CrossMark

Environmental performance evaluation with big data: theories and methods

Ma-Lin Song[1] · Ron Fisher[2] · Jian-Lin Wang[3] · Lian-Biao Cui[1]

Abstract Traditional theories and methods for comprehensive environmental performance evaluation are challenged by the appearance of big data because of its large quantity, high velocity, and high diversity, even though big data is defective in accuracy and stability. In this paper, we first review the literature on environmental performance evaluation, including evaluation theories, the methods of data envelopment analysis, and the technologies and applications of life cycle assessment and the ecological footprint. Then, we present the theories and technologies regarding big data and the opportunities and applications for these in related areas, followed by a discussion on problems and challenges. The latest advances in environmental management based on big data technologies are summarized. Finally, conclusions are put forward that the feasibility, reliability, and stability of existing theories and methodologies should be thoroughly validated before they can be successfully applied to evaluate environmental performance in practice and provide scientific basis and guidance to formulate environmental protection policies.

Keywords Big data · Environmental management · Environmental performance · Data envelopment analysis · Life cycle assessment

1 Introduction

Along with the acceleration of modernization and the huge consumption of energy resources, the emission of hazardous substances is becoming increasingly serious, and the ecological

✉ Ma-Lin Song
  songmartin@163.com

1 Collaborative Innovation Center for Ecological Economics and Management,
  Anhui University of Finance and Economics, Bengbu 233030, Anhui, China

2 Department of International Business and Asian Studies, Griffith Business School,
  Griffith University, Gold Coast Campus, Parklands Drive, Southport, QLD 4222, Australia

3 Center for Industrial and Business Organization, Dongbei University of Finance and Economics,
  Dalian 116025, Liaoning, China

🍁 Springer

environment, in which human beings survive, is facing unprecedented threats (Chevallier and Goutte 2015; Cai et al. 2015). Environmental problems have become important factors that hinder the sustainable development of the economy and society (Bi et al. 2012; Song et al. 2013). In November 2013, after lengthy discussions and negotiations between relevant governmental departments of various countries in Warsaw, Poland, agreements were concluded on some issues including the Durban Platform, green climate funds, reduction of greenhouse gases, and the Warsaw International Mechanism for Loss and Damage. "These agreements were for reducing the losses resulting from environmental changes".[1] The citizens of each country also now pay more attention to environmental problems and exert an increasingly important influence on environmental management decision-making (Glucker et al. 2013; Paco and Raposo 2009). Environmental performance refers to production performance that considers environmental factors. International scholars have reached a consensus that sustainability during the production process should be measured by adopting environmental performance evaluation (Halkos and Tzeremes 2013; Jawahar et al. 2015).

As an effective tool to calculate relative efficiency, data envelopment analysis (DEA), first proposed by Charnes et al. (1978), has attracted significant attention from many scholars and has since been expanded continuously (Cook and Seiford 2009; Ramli et al. 2013). DEA includes the super-efficiency model (Andersen and Petersen 1993) and the cross-efficiency model (Liang et al. 2008; Wu et al. 2016), which can improve efficiency discernment. Recent important achievements in this field include the multiple variable proportionality model (Cook and Zhu 2011), weight restrictions and free production model (Podinovski and Bouzdine-Chameeva 2013), fuzzy efficiency measurement (Kao and Lin 2012), and non-homogeneous decision-making units (DMUs) (Cook et al. 2013). DEA has also been widely used in efficiency evaluations with a consideration of undesirable outputs (Färe et al. 1989). It has gradually become one of the key, widely recognized environmental performance evaluation methods (Song et al. 2012). However, although there exist some studies in this field—such as the SBM model that has network structure (Tone and Tsutsui 2014; Lozano 2015), the environmental efficiency evaluation model with small data (Song and Guan 2014; Arabi et al. 2016), and green supply chain management based on ecological and environmental efficiency (Govindan et al. 2014; Dubey et al. 2015a)—shortcomings such as the specialization of research objects and the weak universality of research methods remain. In addition, as another kind of representative method for environmental performance evaluation, life cycle assessment (LCA) also has the same problem (Reap et al. 2008). Previous studies by Wilson et al. (2013), Fadeyi et al. (2013), Hjaila et al. (2013) and Lozano et al. (2010) had similar drawbacks.

As these environmental performance evaluation technologies have insufficient universality, it is difficult to identify the most appropriate analytic method. Even though some studies have used similar evaluation methods to analyze comparable realistic problems, their theoretical cores may vary greatly. Mohammadi et al. (2013) and Vázquez-Rowe et al. (2012) combined the LCA and DEA methods and evaluated the environmental performance of the grape and soybean production industries, respectively. Oggioni et al. (2011) and Zhou et al. (2014) adopted the DEA method to evaluate the energy efficiency of China's transportation sector and the ecological efficiency of the global cement industry. Our detailed investigation revealed that most studies mainly focused on the specific methods of environmental performance evaluation and their specific application domains, without arriving at conclusions on how to measure environmental performance precisely, and are yet to devise a scientific,

---

[1] The UN Climate Change Conference was held on October 23, 2013 in Warsaw, the capital of Poland. Agreements on some important subjects were reached.

specific, and strongly operable theoretical and methodology system (Liu et al. 2010; Cabeza et al. 2014). In other words, the existing environmental performance evaluation lacks an axiomatized theoretical system. Hence, developing a scientific axiomatized theoretical system and a series of universal evaluation methods based on that system for the environmental performance evaluation field is a critical requirement.

