Environmental performance assessment using the evidential reasoning approach: The case of logistics service providers

Panitas Sureeyatanapas^{1,*} and Korrapat Tawwan¹

¹Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, 123 Moo 16, Mittraphap Road, Khon Kaen 40002, Thailand

Abstract. The demand for environmental performance assessment is increasing among business practitioners, and it has nowadays become one of the key factors for a company's self-improvement as well as for selecting suppliers and logistics providers. The assessment is, in essence, a multiple criteria decision analysis (MCDA) problem comprised of many quantitative and qualitative criteria. Frequently, the assessment data of some criterion is inevitably imprecise and/or incomplete since the nature of environmental assessment relies heavily on professional and complex methods which might not be fully available for every company. Also, qualitative criteria can only be assessed based upon human judgment. This paper, therefore, proposes an application of the evidential reasoning (ER) approach to the assessment of environmental performance for logistics service providers. The lists of criteria and indicators are adapted from ISO 14031. The ER approach is able to logically aggregate all assessment information, although different forms of data (precise or imprecise; complete or incomplete) are obtained. For this paper, assessment data from two logistics companies were gathered and analysed to illustrate the implementation process. The results are in the form of aggregated belief distributions on a unified set of evaluation grades, and the company can use this information for performance improvement and benchmarking.

1 Introduction

Due to growing public awareness of environmental issues and stricter laws and regulations in many countries, product manufacturers nowadays intensively focus on auditing and assessing environmental performance of their suppliers, including logistics partners, in order to improve the overall supply chain performance [1]. Since transportation activities have been claimed as one of the major sources of global warming and the depletion of non-renewable energy, one of the emerging criteria for selecting and auditing suppliers is their abilities to preserve the environment [2-4]. The logistics industry, therefore, has slightly shifted from pricing competition to value-adding services in terms of green and sustainability capabilities [2, 5].

It is generally recognised that performance measurement is crucial for business success. The assessment helps promote continuous improvement, presents achievements of corporate goals of environmental protection, provides data for internal decision making, and ensures that firm's environmental performance meets or goes beyond legislative requirements, as such risks of legal fines and penalties are reduced [4, 6, 7]. ISO 14031 is one of the most well-known standards for environmental management that specifically focuses on performance evaluation. The growing popularity of ISO 14031 can be explained by a number of reasons. First of all, it can be simply adopted

by every single organisation without any limitations in terms of specific sizes and types of industry. Secondly, since it is a kind of standardised framework, comparison with others is possible. It also, thirdly, facilitates the communication of environmental performances in a systematic way as well as provides comprehensive information for policy decision-making [7, 8]. However, the application of ISO 14031 to the logistics industry is still limited among the literature.

Although ISO 14031 could be a choice for logistics firms to start assessing and reporting their environmental performance, a number of criticisms have been made. Firstly, most aspects of performance evaluation suggested in ISO 14031 are still multi-dimensional and qualitative, such as "Wastes" or "Community relations". That means the evaluation still relies on subjectivity and may vary according to personal interpretation. Secondly, since a large number of indicators are exemplified within the standard, it is likely that monitoring too many indicators simultaneously can distract practitioners from the main focuses of their organisations [1]. From this, a method to aggregate various indicators is needed in order for firms to be able to monitor the overall picture of their environmental performance. The obtained aggregated number is also open for business benchmarking in a credible way.

Another issue is that ISO 14031 does not provide a solution to deal with uncertainty and/or incompleteness of the assessment information. A number of studies in

* Corresponding author: panisu@kku.ac.th

green supply management implied that many firms feel uncomfortable when reporting information relating to environmental issues [9, 10]. This might be because such information relates to pollution and negative impacts upon their operations release into the environment. Furthermore, some firms may lack the capability to conduct a self-assessment in relation to some indicators, particularly the ones that require specific technology, laboratory, or skilled personnel. Moreover, firm's performances according to some indicators may fluctuate over time or vary by its branches or sites. In these cases, the average may not always be a good choice since it possibly causes the loss of meaningful information. For some qualitative aspects, in addition, the assessor may lack confidence in giving a precise judgement due to the unavailability or incompleteness of evidence. Although a set of evaluation grades or rating scales can be used to transform subjective opinions into numerical data, each assessor may interpret the meaning of each grade or scale differently. Also, they might not be confident when concluding that a situation being considered matches a particular grade, and one or more grades may better suit the actual practice. Lastly, among the various aspects or criteria in ISO 14031, practitioners within the same organisation, or within the same chain of benchmarking partners, may perceive the importance levels of those aspects differently. This may cause disagreement with regard to the aggregated score obtained. These issues support that the method used to determine the overall environmental performance needs the ability to solve the issues of incompleteness and uncertainty of the information in a sensible way. Furthermore, in the logistics industry, the lack of studies which focus on managers' attitudes toward the importance or the weight of each criterion still represents a gap in the literature.

