


Article

Bayesian Data Fusion Framework for Soil Moisture Interpolation in Entre Ríos, Argentina: Analysis of Topographic Indices

César Augusto Aguirre ^{1,2,3*} , Guillermo Antonio Rondán ¹, Carlos Germán Sedano ¹, Macarena Cardozo ¹ and Tatiana Rut ¹

- 1 Agricultural Sciences School, National University of Entre Ríos, Oro Verde 3101, Argentina; guillermo.rondan@uner.edu.ar (G.A.R.); carlos.sedano@uner.edu.ar (C.G.S.); macacr1368@gmail.com (M.C.); tatau2427@gmail.com (T.R.)
- 2 Paraná Regional School, National Technology University, Paraná 3100, Argentina; cesaraguirre@frp.utn.edu.ar
- 3 National Scientific and Technical Research Council, Paraná 3100, Argentina
- * Correspondence: cesar.aguirre@uner.edu.ar

Abstract: The estimation of soil moisture contents on crop growth is the most important variable and, in many cases, determines yields. Many simulation models predict this variable by considering atmospheric conditions, crop water needs, and soil characteristics. In general, these models simulate soil moisture content for an average plant in the cultivated field without taking into account spatial variability. Relief is one of the characteristics that can explain this variability, so obtaining maps of soil moisture in the root zone has been addressed in this work through spatial interpolations obtained in sampling campaigns, together with the application of the Bayesian data fusion technique. In this work, soil moisture measurements were carried out in the first half of 2021 and the second half of 2022. With these data, several topographic indices were analyzed, finding that the inverse of the topographic wetness index and the digital terrain model best explain the spatial variability of soil moisture. Subsequently, data fusion techniques were applied by combining the results of the Ordinary Kriging interpolation method and these topographic indices. An analysis of the estimation errors was carried out using an independent set of data that did not participate in the spatial interpolations. It is observed that the application of the Bayesian data fusion method, considering these topographic indices, improves the soil moisture estimates compared to the use of the Ordinary Kriging interpolation method alone.

Keywords: soil moisture; spatial variability; geostatistics; Ordinary Kriging; topography



Citation: Aguirre, C. A., Rondán, G. A., Sedano, C. G., Cardozo, M., & Rut, T. (2025). Bayesian Data Fusion Framework for Soil Moisture Interpolation in Entre Ríos, Argentina: Analysis of Topographic Indices. *Agricultural & Rural Studies*, 3(2), 24. <https://doi.org/10.59978/ar03020009>

Received: 12 December 2024

Revised: 4 April 2025

Accepted: 7 April 2025

Published: 28 May 2025



Copyright: © 2025 by the authors. Licensee SCC Press, Kowloon, Hong Kong S.A.R., China. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In agricultural practices aimed at achieving good results in the yield of crops of greater economic interest, one important issue is the estimation of soil moisture in the root zone, as it is essential for the restoration of vegetation, irrigation scheduling, and hydrological modeling (Li, Xu, et al., 2019). It is a vital resource that plays a central role in agricultural areas because it determines the partition of available energy between latent, sensible, and soil heat fluxes and the distribution of precipitation in surface runoff or infiltration, which is a crucial factor for crop growth (Álvarez-Mozos et al., 2005). Different methodologies can be used to obtain three-dimensional soil moisture maps in the cultivated area. The most commonly used are gravimetric measurements at specific sites or the use of soil moisture indices derived from satellite images (Sayago et al., 2017).

Crop growth simulation models are a very useful tool for determining the conditions of growth, development, and the available resources of water and nutrients in the soil. Most of these models are one-dimensional since they simulate the physiological processes of plants considering an average plant in the cultivated area. Heat and moisture fluxes in the soil and at the soil-plant-atmosphere interface are simulated with this criterion (Noilhan & Planton, 1989; Bougeault et al., 1991a, 1991b). However, for several years now, the concept of “precision agriculture” has been in use, with the main objective of incorporating the spatial variability of soil properties in the dosage of products such as fertilizers or spraying and sowing. In this sense, models that allow obtaining estimates of soil temperature and moisture have also evolved, incorporating the resolution of heat and water flow equations in a vertical direction but considering the spatial variability of the agro-hydrological and thermal soil properties in a horizontal direction (Smirnova et al., 1997; Xiu & Pleim, 2001; Chen & Dudhia, 2001a, 2001b; Pleim & Xiu, 2003; Pleim & Guilliam, 2009; Li & Vanapalli, 2021). One of the problems in the use of these models is the initialization of the soil

temperature and moisture values since it needs a three-dimensional matrix of values at the start of the simulation. One solution is to consider that the soil is at “field capacity” when there is abundant and intense rainfall. The problem is that in dry climates this situation is practically impossible, so a large number of soil moisture measurements must be carried out in the region of interest with the consequent cost that this procedure entails. The number of soil moisture samples required to obtain detailed maps is large, making this practice costly in time and resources. An alternative is to sample soil at a few sites and use a spatial interpolation method to obtain soil moisture values at the remaining unsampled sites. Geostatistical concepts (Journel, 1989) are used to apply spatial interpolation methods with rigorous analytical criteria. Karumuri et al. (2023) used this methodology to address uncertainty in input data when experimental measurements of simple and inexpensive physical properties are to be correlated with expensive material properties such as yield strength. Among the most commonly used methods, Ordinary Kriging (OK) can be mentioned (Desbarats et al., 2002; Liao et al., 2016). However, the low density of sampling sites presents some typical configurations such as the so-called “bull’s eyes” that show circles around very low or very high soil moisture values. These configurations are often unrealistic, so the application of secondary data using fusion techniques of primary data (measurements of the variable to be interpolated) and secondary data (such as variables that indirectly explain the behavior of the primary variable). This fusion approach, combining measured data of the primary variable with secondary variables’ data, is widely used in various applications. Among different data fusion methods, the “Evidence theory”, also known as the Dempster-Shafer theory is an imprecise reasoning theory (Zhang et al., 2025), the weighting/voting fusion is often adopted to extract inconsistency features (Feng et al., 2009), neural network fusion, and fuzzy-logic inference. Zamani et al. (2023) use Bayesian Maximum Entropy-based Fusion specifically for the study of water quality variables in a water reservoir. Alizamir et al. (2025) use an innovative data fusion framework based on Bayesian model averaging to monitor dissolved oxygen levels in water. However, in the spatial interpolation of soil moisture, the Bayesian data fusion (BDF) framework is often used for this purpose, taking into account the variations explained by secondary variables related to soil moisture (Zhao et al., 2025) that can be obtained across the entire soil grid—for example, topography (Bogaert & Fusbender, 2007; Fusbender et al., 2008; Peeters et al., 2010; Wang et al., 2018; Li, Wang, et al., 2019) —or updating high-resolution images with coarser image time series using Bayesian data fusion (Fusbender et al., 2009). The limitations in the use of this technique are undoubtedly the availability of soil moisture sample data and a Digital Terrain Model (DTM) with high spatial resolution. Due to this, the DTM is first obtained from aerial photographic survey data carried out in the area with a spatial resolution of 5 m. The objective of this work is to analyze topographic indices to be used as secondary variables within the BDF framework and to use this technique to obtain spatial interpolations of soil moisture in the root zone in the experimental field of the Agricultural Sciences School of Entre Ríos National University (FCA-UNER) during the first months of 2021 and the last months of 2022 that was characterized as “La Niña” period in terms of precipitation variability, with values well below normal.

2. Materials and Methods

2.1. Location of the Study Area and Soil Moisture Sampling Sites

In the experimental farm of the FCA-UNER named Ramón Roldán, located north of the Diamante Department as shown in Figure 1, gravimetric measurements of soil moisture were carried out every 0.2 m depth up to the first meter using a stratified sampling design considering the classes “hill,” “middle hill,” and “low” according to the Topographic Index (TPI) suggested by Weiss (2001) and applied to the experimental field by Tóffoli et al. (2022). Three repetitions of soil moisture data collection were carried out at each sampling site at each visit using the gravimetric method. Soil moisture sampling began on February 25, 2021, and was conducted at 9 sites in the study area. Sampling was conducted at the same sites on March 12 and 19, April 20 and 30, and May 7, 14, and 31, 2021. During the second half of that year, sampling was interrupted due to the health situation caused by the COVID-19 pandemic, which led to the suspension of academic and research activities due to quarantine. Subsequently, during the second half of 2022, soil moisture sampling tasks were resumed at the sites established on August 19, September 9, October 2, 17, and 28, November 18, and December 2, 2022. Each sample was carefully conditioned and taken to the FCA-UNER Soil Laboratory to be weighed, dried in an oven at 105 °C until its weight remained constant, and then weighed again. Gravimetric humidity at location s was obtained by weight difference. The data were subsequently expressed as volumetric moisture using layer-wise bulk density determinations. In addition, soil moisture measurements were carried out in the surface layer (0–0.20 m depth), called AP, during the second half of 2022 with the sole purpose of validating the interpolation methods, that is, these data did not participate in the spatial interpolation process. For this second set of soil moisture data, 5 sites were selected as shown in Figure 2.

The location of the Ramón Roldán farm classified by the TPI (Tóffoli et al., 2022) and both soil moisture sample datasets in the rectangular study area whose dimensions are 1060 m x 880 m are shown in Figure 2. The brown points are located equally in the “hilltop,” “half hill,” and “low” areas according to the TPI, while the white points indicate the sites of additional soil moisture measurements in the AP layer that do not participate in the spatial interpolation processes.

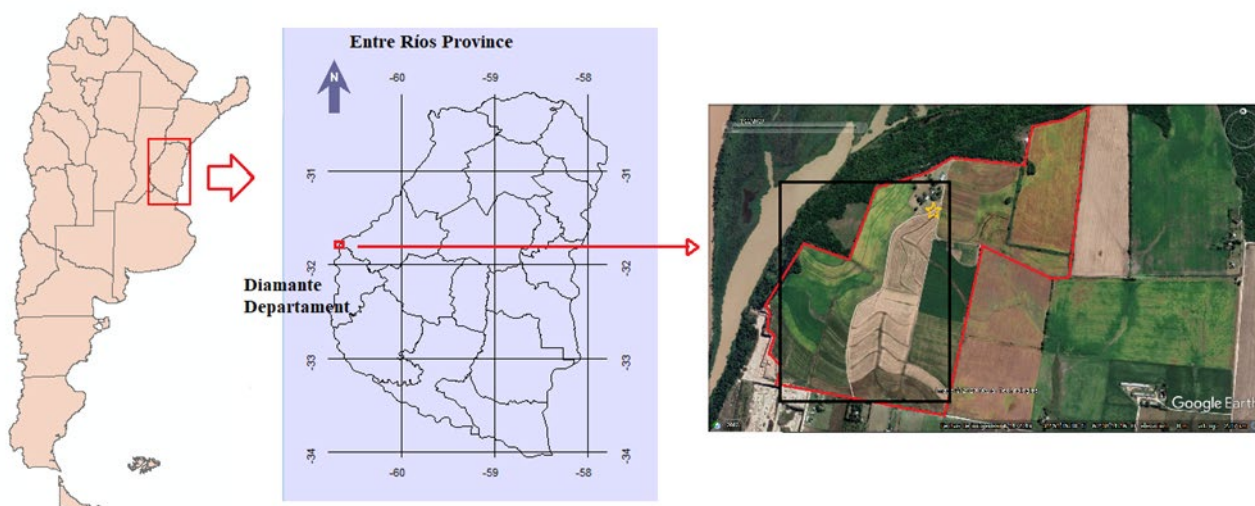


Figure 1. Geographic location of Ramón Roldán experimental farm. The satellite image on the right was obtained on 2019/04/11 from Maxar Technologies Google Earth Pro. The red polygon delimited the perimeter experimental farm and the black rectangle is the study area. ★: Installation site of automatic weather station.

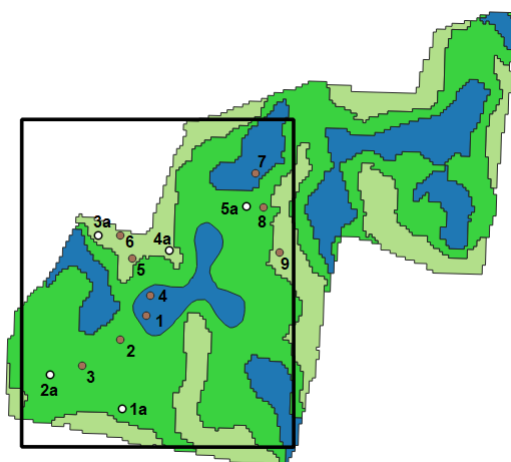


Figure 2. Ramón Roldán experimental farm is classified according to the Topographic Index (TPI), following the methodology of Tóffoli et al. (2022).

- : “Hilltop”; ■ : “Half Hill”; ■ : Low; □ : Area of soil moisture interpolations.
- : sampling sites involved in the interpolations of soil moisture.
- : sampling additional sites for calculating interpolation errors.

2.2. Spatial Interpolation of Soil Moisture

An ordinary Kriging geostatistical interpolation method (OK) was used to perform spatial interpolations of the soil moisture samples. In this case, interpolated soil moisture data grids were generated with a spatial resolution of 10 m × 10 m. This resolution was adopted because the results of the interpolations will be used in combination with data obtained from the analysis of satellite images with this resolution to be used as initial conditions in the soil moisture balance model runs. In addition to the soil moisture values using OK, the standard deviation of the soil moisture interpolations has been calculated for the entire three-dimensional grid and all sampling dates. The OK method was chosen because it has proven useful and popular in many fields. This method produces visually appealing maps from irregularly spaced data. OK attempts to express the suggested trends in the data using the variogram function, which is a mathematical model that indicates

the direction in which the data tend to vary most rapidly. Therefore, the variogram is a function of direction. In this case, one might think that the slope of the land would be associated with a greater variation in soil moisture. It is also logical to assume that the direction of each sampling point towards the lower areas (drainage network) could be represented by a variogram. A detailed description of this method can be found in Cressie (1991) or Journel and Huijbregts (1978). Journel (1989) presents a concise introduction to geostatistics and particularly the Kriging method with its variants. Isaaks and Srivastava (1989) also provide a clear introduction to the subject.

