Automated Interpretation of Soils on Agricultural Lands of Belarus Based on High-Resolution Multispectral Satellite Imagery

Siarhei Myshliakou

Research Republican Unitary Enterprise on Land Management, Geodesy and Cartography "BelNICzem",
Kazintsa str. 86/3,
220108, Minsk, Belarus
sergey_myshl@mail.ru

Abstract. High-resolution multispectral space images QuickBird and Ikonos substantially increase possibilities of land cover/land use remotely studies. Researches and construction of the automated technology of large-scale soil-mapping based on remote sensing data are carried out in Republican Unitary Enterprise "BelNICzem". Basic approaches of multispectral high-resolution imagery processing and thematic interpretation for soil mapping are reviewed in the article. Possibilities of remote studies of water erosion, soil moisture and peat degradation in reclaimed areas are shown by examples of tested areas, located in Braslav and Soligorsk districts of the Republic of Belarus. Major indication properties of soils for automated recognition are considered.

Keywords: Remote sensing, geographic information systems, multispectral imagery, space images, soils, soil maps, automated processing of remote sensing data, recognition, interpretation, mapping.

1 Introduction

Soil maps are important sources of information for efficient land management and land policy. They are being used in such activities as cadastral assessment of agricultural lands, reclamation, agrochemistry and precise agriculture. Soil maps play important role in land management activities directed to abatement of land degradation.

Remote sensing data (RSD) and GIS technologies are powerful tools for advanced soil mapping. There are many works considering applying of RSD for soil properties studies. But problems of digital soil mapping methodologies based on RSD aren't studied well [Kovalev, 1995, Peng, 2003, McBratney, 2003, Sommer, 2003]. Thematic interpretation of images is the principal step of RSD processing need to be done when soil maps are created or updated. Interpretation consists of recognition and assignment of semantic data to contours recognized on image [Guideline…, 1986]. Main complexities of satellite mapping of soils relate just with this process. Visual interpretation of images is subjective. Quality of outputs completely depends on how experienced and skilled an operator is. Automated interpretation based on pattern recognition; therefore it is error-prone regards to reading information encoded in image pixels.

However permanent improving of spatial information processing techniques excludes other methods except automated interpretation of images. When automated processing of images executes main duties of an operator are quality control and manual editing (if necessary). Acceptable result is achieved due to numerous approbations of chosen techniques and algorithms on tested areas.

The main objective of this work is developing the automated technology of satellite image interpretation for the purposes of soil mapping of agricultural lands. Direct research tasks are:

- determination the most appropriate techniques of imagery thematic processing for detection the local features of study areas soil cover;
- testing the techniques of thematic processing of images;
- examination and formalization interpretation keys of study areas.

2 Study areas, input data and preprocessing of images

Interpretation of soils related to different landscapes has specifics depending on physiographic conditions and degree of anthropogenic influence to environment. Methods of landscape indication play key role for interpretation of soils in natural and semi-natural areas such as forests, river valleys, undrained peat bogs [Obuhovskij, 1994, 2005, Weiers, 2004]. For open soils of agricultural lands with no vegetation cover interpretation executes mainly by direct interpretation keys such as hue, color, texture, size and shape of contours. In general the last way provides more accurate results of interpretation in comparison with indication which is always probabilistic. However soil is a complex substance with changeable properties in space and time. This fact determines difficulties in soil contours delineation and thematic interpretation of extracted contours.

Subject of the research are soils of tested areas situated in Braslav district of Vitebsk region (Rozhevo tested area, 810 hectares) and in Soligorsk district of Minsk region (Velichkovichi tested area, 4000 hectares). Masking the territory of agricultural lands was carried out with the National land information system. Masks of agricultural lands include arable lands, fallow lands, grasslands and pastures. Thematic processing of satellite images have been carried out within the bounds of masks.

Both polygons are characterized by specific landscape conditions. Tested area Rozhevo is characterized by dominating of washed loamy luvisols with glacial deposits underlying. Small sizes and complex shapes of pieces of lands and lakes as well as diversity of soils are consequences of high degree of terrain irregularity caused by glacial landforms. Lands are exposed to water erosion; humus horizon in many places is completely washed. Tested area Velichkovichi is a drained peat bog characterized by dominating of hemic histosols and degraded peat-mineral and postpeat soils with fluvioglacial sands and sandy loams underlying. Irreversible modifications in water-physical regime of soils lead to mineralization of peat and drastic changes in soil cover [Romanova, 2004].

