Article

Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture based on Linear Geostatistics

Ricardo M. Llamas[1], Mario Guevara[1], Danny Rorabaugh[2], Michela Taufer[2] and Rodrigo Vargas[1] *

[1] University of Delaware, Dept. of Plant and Soil Sciences, Newark, Delaware, USA; rllamas@udel.edu (R.LL.), mguevara@udel.edu (M.G.)

[2] University of Tennessee, Dept. of Electrical Engineering and Computer Science, Knoxville, Tennessee, USA; dror@utk.edu (D.R.), taufer@utk.edu (M.T.)

* Correspondence: rvargas@udel.edu; Tel.: +1-302-831-1386 (R.V.)

Abstract: Soil moisture plays a key role in the Earth's water and carbon cycles, but acquisition of continuous (i.e., gap-free) soil moisture measurements across large regions is a challenging task due to limitations of currently available point measurements. Satellites offer critical information for soil moisture over large areas on a regular basis (e.g., ESA CCI, NASA SMAP), however, there are regions where satellite-derived soil moisture cannot be estimated because of certain circumstances such as high canopy density, frozen soil, or extreme dry conditions. We compared and tested two approaches--Ordinary Kriging (OK) interpolation and General Linear Models (GLM)--to model soil moisture and fill spatial data gaps from the European Space Agency Climate Change Initiative (ESA CCI) version 3.2 (and compared them with version 4.4) from January 2000 to September 2012, over a region of 465,777 km^2 across the Midwest of the USA. We tested our proposed methods to fill gaps in the original ESA CCI product, and two data subsets, removing 25% and 50% of the initially available valid pixels. We found a significant correlation coefficient ($r = 0.523$, RMSE = 0.092 m^3m^{-3}) between the original satellite-derived soil moisture product with ground-truth data from the North American Soil Moisture Database (NASMD). Predicted soil moisture using OK also had significant correlation coefficients with NASMD data, when using 100% ($r = 0.522$, RMSE = 0.092 m^3m^{-3}), 75% ($r = 0.526$, RMSE = 0.092 m^3m^{-3}) and 50% ($r = 0.53$, RMSE = 0.092 m^3m^{-3}) of available valid pixels for each month of the study period. GLM had lower but significant correlation coefficients with NASMD data (average $r = 0.478$, RMSE = 0.092 m^3m^{-3}) when using the same subsets of available data (i.e., 100%, 75%, 50%). Our results provide support for OK as a technique to gap-fill spatial missing values of satellite-derived soil moisture products across the Midwest of the USA.

Keywords: Soil Moisture; Remote Sensing; Geostatistics; Gap-Filling; Midwestern USA

1. Introduction

Addressing global environmental challenges requires knowledge and information derived from the most accurate and complete available datasets. Soil moisture has an important role in the water and energy cycles, and is regarded as one of the essential terrestrial climate variables [1] due to its influence in soil and atmosphere feedbacks. Furthermore, soil moisture is a critical input variable for applications such as climate modeling [2–4], agricultural planning [5,6], and carbon budget analyses [7,8]. Because of the importance of soil moisture, there are many in situ monitoring networks, organized at the global [9], regional [10,11], or national-scale [12–15]. Despite these national to global efforts, there is still a challenge to represent spatially explicit soil moisture information across large regions related to spatial limitations of point measurements.

Soil moisture can be estimated using remote sensors (e.g., spaceborne radiometers and radar sensors) to provide coarse-scale estimates on a regular basis [9,16]. Examples of remote sensing soil moisture monitoring systems include NASA's Soil Moisture Active Passive (SMAP) [16], ESA's Soil Moisture and Ocean Salinity (SMOS) [17] and the European Space Agency Climate Change Initiative (ESA CCI) [11,18] that deliver publicly available data for a wide range of applications. Despite

doi:10.20944/preprints201909.0126.v1

advances in remote sensing technology, there are still large areas where soil moisture information is not regularly acquired, yielding information gaps in time and space across the world. Missing information arises from certain circumstances such as high canopy density, snow and ice cover, extremely dry surface conditions or frozen soil [11]. These factors hinder radiometers or radar sensors in measuring the dielectric constant in the top layer of soil in order to estimate water content [19].

Consequently, there is a need to develop gap-filling strategies to provide spatially complete satellite-derived soil moisture data across the world. In the most recent version of the ESA CCI product (version 4.4), active and passive sensors are combined by means of a weighted mean, being proportional to the signal-to-noise ratio (SNR) [20]. These ratios are estimated using Triple Collocation Analysis, which is a method that estimates random error variances of three collocated datasets of soil moisture estimates [21]. In areas, where no triple collocation analysis estimates are available, soil moisture values are estimated using a polynomial regression between the signal-to-noise ratios [20]. Other statistical methods (e.g., discrete cosine transformations and singular spectrum analysis) have been applied to fill spatial gaps for satellite-derived geophysical datasets, as well as soil moisture from field measurements [22,23]. These approaches are focused either on the statistical distribution of the data, or in three-dimension information, which includes both space and time. We postulate that alternative gap-filling methods could take advantage on the information contained in the spatial distribution of soil moisture or its relationship with key geophysical variables, such as temperature and precipitation which have been identified in other studies [3,9,24].

In this research, we propose to test the performance of two methods to gap-fill satellite-derived soil moisture in ESA CCI product version 3.2, in which no further gap-filling techniques have been applied. Although version 4.4 includes a gap-filling strategy (as described above), this version still contains gaps across many regions of the world. Our approaches aim to offer alternative strategies to offer spatially complete soil moisture estimates, other than the methods applied in ESA CCI product version 4.4 [21].

We tested two approaches. The first one is based on Ordinary Kriging (OK) spatial interpolation [25–27] to take advantage of the spatial autocorrelation of satellite-derived soil moisture on gridded surfaces. The second one is based on the application of General Linear Models (GLM) to explore the relationship between soil moisture (response variable) with precipitation and minimum air temperature (explanatory variables). We tested these two methods because OK has the advantage of requiring solely spatial soil moisture data, and GLM has the advantage of benefiting from the incorporation of geophysical covariates.

We focused our study over a region in the Midwestern United States (with abundant satellite-data estimates and in situ measurements) between 2000 and 2012. We evaluated the outcome of our gap-filling approaches with ground-truth information using *in situ* measurements from the North American Soil Moisture Database (NASMD) [15]. Although we present monthly results for both OK and GLM, overall correlation coefficient with field data for the period of this study shows that OK performed better than GLM, using as reference the correlation coefficient (e.g. Pearson) between original satellite estimates and ground-truth data from NASMD. In addition to other techniques aimed to gap-filling in satellite-derived geophysical datasets [22,23], our results highlight the potential of spatial interpolation to offer spatially complete soil moisture information. Furthermore, methods based on the spatial distribution of soil moisture, such as OK, might compensate for the lack of geophysical covariates information such as precipitation and air temperature in different regions across the world.

Section 2 provides a description of the region of interest as well as the parameters to select our time frame. Data acquisition, preprocessing, selection of the geophysical covariates, application of proposed gap-filling approaches and validation strategy are also described in Section 2. Section 3 describes the performance of both OK and GLM techniques, as well as the results of cross-validation for both models. Validation using reference correlation between original satellite data and ground-truth soil moisture information is also described in Section 3 and compared with models outputs. Section 3 finally shows the capability of our methods to reproduce the spatial soil moisture patterns shown by the original ESA CCI product. Section 4 proceeds with the discussion of our findings and

 doi:10.20944/preprints201909.0126.v1

their implications in providing spatially complete soil moisture information derived from ESA CCI satellite estimates, version 3.2. Section 5 summarizes important remarks of our work and its implications in providing soil moisture information for specific applications.

2. Materials and Methods

2.1. Region of Interest

The selected region of interest was an area of 465,777 km^2 (Figure 1a) centered in the state of Oklahoma (180,986 km^2) and covering some areas of surrounding states within Midwestern USA: Texas (159,489 km^2), Colorado (11,210 km^2), Kansas (61,343 km^2), Missouri (10,844 km^2), New Mexico (18,550 km^2) and Arkansas (23,356 km^2). The region of interest shows a variety of environmental conditions, both natural and human-driven, that allowed us to test the spatial performance of our gap-filling frameworks. This diversity mitigates bias due to specific environmental conditions (e.g., homogenous land cover, uniform topographic features), which are not the attention of this present study. The region of interest for this study was selected in response to the availability of ground-truth data in that area, mainly over Oklahoma, where MESONET provides a robust set of historical soil moisture records [28]. Additionally, soil moisture data availability in northern Texas and the remaining areas in the region of interest are consistently represented by the NASMD. We highlight that the NASMD integrates data from several monitoring networks including MESONET [15].

The region of interest (Figure 1a) includes a wide variety of land cover types (Figure 1b) dominated by grassland (35.5%), cropland (31.9%) and shrubland (11.0%) in the central and western areas. Whereas forested areas are mostly located in the eastern portion, distributed across needleleaf (2.2%), broadleaf (10.9%) and mixed forests (0.6%) [29].

