Science of the Total Environment 627 (2018) 78–94

Contents lists available at ScienceDirect

Science of the Total Environment

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

Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation

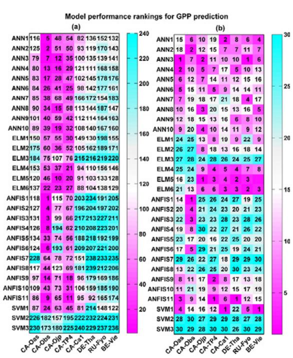
Xianming Dou, Yongguo Yang *

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

HIGHLIGHTS

- Four different data-driven approaches are proposed for estimating daily carbon fluxes.
- Long-term flux tower measurements from eight forest sites in different climates are used.
- The effects of internal parameters on their corresponding methods are assessed together.
- All models with proper functions can precisely simulate and predict carbon fluxes.
- New ELM and ANFIS methods are highly recommended for estimating carbon fluxes.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 November 2017
Received in revised form 18 January 2018
Accepted 20 January 2018
Available online xxxx

Editor: Elena PAOLETTI

Keywords:
Carbon fluxes
Forest ecosystems
Flux towers
Data-driven techniques
Adaptive neuro-fuzzy inference system
Extreme learning machine

ABSTRACT

With the recent availability of large amounts of data from the global flux towers across different terrestrial ecosystems based on the eddy covariance technique, the use of data-driven techniques has been viable. In this study, two advanced techniques, namely adaptive neuro-fuzzy inference system (ANFIS) and extreme learning machine (ELM), were developed and investigated for their viability in estimating daily carbon fluxes at the ecosystem level. All the data used in this study were based upon the long-term chronosequence observations derived from the flux towers in eight forest ecosystems. Both ANFIS and ELM methods were further compared with the most widely used artificial neural network (ANN) and support vector machine (SVM) methods. Moreover, we also focused on probing into the effects of internal parameters on their corresponding approaches. Our estimates showed that most variation in each carbon flux could be effectively explained by the developed models at almost all the sites. Moreover, the forecasting accuracy of each method was strongly dependent upon their respective internal algorithms. The best training function for ANN model can be acquired through the trial and error procedure. The SVM model with the radial basis kernel function performed considerably better than the SVM models with the polynomial and sigmoid kernel functions. The hybrid ELM models achieved similar predictive accuracy for the three fluxes and were consistently superior to the original ELM models with different transfer functions. In most instances, both the subtractive clustering and fuzzy c-means algorithms for the ANFIS models outperformed the most popular grid partitioning algorithm. It was demonstrated that the newly

* 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.scitotenv.2018.01.202
0048-9697/

proposed ELM and ANFIS models were able to produce comparable estimates to the ANN and SVM models for forecasting terrestrial carbon fluxes.

1. Introduction

Carbon cycles in forest ecosystems are strongly driven by the dynamic interactions between biophysical and biogeochemical processes (Delpierre et al., 2012; Hyvönen et al., 2007; Wu et al., 2017). There are several different carbon flux flows of atmospheric trace gases, mainly including gross primary production (GPP), ecosystem respiration (*R*) and net ecosystem exchange (NEE). The ongoing evolution of these carbon fluxes from hourly to decadal time scales (Desai, 2014; Stoy et al., 2005) is affected by global environmental forcing factors, primarily including climate, rising CO_2, nutrient deposition, fires, biotic invasions and different types of human management activities (Frank et al., 2015; Schwartz et al., 2017). Numerous studies have been undertaken to understand the seasonal patterns and environmental control of terrestrial carbon fluxes based on the statistical analysis of micrometeorological measurements from global flux towers using the eddy covariance technique (Hollinger et al., 1999; Jensen et al., 2017; Thum et al., 2008). However, the mechanisms influencing the magnitude and distribution of carbon exchanges between forest ecosystems and the atmosphere remain largely unexplored (Thurner et al., 2016). Therefore, quantifying the carbon fluxes in forest ecosystems is vitally important for offering beneficial suggestions for mitigating future climate change.

Over the past few decades, a number of methods have been proposed and used to estimate the carbon fluxes of terrestrial ecosystems at different spatial patterns from site to regional or global scale. These methods may be classified into four categories: remote sensing-based (e.g., Hilker et al., 2008), process-based (e.g., Huntzinger et al., 2013), atmospheric inversion (e.g., Schneising et al., 2014), and data-driven (e.g., Papale and Valentini, 2003 and Evrendilek, 2013). Among these approaches, process-based models are widely acknowledged as relatively reliable and mature techniques, because these models are able to offer an insight into the complex physical processes that dominate the temporal and spatial variation of carbon fluxes. Nevertheless, recently, several lines of evidence indicate that many uncertainties remain in regard to the estimates of carbon fluxes among different process-based models (Richardson et al., 2012; Schaefer et al., 2012), possibly due to the discrepancies in the structures and/or parameters of applied models (Hilton et al., 2014; Kuppel et al., 2013), the model inputs from different data sources (Keenan et al., 2013), the sparsity or scarcity of estimated sites, and the complexity and heterogeneity of terrestrial ecosystems. To a certain extent, these uncertainties limit our ability to precisely estimate the carbon fluxes of terrestrial ecosystems.

