Robust Statistical Processing of Long-Time Data Series to Estimate Soil Water Content

Mirko Anello[1] · Marco Bittelli[2] · Massimiliano Bordoni[3] ·
Fabrizio Laurini[1] · Claudia Meisina[3] · Marco Riani[1] ·
Roberto Valentino[1]

Abstract

The research presented in this paper aims at providing a statistical model that is capable of estimating soil water content based on weather data. The model was tested using a long-time series of field experimental data from continuous monitoring at a test site in Oltrepò Pavese (northern Italy). An innovative statistical function was developed in order to predict the evolution of soil–water content from precipitation and air temperature. The data were analysed in a framework of robust statistics by using a combination of robust parametric and non-parametric models. Specifically, a statistical model, which includes the typical seasonal trend of field data, has been set up. The proposed model showed that relevant features present in the field of experimental data can be obtained and correctly described for predictive purposes.

✉ Fabrizio Laurini
  fabrizio.laurini@unipr.it

  Mirko Anello
  mirko.anello@unipr.it

  Marco Bittelli
  marco.bittelli@unibo.it

  Massimiliano Bordoni
  massimiliano.bordoni@unipv.it

  Claudia Meisina
  claudia.meisina@unipv.it

  Marco Riani
  marco.riani@unipr.it

  Roberto Valentino
  roberto.valentino@unipr.it

[1] University of Parma, Parma, Italy

[2] University of Bologna, Bologna, Italy

[3] University of Pavia, Pavia, Italy

Ⓐ Springer

1 Introduction

Large-scale quantitative assessment of water resources, which is useful in hydrology, hydrogeology, agriculture, and other fields, is generally carried out using models that take into account soil–atmosphere interaction and the hydraulic behaviour of the soil (Brocca et al. 2007; Koster et al. 2009; Brocca et al. 2014; Mimeau et al. 2021). The shallow part of the soil, which is the most affected by atmospheric variables, is normally unsaturated. Soil water content (SWC) and soil water potential (SWP) are the main variables to be considered in the evaluation of the hydraulic behaviour of unsaturated soil in relation to rainfall events. In fact, such variables are used as input data for different types of physically based models to quantify the soil water balance (Bittelli et al. 2010, 2015).

In particular, SWC is a fundamental property that affects a large variety of biophysical processes, such as seed germination, plant growth, and plant nutrition. Given that it determines water infiltration, percolation, evaporation, and plant transpiration, it is a key variable for computing the soil water budget. Moreover, SWC is an important quantity often required for agricultural practices (tillage, soil fertilization, and irrigation), assessment of drought conditions, estimation of run-off, management of water resources, triggering of shallow landslides, and impact on climatic features of an area (Koster et al. 2004; Liu et al. 2008; Godt et al. 2009; Ahmad et al. 2010).

SWC is also used to model the coupled hydraulic and mechanical behaviour of unsaturated soils in geotechnical problems such as stability analysis of natural slopes, levees, dikes, and dams. With regard to the soil–atmosphere interactions, some researchers demonstrated that SWC might regulate the atmospheric variables that are relevant to the dynamics of storms and occurrence of future rainfall (Eltahir 1998). Soil moisture conditions not only reflect past occurrences of rainfall, but also determine a positive feedback mechanism between soil moisture and subsequent precipitation due to convection-related parameters (Findell and E E, 1997). However, the identification of a relationship between soil moisture and precipitation feedback is not simple, due to a complex interplay between various factors that favour or inhibit convection initiation (Hauck et al. 2011).

Regarding the coupled hydraulic–mechanical behaviour of unsaturated soils in stability analysis of both natural and artificial slopes, many authors have highlighted how small pores in soil induce a strength contribution enabling slope stability even for slopes that are steeper than the soil friction angle. However, such a contribution decreases under increasing water content (Rianna et al. 2014; Leung and Ng 2013). In most slope stability analyses, the behaviour of an unsaturated soil is modelled using the soil water characteristic curve (SWCC), which represents the relationship between SWC and SWP (Rahardjo et al. 2005; Fredlund et al. 2012; Fredlund 2019). In any case, whenever the phenomenon under investigation concerns soil, plants, or atmosphere interactions, the estimation of SWC is very important when direct measurement is not available.

SWC can be measured with a variety of methods in the spatial scale, ranging from a few cubic centimetres (small soil sensors used in the greenhouse or field applications) to kilometres (global microwave satellites). Different time scales can also be employed with measurements that can be performed on a minute-based scale (by using soil sensors) or daily with satellites. When measurement is dependent on the acquisition schedule, it is performed with discontinuous methods, such as ground-penetrating radar (Gerhards et al. 2008). Bittelli (2011) provides a review of the fundamental principles employed for SWC measurement and a discussion about the time and spatial scale measurements. In many practical applications (for instance, irrigation management at the farm scale), soil moisture sensors are not available, and satellite data do not provide the necessary spatial resolution. In this regard, the International Soil Moisture Network aims at collecting data at the global level for a variety of applications in climate science, hydrology, agriculture, and other fields (Dorigo et al. 2021). Additionally, soil moisture modelling and forecast have become important as management tools and require reliable data for model parameterization and testing. Many models are available for quantification of vadose zone processes as discussed in some recent review papers (Vereecken et al. 2016; Zheng et al. 2019).

Prediction methods for the SWC can be grouped into the main categories of data-driven empirical models and process-based models. The data-driven empirical models used for producing soil moisture maps are mostly based on satellite remote sensing data and microwave radar data. They include statistical methods such as Bayesian models (Kim et al. 2017), support vector machines (Yu et al. 2012; Raghavendra and Deka 2014; Liu et al. 2016), multiple linear regression models (Qiu et al. 2003; Jung et al. 2017; Mei et al. 2019; Cai et al. 2019), random forests (Pan et al. 2019), artificial intelligence methods (Nguyen 2022), and artificial neural network algorithms (Zou et al. 2010; Schmidt et al. 2020; Hegazi et al. 2021). Despite the good prediction capabilities of these models, the interpretation of the relationships between one or more predictors and SWC appears rather difficult to interpret from a physical and hydrological point of view (Raghavendra and Deka 2014).

Process-based models focus on the hydrological processes that control the soil moisture transfer mechanisms through physical equations, and calculate the explanatory variables as part of the land surface data assimilation techniques (Dai and Cheng 2022). An extended description of numerical methods and computer code for solving flow equations with process-based models is provided by Bittelli et al. (2015). Observationally obtained factors such as precipitation, atmospheric temperature, and solar radiation can be used for the seasonal dynamic prediction of SWC (Panigrahi and Panda 2003; Bittelli et al. 2010; Valentino et al. 2011; Mo and Lettenmaier 2014).

Process-based models also include numerical models that calculate SWC by solving equations of soil water flow. They are based on water balance parameters and on the main soil hydrological properties, namely the soil water characteristic curve (SWCC) and the hydraulic conductivity function (Van Dam et al. 1997; Šimunek and Van Genuchten 2008). The main advantage of these methods is the physical meaning of the equations used to solve SWC calculations (Lamorski et al. 2013). However, these equations need many soil parameters (hydraulic properties, soil properties, land coverage) that can be difficult to collect over large areas and sometimes require a preliminary calibration of the adopted hydrological parameters (Deng et al. 2011). In

this framework, statistical models based on time series analysis and the adoption of robust statistical analysis are an alternative to process-based modelling and can be used with data that are more easily obtained, such as weather data. Robust statistics is a peculiar branch of statistics: broadly speaking, it is referred to as a collection of methods which provide fully reliable estimates and prediction even in the presence of multiple outliers and large errors in the collected data (Atkinson and Riani 2000; Riani 2004).

The aim of this research is to provide a new statistical model to estimate the SWC within a thickness of 1.4 m from ground level. The rationale is to develop a statistical function linking the quantities involved in both infiltration and evapotranspiration phenomena, namely soil volumetric water content, water potential, air temperature, rainfalls, and solar radiation, but not considering the feedback effect of soil moisture on convection-related parameters. To achieve this goal, a time series of field experimental data was employed. The time series was collected from continuous monitoring over a long period at a test site in Oltrepò Pavese in northern Italy (Bordoni et al. 2021). These data are treated in the framework of robust statistics by using the combination of robust parametric and non-parametric models: a combination of least trimmed squares (LTS) and singular spectrum analysis (SSA).

The paper shows how the proposed model can capture the relevant features present in the data and how it can be used for prediction purposes. The approach is based on models introduced in the paper by Rousseeuw et al. (2019) and uses the MATLAB Flexible Statistics Data Analysis (FSDA) toolbox, which is freely available on the MATLAB marketplace, with fine-tuning on seasonal identifications. Other statistical approaches exist, but none of the available software is sufficiently fine-tuned to handle gross errors or outliers (Hosseini et al. 2015).