Scholars have been exploring and promoting solutions to these unresolved problems (Nahorski and Ravn 2000; Chen and Delmas 2012). The rapid development and wide application of big data have brought new opportunities and challenges to environmental performance evaluation. According to a HACE theorem, big data originates from the distribution and decentralized control of a large volume of heterogeneous and autonomous data. It requires complex and evolving relationships between data (Wu et al. 2014). According to current estimates, from underground physics experiments to retail transactions, security cameras, and GPS systems, about 4 zettabytes of data will be generated each year (Tien 2013). Big data has already permeated every industry and business function field and has become a new production factor that is parallel to labor force and capital. It would drive a new wave of productivity growth and consumer surplus (Manyika et al. 2011). Some developed countries have already constructed their national big data strategy. For example, in 2012, the National Science Foundation (NSF) of America collected key technologies and processes that pushed big data science and projects (BIGDATA)[2] forward. The NSF has also invested a large amount of capital for big data research in five important industries: services, manufacturing, construction, agriculture, and mining. Other countries, including China, have also increased inputs in big data research (Wu et al. 2014). From this, many studies on specific industries, such as Dubey et al. (2015b, c) on a world-class sustainable manufacturing industry, have emerged.

In the environmental management field, a huge amount of high-value information needs to be globally distributed to solve major scientific and social problems (Hampton et al. 2013). By using the big data collected, the US Environment Protection Agency (EPA) and US Energy Information Administration (EIA) have set up the Emissions & Generation Resource Integrated Database (eGRID), which provides almost all carbon emission data resulting from power generation in the US.[3] However, there are few studies on how to establish the methodological system of environmental performance evaluation using big data. Though Cooper et al. (2013) have stated that big data in the context of environmental management has been found, examined, sampled, and applied in LCA, considering both direct and indirect sample data in the open LCA data memory pool, evaluation process proof and statistical tests based on LCA are not available. In fact, big data has the characteristics of volume, velocity, variety, veracity, and valorization (5Vs), which significantly increase the complexity of solving relevant problems (Özdemir et al. 2013). The arrival of the big data era brings unavoidable demands and complex challenges to the imperfect environmental performance evaluation theory and method. This indicates that the preliminary research on environmental performance evaluation with big data has not only high scientific value but also great practical significance.

The remainder of this article is structured as follows. First, we review the literature on environmental performance evaluation, including evaluation theories and relevant methods and applications of DEA, LCA, and the ecological footprint. Second, we describe and comment on the theories and methods of big data and their applications in some fields, as well as

---

[2] NSF has released the plan "Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA)".

[3] These data include emission loads and emission ratios of oxynitride, sulfur dioxide, carbon dioxide, methane, nitrous oxide, net-generated energy, etc.

analyze the associated challenges. Third, we introduce the research progress in the context of environmental management and its big data association established by international scholars. Finally, we summarize the existing achievements, as well as examine the scientific problems that require further study.

2 Theories and methods of environmental performance evaluation and their applications

As the influence of economic activities on the environment has attracted widespread attention, international scholars have put forward many theories and methods to monitor and evaluate environmental performance more effectively (Coelli et al. 2007). A majority of scholars considered the definition of environmental performance to be the economic value borne by a unit environment load (DeSimone and Popoff 2000; Koskela and Vehmas 2012; Liu et al. 2010). A feature of early studies was the attempt to integrate technology, the economy, and environmental performance measurement technology (Scheel 2001; Tyteca 1996). Existing methods of environmental performance evaluation mainly involve the measurement of environmental performance and evaluation of the conditions of material balance (Coelli et al. 2007). A few other evaluation methods, including strategic environmental assessment (SEA) (Zhu and Ru 2008) that evaluates the influences of planning, policies, and schemes on the environment; the ecological footprint (Bagliani et al. 2008) that observes the influences of human consumption on the environment; cost-benefit analysis (Mouter et al. 2013) that focuses on the relationship between costs and benefits of social activities; and material flow analysis (Mouter et al. 2013) that describes the metabolism of social materials (Hashimoto and Moriguchi 2004), are all supplementary to the mainstream DEA and LCA methods.

As a type of nonparametric method, DEA is one of the best methods to measure environmental performance (Bogetoft and Wang 2005). It has the following advantages: it can deal with complex multi-input and output systems and analyze indicators with prices that are difficult to determine and for which weights cannot be decided; it needs no preliminary assumption of relational expressions of the production function, and hence the parametric estimation problem can be avoided; it is useful in that it reveals the hidden and ignored relationships in other methods; and it can quantitatively analyze the root causes of the low efficiency of some DMUs (Liu et al. 2010; Lv et al. 2013).

The core of DEA-based environmental performance evaluation is identifying methods to handle undesirable outputs such as exhaust gas, wastewater, and waste residues generated during production processes. The relevant technologies can be divided into four kinds. The first kind of technology, approved by mainstream scholars, replaces the strong free treatment of undesirable outputs with weak free treatment (Färe et al. 1989, 1993, 2005; Seiford and Zhu 2005; Tone 2003; Zhou et al. 2008, 2007). The second kind of technology takes undesirable outputs as inputs (Dyckhoff and Allen 2001; Hailu and Veeman 2001; Liu and Sharp 1999), and one only needs to determine which indicators are expected to be bigger or smaller. This method is simple and operable, but it cannot reflect real production processes (Seiford and Zhu 2002). The third kind of technology includes a nonlinear monotone decreasing transfer approach (Scheel 2001; Tyteca 1996) and a linear monotone transfer approach (Seiford and Zhu 2002). The former approach uses the reciprocal of an undesirable output as a new output, while the latter adds a sufficiently large positive number to the negative undesirable output to handle this output. The fourth kind of technology is the scale model proposed by You and Yan (2011). This method introduces penalty factors to replace undesirable output

values, and the output of the new system will be the quotient of the original desirable output divided by the penalty factor.

