Neural network models for soil moisture forecasting from remotely sensed measurements

Andrea Marini[1], Loris Francesco Termite[2,3], Alberto Garinei[1,4], Marcello Marconi[1,4], Lorenzo Biondi[1,4]

[1] *Idea-re S.r.l., Via Cornelia, 498, 00166, Rome, Italy*
[2] *Radarmeteo S.r.l., via IV Novembre 119, 35020, Due Carrare (PD), Italy*
[3] *Agrosit S.r.l., Via Briganti 75, 06127, Perugia, Italy*
[4] *Department of Sustainability Engineering, Guglielmo Marconi University, via Plinio 44, 00193, Rome, Italy*

ABSTRACT
Machine learning techniques are employed to describe the temporal behavior of soil moisture using meteorological data as inputs. Three different Artificial Neural Network models, a feedforward Multi-Layer Perceptron, a Long-Short Term Memory and the Adaptive Network-based Fuzzy Inference System, are trained and their results are compared. The soil moisture is expressed in terms of Soil Water Index, derived from satellite retrievals, with the last known value also being used as input. The results are promising as the proposed methodology relies on free-access data with a worldwide coverage, allowing to easily estimate the forthcoming soil moisture. The knowledge of the expected value of this variable could be extremely useful for irrigation scheduling and it is the basis of Decision Support Systems to efficiently manage water resources in agriculture.

1. INTRODUCTION

Knowledge of Soil Moisture (SM) is fundamental in several scientific fields, such as rainfall-runoff modelling, landslide forecasting, soil nutrient cycling processes, drought monitoring, and agriculture [1]. In particular, the importance of having clear control over SM in agriculture is obvious. This has become even greater in recent years due to the climate changes that are increasingly affecting our world. In mid-latitude zones, the effects are visible as both long drought periods (and consequent yield reduction) or extremely intense and localised rainfall events, which could not only ruin a harvest but also lead to crisis when flooding risk occurs. In such a context, it is imperative to properly schedule irrigation practices, which rely on SM as a key variable determining the actual need for water of the crops. Nevertheless, managing irrigation systems could be complicated, especially if irrigation is managed through the help of channels deriving from bigger water bodies, like the Po Valley in Italy. Generally, land reclamation authorities are responsible for the

management of wide networks of interconnected channels, whose flow is regulated by the movement of inline or lateral gates. Generally, a gate is open in order to move irrigation water towards a specific area, and it is closed in order to direct the flow elsewhere. Traditionally, gates are open or closed manually, but in recent years, an automation process has started, using the introduction of remotely controlled actuators for moving the gates [2][3]. However, planning the manoeuvres still remains a complex task. Decision-makers often receive water requests from different stakeholders and must deal with conflicting objectives [4][5]. Sometimes, the schedule of gate-opening and irrigation planning is inconsistent with the actual crop need or with the forthcoming weather conditions. In this context, a Decision Support System (DSS) could be helpful. DSSs are information systems supporting decision-making processes, usually including a software component, and are emerging as powerful tools in several water management contexts [6][7]. According to Power's classification [8], it is proper to consider

Figure 1. Study area.

data-driven, knowledge-driven, and model-driven DSSs (or a mix of the three) for use in irrigation management and scheduling.

The common feature of DSSs in the hydrological context is the collection of in-situ and remotely sensed variables describing the past and current status of the physical processes, and their elaboration aims to forecast their evolution so that decision-makers can act promptly. In this way, real innovative irrigation 'smart grids' can be realised. In a cultivated area, knowledge of SM may give an indication of crop status and help decision-makers decide whether to schedule irrigation or not. Thus, monitoring the values of SM and forecasting its evolution are among the primary features that should be implemented in a DSS aimed at optimally managing irrigation.

SM can be monitored in several ways. The most efficient one is by direct in-situ measurements through specific sensor devices, such as Time Domain Reflectometry (TDR) [9] or Frequency Domain Reflectometry (FDR) probes [10]. However, at the time of writing, the actual diffusion of these sensors is quite limited, and this does not allow a reasonably uniform cover of the variable in extended areas [11]. Paying the price of the lower accuracy and spatial resolution of the measurements, an alternative way of deriving water content information is by means of satellite data, thanks to recently developed Earth observation programmes [12]. The great advantage in using satellite data is that they are easily accessible and also provide worldwide coverage.

