Application of information originating from spatially non-exhaustive, ancillary observations in the spatial predictions targeting specific features of the soil mantle

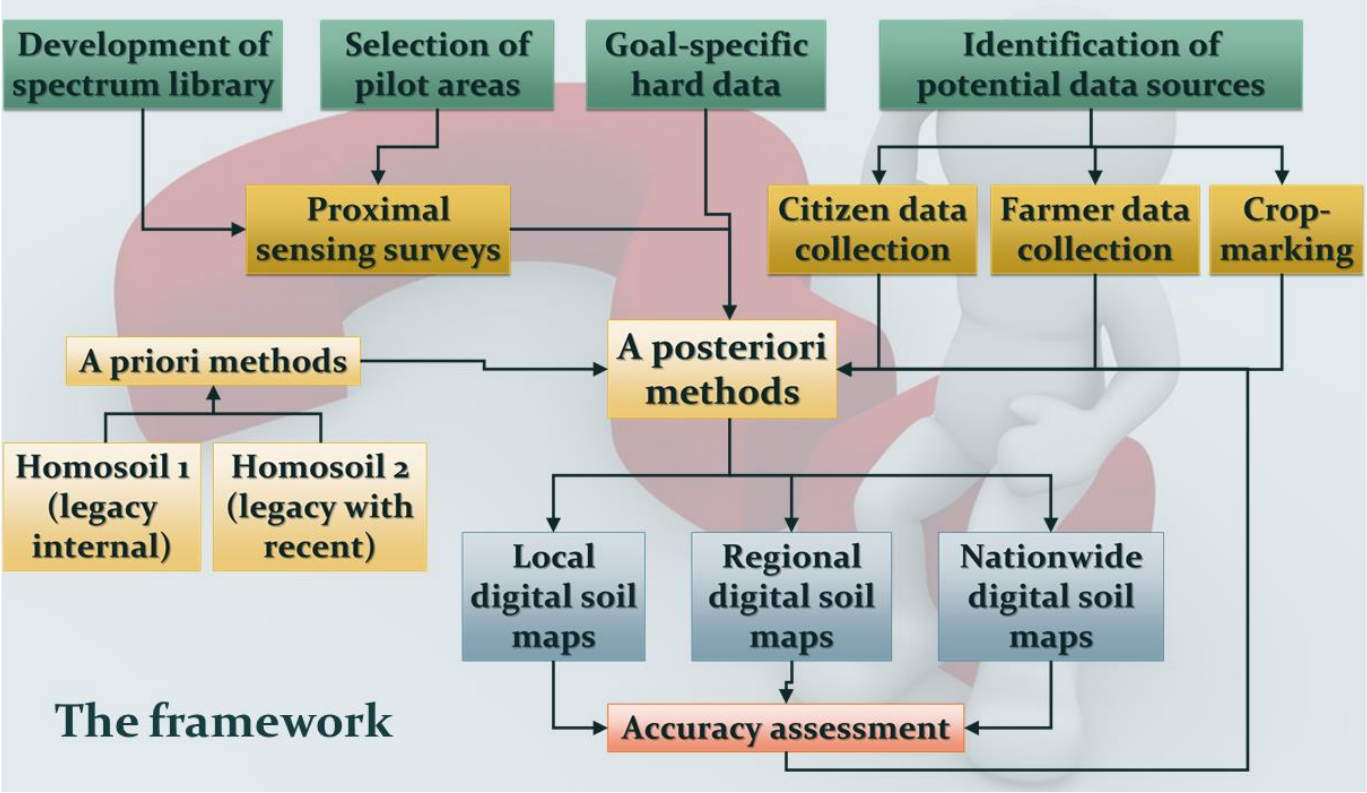
Final report

The activity of the project has heavily built on our former results achieved in the frame of DOSoReMI initiative (NKFIH Grant No K-105167) and its functional extension (NKFIH Grant No KH-126725).

The aim of our work was to investigate,

(i) in which way and with what efficiency spatially non-exhaustive, ancillary information originating from different sources (e.g.: data sequences of various installed field sensors, data collections by proximal sensing techniques, information supply by farmers and land managers as well as citizen science) can be applied in digital soil and environmental mapping, furthermore

(ii) in what manner their application influences (hopefully improves) the results, accuracy and reliability of goal-specific spatial predictions.



During our research we continuously searched for and frequently identified new ways where spatially non-exhaustive, ancillary information may originate from, as well as how it could be utilized efficiently (if at all), and sometimes we concluded in dead-ends. In the present report we go through these directions indicating their success (embodied in publications) or failure by presenting both positive and negative experiences and results. For the sake of clarity we arrange the various case studies along mapping scale.

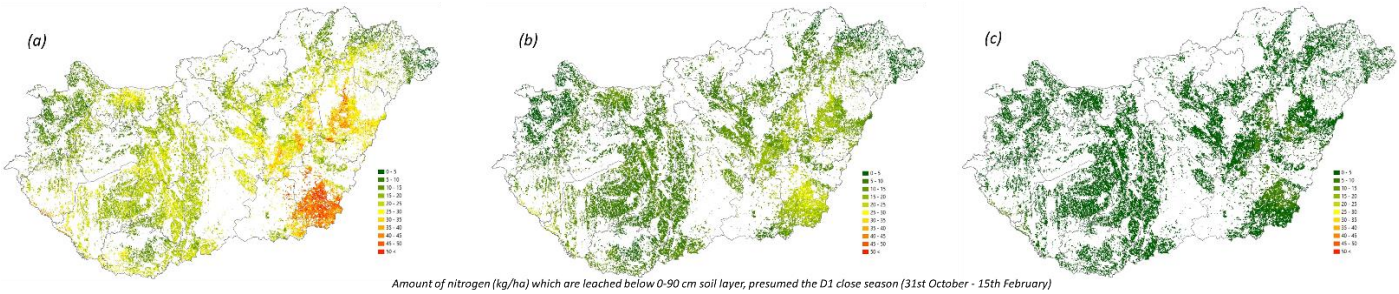
Contents

Mapping soil features at country scale

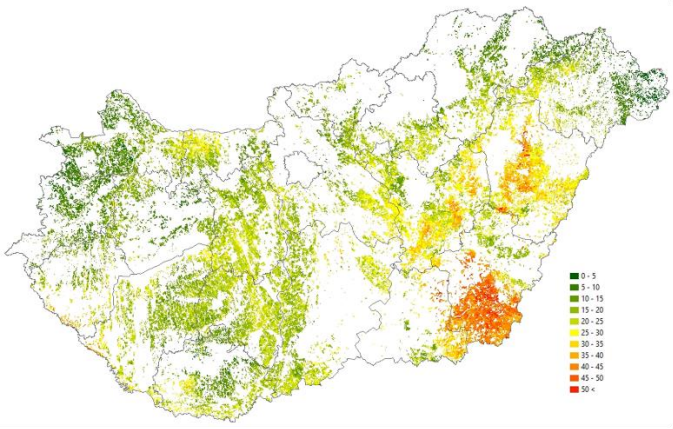
Mapping nitrate leaching hazard assisted by crop simulation based spatial analysis of agricultural soils

Compulsory data acquisition of farmers according to the Nitrates Directive was used to provide input data for the 4M process-based model for estimating yield, crop N uptake, nutrient balances and amount of NO3-N leached under the root zone, all can be considered as various indicators of provisioning and regulating functions of agricultural soils. The Hungarian Nitrate Database uses only one suitable parameter, which can be used for georeferencing its data records, the collected data can be linked to the objects (so called MePAR blocks) of the Agricultural Parcel Identification System. By solving the spatialization of the collected data, the huge amount of information supplied by farmers on load could be linked with spatial environmental ancillary data. The crop simulation based spatial analysis of agricultural soils could be used for the evaluation of the potential pollution effect of nitrate leaching on water quality.

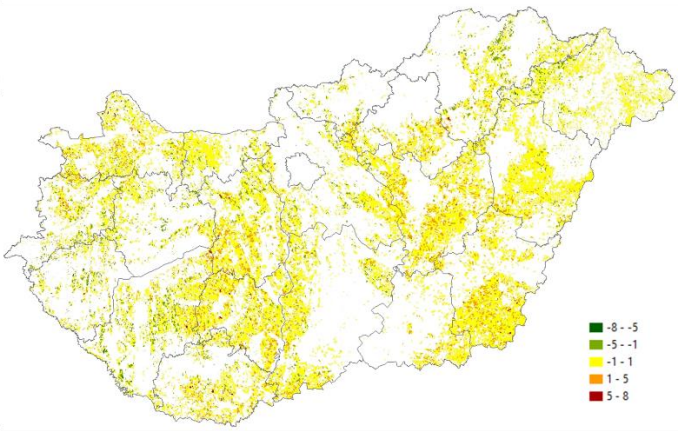
Deliverable: paper published in SUSTAINABILITY 13 : 1 pp. 1-15 (Q1), 2021



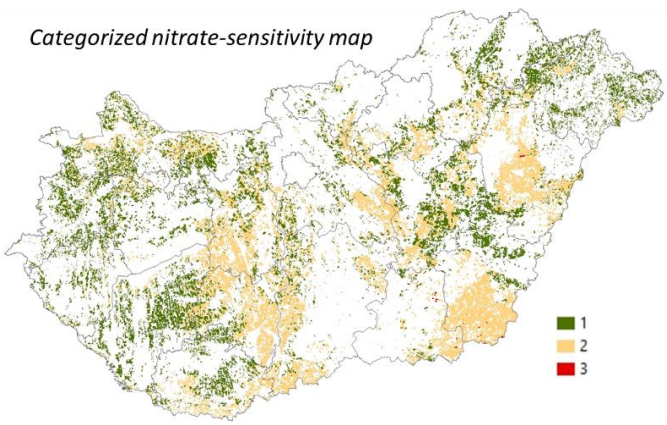
Amount of nitrogen (kg/ha) which are leached below 0-90 cm soil layer, presumed the D1 close season (31st October - 15th February)
(a) actual precipitation for year 2016: wet year, (b) actual precipitation for year 2016 – 200 mm: average year, (c) actual precipitation for year 2016 – 400 mm: dry year

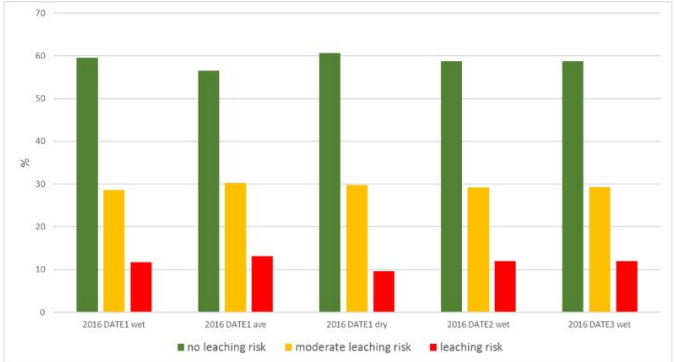


Amount of nitrogen (kg/ha) which are leached below 0-90 cm soil layer, presumed the D3 close season (30th November - 15th February)



Difference of N-balance between the values of D1 and D3 close season (kg/ha), 2016