2.3. Bayesian Data Fusion Framework

The technique of data fusion from topographic indices was used to improve the results of spatial interpolations of soil moisture obtained by ordinary kriging (OK). The technique that allows incorporating another set of data from other sources (such as topographic data) into the analysis of spatial variability is known as Bayesian data fusion (BDF). BDF was proposed within its theoretical framework by Bogaert and Fasbender (2007) and was specifically applied in hydrology by Fasbender et al. (2008) to analyze water table depth data measured in the Dijle River basin in northern Belgium. Peeters et al. (2010) also used this technique to improve spatial interpolations of water table depth data in an aquifer located in central Belgium. This method hypothesizes that topographical data can provide information about soil moisture variations that can help improve spatial interpolations of measured data. It is based on the assumption that part of the variability in soil moisture is explained by topography. To prevent other environmental variables from influencing the result, soil moisture sampling should be carried out on a specific day when there are no changes in weather conditions while the sampling process is taking place. In this way, the variability of soil moisture is limited to topographical characteristics.

The data fusion method proposes the estimation of the variable to be analyzed (in this case the soil moisture η_0) at an unsampled site x_0 based on n observations (sample data) $\eta_s = \{\eta_1, \eta_2, \dots, \eta_n\}$ that were taken at locations $x_s = \{x_1, x_2, \dots, x_n\}$ using some spatial interpolation techniques such as OK and complementing the result of the interpolation with the help of magnitudes obtained indirectly from the entire study area, such as topographic indices. Soil moisture values are obtained from a fitting statistical analysis of a secondary function $g(y)$ that depends on the secondary variable $y = \{y_0, y_1, y_2, \dots, y_m\}$ and that correspond to the sites $x = \{x_0, x_1, x_2, \dots, x_m\}$ where m is the total number of pixels covered by the horizontal extent of the grid for a soil layer. This process must be repeated for all soil layers that make up the soil depth. Thus, at the sampling sites x_s , the soil moisture values obtained from the secondary variable are available. The pairs (η_s, y_s) are used to find the fitting function $g(y)$. Then, under the assumption of mutual independence between the secondary variable data and the measured soil moisture data, Bogaert and Fasbender (2007) show that an “a posteriori” probability density function $pdf = f(\eta_0/y_0)$ at the unsampled site x_0 can be derived given a value of the secondary variable at the same site y_0 from the “a priori” probability density function $pdf = f(\eta_0)$ and the conditional probability density functions $pdf = f(\eta_0/y_{0,j})$:

$$f(\eta_0 / y_0) \propto \frac{1}{|f(\eta_0)|^{p-1}} \prod_{i=1}^p f(\eta_0 / y_{0,i}), \quad (1)$$

where p is the number of variables that can be considered in the conditional probability density function, including the one calculated using the OK interpolation method. Thus, it can be seen that all the variables that are desired can participate in the fusion. However, sometimes a secondary variable that has a low correlation with soil moisture can negatively affect the result. For this reason, it is advisable to first perform a correlation analysis to identify the secondary variables that most accurately explain the variability of soil moisture. Assuming that $f(\eta_0)$, $f(\eta_0/y_s)$ and the $f(\eta_0/y_{0,j})$ have Gaussian distributions, one can calculate the mean μ_{BDF} and the variance σ_{BDF}^2 proposed and used by Bogaert and Fasbender (2007); Fasbender et al. (2008) and Peeters et al. (2010):

$$\mu_{BDF} = \left(\frac{\mu_k}{\sigma_k^2} + \sum_{j=1}^{p-1} \frac{\mu_{y_{0,j}}}{\sigma_{y_{0,j}}^2} - \frac{\mu_s}{\sigma_s^2} \right) \sigma_{BDF}^2, \quad (2)$$

$$\sigma_{BDF}^2 = \left(\frac{1}{\sigma_k^2} + \sum_{j=1}^{p-1} \frac{1}{\sigma_{y0,j}^2} - \frac{1}{\sigma_s^2} \right)^{-1}, \quad (3)$$

where μ_k and σ_k^2 are the mean and variance of the spatial interpolation of data by OK, $\mu_{y0,j}$ and $\sigma_{y0,j}^2$ are the means and variances of the j secondary variables $g_{(y)}$ obtained from the data y and μ_s , σ_s^2 are the mean and variance of the data observed in the sampling characterizing the initial or “a priori” probability density distribution. In this way, the μ_{BDF} value obtained in each pixel of the study area will be the most probable value of the soil moisture η_0 .

2.4. Topographic Indices

Topographic indices were obtained throughout the spatial grid of the study area from a Digital Terrain Model (DTM) calculated from a Digital Elevation Model (DEM) available on the National Geographic Institute (IGN) website:

<https://www.ign.gob.ar/NuestrasActividades/Geodesia/ModeloDigitalElevaciones/Mapa> with a spatial resolution of 10 m. The DEMs were created by this institute based on topographic and photogrammetric surveys of the study area in 2013 (IGN, 2017, 2019; Tocho et al., 2020). From the DTM, four topographic indices were calculated: the Topographic Position Index (TPI) proposed by Weiss (2001), the Topographic Wetness Index (TWI) of Beven and Kirkby (1979), the Euclidean distance to the drainage channels (d_{stream}), and an index that relates d_{stream} with the slope called dDTM, both proposed by Fusbender et al. (2008). All these products were obtained using the QGIS 3.24 software.

2.4.1. Obtaining the Digital Terrain Model (DTM) of the Study Area

A DTM represents the bare ground surface above mean sea level without any ground objects such as plants and buildings. A Digital Surface Model (DSM) captures the natural and built/artificial features of the environment above the DTM. A Digital Elevation Model (DEM) is often used as a generic term for DSM and DTM, thus it only represents height information without any additional definition of the surface (Hirt, 2016). The average height of plants or trees can be obtained from the difference between DSM and DTM. In forest management, a Canopy Height Model (CHM) is a leaf area ceiling model derived from elevation data. In forested areas, the difference between the DSM and the DTM can be viewed as a CHM representing the height of trees above ground level or the canopy height. As mentioned above, the IGN provides DEM data with a spatial resolution of 5 meters in the study area that was obtained from analysis of aerial photographs, so the elements on the ground are included in these products. In rural areas where there are no large buildings, these images resemble a DSM, so the DTM was obtained by first digitizing the tree areas using a very high spatial resolution image such as those provided by Google Earth Engineering™. Subsequently, a field survey was carried out to estimate the height of the canopy. These areas were assimilated into a CHM. A roughness image as proposed by Wilson et al. (2007) can be obtained from the so-called “roughness” algorithm of the GDAL plugins in QGIS (see Figure 3a). This algorithm calculates the largest difference between cells of a central pixel and the values of the surrounding cells. The obtained roughness image allows knowing the maximum difference between the height of a pixel and its neighbors when the DEM is processed (not to be confused with the roughness parameter z_0). In Figure 3a, the ravine can be distinguished in dark tones to the northwest and center-west on the image, with values greater than 20 m. Leaving aside the area that goes from the ravine to the river, a group of trees to the east of the ravine are between 7.5 m and 10 m high (orange area), while another group to the northwest is up to 5 m high (light grey area). The row of trees to the southwest is 10 m high while the group of trees near the university offices and storage sheds is 7.5 m; with some scattered 5 m high. To the west of the ravine, a group of trees with different heights ranging from 7.5 m to 15 m (orange, green and blue) can be seen. Figure 3b shows the satellite image obtained from Maxar Technologies Google Earth Pro on 2022/12/12. Figure 4 shows two zoom views of the Figure 3b. It can be observed that the southern part of the ravine has a steeper slope than the central and northern parts. From the above analysis, an approximate CHM can be derived (Figure 5).

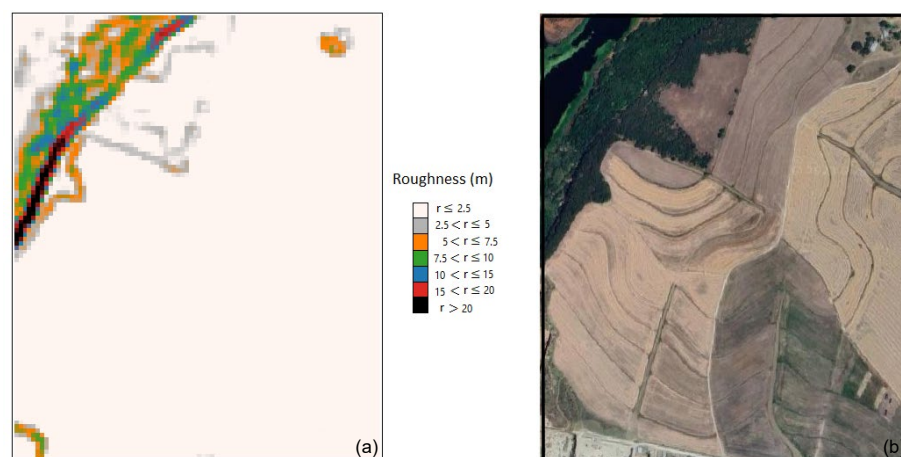


Figure 3. (a): Roughness image (r) obtained from GDAL plugins showing the maximum height differences between each pixel and the surrounding cells. (b): Satellite image obtained from Maxar Technologies Google Earth Pro on 2022/12/12.



Figure 4. Zoom view of Figure 3b. The southern part of the ravine (left) and the central part (right).

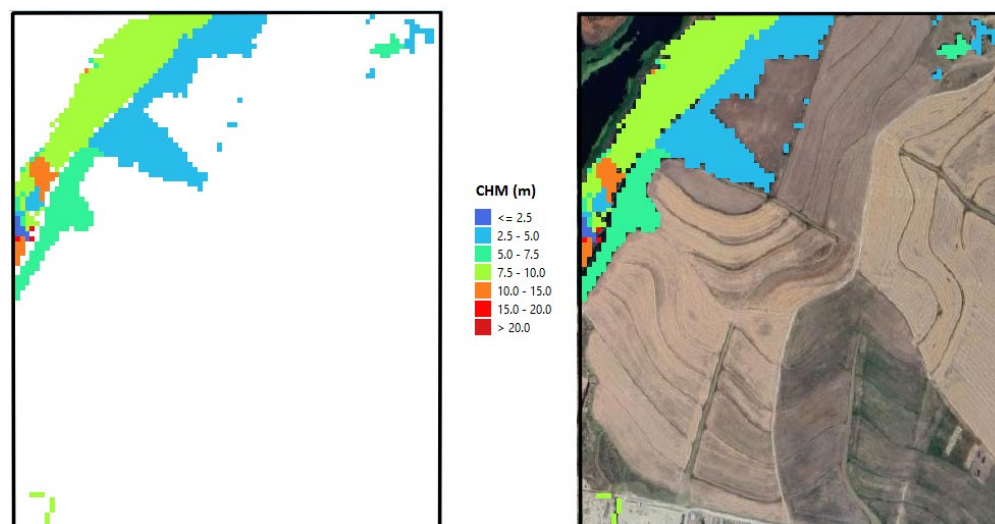


Figure 5. Left: Canopy height model (CHM) obtained from a digitized vector layer using the satellite image (right) from Maxar Technologies Google Earth Pro on 2022/12/12 rasterized using the GDAL tool in QGIS software.

As mentioned above, the DTM was obtained as a difference between the DEM and the CHM. However, due to the error that can occur when estimating the height of tree groups, it is advisable to filter the obtained image using a low-pass filter to remove possible unwanted edges of the tree area. This type of filter was used with a circular kernel of 2 pixels. Figure 6 on the left shows the DTM after performing the difference between DTM and CHM, while the one on the right shows the same image after the filtering process.

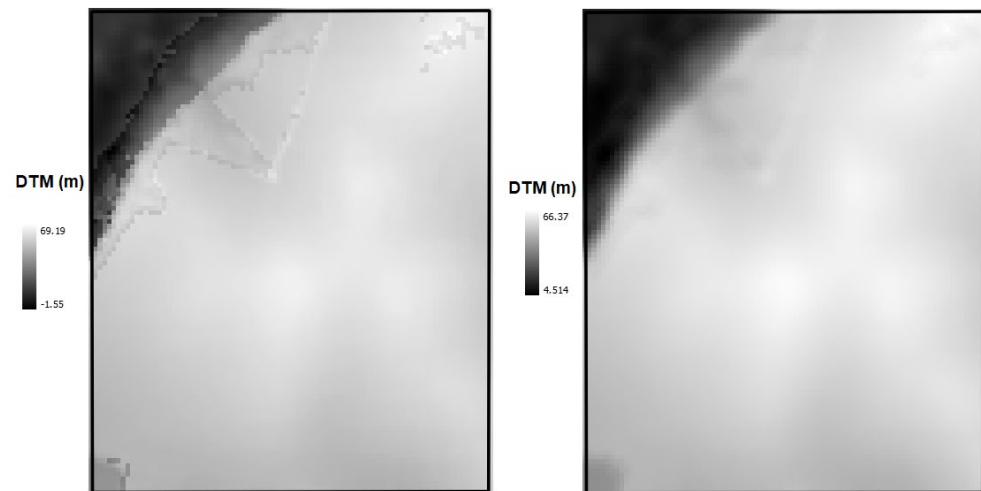


Figure 6. Left: Digital Terrain Model (DTM) obtained from the difference between the DEM and CHM models. Right: Same DTM after having performed the low-pass filtering process.