Moreover, the knowledge of the forthcoming SM evolution could improve irrigation scheduling, e.g. avoiding water supply if the SM is estimated to increase because of weather variables. The temporal dynamic of SM can be estimated through physical, conceptual, or data-driven models [13]. The latter type is used in this study to carry out an exploratory quantitative analysis on the temporal behaviour of SM in agricultural fields.

The use of machine learning has become widespread in many different contexts, ranging from medical [14] to industrial [15]

applications. In the hydrological and environmental fields, machine learning is widely used to predict the future behaviour of several relevant variables [16].

In this study, using machine learning techniques, three Artificial Neural Network (ANN) models are trained to predict SM; subsequently, they are tested, and their results are compared. Among the models' inputs, remotely sensed data are used as a measure of the water content in soil.

The main objective of this article is to develop a model that describes SM evolution in terms of meteorological data, which are clearly among the key factors affecting it. This will allow the forecasting of the future behaviour of the water content in the fields; thus, it will possibly represent the basis of a DSS for water management in irrigation.

2. MATERIALS AND METHODS

A specific case study is considered by selecting a limited geographic area in the Italian region of Veneto. In particular, the area taken into account is located between the cities of Venice and Padua, next to the Venetian Lagoon (Figure 1). This can be considered as a strategic location for the analysis, since it has traditionally been devoted to agriculture and is furnished with a wide and diffused water grid used for irrigation and land reclamation.

The proposed approach makes use of remotely sensed SM data retrieved from Copernicus, the Earth observation program developed by the European Union and European Space Agency. In particular, the SM, which is the target variable to be modelled, is described through the Soil Water Index (SWI), which provides an estimate of water content at various depths in the soil and is computed based on satellite measurements of the Surface Soil Moisture (SSM) [17]. SSM and, consequently, SWI are expressed as relative soil moisture, i.e. percentage of saturation. SWI data can be accessed freely at the Copernicus Global Land Service [18], which is part of the Copernicus programme. The Copernicus SWI data for Europe have a 1 km resolution and are based on SSM data from Sentinel-1/C-band Synthetic-Aperture Radar (SAR) and MetOp/ASCAT sensors [19], [20].

SWI allows for the control of SM at different depths through parameter T: characteristic time length. This parameter describes the temporal dynamics of the water flux below the surface: increasing values of T correspond to deeper soil layers. The great advantage in using SWI is that it requires fixing only one parameter, namely T, which is usually calibrated by means of comparison with probe measurements at the soil depth of interest [21]. Copernicus Global Land Service provides SWI values for eight different characteristic time lengths, varying from 1 to 100. The value of the characteristic time length that was taken into account in this study is $T = 15$. Of course, the same approach presented here can be equally repeated for other choices of T.

The meteorological data taken into account in the analysis are the daily measurements of rainfall; minimum and maximum temperatures; average relative humidity; and wind speed. These data are measured by both ground-based and remote meters (weather radars) and are provided by Radarmeteo. The meteorological data are available for the years 2015-2018, so the analysis is focused on this time period.

The selected methodology involves using machine learning techniques, specifically ANN models, to estimate the future evolution of SWI given the available meteorological information. The last observed SWI value, which, due to processing time, is

Table 1. Input symbols.

Symbol	Description
R_t	Rainfall on day t
T_t^{min}, T_t^{max}	Minimum and maximum temperatures on day t
T_t	Mean temperature on day t [a]
RH_t	Mean relative humidity on day t
ws_t	Mean wind speed on day t
SWI_t	Soil Water Index on day t
$\overline{X}_{t_1,t_2}$	Mean of variable X over days $t_1 \div t_2$
X_{t_1,t_2}^{Σ}	Cumulative value of variable X over days $t_1 \div t_2$

[a] The mean temperature T_t has been computed as the arithmetic mean of the minimum and the maximum daily temperatures.

assumed to be available four days after the observation, is also used as input. Thus, at time t the last available SWI value is assumed to be SWI_{t-4}. The target is the expected difference, ΔSWI, between the present-day SWI and the last observed SWI value. Table 1 summarises the symbolism used for the models' variables.

Three ANN models have been taken into account in this study: a Multi-Layer Perceptron (MLP), a Long Short-Term Memory (LSTM) network [22], and an Adaptive Network-Based Fuzzy Inference System (ANFIS) model [23].

A single hidden layer structure is selected to implement the MLP (Figure 2). The number of units in the hidden layer is 40. The MLP inputs are

$$R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4},$$
$$T_t, \overline{T}_{t-4,t-1},$$
$$RH_t, \overline{RH}_{t-4,t-1},$$
$$ws_t, \overline{ws}_{t-4,t-1},$$
$$SWI_{t-4}.$$

Note that the meteorological variables on day t are also used as inputs of the model. In real predictive applications, forecasted values have to be used for such inputs.

