

Article

In Search of a Soil Moisture Content Simulation Model: Mechanistic and Data Mining Approach Based on TDR Method Results

Andrzej Brandyk ^{1,*}, Bartosz Szela² , Adam Kiczko ³, Marcin Krukowski ³, Adam Koziol ³ , Jerzy Piotrowski ⁴ and Grzegorz Majewski ³ 

¹ Water Centre, Warsaw University of Life Sciences, 02-676 Warsaw, Poland

² Department of Geotechnics, Geomatics and Waste Management, Kielce University of Technology, 25-314 Kielce, Poland; bszelag@tu.kielce.pl

³ Institute of Environmental Engineering, Warsaw University of Life Sciences, 02-787 Warsaw, Poland; adam_kiczko@sggw.edu.pl (A.K.); marcin_krukowski@sggw.edu.pl (M.K.); adam_koziol@sggw.edu.pl (A.K.); grzegorz_majewski@sggw.edu.pl (G.M.)

⁴ Department of Building Physics and Renewable Energy, Kielce University of Technology, 25-314 Kielce, Poland; piotrowski@tu.kielce.pl

* Correspondence: andrzej_brandyk@sggw.edu.pl; Tel.: +48-22-59-35-377



Citation: Brandyk, A.; Szela, B.; Kiczko, A.; Krukowski, M.; Koziol, A.; Piotrowski, J.; Majewski, G. In Search of a Soil Moisture Content Simulation Model: Mechanistic and Data Mining Approach Based on TDR Method Results. *Sensors* **2021**, *21*, 6819. <https://doi.org/10.3390/s21206819>

Academic Editors:
Zbigniew Suchorab and
Grzegorz Łagód

Received: 9 September 2021

Accepted: 6 October 2021

Published: 14 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Soil moisture content simulation models have continuously been an important research objective. In particular, the comparisons of the performance of different model types deserve proper attention. Therefore, the quality of selected physically-based and statistical models was analyzed utilizing the data from the Time Domain Reflectometry technique. An E-Test measurement system was applied with the reflectogram interpreted into soil volumetric moisture content by proper calibration equations. The gathered data facilitated to calibrate the physical model of Deardorff and establish parameters of: support vector machines, multivariate adaptive regression spline, and boosted trees model. The general likelihood uncertainty estimation revealed the sensitivity of individual model parameters. As it was assumed, a simple structure of statistical models was achieved but no direct physical interpretation of their parameters, contrary to a physically-based method. The TDR technique proved useful for the calibration of different soil moisture models and a satisfactory quality for their future exploitation.

Keywords: soil moisture content; dielectric permittivity; modeling; uncertainty; sensitivity analysis

1. Introduction

The importance of soil moisture as an environmental variable is evident from the viewpoint of its role in the hydrological cycle. Soil moisture impacts rainfall-runoff processes, infiltration, groundwater recharge rates, and constrains evapotranspiration as well as photosynthesis. To some extent, it governs water and energy exchange between the land, plants and the atmosphere [1–4], but embraces multiscale feedbacks as well [5]. The importance of soil moisture was stressed particularly in respect of the partitioning between the sensible and latent heat fluxes, and consequently for the temperature of the surface and the lower atmosphere [6,7]. From the viewpoint of mass and energy transfer, the storage and availability of water is crucial for plant growth, and its shortages are considered to be serious limitations to agricultural production. In recent years, the on-going climate changes have resulted in atmospheric and hydrological droughts that give rise to soil and physiological droughts affecting harvest yields and water supplies. The observed rise in air temperature along with precipitation shortages occur more frequently all over the world, e.g., in Europe and Poland, and affect the security of soil water resources and crop yields.

Since soil moisture content plays a significant role to satisfy water requirements not only of agricultural produce, but also in other numerous respects, the TDR method has

become highly developed and applied worldwide for soil moisture estimation [8–12], being available for other materials as well [13–15]. That method involves electrical permittivity as an indirect parameter of soil moisture content estimation. Electrical permittivity ϵ [–] is a measure of matter molecule behavior in the case of external, variable electrical field application [16]. The values of that parameter for individual phases of porous media are the following: air—1 [–], soil solids 1–15 [–], water—80 [–]. The occurring difference between the permittivities of individual phases is the base for dielectrical methods application, including TDR.

Based on the data from the TDR method, the knowledge on temporal variability of soil moisture content in different hydrological conditions, in particular, becomes indispensable for producing environmental forecasts and improving their predictions [17]. It is also of significance to a number of applications, e.g., management practices that may currently employ soil moisture data all over the world. This is also the basis to acknowledge crop plant growth issues but also an effort to better understand and predict soil water storage in short time increments, e.g., on a daily basis.

In this work, we focused on the performance of various models for soil moisture content simulation, taking advantage of the measurement data obtained by the TDR method. In addition to the physically-based models, the literature review [18] shows that potential is offered also by data mining methods [19]. They facilitate finding relatively simple models that involve easily obtainable hydrometeorological variables, such as air relative humidity and temperature, precipitation, or others, in order to simulate the moisture content of the shallow soil layers.

Among all relevant models, artificial neural networks, support vector machines, Random Forests, boosted trees, and k-nearest neighbor should be included [20]. In these methods, at the training stage, the model structure is formulated based on historical data on different input variables. This becomes decisive for the obtainable quality of model simulations, which is usually indicated by the values of mean absolute and relative errors. The main advantage is the non-complex and flexible model structure which was attempted in this work for the selected statistical and physical models. Additionally, the volumetric soil water contents, recorded by a selected type of sensor, were studied to prove their reliability for the calibration of different soil moisture mathematical models [21].

2. Materials and Methods

The analyses involved an algorithm for the selection of the soil moisture content forecast model. Two mathematical models were taken into consideration: a statistical one and a physical model by Deardorff [6,22] which was especially developed and dedicated for shallow soil layers moisture, shaped by daily weather variations. The comparison of the models is a base for their further exploitation in the field of water resources management and related decision-making. The sense of comparative studies and resultant model selection would facilitate to approach either water shortages or excess (i.e., drought or flood), and to mitigate environmental and ecological threats (not overlooking economic issues). With reference to the above-mentioned problems, an appropriate scheme was proposed (Figure 1).

The considerations involved the observation data included: precipitation, air relative humidity, and air temperature and wind speed, to elaborate soil moisture forecast models through data mining methods. The calculated potential evapotranspiration and precipitation were input to a physical model.

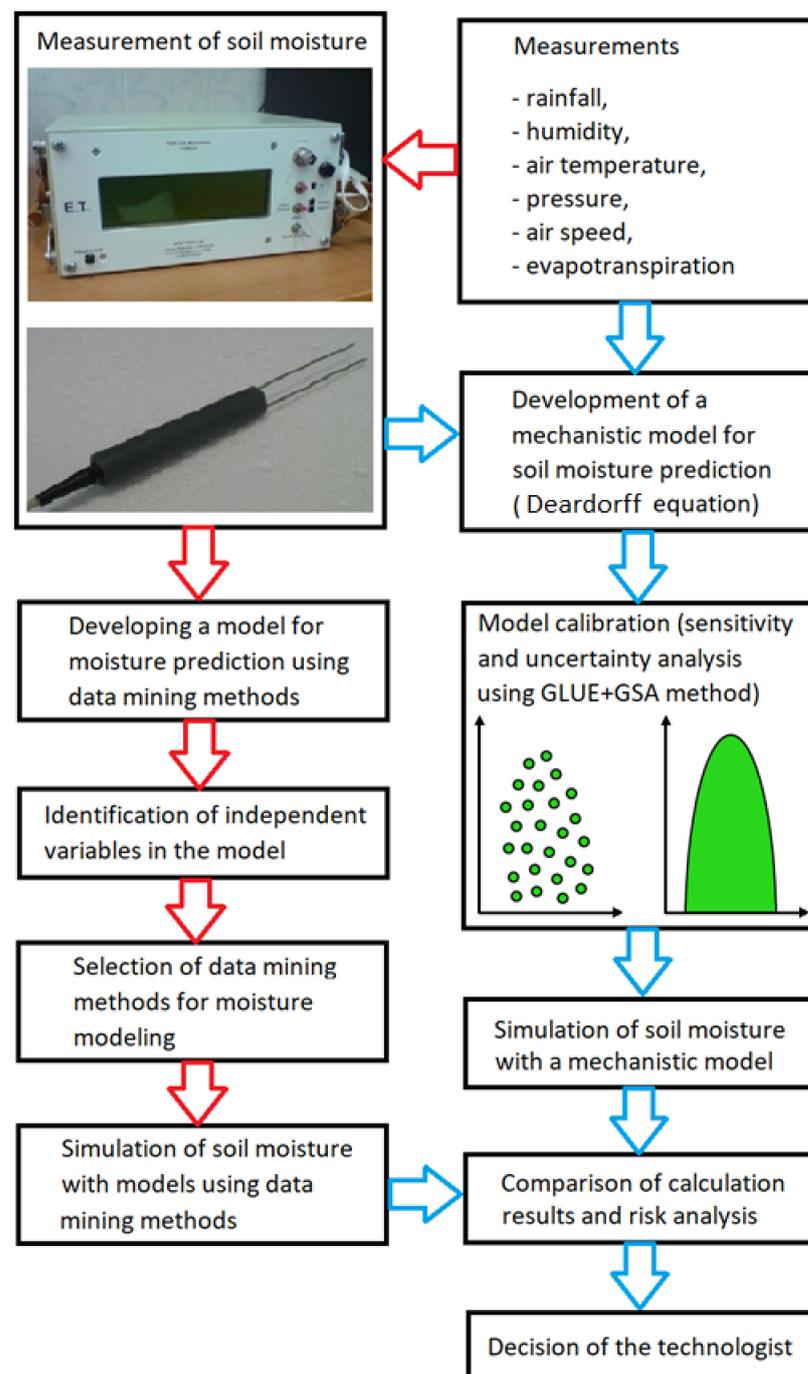


Figure 1. Algorithm for soil moisture model determination.

The main advantage of statistical methods, in comparison to the physical approach, is a much more simple model structure. In a statistical model, the independent variables which significantly influence the phenomenon are distinguished in the first instance. From a practical point of view, a compromise is established between the model complexity, usefulness, and exploitation possibilities. Hence, the analyses performed herein embraced a couple of data mining methods so that the proper model for soil moisture forecast could be found. In reference to the statistical model, the creation of the physical model (mechanistic one) seems to be more complicated. This is due to variability of weather and soil conditions as described by a number of relevant equations [23,24], and the one by Deardorff in particular, which requires the employment of numerical methods as well as proper differential schemes to be solved. Owing to the existence of considerable

interactions between the coefficients of the Deardorff equation there arise problems in their identification that hampers its usefulness.

In order to acknowledge existing constraints in the identification of Deardorff equation parameters a probabilistic estimation was realized through coupled General Likelihood Uncertainty Estimation (GLUE) and Global Sensitivity Analysis (GSA) methodology [25]. Such an approach facilitated to reach reliable distributions of the calibrated parameters in the Deardorff equation, that governed the range of simulation results (95% confidence level).

Based on the determined models (a physical and a statistical one), the simulations of the soil moisture content were performed for the adopted input values. The aforementioned expected model accuracy, resulting from measurement data, would delineate the aftermath of the latter decisions regarding water storage, testing of irrigation efficiency or drainage and irrigation practices, for instance. It would take effect further in evaluation of ecosystem services or ecosystem potential, bringing about uncertainties in the decision-making process.

2.1. Measurements of Meteorological Parameters

The research was performed for the southern district of Warsaw (Ursynów district) within the compounds of Warsaw University of Life Sciences (Figure 2). That area is characterized by a high anthropopressure degree and a high coverage by impervious areas, e.g., the presence of large areas of housing, sidewalks, car parking lots and main streets, that contribute to a fast generation of overland flow.

