How does PTF Interpret Soil Heterogeneity? A Stochastic Approach Applied to a Case Study on Maize in Northern Italy

Angelo Basile 1,2, Antonello Bonfante 1,2, Antonio Coppola 3, Roberto De Mascellis 1, Salvatore Falanga Bolognesi 4, Fabio Terribile 2,5 and Piero Manna 1,2,*

1 Institute for Mediterranean Agriculture and Forest Systems (ISAFOM)-Italian National Research Council (CNR), Via Patacca 85, 80056 Ercolano, Italy; angelo.basile@cnr.it (A.B.); antonello.bonfante@cnr.it (A.B.); roberto.demascellis@cnr.it (R.D.M.)
2 CRISP-Interdepartmental Research Center on the Earth Critical Zone, Via Università 100, 80055 Portici, Italy
3 School of Agricultural, Forestry, Food and Environmental Sciences (SAFE), University of Basilicata, 85100 Potenza, Italy; Antonio.coppola@unibas.it
4 ARIESPACE s.r.l., Centro Direzionale IS A3, 80143 Napoli, Italy; salvatore.falanga@ariespace.com
5 Department of Agriculture, University of Naples Federico II, Via Università 100, 80055 Portici, Italy; fabio.terribile@unina.it
* Correspondence: piero.manna@cnr.it

Received: 27 December 2018; Accepted: 29 January 2019; Published: 5 February 2019

Abstract: Soil water balance on a local scale is generally achieved by applying the classical nonlinear Richards equation that requires hydraulic properties, namely, water retention and hydraulic conductivity functions, to be known. Its application in agricultural systems on field or larger scales involves three major problems being solved, related to (i) the assessment of spatial variability of soil hydraulic properties, (ii) accounting for this spatial variability in modelling large-scale soil water flow, and (iii) measuring the effects of such variability on real field variables (e.g., soil water storage, biomass, etc.). To deal with the first issue, soil hydraulic characterization is frequently performed by using the so-called pedotransfer functions (PTFs), whose effectiveness in providing the actual information on spatial variability has been questioned. With regard to the second problem, the variability of hydraulic properties at the field scale has often been dealt with using a relatively simple approach of considering soils in the field as an ensemble of parallel and statistically independent tubes, assuming only vertical flow. This approach in dealing with spatial variability has been popular in the framework of a Monte Carlo technique. As for the last issue, remote sensing seems to be the only viable solution to verify the pattern of variability, going by several modelling outputs which have considered the soil spatial variability. Based on these premises, the goals of this work concerning the issues discussed above are the following: (1) analyzing the sensitivity of a Richards-based model to the measured variability of \( \theta(h) \) and \( k(\theta) \) parameters; (2) establishing the predictive capability of PTF in terms of a simple comparison with measured data; and (3) establishing the effectiveness of use of PTF by employing as data quality control an independent and spatially distributed estimation of the Above Ground Biomass (AGB). The study area of approximately 2000 hectares mainly devoted to maize forage cultivation is located in the Po plain (Lodi), in northern Italy. Sample sites throughout the study area were identified for hydropedological analysis (texture, bulk density, organic matter content, and other chemical properties on all the samples, and water retention curve and saturated hydraulic conductivity on a sub-set). Several pedotransfer functions were tested; the PTF-Vereecken proved to be the best one to derive hydraulic properties of the entire soil database. The Monte Carlo approach was used to analyze model sensitivity to two measured input parameters: the slope of water retention curve (n) and the saturated hydraulic conductivity (\( k_0 \)). The analysis showed sensitivity of the simulated process to the parameter n being significantly higher than to \( k_0 \), although the former was much less variable. The PTFs showed a smoothing effect of the output variability,
even though they were previously validated on a set of measured data. Interesting positive and significant correlations were found between the \( n \) parameter, from measured water retention curves, and the NDVI (Normalized Difference Vegetation Index), when using multi-temporal (2004–2018) high resolution remotely sensed data on maize cultivation. No correlation was detected when the \( n \) parameter derived from PTF was used. These results from our case study mainly suggest that: (i) despite the good performance of PTFs calculated via error indexes, their use in the simulation of hydrological processes should be carefully evaluated for real field-scale applications; and (ii) the NDVI index may be used successfully as a proxy to evaluate PTF reliability in the field.

**Keywords:** hydraulic properties; simulation model; PTF; Monte Carlo; NDVI; maize; above ground biomass

1. Introduction

The definition of the terms of the water balance in the soil-plant-atmosphere (SPA) system is crucial to many practical applications in both agriculture and environmental issues (e.g., crop production, nitrate leaching, etc.) so as to properly deliver concrete solutions for achieving a large set of goals, such as the Sustainable Development Goals (e.g., SDG2 in [1]). Moreover, the estimate of fluxes and storage of water in the SPA system is fundamental, especially for those applications—from field to landscape scale—which promise to deliver operational answers, as in the case of Decision Support Systems [2].

