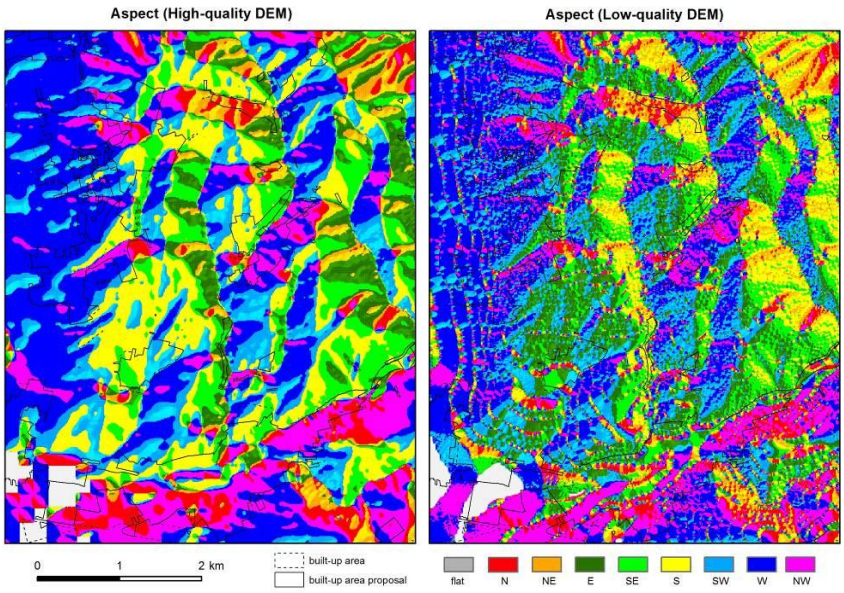



**Fig. 4** Values of the aspect of slopes on the HQ DEM form compact areas. More frequent local changes of the aspects of slopes on the LQ DEM are caused by the concentric arrangement of values of altitude around the entry points; this is a result of interpolation using the IDW method, which causes greater planar curvature

When using both the shaded relief and the visualization of gradients and aspects of slopes, the outputs of the DEM created by the IDW interpolation method show so-called artificial terraces which are caused by unrealistic step changes of values. Gradients of slopes are being underestimated in the flat parts of the terraces and overestimated in the slanting parts. When creating a DEM using the IDW method, the interpolated values often create concentric formations around the entry points; these cause greater curvature of earth's surface in the horizontal direction and therefore more frequent local changes in aspects of slopes.

**Elaboration of Analysis of Slope Aspect and Solar Insolation**

The values of grids of slope gradients derived from the high-quality and the low-quality DEM were divided into 3 categories to suit the requirements of the elaboration of the analysis of conflict with the selected functional areas (residential and recreational). Limit values of gradient are based on limit values of building development which are stated e.g. by [21] or [20].

Subsequently, by employing the zonal histogram function, the output layers were used to establish the pixel frequency of the selected functional area falling within the individual categories of gradient (gradient after reclassification). Via multiplying the pixel frequency by the surface area of one pixel (100 m2), a specific area has been acquired. The use of the zonal histogram function was conditioned by the conversion of the vector layers of the functional areas into a raster grid. Minimum, maximum and average values of gradient of the monitored functional areas were acquired by applying the zonal statistic function to the original grids of slope gradients.

Derivation of the rate of solar insolation was based on the LANDEP methodics (Department of the Environment of the Slovak Republic 2004). The calculation of insolation was based on the values of gradient and aspect of slopes reclassified into categories according to their respective methodics (table 2). A subsequent combination (combine function) of the reclassified grids resulted in the grids of rate of solar insolation. After that the expanse of the selected functional areas falling within the individual categories of rate of solar insolation was established by means of the zonal histogram function.

**Table 2.** Solar insolation rate calculated by combination of aspect and slope (created according to methodical by Ministerstvo životného prostredia SR 2004)

| **Aspect** | **Slope [°]** | | | | |
|---|---|---|---|---|---|
| | **< 3** | **3-7** | **7-12** | **12-17** | **> 17** |
| N | 3 | 3 | 3 | 4 | 5 |
| NE | 3 | 3 | 3 | 4 | 4 |
| E | 3 | 3 | 3 | 3 | 3 |
| SE | 3 | 2 | 2 | 2 | 2 |
| S | 3 | 2 | 2 | 1 | 1 |
| SW | 3 | 2 | 2 | 2 | 2 |
| W | 3 | 3 | 3 | 3 | 3 |
| NW | 3 | 3 | 3 | 4 | 4 |

1 – best solar insolation rate, 2 – good solar insolation rate 3 – average solar insolation rate, 4 – low solar insolation rate 5 – the lowest solar insolation rate

## RESULTS OF MODELLING

### Conflict of Selected Functional Areas with Gradient of Slope

Gradient of slope influences technical and financial exigence of construction of housing and recreational facilities. From the point of view of urban planning and housing development restrictions and control, gradient of slopes can be also used to localize potential dangers of landslides or erosion.

In accordance with the LANDEP methodics, [21] states the limit values of gradient for the following human activities: complex housing development up to 12°, individual housing construction up to 17°, industrial development up to 7°, gardens up to 3°, farm land up to 7°, orchards up to 17°, production of fodder from 7° to 17°, meadows and pasture-lands from 12° to 25°. [20] furthermore states the limit gradient for heavy construction assembly and prevailingly repetitive housing (terraced housing) as 5°. Point housing construction can be, according to [20], carried out up to 12° of gradient (2006); however, problems can be expected concerning construction of utility lines and general access roads. Insolated slopes of up to 15° of gradient can be used for specific types of individual housing construction.

Within the scope of the analysis, the conflict of the values of gradients (derived from both the low-quality and the high-quality DEM) with the functional areas for housing and recreation within the continuous urban development of Potštát was being monitored. The limit values chosen for the analysis are 12° for repetitive and point housing construction and 17° for individual housing construction.

For a typical repetitive or point housing construction, a terrain with a gradient of up to 12° can be used [20]. When using the high-quality DEM, the limit value of gradient of slope was exceeded only by 1,3 % of the current housing developments and by more than 6,9 % of current recreational and sports areas (table 3 and 4). However, when using the low-quality DEM, the extent of the conflicts with gradient of slopes was by 1-3 % higher. The 12° limit of the gradient of slope was, in this case, exceeded by only 1,9 % of the current housing areas and by ca 10,1% of the current sporting and recreational areas (table 4). The reason for these higher values of gradient of slopes while using a low-quality DEM is the presence of small artificial terraces, upper parts of which are almost completely flat (Fig. 2 and 3).

Individual housing construction can be, according to the LANDEP methodics [21], carried out in gradients up to 17°. It is, however, necessary to take into consideration higher technical and financial expenses. The percentage of the current housing and recreational areas which exceed the limit value of 17° is, again, lower when using the high-quality DEM; in case of the functional areas for housing it is only by 0,04%; yet in case of the sporting and recreational areas it is up to 0,9% (table 3 and 4). In this case it is therefore necessary to verify the type of recreational utility and decide, whether there is a conflict or whether it is an intention.

**Table 3.** Areas of slope categories for housing (high- and low-quality of DEM)

| Aspect [°] | Housing – state | | | | | | Housing – proposal | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HQ | | LQ | | difference | | HQ | | LQ | | Difference | |
| | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % |
| 0,0-7,0 | 239047 | 93,02 | 235945 | 91,8 | -3102 | -1,2 | 78112 | 95,9 | 71800 | 93,2 | -6312 | -2,7 |
| 7,1-12,0 | 14517 | 5,65 | 15970 | 6,2 | 1453 | 0,6 | 2934 | 3,6 | 4399 | 5,7 | 1465 | 2,1 |
| 12,1-17,0 | 2749 | 1,07 | 4209 | 1,6 | 1460 | 0,6 | 339 | 0,4 | 747 | 1 | 408 | 0,5 |
| > 17,0 | 668 | 0,26 | 857 | 0,3 | 189 | 0,1 | 74 | 0,1 | 91 | 0,1 | 17 | 0,1 |

**Table 4.** Areas of slope categories for sport and recreation (high- and low-quality of DEM)

| Slope [°] | Sport and recreation – state | | | | | | Sport and recreation – proposal | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HQ | | LQ | | difference | | HQ | | LQ | | Difference | |
| | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % |
| 0,0-7,0 | 29337 | 74,4 | 29055 | 73,6 | -282 | -0,71 | 12826 | 91,9 | 12602 | 90,34 | -224 | -1,6 |
| 7,1-12,0 | 7377 | 18,7 | 6390 | 16,2 | -987 | -2,50 | 1009 | 7,2 | 1045 | 7,49 | 36 | 0,3 |
| 12,1-17,0 | 2149 | 5,4 | 3049 | 7,7 | 900 | 2,28 | 106 | 0,8 | 273 | 1,96 | 167 | 1,2 |
| > 17,0 | 592 | 1,5 | 961 | 2,4 | 369 | 0,94 | 9 | 0,1 | 30 | 0,22 | 21 | 0,2 |

Even the areas suggested for housing, sports and recreation are not entirely conflict–free concerning their gradient. Aprox. 0,5 % of the suggested housing areas and 0,9 % of the sporting and recreational areas exceed the limit value of 12° when using the high-quality DEM. The percentage of areas exceeding this limit value when using the low-quality DEM is again higher, which is caused by the occurrence of numerous artificial terraces along contours. The percentage of suggested functional areas which would exceed the limit value of 17° is absolutely negligible.

### Conflict of Selected Functional Areas with Rate of Solar Insolation

A sufficient rate of solar insolation is one of the housing area health requirements which are necessary to provide a healthy environment for habitation (see [1] or [2]). Czech National Standard 730580-2 [1] focuses primarily on establishing a level of daylighting in interiors. Besides that it also defines values of max. hillshade angle of obstacles or terrain for different types of building developments. The nature of developments in the model area can be generally defined as “standard areas of permanent residence“ or "areas of permanent residence in a continuous repetitive developments in city centres" in which the maximal permitted hillshade angle is 30° and 36° respectively. From this point of view, it is appropriate to place the current and the suggested functional areas for housing or sports and recreation in localities with a higher value of angle of incidence of sunbeams, i.e. at least in the category of low rate of solar insolation, or preferably in the category of medium insolation so that, even in case of hillshade by a new obstacle or terrain formation, there is a sufficient amount of direct sunlight in the particular area.

Within the frame of the continuous development of the Olomouc region no significant conflicts arose between the minimal rate of solar insolation and the current or the suggested functional areas designated for housing or sports and recreation while using both the high-quality and the low-quality background DEM (Fig. 5). The percentage of the functional areas being monitored within the scope of this category is null or completely negligible (table 5 and 6).

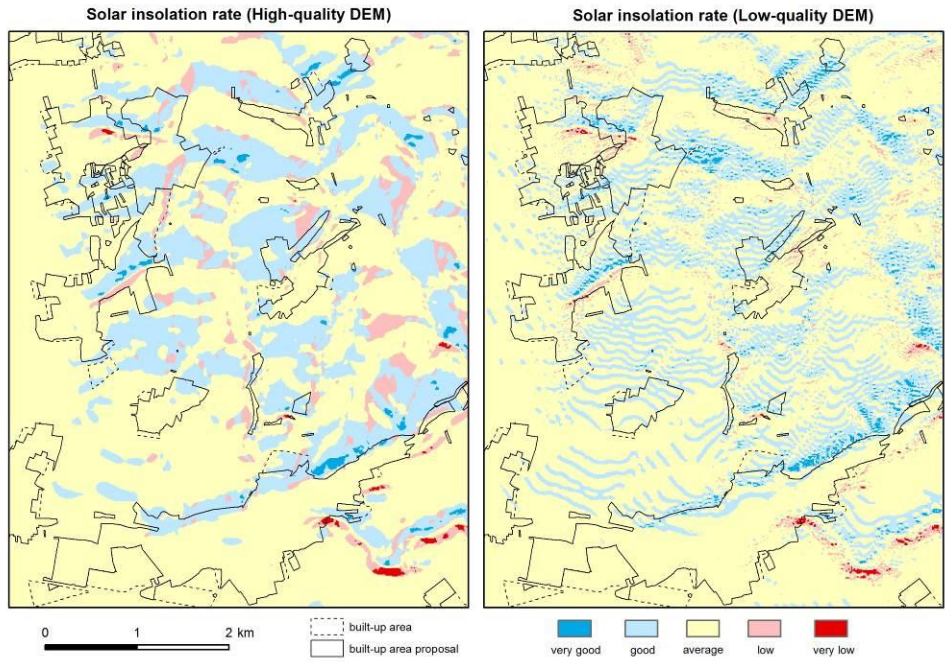



**Fig. 4.** Values of the solar radiation rate.

**Table 5.** Areas of solar insolation rate categories for housing (high- and low-quality of DEM)

| Solar insolation rate | Housing – state | | | | | | Housing – proposal | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HQ | | LQ | | Difference | | HQ | | LQ | | difference | |
| | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % |
| 1 – very good | 273 | 0,1 | 628 | 0,2 | 355 | 0,1 | 133 | 0,2 | 161 | 0,2 | 28 | 0 |
| 2 – good | 17323 | 6,7 | 19503 | 7,6 | 2180 | 0,9 | 6207 | 7,6 | 7316 | 9 | 1109 | 1,4 |
| 3 – average | 233567 | 90,9 | 235190 | 91,5 | 1623 | 0,6 | 73854 | 90,6 | 73597 | 90,4 | -257 | -0,3 |
| 4 – low | 5763 | 2,2 | 1594 | 0,6 | -4169 | -1,6 | 1237 | 1,5 | 355 | 0,4 | -882 | -1,1 |
| 5 – very low | 55 | 0,1 | 66 | 0,1 | 11 | 0 | 28 | 0 | 30 | 0 | 2 | 0 |

**Table 6.** Areas of solar insolation rate categories for sport and recreation (high- and low-quality of DEM)

| Solar insolation rate | Sport and recreation – state | | | | | | Sport and recreation – proposal | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HQ | | LQ | | difference | | HQ | | LQ | | difference | |
| | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % | [ha] | % |
| 1 – very good | 501 | 1,3 | 783 | 2 | 282 | 0,7 | 987 | 7,1 | 1148 | 8,2 | 161 | 1,1 |
| 2 – good | 9850 | 25 | 7800 | 19,8 | -2050 | -5,2 | 0 | 0 | 13 | 0,1 | 13 | 0,1 |
| 3 – average | 27346 | 69,3 | 30056 | 76,2 | 2710 | 6,9 | 12501 | 89,6 | 12650 | 90,7 | 149 | 1,1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 – low | 1662 | 4,2 | 737 | 1,9 | -925 | -2,3 | 462 | 3,3 | 137 | 1 | -325 | -2,3 |
| 5 – very low | 96 | 0,2 | 79 | 0,2 | -17 | 0 | 0 | 0 | 2 | 0 | 2 | 0 |

As mentioned above, a housing development should be placed at least in the category of low insolation, or preferably in the category of medium insolation. Observing only the minimal health requirements for a housing area is not the best solution; therefore, within the frame of the process of urban planning, a housing or recreational development in an area of low insolation should be considered problematic and conflicting. The influence of either the high-quality or the low-quality background DEM on the resulting values (expanse of the conflicting areas) is, within this category, much more significant. When using the high-quality DEM to derive the rate of solar insolation of the relief, 2,2 % of the current housing areas and 4,2 % of the current recreational and sporting areas were detected as having low insolation. When using the low-quality DEM, the percentage of the current areas for housing or sports and recreation which are in conflict with low insolation decreases significantly lower values (i.e. 1,6 % and 2,3 %). The reason probably being the terrace-like structure of the low-quality DEM elaborated by means of the IDW method, which enables incidence of sunbeams also on northwest, north or northeast facing slopes under a more accurate angles. A conflict with the areas of low rate of solar insolation occurs also within the scope of the suggested areas for housing, sports and recreation. Here too the percentage of conflicting areas is higher when using the high-quality DEM. The expanse of the suggested areas is, however, so small that an evaluation of the differences between the influences of the low-quality and the high-quality DEM has almost no practical value.

## CONCLUSION

The primary objective of this article is to evaluate the influences of quality of background DEM (from the point of view of the chosen method of interpolation) on results of modelling of specific spatial conflicts within the frame of urban planning and thus point out potential problems with delimiting necessary measures or further urban development planning. To elaborate the examples, modelling of conflicts of functional areas for housing, sports and recreation with gradient of slopes and rate of solar insolation of slopes was chosen. These aspects, from the perspective of quality and comfort of life, can be considered of an utmost importance. Olomouc region was chosen as the model area.

To evaluate the influence of the background DEM it was necessary to construct so-called high-quality and low-quality DEM. When using identical input data for the DEMs being created, (in this case contours from the DLM 25 data set) the one which achieved the best results being evaluated by non-spatial indicators of metric accuracy was designated as the high-quality DEM. The DEM which achieved the lowest score was then designated as the so-called low-quality DEM. The high-quality DEM for the model area was created through the use of the method of kriging using the exponential model of theoretical semivariogram. The low-quality DEM of the area was elaborated using the IDW method with the weight of $p = 2$. Subsequently, the background DEMs were used to elicit morphometric variables which were later used in the process of modelling the spatial conflicts. The gradient of slopes was a direct

part of the modelling process. It was then combined with the aspect of slopes (according to the LANDEP methodics) to derive the rate of solar insolation.

The results of the elaborated analyses show that in case of employing a high-quality DEM one can always identify places with critical values of a phenomena being monitored (conflicts of functional areas for housing or recreation with values of gradient exceeding 17°, conflicts with a low rate of solar insolation) because such a model creates compact areas. However, when using a low-quality DEM a “fragmentation” occurs and these critical (monitored) values create admittedly larger (in the range of several pixels) but more numerous areas; moreover, they are scattered over a larger area. When using a low-quality background DEM, determination of the necessary measures or planning a further development of a town can become problematical because it is impossible to define “conflicting” parts of a municipality unambiguously. Moreover, because of the scattering of the conflicting places, the whole area of a municipality can become conflicting, which makes any further planning more difficult.

## REFERENCES

[1] Czech technical standard 73 0580-2 (730580) (2007) Daylighting in buildings - Part 2: Daylighting in Residential Buildings. Czech normalization institute, Praha.

[2] Hynvar, V. et al. (2011) Land use limits. http://www.uur.cz/default.asp?ID=2591, [cit. 2011-06-13].

[3] Dai, F. C., Lee, C. F., Zhang, X. H. (2001) GIS-based geo-environmental evaluation for urban land-use planning: a case study. Engineering geology, 61, 257-271.

[4] Vozenilek, V. (2002) Radon risk assessment in GIS. NATO Science series IV: Earth and environmental science, 10, 97-107.

[5] Montz, B. E., Tobin, G. A. (2008) From false sense of security to residual risk: communicating the need for new floodplain development models. Geografický časopis. 60, 1, 3-14.

[6] Gallay, I. (2010) Simple water flow modelling and its use in evaluation of landscape vulnerability. Geografický časopis, 62, 2, 109-125.

[7] Pechanec, V., Burian, J., Kilianova, H., Nemcova, Z. (2011) Geospatial analysis of the spatial conflicts of flood hazard. Moravian Geographical Reports, 11-19.

[8] Krcho, J. (2001) Modelling of georelief and its geometrical structure using DTM: positional and numerical accuracy. Q111 Publishers, Bratislava.

[9] Goodchild, M. F. (2010) Spatial analysis and modeling. In J.D. Bossler, editor, Manual of Geospatial Science and Technology, 575–592. Boca Raton: CRC Press..

[10] Li, Z., Zhu, Q., Gold, Ch. (2005) Digital terrain modeling: principles and metodology. CRC Press.

[11] Longley, P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (2001) Geographic information systems and science. Chichester.

[12] Fencik, R., Vajsablova, M. (2006) Parameters of interpolation methods of creation of digital model of landscape. Proceeding of the 9th AGILE Conference on Geographic Information Science. Visegrád, Hungary, 374-381.

[13] Desmet, P. J. J. (1997) Effects of interpolation errors on the analysis of DEMs. Earth surface processes and landforms, 22, 569-580.

[14] Lopez, C. (2000) Improving the Elevation Accuracy of Digital Elevvation Models: A Comparison of Some Error Detection Procedures. Transactions in GIS, 4, 1, 43-64.

[15] Krcho, J. (2001) Modelling of Georelief and its Geometrical Structure Using DTM – Positional and Numerical Accuracy. Bratislava (Vydavateľstvo Q111).

[16] Hofierka, J., Cebecauert, T., Suri, M. (2007) Optimisation of Interpolation Parameters Using a Cross-validation. Digital Terrain Modelling, Development and Applications in a Policy Support Environment, Series: Lecture Notes in Geoinformation and Cartography, Springer, 67-82.

[17] Li J., Heap A.D. (2008) A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, 23, Canberra.

[18] Svobodova, J., Tucek, P. (2009) Creation of DEM by kriging method and evaluation of the results. Geomorphologia Slovaca et Bohemica, 9, 1, 53-60.

[19] Wilson, J. P., Gallant, J. C., eds. (2000) Terrain Analysis: Principles and Applications. John Wiley & Sons.

[20] Gates R. L., Stout, F. (2004) The Sustainable Urban Development Reader. Routledge London & New York.

[21] Ruzicka, M., Miklos, L., (1982) Landscape-ekological planing (LANDEP) in the process of the teritorial planing. Ekológia (ČSSR), vol. 1, No. 3, 297 – 312

[22] Bracken, I. (2007) Urban planning methods. Routledge.

[23] Yan, J., Stellios, P. (2006) Design for Sustainability. Beijing: China Architecture and Building Press.

[24] Jedlicka, K. (2009) Accuracy of surface models acquired from different sources - important information for geomorphological research. Geomorphologia Slovaca et Bohemica, 9,1, 17-28.

[25] Svobodova, J. (2011) Selection of appropriate interpolation methods for creation DEMs of various types of relief by complex approach to assessment of DEMs. In Proceedings of Symposium GIS Ostrava 2011 [CD-ROM]. VŠB – TUO, Ostrava.

[26] Svobodova, J. (2011) Quality assessment of digital elevation models for environmental applications (in Czech). Ph.D. thesis, Department of Physical Geography and Geoecology, Faculty of Science, University of Ostrava. Ostrava.

# FUZZY SURFACE MODELS BASED ON KRIGING OUTPUTS

Jan CAHA[1], Pavel TUCEK[2], Alena VONDRAKOVA[3] and Lenka PACLIKOVA[4]

Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc,

tr. Svobody 26, 771 46, Olomouc, Czech Republic

*jan.caha@upol.cz[1], pavel.tucek@upol.cz[2], alena.vondrakova@upol.cz[3], lenka.paclikova01@upol.cz[4]*

**Abstract**

Surface and its analyses are important topic in geosciences. Surface models are obtained from data with uncertainty in both position and attribute. The procedure which is used for modelling the surface from the original data results in biased outputs. Fuzzy set theory and fuzzy logic dispose of methods and tools for modelling surfaces based on biased input data. Especially fuzzy numbers are suited for modelling surfaces with uncertainty.

Kriging has been proved to be one of the best interpolation methods and it allows calculation of the standard deviation of the output surface. Based on those inputs, fuzzy surface can be constructed. The fuzzy-surface represents each point of the grid as a fuzzy number. Triangular and trapezoidal fuzzy numbers are the most common ones that are used in applications because of their easy implementation. The kernel of the fuzzy number is the result given by the kriging method and the support of the fuzzy number is calculated from the standard deviation of the kriging method. The output surface represents the interpolation of the data with the uncertainty that was present in the original data as well as the uncertainty that arises from the interpolation procedure. The aim of this paper is the creation of fuzzy surface based on the results of kriging calculation.

Fuzzy surface can be further used in geosciences for analyses of situations where the uncertainty of the result is important for decision making. Knowledge of uncertainty in calculations also allows much better risk management and provides more information for better crisis management.

**Keywords: fuzzy, surface, kriging, fuzzy number, uncertainty**

## INTRODUCTION

All types of surfaces and their analyses have an important role in geosciences [1]. Often they are used as error free models even if significant errors and uncertainty can be and usually is present in those surfaces. Most of the time surfaces are created from data that do not cover the whole area of interest. Interpolation methods are used to create surface from discrete data. Creation of surface through different interpolation methods can lead to significantly different results. For necessary evaluation of the precision and quality of the result surface there are developed several methods that are widely used. These techniques include calculations of different kinds of errors and indexes that show surface quality. The most common known and used method is calculation of root mean square error (RMSE), next are absolute error and Hammock index [1]. Currently the uncertainty in the surface estimation is a topic that has the same importance as the surface quality. There are several ways how the

uncertainty is introduced to the surface; however two of them are the most important: uncertainty in dataset and uncertainty in the interpolation process [2]. There exists also several methods for handling and propagation of uncertainty such as interval arithmetic, Bayesian statistic, Dempsey-Schaffer methods and fuzzy sets and fuzzy logic [3], [4].

Fuzzy sets and logic become widely used theory for handling uncertainty in various disciplines. Most likely it is because of relatively simple concept and the fact that it is very close to the style of human thinking. People unlike machines or mathematical theories do not think in exact values but rather in vague terms. Fuzzy theory allowed using of vague terms in reasoning as well as in calculations. Because of this fuzzy theory became one of the most popular theories for handling uncertainty.

This contribution is related to the approach for creating fuzzy surfaces from the results of interpolation of input data by kriging method. Existing methods used in geosciences for the interpolation have been extended for use on fuzzy data, so there exists fuzzy IDW, fuzzy spline and fuzzy kriging methods [2], [5]. However these methods are not yet implemented in GIS or in mathematical software such as R, Octave or Scilab. Also named methods are intended for use on fuzzy input data. In practical applications such data are not common. But there can be utilized methods of existing techniques of estimation provided by kriging algorithm and combined with expert opinion to create a fuzzy surface.

## FUZZY SET THEORY

Nowadays it is well known that uncertainty is present in almost every information [3]. According to [3], [6] and [7] a lot of phenomena are not statistical by nature and thus probability theory is not well suited for handling their uncertainty. Fuzzy set theory and fuzzy logic were developed as tools for analytical solving of problems that are not suited for probability theory and classic logic [8]. Main purpose of fuzzy theory can be described as precise description of imprecision (or vague) objects or phenomena [9]. Today fuzzy theory is used in wide area of disciplines for handling different data and processes that contain uncertainty.

Fuzzy set theory and fuzzy logic were first introduced by L. A. Zadeh in 1965 [8]. A fuzzy set is a collection of ordered pairs of objects and their membership grades. According to [2] the fuzzy set is defined as:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x))\} : x \in U \qquad (1)$$

Where Ã denotes a fuzzy set, U is a universe on which Ã is defined, x is object from U and μÃ(x) is a degree membership of x to the fuzzy set Ã. Such fuzzy set is characterized by the membership function (fÃ (x)) which associates each x to its degree of membership – number from interval [0,1]. The closer is the membership value to the 1 the more is the object part of the fuzzy set. Value 1 indicates complete membership to the set and value 0 means that the object does not belong to the set at all. Fuzzy set theory is generalization of classic (crisp) set theory, where $x \in A$ or $x \notin A$ as those are the extreme cases of membership degrees associated with values of 1 and 0 [Fig.1]. Fuzzy logic utilizes similar concept using degrees of

truth as a measure of correctness of the predicament. Some important terms connected with fuzzy set theory are:

- kernel – set of all x where $\mu\tilde{A}(x) = 1$
- support – set of all x where $\mu\tilde{A}(x) > 1$
- α-cut – set of all x where $\mu\tilde{A}(x) \geq \alpha$ for $\alpha \in [0,1]$

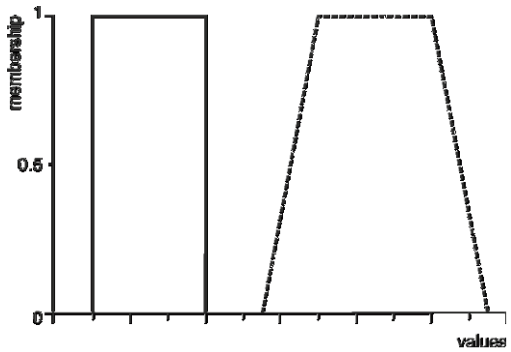



Fig. 1 Crisp set and fuzzy set

It is important to note, that all α-cuts, which special cases are both kernel and support, are crisp sets. This is important for constructing and calculating with fuzzy sets. For details about fuzzy set theory and fuzzy logic see [8].

**Fuzzy Number**

Fuzzy numbers are considered as special cases of fuzzy sets. Fuzzy number is normalized convex fuzzy set used to represent vague value or number. Different types of fuzzy numbers exist i.e. Gaussian, Triangular, Trapezoidal, Piecewise Linear, etc. [fig.2] [2], [5], [10], [14], [16]. Another terminology uses notion fuzzy number for triangular fuzzy number and fuzzy interval for trapezoidal fuzzy number [17], [18], [19]. Because of simple implementation of both definition and calculations the triangular and trapezoidal fuzzy numbers are the most common in geosciences applications [2].

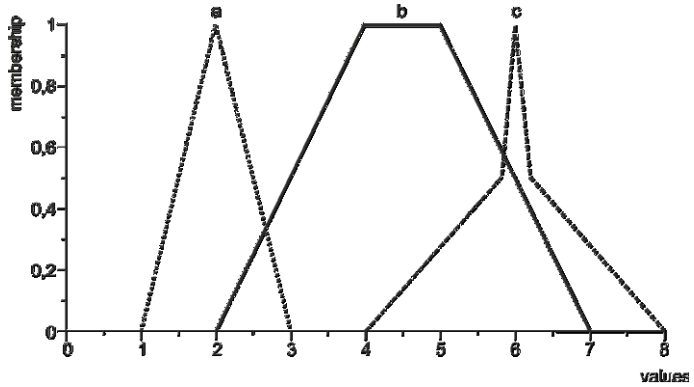



Fig. 1 Fuzzy numbers a) triangular b) trapezoidal c) piecewise linear

Basic triangular fuzzy number is often defined as a triplet [a-,a0,a+] in the same way the trapezoidal number can be defined as a quaternion [a-,a0-,a0+,a+]. Piecewise linear numbers can be defined by ordered pair values and their membership value or for easier computational treatment as set of α-cuts. Each α-cut is represented by interval which allows easier computational operations with such fuzzy numbers [10].

Fuzzy numbers can be used for calculations since all arithmetic operations can be defined for them through the extension principle [7]. However using extension principle was proven to be computationally complicated and methods for calculations with fuzzy numbers using α-cuts were developed [10]. Fuzzy arithmetic is important since it allows propagation of fuzzy numbers through any mathematical operation.

### Fuzzy Surface

As a fuzzy surface can be defined surface that instead of using crisp values uses fuzzy numbers. Such surface has included uncertainty since for each location is possible range of values. This is consistent with so called possibilistic fuzzy set theory [3], [7].

## CONVERTING RESULTS OF KRIGING TO FUZZY SURFACE

### Kriging

Kriging is a set of geostatistical methods used to predict value of variable at location where it was not measured from the set of nearby points where the value of variable is known. Input for kriging is so called random field, that presents points where the variable was measured. For interpolation over the whole area of interest the regularly spaced grid of points is created and the prediction is made for each of those points [1]. Algorithm used by kriging is a linear least square estimator because it minimizes variance of the prediction error.

There are several types of kriging and the most common are simple, ordinary and universal kriging. Each of these methods have different assumptions: simple kriging presumes known constant trend of the variable, ordinary kriging presumes unknown constant trend and universal kriging assumes polynomial trend [11]. Before choosing which method is the most

suitable to use statistical analysis that searches for trends in data have to be performed. However universal kriging seems to be the best choice since most of the geographical data contain some sort of trend [5].

Generally the universal kriging is defined [11]:

$$Z(x) = m(x) + \varepsilon(x) \quad (2)$$

Z(x) is estimation of variable Z at the point x, m(x) is a structural component associated with the trend and ε(m) is regionalized variable which is stochastic and spatially correlated. Universal kriging estimator is defined:

$$Z^*(x) = \sum_{i=1}^{N} \lambda_i Z(x_i) \quad (3)$$

where N is number of neighbors and λ is a vector of kriging coefficients. Results of universal kriging are obtained after minimalization of function:

$$\sum_{j=1}^{N} \lambda_j \gamma_\varepsilon(x_1 - x_1) + \sum_{l=1}^{P} \mu_l f_l(x_i) \quad (4)$$

in the equation γ is a vector of variogram values, μ is vector of Lagrange multipliers and f(x) is a vector of values at sampled locations.

**Kriging Standard Error**

Same as many other statistical methods kriging allows besides estimation of value that have to be predicted also estimation of standard error of this prediction.

In [12] and [5] was proven that kriging standard error has no direct connection to the value that kriging is predicting, while the only connection is to the distance of the neighbour measure points. That is consistent with the kriging definition. The uncertainty (standard error) is higher if the points used for prediction are further away from the prediction location. If the neighbour points are close to the point of estimation the uncertainty is smaller. If the measurement points are located in regular grid the standard error prediction repeats for each part of the grid without any connection to the local variance of data in the grid [12] [fig.3].

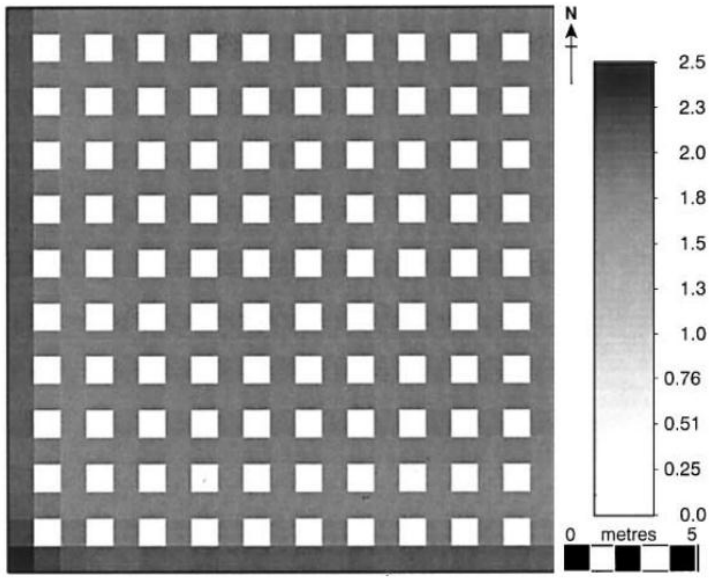



**Fig. 3** Universal kriging standard error for regular grid of input data (source [12])

The value of standard error depends on the type of semivariogram model used for kriging. For different semivariograms the prediction of standard errors should look different, values vary significantly however the general trend that standard error is dependent on distance to measurement points is still significant [12].

**Fuzzification of Kriging Standard Error**

It was defined that the standard error of kriging has no direct connection to the data since it is only function of distance to sampling points and covariance function [13]. This provides important information about the amount of uncertainty in the predicted surface, however this information is still not enough to allow a direct fuzzification of the predicted value to create a fuzzy surface.

All interpolating methods are based on presumption that value of interpolated phenomena is exactly known at set of exact locations. The resulting surface is then often treated as containing no uncertainty in its predicted values [2]. Several studies suggested how information sampled at discrete locations should be fuzzified for further analytical procession in the form of surface that contains uncertainty. In [14] where contour lines are used as a data source it is suggested that in 90% of cases the error in the input data should be smaller than the half of contour lines interval. This error is directly propagated from input data to the surface that was created. Based on this fact triangular number is constructed for each grid cell of the surface according to the equation:

$$\tilde{A} = \left[x_{ij} - \frac{ci}{4}, x_{ij}, x_{ij} + \frac{ci}{4}\right] \quad (5)$$

$x_{ij}$ is predicted value of the surface at coordinates *ij* and *ci* is a contour interval. The output fuzzy number has support range equal to the half of contour lines interval.

Study [15] suggest using value of standard error - σ 0.2, 0.5, 1 and 2 meters as optimistic, two realistic and pessimist presumptions of surface vertical error created from contours lines with density of 1 meter. So in this attitude 95% of data are contained in interval denoted by:

$$I = [x_{ij} - 2\sigma, x_{ij}, x_{ij} + 2\sigma] \quad (6)$$

Another source [16] suggests creating of a triangular fuzzy number from the crisp input data by this formula:

$$\tilde{A} = [x_{ij} - \varepsilon, x_{ij}, x_{ij} + \varepsilon] \quad (7)$$

Where ***ε*** is the error of the measurement. This variant is based on uncertainty in data before the interpolation process, and the uncertainty is equal to the maximal possible error of measurement. The interpolation process is repeated three times to create surfaces: $a^-$,$a^0$,$a^+$. In such case the surface $a^-$ represents the lowest possible surface of variable to predict, $a^0$ the most possible one and $a^+$ the highest possible surface.

The first and second study show similarities as the formulation about 90% of data with error lower then half of the interval size is close match with σ 0.2 for 1 meter density of contour lines, in this case 95% of data has error lower then 40% of the interval size. In second study this serves as an optimistic presumption.

Compromise solution from those studies is to assume that the maximal error in the surface estimation is equal to the measurement error in the input data. This only stands for estimations made by such algorithms that never exceeds the data or at least does not exceed them by much. So this presumption can not be used for spline without tension because spline algorithm tends to exceed the data range in many situations but spline with tension does not exceed the range of data so this presumption can be made for this algorithm. Based on these three studies can be identified the assumption that results of kriging prediction have the maximal possible error equal to the measure precision.

According to the [15] and [16] when assumption about error in data can be made it can be used for the creation of fuzzy surface. However the outlined principle does not incorporate uncertainty in the surface prediction, the assumption of uniformly distributed uncertainty over the surface is made. However kriging provides information about the distribution of uncertainty over the predicted surface. We use feature normalization to rescale the kriging standard error to interval of values [0,1]. Values of 1 indicate the areas with high uncertainty in estimation of the surface while values of 0 indicate areas with low uncertainty.

Based on these presumptions there can be created fuzzy surface that has at each point fuzzy number constructed by this equation:

$$\tilde{A} = [x_{ij} - (E_d + \varepsilon_{ij} * E_e), x_{ij}, x_{ij} + (E_d + \varepsilon_{ij} * E_e)] \quad (8)$$

where $E_d$ is a minimal estimation error that we assume, $E_e$ is an error that arise from the uncertainty of prediction. $\varepsilon_{ij}$ denotes the normalized standard error of kriging at location *ij*. Together $E_d$ plus $E_e$ should be equal to the half of maximal possible error in input data, so that the range of the support of the fuzzy number is equal to the maximal possible error.

## CASE STUDY – RUSAVSKÁ HORNATINA MOUNTAINS

Area Rusavská hornatina mountains is located on the east side of Czech republic [fig.4]. It is a part of Hostýnské vrchy mountains. The area of interest is a square of extent 4x4 kilometres. Input data set were contour lines with density of 5 meters. This quality of altitude data is usual in the Czech Republic.


**Fig. 4** Location of the area of interest

Model area was previously studied and by several testing methods the optimal interpolating algorithm was determined. For Rusavská hornatina mountains the best algorithm with best evaluation was universal kriging. As the most correct setting kriging with second order trend removal with spherical theoretical semivariogram with 20 neighbours was identified. The reason for using spherical semivariogram is the need of preserving the local variability of the surface. Spherical or exponential semivariograms fulfil this need while gaussian preserves more of global trend then local variability. The surface was first created for area 4.1x4.1 km and later just area of interest of size 4x4 km was extracted to avoid errors that occur during interpolation process near the edges of data.

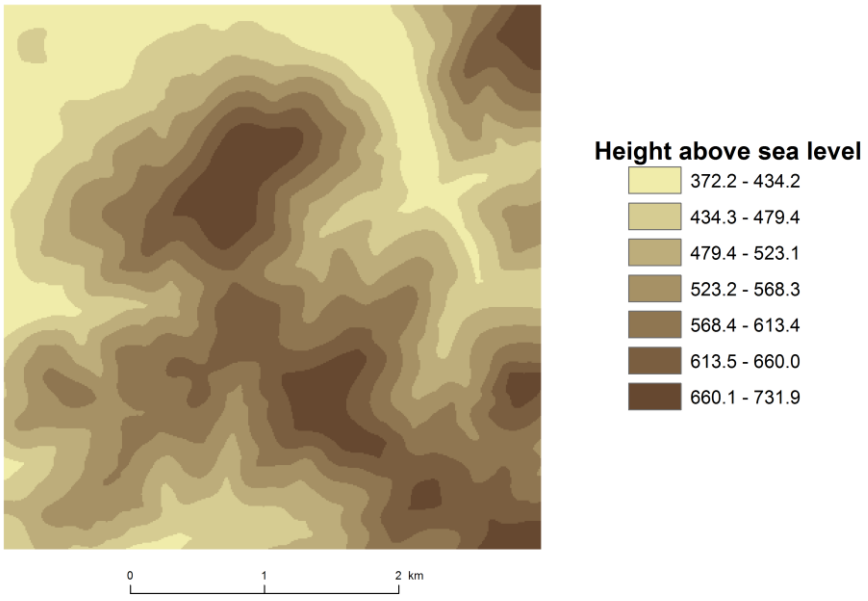



**Fig. 5** Prediction of the terrain elevation

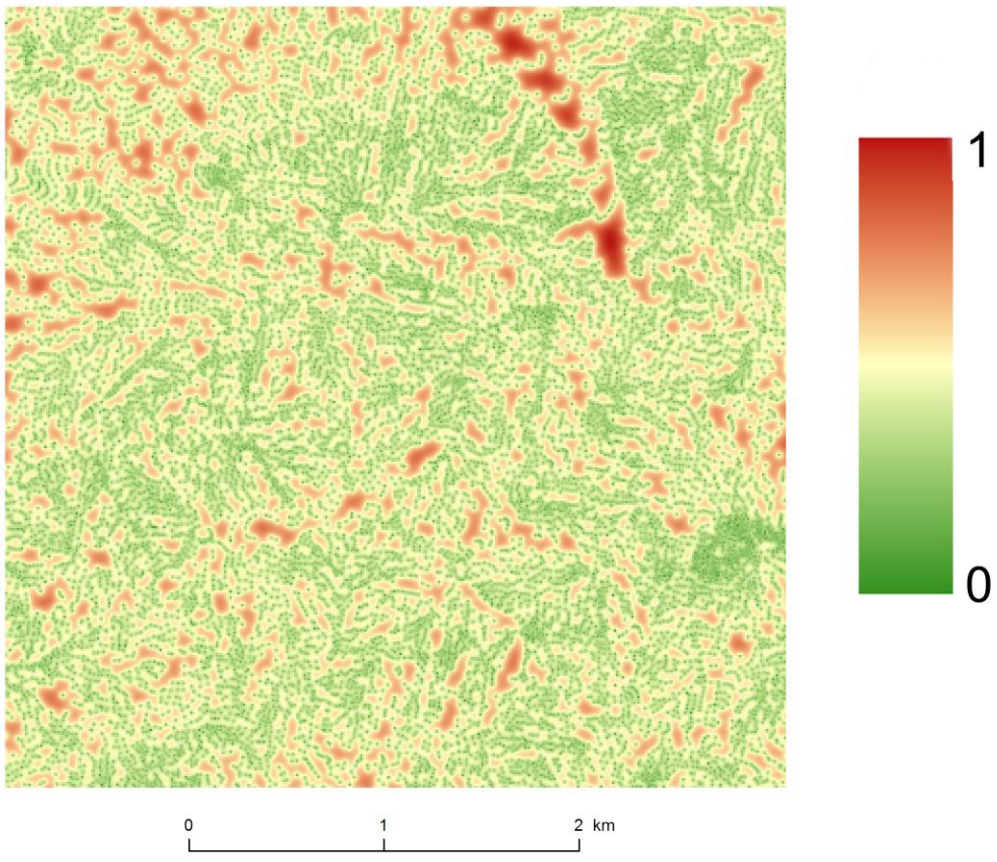



**Fig. 6** Normalized standard error of kriging estimation

It is possible to use standard error in the estimated surface to point out areas that have higher uncertainty in estimation [fig.6]. Values of standard error have normal distribution so their normalization can be done through linear scaling to unit range by the equation:

$$x_n = \frac{x - min(X)}{max(X) - min(X)} \quad (9)$$

where $x_n$ is normalized value of *x, X* is vector of all *x*. Then fuzzification of the result is realized by use of equation [8]. As Ed value of 0.5 meter was chosen and Ee is equal to 0.75 meters, together this gives range of the fuzzy number 2.5 meters as the worst case situation. Result can be visualized in 3D form, where the axis x,y corresponds to the location. Axis z shows estimated value at location which is also the value that has membership degree equal to 1 in the resulting fuzzy surface. Colour shows the maximal possible error in estimation of value at each location [fig.7].

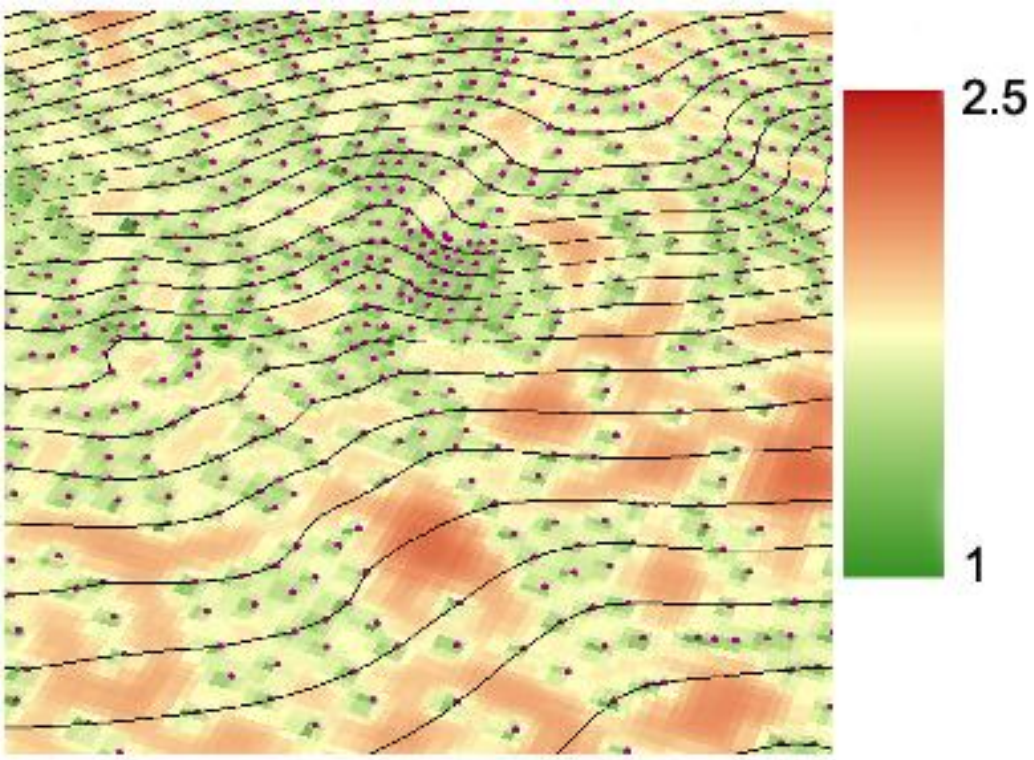



**Fig. 7** Small area of 3D visualization showing input points and their relation to the resulting uncertainty (units are meters)

## DISCUSSION

With the development of modern technologies, measurements are more precise and the effort of scientists is to make the presentation of these data in the most precise form. Therefore it is important in addition to the actual results also present the accuracy of this result. In the field of

surface modelling it is mainly the uncertainty in the interpolation process. Current technologies do not allow efficient visualization of fuzzy surfaces, so it is necessary to find suitable alternatives. Possible way how to provide the most accurate information is presenting of information of surface confidence or possible erroneous of surface based on fuzzy model.

Presentation of possible uncertainty of the model is very important in every discipline such as monitoring of landslides. If the information is low quality because the derived model has a high uncertainty proposed measure can be completely inappropriate. This can occur in a situation where the points in the specific area did not occur at all and the surface model is only the result of interpolation of points more distant. Conversely, if there are many entry points on the measured site, the potential uncertainty of the model is low and the surface model can be used for detailed studies. Example in [fig.7] presents a situation where are entry points of surface model in a large number and perhaps the model uncertainty is very low in this area and very far can be location with very low number of entry points and the model uncertainty can by high and possible error can reach the maximum values.

In various models of natural processes there occurs uncertainty or vagueness arising from the very nature of the monitored phenomenon. Therefore it is very important to avoid bringing an error into other possible calculation. This error can be based on the fact that the model is considered as error free. Such an error could negatively affect the results of each study.

## CONCLUSION

The proposed method allows creation of fuzzy surface containing uncertainty based on output of kriging interpolation method and expert knowledge. Creation of such surface is not computationally demanding and all the necessary methods are implemented in common GIS software. The output fuzzy surface can be further used in operations where the calculation of uncertainty is an important part of the decision making process. Derivative characteristics can be created from such surface and uncertainty of the surface can be propagated to these derivatives. This allows much better treatment of uncertainty during following analysis of both surface and its derivatives.

## REFERENCES

[1] Burrough, P. A. and McDonnell, R. (1998) Principles of geographical information systems. Oxford University Press, Oxford.

[2] Lodwick, W. A. (2008) Fuzzy surfaces in GIS and geographical analysis : theory, analytical methods, algorithms, and applications. CRC Press, Boca Raton.

[3] Bandemer, H. (2006) Mathematics of uncertainty: ideas, methods, application problems. Springer, Berlin.

[4] Lodwick, W. A. and Santos, J. (2003). Constructing consistent fuzzy surfaces from fuzzy data. Fuzzy Sets and Systems, 135(2), 259-277.

[5] Bandemer, H. and Gebhardt, A. (2000) Bayesian fuzzy kriging. Fuzzy Sets and Systems, 112(3), 405-418.

[6] Zadeh, L. A. (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, 90(2), 111-127.

[7] Celikyilmaz, A. and Turksen, I. B. (2009) Modeling uncertainty with fuzzy logic: with recent theory and applications. Springer, Berlin.

[8] Zadeh, L. A. (1965) Fuzzy Sets. Information and Control, 8(3), 338-353.

[9] Zadeh, L. A. (2008) Is there a need for fuzzy logic? Information Sciences, 178(13), 2751-2779.

[10] Dutta, P., Boruah, H. and Ali, T. (2011) Fuzzy Arithmetic with and without using α-cut method: A Comparative Study. International Journal of Latest Trends in Computing, 2, 99-107.

[11] Lopez, C. and Samper, J. (2011) Numerical aspects of the universal kriging method for hydrological applications. http://www.thedigitalmap.com/~carlos/papers/rep96_1/geoenv96.pdf, [cit. 2011-10-31].

[12] Lloyd, C. D. and Atkinson, P. M. (2001) Assessing uncertainty in estimates with ordinary and indicator kriging. Computers & Geosciences, 27(8), 929-937.

[13] Bohling, G. (2005) Kriging. http://people.ku.edu/~gbohling/cpe940/Kriging.pdf, [cit. 2011-10-31].

[14] Fonte, C. and Lodwick, W. (2005) Modelling the Fuzzy Spatial Extent of Geographical Entities. In: Petry, F., Robinson, V. and Cobb, M. (eds.) Fuzzy modeling with spatial information for geographic problems. Springer, Berlin.

[15] Oksanen, J. and Sarjakoski, T. (2005) Error propagation of DEM-based surface derivatives. Computers & Geosciences, 31(8), 1015-1027.

[16] Waelder, O. (2007) An application of the fuzzy theory in surface interpolation and surface deformation analysis. Fuzzy Sets and Systems, 158(14), 1535-1545.

[17] DeBaets, B. and Markova-Stupnanova, A. (1997) Analytical expressions for the addition of fuzzy intervals. Fuzzy Sets and Systems, 91(2), 203-213.

[18] Chalco-Cano, Y., Jimenez-Gamero, M., et al. (2008) An approximation to the extension principle using decomposition of fuzzy intervals. Fuzzy Sets and Systems, 159(24), 3245-3258.

[19] Mesiar, R. (1997) Shape preserving additions of fuzzy intervals. Fuzzy Sets and Systems, 86(1), 73-78.

# EVALUATING HEIGHT DIFFERENCES BETWEEN GLOBAL DIGITAL SURFACE MODELS AND ICESAT HEIGHTS AT FOOTPRINT GEOLOCATION

Fabian ENSLE[1], Johannes HEINZEL[2] and Barbara KOCH[3]

Department of Remote Sensing and Landscape Information Systems, Faculty of Forest and Environmental Sciences, University of Freiburg, Tennenbacherstraße 4, 79106 Freiburg i.Br., Germany

*fabian.enssle@felis.uni-freiburg.de[1] , johannes.heinzel@felis.uni-freiburg.de[2], ferninfo@felis.uni-freiburg.de[3]*

**Abstract**

Three Digital Elevation Models (DEMs) derived from satellite data have been evaluated. These are the SRTM90 and SRTM25 derived from the Shuttle Radar Topography Mission (SRTM) and the Global Digital Elevation Model Version 2 (GDEMV2) computed with data of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). The elevations obtained by the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud and land Elevation Satellite (ICESat) are used to evaluate the vertical accuracy of the aforementioned DEMs. Elevation values are extracted at GLAS footprint centre coordinates. The study area is defined by the external borders of the federal state of Baden-Wuerttemberg, located in south-western Germany. An overall amount of 14615 elevation samples have been analyzed, where altitude information could be derived from all datasets. In addition to the spatial condition of intersecting height information, the heights of GLAS data product (Product GLA14, Version 31) were filtered for unsuitable elevation values. For evaluating the influence of different land use types, the land use map provided by the Office for the Environment, Measurements and Nature Conservation of Baden-Württemberg (LUBW) was used. To estimate the effect on elevation differences by surface slope, the latter was derived from SRTM25 data. Including all 14615 geolocations, results indicate the SRTM90 to have the best correlation with the ICESat heights. Mean differences and standard deviations between ICESat heights and the other elevation sources are: SRTM90 0.4m/10.1, SRTM25 -0.9m/7.8, GDEMV2 0.8m/9.3. Concerning absolute height differences the computed mean values and standard deviations are: SRTM90 6.5m/7.8, SRTM25 5.2m/6.0 and GDEMV2 6.9m/6.3. The slightest increase in mean deviation with increasing terrain slope class was observed in the GDEMV2 dataset.

**Keywords: DEM, DSM, DTM, SRTM25, SRTM90, ASTER GDEM, ICESat, GLAS, elevation, land use**

## INTRODUCTION

Digital Surface Models (DEM) by remotely sensed data are often the solely height information for remote areas. Besides the commercial high quality DEMs, produced by land survey offices or private companies, the freely available DEMs are widely in use. They are utilized for topographic correction of satellite images, calculation of slope or aspect for environmental

analyses or construction planning. Different kinds of DEM's by satellite remotely sensed data do exist. Small scale DEMs with a ground resolution of few km² down to high resolution DEMs of 25m² ground resolution. When asked about the most suitable DEM for a project, it is important to know the vertical accuracy. We will evaluate three common elevation sources that have been produced by satellite data. Two DEMs by the Shuttle Radar Topography Mission (SRTM) with a ground resolution of approximately 90m/25m and the ASTER Global Digital Elevation Model Version 2 (GDEMV2) will be evaluated. The Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud and land Elevation Satellite (ICESat) provides well-suited elevation measurements for evaluating these continental DEMs [1]. Instead of the contiguous DEMs, ICESat elevation information is provided more sparsely in a stripe scanning pattern, but of approved highly precise elevation measurements [1], [2], [3], [4], [5].

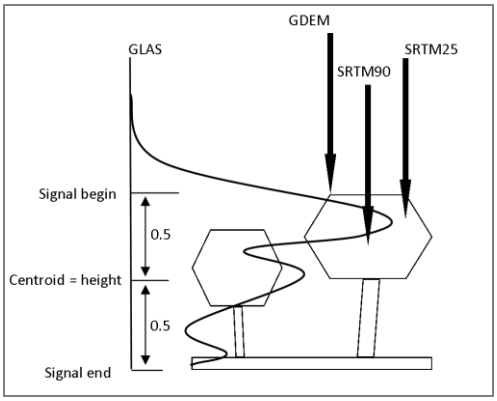



**Fig. 1** Sketchy illustration of forest canopy penetration by GLAS (shown elevation determination is one of seven methods), GDEM, SRTM90 and SRTM25 due to different wavelengths used by remote sensors

For different geographic regions and DEM sources a variety of accuracy assessments have been made. [1] validated elevations from the SRTM mission by ICESat elevations. They found a mean of -0.6m with a standard deviation of 3.46 for low relief and tree cover. Accuracy is deteriorated with increasing slope or increasing vegetation cover. For a study area in the Tibetan Plateau [6] identified a strong correlation of SRTM elevations with ICESat elevations. An assessment of ASTER GDEM using comparison with SRTM and ICESat data was conducted by [7] for central China. An investigation of ASTER GDEM versus SRTM was carried out by [8]. They also used ICESat data for absolute reference. The second release (version 2) of ASTER GDEM was also evaluated against ICESat elevation measurements by [9].

While height differences between single data have already been examined by other authors, until now they have not been evaluated for a complete large area. Besides this, we want to examine the influence of surface slope and the effect of different land use classes. Since the effect of slope on GLAS height measurements is well known [10], we want examine which of the tested DEMs is most sensitive for sloped terrain.

Because of the effects of different wavelengths on elevation computation and canopy penetration (Fig. 1), it is assumed that GLAS surface elevation, compared to the three other DEM heights, will be lower for forested and urban areas. The result of a subtraction of DEM values from ICESat heights should therefore be negative.

## STUDY AREA AND DATA

### Study Area

The extent of the study area for this investigation is the external boarder of the federal state Baden-Wuerttemberg of Germany (Fig. 2). The total land area of 35748km² is covered by mountain as well as by flat regions. The eastern part is built by the plain of the Upper Rhine valley. From this valley to the west the slopes of the Black Forest low mountain range arise. Further west the south-western German cuesta forms a series of plains separated by steep slopes.

### Data

The data collection for the SRTM was started in the year 2000 and comprised an eleven day mission [10]. The shuttle carried two different synthetic aperture radar systems. One of these radar systems operated within C-band and the other within X-band wavelengths (5.6cm and 3.1cm) [11]. The outcome of that mission is a worldwide topographic dataset with near-global coverage. The first release of the C-band derived elevation products was in the year 2003 with a degraded resolution of 3-arc seconds outside the US. The Consortium for Spatial Information (CGIAR-CSI) of the Consultative Group for International Agricultural Research (CGIAR) provides post processed DEM datasets of 3-arc seconds resolution (~ 90m). This product was used in Version 4 within this study. Data is delivered in geographic projection of WGS84 and vertical datum EGM96.

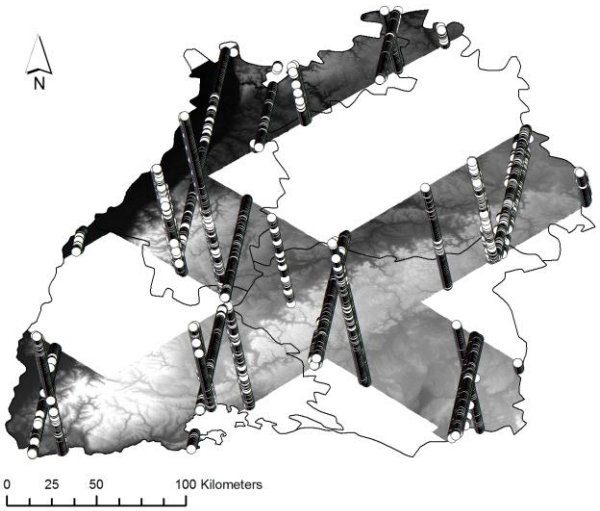



**Fig. 2** Extent of the federal state Baden-Württemberg with circular footprints of ICESat GLAS ground tracks (adjusted data) and available SRTM25 data.

The DEM derived from the X-band radar system is under charge of the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt, DLR) and the Italian Space Agency (ASI) [12]. This SRTM25 was made available for scientific purposes from end of May 2011 by DLR. This product was downloaded with a geographic projection of WGS84 and WGS84 ellipsoidal heights. It has a better ground resolution of approximately 25m x 25m compared to the SRTM90, but only exists in a global stripe pattern. Figure 2 illustrates the coverage by SRTM25 elevation data at the study area.

The ASTER GDEM data is property of the Japanese Ministry of Economy, Trade, and Industry (METI) and the United States National Aeronautics and Space Administration (NASA) [13]. Elevations are determined by an automated stereo-correlating algorithm. From 17th October 2011 on, the second version of ASTER GDEM is available. With a ground resolution of 30m x 30m, the data was obtained from Japan's Ground Data System. The GDEMV2 elevation product is delivered in the same projection and vertical datum as the SRTM90 (WGS84/EGM96).

The ICESat was lunched on 12$^{th}$ January 2003. The GLAS instrument on board ICESat had three lasers with a wavelength of 1064nm for the altimetry measurements. To extend the mission life time, the operating period of the three lasers was non-permanent. The measurement campaigns lasted between 33 to 56 days, several times per year. The last operational period ended on 11$^{th}$ October 2009. Footprint size of the lasers is about 70m in diameter with a spacing of about 170m along record path. This varies with each laser period [14]. Hence the data distribution is point wise and no contiguous full coverage for the study site is available (Fig. 2). These ICESat GLAS data products are distributed by the National Snow and Ice Data Centre (NSIDC). There are 15 different standard data products, GLA01 to GLA15. The GLAS/ICESat L2 Global Land Surface Altimetry Data product GLA14 was used for this study. The most current two releases of NASA ICESat GLAS data products are Release-31 and Release-33. Because only Release-31 is available for all laser campaigns, it was used for this study [15]. The data is provided in scaled integer binary format and can be ordered through the ICESat/GLAS Data Subsetter provided by NSIDC [16]. Due to the reference to the TOPEX/Poseidon ellipsoid and EGM96 geoid, the coordinates need to be further processed.

The land use map applied by the Office for the Environment, Measurements and Nature Conservation of Baden-Württemberg (LUBW) contains 16 land use classes with a spatial resolution of 30mx30m [17]. Land use classes are: (1) urban dense, (2) urban sparsely, (3) industrial, (4) arable land, (5) wine/orchard, (6) mixed orchard, (7) fallow, (8) without vegetation, (9) intensive pasture, (10) extensive pasture, (11) coniferous, (12) deciduous, (13) mixed forest, (14) wind fall, (15) water, (16) wetland. The land use classification is based on Landsat images, captured during the year 2000.

Table 1 summarizes the elevation data sources, their specific geographic projection, the vertical reference system and the ground resolutions.

**Table 1** Summary of elevation sources

| Name | Projection | Ground Resolution |
|---|---|---|
| SRTM90 | WGS84/EGM96 | 90x90m |
| SRTM25 | WGS84/WGS84 | 25x25m |
| GDEMV2 | WGS84/EGM96 | 30x30m |
| ICESat\GLAS | TOPEX-Poseidon | point (x, y, z) |

## METHODOLOGY

The first step comprises a thorough pre-processing of ICESat data. Data from GLAS instrument is stored with unique record indexes where each record index contains 40 laser shots. Some data attributes are stored for the complete record index and others are recorded for each laser shot [18]. Variable names are indicated by unique flags of which the NSIDC provides an altimetry data dictionary with a detailed description [18]. The GLAS land surface elevations are determined by several methods. The method used for each specific waveform is indicated by the 'i_ElvFlg' flag. According to this flag the elevation measurements used in this study have been calculated by the centroid of the received pulse between signal begin and signal end, defined for alternate parameterization (Fig. 1). For further investigations we converted the ICESat GLA14 product from binary to ASCII format. NSIDC offers the tool 'IDL reader' to read data and print all the variables in ASCII format [19]. Since output data structure is not instantly suitable to transfer the records to a data base or preferred table format, order of ASCII data sets have been transposed and new columns with derived values were added. The necessary height transformation from TOPEX/Poseidon to WGS ellipsoidal heights was of importance for this study. A subtraction of geoid undulation values according to [20] was realised. In contrast to [20] we did not use a fixed height offset between TOPEX/Poseidon and WGS ellipsoids of 0.7m, but calculated the offset by the empirically derived formula provided by [21]. Geoid undulation is already given in GLA14 products; however it is only stored for the first and last shot for each record index. We calculated the undulation value for each shot, by linear interpolation between the first and last recorded value. Additional to the height transformation, we reduced the amount of records by filtering invalid or critical values by the following criteria:

a) Elevation use flag 'i_ElvuseFlg' indicating invalid elevations

b) Saturation correction flag 'i_satCorrFlg' indicating invalid correction value

c) Range offset quality flag 'i_rng_UQF' indicating invalid values

d) Cloud contamination flag 'i_FRir_qaFlag' indicating presence of clouds

e) Difference of GLAS height to high resolution DEM 'i_DEM_hires_elv' is more than 100m.

We stored all of the information provided in the GLA14 product in a database, but only latitude, longitude and the transformed height values were used for further evaluation in that study. A total of 14615 geolocations do comply with the requirements of data quality and existing elevation information of all DEMs inside the area under investigation.

The second step requires the transformation of datasets into a unique reference system. The elevation of the SRTM25 dataset refers to the WGS84 ellipsoid. As we want to calculate height differences based on EGM96, we transformed these values. The National Geospatial-Intelligence Agency (NGA) and NASA provide a WGS84 EGM96 15-Minute Geoid Height File and Coefficient File, as well a FORTAN program, named F77, to calculate undulation based on geographic coordinates [22]. For the centre coordinates of each SRTM25 pixel, the undulation was computed by the aforementioned software and finally subtracted from original values.

Height transformation was not necessary for the SRTM90 and GDEMV2 since they are already delivered in EGM96. Coordinate systems have been changed from geographic to map projection WGS 84 /UTM zone 32N. For the land use map a transformation from DHDN / Gauss-Krueger Zone 3 to UTM 32N WGS84 was computed within a Geographic Information System (GIS) using an equation based seven-parameter transformation.

In the third step, the ICESat footprint centre coordinates were used as template to extract elevation and land use information from the other data sources. We did not account for the footprint size or its shape. As well we did not spatially interpolate the raster values according to the location of the GLAS footprint centre inside each corresponding pixel. Finally the DEM values were subtracted from corresponding ICESat elevations.

At the time of this study there was no better possibility to get surface slope information, then to derive it from one of the available DEMs. We decided to take the newly released SRTM25, to calculate surface slope, being aware of the possible autocorrelation of results. Slope values have been grouped into different intervals. They are from 0° to 20° degrees in the range of one degree per interval, between 20° and 65° they are separated in 5° each interval.

## RESULTS

The 14615 sample locations, extracted at centre coordinates of ICESat footprints, are distributed within different land use classes. Within dense urban areas (1) 23 intersecting height values were observed, urban sparsely populated areas (2) 1085, (3) industrial areas 116, (4) arable land 3391, (5) wine/orchard 378, (6) mixed orchard 310, (7) fallow 18, (8) without vegetation 20, (9) intensive pasture 419, (10) extensive pasture 2911, (11) coniferous 1890, (12) deciduous 1765, (13) mixed forest 2057, (14) wind fall 143, (15) water 50 and wetland (16) 39.

Elevation differences for all locations, without any further data manipulation, have a mean value 0.4m with a RMSE 10.1 for SRTM90, -0.9m with RMSE 7.8 for SRTM25 and 0.8m with RMSE 9.3m for the GDEMV2 data set. Mean values and RMSE calculated with absolute values are 6.5m/7.8 for SRTM90, 5.2m/6.0 for SRTM25 and 6.9m/6.3 for GDEMV2. Figure 3 shows four different data plots. The median of each class is represented by the horizontal line inside each box. The dimension of each box indicates inner quartile range (IQR) of the first and third quartile. The upper and lower whiskers are calculated by multiplying the IQR by 1.5. This value is subtracted afterwards from the first quartile. The next higher value of the dataset determines lower whisker line. The upper whisker is calculated by starting from the third quartile. All data values, which exceed the whiskers, are plotted as dot. In figure 3 a-c, the

height differences per land use class can be seen. High IRQs can be observed in the land use class coniferous forest (11), deciduous forest (12), mixed forest (13) and water (15). Lower IRQs and whisker ranges especially occur at land use type of arable land (4) and wine yards/ orchards (5). Figure 3c is of same plot type as the others and is illustrating the distribution of surface slope within the land use classes. Similar to plot 3a-b, land use classes 11, 12, 13 and 15 do have higher median values and higher whisker ranges. Due to small amount of samples inside the land use classes 1, 3, 7, 8, 15 and 16, the corresponding results should not further be considered.

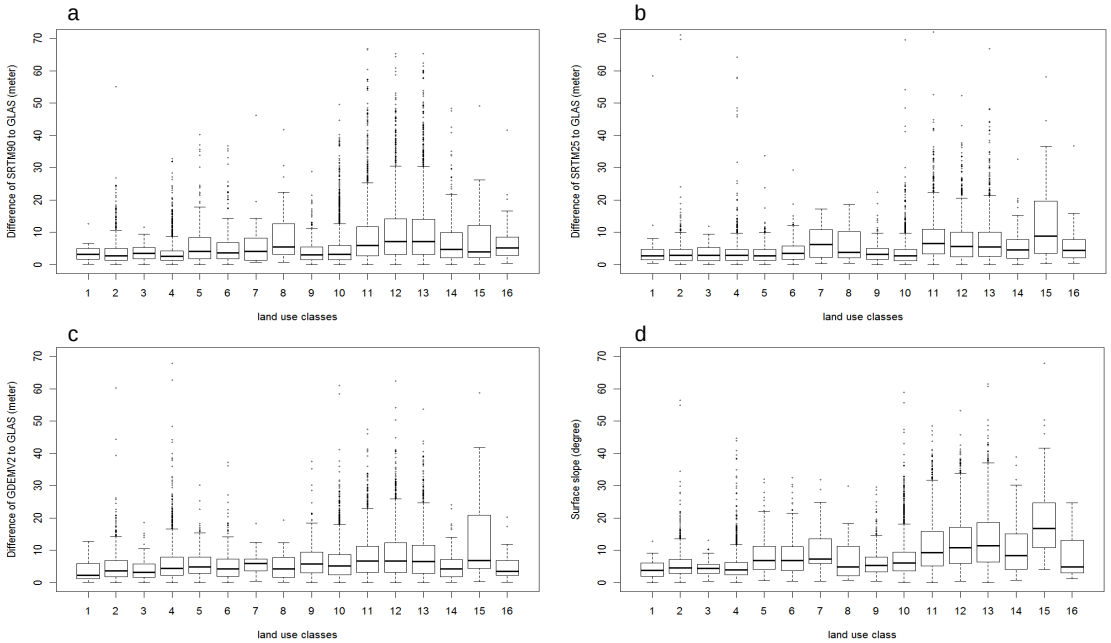



**Fig. 3 (a)** SRTM90 to GLAS elevation difference by land use **(b)** SRTM25 to GLAS elevation difference by land use **(c)** GDEMV2 to GLAS elevation difference by land use **(d)** Surface slope distribution inside land use classes. Land use classes: (1) urban dense, (2) urban sparsely, (3) industrial, (4) arable land, (5) wine/ orchard, (6) mixed orchard, (7) fallow, (8) without vegetation, (9) intensive pasture, (10) extensive pasture, (11) coniferous, (12) deciduous, (13) mixed forest, (14) wind fall, (15) water, (16) wetland

As described in the method section, a grouping of surface slope values into specified intervals was conducted. Fig. 4, Fig. 5 and Fig. 6 illustrate the observed deviance between GLAS heights and DEM heights in relation to surface slope (absolute values in the range of 1° to 20°). Surface slope values are distributed inside the slope classes as follows: (1) 357, (2) 1026, (3) 1337, (4) 1521, (5) 1406, (6) 1227, (7) 1047, (8) 909, (9) 734, (10) 605, (11) 530,(12) 435 ,(13) 411,(14) 349, (15) 291, (16) 274, (17) 247, (18) 215, (19) 188, (20) 165, (25) 654, (30) 368, (35) 185, (40) 37, (45) 21, (50) 11, (55) 4, (60) 3, (65) 2 and 20 observations are above 65 degree. The distribution is right-skewed with a maximum of samples between 3° and 4° in class number 4 (Fig. 7). The high difference between class 20 and 25 occurs because of a change in the interval range. The number of 1341 observations within all classes above 20° surface slope is relative small compared to the other 13274

observations, which are located between 0° and 20°. Linear regression of height differences and ungrouped original slope values in the range of zero to 20 degrees was conducted. With low $r^2$ values between 0.01 and 0.03 a significant relationship could not be found. Gradients for the linear function are 0.46/SRTM90, 0.25/SRTM25 and 0.13/GDEMV2.

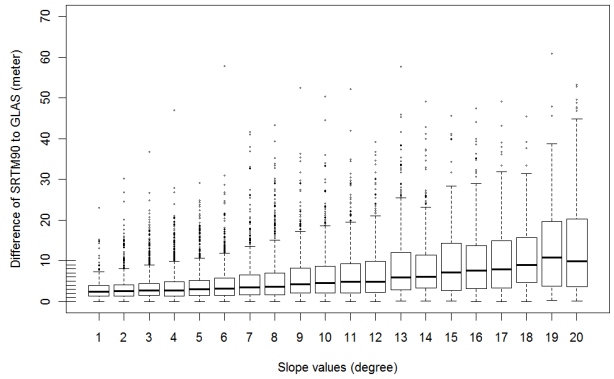



**Fig. 4** Difference of SRTM90 to GLAS elevations by surface slope

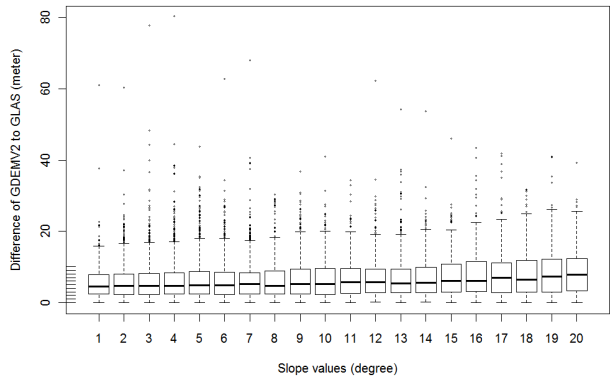



**Fig. 5** Difference of GDEMV2 to GLAS elevations by surface slope

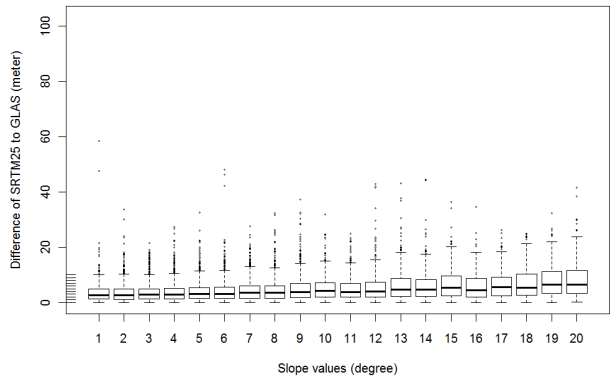



**Fig. 6** Difference of SRTM25 to GLAS elevations by surface slope

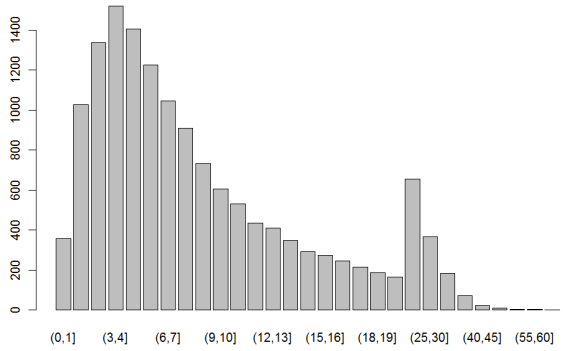



**Fig. 7** Allocation of samples inside slope classes.
Class value is upper range value (0°-1° = class 1)

## DISCUSSION AND CONCLUSION

Elevation differences have been analysed under consideration of two potential influence factors: land use and surface slope. Related to land use classes we observed different values and mean variations of elevation differences. This implies that the land use class does affect elevation differences. Distribution of slope values inside the land use classes (Fig. 3d) is not the same in each class. Compared to other land use classes, the classes 11, 12, and 13 contain higher values of surface slopes. With regard to the forest types it is noticeable that these classes do have similar mean values. This statement can be observed for each DEM. The samples located in forested areas seem to be in more steep areas compared to the other land use classes (Fig. 3d). Two possibilities have to be further investigated. First, elevation measurements are different due to behaviour of different wavelengths (Fig. 1) in relation to the type of land cover. Second, the strong influence of surface slope on height measurements covers any further influence by land use classes. The fact that we used SRTM25 data to calculate surface slope must be viewed critically, because it is well known that radar backscatter behaviour is strongly affected in mountainous areas. Further analyses should be carried out with an independent comparative data set from airborne LiDAR.

With regard to the method used for elevation extraction, we did not account the different resolutions of DEM sources. The extraction of raster values is based on the GLAS footprint centre location. It is not based on the footprint size and therefore not based on a spatial interpolation. For an accurate spatial interpolation, calculation of shape and energy distribution of each ICESat footprint should be done. This will be investigated in further studies with a combination of a high resolution DEM by airborne laser scanning.

Especially the method of GLAS elevation computation should be considered critically. In non-forested areas the method ‘centroid of signal end to signal begin’ for elevation computation may be sufficient. In forested areas the GLAS surface elevation does not represent the top of canopy or the bare ground (Fig. 1). For a more comparable study the behaviour of GLAS waveform on forested areas needs to be considered. A more adapted computation with

respect to forest canopy by GLAS waveforms should make the GLAS elevations a better reference to evaluate the other elevation models in forested areas.

The results of this study are summarized below. Best elevation model for mountainous areas is the GDEMV2. It is least affected by rising surface slope. In flat areas, due to the higher ground resolution, the SRTM25 should be used. As well it has the lowest offset to ICESat data (5.2m), calculated by absolute differences. Further studies should adapt the ICESat elevation computation to the land cover.

## ACKNOWLEDGEMENTS

The authors are grateful to the ReCover consortium and the European Commission for supporting this research within the Seven Frame Work Programme FP7 (grand agreement no. 263075).

## REFERENCES

[1] Carabajal, C. C., Harding, D. J. (2005) ICESat validation of SRTM C-band digital elevation models. Geophys. Res. Lett 32 (22).

[2] Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J.F., Palm, S., Yi, D., and Liiva, P. (2005) Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. Geophys. Res. Lett 32 (21).

[3] Braun, A., Fotopoulos, G. (2007) Assessment of SRTM, ICESat, and Survey Control Monument Elevations in Canada. Photogramm. Eng. Rem. Sens. 73 (12), S. 1333–1342.

[4] Rosette, J. A. B., North, P. R. J., Suarez, J. C., Los, S. O. (2010) Uncertainty within satellite LiDAR estimations of vegetation and topography. Int. J. Remote Sensing 31 (5), S. 1325–1342.

[5] Shtain, Z. and Filin, S. (2011) Accuracy and reliability assessment of GLAS measurement over Israel. ISPRS Workshop, Laser scanning 2011, ISPRS, Volume XXXVIII, Calgary, Canada, conference proceedings.

[6] Huang, X., Xie, H., Liang, T. and Yi, D. (2011) Estimating vertical error of SRTM and map-based DEMs using ICESat altimetry data in the eastern Tibetan Plateau. Int. J. Remote Sensing 32 (18), S. 5177–5196.

[7] Guosong, Z., Huaiping, X. and Feng, L. (2010) Assessment of ASTER GDEM performance by comparing with SRTM and ICESat/GLAS data in Central China, 18th International Conference on Geoinformatics, 1-5, 18-20 June 2010.

[8] Reuter, H.I., Nelson, A., Strobl, P., Mehl, W. and Jarvis, A. (2009) A first assessment of Aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters. Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, vol.5, V-240-V-243, 12-17, July 2009.

[9] Carabajal, C. C. (2011) Aster Global Dem Version 2.0 Evaluation Using Icesat Geodetic Ground Control. https://igskmncnwb001.cr.usgs.gov/aster/GDEM/GDEM_Validation_Documents.zip, [cit. 2011-10-31].

[10] National Aeronautics and Space Administration NASA (2011) Science: ICESat Measurements. http://icesat.gsfc.nasa.gov/icesat/science/measurements.php, [cit. 2011-10-31].

[11] Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S. et al. (2007): The Shuttle Radar Topography Mission. Rev. Geophys 45 (2).

[12] Deutsches Zentrum für Luft- und Raumfahrt (2011) SRTM. https://centaurus.caf.dlr.de:8443/eowebng/ licenseAgreements/DLR_SRTM_Readme.pdf, [cit. 2011-10-31].

[13] Tachikawa, T., Hato, M., Kaku, M. and Iwasaki, A. (2011) The characteristics of ASTER GDEM version 2, IGARSS, July 2011.

[14] Zwally, H., (2004) Overview of the ICESat mission and results, Abstract #C21B-01, Proceedings of the American Geophysical Union Fall Meeting, December, San Francisco.

[15] Zwally, H.J., Schutz, R., Bentley, C., Bufton J., Herring, T., Minster, J., Spinhirne, J. and Thomas, R. (2003) updated current year. GLAS/ICESat L2 Global Land Surface Altimetry Data V031, Acquired: February 2011 to March 2011. Boulder, CO: National Snow and Ice Data Center. Digital media 2011.

[16] National Snow and Ice Data Center (NSIDC) (2011) Distributed Active Archive Center (DAAC). ICESat/GLAS Data Subsetter. http://nsidc.org/data/icesat/order.html, [cit. 2011-10-31].

[17] Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (2011) LUBW. http://www.lubw.baden-wuerttemberg.de, [cit. 2011-10-31].

[18] National Snow and Ice Data Center NSIDC (2011) Data dictionary. http://nsidc.colorado.edu/data/docs/daac/glas_altimetry/ data_dictionary.html, [cit. 2011-10-31].

[19] National Snow and Ice Data Center NSIDC (2011) Tools. http://nsidc.org/data/icesat/tools.html, [cit. 2011-10-31]

[20] Bhang, K.J., Schwartz, F.W. and Braun, A. (2007) Verification of the Vertical Error in C-Band SRTM DEM Using ICESat and Landsat-7, Otter Tail County, MN. IEEE Trans. Geosci. Remote Sensing 45 (1), S. 36–44.

[21] Haran, T. (2004) Ellipsoid. ftp://sidads.colorado.edu/pub/DATASETS/icesat/tools/idl/ellipsoid/README_ ellipsoid.txt, [cit. 2011-10-31].

[22] National Geospatial-Intelligence Agency NGA and NASA (2011) EGM 96. http://earth-nfo.nga.mil/GandG/wgs84/ gravitymod/egm96/egm96.html, [cit. 2011-10-31].

# ANALYSIS OF THE GEOMETRIC PROPERTIES OF DERIVED MULTIRESOLUTION TIN TERRAIN MODELS WITH A VIEW TO MORPHOMETRIC ANALYSIS

Richard FECISKANIN[1]

[1]Department of Cartography, Geoinformatics and Remote Sensing,

Faculty of Natural Sciences, Comenius University in Bratislava,

Mlynská dolina, 84215, Bratislava, Slovakia

*richard.feciskanin@fns.uniba.sk*

**Abstract**

Automatic creation of terrain models by simplifying detailed models become increasingly important with the availability of comprehensive (sometimes redundantly dense) elevation data. In many situations it is expedient to use a modified (simplified) model instead of an original model created from all available data. Simplification plays a key role in the generation of models in various resolutions. We derived TIN terrain models in a number of resolutions by selecting automatic model simplification procedures known from the field of computer graphics. The paper presents an analysis of ability these models for terrain modeling where geometrical accuracy is important – the accuracy of partial derivatives and from them derived morphometric parameters. Therefore, we focused on evaluating the accuracy of partial derivatives computed from the model. The main criterion is the deviation of the triangle normal to normal of a represented surface. We studied the impact of loss of detail in simplifying and impact of random error in input data. The selected simplifying procedures used detailed information about the shape of the terrain surface to adapt individual elements (triangles) to the represented part of the surface to achieve the best possible fidelity. It is shown that we can reach very high efficiency of the terrain model using surface simplification. The model with the fraction of elements (vertices, triangles) from the original one can maintain sufficient accuracy of derived geometric characteristics.

**Keywords: simplification, terrain models, TIN, morphometric variables, comparison**

## INTRODUCTION

Modern methods of obtaining detailed data have brought the possibility to create efficient terrain models with high accuracy. This requires the inclusion of methods that have been previously used rarely, such as simplifying models. Large amount of data are processed to obtain detailed information about the shape of the surface, which will be used for the automated creation of the (derived) model, which efficiently and accurately represents the terrain surface. Simplification methods were developed and improved in other fields such as in the field of computer graphics. Nearly a decade ago,

[1] evaluated simplification methods in computer graphics as mature, enough to use in the field of terrain modeling.

Terrain modeling has its own specifics, which have to be taken into account in applying the procedures. Morphometric variables describe geometric characteristics of terrain and they are often entered into other application models. Therefore, an accuracy of morphometric variables is crucial. This reflects the accuracy of the values of partial derivatives and from them expressed morphometric variables. We will mention inaccuracies of partial derivatives determined from triangulated irregular network (TIN) models (polyhedral models based on 2D triangulation), which are essential for morphometric analysis. Just the used method of model accuracy assessment based on a normal deviation includes the characteristics of these inaccuracies. This part of the problem is treated in the second chapter of this paper.

The origin of numerical and positional deviations of calculating geometric characteristics of the TIN model presented [2], [3]. The work [4] contains extended analysis of the properties of deviations. There were also presented properties of the triangle properly configured to the represented surface, which ensures minimization of deviations of partial derivatives. Accordingly, we analyzed the most common methods and selected those that use the geometric characteristics of the properly configured triangle during the simplification process. We chose two methods under the assumption of a very good geometric fidelity: *quadric error metrics simplification* presented in the work by [5] and *memoryless simplification* presented in the work by [6]. The third chapter describes the basic characteristics of the methods and their relation to the definition of the correctly configured triangle.

We present an analysis of changes in geometrical fidelity measurements among the simplified models in distinct levels of detail in the fourth chapter. We also evaluated the impact of random error in input data. The results showed the positive impact of simplification to eliminate noise. They confirmed that the creation of a simplified model gives the opportunity to excel advantages of TIN, where it can very effectively and yet with sufficient accuracy represent a modeled surface on the basis of detailed information about it. There were obvious significant differences between the behaviors of the tested algorithms, despite their similar nature. This is evidence that the selection of an appropriate method has a major impact on the accuracy of the model.

The paper shows specific terrain analysis issues, specific TIN terrain models issues, and what is the value added to a similar analysis addressed by other research fields, for example [7]. Analysis of algorithms, finding the simplification methods using geometrical properties of the optimal triangle and geometric fidelity analysis of created models by chosen methods are new contributions of this paper.

## ACCURACY OF DERIVED GEOMETRIC PARAMETERS

The next part of analyses focuses on the creation of deviations in the calculation of the values of structural geometric parameters (slope, aspect, curvatures, and so forth) from a TIN model. Deviations present a result of the relationship between surface geometry

and a shape of the triangle that it represents. The primary consequence of an inappropriate relationship is the triangle normal that does not match the terrain surface normal at a given location. Normal deviation cause inaccuracies in determining the partial derivatives and hence the calculation of morphometric variables.

**Relationship Between Surface Geometry and a Shape of a Triangle**

This relationship can be studied on the basis of height differences $\Delta z$ between the plane of the triangle $\sigma_\Delta$ and the terrain surface. For $\Delta z$ is valid $\Delta z = P_z - P_\Delta$, where $P_z$ is the height of the terrain at any point $P(P_x, P_y, P_z)$ and $P_\Delta$ is the third coordinate of the vertical projection of $P$ on the triangle plane $\sigma_\Delta$ with coordinates $(P_x, P_y, P_\Delta)$.

An important role has curve $\Delta z = 0$, which is the intersection of the triangle plane and the terrain. Since the triangle vertices lie on the terrain and in the $\sigma_\Delta$ (a plane they determine), then also lie on the curve $\Delta z = 0$. This curve is a circum curve of the triangle. Its shape is subject to the shape of the terrain in a given location and it is very close to the shape of the Dupin indicatrix, which describes the shape of the terrain surface.

Another important indicator is the position of tangent point $T$ of tangent plane $\sigma_T$ that is parallel to the plane of triangle $\sigma_\Delta$. Then, the triangle normal vector $\boldsymbol{n_\Delta}$ is identical to the normal vector of the tangent plane and then also to the normal vector of terrain $\boldsymbol{n_T}$ in point $T$. Tangent point $T$ has the same values of the first partial derivatives, slope and aspect as calculated from the triangle. [4] provides a more detailed characterization of the relationship between the surface geometry and a shape of the triangle.

**Numerical and Positional Deviations**

[2], [3] identifies a property of the triangle that is not suitable for the representation of terrain as an improper configuration of the triangle. There are deviations in calculations of the values of geometric parameters in improper configured triangles. Almost all triangles forming a triangular network incur less, or greater deviation. This is due to the fact that the real surface normal vector $\boldsymbol{n_G}$ in place of triangle centroid $G$ – with horizontal position $(G_x, G_y)$ – differs from the normal vector of triangle plane $\boldsymbol{n_\Delta}$ (see Fig. 1). Therefore, it is valid $\boldsymbol{n_G \neq n_\Delta}$ and the angle between normals is determined by the relation

$$\cos \delta n = \frac{\boldsymbol{n_G} \boldsymbol{n_\Delta}}{|\boldsymbol{n_G}||\boldsymbol{n_\Delta}|}. \qquad (1)$$

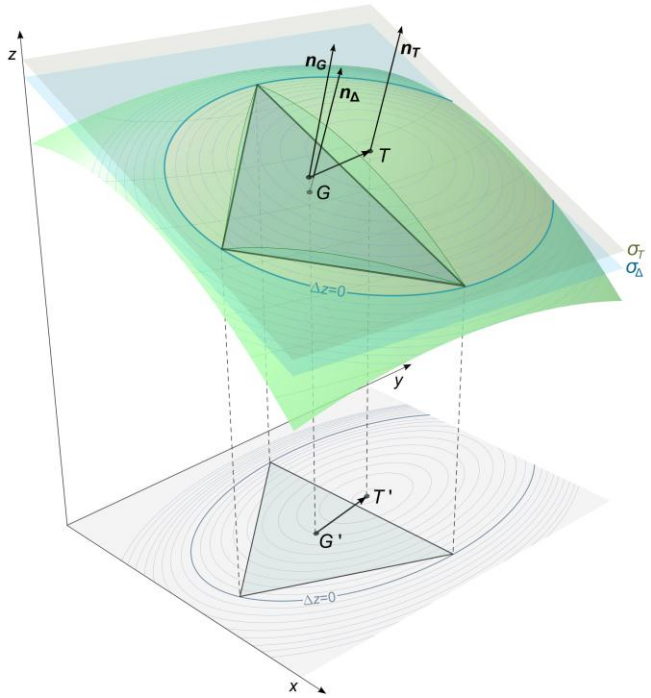



**Fig. 1.** Terrain surface geometry, triangle and deviations

Angle δ*n* represents a normal deviation which, as mentioned above, causes inaccuracies in determining the partial derivatives and hence the calculation of morphometric variables. That is why we chose this parameter as the most important for evaluating geometrical fidelity with a view to morphometric analysis.

