Constructing Meaningful Environmental Indices: A Nonparametric Frontier Approach

P. Zhou [a,*], M.A. Delmas [b,*], A. Kohli [b],

[a] *College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing 211106, China*

[b] *Institute of the Environment and Sustainability, University of California, Los Angeles La Kretz Hall Suite 300, Los Angeles, CA 90095-1496*

Journal of Environmental Economics and Management

2017

Abstract: Environmental information disclosure programs seek to motivate firms to reduce their environmental impact. A variety of environmental impacts are reported in these programs and often this information is aggregated into a composite environmental index (CEI) for easier communication. The challenge is to create a meaningful index that allows environmental performance to be compared over time and space without ambiguity. In this paper, we argue that it is important to develop a cardinally meaningful and standardized CEI and use a nonparametric frontier approach to constructing such an index. This approach has the advantage to handle issues associated with data irregularity and the mixed measurability of underlying variables. We apply this approach to construct a CEI for evaluating the environmental performance of manufacturing facilities in different industrial sectors in Los Angeles based on data from the toxic release inventory. We show how the CEI can be used to improve facility-level environmental performance. A sensitivity analysis is conducted with respect to the uncertainty in data accuracy, which demonstrates the robustness of the nonparametric frontier approach in constructing meaningful environmental indices.

Keywords: Environmental index; Measurability; Toxic Release Inventory (TRI); Environmental performance; Data envelopment analysis

---

Acknowledgments

We would like to thank the Journal Co-Editor-in-Chief Till Requate and two anonymous referees for their invaluable comments and suggestions on an earlier draft of this manuscript. P. Zhou is also grateful to the financial support provided by the National Natural Science Foundation of China (nos. 71625005, 71573119 & 71273005).

[*] Corresponding authors.
E-mail addresses: delmas@ucla.edu (M.A.Delmas), rocy_zhou@hotmail.com (P. Zhou).

1. Introduction

Environmental information disclosure has been touted as a powerful tool that can augment traditional command and control regulation and influence positive environmental performance through public pressure (Konar and Cohen, 1997). A well-known example is the Toxic Release Inventory (TRI), an initiative from the U.S. Environmental Protection Agency (EPA), which imposes mandatory disclosure requirements on large industrial facilities to release information on toxic emissions (Koehler and Spengler, 2007). The success of the TRI hinges on public pressure imposed on plants with poor environmental performance, which provides an incentive for plants to adopt stronger environmental measures and improve their environmental performance (Khanna and Anton, 2002) or, alternatively, positive public recognition for the best performers. Many users of TRI data tend to evaluate environmental performance based on a single metric such as total releases of toxic chemical emissions, while ignoring other potentially relevant dimensions of economic performance such as revenue generated, or employment data, which are crucial aspects of businesses. We argue that a composite environmental index (CEI) must consider environmental performance in conjunction with other measures of corporate performance to identify the "best" plants and practices, those that achieve both environmental and economic success.

Besides the variables used, the CEI must exhibit other properties such as "meaningfulness" and standardization. The terminology of a "meaningful index" originated from the influential study by Ebert and Welsch (2004) that characterized classes of environmental indices. As a fundamental scientific rule, 'meaningfulness' implies that the comparison of environmental performance across time or space based on CEIs must be free of ambiguity (Welsch, 2005). However, when variables with different scale properties, for example, tons of air pollutants (ratio-scale) and temperature measured on the Celsius scale (interval-scale), are combined, it is difficult to aggregate them into a CEI in a meaningful way (Böhringer and Jochem, 2007). In addition, since Tyteca (1996, 1997) scholars have been advocating for the development of a standardized aggregate index between zero and one in order to allow for a proper comparison of environmental performance between firms.[1] Hence it is important to construct a meaningful and standardized CEI that is capable of handling issues of mixed measurability of underlying variables (i.e. both ratio-scale and interval-scale variables are involved) and data irregularity (e.g. the existence of multiple zero entries).

In the existing literature, methods for measuring environmental performance for firms may be broadly classified into two groups. The first group aims to measure environmental performance from the perspective of productive efficiency, which involves classifying underlying variables into inputs and outputs and specifying an environmental production technology for modeling the joint production of good and bad outputs. Within this first group, there are two strategies for the measurement of environmental performance of firms or plants. One strategy is to calculate an adjusted

[1] The study by Tyteca (1996) provided an excellent review of the existing methods for measuring environmental performance of firms, which ranges from simple indicators reflecting only one aspect of the impact of activities to more sophisticated ones reflecting the overall impact on the environment.

measure of efficiency or productivity whereby a firm or plant is credited for simultaneously increasing good output production and reducing bad output production. Pittman (1983) conducted one of the earliest studies initiating this strategy by incorporating pollutants into productivity measurement. Subsequently, the seminal study by Färe et al. (1989) laid an elegant theoretical foundation for using nonparametric frontier methodology to evaluate productive efficiency with undesirable outputs. The framework developed by Färe et al. (1989) has been adopted by a large number of studies, which have focused on not only firms and plants (e.g. Boyd and McClelland, 1999; Färe et al., 1997; Färe et al., 2010; Khanna and Kumar, 2011) but also countries and regions (e.g. Zhou et al., 2010; Hoang and Coelli, 2011; Picazo-Tadeo et al., 2014). The other strategy involves constructing a formal environmental performance index (EPI) by using Shephard or Malmquist distance functions (Färe et al., 2004, 2006, 2010). Its advantage lies in the fact that the resulting EPI holds some desirable index number properties. Both of these strategies can be implemented by utilizing data envelopment analysis (DEA) models.

The second group attempts to aggregate multiple environmental variables into a CEI for performance evaluation and comparison. It allows the use of diverse variables in accordance with the environmental theme being studied. Ebert and Welsch (2004) showed that a geometric mean can lead to a meaningful index when the underlying variables are ratio-scale and strictly positive. Zhou et al. (2006) developed an information loss criterion to assess alternative aggregation rules for constructing CEIs. Munda and Nardo (2009) highlighted the usefulness of non-compensatory aggregation approach. While many previous studies focused on data aggregation, several scholars have examined other important issues such as weighting (e.g. Decancq and Lugo, 2013) and normalization (e.g. Zhou and Ang, 2009; Pollesch and Dale, 2016).

This paper contributes to the existing body of studies on CEIs in the following aspects. First, as an extension to the important work by Ebert and Welsch (2004), we classify 'meaningfulness' into ordinal and cardinal meaningfulness and argue the importance of constructing a cardinally meaningful CEI. Second, in the spirit of the influential studies by Färe et al. (1996, 2004, 2006, 2010) and Tyteca (1996, 1997), we advocate the use of a nonparametric frontier approach for constructing a standardized CEI that simultaneously satisfies cardinal meaningfulness. This approach can also address data irregularity issues that appear in TRI data such as a large number of zeros, which pose challenges in the application of the theoretically meaningful aggregation rules such as geometric mean. Third, we apply this methodology to TRI data for evaluating facility-level environmental performance in Los Angeles County, which provides perspectives on how facilities may improve their environmental performance.

The remainder of this paper is organized as follows. In Section 2, we describe the concept of a meaningful environmental index that is composed of two categories, namely "ordinal meaningfulness" and "cardinal meaningfulness." In Section 3, we present the nonparametric frontier approach and show its desirable theoretical properties as compared to arithmetic aggregation. In Section 4, we describe our empirical study using the nonparametric frontier methodology to evaluate the environmental performance of different facilities from three industrial sectors in Los Angeles County. Concluding remarks are presented in Section 5.