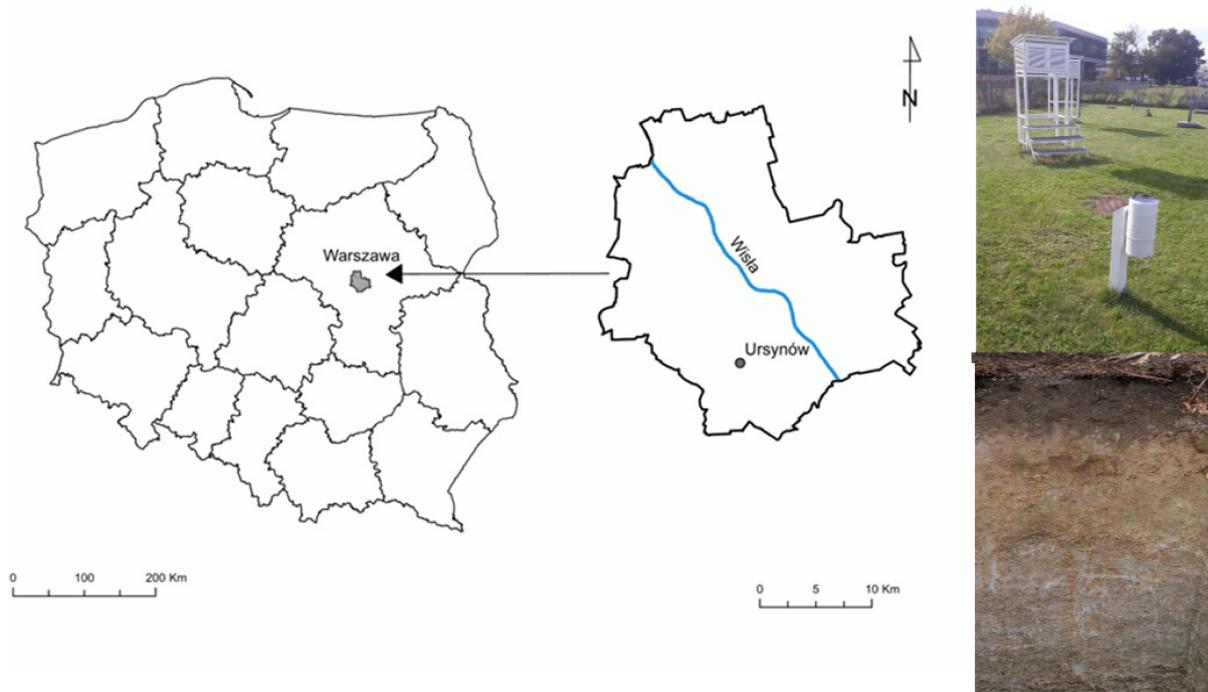


Figure 2. Location of the research site.

However, the percentage of permeable surface includes city parks, and extensive green and recreational areas, which causes water storage in the soil to be essential for the sake of city adaptation to a changing climate. This means there is a need for proper models for the soil moisture content that employ easily obtainable hydrometeorological variables (Table 1). In this respect, the indispensable data range originated from the WULS Ursynów meteorological station ($\lambda E 21^{\circ}02'52''$ $\phi N 52^{\circ}09'38''$, 102.5 m a.s.l.) was used.

Table 1. Sensor types utilized for meteorological parameters monitoring.

Sensor Type	Parameter	Range	Error
Thermohygrometer Vaisala: QMH102 at the height of 2 m.	temperature	−40; +60 °C	<±0.3 °C
	relative humidity	0 to 90% RH	±2%
	relative humidity	90 to 100% RH	±3%
Sonic sensor Vaisala: WS425–B2A1B at the height of 4 m.	wind speed	0–56 m/s	±0.135 m/s or ±3% of the range
	wind direction	0–360°	±2° for wind speed > 2 m/s

Daily precipitation acted as model input for soil moisture simulations, as well as air temperature, relative humidity, and windspeed, that became also indispensable for potential evapotranspiration estimates according to FAO procedure [26] on hourly time step by the Penman–Monteith equation:

$$ET_0 = \frac{0.408 \cdot (R_n - G) + \gamma \cdot \frac{900}{T+273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)} \quad (1)$$

where ET_0 is reference evapotranspiration [mm day^{-1}], R_n is net radiation at the crop surface [$\text{MJ m}^{-2} \text{day}^{-1}$], G is soil heat flux [$\text{MJ m}^{-2} \text{day}^{-1}$], T is air temperature at 2 m height [$^{\circ}\text{C}$], u_2 is wind speed at 2 m height [m s^{-1}], e_s is saturation vapor pressure [kPa], e_a is actual vapor pressure [kPa], $e_s - e_a$ is saturation vapor pressure deficit [kPa], Δ is slope of vapor pressure curve [$\text{kPa } ^{\circ}\text{C}^{-1}$], and γ is psychrometric constant [$\text{kPa } ^{\circ}\text{C}^{-1}$]. Hourly evapotranspiration data were then averaged to obtain mean daily evapotranspiration.

2.2. Measurements of Soil Moisture Content

The selected research site within the WULS campus (the meteorological station) was subject to measurements of the volumetric soil moisture content by the TDR method. The geological setting of that site exhibited generally sandy loams up to 1.9 m depth [27] underlain by sandy clay higher than 10 m in thickness, along with deep ground water table, reaching 11 m beneath the land elevation. The conducted survey revealed soil layers to the depth of 0.2 m, for which moisture was monitored by TDR probes, containing 65% of sand (diameter 2–0.05 mm, including: 24% of coarse sand (1–0.5mm), 16% of medium (0.5–0.25 mm), 13% of fine sand (0.25–0.1 mm) and 12% of very fine sand (0.1–0.05 mm)). The estimated fraction of silt equaled 24% (diam. 0.05–0.002 mm), along with 11% of coarse silt (0.05–0.02 mm) and 13% of fine silt (0.02–0.002 mm), and finally the fraction of clay (diam. < 0.002 mm), reaching 11% [28]. Soil particle size composition was made available by laboratory sieves, that offered diameters from 1 mm to 0.05 mm. However, silt and clay fractions were found through hydrometer (areometric) analyses [29].

Field Operated Multimeter (FOM, E-Test, Lublin, Poland) was used for the soil moisture content measurements, equipped with electromagnetic impulse generator of a falling and rising time equal to about 3×10^{-10} s (300 ps). The generated impulse propagated along the concentric cable (6 m long) of a constant and invariable dielectric permittivity [30]. Then, the impedance changes in the cable–probe system caused the signal reflections, that were next used to determine the time of propagation along the measurement rods (100 mm long, parallel, stainless, 2 mm in diameter with a spacing of 16 mm). The first reflection (the initial one) took place at the cable–probe interface, and the second occurred at the end of the probe. Those reflections were intercepted by the FOM meter in the form of a so-called reflectogram, and automatically interpreted into electrical permittivity and then into soil volumetric moisture content with the use of calibration equation by Malicki [31]. The reliability of the applied equation was proved through comparisons to gravimetric method for sixty soils all over the area of Poland for a range of soil densities. Determination coefficient for the compared variables was equal to 0.98 and the absolute, average error

of moisture content reached the value of 3%. The TDR probes, such as FP/mts, were applied here, that also made it possible to record the soil temperature by AD592CN sensor. They facilitated the measurements performed at two soil depths: 10 and 20 cm at the abovementioned meteorological station, where the land surface was covered by standard, mown grass of 10 cm height. All data were recorded for vegetation periods (01.04–30.09) of 2011, 2013 and 2014.

2.3. Data Mining Methods for Soil Moisture Content Simulation

A literature search [32] revealed a range of data mining methods that can be applied for the sake of environmental processes simulation. An advantage is the simple structure of the models and relatively small number of parameters or coefficients that require identification in a proper manner. On the other hand, a number of cases are thought to employ models of a complex structure with numerous parameters subject to calibration. In principle, more complex models should provide better match of the modeled and observed variables, i.e., higher accuracy. It should be stressed, however, there is no actual rule, since many cases proved relatively simple models to offer high or acceptable accuracy of the results not very different from the complex ones.

Modeling of processes occurring in the natural environment frequently employs multivariate regression methods; however, there exists linear relationships between dependent and independent variables that are found to restrict application possibilities. In response to existing limitations, the required refinements led to MARS (Multivariate Adaptive Regression Spline) determination. The main advantages of MARS lie in its capacity to capture the intrinsic, complicated data mapping in high-dimensional data patterns and produce simpler, easier-to-interpret models, as well as its ability to perform analysis on parameter relative significance. In addition to the complete involvement of the predictors in MARS procedure, there are also all observations of the independent variable being analyzed. Moreover, their variability range becomes divided into intervals where the interval limits are derived from the so-called knots, which, in turn, constitute the threshold values (t) [33]. In this manner, by comparing the values of the independent variable (x_m) with the threshold ones (t_m), the sign of the relationship $y = f(x_i)$ can be determined. The threshold values are utilized to distinguish the values of the independent variable assuming the base functions such as $\beta_i \cdot \max(0, X - t)$. The general form of the MARS model is achieved through summing the base functions and their products as follows:

$$y = \beta_0 + \sum_{m=1}^M \beta_m \cdot H_{k,m} \left(X_{v(k,m)} \right) \quad (2)$$

$$H_{k,m} = \prod_{k=1}^K h_{km} \quad (3)$$

where k is the number of independent variables, M is the number of components of the created model, $m = 1, 2, 3, \dots, M$, and β_m is the weight values.

In order to acknowledge the fact of measurement data falling into clusters, for instance, the procedure for data partitioning was identified through proper rules dependent on the values of independent variables. That finding was successfully implemented in the method of regression and classification trees. However, due to emerging problems with those models' stability, the modification was imposed to replace a single tree with the forest (parallel connected trees—Random Forest) or a serial layout (boosted trees algorithm). The adopted solutions provided better predictive capabilities of the regression tree. Both Random Forest and boosted trees structure require identification of the number of trees, which should not be higher than 300 in practice [34]. On the other hand, an alternative approach to MARS, RF and BT is offered by the neural network method (ANN), in which model structure and the number of estimated coefficients is higher than in the above-mentioned models. Nonetheless, one of the frequently applied ANN models is the multilayer perceptron (MLP) that consists of three layers: an entrance layer, a so-called hidden layer, and the output layer. The neurons of the individual layers are connected by means of proper weights,

which are estimated at the stage of model learning. Because of the significant influence of initial values of the weights on the simulation results, and also problems with determining the global minimum of the analyzed functional relationships, a proper modification of MLP was introduced. If no amendments were introduced, possible limitations to the predictive capabilities of that model may occur, which cause the results of MARS, BF and BT to become comparable or even more precise than the ones achieved by MLP. In the method of SVM (support vectors method) the shortcoming of the neural networks (MLP) was eliminated by replacing the so-called hidden neurons with Kernel functions that facilitated to define the conditions of a global minimum occurrence. The above modification considerably influenced the predictive capabilities of SVM in comparison to MLP. However, in particular the SVM requires the identification of both: the insensitivity threshold (ϵ) and capacity (C) for the coefficients of namely the Kernel core (α_i), stressing that the cases of limited data quantity may affect the optimum values of the coefficients, and finally diminish the predictive capabilities of the model.

The determination of MLP models involved the number of neurons within the range from j to $2j + 1$. In order to optimize the model structure for the adopted number of neurons, there was the need to analyze the proper, so-called activation functions in the hidden and the output layer (assuming particular function types: linear, exponential). Neuron number and activation functions were determined through consecutive iterations, leading to initial model structure. In the MARS method, the recursive algorithm was implemented both for model building and for independent variables selection, while SVM adopted the Gaussian form of Kernel functions, and the values of C, ϵ , and α were obtained through consecutive substitutions method. The GP (genetic programming) models were derived from mathematic operators such as +, −, /, and basically relied on the initial number of individuals $n = 200$, the generations $N = 300$, with the assumed mutation probability $P_m = 0.25$ and the crossing probability P_c at the level of 0.65.