The LSTM network belongs to the wider class of Recurrent Neural Networks (RNNs), i.e. those with feedback connections that allow information to persist over the data sequence. This characteristic makes them naturally suited for modelling time series datasets. The second model taken into account in this
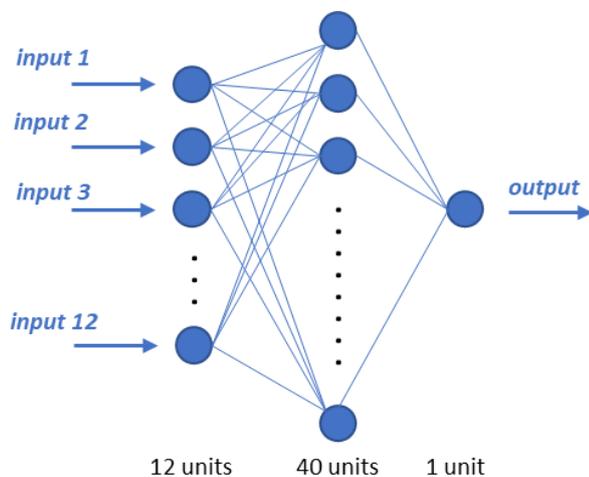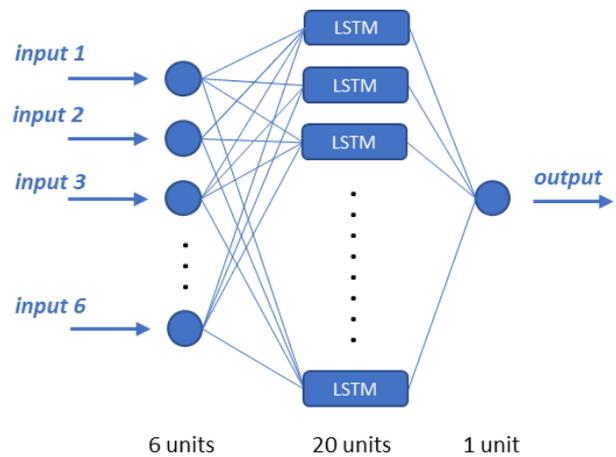

Figure 3. Scheme of the LSTM network.

study is an ANN with one hidden layer formed by 20 LSTM cells (Figure 3). The inputs used in this setup are

$$R_t, T_t^{min}, T_t^{max}, RH_t, ws_t, SWI_{t-4}.$$

In this case, it is sufficient to pass the current values of the variables as inputs, since the model is already capable – by itself – of keeping the relevant past information for the predictions.

ANFIS is a particular type of ANN used to implement Fuzzy Inference Systems (FISs), i.e. inference rules based on the fuzzy logic concept [24][25]. ANFIS has a fixed structure of five layers corresponding to the five FIS steps (input variables fuzzification, antecedent values combination, implication, aggregation of the consequents, and defuzzification of the output variable). Since the number of ANFIS parameters requiring optimization grows exponentially with the number of inputs, the available meteorological data have been aggregated differently with respect to the two previous considered models in order to keep the number of inputs small. Thus, the inputs used in the ANFIS model are

$$R_{t-4,t}^{\Sigma}, \overline{T}_{t-4,t-1}, \overline{RH}_{t-4,t-1}, \overline{ws}_{t-4,t-1}, SWI_{t-4}.$$

A schematic of the ANFIS network is depicted in Figure 4. For each ANFIS input, three membership functions defined in Equation (1) are used:
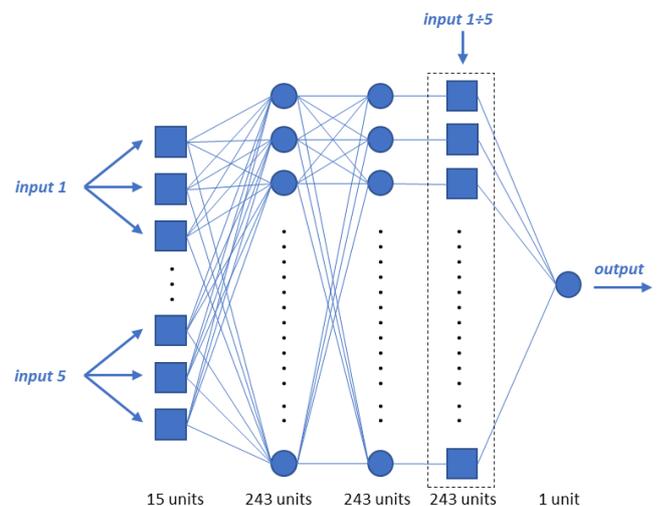

Figure 2. Scheme of the MLP network.

