

Article

Accuracy and Transferability of Artificial Neural Networks in Predicting in Situ Root-Zone Soil Moisture for Various Regions across the Globe

Roiya Souissi * , Ahmad Al Bitar *  and Mehrez Zribi

CESBIO—Centre d'Etudes Spatiales de la Biosphère, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, 18 Avenue Edouard Belin, bpi 2801, CEDEX 09, 31401 Toulouse, France; mehrez.zribi@cesbio.cnes.fr

* Correspondence: roiya.souissi@cesbio.cnes.fr (R.S.); ahmad.albitar@cesbio.cnes.fr (A.A.B.);

Tel.: +33-561-55-8537 (A.A.B.)

Received: 2 October 2020; Accepted: 3 November 2020; Published: 5 November 2020



Abstract: This paper explores the accuracy in using an artificial neural network (ANN) to estimate root-zone soil moisture (RZSM) at multiple worldwide locations using only in situ surface soil moisture (SSM) as a training dataset. The paper also addresses the transferability of the trained ANN across climatic and soil texture conditions. Data from the International Soil Moisture Network (ISMN) were collected for several networks with variable soil texture and climate classes. Several scaling, feature extraction, and training approaches were tested. An artificial neural network employing rolling averages (ANN_{RAV}) of SSM over 10, 30, and 90 days was developed. The results show that applying a standard scaling (SSCA) to the ANN input features improves the correlation, Nash–Sutcliffe efficiency (NSE), and root mean square error (RMSE) for 52%, 91%, and 87%, respectively, of the tested stations, compared to MinMax scaling (MMSCA). Different training sets are suggested, namely, training on data from all networks, data from one network, or data of all networks excluding one. Based on these trainings, new transferability (TranI) and contribution (ContI) indices are defined. The results show that one network cannot provide the best prediction accuracy if used alone to train the ANN. They also show that the removal of the less contributing networks enhances performance. For example, elimination of the densest network (SCAN) from the training enhances the mean correlation by 20.5% and the mean NSE by 42.5%. This motivates the implementation of a data filtering technique based on the ANN's performance. A median, max, and min correlation of 0.77, 0.96, and 0.65, respectively, are obtained by the model after data filtering. The performances are also analyzed with respect to the covered climatic regions and soil texture, providing insights into the robustness and limitations of the approach, namely, the need for complementary information in highly evaporative regions. In fact, the ANN using only SSM to predict RZSM has low performance when decoupling between the surface and root zones is observed. The application of ANN to obtain spatialized RZSM will require integrating remote sensing-based surface soil moisture in the future.

Keywords: soil moisture; root-zone soil moisture; artificial neural networks; ISMN

1. Introduction

Soil moisture is considered an important land parameter that stimulates interactions between the water and energy cycles, since it controls the partitioning of the mass and energy fluxes between land and the atmosphere [1]. Furthermore, soil moisture is integrated into several hydrological applications relevant to water resource decision-making [2]. Surface soil moisture (SSM) (0–5 cm) and root-zone soil moisture (RZSM) (30 cm–1 m), the two components of this variable, are both of interest. SSM is a key parameter that controls various processes in environmental systems [3] and is an important

driver of water and heat fluxes between land and the atmosphere. However, monitoring and forecast applications such as operational agriculture monitoring and crop yield forecasting [4] rely more on RZSM [5]. In addition, RZSM is of interest for short- and medium-range meteorological modelling and hydrological studies over vegetated areas [6]. Its knowledge is crucial for vegetation restoration, runoff and erosion processes [7], as well as climate change [8].

Soil moisture information can be retrieved through three main sources, namely, in situ measurements [9], model-based estimates, and satellite observations (for instance, the Soil Moisture and Ocean Salinity (SMOS) mission [10], the Soil Moisture Active Passive (SMAP) mission [11], the Advanced Microwave Scanning Radiometer (AMSR) [12], and the Advanced Scatterometer (ASCAT) [13]). In situ measurements are crucial for calibrating and validating the latter two [14,15]. Furthermore, land cover change or climate-related trends in the water cycle can be detected through long-term time series of in situ soil moisture (SM) observations [2]. The International Soil Moisture Network (ISMN) is one of the most exhaustive data hosting facilities, providing in situ soil moisture measurements collected from operational networks around the world [2]. Satellites cannot retrieve RZSM information because of the shallow penetration depth of spaceborne data, which is on the order of a few centimeters [16]. RZSM is nonlinearly related to SSM through different hydrological processes, such as diffusion processes [17]. Various computational techniques can be used to retrieve RZSM estimates based on weather forcing and surface information. Land surface models (LSMs) and agronomical models are among the most widely used methods (Surfex [18], ISBA [19], CLM [20], Aquacrop [21], SAFYE [22], etc.). However, the parameter identification and forcing data going into these models may be subject to errors and potentially lead to inaccurate tracking of the long-term evolution of soil moisture. This drawback advances the need for data assimilation techniques [23]. Nevertheless, land data assimilation systems (LDAS) may also propagate errors. Data-driven methods are suggested to overcome these drawbacks, including artificial neural networks (ANNs). ANNs have been widely used in the field of hydrology since the first hydrological implementation of ANN-based modelling by French et al. [24], and they have been used, among other applications, for soil moisture estimation [25]. ANNs are, first, not affected by the errors induced by a potential misconception of the physical relationships, as they do not require explicit configuration of these relationships [26]. In addition, ANNs require just a one-time calibration to be efficient with less heavy computational costs and provide instant estimations of soil moisture once instrument data are loaded [26]. The aforementioned advantages of ANNs, compared to other methods, explain several attempts to estimate RZSM based on surface information using ANNs. However, few studies have assessed the quality of RZSM estimations on a global scale [27].

The aim of this study is to investigate the ability of an ANN to predict RZSM based solely on in situ SSM information. This paper investigates the accuracy of the predicted RZSM over different soil moisture networks. A methodology is also suggested to determine the contribution of a given network to the global results and the transferability of the predictions across different networks. The different steps to reach this objective consist of (1) assessing the impact of the temporal parametrization of the input SSM, the scaling technique, and the impact of the training/validation/test sets; (2) evaluating the transferability and the contribution of a given network in the training process to determine the limitations; and (3) applying a data filtering technique to remove low-quality data.

2. Materials and Methods

2.1. In Situ Soil Moisture Datasets from ISMN

Several areas of the globe with different soil and climate parameters were considered. The measurements of soil moisture used in this study are provided by eight networks from ISMN. Figure 1 illustrates the location of the selected networks in the study.



Figure 1. The International Soil Moisture Network (ISMN) network distribution (adapted from the ISMN web data portal, scale 1 cm:1000 km).

The selected datasets (346 stations) are presented in Table 1 and fill the following criteria:

- Soil moisture data lie within the temporal range (January 2013–December 2019) to maximize common temporal coverage. Some stations do not have data that cover the whole temporal interval (absence of measurements, gaps generated after quality control) but are still selected as long as they fall into that period. The total number of considered records is 10,054,406 hourly values. The representativeness and size of the training dataset is an important criterion since ANNs are data-driven methods [27].
- A station is selected when soil moisture data are available at a depth of 5 cm for SSM and depths ranging between 30 and 60 cm for RZSM. Stations do not always have the same sensor installation and layout. Some stations have horizontal sensors ($\text{depth}_{\text{from}} = \text{depth}_{\text{to}}$), whereas, for other stations, soil moisture sensors are disposed vertically ($\text{depth}_{\text{from}} \neq \text{depth}_{\text{to}}$). In the latter case, stations that fall into the interval [30, 60 cm] were chosen.
- A station is selected if it has at least 3000 hourly soil moisture values (cf. Sections 2.2.2 and 3.2).

Table 1. Overview of selected ISMN networks.

Network	Country	Number of Selected Stations	Selected RZSM Depth (cm)	SM Sensors	Length of Record (Hourly)
AMMA-CATCH	Benin, Niger	5 (3 in Benin +2 in Niger)	40	CS616	191,997
BIEBRZA-S-1	Poland	3	50	GS-3	11,401
CTP-SMTMN	China	54	40	EC-TM/5TM	716,139
HOBE	Denmark	29	55	Decagon-5TE	819,591
FR-Aqui	France	5	30, 34, 50	ThetaProbe ML2X	200,087
OZNET	Australia	19	30	Hydra Probe-CS616	519,938
SCAN	USA	209	50	Hydraprobe-Sdi-12/Ana	6,777,222
SMOSMANIA	France	22	30	ThetaProbe ML2X	818,031

Hourly values of SSM and RZSM at different depths (Table 1) along with their quality flags were extracted from the ISMN data portal. In addition, static variables such as soil texture, land cover, and climate classification were downloaded for each station.

The selected stations have different soil textures (Figure 2) and different climate classes according to the Köppen–Geiger climate classification (Figure 3). The heterogeneity of clay and sand percentages as well as climate classes will help us infer the potential impact of forgoing this information in the training process.