It is generally assumed that flux and storage of water in the unsaturated zone are adequately described by the classical nonlinear Richards’ equation, the reliability whereof depends on accurate definition of the hydraulic properties of the porous medium within the simulated domain (i.e., water retention and hydraulic conductivity). For field scale applications, the inherent spatial variability of soil hydraulic properties has to also be correctly identified and described. The latter task, although not simple, may be managed by characterizing soil hydraulic behavior at several sites in a field both by laboratory and field methods [3]. Of these, (i) laboratory methods, while widely used, often lack representativeness of the real field behavior, unless they are scaled up to the field scale [4]; (ii) classical field methods are the most reliable for real applications despite their evaluation being time consuming and difficult to implement [5]; and (iii) inverse parameter estimation methods coupled with assimilation data from remote sensing are promising but still under development [6,7].

Due to the cons of the methods above mentioned, it is not surprising that for large scale applications indirect methods (i.e., pedotransfer functions, hereafter PTFs) are remarkably widely and increasingly used, being often thought to be characterized by an acceptable cost-benefit ratio. All PTFs in fact are based on widely available and relatively less expensive soil information, such as textural data, soil organic carbon, and bulk density, even if other information may be used in PTF development and application (e.g., soil structure [8,9], terrain attributes [10]). PTFs are nowadays the main technique of estimation when hydraulic properties are required at many points on a landscape scale or when spatial analysis is required [11].

As predictive equations, PTFs are usually evaluated in terms of correspondence between measured and predicted values of hydraulic properties (i.e., soil water content, hydraulic conductivity) or parameters, or functional evaluation [12–15].

Typically, the indexes used are the mean error (ME), mean absolute error (MAE), the coefficient of determination \( R^2 \), the root mean square error (RMSE), the model efficiency (EF) calculated on the residuals [10,16–18], or on observed average or variance [19]. Integrated indexes have also been developed [14,20].

A far more useful approach indeed is functional evaluation, the statistical study of the outcome variability of a simulation model when the variability is due solely to the uncertainty in the
PTFs [21], [22], with the aim of evaluating the performance of PTFs in terms of their application rather than accuracy.

In the context of simulation model applications, field scale soil heterogeneity can be represented through a detailed description of soil hydraulic properties’ variability and simulating the processes as a 2D/3D flow field [23]. However, this approach is highly demanding in computing and physical analysis and its use is confined to specific research studies [24]. Alternatively, soil heterogeneity may be accounted for by considering the flow field as an ensemble of parallel and statistically independent tubes, where the flow processes are assumed to be vertical 1D [25]. This approach is widely applied, being more feasible despite requiring some restrictive assumptions on the flow processes [26–29]. This is known as the stream tube approach and is frequently used in a stochastic framework, where the hydraulic parameters of each tube are randomly drawn by their statistical distribution. In this configuration, the stream tube approach may be used to analyze the sensitivity of the model to each hydraulic parameter, for the specific flow process under study (infiltration, drainage, evapotranspiration), by repeatedly running the flow model with multiple values of that parameter and analyzing the statistics of a selected output (water content, pressure head, etc.) (see, for example [29]). In both cases, one needs the variability of the hydraulic properties to be known at the scale of interest.

In describing a flow process at field scale, PTF-based hydraulic properties may provide acceptable descriptions of the average water content and flow behavior. On the other hand, they have often been found to be inadequate to reproduce actual field spatial patterns, due to the variability of soil hydraulic properties generally observed in natural heterogeneous soils (see, for example, Reference [30]). In other words, this means that even when PTFs prove to be acceptable descriptors of the average behavior in the field, the problem of the representativeness of the field variability (and the related spatial patterns) may still exist. This issue has rarely been dealt with in the literature, and, according to us, deserves to be analyzed in greater depth, in view of the widespread use of PTFs for large scale applications.

To this end, and based on a case study in the Lodi plain (in northern Italy) where forage maize is the primary crop, this paper has a twofold objective:

(i) Evaluating to what extent the variability of the soil hydraulic properties parameters is reflected in the hydrological processes observed at the field scale. To do that, a preliminary analysis of the sensitivity of a physically-based agro-hydrological model to the measured variability of both water retention and hydraulic conductivity parameters will be carried out. This analysis will be based on a stream tube approach used in a stochastic (Monte Carlo) framework;

(ii) Evaluating the effectiveness of selected PTFs in reproducing the field scale hydrological pattern described by the measured hydraulic properties through using independent and spatially distributed information (Normalized Difference Vegetation Index - NDVI) as data quality control.

In other words, we analyze the impact on the process under study of both (1) the variability of soil hydraulic properties parameters, and (2) the effectiveness of PTFs in representing the actual variability of hydraulic properties.

2. Materials and Methods

2.1. Study Area

The study area reported in Figure 1 is located in the north of Italy (Lombardia region) and refers to about 2,000 ha on the eastern alluvial plain of the Adda River, a tributary of the Po River.
The area is comprised of alluvium materials ranging from Würm to present day and the climate is characterized by cold winters (mean temperature of 3.6 °C) and warm humid summers (mean temperature of 22.5 °C). Rainfall ranges between 800 and 1000 mm per year, evenly distributed throughout the study area (www.arpalombardia.it). A large number of drainage channels are characteristic of the landscape. According to a 1:50,000 soil map of the Lombardia region [32], the soil types in the study area include mainly Alfisols (Aquic- and Ultic Haplustalfs) and, with minor extension, Inceptisols (Typic- and Aquic Ustochrepts), characterizing the recent alluvial terraces in the eastern part of the area. Forage maize cultivation (Zea mays) is the major land use type, which is both spatially continuous and uniform (i.e., crop type and crop management). This was directly observed and also confirmed by interviews with local farmers and farmer associations who described relatively homogeneous farming practices, including fertilization (organic and mineral N), pest control, sowing dates, soil management, etc. [31].