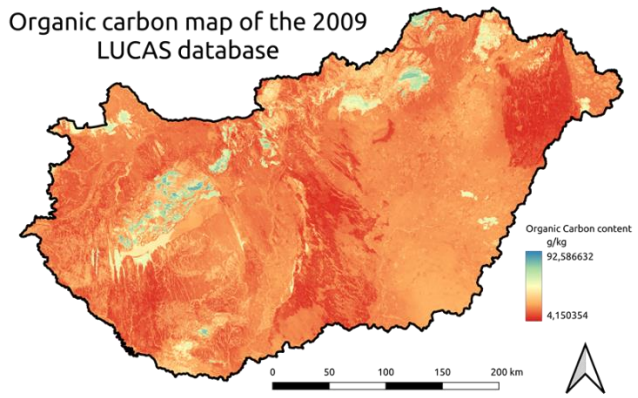
Distribution of three leaching risk categories according to different precipitation and close season scenarios for year 2016

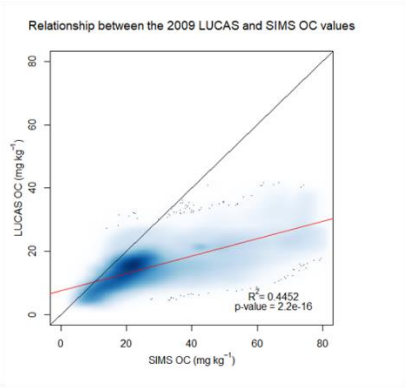
Joint application of the Hungarian Soil Monitoring System and the LUCAS Topsoil dataset as observation data to produce countrywide primary soil property maps

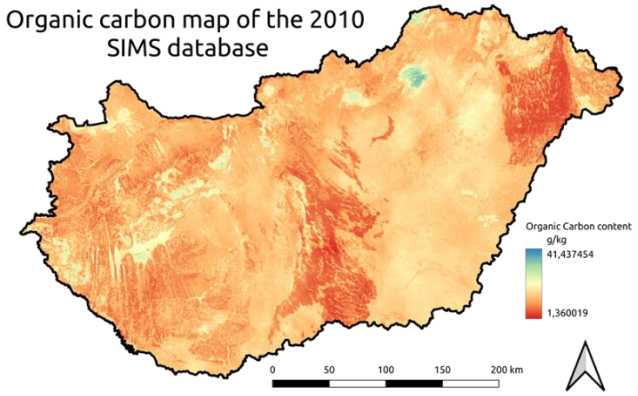
Since soil survey at national scale is very cost- and labour intensive, the compilation of more accurate soil maps relies more and more on combining existing datasets to improve the poor spatial resolution of up-to-date soil surveys. While the continental LUCAS topsoil database is considered up-to-date, it leaves a lot to be desired regarding spatial resolution at national-scale. Using the LUCAS topsoil database to complement the Hungarian Soil Information and Monitoring System (SIMS) is an option to create more detailed soil property maps. The different laboratory analysis methods of the two systems required the data to be converted, so the databases can be directly compared to each other, and later used together for digital soil mapping. To achieve the best results and comparison, two sets of topsoil data were used from 2009 and 2015 for LUCAS and from 2010 and 2016 for the SIMS, respectively. This way, swiftly changing soil properties (pH(H_2O), $CaCO_3$ and SOC) were available for a more accurate side-by-side comparison of the two datasets. We created map products for the soil properties mentioned above from both the LUCAS and SIMS databases with the ancillary data of 28 environmental covariates including parent material, topography, soil type map, and climatic data. Digital soil mapping was performed using random forest kriging with 10-fold cross-validation using the R programming language. The spatial resolution of the maps was 100 x 100 m. The resulting maps were then compared directly to each other using linear regression. Out of the three chemical properties the best correlation was achieved by pH(H_2O) (R^2_{2009} = 0.79 & R^2_{2015} = 0.73), followed by SOC (R^2_{2009} = 0.45 & R^2_{2015} = 0.38) and $CaCO_3$ (R^2_{2009} = 0.38 & R^2_{2015} = 0.36). The difference maps of the two datasets highlight the areas within the country, where further investigation and correction is needed to ensure better quality in future mappings. The results let us to conclude that the LUCAS and national soil databases can and should be harmonized, merged and used together for creating more accurate soil maps at national and continental scale.

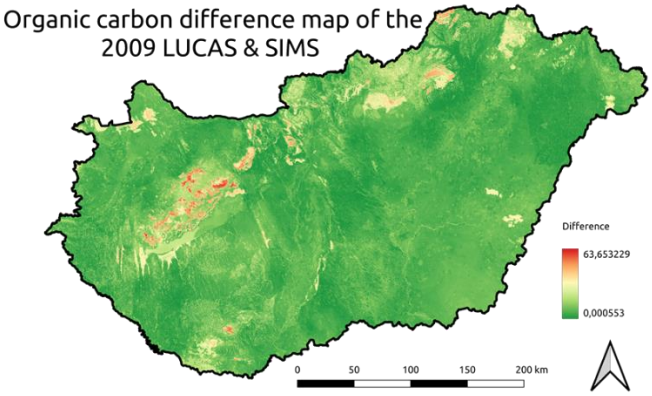
Deliverables:

- submitted manuscript to GEODERMA REGIONAL;
- oral presentations at IUSS DSM/GSM Conference, Orléans, EGU 2024 and 2025 and National GIS Conferences







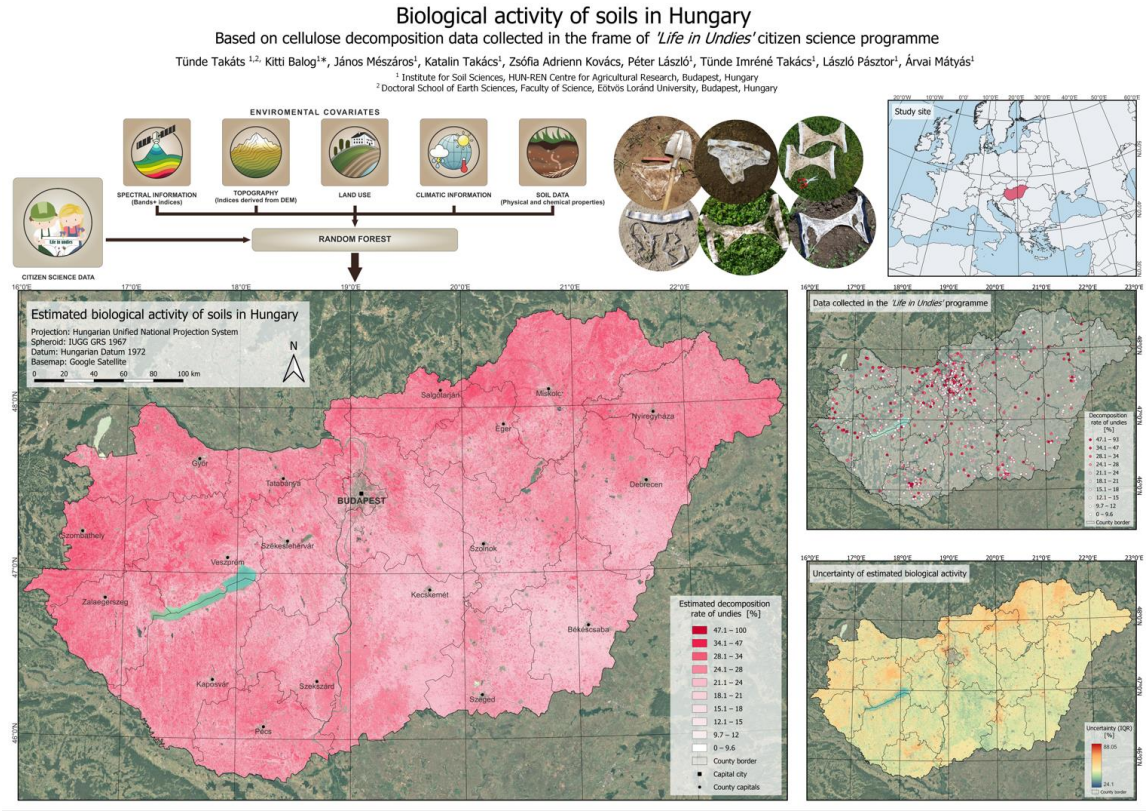


Mapping biological activity of soils in Hungary based on the "Life in Undies" citizen science programme

In 2021 we launched “Life in Undies”, the first Hungarian citizen science program with the double aim: (i) to test the feasibility of a widespread data collection on a selected soil state feature (ii) by obtaining information on the biological activity of soils using a simple estimation procedure. In the Hungarian adaptation of the international “Soil your Undies” programme, nearly 2000 standardised cotton underwear were posted to the participants with a precise step-by-step instruction, to bury them for about 60 days, from May to July in 2021, at a depth of about 20–25 cm. Over the two months in the soil, the underwear began to decompose to varying degrees, influenced by the activity of the soil's biological decomposers. Following the excavation, the participants were asked to send a digital photo of the decomposed underwear and their geographical coordinates with several basic information related to the location and the area (type of cultivation, demographic data etc.). After processing all the incoming images and data, a percentage value of the decomposition joined with spatial position was obtained. While the program proved to be successful in raising environmental awareness, it failed to collect really big, spatially non-exhaustive data to support digital mapping. Thus the countrywide collected biological activity data from nearly 1200 sites were firsly statistically evaluated by spatially aggregating the data both for physiographical and administrative units. The national average decomposition rate was 24.57%, while the maximum decomposition rate reached 93% in a mulched and polyculture utilized garden in sandy soil. Nevertheless, based on the available observation data we made a trial for producing the first national map representing biological activity of the topsoil for the vegetation period in 2021. The aim of the ‘Life in Undies’ national citizen science programme was not only to estimate the microbial activity of soils using a simple method, but also to expand it spatially, allowing to present this information to the community in a mapped format for the first time in Hungary. The nationwide estimation of soil biological activity in Hungary was important not only to test the applicability of citizen science as data collection method in future research but also to identify key background factors that are considered critical variables in mapping biological activity. A method for image processing of the data was developed, and geostatistical method and model were tested to assess how suitable they are for the spatial prediction of biological activity. Not only the result map was generated, but its uncertainty was also provided to aid interpretation of the map.

