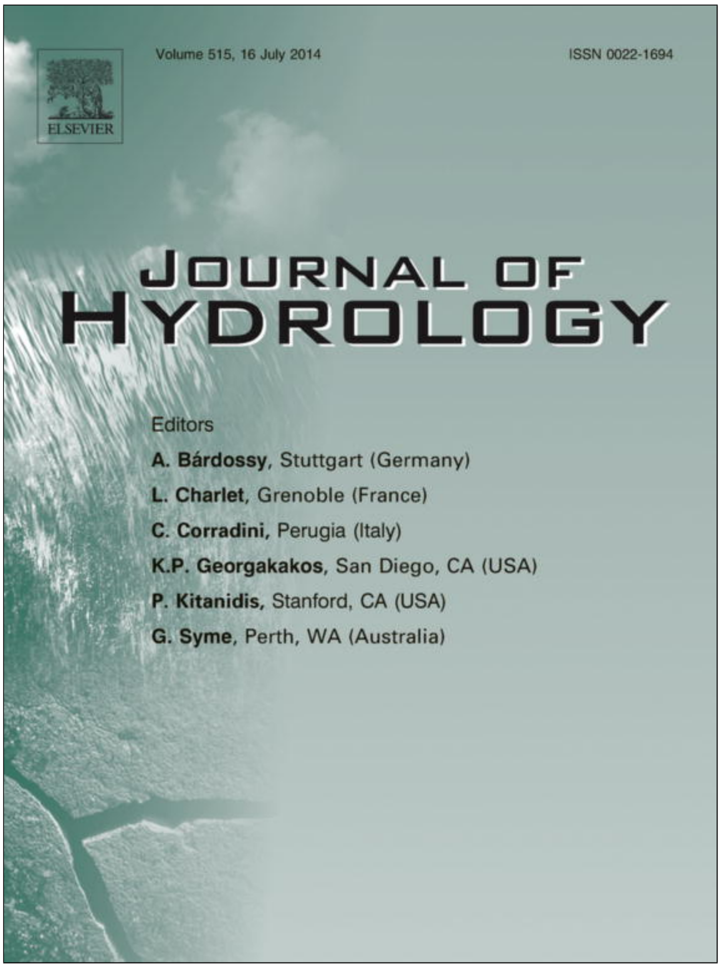
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Journal of Hydrology 515 (2014) 330–344

Contents lists available at ScienceDirect

Journal of Hydrology

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

Assessing artificial neural networks and statistical methods for infilling missing soil moisture records

Gift Dumedah *, Jeffrey P. Walker, Li Chik

Department of Civil Engineering, Monash University, Building 60, Melbourne, 3800 Victoria, Australia

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form 4 April 2014
Accepted 28 April 2014
Available online 9 May 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Attilio Castellarin, Associate Editor

Keywords:
Soil moisture
Temporal stability
Missing values
Artificial neural networks
Rough sets

SUMMARY

Soil moisture information is critically important for water management operations including flood forecasting, drought monitoring, and groundwater recharge estimation. While an accurate and continuous record of soil moisture is required for these applications, the available soil moisture data, in practice, is typically fraught with missing values. There are a wide range of methods available to infilling hydrologic variables, but a thorough inter-comparison between statistical methods and artificial neural networks has not been made. This study examines 5 statistical methods including monthly averages, weighted Pearson correlation coefficient, a method based on temporal stability of soil moisture, and a weighted merging of the three methods, together with a method based on the concept of rough sets. Additionally, 9 artificial neural networks are examined, broadly categorized into feedforward, dynamic, and radial basis networks. These 14 infilling methods were used to estimate missing soil moisture records and subsequently validated against known values for 13 soil moisture monitoring stations for three different soil layer depths in the Yanco region in southeast Australia. The evaluation results show that the top three highest performing methods are the nonlinear autoregressive neural network, rough sets method, and monthly replacement. A high estimation accuracy (root mean square error (RMSE) of about 0.03 m^3/m^3) was found in the nonlinear autoregressive network, due to its regression based dynamic network which allows feedback connections through discrete-time estimation. An equally high accuracy (0.05 m^3/m^3 RMSE) in the rough sets procedure illustrates the important role of temporal persistence of soil moisture, with the capability to account for different soil moisture conditions.

Crown Copyright © 2014 Published by Elsevier B.V. All rights reserved.

1. Introduction

Moisture in the upper layers of the soil is a vital component of the total water balance in the Earth-atmosphere system, playing a crucial role in several hydrological processes. Soil moisture is one of the main factors influencing the partitioning of rainfall into infiltration and runoff (Mahmood, 1996; Thornthwaite, 1961), controlling the exchange of water and energy between the land surface and the atmosphere (Legates et al., 2010; Berg and Mulroy, 2006; Trenberth and Guillemot, 1998; Houser et al., 1998; Reynolds et al., 2002), and the subsurface water drainage that influences the leaching of contaminants to groundwater (Langevin and Panday, 2012; Legates et al., 2010). The reliability of the above mentioned applications usually depends on the availability of a continuous time series of soil moisture record. Typically, soil moisture data acquired through ground (or in situ) measurements have missing values due to equipment malfunction, logger storage overruns, data retrieval problems, and/or severe weather conditions (Dumedah and Coulibaly, 2011; Coulibaly and Evora, 2007). Consequently, the infilling of missing soil moisture values becomes a necessary procedure to generate a continuous time series record.

Several studies have infilled hydrologic variables including precipitation (Mwale et al., 2012; Nkuna and Odiyo, 2011; Coulibaly and Evora, 2007; French et al., 1992; Luck et al., 2000; Abebe et al., 2000; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b), streamflow (Mwale et al., 2012; Ng and Panu, 2010; Ng et al., 2009; Elshorbagy et al., 2000; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b), evapotranspiration (Abudu et al., 2010), air temperature (Coulibaly and Evora, 2007; Schneider, 2001), and soil moisture (Gao et al., 2013; Wang et al., 2012; Dumedah and Coulibaly, 2011). The infilling methods employed in the above studies ranged from statistical methods (Gao et al., 2013; Wang et al., 2012; Dumedah and Coulibaly, 2011) to artificial neural networks (Mwale et al., 2012; Nkuna and Odiyo, 2011; Coulibaly and Evora, 2007), with

* Corresponding author. Tel.: +61 414 273 492.
E-mail address: dgiftman@hotmail.com (G. Dumedah).

http://dx.doi.org/10.1016/j.jhydrol.2014.04.068
0022-1694/Crown Copyright © 2014 Published by Elsevier B.V. All rights reserved.

Author's personal copy

varying levels of accuracy. While several studies have explored different infilling approaches, very few studies have been undertaken to actually reconstruct soil moisture records using both statistical and artificial neural network methods. As a result, this study investigates 5 statistical and 9 artificial neural network methods, a total of 14 methods to estimate missing soil moisture records. The soil moisture monitoring network located in the Yanco region of southeast Australia (Smith et al., 2012) is used as the demonstration data set.

The statistical methods include monthly replacement, weighted Pearson correlation, station relative difference, and a weighted merger of the three statistical methods. Moreover, a method based on the concept of rough sets (Pawlak, 1997; Pawlak et al., 1995; Pawlak, 1982) was used to determine patterns of temporal stability of soil moisture to account for different moisture conditions. The artificial neural networks (ANNs) evaluated in this stud are broadly categorized into feedforward group, dynamic group and radial basis group. Detailed descriptions for the statistical and ANN methods are provided in the methods section. The selected approaches constitute a varied range of methodologies to facilitate a comprehensive inter-comparison between a range of statistical and ANNs with the potential to identify high performing methods to infill missing soil moisture. The infilling methods have been evaluated for their estimation accuracy across 13 soil moisture monitoring stations independently at three different soil layer depths in the Yanco area. Moreover, an evaluation of the soil moisture across the 13 monitoring stations in space and their persistence of relative moisture conditions over several time periods was demonstrated. These space–time distributions are presented for the entire period of the chosen soil moisture data, and also on a month-by-month basis.

2. Study area and soil moisture data

The Yanco area shown in Fig. 1 is a 60 km × 60 km area, located in the western plains of the Murrumbidgee Catchment in southeast Australia where the topography is flat with very few geological outcroppings. Soil texture types are predominantly sandy loams, scattered clays, red brown earths, transitional red brown earth, sands over clay, and deep sands. According to the Digital Atlas of Australian Soils, the dominant soil is characterized by plains with domes, lunettes, and swampy depressions, and divided by continuous or discontinuous low river ridges associated with prior stream systems (McKenzie et al., 2000). The area is traversed by present stream valleys, layered soil or sedimentary materials common at fairly shallow depths; chief soils are hard alkaline red soils, gray and brown cracking clays.

The Yanco area has 13 soil moisture profile stations which form part of the OzNet hydrological monitoring network (www.oznet.org.au) in the Murrumbidgee Catchment. Generally, profile soil moisture monitoring at all the stations in the Yanco area have been in operation since 2004 using Campbell Scientific water content reflectometers (CS615, CS616) and the Stevens Hydraprobe for four soil layers: 0–5 cm (or 0–7 cm), 0–30 cm, 30–60 cm and 60–90 cm (Smith et al., 2012). Sensor response to soil moisture varies with salinity, bulk density, soil type and temperature, so a site-specific sensor calibration has been undertaken using both laboratory and field measurements for both the reflectometers and the Hydraprobes (Western et al., 2000; Western and Seyfried, 2005; Yeoh et al., 2008). As the CS615 and CS616 sensors are particularly sensitive to soil temperature fluctuations (Rüdiger et al., 2010), temperature sensors were installed to provide a continuous record of soil temperature at the midpoint along the reflectometers.