Apart from DEA, LCA is also widely used in the field of environmental performance evaluation (Blengini et al. 2012; Mestre and Vogtlander 2013; Slagstad and Brattebø 2014). In 1969, when Harry E. Teasley Jr. was assigned to manage the packaging of Coca-Cola Company products, he suggested using LCA to evaluate the influences of the life cycle on the environment (Hunt et al. 1996). Currently, the four stages included in this method, namely, the objective and scope, the Life Cycle Inventory (LCI) Analysis, the Life Cycle Impact Assessment (LCIA), and result interpretation, are all included in the ISO 14000 Environmental Management System (ISO 1997, 2006). Relevant guidance relating to the method has already been provided by some studies (Guinée et al. 2002). Because it can effectively resolve the influences of the complexities of the three dimensions of society, environment, and the economy in a sustainable development evaluation system of performance evaluation (Finnveden et al. 2009), as well as consider the diversity of influences of production on the environment (Hauschild and Pennington 2002) and estimate its potential influences (Tiruta-Barna et al. 2007), LCA was acknowledged and applied in industries such as wind energy (Schleisner 2000), waste disposal (Cherubini et al. 2009), and biology (Pérez-López et al. 2014).

However, the scheme selection problem involved in LCA has the challenge of uncertainty and therefore affects the evaluation results (Finnveden et al. 2009). Moreover, it does not fully consider the economic benefits of the production unit (Dong et al. 2014). Therefore, some scholars tried to combine the LCA method with others to avoid these problems. In particular, a combination of LCA and DEA is likely to become widely accepted (Mohammadi et al. 2013; Vázquez-Rowe et al. 2012), as this combination can be used to calculate the composite environmental performance of multiple DMUs (Iribarren et al. 2010) and is more accurate. However, the discriminating capability of this combination remains unsatisfactory (Iribarren et al. 2013), and it cannot effectively settle the problems that emerge during environmental performance evaluation under the condition of big data (Stamp et al. 2013). Hence, evaluation methods based on a combination of DEA and LCA are still in need of further improvement.

3 Fundamental principles of big data and their challenges and breakthroughs

Recently, the quantity of information generated by enterprises, governments, and academic circles has been increasingly rapidly because of science and technology developments. It is estimated that the quantity of data will reach 40 ZB globally in 2020, exceeding the original estimation of 35 ZB (Tien 2013). Moreover, the data in China will reach 8.6 ZB.[4] In May 2011, the McKinsey Global Institute published a research report that analyzed the development prospects of big data in the fields of innovation, competition, and the productivity frontier, among others (Manyika et al. 2011). In May 2012, the United Nations Global Pulse published research illustrating the challenges and opportunities presented by big data and its applications (UN Global Pulse 2012).

At present, no precise and uniform definition of big data exists. Snijders et al. (2012) refer to big data as data collection that cannot be captured, curated, managed, and processed by using traditional data processing tools in a tolerable elapsed time. Some scholars have deliberated

---

[4] International Data Corporation (IDC) presented relevant research in their report "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East".

that big data has five characteristics (the 5Vs), with volume being the most fundamental and principal. The constant increase in data quantity is attributed to the improvement in storage technology, the acquisition of detailed information, and the wide use of digital sensors (Ohlhorst 2012). For example, Walmart processes over 1 million customer transaction records every hour, transmitting about 2.5 petabytes of data, the information quantity of which is 167 times that of the books stored at the Library of Congress of the United States (Johnson 2012). Furthermore, big data is usually generated in the form of dynamic, high-speed data flows (velocity). The value of the contained information decreases rapidly over time, thus requiring that data be tested and analyzed in real time (Schroeck et al. 2012). Variety, which indicates numerous data types and complex structures, is another important feature of big data. Structural data are stored in different tables based on predefined rules, and data access and filtration are relatively simple. However, non-structural data lack uniform and fixed modes or properties and cannot be arranged in the form of a traditional database. This presents challenges concerning the storage and analysis of such data (Ohlhorst 2012). Another problem that needs to be considered about big data is its veracity, that is, the inherent inaccuracy of certain kinds of data. One typical example is that although many countries require that a certain proportion of the yield of renewable resources be utilized in regional energy production, the unpredictability of wind energy makes it difficult to form plans (Schroeck et al. 2012). Some studies have suggested that big data has the problem of valorization. The current negligence in supervision and the imperfections of the incentive and reward mechanism affect the ability of big data to propagate knowledge appreciation and innovation. The problem is especially serious in low- and middle-income countries (LMICs), but the situation is improving (Özdemir et al. 2013).

Big data contains huge values through which we can better understand consumers, optimize supply chains and human resources, and improve financial indexes to bring profound insight to decision-makers (Wamba et al. 2015). Selecting suitable analytic tools according to the above characteristics of big data to acquire the information and knowledge needed from multifarious data quantities is the key to developing big data. Internationally, research is still in the preliminary phase, and highly developed modern information technology is needed to put forward relevant theories that are ready to be tested and modified. The combined use of advanced analytic technologies, including predictive parsing, data mining, statistics, human intelligence, natural language processing, and data visualization will be the main tool to analyze big data (Russom 2011). Enhancing storage capacity and developing technology to counter the rapid growth of data and to analyze its life cycle, evolution, and transmission laws to propagate research on the theory, method, and application in the society, economy, and environment are key challenges in the field of big data that need to be solved urgently. Only by collecting, processing, and acquiring key information and constructing appropriate evaluation theories and methods can big data be transformed into useful information for decision-making.