Deliverables:

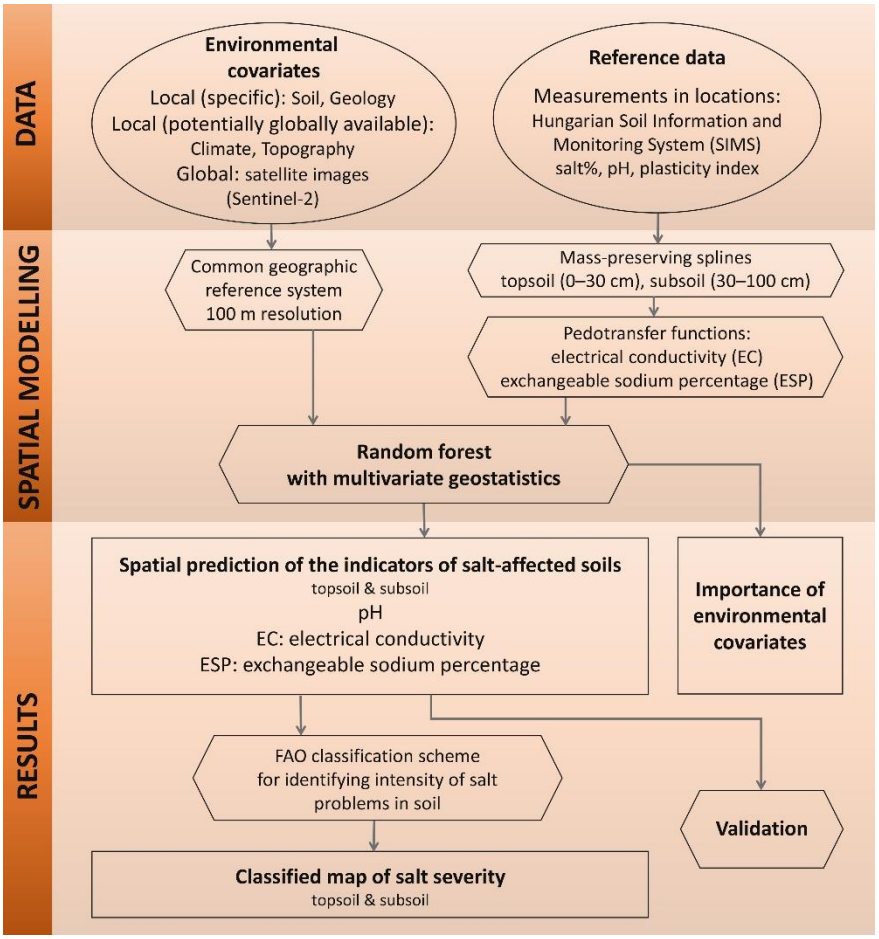
- paper published in AGROKÉMIA ÉS TALAJTAN 72 : 1 pp. 25-43 (Q4), 2023
- manuscript submitted to JOURNAL OF MAPS (under revision)

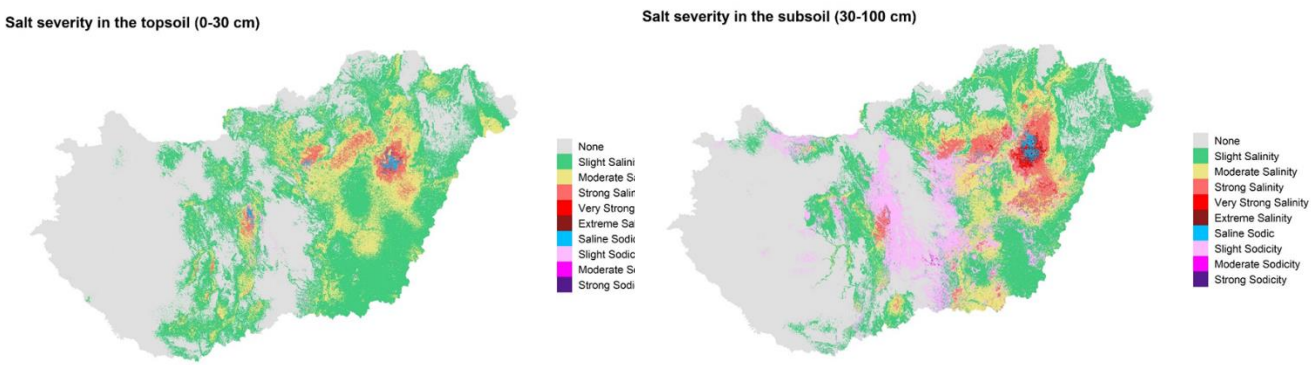


Elaborating the Hungarian segment of the Global Map of Salt-affected Soils

Recently the Global Map of Salt-affected Soils (GSSmap) has been launched that pursued a country-driven approach and aimed to update the global and country-level information on salt-affected soils (SAS). Hungarian contribution to GSSmap was carried out by our research group in the form of preparing national SAS maps using advanced digital soil mapping techniques. A combination of random forest and multivariate geostatistical techniques was used for predicting the spatial distribution of SAS indicators (i.e., pH, electrical conductivity and exchangeable sodium percentage) for the topsoil (0-30 cm) and subsoil (30-100 cm). A large number of indices related to salinization-alkalization were derived from Sentinel-2 satellite images as environmental covariates. The importance plots of random forests showed that in addition to climatic, geomorphometric parameters and legacy soil information, image indices were the most important covariates. The performance of spatial modelling was checked by 10-fold cross validation showing that the accuracy of the SAS maps was acceptable. By this study and by the resulting maps of it, we not just contributed to GSSmap, but also renewed the SAS mapping methodology in Hungary, where we paid special attention to model and quantify the prediction uncertainty that had not been quantified or even taken into consideration earlier. The mapping methodology is suggested to and also planned by us to be used at field scale, where data collection was carried out by proximal sensing techniques (see later).

Deliverable: paper published in REMOTE SENSING 12: (24) 4073 (Q1), 2020



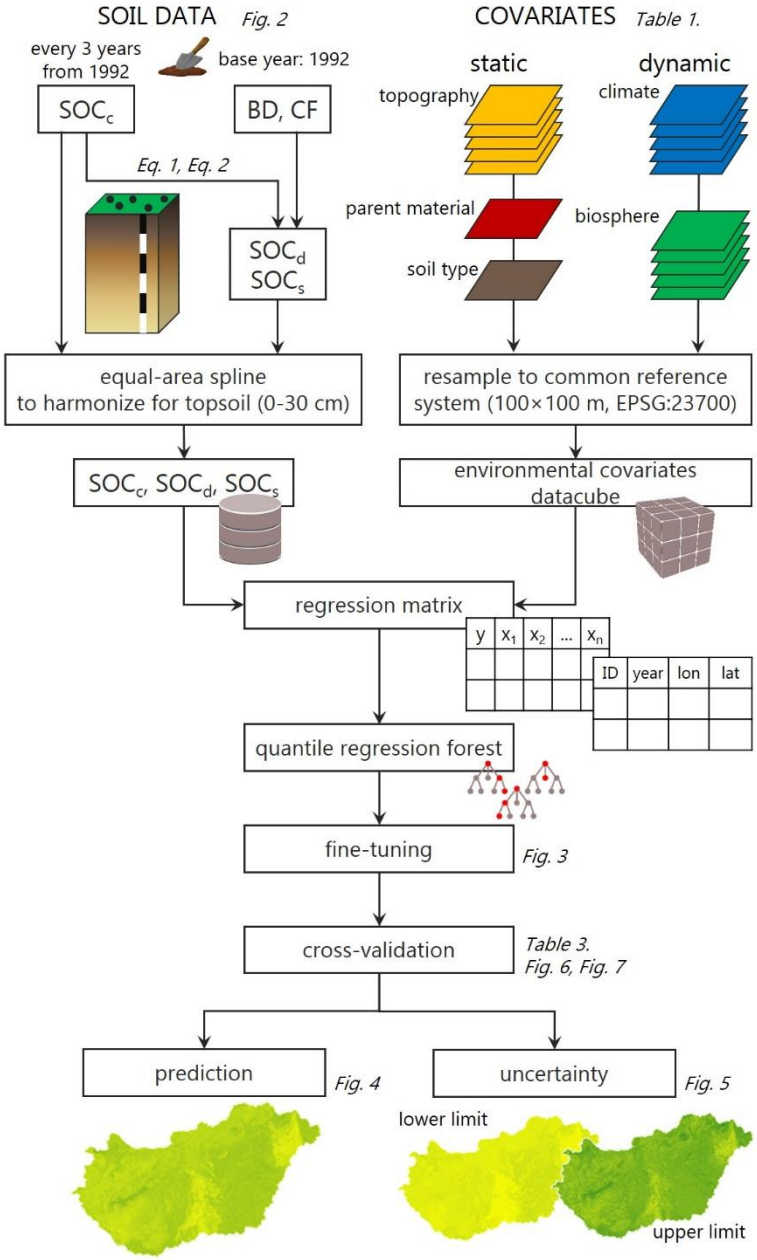


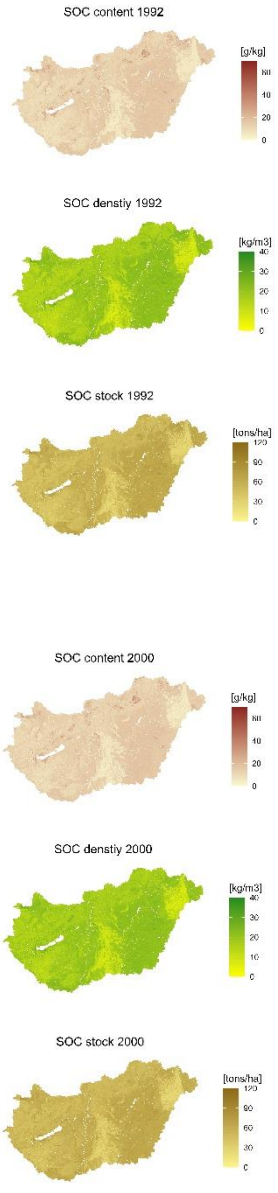
Elaboration of gridded, temporally referenced spatial information on soil organic carbon for Hungary

Soil organic carbon (SOC), known as the most important soil attribute, affects various soil functions and services, essential for nutritious food and clean drinking water. Since recognizing its key role in many environmental challenges, there has been an increasing demand for spatial information on SOC. Our mapping activity aimed at producing spatially exhaustive information on SOC content, density, and stock for the topsoil of Hungary for 1992 and 2000. A time-for-space digital soil mapping approach was pursued to predict and map these SOC properties, with the associated uncertainty, at a resolution of 100×100 m. Particular attention was paid to validating the accuracy of the maps and the reliability of the uncertainty quantifications. The published maps are recommended to be used as baseline maps for Hungary. The spatial resolution makes them suitable for various practical applications (GHG inventory, sustainable agriculture, carbon sequestration). The maps are of interest to researchers, practitioners, and policymakers, helping to achieve scientifically sound results and informed decision-making.

Deliverables:

- paper published in SCIENTIFIC DATA 11, 1312 (2024).
- dataset publicly available stored at Zenodo online repository (https://doi.org/10.5281/zenodo.13236749)