The DTM is also necessary for the simulation of heat and water fluxes at the soil-plant-atmosphere interface since it takes into account terrain variations as a boundary condition in the simulation of airflow over the terrain for the calculation of evapotranspiration. In turn, the CHM is used in Computational Fluid Dynamics (CFD) codes, such as the Advanced Regional Prediction System (ARPS), to determine the average tree height in an area with natural vegetation, in order to estimate the roughness parameter (z_0) of the wind profile.

2.4.2. Calculation and Selection of Topographic Indices

The following topographic variables have been selected to be analyzed as a source of secondary data in the BDF method for estimating soil moisture:

- The digital terrain model (DTM): In this case, the hypothesis is raised that pixels with a lower elevation value are more likely to receive surface drainage water, so they would have greater soil moisture compared to the higher ones.
- The Topographic Position Index (TPI) proposed by Weiss (2001): Computes the elevation of a cell relative to the average of neighboring cells within a specified internal and external radius. $TPI_{(5,25)}$ the internal radius from the cell center is 5 meters and the external radius of the ring is 25 meters. This indicates that the computation area for the average of the elevations of the neighboring cells has been taken as 2 pixels. In this way, if the value is negative it would indicate that the pixel is at a lower elevation than the surrounding area (valley) while a positive value indicates that the pixel is at a higher elevation than the surrounding area (peak). This computation was performed with the “Topographic Position Index” module of SAGATM.
- The Topographic Wetness Index (TWI) proposed by Beven and Kirkby (1979): Describes the tendency of a cell to accumulate surface runoff water from other cells. This index is computed as: $TWI = \ln\left(\frac{SCA}{\tan\theta}\right)$, where SCA is the Specific water Catchment Area for each pixel and θ is the slope of the pixel. The basic concept of TWI is a mass balance: SCA is a parameter of the tendency to receive water, while the local slope together with the drainage (implicit in SCA) describes the tendency to evacuate the received water. Both the SCA and θ are computed from the DTM. There is an algorithm in SAGATM called “TWI one step” that calculates the TWI in a single step as shown by Mattivi et al. (2019).
- The Euclidean distance to water drainage channels (d_{stream}): In this case, the hypothesis is that the cells (or pixels) that are further away from the excess surface water drainage courses would have lower soil moisture than those located very close to them. The location of the segments that represent the channels for evacuating excess surface water must first be computed. These are the pixels that receive the greatest amount of water from their neighbours. Once these channels (stream) have been identified, the Euclidean distance from each pixel in the study area to the corresponding drainage channel is calculated. The drainage channel (or segment) layer can be obtained from the “r.watershed” module included in the GRASSTM plugin. The Euclidean distance layer to these segments is obtained from the “proximity” module of GDAL 3.8.3.
- The slope-penalized distance to water drainage channels (d_{DTM}) proposed by Fasbender et al. (2008) and modified in this work: d_{stream} indicates the Euclidean distance from each sector of the study area to the drainage channels; however, it can be thought that the excess water

would drain superficially faster towards these channels when their slope is greater than in other sites with a lower slope. For example, on the hilltop, a relationship between the terrain elevation and the soil moisture is not justified even when the Euclidean distance from the top to the drainage course is small. On the contrary, in areas with wide valleys, the soil moisture will be close to the ground level, even if the Euclidean distance to the drainage channels is large. Therefore, this indicator hypothesizes that the vertical percolations of water in areas with a high slope are lower and, therefore, result in lower soil moisture. In this way, the distance to the drainage channels can be penalized considering the slope as proposed by Fasbender et al. (2008) and applied by Peeters et al. (2010). However, the authors do not specify the penalty algorithm they use. In this paper, the use of the LS factor is proposed, which takes into account the effect of slope on soil erosion (Wishmeier & Smith, 1978). Moore et al. (1991) present a simplified form: $LS=(n+1) \cdot \left(\frac{A_s}{22,13}\right)^n \left(\frac{\tan\theta}{0,0896}\right)^m$, where $n = 0.4$ and $m = 1.3$. $A_s = 100 \text{ m}^2$ is the pixel area. Penalizing the distance to the drainage courses by this factor gives: $d_{DTM} = d_{stream} (1 + LS)$, then when the slope is zero $d_{DTM} = d_{stream}$.

These topographic indices were obtained using the SAGATM module in the QGIS software. From the analysis of the behavior of these topographic indicators it is observed that, except for the TWI, all the others are inversely proportional to the soil moisture.

2.5. Correlation Analysis of Topographic Indices and Soil Moisture Data Sampling

The study area presents two different soil types: La Juanita and Tezanos Pintos. Figure 7 shows a map of the area for each of them. Because of this, the difference between soil moisture at saturation and measured soil moisture was chosen as the target variable to be analyzed: $HS_{sat} = \eta_{sat} - \eta$. Therefore, this difference is directly proportional to the values taken by the topographic indices except for the TWI. For this index, it is proposed the inverse of TWI, which is named TWI^{-1} . The values of soil moisture at saturation η_{sat} , field capacity η_{fc} and wilting point η_w of the soil layers “Tezanos Pintos” and “La Juanita” are shown in Table 1:

Table 1. Agro-hydrological properties of the Tezanos Pintos and La Juanita soil series.

Layer depth (m)	Tezanos Pintos			La Juanita		
	η_{sat}	η_{fc}	η_w	η_{sat}	η_{fc}	η_w
AP: 0.0 – 0.2	0.4792	0.3959	0.2064	0.6226	0.3180	0.1700
Bt1: 0.2 – 0.4	0.4720	0.4121	0.2321	0.5472	0.4680	0.2700
Bt2: 0.4 – 0.6	0.4870	0.4138	0.2367	0.5472	0.3996	0.2460
BC1: 0.6 – 0.8	0.5104	0.4248	0.3032	0.6226	0.3000	0.1970
BC2: 0.8 – 1.0	0.5162	0.4312	0.2412	0.5472	0.3516	0.2220
> 1.0	0.5162	0.4312	0.2412	0.5472	0.3516	0.2220

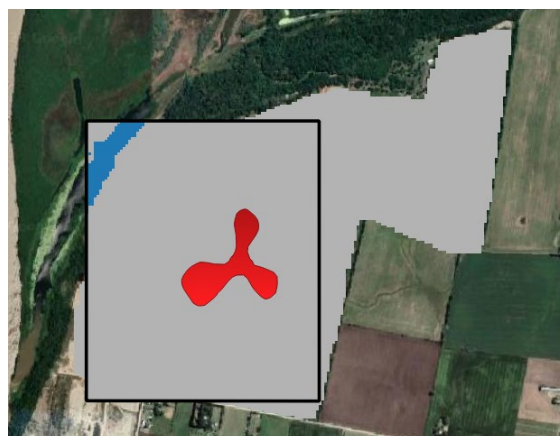


Figure 7. Soil types in the study area. ■: “La Juanita”, ■: “Tezanos Pintos”, ■: Water.

For the sampling dates, the HS_{sat} calculation was performed, which was correlated with the topographic indices to obtain the functions of the secondary variables $g_{sat(y)}$ that allow HS_{sat} to be estimated using topographic indices:

- $g_{sat(DTM)}$: HS_{sat} estimation as a function of DTM.

- $\mathcal{G}_{sat(TPI)}$: HSsat estimation as a function of TPI.
- $\mathcal{G}_{sat(TWI^{-1})}$: HSsat estimation as a function of TWI^{-1} .
- $\mathcal{G}_{sat(d_{stream})}$: HSsat estimation as a function of d_{stream} .
- $\mathcal{G}_{sat(d_{DTM})}$: HSsat estimation as a function of d_{DTM} .

These secondary functions were fitted using soil moisture measurements at the 9 sampling sites. Twenty-seven tuning functions were tested: lineal, Square root-Y, Exponential, Inverse-Y, Y-Square, Square root-X, Double square root, Log-Y Square root-X, Inverse-Y Square root-X, Square-Y Square root-X, Log-X, Square root-Y Log-X, Multiplicative, Inverse-Y Log-X, Y-Square Log-X, Inverse-X, Square root-Y Inverse-X, S-curve, Double Inverse, Y-Square Inverse-X, X-Square, Square root-Y X-Square, Log-Y X-Square, Inverse-Y X-Square, Double Square, Logistic and Log-probit. A correlation analysis has been performed to select the topographic indices that best explain the spatial variability of soil moisture. Equations (2) and (3) can then be applied to obtain maps of the mean value and variance of soil moisture.

2.6. Validation of Interpolation Methods

In order to evaluate different combinations of the BDF method according to the topographic indices with the best correlation of soil moisture as well as to validate the OK results, the relative soil moisture errors have been calculated using the cross-validation method and by an independent set of data sampled in the surface layer $\eta_{s(ad)}$ that does not participate in the interpolations.

2.6.1. Cross-Validation (CV)

The process consisted of estimating soil moisture and its variance using the OK and the secondary functions and their associated variances of the topographic index at observation position x_1 without taking into account the moisture data η_1 . Then, the interpolated soil moisture value $\eta_{1,-1}$ and the measured value η_1 are used to calculate the interpolation error $\hat{e}_{1(CV)} = \eta_1 - \eta_{1,-1}$. The first observation is then reintroduced into the dataset and the second observation is removed. Using the remaining data (including the first observation) and the specified algorithm, a value is interpolated $\eta_{2,-2}$ at the second observation position x_2 . Using the known observation value at this location η_2 , the interpolation error is calculated as before $\hat{e}_{2(CV)} = \eta_2 - \eta_{2,-2}$. The second observation is re-entered into the data set and the process continues in the same way for the third, fourth, fifth observation, etc., up to and including observation N . Then, $N = 9$ ($s = 1, N$) is obtained using this method for each sampling date f and each of the five layers of the soil profile: $\hat{e}_{s(CV)} = \eta_s - \eta_{s,-s}$.

2.6.2. Independent Data Set (Vad)

These data were obtained at $N = 5$ sites during the second half of 2022. From the interpolated grid, the soil moisture value at each position x_0 is obtained, coinciding with the additional measurement of the extra data set that did not participate in the interpolation process $\eta_{0(ad)}$. Since the soil moisture values $\eta_{s(ad)}$ at these sites are known, the interpolation error is calculated for each sample s : $\hat{e}_{s(ad)} = \eta_{s(ad)} - \eta_{0(ad)}$.

2.6.3. Standard Interpolation Error

Subsequently, with the calculations of the interpolation errors $\hat{e}_{s(m)}$, where (m) is the method used: (m) = (CV) or (Vad), the standard root means square error for each date separately ($RMSE_d$) was performed for OK and BDF. In addition, the relative error ($RRMSE_d$) to the spatial mean value of measured data $\bar{\eta}_{s(m)_d}$ equation (4) has been obtained. Equations (5), (6), and (7) are used to calculate the errors of the soil moisture interpolations.

$$\bar{\eta}_{s(m)_d} = \begin{cases} \frac{1}{9} \sum_{s=1}^9 \eta_{s(CV)d} (CV) \\ \frac{1}{5} \sum_{s=1}^5 \eta_{s(ad)d} (Vad) \end{cases} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{s=1}^N (\hat{e}_{s(m)})^2} \quad (5)$$

$$RMSE_d = \sqrt{\frac{1}{N} \sum_{s=1}^N (\hat{\epsilon}_{s(m)d})^2} \quad (6)$$

$$RRMSE_d (\%) = 100 \frac{RMSE_d}{\eta_{s(m)_d}} \quad (7)$$

3. Results and Discussion

3.1. Topographic Indices

Figure 8 shows the TPI using a saturation grey-scale palette $\overline{TPI} \pm 2 S_{TPI} = -0,001258 \pm 2 (0,476)$ to appreciate the variability in the cultivated area. Figure 9a represents the TWI using the same criterion for representing gray tones $\overline{TWI} \pm 2 S_{TWI} = 7,5656 \pm 2 (1,4872)$. It is notable that the highest TWI values coincide with the drainage channels (red lines), indicating that these pixels have a positive water balance because they receive a greater amount of water by surface runoff relative to the rate of water evacuation. This can be observed in Figure 9b. Figure 10a presents the d_stream in grey-scale and the drainage channels of the study area in red. Figure 10b shows the \overline{d}_{DTM} with the drainage channels. The effect of the slope penalty considered in the LS factor on the d_stream is shown.

3.2. Statistical Analysis of Topographic Indices and Obtaining Secondary Functions

A correlation analysis was performed between the sampled *HSsat* soil moisture saturation deficit data and the five topographic indices.

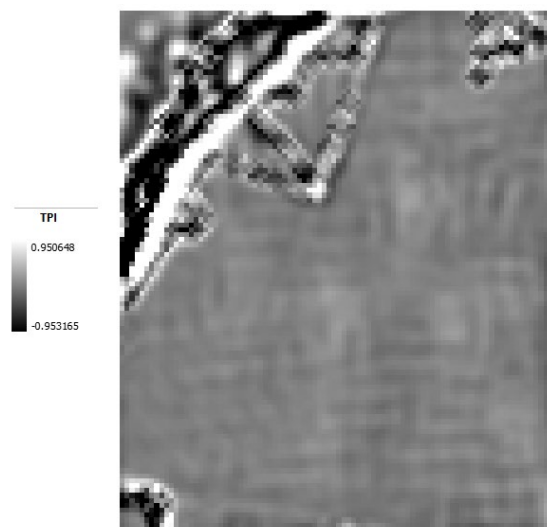


Figure 8. Topographic Position Index (TPI).

