1 # Integrating process-based information into an ANN for root-zone

2 # soil moisture prediction

3 Roiya Souissi[1], Mehrez Zribi[1], Chiara Corbari[2], Marco Mancini[2], Sekhar Muddu[3], Sat Kumar

4 Tomer[4], Deepti B Upadhyaya[3,4], Ahmad Al Bitar[1]

5 [1]CESBIO—Centre d'Etudes Spatiales de la Biosphère, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS,

6 Toulouse, France

7 [2]Department of Civil and Environmental Engineering (DICA), Polytechnic University of Milan, 20133 Milano, Italy

8 [3]Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India

9 [4]Satyukt analytics Pvt Ltd, Sanjay Nagar Main Rd, MET Layout, Bengaluru, Karnataka 560094, India

10 *Correspondence to*: Roiya Souissi (roiya.souissi@cesbio.cnes.fr)

11 **Abstract.** Quantification of root-zone soil moisture (RZSM) is crucial for agricultural applications and soil sciences.

12 RZSM impacts processes such as vegetation transpiration and water percolation. Surface soil moisture (SSM) can be

13 assessed through active and passive microwave remote sensing methods, but no current sensor enables direct RZSM

14 retrieval. Spatial maps of RZSM can be retrieved via proxy observations (vegetation stress, water storage change, and

15 surface soil moisture) or via land surface model predictions. In this study, we investigated the combination of surface

16 soil moisture information with process-based inferred features involving artificial neural networks (ANNs). We

17 considered the infiltration process through the soil water index (SWI) computed with a recursive exponential filter and

18 the evaporation process through the evaporative efficiency computed based on a MODIS remote sensing dataset and

19 simplified analytical model, while vegetation growth was expressed through normalized difference vegetation index

20 (NDVI) time series. Several ANN models with different sets of features were developed. Training was conducted

21 considering in situ stations distributed several areas worldwide characterized by different soil and climate patterns of

22 the International Soil Moisture Network (ISMN), and testing was applied to stations of the same data hosting facility.

23 The results indicate that the integration of process-based features into ANN models increased the overall performance

24 over the reference model level in which only SSM features were considered. In arid and semi-arid areas, for instance,

25 performance enhancement was observed when the evaporative efficiency was integrated into the ANN models. To

26 assess the robustness of the approach, the trained models were applied on observation sites in Tunisia, Italy and South-

27 India that are not part of ISMN. The results reveal that joint use of surface soil moisture, evaporative efficiency, NDVI

28 and recursive exponential filter represented the best alternative for more accurate predictions in the case of Tunisia,

29 where the mean correlation of the predicted RZSM based on SSM only sharply increased from 0.443 to 0.801 when

30 process-based features were integrated into the ANN models in addition to SSM. However, process-based features have

31 no to little added value in temperate to tropical conditions.

32 **Keywords:** root-zone soil moisture, artificial neural networks, evaporative efficiency, exponential filter.

33 **1 Introduction**

34 Soil moisture is a major land parameter integrated into several agricultural, hydrological and meteorological

35 applications (Koster et al., 2004; Anguela et al.,2008) This essential climate variable (ECV) consists of two

36 components, namely, surface soil moisture (SSM) (0–5 cm) and root-zone soil moisture (RZSM) (30 cm to 1 m). The

37 importance of RZSM is mainly highlighted in agricultural applications through vegetation stress and water needs and in

38 carbon and nitrogen cycles, as RZSM influences biogeochemical activities in soil (Martínez-Espinosa et al., 2021).

39 SSM quantification is achieved through three main sources: in situ measurements, model estimates and remote sensing-

40 based products. Microwave remote sensing involving sensors such as the Soil Moisture and Ocean Salinity (SMOS)

41 mission (Kerr et al., 2010), Soil Moisture Active Passive (SMAP) mission (Entekhabi et al.,2010) Advanced

42 Microwave Scanning Radiometer (AMSR) (Owe et al., 2008) and Advanced Scatterometer (ASCAT) (Wagner et al.,

43 2013) has been employed to retrieve SSM at coarse resolutions. Current satellite sensors can only provide surface soil

44 moisture information because of the shallow penetration depth of spaceborne data (on the order of a few centimeters)

45 (Wagner et al., 2007). Fine-spatial resolution synthetic aperture radar (SAR) data can also be applied in synergy with

46 optical data to retrieve soil moisture (Zribi et al., 2011; Hajj et al.,2014; Dorigo et al., 2011), but again for surface soil

47 moisture. The International Soil Moisture Network (ISMN) is an exhaustive data hosting facility focused on soil

48 moisture data and associated ancillary information. The ISMN provides in situ soil moisture measurements collected

49 from operational soil moisture networks worldwide (Dorigo et al., 2011). Various models can be adopted to estimate

50 RZSM, such as land surface models (Surfex (Masson et al., 2013), ISBA (Noilhan et al., 1996), CLM (Oleson et al.,

51 2010), JULES (Best et al., 2011), etc.) or dedicated crop models such as Aquacrop (Raes et al., 2009) or SAFYE

52 (Battude et al., 2017). While these models provide the advantage of physical process-based estimates, these estimates

53 depend on the availability and accuracy of ancillary information. Model predictions are often enhanced by the

54 implementation of data assimilation techniques, such as the land data assimilation system (LDAS) (Sabater et al., 2007;

55 Entekhabi et al., 2020).

56 Data-driven methods such as artificial neural networks (ANNs) have also been commonly applied in hydrology as

57 detailed for instance by the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2020)

58 and in (Tanty el al., 2015). One of their advantages is that these models do not require an explicit model structure to

59 accurately represent the involved hydrological processes but instead construct a relationship between the given inputs

60 and the process of interest. Therefore, ANNs are regarded as dynamic input–output mapping models heavily relying on

61 the provided training data relevant to target values (Pan et al., 2017). Moreover, ANNs only require a one-time

62 calibration to provide soil moisture estimations once instrument data are loaded and thus generate relatively low

63 computational costs (Kolassa et al., 2018). These advantages explain the approach to estimate RZSM based on surface

64 information with ANNs in various methodologies (Pan et al., 2017; Grillakis et al., 2021; Souissi et al., 2020). In this

65 paper, we do not address ANN applications as a model twin where the ANN model is trained on the target for

66 mimicking purposes and subsequently generates predictions while requiring a short computation time or fewer input

67 simplifications. Here, we are instead interested in the adoption of ANNs as independent models trained on in situ

68 observations. Within this context, (Pan et al., 2017) successfully applied an ANN as a model for shallow 20-cm root

69 zone soil moisture prediction with a global correlation coefficient of 0.7. (Grillakis et al., 2021) proposed employing an

70 ANN as a means to calibrate and regionalize the time constant of a recursive exponential filter, which was thereafter

71 applied at the regional scale. A combined implementation of Bayesian probabilistic approach and an ANN to infer

72 RZSM at different depths from optical UAV acquisitions via local training was also applied (Hassan-Esfahani et al.,

73 2017). Multitemporal averaged features to predict RZSM based on only SSM and investigated the transferability of a

74 trained ANN across different climatic conditions globally were proposed in (Souissi et al., 2020). Temporal information

75   can be considered in ANNs through recurrent neural networks (RNNs), long short-term memory (LSTM) architectures

76   (Liu et al., 2021), 1D convolutional neural networks (CNNs), or multitemporal averaging. In (Souissi et al., 2020),

77   median, maximum, and minimum correlation values of 0.77, 0.96, and 0.65 were respectively reported. The use of

78   climatic variables such as precipitation and surface temperature and intrinsic surface properties such as soil texture and

79   land cover has also been considered in ANNs (Liu et al., 2021). The choice of variables depends not only on the data

80   availability but also on the objectives. Finally, ANN-based approaches pertain to the more general term of machine

81   learning (ML) approaches, and within this framework, the random forest approach has been applied to root zone soil

82   moisture prediction (Carranza et al., 2021). The aforementioned studies have investigated the application of multiple

83   information sources to predict root zone soil moisture. The input features are commonly curated for quality, and

84   correlation analysis is conducted to determine the useful inputs, while physical processes are not considered. In this

85   paper, we introduce process-based features based on simplified analytical models representing the major processes

86   contributing to root zone soil moisture dynamics. We investigate the impact of the application of different process-

87   based variables on the precision of RZSM predictions as well as the robustness of our approach. (1) We start from a

88   previously developed MLP model (Souissi et al., 2020), and we extend the feature list to include process-based

89   variables, namely, the soil water index given by a recursive exponential filter, remote sensing-based evaporative

90   efficiency, and NDVI time series. (2) The robustness of the approach is assessed through additional tests involving

91   stations not included in the ISMN database in Tunisia, Italy, and South-India. (3) Climatic analysis is conducted to infer

92   the most indicative process-based features for each climate pattern.