Figure 4. Scheme of the ANFIS network.

$$f_{\text{memb}}(x; \mu, \sigma, \nu, \alpha) = \exp\left[-\frac{|x - \mu|^\nu}{\sigma^2} \frac{1}{1 + e^{-\alpha x}} \right] \qquad (1)$$

Each membership function has four free parameters, μ, σ, ν, and α, and can be seen as a deformation of the gaussian. Indeed, setting $\nu = 2$ and $\alpha = 0$ yields the usual gaussian function, with mean μ and variance σ^2. The parameter ν allows for different peak shapes, while the sigmoid factor $(1 + e^{-\alpha x})^{-1}$ allows for skewness.

It is customary in machine learning applications to split the available dataset into three subsets: the training, the validation, and the test sets. The training set, which usually comprises the most data, is used to feed the learning algorithm. The validation set is used for the selection of the model's hyperparameters or for comparing models. The test set is only used at the end to evaluate the performance of the selected model (observing how well the model works when applied to unseen data). In this analysis, in order to exploit as much data as possible for training, the data has been split only into training and validation sets, disregarding the test set. In practice, the validation set has been used both for the model selection and for the final assessment of the selected model. Though this is not optimal from a statistical point of view, this strategy is quite commonly used when the amount of available data is quite limited.

For all the three models, data from January 2015 to December 2017 are used for training, and data from January to December 2018 are used for validation.

As is usual in such modelling, the dataset has been normalised in order to facilitate the training procedure. The normalisation has been performed in such a way that all the variables have zero mean and unitary standard deviation, i.e. by computing the standard scores.

The mean squared error of the output variable has been used as loss function Regularisation strategies have been exploited in order to keep possible overfitting issues under control. For the MLP and LSTM networks the dropout technique has been used [26]. Instead, in the ANFIS model, an L2 regularisation penalty for the weights of the consequent part of the *if-then* rules has been added to the loss function.

The training has been performed by minimising this loss function by means of ADAM optimisation [27]. The training procedure has been carried out 100 times for all the considered models, and the values of the parameters have been set to the ones giving the least value of the mean squared error for the validation set. The models have been evaluated by means of Root Mean Square Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) index computed for the validation data.

3. RESULTS

At the end of the training phase, the best MLP, LSTM, and ANFIS models have been selected according to the abovementioned criteria, and a deeper analysis of the results has been carried out.

Figure 5 shows the scatter plots of the predicted vs. observed ΔSWI for the three selected models. The blue dots refer to the training set, and the orange ones to the validation set. From these plots, a good agreement of the predictions can be observed when the actual ΔSWI is in the intermediate region (approximately between −7 and 7). Conversely, the predictions seem to be less accurate in the correspondence of the remaining external regions of the observed ΔSWI, where the points deviate more markedly
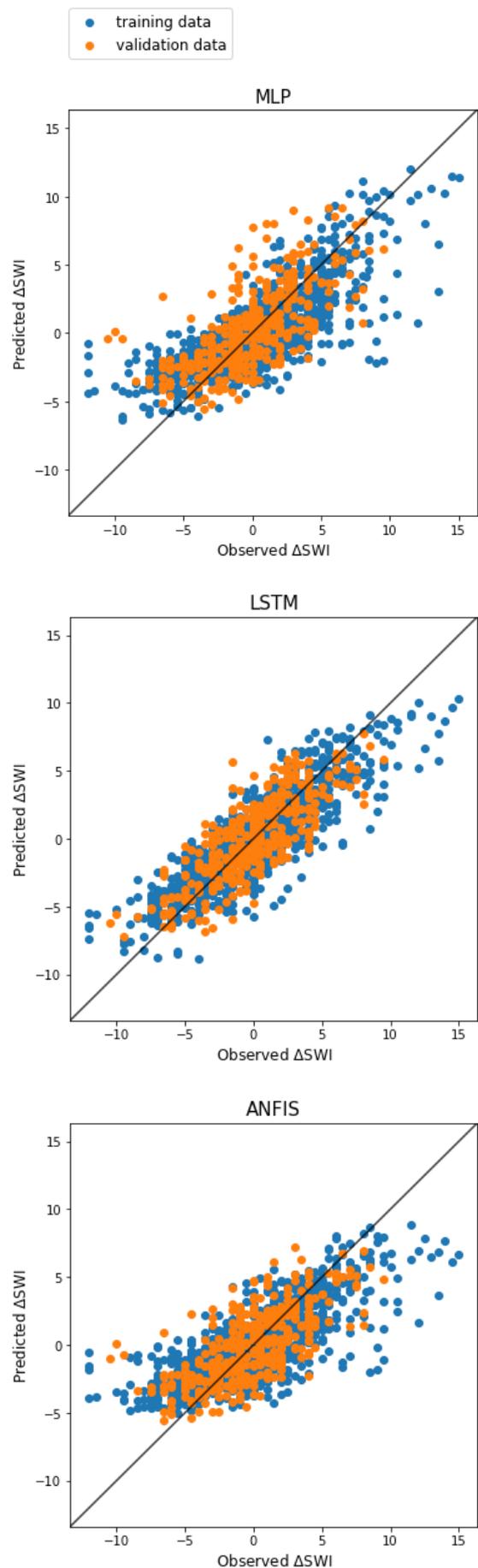

