Catena 207 (2021) 105631

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Soil moisture estimation in two layers for a small watershed with neural network models: Assessment of the main factors that affect the results

Guilherme Kruger Bartels [a,*], Nilza Maria dos Reis Castro [a], Olavo Pedrollo [a], Gilberto Loguercio Collares [b]

[a] *Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul (Institute of Hydraulic Research, Federal University of Rio Grande do Sul -IPH/UFRGS), Av. Bento Gonçalves, 9500, Porto Alegre - RS, Brazil*

[b] *Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas (Center for Technological Development, Federal University of Pelotas - CDTec/UFPel), Rua Gomes Carneiro 01, Pelotas - RS, Brazil*

ARTICLE INFO

Keywords:
Surface and subsurface layer
Soil physical properties
Climatic variables
Topographic variables
Rainfall-related variables

ABSTRACT

Soil moisture, which impacts various hydrological processes, can be estimated from point measurements in a watershed or via remote sensing. As both methods are expensive and complex, efforts have been made to develop empirical models based on data that affect the occurrence of soil moisture. However, these models have been based only on surface-layer data. We present an original approach for investigating regional empirical models of soil moisture, both for the surface (0–10 cm) and subsurface (10–20 cm) layers, and evaluate the main factors which affect the model results. Based on data about the climate, soil properties, topographic features and rainfall, we applied artificial neural network soil-moisture models for a small watershed, in southern Brazil. The models for each layer, with all selected variables, showed Nash-Suitcliffe coefficients of 0.870 and 0.893, respectively, for the surface and subsurface models. We then tested the effects of removing each variable or categories of variables. The most important variables for the surface model were the season, followed by exponential weighted moving average (EWMA) of rainfall. For the subsurface model, the most important variables were the season (although less so than for the surface model), followed by microporosity. All of the variable categories were important in the surface model. In the subsurface model, the soil-related variables were the most important, whereas the rainfall and topography variables were of little importance. It was possible to estimate soil moisture for both layers with good performance. The subsurface model, which used only the soil- and climate-related variables, explained more of the variance in soil moisture than the other models. The subsurface layer is easier to model, because the variation in moisture that is induced by recent climate and precipitation effects is attenuated by the physical features of the soil which control water infiltration.

1. Introduction

Soil moisture is characterized as a key hydrological and ecological variable of the Earth's surface systems (Gao et al., 2013). Above all, it governs the exchange of energy and water between the surface of the earth and the atmosphere, controlling hydrological, meteorological, and ecological processes (Cho and Choi, 2014; Pan et al., 2017). Including soil moisture in hydrological models significantly reduces their uncertainty, enhancing watershed flood prediction (e.g. Alvarez-garreton et al., 2014; Berthet et al., 2009; Massari et al., 2014; Meng et al., 2017; Tayfur et al., 2014; Wooldridge et al., 2003;Zhong et al., 2019). Hydrological models for watershed soil moisture estimation have many potential applications.

However, soil moisture presents high spatial and temporal variability (Gao et al., 2013; Huang et al., 2016; Suo et al., 2018; Zhu et al., 2014), produced by the effects of climate, soil, topography and vegetation

Abbreviations: ANNs, artificial neural networks; NS, Nash-Sutcliffe coefficient; SWC, soil water content; TDRs, Time Domain Reflectometry; SMOS, Soil Moisture and Oceans Salinity; SMAP, Soil Moisture Active Passive; SPI, Standardized Precipitation Index; SPEI, Standardized Precipitation Evapotranspiration Index; TWI, topographic wetness index; EMBRAPA, Brazilian Corporation of Agricultural Research; ETo, reference evapotranspiration; EWMA, exponential weighted moving average; MLP, multilayer perceptron; *r*, correlation coefficient; MAE, mean absolute error; RMSE, root mean squared error.

* Corresponding author.
E-mail addresses: guilhermebartels@gmail.com (G.K. Bartels), nilza@iph.ufrgs.br (N.M.R. Castro), pedrollo.olavo@gmail.com (O. Pedrollo), gilbertocollares@gmail.com (G.L. Collares).

https://doi.org/10.1016/j.catena.2021.105631
Received 20 March 2020; Received in revised form 25 June 2021; Accepted 27 July 2021
Available online 5 August 2021
0341-8162/

(Hagen et al., 2020; Li et al., 2013). Soil properties such as density, organic matter content, texture, structure and macroporosity impact water retention and transport in the soil column (Famiglietti et al., 1998). Features linked to the land cover, root system and plant litter layer impact hydrological aspects such as seepage, surface runoff, interception and evapotranspiration (Jacobs et al., 2004). Topography plays a key role in determining the spatial distribution of soil moisture. In watersheds, steep parts tend to be dryer than flat parts, due to their lower infiltration rates, faster subsurface drainage, and higher surface runoff (Famiglietti et al., 1998). Further, concave surface areas are more humid than convex areas, due to the accumulation of surface and subsurface lateral flow in the surrounding area (Rosenbaum et al., 2012; Zhu et al., 2014). Aspect (hillslope orientation) affects the sun's angle of incidence, and consequently impacts water evaporation (Famiglietti et al., 1998; Moore et al., 1988), altering soil moisture. Grayson et al. (1997) analysed spatial patterns of water distribution during humid periods (when rainfall exceeds evapotranspiration) and dry periods (when evapotranspiration exceeds rainfall) in two Australian catchment areas. In the humid periods, the water movement was mostly surface and subsurface lateral movement, and was mostly determined by topography; in the dry periods, most of the flows were vertical, and the spatial distribution of soil moisture was determined mostly by the soil properties and topography of highly converging areas, such as highly curved depressions (Grayson et al. 1997).

Soil water content is conventionally measured via sample collection and drying (the gravimetric method). However, although this method is reliable, it is impractical because it requires destructive sampling, and has high costs, both in terms of labour and time (Elshorbagy and Parasuraman, 2008; Topp et al., 1984). Devices such as neutron probes, electromagnetic sensors and heat pulse tracers are used to obtain *in situ* point-specific soil moisture data (Gao et al., 2013; Robinson et al., 2008). Time Domain Reflectometry (TDR) is a noteworthy method; this approach determines soil moisture by measuring the dielectric constant following an electromagnetic pulse emission (Topp et al., 1984).

In the last decade, the use of remote sensing applied to large areas (river basins larger than 2,500 km^2) has increased, with the launch of projects to study soil moisture, such as Soil Moisture and Oceans Salinity (SMOS) and Soil Moisture Active Passive (SMAP) (McCabe et al., 2017). The use of point-specific measurements via *in situ* sensors has increased. Nonetheless, there remains a gap in the assessment of intermediate-scale areas, for which information on watershed characteristics, including spatiotemporal soil-moisture dynamics, is lacking (Robinson et al., 2008; Western et al., 2002). For intermediate scales (subwatersheds or small watersheds of 1–80 km^2; Robinson et al., 2008), the understanding of soil moisture is limited, both by the lack of *in situ* soil moisture measurements, and because of the complexity of the watershed environments, which are characterized by a range of soil types, diverse topography and multiple land-use types (Hagen et al., 2020). Therefore, the use of data-driven models such as artificial neural networks (ANNs) are an efficient alternative for soil moisture modelling and for examining related problems at small watersheds.

ANNs for modelling soil moisture typically use remote sensing data for large areas (Cui et al., 2016; Hachani et al., 2019; Rodriguez-Fernandez et al., 2015; Santi et al., 2016; Yao et al., 2017), and point-specific data for small catchment areas (smaller than 1 km^2 area), based on meteorological parameters, soil properties, land use and topographic features (Arsoy et al., 2013; Contador et al., 2006; Elshorbagy and Parasuraman, 2008; Oliveira et al., 2021; Yang et al., 2018). However, very few studies have addressed intermediate-scale catchments (Al-mukhtar, 2016; Gill et al., 2006; Oliveira et al., 2017). For instance, Oliveira et al. (2017) used ANNs to analyse spatiotemporal variation in SWC in a 78 km^2 watershed in Brazil, based on climate data, soil physical properties and topographic variables. The results show that it is possible to estimate SWC efficiently (Nash-Sutcliffe statistic (NS) = 0.770) using topographic data, soil physical properties and rainfall. Alternatively, SWC can be estimated via simplified models using rainfall and topographic information, although with less satisfactory performance (NS = 0.65). However, Oliveira et al. (2017) modelled soil moisture for only the surface layer, leaving a knowledge gap that will be addressed by modelling of the subsurface layer.

The primary purpose of this research was to assess the ability of regional empirical soil moisture models to predict both surface and subsurface layer dynamics, using data on climate, soil properties, topographic features, and rainfall, for a small watershed. Further, we evaluated which variables are most important in determining the performance of these models. To do this, we used ANN modelling with the Multilayer Perceptron (MLP) architecture, due to its relative simplicity and high capacity for approximating nonlinear relationships (Hornik et al., 1989).

2. Materials and methods

2.1. Study area

The study was performed in the Arroio do Ouro watershed, with an area of 17.17 km^2, in the state of Rio Grande do Sul, Brazil (Fig. 1). The elevation ranges from 76 to 326 m above sea level; the mean slope is 7.4°, and it can reach up to 30° (Bartels et al., 2021). The region is located at the Pelotas batholith, a plutonic complex which includes granite, gabbro and diorite, located in the Dom Feliciano belt geotectonic unit, in southern Brazil (Philipp et al., 2016). The soils are considered Acrisols and Regosols, distinguished for being shallow and predominantly containing sandy loam (35%–75% sand), based on World Reference Base ratings (FAO, 2014) (Bartels et al., 2016; see also Fig. 2b). The climate is humid subtropical (Cfa in the Köppen classification), with hot summers and well distributed rainfall throughout the entire year (Peel et al., 2007). Annual rainfall is 1400 ± 299 mm, annual reference evapotranspiration is 1077 ± 33 mm, and mean annual temperature is 18.5 ± 0.5 °C (1971–2018).

Dry/wet variation results from the interaction of precipitation and evapotranspiration (Hu et al., 2018). Thus, drought indices have been widely used for monitoring local dry and wet conditions. We applied two indices – the Standardized Precipitation Index (SPI) (McKee et al., 1993) and the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) – to evaluate the representativeness of the soil moisture monitoring period, using a long time series (1971–2019). Although we tested these indices at various time scales (e.g., 1, 3, 6, and 12 months; supplementary material section), we analysed only at the one-month scale in this study, because of the strong correlations found with soil moisture (Scaini et al., 2015).

2.2. Soil moisture measurement

Surface (0–10 cm) and subsurface (10–20 cm) soil samples were collected at 39 points at the site (Fig. 2), from 28 February 2018 to 3 September 2018, during ten surveys (Fig. 3). In each survey, soil samples were collected over four consecutive days (two surveys) and five consecutive days (eight surveys), totalling 24 soil samples per layer at each of the 39 measurement points. In total, 936 samples were collected from each layer. As it was not possible to sample all points in a single day, it was necessary to divide the 39 sampling points into two sets: set A with 18 points, and set B with 21 points (Fig. 2a).

Gravimetric soil moisture determination (the ratio between water mass and dry soil mass) was performed in a lab, by weighing the wet samples, then drying them in a drying oven at 105 °C for a minimum time of 24 h, and weighing the dried samples. The median soil moisture value obtained from triplicate measurements was used in network training and validation, avoiding possible errors associated with extreme values.