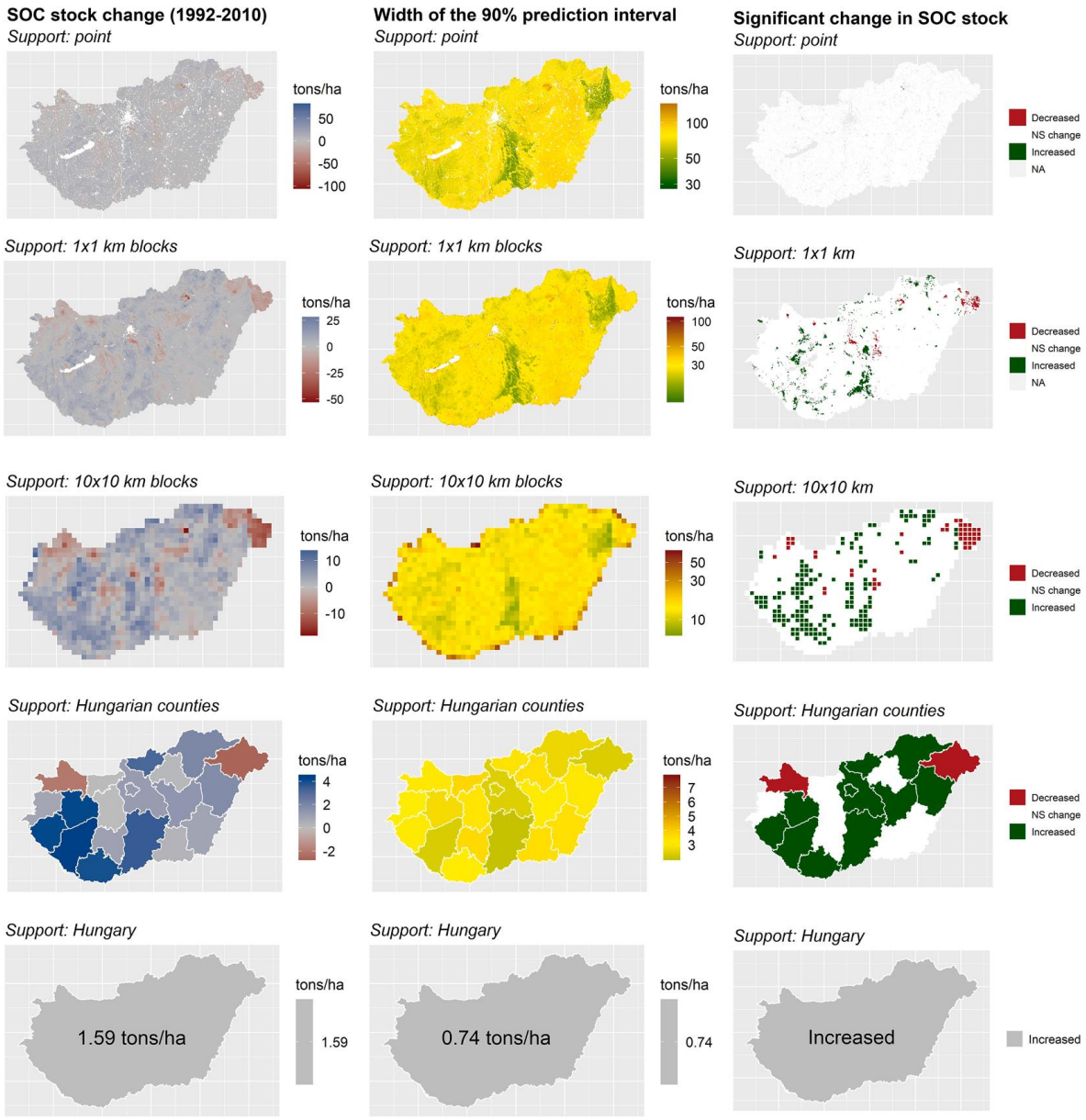




Application of regression cokriging with random forest and spatial stochastic cosimulation to predict soil organic carbon stock change

Many national and international initiatives rely on spatially explicit information on soil organic carbon (SOC) stock change at multiple scales to support policies aiming at land degradation neutrality and climate change mitigation. In this case study we used regression cokriging with random forest and spatial stochastic cosimulation to predict the SOC stock change between two years (i.e. 1992 and 2010) in Hungary at multiple aggregation levels (i.e. point support, 1 × 1 km, 10 × 10 km square blocks, Hungarian counties and entire Hungary). We also quantified the uncertainty associated with these predictions in order to identify and delimit areas with statistically significant SOC stock change. Our study highlighted that prediction of spatial totals and averages with quantified uncertainty requires a geostatistical approach and cannot be solved by machine learning alone, because it does not account for spatial correlation in prediction errors. The total topsoil SOC stock for Hungary was predicted to increase between 1992 and 2010 with 14.9 Tg, with lower and upper limits of a 90% prediction interval equal to 11.2 Tg and 18.2 Tg, respectively. Results also showed that both the predictions and uncertainties of the average SOC stock change were smaller for larger spatial supports, while spatial aggregation also made it easier to obtain statistically significant SOC stock changes. The latter is important for carbon accounting studies that need to prove in Measurement, Reporting and Verification protocols that observed SOC stock changes are not only practically but also statistically significant.

Deliverable: paper published in GEODERMA 403 Paper: 115356 , 12 p. (D1), 2021

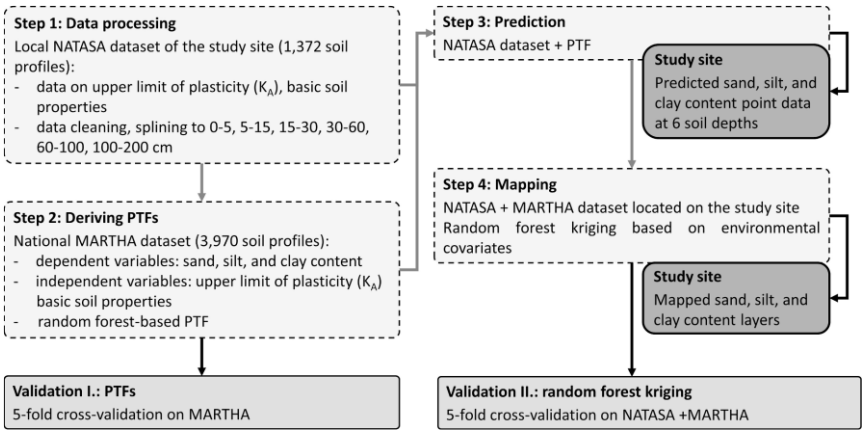


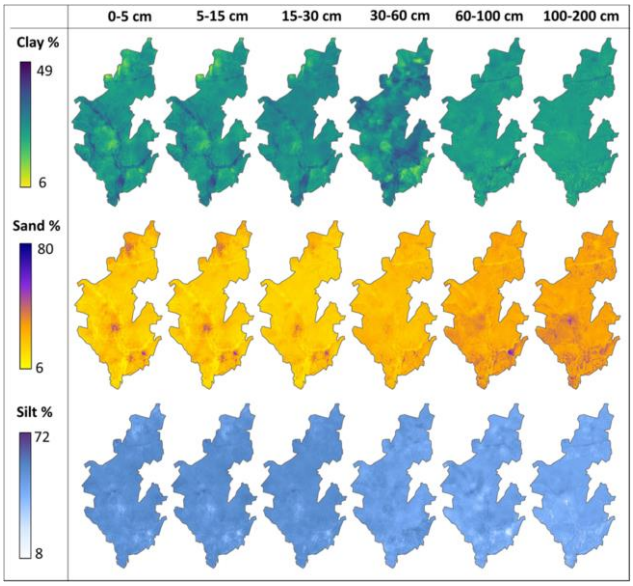
Mapping soil features at regional scale

Large-scale mapping of soil particle size distribution using legacy data and machine learning-based pedotransfer functions

Spatially detailed quantitative data on particle size distribution is not only crucial for assessing soil degradation, hydrology and fertility, but also a basic information to model hydraulic properties, and it is highly demanded for a range of modeling applications. Mapping of sand, silt, and clay content was targeted based on the big data provided by NATASA (Hungarian acronym for Profile-level Database of Hungarian Large-Scale Soil Mapping) initiative. Digital processing of the soil observation records of the still available soil observation legacy data originating from large-scale surveys carried out in Hungary between the 60s and 90s was firstly finalized for the watershed of the Lake Balaton in order to support hydrological modelling studies on the catchment. The digitized soil observations are firstly used in digital mapping of primary soil properties at a scale of 25 meters. Since NATASA includes information on soil taxonomy and basic soil chemical and physical properties, but no direct information on sand, silt and clay content, only an indirect parameter, namely, the upper limit of soil plasticity, pedotransfer functions (PTFs) were developed to compute particle size distribution from soil properties available in the NATASA dataset. The PTFs were trained and tested on the Hungarian Detailed Soil Hydrophysical Database using random forest method. For the prediction model, i) additive log-ratio transformed clay, silt and sand content were used as the dependent variables, and ii) the upper limit of soil plasticity, soil type, calcium carbonate content, organic matter content and pH were included as independent variables. The results indicate that the R^2 values of the PTFs are 0.69 for clay, 0.58 for silt, and 0.74 for sand content. Since the NATASA database contains soil information from different depths, we splined the data into six standard depth layers. The spatial modelling was performed by random forest kriging (RFK) using environmental auxiliary variables. The R2 values of the RFK models range from 0.19 to 0.67 for clay content, from 0.49 to 0.62 for silt content and from 0.69 to 0.74 for sand content.

Deliverable: paper published in GEODERMA, Volume 454, 2025, 117178, ISSN 0016-7061. (D1), 2025

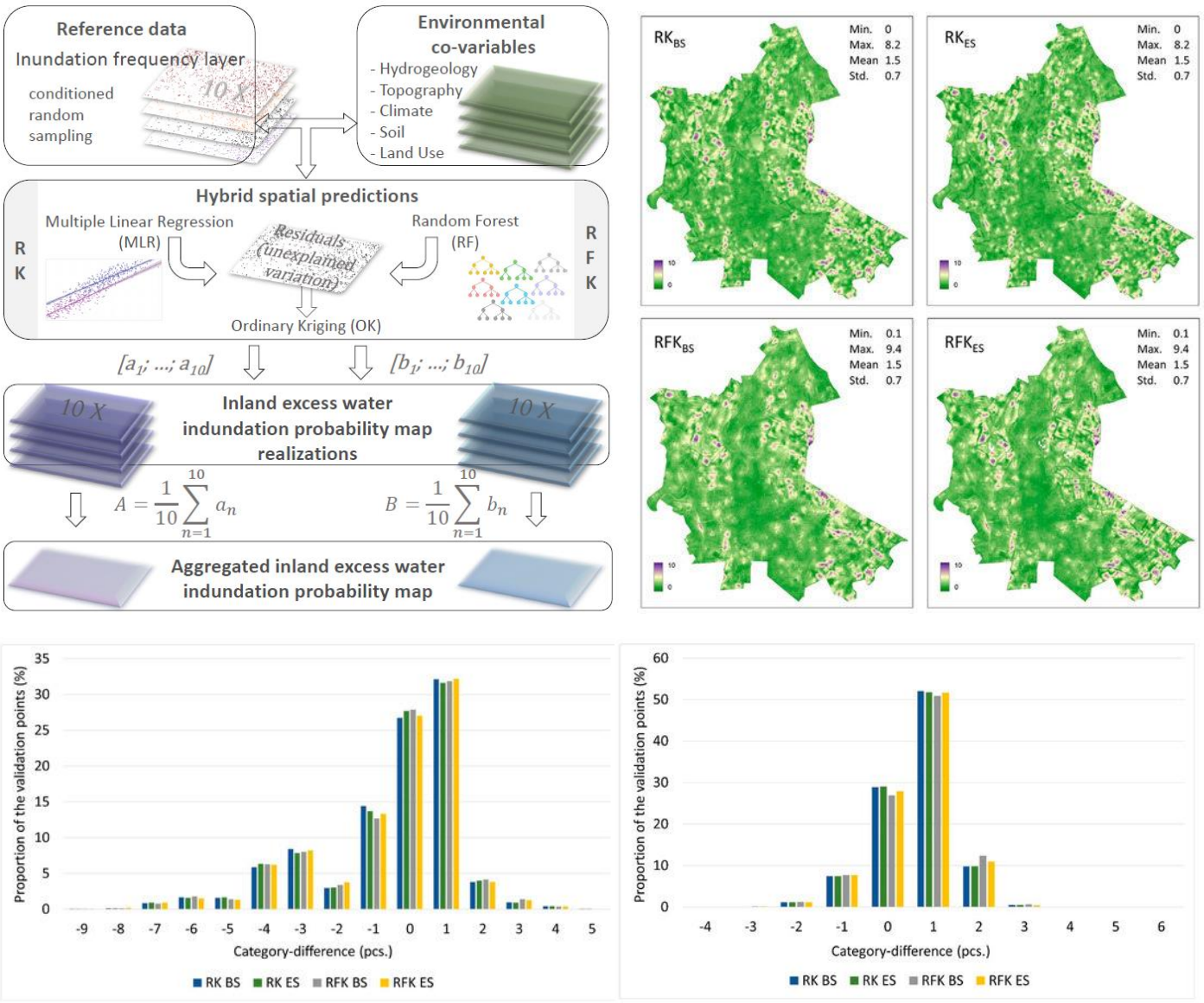




Application of hybrid prediction methods in spatial assessment of inland excess water hazard