Deviations also have positional aspect. It can be usually found a point which has the same normal vector as a triangle plane, and therefore the same values calculated from the first partial derivatives. This point is the mentioned tangent point *T* of tangent plane $\sigma_T$ that is parallel to the triangle plane $\sigma_\Delta$. Positional difference between the point *T* and the triangle centroid *G* creates positional deviation vector.

### Properly Configured Triangle

Due to the above principles and characteristics of numerical and positional deviations we can formulate the conditions for the proper configuration of triangle vertices to terrain surface so that the normal vector $\boldsymbol{n}_{\boldsymbol{\Delta}}$ is identical with the real terrain surface normal vector $\boldsymbol{n}_{\boldsymbol{G}}$ at the place of triangle centroid *G*. If $\boldsymbol{n}_{\boldsymbol{G}} = \boldsymbol{n}_{\boldsymbol{\Delta}}$ then there is no normal deviation and so $\delta n = 0$. This will ensure that values of the partial derivatives and from them expressed morphometric variables are identical as real values of terrain at *G*. This condition will be valid if the centroid *G* and the point *T* have the same horizontal location

$$G_x = T_x,\ G_y = T_y. \qquad (\mathbf{2})$$

Condition (2) is a theoretical condition for proper configuration of the triangle, formulated by [2]. This condition can be fulfilled only under certain conditions and only for the individual triangles, not for groups of triangles in a triangular network. This follows from the geometric relationship between the surface geometry and the shape of a triangle. We can effectively study the properties if the terrain surface, which could be described by the general function of two variables

$$z = f(x, y) \quad (3)$$

is replaced in the small neighborhood of the triangle by a simpler surface. We used an osculating paraboloid with a tangent plane in its vertex identical to the tangent plane of terrain in selected point. The vertex part of the osculating paraboloid in differentially small neighborhood of point is identical to the terrain surface. We will consider small, but finite large neighborhood. Simplification of part of the terrain shape by substituting it with part of osculating paraboloid with vertex $T$ can approximately, but sufficiently accurately reflects the studied relationship. The isolines of height differences $\Delta z$ are then the intersections of osculating paraboloid and the plane parallel to the planes $\sigma_T$ and $\sigma_\Delta$, which are conics. In this case, the Dupin indicatrix gives exact characteristics of isolines of $\Delta z$.

For our needs will be more purposefully to define the conditions of the proper configuration for the plane triangle (local coordinates), not for scalar base ($x, y$). Thus, we will address all aspects of the conditions of the proper configuration in the plane. Then, the condition (2) shows that the triangle centroid $G$ must lie in the center of a conic – intersection of a triangle plane $\sigma_\Delta$ and the osculating paraboloid. This can be achieved only if the conic is an ellipse. Intersection is then described by an ellipse with the center in the triangle centroid $G$, known as the circumscribed Steiner ellipse. In the places where Dupin indicatrix is a hyperbola the condition of proper configuration cannot be fulfilled without residue. The evidence can be found in [4].

We analyzed algorithms using the most common simplification methods and we selected those that use the presented geometric characteristics of the properly configured triangle during the simplification process. This creates a precondition for achieving a very good geometric fidelity of created models.

## SIMPLIFICATION METHODS

There are plenty of works dealing with methods of simplifying models of surfaces and objects, dozens of them are described and categorized in survey [8]. Therefore, we only outline the major works that use the basic types of procedures.

### Types of Simplification Methods

The majority of work presents a method to repeatedly perform a local change in geometry, while the limit reached chosen criterion. There are two basic approaches to the local changes leading to the creation of a simplified model. Methods can be divided into the categories, refinement and decimation. *Refinement methods* are starting with a

few points from the original model in the initial state and more points (triangle vertices) are added while the accuracy of the model is not at the required level. On the opposite, *decimation methods* start with the original model and reduces the number of elements (vertices, edges, triangles), while the level of accuracy of the model according to the chosen criterion is sufficient. Advantage of decimation methods over refinement methods is the ability to compare the influence of geometry changes to the original model (not only the actual state of the model).

Some of the surface simplification methods have been developed specifically for terrain models. They use the fact that the terrain surface is considered under defined conditions as a function (3), where the height is only a function of position. This allows using 2D Delaunay triangulation and deviation measurement only in the direction of the axis *z*. Refinement methods are more common in this category. There are some traditional and still popular methods, namely, *Fowler and Little algorithm* [9]*, VIP (Very Important Points) Algorithm* [10]. From decimation methods is a well-known *Drop heuristic method* [11]. We can overall characterized traditional methods, which are parts of the GIS software solutions, as inefficient, often with insufficient accuracy. They are used mostly in creating a TIN model from a regular grid. For the purpose of creating a very detailed model of the elevation data is better to use other methods of simplifying models.

Simplification methods for more general models of surfaces and objects that cannot be described by the equation (3) are mainly decimation methods and refinement methods are used rarely. Refinement methods use mostly a hierarchical division of triangles, as it stems from a report [8]. Methods vary according to the approach to reduce the number of elements. [1] categorized them into groups based on: vertex (or face) decimation, vertex clustering and vertex merging (edge contraction). Edge contraction (merge its two vertices $V_0$ and $V_1$ to one $V$) is most commonly used process. The advantage is that it is a more atomic operation than vertex decimation or clustering. In one step, one vertex, three edges and two triangles are removed (for non-border edges). It does not require the invocation of a triangulation algorithm.

**Characteristics of Selected Methods**

As mentioned above, we chose for evaluation *quadric error metrics simplification* presented in the work [8] and *memoryless simplification* presented in the work [6]. Both methods use an edge contraction to perform a local change in geometry. Two fundamental decisions affecting the properties of the final model. The first is the location of a new vertex after a contraction and the second is a way of sorting the edges – order of contraction based on selected measurement of edge weight. Despite differently defined conditions, both methods are very similar in the background.

The *memoryless simplification* method determines the weight of edges by calculating the change in volume when contracting edges. Weight value is the sum of volumes of tetrahedrons arising from the surrounding triangles shifting the original vertices $V_0$ and $V_1$ to the new vertex $V$ (see fig. 2). The *quadric error metrics simplification* method determines the weight to the value of the quadratic distance of the vertex to triangle

planes with vertices $V_0$ and $V_1$. Both methods need to first determine the location of the new vertex after edge contraction.

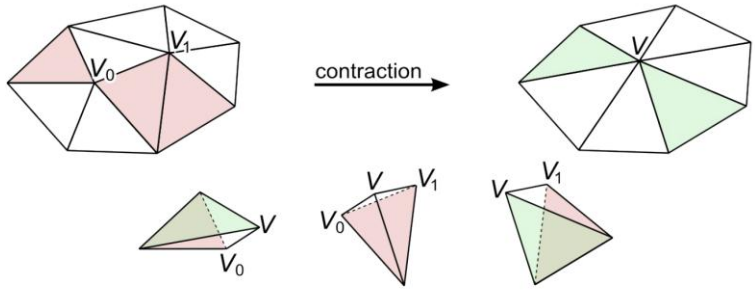



**Fig. 2.** Edge contraction in memoryless simplification

Determining the location of the vertex *V* by these methods is solved by the minimization of a function used to evaluate weight of edges. The process used in the *quadric error metrics simplification* leading in standard circumstances to unambiguously determine the location of *V*. This is the place with a minimum quadratic distance to triangle planes with vertices $V_0$ and $V_1$. These planes are so touching isosurface with a certain quadratic distance from the new vertex. This isosurface is quadric - ellipsoid, which gives the name of the quadric error metrics [5]. Example in two dimensions is shown in fig. 3.

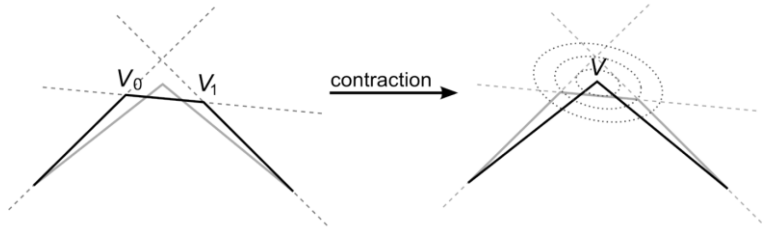



**Fig. 3.** Edge contraction in quadric error metrics simplification in 2D

The evaluation of tetrahedron volume change for determining the location of *V* is not enough. From a geometrical point of view it is the intersection of two planes. The *memoryless simplification* method adds a conservation of volume condition, which further defines a plane. In this case, the individual tetrahedrons have associated sign depending on whether it is an increase or decrease of the volume. This condition also ensures the global preservation of the model volume.

As pointed out by [12], presented basic determination of the optimal position of the new vertex is very similar in both methods. Basically, they are using the same characteristics – can be converted into identical quadratic shape. It differs only in determining the weight of the triangles in a sum. While in the memoryless simplification method is the weight a square area of the triangle, in the quadric error metrics simplification method members are weighted equally, or by the triangle area.

An important difference, which concerns the behavior of the algorithm is that [5] calculated weight for each vertex is stored, while [6], [12] calculate weight on-the-fly. The most important factor that makes the quadric error metrics simplification method one of the fastest among similar methods is an easy determination of the weight of an edge contraction by sum of values stored per vertex.

**Comparison with Properties of Properly Configured Triangle**

Theory of approximation uses various evaluation methods and metrics for deviation measurement of surface approximation by triangles as plane triangular elements of the triangle network. On this basis, there were several works that try to express the optimal shape of a triangle regards the geometric characteristics of the represented surface. We will show that they are equivalent to those that we have defined in the second chapter and to triangles created by conditions in evaluated methods.

As stated by [13], Nadler presented in 1985 that among all triangles with a given area a triangle optimal to replace the quadratic surface function is "long" in the direction of the minimum value of a second directional derivative and "narrow" in the direction of the maximum value of a second directional derivative. The optimum ratio $\rho$ defined by Nadler is

$$\rho = \sqrt{\frac{\lambda_2}{\lambda_1}} \tag{4}$$

where $\lambda_1$ and $\lambda_2$ are eigenvalues of Hessian of (3) [14].

[14] define a triangle aspect ratio as the ratio of principal axes of an ellipse with the smallest area, which passes through the triangle vertices. The smallest area of any circumelipses of the triangle is Steiner ellipse [15]. Although not explicitly stated, the authors handle with parameters of Steiner circumellipse. They also showed that the aspect ratio that results from minimizing the quadric error metric

$$\rho = \sqrt{\frac{\kappa_2}{\kappa_1}} \tag{5}$$

agrees with the optimum determined by Nadler. It is because eigenvalue $\lambda_1$ and $\lambda_2$ are the extreme values of normal curvature $\kappa_1$ and $\kappa_2$ and thus $\lambda_1 = \kappa_1$, $\lambda_2 = \kappa_2$ and equations (4) and (5) are equivalent. As mentioned before, memoryless simplification method minimizes metric with a great deal of similarity to the quadric error metric, so showed properties can be also applied to this method.

Extreme values of normal curvature $\kappa_1$ and $\kappa_2$ correspond to the principal axes of Dupin indicatrix. Their lengths are

$$r_1 = \frac{1}{\sqrt{\kappa_1}} \text{ and } r_2 = \frac{1}{\sqrt{\kappa_2}}. \tag{6}$$

The above facts that it is required to have a triangle shape with Steiner circumellipse corresponding to Dupin indicatrix of modeled surface in the triangle centroid. This

means that the above approach and so tested methods use the same characteristics as presented in chapter 2.3.

## RESULTS

We analyzed geometrical fidelity of the simplified models created by *quadric error metrics simplification* and *memoryless simplification.* We compared deviations in distinct levels of detail and a particular degree of random error in input data.

### Input

Artificial surface defined by mathematical function of two variables represented terrain surface. Its geometrical properties were similar to the geometrical properties of a common terrain surface. We used an artificial surface because of the possibility to calculate exact values of partial derivatives and morphometric variables. Then, calculated deviations are the real values of measurement without any other influence.

Input for simplification algorithms was a polyhedral model based on 2D Delaunay triangulation of a point set containing 68121 points. Each point was horizontally located in neighborhood of a regular grid node. The spatial distribution of input points was irregular, but they were distributed regularly over the surface. Each vertex had height calculated from chosen mathematical function. This point set contained points located directly on surface; there was no noise in their height values.

We created another two input point sets with random error in height value. Random error had normal (Gaussian) distribution with mean in real value. Two sets differed by size of random error determined by the standard deviation $\sigma$ from the mean. For one set standard deviation was $\sigma = 0.1$ m, representing heights with a small amount of noise. The second point set has value of standard deviation $\sigma = 0.5$ m, representing heights with larger noise.

### Simplified Models

We generated simplified models in four levels of detail. Models were consisted of 5,000 vertices in the first level. Models had 1,000 vertices in the second level of detail, while third level was 250. The roughest models had 100 vertices. Althogh the algorithms could not guarantee the exact number of elements in the output, they were within 1-2 vertices of the target.

We used each method for generating model in every level of detail from all three input surfaces (without the noise, with small noise, with large noise). This means that we had 12 models generated by *quadric error metrics simplification* and 12 models generated by *memoryless simplification* for geometric fidelity analysis. Created models are presented in appendix (fig. 5).

### Evaluation Criteria and Comparisons

We chose the normal deviation angle of a triangle $\delta n$ as main criterion which was justified previously. We calculated arithmetic mean $\delta n$ of $\delta n$ for each model where triangles in distance below 50 meters from the border were not counted. This eliminated the border-effect (impact of inappropriate triangles along borders).

The structure of deviations helps detect their weighted mean $\langle \delta n \rangle$. Used weight was the area of triangles. The difference compared to the arithmetic mean suggests whether deviations are caused by small triangles (if $\langle \delta n \rangle < \delta n$) or large triangles (if $\langle \delta n \rangle > \delta n$).

Most common method for evaluating the geometric fidelity of the simplified model compared to the original is *metro* presented in the work [16]. It is based on Hausdorff distance. One of the characteristics given by *metro* method is the mean distance $E_m$ between two surfaces. We used point sets of all input vertices (68121 points) of original model for calculating $E_m$.

We calculated these three parameters of geometrical fidelity for all simplified models. Values are presented in table 1 and fig. 4.

**Table 1.** Deviations of simplified models

| | | Memoryless simplification | | | Quadric error metrics simplification | | |
|---|---|---|---|---|---|---|---|
| Vertices | σ [m] | $\overline{\delta n}$ [°] | $\langle \delta n \rangle$ [°] | $E_m$ [m] | $\overline{\delta n}$ [°] | $\langle \delta n \rangle$ [°] | $E_m$ [m] |
| 100 | 0 | 3.27604 | 2.55572 | 1.79979 | 3.59204 | 2.88599 | 2.10333 |
| | 0.1 | 2.93885 | 2.57074 | 1.81072 | 3.28913 | 2.72489 | 2.05627 |
| | 0.5 | 3.33124 | 2.29989 | 1.70557 | 4.29413 | 4.02402 | 3.44086 |
| 250 | 0 | 2.34934 | 1.72327 | 0.63983 | 2.65791 | 2.04905 | 0.90578 |
| | 0.1 | 2.11402 | 1.72352 | 0.65308 | 2.45305 | 2.14613 | 1.00119 |
| | 0.5 | 2.31724 | 1.92601 | 0.68603 | 3.29024 | 3.13860 | 1.54991 |
| 1000 | 0 | 1.10905 | 0.89683 | 0.15340 | 1.35941 | 1.09395 | 0.22715 |
| | 0.1 | 1.18189 | 0.97909 | 0.16594 | 1.46036 | 1.35587 | 0.29514 |
| | 0.5 | 2.58018 | 1.77606 | 0.26611 | 1.88175 | 1.82220 | 0.43675 |
| 5000 | 0 | 0.53190 | 0.44372 | 0.03226 | 0.63173 | 0.53207 | 0.04846 |
| | 0.1 | 1.07265 | 0.83277 | 0.05657 | 0.90056 | 0.82604 | 0.06971 |
| | 0.5 | 6.22665 | 4.14721 | 0.22374 | 2.35582 | 2.03932 | 0.18641 |

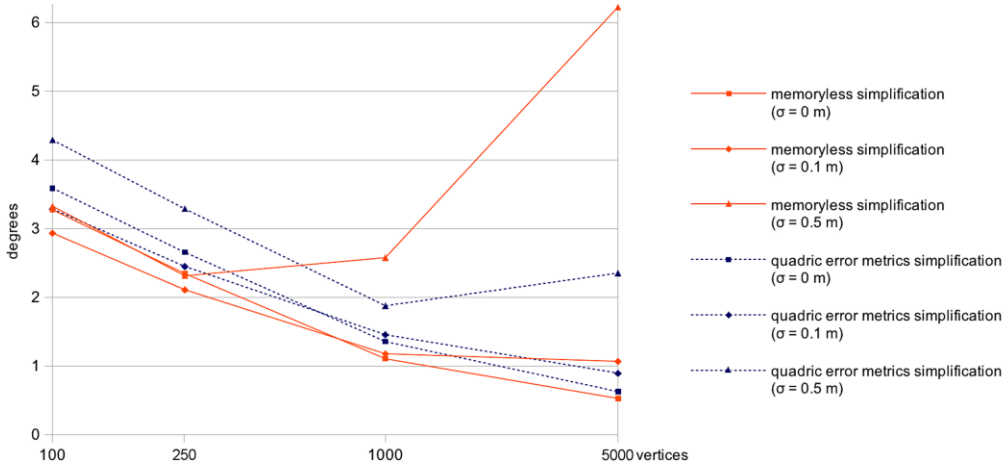



**Fig. 4.** Arithmetic mean of normal deviation angle of simplified models

As presented in the results, it is obvious that reducing the number of elements in the model decreases the geometric fidelity. This basic dependence affects the existence of random errors in input data. The random error of the heights affects derived geometric characteristics negatively at a higher density of points. That is why the same value of noise at the high number of vertices cause normal deviation increase. Input models made from 68121 vertices had mean normal deviations $\delta n = 0.24184\ ^\circ$ for $\sigma = 0$ m, $\delta n = 1.76868\ ^\circ$ for $\sigma = 0.1$ m and, $\delta n = 8.64665\ ^\circ$ for $\sigma = 0.5$ m. From calculated errors it is apparent fundamental difference between the two methods to cope with noise in the input data. The *quadric error metrics simplification* method eliminates the uncertainty caused by random errors significantly better.

The analysis pointed to an interesting finding, such that in both methods, simpler models with low noise achieve the highest accuracy. We believe that this is due to the occurrence of narrow triangles created by simplification in the case of accurate information about the surface shape. This is demonstrated by comparing the arithmetic mean and weighted mean, where all models without random errors of heights had the biggest differences. Small values of random errors do not mean a major loss of information and force the algorithm to approximate the shape of the surface, which is then more smooth. This has a positive impact on the arrangement of triangular elements in the more simplified model.

The most important finding from the analysis of geometric fidelity is that the *memoryless simplification* method achieves higher geometric fidelity than the *quadric error metrics simplification* method. This is reflected by all three parameters used. As mentioned before, the only exception is the ability to deal with random errors in the input data with a higher number of elements.

## CONCLUSION

Experimental results show the possibility to reduce the number of elements in the terrain model, while only to a relatively small extent reduces its geometric fidelity. We

thus obtain a model with a significantly more effective operational capabilities than an original model. On the other hand, with well-used method of simplification will be given only the most relevant elements, allowing reveals the fundamental structure of the terrain. Moreover, it is possible to get different levels of detail and then highlight the hierarchy.

It was shown that simplification also significantly reduces the negative impact of random component of error in the input data. Under certain circumstances, even the simplified model increases the accuracy of derived values describing the geometric structure of the terrain. Results were impressive. Simplified model with approximately 1 % of vertices had the same value of $\delta n$ as original model when $\sigma = 0.1$ m. Moreover, models consisting only of 100 vertices gives better results than the original model when $\sigma = 0.5$ m.

The selection of an appropriate method has a major impact on the accuracy of the model. Although the choice of two very similar methods processing the simplified model was different. Therefore, it is necessary to know the basic properties of algorithms for the proper selection of them. Based on the results, we recommend the *memoryless simplification* method. The same result found the authors in [7]. By their comparison of a number of methods, for application where overall geometric fidelity is required, the most recommended is *memoryless simplification* method and the second choice is *quadric error metrics simplification*. It also confirms our assumption that the methods, which algorithms work with the geometric properties of the optimal triangle will achieve the best geometric accuracy.

New contributions of this paper include: Analysis and finding the simplification methods using properties of the optimal triangle; Analysis of simplifying capabilities of terrain model by selected methods at various levels of detail; Analysis of simplifying capabilities of terrain model by selected methods with various levels of random errors in input data.

In future work, we want to focus on analysis and improvement, or supplement to methods in those places where they show their weaknesses. Overall accuracy would increase the elimination of the most significant deviations rearranging critical triangles. An appropriate tool could be the edge flip optimization with specially selected optimization parameters. The aim is to move closer to creation of process for constructing the triangular network optimized for terrain analysis.

## ACKNOWLEDGEMENT

Presented research was supported by grants VEGA 1/0469/10 and VEGA 1/0457/10 awarded by the Scientific Grant Agency VEGA of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences.

**REFERENCES**

[1] Luebke, D. (2001) A Developer's Survey of Polygonal Simplification Algorithms. IEEE Computer Graphics & Applications, 24-35.

[2] Krcho, J. (1999) Modelling of georelief using DTM - The influence of point configuration of input points field on positional and numeric accuracy. Geografický časopis, 51(3), 225-260.

[3] Krcho, J. (2001) Modelling of Georelief and Its Geometrical Structure Using DTM: Positional and Numerical Accuracy. Q111, Bratislava.

[4] Feciskanin, R. (2009) Optimization of triangulated irre¬gular networks for georelief modelling (in slovak). [Ph.D. thesis], Faculty of Science, Masaryk University, Brno.

[5] Garland, M. and Heckbert, P.S. (1997) Surface simplification using quadric error metrics. Proceedings of SIGGRAPH, 209-216.

[6] Lindstrom, P., and Turk, G. (1998). Fast and memory efficient polygonal simplification. Proceedings of the conference on Visualization '98, Computer Society Press, 279-286.

[7] Surazhsky V. and Gotsman C. (2005) A Qualitative Comparison of Some Mesh Simplification Software Packages. Technical Report, Mar 2005.

[8] Heckbert, P.S. and Garland, M. (1997). Survey of Polygonal Surface Simplification Algorithms. School of Computer Science, Carnegie Mellon University, Pittsburgh, 29.

[9] Fowler, R. J. and Little, J. J. (1979) Automatic extraction of irregular network digital terrain models. Computer Graphics 13(2), 199–207.

[10] Chen, Z. and Guevara, J. A. (1987) Systematic selection of very important points (VIP) from digital terrain model for constructing triangular irregular networks. Proceedings of Auto-Carto 8, Falls Church, ASPRS/ACSM, Baltimore, 50-56.

[11] Lee, J. (1989) A drop heuristic conversion method for extracting irregular network for digital elevation models. GIS/LIS '89 Proceedings, volume 1, ACSM, 30–39.

[12] Lindstrom, P., and Turk, G. (1999). Evaluation of memoryless simplification. Visualization and Computer Graphics, IEEE Transactions, no. 2, 98-115.

[13] Melissaratos, E.A. (1993) Lp Optimal d Dimensional Triangulations for Piecewise Linear Interpolation: A New Result on Data Dependent Triangulations, Utrecht Rijksuniversitteit.

[14] Heckbert, P.S. and Garland, M. (1999) Optimal triangulation and quadric-based surface simplification. Computational Geometry, 14(1-3), 49-65.

[15] Weisstein, E. W. (2005) Steiner Circumellipse. MathWorld. http://mathworld.wolfram.com/SteinerCircumellipse.html [cit. 2011-10-10].

[16] Cignoni, P., Rocchini, C. and Scopigno, R. (1998) Metro: Measuring Error on Simplified Surfaces. Computer Graphics Forum 17, 167–174.

## APPENDIX

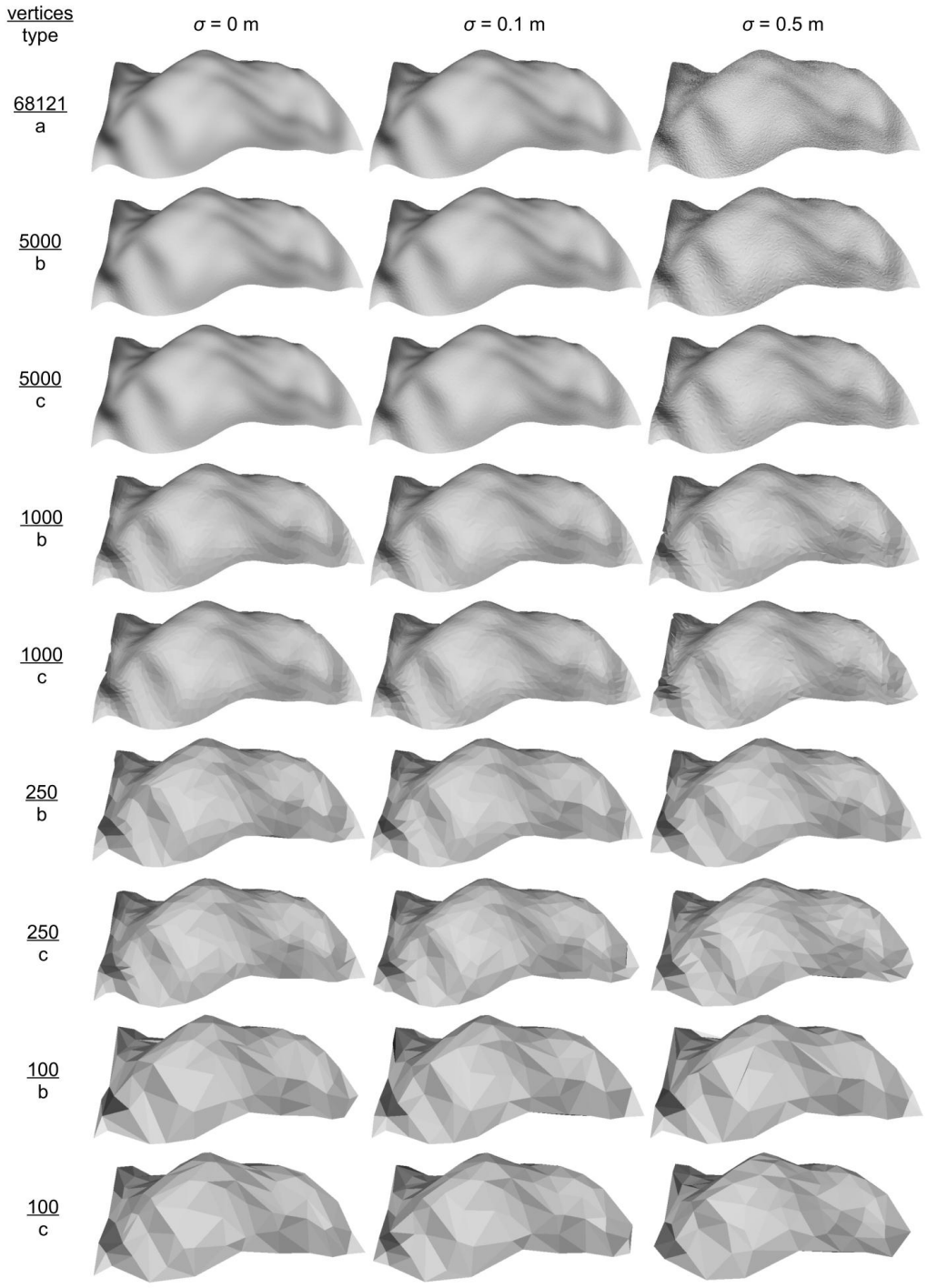



**Fig. 5.** Compared models

# QUALITY ASSESSMENT AND IMPROVEMENT OF ALS DATA ACQUIRED WITHIN THE PROJECT OF NEW HYPSOMETRY GENERATION OF THE CZECH REPUBLIC

Radek FIALA[1]

[1]Department of Mathematics, Faculty of Applied Sciences, University of West Bohemia,

Univerzitní 22, 306 14, Plzeň, Czech Republic

*fialar@kma.zcu.cz*

**Abstract**

Quality assessment of new digital terrain models covering the entire area of the Czech Republic is a prerequisite for their successful application in geosciences and various branches of national economy as well.

The paper presents "method of accuracy analysis of ALS points in close neighbourhood of check points" that is used for quality assessment of ALS data acquired within the Project of new hypsometry generation of the Czech Republic. The method respects considerable disproportion between number of ALS points and geodetically measured check points on control surfaces. Results of the method applied on check surfaces covering one third of the state territory is presented. The results are not related to final products of the Project (neither DMR 5G nor DMP 1G).

Another method leading to homogeneity improvement of ALS data is presented. The method assumes significant cross overlaps of neighbouring ALS strips. Its application removes local systematic height differences of ALS strips so that all details contained in the ALS data are preserved and no smoothing is performed. The same method with a minimal modifications can also be used for adjustment of neighbouring ALS blocks in areas of their overlaps. However the method is not intended for removing of any gross errors in ALS data. Within the Project of generation of new hypsometry of the Czech Republic the method has been applied in almost 20 % of the state territory.

**Keywords: digital terrain model, DTM quality assessment, airborne laser scanning, ALS strip adjustment**

## INTRODUCTION

Up to now the only digital terrain models (DTMs) covering the entire area of the Czech Republic are: DTM of Fundamental Base of Geographic Data (DTM ZABAGED®) maintained by the Czech Office for Surveying, Mapping and Cadastre and Digital Terrain Model of 3rd generation (DMR 3G) maintained by the Ministry of Defense. The DTM ZABAGED® originated from contour lines of Base Map of the Czech Republic in scale 1 : 10,000 periodically updated by means of stereo photogrammetry. The DMR 3G originated from stereo photogrammetric measurement only. Vertical accuracy of both DTMs is very similar: 1–2 m in open terrain and built-up areas, respectively 2–7 m in forested areas [1].

For many purposes, accuracy of the existing DTMs seems to be unsatisfactory. That is why significant improvement or generation of a new DTM was considered. As a result the Project of new hypsometry generation of the Czech Republic (Project) was prepared. There are three participants of the Project: the Czech Office for Surveying, Mapping and Cadastre, the Ministry of Defense and the Ministry of Agriculture. Important results of the Project will be Digital Terrain Model of 4th and 5th generation (DMR 4G, DMR 5G) and Digital Surface Model of 1st generation (DMP 1G). Anticipated vertical accuracy of the most precise DMR 5G is 0.18 m in open terrain and 0.3 m in forested areas. Similar accuracy of the DMP 1G will be 0.3 m in case of objects with well defined surfaces (e.g. building roofs) and 0.7 m in case of other objects [1].

Accuracy assessment of the data acquired within the Project is based on hundreds of control surfaces covered by grid of check points. The control surfaces are almost horizontal and vertically well-defined areas (e.g. football playground). ASPRS recommends at least 20 check points in each control surface [2]. The control surfaces used for accuracy assessment in the Project meets and often surpass this recommendation (100 check points in one control surface is common). Errors computed from data covering the control surfaces have to be considered as errors not including additional errors due to terrain slope and land cover (and coordinate transformation/conversion if applicable).

## METHOD OF ACCURACY ANALYSIS OF ALS POINTS IN CLOSE NEIGHBOURHOOD OF CHECK POINTS

For needs of the Project a new method of accuracy analysis of ALS points in close neighbourhood of the check points has been developed in Department of Mathematics of the University of West Bohemia in Pilsen. It is based on specific properties of control surfaces and check points. Particularly, the method requires a horizontal neighbourhood of a check point within the chosen diameter. The method respects a considerable disproportion between number of ALS points and geodetically measured check points on selected control surfaces. The method assumes that point coordinates are expressed in compound coordinate reference system including heights in any height reference system. Results of the processing are values of systematic error and random error, of course with respect to accuracy of check points.

The computation starts with discovering a set $L_i$ of ALS points in circular neighbourhood of a check point $r_i$ for all check points $r_i$, $I = 1, \ldots, n$, in the control surface. The points $L_i$ must not lay outside the control surface. Then a set $D_i$ of height differences between the check point $r_i$ and ALS points in set $L_i$ is computed. Union $D = D_1 \cup D_2 \cup \ldots \cup D_n$ contains height differences between all check points and ALS points in their neighbourhood. The systematic error $c_h$ is then computed as an average of the set $D$. Random error $\sigma_h$ is computed as a sample standard deviation.

The estimation of random error $\sigma_h$ allows detection of blunders. The method has been intended for fully automatic processing so the blunder detection must be automatic, too. Usual criterion "$k \,.\, \sigma_h$", assuming normal distribution of errors, has been used. The value of $k$ should depend on size of $D$ and on a selected probability of false detection of ALS point to be a

blunder. However, the value of $k = 3$ is usually used. After detection of blunders, the computation of errors is repeated without the blunders.

## METHOD FOR HOMOGENEITY IMPROVEMENT OF ALS DATA

In the course of ALS data processing in the Project, some imperfection in measuring ALS points heights has been detected. It results in local systematic height differences between overlapping ALS strips even after calibration of orientation parameters. The differences are small enough not to violate the accuracy assumptions but it could be serious in some applications, e.g. in visualization of flat areas. Figure 1 shows digital surface model of a storehouse created from two strips of ALS data with 0,08 m average height difference (in most areas the height differences between ALS strips were smaller). Further described procedure eliminates those height differences as illustrated in figure 2.

The process of eliminating local height differences between ALS strips (strip adjustment) is based on a difference model describing those height differences. Such difference model is computed for each pair of adjacent strips. The values of difference model are then split between two strips involved through the system of weight functions. Cross overlap of ALS strips is nearly 50 % in the Project. Therefore, most of the scanned area is covered by data in two adjacent ALS strips. Main steps of strip adjustment are:

- computing of initial values of the difference model,
- cleaning and smoothing of the difference model,
- adjustment of ALS point heights according to values in the difference model.

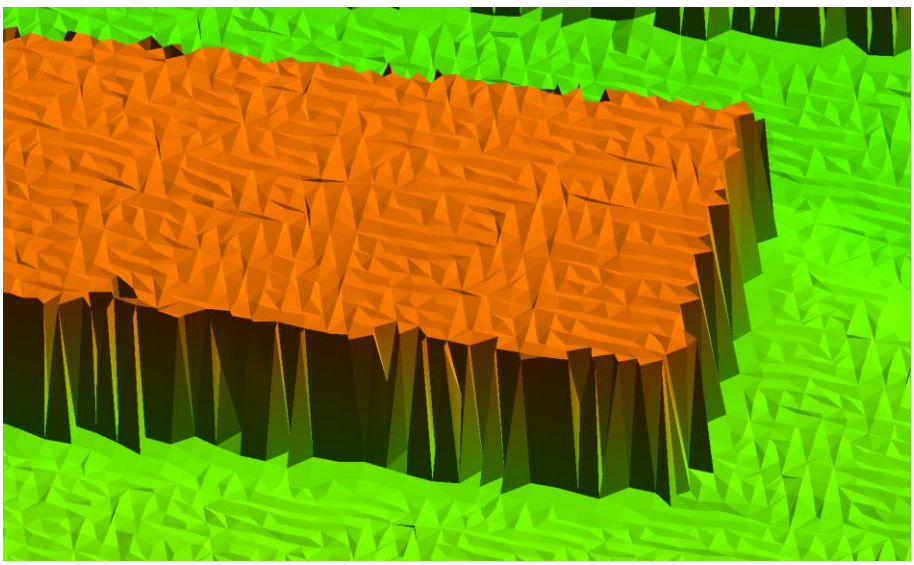


**Fig. 1.** Digital surface model from two overlapping ALS strips

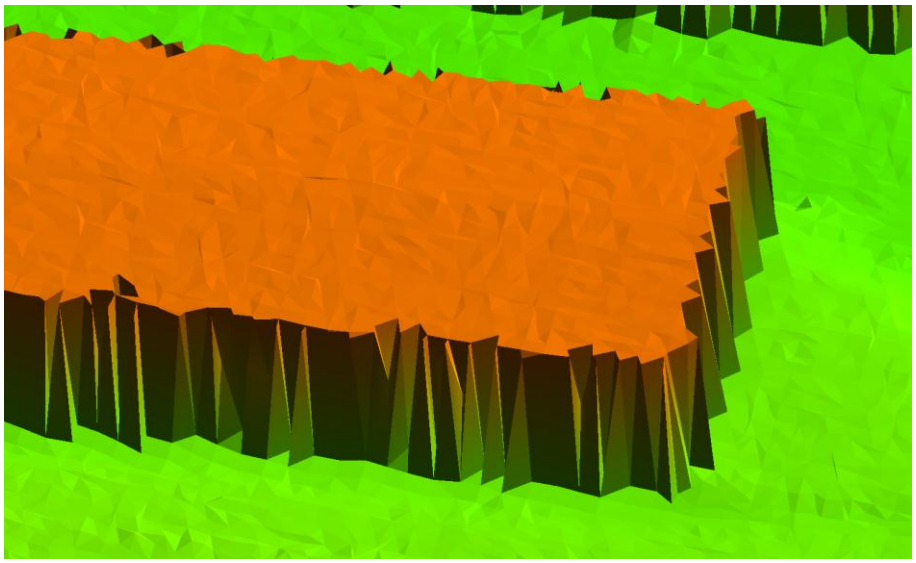


**Fig. 2.** Digital surface model from two overlapping ALS strips with adjusted height differences

**Computing of Initial Values of the Difference Model**

The ALS points in both strips are divided into the identical rectangular grid. For each cell of the grid estimations of mean height and standard deviation of heights of the ALS points is computed. A robust approach using lower and upper quartiles is used assuming symmetric probability distribution of point heights for mean height estimation computation and normal probability distribution for standard deviation estimation computation. Additionally a weight of the cell in each strip is computed depending on point density in the cell and estimation of standard deviation. The initial value of each difference model cell is computed as a difference of estimations of mean ALS point heights in both strips. The weight of the cell is set to a minimum of cell weights in both strips. From now on, the cell values and weights can be regarded as pixels of an image. Example of a difference model with pixel size 5 × 5 m can be seen in figure 3 (a cut 2 km long). Red (respectively blue) colour denotes positive (negative) values, colour saturation denotes absolute value of the cell. In light green areas, the values are too great or too less (> 0.2 m in this case; e.g. in forest areas). In white areas there are no data available. The dotted lines are planned flight lines.

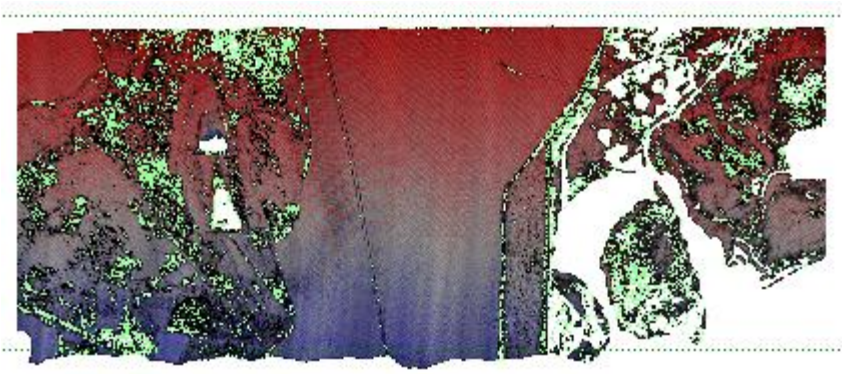


**Fig. 3.** Initial values of a difference model

### Cleaning and Smoothing of the Difference Model

Cleaning and smoothing uses well known algorithms of image processing that were generalized in order to encompass weights of the cells in the calculations. At first, an median filter is applied to the difference model. This filter removes extreme values in the difference model without influencing values of neighbouring cells. After applying the median filter the difference model still contains height difference discontinuities. Hence, a gaussian filter is applied to the difference model. This type of filter smooth out possible height discontinuities. Then pixel weights are removed and the image is filtered using the gaussian filter once again. The resulting difference model is very smooth as can be seen in figure 4.

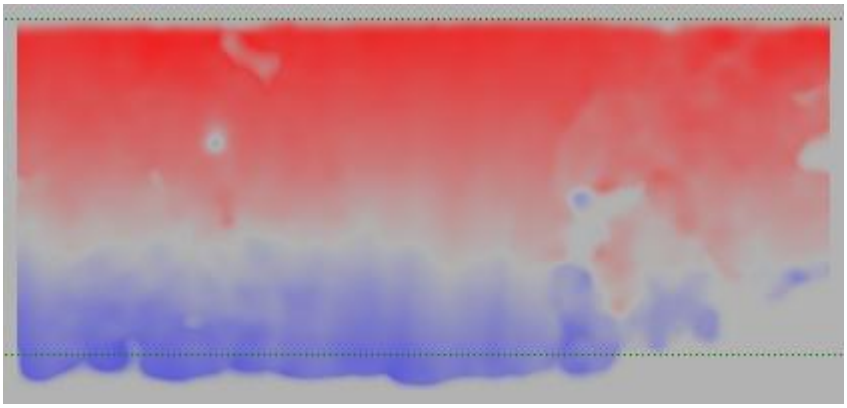


**Fig. 4.** Smoothed difference model

### Adjustment of ALS Point Heights According to Values in the Difference Model

The difference model contains characteristic height differences between two overlapping ALS strips. To eliminate those differences, the heights of points in both strips are corrected so that in particular position the sum of corrections in both strips is equal to value of difference model. Proper splitting of corrections between both strips is ensured by the weight functions. In the center of the strip (more precisely: under the planned flight line) the weight is 0. It means that there are no height correction in the center of the strip. At the edge of the strip (under the flight line of adjacent strip) and farther the weight is 1. The weight function should meet several

requirements to minimize strip height differences and to prevent inducing terrain break lines by adjusting height of ALS points. One of the simplest weight function meeting all requirements can be expressed as a polynomial of degree 3.

As an illustration of the effect of the method, the difference model without any smoothing has been computed after strip adjustment. This difference model is shown in figure 5.

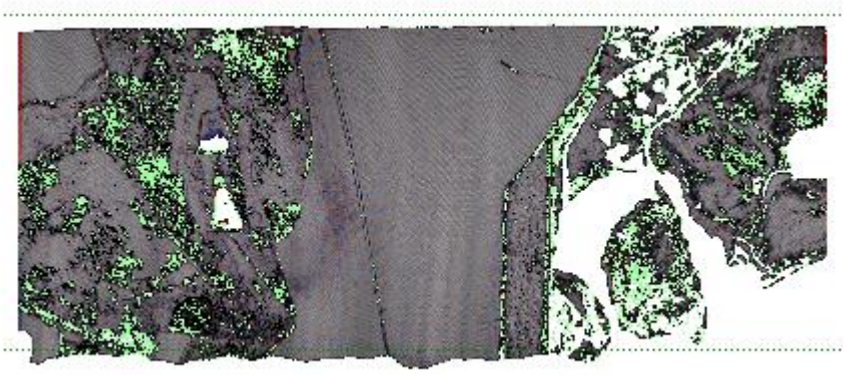


**Fig. 5.** Difference model (non-smoothed) after strip adjustment

This method of strip adjustment was applied by the Land Survey Office in Pardubice in almost 20 % of state territory in the course of the year 2011 with following parameters: pixel size 20 × 20 m, size of median filter 180 m, σ of Gaussian filters 75 m.

## RESULTS AND DISCUSSION

The method of accuracy analysis was implemented and used by the Land Survey Office in Pardubice. Analysis computed for 185 control surfaces covering approximately 1/3 of the state territory (denoted as Pásmo Střed) using 1 m perimeter of neighbourhood results in overall values: systematic error $c_h$ = −0,033 m and random error $\sigma_h$ = 0,048 m.

The strip adjustment method was applied at about 60 % of Pásmo Střed (15,400 km²). Though the strip adjustment method is not intended for removing errors, the values of both overall systematic error $c_h$ and overall random error $\sigma_h$ decreased after strip adjustment as shown in table 1. These results has been computed using the method of accuracy analysis of ALS points in close neighbourhood of check point using 98 control surfaces located in area where adjustment was applied. Changes of the errors for individual control surfaces are depicted in figure 6. As can be seen, random errors $\sigma_h$ calculated for most control surfaces decreased after strip adjustment. Unexpectedly, the value of systematic error $c_h$ decreased too. With respect to weight function role in strip adjustment, a dependency of systematic error $c_h$ on distance from flight line can be the reason for decreasing of systematic error $c_h$.

According to [3] the accuracy of a DTM from ALS data is influenced mainly by ALS point density and terrain slope. Assuming point density of 2.7 points/m² in open terrain and 0° terrain slope (as on control surfaces) the predicted value of random error $\sigma'_h$ = 0,037 m, thah is slightly lower than the computed value. The predicted value of random error is almost same as a computed value of random error after strip adjustment (0,035 m). The results of accuracy

assessment agree with theoretical expectations. According to [4] the projected overall height accuracy of the DMR 5G is reachable considering actual distributions of terrain slopes and forested areas in the Czech republic.

**Table 1.** Overall values of systematic error $c_h$ and average error $\sigma_h$ before and after strip adjustment (98 control surfaces located in area where the strip adjustment was applied)

| | Before strip adjustment | After strip adjustment |
|---|---|---|
| **Overall systematic error $c_h$** | −0,033 m | −0,020 m |
| **Overall random error $\sigma_h$** | 0,046 m | 0,035 m |

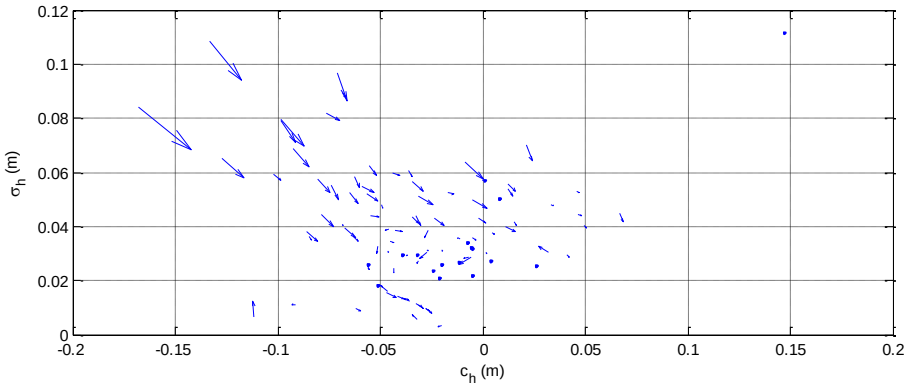



**Fig. 6.** Changes of systematic error $c_h$ and random error $\sigma_h$ before and after strip adjustment (98 control surfaces located in area where the strip adjustment was applied)

## CONCLUSION

The method of accuracy analysis of ALS points in close neighbourhood of check points used for quality assessment of ALS data in the Project of new hypsometry generation of the Czech Republic was introduced. Using this method overal values of systematic error $c_h$ = −0,033 m and random error $\sigma_h$ = 0,048 m was computed for 185 control surfaces covering approximately 1/3 of the state territory (Pásmo Střed).

The principles of method of strip adjustment has been introduced. The method eliminates local height differences between two adjacent strips ensuring that no induced terrain break lines arises. Further, all details of the relief within the ALS data are preserved because no smoothing is applied to the ALS data. Overall random error $\sigma_h$ = 0,046 m computed for 98 control surfaces located in area where the strip adjustment was applied decreased to $\sigma_h$ = 0,035 m after strip adjustment. Unexpectedly, value of overall systematic error decreased from $c_h$ = −0,033 m to $c_h$ = −0,020 m too.

## REFERENCES

[1] Brazdil, K. (2009) New Elevation Model Project of the Czech Republic Territory (in Czech). In Geodeticky a kartograficky obzor, 7, 145–151. ISSN 0016-7096

[2] ASPRS (2004) Guidelines Vertical Accuracy Reporting for Lidar Data. ASPRS. http://www.asprs.org/Division-General/LD-Airborne-Lidar-Committee-Downloads.html.

[3] Karel, W, Kraus, K. (2006) Quality parameters of digital terrain models In EuroSDR Official Publication 51. Utrecht: Gopher, 2006. 87–93. ISBN 90-5179-491-6.

[4] Sima, J. (2011) Contribution to the Accuracy Analysis of Digital Terrain Models Derived from Airborne Laser Scanning Data of Entire Territory of the Czech Republic (in Czech). In Geodeticky a kartograficky obzor, 5, 104–107. ISSN 0016-7096

# URBAN GROWTH MODELING USING GENETIC ALGORITHMS AND CELLULAR AUTOMATA; A CASE STUDY OF ISFAHAN METROPOLITAN AREA, IRAN

Ehsan FOROUTAN[1] and Mahmoud Reza DELAVAR[2]

[1]Dept of Surveying and Geomatics Eng. College of Eng. University of Tehran,
North Kargar St, 11155-4563, Tehran

*eforoutan@ut.ac.ir*

[2]Center of Excellence in Geomatics Eng. and Disaster Management,
Dept of Surveying and Geomatics Eng. College of Eng. University of Tehran,
North Kargar St, 11155-4563, Tehran, Iran

*mdelavar@ut.ac.ir*

**Abstract**

This study integrates cellular automata (CA) and genetic algorithms (GAs) to model urban growth in the Isfahan Metropolitan Area in Iran. The simulation of urban growth through cellular automata models brings improved understanding of the complex dynamic process of land use change, which can not be achieved through conventional models. The cellular automata (CA) as a powerful spatial dynamic modeling tool are designed as a function of parameters whose calibration plays a crucial role for obtaining a suitable set of parameters in order for precise and reliable modeling. Genetic Algorithms are useful tools for decreasing the search space for finding the optimal solution of transition rules in cellular automata and reducing the simulation uncertainties and improving its locational accuracy in urban modeling. The considered objective function in this algorithm is percent correct match (PCM) obtained from error matrix between the simulated and the reference map. Historical land use/cover data of Isfahan Metropolitan Area were extracted from the 1990 and 2001 Landsat ETM$^+$ images at 30m spatial resolution. Three different Moore neighborhood sizes have been considered for cellular automata model and simulation of urban growth for the year 2001 is performed. The simulation outcomes, evaluated with kappa statistic of 74.15% demonstrate that the integration of GA and CA could be suitable for dynamic urban growth modeling.

**Keywords: cellular automata, genetic algorithms, calibration, neighborhood size**

## INTRODUCTION

Developing methods for assessing different urban growth planning scenarios regarding the future consequences for land use and the progress of current spatial plans and policies is critically important for urban and regional planners [1]. An urban land use system is dominated by human activities with complex spatio-temporal dynamics [2]. Many efforts have been made to model urban growth using different models and algorithms but amongst them cellular automata models have proved rather popular as frameworks for modeling and simulating the physical growth of cities. The simulation of urban growth through cellular automata models brings improved understanding of the complex dynamic process of land use change, which cannot be achieve through conventional models [3]. CA was originally developed for

simulating complex systems in physics and biology. CA systems were first developed in the late 1940s by S. Ulam and J. von Neumann and then at first serious research developed by [4] demonstrated that complex natural phenomena can be modeled by CA models. The first effort to employ CA for modeling urban growth refers to a pioneer work of [5] who proposed the application of cellular space models to geographic modeling. Because of spatial nature of CA they could gain attention between researchers and urban planners, as a large and increasing volume of work shows that CA are proper tools for modeling spatial dynamics (e.g. [1], [6], [7], [8], [9], [10], [11], [12] and [13]. However, even though technically there is little limitation developing CA models within a GIS environment, it remains a research issue to urban modelers to identify suitable transition rules and their defining parameters in CA based urban modelling [14]. In CA, many variables are involved and each variable is usually associated with a parameter that indicates its importance in simulation [15]. These parameters significantly affect the outcomes of urban simulation [13]. It is essential to define proper parameter values when CA is used to simulate realistic cities. Only through calibration the cellular automata model can produce an urban level and urban pattern close enough to reality [16]. In other words, the purpose of calibration is to establish the relationship between land use change and the factors that affect probability of land conversion [17]. Calibration in cellular automata modeling intends to find the best value combination for transition rules to match simulated and real urban phenomenon quantitatively and qualitatively [18]. A number of cellular automata calibration methods have been developed for urban growth modeling. They achieved various levels of success and efficiency. Generally, calibration methods are divided into three categories including statistical, visual and artificial intelligence tools [18]. [19] calibrated the SLEUTH model by using visual and statistical tests to find the best values for the five growth parameters (slope, land use, exclusion, urban extent, transportation, and hill shade). In addition [20] uses brute force method to find the best parameter of SLEUTH model of urban growth. Such calibration methods take expanded CPU time to reach the most appropriate parameter set in the search space and may get trapped in a local maximum. In recent years, artificial intelligence techniques have attracted considerable attention research in urban modeling for many reasons. Intelligence techniques such as artificial neural networks (ANNs), fuzzy logic, and genetic algorithms (GAs) are popular tools, since they can deal with complex engineering problems which are difficult to solve by classical methods. In case of ANNs, [13] proposed a model which employed ANNs as calibration tool of cellular automata but unlike traditional CA models it generate implicit rules that could not be interpretable. In case of Gas, initial efforts tried to attach genetic algorithms to cellular automata urban model design for performance improvement. [21] modeled the changes in land-use for Rome, Italy using genetic algorithms to produce a new set of rules for the cellular automata model. In addition Genetic algorithms were used to find the optimal set of possibilities of land-use planning for Provo, Utah [22]. A recent study tried to formalize genetic algorithms as a calibration tool for the SLEUTH model [20]. In [16] proposed a model to enhance the efficiency of transition rule calibration in cellular automata urban growth modeling using GA for Indianapolis.

This paper applies genetic algorithm as an artificial intelligence technique for calibration of cellular automata transition rules for urban growth modeling in Isfahan Metropolitan Area of

Iran from 1990 to 2001. The model validation is performed employing kappa statistic and overall accuracy to test the model ability to simulate urbanization pattern of reality.

## STUDY AREA

The Isfahan Metropolitan Area in the center of Iran is considered in this study for modeling urban growth. The data that have been used for calibration and simulation included two historical satellite images covering a period of ten years. These raw images include two 28.5 m resolution $ETM^+$ images 1990 and 2001.

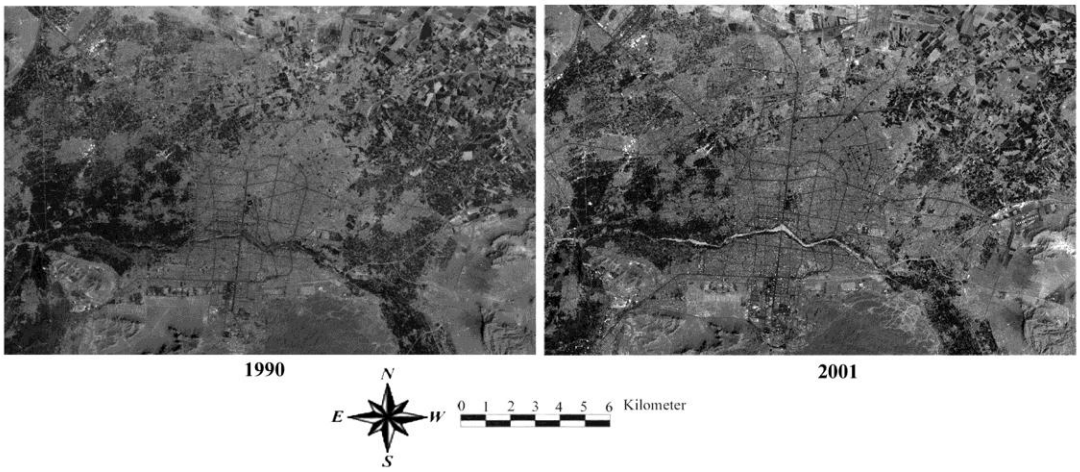



Fig. 1. Satellite images of Isfahan in 1990 and 2001, respectively.

Both images were geometrically rectified and registered to the Universal Transverse Mercator (UTM) WGS 1984 zone 39N. Registration errors were about 0.50 pixels. In addition, combinations of RGB bands of Landsat images were performed to prepare satellite imageries for better classification. Fig. 1 shows the final results for the 1990 and 2001 Landsat images of Isfahan. After rectifying and registering the images, the next step is classification of the outcomes which are the inputs to a cellular automata model. All land use classes of Isfahan were also reclassified from their original classification to Anderson Level I [23] for the modeling exercises. Four classes are defined based on maximum likelihood classification system namely water, road, urban, vegetation area and barren. These land uses considered as effective land uses for urban growth as shown in Fig. 2.

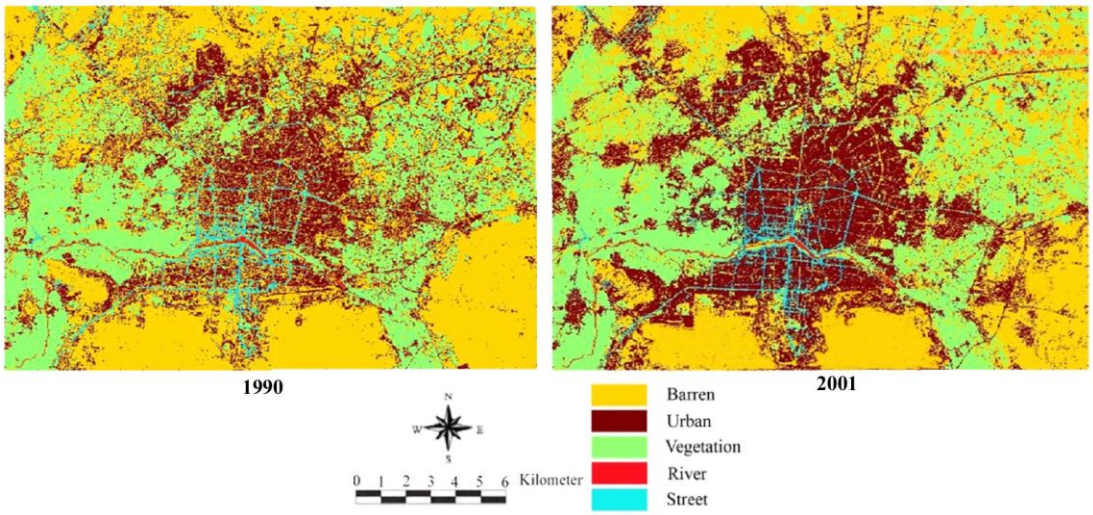



Fig. 2. Isfahan classified land use map in 1990 and 2001, respectively

## METHODOLOGY

Calibration of cellular automata plays a crucial role in a precise and reliable modeling. In this section we address the structure and function of the proposed model. Land uses extracted from satellite images in previous section are: urban, road, vegetation, water and non-urban (barren) areas. Each of these land uses has its effect on urban growth that importance of each, varies from a region to another. These preferences have been considered in a set of rules. For example some people prefer to inhabit in areas near to the vegetation areas while the other prefers have an access to streets. Using cellular automata transition rules, we will be able to extract these rules and use them for modeling and forecasting urban growth. The proposed transition rules for the cellular automata urban growth model are designed using the above mentioned input to identify the minimum urbanization conditions needed in a Moore neighborhood (square) for a test pixel to become urban or not. These transition rules are summarized as follows:

1. IF the test pixel is water, road, vegetation OR urban THEN no change.
2. IF the test pixel is non-urban (barren) THEN it becomes urban if:
   - Its neighboring URBAN pixel count is ≥ No_u,

     OR
   - Its neighboring ROAD pixel count is ≥ No_s,

     OR
   - Its neighboring VEGETATION pixel count is ≥ No_v,

     OR
   - Its neighboring WATER pixel count is ≥ No_w.

where No_r, No_s, No_P and No_l are the minimum number of urban, street, vegetation and water pixels, respectively in a square neighborhood which are require for urbanization of one non-urban pixel. In this research we consider three different neighborhood sizes (3×3, 5×5 and 7×7) and finally compare the obtained results from them and select the best one for our study area. There are a total $9^4$, $25^4$ and $49^4$ combinations of possible rule values for 3×3, 5×5 and 7×7 neighborhood sizes, respectively. These extensive search spaces entail us to use optimum search methods such as genetic algorithms for reducing the search spaces and finding the optimum solutions.

For evaluating each generated population, we have to use model validation parameters. The objective function used in this study is percent correct match (PCM) and we are going to minimize (100-PCM). The percent correct match (PCM) is a parameter that indicates the accuracy of the model using confusion matrix. This parameter assesses the following parameters: (1) the urban growth is occurred in real world and the model illustrates it, (2) the urban growth is not occurred in real world while the model has shown the growth. The confusion matrix can summarize the results by overlaying the map of a simulated land use on the map of reality, as shown in Table 1 [24].

**Table 1**. Confusion Matrix

| Model | Reality | | |
|---|---|---|---|
| | Change | Non-change | Total |
| Change | A | B | A+B |
| Non-change | C | D | C+D |
| Total | A+C | B+D | A+B+C+D |

The percent correct match (PCM) is calculated from (1) (Pontius and Schneider, 2001):

$$PCM = \frac{A + D}{A + B + C + D} \quad (1)$$

Genetic algorithms (GAs) were invented by John Holland in the 1960s [25] and were further developed by him and his colleagues at the University of Michigan in the 1960s and the 1970s [26]. The general algorithm of this approach is as follows:

```
Algorithm: Genetic Algorithms
```

- `Initialize population`
- `Evaluate population`
- `Do while (termination-criteria is not satisfied)`
- `Select parents for reproduction`
- `Perform recombination and mutation`
- `Evaluate`

```
Loop
```

The first step towards applying genetic algorithms is generating initial population. Population for calibrating the proposed cellular automata model includes the number of urban, street, vegetation and water pixels in a defined square neighborhood. As an example, a rule as a member of population could be 2, 4, 5, 7 which means that a non-urban pixel will become urban if in its neighborhood there are at least 2 or 4 or 5 or 7 urban, street, vegetation or water, respectively. Proposed population size is 20 which mean that the population includes 20 of the mentioned rules that are randomly generated. Then the cellular automata model is run for each member in the population and the fitness of the population is calculated through the objective function. If each of the rules satisfies the termination criterion, it will be selected as the best answer and algorithm will be stopped; otherwise the rules with the lower objective function are given a lower rank and vice versa. For using genetic operation such as elitism, crossover and mutation, population must be sorted descending in term of its fitness. The next population will be generated according to elitism selection, crossover and mutation. Based on the proposed algorithm, 10% of the sorted population in the previous step is copied directly into the next generation before crossover and mutation. This step allows us to keep the best answers. The next step in genetic algorithms is crossover and mutation applied to population from the previous step. Crossover is the process in which two chromosomes (strings) combine their genetic materials to produce a new offspring which possesses both their characteristics. In Fig. 3 crossover operation on the two parents are shown that leads to producing two offsprings. In this paper crossover rate was 0.60 which means that crossover applies on the 60% of population.

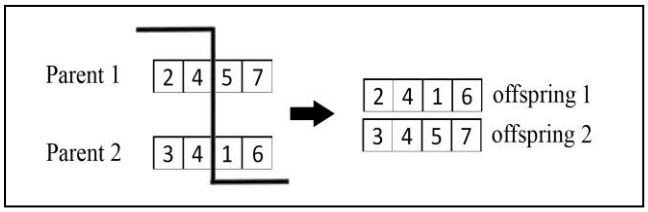



Fig. 3. Cross Over operator

The last step in finalizing a new population is mutation operation. It is the process by which a string is deliberately changed so as to maintain diversity in the population set and prevents the solution from local minima. In this study, mutation applied to the 30% of the population by generating a set of random solutions. By applying elitism selection, crossover and mutation, new population is produced and could be evaluated using the objective function. This process is repeated until termination criterion is satisfied.

## RESULTS AND DISCUSSION

Table 3 shows the results acquired through applying genetic algorithms as a tool for calibrating cellular automata transition rules. Calibrated rules column in Table 3 indicate the minimum required number of urban, street, vegetation and water pixels in a defined square neighborhood, respectively.

Aside from calibration of cellular automata, validation is other issues that must be addressed. Validation procedure measure the performance of CA model to simulate a reliable model that can be reproduced urbanization pattern of reality. A well known measurement procedure is based on the use of error matrices for mapping comparisons [27]. For evaluating the model accuracy, Kappa statistics (KS) is used. The Kappa statistics is much used to assess the similarity between the observed and predicted results. The calculation of Kappa is based on contingency table [28] (Table 2).

**Table 2**. The Contingency Table

| | | Model | | | | Total |
|---|---|---|---|---|---|---|
| | Class | 1 | 2 | … | c | |
| Reality | 1 | $P_{11}$ | $P_{12}$ | … | $P_{1c}$ | $P_{1T}$ |
| | 2 | $P_{21}$ | $P_{22}$ | … | $P_{2c}$ | $P_{2T}$ |
| | … | … | … | … | … | … |
| | c | $P_{c1}$ | $P_{c2}$ | … | $P_{cc}$ | $P_{cT}$ |
| Total | | $P_{T1}$ | $P_{T2}$ | … | $P_{Tc}$ | 1 |

On the basis of the contingency table, many statistics can be derived as follow [29]:

$$P(A) = \sum_{i=1}^{c} P_{ii}$$
$$P(E) = \sum_{i=1}^{c} P_{iT}.P_{Ti} \quad (2)$$

And finally Kappa statistics could be calculated from (3) [29]:

$$KS = \frac{P(A) - P(E)}{1 - P(E)} \quad (3)$$

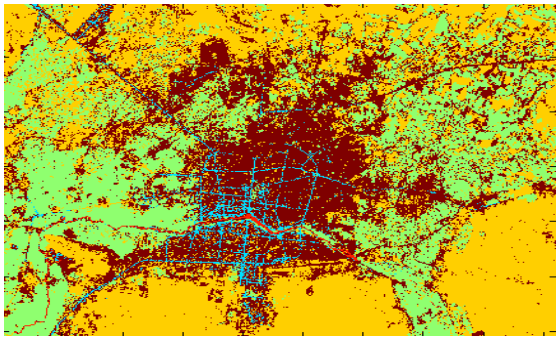
It is generally considered that Kappa values for map agreement are: >0.8 is excellent; 0.6-0.8 is very good; 0.4-0.6 is good; 0.2-0.4 is poor and <0.2 very poor [30].

**Table 3**. Results of Cellular Automata Calibration

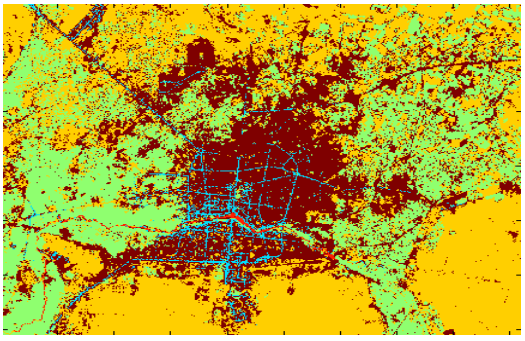
| Windows Size | PCM% | KS% | Calibrated Rules |
|---|---|---|---|
| 3 | 86.28 | 72.38 | 4 , 1 , 8 ,2 |
| 5 | 87.17 | 74.15 | 11 , 3 , 23 , 12 |
| 7 | 86.49 | 72.65 | 25 , 12 , 43 , 25 |

As the results show (Table 3), percent correct match (PCM) and Kappa statistics calculated from confusion matrix and contingency table, indicates that using a square neighborhood size of 5×5 i.e. in a neighborhood of 150×150m is more suitable than 3×3 or 7×7 neighborhood size. Based on the earlier work [30] Kappa values calculated for simulated land use map

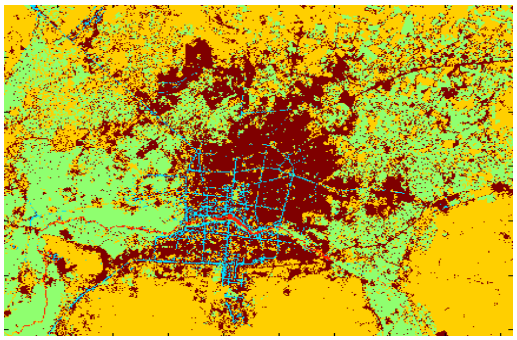
seems very good. The simulated land use map of Isfahan Metropolitan Area is shown in Fig. 4.



**Simulated land use map using 3×3 neighborhood size**



**Simulated land use map using 5×5 neighborhood size**



**Simulated land use map using 7×7 neighborhood size**

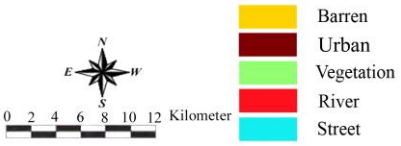



Fig. 4. Isfahan simulated land use maps in 2001 using different neighborhood sizes.

As mentioned, the calibrated rules are the minimum required number of pixels of land uses for urbanization of a non-urban pixel. These rules as a set of numbers can be interpreted as follows: In case of window size of 5 the calibrated cellular automata transition rules are 11, 3, 23, and 12. It means that in the Moore neighborhood of a non-urban pixel must be at least 11 pixels of urban land use, or 3 pixels of street land use, or 23 pixels of vegetation land use and or 12 pixels of water land use. Therefore, it can be drawn that the order of urbanization preference is street, urban, water and vegetation land uses.

## CONCLUSIONS

In this paper, the calibration procedure of the cellular automata urban model was outlined along with the results for the Isfahan Metropolitan Area, Iran. Since cellular automata urban models involve large number of variables, it needs an effective method for calibration. This exhaustive search for finding the optimum solutions will be reduced only through

metaheuristics methods such as genetic algorithms. Genetic algorithms due to the effective search and high performance could be a suitable method for calibrating urban cellular automata. The paper verified that the neighborhood size of the cellular automata model depends on the urban structure of study area and varies from one region to another. Also a larger neighborhood size generates a smoother pattern compared to the smaller one. Future works encourage doing the calibration process in spatial units smaller than townships and acquiring the model parameters to test the effect of spatial modeling unit size on the reliability of modeling.

## REFERENCES

[1] Al-Ahmadi, K., See, L., Heppenstall, A. and Hogg, J. (2009) Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. Ecological Complexity, 6(2), 80-101.

[2] Hu, Z., and Lo, C. (2007) Modeling urban growth in Atlanta using logistic regression. Computers, environment and urban systems, 31(6), 667-688.

[3] Batty, M., Xie, Y. and Sun, Z. (1999) Modeling urban dynamics through GIS-based cellular automata. Computers, environment and urban systems, 23(3), 205-233.

[4] Wolfram, S. (1984) Cellular automata as models of complexity. Nature, 311(5985), 419-424.

[5] Tobler, W. R. (1970) A computer movie simulating urban growth in the Detroit region. Economic geography, 46, 234-240.

[6] Alkheder, S. and Shan, J. (2005) Cellular Automata Urban Growth Simulation and Evaluation-A Case Study of Indianapolis, Geomatics Engineering, School of Civil Engineering, Purdue University.

[7] Couclelis, H. (1985) Cellular worlds: a framework for modeling micro-macro dynamics. Environment and Planning A, 17(5), 585–596.

[8] Dragicevic, S. (2004) Coupling fuzzy sets theory and GIS-based cellular automata for land-use change modeling. Paper presented at the Fuzzy Information.
Processing NAFIPS '04. IEEE Annual Meeting of the Processing NAFIPS'04, Banff, Canada.

[9] Itami, R. M. (1994) Simulating spatial dynamics: cellular automata theory. Landscape and Urban Planning, 30(1-2), 27-47.

[10] Portugali, J. and Benenson, I. (1995) Artificial planning experience by means of a heuristic cell-space model: simulating international migration in the urban process. Environment and Planning A, 27, 1647-1647.

[11] Wu, F. (1998) An experiment on the generic polycentricity of urban growth in a cellular automatic city. Environment and Planning B: Planning and Design, 25(5), 731-752.

[12] Wu, F. (1998) Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. Journal of Environmental Management, 53(4), 293-308.

[13] Yeh, A. G. O. and Li, X. (2002) Urban simulation using neural networks and cellular automata for land use planning. Paper presented at the Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada

[14] Feng, Y., Liu, Y., Tong, X., Liu, M. and Deng, S. (2011) Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. Landscape and Urban Planning, 102(3), 188-196. doi: 10.1016/j.landurbplan.2011.04.004

[15] Li, X., Yang, Q. S. and Liu, X. P. (2007) Genetic algorithms for determining the parameters of cellular automata in urban simulation. Science in china series d Earth sciences-english edition, 50(12), 1857.

[16] Shan, J., Alkheder, S. and Wang, J. (2008) Genetic algorithms for the calibration of cellular automata urban growth modeling. Photogrammetric Engineering & Remote Sensing, 74(10), 1267-1277.

[17] Wu, F. (2002) Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science, 16(8), 795-818.

[18] Al-Kheder, S. A. (2007) Urban growth modeling with artificial intelligence techniques. PhD Thesis, Purdue University.

[19] Clarke, K. C., Hoppen, S. and Gaydos, L. (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning B, 24, 247-262.

[20] Goldstein, N. C. (2004) Brains vs. Brawn - comparative strategies for the calibration of a cellular automatabased urban growth model (Vol. 3): Boca Raton, FL: CRC Press.

[21] Colonna, A., Di Stephano, V., Lombardo, S., Papini, L. and Rabino, A. (1998) Learning urban cellular automata in a real world: The case study of Rome metropolitan area. Paper presented at the ACRI'98 Third Conference on Cellular Automata for Research and Industry, Trieste.

[22] Balling, R. J., Taber, J. T., Brown, M. R. and Day, K. (1999) Multiobjective urban planning using genetic algorithm. Journal of urban planning and development, 125, 86.

[23] Anderson, J. R., Hardy, E. E., Roach, J. T. and Witmer, R. E. (1976) A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964. US Geological Survey Circular, 671, 7-33.

[24] Pontius, R. G. and Schneider, L. C. (2001) Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment, 85(1-3), 239-248.

[25] Holland, J. H. (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence: The MIT press.

[26] Goldberg, D. E. (1989) Genetic algorithms in search, optimization, and machine learning: Addison-wesley.

[27] Norte Pinto, N. and Pais Antunes, A. (2007) Cellular automata and urban studies: a literature survey. ACE: Arquitectura, Ciudad y Entorno, núm. 3, Febrero 2007.

[28] Monserud, R. A. and Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. Ecological Modelling, 62(4), 275-293.

[29] Hagen, A. (2003) Multi-method assessment of map similarity. International Journal of Geographical Information Science, 17(3), 235-249.

[30] Pijanowski, B. C., Pithadia, S., Shellito, B. A. and Alexandridis, K. (2005) Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. International Journal of Geographical Information Science, 19(2), 197-215.

# OPTIMAL INTERPOLATION OF AIRBORNE LASER SCANNING DATA FOR FINE-SCALE DEM VALIDATION PURPOSES

Michal GALLAY[1], Chris LLOYD[2] and Jennifer MCKINLEY[2]

[1]Ústav geografie, Prírodovedecká fakulta, Univerzita Pavla Jozefa Šafárika,
Jesenná 5, 040 01 Kosice, Slovakia
*michal.gallay@upjs.sk*

[2]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast,
BT7 1NN Belfast, Republic of Ireland
*c.lloyd@qub.ac.uk, j.mckinley@qub.ac.uk*

**Abstract**

Airborne laser scanning (ALS) is becoming a widely used method for detailed remote sensing across European and other countries. It provides dense sets of highly accurate data points which can be further used for creation of digital elevation models (DEMs). However, the high density also determines higher processing power demands. In case grid-based DEMs are concerned and interpolation is required, it is therefore useful to test the performance of particular interpolation methods using smaller samples in order to find the suitable approach and provide guidelines for other users. In this paper, inverse distance weighting (IDW) and regularized spline with tension and smoothing (RST) were tested to find the optimal DEM based on last return airborne laser scanning (ALS) data. Data from four sites differing in terrain complexity were used. Summary statistics of cross-validation errors were used to select the parameters for generating the optimal ALS DEM. The results show that IDW and RST generate very similar cross-validation statistics. Cross-validation errors increase with increasing short-range variation of elevation for IDW as well as for the RST interpolation method. Rescaled tension between 200 – 400 with smoothing provided acceptable cross-validation statistics for RST. Supplemental validation of the interpolated ALS DEMs against ground surveyed measurements provided means for absolute accuracy assessment. It suggests that both IDW and RST are equivalent for the outlined purpose. For IDW being a simpler method demanding less processing time, this was chosen for deriving the optimal ALS based DEM. The choice of interpolation method is less influential when a surface is interpolated to coarser or similar resolutions than the resolution of the input data which supports finding by other authors.

**Keywords: inverse distance weighting, spline, laser scanning, lidar, interpolation**

## INTRODUCTION

Airborne laser scanning (ALS) can be considered as the most accurate method for mapping land surface by remote sensing. It provides rapid and dense collection of data points with submeter to subdecimeter measurement precision. Such data properties are difficult to achieve in an efficient way by other remote sensing methods such as photogrammetry,

synthetic aperture radar (SAR). ALS is capable of collecting altitude of several surface levels depending on the penetration of laser beam down through the ground. The height data are further used to generate digital elevation models (DEM) and the recorded intensity of the backscattered laser beam can be used for classification of surface objects. Generally in vegetated areas, the first returns correspond to the upper landscape canopy level (e.g. vegetation tops) and the last returns to the ground (terrain surface). While the first returns are used to generate digital surface models (DSM) the last returns are used for generation of digital terrain models (DTM). In cases, where impenetrable objects such as buildings are present or the ground is exposed the first returns refer to the last returns. Detailed background on ALS can be found in [1] or [2]. Nowadays, ALS as a technology is becoming more accessible for a wide range of users who deal with the problem of effective processing of millions of data points. This has also stimulated research on suitable approaches for generation of digital elevation model (DEM) from the ALS data [3], [4] and [5]. The studies compared performance of several interpolation methods used to derive gridded DEMs for a certain region. Comparison of various kinds of interpolation methods for DEM generation for various terrain types is presented in [6] or [7] who, however, used digitized contour line data. A systematic methodology for assessment of ALS DEMs is outlined for example in [8].

The paper presented here extends the findings of the previously mentioned authors by (i) considering four different types of terrain complexity (flat, undulating, hilly, mountainous), (ii) comparing inverse distance weighting and regularized spline with tension and smoothing as described in [9] and [10]. The aim is to find an optimal DEM derived from the ALS last return data which can be further used as a reference DEM ('ground truth') for assessing vertical accuracy of DEMs derived from data acquired by different methods (contour digitizing, photogrammetry, SAR interferometry) which was done in [11].

## STUDY SITES AND DATA

It is fortunate that different types of fine-scale digital elevation datasets exist for the territory of Great Britain which is why this European region is the focus. Not only the terrain but also topographic surface data exist on a national basis. Up to date ALS, SAR interferometry, photogrammetry and digitizing of topographic maps have been used to generate nation-wide relatively high-detail DEMs. For that reason it was possible to compare data of a similar scale overlapping over different types of landscapes which is presented in [11]. As Fig. 1 shows four sites were selected in England where the data generated using the different acquisition methods overlap in space. In this paper ALS last return data for the sites were used for the object of the analyses. Their extent was adjusted so that all data types overlap at the site and variability of terrain is represented sufficiently. The maximal extent of the sites was limited to 2 km by 2 km with respect to the intended working scale of 1:10 000 (5 metre cell size) for the analyses. This is a manageable size if one calculates that full coverage at 5 m cell size means million data points to be analysed bearing in mind that permutations of this number will be involved.

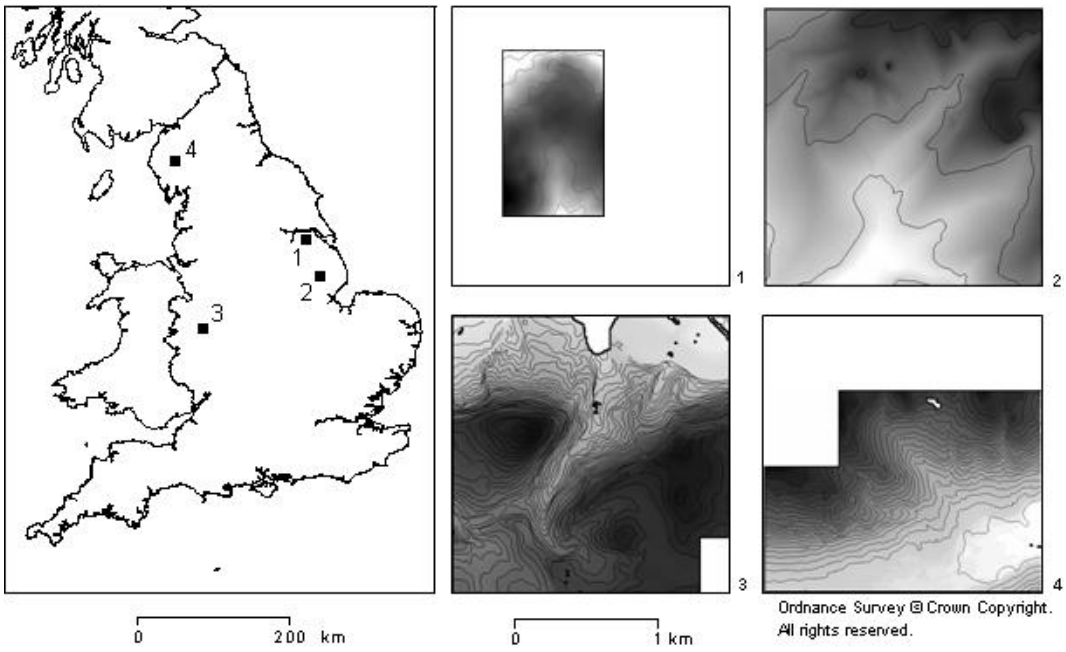



**Fig. 1.** The location and extent of the study sites in Great Britain. 1 - Barton upon Humber (flat terrain), 2 – Lincoln (undulating terrain), 3 – Telford (hilly terrain), 4 - Great Langdale (mountainous terrain). Contour interval 1, 2, 3 is 5 metres, for 4 it is 20 metres. Elevation increases from white to black shade.

The data were provided by the Environment Agency UK (http://www.environment-agency.gov.uk) and they originated during missions flown by the in July 2003, March 2001, August 1999, December 2000, respectively for the site 1 to site 4.

The data represent last return echoes which are considered in the paper as terrain surface samples. The Agency is responsible for official management of flood defence, visualising flood risk and flood forecast. For that reason, the Agency opted for ALS acquisition of high accuracy elevation data. The ALS missions have been flown since 1998 by plane.The measurement accuracy stated by the provider is 0.25 meters (1σ) for a flat open ground and the points have 2 meters approximate spacing. The main constraint of the study sites selection was the spatial coverage of the datasets at the time of the research. The coverage of the airborne ALS data was the main limitation in choosing the particular sites. The ALS data acquisition by the Environment Agency was limited to the regions of the Great Britain under potential risk of flooding such as coastal zones, valley floors, villages, towns and cities. For that reason, the sites were chosen on the basis of expert judgement considering the condition of differing landscape types. Providing all DEM data types have full national coverage stratified random sampling could be used for the selection.

Another dataset used for the purposes of this paper comprised elevation data collected with a total station in the Great Langdale valley (site 4). The measurements were used to assess the accuracy of the ALS points of the last return (see [12]) and to assess the DTMs produced

within the tuning of IDW and RST parameters. The ground survey data were acquired for two sections representing different types of terrain situated at the Middle Fell farm (inclined slope) and at the Rossett Bridge (flat alluvial plain). The height was measured every 2-5 metres on approximately a regular pattern with a denser sampling on terrain break lines. Specifications of the datasets can be found in [12] or [11]. However, it would be more appropriate to use data spread across the whole area, since the original purpose of the ground survey was to look at the performance of the ground survey methods on a finer scale rather than collect the ground truth for validating the airborne data.

## METHODS

Finding the optimal interpolation parameters for generation of ALS DTM involved tuning the parameters of two interpolation methods. Inverse distance weighting (IDW) and regularized spline with tension and smoothing (RST) were chosen and tested to find the optimal DTM. IDW is a standard and fast, although not a very flexible method of local spatial prediction. The main principle is in assigning weight in the interpolation equation based on the distance to the prediction location. More details on the method can be found for example in [4]. On the other hand, calculations with RST are more complex than with IDW and for that reason the method has higher processing time demands. RST is explained in detail by [9] and [10]. The method is fully implemented in GRASS GIS software while IDW is an integral part of a wide range of analytical software ArcGIS, QGIS, Surfer. For both IDW and RST interpolation methods, there is the option to conduct the 'leave-one-out' version of K-folded cross-validation procedure, thus calculate prediction errors at the locations of the original points (see e.g. [4]). In the special form where K equals the number of data points, the method is embedded in the Geostatistical Analyst tool of ArcGIS as well as in *v.surf.rst* in GRASS GIS which were both employed in this research. Summary statistics of these errors were used to select the parameters for generating the optimal ALS DTM. The most important criterion was to find the interpolation settings which cross-validation error statistics approximate to zero. Mean error, interquartile range, root mean squared error and mean absolute error were considered the most indicative statistics. If the requirement of cross-validation errors approaching zero is met the interpolation is as robust and reliable as possible but it does not imply that it is producing the most accurate DTM surface. For this reason the validation of the interpolated ALS DEMs was supplemented by an assessment of absolute accuracy against ground surveyed measurements on site 4.

### IDW Interpolation and Cross-Validation

IDW based interpolation and its cross-validation was conducted in Geostatistical Analyst of ArcGIS 9.0 (© ESRI). Squared distances ($p=2$) were used in the weighting function and the only varying parameter was the number of neighbouring points entering the prediction at the grid nodes. The initial setting of the number of the closest points was four and it was further incremented by 4 up to 32 points, so that eight DEMs were produced for each site. With the increment of 4 it is more likely that the number of neighbours increases evenly around the predicted location. This condition is applicable as the distribution of the raw ALS points follows approximately a regular pattern.

**RST Interpolation and Cross-Validation**

Finding the optimal interpolation parameters for the RST involved tuning of tension and smoothing, while the other parameters remained constant and were used as follows:

```
v.surf.rst -t segmax=40 npmin=300 dmin=1.0 dmax=5.026359 zmult=1.0.
```

Due to the settings the total number of input points was reduced to 33 – 50% of the original, but the cross-validation and the DEM interpolation still required several hundred times longer processing time than IDW. Tension and smoothing parameters were incremented and cross-validation errors were calculated for each setting. In our case, rescaled tension ranging from 100 to1000 and smoothing values of 0.0, 0.1, 0.2, 0.5 were applied, thus 40 different DEMs were generated. Similar RST tuning procedures were carried out by [5] and [13].

**Masking the ALS data**

The raw ALS data used in the presented research represent point height measurements of the surface hit by a single laser pulse as the last, hence the term “last returns”. In cases, where the laser light penetrated down to the ground, these elevations should be the most accurate samples of the ground surface amongst the DEMs analysed. However, this assumption is not applicable for dense vegetation cover or man-made structures such as buildings, roads, cars, etc. For that reason, the assessment is carried out in two parallel workflows for the unmasked and masked above-ground surface features. The objective was to observe the effect of the presence of the masked objects to the differences in the analyses with regard to the outputs of the statistical and geostatistical analyses. To identify the problematic objects, a DTM based on the last return ALS points was subtracted from an ALS based DSM which was also provided by the Environment Agency. The above-ground surface objects were expected to stand out in the residual surface which was visually analysed and compared against a detailed aerial orthophotograph of 25 cm spatial resolution. The objects identified as the above-ground surface features were manually traced and on-screen vectorized over the orthophotograph of each site. Further the polygons were converted to raster of 5 m cell size and the mask regions was expanded by 2 pixels (10 m) to avoid possible edge effects of the points capturing non-ground surface heights in the subsequent analyses.

**Supplemental Validation Against Ground Surveyed Data**

This validation should be regarded as supplementary to the cross-validation and not equivalent. The heights measured by surveying with total station were subtracted from DEMs derived from the ALS data. Ideally, one would prefer TPS measurements at the locations of ALS points or within a radius corresponding to the horizontal accuracy of the ALS data (ca. 15-30 cm) depending on the height of flight). However, there were only few TPS locations within such a close distance. For that reason the following approach was applied. The DEMs produced during the tuning of IDW and RST.were generated with the spatial resolution of 5 metres at which the accuracy assessment of the airborne datasets was carried out in [11]. Elevations at TPS points were subtracted from the IDW based DEMs as well as from the RST based DEMs. Thus, elevation residuals at TPS locations were produced at 5 m spatial resolution and their statistics are provided in Table 5.3-4. It can be argued that if the elevation

is sampled exactly at the locations of the grid nodes the values would differ. Interpolation of the TPS to the grid nodes would provide a solution but it would also increase number of variables in the assessment.

## RESULTS

### IDW Interpolation

The outcomes of the IDW cross-validation displayed in Fig. 2 suggest that the varying number of neighbours has a relatively small effect on the statistics of each site. The differences of the central values are of the order of millimetres both for the unmasked and for the masked datasets.

In general for all sites, the mean and median error values around zero suggest unbiased predictions. The statistics fall below a centimetre level with the exception of Telford site where the predictions are systematically underestimated by less than 2 cm if mean errors are taken into account. The statistics deviate more from zero when no mask is applied. They are rather similar for Barton, Lincoln and Great Langdale datasets but higher for Telford with no masking. Although mountainous Great Langdale has the highest overall variation of elevation, it was identified in variograms that the short range variation at Telford is higher due to presence of trees (forest) and the high buildings of the power station with tall cooling towers.