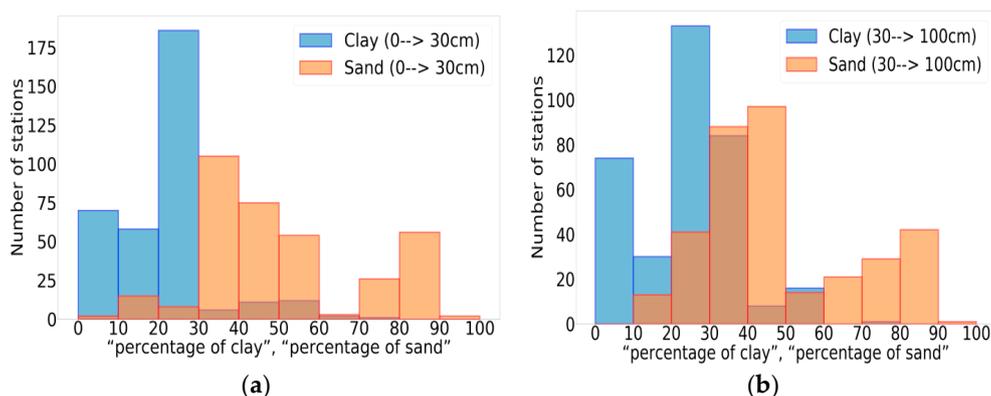


Figure 2. Clay/sand percentages for all of the stations. (a) Clay/sand percentages for the depth interval [0, 30 cm]. (b) Clay/sand percentages for the depth interval [30, 100 cm].

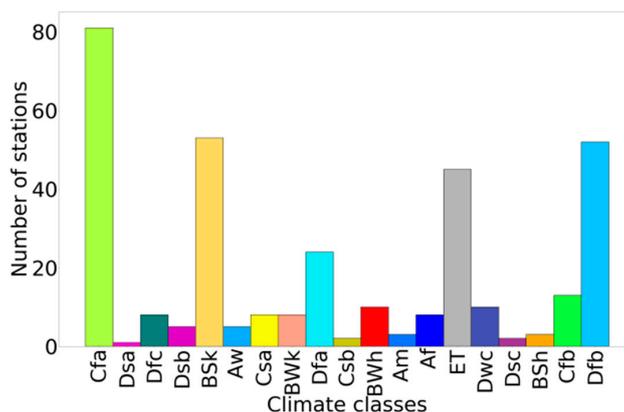


Figure 3. Climate class repartition for the SM stations (the color code is the same as that used in the updated world map of the Köppen–Geiger climate classification [28]).

2.2. Methods

2.2.1. Configuration of the Artificial Neural Network

ANNs can be single or multilayered. The multilayer perceptron (MLP), which is a multilayer feed-forward ANN, is one of the most commonly used ANNs and is considered as the most popular in water resources. A multilayer perceptron is a variant of the original model proposed by Rosenblatt in the 1950s and it has one or more hidden layers between its input and output layers. The neurons are organized in layers such that neurons of the same layer are not interconnected and that the connections are directed from lower to upper layers [29]. Each neuron returns an output based on a weighted sum of all inputs and according to a nonlinear function called the transfer or activation function [30]. The input layer, made up of SSM values, is connected to the hidden layer(s), which is made up of hidden neurons. The final estimates of the ANN are given by an activation function associated with the final layer called the output layer, using a sum of the weighted outputs of the hidden neurons.

Under the assumption of an ANN with one hidden layer, the whole process can be summarized by the following equation:

$$Y = f_2\left(\sum_{j=1}^L w_j f_1\left(\sum_{i=1}^N X_i w_{ij} + b_1\right) + b_2\right) \quad (1)$$

where Y is the output of the ANN and f_1 and f_2 are the activation functions of the hidden layer and the output layer, respectively. w_{ij} and w_j are the weights given to the neurons in the input layer and hidden layer, respectively. b_1 and b_2 are the biases of the input layer and hidden layer, respectively. L and N are the number of hidden neurons and inputs, respectively. Figure 4 includes a simplified diagram of a fully connected ANN with one hidden layer.

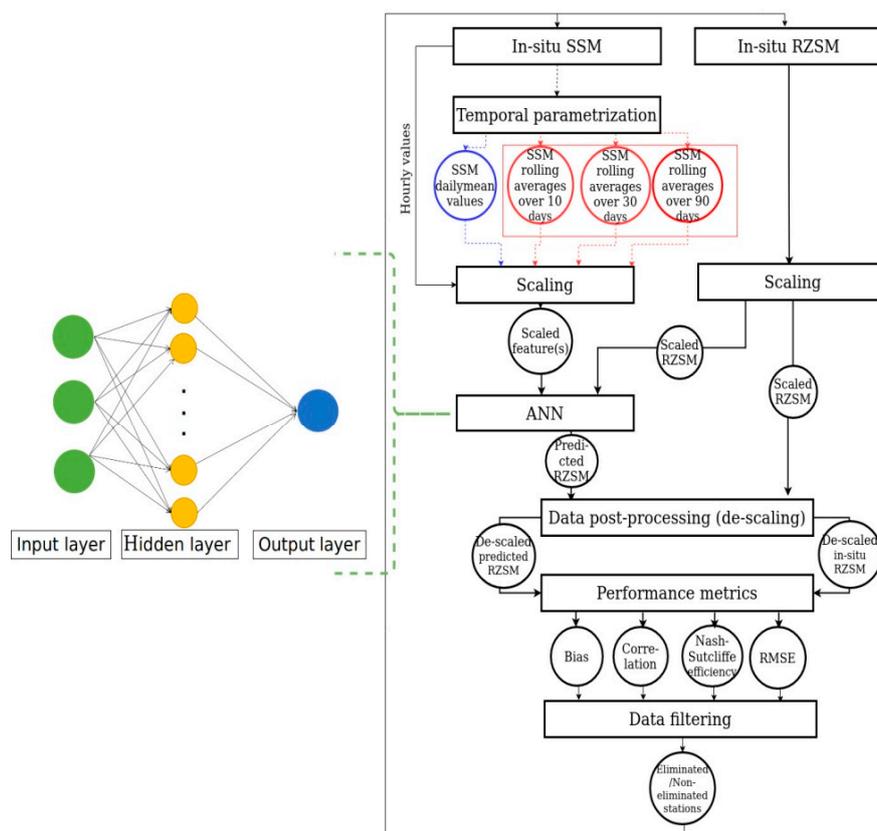


Figure 4. Data processing scheme.

Considering that problems with two hidden layers are rarely encountered and even if the corresponding ANN configuration can represent functions regardless of shape [31], we tested a one- and two-hidden layer ANN architecture. For the number of hidden neurons, a small number leads to underfitting, which may lead to inaccurate detection of complicated signals within the data [32]. In contrast, too many hidden neurons lead not only to overfitting that makes the information contained in the training set insufficient to train all of the hidden neurons but also to a longer training time [32]. Given this information and based on preliminary analysis of the output performances in terms of root mean square error (RMSE) not shown in this paper, an ANN architecture of one hidden layer with 20 hidden neurons was adopted for the remainder of the paper. A tangent sigmoid function was selected as the activation function of the hidden layer due to its anti-symmetry feature, which may accelerate the learning process [27]. A linear function was associated with the output layer. This can be justified by the experiments in [33], where they show that MLPs made up of one input layer, one hidden layer with a nonlinear transfer function, and one output layer with a linear transfer function can

approximate any function with a finite number of discontinuities. A quadratic cost function is used as a loss function, and stochastic gradient descent (SGD) is used as the optimization algorithm.

2.2.2. Features and Scaling

The input and target datasets are preprocessed such that only dates with both SSM and RZSM measurements available are kept. All other dates not filling this condition are dropped. The observed proxy variable (dielectric constant) by in situ instruments is in some cases corrected with soil temperature. Since our objective in these exercises is to test the capacity of SSM to predict RZSM, surface temperature was not considered in the feature construction. Land cover and climate conditions have a high impact on the variability of SSM, and RZSM is mainly linked to SSM through diffusion in porous media and evapotranspiration. These processes present variable specific time scales based on soil properties. Based on that, different temporal configurations were assessed for ANN input features:

- ANN_H: A one-feature ANN such as the feature is the hourly values of SSM.
- ANN_D: A one-feature ANN such as the feature is the daily mean values of SSM.
- ANN_{RAV}: A three-feature ANN such as the three features is the SSM backward rolling average values over 10, 30, and 90 days.

The target dataset (i.e., the RZSM dataset) is truncated for each station; for example, the first value fitted in the neural network is the 2160th available hourly RZSM value (applying the rolling average over 90 days on SSM requires the truncation of input and target data at the 2160th value, which corresponds to 90 days of hourly soil moisture retrievals). The target and input data are then scaled to fall into the same range of values. Non-scaling training was performed, and two scaling methods were tested:

- SSCA (Standard scaling): Standard scaling or Z-score normalization transforms the distribution of a dataset such that the mean and standard deviation of the observations are 0 and 1, respectively, using Equation (2):

$$Z_{\text{norm}} = \frac{x - \bar{x}}{\sigma_x} \quad (2)$$

where Z_{norm} is the normalized data, x is the input, \bar{x} is the mean, and σ_x is the standard deviation of the input data [32].

- MMSCA (MinMax scaling): This scaling scheme constrains the range of each input feature or each output of a neural network. This is usually performed by rescaling the features or outputs from one range of values to a new range of values. Generally, the features are rescaled to lie within a range of 0 to 1 or from -1 to 1 . The rescaling is often accomplished by using a linear interpolation formula such as [34]:

$$x'_i = (\max_{\text{target}} - \min_{\text{target}}) \times \left[\frac{x_i - \min_{\text{value}}}{\max_{\text{value}} - \min_{\text{value}}} \right] + \min_{\text{target}} \quad (3)$$

where x'_i is the scaled data, x_i is the input, \max_{target} and \min_{target} are the new maximum and minimum values, respectively, and \max_{value} and \min_{value} are the original maximum and minimum values of the input data, respectively.

The data are scaled and more precisely standardized before the training step. The result vector leaving the ANN (i.e., the vector of predicted RZSM) is in the standardized format and has to be “de-standardized”. The same goes for the standardized in situ RZSM [32]. Subsequently, performance metrics are computed.