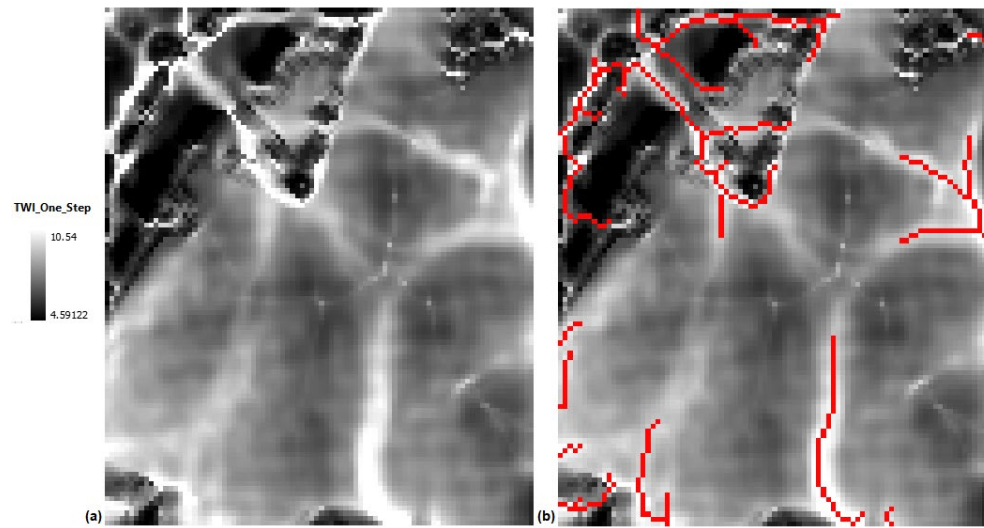


Figure 9. (a) Topographic moisture index (TWI) and (b) overlay of TWI with excess water surface drainage channels.

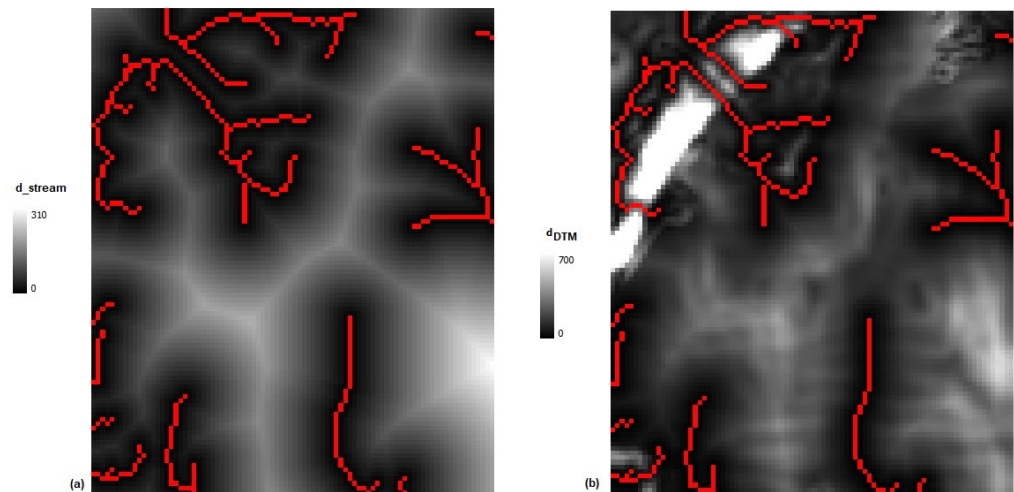


Figure 10. (a) Euclidean distance to excess water drainage channels (d_{stream}) and (b) distance penalized by the LS factor that takes into account the d_{DTM} slope.

From the analysis of the correlation coefficients, it was observed that the topographic indices that best explain the variability of the HS_{sat} for all sampling dates are the DTM and the TWI^{-1} . These two indices were used as secondary variables for the fusion with the interpolation of spatial data using the OK method. For both the DTM and the TWI^{-1} indices, the most frequent HS_{sat} fitting function on the observation dates is Double-Square. However, it was observed that the intercept coefficient of the Double-Square function is generally negative for DTM, which implies that the square root of the function has no solution when the value of DTM is low. For this, X-Square function was chosen for this topographic index after verifying that the values of the correlation coefficient are slightly lower than those obtained with the Double-Square function. Table 2 shows the correlation coefficients. The secondary variable functions for both indices are:

$$g_{sat(DTM)} = a + b \cdot DTM^2 \quad (8)$$

$$g_{sat(TWI^{-1})} = \sqrt{a + b \cdot (TWI^{-1})^2} \quad (9)$$

It can be observed in Table 2 that to homogenize the calculation method of the secondary functions, it was decided to use X-Square and Double-Square functions fitting for DTM and TWI^{-1} topographic indices respectively on all sampling dates.

Table 2. Correlation coefficients of X-Square function for the DTM index and Double-Square function for the TWI^{-1} index.

Sampling dates	Topographic Index			
	DTM		TWI^{-1}	
	$g_{sat}(DTM)$	ρ	$g_{sat}(TWI^{-1})$	ρ
2021/02/25	X-Square	0.6325	Double-Square	0.6931
2021/03/12	X-Square	0.6382	Double-Square	0.6604
2021/03/19	X-Square	0.7094	Double-Square	0.7069
2021/04/20	X-Square	0.7007	Double-Square	0.7009
2021/04/30	X-Square	0.7297	Double-Square	0.6849
2021/05/07	X-Square	0.7200	Double-Square	0.7345
2021/05/14	X-Square	0.7392	Double-Square	0.7576
2021/05/31	X-Square	0.7172	Double-Square	0.7007
2022/08/19	X-Square	0.7367	Double-Square	0.7440
2022/09/09	X-Square	0.6532	Double-Square	0.6282
2022/10/02	X-Square	0.7663	Double-Square	0.8041
2022/10/17	X-Square	0.7959	Double-Square	0.8438
2022/10/28	X-Square	0.8257	Double-Square	0.8085
2022/11/18	X-Square	0.5952	Double-Square	0.5476
2022/12/02	X-Square	0.7411	Double-Square	0.7192
AVERAGE		0.7134		0.7156

The BDF method by using OK spatial interpolation and two secondary variables, DTM and TWI^{-1} , equation (1) yield:

$$f(\eta_0 / y_0) \propto \frac{1}{|f(\eta_0)|^2} \left[f(\eta_0 / y_s) f(\eta_0 / g_{(DTM)}) f(\eta_0 / g_{(TWI^{-1})}) \right] \tag{10}$$

where $f(\eta_0 / y_s)$ is the Probability Density Function (PDF) of the soil moisture obtained at the un-sampled site x_0 based on observations x_s through the OK spatial interpolation method and the other PDFs in the second member are the probability density functions of soil moisture for the same site obtained from the secondary functions $g_{(DTM)}$ and $g_{(TWI^{-1})}$. From equation (10), the mean value and variance of the BDF are obtained with equation (2) and equation (3) assuming Gaussian distribution. It is observed that not only the mean values of the secondary variables $\mu_{y_0,j}$, ($g_{(y)}$ functions) are involved but also their variances ($\sigma_{y_0,j}^2$). These can be calculated by considering that the secondary functions respond to a normal probability distribution. From the statistical analysis of the sampled data, it is possible to use normal distribution models to calculate the variances as a function of the secondary variable. This calculation has been carried out for all sampling dates by finding the piecewise fit functions of the standard deviations using parabolic equations for DTM and TWI^{-1} variables. Figure 11 shows the $g_{(y)}$ functions, standard deviation, and 95% confidence for DTM and TWI^{-1} (respectively calculated from soil profile samples) during two 2021 sampling dates (February 25 [summer] and May 31 [winter]) and two 2022 sampling dates (August 19 [summer] and December 2 [winter]), illustrating the analytical approach. Figure 12 shows the standard deviation σ_{DTM} (left) and $\sigma_{TWI^{-1}}$ (right).

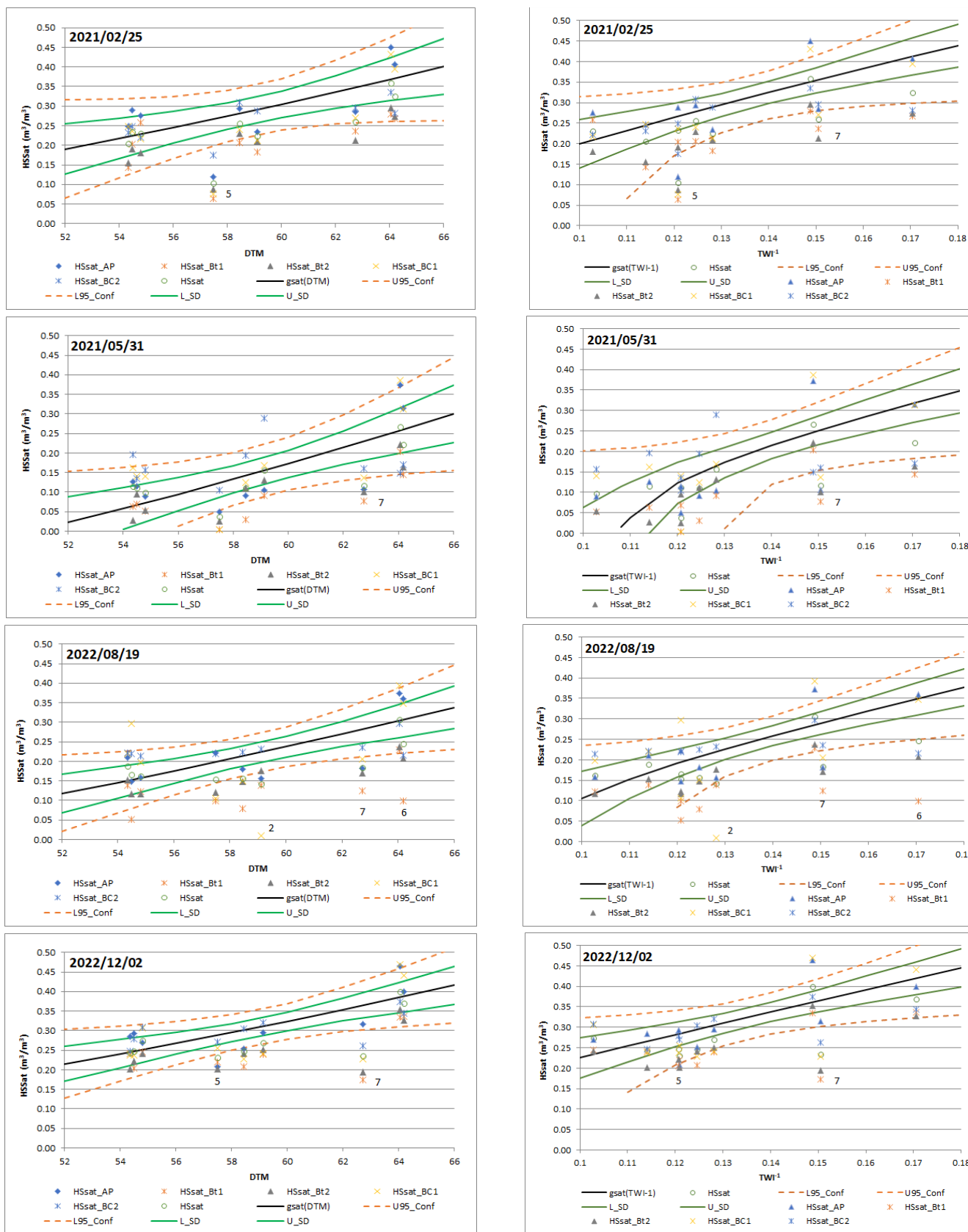


Figure 11. Curves of $g_{sat}(DTM)$ (left) and $g_{sat}(TWI^{-1})$ (right) functions, Standard Deviation Lower and Upper limits (L_SD; U_SD), 95% Confidence Lower and Upper limits (L95_Conf; U95_Conf), and $HSsat$ values (symbols) were obtained from soil moisture measurements on 25 February 2021, 31 May 2021, 19 August 2022, and 2 December 2022.

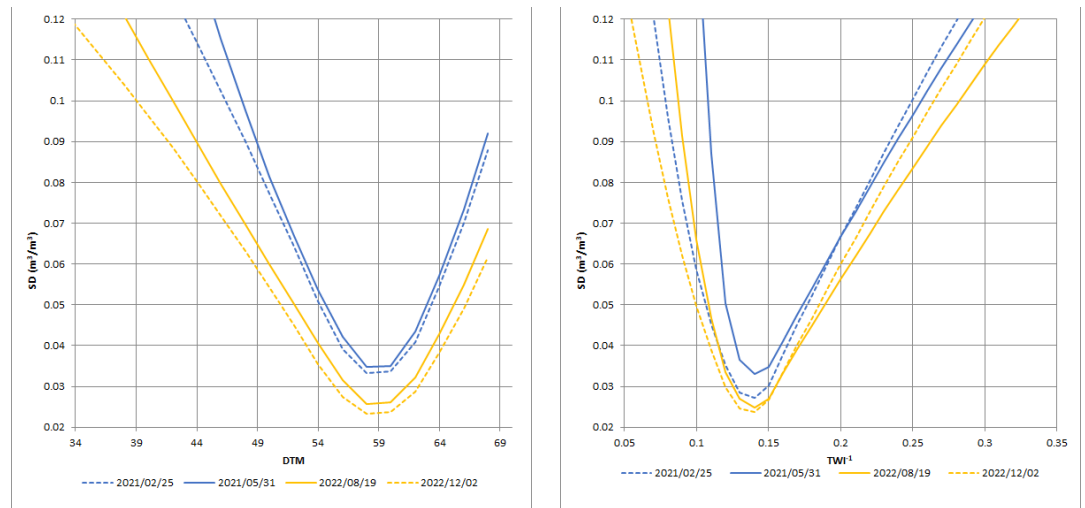


Figure 12. Fitted Standard Deviation curves of σ_{DTM} (left) and $\sigma_{TWI^{-1}}$ (right) using piecewise parabolic equations for summer days (dashed lines) and winter days (solid lines) in the 2021 and 2022 years.