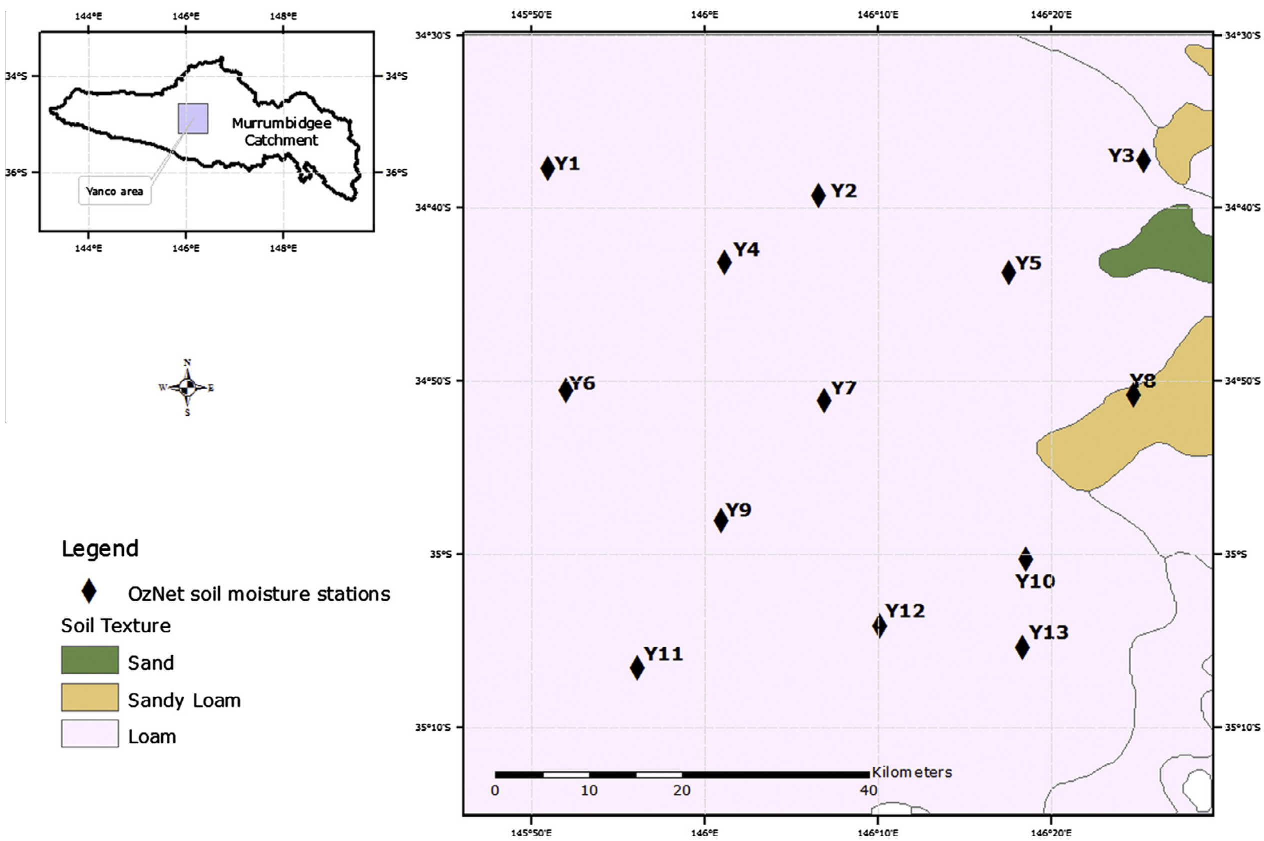


Fig. 1. Yanco study area in south-east Australia showing the location of soil moisture stations and the soil texture distribution.

Author's personal copy

Calibration relationships between the sensor observations, coinciding with traditional Time-Domain Reflectometer (TDR) measurements and thermo-gravimetric measurements have been established for both sensors. The average root mean square error was found to be 0.03 m^3/m^3 for both the Campbell Scientific (Yeoh et al., 2008) and Hydraprobe (Merlin et al., 2007) sensors. The soil moisture observations were made at half-hourly time intervals, but it is noted that the infilling procedure was applied to an hourly version of the data.

3. Methods for infilling missing soil moisture records

Using the raw soil moisture data, a complete data set was retrieved by removing all periods with missing records, such that all soil moisture records were temporally consistent (or common) to all stations for a specific soil layer (Dumedah and Coulibaly, 2011; Coulibaly and Evora, 2007). In other words, the complete data set is spatially complete in a way that each record in the data set at any one station has corresponding records available across the remaining 12 stations for the specified soil layer. As a result, three complete data sets were derived, with each data set corresponding to one of the first three soil layer depths. The rationale to generate the complete data set in this manner is partly because soil moisture records are usually missing at the same station for all three soil depths at any given time period. This approach to generate the complete data record allows the infilling of soil moisture for any soil depth using information from other monitoring stations, when records from other soil depths are often missing. That is, when soil moisture is missing at the one (e.g., surface) soil layer for any given monitoring station, there is high probability that the soil moisture is also missing at the other (e.g., second and third) soil layers for the same station. In practice, it is easier to find data at other monitoring stations to infill soil moisture for the station with the missing record for the specified time period.

It is noted that soil moisture data from only the first three soil layers are used in this study due to limited records available across stations for the 60–90 cm soil layer. The complete data set was divided into two: a training data set, and a validation data set. Using the complete data set, 20% of its records were randomly removed (temporally and independent of monitoring station) to makeup the validation data set, with the remaining 80% constituting the training data set for model development. As a result, all the infilling methods were developed or trained using the training data set, and subsequently evaluated against known values in the validation data set. The validation data of 20% represents a considerable proportion of missing soil moisture compared to the proportion of missing values found in past studies, including Gao et al. (2013); Wang et al. (2012); Dumedah and Coulibaly (2011); and Coulibaly and Evora (2007). The descriptive statistics along with the number of records for the training and validation data sets are summarized in Table 1. A time series plot of the training and validation data sets for the first and second soil layers at station Y1 is shown in Fig. 2. This plot also illustrates the huge disparities in the number of records and the time periods when soil moisture is available between different soil layer depths. It is noted that the number of records in the training and validation data sets for specific soil layers is the same across all the monitoring stations, in accordance with the definition of the complete data. Although the number of records in the validation data set is the same for a specific soil layer, their time periods are unique as they were randomly generated for each station independently. The infilling methods applied for estimating the missing soil moisture are described in the following sections.

3.1. Station layer relative difference method

The station layer relative difference (SLRD) method is based on the concept of temporal stability of soil moisture, and uses parametric test of relative difference, which was proposed by

Table 1
Number of records, mean and standard deviation of soil moisture (m^3/m^3) for training and validation data sets for the three soil layers across all 13 soil moisture monitoring stations.

Station	Statistic	0–5 cm		0–30 cm		30–60 cm	
		Training	Validation	Training	Validation	Training	Validation
1–13	Records	5500	1375	15577	3895	15577	3895
1	Mean	0.083	0.079	0.146	0.147	0.264	0.263
	STD	0.045	0.043	0.036	0.037	0.049	0.047
2	Mean	0.184	0.183	0.215	0.215	0.248	0.246
	STD	0.068	0.067	0.075	0.077	0.106	0.104
3	Mean	0.125	0.127	0.125	0.125	0.153	0.152
	STD	0.033	0.033	0.041	0.041	0.042	0.040
4	Mean	0.171	0.173	0.272	0.271	0.223	0.223
	STD	0.077	0.078	0.091	0.091	0.071	0.071
5	Mean	0.147	0.144	0.167	0.166	0.285	0.285
	STD	0.066	0.064	0.040	0.039	0.045	0.045
6	Mean	0.158	0.154	0.177	0.181	0.269	0.267
	STD	0.104	0.105	0.105	0.107	0.080	0.077
7	Mean	0.112	0.112	0.187	0.187	0.361	0.361
	STD	0.052	0.052	0.064	0.063	0.037	0.038
8	Mean	0.101	0.103	0.106	0.106	0.235	0.236
	STD	0.073	0.073	0.038	0.038	0.024	0.025
9	Mean	0.206	0.202	0.174	0.175	0.377	0.377
	STD	0.090	0.090	0.046	0.044	0.057	0.058
10	Mean	0.155	0.152	0.221	0.220	0.326	0.327
	STD	0.076	0.076	0.102	0.101	0.079	0.080
11	Mean	0.113	0.115	0.267	0.269	0.403	0.402
	STD	0.073	0.074	0.118	0.119	0.111	0.111
12	Mean	0.143	0.145	0.236	0.237	0.337	0.343
	STD	0.066	0.066	0.089	0.089	0.091	0.089
13	Mean	0.166	0.169	0.206	0.204	0.247	0.247
	STD	0.075	0.077	0.107	0.106	0.077	0.075

Author's personal copy

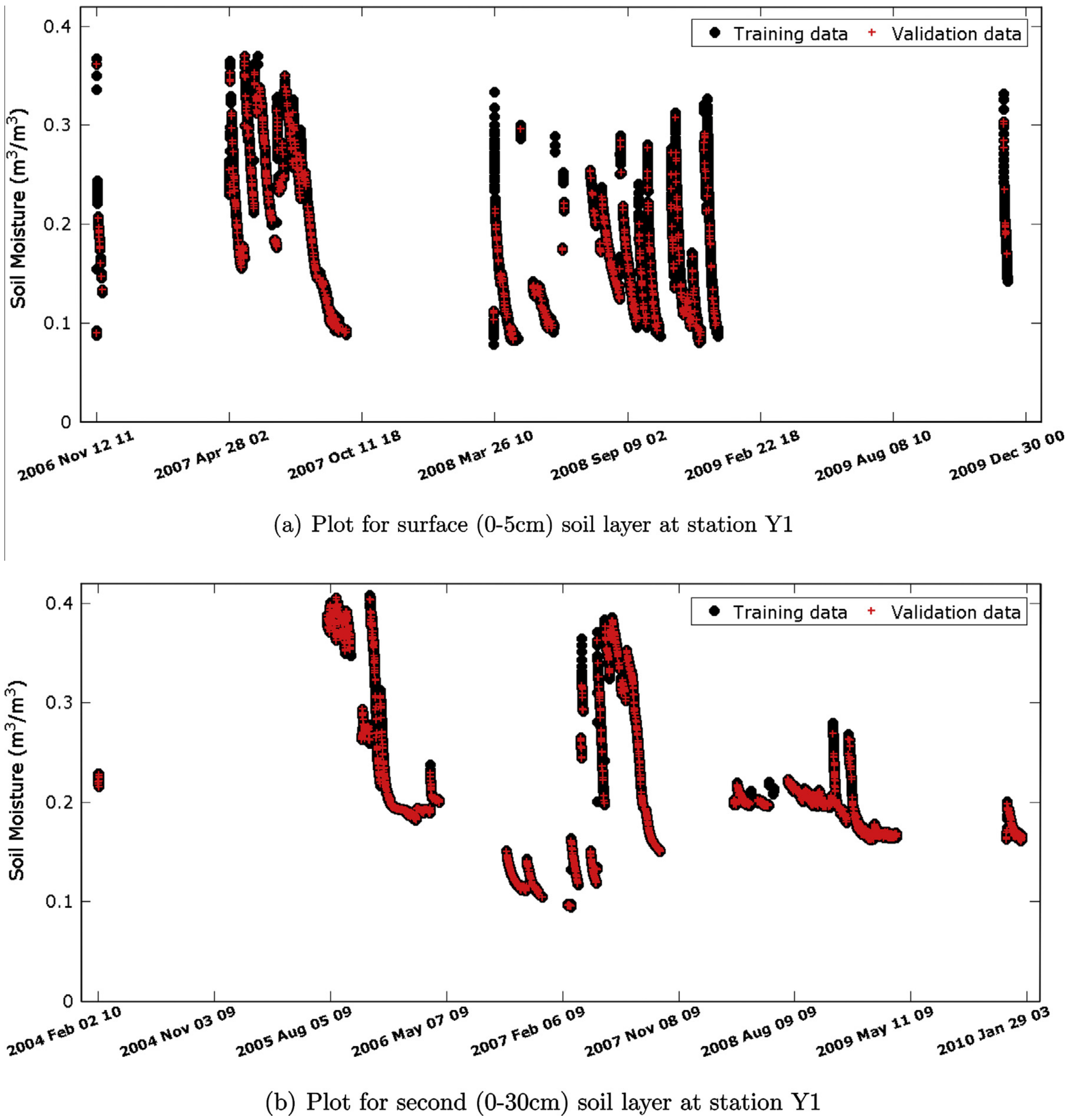


Fig. 2. Time series plots (with horizontal axis in 'year month day hour' date format) of training and validation data sets for station Y1 for surface and second soil layers.