2.4. Deardorff Equation

In order to describe relationships between meteorological parameters and soil moisture, the quasi-physical model was utilized, based on the concept presented by Deardorff [6]. The choice of that methodology was conditioned by simplicity aspects and appropriateness to represent diurnal variation influence on water regime of the shallow soil system. According to original assumptions, the variation of near-surface volumetric soil water content can be approached through mass exchange between the atmosphere and two distinguished soil layers: upper and lower. The first one, namely the surface layer, constitutes the transitional zone of the soil moisture controlled by daily weather variations to a large extent. Its thickness is assumed to be 10 cm. In the second layer, the moisture content changes seem to be visibly lower. The Deardorff model incorporates then only two time-dependent variables: the soil-surficial moisture content, for a first layer, and the bulk moisture content of the second one:

$$\frac{\partial w_1}{\partial t} = -C_1 \cdot \frac{E_g - P}{h_1} - C_2 \cdot \frac{w_1 - w_2}{\tau}, \quad 0 < w_1 \leq w_{max} \quad (4)$$

$$\frac{\partial w_2}{\partial t} = -C_1 \cdot \frac{E_g - P}{h_1} - C_2 \cdot \frac{w_1 - w_2}{\tau}, \quad 0 < w_1 \leq w_{max} \quad (5)$$

where w_1 , w_2 denotes the soil water content (–) at upper and lower layer, respectively, h_1 and h_2 are layer thickness (m), P is precipitation rate (m/s), τ is diurnal time period (1 day expressed in seconds), C_1 , C_2 are dimensionless proportionality coefficients, E_g is the bare-soil evaporation (m/s), μ is the constant used to account for the moisture transport below a lower layer and w_{max} is the soil field capacity. The C_1 and C_2 parameter values rely on the soil moisture and texture as well as the functional relationships reported by Noilhan and Planton [35]. In the Deardorff approach the C_2 parameter is treated as a constant and C_1 becomes a piecewise function of effective saturation ($C_1 = f(w_1/w_{max})$). The present model follows the original Deardorff formulation; however, the authors adopted another

type of function for C_1 . Eventually, its value is derived against effective saturation, but as a power function of a single parameter:

$$C_1 = \left(\frac{w_1}{w_{max}} \right)^m \quad (6)$$

where m denotes the model parameter. That modification was imposed to reduce the number of parameters, as a piecewise, original relationship requires two more. Assuming $m < 0$, the proposed equation preserved the decreasing character of the C_1 coefficient with saturation. The exponential form of that equation also showed better agreement with an experimental set proposed by Noilhan and Planton [35]. With Equation (5), the dependency given with the data points for all three soil classes, sand, loam and clay, was successfully reproduced, and the coefficient of a linear correlation became higher than 0.95.

The soil moisture transport was not involved in the Deardorff model as it was limited to a time period of a few weeks only. Here, to extend the time scale, the sink term was added to Equation (4), that permitted taking account of the moisture transport outside the model's domain. That term was utilized as linear relationship, applied before in the similar studies [36]. The bare-soil evaporation rate (E_g) was linked to the potential evaporation (E_p) with the following Equation (7):

$$E_g = \begin{cases} \left(\frac{w_1}{w_{sat}} \right) \cdot E_p & w_1 < w_s \\ E_p & w_1 \geq w_s \end{cases} \quad (7)$$

where $w_s \approx 0.75 w_{max}$.

The solutions of Equations (4) and (5) were achieved through explicit Runge–Kutta procedure, offered by the ode23 solver as the component of the Matlab computation software [37]. The variables m , C_2 , μ and w_{max} were considered to be model parameters.

2.5. Coupled GSA-GLUE Methodology

The basics of the GSA-GLUE technique lie in the probabilistic identification of parameters according to original GLUE framework. A common practice is to search for one deterministic parameter set that provides the closest match of the model-simulated and measured values. Probabilistic approach facilitates the case of models with a considerable number of parameters since the inverse modeling may lead here to many equivalent solutions. The existence of different parameter combinations, that allow to sufficiently represent the process dynamics, results from lack of reliable data that provide ambiguous parameter values. Therefore, the solution is offered by adopting probabilistic parameter identification. In this case the existence of only one optimum combination of parameter values is dismissed on account of random distribution, which properties are to be determined in the model identification process. Probabilistic description of parameter space in the GLUE method provides also for the description of model uncertainty itself, being attributable to insufficient data number, but also imperfect representation of the real process (e.g., inaccurate data, inadequate mathematical description, simplifying assumptions) [38]. The problem of GLUE parameter estimation was formulated as Bayesian analysis:

$$M(X|w) = \frac{M(X)M(w|X)}{M(w)} \quad (8)$$

where X is the vector of model parameters $[m \ C \ \mu \ w_{max}]$, $M(X)$ is a priori distribution of parameters, w is the vector of measured soil moisture contents, $M(w|X)$ is the reliability function, $M(X|w)$ is resultant distribution a posteriori, and $M(w)$ is the scaling factor (such that $\sum M(w|X) = 1$).

A priori distribution $M(X)$ reflects initial assumptions on parameters' variability. In the case of models of surface runoff either soil moisture content or the structure of the parameters' distribution is not known, but only their range or allowable limits resulting

from their physical interpretation. A constant character of the parameter distribution was adopted for soil moisture content modeling herein, which is a frequent case in such or similar applications [38]. Moreover, GLUE parameter identification embraced the transformation of a priori distribution to a posteriori one through reliability function as conditioning the probability of parameter combination due to the fitting quality of the modeled and measured values. In the analyses herein the following form of the reliability function was applied:

$$M(X|w) = p \cdot \exp\left(\frac{-r}{k \cdot V(r)}\right) \quad (9)$$

where p is scaling parameter, k is a shape factor used to control a variance of function value, and $V(\cdot)$ is the variance r -averaged squared deviation of modeled and simulated values:

$$r = \frac{1}{N} \cdot \sum_{k=1}^N (w_{m,k} - w_{s,k})^2 \quad (10)$$

where $w_{m,k}$ is the measured soil moisture content, and $w_{s,k}$ is the simulated soil moisture content.

According to GLUE procedure assumptions Equation (8) is solved by means of the Monte Carlo method. In the first stage, the parameter sample is created based on the adopted a priori distribution. Then, the model is launched for either combination of parameters, that forms the following vector, $X = [x_1, x_2, \dots, x_i]$, and based on the calculated and measured soil moisture contents the reliability function values (w) and resultant a posteriori distribution of parameters are determined.

The supplementary element to the GLUE method was introduced as GSA (global sensitivity analysis). It facilitated to quantify the impact of parameter variance on the variance of model results, and finally to distinguish those parameters that significantly influence the model uncertainty. GSA belongs to variance-based methods and relies on variance decomposition, as follows:

$$V(Y) = V[E/X_i = x_i^*] + E[V(Y/X_i = x_i^*)] \quad (11)$$

where Y is the model output and $E(Y|X_i = x_i^*)$ and $V(Y|X_i = x_i^*)$ is the model expected value and variance, respectively, for the i -th parameter X_i to x_i^* .

The functions $E(\cdot)$ and $V(\cdot)$ are defined for X_i parameter that takes different values in its variability range. Once a particular sensitive parameter has been revealed, the output variance $E[V(Y|X_i = x_i^*)]$ takes relatively small values for different X_i since the component $V(Y|X_i = x_i^*)$ does not involve the variability of that sensitive parameter. The significance of a parameter, on the other hand, will be manifested through high variance values of $V[E|X_i = x_i^*]$ on the condition that X_i parameter changes along with output expected value. The model output variance is described by the second component of Equation (11) and is related with the direct impact of the X_i parameter variance. That was the base to propose the model sensitivity measure to i -th parameter in GSA procedure, the so-called first order sensitivity indicator (S_i):

$$S_i = \frac{E[V(Y/X_i = x_i^*)]}{V(Y)} \quad (12)$$

Variance decomposition according to Equation (11) can be executed against all other parameters, ($X_{-i} = x_{-i}^*$) than only on X_i , as follows:

$$V(Y) = V[E/X_{-i} = x_{-i}] + E[V(Y/X_{-i} = x_{-i})] \quad (13)$$

In Equation (13), in the same manner as (11), the first component describes part of the model variance, being a direct effect of the parameters X_{-i} on the model output Y . The second component, however, represents the influence of X_i parameter being in interaction

with other parameters (X_{-i}) [38]. On this basis, the total sensitivity index is defined (S_{Ti}) for the model against X_i parameter as follows:

$$S_{Ti} = \frac{E[V(Y/X_i = x_i^*)]}{V(Y)} \quad (14)$$

The general problem of sensitivity indexes determination by Equations (12) and (14) is the necessity to know the variance of the parameters. A solution is offered by the coupled sensitivity analysis and parameter identification in GSA-GLUE. Sensitivity analysis is performed on the index that takes the same form as the reliability function in the GLUE method and enables to analyze the model variance as identified with the uncertainty herein. Variance decomposition is realized by means of Monte Carlo technique with the adoption of proper structure of a sample to determine the values of parameter ensemble $X_i = x_i^*$ and $X_{-i} = x_{-i}$, according to Equations (11) and (13). Most frequently, the properly prepared sample is utilized both for GLUE parameter identification and sensitivity analysis.

In this paper the probabilistic identification of Deardorff model parameters m , C_2 , μ , and w_{max} was accomplished by coupled GLUE-GSA methodology. Constant (a priori) distributions of the parameters were assumed according to literature data [24]. Parameter ranges are given in Table 2, and based on their values as well as assumed distributions the coupled GLUE-GSA analyses were performed aiming at the estimation of first order sensitivity indexes (S_i) as well as total sensitivity (S_{Ti}).

Table 2. Assumed parameter ranges (m , C_2 , μ , w_{max}).

Variables	Unit	Min	Max
m	-	0	2
C_2	-	0	2
μ	-	0	10^{-7}
w_{max}	-	0.24	0.42

Based on the above-mentioned analyses, a multidimensional distribution of the parameters m , C_2 , μ , and w_{max} were determined so as to describe the reliability of the Deardorff model. That model was calibrated on soil moisture content data for the first vegetation season (first period) and validated on the available dataset for the second and third period, so as to test the accuracy of the model in different meteorological conditions.

2.6. Aspects of Soil Moisture Content Estimation

Since soil moisture content may play a role as an indicator in decision-making processes (e.g., catchment management planning, water resources distribution and allocation) the offered accuracy of the relevant, applied mathematical model seems to be crucial. Taking final decisions is directly dependent on the achieved simulation results (y_{mod}).

The measure of the decisions' reliability and their potential effects can be expressed as follows:

$$\begin{cases} Rs(t) = y_{mes}(t) - y_{mod}(t) \\ Rs(t)_d < 0 \\ Rs(t)_g > 0 \end{cases} \quad (15)$$

where y_{mes} is measured soil moisture content, y_{mod} is modeled soil moisture content, and $Rs(t)$ is a function of either potential ecological or economic effects based on discrepancies between the modeled and observed variables (data). The $Rs(t)$ function takes different values depending on meteorological conditions (e.g., the season of the year, vegetation period, drought, etc.).