Figure 5. Scatter plots of the predicted vs. observed ΔSWI for the three models.
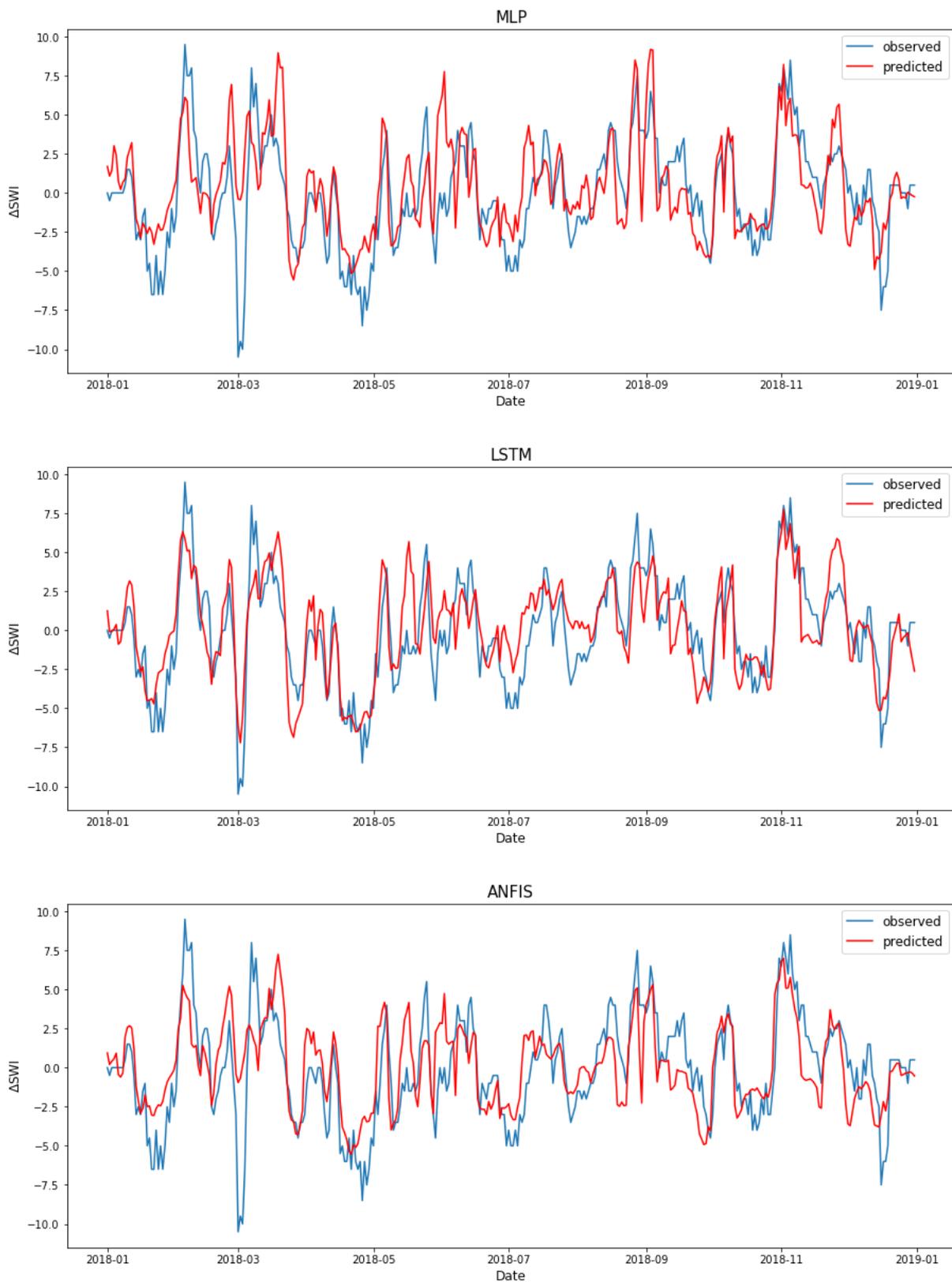
Figure 6. Predicted and observed ΔSWI for the validation set.