Vachaud et al. (1985). The relative difference, δ_{ij} is estimated according to Eq. (1); where θ_{ij} is soil moisture content at location i on time j; n is the number of sampling locations, and $\bar{\theta}_j$ is the spatial average of soil moisture content at time j which is defined in Eq. (2). The estimation is summarized by finding the mean, $\bar{\delta}_i$ and standard deviation, $\sigma(\delta_i)$.

$$\delta_{ij} = \frac{\theta_{ij} - \bar{\theta}_j}{\bar{\theta}_j} \quad (1)$$

$$\bar{\theta}_j = \frac{1}{n} * \sum_{j=1}^{n} \theta_{ij} \quad (2)$$

The SLRD was originally used to infill missing soil moisture records in Dumedah and Coulibaly (2011), and it is applied in this study to estimate the mean and standard deviation of relative difference for each soil moisture station using each of the three soil layers. That is, for each soil layer the overall relative difference is estimated for each station using data from a corresponding soil layer across all stations. The SLRD procedure estimates the relative wetness between the stations, indicating temporal persistence of spatial pattern for soil moisture across stations. The relative difference was estimated month-by-month and stored in a lookup (or reference) table for use in estimating missing soil moisture according to Eqs. (3) and (4).

$$\theta_{est} = \theta_i + \theta_i * \bar{\delta}_i \quad (3)$$

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_i^2} + \frac{1}{\sigma^2_{(\delta_i)}} \quad (4)$$

where θ_{est} is the estimated soil moisture, σ^2 is the variance of the estimated soil moisture, θ_i is the average soil moisture from other stations for the current time with its associated variance σ_i^2, and $\bar{\delta}_i$ is the lookup (or reference) relative difference with its variance $\sigma^2_{(\delta_i)}$.

3.2. Monthly average replacement method

Using the training data, the monthly (January through December) average and standard deviation are estimated for each monitoring station and soil moisture layer. The monthly average

Author's personal copy

replacement (MAR) uses the estimated monthly average and standard deviation from the training data to replace missing values for the corresponding month in the validation data.

3.3. Weighted Pearson correlation coefficient method

The weighted Pearson correlation coefficient (WPCC) uses Pearson correlation coefficient (PCC) as an estimate of the temporal association between monitoring stations. The PCC values are used as a weighting factor to estimate missing soil moisture values such that high PCC values are weighted high and low PCCs are weighted low. The PCC weight for the ith station is determined according to Eq. (5), and the missing soil moisture value is estimated using Eq. (6).

$$w_i = \frac{PCC_i}{\sum_{j=1}^{n} PCC_j} \tag{5}$$

$$\theta_{est} = \sum_{i=1}^{n} \theta_i * w_i \tag{6}$$

where θ_{est} is the estimated soil moisture, and θ_i is the soil moisture value at the ith neighboring station.

3.4. Merged method

The merged method is an approach adopted in Dumedah and Coulibaly (2011) to assemble the estimated soil moisture from other methods based on their variance errors, such that estimates with high variance errors are weighted less or highly penalized. In this study, the merged method is used to combine three soil moisture estimates from SLRD, MAR, and WPCC. The weight for the ith method is determined according to Eq. (7), and the merged soil moisture is estimated in Eq. (8) with its variance error in Eq. (9).

$$w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{n} \left(1/\sigma_j^2\right)} \tag{7}$$

$$\theta_{est} = \sum_{i=1}^{n} \theta_i * w_i \tag{8}$$

$$\frac{1}{\sigma^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \tag{9}$$

where n is the number of soil moisture monitoring stations; σ_i is the variance error for ith method, θ_{est} is the estimated soil moisture, and θ_i is the soil moisture value for ith method.

3.5. Method based on rough sets

In the SLRD approach, an estimate of the overall relative difference for each station in relation to the average soil moisture representing the entire spatial area was used. While the overall relative difference represents the rank persistence of soil moisture for each station over long time periods, the unique properties of certain soil moisture conditions (e.g., extreme dry or wet seasons) may be over generalized or smoothed. To account for unique properties of soil moisture conditions, the rough sets theory is used to determine unique soil moisture groupings across the monitoring stations, where patterns of relative difference are determined for each moisture category. A rough set is a classical set with a non-empty boundary when approximated by another set (Pawlak et al., 1995; Pawlak, 1982). The concept of rough sets is proficient for the discovery of hidden patterns and the characterization of relationships (Dumedah and Schuurman, 2008; Dumedah et al., 2008; Ohrn, 1999; Pawlak, 1997), which is suitable to categorize patterns of temporal persistence of soil moisture for different moisture conditions.

The rationale to employ rough sets is to examine whether a patterned relative difference for different moisture conditions will provide a better estimate of the missing soil moisture record than the overall relative difference used in the SLRD. The rough sets approach is applied using the following procedure.

(i) Using the training data, the relative difference is computed temporally at each station for each of the three soil layers using Eq. (1).
(ii) A number of rough sets categories x (e.g., 4) is chosen to group the relative difference values. This number x is used to generate x categories of relative difference groups for each monitoring station. The grouping intervals are not necessarily the same for different monitoring stations but they are dependent on the distribution of the relative difference values at each station.
(iii) For each x category at each station, the relative difference values at the other stations which correspond temporally to the current station are extracted. The pattern between the current station and each of the other remaining stations is determined by finding the dominant or the most frequently occurring group. The derived pattern is a frequency-based probability with the group having the highest probability representing the lower approximation of the rough sets. This procedure is repeated to determine the pattern for each of the x categories at each station; the patterns are stored in a lookup (i.e., a reference) table for use in estimating missing soil moisture records.
(iv) It is noteworthy that for each pattern, evaluation measures including support, strength in Eq. (10), certainty or accuracy in Eq. (11), and coverage in Eq. (12) are determined to assess the reliability of the derived pattern. These pattern evaluations are also used as a basis to choose the number of categories (x) referred to in step (ii) above.

$$Strength = \sigma(\lambda \to \beta) = \frac{support(\lambda, \beta)}{|U|} = \frac{|C(x) \cap D(x)|}{|U|} \tag{10}$$

where λ is a combination of descriptors or condition attributes denoted $C(x)$ (e.g. relative difference from other stations), β is a decision value denoted $D(x)$ (e.g. relative difference category for current station), $\lambda \to \beta$ denotes a decision rule, $support(\lambda, \beta)$ which is read as the support of the decision rule $\lambda \to \beta$ represents the number of objects (i.e. soil moisture records) which have both properties of λ in condition attributes and β in decision values. The $|U|$ is the cardinality of U, and U is a non-empty set of objects (or records) representing the full universe of objects in the decision table (Ohrn, 1999; Pawlak et al., 1995). The $C(x)$ is the pattern for condition attributes; $D(x)$ is the pattern in decision attribute.

$$Accuracy(\lambda \to \beta) = Pr(\beta|\lambda) = \frac{support(\lambda, \beta)}{support(\lambda)} = \frac{|C(x) \cap D(x)|}{|C(x)|} \tag{11}$$

where $support(\lambda)$ and $C(x)$ represent the pattern for condition attributes.

$$Coverage(\lambda \to \beta) = Pr(\lambda|\beta) = \frac{support(\lambda, \beta)}{support(\beta)} = \frac{|C(x) \cap D(x)|}{|D(x)|} \tag{12}$$

where $support(\beta)$ and $D(x)$ represent the pattern for decision attributes.

Author's personal copy

(v) To infill a missing record at a monitoring station, the relative difference is determined for the other monitoring stations with available soil moisture record for the specified time. The estimated relative difference values are compared to the lookup (i.e. the reference) patterns to find a matching category for the current station.

(vi) The matching category is then applied to estimate the missing soil moisture record. The rough sets infilling procedure is such that different (or unique) relative difference values will be used for the same station depending on the matching between the current pattern and the lookup pattern found during the training stage. That is, the rough sets procedure provides several lookup patterns to account for a specific soil moisture condition (wet, dry, etc).

3.6. Artificial neural network models

Artificial neural networks have been used to infill hydrological variables including evapotranspiration, precipitation, air temperature, and streamflow (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b). The three broad groups of ANN which have been widely used in several studies, and used in this study to infill missing soil moisture records include: feedforward network, dynamic network, and radial basis network. Feedforward neural networks represent nonlinear static configurations with no feedback or delay components such that the output is derived from the input through a feedforward connection (Beale et al., 2012; Hagan et al., 1996). Dynamic networks, in contrast, use the direct input–output relationships together with feedbacks from current or previous inputs, outputs, or states of the network (Beale et al., 2012; Coulibaly and Evora, 2007; Hagan et al., 1996). Radial basis networks have a similar configuration as feedforward networks but use memory-based learning for their design in a way that learning is viewed as a curve-fitting problem in a high-dimensional space (Beale et al., 2012; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a; Hagan et al., 1996). Radial basis networks specifically have a single hidden layer with linear output layer and are considered local approximations. The individual ANNs are briefly described below.

1. *Feedforward neural networks (FF):* the feedforward neural network (NN) uses a basic configuration shown in Fig. 3 to connect an input layer of source nodes to an output layer of neurons (i.e., computation nodes) through a series of weights (Ananda Rao and Srinivas, 2003; Haykin, 1999). Between the input and output layer is one or more layers of hidden nodes which extract important features contained in the input data. The hyperbolic tangent sigmoid transfer function is used to generate the output from the *j*th node in the hidden layer according to Eq. (13).

$$x_j^2 = S\left(k_j^2\right) = \frac{e^{k_j^2} - e^{-k_j^2}}{e^{k_j^2} + e^{-k_j^2}} \tag{13}$$

where x_j^n is the information from node j in the nth layer, and k_j^2 is the weighted sum of the information going into jth node in the hidden layer, expressed according to Eq. (14).

$$k_j^2 = \sum_{i=1}^{l_1} w_{ij}^1 x_i^1 + b_j^2 \tag{14}$$

where x_i^n is the information from node i in the nth layer, w_{ij}^n is the weight of the connections between the ith node from the nth layer and the jth node from the $n+1$th layer, l_1 is the total number of nodes in the first layer, and l_2 is the total number of nodes in the second layer.

A linear transfer function is used to generate the output according to Eq. (15).

$$x_1^3 = \sum_{i=1}^{l_2} w_{ij}^2 x_i^2 + b_1^3 \tag{15}$$

The training process is supervised as input and output data are supplied to the network for reference. The algorithm used in the training process to adjust the weight between the connections is the Levenberg–Marquardt back propagation algorithm (Hagan et al., 1996) and the performance of the network is evaluated using the mean square error function.