4 Big data research that relates to environmental management

Currently, environmental management data generated by remote sensing, network-based investigation, and computer modeling are increasing rapidly, and even social contact media have attracted researchers' attention (Jang and Hart 2015). For example, different kinds of production enterprises and merchants can directly release various kinds of information through network platforms; consumers can obtain information quickly and consequently

select more environment-friendly products owing to the aspects of maintaining individual health and protecting the environment according to the obtained information. Moreover, consumer selections will be transmitted or fed back through these network platforms, thus encouraging merchants and production enterprises to improve the environment-friendly quality of products. During this process, big data contains abundant information. If relevant data in the field of environmental management could be passed to the government, the concerned government officials would be motivated to improve the level of environmental management. Thus, analytic tools must be developed to handle these data with different structures for environmental evaluation and prediction. The linked open data method is one tool for data mining and analysis that is more favorable for interdisciplinary analyses, especially those that involve environmental analysis (Lausch et al. 2015). Some scholars have suggested that different countries and regions need to cooperate more broadly to collect and sort through big data in the areas of energy resources and the environment, and then test the level of global sustainable development through modeling (Gijzen 2013). The capacity of big data lies in accelerated growth, which will raise increasingly complicated questions for scientific researchers, including those concerning space-time dependence in multiple scales and multiple social aspects. Thus, the traditional data processing approach is no longer applicable. Given the large quantity of data in the environmental evaluation, one available option is to reduce the dimensions. First, the huge quantity of data can be divided into several subdata sets by using a sampling technique based on data types. The optimal data mining technology can then be employed to integrate these subdata sets. Finally, the environmental evaluation indexes may be divided into several equivalence classes according to the required accuracy, and the subdata may be evaluated accordingly. It becomes progressively more important to find methodological solutions (Wikle et al. 2013).

The exploration of big data confronts researchers with many difficulties and challenges (Bizer et al. 2012). However, numerous opportunities are also provided for the development of the advanced sciences, including ecological science and information resource management (Hampton et al. 2013). To improve the research efficiency, some scholars proposed the concept of big science (Aronova et al. 2010), which is based on a long-term ecological research network. Big science depends on a big-enough database system set up by governments and financial groups, and the data it contains can fully cope with the problems likely to be faced during scientific research processes; however, not all the important data are included (Hampton et al. 2013). In social ecology, for the construction of big data, the sources of data and characteristic analyses are of great significance to the scientific nature of management decision-making (Reichman et al. 2011).

Given that data in various dimensions are encountered in the analysis of big data of ecological science and sustainability, higher demands for synergy and sharing have emerged. For example, the US Department of Agriculture and the US EPA facilitated the synergy between agricultural development and environmental protection through data exchange and sharing (Hawkins et al. 2013). One important example about data sharing is the focus on public feedback, such as sharing environmental data to the public through the Internet. This will stimulate public participation in environmental management and could further generate a dynamic, complex, and large amount of feedback data on environmental evaluation. Integrating data sharing into environmental performance evaluation will be beneficial in formulating unified priority objectives among the government, enterprises, and the public, as well as enhancing the accuracy of environmental performance evaluation and improving the environmental management level. However, as the study of social ecology has been in the long tail of science for a long time (Heidorn 2008), the sharing and synergetic collection of data resources in practical scientific research cannot be realized smoothly (Ellison 2010).

Although scholars have focused on LCA because a different selection of samples will result in different evaluation results, the evaluation methods that are based on LCA are difficult to popularize. For example, when performing environmental performance evaluation from the perspective of nations, there may be significant differences between evaluation results and reality (Cooper et al. 2013). Some scholars also use the infinite dimensional spectral theory in the functional analysis of referencing and adopting eigenvalues and eigenvectors to avoid such mistakes (Cooper et al. 2013). Some scholars propose the idea of Climate Analytics-as-a-Service, inquiring whether computing that stimulates innovation and technology transfer can be applied to the big data analysis relating to the climate field, but its potential renewability and capability needs to be further assessed (Schnase et al. 2014). Dubey et al. (2015c) investigated the effects of big data on a world-class sustainable manufacturing industry and presented a big data analytic framework for the reduction of gathered data. They applied this framework to big data that satisfies the 5Vs. Some scholars have proposed the Artificial Neural Network (ANN) (Millie et al. 2013) or a combination of ANN and Geographic Information Systems (Pijanowski et al. 2014) to evaluate the ecological environment. However, the actual ecological relevance of ANN requires further verification.

The complexity and huge values of big data research in the field of environmental management will inevitably facilitate the further improvement and innovation of existing evaluation theories and methods. Although DEA, which is one of the most commonly used methods for environmental performance evaluation, has already been used in the efficiency evaluation of large-scale data sets by researchers such as Emrouznejad and Shale (2009) and Medina-Borja et al. (2007), there are only a few cases of its application to environmental performance evaluation facing big data. Not only does this indicate the commencement of big data theories, methods, and applications but it also predicts that the research of environmental performance evaluation with big data, on the basis of DEA, will develop widely in the future.

Finally, we will take environmental performance evaluations of thermal power plants as examples to explain the applications of big data. The thermal power industry is a typical high-emission industry, with the primary undesirable outputs being total suspended particulates with diameters smaller than $100\,\mu m$ and dust particles with diameters larger than $10\,\mu m$, both of which received considerable attention recently. Other emissions include respirable particulates, including PM10 with diameters smaller than $10\,\mu m$ and PM2.5 with diameters smaller than $2.5\,\mu m$, which are harmful to human health, as well as exhaust gases, such as SO_2 and NOX. These emission data are acquired mainly by installing environmental monitoring devices. However, not all thermal power plants are capable of installing such devices. The bulletin published by the State Ministry of Environmental Protection at the end of 2013 showed that 3127 thermal power plants had been included in the scope of the key-point investigation and statistics. Along with the enhancement of environmental protection in China, an increasing number of thermal power plants will be monitored. In addition to governmental monitoring, the public is highly likely to pay more attention to the pollution discharge conditions of enterprises. Meanwhile, information is freely available on the Internet, and public evaluations will be reflected on such media. Data acquired through monitoring devices and public feedback satisfy the five basic features of big data. By processing and sorting these data, we can screen out the valuable information that we need. However, the originally collected emission data are mixed with many inefficient, time-varying, inaccurate, and unstable data or data in the form of extreme values. Relationships between inputs, desirable outputs, and undesirable outputs are very complicated. Thus, we cannot directly analyze these data but instead have to apply a dimension reduction process. One suggested method is the developed DEA + LCA approach. First, each input is analyzed during the life cycle of repeatedly usable inputs by using the DEA method to perform environmental performance analysis with

consideration of undesirable outputs. Then, the final environmental performance evaluation results are obtained through the LCA method (Vázquez-Rowe et al. 2012). In sum, only through the continuous comparison and development of new reliable analytic methods of big data can decision-making suggestions for improving environmental management levels of thermal power plants be provided.

