Multidiszciplináris tudományok, 11. kötet. (2021) 1 sz. pp. 22-37 **https://doi.org/10.35925/j.multi.2021.1.3**

ESTIMATION OF SURFACE SOIL MOISTURE BY INTEGRATING ENVIRONMENTAL DATA AND REMOTE-SENSING SATELLITES

Daniel Kibirige
PhD student, Institute of Geography and Geoinformatics, University of Miskolc Hungary
Address: 3515 Miskolc, Miskolc-Egyetemváros, e-mail: ecodan@uni-miskolc.hu

Endre Dobos
Associate professor, Institute of Geography and Geoinformatics, University of Miskolc Hungary
Address: 3515 Miskolc, Miskolc-Egyetemváros, e-mail: ecodobos@uni-miskolc.hu

Abstract

Soil moisture (SM) or soil water content is a critical variable in the climate system and a key parameter in earth surface processes. This study aimed to assess citizen observatory (CO) data's suitability to develop a method to estimate surface SM distribution using Sentinel-1B and Landsat 8 data; acquired between January 2019 and June 2019. Three approaches were developed and compared using multiple linear regression (MLR), regression-kriging (RK) and cokriging (CK). MLR provided more realistic spatial patterns over the landscape, even in a data-poor environment. RK was found to be a potential tool to refine the results, while CO was found to be less effective. The obtained results showed that CO data harmonised with Sentinel-1B SAR, Landsat 8, and terrain data could estimate and map soil moisture content.

Keywords: *soil moisture; digital soil mapping; Sentinel-1; synthetic aperture radar (SAR).*

1. Introduction

Soil moisture (SM) or soil water content is a critical variable in the climate system and a key parameter in earth's surface processes, especially in water and energy cycles. It influences the overall amount of water that follows on the earth's surface and below the surface [1]. SM is described as the amount of water content (vol/vol) within a given region of the soil profile at a given time [2]. While a significant number of local and regional soil moisture networks exist worldwide, there are no general requirements, e.g. observed variables, sensor types, sensor configuration, and the data produced are not readily available in the appropriate scale [3]. The International Soil Moisture Network (ISMN) is an international initiative that attempts to solve data gaps of soil moisture [4, 5]. In situ, soil moisture measurements are collected from various globally harmonised networks in terms of sample periods, units and data types, and made available to the public free of charge via the web [6]. It is also clear that the COs (e.g. the GROW Observatory) have a great capacity to add to the unprecedented flow of data from thousands of sensors. In the 21st century, citizen science research is increasingly evolving, whereby data collection is carried out by or through non-experts [7]. Despite the long history of the 'citizen observatory' (CO), there is still a lack of quality assurance for the scientific use of data produced by citizen science [8]. Biodiversity topics [9], and nature and climate topics [10] have overshadowed citizen science programs. The GROW Observatory tracks soil resources to engage a target audience of smallholder farmers and community groups practising sustainable cultivation [3].

The project's goal was to demonstrate a complete 'citizens' observatory' framework for measuring soil moisture, land use, and land cover. It also contributes in situ satellite validation data; producing valuable data products and applications, and finally supplies local farmers with a local soil context [3, 11]. In particular, precision farming, scientific studies to test CO data analysis and SM mapping on a wide scale are required to help agriculture development. This study aimed to support this process by developing and testing several digital soil mapping (DSM) procedures.

DSM [12, 13] has helped researchers quantitatively map soil properties using numerous data sources and several different spatial scales [14, 15]. Compared to more conventional mapping approaches, DSM needs a smaller soil sampling effort [14, 16]. Once the lower quantity of soil data sampled has been harmonised with environmental covariates from various sources, DSM can be a useful method for estimating soil moisture and different soil properties [17]. The SM measurements and tracking sites that are available are limited and not particularly useful for local users. The citizen observatory is a possible way to address this data void. A relatively large geo-referenced upper 10 cm SM dataset was obtained by GROW Observatory, to generate observed data for Sentinel 1 applications [3].

A significant proportion of SM forecasts based on satellite imagery are focused on global and national scales to date [18,19]. Freely available satellite radar data, such as Sentinel-1, have tremendous potential for SM estimation [20], but authentic products need field calibration data; one of the most restricting variables. Systems capable of calculating diverse soil properties, such as SM, are rare and are limited to a few measurement sites and spatial interpolation techniques are mostly absent [21, 22]. The number of scientific articles addressing SM estimation using radar satellite data for irrigation and farming applications is growing, considering the difficulties [23, 24].

Zribi et al. [20], reviewed emerging research of SM estimation. To extract and test SM maps for an arid region, Gharechelou et al. [25] ran multiple interpolation methods (inverse distance weighting, kriging, and co-kriging). The maps created were compared and validated using ground-truth data. For spatial interpolation of many soil properties, Keskin and Grunwald [26] used Regression Kriging (RK). Han et al. [27] used data from MODIS to estimate SM. Zeng et al. [28] used cokriging to measure soil moisture on bare soil using temporal backscatter ratios using three variables dependent on cross-semivariogram and cross-variance over 91 sampling points. The results revealed a root mean square error (RMSE) value of between 2,669 and 2,701 (December to April) for five months. Two analytical models were tested by Chatterjee et al. [29]: a multiple linear regression model against the cubist model using SM measurements in situ soil sensor, Sentinel-1 data, and separate ancillary data (USGS DEM, Sentinel 1, US National Ground Cover Database, and Soil Survey Spatial Probabilistic Remapping (POLARIS) data set). Coefficient of determination (R^2) was the mathematical metric used for evaluating the model.

In soils with a particular crop class, the Cubist model (0.56) worked better than the multiple linear regression (0.24). The MLR findings are close to the results obtained in our analysis and indicate that the dynamic interaction between soil moisture and satellite imagery can often not be accounted for by MLR. Three kriging techniques (extended, ordinary, and cokriging) were used by Zhang et al. [30]; interpolation algorithm to interpolate near-surface soil moisture data measured by wireless sensor networks using remote sensing-derived spectral variables, NDVI and albedo. The visual findings showed that more spatial information was revealed by the product of extended kriging than ordinary kriging or cokriging. However, the Expanded Kriging (2.41) root mean square error (RMSE) was observed to be the lowest of the four interpolation effects (RMSE ranged from 2.62 to 3.01).