In exploitation practice two cases may occur, whether the modeled values are higher than measured ones or vice versa: $Rs(t)_d < 0$ and $Rs(t)_g > 0$, respectively. In the first case, the decision-making needs to acknowledge that the managed amount of water is actually lower than the theoretical, calculated one (y_{mod}). This can be a constraint on, e.g., drainage-

irrigation scheduling or other solutions (retention, temporary water damming) in order to maintain a required soil moisture content. As a consequence, the environmental or ecological problems may occur related with the nature of the analyzed catchment and the discussed, relevant management issues. The value of $Rs(t)_g > 0$ indicates additional soil water reserves that were not taken into account in the calculation phase. However, in this case the foreseen water requirements in the catchment can be treated as satisfied. Existing water excess, on the other hand, could be subject to reservoir retention or energy production. Whatever the solution, usually dependent on the existing infrastructure status, the proper maintenance of catchment resources would require modifying Equation (15) as follows:

$$\begin{cases} Rs(t) = y_{mes}(t) - y_{mod}(t, \theta) \\ Rs(t, \theta)_{d,g} \rightarrow \min \\ \theta = f(\varkappa, x_1, x_2, \dots, x_j) \end{cases} \quad (16)$$

where θ is the function for the simulations of the soil moisture content dependent on meteorological conditions (in the situation of a deep groundwater table). Two models were adopted here: a mechanistic one, $\varkappa = 1$, and a statistical one, $\varkappa = 2$, which are chosen to execute the simulation due to prevailing local conditions in such a manner that the available expert knowledge helps to minimize the differences between the model and observations.

At last, the analyses of predictive capabilities of the models were performed in order to reveal the periods of simulation results over- or underestimation, or if the measured data shows considerable mismatch against the modeled variables. In order to achieve that aim the so-called “heat maps” were prepared so as to classify modeling errors on a properly assumed scale.

3. Results and Discussion

3.1. Coupled GLUE-GSA Method

Based on (C_2, m, μ, w_{max}) parameter variability ranges the simulations of their numerical values were performed by means of Monte Carlo procedure in the first instance. Obtained vectors (C_2, m, μ, w_{max}) for the measured values (the values of precipitation and evapotranspiration) were incorporated into the Deardorff equation and then the probabilistic solution of the soil moisture content w_1 and w_2 was achieved. Taking into consideration the measurement data and the results achieved through Equations (11) and (13), the first order (S_i) and total (S_{T_i}) sensitivities were calculated (Figure 3). At the same time, the reliability functions for the consecutive, calibrated parameters of the Deardorff equation were determined and are given in Figure 4.

Taking advantage of a posteriori parameter distributions, the 95% confidence ranges were also determined for soil moisture contents w_1 and w_2 throughout the measurement period (Figure 5).

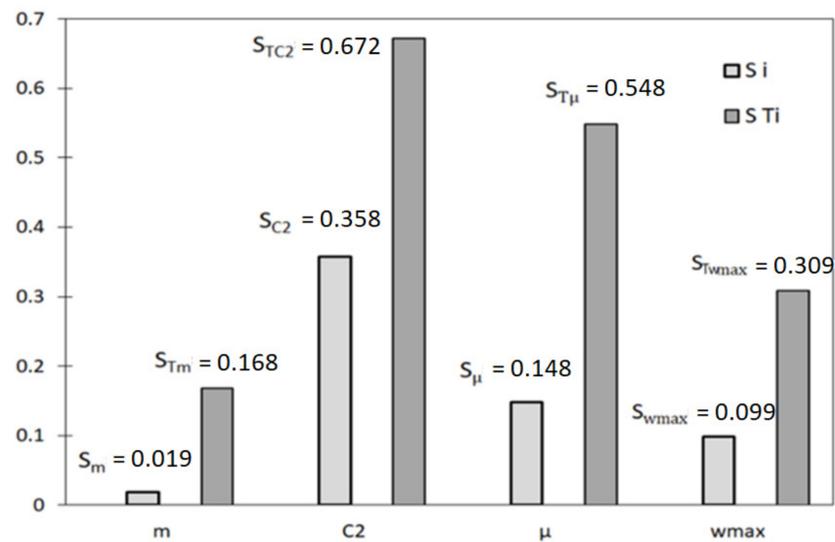


Figure 3. First order (S_i) and total (S_{Ti}) sensitivities for the analyzed parameters (m , C_2 , μ , w_{max}) of Deardorff equation.

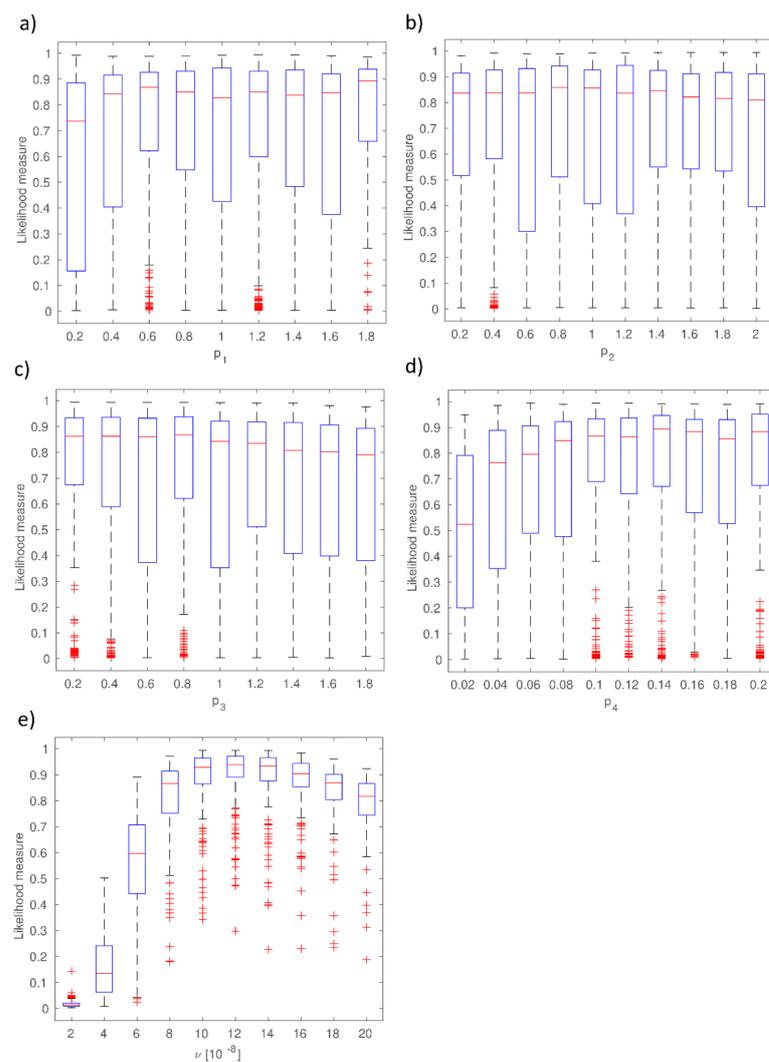


Figure 4. Reliability functions (p_1 , p_2 , p_3 , p_4 , ν) for the calibrated parameters of Deardorff model: (a) reliability of C_2 parameter, (b) reliability of m parameter, (c) reliability of μ parameter, (d) reliability of w_{max} , (e) model reliability function.

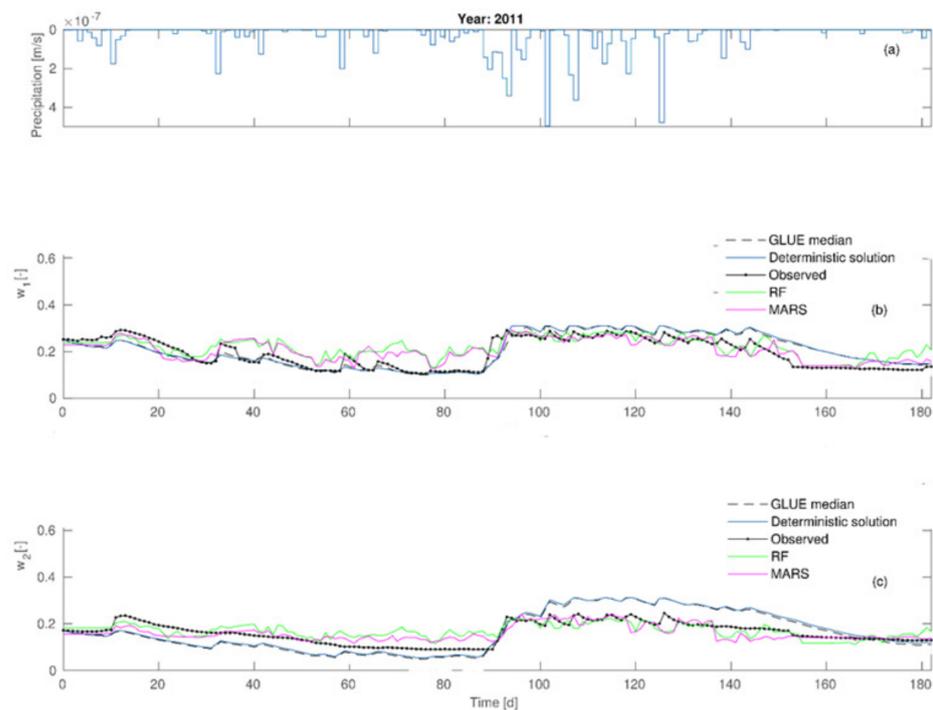


Figure 5. Comparison of the modeled and observed soil moisture contents by Deardorff model for the first calibration stage: (a) precipitation rates (b) w_1 -soil moisture content of the first layer (c) w_2 -soil moisture content of the second layer.

The data given in Figure 3 proves the most considerable influence of the C2 parameter on the agreement between the modeled and observed values. It is made evident by the highest values of the first order sensitivity ($S_{C2} = 0.358$) and the total sensitivity ($S_{TC2} = 0.548$) in comparison to other parameters' sensitivities. Lesser influence on calibration results was shown by μ parameter which is supported by S_{μ} and $S_{T\mu}$ values. Among all the analyzed parameters of the Deardorff model, the minor impact of m parameter was noticed. In this case, the values of S_i and S_{Ti} were equal to 0.019 and 0.168, respectively. Data shown in Figure 4 prove that the increase of p_1 from 0 to 0.2 leads to the average value of reliability function range from 0.75 to 0.85. However, the case of p_2 and p_3 proves that the increase of those parameters' values leads to decline of mean values of the reliability function and, moreover, the increase of p_4 value considerably influences the agreement of the model and observations. It is also a fact that the increase from $p_3 = 0.02$ to $p_3 = 0.08$ caused the reliability function to range from 0.52 to 0.85. A considerable influence of ν parameter on reliability function was also noted, as the increase from $\nu = 2 \cdot 10^{-8}$ to 10^{-7} results in that function value to range from 0.02 to 0.93. Further increase of ν to $2 \cdot 10^{-7}$ causes the decrease of reliability function value to 0.82. The existence of strong interactions between consecutive calibrated parameters (p_1, p_2, p_3, p_4, ν) is indicated by the achieved reliability function ranges. Calibration results of the Deardorff model (soil moisture content w_1, w_2) by the coupled GLUE-GSA method for the analyzed second period are given in Figures 5 and 6.