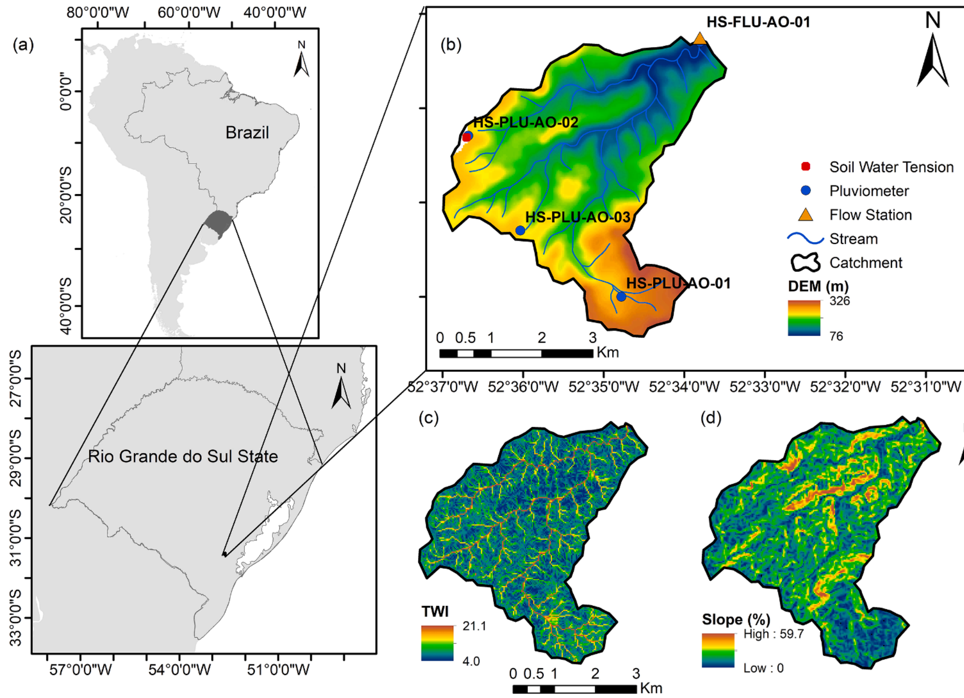

Fig. 1. (a) Study area in the State of Rio Grande do Sul, Brazil; (b) Arroio do Ouro watershed showing digital elevation model, streams (blue lines) and the locations of the tensiometer (red point), rain gauges (blue points) and flow station (yellow triangle); (c) Topographic wetness index (TWI); (d) Slope.

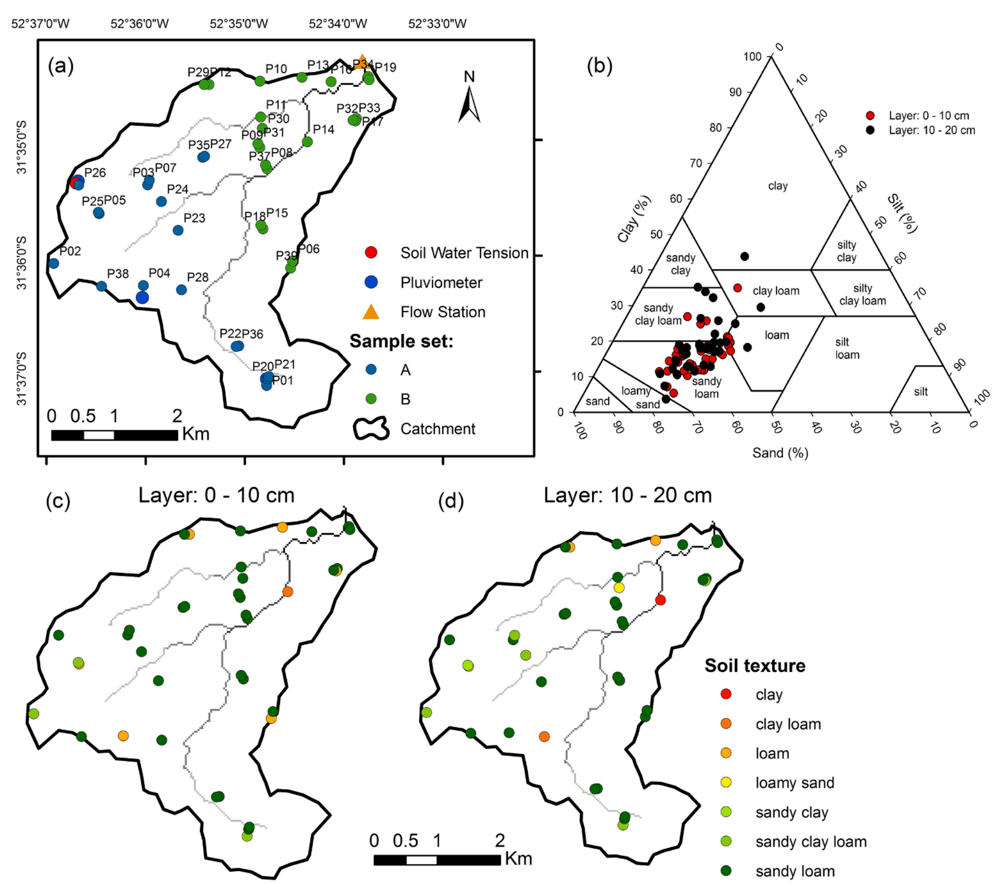

Fig. 2. (a) Soil moisture at the 39 monitoring points, divided spatially into two sets (A and B), so that one set could be sampled on one day; (b) Soil texture at the 39 points at the surface (red dots) and subsurface (black dots); (c, d) Soil texture rating at the 39 points for the surface and subsurface, respectively.

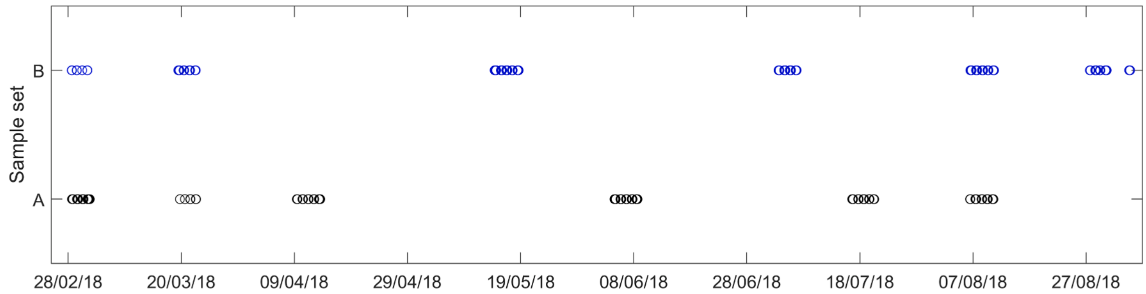

Fig. 3. Soil moisture monitoring for the samples from sets A and B, from 28 February to 3 September 2018. Sets A and B were divided spatially to enable sampling of one set on one day.

2.3. Model input variables

In total, 48 input variables were tested using ANN models, to select those variables that improved the estimation of soil moisture spatial and temporal variability. The variables were divided in four categories: (i) Topography (6 variables); (ii) Soil properties and land use and cover (14); (iii) Climatic variables (17); and (iv) rainfall-related variables (11).

2.3.1. Topographic variables

Six topographic variables were tested; four reflect point-related features (elevation, slope, distance from sample point to the closest river reach, and difference in elevation between the sampling point and the closest river reach). The two others, terrain curvature and topographic wetness index (TWI), are associated with the water mass, which is proportional to the contributing area, and its momentum, which is proportional to the slope (Contador et al., 2006). TWI was originally drafted by Beven and Kirkby (1979), calculated as follows:

$$TWI = ln\left(\frac{\alpha}{tan\beta}\right) \tag{1}$$

where α represents the upstream contribution area per unit contour length (specific area) and β is the local slope.

All topographic variables were extracted from the digital elevation model, with a spatial resolution of 30 m. The topographic variables are

Table 1
Topography, climate, and rainfall-related variable categories that were used as input for the artificial neural network models, and the model output (soil moisture), for the surface and subsurface soil layers.

N°	Variable	Categories	Minimum	Maximum	Mean	Median	SD
1	Elevation (m)	Topography	94.00	269.00	188.26	188.00	48.45
2	Slope (%)		1.37	39.17	13.47	10.70	9.15
3	[1]TWI (-)		4.34	13.37	7.01	6.41	2.00
4	Curvature (-)		−1.32	1.44	0.03	0.12	0.63
5	[2]DTR (m)		15.70	533.50	193.96	157.80	129.68
6	[3]DNR (m)		0.00	84.00	22.59	21.00	20.16
7	[4]Season (-)	Climate	–	–	–	–	–
8	Min. air temp. (°C)		2.40	20.90	11.84	11.10	4.83
9	Max. air temp. (°C)		9.20	32.60	20.89	21.30	6.39
10	Mean air temp. (°C)		6.33	25.80	15.91	15.34	5.17
11	Min. Rel. humid. (%)		24.80	99.70	64.00	61.40	18.51
12	Max. Rel. humid. (%)		82.30	100.00	98.91	100.00	3.44
13	Mean Rel. humid. (%)		55.71	100.00	86.14	87.85	10.58
14	Soil Temp. 5 cm (°C)		9.10	28.30	17.59	16.20	5.48
15	Global solar radiation (cal.cm^{-2} day^{-1})		27.00	564.70	273.77	260.80	158.34
16	[5]ETo (mm)		0.40	5.00	2.30	2.35	1.32
17	[6]Cum. ETo with 5 days (mm)		2.70	21.20	11.68	10.80	5.00
18	[6] Cum. ETo with 7 days (mm)		6.30	27.90	16.12	14.15	6.79
19	[6] Cum. ETo with 14 days (mm)		15.10	60.10	31.66	27.25	14.26
20	[6] Cum. ETo with 21 days (mm)		23.00	90.30	47.28	40.20	21.49
21	[6] Cum. ETo with 30 days (mm)		38.00	139.70	70.87	60.00	32.34
22	[6] Cum. ETo with 45 days (mm)		60.00	207.80	110.38	86.50	52.06
23	[6] Cum. ETo with 60 days (mm)		83.90	287.20	154.92	131.15	72.24
24	[7] Cum. rainfall with 6 h (mm)	Rainfall	0.00	18.45	1.24	0.00	3.38
25	[7] Cum. rainfall with 12 h (mm)		0.00	26.79	2.33	0.00	5.29
26	[8] Cum. rainfall with 1 day (mm)		0.00	29.73	4.93	0.11	8.01
27	[8] Cum. rainfall with 2 days (mm)		0.00	100.78	11.07	3.27	18.68
28	[8] Cum. rainfall with 3 days (mm)		0.00	145.39	16.80	7.00	25.73
29	[8] Cum. rainfall with 4 days (mm)		0.00	175.27	21.24	11.80	30.32
30	[8] Cum. rainfall with 5 days (mm)		0.00	196.90	26.51	17.54	34.58
31	[8] Cum. rainfall with 10 day (mm)		1.89	198.31	43.94	42.61	37.01
32	[8] Cum. rainfall with 15 days (mm)		10.02	255.66	72.38	61.82	44.31
33	[8] Cum. rainfall with 25 days (mm)		43.89	270.40	124.65	106.35	60.45
34	[9]EWMA of past hour rainfall (mm)		0.00	1.53	0.21	0.06	0.34
–	Soil Moisture (g g^{-1}) – Layer: 0–10 cm	Output	0.013	0.438	0.193	0.192	0.074
–	Soil Moisture (g g^{-1}) – Layer: 10–20 cm		0.014	0.383	0.178	0.179	0.063

N°: Identification of the respective variable; SD: Standard deviation; [1] Topographic wetness index; [2] Distance from sampled point to the closest river stretch; [3] Difference in elevation between the sampling point and the closest river stretch; [4] Season (summer, 1; autumn, 2; winter, 3); [5] Reference evapotranspiration (ETo); [6] Cumulative Eto over the preceding 5 to 60 days; [7] Cumulative rainfall over the preceding 6 to 12 h; [8] Cumulative rainfall over the preceding 1 to 25 days; [9]EWMA: Exponential weighted moving average.

described in Table S1, and their descriptive statistics are presented in Table 1.