For field-scale SM estimation in a semi-arid area, Bousbih et al. [23] used combined Sentinel 1 and 2 datasets using an integrated decision tree estimation technique. Two models, a classification of the neural network and a direct water cloud model, were compared. They concluded the method is promising and that the values of the RMSE ranged from 4.8 to 6.4, and R^2 values ranged from 0.4 to 0.58. Xing et al. [24] derived SM using Radarsat-2 data applied an updated water cloud model to eliminate vegetation's influence, applied the Dubois model for dielectric constant retrieval and finally measured the volumetric soil's moisture. Xing et al. [24] final results yielded an RMSE value of 4.43, and an R^2 of 0.71.

Therefore, this research aimed to develop a spatially and temporally estimation system, analysing and assessing soil moisture by combining terrain and remote sensing data, such as NDVI products derived from Sentinel-1 and Landsat 8. The analysis was carefully conducted through the following steps: (1) defining an in situ soil moisture sampling design for low-cost soil sensors from the GROW example; (2) selecting environmental covariates with the highest statistical correlation with soil moisture; (3) developing and simulating a simple quantitative approach to soil moisture estimation; and (4) creating and simulating a quantitative approach to soil moisture estimation. Finally, the study explored whether low-cost sensors are sufficient instruments for estimating soil moisture data across various geomorphological units to support agricultural and scientific research activities.

2. Study Area

The study area consisted of 1700 hectares (ha) of agricultural land. It is situated on the southern tip of the mountain range of the North-Western Carpathians and the southern slopes of the Bükk Mountains (Figure 1). The site is located near the village of Tard (47°52′33.67″ N, 20°35′56.53″ E). Geomorphologically, the region consists of two south-facing plateaus, which are separated by the Tardi stream. The average elevation is between 150 and 200 m above the mean sea level, and the relatively dry climate characterises the area.

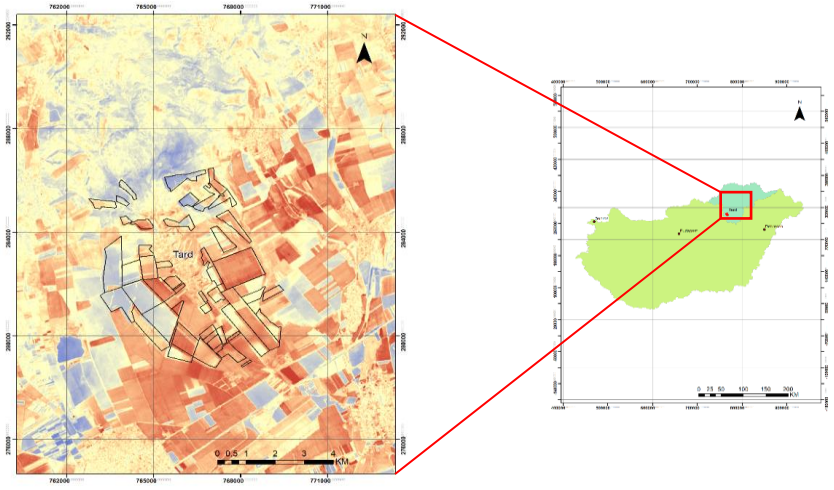

Figure 1. *Study area Tard village located in the western county of Borsod-Abaúj-Zemplén in Hungary*

The soils of the region are predominantly mollic Vertisols [32] (Udic Haplustert [33]), which cover 80 per cent of the region, including high plateaus. The bottom of the valley has a loam texture and the

Gleyic Chernozems [32] (Cumulic Endoaquols [33]) while the steep side of the valley has a loamy sand texture. The northern edge becomes a steep, sandy loam of Luvisol [32] (Alfisol [33]) with a lot of rhyolite tuff near the surface; this is where land use is transitioning from arable land to forest to mountain range.

3. Materials and Methods

3.1. Materials

3.1.1. Field Data

Measurements of soil moisture were conducted using Flower Power sensor (by Parrot S.A., Paris, France) developed low-cost soil moisture sensor) at a depth of 0–10 cm below the soil's surface. Low-cost sensors include SM data in the upper 10 cm range, soil temperature, light intensity and conductivity. The explanation for the GROW project collection of the upper 10 cm was to theoretically provide validation data for Sentinel-1 images with a comparable depth of penetration.

The observations were logged every 15 minutes. The SM lines of drying—no precipitation with stable circumstances—showed consistent patterns with individual daily variations, which we found artefacts due to the abrupt temperature rise. These temperature values have been used as input variables for the SM calculation, resulting in an inaccurate estimation. Soil moisture measurements were selected at 2 am to minimise the temperature effect, of the day chosen because, at this time, there was so little impact on environmental variables, such as light, temperature, and evapotranspiration. The dates were selected based on the availability of Sentinel-1 data. NDVI was determined from the nearest cloud-free dates of the available satellite items, which happened to be Landsat 8 results because Sentinel 2 images have a lot of cloud cover. Thus, Sentinel 1B and field data were from the same day, and the NDVI was 1 to 8 days after the chosen dates, depending on the month. A systemic network was designed to collect geomorphological and soil variations to establish an optimum configuration, which defined the sensor's efficiency. Sensors were installed in all representative geomorphological units of the study area to ensure the thematic coverage.

The pilot site's total area, where in situ soil moisture sensors were installed, was roughly 1,200 hectares, out of 1,700 hectares. This pointed out that more than 70% of the area was included, suggesting a good variance in soil type and land usage. To achieve an effective spatial distribution, the sensors were installed at representative positions in all geomorphological units. Catenas/toposequences have also been identified for the analysis of transition zones. Geomorphological units have been identified by the e–SOTER methodology [34] for soil and digital soil databases.

The total number of sensors installed in the study area ranged from 44 to 76 sensors per month. A total of 36 parcels were then selected per month for in situ soil moisture measurements. However, the main emphasis was on 12 of them having winter wheat since it was the only crop to have been in the field consecutively for the study duration. The vegetation effect was more prominent at the end of the study period than the beginning due to the growth of wheat, which resulted in more noise from the backscatter coefficient. Fieldwork was performed from January 2019 to June 2019 to consider variations in soil moisture during various seasons.