2.2. Hydraulic Properties

In order to consider the overall soil spatial variability, 105 locations were identified according to a stepwise sampling strategy based on annealing procedures [31]. Two soil horizons (upper Ap horizon and the underlying Bw horizon) were sampled at each location in 2004. Soil textural analysis was carried out on these samples using the hydrometer-sieving technique [33]. Moreover, 62 undisturbed soil samples were collected from the upper horizon using an aluminum cylinder (Φ = 7 cm, l = 7 cm). The samples were slowly wetted in laboratory from the bottom; after saturation, saturated hydraulic conductivity \( k_0 \) was determined by the variable head technique [34]. Soil water retention curve \( \theta(h) \) was measured starting from the saturation by using the suction table (9 points) and the pressure plate apparatus (3 points) [35].

Soil hydraulic properties were also estimated by means of PTFs, using textural data and soil organic matter content as input parameters, in 43 samples corresponding to the remaining locations (out of the total of 105), plus 31 samples randomly selected from among those already analyzed.
in laboratory. These last were used to evaluate the PTFs’ performance. Three PTFs were tested: Vereecken [22], Hypres [21], and ROSETTA [36].

2.3. Evaluation of PTFs’ Performance

Agreement between the measured and estimated values was expressed by the following indexes: the root mean squared error (RMSE), the root mean error (RME), and the modelling efficiency index (EF). For all the indexes, $M_i$ is the ith measured value, $E_i$ is the estimated ith value, and $n$ is the number of soil water content pairs. $\bar{M}$ is the mean of measured soil water content.

The root mean error (RME) ranges between $-\infty$ and $+\infty$, with the optimum $= 0$. If positive, it indicates that the model overestimates the prediction, if negative indicates underestimation, and when close to zero, it indicates the absence of trends:

$$RME = \frac{\sum_{i=1}^{n} E_i - \sum_{i=1}^{n} M_i}{\sum_{i=1}^{n} M_i}$$  \hspace{1cm} (1)

The root mean square error (RMSE) has a minimum and optimum value at 0. It is a difference-based measure of performance of the model expressed in quadratic form, and is fairly sensitive to outliers.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (E_i - M_i)^2}{n}}$$  \hspace{1cm} (2)

Modelling efficiency (EF) can take either positive or negative values; 1 being the upper limit, while negative infinity is the theoretical lower bound. EF values lower than 0 result from a worse fit than the average of measurements:

$$EF = 1 - \frac{\sum_{i=1}^{n} (E_i - M_i)^2}{\sum_{i=1}^{n} (M_i - \bar{M})^2}$$  \hspace{1cm} (3)

2.4. Soil-Plant-Atmosphere Model

The soil-plant-atmosphere water balance model was investigated using a numerical code written in MATLAB for studying water and solute processes in both one and two domains [29,37,38]. The model calculates the soil water flow by solving the Richards’ equation using an implicit, backward, finite differences scheme with explicit linearization.

Several water retention and hydraulic conductivity models can be selected. In this study, the soil water retention was described by the unimodal $\theta(h)$ relationship proposed by Reference [38] and expressed here in terms of the effective saturation $Se$, as follows:

$$Se = \left[ \frac{1}{1 + (\alpha |h|)^n} \right]^m$$  \hspace{1cm} (4)

with $Se = (\theta - \theta_r)/(\theta_0 - \theta_r)$, $\theta_r$ and $\theta_0$ being the residual water content and the water content at $h = 0$, respectively, and in which $\alpha$ (cm$^{-1}$), $n$, and $m$ are curve-fitting parameters.

Mualem’s expression was applied to calculate relative hydraulic conductivity $k_r$ [39]. Assuming $m = 1-1/n$, a closed-form analytical solution to predict $k_r$ at a specified volumetric water content was obtained:

$$k_r(Se) = \frac{k(Se)}{k_0} = Se^{\tau} \left[ 1 - \left( 1 - Se^{\frac{1}{m}} \right)^{m-2} \right]$$  \hspace{1cm} (5)

in which $k_0$ is the hydraulic conductivity at $\theta_0$, and $\tau$ is a parameter which accounts for the tortuosity and partial correlation between adjacent pores.

Flow rates and pressure heads, whether constant or variable over time, can be assumed as the upper boundary condition. Gradients of different value, pressure heads, or flow rates, again whether
The model discretizes the spatial flow field in a prescribed number of nodes (usually 100) of constant width (Dz). Time discretization starts with a prescribed initial time increment (Dt). This time increment is automatically adjusted at each time level according to the criteria proposed by Reference [40].