2.3.2. Soil variables

The soil variable category comprises 14 variables: land use and cover; soil bulk density; macroporosity; microporosity; total porosity; clay; silt; total sand content; very coarse sand; coarse sand; medium sand; fine sand; very fine sand; and soil water tension. Except for land use and cover, these variables were collected in three replicates, of which we used the median, from the surface (0–10 cm) and subsurface (10–20 cm) layers at each point (Table S2).

Land use and cover was scored as follows: native forest, 1; native grassland, 2; fruit crops, 3; annual crops with vegetable covering, 4; annual crops without vegetable covering, 5; commercial forests, 6. These scores reflect the spatiotemporal variability in land use and cover.

Undisturbed soil samples were collected in metal tubes (0.076 m diameter; 344.1 cm^3 core volume) using a Uhland soil sampler, to determine soil bulk density, total porosity, macroporosity and microporosity. Soil microporosity corresponds to water retained at a pressure potential of 6 kPa (pore diameter equivalent to 50 µm), and macroporosity was calculated as the difference between total porosity and microporosity (Danielson and Sutherland, 1986). Samples were then dried for 24 h at 105 °C to determine soil bulk density (Blake and Hartge, 1986). Mean total porosity was slightly higher for the surface layer than for the subsurface layer. However, the subsurface layer exhibited higher amplitude in variation (between the maximum and minimum microporosity) than the surface layer (Table 2). For the surface layer, native forest had the lowest soil bulk density and highest total porosity.

Disturbed soil samples were air-dried, sieved through a 2 mm sieve, and used to analyse soil granulometry (clay, silt, and sand percentages) using the pipette method (Gee and Bauder, 1986). Sandy loam predominated in both layers (Fig. 2b). Soil water tension was measured daily throughout the study period using tensiometers installed in the watershed (Fig. 1b). The tensiometers were installed with two replicates, at depths of 7 cm and 15 cm from the soil surface, to measure the surface and subsurface layers, respectively. As input into the neural network models, we used soil water tensions obtained on the same days as the sample surveys. The amplitude of variation in soil water tension was higher in the surface than in the subsurface layer (Table 2).

2.3.3. Climatic variables

Seventeen climatic variables, measured daily, were tested as inputs for the ANN models. The season variable was used, based on the season in which the soil moisture samples were collected: summer, 1; autumn, 2; and winter, 3. This approach has been performed with satisfactory results (Oliveira et al., 2017). Air relative humidity and air temperature (minimum, medium, and maximum) were obtained from two automatic stations installed in the watershed (HS-PLU-AO-01 and HS-PLU-AO-03, Fig. 1b). The other variables – soil temperature at 5 cm depth, global solar radiation and reference evapotranspiration (ETo) – were obtained from a weather station of the Empresa Brasileira de Pesquisa Agropecuária (Brazilian Corporation of Agricultural Research - EMBRAPA), located about 17 km from the watershed (Latitude 31° 41′ S; Longitude 52° 26′ O; elevation: 57 m). ETo was calculated via the Penman-Monteith equation, as recommended by the FAO (Allen et al., 1998). Along with ETo values, another eight input variables were tested in the model (ETo on the day of soil moisture measurement, and the cumulative ETo from day 5 before soil moisture sampling (hereafter “cumulative 5-d Eto”); the same naming convention is then applied to cumulative ETo on days 7, 14, 21, 30, 45, and 60 before soil moisture sampling). The sample surveys were conducted on days with low and high global solar radiation, which affects ETo; ETo exhibited a large range (0.4 to 5.0 mm day^{-1}) during the monitoring period (Table 1).

2.3.4. Rainfall variables

We selected 11 rainfall-related variables: cumulative rainfall in the 6 h and 12 h before soil moisture measurement (hereafter “cumulative 6-h rainfall” and “cumulative 12-h rainfall”); cumulative rainfall on day 1 before soil moisture measurement (hereafter “cumulative 1-d rainfall”; the same naming convention is then applied to cumulative rainfall on days 2, 3, 4, 5, 10, 15, and 25 before soil moisture measurement); and the exponential weighted moving average (EWMA) of rainfall. EWMA was introduced by Moore (1980) to represent soil moisture in rainfall-based models. EWMA assigns a higher weight to more recent rainfall events; it has been used (Oliveira et al. 2017) as an important variable in estimating soil moisture via ANNs.

Rainfall data were obtained from three tipping-bucket rain gauges installed in the watershed (Fig. 1b). We calculated average rainfall over the watershed using the Thiessen polygon method, and used this in the model.

Table 1 describes the topography, climate, and rainfall variable categories that were used as inputs for the ANN models, as well as the model output (soil moisture). Table 2 describes the soil-related variables used as input for the ANN models.

2.4. ANN models

ANNs emerged with the artificial neuron concept of McCulloch and Pitts (1943). However, it was only after the 1990 s that they were applied with success to hydrology and related areas (ASCE, 2000; Dawson and Wilby, 2001) . They achieved relevance following the development by Rumelhart et al. (1986) of the algorithm for training

Table 2
Soil characteristics used as input for the artificial neural network models.

N°	Variable	Minimum	Maximum	Mean	Median	SD	Minimum	Maximum	Mean	Median	SD
		Layer: 0–10 cm					Layer: 10–20 cm				
35	[1]Land use and cover (-)	–	–	–	–	–	–	–	–	–	–
36	[2]BD (g cm^{-3})	0.886	1.663	1.401	1.468	0.207	1.064	1.710	1.433	1.449	0.146
37	[3]Macro (cm^3 cm^{-3})	0.050	0.313	0.156	0.132	0.079	0.039	0.305	0.146	0.135	0.074
38	[4]Micro (cm^3 cm^{-3})	0.104	0.413	0.296	0.303	0.071	0.066	0.472	0.281	0.292	0.076
39	[5]TP (cm^3 cm^{-3})	0.356	0.611	0.451	0.430	0.066	0.329	0.574	0.427	0.424	0.052
40	Soil water tension (cm Hg)	0.00	22.00	3.84	2.00	4.34	0.00	9.50	3.26	2.38	2.63
41	Clay (%)	5.37	34.92	16.42	16.19	5.51	3.67	43.82	18.86	17.82	7.87
42	Silt (%)	15.42	31.16	23.10	22.67	4.42	13.90	34.90	22.87	22.38	4.59
43	Total sand (%)	41.03	72.72	60.48	61.73	7.35	34.78	74.73	58.28	58.75	8.88
44	Very coarse sand (%)	2.56	37.79	17.30	15.16	9.31	1.34	37.51	16.29	15.72	9.28
45	Coarse sand (%)	5.69	27.21	14.22	13.62	3.52	4.27	24.54	13.78	13.74	3.50
46	Medium sand (%)	6.97	19.55	11.88	11.56	3.00	6.20	21.78	11.73	10.70	3.58
47	Fine sand (%)	8.11	20.26	13.00	12.16	3.12	3.64	21.66	12.22	11.98	3.33
48	Very fine sand (%)	0.40	9.51	4.07	3.94	2.23	0.52	8.07	4.27	4.38	2.30

N°: Identification of the respective variable; SD: Standard deviation; [1] Land use and cover (native forest, 1; native grassland, 2; fruit crops, 3; annual crops with vegetable covering, 4; annual crops without vegetable covering, 5; commercial forests, 6); [2] Soil bulk density; [3] Macroporosity; [4] Microporosity; [5] Total porosity.

MLP networks; according to this algorithm, the set of artificial neurons is arranged into a layered structure, and the outputs from previous neural layers are used as inputs by the following neurons in determining the next output layer. An MLP network with a three-layer architecture is considered capable of approximating any continuous function to any desired degree of accuracy, if it is appropriately trained and relies on a sufficient number of neurons in the inner layer (Hornik et al., 1989). We chose it for this research because of these features.

The backpropagation algorithm (Rumelhart et al., 1986) and convergence acceleration techniques of Vogl et al. (1988) were applied to network training. However, because of their approximation capabilities, ANNs may be overly adjusted to the data, which would make the model unviable for other applications. To prevent this overfitting, the cross-validation technique (Hecht-Nielsen, 1990) was implemented, with the available dataset divided into three parts, for training, validation, and verification. In this way, one can identify the training cycle in which performance with training samples continues to improve while performance with data other than the training data decreases. Such a cycle indicates model overfitting, and the training of the network should be interrupted. Finally, the trained network is submitted to a verification sample which was not included in the training stage, to guarantee the model's capacity to generalize (Hecht-Nielsen, 1990).

Through systematic sampling, the same frequency distribution of the complete original series was maintained for training, validation, and verification sub-samples. For the training subsample, additional care was taken to ensure the representativeness of the entire data domain (extreme values), both in terms of the input and output variables (soil moisture).

At the initialization of the neural network, synaptic weights are randomly assigned to the neurons, which may result in an unfavourable beginning. Therefore, a series of repetitions is used to identify the ANN whose training results in the best validation performance.

Due to the high number of variables (48) that could be included in the input layer, we adopted an initial method to select the variables to be tested. The selection method departed from the Pearson correlation coefficient (r) between the input variables and soil moisture (Table 3). Variables with higher correlations with soil moisture were chosen; among them, those with a correlation among themselves below a given limit ($r = 0.9$) were selected. Linear correlation applied to variable selection provides an indication of which variables are worth including as inputs in the model for soil moisture estimation. However, using strongly correlated variables duplicates the information used by the model, which confuses the ANNs, often reducing network performance, as observed by Oliveira et al. (2017). We used this method for initial model selection, and included variables that were not highly correlated with soil moisture, but that are known to be affect soil moisture dynamics.

We used a method first presented by Sari et al. (2017) to estimate the number of inner layer neurons. From an oversized network, the number of inner-layer neurons is progressively reduced until a reduction in its generalization capacity, due to the reduced degrees of freedom, is observed. The lowest number of neurons observed before the reduction in generalization capacity is chosen. This must be assessed using the validation sample, since the verification sample must not be used in the training of synaptic weights or in choosing the ANN architecture (Hecht-Nielsen, 1990).

The ANN models were developed, trained, and verified by the authors using MATLAB® R2012b. In total, 144 ANN models were tested, with different input variable combinations.

2.5. Evaluation of model performance

After training, validation and verification, the statistical indicators were calculated based on the errors between the observed and simulated values: mean absolute error (MAE); root mean squared error (RMSE); Nash-Sutcliffe efficiency coefficient (NS); and the quantiles of the error distribution (10, 50, and 90%). MAE, RMSE, and NS are calculated as follows:

Table 3
Pearson linear correlation *(r)* between the input variables and soil moisture at the surface (0–10 cm) and subsurface (10–20 cm) layers.