The aim of this study is therefore to propose a logical method to assess and analyse environmental performance for logistics operations. The tool must be able to perform even when uncertainty and/or incompleteness of the assessment information, as well as uncertainty of the criteria weights, exist. To respond to these, the evidential reasoning approach (ER) is employed. ER can theoretically cope with the limitations of other MCDA methods generally used to determine a composite index. For example, the simple additive weighting (SAW) method which relies on the additive approach assumes that all indicators are preferentially independent [11, 12]. This assumption may not be realistic in conducting an environmental performance assessment. The aggregation based on the weighted geometric mean (WGM) method [13], on the other hand, does not allow for any compensation among indicators. That means the overall performance will be zero when at least one indicator has zero value. This assumption may not always be acceptable in business performance evaluation since a company's weakness in one indicator may be compensated by a very strong point in another, which helps to build a competitive advantage. For the data envelopment analysis (DEA) [14] and the technique for order preference by similarity to an ideal solution (TOPSIS) [15], the composite score or the overall performance for a company is determined by using a

comparison of the performance of other companies. They might be therefore applicable only for ranking alternatives but not for a single company's self-assessment [16]. SAW, WGM, DEA, or TOPSIS are also not designed to deal with qualitative indicators, nor any uncertainty of the data.

This paper is organised as follows: Following the introduction, ISO 14031 is briefly described in Section 2. Section 3 then explains the algorithms of the ER approach. Section 4 gives details of the research methodology. The applicability of the ER approach is then demonstrated using data from two logistics companies. The results and discussion are given in Section 5, and Section 6 is the conclusion.

2 ISO 14031

ISO 14031 is an international standard that provides a guideline for assessing and monitoring an organisation's environmental performance [1, 17]. It describes two dimensions of performance indicators: (i) Management performance indicators (MPIs) and (ii) Operational performance indicators (OPIs). MPIs indicate management capability and efforts to improve an organisation's environmental performances while OPIs provide direct measurements about such performances (e.g., air emission; the use of energy and water). MPIs are also linked to economic and social dimensions, such as indicators that show an organisation's capabilities to educate or encourage staff to follow environmental protection guidelines; costs and investments spent on this matter. MPIs are divided into four criteria: (i) Implementation of environmental policies and programmes, (ii) Regulatory compliance, (iii) Financial performance associated with environmental performance, and (iv) Community relations. OPIs, on the other hand, are broken down into nine criteria: (i) Materials, (ii) Energy, (iii) Services supporting organisation's operations, (iv) Physical facilities and equipment, (v) Supply and delivery, (vi) Products, (vii) Services provided by the organisation, (viii) Wastes, and (ix) Emissions [18]. For this study, however, the initial set of OPIs was adapted to suit the characteristics of the logistics industry. Finally, only seven criteria remain to represent OPIs, including (i) Materials, (ii) Energy, (iii) Services supporting organisation's operations, (iv) Physical facilities and equipment, (v) Products and services, (vi) Wastes, and (vii) Emissions.

As suggested by ISO 14031, it is not necessary for practitioners to include all criteria of MPIs and OPIs in their evaluation system. Each firm normally has its own interests or concerns that relate to the environment. To implement ISO 14031, for each firm, suitable indicators (and measurement units) under each criterion should be identified [18].

3 The evidential reasoning approach

The ER approach is one MCDA technique that has been applied to many areas of performance assessment and decision making, such as the self-assessment following

the European Foundation for Quality Management (EFQM) Model [19], the assessment of new product design [20], the assessment based upon the balanced scorecard [21], the evaluation and selection of research projects [22], and the prioritisation of energy sources [23].

The ER approach was firstly proposed by Yang and Singh [24] as a procedure to combine multiple criteria based upon the evidence theory that employs a belief distribution to describe the assessment. The algorithm developed in their article is based on the recursive approach that requires pairwise aggregations. Wang, et al. [25] then proposed an analytical ER algorithm to present an explicit aggregation function by combining all indicators simultaneously. The assessment of each indicator, e_i , is conducted towards a set of grades H , $H = \{H_1, H_2, \dots, H_n, \dots, H_N\}$, where H_{n+1} is preferred to H_n (H_1 and H_N denote the worst and the best grades, respectively). All grades must be defined to be mutually exclusive and collectively exhaustive [26].

Suppose that the total performance could be assessed towards L indicators, or a set of indicators e_i ($i=1, \dots, L$), the assessment result (S) for a company M in terms of each indicator e_i can be expressed in the form of a belief distribution, as shown below.

$$S(e_i(M)) = \{(H_n, \beta_{n,i}(M)), n = 1, \dots, N; (H, \beta_{H,i}(M))\}, \\ i = 1, \dots, L \quad (1)$$

$\beta_{n,i}(M)$ represents a belief degree in which a company M is assessed to a grade H_n when an indicator e_i is considered, where $\beta_{n,i}(M) \geq 0$ and $\sum_{n=1}^N \beta_{n,i}(M) \leq 1$. $\beta_{H,i}(M)$ is the degree of global ignorance or the portion of belief unassigned to subsets of H , $\beta_{H,i}(M) \geq 0$ and $\sum_{n=1}^N \beta_{n,i}(M) + \beta_{H,i}(M) = 1$ [26].