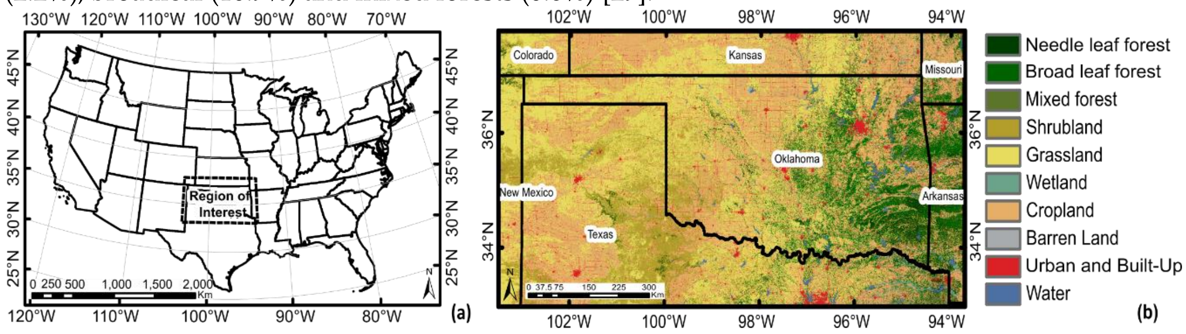

Figure 1. **(a)** Region of interest in the Midwestern USA, where soil moisture gap-filling methods were performed; **(b)** Land cover types over the region of interest (30 m), level 1 NALCMS classification [29].

2.2. Data

2.2.1. Satellite-Derived Soil Moisture

For this study, we used the ESA CCI soil moisture product version 3.2 (Table 1) that has gathered historical records from active and passive remote sensors [11,18]. This product provides soil moisture estimates at 0.25 degrees of spatial resolution on a daily basis, from November 1978 to December 2016. Passive sensors were used to generate this product up to 1991, and the combination of both active and passive sensors was incorporated since then [11]. The ESA CCI product was developed in collaboration with Vienna University of Technology (TU Wien) and focuses on the use of data derived from C-band scatterometers. Such as European Remote Sensing Satellites (ERS-½) and METOP, as well as the use of data from multi-frequency radiometers such as Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave Imager (SSM/I), Microwave Imager (TMI), Advanced Microwave Scanning Radiometer (AMSR-E) and Windsat [3]. These sensors are characterized for the suitability for soil moisture retrieval [3].

ESA CCI soil moisture product version 3.2 has not been subject to any further gap-filling technique (as compared to version 4.4), and represents a suitable product for testing the proposed satellite-derived soil moisture gap-filling approaches in this work. Although version 4.4. does mask areas of dense vegetation using Vegetation Optical Depth Layers, as well as flag measurements taken under frozen conditions [21], the product still shows a large number of gaps in our region of interest. Additional analyses of our approaches applied to ESA CCI soil moisture version 4.4 are described in supplementary material (S1).

Table 1. Main characteristics of ESA CCI soil moisture version 3.2 [18].

Title ESA CCI Soil Moisture Version 3.2	
Release date	February 2017
Available products	Active Passive Combined
Scatterometer sensors used	ERS- 1/2 AMI WS ERS-2 AMI WS MetOp-A + B ASCAT SMMR SSM/I TMI
Radiometer sensors used	Windsat AMSR-E AMSR2 SMOS
Available time span	November 1978 - December 2015 (Active and Passive products) August 1991 - December 2015 (Active product)

Daily soil moisture global records from the ESA CCI product were acquired and then cropped to the region of interest. Daily estimates were merged into monthly soil moisture spatial layers using mean and median values, this way; we tackled the lack of daily coverage in areas out of the satellites swath. Monthly mean values initially reduced the number of gaps in daily products but still provided reliable information to identify spatial patterns and trends in our study period. These values then were used to explore their relationship with different geophysical covariates (supplementary material S2). Monthly values can describe soil moisture variability over a few weeks due to soil moisture memory effects, as water content derived from sudden excessive rains or lack of water onset, can generate wetness or dryness conditions that might last for a couple of weeks [2].

An important step in preparing the soil moisture data for analysis is identifying the most relevant summary statistics, such as mean or median. The median value is more useful when data is concentrated on a brief period of the month (because of long data gaps) with an uneven distribution of data [30]. However, mean monthly soil moisture values showed higher correlation with the tested set of geophysical covariates (supplementary material S2). For our study period (January 2000 to September 2012), Figure 2 shows the frequency of spatial and temporal monthly soil moisture data gaps, where no mean values were calculated due to lack of valid pixels. A pixel is considered valid when soil moisture estimates are available from the ESA CCI product over the region of interest.

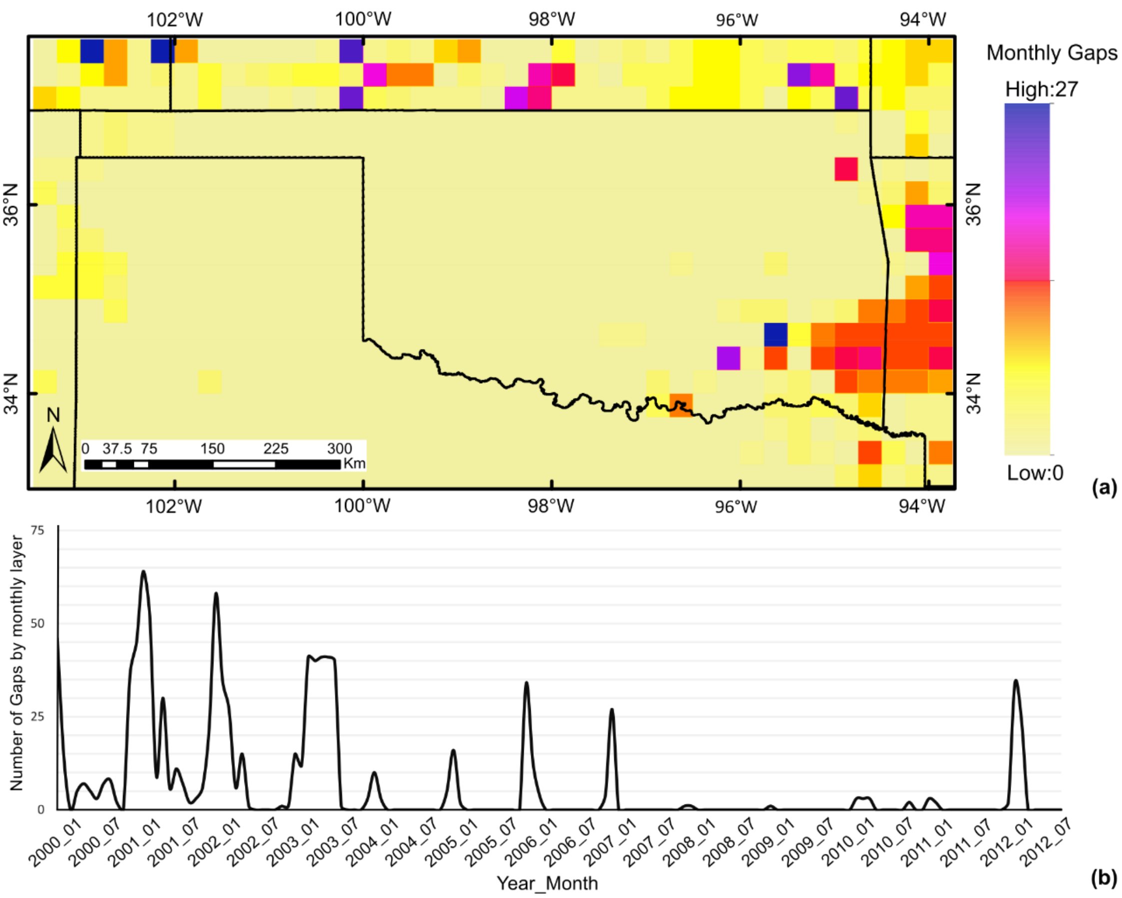

Figure 2. **(a)** Number of gaps in monthly soil moisture estimates over the region of interest, derived from ESA CCI product (version 3.2), January 2000 – September 2012; **(b)** Distribution of gaps along the study period in monthly steps, each number represents the quantity of pixels without data, out of 741 pixels in the region of interest.

2.2.2. Soil Moisture Covariates

For the ESA CCI gap-filling approach using GLM, we explored the relationships between soil moisture and some geophysical variables. Ancillary monthly layers were generated for precipitation, atmospheric temperature, as well as static values of soil texture and topographic wetness index (TWI). These selected variables are known to work as drivers for water input in soil [2,3].

Meteorological data was acquired at 1-km spatial resolution monthly layers produced by the Daily Surface Weather and Climatological Summaries (DAYMET) [31]. Total monthly precipitation, as well as monthly averages of minimum and maximum air temperature raster layers from January 2000 to September 2012 were cropped to the region of interest, projected to the WGS84 Lat-Long coordinate system and resampled to 0.25 degrees by means of nearest neighbor method (ngb) [32].

Soil texture was obtained from the US soil survey geographic database [33] and we classified all classes into four general categories based on the texture triangle from US Department of Agriculture (USDA) [34]: coarse, medium, medium fine, and fine. Soil texture then was resampled to 0.25 degrees resolution using ngb [32]. We calculated TWI using SAGA GIS [35] with a digital elevation model at 250 meters resolution [25], then resampling the output to 0.25 degrees using ngb [32]. Detailed information on the definition of geophysical variables for this work and their further processing are given in the supplementary material (supplementary material S2).

2.2.3. Validation Data

In order to establish a reference value that describes the spatial distribution pattern of soil moisture over our region of interest, we acquired records from the North American Soil Moisture Database (NASMD). NASMD provides the densest possible soil moisture network that integrates field measurements across North America [15]. By 2015, the NASMD had integrated 33 observation networks and two short-term soil moisture campaigns, providing ground-truth data for over 1,800 observation sites in the USA, Canada and Mexico [15]. Some of the densest regional networks integrated by NASMD offer soil moisture data in our region of interest (e.g., MESONET), and records at 5-cm depth, where the soil layer closely interacts with the atmosphere and it is sensed by satellites [36]. We extracted all information available from the NASMD over our region of interest that comprised records at 5-cm depth, from January 2000 to September 2012. Finally, we transformed this data to georeferenced point layers to be integrated in our ground-truth validation approach.