Variable	r (0–10 cm)	r (10–20 cm)	Variable	r (0–10 cm)	r (10–20 cm)
Elevation (m)	0.163	0.155	Min. Rel. humid. (%)	0.207	0.195
Slope (%)	−0.223	−0.321	Max. Rel. humid. (%)	−0.039	−0.038
TWI (-)	0.170	0.222	Rel. humid. (%)	0.148	0.137
Curvature (-)	−0.078	−0.102	Soil Temp. 5 cm (°C)	−0.334	−0.298
DTR (m)	−0.120	−0.178	Global solar radiation (cal cm^{-2} day^{-1})	−0.338	−0.322
DTR (m)	−0.168	−0.248	ETo (mm)	−0.332	−0.308
Land use and cover (-)	−0.419	−0.271	Cum. ETo of 5 days (mm)	−0.369	−0.349
BD (g cm^{-3})	−0.533	−0.442	Cum. ETo of 7 days (mm)	−0.379	−0.354
Macro (cm^3 cm^{-3})	−0.162	−0.371	Cum. ETo of 14 days (mm)	−0.385	−0.370
Micro (cm^3 cm^{-3})	0.651	0.710	Cum. ETo of 21 days (mm)	−0.400	−0.384
TP (cm^3 cm^{-3})	0.503	0.507	Cum. ETo of 30 days (mm)	−0.408	−0,385
Soil water tension (cm Hg)	−0.344	−0.390	Cum. ETo of 45 days (mm)	−0.419	−0.390
Clay (%)	0.492	0.557	Cum. ETo of 60 days (mm)	−0.416	−0.377
Silt (%)	0.422	0.342	Cum. rainfall of 6 h (mm)	0.148	0.151
Total sand (%)	−0.624	−0.671	Cum. rainfall of 12 h (mm)	0.229	0.232
Very Coarse sand (%)	−0.578	−0.648	Cum. rainfall of 1 day (mm)	0.256	0.256
Coarse sand (%)	−0.126	−0.193	Cum. rainfall of 2 days (mm)	0.233	0.214
Medium sand (%)	0.096	0.054	Cum. rainfall of 3 days (mm)	0.234	0.211
Fine sand (%)	0.190	0.134	Cum. rainfall of 4 days (mm)	0.220	0.195
Very Fine sand (%)	0.169	0.038	Cum. rainfall of 5 days (mm)	0.205	0.175
Season (-)	0.402	0.361	Cum. rainfall of 10 days (mm)	0.164	0.160
Min. air temp. (°C)	−0.230	−0.197	Cum. rainfall of 15 days (mm)	0.279	0.260
Max. air temp. (°C)	−0.276	−0.252	Cum. rainfall of 25 days (mm)	0.251	0.230
Mean air temp. (°C)	−0.266	−0.236	EWMA of past hour rainfall (mm)	0.275	0.264

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|y_j - \hat{y}_j\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2} \tag{3}$$

$$NS = 1 - \sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2 / \sum_{j=1}^{N}\left(y_j - \bar{y}\right)^2 \tag{4}$$

where y_j is the jth observed variable;$\hat{y}_j$ is the jth simulated variable assessed;$\bar{y}$ is the average of all observed values, and N is the total number of measurements.

Since the NS coefficient represents the proportion of the variance explained by the model, the difference in NS between the complete model and a simpler model reflects the proportion of explained variance lost by using the simpler model. Therefore, this difference can be used to quantify comparisons.

3. Results

3.1. Analysis of correlation with soil moisture

Table 3 presents the correlation coefficient (r) between the 48 variables and soil moisture. In both layers, soil moisture was most strongly correlated with microporosity (0.651 and 0.710) and total sand content (−0.624 and − 0.671), for the surface and subsurface layers, respectively. The strong positive correlation for microporosity is understandable, because it represents the soil's water retention capacity. The opposite correlation, for the proportion of sand, makes sense because higher sand content improves soil infiltration and reduces its moisture retention capacity.

In terms of topography, soil moisture was most highly correlated with slope, followed by the difference in elevation between the sampling point and the closest river reach (negative correlations), and TWI (positive correlation; Table 3). A steep slope increases runoff to lower-lying regions. The highest points in the watershed had lower soil moisture than points at similar elevation as the channel network, reflecting water movement from the top hillslope towards the channel network. Further, as the upstream contributing area increases (resulting in more cumulative flow), the downstream soil moisture increases.

The climatic variables were more strongly correlated with soil moisture in the surface layer than in the subsurface layer. Compared to the climatic variables, the rainfall-related variables were less strongly correlated with soil moisture (Table 3). This suggests that these climatic variables affect soil moisture more than rainfall-related variables, especially at the surface layer.

3.2. Temporal and spatial soil moisture variation in the watershed

During the monitoring period (February to September 2018), soil moisture ranged from 0.013 to 0.438 g g^{-1} for the surface and 0.014 to 0.383 g g^{-1} for the subsurface, presenting very similar mean values throughout the period (Table 1). Average soil gravimetric moisture was lowest during the first week of surveys, at 0.123 g g^{-1} for the surface and 0.114 g g^{-1} for the subsurface.

Temporal soil moisture variation was influenced by rainfall and evapotranspiration seasonality (Fig. 4a). From summer to the end of autumn (21 December 2017 to 20 June 2018), there was 605.2 mm cumulative rainfall and 624.2 mm cumulative ETo. In the winter (21 June to 22 September 2018), cumulative rainfall was 646.8 mm, while ETo was only 156.8 mm. In both layers, soil moisture was lowest during the first five surveys (summer and autumn; Fig. 4b, c), presenting averages of 0.167 g g^{-1} for the surface and 0.159 g g^{-1} for the subsurface. In winter, average soil moisture was higher, at 0.218 g g^{-1} for the surface and 0.195 g g^{-1} for the subsurface, respectively.

These findings reflect the fact that soil moisture responds to rainfall (which raises it) and evapotranspiration (which lowers it). SPI (SPEI) ranged from − 1.0 to 1.56 (−1.1 to 1.6) at the one-month time scale (Figs S1 and S2). SPI and SPEI can be classified as near-normal (−0.99 to 0.99), moderately wet (1.0 to 1.49), and severely wet (1.50 to 1.99). For the period 1971 to 2019, the proportions of samples falling into these categories, for SPI (SPEI), were 70% (65%) for near-normal, 10.0% (10.5%) for moderately wet, and 4.1% (5.8%) for severely wet (Fig. S3). Overall, the frequency of SPI (SPEI) explained 81% (80%) of the variation in the time series. This indicates that the conditions during the monitoring period were near-normal, both in terms of dry and wet conditions.

High soil moisture was observed even during the driest period (the first five weeks), and low soil moisture even during the wet period (the last four weeks), particularly in the subsurface (Fig. 4b, c). This becomes more evident when comparing the two sets of sample points: set A points have lower soil moisture variability than set B points, in both layers (Fig. 5). Point P14 had the highest median surface and subsurface soil moisture (at 0.322 and 0.357 g g^{-1}, respectively; Fig. 5b, d). This point had the highest proportion of clay, and highest microporosity, in the surface layer (Table S2). In contrast, point P30 had the lowest median soil moisture (0.038 and 0.031 g g^{-1}), lowest proportion of clay and lowest microporosity, in both layers. This reflects the importance of soil granulometry and porosity for water retention.

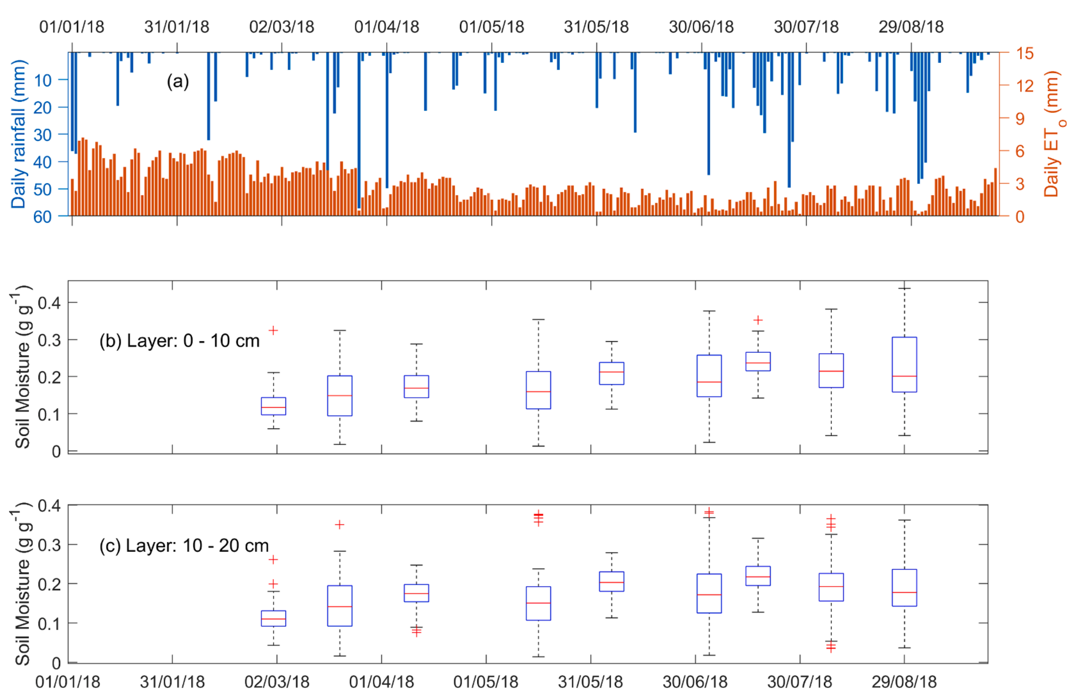

Fig. 4. (a) Rainfall and reference evapotranspiration (ETo) variation during the study period. Box-plot of gravimetric moisture measurements for the ten surveys at (b) the surface layer (0–10 cm) and (c) the subsurface layer (10–20 cm). A, B, and A/B: group of points displayed, according to Fig. 2a. Box edges: 25th and 75th percentiles; Central line: median; Whiskers: lower and upper non-divergent limits; Crosses: outliers.