The main novelty of the proposed model is its ability to accurately predict the SWC at various soil depths based on daily rainfall data. Among the evaluated meteorological variables that were available in our study, it was found that daily air temperature paired with prior rainfall accumulation was the most important. Therefore, the model is able to self-tune and predict seasonal fluctuations using very few field data. Compared with other models, the proposed model requires very little computational effort and uses readily available input data. These characteristics make it particularly suitable for large-scale implementation in areas with scarce experimental data.

The structure of the paper is as follows: Section 2 illustrates the test site, the available observations, and the processing of field data, while Sect. 3 introduces the model and the methodology for analysing a time series which contains a trend, time-varying multiple seasonal components, and isolated or consecutive outliers. Section 4 shows the results of the methodology application and the comparison between model results and time series of field measurements. The relevant aspects of the methodology and results are discussed in Sect. 4 as well. Finally, Sect. 5 presents the concluding remarks.

2 Data and Methods

2.1 Monitoring Test Site

The selected test site is located near the village of Montuè (Fig. 1) in the north-eastern Oltrepò Pavese (northern Italian Apennines, Lombardy region, northern Italy), within the catchment of Scuropasso creek. The test site is 0.02 km^2 wide and is representative of the main geological and geomorphological features of the study area.

The bedrock is made of gravel, sand, and poorly cemented conglomerates, overlying marls and gypsum (Vercesi and Scagni 1984). The groundwater is characterized by deep water circulation, which is confined in fractured levels located at different depths in the bedrock, without forming a continuous aquifer. The test site faces east, at altitudes ranging between 170 and 210 m a.s.l. The slope steepness is between 26° and 35°, in a very steep range all along the hillslope. The top of the slope is mostly covered by grass and shrubs, while the slope toe is covered by a woodland of black robust trees.

According to Koppen's classification of world climates, the climatic regime is temperate/mesothermal (Csa: Mediterranean hot summer climate), with a mean yearly temperature of 13° C and mean yearly rainfall around 694 mm (Canevino meteorological station, ARPA Lombardia monitoring network).
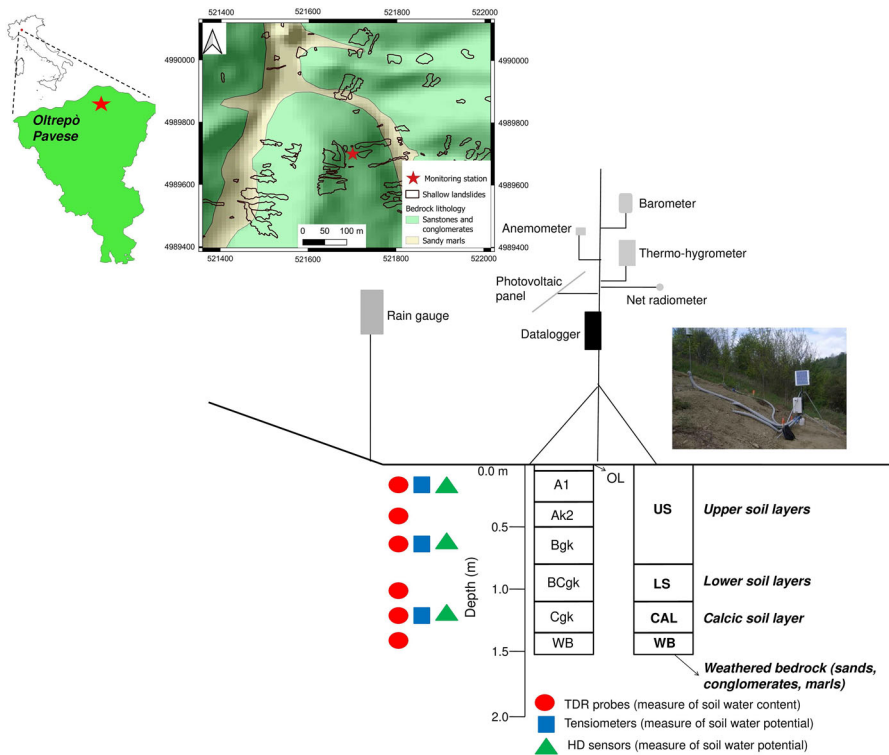

Fig. 1 Location of the site, scheme of the devices, and soil composition

The test site is located in a catchment very prone to shallow landslides. In particular, an extreme rainfall event (160 mm accumulated rain in 62 h) that occurred on 27 and 28 April 2009 triggered many shallow landslides (mean density of 29 landslides per km^2) in the surrounding area (Bordoni et al. 2015) (Fig. 1). The same event caused nine shallow landslides in the test site. This slope was affected by a further shallow failure that occurred between 28 February and 2 March 2014 as a consequence of rainfall of 68.9 mm in 42 h (Bordoni et al. 2015). Shallow landslides on this slope involved areas of a few hundred square metres, with sliding surfaces at 1 m from ground level, mostly corresponding to slope steepness between 30° and 35°.

The shallow landslides involved clayey-sandy silts and clayey-silty sands, which derive from bedrock weathering and are characterized by three main layers (Fig. 1). In the first layer (US), from the ground surface down to 0.7 m, the soil is clayey-sandy silt with low plasticity, high carbonate content, and unit weight between 16.7 and 17.0 kN/m^3. The second soil layer (LS), between 0.7 and 1.1 m from the ground level, has similar characteristics as the US layer but a higher unit weight of 18.6 kN/m^3. At a depth between 1.1 and 1.3 m, the soil has the same textural, plasticity, and density features of the LS layer, but it is characterized by a significant increase in carbonate content up to 35.3%. This layer can be classified as a calcic horizon (CAL), where the carbonate concretions have higher density than in other levels. The weathered bedrock (WB), composed of sand and poorly cemented conglomerates, is positioned 1.3 m below the ground surface. These soil layers are characterized by hydraulic conductivity that decreases as depth increases. Hydraulic conductivity was measured in the field through a compact constant head permeameter (Amoozemeter; Amoozegar 1989). The US layer has the highest value, in the order of 10^{-5} m/s, while LS and CAL are characterized by a saturated hydraulic conductivity equal to 10^{-6} m/s and 10^{-7} m/s, respectively. With regard to the mechanical features of the soils, the peak shear strength parameters were obtained through triaxial tests. The US and LS layers are characterized by similar friction angles between 31° and 33°, and by zero effective cohesion. The CAL layer has a smaller friction angle (26°) than the other layers, but it has effective cohesion of 29 kPa. Moreover, all the soil layers are over-consolidated, as demonstrated by oedometric tests. Table 1 summarizes the main soil features at the Montuè test site.

A monitoring station, which integrates meteorological and hydrological sensors, was installed at the test site in March 2012 (Fig. 1). The meteorological sensors measure rainfall, air temperature, air humidity, atmospheric pressure, wind speed and direction, and net solar radiation. The soil probes measure water content, water potential, and soil temperature. Details on the devices are reported in Table 2.

Hydrological sensors included six time-domain reflectometer (TDR) probes installed at different depths, three jet-fill tensiometers, and three heat dissipation (HD) sensors installed in pairs at three different depths based on the characteristics of the soil layers. Jet-fill tensiometers and HD sensors are in pairs because the jet-fill tensiometer measures soil–water potential higher than −10 J/kg (fewer negative values, lower absolute values), whereas the HD sensor allows one to obtain soil–water potential lower than −10 J/kg (more negative values, higher absolute values). The HD sensor is based on the Flint et al. (2002) equation to convert the measured change in soil temperature after a constant heating period (Bittelli et al. 2012). All field data were collected

Table 1 Description of different soil layers

Layer (-)	Depth (m)	Gravel (%)	Sand (%)	Silt (%)	Clay (%)	w_L (%)	P_I (%)	USCS (-)	γ (kN/m³)	ϕ' (°)	c' kPa	CaCO₃ (%)	K_s (m/s)
US	0.2	12.3	12.5	53.9	21.3	39.8	17.2	CL	17	31	0	14.1	10^{-5}
US	0.4	1.6	11	59.5	27.9	38.5	14.3	CL	16.7			15.7	
US	0.6	8.5	13.2	51.1	27.2	40.3	15.7	CL	16.7			14.1	
LS	1	2.4	12.2	56.4	29	39.2	15.9	CL	18.6	33	0	16.1	10^{-6}
CAL	1.2	0.5	7.5	65.6	26.4	41.8	16.5	CL	18.2	26	29	35.3	10^{-7}
WB	1.4	0.2	75	24.8	0	–	–	SM	18.1	–	–	13.7	–

Layers: US = upper soil; LS = lower soil; CAL = calcic soil; WB = weathered bedrock. Notation: w_L = liquid limit; P_I = plasticity index. USCS (Unified Soil Classification System): CL = clay of low plasticity; SM = silty sand. Physical and mechanical properties: γ = unit weight; ϕ' = friction angle; c' = effective cohesion; CaCO₃ = carbonate content; K_s = saturated hydraulic conductivity

Table 2 Devices and sensors for hydrological monitoring with data logger: No. 1 CR1000X (Campbell Scientific, Inc.)