The setting of the boundary conditions at the top and the bottom of the soil profile for model application at field scale, including irrigation, precipitation, and potential evapotranspiration, is described in Reference [30], where potential evapotranspiration $ET_p$ is calculated using the Penman-Monteith equation [41] based on daily data (i.e., air temperature and humidity, net radiation, wind speed). A simplified approach of the Penman-Monteith equation can be used on a large scale, due to the lack of detailed daily information required for its application. In this study, specifically, potential evapotranspiration was calculated from reference evapotranspiration ($ET_0$), estimated by means of the Hargreaves and Samani formula [42], and $K_c$, a crop-specific factor ($ET_p = K_c \times ET_0$).

$ET_p$ is partitioned into $T_p$ and $E_p$ on the basis of Beer’s law [43]:

$$E_p = ET_p e^{-k^* LAI}$$
$$T_p = ET_p - E_p$$

where $k^*$ is the light extinction coefficient and LAI is the leaf area index. Unit gradient ($dH/dz = -1$) was assumed at lower boundary condition.

Water uptake, considered an additive term in the Richards equation, and actual transpiration are simulated according to the model proposed by Reference [44], where root water uptake $S$ is described as a function of the pressure head, $h$:

$$S(h) = \alpha(h) \cdot S_{\text{max}} = \alpha(h) \cdot T_p / |z_r|$$

with $z_r$ (cm) being the thickness of the root zone and $\alpha(h)$ a semi-empirical function of pressure head $h$, varying between 0 and 1. The shape of the function $\alpha(h)$ depends on four critical values of $h$, which are related to crop type and to potential transpiration rates. The actual transpiration rate $T_{\text{act}}$ (cm d$^{-1}$) is computed by the integration of $S$ over the root layer.

To get a reliable soil water balance, proper parameterization of soil hydraulic properties through a calibration procedure is essential. The model applied (similar to the SWAP model [45]) was calibrated and validated during the years 2002 and 2003, respectively, by comparing measured by Time Domain Reflectometry and simulated soil water content for several soil depths at one site representative of the study area. Specifically, twelve probes were installed: (i) two, vertically at depth of 0–0.15 m, one within-row and another between-row; (ii) eight, horizontally at 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 m; and (iii) two, vertically at 0.9–1.1 m and 1.1–1.3 m below soil surface. Details can be found in Reference [46], in which the calibration site is named Caviaga district.

### 2.5. Monte Carlo Simulations

The widely used Monte Carlo technique, which treats any uncertain parameter as a random variable that obeys a given probabilistic distribution, is suitable for analyzing probabilistic uncertainty about the specific flow process being studied [29]. As an alternative, it can be used to perform sensitivity analysis of soil hydraulic parameters.

The technique assumes that stochastic processes can be described by a large number of equally probable realizations. N successive sets of realizations are created by randomly drawing one value from the assumed distribution of each of the $P$ input parameters ($P=pi$ for $i=1, 2, \ldots, P$) or from a joint multivariate density function for correlated parameters. A deterministic response vector $Y (Y=\{P\})$ is calculated by running the deterministic model N times, such that the statistical moments of the desired output dependent variables can be estimated. To cite an example, by solving the Richards’ equation N
times, one can obtain the mean and variance of the water content due to the fluctuations of the p-th random parameter according to its distribution:

\[
< \theta_p(z, t) > \geq \frac{1}{N} \sum_{j=1}^{N} \theta_p(z, t)
\]  

(8)

\[
\sigma^2_p = \frac{1}{N-1} \sum_{j=1}^{N} [\theta_p(z, t) - < \theta_p(z, t) >]
\]  

(9)

where \(< >\) denotes ensemble averaging.

In order to perform the Monte Carlo numerical simulation, assuming uncorrelated parameters, the statistical moments of the probability density function of model input parameters are required. The statistical moments were obtained from the 62 measured hydraulic properties, assuming that soil hydraulic variability can be described by the statistical distribution of such parameters.

500 simulations of the soil water content were carried out, assuming the soil profile with parameters of hydraulic functions to be generated randomly in the first layer and fixed at the value recognized where the model was calibrated. \(N=500\) was chosen as an acceptable compromise between the degree of convergence of the output statistics (stability of ensemble mean and variance estimates of water content) and time consumed.

2.6. Remote Sensing Data

NDVI as proxy for model validation:

Following the approach reported in [31], NDVI was used as an estimate of the forage maize green biomass production, NDVI being strongly related to photosynthetic activity and widely used to estimate landscape patterns of primary production [47–49].

Specifically, field observations and interviews with farmers verified the seeding date and confirmed that July represented the period of maximum crop growth. During this month, green biomass was randomly sampled at 10 different sites in the study area to test NDVI-maize green biomass production. At each of the 10 sites, the samples were geo-referenced and eight maize plants were collected then weighed to estimate total dry biomass. Then, time-lapse (March–September 2004) NDVI images were obtained by MODIS VI and processed in ArcView software (i.e., image overlay, checking artefacts/outliers) in order to produce pixel based NDVI curves, corresponding with the sampling sites, for the period 10 June–28 July 2004. For each curve, the integral was calculated and the data were matched with the green biomass measurements. Results revealed a significant positive correlation \(r^2:0.83\) between these MODIS NDVI data and total dry biomass data [31].