3.1.2. Sentinel 1B Satellite Imagery

The Sentinel-1 constellation part of the European Space Agency (ESA) provides Sentinel 1A and 1B satellites. In this study, satellite imagery of Sentinel-1B (Figure 2) was used to measure soil moisture because it carries the C-SAR sensor, which provides high-resolution imagery regardless of the weather conditions. Sentinel 1B provides images in both singular and dual polarisations within a 12-day cycle, and the specifications of the images used in this study are in Table 1. Hosseini et al. [35] recommended using co-polarized backscatter coefficients that transmit-receive polarisation (VV) because they are less susceptible to system noise and cross-interference than the weaker cross-polarised coefficients (HV and VH). The Synthetic Aperture Radar (SAR) data collection date was then downloaded to ensure accuracy during soil moisture sensor modelling. The dates that fit the field measurement data were picked, i.e. the measured soil moisture date and period (2 am). While this time (2 am) was different of each satellite passing time, the chosen time displayed a consistent night reading that did not affect the temperature, which plays a significant role in measuring soil moisture.

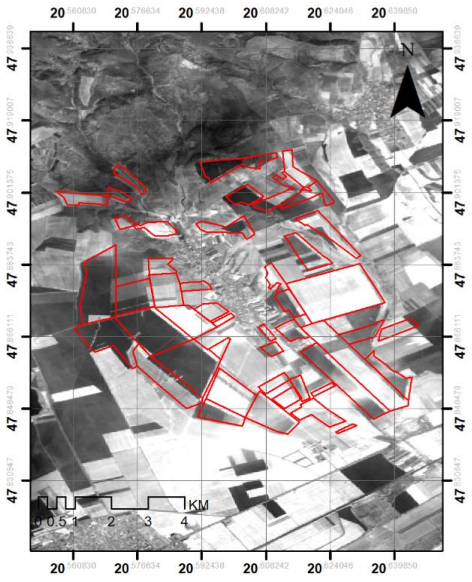

Figure 2. *Sentinel 1B satellite imagery of Tard village with the parcels used for the study delineated in red.*

The specifications of the Sentinel-1B satellite images used are listed below:

Table 1. Specifications of the Sentinel-1B data used in the study

Specifications	Sentinel-1B
Acquisition times	January 2019–June 2019
Imaging Mode	IW.
Imaging frequency	C-Band (5.4 GHz)
Polarization	VV-VH
Data product	Level 1—GRD
Resolution mode	10 m

3.1.3. Landsat 8 Imagery

Landsat 8 comprises Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments that provide high-resolution images. Multi-spectral images have a resolution of 30 m and a revisiting period of 16 days. In this study, OLI was used and of the nine spectral bands; band 4 (Red: 0.64 – 0.67 μm) and band 5 (Near-Infrared: 0.85 – 0.88 μm) were used for deriving NDVI. Four Landsat 8 images were downloaded from January to June 2019 (Table 2). The measured soil moisture dates were matched by the nearest acquisition date of the Landsat 8 image.

Table 2. *Landsat 8 acquisition dates*

Measured Soil Moisture Acquisition Date	Landsat 8 Image Acquisition Date
30 January 2019	21 January 2019
6 February 2019	6 February 2019
10 May 2019	13 May 2019
6 June 2019	14 June 2019

3.2. Methods

3.2.1. Data Collection

Data were collected bi-weekly in the field using a 2-step synchronisation process. Firstly, the sensor was connected to a Bluetooth enabled device, e.g. smartphone or tablet via a preinstalled flower power app. Data collection started automatically once a connection was successful. Secondly, data were then uploaded to the cloud (data synchronisation) either using an internet connection (3G or Wi-Fi).

Finally, data was downloaded from the cloud to an excel CSV format, which was reviewed for any irregularities. If any anomalies were detected, the data point was removed from the raw data set at a given time. Examples of anomalies included missed data as the sensor neglected to register, and irregular readings mostly attributable to low batteries or cracking disruptions. In this study, less than 5 per cent of the data contained anomalies; thus, the data measured were appropriate for review.

3.2.2. Extraction of Covariates

For the selection of multiple covariates, a mixture of literature and statistical processing was used. Based on the literature, covariates had to follow three criteria; firstly, they had to be descriptive of soil-forming factors; secondly, they had to have a clear association with SM; thirdly, they had to be generally accessible [34]. Since many environmental covariates can be obtained from DEM, only a few have been selected, specifically those with a high correlation to soil moisture (Table 3)

Table 3. *Pearson's correlation of field SM data and the selected four terrain variables (2019 March)*

Covariate	Pearson's Correlation
DEM	0.627
Slope	0.552
Aspect	0.489
Relief	0.571

The entire study area's DEM source was extracted from the EU-DEM 1:10 000 using the clipping tool in ArcGIS [37]. The DEM was the most important parameter of this analysis since three environmental covariates were extracted from it. The slope was the first environmental covariate

extracted from the DEM. The slope affects the drainage of surface water, the rate of penetration, and the rate of erosion, both of which are important to soil moisture analysis. In ArcGIS, the slope was measured using the DEM as input raster using the Spatial Analyst Tools/Surface/Slope. The method identifies the slope (gradient or rate of maximum change in z-value) of each raster surface cell.

The next environmental covariate to be derived was the aspect ratio. The orientation of the slopes determines the soil moisture content of the soil. For example, areas with southern slopes were more exposed to sunlight radiation, rendering their surface-level more dry. On the other side, the northern slopes tended to be somewhat wet and more humid. The aspect was extracted using the Spatial Analyst Tools/Surface/Aspect.

Relief intensity (RI) is one of the most widely used surface characterisation environmental covariate. RI is calculated as the difference between the highest and lowest points within an assigned unit of area, i.e. the amount of potential energy. The value is generally expressed in m/area of a circle. Due to the study area's small size, a 500-m diameter circle was calculated using the "focalrange" command in the Focal Statistics toolbox in ArcGIS. Then, flow-accumulation/contribution area was the fourth component since this derivative describes the flow of surface water and the possible transfer of surface water and soil moisture along the slope.

3.2.3. Site geology

The study area is characterised by a dominant NW-SE path valley and range structure, which explains the high significance of the aspects. The agricultural area has heavy textures of clay and loamy sand. The interfluves/plateaus are covered with heavy clay, while the valley slopes with higher gradients are mostly sandy materials beneath the clay strata and vulnerable to the erosion of the covering clay. As a result, the landscape describes much of the soil parent material shifts and soil forms. Of course, this connection is only real geographically, but these kinds of locally valid terrain variables can be extracted from other relatively limited scale places. Similar field data can be extracted from precision farm machinery data so that the method can be extended conveniently anywhere precision farming with real-time kinematic (RTK) navigation is used.