Meteorological Hydrological indic.	Device	Model	Depth (m)	Accuracy	Range of measure
SWC-Soil temp - SW elect. conduct	TDR probes	CS610, Campbell Sci Inc., Logan, UT	0.2, 0.4, 0.6 1.0, 1.2, 1.4	$0.01-0.02$ m^3/m^3	$0-1\,m^3/m^3$
SW potential (< -10 J/kg)	HD sensors	HD229, Campbell Sci Logan, UT	0.2, 0.6, 1.2	$1.5-2.0$ J/kg	$-10{,}000/ -10$ J/kg
SW potential (> -10 J/kg)	Tensiometer	JetFill 2725, Soilmoist Equipment Corp. Santa Barbara, CA	0.2, 0.6, 1.2	$1.5-2.0$ J/kg	$-80/15$ J/kg
Rainfall	Rain gauge	52,203, Young Comp. Traverse City, MI	–	0.01 mm	0- mm
Air temp./Air hum	Thermo hygrometer	HMP155A, Campbell Sci Inc., Logan, UT	–	0.2°C 1%	$-80/60$°C 0–100%
Atm. pressure	Barometer	CS100, Campbell Sci Inc., Logan, UT	–	0.1 hPa	600–1100 hPa
Wind speed direction	Anemometer	WINDSONIC, Campbell Sci. Inc., Logan, UT	–	0.01 m/s 3°	0–60m/s 0-359°
Net solar radiation	Net radiometer	NR-LITE 2, Kipp&Zonen Delft, Netherlands	–	$3\,W/m^2$	–

SW/SWC stands for "soil water/soil water content"; HD stands for "heat dissipation"

by a data logger powered by a photovoltaic panel and recorded with a frequency of 10 min. A more detailed description of the monitoring station and the probes is reported in Bordoni et al. (2015). As described in the following sections, field-measured data over 8 years relating to both soil hydrological quantities and atmospheric variables (Bordoni et al. 2021) were taken into account for the development of the proposed model.

2.2 Field Data Processing

Field measurements of both soil and atmospheric variables were recorded with a frequency of 10 min, but for the purpose of this research, accumulated hourly data were deemed more appropriate. The final hourly time series presented randomly scattered missing values. This was the first issue to be solved. There are several methods for performing missing replacement, and an interpolation is a common choice. A more robust alternative is to replace the missing data points with the median of a small block of data, using some of the previous and subsequent records. Additional jittering taken from a uniform distribution could be considered if data replacement involves a large chunk of data that would be constant over time.

In the subsequent analysis, daily data are obtained by aggregating or averaging hourly data. Obviously, data with shorter frequency alleviate the arbitrariness underlying the missing data replacement, and both alternatives discussed above result in similar outcomes once daily data are considered.

3 The Statistical Model

Based on time series of field data discussed in Bordoni et al. (2021), the aim of this research is to provide a unified statistical framework for modelling and prediction of SWC at different soil depths. In this section, the statistical features of the data and the structure of the proposed model are discussed. A preliminary discussion is related to the approach followed to validate the model. We split the data into two parts: in the so-called training part, daily time series (21/11/2012 to 31/12/2019) are used to estimate all parameters of the model. Diagnostics in-sample are assessed via residual analysis (see Sect. 4.3). Subsequently, in the testing part, the validation of the model is explored using daily out-of-sample forecasts for the year 2020, with details reported below (see Sect. 4.4). We recall that we have daily data, properly cleaned with robust filters discussed in Sect. 2.2. Field SWC data measured at depths of 0.2 m and 1.2 m are plotted in Fig. 2 in black and blue, respectively. The red vertical lines of Fig. 2 denote the daily cumulative precipitation. A similar plot is presented in Fig. 3, where the red line denotes the daily average temperature.

From visual inspection of both Figs. 2 and 3, it is clear that there is a seasonal variation in SWC at all depths, but whether there is a clear direct link between SWC and atmospheric variables is far from obvious.

Table 3 lists all the variables that were originally available in the data loggers. The superscript in $Y_t^{(m)}$ denotes the value of the outcome at soil depth of m metres. A similar notation is used for the explanatory variables $X_{t,j}^{(m)}$ (with $j = \{1, 2, \ldots, 9\}$). Our aim is to model SWC at a specific soil depth via a minimal set of explanatory
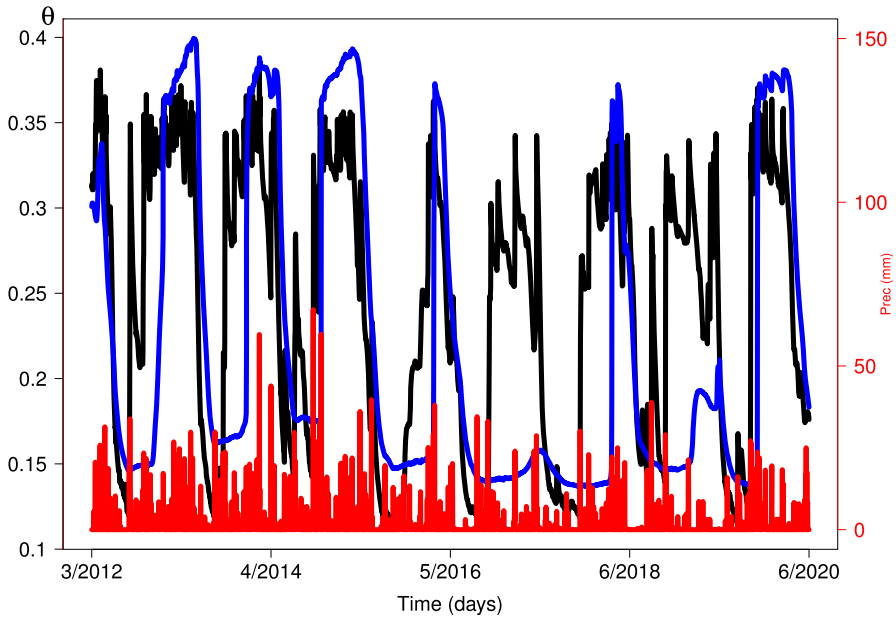
Fig. 2 Time series of daily values for SWC (first axis) at soil depth of 0.2 m (black line), 1.4 m (blue line), and the daily accumulated rain (red line on the second axis)
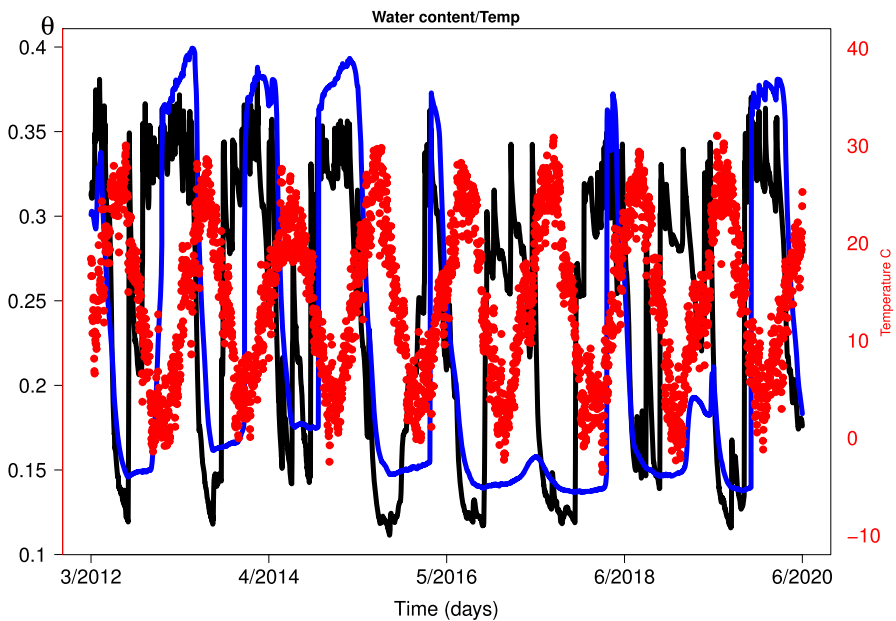

