Journal of Hydrology 561 (2018) 764–779

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information

Wei Fang[a], Shengzhi Huang[a,*], Qiang Huang[a], Guohe Huang[b], Erhao Meng[a], Jinkai Luan[a]

[a] State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, China
[b] Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Purna Chandra Nayak, Associate Editor

Keywords:
Evapotranspiration
Partial mutual information
Climatic indices
Teleconnection

ABSTRACT

In this study, reference evapotranspiration (ET_0) forecasting models are developed for the least economically developed regions subject to meteorological data scarcity. Firstly, the partial mutual information (PMI) capable of capturing the linear and nonlinear dependence is investigated regarding its utility to identify relevant predictors and exclude those that are redundant through the comparison with partial linear correlation. An efficient input selection technique is crucial for decreasing model data requirements. Then, the interconnection between global climate indices and regional ET_0 is identified. Relevant climatic indices are introduced as additional predictors to comprise information regarding ET_0, which ought to be provided by meteorological data unavailable. The case study in the Jing River and Beiluo River basins, China, reveals that PMI outperforms the partial linear correlation in excluding the redundant information, favouring the yield of smaller predictor sets. The teleconnection analysis identifies the correlation between Nino 1 + 2 and regional ET_0, indicating influences of ENSO events on the evapotranspiration process in the study area. Furthermore, introducing Nino 1 + 2 as predictors helps to yield more accurate ET_0 forecasts. A model performance comparison also shows that nonlinear stochastic models (SVR or RF with input selection through PMI) do not always outperform linear models (MLR with inputs screen by linear correlation). However, the former can offer quite comparable performance depending on smaller predictor sets. Therefore, efforts such as screening model inputs through PMI and incorporating global climatic indices interconnected with ET_0 can benefit the development of ET_0 forecasting models suitable for data-scarce regions.

1. Introduction

Evapotranspiration is a crucial component in the hydrological cycle, simultaneously transferring water from land, oceans and plants to the atmosphere through evaporation and transpiration (Tabari et al., 2013). Estimating the reference evapotranspiration (ET_0) is essential for engineering applications like the irrigation scheduling as well as scientific research like the hydrological modelling. The FAO-56 Penman-Monteith (FAO-PM) equation (Allen et al., 1998) is recommended by the Food and Agriculture Organization (FAO) to be a standard model for estimating ET_0. Benefiting from a solid physical foundation, the FAO-PM equation with related adjustments can be used as a good estimator (Jato-Espino et al., 2016). Its main drawback, however, lies in its relatively high data requirement, which limits its application in many regions, especially in the least economically developed countries, where sufficient meteorological stations and reliable observations are often unavailable (Droogers and Allen, 2002). Therefore, it is of important significance to develop alternative models with lower data burden and computationally suitable for forecasting ET_0 in data-scarce regions.

The aforementioned limitation of the FAO-PM equation has led researchers to turn to numerous empirical models with reduced data requirements. Empirical models mainly include temperature-based (Hargreaves, Blaney-Criddle and Thornthwaite) equations and radiation-based (Priestley-Taylor, Makkink and Jensen-Haise) equations, some of which the FAO-PM equation evolved from. As no universal consensus has been achieved on their global applicability, additional parameter estimation is an indispensable step in applying empirical models to different climatic conditions (Droogers and Allen, 2002; Nandagiri and Kovoor, 2006). The other category of alternative models manages to capture the mapping relationship between selected inputs and ET_0 by means of statistical methods or artificial intelligence approaches covering from multiple linear regression, autoregressive moving average and support vector regression (Jato-Espino et al., 2016; Psilovikos and Elhag, 2013; Tabari et al., 2012; Cheng et al., 2016) to various neural networks and evolutionary algorithms (Falamarzi et al.,

* Corresponding author.
E-mail address: huangshengzhi@xaut.edu.cn (S. Huang).

https://doi.org/10.1016/j.jhydrol.2018.04.038
Received 6 January 2018; Received in revised form 12 April 2018; Accepted 14 April 2018
Available online 19 April 2018
0022-1694/

2014; Shiri et al., 2014; Traore et al., 2016; Fang et al., 2017). For all these models, identifying the optimal input is a fundamental task and is a necessity to reduce the model data requirements. The conventional solution is to test several input combinations comprising only a portion of the meteorological variables available and then derive the optimal input set according to predetermined evaluation criteria (Parasuraman et al., 2007; Partal, 2016; Traore et al., 2016). Though a computationally efficient searching strategy, examining a fraction of all possible combinations instead of an exhaustive search still leaves doubt as to whether there are some combinations with lower data requirements outperforming the ‘optimal’ input set selected. The other strategy for screening model inputs is based on calculating the linear correlation coefficient, which statistically quantifies the linear dependence between each meteorological variable and ET_0 (Jain et al., 2008; Kişi, 2006). Meteorological variables with strong linear correlation with ET_0 are included in the model input set. This strategy, however, is argued to likely select redundant inputs that provide the same amount of information regarding ET_0. Afterward, the partial linear correlation is introduced to further eliminate the redundant information from the input set (Mallikarjuna et al., 2012). On the other hand, evapotranspiration is universally considered a nonlinear process dependent on interacting climatological variables. As a result, the nonlinear dynamics of the evapotranspiration process may not be well captured by only examining the linear correlation.