93   **2   Materials and Methods**

94   The proposed methodology entails the construction of several ANN models with both direct (SSM, surface temperature,

95   and NDVI) and intermediate sets of features (soil water index and evaporative efficiency) computed based on

96   simplified analytical models. An overview of the processing configuration is shown in Figure 1.

97

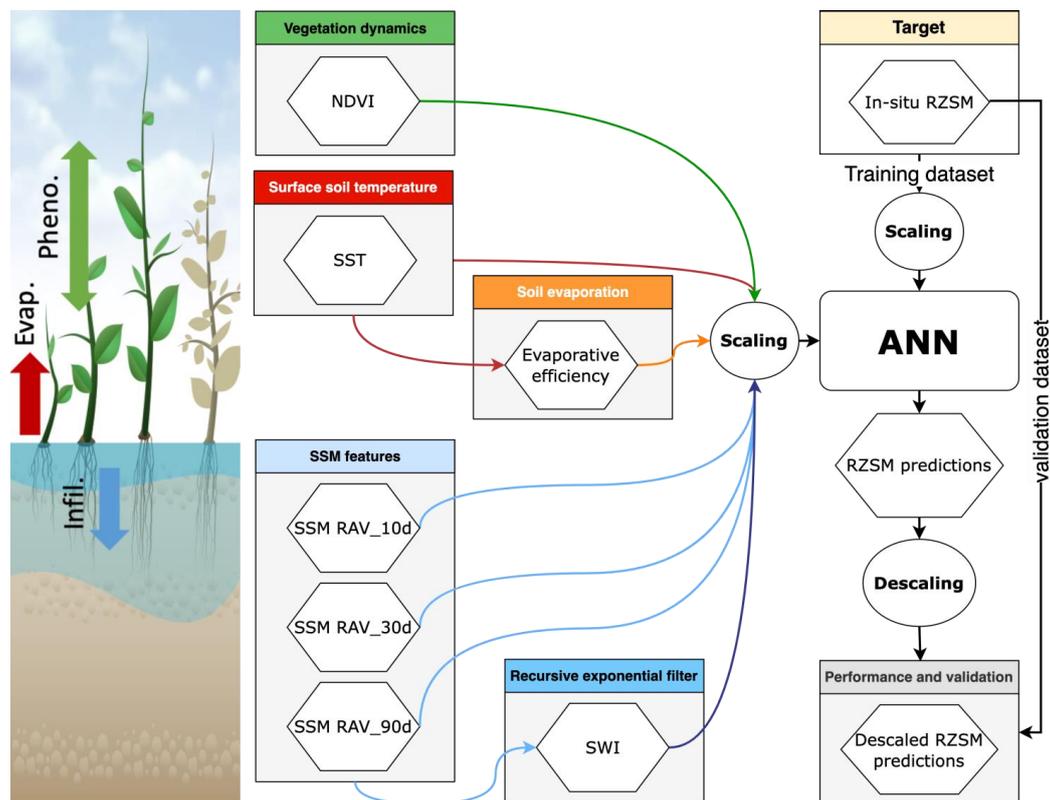



Figure 1. Overview of the processing configuration.

99   This approach results in a combination of ANN models (Table 1). Each model has one physical-process based or a

100 geophysical feature in addition to the three SSM features. All the ANN model hyperparameters remain the same except

101 the number of input features, as described at the end of this section.

102   **Table 1.** ANN model configurations with the respective input variables ([*]: rolling averages of SSM over 10 days; [**]: rolling averages

103 of SSM over 30 days; [***]: rolling averages of SSM over 90 days).

Model Features	SSM_10d_RAV[*]	SSM_30d_RAV[**]	SSM_90d_RAV[***]	SST	NDVI	SWI	EVAP
ANN_SSM	X	X	X				
ANN_SSM_TEMP	X	X	X	X			
ANN_SSM_NDVI	X	X	X		X		
ANN_SSM_EXP-FILT_T5	X	X	X			X	
ANN_SSM_EVAP-EFF_B60	X	X	X				X

ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5	X	X	X	X	X	X

104

The model with the simplest starting point is ANN_SSM based on (Souissi et al., 2020). The most complex model includes the full set of inputs. Intercomparison of the model performance provides information on the added value of each input. All input features are scaled, and training is performed on each of these features based on scaled in situ RZSM data retrieved from the ISMN. The RZSM model predictions are validated against an independent set of observations.

2.1 Datasets

2.1.1 ISMN soil moisture data

The first training and test operations were conducted on eight ISMN networks previously considered in (Souissi et al., 2020). Figure 2 shows the distribution of the considered soil moisture networks with different soil textures and climatic parameters. The selected stations exhibit a root zone depth varying between 30 and 60 cm (Table 2).

115

Figure 2. International Soil Moisture Network (ISMN) network distribution (adapted from the ISMN web data portal (https://www.geo.tuwien.ac.at/insitu/data_viewer/); scale: 1 cm=1000 km).

Table 2. Overview of the considered ISMN and external networks.

Network	Country	Number of Selected Stations	Selected RZSM Depth (cm)	SM Sensors
AMMA-CATCH	Benin, Niger	5 (3 in Benin and 2 in Niger)	40	CS616
BIEBRZA-S-1	Poland	3	50	GS-3
CTP-SMTMN	China	54	40	EC-TM/5TM

HOBE	Denmark	29	55	Decagon-5TE
FR-Aqui	France	5	30, 34, 50	ThetaProbe ML2X
OZNET	Australia	19	30	Hydra Probe-CS616
SCAN	USA	209	50	Hydraprobe-Sdi-12/Ana
SMOSMANIA	France	22	30	ThetaProbe ML2X

119

2.1.2 External soil moisture data

121 The external networks only considered to assess the transferability and robustness of the approach were employed for
122 validation. The trained models are run for predictions only over these sites. They have been selected to cover semi-arid,
123 moderate and tropical semi-arid climates.

124 • Tunisian site: The Merguellil site is located in central Tunisia (9°54 E; 35°35 N). This site is characterized by a
125   semiarid climate with highly variable rainfall patterns, very dry summer seasons, and wet winters. The
126   Merguellil site represents an agricultural region where croplands, namely, olive groves and cereal fields,
127   prevail (Zribi et al., 2021). At this study site, a network of continuous thetaprobe stations installed at bare soil
128   locations provided moisture measurements at depths of 5 and 40 cm. All measurements were calibrated against
129   gravimetric estimations. Data were obtained from the Système d'Information Environmental (SIE) web
130   application catalog.

131 • Italian site: The Landriano site is located in northern Italy. This station is located in a maize field, which was
132   monitored in 2006 and from 2010 to 2011 (Masseroni et al., 2014). The soil texture is sandy loam, and the field
133   is irrigated by the border method with an average irrigation amount of approximately 100 to 200 mm per
134   application with one to two applications per season due to the presence of a shallow groundwater table. Soil
135   moisture measurements were performed with time domain reflectometer (TDR) soil moisture sensors. Five
136   TDR soil moisture sensors were installed along a profile at depths of 5, 20, 35, 50, and 70 cm.