Fig. 3 Time series of daily values for SWC (first axis) at soil depth of 0.2 m (black line), 1.4 m (blue line), and the daily average air temperature

Table 3 Variables included in our dataset

Soil/atmosphere indicator	Measurement levels (soil depth)	Notation used in the paper	Specification daily summary
Soil water content	0.2 m; 0.4 m; 0.6 m	$Y_t^{(02)}; Y_t^{(04)}; Y_t^{(06)}$	Average WC in a day t
(SWC)	1 m; 1.2 m; 1.4 m	$Y_t^{(10)}; Y_t^{(12)}; Y_t^{(14)}$	At specific soil depth
Soil temperature	0.2 m; 0.6 m; 1.2 m	$X_{t,1}^{(02)}; X_{t,1}^{(06)}$	Average TSOIL in a day t
(TSOIL)		$X_{t,1}^{(12)}$	at specific soil depth
Soil water potential	0.2 m; 0.6 m; 1.2 m	$X_{t,2}^{(02)}; X_{t,2}^{(06)}$	Average Pwp in a day t
(Pwp)		$X_{t,2}^{(12)}$	At specific soil depth
Atmospheric pressure	Ground level	$X_{t,3}$	Average atmospheric
(Patm)			pressure in a day t
Air temperature	Ground level	$X_{t,4}$	Average temperature
(Tair)			in a day t
Air humidity	Ground level	$X_{t,5}$	Average air humidity
(Hum)			in a day t
Wind speed	Ground level	$X_{t,6}$	Average wind speed
(WSp)			in a day t
Precipitation	Ground level	$X_{t,7}$	Cumulative mm of rains
(Cum. rain)			in a day t
Wind direction	Ground level	$X_{t,8}$	Most prevalent
(Wdir)			direction in a day t
Solar radiation	Ground level	$X_{t,9}$	Average solar radiation
(Solar)			in a day t

For each variable we have daily data, and the subscript t indicates that a time series is available for all variables. Many variables will not be included in our model

variables that are easy to obtain. By "easy to obtain", we mean that such variables do not require the installation of specific devices in the soil.

The pairwise scatter of daily data does not suggest any specific relationship between the available variables. On the contrary, the time series plot shows some regularity, mostly related to seasonal factors and common trends among the variables. The building bricks of the proposed model are formulated by the regression-like expression

$$Y_t^{(m)} = c_0 + \sum_{a=0}^{A} \alpha_a t^\alpha + \sum_{j=1}^{P} \theta_j X_{t,j}^{(m)} + \left[\sum_{b=1}^{B} \beta_{b,1} \cos(\omega_b t) + \sum_{b=1}^{B} \beta_{b,2} \sin(\omega_b t) \right]$$

$$\left(1 + \sum_{g=1}^{G} \gamma_g t^g \right) + \delta_1 \mathbf{I}(t \geq \delta_2) + W_t. \tag{1}$$

Details and rationales of model (1) are discussed for monthly data in Rousseeuw et al. (2019), and here we revise the most important features. The model has four main components: polynomial time parameters for long-term trends, denoted by α_a; linear effect of time-varying explanatory variables with coefficients θ_j, and the same notation

is used when the explanatory variables of Table 3 do not have the superscript (m) or they have a lag k effect, that is, when $X_{t-k,j}$ is considered; seasonality term modelled by trigonometric waves with coefficients $\beta_{b,1}$ and $\beta_{b,2}$, having time-varying magnitude driven by γ_g; and finally, a level shift is included in the case of a major sudden level break located at time δ_2, with magnitude δ_1. A minor comment is warranted for $\omega_b = 2b\pi/T$, where T is the length of the time period (1 year of daily data, so $T = 365.25$), implying that ω_b is driven by the time-frequency of the recorded data.

For the random disturbance W_t we assume a Gaussian-like distribution with 0 mean and finite variance σ_W^2. Despite the non-linear structure, the model introduced in Eq. (1) can be recast into a regression-like framework and enjoy simplicity of estimation coupled with robustness (see Sect. 2.2 in Rousseeuw et al. (2019) for further details). One can note the presence of an intercept c_0. Additionally, it can happen that there is a "lag effect" of the explanatory variables on the $Y_t^{(m)}$, and in that case, the explanatory variable will be written, for example, like $X_{t-k,j}^{(m)}$, with integer $k > 1$ (with superscript (m) removed when the explanatory variable is related to ground-level measurements).

Model (1) is fitted to all soil depths of $Y_t^{(m)}$ and, for each single analysis, a careful variable selection is performed. A relatively common structure considers as significant only two predictors: the daily average air temperature $X_{t,4}$ and the cumulative daily lagged rain $X_{t-k,7}$, with the value of k depending on the soil depth under investigation. The seasonal sine/cosine waves are significant for values of B in the set $\{1, 2, 3\}$, depending on the soil depth. At first glance it seems that the interaction term between seasonal sine/cosine and polynomial components is unnecessary. Finally, in some cases we also found a significant linear trend, with negative drift, which might suggest global warming issues.

In other words, based on our experimental data, the model introduced in Eq. (1) reduces to the following special case

$$Y_t^{(m)} = c_0 + \sum_{a=0}^{A} \alpha_a t^\alpha + \sum_{j=1}^{P} \theta_j X_{t,j} + \left[\sum_{b=1}^{B} \beta_{b,1} \cos(\omega_b t) + \sum_{b=1}^{B} \beta_{b,2} \sin(\omega_b t) \right] + W_t.$$

(2)

The focus is now on the specific values of unknown parameters for all studied soil depths. Before discussing the features of significant coefficients in each sub-model at a specific depth, we anticipate that the relevant predictors are a mixture of trend-seasonal deterministic components (low-degree polynomial functions and sine/cosine waves) and atmospheric stochastic components, driven by rain and temperature. These findings have important practical implications, as the water content can be estimated with a very minimal set of explanatory variables for which data values are easily retrievable (simple devices installed on the surface). Additionally, due to the availability of existing software such as Weather Generator (Tomei et al. 2022), future scenarios can be easily simulated for long-term assessment.

3.1 Hints from Singular Spectrum Analysis for Seasonal Components

In this work, a very powerful signal processing technique (singular spectrum analysis, SSA) is used to reduce the impact of noise on the measured data and to detect structural
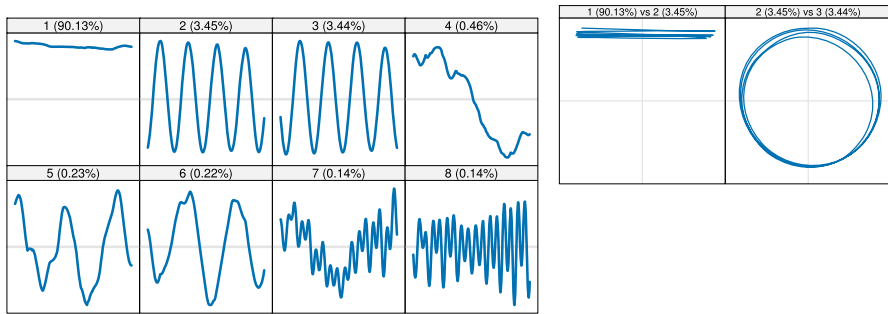
Fig. 4 Eigenvectors of the SSA decomposition for each component and pairwise comparisons. The first three eigenvectors are responsible for almost all signal in the data (about 97% of the signal) and display a steady long-term and marked seasonality

variations in the data (Huffaker et al. 2017). SSA separates time series data into structured variation (signal), including trend and oscillatory components, and unstructured variation (noise). Since the proposed model can be implemented with a different number of oscillatory components (periods), we used the SSA to enable optimal selection of the number of periods. After identifying the proper number of periods contributing to the signal, the result was used to fine-tune the structure of model (1) and to obtain a statistical estimation of the associated parameters.

By using SSA, it was possible to obtain information about which seasonal effects are overwhelming and which are, instead, negligible. From the visual inspection of eigenvectors (individual and pairwise comparisons) of the SSA for SWC at 0.2 m, it appears that there is a strong seasonal pattern and a long-term trend, suggesting that the location under investigation is potentially subject to long-term climate changes. All these findings are visible from inspection of both panels of Fig. 4. Similar results hold for all other SWC depths (not reported).

It is possible to extract the components of SSA for convenient visual inspection of any regularity. As an illustration, we show the extraction of the long-term trend and seasonal components in Fig. 5. In particular, the four panels represent (i) the original series of SWC recorded at 0.2 m; (ii) the trend (whose decline looks linear at first glance and consistent with findings reported in Table 4—see the sign of the estimate of α_1); (iii) the overall effect of the two seasonal components associated with eigenvectors 2 and 3; and (iv) the "residual" part from the decomposition, which still appears to be far from white noise. As stated previously, this issue will be investigated below, in Sect. 4, where some model improvements will be discussed, but other adjustments are subject to further research.