2.3. Gap-filling Methods

Our first gap-filling approach was based on OK interpolation. This technique lead to high uncertainty over areas with very large spatial gaps because it relies on the spatial autocorrelation of available data. Consequently, we also tested a second approach based on GLM to test the relationship between soil moisture and geophysical covariates.

OK interpolation strategy depends solely on separation distance between sampled locations and not on an absolute position [27]. This offers a feasible strategy to fill spatial gaps in areas where no other information is available to be included in similar interpolation methods such as CoKriging or Regression Kriging. This is the most popular among all Kriging methods, as it works in almost any situation and its assumptions are easily filled [27].

General Linear Models (GLM) on the other hand, represent multivariate regression models [37]. In this approach, we assume linear relationships between the dependent variable (soil moisture) and the predefined predictors (precipitation, minimum air temperature) before considering relationships that are more complex. These relationships have been also explored in previous studies, integrating predictors such as vegetation indices, precipitation and temperature [38,39].

Soil moisture spatial-gaps in the region of interest are not always sufficient to test interpolation methods, as in some months there are no gaps over the region of interest. Thus, we decided to randomly remove valid data from each soil moisture monthly layer as well as their correspondent locations on the geophysical covariates layers. Then both OK and GLM were performed on 100%, 75% and 50% of available valid pixels in each month, similar to gap-filling analyses in previous studies [22].

The overall process for soil moisture prediction (Figure 3), derived from the proposed modeling techniques, was evaluated using cross-validation and ground-truth data from the NASMD available from January 2000 to September 2012. An extensive description of workflow as well as a sample process for one month are provided in supplementary material (S3).

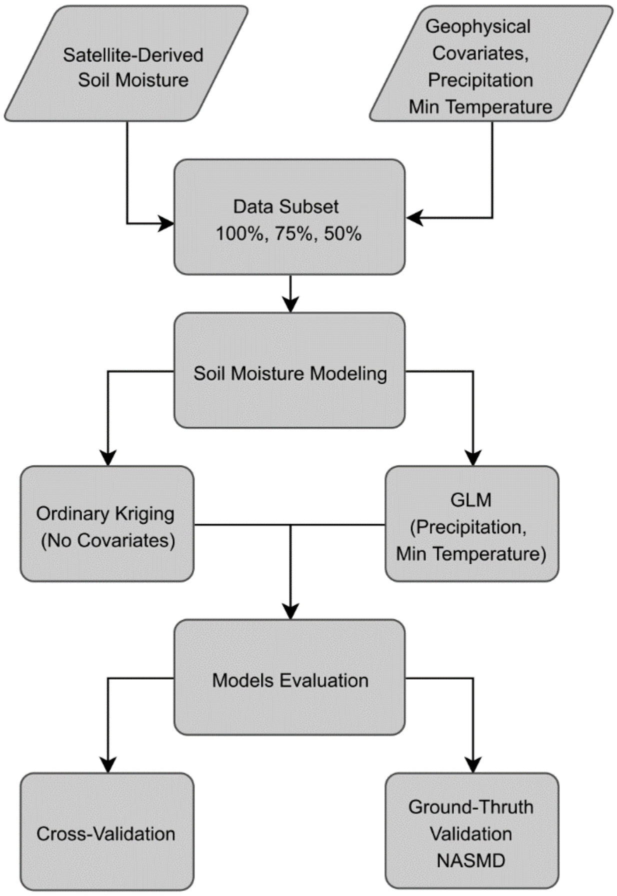

Figure 3. Workflow for soil moisture modeling and the gap-filling over the region of interest, regarding 100%, 75% and 50% of available valid pixels in each monthly layer. Cross-validation as well as ground-truth validation is also described.

2.3.1. Ordinary Kriging

OK was performed using the AutoMap package developed for R statistical platform [40]. By means of the autofit-variogram tool, the best-fitted variogram model was automatically selected to generate independent predictions over each month. Five different variogram models (i.e., Spherical, Exponential, Gaussian, Matern and Stein's parameterization) were evaluated and the one with the smallest residual sum of the squares was selected [40]. The prediction of values at unsampled locations is the linear combination of N variables, as expressed in Equation 1,

$$Z(u) = \sum_{i=1}^{N} \lambda_i \, Z(u_i) \tag{1}$$

where λ_i represents the original weighted values. Weights are calculated as function of distance between sampled and unsampled locations to be predicted. The weights sum must be equal to 1, thus estimations fulfill unbiasedness requirement [41],

After OK spatial interpolation, predicted values as well as their standard errors were obtained for each month, derived in three different cases from 100%, 75% and 50% of available valid pixels. We applied 10-fold cross-validation [40] to OK outputs, for the above mentioned percentages of valid pixels using autoKrige.cv [40]. Finally, we assessed the spatial dependence found in each monthly layer using the nugget-sill ratio. Ratios of at most 0.25 represented strong spatial dependence, between 0.25 and 0.75 moderate spatial dependence, and at least 0.75 weak spatial dependence, as previously reported [42].

2.3.2. General Linear Models

For GLM, we first tested the overall correlation between soil moisture (monthly mean and median values) and each one of the geophysical covariates (monthly precipitation, monthly maximum and minimum air temperature, soil texture and TWI). Secondly, we extracted a time series

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

for each valid pixel along the 153 monthly soil moisture layers, and tested the pixel-individual correlation with each one of the covariates. Finally, we calculated the correlation coefficients of all valid pixels available for each monthly layer with the corresponding temporal layer for each one of the covariates. Based on these analyses we established that the spatial values of mean monthly precipitation and minimum air temperature were the variables with the highest absolute correlation coefficient with mean monthly soil moisture (supplementary material S2). These geophysical covariates were used to predict soil moisture based on GLM, as shown in Equation 2,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (2)$$

where Y_i represents the response variable, X_{i1} and X_{i2} represent the predictor variables, β_0, β_1 and β_2 are the paaameters of the model, and ε_i is the error term [37].

Predictions were also performed for the three predefined subsets (100%, 75% and 50%) of available valid data over the region of interest in each month of the study period. We used the GLM tool from caret statistical package in R [43] to generate independent models for each month, as well as a 10-fold cross-validation process. For this purpose, we used 75% of the data in each independent monthly dataset as training data and 25% as test data.

2.4. Ground-Truth Validation

2.4.1. Reference correlation between NASMD and satellite-derived soil moisture

First, we established a reference correlation value between original satellite-derived soil moisture and data from the NASMD. We extracted all available data from NASMD over the region of interest for each month during the study period and calculated the mean monthly value of soil moisture at 5-cm depth, for each field station. Thus, capturing as much variation as possible from the upper soil layers sensed by the satellites. We tested the correlation between satellite-derived values over each spatially correspondent pixel with soil moisture information derived from the NASMD. This process was performed over the layers using 100%, 75% and 50% of available valid pixels. When there was more than one NASMD station within one corresponding pixel of satellite-derived soil moisture, every station value from within the pixel area was accounted for the correlation analysis with the satellite data. Overall, we used data from 144 stations in the months with the highest availability of field soil moisture records. The use of all NASMD available stations allowed us to retain the overall observation-estimation pairs. Figure 4 shows the number of available NASMD stations available over the region of interest in each month. Across the entire study period, all available stations provided 19,336 points to compare satellite-derived soil moisture estimates and ground-truth data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