It can be observed in Figure 11 for the 2021 sampling dates that there is variability that is not explained by the secondary variable DTM at sampling site 5, where all soil moisture values obtained from samples in February and a large part of them in May, are close to saturation beyond the 95% confidence limit. This behavior can also be seen on site 7 from mid-April 2021 and in the second half of 2022. This may be due to different land covers between the sampled sites. At sites 7 and 5 there were corn and soybean crops that were planted in November 2020 respectively with dense leaf cover at the time of soil moisture sampling and that completely covered the soil, unlike the other sites where soybeans were planted in February 2021, so the plant cover did not completely cover the soil because the crop was in its early vegetative stages. This year, it is also observed that during March and April, the precipitation was greater than the reference evapotranspiration (computed by Penman-Monteith method modified by Allen et al., 1998; Figure 13, left panel). Consequently, the soil moisture curves are closer to saturation. During the second half of 2022, there was a marked precipitation deficit (Figure 13, right panel), so the curves in this semester indicate a drier soil profile (see Figure 11, December 2022). For both the topographic variables DTM and TWI^{-1} , there are deviations towards wetter soil situations in sites 2, 6, and 7 in August 2022. However, a difference is observed when comparing Figure 11 in this semester. The TWI^{-1} index does not show a significant deviation at site 5 on December 2022, while the DTM has greater deviations towards low values outside the 95% confidence level limit, which may indicate that TWI^{-1} tends to depict better than the DTM part of the differences in coverage between sampling sites.

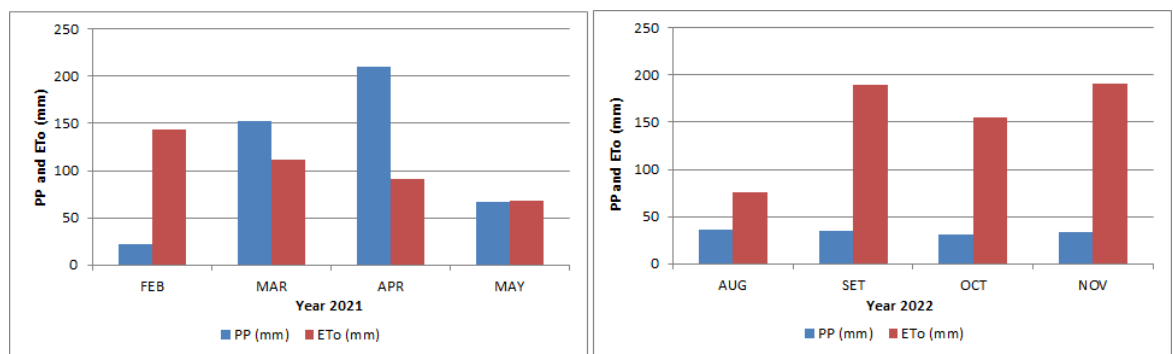


Figure 13. Relationship between Precipitation (PP) and Reference Evapotranspiration (ETo) for the first months of 2021 and the last months of 2022 in the study area.

3.3. Selected Secondary Functions Maps

The mean and standard deviation models of the secondary functions $g_{sat(y)}$ were used to obtain the soil moisture maps predicted by the variables $g_{(y)} = \eta_{sat} - g_{sat(y)}$. Maps of mean soil moisture for the AP layer depicted by the secondary functions $g_{(DTM)}$ and $g_{(TWI^{-1})}$ have been obtained for the 15 sampling dates. As an example, those corresponding to February 25, 2021; May 31, 2021; August 19, 2022; and December 2, 2022, are presented in Figure 14. Figure 15 presents

the standard deviation maps for the AP surface layer at the same sampling dates. It is worth noting that the standard deviations of the variables $g_{sat(y)}$ are the same as those of the variables $g(y)$ since both differ by a constant η_{sat} . This is due to the property of the linear combination of the variances: $\sigma_{(a+bx)}^2 = \frac{1}{b} \sigma_{(x)}^2$ where in this case: $a = \eta_{sat}$ and $b = 1$.

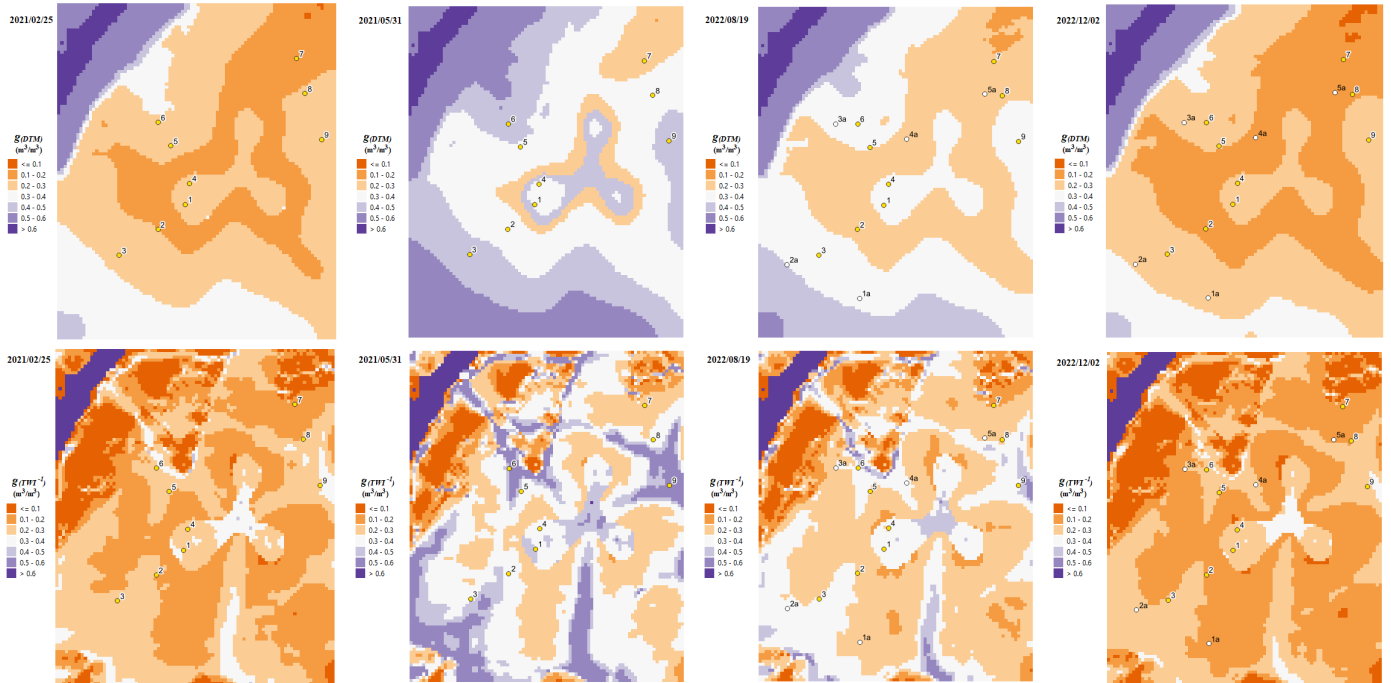


Figure 14. Maps of mean soil moisture for the AP layer depicted by the secondary functions $g_{(MDT)}$ (up) and $g_{(TWI^{-1})}$ (down). Sampling dates: 2021/02/25; 2021/05/31; 2022/08/19; 2022/12/02. The points indicate the soil moisture sampling sites.

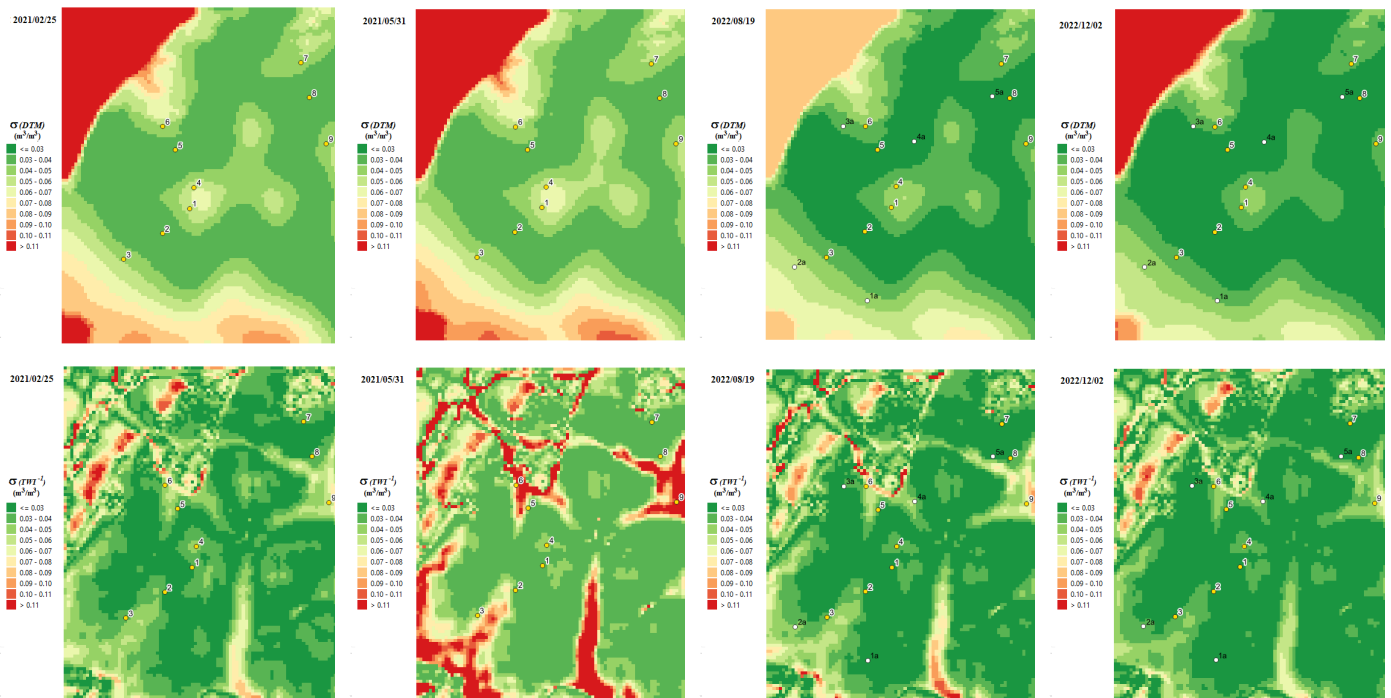


Figure 15. Maps of Standard Deviation soil moisture for the AP layer depicted by the secondary functions $g_{(MDT)}$ (left) and $g_{(TWI^{-1})}$ (right). Sampling dates: 2021/02/25, 2021/05/31, 2022/08/19, 2022/12/02. The points indicate the soil moisture sampling sites.

The temporal evolution of the mean soil moisture value maps (depicted by both secondary variables in Figure 14, $\mu_y: g_{(MDT)}$ and $g_{(TWI^{-1})}$) shows that in the summer months (February and March 2021,) the values are low, possibly due to the influence of solar radiation and temperature in the AP soil surface layer. In April and May, a moisture recharge is shown in this profile, possibly due to the lower demand for water by evapotranspiration. In 2022, a progressive drying of the layer is observed as the year progresses. It is seen that the soil moisture explained by TWI^{-1} follows the trace of the excess water surface drainage channels in greater detail than the DTM. Both secondary variables show the difference in moisture by soil type. The differences in spatial variability of soil moisture between the years 2021 and 2022 are more evident in Figure 15 than in Figure 14. As shown previously, water stress situations in 2022 cause the spatial variability in soil moisture to be much lower than in 2021. Comparing the spatial variability of soil moisture for samples taken in the first half of 2021 with those of the second half of 2022, it is observed that these differences are larger for the secondary variable TWI^{-1} .

3.4. Spatial Interpolation of Soil Moisture Measurements Using Ordinary Kriging (OK)

Spatial interpolations using the OK method for all sampling dates were performed using Surfer 9.0 software. This software also allows obtaining standard deviation maps using the experimental variograms. Linear models of variograms were chosen because the number of soil moisture samples in the sub-basins is insufficient for the plotting of local directional variograms based on the terrain slopes. As an example, Figure 16 and Figure 17 show the mean value (left) and standard deviation (right) maps of the soil moisture interpolations using the OK method (μ_k, σ_k) for same dates.