4 Results, Diagnostics, and Validation

4.1 SWC at Superficial Levels: Depth <1 m

We report the results of model fitting for soil depths of SWC located at 0.2, 0.4, and 0.6 m, which we refer to as "superficial levels". We report the estimated parameters of the model (2) after a careful, statistically motivated variable selection in Table 4.
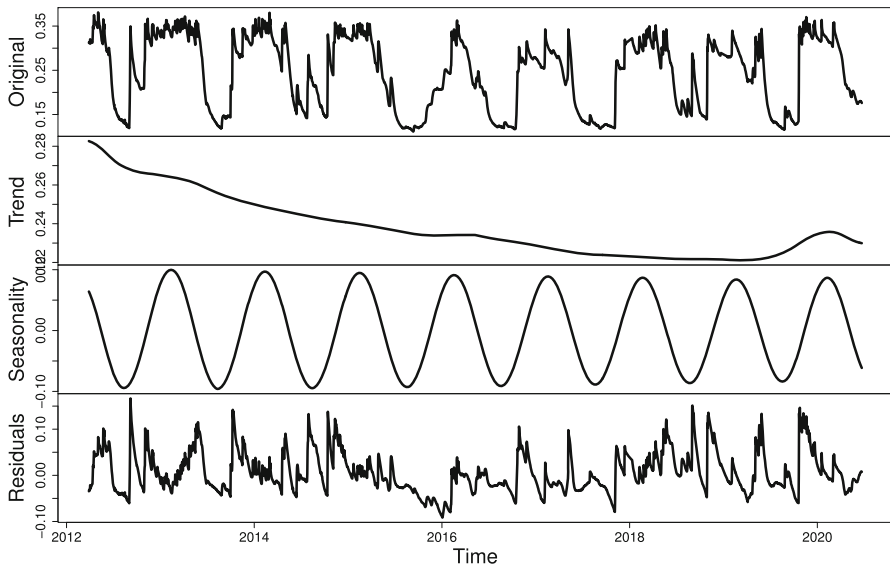
Fig. 5 Plot of original series and reconstruction of components after SSA for SWC at 0.2 m. The components responsible for the overall signal are the long-term trend and two seasonal components

Table 4 Significant variables included for SWC at depth of 0.2 m

Depth 0.2 m	Estimate	S.E.	t-stat	p-value
c_0	2.3E−01	3.7E−03	6.2E+01	0.0E+00
α_1	−3.3E−06	9.7E−07	−3.4E+00	7.5E−04
$\beta_{1,1}$	5.4E−02	2.6E−03	2.0E+01	1.3E−85
$\beta_{1,2}$	4.2E−02	1.0E−03	4.1E+01	9.8E−281
θ_7 for $X_{t-51,7}$	4.3E−04	1.2E−05	3.7E+01	7.3E−241
θ_4 for $X_{t-1,4}$	−1.5E−03	2.3E−04	−6.4E+00	2.2E−10

Using the training data, the adjusted R^2 value for all the fitted models considered here is around 0.7 (or even larger), with better performance at more superficial levels. In all cases, there is a temporal correlation in the residuals, and we provide comments on this evidence below.

From a temporal viewpoint, the most important findings are the presence of a negative linear trend and the presence of a single sine/cosine wave, implying one strong seasonality pattern. At the superficial depth of 0.2 m, there is a positive effect of accumulated rainfall, which lagged at about 50 days. In other words, the contribution of accumulated rainfall is strongest with a lag of approximately 50 days, implying that the amount of SWC at day t is mostly driven by the accumulated rainfall over the prior 50 days. This last piece of evidence indicates a positive effect and relatively long persistence of accumulated rainfall, holding constant the effect of all the other explanatory variables. This finding is not new, and one of the first attempts of modelling this persistence dates to Yu and Cruise (1982).

Table 5 Significant variables included for SWC at depth of 0.4 m

Depth 0.4 m	Estimate	S.E.	t-stat	p-value
c_0	2.5E−01	3.4E−03	7.3E+01	0.0E+00
α_1	−1.2E−05	9.0E−07	−1.3E+01	2.9E−38
$\beta_{1,1}$	7.4E−02	2.4E−03	3.1E+01	9.0E−176
$\beta_{1,2}$	2.7E−02	8.5E−04	3.2E+01	1.1E−185
$\beta_{2,1}$	−8.2E−03	8.8E−04	−9.3E+00	4.6E−20
$\beta_{2,2}$	−9.7E−03	8.8E−04	−1.1E+01	1.2E−27
$\beta_{3,1}$	−8.3E−03	8.5E−04	−9.8E+00	2.1E−22
$\beta_{3,2}$	−2.7E−03	8.3E−04	−3.3E+00	1.1E−03
θ_7 for $X_{t-63,7}$	2.9E−04	1.0E−05	2.8E+01	5.9E−147
θ_4 for $X_{t-1,4}$	−6.0E−04	2.1E−04	−2.9E+00	4.3E−03

The temperature at a superficial depth of 0.2 m has a negative effect on the SWC. Stated more precisely, the value of the average air temperature at day $t - 1$ negatively influences the level of SWC. The choice of lagged temperature at $t - 1$ rather than t is for practical use of the model: using the temperature recorded "yesterday" gives no uncertainty on such explanatory variable when daily predictions are sought. Additionally, we report that using $X_{t,4}$ instead of $X_{t-1,4}$ yielded very marginal model improvements.

Similar comments hold for models fitted at depths 0.4 and 0.6 m, reported in Table 5 and Table 6, respectively. The main differences rely on the selection of more involved seasonal effects, as three waves of sine/cosine are found by our variable selection algorithm. The negative gradient of the long-term trend is significant at a depth of 0.4 m and no longer significant at a depth of 0.6 m. We note the longer persistence effect of the accumulated rainfall, which is always positively related to the amount of SWC, but with longer-lasting effects as depth increases, suggesting a longer time span needed for drying the soil. At depths of both 0.4 m and 0.6 m, the effect of the average daily surface temperature is negative, with magnitude decreasing with increasing depth, following the results obtained at 0.2 m. This feature anticipates that the average air temperature might reverse its effect at some stage.

4.2 SWC at Deeper Levels: Depth of 1 m and More

For deeper levels, the structure of the best-fitted model is still in the form of expression (2). Using our robust fit and robust variable selection algorithm, the coefficients are reported in Tables 7, 8, and 9.

The main finding is that the coefficient associated with the air temperature has a positive sign, as we highlighted earlier, and this feature seems to have a natural physical explanation in the interaction between SWC and air temperature. The effect of cumulative rainfall is still significant, but the time lag at which the most important peak is found is longer for this soil depth, suggesting a longer persistence effect at deeper levels than at superficial levels (we find this very sensible). The number of the

Table 6 Significant variables included for SWC at depth of 0.6 m

Depth 0.6 m	Estimate	S.E.	t-stat	p-value
c_0	2.1E−01	2.6E−03	8.0E+01	0.0E+00
$\beta_{1,1}$	6.7E−02	1.8E−03	3.7E+01	8.2E−238
$\beta_{1,2}$	6.2E−02	9.3E−04	6.6E+01	0.0E+00
$\beta_{2,1}$	2.0E−02	7.3E−04	2.7E+01	9.8E−140
$\beta_{2,2}$	7.3E−03	6.8E−04	1.1E+01	2.2E−26
$\beta_{3,1}$	1.2E−02	6.8E−04	1.7E+01	6.7E−62
$\beta_{3,2}$	−1.7E−02	6.8E−04	−2.5E+01	1.6E−127
θ_7 for $X_{t-81,7}$	4.1E−04	6.6E−06	6.2E+01	0.0E+00
θ_4 for $X_{t-1,4}$	−5.0E−04	1.7E−04	−2.9E+00	3.5E−03

Table 7 Significant variables included for SWC at depth of 1 m

Depth 1 m	Estimate	S.E.	t-stat	p-value
c_0	9.5E−02	4.0E−03	2.4E+01	6.1E−115
α_1	−2.7E−06	1.0E−06	−2.6E+00	9.9E−03
$\beta_{1,1}$	6.1E−02	2.2E−03	2.7E+01	3.5E−144
$\beta_{1,2}$	3.5E−02	1.6E−03	2.2E+01	2.6E−99
θ_7 for $X_{t-92,7}$	4.2E−04	9.0E−06	4.7E+01	0.0E+00
θ_4 for $X_{t-1,4}$	1.4E−03	2.2E−04	6.2E+00	5.1E−10

Table 8 Significant variables included for SWC at depth of 1.2 m

Depth of 1.2 m	Estimate	S.E.	t-stat	p-value
c_0	1.4E−01	4.5E−03	3.2E+01	4.5E−188
$\beta_{1,1}$	7.3E−02	2.2E−03	3.3E+01	4.4E−202
$\beta_{1,2}$	8.6E−03	2.6E−03	3.3E+00	1.0E−03
θ_7 for $X_{t-112,7}$	4.9E−04	8.8E−06	5.6E+01	0.0E+00
θ_4 for $X_{t-1,4}$	2.3E−03	2.8E−04	8.3E+00	2.3E−16

multiple seasonal cycles is generally lower than those found at superficial levels, as it appears that only long-term seasonality is found. We found a negative linear trend at 1 m, the magnitude of which is similar to what we have at a depth of 0.6 m. The actual presence of a significant long-term trend would require further investigation, perhaps including more data from several nearby sites.

