Journal of Cleaner Production xxx (2016) 1–11

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Evaluation on China's forestry resources efficiency based on big data

Lei Li [*], Tongtong Hao, Ting Chi

School of Management and Economics, Tianjin University, Tianjin, China

ARTICLE INFO

Article history:
Received 16 June 2015
Received in revised form
11 February 2016
Accepted 16 February 2016
Available online xxx

Keywords:
Big data
China's forestry resources
DEA
Malmquist total factor productivity index

ABSTRACT

The development of China's forestry resources has never been more challenging due to serious problems such as shortage, inferiority and uneven distribution of forestry resources. Therefore, the study and evaluation of China's forestry resources has a great significance to improve efficiency and ensure the sustainable development of the forestry resource. Meanwhile, the vast territory, huge population and widespread forest landscape of China have led to the numerous indexes and the huge data. To perform the research and evaluation accurately, this paper utilized the big data theory to analyze the relevant data of China's forestry resources. This study collected the data from 31 inland provinces and municipalities of China from 2005 to 2013, after which we carefully examined economic, social and ecological factors to choose assessment indexes and processed data accordingly. Firstly, we performed a cross-sectional dataset analysis using the method of data envelopment analysis to investigate the forestry resources efficiency in 31 inland provinces and municipalities of China in years 2008, 2012 and 2013. Secondly, we analyzed time series data of the 31 inland provinces and municipalities from 2005 to 2013 using the Malmquist total factor productivity index method. Our results showed the dominant factor that restraining forestry resources efficiency for the 31 inland provinces and municipalities is the implemented technology. So we suggest increasing the investment in science and technology to improve the overall efficiency of forestry resources, along with improvement of operation and management by relevant administrative departments to improve technology utilization. The innovation of this paper lies in the dynamic process of analysis.

1. Introduction

The economic reform and opening policy and entering the WTO have resulted in China's rapid economic development, social progress and the population growing (Song et al., 2013). However, this has resulted in enormous consumption of natural resources and will bring greater pressure on them in the future. Meanwhile, the ecological environment is severely deteriorated (Cherniwchan, 2012). The whole world is facing many ecological problems, such as the decreasing forest areas, the increasing desertification, the constantly emerging crisis of water resources and the extinguished species. Therefore, it is important to increase the efficiency of all the natural resources and to keep healthy and sustainable development of the ecosystems. Zhelev (2005) addressed the opportunities to improve profit and reduce the investment cost by studying the integrated management of resources such as electrical energy, steam and fresh water. Napoles-Rivers et al. (2013) proposed a mathematical programming model for sustainable water management in macroscopic systems and found that implementing strategies that include alternative water sources might reduce fresh sources consumption and waste generation and still be economically attractive. Gonzalez et al. (2014) evaluated the environmental profile of the production of maritime pine biomass in France under two different management scenarios. Giuntoli et al. (2015) used the dataset from the latest European Commission document to study the domestic heating from forest logging residues. Ma et al. (2015) presented a set of tools to help decision makers better solve the resource and ecological efficiency problems and to guide resource-based chemical industries in sustainable development. Li et al. (2013) used the improved super-efficiency data envelopment analysis to promote the improvement of production efficiency in energy companies and the enhancement of China's energy supply.

Among all the researches, only a few have considered the analysis of China's forestry resources. It is generally believed that the forests can sequester carbon dioxide and mitigate climate change. However, it also needs the efficient management and

* Corresponding author.
E-mail address: lilei@tju.edu.cn (L. Li).

http://dx.doi.org/10.1016/j.jclepro.2016.02.078
0959-6526/

operation (Schulze and Schulze, 2010). At the same time, China's forestry resources are faced with serious problems, such as the shortage, low quality and uneven distribution. As a result, the development of China's forestry resources has huge challenge. According to the statistics of State Forestry Administration, China's current forest coverage rate is lagged far behind the world average level of 31 percent and China's per-capita forest area is only one-fourth of the world average. At the APEC Economic Leader's Meeting in 2009, China promised to the international community that it will realize a double-increase of forest area and storage. Therefore, China's forestry resources are faced with unprecedented difficulties pressure and how to evaluate and improve the efficiency is particularly important.

However, China has a vast land area and widespread forest, and the forestry indexes are numerous and the data is huge. It leads the data-collection, data storage and data analysis very difficult. To do the research and evaluation accurately, this paper brings in the big data theory and analyzes annual statistical data from 2005 to 2013 of China's forestry resources based on the big data.

