Forest Ecology and Management 432 (2019) 121–131

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Analyzing urban forest coverage variation in Guangzhou-Foshan region using factorial analysis based multivariate statistical prediction models

J. Wang[a,b], Y.P. Li[a,c,*], J. Sun[a,b], Y.T. Lin[a,b]

[a] School of Environment, Beijing Normal University, Beijing 100875, China
[b] Sina-Canada Energy and Environmental Research Center, North China Electric Power University, Beijing 102206, China
[c] Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Sask S4S 7H9, Canada

ARTICLE INFO

Keywords:
Factorial analysis
Forest coverage area
Human activity
Interaction
Multivariate statistical prediction

ABSTRACT

In this study, factorial analysis based multivariate statistical prediction (FAMSP) models are developed to analyze the variation of urban forest coverage area (FCA). Through incorporating techniques of multivariate linear regression (MLR), multivariate quantile regression (MQR), stepwise cluster analysis (SCA), and support vector machine (SVM) within factorial analysis framework, four FAMSP models are advanced. The developed models have advantages in reflecting the complex relationships (e.g., linear/nonlinear and/or continuous/discrete) among urban FCA, human activity, and natural factors. Factorial analysis is used for exploring the interactions among multiple factors on FCA variation. The FAMSP models are also applied to Guangzhou-Foshan region for illustrating their applicabilities in FCA variation analysis. Results reveal that different multivariate statistical prediction methods lead to different performances for FCA variation. SCA and SVM can get more satisfactory performances than MLR and MQR due to their superior ability in characterizing the nonlinear features of FCA variation. Population is one of the key drivers for FCA variation due to its high sensitivity to timber consumption and stock; population would affect the regional climatological condition (e.g., precipitation), which consequently alters forest growth. The factors of Guangzhou would primarily impact regional FCA variation due to its higher population and higher timber demand than those in Foshan. These findings are helpful for the urban forest sustainable development and timber resources management.

1. Introduction

Forests, as important terrestrial ecosystems, are vital habitats for biodiversity and they are also critical for the provision of important ecosystem services to human well-being, such as regulating climate, providing food and fiber, and maintaining water availability and quality (Masek et al., 2015; Brockerhoff et al., 2017; Correia et al., 2018). However, in recent years, due to the rapid growth of population, the speedy development of agriculture and urban construction, as well as improper forest management, the degradation of forest resources has been accelerating (Li, 2004; Omrani et al., 2017; Nazir and Ahmad, 2018). For example, the global forest area has decreased by 1.3 million km^2 since 1990s; especially for China, 9.3 million ha of natural forests are destroyed annually, causing the loss of more than 10 million m^3 of timber. As a result, a series of ecological and environmental problems (e.g., soil erosion, biodiversity loss, desertification, flooding and disease) have occurred with increasing frequency, further threatening human health (Caputo et al., 2016; Pastori et al., 2017; Zulueta et al., 2017). Such degradation affects the structures and functions of forest ecosystem, and brings about great challenges to the balance between economic development and ecological sustainability. Understating the drivers of forest variation and the impacts of human activities and natural processes is critical for forest ecosystem management.

Forest coverage area (FCA), a commonly used indicator to evaluate terrestrial environmental conditions, is closely related to anthropogenic activities, livestock breeding, deforestation monitoring, and climate change (Purevdory et al., 1998; Wang et al., 2015). A number of models were developed for analyzing FCA, including physically based models and statistical models (Zhou et al., 2011; Seidl et al., 2012; Potapov et al., 2015). Physically based models have complex structures and need many inaccessible types of inputs, making modeling process costly and time-consuming (Augustin et al., 2009; Nishii and Tanaka, 2013; Kumar et al., 2014; Li et al., 2017). Previously, some statistical approaches with small data requirement were widely used for ecosystem simulation, such as artificial neural network (ANN), support vector machine (SVM), Mann-Kendall test, quantile regression, and stepwise-cluster

* Corresponding author at: School of Environment, Beijing Normal University, Beijing 100875, China.
E-mail address: yongping.li@iseis.org (Y.P. Li).

https://doi.org/10.1016/j.foreco.2018.09.010
Received 7 July 2018; Received in revised form 21 August 2018; Accepted 9 September 2018

analysis (SCA). In detail, Nishii and Tanaka (2015) used SVM method for predicting forest area in Hiroshima with indicators of human population density and relief energy; results indicated that human population density is the primary indicator influencing forest area. Zheng et al. (2017) proposed a vegetation dynamics SCA model to establish the relationship between Normalized Difference Vegetation Index (NDVI) and climate condition, revealing that precipitation is the main factor that could lead to the change of NDVI values. Reis et al. (2018) employed ANN method for estimating the survival and mortality of individual trees in Amazon rain forest; results showed that the ANN method could accurately estimate the individual survival of trees. Most of the approaches could effectively reflect the relationship between forest and the associated drivers. However, it is still not clear which method possesses the best performance, particularly for more and more complex urban ecosystems. Besides, the previous studies mainly emphasized on one single city or region, neglecting to explore the interactive effects of urban agglomeration containing multiple cities.