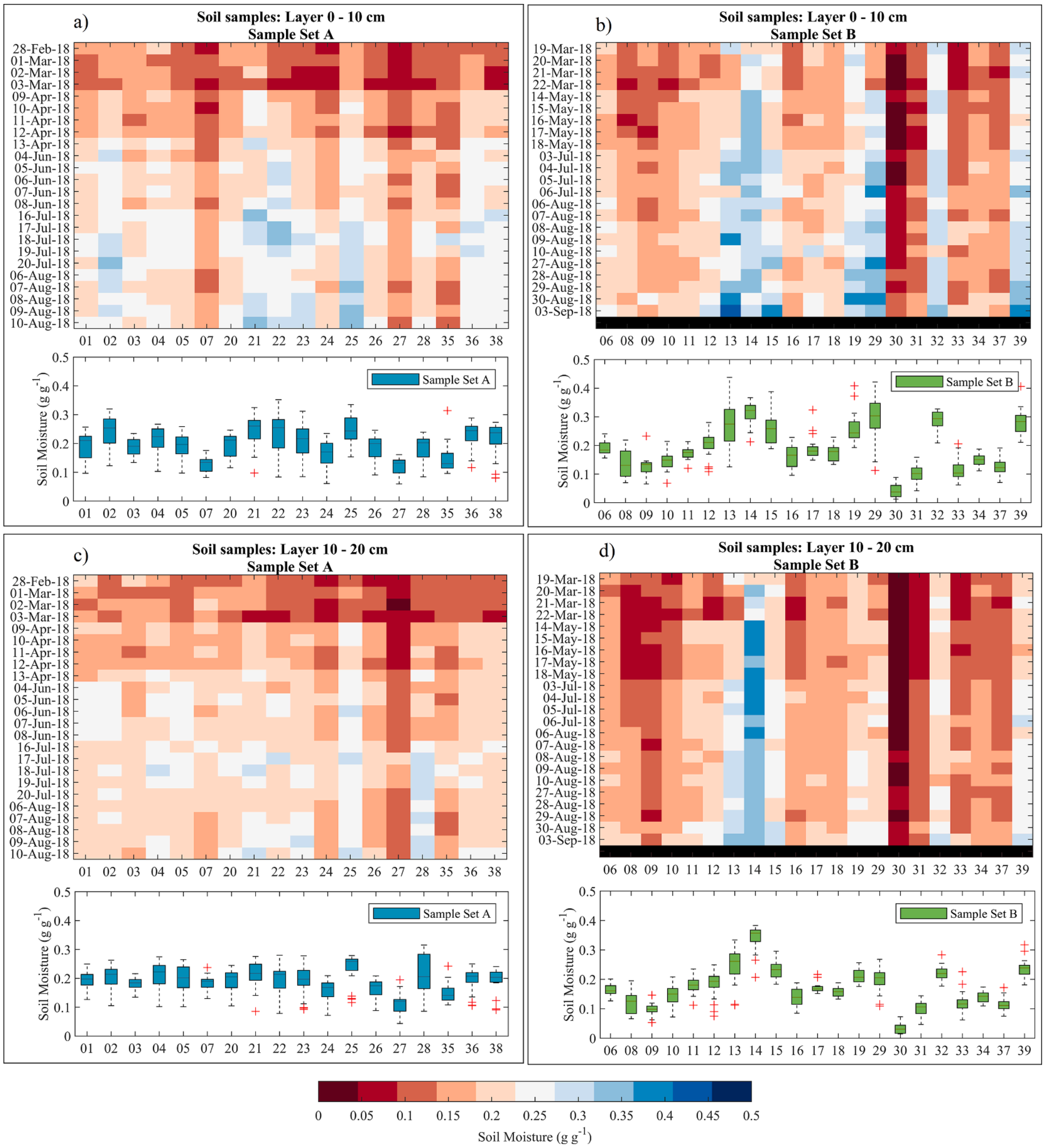

Fig. 5. Spatial and temporal distribution of soil moisture for the 39 measurement points. (a, b) Surface-layer soil samples from sets A and B, respectively; (c, d) Subsurface soil samples from sets A and B, respectively. The numbers on the x-axes indicate the soil sampling points.

Topography and land use and cover may also explain the differences in soil moisture between sample sets A and B. For both layers, spatial heterogeneity of soil moisture was higher in set B (CV = 0.44) than in set A (CV = 0.29). The average slope of set B points (17.5%) was higher than that of set A points (8.8%). Further, land use and cover affected soil moisture: points in native forest had greater variability (CV = 0.35) than those in grassland (CV = 0.28). One-third of the area of sample set B is native forest.

3.3. Soil moisture estimation via ANN models

We used one-third of the records from each subsample for cross-validation training, and the same training configuration was used for all models (20 repetitions and a maximum of 90,000 cycles). Starting with an oversized initial network, with 20 neurons in the hidden layer, each model was tested to determine its optimal complexity (Section 2.4). All models whose complexity was researched resulted with up to six neurons in the inner layer. Since, in all cases, the possible excess of complexity was contained by means of cross-validation, the number of six neurons was adopted as a reasonable standard for the final complexity of all models.

By assessing their linear correlations (Section 2.4), we selected 14 variables for the surface and 15 for the subsurface. We then tested the effects of including or removing certain variables. The initial selection of input variables had the greatest impact in selecting the best models for each layer (Table S3). The best models contained more topographical variables and fewer soil-related variables. Tables S4 and S5 present data on model performance evaluation during network verification, for the 39 models analysed.

For the subsequent analysis, we used the verification subsample (one-third of the samples). The main models for this stage of the analysis had NS coefficients ranging from 0.477 to 0.887 for the surface and 0.213 to 0.893 for the subsurface (Table 4). The two best models were M38 (surface, with 11 network input variables) and M49 (subsurface, with 13 network input variables), representing all of the variable categories (topography, soil, climate, and rainfall; Table S4). Model M38 (NS = 0.870) had the following input variables: elevation, slope, TWI, land use and cover, soil bulk density, microporosity, total sand content, season, cumulative 6-h rainfall, EWMA, and cumulative 7-d ETo (model statistics, in g g^{-1}: RMSE = 0.026, MAE = 0.02, E10 = − 0.03, E50 =

Table 4
Statistical verification performance of the best models for the surface and subsurface layers.

Model	38	51	52	53	54	55	56	49	64	65	66	67	68	69
Layer	Surface (0–10 cm)							Subsurface (10–20 cm)						
J[1]	18,273	62,377	48,001	16,781	13,005	34,215	5852	16,618	61,888	49,176	3784	53,695	13,096	1193
Input variables[2]	1, 2, 3, 35, 36, 38, 43, 7, 24, 34, 18	35, 36, 38, 43, 7, 24, 34, 18	1, 2, 3, 7, 24, 34, 18	1, 2, 3, 35, 36, 38, 43, 24, 34	1, 2, 3, 35, 36, 38, 43, 7, 18	1, 2, 3, 35, 36, 38, 43	35, 7, 24, 34, 18	1, 2, 3, 35, 36, 38, 40, 41, 43, 7, 26, 34, 22	35, 36, 38, 40, 41, 43, 7, 26, 34, 22	1, 2, 3, 7, 26, 34, 22	1, 2, 3, 35, 36, 38, 40, 41, 43, 26, 34	1, 2, 3, 35, 36, 38, 40, 41, 43, 7, 22	1, 2, 3, 35, 36, 38, 41, 43	35, 40, 7, 26, 34, 22
	Verification													
E10	−0.03	−0.033	−0.038	−0.046	−0.036	−0.054	−0.061	−0.021	−0.029	−0.042	−0.027	−0.025	−0.045	−0.07
E50	0.001	0.0014	−0.003	0.0033	−0.0045	0	0.0033	0.0002	0.0025	−0.0005	0.0044	−0.0013	0.0018	0.0036
E90	0.033	0.041	0.046	0.064	0.036	0.049	0.063	0.025	0.03	0.043	0.041	0.028	0.047	0.057
MAE	0.02	0.023	0.027	0.035	0.023	0.034	0.04	0.015	0.018	0.026	0.021	0.017	0.03	0.043
RMSE	0.026	0.031	0.035	0.045	0.031	0.044	0.052	0.020	0.024	0.034	0.027	0.022	0.040	0.055
NS	0.870	0.824	0.770	0.617	0.824	0.635	0.497	0.893	0.858	0.708	0.816	0.880	0.594	0.224

[1] J: number of cycles used in model selection. [2] Input variables: 1, elevation; 2, Slope; 3, TWI; 7, Season; 18, Cumulative 7-day ETo; 22, Cumulative 45-day ETo; 24, Cumulative 6-h rainfall; 26, Cumulative 1-day rainfall; 34, EWMA of rainfall in the hour before sampling; 35, Land use and cover; 36, Soil bulk density; 38, Microporosity; 40, Soil water tension; 41, Clay; 43, Total Sand.

0.001, and E90 = 0.033; Fig. 6a, Table 4). The symmetrical error distribution (E10 versus E90), its MAE, and the E50 close to zero, indicates good neural network adjustment during the verification stage.

In contrast, model M49 (the best subsurface model; N = 0.893) included soil water tension and clay percentage; further, rather than cumulative 7-d ETo, it included cumulative 45-d ETo; and rather than cumulative 6-h rainfall, it included cumulative 24-h rainfall (model statistics, in g g^{-1}: RMSE = 0.020, MAE = 0.015, E10 = − 0.021, E50 = 0.0002, and E90 = 0.025 (Fig. 6b, Table 4). M49 also had symmetrically distributed errors, with low MAE and E50 values, indicating good neural network adjustment, even better than for the surface model (M38).

The good verification performance of both M38 and M49 demonstrates that ANNs can be used to model soil moisture, and to make predictions even for situations not presented during training, thus confirming their capacity to generalize. Given that M49 presented better verification performance statistics, the performance statistics cannot be used alone to compare the surface and subsurface models. Nonetheless, it may be important to address large differences in model performance statistics, particularly in NS.

3.3.1. *Importance of input variables in the ANN models*

For the surface model (M38), removing the season variable caused the largest reduction in network predictive capacity, reducing NS by 0.076, producing a model with NS = 0.794, RMSE = 0.033 g g^{-1}, MAE = 0.026 g g^{-1}. EWMA of rainfall in the hour before sampling was the second most important input variable: its removal reduced the model's predictive capacity, reducing NS by 0.053, producing a model with NS = 0.817, RMSE = 0.031 g g^{-1}, MAE = 0.023 g g^{-1}. Therefore, EWMA and its interactions with the other variables were important predictive factors, although without a strong linear relationship with soil moisture.

For the subsurface layer model (M49), removing microporosity and the season variable caused the largest reductions in predictive performance (NS = 0.846 and 0.842, RMSE = 0.025 for both, with NS reductions of 0.047 and 0.051 g g^{-1}, MAE = 0.019 and 0.018 g g^{-1}, respectively). These findings indicate that, for the subsurface soil moisture, microporosity is as important as the season variable. Further, the season variable was less important in the subsurface model than in the surface model. These results highlight the relevance of easily obtainable input variables for predicting both surface and subsurface soil moisture.

3.3.2. *Importance of the variable categories for the best models*

The performance statistics of the M38 and M49 models, and of those obtained by removing all of the variables in each of the four variable categories, are shown in Fig. 7. The reductions in NS are relative to the complete models.

Removing the topographic variables reduced the verification performance of both models, noticeably increasing the dispersion of errors, with reductions in NS of 0.046 (M51, surface) and 0.035 (M64, subsurface).

Removing soil-related variables caused larger reductions in performance than removing topography-related variables, reducing NS by 0.100 (M52, surface) and 0.185 (M65, subsurface). This indicates that soil-related variables affected the subsurface model more than the surface model. M52 and M65 had similar E10 and E90 values: E10 was slightly higher in M65 (−0.042 g g^{-1}) than in M52 (−0.038 g g^{-1}).

Removing climate-related variables reduced the predictive performance of the surface and subsurface models (M53 and M66, respectively). For the surface model, removing climatic variables had more effect than removing the other variable categories, reducing NS by 0.253; for the subsurface model, the reduction in NS was lower, at 0.077. The errors were larger, and the error distribution was more asymmetric, for M53 (E10 = − 0.046 g g^{-1}, E90 = 0.064 g g^{-1}) than for M66 (E10 = − 0.027 g g^{-1}, E90 = 0.041 g g^{-1}). M53 tended to overestimate low soil moisture values and underestimate high soil moisture values (Fig. 7).