5 Summary and prospects

The collection and arrangement of big data in the context of environmental management and the construction of proper scientific models for performance evaluation will provide the basis for establishing an environmental protection platform, as well as for improving the effects and efficiency of environmental protection. This will also provide a reference for the improvement of environmental treatment schemes. However, as the relevant literature shows, the theory and method of environmental performance evaluation with big data can still be significantly improved in at least three aspects:

(a) The traditional theoretical system of the axiomatization of environmental performance evaluation needs to be improved. A scientific, perfect, and feasible environmental performance evaluation system adaptable to big data should be set up, such that environmental performance evaluation can be conducted more accurately in terms of collaboration with increasing indicator data, and timely and accurate information can be provided for environmental management decision-making.

(b) An environmental performance evaluation system using big data is based on different scientific fields, including management science, computer science, statistics, and environmental science, and specific evaluation processes involve the improvement and integration of DEA, LCA, and artificial intelligence methods. Deciding the ways by which to select, extend, combine, and test these methods is the key to improving the evaluation confidence coefficient.

(c) Unstructured information should be collected, arranged, grouped, and summarized during the evaluation processes to minimize the loss of information. Moreover, when, after treatment, big data expresses conditions of infinite samples and finite indicators, rules about sample homogeneity should be set up, and homogenization should be applied to non-homogeneous samples by feasible means.

(d) New extended models must be designed based on traditional environmental performance evaluation theories and methods for constructing the theory and method system of environmental performance evaluation using big data. This requires the effective identification and complete sequencing of a huge number of DMUs, an analysis of the relationship between undesirable outputs and between inputs and desirable outputs, the processing of dynamic and unstructured information, an effective measurement of inaccurate and unstable data, a homogenization technology for non-homogeneous DMUs, and the consideration of combined performance evaluation that can make multiple use of inputs. After verifying the applicability, reliability, and stability of such a theory and method system, it can be applied to practical environmental performance evaluations to provide a scientific basis for designing environmental protection policies in the new era.

Apart from the focus on innovation and the development of theories, another important application direction of environmental performance evaluation methods is the performance evaluation of the environmental supply chain with the help of big data. Currently, the supply chain development strategy has become an important tool for industrial competition.

The concept and requirement of the environmental supply chain encompass "green" and "environmental protection," both of which should run through the whole process of supply chain management. Energy consumption and pollutant emission throughout the supply chain should be minimized, and the industry should be able to develop sustainably. The manner by which to establish a big data analytic system that supports the environmental supply chain and to integrate data resources in the big data era for evaluating environmental efficiencies and resource efficiencies in the industrial supply chain will be important application directions of theories on big data environmental performance evaluation.

References

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*(10), 1261–1264.

Arabi, B., Doraisamy, S. M., Emrouznejad, A., & Khoshroo, A. (2016). Eco-efficiency measurement and material balance principle: An application in power plants Malmquist Luenberger Index. *Annals of Operations Research*. doi:10.1007/s10479-015-1970-x.

Aronova, E., Baker, K. S., & Oreskes, N. (2010). Big science and big data in biology: From the international geophysical year through the international biological program to the long term ecological research (LTER) network, 1957-present. *Historical Studies in the Natural Sciences*, *40*(2), 183–224.

Bagliani, M., Galli, A., Niccolucci, V., & Marchettini, N. (2008). Ecological footprint analysis applied to a sub-national area: The case of the province of Siena (Italy). *Journal of Environmental Management*, *86*(2), 354–364.

Bi, G., Luo, Y., Ding, J., & Liang, L. (2012). Environmental performance analysis of Chinese industry from a slacks-based perspective. *Annals of Operations Research*, *228*(1), 65–80.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: Four perspectives-four challenges. *ACM SIGMOD Record*, *40*(4), 56–60.

Blengini, G. A., Busto, M., Fantoni, M., & Fino, D. (2012). Eco-efficient waste glass recycling: Integrated waste management and green product development through LCA. *Waste Management*, *32*(5), 1000–1008.

Bogetoft, P., & Wang, D. (2005). Estimating the potential gains from mergers. *Journal of Productivity Analysis*, *23*(2), 145–171.

Cabeza, L. F., Rincón, L., Vilariño, V., Pérez, G., & Castell, A. (2014). Life cycle assessment (LCA) and life cycle energy analysis (LCEA) of buildings and the building sector: A review. *Renewable & Sustainable Energy Reviews*, *29*, 394–416.

Cai, Y., Lu, Y., Stegman, A., & Newth, D. (2015). Simulating emissions intensity targets with energy economic models: Algorithm and application. *Annals of Operations Research*. doi:10.1007/s10479-015-1927-0.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Chen, C.-M., & Delmas, M. A. (2012). Measuring eco-inefficiency: A new frontier approach. *Operations Research*, *60*(5), 1064–1079.

Cherubini, F., Bargigli, S., & Ulgiati, S. (2009). Life cycle assessment (LCA) of waste management strategies: Landfilling, sorting plant and incineration. *Energy*, *34*(12), 2116–2123.

Chevallier, J., & Goutte, S. (2015). Estimation of Lévy-driven Ornstein–Uhlenbeck processes: Application to modeling of CO_2 and fuel-switching. *Annals of Operations Research*. doi:10.1007/s10479-015-1967-5.

Coelli, T., Lauwers, L., & Van Huylenbroeck, G. (2007). Environmental efficiency measurement and the materials balance condition. *Journal of Productivity Analysis*, *28*(1–2), 3–12.