2. *Fitting problem neural network (FP):* The fitting neural network is a type of feedforward neural network that is used to fit an input–output relationship (Beale et al., 2012; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a; Hagan et al., 1996). The FP

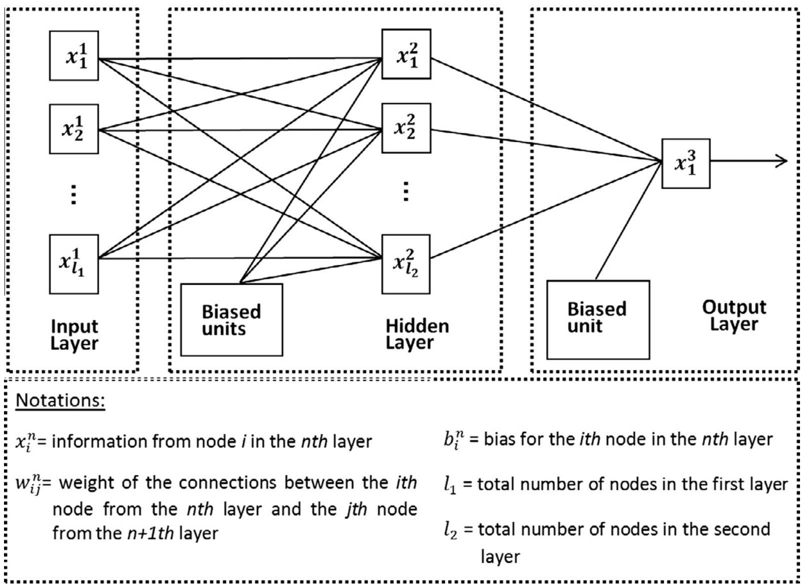


Fig. 3. Schematic configuration of the feedforward neural network.

Author's personal copy

establishes a relationship between the station under investigation (output data) and the remaining stations (input data). The output layer consists of one node which represents the station under investigation, while the input layer contains 12 nodes, each representing one of the remaining 12 stations. The network configuration and the transfer functions in the hidden and output layer are exactly the same as the feedforward neural network.

3. *Cascade-forward neural networks (CF):* The cascade-forward neural network has a similar configuration as the FF network, but includes additional layers to learn complex input–output relationships more quickly (Beale et al., 2012; Hagan et al., 1996). The additional layers allow weight connection from the input to each layer, and from each layer to the successive layers (Omaima AL-Allaf, 2012; Beale et al., 2012). The transfer function for the hidden layer in the FF layout is also used in the CF, but the transfer function for the output layer is no longer a linear function but a hyperbolic tangent sigmoid transfer function as in the hidden layer.
4. *Pattern recognition neural network (PR):* The pattern recognition neural network uses the same configuration as the FF network but includes a classification of the input data into specific classes (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a; Hagan et al., 1996). In reference to the FF configuration, the PR uses a hyperbolic tangent sigmoid transfer function for both the hidden and output layer. The hyperbolic tangent sigmoid transfer function is used in the output layer because the output of the function ranges between −1 and +1, for classification of input into different categories.
5. *Time delay neural network (TD):* The time delay network as a dynamic configuration uses feedback and delay connections to provide a dynamic interaction between layers (Beale et al., 2012; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a). The TD uses the feedforward configuration with a tapped delay line at the input. A schematic layout of the TD network is shown in Fig. 4. The tapped delay line used in this case has a delay from 1 to 2, meaning that the maximum delay of the network is a 2 time step with the prediction done for one time step ahead. The transfer functions, training algorithm and the performance function used for the network is the same as in the feedforward neural network. Due to the tapped delay line used to hold past information, a slight adjustment to the variables in the transfer function is necessary. The weighted sum of the information going into the *j*th node in the second layer (hidden layer) at time *t* is determined according to Eq. (16).

$$k_j^2(t) = \sum_{i=1}^{l_1}\sum_{d=0}^{m} w_{ij}^1(d)x_i^1(t-d) + b_j^2 \tag{16}$$

where m is the memory length of the tapped delay line, and d represents the delayed time step. The output from the transfer function of the hidden layer is estimated using Eq. (17).

$$x_j^2(t) = S\left(k_j^2(t)\right) = \frac{e^{k_j^2(t)} - e^{-k_j^2(t)}}{e^{k_j^2(t)} + e^{-k_j^2(t)}} \tag{17}$$

Since the transfer function of the output layer is a linear function and has only one output node present, the output is estimated according to Eq. (18).

$$x_1^3(t) = \sum_{i=1}^{l_2} w_{ij}^2 x_i^2(t) + b_1^3 \tag{18}$$

6. *Nonlinear autoregressive neural network (NA):* The NA uses the same network configuration as the TD network, but employs nonlinear autoregression to represent the forward dynamics of the output through a general discrete-time nonlinear estimation (Beale et al., 2012; Omaima AL-Allaf and AbdAlKader, 2011; Chow and Leung, 1996). Whereas the feedback in the TD network is limited to the input layer and feedforward networks, the feedback in the NA network is recurrent such that it encloses several layers of the network (Beale et al., 2012; Safavieh et al., 2007; Chow and Leung, 1996; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a). The NA, as a time series model, regresses the next value of the dependent output on previous values of the output and previous independent input values. According to the NA model

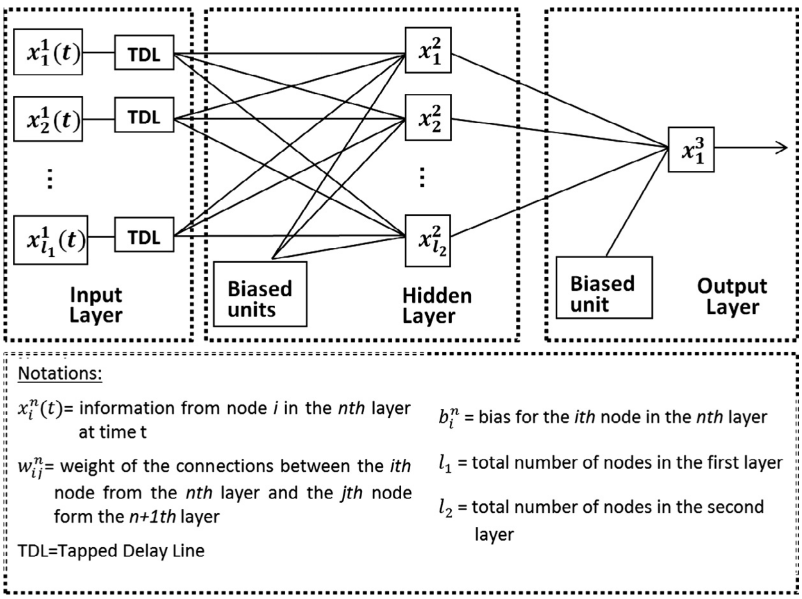


Fig. 4. Schematic configuration of the time delay neural network.

Author's personal copy

defined in Eq. (19), an accurate estimation of the output is a function of the information on previous values on the output variable y and the exogenously determined variable u such that.

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-D_y}, u_t, u_{t-1}, u_{t-2}, u_{t-3}, \ldots, u_{t-D_u}) \quad (19)$$

where y_t is the output variable of interest to be predicted, F is the nonlinear network; and u_t is the exogenously determined input variable associated with y_t. The exogenous inputs, $u_t, \ldots, u_{t-D_u}$ can be determined with an input delay line with memory of order D_u. Similarly, the endogenous inputs, $y_{t-1}, \ldots, y_{t-D_y}$ can be estimated with an input delay line with memory of order D_y. That is, the NA relates the current output value to be estimated with (i) previous values of the output variable and (ii) current and previous values of the exogenous inputs.

7. *Nonlinear autoregressive neural network with external input (NAE):* The NAE has a similar configuration as the NA neural network, with the same transfer function, training algorithm and performance function (Beale et al., 2012; Omaima AL-Allaf and AbdAlKader, 2011; Chow and Leung, 1996). While the NA does not use exogenous inputs the NAE specifically employs the external inputs to estimate the output.
8. *Exact radial basis network (RBE):* The radial basis neural network is a three layer: input, hidden and output network proposed by Broomhead and Lowe (1988). The RBE has a similar configuration as the feedforward network, but its transfer function in the hidden layer is a radial basis function. The radial basis function estimates the output by using the standard Euclidean distance between the input and its corresponding weight (Beale et al., 2012; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a; Hagan et al., 1996). That is, for each node the Euclidean distance between the center and the input vector is estimated and subsequently transformed by a nonlinear function that determines the output of the nodes in the hidden layer (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a).
9. *Generalized regression neural network (GR):* The GR network uses the same network configuration as the radial basis layer but with additional special linear layer. In the GR, the number of nodes can be as many as the number of input vectors, and nodes weighted input is the standard Euclidean distance between the input vector and its weight vector (Beale et al., 2012).

To allow the network configurations to function at their optimal level, the number of nodes in each of the layers for all the networks was determined. Typically, the number of input nodes in the input layer is determined by the number of soil moisture monitoring stations available. The NA and NAE neural networks used 13 input nodes; all the remaining networks used 12 nodes in the input layer. The optimal number of nodes in the hidden layer of the neural network is determined using a trial-and-error procedure for all networks except the exact radial basis neural network and the generalized regression neural network. The number of nodes tested ranged from 2 to 30; the optimal number of nodes and its associated trained network were used to infill the missing soil moisture record for subsequent validation period with known data.

4. Results and discussion

The application of the 14 infilling methods is presented in two stages. The first stage illustrates the spatial–temporal variation of soil moisture for the 13 monitoring stations; whereas the second stage presents the evaluation of the infilling methods to estimate the soil moisture records. The infilling methods are evaluated in three phases: surface soil layer, second soil layer, and the third soil layer independently.