3.2.4. Pre-Processing and Selection of Sentinel 1B satellite Imagery

The Sentinel-1B satellite image Ground Range-Detected (GRD) product was selected because its products consist of SAR data that has been detected, multi-looked and mapped to the ground range using the Earth ellipsoid model as opposed to a single complicated look (SLC). For the GRD products, their ellipsoid projection is corrected using the terrain height defined in the product's general annotation. The downloaded Sentinel 1B images were subjected to analysis to eliminate noise from the images that may affect the final backscatter coefficient. The study was conducted using specially planned pre-processing measures (Figure 3), including the implementation of the orbit file, thermal noise reduction, calibration and ground correction [39], which were performed using SNAP program version 7.03. These procedures were followed to acquire the coefficient of backscatter of Sigma " σ" The unitless backscatter coefficient is then transformed to σ(dB) using a logarithmic transformation using the following formula [40]:

$$\sigma(dB) = log10\,(DN) \tag{1}$$

whereby DN is a raw digital number.

Figure 3. *Sentinel 1B ground range detection (GRD) pre-processing workflow, adapted from [39].*

The selection of Sentinel 1B images was chosen based on the resulting image produced after the pre-processing process. In ArcGIS, corrected raster images were visually analysed to see which image had the least "noise" interference after GRD pre-processing. The degree of noise was determined by comparing the deviation of the pixel value in the test area against Sentinel 1 GRD processed image. The Sentinel-1B images with minimal noise and good backscatter values were chosen on four dates: 30th January, 2nd February, 10th May and 6th June 2019.

3.2.5. Spatial Interpolation Methods

Three interpolation methods were chosen to describe and analyse soil moisture content. The geostatistical interpolation techniques used were multiple linear regression, regression kriging and cokriging. Kriging is a stochastic model that characterises unknown spatial variability. Dobos et al. [41] described kriging's ability to produce spatial prediction in digital soil mapping.

Multiple Linear Regression (MLR): MLR is capable of predicting physical soil properties using multi-spectral data. This study used it to model the relationship between environmental covariates (explanatory variables) and observed soil moisture at a 0.01 significance. The regression model's predicted value included an error variable that showed the difference between the predicted value and the data point's observed value.

Regression Kriging (RK): RK also referred to as universal kriging or kriging with external drift, integrates the regression of the target variable on environmental covariates with the regression residuals [31]. RK was calculated as the sum of the MLR projection and the ordinary kriging of the regression residuals. RK's benefit is its potential to expand the approach to a wider variety of regression techniques and interpret the specified interpolated components [42].

Cokriging (CK): Cokriging is a multivariate variant of ordinary kriging, which uses two or more variables to estimate or predict a primary variable. Cokriging calculates estimates or predictions for a poorly sampled variable and usually reduces the prediction error variance and outperforms the kriging process if the second variable is strongly correlated with the primary variable [43]. This study chose three combinations: DEM + Sentinel, NDVI + Sentinel, and DEM + Sentinel + NDVI.

4. Results and Discussion

4.1. Results

4.1.1. Multiple Linear Regression (MLR.)

An MLR model was used to estimate soil moisture in the upper 10 cm of the soil profile, and the resulting equations are shown in Table 4.

Table 4. *Multiple Linear Regression Equations*

Date	Intercept	Slope	Aspect	Relief	Flow Acc.	Sigma (σVV)	NDVI	No. of Observations
30 January 2019	46.981453	−0.020519	0.007187	−0.004488	−0.000907	−0.337625	−65.115101	76
p-value	*0.000000* *	*0.974165*	*0.628897*	*0.396057*	*0.561509*	*0.243762*	0.430837	
6 February 2019	33.61278	0.031734	−0.005617	0.000073	−0.001149	0.30704	53.4524	76
p-value	*0.000001*	*0.953766*	*0.669784*	*0.988039*	*0.405165*	*0.130149*	0.012407	
10 May 2019	47.490637	−0.400499	−0.014023	−0.000658	0.000131	0.531087	−3.42735	46
p-value	*0.000000* *	*0.461386*	*0.390445*	*0.896895*	*0.931458*	*0.05184*	0.555235	
6 June 2019	47.719149	−0.071528	0.000421	0.001496	0.000924	0.289084	−3.045411	47
p-value	*0.000113*	*0.928151*	*0.982766*	*0.823476*	*0.60737*	*0.376018*	0.804289	

The predicted SM images' spatial structure's interpretation describes weather conditions trends and SM's redistribution due to soil-landscape processes. Figure 4 displays eight SM images of the study area using the same colour scheme.

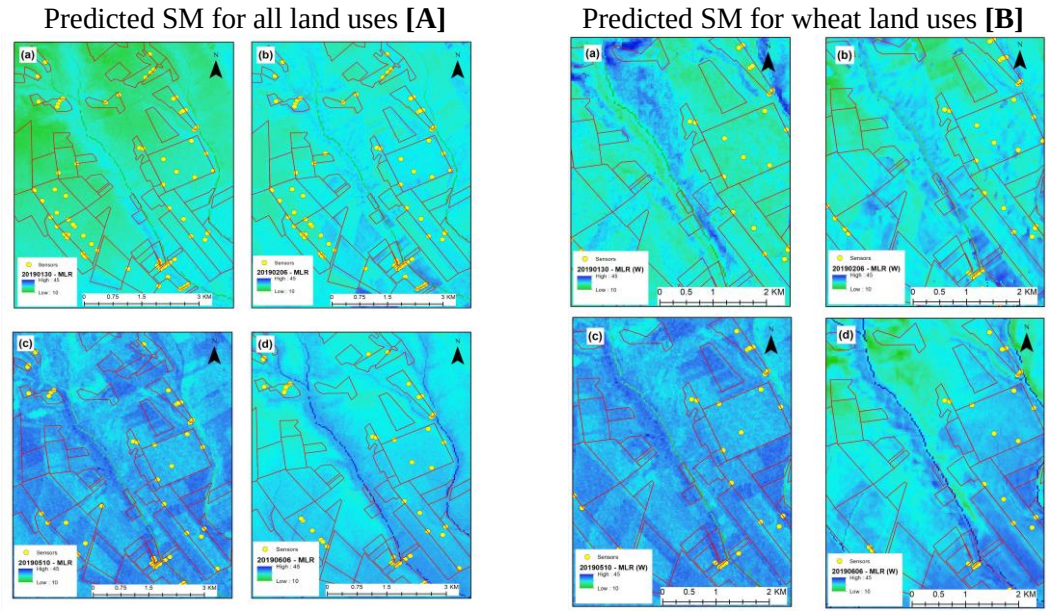

Figure 4. *MLR - All land use* ***[A]*** *and Wheat* ***[B]****: Predicted soil moisture maps 2019 (V%) — (a) January, (b) February, (c) May, and (d) June, respectively.*