4.3 Diagnostic Check and Analysis of Residuals

In this section we analyse residuals $e_t = y_t - \hat{y}_t$, $t = 1, 2, \ldots, N$, where N is the sample size used in the fit, and $\hat{y}_t$ are the fitted values after estimating the parameters

Table 9 Significant variables included for SWC at depth of 1.4m

Depth of 1.4 m	Estimate	S.E.	t-stat	p-value
c_0	7.2E−02	4.5E−03	1.6E+01	6.0E−54
$\beta_{1,1}$	3.4E−02	2.8E−03	1.2E+01	7.6E−32
$\beta_{1,2}$	2.5E−02	2.1E−03	1.2E+01	4.8E−32
$\beta_{2,1}$	1.7E−02	1.3E−03	1.3E+01	2.8E−38
$\beta_{2,2}$	1.8E−02	1.2E−03	1.6E+01	7.0E−54
θ_7 for $X_{t-94,7}$	7.9E−04	1.0E−05	7.6E+01	0.0E+00
θ_4 for $X_{t-1,4}$	1.4E−03	2.9E−04	4.9E+00	1.2E−06
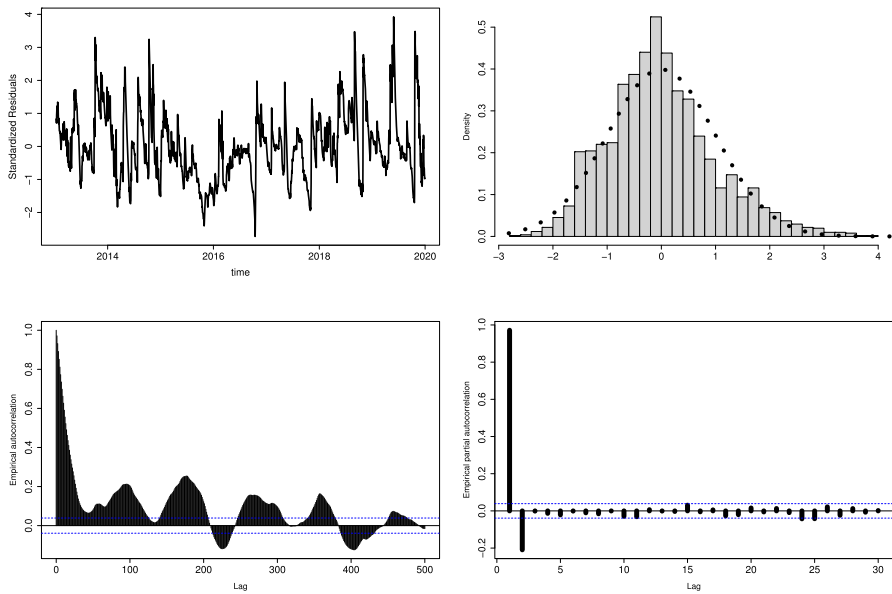
Fig. 6 Time series of daily SWC at 0.2 m split into training (up to the end of 2019) and testing, using a scenario for simulation of precipitation and air temperature

of model (2). We comment only on residuals of SWC at 0.2 m, but results are similar for other depths. Estimated coefficients are reported in Table 4. Residuals are standardized so they have zero mean and unit variance, and it is simpler to contrast their values against quantiles of a standard normal distribution. The comparison against a standard normal is useful for checking marginal features of residuals. Another feature to inspect is the temporal correlation of residuals via the analysis of the empirical autocorrelation and the empirical partial autocorrelation; these diagnostic checks are routinely performed to assess a model's mis-specification, and are all summarized, for example, in Brockwell and Davies (2016)[Sect. 5.3, pp. 144 to 147].

The four panels of Fig. 6 highlight some interesting findings. The plot of residuals over time (top left panel of Fig. 6) shows a pattern that displays some time dependence. Therefore, residuals are not white noise. This is confirmed by the estimates

of both autocorrelation and partial autocorrelation of residuals (bottom left and bottom right panels, respectively, of Fig. 6): these two diagnostic plots suggest that an auto-regressive model should provide some improvement of the fit. The robust fit of model (2) with auto-regressive moving average (ARMA) is still under construction in the FSDA toolbox that we have used for this research. Finally, from the top right panel of Fig. 6 we observe that residuals are not Gaussian, but deviation from normality appears very minor (see the reference dotted line, which is the density of a standard normal). For deeper soil levels, results are broadly similar and thus not reported, but are available upon request.

A concern that needs to be investigated is associated with the "direction" of the model's errors: it seems that the model underestimates some large values, as observed standardized residuals larger than 3 occur quite often (compared with the theoretical normal assumption). As a final summary, the model diagnostics suggest including ARMA components and some adjustment for the possible presence of heavy tails; this will be the subject of further ongoing research.

4.4 Forecast Scenario for 2020

Despite its simplicity, the model demonstrates good performance at all depths, with an average adjusted R^2 exceeding 0.7 for the observed data up to 31 December 2019. As already highlighted in Sect. 4.3, we also noted some serial correlation in the residuals, and approaches for handling this feature will be better investigated and suitably addressed in further research. We now turn to the investigation of a genuine forecast scenario using generated climate data of precipitation and air temperature. The climate data are generated via scenario simulation on a daily basis for precipitation and air temperature for all of 2020. We used the Weather Generator software developed by Tomei et al. (2022) to perform a scenario generation. Here, we discuss in some detail the two most "extreme" cases (i.e. shallowest and deepest soil levels) for illustrative purposes. For all intermediate depths, we show all results and give some comments.

Figure 7 includes the training part and the forecast part, distinguished by a dotted vertical line. In the training part, the agreement between the observed data and the model results appears convincing. In the forecast part (2020), it is possible to see that some sharp observed peaks are not accurately predicted by the model (mostly during the dry period). This is probably due to inappropriate functioning of the field device. This lack of accuracy is still visible at depths of 0.4 m and 0.6 m (see top left and top right panels of Fig. 9).

Subsequently, we used the same generated weather to perform a similar check but at a deeper level (SWC at 1.4 m). For this level, the training part is not fully satisfactory, especially in 2017 and 2019 when periods of severe drought were observed throughout northern Italy. In this case, the forecast is relatively smooth, and the real observed values are in agreement with the model results. Similar arguments hold for levels at 1 m and 1.2 m, which are displayed in the bottom panels of Fig. 9.
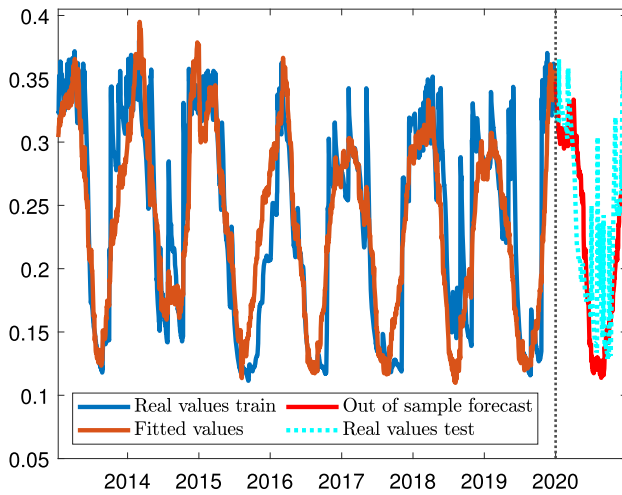
Fig. 7 Time series of daily SWC at 0.2 m split into the training (up to the end of 2019) and testing, using scenario for simulation of precipitation and air temperature
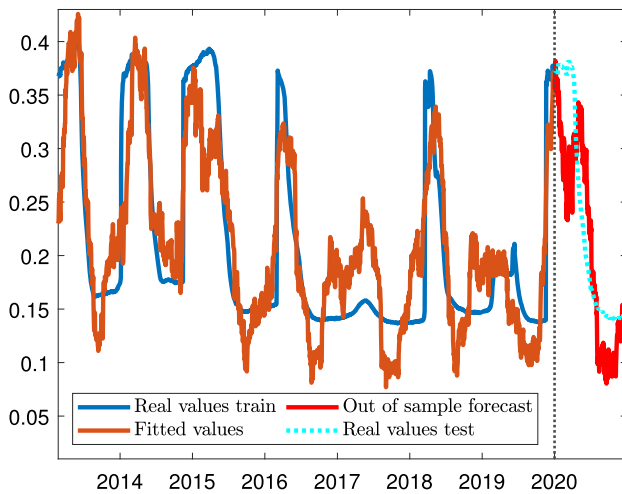