4.1. Space–time variation of station soil moisture

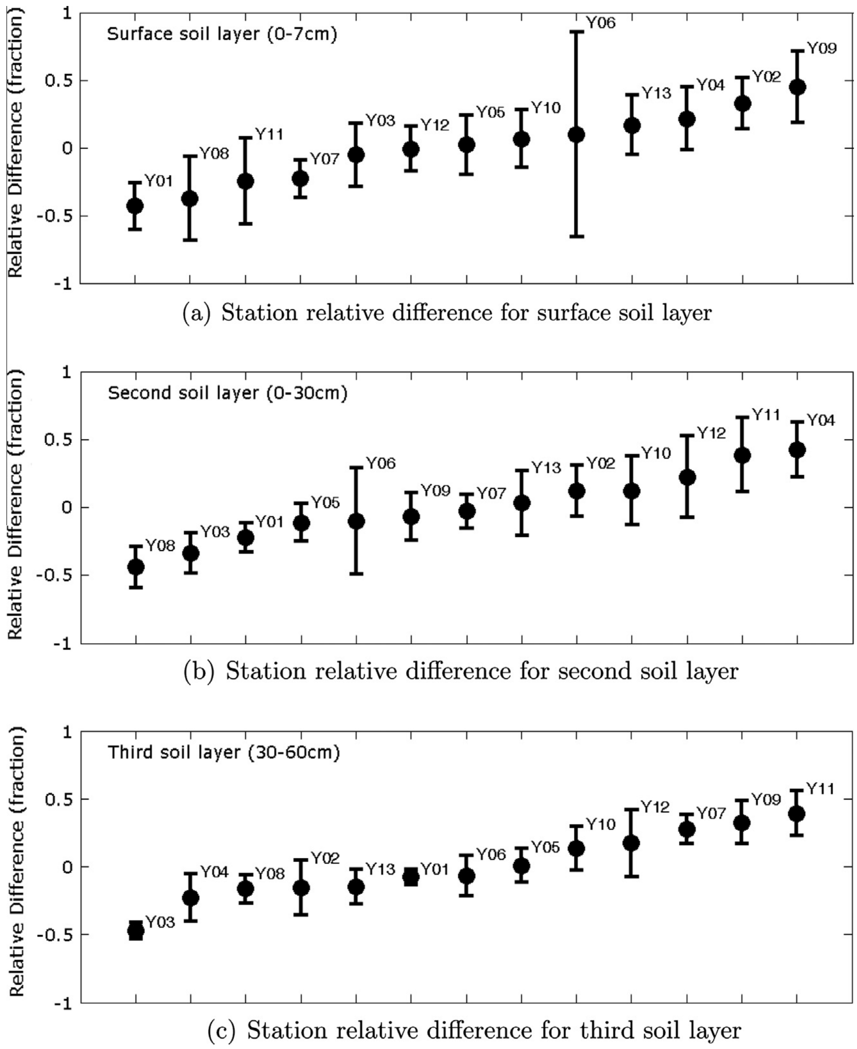
The space–time variation of soil moisture is examined through evaluation of the seasonal distribution based on the relative difference across the 13 monitoring stations for each soil layer. The overall relative difference for each monitoring station which was estimated using the entire period of data in the training data set is shown for the three soil layers in Fig. 5. The level of variability (based on the standard deviation) in the estimated relative differences is highest at the surface soil layer due to its dynamic interaction with the atmosphere, and lowest at the deeper (i.e. the third) soil layer. These differences in the level of variability across soil layers are consistent with findings in Dumedah and Coulibaly (2011); Brocca et al. (2010); Brocca et al. (2009); Cosh et al. (2008); Cosh et al. (2006); Martinez Fernandez and Ceballos (2005); and Martnez Fernndez and Ceballos (2003). Moreover, the result shows and identifies for each soil layer, the relative soil moisture content of the monitoring stations. Across the three soil layers, there are monitoring stations including Y05, Y08, and Y10 with varied soil moisture conditions which have retained temporally consistent moisture conditions relative to other stations. However, inconsistent relative moisture conditions are observed at monitoring stations including Y01, Y04, and Y09. Overall, the pattern of relative soil moisture conditions across the monitoring stations is not consistent between different soil layers when the entire period of the training data set was used.

To further examine the differences in moisture conditions for the soil layers, the seasonal distribution of the relative moisture conditions is examined through a month-by-month evaluation of the relative differences. The month-by-month variations of the station relative differences are presented for the second and third soil layers in Figs. 6 and 7. Due to the overlapping soil depths between the surface (0–7 cm) layer and the second (0–30 cm) soil layer, the plot for the surface soil layer is not shown here. The relative moisture conditions in these results remain consistent from month-to-month, with relatively wetter and drier stations retaining their moisture conditions for both the second and third soil layers. In addition, the relative moisture conditions remain consistent between the two soil layers, such that relatively drier stations such as Y03 and Y08, and wetter stations including Y11 and Y12 have retained their moisture conditions across both layers. Overall, relative moisture conditions are far more consistent between the two soil layers in this month-by-month evaluation compared to the lumped evaluation based on the entire period of the training data set. The consistency of relative soil moisture for the month-by-month evaluation suggests that the soil moisture memory (i.e., the persistence of wet or dry condition) is strongly dependent on seasonality. That is, the soil moisture content at a given station in relation to other stations is highly stable when the overall soil moisture condition across all stations is either dry or wet. But the entire data period masks or conceals the soil moisture dynamics associated with the different seasons.

4.2. Evaluating the soil moisture estimation accuracy for infilling methods

The infilling methods were applied to estimate missing soil moisture records separately for each of the three soil layers. The infilling methods were assessed using both the training and the validation data sets. The estimated soil moisture values are

Author's personal copy



(a) Station relative difference for surface soil layer

(b) Station relative difference for second soil layer

(c) Station relative difference for third soil layer

Fig. 5. Comparison between overall station relative differences for the three soil layers. The relative differences were estimated using only the training data set.

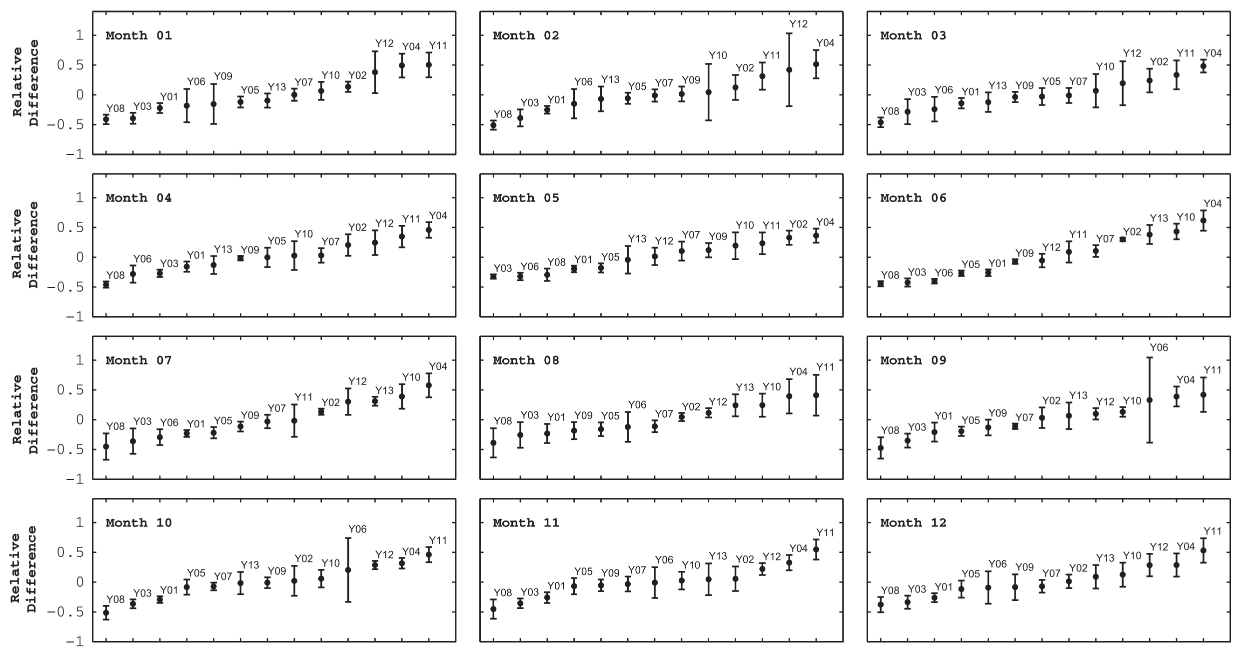


Fig. 6. Month-by-month variation of station relative differences for the second soil layer. The relative differences were estimated using only the training data set.

Author's personal copy

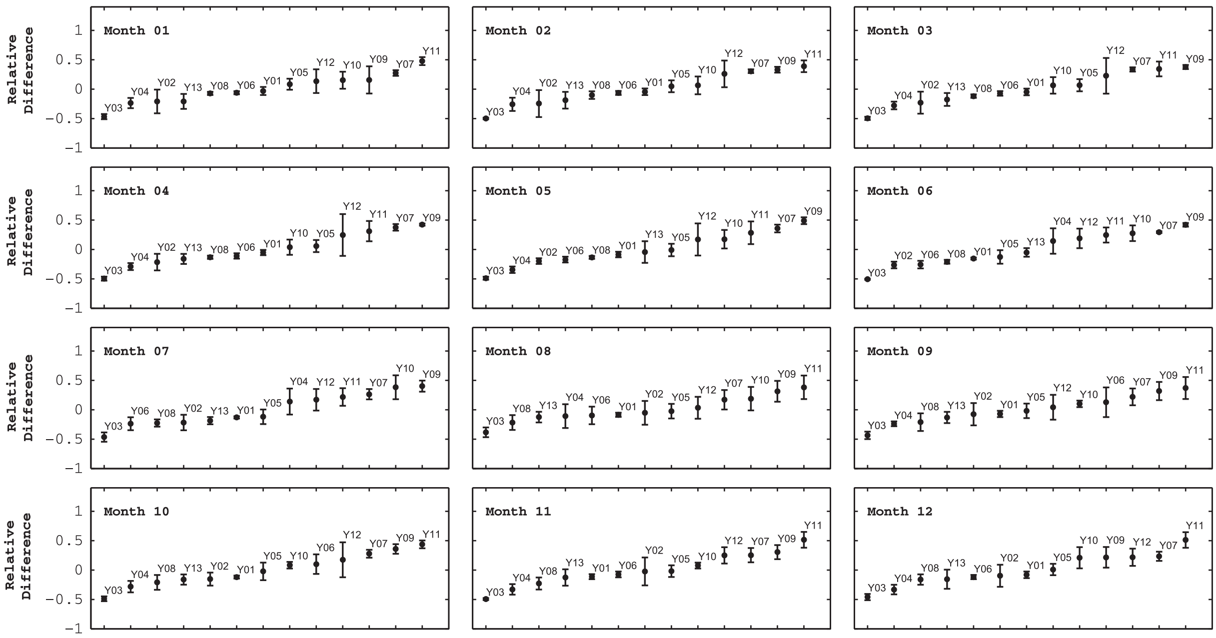


Fig. 7. Month-by-month variation of station relative differences for the third soil layer. The relative differences were estimated using only the training data set.

assessed using two widely used evaluation measures: the root mean square error (RMSE), and the coefficient of determination (R^2). The RMSE quantifies the overall predictive error such that soil moisture estimates with RMSE values closer to zero indicate accurate estimates, and higher RMSE values represent inaccurate soil moisture estimation. The R^2 measures the predictive power or the goodness of fit of the infilling method; its values vary between 0 and 1, with values closer to 1 indicating accurate estimates, and those closer to zero representing inaccurate soil moisture estimation.