NDVI as a proxy for hydraulic properties variability detection:

To this end, we used the following layers:

- An NDVI map at high resolution from a visible RGB and near IR Quickbird image for 21 July 2004 (spatial resolution 2.4 m).
- NDVI maps derived from free Landsat 5 TM and Landsat OLI-8 scenes Collection 1 Level-2 on-Demand—path 193–194/row 28–29—at spatial resolution of 30 m, atmospherically corrected [50,51], including a cloud, shadow, water, and snow mask produced using CFMASK [52], as well as a per-pixel saturation mask for years 2009, 2010, 2013, 2014, 2015, 2016, 2017, and 2018 ranging from 1 to 25 July, in order to strengthen our assumptions on the relationship between NDVI and hydraulic properties. The selection of these layers (years and periods) was done considering their availability, the need to observe NDVI data within July as was done for 2004, and the presence of masking clouds during specific days.

NDVI was calculated using the Equation (10), where \(\rho_{RED}\) and \(\rho_{NIR}\) are reflectance into band 3 (red, 0.63–0.69 µm) and band 4 (near-infrared, 0.76–0.90 µm) for Landsat 5 TM; into band 4 (red, 0.64–0.67 µm) and band 5 (near-infrared, 0.85–0.88 µm) for Landsat OLI-8.
NDVI = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} \quad (10)

Firstly, all areas (other land uses, roads, buildings, outliers, etc.) different from maize crop were masked by applying a supervised classification of the study area using ERDAS Imaging 8.4 software (ERDAS, Inc.- www.erdas.com); secondly, the values (pixel extraction) of NDVI recorded corresponding to (same coordinates) the 105 maize sites where soil samples were collected were extracted from the satellite images by using the ArchMap 10.3.1 (Environmental Systems Research Institute – ESRI, Redlands, CA) spatial analyst tool. During the analysis, the number of pairs of data valid for comparison varied from 19 in 2015 to 26 in 2009, due to masking procedures, outliers, and possible cloudiness.

The result of these operations was a dataset from 2004 to 2018, reporting the maize NDVI values corresponding to the coordinates of the soil sample sites. From this dataset, the data relating to the sample sites where both soil hydrological measurements and PTFs estimates were carried out were successively extrapolated.

2.7. Climate Data

Daily climate data of Lodi plain (rain and temperatures from historical database at www.arpalombardia.it) were acquired for the years considered in the study. Statistics on the data were calculated (average temperatures and cumulated rains) for May and June, with the scope of interpreting the effects on maize growth of climatic conditions during a critical period for the crop (shooting phase) before the acquisition of the NDVI data.

3. Results

3.1. Variability of Soil Hydraulic Properties

The 62 soil water retention experimental data were fitted to the Van Genuchten equation [38] assuming m=1−1/n and fixing θ_r=0. The mean of the parameters along with standard deviation, skewness, and coefficient of variation are reported in Table 1. Furthermore, textural information is also reported.

<table>
<thead>
<tr>
<th></th>
<th>(\theta_s)</th>
<th>(\alpha)</th>
<th>(n)</th>
<th>(k_0) (cm d(^{-1}))</th>
<th>Clay (%)</th>
<th>Silt (%)</th>
<th>Sand (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.42</td>
<td>0.02</td>
<td>1.32</td>
<td>486</td>
<td>16.4</td>
<td>34.0</td>
<td>49.6</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.06</td>
<td>0.02</td>
<td>0.10</td>
<td>897</td>
<td>3.8</td>
<td>7.1</td>
<td>8.5</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
<td>13.9</td>
<td>109</td>
<td>7.74</td>
<td>185</td>
<td>23.0</td>
<td>20.8</td>
<td>17.1</td>
</tr>
<tr>
<td>Type of distribution (n: normal; Log-n: Lognormal)</td>
<td>n</td>
<td>Log-n</td>
<td>n</td>
<td>Log-n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
</tbody>
</table>

What immediately comes up is the unexpectedly low variability of the \(n\) parameter, whereas very high variability is shown by the \(\alpha\) parameter. Moreover, very large variability—as expected—is reflected in the saturated hydraulic conductivity \(k_0\). \(\theta_s\) and \(n\) follow a normal distribution, and \(\alpha\) and \(k_0\) a log-normal distribution.

3.2. Model Validation

Parameters for the point-based calibration and validation of the model in the same soils of the area under study are reported in Reference [46]. We performed an analysis to check: (i) if the site of measurement chosen for the point-based model calibration is representative of the hydraulic properties variability, and (ii) the sensitivity of the model to the variability of the hydraulic properties. In this respect, the average and standard deviation of the simulated \(\theta(t)\) at \(z = 0\text{–}15\) cm depth were calculated.
following 500 model runs performed by the Monte Carlo technique, with 500 random vectors of the parameters of the van Genuchten–Mualem model generated according to the distributions reported in Table 1. Crop and other model parameters are those of the model SWAP reported in [46]. The results are shown in Figure 2., where, for the sake of simplicity, we have reported the sensitivity analysis of only the n and k₀ parameters. In addition, in Figure 2, the soil water content measurements and the simulations coming from the calibrated parameters are also shown (red squares and red dashed line, respectively). What comes up is the good agreement between measured and simulated soil water content. The agreement is rather good during the drying period after the rain, with the exception of the period lasting from DOY 335 to DOY 342. The differences can be attributed to the amount of the rain inputted in the model. The gauge at the Caviaga site was located less than 1 km away from the experimental station, therefore we presume that the spatial variability of the rain played a role in our simulation. Therefore, we attribute the underestimation of soil water content simulation to the underestimation of few millimeters of rain in the experimental farm with respect to the measurement gauge.