Wang, et al. [27] extended the ability of the ER approach to handle interval belief degree as a result of the inclusion of interval data from the assessment. Guo, et al. [28] then demonstrated a case in which interval weights of criteria were inserted into the ER algorithm. When such kinds of uncertainty exist, the belief distribution of company M towards the indicator e_i is also expressed in the interval form, as shown below.

$$S(e_i(M)) = \{(H_n, \beta_{n,i}(M) \in [\beta_{n,i}^-(M), \beta_{n,i}^+(M)]), n = \\ 1, \dots, N; (H, \beta_{H,i}(M) \in [\beta_{H,i}^-(M), \beta_{H,i}^+(M)])\}, i = 1, \dots, L \quad (2)$$

In response to a case in which interval belief degrees and interval weights are involved, the analytical ER algorithm is handled through the optimisation functions, as shown below. The next explanations of the ER algorithm follow Wang, et al. [27], Guo, et al. [28], and Sureeyatanapas, et al. [16]. Note that $n = 1, \dots, N$ and $i = 1, \dots, L$.

Objective functions:

$$\text{Max/Min } \beta_n(M) = \frac{m_n}{1 - \bar{m}_H} \quad (3)$$

$$\text{Max/Min } \beta_H(M) = \frac{\bar{m}_H}{1 - \bar{m}_H} \quad (4)$$

Subject to (s.t.):

$$m_n = k \left[\prod_{i=1}^L (m_{n,i} + \bar{m}_{H,i} + \tilde{m}_{H,i}) - \prod_{i=1}^L (\bar{m}_{H,i} + \tilde{m}_{H,i}) \right] \quad (5)$$

$$\tilde{m}_H = k \left[\prod_{i=1}^L (\bar{m}_{H,i} + \tilde{m}_{H,i}) - \prod_{i=1}^L \bar{m}_{H,i} \right] \quad (6)$$

$$\bar{m}_H = k \left[\prod_{i=1}^L \bar{m}_{H,i} \right] \quad (7)$$

$$k = \left[\sum_{n=1}^N \prod_{i=1}^L (m_{n,i} + \bar{m}_{H,i} + \tilde{m}_{H,i}) - (N - 1) \prod_{i=1}^L (\bar{m}_{H,i} + \tilde{m}_{H,i}) \right]^{-1} \quad (8)$$

$$m_{n,i} = w_i \beta_{n,i}(M) \quad (9)$$

$$\tilde{m}_{H,i} = w_i \beta_{H,i}(M) \quad (10)$$

$$\bar{m}_{H,i} = 1 - w_i \quad (11)$$

$$w_i^- \leq w_i \leq w_i^+ \quad (12)$$

$$\sum_{i=1}^L w_i = 1 \quad (13)$$

$$\beta_{n,i}^-(M) \leq \beta_{n,i}(M) \leq \beta_{n,i}^+(M) \quad (14)$$

$$\beta_{H,i}^-(M) \leq \beta_{H,i}(M) \leq \beta_{H,i}^+(M) \quad (15)$$

$$\sum_{n=1}^N \beta_{n,i}(M) + \beta_{H,i}(M) = 1 \quad (16)$$

$$\sum_{n=1}^N m_{n,i} + \bar{m}_{H,i} + \tilde{m}_{H,i} = 1 \quad (17)$$

where w_i represents the relative weight of the indicator e_i . The weights of all indicators must be normalised to sum to one; $m_{n,i}$ is the weighted belief which supports the assessment of M to the grade H_n based on the indicator e_i ; The probability mass unassigned to any grade after all N grades have been considered is a combination of two parts: $\bar{m}_{H,i}$ and $\tilde{m}_{H,i}$. $\bar{m}_{H,i}$ is the remaining probability mass caused by the fact that an indicator e_i only partly contributes to the assessment relative to its weight. $\bar{m}_{H,i}$ is equal to zero if e_i dominates the assessment or its relative weight is equal to one. $\tilde{m}_{H,i}$ is an unassigned probability mass resulting from the incompleteness of the assessment $S(e_i)$. $\tilde{m}_{H,i}$ is zero if the assessment is complete, or $\sum_{n=1}^N \beta_{n,i} = 1$.

For this model, two sets of variables, including the intervals for (i) criteria weights and (ii) belief degrees, are the inputs of the optimisation process. When the precise assessment is given for some indicators, the minimum and maximum values of the interval are equal to that precise value. After combining all L indicators, the total belief degrees, β_n and β_H , can be generated. β_H denotes the degree of belief unassigned to any single grade, or it reflects the total degree of incompleteness in the assessment. The aggregated result for each company M , or $S(y(M))$, can be described as follows.

$$S(y(M)) = \{(H_n, \beta_n(M) \in [\beta_n^-(M), \beta_n^+(M)]); (H, \beta_H(M) \in [\beta_H^-(M), \beta_H^+(M)])\} \\ \text{with } \sum_{n=1}^N \beta_n(a_i) + \beta_H(a_i) = 1 \quad (18)$$

The number of nonlinear optimisation models required to aggregate belief degrees of each company is equal to $2(N+1)$ for one combination process. The obtained result shows a belief distribution presenting a panoramic view of the overall performance. However, if the aim of the assessment is for ranking or comparing alternatives, the aggregated belief distribution cannot straightforwardly support those purposes. The determination of the expected utilities is then suggested. Utility reflects the assessor's preference for the value of the indicator being considered [12]. The utility of a grade H_n , or $u(H_n)$, could be between zero to one, and $u(H_{n+1}) > u(H_n)$. The expected utility of M can be determined, without any kind of uncertainty, through equation (19).

$$u(\mathcal{S}(y(M))) = \sum_{n=1}^N \beta_n(a_l)u(H_n) \quad (19)$$

Nevertheless, the expected utility of M will be in the form of an interval if the assessment involves incomplete or uncertain information. The interval is determined by simply assuming that the unassigned belief degree, (β_H) , is transferred to the best and the worst grades to represent the upper and lower bounds of the interval, respectively. To determine the maximum and minimum expected utilities, denoted by $U_{max}(M)$ and $U_{min}(M)$, the following pair of optimisation models are employed. The comparison of alternatives can be conducted based on the average utility, or $U_{avg}(M)$, which can be defined as the midpoint between $U_{max}(M)$ and $U_{min}(M)$. The input variables for this model are the aggregated degrees of belief and the utilities of each grade.