Inland excess water is temporary water inundation that occurs in flat-lands due to both precipitation and groundwater emerging on the surface as substantial sources. Inland excess water is an interrelated natural and human induced land degradation phenomenon. Identification of areas with high risk requires spatial modelling, that is mapping of the specific natural hazard. Various external environmental factors determine the behaviour of the occurrence, frequency of inland excess water. Spatial auxiliary information representing inland excess water forming environmental factors were taken into account to support the spatial inference of the locally experienced inland excess water frequency observations. The applied hydrological and hydrometeorological predictors were represented by spatially non-exhaustive co-variables. Two hybrid spatial prediction approaches were tested to construct reliable maps, namely Regression Kriging (RK) and Random Forest with Ordinary Kriging (RFK) using spatially exhaustive auxiliary data on soil, geology, topography, land use, and climate. Comparing the results of the two approaches, we did not find significant differences in their accuracy. Although both methods are appropriate for predicting inland excess water hazard, we suggest the usage of RFK, since (i) it is more suitable for revealing non-linear and more complex relations than RK; (ii) it requires less presupposition on and preprocessing of the applied data (iii) keeps the range of the reference data, while RF RK tends more heavily to smooth the estimations and (iv) it provides a variable rank, providing explicit information on the importance of the used predictors.

Deliverable: paper published in ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION 2020: (9) p. 4. (Q1), 2020

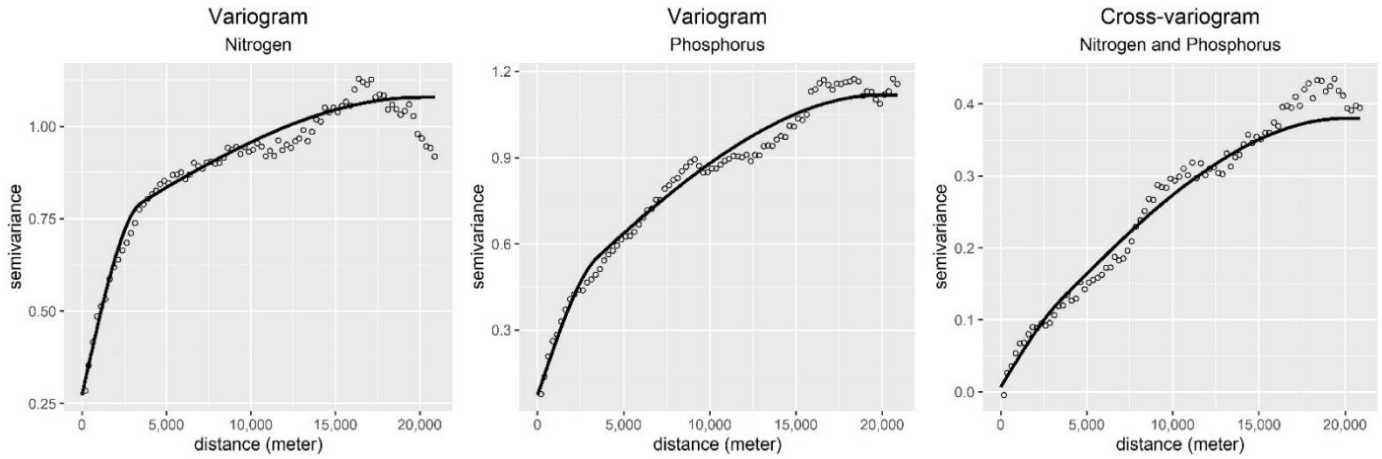


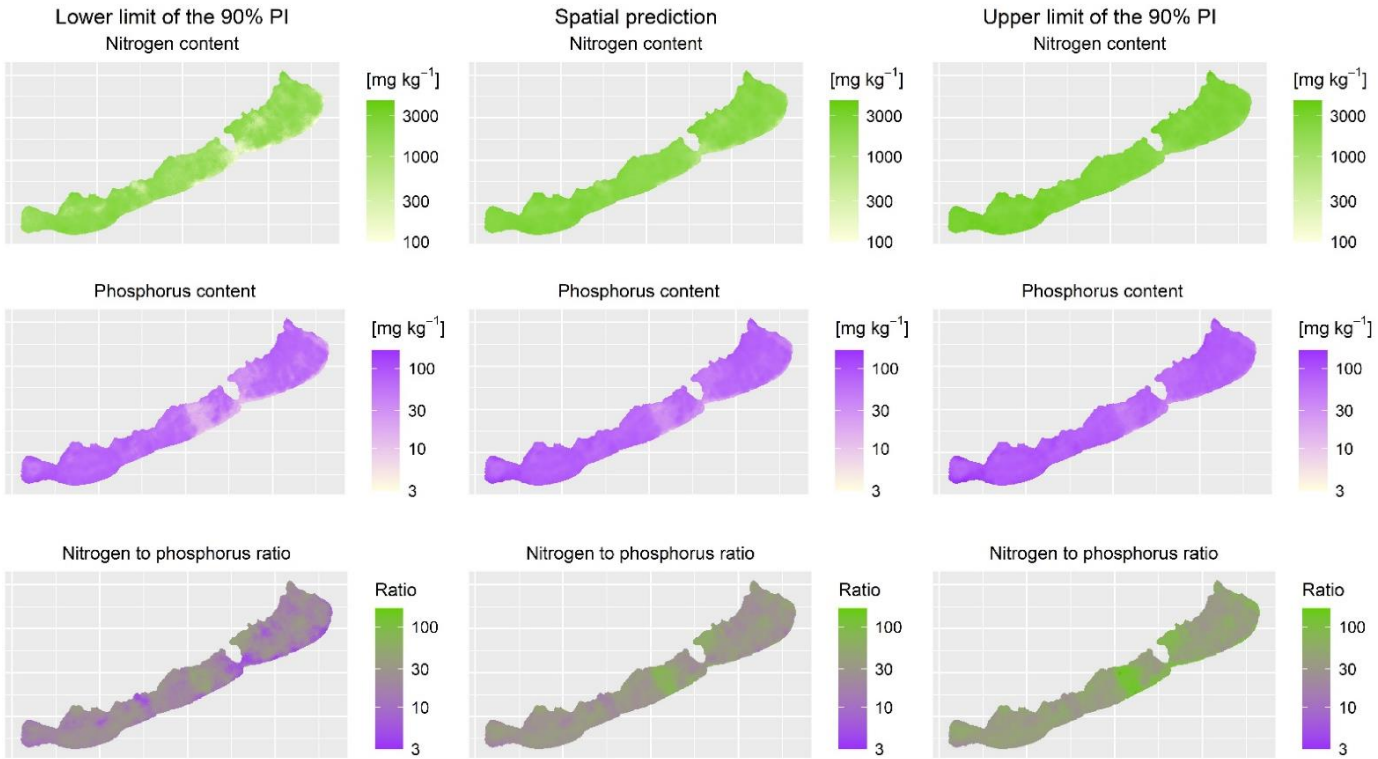
a) Category-difference between the independent validation dataset, and the categorized predicted values. b) Difference between the legacy map's inundation frequency and the categorized inundation result maps.

Application of multivariate geostatistics for spatial modelling of more than one variable

Based on a digital database formerly elaborated by digitizing the data of a full survey of the Lake Balaton's sediments, the applicability of multivariate geostatistics was tested for spatial modelling of more than one variables. We jointly modelled the spatial distribution of the nutrients nitrogen (N) and phosphorus (P), and their ratio (i.e. N:P) in the sediments of the lake and then provided spatial predictions at different supports (i.e. point, basin and entire lake) with the associated uncertainty. Variography confirmed that there is a spatial interdependency between the nutrients. The results revealed that the application of multivariate geostatistics allows this interdependency to be taken into account and exploited to provide coherent and accurate spatial predictions. Moreover, stochastic realizations, reproducing the joint spatial variability, can be generated, which allow to provide spatially aggregated predictions with the associated uncertainty at various size of supports. Our study highlighted that it is worthy of applying multivariate geostatistics for the spatial modelling of two or more variables, which jointly vary in space.

Deliverable: paper published in WATER 14 : 3 Paper: 361 (Q1), 2022

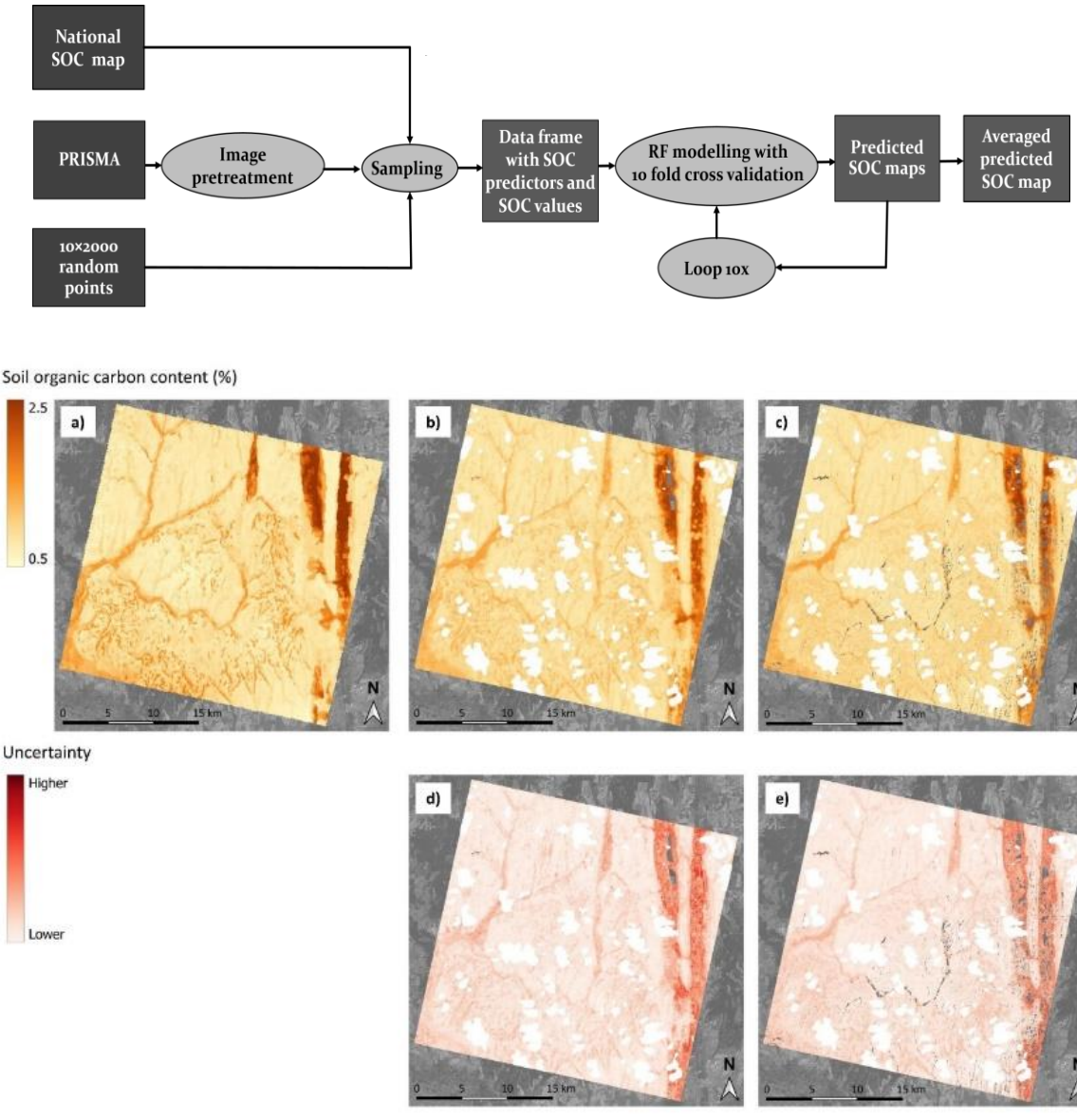




Machine learning based downscaling of a national scale soil organic map using hyperspectral satellite imagery to support carbon accounting at plot scale