137 • Indian site: The Berambadi watershed is located in Gundalpet taluk, Chamarajanagara district, in the southern
138   part of Karnataka state in India and covers an area of approximately 84 km². The aridity index (P/PET) is 0.7,
139   with an average rainfall of 800 mm/year and a PET value of 1100 mm/year. The climate is classified as Aw,
140   and the major soil types in the region vary between sandy loam, sandy clay loam and sandy clay. Hydrological
141   variables have been intensively monitored since 2009 in the Berambadi watershed by the Environmental
142   Research Observatory ORE BVET and AMBHAS Observatory. The soil moisture levels at the surface (5 cm)
143   and root zone (50 cm) are monitored with a HydraProbe sensor at different agricultural sites across the
144   watershed, and in the current study, 4 stations were chosen. The 3 major cropping seasons include kharif (June
145   to September), during which the first crop is grown, which is usually rainfed during the rabi season (October to
146   January), and summer (February to May), during which the second and third crops are grown, which are usually
147   irrigated. The major crops grown in the region include turmeric, maize, sunflower, marigold and vegetables.

148 **2.1.3 Surface soil temperature**

149 In addition to in situ soil moisture, the ISMN optionally includes meteorological and soil variables that are available

150 over specific time periods. Values of the situ surface soil temperature among these variables can be employed as a

151 useful indicator of the soil moisture data quality. The soil temperature was provided in Celsius, and the plausible values

152 range from -60 to 60 °C. Regarding soil moisture data, surface soil temperature data were also provided with quality

153 flags (Dorigo et al., 2011). However, the drawback is that this variable is not available in all networks, which is the

154 case with the AMMA-CATCH network.

155 **2.1.4 Normalized difference vegetation index**

156 We considered the remote sensing-based normalized difference vegetation index (NDVI) to quantify vegetation

157 dynamics. We extracted this index from the Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation

158 Indices product (MOD13Q1 version 6). MODIS Vegetation Indices (MOD13Q1) version 6 data are generated at 16-

159 day intervals and a 250-m spatial resolution as a Level 3 product. This product provides two primary vegetation layers.

160 The first vegetation layer is the NDVI, which is referred to as the continuity index of the existing National Oceanic and

161 Atmospheric Administration-Advanced Very High Resolution Radiometer (NOAA-AVHRR)-derived NDVI. The

162 algorithm chooses the best available pixel value from all the acquisitions over the 16-day period. The criteria

163 considered are low cloud coverage, low view angle, and highest NDVI value (Didan, 2015). To obtain daily NDVI

164 values, we conducted linear interpolation of the 16-day product.

165 **2.1.5 Potential evapotranspiration**

166 Similarly, we assessed the impact of considering a remote sensing-based evaporative efficiency on RZSM prediction.

167 The computation details of this variable will be detailed later (cf. Section 3). We employed the remote sensing-based

168 potential evapotranspiration (PET) to compute the evaporative efficiency. We extracted the PET from the MOD16A2

169 Evapotranspiration/Latent Heat Flux version 6 product, which is an 8-day composite dataset produced at a 500-m pixel

170 resolution. The algorithm used for MOD16 data product collection is based on the logic of the Penman–Monteith

171 equation, which employs inputs of daily meteorological reanalysis data along with MODIS remote sensing data

172 products such as vegetation property dynamics, albedo, and land cover. The MOD16A2 product provides layers for the

173 composite evapotranspiration (ET), latent heat flux (LE), potential ET (PET) and potential LE (PLE). The pixel values

174 for the PET layer include the sum of all eight days within the composite period (Running et al., 2017). To obtain daily

175 PET values, we performed linear interpolation of the 8-day product.

176 **2.2 Methods**

177 **2.2.1 Recursive exponential filter**

178 Two ANN models presented in Table 1 contained extra knowledge on infiltration process information based on the

179 outputs of the recursive exponential filter (Stroud, 1999) as a feature. The recursive exponential filter was first

180 introduced by (Wagner et al., 1999) to estimate the soil water index (SWI) from surface soil moisture. The equation for

181 the recursive formulation can be written as follows:

182 $$SWI_{m(n)} = SWI_{m(n-1)} + K_n(ms(t_n) - SWI_{m(n-1)}) \ (1)$$

183   where:

184   - $SWI_{m(n)}$ is the soil water index at time t_n,

185   - $ms(t_n)$ is the estimated surface soil moisture at time t_n,

186   - K_n is the gain at time t_n, which occurs in [0,1] and is given by:

187
$$K_n = \frac{K_{n-1}}{K_{n-1} + e^{-\frac{(t_n - t_{n-1})}{T}}} \quad (2) \text{ and}$$

188   -   T is a time constant and is the only required tuning parameter to compute the recursive

189   exponential filter.

190 Regarding T values, we considered an empirical list ([1,3,5,7,10,13,15,20,40,60]), which was partly inspired by (Paulik

191 et al., 2014) ($T \in [1,5,10,15,20,40,60,100]$). Given the list of T values, recursive exponential filter outputs were

192 computed for all of the stations (346 stations) given each T value. Based on the correlation values between the in situ

193 RZSM values and the recursive exponential filter-based RZSM pre-estimates, we established the optimal time variable

194 T, hereafter referred to as T_{best}, for each station.

195 A large proportion of the stations attained an optimal time constant (T_{best}) value equal to 60 days which suggests an

196 abnormally long infiltration time. These stations belong to the SCAN network and exhibit an RZSM acquisition depth

197 of 50 cm, in contrast other networks such as SMOSMANIA, for instance, where RZSM is retrieved at 30 cm. The high

198 values correspond to correlation with seasonal dynamics rather than infiltration processes. This depth could explain the

199 anomalously long infiltration time. This has been demonstrated in (Paulik et al., 2014), who demonstrated that the

200 average T value with the highest correlation (T_{best}) increased with increasing depth of the in situ observations.

201 For comparison purposes, (Paulik et al., 2014) found that 23.98% of the stations achieved T_{best}=5 days, while 21.58% of

202 the stations achieved $T_{best} \geq 60$ days (60 or 100 days).

203 (Albergel et al., 2008) considered an average T_{best} value of 6 days for the SMOSMANIA network. This value

204 represented the average T_{best} value for all stations belonging to the SMOSMANIA network. In our case, the average

205 T_{best} value for all stations of the SMOSMANIA network reached 9 days. In this study, an average T_{best} value could be

206 established for each station or each network. However, this is not relevant to our work because we aimed to evaluate

207 maps of remote sensing data, and thus, we did not compute T_{best} at each location. We fixed the value of T to 5 days as a

208 median infiltration time.

209 **2.2.2 Evaporative efficiency**

210 An ANN model with evaporative efficiency input was also developed. This variable, which is defined as the ratio of the

211 actual to potential soil evaporation, was first introduced in (Noilhan, J. and Planton, 1989; Jacquemin et al., 1990; Lee

212 et al., 1992) and thereafter readapted in (Merlin et al., 2011) to include the soil thickness and is expressed as follows:

213

214
$$\beta_3 = [\tfrac{1}{2} - \tfrac{1}{2}\cos(\pi\theta_L/\theta_{max})]^P \quad for \ \theta_L \leq \theta_{max} \ (3)$$

215
$$\beta_3 = 1 \ for \ \theta_L > \theta_{max}$$

216 where: - θ_L is the water content in the soil layer of thickness L.

217   - P is a parameter computed as follows:

218
$$P = (\tfrac{1}{2} + A_3 \frac{L - L_1}{L_1}) \frac{LE_p}{B_3} \ (4)$$

219 - θ_{max} is the soil moisture at field capacity, as reported in (Noilhan, J. and Planton, 1989; Jacquemin et al.,
220 1990; Lee et al., 1992), or the soil moisture at saturation, as considered in (Merlin et al., 2011). In our case,
221 this variable denotes the maximum soil moisture at each station.