2.2.3. Training and Test Configuration

As mentioned above, one of the objectives of this paper is to assess the genericness of the model. Another point to investigate, at this level, is the training set and assess the impact of its density and its data quality, for instance. For this, different configurations were considered and termed as follows:

- ANN-TOT refers to a training/test approach where 70% of the whole global dataset (70% of the stations of all networks) forms the training set, the remaining 30% of the global dataset consists of a validation set, and the test set is made up of the whole dataset.
- ANN-Net_i refers to a training/test approach where 70% of the values belonging to the stations of a given network (Net_i) form the training set, the remaining 30% of values remaining in Net_i serve as a validation set, and the test set is made up of the whole dataset.
- ANN-(TOT-Net_i) refers to a training/test approach where 70% of the whole global dataset minus the values of a given network (Net_i) form the training set, the remaining 30% of the global dataset minus measurements of Net_i serve as a validation set, and the test set is made up of the whole dataset.

2.2.4. Performance Indicators

Individual Station Performance Metrics

The model is assessed through the following performance metrics: bias, Pearson correlation coefficient, Nash–Sutcliffe efficiency (NSE) (Equation (4)), and RMSE. The final step of the processing (Figure 4) consists of the comparison of the actual values of RZSM with the predicted values and outputting individual performance metrics of each station.

$$NSE = 1 - \frac{\sum_1^N (RZSM_{insitu} - RZSM_{predicted})^2}{\sum_1^N (RZSM_{insitu} - \overline{RZSM_{insitu}})^2} \quad (4)$$

where N is the length of the SM dataset of the considered station.

In addition to the individual performance metrics generated for each station, performance metrics are also generated for all the stations per network.

Skill Indices

Different skill indices are computed to assess the transferability and the contribution of a given network to the training process. First, the performance differences between ANN-TOT and ANN-Net_i are assessed through a coefficient termed TranI_{Net_i} (Transferability Index), which is based on the correlation values yielded by each test network (Net_j) (Equation (5)).

$$\text{TranI}_{\text{Net}_i-\text{Net}_j} = 100 \times \frac{\text{corr}_{\text{ANN-Net}_i}(\text{Net}_j) - \text{corr}_{\text{ANN-TOT}}(\text{Net}_j)}{\text{corr}_{\text{ANN-TOT}}(\text{Net}_j)} \quad (5)$$

Subsequently, the contribution of a given network can be assessed when the performance results yielded by ANN-(TOT-Net_i) are compared with those yielded by ANN-TOT. Consequently, we computed the coefficient ContI_{Net_i} (Contribution Index) (Equation (6)).

$$\text{ContI}_{\text{Net}_i-\text{Net}_j} = 100 \times \frac{\text{corr}_{\text{ANN-(TOT-Net}_i)}(\text{Net}_j) - \text{corr}_{\text{ANN-TOT}}(\text{Net}_j)}{\text{corr}_{\text{ANN-TOT}}(\text{Net}_j)} \quad (6)$$

Both indices are based on correlation values. This choice can be justified by the importance of this indicator, which is often used in the assessment of level agreement between soil moisture products [35]. The correlation is sensitive to both the skill of retrievals with regard to short-term soil

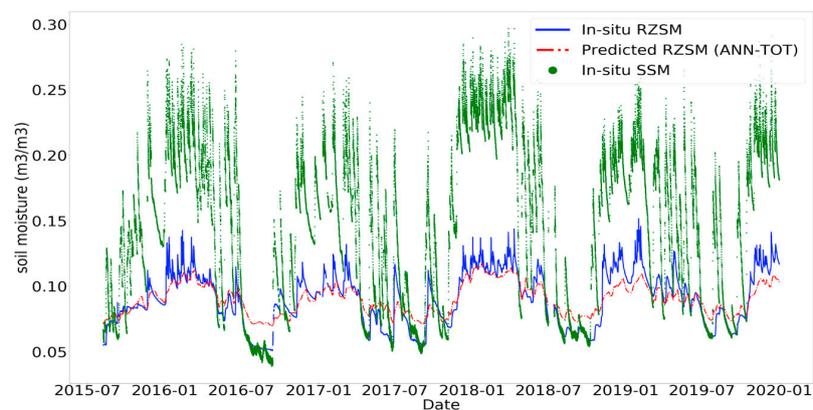
moisture anomalies and their ability to capture typical soil moisture seasonal cycling [36]. Moreover, selection of the SSCA removes the bias.

2.2.5. Data Filtering

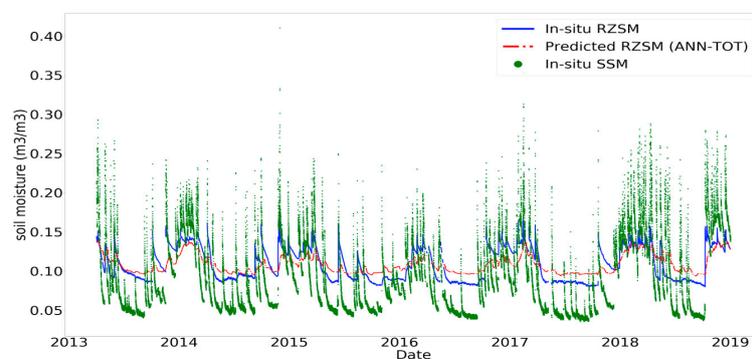
A filtering method was developed to detect underperforming stations and eliminate them from the training dataset. The filtering is based on setting q_{th} quantiles of the correlation values yielded by each test station using the ANN-TOT approach. Once the training/test process is over and the performance metrics for each station are retrieved, a loop runs through the stations one by one and selects those whose correlation is less than the q_{th} quantile of correlation. The training/test process is then reconducted such as the training set is made up of 70% of the non-eliminated stations, the validation set is made up of the remaining 30% of the non-eliminated stations, and the test set is formed by both eliminated and non-eliminated stations. This operation is repeated q times. This new training/test approach is hereafter referred to as ANN-TOT-Qual-Stat (“Qual” represents quality since this method aims to improve the quality of results).

3. Results and Discussion

Figure 5 illustrates the RZSM outputs of the ANN model through two selected examples over French networks. The time series shown below present in situ RZSM in blue, ANN-predicted RZSM (with ANN-TOT) in red, and in situ SSM in green over the stations “Hillan2” (network “FR-Aqui”) and “Lezignan-Corbieres” (network “SMOSMANIA”). We can see that RZSM predictions follow up the evolution of in situ RZSM with almost a positive bias during dry events and a negative bias during wet events. Some fake peaks are sometimes generated after an abrupt increase or decrease in SSM.



(a)



(b)

Figure 5. In situ SSM, in situ RZSM, and predicted RZSM times series. (a) Station “Hillan2” (“FR-Aqui”). (b) Station “Lezignan-Corbieres” (“SMOSMANIA”).

3.1. Impact of Scaling

The three scaling schemes presented in the methods section were tested using the different training approaches (cf. Section 2.2.2). Figure 6 displays the statistical distributions as histogram plots yielded by the three scaling schemes for the training approach ANN-TOT.

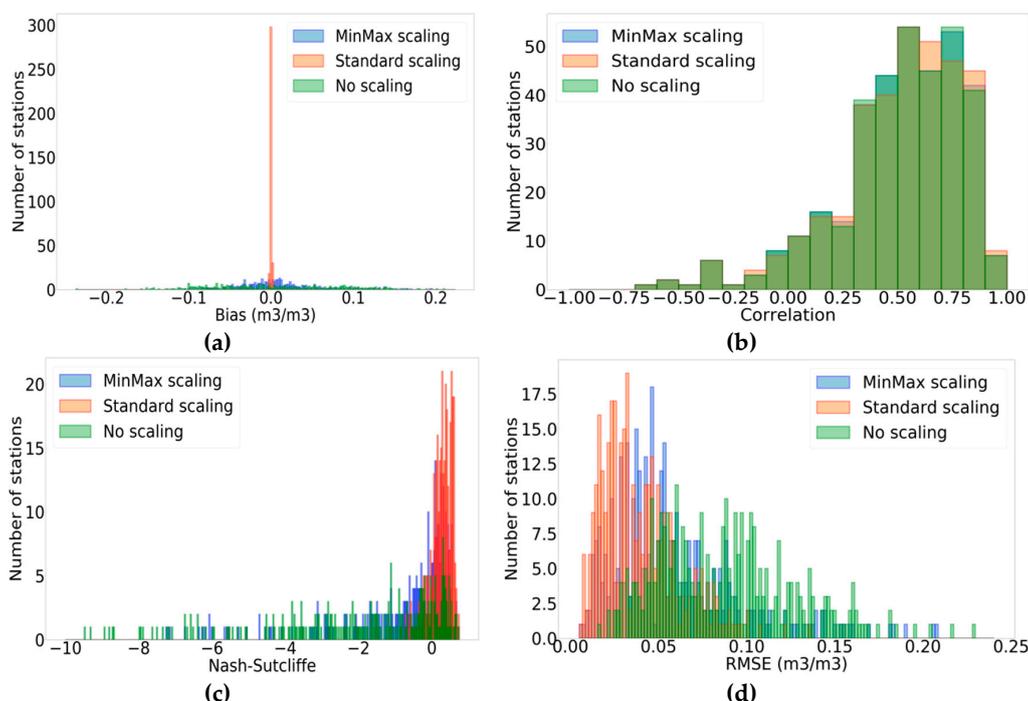


Figure 6. Performance metrics for all of the stations with the different scaling schemes; blue—MMSCA; red—SSCA; green—no scaling (training approach: ANN-TOT). (a) Bias. (b) Correlation. (c) NSE (Nash–Sutcliffe efficiency; NSE values less than 10 were replaced by -10 for better readability). (d) RMSE.