![Figure 2](image)

**Figure 2.** Soil water content measured and simulated (red squares and red dashed line) in the reference soil. Black and grey lines are the average soil water content applying 500 Monte Carlo runs of the n and k₀ parameters, respectively. Error bars represent the standard deviation of the two set of 500 Monte Carlo runs.

The two sets of 500 Monte Carlo simulations confirm that the site chosen in 2002 [46] adopting classical pedological rules was corrected, representing the variability of the hydraulic properties of the area. This result is not trivial, because the assumption that the representative soil profile is also representative from a hydrological point of view is always implicitly assumed in hydrological studies but seldom proved [53].

### 3.3. Predictive Capability of PTFs

Average soil water retention curves were calculated for the 31 samples for which we could use both measured data and data estimated by PTFs (Hypres, Rosetta, Verecken). Table 2 shows the statistics of the fitting. Hypres and Verecken PTFs seem to better reproduce the measured soil water retention curve. Of the two, we decided to apply the PTF by Verecken for its better performance on...
EF and because its shape is closer to the measured one, whereas the Hypres PTF crosses the measured curve, thus showing a lower value of ME.

Table 2. Statistics of pedotransfer functions (PTF)s.

<table>
<thead>
<tr>
<th></th>
<th>Root Mean Error (RME)</th>
<th>Root Mean Square Error (RMSE)</th>
<th>Model Efficiency (EF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPRES</td>
<td>0.09</td>
<td>5.51</td>
<td>-1.56</td>
</tr>
<tr>
<td>VERECKEEN</td>
<td>-2.86</td>
<td>5.47</td>
<td>-1.25</td>
</tr>
<tr>
<td>ROSETTA</td>
<td>-7.03</td>
<td>5.76</td>
<td>-2.56</td>
</tr>
</tbody>
</table>

Figure 3 shows the average soil water retention curve both as measured and as estimated by the PTF-Vereeken; the vertical bars represent ±1 standard deviation. As also reported in Table 2, agreement between the measured and estimated curves is very good, with rather similar shapes of the two average curves. What is also evident, as expected, is the smoothing effect of the PTF, showing a very narrow variability if compared with the measured data.

Figure 3. Average and error bars (standard deviation) of both measured (black lines) and estimated (grey lines) SWRCs.

3.4. Soundness of PTF Estimations Based on NDVI Data

We evaluated the relationship between soil hydraulic parameters and active green biomass data as estimated in 2004 by the locally calibrated, Quickbird NDVI image (2.4m x 2.4m Normalized Difference Vegetation Index at the maximum crop growth period), used here as a proxy to verify the PTF reliability. This evaluation was done considering the 31-sample dataset consisting of both measured data and data estimated by PTFs. We assumed that maize was well irrigated without nutrient deficiency (i.e., nitrogen). Details of this procedure of calibration are reported in Reference [31].

Among the hydraulic parameters, only the n parameter, and that only for the measured data, appeared significantly correlated with the NDVI in 2004 ($r^2 = 0.63$ at 0.01 level of significance); all the other parameters were uncorrelated (data not shown). Then, in order to verify if this finding was consistent over time, the same procedure was applied using the NDVI data derived from free Landsat 8 images (30m x 30m spatial resolution) for the year 2004 again, and also for years 2009, 2010, 2013, 2014, 2015, 2016, 2017, and 2018. The use of the multitemporal Landsat images was justified because of their free availability as against the Quickbird ones that are available for a fee, thus making the procedure more easily repeatable. In Table 3 the relationships between measured and estimated parameters vs NDVIs are shown. The estimated parameters are all uncorrelated to the NDVI, while at this coarser spatial resolution, positive and statistically significant correlations were also found, again on n parameter derived from measured water retention curves for the years 2004, 2010, 2013, 2016, and
2017. For the same years, significant negative correlation was also found for the \( \alpha \) parameter. Only for 2015 was \( k_0 \) positively correlated with NDVI.

**Table 3.** Square of the correlation coefficients (\( r^2 \)) between measured and estimated hydraulic parameters vs NDVI estimated by Landsat-8 images.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Measured</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_s )</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>( k_0 ) (cm/d)</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>( \alpha ) (1/cm)</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>( n )</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

(2004) (significant correlation at **0.01 and *0.05 level)

During the analysis, the number of pairs of data valid for comparison varied from 19 in 2015 to 26 in 2009, 2017, and 2018, due to masking procedures, outliers, and possible cloudiness.

Because the \( n \) parameter is the only one correlated to the NDVI, we accordingly focused our analysis on this parameter. We left out the analysis of the \( \alpha \) parameter because it is correlated to a certain extent to the \( n \) parameter (\( r = -0.62 \) for measured parameters). In Figure 4, as an example, are plotted the measured and estimated data of the \( n \) parameters vs NDVIs pertaining to two years, which show very different behavior: 2004 correlated and 2009 uncorrelated.