The estimated values for the two evaluation measures (RMSE and R^2) based on the training data set for all 13 monitoring stations for each soil layer are shown in Table 2. All the infilling methods have an overall RMSE of about 0.03 m^3/m^3 or better, except the radial basis neural network method which has unacceptably high RMSE values greater than 0.2 m^3/m^3 for the difference soil layers. The top three high performing methods are the nonlinear auto regressive neural network, the rough sets method, and the generalized regression neural network. Since the majority of the infilling methods show a high estimation accuracy for the training data set, they can be safely used to infill missing values represented by the validation data set. Moreover, the soil moisture sensor has an accuracy limit of about 0.03 m^3/m^3, which is equivalent to the overall RMSE obtained by most of the infilling methods.

Using the validation data set, the known soil moisture records and the estimated values determined from the infilling methods are compared for the 13 monitoring stations in Fig. 8 for the surface soil layer, Fig. 9 for the second soil layer, and Fig. 10 for the third soil layer. In the surface soil layer estimations, 8 of the infilling methods have RMSE values less than 0.1 m^3/m^3 and R^2 values greater than 0.5. The highest performing infilling method is the nonlinear autoregressive neural network showing a highly accurate prediction of the missing soil moisture with RMSE of 0.03 m^3/m^3 and R^2 of 0.86. The remaining ANN methods performed poorly, on the basis of RMSE and R^2 measures compared to the statistical methods. The merged method as a weighted combination of the station layer relative difference, monthly replacement method and the weighted Pearson correlation, performed

Table 2
Values of evaluation measures for all infilling methods based on the training data set for the first three soil layers.

Infilling methods	0–5 cm		0–30 cm		30–60 cm	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Station relative difference	0.0331	0.886	0.0365	0.868	0.0397	0.848
Pearson correlation method	0.0394	0.867	0.0552	0.820	0.0789	0.498
Monthly replacement method	0.0250	0.857	0.0224	0.892	0.0216	0.856
Merged method	0.0225	0.898	0.0216	0.904	0.0257	0.717
Rough sets method – 4 sets	0.0276	0.961	0.0268	0.935	0.0199	0.962
Rough sets method – 5 sets	0.0273	0.957	0.0254	0.953	0.0233	0.960
Rough sets method – 6 sets	0.0303	0.948	0.0272	0.950	0.0232	0.964
Cascade forward neural network	0.0005	0.680	0.0005	0.786	0.0436	0.742
Feedforward neural network	0.0005	0.709	0.0003	0.768	0.0961	0.632
Fitting problem with neural network	0.0005	0.724	0.0004	0.730	0.0762	0.772
Nonlinear auto regressive NN	0.0001	0.978	0.0000	0.997	0.0025	0.999
Nonlinear auto regressive, exact input NN	0.0006	0.577	0.0003	0.827	0.0412	0.949
Generalized regression neural network	0.0003	0.905	0.0003	0.921	0.0221	0.961
Pattern recognition neural network	0.0005	0.768	0.0005	0.782	0.0395	0.747
Time delay neural network	0.0006	0.652	0.0000	0.652	0.0528	0.757
Exact radial basis neural network	0.2717	0.432	0.1466	0.323	0.2928	0.306

Author's personal copy

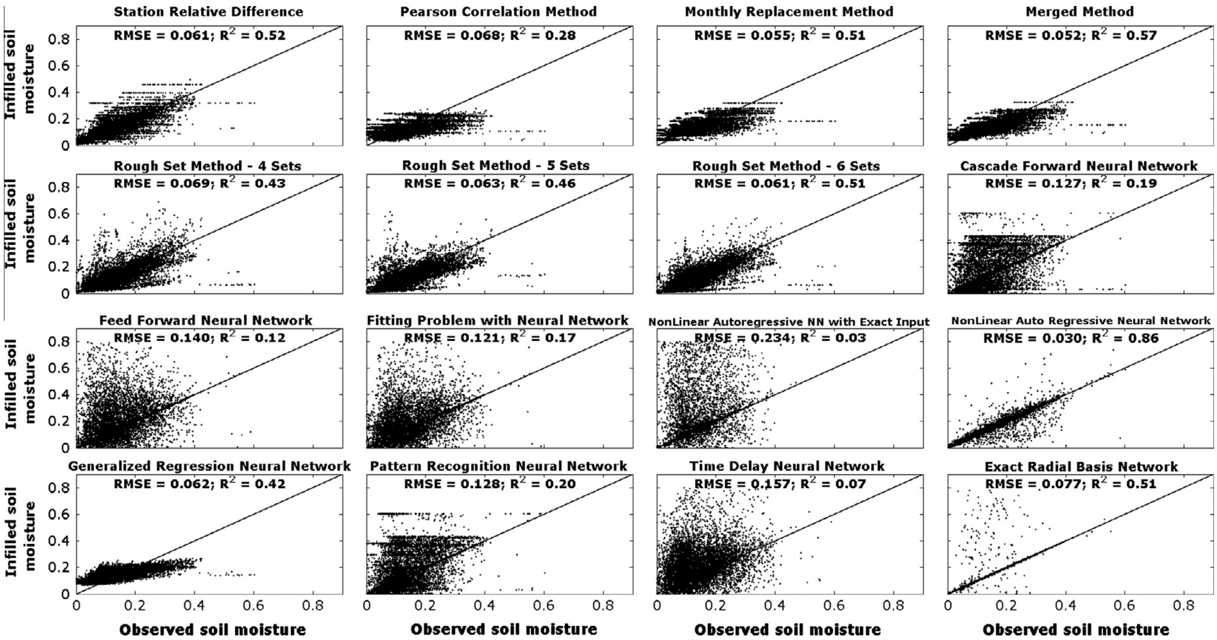


Fig. 8. Validation of infilling methods through a comparison between observed soil moisture (in m^3/m^3) and estimated soil moisture (in m^3/m^3) infilled using the various methods for the surface soil layer across all 13 monitoring stations.

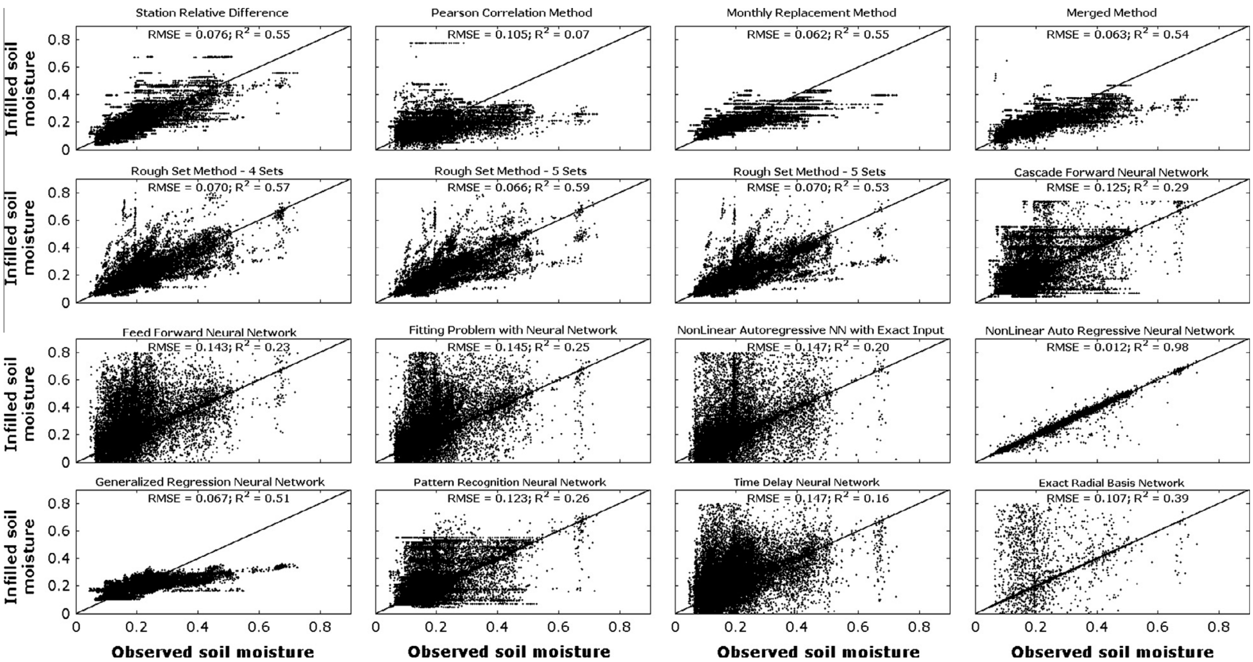


Fig. 9. Validation of infilling methods through a comparison between observed soil moisture (in m^3/m^3) and estimated soil moisture (in m^3/m^3) infilled using the various methods for the second soil layer across all 13 monitoring stations.

better than the individual methods that were combined (based on RMSE and R^2). The statistical methods show stratification in their estimations which is due to loss of soil moisture variability for different times within months with multiple missing records. The performance of the rough sets method (for set categories of 4, 5, and 6) is notable with high temporal consistency between the known and estimated soil moisture values. The set categories represent the number of relative difference patterns subject to a specific soil moisture condition that were used to estimate the missing records. The rough sets estimations show no stratifications as found in the other 4 statistical methods. It is noteworthy that different set categories ranging from 2 to 8 have been evaluated for the rough sets method, and that the most competitive set categories are 4, 5, and 6 which are the ones presented for evaluation.

The evaluation of the infilling methods for the second and third soil layers show similar levels of performance between the methods, indicating that different soil layers have limited effect on the estimation accuracy of the infilling methods. The overall level of performance based on RMSE and R^2 values for the infilling methods is shown in Fig. 11 for each of the three soil layers. The result shows the decreasing order of performance for the methods, with the nonlinear autoregressive neural network remaining the highest