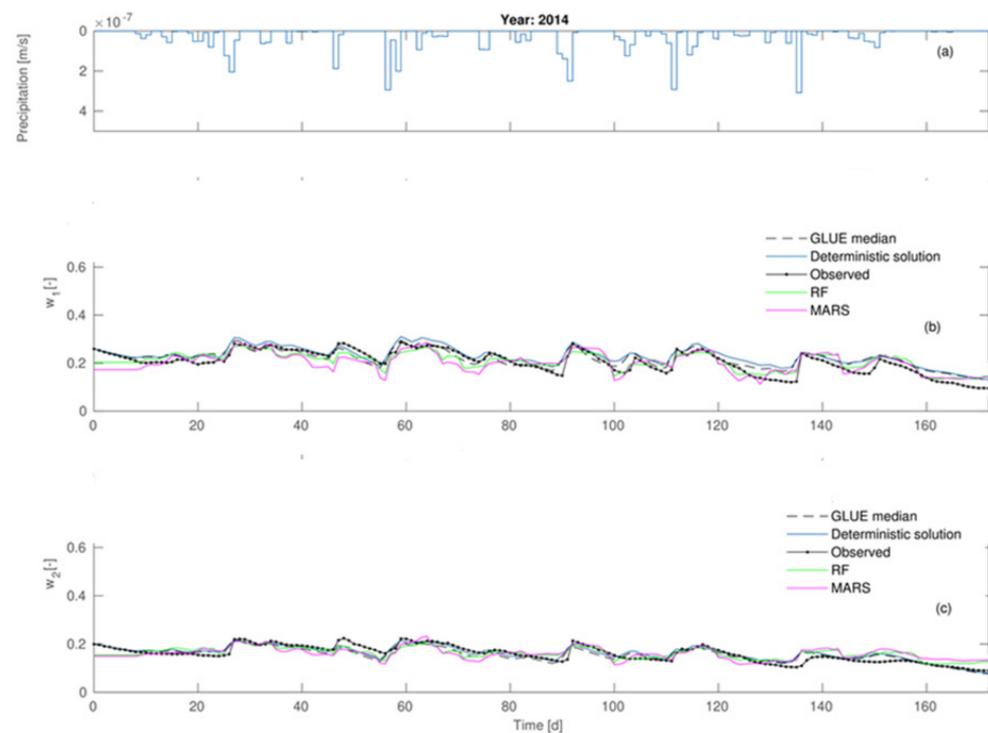


Figure 6. Comparison of the modeled and observed soil moisture contents by Deardorff model for the second calibration stage: (a) precipitation rates (b) w_1 -soil moisture content of the first layer (c) w_2 -soil moisture content of the second layer.

The results show w_1 (average soil moisture content in layer1) to be underestimated by as much as 0.03 maximum in comparison to measured values for the period 0–85 days of the first calibration stage. However, the soil moisture content at 0.20 m (w_2) was underestimated by 0.03 on average for the whole period of 0–85 days, where the daily average precipitation was equal to 1.64 mm and the total amounted to 143.1 mm. The next period, encompassing days 89–120, showed the moisture contents w_1 and w_2 to be overestimated in comparison with the measurements, and involved more intense daily average precipitation of 5.77 mm and a total of 369.3 mm. However, the period between day 120 and 180 exhibited w_1 values to be overestimated by 0.08 maximum, and the observed difference between the model and observations decreased within the period, leading to their final equality. It was also noted that the simulated soil moisture content values in the second layer (w_2) became lower than the observed ones after the 170th day. The second calibration phase (2014) was subject to considerably smaller errors. The achieved results pointed at predictive capabilities of the model to be dependent on dynamic meteorological conditions. This was the reason for the realization of additional soil moisture content simulations (w_1, w_2) aimed at detailed analyses of model predictive capabilities in order to select the optimum calculation method.

3.2. Statistical Models for Soil Moisture Content Forecast

Based on Section 2.3 and the presented relevant principles for model building by statistical methods BT, RF, MLP, MARS, and SVM, the simulations of soil moisture content were realized for both layers (w_1, w_2). The calculations involved numerous combinations of independent variables, describing precipitation totals (P), air temperature (T, T_{sr}), and humidity (Wilg, $Wilg_{sr}$). Tables 3 and 4 present fitting quality (R, MAE, MAPE) of modeling results and observations for those combinations, achieved by BT, RF, MARS, MLP, GP, and SVM models. The lowest error values of moisture content simulations (w_1, w_2) were achieved by MARS model, in the following form:

$$w_1 = 0.362 - 4.133 \cdot \max(0; 4.87 - \text{Psr}) + 0.257 \cdot \max(0; 65.07 - \text{Wilgsr}) + 2.182 \cdot \max(0; \text{Tsr} - 17.33) - 2.66 \cdot \max(0; \text{Psr} - 1.028) + 2.752 \cdot \max(0; \text{Psr} - 3.343) - 3.369 \cdot \max(0; \text{Tsr} - 19.97) - 1.157 \cdot \max(0; \text{Tsr} - 12.70) + 3.995 \cdot \max(0; \text{Tsr} - 21.671) \quad (17)$$

$$w_2 = 2.08 - 0.951 \cdot \max(0; \text{Psr} - 6.91) - 0.771 \cdot \max(0; 6.914 - \text{Psr}) - 0.174 \cdot \max(0; 10.586 - \text{Tsr}) + 0.132 \cdot \max(0; 65.07 - \text{Wilgsr}) + 2.369 \cdot \max(0; \text{Tsr} - 17.328) + 1.064 \cdot \max(0; \text{Psr} - 3.342) - 2.673 \cdot \max(0; \text{Tsr} - 19.38) - 1.148 \cdot \max(0; \text{Tsr} - 13.385) + 2.107 \cdot \max(0; \text{Tsr} - 21.071) \quad (18)$$

Table 3. Comparison of predictive capabilities for MARS, SVM, MLP models.

Parameter	First Layer						Second Layer					
	Learning			Testing			Learning			Testing		
	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE
MARS												
P (t = 1)	0.60	4.00	20.58	0.43	5.10	31.20	0.58	2.79	18.14	0.50	3.21	22.77
P (t = 2)	0.58	3.82	19.95	0.51	4.37	29.04	0.62	2.73	17.95	0.52	3.08	22.05
P (t = 3)	0.60	3.66	19.37	0.51	4.34	26.66	0.60	2.64	17.58	0.55	2.95	21.15
P (t = 4)	0.46	3.52	18.87	0.57	3.92	24.16	0.59	2.64	17.78	0.58	2.78	20.09
P (t = 5)	0.63	3.38	18.05	0.61	3.67	22.63	0.60	2.62	17.59	0.62	2.63	19.10
P (t = 6)	0.66	3.30	17.49	0.65	3.41	21.13	0.62	2.55	17.07	0.65	2.50	18.24
P (t = 7)	0.67	3.20	16.90	0.66	3.31	20.41	0.66	2.48	16.63	0.67	2.39	17.50
P (me)	0.68	3.19	16.90	0.65	3.36	20.57	0.66	2.48	16.63	0.68	2.38	17.45
T (t−1)	0.60	3.89	20.04	0.39	5.45	33.19	0.61	2.69	17.55	0.48	3.31	23.27
Wilg (t−1)	0.60	4.03	20.64	0.38	5.58	34.14	0.62	2.73	17.68	0.47	3.35	23.55
P (me), T _{sr}	0.81	2.48	12.05	0.81	2.50	15.90	0.79	1.93	12.43	0.74	2.15	15.53
P (me), Wilg _{sr}	0.77	3.15	15.58	0.83	2.44	16.05	0.68	2.33	15.04	0.82	1.89	14.68
P (me), Wilg _{sr} , T _{sr}	0.85	2.45	12.03	0.92	2.07	12.22	0.82	1.96	12.62	0.94	1.55	11.28
MLP												
P (t = 1)	0.41	4.94	28.71	0.55	4.30	24.98	0.39	3.02	20.69	0.34	3.51	27.94
P (t = 2)	0.53	4.57	26.63	0.52	4.37	24.59	0.42	3.16	22.94	0.52	3.13	22.23
P (t = 3)	0.60	3.94	23.18	0.61	3.95	22.90	0.50	2.99	22.14	0.58	2.86	20.44
P (t = 4)	0.67	3.46	20.03	0.71	3.64	20.08	0.58	2.39	16.71	0.59	2.86	21.16
P (t = 5)	0.69	3.46	19.83	0.71	3.20	17.93	0.61	2.63	17.72	0.65	2.81	22.20
P (t = 6)	0.76	2.84	16.76	0.72	3.00	17.00	0.66	2.16	15.53	0.63	2.79	20.53
P (t = 7)	0.69	3.20	19.03	0.77	2.92	16.50	0.62	2.66	20.34	0.71	2.19	14.11
P (me)	0.75	2.99	17.33	0.76	2.85	16.39	0.68	2.40	18.02	0.69	2.37	18.10
T (t−1)	0.14	5.38	31.86	0.18	5.38	28.98	0.17	3.33	23.06	0.20	3.40	24.29
Wilg (t−1)	0.21	5.20	30.09	0.29	5.59	33.47	0.28	3.38	23.26	0.17	4.17	30.34
P (me), T _{sr}	0.81	2.54	16.90	0.80	2.71	15.50	0.65	2.56	18.47	0.73	2.37	17.24
P (me), Wilg _{sr}	0.76	2.89	17.17	0.83	2.33	13.18	0.65	2.24	16.33	0.65	2.31	18.07
P (me), Wilg _{sr} , T _{sr}	0.82	2.50	16.18	0.84	2.15	15.50	0.67	2.18	15.72	0.68	2.20	16.00
SVM												
P (t = 1)	0.43	4.57	23.43	0.31	4.75	26.12	0.33	3.11	21.38	0.13	3.08	21.53
P (t = 2)	0.51	4.22	21.62	0.54	4.01	22.08	0.39	2.99	20.87	0.39	2.77	19.65
P (t = 3)	0.56	3.89	20.18	0.68	3.45	19.52	0.45	2.87	20.50	0.52	2.57	18.63
P (t = 4)	0.62	3.62	19.34	0.76	2.92	17.24	0.49	2.78	20.29	0.61	2.37	17.83
P (t = 5)	0.67	3.32	18.01	0.74	3.05	18.02	0.56	2.62	19.34	0.56	2.42	18.21
P (t = 6)	0.71	3.13	17.15	0.73	2.96	17.63	0.61	2.51	18.55	0.59	2.32	17.81
P (t = 7)	0.73	2.98	16.35	0.72	2.99	17.53	0.65	2.41	17.68	0.61	2.29	17.16
P (me)	0.74	2.92	16.35	0.73	2.95	17.05	0.68	2.32	17.15	0.62	2.26	16.56
T (t−1)	0.20	5.14	28.43	0.18	5.74	33.92	0.19	3.26	23.50	0.13	3.28	24.62
Wilg (t−1)	0.19	5.30	30.56	0.18	5.38	33.14	0.07	3.42	22.46	0.02	3.09	21.60
P (me), T _{sr}	0.78	2.86	15.15	0.79	2.56	14.26	0.67	2.34	17.61	0.62	2.33	16.70
P (me), Wilg _{sr}	0.76	2.86	15.79	0.76	2.86	16.11	0.72	2.21	16.47	0.67	2.11	15.31
P (me), Wilg _{sr} , T _{sr}	0.83	2.42	13.74	0.84	2.39	13.47	0.67	2.35	17.46	0.64	2.16	15.85

Table 4. Comparison of predictive capabilities for BT, RF models.