To this end, entropy and mutual information (MI), two important notions in information theory, are introduced to quantify more general (both linear and nonlinear) dependence. Entropy is known to be a measure of uncertainty for given variables and it is through the notion of entropy that MI is derived (Quilty et al., 2016). MI, also termed transinformation, is defined as the information content of one variable that is also contained by another variable and is formulated as the difference between total entropy of the two random variables and their joint entropy (Ahmadi et al., 2009; Yang et al., 2000). Ahmadi et al. (2009) and Nourani et al. (2015) have applied these two information-content-based criteria (namely, entropy and MI) to input selection for solar radiation estimation and rainfall-runoff modelling, respectively. Evaluating entropy and MI makes it possible for input selection to consider both linear and nonlinear dependence between input candidates and model output. However, as in the case of selecting input through the linear correlation coefficient, there is a disadvantage when using entropy and MI to screen meaningful inputs. This is, an input strongly correlated with the model output might provide redundant information that has been explained by previously selected inputs. To overcome this shortcoming, Sharma (2000) proposed partial mutual information (PMI) for evaluating the additional mutual information attained by adding a potential input to the model input set. In this study, the utility of the partial mutual information to identify relevant predictors for ET_0 is investigated and is compared with that of the partial linear correlation.

The past two decades have witnessed an increasing number of studies on the interconnections between hydrological variables and global climate patterns at multiple timescales. For precipitation, streamflow and groundwater levels, numerous research has identified their delayed response to variability in climatic indices, such as the North Atlantic Oscillation (NAO), Southern Oscillation Index (SOI) and Pacific-North American pattern (PNA) (Cai et al., 2010; Coleman and Budikova, 2013; Tremblay et al., 2011, Huang et al., 2018, Liu et al., 2018). Wang et al. (2006) revealed the strong influence of El Niño–Southern Oscillation (ENSO) events on regional precipitation in the Yellow River Basin, China, which resulted in a 51% decrease in runoff to the sea. Zhang et al. (2007) reported the spatially changing (in-phase or anti-phase) interconnection between ENSO and the annual maximum streamflow from the upper to the lower Yangtze River Basin, China. It was found by Xu et al. (2007) that approximately 20% of 481 gauging stations in China showed a significant correlation between precipitation and SOI, and a more negative correlation than positive was observed. Such interconnections have been exploited by forecast practices involving these hydrological variables successfully (Fan et al., 2015; Schepen et al., 2012; Yang et al., 2017). With respect to ET_0, Meza (2005) found that ET_0 variation in the Maipo River Basin, Chile, was influenced by phases of ENSO, concluding that during the winter and spring, there was up to a 30% difference in ET_0 between the El Niño and La Niña years. Sabziparvar et al. (2011) analysed the ET_0-SOI interconnection at 13 meteorological station sites in Iran. At most of the studied sites, winter and spring ENSO events influenced the ET_0 values of the following summer and autumn. Spatially, more significant impacts of ENSO forcing on ET_0 variability were observed at warm arid sites than at humid sites. Tabari et al. (2014) examined the ET_0-NAO interconnection during winter at 41 Iranian meteorological stations. The results disclosed the negative correlation between winter ET_0 and NOA index, and a negative phase of NAO led to a 3% increase in ET_0 values relative to those during a positive phase. In spite of studies reporting the apparent interconnection between regional ET_0 and global climate patterns, little attention has been paid to incorporating influential climatic indices into ET_0 forecasting practices. Therefore, this study employs global climatic indices as additional potential inputs of forecasting models to analyse their correlation with ET_0 in the study area and investigate their role in yielding a higher forecasting accuracy. The merit lies in that these climatic indices can be easily acquired from related research institutions and do not increase the data collection burden, and they can be universally applied to regions with meteorological data scarcity.

This study aims to (1) investigate the utility of partial mutual information to identify meaningful predictors for ET_0 through a comparison with the partial linear correlation, which merely measures the linear dependence; (2) recognize the interconnection between global climate indices and regional ET_0; and (3) recommend the optimal ET_0 forecasting models having both favourable performance and lower data requirements for regions subject to data scarcity. An appropriate input variable selection (IVS) technique benefits models through effectively decreasing the data requirements. In addition, introducing climatic indices may favour the explanation of variability in ET_0, which ought to be interpreted by the missing meteorological variables. Therefore, the study could have important implications for developing ET_0 forecasting models suitable for the least economically developed countries.

2. Model developments

2.1. An overview of ET_0 forecasting models

The procedure for developing ET_0 forecasting models is organized into four parts, which are depicted in Fig. 1.

2.1.1. Input candidate pools

Scenario 1 is utilized to compare the utility of the partial mutual information and partial linear correlation to screen predictors for ET_0. Under Scenario 1, the input candidate pool comprises all local meteorological variables characterizing variations in air temperature, air pressure, precipitation, humidity, solar radiation and wind speed. It is a prevailing means of composing the input candidate pool and has been used in many previous studies (Chatzithomas and Alexandris, 2015; Kumar et al., 2002; Tabari et al., 2012). Scenario 2 further comprises global climatic indices, in addition to the meteorological variables of Scenario 1, and can provide a comparison with Scenario 1 for investigating the effectiveness of climatic indices in enhancing model performance. Scenario 3 is used for developing ET_0 forecasting models suitable for the least economically developed regions. With consideration of the meteorological data scarcity in many such regions, the input candidate pool under the latter scenario only includes routinely measured meteorological variables (air temperature and sunshine duration), which are available at nearly all meteorological stations. Global climatic indices are further introduced as potential model inputs to