from the perfect forecast line: in these regions, the models tend to underestimate the amplitude of the real ΔSWI. It can be clearly observed that this effect is sharper for the MLP and the ANFIS models than for the LSTM. A possible explanation of this issue may be related to the small number of occurrences of such events in the available dataset. Note, however, that the difficulty of

correctly reproducing extreme observations is also a known problem in machine learning modelling [28].

The plots in Figure 6 show the temporal behaviour of the ΔSWI predicted by the MLP, LSTM, and ANFIS models on the validation set, along with the corresponding observed values.

Table 2. Model results.

	Training		Validation	
	NSE	*RMSE*	*NSE*	*RMSE*
MLP	0.590	2.579	0.397	2.599
LSTM	0.746	2.027	0.606	2.113
ANFIS	0.558	2.677	0.468	2.440

The evaluation metrics, NSE and RMSE, are shown in Table 2 for both the training and validation data. In particular, the metrics for the validation set are the most interesting ones for an assessment of the models.

The results in Table 2 shows that among the three models taken into account, the one that has the best performance, both in terms of the least RMSE and higher NSE, is the LSTM one. The ANFIS model slightly outperforms the MLP.

It is interesting to note that the RMSE on the validation set and the RMSE on the training set are quite similar. This is quite surprising, since one usually expects a worsening of the performance metrics on the validation set compared to that of the training set. Indeed, for the ANFIS model, the validation RMSE is even better than the training RMSE. This 'strange' behaviour is probably due to the presence of a higher number of outliers in the training set with respect to the validation set.

4. CONCLUSIONS

This article presented an approach for modelling SM using machine learning techniques. SM is expressed in terms of SWI retrieved from satellite measurements. These products are released on the Copernicus website with a time delay of two days after observation. To be cautious, it has been assumed that the last available SWI value dates back to four days earlier. Thus, the variation of SWI in a four-day period has been modelled through three different network-based models: a single hidden layer feedforward MLP, an LSTM network, and an ANFIS. Meteorological data, comprising rainfall, temperature, relative humidity, and wind speed records, have been used as inputs in both models along with the last known SWI value.

In the qualitative analysis, the temporal behaviours of the predicted and the observed ΔSWI appear to correlate well. Yet, at the quantitative level, the agreement is not always satisfying,

especially when the observed value lies in the tail of the ΔSWI distribution. However, when modelling a complex phenomenon using only a limited fraction of the actual set of variables influencing it, it is essential for the measured data to be extremely accurate and numerous. On the contrary, in this research, the available dataset spans over just four years (only three were used for training), and it could be that it has been insufficiently extended so as to guarantee an adequate generalisation capability to the models. Moreover, while they have been extremely useful thanks to their worldwide coverage and daily frequency, the SWI data obviously do not have the same accuracy as in-situ measurements. Finally, due to lack of historical records, all the variables influencing the system status related to human interventions, e.g. irrigation supplies, have been neglected.

The comparison of the results of the models that have been taken into account indicates that the LSTM model outperforms the other two selected ANNs.

The output of the models, i.e. ΔSWI, may be added to the last known SWI value, SWI_{t-4}, in order to obtain the predicted SWI, SWI_{pred}, which is what the decision-makers are eventually interested in. The time series of the observed SWI and the LSTM SWI_{pred} are shown in Figure 7.

The results of the research are encouraging, as they show that in the absence of networks of in-situ sensors, satellite-derived measurements of SM can be forecasted by machine learning models, using simple meteorological data as inputs. This could be extremely useful in managing water resources, since SM is a fundamental variable to be considered when supplying irrigation.