The research is directed to making and updating soil maps at a scale 1:10 000. To make such maps RSD of very high resolution is needed [Kravtsova, 2005]. Satellite images of very high resolution QuickBird (received July, 3, 2006) and Ikonos (received September, 17, 2006) were used in work. Spatial resolution of QuickBird is 0,6 m in panchromatic band and 2,4 m in spectral bands. Ikonos has a spatial resolution of 1 and 4 m in panchromatic and spectrozonal modes respectively. Both sensors have sufficient spatial resolution to satisfy accuracy requirement to maps at a scale 1:10 000.

Acquired satellite images need to be preprocessed. Input images were georeferenced and orthorectified in UTM projection and WGS-84 coordinate system. Histogram’s transformations of brightness and contrast (equalization, normalization, and manual correction) were executed with the purpose of spatial enhancement of images. Input digital numbers of pixels were modified by sum high-frequency filtering and median filtering.

3 Thematic processing of images and identification of soils by spectral properties

Automated interpretation is based on classification of objects according their reflectance capacity. Classification of multiband images implies dividing into classes in multidimensional feature space. A dimension of this feature space is determined by number of analyzed image bands. Each pixel of multiband image with a set of spectral values is a point in multidimensional feature space [ERDAS…, 1999]. Modification of feature space supposed to be optimal way to increase interpretation accuracy. Bands with high degree of recognition abilities of analyzed objects include in multidimensional feature space whereas low-informative bands exclude from it. Bands which can substantially enhance quality of interpretation are derivative pseudoimages produced by multivariate statistical analysis and mathematical transformations of feature spaces. Following types of thematic image processing were tested in the course of the work: principal component analysis, “tasseled cap” transformation, NDVI calculation. Each method of thematic processing allows to reveal some features of landscapes and structure of soil cover of study areas.

Satellite image of Rozhevo tested area illustrates state of landscape in the beginning of July. Crops in various vegetative stages as well as grassland vegetation have the primary influence to spectral reflectance and digital number values in this season. Spectral reflectance of soils disable for observing, hence interpretation with direct interpretation keys is impossible. In this case finding appropriate indirect indicators and selection the most informative images – are principal tasks need to be solved for successful interpretation.

Recognition of water erosion patches on loamy soils of Rozhevo tested area is possible for grasslands. Water erosion characterizes by humus horizon thinning and organic matter capacity decreasing due to washing and sedimentation in lowlands. In accordance with rate of water erosion progress soils is divided into untruncated, aggradational, weakly-, moderately- and heavily truncated [Field investigation…, 1990]. Prior assumption about dependence between spectral brightness of pixels and degree of soils washing was taken as indirect interpretation key. Such dependence reveals on the image received in summer time due to state of vegetation. Sparse or suppressed vegetation cover is the cause of increasing spectral brightness in visual bands (0,45 – 0,69 mcm). The type of vegetation indicates patches of progressive water erosion. Dense vegetation cover determines low spectral reflectance in visual bands and indicates untruncated and aggradational soils. Pseudoimages derived by algorithms of thematic processing develop prior assumption described above.

Image of Velichkovichi tested area received in autumn time is characterized by open ground agricultural fields. A total area of open soils with no vegetation on the satellite image amounts 1046 hectares. Visual analysis reveals light patches of mineral soils and degraded peats on a dark background of drained peats of various thicknesses. One of the main tasks for this case study is separation of degraded peats from mineral soils (luvisols and albeluvisols). These two soil types are visually similar. The last ones are associated with former mineral islands among pet bog and with sandy bars of former floodplain. Degraded peats are characterized by topsoil thinning, mineralization of peats and wind erosion (deflation). All these negative processes lead to decreasing of organic matter capacity and increasing of ash value in peat. Organic matter capacity is a major factor determining spectral reflectance of soils [Kravtsova, 2005, Selige, 2005].

A regression analysis verifies spectral dependence on ash value of peat and humus capacity of soils. The analysis was performed for 19 soil samples which have been taken during field works in September 2008. Ash value of peat and humus capacity was measured in laboratory. The highest pair correlation is revealed for green band of Ikonos sensor system (0,52 – 0,61 mcm). Pearson’s correlation coefficient is 0,718 for all open soils in Velichkovichi tested area including peat, degraded and mineral soils. Executed analysis has revealed a nonlinear dependence of spectral reflectance on organic matter capacity with determination coefficient R^2 = 0,769. The dependence describes by inverse quadratic function with five regression’s coefficients.