We had a dry January and a little wetter February. May and June were much wetter with high rainfall, and increased evapotranspiration attributed to increased biomass and the mild summer weather. This indicates that June's water balance shifted from positive to negative, and the area began to dry up. This mechanism is apparent, with the reintroduction of the greenish colours on the northern part of Figure 4B(d) resulting in a textbook-like example of SM redistribution due to water seepage streaming.

The two sets of images (Figures 4A and 4B) are so different because the "all land uses"(Figure 4A) images are well adjusted to the terrain, due to the more general and variable land-use sampling, so the land use effect is "dissolved". While the images in Figure 4B, where only the wheat land use was

sampled, show a misleading impact on land use. The strong blue areas of Figure 4B(a) and the green ones of Figure 4B(c) are not due to higher or lower soil moisture but instead represent the pastures, forests, orchards and non-agricultural areas around Tard village. Therefore, a general conclusion is deduced that the resulting images show a realistic spatial pattern for the sampled areas in "all land uses" (Figure 4A) but more so in the wheat land use (Figure 4B), which matches expert judgements. The spatial patterns showed the model overestimates the low values and underestimates the high ones while matching values close to the mean. Then, the smoothing impact on the image explains the good realistic spatial patterns, but the values are again over- and underestimated towards extreme values.

4.1.2. Regression Kriging

The MLR model results presented a good spatial structure which explained the deterministic part model and the variability of the model was described as fair. To illustrate the stochastic part by decreasing deviation from the observed values and the spatial SM values, all observation sites' errors were calculated and krigged to develop continuous layers of error distributions study area (Figures 5A & 5B). These layers' values were added to the regression results to correct the deviation from the observed values.

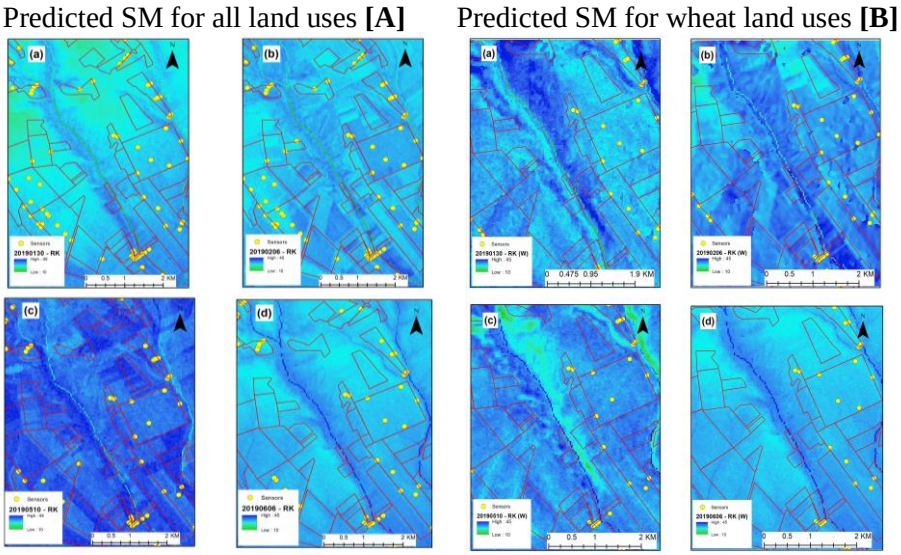

Figure 5. *RK - All land use **[A]** and Wheat **[B]**: Predicted soil moisture maps 2019 (V%)—(a) January, (b) February, (c) May, and (d) June, respectively.*

4.1.3. Ordinary Cokriging

Cokriging using the ordinary kriging method was split into three different covariate setups— one, DEM + Sentinel, two; NDVI + Sentinel, and three; DEM + Sentinel + NDVI which was done for the four dates and two land-use classes. During the months of January and May 2019, cokriging yielded slightly better results in the wheat land-use class than the 'all land-use" class (Figures 6A & 6B). On the contrary, in February and June 2019, cokriging increased the performance of the model of only the "all land use" class (Figure 6A), and the wheat class (Figure 6B) performances always performing

were poorer. This could be attributed to the low number of observation points. For maximum performance, kriging or any spatial interpolation method requires a fair density of the observation sites covering all geomorphologic soil-landscape units' representative points. If there is any gap in the observation network, an unrealistic estimation is possible if real spatial autocorrelation between the neighbouring points is non-existent.

The regression method was not as sensitive to changes; because it uses a point vector base. This is why it performed better in a limited data density scenario. So, the best results in the "all-land use" class were from the NDVI + Sentinel setup, while for the wheat class was always the DEM + Sentinel +NDVI setup. The three cokriging setups' results showed marginal differences in both classes, meaning that the cokriging procedure's covariates could not add much extra information to the kriging-based estimation.

