Computers and Electronics in Agriculture 148 (2018) 95–106

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems

Xianming Dou, Yongguo Yang*

Key Laboratory of Coalbed Methane Resources and Reservoir Formation Process of Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China

School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Keywords:
Evapotranspiration
Machine learning
Adaptive neuro-fuzzy inference system
Extreme learning machine
Terrestrial ecosystems
Eddy covariance

ABSTRACT

Elucidating the biophysical mechanisms governing the exchange of water vapor between land and the atmosphere is particularly crucial for addressing water scarcity under climate change. Owing to the rapid development of machine learning techniques, a series of powerful tools have been proposed over the past two decades, allowing the scientific community to obtain new insights into the patterns of evapotranspiration (ET) on different spatial scales ranging from ecosystem to global. The primary focus of this study was to investigate the feasibility and effectiveness of both extreme learning machine (ELM) and adaptive neuro-fuzzy inference system (ANFIS) to estimate the daily ET with flux tower observations in four main types of ecosystems. A comparative research was undertaken to evaluate the potential of the models compared with the conventional artificial neural network and support vector machine models. All the developed models were evaluated according to the following performance indices: coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE), root mean square error and mean absolute error. The results showed that all the applied models had high performance for modeling daily ET (e.g., R^2 = 0.9398–0.9593 and NSE = 0.8877–0.9147 in forest ecosystem). Among the applied ELM models, the three hybrid ELM methods outperformed the original ELM method in most cases at the four sites and the computational time required for learning these ELM models has been considerably reduced. The subtractive clustering and fuzzy c-means clustering algorithms for ANFIS generally performed better than the grid partitioning algorithm. It was concluded that the advanced ELM and ANFIS models can be recommended as important complements to traditional methods due to their robustness and flexibility. Moreover, significant difference regarding the modeling performance existed among the four major ecosystems types. The models generally achieved the best performance in forest ecosystem, while provided the worst in cropland ecosystem.

1. Introduction

Evapotranspiration (ET) of terrestrial ecosystems is a major contributor to water balance in both regional and global scales (Huntington, 2006; Ukkola and Prentice, 2013), and it is widely employed to measure the amounts of total water loss through several key processes between land and the atmosphere, including plant transpiration, canopy evaporation and soil evaporation (Wang and Dickinson, 2012). In addition, terrestrial ET is strongly related to the carbon assimilation of terrestrial ecosystems. Due to their close coupling, special attention has been placed on exploring the complex interaction between carbon and water cycles in recent years (Halladay and Good, 2017; Medlyn and De Kauwe, 2013). Moreover, numerous studies have suggested that a variety of ecosystem processes and relevant parameters, such as ecosystem productivity, water and energy balances, and soil water content are substantially effected by ET (Halladay and Good, 2017; Mystakidis et al., 2016). Considerable progress in understanding the magnitudes of ET in different terrestrial ecosystems and the relevant mechanisms of ongoing variations in ET on daily, seasonal, annual and inter-annual time scales has been made, primarily based on the global flux tower measurements using the eddy covariance technique (Baldocchi et al., 2001). Despite these findings, quantifying the nonlinear processes that dominate ET remains a great challenge, partly because that the structure and functioning of terrestrial ecosystems can be appreciably altered by the varying global environmental change, mainly including climate change and human

* Corresponding author at: Key Laboratory of Coalbed Methane Resources and Reservoir Formation Process of Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China.

E-mail address: yongguoyang@hotmail.com (Y. Yang).

https://doi.org/10.1016/j.compag.2018.03.010
Received 27 November 2017; Received in revised form 21 February 2018; Accepted 10 March 2018
0168-1699/

practical activities (Mao et al., 2015; Polhamus et al., 2013). In this context, precisely estimating ET in terrestrial ecosystems is of particular importance for water resource management and this especially critical issue needs to be addressed in the present study.

Numerous methods have been proposed in recent years to estimate terrestrial ET at different spatial scales ranging from site to global scale (Hirschi et al., 2017; Wang and Dickinson, 2012). These approaches may be commonly divided into two categories. One major model type involves many different remote sensing-based ET models using the observations of several controlling variables from satellites that can be directly correlated with the land surface ET. These remotely sensed ET models mainly include surface temperature and vegetation index space, surface energy balance, Priestley-Taylor and Penman-Monteith methods (Zhang et al., 2016a). A number of studies have provided evidence that the predictive accuracy of these satellite-based ET mapping methods relies heavily on the retrieval quality of biophysical parameters (e.g., meteorological variables and vegetation indices) obtained from the remote sensing techniques (Glenn et al., 2010; Verstraeten et al., 2008). The other major type to quantify variations in terrestrial ET refers to land surface models that are organized by carefully elucidating the complex hydrological processes of water flux interactions between land and the atmosphere (Polhamus et al., 2013). In general, these well-established models for ET modeling can be further utilized in the application of upscaling ET from site to regional or global level. However, it is not easy to obtain the key plant parameters required by land surface models in relation to both transpiration and evaporation processes, as well as some important driving variables influencing the variation of ET (Abramowitz et al., 2007; Polhamus et al., 2013). Furthermore, some key plant parameters associated with ET may vary temporally and spatially between and within plant functional types due to the heterogeneous environments over the land surface (Brümmer et al., 2012; Williams et al., 2012), which can substantially increase the complexity of characterizing the ongoing process of ET and thus may cause great hindrance to the development of accurate land surface models.