Parameter	First Layer						Second Layer					
	Learning			Testing			Learning			Testing		
	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE
BT												
P (t = 1)	0.49	4.63	27.64	0.33	5.13	27.93	0.47	2.99	21.66	0.27	3.63	23.33
P (t = 2)	0.59	4.09	24.43	0.52	4.43	24.15	0.55	2.83	20.54	0.48	3.32	21.50
P (t = 3)	0.62	3.92	22.85	0.55	3.93	21.91	0.56	2.71	19.17	0.50	2.83	18.62
P (t = 4)	0.69	3.39	19.49	0.65	3.44	18.71	0.63	2.51	18.06	0.62	2.48	16.44
P (t = 5)	0.74	3.08	17.58	0.71	3.00	16.31	0.64	2.48	17.85	0.61	2.56	17.17
P (t = 6)	0.74	3.10	17.83	0.78	2.71	15.34	0.66	2.40	17.05	0.72	2.25	15.26
P (t = 7)	0.78	2.82	16.21	0.79	2.66	15.11	0.67	2.30	16.34	0.75	2.14	14.56
P (me)	0.78	2.85	16.40	0.80	2.67	14.50	0.69	2.24	15.82	0.78	2.01	13.94
T (t−1)	0.22	5.29	30.70	0.26	5.09	28.43	0.42	3.05	21.19	0.31	3.02	19.87
Wilg (t−1)	0.25	5.31	30.10	0.12	5.32	29.43	0.29	3.31	22.39	0.10	3.48	22.29
P (me), T _{sr}	0.81	2.66	15.51	0.80	2.80	15.68	0.71	2.20	15.66	0.72	2.33	15.61
P (me), Wilg _{sr}	0.79	2.79	16.07	0.80	2.81	15.42	0.71	2.23	15.99	0.70	2.25	14.78
P (me), Wilg _{sr} , T _{sr}	0.81	2.68	15.36	0.81	2.72	14.71	0.73	2.15	15.09	0.69	2.20	14.00
RF												
P (t = 1)	0.50	4.66	27.72	0.35	4.89	27.13	0.40	3.16	22.38	0.12	3.62	24.96
P (t = 2)	0.61	4.11	24.25	0.41	4.72	26.23	0.57	2.77	19.43	0.12	3.77	26.27
P (t = 3)	0.63	3.76	21.95	0.59	3.97	22.70	0.58	2.58	18.45	0.38	3.02	21.12
P (t = 4)	0.69	3.37	19.94	0.70	3.39	19.60	0.62	2.47	17.79	0.57	2.61	18.43
P (t = 5)	0.75	2.92	17.38	0.72	3.17	18.15	0.67	2.28	16.66	0.63	2.53	17.95
P (t = 6)	0.77	2.83	17.05	0.75	2.98	16.68	0.71	2.15	15.94	0.59	2.65	18.51
P (t = 7)	0.81	2.54	15.28	0.77	2.86	16.34	0.74	1.98	14.61	0.60	2.64	18.23
P (me)	0.83	2.34	14.03	0.78	2.79	14.94	0.74	2.04	14.59	0.61	2.48	17.07
T (t−1)	0.37	4.83	28.59	0.06	5.38	30.42	0.36	3.03	21.31	0.20	3.38	23.06
Wilg (t−1)	0.32	5.13	29.99	0.18	5.49	30.50	0.32	3.27	22.44	0.06	3.55	24.20
P (me), T _{sr}	0.87	2.72	16.34	0.75	3.30	18.66	0.82	2.06	14.78	0.59	2.67	19.14
P (me), Wilg _{sr}	0.85	2.96	17.66	0.80	3.30	18.89	0.80	2.05	14.92	0.58	2.70	19.50
P (me), Wilg _{sr} , T _{sr}	0.86	2.93	17.40	0.83	2.96	17.80	0.78	2.01	14.50	0.68	2.20	15.20

In the case of BT and RF models, elaborated for w_1 and w_2 forecasting, the obtained number of trees ranged from 40 to 200, which means the models were not over-learned. For the MLP method the number of neurons ranged from 2 to 6, indicating that their maximum value ($2 \cdot j + 1$) was not exceeded. The proper activation functions in the hidden layer and the output layer were most frequently in the form of hyperbolic tangent and linear relationships. Moreover, the models achieved through the SVM method yielded the C parameter range 100–900 and $\alpha = 0.10$ –0.80, respectively.

All in all, the data given in Tables 3 and 4 proved that precipitation, air temperature, and humidity influenced the soil moisture content most significantly. The smallest forecast errors for w_1 and w_2 , taking into consideration only the precipitation totals, were obtained for average daily rainfall on a weekly basis. What is more, the achieved results point at more considerable impact of air temperature than humidity on soil moisture content, which is made evident by the estimated error values (MAE, MAPE). It points at a major influence of meteorological conditions which is also reflected by other research [33]. Among all analyzed methods, the smallest forecast errors of w_1 and w_2 were obtained by the MARS approach. More considerable errors were found for the SVM, BT and RF models. The analyzed combinations of independent variables provided the smallest error values of w_1 and w_2 forecasting for the P_{sr}, Wilg_{sr}, and T_{sr} (weekly averaged values) combination set.

All the above-mentioned models served to find the independent variables, which exhibited the best match of the modeled and observed values. Those variables seem to prove the soil moisture content to be conditioned by the dynamics of precipitation phenomena and meteorological conditions, first of all. Other impacts, such as soil type

and its physical properties, may be found as important and cannot be omitted. Although these other impacts were not deeply considered herein, they deserve more insight in a future application of elaborated models. However, additional validation of these models in different precipitation and meteorological conditions should be treated as the most urgent issue.

3.3. Simulations of Soil Moisture Content by Deardorff and Statistical Models—Heat Maps

Taking into consideration the simulation results by statistical models on one hand, and the physical model results on the other hand, additional calculations of soil moisture content (w_1 , w_2) were realized for vegetation seasons of 2013 and 2014 (calibration stages). This aimed to provide some insight into the models' performance in slightly different development of hydrological and meteorological conditions (e.g., temporarily variable precipitation seasons). The achieved results, given in Figures 7–10, prove that from all the applied methods (Deardorff model, BT, RF, MLP, SVM and MARS method) the best match of the modeled and the observed soil moisture contents was provided by Deardorff equation (physical one, Equations (4) and (5)). For most of the analyzed periods, the outcome of all applied models shows the overestimation of the simulated values in comparison to measurement data. This fact can be treated as a limitation in respect of the utilization of these models for soil moisture control. In realistic exploitation conditions, the soil moisture content would be lower than the mode result, which can lead to water shortages at the stage of decision making.

Based on the performed calculations by statistical models (RF, BT, SVM, MLP, MARS), and also by the Deardorff model, it was found that soil moisture contents (w_1 , w_2) for no-rain periods were overestimated by 10% (SVM, MLP) or even by 15% (BT, RF, MARS, Figures 7 and 8). The same tendency was revealed for 2014 in the second part of the analyzed period—mainly in September. The highest error values, exceeding 5%, were noticed for the BT method. Relatively small errors of w_1 and w_2 forecast were achieved by RF, BT, SVM, MLP, and MARS determined by the choice of independent variables. It was also noticeable that the simulated soil moisture content at the depth of 0.10 m (w_1) in the case of intense precipitation (period between 30th and approximately 90th day) was locally lowered even by 20% (Figure 8) by the applied statistical models: MLP, SVM, MARS, and RF.

Simulation results of the soil moisture content for the depth of 0.20 m (w_2), taking into account the single precipitation events, tended to be lower by about 5 to 10% in the adopted statistical models (Figure 7). In particular, this pertains to the period from the 30th to the 90th day for which the moisture contents were lower by as much as 5 to 10% in comparison with the observations (Figure 8).

Such values of the modeled soil moisture content result from the applied calculation procedures (models) and rely on the selection of the independent variables for the analyzed case study (P_{sr} , $Wilg_{sr}$, and T_{sr}). Only the computations for MLP and MARS for 2013 showed that for single precipitation events the w_2 values (moisture contents for the second layer) became overestimated by not more than 5%. However, for 2014, the period between the 30th and 90th day exhibited visibly lower simulated moisture contents for most of the precipitation events. On the other hand, taking into consideration such precipitation rates that are not of a more intensive character (e.g., the rates of $p < (2-3) \cdot 10^{-7}$ m/s), the simulated values of the soil moisture content were underestimated by no more than 5%. Only in the case of lower precipitation intensity was a tendency of overestimated modeling results observed, especially for RF, MARS and SVM method (Figures 9 and 10). Moreover, the application of elaborated statistical models through RF, BT, MLP, SVM, and MARS (Figures 7 and 9) found that the simulated soil moisture contents w_1 and w_2 (both layers), preceded by a 23-day no-rain period in particular, were underestimated by about 15% maximum for SVM and MLP, 12% by MARS and RF, and 10% by BT. It was also noted, the second layer 0.20-m (w_2) moisture contents were underestimated by as much as 5 to 10% by all the above-mentioned methods, relying on the choice of independent variables. It

was found for the BT method, in particular, that for the period from July to August 2014 the soil moisture content was overestimated by 5 to 10%. Furthermore, the results provided by RF, SVM, MLP and MARS for the analyzed period, preceded by rainfall events of daily intensity ranging from $(0.1-1) \cdot 10^{-7}$ m/s, tended to be overestimated by as much as 5%. In summary, the predictive behavior of all the elaborated statistical models in respect of the soil moisture contents seem to become strictly dependent on the character (particularly the actual values) of the input data (P_{sr} , $Wilg_{sr}$, and T_{sr}).

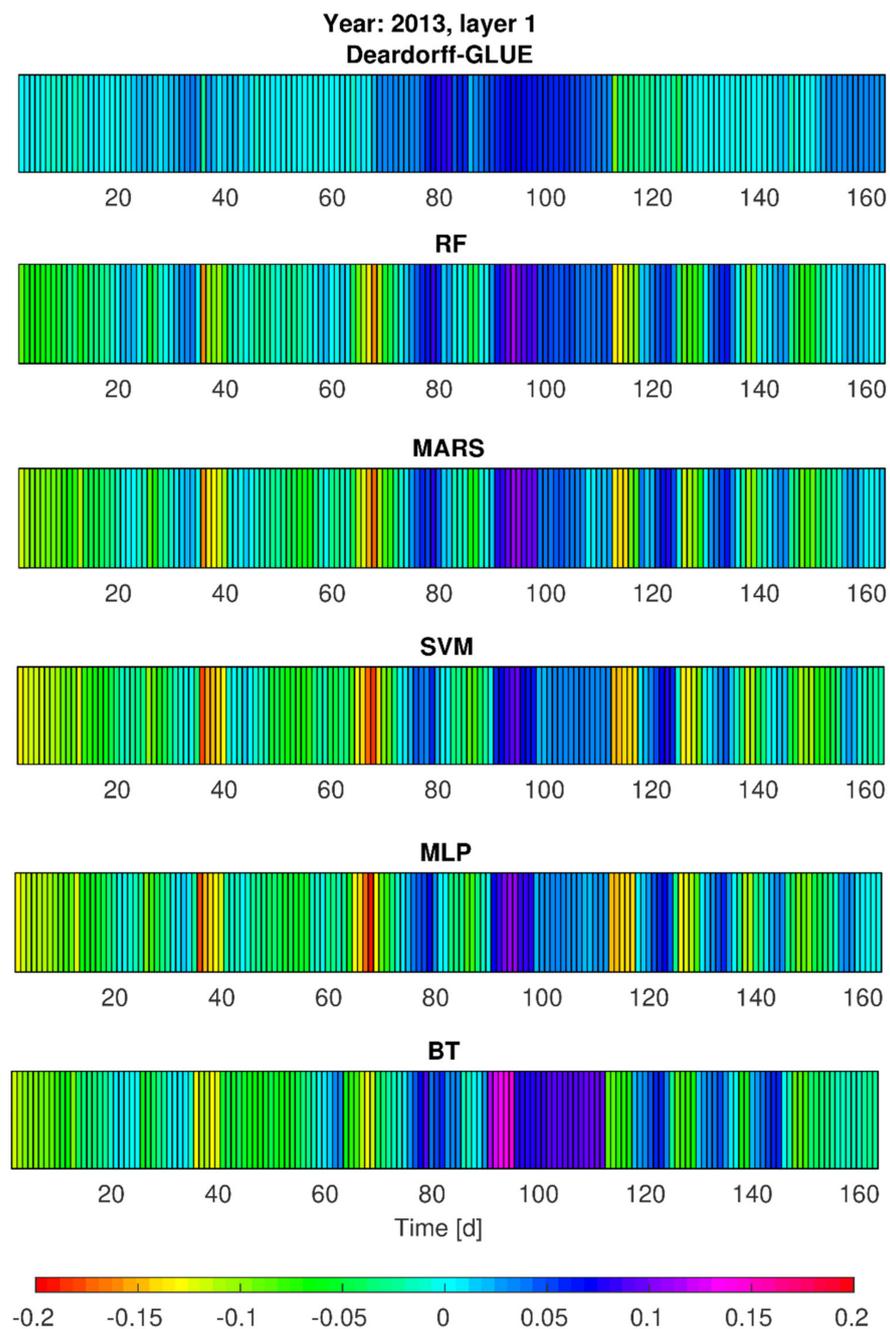


Figure 7. Error variability of soil moisture content forecast (w_1 , first layer) in 2013 by Deardorff model and statistical methods (BT, RF, SVM, MLP, MARS).