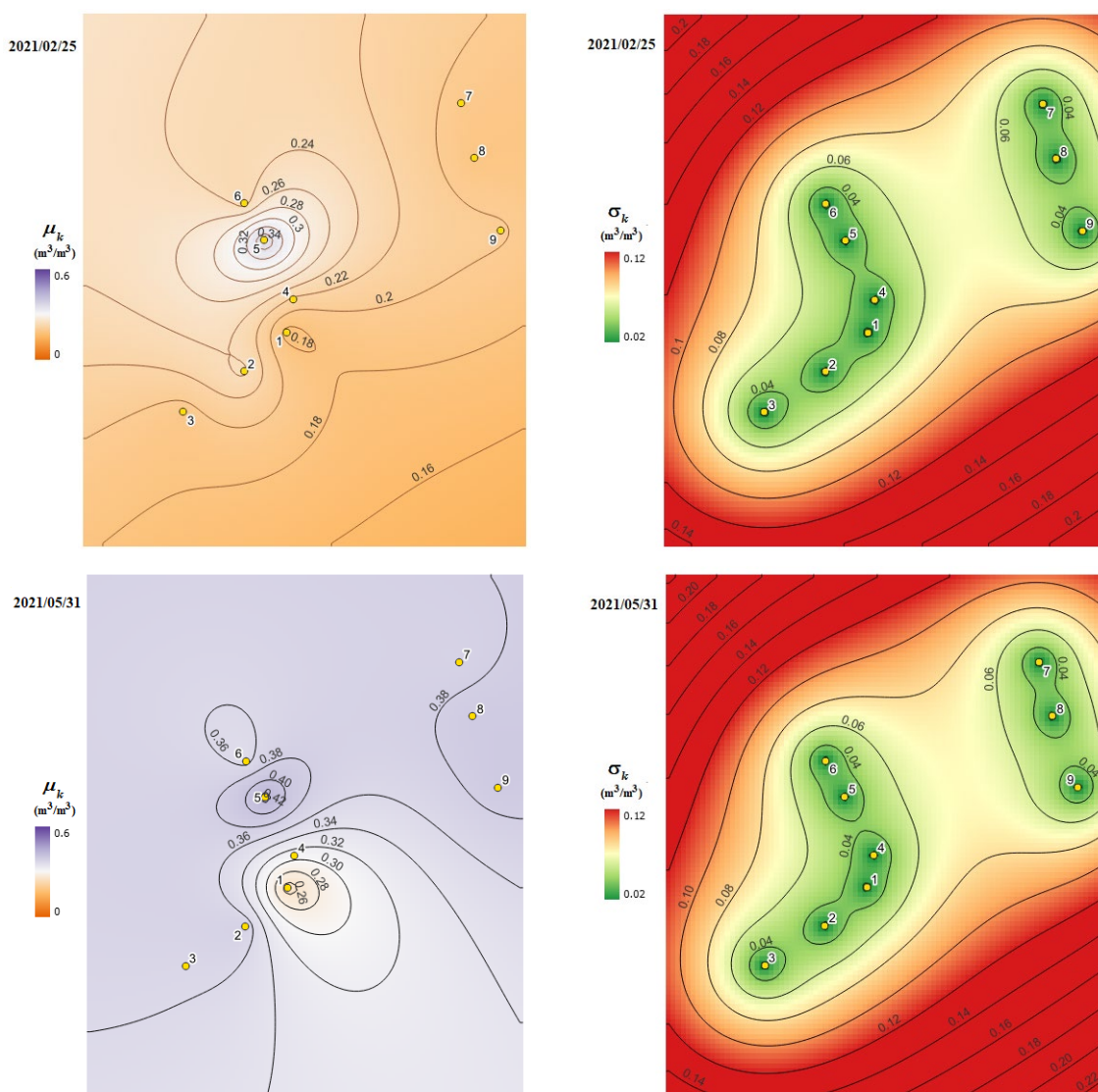


Figure 16. Soil moisture values in AP interpolated using OK: μ_k (left) and σ_k (right). Sampling dates: 2021/02/25 and 2021/05/31. The points indicate the soil moisture sampling sites.

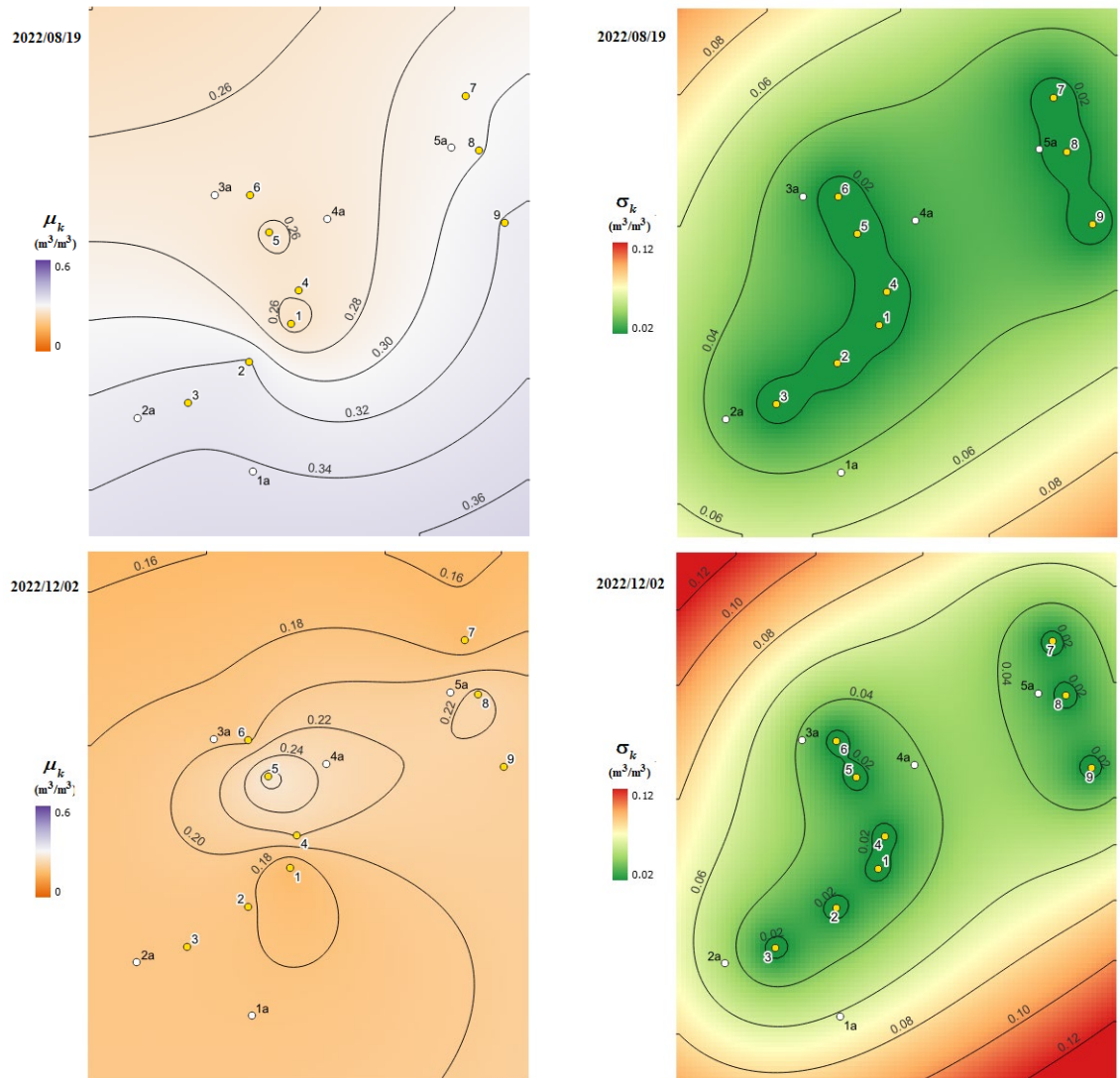


Figure 17. Soil moisture values in AP interpolated using OK: μ_k (left) and σ_k (right). Sampling dates: 2022/08/19 and 2022/12/02. The points indicate the soil moisture sampling sites.

Analysis of the interpolation maps using the OK method (see Figures 16 and 17), reveals that the low sampling site density presents the typical configurations called “bull’s eyes” that correspond to circles around very low or very high soil moisture values. These configurations are unrealistic, which justifies the application of the BDF method to improve the OK results. The problem of low sample density can also be observed in the standard deviation maps since the dispersion around the interpolated values increases very rapidly as one moves away from the sampling sites. This is observed very markedly in 2021 (Figure 16, right panel) and less evident in 2022 (Figure 17, right panel), because in this year the spatial variability of soil moisture in the study area was lower due to the occurrence of low rainfall throughout the region (see Figure 13).

3.5. Validation of the Bayesian Data Fusion Method

Using equations (4), (5), (6), and (7) the Relative Root Mean Square Standard Error $RRMSE_d$ was calculated for the soil moisture data sampling dates using the Cross-validation method (CV) and data set of independent samples (Vad) that were not included in the OK spatial interpolation. In the last method, these data marked on the previous maps of the year 2022 as 1a to 5a were taken in the surface layer of the AP soil profile. This validation methodology was used to compare the results of the OK and BDF interpolations using equations (2) and (3) considering three data fusion alternatives:

- (1) OK and DTM;
- (2) OK and TWI^{-1} ;
- (3) OK and both secondary variables (DTM and TWI^{-1}).

Table 3 shows $RRMSE_d$ and the average over the sampling dates using the (CV) method. An improvement is observed in the interpolation errors of secondary topographic variables in the estimation of soil moisture using BDF. When the OK spatial interpolation method is compared with the BDF by using (CV) error estimation, it can be observed that in general the estimation error is greater in the summer months and is lower in the winter months. Considering the average of the soil moisture estimation errors in all the observations, it can be seen that the BDF method has a lower $RRMSE$ than the OK method, although the differences are very small. Comparing the different combinations in the BDF, it is observed that the (OK-DTM) has the lowest $RRMSE$. When analyzing the years 2021 and 2022 separately, it can be seen that in 2021 the $RRMSE$ is higher than in 2022, which affirms that the spatial variability is greater in 2021, possibly due to the fact that rainfall was greater than in the second half of 2022. In the comparison between the different interpolation methods, it is observed that for the first half of 2021, the FDB method with the combination (KO-DTM) is the one with the lowest $RRMSE$. In the second half of 2022, the one with the lowest $RRMSE$ is the FDB method with the combination (OK-DTM-TWI⁻¹).

Table 3. Relative root mean square errors $RRMSE_d$ calculated using the interpolation errors by Cross-validation (CV) method for the Ordinary Kriging (OK) and Bayesian Data Fusion (BDF).

Sampling dates (<i>d</i>)	$RRMSE_d$ (%)			
	OK	BDF		
		OK-DTM	OK-TWI ⁻¹	OK-DTM-TWI ⁻¹
2021/02/25	32.59689	31.97403	31.57704	31.21643
2021/03/12	16.63304	16.69337	16.79643	16.94355
2021/03/19	18.91018	18.38758	18.51472	18.28422
2021/04/20	14.92306	14.96444	15.62951	15.71715
2021/04/30	12.89757	12.83425	13.47868	13.42929
2021/05/07	16.68387	16.67366	17.62583	17.60853
2021/05/14	10.19849	10.43183	11.34280	11.68066
2021/05/31	15.78390	15.56965	15.97378	15.98117
2022/08/19	7.30308	7.43519	7.72497	7.89307
2022/09/09	16.99758	17.14755	17.35412	17.52723
2022/10/02	13.19149	12.97038	12.63427	12.56903
2022/10/17	13.72735	13.68217	13.28364	13.43625
2022/10/28	15.35997	14.64658	14.32769	14.10953
2022/11/18	18.35396	18.24952	18.19721	18.16677
2022/12/02	22.84751	22.01082	21.90845	21.43883
AV. 2021–2022	16.42720	16.24473	16.42461	16.40011
AV. 2021	17.32837	17.19110	17.61735	17.60763
AV. 2022	15.39728	15.16317	15.06148	15.02010

Table 4 shows the $RRMSE_d$ and their average of the sampling dates using the independent set of soil samples (Vad). It can be seen that the lowest average $RRMSE_d$ is observed when using the BDF method with the variables (OK-TWI⁻¹) combination, even though on certain dates, the BDF method using (OK-DTM) variables has the best result (three of the seven dates analyzed: August 19, November 18, and December 2).

Figure 18 (left) shows a scatter diagram of the errors calculated by CV for the estimation of soil moisture in the AP layer using the OK methods and the topographic variables DTM and TWI⁻¹. Figure 18 (right) presents the scatter diagram of the errors considering the OK method and the FDB method with the combinations (OK-DTM), (OK-TWI⁻¹), and (OK-DTM-TWI⁻¹). The measured data from all sampling dates are represented on the abscissas and the simulated data with spatial interpolation methods are represented on the ordinate. These figures show the lowest error dispersion for the BDF interpolation methods compared to the soil moisture estimation using only the topographic variables separately. The fusion of these with the OK is evidenced in these figures.

Table 4. Relative root mean square errors $RRMSE_d$ calculated using the interpolation errors by Validation with an additional data set (V_{ad}) for the Ordinary Kriging (OK) and Bayesian Data Fusion (BDF).

Sampling dates (d)	$RRMSE_d$ (%)			
	OK	BDF		
		OK-DTM	OK-TWI ⁻¹	OK-DTM-TWI ⁻¹
2022/08/19	13.68492	13.09942	15.70013	14.80832
2022/09/09	18.33629	17.09313	14.23506	14.81375
2022/10/02	12.01752	14.13030	6.41038	7.39070
2022/10/17	13.26676	20.86311	14.43126	18.16939
2022/10/28	21.29076	29.69339	17.26534	24.03038
2022/11/18	9.80441	7.85816	13.26226	12.26633
2022/12/02	14.68534	10.74759	15.96717	10.76994
AV. 2022	14.72657	16.21216	13.89594	14.60697

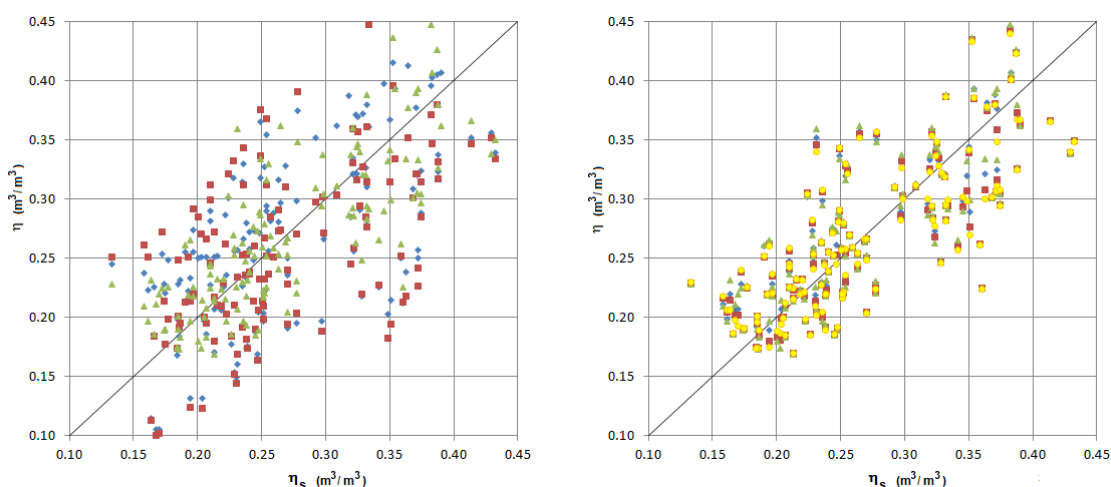


Figure 18. Scatter diagram of the errors calculated by CV method in the estimation of soil moisture. Left: \blacktriangle OK, \blacklozenge $g(DTM)$, \blacksquare $g(TWI^{-1})$. Right: \blacktriangle OK, \blacklozenge BDF(OK-DTM), \blacksquare BDF(OK-TWI⁻¹), \bullet BDF(OK-DTM-TWI⁻¹).