222 - LE_p is the potential evaporation. In our case, we replaced this variable with the potential evapotranspiration
223 (PET) extracted from the MODIS 500-m 8-day product (MOD16A2). P was then replaced by proxy P*. As the
224 ANN model performed its own calibrations on the set of features, this adaptation of the P term did not impact
225 the process.
226 - L_1 is the thinnest represented soil layer, and A_3 (unitless) and B_3 (W/m²) are the two best-fit parameters a
227 priori depending on the soil texture and structure, respectively. As we were interested in the evaporative
228 efficiency at the surface, L=L1=5 cm, P* is thus expressed as:

229
$$P^* = \frac{PET}{2B_3} \quad (5)$$

230 **2.2.3 Artificial neural network implementation**

231 The multilayer perceptron (MLP), which is a multilayer feed-forward ANN, is one of the most widely applied ANNs,
232 mainly in the field of water resources (Abrahart and See, 2007) The multilayer perceptron contains one or more hidden
233 layers between its input and output layers. Neurons are organized in layers such that the neurons of the same layer are
234 not interconnected and that any connections are directed from lower to upper layers (Ramchoun et al., 2016). Each
235 neuron returns an output based on the weighted sum of all inputs and according to a nonlinear function referred to as
236 the transfer or activation function (Oyebode and Stretch, 2019). The input layer, consisting of SSM values and/or other
237 process-based variables, is connected to the hidden layer(s), which comprises hidden neurons. The final ANN-derived
238 estimates of the ANN are given by an activation function associated with the final layer denoted as the output layer,
239 based on the sum of the weighted outputs of the hidden neurons.

240 We started with the ANN model developed in (Souissi et al., 2020), whose architecture consists of one hidden layer of
241 20 hidden neurons, a tangent sigmoid function as the activation function of the hidden layer, a quadratic cost function
242 as the loss function and the stochastic gradient descent (SGD) technique as the optimization algorithm. This model was
243 developed to estimate RZSM based on only in situ SSM information. SSM was not applied as a feature of hourly values
244 but was employed in the form of three features, namely, SSM rolling averages over 10, 30 and 90 days. Two
245 additional ANN models were developed to study, through each model, the impact of the application of the
246 NDVI, which describes vegetation dynamics and the surface soil temperature as features. A model
247 combining surface soil moisture, NDVI, evaporative efficiency and recursive exponential filter was further
248 considered. These ANN models were trained on the 122 stations among the 346 stations of the ISMN based
249 on (Souissi et al., 2020). Training of the above ANN models was conducted considering 70% of these 122
250 stations. Thirty percent was reserved for validation, and testing was conducted at all stations.

251 **3   Results**

252 **3.1 Intercomparison of the ANN models**

253 The generated correlation histograms (Fig. 3) and performance metrics presented in Table 3 demonstrate that

254 integration of the considered process-based features improved the prediction accuracy in certain cases compared to the

255 reference. In terms of the NDVI, 55.56% of the stations attained better correlation values with ANN_SSM_NDVI than

256 those obtained with ANN_SSM. Additionally, 44.44% of the stations achieved a correlation value higher than 0.7 with

257 model ANN_SSM_NDVI, versus 38.41% of the total stations achieving a similar correlation value with model

258 ANN_SSM.

259

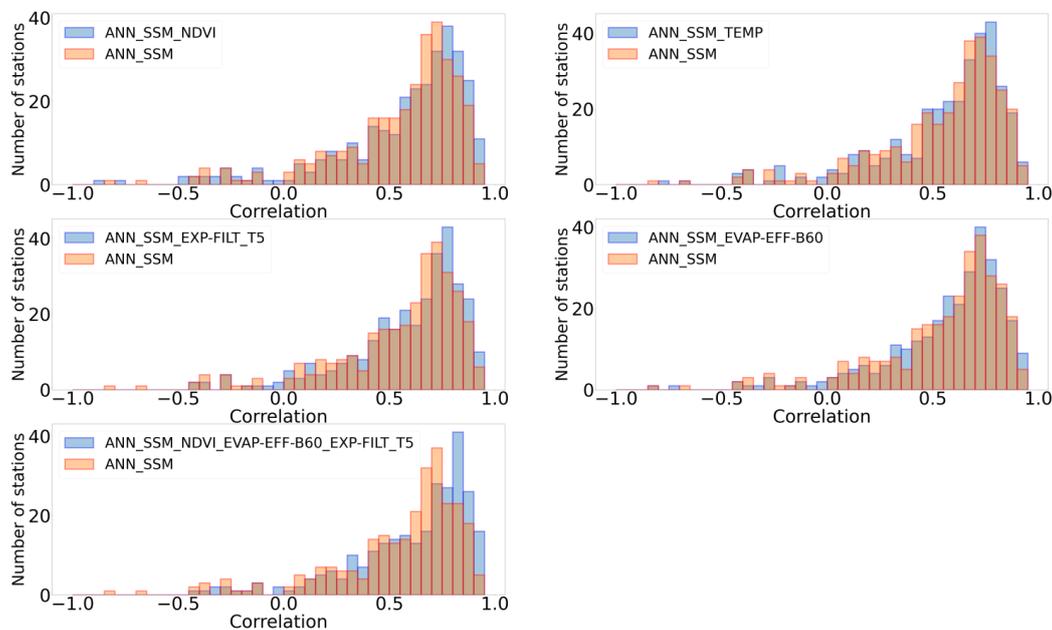



Figure 3. Correlation histograms of (a) ANN_SSM_NDVI, (b) ANN_SSM_TEMP, (c) ANN_SSM_ EXP-FILT_T5, (d)
ANN_SSM_EVAP-EFF-B60, and (e) ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 compared to ANN_SSM.

262 In regard to the ANN_SSM_TEMP model that integrates the soil surface temperature, 54.35% of the stations (except

263 the stations of the AMMA-CATCH network, as no surface temperature data were available) exhibited higher

264 correlation values than those obtained with the ANN_SSM model. Additionally, 40.24% of the stations achieved a

265 correlation value higher than 0.7 with model ANN_SSM_TEMP versus 36.94% of the stations with model ANN_SSM.

266 In addition, model ANN_SSM_EXP-FILT-T5 that integrates the simplified infiltration based features yielded slightly

267 better correlations, and 62.62% of the total stations attained better correlations than those obtained with model

268 ANN_SSM. A total of 45.37% of the stations achieved a correlation value higher than 0.7 with model

269 ANN_SSM_EXP-FILT-T5, in contrast to 38.98% of the stations achieving a similar correlation value with model

270 ANN_SSM.

271 Regarding the evaporative efficiency, we considered different values of fitting parameter B_3 (Eq. 4) such that B_3

272 remained within the [50,60] interval. This parameter can be fitted different variables, such as the wind speed or relative

273 humidity. Comparisons based on the correlation values provided by the different models for each B_3 value indicated

274 that the performance was insensitive to the B_3 value. Thus, we fixed the B_3 value to 60 W m^{-2}. Comparison of models

275 ANN_SSM and ANN_SSM_EVAP-EFF-B60 revealed that 57.89% of the stations attained higher correlation values

276 with the latter model. A total of 41.12% of the stations exhibited a correlation value higher than 0.7 with model

277 ANN_SSM_EVAP-B60 versus 38.48% of the stations with model ANN_SSM.

278 Finally, we investigated the impact of the joint application of the NDVI, recursive exponential SWI (T= 5 days) and

279 evaporative efficiency (B_3=60 W m^{-2}) in the ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 model. The surface

280 soil temperature was not included, as its effect is included in the evaporation process. At 64.6% of the stations, the

281 correlation value obtained with this model was higher than that obtained with the ANN_SSM model. In addition, 51.1%

282 of the stations achieved a correlation value higher than 0.7 with model ANN_SSM_NDVI_EVAP-EFF-B60_EXP-

283 FILT_T5, in contrast to 39.42% of the stations with model ANN_SSM.

284 **Table 3.** Proportion of the stations exhibiting performance enhancement using the ANN models with the process-based features

285 model compared to model ANN_SSM.