2. Meaningful Composite Environmental Indices

The concept of "meaningfulness" given by Ebert and Welsch (2004) is built upon the invariance of preference orderings with respect to the measurement units of underlying variables. In addition to orderings, the relative performance gaps of CEI values may also carry valuable information for performance comparison and improvement. As such, we classify 'meaningfulness' into 'ordinal meaningfulness' and 'cardinal meaningfulness' and use a simple example to illustrate the importance of cardinal meaningfulness in this section.

2.1 Definitions

Let $V_k = (v_{k1}, \cdots, v_{kn})$ denote a vector of n underlying environmental variables for entity k ($k = 1, \cdots, K$). Our task is to construct a CEI for each entity based on the n variables. One usage of constructing the CEI is to provide a ranking of different entities in environmental performance, which can be characterized by the preference ordering $\succeq$ defined on $\Re^n$. Thus a CEI can be represented by a mapping function $I : \Re^n \to R$ that satisfies

$$V_k \succeq V_l \Leftrightarrow I(V_k) \geq I(V_l) \quad \forall k, l \in \{1, \cdots, K\} \tag{1}$$

Note that the measurement units of the underlying n variables may be changed, which can be represented by a transformation function $F = (f_1, \cdots, f_n)$ such that

$$F : (v_{k1}, \cdots, v_{kn}) \to (f_1(v_{k1}), \cdots, f_n(v_{kn})) \tag{2}$$

As described in Ebert and Welsch (2004) and Welsch (2005), an admissible transformation involves expansion as well as translation, i.e. $f_i(v_{ki}) = \alpha_i v_{ki} + \beta_i, \alpha_i > 0$. With reference to CEIs, the orderings of different entities are expected to be invariant with respect to any admissible transformation of underlying variables (Ebert and Welsch, 2004; Welsch, 2005), i.e.

$$V_k \succeq V_l \Leftrightarrow F(V_k) \succeq F(V_l) \quad \forall k, l \in \{1, \cdots, K\} \tag{3}$$

Definition 1 (*Ordinal meaningfulness*). I is an ordinally meaningful index if it satisfies

$$I(V_k) \geq I(V_l) \Leftrightarrow I(F(V_k)) \geq I(F(V_l)) \quad \forall k, l \in \{1, \cdots, K\} \tag{4}$$

It should be pointed out that Ebert and Welsch (2004) and Welsch (2005) termed I satisfying Eq. (4) as a meaningful index while we refer to it as an ordinally meaningful index in this paper. Ebert and Welsch (2004) showed that geometric aggregation would yield an ordinally meaningful CEI when the underlying variables are ratio-scale noncomparable. Despite the importance of ordinal meaningfulness, as discussed by Böhringer and Jochem (2007), many popular CEIs for sustainability did not take it into account and were therefore misleading with respect to policy practice.

Acknowledging the importance of ordinal meaningfulness, we argue that it is valuable for a CEI to preserve a relative performance gaps between entities. This may be illustrated by the case of a city-level air pollutant index derived from several air pollutants. If the index values are respectively 150, 140 and 50, it says that the last city shows the best while the first city shows the worst air pollution level. When the index values become 80, 60 and 50, the same message is transmitted regarding their orderings in air pollution level. Beyond it, we observe that the performance gaps between the first two cities and the last one become smaller. It implies that the index values also carry valuable information through their relative performance gaps between entities. Thus we have

Definition 2 (*Cardinal meaningfulness*). I is a cardinally meaningful index if it satisfies

$$I(V_k) = \alpha I(F(V_k)) \quad \forall k \in \{1, \cdots, K\} \tag{5}$$

where α is a positive constant.

Eq. (5) says that a cardinally meaningful CEI preserves the relative performance gaps between entities for any admissible transformation of underlying variables. $\alpha = 1$ implies that the CEI values will not change, which is not necessary for satisfying cardinal meaningfulness but still desirable as the resulting CEI looks more standardized. Obviously, cardinal meaningfulness represents a stronger requirement than ordinal meaningfulness. That is to say,

Proposition 1. *A cardinally meaningful index must be an ordinally meaningful index; not vice versa*.

Once a cardinally meaningful CEI is theoretically defined, the next task is to identify a way to compute its values. While adequate weighting, normalization and aggregation of underlying variables are often regarded as pre-requisites for the practice, Böhringer and Jochem (2007) pointed out that there are no unambiguous rules for data weighting and normalization as they often imply value judgements.[2]

[2] Normalization is the process of transforming the different measurement units of underlying variables into a common unit or dimensionless. The recent study by Pollesch and Dale (2016) provides a comprehensive

Regarding data aggregation, Welsch (2005) and Böhringer and Jochem (2007) described several meaningful aggregation methods dependent on the scale and comparability characteristics of underlying variables. An important finding of their studies is that the arithmetic mean is meaningful for variables satisfying interval scale and full comparability. While normalization can help achieve comparability and may internally be linked to meaningful aggregation, in this paper we only focus on the aggregation of underlying variables without explicitly discussing the normalization scheme.

2.2 An illustrative example

We use a simple example to illustrate the issue of data irregularity occurring in the TRI database and explain why arithmetic and geometric means aggregation rules are inappropriate for the application. Table 1 shows the data on two environmental variables for four selected facilities in the Chemicals industry in the Los Angeles County (Delmas and Kohli, 2014). Clearly, one data irregularity is that there exist multiple zero entries.

Table 1. Data for the illustrative example

Facility	Total toxic releases (Pounds)	Toxicity of on-site releases (Pounds-toxicity)
A	0	0
B	2000	6000
C	200	22000
D	180	0

As shown in Table 1, facility A obviously has the best environmental performance since it represents the best practice for the total toxic releases and the toxicity of the

discussion on alternative normalization methods in the context of sustainability assessment.

releases. If an arithmetic mean is applied to evaluate the performance for facilities B and C, their CEI values are respectively (2000+6000)/2=4000 and (200+22000)/2=11000 respectively. However, if the measurement unit of the second variable is changed to "thousand pounds-toxicity", the CEI values of B and C computed by arithmetic mean will become (2000+6)/2=1003 and (200+22)/2=111. The preference orderings of the two facilities are reversed, which verifies the conclusion drawn by Ebert and Welsch (2004) that an arithmetic mean cannot yield an ordinally meaningful CEI without normalization. Note that the two variables are ratio-scale noncomparable so that a geometric mean would yield a meaningful CEI for facilities B and C as shown by Ebert and Welsch (2004). Based on geometric means, their CEI values are respectively $(2000\times6000)^{0.5}=3464$ and $(200\times22000)^{0.5}=2098$. Nevertheless, if the entire dataset is considered, facilities A and D have at least one variable equal to zero. This violates the condition given by Ebert and Welsch (2004) that the observations of underlying variables are strictly positive. If the geometric mean is applied, the CEI values for facilities A and D are equal to zero indicating that the aggregation rule is not zero robust. While only ratio-scale variables are considered in this example, both ratio-scale and interval-scale variables might be involved in the application (i.e the mixed measurability of underlying variables). In this circumstance, even if all the observations are strictly positive, the geometric mean aggregation rule will not yield a meaningful index (Ebert and Welsch, 2004; Welsch, 2005).

3. Method

In this section, we introduce a nonparametric frontier approach, DEA, for constructing a cardinally meaningful and standardized CEI. This approach can easily address the issues of data irregularity that frequently appear in TRI data as well as the mixed measurability of the underlying variables.

3.1 DEA model

As a nonparametric frontier methodology, DEA employs linear programming to identify the best practice frontier and evaluate the relative performance of each entity based on the observations of inputs and outputs for a group of comparable entities (Coelli et al., 2005). Since the seminal study by Färe et al. (1989) and the influential work by Tyteca (1996), DEA has been widely applied to the measurement of environmental performance or pollutant-adjusted efficiency/productivity of different entities. Examples of such studies include Färe et al. (1996, 2004, 2006, 2007, 2010), Tyteca (1997), Boyd and McClelland (1999), Zhou et al. (2010), Hoang and Nguyen (2013) and Picazo-Tadeo et al. (2014).