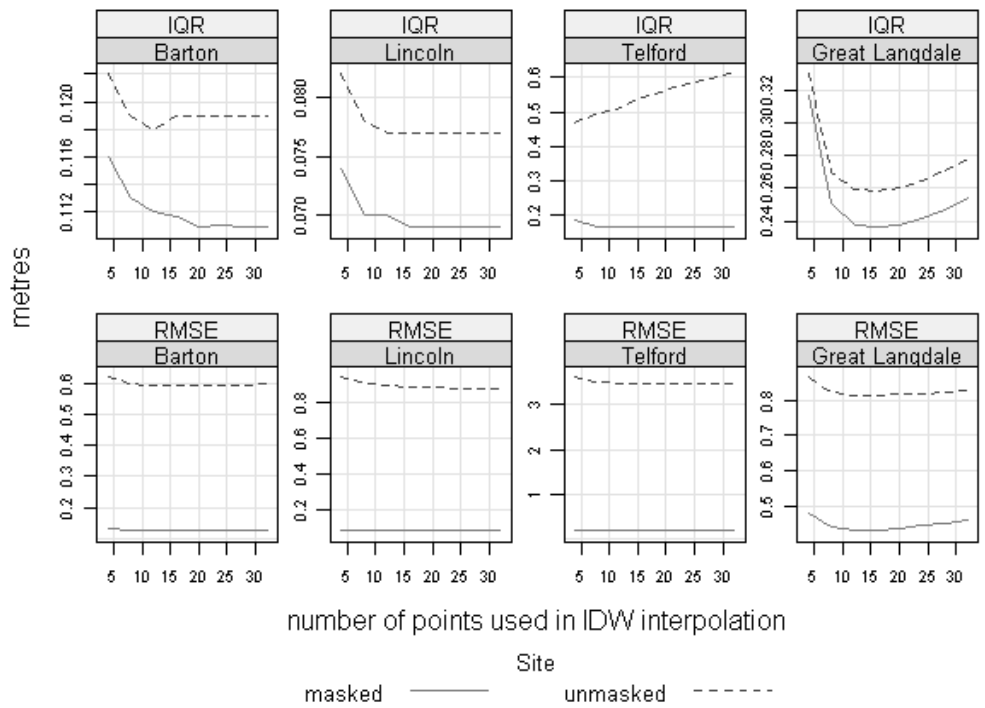



**Fig. 2.** IDW cross-validation statistics of interpolation of ALS data with different number of neighbours when the above-ground surface objects were masked and when they were not. Note the scale of y axis is different for each site to highlight the trends. Interquartile range (IQR) and root mean square (RMSE) of cross-validation errors.

Predicted values react more sensitively when the variation of the input values is larger at the given resolution of prediction. The above-ground surface objects are captured by the ALS on each site which biases the performance of the interpolation in the areas of bare ground surface.

One can conclude that fitting the surface by the IDW interpolation becomes less reliable in the areas of sudden increase of short range variation (higher slope angle) which are the edges of the masked areas (building walls, forest edge) or in their interiors (surface height variation within a forest, car park, settlement). Also the number of neighbours entering the interpolation with which the cross-validation statistics are favourable differs at each site. Thus, masking these objects would better describe how well the interpolation method fits the ground surface. Fig.2 also reveals that the cross-validation errors are markedly less for all sites after the masking. Overall, the RMSEs are lower for smooth terrains of Barton (lowest: 0.124 m) and Lincoln (lowest: 0.085 m) while the RMSE is higher for rougher terrain at Telford (lowest: 0.184 m) and Great Langdale (lowest: 0.431 m). There appears to be no obvious preference for optimal number of points used in the IDW prediction as the differences among the settings are very subtle in the order of millimetres. However, a local minimum of RMSE can be observed between 10 to 24 neighbours for all sites. Based on the statistics after masking the datasets, the interpolation with 12 neighbours appeared to be the optimal solution as the RMSE, MAE and ME are the closest or at least the second closest to zero for all four areas. Since it was found that this parameter does not vary for any of the sites the number of variables was not increased. The purpose of the DEMs was to be a representation of the real surface and for that reason fewer number of input points was preferred if the statistics are very similar as the variation of the points further apart would not influence the predicted values at the grid nodes.

**RST Interpolation and Cross-Validation**

The RST cross-validation statistics are displayed in Fig. 3. The residuals seem to be unbiased for all sites and follow similar tendencies and order as the IDW cross-validation residuals. However, there are extremely large errors (above 1 metre) produced until tensions reach 200. Increasing tension values keep the residuals at a similar level with a slight sinusoidal fluctuation in the order of centimetres depending on the dataset. Local minimum is reached at tension value of 300 and local maximum tension of 600. This tendency can be seen when the above-ground surface features are unmasked and when they are excluded. Introduction of smoothing has a positive effect on reducing the residuals. There is a marked difference especially at lower tensions when no smoothing and smoothing of 0.1 is introduced. A larger smoothing value makes RMSEs at higher tensions to be more similar. Masking the above-ground surface objects also reduces the residuals. The RMSE is several times lower than for the unmasked datasets and the order is again comparable to the case of IDW cross-validation. The RMSE and the other statistics of the RST cross-validation suggest that tension of 300 and smoothing of 0.5 are optimal to be used for all sites. Even though tensions close to 1000 generate comparable RMSEs, visually such DEMs look too noisy.

### Supplemental Validation Against Ground Survey Data

The lowest RMSEs for given IDW and RST interpolation settings at 5 m spatial resolution are similar for the flat meadow (about 0.30 m) and very similar on the inclined slope (IDW: 0.62 m, RST: 0.60 m). However, a greater difference was expected not only due to the use of different prediction methods but also because of the increased uncertainty of the measured values by interpolating into a coarser grid of 5 metres cell size. In fact overall, they are similar and the lowest RMSEs for a given IDW and RST settings are about 0.30 m for the flat meadow at the Rossett Bridge and 0.60 m for the slope at the Middle Fell Farm, respectively. RST RMSE as well as other statistics are the lowest when smoothing is applied and are most similar. RST with no smoothing generates larger errors which tend to decrease as tension increases. When slight smoothing is applied the residuals reach the lowest values and level out for tensions of 300 and higher.

## DISCUSSION AND CONCLUSIONS

In summary, the data used with IDW and RST and smoothing generate very similar cross-validation statistics when only tensions are higher than 100 are considered only. However with more detail, comparison of the lowest RMSE achieved in RST cross-validation and the lowest RMSE of the IDW cross-validation suggests better performance of RST when the datasets were unmasked. The RMSE is smaller by a few millimetres (with the exception of Barton, site 1). Conversely, the IDW performs slightly better with the masked datasets generating RMSEs lower by a few millimetres (with the exception of Great Langdale). This implies that RST generates a slightly more reliable surface compared to IDW when the short-range variation is high due to presence of the above-ground surface objects. The main drawback of RST is the processing time which is several hundred times longer for the RST than with IDW. The RST calculation used 300 neighbours to fit surface per one prediction whereas IDW worked with a ten times smaller number. The number of points entering the RST per prediction can be reduced by setting the *npmin*, *dmin* and *segmax* parameters to lower values, but tests would be required to see how RMSE reacts to such settings. Unarguably, RST can produce smooth surfaces if the settings are adjusted to do so. This kind of DEMs is desirable for many types of terrain related analysis and IDW is not capable to fine control the surface properties. For the purposes of further analysis, it was decided that IDW interpolated using 12 closest neighbours would be used to generate the DEM to be used as the ground truth. The main reasons were that to avoid smoothing with RST and to produce exact interpolation function in case both methods produce very similar RMSE. IDW is an approachable method in proprietary as well as open source GIS systems or statistical software and it is simple to code.

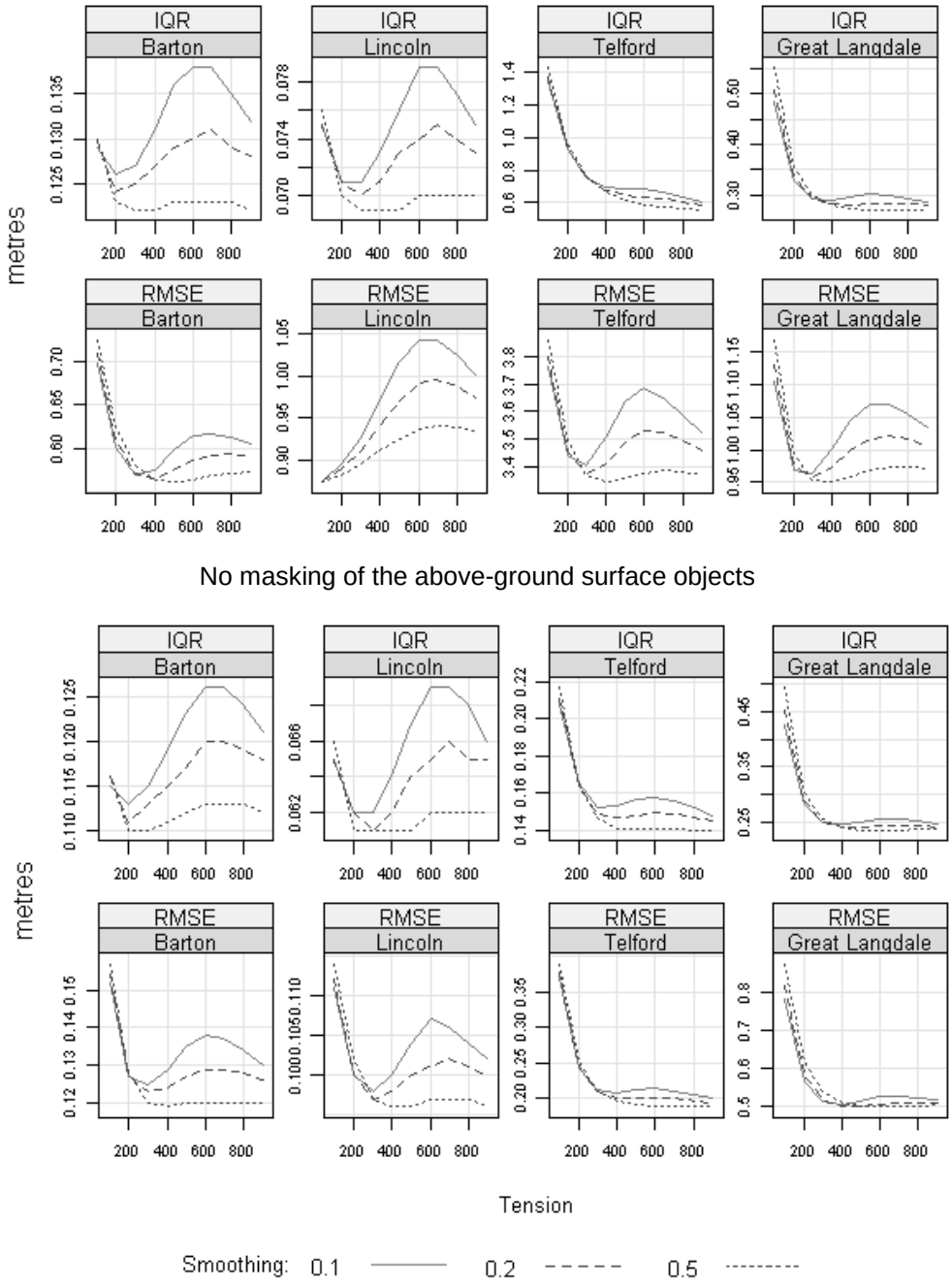



**Fig. 3.** RST cross-validation statistics of interpolation of ALS data with different tension and smoothing values *without and with applied mask* to the above-ground surface objects. Note the scale of the y axis is different for each site to highlight the trends. Values of zero smoothing are not displayed due to larger range of y axis required. Interquartile range (IQR) and root mean square (RMSE) of cross-validation errors.

Thus, the results produced can be tested more easily by other researchers. RST is recommended if both smooth and at the same time still reliable surfaces are required. The properties of a DEM can be finely tuned and controlled with RST. Testing the RST based DTM would be advised for expanding on the current outcomes of this paper. Additionally, there are further conclusions which can be drawn from this analysis with respect to the method and data used:

(i) There is a very small difference in applying IDW or RST on the ALS data in terms of the accuracy level. This just supports the findings of [3], [4], [14] and [15] in that the choice of interpolation method is less influential when a surface is interpolated to coarser resolutions than the resolution of the input data.

(ii) Cross-validation errors increase with increasing short-range variation of elevation for IDW as well as for the RST interpolation method. Most important is the variation within the range encompassing the neighbours which enter the prediction at a location. This relationship is suggested by the decrease of the cross-validation residuals and their statistics after the above-ground surface objects captured by the ALS were excluded. Variograms ranges changed accordingly. The user should be aware that even if the terrain is flat the presence of any buildings, walls or trees will bias the cross-validation statistics and the summary statistics should be calculated also after excluding the above-ground surface objects. It depends on the purpose of the DEM, whether the interpolation should model every feature with the highest accuracy possible or just the ground surface.

(iii) Rescaled tension (flag –t in v.surf.rst command) between 200 – 400 with smoothing provides acceptable cross-validation statistics when data of comparable spacing and accuracy are used as inputs to RST. Higher tensions cause the presence of pits and peaks around the input points even though the cross-validation errors are low. Lower tension than 200 produces overshoots, surface is very flexible and ‘free to bend’ unless smoothing is applied too. Similar findings were published in [5]

(iv) Smoothing greater than zero decreases cross-validation errors of RST. It should be supported by validation against precise data for establishing the parameter value.

The supplemental comparison of ALS point datasets against the ground surveyed measurements with a total station provided insight into the absolute accuracy of the point data and DEM further used as the ground truth. It suggests that both IDW and RST are equivalent in terms of representation of the real surface for the outlined purpose. For IDW being a simpler method it was chosen for further analysis.

## REFERENCES

[1] Baltsavias, E.P. (1999) Airborne laser scanning: basic relations and formulas. ISPRS Journal of Photogrammetry and Remote Sensing, 54(2-3), 199-214.

[2] Pfeifer, N. and Briese, C. (2007) Geometrical aspects of airborne laser scanning and terrestrial laser scanning. In Procceedings of the ISPRS Workshop on Laser Scanning and SilviLaser 2007, Espoo, Finland, ISPRS, 311-319.
http://foto.hut.fi/ls2007/final_papers/Pfeifer_2007_keynote.pdf

[3] Rees, W.G. (2000) The accuracy of Digital Elevation Models interpolated to higher resolutions. International Journal of Remote Sensing, 21(1), 7-20.

[4] Lloyd, C.D. and Atkinson, P. M. (2002) Deriving DSMs from ALS data with kriging. International Journal of Remote Sensing, 23(12), 2519-2524.

[5] Mitasova, H., Mitas, L. and Harmon, R.S. (2005) Simultaneous spline approximation and topographic analysis for ALS elevation data in open source GIS. IEEE Geoscience and Remote Sensing Letters, 2(4), 375-379.

[6] Carrara, A., Bitelli, G. and Carla, R. (1997) Comparison of techniques for generating digital terrain models from contour lines. International Journal of Geographical Information Science, 1997. 11(5), 451-473.

[7] Wise, S. (2011) Cross-validation as a means of investigating DEM interpolation error. Computers & Geosciences, 37(8), 978-991.

[8] Hohle, J. and Hohle, M. (2009) Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS Journal of Photogrammetry and Remote Sensing, 64(4), 398-406.

[9] Mitášová, H. and Mitáš, L. (1993) Interpolation by regularized spline with tension I. Theory and implementation. Mathematical Geology. 25, 641-655.

[10] Mitasova, H. and Hofierka, J. (1993) Interpolation by regularized spline with tension II. Application to terrain modeling and surface geometry analysis. Mathematical Geology, 25, 657-659.

[11] Gallay, M. (2010) Assessing alternative methods of acquiring and processing digital elevation data, in School of Geography, Archaeology and Palaeoecology (PhD thesis). Queen's University Belfast, Belfast, 389 p.

[12] Gallay, M., Lloyd, C. D., McKinley, J. and Barry, L. (2011) Comparing the vertical accuracy of digital elevation models derived using modern ground survey and airborne laser scanning (in Slovak). Kartografické listy, 19, 61-71.

[13] Hofierka, J., Cebecauer, T. and Suri, M. (2007) Optimisation of Interpolation Parameters Using a Cross-validation. In Peckham, R.J., Jordan, G. (eds.) Digital terrain modelling: development and applications in a policy support environment, Springer, New York. 67-82.

[14] Liu, X., Zhang, Z., Peterson, J. and Chandra, S. (2007) The effect of LiDAR data density on DEM Accuracy. In Oxley, L. and Kulasiri, D. (eds.) MODSIM07 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand Inc.: Christchurch, New Zealand, 1363-1369.

[15] Smith, S. L., Holland, D. A. and Longley, P. A. (2005) Quantifying Interpolation Errors in Urban Airborne Laser Scanning Models. Geographical Analysis, 37, 200-224.

# DRAINAGE AREA ESTIMATION IN PRACTIC
# HOW TO TACKLE ARTEFACTS IN REAL WORLD DATA

Abdulghani HASAN[1], Petter PILESJO[2] and Andreas PERSSON[2]

[1]Department of Earth and Ecosystem Sciences, Faculty of science, Lund University, Sölvegatan 12, 22362, Lund, Sweden

*abdulghani.hasan@nateko.lu.se*

[2]GIS Centre, Lund University, Solvegatan 10, 22362, Lund, Sweden

**Abstract**

Large improvements in flow estimation have been made the last decades. However, most, if not all, of the proposed algorithms have been developed and tested on mathematical, or manipulated natural, surfaces. There is an urgent need to develop algorithms dealing with natural artefacts in the landscape, like flat areas and depressions (sinks), caused by man, generalisation (of data type), or by errors in e.g. interpolation. The aim of this study is to present and test practical solutions making it possible to estimate drainage area over natural surfaces with a special focus on sinks and flat areas.

Compared to other studies the main contributions of this research are that we have: adapted surface flow routing algorithms over flat areas to multiple flow algorithms, and developed algorithms letting the user decide which sinks to remove, either by beaching or filling. In both cases the user has the possibility to influence the result, by defining breaching points and deciding thresholds regarding area, volume or depth when filling sinks.

In this study algorithms making it possible for the user to fill sinks depending on area, depth or volume have been developed. This increases the possibility to separate natural sinks from ones coursed by data or analysis errors. Also an interactive semi-automatic way of breaching break lines in the terrain is presented. This is needed instead of filling sinks caused by man-made artefacts, like road banks and train lines. Existing culverts are then replaced by used defined breach lines.

Multiple flow functions to route overland flow over flat surfaces, caused by filled sinks or generalisation, are presented. The flat areas are classified as either ‘flow-out’ or ‘flow-in’. Flow-out occurs when one or more cells on the flat area border have an elevation value lower than the flat area cells. A flat area is classified as ‘flow-in’ when all cells on the border of flat area have elevations higher than the flat area cells. These functions, together with the ones for sink removal, are exemplified and tested on real-world data. Results clearly indicate the benefits of the presented algorithms, making it possible to model overland flow and estimate flow accumulation/drainage area continuously in digital elevation models, without imposing vector hydrology covers (streamlines, lakes etc.).

**Keywords: drainage area, flow accumulation, DEM, multiple flow distribution, flat areas, sinks**

## INTRODUCTION

Catchment topography is critical for models of distributed hydrological processes [1], [2]. Slope controls flow pathways for surface as well as near surface flow, and influences the sub surface flow pattern substantially. The key parameter derived from catchment topography is flow distribution, which tells us how overland flow is distributed over the catchment area.

The fact that flow distribution over a land surface is a crucial parameter in hydrological modelling, in combination with available digital data, has rapidly increased the use of Digital Elevation Models (DEM). Modelling has made it possible to estimate flow on each location over a surface, normally by the use of gridded raster DEMs (see e.g. [3]). Based on the flow distribution estimation on each location represented by a DEM, the drainage pattern over an area, as well as various hydrological parameters such as catchment area and up-stream flow accumulation, can be modelled.

Generally, surface flow and flow accumulation are estimated by the use of two different types of raster algorithms; either approximating to a single, or multiple, flow directions (see e.g. [4], [5] and [3]). If working with raster data, multiple flow algorithms assume transport to more than one adjacent cell, while a single flow algorithm only distributes water to one neighbour cell at a time in the raster. In many cases, single directional flow algorithms produce satisfying results over concave surfaces, while it is often more appropriate to divide the flow into two or more directions if the surface is flat or has a convex form [6], [3]. Combinations of the two types are often to prefer when modelling e.g. water flow over natural surfaces [7].

The single flow algorithm was described by [8]. It assumes that flow follows only the steepest downhill slope. Using a raster DEM, implementation of this method resulted in that hydrological flow at a point only follows one of the eight possible directions corresponding to the eight neighbouring grid cells [9], [10], [11], [8]. However, for the quantitative measurement of the flow distribution, this over-simplified assumption must be considered as illogical and would obviously create significant artefacts in the results, as stated by e.g. [12], [13], [14], [6], [15], [3].

Attempts to solve the problem connected to the single flow algorithms have led to several proposed ‘multiple flow direction’ algorithms [12], [13], [14], [7], [16], [6], [3]. Many of these algorithms estimate the flow distribution values proportionally to the slope gradient, or risen slope gradient, in each direction from the centre cell.

Even if large improvements in flow estimation have been made during the last decades, most, if not all, of the proposed algorithms have been developed and tested on mathematical, or manipulated natural, surfaces. There is an urgent need to develop algorithms (or practical solutions) dealing with natural artefacts in the landscape, like flat areas and depressions (sinks), caused by man, generalisation (of data type), or by errors in e.g. interpolation.

Most flow distribution algorithms available imply sink removal before estimating flow distribution from each cell to its surrounding cells. Normally this is done by ‘filling’ the depression. The result in the DEM is that all cells in the former sink get elevation values equal to the lowest cell value of the border cells. This of course creates a flat area. It can be discussed if, when and how sinks will be removed. Depending on origin (natural or a result of

errors, generalisation and/or man) some sinks might not be removed. There might also be reasons not to fill a sink to remove it, but rather breach the barrier creating the sink. Additionally, it is not rare that flow routing algorithms are not able to estimate flow over flat areas (see e.g. [17], [18], [8], [19], [7]).

Sink removal and surface flow routing have been discussed by many researchers during the last decades [20], [21], [22], [23], [24]. However, these are not adapted to multiple flow algorithms, and/or do not allow the user to interact in the decision if and how to remove a sink. In this research we have combined a well-functioning multiple flow algorithm (see [25]) with modified and developed algorithms for surface flow routing over flat areas. We also discuss and propose interactive solutions for man-made artefact in DEM data.

## COMMON ARTEFACTS IN REAL WORLD DIGITAL TOPOGRAPHIC DATA

Below we briefly present and discuss some common topographic artefacts, influencing flow estimation, in real world data.

### Depressions/Sinks

A sink is an isolated drainage basin, not directly linked to the general drainage pattern where all overland flow ends up the oceans. In a DEM it is characterized by one or more cells surrounded by cells with higher elevation values. Sinks are natural features that frequently occur in the terrain. The water in natural sinks is further transported as subsurface flow until it reaches the ground water table (in oceans, streams. lakes, or as ground water).

Non-natural sinks also often occur in a DEM. These can be a result of data sampling/incomplete data, missing depressions/channels transporting surface flow, and/or inappropriate interpolation algorithms chosen when creating gridded raster DEMs based on point or line data. Also scale/cell size can cause sinks. If the resolution in the DEM is relatively low, small streams and channels, transporting water, can be ‘hidden’. Kenny et al. (2008) highlight the presence of sinks in real world’s data, and also the need for well-functioning algorithms for sink removal.

The logical way of thinking when carrying out hydrological modelling is that sinks should not be removed (by filling or breaching) unless we are sure that they are non-natural artefacts. However, depending on the size of the sink there might be good reasons to assume that also natural sinks are directly linked to the down-hill drainage areas, through e.g. wetlands. If so, also some of the natural sinks should be removed.

Sinks need to be filled in order to estimate flow directions from all cells and connects them into flow distribution and flow accumulation. It should be noted that any unfilled sink will result in stopping flow at that cell. Sometimes this is not desirable.

Based on the discussion about natural and non-natural sinks above we can conclude that there often is a need for sink removal. It would be desirable if the cell removal, additional to other knowledge, could be based on size and form of the sinks. Area, volume as well as depth of a sink might help the user to decide if it should be removed or not (see Fig. 1). Also the

form of the sink can be a help for the user to decide if it should be removed by filling or breaching /see below).

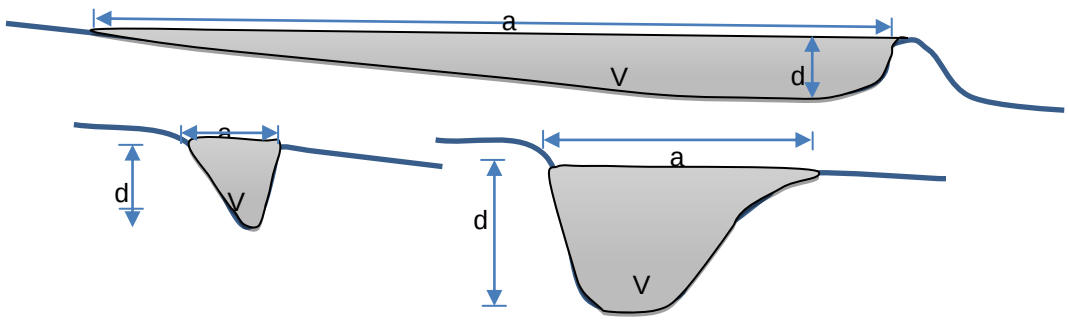



**Fig. 1.** Examples of different forms and sizes of sinks. These parameters are a help when judging if the sink is natural or not, as well as if natural sinks should be removed. V = volume, d = depth, and a equals the area of the sink.

**Man-made Structures (Non-Natural Break Lines)**

Many man-made structures disturb or block estimated flow in a DEM. The main reason for this is that e.g. tunnels and culverts are not captured within the data collection, and thus not included in the terrain model. A DEM covering an area including man-made barriers may thus result in non-natural sinks close to these structures. The barriers are often roads and railway lines, where structures like culverts, siphons and tunnels have been 'hidden' when constructing the DEM. The resulting sinks should be removed since, in these cases, all cells are actually connected and the flow should continue through the man-made artefact. It is thus crucial to identify cells on both 'sides' of the artefact, and let water flow between these points e.g. by breaching the barrier. Such connecting flow structures cannot be detected in the data when creating DEMs. Therefore, it's crucial to find solution how to handle this before estimating flow, flow accumulation and drainage area from a DEM.

**Flat Areas**

A flat area cell can be defined as a cell surrounded by one or more cells with the same elevation, and no cell with lower elevation (see Fig. 2). Flat cells do not exist in the reality, but sometimes in DEMs depending on generalisation of data type (maybe only integer is used, or a limited number of decimals) as well as filled sinks. This implies that flow should be modelled over flat areas, also justified by the fact that water flows over flat areas due to the change in the energy when water is accumulated (depth increases). Water naturally always flows to places with lower elevation, even if the terrain is 'terraced'.

The question is of course how flow should be modelled over these flat areas. The flow can be defined as either 'flow-out' or 'flow-in'. Flow-out implies that the flow from the flat area cells is directed out of the flat area, while flow-in occurs when the flow is converging in the centre of

the flat area. The cell in the centre of the flat area will then have no flow out, and is treated as a sink.

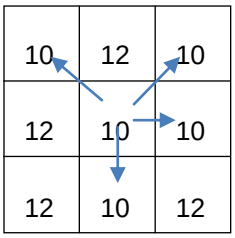



**Fig. 2.** A 3 by 3 cell window exemplifying a flat area cell (the centre cell). The flat cell is surrounded by one or many cells with the same elevation, and no cells with lower elevation. Flow from a flat cell can be none, or outflow to one or more of the neighbouring cells with same elevation.

## AIM

It is a fact that the presence of sinks, man-made structures, as well as flat areas in digital elevation models, does course great difficulties in hydrological modelling. No operational solutions how to tackle these problems in combination with a multiple flow routing algorithm have yet been presented.

The aim of this study is to develop, present and test practical solutions making it possible to estimate flow accumulation/drainage area over natural surfaces, represented as gridded raster DEMs, in an effective and usable way. The solutions will be applied on a real-world data set in order to exemplify their effects and usability.

## METHODOLOGY

Below follows a step-by-step description of how to tackle the above-mentioned artefacts in real world data. All software referred to is available through www.gis.lu.se/petter.

### Cell Classification

In order to separate and deal with the artefacts in the real-world DEMs we decided to label/classify all cells in the DEM. Three different classes were used, namely ‘Flat’, ‘Sink’, or ‘Undisturbed’ cells. Fig. 3 is illustrating the different classes using a 3 by 3 cell window.

In a 3 by 3 window, the centre cell is classified as flat if at least one of the eight surrounding cells has the same elevation as the centre cell, while all other cells have a higher elevation value (see Fig. 3a). A flat area is a group of cells consisting of at least two flat cells neighbouring each other. A cell is classified as sink when all its neighbours have higher elevations (see Fig. 3b), while all remaining cells, having at least one cell with a lower elevation value are classified as undisturbed (see Fig 3c).

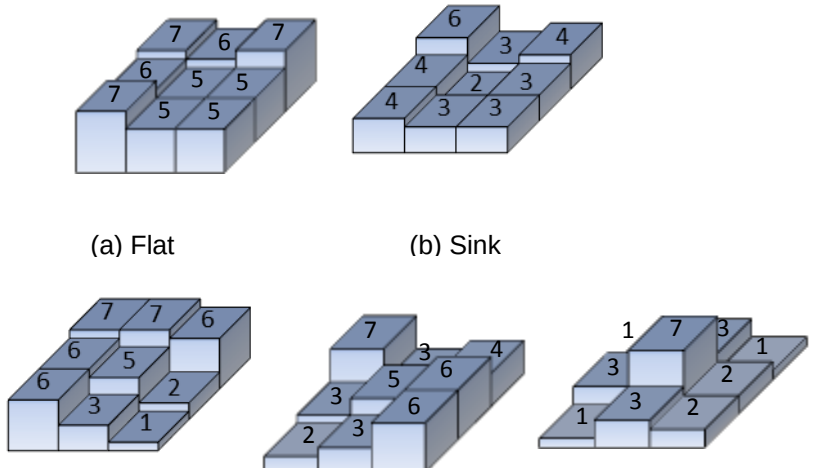



**Fig. 3.** Illustrations of the different topographic forms classified by the use of a 3 by 3 cell window. A flat cell (a) is surrounded by cells with equal and higher elevations; a sink cell (b) is surrounded only by cells with higher elevations, while undisturbed cells (c) are neighboring at least one cell with lower elevation.

**Filling Sinks**

As stated above some sinks in the DEM data might be a result of man-made artefacts, and need to be removed. In order to do this, and also include flexibility regarding area, depth and volume of the sink, a function was created in MATLAB [26]. This function, attached in Appendix 1, makes it possible for the use to select threshold values (area, depth and volume) for sink removal through filling.

The idea behind the function is rather straight forward; Elevation values in flat areas (consisting of neighbouring cells classified as flat, see above) are compared to the elevation value of the lowest adjacent 'non-flat area' cell. The number of cells in the flat area multiplied by the cell size equals the area, the maximum difference in elevation equals the depth of the sink, and the sum of all differences in elevation multiplied by the cell size equals the volume.

The user is presented statistics (area, depth and volume) of the sinks, and then has the possibility to eliminate sinks of decided form and size by filling. The function also makes it possible to identify sinks to be removed by breaching (see Section Breaching break lines below).

**Breaching Break Lines**

If we know that man-mad artefacts, like roads, train lines or other walls are present in the data or suspected by visual interpretation of the DEM or analysis of the form and sizes of sinks (see above) there is a need for breaching of these artefacts. Thus a function (see Appendix 2)

to breach the cells was added, in order to enable users to deal with e.g. man-made flow barriers like roads and railway lines or any other type of break lines. This function breaches the barrier by connecting two user-defined points/cells on the opposite sides of the obstruction. This is done in a semi-automatic way, where the user selects the approximate location of the two end points of the breach line, and the program searches and proposes suitable points/cells (to be confirmed or changed by the user). Rules used to propose these points are that, the starting point (higher elevation) should be the lowest point on the up-slope side of the barrier and higher than the end point, being the highest point on the other side of the barrier. When the points are selected the elevation of all cells in between the points will be changed according to a linear regression line between the points. The function as partly illustrated in Fig. 4, where it can be seen how the elevation values are changed along the breach line to let the flow pass the barrier.

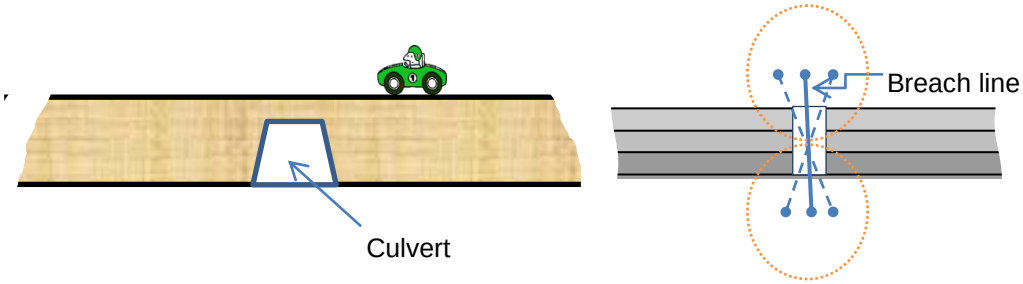



**Fig. 4.** Illustration of how a man-made artefact like a road can be breached. To the left we see a culvert connecting the two sides of the barrier. This is however normally not represented in the DEM. To the right we see the result of the breaching function, where the user have identified the end point of a line (solid line) then the software search for the optimal ends within defined search radius (dashed lines). The breached line will enable water to pass through the barrier.

### Flow Distribution

The amount and direction of flow from each cell in a DEM to its surrounding cells depend on which flow distribution algorithm is used. However, in this study it is not our aim to explain and discuss estimated flow distribution using different flow algorithms, or to compare them. We are instead focusing on how to distribute flow and estimate drainage area/flow accumulation over flat surfaces. However, we want to stress that it is important to use a so called multiple flow algorithm (see e.g. [3]). We have used a modification of the algorithm presented by [3], based on facets with the cell centres as corner points (see below).

In Fig. 2 we illustrated the main problem connected to flat cells; How to decide to which cells water should be distributed if they are all equal in elevation? In Fig. 5 we present a proposed solution of the scenario of Fig. 2. Fig. 5a illustrates the possible flow paths from the centre cell, not violating the rule that no water is allowed to flow up-hill. In Fig. 5b we have extended the window to become 5 by 5 cells, and can logically explain why water is distributed to the upper right cell. The idea behind the algorithm, presented as a MATLAB function in Appendix 3, is that flow is directed in the direction of a lower cell if possible. In the case of Fig. 5 the

upper right cell in the 5 by 5 cell window has an elevation of 8, indicating that the flow is directed in this direction (blue arrows), in not in the other direction (red arrow).

Fig. 5 also reveals the importance of only distributing flow to cells with lower elevation values in 'Undisturbed' terrain. This since cells with an elevation equal to the centre cell, if allocated water, might not represent a 'way out' for the water, but a 'dead end', 'trapping' the water in the sink.

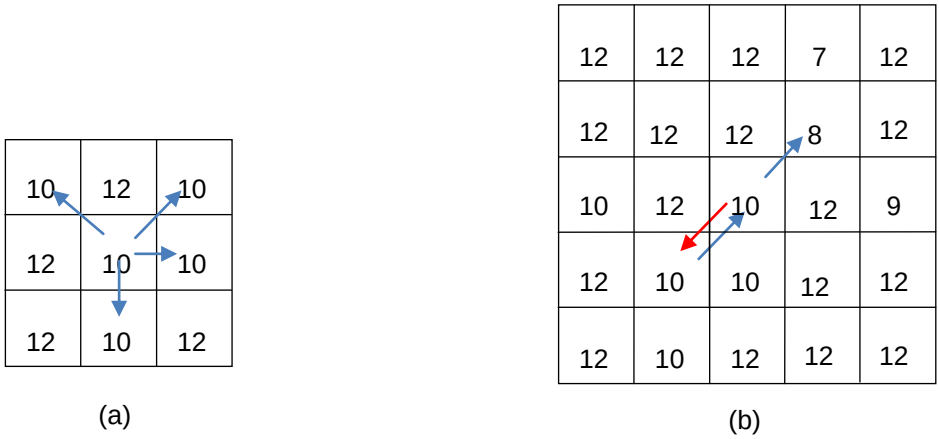



**Fig. 5.** An example of flow distribution from a flat cell. In (a) all possible flow directions are identified, and in (b) the optimal flow path, directed (via another flat cell) to a cell with a lower elevation value is identified.

Two different functions has been developed and added to the main flow distribution program in order to handle flow over flat areas. The functions are attached in Appendix 3 and 4.

FLAT_Flow_out function (see Appendix 3) is a function directing the flow from flat cells where a way out of the flat area can be found. Examples are illustrated in Fig. 5 above and Fig. 6 below. Estimating the flow directions is done according to the following steps:

1. Identify outlets, i.e. cells just outside the border of the flat area with elevation values less than the flat area.
2. Assign neighbouring flat cells flow directions pointing at the outlet cells (by vector addition and splitting, see below).
3. Assign neighbouring flat cells flow directions pointing at the cells assigned in step 2.
4. Continue like above until all flat cells have been assigned flow directions.

So, the procedure starts by giving neighbouring cells to the 'out flow' cell flow directions 'pointing' to that cell, and then stepwise move further and further away from the out flow cell assigning flat cells flow directions to neighbours with a defined direction. Cells with the highest number of neighbours with defined flow directions are processed first, followed by cells with lower numbers of "defined neighbours".

The flow routing from a flat cell is done by vector addition. Since we are using a multiple flow algorithm a non-flat cell normally distributes water to more than one neighbouring cell. The

flow distribution can be described as vectors in different directions (maximum eight, equivalent to the directions to the eight neighbour cells), where each vector has a length corresponding to the amount of water directed in that direction (0 – 100%). By adding all these vectors for all non-flat neighbour cells (maximum eight neighbor cells with maximum eight vectors each) a new vector is created (see [7]). This vector is supposed to represent the flow from the flat cell. Normally the direction of this vector falls in between to neighbouring cells. If so it is split proportionally (see [7]), resulting in multiple flow also over flat areas.

The FLAT_Flow_in function (see Appendix 4) is used when there is no way out from the flat area. An example is presented in Fig. 6b below. All cells just outside the border of the flat area have elevations higher than the flat area cells, and this result in a converging flow in the centre. This centre cell will have no defined flow and will be treated as a sink (see above). The flow directions of the surrounding cells will be estimated (by vector addition) starting from the flat cells that have the maximum number of known distributed flow directions cells (i.e. the border cells of the flat area).

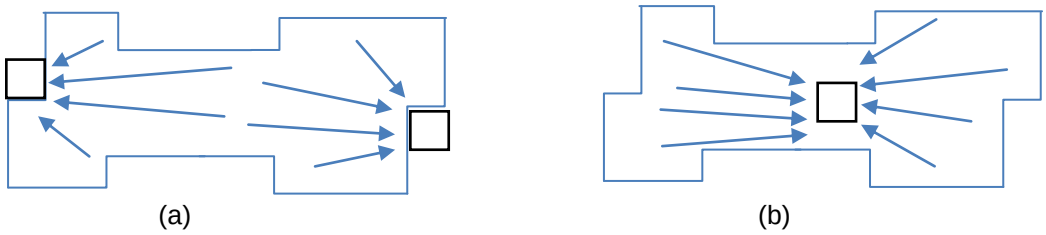



**Fig. 6.** Illustrations how flow distribution is assigned to cells in two different types of flat areas. In example (a) the flat area has two outflow cells, resulting in a continuous flow pattern from the centre of the flat area, while in (b) the flat area has no outflow, resulting in a sink (or actually a cell treated as a sink) in the middle of the flat area.

As mentioned above, estimated flow directions from undisturbed cells will be dependent on the used algorithm. In our study, we have used a development of a facet-based algorithm presented by [3].

Above we have discussed flow distribution over flat areas. If an area is undisturbed, there are a large number of flow algorithms to choose between (see e.g. [8], [13], [11], [12], [7]), all capable of estimating water flow. We have used an improved version of the facet-based presented in ([3]). This method starts by creating triangular facets around centre cells in 3 by 3 cell windows in the regular gridded DEM (see Fig. 7).On these eight facets water flow is modelled by using slope and slope direction of each facet. Incoming (from other cells/facets) as well as produced (precipitation) water is tracked over a facet, all the way to the facet border where it enters another facet.

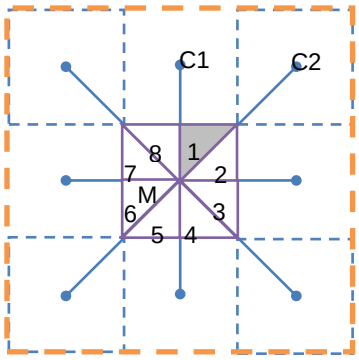



**Fig. 7**. In the used flow distribution algorithm facets are created around each centre cell. The slopes and slope directions of these facets are used to track the theoretical water flow, and estimate to which neighboring cell/cells water will be transported.

It should be noted that, independently of flow algorithm, estimated flow from sink cells is estimated to zero. This means that an unfilled sink interrupts the flow pattern within the DEM, creating an isolated drainage area

### Flow Accumulation/Drainage Area

In order to estimate flow accumulation (often referred to as drainage area) the drainage paths over the surface have to be traced, and the number of up-slope cells (area) transporting water to each and every cell has to be calculated. This is normally (see [24], [23], [9], [20], [21]) done by starting a 'search function' from peak cells, only transporting water to surrounding cells but not receiving water from neighbour cells. The drainage area for these peak cells will be one cell, and the drainage areas for the down-slope neighbour cells that they are transporting water to will be proportionally updated. For example, if water is transported to two neighbour cells on an equal basis both these cells will get an updated drainage area of +0.5. Then these peak cells are flagged as visited, and be neglected in further processing. After this the software will again search for cells with only out flow. Since the original peaks are excluded this will be cells neighbouring the peaks. Flow to down-slope neighbouring cells will be estimated, and then next iteration will start one 'level' down. This continues until all cells have been examined. A detailed description of the estimation of flow accumulation can be found in ([7]). The method is equivalent to counting the flow paths/flow packages in the parts of the eight facets covering the centre cell (see Fig. 7).

## EXPERIMENT

The proposed methods have been tested using real-world data. The newly developed drainage area algorithms focusing on sink removal and flow estimation on flat areas have been applied to a real-world DEM, and the effects have been studied and compared. The spatial pattern of derived flow distribution is visually examined to detect significant artefacts.

**Real-World DEM**

The real-world earth surface elevation data used are from the Stordalen mire and its catchment area. Stordalen is a peatland area in the Arctic region, 10 km west of Abisko (68º 20' N, 19º 03' E) in northern Sweden. The Stordalen mire is a very flat area with many ponds and water bodies, resulting in a large number of flat areas and sinks. Moreover, a road and a railway line are constructed in this area, with many culverts to connect the flow of water between the two sides of the road or the two sides of the railway line. This site was thus judged as suitable for testing of the new methods dealing with flat areas, sinks and culverts.

An airborne LiDAR device has been used to measure the surface elevation. LiDAR is an acronym for “Light Detection And Ranging”, and is a laser-based, remote sensing system used to collect various kinds of environmental data, including topographic data [27]. Over the area defined above the total number of measured elevation data points (the raw data) is 76 940 341. This results in a high resolution data set with an average spatial distribution of approximately 13 points/m2. The LiDAR data in the present study were retrieved with a TOPEYE S/N 425 system mounted on Helicopter SE-HJC. The altitude when sampling was 500 m. The LiDAR data have been post processed and adjusted against 54 known points connected to the national geodetic network. The mean vertical error after post-processing corrections is +0.004 m and the average magnitude of errors is 0.022 m. The RMSE is 0.031 m and the standard deviation is 0.031 m. For more details see [25].

The raw LiDAR data were interpolated into a raster DEM by using inverse distance interpolation [28]. An accuracy assessment has been conducted by [25] on the created DEMs to find the best combination of Search radius and cell size. The selected optimum search radius was set to 1 m and the cell size set to 1 m also (see [25]). The DEM is presented in Fig. 8 below.

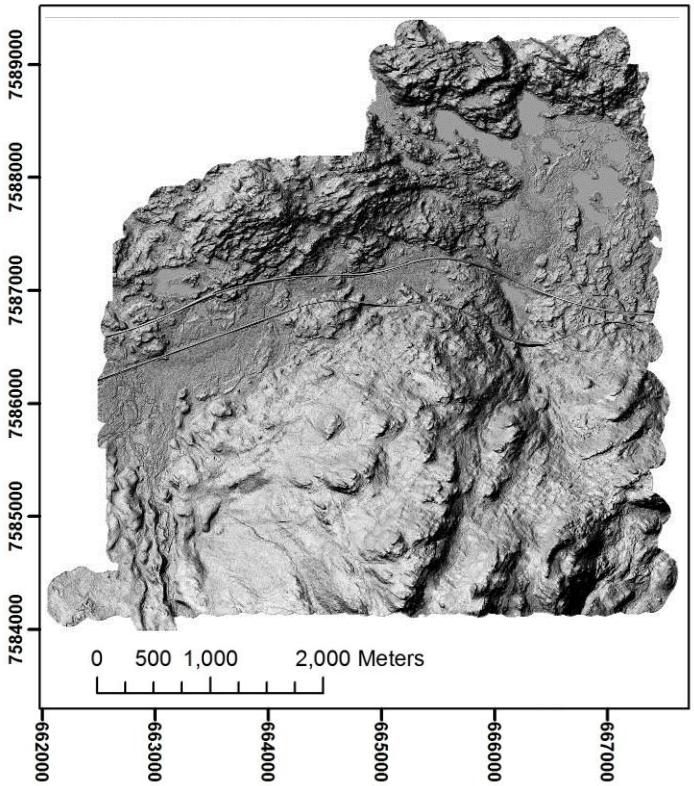



**Fig. 8**. Hill-shade of the used DEM covering the Stordalen area in northern Sweden.

## RESULTS

Fig. 9a shows the sinks in the real-world DEM while Fig. 9b shows the remain sinks after breaching all man-made structures using the culvert breaching approach presented in this study. Large sinks are removed by breaching few cells. Referring to Fig. 9 one can clearly identify the problem with a large number of sinks, which would heavily influence estimated drainage area if not removed.

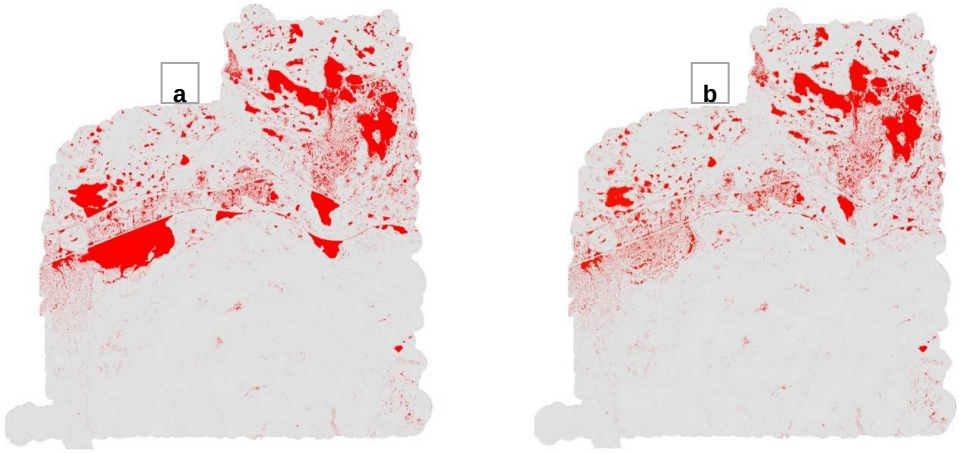



**Fig. 9**. Fig. 9a illustrates the sinks (coloured red) in the Stordalen area. Fig. 9b illustrates the remain sinks after breaching all man-made structures with the culvert approached presented in this study.

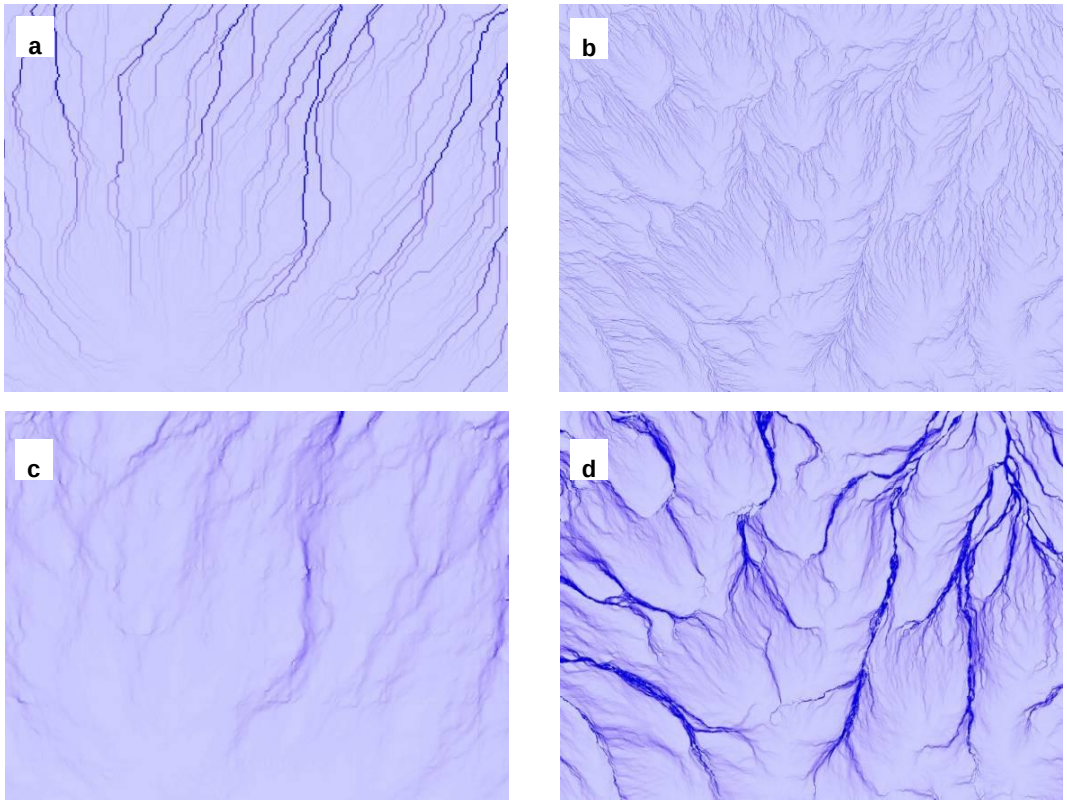



**Fig. 10.** Shows estimated drainage area, using two different algorithms and two different zooming views. In Fig. 10a and Fig. 10b the single flow D8 algorithm (Jenson, et al. (1988)) has been applied, while the result of the newly developed algorithms presented in this paper is illustrated in Fig. 10c and 10d.

Fig. 10 a-d shows show estimated drainage area, using two different algorithms and two different zooming views. In Fig. 10a and Fig. 10b the single flow D8 algorithm (Jenson, et al. (1988)) has been applied, while the result of the newly developed algorithms presented in this paper is illustrated in Fig. 10c and 10d.

In Fig. 11 we see the effect of breaching a barrier. A sink created by a train line is filled, and the resulting drainage area estimation is presented in Fig. 11a. If, instead of filling the sink, the barrier (in this case the train line) is breached a totally different flow and drainage area pattern is created (Fig. 11b). This highlights the importance of flexibility and choice in sink removal, made possible by the introduction of the above-mentioned algorithms.

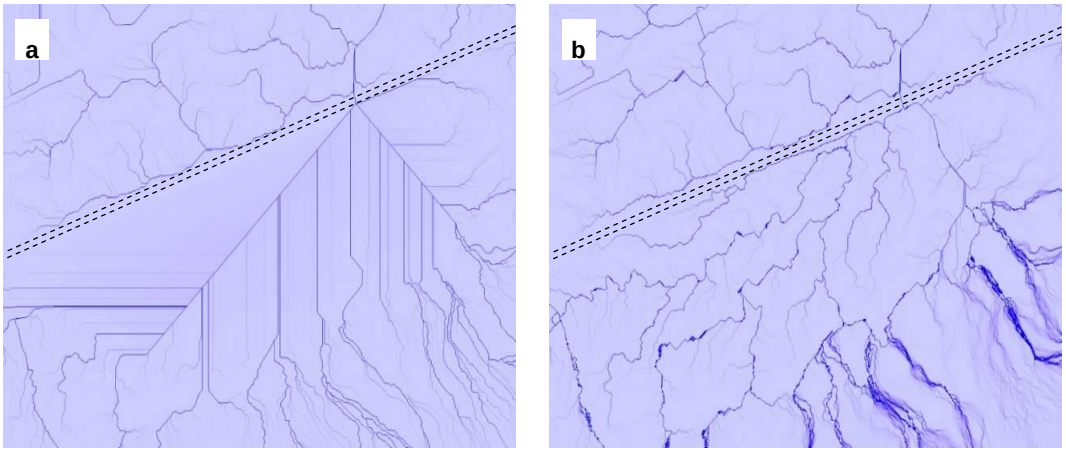



**Fig. 11.** Illustrations of the result of using and not using the function to breach barriers and connect the flow between two sides of a barrier, in this case a railway line. In Fig. 11a (left) the sink has been filled, and the flow accumulation/drainage area has been estimated. This results in a large area of 'regular' flow. In Fig. 11b (right) a few cells in the barrier have been breached, resulting in a more 'natural' estimated flow.

## DISCUSSION AND CONCLUSIONS

In this research paper we have presented new and modified solutions for surface water routing based on a digital elevation model (DEM). These include different ways of sink removal, as well as water routing over flat surfaces. The related problems are not at all new, but have been discussed by several authors during the last decades (see e.g. [24], [23], [4], [21], [22]). Our contributions to the discussions and solutions are mainly that:

- Our solutions are designed to be applied in combination with multiple flow algorithms, proved to be superior to the "normally used" single flow approach (see e.g. [3]).
- We allow the user to interactively choose which sinks that should be filled or breached.
- We allow the user to interactively set thresholds for sinks to be filled, referring to sink area, sink volume, and sink depth.
- We allow the user to interactively define points/cells to breach barriers.

- We are only using the DEM for the flow routing, and no vector layers, like streams and lakes are needed.

It should also be noted that elevation (cell) values are not automatically updated, but the user is given the option to keep the original values, in order not to disturb e.g. estimations of slope and aspect.

It is well known that multiple flow algorithms are more realistic than e.g. the D8 solution ([8]). Thus, multiple flows over flat areas are logical. By summarizing the flow vectors (in different directions) of neighbouring cells, with defined flows, around a flat cell, and then split the summarized vector into two adjacent cells, a more realistic flow is modelled.

The question of which sinks that is results of errors and/or generalization is almost impossible to reply. However, by letting the user decide this, based on expert knowledge and/or visual interpretation of the DEM, and define thresholds for the area, volume or depth of these, we might get a more realistic representation of the reality after a sink removal.

The proposed culvert function can be cost effective. Breaching a few cells representing e.g. a culvert costs less than filling a large area around the sink. Moreover the culvert function can be used by civil engineers to help select the best locations for constructing culverts, tunnels or siphons.

We suggest that the methods presented in this paper are used together with other solutions, e.g. breaching according to [29], and sink flow routing according to [2]. [2] present an “elegant solution”, where flow from a sink is transported as subsurface flow until it reaches the closest cell with a lower elevation than the sink bottom. From there the flow continues on the surface. Of course the results will also improve if methods like the one presented by [22], including vector features if streams, lakes etc., will be used. However, integration of these methods can be time consuming, and they should preferably be adapted to multiple flow solutions of surface flow routing.

When estimating flow accumulation/drainage area and wetness index in order to increase accuracy in e.g. carbon modelling, data are often relatively course and generalized (see e.g. [25]). Detailed stream networks are seldom available and the users have to rely on the DEM only. Under such circumstances the methods presented above are very useful.

## REFERENCES

[1] Quinn, P., Beven, K., Chevallier, P. and Planchon, O. (1991) The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5, 59-79.

[2] Seibert, J., Bishop, K. H. and Nyberg, L. (1997) A test of TOPMODEL's ability to predict spatially distributed groundwater levels. Hydrological Processes, 11, 1131-1144.

[3] Zhou, Q., Pilesjo, P. and Chen, Y. (2011) Estimating surface flow paths on a digital elevation model using a triangular facet network. Water Resour. Res., 47, W07522.

[4] Beven, K. J. and Moore, I. D. (1993) Terrain analysis and distributed modelling in hydrology / edited by K.J. Beven and I.D. Moore. Wiley & Sons, Chichester, England ; New York.

[5] Wilson, J. P. and Gallant, J. C. (2000) Terrain Analysis: Principles and Applications. Wiley,

[6] Seibert, J. and McGlynn, B. L. (2007) A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. Water Resour. Res., 43, W04501.

[7] Pilesjo, P., Zhou, Q. and Harrie, L. (1998) Estimating flow distribution over digital elevation models using a form-based algorithm. Geographic Information Sciences, 4, 44-51.

[8] O'Callaghan, J. F. and Mark, D. M. (1984) The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing, 28, 323-344.

[9] Band, L. E. (1986) Topographic partition of watersheds with digital elevation models. Water Resources Research, 22, 15-24.

[10] ESRI. (1991) Cell-based Modelling with GRID, Environmental System Research Institute.

[11] Mark, D. M. (1984) Automated detection of drainage networks from digital elevation models. Cartographica, 21, 168-178.

[12] Freeman, T. G. (1991) Calculating catchment area with divergent flow based on a regular grid. Computers and Geosciences, 17, 413-422.

[13] Holmgren, P. (1994) Multiple flow direction algorithms for runoff modelling in grid based elevation models: an empirical evaluation. Hydrological Processes, 8, 327-334.

[14] Pilesjo, P. and Zhous, Q. (1996), A multiple flow direction algorithm and its use for hydrological modelling, Geoinformatics´96, West Palm Beach, FL, April 26-28, 366-376.

[15] Wolock, D. M. and McCabe, G. J. (1995) Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. Water Resources Research, 31, 1315-1324.

[16] Quinn, P. F., Beven, K. J. and Lamb, R. (1995) The ln(a/tan-beta) Index - how to calculate it and how to use it within the TOPMODEL framework. Hydrological Processes, 9, 161-182.

[17] Costa-Cabral, M. C. and Burges, S. J. (1994) Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. Water Resources Research, 30, 1681-1692.

[18] Fairfield, J. and Leymarie, P. (1991) DRAINAGE NETWORKS FROM GRID DIGITAL ELEVATION MODELS. Water Resources Research, 27, 709-717.

[19] Tarboton, D. G. (1997) A new method for the determination of flow directions and upslope areas in grid DEMs. Water Resources Research, 33, 309-319.

[20] Beven, K. J. and Kirkby, M. J. (1979) A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Journal, 24, 43-69.

[21] Jenson, S. K. and Domingue, J. O. (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing, 54, 1593-1600.

[22] Kenny, F., Matthews, B. and Todd, K. (2008) Routing overland flow through sinks and flats in interpolated raster terrain surfaces. Computers & Geosciences, 34, 1417-1430.

[23] Lindsay, J. B. and Creed, I. F. (2005) Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEMs. Photogrammetric Engineering and Remote Sensing, 71, 1029-1036.

[24] Oimoen, M. J. (2000), An effective filter for removal of production artifacts in US Geological Survey 7.5-minute digital elevation models, Proceedings of the 14th International Conference on Applied Geologic Remote Sensing, Las Vegas, NV, USA. Veridian ERIM International, Ann Arbor, MI, 311-319.

[25] Hasan, A., Pilesjo, P. and Persson, P. (2011) The use of LiDAR as a data source for digital elevation models - DEM resolution versus accuracy and estimated slope, drainage area and wetness in northern peatlands. Hydrology and Earth System Sciences, discussion paper.

[26] MathWorks. (2008) MATLAB R2008B. R2008B.

[27] Fowler, R. (2001) Topographic Lidar Maune, Digital Elevation Model Technologies And Applications, American Society for Photogrammetry and Remote Sensing, Maryland.

[28] Shepard, D. (1968) A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM national conference.

[29] Garbrecht, J. and Martz, L. W. (1999) TOPAZ: an automated digital landscape analysis tool for topographic evaluation, drainage identification, watershed segmentation and subcatchment parameterization; TOPAZ overview. United States Department of Agricultural, Agricultural Research Service (USDA-ARS) Publication No. GRL 99-1, Washington, DC, USA.

**Appendixes**

Appendixes 1 – 4 are available in the on-line version of the proceedings at http://gis.vsb.cz/gis2012/.

# DEFORMATION MONITORING USING INSAR AND ARTIFICIAL REFLECTORS

Ivana HLAVACOVA[1], Lena HALOUNOVA[1] and Kvetoslava SVOBODOVA[1]

[1]Department of Mapping and Cartography, Faculty of Civil Engineering, Czech Technical University, Thákurova 7, 166 21, Prague, Czech republic

*ivana.hlavacova@fsv.cvut.cz, lena.halounova@fsv.cvut.cz, kvetoslava.svobodova@fsv.cvut.cz*

**Abstract**

The SAR interferometry (InSAR) method can be used to create digital elevation model (DEM), or to monitor terrain deformation or some features of the atmosphere. The method uses several satellite radar images, to form interferograms from pairs of them. The interferogram always contains several components: the flat Earth, topography, deformations, atmospheric influence and noise. The aim is typically to extract one of the components and therefore, it is possible to estimate and subtract (or neglect) the others.

A big problem of this method is so-called decorrelation, i.e. if the noise is significantly higher than other components. Mostly, it happens in vegetated areas - forests, agricultural fields, meadows, mostly during late spring, summer and early autumn. Such vegetated areas can be monitored only during winter (if no snow is laying), but with lower accuracy. Much better results can be gained in urban areas, which are stable in time (in the order of centimetres). The decorrelation can be also found if there is a long time or long distance between the acquisitions of the scenes in a pair.

We deal with monitoring of deformations, which occur due to undermining or subsidence (on waste dumps), or due to landslides, mostly in the Nothern-Bohemian coal basin. We cooperate with the Coal Services, a.s., company, and within this project there were 11 artificial corner reflectors installed in the mines surroundings (near the Most city). The places to install the reflectors were selected by the company as the most important places to monitor. This is the only way to monitor areas containing a very small amount of artificial objects.

Now, the corner reflectors are monitored by German satellite TerraSAR-X, which enables acquisition of images with resolution of approx. 3x3 m. The corner reflectors can be found in the scene as very light points (with respect to their surroundings), about 2x2 pixels large. It is impossible to find one of the reflectors, because it is placed in the area of radar foreshortening/layover. The period of this satellite is 11 days, the acquisition is performed (due to financial reasons and with respect to the deformation rate) once in 33 days.

The basic interferometric processing of 4 available scenes was performed, both for the whole area and the corner reflectors alone.

**Keywords: synthetic aperture radar, interferometry, subsidence, Northern Bohemian coal basin, artificial corner reflectors**

## INTRODUCTION, AREA AND PROJECT DESCRIPTION

The Northern-Bohemian coal basin is full of brown coal. Both in history and today, it is being exploited. In history, the most frequent way was deep-mining, with the risk of subsidence of undermined areas. Today, there are many huge open-pit mines and waste dumps, with a similar risk of subsidence - however, reclamation was mostly performed in a way in order to minimize the impact of the subsidence - most of the area is forested, there can also be found water surfaces, horse racing grounds, and only rarely, when there is no other possibility, a road, railway etc. is built on a waste dump. New family houses are being built on old waste dumps.

In cooperation with Czech Coal, a.s. company, there were 11 artificial corner reflectors installed in the area. Their list can be found in table 1. Last spring, the reflectors started to be monitored by the TERRASAR-X satellite. List of available data due to date of paper submission can be found in table 2. The TERRASAR-X [1] satellite is a German satellite, with a SAR on-board. Its period of overpass is only 11 days, and the wavelength is only 3.1 cm (X-band). Due to the fact that fast deformations are not expected in the area of interest (in built-up areas which can be monitored), we decided to order one scene in 33 days. The incidence angle for these scenes is about $30^{o}$. The scenes were acquired in the StripMap mode with resolution of about 3x3 metres.

**Table 1.** Corner reflectors, placed in the area of interest. Point 9, at the dam near the Vysoká Pec village, is invisible in the scene due to the radar shadow (scenes are from descending pass). The intensity is only approximate and is computed as a mean for all scenes.

| Point | Name | Y | X | range | az. | intensity |
|---|---|---|---|---|---|---|
| 1 | ČSA | 797910.59 | 986137.97 | 1568 | 2700 | $6 \times 10^7$ |
| 2 | ČSA lom | 798828.08 | 984351.49 | 2474 | 2030 | $5 \times 10^6$ |
| 3 | Centrum | 793904.55 | 982765.86 | 49 | 509 | $5 \times 10^6$ |
| 4 | Horní Jiřetín | 796600.26 | 980927.46 | 1842 | 103 | $2 \times 10^6$ |
| 5 | Černice | 798208.87 | 981182.20 | 2666 | 479 | $2 \times 10^6$ |
| 6 | monitoring | 799860.73 | 982529.26 | 3266 | 1362 | $7 \times 10^7$ |
| 7 | Jezeří | 800040.38 | 982355.30 | 3310 | 1312 | $3 \times 10^6$ |
| 8 | ČS - ČSA | 801088.68 | 983470.47 | 3754 | 1991 | $1 \times 10^6$ |
| 9 | přehrada u Vys. Pece | 802050.76 | 982886.82 | invisible (shadow) | | |
| 10 | Vysoká Pec ob. úřad | 802976.67 | 985356.52 | 4387 | 3159 | $4 \times 10^6$ |
| 11 | mont. místo Vrskmaň | 801578.94 | 988818.49 | 2989 | 4518 | $5 \times 10^6$ |

**Table 2.** List of available scenes. All scenes are acquired in descending mode, at approx. 5:25 UTC.

| Date | Temp. baseline [days] | Perp. baseline [m] |
|---|---|---|
| *2011-06-17 (master)* | *0* | *0* |
| 2011-07-20 | 33 | -145 |
| 2011-08-22 | 66 | 3.5 |
| 2011-09-24 | 99 | -143 |

An artificial reflector is expected to be found in the scene as a very high-intensity point, in comparison to its surroundings. It is advised that the reflectors are not installed close to object which can also have high reflectivity (buildings, bridges over roads or rivers).

All points were found in all the scenes, except for point 9, which is, unfortunately, situated in radar shadow. In most cases, the reflectors are visible as a light square (mostly 2x2 points), in some cases it is a cross, in some cases there are also other light pixels.

Using intensity, the position of the reflectors were estimated with a sub-pixel accuracy, and also the intensity was interpolated for such an estimated position. The estimation is performed in the GAMMA software [2]. In table 1, the intensity is given only approximately in order to allow comparison among individual reflectors. Comparing the estimated reflector positions, standard deviation for each point is 0.1 pixel at maximum (for both range and azimuth coordinates).

We tried both types of processing, to estimate Earth surface deformations either only for the reflectors, or for the whole scene crop, and present preliminary results (we are still waiting for more scenes to be acquired) of both types.

## PROCESSING OF THE WHOLE SCENE CROP

The four scenes acquired were co-registered to a selected master (the first acquisition), and then cropped. Six interferograms were formed of them. The interferograms are mostly decorrelated due to vegetation, but artificial objects, such as buildings, roads, bridges, railway etc. can be well recognized.

For topography subtraction, we use the SRTM (Shuttle Radar Topography Mission) DEM (digital elevation model), acquired in 2000 using X-band antenna [3]. Because SRTM was primarily designed for C-band, the X-band data contain "holes", places which were not mapped due to shorter swath. In our case, this is the case of the bottom-left corner; fortunately, there is no corner reflector placed in this area.

The SRTM X-band DEM is classified as precise (however, it is 11 years old, so not very up-to-date in open-pit mines). However, the resolution of the scenes is so good that in some cases, we can see a phase gradient over a wall of a building - and this cannot be covered in the DEM with resolution of 1'' (about 20-30 m for our latitude).

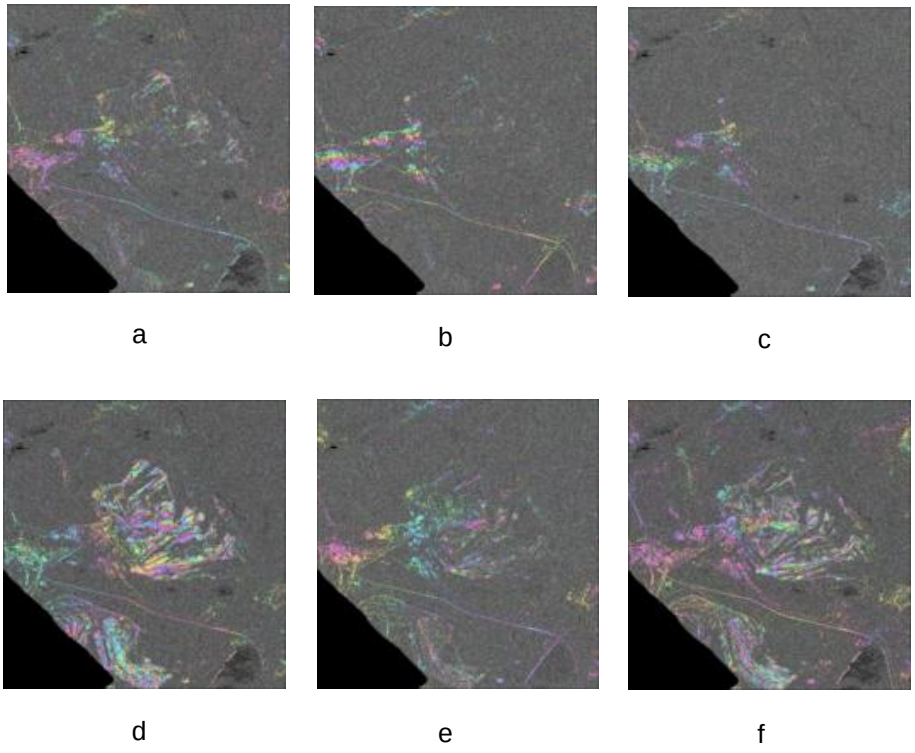



**Fig. 1.** Interferogram phase maps, the colorsale is displayed in Fig. 2 and corresponds to $2\pi$. Created from scenes acquired (a) 11-06-17 and 11-07-20, (b) 11-06-17 and 11-08-22, (c) 11-06-17 and 11-09-24, (d) 11-07-20 and 11-08-22, (e) 11-07-20 and 11-09-24, (f) 11-08-22 and 11-09-24. Interferograms (b) and (e) are of very short perpendicular baseline, i.e. no DEM error is expected. However, interferogram (b) contains visible fringes (in the azimuth direction), which can be attributed only to atmoshperic delay or orbit errors. Within the open mines (coherent only in (d), (e), (f)), fringes are expected due to the out-of-date DEM. The line between the open mines is the Ervěnice corridor - railway, road and pipeline built on a waste dump in 1980s (the coherent line is the railway). The line approx. perpendicular to the Ervěnice corridor in its lower part is the Kyjická dam. The built-up area on the left part of the scene is Komořany.

**Fig. 2.** Colorscale used for the interferogram and deformation maps.

Among the six interferograms, we can find two with a very short perpendicular baseline (shorter than 5 m) - both with temporal baseline of 66 days. In the interferograms with the perpendicular baseline of about 145 m, only 1 m error in the DEM causes a phase error of about 11 degrees.

However, as an additional error source, besides the DEM error, we must consider the atmospheric delay variation. This is expected to be of a long wavelength, i.e. should cause phase differences on long distances; and therefore short distance phase differences are to be considered deformations.

In all interferograms in Fig. 1, one can see phase trends. In interferogram (b), the fringes are probably caused by atmospheric delay variation, however, if this was the case, we should see it also in some other interferograms. The same applies for an orbit error. A small DEM error is expected in this case due to the very short perpendicular baseline.

Interferograms (a), (c), (d) and (e) present a slow trend (let us neglect the colour variations of the open mines, where there is certainly up-to-date DEM). This trend can be probably attributed to atmospheric delay variations. In interferograms (c) and (e), one can find a phase (colour) change at the beginning of the corridor itself (at the double bend). This was not investigated yet, and we plan to prepare a mask (to mask out the open pit mines) and estimate the atmospheric delay as the trend.

We also tried the IPTA (Interferometric Point Target Analysis) processing. IPTA is implemented in the GAMMA software [2] and is similar to other (better-documented) methods of Permanent Scatterers [4] or Persistent Scatterers [5]. However, the recommended number of scenes for this method is at least 10. For a small number of scenes, it is more difficult to select the points to be processed , and also estimate a suitable threshold for a posteriori standard deviation. In addition, during the adjustment, phase ambiguities are estimated, and there are many more possibilities to estimate them with just a small change of the resulting standard deviation (due to the small number of inputs).

A priori point selection is performed on the basis of MSR (mean-to-sigma-ratio), described in [5]. Both mean and standard deviation (sigma) are computed from the intensity information of each pixel (from all scenes available). However, such a measure becomes statistically more precise for a higher number of scenes. Software GAMMA [1], however, during the evaluation also takes into account the intensity of the neighbouring points, and therefore only pixels that are much brighter than their surroundings are really selected. On the other hand, allows to perform the selection on a different basis, based only on the surroundings of each pixel (for each scene separately), but according to our older tests, better results were achieved using the statistical method even for small number of scenes.

After processing, i.e. the adjustment with the ambiguity estimation, a posteriori point selection is performed, on the basis of the a posteriori standard deviation. The threshold recommended in the GAMMA software does not work here, it gives too many points with irrelevant results, and even if lowering the threshold helps, the resulting deformation map (see fig. 3) still contains many points which do not seem to be appropriate. This problem is expected to be improved with a higher number of scenes to be acquired in the future.

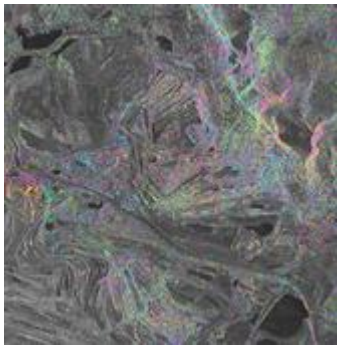


Fig. 3. Deformation map from the IPTA method. The colorscale is displayed in Fig. 2 and corresponds to and corresponds to 50 cm/year.

## PROCESSING OF THE REFLECTOR INFORMATION ALONE

In this case, after finding the reflectors, it is necessary to estimate their precise (sub-pixel) positions and interpolate the phase to this precise position. Software GAMMA implements two ways of doing that:

- the *ptarg* script, suitable to be used on original SLC (single-look complex) data (containing intensity and phase, but not yet paired and processed into interferograms)
- the *xpt_slc* script, which is part of the IPTA package. To be able to use it, one first need to make the point selection and then work only with the selected points.

### The *ptarg* script

In our project, we are interested in the interferogram phase of the reflectors; however, we do not need the phase information of the original SLC data. Interferogram phase is the difference between phases of two scenes; however, with some adjustments: flat-Earth phase subtraction and DEM subtraction. The phase should not be filtered spatially, and we are not sure if spectral filtering is recommended (however; due to short baselines and zero Doppler centroid, filtering is expected to be almost useless).

Interferogram creation, flat-Earth and DEM subtraction is performed within the GAMMA software and we tried to apply the *ptarg* script on the result. However, the resulting interferogram contains only the phase component, with the intensity equal to 1 for all points. In this case, it is impossible to estimate the sub-pixel position of the reflector, because for the estimation, the intensity information is used. We therefore multiplied all interferograms with the intensity of the master scene (acquired on 2011-06-17) in order to estimate the precise position of the reflectors and its phase. The estimated intensity (approximate) can be found in table 1.

Using *ptarg*, we estimated the reflector positions and phase for all six interferograms (displayed in Fig. 1) and for all reflectors (except for reflector 9, which cannot be found in the image due to the radar shade). In all cases, the *ptarg* script was successful, the sub-pixel position was estimated and the phase interpolated. The *ptarg* script computes two values of

interpolated phase: *peak phase* and *peak phase with phase gradients*. We did not know which one is better to use in this case, and tried to use both.

The estimated positions for each reflector differ by 0.02 pixel at maximum, which seems to be appropriate. To check the phase accuracy, we compute triangular sums: we select three scenes, make three interferograms containing only these scenes, and the phase sum of all of them is expected be 0 (except for noise, orbit errors and position estimate and interpolation errors).

However, this is not always the case. The triangular sums can be found in table 3.

**Table 3.** Triangular sums for the *ptarg* outputs (in radians), reduced for integer multiples of $2\pi$. The triangle is constructed from the three interferograms created from the three scenes disclosed in the first column. For all points, numbered 1 to 11 (except for 9, which is invisible), the sum of estimated peak phases and sub of estimated "peak phases with phase gradients" is computed. It can be seen that only for the last triangle the "peak phase with phase gradients" is approximately constant for all points, however not close to zero. Unfortunately, this is not the case of previous triangles, but this problem can be solved only with more scenes.

| Triangle | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 110617-110720-110822 | peak phase | -1.33 | -2.60 | -1.59 | 2.00 | 1.35 | -2.84 | -2.92 | 0.81 | 3.00 | 1.60 |
| | with ph. gr. | -1.23 | -1.98 | -2.51 | -2.21 | 1.61 | -2.46 | -2.88 | 0.26 | -3.00 | 0.15 |
| 110617-110720-110924 | peak phase | -0.12 | 0.55 | 3.06 | 1.96 | 0.79 | -1.75 | -1.21 | 2.97 | 0.19 | -3.07 |
| | with ph. gr. | -0.10 | 0.61 | 3.08 | 1.14 | 0.82 | -1.77 | -1.37 | 2.66 | -0.86 | 0.35 |
| 110617-110822-110924 | peak phase | 0.30 | 1.87 | -0.72 | -2.88 | -1.87 | -0.13 | 0.78 | 2.12 | 1.99 | 1.35 |
| | with ph. gr. | -0.03 | 1.38 | -1.85 | 2.09 | -2.04 | -0.72 | 0.34 | 0.98 | 0.85 | -1.03 |
| 110720-110822-110924 | peak phase | -0.91 | -1.28 | 0.92 | -2.83 | -1.31 | -1.22 | -0.93 | -0.03 | -1.48 | -0.25 |
| | with ph. gr. | -1.16 | -1.21 | -1.15 | -1.27 | -1.25 | -1.41 | -1.17 | -1.43 | -1.30 | -1.23 |

The triangular sums, disclosed in table 3, seem to be random. Only in the case of the last row, these sums are approximately constant. This could mean that it is more suitable to use "peak phase with phase gradients" for the following processing; however, it is not evident why this does not apply to the previous three triangles. We hope that this will be more clear in future, with more available scenes.

Phases with such high (and random) triangular sums are useless to be processed.

### The *xpt_slc* script

The *xpt_slc* script is a part of the IPTA package, and – similarly to *ptarg* – estimates the position of the reflectors (point scatterers) with a sub-pixel accuracy. And although it also works for original SLC data, the resulting format is compatible with the following processing - interferogram creation, flat-Earth and DEM phase subtraction (both are estimated on the basis of the precise position). That means, that the results are expected to be more accurate than that of *ptarg*, and also that triangular sums should be close to zero or at least close to constant for all reflectors.

However, in the case of *xpt_slc*, not all reflectors were successfully identified as point scatterers in all scenes (reflectors 6 and 7 are missing in only one scene, but the others even in more of them). Reflector 11 was even not selected for processing in the a priori point selection, described above.

However it may seem that both scripts have the same basis, it is no way to override the inability of the *xpt_slc* script to detect some reflectors as point scatterers. This is the case of the reflectors with the lowest intensity (see table 1). Also, however the differences among the estimated positions of each reflector were usually less than 0.2 pixels for ptarg, for xpt_slc are the differences almost 1 pixel, which is probably the reason why some scenes are missing for a reflector.

Research in the near future will be devoted to comparison of the phase and intensity values between both scripts, in order to use the values estimated by *ptarg* for the subsequent processing.

## CONCLUSIONS

There are still too few scenes in order to present any reasonable results. However, there are some notes about what the further processing may give:

- Individual reflectors can be well found in the images, except for reflector 9, which is situated in radar shadow. Reflectors 1, 2, 3, 10 are in all available scenes detected as point scatterers, reflectors 6 and 7 are missing only in one scene. The other reflectors may be inaccurately oriented or surrounded by other well-reflecting objects. Reflector 11 was even not selected as a point for IPTA processing.
- The *ptarg* script does not give satisfactory results. The source of inaccuracy may be the fact that the interferogram phase is then manually multiplied by image intensity (of the same image for all interferograms), but the phase reductions (flat-Earth and DEM phase) are computed not for the reflector position, but for the cell centre.
- More hopeful is the *xpt_slc* script. However, it does not detect all points as point scatterers, and also the variations in the estimated position is higher.
- Conventional interferometric processing gives reasonable results, there are scene pairs with very short perpendicular baselines; however, the interferograms made from such a pair are not as unicolour as expected, probably due to DEM errors (the resolution of the scene is much higher than the resolution of the available DEM).

- IPTA processing is not very promising yet; however, we still hope that this will improve with a higher amount of scenes available.

## REFERENCES

[1] Infoterra (2011) TerraSAR-X Radar Satellite Imagery. http://www.infoterra.de/terrasar-x-satellite, [cit 2011-10-31].

[2] GAMMA Remote Sensing and Consulting AG (2011) Gamma Remote Sensing. http://www.gamma-rs.ch, [cit 2011-10-31].

[3] Deutsches Zentrum für Luft und Raumfahrt (2011) SRTM. http://centaurus.caf.dlr.de:8080/eoweb-ng/licenseAgreements/DLR_SRTM_Readme.pdf, [cit 2011-10-31].

[4] Kampes, B (2006) Radar Interferometry: Persistent Scatterer Technique. Springer, Dordrecht, the Netherlands.

[5] Ferreti, A., Prati, C. and Rocca, F. (2001) Permanent Scatterers in SAR Interferometry. IEEE Transaction on Geoscience and Remote Sensing, 39, 8-20.

# ACCURACY ASSESSMENT OF MDL LASER ACE®300 FOR POINT POSITIONING AND 3D MODELING IN FORESTRY APPLICATION

Jasmee JAAFAR[1], Juazer Rizal ABDUL HAMID[2],
Roslina IDRIS[3] and Azman Ali M. HABIB[4]

[1,2]Department of Surveying Sc. & Geomatics, Faculty of Architecture, Planning & Surveying
Universiti Teknologi MARA, 40450, Shah Alam, Malaysia

*[1]jasmee@salam.uitm.edu.my, [2]juazer@salam.uitm.edu.my*

[3,4]Postgradute, Department of Surveying Sc. & Geomatics
Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, 40450
Shah Alam, Malaysia

*[3]jaslina01@yahoo.com,[4]the_engineer@hotmail.com*

**Abstract**

The process of automation towards gathering of tree inventory and the creation of a 3D model for forestry applications are highly demanded. In the present study, a MDL LaserAce®300 was the main equipment used in acquiring three dimensional (3D) surface data of points and tree feature in a forested environment. The results from the field observations conducted, was assessed based on the accuracy attained of the horizontal positions and the constructed 3D model of the forest floor. The planimetric assessment will be based on the standard traversing and tacheometric surveying data sets respectively. In the tree inventory exercise, tree positions are gathered from a known coordinated point. The points are part of a traverse survey station. In this study, the accuracy of the traverse points is evaluated based on the coordinates of eight survey stations forming a closed traverse. The survey stations coordinates are established using MDL LaserAce®300 and a total station (Topcon GTS 235N) respectively. The coordinate for the first survey station was assumed to be N1000, E1000 and the respective coordinates forming the traverse were then computed. The linear misclosure (accuracy) for the traverses are 1: 107 and 1:55017 for MDL LaserAce®300 and total station respectively. On the other hand, the RMSE (Root Mean Square Error) for the computed coordinates established using MDL LaserAce®300 and total station is found to be 1.131 m. With this accuracy, traversing using MDL LaserAce®300 is obviously adequate for forestry application. For the vertical accuracy assessment, an area of approximately 200 x 200 meter within a forest reserve compartment with varying topography of 95 to 102 m above MSL (Mean Sea Level) was chosen. Two sets of DEMs (Digital Elevation Models) were constructed using 235 and 250 points gathered using MDL LaserAce®300 and total station respectively. The targeted 235 points using MDL LaserAce®300 was directed at the middle of a tree trunk at approximately 1.5 m above ground level, which represent the position of observation during the tree inventory exercise. For the observed 250 points using total station, the points are represented by the position of the prism pole, randomly placed to capture the surrounding topography of the study area. In this study, it was found that the accuracy of the constructed MDL LaserAce®300 DEM is ±1.95 m. With these accuracies (planimetric and constructed

DEM) obtained, the usage of the MDL LaserAce®300 to support various forest mapping activities such as gathering of tree inventory data, and the construction of the 3D model is justifiable.

**Keywords: measuring laser, tree inventory, 3D model**

## INTRODUCTION

The rainforest is the world's greatest natural resource and needs to be preserved and managed sustainably. Malaysia, being a tropical rainforest the conservation programme is understood as being part of the activities under the jurisdiction of the Forestry Department, and one of their challenging tasks is to obtain a tree inventory data on selected trees, such as the trees' geo-location with respect to a local reference coordinate system, tree height, diameter breast height (dbh), species type and other associated tree attributes for the development of the Geographic Information System (GIS) database. The process is tedious, time consuming, costly and labour intensive. The identification of individual trees plus tree count is one of the important agendas for inputs to the GIS database.

No doubt, with the advent of GPS (Global Positioning System), positioning of Earth's surface features could be carried out with ease; however, in a forested environment, the effect of tree canopy cover will usually result in blockage of the satellite signal and certainly, would give some restrictions for a good observational record, which ultimately could degrade the accuracy of the position obtained. Furthermore, the height component obtained from handheld GPS is of low accuracy to support the creation of a reliable 3D model. The normal surveying method of gathering tree inventory variables, which involved sophisticated surveying equipments such as the total station and other surveying accessories need trained personnel to execute the task. With the advent of handheld laser equipments such as the MDL LaserAce®300, acquisition of tree position using this equipment seems appealing. MDL LaserAce®300 is equipped with a fluxgate compass that could determine the direction of a sight with respect to the magnetic North. Horizontal and slope distance between the observer and target could also be measured with ease using the MDL LaserAce®300. With this information, spatial information such as, positions of respective points and difference in height between observed and targeted points can be computed using the direction and height information obtained from the MDL LaserAce®300.

Tree counting using laser scanning technology and remotely sensed data sets is an active research agenda. However, inside a forested area, especially in a tropical forest setting consisting of trees of different tree species, the task is quite daunting and cumbersome. Although utilizing techniques that involved the integration of airborne remotely sensed datasets, such as the utilization of hyperspectral and LiDAR (Light Detection and Ranging) data sets could be giving reliable information of the tree attributes that could aid towards creating accurate GIS database of the tree inventory, for instance [1], they may not be so successful in the determination of tree species, *dbh* and height. This is due to the diversity of tree species in a heterogeneous environment which complicate the reflectance response and ultimately do not facilitate the tree identification process via image-based processing analysis techniques [2]. Hence, the success of identifying tree species using remotely sensed data

sets is debatable. On the other hand, the construction of a Digital Surface Model (DSM), which depicts the surface and above-surface features for various forest applications, is highly demanded. By having surface models, the tree height component may be estimated accordingly. Laser scanning technology or commonly known as LiDAR seems to be the-state-of-the-art approached in the creation of the digital model, for instance [3]. The construction of Digital Elevation Model (DEM) of the forest floor in a rainforest environment using LiDAR is not an easy or straightforward task since one needs adequate and reliable samples of point-cloud LiDAR data that actually hit the ground or the forest floor. This was made difficult by the obstruction of the canopy cover against the travelling laser pulse signals, the acquisition of these data sets by no means, is cheap and easy. Thus, the development of tree inventory via field-based survey measurement of the tree parameters remains to be ever important for foresters in their daily routine activities.

The acquisition of tree inventory parameters by field-based survey methods is the common practice in Malaysia. The practice, however has some drawbacks, which include the need of more manpower and trained personnel. However, lack of control points, difficulties in determining line of sight coupled with rugged terrain conditions are among the obstacles often faced by foresters and land surveyors. Among other, drawbacks include, acquisition of the information using conventional surveying techniques needs trained personnel, incurred high cost and time consuming. Field's automation in a tree inventory process seems to be the solution in need to speed up the process of tree inventory survey and data management. The advent of surveying technology such as, hand held laser, namely the MDL LaserAce®300 [4] that could measure distances, the difference in height and bearing automatically with reasonable accuracy. Integrating this equipment with computing technology towards automation in acquiring tree inventory seems feasible. In this study, the capability of the MDL LaserAce®300 towards acquiring tree inventory and Reduced Level (height) in an automated system is investigated. The accuracy of the horizontal and vertical position of points gathered will then be critically assessed with respect to forest application.

## METHODOLOGY

In this study, an in-house software known as Automated Spatial Survey Information System (ASSIST) that automates the process of acquiring tree inventory is developed using Visual Basic and C++ high level programming language [5]. This system utilized MDL LaserAce®300 as main hardware. Figure 1.0 shows main menu for ASSIST. Among the main modules incorporated within ASSIST software are shown in Figure. 1.

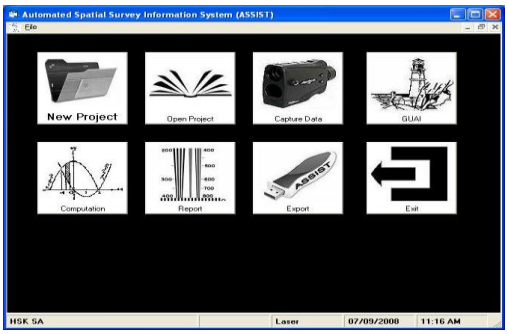



**Fig 1.** ASSIST Main Menu

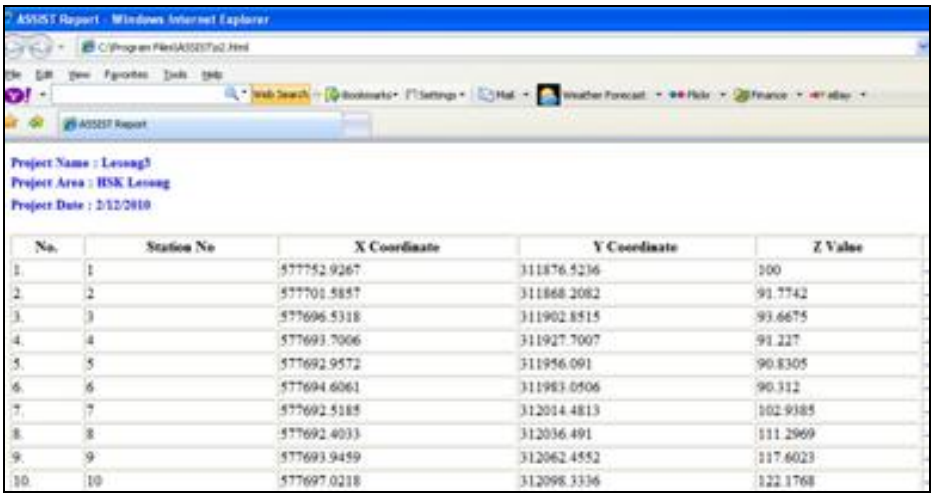
Table 1 gives brief explanation on the associated modules within the software. There are eight (8) main modules with dedicated functions to ensure the process of acquiring tree information will be carried out with ease.

Figure 2, shows an example of the acquired survey stations coordinates from a tree inventory exercise using ASSIST. The information that will be reported includes, the traverse station number and the three-dimensional rectangular coordinates of a survey station (X,Y and Z). It should be pointed out that, in a tree inventory exercise using ASSIST, the first coordinate (X,Y,Z) for the starting Survey Station (Station 1) of the traverse loop should be known.

