

# Predikce biomasy lesních porostů pomocí metod dálkového průzkumu Země, Google Earth Engine a pokročilých metod strojového učení

Petr Lukeš  
Ústav výzkumu globální změny AV ČR

**Spoluautoři / Co-authors:** Olga Brovkina, Emil Cienciala, Růžena Janoutová

**Sekce / Topic:** Dálkový průzkum Země

**Abstrakt:** Lesy důležitou roli při fixaci atmosférického uhlíku ve formě biomasy, čímž zmírňují globální změnu klimatu. Lesní biomasa a její růst se tradičně měří polními metodami a statistickým vzorkováním reprezentativních lesních ploch. Typicky se měří průměr kmene v prsní výšce (DBH) a výška stromu (H) pro každý strom ve zkoumaném území, přičemž biomasa je vypočtena pomocí druhově specifických allometrických rovnic. Tyto metody jsou relativně přesné (v závislosti na použité allometrické rovnici), ale pracné (inventarizace lesních ploch na úrovni celého státu vyžaduje vzorkování tisíců lesních ploch) a jsou reprezentativní pouze k danému období. Kvůli statistickému vzorkování bodů jsou odhady biomasy platné pouze pro větší regionální jednotky (např. okresy, regiony nebo dokonce státy) a nemohou na úrovni porostů popsat lokální variabilitu biomasy. Data dálkového průzkumu poskytují neocenitelné informace o našich ekosystémech, pozemních i vodních. Satelitní dálkový průzkum je zvláště vhodný pro systematické a opakované monitorování ekosystémů. Současné satelitní snímací systémy pro monitorování životního prostředí provozované NASA (Landsat) a ESA (Sentinel) jsou schopny pozorování na úrovni jednotlivých lesních porostů ve více spektrálních kanálech pokrývajících viditelné a infračervené pásma, které jsou potenciálně citlivé na lesní biomasu a její změny. Aktivní radarový systém (SAR) provozovaný na satelitech Sentinel-1 je navíc atraktivním zdrojem údajů o dálkovém průzkumu pro jeho schopnost snímkovat povrch Země za každého počasí. Potenciál dat družicového dálkového průzkumu lze plně využít pouze při práci s časovou řadou pozorování. To nevyhnutelně představuje výzvu kvůli velkému objemu dat a výpočetním prostředkům potřebným pro jejich zpracování. Novým trendem ve zpracování dat dálkového průzkumu Země je cloudové zpracování, kde archiv dat i algoritmy dat pro jejich zpracování jsou prováděny na dálku na vysoce výkonných počítačích. Při vývoji predikčního modelu mezi jakoukoli pozorovanou veličinou dat dálkového průzkumu (odrazivost povrchu a jejich odvozené transformace, koeficient zpětného rozptylu radarových dat) a proměnnými souvisejícími s lesní biomasou může výběr konkrétního statistického modelu ovlivnit predikční výkon a závěry o použitelnosti konkrétních metod. V posledních desetiletích byly vyvinuty četné statistické metody, mimo jiné varice klasického lineárního regresního modelování, umělé neuronové sítě, Support vector machines nebo rozhodovací stromy. V této studii používáme přístup cloudového zpracování v aplikaci Google Earth Engine ke zpracování a interpretaci bohaté časové řady jak optických (Landsat-8 a Sentinel-2), tak radarových (Sentinel-1) dat. Díky tomu získáváme vysoce kvalitní prediktory lesní biomasy. Byly využito několik databází lesní inventarizace, včetně inventury biomasy prováděné v rámci inventarizace krajiny CzechTerra, defoliace porostů z panevropské databáze ICPForests a vlastního měření Indexu listové plochy. Různé pokročilé statistické metody strojového učení byly využity pro predikci biomasy a s ní souvisejících veličin a výsledky navzájem porovnány. Nejlepší predikční modely byly aplikovány v prostředí Google Earth Engine k vytvoření dynamických map biomasy pro území České republiky.

**Title:** Predicting forest biomass from satellite remote sensing using Google Earth Engine and advanced machine learning techniques

**Abstract:** Among terrestrial ecosystems, forests are playing an important role in fixing the atmospheric carbon in the form of biomass, thus mitigating climate change. Forest biomass and its growth is traditionally measured using field methods and statistical sampling of representative forest plots. Typically, diameter at breast height (DBH) and tree height (H) are measured for each tree within sampled area, and the biomass related estimates for studied plot are calculated using species-specific allometric equations. Such methods are relatively accurate (depending on the allometric equation used), but laborious (country-level forest inventories requires thousands of forest plots for representative estimates), and representative for given period of sampling only. Due to statistical point sampling, biomass estimates are valid only for larger regional units (e.g. districts, regions, or even states) and can not describe biomass variability on stand level. Remote sensing data provides invaluable information about our ecosystems, both terrestrial and aquatic. The satellite remote sensing is particularly suitable for systematic and repetitive monitoring of ecosystems. Contemporary satellite remote sensing systems for environmental monitoring operated by NASA (Landsat) and ESA (Sentinel) are capable of stand-level observations of multiple spectral channels covering visible and infrared bands, which are potentially sensitive to forest biomass and its change. In addition, active radar system (SAR) operated on Sentinel-1 satellites is an attractive source of remote sensing data for its all-weather capability. Full potential of satellite remote sensing data can be only explored when working with time series of observations. This inevitably possess a challenge due to large volumes of data and computational resources needed for digesting the data. An emerging trend in remote sensing data processing is so-called cloud based processing, where both the remote sensing data archives and algorithms are stored and executed remotely on high-performance computers. When developing a prediction model between any remote sensing observation quantity (surface, reflectance and its transformations, radar backscattering coefficient) and forest biomass related variables, the selection of particular statistical model may influence the prediction performance and conclusion drawn about the usability of remote sensing data studied. Numerous statistical methods were developed in recent decades, including variations of classical linear regression modelling, artificial neural networks, support vector regression, or decision trees, among others. In this study, we apply a cloud-based processing approach of Google Earth Engine to digest a rich time series

of both optical (Landsat-8 and Sentinel-2) and radar (Sentinel-1) data, yielding high quality remote sensing predictors of forest biomass-related variables. Several forest inventories were used, including biomass inventories performed within CzechTerra landscape inventory, ICPForests defoliations and leaf area index. Various statistical methods were used, cross-compared and the best ones were applied in Google Earth Engine to produce a wall-to-wall continuous maps of biomass-related variables for the area of Czech Republic.