Objective functions:

$$\begin{aligned} \text{Max } U_{max}(M) &= \sum_{n=1}^{N-1} u(H_n)\beta_n(M) \\ &\quad + u(H_N)(\beta_N(M) + \beta_H(M)) \\ \text{Min } U_{min}(M) &= u(H_1)(\beta_1(M) + \beta_H(M)) \\ &\quad + \sum_{n=2}^N u(H_n)\beta_n(M) \end{aligned} \quad (20)$$

s.t.

$$\beta_n^-(M) \leq \beta_n(M) \leq \beta_n^+(M) \quad (22)$$

$$\beta_H^-(M) \leq \beta_H(M) \leq \beta_H^+(M) \quad (23)$$

$$\sum_{n=1}^N \beta_n(M) + \beta_H(M) = 1 \quad (24)$$

4 Research methodology

The assessment process starts by investigating the importance (or the weights) of each criterion in respect of the logistics industry. It is considered that the weights should not be individually determined by a single organisation, although the assessment results are only monitored internally for that firm. This is because, for each firm, the image of 'being green' or sustainability does not depend only on their own perception, but it is also judged by the external community. Taking into account how people within the same industry perceive

which criteria significantly contribute to the overall environmental image can help firms improve their own performance in a more efficient way. For this study, the weights were elicited from 10 industry experts (hereafter referred to as the "decision makers", or "DMs") based on interviews and the direct rating technique. The 10 DMs include two academic lecturers in logistics management and eight senior managers from different truck fleet logistics companies in Thailand. There are several product types delivered by the eight companies, including metals and steel, cement, concrete, parcels, agricultural products, frozen foods, beverages, etc.

The direct rating is one of the most straightforward weighting methods available. It is highly recommended when the performance evaluation relies on a large number of criteria and indicators, and when a respondent does not feel comfortable using complex weighting methods [29, 30]. The direct rating only requires a DM to directly assign the weight to each criterion through a simple scale, such as 0-10 points ranging from 'extremely unimportant' to 'extremely important', respectively [31]. For this study, each DM was asked to assign the weights to each MPI criterion, followed by each criterion under OPIs. Finally, the weights were assigned to the overall picture of MPIs and then to that of OPIs. The DMs were also asked to provide supporting reasons. The weights of criteria belonging to the same group were then normalised to be relative to each other (they were sum to one). Indicators being under the same criterion were assumed to take equal weight in the aggregation processes. This is because the list of indicators under each criterion is likely to be changeable depending on a firm's current focuses and strategies.

To demonstrate the applicability of the ER approach to the ISO 14031-based environmental assessment, indicators representing each criterion were identified. Each criterion was able to be assessed through either quantitative or qualitative indicators, or both, and the numbers of indicators were not necessary to be equal among criteria. A measurement unit was specified for each quantitative indicator, and a set of evaluation grades was determined for each qualitative one. A clear definition referring to objective evidence or feasible situations was attached to each grade in order to standardise the assessment or to minimise subjective bias. For each qualitative indicator, the set of grades should cover all feasible practices. The number of grades might be different among different indicators. This depends on how many distinct levels of performances can be classified for each indicator [32]. Then, actual data sets from two logistics companies were employed to demonstrate the analysis. The data was gathered through the interviews of the company managers. They were two of the eight managers giving the weight information. The interviews were conducted during January 2018. For each quantitative indicator, the managers were asked to provide the data of the year 2017. They, however, were informed that they could give interval values if they were uncertain about the answers, or if the data fluctuated over a year. Another option was to skip some indicators if they completely lacked information about such indicators, or if they did not want to disclose such

information. In terms of qualitative assessment, the managers were asked to describe the extent of their practices regarding each particular indicator, without showing them the evaluation grades. Belief degree was then assigned to each grade by the researchers based upon the evidence or the stories provided by the managers. Since the aggregation of indicators according to the ER approach requires that all indicators must be assessed towards the same scale, five general grades (None, Poor, Fair, Good, and Excellent) were employed for this study. To unlock such condition, the rule-based transformation technique, introduced by Yang [26] and further developed by Wang, et al. [27] and Guo, et al. [28], was employed to transform various forms of original assessment data into belief distributions attached to the general grades. Due to the word limit, the transformation processes cannot be fully presented in this paper. After the transformation, the results of all indicators were arranged into a form that was applicable for the ER to derive the overall performance for each company, using the weight ranges determined by the 10 industry experts. Note that the nonlinear optimisation here was performed through LINGO software.

5 Results and discussion

5.1 The importance of the criteria (the weights)

Table 1 shows the relative weights of the criteria elicited from the 10 experts. The weights are presented in the form of the intervals to reflect different attitudes towards the importance or the contribution of each criterion to the overall environmental performance. The averages are also presented.