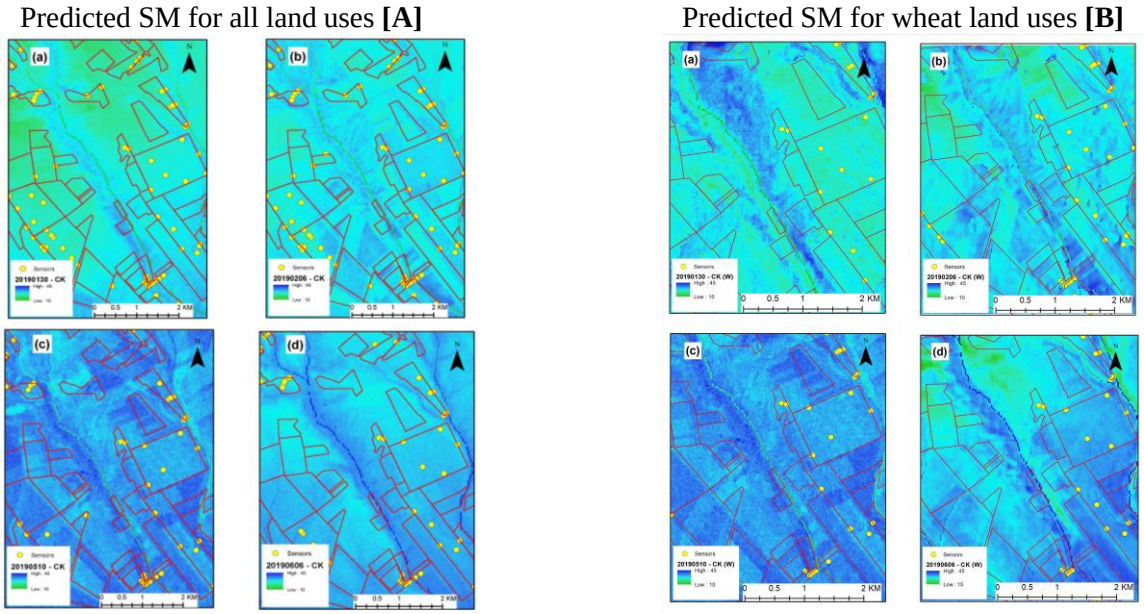

Figure 6. *CK - All land use **[A]** and Wheat **[B]**: Predicted soil moisture maps 2019 (V%)—(a) January, (b) February, (c) May, and (d) June, respectively.*

4.2. Discussion

The results of the three different approaches were developed, tested and compared - per monthly, per land use to determine the best method per month, and the RMSE values were presented (Table 5). The overall results are summarised below for each approach:

1. MLR was comparatively good compared to the cokriging approach, significantly when the results were corrected using the krigged error layers, i.e. the regression kriging approach. The linear regression model provided a very plausible SM distribution across the landscape that suited the established driving forces of the SM distribution, where the RMSE ranged from 3.77 to 5.85. The findings were consistent with those reported by Chatterjee et al. [29]. The R^2 values ranged from 0.19 to 0.35, which are relatively low, but more or less correlate to the literature's findings at this scale. These results were similar to Chatterjee et al. [29], who reported an R^2 of 0.24. The reason

for low R^2 values was that the model overestimated the low values and underestimated the high vales. Therefore, it decreased the estimated values' range down to approximately 60% of the measured values. This approach's major advantage was that it is less sensitive to the number of data points and the data density.

Table 5. *Summary of results.*

Date	Land use	No. Observations	RMSE—MLR	RMSE—RK	RMSE—CK (Various)
30 January 2019	All	76	5.85	4.39 *	6.11 (NDVI, σVV)
	Wheat	32	4.67	3.05 *	6.18 (DEM, σVV, NDVI)
6 February 2019	All	76	**5**.18	1.30 *	5.65 (DEM, σVV, NDVI)
	Wheat	41	4.60	2.81 *	5.72 DEM, σVV, NDVI
10 May 2019	All	46	4.14	2.81 *	4.72 (DEM, σVV, NDVI)
	Wheat	31	3.77	1.92 *	4.61 (DEM, σVV, NDVI)
6 June 2019	All	49	5.86	4.18 *	5.37 (DEM, σVV, NDVI)
	Wheat	27	4.76	2.52 *	5.73 (DEM, σVV, NDVI)

2. Regression kriging (RK), due to its refinement of findings, is one of the most common spatial interpolation techniques in soil science [26]. RK results indicated better performance of the empirical model in predicting soil moisture content which recorded an RMSE range of 1.30 to 4.39. These results were dependant on the standard deviation of the measured values, which was considered acceptable considering the sample size, and area mapped. To clarify the unexplained stochastic component of variation, the observation sites' error values were calculated and krigged, assuming a spatial dependence structure in the values-driven by some unexplained landscape factors. After that, the values were then used to correct the regression estimation, and the process was validated with the leave-one-out cross-validation approach in ArcGIS. The calculated RMSE values were significantly lower and often resulted in a considerable accuracy improvement in all land class cases. Despite this, the statistics' improvement depends on the spatial setup and representativity of the sampling design and does not always correct the error evenly over the landscape and land uses.
3. Cokriging results recorded the weakest performance out of the three methods. The covariates did not have a traceable impact, which could be attributed to the low impact of combined covariates kriging approach. There is a rich literature on cokriging as an interpolation method for soil moisture data estimation. For example, Zhang et al. [30] reported RMSE values between 2.62 and 3.01 using different kriging approaches, like ordinary, cokriging, and extended kriging. Also, Zeng et al. [28] reported RMSE in the same range, but their study was performed on a bare soil surface, lacking the vegetation effect on the estimation unlike this study what comprised mainly of winter wheat. On the other hand, Gharechelou et al. [25] reported even better RMSE for an arid area, where the original value range was between 0% and 10% difference in SM with an of RMSE

around 1.8. These examples show that even though the cokriging results were poor, studies highlight other variables that need to be considered to ensure the model's improvement.

5. Summary

In this study, three statistical methods were applied to map soil moisture content using Sentinel-1B satellite data, NDVI images, and terrain information. Multiple linear regression, regression kriging, and cokriging were tested for quick SM mapping using low-cost sensor data from CO. MLR provided a more realistic spatial pattern over the landscape, even in a data-poor environment. RK was found to be a potential tool to refine the results, while cokriging was found to be less effective than the other methods. In relation to our aim of whether CO data can be used as a reliable source, it was concluded that the CO data has a real potential to provide inputs for temporal SM mapping using easy-to-access freely available datasets, such as DEM, Sentinel-1 and Landsat 8 satellite imagery. The recorded RMSE values of the regression kriging are comparable or even better than those reported for similar digital soil mapping approaches and ranged between 1.3 and 4.39. The models always identified the trends, but the value ranges were consistently lower than the original values though the overall data represents major SM distribution trends over the landscape. A positive finding was that the trend describing the moist areas was well explained, whereas the model needed improvement in the dry regions. At the same time, the deviations of the estimation from the observed values increased towards the extremes. Therefore, to fully understand these trends, further studies are needed to overestimate and underestimate predicted values, and the comparisons with capacitance probes could even enhance the models.