Removing rainfall-related variables reduced model performance less

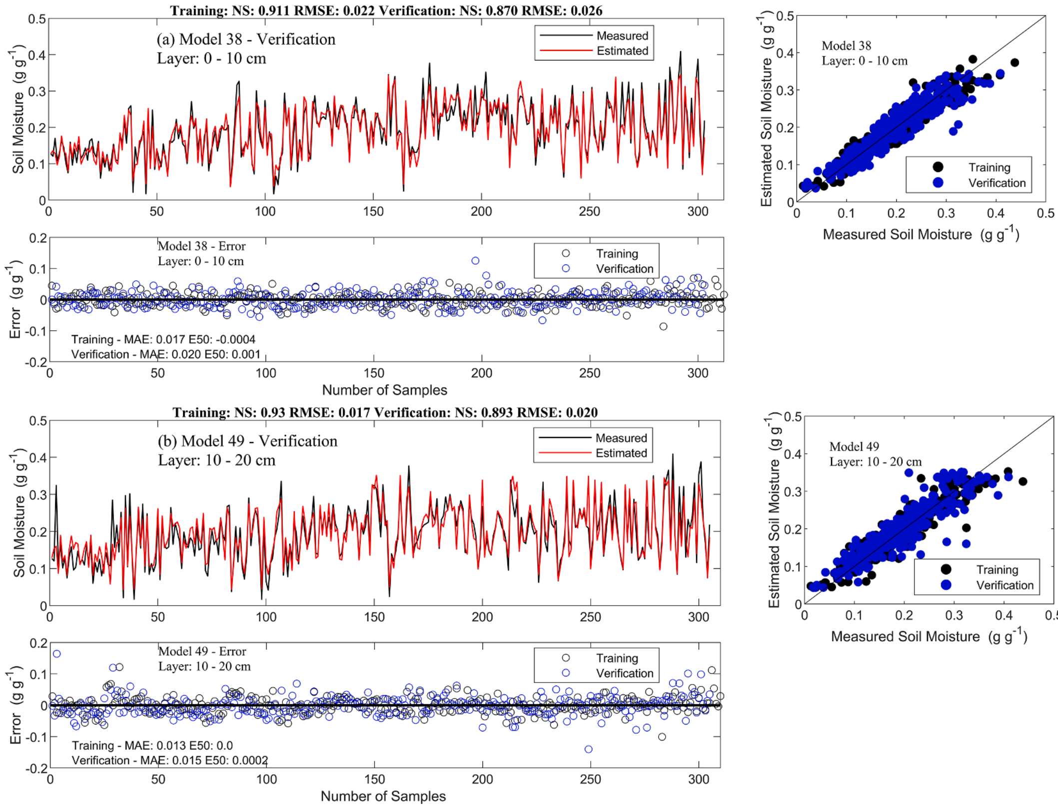

Fig. 6. Model performance during the verification process. (a) Model M38, surface layer; (b) Model M49, subsurface layer. Soil moisture measurements and estimates in relation to the ideal adjusted values (1:1 line) for training (black circles) and verification (blue circles) for both models. Error: difference between the measurements and estimates for training (black circles) and verification (blue circles) for both models.

for the subsurface model (M67) than for the surface model (M54), with NS reductions of 0.013 and 0.046, respectively. Although both models had symmetrical error distributions, the errors more dispersed for M54 (E10 = $-$ 0.036 g g^{-1}, E90 = 0.036 g g^{-1}) than for M67 (E10 = $-$ 0.025 g g^{-1}, E90 = 0.028 g g^{-1}).

In summary, although all four variable categories are important for the surface model, those related to climate are the most important, followed by those related to soil. For the subsurface model, soil-related variables are the most important, followed by those related to climate; rainfall and topographical variables are of little importance for the subsurface model.

3.3.3. *Importance of spatial and temporal variables for the best models*

The input variables were separated based on spatial features (for elevation, slope, TWI, land use and cover, soil bulk density, microporosity, clay content, and total sand content) and temporal features (land use and cover, soil water tension, climate, cumulative 6-h rainfall, cumulative 1-day rainfall, EWMA of rainfall in the hour before sampling, and cumulative 7-day and 45-day ETo). Because it is a spatial variable with patterns that change over time, land use and cover was considered as both spatial and temporal. Not surprisingly, excluding either the spatial or temporal information from the models reduced their predictive potential (Fig. 8, Tables S4 and S5).

Removing the spatial features from the models reduced the predictive performance of the subsurface model (M69) more than that of the surface model (M56). Although they had similar median errors (E50 = 0.0036 and 0.0033 g g^{-1} for M69 and M56, respectively), the error distribution of M69 was more asymmetrical (E10 = $-$ 0.07 g g^{-1}, E90 = 0.057 g g^{-1}) than that of M56 (E10 = $-$ 0.061 g g^{-1}, E90 = 0.063 g g^{-1}). This confirms our earlier finding that removing the soil-related variables reduces predictive performance more for the subsurface model than the surface model.

Removing temporal variables similarly reduces the predictive performance of both the subsurface model (M68: NS = 0.594, RMSE = 0.040 g g^{-1}, and MAE = 0.03 g g^{-1}) and the surface model (M55: NS = 0.635, RMSE = 0.044 g g^{-1} and MAE = 0,034 g g^{-1}). The errors of M55 were larger and more asymmetrically distributed (E10 = $-$ 0.054 g g^{-1}, E90 = 0.049 g g^{-1}) than those of M68 (E10 = $-$ 0.045 g g^{-1}, E90 = 0.047 g g^{-1}). Although the climate and rainfall-related variables had little influence on the performance of the subsurface model, removing all of the temporal variables (climate and rainfall-related variables, land use and cover, and soil water tension) reduced the model's predictive performance. This is because the soil-related variables have a strong influence on model performance.

For the surface model, removing spatial variables reduced predictive performance more than removing temporal variables. However, the effect of removing spatial variables was less strong for the surface than for the subsurface model. Although spatial variables are important for the surface model, they have a larger effect on the subsurface model (M69). This provides further evidence that the climatic variables are more relevant to the performance of the surface model.

4. Discussion

The effects of topography, soil, climate, land use, and land cover on soil moisture have been widely investigated (e.g. Hu and Cheng, 2014; Korres et al., 2015; Liang, 2017; Yang et al., 2017). In the studied watershed, for both layers, spatial heterogeneity of soil moisture was higher for set B than set A samples, which can be attributed to the higher average slope of set B. Further, the higher CV of soil moisture for the native forest than for the grassland indicates that land use and cover also influenced soil moisture variability. The greater heterogeneity of soil moisture on steep slopes and in native forest is consistent with previous findings (Korres et al., 2015; Yang et al., 2017).

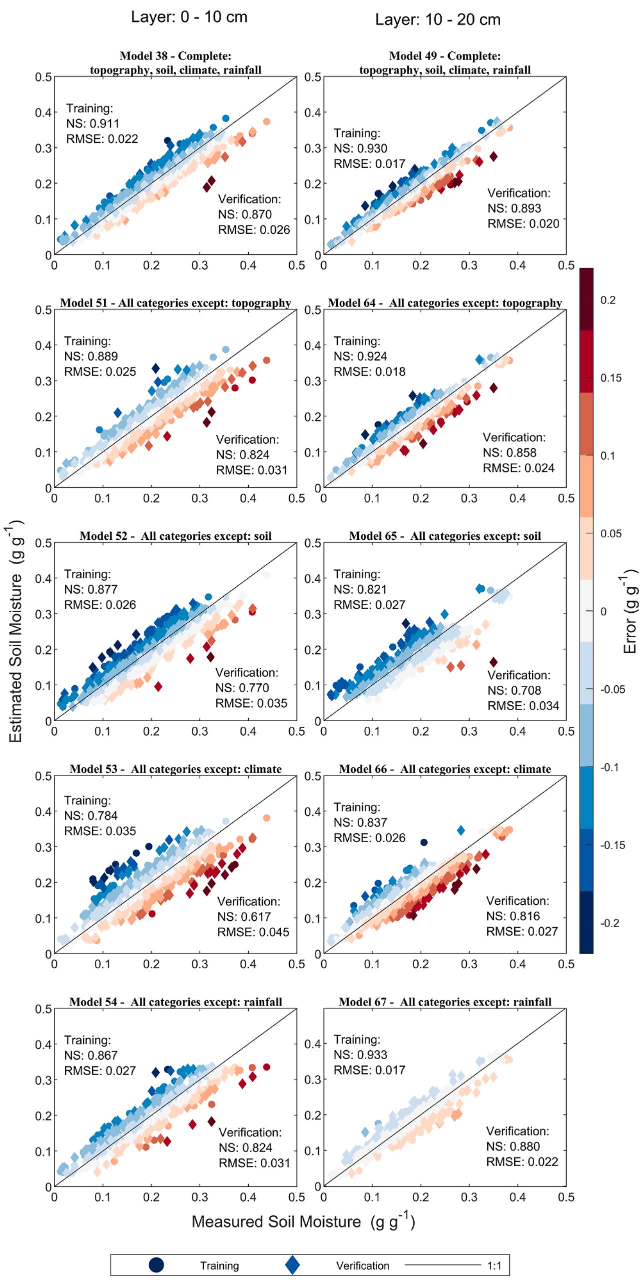

Fig. 7. Model performance testing by removing each variables category (topography, soil, climate and rainfall). The 1:1 line depicts the ideal adjusted values. For soil moisture, the colder (blue) and hotter (red) colors depict higher negative errors (overestimates) and higher positive errors (underestimates), respectively.

All four of the input variable categories (topography, soil, climate, and rainfall) were included in the best-performing ANN-based soil moisture models (M38 and M49). This reflects the complex spatiotemporal dynamics that determine soil moisture, at the watershed scale. The fact that variables with low linear correlations with soil moisture (e.g., elevation, EWMA; Table 3) were included in the best models indicates that even low-correlation variables can affect ANN performance. This is because ANNs build nonlinear relationships among input and output variables (Oliveira et al., 2017).

Both of the best-performing models (M38 and M49, for surface and subsurface, respectively) performed well during the verification stage. The slightly superior performance of the subsurface model probably reflects the fact that the soil moisture has lower coefficients of variation in this layer. Similarly, Contador et al. (2006) obtained good results when using ANN modelling to estimate soil moisture in a Spanish watershed, emphasizing the effects of changes in land cover on soil moisture. Further, Kornelsen and Coulibaly (2014) reported that ANNs can explain nonlinear soil moisture dynamics; however, they achieved good results for deeper layers only when using surface soil moisture as a network input. Using the same variable categories that we used, Oliveira et al. (2017) also achieved good results, for a watershed of the same climatic type but with very different land use, land cover, soil hydro-physical features, and topography. In the present study, cumulative 6-h rainfall was more important in predicting surface soil moisture, whereas cumulative 24-h rainfall was more important in predicting subsurface soil moisture, indicating that rainfall affected surface soil moisture faster than subsurface soil moisture. Similarly, cumulative 7-day ETo was more important for the surface model, whereas cumulative 45-day Eto was more important for the subsurface model, indicating that the surface soil dries out faster than the subsurface soil. This suggests that surface soil moisture responds to the most recent rainfall and climatic conditions, whereas subsurface soil moisture is attenuated by the soil's hydro-physical features, which control water infiltration and delay the effects of rainfall and climate conditions, as observed by Lv et al. (2019). Furthermore, the linkage of soil moisture in surface and subsurface is dependent on the kind of transition between soil horizons (Hagen et al., 2020). Water reallocation to greater depths causes soil moisture to be more stable at greater depths (Rosenbaum et al., 2012).

Excluding the input variable categories had different effects on the performance of the surface and subsurface ANN models. In both layers, excluding the soil-related variables caused greater loss of performance than removing the topographic variables. Gwak and Kim (2017) and Hu and Cheng (2014) report that soil-related properties are more important than topography in determining the soil moisture distribution. Various other studies have reported that topography, land use and land cover are essential in characterizing catchment-scale soil moisture variability (Liang, 2017; Yang et al., 2017; Yu et al., 2018).

For the surface layer, climate-related variables were more important than topography, soil, and rainfall-related variables in predicting soil moisture. This probably related to the observed differences in cumulative rainfall and ETo between dry periods (summer and autumn) and humid periods (winter). By assigning a numeric value to each season, this information is indirectly included in the model. Further, for small watersheds, it has been reported that changes in soil moisture over time may influence processes which control spatial patterns of soil moisture (e.g. Hu and Cheng, 2014; Liang, 2017; Western et al., 2004). In this context, for a small watershed in Germany, Rosenbaum et al. (2012) observed that temporal changes in the surface layer (0–5 cm) were strongly influenced by climatic forcing.