Fig. 8 Time series of daily SWC at 1.4 m split into the training (up to the end of 2019) and testing, using scenario for simulation of precipitation and air temperature

5 Discussion and Final Remarks

We have developed a statistical model to describe the temporal pattern of water content at different depths in soil. SWC is a fundamental variable of water balance in soil, influencing several agronomic, geological, and hydrological processes. The model was developed starting from a dataset of meteorological and hydrological parameters measured by a monitoring station on a hillslope very prone to shallow landslides (Bordoni et al. 2015). In fact, shallow landslide triggering depends strongly on SWC
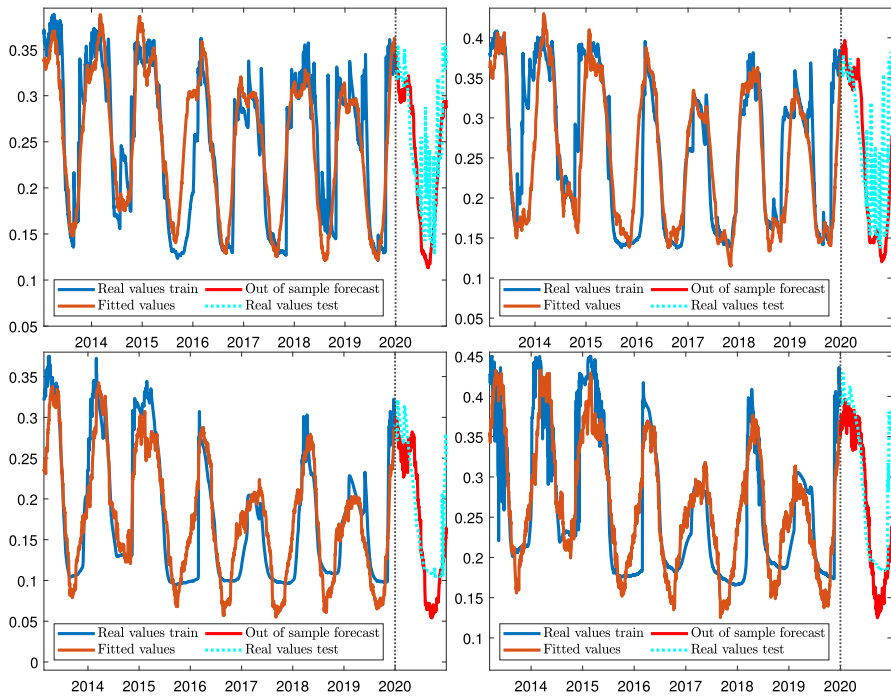
Fig. 9 Time series of daily SWC at several depth, ranging from 0.4 to 1.2 m, with data split into training (up to the end of 2019) and testing, using a scenario for simulation of precipitation and air temperature

values. Shallow failures are triggered when soil approaches or reaches saturated conditions, namely values of SWC close to or equal to the total volume of voids, during or immediately after intense rainfall events (Godt et al. 2009).

Despite the simplicity of the underlying mathematical model, the results obtained are very satisfactory. The use of the proposed model might have benefits in water management and other effects on shallow landslide predictions. We tested our methods using standard goodness-of-fit measures and via a long-term scenario (1 year of daily data).

One of the major benefits of our data-driven approach is the possibility of obtaining accurate daily predictions relying on past data only (i.e., on data that are known without uncertainty). Another benefit is that we require very little physical instrumentation, none of which is located underground, making the water content estimation feasible for very large audiences.

There are limitations in our study, and we left some issues open to further research, some of which are currently under investigation in parallel research projects. From a statistical viewpoint, the selected models all display some correlation in the residuals, and this suggests a more involved time series modelling. We try to fix this feature by adding some with an ARMA structure (and their seasonal generalization), but that quickly turns into an over-fitting. Additionally, fitting seasonal ARMA models when

outlying observations are included requires specific software, which is not yet available for multiple seasonalities, as we have found in our data via the SSA.

Another feature that we have overlooked is the mutual interaction of SWC at different depths and at different temporal lags. Addressing this multivariate response problem requires methods that generalize those illustrated in Lowther et al. (2020), which for our perspective require some fine-tuning for robustness checks. We believe that the joint modelling of water content at different depths, robust fitting, and software development open an avenue for further research.

In terms of the usability of our approach, we are investigating other sites with different soil types and soil use, retrieving data from official worldwide sources. At the moment, we have evidence that different soil compositions and plants have an effect on SWC and on the speed of drying of the soil. The ability to make valid inferences regarding the specific soil composition and plant coverage would require a larger set of data, which are currently being collected.

Declaration

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Ahmad S, Kalra A, Stephen H (2010) Estimating soil moisture using remote sensing data: a machine learning approach. Adv Water Resour 33:69–80

Amoozegar A (1989) A compact constant-head permeameter for measuring saturated hydraulic conductivity of the vadose zone. Soil Sci Soc Am J 53:1356–1361

Atkinson AC, Riani M (2000) Robust diagnostic regression analysis. Springer, New York

Bittelli M (2011) Measuring soil water content: a review. Hort Tech 48:1–15

Bittelli M, Campbell GS, Tomei F (2015) Soil physics with Python: transport in the soil-plant-atmosphere system. OUP Oxford

Bittelli M, Tomei F, Pistocchi A, Flury M, Boll J, Brooks E, Antolini G (2010) Development and testing of a physically based, three-dimensional model of surface and subsurface hydrology. Adv Water Resour 33:106–122

Bittelli M, Valentino R, Salvatorelli F, Rossi Pisa P (2012) Monitoring soil-water and displacement conditions leading to landslide occurrence in partially saturated clays. Geomorphology 173–174:161–173

Bordoni M, Bittelli M, Valentino R, Vivaldi V, Meisina C (2021) Observations on soil-atmosphere interactions after long-term monitoring at two sample sites subjected to shallow landslides. Bull Eng Geol Environ 80(10):7467–7491

Bordoni M, Meisina C, Valentino R, Lu N, Bittelli M, Chersich S (2015) Hydrological factors affecting rainfall-induced shallow landslides: from the field monitoring to a simplified slope stability analysis. Eng Geol 193:19–37

Brocca L, Camici S, Melone F, Moramarco T, Martínez-Fernández J, Didon-Lescot JF, Morbidelli R (2014) Improving the representation of soil moisture by using a semi-analytical infiltration model. Hydrol Process 28(4):2103–2115

Brocca L, Morbidelli R, Melone F, Moramarco T (2007) Soil moisture spatial variability in experimental areas of central Italy. J Hydrol 333(2–4):356–373

Brockwell PJ, Davies RA (2016) Introduction to time series and forecasting, 3rd edition. Springer

Cai Y, Zheng W, Zhang X, Zhabgzhong L, Xue X (2019) Research on soil moisture prediction model based on deep learning. PLoS ONE 14:e0214508

Dai J, Cheng S (2022) Modeling shallow soil moisture dynamics in mountainous landslide active regions. Front Environ Sci 10:913059

Deng J, Chen X, Du Z, Zhang Y (2011) Soil water simulation and predication using stochastic models based on LS-SVM for red soil region of China. Water Resour Manage 25:2823–2836

Dorigo W, Himmelbauer I, Aberer D, Schremmer L, Petrakovic I, Zappa L, Preimesberger W, Xaver A, Annor F, Ardö J, Baldocchi D, Blöschl G, Bogena H, Brocca L, Calvet JC, Camarero JJ, Capello G, Choi M, Cosh MC, Demarty J, van de Giesen N, Hajdu I, Jensen KH, Kanniah KD, de Kat I, Kirchengast G, Rai PK, Kyrouac J, Larson K, Liu S, Loew A, Moghaddam M, Martínez Fernández J, Mattar Bader C, Morbidelli R, Musial JP, Osenga E, Palecki MA, Pfeil I, Powers J, Ikonen J, Robock A, Rüdiger C, Rummel U, Strobel M, Su Z, Sullivan R, Tagesson T, Vreugdenhil M, Walker J, Wigneron JP, Woods M, Yang K, Zhang X, Zreda M, Dietrich S, Gruber A, van Oevelen P, Wagner W, Scipal K, Drusch M, Sabia R (2021) The international soil moisture network: serving earth system science for over a decade. Hydrol Earth Syst Sci Discussions 2021:1–83