In the past few decades, machine learning techniques have been increasingly utilized to estimate hydrological variables (Abrahart et al., 2012; Keshtegar et al., 2016; Kousari et al., 2017; Moghaddamnia et al., 2009; Park et al., 2016), ecological variables (Crisci et al., 2012) and renewable energy variables (Voyant et al., 2017). According to these applications, the predictive ability of machine learning techniques has been extensively demonstrated, partly due to their powerful advantage of automatically capturing the complex nonlinear structures and characteristics. Especially for the forecasts of reference ET and evaporation, a growing number of studies suggest that machine learning methods can provide better estimates than empirical equations based on different meteorological driving factors (Antonopoulos et al., 2016; Gocić et al., 2015; Kisi et al., 2015; Mehdizadeh et al., 2017; Misaghian et al., 2017; Petković et al., 2016; Tabari et al., 2013; Yassin et al., 2016). More recently, much attention has been paid to the estimation of ET in terrestrial ecosystems using machine learning modeling approaches, referring mainly to artificial neural network (ANN) and support vector machine (SVM). These approaches are able to identify some previously unknown processes dominating the variation of ET at the ecosystem level and thereby can accurately estimate ET (Chen et al., 2014; Dou et al., 2015; Gocić et al., 2015; Shrestha and Shukla, 2015; Yang et al., 2006). In addition, these methods have also been fluently used to solve other different issues as to ET, such as filling the missing data of ET, sensible and latent heat fluxes as well as climatic variables based on the measurements of flux tower sites (Abudu et al., 2009; Chen et al., 2012), exploiting the relative contributions of the environmental variables influencing the mechanisms underlying of ET variation (Huo et al., 2012; Jain et al., 2008), and revising the estimated errors of ET from the land surface models (Abramowitz et al., 2007).

Recently, two state-of-the-art machine learning techniques, namely extreme learning machine (ELM) and adaptive neuro-fuzzy inference system (ANFIS), are widely utilized in the hydrologic time series modeling and forecasting (Alizadeh et al., 2017; Gocic et al., 2016; Hashim et al., 2016; Petković et al., 2015; Pour-Ali Baba et al., 2013; Shamshirband et al., 2016; Shoaib et al., 2014, 2018; Tabari et al., 2013). Especially for ELM, as a relatively novel approach, its ability and strengths over other conventional methods in terms of dealing with nonlinear problems in other fields have been confirmed by many previous studies. For instance, Abdullah et al. (2015) used ELM method to simulate reference ET based on a set of meteorological data, and found it was simple, efficient and effective in practical terms and could be recommended as an alternative tool to traditional methods for estimating reference ET. Yaseen et al. (2016) used ELM, SVM and generalized regression neural network (GRNN) method to imitate monthly stream-flow in a semi-arid region in Iraq, and reported that the ELM performed better than the SVM and GRNN methods in terms of all the performance metrics. Furthermore, prior studies have also proved the usability and potential of ANFIS method for forecasting evaporation and these developed ANFIS models generally provided better estimates than conventional data-driven approaches (Kişi and Tombul, 2013; Malik and Kumar, 2015; Shiri et al., 2011). To the best of our knowledge, however, for ecosystem-level ET modeling and prediction based on the measurements with the eddy covariance technique, no attempt has been made to test the practicability and ability of these two advanced approaches (ELM and ANFIS).

The present study therefore seeks to estimate the ET using four common machine learning approaches based on continuous six-year daily observation data from the flux tower sites across four different terrestrial ecosystems. The specific objectives are the following: (1) to examine the feasibility of both the ELM and ANFIS approaches for modeling the ET at the ecosystem scale; (2) to compare the simulation performance of our newly proposed approaches with traditional ANN and SVM methods; and (3) to investigate the modeling difference among four representative ecosystems (grassland, forest, cropland and wetland).

2. Materials and methods

2.1. Site description

In this study, four flux measurement sites were employed to examine the generalization ability of our proposed models across different ecosystems. We show the details of each site in Table 1. These sites covering four primary ecosystem types included grassland flux observation site from Vaira Ranch in the United States (US-Var),

Table 1
Site characteristics used in this study. The mean annual temperature (MAT, °C) and total annual precipitation (TAP, mm $year^{-1}$) are for the time period (Period). Vegetation types include grassland (GRA), deciduous broadleaf forest (DBF), cropland (CRO) and wetland (WET).

Site	Latitude	Longitude	MAT	TAP	Climate	Vegetation	Period	Reference
US-Var	38.41	−120.95	15.76	533	Mediterranean	GRA	2002–2007	Ma et al. (2007)
DE-Hai	51.08	10.45	8.19	833	Temperate	DBF	2001–2002, 2004–2007	Knohl et al. (2003)
BE-Lon	50.55	4.74	11.33	699	Temperate	CRO	2004–2009	Moureaux et al. (2006)
SE-Deg	64.18	19.55	2.90	436	Boreal	WET	2004–2009	Lund et al. (2010)