Table 1. Relative weights of the criteria.

Criteria	Weight ranges	Weight average
MPI1 - Implementation of environmental policies and programmes	0.21-0.30	0.25
MPI2 - Regulatory compliance	0.26-0.50	0.30
MPI3 - Financial performance associated with environmental performance	0.10-0.27	0.23
MPI4 - Community relations	0.10-0.24	0.22
OPI1 - Materials	0.00-0.15	0.09
OPI2 - Energy	0.15-0.30	0.19
OPI3 - Services supporting organisation's operations	0.10-0.16	0.12
OPI4 - Physical facilities and equipment	0.10-0.25	0.15
OPI5 - Products and services	0.00-0.17	0.10
OPI6 - Wastes	0.13-0.20	0.16
OPI7 - Emissions	0.14-0.30	0.19
MPIs	0.30-0.56	0.46
OPIs	0.44-0.70	0.54

For MPIs, in general, most DMs assigned a high weight to MPI2 indicating that they greatly realised the impact of legal fines and penalties on the image of their environmental performance. Many of them stressed that,

for a firm to sustain itself in the logistics business, regulations relating to freight and transportation must be strictly considered. A bad image regarding this point would destroy customers' loyalty and decrease prospective sales. MPI1, based on an average basis, seemed to be the second-most important management criterion. Many DMs stated that, to be successful in green programs, environmental concerns must be embedded into a company's policies and strategies.

Energy consumption (OPI2) and Emissions (OPI7) became the two most important operational criteria for the logistics industry. From the interviews, this sector consumed a large amount of diesel, and this led to the emission of various greenhouse gases. Most of the green logistics practices, therefore, focused on minimising fuel consumption and air pollution. The cost of diesel also represented the main cost of a transportation service, which needed to be minimised. Wastes (OPI6) also received a high weight from many DMs due to a strong association between logistics operations and many kinds of solid wastes, such as used tyres, pallets, paper boxes.

Regarding the two main aspects of environmental performance, based on an average basis, the weight of OPIs seemed to be only slightly higher than that of MPIs. Many DMs believed that actual outputs and outcomes better reflected a company's successful implementation of green initiatives than merely displaying management strategies. On the other hand, a number of DMs expressed a contrastive opinion that management should be the most important since it is the driving mechanism for a company to reach an expected outcome. However, the weight ranges of MPIs and OPIs greatly overlapped each other, indicating that the ranking based on the averages was not generalised. The intervals of the weights were then inputted into the optimisation model in order to analyse the environmental performance of the two companies.

5.2 Indicators and the transformation into belief distributions

To demonstrate the applicability of the ER approach, 20 indicators (belonging to the four MPIs and the seven OPIs) were generated, as shown in Table 2. They were adapted from the examples of indicators listed in ISO 14031 [18]. A measurement unit was specified for each quantitative indicator, and a set of evaluation grades (with clear definitions attached to each grade) was determined for each qualitative one. Equivalent rules for the transformation were also established for each indicator by referring to the industry's best and worst cases, based mainly on the review of the literature and the interviews. Unfortunately, these kinds of information were not displayed in this article due to the word limit.

For this paper, MPI3.1 and MPI1.3 were used to respectively demonstrate the transformation of assessment data into belief distributions for quantitative and qualitative indicators. For MPI3.1, its measurement unit was a percentage (%). Figure 1 shows the equivalent rule used for the transformation. The value 20% was considered here to be equivalent to the grade 'Excellent',

while ‘None’ was assigned to 0%. The concept of linear utility function was then applied to the remaining grades.

Table 2. List of indicators under each criterion.

Criteria	Indicators
MPI1	MPI1.1: The number of environmental initiatives per year (number)
	MPI1.2: Percentage of employees participating in environmental initiatives (%)
	MPI1.3: Environmental performance assessment (evaluation grades A-D)*
	MPI1.4: Monitoring of fuel consumption (evaluation grades A-D)*
MPI2	MPI2.1: Number of non-compliance with laws and regulations per year (number)
MPI3	MPI3.1: Percentage of total expenditure on environmental research and initiatives relative to the net profit per year (%)
MPI4	MPI4.1: The number of complaints from the local community per year (number)
	MPI4.2: Progress on local remediation and social development activities (evaluation grades A-E)*
	MPI4.3: Progress on a social survey about corporate community relations (evaluation grades A-E)*
OPI1	OPI1.1: The use of recycled or reused materials (evaluation grades A-E)*
OPI2	OPI2.1: Diesel used relative to a kilometre of transportation per year (litres per kilometre)
OPI3	OPI3.1: Waste management related to supporting operations (evaluation grades A-D)*
OPI4	OPI4.1: Percentage of vehicles in fleet with pollution-abatement technology (%)
	OPI4.2: Progress on vehicle preventive maintenance program (evaluation grades A-C)*
OPI5	OPI5.1: Ability to launch environmentally-friendly products/services (evaluation grades A-D)*
OPI6	OPI6.1: Management of used tyres (evaluation grades A-C)*
	OPI6.2: Management of used lubricating oil (evaluation grades A-C)*
	OPI6.3: The weight of solid waste disposed of per year (tonnes)
OPI7	OPI7.1: Carbon dioxide (CO ₂) emission per a tonne and a kilometre of transport (kilogram of CO ₂ equivalent/tonne-kilometre)
	OPI7.2: Emission of particulate matter under 10 microns diameter per a kilometre of transport (grams/kilometre)

*Qualitative evaluation

Fig. 1. Assessment data of company *M* on MPI3.1 and the equivalent rule.