Model	% of stations at which the correlation improves over the model ANN_SSM level	% of stations at which RMSE improves over the model ANN_SSM level
ANN_SSM_NDVI	55.56	48.25
ANN_SSM_TEMP'	54.35	46.25
ANN_SSM_EXP-FILT_T5	62.62	51.44
ANN_SSM_EVAP-EFF-B60	57.89	48.68
ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5	64.6	57.3

286

287 **3.2 Robustness of the approach**

288 To assess the robustness of our approach, which involves RZSM prediction using the various ANN models with

289 different features, we predicted RZSM at 40 cm at sites not previously considered in previous parts of the study. The

290 selected stations are located in: the Kairouan Plain, a semiarid region in central Tunisia, Landriano site located in the

291 North of Italy, and the Berambadi watershed located in Gundalpet taluk, South-India. In the case of the Kairouan

292 Tunisia, model ANN_SSM yielded moderate- to low-precision predictions, as highlighted by the performance metrics

293 listed in Table 4. The time series indicated that the RZSM predictions followed the SSM seasonality, which was

294 reflected by the false peaks generated in the RZSM predictions whenever a sharp increase or decrease occurred in the

295 SSM values. This observation was already demonstrated by (Souissi et al., 2020). Actually, the Kairouan Plain is

296 characterized by a semiarid environment where rainfall events infrequently occur and the level of evaporation is high.

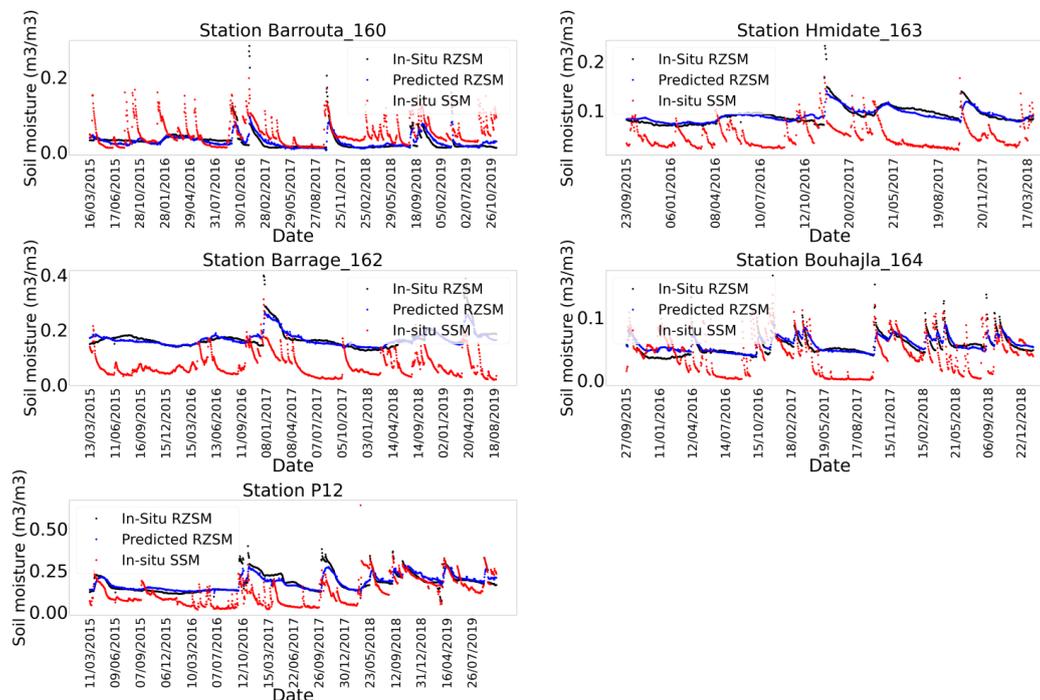



Figure 4. In situ SSM, in situ RZSM, and predicted RZSM series at the stations in the Kairouan Plain (Tunisia) with model ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5.

However, the consideration of additional features, namely, the NDVI, evaporative efficiency and recursive exponential filter SWI, in the ANN models resulted in a good agreement between the in situ and predicted RZSM values (Fig. 4). The correlation values were improved by 60.04%, 169.5%, 112.02%, 80.23% and 53.7% at stations Barrouta-160, Hmidate_163, Barrage_162, Bouhajla_164 and P12, respectively, with the ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 model over ANN_SSM model values. Similarly, RMSE values were reduced (Table 4).

At the South-Indian stations, the ANN_SSM model yielded a good agreement even without the integration of process-based features. The NDVI added little to nonsignificant improvement at station Bheemanbidu. The same observation was made at the Italian site. The application of multiple features performed the best under arid conditions, e.g., in Tunisia. In the tropical and temperate climate regions, this was not the case. The presence of clouds in the MODIS NDVI and potential evapotranspiration products could explain this observation at sites of South-India and North-Italy. In South-India, for instance, the maximum variability in soil moisture occurred during the monsoon season, which is characterized by a large amount of clouds.

Table 4. Performance metrics of models ANN_SSM, ANN_SSM_NDVI and ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 at the sites in central Tunisia, northern Italy and South-India.

Model	ANN_SSM	ANN_SSM_NDVI	ANN_SSM_NDVI_EVAP-EFF_B60_EXP-FILT_T5
		TUNISIA	

Station	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE
Barrouta_160	0.463	0.021	0.395	0.023	0.714	0.016
Hmidate_163	0.318	0.019	0.4	0.018	0.834	0.011
Barrage_162	0.416	0.035	0.457	0.035	0.864	0.019
Bouhajla_164	0.435	0.016	0.385	0.017	0.733	0.01
P12	0.581	0.047	0.578	0.048	0.861	0.029
INDIA						
Station	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE
Madyanahundi	0.813	0.04	0.78	0.042	0.744	0.044
Bheemanbidu	0.76	0.046	0.784	0.044	0.763	0.046
Beechanalli2	0.825	0.038	0.787	0.04	0.743	0.044
Beechanalli1	0.713	0.024	0.713	0.024	0.633	0.025
Italy						
Station	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE
Landriano	0.861	0.038	0.827	0.041	0.841	0.038

314

4 Discussion

316 Climate analysis of the results yielded by the different models indicated that among all models, the climate class with
317 the highest mean correlation change rate (Fig. 5) was class BWk, which regroups desert areas where the link between
318 SSM and RZSM is weak due to high evaporative rates. Class Dfa, which includes areas experiencing harsh and cold
319 winters, also yielded a high mean correlation change rate (>100%). Similarly, at stations of this climate type, the link
320 between the surface and root zone is poor. In regard to class Cfa, in which 88.6% of the total stations belongs to SCAN
321 network, the high mean correlation change rate could be explained by the surface-subsurface decoupling phenomena
322 detected within this network, as previously reported in (Souissi et al., 2020). The model with the largest number of
323 stations with improved predictions over the ANN_SSM model predictions was ANN_SSM_NDVI_EVAP-EFF-
324 B60_EXP-FILT_T5. Actually, the coupled use of process-based features in the ANN models exerted a greater impact
325 on the prediction accuracy than that exerted by the one-at-a-time application of these features. In model
326 ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5, the three process-based features jointly employed seemed to
327 counterbalance the weight of these three SSM features. In this model, the process-based features were equally
328 represented versus the SSM information depicted by these three features. The redundancy of the considered SSM
329 information could explain the limited impact of the one-at-a-time addition of process-based features the joint addition
330 of the three process-based features.