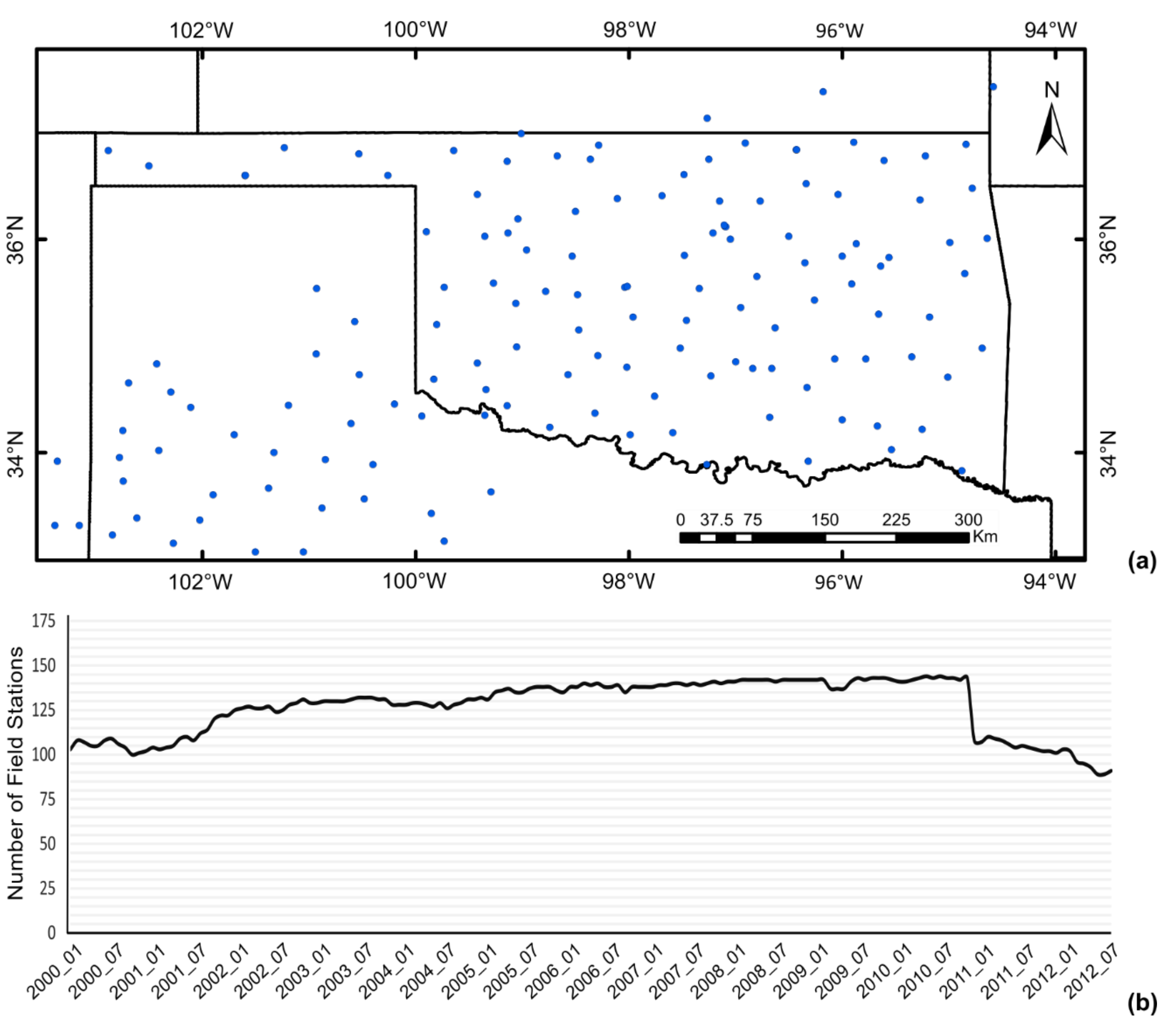

Figure 4. All NASMD stations available for the study period (144); **(a)** Stations are broadly distributed over Oklahoma and northern Texas, however, ground-truth data is scarce over the surrounding states within the region of interest; **(b)** Number of ground-truth stations available per month for validation.

2.4.2. Correlation between predicted soil moisture and NASMD

In order to validate our soil moisture predicted values, we looked for the closest similar correlation coefficient from our outputs and the NASMD, to the correlation coefficient between the original ESA CCI estimates with NASMD. Thus, repeating the same value of a satellite estimate or predicted value for each field station that is located within the same cell. This way we take advantage of as much validation information as possible over our region of interest. We followed the same approach as in section 2.4.1 to evaluate the soil moisture values derived from the modeling approaches with the NASMD. This allowed us to evaluate 19,411 pixels where we calculated the overall correlation coefficient (all months), and monthly correlation coefficients.

3. Results

3.1. OK models selected for soil moisture predictions

Variograms using Stein's parameterization [44] were the most common across the 459 monthly layers (n=382). Exponential (n=69), Spherical (n=6), and Gaussian (n=2) were used in a substantially lower number of predicted soil moisture layers. We found strong spatial dependence in 454 of the monthly layers (nugget-sill < 0.25) and moderate spatial dependence in the remaining 5 layers (0.25 < nugget-sill < 0.75) (Figure 5a). Stein's parameterization [44] had an overall relatively small sum of squared errors when compared to most variogram models and when considered the large sample size (Figure 5b).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

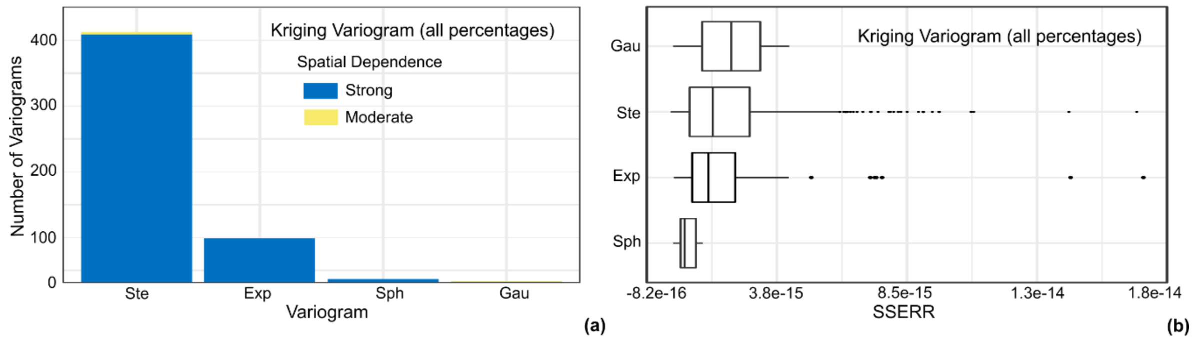

Figure 5. (a) Most fitted variogram models used to predict soil moisture, 459 models generated for 153 monthly layers derived from all percentages (100%, 75% and 50%) of valid pixels; **(b)** Boxplots of the sum of squared error (SSERR) for each set of variogram model.

3.2. Cross-Validation of predicted values

Overall, both models had good cross-validation results, but OK had consistently higher correlation coefficients and lower RMSE (Table 2). These results were consistent as different percentage of available data was used.

Table 2. Cross-validation outputs for OK and GLM, all predicted and observed values along the 153 monthly layers.

Method	Percentage of data	Correlation	RMSE
	100%	0.968	0.012 m^3m^{-3}
OK	75%	0.967	0.012 m^3m^{-3}
	50%	0.963	0.013 m^3m^{-3}
	100%	0.813	0.028 m^3m^{-3}
GLM	75%	0.813	0.028 m^3m^{-3}
	50%	0.814	0.028 m^3m^{-3}

Additional cross-validation between predicted and observed values by month (January-December) is reported using Taylor diagrams (Figure 6), which simultaneously report correlation coefficient, normalized standard deviation and centered root mean squared error [45]. The Taylor diagrams [46] consistently show that OK had higher correlation coefficient and lower centered RMSE and standard deviations, and consequently were closer to the observations. These results were consistent regardless the percentage of available data used. Overall, OK had correlation coefficients ranging from 0.924 to 0.971, whereas GLM values ranged between 0.618 and 0.83. Centered RMSE values between observed and predicted values with OK ranged between 0.227 and 0.377, whereas GLM ranged between 0.558 and 0.786.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

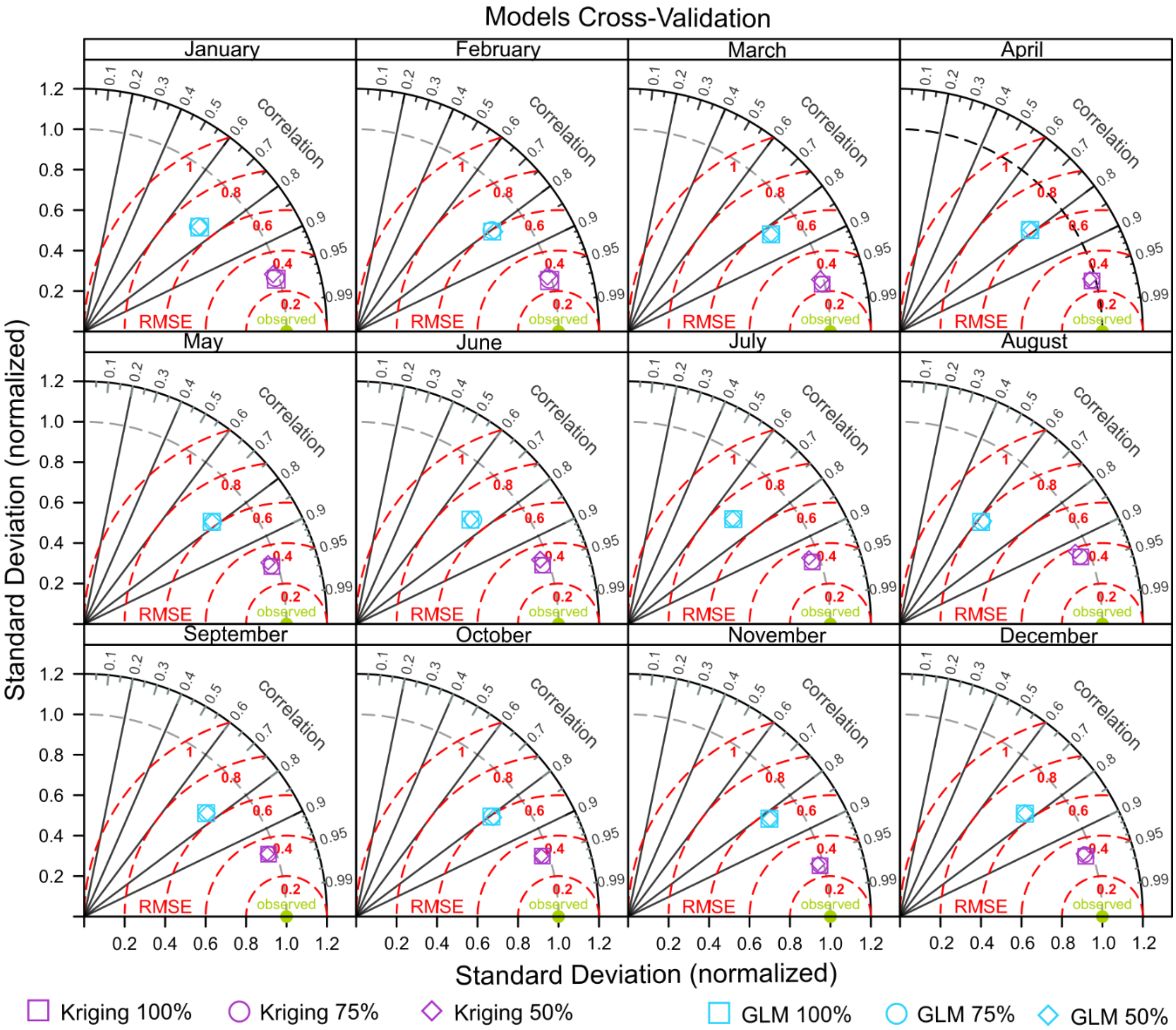

Figure 6. Taylor diagrams based on Cross-validation by month across years, OK and GLM using 100%, 75% and 50% of available valid data. Standard deviation and RMSE values for each output have been normalized using the standard deviation of the observed values.