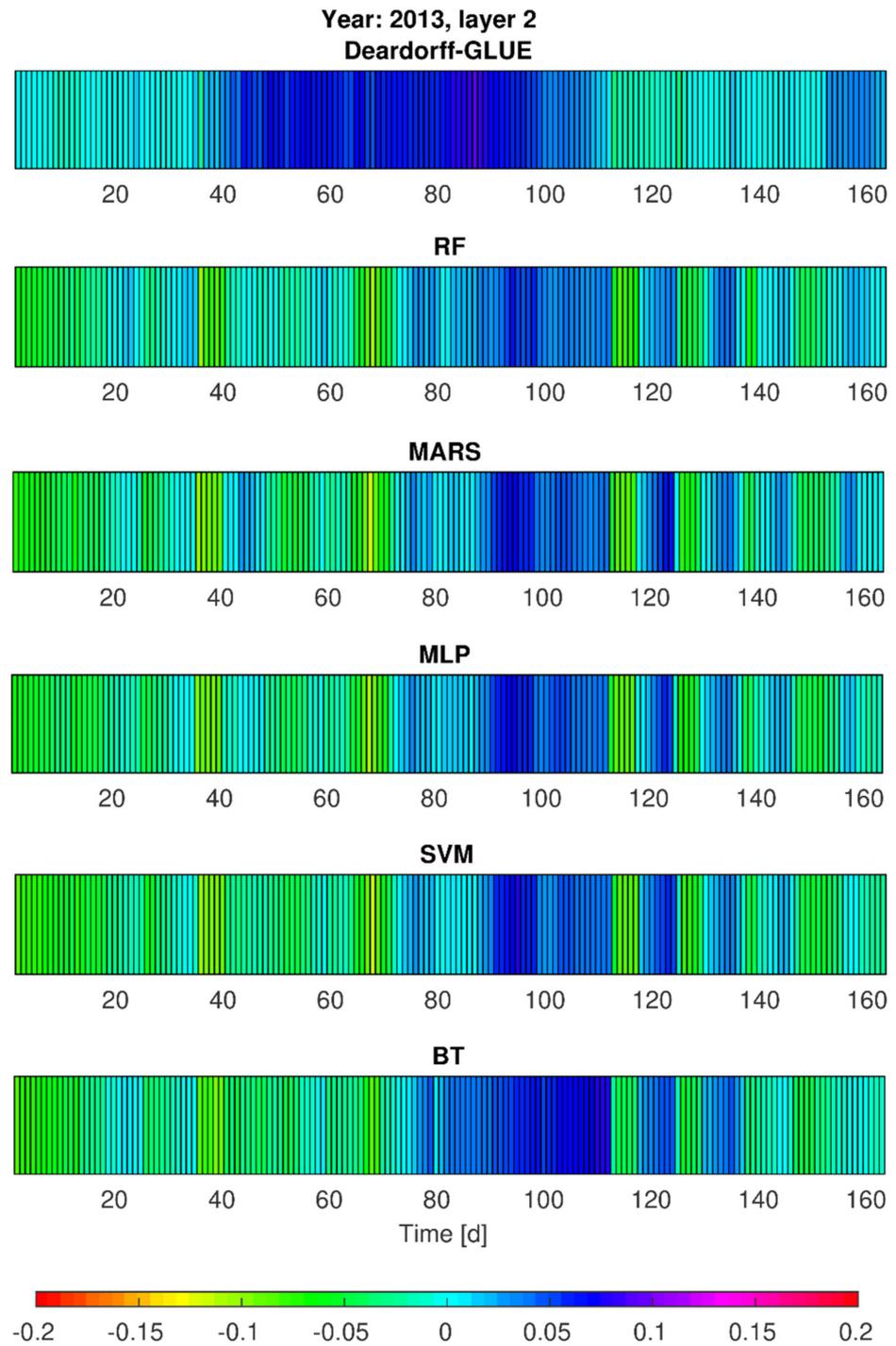


Figure 8. Error variability of soil moisture content forecast w_2 (second layer) by Deardorff model and statistical methods (BT, RF, SVM, MLP, MARS) for 2013.

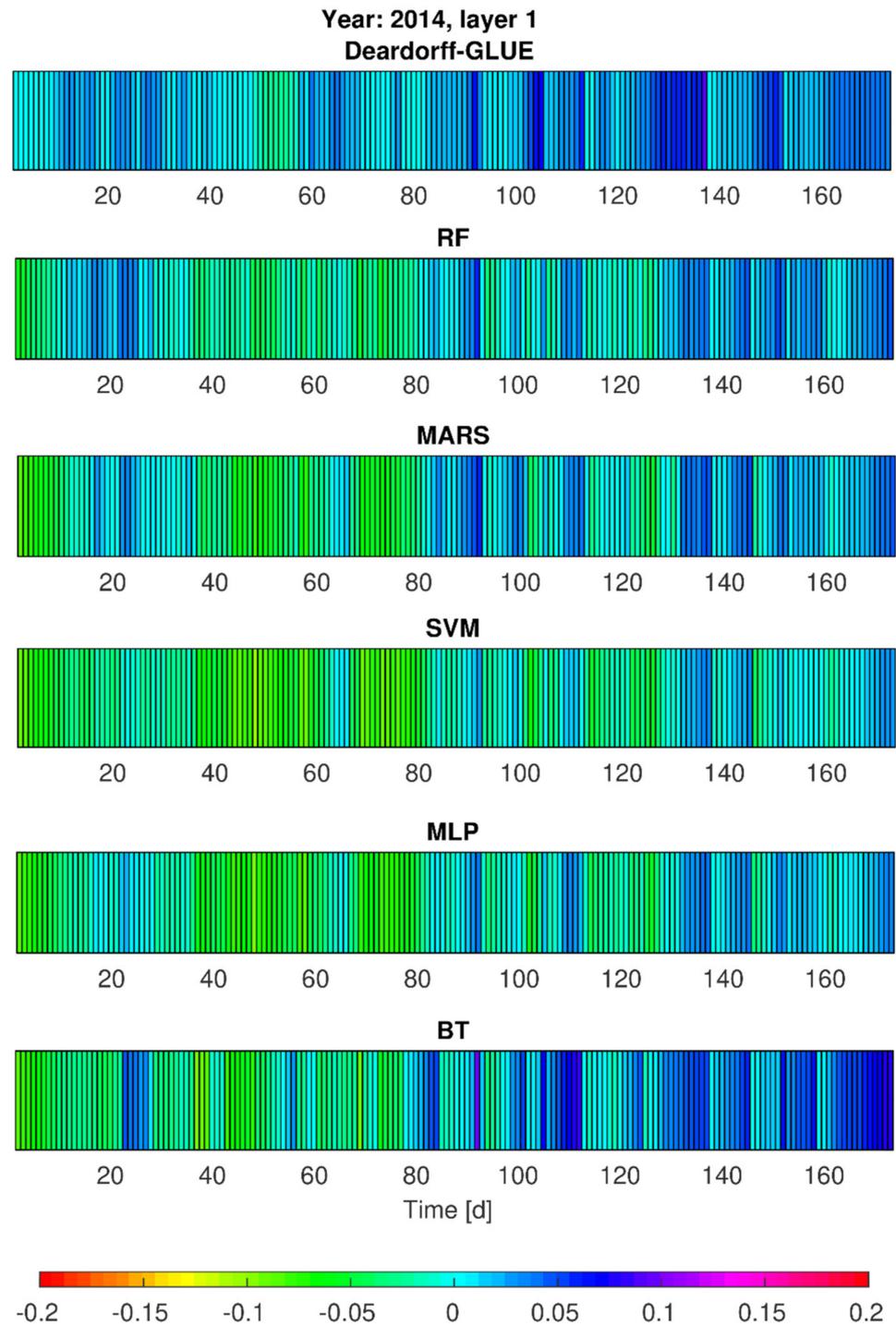


Figure 9. Error variability of the soil moisture content w_1 (first layer) for 2014 by Deardorff model and statistical methods (BT, RF, SVM, MLP, MARS).

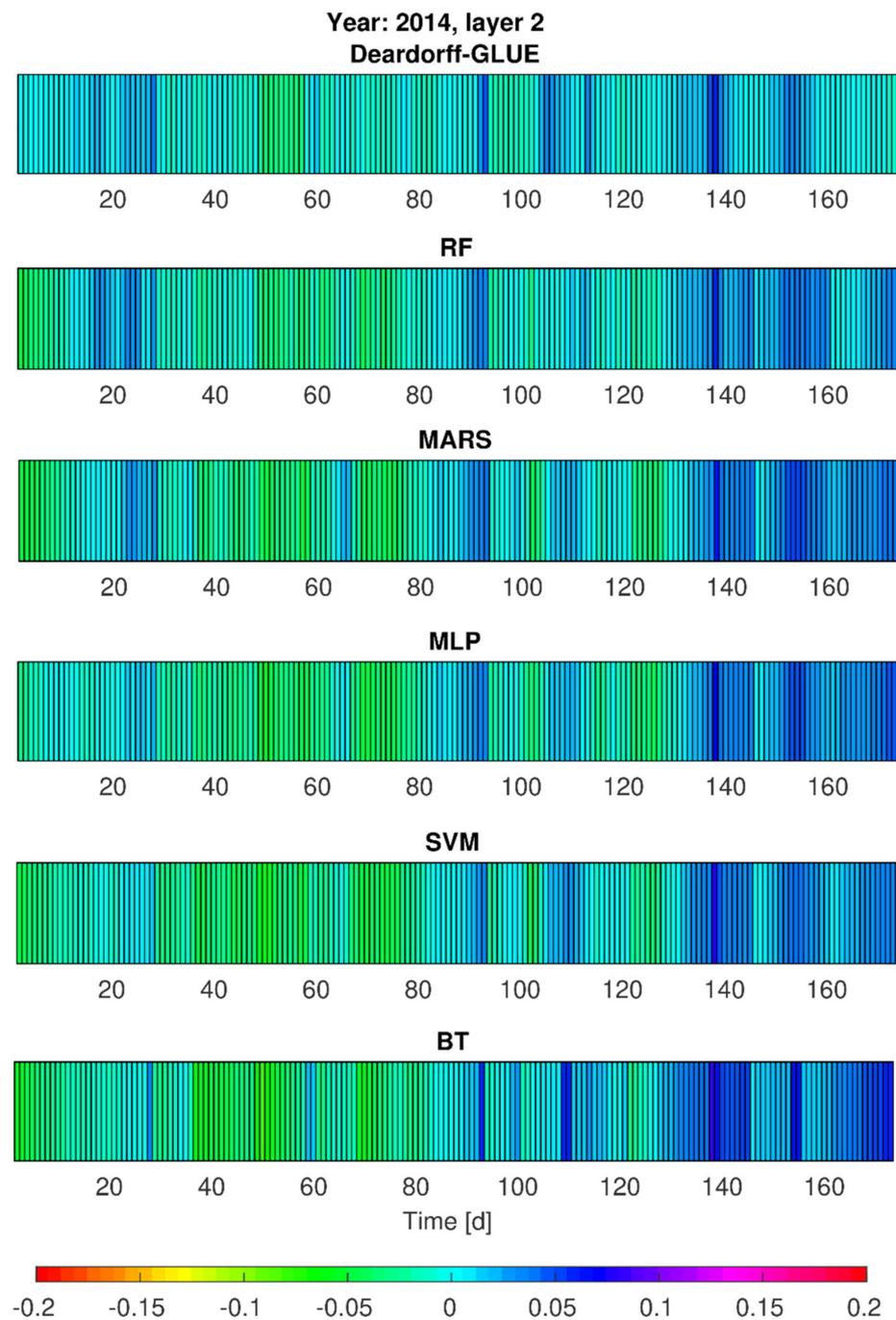


Figure 10. Error variability of the soil moisture content w_2 (second layer) for 2014 by Deardorff model and statistical methods (BT, RF, SVM, MLP, MARS).