Since mineral soils and peat influence spectral properties in different manner more detailed analysis was performed for peat and degraded peat soils only. Key site of small area (80 by 50 meters) was selected in east part of Velichkovichi tested area. 40 soil samples were taken regularly with a range of 10 meters across the area of key site. The image of 1st principal component was determined as the most appropriate for analysis with Pearson’s correlation coefficient equal 0,901. Exponential function describes the dependence of spectral reflectance on organic matter content with determination coefficient R^2 = 0,769 and two regression’s coefficients.

The technology based on regression analysis provides automated mapping of mineral soils, peat with no degradation signs (more than 50% of organic matter), degraded peat-mineral soils (20 – 50% of organic matter) and degraded sandy peat and mineral postpeat soils (less than 20% of organic matter).The most appropriate images for abovementioned soils extraction are red (0, 64 – 0,72 mcm) and green (0,52 – 0,61 mcm) bands of Ikonos sensor system, pseudo images of 1nd principal component and brightness index of “tasseled cap” transformation. Except organic mater capacity red and green bands provide a possibility of extraction peaty-gley and peat-gley soils notable for weak thickness of peat horizons (less than 30 cm and less than 50 cm respectively). At the same time recognition and extraction of peat soils with more than 50 cm of peat horizons thickness is impossible [Kravtsova, 2005, Kovalev, 1999]. Principal component analysis provides an extraction of mineral soils according their moisture content as well.

4 Analysis and formalization of soils interpretation keys

Accuracy of interpretation strongly depends on a set of interpretation keys being used. None of automated interpretation techniques can set off the effect of poorly chosen criteria for objects extraction. The principal tasks of automated interpretation are formalization of interpretation keys and algorithmization of the process. Formalization of interpretation keys assumes their representation in digital view suitable for computer data processing. Automated interpretation of soils requires visual and descriptive characteristics being transformed into a set of variables for mathematical and statistical operations.

Since such interpretation keys as shape, size and texture describing landscape's view are complicatedly formalized, automated interpretation of soils has been carried out by spectral parameters only. Information about spectral properties of soils was obtained on the basis of field works results. 27 soil profiles were digged out in Velichkovichi tested area for precise diagnostics of soils. In Rozhevo tested area 36 profiles were digged out. Besides field reconnoitring of soils was executed for determination areas with obvious soils. All places of profiles lying as well as reconnaissance routes were positioned with GPS.

The interpretation process requires establishing relations between physical conditions of the surface and spectral properties [Tompkins, 1997]. Thematic classifiers need to be composed before interpretation process. Training the classifier consists of construction and development the knowledge base and the database considering study objects (soils in this case). Classifier can be trained on the basis of priori values of variables corresponding to certain soils. The other way of classifier's training is signature or training samples collection on the image. The first case assumes applying spectral libraries describing characteristics of soils in digital manner. In the second case operator creates input data for interpretation in interactive mode. The operator should be aware with soil cover structure of study area to delineate referenced samples on image with maximal reliability.

Since optical properties of landscapes change permanently, construction the universal spectral library of soils is hardly possible. In visual interpretation process tables of interpretation keys and landscape indicative tables are used for interpretation key definition as shown in works [Obuhovskij, 1994, Kovalev, 1995]. A direct training of classifiers is more advisable for this case study.

Representativeness analysis of training samples is compulsory stage of the work. Representativeness of training samples is achieved when next conditions are accomplished:

- adequate area;
- plurality of samples;
- stability of spectral reflectance values between image bands inside the training sample;
- spectral homogeneity of the training sample.

Abovementioned conditions hold true if initial samples are chosen correctly and their statistical analysis is performed. General framework of classifier's training is shown on the fig. 1. Statistical analysis allows to enhance representativeness of signatures excluding untypical and error values from classifier. Besides there is an opportunity of approximate estimation the efficiency of interpretation.

Statistical analysis of recognition abilities of soils has two goals. The first one is determination bands and derived pseudoimages suitable for interpretation. The second one is determination interpretation keys which can be potentially used in an automated interpretation process. Interpretation keys are formalized in accordance with threshold values of digital numbers for each class. In this research threshold values set up using standard deviation parts for each class

5 Classification, accuracy assessment and map making

Classification is the next step of automated processing of images. It is well known there are no universal approaches to classification of images. Unsupervised, supervised and expert object-based classifications were tested in the research. Satisfied results have been derived when maximal likelihood method of supervised classification was applied. Accuracy assessment was performed for classification results. 76 samples with known soils were selected for checking in the course of field works in Velichkovichi tested area. All of them were positioned with GPS. Error matrix was constructed in order to estimate the accuracy of classification results (table 1).