The results highlight the importance of scaling to improve the performance metrics, given the poor performance when the data are unscaled (very negative NSE reaching -285.76 , high RMSE values with an average value of $0.0872 \text{ m}^3/\text{m}^3$). This confirms the statement that the application of preprocessing transformations to the input data is always profitable in practice before presenting data to the neural network [37] and that scaling techniques enhance the reliability of the trained network [38]. The outputs are likewise post-processed to obtain the required output values. It is, then, more relevant to only compare MMSCA with SSCA.

- Bias is considerably reduced with the application of SSCA. This is expected, as the SSCA method by construction tends to eliminate bias. These values ranged between -0.002 and $0.002 \text{ m}^3/\text{m}^3$ for SSCA, whereas MMSCA yielded bias values between -0.105 and $0.196 \text{ m}^3/\text{m}^3$.
- Correlation values are quite similar for the two scaling methods. An insignificant difference of less than 0.001 for correlation values is obtained by MMSCA and SSCA for approximately 60% of the stations (206 stations). Approximately 52% of the stations (181 stations) have higher correlation values with SSCA, approximately 6% of the stations (23 stations) have the same correlation values for both scaling methods, and the remaining stations (142 stations) have higher correlation values with MMSCA.
- RMSE values are also improved with SSCA in comparison with MMSCA mainly due to the enhancement of bias correction. Approximately 87% of the stations (302 stations) show lower RMSE values with SSCA, approximately 7% of the stations (25 stations) have invariable RMSE values, and the remaining stations (19 stations) have better RMSE values with MMSCA. The maximum

decrease (and thus, improvement) in RMSE is recorded for the “Reynolds Homestead” station (“SCAN” network) with SSCA such that the decrease is equal to $0.145 \text{ m}^3/\text{m}^3$. RMSE values yielded by SSCA and no scaling are consistent with previous results advanced in [27] for RZSM estimates at a depth of 50 cm in the case of the “SCAN” network. Actually, the authors in [27] used linear rescaling to compare ANN-simulated soil moisture (generated by SMOS data) to the reference datasets (GLDAS-1/Noah output). The ANN-simulated RZSM values were bias-corrected to match the mean and standard deviation of the reference set. The authors in [27] obtained a mean RMSE of $0.054 \text{ m}^3/\text{m}^3$ following bias correction against a mean RMSE of $0.082 \text{ m}^3/\text{m}^3$ without bias correction. In our case, for the network “SCAN”, SSCA gives a mean RMSE equal to $0.042 \text{ m}^3/\text{m}^3$ against a mean RMSE of $0.090 \text{ m}^3/\text{m}^3$ without scaling. For SSCA, RMSE is equal to the unbiased root mean square error (ubRMSE) since bias is eliminated by construction. In fact, the relation between these two metrics is as follows:

$$\text{RMSE}^2 = \text{ubRMSE}^2 + \text{bias}^2 \quad (7)$$

- NSE values are drastically improved when the SSCA is applied. Approximately 91% of the stations (315 stations) have better NSE values. The best improvements are recorded for stations “PrairieView#1” and “GuilarteForest”, which belong to the network “SCAN”, such as NSE differences (SSCA-MMSCA), which are equal to 86.827 and 85.483, respectively. The difference in behavior between correlation and NSE can be explained by the fact that NSE is a function of RMSE (Equation (8)). Given that RMSE is considerably reduced for most stations with SSCA, NSE is improved.

$$\text{NSE} = 1 - \frac{\text{RMSE}^2}{\sigma_0^2} \quad (8)$$

where the symbol “o” refers to the observation, i.e., in situ RZSM.

While the results in terms of correlation are close, the enhancement in bias correction justifies the choice of SSCA as the scaling method. For this reason, it is adopted for the rest of the paper.

3.2. Impact of the Temporal Information

The three-temporal preprocessing approaches for feature extraction, presented in the methods section, yield close results with slightly better results for the backward rolling average (ANN_{RAV}) (Figure 7). The mean correlation is equal to 0.509, 0.511, and 0.561 with the hourly, daily mean, and rolling average SSM values, respectively. Similarly, the mean NSE is equal to 0.260, 0.263, and 0.325, and the mean RMSE is equal to 0.0392, 0.0391, and $0.0359 \text{ m}^3/\text{m}^3$ with the hourly, daily mean, and rolling average SSM values, respectively. In light of the results above, the backward rolling average approach (ANN_{RAV}) is adopted for the rest of the paper.

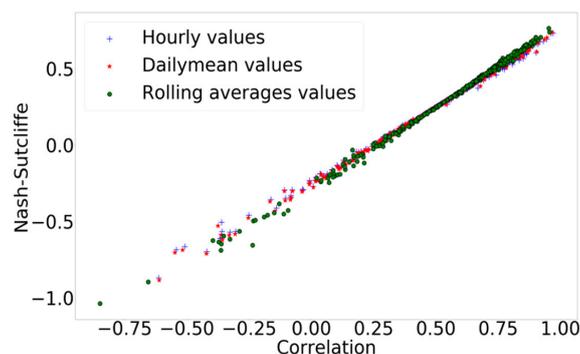


Figure 7. Correlation and NSE scatter plots (training approach: ANN-TOT); blue cross— ANN_{H} ; red star— ANN_{D} ; green circle— ANN_{RAV} .

3.3. Impact of the Training Approach

To assess the transferability of the trained ANN across networks, we suggested the training approach ANN-Net_i, which corresponds to training over one network. Table 2 presents the TranI_{Net_i} values as introduced in Section 2.2.4. Columns indicate the training approach, and rows specify the network on which the test was done. A positive cell value means that ANN-Net_i outperforms ANN-TOT and vice versa for negative values.

Table 2. Transferability index (TranI) for the selected networks.

Training Test	ANN-AMMA-CATCH	ANN-BIEBRZA-S-1	ANN-CTP-SMTMN	ANN-FR-Aqui	ANN-HOBE	ANN-OZNET	ANN-SCAN	ANN-SMOSMANIA
AMMA-CATCH	+1.12%	+0.10%	+0.61%	+0.61%	0%	0%	−1.02%	+0.51%
BIEBRZA-S-1	−0.66%	+3.53%	−2.21%	−0.55%	−0.55%	−3.31%	−1.88%	+0.99%
CTP-SMTMN	−0.88%	−3.62%	+0.77%	−0.33%	+0.33%	+0.11%	−0.99%	−0.21%
FR-Aqui	+0.46%	−3.56%	−1.26%	+2.53%	−1.49%	−3.1%	−2.76%	−2.07%
HOBE	−2.40%	−1.49%	−1.03%	−1.83%	+0.34%	−0.92%	−1.26%	−0.34%
OZNET	−5.03%	−6.42%	−1.51%	−5.28%	−0.50%	+1.26%	−1.89%	−3.02%
SCAN	−1.5%	−1.39%	−1.07%	−1.07%	−0.43%	−0.64%	+0.11%	−1.28%
SMOSMANIA	+0.57%	−1.82%	+0.11%	−0.57%	+1.82%	−1.25%	−3.65%	+3.53%

The first result that can be drawn when comparing ANN-TOT with ANN-Net_i is that the latter gives slightly better results when the test network is Net_i, i.e., the model works better for a given network when the training is solely processed on that network. The positive TranI_{Net_i} coefficients displayed in the diagonal element of Table 2 demonstrate that.

Further observations can be drawn from Table 2. As expected, ANN-Net_i performs worse than ANN-TOT when applied to the networks on which the training has not been performed. The training approach ANN-BIEBRZA-S-1 (i.e., training processed on the BIEBRZA-S-1 network) displays the maximum performance loss compared to ANN-TOT (average loss of −1.83%). Actually, the “BIEBRZA-S-1” network has only three usable stations for our study (i.e., which satisfy the conditions established in Section 2.1), which contain little data (Table 1). The stations of this network have high organic carbon content (39.4%), as provided by ISMN based on the Harmonized World Soil Database v1.1 by IIASA, unlike the rest of the stations where organic carbon content <10%. Besides, the grassland site of “BIEBRZA-S-1” network is located on an intensively mowed, drained meadow with semi-organic soil (muck-peat soil). There, the surface soil layers featured a strong annual cycle with a maximum amplitude of around 60 vol. % [39]. These observations may explain the behavior of the BIEBRZA-S-1 network.

The “OZNET” test network delivered the worst performance compared to the other test networks when the training was run on the other networks. Figure 8 displays the correlation and NSE values for the stations of the “OZNET” network with different training approaches. The behavior of the “OZNET” test network may be explained by the climate specificities of this region of the world, which are characterized by reversed seasons compared to the Northern Hemisphere.

Moreover, some networks are not representative of other networks, i.e., ANN-Net_i performs worse than ANN-TOT for the Net_i test network and vice versa (ANN-Net_i performs worse than ANN-TOT for test network Net_j). If we separately consider the “OZNET” and “FR-Aqui” test networks, we see that ANN-TOT gives better correlation values than ANN-FR-Aqui and ANN-OZNET. Actually, the FR-Aqui network is situated in southwestern France (Figure 1), and its sites cover “the Les Landes” forest of the Bordeaux-Aquitaine region with one additional site (Parcmeteo) in Bordeaux city. The soil texture in the “Les Landes” forest is mainly sandy and characterized by the presence of dark organic matter to a depth of 30 cm. The “OZNET” network lies within the Murrumbidgee River Catchment in Australia. The soil texture in the top layer is predominantly silty loam, loamy sand, and sandy loam. The study area of network “OZNET” covers farms of flood irrigation and dryland cropping (Coleambally Irrigation Area (CIA)) and pastures of grazing.