Cook, W. D., Harrison, J., Imanirad, R., Rouse, P., & Zhu, J. (2013). Data envelopment analysis with nonhomogeneous DMUs. *Operations Research*, *61*(3), 666–676.

Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA)-thirty years on. *European Journal of Operational Research*, *192*(1), 1–17.

Cook, W. D., & Zhu, J. (2011). Multiple variable proportionality in data envelopment analysis. *Operations Research*, *59*(4), 1024–1032.

Cooper, J., Noon, M., Jones, C., Kahn, E., & Arbuckle, P. (2013). Big data in life cycle assessment. *Journal of Industrial Ecology*, *17*(6), 796–799.

Cooper, J. S., Kahn, E., & Ebel, R. (2013). Sampling error in US field crop unit process data for life cycle assessment. *The International Journal of Life Cycle Assessment*, *18*(1), 185–192.

DeSimone, L. D., & Popoff, F. (2000). *Eco-efficiency: The business link to sustainable development*. Cambridge, MA: MIT Press.

Dong, J., Chi, Y., Zou, D., Fu, C., Huang, Q., & Ni, M. (2014). Energy-environment-economy assessment of waste management systems from a life cycle perspective: Model development and case study. *Applied Energy*, *114*, 400–408.

Dubey, R., Gunasekaran, A., & Ali, S. S. (2015a). Exploring the relationship between leadership, operational practices, institutional pressures and environmental performance: A framework for green supply chain. *International Journal of Production Economics*, *160*, 120–132.

Dubey, R., Gunasekaran, A., & Chakrabarty, A. (2015b). World-class sustainable manufacturing: Framework and a performance measurement system. *International Journal of Production Research*. doi:10.1080/00207543.2015.1012603.

Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2015c). The impact of big data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology*. doi:10.1007/s00170-015-7674-1.

Dyckhoff, H., & Allen, K. (2001). Measuring ecological efficiency with data envelopment analysis (DEA). *European Journal of Operational Research*, *132*(2), 312–325.

Ellison, A. M. (2010). Repeatability and transparency in ecological research. *Ecology*, *91*(9), 2536–2539.

Emrouznejad, A., & Shale, E. (2009). A combined neural network and DEA for measuring efficiency of large scale datasets. *Computers & Industrial Engineering*, *56*(1), 249–254.

Fadeyi, S., Arafat, H. A., & Abu-Zahra, M. R. (2013). Life cycle assessment of natural gas combined cycle integrated with CO_2 post combustion capture using chemical solvent. *International Journal of Greenhouse Gas Control*, *19*, 441–452.

Färe, R., Grosskopf, S., Lovell, C. K., & Pasurka, C. (1989). Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. *The Review of Economics and Statistics*, *71*(1), 90–98.

Färe, R., Grosskopf, S., Lovell, C. K., & Yaisawarng, S. (1993). Derivation of shadow prices for undesirable outputs: A distance function approach. *The Review of Economics and Statistics*, *75*(2), 374–380.

Färe, R., Grosskopf, S., Noh, D.-W., & Weber, W. (2005). Characteristics of a polluting technology: Theory and practice. *Journal of Econometrics*, *126*(2), 469–492.

Finnveden, G., Hauschild, M. Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., et al. (2009). Recent developments in life cycle assessment. *Journal of Environmental Management*, *91*(1), 1–21.

Gijzen, H. (2013). Development: Big data for a sustainable future. *Nature*, *502*(7469), 38–38.

Glucker, A. N., Driessen, P. P., Kolhoff, A., & Runhaar, H. A. (2013). Public participation in environmental impact assessment: Why, who and how? *Environmental Impact Assessment Review*, *43*, 104–111.

Govindan, K., Sarkis, J., Jabbour, C. J. C., Zhu, Q., & Geng, Y. (2014). Eco-efficiency based green supply chain management: Current status and opportunities. *European Journal of Operational Research*, *2*(233), 293–298.

Guinée, J., Gorrée, M., Heijungs, R., Huppes, G., Kleijn, R., de Koning, A., et al. (2002). *Handbook on life cycle assessment operational guide to the ISO standards*. Dordrecht: Kluwer.

Hailu, A., & Veeman, T. S. (2001). Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. *American Journal of Agricultural Economics*, *83*(3), 605–616.

Halkos, G. E., & Tzeremes, N. G. (2013). A conditional directional distance function approach for measuring regional environmental efficiency: Evidence from UK regions. *European Journal of Operational Research*, *227*(1), 182–189.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., et al. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162.

Hashimoto, S., & Moriguchi, Y. (2004). Proposal of six indicators of material cycles for describing society's metabolism: From the viewpoint of material flow analysis. *Resources, Conservation and Recycling*, *40*(3), 185–200.

Hauschild, M., & Pennington, D. (2002). Indicators for ecotoxicity in life cycle impact assessment. In H. A. Udo de Haes, G. Finnveden, M. Goedkoop, M. Hauschild, E. Hertwich, P. Hofstetter, W. Klöpffer, W. Krewitt, E. Lindeijer, O. Jolliet, R. Müller-Wenk, S. L. Olsen, D. Pennington, J. Potting, & B. Steen (Eds.), *Life-cycle impact assessment: Striving towards best practice*. Pensacola, FL: SETAC Press.

Hawkins, T., Ingwersen, W., Srocka, M., Transue, T., & Ciroth, A. (2013). Tools to support the widespread application of life cycle assessment: Development of the life cycle assessment harmonization tool and

the improvement of openLCA. *2013 International symposium on sustainable systems and technology*, May 15–17,2013. http://lcacenter.org/lcaxiii/abstracts/abstract-dynamic.php?id=946. Accessed 17 September 2013.

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, *57*(2), 280–299.

Hjaila, K., Baccar, R., Sarrà, M., Gasol, C., & Blánquez, P. (2013). Environmental impact associated with activated carbon preparation from olive-waste cake via life cycle assessment. *Journal of Environmental Management*, *130*, 242–247.