**Figure 4.** Measured (●) and estimated (△) values of \( n \) parameter vs NDVI values estimated by Landsat 8 images.

In 2004, NDVI and \( n \) are significantly positively correlated (\( r: 0.48 \) at 0.01 level) only when analyzing \( n \) obtained from measured data, while no correlation is observed when analyzing the \( n \) values estimated from PTF. In 2009, NDVI and \( n \) do not show any consistent correlation, regardless of whether \( n \) comes from measured or estimated data.

With the aim of better interpreting our different findings for different years, we investigated the climate (temperature and rain) for the years considered in our study. The results are shown in Figure 5, which reports for those years the average daily temperature versus the cumulated rain recorded during the period May–June, when intense maize growth is recognized.
It is quite evident that lower average temperatures (just below 20 °C) coupled with major rains were recorded during the years whose values of NDVI in July appeared significantly correlated with the n parameter of the measured soil water retention curves (area under dashed line).

We interpret these findings assuming that during the period May-June irrigation coupled with major rains and slightly lower temperatures is a condition that can result in the phenomenon of crops suffering due to waterlogging, mainly in soils characterized by less than optimal drainage (lower value of n parameter). In fact, the SWRC is the cumulative of the pore-size distribution and the parameter n is an indicator of the shape of pore-size distribution. It represents the slope of SWRC in the capillarity region and it measures how rapidly the water is drained for a given change of pressure head. The higher n is, the higher the amount of water drained in that interval of pressure head. This is mathematically reflected in the van Genuchten–Mualem model. By looking at Equation (5), the lower n is (and thus m), the lower is k(θ) and thus the slower the draining process. This may be seen graphically in the Figure 1 and Figure 7 reported in Reference [54].

4. Discussion

When dealing with flow processes on a large scale (i.e., field, regional), the role of soil heterogeneity and the related variability of hydraulic properties must necessarily be considered. In this context, the widespread use of simplified methods (i.e., PTFs) for estimating soil hydraulic properties is an attempt to overcome the frustrating difficulty of finding adequate sets of hydraulic properties, especially for large scale applications.

However, in doing so, clarity on two crucial aspects is essential: (i) the variability of soil hydraulic properties parameters may have a significant impact on the process under study; (ii) the PTFs may not provide reliable estimation of the actual hydraulic properties’ variability. These two points are described in detail below.

4.1. Variability of Soil Hydraulic Properties Parameters and its Impact on the Process under Study

Results produced after the analysis of the statistical variability of the parameters seem to be very interesting. For the sake of simplicity, we limit our discussion here to the two parameters having divergent behavior (i.e., n and k₀). Indeed, despite the fact that variability of the parameter n is considerably lower than that shown by k₀ [CV (n) = 7.74%; CV (k₀) = 185%] (see Table 1), its variation affects the variability of the simulated process to an extent equal to or even larger than for k₀.
This is evident in Figure 2, where the width of the ±1 standard deviation vertical bars is displayed. Although on average they show similar variability [SD (n) = 0.04; SD (k₀) = 0.03], the model is more sensitive to the variability of the n parameter during the drainage processes (i.e., from day 310 to 345), whereas it is more dependent on k₀ variability during the infiltration / redistribution processes (i.e., from day 290 to 310). This agrees with the findings in Reference [29], who observed a similar behavior for a field process simulated with bimodal hydraulic parameters, and with Reference [55], who found during an infiltration process that predictive uncertainty is most sensitive to uncertainty in the saturated hydraulic conductivity k₀.

Therefore, the impact of the parameters’ variability on the flow process under study is much more affected by the sensitivity of the model to the parameter itself than by its coefficient of variation.

4.2. Effectiveness of PTFs in Representing the Actual Variability of Hydraulic Properties and its Impact on the Process under Study

Assessment of the deviation between measured and predicted values of both parameters and variables is required as a preliminary step in the evaluation of the performance of the PTF. In our case study, the statistics of agreement between measured and estimated-by-PTF soil water retention curves are satisfactory, indicating that the chosen PTF performs well—on average—in reproducing measured soil water retention curve.

However, because the PTFs are generally used as input in models simulating non-linear processes, it is highly desirable to carry out their functional evaluation in respect of the model’s performance rather than just the PTF performance [11,56]. This becomes even more important in the presence of soil heterogeneity.

As reported in Table 3 and shown in Figure 4, the bio-physical process (i.e., above-ground biomass estimation by means of NDVI) is strongly related to soil heterogeneity and specifically to the variability of the n parameter (and partially to the α parameter, correlated to n), while the variability of the other parameters is irrelevant for the AGB estimation.

Here, it needs to be stressed that this soil hydrological analysis is related to the plowed Ap horizon, where the roots are more active and therefore strongly influence crop growth. Moreover, the area under study is rather homogeneous in terms of soil horizon sequences below the Ap.

This result, in conjunction with the outcome of the sensitivity analysis previously discussed, induced our further focus on analysis of the n parameter.