Author's personal copy

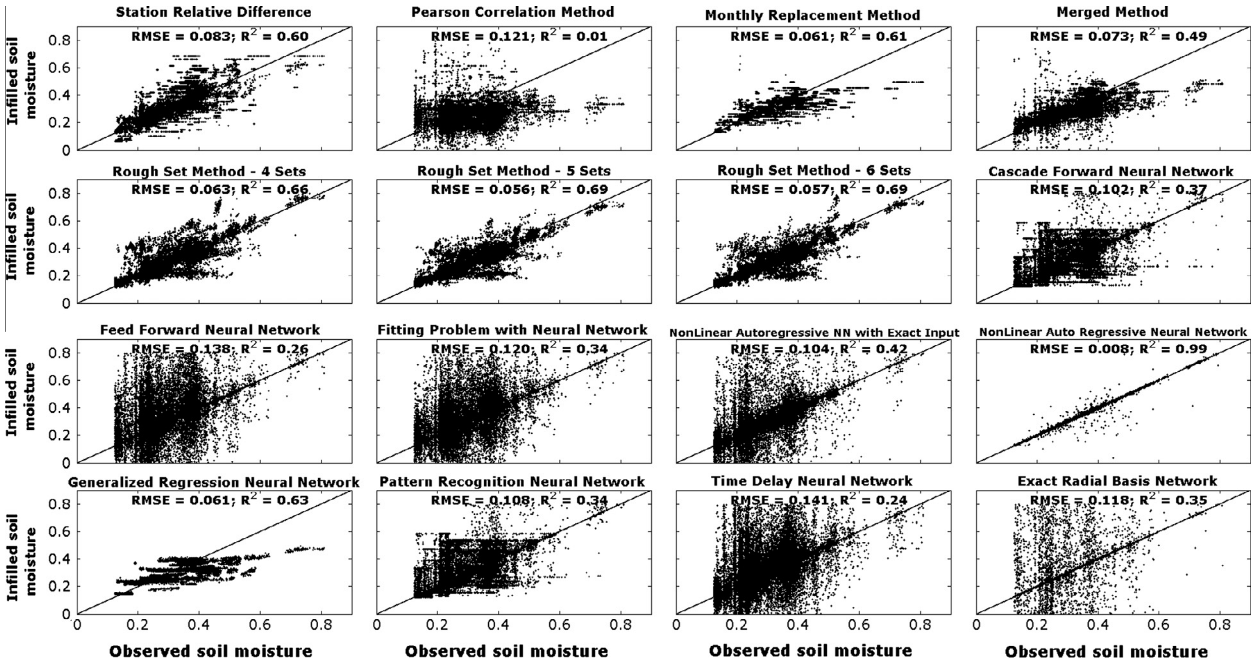


Fig. 10. Validation of infilling methods through a comparison between observed soil moisture (in m^3/m^3) and estimated soil moisture (in m^3/m^3) infilled using the various methods for the third soil layer across all 13 monitoring stations.

performing method. Overall, the rough sets method is the second highest performing method followed by the monthly replacement, merged method, station relative difference, and so on to the exact radial basis neural network representing the least performing method. In addition, the RMSE values for the infilling methods remain similar across the three soil layers, whereas the temporal accuracy based on R^2 values has increased from the first layer through to the third layer. The increased temporal accuracy with increasing soil depth is due to less variability in soil moisture for the deeper soil layer, whereas the surface soil layer is highly variable because of its dynamic interaction with the atmosphere. Mostly, infilling methods tend to accurately estimate missing values in less varied data sets than highly dynamic data sets such as that represented in the surface soil layer.

4.3. Implications on using the infilling methods

Together with the evaluation of infilling methods presented in the previous section, it is important to outline the implications of using the above methods for estimating missing soil moisture records. In the three categories of ANNs examined, the feedforward group (FF, CF, FP and PR), the dynamic group (TD, NA, and NAE), and the radial basis group (RBE, GR) show different levels of accuracy to estimate the missing soil moisture records. Across the three soil layers, the dynamic group has the highest soil moisture estimation accuracy followed by the feedforward group, and the radial basis representing the least performing configuration.

The high estimation accuracy found in the dynamic network is due to its evaluation of feedback between layers through discrete-time nonlinear estimation, with missing values estimated from distinct time delays. Soil moisture as a spatial–temporal variable, dynamically responds to changes in antecedent moisture conditions and the atmosphere. The decreasing order of performance in the dynamic group is NA, NAE, and TD; where NA is the highest performing of all the infilling methods evaluated. The differences in estimation accuracy show that the NA method best represents the dynamics of the missing soil moisture record.

In the feedforward group the decreasing order of performance is CF, PR, FP and FF. These four networks in the feedforward group all use the same transfer function for the hidden layer, but with different transfer functions for the output layer. The FF and FP use a linear transfer function whereas the CF and PR use a hyperbolic tangent sigmoid transfer function. Given the relatively higher estimation accuracy for CF and PR compared to FF and FP, the hyperbolic tangent sigmoid transfer function is better suited for infilling missing soil moisture records, as the linear transfer function is unable to accurately represent the complex relationships in the missing soil moisture records. The accurate estimation in the CF is due in part to additional hidden layers used in its training stage.

In the radial basis group, the GR has a higher estimation accuracy than the RBE network which was found to be the least performing network of all methods evaluated. The performance of the GR network is notable, ranking 6th and 8th for RMSE and R^2 respectively across all 16 infilling approaches. Overall, regression-based neural networks have higher estimation accuracy and are preferred to the other neural networks when infilling missing soil moisture records.

It is important to outline drawbacks of using ANN to infill missing soil moisture records. Generally, ANNs require training to generate the prediction in a way that the trained network is highly dependent on the amount of data records added to the training data set. For practical purposes, the number of records in the training data set may change due to additional temporal data records, or soil moisture monitoring stations. Due to the computational procedure of network training, the general neural network is expected to be considerably affected by the amount of data in the training data set compared to statistical methods. The required training of the neural network prior to prediction and its dependency on changes in the training data set is an important limitation on the estimation procedure. Note that the changes in training data set referred to in this case are limited to the additional soil moisture data added to the training data set from sources such as additional temporal records or data records from additional monitoring stations.

Author's personal copy

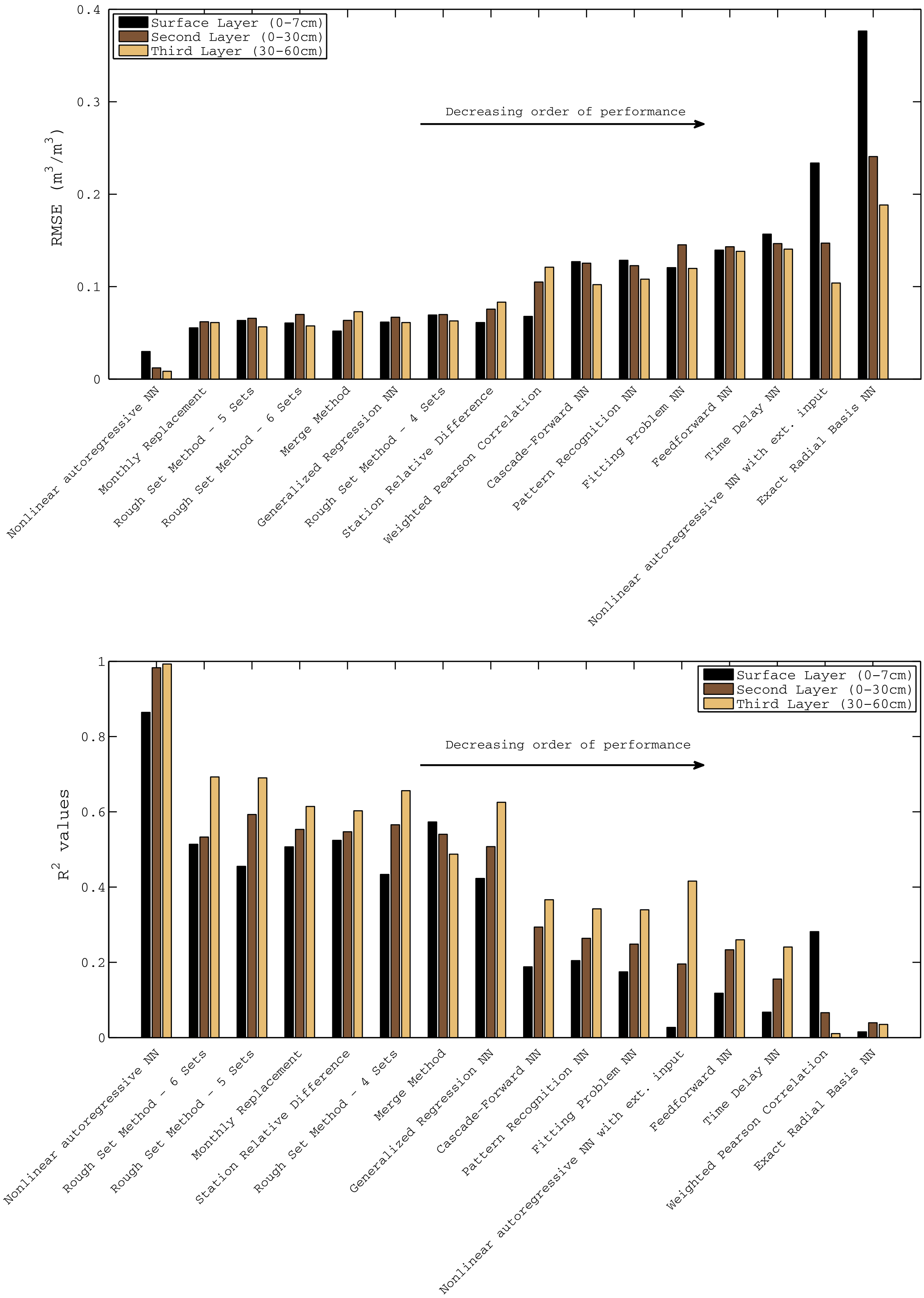


Fig. 11. Overall performance of the infilling methods based on the RMSE (the top panel) and R^2 (the bottom panel) evaluations across the 13 monitoring stations for the three soil layers.

The statistical methods along with the rough sets technique have estimation accuracy competitive to those obtained from the neural network configurations. In the statistical and rough sets methods, the decreasing order of performance is: rough sets, monthly replacement, merged method, station relative difference, and weighted Pearson correlation method. The high estimation accuracy in the rough sets method shows that temporal persistence of soil moisture when properly classified can provide a highly competitive estimation of missing soil moisture record. The rough sets approach consistently has a higher estimation

Author's personal copy

accuracy than the station layer relative difference for each of the three soil layers, meaning that the rough sets grouping which accounts for the unique moisture conditions provides a better estimation procedure than the lumped station relative difference across the entire time period. Specifically, the rough sets method has an accuracy increase of 5% in RMSE and 3% in R^2 when compared to the station relative difference method, thus demonstrating the significance of the rough sets approximation.

5. Conclusions

This study has evaluated 14 infilling methods including artificial neural network and statistical techniques, to estimate missing soil moisture records at 13 monitoring stations independently for three different soil layers. An evaluation of the estimated soil moisture values against known records showed that the top three highest performing methods are the nonlinear autoregressive neural network, the rough sets method, and the monthly replacement. The high estimation accuracy (RMSE of 0.03 m^3/m^3) found in the NA network was the result of its regression based dynamic network, which allows feedback connections through discrete-time estimation. Despite the high estimation accuracy of the NA network, the ANNs in general lack a space–time explanation or any insight into the relative soil moisture conditions at the monitoring stations. Hence, the NA method is best suited for soil moisture estimations for cases where the physical space–time relationships between monitoring stations are not of primary focus.

The rough sets approach is advantageous because of its equally high estimation accuracy (RMSE of 0.05 m^3/m^3) associated with its pattern-based and space–time explanation of relative moisture conditions across monitoring stations. The equally high estimation accuracy in the rough sets procedure illustrates the important role of temporal persistence in soil moisture and its grouping to account for different soil moisture conditions (e.g. wet, dry, etc). The estimation procedure also illustrates the utility of the rough sets approximation to determine patterns of temporal persistence of soil moisture that are relevant to different moisture conditions. Consequently, the rough sets method is the preferred approach to extrapolating soil moisture in space and time for other locations, with the capability to account for seasonality. It is noteworthy that while the monthly replacement provides an accurate estimation of the missing soil moisture, its infilled values are usually stratified into layers due to multiple missing values in the same month.