3.6. Soil Moisture Maps Obtained with Bayesian Data Fusion Combination (OK-DTM-TWI⁻¹)

The soil moisture maps of all layers and all sampling dates are obtained by using the BDF for (OK-DTM-TWI⁻¹) combinations. Figures 19–22 show the AP layer maps of all sampling dates.

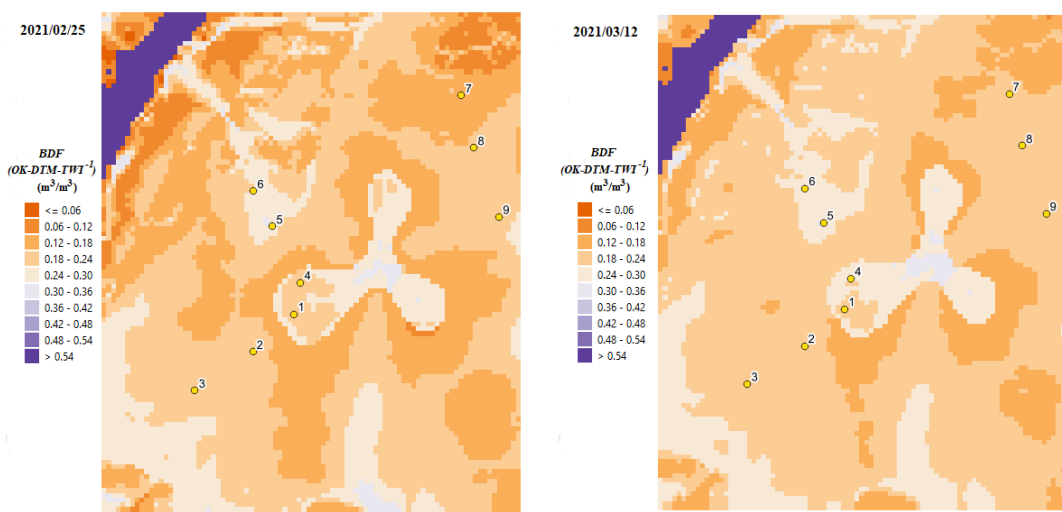


Figure 19. Maps of soil moisture on the AP layer estimated by Bayesian Data Fusion method BDF(OK-DTM-TWI⁻¹). Soil moisture sampling date: February 25 and March 12, 2021. The points indicate the soil moisture sampling sites.

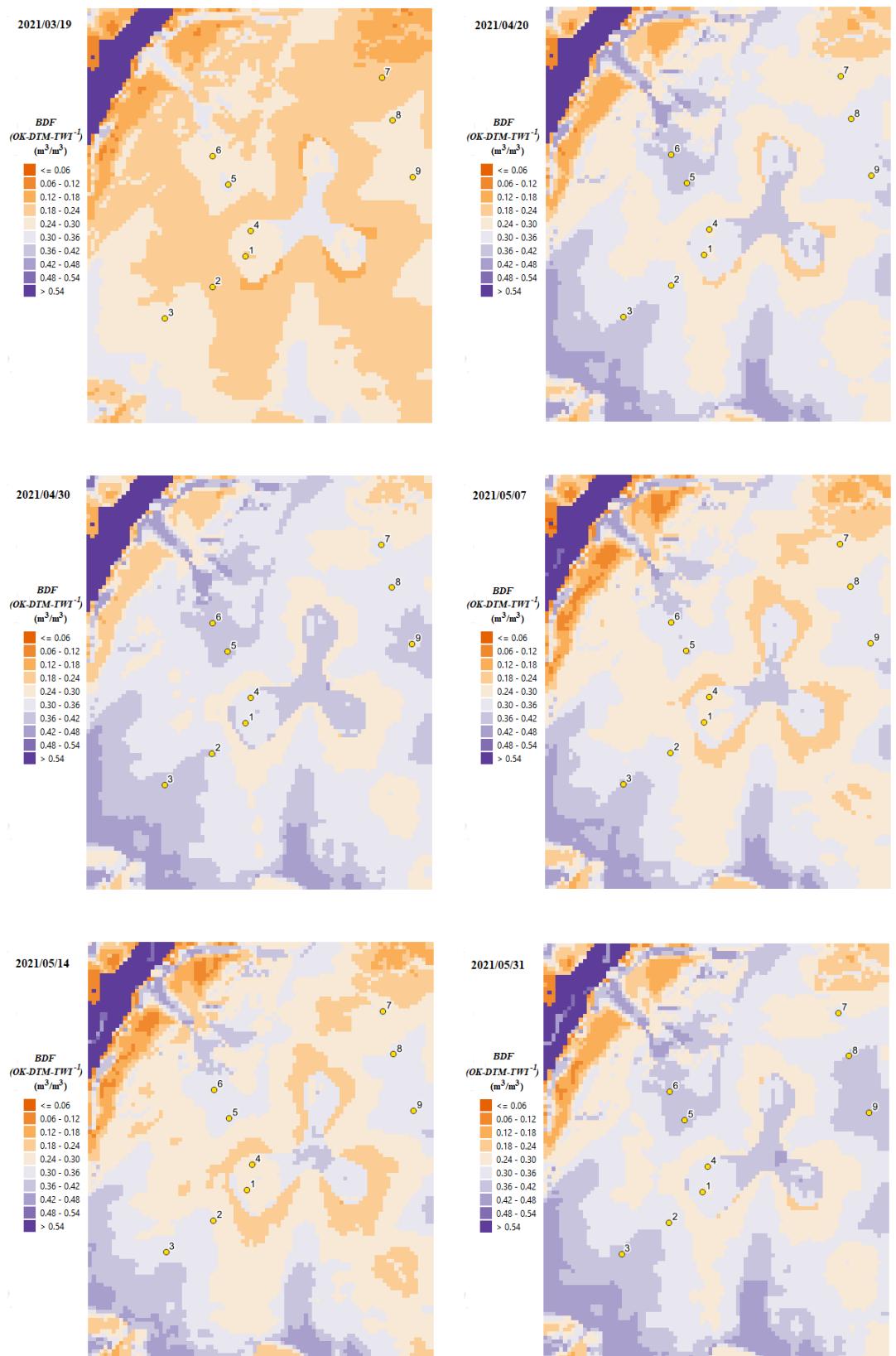


Figure 20. Maps of soil moisture on the AP layer estimated by Bayesian Data Fusion method BDF(OK-DTM-TWI⁻¹). Soil moisture sampling dates: March 19, April 30, May 5, 14, and 31, 2021. The points indicate the soil moisture sampling sites.

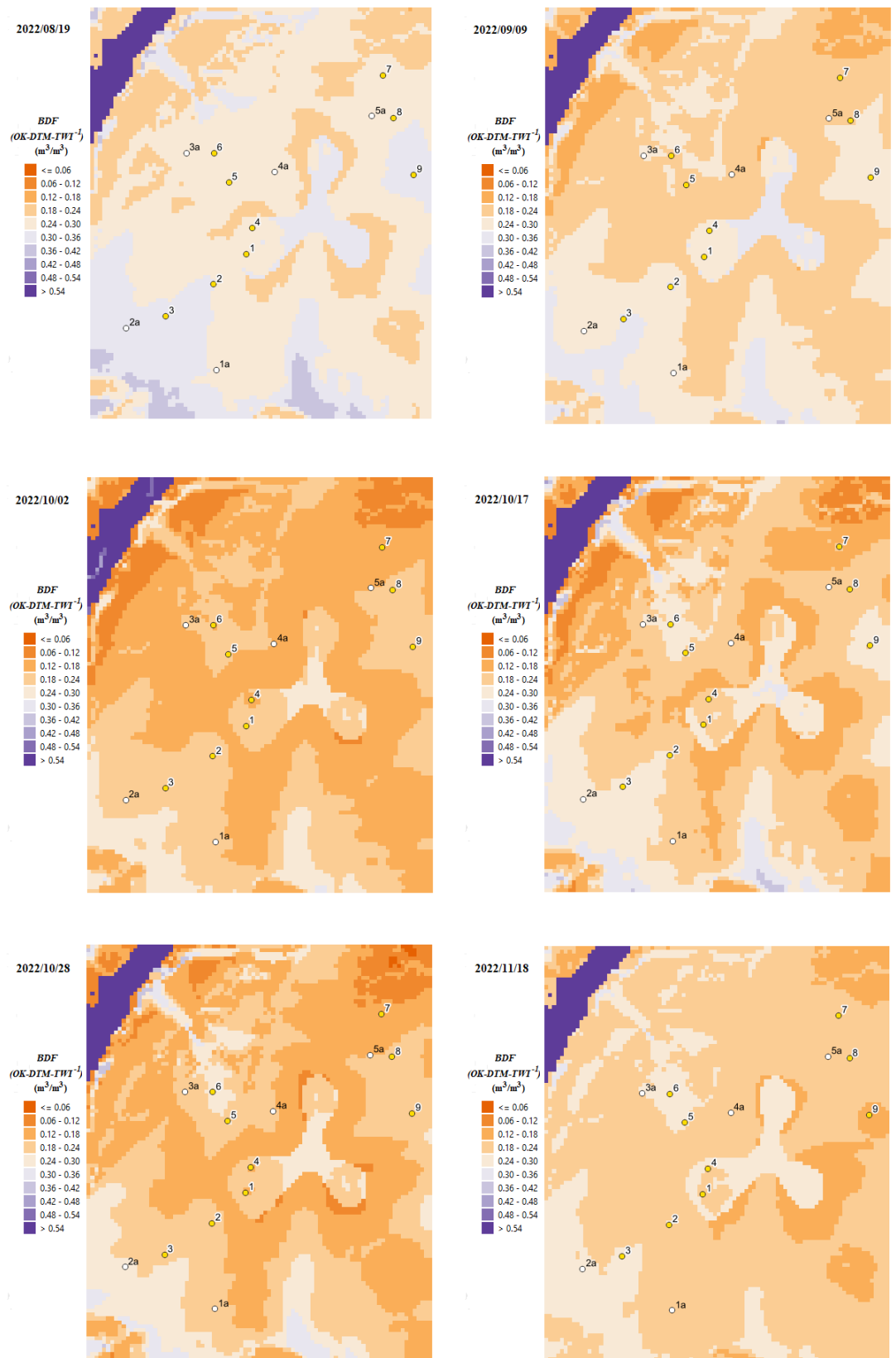


Figure 21. Maps of soil moisture on the AP layer estimated by Bayesian Data Fusion method BDF(OK-DTM-TWI⁻¹). Soil moisture sampling dates: August 19, September 9, October 2, 17, 28, and November 18, 2022. The points indicate the soil moisture sampling sites.

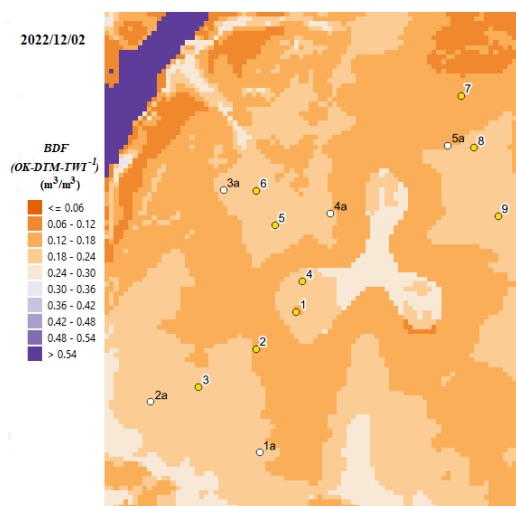


Figure 22. Maps of soil moisture on the AP layer estimated by Bayesian Data Fusion method BDF(OK-DTM-TWI⁻¹). Soil moisture sampling dates: December 2, 2022.

From Figures 16, 17, and 19–22, it can be seen that data fusion substantially improves the results of spatial interpolations of soil moisture when used instead of the OK method alone. By comparing the combinations of the BDF method from the estimation errors, it can be seen that in the months of the second half of 2022, the BDF(KO-TWI⁻¹) has lower relative errors than the others, but during the first half of 2021 the BDF(KO-DTM) combination has the lowest error. In any case, the differences are not very significant. Considering a classification in the performance of the estimation methods proposed by Despotovic et al. (2016), all three combinations are classified as “GOOD” with 10 % > *RRMSE* % < 20 % on average. In addition, the combination of both secondary variables could be considered redundant since the TWI⁻¹ index is obtained from the DTM with the addition of a water flow balance on the sub-basins.

4. Conclusions

A methodology that combines different data sources based on Bayesian Data Fusion (BDF) is applied to incorporate topographic information into the spatial interpolation of soil moisture data.

A statistical analysis of 5 topographic indices derived from a digital terrain model (DTM) was performed to evaluate the capacity to explain the spatial variability of soil moisture.

The topographic wetness index (TWI) and the DTM presented better results according to a statistical correlation analysis.

In this paper, it was observed that both topographic indices reflect well the spatial variability of soil moisture when combined with spatial interpolations using OK in the BDF framework.