331   In addition, (Karthikeyan and Mishra, 2021) demonstrated that at root depths beyond 20 cm, the importance of SSM

332   was notably lower than that at the 20-cm depth, signifying decorrelation between surface and deeper SM values, which

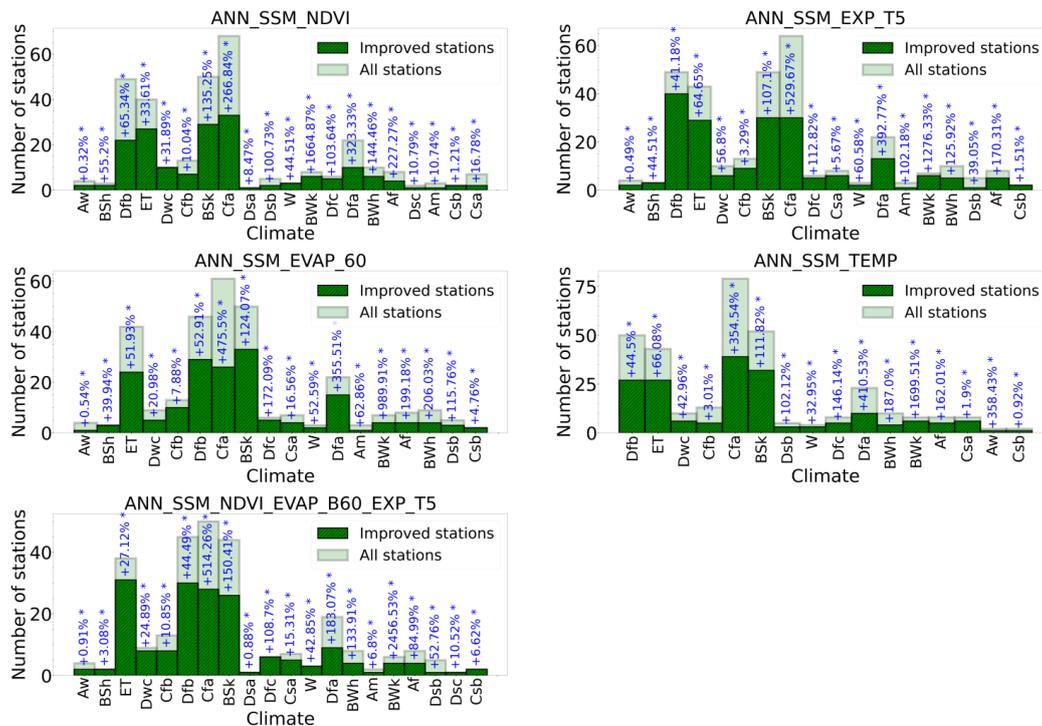
333   is in accordance with the findings in (Souissi et al., 2020), and it was further revealed that vegetation exhibits a higher

334   importance than that of meteorological predictors LST and precipitation. (Kornelsen and Coulibaly, 2014) indicated

335   that evapotranspiration is the most important meteorological input for the prediction of soil moisture in the root zone

336   with the MLP, which reflects the importance of the water vapor flux in soil moisture state determination.

337

338

339   **Figure 5.** Climate classification of the stations performing better with models (a) ANN_SSM_NDVI (b) ANN_SSM_EXP-T5 (c)

340   ANN_SSM_EVAP-60 (d) ANN_SSM_TEMP and (e) ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 compared to model

341   ANN_SSM.

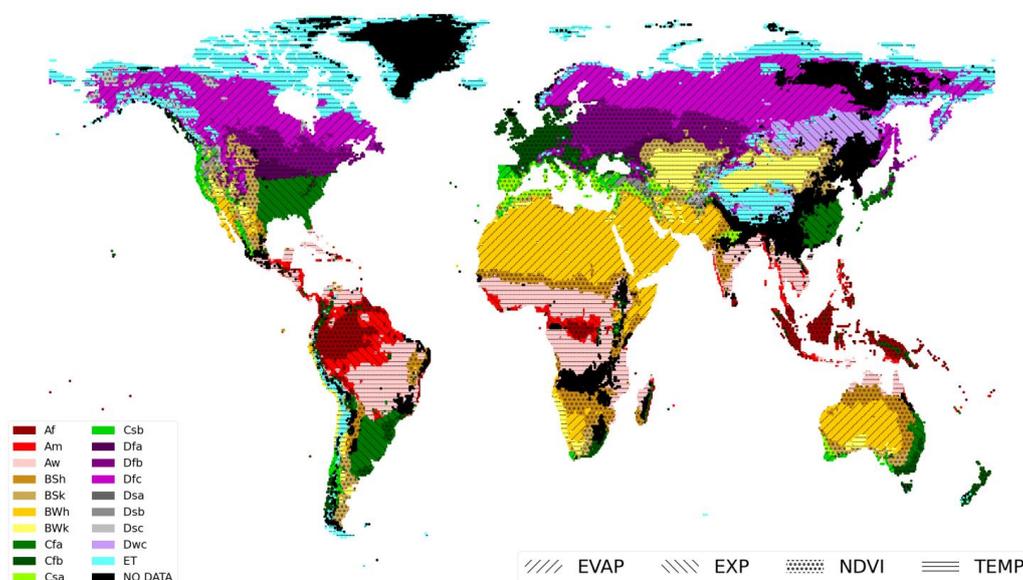
342    $* corr\_change\_rate = mean(\frac{corr_{ANN\_SSM\_X} - corr_{ANN\_SSM}}{corr_{ANN\_SSM}} * 100)$ (6)

343   where X denotes a process-based variable (X ∈ ['NDVI', 'EXP-FILT-T5', 'EVAP-B60', 'TEMP']

344   The map illustrated in Fig. 6, shows the best-performing ANN models based on the mean correlation change rate (Eq.

345   6). We assumed that the results in a given area of a specific climate class could be extended to other areas of the same

346   climate class even if we did not consider the data for these areas. The climate classes without at least one station were

347   marked in black and labeled with 'NO DATA'.

348

349 **Figure 6.** Best-performing ANN models per climate class based on the mean correlation change rate.

350 In arid areas such as the eastern and western sides of the USA with high evaporation rates, ANN_SSM_EVAP-EFF-60

351 was the best performing model. Similarly, in bare areas of Africa, the Middle East and Australia where the Bwh climate

352 class prevailed (arid desert hot climate), the evaporative efficiency was the best informative variable.

353 In the internal part of continental Europe and near the Mediterranean Basin, the NDVI was the most relevant indicator

354 for RZSM estimation, where agricultural fields dominated. Similarly, the Great Plains region in the USA was deeply

355 affected by the NDVI, as this region is a cultivated area. The same result could be obtained for regions belonging to

356 climate class Bsh (arid steppe hot) and mainly covered by grassland and shrubland areas according to ESA CCI land

357 cover maps.

358 In Nordic areas characterized by the ET climate class, the soil temperature was the most important root zone soil

359 moisture indicator mainly because of the freeze–thaw events encountered in these regions. In tropical savannah wet

360 areas (class Aw), the ANN_SSM_TEMP model was the best-performing model.

361 This classification definitely suffered limitations mainly provoked by the generalization of the climatic analysis results

362 to areas not considered in this study. For instance, in regions of climate class Dfc (cold dry without a dry season, cold

363 summer climate), we expect the temperature to serve as the most relevant indicator instead of the evaporative

364 efficiency.

365 **5   Conclusion**

366 In this study, we developed several ANN models to estimate RZSM based either on solely in situ SSM information

367 or on a group of process-based features, in addition to SSM, namely, the soil water index computed with a recursive

368 exponential filter, evaporative efficiency, NDVI and surface soil temperature. Different regions across the globe with

369 distinct land cover and climate patterns were considered. The main conclusion of this study was that the consideration
370 of more features in addition to SSM information could enhance the accuracy of RZSM predictions mainly in regions
371 where the link between SSM and RZSM is weak.