4. Summary and Conclusions

The case of physically-based models results in more straightforward estimations of the consecutive, independent variables' influence on the simulation results. It involves particular equations which make it possible to analyze and prove the impact of selected variables on the process dynamics. Statistical models, however, show more difficulties in this respect. All of the models that were approached in this work, as well as the coefficients describing their structure, encountered difficulties in their physical interpretation, but this was found only by this particular study and cannot be generalized. It should be mentioned that statistical models involved the interpolation of soil moisture contents on the basis of the data, which was utilized at the stage of learning and testing. The data used for

calculations may involve identical or similar combinations of the independent variables which, in turn, influence the results of the forecasts. Physical models can avoid such problems. However, taking into account the fact that the Deardorff model overestimated the soil moisture contents, potential problems in decision-making may arise, affecting the determined goals that were suggested by this particular study, such as water reservoirs and hydraulic structures operation. It is based on the assumption that there is more water stored in the soil than there actually is, which can be considered a shortcoming in the management of water resources in a hypothetical area. This may lead to potential negative ecological and economic consequences from the viewpoint of decision-making and planning, therefore the need arises to introduce, for instance, safety factors. The value of that factor, assumed to be lower than one, would reduce the actual amount of water resulting from the simulations.

Nonetheless, the statistical models applied to TDR data (soil moisture contents) proved to have both a simple structure and interpretation character, showing considerable exploitation potential. The physically-based Deardorff model required additional assumptions on the formal relationships between the parameter values. Eventually, the measurement data gathered by TDR technique facilitated the uncertainty analysis on different types of models. Evidence was found for sufficient reproduction of the soil moisture contents for two layers (at 10 cm and 20 cm depth) of a loamy soil at a particular sampling point through a selection of statistical models and a physical one. Even if the proposed methods can be considered as universal, their future development and utilization requires calibration on a range of data coming from similar research sites.

Author Contributions: Conceptualization, A.B. and B.S.; methodology, A.K. (Adam Kiczko) and A.K. (Adam Koziol); validation, A.B., G.M. and J.P.; formal analysis, B.S.; investigation, A.B.; resources, M.K. and G.M.; data curation, A.K. (Adam Kiczko); writing—original draft preparation, A.B.; writing—review and editing, A.K. (Adam Kiczko) and B.S.; visualization, B.S.; supervision, A.B.; project administration, A.B., M.K., A.K. (Adam Kiczko) and A.K. (Adam Koziol). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project “Technological innovations and system of monitoring, forecasting and planning of irrigation and drainage for precise water management on the scale of drainage/irrigation system (INOMEL)” under the BIOSTRATEG3 program, funded by the Polish National Centre for Research and Development. Contract number BIOSTRATEG3/347837/11/NCBR/2017.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This study utilized the meteorological station and TDR probes as the equipment of the Water Center of Warsaw University of Life Sciences.

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

References

1. Albergel, C.; de Rosnay, P.; Gruhier, C.; Muñoz-Sabater, J.; Hasenauer, S.; Isaksen, L.; Kerr, Y.; Wagner, W. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sens. Environ.* **2012**, *118*, 215–226. [[CrossRef](#)]
2. Brocca, L.; Hasenauer, S.; Lacava, T.; Melone, F.; Moramarco, T.; Wagner, W.; Dorigo, W.; Matgend, P.; Martínez-Fernández, J.; Llorens, P.; et al. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sens. Environ.* **2011**, *115*, 3390–3408. [[CrossRef](#)]
3. Ainiwaer, M.; Ding, J.; Kasim, N.; Wang, J.; Wang, J. Regional scale soil moisture content estimation based on multi-source remote sensing parameters. *Int. J. Remote Sens.* **2020**, *41*, 3346–3367. [[CrossRef](#)]
4. Whan, K.; Zscheischler, J.; Orth, R.; Shongwe, M.; Rahimi, M.; Asare, E.O.; Seneviratne, S. Impact of soil moisture on extreme maximum temperatures in Europe. *Weather Clim. Extrem.* **2015**, *9*, 57–67. [[CrossRef](#)]
5. Seneviratne, S.I.; Corti, T.; Davin, E.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.* **2010**, *99*, 125–161. [[CrossRef](#)]
6. Deardorff, J.W. A parameterization of ground-surface moisture content for use in atmospheric prediction models. *J. Appl. Meteorol.* **1977**, *16*, 1182–1185. [[CrossRef](#)]

7. Ban-Weiss, G.; Bala, G.; Cao, L.; Pongratz, J.; Caldeira, K. Climate forcing and response to idealized changes in surface latent and sensible heat. *Environ. Res. Lett.* **2011**, *6*, 034032. [[CrossRef](#)]
8. Lee, J.; Horton, R.; Noborio, K.; Jaynes, D.B. Characterization of preferential flow in undisturbed, structured soil columns using a vertical TDR probe. *J. Contam. Hydrol.* **2001**, *51*, 131–144. [[CrossRef](#)]
9. Yanful, E.K.; Mousavi, S.M.; Yang, M. Modeling and measurement of evaporation in moisture-retaining soil covers. *Adv. Environ. Res.* **2002**, *7*, 783–801. [[CrossRef](#)]
10. Udawatta, R.P.; Anderson, S.H.; Motavalli, P.P.; Garrett, H.E. Calibration of a water content reflectometer and soil water dynamics for an agroforestry practice. *Agrofor. Syst.* **2010**, *82*, 61–75. [[CrossRef](#)]
11. Skierucha, W.; Wilczek, A.; Szyplowska, A.; Sławiński, C.; Lamorski, K. A TDR-Based Soil Moisture Monitoring System with Simultaneous Measurement of Soil Temperature and Electrical Conductivity. *Sensors* **2012**, *12*, 13545–13566. [[CrossRef](#)]
12. Wilczek, A.; Szyplowska, A.; Nosalewicz, A.; Skierucha, W.; Wilczek, W. Project of an automatic system for soil moisture regulation using TDR technique. In Proceedings of the VI International Scientific Symposium Farm Machinery and Processes Management in Sustainable Agriculture, Lublin, Poland, 20–22 November 2013; pp. 197–200.
13. Suchorab, Z.; Jedut, A.; Sobczuk, H. Water content measurement in building barriers and materials using surface TDR probe. In Proceedings of the 16th Central European Conference, Jamrozowa, Polana, 18–20 November 2007.
14. Suchorab, Z.; Widomski, M.K.; Łagód, G.; Barnat-Hunek, D.; Majerek, D. A Noninvasive TDR Sensor to Measure the Moisture Content of Rigid Porous Materials. *Sensors* **2018**, *18*, 3935. [[CrossRef](#)]
15. Suchorab, Z.; Majerek, D.; Kočí, V.; Černý, R. Time Domain Reflectometry flat sensor for non-invasive monitoring of moisture changes in building materials. *Measurement* **2020**, *165*, 108091. [[CrossRef](#)]
16. Topp, G.C.; Davis, J.L.; Annan, A.P. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.* **1980**, *16*, 574–582. [[CrossRef](#)]
17. Alley, R.B.; Emanuel, K.A.; Zhang, F. Advances in weather prediction. *Science* **2019**, *363*, 342–344. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, W.; Zhang, Y.; Goh, A. Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation. *Tunn. Undergr. Space Technol.* **2017**, *64*, 24–33. [[CrossRef](#)]
19. Jeihouni, M.; Alavipanah, S.; Toomanian, A.; Jafarzadeh, A. Digital mapping of soil moisture retention properties using solely satellite-based data and data mining techniques. *J. Hydrol.* **2020**, *585*, 124786. [[CrossRef](#)]
20. Fernández, J.R.A.; Nieto, P.J.G.; Muñoz, C.D.; Antón, J.C. Álvarez Modeling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecol. Eng.* **2014**, *68*, 80–89. [[CrossRef](#)]
21. Mathias, S.; Skaggs, T.; Quinn, S.; Egan, S.; Finch, L.; Oldham, C. A soil moisture accounting procedure with a Richards' equation-based soil texture-dependent parametrization. *Water Resour. Res.* **2015**, *51*, 506–523. [[CrossRef](#)]
22. Brandyk, A.; Kiczko, A.; Majewski, G.; Kleniewska, M.; Krukowski, M. Uncertainty of Deardorff's soil moisture model based on continuous TDR measurements for sandy loam soil. *J. Hydrol. Hydromech.* **2016**, *64*, 23–29. [[CrossRef](#)]
23. Lamorski, K.; Pastuszka, T.; Krzyszczak, J.; Sławiński, C.; Witkowska-Walczak, B. Soil Water Dynamic Modeling Using the Physical and Support Vector Machine Methods. *Vadose Zone J.* **2013**, *12*, 1–12. [[CrossRef](#)]
24. Samaneh, E.; Vahidreza, J.; Majid, M.; Khashei, A.; Bilondi, M.; Pourreza, M. Assessing an efficient hybrid of Monte Carlo technique (GSA-GLUE) in Uncertainty and Sensitivity Analysis of van Genuchten. *Comput. Geosci.* **2021**, *25*, 503–514.
25. Ratto, M.; Tarantola, S.; Saltelli, A. Sensitivity analysis in model calibration: GSA-GLUE approach. *Comput. Phys. Commun.* **2001**, *136*, 212–224. [[CrossRef](#)]
26. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper 56*; FAO: Rome, Italy, 1998.
27. Matusiewicz, W. Foundation walls drying and dewatering of soil adjacent to the Ursyn Niemcewicz Palace. *Sci. Rev. Eng. Env. Sci.* **2013**, *60*, 208–221.
28. Szejba, D.; Gnatowski, T.; Oleszczuk, R. *Soil Physical and Hydraulic Properties Estimation for the Drainage of WULS Sport Facilities*; WULS Expertise: Warsaw, Poland, 2003; p. 15. (In Polish)
29. Ryzak, M.; Bieganowski, A.; Bartmiski, P. Methods for determination of particle size distribution of mineral soils. *Acta Agrophys.* **2009**, *175*, 60–72.
30. Malicki, M.; Skierucha, W. A manually controlled TDR soil moisture meter operating with 300 ps rise-time needle pulse. *Irrig. Sci.* **1989**, *10*, 153. [[CrossRef](#)]
31. Malicki, M.; Plagge, R.; Roth, C. Improving the calibration of dielectric TDR soil moisture determination taking into account the solid soil. *Eur. J. Soil Sci.* **1996**, *47*, 357–366. [[CrossRef](#)]
32. Carranza, C.; Nolet, C.; Peziz, M.; van der Ploeg, M. Root zone soil moisture estimation with Random Forest. *J. Hydrol.* **2021**, *593*, 125840. [[CrossRef](#)]
33. Achieng, K.O. Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. *Comput. Geosci.* **2019**, *133*, 104320. [[CrossRef](#)]
34. Brunetti, G.; Šimůnek, J.; Piro, P. A comprehensive numerical analysis of the hydraulic behavior of a permeable pavement. *J. Hydrol.* **2016**, *540*, 1146–1161. [[CrossRef](#)]
35. Noilhan, J.; Planton, S. A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.* **1989**, *117*, 536–549. [[CrossRef](#)]

-
36. Huang, J.; Dool, H.M.V.D.; Georgarakos, K.P. Analysis of Model-Calculated Soil Moisture over the United States (1931–1993) and Applications to Long-Range Temperature Forecasts. *J. Clim.* **1996**, *9*, 1350–1362. [[CrossRef](#)]
 37. Shampine, L.F.; Reichelt, M.W. The MATLAB ODE Suite. *SIAM J. Sci. Comput.* **1997**, *18*, 1–22. [[CrossRef](#)]
 38. Yan, Y.; Liu, J.; Zhang, J.; Li, X.; Zhao, Y. Quantifying soil hydraulic properties and their uncertainties by modified GLUE method. *Int. Agrophys.* **2017**, *31*, 433–445. [[CrossRef](#)]