Hence, the question that arises is: how much does the variability of the n parameter—derived from PTF—reproduce the process variability? In fact, although the typical PTF evaluation procedure has provided satisfactory results (see Table 2 and Figure 3), the functional evaluation of the PTF performance on NDVI data has clearly highlighted different behaviour between measured and estimated data. The reason for this lies mainly in the fact that the PTFs are smoothing models of the real hydraulic properties, and therefore are not able to reproduce the actual process variability. Smoothing the inherent soil heterogeneity without taking into account the parameter estimation uncertainty [57] will reduce the PTF performance in reproducing soil heterogeneity [58,59].

This can be a major problem, since currently PTFs are “the reference” for many practical applications dealing with agriculture and environment, whereas soil field measurements have become very rare and their use is restricted to a small community of specialized scientists. This problem is exacerbated by the widespread use of models simulating soil water balance in order to get different outcomes (e.g., nitrate leaching, crop productivity, organic matter balance, etc.).

Here we suggest that PTF has to be used with caution, especially when dealing with soil heterogeneity.

In fact, a proper acknowledgement of the occurrence of different soil properties (e.g., diverse n values) within the same landscape may further enhance the impact of sustainable land management practices, since they would be tuned to the specific soil properties and soil types; this outcome is even further enlarged when considering actions for climate change adaptation.
5. Conclusions

In this research, an area of about 2000 hectares was characterized for hydraulic properties through 105 soil sampling instances in the root zone layer. The dataset, composed of measured and estimated data, was used to investigate the effects of the heterogeneity of the hydraulic properties parameters on a process simulated through a modelling application. The results obtained have shown higher sensitivity of the model to the parameter n, which variability can significantly impact the model outcomes.

This evidence led us to wonder if, in our case study, methods of estimating hydraulic parameters, such as the PTFs, could be applied, preserving at the same time the information on soil heterogeneity. We first assessed the reliability of some PTFs by calculating three error indexes, which led to promising results indicating the PTF of Vereecken as the best for our case study. However, these laboratory-based results were not fully reflected in the processes occurring at field scale. In fact, the use of the NDVI as independent information to indirectly verify the PTFs’ reliability showed no correlation between the vegetation index and the parameter n when estimated, while the measured data appeared positively and significantly correlated.

Why is NDVI, and hence the estimation of the total green biomass, positively correlated with n and k0? The study area is characterized by intensive agriculture—mainly livestock and maize farming—with old irrigation systems where over-irrigation practices are widely applied (i.e., flooding irrigation). From a hydrological point of view, k0 and n are the two parameters mainly affecting the drainage process [27], i.e., the higher their value, the more hydraulically conductive the soil. Therefore, for this case study conductive Ap horizons—where the majority of the water uptake takes place—provided better conditions for crop growth, reducing the effect of saturation (e.g., lack of oxygen for root uptake functioning [60]). Therefore, despite several factors that could influence crop responses, such as slopes, soil depth, rainfall, temperatures, and management practices, in this case study (plain territory, not very extended, similar management) the contribution of spatial variability of soil hydraulic properties is considered the main factor.

These results for our case study provided two main conclusions: (i) despite the good performance of PTFs calculated via error indexes, their use in the simulation of hydrological processes should be carefully evaluated for real field-scale applications; (ii) the NDVI index may be used successfully as a proxy to evaluate PTF reliability in the field.

We believe that additional research is required to develop a functional evaluation of PTFs, before their general application to deliver operational answers.

Author Contributions: Conceptualization, A.B. (Angelo Basile); Data curation, R.D.M., S.F.B. and P.M.; Funding acquisition, A.B. (Angelo Basile); Methodology, A.B. (Angelo Basile) and P.M.; Software, A.C.; Writing—original draft, A.B. (Angelo Basile); Writing—review & editing, A.B. (Antonello Bonfante), A.C., F.T. and P.M.

Funding: This research was funded by EC H2020 LANDSUPPORT project, grant number 774234.

Acknowledgments: The authors wish to thank Dr. Loredana Randazzo who helped in writing the numerical code.

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

References

2. Terribile, F.; Agrillo, A.; Bonfante, A.; Buscemi, G.; Colandrea, M.; D’Antonio, A.; De Mascellis, R.; De Michele, C.; Langella, G.; Manna, P. A web-based spatial decision supporting system for land management and soil conservation. Solid Earth 2015, 6, 903. [CrossRef]


7. Valdes-Abellan, J.; Pachepsky, Y.; Martinez, G. Obtaining soil hydraulic parameters from soil water content data assimilation under different climatic/soil conditions. *Catena* 2018, 163, 311–320. [CrossRef]


43. Prabhakara, K.; Hively, W.D.; McCarty, G.W. Evaluating the relationship between biomass, percent ground cover and remote sensing indices across six winter cover crop fields in Maryland, United States. *Int. J. Appl. Earth Obs. Geoinf.* 2015, 39, 88–102. [CrossRef]


55. Wang, P.; Tartakovsky, D.M. Probabilistic predictions of infiltration into heterogeneous media with uncertain hydraulic parameters. Int. J. Uncertain. Quantif. 2011, 1, 35–47. [CrossRef]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).