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.

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/

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.

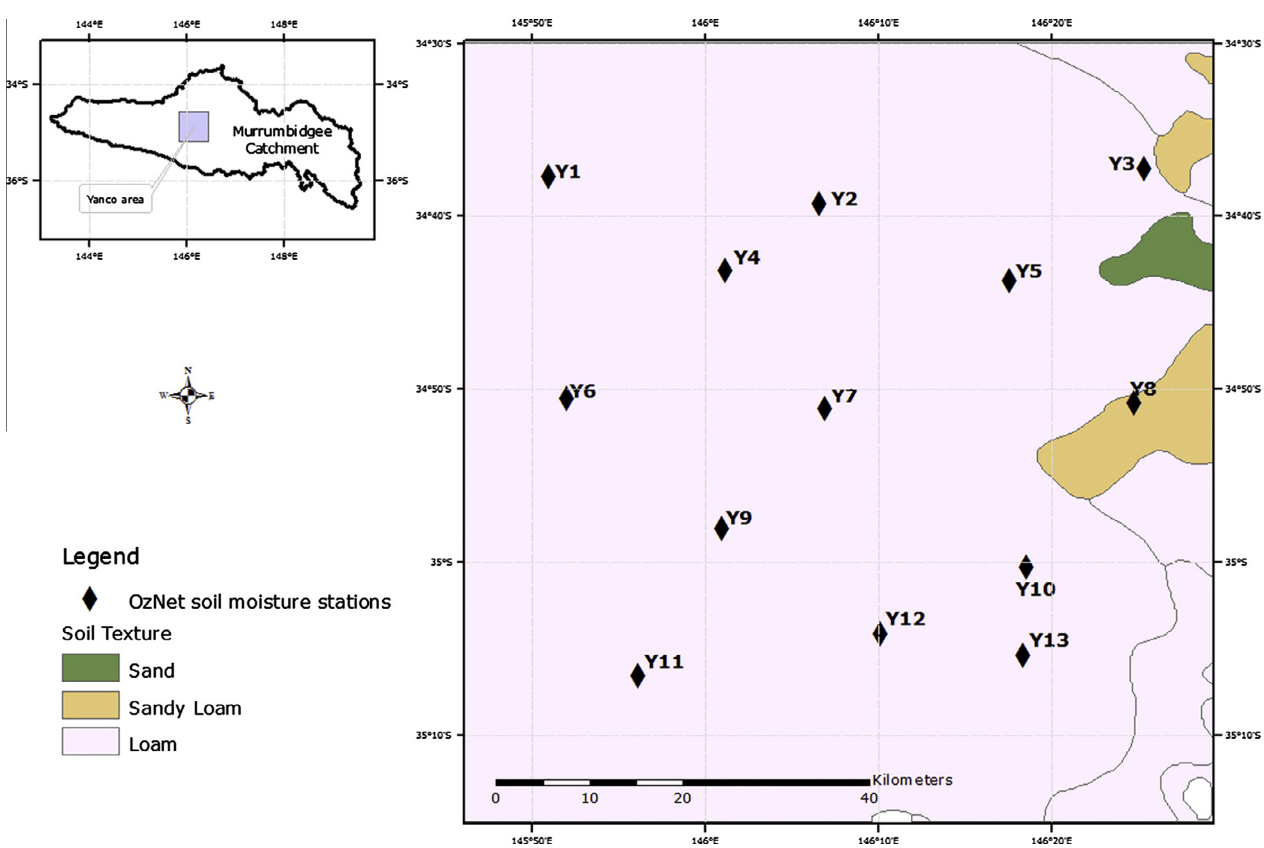

Fig. 1. Yanco study area in south-east Australia showing the location of soil moisture stations and the soil texture distribution.