By knowing the coordinate of Survey Station 1, the coordinate of the next Survey Station (i.e. Station 2) of the traverse can be computed based on the recorded bearing and distance measured by the MDL LaserAce®300. After establishing Survey Station 2, the tree inventory parameter is collected at Survey Station 1.

**Table 1.** System (modules) within ASSIST

| No | Module | Purpose/Capability |
|---|---|---|
| 1 | New Project | Creation of a new project, initialisation for the project is carried out in this module. |
| 2 | Open Project | Open an existing project from the database. |
| 3 | Capture data | Integration with the MDL LaserAce® 300. Tree inventory database will be created in this module. |
| 4 | GUAI | Graphic User Advanced Interface (GUAI) for online viewing, help user to navigate in the rainforest and viewing of the tree inventory coverage. |
| 5 | Computation | Standard surveying adjustments on surveyed points and tree positions. |
| 6 | Report | Generation of survey report and tree inventory. |
| 7 | Export | Export to GIS ready format (shape file). |
| 8 | EXIT | Exit the software. |



Project Name : Lesong3
Project Area : HSK Lesong
Project Date : 2/12/2010

| No. | Station No | X Coordinate | Y Coordinate | Z Value |
|---|---|---|---|---|
| 1. | 1 | 577752.9267 | 311876.5236 | 100 |
| 2. | 2 | 577701.5857 | 311868.2082 | 91.7742 |
| 3. | 3 | 577696.5318 | 311902.8515 | 93.6675 |
| 4. | 4 | 577693.7006 | 311927.7007 | 91.227 |
| 5. | 5 | 577692.9572 | 311956.091 | 90.8305 |
| 6. | 6 | 577694.6061 | 311983.0506 | 90.312 |
| 7. | 7 | 577692.5185 | 312014.4813 | 102.9385 |
| 8. | 8 | 577692.4033 | 312036.491 | 111.2969 |
| 9. | 9 | 577693.9459 | 312062.4552 | 117.6023 |
| 10. | 10 | 577697.0218 | 312098.3336 | 122.1768 |

**Fig 2**. Information for the traverse

This is done by measuring distances (horizontal, slope and difference in height) to the surrounding trees at a known height (Figure 3) using the MDL LaserAce®300 equipment. The diameter-at-breast height (*dbh*) is computed based on relaskop reading couple with the distance measured by MDL LaserAce®300 to the tree. The tree species are identified manually and chosen from the database, which is embedded within the ASSIST.

The procedure is then repeated on the next survey station until the survey process is completed. The process of traversing that forms a closed loop (start and end at the same survey station) is recommended.

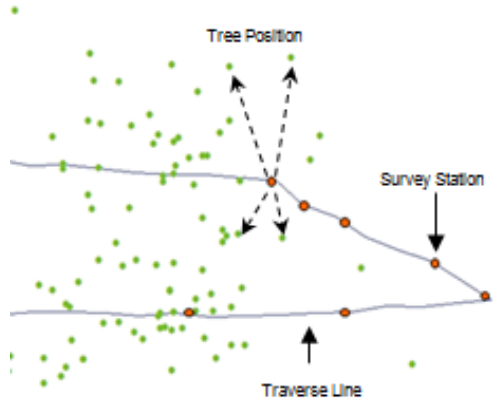
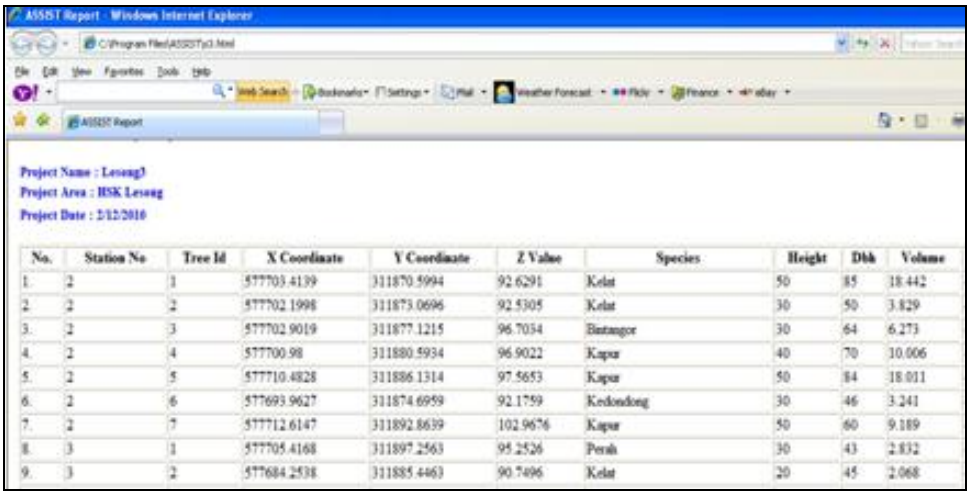



**Fig 3.** Acquisition of tree inventory from an established survey station

Figure 4, depicts the attribute data for the tree inventory acquired using ASSIST and they include, the Survey Station (from which the coordinates of the tree is acquired), the computed coordinates for the tree position (X-Coordinate, Y-Coordinate), the Reduced Level of the tree position (Z-Value), the tree species, height, *dbh* and Volume (computed).

Referring to Figure 4, based on the computed reduced level (Z Value) of points collected during the tree inventory exercise, Digital Elevation Model for the area could be easily constructed. Fig 5 shows an example of a DEM constructed using the reduced level (Z Value) with the aid of ArcGIS software [6].



Project Name : Lesung3
Project Area : HSK Lesung
Project Date : 2/12/2010

| No. | Station No | Tree Id | X Coordinate | Y Coordinate | Z Value | Species | Height | Dbh | Volume |
|---|---|---|---|---|---|---|---|---|---|
| 1. | 2 | 1 | 577703.4139 | 311870.5994 | 92.6291 | Kelat | 50 | 85 | 18.442 |
| 2. | 2 | 2 | 577702.1998 | 311873.0696 | 92.5305 | Kelat | 30 | 50 | 3.829 |
| 3. | 2 | 3 | 577702.9019 | 311877.1215 | 96.7034 | Bintangor | 30 | 64 | 6.273 |
| 4. | 2 | 4 | 577700.98 | 311880.5934 | 96.9022 | Kapur | 40 | 70 | 10.006 |
| 5. | 2 | 5 | 577710.4828 | 311886.1314 | 97.5653 | Kapur | 50 | 84 | 18.011 |
| 6. | 2 | 6 | 577693.9627 | 311874.6959 | 92.1759 | Kedondong | 30 | 46 | 3.241 |
| 7. | 2 | 7 | 577712.6147 | 311892.8639 | 102.9676 | Kapur | 50 | 60 | 9.189 |
| 8. | 3 | 1 | 577705.4168 | 311897.2563 | 95.2526 | Perah | 30 | 43 | 2.832 |
| 9. | 3 | 2 | 577684.2538 | 311885.4463 | 90.7496 | Kelat | 20 | 45 | 2.068 |

**Fig 4**. Tree Inventory

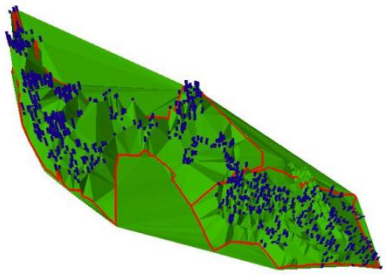


**Fig 5**. Constructed DEM using Reduced Level information

**Instrument and Data**

Table 2, shows the equipments used in this study, among the equipments used are Total Station (Topcon GTS 235N) and its accessories, which include tripod and reflector. The Total Station is used to acquire survey points with high accuracy and the points collected will act as the control during the assessment process. Apart from Total Station, the MDL LaserAce®300 (with pole and prism) was also used. MDL LaserAce®300 is the main equipment used in the tree inventory exercise where the points gathered using this equipment will be critically assessed compared to the point's collected using Total Station.

As mentioned, the coordinates of related points can be acquired via measurement made by MDL LaserAce®300. The accuracy of the observed coordinates (X,Y,Z) in supporting forest application need to be checked and determined. In this study, the datasets derived are evaluated based on two components, namely;

i) Horizontal Accuracy, and

ii) Vertical Accuracy

The datasets used to evaluate the horizontal accuracy is based on a closed traverse consisting of eight (8) survey station. The first set (control) survey is carried out using accurate surveying equipment consisting of a Total Station.

**Table 2**.Major Equipments used in the study

| No | Instrument | General Description | Comments |
|---|---|---|---|
| 1 |  Total Station (Topcon) | Angular Accuracy 1"<br>Distance Accuracy .001m | Acquisition of survey points for traverse and tacheometric points to act as control for assessment |
| 2 |  EDM Prism & Tripod | Mount the Total Station (Tripod) and for EDM reflector (Prism) | Part of Total Station accessories |
| 3 |  LaserAce® 300 | Angular Accuracy 0.1degree<br>Distance Accuracy 0.1m | Acquisition of survey points for traverse and tacheometric points to be accessed its accuracy |
| 4 |  Prism With Pole | Adjustable heights, act as LaserAce®300 reflector. | Reflector for MDL LaserAce®300 |

The second dataset for the closed traverse was measured using the MDL LaserAce®300. Table 3, shows the dataset gathered using Total Station and MDL LaserAce®300.

**Table 3.** Bearing and distance between total survey station and MDL LaserAce®300

| Total Station | | | | MDL Laser | | | |
|---|---|---|---|---|---|---|---|
| Station | | Bearing | Distance | Station | | Bearing | Distance |
| From | To | (Degree, Min & Sec) | | From | To | (Degree, Min & Sec) | |
| 1 | 2 | 176 59 26 | 40.489 | 1 | 2 | 178 00 00 | 38.2 |
| 2 | 3 | 258 19 23 | 58.729 | 2 | 3 | 260 06 00 | 57.0 |
| 3 | 4 | 311 36 01 | 80.215 | 3 | 4 | 312 30 00 | 79.8 |
| 4 | 5 | 307 29 00 | 100.168 | 4 | 5 | 306 00 00 | 99.5 |
| 5 | 6 | 354 55 05 | 120.256 | 5 | 6 | 356 00 00 | 119.1 |
| 6 | 7 | 87 00 13 | 150.990 | 6 | 7 | 85 30 00 | 152.0 |
| 7 | 8 | 81 05 03 | 38.768 | 7 | 8 | 83 06 00 | 37.8 |
| 8 | 1 | 175 11 58 | 196.252 | 8 | 1 | 174 12 00 | 197.5 |

For the vertical accuracy assessment, an undulating area of approximately 200 x 200 meters with elevation varying from 95 to 102 meters above Mean Sea Level (MSL) inside a forest reserve compartment is chosen. The selected area represents a natural condition where tree inventory tasking was undertaken. Altogether 250 points were taken as samples and their coordinates (X,Y,Z) were measured and recorded using a Total Station instrument utilizing the

Tacheometric Survey technique (act as control) [7]. Whilst another 235 points were gathered using the MDL LaserAce®30, by observing the tree trunk (middle) at approximately 1.5 meters above ground level. Fig 6 shows the distribution of those points across the study area with the point samples representing the Tacheometric and MDL LaserAce@300 respectively. Based on these 3-D point samples, the digital elevation models (DEMs) were generated for both acquired survey techniques using TIN Interpolation method. The digital elevation models are constructed based on the 3D coordinates gathered.

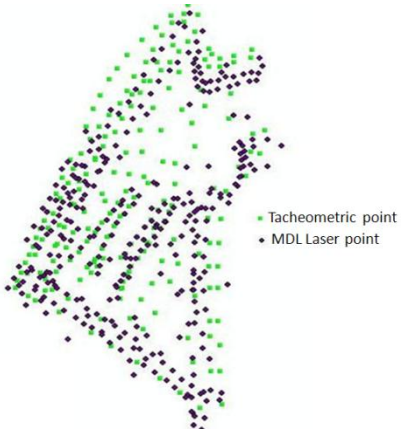



**Fig 6.** Points distribution acquired using total station and and MDL LaserAce@300

## RESULTS AND DISCUSSION

The horizontal and vertical accuracy for the datasets gathered is analyzed quantitatively under the following headings;

### Horizontal Accuracy

The accuracy (Linear Misclosed) [7] for both traverse, namely using the Total Station and MDL LaserAce@300 are first determined. Referring to Table 4, the accuracy (Linear Misclosed) for the traverse are 1: 55017 and 1:107 representing standard traversing using Total Station and MDL LaserAce®300 respectively. Since the accuracy for the standard traverse using Total Station exceed 1:8000 for the first class survey [7], the standard traverse will then act as a control in the assessment process.

**Table 4.** Linear Misclosed

Total Station

| Station | | Bearing | Distance | Latit | Depart |
|---|---|---|---|---|---|
| From | To | (Degree, Min & Sec) | | | |
| 1 | 2 | 176 59 26 | 40.489 | -40.433 | 2.126 |
| 2 | 3 | 258 19 23 | 58.729 | -11.886 | -57.514 |
| 3 | 4 | 311 36 01 | 80.215 | 53.257 | -59.984 |
| 4 | 5 | 307 29 00 | 100.168 | 60.955 | -79.486 |
| 5 | 6 | 354 55 05 | 120.256 | 119.783 | -10.652 |
| 6 | 7 | 87 00 13 | 150.990 | 7.893 | 150.784 |
| 7 | 8 | 81 05 03 | 38.768 | 6.008 | 38.300 |
| 8 | 1 | 175 11 58 | 196.252 | -195.564 | 16.424 |
| | | Total | 785.867 | 0.014 | -0.004 |
| | | Linear Misclosed | 1: 55017 | | |

MDL Laser

| Station | | Bearing | Distance | Latit | Depart |
|---|---|---|---|---|---|
| From | To | (Degree, Min & Sec) | | | |
| 1 | 2 | 178 00 00 | 38.2 | -38.177 | 1.333 |
| 2 | 3 | 260 06 00 | 57.0 | -9.800 | -56.151 |
| 3 | 4 | 312 30 00 | 79.8 | 53.912 | -58.835 |
| 4 | 5 | 306 00 00 | 99.5 | 58.485 | -80.497 |
| 5 | 6 | 356 00 00 | 119.1 | 118.810 | -8.308 |
| 6 | 7 | 85 30 00 | 152.0 | 11.926 | 151.531 |
| 7 | 8 | 83 06 00 | 37.8 | 4.541 | 37.526 |
| 8 | 1 | 174 12 00 | 197.5 | -196.489 | 19.959 |
| | | Total | 785.9 | 3.208 | 6.558 |
| | | Linear Misclosed | 1: 107 | | |

Referring to Table 4, the *Latit* and *Depart* [7,8] component are computed based on Equation 1 and 2.

Figure 7 shows an example for the position of *Latit* and *Depart* on a surveyed line.

$$Latit = Distance * Cos\ (Bearing) \quad (1)$$

$$Depart = Distance * Sin\ (Bearing) \quad (2)$$

Referring to Fig 7, the *Latit* and *Depart* component for the surveyed line is in the North (+ve) and East (+ve) direction. Bearing is referred to the horizontal angle measured from reference North to the line of sight. In Fig 7, the line of sight is from Station 1 to Station 2.

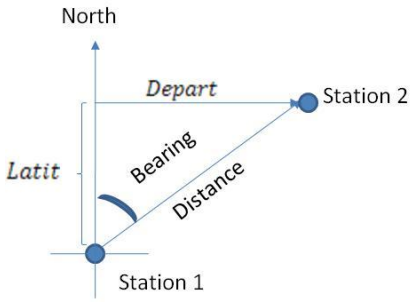



**Fig 7.** *Latit* and *Depart* for a survey line

Adjustment on the *Latit* and *Depart* obtained (Table 5) was then carried out using the Bowditch Method [7,8] The coordinates for the traverse are then computed by assuming the first point coordinate as N1000, E1000 (Northing, Easting).The different in coordinates between established using Total Station and MDL LaserAce®300 in Northing and Easting are computed (Table 5). The Root Mean Square Error for the different in coordinate is then computed based on Equation 3.

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(\nabla)^2\right]} \quad (3)$$

$RMSE$ = Root Mean Square Error; $n$ is the total number of points and *i* is the point number;

$\nabla$ is the different in Northing (dN) or Easting (dE)

**Table 5.** Difference in established coordinates

| Point | Coord Total Station | | Coord MDL Laser | | dN | dE |
|---|---|---|---|---|---|---|
| | Northing | Easting | Northing | Easting | | |
| 1 | 1000 | 1000 | 1000 | 1000 | | |
| 2 | 959.566 | 1002.126 | 961.667 | 1001.012 | -2.101 | 1.114 |
| 3 | 947.679 | 944.613 | 951.633 | 944.383 | -3.954 | 0.230 |
| 4 | 1000.934 | 884.629 | 1005.218 | 884.877 | -4.284 | -0.248 |
| 5 | 1061.888 | 805.143 | 1063.295 | 803.545 | -1.407 | 1.598 |
| 6 | 1181.669 | 794.491 | 1181.609 | 794.237 | 0.060 | 0.254 |
| 7 | 1189.559 | 945.275 | 1192.912 | 944.491 | -3.353 | 0.784 |
| 8 | 1195.567 | 983.575 | 1197.298 | 981.7 | -1.731 | 1.875 |
| | | | | ***RMSE*** | 2.793 | 1.076 |
| | | | | $RMSE_{dNdE}$ | 1.131 | |

Referring to Table 5, it is shown that using MDL LaserAce®300 for forest mapping seems feasible, the accuracy of the laser in establishing horizontal points is ±1.131m ($RMSE_{dNdE}$). For tree inventory purposes, where the position of the tree should be established with less than ±5 meter [9] of accuracy, MDL LaserAce®300 could play an important role towards forest inventory mapping.

### Vertical Accuracy

The vertical accuracy is based on the different in height between the constructed DEMs (Grid size 2m) as shown on the Residual Surface (Figure 8) [10]. In this study, the error seems to occur in step slopes areas, and this might due to the uncertainty in the creation of the model due to the position and number of points gathered to represent the surface. However, further studies on the effect will be an ongoing research agenda.

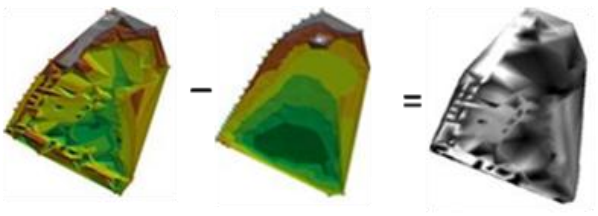



**Fig 8.** Residual surface

Table 6, shows the computed mean error for the residual surface. The mean is computed based on the total summation of the error with respect to grid count (occurrence) for the surveyed area.

**Table 6.** Error Count of the Residual surface

| Number of Grid Count | Error (Different) in meter | Total Error (Meter) |
|---|---|---|
| 38 | 0.5 | 19 |
| 68 | 1.7 | 115.6 |
| 43 | 2.2 | 94.6 |
| 18 | 3.3 | 59.4 |
| 8 | 4.3 | 34.4 |
| 5 | 5.5 | 27.5 |
| *Total* = 180 | | (350.5/180)<br>*Mean Error* = 1.95 m |

It was known that, Digital Elevation Model (DEM) depicting the forest environment is of high value to a variety of forestry applications, such as designing of logging routes, biomass study and the simulation of flood plain studies. In this study, it was shown that, using MDL LaserAce®300, heights of points can be gathered with reasonable accuracy of less than ±2 meter [9]. In forestry applications, DEM constructed with less than ±5 meter accuracy is sufficient to support a diversity of forestry applications [9].

## CONCLUSION

Tree inventory is of high importance in managing forest activities. This study has depicted the capability of MDL LaserAce®300 for tree inventory exercise. It should be pointed out that, during a tree inventory exercise, the position of the tree needs to be determined, therefore, the accuracy of the instrument in used need to be assessed. It was found that the accuracy of the horizontal points (survey station) and the constructed DEM is less than ±2 meters by utilizing the MDL LaserAce®300. The findings from this study are adequate to support data collection for tree inventory and other forestry applications.

## ACKNOWLEDGEMENT

Computing facilities were made available by the Department of Surveying Science & Geomatics, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA Malaysia, and the Research Management Institute, Universiti Teknologi MARA, Malaysia for the research grant.

## REFERENCES

[1] Abdul Hamid, J. R., Mather, P. M. and Hill, R. A., (2003) Mapping of conifer forest plantations using airborne hyperspectral and Lidar data, (Ed.). Proceedings of the 23rd Symposium of the European Association of Remote Sensing Laboratories, Ghent, Belgium, 2-5 June, Millpress.

[2] Nafisah, K., Juazer, R. A. H., Abdul, M. M. N. and Zulkiflee, A. L.. (2011) Synergistic Utilization of Multispectral and LiDAR Data for Tree Crown Delineation. Proceeding of the 10th International Symposium & Exhibition on Geoinformation 2011 ( ISG & ISPRS 2011). Shah Alam, Selangor, Malaysia.

[3] Roslina, I., Zulkiflee A. L. and Jasmee, J.,(2011) Accuracy Assessment of Elevation Models Derived from LiDAR, International Symposium & Exhibition on Geoinformation 2011 ( ISG & ISPRS 2011). Shah Alam, Selangor, Malaysia.

[4] MDL (2011) Measurement Device Ltd. http://www.mdl.co.uk/, [cit. 2011-11-25].

[5] Azman, A. (2010) Tree Mapping System. Unpublished MSc. thesis, University Teknologi MARA, Shah Alam, Malaysia.

[6] ESRI (1997) Using ArcView 3D Analyst. ESRI Inc, California, USA.

[7] Uren, J. and Price, W. F. (1994) Surveying for Engineers. London: The Macmillan Press Ltd.

[8] Irvine, W. (1995) Surveying for Construction. McGraw-Hill International (UK) Limited. 4th Edition.

[9] Khali, A., Ibrahim, S. and Norizan A. P. (2001) Tree mapping using DGPS as part of precision forestry practices. Conference on Remote Sensing Technology & GIS in Forestry Application, Malaysian Centre for Remote Sensing (MACRES), Kuala Lumpur, Malaysia.

[10] Jaafar, J. (2000) An evaluation of the generation and potential applications of digital surface models. Unpublished Ph.D. thesis, The University of Nottingham, Nottingham, UK.

# ADVANCED DATA STRUCTURES FOR SURFACE STORAGE

Karel JANECKA[1] and Michal KARA[2]

[1]Department of Mathematics, Faculty of Applied Sciences, University of West Bohemia, Univerzitni 22, 323 00, Pilsen, Czech Republic

*kjanecka@kma.zcu.cz*

[2]Georeal, Halkova 12, 301 00, Pilsen, Czech Republic

*michal.kara@georeal.cz*

**Abstract**

Nowadays the amount of spatial data is rapidly increasing. One of the reasons is the modern technique for collecting data like laser scanning etc. The demand for an efficient data processing brings in also the requirement for their effective storage. Due to the amount of data it seems useful to store them in a spatial database. Modern database management systems are more and more supporting the processing multidimensional spatial data. Most of them allow using an object-relational data model. It means that multidimensional data can be stored as an object using the appropriate abstract data type. The crucial factor for maintaining a large volume of spatial data inside a spatial database is an availability of special data structures which are suitable for such kind of data. The paper introduces the advanced data structures which are designed to support storage of surface data inside the spatial database. The focus is put on the extension of Oracle database management system for processing spatial data – Spatial as there are special data structures allowing direct storage of data representing the surface as Triangulated Irregular Network or Point Clouds. These structures are complicated in comparison with the standard Open Geospatial Consortium data types for multidimensional data. They provide the concept which seems to be useful but there are some bottlenecks for usage of real data. The paper presents the finding out about the underbellies and proposes some possible improvements.

**Keywords: Data structures, spatial databases, surface storage**

## INTRODUCTION

The many initiatives have been done aiming to propose suitable data structures for a surface storage. Many earlier proposed data structures could be found e.g. in [1] or [2]. In such cases a surface is stored mostly in a proprietary file format. Some of the proposed structures could be stored in relational database management systems using the atomic data types. The Fig. 1 illustrates a vertex-based Triangulated Irregular Network (TIN) in which points include a clockwise list of neighbouring vertices. This structure could be simply implemented using the atomic *number* data type. Nowadays there are several important reasons why to use spatial database management systems (SDBMS) instead of files. The main ones could be that SDBMS support transactions, multiple users and editing and Structured Query Language (SQL).

The spatial database management systems can be characterized by many factors like a prize, support, speed of queries or memory consumption. Such comparisons could be found for example in [3] or [4]. Another approach can be to compare SDBMS according to the supported geometries and functions. This criterion is more related to the aim of this paper to present the advanced data structures which are suitable for direct surface storing. In this matter the most mature implementations of a spatial type system [5] were explored and only in Oracle Spatial the appropriate specialized data structures for storing surface were discovered.

Open Geospatial Consortium (OGC) delivered the standard [6] which defines Geometry object model. This model consists of many geometry classes supporting the storage of various types of geometries. One of the main classes is the *Surface* class. The *PolyhedralSurface* class is the instantiable subclass of *Surface*. It is a simple surface consisting of some number of Polygon patches or facets. OGC Geometry object model also contains a class *TIN* which aims to support storage of TIN inside spatial databases. According to the OGC Geometry object model TIN could be defined as a contiguous collection of triangle patches which share common boundary segments.

For practical database implementation OGC proposed several ways how to implement Geometry object model in SQL:

- objects are implemented in SQL92 using numerical types for storage of geometric objects;
- objects are implemented in SQL92 using binary types for storage of geometric objects;
- objects are implemented in SQL92 with geometric types.

In SQL92 with geometric types the columns in a feature table are defined by feature attributes; one or more of the feature attributes will be a geometric attribute. The geometry of a feature is one of the feature attributes and is a SQL geometry type as defined in OGC Simple Feature Access standard [7]. These SQL geometry types are based upon the OGC Geometry object model.

Mostly all the main spatial database management systems support this OGC Simple Feature Access standard. Some database systems (e.g. MySQL) support only version 1.0 but most of them (e.g. PostGIS, Microsoft SQL Server, Oracle Spatial) provide user with version 1.2 allowing modelling 3D geometries. ESRI Geodatabase data format allows storing TIN as a planar graph where nodes are connected by edges to form triangles. Edges connect nodes that are close to one another. PostGIS spatial extension of PostgreSQL database management system has constructors for creating 3D geometry. TIN is modelled as a special case of polyhedral surface which is collection of adjacent triangles. Very similar situation is for Microsoft SQL Server database system. From this point of view Oracle Spatial is an example of SDBMS providing data structures and mechanisms suitable directly for TINs or point clouds.

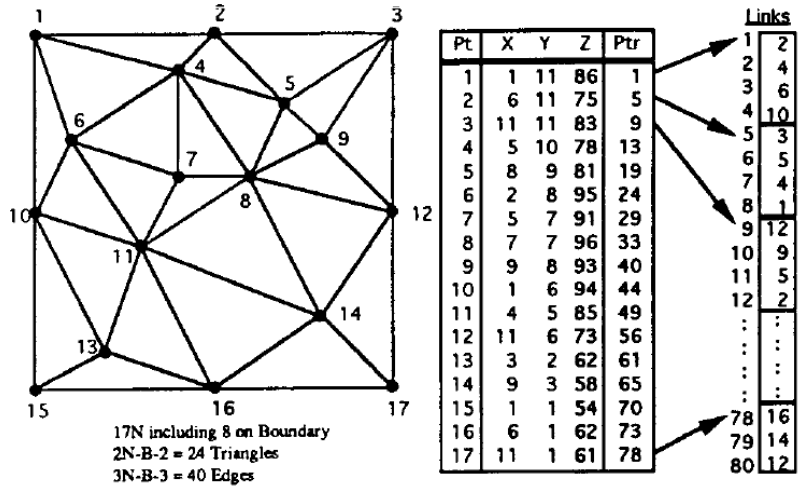



Fig. 1. TIN data structure with a clockwise list of neighbouring vertices [8].

The two advanced data structures suitable for surface storage in the spatial database (here specifically in Oracle Spatial) will be described. Both structures are based on abstract data type mechanism. The first one is designed to store surface as a point cloud. The second one supports storage of surface as a TIN.

## ADVANCED DATA STRUCTURES

In order to add a new data type to a relational database system and to take full advantage of a new data type there are generally three different types of operations that may be specified. The first type is the primitive operations which define constants for the new type, comparisons, inserting and updating values (or how the values are actually stored). The second type of operation is the aggregate. Aggregates are operations such as the COUNT, SUM, MAX, MIN provided by the relational database management systems. Aggregate is defined to be any mapping from a single relation to a single value of a specified (possibly programmer-defined) data type. The third type of operation, the transformation, takes a relation as its parameter and returns a relation as the result [9]. Transformations could be used to convert a relation containing a surface in one representation to a relation containing a different representation for the surface. For effective surface storing and processing in the spatial database some additional conditions must be fulfilled. The specialized data structure for a surface storage must support creation of the spatial index (e.g. R-tree) as real datasets consist of large volume of data. This is also requirement on the spatial database management system to effectively build a spatial index upon these structures.

### Object-Relational Approach

The advanced data structures proposed for surface data and related information storing are defined using several kinds of data types. First standard database data types like NUMBER or VARCHAR are used. Due to the real spatial data bulkiness the coordinates of single points are stored in attributes of the binary large data type. For modelling other surface's attributes

like minimum bounding objects the data types based on the abstract data types are used. Abstract data types extend the classical relational paradigm about the possibilities and advantages of object oriented approach to spatial data modelling. This is a hierarchal database design that is based on objects and parent-child relationships. Within object database principles and abstract data type can take on multiple forms. This allows multiple types of data to be stored in the database structure. An example of such type is SDO_GEOMETRY object data type used for storage of the minimum bounding objects. These bounding objects have the crucial role for spatial queries evaluation. The SDO_GEOMETRY object data type is defined as follows:

```
CREATE TYPE SDO_GEOMETRY AS OBJECT (

       SDO_GTYPE     NUMBER

       SDO_SRID      NUMBER

       SDO_POINT     SDO_POINT_TYPE

       SDO_ELEM_INFO MDSYS.SDO_ELEM_INFO_ARRAY
       SDO_ORDINATES MDSYS.SDO_ORDINATE_ARRAY );
```

The *sdo_gtype* attribute indicates the type of the geometry. The possible geometries are those as specified in [7]. The *srid* attribute identifies a coordinate system. The *sdo_point* attribute stores the coordinates if a spatial object has the point geometry. For line and polygon geometries the attributes *sdo_elem_info* and *sdo_ordinates* are use. Both attributes are defined using the user defined data types. The *sdo_elem_info* attribute contains information about how to interpret coordinates stored in the *sdo_ordinates* attribute.

The data structures based on abstract data types allow an object-relational modelling spatial objects like a surface. Then the advantages of both attitudes can be used. It is possible to store the entire description of the surface in one table in one column and make relations with other objects which are stored in a spatial database.

Some practical solutions use SDO_GEOMETRY object data type only for simple storage of surface. The example of such approach could be some digital hypsography databases where however database management system serves only as a data repository [10] for storage of coordinates of points. The analysis of hypsography data and creation of surfaces are done outside the database. In case when the spatial database is used only to keep surface data (e.g. TIN) it is possible for a surface storage to use also the Well Known Text [6] or Well Known Binary [6] data formats.

**Surface as a Point Cloud in Spatial Database**

Current methods for collecting geographic data such as aerial and terrestrial laser scanning allow to receive a large amount of detailed data in a relatively short time. The laser scanning data are obtained in the form of point clouds. An important factor for the usage of the obtained data is their further processing. According to the scope of the data it is suitable to store and analyze such data in a spatial database management system. The number of points in a point cloud is usually thousands or millions. For such amount of data the SDO_GEOMETRY object data type is not much suitable. The reason is a restriction for the maximum of the stored

numbers (vertices). Furthermore the disadvantage of using SDO_GEOMETRY could be that all points are stored in one array. It is not possible to query only a part of this array of vertices. The whole array must be loaded into a memory and due to the volume of the data this can cause performance difficulties. That requires special data structures which enable effective point clouds management directly in the database. It means that each point in point cloud represents the point of surface. This allows generating a surface upon such stored data.

From the logical data model point of view each point cloud is stored using two tables - *Base Table* and *Block Table*. The base table is for storage of user information about a point cloud. As such it contains user defined columns. The block table serves as a place for storage of the point cloud. This point cloud placement in the block table can be done in several ways. The entire point cloud can be stored in one row of the block table. However for the real data dataset containing a huge amount of points the point cloud is stored in the more blocks (rows). Fig. 2 captures this point cloud division.

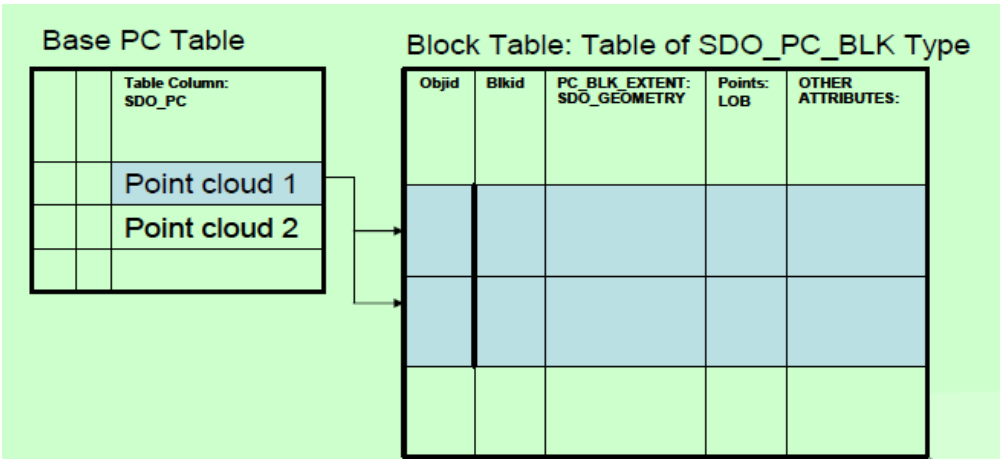



**Fig. 2.** The point cloud theoretically consists of the several blocks [11].

Physically the surface as a point cloud is stored and modelled using the three main data object structures:

- SDO_PC
- SDO_PC_BLK
- SDO_PC_BLK_TYPE

*SDO_PC object data type*

SDO_PC stores reference to smaller blocks covering the points according to their placement in a space. This allows selecting only these blocks (set of data) which are really requested. It is not needed to load all point cloud and so the time complexity won't usually be so high. The information regarding the point cloud is stored in a table which is called *Base Table* containing the SDO_PC column. The particular blocks containing the point cloud data are stored in a separate table. This table is referenced as *Block Table* and is defined using the specialized SDO_PC_BLK data type. The relation between the base table and block table is captured in Fig. 3. The number of points per block is given by user and depends on the size of input data

set. The setting maximum number of points per block will influence the effectiveness of analysis and for example the visualization [11].

SDO_PC is further used for storage of the main characteristics of the point cloud. There the condition for a division of the points into smaller logical data blocks is stored. On principle this condition is expressed as the number of points in one block. For a spatial query evaluation this condition has a fundamental meaning. It is more effective to work with smaller units as these blocks could be spatially indexed. For this reason each logical data block is bounded by the minimum bounding rectangle (of SDO_GEOMETRY object data type). This leads to the faster spatial query evaluation. There is no possibility to set up the boundaries of block manually. It means that in case where the points will be more irregularly distributed this cannot be technically taken into consideration. For points splitting into particular blocks some heuristic value can be used. In case of Oracle Spatial the spatial database engine aims to split the points uniformly that each logical block will cover the same number of points. Furthermore attribute *pc_extent* contains object representing the spatial extent of the point cloud i.e. the minimum bounding object enclosing all objects in the point cloud. This can be used is some spatial analysis.

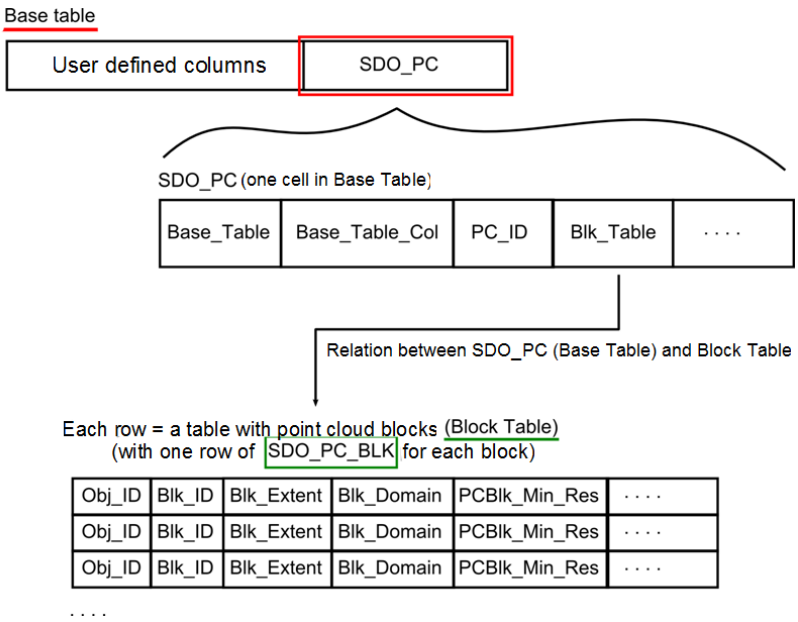



**Fig. 3.** The relation between the Base table and Block table [11].

*SDO_PC_BLK object data type*

This is the main building block of the point cloud database structure. The blocks of points and their coordinates are stored using this object data type. For each point both the identifier of the particular block and identifier of point are stored. Due to the expected number of points in real data sets the binary data type is used for storage of coordinates as can be also seen from the Fig. 2. This data type stores among others information about the minimum and maximum resolution level at which the block is visible in a query.

The division of points into the blocks is done in several steps:

- computation of the total number of points,
- determination of blocks which will be fully filled by points up to the maximum (threshold value),
- determination of two blocks which will be equally filled by the rest of points (see Fig. 4).

| | NUM_POINTS |
|---|---|
| 1 | 7115 |
| 2 | 7115 |
| 3 | 10000 |
| 4 | 10000 |
| 5 | 10000 |

**Fig. 4.** The principle of distribution of points into the blocks [12].

The default threshold value for the maximum number of points in one block is 5000. According to the size of input data it is possible to set a different value as illustrated in Fig. 4 where the threshold value was set on 10000 points per block. The next very important attribute for spatial queries is the *Blk_extent* attribute which stores a block spatial extent as an SDO_GEOMETRY object. The spatial index is build upon this attribute which enables faster query evaluation. This situation is captured in Fig. 5. If a clip operation returning the points that are within a specified query window is done an object of SDO_PC_BLK_TYPE is returned which is defined as table of SDO_PC_BLK [13].

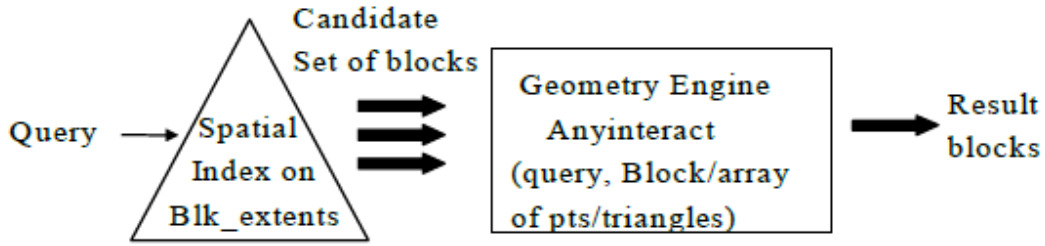



**Fig. 5.** A spatial query evaluation uses the attribute Blk_extent and spatial index [12].

The block is retrieved only in a case that the query window intersects the particular spatial extent and if the minimum (maximum) resolution interval of the block intersects the minimum (maximum) resolution interval of the query. The detail description of all available attributes (for SDO_PC and SDO_PC_BLK) mentions [13].

**Surface as a TIN in Spatial Database**

Oracle Spatial supports direct storage of TIN in the special data structure which is based on abstract data type mechanism. Fig. 6 shows the principle of loading and processing input data into Spatial. The real datasets forming the TIN are loaded from the source flat files or some

existing point tables. Furthermore nowadays the data from laserscanning are the source of very precise surface data. Spatial database management system can transform the source point (cloud) data into specialized TIN data structure. There can be done some spatial operations like clipping upon the TIN data directly on the side of spatial database.

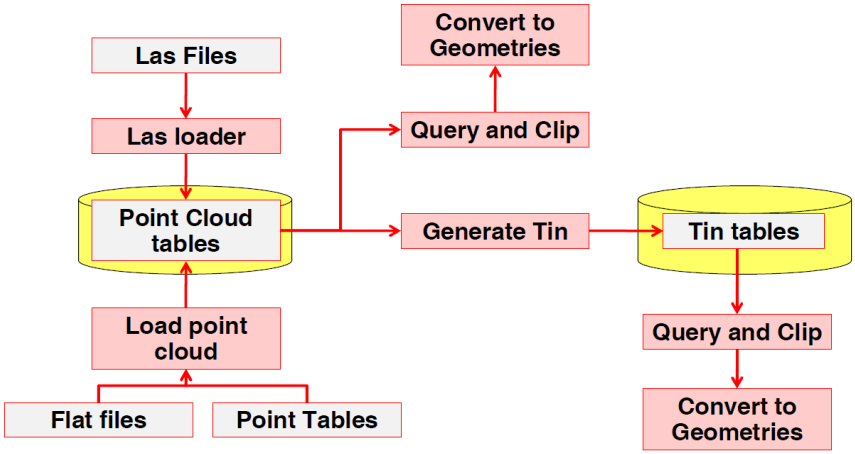



**Fig. 6.** The principle of processing input surface data and storage in a spatial database [14].

In Spatial a TIN is logically structured into blocks of data as Fig. 7 illustrates. This division allows more effective storage and retrieval of data. A spatial query on select of only some restricted area will select only particular blocks for which the query is relevant. In Spatial for a TIN creation the Delaunay triangulation algorithm is used. This algorithm calculates the most appropriate combination of triangles from the input data set. The aim at best is to create the equilateral triangles containing no other points. The input data sets are in form of a point cloud.

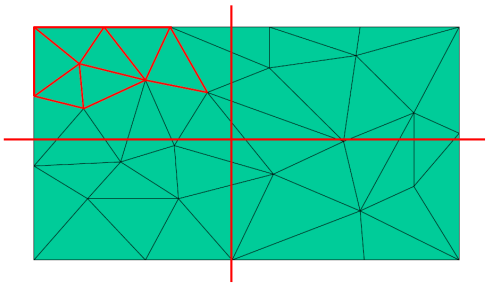


**Fig. 7.** TIN is logically structured into smaller blocks [15]

The TIN surface is stored and modelled using the three main data object structures:

- SDO_TIN
- SDO_TIN_BLK
- SDO_TIN_BLK_TYPE

*SDO_TIN object data type*

The description of a TIN is stored in a single row in a single column of object type SDO_TIN in a user-defined table (also called *base table*). This table then contains all attribute information related to the TIN. SDO_TIN holds the parameters for partition of the TIN. For spatial indexing purposes and faster spatial queries evaluation the minimum bounding object enclosing all objects in the TIN is available. The minimum bounding rectangle is of SDO_GEOMETRY object data type. The particular blocks of data are stored in a separate table modelled using SDO_TIN_BLK type. Some other attributes which are currently available in the SDO_TIN object data type are not yet supported for practical work. For example the break lines are not fully supported [13].

The relation between the table containing the SDO_TIN object and TIN block table is captured in Fig. 8.

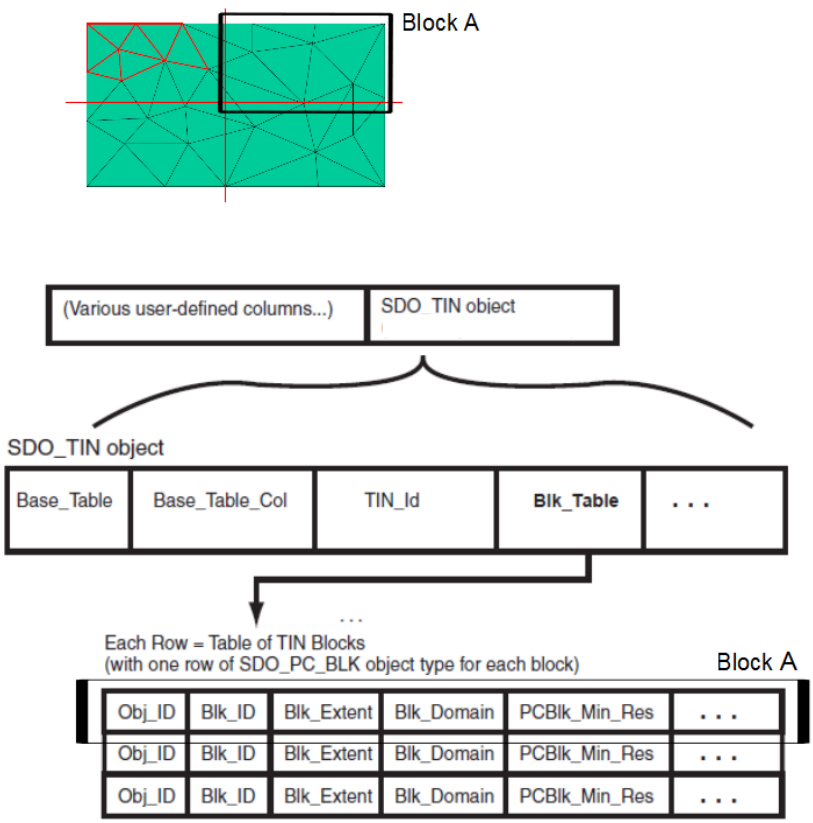



**Fig. 8.** Storage model for TIN data.

The attribute *tin_stop_lines* represents the stop line or lines in the TIN are the line string or multiline string stored as SDO_GEOMETRY object. Stop lines typically indicate places where the elevation lines are not continuous such as the slope from the top to the bottom of a cliff. Such regions are to be excluded from the TIN.

*SDO_TIN_BLK and SDO_TIN_BLK_TYPE object data types*

The TIN block table is modelled as a table with rows of the SDO_TIN_BLK type. It contains information about a spatial extent of the block, the minimum (maximum) resolution level at which block is visible in a spatial query. The spatial extent is modelled as SDO_GEOMETRY object. In the block table for each block the information about the number of points in the block is stored. The triangles shaping the TIN and points are stored in different attributes. Points' coordinates are stored using a binary data type. This binary attribute consists of an array of points storing both identifier of particular block and point as such. The triangles are stored also as a binary sequence of data. Each vertex is specified by the identifier of the block and identifier of the point. If a clip operation is done then an object of SDO_TIN_BLK_TYPE is returned. This object data type is defined as table of SDO_TIN_BLK [13].

When the existing TIN is modified then the entire TIN must be rebuild. In case of the removing the point from the existing TIN it must be done explicitly from the base table. The standard Spatial SDO_TIN_PKG package allows transforming the TIN data into the SDO_GEOMETRY object data. Then it is easy to visualize the TIN data without any other data transformation as illustrated in Fig. 9.

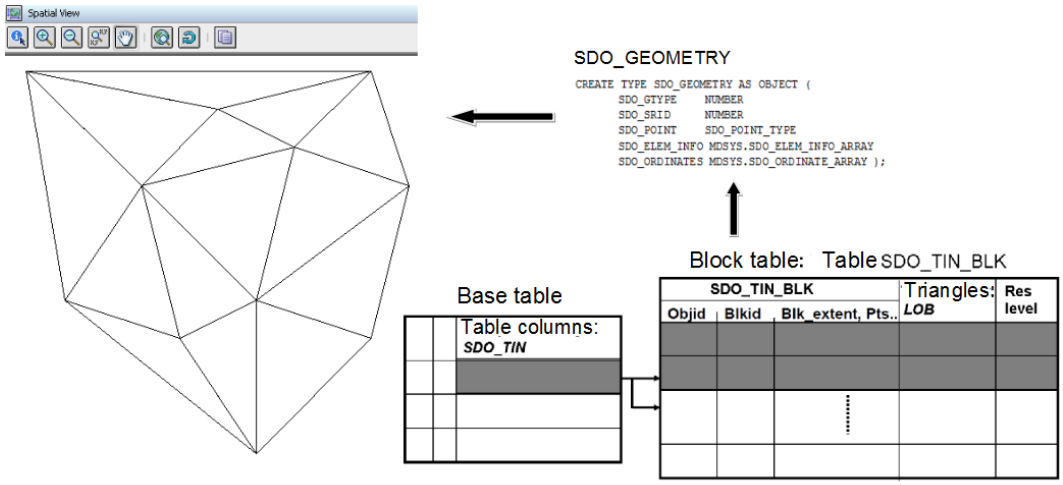



Fig. 9. Example of TIN visualization.

**Further Processing of the Surface Data in Spatial Database**

From the spatial analysis point of view the main SDMBSs support mainly such data types like SDO_GEOMETRY. The reason is that it is relatively easy to create a function working with features which are in spatial database modelled as objects. However if the surface is stored using the appropriate data structure the best for spatial queries (analyses) is if a spatial database would support operations also upon these advanced data structures. Due to the novelty and particularity of the presented data structures there is relatively weak build-in support in Spatial. For spatial analyse which should be done directly in a spatial database the possible solution seems to be either to use the implicit function transforming the surface data or its part into the data type like SDO_GEOMETRY or using a function-based mechanism.

Then it is possible to apply all available standardized functions working with this object data type as illustrates the following example showing the selection of all blocks within a specified distance from the reference object.

```
SELECT blk_id, obj_id
FROM   pc_blktab
WHERE  SDO_WITHIN_DISTANCE(
         VRAT_GEOM(points,num_points),   --function returning the geometry
       description
         SDO_GEOMETRY(2003, NULL, NULL,
           SDO_ELEM_INFO_ARRAY(1,1003,3),
           SDO_ORDINATE_ARRAY(-812783,-1079712,-812782,-1079712)),
       'distance=100 min_resolution=10') = 'TRUE';
```

The main advantage of using a function-based index is that the data remain store in the specialized data structure and are ad-hoc materialized and hence no other space for storage of the surface data is needed. Table 1 gives an overview about limitations for storing.

**Table 1.** Limitations for particular object types.

| Object Type | Attribute | Limitation |
|---|---|---|
| “classic” SDO_GEOMETRY | SDO_ORDINATE_ARRAY | 349 525 of 3D points |
| “uber” SDO_GEOMETRY | SDO_ORDINATE_ARRAY | 3 333 333 of 3D points |
| SDO_PC_BLK | POINTS | BLOB - 32 bytes for each point |
| SDO_TIN_BLK | SDO_POINT | for point cloud data - 12 MB (for Oracle Database Release 11.1.0.7 or later); approximately 393k of 3D points |
| SDI_TIN_BLK | TRIANGLES | BLOB - limitation is not documented |

## CONCLUSION

The trend of storage of large amounts of data into spatial databases is evident. If the spatial database should serve not only as data storage then there is the need for specialized data structures. These structures should allow to model, store and process the surface data effectively. The structure should provide a possibility to store also user-defined information. If the focus is put on modern spatial database management systems then the more specialized data structures based on abstract data types can be found in Spatial. As the structures implementations are relatively new there is a lack of experience with real data but some future

improvements are obvious. The detailed information about how to use the presented data structures implemented in Oracle Spatial for the storage and processing of the real surface data can be found mainly in [11]. These advanced data structures are fully integrated into the spatial database management system and support spatial indexing and spatial analysis as well.

The concept of building such structures upon abstract data types seems to be suitable for further development of both data structures and related functions as computation of inclination of slope etc. The weakness for the usage of the current TIN data structure implementation can be no support for direct data visualization (the same for the point cloud data structure). In case that some spatial analysis will be done on the side of spatial database no presence of a native visualization tool could be limiting but it is only a minor issue. As some points may lay directly on the boundary between two blocks the data could be stored redundantly. Some triangles can also belong to more than one block as illustrated in Fig. 7.

For the point cloud data structure the storage of additional non-spatial attributes related to the points will require an extension of this data structure. At this moment it is possible to store only information related to the point cloud as a whole in user defined attributes (columns). In case that it will be necessary to update the point cloud then this new actualized point cloud must be created from the beginning. For example there is no default function for adding one or several new points into the existing point cloud. There is also no exact algorithm about how to set the threshold value for maximum number of points in one block. The result from the testing is that this should be done according to the size of input data with aim to receive acceptable performance results for spatial analysis.

The further improvement and development of the proposed data structures should focus on an additional support for surface modification and analyses. This fact should be taken into account for example by architects who are responsible for creation of geoinformation infrastructure or GIS specialists who would like to use the spatial database in their work flow processing the surface data. The procedural extension of the Structure Query Language allows to propose and implement own functions and procedures which will use the built-in spatial functions. Due to the fact that surface data are stored directly in spatial database it will enable not only to create a surface and storing it in the appropriate advanced data structure but also e.g. to make the surface generalization and further spatial analysis directly by means of SDBMS.

## REFERENCES

[1] Kidner, D.B. (1991) Digital Terrain Models for Radio Path Loss Calculations, Unpublished Ph.D. Thesis, Department of Mathematics & Computer Studies, The Polytechnic of Wales, U.K.

[2] Packer, T.K., Fowler, R. J., Little, J.J. and Mark, D.M. (1978) The Triangulated Irregular Network, Proceedings of the Digital Terrain Models (DTM) Symposium, ASP/ACSM, May 9-11, 516-540.

[3] Boston Geographic Information Systems (2011) Compare SQL Server 2008 R2, Oracle 11G R2, PostgreSQL/PostGIS 1.5 Spatial Features. http://www.bostongis.com/PrinterFriendly.aspx?content_name=sqlserver2008r2_oracle11gr2_postgis15_compare, [cit. 2011-12-03]

[4] Boston Geographic Information Systems (2011) Cross Compare SQL Server 2008 Spatial, PostgreSQL/PostGIS 1.3-1.4, MySQL 5-6. http://www.bostongis.com/PrinterFriendly.aspx?content_name=sqlserver2008_postgis_mysql_compare, [cit. 2011-12-03]

[5] Greener, S. (2011) Oracle Spatial for PostGIS Users – Understand, Isolate and Migrate. The SpatialDB Advisor. http://download.osgeo.org/osgeo/foss4g/2009/SPREP/2Thu/Parkside%20Auditorium/1300/, [cit. 2011-12-03]

[6] OGC 06-103r4 (2006). OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture. Open Geospatial Consortium Inc.

[7] OGC 06-104r4 (2006). OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option. Open Geospatial Consortium Inc.

[8] Kidner, D. B. and Smith, D. H. (1993) Data structures for terrain modelling and ground cover data. Proceedings of Terrain Modelling and Ground Cover Data for Propagation Studies.

[9] Oosborn, S. L. and Heaven, T. E. (1986) The Design of a Relational Database System with Abstract Data Types for Domains. ACM Transactions on Database Systems. Volume 11 Issue 3.

[10] Potrebova, L. (2011) Technology solution of the national hypsography databases. University of West Bohemia. Pilsen.

[11] Kara, M. (2011) Storage and analyses of the point cloud in Oracle Spatial. University of West Bohemia. Pilsen.

[12] Ravada, S., Horhammer, M. and Kazar, B. (2011) Point Cloud. Storage, Loading, and Visualization. http://www.cigi.illinois.edu/cybergis/docs/Kazar_Position_Paper.pdf, [cit. 2011-04-15].

[13] Murray, C. (2009) Oracle Spatial Developer's Guide 11g Release 1. Oracle, 2009, http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28400.pdf , [cit. 2010-03]

[14] Kara, M. (2010) TIN in Oracle Spatial. University of West Bohemia. Pilsen.

[15] Ravada, S., Kazar, B. M., Kothuri, R. (2009) Query Processing In 3-D Spatial Databases. 2008. http://3dgeoinfo.Uos.Ac.Kr/Data/Presentation/All/Session02/1_3d_Query_Processing_3dgeoinfo08_Presentation.Pdf, [cit. 2009-12-10].

# SEGMENTATION OF LIDAR DATA FOR EXTRACTING BUILDING'S ROOF SHAPES, USING FUZZY LOGIC CONCEPTS

Nikrouz MOSTOFI[1] and Farhad SAMADZADEGAN[2]

[1]Dept. of surveying and Geomatics Engineering, Faculty of Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

*n_mostofi@azad.ac.ir*

[2] Dept. of surveying and Geomatics Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

*samadz@ut.ac.ir*

**Abstract**

Building extraction becomes one of important fields of research in photogrammetry and machine vision. Involving airborne laser scanner (LiDAR) in geomatics, as an active sensor of 3D data acquisition, great evolution happened in data preparing to 3D outdoor object modeling. In consequence a great trend in developers and geomatics science researcher's communities leaded them to develop techniques in extraction of interesting objects from point cloud such as buildings. The LiDAR data don't explicitly contain any geometric information and feature of objects. The feature such as planes, lines and corners can be only indirectly extracted by segmentation algorithms.

In this research, a method for segmentation of LiDAR data for extracting buildings is proposed. In the proposed method, input data is an irregular LiDAR data and an adaptive clustering method for roof plane extraction from LiDAR data is implemented. So, to extract the roof structure, an assumption of planarity has been made. It is assumed that the roof can be modeled by a set of planar segments. In the proposed method the clustering is done without having any prior information about the number of the clusters. The clustering is done using FCM algorithm with considering maximum cluster number and then the extracted cluster will be modified using an iteratively split and merge technique. This approach was evaluated using some buildings of used data set and the results prove high efficiency and reliability of this method for extraction of buildings.

**Keywords: Clustering, Segmentation, LiDAR data, Building extraction, Fuzzy Logic.**

## INTRODUTION

Airborne laser scanning is an established technology for highly automated acquisition of digital surface models. Furthermore, in recent years LIDAR data has become as a highly acknowledged data source for interactive mapping of 3D man-made and natural objects from the physical earth's surface. The dense and accurate recording of surface points have encouraged research in processing and analysing the data to develop automated processes for feature extraction, object recognition and object reconstruction [1]. Also, the need of achieving 3D model of buildings with acceptable precision, become important information in several fields in engineering. For example monitoring of changes in urban area or in the case

of managing crisis, need the accurate 3D model of cities. But achieving a reliable automatic method needs so much tries. Still a challenge of surface extraction from laser scanning is the high amount of data to process. Strategies for efficient data processing such as segmentation are needed to extract features from high resolution LIDAR data. LIDAR data is dense, with high accuracy, but one still needs to extract higher level features from it [2]. Building representations are needed in cartographic analysis, urban planning, and visualization. We propose "Split and Merge" method for clustering to find building segments using fuzzy reasoning strategy [3].

## PROPOSED METHOD

The simple way for extraction of building information from LiDAR data is partial and local analysis. This means that a part of point cloud should be analysed which is related to a building in object space [4]. So, it is possible to use a 2D map of area for selecting coincident points between a building and point cloud. In the other hand it is known that building roofs have clearly different height comparing with surrounding objects. This knowledge will be helpful for segmentation of buildings from ground points. So, after extracting clusters from probable building points, an efficient segmentation method is needed to generate parts of buildings from extracted clusters. So, after clustering step, a novel method will be implemented for segmentation of extracted clusters.

### Clustering Technique

Clustering is a process of assigning pixels to categories or clusters based on some logic which acts on similarity of the pixels feature vectors. Three clustering techniques are popular to use in photogrammetry are K-means (or hard C-means) clustering, fuzzy C-means clustering and competitive learning networks. K-means is a representative for a classical and well explored unsupervised classification algorithm [5]. Its counterpart in the fuzzy techniques is the fuzzy C-means algorithm which considers each cluster as a fuzzy set, while a membership function measures the possibility that each feature vector belongs to a cluster. With respect to the advantage of fuzzy concept, FCM algorithm will be used in the experiments [6,7].

### Fuzzy C-Means Clustering Algorithm:

Fuzzy C-means clustering (FCM), also known as fuzzy ISODATA, is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by its membership grade. Bezdek proposed this algorithm as an alternative to earlier (hard) K-means clustering [6]. FCM partitions a collection of n vector $x_i$ , $i = 1, \ldots, n$ into $c$ fuzzy groups, and finds a cluster centre in each group such that an objective function of a dissimilarity measure is minimized. The major difference between FCM and K-means is that FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1[8]. In FCM, the membership matrix $U$ is allowed to have not only 0 and 1 but also the elements with any values between 0 and 1 [9].

$$\sum_{i=1}^{c} U_{ij} = 1, \forall j = 1,..., n \quad (1)$$

The objective function for FCM is then a generalization of K-Means Equation as following:

$$J(U, c_1,....,c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \left\| x_j - c_i \right\|^2 \quad (2)$$

Where $u_{ij}$ is between 0 and 1; $c_i$ is the cluster centre of fuzzy group $i$. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership $u_{ij}$ and the cluster centres $c_i$ by:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}, \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{\left\| x_j - c_i \right\|}{\left\| x_j - c_k \right\|} \right)^{2/(m-1)}} \quad (4)$$

The FCM clustering algorithm is composed of the following steps:

Step 1: Initialized the membership matrix $U$ with random values between 0 and 1 such that the constraints in Equation (1) are satisfied.

Step 2: Calculate $c$ fuzzy cluster centres $c_i$ , $i$= 1, ... , $c$. using Equation (3).

Step 3: Compute the cost function (objective function) according to Equation (2). Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Compute a new $U$ using Equation (4). Go to step 2.

The cluster centres also can be initialized firstly and then iterative process carried out [10], [11], [12].

**Segmentation with Split and Merge Technique**

It is obtainable that LIDAR data from a roof plane must lie in a same group after transferring to object space. Normal segmentation algorithms for LIDAR data such as region growing satisfy this condition. But, in this research another condition will be used that called "split & merge". In case of splitting, for each point group of extracted cluster, a plane will be fitted using least square method. So, the angle between two planes will be calculated using normal vector of generated plane as shown in equation (5). If the value of the angle was smaller than threshold ( $\varepsilon$ ), two planes will be merged and a new plane will be generated.

$$arc\ \cos(nv_i . nv_j) < \varepsilon \quad (5)$$

After merging planes with discussed property, many of extracted classes will merge together and then the number of new class members will increase. so, the new normal vector will slightly change. This process will continue until satisfy tree main conditions:

- None of the contiguous planes should not have angle smaller than ten degrees.
- Distance between points and new plane should be smaller than a threshold.
- Population of a group member must be larger than a threshold.

The last condition, set for avoiding from generating unnecessary planes. It is desirable that the segments achieved from this method will be pure and homogenous. Flowchart of this method for development is shown in Fig. 1.

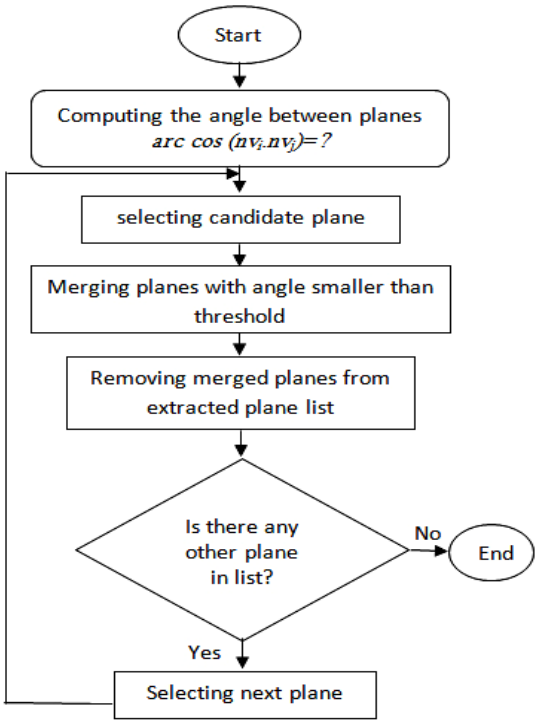



Fig. 1.Flowchart of merging technique.

**Using 2d Map as Additional Data**

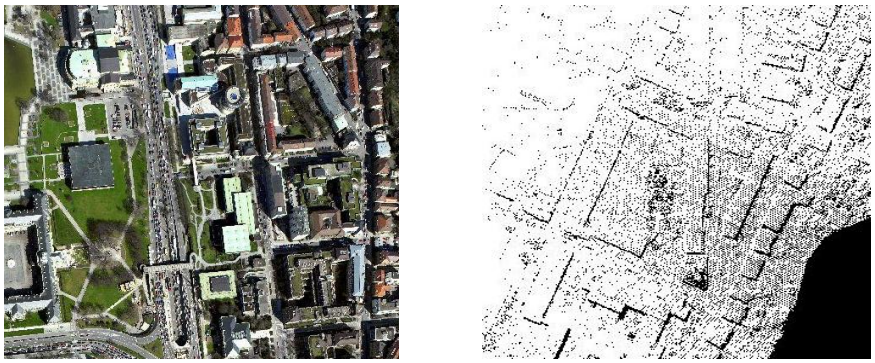
Two dimensional map of the related area is an important data source for validating clustering process with providing reliable point candidates for buildings. This additional data helps the algorithm for finding probable building position in shorter time interval and with high level of reliability. Therefore just probable areas positions from object space will be candidate for finding buildings. So, relation between object space and image space will be defined by Direct Linear Transformation (DLT) as shown in equation (6).

$$x = \frac{L_1X + L_2Y + L_3Z + L_4}{L_9X + L_{10}Y + L_{11}Z + 1}$$
$$y = \frac{L_5X + L_6Y + L_7Z + L_8}{L_9X + L_{10}Y + L_{11}Z + 1} \quad (6)$$

Where, $L_1,...,L_{11}$ are the coefficients between map space (x, y) and object space (X,Y,Z).

## EXPERIMENTAL INVESTIGATIONS

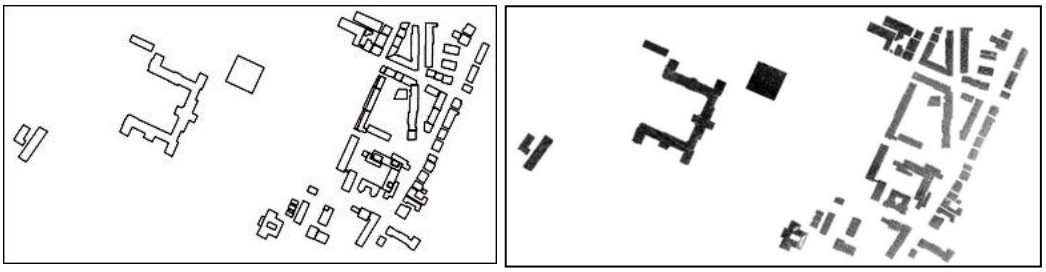
The airborne LIDAR data used in the experimental investigations have been recorded with TopScan's Airborne Laser Terrain Mapper system ALTM 1225 (TopScan, 2004). The pixel size of the range images is one meter per pixel. This reflects the average density of the irregularly recorded 3D points which is fairly close to one per $m^2$. Fig. 2 shows the point cloud and image of test area [4].



(a) (b)

Fig. 2.Test area, aerial image (a), point cloud (b)

For evaluating the proposed methods in this research, a guide user interface (GUI) developed in MATLAB programming software. In the first step the point cloud related to buildings truncated from data source using 2D map of the area using DLT equation. Fig. 3 shows, 2D map of the selected area and correspondence point cloud from source data set that selected with above approach.

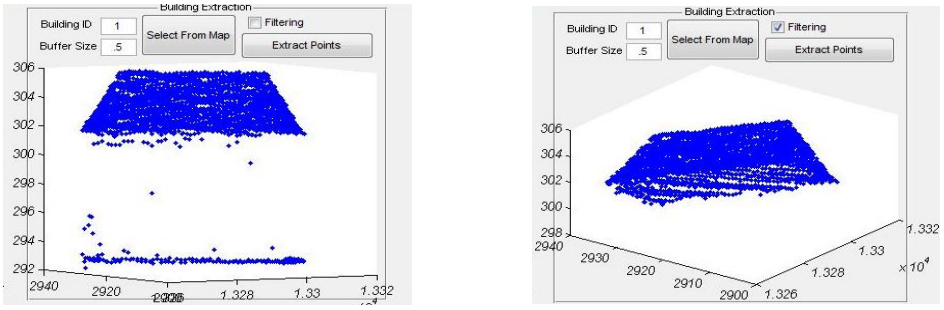


(a) (b)

**Fig. 3.** 2D map (a), and selected point clouds (b)

In the second step, pre-process stage for preparing data, selection of points from buildings polygons have been done by filtering non-roof points from data source as shown in Fig. 4. Also, selecting candidate points for probable buildings have been done by using an extra buffer area for being points involving roof and ground data. This technique will help the algorithm for separating roof points with the knowledge that roof points have more height in comparison with ground points. So, roof points will be separated from non-roof points.

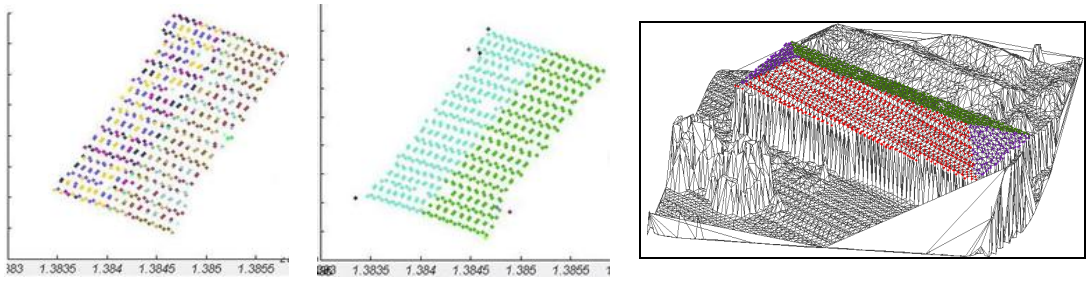
As discussed above, the link between image space and object space is obtained by DLT transformation. By comparing the elevation of roof and ground points with a logical buffer area, filtering is performed as shown in Fig. 4.




(a) (b)

**Fig. 4.** Pre-process of data, selecting points with 0.5 meter buffer (a), filtering of non-roof points (b)

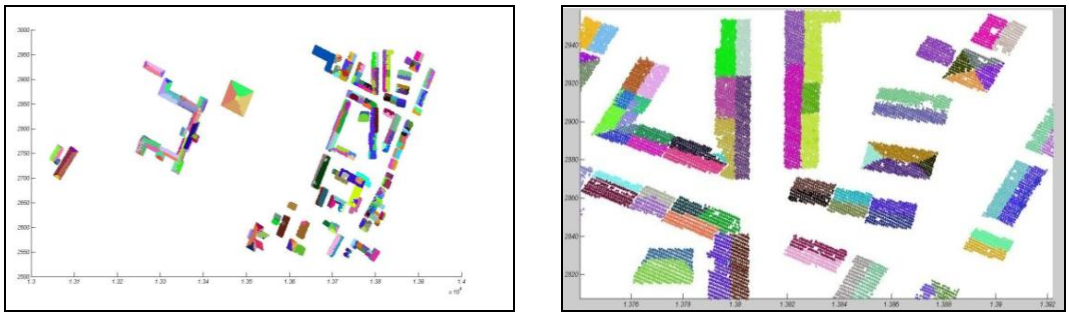
FMC method is used for clustering as mentioned. For automating this method, the maximum class number for clustering algorithm is set to fifteen classes desiring that the numbers of classes are so smaller than fifteen. So, it is claimable that this algorithm is independent from number of initial class number. After extracting features, the proposed method called "Split & merge" is implemented segmentation of roof data as shown in Fig. 5.



(a) (b) (c)

**Fig. 5.** Feature extraction result (a), segmentation with “split and merge” method (b), overlay with TIN (c)

Final result of segmentation in selected area is shown in Fig. 6. With respect to the density of point cloud in this research, extractions of the small parts of the roofs such as chimneys and dormers are neglected.



(a) (b)

**Fig. 6.** Result of segmentation on selected area (a), and in closer view (b).

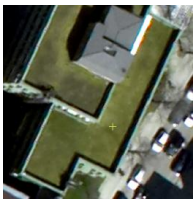
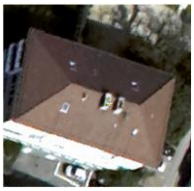
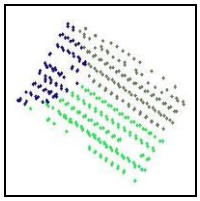
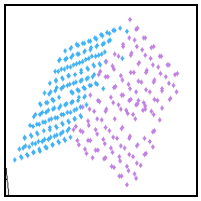
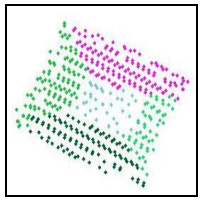
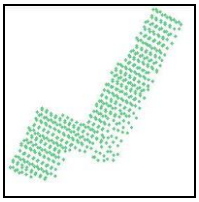
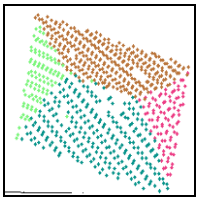
Accuracy of segmentation is computed by using three criterions for each building individually and final accuracy computed from getting mean value of these three values [12]. Criterions which used are Purity, F-measure and NMI, that for computing them, it is necessary to have the reference data [13]. Reference data is provided with visual extracting method for each roof plane. Numerical and graphical assessments of accuracy for each type of roofs are presented in Table 1 and Fig. 7.

**Table 1.** Accuracy assessment of segmentation considering roof types

| Roof type | Purity | NMI | F-measure | MEAN VALUE |
|---|---|---|---|---|
| **Flat** | 0.834 | 0.812 | 0.612 | **0.752** |
| **Gable** | 0.956 | 0.839 | 0.681 | **0.825** |
| **Hipped** | 0.829 | 0.787 | 0.583 | **0.733** |
| **Complex** | 0.667 | 0.423 | 0.482 | **0.524** |

Proposed method is tested in several roof types that are in test area and the result of segmentation is shown for each image (Table. 2). After generating valid segments of roof's structure, it is possible to fit a proper plane for each segment) [14], [15], [16]. So, reconstructed shape of the building will be the topic of future researches [16].

**Table 2.** Segmentation result with several types of roofs.

| | | | | | |
|---|---|---|---|---|---|
| Image |  |  |  |  |  |
| Segmentation Result |  |  |  |  |  |

## CONCLUSION

This paper presented and tested a new segmentation method based on the iterative "split & merge" method in extracting buildings and roof shapes form LiDAR data based on fuzzy C-means method. One of the advantages of this method is that it doesn't need the initial number of classes while it is possible to set this value to maximum number of desirable classes. The other advantage is the efficiency of segmentation technique that proposed in this research. Numerical and visual results, demonstrate high capability of segmentation process that specially achieved in gable type roofs because of distinguished slopes in these kinds of roofs. But, in the other hand because of the repeated filtering on clusters, small side effects such as noise, chimneys and dormers will have not a significant effect on final outcome.

But, in some cases non-homogeneous points also have been seen in some segments. As is shown in Fig. 8, despite the successive filtering and applied "split & merge" method, three points are located within the heterogeneous class. It should be noted that the number of these points doesn't make a large negative effect on the final results but removing these wrong points will make the proposed method more reliable. Furthermore, if it is necessary to extract dormers or chimneys from roofs, obtained methods may change in setting of thresholds. In such case, the resolution of the LiDAR data source is significant parameter for extracting accurate shape of roof details.

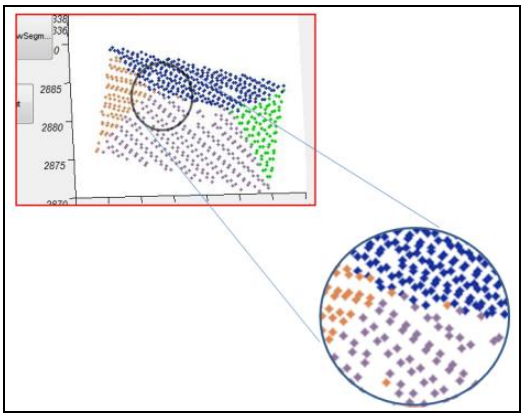



Fig. 8. There are three points in the heterogeneous classes

**ACKNOWLEDGMENT**

The authors would like to thank the Islamic Azad University's (South Tehran Branch) research deputy Office for supporting this work. We also would like to thank Dr. Michael Hahn from Stuttgart University of Applied Sciences for providing the data set used in the paper.

**REFERENCES**

[1] Arefi, H., Hahn, M. and Lindenberger, J. (2003) LIDAR data classification with remote sensing tools. Joint ISPRS Commission IV Workshop Challenges in Geospatial Analysis, Integration and Visualization II, Stuttgart, September 8- 9 2003.

[2] Elbernik, S. O. (2008) Problem in automated building reconstruction based on dense airborne laser scanner data,. The international archive of photogrammetry, Remote sensing and Spatial Information Sciences. Vol. XXXVII. Part B3a. Beijing 2008.

[3] Arefi, H., Hahn, M., Samadadzadegan, F. and Lindenberger, J. (2004) Comparision of Clustering Techniques Applied to Laser Data. ISPRS congress, Turkey, Istanbul, comm7.

[4] Samadzadegan, F. (2004) Object Extraction and Recognition from LiDAR Data Based on Fuzzy Reasoning and Information Fusion Technique, ISPRS congress, Turkey, Istanbul.

[5] Saeedi, S., Samadzadegan, F. and El-sheimy, N. (2009) Object Extraction From LiDAR Data Using an Artificial Awarm bee. CMRT09. IAPRS, Vol. XXXVIII, Part 3/W4, Paris, France, 3-4 September 2009.

[6] Teo, T.A. (2008) Parametric reconstruction for complex building from LiDAR and vector map using a dived and conqer strategy. IEEE Transaction on Geoscience and Remote Sensing, 63(11): 2084-2069.

[7] Lodha, S. K. (2005) Semi automatic roof reconstruction from aerial LiDAR data using K-Mean with refined seeding. ASPRS 2005 Annual Conference "Geospatial Goes Global: From Your Neighborhood to the Whole Planet" March 7-11, 2005 Baltimore, Maryland.

[8] Bezdek, J. C. (1981) Pattern Recognition with Fuzzy Objective Function. Plenum Press.

[9] Bezdek, J. C. (1987) Some non-standard clustering algorithms. In: Legendre, P. & Legendre, L. Developments in Numerical Ecology. NATO ASI Series, Vol. G14. Springer-Verlag

[10] Jain, K., Richard, C. and Dubes, C. (1998) Algorithms for Clustering Data. Prentice Hall.

[11] Forlani, G., Nardinocchi, C., Scaioni, M. and Zingaretti, P. (2003) Building reconstruction and visualization from LiDAR data. IAPRS International Archive of Photogrammetry and Remote Sensing and Spatial Information Sciences. Vol. 34, Part 5/W12, 151-156.

[12] Sampath, A. and Shan, J. (2008) Clustering based planner roof extraction from LiDAR data. ISPRS 2008 Annual Conference.

[13] Hoffman, A. D. (2007) An Approach to 3D Building Model Reconstruction from Airborne Laser Scanner Data Using Parameter Space Analysis and Fusion of primitives. PhD thesis: Munchen University, Geodesy Commission.

[14] Hofmann, A. D., Maas, H. and Streilein, A. (2003) Derivation of roof types bye cluster analysis in parameter space of airborne lasers canner point cloud. ISPRS Commission III WG3, Workshop, "3-D reconstruction from airborne laser scanner and InSAR data". IAPRS, International archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. 34, Part 3/W13, 112-117.

[15] Ma, R. (2004) Building model reconstruction from LiDAR data and aerial photographs. PhD thesis. The Ohio State University, Geodetic Science.

[16] Hofmann, A. D., Maas, H. and Streilein, A. (2003) Derivation of roof types bye cluster analysis in parameter space of airborne lasers canner point cloud. ISPRS Commission III WG3, Workshop, 3-D reconstruction from airborne laser scanner and InSAR data. IAPRS, International archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. 34, Part 3/W13, 112-117.

[17] Fostner, W. (1999) 3D-City Models: Automatic and Semi-automatic Acquisition Methods. Photogrammetrische Woche, Stuttgart.

# MODELLING THE UNCERTAINTY OF SLOPE ESTIMATION FROM LIDAR-DERIVED DEM: A CASE STUDY FROM LARGE-SCALED AREA IN CZECH REPUBLIC

Ivan MUDRON[1], Michal PODHORANYI[2], Juraj CIRBUS[1], Branislav DEVECKA[1], Peter BOBAL[1] and Jozef RICHNAVSKY[1]

[1]Institute of Geoinformatics, Faculty of Mining and Geology, VSB-TU of OSTRAVA, 17.listopadu 15/2172, 70833, Ostrava, Czech Republic

*Ivan.mudron@vsb.cz, Juraj.Cirbus@vsb.cz, Branislav.devecka@vsb.cz, Peter.bobal@vsb.cz*

[2]IT4Inovation Centre of Excellence VSB-TU of OSTRAVA, 17.listopadu 15/2172, 708 33, Ostrava, Czech Republic

*Michal.podhoranyi@vsb.cz*

**Abstract**

This paper summarizes the approach and results of error propagation analyses in the Olše and Stonávka confluence area. In terrain analyses the outputs of the aforementioned analysis are always a function of input. Four different digital elevation model (DEM) resolutions (0.5, 1, 5 and 10 meters from LIDAR cloud points) were examined with Root Mean Square Error (RMSE) rating up to 0.317 meters (10 m DEM). They all originate in LIDAR survey. In the analyses was performed a stochastic Monte Carlo simulation with 100 iterations. Article targets the error propagation for large-scaled area using high quality input DEM. The DEM data uncertainty (RMSE) was computed from samplings and ground research (RTK GPS). According to empirical error distribution it was used semivariogram to model spatially autocorrelated uncertainty in elevation. Second procedure modelled the uncertainty without autocorrelation using random N (0, RMSE) error generator. Based on Monte Carlo simulation method the initial DEM was repeatedly perturbed by the uncertainty. Finally, statistical summaries were drawn to investigate the expected hypothesis. As expected; the error in slopes is increased with the vertical error in input DEM. According to similar studies using different DEM input data, high qualitative LIDAR input data decreases the output uncertainty. Errors without spatial autocorrelation do not result to greater variance in resulting slope error. Therefore it should be challenged, if error propagation without spatial autocorrelation represents sufficiently the true state of the nature of the error representation. In this case; although the slope error results (comparing random uncorrelated and empirical autocorrelated error fields) did not showed statistical significant difference, the input elevation error pattern has not been normally distributed and therefore the random error generator realization is not suitable interpretation of true state of elevation errors. The normal distribution was rejected because of the high kurtosis and extreme values (outliners).

**Keywords: uncertainty, error propagation, monte carlo simulation, LIDAR-derived DEM, autocorrelation, RMSE, slope estimation.**

## INTRODUCTION

Although many studies and research in field of digital elevation model uncertainty and its error propagation were done, still there are some unacceptable assumptions about the error expected. Firstly, the DEM error disappears with precise data acquisition and optimal interpolation algorithm. Secondly the DEM error is thought to be as small as not affecting the outputs of the analyses using DEM input. Last but not least DEMs are assumed and used as error-free models of reality, even though the existence of elevation uncertainty and gross errors are widely recognized [1], [2]. In the last decades geomorphometry based on fine topscale DEMs have become popular in environmental science [3]. The accuracy of a digital elevation model is particularly important with its intended use [4]. So the misjudgements increased the importance of solving DEMs uncertainty and the error propagation problem. The awareness that uncertainty propagates through spatial analyses and may produce poor results that lead to wrong decisions has triggered a lot of research on spatial accuracy assessment and data quality management in GIS (e.g. [5], [6], [7], [8]). Information on the uncertainties in results from Geographic Information Systems (GIS) is needed for effective decision-making. Current GISs, however, do not provide this information [9], [10]. Furthermore there is the demand for presenting a level of accuracy (precision) [10]. Thus the long term vision in the research in spatial data uncertainty, accordingly DEM as well, has been to develop a general purpose “error button” for generating information systems (GIS) [11]. There are two main ideas how to implement this button. GIS could be incorporating the button into the product metadata [12] or in more sophisticated solution is seen the button as user-dependent, which offers various possibilities for refining the error model according to the user’s level of expertise [13]. The first steps towards the vision became a reality with building a data uncertainty engine, which implements the general framework for characterising uncertain environmental variables with probability models [14]. According to the authors many other research groups have worked on the design of an ‘error-aware GIS’, but very few have reached the operational stage. After the call for the development of geographical information systems that can handle uncertain data lasted at least for twenty years, Heuvelink developing the Data Uncertainty Engine (DUE) engine filled the gap [8]. Just the first step towards the solution of the error propagation problem has been made. The DUE must be further elaborated and improved. Sustained development of science and technology brought and will bring new methods of data collection and processing. As well DUE as other potential software, using different or same approaches, have to adjust to the changes. The usage of massive high-resolution DEMs based on airborne light detection and ranging (LIDAR) has renewed some assumptions. Two important factors appear to explain the lack of scientific knowledge about the use of LIDAR DEMs in uncertain-aware terrain analysis. Firstly, the common belief has been that the high quality of LIDAR DEMs [15], [16], [17] will make the uncertainty-aware terrain analysis unnecessary. Secondly, uncertainty propagation studies have typically made use of simulation methods, such as simulated annealing and sequential Gaussian simulation [18], that are unsuitable for massive data sets because of their poor scalability [19]. The aim of this paper is to analyze the aforementioned problems.

**DEM ERROR**

Spatial uncertainty is defined as the difference between the contents of a spatial database and the corresponding phenomena in the real world. Because all contents of spatial databases are representations of the real world, it is inevitable that differences will exist between them and the real phenomena that they purport to represent [20]. Error is defined as the difference between reality and a representation of reality. In practice, errors are not exactly known. At best, there is known the distribution of values. The chances are equal that the error is positive or negative. [14]. The paper follows the taxonomy in which error is a measurable part of the uncertainty and is well-defined (probability density function is well known etc.) [21]. This choice is justifiable because the semantics of elevation do not suffer from the conceptual ambiguities common in, for example, defining the error in area class maps [1].The detailed process by which the errors in a DEM are created depends on the type of DEM and how it was created. Whatever method is used, DEM estimates are affected by several error sources, which can be grouped generally under three main classes: accuracy, density, and distribution of data, surface characteristics, and interpolation algorithms [22], [23]. Uncertainty in DEMs originates from two sources, errors in the lattice (gross, systematic, random) and accuracy loss due to lattice representation of the terrain. It has been distinguish between positional and attribute uncertainty. Attribute uncertainty represents the deviation from true state of height and positional the shift in the object's position. Understanding the uncertainty is essential to correct modelling. Most frequently error in standard DEM products is reported as the Root Mean Squared Error (RMSE). Various methods have been used for estimating the RMSE. Most recently it is supposed to be estimated by comparison of elevations between the well located sites in a survey of higher accuracy with the elevation recorded in DEM at a minimum of 20 test points. The test points may be contour lines, bench marks, or spot elevations [Fisher 1992]. RMSE is based on the following formula:

$$RMSE = \sqrt{\frac{\sum (z-h)^2}{n}} \qquad \textbf{(1)}$$

where $z$ is the elevation recorded in the DEM; $h$ is the elevation measured at the higher precision and $n$ is the total number of tested locations (at least 20). The Gaussian error model (mean is the estimate of the true values and standard deviation is a measure of the uncertainty) makes only the most general assumptions about the processes by which the error has accumulated [24]. To achieve an improved estimate of the error for any particular area, a set of measurements made at a higher precision is required, at best having another DEM of the same area at a higher precision. In this case it is possible to compare all values [23]. The spot heights and DEM or both DEMs have to be constructed separately, independence is strictly required. When additional information is available about the structure of errors in data set, the Gaussian model should be replaced with a substituting more accurate pattern of error (non-stationary or stationary spatial dependent random error field). According to previous studies (e.g. [6], [24], [23], [13], [19], [25]) DEM errors are spatially correlated, autocorrelation is a natural characteristic of the error data.

Hunter distinguished three cases of spatial dependence. Case one is spatial independence (r = 0). The elevation of each point is considered to be spatially independent of its neighbours (r = 0). In other words, knowledge of the error present at one point provides no information on the errors present at neighbouring points, even though the elevation themselves may have similar values. The elevation realization *h* at location *x, y* is achieved by disturbing each observed elevation *z* at same location by an independent disturbance term *N (0, RMSE)*, which is normally distributed random variable with mean 0and standard deviation RMSE (**Eq. 2**):

$$h_{(x,y)} = z_{(x,y)} + N(0, RMSE) \quad \textbf{(2)}$$

Case two is spatial dependence (limit r =1). At the other extreme, spatial autocorrelation reaches maximum. All errors are perfectly correlated, and there is in effect only 1 degree of freedom in disturbance field being applied to the DEM. It is unlikely, that any DEM production process would generate systematic error in elevations. Case three is the spatial dependence (0 < R < 1). The case of positive correlation less than 1 is clearly most realistic [24] and the disturbance *N(0,RMSE)* is spatially correlated to certain range following the fitted error model. Exponential and Gaussian [1] spatial autocorrelation models were selected to represent the correlation of the DEM error in the DEM uncertainty propagation studies. First exponential and later Gaussian model has been found to be realistic and suitable for topography [18]. The study investigates the type of model, range and the spatially independent random error pattern.

**Error Propagation**

There are two main approaches in error propagation of a continuous variable: the analytical and the numerical error propagation. The analytical error propagation method uses an explicit mathematical model to describe the mechanisms of error propagation for a particular multi-criteria decision rule [26]. In numerical methods, the calculations are not made with exact numbers. Instead of exact numbers are processed numerically generated random data sets. Usually they are generated on a computer and in a case of complicated data or physical model for analytical approach. Simulation of the error is made stochastically using Monte Carlo simulation, which is further subdivided into unconditioned and conditioned [27]. Unconditional error simulation models are based on number of realizations of random functions. At their most basic, they comprise an algorithm to select independent and uncorrelated values drawn from a normal distribution which can be added to the original DEM. The problem with unconditioned simulation is that it still makes the assumption that the pattern of error is uniform over the study area or a wider region. Conditioned error models directly honour observations of error at the sample locations. Such observations might have been obtained by comparison between the DEM and a higher accuracy reference data set collected from the same area [27]. In other words, the parameters of error model vary depending on the specific location. Comparing the results of using different methods of error modelling, the best method, which gives widely implementable and defensible results, is that based on conditional stochastic simulation [23]. The most common uncertainty propagation

analysis approach makes use of Monte Carlo stochastic simulation [28].The utilisation of Monte Carlo simulation, which is the most flexible method for investigation the propagation of uncertainty in terrain analysis, is time-consuming [19]. Thus in this paper was used unconditional Monte Carlo simulation to propagate the error. Although the area is relatively small (11.26 $km^2$ respectively 1.25 $km^2$) and the relative difference in elevation less than 45 meters, it has been investigated the empirical error pattern to find anomaly or trend within (chapter 5.1).The outline of the Monte Carlo simulation is shown in the **Fig. 1**.