6. Acknowledgements

The research reported here was carried out as part of the EFOP-3.6.1-16-2016-00011 "Younger and Renewing University – Innovative Knowledge City – Institutional development of the University of Miskolc aiming at intelligent specialisation" project implemented in the framework of the Szechenyi 2020 program. The realisation of this project is supported by the European Union, co-financed by the European Social Fund.

References

[1] Bablet, A., Viallefont, F., Fabre, S., Briottet, X., Jacquemoud, S.: *Assessment of Soil Moisture Content Using a Multilayer Radiative Transfer Model of Soil Reflectance (MARMIT) in the Solar Domain*. Geophysical. Research Abstracts 2018, 20:17709.

[2] Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Van Zyl, J.: *The Soil Moisture Active Passive (SMAP) Mission*, Proceedings IEEE 2010, 98:704-716. **https://doi.org/10.1109/JPROC.2010.2043918**

[3] Kovács, K.Z., Hemment, D., Woods, M., Van Der Velden, NK., Xaver, A., Giesen, R.H., Burton, V.J., Garrett, N., Zappa, L., Long, D.: *Citizen Observatory Based Soil Moisture Monitoring-The Grow Example*. Hungarian Geographical Bulletin 2019, 68:119-139. **https://doi.org/10.15201/hungeobull.68.2.2**

[4] Dorigo, W.A., Oevelen, P., Wagner, W., Drusch, M., Mecklenburg, S., Robock, A., Jackson, T.: *A New International Network for in Situ Soil Moisture Data*. Eos, Trans. American Geophysics Union 2011,92:141-142. **https://doi.org/10.1029/2011EO170001**

[5] Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Robock, A.: *The International Soil Moisture Network: A Data Hosting Facility for Global in Situ Soil Moisture Measurements*. Hydrology Earth Systems. Science 2011, 15:1675-1698. **https://doi.org/10.5194/hess-15-1675-2011**

[6] Indian People Organising for Change. Climate Change: The Physical Science Basis. Agenda 6, (2007), pp333.

[7] Xaver, A., Zappa, L., Rab, G., Pfeil, I., Vreugdenhil, M., Hemment, D., Dorigo, WA.: *Evaluating the Suitability of the Consumer Low-Cost Parrot Flower Power Soil Moisture Sensor for Scientific Environmental Applications*, Geosciences Instruments Methods Data Systems 2020, 9:117-139. **https://doi.org/10.5194/gi-9-117-2020**

[8] Freitag, A., Meyer, R., Whiteman, L.: *Strategies Employed by Citizen Science Programs to Increase the Credibility of Their Data, Citizen Science: Theory and Practice*. Citizen Science 2016, 1. **https://doi.org/10.5334/cstp.91**

[9] Pocock, M.J.O., Tweddle, JC., Savage, J., Robinson, L.D., Roy, H.E.: *The Diversity and Evolution of Ecological and Environmental Citizen Science*. PLoS ONE 2017, 12:0172579. **https://doi.org/10.1371/journal.pone.0172579**

[10] Gharesifard, M., Wehn, U., Van Der Zaag, P.: *Towards Benchmarking Citizen Observatories: Features and Functioning of Online Amateur Weather Networks*. Journal Environmental Management 2017, 193:381-393. **https://doi.org/10.1016/j.jenvman.2017.02.003**

[11] Woods, M., Hemment, D., Ajates, R., Cobley, A., Xaver, A., & Konstantakopoulos, G.: *GROW Citizens' Observatory: Leveraging the power of citizens, open data and technology to generate engagement, and action on soil policy and soil moisture monitoring*. In IOP Conference Series: Earth and Environmental Science 2020, 509(1):012060. **https://doi.org/10.1088/1755-1315/509/1/012060**

[12] Al-Yaari, A., Wigneron, J.P., Kerr, Y., Rodríguez-Fernández, N.J., O'Neill, P.E., Jackson, T.J., Yueh, S.: *Evaluating Soil Moisture Retrievals from Esa's Smos and Nasa's SMAP Brightness Temperature Datasets*. Remote Sensing Environment 2017, 193:257-273. **https://doi.org/10.1016/j.rse.2017.03.010**

[13] Zhang, G.-L., Liu, F., Song, XD: *Recent Progress and Future Prospect of Digital Soil Mapping: A Review, Journal Integrated*. Agriculture 2017, 16:2871-2885. **https://doi.org/10.1016/S2095-3119(17)61762-3**

[14] McBratney, A.B., Santos, M.M., Minasny, B.: *On Digital Soil Mapping*. Geoderma 2003, 117:3-52. **https://doi.org/10.1016/S0016-7061(03)00223-4**

[15] Boettinger, J., Howell, D., Moore, A., Hartelink, A., Kienast-brown, S.: *Digital Soil Mapping, Springer Netherlands: Dordrecht*, The Netherlands, 2010. **https://doi.org/10.1007/978-90-481-8863-5**

[16] Bakker, A.: *Soil Texture Mapping on a Regional Scale with Remote Sensing Data*, 2012. URL: https://edepot.wur.nl/246954.

[17] Jenny, H.: Factors of Soil Formation: *A System of Quantitative Pedology, Soil Science.: Dover Publications: Mineola*, NY, USA, 1941. **https://doi.org/10.1097/00010694-194111000-00009**

[18] Rossiter, D.G.: *Digital Soil Resource Inventories: Status and Prospects, Soil Use Management* 2006, 20:296-301. **https://doi.org/10.1111/j.1475-2743.2004.tb00372.x**

[19] Malone, B.P., Minasny, B., McBratney, A.B.: *Digital Soil Mapping*, Springer: Cham, Switzerland, (2017); pp1-5. **https://doi.org/10.1007/978-3-319-44327-0_1**

[20] Zribi, M., Albergel, C., Baghdadi, N.: *Editorial for the Special Issue "Soil Moisture Retrieval us-ing Radar Remote Sensing Sensors"*. Remote Sensing 2020, 12:1100. **https://doi.org/10.3390/rs12071100**

[21] Montanarella, L., Pennock, D.J., McKenzie, N., Badraoui, M., Chude, V., Baptista. I.: *World's Soils are Under Threat*. Alice 2016, 2:79-82. **https://doi.org/10.5194/soil-2-79-2016**