Due to the extensive communications (e.g., flows of material, energy and information) among different cities, urban agglomeration ecosystem is plagued with interactions and complexities. For example, due to the difference in timber demand and supply, adjacent cities could import/export timber resources from each other. Consequently, forest coverage can be influenced by adjacent city, leading to more complex interactions among cities. Moreover, forest ecosystems are characterized by diverse components (e.g., social factors and natural processes). These components are interacted with each other, leading to ecosystem being nonlinear, and spatiotemporally heterogeneous (Wu and Marceau, 2002; Sun et al., 2018). Consequently, urban FCA variation is a complicated process that is impacted by individual and interactive effects of multiple factors. When different factor combinations give rise to equally good predictions, the unique optimal factor set cannot be determined, leading to increased parameter uncertainty and amplified prediction error (Mortier et al., 2013; Liu et al., 2017). Factorial analysis (FA), as a multivariate inference method, is based on the theory of design of experiment (DOE) and is used for quantifying the individual and interactive effects of parameters on model response. FA can facilitate the exploration of (i) the main effects by measuring the variation ranges of forest coverage area under the impact of individual parameter, and (ii) parameter interactions via disclosing each parameter's effect under the impact of other parameters (Zhang et al., 2016). FA has been widely used in environmental systems management, such as water resources management, energy system planning, and hydrological simulation (Wang et al., 2015; Zhang et al., 2016; Liu et al., 2017). Unfortunately, few studies have been reported in the field of forest ecosystem.

Therefore, this study is to develop factorial analysis based multivariate statistical prediction (FAMSP) models for managing forest ecosystem of urban agglomeration, through introducing multivariate linear regression (MLR), multivariate quantile regression (MQR), stepwise cluster analysis (SCA), and support vector machine (SVM) into factorial analysis framework. FAMSP models have advantages in facilitating the exploration of (i) the potential relationship (e.g., linear/nonlinear and/or continuous/discrete) between FCA and associated drivers, (ii) the individual and interactive effects of human activity and natural factors, and (iii) identifying interactions between adjacent cities. Then, the developed FAMSP models are applied to analyzing the variation of urban forest coverage area (FCA) in Guangzhou-Foshan region. Results obtained are expected to provide more valuable information for disclosing the variation of urban FCA as well as more effectively managing timber resources in the Pearl River Delta urban agglomeration.

2. Study area

Guangzhou (112°27′-114°03′E, 22°26′-23°56N) and Foshan (112°28′-114°24′E, 22°38′-23°27N) are located in the Pearl River Delta, the south-central part of Guangdong province (as shown in Fig. 1). In 2017, Guangzhou covered 7434.4 km^2 with population of 13.5 million. Its annual precipitation reaches 1736 mm and average temperature is about 22.1 °C. The forest coverage area occupies 42% of the total area of Guangzhou's districts (Jing et al., 2018). Foshan, with the administrative area of 3797.7 km^2, owned population of 7.4 million in 2017. Its annual precipitation is about 1521 mm and average temperature is about 23.2 °C. The forest coverage ratio is only 22% in 2017 (Pan et al., 2017).

Due to rapid population growth and economic development, Guangzhou and Foshan are both confronted with tremendous pressure to satisfy their increasing timber resources demands. On one hand, speedy urbanization and industrialization has led to the decline of forest coverage. The forest coverage in Guangzhou was 3122 km^2 in 2004, more abundant than that in 2008 (i.e. 2362 km^2). On the other hand, the local governments have paid much attention to protecting forest resources for building green-forest cities in recent years, which further brings about conflicts between timber resources demand and supply. As a result, to satisfy the demand for timber resources, Guangzhou mainly relies on timber import. For example, in 2015, about 1.05 million tons of timbers were imported to Guangzhou. The City of Foshan owns massive furniture factories. Bordering on Foshan, Guangzhou is largely interacted with Foshan in terms of furniture markets, causing complex communications of timber resources between the two cities. For sustainable development of forest ecosystem and effective management of timber resources, it is imperative to understand how the forest ecosystem is changed and what the main drivers of forest coverage variation are.

3. Framework of FAMSP models

In this study, factorial analysis based multivariate statistical prediction (FAMSP) models are developed for analyzing the variation of urban forest coverage area (FCA). The framework of FAMSP models is displayed in Fig. 2, where four statistical methods (MLR, MQR, SCA and SVM) are incorporated within factorial analysis framework. The developed models have advantages in: (1) effectively incorporating human activity and natural factors into FCA variation analysis, (2) simply presenting the relationships between predictors and predictands, (3) requiring less data than physical models, and (4) identifying the interactions of two adjacent cities as well as the interactions among multiple factors.

3.1. Multivariate linear regression

Multivariate linear regression (MLR) attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. MLR is a classic technique providing several advantages: simplicity, interpretability and possibility of being adjusted over the transformations of the variables (Min and Lee, 2005). Models are characterized by an equation taking the form of

$$Y = \beta_0 + \beta_1 X_1 + \cdots \beta_p X_p \quad (1)$$

where Y is the predicted or the expected value of dependent variable, X_1 through X_p are predictor variables, β_0 is the value of Y when all of the independent variables (X_1 through X_p) are equal zero, β_1 through β_p are the estimated regression coefficients. Each regression coefficient represents the change in Y relative to a one unit change in the respective variable.

3.2. Multivariate quantile regression

Multivariate quantile regression (MQR) supplements multiple regressions for the mean by supplying information about the relationship between the response and the covariates at the tails of the response