The conventional use of DEA for environmental performance measurement starts from a differentiation between good and bad outputs as well as the specification of a production technology for modeling their joint production. In this line of research, Färe et al. (1989) has laid an elegant theoretical foundation, which makes the nonparametric frontier methodology popular for performance measurement with bad outputs such as pollutants. In constructing CEIs, however, there may not exist a productive relationship between underlying variables (e.g. the case of the air pollutant index). Nevertheless, the variables may also be divided into “inputs” and “outputs” which respectively satisfy the properties of “the smaller the better” and “the larger the better” from the perspective of performance improvement. To differentiate between inputs and outputs, we replace the vector $V_k = (v_{k1}, \cdots, v_{kn})$ given in Section 2 by $V_k = (X_k, Y_k) = (x_{k1}, \cdots, x_{km}, y_{k1}, \cdots, y_{ks})$ where X_k and Y_k are respectively the input and output vectors. Using all the observations, we can construct a quasi-reference technology as follows:[3]

$$
\begin{aligned}
\mathrm{S} = \{(X, Y) : & \sum_{k=1}^{K} x_{ki} z_k \le x_i, \quad i = 1, \cdots, m \\
& \sum_{k=1}^{K} y_{kr} z_k \ge y_r, \quad r = 1, \cdots, s \\
& \sum_{k=1}^{K} z_k = 1 \\
& z_k \ge 0, \ k = 1, \cdots, K \}
\end{aligned}
\tag{6}
$$

With Eq. (6) as the constraint, we can formulate the following range adjusted DEA model (Aida et al., 1998; Cooper et al., 1999):

$$
\begin{aligned}
& \max \frac{1}{m+s} \left(\sum_{i=1}^{m} \frac{s_i^-}{R_i^-} + \sum_{r=1}^{s} \frac{s_r^+}{R_r^+} \right) \\
& \text{s.t. } \sum_{k=1}^{K} x_{ki} z_k + s_i^- = x_{oi}, \quad i = 1, \cdots, m \\
& \qquad \sum_{k=1}^{K} y_{kr} z_k - s_r^+ = y_{or}, \quad r = 1, \cdots, s \\
& \qquad \sum_{k=1}^{K} z_k = 1 \\
& \qquad z_k \ge 0, s_i^- \ge 0, s_r^+ \ge 0
\end{aligned}
\tag{7}
$$

where x_{oi} and y_{or} respectively denote the *i*-th input and *r*-th output for entity o ($o \in \{1, \cdots, K\}$); R_i^- and R_r^+ denote the ranges for input *i* and output *r*, which are

[3] The term “quasi-reference technology” implies that it looks like a reference technology externally but the productive relationship between inputs and outputs may not exist. Since the choice of inputs and outputs for constructing a CEI is dependent on the environmental theme concerned, the commonly used inputs such as capital, labor and energy may not be included in the construction of CEIs. Actually, Färe et al. (2006, 2010) also excluded such inputs in developing an environmental performance index that has an advantage of crediting a producer for adopting processes generating more good output per unit of bad output produced. A difference is that Färe et al. (2006, 2010) classified outputs into good and bad outputs while we treat bad outputs as inputs since they both follow the property of “the smaller the better” (Hailu and Veeman, 2001).

respectively defined as $R_i^- = \max\{x_{ki}, k=1,\cdots,K\} - \min\{x_{ki}, k=1,\cdots,K\}$ and $R_r^+ = \max\{y_{kr}, k=1,\cdots,K\} - \min\{y_{kr}, k=1,\cdots,K\}$.

Eq. (7) belongs to the family of additive DEA models.[4] Its objective function, often referred to as range adjusted inefficiency measure, represents the average of slacks-based inefficiency measure for entity o. The constraints determine the best practice frontier from which the maximally potential reduction and expansion in inputs and outputs are identified. When the range for a variable is zero, it indicates that all the entities have the same value for the variable so that the variable may be excluded in environmental performance evaluation. In that circumstance, the relevant component in the objective function of Eq. (7) and the corresponding constraint need to be removed. The last constraint $\sum_{k=1}^{K} z_k = 1$ as a convexity condition guarantees that the measurement units of ratio-scale variables will not change the optimal solution (Cooper et al., 1999). For any admissible transformation of the original variables, i.e. $f_i(v_{ki}) = \alpha_i v_{ki} + \beta_i$, the focus on slacks (or gaps) accommodates the shift parameter (α_i) and the scaling factor (β_i) is handled by means of range adjustment.

Eq. (7) as a simple linear programming model can be easily solved by any linear programming software package. Once the optimal solution to Eq. (7) is derived, we can define a CEI as

[4] As a slacks-based DEA model, Eq. (7) has a close relationship with the non-radial directional distance function (DDF) that has gained much popularity in efficiency and productivity analysis (Chambers et al., 1996; Zhou et al., 2012). In environmental economics, DDF has been widely used to assess environmental performance and the impact of environmental regulation. See, for example, Boyd and McClelland (1999), Hoang and Coelli (2011) and Picazo-Tadeo et al. (2014).

$$CEI(V_o) = CEI(X_o, Y_o) = 1 - \frac{1}{m+s}\left(\sum_{i=1}^{m}\frac{s_i^{*-}}{R_i^-} + \sum_{r=1}^{s}\frac{s_r^{*+}}{R_r^+}\right) \quad \forall o \in \{1, \cdots, K\} \qquad (8)$$

where * denotes the corresponding optimal slack variable. It can be shown that a C*EI* derived from Eq. (8) satisfies the following properties (Cooper et al., 1999):

P1. $0 \le CEI(V_o) \le 1$.

P2. $CEI(V_o) = 1 \Leftrightarrow$ Entity o is located on the best practice frontier; $CEI(V_o) < 1 \Leftrightarrow$ Entity o is not located on the best practice frontier and can be improved in certain dimensions.

P3. $CEI(V_o)$ is invariant to the measurement units of inputs and outputs.

P4. $CEI(V_o)$ is strongly monotonic.

P5. $CEI(V_o)$ is translation invariant.

P1 indicates that Eq. (8) yields a standardized index lying between zero and one, and a larger index value is linked to better environmental performance. P2 implies that the entities that do not play a role in constructing the best practice frontier have index values less than unity. From Eq. (7), we can easily identify the entities forming the best practice frontier as those associated with nonzero z_k. P3 indicates that the index is invariant with the measurement units of ratio-scale variables. The implication of P4 is that a reduction in any input or an increase in any output leads to an increase in the index value. P5 means that additions and subtractions of constants by any variables will not affect the index value, which is particularly useful when some interval-scale indicators are involved in constructing CEIs.[5] Combining P3 to P5, we

[5] It should be pointed out that the range adjusted DEA model, i.e. Eq. (7), is not the only choice for generating a CEI satisfying P1 to P5. For example, the bounded adjusted DEA model proposed by Cooper et al. (2011) may also be used to construct a cardinally meaningful CEI, which might be worth further investigating in future research.

have

Proposition 2. *The CEIs derived from Eqs. (7) and (8) are cardinally meaningful, i.e.* $CEI(V_k) = CEI(F(V_k)) \quad \forall k \in \{1, \cdots, K\}$.

Proposition 3. If $X_k \le X_l$, $Y_k \ge Y_l$ and there is at least one i (s) such that $x_{ki} < x_{li}$ ($y_{ks} > y_{ls}$), then $CEI(V_k) > CEI(V_l)$.