"Big data" was first emerged in 1980s. In 2011 the concept of "big data" was officially launched on the EMC World titled of "Cloud Meets Big Data". And then, McKinsey Global Institute delimited the big data "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011). After this, big data has become the hot issue in the computer industry and the financial business, and it also draws attention of the governments. Obama unveils the "Big Data Research and Development Initiative" in 2012 (Weiss and Zgorski, 2012), and the big data strategy competition is opened. Big data is generally considered as volume, velocity, variety and value.

Although big data has caused closed attention and heated discussion among the governments, business industry and scientific circles, it is scarcely used in academia. It is because that the data processing needs a significant amount of storage and relevant equipment and software. So this paper draws support from the big data and uses DEA and Malmquist total factor productivity index to do a cross-sectional dataset analysis and time series analysis to evaluate and improve China's forestry resources efficiency.

2. Previous literature review

As mentioned above, there has been an increasing interest in big data. However, it still has not reached to a common definition. Garlasu (2013) thought that big data is a kind of data which has three characteristics: a great volume of data; can't be structured into regular database; produced with great velocity. Gandomi and Haider (2015) integrated definitions from practitioners and academics and offered a broader definition of big data that captures its other unique and defining characteristic. Baru et al. (2013) imposed that big data has created a new set of challenges and opportunities due to its volume, velocity, variety and value. What's more, many researches are concerned about the big data security. For example, Lu et al. (2013) think that the Information and Telecommunication supply chain faces many security problems. Kim et al. (2013) proposed attribute selection methodology so as to protect the value of big data and guarantee the security. At present, the research about big data is focused on the thorough big data state of art (Camargo-Vega et al., 2015).

As for forestry resources, the present researches are concentrated on the reform of management system (Emi et al., 2005), the legislation of real right system (Hyde et al., 1991) and the property rights institution (Adhikari, 2001). Some scholars studied the forestry efficiency: Chen et al. (2015) use input–output model to examine the current state of forest resource utilization in China. Mattila et al. (2011) investigated the effects of sustainability of the entire forestry industry considering environmental factor. Alexander and Palmer (1999) presented an overview of the indicators and methods developed for the Forest Health Monitoring program, and also reviewed some results after four years of monitoring and research. Viitala and Hanninen (1998) investigated the efficiency of 19 state-funded regional Forestry Boards and found that the input of the Forestry Boards can save in the range of 20%. Lee (2005) measured the relative efficiency of 97 global forest and paper companies in 2001, and found that most companies were operating in regions of decreasing returns to scale and relatively inefficient companies could adjust their variable indexes to achieve the efficient production frontier.

Despite the researches on China's forestry resources efficiency before have measured the efficiency situation effectively, it still has many shortcomings. Firstly, the study subjects are most centralized on a certain state or on a forestry authority. However, this kind of research cannot evaluate China's whole forestry resources efficiency because of the outstanding areal variation in China. Next, as for the research method, the indexes are numerous and the data is huge, which results in the long and costly qualitative analysis that is not feasible any longer, nor the grey model can solve the multiple inputs–outputs problem. Although some specialists have put data envelopment analysis into the evaluation of forestry resources, this method can only study one particular year and cannot study the past years vertically. To overcome the defects above, this study firstly uses DEA to evaluate the relatively efficiency of China's 31 inland provinces and municipalities in the same year. Then, it analyzes the efficiency varying of the 31 inland provinces and municipalities from 2005 to 2013 using the Malmquist total factor productivity index method.

3. Methodology

Because of the numerous indexes and the huge data we firstly perform big data methodology. We collect and analyze relevant data and provide the valuable information by means of data analysis, such as data sampling, data scrubbing and so on. Then on the basis of big data methodology, we adapt the DEA model and Malmquist total factor productivity index.

3.1. The DEA model

Data envelopment analysis (DEA) was put forward in 1978 by Charnes et al. (1978) and it is a non-parametric method. DEA uses linear programming models to give weights to the inputs and outputs and calculate the relative efficiency of every decision making units (DMU) (Cooper et al., 2004). The biggest advantage of DEA is that it doesn't have to determine the functional relationships between the inputs and the outputs before. The DEA model avoids the dependent on specific mathematical model and doesn't need the non-dimensional quantities of original data. So it deserves widespread application and rapid development and is especially applied to a model with multiple outputs and multiple inputs. The most fundamental models of DEA are the CCR model and the BCC model.

1. The CCR model
   The CCR model is put forward by Charnes et al. (1978), and is also called Constant Return Scale model (CRS). As the most fundamental DEA model, the CCR model is based on the assumption that the return to scale is constant. The CCR model assumes that there are n decision making units and each decision making unit has m kinds of inputs and s kinds of outputs. x_{ij} is the ith kind input of jth kind decision making unit, and y_{rj} is the rth kind output of the jth decision making unit. λ_j is the