3.3. Ground-Truth Validation with NASMD

We found an overall correlation coefficient of r = 0.523 and a RMSE of 0.093 $cm^3\,cm^{-3}$ between the original ESA CCI data and the available NASMD stations across the study period (153 months). These values served as baseline and showed that values generated using OK were closer to the reference than those using GLM (Table 3).

Table 3. Overall correlation coefficients between all ground-truth validation points and CCI soil moisture product, as well as gap-filled outputs. Percentages show the data subset used to predict soil moisture values over the region of interest.

Method	Percentage of data	Correlation	RMSE
CCI	100%	0.523	0.092 m^3m^{-3}
	100%	0.522	0.092 m^3m^{-3}
OK	75%	0.526	0.092 m^3m^{-3}
	50%	0.530	0.092 m^3m^{-3}
	100%	0.478	0.092 m^3m^{-3}
GLM	75%	0.478	0.092 m^3m^{-3}
	50%	0.478	0.092 m^3m^{-3}

We explored the temporal dynamics of the correlation coefficients and RMSEs by month throughout the study period. Figure 7 shows the R-squared values between the monthly correlation coefficients from ground-truth data and CCI products, and coefficients from ground-truth data and predicted values by both OK and GLM. RMSE is reported in the same manner (Figure 7). Consistently, OK correlation coefficients with ground-truth data, are closer to correlation coefficients used as reference between validation data and CCI product (Figure 7a). On the other hand, GLM outputs show lower general R-squared values between the outputs and the reference, and are hardly fitted to the regression line (Figure 7b). In a similar way, R-squared between RMSE values from CCI product and ground-truth data, as well RMSE from model outputs and ground-truth data, show closer relation for OK (Figure 7a) outputs rather than GLM (Figure 7b).

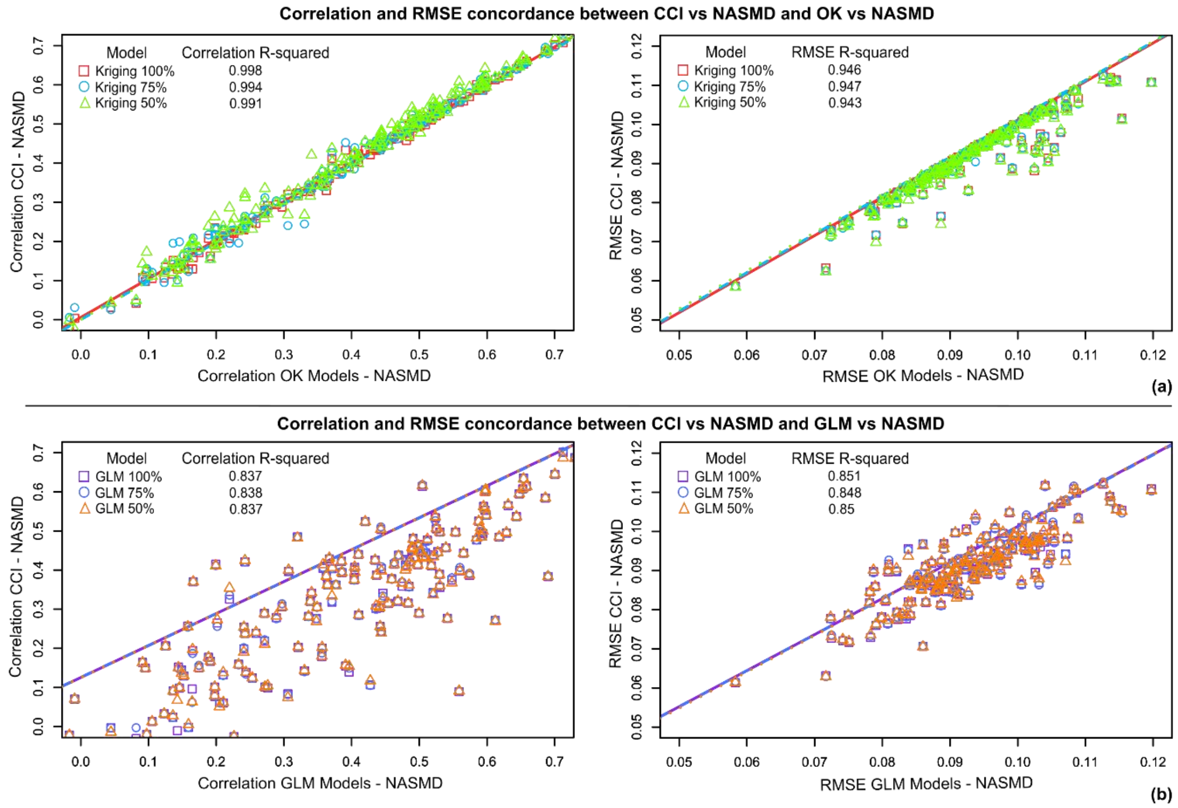

Figure 7. R-squared values showing the concordance between the reference validation (CCI-NASMD) and the validation of proposed gap-filling methods with NASMD. Each point represents one month, using 100%, 70%, or 50% of the available data, with the percentage indicated by the shape used. (a) Validation of Kriging with NASMD, with correlation on the left and RMSE on the right; (b) Validation of GLM with NASMD, with correlation on the left and RMSE on the right. Regression lines between correlated datasets are shown in each plot.

3.4. Spatial Gap-Filling performance of modeling methods

The comparison between the outputs of our modeling methods in contrast with the original ESA CCI soil moisture product shows that OK approach better reproduces the spatial pattern captured by satellite estimates. Figure 8a shows the mean soil moisture estimates from ESA CCI product derived from 153 monthly layers in our region of interest, without any gap-filling technique. In comparison to the original spatial distribution of soil moisture, OK visually shows a more simmilar patterns, independently of the percentage of valid pixels used for modeling (Figure 8b,c,d). On the other hand, GLM clearly shows a less effective performance in reproducing soil moisture spatial patterns, regardless the percentage of valid pixels included in the modeling process (Figure 8e,f,g). We also found that density distribution that depicts the mean soil moisture values along the study perido in the original ESA CCI is better reproduced by our OK approach The perfomarce of OK is similar, either using 100% 75% or 50 % of available valid data (Figure 9a). Meanwhile, GLM density

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

distribution falls apart from the values showed in the original ESA CCI product (Figure 9b). By means of this analysis we confirm the better performance of OK, not only in a direct modeling approach (Figure 6) and its better relationship with validation data (Figure 7), but also its higher capacity in reproducing original ESA CCI spatial soil moisture patterns.

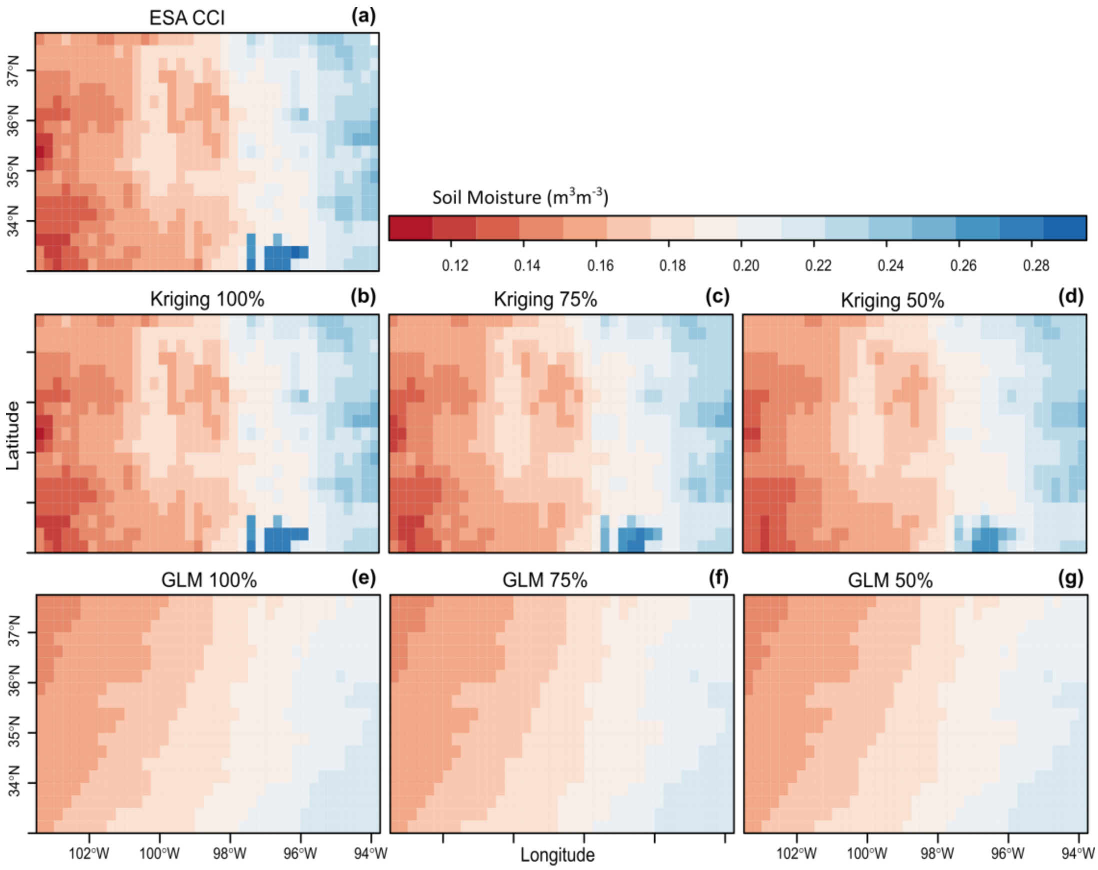

Figure 8. Mean soil moisture values along the study period (January 2000 – September 2012) over the region of interest. **(a)** Mean values of original ESA CCI soil moisture estimates, no gap-filling methods applied; **(b)** Soil moisture mean values modeled using OK and 100% of available valid data; **(c)** Soil moisture mean values modeled using OK and 75% of available valid data; (d) Soil moisture mean values modeled using OK and 50% of available valid data; (e) Soil moisture mean values modeled using GLM and 100% of available valid data; (f) Soil moisture mean values modeled using GLM and 75% of available valid data; (g) Soil moisture mean values modeled using GLM and 50% of available valid data.