372 In arid areas with high evaporation rates, the most informative feature was the evaporative efficiency. In regions with
373 agricultural fields, the NDVI was, for example, the most relevant indicator to predict RZSM. Overall, the best
374 performing model included the surface soil moisture, NDVI, recursive exponential filter and evaporative efficiency as
375 features. Approximately 61.68% of the tested stations experienced correlation enhancement due to the joint
376 consideration of process-based features over RZSM model predictions based on only surface soil moisture information.

377 The robustness of the approach was further assessed through additional tests considering external sites in central
378 Tunisia, India and Italy. Similarly, the process-based features exerted a positive impact on the prediction accuracy when
379 combined with surface soil moisture in the case of Tunisia. The mean correlation across the five Tunisian stations
380 sharply increased from 0.44 when only SSM was considered to 0.8 when all process-based features were combined
381 with SSM. In India and Italy, the correlations were already high with the reference model ANN_SSM, and the addition
382 of process-based features, namely, NDVI, did not improve the performance potentially because of the cloudy conditions
383 in India and noisy MODIS products.

384 Future work will examine the ability of the developed model to estimate RZSM across larger areas based on remote
385 sensing global soil moisture products. The use of remote sensing derived soil moisture products may yield lower
386 correlations with the reference model ANN_SSM which potentially implies further improvement when process-based
387 features are added.

388 **Acknowledgments**

389 The PhD thesis of R. Souissi was financed by the ERANET RET-SIF project, and complementary financing was
390 provided by the PRIMA Programme SMARTIES project. The authors thank the International Soil Moisture Network
391 (ISMN) and supporting networks for providing the soil moisture data.

392 **References**

393 Abrahart, R. J. and See, L. M.: Neural network modelling of non-linear hydrological relationships, 11, 1563–1579,
394 https://doi.org/10.5194/hess-11-1563-2007, 2007.

395 Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and
396 Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based
397 on in-situ observations and model simulations, 12, 1323–1337, https://doi.org/https://doi.org/10.5194/hess-12-1323-
398 2008, 2008.

399 ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, Artificial Neural Networks in
400 Hydrology. II: Hydrologic Applications, 5, 124–137, https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124), 2000.

401 Battude, M., Al Bitar, A., Brut, A., Tallec, T., Huc, M., Cros, J., Weber, J.-J., Lhuissier, L., Simonneaux, V., and
402 Demarez, V.: Modeling water needs and total irrigation depths of maize crop in the south west of France using high

403   spatial and temporal resolution satellite imagery, Agricultural Water Management, 189, 123–136,

404   https://doi.org/10.1016/j.agwat.2017.04.018, 2017.

405   Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A.,

406   Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and

407   Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water

408   fluxes, 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

409   Carranza, C., Nolet, C., Pezij, M., and van der Ploeg, M.: Root zone soil moisture estimation with Random Forest,

410   Journal of Hydrology, 593, 125840, https://doi.org/10.1016/j.jhydrol.2020.125840, 2021.

411   Didan, K., MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set], NASA

412   EOSDIS Land Processes DAAC, 2015. Available: https://doi.org/10.5067/MODIS/MOD13Q1.006, last access: 2

413   december 2021.

414   Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Drusch, M., Mecklenburg, S., van Oevelen, P.,

415   Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil

416   moisture measurements, Hydrol. Earth Syst. Sci. Discuss., 8, 1609–1663, https://doi.org/10.5194/hessd-8-1609-2011,

417   2011.

418   Entekhabi, D., Nakamura, H., and Njoku, E. G.: Retrieval of soil moisture profile by combined remote sensing and

419   modeling, in: Retrieval of soil moisture profile by combined remote sensing and modeling, De Gruyter, 485–498, 2020.

420   Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S.

421   D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam,

422   M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture

423   Active Passive (SMAP) Mission, Proc. IEEE, 98, 704–716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

424   Fieuzal, R., Baup, F., and Marais-Sicre, C.: Monitoring Wheat and Rapeseed by Using Synchronous Optical and Radar

425   Satellite Data—From Temporal Signatures to Crop Parameters Estimation, ARS, 02, 162–180,

426   https://doi.org/10.4236/ars.2013.22020, 2013.

427   Grillakis, M. G., Koutroulis, A. G., Alexakis, D. D., Polykretis, C., and Daliakopoulos, I. N.: Regionalizing Root-Zone

428   Soil Moisture Estimates From ESA CCI Soil Water Index Using Machine Learning and Information on Soil,

429   Vegetation, and Climate, 57, e2020WR029249, https://doi.org/10.1029/2020WR029249, 2021.

430   Hajj, M., Baghdadi, N., Belaud, G., Zribi, M., Cheviron, B., Courault, D., Hagolle, O., and Charron, F.: Irrigated

431   Grassland Monitoring Using a Time Series of TerraSAR-X and COSMO-SkyMed X-Band SAR Data, Remote Sensing,

432   6, 10002–10032, https://doi.org/10.3390/rs61010002, 2014.

433   Han, H., Choi, C., Kim, J., Morrison, R. R., Jung, J., and Kim, H. S.: Multiple-Depth Soil Moisture Estimates Using

434   Artificial Neural Network and Long Short-Term Memory Models, Water, 13, 2584,

435   https://doi.org/10.3390/w13182584, 2021.

436 Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., and Mckee, M.: Spatial Root Zone Soil Water Content Estimation in
437 Agricultural Lands Using Bayesian-Based Artificial Neural Networks and High- Resolution Visual, NIR, and Thermal
438 Imagery, 66, 273–288, https://doi.org/10.1002/ird.2098, 2017.

439 Jacquemin, B. and Noilhan, J.: Sensitivity study and validation of a land surface parameterization using the HAPEX-
440 MOBILHY data set, Boundary-Layer Meteorol, 52, 93–134, https://doi.org/10.1007/BF00123180, 1990.

441 Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over
442 the United States, Remote Sensing of Environment, 266, 112706, https://doi.org/10.1016/j.rse.2021.112706, 2021.

443 Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N.,
444 Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., and Mecklenburg, S.: The SMOS Mission:
445 New Tool for Monitoring Key Elements ofthe Global Water Cycle, 98, 666–687,
446 https://doi.org/10.1109/JPROC.2010.2043032, 2010.

447 Kolassa, J., Reichle, R. H., Liu, Q., Alemohammad, S. H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T.,
448 Colliander, A., Cosh, M., Holifield Collins, C., Jackson, T. J., Martínez-Fernández, J., McNairn, H., Pacheco, A.,
449 Thibeault, M., and Walker, J. P.: Estimating surface soil moisture from SMAP observations using a Neural Network
450 technique, Remote Sensing of Environment, 204, 43–59, https://doi.org/10.1016/j.rse.2017.10.045, 2018.

451 Kornelsen, K. C. and Coulibaly, P.: Root-zone soil moisture estimation using data-driven methods, Water Resour. Res.,
452 50, 2946–2962, https://doi.org/10.1002/2013WR014127, 2014.

453 Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E.,
454 Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman,
455 A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of Strong Coupling Between
456 Soil Moisture and Precipitation, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

457 Lee, T. J. and Pielke, R. A.: Estimating the Soil Surface Specific Humidity, 31, 480–484, https://doi.org/10.1175/1520-
458 0450(1992)031<0480:ETSSSH>2.0.CO;2, 1992.

459 Liu, Y., Chen, D., Mouatadid, S., Lu, X., Chen, M., Cheng, Y., Xie, Z., Jia, B., Wu, H., and Gentine, P.: Development
460 of a Daily Multilayer Cropland Soil Moisture Dataset for China Using Machine Learning and Application to Cropping
461 Patterns, 22, 445–461, https://doi.org/10.1175/JHM-D-19-0301.1, 2021.

462 Martínez-Espinosa, C., Sauvage, S., Al Bitar, A., Green, P. A., Vörösmarty, C. J., and Sánchez-Pérez, J. M.:
463 Denitrification in wetlands: A review towards a quantification at global scale, Science of The Total Environment, 754,
464 142398, https://doi.org/10.1016/j.scitotenv.2020.142398, 2021.