Consistent with Oliveira et al. (2017), we found that including simple, accessible, and low cost variables (such as land use and cover, and season) improved network performance in estimating soil moisture. Excluding rainfall-related variables caused small losses in predictive performance, especially for the subsurface layer. This is consistent with other experimental and modelling studies. For instance, Metzger et al. (2017), in a forest-parcel experiment, observed that soil wetting and rainfall patterns were weakly associated; they attributed this to the rapid drying of soil after rainfall, with dry soil being the stable condition over time. Using a Richards equation-derived 3D model for a hillslope, Coenders-Gerrits et al. (2013) observed that rainfall influences soil moisture predictions, but only during and shortly after a rainfall event, with bedrock topography being the limiting factor most of the time.

5. Conclusions

For a small watershed, we investigated the capacity of ANN models to predict regional soil moisture, both for surface and subsurface layers, and evaluated the main driving factors. The models were configured using inputs in four categories (topography, soil properties, climate, and

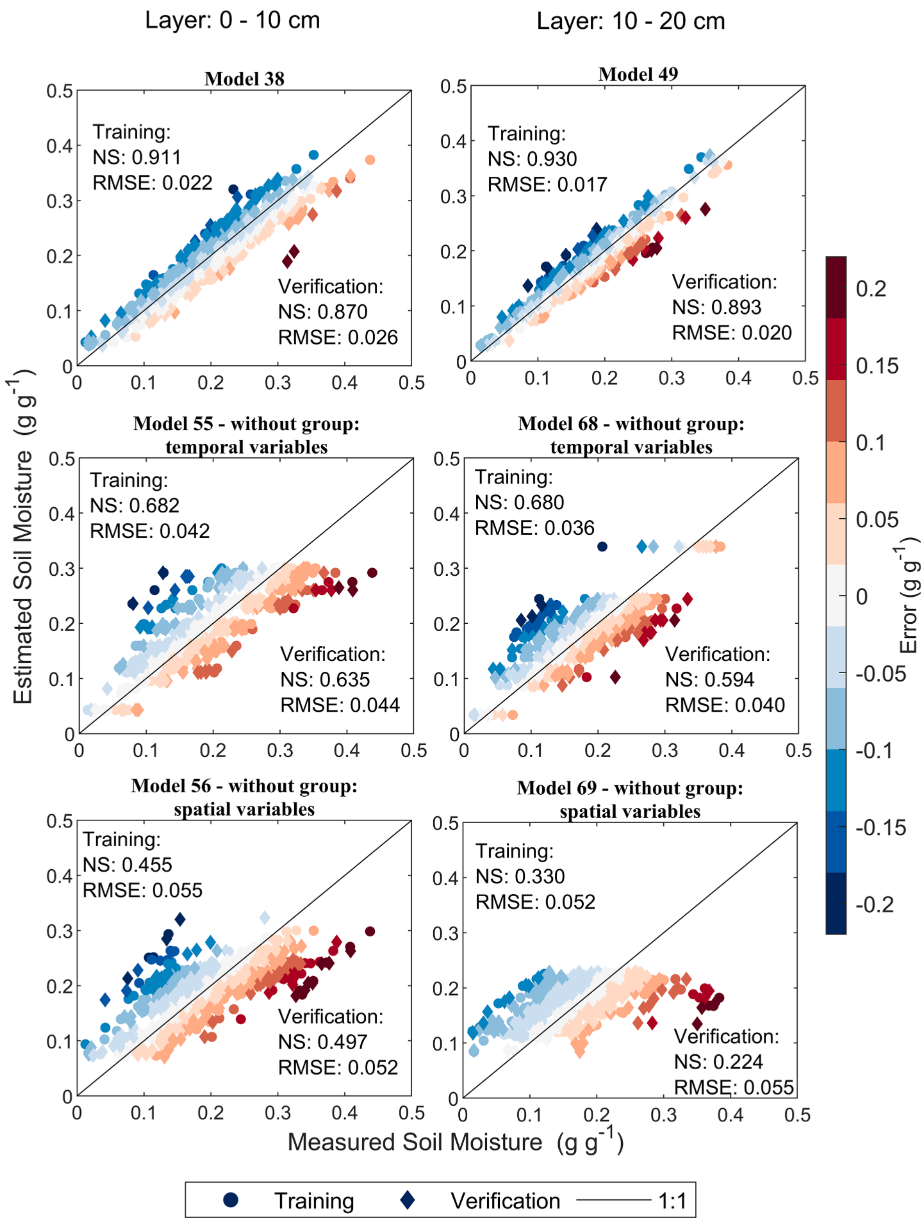

Fig. 8. Model performance following removal of temporal and spatial variables. The 1:1 line depicts the ideal adjusted values. For soil moisture, the colder (blue) and hotter (red) colors depict higher negative errors (overestimates) and higher positive errors (underestimates), respectively.

rainfall) and were classified as spatial (having invariant physical characteristics) and temporal (varying over time, such as rainfall and ETo).

For both layers, the complete models showed excellent performance. We then evaluated model performance by removing each one of the variables, categories, or spatiotemporal classes, in turn. The most important variable for the surface model was climate, followed by the EWMA of rainfall. For the subsurface model, climate was also the most important variable (although less so than for the surface model), followed by microporosity. Although all four categories were important for the surface model, the most important was climate, followed by soil properties. For the subsurface model, the most important categories were soil properties, followed by climate; rainfall and topography were of little importance. For both layers, the models were more sensitive to exclusion of spatial than of temporal variables.

In conclusion, it is possible to estimate soil moisture for both layers with good performance, using the selected variables, which represent the physical conditions affecting soil moisture. However, the surface model requires more input variables to achieve good performance. In contrast, for the subsurface model, more variance in soil moisture can be explained using only soil and climate-related variables (in particular, season and microporosity). This is because the most recent rainfall and climate conditions determine changes in surface soil moisture, whereas subsurface soil moisture is attenuated by soil properties, which control water infiltration and delay the effects of rainfall and climate.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank NEPE-HidroSedi at the Federal University of Pelotas (UFPel) for providing facilities to perform laboratory analyses and field surveys, and the Agricultural Meteorology Laboratory of the Brazilian

Corporation of Agricultural Research (EMBRAPA) for providing the meteorological data used in this study.

Funding

We thank the National Council for Scientific and Technological Development (CNPq) for financing a PhD fellowship [grant number 141235/2017-9] for the first author, and research productivity fellowships for the second and third authors. The funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.catena.2021.105631.

References

Al-mukhtar, M., 2016. Modelling the root zone soil moisture using artificial neural networks, a case study. Environ. Earth Sci. 75, 1–12. https://doi.org/10.1007/s12665-016-5929-2.

Alvarez-garreton, C., Ryu, D., Western, A.W., Crow, W.T., Robertson, D.E., 2014. The impacts of assimilating satellite soil moisture into a rainfall – runoff model in a semi-arid catchment. J. Hydrol. 519, 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041.

Arsoy, S., Ozgur, M., Keskin, E., Yilmaz, C., 2013. Geoderma Enhancing TDR based water content measurements by ANN in sandy soils. Geoderma 195–196, 133–144. https://doi.org/10.1016/j.geoderma.2012.11.019.

ASCE, 2000. Artificial Neural Networks in Hydrology I: Preliminary Concepts. J. Hydrol. Eng. 5, 115–123. https://doi.org/10.5121/ijsc.2012.3203.

Bartels, G.K., Castro, N.M.dos R., Collares, G.L., Fan, F.M., 2021. Performance of bedload transport equations in a mixed bedrock–alluvial channel environment. Catena 199, 105108. https://doi.org/10.1016/j.catena.2020.105108.

Bartels, G.K., Terra, V.S.S., Cassalho, F., Lima, L.S., Reinert, D.J., Collares, G.L., 2016. Spatial variability of soil physical and hydraulic properties in the southern Brazil small watershed. African J. Agric. 11, 5036–5042. https://doi.org/10.5897/AJAR2016.11812.

Berthet, L., Andréassian, V., Perrin, C., Javelle, P., 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting ? Comparison of event-based and continuous approaches on 178 catchments. Hydrol. Earth Syst. Sci. 13, 819–831. https://doi.org/10.5194/hess-13-819-2009.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834.

Blake, G.R., Hartge, K.H., 1986. Bulk Density. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods, Madison, pp. 363–375.

Cho, E., Choi, M., 2014. Regional scale spatio-temporal variability of soil moisture and its relationship with meteorological factors over the Korean peninsula. J. Hydrol. 516, 317–329. https://doi.org/10.1016/j.jhydrol.2013.12.053.

Coenders-Gerrits, A.M., Hopp, L., Savenije, H.H., Pfister, L., 2013. The effect of spatial throughfall patterns on soil moisture patterns at the hillslope scale. Hydrol. Earth Syst. Sci. 17, 1749–1763. https://doi.org/10.5194/hess-17-1749-2013.

Contador, J.F.L., Maneta, M., Schnabel, S., 2006. Prediction of Near-Surface Soil Moisture at Large Scale by Digital Terrain Modeling and Neural Networks. Environ. Monit. Assess. 121, 213–232. https://doi.org/10.1007/s10661-005-9116-2.

Cui, Y., Long, D., Hong, Y., Zeng, C., Zhou, J., Han, Z., Liu, R., Wan, W., 2016. Validation and reconstruction of FY-3B/MWRI soil moisture using an artificial neural network based on reconstructed MODIS optical products over the Tibetan Plateau. J. Hydrol. 543, 242–254. https://doi.org/10.1016/j.jhydrol.2016.10.005.

Danielson, R.E., Sutherland, P.L., 1986. Porosity. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods, Madison, pp. 443–461.

Dawson, C.W, Wilby, R.L, 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geog. 25, 80–108. https://doi.org/10.1191/030913301674775671.

Elshorbagy, A., Parasuraman, K., 2008. On the relevance of using artificial neural networks for estimating soil moisture content. J. Hydrol. 362, 1–18. https://doi.org/10.1016/j.jhydrol.2008.08.012.

Famiglietti, J.S., Rudnicki, J.W., Rodell, M., 1998. Variability in surface moisture content along a hillslope transect : Rattlesnake Hill. Texas. J. Hydrol. 210, 259–281. https://doi.org/10.1016/S0022-1694(98)00187-5.

Food and Agriculture Organization of the United Nations, 2014. World reference base for soil resources 2014: International soil classification system for naming soils and creating legends for soil maps. FAO, Rome.

Gao, X., Wu, P., Zhao, X., Wang, J., Shi, Y., Zhang, B., 2013. Estimation of spatial soil moisture averages in a large gully of the Loess Plateau of China through statistical and modeling solutions. J. Hydrol. 486, 466–478. https://doi.org/10.1016/j.jhydrol.2013.02.026.

Gee, G.W., Bauder, J.W., 1986. Particle-size Analysis. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods, Madison, pp. 383–411.

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. SOIL MOISTURE PREDICTION USING SUPPORT VECTOR MACHINES. J. Am. Water Resour. Assoc. 42, 1033–1046. https://doi.org/10.1111/j.1752-1688.2006.tb04512.x.

Grayson, R.B., Western, A.W., Chiew, F.H.S., Blöschl, G., 1997. Preferred states in spatial soil moisture patterns : Water Resour. Res. 33, 2897–2908.

Gwak, Y., Kim, S., 2017. Factors affecting soil moisture spatial variability for a humid forest hillslope. Hydrol. Process. 31, 431–445. https://doi.org/10.1002/hyp.11039.