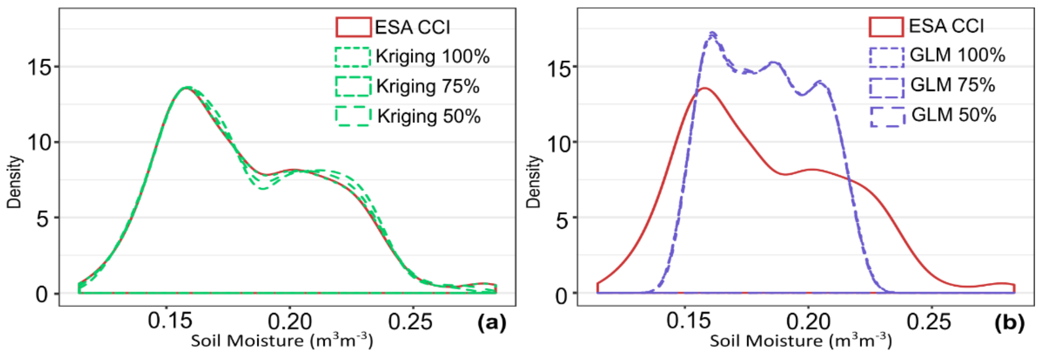

Figure 9. Density distribution of mean soil moisture values along the study period for 741 pixels over the region interest. **(a)** ESA CCI and modelled data using OK with 100%, 75% and 50% of available valid data; **(b)** ESA CCI and modelled data using GLM with 100%, 75% and 50% of available valid data.

4. Discussion

Our results showed that OK and GLM techniques could be used as alternative approaches to gap-filling in soil moisture data derived from ESA CCI product version 3.2. Our proposed methods can be used either in conjunction with ancillary data as precipitation and temperature or using solely the spatial distribution of soil moisture estimates derived from ESA CCI product. Furthermore, our results show that spatial patterns and temporal relations between satellite and ground-truth data are preserved by using OK, and can benefit from GLM approach, although the input data for each method may be different.

Precipitation and minimum air temperature were the strongest correlated environmental covariates with soil moisture (supplementary material S2). These relationships are likely influenced by the grid size (0.25 degrees), as the spatial influence of precipitation and air temperature represents regional and mesoscale climatic patterns [47]. Previous research showed that the increasing in spatial resolution yields more detail in the meteorological information but just limited impacts on its forecasting skill [48]. It is known that from plot to watershed scale soil texture and topography are highly correlated with soil moisture [2,3], but these relationships may change at the coarse scale of the ESA CCI soil moisture product. Thus, these features were not included as geophysical covariates in our GLM approach.

Overall, our results provide support for OK and GLM as techniques to gap-fill spatial missing values of satellite-derived soil moisture products. However, overall performance indicates that OK represents a more reliable method for soil moisture gap-filling in comparison with GLM. Previous studies have compared the advantages of OK for interpolation of spatial soil moisture and other soil properties [49–52] but most analyses have been performed for spatially interpolating soil properties based on field data [24,52–54]. OK has been regarded as an unbiased linear estimator [41] and our results support it as a feasible approach due to the spatial scale of the original ESA CCI estimates (0.25 degrees) under the gap scenarios tested in this work. At this coarse scale, soil moisture values represent a quasi-continuous matrix that meets basic assumptions of Kriging analysis such as stationarity [41] and spatial dependence [52]. OK also incorporates spatial autocorrelation by using the variogram and providing the error variance estimation from predicted values, offering some advantages over deterministic methods such as Inverse Distance Weighting (IDW), which may create noisy fields in interpolation processes. As other Kriging methods, OK is an exact interpolator, which ensures that values at sampled locations are exactly preserved. Thus, we aim to fill the spatial gaps by modeling the entire region of interest, while preserving original values where data existed previously. Additionally, OK performs value predictions based solely on spatial data distribution, offering a suitable approach in cases where no well represented covariates datasets are available over the region of interest as well as it compensates for data clustering [55]. Another evidence in support for OK is the fact that the nugget-sill ratio was less than 0.25 in 99% of the fitted variograms, which implies strong spatial dependence as discussed elsewhere [42].

In the case of GLM, this approach allowed us to explore the most evident relationships between soil moisture in the upper layer of soil and the atmospheric factors that we found to be better correlated (supplementary material S2). We followed a parsimonious principle by means of GLM technique, applying the simplest model with the least assumptions before assuming relationships that are more complex. This parsimonious reasoning and its applications to multivariate models has been explored in other studies [56].

The evaluation of both OK and GLM approaches by means of cross-validation regarding their prediction capacity for actual satellite data, shows similar correlation coefficients as reported by [53] in spatial interpolation of soil moisture, and similar RMSE as reported by [52] for other soil properties. Cross-validation technique has been commonly used in other similar studies [52,53] and offers and initial insights of modeling techniques without considering ground-truth data for validation. Our cross-validation strategy showed that OK better predicted soil moisture values rather than GLM, in spite of pixel removal at different percentages. Regarding cross-validation for monthly grouped values, neither OK nor GLM show an evident bias due to seasonality, as monthly correlation coefficients and RMSE values systematically describe the same patterns found when using data from the entire study period in a single dataset.

In spite of cross-validation results, ground-truth validation was performed to evaluate the suitability of each method (OK and GLM) predicting missing values in the ESA CCI product. We acknowledge the conceptual challenge of this data matching, and the need of balancing ground-truth information in order to be representative of satellite-derived estimates. Representativeness challenges in validation ESA CCI product have been also acknowledged previously [36]. Two main problems are identified [36]. 1) Satellite sensors retrieve ground information from the upper soil layer (0.5 – 5-cm depth), this layer is directly exposed to the atmosphere, therefore, its physical characteristics may differ from the information provided by soil moisture sensors placed at 5-cm depth or deeper. Thus, satellite estimates represent a more variable soil layer, different from soil at deeper layers. 2) Even a spatially extensive soil moisture network cannot cover any area widely enough, to provide scaling representativeness between point-scale measurements and satellite estimates. Field measurements depict soil characteristics in the range of a few square decimeters, while satellite products commonly cover a few kilometers per pixel (~27 km pixel sixes in ESA CCI product). In this regard, our work does not aim to provide strategies of accuracy assessment between field measurements and satellite estimates as explored by [57]. We seek to reproduce the spatial soil moisture patterns expressed by the satellite-derived soil moisture and its actual correlation with ground-truth data.

As proposed by [57], the selection of reliable ground-truth stations and the definition of Core Validation Sites (CSV) represent a step forward in the evaluation of remotely sensed soil moisture. However, regarding the limited availability of ground stations providing soil moisture information, we integrated all available ground-truth data for our region of interest instead of defining CSV. This way, we took advantage of all available field soil moisture records over the region of interest. This approach might introduce uncertainty, as neighboring stations within the same 0.25 degrees pixels in some cases could be affected by different moisture conditions in a large area. However, as our approach aims to reproduce the spatial distribution of soil moisture showed by the satellite estimates based on the correlation with ground-truth data, we aim to retain all the variation offered by NASMD stations.

In order to define the best-tested soil moisture prediction model to fill the gaps in ESA CCI product, correlation founded with ground-truth data was set as reference for our proposed models in every month of the study period. This yielded a more specific way to validate our proposed methods regarding different soil moisture estimates conditions in every month of the ESA CCI product. Given that our research aims to complete spatial information of ESA CCI, reference correlation coefficients helped us to define which model, best reproduces the spatial pattern of the original product. OK showed better results over GLM, as we found, higher the number of valid pixels to shape the variogram parameters, closer the correlation coefficient to the reference. Furthermore, OK performance does not significantly decreases even though valid pixel are artificially removed. On the other hand, GLM correlation with ground-truth data, showed less similar values to the reference, independently of the percentage of valid pixels removed.

Given that, OK and GLM performance for our region of interest does not fall very apart, GLM can be an alternative approach in similar regions where satellite-derived soil moisture estimates are spatially scarce or high clustered. As GLM relies more on predictors availability rather than spatial distribution. Besides, when OK does not meet the best requirements, GLM can ingest data from robust meteorological datasets [58,59] to obtain the geophysical covariates that we used in our analysis. Based on the correlation coefficient between ESA CCI soil moisture product and NASMD ground-truth data, we found that OK consistently better reproduces reference correlation coefficients and RMSE values. Nevertheless, GLM correlation coefficients and RMSE values with NASMD do not significantly decrease from the reference, which still makes this method an alternative approach to gap-filling. Finally, the analysis of the mean soil moisture spatial patterns along the study period showed that OK outputs consistently better reproduced the spatial patterns in the original ESA CCI

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2019 doi:10.20944/preprints201909.0126.v1

product. This can be visually distinguished on the mean soil moisture maps, as well as in the density distribution of the original product in comparison with both OK and GLM outputs.