[22] Colliander, A., Cosh, MH, Misra, S., Jackson, T.J., Crow, W.T.; Chan, S.; Bindlish, R.; Chae, C., Collins, C.H.; Yueh, SH.: *Validation and Scaling of Soil Moisture in a Semi-Arid Environment: SMAP Validation Experiment 2015 (SMAPVEX15)*. Remote Sensing Environment 2017, 196:101-112. **https://doi.org/10.1016/j.rse.2017.04.022**

[23] Bousbih, S., Zribi, M., El Hajj, M., Baghdadi, N., Chabaane, Z.L., Gao, Q., Fanise, P.: *Soil Moisture and Irrigation Mapping in A Semi-Arid Region, Based on the Synergetic Use of Sentinel-1 and Sentinel-2 Data*. Remote Sensing 2018, 10:1953. **https://doi.org/10.3390/rs10121953**

[24] Xing, M., He, B., Ni, X., Wang, J., An, G., Shang, J., Huang, X. *Retrieving Surface Soil Mois-ture over Wheat and Soybean Fields during Growing Season Using Modified Water Cloud Model from Radarsat-2 S.A.R. Data*. Remote Sensing 2019, 11:1956. **https://doi.org/10.3390/rs11161956**

[25] Gharechelou, S., Tateishi, R., Sharma, R.C., Johnson, B.A.: *Soil Moisture Mapping in an Arid Area Using a Land Unit Area (LUA) Sampling Approach and Geostatistical Interpolation Techniques.* International Journal Geo-Informatics 2016, 5:35. **https://doi.org/10.3390/ijgi5030035**

[26] Keskin, H., Grunwald, S.: *Regression Kriging as a Workhorse in the Digital Soil Mapper's Toolbox*. Geoderma, 326, (2018), pp22-41. **https://doi.org/10.1016/j.geoderma.2018.04.004**

[27] Han, Y., Bai, X., Shao, W., Wang, J.: *Retrieval of Soil Moisture by Integrating Sentinel-1A and MODIS Data over Agricultural Fields*. Water 2020, 12:1726. **https://doi.org/10.3390/w12061726**

[28] Zeng, L., Shi, Q., Guo, K., Xie, S., Herrin, J.S.: *A Three-Variables Cokriging Method to Esti-mate Bare-Surface Soil Moisture Using Multi-Temporal, VV-Polarization Synthetic-Aperture Radar Data*. Hydrogeology Journal 2020, 1:1-11. **https://doi.org/10.1007/s10040-020-02177-z**

[29] Chatterjee, S., Huang, J., Hartemink, AE: *Establishing an Empirical Model for Surface Soil Moisture Retrieval at the US Climate Reference Network Using Sentinel-1 Backscatter and Ancillary Data*. Remote Sensing 2020, 12:1242. **https://doi.org/10.3390/rs12081242**

[30] Zhang, J., Li, X., Yang, R., Liu, Q., Zhao, L., Dou, B.: *An Extended Kriging Method to Interpo-late Near-Surface Soil Moisture Data Measured by Wireless Sensor Networks*. Sensors 2017, 17:1390. **https://doi.org/10.3390/s17061390**

[31] Hengl, T., Heuvelink, G.B.M., Stein, A.: *A Generic Framework for Spatial Prediction of Soil Variables Based On Regression-Kriging*. Geoderma 2004, 120:75-93. **https://doi.org/10.1016/j.geoderma.2003.08.018**

[32] WRB, IWG. *World Reference Base for Soil Resources* 2006, In World Soil Resources Reports; WRB, IWG: 2007. URL: http://www.fao.org/3/a-a0510e.pdf

[33] USDA-NRCS.: *Keys to Soil Taxonomy*, 128th ed., 1998 USDA-NRCS. URL: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/class/taxonomy/?cid=nrcs142p2_053580

[34] Dobos, E., Daroussin, J., Montanarella, L.: *A Quantitative Procedure for Building Physiographic Units Supporting a Global Soter Database*. Hungarian Geographical Bulletin 2010, 59:181-205.

[35] Hosseini, R., Newlands, N.K., Dean, C.B., Takemura, A.: *Statistical Modeling of Soil Moisture, Integrating Satellite Remote-Sensing (SAR) and Ground-Based Data*. Remote Sensing 2015, 7:2752-2780. **https://doi.org/10.3390/rs70302752**

[36] Dobos, E., Michéli, E., Baumgardner, M.F., Biehl, L., Helt, T.: *Use of Combined Digital Elevation Model and Satellite Radiometric Data for Regional Soil Mapping*. Geoderma 2000, 97:367-391. **https://doi.org/10.1016/S0016-7061(00)00046-X**

[37] EEA.: *Digital Elevation Model over Europe (EU-DEM)*. 2013. URL: https://www.eea.europa.eu/data-and-maps/data/eu-dem.

[38] Dobos, E., Daroussin, J.: *Calculation of Potential Drainage Density Index (PDD) Potential Drainage Density Index (PDD)*. In Digital Terrain Modelling, Springer, Berlin Heidelberg, (2007), pp. 283-295. **https://doi.org/10.1007/978-3-540-36731-4**

[39] Filipponi, F.: *Sentinel-1 GRD Pre-processing Workflow*. Proceedings of 3rd International Electronic Conference on Remote Sensing 2019, 18(1):1-4. **https://doi.org/10.3390/ECRS-3-06201**

[40] SNAP Software.: *SAR Basics TutorialHelp Document*, SNAP Software, 2019. URL: http://step.esa.int/docs/tutorials/S1TBX%20SAR%20Basics%20Tutorial.pdf

[41] Dobos, E., Carré, F., Hengl, T., Reuter, H.I., Tóth, G.: *Digital Soil Mapping. as a Support to Production of Functional Maps*. EUR 22123 EN, Office for Official Publications of the European Communities: Luxemburg, 2006, pp.68. URL: https://esdac.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/other/EUR22123.pdf

[42] Hengl, T., Heuvelink, G.B.M., Rossiter, D.G.: *About Regression-Kriging: from Equations to Case Studies*. Computers & Geosciences 2007, 33:1301-1315. **https://doi.org/10.1016/j.cageo.2007.05.001**

[43] Adhikary, S.K.; Muttil, N.; Yilmaz, A.G.: *Cokriging for Enhanced Spatial Interpolation of Rainfall in Two Australian Catchments*. Hydrological Process 2017, 31:2143-2161. **https://doi.org/10.1002/hyp.11163**