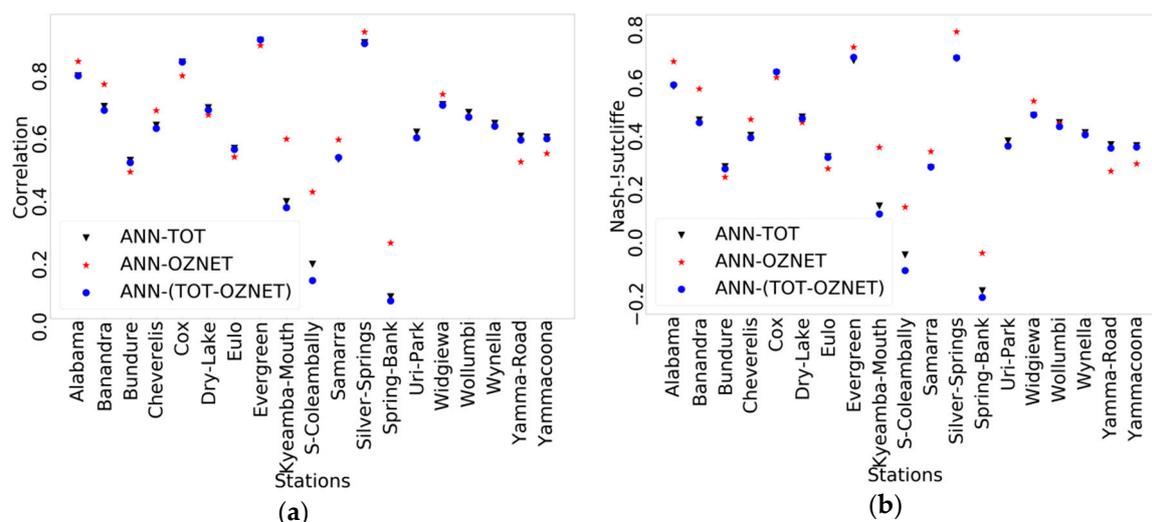


Figure 8. Performance metrics for the stations of the “OZNET” network with the different training approaches. (a) Correlation. (b) NSE.

In this paragraph, the ContI indices obtained from the ANN-TOT and ANN-(TOT-Net_i) setups are presented. The aim of ContI is to help assess the potential influence of a given network Net_i. Table 3 presents the ContI values as introduced in Section 2.2.4. Columns indicate the training approach, and rows specify the network on which the test was performed. A positive cell value indicates that ANN-(TOT-Net_i) outperforms ANN-TOT and vice versa for negative values. The first observation that can be drawn from Table 3 is the positive impact the extraction of the “SCAN” network from the training process would have on all of the test networks except for “OZNET” (loss of −0.13% against ANN-TOT) and “SCAN” (loss of −0.53% against ANN-TOT). This is an interesting case study since “SCAN” is the densest network (Table 1). The negative impact induced by the elimination of the “SCAN” network from the training process on the “OZNET” network can be explained by the climate classification of the stations of both networks. Actually, 7 stations of the “OZNET” network have a common climate class (“cfa”) with approximately 30% of the stations of the “SCAN” network (66 stations). The remaining 12 stations of the “OZNET” network share the climate class (“Bsk”) with approximately 20% of the stations of the “SCAN” network (41 stations).

Table 3. Contribution index (ContI) for the selected networks.

Training Test	ANN-(TOT-AMMA-CATCH)	ANN-(TOT-BIEBRZA-S-1)	ANN-(TOT-CTP-SMTMN)	ANN-(TOT-FR-Aqui)	ANN-(TOT-HOBE)	ANN-(TOT-OZNET)	ANN-(TOT-SCAN)	ANN-(TOT-SMOSMANIA)
AMMA-CATCH	−0.20%	−0.10%	−0.31%	−0.20%	0%	0%	0.92%	0%
BIEBRZA-S-1	−0.44%	−0.44%	−0.66%	−0.22%	−0.44%	−0.33%	−0.33%	−0.11%
CTP-SMTMN	0%	0%	−0.33%	0.11%	0%	0%	0.66%	0.22%
FR-Aqui	−0.46%	−0.35%	−0.46%	−0.58%	−0.12%	−0.12%	1.61%	−0.12%
HOBE	−0.11%	−0.11%	−0.23%	−0.11%	−0.23%	−0.11%	0.34%	0.11%
OZNET	0%	−0.13%	−0.38%	0%	−0.13%	−0.38%	−0.13%	0.25%
SCAN	0%	0%	0.11%	0%	0%	0%	−0.53%	0%
SMOSMANIA	−0.12%	−0.23%	−0.81%	0%	0%	0.12%	2.77%	0.69%

In addition, the aforementioned observation demonstrates the impact of data quality on performance. Although “SCAN” is the densest network, its elimination refines the results (positive ContI values). Figure 9 helps assess the data quality of the “SCAN” network. When considering the training approach ANN-TOT, negative values of NSE and correlation are yielded by approximately 19% (41 stations) and 7% (16 stations) of the stations belonging to “SCAN”. NSE and correlation values less than 0.5 are obtained for approximately 80% (166 stations) and 40% (87 stations) of the stations in the “SCAN” network. Similarly, with the training approach ANN-SCAN, negative values of NSE and correlation are recorded for approximately 18% (38 stations) and 8% (18 stations) of “SCAN”

stations. NSE and correlation values less than 0.5 are given by 85% (179 stations) and 40% (85 stations) of its stations.

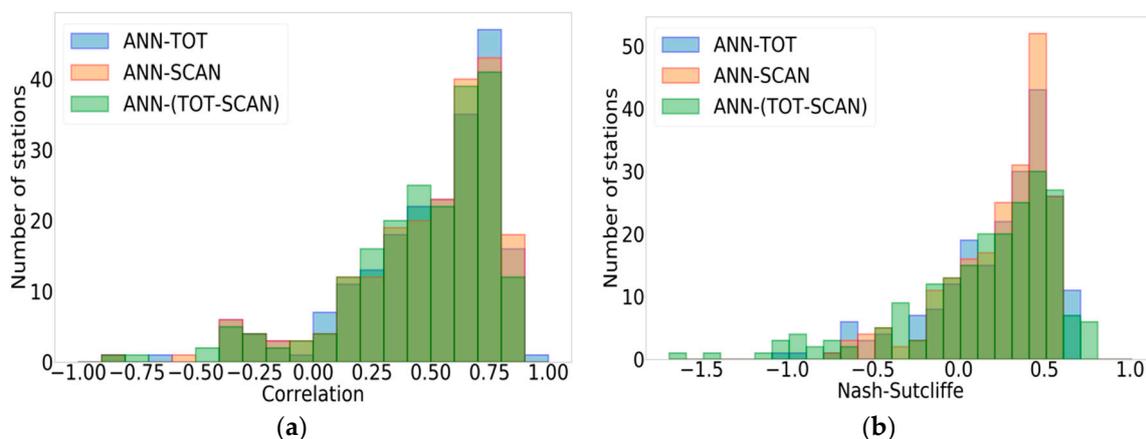


Figure 9. Performance metrics for the “SCAN” test network with ANN-TOT, ANN-SCAN, and ANN-(TOT-SCAN). (a) Correlation. (b) NSE.

Examining “SCAN” stations one by one shows that station “LyeBrook” (2042) gives the lowest NSE and correlation values: -1.037 and -0.849 . A closer look into the soil moisture time series of this station (Figure 10) reveals, on the one hand, data gaps that were well identified in the ISMN quality flag and, on the other hand, constantly low SSM values over a long period of time. Many phenomena may be behind the registration of a constant value over time, such as frost periods and longer sensor dropouts [40]. These constant values lead to an inaccurate prediction of RZSM by the ANN that automatically predicts constant RZSM values in the period overlapping with constant SSM values.

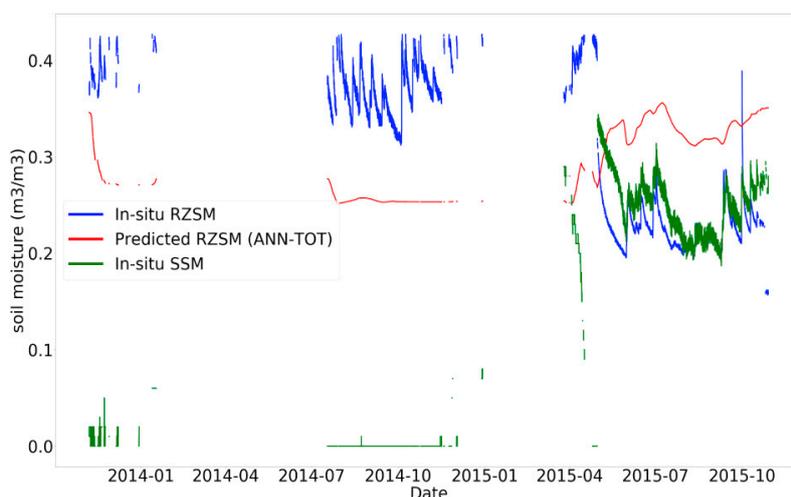


Figure 10. In situ SSM, in situ RZSM, and predicted RZSM of station “LyeBrook” (“SCAN” network).

From another perspective, the “SCAN” test network is not influenced by the elimination of the other networks from the training dataset (0% loss), except for “CTP-SMTMN” and “SCAN”. This can be explained, first, by the density of the “SCAN” network, which represents 67.4% of the whole dataset. This dominant proportion makes the weight of other networks such as “BIEBRZA-S-1” (0.11% of the whole dataset) or “AMMA-CATCH” (1.9% of the whole dataset) not relevant against the density of the “SCAN” network in the training approach ANN-TOT. The elimination of the “CTP-SMTMN” network (7.12% of the whole dataset), i.e., the application of the training approach ANN-(TOT-CTP-SMTMN), leads to worse results compared with ANN-TOT. Actually, the “CTP-SMTMN” network stations are

either located in the “ET” (tundra) (83% of the stations, i.e., 45 stations) or “Dwc” (9 stations) climate classes. Both climate classes are solely covered by this network. This shows the importance of a good sampling of climate classes to perform accurate estimates.