Propositions 2 implies that the CEIs derived from Eqs. (7) and (8) can easily handle data irregularity issues such as multiple zero entries and mixed measurability of the underlying variables. The property of cardinal meaningfulness also facilitates the computation of CEIs when other data irregularity issues exist. In the case of the TRI dataset, the range of values for certain variables can be rather large, which may pose challenges in solving linear programming models due to computer rounding errors. Owing to the properties of unit and translation invariance of Eq. (7), we may rescale the variables to remove the zero values and force the variables to be comparable. Despite the advantages, a concern (and possible weakness) of the use of DEA to construct CEIs is that multiple entities may have index values of unity preventing them from being compared with each other. This, however, also indicates that each of these entities has its particular strengths in certain dimensions, allowing them to serve as benchmarks for similar entities.

3.2 Linkage with arithmetic mean aggregation

The derivation of CEIs by Eqs. (7) and (8) requires solving a series of linear programming models. However, when there is a “super-entity” dominating all the other entities in all dimensions, we may directly derive the CEIs without solving linear programming models. In this circumstance, other entities will automatically identify the "super-entity" as their benchmark, and the optimal slack in a variable for other entities will be equal to their distances from the "super-entity". Mathematically, the CEI can be derived by

$$CEI(V_k) = 1 - \frac{1}{m+s}\left(\sum_{i=1}^{m} \frac{x_{ki} - \min_k\{x_{ki}\}}{R_i^-} + \sum_{r=1}^{s} \frac{\max_k\{y_{kr}\} - y_{kr}}{R_r^+} \right)$$

$$= \sum_{i=1}^{m} \frac{1}{m+s}\left(\frac{\max_k\{x_{ki}\} - x_{ki}}{\max_k\{x_{ki}\} - \min_k\{x_{ki}\}} \right) + \sum_{r=1}^{s} \frac{1}{m+s}\left(\frac{y_{kr} - \min_k\{y_{ki}\}}{\max_k\{y_{kr}\} - \min_k\{x_{kr}\}} \right) \tag{9}$$

Eq. (9) is a weighted sum of the normalized variables for all the entities for which a linear min-max normalization scheme is adopted. It suggests that the weighted sum method could lead to a meaningful index when there exists a "super-entity". However, in reality a "super-entity" is unlikely to exist when multi-dimensional environmental performance is concerned. One may imagine that a "super-entity" could be artificially generated by taking the highest values for the outputs and the lowest values for the inputs. Indeed, this practice makes the use of Eq. (9) feasible, which simplifies the computation of CEIs. However, as Munda and Nardo (2009) discussed, the weighted sum aggregation rule assumes full compensability between different variables, implying that the variables are completely substitutable with each other. Since different dimensions cannot be fully substituted with each other, the assumption might not be appropriate for scientifically assessing environmental performance (Munda and Nardo, 2009). In the range adjusted DEA model, the substitution between the optimal slacks for different variables is allowed when there exist multiple optimal solutions.

However, this kind of substitution indicates that an entity may have multiple choices to reach the best practice frontier. Different optimal solutions only imply different pathways while the ultimate goal is common – improving environmental performance!

4. Empirical study

4.1 Background

The public availability of the TRI database allows stakeholders as well as researchers to make comparisons of environmental performance across and between firms/plants over time for different purposes (Khanna et al., 1998). Prior studies using the TRI database have specified a variety of environmental indicators, but there is no consensus on which indicator represents an ideal proxy for the measurement of environmental performance (Toffel and Marshall, 2004). The environmental indicators used include aggregate toxic releases (Bui and Kapon, 2012), toxic releases weighted by toxicity factors (Cole et al., 2013), on-site toxic releases and off-site transfers (Khanna and Damon, 1999), the ratio of toxic releases to net sales (Konar and Cohen, 1997), and the toxic releases adjusted by distance (Hanna, 2007). Often, various indicators have been separately used, while it has been argued that the measurement of environmental performance based on TRI data needs to consider not only toxic releases but also other indicators such as revenue and toxicity factors (Gerde and Logsdon, 2001).

Scholars have shown that providing facility-level specific information allows greater transparency and can influence pollution abatement positively (Konar and

Cohen, 1997). If environmental performance is evaluated only at the firm level, rather than the facility level, some facilities of a parent company performing above or below average might not be recognized depending on the environmental regulations of the state in which they are located. Despite the availability of this facility specific information, only a few studies, such as Färe et al. (2010) and Bui and Kapon (2012), assessed facility level environmental performance. Specifically, the interesting study by Färe et al. (2010) used Malmquist quantity index and DEA to develop a formal environmental performance index for assessing the performance of coal-fired power plants in releasing toxic chemicals. In this section, we shall employ the nonparametric methodology described in Section 3 to construct a meaningful and standardized CEI for assessing the facility-level environmental performance in toxic releases in Los Angeles County. The empirical analysis not only demonstrates the robustness of the CEI but also shows how the CEI can provide perspectives on the improvement of facility-level environmental performance.

4.2 Data description

Our analysis is based on 150 facilities from three major industries - Primary Metals, Fabricated Metals and Chemicals, which respectively have 29, 54 and 67 facilities and as a whole accounted for 59% of the total toxic releases reported to the TRI database in Los Angeles County for 2012 (Delmas and Kohli, 2014). As pointed out by Delmas and Blass (2010), it is inappropriate to compare the environmental performance of firms or plants from different industries due to their different operating characteristics. As such, the facilities are evaluated and compared with those in the same industry. Since the sample size varies across different industries, our analysis may also shed some insights on how the construction of the CEI is affected by sample size.

Our CEI is derived from four variables, which represent a facility's effort in generating revenue while simultaneously preventing toxics releases into the natural environment. The first variable is the *Quantity of Total Toxic Releases* (QTTR), which includes both on-site and off-site releases to the environment but excludes the toxic releases arising from catastrophic and extreme events. The second is the *Toxicity of Total On-site Toxic Releases* into the atmosphere (TTTR), which is the sum of chemical-specific toxic releases weighted by their corresponding toxicity factors. TTTR accounts for the varying toxicity of chemicals releases and is valuable in measuring the local health-related impacts of different facilities, which cannot be captured by QTTR alone. The third variable is referred to as the *Percentage of Waste Managed through Recycling, Energy Recovery and Treatment* (PWM), which is the ratio of waste managed through recycling, energy recovery and treatment to the total waste including released and managed waste. The forth variable is the *Gross Revenue* (GR), which is a financial indicator that highlights each facility's ability in generating revenue given a certain amount of toxic releases. Of the four variables, QTTR and TTTR are used as inputs while PWM and GR are used as outputs, following the scheme that QTTR and TTTR are "the smaller the better" and PWM and GR are "the larger the better." Table 2 lists the summary statistics of the four variables by industry. A detailed description of these and additional TRI variables as well as data sources can be found in Delmas and Kohli (2014).

Table 2 Summary statistics of four variables by industry

Industry		**Quantity of Total Releases (in pounds)**	**Toxicity of Total On-site Toxic Releases (in pounds-toxicity)**	**Percentage of Waste Managed (%)**	**Gross Revenue (in 10^6 US$)**
Chemicals (67 facilities)	Mean	5500	427707	61	79.64
	Std. Dev.	10697	1706483	44	213.21
	Min	0	0	0	0.09
	Max	55170	13616910	100	1310.00
Primary Metals (29 facilities)	Mean	109161	55936001	58	44.65
	Std. Dev.	533393	292429498	48	76.79
	Min	0	0	0	1.14
	Max	2823311	1548000000	100	319.44
Fabricated Metals (54 facilities)	Mean	21847	293951	79	44.65
	Std. Dev.	85782	1117620	36	62.78
	Min	0	0	0	0.15
	Max	496159	6475100	100	377.91

As shown in Table 2, the minimum values of QTTR, TTTR and PWM are zero for all the three industries and the ranges for the first two variables are extremely large. This may give rise to certain computational problems due to computer rounding errors if DEA models are directly solved. Fortunately, the range adjusted DEA models given by Eq. (7) are not affected by any linear transformations of the variables, which facilitate the calculation of our CEIs. For comparison purposes, we also compute the CEI values by using the weighted sum aggregation rule cum min-max linear normalization, i.e. Eq. (9). While a geometric aggregation can lead to a meaningful CEI given the fact that four ratio-scale variables are aggregated, we do not use the aggregation rule due to the existence of multiple zeros in the dataset.