Given h is the assessment data of MPI3.1 of company *M*, and $h \in [3.9, 4.9]$, which is an interval value covered by two adjacent grades (H_{None} and H_{Poor}), as shown in Figure 1. Following Wang, et al. [27], the degrees of belief assigned to H_{None} and H_{Poor} are then derived as the intervals, $[\beta_{None}^-, \beta_{None}^+]$ and $[\beta_{Poor}^-, \beta_{Poor}^+]$, as follows:

$$\beta_{None}^- = \frac{5 - 4.9}{5 - 0} = 0.02, \quad \beta_{None}^+ = \frac{5 - 3.9}{5 - 0} = 0.22$$

$$\beta_{Poor}^- = \frac{3.9 - 0}{5 - 0} = 0.78, \quad \beta_{Poor}^+ = \frac{4.9 - 0}{5 - 0} = 0.98$$

The data was finally transformed into the general form: $S(MPI3.1(M)) = \{(H_{None}, [0.02, 0.22]); (H_{Poor}, [0.78, 0.98])\}$.

For MPI1.3, an example of qualitative indicators, it was assessed through four grades (A-D). The equivalent rule is displayed by Figure 2.

Fig. 2. Assessment data (belief degree) of company *N* on MPI1.3 and the equivalent rule.

Figure 2 shows that the local grade ‘C’ was equivalent to the general grade ‘Poor’ for only around 0.666 degree of equivalent, while another 0.334 degree falls within the grade ‘Fair’. The same concept is also applied to the grade ‘B’. According to the story given by the manager of company *N* and the definitions attached to the assessment grades A-D for MPI1.3, the company’s performance towards this indicator falls within the grade ‘B’ for 60% and another 40% is in accordance with grade ‘C’. Then, following Yang [26], the assessment data was transformed into the general form: $S(MPI1.3(N)) = \{(H_{Poor}, 0.27); (H_{Fair}, 0.33); (H_{Good}, 0.4)\}$. At the end, the transformed belief distributions are summarised in Table 3. Note that $(H_H, 1)$ means that there was a lack of information due to sufficient data not being available or completely unknown, or the company did not want to disclose such information.

5.3 Application of the ER approach

The data from Table 3 was then used to analyse the environmental performance of each company under uncertain and incomplete information through the ER nonlinear optimisation model, or Equations (3)-(17). For each company, the aggregation of all indicators under the same criterion was firstly performed. This was followed by the aggregation of all MPIs. Then, all OPIs were combined. Finally, the overall performance was

determined by aggregating the two main aspects. Table 4 shows the aggregated interval degrees of belief for the overall MPI, OPI, and the environmental performance of the two companies. The intervals of the expected utilities were derived by Equations (20)-(24). The results are shown in Table 5. For this study, the utilities of the five general grades; $u(H_{None})$, $u(H_{Poor})$, $u(H_{Fair})$, $u(H_{Good})$, and $u(H_{Excellent})$; were set to be 0, 0.25, 0.5, 0.75, and 1, respectively.

Table 5 shows that company *M* receives the highest average utility in terms of the overall environmental performance. However, when considering the ranges of the utilities, the maximum utility of company *N* (0.74) is greater than the minimum of company *M* (0.42). This means *M* does not absolutely dominate *N*. If all information is precisely known, it is possible that the *N*'s performance might be greater than *M*'s.

Based on the average basis, *M* is ranked first in both MPIs and OPIs. When considering MPIs, the expected utility of *M* completely dominates that of *N* (the minimum of *M* is greater than the maximum of *N*). This is mainly due to the fact that *M* performs much better than *N* in both MPI1 and MPI2 which receive high weights from the industry representatives. For *M*, there are plenty of programmes and projects that relate to the environment. The manager of *M* also stressed, during the interview, that the company strictly conformed to legal requirements, and, as such, the amount of fines and the number of sanctions/warnings were considerably lower than those of *N*. There are a few indicators that *N* performs better than *M*, such as MPI4.2 which is assessed towards a linguistic scale (grades *A-E*). Company *N* is able to show evidence of their efforts to improve the quality of life of the community, such as sponsoring local religious activities and contributing to public infrastructure. CSR activities have been embedded in the company *N*'s policies whereas they have only been conducted unofficially by company *M*. The weight of MPI4, unfortunately, is not high enough to push *N* to be greater than *M* in terms of the overall management performance.

For OPIs, the ranges of the expected utilities of the two companies greatly overlap each other, and the average of *M* (0.67) is just slightly higher than that of *N* (0.50). In general, company *M* seems to perform better than *N*, particularly in OPI3 and OPI4. *M* has a policy of waste segregation. Food waste is used to produce organic fertilisers. Recyclable waste is sold to external parties. There is also a campaign to replace the use of printed documents with email. *M*, moreover, has a greater percentage of trucks that utilise liquefied petroleum gas (LPG), a kind of low-carbon energy source, when compared to *N*. On the other hand, when considering OPI2, one of the OPI criteria receiving the highest weight, the manager of company *N* gave information indicating that the average rate of companies *N*'s fuel consumption (diesel) fluctuated between 0.167-0.25 litres/kilometre while the manager of *M* only gave a single number (around 0.24 litres/kilometre). This implies that *M* performs slightly poorer than *N* in terms of energy consumption.