Data-driven techniques have been attracting great interest in different fields, which is largely due to their advantage of adaptively extracting the crucial information contained in a large amount of data without making any assumption about the given application. During the last decade, these techniques have been broadly employed to deal with various nonlinear problems, such as modeling and prediction of ecological and biophysical variables (Dengel et al., 2013; Vahedi, 2016), and forecasting of water resource variables (Maier et al., 2010). More importantly, regarding the estimates of carbon fluxes in terrestrial ecosystems, predictive ability of these techniques has been increasingly recognized in recent years for addressing a variety of issues, primarily including identifying the relative importance of environmental factors that affect the underlying mechanisms of carbon exchanges (Gevrey et al., 2003), elucidating the nonlinear processes of carbon interactions between land and the atmosphere (Evrendilek, 2013), providing a set of helpful tools for upscaling the carbon fluxes from site to regional or global scale (Papale et al., 2015), coupling with different process-based models with the objective of reducing their predictive uncertainties (Wang et al., 2012), and interpolating the missing data of carbon fluxes, energy fluxes as well as climatic variables based on the measurements from global flux towers (Menzer et al., 2015). It should be noted that most of the studies above-mentioned were undertaken almost exclusively through two traditional data-driven approaches, namely artificial neural network (ANN) and support vector machine (SVM).

It is well-known that both ANN and SVM methods have received a great deal of attention in the last decade and have been extensively utilized in diverse fields (Abrahart et al., 2012; Mountrakis et al., 2011). Nevertheless, these two approaches still have some shortcomings, which have been revealed by previous studies Maier et al. (2010) and Raghavendra. Raghavendra and Deka (2014), respectively. In general, the ability of ANN method is limited by several disadvantages, such as slow learning speed, over-fitting and local minima. Additionally, it is also relatively difficult to determine some key parameters, such as training function and activation function. SVM also exists several drawbacks, such as high memory requirement and a large amount of computing time during learning process. In order to overcome the disadvantages of these two approaches, many new modeling techniques have been proposed in recent years. It is impractical to consider all approaches in this work. Therefore, extreme learning machine (ELM) and adaptive neuro-fuzzy inference system (ANFIS) approaches were exclusively taken into account due to their respective strengths. The former is characterized by apparent simplicity of model design, fast learning and convergence speed (Huang et al., 2006), and the latter takes advantage of the strengths of both ANN and fuzzy inference methods (Jang, 1993). To date, the potential of these two advanced approaches for solving forecasting and prediction issues has been demonstrated by numerous recent studies in other different fields, such as water resource management (Kişi et al., 2011; Kisi et al., 2012; Petković et al., 2016; Shiri et al., 2011; Shiri et al., 2016; Yaseen et al., 2016), sustainable energy engineering (Hosseini Nazhad et al., 2017; Mohammadi et al., 2015; Shamshirband et al., 2015), thermal engineering (Jović et al., 2016; Kariminia et al., 2016a; Kariminia et al., 2016b; Kariminia et al., 2016c; Shamshirband et al., 2016) and environmental science (Kisi et al., 2013a; Kisi and Shiri, 2014; Mohammadi et al., 2016; Raos et al., 2016). To the best of our knowledge, however, no study has been reported that probes into the validity of these two innovative approaches for estimating terrestrial carbon fluxes.

Especially noteworthy is the fact that all the methods above-mentioned are significantly affected by their respective inner parameters or functions according to a growing number of studies (Citakoglu, 2015; Maier and Dandy, 1998; Tabari et al., 2013). For instance, Zounemat-Kermani et al. (2016) compared and evaluated the capability of both ANN with three different training algorithms, including Broyden–Fletcher–Goldfarb–Shanno (BFGS), gradient descent and conjugate gradient, and SVM with four different kernel functions, including radial basis function (RBF), linear, polynomial, and sigmoid, for predicting daily suspended sediment concentrations. They found that the BFGS and RBF algorithms respectively provided the most accurate estimates and should be recommended as helpful options for hydrologic time series forecasting. Cobaner (2011) used two types of algorithms (grid partition and subtractive clustering) to generate different fuzzy inference systems (FISs) for establishing ANFIS models for modeling reference evapotranspiration and found that subtractive clustering-based ANFIS model required fewer computations and obtained higher accuracy in comparison with subtractive clustering-based ANFIS model.