4.3 Main results and discussions

Table 3 shows the summary statistics of the CEI values calculated from the nonparametric frontier approach, i.e. Eqs. (7)-(8), and the weighted sum aggregation rule cum min-max linear normalization, i.e. Eq. (9). In terms of variance, the two sets of CEIs are quite close to each other for the sectors of Primary Metals and Fabricated Metals, while for the Chemicals sector the CEI obtained from the nonparametric frontier approach showed a slightly larger variance than that from the weighted sum aggregation rule. In addition, the ranges of CEI values from the two methods are also very close to each other.

Table 3 Summary statistics of the CEIs by two aggregation methods

Industry		**Nonparametric frontier approach**	**Weighted sum aggregation cum min-max normalization**
Chemicals	Mean	0.75	0.63
	Std. Dev.	0.18	0.13
	Median	0.72	0.68
	Min	0.22	0.22
	Max	1.00	0.99
Primary Metals	Mean	0.84	0.66
	Std. Dev.	0.20	0.18
	Median	0.96	0.75
	Min	0.27	0.26
	Max	1.00	0.98
Fabricated Metals	Mean	0.83	0.70
	Std. Dev.	0.12	0.12
	Median	0.88	0.75
	Min	0.52	0.40
	Max	1.00	0.91

Four hypotheses are proposed and tested to investigate whether there exist significant differences in the CEIs computed by the two aggregation methods and for the three industries when only the nonparametric frontier approach is employed. The proposed null hypotheses are described as follows:

(1) The choice between the nonparametric frontier approach and the weighted sum aggregation does not affect the CEIs;

(2) The chemicals sector has the same environmental performance as primary metals sector in toxic releases;

(3) The primary metals sector has the same environmental performance as fabricated metals sector in toxic releases;

(4) The chemicals sector has the same environmental performance as the fabricated metals sector.

Since the two sets of CEI values as well as the differences derived do not follow a normal distribution, we employ the commonly used Wilcoxon-Mann–Whitney rank-sum-test to test the four hypotheses. To test hypothesis (1), the CEI values for the three sectors are separately used, which leads to three sets of testing results. For testing hypotheses (2) to (4), we only use the CEI values computed by the nonparametric frontier approach.

Table 4 shows the results of the hypothesis tests. It can be observed that the three null hypotheses for comparing two different aggregation methods are all rejected at the 0.01 level of significance implying that the nonparametric frontier approach yields larger CEI values than the weighted sum aggregation rule. In addition, the CEI results obtained from the nonparametric frontier approach suggest that the primary metals industry might show better environmental performance than the chemicals industry.

However, there is no statistical evidence for rejecting the last two hypotheses at the 0.01 level of significance, which indicates the differences between the chemicals and fabricated metals industries as well as between the primary metals and fabricated metals industries in environmental performance are not significant.[6]

Table 4 Summary of hypothesis test results

Null hypothesis	**Mann-Whitney U**	***p*-value**
H_{01a}: Mean($CEI_{Ch\text{-}RAM}$)=Mean($CEI_{Ch\text{-}WS}$)	5225	0.0018
H_{01b}: Mean($CEI_{PM\text{-}RAM}$)=Mean($CEI_{PM\text{-}WS}$)	1062	0.0013
H_{01c}: Mean($CEI_{FM\text{-}RAM}$)=Mean($CEI_{FM\text{-}WS}$)	3839	0.0000
H_{02}: Mean(CEI_{Ch})=Mean(CEI_{PM})	1752	0.0059
H_{03}: Mean(CEI_{PM})=Mean(CEI_{FM})	1451	0.0263
H_{04}: Mean(CEI_{Ch})=Mean(CEI_{FM})	3650	0.0638

We also investigate the correlation between the CEIs derived from the two different aggregation rules. The Pearson and Spearman rank correlation coefficients obtained are shown in Table 5. There exist significant positive correlations between the CEI values derived from the two alternative aggregation rules. In particular, we find that both the Pearson and Spearman rank correlation coefficients of the two sets of CEIs for Fabricated Metals sector are as high as 0.998, which might be an indication of the robustness of the rankings to the choice of aggregation rule for different facilities in this sector.

Further comparisons between the two aggregation rules in deriving CEIs may be

[6] It is worth pointing out that the environmental performance comparisons between different groups of facilities should be performed with caution when the CEIs are constructed by the nonparametric frontier approach. Since the CEI values are relative and not absolute ones, another possible reason for the between-group differences might be that the facilities in an industry are closer – on average – to the frontier for the industry compared to the facilities in another industry.

conducted by looking through facility-level CEI values. When the nonparametric frontier approach is used, more than one facility will achieve a CEI value of unity unless there exists a super-facility dominating all the other entities in all the dimensions. While the weighted sum aggregation usually has higher discriminating power, the CEI values derived from the method could sometimes be misleading since the full compensability between all the variables is implicitly assumed. To examine this point, in Table 6 we summarize the CEI values from the weighted sum aggregation as well as the original data for the facilities with a CEI value of unity using the nonparametric frontier approach.

Table 5 Correlation coefficients of CEIs derived from two aggregation methods

Correlation type	Chemicals	Primary Metals	Fabricated Metals
Pearson	0.782*	0.913*	0.998*
Spearman	0.791*	0.902*	0.998*

*: Correlated at the 1% significance level

Table 6 CEIs from weighted sum aggregation for the facilities with unity values by the nonparametric frontier approach

Industry	Facility	Quantity of Total Releases	Toxicity of Total On-site Toxic Releases	Percentage of Waste Managed	Gross Revenue	CEI (Weighted sum)
Chemicals	No. 27	1085	488	99	1310	0.99
	No. 60	1	1760	0	93	0.52
	No. 63	0	0	100	91	0.77
Primary Metals	No. 7	128	17600	100	54	0.79
	No.12	21	27896	67	129	0.77
	No.13	868	295434	93	319	0.98
	No.16	541	565800	87	266	0.92
	No. 23	0	0	100	49	0.79
	No. 24	5	37460	0	86	0.57

	No. 29	15823	73472	99	102	0.83
Fabricated	No. 10					0.91
Metals		16200	1592210	93	378	
	No. 54	0	0	100	178	0.87

It can be observed from Table 6 that three facilities from the chemical industry and two facilities from the fabricated industries are located on the best practice frontiers. For the primary metals industry, however, seven facilities constitute its best practice frontier. The variation should mainly be attributed to the differences in the numbers of facilities in the three sectors. While the chemicals and fabricated metals industries respectively consist of 67 and 54 facilities, the primary metals industry has only 29 facilities. Although the higher discriminating power of weighted sum aggregation rule in constructing CEIs is insensitive to sample size, the meaningfulness of the CEIs based on this rule is questionable. For example, in the chemicals sector facility no. 63 had no toxic releases, which should be an indication of better environmental performance. However, since its gross revenue is substantially less than that of facility no. 27, the CEI value of facility no. 63 is much smaller than that of facility no. 27 although the latter produced 1085 pounds of toxic releases. The same cases occur for the facility no. 23 in the primary metals sector and the facility no. 54 in the fabricated metals sectors. In evaluating the facility-level environmental performance, it is logical to reward the facilities producing higher revenues with the same impact of toxic releases by giving them higher CEI scores. But it may not be reasonable to give better evaluation of a facility with relatively higher revenue and toxic releases than another facility without any toxic releases. Compared to the weighted sum aggregation rule, the nonparametric frontier approach treats the facilities with distinctive strengths in various dimensions indifferently by giving them the same CEI values of unity, which seems to be more reasonable for an appropriate environmental performance assessment.