Hunt, R. G., Franklin, W. E., & Hunt, R. (1996). LCA-how it came about. *The International Journal of Life Cycle Assessment*, *1*(1), 4–7.

Iribarren, D., Martín-Gamboa, M., & Dufour, J. (2013). Environmental benchmarking of wind farms according to their operational performance. *Energy*, *61*, 589–597.

Iribarren, D., Vázquez-Rowe, I., Moreira, M. T., & Feijoo, G. (2010). Further potentials in the joint implementation of life cycle assessment and data envelopment analysis. *Science of the Total Environment*, *408*(22), 5265–5272.

ISO 14040:1997. (1997). *Environmental management-life cycle assessment-principles and framework*. http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=23151. Accessed 15 September 2012.

ISO 14044:2006. (2006). *Environmental management-life cycle assessment-requirements and guidelines*. http://www.iso.org/iso/catalogue_detail?csnumber=37456. Accessed 9 September 2013.

Jang, S. M., & Hart, P. S. (2015). Polarized frames on climate change and global warming across countries and states: Evidence from twitter big data. *Global Environmental Change*, *32*, 11–17.

Jawahar, N., Pandian, G. S., Gunasekaran, A., & Subramanian, N. (2015). An optimization model for sustainability program. *Annals of Operations Research*. doi:10.1007/s10479-015-1928-z.

Johnson, J. E. (2012). Strategy: Big data + big analytics = big opportunity. *Financial Executive*, *28*(6), 50.

Kao, C., & Lin, P.-H. (2012). Efficiency of parallel production systems with fuzzy data. *Fuzzy Sets and Systems*, *198*, 83–98.

Koskela, M., & Vehmas, J. (2012). Defining eco-efficiency: A case study on the Finnish forest industry. *Business Strategy and the Environment*, *21*(8), 546–566.

Lausch, A., Schmidt, A., & Tischendorf, L. (2015). Data mining and linked open data-new perspectives for data analysis in environmental research. *Ecological Modelling*, *295*, 5–17.

Liang, L., Wu, J., Cook, W. D., & Zhu, J. (2008). The DEA game cross-efficiency model and its Nash equilibrium. *Operations Research*, *56*(5), 1278–1288.

Liu, W., Meng, W., Li, X., & Zhang, D. (2010). DEA models with undesirable inputs and outputs. *Annals of Operations Research*, *173*(1), 177–194.

Liu, W., & Sharp, J. (1999). DEA models via goal programming. In G. Westermann (Ed.), *Data envelopment analysis in the service sector* (pp. 79–101). New York: Springer.

Lozano, S. (2015). Technical and environmental efficiency of a two-stage production and abatement system. *Annals of Operations Research*. doi:10.1007/s10479-015-1933-2.

Lozano, S., Iribarren, D., Moreira, M. T., & Feijoo, G. (2010). Environmental impact efficiency in mussel cultivation. *Resources, Conservation and Recycling*, *54*(12), 1269–1277.

Lv, W., Zhou, Z., & Huang, H. (2013). The measurement of undesirable output based-on DEA in E&E: Models development and empirical analysis. *Mathematical and Computer Modelling*, *5*(58), 907–912.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). Big Data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. Accessed 9 October 2014.

Medina-Borja, A., Pasupathy, K., & Triantis, K. (2007). Large-scale data envelopment analysis (DEA) implementation: A strategic performance management approach. *Journal of the Operational Research Society*, *58*(8), 1084–1098.

Mestre, A., & Vogtlander, J. (2013). Eco-efficient value creation of cork products: An LCA-based method for design intervention. *Journal of Cleaner Production*, *57*, 101–114.

Millie, D. F., Weckman, G. R., Young, W. A., Ivey, J. E., Fries, D. P., Ardjmand, E., et al. (2013). Coastal 'big data' and nature-inspired computation: Prediction potentials, uncertainties, and knowledge derivation of neural networks for an algal metric. *Estuarine, Coastal and Shelf Science*, *125*, 57–67.

Mohammadi, A., Rafiee, S., Jafari, A., Dalgaard, T., Knudsen, M. T., Keyhani, A., et al. (2013). Potential greenhouse gas emission reductions in soybean farming: A combined use of life cycle assessment and data envelopment analysis. *Journal of Cleaner Production*, *54*, 89–100.

Mouter, N., Annema, J. A., & Van Wee, B. (2013). Ranking the substantive problems in the Dutch cost-benefit analysis practice. *Transportation Research Part A: Policy and Practice*, *49*, 241–255.

Nahorski, Z., & Ravn, H. F. (2000). A review of mathematical models in economic environmental problems. *Annals of Operations Research*, *97*(1–4), 165–201.

Oggioni, G., Riccardi, R., & Toninelli, R. (2011). Eco-efficiency of the world cement industry: A data envelopment analysis. *Energy Policy*, *39*(5), 2842–2854.

Ohlhorst, F. J. (2012). *Big data analytics: Turning big data into big money*. New York: Wiley.

Özdemir, V., Badr, K. F., Dove, E. S., Endrenyi, L., Geraci, C. J., Hotez, P. J., et al. (2013). Crowd-funded micro-grants for genomics and big data: An actionable idea connecting small (artisan) science, infrastructure science, and citizen philanthropy. *OMICS: A Journal of Integrative Biology*, *17*(4), 161–172.

Paco, A., & Raposo, M. (2009). Identifying the green consumer: A segmentation study. *Journal of Targeting, Measurement and Analysis of Marketing*, *17*(1), 17–25.

Pérez-López, P., González-García, S., Ulloa, R. G., Sineiro, J., Feijoo, G., & Moreira, M. T. (2014). Life cycle assessment of the production of bioactive compounds from tetraselmissuecica at pilot scale. *Journal of Cleaner Production*, *64*, 323–331.