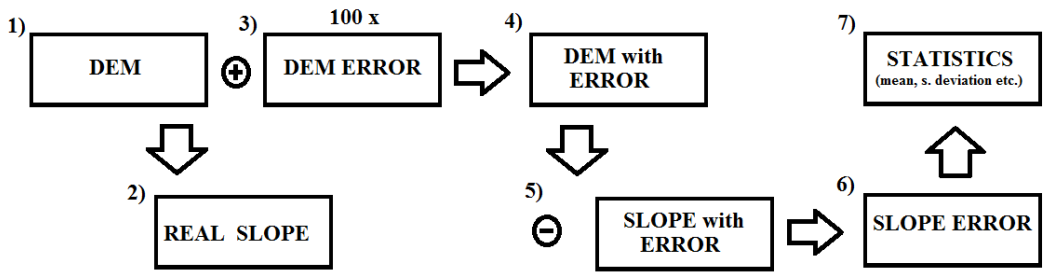



**Fig. 1:** Outline of Monte Carlo simulation, here 1) denotes the input DEM, 2) SLOPE calculated from 1),3) generated DEM ERROR,4) Alternative DEM, 5) Alternative slope, 6) Error in slope, 7) Statistics.

In simulations were used particular initial DEM. This DEM has been considered as error free representation of the true state of the elevation. Next has been calculated the “error free slope”. Then DEM error patterns have been generated according to initial DEM and error model attributes. Initial DEM has been perturbed with the generated random error field. The resulting DEM has the essential properties of both the error pattern and initial raster. Thus have been generated 100 realizations of DEM and subsequently slope estimates derived from alternative DEMs. Set of error patterns in slopes has been calculated as the difference between the error free slope and the particular alternative slope. Using appropriate statistics has been derived the result of the simulation. In some cases has to be used absolute error value instead of the error value.

**Slope Computation Algorithm**

A variety of methods can be used to estimate slope from DEM. Weighted least squares fit of a plane to a 3x3 neighbourhood centred on each point is the most amenable to a mathematical analysis of error propagation [24]. Including the most used GIS software (SW) ArcGIS most of the GIS SW use this method to compute the slope from DEM. In this paper we decided to follow the aforementioned method’s algorithm. The output slope derivate can be calculated in degrees (angular unit **Eq. 8**) or percentage (**Eq. 7**). Degrees are the units chosen in the paper. Slope in degrees is calculated multiplying the slope in radians with 57.29578. Slope calculation (**Fig. 2**) is based on the change of height (rise) in the direction of x and y direction

(run) - mathematically the first partial derivation of z in x and y axes. Thus the slope (**Eq. 5**) is determined by the rate of change (*Beta*) in both horizontal (*HD* **Eq. 3**) and vertical (*VD* **Eq. 4**) direction from the centre cell (*E*).

$$HD = \frac{\partial z}{\partial x} \quad \textbf{(3)}$$

$$VD = \frac{\partial z}{\partial y} \quad \textbf{(4)}$$

The approximation of the partial derivatives was made by third-order finite difference method (**Eq. 5** and **6**) [19]. The method uses the 3x3 neighbourhood (**Fig. 3**) of the elevation values obtained in the raster around the centre cell. The distance between the elevation points is denoted *w*and represents also the cell (pixel) size of raster.

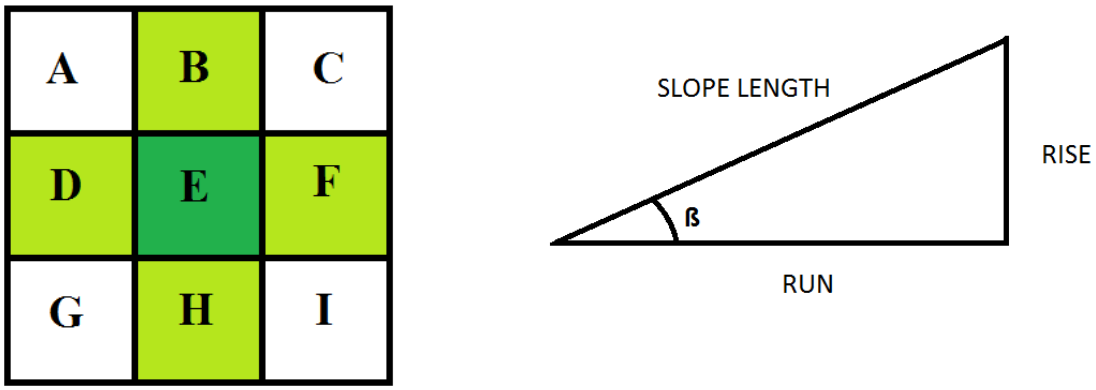



**Fig. 2.** Left the 3x3 neighbourhood window of the centre cell E and right the rise, run and beta description.

$$HD \approx \frac{(C + 2F + I) - (A + 2D + G)}{8 * w} \quad \textbf{(5)}$$

$$VD \approx \frac{(A + 2B + C) - (G + 2H + I)}{8 * w} \quad \textbf{(6)}$$

$$S = \sqrt{HD^2 + VD^2} \quad \textbf{(7)}$$

$$\beta = \arctan\left(\sqrt{HD^2 + VD^2}\right) \quad \textbf{(8)}$$

The influence of data precision on derived slope is highly related to the grid resolution. While using a high-resolution DEM (e.g. 1m grid resolution), the influence of data precision becomes

quite significant. DEM resolution determines the level of details of the surface being described. It naturally influences the accuracy of derived surface parameters. On the other side usually the DEM error caused by data precision level is quite minimal, except in flat areas where the rounding errors could be significant [30]. The precision significance has been investigated also, to prove or reject. We tried to minimize the rounding error, because of flat areas.

## STUDY AREA

Error propagation was carried out along a 5.9 km stretch of the Olše River and a 3.2 km stretch of the Stonávka River. Both river sections are located in the northeast region of the Czech Republic near its border with Poland [31].The area is located south of the city Karviná in the north-eastern part of the Moravian-Silesian Region. The area is 5.544 km in length and 2.281 km in width spaced. After the area affected with gross error has been eliminated, it remained a total area of 11.262 $km^2$. Because of gross errors and uncertainty in data collection process caused by atmosphere, three parts of the area (west) had to be clipped. Due the time-consuming computational method the 1.250 $km^2$ large study are have been used in case of higher precision data input (**Fig. 3**). The elevation of the area is varying between 211 and 256 (respectively 216 to 227 for small area) meters over the sea level. The slope is varying from 0°to 85° (respectively 0 to 67 degrees). The average slope values (1.95° to 3.9° respectively 3° to 3.5°) and the data histograms revealed the flat characteristics of the surface with few steep slopes

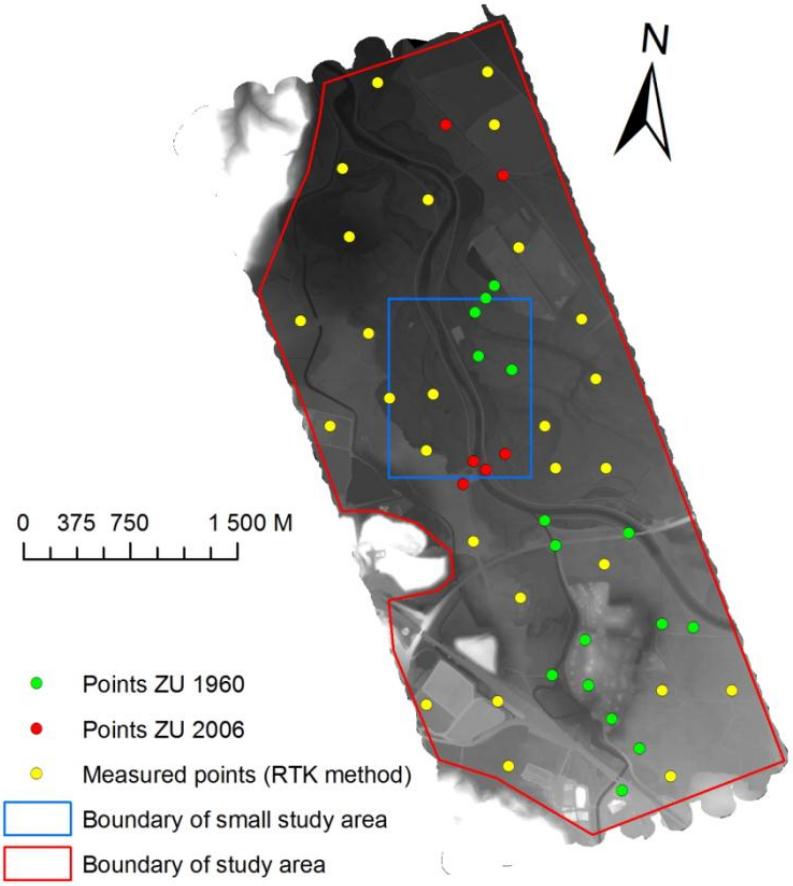



**Fig. 3.** Study area and measurement point locations for RMSE computation.

## DATABASE CREATION

GIS database comes from various sources, each having its own level of uncertainty, depending on the specific technique used to acquire it. [9] The input data used to create the DEM in this study were obtained using the LIDAR method (Light Detection and Ranging). The Swedish company TopEyeAB, working with the MK-II laser system of its own design, carried out flights over the research area. The system consisted of a laser scanner with a50 kHz frequency, Inertial Navigation System (INS) and Global Positioning System (GPS) systems. The optical portion of the scanner deviated the laser beam into circular traces. The system was supplemented by a Rollei digital air camera with a 16-megapixel resolution (4080 x 4076 pixels). Scanning was carried out on the D-Hahn helicopter carrying the MKII-S/N 804 system at an altitude of 300 m. DEMs (0.5, 1, 5 and 10 m. resolution) were computed independently of each other from particular acquired LIDAR data point cloud. The RMSE in input data has been calculated two times for every DEM to make comparison of possible inputs. First have been calculated the error values subtracting the DEM from the DEM with higher precision

(resolution). 0.2 m resolution DEM has been used for 0.5 m resolution DEM. Then the RMSE (0.317 for 10m, 0.156 for 5m, 0.04 for 1m and 0.035 for 0.5 meter resolution) was calculated from the error values of the whole area. This RMSE values have been compared with the result of the second computation, which has been computed from RMSE of 49 point measurements in the study area (**Fig. 3**). 22 of 49 points were created by CUZK (Land Survey Office of Czech Republic) without any given information of data gathering method and accuracy. The second RMSE computation has a higher RMSE, which was effected by the location of the 49 points. They are not representative for the whole area. 49 points were located more in error prone surface (roadsides, river bank sides) as is their proportion of the total area. The 10 m resolution RMSE difference takes 5.7 cm (0.374 for 49 points and 0.317 for LIDAR), what is 17% of the total value of LIDAR RMSE. In other resolutions cases it was even worse (5 m – 14.1 cm, 1 an 0.5 m – 24.9 cm). It is necessary to mention, that the LIDAR DEM of higher accuracy inherent a certain uncertainty too. LIDAR RMSE results have been taken to fit the spatially uncorrelated error pattern as consequence of better representation of the continuous empirical error pattern. The autocorrelated error pattern has been made by investigating the empirical elevation error (*Chapter 5.1*).

**Simulation of Random Fields**

The input error field has been made by investigating the empirical error pattern obtained with aforementioned method (*Chapter 2*). Error propagation has been modelled with and without spatially autocorrelated error field. The real state of nature is other than the expected theoretical state. First, there is an unjustified assumption that the mean error is zero [32]. The error mean statistics were close to zero, but all of them were rejected as statistical zeroes using t-test hypothesis test in SW Statgraphics (**Tab. 1**).

**Tab. 1** DEM error statistics (Number of Elevation Points, Error Mean [meters], Standard Deviation of Error [meters], Maximum Absolute Error [meters]).

| DEM resolution | NUM. POINTS | MEAN | STD. DEVIATION | MAX ABS ERROR |
|---|---|---|---|---|
| 10 x 10 | 263 520 | $-3.2\ 10^{-2}$ | 0.692 | 11.942 |
| 5 x 5 | 1 051 997 | $-1.2\ 10^{-3}$ | 0.362 | 12.053 |
| 1 x 1 | 26 289 516 | $-2.3\ 10^{-3}$ | 0.085 | 9.567 |
| 0.5 x 0.5 | 83 963 724 | $1.0\ 10^{-5}$ | 0.008 | 1.597 |

Thus the best fit is to follow the empirical model [23]. If the difference between the elevation in the DEM and the actual surface (which equals the error surface) is done, the error surface should have a large positive autocorrelation [33]. It is assumed that the RMSE over the study area is constant or spatially autocorrelated, what was confuted in previous researches (Fisher, Oksanen etc.). Although the total area is 11.262 $km^2$ small and according to the terrain surface and aforementioned research results (RMSE should be constant), it was necessary to divide it into smaller subareas, where this statement was proved. There have not been find a significant difference in parameters (range, partial sill and nugget). The area has been

searched for trends. But none have been found. The best fitted model was the Stable one. According to previous researches it has been chosen Exponential and Gaussian to fit the pattern. Gaussian and Spherical had almost the same results, but the Gaussian better fitted the closest averaged values, so it has been chosen (Tab. 2, Fig. 4, 5, 6, 7). The appropriate shape of the spatial autocorrelation model was not critical as the computed autocorrelation parameters.

**Tab. 2** Gaussian error model parameters.

| DEM resolution | Lag Size [m] | Num. of Lags | Nugget [m] | Partial Sill [m] | Range [m] |
|---|---|---|---|---|---|
| 10 x 10 | 6.480 | 12 | 0.246 | 0.165 | 49.025 |
| 5 x 5 | 5.054 | 12 | 0.068 | 0.042 | 31.090 |
| 1 x 1 | 3.610 | 12 | 0.001 | 0.005 | 13.752 |
| 0.5 x 0.5 | 0.544 | 12 | $3.2\ 10^{-5}$ | $2.7\ 10^{-5}$ | 3.897 |

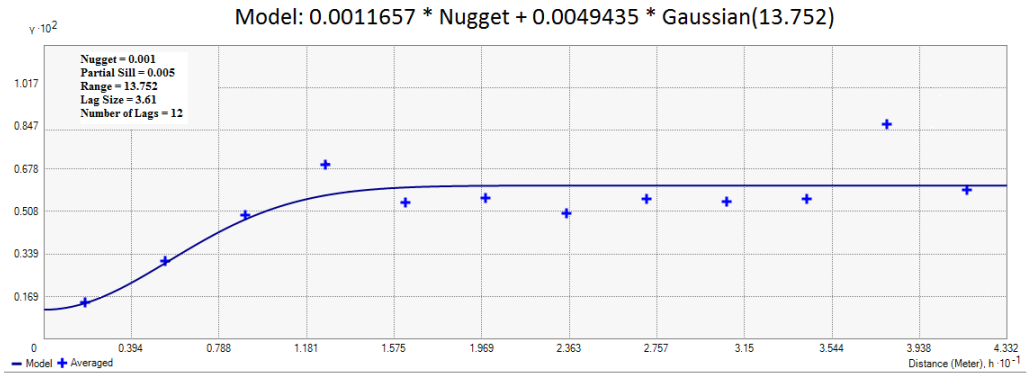



**Fig. 4.** Gaussian error model for 1x1 m resolution DEM.

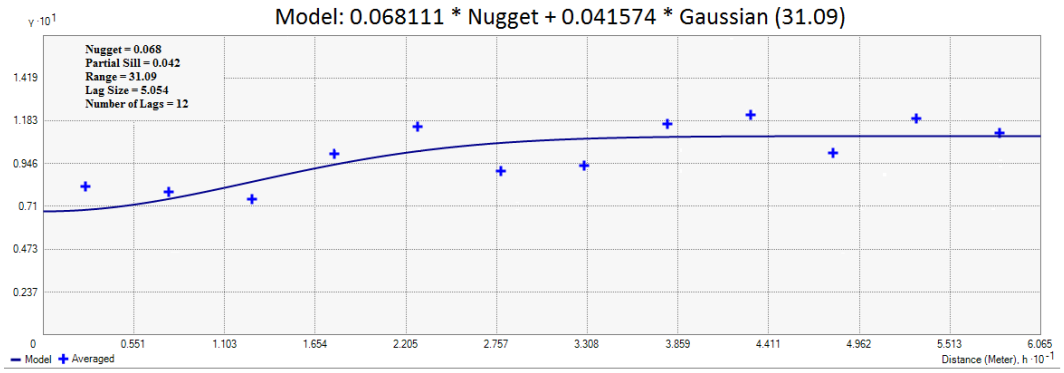



**Fig. 5.** Gaussian error model for 5x5 m resolution DEM.

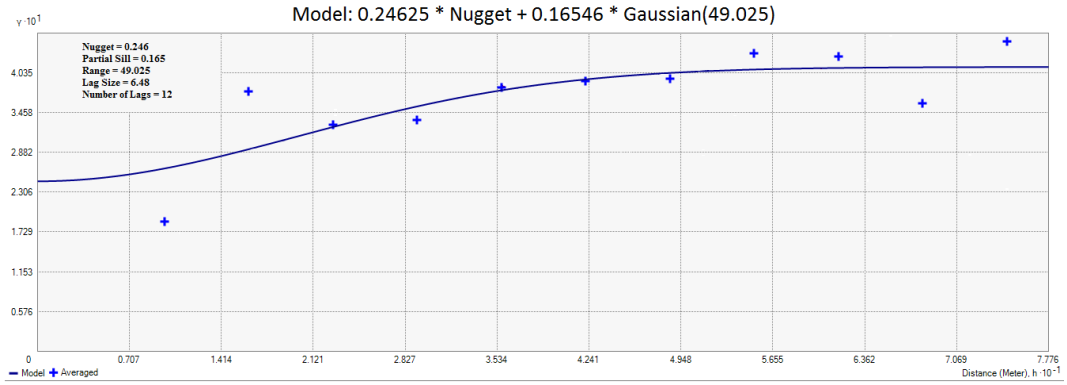



**Fig. 6.** Gaussian error model for 10x10 m resolution DEM.

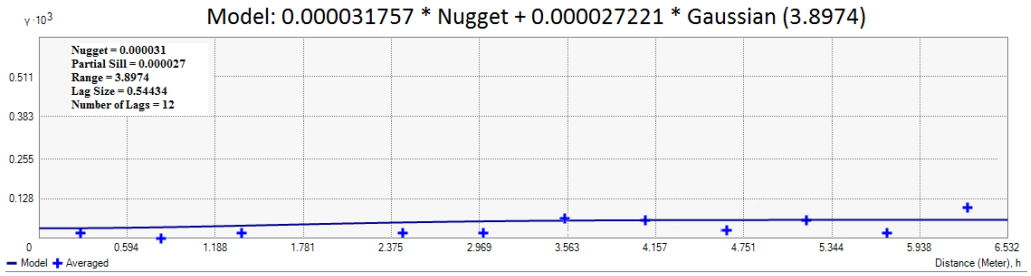



**Fig. 7.** Gaussian error model for 0.5x0.5 m resolution DEM.

The theoretical Gaussian models have been used to model the fields, **Fig. 8** depicts the difference between the spatially correlated and uncorrelated random field (10 m DEM). The error fields have been modelled 100 times for each DEM to perform Monte Carlo simulation. The outputs of aforementioned stochastic error propagation (**Fig. 1**) are mentioned in the following chapter results. The theoretical 0.5 m resolution Gaussian error model (**Fig. 7**) opens a question about the threshold, when is reasonable use spatially autocorrelated model and when just white noise.

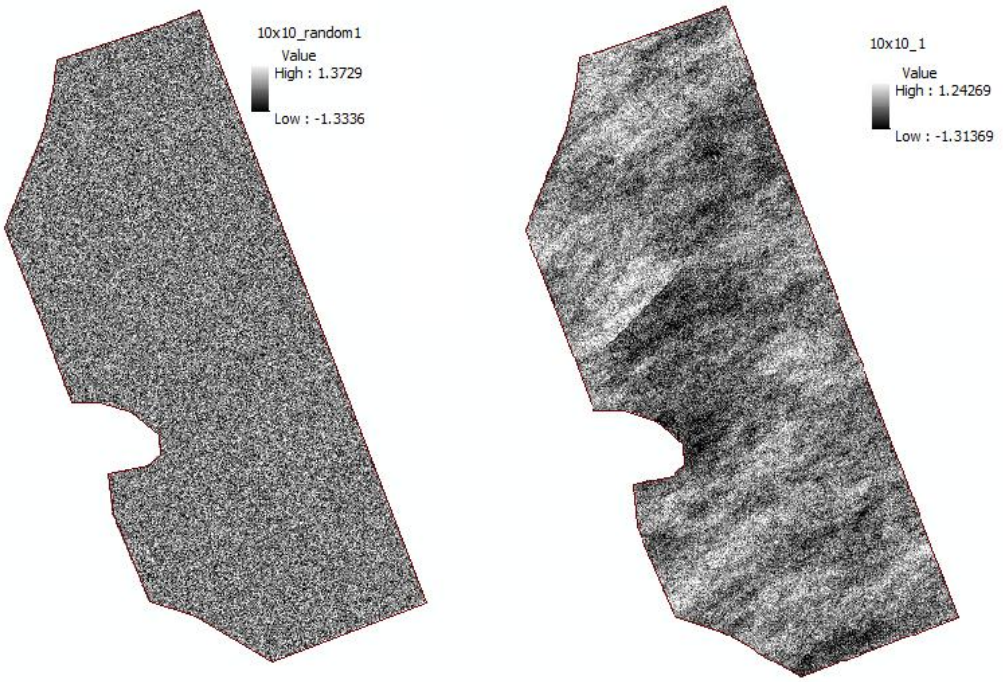



**Fig. 8.** Left uncorrelated white noise and right spatially correlated random error field of 10 m DEM.

## RESULTS AND CONCLUSIONS

The error propagation results are summarized in table 3. For example in case of 10x10 m DEM error input we expect 0.680° (respectively 0.657° without spatial autocorrelation) error (mean of the means in column 5). For 5x5 m it is 0.657° (0.635°), 1x1 m 0.796° (0.776°) and for 0.5x0.5 m 0.272° (0.395°). The results are represented in absolute values. The behaviour of the error, that the value $x$ and its opposite value $-x$ represent the same deviation from the real state of nature, made possible this representation. It is more natural to see the errors in positive values and it enables better interpretations. If one number is most representative of the error evaluation, then it is the mean. It has been randomly selected 50 spot samples to prove the insignificant difference between the slope error result derived from inputs with and without autocorrelation.

**Tab. 3**: Error propagation results [m or °] (Particular DEM, Input DEM error std. deviation for all elevation values, DEM RMSE, Output Slope absolute error statistics according to cells – mean, min, max, std. dev.).

| DEM | Auto- | Error in DEM | | Output Slope absolute error [degrees] | | | |
|---|---|---|---|---|---|---|---|
| resolution | Correl | s. d. | RMSE | mean | Min | Max | std. dev. |
| 10 x 10 | Yes | 0.692 | 0.317 | 0.449-1.302 | 2.98 $10^{-8}$-0.392 | 0.979-4.312 | 0.373-0.881 |
| 10 x 10 | No | 0.692 | 0.317 | 0.379-1.296 | 0-0.273 | 1.169-4.925 | 0.283-0.829 |
| 5 x 5 | Yes | 0.362 | 0.156 | 0.296-1.255 | 0-0.366 | 0.580-4.473 | 0.11-0.882 |
| 5 x 5 | No | 0.362 | 0.156 | 0.216-1.286 | 0-0.313 | 0.631-4.543 | 0.167-0.831 |
| 1 x 1 | Yes | 0.085 | 0.04 | 0.276-3.017 | 0-0.469 | 0.805-13.205 | 0.202-2.276 |
| 1 x 1 | No | 0.085 | 0.04 | 0.286-1.441 | 0-0.361 | 0.919-5.747 | 0.210-0.871 |
| 0.5 x 0.5 | Yes | 0.008 | 0.035 | 0.043-0.510 | 0-0.134 | 0.126-2.438 | 0.029-0.385 |
| 0.5 x 0.5 | No | 0.008 | 0.035 | 0.027-2.587 | 0-0.164 | 0.120-10.651 | 0.073-1.557 |

Every spot sample has 100 alternative values, which have been used to compute mean and standard deviation. Two sample F test (st. deviation) and two sample t rest (mean) have been used. Null hypothesis set to: There is no difference in standard deviation (respectively means) and alternative hypothesis to: There is statistically significant difference between the std. deviations (means). For example for 1x1m resolution; we discovered that 49 in 50 cases for mean, respectively 47 in 50 for std. deviation are not significantly different (**Tab 4** showing 5 examples). Errors without spatial autocorrelation do not result to greater variance in resulting slope error (Oksanen got same results). Therefore it should be challenged, if error propagation without spatial autocorrelation represents sufficiently the true state of the nature of the error representation. In else we proved, that DEM error input without autocorrelation does not result (few exceptions) to greater error estimate of slope. Critical is the 0.5x0.5 resolution DEM error input, which leads to more inequalities. This phenomenon should be further investigated to understand the reason.

**Tab. 4**: Two sample F-test respectively t-test for 5 spots. Hypothesis ($H_0$) *(sigma1/sigma2 = 1.0)* concerning the ratio of the standard deviations of one spot sample of 100 observations for F-test, and hypothesis concerning the difference between the means (mean1-mean2 = 0.0, sigma1 and 2 input needed too) for t-test. (both 95.0% confidence level, P-value 0.05 and less rejects $H_0$).

| | | | | **Hypothesis test:** | | |
|---|---|---|---|---|---|---|
| **Random Sample** | **Slope** | **Autocor. Value** | **White noise Value** | **F(t) statistics** | **P-v.** | **Null Hypothesis** |
| 1 std. deviation | 17.536° | 0.322 | 0.314 | (F) 1.052 | 0.803 | Do not reject, ratio = 1 |
| 2 std. deviation | 11.232° | 0.396 | 0.400 | (F) 1.051 | 0.803 | Do not reject, ratio = 1 |
| 3 std. deviation | 5.950° | 0.407 | 0.416 | (F) 0.957 | 0.828 | Do not reject, ratio = 1 |
| 4 std. deviation | 0.137° | 0.442 | 0.392 | (F) 1.271 | 0.234 | Do not reject, ratio = 1 |
| 5 std. deviation | 0.226° | 0.502 | 0.322 | (F) 2.435 | $1.10^{-5}$ | Do reject, ratio <> 1 |
| 1 mean | 17.536° | 0.798 | 0.795 | (t) 0.067 | 0.947 | Do not reject, difference = 0 |
| 2 mean | 11.232° | 0.787 | 0.778 | (t) 0.200 | 0.841 | Do not reject, difference = 0 |
| 3 mean | 5.950° | 0.769 | 0.817 | (t) -0.825 | 0.410 | Do not reject, difference = 0 |
| 4 mean | 0.137° | 1.117 | 1.155 | (t) -0.643 | 0.521 | Do not reject, difference = 0 |
| 5 mean | 0.226° | 1.008 | 0.894 | (t) 1.911 | 0.057 | Do not reject, difference = 0 |

Although the result of input error without autocorrelation did not showed greater aberration, it is not suitable for elevation error pattern modelling. In fine topscale and microscale [1] scale has the error pattern large positive autocorrelation. Furthermore in our case the outliers are responsible for rejection of Gaussian distribution. The outliners have to be also incorporated to the error model, what has not been done due to the lack of time. The average variance and mean of the errors in slopes is not strictly increasing with steepness of the slope (e.g. **Fig.9**). This causality should be further investigated; one of the reasons is the insufficient number of samples with steeper slope. The prevailing spatial distribution of slopes in study are is partially captured also in mean slope error (**Fig. 10**, **Fig. 11**). Input based on empirical elevation error (AC) describes better the error pattern and leads to more realistic and accurate spatial distribution of slope errors according to slope in study area. White noise (WN) input error field is closest to AC in minimum slope error distribution. Linear planar surfaces (roads etc.) are inadequately propagated. Planar surface is the most error prone type. According to similar studies (Fisher, Goodchild etc.) using different DEM input data, high quality LIDAR input data decreases the output uncertainty. In our case, the autocorrelated model fitted the error surface with exception of its outliers. There is a need to find a way how to include them. Extreme values are higher in case of theoretical model with autocorrelation; random number

generator produces smaller extreme values also. Autocorrelation also expands the std. deviation of extreme values. On the one side the extreme elevation error values were found to be clustered around the steepest slopes, on the other side the steeper slopes has smaller slope error result with same elevation error input. Range of fitted empirical error model (49.6 for 10x10, 31.1 for 5x5, 13.8 for 1x1 and 3.9 for 0.5x0.5) was decreasing with higher resolution. We do assume that there should be a specific resolution limit value, where range is close to 0. Geostatistical modelling is very time consuming. We had to decrease the extent for the 0.5x0.5 and 1x1 meter resolution inputs. To compute one 1x1 meter DEM resolution error pattern (21 983 304 values in 5964 rows and 3686 columns) took 12 days and 17 hours (using 30 GB RAM and 4 processors Intel(R) Core(TM2)2 Quad CPU Q9300, 2.5 GHz). This computation requires super-computer.

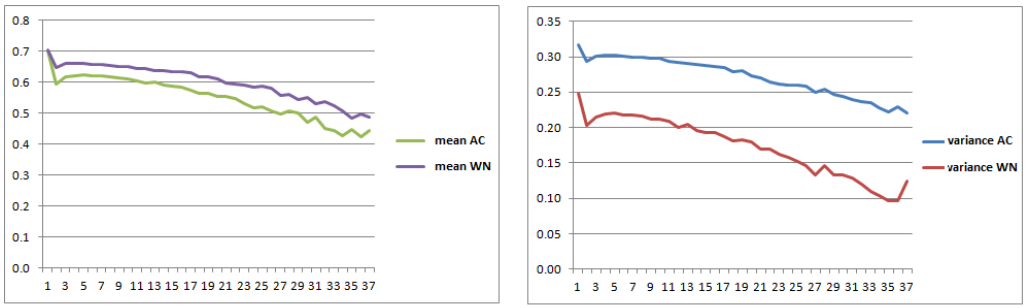



**Fig. 9**: Slope Error dependent variable (vertical axis) vs. Slope independent variable (horizontal axis) (WN randomly generated white noise, AC autocorrelation input according to empirical error pattern) showing the decrease in slope error with increasing slope.

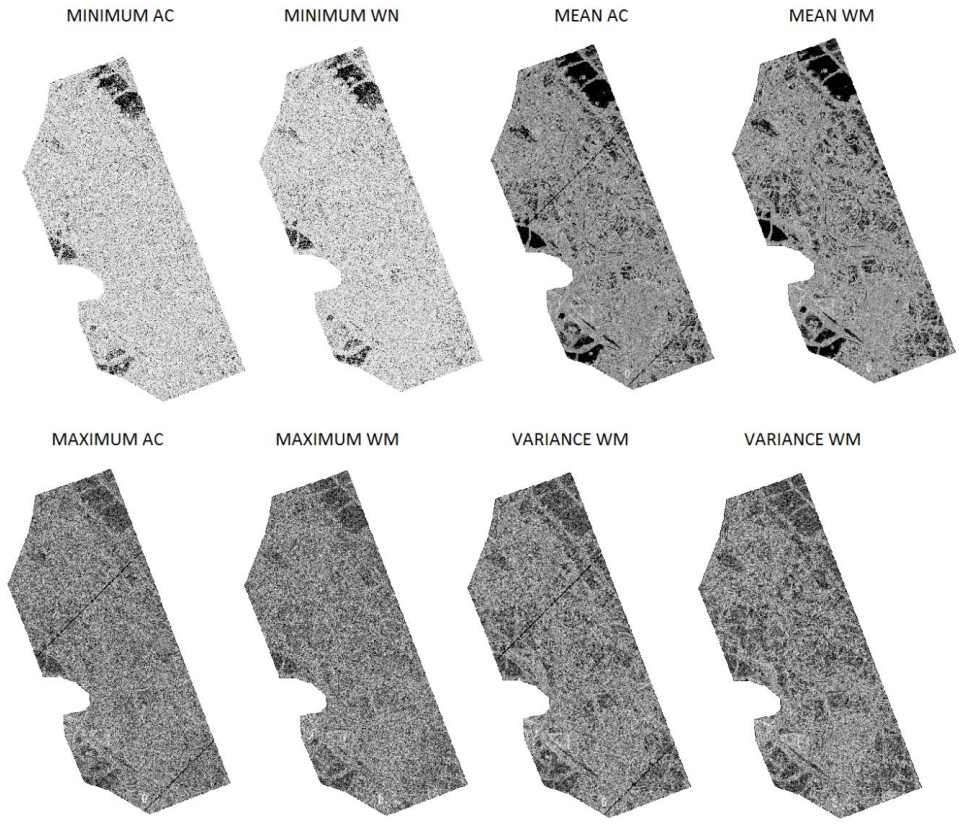



**Fig.10**: Mean, Variance, Minimum and Maximum statistics for 10x10m slope errors; if darker the colour then higher the slope error value and more planar the surface.

## DISCUSSION

Although a lot of research has been made in the uncertainty and error propagation field over the last decades, still many questions left unanswered. In this study we focused to clear antagonistic results provided by Oksanen and Fisher. Oksanen declared that slope errors modelled without autocorrelation do not get worse result. In else, the slope derivate has not maximum variation with spatially uncorellated random error. On the other side Fisher declared, that the slope derivate computed from uncorrelated random error is worse, because of the poor input elevation error representation. We have found that Oksanen is right. Fisher is correct about the poor representation and the research area should be always investigated before analysed. We were not able to completely ascertain the character of the pattern error. Definitely it has been found the underlying error pattern. Some irregular outliners appeared which have to be incorporated. The next step should be investigation of the outliners. The empirical error model and the modelled error model have to be subtracted and the product investigated (external data may help too – underlying geology, terrain roughness, land use etc.). The resulting pattern is an addendum to the underlying error pattern. There can be more functions describing local shapes of error pattern. Sum of all functions (patterns) gives the

resulting error pattern. We have found that there should be a threshold value, which in case of high precision and resolution data do not require the usage of autocorrelation in error surface (in case of high precision LIDAR data input and relatively small area).

It is true, that any given input data is carrying error value significant enough to change the resulting slope – even the high precision micro-scale LIDAR DEM. Results obtained with DEM inputs of same resolution and acquired with other methods (photogrammetric) could be used for better comparison and calculation of exact LIDAR improvement in slope error estimation. Other software tools should be used to prove simulated reality with gstat. According to time demanding computational process, it should be less consuming processes investigated for error pattern simulation, e.g. fuzzy approach. Software development and new supercomputers could be another solution. There is still a doubt, pros and cons, if unconditional Gaussian or sequential Gaussian simulation has to be used, how to model non-stationary error field in larger areas and what it is dependent on?

It is necessary to remember the main reason of dealing the uncertainty: decreasing the risk that the outcome will be incorrect and will lead to wrong decisions. This study has been made as error propagation background to inundation area delineation with GLUE method in the area. The processing of airborne hyperspectral data introduces uncertainty, which is enough to change the product. To know the uncertainty in the result is important in crisis management and other fields. Sometimes even one degree in slope can change the situation and flooded area.

## ACKNOWLEDGEMENT

This paper has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme 'Research and Development for Innovations' funded by Structural Funds of the European Union and state budget of the Czech Republic.

## REFERENCES

[1] Oksanen, J. and Sarjakoski, T. (2005) Error propagation of DEM-based surface derivates in Computers & Geoscience, Vol. 31, 1015-1027.

[2] Trolegart, K., Ostman, A. and Lindgren, R. (1986) A comparative test of photogrammetrically sampled digital elevation models in Photogrammetria, Vol. 41, 1 -16.

[3] Hutchinson, M. F. and Gallant, J. C. (2000) Digital elevation models and representation of terrain shape in Wilson, J. P., Gallant, J. C. (eds.) Terrain Analysis: Principles and Applications, Willey, New York, NY, 29-50, 978-0471321880.

[4] Erdogan, S. (2010) Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study in Computers & Geoscience, Vol. 36, 34 - 43.

[5] Heuvelink, G. B. M. (1998) Error Propagation in Environmental Modelling with GIS. Taylor & Francis, London.

[6] Lee, J., Snyder, P. K. and Fisher, P. F. (1992) Modeling the effect of data errors on features extraction from digital elevation models in Photogrammetric engineering and Remote Sensing, Vol. 58, No. 10, 1461–1467

[7] Shi, W., Fisher, P.F. and Goodchild, M.F. (eds.) (2002) Spatial Data Quality. Taylor & Francis, London. ISBN-13: 978-0415258357.

[8] Heuvelink, G. B. M. (2007) Error-Aware GIS at Work: Real-World Applications of the Data Uncertainty Engine in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34.

[9] Hwang, D., Karimi, H. A. and Byun, D. W. (1998) Uncertainty analysis of environmental models within GIS environments in Computers & Geosciences, Vol. 24, No. 2, 119-130.

[10] Burrough, P. A. (1993) Principles of Geographic Information Systems for Land Resources Assessment. Clarendon Press, Oxford.

[11] Openshaw, S., Charlton, M. and Carver, S. (1991) Error propagation: A monte Carlo simulation in Masser I., Blakemore, M. (eds.) Handling Geographical Information: Methodology and Potential Applications, Longman Scientific and Technical, London, UK, 78-101, ISBN: 978-0470217924.

[12] Goodchild, M. F., Shortridge, A. M. and Fohl, P. (2000) Encapsulating simulation models with geospatial data sets in Lowell K., Jaton, A. (eds.) Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources, Ann Arbor Press, Chelsea, MI, 123-130, ISBN: 978-1575041193.

[13] Heuvelink, G. B. M. (2003) Analysing uncertainty propagation in GIS: Why is it not so simple? in Foody G. M., Atkinson P. M. (eds.) Uncertainty in Remote Sensing and GIS, John Wiley and Sons, Chichester, UK, 155-165, ISBN 978-0470844083.

[14] Heuvelink, G. B. M. and Brown, J. D. (2007) Uncertain Environmental Variables in GIS in Shashi, S., Hui, X. (eds.) Encyclopedia of GIS, Springer, 1184-1189, ISBN 978-0387308586.

[15] Hodgson, M. E., Jensen, J., Raber, G., Tullis, J., Davis, B. A., Thompson, G. and Schuckman, K. (2005) An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions, Photogrammetric engineering and Remote Sensing, 71 (7), 817-823.

[16] Barber, C. P. and Shortage, A. (2005) Lidar elevation data for surface hydrologic modeling: Resolution and representation issues. Cartography and Geographic Information Science, 32 (4), 401-410.

[17] Vaze, J. and Teng, J. (2007) High resolution LIDAR DEM – How good is it? in Oxley, L., Kulasiri, D. (eds.) MODSIM 2007 international congress on modelling and simulation, 692-698.

[18] Goovaerts, P. (1997) Geostatistics for natural resources evaluation. New York: Oxford University Press, pp. 483, ISBN 978-0195115383.

[19] Oksanen, J. and Sarjakoski, T. (2010) Non-stationary modelling and simulation of LIDAR DEM uncertainty in Accuracy 2010 Symposium, July 20 – 23, Leicester, UK.

[20] Goodchild, M. F. (2007) Imprecision and Spatial Uncertainty in Shashi, S., Hui, X. (eds.) Encyclopedia of GIS, Springer, 480-483, ISBN 978-0387308586.

[21] Fisher, P. F. (1999) Models of uncertainty in spatial data in Longley, P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (eds.) Geographical Information Systems: Principles, Techniques, Management, and Applications, John Wiley nad Sons, New York, 191-205, ISBN: 978-0471735458.

[22] Gong, J., Zhilin, L., Zhu, Q., Sui, H.G. and Zhou, Y., (2000) Effect of various factors on the accuracy of DEMs: an intensive experimental investigation in Photogrammetric Engineering and Remote Sensing, Vol. 66, No. 9, 1113–1117.

[23] Fisher, P. F. (1998) Improved Modeling of Elevation Error with Geostatistics in Geoinformatica, Vol. 2:3, 215-233.

[24] Hunter, G. J. and Goodchild, M. F. (1997) Modelling the uncertainty of slope and aspect estimates derived from spatial database in Geographical Analyses, Vol. 29, No. 1, 35-49.

[25] Caers, J. (2011) Modeling Uncertainty in the Earth Science, John Wiley & Sons, Oxford, UK, ISBN 978-1119992622.

[26] Eastman, J. R., Kyem, P. A. K., Toledano, J. and Jin, W. (1993) GIS and decision making, Explorations in Geographical Information Systems Technology, Vol. 4, Worcester, MA, Unitar, Clark Labs.

[27] Fisher, P. F. and Tate, N. J. (2006) Causes and consequences of error in digital elevation models, Progress in Physical Geography, Vol. 30, No. 4, pp. 467-489.

[28] Beekhuizen, J., Heuvelink, G. B.M., Reusen, I. and Biesemans, J. (2009) Uncertainty Propagation Analysis of the Airborne HyperspectralData PROCESSING CHAIN in Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009, Whispers, ISBN 978-1-4244-4686-5.

[29] Skidmore, A. K. (1989) A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model in International Journal of Geographical Information Systems, Vol. 3, 309 – 318.

[30] Zhou, Q. and Liu, X., (2004) Analysis of errors of derived slope and aspect related to DEM data properties. Computers & Geosciences 30, 369–378.

[31] Podhoranyi, M., Unucka, J., Bobal, P. and Rihova, V. (2011) Effects of Lidar DEM Resolution in Hydrodynamic Modelling: Model Sensitivity for Cross-Sections. in International Journal of Digital Earth. ISSN: 1753-8947

[32] Li, Z. (1988) On the measure of digital terrain model accuracy in Photogrammetric Record, Vol. 12, 873-877.

[33] Goodchild, M. F. (1995) Attribute accuracy in Guptill, S. C., Morrison, J. (eds.) Elements of Spatial Data Quality, Oxford, Pergamon, 59-79, ISBN:978-0080424323.

# COMPARING VULNERABILITY DELINEATIVE OF AQUIFER USING DRASTIC AND FUZZY LOGIC METHODS (CASE STUDY: GULGIR PLAIN OF MASJED SOLIEMAN, IRAN)

Hamid Reza NASSERY[1] and Ferdows SAREMI NEJAD[2]

[1]Associate Prof. Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

*h-nassery@sbu.ac.ir*

[2]M.Sc. Student in Hydrogeology, Islamic Azad University, Science and Research, Tehran, Iran

*Saremi1603@Yahoo.com*

**Abstract**

Various vulnerability evaluation systems are developed for proper and effective prevention of future groundwater pollutions. Presently, there are different methods to detect the contamination potentiality of the ground water. The present research evaluates the vulnerability delineative of Gulgir Plain aquifer. This plain is located on the east of Khuzestan Province. Due to its rich groundwater resources and fertile soil the plain enjoys a flourishing agriculture. However, due to massive agricultural activities the quality of the groundwater has been exposed to agricultural pollutions particularly Nitrates. The DRASTIC model and fuzzy Inference were chosen from the myriad of vulnerability evaluation methods. Principles of the DRASTIC model are based on the combination of hydrological and hydro-geological indexes that affect the transmission or non-transmission of contaminations and Boolean Logic is used to calculate the indexes. The Boolean logic may result in erroneous conclusions about the quantities that proximate the borders. The fuzzy logic, however can improve the accuracy of results particularly concerning the border quantities. Therefore, this method has been used in the present research and the results are compared to those of the Boolean logic. The results confirm the power of fuzzy modeling in determining hydrogeological parameters which lack assurance.

**Keywords: vulnerability, delineative, DRASTIC, fuzzy logic**

## INTRODUCTION

Ground water is considered as important resource due to the lower potentiality of contamination as well as a high reserve capacity compared to the surface water [1]. The major sources of spot and dispersive contaminants arising from human activities on the ground and penetration of these contaminants into the ground tend to reduce the quality of the ground water. Therefore, preventing the ground water from contamination is essential to the management of ground water resources [2]. Evaluation of vulnerability is a proper and low cost method for detection of the areas which are apt to contamination. The concept of vulnerability was proposed in France to give information on contamination of ground water for the first time at the end of 1960 [3]. Vulnerability may be defined as the possibility of penetration and dispersion of contaminants from the earth surface into the ground water system. Vulnerability is considered as an inherent apt of the ground water which depends on

the extent of sensitivity of this system to the impacts of the human or natural activities. Vulnerability is divided into two inherent and special categories in term of meaning. The inherent vulnerability means sensitivity of ground water to the natural elements while special vulnerability studies the inherent vulnerability together with the probability of exposure of the ground waters to penetration of contaminants.

One of the methods that greatly help the management of ground waters is to prepare maps in which the areas vulnerable or sensitive to contamination have been located. Moreover, an index is also defined to identify those areas that are more exposed to contaminants than others. In this way we can compare different areas with each other and find a single criterion for evaluation.

The maps and data should be combined and classified in order to Boolean the results and calculate the index of vulnerability. These classifications are usually based on the Boolean Method. In this classification the border between the classes should be specified. Whereas the vulnerability indexes are of spectral and extensive nature, the classification on the basis of the Bolin method causes an area to be relocated from one class to an upper class or a lower class with a slight change, something which is not principally acceptable and justifiable. However, by using the foundations of the fuzzy theory in which each and every subject and concept has a membership grade a suitable method may be offered for classification and rating as compared with the Boolean method [4]. The present analysis is aimed at identification of the areas in the subterranean water of Gulgir Plain that are apt to contamination, and examining the possibility of applying the fuzzy logic for classification of the vulnerability maps.

**Situation and General Features of the Study Plain**

The Gulgir plain is situated in Masjed Soleiman region within the geographical longitude 49° 27′ 30″and 49° 31′ 19″ and the geographical latitude 31° 44′ 05″ and 31° 47′ 01″ with an area of about 1600 hectares. It is located in the middle of Karoun basin (Fig. 1).

The area of the alluvial subterranean water of Gulgir Plain is approximately 24.48 square kilometers. Gulgir, Tambi, Bahramabad, Hajiabad, Shamsabad, and Sazbabad villages are among the major population center in Gulgir Plain. Access to the plain is possible through Ahwaz-Izeh-Masjed Soleiman-Haftgel main roads. Regional streams include Shour Tambi, Ab Gulgir, Ab Lashgar, Shour Barik, Bid Zard and Darreh Kouh la.

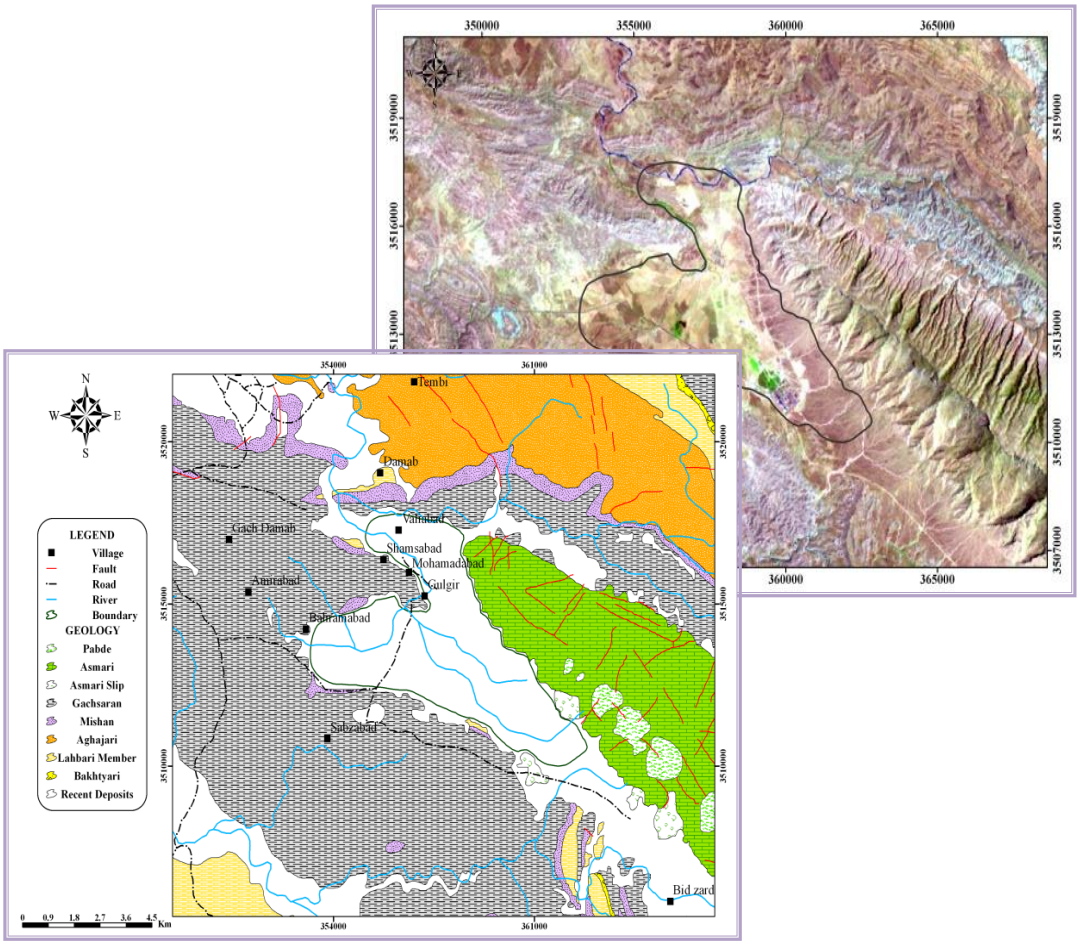



**Fig. 1**. Geological map and satellite image of Gulgir Plain

## STUDY METHOD

The DRASTIC method and fuzzy model have used in order to calculate the extent of vulnerability of the ground water of Gulgir Plain. The extent has been calculated in the GIS domain. The research has used various information sources including topographic maps with a scale of 1:25000 (by Army Geographical Organization), geological map with a scale of 1:100000 (by state Geological and Mineral Explorations Organization), report of the pedological studies of Gulgir Masjed Soleiman (state Water and Soil Engineering Service Company-Management of Karkheh Studies), data on the level of ground water, results of the pumping tests, excavation log of the observation-exploration wells and maps of geo-physical studies. The map information (like the topographic map) was described in numbers and put into the geographical information system while table information (such as depth to water table) was changed into the data bank format and loaded into the data base. The fuzzy model was employed in this research using MATLAB 7.6.0 software based on the Mamdani minimum - maximum fuzzy Inference, which has the widest application in scientific problems due to its

simple and effective structure, and defuzzifier making Center of gravity method. The Gaussian membership function was used for fuzzifier making due to its simplicity.

**DRASTIC Model**

There are different methods for estimation of the vulnerability of ground water. One of these methods is overlay – index method which combines the elements controlling the movement of the contaminants from the ground surface to the saturated zone and shows the result as a vulnerability index in different spots of the region under the study. One of the models that are widely used to evaluate the vulnerability of the ground water regarding a wide range of potential contaminants is DRASTIC model. The model is the most common overlay – index method which has been developed by the United States Environment Protection Agency (USEPA) and American Water Well Association (AWWA) to determine the contamination potentiality of the ground water. By collecting the key factors affecting the transfer of soluble materials, the model evaluates the contamination potentiality of a region. DRASTIC model calculates the potential contamination of the ground water using seven factors affecting them. These factors are: Depth to water table, Net Recharge, Aquifer media, soil media, Topography, Impact of vadose zone media, and Hydraulic conductivity [5]. These seven parameters do not have equal impacts. As for the topic of vulnerability some of these parameters are more important than others. Consequently, each factor is given a relative weight between 1 and 5 based on its relative importance compared to other factors. Among the cited depth to water table and vadose zone media have the most value and Topography slop has the lowest value parameters. Each parameter is also divided into spans with different effect on contamination potentiality. Each span is given a rate between 1 and 10 (10 for highest effect and 1 for lowest effect). The result of the DRASTIC model is a numerical index which is deducted from the rates and weights allocated to the parameters of the model. To calculate the DRASTIC index equation 1 is used:

$$\text{DRASTIC Index} = D_R D_W + R_R R_W + A_R A_W + S_R S_W + T_R T_W + I_R I_W + C_R C_W \quad (1)$$

Where weight **(W)** and rate **(R)** relate to each of the model's parameters. Upon calculation of the DRASTIC index the vulnerable areas of the ground water are located. We should notice that this index would offer only a relative evaluation and would only distinguish the seriously vulnerable areas from those with lower vulnerability without the capability of an absolute assessment.

**Preparation of Standard Maps of the Model**

A standard map shows the extent of relativity of a feature to the target under consideration. These maps are prepared with geographical data input, storage, process and analysis of the output. In other words, standard maps can be drawn as the output of the data processing based on GIS [6].

**- Depth to water table:** The distance between the earth surface and level of the ground water determines the depth of impact, i.e. it determines the thickness of the non-saturated section. The higher the depth to water table is, the more time it takes the materials to reach the ground water with the possibility of distribution, dilution and absorption of contaminants by soil. Data of the water level in piezometers of the Plain were used to prepare the map of the depth to

water table. To this end, a table was primarily prepared of the water level information including the name of piezometers, situation of piezometers based on UTM and depth of the water in the statistical period between 2007-08 and 2008-09 in Excel domain. The information was then changed into a format acceptable for (*.xls) Arc GIS software. Internal detection through IDW was then used as a suitable method to change the mentioned spot data into the level. In this way the coaxial map of the depth to water table was prepared to be combined with other classes and was classified and rated based on the table 1. The rating map of the depth to water table which has been used in DRASTIC model of Gulgir Plain is shown in fig. 2-a.

**- Net Recharge**: The recharge of the ground water causes the contaminants to transfer vertically, reach the water table, and move along the aquifer horizontally. The DRASTIC model presumes that the main recharge sources are rain and snow falls. In order to examine the volume of the ground water reserve of Gulgir Plain, the overlap method of the map of balance changes of water table and map of distribution of reserve capacity of the ground water were used. In the method of water balance fluctuations of the ground water a direct estimation of the recharge is resulted by multiplication of the amount of the rise of stagnant water level by a special yield [7]. The method in based on the hypothesis that the rise of the balance of water level in unbounded subterranean water is the result of the ground water recharge. The amount of rise of the ground water due to this method is proportionate to the recharge through the soil surface. To this end, the unit hydrograph of Gulgir Plain drawn by specifying the wet season (maximum water table balance) and dry season (minimum water table balance) and the co-balance map of the water table was prepared for the wet and dry seasons. The data resulting from the model of ground water current of the Plain [8] was used to prepare the argillous layer of reserve capacity. Then the amount of recharge was calculated according to the relation 2 [9] by multiplication of map resulting from subtraction of minimum balance from maximum balance by the reserve capacity map and rated based on the rating table 1(Fig. 2-b).

$$R = S_y \, dh / dt = S_y \, \Delta h / \Delta t \qquad (2)$$

Where **R** is the rate of recharge, **Sy** is special yield, **h** is the water balance height and **t** stands for time.

**- Aquifer media:** The layer of the aquifer media is a descriptive layer which depends on the materials forming this aquifer. The aquifer media and the materials forming it determine the process of movement in the current system of the ground water. In order to prepare the layer of the ground water of Gulgir Plain the log of the observation-exploration wells was used. To prepare the layer of the aquifer media on the basis of the ratio of the forming materials a numerical value between 1 and 10 was allocated to each well (form the water table to the bed rock) based on the table 1. The rated map of the aquifer media was then prepared (Fig.2-c).

**- Soil media:** This section includes the upper part of the non-saturated area which extends up to the level of penetration of the roots of plants or activity of microorganism. The soil texture is used as a factor affecting the contamination potentiality in the DRASTIC model. The soil map of the region has been prepared based on the report of semi-detailed pedological studies of the lands of Gulgir region of Masjed Soleiman, plan of supply and transfer of water and

pressured irrigation [10]. On this basis, upon determination of the sampling spots and identification of various types of soil in the region, the soil layer of the region was taken and stored as an argillous layer and was eventually rated based on the table 1 (Fig. 2-d).

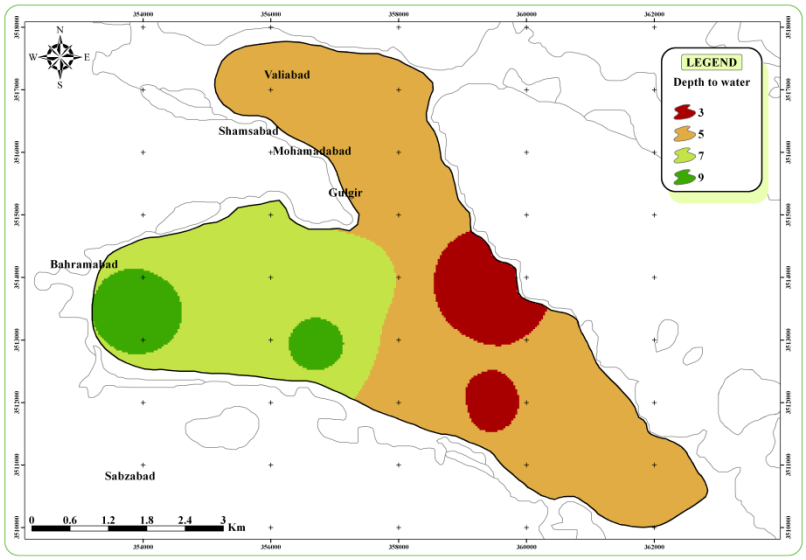
**- Topography**: A rise in the topography slope causes reduction of endurance of water on the earth surface and reduction of the penetration rate. A mild slope causes the rise of penetration rate leading to a higher contamination potentiality. The elevation-numerical files 1:25000 of the State Surveying Organization were used to prepare the slope map. These files contain baselines, elevated points and waterways. The elevation-numerical model of the region was prepared by using these factors. The DEM achieved by editing in the Arc GIS domain, in Spatial Analyst section and by using the Analysis Surface function the slope map of the region was achieved. The topographic layer of the region was then prepared based on the slope rates of the DRASTIC model in table 1. (Fig. 2-e).

**- Impact of vadoze zone:** The non-saturated region includes the existing residues from the water table to the soil zone. This area is non-saturated or alternatively saturated. The log of the observation - exploration wells and a method similar to the layer of the aquifer media were used to prepare this layer with the difference that the thickness and material of the upper layers of the stagnant water level are included in the log of the afore-mentioned wells (Fig. 2-f).

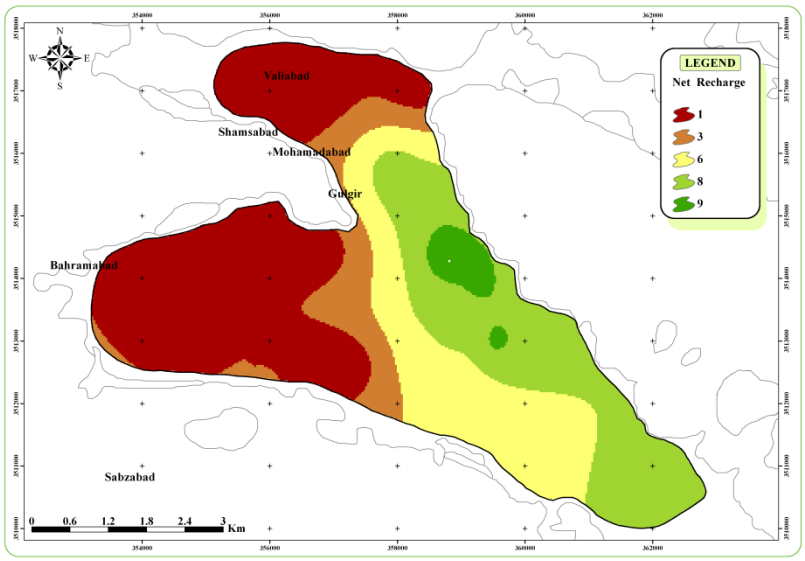
**- Hydraulic conductivity:** The hydraulic conductivity depends on the material of the soil, the intermediary zone and the aquifer. This parameter implies the capacity of water conduction and the contaminations solved in it. The higher the hydraulic conductivity is the lower the contamination mortality would be. The maps of transfer capacity and thickness of the ground water are required for preparation of the hydraulic conductivity layer. The map of hydraulic conductivity was prepared through dividing the transfer capacity layer to the thickness of alluvium, and rated based on table 1. In this way the hydraulic conductivity layer was achieved in order to be combined with other layers (Fig. 2-g)

**Table 1.** Classification and rating of parameters of DRASTIC model in the Gulgir plain

| Rating | Parametr | Rating | Parametr | Rating | Parametr |
|---|---|---|---|---|---|
| | Topography (%) | | Net Recharge (Cm/Year) | | Depth to water (m) |
| 10 | 0 – 2 | 1 | 1.11 - 4.4 | 9 | 2.6 – 4.5 |
| 9 | 2 - 6 | 3 | 4.4 - 7.6 | 7 | 4.5 – 9 |
| 5 | 6 - 12 | 6 | 7.6 - 10.9 | 5 | 9 – 15 |
| 3 | 12 - 13 | 9 | 10.9 - 14.1 | 3 | 15 – 19.8 |
| | | 8 | 14.1 - 16.56 | | |
| **Rating** | Hydraulic conductivity(m/d) | **Rating** | Impact of Vadoze zone | **Rating** | Aquifer media |
| 1 | 7.7 – 8.6 | 9 | Gravel with Sand | 10 | Gravel |
| 2 | 8.6 – 9.49 | 7 | Limestone | 9 | Gravel with Sand |
| 4 | 9.49 – 10.39 | 7 | Silt with Sand & fine Gravel | 8 | Clay with Sand & fine Gravel |
| 6 | 10.39 – 11.28 | 6 | Silt with interbed of Sand | 8 | Sand |
| 8 | 11.28 – 12.18 | 6 | Debrise Limestone & fine Sand | 7 | Silt with Sand & fine Gravel |
| 10 | 12.18 – 13.07 | 5 | Marl & Limestone | 7 | Limestone |
| **Rating** | Soil media | 5 | Silty Clay with Sand | 7 | Silt and Pebble |
| | | 5 | Clay with Sand | 6 | Clay with Gravel |
| 10 | Silty Loam | 4 | Silt | 6 | Silt with interbed of Sand |
| 8 | Loam | 3 | Silty Clay | 6 | Debrise Limestone & fine Sand |
| 6 | Clay Loam | 2 | Marl | 5 | Silt |
| 4 | Silty Clay Loam | 2 | Clay with Silt | 5 | Silty Clay with Sand |
| 2 | Silty Clay | | | 5 | Marl & Limestone |
| | | | | 4 | Clay with Sand |
| | | | | 3 | Silty Clay |
| | | | | 2 | Clay with Silt |
| | | | | 2 | Marl |
| | | | | 1 | Clay |

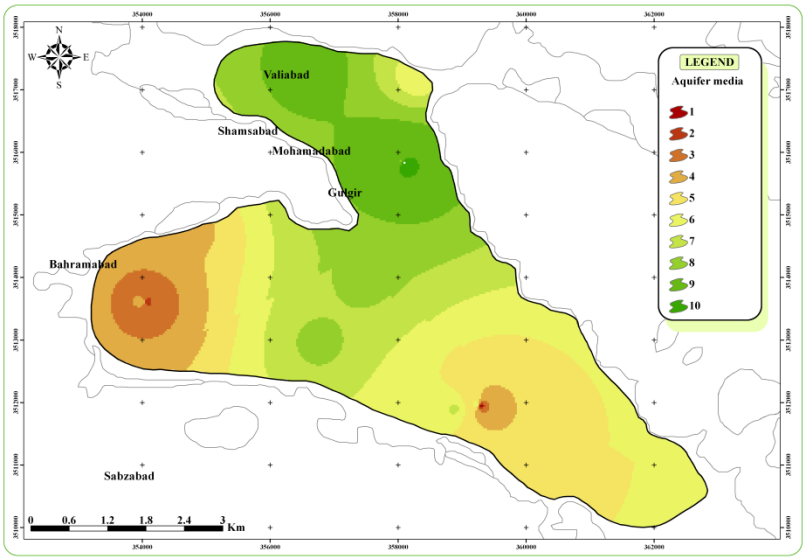



(a)

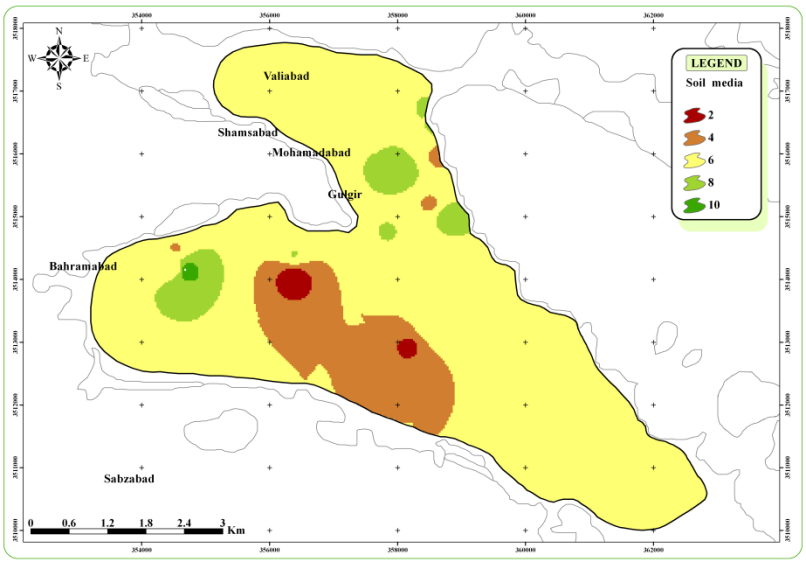



(b)

The beginning of **fig. 2.**

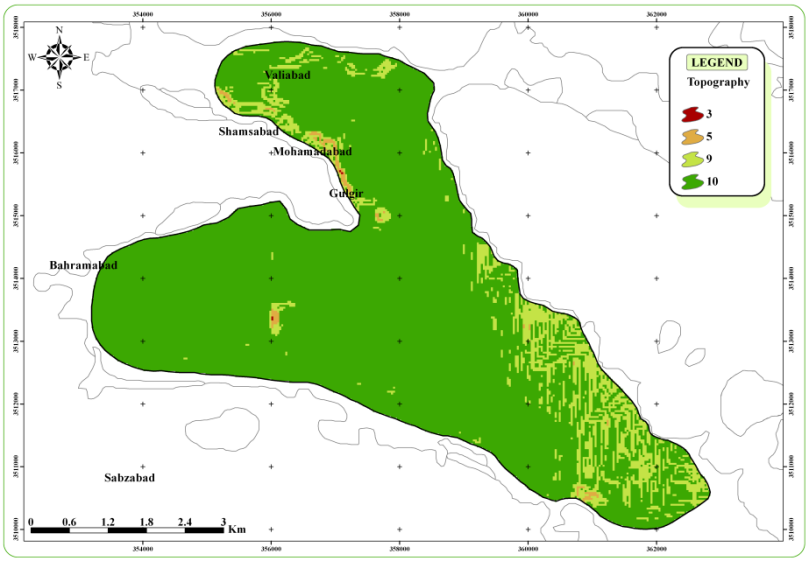



(c)

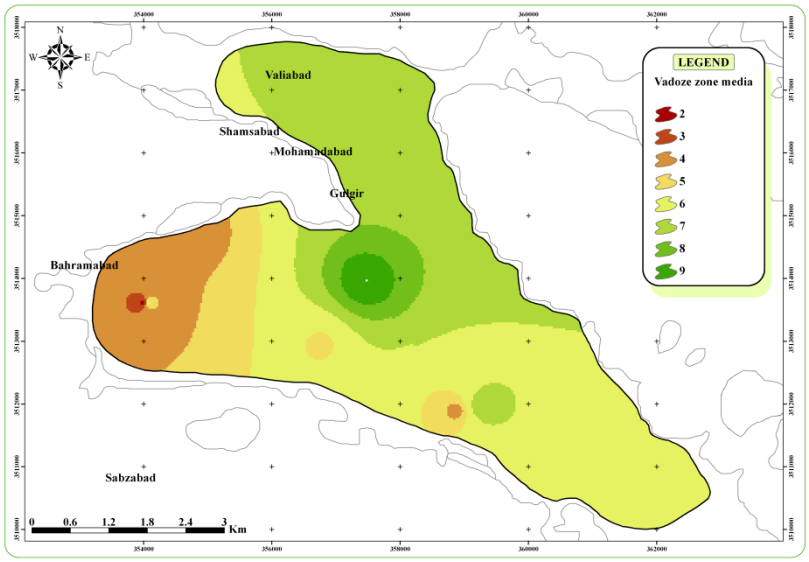



(d)

The continue of **fig. 2.**

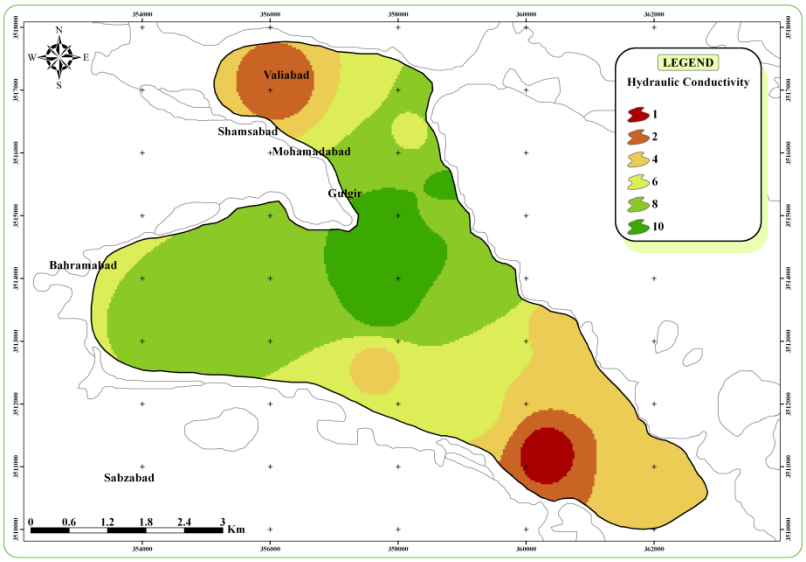



(e)




(f)

The continue of **fig. 2.**




(g)

**Fig. 2**. Map of layers of DRASTIC model (**a**-Depth to water table, **b**- Net Recharge, **c**- Aquifer media, **d**- Soil media, **e**- Topography, **f**- Impact of vadoze zone, **g**- Hydraulic conductivity)

**Fuzzy Model**

Evaluation of the vulnerability of ground water or description of contaminated areas is not an easy task since it depends on many complicated parameters. The lack of certainty and assurance is innate to all methods of evaluation of vulnerability which is the result of information error and variability of the hydro-geological parameters to the time and place [11]. Therefore, preparation of a flexible model that can respond in an uncertain condition with the least entries is a proper managerial tool for evaluation of vulnerability of the subterranean water. Use of fuzzy logic has become highly widespread in many scientific branches which need information classification. Whereas classification of information in evaluation of the vulnerability of the ground water and determination of the border between these classifications are of particular importance, the fuzzy logic would evaluate the vulnerability better than common methods [12]. Stages of making a fuzzy control system is generally expressed as follows (Fig. 3):

1 - Fuzzifier the definite values of the entries

2 - Extracting the rules base and method of fuzzy deduction

3 - Defuzzifier the output fuzzied values.

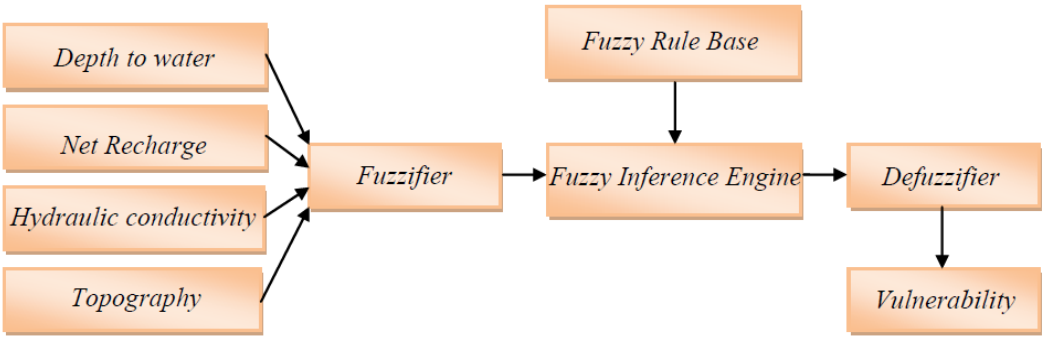



**Fig. 3**. Structure of fuzzy model for vulnerability evaluation

**To fuzzy the definite entry values:** The first step in establishing a fuzzy model is to define the inputs and membership functions. The input parameters of fuzzy deduction model include depth to water table, net recharge, topography and hydraulic conductivity. There is no possibility to make a fuzzy deduction for other three parameters of the DRASTIC model, namely aquifer media, vadoze zone media and the soil media because they lack moderate values. The aforementioned parameters have become fuzzy using of Gaussian membership function. The fuzzy chart of each parameter was drawn by programming in MATLAB software domain. Whereas the parameters are independent they would have separate fuzzy. For the variables of depth to water table, amount of net recharge and Hydraulic conductivity membership functions such as very high, high, moderate, low and very low have been defined. For the topography variable the membership functions of high, moderate, low and very low have been considered (Fig. 4).

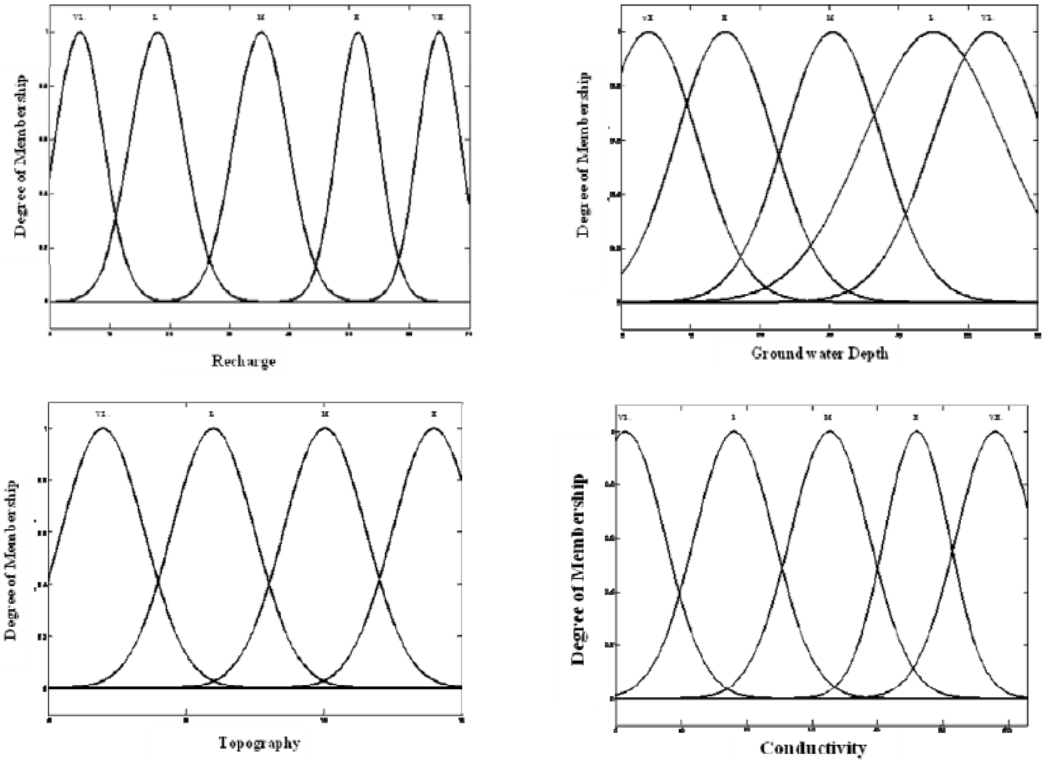



**Fig. 4.** Input membership functions of fuzzy model

The extent of membership in these charts is achieved from relation 3:

$$f = (x; \delta, c) = e^{\frac{-(x-c)^2}{2\delta^2}} \quad (3)$$

Where **c** is the average or center of the Gaussian curve, **δ** is deviation of criterion of membership function and **x** is the value of the parameters. The value of these parameters for the membership functions of the fuzzy model is as follows:

| | | |
|---|---|---|
| c = [2,2,2,2,2.5] | δ = [5.5,9.96,16.5,22,27] | Depth to water table |
| c = [1.5,1.5,1.5,1.5,1.5] | δ = [3.5,7.10.5,14,17.5] | Net Recharge |
| c = [4,4,4,4,4] | δ = [2,8,15,22,28] | Hydraulic conductivity |
| c = [1,1,1,1] | δ = [4,8,14,18] | Topography |

**Extraction of rules base and fuzzy deduction method:** After fuzzify the input parameters; the fuzzy rules base is made. The fuzzy rules evaluates the vulnerability through fuzzy phrases consisting of "If -.then" and in each rule the combined effects of the used indexes are determined in terms of the intended viewpoint. The number of the required rules depends on the number of indexes and number of classes of each index which is calculated on the basis of relation 4.

$$I = K_1 \times K_2 \times \ldots \times K_n \quad (4)$$

Where **I** am the number of rules, **n** is the number of index and **K** is the number of classes in each index. In this research and based on this relation 500 rules may be defined in view of the number of variables and the classes thereof however, 100 rules have been defined based on the studies (table 2).

One membership function is created for the output of each existing rule in view of Mamdani fuzzy Inference method. In this method which is used for conjunctive rules, the lowest membership degree is chosen from the membership degrees of the inputs of a fuzzy rule in each range and transferred to the output. This is repeated for all the rules in all the ranges of changes in the variables in order to achieve the output fuzzy variable.

**Non - fuzzying the output fuzzy values:** Output fuzzy values should be changed into the actual number by using the defuzzifier method and based on the output membership function. There are different methods for non - fuzzying the output. These may include height, average, maximum, Center of gravity, the present research has used the defuzzifier Center of gravity method to calculate the actual value of the output index. This is the most common method for changing the fuzzy quantity into the classic quantity. To calculate the value, the relation 5 has been used [13].

$$\bar{Y} = \frac{\int y\mu(y)dy}{\int \mu(y)dy} \quad (5)$$

Where **y** is the output quantity, **μ(y)** is the output membership grade of y and **Y** is the actual quantity of the output.

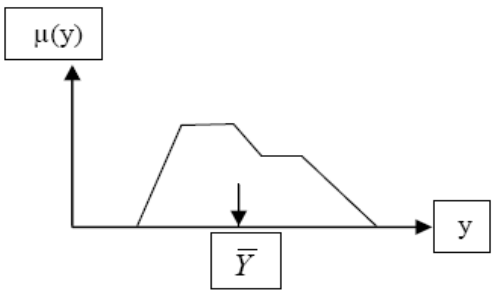



**Fig. 5.** Center of gravity Defuzzifier

**Evaluation of Vulnerability of Aquifer**

Results of the DRASTIC model (Fig. 6-a) show that the highest rate of vulnerability is observed in the eastern part of Gulgir Plain. The high rate of recharge, large granulation of the aquifer, vadoze zone media, soil media and hydraulic conductivity seem to be involved in the

high rate of vulnerability in this part of the Plain. The lowest rate of vulnerability is seen in the northern, western, northwestern and southwestern parts of the Plain. On this basis 15.54 % of the Plain has a very high vulnerability, 39.25 % has a high vulnerability and 45.21 % has a moderate vulnerability. The result of the fuzzy deduction system (Fig. 6-b) also confirms that the eastern part of the Plain has the highest rate of vulnerability. This area enjoys the highest rate of feeding. On this basis, 4.69 sq. kilometers (15.36 %) of the Plain has a very high rate of vulnerability, 2.32 sq. kilometers (7.58 %) has a high rate of vulnerability, and 23.56 sq. kilometers (77.05 %) has a medium rate of vulnerability. Based on this method an extensive part of the Plain would be moderately vulnerable

**Table 2.** Rules base of fuzzy model of vulnerability

| Row | Depth to water | Net Recharge | Hydraulic conductivity | Topography | Vulnerability |
|---|---|---|---|---|---|
| 1 | V.L. | V.H. | M. | H. | M. |
| 2 | V.L. | H. | H. | M. | L. |
| 3 | V.L. | M. | H. | L. | V.L. |
| 4 | L. | V.H. | M. | L. | L. |
| 5 | M. | V.H. | M. | M. | H. |
| 6 | M. | V.H. | L. | L. | M. |
| 7 | H. | H. | M. | L. | H. |
| 8 | H. | H. | M. | M. | H. |
| 9 | H. | L. | L. | V.L. | L. |
| 10 | H. | L. | M. | L. | M. |
| 11 | M. | L. | V.H. | V.L. | M. |
| 12 | M. | V.L. | H. | V.L. | M. |
| 13 | H. | V.L. | L. | V.L. | M. |
| 14 | V.H. | H. | M. | L. | H. |
| 15 | V.H. | M. | M. | V.L. | H. |
| 16 | V.H. | L. | M. | V.L. | M. |
| 17 | V.H. | V.L. | L. | V.L. | L. |
| 18 | V.H. | L. | L. | L. | L. |
| 19 | V.H. | V.L. | M. | V.L. | M. |
| 20 | H. | V.L. | V.L. | H. | L. |
| 21 | H. | L. | V.L. | M. | L. |
| 22 | M. | V.L. | V.L. | L. | V.L. |
| 23 | M. | V.L. | M. | V.L. | L. |
| 24 | M. | V.L. | L. | L. | L. |
| 25 | H. | L. | V.L. | M. | L. |
| 26 | H. | L. | V.L. | H. | L. |
| 27 | M. | L. | L. | M. | V.L. |
| 28 | L. | L. | L. | L. | V.L. |
| 29 | L. | M. | L. | H. | V.L. |
| 30 | L. | M. | L. | L. | L. |
| 31 | L. | L. | V.L. | M. | V.L. |
| 32 | M. | M. | M. | V.L. | M. |
| 33 | M. | H. | M. | L. | M. |
| 34 | H. | H. | L. | L. | H. |
| 35 | H. | M. | L. | L. | H. |
| 36 | H. | L. | L. | L. | M. |
| 37 | M. | V.L. | V.H. | V.L. | M. |
| 38 | M. | L. | H. | V.L. | M. |
| 39 | H. | V.H. | M. | M. | V.H. |
| 40 | H. | H. | M. | M. | H. |
| 41 | L. | V.H. | H. | H. | M. |
| 42 | L. | V.H. | H. | M. | M. |
| 43 | L. | H. | H. | L. | M. |
| 44 | L. | H. | M. | L. | M. |
| 45 | H. | V.H. | H. | L. | V.H. |
| 46 | M. | V.H. | M. | V.L. | H. |
| 47 | V.L. | M. | H. | M. | L. |
| 48 | V.L. | H. | H. | L. | L. |
| 49 | L. | V.H. | H. | L. | M. |
| 50 | M. | V.H. | H. | V.L. | H. |

| Row | Depth to water | Net Recharge | Hydraulic conductivity | Topography | Vulnerability |
|---|---|---|---|---|---|
| 51 | M. | V.H. | H. | H. | V.H. |
| 52 | L. | V.H. | H. | M. | V.H. |
| 53 | V.L. | M. | H. | H. | M. |
| 54 | V.L. | H. | M. | H. | L. |
| 55 | V.L. | M. | M. | V.L. | M. |
| 56 | V.L. | H. | M. | L. | L. |
| 57 | V.L. | M. | H. | V.L. | L. |
| 58 | V.L. | H. | M. | M. | L. |
| 59 | V.L. | M. | M. | H. | L. |
| 60 | L. | H. | M. | H. | V.L. |
| 61 | L. | M. | H. | M. | V.L. |
| 62 | L. | L. | M. | L. | V.L. |
| 63 | L. | M. | M. | H. | V.L. |
| 64 | M. | H. | M. | H. | L. |
| 65 | M. | M. | H. | H. | L. |
| 66 | M. | M. | M. | L. | M. |
| 67 | M. | M. | V.H. | H. | M. |
| 68 | L. | H. | M. | M. | L. |
| 69 | L. | M. | H. | L. | V.L. |
| 70 | L. | L. | L. | H. | V.L. |
| 71 | V.L. | L. | M. | M. | V.L. |
| 72 | V.L. | M. | L. | H. | V.L. |
| 73 | L. | V.L. | M. | H. | V.L. |
| 74 | V.L. | L. | L. | H. | V.L. |
| 75 | H. | L. | M. | L. | H. |
| 76 | V.H. | M. | L. | V.L. | H. |
| 77 | V.H. | H. | H. | M. | V.H. |
| 78 | V.H. | H. | M. | L. | V.H. |
| 79 | H. | M. | H. | V.L. | H. |
| 80 | H. | H. | M. | L. | H. |
| 81 | H. | V.H. | H. | L. | V.H. |
| 82 | H. | H. | M. | L. | H. |
| 83 | H. | L. | H. | L. | M. |
| 84 | H. | M. | M. | M. | M. |
| 85 | M. | H. | H. | L. | M. |
| 86 | H. | M. | M. | V.L. | H. |
| 87 | V.H. | M. | H. | M. | H. |
| 88 | V.L. | L. | M. | H. | V.L. |
| 89 | V.L. | V.L. | H. | L. | V.L. |
| 90 | V.L. | M. | M. | H. | V.L. |
| 91 | V.L. | M. | M. | M. | V.L. |
| 92 | V.L. | L. | L. | V.L. | V.L. |
| 93 | V.L. | L. | V.L. | L. | V.L. |
| 94 | V.L. | M. | L. | V.L. | V.L. |
| 95 | V.L. | M. | H. | V.L. | L. |
| 96 | V.L. | L. | V.L. | H. | V.L. |
| 97 | V.L. | L. | L. | V.L. | V.L. |
| 98 | V.L. | M. | V.L. | V.L. | V.L. |
| 99 | M. | V.L. | V.L. | L. | V.L. |
| 100 | M. | M. | V.L. | L. | M. |

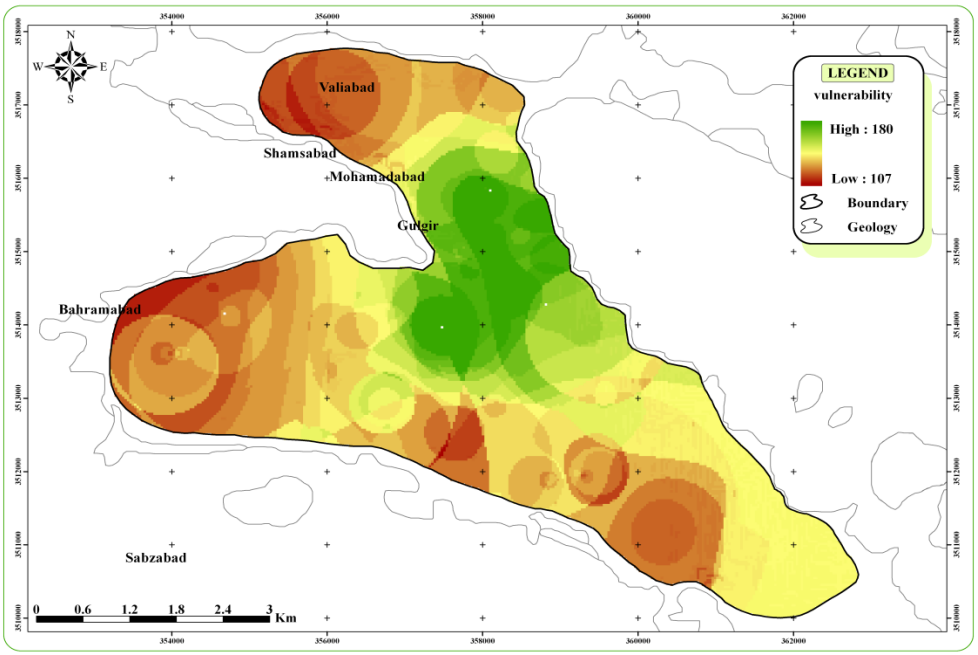
« Very Low: V.L. ‹Low : L. ‹Moderate : M. ‹ high : H. ‹Very high :V.H.»

**Verification of the Model**

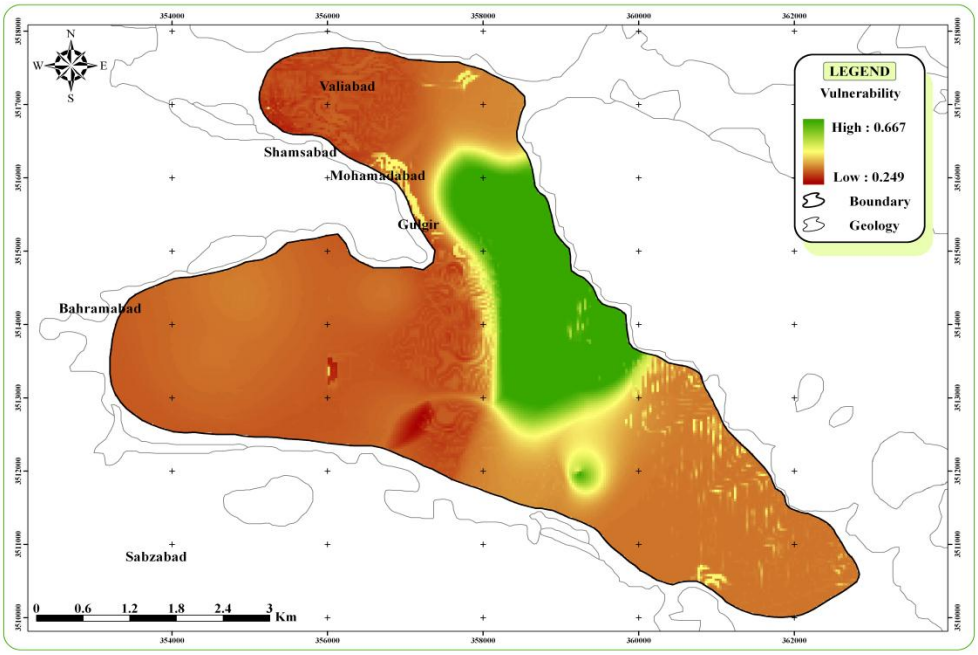
In order to verify the results of the model regarding the nitrate density, data samples taken from the water well in March 2006 were used. Presence of nitrate in the subterranean water implies the destruction of the water quality. Due to the use of fertilizers, in the region which is mainly manures or nitrated chemical fertilizers the nitrate ion existing in the ground water is analyzed.

Based on the nitrate coaxial map (Fig. 7), the volume of nitrate in the ground water is higher in the eastern and northern parts than other parts of the Plain.

In order to specify the parameter with the highest effect on vulnerability of the ground water in the Gulgir plain, the correlative coefficient between the parameters of the DRASTIC model and the nitrate layer was calculated (table 3). The results show that net recharge has the most correlation with the nitrate layer and therefore it has a higher impact on the vulnerability of the subterranean water as compared to other parameters of the DRASTIC model, followed by hydraulic conductivity, aquifer media and vadoze zone media.