In principle, the proposed approach can be applied in all those settings where measurements from different sources over large areas (whether ground-based or remote sensing), meteorological data, and any quantitative information need to be processed to provide synthetic outputs for the final user. The exploitation of artificial intelligence techniques allows for modelling complex and highly nonlinear processes. Furthermore, this factor makes it possible to avoid the use of those parameters that would be necessary in physical modelling and would require extensive field campaigns to characterise the study area or should be estimated during model calibration. For example, no information about the hydraulic and pedologic characteristics of the soil is provided to the ANN models presented in this study.

One limitation of the proposed approach is that it allows for the modelling of SM on a scale that is larger than the typical plot
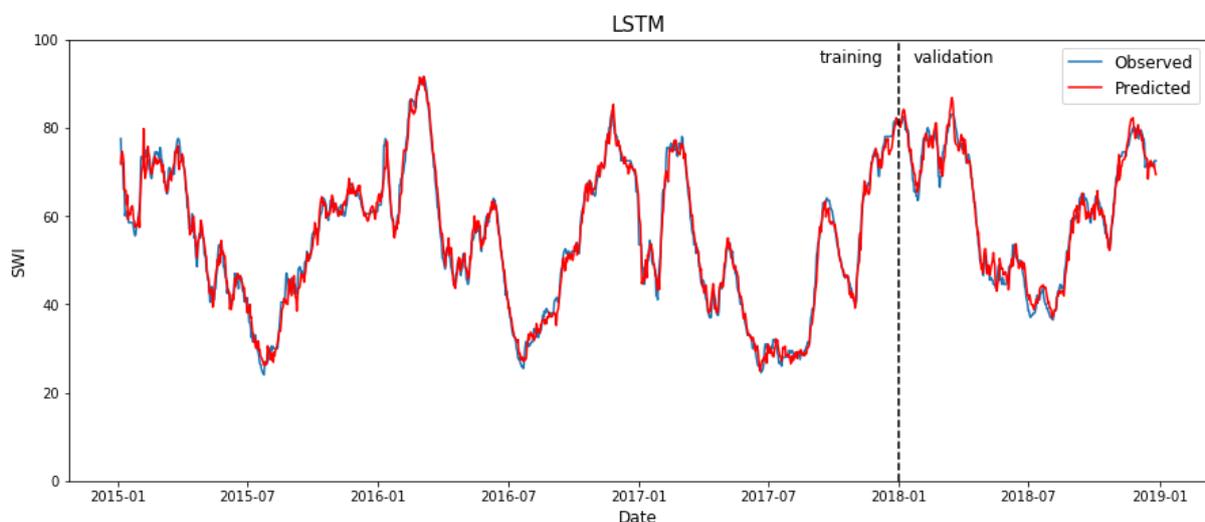

Figure 7. SWI time series reconstructed from the LSTM model.

dimension. The resolution of this approach is indeed set by that of the satellite data, namely 1 km. However, this resolution is sufficiently high so as to provide a decision support tool for the authorities that are responsible for irrigation scheduling over large areas.

Moreover, even if the spatial resolution of this kind of model is coarse compared to the typical field dimensions, valuable information can be applied also on a smaller scale. Indeed, it has been proven that SM follows a similar behaviour at global and local scales, and this allows for the downscaling of satellite measurements of this variable [29].

ACKNOWLEDGEMENT

REFERENCES

[1]   J. Dari, R. Morbidelli, C. Saltalippi, C. Massari, L. Brocca, Spatial-temporal variability of soil moisture: Addressing the monitoring at the catchment scale, Journal of Hydrology 570 (2019) pp. 436-444.

[2]   S. Choy, E. Weyer, Reconfiguration schemes to mitigate faults in automated irrigation channels, Control Engineering Practice 16(10) (2008) pp. 1184-1194.

[3]   H. A. Nasir, M. Cantoni, E. Weyer, An efficient implementation of stochastic mpc for open channel water-level planning, Proc. of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 2017, pp. 511-516.

[4]   J. R. Ribas, An assessment of conflicting intentions in the use of multipurpose water reservoirs, Water Resources Management 28(12) (2014) pp. 3989-4000.

[5]   S. H. Zyoud, L. G. Kaufmann, H. Shaheen, S. Samhan, D. Fuchs-Hanusch, A framework for water loss management in developing countries under fuzzy environment: Integration of fuzzy ahp with fuzzy topsis, Expert Systems with Applications 61 (2016) pp. 86-105.