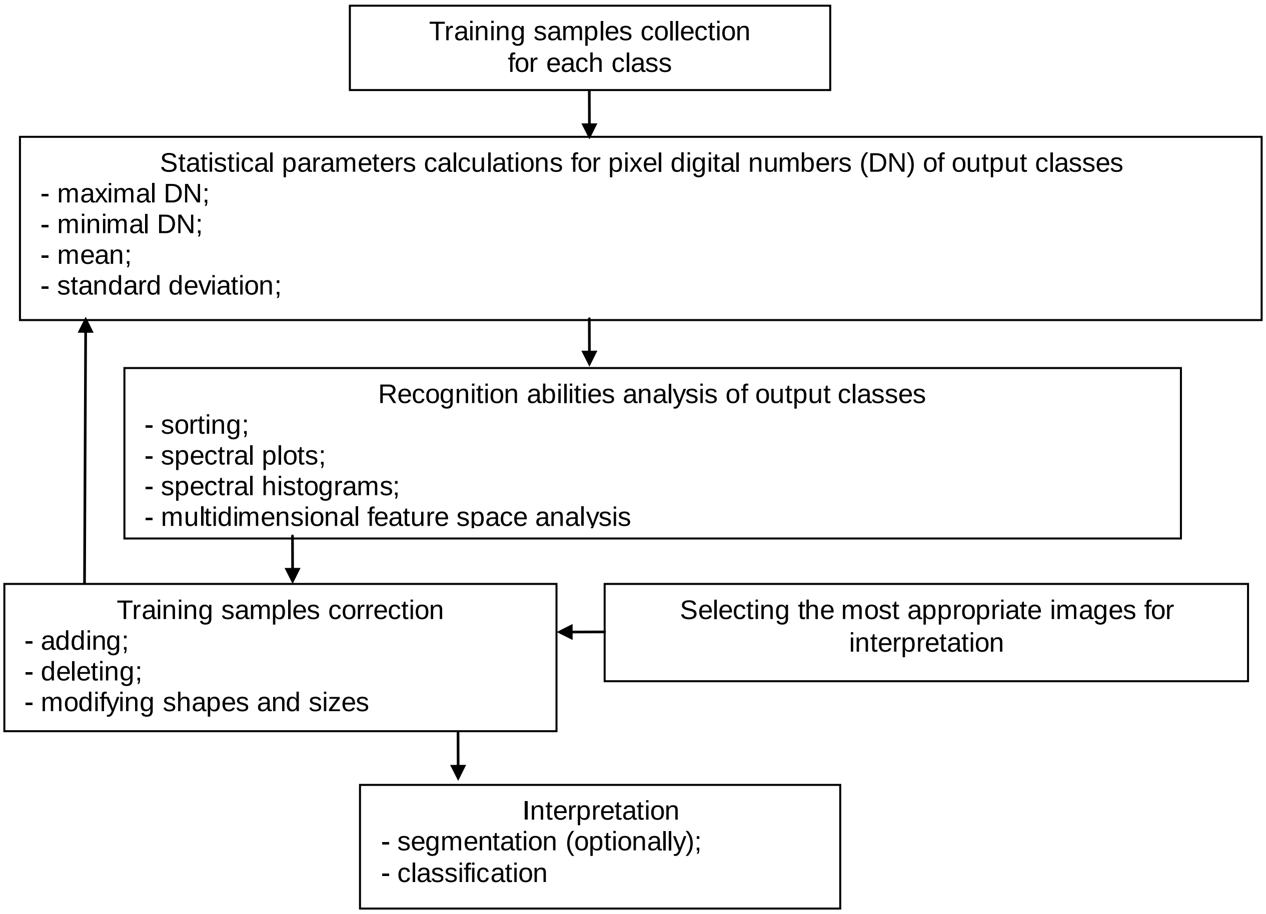


Fig. 1. General framework of classifier training

The total producer's accuracy is 65,8% and 59,4% with Kappa coefficient. In general it can be considered as satisfied result because soil is a complex substance with no exactly delineated borders between types or any other taxonomic units. Various soils with different properties are often similar spectrally and spatially. And the same soils may look differently due to any temporal conditions and external factors. Sod-gleyey and sod-gley soils which are notable for gley horizon thickness and moistening conditions have the lowest accuracy (31,3% and 25,6% respectively). Therefore accuracy level can be increased if two classes of gleysols will be merged into joint class. Besides these two classes are not widely spread on the study area and cannot be adequately represented in interpretation process. Accuracy assessment reveals high degree of reliability of peat (83,3%), degraded peat (69,9%) and mineral soils (69,9%). Some inaccuracies take places when peat-gley soils separate from peat due to difficulties of peat thickness recognition. Inaccuracies in two classes of degraded soils appear because in some cases degraded soils (especially of high degree of peat mineralization) are similar to mineral ones (luvisols, albeluvisols and gleysols).

Classified images usually contain noises and difficult for visual perception. Hence they need to be generalized. Generalization was performed in automatic mode; general aims of generalization of soil contours are:

- deleting small contours causing noise on the image;
- eliminating contours;
- smoothing and simplifying contours.

Vectorization and map designing are final stages of the process. After vectorization outputs need to be generalized again in order to get maps satisfied the accuracy requirements at 1:10 000 scale. Legend and map design are chosen randomly since output results of the work are input data for the official soil map creating and updating. The fragment of output soil map of Velichkovichi tested area is shown on the fig. 2.