As described in Section 3, an additional strength of the nonparametric frontier approach is that it may help each facility identify its benchmark (groups) as well as the directions of potential improvement in different dimensions. Take the variable QTTR as an example. We can use Eq. (7) to compute the optimal slack in QTTR for each facility within the three sectors, which indicates that facility's potential reduction in QTTR to reach the level of its benchmark. Fig. 1 shows the sectoral potential reduction in QTTR as a percentage of the actual QTTR for each of the three sectors. It is observed that almost 90% of QTTR for Chemicals sector could be reduced if all the facilities reach the levels of their benchmarks. For primal metals and fabricated Metals sectors, the potential reductions in QTTR are more than 95% of the actual quantities! This result might be explained by the fact that there are some facilities with poor environmental performance and very high QTTR. For example, facility no. 21 in the primary metals sector could reduce 2,822,677 pounds of toxic releases with reference to a convex combination of facility no. 13 and no. 23, which accounts for 97% of potential reductions in QTTR for the whole sector. While it shows a huge potential in reducing toxic releases, this might be unrealistic due to the scale discrepancy between facility no. 21 and its benchmark group. Nevertheless, it at least offers a direction along which facility no. 21 may improve its environmental performance through managerial efforts.

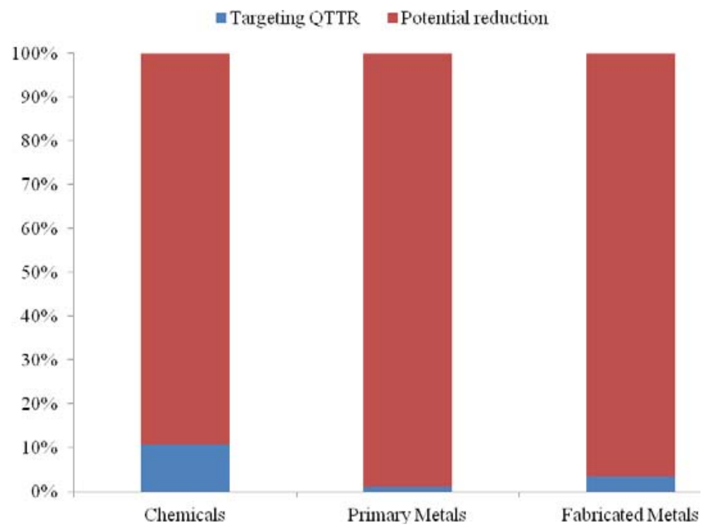

Fig. 1 Percentages of potential reductions in QTTR for three industries

Table 7 summarizes the appearance frequencies of the facilities forming the best practice frontiers of the three industries. It can be seen that facility no. 27 in the Chemicals, no. 23 in the Primary Metals and no. 54 in the Fabricated Metals are most frequently identified as the benchmarks. Referring to Table 6, we find that the latter two are indeed the best performers in toxic releases with reasonable revenues. In terms of facility no. 27 in the chemical industry, 99% of those toxic releases were properly handled. Meanwhile, this facility generates a gross revenue that is over ten times the revenue of the other two facilities forming the best practice frontier of the Chemicals industry. In view of these features, it is not surprising that facility no. 27 has been identified as a benchmark most frequently. On the contrary, although facility no. 60 in the chemicals industry and nos. 12 & 24 in the primary metals industry also have CEI values of unity, they are not used to evaluate any other facilities except themselves. It implies that the three facilities, which did not perform well in managing toxic releases, cannot be dominated by any convex combination of other facilities, and therefore should not be set as the benchmarks in environmental performance assessment.

Table 7 Appearance frequencies of the facilities forming the best practice frontiers

Industry	Facility no.	Frequency of appearance in the best practice frontier
Chemicals	**#27**	**53**
	#60	*1*
	#63	35
Primary Metals	#7	5
	#12	*1*
	#13	13
	#16	9
	#23	**19**
	#24	*1*
	#29	5
Fabricated Metals	#10	30
	#54	**52**

4.4 Sensitivity analysis

Our CEI is derived through solving a series of linear programming models and may be affected by some uncertainty factors. First, the best practice frontier, formed by the existing facilities, is an estimate of the "true" frontier, which might be subject to uncertainty arising from the sampling variation of the obtained frontier. Although the uncertainty can be handled by using bootstrap methods for assessing the sampling variation (Simlar and Wilson, 2015), it should be noted that the best practice frontier in the context of constructing CEIs is somewhat different from the production frontier in efficiency and productivity analysis. One main usage of our CEI is to conduct cross-sectional comparison or monitor the environmental performance over time. Such an application context allows us to use the observations from all the comparable entities at different time points to form the best practice frontier and construct CEIs.

As such, the uncertainty due to the sampling variation of the obtained frontier will not be studied in this paper.

Data accuracy is another important source of uncertainty. In the case of TRI data, as pointed out by Toffel and Marshall (2004), the uncertainty in data accuracy made the development of environmental performance metrics difficult. While the sensitivity of the nonparametric frontier approach with respect to data perturbation could be theoretically examined, in this paper we conduct a sensitivity analysis of the CEIs by artificially changing the observations in a random way. It is assumed that the data errors for all the observations are within ±10% of the data observed. By generating random numbers within [-10%, 10%], we create 50 datasets for each of the three industries based on which 50 sets of CEIs can be derived. Using the 50 sets of CEIs, we first compute the average CEI values as well the corresponding standard derivations for each of three industries. Fig. 2 shows the box plots of the averages and standard deviations of CEI values by industry.

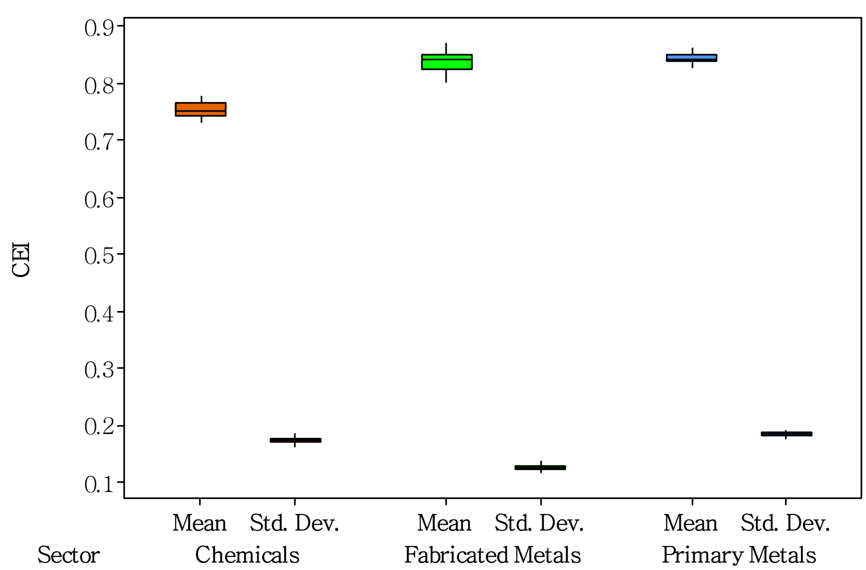

Fig. 2 Boxplots of the average and standard deviation values of CEIs by industry

It can be observed from Fig. 2 that the impact of data uncertainty in the specified range on the average CEI value of each industry is relatively weak, especially for the case of the primary metals sector. Meanwhile, data variation has very little impact on the standard derivation of the sectoral average CEI values. It suggests that the sectoral average CEI values are quite insensitive to the uncertainty in data accuracy (within the specified range). To investigate whether the dispersion of CEI values for all the facilities in each of the three industries varies significantly when the uncertainty in data accuracy is considered, we show the box plots of the average CEI values from the simulated data for all the facilities in each of the three industries in Fig. 3. For comparison purposes, the box plots of the CEI values derived from the original dataset are also provided. It can be seen from Fig. 3 that there are few changes in the median and ranges of CEI values for the chemical and fabricated metals industries. However, the change in the median of CEI values for the primary metals industry seems to be slightly larger, which could be due to the relatively small number of facilities in this industry. It might be an indication that CEI values are insensitive to the uncertainty in data accuracy when the sample size is relatively large.