[6]   E. Giusti, S. Marsili-Libelli, A fuzzy decision support system for irrigation and water conservation in agriculture, Environmental Modelling & Software 63 (2015) pp. 73-86.

[7]   H. Zamani Sabzi, S. Abudu, R. Alizadeh, L. Soltanisehat, N. Dilekli, J. P. King, Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 40(11) (2018) pp. 1398-1416.

[8]   D. J. Power, Decision Support Systems: Concepts and Resources for Managers. Greenwood Publishing Group, 2002, ISBN 13: 978-1567204971.

[9]   G. C. Topp, J. Davis, A. P. Annan, Electromagnetic determination of soil water content: Measurements in coaxial transmission lines, Water Resources Research 16(3) (1980) pp. 574-582.

[10] S. U. Susha Lekshmi, D. Singh, M. S. Baghini, A critical review of soil moisture measurement, Measurement 54 (2014) pp. 92-105.

[11] L. Brocca, L. Ciabatta, C. Massari, S. Camici, A. Tarpanelli, Soil moisture for hydrological applications: open questions and new opportunities, Water 9(2) (2017) p. 140.

[12] A. S. Belward, J. O. Skøien, Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites, ISPRS Journal of Photogrammetry and Remote Sensing 103 (2015) pp. 115-128.

[13] L. Brocca, S. Camici, F. Melone, T. Moramarco, J. Martínez-Fernández, J. -F. Didon-Lescot, R. Morbidelli, Improving the representation of soil moisture by using a semi-analytical infiltration model, Hydrological Processes 28(4) (2014), pp. 2103-2115.

[14] J. K. Kueper, A. L. Terry, M. Zwarenstein, D. J. Lizotte, Artificial intelligence and primary care research: a scoping review, The Annals of Family Medicine 18(3) (2020) pp. 250-258.

[15] S. Ilić, A. Selakov, S. Vukmirović, A. Erdeljan, F. Kulić, Short-term load forecasting in large scale electrical utility using artificial neural network, Journal of Scientific and Industrial Research 72(12) (2013) pp. 739-745.

[16] S. Ardabili, A. Mosavi, M. Dehghani, A. R. Várkonyi-Kóczy, Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review, Proc. of the International Conference on Global Research and Education, 2019, pp. 52-62.

[17] W. Wagner, G. Lemoine, H. Rott, A method for estimating soil moisture from ers scatterometer and soil data, Remote sensing of Environment 70(2) (1999) pp. 191-207.

[18] Soil Water Index, Copernicus Global Land Service. [Online]. Available:
https://land.copernicus.eu/global/products/swi

[19] B. Bauer-Marschallinger, C. Paulik, S. Hochstöger, T. Mistelbauer, S. Modanesi, L. Ciabatta, C. Massari, L. Brocca, W. Wagner, Soil moisture from fusion of scatterometer and SAR: Closing the scale gap with temporal filtering, Remote Sensing 10(7) (2018) p. 1030.

[20] Copernicus Global Land Operations, Algorithm theoretical basis document – Soil Water Index collection 1km, Version 1 (2019). [Online] Available:
https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_SWI1km-V1_I1.20.pdf

[21] C. Paulik, W. Dorigo, W. Wagner, R. Kidd, Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network, International Journal of Applied Earth Observation and Geoinformation 30 (2014) pp. 1-8.

[22] S. Hochreiter, S. Schmidhuber, J. Schmidhuber, Long short-term memory, Neural Computation 9(8) (1997) pp. 1735-1780.

[23] J.-S. Jang, Anfis: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics 23(3) (1993) pp. 665-685.

[24] L. A. Zadeh, Fuzzy sets, Information and Control 8(3) (1965) pp. 338-353.

[25] L. A. Zadeh, Fuzzy logic and approximate reasoning, Synthese 30(3-4) (1975), pp. 407-428.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15(1) (2014) pp. 1929-1958.

[27] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, Proceedings of the 3rd International Conference on Learning Representations (2015), arXiv:1412.6980.

[28] C. Imrie, S. Durucan, A. Korre, River flow prediction using artificial neural networks: generalisation beyond the calibration range, Journal of Hydrology 233(1-4) (2000) pp. 138–153.

[29] L. Brocca, T. Tullo, F. Melone, T. Moramarco, R. Morbidelli, Catchment scale soil moisture spatial–temporal variability, Journal of Hydrology 422 (2012) pp. 63-75.