The precise estimation of soil organic carbon (SOC) content is significant in carbon se-questration, contributing to the broader efforts in addressing climate change and alleviating soil degradation. Reflectance spectroscopy has emerged as a promising method for estimating SOC content. Using hyperspectral imagery from air- and spaceborne platforms presents a significant opportunity for the larger-scale mapping of soil physico-chemical parameters. Our study targeted the recently compiled national SOC map to be downscaled by digital soil mapping using mainly PRISMA hyperspectral satellite imagery as environmental covariate to two spatial resolution (30 and 5 m). Estimating SOC levels within the study area was accomplished by applying machine-learning based models. We evaluated the prediction accuracy of trained models augmented with a variety of environmental data in three scenarios at two spatial resolutions: (i) PRISMA spectral bands, (ii) PRISMA spectral bands and spectral indices, (iii) PRISMA spectral bands together with spectral indices and topomorphometric covariates. The best prediction (CCC_{5m} = 0.71 and CCC_{30m} = 0.76) was achieved with scenarios including all three types of predictors for both spatial resolutions. The results indicate that pure spectral features, in case of bare soil and hyperspectral imagery, are capable relatively high reliable predictions.

Deliverable: manuscript under preparation for submission to Earth Systems and Environment



(a) The 100 meter resolution soil organic carbon map used as reference; (b) The best (Spectral bands+indices+DEM) SOC map at 100 m resolution; (c) The best (Spectral bands+indices+DEM) downscaled SOC map at 5 m resolution; (d) Uncertainty map for the best performing (Spectral bands+indices+DEM) SOC map at 30 m resolution; (e) Uncertainty map for the best (Spectral bands+indices+DEM) downscaled soil organic map at 5 m resolution (white areas are masked due to cloud cover).

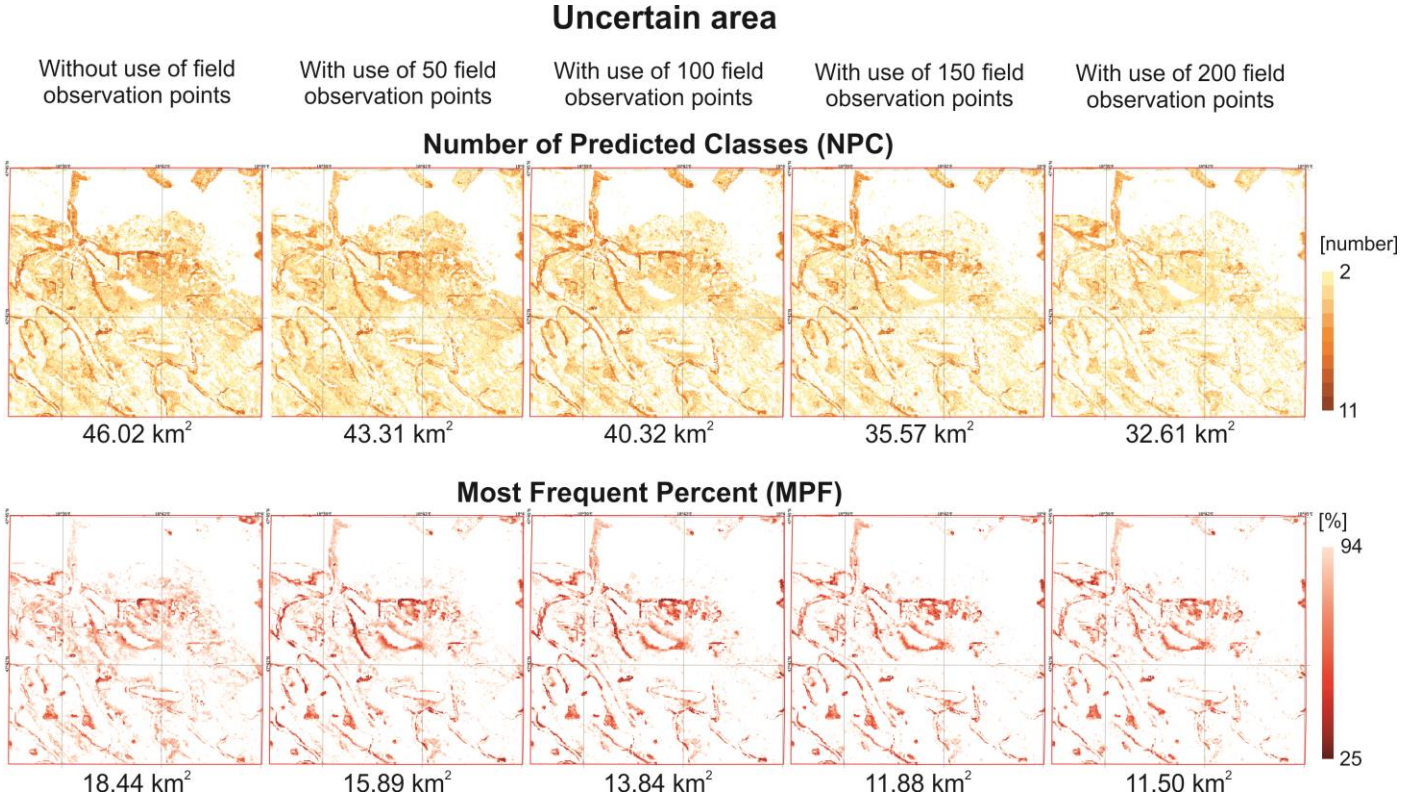
Integrating legacy geological map, Earth Observation imagery and auxiliary spatial data with successive ground truth observation and machine learning for digital mapping of parent material (*proved to be dead-end*)

In this study we tested the feasibility of DSM-based disaggregation of a geological map to map soil parent material in a pilot area, the Dorog Basin, Hungary. The available sources made the collection of ground truth data in limited extent, which fact stimulated us to test the effect of increasing number of field observation data on the performance on the disaggregated map. The legacy geological map was correlated with the FAO code system characterizing soil parent material and then disaggregated by sequentially sampling the map to get virtual observation data for training and testing the classification of the lithological composition. Various machine learning methods (Random Forest (RF), Gradient Boosting Machine (GBM), Deep Learning (DL), Generalized Linear Model (GLM), Naive Bayes Classifier (NB)) were applied. The models were built on a large set of environmental covariates consisting of satellite imagery data used both in form of native spectral bands and derived spectral indices; various derivatives of SRTM digital elevation model (DEM) together with digital primary soil property maps. To examine the effects of increasing number of ground truth data, 200 sites were visited to collect visual field observation data. The testing was built up according to a scenario, in which the number of field observation points were increased iteratively with 50 in subsequent steps by adding them to the randomly generated point set. Validation of classified maps was carried out by using different measures to evaluate the usage of increasing number of field observation data in the predictions:

- overall accuracy of the predicted maps,
- the number of predicted predicted classes of each pixel and
- the percentage of the most frequently predicted class.

The various evaluation methods give different results emphasising differing perspective of the predictive maps. While the overall mean accuracy of the models systematically decreases as the number of field observation points increases, definite reduction in the size of the uncertainly estimated areas was experienced. Our suggestion is, that overall mean accuracy cannot be considered as the primary way to determine the reliability of the predictive models and the digital maps inferred by them, it should be supplemented with further measures.

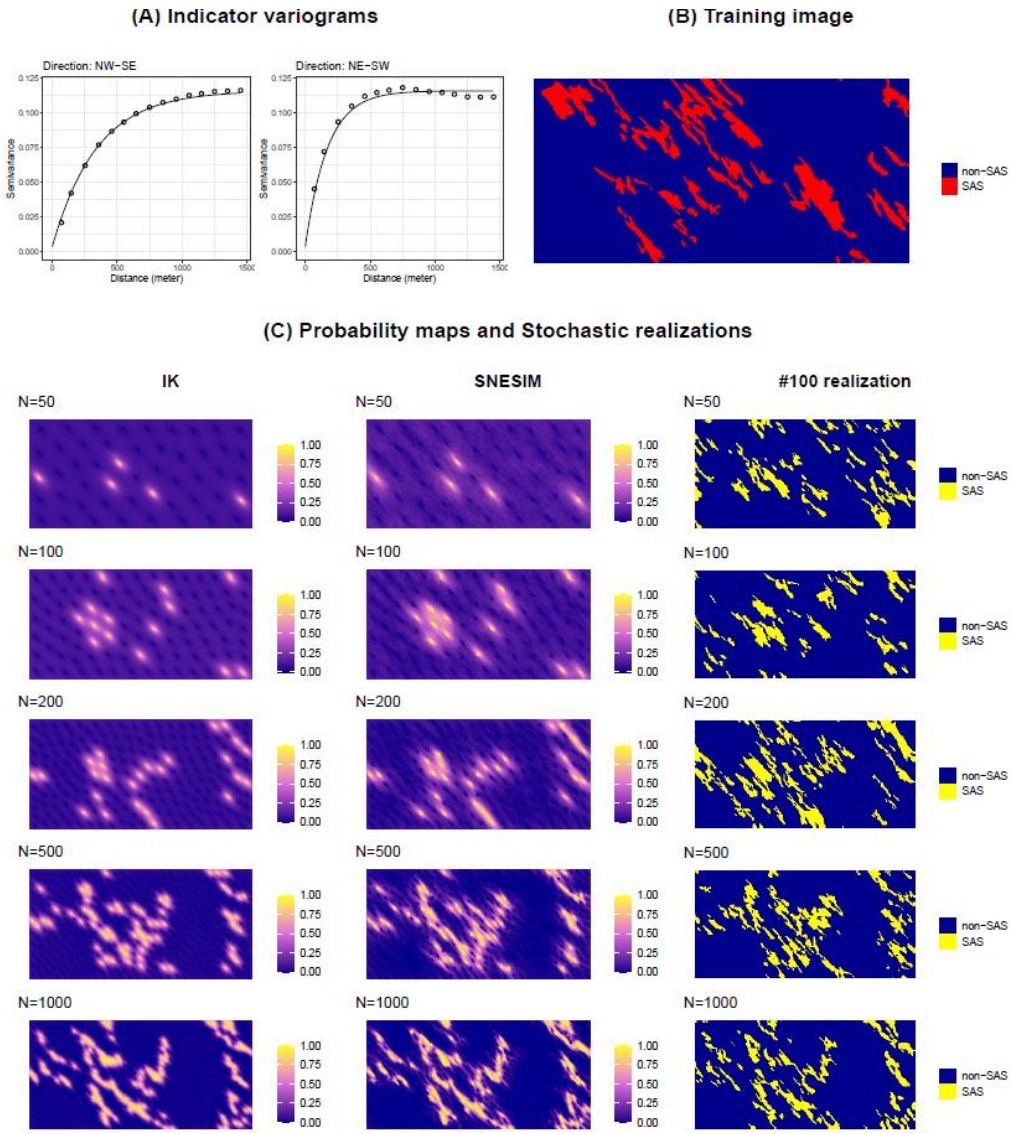
Deliverable: In the deficiency of real results the work was not published.