Hachani, A., Ouessar, M., Paloscia, S., Santi, E., Pettinato, S., 2019. Soil moisture retrieval from Sentinel-1 acquisitions in an arid environment in Tunisia: application of Artificial Neural Networks techniques. Int. J. Remote Sens. 40, 9159–9180. https://doi.org/10.1080/01431161.2019.1629503.

Hagen, K., Berger, A., Gartner, K., Geitner, C., Kofler, T., Kogelbauer, I., Kohl, B., Markart, G., Meißl, G., Niedertscheider, K., 2020. Event-based dynamics of the soil water content at Alpine sites (Tyrol, Austria). CATENA 194, 104682. https://doi.org/10.1016/j.catena.2020.104682.

Hecht-Nielsen, R., 1990. Neurocomputing. Addison - Wesley Publishing Company, Boston.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.

Hu, Q., Pan, F., Pan, X., Hu, L., Wang, X., Yang, P., Wei, P., Pan, Z., 2018. Dry-wet variations and cause analysis in Northeast China at multi-time scales. Theor. Appl. Climatol. 133, 775–786. https://doi.org/10.1007/s00704-017-2222-6.

Hu, W., Cheng, B., 2014. Revealing the relative influence of soil and topographic properties on soil water content distribution at the watershed scale in two sites. J. Hydrol. 516, 107–118. https://doi.org/10.1016/j.jhydrol.2013.10.002.

Huang, X., Shi, Z.H., Zhu, H.D., Zhang, H.Y., Ai, L., Yin, W., 2016. Soil moisture dynamics within soil pro fi les and associated environmental controls. Catena 136, 189–196. https://doi.org/10.1016/j.catena.2015.01.014.

Jacobs, J.M., Mohanty, B.P., Hsu, E., Miller, D., 2004. SMEX02: Field scale variability, time stability and similarity of soil moisture. Remote Sens. Environ. 92, 436–446. https://doi.org/10.1016/j.rse.2004.02.017.

Kornelsen, K.C., Coulibaly, P., 2014. Root-zone soil moisture estimation using data-driven methods. Water Resour. Res. 50, 2946–2962. https://doi.org/10.1002/2013WR014127.

Korres, W., Reichenau, T.G., Fiener, P., Koyama, C.N., Bogena, H.R., Cornelissen, T., Baatz, R., Herbst, M., Diekkrüger, B., Vereecken, H., Schneider, K., 2015. Spatio-temporal soil moisture patterns – A meta-analysis using plot to catchment scale data. J. Hydrol. 520, 326–341. https://doi.org/10.1016/j.jhydrol.2014.11.042.

Li, X., Zhang, S., Peng, H., Hu, X., Ma, Y., 2013. Agricultural and Forest Meteorology Soil water and temperature dynamics in shrub-encroached grasslands and climatic implications : Results from Inner Mongolia steppe ecosystem of north. Agric. For. Meteorol. 171–172, 20–30. https://doi.org/10.1016/j.agrformet.2012.11.001.

Liang, W.L., 2017. Analysis of the contributions of topographic, soil, and vegetation features on the spatial distributions of surface soil moisture in a steep natural forested headwater catchment. Hydrol. Process. 31, 3796–3809. https://doi.org/10.1002/hyp.11290.

Lv, L., Liao, K., Zhou, Z., Zhu, Q., Shen, C., 2019. Catena Determining hot moments / spots of hillslope soil moisture variations based on high-resolution spatiotemporal soil moisture data. Catena 173, 150–161. https://doi.org/10.1016/j.catena.2018.10.012.

Massari, C., Brocca, L., Moramarco, T., Tramblay, Y., Lescot, J.D., 2014. Advances in Water Resources Potential of soil moisture observations in flood modelling : Estimating initial conditions and correcting rainfall. Adv. Water Resour. 74, 44–53. https://doi.org/10.1016/j.advwatres.2014.08.004.

McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E.C., Franz, T.E., Shi, J., Gao, H., Wood, E.F., 2017. The Future of Earth Observation in Hydrology. Hydrol. Earth Syst. Sci. Discuss. 21, 1–55. https://doi.org/10.5194/hess-2017-54.

McKee, T.B., Doesken, N., Kleist, J., 1993. THE RELATIONSHIP OF DROUGHT FREQUENCY AND DURATION TO TIME SCALES, in: 8th Conference on Applied Climatology. Anaheim, California, pp. 179–184. https://doi.org/10.1002/joc.846.

McCulloch S., Warren, Pitts, Walter, 1943. A logical calculus of the ideas immanent in nervous activity. The B. Math. Biophys. 5 (4), 115–133. https://doi.org/10.1007/BF02478259.

Meng, S., Xie, X., Liang, S., 2017. Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. J. Hydrol. 550, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024.

Metzger, J.C., Wutzler, T., Valle, N.D., Grauer, C., Lehmann, R., Roggenbuck, M., Schelhorn, D., Weckmüller, J., Küsel, K., Uwe, K., Susan, T., Hildebrandt, A., 2017. Vegetation impacts soil water content patterns by shaping canopy water fluxes and soil properties. Hydrol. Process. 31, 3783–3795. https://doi.org/10.1002/hyp.11274.

Moore, I.D., Burch, G.J., Mackenzie, D.H., 1988. Topographic Effects on the Distribution of Surface Soil Water and the Location of Ephemeral Gullies. Trans. ASAE 31, 1098–1107. https://doi.org/10.13031/2013.30829.

Moore, R.J., 1980. Real-time Forecasting of Flood Events Using Transfer Function Noise Models: report, Part 2. Institute of Hydrology, Wallingford.

Oliveira, M.H.C., Sari, V., Castro, N.M.dos R., Pedrollo, O.C., 2017. Estimation of soil water content in watershed using artificial neural networks. Hydrol. Sci. J. 62, 2120–2138. https://doi.org/10.1080/02626667.2017.1364844.

Oliveira, V.A., Rodrigues, A.F., Morais, M.A.V., Terra, M.de C.N.S., Guo, L., Mello, C.R., 2021. Spatiotemporal modelling of soil moisture in an Atlantic forest through machine learning algorithms. Eur. J. Soil Sci. ejss.13123 https://doi.org/10.1111/ejss.13123.

Pan, X., Kornelsen, K.C., Coulibaly, P., 2017. Estimating Root Zone Soil Moisture at Continental Scale Using Neural Networks. JAWRA J. Am. Water Resour. Assoc. 53, 220–237. https://doi.org/10.1111/1752-1688.12491.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classificatio. Hydrol. Earth Syst. Sci. 1, 1633–1644. https://doi.org/10.1127/0941-2948/2006/0130.
Philipp, R.P., Bom, F.M., Pimentel, M.M., Junges, S.L., Zvirtes, G., 2016. SHRIMP U-Pb age and high temperature conditions of the collisional metamorphism in the V??rzea do Capivarita Complex: Implications for the origin of Pelotas Batholith, Dom Feliciano Belt, southern Brazil. J. South Am. Earth Sci. 66, 196–207. https://doi.org/10.1016/j.jsames.2015.11.008.
Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., Wendroth, O., 2008. Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. Vadose Zo. J. 7, 358–389. https://doi.org/10.2136/vzj2007.0143.
Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil Moisture Retrieval Using Neural Networks: Application to SMOS. IEEE Trans. Geosci. Remote Sens. 53, 5991–6007. https://doi.org/10.1109/TGRS.2015.2430845.
Rosenbaum, U., Bogena, H.R., Herbst, M., Huisman, J.A., Peterson, T.J., Weuthen, A., Western, A.W., Vereecken, H., 2012. Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. Water Resour. Res. 48, 1–22. https://doi.org/10.1029/2011WR011518.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.
Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., 2016. Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors. Int. J. Appl. Earth Obs. Geoinf. 48, 61–73. https://doi.org/10.1016/j.jag.2015.08.002.
Sari, V., dos Reis Castro, N.M., Pedrollo, O.C., 2017. Estimate of Suspended Sediment Concentration from Monitored Data of Turbidity and Water Level Using Artificial Neural Networks. Water Resour. Manag. 31, 4909–4923. https://doi.org/10.1007/s11269-017-1785-4.
Scaini, A., Sánchez, N., Vicente-Serrano, S.M., Martínez-Fernández, J., 2015. SMOS-derived soil moisture anomalies and drought indices: a comparative analysis using in situ measurements. Hydrol. Process. 29, 373–383. https://doi.org/10.1002/hyp.10150.
Suo, L., Huang, M., Zhang, Y., Duan, L., Shan, Y., 2018. Soil moisture dynamics and dominant controls at different spatial scales over semiarid and semi-humid areas. J. Hydrol. 562, 635–647. https://doi.org/10.1016/j.jhydrol.2018.05.036.
Tayfur, G., Zucco, G., Brocca, L., Moramarco, T., 2014. Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. J. Hydrol. 510, 363–371. https://doi.org/10.1016/j.jhydrol.2013.12.045.
Topp, G.C., Davis, J.L., Bailey, W.G., Zebchuk, W.D., 1984. The measurement of soil water content using a portable TDR hand probe. Can. J. Soil Sci. 64, 313–321.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.
Vogl, T.P., Mangis, J.K., Rigler, A.K., Zink, W.T., Alkon, D.L., 1988. Accelerating the convergence of the back-propagation method. Biol. Cybern. 59, 257–263. https://doi.org/10.1007/BF00332914.
Western, A.W., Grayson, R.B., Blöschl, G., 2002. Scaling of Soil Moisture: A Hydrologic Perspective. Annu. Rev. Earth Planet. Sci. 30, 149–180. https://doi.org/10.1146/annurev.earth.30.091201.140434.
Western, A.W., Zhou, S., Grayson, R.B., Wilson, D.J., Mcmahon, T.A., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. J. Hydrol. 286, 113–134. https://doi.org/10.1016/j.jhydrol.2003.09.014.
Wooldridge, S.A., Kalma, J.D., Walker, J.P., 2003. Importance of soil moisture measurements for inferring parameters in hydrologic models of low-yielding ephemeral catchments. Environ. Model. Softw. 18, 35–48.
Yang, J., Feng, J., He, Z., 2018. Improving soil heat and moisture forecasting for arid and semi-arid regions : A comparative study of four mathematical algorithms. Arid L. Res. Manag. 32, 149–169. https://doi.org/10.1080/15324982.2017.1408716.
Yang, Y., Dou, Y., Liu, D., An, S., 2017. Spatial pattern and heterogeneity of soil moisture along a transect in a small catchment on the Loess Plateau. J. Hydrol. 550, 466–477. https://doi.org/10.1016/j.jhydrol.2017.05.026.
Yao, P., Shi, J., Zhao, T., Lu, H., Al-Yaari, A., 2017. Rebuilding Long Time Series Global Soil Moisture Products Using the Neural Network Adopting the Microwave Vegetation Index. Remote Sens. 9, 35. https://doi.org/10.3390/rs9010035.
Yu, B., Liu, G., Liu, Q., Wang, X., Feng, J., Huang, C., 2018. Soil moisture variations at different topographic domains and land use types in the semi-arid Loess Plateau, China. CATENA 165, 125–132. https://doi.org/10.1016/j.catena.2018.01.020.
Zhong, M., Jiang, T., Hong, Y., Yang, X., 2019. Performance of multi-level association rule mining for the relationship between causal factor patterns and flash flood magnitudes in a humid area. Geomatics. Nat. Hazards Risk 10, 1967–1987. https://doi.org/10.1080/19475705.2019.1655102.
Zhu, H.D., Shi, Z.H., Fang, N.F., Wu, G.L., Guo, Z.L., Zhang, Y., 2014. Soil moisture response to environmental factors following precipitation events in a small catchment. Catena 120, 73–80. https://doi.org/10.1016/j.catena.2014.04.003.