465 Masseroni, D., Corbari, C., and Mancini, M.: Validation of theoretical footprint models using experimental
466 measurements of turbulent fluxes over maize fields in Po Valley, Environ Earth Sci, 72, 1213–1225,
467 https://doi.org/10.1007/s12665-013-3040-5, 2014.

468 Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel,
469 F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L.,
470 Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C.,

471   Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F.,

472   Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface

473   platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960,

474   https://doi.org/10.5194/gmd-6-929-2013, 2013.

475   Merlin, O., Bitar, A. A., Rivalland, V., Béziat, P., Ceschia, E., and Dedieu, G.: An Analytical Model of Evaporation

476   Efficiency for Unsaturated Soil Surfaces with an Arbitrary Thickness, 50, 457–471,

477   https://doi.org/10.1175/2010JAMC2418.1, 2011.

478   Noilhan, J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, Global and Planetary Change, 13,

479   145–159, https://doi.org/10.1016/0921-8181(95)00043-7, 1996.

480   Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, 117,

481   536–549, https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

482   Oleson, W., Lawrence, M., Bonan, B., Flanner, G., Kluzek, E., Lawrence, J., Levis, S., Swenson, C., Thornton, E., Dai,

483   A., Decker, M., Dickinson, R., Feddema, J., Heald, L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian,

484   T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stockli, R., Wang, A., Yang, Z.-L., Zeng, X., and Zeng, X.:

485   Technical Description of version 4.0 of the Community Land Model (CLM), https://doi.org/10.5065/D6FB50WZ,

486   2010.

487   Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface

488   moisture, J. Geophys. Res., 113, F01002, https://doi.org/10.1029/2007JF000769, 2008.

489   Oyebode, O. and Stretch, D.: Neural network modeling of hydrological systems: A review of implementation

490   techniques, 32, e12189, https://doi.org/10.1111/nrm.12189, 2019.

491   Pan, X., Kornelsen, K. C., and Coulibaly, P.: Estimating Root Zone Soil Moisture at Continental Scale Using Neural

492   Networks, 53, 220–237, https://doi.org/10.1111/1752-1688.12491, 2017.

493   Paris Anguela, T., Zribi, M., Hasenauer, S., Habets, F., and Loumagne, C.: Analysis of surface and root-zone soil

494   moisture dynamics with ERS scatterometer and the hydrometeorological model SAFRAN-ISBA-MODCOU at Grand

495   Morin watershed (France), 12, 1415–1424, https://doi.org/10.5194/hess-12-1415-2008, 2008.

496   Paulik, C., Dorigo, W., Wagner, W., and Kidd, R.: Validation of the ASCAT Soil Water Index using in situ data from

497   the International Soil Moisture Network, International Journal of Applied Earth Observation and Geoinformation, 30,

498   1–8, https://doi.org/10.1016/j.jag.2014.01.007, 2014.

499   Raes, D., Steduto, P., Hsiao, T. C., and Fereres, E.: AquaCrop—The FAO Crop Model to Simulate Yield Response to

500   Water: II. Main Algorithms and Software Description, 101, 438–447, https://doi.org/10.2134/agronj2008.0140s, 2009.

501   Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., and Ettaouil, M.: Multilayer Perceptron: Architecture Optimization

502   and Training, IJIMAI, 4, 26, https://doi.org/10.9781/ijimai.2016.415, 2016.

503   Running, S., Q. Mu, M. Zhao. MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid

504   V006. 2017, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD16A2.006,

505   last access: 2 december 2021.

506   Sabater, J. M., Jarlan, L., Calvet, J.-C., Bouyssel, F., and De Rosnay, P.: From Near-Surface to Root-Zone Soil

507   Moisture Using Different Assimilation Techniques, J. Hydrometeor., 8, 194–206, https://doi.org/10.1175/JHM571.1,

508   2007.

509   SIE, Available: https://osr-cesbio.ups-tlse.fr/, last access: 8 december 2021.

510   Souissi, R., Al Bitar, A., and Zribi, M.: Accuracy and Transferability of Artificial Neural Networks in Predicting in Situ

511   Root-Zone Soil Moisture for Various Regions across the Globe, 12, 3109, https://doi.org/10.3390/w12113109, 2020.

512   Stroud, P. D.: A Recursive Exponential Filter For Time-Sensitive Data, 1999.

513   Tanty, R., Desmukh, T. S., and MANIT BHOPAL: Application of Artificial Neural Network in Hydrology- A Review,

514   IJERT, V4, IJERTV4IS060247, https://doi.org/10.17577/IJERTV4IS060247, 2015.

515   Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational

516   readiness of microwave remote sensing of soil moisture for hydrologic applications, Hydrology Research, 38, 1–20,

517   https://doi.org/10.2166/nh.2007.029, 2007.

518   Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A.,

519   Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L.,

520   Wang, Y., Blöschl, G., Eitzinger, J., and Steinnocher, K.: The ASCAT Soil Moisture Product: A Review of its

521   Specifications, Validation Results, and Emerging Applications, 5–33, https://doi.org/10.1127/0941-2948/2013/0399,

522   2013.

523   Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil

524   Data, Remote Sensing of Environment, 70, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X, 1999.

525   Zribi, M., Chahbi, A., Shabou, M., Lili-Chabaane, Z., Duchemin, B., Baghdadi, N., Amri, R., and Chehbouni, A.: Soil

526   surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation,

527   15, 345–358, https://doi.org/10.5194/hess-15-345-2011, 2011.

528   Zribi, M., Foucras, M., Baghdadi, N., Demarty, J., and Muddu, S.: A New Reflectivity Index for the Retrieval of

529   Surface Soil Moisture From Radar Data, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 14, 818–826,

530   https://doi.org/10.1109/JSTARS.2020.3033132, 2021.

531

532

533

534

535

536 **APPENDIX**

537 **Climate classes (Köppen classification):**

538 • Af: Tropical Rainforest
539 • Am: Tropical Monsoon
540 • As: Tropical Savanna Dry
541 • Aw: Tropical Savanna Wet
542 • BWk: Arid Desert Cold
543 • BWh: Arid Desert Hot
544 • BWn: Arid Desert with Frequent Fog
545 • BSk: Arid Steppe Cold
546 • BSh: Arid Steppe Hot
547 • BSn: Arid Steppe with Frequent Fog
548 • Csa: Temperate Dry Hot Summer
549 • Csb: Temperate Dry Warm Summer
550 • Csc: Temperate Dry Cold Summer
551 • Cwa: Temperate Dry Winter, Hot Summer
552 • Cwb: Temperate Dry Winter, Warm Summer
553 • Cwc: Temperate Dry Winter, Cold Summer
554 • Cfa: Temperate without a Dry Season, Hot Summer
555 • Cfb: Temperate without a Dry Season, Warm Summer
556 • Cfc: Temperate without a Dry Season, Cold Summer
557 • Dsa: Cold Dry Summer, Hot Summer
558 • Dsb: Cold Dry Summer, Warm Summer
559 • Dsc: Cold Dry Summer, Cold Summer
560 • Dsd: Cold Dry Summer, Very Cold Winter
561 • Dwa: Cold Dry Winter, Hot Summer
562 • Dwb: Cold Dry Winter, Warm Summer
563 • Dwc: Cold Dry Winter, Cold Summer
564 • Dwd: Cold Dry Winter, Very Cold Winter
565 • Dfa: Cold Dry without a Dry Season, Hot Summer
566 • Dfb: Cold Dry without a Dry Season, Warm Summer
567 • Dfc: Cold Dry without a Dry Season, Cold Summer
568 • Dfd: Cold Dry without a Dry Season, Very Cold Winter
569 • ET: Polar Tundra
570 • EF: Polar Eternal Winter
571 • W: Water

572