Testing multiple-point geostatistics and indicator approach for the mapping of highly structured spatial patterned soils (*proved to be dead-end*)

An important property of salt-affected soils (SAS) is their highly structured spatial pattern, which could pose a real challenge when SAS mapping is targeted using geostatistics and environmental covariates are not available. This can be attributed to the fact that the multivariate normal distribution is in the center of most geostatistical techniques, which maximizes spatial disorder, i.e. dissolves the highly structured SAS pattern. Therefore, our objective was to propose two geostatistical approaches, namely the indicator approach and multiple-point geostatistics (MPS), which can handle the highly structured SAS pattern in theory. Indicator kriging (IK) and single normal equation simulation (SNESIM) were tested in a pilot area (190.4 km^2), Hungary, where spatially exhaustive information on SAS pattern was available due to the Digital Kreybig Soil Information System. This exhaustive information was used as ground truth and virtually sampled by spatial coverage sampling with various sample sizes. Our study illustrated that both geostatistical approaches can capture the highly structured SAS pattern. Although SNESIM provided promising stochastic realizations at first glance, the validation technique developed to evaluate the probability maps' accuracy pointed out that IK outperformed SNESIM at each sample size. Additionally, SNESIM consistently underestimated the probability of the presence of SAS, especially in areas where SAS was more likely to be found. This is a serious shortcoming of the algorithm. However, we think that MPS should not be abandoned, rather more research and studying further MPS algorithms are needed, as the merits of MPS over variogram-based techniques can be still useful in the assessment of the highly structured SAS pattern.

Deliverable: In the deficiency of real results the work was not published.

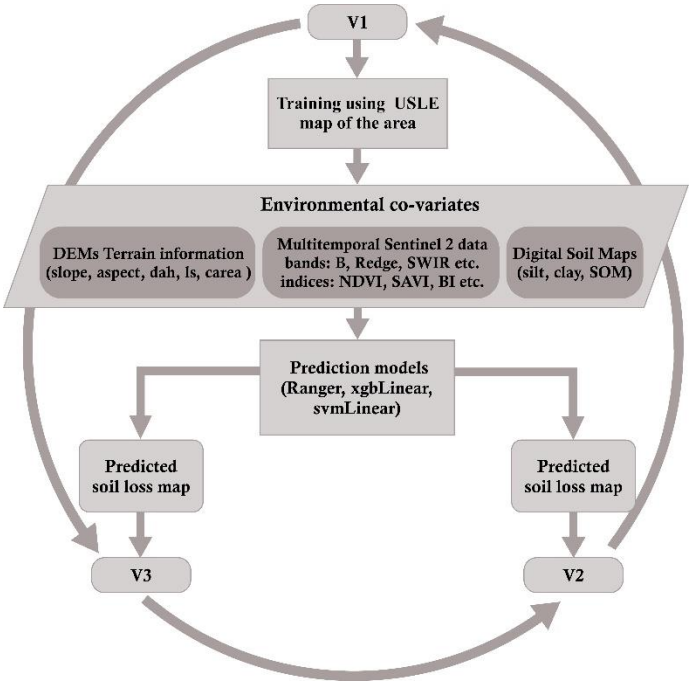


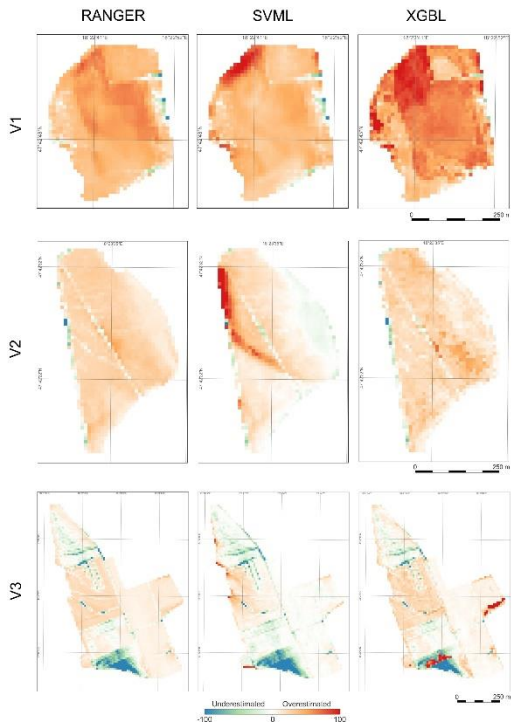
Mapping soil features at plot scale

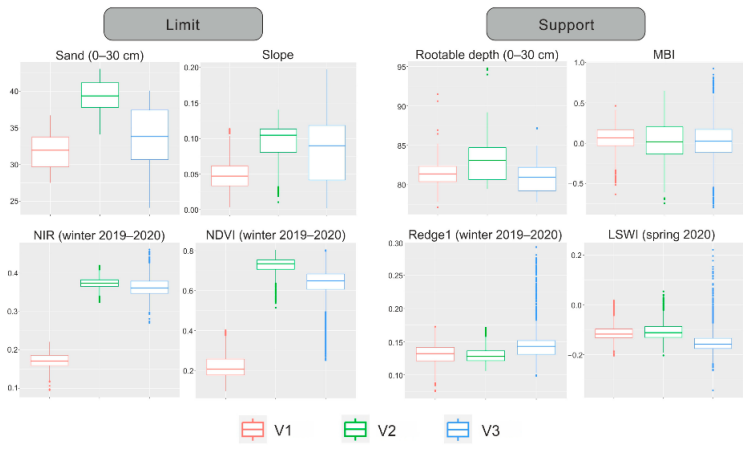
Testing the applicability and transferability of data-driven geospatial models for predicting soil erosion in vineyards

Empirically based approaches, like the Universal Soil Loss Equation (USLE), are appropriate for estimating mass movement attributed to rill erosion. USLE and its associates become widespread even in spatially extended studies in spite of its original plot-level concept, as well as with certain constraints on the supply of suitable input spatial data. At the same time, there is a continuously expanding opportunity and offer for the application of remote sensing (RS) imagery together with machine learning (ML) techniques to model and monitor various environmental processes utilizing their versatile benefits. We tested the applicability of data-driven geospatial models for predicting soil erosion in three vineyards, considering the seasonal variation in influencing factors. Soil loss was formerly modeled by USLE, thus providing non-observation-based reference datasets for the calibration of parcel-specific prediction models using various ML methods (Random Forest, eXtreme Gradient Boosting, Regularized Support Vector Machine with Linear Kernel), which is a well-established approach in digital soil mapping. Predictions used spatially exhaustive, auxiliary, and environmental covariables. RS data were represented by multi-temporal Sentinel-2 satellite imagery data, which were supplemented by (i) topographic covariates derived from a UAV-based digital surface model and (ii) digital primary soil property maps. In addition to spatially quantifying soil erosion, the feasibility of transferring the inferred models between nearby vineyards was tested with ambiguous outcomes. Our results indicate that ML models can feasibly replace the empirical USLE model for erosion prediction. However, further research is needed to assess model transferability even to nearby parcels.

Deliverable: paper published in LAND, 14, 163. https://doi.org/10.3390/land14010163 (Q1), 2025



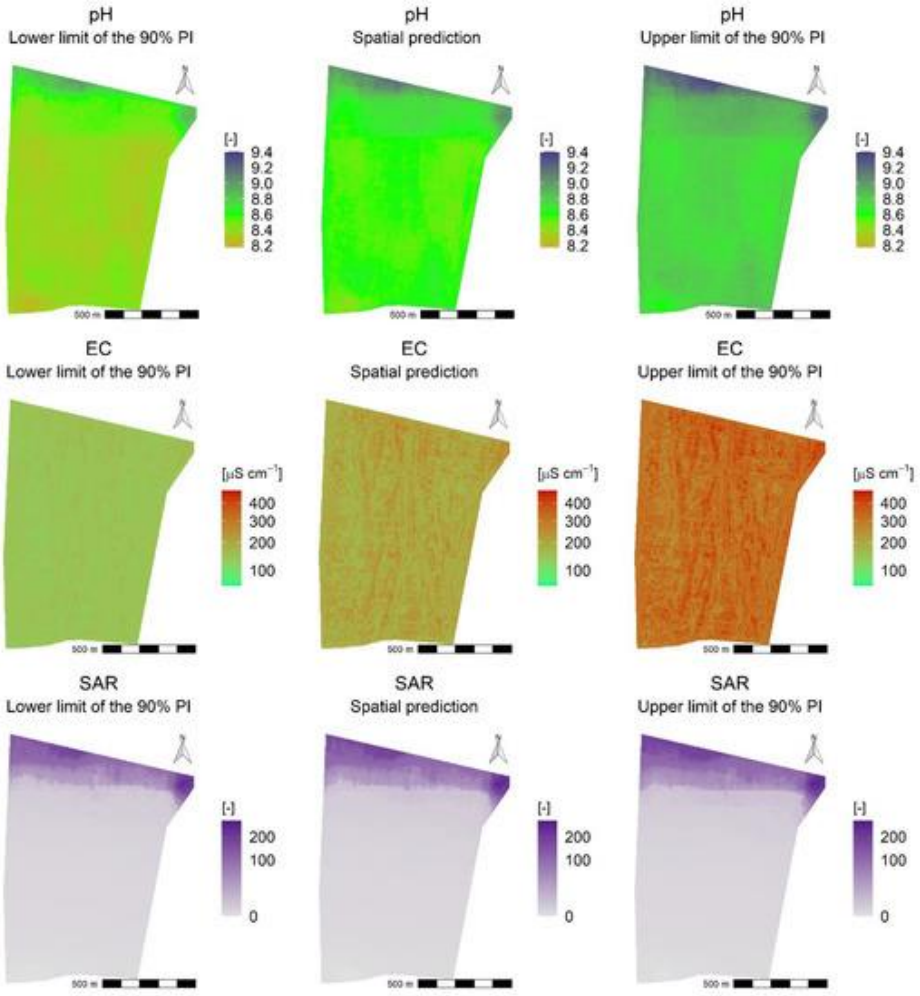




High-resolution mapping and assessment of salt-affectedness by the combination of ensemble learning and multivariate geostatistics

Salt-affectedness on an arable land (0.85 km^2) was mapped with high spatial resolution, using a combination of ensemble machine learning and multivariate geostatistics on three salt-affected soil indicators (i.e., alkalinity, electrical conductivity, and sodium adsorption ratio) based on 85 soil samples. Ensemble modelling with five base learners (i.e., random forest, extreme gradient boosting, support vector machine, neural network, and generalized linear model) was carried out and the results showed that ensemble modelling outperformed the base learners for alkalinity and sodium adsorption ratio with R^2 values of 0.43 and 0.96, respectively, while only the random forest prediction was acceptable for electrical conductivity. Multivariate geostatistics was conducted on the stochastic residuals derived from machine learning modelling, as we could reasonably assume that there is spatial interdependence between the selected salt-affected soil indicators. We used 10-fold cross-validation to check the performance of the spatial predictions and uncertainty quantifications, which provided acceptable results for each selected salt-affected soil indicator (for pH value, electrical conductivity, and sodium adsorption ratio, the root mean square error values were 0.11, 0.86, and 0.22, respectively). Our results showed that the methodology applied in this study is efficient in mapping and assessing salt-affectedness on arable lands with high spatial resolution. A probability map for sodium adsorption ratio represents sodic soils exceeding a threshold value of 13, where they are more likely to have soil structure deterioration and water infiltration problems. This map can help the land user to select the appropriate agrotechnical operation for improving soil quality and yield.