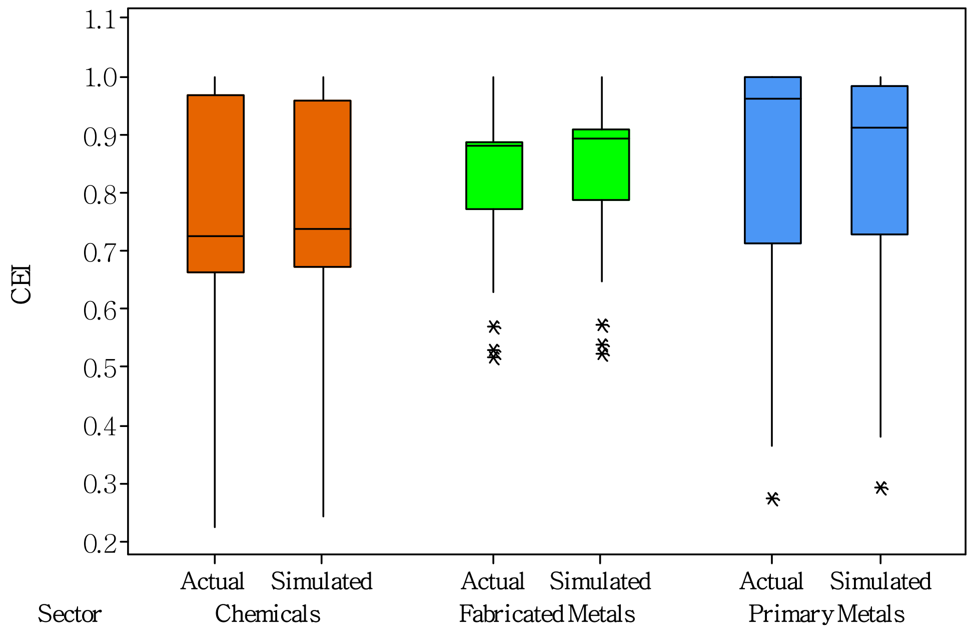

Fig. 3 Boxplots of the CEI values for all the facilities derived from both actual and simulated data

As the uncertainty in data accuracy has a relatively larger impact on the CEI values of facilities in the primary metals industry, it is worthwhile looking through the types of facilities that would be more easily affected in their CEI values. Fig. 4 shows the CEI value of each facility in the primary metals industry as well as the corresponding average CEI value derived from the simulated data. It is found that most of the facilities have small changes in CEI when data variation exists. In particular, five facilities, i.e. nos. 12, 13, 16, 23 and 24, are always located at the best practice frontier which might be an indication of the robustness of the best practice frontier with respect to the uncertainty in data accuracy. However, several facilities such as nos. 8, 11 and 20 show relatively larger gaps between the actual CEI value and the CEI value from simulated data.

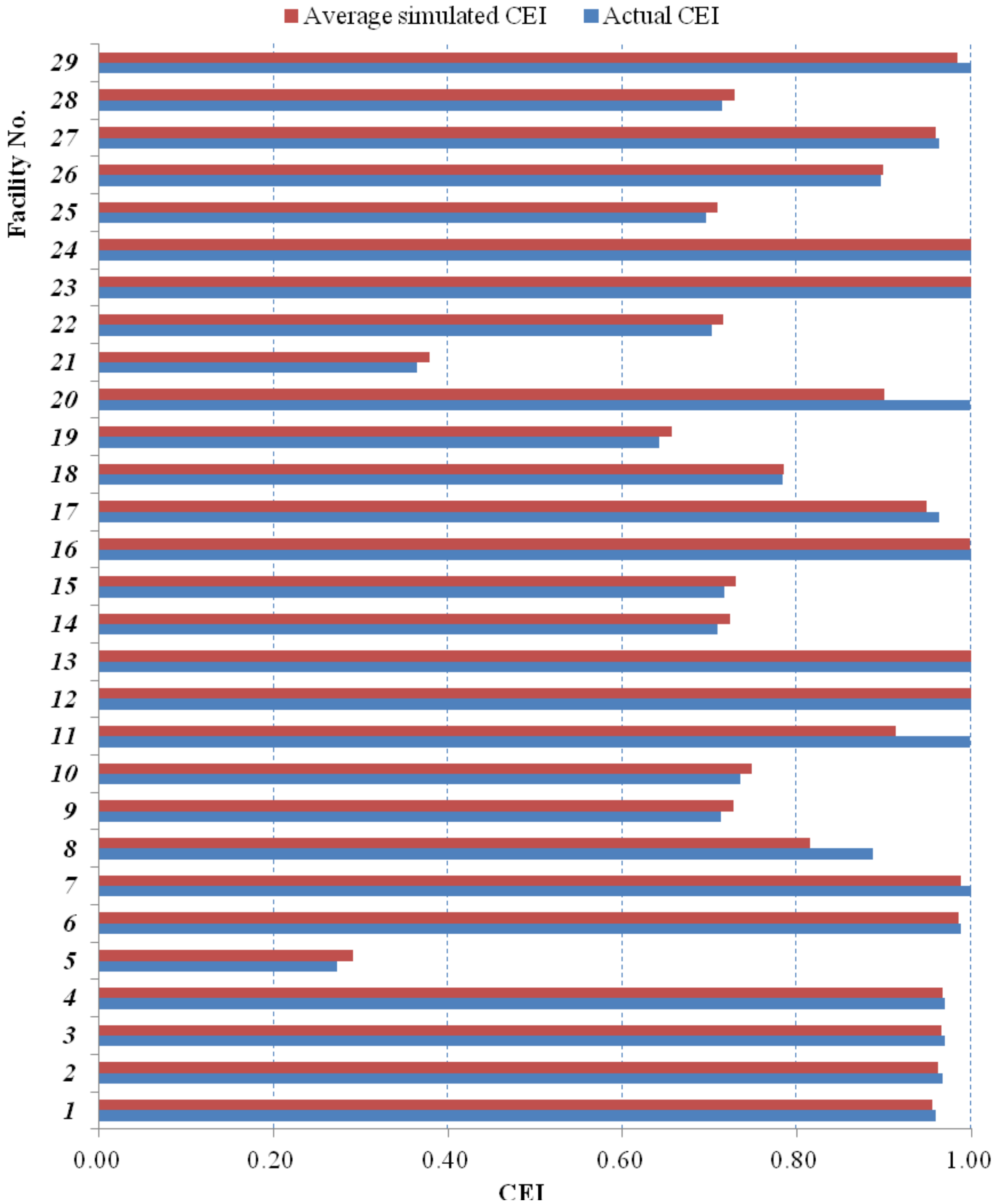

Fig. 5 Comparison between the CEIs from original and simulated datasets for the facilities in the Primary Metals industry

5. Conclusions

This paper argues that it is important to construct a cardinally meaningful and standardized CEI for the measurement of environmental performance. A CEI is said to be cardinally meaningful if its values are invariant with respect to the changes in the measurement units of underlying variables. This concept is particularly important when the cardinality characteristics of CEIs are concerned. The commonly used aggregation methods, e.g. arithmetic and geometric aggregation methods, cannot yield a cardinally meaningful CEI when mixed measurability of underlying variables is involved. We propose to use a nonparametric frontier approach, i.e. range adjusted DEA model, to construct a cardinally meaningful CEI, which can easily handle the issues of mixed measurability of underlying variables and data irregularity such as the existence of multiple zeros.