We acknowledge that OK represents the best-tested method for soil moisture prediction and gap-filling of the ESA CCI product over our region of interest, based on the analysis of the monthly mean values from January 2000 to September 2012. The application of this method in other regions and under different conditions, should consider availability and distribution of soil moisture estimates. Since in large discontinuous areas, stationary can be wrongly assumed, yielding high uncertainty in predicted values. Although ESA CCI soil moisture product has released a new version (4.4), we consider the approaches proposed in this work as an alternative strategy to improve the gap filling process, as the new version still shows large areas with gaps (supplementary material S1).

In future research, it is necessary to explore ESA CCI gap-filling over larger areas such as the conterminous United States, where well spatially represented meteorological datasets are available and different scenarios of gaps distribution can be tested. Daily data must be also incorporated, as this is the temporal resolution in which original soil moisture estimates are delivered. Thus, opening the possibility to operationally filling the gaps in the original soil moisture estimates provided by ESA CCI product. These implementations represent an upscaling need in computational capacities, therefore, High Performance Computing (HPC) techniques must be considered.

5. Conclusions

For the region of interest, linear geostatistics techniques offer a suitable approach to fill the soil moisture spatial gaps of the ESA CCI product (version 3.2). Although version 4.4 follows different strategies to fill data gaps, our research highlights the incorporation of spatial distribution of soil moisture, as well as the use of geophysical covariates to model-missing values. Selected geophysical covariates to model soil moisture in this study, i.e. precipitation and minimum air temperature, can be easily integrated due to their historical availability across larger regions e.g. conterminous United States (CONUS). Selected region of interest provided a spatially extent set of valid pixels from January 2000 to September 2012, which allowed us to test our proposed methods under different scenarios of gap presence, due to natural conditions as well as artificial pixel removal.

Ordinary Kriging method does not need to use any additional covariates as it is built upon the spatial distribution of soil moisture data, however this method can be inconclusive over areas where reference data is highly sparse or clustered (i.e., data scenarios where we found weak spatial structure on satellite soil moisture). General Linear Models on the other hand, offer an alternative to spatially model soil moisture and fill the gaps in the ESA CCI product, however their performance is lower than OK. As we found in this research, soil moisture at coarse scale can be significantly correlated to covariates such as precipitation and minimum air temperature, which can be easily ingested by predicting models over most of CONUS.

Derived from cross-validation for each method and specific percentage of available data, both methods—Ordinary Kriging and General Linear Models—showed significant prediction performance for soil moisture data. However, as we intended to reproduce the soil moisture spatial patterns of the ESA CCI product and its relationship with ground-truth soil moisture data, we considered field validation as the best approach to find the most suitable gap-filling method.

Besides offering information for a wide variety of applications by itself, spatially complete soil moisture information covering large areas can also be related to point-based soil moisture networks to jointly monitoring ecological processes. Thus, soil moisture gap-filled data can yield a better understanding of its role in water and carbon cycles, with important implications in plant and soil respiration, or plant growth. Therefore, influencing our capacity to predict climate change signals in soil moisture estimates from the regional to the global scale.

Supplementary Materials: Supplementary materials S1, S2 and S3 are submitted with this manuscript.

Author Contributions: R.LL., M.G. and R.V. conceived and designed the research. R.LL. and M.G. developed the code for processing and analyzing the data. R.LL. wrote the first draft of the manuscript with input from

R.V. All authors contributed with interpretation of the results, reviewed, and approved the manuscript. R.V. and M.T. supervised and coordinated the research team.

Funding: This study was funded by a University of Delaware Strategic Initiative research grant and the National Science Foundation (OAC grant#1724843).

Acknowledgments: We thank Inder Tecuapetla for his input on the conceptual analysis of spatio-temporal correlation, and Paula Olaya and Joe Teague for their comments to improve this work.

Conflicts of Interest: “The authors declare no conflict of interest.”

References

1. Ward, S. *The Earth observation handbook*; ESA Communication Production Office: Noordwijk, The Netherlands, 2008.
2. Koster, R. D.; Suarez, M. J. Soil Moisture Memory in Climate Models. *J. Hydrometeorol*. **2001**, 2, 558–570, doi:10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2.
3. Seneviratne, S. I.; Corti, T.; Davin, E. L.; Hirschi, M.; Jaeger, E. B.; Lehner, I.; Orlowsky, B.; Teuling, A. J. Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Rev*. **2010**, 99, 125–161, doi:10.1016/j.earscirev.2010.02.004.
4. Meehl, G. A.; Washington, W. M. A Comparison of Soil-Moisture Sensitivity in Two Global Climate Models. *J. Atmos. Sci*. **1988**, 45, 1476–1492.
5. Engman, E. T. Applications of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sens. Environ*. **1991**, 35, 213–226, doi:10.1016/0034-4257(91)90013-V.
6. Narasimhan, B.; Srinivasan, R. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agric. For. Meteorol*. **2005**, 133, 69–88, doi:10.1016/j.agrformet.2005.07.012.
7. Davidson, E. A.; Belk, E.; Boone, R. D. Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. *Glob. Chang. Biol*. **1998**, 4, 217–227, doi:10.1046/j.1365-2486.1998.00128.x.
8. Williams, C. A.; Albertson, J. D. Soil moisture controls on canopy-scale water and carbon fluxes in an African savanna. *Water Resour. Res*. **2004**, 40, 1–14, doi:10.1029/2004WR003208.
9. Dorigo, W. A.; Wagner, W.; Hohensinn, R.; Hahn, S.; Paulik, C.; Xaver, A.; Gruber, A.; Drusch, M.; Mecklenburg, S.; Van Oevelen, P.; Robock, A.; Jackson, T. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci*. **2011**, 15, 1675–1698, doi:10.5194/hess-15-1675-2011.
10. Dorigo, W. A.; Xaver, A.; Vreugdenhil, M.; Gruber, A.; Hegyiová, A.; Sanchis-Dufau, A. D.; Zamojski, D.; Cordes, C.; Wagner, W.; Drusch, M. Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. *Vadose Zo. J*. **2013**, 12, 0, doi:10.2136/vzj2012.0097.
11. Dorigo, W. A.; Gruber, A.; De Jeu, R. A. M.; Wagner, W.; Stacke, T.; Loew, A.; Albergel, C.; Brocca, L.; Chung, D.; Parinussa, R. M.; Kidd, R. Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ*. **2015**, 162, 380–395, doi:10.1016/j.rse.2014.07.023.
12. Robock, A.; Mu, M.; Vinnikov, K.; Trofimova, I. V.; Adamenko, T. I. Forty-five years of observed soil moisture in the Ukraine: No summer desiccation (yet). *Geophys. Res. Lett*. **2005**, 32, 1–5, doi:10.1029/2004GL021914.
13. Martínez-Fernández, J.; Ceballos, A. Temporal Stability of Soil Moisture in a Large-Field Experiment in Spain. *Soil Sci. Soc. Am. J*. **2003**, 67, 1647, doi:10.2136/sssaj2003.1647.
14. Schaefer, G. L.; Cosh, M. H.; Jackson, T. J. The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Ocean. Technol*. **2007**, 24, 2073–2077, doi:10.1175/2007JTECHA930.1.
15. Quiring, S. M.; Ford, T. W.; Wang, J. K.; Khong, A.; Harris, E.; Lindgren, T.; Goldberg, D. W.; Li, Z. The North American Soil Moisture Database: Development and Applications. *Bull. Am. Meteorol. Soc*. **2016**, 97, 1441–1459, doi:10.1175/BAMS-D-13-00263.1.
16. Entekhabi, D.; Njoku, E. G.; O'Neill, P. E.; Kellogg, K. H.; Crow, W. T.; Edelstein, W. N.; Entin, J. K.; Goodman, S. D.; Jackson, T. J.; Johnson, J.; Kimball, J.; Piepmeier, J. R.; Koster, R. D.; Martin, N.; McDonald, K. C.; Moghaddam, M.; Moran, S.; Reichle, R.; Shi, J. C.; Spencer, M. W.; Thurman, S. W.; Tsang, L.; Van Zyl, J. The soil moisture active passive (SMAP) mission. *Proc. IEEE*. **2010**, 98, 704–716, doi:10.1109/JPROC.2010.2043918.