As a conclusion from the results above, the “SCAN” network was removed from both the training and the test datasets (Figure 11). The mean correlation and mean NSE are improved by 20.49% and 42.46%, respectively. Correlation values less than 0.5 are yielded by 31.21% of the stations (108 out of a total of 346 stations) and 15.33% of the stations (21 out of a total of 137 stations) before the elimination of “SCAN” and after the elimination of “SCAN”, respectively. Correlation values greater than 0.7 are recorded for 40.75% of the stations (141 out of a total of 346 stations) and 55.47% of the stations (76 out of a total of 137 stations) before the elimination of “SCAN” and after the elimination of “SCAN”, respectively. Negative NSE values are obtained by 14.45% of the stations (50 out of a total of 346 stations) and 6.57% of the stations (9 out of a total of 137 stations) before the elimination of “SCAN” and after the elimination of “SCAN”, respectively. NSE values greater than 0.5 are recorded for 33.53% of the stations (116 out of a total of 346 stations) and 53.28% of the stations (73 out of a total of 137 stations) before the elimination of “SCAN” and after the elimination of “SCAN”, respectively.

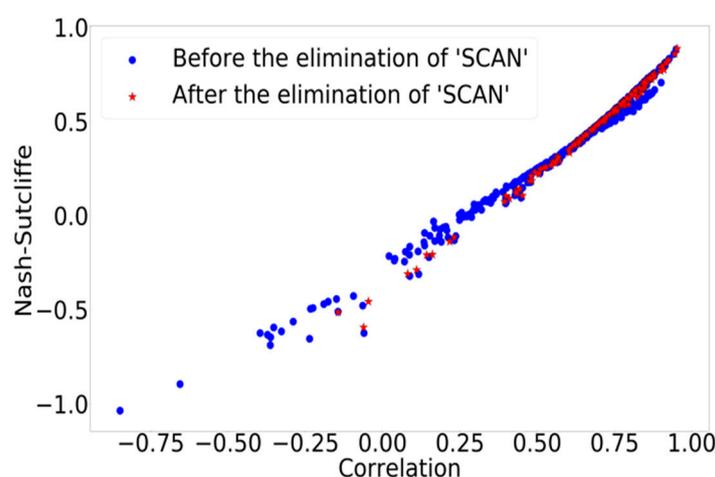


Figure 11. Correlation and NSE scatter plots. Blue circle—before the elimination of the “SCAN” network; red star—after eliminating the “SCAN” network.

3.4. Data Filtering

As previously described in Section 2.2.5, the developed filtering method is intended to identify the underperforming stations and remove them from the training process. The method is a straightforward exclusion using q_{th} quantiles of the correlation vector given by the test stations when ANN-TOT is adopted. Table 4 presents the number of eliminated stations (ES) and non-eliminated stations (NES) in accordance with the value of q .

Table 4. Number of eliminated stations (ES) and non-eliminated stations (NES) based on q_{th} quantiles.

q	Number of ES	Number of NES
0.9	308	38
0.8	275	71
0.75	254	92
0.65	224	122
0.5	170	176
0.4	141	205
0.3	105	241
0.2	71	275
0.1	38	308

After selecting the stations to remove, the training approach ANN-TOT-Qual-Stat is run as described in Section 2.2.5, and performance metrics are yielded for all of the stations based on the value of the q^{th} quantile. Figure 12 shows that the poorest performances (negative NSE and negative correlation values) are recorded for the stations that were eliminated from the training process (regardless of q). Such a result is expected. More importantly, for the non-eliminated stations, q value 0.1 yields better performance metrics than the rest of q values until the level where the correlation is equal to 0.5 and NSE is equal to 0. Beyond that level, q values yield quite similar performance metrics with a slight enhancement for $q = 0.9$ (a maximum correlation of 0.963 against 0.955 with $q = 0.9$ and $q = 0.1$, respectively, and a maximum NSE of 0.922 against 0.809 for $q = 0.9$ and $q = 0.1$, respectively). The maximum correlation value is recorded for the station “Nalohou-Mid” (“AMMA-CATCH” network) with both $q = 0.9$ and $q = 0.1$. The correlation value yielded for the same station before the application of this data filtering technique is equal to 0.856. Similarly, the maximum NSE value is obtained by the station “Nalohou-Mid” with both q values, whereas it is equal to 0.593 before the application of the data filtering method. This station has a tropical savanna climate and is characterized by strong seasonal dynamics that the ANN model manages to capture.

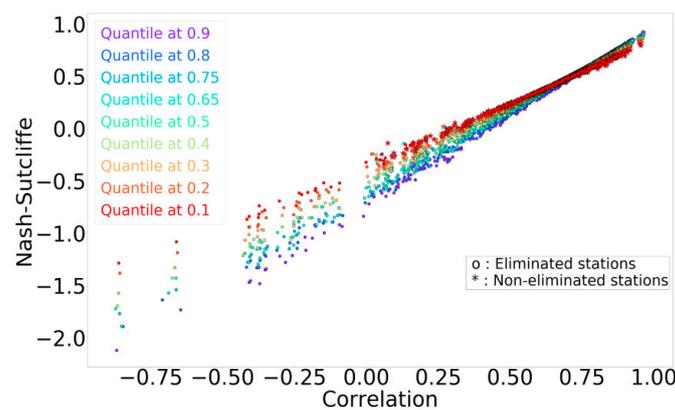


Figure 12. Correlation and NSE values after data filtering.

Table 5 presents the improvement rate in correlation, NSE, and RMSE for the eliminated and non-eliminated stations based on the q^{th} quantile. Of the non-eliminated stations, 60.98% ($q = 0.4$) to 73.68% ($q = 0.9$) show better correlation values. A total of 67.85% ($q = 0.1$) to 100% ($q = 0.9$) of the non-eliminated stations give better NSE values, and 39.94% ($q = 0.1$) to 100% ($q = 0.9$) yield better RMSE.

Table 5. Improvement rates in the individual performance metrics for the eliminated stations (ES) and non-eliminated stations (NES) based on the q^{th} quantiles.

Q	Number of ES	Number of NES	Correlation	NSE	RMSE
0.9	308	38	48.7% of ES	28.57% of ES	34.41% of ES
			73.68% of NES	100% of NES	100% of NES
0.8	275	71	44.72% of ES	26.18% of ES	36.72% of ES
			63.38% of NES	97.18% of NES	97.18% of NES
0.75	254	92	47.24% of ES	24.8% of ES	17.71% of ES
			70.65% of NES	95.65% of NES	88.04% of NES
0.65	224	122	41.07% of ES	19.19% of ES	11.16% of ES
			63.93% of NES	88.53% of NES	78.69% of NES
0.5	170	176	47.06% of ES	14.71% of ES	10.59% of ES
			66.48% of NES	88.07% of NES	73.86% of NES
0.4	141	205	41.13% of ES	14.18% of ES	7.09% of ES
			60.98% of NES	78.05% NES	63.41% of NES
0.3	105	241	39.05% of ES	13.33% of ES	7.62% of ES
			66.39% of NES	78% of NES	60.17% of NES
0.2	71	275	25.35% of ES	11.26% of ES	0% of ES
			60% of NES	73.45% of NES	50.18% of NES
0.1	38	308	23.68% of ES	13.16% of ES	0% of ES
			63.31% of NES	67.85% of NES	39.94% of NES