The findings from this study point to the potential of these methods to infill other hydrologic variables such as air and soil temperature, precipitation, and evapotranspiration, which are also plagued with missing records. Moreover, the findings are directly applicable to soil moisture time series data from remote sensing, which can be affected by missing records due to radio frequency interference, frozen and ice conditions, problems with retrieval algorithm convergence, and satellite orbital issues.

Acknowledgements

This work was supported by funding from the Australian Research Council (DP0879212). The program for the infilling methods used herein is available upon request from the first author at: dgiftman@hotmail.com.

References

Abebe, A., Solomatine, D., Venneker, R., 2000. Application of adaptive fuzzy rule – based models for reconstruction of missing precipitation events. Hydrol. Sci. J. 45 (3), 425–436.
Abudu, S., Salim Bawazir, A., Phillip King, J., 2010. Infilling missing daily evapotranspiration data using neural networks. J. Irrig. Drain. Eng. 136 (5), 317–325.
Ananda Rao, M., Srinivas, J., 2003. Neural Networks: Algorithms and Applications. Alpha Science International, Pangbourne, ISBN-13: 978-1842651315. p. 239.
ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology I: preliminary concepts. J. Hydrol. Eng. 5 (2), 115–123.
ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology II: hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.
Beale, M.H., Hagan, M.T., Demuth, H.B., 2012. Neural Network Toolbox: Users Guide. The MathWorks, Inc.
Berg, A.A., Mulroy, K., 2006. Streamflow predictability given macro-scale estimates of the initial soil moisture status. Hydrol. Sci. J. 51 (4), 642–654.
Brocca, L., Melone, F., Moramarco, T., Morbidelli, R., 2009. Soil moisture temporal stability over experimental areas in central italy. Geoderma 148, 364–374.
Brocca, L., Melone, F., Moramarco, T., Morbidelli, R., 2010. Spatial–temporal variability of soil moisture and its estimation across scales. Water Resour. Res. 43, W02516.
Broomhead, D.S., Lowe, D., 1988. Multivariable functional interpolation and adaptive networks. Complex Syst. 2, 321–355.
Chow, T.W.S., Leung, C.T., 1996. Nonlinear autoregressive integrated neural network del for short-term load forecasting. IEEE Proc. – Gener., Trans. Distrib 43 (5), 500–506.
Cosh, M.H., Jackson, T.J., Starks, P., Heathman, G., 2006. Temporal stability of surface soil moisture in the little washita river watershed and its applications in satellite soil moisture product validation. J. Hydrol. 323 (1-4), 168–177.
Cosh, M.H., Jackson, T.J., Moran, S., Bindlish, R., 2008. Temporal persistence and stability of surface soil moisture in a semi – arid watershed. Remote Sens. Environ. 112, 304–313.
Coulibaly, P., Evora, N., 2007. Comparison of neural network methods for infilling missing daily weather records. J. Hydrol. 341 (1-2), 27–41.
Dumedah, G., Coulibaly, P., 2011. Evaluation of statistical methods for infilling missing values in high-resolution soil moisture data. J. Hydrol. 400 (1-2), 95102. http://dx.doi.org/10.1016/j.jhydrol.2011.01.028.
Dumedah, G., Schuurman, N., 2008. Minimizing the effects of inaccurate sediment description in borehole data using rough set theory and transition probability. J. Geogr. Syst. 10 (3), 291–315.
Dumedah, G., Schuurman, N., Yang, W., 2008. Minimizing effects of scale distortion for spatially grouped census data using rough sets. J. Geogr. Syst. 10 (1), 47–69.
Elshorbagy, A.A., Panu, U.S., Simonovic, S.P., 2000. Group – based estimation of missing hydrological data: Ii. Application to streamflows. Hydrol. Sci. J. 45 (6), 867–880.
French, M., Krajewski, W., Cuykendall, R., 1992. Rainfall forecasting in space and time using a neural network. J. Hydrol. 137 (1-4), 1–31.
Gao, X., Wu, P., Zhao, X., Wang, J., Shi, Y., Zhang, B., Tian, L., Li, H., 2013. Estimation of spatial soil moisture averages in a large gully of the loess plateau of china through statistical and modeling solutions. J. Hydrol. 486, 466–478, <http://www.sciencedirect.com/science/article/pii/S0022169413001492>.
Hagan, M., Demuth, H., Beale, M., 1996. Neural network design. Electrical Engineering Series. Brooks/Cole, <http://books.google.com.au/books?id=cUNJAAAACAAJ>.
Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice-Hall, Englewood Cliffs, NJ, vol. 2nd ed.
Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H., Goodrich, D.C., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour. Res. 34 (12), 3405–3420.
Langevin, C.D., Panday, S., 2012. Future of groundwater modeling. Ground Water 50, 334–339. http://dx.doi.org/10.1111/j.1745-6584.2012.00937.x.
Legates, D.R., Mahmood, R., Levia, D.F., DeLiberty, T.L., Quiring, S.M., Houser, C., Nelson, F.E., 2010. Soil moisture: a central and unifying theme in physical geography. Prog. Phys. Geogr. 34 (6). http://dx.doi.org/10.1177/0309133310386514.
Luck, K., Ball, J., Sharma, A., 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J. Hydrol. 227 (1-4), 56–65.
Mahmood, R., 1996. Scale issues in soil moisture modeling: problems and prospects. Prog. Phys. Geogr. 20 (3), 273–291.
Martinez Fernandez, J., Ceballos, A., 2005. Mean soil moisture estimation using temporal stability analysis. J. Hydrol. 312, 28–38.
Martnez Fernndez, J., Ceballos, A., 2003. Temporal stability of soil moisture in a large – field experiment in spain. Soil Sci. Soc. Am. J. 67, 1647–1656.
McKenzie, N., Jacquier, D., Ashton, L., Cresswell, H., 2000. Estimation of Soil Properties Using the Atlas of Australian Soils. Tech. Rep., CSIRO Land and Water Technical Report 11/00. <http://www.clw.csiro.au/publications/technical2000/>.
Merlin, O., Walker, J., Panciera, R., Young, R., Kalma, J., Kim, E., December 2007. Calibration of a soil moisture sensor in heterogeneous terrain with the National Airborne Field Experiment (NAFE) data. In: MODSIM2007, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 2604–2610.
Mwale, F.D., Adeloye, A.J., Rustum, R., 2012. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi - A self organizing map approach. Phys. Chem. Earth, 34–43.
Ng, W., Panu, U.S., 2010. Comparisons of traditional and novel stochastic models for the generation of daily precipitation occurrences. J. Hydrol. 380 (1-2), 222–236.

Author's personal copy

Ng, W., Panu, U.S., Lennox, W., 2009. Comparative studies in problems of missing extreme daily streamflow records. J. Hydrol. Eng. 14 (1), 91–100.
Nkuna, T.R., Odiyo, J.O., 2011. Filling of missing rainfall data in Luvuvhu River Catchment using artificial neural networks. Phys. Chem. Earth 36, 830–835.
Ohrn, A., 1999. Discernibility and Rough Sets in Medicine: Tools and Applications, Computer and Information Science, Norwegian University of Science and Technology, Trondheim.
Omaima AL-Allaf, N.A., 2012. Cascade-Forward vs. Function Fitting Neural Network for Improving Image Quality and Learning Time in Image Compression System. In: World Congress on Engineering 2012 (WCE 2012), The 2012 International Conference of Signal and Image Engineering (ICSIE2012), Imperial College London, in London, UK, 4-6 July, 2012.
Omaima AL-Allaf, N.A., AbdAlKader, S.A., 2011. Nonlinear Autoregressive neural network for estimation soil temperature: a comparison of different optimization neural network algorithms. UBICC J. Issue ICIT2011 6 (4), 43–51.
Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inform. Sci. 11, 341–356. http://dx.doi.org/10.1007/BF01001956.
Pawlak, Z., 1997. Rough set approach to knowledge-based decision support. Eur. J. Oper. Res. 99, 48–57. http://dx.doi.org/10.1016/S0377–2217(96)00382–7.
Pawlak, Z., Grzymala-Bausse, J., Slowinski, R., Ziarko, W., 1995. Rough sets. Commun. ACM 38 (11), 89–95. http://dx.doi.org/10.1145/219717.219791.
Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. J. Clim. 15 (13), 1609–1625.
Rüdiger, C., Western, A., Walker, J., Smith, A., Kalma, J., Willgoose, G., 2010. Towards a general equation for frequency domain reflectometers. J. Hydrol. 383, 319–329.
Safavieh, E., Andalib, S., Andalib, A., 2007. Forecasting the unknown dynamics in NN3 database using a nonlinear autoregressive recurrent neural network. In: Proceedings of International Joint Conference on Neural Networks (IJCNN), Orlando, FL, pp. 2105–2109. <http://dx.doi.org/10.1109/IJCNN.2007.4371283>.
Schneider, T., 2001. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. J. Clim. 14 (5), 853–871.
Smith, A.B., Walker, J.P., Western, A.W., Young, R.I., Ellett, K.M., Pipunic, R.C., Grayson, R.B., Siriwardena, L., Chiew, F.H.S., Richter, H., 2012. The murrumbidgee soil moisture monitoring network data set. Water Resour. Res. 48 (W07701). http://dx.doi.org/10.1029/2012WR011976.
Thornthwaite, C.W., 1961. The task ahead. Ann. Assoc. Am. Geogr. 51 (4), 345–356.
Trenberth, K.E., Guillemot, C.J., 1998. Evaluation of the atmospheric moisture and hydrological cycle in the ncep/ncar reanalyses. Clim. Dyn. 14 (3), 213–231.
Vachaud, G., Passerat De Silans, A., Balabanis, P., Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function. Soil Sci. Soc. Am. J. 49, 822–828.
Wang, G., Garcia, D., Liu, Y., de Jeu, R., Dolman, A.J., 2012. A three-dimensional gap filling method for large geophysical datasets: application to global satellite soil moisture observations. Environ. Model. Softw. 30, 139–142, <http://www.sciencedirect.com/science/article/pii/S1364815211002453>.
Western, A., Seyfried, M., 2005. A calibration and temperature correction procedure for the water-content reflectometer. Hydrol. Proces. 19, 3785–3793.
Western, A., Young, R., Chiew, F., 2000. Murrumbidgee Soil Moisture Monitoring Network Field Calibration. Cooperative Research Centre for Catchment Hydrology, Dept. of Civil and Environmental Engineering, The University of Melbourne, Australia.
Yeoh, N., Walker, J., Young, R., Rüdiger, C., Smith, A., Ellett, K., Pipunic, R., Western, A., 2008. Calibration of the Murrumbidgee Monitoring Network CS616 Soil Moisture Sensors. Master's Thesis, Dept. of Civil and Environmental Engineering, The University of Melbourne, Australia.