(a)




(b)

**Fig. 6**. The vulnerability map of Gulgir Plain (**a**. DRASTIC model, **b**. fuzzy model)

**Table 3**. Correlation between layers of DRASTIC model and nitrate density

| Hydraulic conductivity | Vadoze zone media | Topography | Soil media | Aquifer media | Net Recharge | Depth to water | Layers |
|---|---|---|---|---|---|---|---|
| **0.3881** | **0.2791** | **0.2081** | **0.2737** | **0.3145** | **0.5138** | **0.2236** | Nitrate |

## CONCLUSION

Results of the present research generally show that the general trend of vulnerability is the same in using the definite and fuzzy layers. However, when the fuzzy layers are used, the gradual mode of change appears more clearly. Whereas layers are combined through Boolean logic in the DRASTIC method, errors emerge in border values and a slight change in the quantity of the factors causes relocation from one class to another which is not justifiable. Therefore, use of the fuzzy logic improves the accuracy of the results particularly concerning the border values.

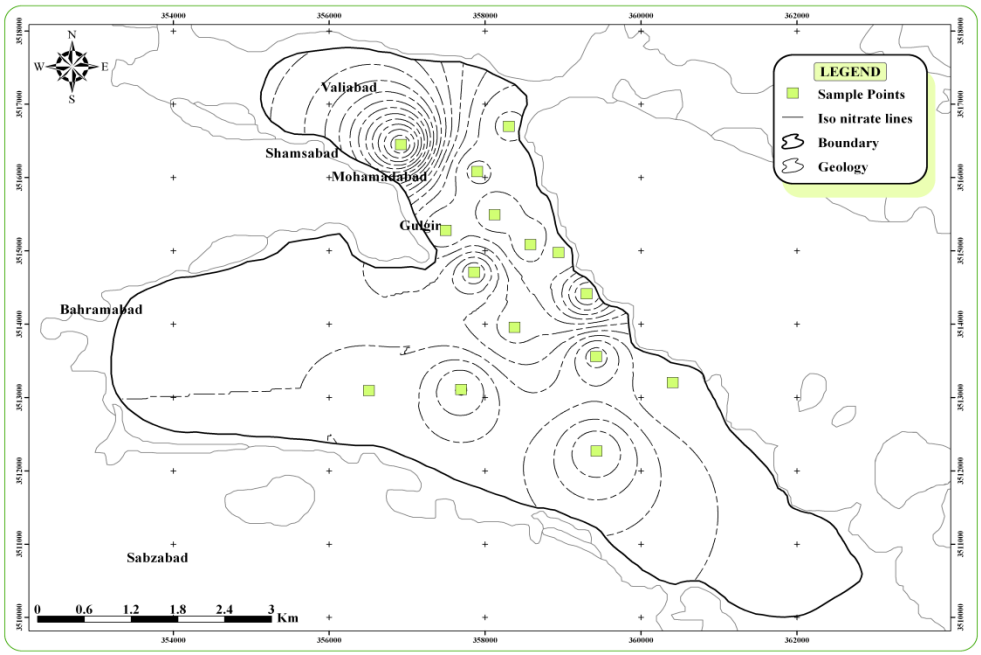



**Fig. 7**. Coaxial nitrate map in subterranean water of Gulgir Plain

## REFERENCES

[1] US EPA. (1996) Environmental indicators of water quality in the United States, Washington, D.C., Office of water, EPA, Washington, D.C, EPA 841– R – 96 – 002.

[2] Melloul, A. and Collin, M. (1994) Water quality factor identification by the principal component's statistical method. J. water Sci. technol., 34, 41 – 50.

[3] Vrba, J. and Zaporozec, A. (1994) Guidebook on mapping groundwater vulnerability, IAH International Contribution for Hydrogeology 16, 131.

[4] Malano, H.M. and Gao, G., (1992) Ranking and Classification of irrigation system performance using fuzzy set theory: Case studies in Australia and China. Irrigation and Drainage Systems, (6): 129 – 148 .

[5] Aller, L., Bennet, T., Leher, J. H., Petty, R. J., and Hackett, G. (1987) DRASTIC: A standardized System for Evaluating Groundwater Pollution Potential Using Hydrogeologic Setting. EPA – 600/ 2 – 87 – 035. Ada, Oklahoma :U. S. Environmental protection Agency.

[6] Malczewski, J. (1999) GIS and Multicriteria Decision Analysis. John Willey & Sons, Inc.

[7] Rushton, K. R. (2003) Groundwater Hydrology, Conceptual and Computational Models. John Wiley & Sons Inc., 416.

[8] Khuzestan Water and Electricity Organization, deputy for basic studies and comprehensive plans of water resources, management of basic studies of water resources, 2009-10. Report on studies to locate water hardening resources (Karts) of Asmari anticline.

[9] Scanlon, B.R., Healy, R.W. and Cook, P.G. (2002) Choosing appropriate techniques for quantifying groundwater recharge, Hydrology Journal 10(1), 18 – 39.

[10] State Water and Soil Engineering Service, management of Karkheh studies(2010) Report on semi - detailed accurate pedological studies of the lands in Gulgir region of Masjed Soleiman.

[11] EPA (1993) A methodology and decision support System for evaluating the leaching potential of pesticides, U.S. Environmental protection Agency, EPA, 600/S – 93/010.

[12] Dixon, B., Scott, H.D, Dixon, J.C. and Steele, K.F. (2002) Prediction of aquifer vulnerability to pesticides using fuzzy rule – based models at the regional scale, Physical Geography 23

[13] Teshnehlab, M., Saffrapour, N. and Afyouni, D. (2009) Fuzzy Systems and Fuzzy Control, Publication of Khajeh Nassireddin Tousi University of Technology

# ANALYSIS OF ELEVATION DATA WITH TIME ASPECT FOR ATHLETES

Rostislav NETEK[1] and Jaroslav BURIAN[1]

[1]Department of geoinformatics, Faculty of Science, Palacký University, Olomouc, tř. Svobody 26, 771 46, Olomouc, Czech Republic

*rostislav.netek@upol.cz, jaroslav.burian@upol.cz*

**Abstract**

This paper presents specialized analysis based on 4D data for athletes. Generally the data received from GPS (Global Positioning System) devices contain elevation, spatial and temporal aspects. This type of data allows localizing current position of athlete in every moment. Evaluating and planning performance for professional athletes is based on elevation data primarily, because just elevation plays very important role for evaluating and planning for professional athletes. There are many variables in different sports which are tracked and monitored during the performance; like cadence, power, speed, heart-rate and many others. All of these variables are closely dependent on both elevation and time, and periodic monitoring is strictly required for training. There exists some specialized software which enables complex analysis of telemetry data from athlete's sensors depending on elevation with possibility of visualization in tabular, attribute, 2D or 3D graphic form.

The paper discusses different types and restrictions of elevation data received from GPS devices, especially possibilities of analysis which are really used by professional athletes and real experiences with using Garmin Edge 800 and Forerunner 310XT devices as well. Created application for upload data in GPX and/or KML file for visualization elevation data in time line is described. The application is based on interactive elevation profile, which allows analyzing information in 2D, 3D or 4D point of view, with manually and automatically shifting in time. For every moment are available elevation and other dependent variables shown in graphic, attribute and spatial form with focus on altitude.

**Keywords: elevation data, sport, web application**

## INTRODUCTION

There are many fields, where elevation data are recorded and analyzed nowadays, because in last few years GPS devices are widespread for common users as well as for experts. GPS sensors are implemented in lot of kinds of equipment; users can record their route by mobile-phones, laptops, tablets, watches, cameras and of course by standard GPS receivers. In most cases the data are used, analyzed and visualized only in three dimensions, where on X and Y axis is captured horizontal location, and Z axis contains vertical value (in other words the elevation). In fact not only 3D data are captured, because all GPS devices allow 4D data capturing. The last fourth dimension means, that data are captured in some predefined time period, and typical positioning data can be analyzed also in time in addition. Actually, the sport is typical field in which analyzing data in time-line is very necessary. In addition there are many variables used in different sports, which are monitored during the performance, like

cadence, power, speed, heart rate, temperature, wind-speed and many others. It means that athletes can evaluate tracked values from sensors in every moment of activity, and compare each activity with others (for example the same route during the year). The elevation plays very important role because power of athletes in every sport activity is strictly dependent on elevation profile in time. These analyses are very important for evaluation trainings and/or races with aim to improve the performance in the future. Only elevation data with time aspect gives possibility of objective evaluation, which is only one way to success for professional athletes nowadays.

## REVIEW OF GEOSPATIAL TECHNOLOGIES FOR SPORT

In the last few years we have seen a huge boom in the using of modern geographic information system (GIS) and geospatial technologies (GIT) for sport activities. On the other side, it is still not much widespread like in other fields (compare with traffic, military, archaeology or geology). The most common use (tourism, cycling, sailing, orienteering, rally, in-line skating, cross-country running, hiking, aviation …) of GIT is navigation and/or tracking routes by GPS devices during the trip. The Internet is currently the most widely used medium for sharing information, so the other use is upload and share source data into specialized map application like e.g. OpenCycleMap, MapMyRide or Service Garmin Connect [1]), with focused on both spatio-temporal and elevation component of data.

The location-based service (LBS) gives geospatial information, where geospatial means that it is based on the current geographic location. In all fields of human activity, including leisure-time activities and sport, is possible to localize position in time and tracking the movement. The sport gives us great opportunity to implement modern technologies like LBS and make some scientific research, because in every kind of sport is possible to analyze some movement of athletes [2], [3]. There exist many advantages why to track athletes – analyze and evaluate their training, locate their position in unknown surroundings, protect their safety, check their physiological characteristics because of health and obesity reasons or increase attractiveness of TV broadcast. The main principle is to locate the athletes by GPS sensor with high precision and to transmit consecutive data into spatial database. Data can be transmitted by Wi-Fi, GSM, SMS or modem in real time, which allows displaying position of athletes in real-time on screen [4]. Loggers have internal memory, which allows downloading data from device later into computer. In current years GPS tracking is more and more implemented in all kinds of sport. It is integral component of present sport activities all over the world.

### Professional vs. Amateur Sport

From the principle of sport, this issue can’t be viewed integrally, but it is also necessary to separate recreational and amateur sports with a genuine professionals. Professionals have a lot of possibilities to use the latest technologies and products, which are often very expensive and not available for the public. Selected professionals are often so called "testing racers" for example in the skiing, the cycling, or the Formula 1 drivers. They are testing prototypes of equipment, which is made in only one or several pieces in special version. Only on the results of these testing companies decide about the future of product development. The professionals

have often access to extra function and/or tools which are not available in standard mode. This serves especially for testing and developing special devices.

In cycling, two of the best professional teams are sponsored by Garmin and HTC companies. The racers of Garmin-Cervélo Pro-Tour team and HTC-Highroad Pro-Tour team keep at disposal GPS devices about one year earlier than other teams, because Garmin and HTC companies testing their new devices just on team members. One season later, this device is available for other professional cyclists and earlier two seasons later, the same device is available in the shops for public. Generally at this moment the price is very high and it takes about two years to wide expansion into amateur cyclist. It can take about five years between the first use by professional and spread to the public. This is the reason, why focusing only at professionals athletes in this article is more relevant.

**Role of Elevation and Time During the Performance**

In every sport the general aim is to be better and/or be faster than others. Because of physiologic aspect each person is different and each person has different presumptions for performance of course. Due to this individual distinction there are some physiologic variables which are measured by professional athletes. The most objective two characteristics are Power (in Watts; rate of work performed by an athlete in one moment) and Lactate (in millimoles per litre of blood; value of lactic acid produced by muscles during the performance), but both of them require very sophisticated devices. Often this kind of data can be captured it only in medical laboratories. Due to this reason is more often used heart-rate monitor (unit: Beats-per-minute) for everyday individual necessities, in addition to speed and/or cadence sensors in some kinds of sports (Fig. 1).

The elevation plays most important role in measuring by sensors, because every of mentioned characteristic is strictly dependent on altitude and elevation changes – in higher elevation the power goes down, muscles produce more Lactate, depending on previous both the heart-rate value goes up, the speed goes down. Very similar situation like increasing in elevation bring increasing in time of performance, especially in long-time endurance.

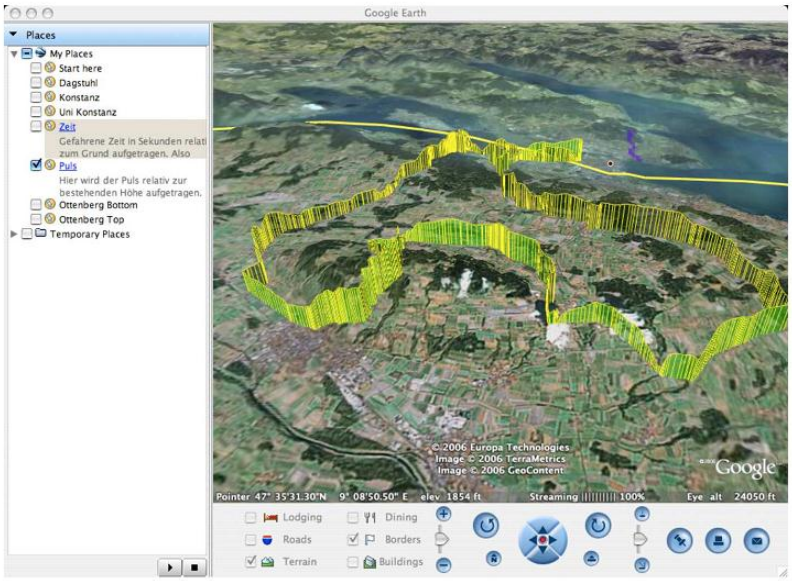



**Fig. 1.** Visualization of heart-rate values above terrain [5]

## Cycling

A lot of cyclists use every day GPS receiver, but for professional riders are beneficial only connection with the speedometer, cadence-meter, heart-rate monitor and power-meter (see Fig. 2). Only comprehensive data obtained from all these sensors allow professionals to fully analyze their training. The itinerary can be evaluated in the numerical statistics in a graphical time record, including tabular information on all measured variables for every moment (see Fig. 3). Specialized devices produced by companies Garmin, Suunto or Polar are used in Pro-tour races. Especially Finnish company Polar (2011) is the world leader in heart-rate sensors. In every cases, the process of analysis of races output is allowed only by coherent multidimensional telemetry data received from GPS and other sensors.

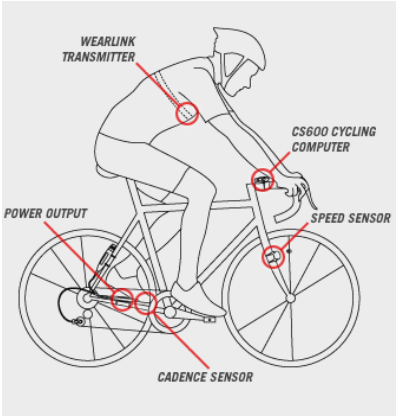



**Fig. 2.** Sensors used in professional cycling (Polar: http://www.heartratemonitor.co.uk/polar_wind_power_kit.html)

According to Saupe´s research [5] GPS device Garmin Edge recorded longitude, latitude and elevation simultaneously with other cycle sensors. Fig. 1 visualizes heart-rate values above 2.5D terrain data for a 50km training ride in Switzerland. The heart rate is displayed as the height of the yellowish band along the cycling track. The output of heart rate, cadence, time stamp, longitude, latitude, and altitude per sample is given in the form of an XML file. For the synchronization it suffices in this case to manually identify the time origins in both sequences and then to interpolate positions assuming a constant frame rate for the video. In this application time series synchronization allows replay of any given training session on the same training course, given a corresponding sequence of recorded GPS parameters. Moreover, for biofeedback experiments in the laboratory, computer controlled traction of an ergometer station can be accurately rendered. This setup could open a number of new possibilities for athletes, coaches and scientist [5].

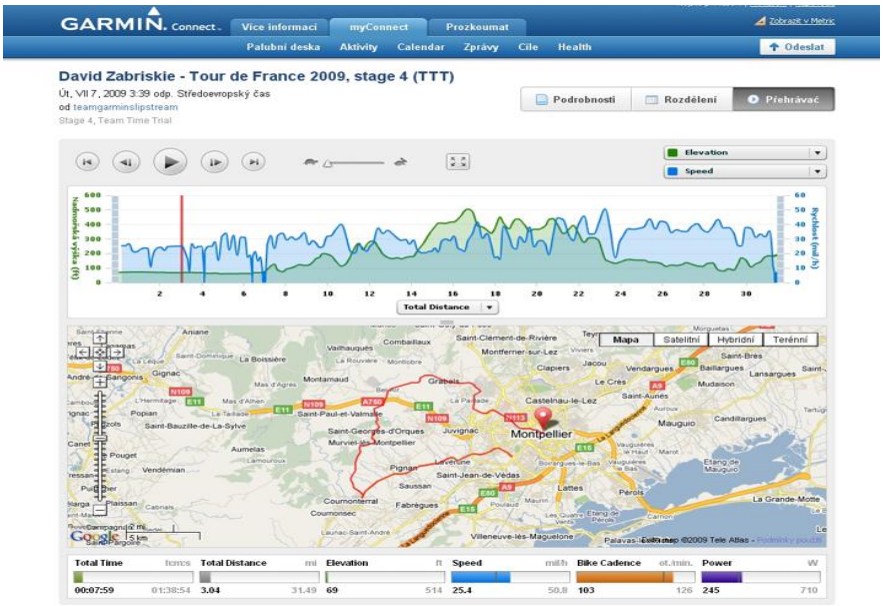



**Fig. 3.** Visualization of telemetry data, David Zabriskie 4th stage of the Tour de France 2009 (Service Garmin Connect: http://connect.garmin.com)

Some special devices allow so-called „virtual training partner" [6], [5]. The device can load another route file with both spatio-temporal and elevation characteristics (race of opponents and own race from another year) and the device shows deviations from the trajectory and time of the virtual opponent, in addition some stationary bicycle can regulate the resistance depending on elevation in real route (it means, that if in real route go up to hill, the stationary bicycle regulate the resistance higher, due to value of elevation in time). With synchronized video which is displayed at the same time with satellite images, is possible to simulate a real race track. It is possible to download all outputs from pro-cycling-team websites and simulate any route at your home.

Already established attractive technology is GPS tracking with real-time visualization on TV screen. Competitors are localized in real-time using the GPS chip and simultaneously are

their position display on the TV screen. In addition to TV broadcasts, the technology is being widely used in stage races (Tour de France, Giro d´Italia, Vuelta de Espana, etc.), where visitors have opportunity to watch events on the big screens around the track or in TV at home. Competitors receive sensitive GPS sensor which capture data from GPS and other devices (heart-rate monitor, speedometer etc.) and stores it on SIM card. Then are data send via GPRS to the central server, which graphically displays the actual position, elevation and all other variables of the rider on the screen during live the races [4]. Fig. 4 illustrates usage of elevation data with time aspect in real case. Beside real-time “2D” location of rider Maxime Monfort (yellow dot in map) and route of stage, there is actual position in elevation profile accompanied by actual values from four bicycle sensors (Speed, Power, Cadence, Heart-rate). The 2D/ 3D position of riders and all values have time aspect as well, because they are updated every two seconds. The only technical restriction can be surroundings of the rider (like dense forest), which eliminates the signal from satellites, but it is a minor disruption which takes only a few seconds, so on the screen is still possible compare leaders. In some special events the link can be combined with a video camera sees from helmet´s camera. GPS tracking for the "biggest" races is used since spring 2008, where was first time used at Giro d´Italia. During the Tour de France 2010, this technology was used for the first time such a general classification in real time. But it is necessary to say, there are some special occasions, when it cannot be considered 100% reliable for classification. For example in 3rd stage of Tour de France 2011, had Lance Armstrong 3 defects in a single stage. In the case of emergency, riders with no ambitions give own bike to the captain, so Lance Armstrong rode more than one bike (with different sensor) during a stage. In all cases technology of GPS tracking is very attractive for fans [7].

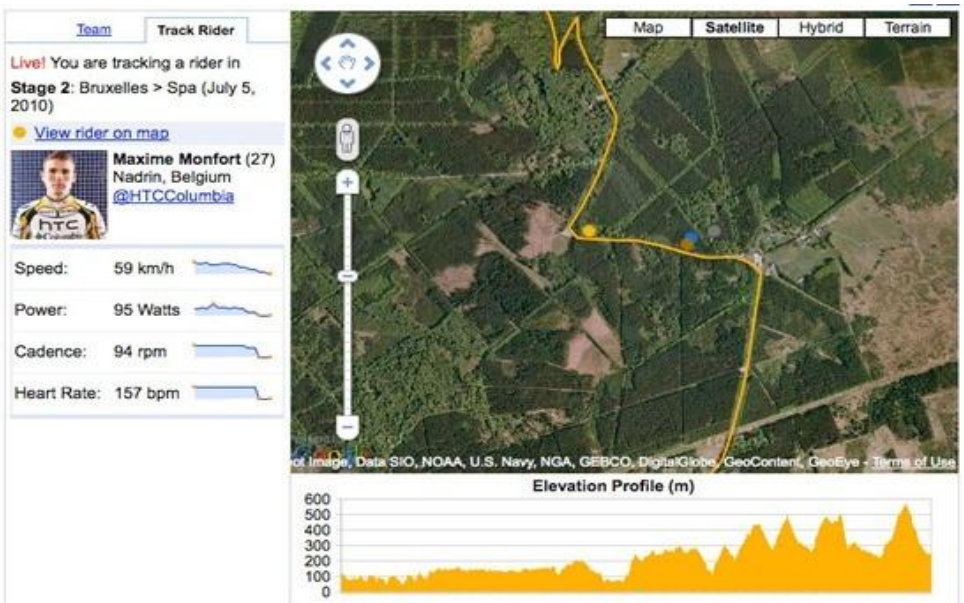



**Fig. 4.** GPS tracking – yellow dot indicates position of Maxime Monfort during the 2nd stage of the Tour de France 2010 (brown, gray and blue dots indicate other racers) [6]

**Running, Orienteering, Cross Country Skiing, Inline Skating**

In the northern countries the orienteering is very popular outdoor sport. In this sport the issues discussed in this paper was used for the first time. The principle is very similar to the cycling, only with some technology restriction because orienteering take place in the forest. Also only selected athletes are equipped by GPS receiver, because competitors start in intervals, not all together at the same time. Because of this reason is not possible to use it for general classification in real-time, but it is very popular for screening the competitor's current position directly on the map. On the other side, there is great opportunity to analyze every part of run after the race and make complex analysis session (identify mistakes, time-losses, the reasons for the time-losses, better solution for the future, combination of different solutions etc.) [8]. More than in any other sport branch in orienteering is elevation significant for athletes. Due to this reason, every application for visualization contains elevation profile.

In some running shoes is used technology for localization athletes by GPS shoe chip (miCoach by Adidas; and Nike+ by Nike). In both of systems is supported connection with Polar heart-rate monitor. Just complex data are valuable to analyze training by professional athletes. The advantages of the using this system is almost the same as in cycling. The runners need to evaluate attributes, because only comparing time/distance with their hear-rate in every second together allow them place for improving. There is function of virtual training partner and audio coaching by synchronization with mobile device. Based on predefined times/distance in every moment during route, the device compares your time/distance with "partner" and via mobile device give you signal about current situation.

There is a similar situation in sports like cross-country skiing or inline-skating. Nowadays the GPS tracking is required in a lot of competitions, but very interesting are applications built on a ski routes. Some volunteers or snowmobiles (vehicle which fixes cross-country routes), have own GPS receiver connected with thermometer and communicate via GPRS with central server. In the result, on the web application visitors can see completely up-to-date information about terrain covered by snow and weather conditions around [2].

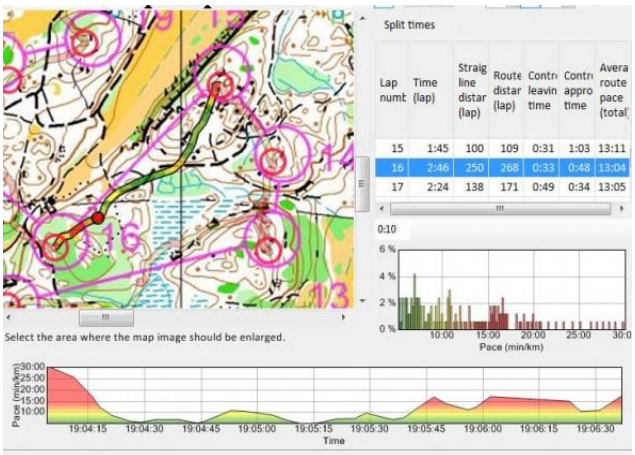



**Fig. 5.** GPS tracking output for orienteering (O-training: http://o-training.net/blog/2011/04/13/gps-analysis-for-orienteering-the-basics/)

Some special devices are used also for safe extreme skiing and ski-touring. Every skier has his own device (avalanche transceiver) that is used during the activity like active device (sender). In dangerous situation (avalanches) the device can be switched to the searching mode (receiver) and allows finding person trapped in avalanche. These devices use radio frequency. There are two basic transceiver types, analogue and digital. Digital devices give a range and direction; analogue devices provide an audio indication of distance and direction.

### Sailing

Not orienteering, but sailing implemented GPS tracking for the first time at global scope. It was great to see live GPS tracking used in the Olympic Games Beiing 2002, where BBC made the first global broadcast with GPS tracking and show outputs in real-time on the screen. Today is GPS tracking on boats for navigation and because of emergency cases, however for comparing of boat trajectories is time-line very significant [8].

## THE "PRAŽSKÉ SCHODY" RACE APPLICATION

The aim of our research was to create a specialized web mapping site, for organizers, visitors and racers, working with spatial and time aspects of real data on some real route in Czech Republic. The application is based on interactive elevation profile, connected with attribute information and photos in map field, moreover accompanied by video from biker's perspective. It allows analyzing information in 2D, 3D or 4D point of view, with manually and automatically shifting in time. For every moment are available elevation and other dependent variables shown in graphic, attribute and spatial form with focus on altitude. It was chosen "Pražské schody" race, which allows applying a lot of interactive functions. The application is available at: http://gislib.upol.cz/app/netek10/schody/map.html .

### Elevation Data

The "Pražské schody" race is the most attractive mountain bike (MTB) race in the Czech Republic, with an attractive downhill on the Old Castle stairs at the Prague Castle with the participation of the world's best cyclists.

Data were firstly captured by GPS device Garmin Edge 800 by one of the racer on the same route in previous year. Longitude, latitude and elevation in time with a resolution of one sample per second were recorded. It means that were collected real data from real route on real race!

Because one of the aims was to connect every edge-point with some attribute information and photo, original GPX file from Garmin device were modified into more than 30 edge-points To each of edge-point was manually added into database textual information and photo describing this place. Moreover, for each edge-point were automatically extracted or calculated from GPX file values of azimuth, partial and total distance, position into map field (for connection and synchronization between database and map) and two specific number values: elevation and time aspect. Elevation value ensures correct line of elevation profile (in GPX file it can be recognized by element <ele>). The second one, time value provides

synchronization of profile animation with map animation in map field (recognized by element <time>). The data were stored in MySQL database on the server.

### GPX File

All Garmin devices store recorded data in *.GPX files. There is great opportunity to use GPX files for spatio-temporal and elevation data, because of some typical elements. GPX (GPS eXchange Format [9]) is an XML schema designed as a common GPS data format for software applications. It can be used to describe waypoints, tracks, and routes. Its tags store location, elevation, and time and can in this way be used to interchange data between GPS devices and software packages. Latitude and longitude are expressed in decimal degrees using the WGS84 coordinate system. Dates and times are not local time, but instead are Coordinated Universal Time (UTC) using ISO 8601 format (Wikipedia). The structure of GPX file corresponds to XML file, it contain some elements, like header or metadescription *(<metadata>* element) with some variables about the route. The main part of GPX file takes element *<trk>* and *<trgseg>* contains segment of route or whole routes, which is divided in particular edge-points called trackpoints *<trkpt>.* Each of trackpoint contains values about position and two characteristics about elevation and time (*<ele>* and *<time>).*

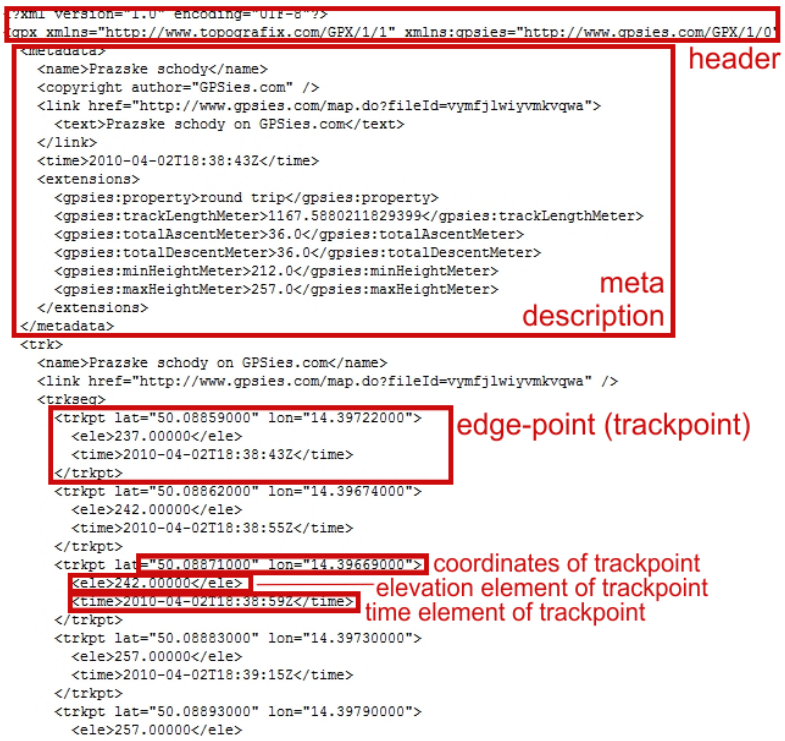



**Fig. 7.** Structure of GPX file with <ele> and <time> elements

### Methods

In the application a range of technologies was implemented. Especially Google Maps API using JavaScript, PHP for dynamic pages generation, which communicates with MySQL

database, and of course CSS and HTML for formatting own Web site were used. With Google Maps API is possible to create complex mapping applications using a combination of pre-programmed functions and/or scripts (e.g. JavaScript). This process is called a mashup. The result of this process is an application (so-called Web 2.0), which combines content from more than one source and allows the simultaneous use of different types of information and/or source combined together. This principle gives programmers the ability to create virtually low-cost applications, which are qualitatively comparable with the best commercial products. Besides the theoretical summary of the latest technologies using in cycling is the main result implementation these functions into real application. The application supports both KML/KMZ and GPX files for track, but because of native implementation of <ele> and <time> elements in GPX file, is better to use GPX file due to synchronization with animation.

**Application**

The prototype application demonstrates wide possibilities of the GI technologies. The application was focused and built on the interactive and dynamic elevation profile of the race, which is also linked to the map field. Dynamic means that profile can be run both manually and automatically as animation in three levels of speed. In the tools menu can be found in a wide range functions: deactivation/activation of layers (difficulty, type of surface, incline, critical points, critical points, audience sectors, parking), animation of track with information to the individual sections, export in 22 formats, "onscreen" search the route, an interactive itinerary and connected with animation and video tracks from the perspective of the rider, which simultaneously displays the location of the point on elevation profile, and in the map animation. There are administration possibilities with expanded functions to add/edit/delete other events and information about them.

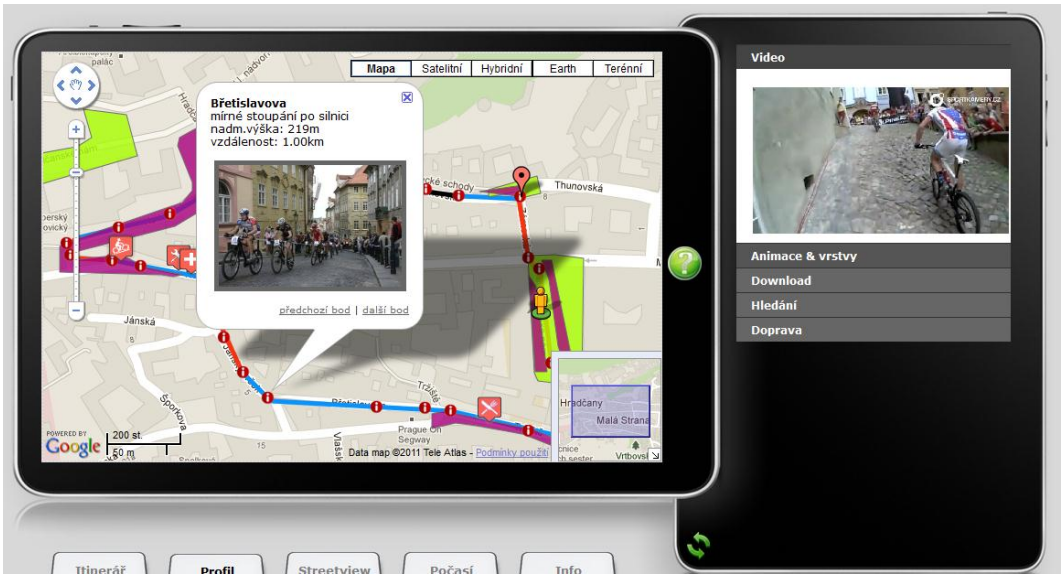



**Fig. 8.** Application “Pražské schody” Race

### Interface

Web interface is divided into 3 main parts - the main map field, the primary tools menu on the right side of the map field and the other tools under the map field (Fig. 8). There are also “Help” and “Page refresh” icons with adequate functions. After many disappointing attempts were for interactive profile used Flash technology, map field is based on Google Maps API - JavaScript version. The application was optimized and correctly viewed in browsers: Firefox 3.6 and 4.0, Opera 10.60, Chrome 5.0. All photos are used with permission of the author Michal Červený.

### Tools and Functions

Map field use standard view of Google Maps, scale and pan tools are located in the top left corner, switcher between 5 layers (standard, satellite, satellite combined with the street network and descriptions, Google Earth, terrain) is in the top right corner. In the bottom left corner the graphic scale is located; on the opposite corner is small preview map, which can be minimized. Map movement is allowed by press the left mouse button while dragging the mouse or by the arrows in the top left corner. Zoom can be changed by the mouse wheel; function zoom by double click was disabled, because this process is assigned to searching by entry points. There is 34 “info-points” attached to each edge-point of the route and each of them includes the name, description, elevation, distance from the start, photography and links to the next or previous “info-point”. Click by mouse on the red “info-point” icon displays the appropriate information.

In the tool menu on the right side five tabs are located. The first one contains video captured from the perspective of bikers (camera was located on the biker´s helmet). In the second tab user can (de)activate layers (difficulty, type of surface, climb/descent, important points, critical points, spectator and parking sectors) and show its legend. There is also command to deactivate all layers at once. Very interesting is possibility to control animation by user (start / pause / resume / first-point/ last-point buttons) which automatically step-by-step displays information to each info-point. The third tab gives possibility to download the track in 22 different digital formats supported elevation data: KML, KMZ, PCX5, GPX Track, GPX Route, Garmin CRS, Garmin TCX, Garmin Logbook, OVL, Fugawi, KOMPASS Verlag, GeoRSS Track, Excel, OziExplorer, MagicMaps IKT, PathAway, Navigon RTE 5.x, Navigon RTE 6.x, Navigon Freshroute, TomTom ITN, Magellan, qpeGPS. In the fourth tab is standard search form, and the last one contains opportunity to find a connection with itinerary to the track for visitors.

The second tool menu under the map contains another five tabs. The first one is a detailed interactive itinerary connected with map (the map is centred on the selected item), in addition it allows searching and/or sorting records. The second tab, probably the most important one, contains an *interactive elevation profile* – particular points on the line profile corresponding with the location in the map, the map is centred when user pass through the profile-line. There is also available animation in three levels of speed; it shows the location in elevation profile and map continuously. The animations are allowed by <ele> and <time> elements in GPX file. Animation can be run at three speeds or manually moved. The third tab is about realistic look into the street allowed by StreetView technology. Localization of the projection point in the

map is indicated by yellow-icon. The next two tabs show the current weather situation and basic information about the circuit. A complete list of features is available in Help.

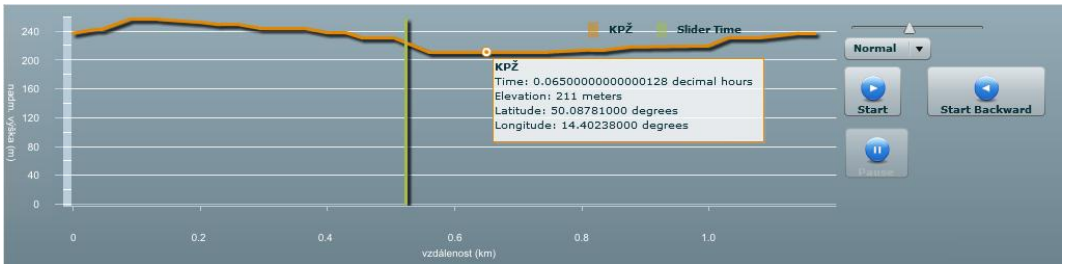



**Fig. 9.** Interactive elevation profile with animation of "Pražské schody" Race

**Curent Situation**

The application has been tested, runs on the server and it is fully functional now. Due to a simple installation, is ready for immediate launch, regardless of its location, on other server and/or for any kind of sport. It is for free and fully independent on original author. There are number of really great similar applications, but mostly they are focused on concrete activity only. It provides great potential for widespread usage of similar technologies in any sport activities (for organizers, racers and viewers as well). Particularly it is used for racers just now (web applications for performance's visualization), unfortunately in very limited form compared with wide opportunities. The behavior of the organizers is the biggest restriction, because they are conservative and often they don't want to include new technologies despite the advantages. First of all organizers must be opened for technologies of 21$^{st}$ century. When approach of organizers will turn to positive status, then we will meet GI technologies in every sport field everyday.

**CONCLUSION**

The paper shows an overview of elevation data with time aspect used by athletes for the training and competing activities. Because it takes a long time, from the first usage by professionals to spread to the public, is more relevant be focused only on professionals. Typical usage of 4D data for sport is so-called GPS tracking implemented in wide range of sports, both of individual and collective sports. There are specialized GPS tracking technologies and applications in sports such as cycling, orienteering, sailing, cross-country skiing or running as well. Field of professional sport and leisure-time activities provide a great potential for implementation both spatio-temporal and elevation data for analysis, visualization, evaluation and planning performance. Online web map application focused on mountain bike race exhibition "Pražské schody" was created. Together 22 formats of the race track are available; in addition to basic information about the route profile is supplemented with interactive animations, itineraries, videos, animations and other tools. It is obvious, that multi-spatial data together with modern geospatial technologies will play important role in field of sport in future as well. Geospatial technologies are becoming natural part of professional sport and bring big advantages for athletes and fans.

## REFERENCES

[1] Garmin (2010) Service Garmin Connect. http://connect.garmin.com/, [cit. 2010/08/04]

[2] Ahtinen, A., Isomursu, M., Huhtala, Y., Kaasinen, J., Salminen, J. and Hakkila, J. (2008) Tracking Outdoor Sports – User Experience Perspective. In: (ed.) E. Aarts et al.: Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, Volume 5355/2008, 192-209.

[3] Antikainen H., J. Rusanen, S. Vartiainen, M. Myllyaho, J. Karvonen, M. Oivo, J. Similaja and K. Laine (2006) Location-based Services as a Tool for Developing Tourism in Marginal Regions. Finland, Nordia Geographical Publications 35:2, p 39-50. NGP Yearbook 2006.

[4] Brown, G. (2007) GPS tracking for athletes? http://www.cyclingnews.com/news/gps-tracking-for-athletes, [cit. 2010/08/03].

[5] Saupe, D. et al. (2007). Analysis and visualization of space-time variant parameters in endurance sports training. Proceedings of the 6th International Symposium on Computer Science in Sports (IACSS 2007).

[6] Netek, R. (2010) Implementation of geoinformatics technologies for mountain bike competition. Palacky University in Olomouc, Department of Geoinformatics, Olomouc.

[7] Yao, A., Uebersax, D., Gall, J. and Van Gool, L. (2008) Tracking People in Broadcast Sports. In: (ed.) M. Goesele et al.: 32nd Annual Symposium of the German Association for Pattern Recognition. Darmstadt, Germany, 151 – 161.

[8] Zentai, L. (2005) LBS in orienteering sport. In: (ed.) Georg Gartner: Location based services & telecartography Symposium, Technische Universität, Wien, ISSN 1811 8380, 249-252.

[9] Wikipedia (2011) GPS eXchange Format. Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/GPS_eXchange_Format, [cit. 2011/10/26].

# INTERACTIVE SAND-COVERED TERRAIN SURFACE MODEL WITH HAPTIC FEEDBACK

Vaclav PURCHART[1], Ivana KOLINGEROVA[2] and Bedrich BENES[3]

[1,2]Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitni 8, 306 14, Pilsen, Czech Republic

*vpurch@kiv.zcu.cz, kolinger@kiv.zcu.cz*

[3]Department of Computer Graphics Technology, Knoy Hall of Technology, 401. N.Grant Street, 47907-2021, West Lafayette, IN, U.S.A

*bbenes@purdue.edu*

**Abstract**

We describe a method for an intuitive haptic rendering and modeling of GIS terrain models represented as triangular irregular networks (TIN). The TIN terrain model does not allow representation of concavities and overhangs but it is sufficient for most GIS applications. Model editing is achieved by a set of virtual tools with various shapes which are controlled by a haptic device with force feedback. The haptic feedback makes manipulation with the terrain model much more intuitive than by the mouse, because it operates in 3D and it also provides better control with the force. The haptic device is point-based, so the forces provided by different tools must be mapped to a point. This, in effect, helps eliminating certain vibrations that could be caused by numerical errors and insufficient scene sampling. Our approach allows an adaptive adjustment of the level of detail of the model according to the terrain properties and user requirements. Haptic visualization in GIS models has been already used but such solutions are not common. Unlike existing approaches, our solution uses a general shape of the tool, it allows interactive editing of the TIN model, and it utilizes parallelism and coherence between consecutive states of the model that make haptic visualization faster and smoother.

**Keywords: geovisualization, haptic visualization, triangular irregular network terrain model, interactive manipulation**

## INTRODUCTION

Various approaches to exploration, visualization, and intuitive editing of GIS data have been developed last years, but there are still many open problems. One of them is the haptic visualization and editing of terrains. The haptic visualization has the potential to provide additional information to the user (sense of touch) in addition to visual output only. And, unlike the mouse or keyboard, the haptic devices provide native 3D manipulation, which is also more intuitive. Haptic devices were very expensive but, nowadays, they are becoming affordable for general public. New falcon devices, such as Falcon Novint are priced at $250 in December 2011.

The most commonly used terrain representation in existing haptic applications is digital elevation model (DEM) based on a regular grid or volume data. Volumetric-based or grid-based methods are memory consuming, because the resolution of the grid must correspond to the smallest detail of the the model. To the best of our knowledge, there is no suitable haptic method which works natively with TIN and allows both editing and simulation of physical processes on the terrain.

We have developed a haptic visualization method which works natively with TIN and allows interactive editing of the terrain and erosion simulation using a haptic device. User can imprint a tool shape into the terrain or he/she can drag a tool across the terrain. The physical model of the terrain represents sand but our approach is general enough to simulate different materials. All changes in the terrain are smoothened by thermal erosion which can be also controlled, e.g., allows varying the sand humidity. This limitation of our technique is that it does not allow shapes with overhangs. The goal of our haptic visualization is to be realistic, easy to use, and reasonably fast to be usable.

Our method uses the Delaunay triangulation (DT) [1] for the TIN representation and the Constrained Delaunay triangulation (CDT) for editing [2]. We have created various editing tool shapes, such as rectangle, circle, sphere, ellipsoid, and man foot print. As the haptic device provides point-based feedback, all forces exerted on our shapes must be mapped to a single point. Because of different frame rate requirements, our implementation uses multiple CPU threads; one for the graphics feedback and one for the haptic visualization.

The paper is structured as follows. The following section will briefly introduce the existing work. Next, our method is described. Next section summarizes our experiments and the last section concludes the paper.

## RELATED WORK

Haptic visualization methods of geospatial data are quite rare. The previous work is focused on haptic exploration and haptic visualization of volumetric data sets or point sets.

Faeth at al. present a framework for haptic visualization of geospatial data using X3D in [3], their framework also allows making deformation of the model. They use the ArcGIS program to convert various GIS formats into X3D indexed mesh and grid image. Vertices changed their heights after deformation and, additionally, they are rearranged according to Generalized Chainmail deformation algorithm [4]. It results in a cloth-like behavior of the terrain. "Soft" and "hard" materials are distinguished. Force response can map terrain properties.

Guo at al. [5] present a method for interactive haptic manipulation of point sets. User edits are described with implicit functions. The model consists of two representations. The first is a global scalar field which stores a distance from the zero plane. The other representation is a voxel regular grid. A global mass-spring system is created between voxels. This method allows simulation of voxel-based physics processes. Haptic visualization uses the exponential weighted average extrapolation scheme to smooth the resulting force.

Beneš at al. have described an algorithm for haptic visualization of the sand surface in [6]. The user can interactively edit the terrain using a haptic device and the terrain changes are

smoothened by erosion. The terrain model is described as a regular height-field. This is significantly better than voxel-based approaches in terms of memory requirements.

Grid-based methods have high memory requirements and have implicit limits on the details they can represent; TIN-based methods do not allow editing but only a deformation of existing terrain. No new vertices can be added to the TIN. Our method provides such features.

## PROPOSED METHOD

Our algorithm consists of three main parts: (i) the geometric model, (ii) physics simulation, and (iii) haptic interaction. The input of our algorithm is a set of 3D points [*x*, *y*, *z*], where *x* and *y* defines the position in terrain and *z* is an elevation of the terrain. The points are given by a height map, by an image, or a bivariate function. The output of this process is the TIN modeled as a Delaunay triangulation of the given set of points, which we use for physics simulation. DT is computed using the incremental insertion algorithm [1]. This algorithm is very useful for our purpose, because it can add vertices into the triangulation on the fly. So the terrain modification could be easily handled. Terrain could be modified by erosion or by haptic device. As a testing algorithm for physics simulation, we have developed an algorithm for thermal erosion in [7].

We use a set of virtual tools for the haptic manipulation. The tool consists of a set of control points and edges between these points. The edges are used as constraints for the CDT. The terrain is deformed inside the virtual tool shape according to this shape and according to the user requirements. The height of these points adapts to the terrain height in the current tool position when the tool is imprinted into the terrain, (see Fig. 1). The amount of material pushed away by the tool is defined as a difference between heights of the tool depth level, and height of the control points in the direction of tool movement.

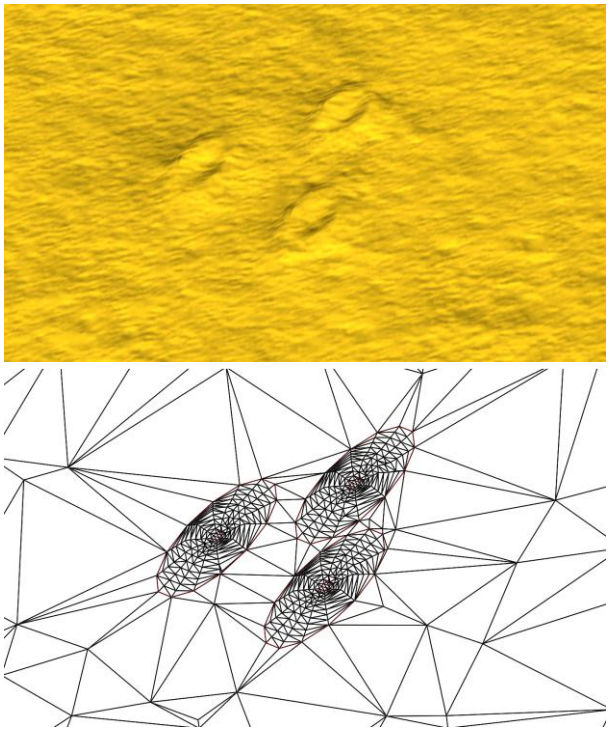


**Fig. 1.** Demonstration of terrain editing. Left: Three imprints of elliptic shape tool on a nearly flat terrain covered by yellow sand. Right: The same scene displayed as a wireframe model. Border edges of ellipses are constrained.

**Haptic Visualization**

The haptic manipulation is initiated by a haptic tool handled by the user and it causes geometric changes of the TIN. The erosion algorithm is applied to the changed vertices. The challenge is to maintain a correct haptic feedback when the TIN is changing its geometry even though the TIN is not in a stable state for a while.

The graphics visualization refresh rate is about 60Hz that is way below the required refresh rate for haptic rendering that is about 1,000Hz. The TIN is constantly changing because while the user edits it and the erosion smoothes it, the haptic thread must update its data only if the slow graphics and geometry thread reaches a stable state. The only stable state is the end of graphics thread loop (all terrain edits have been performed and the TIN has been repaired).

The data, the update of which is fast enough to be directly handled in the haptic thread, are the position of haptic cursor and the force feedback. The haptic device sends a 3D position of the cursor as a point [$x$, $y$, $z$] 1,000 times per second. The virtual tool is correctly updated based on this position. The tool movement (direction as well as its velocity) is acquired as a difference of two consecutive tool's positions. We must set the force feedback $F_{out}$ to the device (3D vector in newtons). This value must be also set 1,000 times per second. The device acts in the direction of force vector with the given force amount.

As it is mentioned above, visualization and haptic threads have different refresh rates. Usually it is 60Hz versus 1kHz. So the haptic thread makes $1{,}000/60 \approx 17$ iterations for each graphics render. We call this number of iterations $it_p$ and we assume it to be constant. Slow communication with the haptic device would be suboptimal – the haptic perception would change very slowly, only with 60Hz refresh rate instead of 1kHz. The force feedback feels runny for low framerate. The speedup in our approach is achieved by interpolation of the forces between two consecutive frames. Our algorithm stores the old force feedback state before an update and the new one. The force is interpolated in time in each haptic iteration, so that the resulting force feedback is smoother.

**Force Feedback Components**

The force feedback $F$ consists of two forces (see Fig. 2). The first one is a penetration force $F_p$ and it depends on the tool depth level – a deeper penetration results in a larger force which acts strictly in the vertical direction for sand and our set of tools. Sand settles down quickly because of erosion, so we can expect a nearly horizontal surface [6]. It is impossible to feel the difference between our simplified strictly vertical $F_p$ and the real, physically correct $F_p$.

The second part of the force feedback is the friction force $F_r$, which is physically defined by Equation 1. It depends also on the penetration depth in the direction of tool movement and on the speed of tool movement – a higher velocity means a larger force which acts against the tool movement direction.

$$F_r = \mu \cdot F_n, \qquad (1)$$

where $\mu$ is the coefficient of friction, which is an empirical property which differs for each material, and $F_n$ is the normal force which is perpendicular to the surface and acts against the tool. In our method we use an approximation of this physical law as explained below.

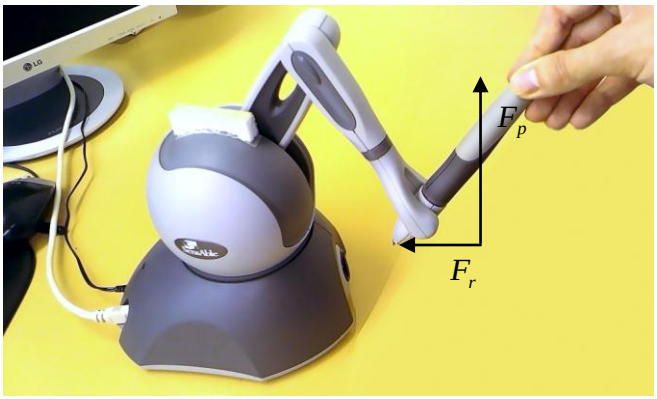



**Fig. 2.** Low end haptic device (Phantom Omni) used for the test.

The force feedback is a bit more complicated as the user editing tools can have an arbitrary shape rather than a single point. We have created tools with various shapes, such as a rectangle, a circle, a sphere, an ellipsoid, and a human footprint, but the haptic cursor is a single point for the haptic device we use. We need to map all the forces on the circumference

of the tool into the single point. This can be roughly computed as a mean value of forces in all control points of the tool (Equation 2). Such approach is not physically accurate for non-circular tools, but it gives a good force feedback for tools we have tested.

$$F = \frac{\sum_{i=1}^{N}(F_{ri}+F_{pi})}{N} + F_G, \quad (2)$$

where $F_{ri}$ is the friction force for the i-th control point, $F_{pi}$ is the penetration force for the i-th control point, $N$ is the number of control points, and $F_G$ is the force caused by the so called gel effect.

The force $F_G$ is an additional horizontal force. It is a haptic effect [8] known as a “gel effect”. This force depends on the velocity of the tool and acts as a tangent viscosity; higher velocity causes that a larger force acts against the moving direction. This force very well describes behavior of the real sand and simulates a sand resistance against the fast movement.

**Force Computation**

Let $z_{tool}$ be a current tool depth level, $c_i$ is a position of *i-th* control point of the tool, $z_{ci}$ is a $z$ coordinate of *i-th* tool control point, $P_{act}$ is a current position of the tool, $P_{old}$ is a previous tool position, and $T$ is a vertical force tolerance. If the tool is in vicinity of the terrain surface (tool's distance from the surface is smaller than $T$) the force starts to act.

The height difference for *i-th* control point is computed by Equation 3. Note that the points coordinates have to be normalized into the interval [0; 1].

$$\Delta h_i = z_{tool} - z_{ci}. \quad (3)$$

The force feedback starts to act only when the inner part of the tool (a tool depth level) is below the terrain level or slightly above it to avoid numerical errors ($\Delta h_i > -T$). The direction $dir_i$ of the tool movement from the tool position to the *i-th* control point is described by Equation 4:

$$dir_i = \frac{c_i - P_{act}}{|c_i - P_{act}|}. \quad (4)$$

The friction force is computed by Equation 5:

$$F_{ri} = -dir_i \cdot \Delta h_i \cdot \frac{F_M}{2}, \quad (5)$$

where $F_M$ is the maximum force allowed by the haptic device (10N in our case).

The penetration force for the i-th control point is computed by Equation 6:

$$F_{pi} = \Delta h_i \cdot \frac{F_M}{2} \quad (6)$$

To obtain the "gel effect" force, we compute a tool velocity $v$ between frames by Equation 7. Note that we expect a constant time interval between frames (equal to one).

$$v = P_{old} - P_{act}. \quad (7)$$

The force for the "gel effect" is then defined by Equation 8. Zero *z* coordinate restricts the $F_G$ to the strictly horizontal direction.

$$F_G = [v_x, v_y, 0] \cdot F_M \quad (8)$$

The final force feedback $F_{out}$, which is sent to the device, is interpolated between two consecutive frames in order to smooth any abrupt changes during haptic frames between the geometry updates. In the haptic thread, there are two counters: the total frame count in the previous haptic rendering sequence $it_p$ and the frame count rendered since the last geometry update *it*. The resulting force is computed from Equation 9:

$$F_{out} = F_{old} + \frac{F - F_{old}}{it_p} \cdot it, \quad (9)$$

where *F* is the force from Equation 2 and $F_{old}$ is the force *F* computed before the last geometry update.

## EXPERIMENTS AND RESULTS

We have implemented our algorithm in C++ and used OpenGL library, GLSL, and OpenHaptics® library [9] for haptics rendering. All data were measured on a standard desktop computer running Windows 7, with 12GB RAM, Intel Core i7 clocked at 1.67GHz, and NVIDIA GeForce GTX 295, with 1.5GB of memory.

We have experimentally verified our algorithm by drawing into real box with two types of real sand and comparing visually the results as shown in Fig. 3a. So far, the haptic feedback has been tested in the scope of our working group.

As a demonstration of haptic manipulation we have simulated the same effect using our system (see Fig. 3b). Force feedback is smooth and it is almost the same as the real sand. The penetration force is slightly increased with increasing depth of the tool. The user must make a significant effort to push the tool to the sides if its depth is high. If the tool rests, the force components are mutually balanced.

(a) (b)

Fig. 3. The word GIS written in real sand (a); simulated using our system (b).

## CONCLUSIONS

The force feedback in our model is not physically accurate because we have a haptic device with only a point cursor, but the touch sense feels very much like a real sand as we verified in the experiment. As haptic device is nowadays rather cheap and available, the described approach can be useful for the reader.

There are many possible ways how the algorithm can be extended. One apparent way is by using a different type of triangulation.

Another possible avenue for future work is to perform a user study to measure the accuracy of force feedback or to extend the haptic visualization with materials with different granularity.

## ACKNOWLEDGEMENTS

We would like to thank to Petr Vaněček (University of West Bohemia, Plzen) for his valuable comments and Jan Flasar (Masaryk University, Brno) for his help with introduction to haptic.

This work has been supported by the Czech Science Foundation under the research project project 201/09/0097, and project SGS-2010-028 and by NSF IIS-0964302, NSF OCI-0753116 Integrating Behavioral, Geometrical and Graphical Modeling to Simulate and Visualize Urban Areas.

## REFERENCES

[1] Berg, de M., Kreveld, van M., Overmans, M. and Schwarzkopf, O. (1997) Computational Geometry: Algorithms and Applications. Springer-Verlag.

[2] Sloan, S. W. (1993) Fast Algorithm for Generating Constrained Delaunay Triangulations. Computers and Structures. Vol. 47, no. 3, 441-450.

[3] Faeth, A., Oren, M. and Harding C. (2008) Combining 3-D Geovisualization with Force Feedback Driven User Interaction. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (GIS '08). ACM, New York, USA, Article 25, 9 pages.

[4] Li, Y. and Brodlie, K. (2003) Soft Object Modeling With Generalized Chainmail, Computer Graphics Forum, 22(4): 717–7.

[5] Guo, X., Hua J. and Qin, H. (2004) Touch-Based Haptics for Interactive Editing on Point Set Surfaces. IEEE Computer Graphics Applications, 31-39.

[6] Benes, B., Dorjgotov, E., Arns, L. and Bertoline, G. (2006) Granular Material Interactive Manipulation: Touching Sand with Haptic Feedback. Proceedings of the 14-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 295-304.

[7] Purchart, V., Kolingerova, I. and Benes, B. (2011) Interactive Manipulation of an Adaptive Triangulation-based Sand covered Terrain. The Visual Computer.

[8] Lin, M. C. and Otaduy, M. (2008) Haptic Rendering: Foundations, Algorithms and Applications. A. K. Peters, Ltd., Natick, MA, USA.

[9] Sensable (2011) OpenHaptics Toolkit . http://www.sensable.com/products-openhaptics-toolkit.htm, [cit. 2011-12-8].

# SENSITIVITY ANALYSIS OF ANALYSIS OF VISIBILITY FROM LINE

Katerina RUZICKOVA, Martina DOHNALOVA and Jan RUZICKA

Institute of Geoinformatics, VSB-Technical University of Ostrava,
17. listopadu 15, 708 33, Ostrava, Czech republic

*katerina.ruzickova@vsb.cz, martina.dohnalova.st1@vsb.cz, jan.ruzicka@vsb.cz*

**Abstract**

An analysis of visibility is one of the basic analyses with digital surface models. Different programs solve this task in a different way, with a different algorithm. For example ArcGIS offers to find visible areas from a point or from a line feature. The base algorithm calculates visible places from one point. In a case of counting visible areas from more points, the merge of visible areas for individual points is performed. A similar method is applied for evaluation of visible areas from a line feature. There are provided analyses of visibility from all line nodes and afterwards the merge of results into one layer. The result of analysis of visibility from a line is responsive to a line geometry. A placement and a number of line nodes are important and they can influence analysis result.

The goal of this paper is to present an influence of a line geometry to a result of visibility analysis. The sensitivity analysis was executed by many analyses of visibility from a different line geometry and for a different resolution of a digital surface model also. It was realized for ArcGIS and for GRASS GIS. This process was automatized. The results were compared and analysed.

**Keywords: sensitivity analysis, analysis of visibility, DEM**

## INTRODUCTION

Analysis of visibility is one of the basic analyses with digital surface models. It is possible to find it in a whole series of GIS software. Solving of the accuracy and fastness of viewshed computation is the focus of research for many years. Last years is often analysed the uncertainty associated to the output GIS-based models. Next important aspect of spatial modelling is sensitivity analysis. According to Satelli the sensitivity analysis is the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input. It could answer the questions: “Which of the uncertain input factor(s) is so non-influential that we can safely fix it/them?” or “If we could eliminate the uncertainty in one of the input factors, which factor should we choose to reduce the most the variance of the output?” [1].

### Viewshed Analysis

Sense of viewshed analysis is to determinate visible area from the observer site. The most of GIS software offer a tool to calculate visibility from line. It is extension of basic function calculating visible places from one point. In this case the merge of visible areas for individual points – vertices of line is performed (**Fig. 1**).

The quality of viewshed result is determined by accuracy and complexity of input data and the quality of viewshed algorithm. The input data is DEM (Digital Elevation Model), observer site and parameters of visibility (e.g. the direction of the viewshed and the search distance).

Research coped with computing viewshed following:

- Viewshed algorithm improvement and development.
- The quality of the objects representation depends on the resolution of sampled data and its influence to viewshed analysis. In this context the terrain complexity is solved.
- Computational time – better objects representation means more computational time and more memory.

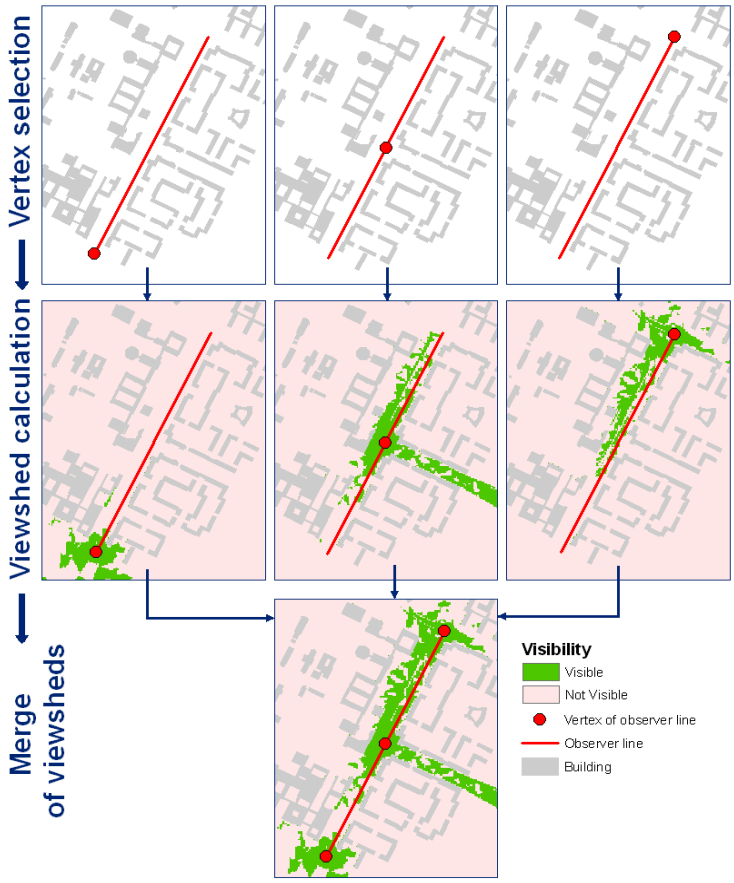



**Fig. 1.** Principe of visibility calculation form line

Viewshed could be managed over the Triangulated Irregular Networks (TIN) or Regular Square Grids (RSG) terrain. The algorithms of viewshed for TIN and RSG are different. For TIN is usually implemented by computing the visible parts of triangle element and the method of building line-of-site profiles is usually used for RSG. Methods based on RSG are widely expansed and used, because data structure is simpler and more convenient to manipulate.

These methods are preferred in GIS, they are used for the research described in the paper. There are many algorithms, based on this principle, such as Janus, Dynatacs, Modsaf and Brasenham [2]. The differences between them mainly lie in the different methods of DEM interpolation a and the principle of visible judgement. The different approach is based on Reference Planes (RP) [3], [4]. An overview and comparison of the algorithms were published in [2], [5], [6]. There are mainly described Xdraw, R2, R3, WS and RP algorithms. The most accurate solution is obtained by the exact algorithm R3, other approximative algorithms work with relatively small error [6]. These algorithms are consecutively specified and improved [2], [4]. Recent developers use also different approaches like visibility graph theory, statistical sampling, reverse viewshed analysis and least-cost computation methods for determining optimal paths [7].

Unfortunately the specification of concrete visibility algorithms in GIS software is often part of producer's intellectual property and it is not released. For example ESRI Inc. does not publish viewshed algorithm of ArcGIS. But there are also open source software like GRASS, which viewshed tool was written by Khawaja, K.O. and Neteler, M. GRASS offers the tool for viewshed from point only.

The quality of the input data representation (depends on the resolution) is important. The influence of terrain resolution and complexity was solved. The first experiments was done by Fisher in paper Propagating effects of database generalization on the viewshed. In this paper the effects of generalizing a Digital Elevation Model on the area which is determined to be visible is examined [5]. There was demonstrated, that various conversions of DEM yield variances in final raster of visibility. The accuracy of viewshed techniques proved Maloy et al. [8]. They compared actual viewshed to predicted viewshed generated using various commonly used data elevation data sets, various data models and various visibility criteria. It was found that some data sets produced predicted viewsheds that mimicked surveyed viewshed much more accurately than other data sets, and that some combinations of data models and visibility criteria produced much greater matching accuracies than other combinations. The influence of a terrain generalisation on a viewshed generation examined also Nielsen and Costopoulos [9]. They compared Idrisi, ArcMap and GRASS. Unfortunately their method evaluated differences between ArcMap-Idrisi and Idrisi-GRASS, but not relation between ArcMap-GRASS. But the conclusion presents, that GRASS-Idrisi viewsheds are in close agreements, and some of ArcMap-Idrisi comparisons showed considerable differences especially in high elevation range. The Idrisi viewsheds were generally larger. The viewsheds disagree mostly where the slope is at a threshold, such as at ridgelines and other edges. Next example of viewshed analysis techniques optimisation was published by Kim et al. [7]. They used two strategies. The first was to restrict the search to key topographic points in the landscape such as peaks, pits and passes. The second was to use heuristic which have been applied to other maximal coverage spatial problems such as location–allocation. The results show that the use of these two strategies results in a reduction of the computing time necessary by two orders of magnitude, but at the cost of a loss of 10% in the area viewed.

Because of more time exacting execution the parallel processing is suitable for viewshed analyse for RSG. Teng et al. [10] presents solution, in which every line-of-sight within terrain was processed using parallel solution.

**Sensitivity Analysis**

The Sensitivity Analysis (SA) is the study of how the variation (uncertainty) in the output of a statistical model can be attributed to different variations in the inputs of the model. It is a technique for systematically changing variables in a model to determine the effects of such changes [11]. There are different ways of classifying sensitivity analysis methods. For example, these methods may be broadly classified as mathematical, statistical (or probabilistic), and graphical [12].

Properties of an ideal global sensitivity analysis method are [13]:

- Cope with the influence of scale and shape.
- Include multidimensional averaging.
- Be model independent.
- Be able to treat grouped factors as if they were single factors.

Fulfilment of these properties is showed in **Tab. 1.**

**Table 1.** Properties of sensitivity measurements [13]

| | **Scale and shape** | **Multidimensional averaging** | **Model independent** | **Grouping of factors** |
|---|---|---|---|---|
| Derivates, Local methods | N | N | N | Y |
| Regression method, e.g. Standardised regression coefficients | Y | Y | N | N |
| Morris | N/Y | Y | Y | Y |
| Variance based methods | Y | Y | Y | Y |
| Monte Carlo Filtering | Y | Y | Y | N |

There are many different ways to perform sensitivity analyses. Basic sensitivity techniques presents for example [12], [13], [14], [15]. According this papers the most fundamental of sensitivity techniques is the direct method – calculating partial derivates of the model. But the using is limited and the output function must be known. Without knowing this function the standard regression analysis could be the most comprehensive technique and is relatively easy to perform with available software. Correlation, rank correlation, and partial rank correlation can be calculated to determine relationship between the independent and dependent variables.

The Morris method is a one-factor-at-a-time method using randomized sampling matrices, which allow direct observation of elementary effects [15].

Variance based methods such as “importance measure” or “correlation ratio”, are capable of estimating the main effect contribution of each parameter to the output variance. Next based methods are Fourier Amplitude Sensitivity Test (FAST) and Sobol’s methods . FAST can not only measure the main effect, they can also compute the so-called “Total Sensitivity Indices” (TSIs). The TSIs measure the main (first order) effect of each individual or group of inputs on

the model output, as well as all higher order effects that can be attributed to that parameter. Sobol's methods is similar to FAST, but uses a unique decomposition of the model into summations of increasing dimensionality [15].

Monte Carlo Filtering is the process of rejecting sets of model simulations that fail to meet prespecified criteria of model performance.

Using graphical methods for a sensitivity analysis could be also appropriately. Graphs, charts, or surfaces are used to give a visual indication of how an output is affected by various inputs.

The selection of the sensitivity technique depends on the used model, reasons for SA including the needs to determine, problem complexity and users experience. [13] aims at guiding a non-expert user in the choice of the method to adopt for the user own problem. The methods recommended include the variance based methods, the method of Morris, and Monte Carlo filtering, e.g. some effective methods for global sensitivity analysis.

Interesting application of UA and SA was created by Crosetto and Tarantola [16]. They developed the tool (for hydrology model), witch implements UA and SA procedures for input data optimization. There were used FAST (Fourier Amplitude Sensitivity Testing) and Morris techniques to cope with computationally expensive models numerous input factors.

**Problem Presentation**

Intent of the paper is visibility analysis from line feature. Complexity of line feature is done by its curvature. In the parts of rugged feature the line are present by higher number of vertices. There could be less vertices for straight feature for the same data resolution. For example the line for highways will be probably described with less vertices than line for winding stream. The main highway D1 in the Czech Republic (between towns Prague and Brno) has the average line segment length 6.3 km (dataset ArcČR500, 1:500 000).

The Principe of viewshed calculation from line (presented on **Fig.1**) in ArcGIS is that the nodes and vertices of lines are used as observation points. Final output is the merge of viewsheds from these observation points. Therefore the viewshed output depends on the density and placement of line vertexes. The influence of this method is presented on **Fig.2**. There was used digital surface model including the buildings. It is not the typical using of the viewshed tool. The buildings were included in to highlight the effect of terrain barriers and consequently to highlight the influence of line vertices density and placement on visibility output. There is obvious that a higher density of vertices causes larger visibility areas. The difference between visibility from 5 and 6 vertices shows, that the placement of vertices is important as well.

The goal of the study is the exam how the density of vertices on line feature can induce changes in results of viewshed analysis. The default hypothesis is increasing visible area with higher density of vertices on the line. The size of the change has been researched. It will be possible to generalize the observer line, in a case of small influence. It could be prosperous in a case of arduous visibility processing over the large areas. If the change of density of vertices yields significant change of visible area, the observer line resolution must be prepared carefully.

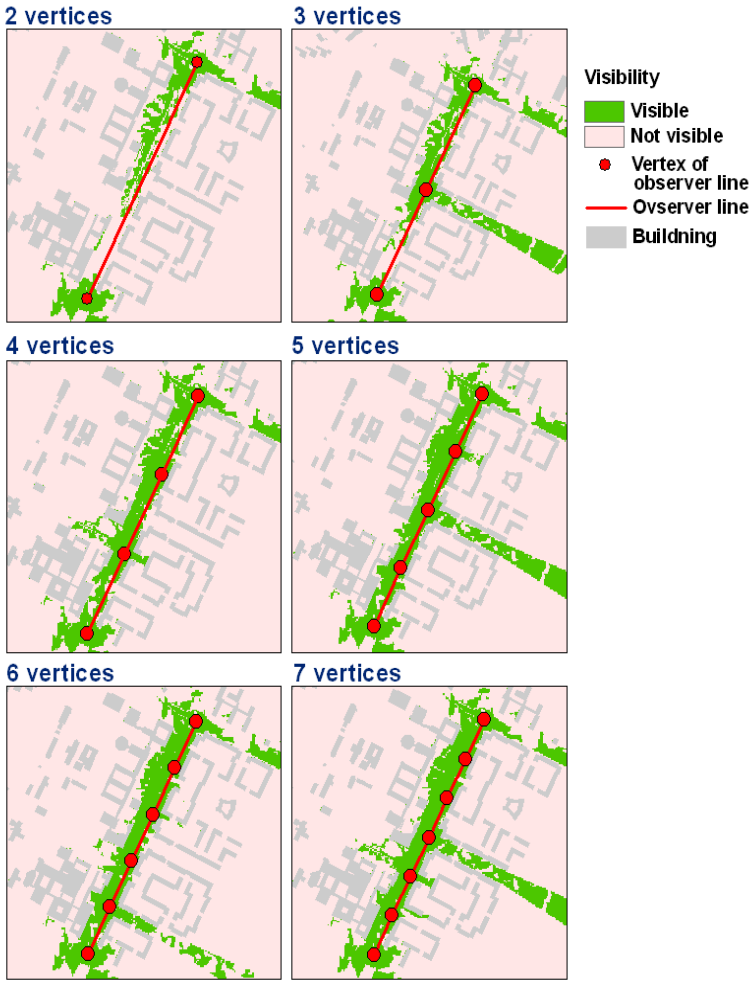



**Fig. 2.** Influence of a placement and a number of line nodes on viewshed result

## DATA AND METHODS

There was used ASTER GDEM terrain [17]. The tile N49E017 was selected. It presents part of Moravian-Silesian Region (in the Czech Republic). The planar resolution is approximately 30m and the height resolution 20m. The data are distributed in WGS84 / EGM96. Afterwards the data acquisition it was transformed into JTSK (EPSG:2065). After transformation the raster size was 3175x4514 pixels and resolution 26m. Heights range is 88-1163m, mean height 625m and mean slope 7.8°.

The observer line was manually digitalised over the terrain. The vertex distances are approximately 3km.

The total length of line was 16.5km.

The viewshed analysis were executed in GRASS 6.0 and ArcGIS 10. For statistical evaluation STATGRAPHIC Plus 5.0 was utilized.

The observer line was split in several distances. The distance was changed from 1000 to 50m stepping by 50m. Densify function in ArcGIS was used for line splitting at the defined distance. The module v.split was used in GRASS. The representation of digital terrain was generalized from 50 to 1000m stepping by 50m. There was selected bilinear interpolation. ArcGIS has the tool Viewshed, GRASS performs visibility analysis via r.los module. There was module r.viewshed in older versions of GRASS, but it is not included in the new versions of the software. The using of this module is recommended for DEM size under 1000x1000 pixels. Processing larger DEMs is time consuming.

For both used software the script was developed. Python was used for ArcGIS and Bash for GRASS. It takes the viewsheds as it is determined in changing terrain resolution and changing vertices spacing observer line. There was used only one limit. The search distance was set up on 20km. Small differences between the scripts were caused by different input and output software requests. GRASS calculate the viewshed only form points, so the effect of viewsheds for line vertices merging was append. The script create output file with overview of computed set-ups. For each viewshed was assigned: used terrain resolution, used vertices spacing, number of vertices, number of visible pixels, visible area, computing time.

For final comparison with derived viewsheds at multiple levels of terrain generalization and vertices spacing were evaluated. Suitable methods for SA was consulted with specialist. Unavailable output function (mathematical model description) excluded the using of direct method – calculating partial derivations of the model. The standard regression analysis and variability parameters were advised.

## RESULTS

Variability of viewsheds was evaluated separately for GRASS and ArcGIS (**Tab. 2.**). All obtained results of viewshed were analysed. There was not distinguished effect of single inputs.

**Table 2.** Variability of computed visible area

| | Min. visible area [km2] | Max. visible area [km2] | Mean visible area [km2] | Stnd. skewness | Stnd. kurtosis | Variability [E10km2] |
|---|---|---|---|---|---|---|
| GRASS | 98.543 | 944.000 | 633.739 | -4.676 | -1.812 | 3.958 |
| ArcGIS | 77.908 | 794.000 | 497.289 | -3.948 | -2.451 | 2.987 |

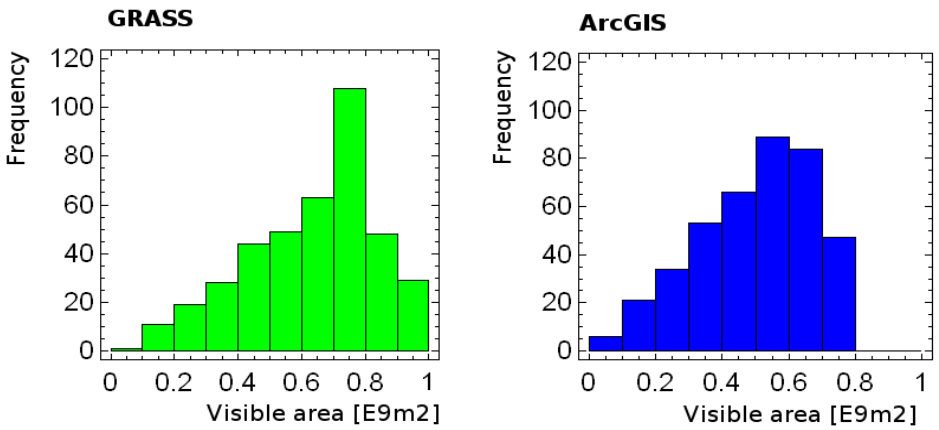



Fig. 3. Histograms of Visible area

The values of standardised skewness and standardised Kurtosis in **Tab.2.** and histograms on **Fig. 3.** show, that distributions of results do not fit normal distribution. It is the reason why we could not use evaluation by using standard deviations. Also commonly used coefficient of variation is evaluated due standard deviations. The values of variability in **Tab.2.** show, that GRASS embody higher elevation of variability. Next mentioned differences between calculated visibility area in ArcGIS and GRASS are presented by minimal, maximal and mean visible area (in **Tab.2.**). It could be said, that GRASS in general reach larger visible area then ArcGIS.

Variability invoked by changing observer line vertex spacing is summarised on **Fig. 4**. There were separately qualified the variabilities for fixed DEM resolution. Variability of visibility area expansion in GRASS predicates about faintly reducing effect of changing observer line vertex spacing while downsizing DEM resolution. This trend is not perceptible in ArcGIS.

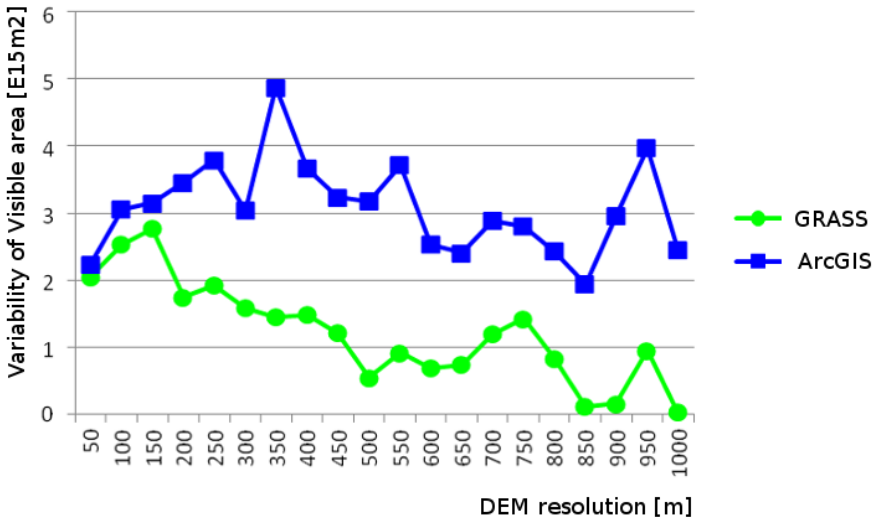



**Fig. 4.** Variability of Visible area owing to change of vertex spacing

At using the highest DEM resolution 1000x1000km GRASS provides almost the same visible areas for different vertex spacing in observer line (variability tent to 0). This fact is also perceptible on **Fig. 5.** The influence of vertex spacing on the size of visible area (for selected fixed DEM resolution) is plotted there.

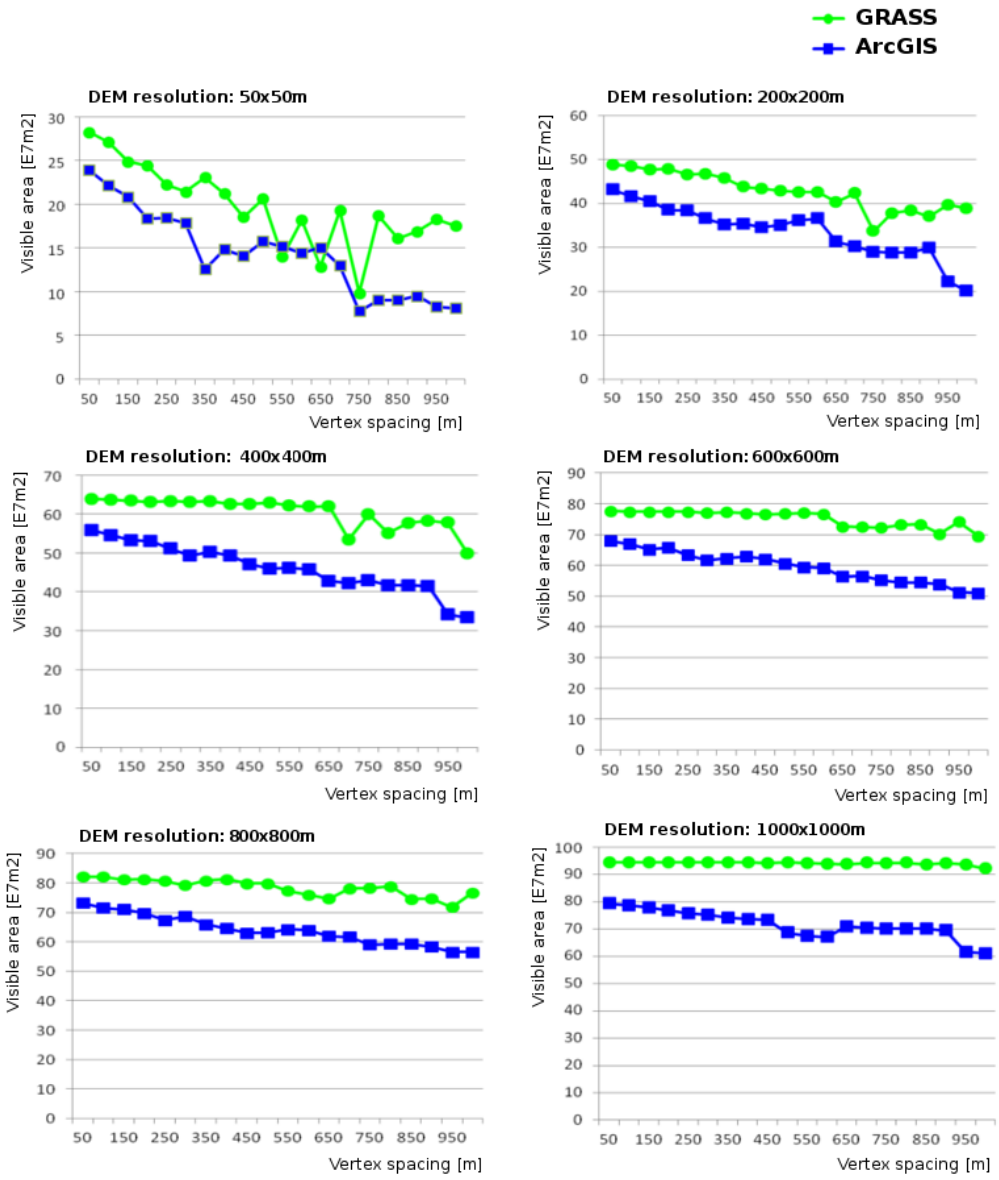



**Fig. 5.** The influence of vertex spacing – for fixed DEM resolution

Next the relationship between watched inputs a output was evaluated by regression analysis. The linear regression model was tested to describe this relationship between dependent variable (visible area) and independent variables (DEM resolution and vertex spacing).

**Table 3.** Relationship DEM resolution - visible area and vertex spacing - visible area

| | | **Correlation coefficient [-]** | **Coefficient of determination [%]** | **P-value** |
|---|---|---|---|---|
| GRASS | DEM resolution - visible area | 0.958 | 91.795 | 0.000 |
| GRASS | Vertex spacing - visible area | -0.134 | 1,797 | 0.007 |
| ArcGIS | DEM resolution - visible area | 0.922 | 85.015 | 0.000 |
| ArcGIS | Vertex spacing - visible area | -0.306 | 9,370 | 0.000 |

According to correlation coefficient values (**Tab. 3.**) the impact of DEM resolution is significantly stronger. Signs of values confirm the hypothesis that DEM resolution influences visible area directly, vertices spacing of observer line influences it indirectly. It means, that visible area is larger with bigger pixel size of DEM. The graphs on **Fig. 6.** affirm it. On the other hand longer distance between vertices of observer line produces smaller visible area (**Fig. 5.**). The coefficient of determination says, that approximately 92% of variation of visible area value causes variation of DEM resolution in GRASS. In ArcGIS it is 85%. Impact of vertex spacing is 1.8% (GRASS) and 9.4% (ArcGIS) inducted variety of visible area. But due to small P-values, we can not omit the influence of vertex spacing of observer line. The influence of observer line vertex spacing seems to be greater in ArcGIS then in GRASS.

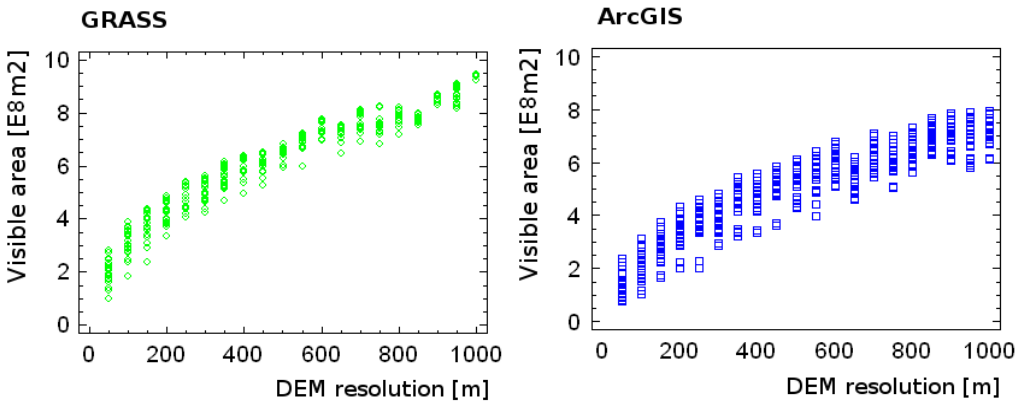



**Fig. 6.** The influence of DEM resolution – aggregated

There was willingness to deduce some evaluation predicate about witch combination inputs settings leads to similar results. But there are not any clusters in values. Therefore there is only plotted inverse relation between calculated visible area and observer line vertex spacing (and DEM resolution) on **Fig. 7.** There can be only find out which DEM resolution leads to larger or smaller visible area. The influence of observer line vertex spacing is not distinctive very well. We just say, that in ArcGIS are the smallest visible areas generated upon the

smallest DEM pixel size and the largest distance of vertex spacing in observer line. The biggest visible areas are as a result of the biggest DEM pixel size and the smallest observer line vertex spacing. In GRASS is the same trend, but it is not so obvious.

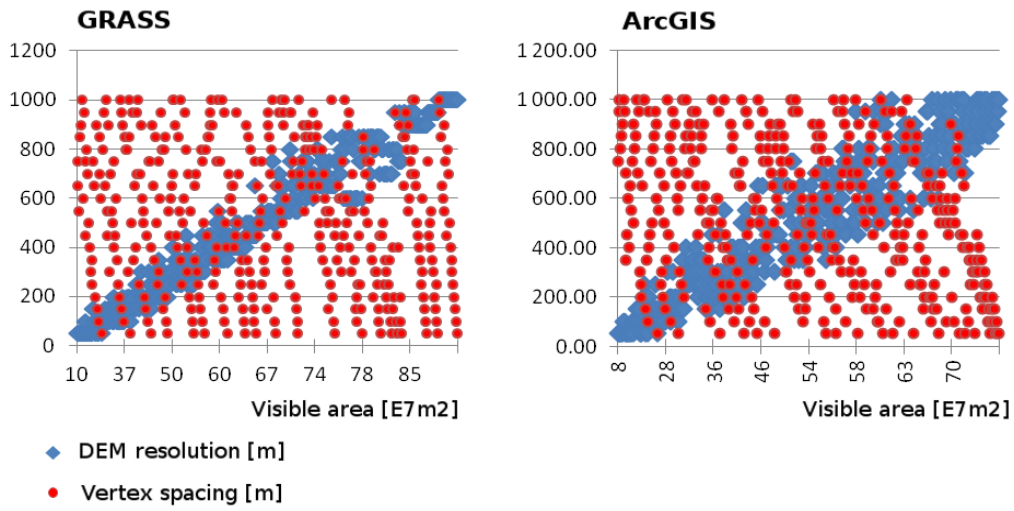



**Fig. 7.** The influence of DEM resolution

Finally the results of viewsheds could be used also for comparing the behaviour of viewshed calculating in GRASS and ArcGIS. GRASS produce more extensive visible area and it is more sensitive to changing watched input parameters in general. But ArcGIS is more sensitive to changing especially observer line vertex spacing. Next software confrontation was executed via scatter plots. Relationship between difference of visible area (difference between GRASS and ArcGIS results yield for the same conditions) and vertex spacing or DEM resolution were plotted (**Fig. 8.**). The graphs show enlarging differences with growing vertex spacing and DEM pixel size. It agree with progression of values of visible area on **Fig. 5.** and with progression of variability values of visible area on **Fig. 4.**

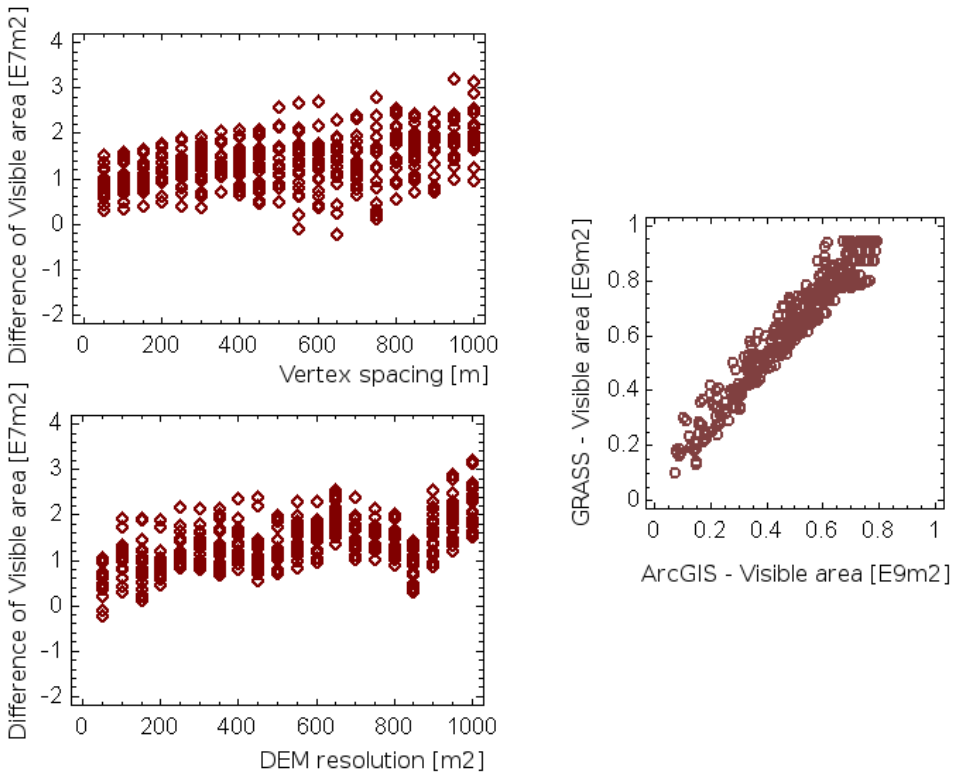



**Fig. 8.** Scatter plots, Difference of Visible area between GRASS and ArcGIS

## CONCLUSIONS

We preliminarily affirm the hypothesis that shorter distance between vertices of observer line leads to larger visible area produced by viewshed analysis. This effect is more significant with calculating over more detailed DEM in GRASS. On the other hand via to regress analysis the influence of this parameter is not so expressive like influence of DEM resolution. In GRASS the change of observer line vertex spacing yields only 1.8% of change of visible area. In ArcGIS is 9.4% participation on the viewshed result changing. Therefore the observer line does not need to be segmented with hight frequency, especially when processing in GRASS.