Table 3. Distribution assessment matrix of the two companies.

Indicators	Company <i>M</i>	Company <i>N</i>
MPI1.1	$(H_{Fair}, 1)$	$(H_{Fair}, 0.5); (H_{Good}, 0.5)$
MPI1.2	$(H_{Excellent}, 1)$	$(H_{Excellent}, 1)$
MPI1.3	$(H_{Fair}, 0.33); (H_{Good}, 0.67)$	$(H_{Poor}, 0.27); (H_{Fair}, 0.33); (H_{Good}, 0.4)$
MPI1.4	$(H_{Fair}, 0.33); (H_{Good}, 0.67)$	$(H_{Fair}, 0.33); (H_{Good}, 0.67)$
MPI2.1	$(H_{Good}, 0.2); (H_{Excellent}, 0.8)$	$(H_{None}, 1)$
MPI3.1	$(H_{None}, [0.02, 0.22]); (H_{Poor}, [0.78, 0.98])$	$(H_{Poor}, [0.32, 0.78]); (H_{Fair}, [0.22, 0.68])$
MPI4.1	$(H_{Good}, 0.6); (H_{Excellent}, 0.4)$	$(H_{Good}, 1)$
MPI4.2	$(H_{Fair}, 0.5); (H_{Good}, 0.5)$	$(H_{Good}, 1)$
MPI4.3	$(H_{None}, 1)$	$(H_{None}, 0.9); (H_{Poor}, 0.1)$
OPI1.1	$(H_{Good}, 0.7); (H_{Excellent}, 0.3)$	$(H_{Fair}, 1)$
OPI2.1	$(H_{Fair}, 0.4); (H_{Good}, 0.6)$	$(H_{Fair}, [0, 0.5]); (H_{Good}, [0, 1]); (H_{Excellent}, [0, 0.33])$
OPI3.1	$(H_{Excellent}, 1)$	$(H_{None}, 0.7); (H_{Poor}, 0.2); (H_{Fair}, 0.1)$
OPI4.1	$(H_{Fair}, 0.4); (H_{Good}, 0.6)$	$(H_{None}, 0.9); (H_{Poor}, 0.1)$
OPI4.2	$(H_{Poor}, 0.065); (H_{Fair}, 0.07); (H_{Good}, 0.065); (H_{Excellent}, 0.8)$	$(H_{Poor}, 0.33); (H_{Fair}, 0.34); (H_{Good}, 0.33)$
OPI5.1	$(H_{None}, 1)$	$(H_{None}, 0.5); (H_{Poor}, 0.33); (H_{Fair}, 0.17)$
OPI6.1	$(H_{Excellent}, 1)$	$(H_{Excellent}, 1)$
OPI6.2	$(H_{Excellent}, 1)$	$(H_{Excellent}, 1)$
OPI6.3	$(H_H, 1)$	$(H_H, 1)$
OPI7.1	$(H_H, 1)$	$(H_H, 1)$
OPI7.2	$(H_H, 1)$	$(H_H, 1)$

Table 4. Aggregated interval belief degrees of the two companies.

Performances	Company <i>M</i>	Company <i>N</i>
MPI1	$(H_{Fair}, 0.44); (H_{Good}, 0.33); (H_{Excellent}, 0.23)$	$(H_{Poor}, 0.06); (H_{Fair}, 0.30); (H_{Good}, 0.42); (H_{Excellent}, 0.22)$
MPI2	$(H_{Good}, 0.2); (H_{Excellent}, 0.8)$	$(H_{None}, 1)$
MPI3	$(H_{None}, [0.02, 0.22]); (H_{Poor}, [0.78, 0.98])$	$(H_{Poor}, [0.32, 0.78]); (H_{Fair}, [0.22, 0.68])$
MPI4	$(H_{None}, 0.32); (H_{Fair}, 0.16); (H_{Good}, 0.4); (H_{Excellent}, 0.12)$	$(H_{None}, 0.26); (H_{Poor}, 0.03); (H_{Good}, 0.71)$
MPIs	$(H_{None}, [0.02, 0.14]); (H_{Poor}, [0.05, 0.26]); (H_{Fair}, [0.08, 0.17]); (H_{Good}, [0.2, 0.28]); (H_{Excellent}, [0.3, 0.56])$	$(H_{None}, [0.29, 0.66]); (H_{Poor}, [0.03, 0.23]); (H_{Fair}, [0.06, 0.29]); (H_{Good}, [0.12, 0.31]); (H_{Excellent}, [0.04, 0.07])$
OPI1	$(H_{Good}, 0.7); (H_{Excellent}, 0.3)$	$(H_{Fair}, 1)$
OPI2	$(H_{Fair}, 0.4); (H_{Good}, 0.6)$	$(H_{Fair}, [0, 0.5]); (H_{Good}, [0, 1]); (H_{Excellent}, [0, 0.33])$
OPI3	$(H_{Excellent}, 1)$	$(H_{None}, 0.7); (H_{Poor}, 0.2); (H_{Fair}, 0.1)$
OPI4	$(H_{Poor}, 0.03); (H_{Fair}, 0.24); (H_{Good}, 0.34); (H_{Excellent}, 0.39)$	$(H_{None}, 0.45); (H_{Poor}, 0.23); (H_{Fair}, 0.16); (H_{Good}, 0.16)$
OPI5	$(H_{None}, 1)$	$(H_{None}, 0.5); (H_{Poor}, 0.33); (H_{Fair}, 0.17)$
OPI6	$(H_{Excellent}, 0.79); (H_H, 0.21)$	$(H_{Excellent}, 0.79); (H_H, 0.21)$
OPI7	$(H_H, 1)$	$(H_H, 1)$
OPIs	$(H_{None}, [0, 0.18]); (H_{Poor}, [0, 0.01]); (H_{Fair}, [0.08, 0.20]); (H_{Good}, [0.12, 0.39]); (H_{Excellent}, [0.23, 0.49]); (H_H, [0.12, 0.3])$	$(H_{None}, [0.1, 0.34]); (H_{Poor}, [0.04, 0.15]); (H_{Fair}, [0.02, 0.39]); (H_{Good}, [0.04, 0.4]); (H_{Excellent}, [0.09, 0.29]); (H_H, [0.13, 0.3])$
Overall environmental performance	$(H_{None}, [0.01, 0.17]); (H_{Poor}, [0.01, 0.17]); (H_{Fair}, [0.07, 0.20]); (H_{Good}, [0.13, 0.38]); (H_{Excellent}, [0.25, 0.58]); (H_H, [0.04, 0.22])$	$(H_{None}, [0.14, 0.59]); (H_{Poor}, [0.03, 0.21]); (H_{Fair}, [0.03, 0.38]); (H_{Good}, [0.05, 0.40]); (H_{Excellent}, [0.05, 0.25]); (H_H, [0.04, 0.22])$