17. ESA SMOS Mission summary Available online: https://earth.esa.int/web/guest/missions/esa-eo-missions/smos/mission-summary (accessed on Jul 13, 2018).
18. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; Haas, E.; Hamer, P. D.; Hirschi, M.; Ikonen, J.; de Jeu, R.; Kidd, R.; Lahoz, W.; Liu, Y. Y.; Miralles, D.; Mistelbauer, T.; Nicolai-Shaw, N.; Parinussa, R.; Pratola, C.; Reimer, C.; van der Schalie, R.; Seneviratne, S. I.; Smolander, T.; Lecomte, P. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ*. **2017**, 203, 185–215, doi:10.1016/j.rse.2017.07.001.
19. Topp, G. C.; Davis, J. L.; Annan, A. P. Electromagnetic Determination of Soil Water Content: Measruements in Coaxial Transmission Lines. *Water Resour. Res*. **1980**, 16, 574–582, doi:10.1029/WR016i003p00574.
20. Chung, D.; Dorigo, W.; De Jeu, R.; Kidd, R.; Wagner, W. Product Specification Document (PSD) D.1.2.1 Version 4.4, 2018, 49.
21. Gruber, A.; Dorigo, W. A.; Crow, W.; Wagner, W. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Trans. Geosci. Remote Sens*. **2017**, 55, 6780–6792, doi:10.1109/TGRS.2017.2734070.
22. Wang, G.; Garcia, D.; Liu, Y.; de Jeu, R.; Johannes Dolman, A. A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environ. Model. Softw*. **2012**, 30, 139–142, doi:10.1016/j.envsoft.2011.10.015.
23. Kondrashov, D.; Ghil, M. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Process. Geophys*. **2006**, 13, 151–159.
24. DeLiberty, T. L.; Legates, D. R. Spatial Variability and Persistence of Soil Moisture in Oklahoma. *Phys. Geogr*. **2008**, 29, 121–139, doi:10.2747/0272-3646.29.2.121.
25. 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, doi:10.1016/j.geoderma.2003.08.018.
26. Cressie, N. The origins of kriging. *Math. Geol*. **1990**, 22, 239–252, doi:10.1007/BF00889887.
27. Oliver, M. A.; Webster, R. *Basics Steps in Geostatistics: The Variogram and Kriging*; Springer: New York, NY, 2015.
28. Brock, F. V.; Crawford, K. C.; Elliott, R. L.; Cuperus, G. W.; Stadler, S. J.; Johnson, H. L.; Eilts, M. D. The Oklahoma Mesonet: A Technical Overview. *J. Atmos. Ocean. Technol*. **1995**, 12, 5–19.
29. Commission for Environmental Cooperation. 2010 Land Cover of North America at 30 meters, Ed. 1. 2017. http://www.cec.org/tools-and-resources/north-american-environmental-atlas/land-cover-and-land-cover-change.
30. Llamas, R. M. R. M.; Bonifaz, R.; Valdés, M.; Riveros-Rosas, D.; Leyva-Contreras, A. Spatial and temporal variations of atmospheric aerosol optical thickness in northwestern Mexico. *Geofis. Int*. **2013**, 52, 321–341, doi:10.1016/S0016-7169(13)71480-6.
31. Thornton, M. M.; Thornton, P. E.; Wei, Y.; Mayer, B. W.; Cook, R. B.; Vose, R. S. Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA, 2018. https://doi.org/10.3334/ORNLDAAC/1345.
32. Parker, J. A.; Kenyon, R. V.; Troxel, D. E. Comparison of Interpolating Methods for Image Resampling. *IEEE Trans. Med. Imaging*. **1983**, MI-2, 31-39, doi:10.1109/TMI.1983.4307610.
33. Soil Survey Staff. Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States. United States Department of Agriculture, Natural Resources Conservation Service, 2016. Available on line at https://gdg.sc.egov.usda.gov/.
34. Bertermann, D.; Bialas, C.; Morper-Busch, L.; Klug, H.; Rohn, J.; Stollhofen, H.; Psyk, M.; Jaudin, F.; Maragna, C.; Einarsson, G. M.; Vikingsson, S.; Orosz, L.; Jordan, G.; Vîjdea, A.-M.; Lewis, M.; Lawley, R. S.; Latham, A.; Declercq, P.-Y.; Petitclerc, E.; Zacherl, A.; Arvanitis, A.; Stefouli, M. ThermoMap - An Open-Source Web Mapping Application for Illustrating the very Shallow Geothermal Potential in Europe and selected Case Study Areas. *Eur. Geotherm. Congr*. **2013**, 1–8.
35. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v.2.1.4. *Geosci. Model. Dev*. **2015**, 8, 1991–2007, doi:10.5194/gmd-8-1991-2015.
36. ESA CCI ESA Soil Moisture CCI Validation, Available online: https://www.esa-soilmoisture-cci.org/validation (accessed on Jul 1, 2019).
37. Kutner, M. H.; Nachtsheim, C. J.; Neter, J.; Li, W. *Applied Linear Statistical Models*; 5th ed.; McGraw-Hill: New York, 2005.

38. Jung, C.; Lee, Y.; Cho, Y.; Kim, S. A study of spatial soil moisture estimation using a multiple linear regression model and MODIS land surface temperature data corrected by conditional merging. *Remote Sens*. **2017**, 9, 1–20, doi:10.3390/rs9080870.
39. Cai, Y.; Zheng, W.; Zhang, X.; Zhangzhong, L.; Xue, X. Research on soil moisture prediction model based on deep learning. *PLoS One*. **2019**, 14, 1–19, doi:10.1371/journal.pone.0214508.
40. Hiemstra, P. H.; Pebesma, E. J.; Twenhöfel, C. J. W.; Heuvelink, G. B. M. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci*. **2009**, 35, 1711–1721, doi:10.1016/j.cageo.2008.10.011.
41. Isaaks, E. H.; Srivastava, R. M. Ordinary Kriging. In Applied Geostatistics; Oxford University Press: New York, USA, 1989; pp. 278–322.
42. Cambardella, C. A.; Moorman, T. B.; Parkin, T. B.; Karlen, D. L.; Novak, J. M.; Turco, R. F.; Konopka, A. E. Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Sci. Soc. Am. J.* **1994**, 58, 1501-1511, doi:10.2136/sssaj1994.03615995005800050033x.
43. Kuhn, M. caret: Classification and Regression Training. R package version 6.0-71. 2016. https://CRAN.R-project.org/package=caret
44. Stein, M. L. *Interpolation of Spatial Data, Some Theory for Kriging*; Springer: Chicago, IL, USA, 1999.
45. Taylor, K. E. Summarizing multiple aspects of a model performance in a single diagram. *J. Geophys. Res.* **2001**, 106, 7183–7192, doi:10.1029/2000JD900719.
46. Lemon, J. Plotrix: a package in the red light district of R. *R-News*. **2006**, 6, 8–12.
47. Tang, C.; Chen, D. Interaction between Soil Moisture and Air Temperature in the Mississippi River Basin. *J. Water Resour. Prot.* **2017**, 9, 1119–1131, doi:10.4236/jwarp.2017.910073.
48. Salathé Jr., E. P.; Steed, R.; Mass, C. F.; Zahn, P. H. A High-Resolution Climate Model for the U.S. Pacific Northwest: Mesoscale Feedbacks and Local Responses to Climate Change. *J. Clim.* **2008**, 21, 5708–5726, doi:10.1175/2008JCLI2090.1.
49. Ding, Y.; Wang, Y.; Miao, Q. Research on the spatial interpolation methods of soil moisture based on GIS. Proc. of Intl. Conf. on. Sci. and Tech., Nanjing, China, March 26-28, 2011, 709–711, doi:10.1109/ICIST.2011.5765344.
50. Chen, H.; Fan, L.; Wu, W.; Liu, H-B. Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought. *Environ. Monit. Assess*. **2017**, 189, doi:10.1007/s10661-017-6244-4.
51. Zhang, M.; Li, M.; Wang, W.; Liu, C.; Gao, H. Temporal and spatial variability of soil moisture based on WSN. *Math. Comput. Model.* **2013**, 58, 826–833, doi:10.1016/j.mcm.2012.12.019.
52. Cruz-Cárdenas, G.; López-Mata, L.; Ortiz-Solorio, C. A.; Villaseñor, J. L.; Ortiz, E.; Silva, J. T.; Estrada-Godoy, F. Interpolation of Mexican soil properties at a scale of 1:1,000,000. *Geoderma*. **2014**, 213, 29–35, doi.org/10.1016/j.geoderma.2013.07.014.
53. De Benedetto, D.; Castrignanò, A.; Quarto, R. A Geostatistical Approach to Estimate Soil Moisture as a Function of Geophysical Data and Soil Attributes. *Procedia Environ. Sci.* **2013**, 19, 436–445, doi:10.1016/j.proenv.2013.06.050.
54. Abuduwaili, J.; Tang, Y.; Abulimiti, M.; Liu, D.; Ma, L. Spatial distribution of soil moisture, salinity and organic matter in Manas River watershed, Xinjiang, China. *J. Arid Land*. **2012**, 4, 441–449, doi:10.3724/SP.J.1227.2012.00441.
55. Brusilovskiy, E. Spatial Interpolation: A Brief Introduction Available online: http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Interpolation.php (accessed on Oct 22, 2018).
56. Seasholtz, M. B.; Kowalski, B. The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta*. **1993**, 277, 165–177, doi:10.1016/0003-2670(93)80430-S.
57. Colliander, A.; Jackson, T. J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S. B.; Cosh, M. H.; Dunbar, R. S.; Dang, L.; Pashaian, L.; Asanuma, J.; Aida, K.; Berg, A.; Rowlandson, T.; Bosch, D.; Caldwell, T.; Caylor, K.; Goodrich, D.; al Jassar, H.; López-Baeza, E.; Martínez-Fernández, J.; González-Zamora, A.; Livingston, S.; McNairn, H.; Pacheco, A.; Moghaddam, M.; Montzka, C.; Notarnicola, C.; Niedrist, G.; Pellarin, T.; Prueger, J.; Pulliainen, J.; Rautiainen, K.; Ramos, J.; Seyfried, M.; Starks, P.; Su, Z.; Zeng, Y.; van der Velde, R.; Thibeault, M.; Dorigo, W.; Vreugdenhil, M.; Walker, J. P.; Wu, X.; Monerris, A.; O'Neill, P. E.; Entekhabi, D.; Njoku, E. G.; Yueh, S. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* **2017**, 191, 215–231, doi:10.1016/j.rse.2017.01.021.

doi:10.20944/preprints201909.0126.v1

58. Menne, M. J.; Durre, I.; Vose, R. S.; Gleason, B. E.; Houston, T. G. An Overview of the Global Historical Climatology Network-Daily Database. *J. Atmos. Ocean. Technol.* **2012**, 29, 897–910, doi:10.1175/JTECH-D-11-00103.1.
59. Fick, S. E.; Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, 37, 4302–4315, doi:10.1002/joc.5086.