Deliverable: paper published in AGRONOMY 12 : 8 Paper: 1858 (Q1), 2022



Pedometric management zone delineation based on various proximal sensing datasets

The importance of optimal management zone delineation is growing, yet existing methods often overlook digital soil maps. We address this gap by proposing a method that leverages these maps to highlight field-scale soil properties. In our research, an agricultural plot was surveyed with various proximal sensing methods: with three different geophysical tools, as well as a drone based hyperspectral sensor generating spectral information and digital elevation model (DEM). Evenly distributed field samples were also collected across the plot to analyse six soil parameters (K_2O, P_2O_5, Nitrate and Soil Organic Material (SOM) contents, pH and Plasticity index). The 126 spectral bands, DEM together with its 10 topo-morphometric derivatives and 18 rasterized, interpolated geophysical measurements (Ground Penetrating Radar, electrical conductivity using a GF Instruments CMD-Explorer and CMD-Explorer Mini, resistivity data measured, Gamma-ray data collected using a GF Instruments Gamma Surveyor Vario 2048-channel geophysical gamma-ray spectrometer) were allocated into three co-variate datasets, which were then used independently and in all possible combinations to create soil maps. A Random Forest algorithm, optimized with the Caret package, was used to evaluate the data. 70% of the observations were used for model building, and the remaining 30% for testing. Performance of the predictions were measured by Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Numerical Standard Error (NSE) and Lin's Concordance Correlation Coefficient (CCC). Best performing primary soil property maps were clustered using K-means method targeting the identification of their optimum classification, thus deriving pedometric zones. The number of zones varied between four and six for the seven versions. The generated zonal maps were assessed independently for similarity with ANOVA-test comparing them with the yield of crop present on the plot. Finally, all models of ANOVA-testing were compared using the Akaike information criterion, showing the dataset built up on geophysical and spectral information as the best and explaining the highest variance. Our results showed the usability of geophysical tools for mapping in the deeper soil levels and of spectral information for the top-soil layer.

Deliverables:

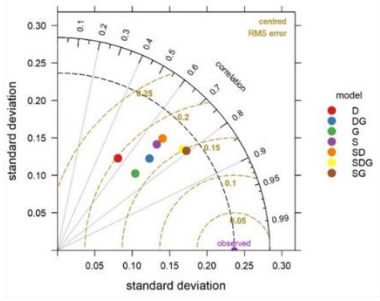
- poster presentation at The Sixth Global Proximal Soil Sensing Workshop
- manuscript in preparation to be submitted to SMART AGRICULTURAL TECHNOLOGY

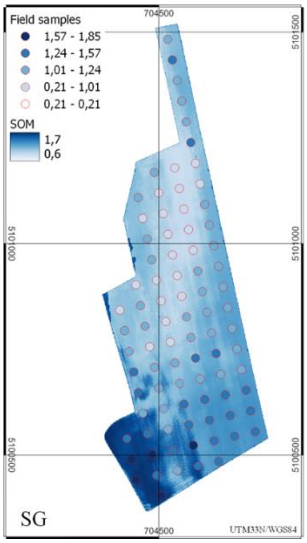
Results

Good results (R > 0.75)

SOM [m/m%]

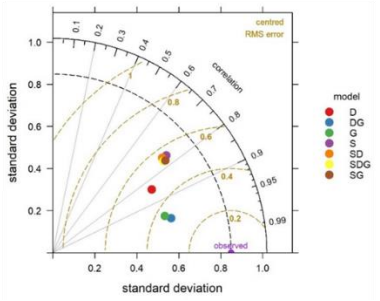
methodname	R	RMSE	CCC
S	0,684	0,173	0,667
SD	0,686	0,174	0,678
SDG	0,780	0,150	0,772
SG	***0,792***	***0,147***	***0,784***
D	0,550	0,196	0,489
DG	0,710	0,169	0,662
G	0,712	0,170	0,620

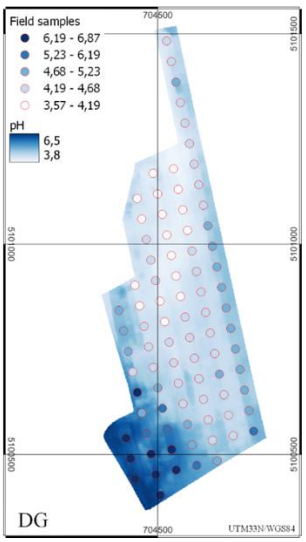




pH (KCl)

methodname	R	RMSE	CCC
S	0,759	0,554	0,741
SD	0,756	0,564	0,727
SDG	0,770	0,544	0,743
SG	0,774	0,542	0,748
D	0,843	0,474	0,773
DG	***0,960***	***0,322***	***0,898***
G	0,951	0,357	0,872





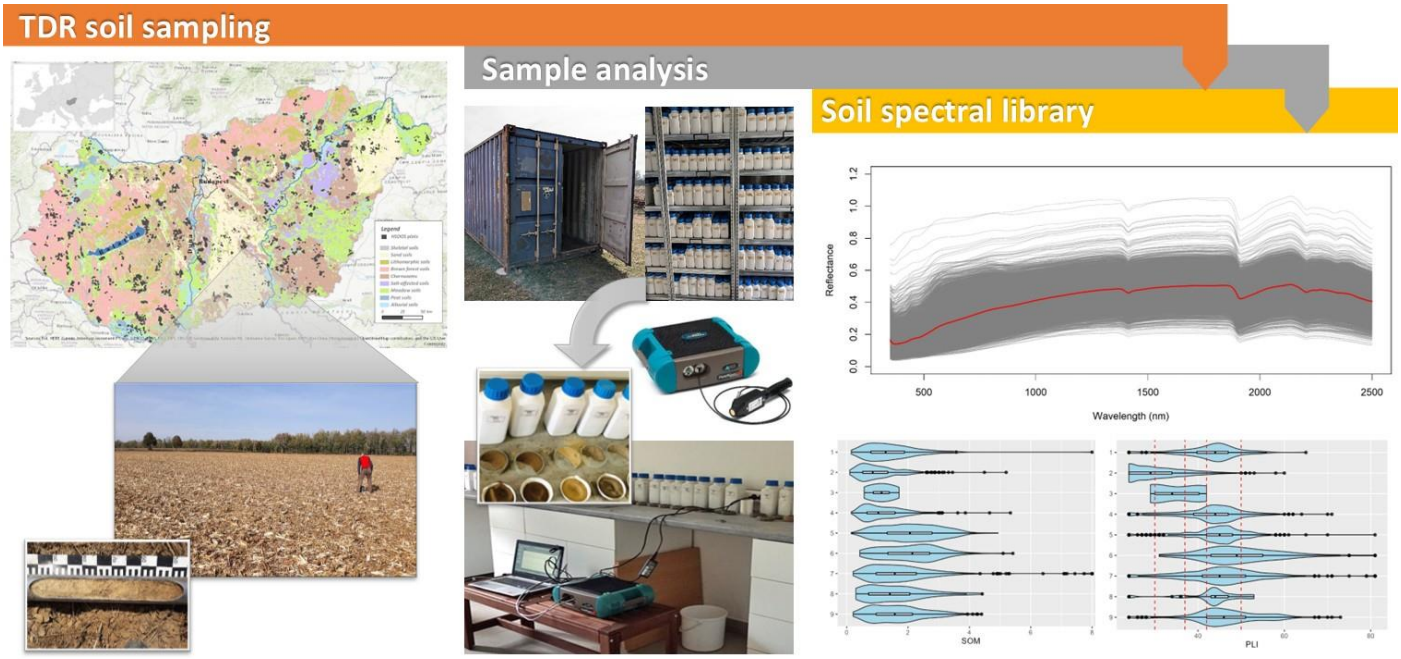
Specific, spatially non-exhaustive, ancillary observation data sets to support digital soil mapping

Setting up the Vis-NIR soil spectral library of the Hungarian Soil Degradation Observation System

Since soil spectroscopy is considered to be a fast, simple, accurate and non-destructive analytical method, its application can be integrated with wet analysis as an alternative. Therefore, development of national-level soil spectral libraries containing information about all soil types represented in a country is continuously increasing to serve as a basis for calibrated predictive models capable of assessing physical and chemical parameters of soils at multiple spatial scales. The laboratory and visible-near infrared spectral data of legacy soil samples from the Hungarian Soil Degradation Observation System (HSDOS) were merge to establish a data set, which includes the following parameters measured in 5,490 soil samples: pH_{KCl}, soil organic matter (SOM), calcium carbonate (CaCO3), total salt content (TSC), total nitrogen (N_{total}), soluble phosphorus (P_2O_5-AL), soluble potassium (K_2O-AL), plasticity index according to Hungarian standard (PLI), soil profile depth and reflectance data between 350 and 2,500 nm wavelength. The most promising benefit of the united data sets is the significant extension of prediction possibilities to estimate important primary soil properties, which than can be efficiently applied in digital soil mapping activities.

Deliverables:

- paper published in SCIENTIFIC DATA 12, 363 (D1), 2025;
- dataset publicly available stored at Zenodo online repository (https://doi.org/10.5281/zenodo.14610222)
- manuscript in preparation on the detailed presentation of the elaborated predictions
- oral presentation at The Sixth Global Proximal Soil Sensing Workshop



Cropmarks in aerial archaeology as special spatial indicators of soil features

Cropmarks are a major factor in the effectiveness of traditional aerial archaeology. The positive and negative features shown up by cropmarks are the role of the different cultivated plants and the importance of precipitation and other elements of the physical environment. In co-operation with the experts of the Eötvös Loránd University a new research was initiated to compare the pedological features of cropmark plots (CMP) and non-cropmark plots (nCMP) in order to identify demonstrable differences between them. For this purpose, the spatial soil information on primary soil properties provided by DOSoReMI.hu was employed. To compensate for the inherent vagueness of spatial predictions, together with the fact that the definition of CMPs and nCMPs is somewhat indefinite, the comparisons were carried out using data-driven, statistical approaches. Three pilot areas were investigated, where Chernozem and Meadow type soils proved to be correlated with the formation of cropmarks. Kolmogorov-Smirnov tests and Random Forest models showed a different relative predominance of pedological variables in each study area. The geomorphological differences between the study areas explain these variations satisfactorily. The identified relationships between cropmarking and soil features were planned to be utilized in the spatial inference of soil properties, where cropmarking sites would have represented a unique, spatially non-exhaustive auxiliary information. The work has been suspended, but hopefully will be restarted in the (near) future.

Deliverable: paper published in REMOTE SENSING 13(6), 1126 (Q1), 2021