Eltahir E (1998) A soil moisture-rainfall feedback mechanism. Water Resour Res 34:765–766

Findell KL, Eltahir EA (1997) An analysis of the soil moisture-rainfall feedback, based on direct observations from illinois. Water Resour Res 33:725–735

Flint A, Campbell G, Ellett K, Calissendorf C (2002) Calibration and temperature correction of heat dissipation matric potential sensors. Soil Sci Soc Am J 66:1439–1445

Fredlund D (2019) State of practice for use of the soil-water characteristic curve (SWCC) in geotechnical engineering. Can Geotech J 56:1059–1069

Fredlund D, Rahardjo H, Fredlund M (2012) Unsaturated soil mechanics in engineering practice. Wiley, London

Gerhards H, Wollschlager U, Yu Q, Schiwek P, Pan X, Roth K (2008) Continuous and simultaneous measurement of reflector depth and average soil-water content with multichannel ground-penetrating radar. Geophysics 73:J15–J23

Godt J, Baum R, Lu N (2009) Landsliding in partially saturated materials. Geophys Res Lett 36:1–5

Hauck C, Barthlott C, Krauss L, Kalthoff N (2011) Soil moisture variability and its influence on convective precipitation over complex terrain. Q J R Meteorol Soc 137:42–56

Hegazi E, Yang L, Huang J (2021) A convolutional neural network algorithm for soil moisture prediction from sentinel-1 sar images. Remote Sensing 13

Hosseini R, Newlands NK, Dean CB, Takemura A (2015) Statistical modeling of soil moisture, integrating satellite remote-sensing (SAR) and ground-based data. Remote Sensing 7(3):2752–2780

Huffaker R, Bittelli M, Rosa R (2017) Non linear time series analysis with R. Oxford University Press, Oxford

Jung C, Lee Y, Cho Y, Kim S (2017) A study of spatial soil moisture estimation using a multiple linear regression model and modis land surface temperature data corrected by conditional merging. Remote Sensing 9:870

Kim J, Mohanty B, Shin Y (2017) Effective soil moisture estimate and its uncertainty using multimodel simulation based on bayesian model averaging. J Geophys Res: Atmospheres - AGU J 120:8023-8042. ISSN 2169-8996

Koster R, Dirmeyer P, Guo Z, Bonan G, Chan E, Cox P, Gordon C, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu C, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud Y, Taylor C, Verseghy D, Vasic R, Xue Y, Yamada T (2004) Regions of strong coupling between soil moisture and precipitation. Science 305:1138–1140

Koster R, Guo Z, Yang R, Dirmeyer P, Mitchell K, Puma M (2009) On the nature of soil moisture in land surface models. J Clim 22(16):4322–4335

Lamorski K, Pastuszka T, Krzyszczak J, Sławiński C, Witkowska-Walczak B (2013) Soil water dynamic modeling using the physical and support vector machine methods. Vadose Zone J, 42–56

Leung A, Ng C (2013) Seasonal movement and groundwater flow mechanism in an unsaturated saprolitic hillslope. Landslides 10:455–467

Liu D, Mishra A, Yu Z (2016) Evaluating uncertainties in multi-layer soil moisture estimation with support vector machines and ensemble Kalman filtering. J Hydrol 538:243–255

Liu H, Xie D, Wu W (2008) Soil water content forecasting by ANN and SVM hybrid architecture. Environ Monit Assess 143:187–193

Lowther A, Fearnhead P, Nunes M, Jensen K (2020) Semi-automated simultaneous predictor selection for regression-SARIMA models. Stat Comput 30:1759–1778

Mei X, Ma L, Zhu Q, Li B, Zhang D, Liu H, Zhang Q, Gou Q, Shen M (2019) The variability in soil water storage on the loess hillslopes in china and its estimation. CATENA 172:807–818

Mimeau L, Tramblay Y, Brocca L, Massari C, Camici S, Finaud-Guyot P (2021) Modeling the response of soil moisture to climate variability in the mediterranean region. Hydrol Earth Syst Sci 25(2):653–669

Mo K, Lettenmaier D (2014) Hydrologic prediction over the conterminous united states using the national multi-model ensemble. J Hydrometeorol 15:1457–1472

Nguyen T (2022) A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Sci Total Environ 833(155066)

Pan J, Shangguan W, Li L, Yuan H, Zhang S, Lu X, Wei N, Dai Y (2019) Using data-driven methods to explore the predictability of surface soil moisture with fluxnet site data. Hydrol Process 33:2978–2996

Panigrahi B, Panda SN (2003) Field test of a soil water balance simulation model. Agric Water Manag 58:223–240

Qiu Y, Fu B, Wang J, Chen L (2003) Spatiotemporal prediction of soil moisture content using multiple-linear regression in a small catchment of the loess plateau, china. CATENA 54:173–195

Raghavendra S, Deka P (2014) Support vector machine applications in the field of hydrology: a review. Appl Soft Comput 19:372–386

Rahardjo H, Lee T, Leong E, Rezaur R (2005) Response of a residual soil slope to rainfall. Can Geotech J 42:340–351

Riani M (2004) Extensions of the forward search to time series. Stud Nonlinear Dyn Econom, 8

Rianna G, Pagano L, Urciuoli G (2014) Investigation of soil-atmosphere interaction in pyroclastic soils. J Hydrol 510:480–492

Rousseeuw P, Perrotta D, Riani M, Hubert M (2019) Robust monitoring of time series with application to fraud detection. Econom Stat 9:108–121, ISSN 2452-3062

Schmidt A, Mainwaring D, DA M (2020) Development of a tailored combination of machine learning approaches to model volumetric soil water content within a mesic forest in the pacific northwest. J Hydrol 588(125044)

Tomei F, Volta A, Costantini L (2022) Weathergenerator. Technical report, ARPAE-Emilia Romagna, https://github.com/ARPA-SIMC/WeatherGenerator

Valentino R, Montrasio L, Losi GL, Bittelli M (2011) An empirical model for the evaluation of the degree of saturation of shallow soils in relation to rainfalls. Can Geotech J 48:795–809

Van Dam J, Huygen J, Wesseling J, Feddes R, Kabat P, Van Walsum P, Groenendijk P, Van Diepen C (1997) Theory of swap, version 2.0. simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment. Tech Rep Dep Water Resources

Vercesi P, Scagni G (1984) Osservazioni sui depositi conglomeratici dello sperone collinare di stradella. Rendiconti della Società Geologica Italiana 7:23–26

Vereecken H, Schnepf A, Hopmans J, Javaux M, Or D, Roose T, Vanderborght J, Young M, Amelung W, Aitkenhead M, Allison S, Assouline S, Baveye P, Berli M, Brüggemann N, Finke P, Flury M, Gaiser T, Govers G, Ghezzehei T, Hallett P, Hendricks Franssen H, Heppell J, Horn R, Huisman J, Jacques D, Jonard F, Kollet S, Lafolie F, Lamorski K, Leitner D, McBratney A, Minasny B, Montzka C, Nowak W, Pachepsky Y, Padarian J, Romano N, Roth K, Rothfuss Y, Rowe E, Schwen A, Šimůnek J, Tiktak A, Van Dam J, van der Zee S, Vogel H, Vrugt J, Wohling T, Young I (2016) Modeling Soil Processes: Review, Key Challenges, and New Perspectives. Vadose Zone J 15(5)

Yu SL, Cruise JF (1982) Time series analysis of soil moisture data. In: El-Shaarawi A, Esterby S (eds) Time series methods in hydrosciences, volume 17 of developments in water science. Elsevier, pp 600–607

Yu Z, Liu D, Lu H, Fu X, Xiang L, Zhu Y (2012) A multi-layer soil moisture data assimilation using support vector machines and ensemble particle filter. J Hydrol 475:53–64

Zheng W, Zhangzhong L, Zhang X, Wang C, Zhang S, Sun S, Niu H (2019) A review on the soil moisture prediction model and its application in the information system. In: Li D, Zhao C (eds) Computer and computing technologies in agriculture XI. Springer, Cham, pp 352–364

Zou P, Yang J, Fu J, Liu G, Li D (2010) Artificial neural network and time series models for predicting soil salt and water content. Agric Water Manag 97:2009–2019

Šimunek J, Van Genuchten M (2008) Modeling nonequilibrium flow and transport with hydrus. Vadose Zone J 7:782–797