Table 1. Error matrix of classification results of Ikonos satellite image (Velichkovichi tested area)

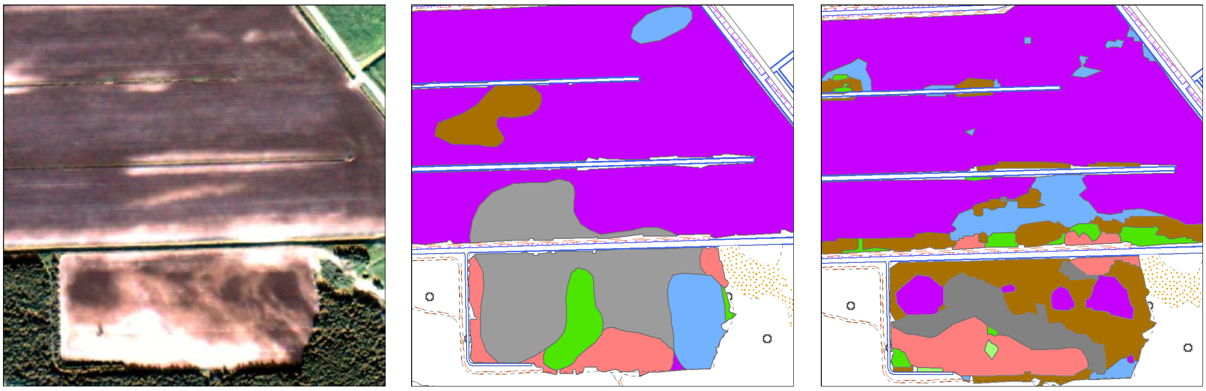
Ground truth data / Recognition results	Sod-podzol (luvisols and albeluvisols)	Sod-gleyey (gleysols)	Sod-gley (gleysols)	Peat-gley (sapric histosols)	Peat (hemic histosols)	Degraded peat-mineral (antrosols)	Degraded sandy peat and mineral postpeat (antrosols)	Total	%	% with Kappa
Sod-podzol (luvisols and albeluvisols)	5		1				1	7	71,4	69,9
Sod-gleyey (gleysols)		1					2	3	33,3	31,3
Sod-gley (gleysols)		2	2			3		7	28,6	25,6
Peat-gley (sapric histosols)				7	6			13	53,8	48,1
Peat (hemic histosols)				3	22			25	88,0	83,3
Degraded peat-mineral (antrosols)			1	1		9	1	12	75,0	69,9
Degraded sandy peat and mineral postpeat (antrosols)						5	4	9	55,6	39,6
Total	5	3	4	11	28	17	8	76	65,8	59,4

6 Conclusion, summary

Interpretation of RSD is a laborious process. Automation of this process means hierarchical approach and sequential execution of operations. Therefore necessity of algorithmization and models construction appears. By means of special development environments such as Model Builder for ArcGIS and Spatial Modeler for ERDAS IMAGINE models of preprocessing, thematic processing, classification, generalization and vectorization of images were developed. Models are constructed in a manner when output data of previous process become input data for next process. Hence uninterrupted execution of commands and minimization of operator's work are provided. Input data for models are: land cover layer from National land information system, 4 spectral bands of QuickBird or Ikonos image and layer of training samples. Vector layer of soil map is an output of modeling. Operator's work is required just on the stage of recognition ability analysis of output classes' training samples.