We apply the nonparametric frontier approach to constructing a CEI for evaluating the facility-level environmental performance of toxic releases in three industries (i.e. chemical, primary metals and fabricated metals) in Los Angeles County based on the latest TRI data. At the industry level, we find that the primary metals industry shows better environmental performance than the chemical industry while other pairwise comparisons do not show statistically significant differences. In addition, we summarize the benchmark facilities in every industry as well as their appearance frequency in forming best practice frontiers, which represent targets for other facilities to improve their environmental management practices. Finally, we investigate whether the uncertainty in data accuracy has a significant effect on the CEI results. Our results show that the distributions of the CEI values change very little when confronted with data errors of 10%, which might be an indication of the robustness of our CEI in evaluating facility-level environmental performance.

References

Aida, K., Cooper, W.W. Pastor, J.T., Sueyoshi, T., 1998. Evaluating water supply services in Japan with RAM: A range-adjusted measure of inefficiency. Omega 26, 207-232.

Böhringer, C., Jochem, P.E.P., 2007. Measuring the immeasurable – A survey of sustainability indices. Ecological Economics 63, 1-8.

Boyd, G.A., McClelland, J.D., 1999. The impact of environmental constraints on productivity improvement in integrated paper plants. Journal of Environmental Economics and Management 38, 121-142.

Bui, L.T.M., Kapon, S., 2012. The impact of voluntary programs on polluting behavior: Evidence from pollution prevention programs and toxic releases. Journal of Environmental Economics and Management 64, 31-44.

Chambers, R.G., Chung, Y., Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory 70, 407-419.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., Battese, G.E., 2005. An Introduction to Efficiency and Productivity Analysis (2nd ed.). Springer.

Cooper, W.W., Park, K.S., Pastor, J.T., 1999. RAM: A range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. Journal of Productivity Analysis 11, 5-42.

Cooper, W.W., Pastor, J.T., Borras, F., Aparicio, J., Pastor, D., 2011. BAM: A bounded adjusted measure of efficiency for use with bounded additive models. Journal of Productivity Analysis 35, 85-94.

Cole, M.A., Elliott, R.J.R., Khemmarat, K., 2013. Local exposure to toxic releases: examining the role of ethnic fractionalization and polarization. Ecological Economics 93, 249-259.

Delmas, M.A., Blass, V.D., 2010. Measuring corporate environmental performance: The trade-offs of sustainability rating. Business Strategy and the Environment 19, 245-260.

Delmas, M.A., Kohli, A., 2014. UCLA-US EPA TRI university challenge: Assessing and communicating the environmental impact of TRI facilities in Los Angeles County. Project report, UCLA Institute of the Environment and Sustainability.

Ebert, U., Welsch, H., 2004. Meaningful environmental indices: A social choice approach. Journal of Environmental Economics and Management 47, 270-283.

Färe, R., Grosskopf, S., Hernandez-Sancho, F., 2004. Environmental performance: An index number approach. Resource and Energy Economics 26, 343-352.

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

Färe, R., Grosskopf, S., Pasurka, C., 2006. Social responsibility: U.S. power plants 1985-1998. Journal of Productivity Analysis 26, 259-267.

Färe, R., Grosskopf, S., Pasurka, C., 2007. Environmental production functions and environmental directional distance functions. Energy 32, 1055-1066.

Färe, R., Grosskopf, S., Pasurka, C.A., 2010. Toxic releases: An environmental performance index for coal-fired power plants. Energy Economics 32, 158-165.

Färe, R., Grosskopf, S., Tyteca, D., 1996. An activity analysis model of the environmental performance of firms – application to fossil-fuel-fired electric utilities. Ecological Economics 18, 161-175.

Gerde, V.W., Logsdon, J.H., 2001. Measuring environmental performance: Use of the toxics releases inventory (TRI) and other US environmental databases. Business Strategy and the Environment 10, 269-285.

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, 605-616.

Hanna, B.G., 2007. House values, incomes, and industrial pollution. Journal of Environmental Economics and Management 54, 100-112.

Hoang, V.N., Coelli, T., 2011. Measurement of agricultural total factor productivity growth incorporating environmental factors: A nutrients balance approach. Journal of Environmental Economics and Management 62, 462-474.

Khanna, M., Damon, L.A., 1999. EPA's voluntary 33/50 program: Impact on toxic releases and economic performance of firms. Journal of Environmental Economics and Management 37, 1-25.

Khanna, M., Quimio, W.R.H., Bojilova, D., 1998. Toxics release information: A policy tool for environmental protection. Journal of Environmental Economics and Management 36, 243-266.

Koehler, D.A., Spengler, J.D., 2007. The toxic release inventory: Fact or fiction? A case study of the primary aluminum industry. Journal of Environmental Management 85, 296-307.

Konar, S., Cohen, M.A., 1997. Information as regulation: The effect of community right to know laws on toxic emissions. Journal of Environmental Economics and Management 32, 109-124.

Munda, G., Nardo, M., 2009. Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. Applied Economics 41, 1513-1523.

Picazo-Tadeo, A.J., Castillo-Giménez, J., Beltrán-Esteve, M., 2014. An intertemporal approach to measuring environmental performance with directional distance functions: Greenhouse gas emissions in the European Union. Ecological Economics 100, 173-182.

Pittman, R.W., 1983. Multilateral productivity comparisons with undesirable outputs. Economics Journal 93, 883-891.

Pollesch, N.L., Dale, V.H., 2016. Normalization in sustainability assessment: Methods and implications. Ecological Economics 130, 195-208.

Simar, L., Wilson, P.W., 2015. Statistical approaches for non-parametric frontier models: A guided tour. International Statistical Review 83, 77-110.

Toffel, M.W., Marshall, J.D., 2004. Improving environmental performance assessment: A comparative analysis of weighting methods used to evaluate chemical release inventories. Journal of Industrial Ecology 8, 143-172.

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, 281-308.

Tyteca, D., 1997. Linear programming models for the measurement of environmental performance of firms – Concepts and empirical analysis. Journal of Productivity

Analysis 8, 183-197.

Welsch, H., 2005. Constructing meaningful sustainability indices. In: C. Böhringer, A. Lange (eds.), Applied Research in Environmental Economics, Physica-Verlag, Heidelberg, pp. 7-22.

Zhou, P., Ang, B.W., 2009. Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. Social Indicators Research 94, 83-96.

Zhou, P., Ang, BW., Han, J.Y., 2010. Total factor carbon emission performance: A Malmquist index analysis. Energy Economics 32, 194-201.

Zhou, P., Ang, B.W., Poh, K.L., 2006. Comparing aggregating methods for constructing the environmental index: An objective measure. Ecological Economics 59, 305-311.

Zhou, P., Ang, B.W., Wang, H., 2012. Energy and CO_2 emission performance in electricity generation: A non-radial directional distance function approach. European Journal of Operational Research 221, 625-635.