3.5. Impact of Climate and Soil Texture

To investigate the model's genericness and transferability, the model's predictions are analyzed across climatic regions and soil texture. For this exercise, data filtering is applied with a value of q equal to 0.65, ensuring good screening of underperforming stations while providing good coverage of the climate classes and soil properties. The new training was run on 70% of the previously detected NES (122 stations), 30% of the remaining NES were used for validation, and the test was run on all of the NES. Figure 13a,b present the correlation distribution with respect to the climate classes and the percentage of subsurface clay. Clearly, data filtering with a threshold of 0.65 leads to under-sampling in some cases, namely, for climate classes "Af", "Am", "Bwk", and "Csb" and for the clay fraction interval (30%, 40%), where only one sample was available. Figure 13a provides relevant insights into the impact of climate regions on the results. It is clear that the stations belonging to "Aw", the tropical savanna climate class, yield the best correlation values. This observation can be explained by the strong seasonal dynamics and the presence of wet/dry cycles for the stations of this particular class. Group "B" ("BSk", "BWh", "BWk") includes desert areas where the link between SSM and RZSM is weaker than elsewhere because of the evaporation rates and sporadic rainfall, which reduce the link between observed SSM and RZSM and thus, the performances. This result is consistent with [35], who worked on the assessment of the level of agreement between different LSM products. In fact, they obtained low correlations in the deserts that have, by definition, low mean precipitation and a correspondingly low precipitation variance. Reference [35] confirmed that model agreement should be largest in regions with large variations in precipitation forcing because a larger precipitation variance suggests a larger variation in the moisture storage that all of the models can more easily capture. Moreover, an equilibrium and a regression approach were applied in [41] to establish a relationship between SSM and RZSM. They confirmed that errors in the RZSM estimations are encountered more for the first approach during periods of high surface evaporation or intermittent rainfall and when there is significant evapotranspiration. This shows the limitations of predicting the RZSM from SSM only, and in these specific conditions, it is of interest to include evapotranspiration-related observation variables as input features to the ANN model. Evapotranspiration was identified as a primary variable to predict RZSM in [42]. They showed that an ANN model trained with the dataset of soil moisture profiles generated by the HYDRUS-1D model using meteorological data from the lower Great Lakes region and tested on the same region was sensitive to evapotranspiration because of its role in extracting moisture from the soil. While their results may be dependent on the model's physical assumptions and uncertainty in inputs, in our results using a statistical model with no a priori assumptions, we reach the same conclusions. In Figure 13a, the "C" group, which covers areas of good quality data (mainly "SMOSMANIA" and "FR-Aqui") and is distinguished by dry/wet cycles, yields good performance. These regions are of interest because they hold agricultural areas in Europe, such as the southwest plains in France, where the knowledge of RZSM is of interest for sunflower and maize crops. Mechanistic or physical modeling of the water movement in the soil in the current state of knowledge is governed by the Richards equation. These approaches are very dependent on soil hydrodynamic parameters. Several parsimonious approaches were utilized to counter these drawbacks. Reference [6] used the recursive formulation of the exponential filter [43,44] to retrieve the root-zone soil moisture index (SWI_m) from the in situ SSM of the SMOSREX network in France and the SIM model outputs. The seasonal and interannual variability of SWI_m were also captured after the optimization of the characteristic time length of the filter (T_{opt}). Reference [6] found that over the tested sites, no link could be established between soil texture and the characteristic time length T and highlighted that there is a potential climatic effect that may exist but requires further investigation. The exponential model can be assimilated in fluid mechanics to apply mass conservation equations to an emptying bucket with a transfer function. Alternatively, the ANN does not require assumptions on the model structure (non-linearity is addressed by increasing the complexity of the Neural Network), and because of this, it can be considered more suitable to study its transferability. On the other hand, CDF matching [45] and ANN [27,42] are two statistical methods that have been used to derive RZSM

from SSM. While CDF matching determines the RZSM from SSM by correcting the SSM probability density function to match the observed RZSM, ANNs do not require a priori knowledge of probabilities. As such, they provide a more general framework and the trained ANN model can be applied outside of the training dataset. However, they have some drawbacks as they require a larger dataset than CDF matching to determine the network weighting coefficients. If not available, a risk of overfitting can exist. In the current study, this risk is not present considering the large number of available SSM and RZSM datasets. Nonetheless, as shown in this paper, the results cannot be completely generalized in areas of high evaporation, for instance. Figure 13a also presents the performance for the “Dfa” class, which covers northern areas characterized by harsh cold winters. The presence of frost events may explain a weaker link between SSM and RZSM and thus, weaker correlations. Reference [35] obtained low average correlation values between the different LSM products in high northern latitudes and explained that by the differences in the parameterization of snow and frozen soil for each product. Overall, the performances across climate conditions obtained in our paper are coherent with the results over the continental United States in [27]. In fact, the authors in [27] developed several ANN models to retrieve RZSM at depths of 20 and 50 cm using data from sites located in the continental United States. Each ANN model used a combination of soil texture, SSM, and cumulative values of air temperature, surface soil temperature, rainfall, and snowfall for the input features. Reference [27] confirmed that the retained soil moisture sites could not be considered representative of all soil and climate conditions at a global scale and showed that the ANNs were effective at retrieving RZSM at a depth of 20 cm with a correlation coefficient above 0.7 in most cases, whereas they were less effective at predicting RZSM at 50 cm. This can be explained by surface–subsurface decoupling. Reference [41,46] showed that for a given surface zone depth, the deeper the profile is, the less the correlation between surface and profile soil moisture. Reference [47,48] also confirmed that this surface–subsurface decoupling, controlled by the soil’s hydraulic properties, may occur in coarse-textured and stratified soils as well as dry conditions. Reference [6] also showed that soil depth or thickness is the main factor impacting RZSM retrieval. This exposes a second limitation in addition to the impact of evapotranspiration mentioned above. Figure 13b shows that the low clay fractions present a larger dispersion of correlation in comparison with percentages greater than 30%. In our case, the result can be explained by the small number of stations having such clay percentages. In general, no direct relation between soil texture and model performances can be concluded, which is in agreement with [6].

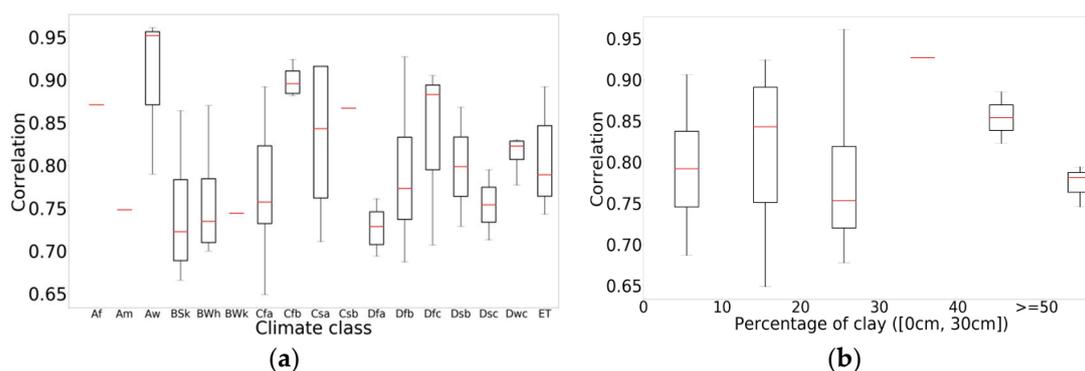


Figure 13. Correlation boxplots after the application of the data filtering technique ($q = 0.65$) with respect to climate classes (a) and subsurface soil clay percentage (b).

4. Conclusions

Throughout this study, we developed an ANN model to estimate RZSM based only on in situ SSM information in several regions across the globe. The main conclusion of the study is that an ANN of 1 hidden layer and 20 hidden neurons can provide accurate predictions of RZSM, provided that a specific ANN configuration is considered. For instance, testing two scaling approaches, we found that SSCA provided the best results, as it minimizes the bias by construction and improves the correlation,

Nash–Sutcliffe, and RMSE when compared with MMSCA. Moreover, a neural network of three features employing rolling averages of SSM over 10, 30, and 90 days is recommended over a single SSM estimate. We assessed the transferability of the trained ANN across observation networks and the contribution of each network to the model learning skills by developing two new indices (TranI and ContI). As expected, the results show that training with data from a single network cannot provide the best predictions. More interestingly, our experiment showed the impact of moderate to low-quality data on the performance of the model through the example of the network “SCAN”, which, while being the densest (67.4% of the whole global dataset), deteriorated the performance of the model. Based on this, we applied a statistical filtering method to eliminate underperforming stations. We analyzed the model performances across climate classes and soil textures. The results showed that the model performs best in regions with alternating wet and dry conditions, while performances were lower over very dry areas with high evaporation rates and sporadic rainfall. This has been depicted by several studies, and the results suggest that to enhance the accuracy over these regions, input features related to the evapotranspiration process need to be added. We did not find relevant results about the impact of soil texture on the model performance, but this should be further investigated with spatial data. We also identified that decoupling between the surface and subsurface deteriorated the predictions of RZSM over 50 cm in depth. In the future, SSM from Earth Observation (EO) will be tested with the current approach to provide spatially distributed RZSM over different climate regions.

Author Contributions: Methodology, A.A.B., R.S. and M.Z.; software, R.S.; writing—original draft preparation, R.S.; writing—review and editing, A.A.B. and M.Z.; supervision, M.Z. and A.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: Souissi’s PhD is co-funded by the ERANET RET_SIF ANR and French National Space Agency CNES Thesis programme.

Acknowledgments: This work was supported by the French National Research Agency (ANR) and the French National Space Agency (CNES). The authors thank the International Soil Moisture Network (ISMN) and the supporting networks for the availability of soil moisture data.

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

References

1. Brocca, L.; Ciabatta, L.; Massari, C.; Camici, S.; Tarpanelli, A. Soil Moisture for Hydrological Applications: Open questions and New Opportunities. *Water* **2017**, *9*, 140. [[CrossRef](#)]
2. Dorigo, W.A.; Wagner, W.; Hohensinn, R.; Hahn, S.; Paulik, C.; Xaver, A.; Gruber, A.; Drusch, M.; Mecklenburg, S.; van Oevelen, P.; et al. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1675–1698. [[CrossRef](#)]
3. Haubrock, S.-N.; Chabrillat, S.; Lemmertz, C.; Kaufmann, H. Surface soil moisture quantification models from reflectance data under field conditions. *Int. J. Remote Sens.* **2008**, *29*, 3–29. [[CrossRef](#)]
4. Mladenova, I.E.; Bolten, J.D.; Crow, W.; Sazib, N.; Reynolds, C. Agricultural Drought Monitoring via the Assimilation of SMAP Soil Moisture Retrievals Into a Global Soil Water Balance Model. *Front. Big Data* **2020**, *3*, 10. [[CrossRef](#)]
5. Manfreda, S.; Brocca, L.; Moramarco, T.; Melone, F.; Sheffield, J. A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 1199–1212. [[CrossRef](#)]
6. Albergel, C.; Rüdiger, C.; Pellarin, T.; Calvet, J.-C.; Fritz, N.; Froissard, F.; Suquia, D.; Petitpa, A.; Pignatelli, B.; Martin, E. From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 1323–1337. [[CrossRef](#)]
7. Gao, X.; Wu, P.; Zhao, X.; Zhang, B.; Wang, J.; Shi, Y. Estimating the spatial means and variability of root-zone soil moisture in gullies using measurements from nearby uplands. *J. Hydrol.* **2013**, *476*, 28–41. [[CrossRef](#)]
8. Han, E.; Merwade, V.; Heathman, G.C. Application of data assimilation with the Root Zone Water Quality Model for soil moisture profile estimation in the upper Cedar Creek, Indiana. *Hydrol. Process.* **2012**, *26*, 1707–1719. [[CrossRef](#)]