Potential of multispectral high-resolution satellite images for soil mapping and soils properties examination is shown in the research. For areas with high degree of terrain irregularity in Belarusian Poozerje (Rozhevo tested area) a visual interpretation of water erosion is possible but mapping cannot be realized because of spectral similarity of signatures describing different types of eroded soils. For open vegetation-free arable lands situated in reclaimed areas (Velichkovichi tested area) multispectral satellite images of very high resolution give a possibility of estimation the degree of peat degradation process. Besides, mapping of several types of soils is possible with sufficiently high accuracy (59,4% of total producer's accuracy for Velichkovichi tested area).

GIS-based technologies of RSD processing increase an efficiency of digital soil mapping first of all in part of precision of soil contours localization. Besides time consuming for field surveys and map making decreases three times on average (depends on soil cover complexity).



a) b) c)

Legend

Soils
sod-podzol and sod-podzol boggy (luvisols and albeluvisols)
sod-gleyey (gleysols)
sod-gley (gleysols)
peaty-gley and peat-gley (sapric histosols)
peat (hemic histosols)
degraded peat-mineral (antrosols)
degraded sandy peat and mineral postpeat (antrosols)

Fig. 2. Result of automated interpretation of Ikonos satellite image (Velichkovichi tested area)
a) fragment of Ikonos satellite image, 17 September 2006;
b) fragment of digitized soil map of “Novoe Polessje” agriculture produser’s co-operative (Belgiprozem, 2004);
c) fragment of digital soil map as an output of automated interpretation process of Ikonos satellite image.

Further researches on this subject will be directed to optimization of model construction, model adaptation for various RSD (various in spatial and spectral resolution, acquiring time, thematic content etc), applying of digital elevation models in interpretation process. The described technology will be tested in other places in Belarus characterized by land degradation processes.

Reference

ERDAS Field Guide. ERDAS Inc, Atlanta, Georgia, 1999.

Field investigation and mapping of soils of BSSR. Uradzhaj, Minsk, 1990 (In Russian).

Guideline of aerophotographic materials using in large-scale mapping of soils. Belarusian Research Institute of soil science and agrochemistry, Republican planning institute on land management “Belgiprozem”, Minsk, 1986 (In Russian).

Kovalev A.A. [et al] *Remote mapping of environment.* Institute of geological science of National Academy of science of Belarus, Minsk, 1995 (In Russian).

Kovalev A.A., Nichiporovich Z.A., Shalkevich F.E., Krivonos O.V. *Methodology of remote diagnostics of meliorated peat soils of Polessje on the basis of aerial photography,* Minsk, 1999 (In Russian).

Kravtsova V.I. *Space methods of soil study.* AspectPress, Moscow, 2005 (In Russian).

McBratney A.B., Mendoca Santos M.L., Minasny B. *On digital soil mapping.* Geoderma 117 (2003) p. 3-52.

Obuhovskij J.M., Gubin V.N., Marcinkevich G.I. *Aerospace researches of Belarusian landscapes.* Navuka i tekhnika, Minsk, 1994 (In Russian).

Peng W., Wheeler D.B., Bell J.C., Krusemark M.G. Delineating patterns of soil drainage class on bare soils using remote sensing analyses. Geoderma 115 (2003), p. 261-279.

Romanova T.A. *Diagnostics of soils of Belarus and their classification in FAO-WRB,* Belarusian Institute of soil science and agrochemistry, Minsk, 2004 (In Russian)

Selige T., Bohner J., Schmidhalter U. *High resolution topsoil mapping using hyperspectral image and fielddata in multivariate regression modeling procedures*. Geoderma 136 (2006), p. 235-244.

Sommer M. [et al] *Hierarchical data fusion for mapping soil units at field scale.* Geoderma 112 (2003), p. 179-196.

Tompkins S., Mustard J. F., Pieters J.M., Forsyth D.W. *Optimizations of Endmembers for Spectral Mixture Analysis.* Remote Sensing of Environment 59 (1997), p. 472-489.

Vireling A. *Satellite remote sensing for water erosion assessment: A review.* Catena 65 (2006), p. 2-18.

Weiers S., Bock M., Wissen M., Rossner G. *Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS method.* Landscape and Urban Planning 67 (2004) 43–65