A soil moisture data set independent of spatial interpolation was used to calculate the interpolation errors by the OK method and the alternative applications of the BDF: OK-DTM, OK-TWI⁻¹, and OK-DTM-TWI⁻¹.

The interpolation methodology presented and applied in this paper shows that using different data sources for soil moisture interpolation within the Bayesian data fusion framework, even with limited data, allows obtaining more realistic maps by incorporating topographic information.

The fusion of multiple sources of information provides significant advantages over data from a single source, as the estimation is improved relative to the application of a simple method of spatial interpolation of soil moisture data obtained from soil samples.

The practical innovation lies in the use of topographic indices as secondary variables to obtain improved soil moisture maps and can be used in a Bayesian data fusion framework as a valuable tool to reliably predict the most likely soil moisture. Its theoretical value lies in the application of these indices in the Bayesian data fusion framework, both individually and in combination.

The results obtained will serve as data for setting the initial conditions of multi-layer soil moisture simulation models.

5. Future Works

An extension of the research conducted in this paper will be the use of these topographic indices within the Bayesian fusion framework at other sites with different soil properties and topography characteristics.

Furthermore, the aspect and curvature of the relief as secondary variables will be the subject of research in future work.

The use of other interpolation methods for primary soil moisture data such as Co-Kriging, Universal Kriging, and Radial-Basis Function will be another subject of future works.

The results of this work will be used as input for soil temperature models and estimates of the thermal profile in subsurface layers for the initialization of multi-layer simulation models of soil moisture and temperature.

Supplementary Materials: The following supporting information can be downloaded: All Figures https://drive.google.com/file/all_Figures, Figures.rar.

The collection soil moisture data and analysis can be downloaded at: https://docs.google.com/spreadsheets/Atanor_Data_Soil_Moisture: HS_Atanor_Data_and_analysis.xlsx.

CRedit Author Statement: César Augusto Aguirre: Conceptualization, Methodology, Writing-original draft, Software, and Supervision; Carlos Germán Sedano: Visualization, Investigation, and Validation; Guillermo Antonio Rondán: Data collection, Data curation, and Investigation; Macarena Cardozo: Investigation and Software; Tatiana Rut: Investigation and Software.

Data Availability Statement: The Digital Elevation Model can be obtained from Instituto Geográfico Nacional at: <https://www.ign.gov.ar/NuestrasActividades/Geodesia/ModeloDigitalElevaciones/Mapa>

Funding: This research was funded by Fund for Scientific and Technological Research, Argentine (FONCyT), Argentina and National University of Entre Ríos (FCA-UNER), Argentine.

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

Acknowledgments: This study was financially supported by PICTO UNER-UADER 011 project of Fund for Scientific and Technological Research, Argentine and PID UNER 2242 project of National University of Entre Ríos, Argentine.

References

- Alizamir, M., Moradveisi, K., Othman Ahmed, K., Bahrami, J., Kim, S., & Heddami, S. (2025). An efficient data fusion model based on Bayesian model averaging for robust water quality prediction using deep learning strategies. *Expert Systems with Applications*, 261, 125499. <https://doi.org/10.1016/j.eswa.2024.125499>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. Food and Agriculture Organization of the United Nations. <https://www.fao.org/4/X0490E/x0490e00.htm>
- Álvarez-Mozos, J., Casali, J., & González-Audicana, M. (2005). Teledetección radar como herramienta para la estimación de la humedad superficial del suelo en cuencas agrícolas [Radar remote sensing as a tool for surface soil moisture estimation in agricultural watersheds]. *Revista de Teledetección*, (23), 27–42. <https://www.aet.org.es/revistas/revista23/AET23-03.pdf>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Bogaert, P., & Fasbender, D. (2007). Bayesian data fusion in a spatial prediction context: A general formulation. *Stochastic Environmental Research and Risk Assessment*, 21, 695–709. <https://doi.org/10.1007/s00477-006-0080-3>
- Bougeault, P., Noilhan, J., Lacarrère, P., Mascart, P. (1991a). An experiment with an advanced surface parameterization in a Mesobeta-Scale Model. Part I: Implementation. *Monthly Weather Review*, 119(10), 2358–2373. [https://doi.org/10.1175/1520-0493\(1991\)119<2358:AEWAAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<2358:AEWAAS>2.0.CO;2)
- Bougeault, P., Bret, B., Lacarrère, P., Noilhan, J. (1991b). An experiment with an advanced surface parameterization in a Mesobeta-Scale Model. Part II: The 16 June 1986 Simulation. *Monthly Weather Review*, 119(10), 2374–2392. [https://doi.org/10.1175/1520-0493\(1991\)119<2374:AEWAAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<2374:AEWAAS>2.0.CO;2)
- Chen, F., & Dudhia, J. (2001a). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129(4), 569–585. [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2)
- Chen, F., & Dudhia, J. (2001b). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part II: Preliminary model validation. *Monthly Weather Review*, 129(4), 587–604. [https://doi.org/10.1175/1520-0493\(2001\)129<0587:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0587:CAALSH>2.0.CO;2)
- Cressie, N. A. C. (1991). *Statistics for spatial data*. John Wiley and Sons, Inc. <https://doi.org/10.1002/9781119115151>
- Desbarats, A. J., Logan, C. E., Hinton, M. J., & Sharpe, D. R. (2002). On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, 225(1–4), 25–38. [https://doi.org/10.1016/S0022-1694\(01\)00504-2](https://doi.org/10.1016/S0022-1694(01)00504-2)
- Despotovic, M., Nedjic, V., Despotovic, D., & Cvetanovic, S. (2016). Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews*, 56, 246–260. <https://doi.org/10.1016/j.rser.2015.11.058>
- Fasbender, D., Obsomer, V., Bogaert, P., & Defourny, P. (2009). Updating scarce high resolution images with time series of coarser images: A Bayesian data fusion solution. In N. Milisavljević (Ed.), *Sensor and Data Fusion*. I-Tech Education and Publishing. <https://doi.org/10.5772/6581>
- Fasbender, D., Peeters, L., Bogaert, P., & Dassargues, A. (2008). Bayesian data fusion applied to water table spatial mapping. *Water Resource Research*, 44(12), W12422. <https://doi.org/10.1029/2008WR006921>
- Feng, F., Hu, X., Hu, L., Hu, F., Li, Y., & Zhang, L. (2009). Propagation mechanisms and diagnosis of parameter inconsistency within Li-Ion battery packs. *Renewable and Sustainable Energy Reviews*, 112, 102–113. <https://doi.org/10.1016/j.rser.2019.05.042>
- Hirt, C. (2016). Digital Terrain Models. In E. Grafarend (Ed.), *Encyclopedia of Geodesy* (pp. 1–6). Springer Cham.

- https://doi.org/10.1007/978-3-319-02370-0_31-1
- Instituto Geográfico Nacional. (2017). Modelo Digital de Elevaciones Aerofotogramétrico de Río Paraná (Sector 3.5) [Computer software]. <https://www.ign.gov.ar/NuestrasActividades/Geodesia/ModeloDigitalElevaciones/Mapa>
- Instituto Geográfico Nacional. (2019). Modelo Digital de Elevaciones de la República Argentina versión 2.0 [Computer software]. Dirección de Geodesia.
- Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to Applied Geostatistics*. Oxford University Press.
- Journel, A. G. (1989). *Fundamentals of geostatistics in five lessons*. American Geophysical Union. <https://doi.org/10.1029/SC008>
- Journel, A. G., & Huijbregts, C. (1978). *Mining geostatistics*. Academic Press.
- Karumuri, S., McClure, Z. D., Strachan, A., Titus, M., & Bilonis, I. (2023). Hierarchical Bayesian approach to experimental data fusion: Application to strength prediction of high entropy alloys from hardness measurements. *Computational Materials Science*, 217, 111851. <https://doi.org/10.1016/j.commatsci.2022.111851>
- Liao, K., Lai, X., Liu, Y., & Zhu, Q. (2016). Uncertainty analysis in near-surface soil moisture estimation on two typical land-use hillslopes. *Journal of Soils and Sediments*, 16, 2059–2071. <https://doi.org/10.1007/s11368-016-1405-6>
- Li, D.-Q., Wang, L., Cao, Z.-J., & Qi, X.-H. (2019). Reliability analysis of unsaturated slope stability considering SWCC model selection and parameter uncertainties. *Engineering Geology*, 260, 105207. <https://doi.org/10.1016/j.enggeo.2019.105207>
- Li, X., Xu, X., Liu, W., Xu, C., Zhang, R., & Wang, K. (2019). Prediction of profile soil moisture for one land use using measurements at a soil depth of other land uses in a karst depression. *Journal of Soils and Sediments*, 19, 1479–1489. <https://doi.org/10.1007/s11368-018-2138-5>
- Li, Y., & Vanapalli, S. K. (2021). A novel modeling method for the bimodal soil-water characteristic curve. *Computers and Geotechnics*, 138, 104318. <https://doi.org/10.1016/j.compgeo.2021.104318>
- Mattivi, P., Franci, F., Lambertini, A., & Bitelli, G. (2019). TWI computation: A comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, 4, 6. <https://doi.org/10.1186/s40965-019-0066-y>
- Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modeling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3–30. <https://doi.org/10.1002/hyp.3360050103>
- Noilhan, J., & Planton, S. (1989). A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117(3), 536–549. [https://doi.org/10.1175/1520-0493\(1989\)117<0536:ASPOLS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2)
- Peeters, L., Fasbender, D., Batelaan, O., & Dassargues, A. (2010). Bayesian data fusion for water table interpolation: Incorporating a hydrogeological conceptual model in kriging. *Water Resources Research*, 46(8), W08532. <https://doi.org/10.1029/2009WR008353>
- Pleim, J. E., & Guilliarm, R. (2009). An indirect data assimilation scheme for deep soil temperature in the Pleim–Xiu land surface model. *Journal of Applied Meteorology and Climatology*, 48(7), 1362–1376. <https://doi.org/10.1175/2009JAMC2053.1>
- Pleim, J. E., & Xiu, A. (2003). Development of a land surface model. Part II: Data assimilation. *Journal of Applied Meteorology*, 42(12), 1811–1822. [https://doi.org/10.1175/1520-0450\(2003\)042<1811:DOALSM>2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042<1811:DOALSM>2.0.CO;2)
- Sayago, S., Ovando, G., & Bocco, M. (2017). Landsat images and crop model for evaluating water stress of rainfed soybean. *Remote Sensing of Environment*, 198, 30–39. <https://doi.org/10.1016/j.rse.2017.05.008>
- Smirnova, T. G., Brown, J. M., & Benjamin, S. G. (1997). Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Monthly Weather Review*, 125(8), 1870–1884. [https://doi.org/10.1175/1520-0493\(1997\)125<1870:PODSMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2)
- Tocho, C. N., Antokoletz, E. D., & Piñón, D. A. (2020). Towards the realization of the International Height Reference Frame (IHRF) in Argentina. *International Association of Geodesy Symposia*, 152. https://doi.org/10.1007/1345_2020_93
- Tóffoli, M. B., Hernández, J. P., Kindeknecht, L. E., Befani, M. R., & Bressan, M. P. (2022). Caracterización edáfica y topográfica para la determinación de ambientes en función del agua disponible [Edaphic and topographic characterization for the determination of environments according to available water.]. *Revista Científica Agropecuaria*, 25(2), 30–40. <https://pcient.uner.edu.ar/index.php/rca/article/view/1535>
- Wang, L., Cao, Z.-J., Li, D.-Q., Phoon, K.-K., & Au, S.-K. (2018). Determination of site-specific soil-water characteristic curve from a limited number of test data – A Bayesian perspective. *Geoscience Frontiers*, 9(6), 1665–1677. <https://doi.org/10.1016/j.gsf.2017.10.014>
- Weiss, A. (2001, July). Topographic position and landform analysis [Poster presentation]. ESRI users Conference, San Diego, CA.
- Wilson, M. F. J., O’Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30(1–2), 3–35. <https://doi.org/10.1080/01490410701295962>
- Wishmeier, W. H. & Smith, D. D. (1978). *Predicting rainfall erosion Losses. A guide to conservation planning*. Science and Education Administration, United States Department of Agriculture. <https://www.govinfo.gov/content/pkg/GOVPUB-A-PURL-gpo31516/pdf/GOVPUB-A-PURL-gpo31516.pdf>
- Xiu, A., & Pleim, J. E. (2001). Development of a land surface model. Part I: Application in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, 40(2), 192–209. [https://doi.org/10.1175/1520-0450\(2001\)040<0192:DOALSM>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2)
- Zamani, M. G., Reza Nikoo, M., Niknazar, F., Al-Rawas, G., Al-Wardy, M., & Gandomi, A. H. (2023). A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy. *Journal of Cleaner Production*, 416, 137885. <https://doi.org/10.1016/j.jclepro.2023.137885>
- Zhang, Q., Zhang, P., & Li, T. (2025). Information fusion for large-scale multi-source data based on the Dempster-Shafer evidence theory. *Information Fusion*, 115, 102754. <https://doi.org/10.1016/j.inffus.2024.102754>
- Zhao, T., Yang, Y., Xu, L., & Sun, P. (2025). Bayesian identification of the optimal soil-water characteristic curve (SWCC) model and reliability analysis of unsaturated loess slope from extremely sparse measurements. *Engineering Geology*, 349, 107975. <https://doi.org/10.1016/j.enggeo.2025.107975>

Disclaimer: The views, statements, and data presented in *Agricultural & Rural Studies (A&R)* reflect solely the perspectives of the individual authors and contributors, and do not represent the official positions of SCC Press and/or the editorial team. SCC Press and/or the editorial team assume no liability for any harm, injury, or damage to persons or property arising from the ideas, methodologies, instructions, or products referenced herein.