Unfortunately these conclusions result from one observer line and DEM only. They were not validated various lines and DEMs. It is subject of the ongoing researching. Afterwards we can finalize more independance evaluation.

Processing in two software allowed also their comparing. ArcGIS is more sensitive to changing observer line vertex spacing. But GRASS is more sensitive to watched input parameters et al.l and produces larger visible area. We can also acknowledge than processing visibility analysis over to large DEMs in GRASS is very time consuming.

**REFERENCES**

[1] Saltelli, A. (2005) Global Sensitivity Analysis: An Introduction. 2005. Online http://sensitivity-analysis.jrc.ec.europa.eu/tutorial/Saltelli_tutorial.pdf. [cit. 2011-08].

[2] Xu, Z-Y. and Yao, Q. A (2009) Novel Algorithm for Viewshed Based on Digital Elevation Model. Asia-Pacific Conference on Information Processing, Shenzhen, China, 18-19 July, IEEE Computer Society Press, Los Alamitos, 294 – 297. ISBN 978-0-7695-3699-6.

[3] Wang, J., Robinson, G. J. and White, K. (2000) Generating Viewsheds Without Using Sightlines. Photogrammetric Engineering & Remote Sensing, 66, 87-90. ISSN 0099-1112.

[4] Zhi, Y., Wu, L. and Cai, H. (2011) An improved algorithm for computing viewshed based on reference planes. 19th International Conference on Geoinformatics, Shanghai, China, 24-26 June, IEEE Computer Society Press, 1799 – 1804. ISBN 978-1-61284-849-5.

[5] Fisher, P. (1996) Propagating effects of database generalization on the viewshed. Transactions in GIS, 1, 69-81.

[6] Kaucic, B. and Zalik, B. (2002) Comparison of viewshed algorithms on regular spaced points. Proceeding SCCG '02 Proceedings of the 18th spring conference on Computer graphics ACM, New York, 2002, ACM Press. ISBN 1-58113-608-0.

[7] Kim, Y-H., Rana, S. and Wise, S. (2004) Exploring multiple viewshed analysis using terrain features and optimisation techniques. Computer & Geosciences, 30, 1019-1032.

[8] Maloy, A. M. and Dean, D. J. (2001) An Accuracy Assessment of Various GIS-Based Viewshed Delineation Techniques. Photgrammetric Engineering & Remote Sensing, 67, 1293-1298.

[9] Nielsen, C. and Costopoulos, A. (2005) The impact of terrain severity on variation in viewshed generation. Comparing Idrisi, ArcMap and GRASS. Archaeological Computing Newsletter, 62, 17-26. ISBN 88-7814-491-6.

[10] Teng, Y. A., Dementhon, D. and Davis, L.S. (1993) Region-to-region visibility analysis using data parallel machines. Concurrency: Practice and Experience, 5, 379–406.

[11] Wikipedia (2011) Sensitivity analysis. http://en.wikipedia.org/wiki/Sensitivity_analysis, [cit 2011-08].

[12] Frey. H.C. and Patil, S. R. (2002) Identification and review of sensitivity analysis methods. Risk Analysis, 22, 553-578.

[13] Joint Research Centre of Ispra (2011) Tutorial on Sensitivity Analysis. http://sensitivity-analysis.jrc.ec.europa.eu/, [cit. 2011-09].

[14] Hamby, D. M. (1995) A comparison of sensitivity analysis techniques. Health Physics, 68, 195-204.

[15] Ravalico, J. K., Maier, H. R., Dandy, G. C., Norton, J. P. and Croke, B. F. W. (2005) A comparison of sensitivity Analysis Techniques for Complex Models for Environmental Management. Internetional Congress on Modelling and Simulation, Melbourne, 12-15 December, Modelling and Simulation Society of Australia and New Zealand, 2533-2539. ISBN 0-9758400-0-2.

[16] Crosentto, M. and Tarantola, S. (2001) Uncertainty and sensitivity analysis: tools for GIS model implementation. Geographical information science, 5, 415-437.

[17] Earth Remote Sensing Data Analysis Center (2011) ASTER GDEM. http://www.gdem.aster.ersdac.or.jp/, [cit. 2011-09].

# CLUSTERING APPROACHES FOR HYDROGRAPHIC GENERALIZATION

Alper SEN and Turkay GOKGOZ

Geomatics Engineering Department, Civil Engineering Faculty, Yildiz Technical University, Davutpasa Street, 34220, Esenler, Istanbul, Turkey

*{alpersen,* gokgoz*}@yildiz.edu.tr*

**Abstract**

In many GIS applications users need to visualize and inspect data in different scales. This case requires that the different representations are stored at different levels of detail. One possibility is to store maps from the representation levels in a multiple representation database. Generalization methods are used to maintain and update the spatial databases.

Hydrographic data at different resolutions or scales are needed in various spatial studies. This paper presents a generalization operation, selection/elimination, for river networks based on clustering methods, namely k-means and Self Organizing Maps (SOM).

This is a model generalization operation, that considers geometric, topological and semantic river attributes as input. Clustering methods group all rivers into different categories according to similarities of the attributes, and then select the rivers for the reduced map scale based on these categories.

The clustering methods are compared using the data of United States Geological Survey (USGS) National Hydrography Dataset (NHD) at scales of 1:24000 and 1:100000. Clusters are eliminated based on the coefficient of line correspondence (CLC) is used to evaluate how well the features of clusters match the target 1:100000-scale. If the features of a cluster do not match any features of the target 1:100000-scale (CLC value is zero) or CLC value is different from zero, yet the selection of the cluster decreases the total CLC value of the derived network, then the cluster is eliminated.

The derived networks are visually compared to USGS 1:24000-scale and 1:100000-scale NHD. Töpfer's "Radical Law" is also used for a quantitative comparison. The case study applied to the network illustrates that both clustering approaches for hydrographic generalization have pleasing visual impact. However, SOM-based approach can be used as an effective method for the selection of rivers via data visualization and exploration for multi dimensional geospatial data.

**Keywords: hydrographic generalization, cluster, self organizing maps, k-means, river network**

## INTRODUCTION

By the development of the technology and science, substantial investments towards Spatial Data Infrastructure (SDI) projects are occurred at local, national, regional/multi-national and global levels. Although various definitions exist, the term Spatial Data Infrastructure (SDI) generally describes a framework of components that collectively create an environment in

which spatial data can be accessed and widely used; often it is the co-ordinating infrastructure underpinning the data assets of a country. The earliest and perhaps most prominent examples of SDIs are the US National Spatial Data Infrastructure (NSDI) in USA, The Authoritative Topographic-Cartographic Information System (ATKIS) in Germany and the Ordnance Survey Digital National Framework in Great Britain. The central task of National Mapping Agencies has been to establish and standardize digital geographic databases from which to produce maps. Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV) defines mapping agencies in a common and harmonized data model, called the AAA data model (AFIS-ALKIS-ATKIS Data Model) can generally be used for standardization of the spatial infrastructure.

The maintenance and updating of these databases has become an urgent problem for which there remains no uniform solution. Since problems in updating digital geographic databases have become a major impediment to the effective use of geographic data in production environments, multiple representation databases have become one of the most important researches. The different representations are stored at different levels of detail. Its flexibility lies in its ability to derive different types of maps from the representation levels of a multiple representation database, using generalization methods. In this respect, automated spatial data generalization techniques are very important, besides the methods such as data modeling, data management and data distribution [1], [2].

Generalization is a process used for reducing the volume of data of a spatial data set while preserving important structures [3]. Map generalization operations concerned with the abstraction of the database come under the heading of “model generalization”, whilst the set of operations concerned with the optimal visualization of the selected data are grouped under “cartographic generalization”. Model generalization is relevant to activities other than the visual. In particular it has relevance to data mining [4].

Model and cartographic generalization is used for deriving Digital Cartographic Model (DCM) from Digital Landscape Model (DLM). Both DLM and DCM are spatial databases in GIS. DLM is a generic term for a comprehensive description of the landscape, usually in the form of a topographic basic scale map. Databases derived from the primary DLM, through model generalization, are special-purpose secondary models of reality. Both primary and secondary DLMs may be used to create Digital Cartographic Model (DCM), through the process of cartographic generalization.

Model generalization is mainly a filtering process to obtain a subset of an original database for data analysis. Cartographic generalization is used for graphic display and it aims to improve the visual effectiveness and readability of a map. GIS need to perform model and cartographic generalization in order to satisfy both analysis and display purposes, respectively [5].

Selection and elimination operation has primary importance in map generalization. Competition for map space is a fundamental principle of map design. Unnecessary features are eliminated and important features are retained. The completeness of a map is affected by the elimination of features due to generalization procedures. [6] defines that the selection is a generalizing, but it is not part of cartographic generalization included simplification, classification and symbolization. [7] suggest a mathematical formula, the Principle of Selection

also known as Radical Law, is one of the well known selection method. From the study of atlases of different countries, the Principle of Selection proves to be a great potential value for the derivation of smaller scale maps from larger scale source material. The Principle of Selection is expressed by an equation that relates the number of occurrences of a particular feature at a source map scale and at a derived map scale. The principle can be expressed in its simplest form as

$$n_f = n_a \sqrt{M_a / M_f} \quad (1)$$

where $n_f$ is the number of objects that can be shown at the derived scale, $n_a$ is the number of objects shown on the source material, $M_a$ and $M_f$ are the scale denominators of the source and the derived map, respectively. The formula yields the number of symbols to be displayed, but it does not reveal which of the symbols should be chosen. Töpfer's law is the only quantitative rule in the selection and elimination of the features.

Stream order is a common technique to assign a hierarchy to the components of a river network is used for selection [8], [9], [10]. [11] uses the Radical Law and weighted values for selecting the cities on Israel maps. [12] applies the Radical Law to towns in England. [13] examines the selection of base information for thematic maps. [14] suggests a weighted network data for river generalization. [15] presents a method to select rivers by creating a fine hierarchy and filtering. [16] focuses on reduction of roads derived from axial maps. [17], [18] uses the building density as a threshold value for reducing the buildings in a block. The good continuation grouping principle derived from the Gestalt theory can serve as the basis for analyzing road and river networks into a set of linear elements called as strokes. [19] apply the perceptual grouping principle of good continuation to build strokes in road network selection. [20] also use the good continuation principle to build strokes in road and river networks in order to perform the selection. [21] and [22] introduce the centrality measures (i.e. degree, closeness, and betweenness) based on the connectivity graph for characterizing the structural properties of an urban street network, and for the selection of important streets using SOM algorithm. [23] present a method to select the river network by the watershed area threshold based on the watershed hierarchical partitioning. [24] uses Critical Path Method for selecting the rivers according to the weights. [25] focuses on model generalization and uses the principle of good continuation to enrich the database with river strokes. In Touya's study, in order to determine which strokes are to be selected, the main criterion used is a hierarchical organization of the strokes. [26] uses a pruning algorithm considering the upstream drainage area. [27] compare two pruning approaches which are stratified pruning, and the length and density pruning for hydrographic networks. [28] suggest a block-based selection method for road network generalization.

In this study, an approach for a model generalization of a river network at scale of 1:24000 is proposed in order to obtain the river network at scale of 1:100000. In this approach, geometric (length and sinuosity), topological (degree centrality, betweeness centrality and closeness centrality) and semantic (type, stream level, lake connection and flow accumulation) attributes are considered as inputs, and clustering methods, k-means which looks for a partition based on a given number of clusters (*k*) and an unsupervised learning method of artificial neural networks, self-organizing maps (SOM) are used. SOM is different

from the more popular approach k-means by reason of using topological relationships between clusters and geographic visualization environment. Some studies discusses about the relationship between GIScience and the SOM method. [29] suggests SOM for the typification of buildings. [30] edit a book about SOM applications in GIS.

Clustering methods group all rivers into different categories according to similarities of various attributes, and then selects rivers at reduced map scale based on these categories. The basic idea behind clustering is the attempt to organize objects into groupings based on certain shared characteristics. The clustering methods are compared using the data of United States Geological Survey (USGS) National Hydrography Dataset (NHD) at scales of 1:24000 and 1:100000. Clusters are eliminated based on the coefficient of line correspondence (CLC) is used to evaluate how well the features of the clusters match the target 1:100000-scale. If the features of a cluster do not match any features of the target 1:100000-scale (CLC value is zero) or CLC value is different from zero, yet the selection of the cluster decreases the total CLC value of the derived network, then the cluster is eliminated.

Finally, the derived networks are visually compared to USGS 1:24000-scale and 1:100000-scale NHD. Töpfer's "Radical Law" is also used for a quantitative comparison. The case study applied to the network illustrates that both clustering approaches for hydrographic generalization have pleasing visual impact. However, SOM-based approach can be used as an effective method for the selection of rivers via data visualization and exploration for multi dimensional geospatial data.

## METHODOLOGY

### Cluster Analyses

The complexity of the problem of generalization is that the knowledge is difficult to formalize, where objects, attributes, tools and objectives are of a multiple variety. There are complex relations between geometric, topological and semantic attributes of the features in the level of details and multidimensional spatial data space requires spatial data mining techniques to analyze.

Unlike classification and prediction, which analyze class-labeled data object, clustering analyzes data objects without consulting a known class label and it can be used for the selection process of model generalization. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such selection/elimination labels.

The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters.

Since clustering is the most frequent interpretation and implementation of the SOM method, it is useful to compare it to one of the most popular approach, k-means clustering which look for

a partition based on a given number of clusters (*k*) [32]. There is a disadvantage of using clustering approach for the selection, since clustering methods used are not deterministic.

**Self Organizing Maps (SOM) Method**

SOM is part of a large group of techniques known as artificial neural network (ANN). One quickly realizes that, apart from seeing the SOM only in the context of other ANN methods, depending on its purpose and training parameters one could also interpret it primarily as a clustering or dimensionality reduction technique. SOM adjusts weights of neuron vectors for minimizing the distance and maximizing the similarity to input vectors; thus the similar inputs are associated with closely positioned neurons. In SOM, the input vectors do not correspond to classes known a priori. Output nodes compete for the input vectors on the basis of certain similarity functions and the weights of winning nodes are adjusted according to the weights of respective input nodes. The basic idea behind clustering is the attempt to organize objects into groupings based on certain shared characteristics [30].

Input data as consisting of n-dimensional vectors:

$$x = [\varepsilon_1, ....., \varepsilon_n]^T \in R^n \quad (2)$$

Each of k neurons has an associated reference vector:

$$m_1 = [\mu_1, ....., \mu_n]^T \in R^n \quad (3)$$

During training, one $x$ is compared with all $m_i$ to find the reference vector $m_c$ that satisfies a minimum distance or maximum similarity criterion. Through a number of measures are possible, in this study Euclidean distance is used.

$$\|x - m_c\| = \min_i \{\|x - m_c\|\} \quad (4)$$

The best-matching unit and neurons within its neighbourhood are modified by;

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \quad (5)$$

Neighborhood function defines a distance-weighed model for adjusting neuron vectors. The Gaussian model:

$$h_{ci}(t) = \alpha(t).e^{-d_{ci}^2 / 2\sigma_i^2(t)} \quad (6)$$

$d_{ci}$ : Distance between the best matching unit (BMU) and the respective neuron

In the Gaussian model, that neighborhood's size appears as kernel width ($\sigma$) and is not a fixed parameter. Similarly, the initial learning rate ($\alpha_0$) is an input parameter, which is then gradually decreased as $t$ progresses. SOM training stops when a predetermined number of training cycles ($t_{max}$) are completed.

In this study, Matlab SOM tool is used for clustering the river network. 5-by-5 SOM in a two dimensional grid is used for clustering the 272-by-9 inputs (272 features and 9 attributes). The output vectors are trained based on the input vectors. During the training process, a Gaussian neighborhood function is chosen. Empirically chosen parameter settings are listed in Table 1.

**Table 1.** Parameter settings for the SOM training

| Parameter | Value |
|---|---|
| Size | 5x5 |
| Dimensionality | 2 |
| Inputs | 272x9 |
| Topology | Hexagonal |
| Neighbourhood | Gaussian |
| Distance | Euclidean |
| Initial learning rate | 0.5 |
| Training cycles | 20000 |

The first of Fig.1 shows how many of the training data are associated with each of the neurons (hexagons). The topology is a 5-by-5 hexagonal, so there are 25 neurons. Totally 14 clusters are occurred in 25 neurons according to their similarities after the SOM training as shown by hexagons in different sizes. The larger hexagon is associated with more input vectors. The maximum number of hits associated with any neuron is 45. Thus there are 45 input vectors in that cluster. The second of Fig.1 shows the U-matrix, on which distances between input vectors can easily be identified, developed by [33]. In this figure, the blue hexagons represent the neurons. The red lines connect neighboring neurons. The colors in the regions containing the red lines indicate the distances between neurons. The darker colors represent larger distances, and the lighter colors represent smaller distances. Any boundaries cannot be recognized in the U-matrix except the darker one at the upper right corner that neuron represents the longest central stem (main stem), and very different from the other vectors which similar to each other.

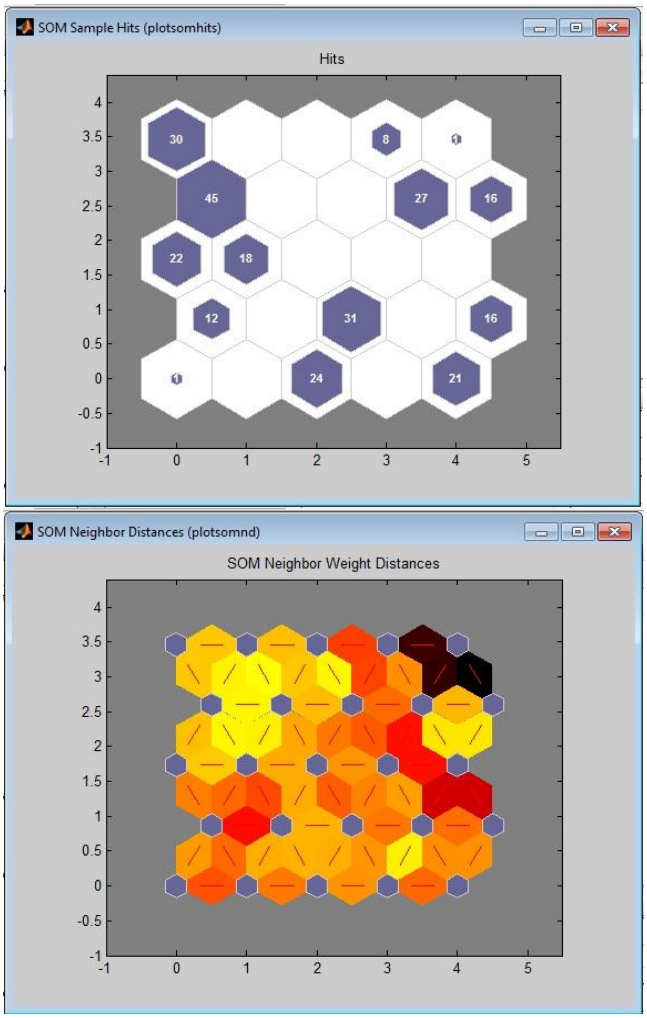



**Fig. 1.** 5-by-5 SOM network: Clustered samples in neurons (first) and neighbor weight distances / U-matrix (second)

**K-means Method**

Given a database of *n* objects and *k*, the number of clusters to form, a partitioning algorithm organizes the object into *k* partitions *(k≤n),* where each partitions represents a cluster. The clusters are formed to optimize an objective partitioning criterion, often called a similarity function, such as distance, so that the objects within a cluster are “similar”, whereas the objects of different clusters are “dissimilar” in terms of the database attributes.

The k-means algorithm for partitioning based on the mean value of the objects in the cluster. The number of clusters *k* and a database containing *n* objects are the inputs. A set of *k* clusters that minimizes the squared-error criterion is the output.

In this method: (1) arbitrarily *k* objects as the initial cluster centers are chosen; (2) each object is assigned to the cluster to which the object is the most similar based on the mean value of

the objects in the cluster (3) the cluster means are updated (4) the iterations are continued until no change [32].

In this study, Matlab is also used for k-means clustering which partitions 272 features-by-9 dimensions into 14 and 25 clusters using the squared Euclidean distance. Each centroid is the mean of points in the cluster. The numbers of clusters (*k*), 14 and 25 are determined based on the SOM results and the total sum of point-to-centroid distances, respectively. In order to decrease the total sum of point-to-centroid distances *k* number is changed from 14 to 25.

## THE CASE STUDY

The National Hydrography Dataset (NHD) is a vector geospatial data layer of the National Map, being developed by the United States Geological Survey (USGS) and created from many data sources and Web Map Services, representing the surface water hydrography of the United States. It is available nationwide as medium resolution at 1:100000-scale, and as high resolution at 1:24000-scale or better.

The study area is the subbasin 102901070403 in the 1:24000 The National Hydrography Dataset, forms the watershed for the Little Pomme de Terre River, in the Midwest United States, in Missouri. The basin area is 159.56 $km^2$.

### Input Data

The geometric, topological and semantic attributes of the rivers are derived from the geodatabase files, namely NHDH1029 (1:24000-scale) and NDHM1029 (1:100000-scale) and National Elevation Dataset 1/3 arc-second (approx. 10m) DEM and used as the inputs of the clustering methods, k-means and SOM.

During selection it is important that a river is processed as a whole, and segments are not separately eliminated, which may disconnect the graph [27]. Before deriving the input data, river segments are combined according to the stream levels. Since the limitations of the elimination of some river segments between the confluences in order to protect the main stems [34], note that river segments are combined according to the stream levels. After combining the segments of rivers, there are totally 272 features included tributaries and stems in the river network.

A stream level is assigned to each reach in the drainage network. Reaches are delineated between confluences in the drainage network, and each reach is assigned a unique and permanent address called a reach code. The stream level value is a numeric code that identifies a hierarchy of main paths of water flow through the network. Level values are established for the purpose of computationally traversing the drainage network through flow relations identified between the reaches. The level coding system appears to be the upstream routine of the Horton and Strahler stream ordering system (NHD standards) [31] (Fig. 2). The derived attributes are normalized and all become in the range of [0,1].

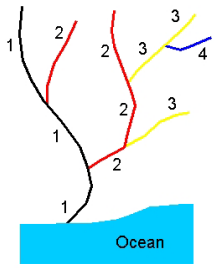



**Fig. 2.** Stream levels [31]

**Geometric Attributes:**

*Length (ratio-scaled)***:** The lengths of tributaries and stems.

*Sinuosity (ratio-scaled)***:** The ratio of the Euclidean distance between the end points to the stream length.

**Topological Attributes:**

Pajek software is used to calculate the topologic attributes via the connectivity graph of the river network [35]. A graph is a set of points (vertices) in a mathematical space, which are interconnected by a set of lines (links). A graph-theoretic representation of river networks uses streams as vertices and stream junctions as links of a connectivity graph to understand the topology of river networks. Centerlines of the lake polygons and the river lines are not in the same feature type. Therefore, centerline of the lake polygon represents different vertex and it is not included into the clustering. Fig. 3 illustrates the Little Pomme de Terre River network at 1:24000-scale and its connectivity graph, and the representative vertex of the lake polygon centerline shown in the circle.

*Degree centrality (ratio-scaled)* of a vertex is defined as the number of links incident upon a vertex (i.e., the number of ties that a vertex has).

$$C_D(V_i) = \frac{\sum_{k=1}^{n} r(V_i, V_k)}{n-1} \tag{7}$$

*Closeness centrality (ratio-scaled)* of a vertex is based on the inverse of the distance of each vertex to every other vertex in the network.

$$C_C(V_i) = \frac{n-1}{\sum_{k=1}^{n} d(V_i, V_k)} \tag{8}$$

*Betweenness centrality (ratio-scaled)* of a vertex counts the number of shortest paths between *i* and *k* that vertex *j* resides on. Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not.

$$C_B(V_i) = \sum_{j=1}^{n} \sum_{k=1}^{j-1} \frac{p_{ikj}}{p_{ij}} \tag{9}$$

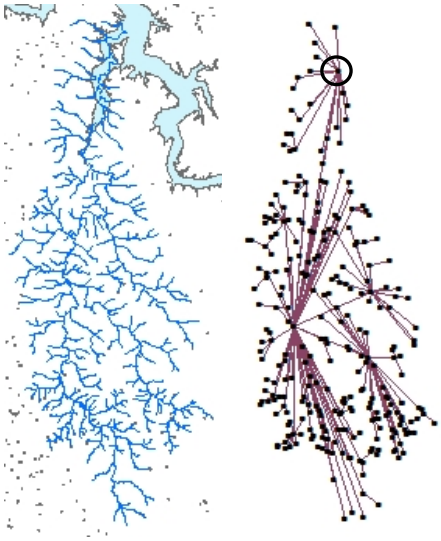


**Fig.3.** Little Pomme de Terre River network at 1:24000-scale (left) and its connectivity graph (right)

**Semantic Attributes:**

*Feature code (binary)***:** Features are classified by type codes. These feature types are such as intermittent and perennial stream/river.

*Stream level (ordinal):* A numeric code identifies each main path of water flow through a drainage network based on Horton and Strahler order. The lowest value of stream level represents the highest hierarchy in the NHD database.

*Lake connection (binary)***:** If a stream/river intersects a lake/pond at 1:24000-scale then the value gets one, else the value is zero.

*Average flow accumulation (ratio-scaled)***:** ArcHydro tools functions are used for the hydrologic computations. The flow accumulation grid that contains the accumulated number of cells upstream of a cell is computed for each cell in the input grid. Then the average flow accumulations of the linear rivers located on the grid cells are determined. In order to compute the average flow accumulation, “DEM Reconditioning” function which modifies the DEM by imposing linear river network (burning/fencing), “Fill Sinks” function which modifies the elevation value to eliminate the problems about cells surrounded with higher elevation and resumes the water flow, and “Flow Directions” using D8 method are used.

**Selection and Elimination of Clusters**

The coefficient of line correspondence (CLC) estimates how well two sets of lines, representing similar features on the ground, overlap each other. Clusters are eliminated based on the coefficient of line correspondence (CLC) is used to evaluate how well the features of the clusters match the target 1:100000-scale. If the features of a cluster do not match any features of the target 1:100000-scale (CLC value is zero) or CLC value is different

from zero, yet the selection of the cluster decreases the total CLC value of the derived network, then the cluster is eliminated.

Totally 9 clusters are eliminated from 14 SOM clusters, whilst 9 and 16 clusters are eliminated from 14 and 25 k-means clusters, respectively. Selected clusters with CLC values are given in Table 2.

CLC value is calculated by the given equation (10).

$$M / (O + C + M) \quad (10)$$

where $M$ is the sum of the lengths of matching target lines, $O$ is the sum of the lengths of target lines that are omitted from the generalized data set, and $C$ is the sum of the length of lines in the generalized data set that do not have a match in the target data set (commission errors), which is divided by the 1:100000 – 1:24000 scales length expansion factor. Reducing line lengths in $C$ by the expansion factor puts all values on a common scale [26]. The expansion factor of 1.03 is determined as the average ratio of the 1:24000-scale to 1:100000-scale lengths for the 15 matching rivers, which are distributed over the study area. Clusters by SOM and two variations of k-means method, and selection and elimination of clusters according to CLC values are given in Table 2 and Table 3.

**Table 2.** Selected clusters with CLC values

| SOM | | k-means (k=14) | | k-means (k=25) | |
|---|---|---|---|---|---|
| Cluster no | CLC | Cluster no | CLC | Cluster no | CLC |
| 10 | 0.09 | 3 | 0.09 | 3 | 0.05 |
| 11 | 0.24 | 8 | 0.07 | 4 | 0.06 |
| 19 | 0.18 | 10 | 0.09 | 6 | 0.02 |
| 24 | 0.05 | 12 | 0.37 | 7 | 0.11 |
| 25 | 0.22 | 13 | 0.18 | 13 | 0.12 |
| | | | | 18 | 0.04 |
| | | | | 20 | 0.08 |
| | | | | 21 | 0.07 |
| | | | | 22 | 0.22 |

CLC value of 0.02 is quite low in Table 2. However, the selection of these clusters increases the total CLC value of the derived network (M>C).

**Table 3.** Eliminated clusters with CLC values

| SOM | | k-means (k=14) | | k-means (k=25) | |
|---|---|---|---|---|---|
| Cluster no | CLC | Cluster no | CLC | Cluster no | CLC |
| 3 | 0.06 | 2 | 0.01 | 2 | 0.01 |
| 6 | 0.01 | 7 | 0.06 | 9 | 0.07 |
| 12 | 0.07 | 11 | 0.06 | 10 | 0.02 |
| 1, 5, 8, 16, 20, 21 | 0 | 1, 4, 5, 6, 9, 14 | 0 | 12 | 0.06 |
| | | | | 14 | 0.01 |
| | | | | 1, 5, 8, 11, 15, 16, 17, 19, 23, 24, 25 | 0 |

In Table 3, some clusters, whose CLC values are different from zero, are eliminated, because of decreasing the total CLC value of the derived networks (M<C). The total CLC values of 0.65, 0.63 and 0.73 result from SOM, k-means whose cluster numbers are 14 and 25, respectively.

## RESULTS

Results are given in several tables and figures below. The visual and quantitative comparisons are done among the derived networks using clustering methods for selection/elimination, USGS 1:24000-scale and 1:100000-scale NHD maps. Töpfer's "Radical Law" which links the scale of the map to the amount of details it should contain is also used for quantitative comparison.

Since the limitations of the elimination of some rivers with respect to protection of main stems, note that stream segments are combined according to the stream levels based on upstream routine of the Horton and Strahler stream ordering system. Therefore, the number of combined river segments (tributaries and stems) is not equal to the number of river segments in the NHD. Combining the river segments task increases the quality of clustering [34].

According to the results, SOM and k-means clustering are very similar when the number of clusters is determined based on the SOM results. The number of selected hydrographic features by SOM and k-means (k=14) are closer to the Töpfer's number (Table 4). The number of features of the target 1:100000-scale NHD is far from the Radical Law number because of the high elimination rate of the short rivers and some of them are in high

hierarchical order. In order to compare the original NHDs in a qualitative manner, the river networks are overlapped in 1:24000 and 1:100000 scales (Fig. 4). Some short rivers in high hierarchical order are eliminated at target 1:100000-scale as shown in Fig. 4 close-up.

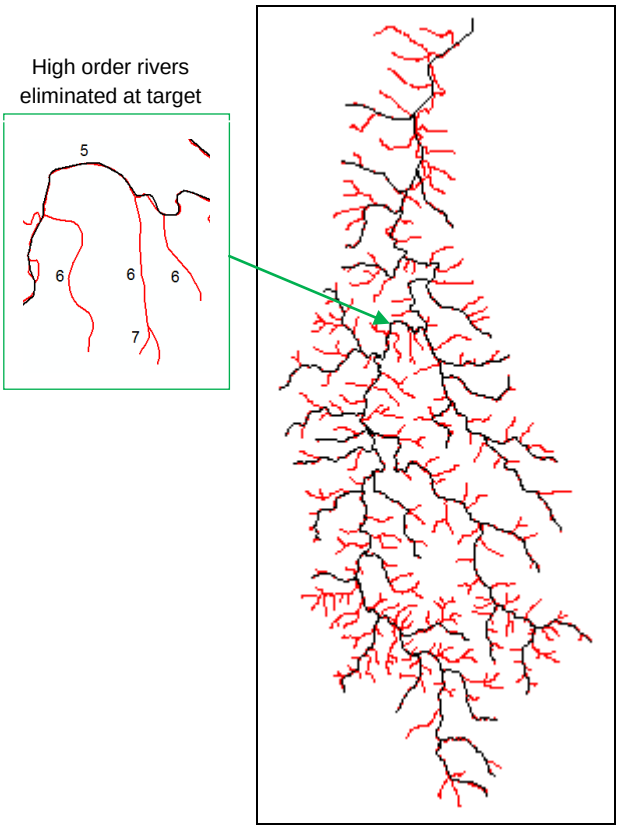



**Fig. 4.** 1:24000 NHD (red), 1:100000 NHD (black)

In Table 4, total network length and drainage density values are given as statistical information. Selection percentages of the total network length and features are also given in brackets. The drainage density is average length of rivers within the basin per unit of area. The drainage density is expressed as:

$$D_d = \sum L / A \qquad (11)$$

where $\sum L$ is the total length of the rivers and $A$ is the area, both in units of the same system [8]. Drainage densities of derived networks using SOM and k-means (k=14) methods are almost same. However, the network derived from 25 k-means clusters is close to the target drainage density.

Since the minimum river length is 30 meters according to the minimal dimensions of data displayed on a screen [36], the length of 0.05 km can be accepted. Selection by k-means (k=25) method is better than SOM considering the minimum lengths.

**Table 4.** Total network length, minimum river length, drainage density and the number of features on the original and derived maps, the number of features at 1:100000-scale by the Radical Law, and the selection percentages of the total network length and features in brackets

| Scale | Number | ∑Length (km) | Min. Length (km) | Density (km/km$^2$) |
|---|---|---|---|---|
| 1:24000 | 272 | 275.047 | 0.02 | 1.72 |
| 1:100000 | 30 (11%) | 122.388 (44%) | 1.58 | 0.77 |
| SOM 1:100000 | 74 (27%) | 140.768 (51%) | 0.05 | 0.88 |
| k-means 1:100000 (k=14) | 77 (28%) | 144.442 (52%) | 0.05 | 0.9 |
| k-means 1:100000 (k=25) | 35 (13%) | 111.45 (40%) | 0.21 | 0.7 |
| Töpfer 1:100000 | 133 (49%) | - | - | - |

As it is given in Table 5, SOM and k-means (k=14) methods retain all the rivers in the highest hierarchical level, namely 5. However, k-means (k=25) retains half of them. The percentage of selected features decreases in the lower hierarchy. All of the features are eliminated in the low stream levels of 8 and 9.

**Table 5.** Numbers of features in the hierarchy classes of the original and derived networks, and selection percentages of the features in brackets

| Stream level | 1:24000 | 1:100000 | SOM 1:100000 | k-means (k=14) | k-means (k=25) |
|---|---|---|---|---|---|
| 5 | 18 | 5 (28%) | 18 (100%) | 18 (100%) | 9 (50%) |
| 6 | 69 | 17 (25%) | 32 (46%) | 35 (51%) | 18 (26%) |
| 7 | 126 | 8 (6%) | 24 (19%) | 24 (19%) | 8 (6%) |
| 8 | 53 | - | - | - | - |
| 9 | 6 | - | - | - | - |

The clustering methods consider the geometric, topological and semantic attributes and try to group all rivers into different categories according to the similarities of attributes. Fig. 5 and 6 are given for visual comparison between derived networks using clustering methods for selection/elimination and NHD networks. Hydrographic generalization using clustering approaches has pleasing visual impact. The proposed k-means method using 25 clusters to select the rivers matches the target scale of 1:100000 better.

In Fig. 5.a and 5.b, derived network using k-means method with 14 clusters and original networks are overlapped and compared. The derived networks of k-means (k=14) and SOM

are almost same. The blue circles show the rivers which are retained different from the derived network using SOM. However, it is not recognized a specific characteristic of these three rivers (Fig. 5.b).

Derived network using k-means method with 25 clusters and original networks are overlapped and compared in Fig. 5.c and 5.d. The blue frame in Fig. 5.c shows the elimination of intermittent rivers in the highest hierarchy, but SOM and the other variation of k-means retained them. The perennial rivers in $7^{th}$ stream level are retained only if they intersect with lakes and betweeness centrality is different from zero, as shown in the green framed close-up. However, the intermittent rivers in $7^{th}$ stream level are eliminated. The intermittent rivers in $6^{th}$ stream level are retained only if they intersect with lakes as shown in red framed close-up (Fig. 5.d).

Finally, derived network using SOM method and the original NHD networks as shown in red are overlapped for comparison in Fig. 6. The perennial rivers in $7^{th}$ stream level are retained only if they intersect with lakes as shown in the green framed close-up. However, the intermittent rivers in $7^{th}$ stream level are eliminated. The intermittent rivers in $6^{th}$ stream level are retained only if they intersect with lakes as shown in red framed close-up. Also, the perennial rivers in $6^{th}$ stream level which intersect with lakes and the high accumulated ones are retained as shown in the blue framed close-up (Fig. 6.b).

As a result, two clustering methods can be used in hydrographic generalization. Two methods organize objects into groupings putting forward the different characteristic. SOM-based approach is an effective method for the selection of rivers via data visualization and exploration for multi dimensional geospatial data and works better for the selection of the rivers considering Töpfer's "Radikal Law". However the derived network using k-means (k=25) matches the target scale of 1:100000 better.

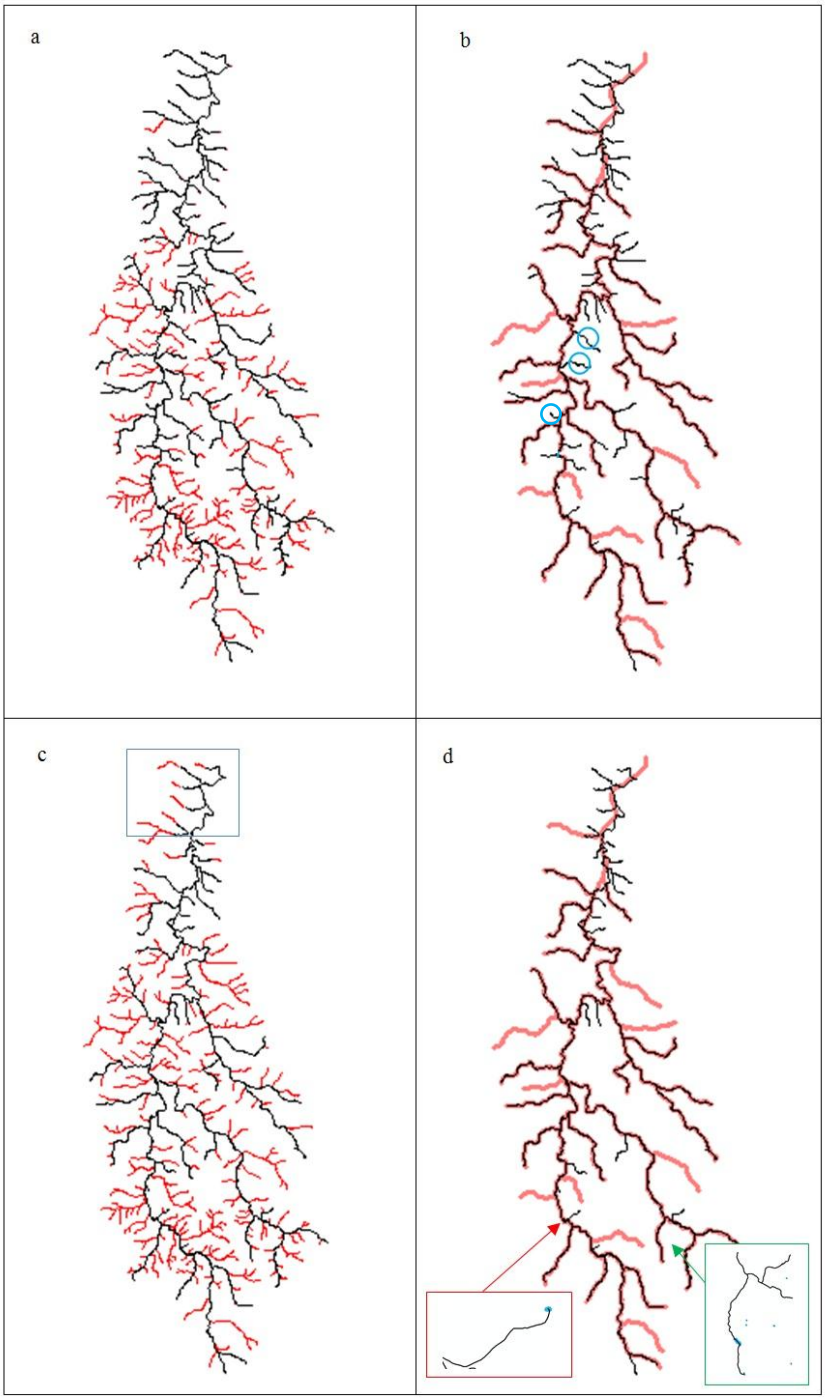



**Fig. 5.** a) 1: 100000 k-means (k=14) clustering (black), 1:24000 NHD (red), b) 1: 100000 k-means (k=14) clustering (black), 1:100000 NHD (red), c) 1: 100000 k-means (k=25) clustering (black), 1:24000 NHD (red), d) 1: 100000 k-means (k=25) clustering (black), 1:100000 NHD (red)

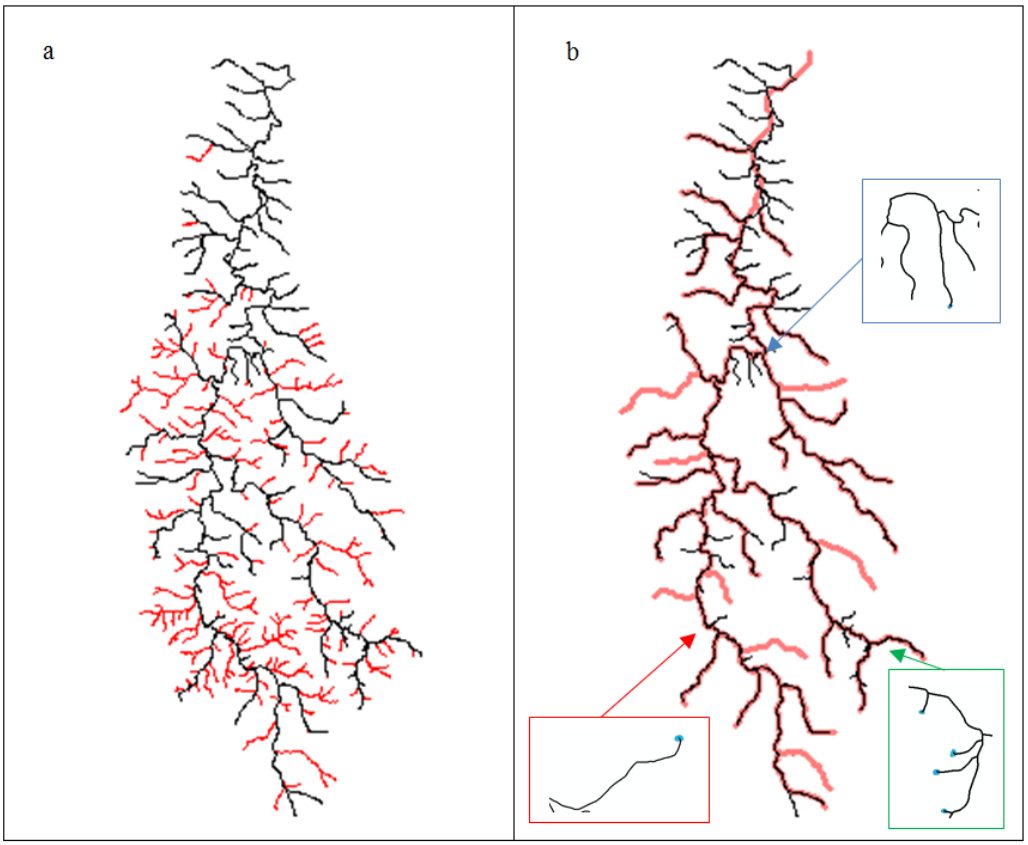



**Fig. 6.** a) 1:100000 SOM clustering (black), 1:24000 NHD (red) b) 1:100000 SOM clustering (black), 1:100000 NHD (red)

## CONCLUSIONS AND FUTURE STUDY

The clustering approaches and CLC values provide an automated model generalization. There is a disadvantage of using the clustering approach for selection. Clustering methods used are not deterministic, which means that different runs of the algorithm lead to different result. This is due to the random selection at the beginning of the process. However, the results are still reflecting the density and distribution of the original situation.

The number of selected hydrographic features by SOM generalization is closer to the Töpfer's number than k-means (k=25) clustering and target 1:100000 NHD. However the derived network by k-means (k=25) matches the target scale of 1:100000 better considering the total CLC value. Moreover, the shortest river in this network is more legible.

The case study applied to the network illustrates that the SOM-based approach can be used as an effective method for the selection of rivers via data visualization and exploration for multi dimensional geospatial data. It is not effective to define the number of clusters by examining the total sum of point-to-centroid distances in k-means clustering.

In the future study, different and more effective attributes will be researched and applied to the net in order to determine the clusters more definite. Some optimization algorithms such as Genetic algorithms will be used as a determiner of the parameters in the SOM architecture.

Support vector machines (SVM) which is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns will be used for training the input data to select the rivers in model generalization.

## ACKNOWLEDGEMENTS

The authors would like to thank Prof. Dr.-Ing. habil. Monika Sester, the director of the Cartography and Geoinformatic Institute, Leibniz University Hannover, for the valuable comments and TÜBİTAK (The Scientific and Technological Research Council of Turkey) for their supports.

## REFERENCES

[1] Kilpalainen, T. (1997) Multiple representation and generalization of geo-databases for topographic maps. Dissertation, Helsinki University of Technology, Publications of the Finnish Geodetic Institute.

[2] Sarjakoski, L. T. (2007) Conceptual models of generalisation and multiple representation. In: Mackaness, W. A., Ruas, A. and Sarjakoski, L. T., (Eds.), Generalization of Geographic Information: Cartographic Modelling and Applications, Elsevier, Amsterdam, 11-35.

[3] Sester, M. (2008) Self-Organizing Maps for density-preserving reduction of objects in cartographic generalization. In: Agarwal, P. and Skupin, A., (Eds.), Self-Organizing Maps Applications in Geographic Information Science, John Wiley & Sons, England, 107-120.

[4] Mackaness, W. A. (2007) Understanding geographic space. In: Mackaness, W. A., Ruas, A. and Sarjakoski, L. T., (Eds.), Generalization of Geographic Information: Cartographic Modelling and Applications. Elsevier, Amsterdam, 1-10.

[5] Joao, E.M. (1998) Causes and Consequences of Map Generalisation. Taylor & Francis Ltd, London, 35-37.

[6] Robinson, A. H., Morrison, J. L., Muehrcke P. C., Kimerling A. J. and Guptill, S. C. (1995) Elements of Cartography, Sixth Edition, Wiley, USA.

[7] Topfer, F. and Pillewiser, W. (1966) The principles of selection. Cartographic Journal, 3(1), 10-16

[8] Horton, R.E. (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Bulletin of the Geological Society of America, 56(3), 275-370.

[9] Strahler, A.N. (1952) Dynamic basis of geomorphology. Bulletin of Geological Society of America, 63(7), 923-938.

[10] Shreve, R.L. (1966) Statistical law of stream numbers. Journal of Geology, 74, 17-37.

[11] Kadmon, N. (1972). Automated selection of settlements in map generalization. The Cartographic Journal, 9(2), 93-98.

[12] Stenhouse, H. (1979) Selection of towns on derived maps. The Cartographic Journal, 16(1), 30-39.

[13] Fitzsimons, D. E. (1985) Base data on thematic maps. The American Cartographer, 12(1): 57-61.

[14] Wolf, G.W. (1988) Weighted surface networks and their application to cartographic generalization. In: Barth, W. (ed) Visualization Technology and Algorithm. Springer-Verlag, Berlin, 199–212.

[15] Richardson, D. (1994) Generalisation of spatial and thematic data using inheritance and classification and aggregation hierarchies. In: Advances in GIS Research 2, Taylor and Francis, 957-972.

[16] Mackaness, W. A. (1995) Analysis of urban road networks to support cartographic generalization. Cartography and Geographic Information Systems, 22(4), 306-316.

[17] Ruas, A. (1998) O-O Constraints modelling to automate urban generalisation process. Proceedings of SDH'98, Vancouver, 225-235.

[18] Ruas, A. (1999) The role of meso level for urban generalisation. Workshop on Progress in Automated Map Generalisation, ICA, Ottowa.

[19] Thomson, R.C. and Richardson, D.E. (1999) The 'good continuity' principle of perceptual organisation applied to the generalisation of road networks. Proceeding of the 19th International Cartographic Conference, Ottawa, Canada, 1215-1225.

[20] Thomson, R.C. and Brooks, R. (2000) Efficient generalization and abstraction of network data using perceptual grouping. Proceedings of the 5th International Conference on GeoComputation: Greenwich, August 23-25.

[21] Jiang, B. and Claramunt, C. (2004) A structural approach to the model generalization of an urban street network. GeoInformatica, 8(2), 157-172.

[22] Jiang, B. and Harrie, L. (2004) Selection of streets from a network using self-organizing maps. Transactions in GIS, 8(3), 35-350.

[23] Ai, T., Liu, Y. and Chen, J., (2006) The hierarchical watershed partitioning and data simplification of river network. In: Riedl, Andreas; Kainz, Wolfgang; Elmes, Gregory A. (Eds.), Progress in Spatial Data Handling, Springer, Part 11, 617-632.

[24] Cetinkaya, B. (2006) Topografik haritaların üretiminde eş yükseklik eğrileri, akan su ve su iletim hatları coğrafi verilerin otomasyon süreçleri. Dissertation, Istanbul Technical University, Fen Bilimleri Enstitüsü, Istanbul (in Turkish).

[25] Touya, G. (2007) River network selection based on structure and pattern recognition. Proceedings of the 23rd International Cartographic Conference, 4-9 August 2007, Moscow, Russia.

[26] Stanislawski, L.V. (2009) Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. Computers, Environment and Urban Systems, 33, 325-333.

[27] Stanislawski, L.V. and Savino S. (2011) Pruning of hydrographic networks: A comparison of two approaches. 14th ICA/ISPRS Workshop on Generalisation and Multiple Representation, 2011, Paris.

[28] Gulgen, F. and Gokgoz, T. (2011) A block-based selection method for road network generalization. International Journal of Digital Earth, 4(2), 133-153.

[29] Sester, M. (2005) Optimization approaches for generalization and data abstraction, International Journal of Geographical Information Science,19(8), 871 – 897.

[30] Agarwal, P. and Skupin, A. (2008) Self-Organising Maps : Applications in Geographic Information Science, Wiley.

[31] USGS (2000) National Hydrography Dataset.

[32] Han, J. and Kamber, M. (2001) Data mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001, USA.

[33] Ultsch, A. and Siemon, H. (1989) Technical report 329, University of Dortmund, Dortmund, Germany.

[34] Sen, A. and Gokgoz, T. (2011) Hydrographic selection / elimination in automated generalization by using artificial neural networks. (Poster), in Proceeding of the 25th International Cartographic Conference, 3-8 July 2011, Paris, France.

[35] Nooy, Mrvar and Batagelj, (2005), Exploratory social network analysis with Pajek review. Cambridge University Press.

[36] Spiess, E. (1995) The need for generalization in a GIS environment. In: Gıs and Generalization, Methodology and Practise, Eds. Müller, J.C., Lagrange, J.P. and Weibel, R., Taylor and Francis London.

# ANALYSIS OF THE RELATIONSHIPS AMONG ERROR VALUES AND VALUES OF MORPHOMETRIC PARAMETERS DERIVED FROM THE DEM

Jana SVOBODOVA[1], Lukas MAREK[1] and Pavel TUCEK[1]

[1]Department of Geoinformatics, Faculty of Science, Palacky University,
tr. Svobody 26, 771 46, Olomouc, Czech Republic

*j.svobodova@upol.cz, lukas.marek@upol.cz, pavel.tucek@upol.cz*

**Abstract**

A fundamental tool for the exploration of the relations (dependencies) among two or more variables is a correlation analysis. Main goal of the correlation is to analyse an existence or absence of the relation among chosen variables and also quantification of the strength of this (in)dependence. A dependent variable as well as an independent variable is not determined during the correlation analysis. This determination of dependency is expressed subsequently during the regression analysis, when the proved relations are mathematically expressed.

The research is aimed at the exploration of the relations among values of errors and values of morphometric characteristics or their changes (derived from digital elevation models). Firstly, the correlation analysis is used. Then, correlated variables entered the regression analysis. Correlation analysis as well as regression analysis has been applied on values (or their changes) of morphometric parameters which were derived from so-called “high-quality” and “low-quality” DEMs of modelled areas. The aim is to find out the variability of relations using the most different DEMs according to the non-spatial evaluation of the metric accuracy. Different terrain configurations were taken into account: three rugged or flat highlands, uplands, hilly areas and three flat lands. Using the multiple areas with the same type of relief allows us to observe a presence of the trend in the relationships among values of errors and values of morphometric characteristics. The paper describes not only the calculation of correlation and regression analyses but also data pre-processing.

**Keywords: DEM, morphometric parameters, correlation, linear regression, GWR**

## INTRODUCTION

Correlation and regression analyses serve for a research of dependencies between variables. The logical procedure of statistical investigation is as follows. Firstly, the correlation analysis is carried out in order to find out whether there are any dependencies between variables. If the dependency among variables has been proven, it is possible to express it mathematically by the regression analysis.

Within the research, a correlation has been used to investigate existence and closeness of relationships between error values and morphometric parameter values. The variables where the relationship has been proven then entered into the regression analysis. The correlation and regression analysis were applied to the error values and the values of morphometric parameters derived from the so-called low-quality of the DEM for each sample area. The aim

was to determine the trend of changes in values of morphometric parameters using low-quality DEM evaluated by global metric accuracy methods.

## STATE OF ART IN RESEARCH OF STUDIED RELATIONSHIPS

At present, very few authors deal with the research of the relationship between error values and values of morphometric parameters using e.g. correlation and regression analysis. As the most important works in this area research by the authors [1], [2] and [3], [4] could be mentioned.

[1] used 4 basic morphometric parameters for regression analysis of the dependence of errors: slope, aspect, total curvature and elevation. [5] replaced the total curvature by planar and profile curvature and together with the other parameters he considered them as adequate for the exhaustive quantification of surface shapes. [3], [4] is not limited in his work to these basic morphometric parameters, but he complements them by the relative relief, texture, mean extremity, minimum extremity, maximum extremity and vertex distance. For most parameters he tested also the second and third derivatives of these variables, in order to investigate non-linear relationships between error values in the DEM and values of morphometric parameters.

To assess the basic relationship between the error values in the DEM and values of morphometric parameters the calculation of the correlation coefficients were separately performend for each parameter. The above-mentioned authors (e.g. [4], [2]), however, consistently state that it is inappropriate to study the strength of the relationships of individual morphometric parameters and errors separately. On the other hand, the scales of the error values cannot be adequately explained only by one parameter alone, it is necessary to use more independent (explanatory) parameters simultaneously. This can be achieved using regression methods.

In the frame of the study, only six basic values of morphometric parameters (slope, aspect, total, profile or planar curvature and elevation) were used to build the regression model. The aim was to find out to what extent are these parameters, which are according to [5] sufficient for an exhaustive description of surface shapes, able to predict error values. The use of further explanatory variables (e.g. according to [4]) was tested on one selected area. Their inclusion in the regression analysis, however, has brought hardly any improvement of quality of a model if the coefficient of determination is regarded.

## METHODS

### Model Areas and Data

The study is focused on the research among the error values (Fig. 1) and the values of morphometric parameters. The research was implemented through the three sample areas (flat highlands): Studenská, Uhřická and Divácká highlands. The aim of this study was to describe the relationships among the error values and the values of morphometric parameters derived from the low-quality DEMs. It is possible to suppose a greater degree of dependence of monitored variables than in the high-quality DEMs. The investigation of the relationships

among these variables derived from the high-quality DEM (even for other types of relief) is subject for further research of the authors.

A more detailed review of methods of DEM quality assessment can be found in [6]. Svobodova also stated that for the original data with high density of points only non-spatial methods of DEM evaluation together with a visual inspection of DEM can be suitable. The mentioned work provides the summary of interpolation methods and their settings for the creation of quality DEMs. Because a larger set of different DEMs, generated by different interpolation methods, was selected in [6], the work contains also an overview of methods and setting which are inappropriate for selected terrain configurations. These inappropriate methods were then used in this study for creation of low-quality DEMs with pixel size of 10 m. List of employed methods and setting of their parameters is shown in Tab. 1.

**Tab. 1.** Interpolation methods and setting of their parameters for creation of low-quality DEM of model areas and the resulting values of root mean square error (RMSE), total absolute error (AE) and hammock index (H) (source: [6])

| Model area | Interpolation method | Power | No. of input points | RMSE | AE | H |
|---|---|---|---|---|---|---|
| Divácká highlands | IDW | 0,5 | 10 | 2,91 | 4072,29 | 0,26 |
| Studenská highlands | IDW | 2 | 20 | 3,04 | 2012,62 | 0,68 |
| Uhřická highlands | IDW | 0,5 | 20 | 3,71 | 4686,53 | 0,24 |

Contour lines with 5-m equidistance from the data model DMÚ25 were chosen as the input data for DEM creation. The contour lines were firstly converted to elevation points. According to the principle of split-sample validation, 15 % of points had been removed before the interpolation. These points were set aside points of reference because of the DEM evaluation. When the DEMs were created (based on the Tab. 1) the raster layers of morphometric characteristics could have been derived. Particularly – slope, aspect, plan, profile and overall curvature were extracted, but only the values of characteristics in the reference points were used further for the correlation and regression analysis because the reference points were the only places where real error of certain DEM could have been calculated. Beside morphometric characteristics, the elevation in reference points was also used in the analyses.

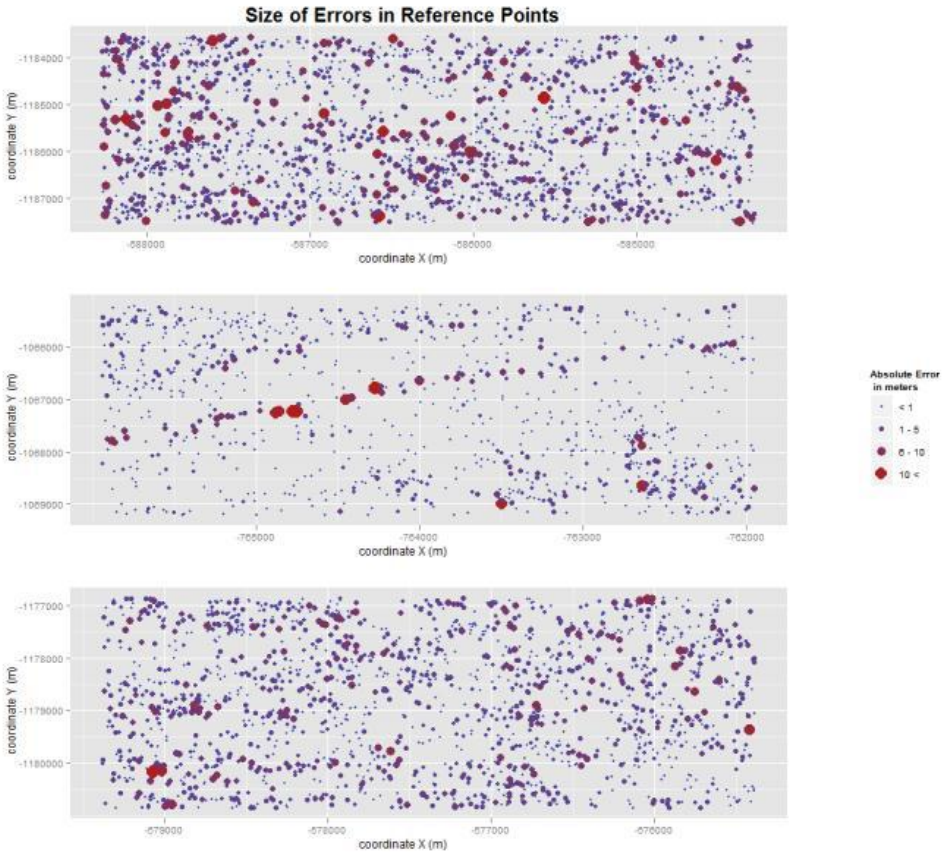



**Fig. 1.** Spatial distribution of reference points and size of DEM errors. Top position - Divácká highlands (DH), middle position - Studenská highlands (SH), bottom position - Uhřická highlands (UH).

## Software

Several softwares were used for analyses of spatial data. ArcMap 10.0 was utilized for creation of DEMs and derivation of morphometric parameters and also for basic operations such as extraction of coordinates from the shapefile, but RStudio – IDE for R, was the main processing software. The R project disposes plenty of packages (libraries) for analyses of geodata. Because we dealt with spatial autocorrelation and geographically weighted regression the most important were packages *spdep* (Spatial dependence: weighting schemes, statistics and models) and *spgwr* (Geographically weighted regression), that provide us the proper analytical capabilities. Alternatively, the software OpenGeoDa that serves as an introduction to spatial data analysis can be used.

## Correlation and Spatial Autocorrelation

Initial work by geographers on measuring the strength of a linear relationship between two variables (x and y) relied upon the Pearson product–moment correlation coefficient ($r_{xy}$), which values lie in the <-1;1> interval [7]. A value of -1 represents perfect negative relationship, $r_{xy}$ = *0* indicates the absence of any statistical relationship between two variables while $r_{xy}$=*1*

means perfect positive relationship. If the variables are not normally distributed, than the Pearson product–moment correlation coefficient is replaced with Spearman's rank coefficient ($r_s$).

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(i_x - i_y)^2}{n \cdot (n^2 - 1)} \quad (1)$$

where expression $(i_x - i_y)$ means differences in ranks of corresponding values of *x* and *y* ; and *n* = numbers of pairs of *x* and *y* values. Pair correlation among more than two variables can be expressed by a correlation matrix. This matrix is squared and its every entry is $r_{xy}$ or $r_s$. The correlation matrix can be visualized in the form of simple level plot, where intensity of linear relationship is described by the colour on the continuous scale [8] (Fig. 4a). Second alternative of visualization is a correlation ellipse (Fig. 4b), where correlation is expressed with the angle of ellipse and the size of the minor semiaxis and furthermore with colour [9].

Because the study dealt with spatial data, we needed to include the spatial extension of correlation. Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional (2-D) surface, introducing a deviation from the independent observations assumption of classical statistics [10]. Spatial autocorrelation exists because real-world phenomena are typified by orderliness, (map) pattern, and systematic concentration, rather than randomness. In other words, spatial autocorrelation means a dependency exists between values of a variable in neighbouring or proximal locations, or a systematic pattern in values of a variable across the locations on a map due to underlying common factors [10]. Positive spatial autocorrelation refers to the patterns where nearby or neighbouring values are more alike; while negative spatial autocorrelation refers to the patterns where nearby or neighbouring values are dissimilar [11].

Most often used global analysis of spatial autocorrelation are *Moran's I statistics*, Getis-Ord G statistics and Geary's C statistics [12]. We focused mainly on Moran's I statistics (2) - an index of spatial autocorrelation, which involves the computation of cross products of mean adjusted values that are geographic neighbours (i.e., covariations). It ranges from roughly (–1, –0.5) to nearly 0 for negative, and nearly 0 to approximately 1 for positive, spatial autocorrelation, with an expected value of *–1/(n – 1)* for zero spatial autocorrelation, where *n* denotes the number of areal units [10]. [11] defined Moran's I as:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(i_x - i_y)^2}{n \cdot (n^2 - 1)} \quad (1)$$

$$I_k = \frac{n * \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{(k)} * (z_i - \bar{z}) * (z_j - \bar{z})}{\left(\sum_{i=1}^{n}(z_i - \bar{z})^2\right) * \left(\sum_{i \neq j}\sum w_{ij}^{(k)}\right)} \quad (2)$$

where $I_k$ is resulting coefficient, $w_{ij}^{(k)}$ is an indication of distance between areas *i* and *j* for step *k*, $z_i$ is observed characteristic and  is a mean of features. The local versions of those global indices are grouped under the term LISA, which means Local Indicators of Spatial Association. Graphical output of local version of the Moran's I is called Moran scatterplot (Fig. 2).

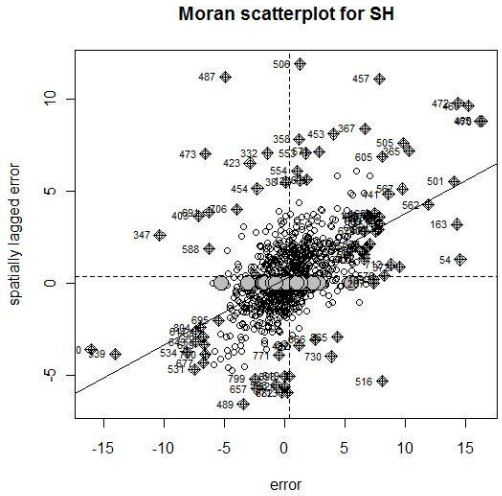



**Fig. 2** Moran scatterplot

**Multiple Linear Regression and Geographically Weighted Regression (GWR)**

Regression analysis is a useful tool for analyses where the aim is to examine whether a particular outcome (the dependent variable) is in some way linked to variations in another phenomenon (the independent or explanatory variable). The simplest case deals only with one dependent and one independent variable and it can be thought as extension of correlation. But frequently it is demanded to examine a relationship among one depend variable and more independent (explanatory) variables. This is called multiple or multivariate regression (3). A multiple regression equation has the general form:

$$y = \alpha + \beta_1 x_1 + ... + \beta_n x_n + e \qquad (3)$$

In the above equation *y* is dependent variable, $\alpha$ is intercept, each value of $\beta_1$ to $\beta_n$ is called the partial regression coefficient, $x_1$ to $x_n$ are explanatory variables and *e* is residual. Residual is the difference between the value of the dependent variable predicted by the model and the true value of the dependent variable. The method of estimation that we used is ordinary least squares.

There exist several assumptions for the regression model [13]:

1. There is a linear relationship between the explanatory variable and the response variable.
2. Residuals are normally distributed.
3. Mean of residuals equals zero.
4. Homoscedasticity of residuals.
5. Autocorrelation: the residuals should be independent of each other.
6. Lack of measurement error.

If the model complies with assumptions mentioned above, then one can discuss its reliability. The most common measure of how well future outcomes are likely to be predicted by the model is called the coefficient of determination $R^2$. It is the proportion of variability in a data set that is accounted for by the statistical model [14]. The $R^2$ value provides us with details about the overall model fit and its values range from 0 to 1 (or 0 – 100 %), in the case of perfectly fitted model.

In the case of multivariate model that contains bigger amount of explanatory variables, it is very likely that dropping-off some of these variables would not cause significant loss of the model's precision. The automatic procedure of the choice of explanatory variables, which remain in the model, is called stepwise regression. Usually, this takes the form of a sequence of F-tests, but other techniques are possible, such as t-tests, $R^2$, Akaike information criterion (*AIC*), Bayesian information criterion (*BIC*), Mallows' *Cp*, or false discovery rate [15], [16].

Geodata can be processed using spatial extension of (multiple) linear regression - Geographically Weighted Regression (GWR). While multiple regression model is rather global method, GMR is local regression technique that allows the model parameters to vary across the space. Although the local technique does not allow extrapolation beyond the region in which the model was established, it allows the parameters to vary locally within the study area, and may provide a more appropriate and accurate basis for descriptive and predictive purpose [17].

The local estimation of the parameters with GWR is given by the equation (for two independent variables) [18]:

$$y = \beta_0(\mu,v) + \beta_1(\mu,v)x_1 + \ldots + \beta_n(\mu,v)x_n + e \quad (4)$$

This regression equation orders the regression parameters to be estimated at a location for which the spatial coordinates are provided by the variables $\mu$ and $v$ and parameters can also be estimated at locations where there are no data [17]. In GWR, the regression and its parameters in each point of space is quantified separately and independently from other points. The regression model is calibrated on all data that are positioned within the region described around a regression point and the process is repeated for all regression points [19]. GWR works in the way that each data point is weighted by its distance from the regression point. It means that the closer a data point is to the regression point, the more weight it

receives [17]. In fact, there exist two main approaches for estimation of regression parameters on individual points. The first approach is application of kernel with a *fixed* bandwidth; the second is then *adaptive* bandwidth (Fig. 3). A fixed bandwidth (e.g. based on Gaussian function) is a suitable choice for modelling if the sample points are reasonably regularly spaced in the study area, otherwise adaptive bandwidth is recommended [20].

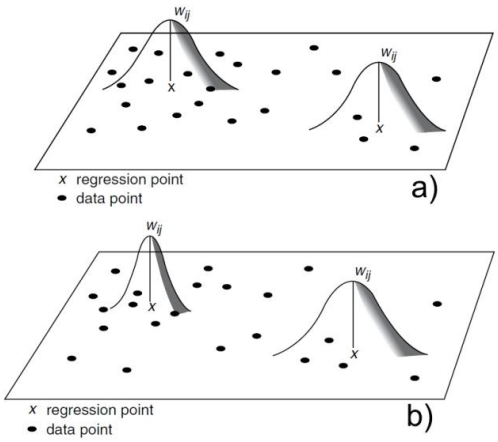



**Fig. 3.** GWR with fixed kernels (a) and adaptive kernels (b) [18]

## RESULTS AND DISCUSSION

### Correlation and Spatial Autocorrelation

Firstly, correlation coefficients between all pairs of characteristics were computed. Results of correlation are shown in the Fig. 4, where two possible ways of the visualization of correlation are used. One can easily find where the relationship between two characteristics exists. The strongest relations were discovered between all types of terrain curvatures. Strong positive correlation ($r_{s\,=}$ 0.9) was found between overall and planimetric curvature (dark red square in the Fig. 4a and narrow ellipses in the Fig. 4b). Strong negative correlations were found between profile and planimetric curvature ($r_{s\,=}$ 0.7) and also profile and overall curvature ($r_{s\,=}$ 0.9) (dark blue color in Fig. 4a). These tight relations between characteristics assume possible removing of some characteristics from the subsequent linear model. Values of correlation coefficients (Fig. 4) are results for the area of Divácká highlands but results for the other two highlands are very similar.

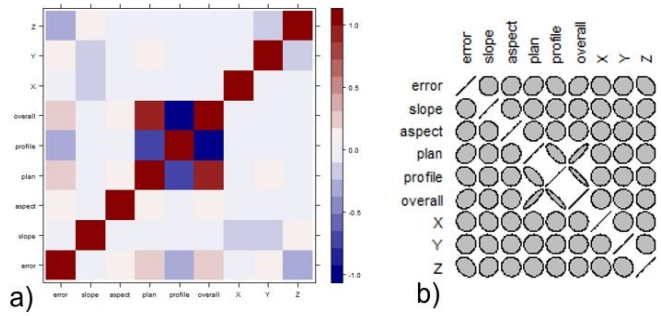



**Fig. 4.** Level plot of correlations (a) and correlation ellipses (b)

Then the spatial autocorrelation of DEMs errors were investigated. Because measures of the spatial autocorrelation are based on spatial weights in the neighbourhood, earlier than their computation, spatial weights needed to be defined. Points in the range from 0 to 150 m were chosen as the neighbourhood (Fig. 5a). Moran's I was then computed using not only the spatial weights but also the Monte Carlo simulations. Values of Moran's I statistics were $I_k$ = 0.21 for Divácká highlands, $I_k$ = 0.37 for Studenská higlands and $I_k$ = 0.35 for Uhřická highlands. These values proved an existence of some at least weak spatial autocorrelation. Moran's plot for Divácká highlands is shown in the Fig. 5b.

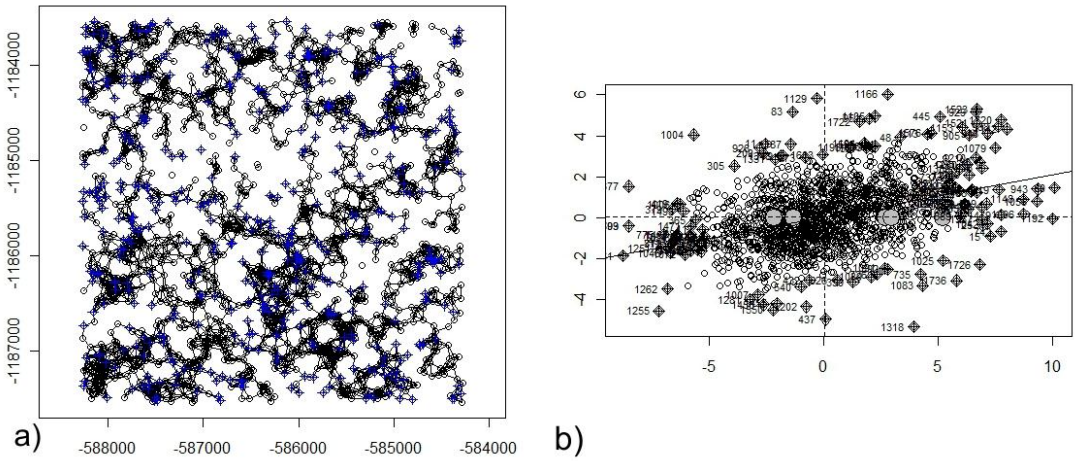



**Fig. 5.** Visualization of neighbours for each points (a) and Moran scatterplot for DH (b)

**Multiple Linear Regression**

The main aim of this study was to prove whether the linear relationship between the value of error and morphometric parameters exists. The multiple linear regression was used for the first assessment of the data. The value of error of created DEM was used as dependent variable and morphometric parameters and elevation as independent (explanatory) variables. The intercept and partial regression coefficients of linear models were calculated using

function for fitting linear models in R. Because some correlations between explanatory variables exist, the reduction of the model can be expected. The simplifications of linear models were performed using the stepwise algorithm with combined forward and backward direction. Finally, the formula with generally usable variables (regression coefficient depends on the particular situation) was chosen. This formula has form: *Error ~ Intercept + Slope + Ascent + Profile Curvature + Elevation.* Resulting coefficients, *AIC* and $R^2$ of the computed models are shown in the Tab. 2. The coefficients of determination ($R^2$), which expresses the quality of the model, are very low $R^2 = 0.06 – 0.17$. That means that these linear models are not suitable for the modelling of errors using morphometric characteristics and elevation.

**Tab. 2.** Values of intercept, partial regression coefficients and AIC and coefficient of determination ($R^2$). DH is original multiple regression model for Divácká highlands containing all explanatory variables, DH step is same model after stepwise regression procedure and General DH is finally chosen general model which uses only four explanatory variables. The same is for Studenská highlands (SH) and Uhřická highlands (UH).

| | | Partial regression coefficients | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Intercept | Slope | Ascent | Planimetric Curvature | Profile Curvature | Overall Curvature | Elevation | *AIC* | $R^2$ |
| DH | 6.492 | -0.016 | 0.004 | 55.558 | -56.026 | -55.560 | -0.024 | 8128.873 | 0.17 |
| DH step | 6.392 | - | 0.004 | - | -0.455 | - | -0.024 | 8125.068 | 0.17 |
| General DH | 6.491 | -0.016 | 0.004 | - | -0.455 | - | -0.024 | 8125.245 | 0.17 |
| SH | 3.374 | 0.071 | 0.003 | -243.273 | 242.902 | 243.421 | -0.008 | 4720.827 | 0.06 |
| SH step | 3.374 | 0.071 | 0.003 | -243.273 | 242.902 | 243.421 | -0.008 | 4720.827 | 0.06 |
| General SH | 3.425 | 0.071 | 0.003 | - | -0.631 | - | -0.008 | 4720.834 | 0.06 |
| UH | 11.431 | -0.024 | 0.003 | 434.896 | -435.033 | -434.935 | -0.042 | 7999.138 | 0.16 |
| UH step | 11.277 | - | 0.003 | 436.557 | -436.686 | -436.591 | -0.042 | 7998.318 | 0.16 |
| General UH | 11.527 | -0.024 | 0.003 | - | 0.045 | - | -0.042 | 8006.681 | 0.16 |

## Geographically Weighted Regression

It was proved earlier, that a spatial autocorrelation exists in the spatial distribution of the error. Although the autocorrelation is not very strong, certain spatial trend can be assumed. Using spatial extension of linear regression – GWR we tried to improve the general multiple linear models. GWR needed to define spatial weights for each point. Both methods of spatial kernels (fixed bandwidth and adaptive bandwidth) were tested and the adaptive kernel was chosen. Sizes of the adaptive kernels were chosen using the automated procedure *gwr.sel* in the R package *spgwr*, which was used also for the computation of GWRs. Parameters of adaptive kernels for GWR as well as their *quasi-global* $R^2$ and *AIC* are shown in the Tab. 3. We can roughly compare resulting values of GWR with matching multiple linear regression and see the improvement of the model quality. Tab. 3 also shows that coefficient of determination

increased significantly in GWR in comparison with linear models and by contrast the value of *AIC* decreased which also proved enhancement of model by using GWR.

**Tab. 3.** Size of adaptive kernels for GWR, *AIC* and coefficient of determination ($R^2$).

| | Adaptive kernel (quantile) | Average no. of points in adaptive kernel | AIC | $R^2$ |
|---|---|---|---|---|
| DV | 0.006 | 10 | 7412.95 | 0.54 |
| SH | 0.013 | 12 | 4309.20 | 0.48 |
| UV | 0.005 | 8 | 7224.43 | 0.59 |

## CONCLUSIONS

Firstly we tried to prove that it is possible to model and also predict errors of DEM using correlation and multiple linear regression. But after the computing of both we discovered that it is not suitable to use these non-spatial methods. We also computed the Moran's I statistic of spatial autocorrelation and proved existence of certain type of spatial trend in data sets. That is why we used local method of linear regression – GWR. GWR significantly improved our linear models ($R^2$ increased about 0.3). Notwithstanding these improvements of models, we cannot recommend linear regression and even GWR for the modelling of DEMs errors spatial distribution (at least for the highlands). While using the multiple linear regression we computed particular formula for the prediction of error, using the GWR the situation and also the interpretation is more difficult, because of the nature of GWR, which is a local method. It means that a basic multiple linear regression is computed for all points using the defined kernel and also partial regression parameters are calculated for this particular point.

## REFERENCES

[1] Erdogan, S. (2009) A comparison of interpolation methods for producing digital elevation models at the field scale. Earth surface processes and landforms, 34, 366-376.

[2] Erdogan, S. (2010) Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study. Computer & Geosciences, 36, 34-43.

[3] Carlisle, B. H. (2002) Digital elevation model quality and uncertainty in DEM-based spatial modelling [online]. [cit. 2010-01-25]. Greenwich, United Kingdom: University of Greenwich. Ph.D. thesis, http://www.numyspace.co.uk/~unn_szbc1/PhD/index.htm

[4] Carlisle, B. H. (2005) Modelling the Spatial Distribution of DEM Error. Transaction in GIS, 9, 4, 521-540

[5] Evans, I. S. (1972) General geomorphometry, derivatives of altitude, and descriptive statistics. In: Spatial Analysis in Geomorphology. Londýn: Methuen, 17-90.

[6] Svobodova, J. (2011) Quality assessment of digital elevation models for environmental applications (in Czech). Ph.D. thesis. Department of Physical Geography and Geoecology, Faculty of Science, University of Ostrava. Ostrava.

[7] Robinson, G. M. (2009) Statistics - Overview, In: Kitchin, R. and Thrift, N. (eds), International Encyclopaedia of Human Geography, 436-451.

[8] Sarkar, D. (2008) Lattice: Multivariate Data Visualization with R, Springer. http://lmdvr.r-forge.r-project.org/

[9] Murdoch, D.J. and Chow, E.D. (1996) A graphical display of large correlation matrices. The American Statistician 50, 178-180.

[10] Griffith, D. A. (2009) Spatial Autocorrelation, In: Kitchin, R. and Thrift, N. (eds), International Encyclopaedia of Human Geography, 308-316.

[11] Lu, Y. and Thill, J.-C. (2009) Assessing the cluster correspondence between paired point locations. Geographical Analysis, 35, 290-309.

[12] Anselin, L. (1995) Local indicators of spatial association – LISA, Geographical Analysis, 27, 93-115.

[13] Pearce, J. (2009) Regression, Linear and Nonlinear, In: Kitchin, R. and Thrift, N. (eds), International Encyclopaedia of Human Geography, 302-308.

[14] Steel, R. G. D. and Torrie, J. H. (1960) Principles and Procedures of Statistics, New York: McGraw-Hill, 187-287.

[15] Hocking, R. R. (1976) The Analysis and Selection of Variables in Linear Regression, Biometrics, 32.

[16] Draper, N. and Smith, H. (1981) Applied Regression Analysis, 2d Edition, New York: John Wiley & Sons, Inc.

[17] Propastin, P., Kappas, M. and Erasmi, S. (2008) Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate. Journal of Spatial Data Infrastructures Research, 3, 73–94.

[18] Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002) Geographically weighted regression: the analysis of spatially varying relationships. Chichester, Willey.

[19] Paez, A. and Wheeler, D.C. (2009) Geographically Weighted Regression, In: Kitchin, R. and Thrift, N. (eds), International Encyclopaedia of Human Geography, 407-414.

[20] Charlton, M. and Fotheringham, A. S. (2009) Geographically Weighted Regression: White Paper. http://ncg.nuim.ie/ncg/GWR/GWR_WhitePaper.pdf

# HARDNESS DATA SYNTHESIS FOR HEIGHT-FIELD BASED LANDSCAPE MODELS

Korneliusz WARSZAWSKI[1], Slawomir NIKIEL[2] and Tomasz ZAWADZKI[1]

[1]Faculty of Electrical Engineering, Computer Science and Telecommunications,
University of Zielona Gora, ul. Podgorna 50, 65-246 Zielona Gora, Poland

*k.warszawski@weit.uz.zgora.pl*
*t.zawadzki@weit.uz.zgora.pl*

[2]Institute of Control and Computation Engineering,
University of Zielona Gora, ul. Podgorna 50, 65-246 Zielona Gora, Poland

*s.nikiel@issi.uz.zgora.pl*

**Abstract**

Landscape modelling is a key element of synthetic environments in virtual reality systems such as: military training systems, virtual reconstruction of cultural heritage, digital entertainment and game development. Most popular and simplest to implement data structures for terrain models are based on the height-field representation. Unfortunately, such models do not contain information about hardness of terrain or it susceptibility for erosion forces in different landscape areas. It is possible to derive hardness data from actual geological records, but for virtual environments we propose the fully automated synthesis of hardness information for terrain models. Our method stems from classical Poisson Faulting algorithm, that was originally used by Mandelbrot and Voss to model ragged landscapes and whole planets. The proposed technique can help to generate hardness data under the height-field base terrain, with regard to different geological materials. The algorithm ensures that those materials are not randomly displaced but form clusters throughout the entire virtual terrain. In addition, the technique can be simply parallelized and implemented in CUDA programming environment. Obtained data structure can be used for further synthesis of eroded landscapes or can work as a test bed for comparison of different geological erosion simulating some models.

**Keywords: Hardness generation, landscape modelling, erosion model, virtual reality**

## INTRODUCTION

Remarkable forms of virtual terrains are possible when developers, artists and virtual world builders spend a lot of time manually deforming polygon meshes. Alternatively, an acceptable visual level of recognized geological structures can be obtained much faster by some automated methods of terrain modelling. We can observe constant development e.g. in chaotic and dynamical systems that may open some paths to modern automated modelling procedures [1], [2].

Applications of those techniques have been used as elements of virtual environments in military and civilian training courses, simulations of cultural heritage reconstructions, digital entertainment, game development and virtual reality systems [3], [4], [5], [6], [7].

Landscapes described as voxel-based model *(voxel maps)* enable modelling terrain shapes and their internal structure, e.g. caves. Such models offer the most accurate representation for complex terrain structures, however the major inconvenience is their complexity and 'greed' for memory storage and computing resources. For most models that are used in simulations, where internal structure is not important case, a classical height-field *(height map)* representation is sufficient. Such a model contains only information about elevation of terrain nodes at coordinates defined by index of row and column of this data structure [8], [9].

This kind of data is a challenge when we want to simulate erosion processes. Implementation of hardness data generation technique, may offer effective solution for this types of inconveniences.

## TERRAIN MODELLING FOR VIRTUAL LANDSCAPES

### The Classic Model

The height-field representation is the most popular structure for landscape shape description. Major advantage of such models is easy implementation of the underlying data structure and relatively simple algorithms for rendering, Level Of Detail *(LOD)*, texturing and shading. The discreet mathematical model is organized as *m-by-n* matrix where cells stores records about elevation at position indicated by indexes of row and column. This kind of model is suitable for description of landscape surface without complex elements e.g. rock shelves, caves, cliffs, etc. [8], [9], [10], [11], [12].

### The Voxel Model

Limitations of landscape models based on height-field structure result from planar space representation. It makes them unacceptable for high visual realism. Any structure similar to caves, cliffs, rock shelves or any vertical surface cannot be described by that model. This inconvenience can be solved with defining landscape as scalar function over three-dimensional grid *(voxel map)*. Major weakness of the voxel-based method is that it consumes lots of system memory resources. Also the rendering process needs complex algorithms with less performance than in the classical approach [9], [10], [13].

### The LDR Model

Benes and Forsbach proposed a Layered Data Representation *(LDR)* of terrain model, that was a compromise between the methods of height-field and voxel landscape description. Similar to the classic approach, the discreet mathematical model is organized as k-layered m-by-n matrixes, where cells contains information about elevation of the layer and other attributes e.g. density, water or gas depositions. Total height of terrain at given coordinates is a sum of heights at each layers on the corresponding positions. In modelling process, those authors assumed that zero-height layer do not exist in general model, which results in

increased performance of the algorithm. In addition, when cells in the layers contains zero-density or include water or gas, then those cells represent caves or holes in the terrain structure. This property is a major advantage over the classic height-field depiction of landscape models and is similar to the voxel representation [14], [15], [16].

## TERRAIN MODEL WITH HARDNESS DATA

In our research, we assumed that the mathematical models for virtual landscapes are based on two-layered representation. Both layers are discrete models organized as *m-by-n* matrix, where $m, n \in N$.

### The Height-Field Layer

The first layer *(H)* represents a classic height-field and is a function defined in equation 1, where each cell represents the elevation value at coordinates defined by indexes of a given row and column in the matrix.

$$H : \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \rightarrow R \quad (\mathbf{1})$$

The initial terrain data can be a plain height-field or it can be imported from a file of real landscape data, e.g., a Digital Elevation Model *(DEM)* or a Geographic Information System *(GIS)*. It can be also modelled by any automated method, e.g., Midpoint Displacement, Poisson Faulting or Particle Systems-based [8], [11], [12], [17], [18], [19], [20], [21], [22].

### The Hardness-Field Layer

The second layer *(D)* stores data describing the hardness of materials from which a given landscape was constructed and is a function defined in equation 2.

$$D : \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \rightarrow R \quad (\mathbf{2})$$

This data can be obtained from geological records. Alternatively, for the virtual environment we propose to simulate it by modified *Poisson Faulting* (also known as *Fault Formation*) algorithm [8], [11], [17].

## HARDNESS SYNTHESIS

As in classic approach, we start generating hardness-field layer from ‘flat ground’ stored in a two-dimensional array. In the first phase we must select a value for height of the fault and a total number of them to be generated over the layer. Less number of faults bring more ‘crispy’ material clusters. To generate a fault, it is necessity to pick two different points with coordinates $(x_1, y_1)$ and $(x_2, y_2)$ exist in the hardness layer. Whit this points we can describe the line of the fault. Next, for each cell coordinates *(i, j)* in the hardness layer, we calculate distance from that line of fault (see equation 4) as determinant of the distance matrix *(S)* constructed as follow in equation 3:

$$S_{ij} = \begin{bmatrix} x_2 - x_1 & i - x_1 \\ y_2 - y_1 & j - y_1 \end{bmatrix} \quad (\mathbf{3})$$

that enables:

$$\det(S_{ij}) = (x_2 - x_1) * (j - y_1) - (y_2 - y_1) * (i - x_1) \quad (4)$$

The distance factor determines on which side of the given fault, the cell of the hardness layer is located (see Fig. 1).

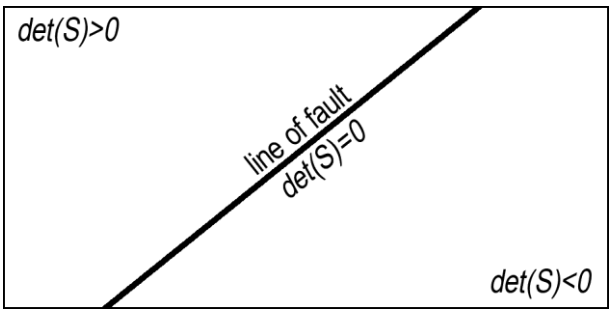



**Fig. 1.** Distance factor

After that step, the layer data is submitted to partial modifications with regard to distance factor *(det(S))*. If this factor is positive then value in given cell *(d)* increases with the previously selected height value of the single fault *(Δh)*. Otherwise, the cell value remains unchanged (see equation 5). The fault generation procedure is repeated until all faults are applied to the hardness-field layer.

$$d_{ij} = \begin{cases} d_{ij} + \Delta h & , \text{if } det(S_{ij}) > 0 \\ d_{ij} & , \text{if } det(S_{ij}) \leqslant 0 \end{cases} \quad (5)$$

The fault forming process results in a fractional Brownian surface *(fBs)* that is usually a base for further modifications [11].