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environmental Modelling & Software*, *51*, 250–268.

Podinovski, V. V., & Bouzdine-Chameeva, T. (2013). Weight restrictions and free production in data envelopment analysis. *Operations Research*, *61*(2), 426–437.

Ramli, N. A., Munisamy, S., & Arabi, B. (2013). Scale directional distance function and its application to the measurement of eco-efficiency in the manufacturing sector. *Annals of Operations Research*, *211*(1), 381–398.

Reap, J., Roman, F., Duncan, S., & Bras, B. (2008). A survey of unresolved problems in life cycle assessment. *The International Journal of Life Cycle Assessment*, *13*(5), 374–388.

Reichman, O., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, *331*(6018), 703–705.

Russom, P. (2011). *Fourth quarter: Big data analytics*. TDWI Best Practices Report.

Scheel, H. (2001). Undesirable outputs in efficiency valuations. *European Journal of Operational Research*, *132*(2), 400–410.

Schleisner, L. (2000). Life cycle assessment of a wind farm and related externalities. *Renewable Energy*, *20*(3), 279–288.

Schnase, J. L., Duffy, D. Q., Tamkin, G. S., Nadeau, D., Thompson, J. H., Grieg, C. M., et al. (2014). Merra analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. *Computers, Environment and Urban Systems*. doi:10.1016/j.compenvurbsys.2013.12.003.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P., (2012). Analytics: The real-world use of big data. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html. Accessed 3 May 2014.

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *142*(1), 16–20.

Seiford, L. M., & Zhu, J. (2005). A response to comments on modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *161*(2), 579–581.

Slagstad, H., & Brattebø, H. (2014). Life cycle assessment of the water and wastewater system in Trondheim, Norway—A case study. *Urban Water Journal*, *11*(4), 323–334.

Snijders, C., Matzat, U., & Reips, U.-D. (2012). Big data: Big gaps of knowledge in the field of internet science. *International Journal of Internet Science*, *7*(1), 1–5.

Song, M., An, Q., Zhang, W., Wang, Z., & Wu, J. (2012). Environmental efficiency evaluation based on data envelopment analysis: A review. *Renewable and Sustainable Energy Reviews*, *16*(7), 4465–4469.

Song, M., & Guan, Y. (2014). The environmental efficiency of Wanjiang demonstration area: A Bayesian estimation approach. *Ecological Indicators*, *36*, 59–67.

Song, M., Zhang, L., An, Q., Wang, Z., & Li, Z. (2013). Statistical analysis and combination forecasting of environmental efficiency and its influential factors since China entered the WTO: 2002–2010–2012. *Journal of Cleaner Production*, *42*, 42–51.

Stamp, A., Althaus, H.-J., & Wäger, P. A. (2013). Limitations of applying life cycle assessment to complex co-product systems: The case of an integrated precious metals smelter-refinery. *Resources, Conservation and Recycling*, *80*, 85–96.

Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*, *22*(2), 127–151.

Tiruta-Barna, L., Benetto, E., & Perrodin, Y. (2007). Environmental impact and risk assessment of mineral wastes reuse strategies: Review and critical analysis of approaches and applications. *Resources, Conservation and Recycling*, *50*(4), 351–379.

Tone, K. (2003). *Dealing with undesirable outputs in DEA: A slacks-based measure (SBM) approach*. GRIPS Research Report Series. http://ci.nii.ac.jp/naid/120005571101/. Accessed 5 January 2014.

Tone, K., & Tsutsui, M. (2014). Dynamic DEA with network structure: A slacks-based measure approach. *Omega*, *42*(1), 124–131.

Tyteca, D. (1996). On the measurement of the environmental performance of firms—A literature review and a productive efficiency perspective. *Journal of Environmental Management*, *46*(3), 281–308.

UN Global Pulse. (2012). *Big data for development: Challenges and opportunities*. http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf. Accessed 3 May 2014.

Vázquez-Rowe, I., Villanueva-Rey, P., Iribarren, D., Moreira, M. T., & Feijoo, G. (2012). Joint life cycle assessment and data envelopment analysis of grape production for vinification in the RíasBaixas appellation (NW Spain). *Journal of Cleaner Production*, *27*, 92–102.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246.

Wikle, C., Holan, S., & Hooten, M. (2013). Guest editor's introduction to the special issue on modern dimension reduction methods for big data problems in ecology. *Journal of Agricultural, Biological, and Environmental Statistics*, *3*(18), 271–273.

Wilson, B. P., Lavery, N. P., Jarvis, D. J., Anttila, T., Rantanen, J., Brown, S. G., et al. (2013). Life cycle assessment of gas atomised sponge nickel for use in alkaline hydrogen fuel cell applications. *Journal of Power Sources*, *243*, 242–252.

Wu, J., Chu, J., Sun, J., & Zhu, Q. (2016). DEA cross-efficiency evaluation based on Pareto improvement. *European Journal of Operational Research*, *248*(2), 571–579.

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107.

You, S., & Yan, H. (2011). A new approach in modelling undesirable output in DEA model. *Journal of the Operational Research Society*, *62*(12), 2146–2156.

Zhou, G., Chung, W., & Zhang, Y. (2014). Measuring energy efficiency performance of China's transport sector: A data envelopment analysis approach. *Expert Systems with Applications*, *41*(2), 709–722.

Zhou, P., Ang, B., & Poh, K. (2008). Measuring environmental performance under different environmental DEA technologies. *Energy Economics*, *30*(1), 1–14.

Zhou, P., Poh, K. L., & Ang, B. W. (2007). A non-radial DEA approach to measuring environmental performance. *European Journal of Operational Research*, *178*(1), 1–9.

Zhu, D., & Ru, J. (2008). Strategic environmental assessment in China: Motivations, politics, and effectiveness. *Journal of Environmental Management*, *88*(4), 615–626.