9. Lekshmi, S.U.S.; Singh, D.N.; Baghini, M.S. A critical review of soil moisture measurement. *Measurement* **2014**, *54*, 92–105.
10. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.-P.; Delwart, S.; Cabot, F.; Boutin, J.; Escorihuela, M.-J.; Font, J.; Reul, N.; Gruhier, C.; et al. The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle. *Proc. IEEE* **2010**, *98*, 666–687. [[CrossRef](#)]
11. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE* **2010**, *98*, 704–716. [[CrossRef](#)]
12. Owe, M.; de Jeu, R.; Holmes, T. Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res. Earth Surf.* **2008**, *113*. [[CrossRef](#)]
13. Wagner, W.; Hahn, S.; Kidd, R.; Melzer, T.; Bartalis, Z.; Hasenauer, S.; Figa-Saldaña, J.; de Rosnay, P.; Jann, A.; Schneider, S.; et al. The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. *Metz* **2013**, *22*, 5–33. [[CrossRef](#)]
14. Ceballos, A.; Scipal, K.; Wagner, W.; Martínez-Fernández, J. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain. *Hydrol. Process.* **2005**, *19*, 1549–1566. [[CrossRef](#)]
15. Wagner, W.; Naeimi, V.; Scipal, K.; de Jeu, R.; Martínez-Fernández, J. Soil moisture from operational meteorological satellites. *Hydrogeol. J.* **2007**, *15*, 121–131. [[CrossRef](#)]
16. Wagner, W.; Blöschl, G.; Pampaloni, P.; Calvet, J.-C.; Bizzarri, B.; Wigneron, J.-P.; Kerr, Y. Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrol. Res.* **2007**, *38*, 1–20. [[CrossRef](#)]
17. Sabater, J.M.; Jarlan, L.; Calvet, J.-C.; Bouyssel, F.; De Rosnay, P. From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques. *J. Hydrometeor.* **2007**, *8*, 194–206. [[CrossRef](#)]
18. Masson, V.; Le Moigne, P.; Martin, E.; Faroux, S.; Alias, A.; Alkama, R.; Belamari, S.; Barbu, A.; Boone, A.; Bouyssel, F.; et al. The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.* **2013**, *6*, 929–960. [[CrossRef](#)]
19. Noilhan, J.; Mahfouf, J.-F. The ISBA land surface parameterisation scheme. *Glob. Planet. Chang.* **1996**, *13*, 145–159. [[CrossRef](#)]
20. Oleson, W.; Lawrence, M.; Bonan, B.; Flanner, G.; Kluzek, E.; Lawrence, J.; Levis, S.; Swenson, C.; Thornton, E.; Dai, A.; et al. *Technical Description of version 4.0 of the Community Land Model (CLM)*; NCAR: Boulder, CO, USA, 2010. [[CrossRef](#)]
21. Raes, D.; Steduto, P.; Hsiao, T.C.; Fereres, E. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description. *Agron. J.* **2009**, *101*, 438–447. [[CrossRef](#)]
22. Battude, M.; Al Bitar, A.; Brut, A.; Tallec, T.; Huc, M.; Cros, J.; Weber, J.-J.; Lhuissier, L.; Simonneaux, V.; Demarez, V. Modeling water needs and total irrigation depths of maize crop in the south west of France using high spatial and temporal resolution satellite imagery. *Agric. Water Manag.* **2017**, *189*, 123–136. [[CrossRef](#)]
23. Pleim, J.E.; Xiu, A. Development of a Land Surface Model. Part II: Data Assimilation. *J. Appl. Meteor.* **2003**, *42*, 1811–1822. [[CrossRef](#)]
24. Tanty, R.; Desmukh, T.S. MANIT BHOPAL Application of Artificial Neural Network in Hydrology—A Review. *IJERT* **2015**, *V4*, IJERTV4IS060247. [[CrossRef](#)]
25. Elshorbagy, A.; Parasuraman, K. On the relevance of using artificial neural networks for estimating soil moisture content. *J. Hydrol.* **2008**, *362*, 1–18. [[CrossRef](#)]
26. Kolassa, J.; Reichle, R.H.; Liu, Q.; Alemohammad, S.H.; Gentine, P.; Aida, K.; Asanuma, J.; Bircher, S.; Caldwell, T.; Colliander, A.; et al. Estimating surface soil moisture from SMAP observations using a Neural Network technique. *Remote Sens. Environ.* **2018**, *204*, 43–59. [[CrossRef](#)] [[PubMed](#)]
27. Pan, X.; Kornelsen, K.C.; Coulibaly, P. Estimating Root Zone Soil Moisture at Continental Scale Using Neural Networks. *J. Am. Water Resour. Assoc.* **2017**, *53*, 220–237. [[CrossRef](#)]
28. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [[CrossRef](#)]
29. Ramchoun, H.; Amine, M.; Idrissi, J.; Ghanou, Y.; Ettaouil, M. Multilayer Perceptron: Architecture Optimization and Training. *IJIMAI* **2016**, *4*, 26. [[CrossRef](#)]
30. Oyeboode, O.; Stretch, D. Neural network modeling of hydrological systems: A review of implementation techniques. *Nat. Resour. Modeling* **2019**, *32*, e12189. [[CrossRef](#)]

31. Heaton, J. *Introduction to Neural Networks with Java*; Heaton Research, Inc.: St. Louis, MO, USA, 2008; ISBN 9781604390087.
32. Chai, S.-S.; Walker, J.; Makarynsky, O.; Kuhn, M.; Veenendaal, B.; West, G. Use of Soil Moisture Variability in Artificial Neural Network Retrieval of Soil Moisture. *Remote Sens.* **2009**, *2*, 166–190. [[CrossRef](#)]
33. Yonaba, H.; Ancil, F.; Fortin, V. Comparing Sigmoid Transfer Functions for Neural Network Multistep Ahead Streamflow Forecasting. *J. Hydrol. Eng.* **2010**, *15*, 275–283. [[CrossRef](#)]
34. Bishop, C.M.; Bishop, P. *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, UK, 1995; ISBN 9780198538646.
35. Koster, R.D.; Guo, Z.; Yang, R.; Dirmeyer, P.A.; Mitchell, K.; Puma, M.J. On the Nature of Soil Moisture in Land Surface Models. *J. Clim.* **2009**, *22*, 4322–4335. [[CrossRef](#)]
36. Crow, W.T.; Miralles, D.G.; Cosh, M.H. A Quasi-Global Evaluation System for Satellite-Based Surface Soil Moisture Retrievals. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2516–2527. [[CrossRef](#)]
37. Priddy, K.L.; Keller, P.E. *Artificial Neural Networks: An Introduction*; SPIE Press: Bellingham, WA, USA, 2005; ISBN 9780819459879.
38. Jayalakshmi, T.; Santhakumaran, A. Statistical Normalization and Back Propagation for Classification. *IJCTE* **2011**, 89–93. [[CrossRef](#)]
39. Dabrowska-Zielinska, K.; Musial, J.; Malinska, A.; Budzynska, M.; Gurdak, R.; Kiryla, W.; Bartold, M.; Grzybowski, P. Soil Moisture in the Biebrza Wetlands Retrieved from Sentinel-1 Imagery. *Remote Sens.* **2018**, *10*, 1979. [[CrossRef](#)]
40. Dorigo, W.A.; Xaver, A.; Vreugdenhil, M.; Gruber, A.; Hegyiová, A.; Sanchis-Dufau, A.D.; Zamojski, D.; Cordes, C.; Wagner, W.; Drusch, M. Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. *Vadose Zone J.* **2013**, *12*, vzj2012.0097. [[CrossRef](#)]
41. Jackson, T.J.; Hawley, M.E.; O'Neill, P.E. Preplanting Soil Moisture Using Passive Microwave Sensors1. *Jawra J. Am. Water Resour. Assoc.* **1987**, *23*, 11–19. [[CrossRef](#)]
42. Kornelsen, K.C.; Coulibaly, P. Root-zone soil moisture estimation using data-driven methods. *Water Resour. Res.* **2014**, *50*, 2946–2962. [[CrossRef](#)]
43. Wagner, W.; Lemoine, G.; Rott, H. A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. *Remote Sens. Environ.* **1999**, *70*, 191–207. [[CrossRef](#)]
44. Stroud, P.D. *A Recursive Exponential Filter For Time-Sensitive Data*; Los Alamos National Laboratory: Los Alamos, Mexico, 1999.
45. Gao, X.; Zhao, X.; Brocca, L.; Pan, D.; Wu, P. Testing of observation operators designed to estimate profile soil moisture from surface measurements. *Hydrol. Process.* **2019**, *33*, 575–584. [[CrossRef](#)]
46. Arya, L.M.; Richter, J.C.; Paris, J.F. Estimating profile water storage from surface zone soil moisture measurements under bare field conditions. *Water Resour. Res.* **1983**, *19*, 403–412. [[CrossRef](#)]
47. Walker, J.P.; Willgoose, G.R.; Kalma, J.D. Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application. *Water Resour. Res.* **2002**, *38*, 37-1–37-13. [[CrossRef](#)]
48. Hirschi, M.; Mueller, B.; Dorigo, W.; Seneviratne, S.I. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. *Remote Sens. Environ.* **2014**, *154*, 246–252. [[CrossRef](#)]

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



© 2020 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/>).