To achieve that is very important to normalize each data to the interval *[0, 1]*. For that purpose, we use the equation 6, a simple normalization method, where normalized values *(d')* are calculated by modifications of initial records *(d)* at given coordinates *(i, j)* by minimal *(min)* and maximal *(max)* values derived from the entire non-normalized data structure.

$$d'_{ij} = \frac{d_{ij} - min}{max - min} \quad (6)$$

The next phase is to choose the total number of materials of which given landscape model is constructed. We called this factor *the class of hardness of terrain model*. When normalized data is less than value *(1.0)* then hardness value *(d")* at given coordinates *(i, j)* is the floor function for multiplications of normalized data *(d')* and class of hardness *(λ) divided by it*, with restrictions stated in equation 7, for initial data equal to *(1.0)* value. This limitation eliminates artefacts data that creates excessed hardness records.

$$d''_{ij} = \begin{cases} \frac{\lfloor d'_{ij} * \lambda \rfloor}{\lambda} & , \text{if } d'_{ij} < 1 \\ \frac{\lambda - 1}{\lambda} & , \text{if } d'_{ij} = 1 \end{cases} \quad (7)$$

Finally, we achieve clusterized hardness data on the second layer of the terrain model. The general structure of presented technique is shown in Fig. 2.

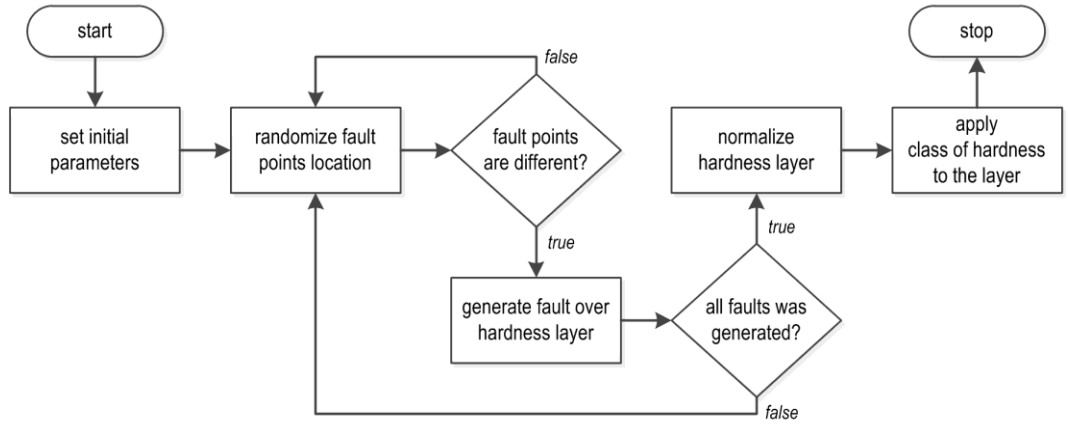



**Fig. 2.** Structure of the hardness synthesis algorithm

## THE IMPLEMENTATION

The overall processing of the data structure is ideal for parallelization and implementation in CUDA environment. This 'parallelism' of the proposed technique results in near real-time synthesis of hardness model for height-field based landscapes. General CUDA implementations of main algorithm phases are shown in follow listings:

**Listing 1.** Poisson Faulting (phase one).

```
__global__ static void Faulting(float fSize,            // height of fault
                                float p1_X, float p1_Y, // position of fault 1st point
                                float p2_X, float p2_Y, // position of fault 2nd point
                                int cols, int rows,     // resolution of layer
                                float* layer)           // records of layer
{
    int index = blockIdx.x * blockDim.x + threadIdx.x;  // cell index

    if (index < cols * rows)                          // index cannot exceed layer size
    {
        float p0_X = index % cols;                      // cell coordinates
        float p0_Y = index / cols;                      // in 2D space

        float det_S = (p0_X - p1_X) * (p0_Y - p2_Y)     // distance factor
                    - (p0_X - p2_X) * (p0_Y - p1_Y);    // and it calculation

        if (det_S) > 0)                                 // for each cell
                                                        // with positive data
        {
            *(layer + index) += fSize;                  // increase value of cell
        }
    }
}
```

**Listing 2.** Normalization (phase two).

```
__global__ static void Normalize(float min, float max,    // extreme values of layer
                                 int cols, int rows,      // resolution of layer
                                 float* layer)            // records of layer
{
    int index = blockIdx.x * blockDim.x + threadIdx.x;    // cell index

    if (index < rows * cols)                          // index cannot exceed layer size
    {
        *(layer + index) = (*(layer + index) - min)       // modify cell data
                         / (max - min);                   // to normalized value
    }
}
```

**Listing 3.** Applying class of hardness (phase three).

```
__global__ static void Hardenize(int hClass,              // class of hardness
                                 int cols, int rows,      // resolution of layer
                                 float* layer)            // records of layer
{
    int index = blockIdx.x * blockDim.x + threadIdx.x;    // cell index

    if (index < rows * cols)                          // index cannot exceed layer size
    {
        if (*(layer + index) < 1.0f)                      // for each cell with
                                                          //value less than 1.0
        {
            *(layer + index) =                            // apply
                   ((int) (*(layer + index) * hClass))    // this modification
                   / (float) hClass;                      // to cell value
        }
        else                                              // in other case
        {
            *(layer + index) = (hClass - 1.0f)            // apply
                             / hClass;                    // this modification
        }                                                 // to cell value
    }
}
```

## RESULTS

The resulting data structure can be simply assigned to the hardness values collected from Friedrich Mohs scale of mineral hardness. It can be used to simulate the erosion processes and their results reflected by virtual landscapes.

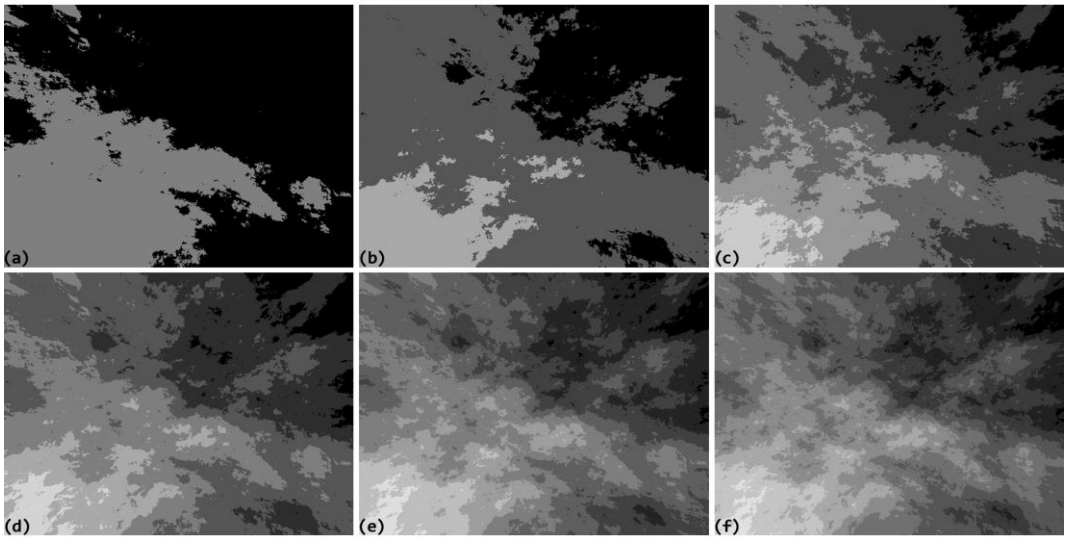



**Fig. 3.** Classes of hardness – graphical representation
(a) 2nd class model, (b) 3rd class model, (c) 5th class model,
(d) 6th class model, (e) 8th class model, (f) 9th class model

As is presented in Fig. 3, different materials are drawn in different tones of greyscale. The results of our simulations depicts formations of mineral structures similar to horizontal cross-section of a land. In direct relation to the class of hardness layer, we achieve more or less varied geological materials

The numerical complexity of the algorithm depends mostly on the number of used faults and decreases when this number increasing. The experimental results, as is shown in Fig. 4, are similar to the results received with classic Poisson faulting technique and confirm the linear nature of our method. The both normalization and mineralization phases have minimal influence (bellow 0,01 sec) to the overall performance of the hardness synthesis process and may be omitted in the performance calculations.

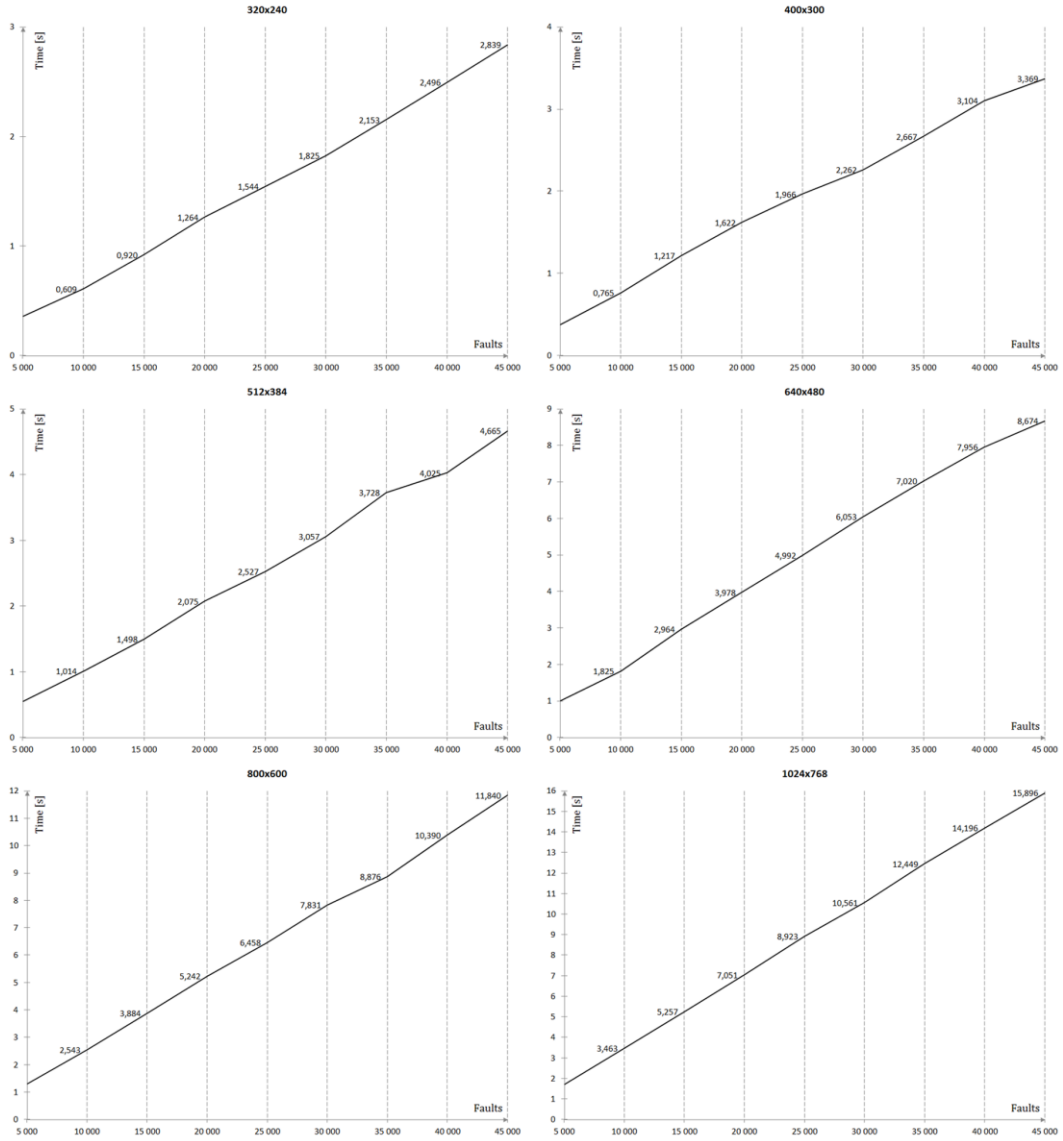



**Fig. 4.** Performance diagrams of hardness synthesis technique

All simulations processes were performed on CUDA-supported nVidia GeForce GTX 560 Ti class GPU. For selected resolution of hardness-field layer, the experimental procedure includes 1 000 generation of faults between 5 000 and 45 000. The diagrams correspond to the average values obtained in those simulations.

## CONCLUSION

In the paper, we proposed the technique for automated hardness synthesis for height-field based terrain models. Such a model opens efficient way for implementing mineral base in the general landscape structure. The performance of the algorithm shows the possibility to make numerous generations of different internal geological structures for a single terrain model. It

can serve as a basis for further simulations of erosion forces acting on that terrain with regard to different mineral depositions. Wide class of hardness data can be done in near-real time process. The use of Poisson Faulting algorithm as a mathematical background of our proposition helps to generate similar hardness data in areas close to each other. The main advantage is clusterized data instead of randomly displaced 'noisy' depositions, thus it is similar to natural structure which can be observed in cross-sections of a land. Additionally, the hardness-field can be directly adopted to terrain models based on the LDR data structure.

Our further research will be focused on the development of hardness synthesis method for landscapes based on the voxel representation, that can be used for improvement of erosion simulation on fully three-dimensional terrain models.

## REFERENCES

[1] Clempner, J. B. and Poznyak, A. S. (2011) Convergence method, properties and computational complexity for Lyapunov games. International Journal of Applied Mathematics and Computer Science, 2 (21), 349-361.

[2] Di Trapani, L. J. and Inanc, T. (2010) NTGsim: A graphical user interface and a 3D simulator for nonlinear trajectory generation methodology. International Journal of Applied Mathematics and Computer Science, 2 (20), 305-316.

[3] Bohemia Interactive Simulations (2011) Bohemia Interactive Simulations. http://www.bisimulations.com, [cit. 2011-00-00]

[4] Bonk, C. J. and Dennen, V. P. (2005) Massive multiplayer online gaming: A research framework for military training and education, TR 2005-1, Office of the Under Secretary of Defense for Personnel and Readiness, Washington, USA.

[5] Rickel, J. and Johnson, W. L. (1999) Virtual humans for team training in Virtual Reality. Proceedings of 9th World Conference on AI in Education (AIED '99), Le Mans, France, 19-23 July, IOS Press, 578-585.

[6] Smelik, R., Tutenel, T., De Kraker, K. J. and Bidarra, R. (2010) Declarative terrain modeling for military training games. International Journal of Computer Games Technology, 2010, 2:1-11.

[7] Wells, W. D. and Darken, C. J. (2005) Generating enhanced natural environments and terrain for interactive combat simulations (GENETICS). Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '05), Monterey, USA, 7-9 November, ACM Press, 184-191.

[8] Musgrave, F. K. (1993) Methods for realistic landscape imaging (PhD thesis). Yale University, New Haven, USA.

[9] Lengyel, E. S. (2010) Voxel-based terrain for real-time virtual simulations (PhD thesis). University of California, Davies, USA.

[10] Frade, M., De Vega, F. F. and Cotta, C. (2009) Breeding terrains with genetic terrain programming: The evolution of terrain generators. International Journal of Computer Games Technology, 2009, 1-13.

[11] Mandelbrot, B. B. (1982) The fractal geometry of nature (2nd ed.). H. W. Freeman and Co., San Francisco, USA.

[12] Musgrave, F. K, Kolb, C. E. and Mace, R. S. (1989), The synthesis and rendering of eroded fractal terrains. Proceedings of 16th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH ’89), Boston, USA, 31 July - 4 August, ACM, 41-50.

[13] Wan, M., Qu, H. and Kaufman, A. (1999) Virtual flythrough over a voxel-based terrain. Proceedings of the 1st IEEE Virtual Reality Conference (VR '99), Huston, USA, 13-17 March, 53-60.

[14] Benes, B. and Forsbach, R. (2001) Layered data representation for visual simulation of terrain erosion. Proceedings of 17th Spring Conference on Computer Graphics (SCCG ’01), Budmerice, Slovakia, 25-28 April, IEEE Computer Society, 80-85.

[15] Benes, B. and Forsbach, R. (2002) Visual simulation of hydraulic erosion. Proceedings of 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG ‘02), Plzen-Bory, Czech Republic, 4-8 February, 79-86.

[16] Benes, B. (2007) Real-time erosion using shallow water simulation. Proceedings of 4th Workshop in Virtual Reality Interactions and Physical Simulation (VRIPHYS ‘07), Dublin, Ireland, 9 November, 43-50.

[17] Fournier, A., Fussell, D. and Carpenter, L. (1982) Computer rendering of stochastic models. Communications of the ACM, 6 (25), 371-384.

[18] Shankel, J. (2000) Fractal terrain generation - Fault formation. In: DeLoura, Mark A. (eds), Game Programming Gems (vol. 1). Charles River Media, Hingham, USA, 499-502.

[19] Shankel, J. (2000) Fractal terrain generation – Midpoint displacement. In: DeLoura, Mark A. (eds), Game Programming Gems (vol. 1). Charles River Media, Hingham, USA, 503-507.

[20] Shankel, J. (2000) Fractal terrain generation – Particle deposition. In: DeLoura, Mark A. (eds), Game Programming Gems (vol. 1). Charles River Media, Hingham, USA, 508-511.

[21] Warszawski, K. (2009) Ground from smoke: Using particle systems in C#. Game Developer Magazine, 3 (16), 15-21.

[22] Warszawski, K. and Nikiel, S. (2009) A proposition of particle system-based technique for automated terrain surface modeling. Proceedings of 5th International North American Conference on Intelligent Games and Simulation (Game-On-NA ‘09), Atlanta, USA, 26-28 August, EUROSIS, 17-19.

# HILL-SHADING BASED ON ANISOTROPIC DIFFUSE ILLUMINATION

Klemen ZAKSEK[1,2], Kristof OSTIR[1,3], Peter PEHANI[1,3],
Ziga KOKALJ[1,3] and Ekkehard POLERT[4]

[1]Centre of Excellence Space-Si, Askerceva 12, SI-1000 Ljubljana, Slovenia

[2]Institute of Geophysics, University of Hamburg, Bundesstr. 55, D-20146 Hamburg, Germany

*klemen.zaksek@zmaw.de*

[3]Institute of Anthropological and Spatial Studies, ZRC SAZU,
Novi trg 2, SI-1000 Ljubljana, Slovenia

*kristof@zrc-sazu.si, peter.pehani@zrc-sazu.si, ziga.kokalj@zrc-sazu.si*

[4]Retired, Hamburg, Germany

*d.dreysacz@arcor.de*

**Abstract**

Analytical hill-shading is a computer-based process of generating a shaded surface from a digital surface model. There are numerous analytical hill-shading techniques; however only the method based on the cosine of the incidence angle of an imaginative direct illumination source became a standard feature in most GIS software. Although widely accepted, this method has two major drawbacks: identifying details in deep shades and inability to properly represent linear features lying parallel to the light beam. Several authors have tried to overcome these limitations by changing the position of the light source.

Our contribution explores alternative terrain visualization techniques based on diffuse, rather than direct, illumination. We first compare three published methods for terrain visualization based on diffuse illumination: openness [1], sky-view factor [2], and uniform sky illumination [3]. These are all based on assumption that illumination is isotropic – the brightness of the light source is direction independent. This visualizes the concave surface forms as dark regions and convex surface forms as bright features. Therefore, these three methods give better insight on the relative elevation of each point than the classical analytical hill-shading.

Yet the analytical hill-shading seems to be more intuitive for most users. Thus we upgraded the previous studies on diffuse surface illumination – instead of considering an isotropically bright sky as the light source we tested various sky brightness distributions. The results can be used as a general terrain visualization because the method combines advantages of direct and diffuse light sources. In particular, we show that such a visualization improves the recognition of local scale features on high resolution DEMs, which is very important in different applications: visual inspection of DEM, recognition of archaeological features, etc.

**Keywords: hill-shading, diffuse illumination, sky-view factor,
anisotropic sky brightness, terrain**

## INTRODUCTION

Development of GIS software made analytical hill-shading one of the most common methods for terrain visualization. Analytical hill-shading is a computer-based process of generating a shaded relief from a raster digital elevation model (DEM). It is a description of how the relief surface reflects the incoming illumination based on physical laws or empirical knowledge. There are numerous analytical hill-shading techniques [4], however only the method developed by Yoëli [5] became a standard feature in common GIS software. Therefore, when we speak about analytical hill-shading (SHD) we consider his method. The computed gray value of the classical SHD is proportional to the cosine of the illumination incidence angle on the relief surface. The method is fast and its results are intuitive to interpret, however it also has two major drawbacks. Direct illumination restricts the visualization in dark shades and very brightly lit areas, where no or very little detail can be perceived. A single light beam also fails to unveil linear structures that lie parallel to it [6].

Another approach to visualize the terrain is employing the uniform diffuse illumination. Diffuse illumination was systematically studied in the field of computer graphics (video games, photorealistic rendering) already in the 1990s when Zhukov et al. [7] introduced obscurances. Because diffuse illumination requires long processing time some studies aimed to simplify it. Proxies for such an illumination model can be extracted from SHD from multiple directions: mean of multiple shading from different azimuths [8], combination of standard SHD (azimuth 315°) with vertical illumination [9], RGB image from hill-shades from three different directions [6, 10]. More advanced methods are based on the horizon estimation. Yokoyama et al. [1] described openness (OPN) – a method that is based on estimating the mean horizon elevation angle in eight directions within a defined search radius. Zakšek et al. [1] proposed to use sky-view factor (SVF) which is a solid angle of the visible sky hemisphere obscured by the terrain. Kennelly and Stewart [3] considered Lambertian reflection while using a uniform sky illumination (USI) instead of point source illumination.

Such approaches significantly enhance perception of relative elevation but they lack the ease of reading the results for non GIS experts. A logical step is merging classical SHD with one of the mentioned methods [1], [2], [3]. In this paper we propose to use an anisotropic diffuse illumination that has advantages of classical SHD and the previously cited methods: it provides the intuitive perception of the terrain, it produces sharper images than the classical SHD (for instance the ridges are easier to follow), even subtle features are easy to detect, etc. In the following, we therefore shortly compare some existing methods based on diffuse illumination and extend them by considering anisotropic diffuse illumination. We propose two simple solutions for generating anisotropic diffuse illumination source. Before the conclusion we discuss the proposed methods and compare them to SHD and SVF.

## TERRAIN VISUALIZATION USING ISOTROPIC DIFFUSE ILLUMINATION

OPN [1], SVF [2], and USI [3] have their origin in a horizon determination (fig. 1), thus their results are similar (fig. 2). There is, however a significant geometric difference between the OPN and the other two methods. OPN does not limit the estimation of each horizon elevation angle ($\gamma$) with the mathematical horizon as other two methods do. In other words, OPN

considers the whole sphere and not only the celestial hemisphere like SVF and USI. Therefore, the maximum OPN value can be greater than π/2. In addition, interpreting OPN is sometimes difficult because a slope is visualized in the same manner as a horizontal plane (both a long slope and a horizontal plane without any obstacles have an openness of π/2). Using an appropriate histogram stretch, it is possible to visualize OPN in a similar way as SVF or USI, but only in the concave areas (with horizon zenith angles in all search directions less than π/2). In other areas, OPN gives significantly different results. We can therefore conclude that OPN is not as intuitive as SVF or USI, but it is more appropriate in automatic detection of linear structures because the method exposes many edges.

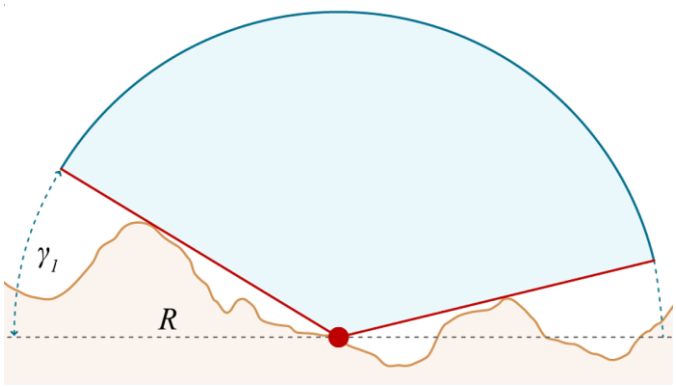



**Fig. 1.** For each direction i the methods OPN, SVF, and USI search for the horizon elevation angle $\gamma_i$ within the vicinity of a stand-point (red dot) limited by the radius R. The blue line represents the visible sky (limited by terrain) that illuminates the stand-point. Since the isotropic illumination is assumed, the brightness of the sky is direction independent.

There are also further differences between these methods. OPN merely averages the horizon elevation angle in n (usually 8) directions (eq. 1). SVF estimates the solid angle of the visible sky in each direction; then it averages the solid angles and normalizes the mean value with the solid angle of a hemisphere (eq. 2).

$$OPN = \frac{\sum_{i=1}^{n} \gamma_i}{n} \quad \textbf{(1)}$$

$$SVF = 1 - \frac{\sum_{i=1}^{n} sin\gamma_i}{n} \quad \textbf{(2)}$$

OPN and SVF therefore merely provide information about the illumination availability. They estimate how much of the whole available light source can be used to illuminate the terrain. They do not, however, provide any information about the interaction of the illumination with the terrain surface. This is what USI does – it assumes that the terrain is perfectly matte – a Lambertian surface, where is the reflected light intensity a function of the cosine of illumination incidence angle δ. The brightness of the Lambertian surface is independent of the viewing

angle. USI therefore integrates the cosine of incidence angle from each infinitesimal solid angle of the visible sky (eq. 3).

$$USI = \frac{1}{2\pi} \cdot \sum_{i=1}^{n} \left[ \frac{2\pi}{n} cos\varepsilon \left( \frac{\pi}{2} sin\delta - scos\delta cos^2 \gamma_i - sin\delta sin\gamma_i cos\gamma_i - \gamma_i sin\delta \right) \right] \quad \textbf{(3)}$$

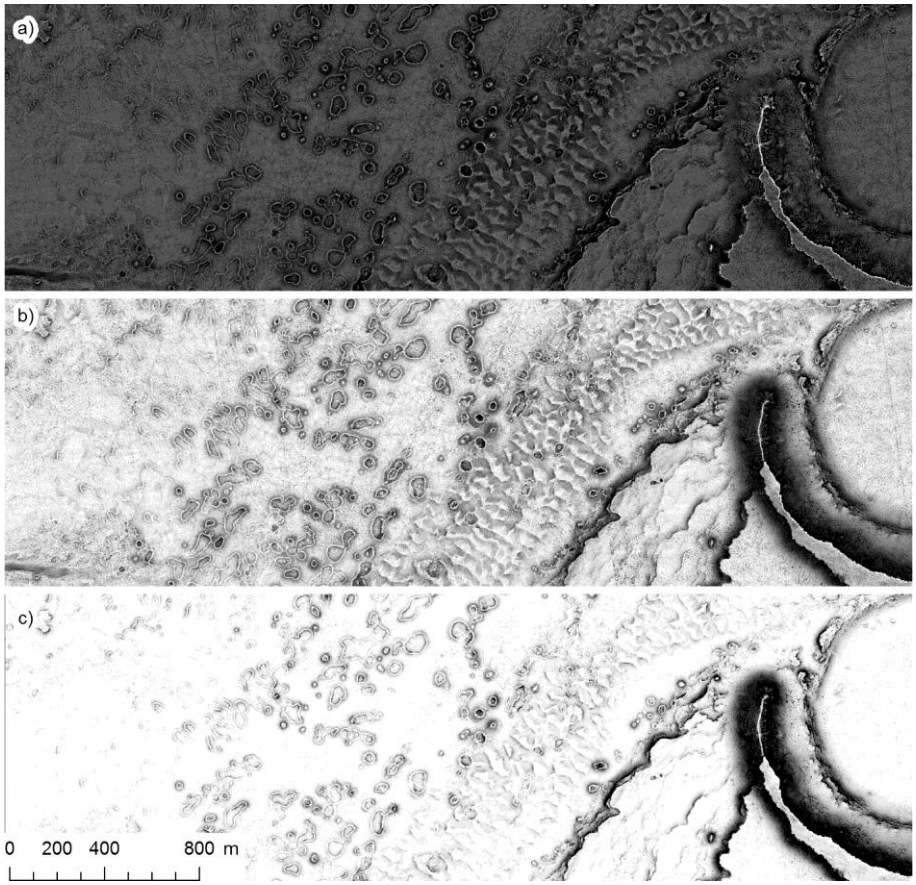



**Fig. 2.** An example for OPN (a) with linear stretch between -10° and 10°, SVF (b) with linear stretch between 0.7 and 1.0, and USI (c) with linear stretch between 0.85 and 1.0. The image shows ring and cone structures, and dunes in Athabasca Valles, Mars (area of 3800 × 1200 m; data from HiRISE instrument aboard Mars Reconnaissance Orbiter [11]; credits to NASA/JPL/University of Arizona). All data were processed in 8 directions with the maximum radius of 10 pixels (30 m).

Because of the OPN's limitations that originate in considering the whole hemisphere we do not consider OPN any further (fig. 2). Linear histogram stretch gives USI a smoother looking surface (fig. 2) but SVF is easier and much faster to compute and exposes more details. In the following text we therefore only consider SVF as a representative technique to try with anisotropic diffuse illumination.

## HILL-SHADING TERRAIN FROM ANISOTOPIC DIFFUSE ILLUMINATION SOURCE

The published methods OPN [1], SVF [2], and USI [3] assume that the intensity of the light source is direction independent. In the field of solar radiation, many methods have been developed that account for anisotropic sky brightness [12], [13], [14], [15], [16]. They claim that the light intensity of each sky element depends on the zenith distance of the sky element and the distance (arc distance across the celestial hemisphere) from the sun. For the visualization purposes we do not need to follow the physical models completely. To enhance the perception of the terrain using diffuse illumination we can simplify to the statement that illumination is azimuth dependent. We thus provide two simple enhancements for SVF based visualization.

### Azimuth dependent diffuse illumination source

Azimuth dependent illumination source is the easiest way to account for anisotropy in the sky brightness distribution. Assuming that the sky is brighter in some directions than in others, we can redefine SVF into $SVF_a$ (equation 4). Please note that such an illumination could theoretically lead to problems on the celestial pole – there each azimuth has different weight.

$$SVF_a = 1 - \frac{\sum_{i=1}^{n} (sin\gamma_i \cdot p_i)}{\sum_{i=1}^{n} p_i} \qquad \textbf{(4)}$$

where $p_i$ is the weight defined in the equation 5. We propose the weight to be based on the cosine function of half angle. The function exponent c has to be an even number (for instance 2 or 4); the higher it is the larger is anisotropy. The two parameters that have to be set for estimation of the weights are: the azimuth of the highest weight $\lambda_{max}$, and the minimal possible weight $p_{min}$ (it is reached when the azimuth is opposite to $\alpha_{max}$). Theoretically, the $p_{min}$ has values between 0 and 1. If it is set to 0 then the terrain is not illuminated from the direction opposite to $\alpha_{max}$. If it is set to 1 then the sky brightness is isotropic. An example for such a weight definition is shown in fig. 3.

$$p_i = (1 - p_{min}) \cdot cos^c \frac{\lambda_i - \lambda_{max}}{2} + p_{min} \qquad \textbf{(5)}$$

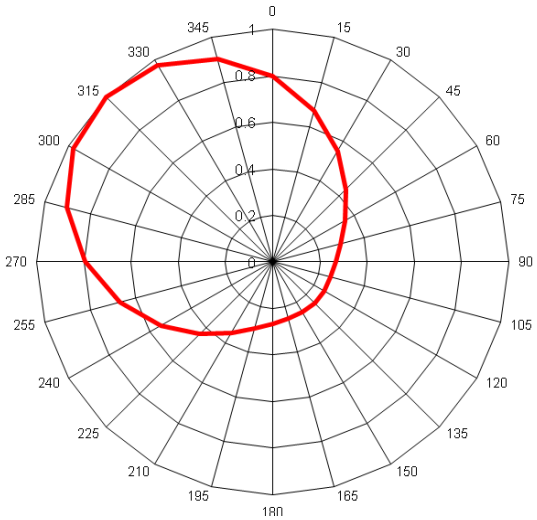



**Fig. 3.** An example of azimuth weight distribution; the largest weight (1.0) is reached at azimuth of 315°, and the lowest (0.25) at the azimuth of 135°; the cosine fuction exponent c is set to 4.

According to our experience the optimal setup for the weight parameters are: c = 4, $\lambda_{max}$ = 315°, and 0.2 < $p_{min}$ < 0.5. Fig. 4 compares the classical SHD, SVF, and $SVF_a$ for the area of Ngorongoro (Tanzania). The obvious difference between the classical SHD and the other two visualisations is larger contrast in the latter two. $SVF_a$ also gives the reader a more intuitive perception of the terrain than the basic SVF. This is evident especially in the lower right corner where a hilly region is much more plastic in $SVF_a$ than in SVF. When compared to the classical SHD, $SVF_a$ preserves the advantages of basic SVF: it gives much more detail in the flat terrain and it “sharpens” the ridges. This enhances especially the perception of the round features (for instance the caldera in the left upper corner).

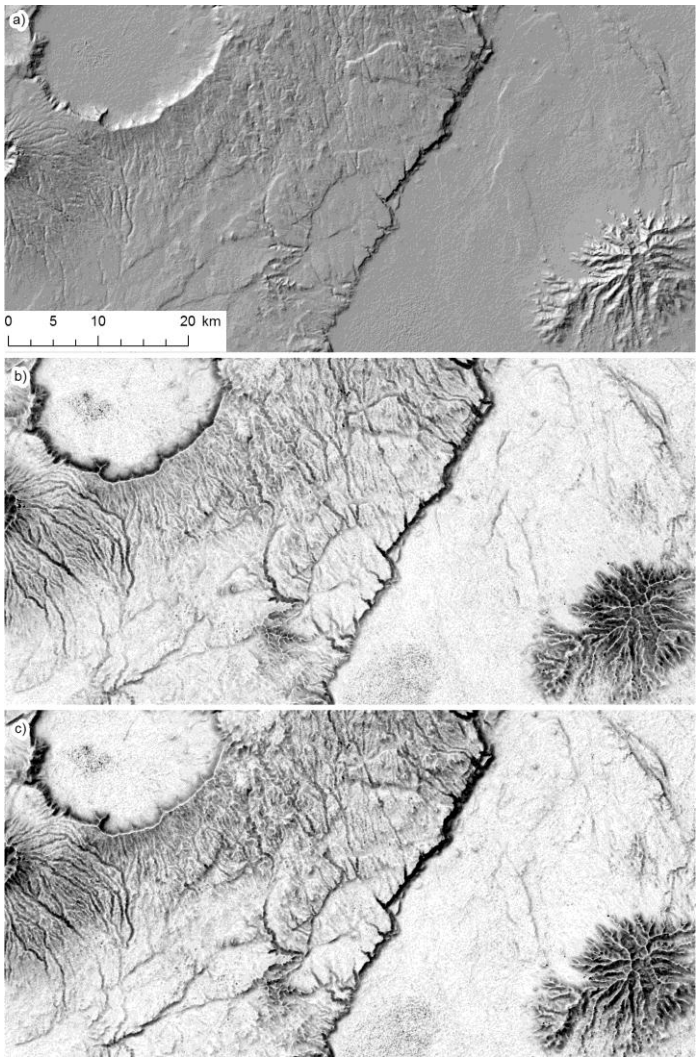



**Fig. 4.** An example for classical SHD (a), with linear stretch between 0.0 and 1.0, SVF (b) with linear stretch between 0.7 and 1.0, and SVFa (c) with linear stretch between 0.7 and 1.0. The image shows Ngorongoro crater and its surrounding area in Tanzania (area of 72 × 36 km; the source of data is ASTER GDEM data [17] obtained from http://lpdaac.usgs.gov, maintained by the NASA). SVF and $SVF_a$ were processed in 8 directions with the maximum radius of 10 pixels (300 m).

**Solar Distance Dependent Diffuse Illumination Source**

From the physical point of view solar distance dependent illumination source has a more reasonable sky brightness distribution than azimuth dependent illumination source. We assume that the sky is the brightest around the position of the sun. In addition, we assume that the sun is so small that we can shade its direct illumination of the terrain. If the scattering in the atmosphere is direction independent, then the sky brightness depends merely on the

arc distance (across the celestial sky) from the sun to the sky element. The easiest way to compute the arc distance is the dot product of unit vectors pointing from the stand point to the sun and from the stand point to the sky element; this gives us the cosine of the arc distance. We redefine SVF into solar distance dependent $SVF_d$ in the equation 6.

$$SVF_d = \int_{VS} f(d_{S-\Omega}) \cdot d\Omega \qquad \textbf{(6)}$$

An appropriate function $f$ is an exponential function (equation 7). We manipulate the visualization with four parameters. The first two define the solar position: solar elevation angle $\varphi_S$, and solar longitude $\lambda_S$. For the sky element that covers the position of the sun we set the weight to 1. The most distant sky element receives a predefined minimal weight $p_{min}$. This can range between 0 and 1 (in the latter case the illumination is isotropic). The decrease of the function value from 1 to $p_{min}$ depends on the exponent c.

$$f(d_{S-\Omega}) = \frac{\left[max(d_{S-\Omega}) - d_{S-\Omega}\right]^c + w_{min}}{\left[max(d_{S-\Omega})\right]^c + w_{min}}$$
$$w_{min} = \frac{\left[max(d_{S-\Omega})\right]^c \cdot p_{min}}{1 - p_{min}} \qquad \textbf{(7)}$$
$$d_{S-\Omega} = arccos([cos\varphi_S cos\lambda_S, cos\varphi_S sin\lambda_S, sin\varphi_S] \cdot [cos\varphi_\Omega cos\lambda_\Omega, cos\varphi_\Omega sin\lambda_\Omega, sin\varphi_\Omega])$$

In the equation 6 we integrate the product of the function $f$ of the arc distance between the sun and to the sky element $d_{S-\Omega}$ with the area of the corresponding sky element $d\Omega$ over the visible sky VS. We solve this integral numerically by generating a look up table (LUT) where each sky element (with spherical coordinate $\varphi_\Omega$, $\lambda_\Omega$) has an assigned value $SVF_d$ value. We divide the whole hemisphere in longitude ($\lambda$) direction into predefined number of lunes where each lune corresponds to one horizon search direction. Each lune is in the latitude direction separated by the previously chosen resolution. For each sky element we then compute the arc distance from the sun (fig. 5 above). This is the basis to estimate a brightness weight for this element. Then we multiply the computed arc distance by the area of the sky element. Summing all these values from zenith to each tangent gives us a look up table (LUT) that is used to solve equation 6 (an example fig. 5 below).

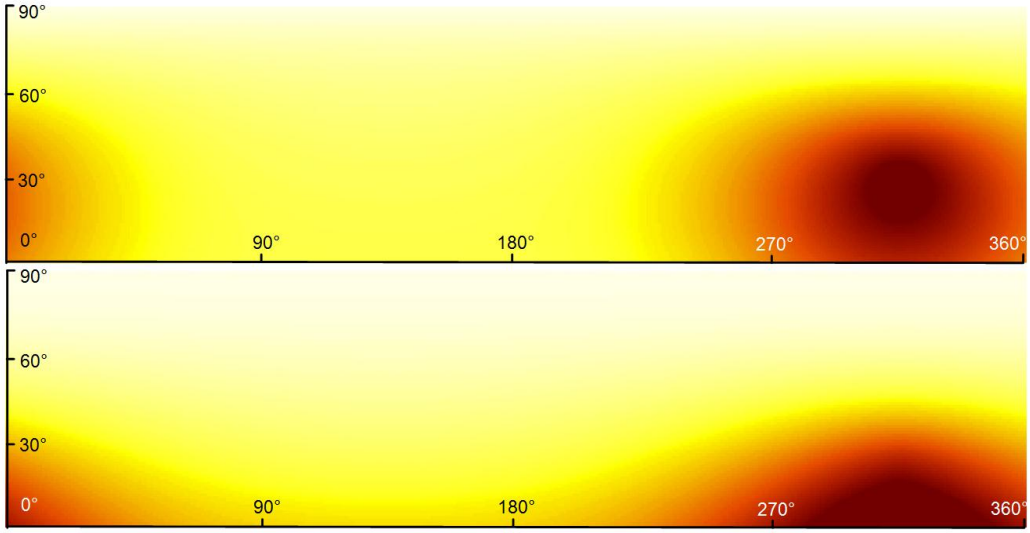



**Fig. 5.** An example of the visualized weight (according to spherical longitude on x axis and spherical latitude on y axis) based on the arc distance between the sun and the sky element (above) and the final LUT (below). The dark colour represents higher values and bright colour lower values. Parameters were set to: $\varphi_S = 30°$, $\lambda_S = 315°$, $p_{min} = 0.2$, and $c = 3$.

Instead of presenting another high resolution example we decided to show that the described procedure works successfully also on low resolution data (200 m); fig. 6 compares the classical SHD, SVF, and $SVF_d$ for the area of the South-Eastern Alps (in Austria and Slovenia). As in fig. 4 we can observe larger contrast in SVF and $SVF_d$ than in the classical SHD. $SVF_d$ was computed using intense illumination anisotropy: $\varphi_S = 10°$, $\lambda_S = 315°$, $p_{min} = 0.0$, and $c = 9$. Such a setup gives the reader a very plastic perception of the terrain. This is evident especially in the high mountains region. No significant improvement comparing to basic SVF can be observed in the flat regions. The disadvantages of such a strong anisotropy are dark shades in the Alpine region. The classic SHD also reveals a very similar problem.

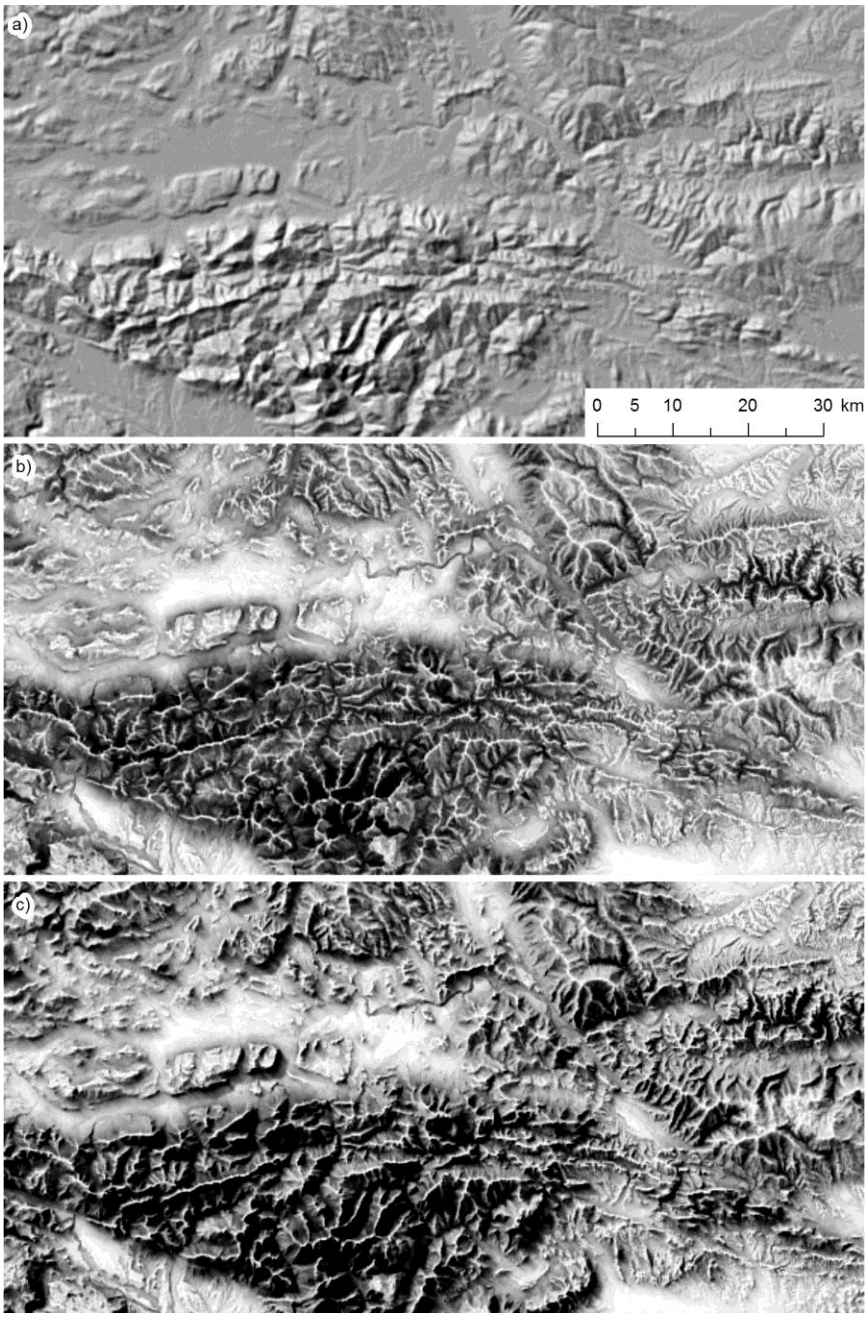



**Fig. 6.** An example for classical SHD (a) with linear stretch between 0.0 and 1.0, SVF (b) with linear stretch between 0.7 and 1.0, and $SVF_d$ (c) with linear stretch between 0.7 and 1.0. The image shows the South-Eastern Alps in Austria and Slovenia (area of 110 × 55 km; the source of data is SRTM data [18] resampled to 200 m resolution). SVF and $SVF_d$ were processed in 8 directions with the maximum radius of 10 pixels (2000 m). The anisotropy parameters were set to: $\varphi_S = 10°$, $\lambda_S = 315°$, $p_{min} = 0.0$, and $c = 9$.

## DISCUSSION AND CONCLUSION

Fig. 7 compares all discussed terrain visualization methods: the classical SHD, SVF, $SVF_a$, and $SVF_d$ for the karst terrain in South-West Slovenia. To make the comparison easier we chose the azimuth of the directional light $\lambda_S = 315°$ (SHD, $SVF_a$, $SVF_d$), elevation of directional light 45° (SHD, $SVF_d$), exponential coefficient $c = 4$ ($SVF_a$, $SVF_d$), and minimum weight $p_{min} = 0.2$ ($SVF_a$, $SVF_d$). The dolines (karstic sinkholes) in lower left corner are notable in all images, but might be misinterpreted to be convex and not concave features – perception varies from person to person. Railway that makes a large curve in the North-South direction is clearly visible on all visualizations. Terraces in the village of Štanjel (upper centre of the image) are the most notable in a) and b). The same goes for the archaeological remains east from the village (round linear form). For the general terrain (right part of the image) the most appealing from the presented visualizations is SHD (a). On terraces in the rough terrain (upper right corner) $SVF_d$ (d) shows more local detail than SHD and still provides enough general terrain clues. On the steep slopes facing away from the sun $SVF_a$ (c) is too dark.

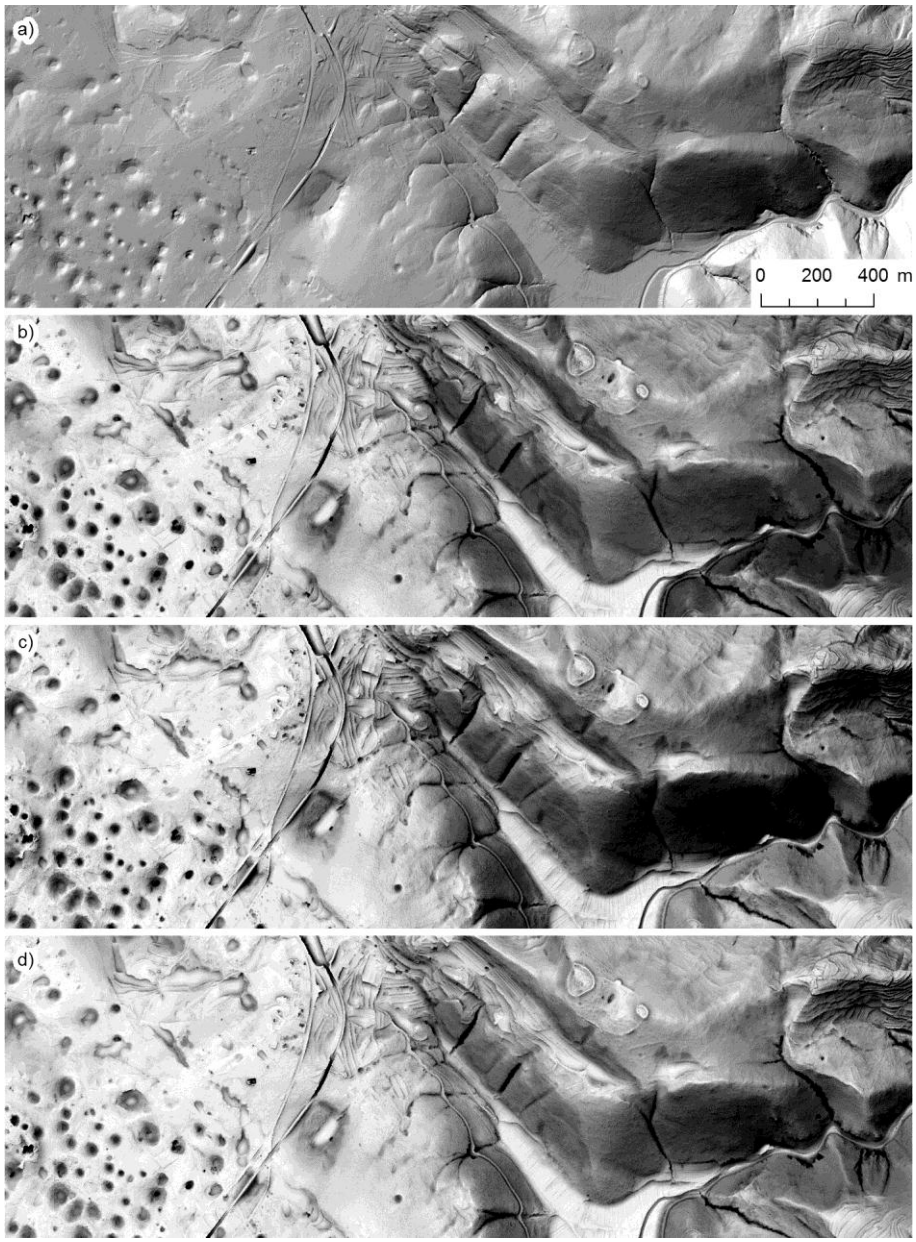



**Fig. 7.** A comparison between classical SHD (a) with linear stretch between 0.0 and 1.0, SVF (b), $SVF_a$ (c), and $SVF_d$ (d). All SVF based results are presented with linear stretch between 0.7 and 1.0. They were processed in 8 directions with maximum search radius 20 pixels (20 m). The area near Štanjel and Kobdilj (South-West Slovenia) covers an area of 3200 × 1100 m. The data source is lidar DEM of 1 m resolution.

Fig. 8 presents the comparison between image histograms (bin size of 0.01) of the data used for fig. 7. SHD has a smooth frequency function with modus just below 0.7 (this value corresponds to the cosine of the flat area using directional illumination source with elevation

angle of $\varphi_S$ = 45°). Small local peak of the SHD curve (on the value of 0) shows that some saturated values were in the data. All three SVF based visualizations have significantly different histogram from the SHD histogram. Their modus is on the value of about 0.95 (such a value corresponds to mostly flat terrain with some small obstacles on the horizon). Because of high anisotropy characteristics of the illumination $SVF_a$ has higher "left tail" of the frequency distribution; the values are transferred (comparing to SVF) from the middle part of the distribution to the left (low values) part of the distribution. This makes all South-Eastern slopes very dark. The distribution of $SVF_d$ has even smaller range that makes it (using the same stretch parameter) in general brighter than the others.

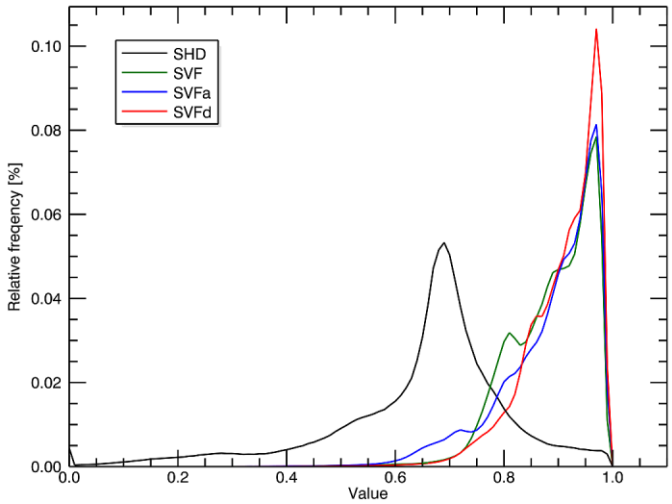



**Fig. 8.** A comparison between image histograms of data used in fig. 7.

The most significant outcome of the histogram comparison (fig. 8) is, however, the fact that the SVF based curves are not as smooth as SHD curve. This means that with these methods we can most likely identify more terrain features than with SHD. We also showed that the anisotropic diffuse illumination can enhance the terrain perception (figs. 4, 6, 7). The question remains: how far can we go by the anisotropy setup? Such a strong effect as we used in fig. 6 diminishes the advantages of diffuse over direct illumination because the shades (southern Alpine slopes) are too dark to expose any detail. These shades can be enhanced using an appropriate histogram manipulation. In the future we will thus continue our work with systematic study of parameters defining the illumination anisotropy models. We intend to define the most critical visualization parameter and its suitable values in order to make it easier to use for non-GIS experts. We also mentioned the importance of the histogram stretch. Here we always used only linear stretch in order to provide easier comparison between different visualizations. The appearance of the image can be significantly improved if an image stretch is optimized, thus this is the second object of our future work.

We can conclude that with introduction of anisotropic illumination it is in general possible to enhance the perception of the terrain illuminated by a diffuse light source, but it might be that the shades are too dark (the same problem as with the classical SHD). The proposed solutions can improve especially the interpretation of terrain data in hydrological analyses,

enhance feature extraction, and are very suitable for visual validation of DEM geomorphologic quality. The two proposed methods result in slightly different visualizations, thus we suggest using $SVF_a$ because it is faster and to easier to manipulate.

## ACKNOWLEDGMENTS

The Centre of Excellence for Space Sciences and Technologies SPACE-SI is an operation partly financed by the European Union, European Regional Development Fund, and Republic of Slovenia, Ministry of Higher Education, Science and Technology.

The ASTER GDEM data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (http://lpdaac.usgs.gov/get_data).

## REFERENCES

[1] Yokoyama, R., Shlrasawa, M. and Pike, R.J. (2002) Visualizing topography by openness: A new application of image processing to digital elevation models. Photogrammetric Engineering and Remote Sensing 68, p. 251-266.

[2] Zaksek, K., Ostir, K. and Kokalj, Z. (2011) Sky-View Factor as a Relief Visualization Technique. Remote Sensing 3, p. 398-415.

[3] Kennelly, P.J. and Stewart, A.J. (2006) A Uniform Sky Illumination Model to Enhance Shading of Terrain and Urban Areas. Cartography and Geographic Information Science 33, p. 21-36.

[4] Horn, B.K.P. (1981) Hill shading and the reflectance map. Proceedings of the IEEE 69, p. 14-47.

[5] Yoeli, P. (1965) Analytische Schattierung. Ein kartographischer Entwurf. Kartographische Nachrichten 15, p. 141–148.

[6] Devereux, B.J., Amable, G.S. and Crow, P. (2008) Visualisation of LiDAR terrain models for archaeological feature detection. Antiquity 82, p. 470-479.

[7] Zhukov, S., Inoes, A. and Kronin, G. (1998) An Ambient Light Illumination Model. In: Rendering Techniques ’98: p. 45-56. Springer-Verlag Wien New York.

[8] Hobbs, K.F. (1995) The rendering of relief images from digital contour data. The Cartographic Journal 32, p. 111-116.

[9] Imhof, E. (1982) Cartographic relief presentation. Berlin, New York, Walter de Gruyter.

[10] Hobbs, K.F. (1999) An investigation of RGB multi-band shading for relief visualisation. International Journal of Applied Earth Observation and Geoinformation 1, p. 181-186.

[11] Kirk, R.L. et al. (2008) Ultrahigh resolution topographic mapping of Mars with MRO HiRISE stereo images: Meter-scale slopes of candidate Phoenix landing sites. Journal of Geophysical Research 113. Available at: http://dx.doi.org/10.1029/2007JE003000.

[12] Gracia, A., Torres, J.L., De Blas, M., Garcia, A. and Perez, R. (2011) Comparison of four luminance and radiance angular distribution models for radiance estimation. Solar Energy 85, p. 2202-2216.

[13] Igawa, N., Koga, Y., Matsuzawa, T. and Nakamura, H. (2004) Models of sky radiance distribution and sky luminance distribution. Solar Energy 77, p. 137-157.

[14] Li, D.H.W. and Cheung, G.H.W. (2005) Study of models for predicting the diffuse irradiance on inclined surfaces. Applied Energy 81, p. 170-186..

[15] Kocifaj, M. (2009) Sky luminance/radiance model with multiple scattering effect. Solar Energy 83, 1914-1922.

[16] Perez, R., Seals, R. and Michalsky, J. (1993) All-weather model for sky luminance distribution—Preliminary configuration and validation. Solar Energy 50, p. 235-245.

[17] NASA Land Processes Distributed Active Archive Center (LP DAAC) at the USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. (2003) ASTER Global DEM. http://lpdaac.usgs.gov/, [cit. 2011-10-27].

[18] Jarvis, A., Reuter, H.I., Nelson, A. and Guevara, E. (2008) Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org/, [cit. 2010-12-27].

# TERRAIN AND CANOPY SURFACE MODELLING FROM LIDAR DATA FOR TREE SPECIES CLASSIFICATION

Zhenyu ZHANG[1, 2], Xiaoye LIU[2], Jim PETERSON[1] and Wendy WRIGHT[3]

[1]Centre for GIS, School of Geography and Environmental Science, Monash University, Clayton, Victoria 3800, Australia

*Zhenyu.Zhang@usq.edu.au; Jim.Peterson@monash.edu.au*

[2]Australian Centre for Sustainable Catchments, and Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland 4350, Australia

*Xiaoye.Liu@usq.edu.au*

[3]School of Applied Sciences and Engineering, Monash University, Churchill, Victoria 3842, Australia

*Wendy.Wright@monash.edu.au*

**Abstract**

It has been recognised that airborne LiDAR (light detection and ranging) offers advantages over the interpretation of aerial photographs and processing of multi-spectral and/or hyper-spectral remote sensing data in forest classification. LiDAR with capability of canopy penetration yields such high density sampling that detailed terrain and canopy surface models can be derived. Recent success in forest classification using LiDAR derived products including terrain and canopy surface models has been reported in many studies. However, there is still considerable scope for further improvement in classification accuracy by taking maximum advantage of the information extracted from LiDAR data and by employing more efficient classifiers such as support vector machines (SVMs). This study aims to use LiDAR data to generate digital terrain and canopy surface models to identify the location and crown size of individual trees for the species classification of Australian cool temperate rainforest dominated by the Myrtle Beech (*Nothofagus cunninghamii*) and neighbouring Silver Wattle (*Acacia dealbata*). The tree species classification was achieved by employing LiDAR-derived structure and intensity variables via linear discriminant analysis (LDA) and SVMs. The results showed that the inclusion of LiDAR-derived intensity variables improved the accuracy of the classification of the Myrtle Beech and the Silver Wattle species in the study area. It demonstrated that the SVMs have significant advantages over the traditional classification methods such as the LDA methods in terms of classification accuracy.

**Keywords: LiDAR, LiDAR intensity, canopy surface model, support vector machines, forest classification**

## INTRODUCTION

Airborne light detection and ranging (LiDAR), also referred to as airborne laser scanning (ALS), is one of the most effective means of terrain data collection. Using LiDAR data for the generation of digital terrain model (DTM) (or digital elevation model (DEM)) is becoming a standard practice in the spatial science community [1]. There has been increasing interest in the application of airborne LiDAR for the analysis of forest structures and the classification of forest type over the last decade [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]. LiDAR applications in forest classification mostly refer to either the structural description at plot or stand level or to single tree identification [12], [13], [14], [15]. As high-density LiDAR data became more readily available, the scope to single tree identification and species classification became more obvious. From dense LiDAR points, individual trees can be identified. Typically, individual tree crown outlines are extracted from a LiDAR-derived CHM (canopy height model) [16], [17], [18], [5] such that the LiDAR point cloud within the individual crown outlines become the key to deriving the variables relevant to tree species classification. These variables are derived either from LiDAR height measurement or LiDAR intensity values. The basic idea of using structure/height and intensity variables for tree species classification is that different species have different crown properties [5]. The vertical distribution and configuration of the forest components are forest type dependent [16]. The structural differences will affect the distribution of the laser returns from the forests [5]. Therefore, the variables derived from the LiDAR data can be used for tree species identification and forest type classification. Height (or structure) related variables usually included crown height, crown depth, mean crown height and crown density [19].

In addition to the three-dimensional coordinates, most LiDAR systems also capture the intensity of the backscattered laser pulses [20], [21]. The back-scattered laser signal is converted to an electrical signal by a photodetector (typically an avalanche photodiode). The generated photocurrent or voltage is then quantized to a digital number (usually expressed in percent value, representing the ratio of strength of reflected light to that of emitted light or by scaling the minimum and maximum values to a 8-bit (0-255) grayscale palette) which is referred to as the LiDAR intensity value for a particular return [21], [22]. Most commercial LiDAR systems used for topographic and forest mapping use lasers that emit energy in the near infrared range of the electromagnetic spectrum. Vegetation reflects this wavelength well [23], [29]. It has great potential that the LiDAR intensity values could assist tree species classification [5]. The LiDAR intensity of forest canopies is influenced by many factors including the reflectance, density, size and orientation distribution of foliage and system settings [22]. Applications of LiDAR intensity data in tree species identification have been attempted in several studies [17], [25], [26], [24], [5], [22]. It has shown great potential that the LiDAR intensity values could assist tree species classification [5].

Recent success in forest classification and tree species identification using LiDAR-derived variables has been reported in a few published articles. However, there is still considerable scope for further improvement in classification accuracy. Furthermore, the problems associated with satisfying the assumptions that underlie traditional classification methods have also driven research into nonparametric alternatives such as neural networks, decision trees, and more recently support vector machines (SVMs) [27]. The SVMs have recently

attracted the attention of the remote sensing community [28], [27], [29], [30], [31] and have been attempted to LiDAR-derived variables in forest applications [32], [33], [34], [35], [36]. SVMs originated from the principle of statistical learning theory and its foundations were introduced by Vapnik [38], [39]. The simplest nature of classification with a SVM is for the situation in which the two classes are linearly separable [40]. The core operation of SVMs is to construct an optimal hyperplane that separates the two classes in such a way that the distance from the hyperplane to the closest training data points in each of the classes is as large as possible [38], [29]. The hyperplane is determined in a SVM to maximize the generalization ability. However, if the input data are not linearly separable, the obtained classifier may not have high generalization ability [41]. For nonlinear decision surfaces, the input data are mapped into a high-dimensional feature space via nonlinear vector mapping function by introducing a variety of kernels, in order to spread the distribution of the data in a way that facilitates the fitting of a linear hyperplane [29], [41]. The selection of kernel function and appropriate values for corresponding kernel parameters, referred to as kernel configuration, may affect the performance of the SVMs [28]. Among many developed kernels [42], the radial basis function (RBF) is widely used [29]. The only one parameter to be predefined for RBF is $\gamma$ which controls the width of the kernel. The accuracy of classification by a SVM is dependent on the proper selection of the magnitude of the parameters $C$ and $\gamma$ [27]. $C$ is a penalty parameter (or cost parameter) to be determined by the user, which determines the trade-off between the maximization of the margin and the minimization of the classification error [41]. For details of SVMs and applications in remote sensing, readers are referred to [38], [41], [43], [27] and [29]. The overall objective of this study is to use LiDAR data to generate digital terrain and canopy surface models to identify the location and crown size of individual trees for the species classification. The specific objectives are to examine the contribution of the LiDAR intensity variables to the classification results and evaluate the performance the SVMs for the tree species classification.

## MATERIALS AND METHODS

### Study Area

The study area is in the eastern Strzelecki Ranges, southeast Victoria, Australia (Fig. 1). The Strzelecki Ranges (formerly known as the Great Forest of South Gippsland) are an isolated series of mountains in the southern section of the Gippsland region. Prior to European settlement the Strzelecki Ranges were densely vegetated by wet forest (also referred to as wet sclerophyll forest) and cool temperate rainforest. Wet forest is most commonly dominated by Mountain Ash (*Eucalyptus regnans*) [44], characterised by a tall eucalypt overstorey, a broad-leaved shrubby understorey and a moist, shaded, fern-rich ground layer that is usually dominated by tree-ferns [45]. In eucalypt-free areas, Silver Wattle may be locally dominant [44]. Cool temperate rainforest is defined as a closed, non-eucalypt forest, which occurs in high rainfall areas and within wet forest areas which have not been exposed to fire. Myrtle Beech is the dominant species of cool temperate rainforest in the study area. These forests have experienced widespread land clearing since European settlement. Subsequent agricultural abandonment and a frequent wildfire history have resulted in severely disturbed landscape in the Strzelecki Ranges [46], [47], [48]. Rainforest is sensitive to fire and, following

fire, is often replaced by forest types dominated by fire tolerant species such as some eucalypts, which rely on fire to open their protective seed pods so that their seeds can germinate [49], [50]. For example, in the study area, there was extensive regeneration of eucalypt forest following catastrophic wildfires in 1939 and 1944. The landscape has undergone further significant changes with the establishment of large scale plantations in the area since the mid-twentieth century [51], [46]. Currently, areas bordering cool temperate rainforest in the Eastern Strzeleckis are a mosaic of different land use histories involving both natural and human disturbances, and so a very complex forest structure in the remnant patches of cool temperate rainforest and adjacent forests including wet sclerophyll and plantation forests prevails. This study focuses on an area with cool temperate rainforest distribution in the Eastern Strzeleckis, which covers an area of 1.82 km2 with elevations ranging between 322 m and 448 m.

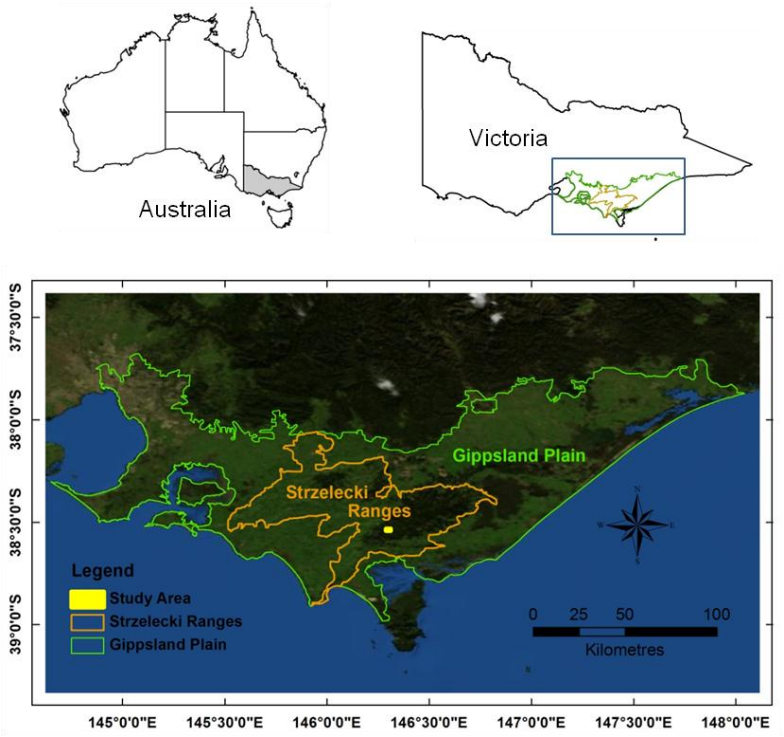



**Fig. 1.** Study area

**Data**

LiDAR data were collected using an Optech ALTM Gemini LiDAR system at a flying height of 1,100 m above ground between 11 and 23 October 2009 (for the whole Strzelecki Ranges). The laser pulse repetition frequency is 70 kHz. The laser scanner was configured to record up to 4 returns for one laser pulse. The average point density was 4 points per square metre, and the laser footprint diameter was 0.3 m. The LiDAR data used for this project was documented as 0.20 m for vertical accuracy and 0.25 m for horizontal accuracy. The LiDAR data were

classified into ground and non-ground points by the vendor and were delivered in binary LAS 1.2 file format.

Ecological Vegetation Classes (EVCs), which describe the spatial extent of native vegetation types, were introduced by the Victoria Department of Sustainability and Environment in the 1990s. EVC mapping was implemented as part of the regional forest agreements, driven by a need to determine boundaries for a forest reserve system. The EVC mapping was undertaken first by the interpretation of aerial photographs and the process was designed to outline native vegetation patches and any obviously related patterns. The range of aerial photograph patterns was then field checked and lists of plant species were recorded [44], [52]. EVCs are the basic regional scale mapping unit used for forest ecosystem assessments, biodiversity planning and conservation management in Victoria. The EVC data provided by the HVP Plantations Pty Ltd were used as reference data in this study.

**Methods**

A DEM with one metre horizontal resolution (grid size) was generated using the LiDAR ground data. The highest points of the first returns of non-ground LiDAR data within each one metre square which represent the laser returns from tree canopy were used to generate a canopy surface model (CSM) for the study area. A canopy height model (CHM) was computed by subtracting the DEM from the CSM. The CHM represents the height variance of the top canopy. The TreeVaW software developed by [53] was used to identify the location and crown size of individual trees from the CHM in the study area. It operates on the assumption that the local maximum height in a spatial neighbourhood represents the tip of a tree crown. Detailed descriptions of the TreeVaW algorithm performance are found in [2], [53], [18]. For each identified tree, the crown diameter for each identified tree was calculated by averaging two lengths measured along two perpendicular directions from the location of the tree top [18]. The LiDAR points were extracted for each of the individual trees using a cylinder defined by that tree crown diameter obtained from the output of the TreeVaW. Extracted LiDAR points were used to create tree height profiles representing the spatial distribution of the vertical structure of individual trees. An important tree attribute is the crown base height: the distance measured from the ground to the bottom of the tree crown. The *k*-means clustering algorithm that produces a partition of the data into the *k* different clusters in such a way that all individuals in a cluster are closer to their own cluster mean [54], [55] was performed on the height profile to determine the crown base height of each individual tree. It is the LiDAR point data from above the crown base that are used in deriving canopy variables. The variable names and descriptions derived from heights and intensity values of laser returns within tree crowns are listed in Table 1.

One-way ANOVA (analysis of variance), a statistical technique to test whether the observed differences between the sample means are of such magnitude as to indicate that they could have come from the same or different populations [56], was performed on these variables to see if each of these variables can be used to distinguish one tree species from the other.

The hypotheses to be tested for a variable in the one-way ANOVA are [57]:

$H_0$: $\mu_1 = \mu_2$ (the means of a variable from the two tree species are equal)

$H_1$: $\mu_1 \neq \mu_2$ (the means of a variable from the two tree species are not equal)

The *P*-value in an *F*-statistic for the one-way ANOVA test of $H_0$ is $P(F < F_{df1,\ df2})$, where $df_1$ and $df_2$ are the degrees of freedom and are equal to 1 and 592 in this study (2 tree species and 594 sampled trees), $F_{df1,\ df2}$ is the *F-critical* value with the degrees of freedom $df_1$ and $df_2$, and *F* is the observed value of the *F*-statistic [57]. For a given significance level, usually at $\alpha = 0.05$ level (so-called at 95% confidence level), the *F-critical* value is the *F* value above which $100\alpha\%$ of the null sampling distribution occurs [58]. Consulting the *F*-distribution table with 1 and 592 degrees of freedom, the *F-critical* value at $\alpha = 0.05$ level was obtained as 3.841. To retain the null hypothesis at 95% confidence level, the *F* value must be less than 3.841. For any one of the LiDAR-derived variables, if the *F* value is greater than 3.841 (the *P*-value is less than α), the null hypothesis will be rejected. This indicates that for this variable, the mean values from two tree species are significant different. In other words, one tree species can be distinguished from the other by this LiDAR-derived variable.

**Table 1.** LiDAR-derived structure and intensity variables and description.

| **Variable** | **Description** |
|---|---|
| MaxH | Maximum crown height |
| Depth | Depth (or extent) of tree crown |
| MeanH | Mean crown height |
| StdDev | Standard deviation of heights of laser returns within a crown |
| Density | Ratio of the number of laser returns in the crown to the total number of laser returns within the area defined by a crown diameter |
| MeanI | Mean intensity value of laser returns within a crown |
| StdDevI | Standard deviation of intensity value of laser returns within a crown |
| MeanIF | Mean intensity value of first laser returns within a crown |
| StdDevIF | Standard deviation of intensity value of first laser returns within a crown |

The RBF kernel is used to fit SVMs for the classification of Myrtle Beech and Silver Wattle using LiDAR-derived variables. The RBF kernel parameter $\gamma$ and the penalty term *C* must be properly determined to produce the best classification accuracy. The grid search algorithm (part of the SVM package LIBSVM) [43] was implemented to find out the suitable parameter values. A grid search over a bounded space (x, y) begins at one corner of the grid and is evaluated with cross-validation at every grid point separated by a value δ until the opposite corner of the grid is reached [29]. An exponentially growing sequences of *C* and $\gamma$ (for example, $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$, $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$) was recommended in practice [59]. Furthermore, in order to reduce the search time, the algorithm searches and evaluates the parameter values over a coarse grid (i.e., bigger δ). Once identifying a better region on the

grid, the search then focuses on a finer grid (smaller δ) on that region [29], [59]. The selected parameter values were used in SVMs for tree species classification. The procedures of selection of parameter values and implementation of SVMs for tree species classification with cross-validation were carried out using LiDAR-derived structure variables only, intensity variables only, and both structure and intensity variables. The classification results were compared with those from the LDAs.

## RESULTS AND DISCUSSION

The *F*-statistics and p-values from the one-way ANOVA for all LiDAR-derived variables are shown in Table 2. All the *p*-values from one-way ANOVA are less than 0.05. All the *F* values in the one-way ANOVA are greater than the *F*-critical value 3.841. Therefore, the null hypotheses in the one-way ANOVA for all 9 variables should be rejected. The results revealed that it is possible to discriminate two tree species if using any one of the 9 LiDAR-derived variables.

The overall accuracy of the classification results from the LDAs and the SVMs are shown in Fig. 2. It is observed that using LiDAR-derived structure variables, 83.2% of the individual trees of the Myrtle Beech and Silver Wattle were correctly classified using the LDA. If using only the intensity variables in the LDA, relatively low classification accuracy, 74.6% was obtained. However, if including both structure and intensity variables in the LDA, the overall classification accuracy increased from 83.2% (using only structure variables) to 86.4% (using both structure and intensity variables). The results of this study demonstrated the contribution of LiDAR-derived intensity variables to the identification of the Myrtle Beech and the Silver Wattle tree species at individual tree level. Although relatively low classification accuracy was obtained when only using LiDAR-derived intensity variables, the combination of both the structure and intensity variables in the discriminant analysis allowed individual Myrtle Beech and the Silver Wattle trees to be identified with high accuracy.

**Table 2.** Statistics results for variables from one-way ANOVA and independent t tests

| | **F** | **p-value** |
|---|---|---|
| **MaxH** | 192.894 | < 0.001 |
| **Depth** | 53.319 | < 0.001 |
| **MeanH** | 17.851 | < 0.001 |
| **StdDev** | 143.317 | < 0.001 |
| **Density** | 27.424 | < 0.001 |
| **MeanI** | 121.217 | < 0.001 |
| **StdDevI** | 90.191 | < 0.001 |
| **MeanIF** | 92.909 | < 0.001 |
| **StdDevIF** | 108.138 | < 0.001 |

In comparison with the results of the LDAs, 88.6% of the individual trees of the Myrtle Beech and Silver Wattle were correctly classified by the SVM if using LiDAR-derived structure variables only. An overall accuracy of 75.4% was achieved from the SVM classifier using the intensity variables only, without significant increase compared with the results from the LDA, indicating that the tree species classification results are not promising if using only the intensity variables no matter which classification approach (SVM or LDA) was used. If using both structure and intensity variables, however, the overall classification accuracy increased from 88.6% (using only structure variables) to 92.8%, also indicating the contribution of intensity variables to the improvement of classification results from the SVM classifier.

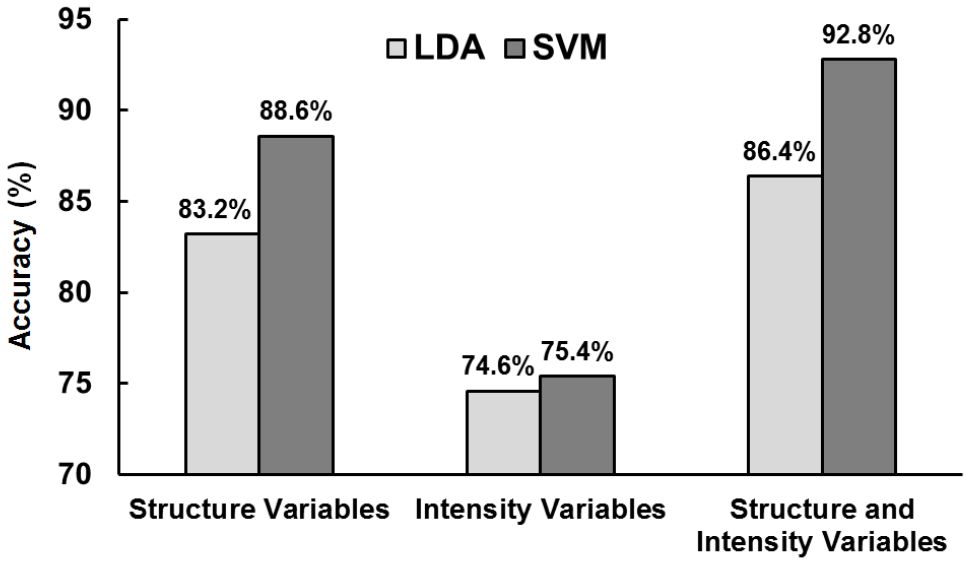



**Fig. 2.** Overall classification accuracy from support vector machines (SVM) and linear discriminant analysis (LDA) using LiDAR-derived structure variables only, intensity variables only, and both structure and intensity variables.

The acquisition of LiDAR intensity is usually at no extra cost in a LiDAR mission. Therefore, a growing attention has been paid to the application of these valuable intensity data. However, the LiDAR intensity values are affected by many factors such as the flying height, atmospheric conditions, directional reflectance properties, the reflectivity of the target, and the laser settings [60], [61], [24]. Accordingly, what each intensity dataset can offer to the analyst is not easy to define without at least preliminary exploration after due consideration of LiDAR intensity calibration. It is worth noting that so far most of studies used raw LiDAR intensity data for forest classification [16], [17], [24]. It is generally believed that as long as the relative intensity values are different between different forest types, the intensity values might be applicable for forest classification [62]. It is no doubt that if the intensity data were calibrated properly, there will be a considerable potential to use intensity to improve the application of LiDAR data in forest classification and tree species identification [63].

It is evident that the SVMs produced higher accuracy than the LDA did. Compared to the LDA method, significant increases in overall classification accuracy were observed by using LiDAR-derived structure variables only or using both structure and intensity variables in the SVMs. If using LiDAR-derived intensity variables only, the overall classification accuracy from the SVM is just slightly higher than that from the LDA. However, with combined use of the intensity and structure variables, the overall accuracy from the SVMs increased from 88.6% to 92.8%, indicating the contribution of the intensity variables to the improvement of classification results.

## CONCLUSIONS

This study examined the applicability of LiDAR-derived terrain and canopy surface models for the identification of the location and crown size of individual trees for species classification using LiDAR structure and intensity variables. It demonstrated the success of the SVMs for the identification of the Myrtle Beech (the dominant species of the Australian cool temperate rainforest in the study area) and adjacent tree species – notably, the Silver Wattle at individual tree level using LiDAR-derived structure and intensity variables. An overall accuracy of 92.8% was achieved from the SVM approach. Compared to the overall accuracy of 86.4% from the linear discriminant analysis, it is evident that the SVMs have significant advantages over the traditional classification methods such as the LDA method in terms of classification accuracy. Although the overall accuracy of classification results from both SVM and LDA was relatively low when just using the intensity variables in the analysis, combination of the structure and intensity variables in the discriminant analysis did improve the accuracy of classification results, indicating the contribution of the LiDAR intensity variables to the classification results.

## ACKNOWLEDGEMENTS

We would like to thank the HVP Plantations Pty Ltd for providing the LiDAR and EVC data. We also thank Associate Professor Sorin Popescu at Texas A&M University for providing the TreeVaW software for use in this study.

## REFERENCES

[1] Liu, X. (2011) Accuracy assessment of LiDAR elevation data using survey marks. Survey Review, 43(319):80-93.

[2] Popescu, S. C., Wynne, R. H. and Nelson, R. F. (2002) Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. Computers and Electronics in Agriculture, 37:71-95.

[3] Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C. and Gerard, P. D. (2003) Characterizing vertical forest structure using small-footprint airborne LiDAR. Remote Sensing of Environment, 87:171-182.

[4] Goodwin, N. R., Coops, N. C. and Culvenor, D. S. (2006) Assessment of forest structure with airborne LiDAR and the effects of platform altitude. Remote Sensing of Environment, 103:140-152.

[5] Orka, H. O., Naesset, E. and Bollandsas, O. M. (2009) Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sensing of Environment, 113:1163-1174.

[6] Suratno, A., Seielstad, C. and Queen, L. (2009) Tree species identification in mixed coniferous forest unsing airborne laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing, 64:683-693.

[7] Vauhkonen, J., Tokola, T., Packale, P. and Maltamo, M. (2009) Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. Forest Science, 55(1):37-47.

[8] Junttila, V., Kauranne, T. and Leppanen, V. (2010) Estimation of forest stand parameters from airborne laser scanning using calibrated plot databases. Forest Science, 56(3):257-270.

[9] Korpela, I., Orka, H. O., Maltamo, M., Tokola, T. and Hyyppa, J. (2010b) Tree species classification using airborne LiDAR - effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica, 44(2):319-339.

[10] Korhonen, L., Korpela, I., Heiskanen, J. and Maltamo, M. (2011) Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. Remote Sensing of Environment, 115(4):1065-1080.

[11] Zhang, Z., Liu, X., Peterson, J. and Wright, W. (2011) Cool temperate rainforest and adjacent forests classification using airborne LiDAR data. Area, 43(4):438-448.

[12] Holmgren, J., Nilsson, M. and Olsson, H. (2003) Estimation of tree height and stem volume on plots using airborne laser scanning. Forest Science, 49(3):419-428.

[13] Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, X. and Maltamo, M. (2008) Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing, 29(5):1339-1366.

[14] Hawbaker, T. J., Gobakken, T., Lesak, A., Tromborg, E., Contrucci, K. and Radeloff, V. (2010) Light detection and ranging-based measures of mixed hardwood forest structure Forest Science, 56(3):313-326.

[15] Lindberg, E., Holmgren, J. and Olofsson, K. (2010) Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. International Journal of Remote Sensing, 31(5):1175-1192.

[16] Brandtberg, T., Warner, T. A., Landenberger, R. E. and McGraw, J. B. (2003) Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment, 85:290-303.

[17] Holmgren, J. and Persson, A. (2004) Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90:415-423.

[18] Popescu, S. C. and Wynne, R. H. (2004) Seeing the trees in the forest: using Lidar and multispectral data fusion with local filtering and variable window size for estimation tree height. Photogrammetric Engineering and Remote Sensing, 70(5):589-604.

[19] Heurich, M. and Thoma, F. (2008) Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (Ragus sylvatica) and Norway spruce (Picea abies) forests. Forestry, 81(5):645-661.

[20] Wehr, A. and Lohr, U. (1999) Airborne laser scanning - an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing, 54(2-3):68-82.

[21] Liu, X. (2008) Airborne LiDAR for DEM generation: some critical issues. Progress in Physical Geography, 32(1):31-49.

[22] Korpela, I., Orka, H. O., Hyyppac, J., Heikkinen, V. and Tokola, T. (2010a) Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. ISPRS Journal of Photogrammetry and Remote Sensing, 65:369-379.

[23] Lillesand, T. M., Kiefer, R. W. and Chipman, J. W. (2008) Remote Sensing and Image Interpretation. John Wiley, New York.

[24] Kim, S., McGaugher, R. J., Andersen, H.-E. and Schreuder, G. (2009) Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sensing of Environment, 113:1575-1586.

[25] Moffiet, T., Mengersen, K., Witte, C., King, R. and Denham, R. (2005) Airborne laser scanning: exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. ISPRS Journal of Photogrammetry and Remote Sensing, 59:289-309.

[26] Brandtberg, T. (2007) Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS Journal of Photogrammetry and Remote Sensing, 61:325-340.

[27] Foody, G. M. and Mathur, A. (2004a) A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing, 42(6):1335-1343.

[28] Huang, C., Davis, L. S. and Townshend, J. R. G. (2002) An assessment of support vector machines for land cover classification. Remote Sensing of Environment, 23(4):725-749.

[29] Tso, B. and Mather, P. M. (2009) Classification Methods for Remotely Sensed Data. CRC Press, Boca Raton, FL.

[30] Heikkinen, V., Korpela, I., Tokola, T., Honkavaara, E. and Parkkinen, J. (2011) An SVM classification of tree species radiometric signatures based on the Leica ADS40 Sensor. IEEE Transactions on Geoscience And Remote Sensing, 49(11):4539-4551.

[31] Mountrakis, G., Im, J. and Ogole, C. (2011) Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66:247-259.

[32] Koetz, B., Morsdorf, F., van der Linden, S., Curt, T. and Allgower, B. (2008) Multi-source land cover classification for forest fire management based on imaging spectromery and LiDAR data. Forest Ecology and Management, 256:263-271.

[33] Korpela, I., Koskinen, M., Vasander, H., Holopainen, M. and Minkkinen, K. (2009) Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. Forest Ecology and Management, 258:1549-1566.

[34] Angelo, J. J., Duncan, B. W. and Weishampel, J. F. (2010) Using lidar-derived vegetation profiles to predict time since fire in an Oak Scrub landscape in East-Central Florida. Remote Sensing, 2:514-525.

[35] Chen, G. and Hay, G. J. (2011) A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and quickbird data. Photogrammetric Engineering and Remote Sensing, 77(7):733-741.

[36] García, M., Riano, D., Chuvieco, E., Salas, J. and Danson, F. M. (2011) Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. Remote Sensing of Environment, 115:1369-1379.

[37] Zhao, K., Popescu, S., Meng, X., Pang, Y. and Agca, M. (2011) Characterizing forest canopy structure with llidar composite metrics and machine learning. Remote Sensing of Environment, 115:1978-1996.

[38] Vapnik, V. N. (1998) Statistical Learning Theory. John Wiley & Sons, New York.

[39] Vapnik, V. N. (2000) The Nature of Statistical Learning Theory. Springer-Verlag, New York.

[40] Foody, G. M. and Mathur, A. (2004b) Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sensing of Environment, 93:107-117.

[41] Abe, S. (2010) Support Vector Machines for Pattern Classification. Springer-Verlag, London, UK.

[42] Scholkopf, B. and Smola, A.J. (2002) Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.

[43] Chang, C. C. and Lin, C. J. (2011) LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(27):1-27.

[44] Davies, J. B., Oates, A. M. and Trumbull-Ward, A. V. (2002) Ecological vegetation class mapping at 1:25000 in Gippsland. Victorian Department of Natural Resources and Environment, Melbourne, Australia.

[45] DSE. (2005) EVC Benchmarks - Strzelecki Ranges bioregion. Victorian Department of Sustainability and Environment, Melbourne, Australia.

[46] Noble, W. S. (1978) The Strzeleckis: a New Future for the Heartbreak Hills. Victoria Forests Commission, Melbourne, Australia.

[47] Gullan, P. K., Earl, G. E., Forbes, S. J., Barley, R. H. and Walsh, N. G. (1984) Sites of botanical significance in Central Gippsland. Department of Conservation, Forests and Lands, Melbourne, Australia.

[48] Legg, S. (1986) Farm abandonment in South Gippsland's Strzelecki Ranges, 1870-1925: challenge or tragedy. Gippsland Heritage Journal, 1(1):14-22.

[49] Reichl, P. (1966) Mountain forests of Gippsland. Thomas Nelson (Australia) Ltd, Melbourne, Victoria, Australia.

[50] Langkamp, P. (1987) Germination of Australian native plant seed. Inkata Press, Melbourne, Australia.

[51] Littlejohn, E. J. (1978) Seventy years around Ryton Junction in the Strzelecki Ranges: 1904 virgin forest, 1920-30 farmland, 1977 replanted forest. Enterprise Press, Sale, Victoria, Australia.

[52] Boyle, C. and Lowe, K. W. (2004) Biodiversity action planning strategic overview for the Strzelecki Ranges bioregion. Victorian Department of Sustainability and Environment, Melbourne, Australia.

[53] Popescu, S. C., Wynne, R. H. and Nelson, R. F. (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Canadian Journal of Remote Sensing, 29(5):564-577.

[54] Landau, S. and Everitt, B. S. (2004) A Handbook of Statistical Analyses using SPSS. CRC Press LLC, Boca Raton, London, New York and Washington, D.C.

[55] Burns, R. P. and Burns, R. (2008) Business Research Methods and Statistics using SPSS. SAGE Publications Ltd, London.

[56] Walford, N. (2011) Practical Statistics for Geographers and Earth Scientists. Wiley-Blackwell, Oxford, UK.

[57] Cabrera, J. and McDougall, A. (2002) Statistical Consulting. Springer, New York, Berlin and Heidelberg.

[58] Seltman, H. J. (2010) Experimental Design and Analysis. Carnegie Mellon University, Pittsburgh, PA, USA.

[59] Hsu, C.-W., Chang, C. C. and Lin, C. J. (2010) A practical guide to support vector classification. Department of Computer Science, National Taiwan University, Taipei.

[60] Baltsavias, E. P. (1999) Airborne laser scanning: basic relations and formulas. ISPRS Journal of Photogrammetry and Remote Sensing, 54:199-214.

[61] Liu, X., Zhang, Z., Peterson, J. and Chandra, S. (2007) LiDAR-derived high quality ground control information and DEM for image orthorectification. Geoinformatica, 11(1):37-53.

[62] Lovell, L. L., Jupp, D. L. B., Culvenor, D. S. and Coops, N. C. (2003) Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Canadian Journal of Remote Sensing, 29(5):607-622.

[63] Donoghue, D. N. M., Watt, P. J., Cox, N. J. and Wilson, J. (2007) Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. Remote Sensing of Environment, 110:509-522.

## Author Index