Table 5. Expected utilities of the two sample companies
($U_{min}, U_{avg}, U_{max}$).

Performance aspects	Company <i>M</i>	Company <i>N</i>
MPIs	(0.56, 0.70 , 0.83)	(0.19, 0.32 , 0.46)
OPIs	(0.41, 0.67 , 0.93)	(0.19, 0.50 , 0.80)
Overall environmental performance	(0.42, 0.67 , 0.92)	(0.12, 0.43 , 0.74)

5 Conclusion

This paper presents a unique application of the ER approach to the assessment of environmental performance towards ISO 14031 for logistics operations. It shows that the ER algorithm is able to facilitate the combination of various criteria and indicators even when uncertainties in criteria weights and belief assignments (assessment data) are involved. Uncertainties and incompleteness in the assessment data are still preserved in the aggregated results. This enhances transparency and minimises any inaccuracies of the results that might be caused by too much effort being put into the process of converting imprecise or uncertain data into crisp numbers. The ER approach also suggests the design of evaluation grades in the assessment of qualitative criteria by linking the grades to pieces of evidence or feasible practices in order to minimise subjectivity. The relative weights of criteria are employed in the ER algorithm, and this indicates a compensatory approach among multiple criteria which is consistent with the nature of business performance evaluation and decision-making.

This study, also, is the first effort of its kind to elicit the weights or the importance of ISO 14031 criteria perceived by the logistics industry. Although it is difficult to reach a complete consensus within the industry regarding this matter, the obtained ranges of the weights are able to reveal some general viewpoints with major agreement. Overall, in terms of the management performance (MPIs), the industry tends to praise high contributions of conformance to legal regulations and the establishment of official environmental policies and programmes to the image of environmental management. In contrast, focusing only on building community relations tends to have only a small impact. In terms of operational performance (OPIs), fuel consumption and air emissions seem to be at the top of the most important criteria used to reflect the performance of a logistics operator. Understanding major concerns and perceptions of the industry could help create better guidelines for logistics companies to initiate their environmental improvement programmes in a way that is powerful and well-recognised by the community.

The ER approach is effective not only for performance comparison across different companies but a single company can also employ such method for their own self-assessment. The obtained aggregated degrees of belief facilitate the companies to establish improvement plans. For example, regarding the two sample companies, company *M* may choose to focus on indicator MPI1.1 (due to its high weight and the fact that the company itself still has only a 'Fair' performance) by establishing a team to investigate the feasibilities of

conducting more environmental initiatives. Company *N*, on the other hand, still has a low performance in many indicators. For this case, high-weight indicators, such as MPI2.1 and OPI7, should be firstly considered. For MPI2.1, the company should place emphasis on educating their staff and drivers of the updated laws and regulations that relate to their operations in order to avoid being subject to legal fines and penalties. In terms of OPI7, both companies completely lack the capability of properly estimating their carbon footprints of transportation services, which, nowadays, has become one of the key aspects considered by shippers or product manufacturers in their selection-process of logistics service providers. For unknown information, the ER approach assumes that their actual performance could be any value within the entire range of possible cases, and this lowers the minimums of their expected utilities.

This study does not aim to standardise the list of indicators under each ISO 14031 criterion for the logistics industry. The list of indicators shown in Table 2 only represents the examples for demonstrating the implementation process of the ER approach. Subsequent studies may consider investigating practical indicators that are generally recognised and implemented among logistics service providers in order to promote better performance comparison within the industry, using a general set of indicators.

Acknowledgements

This work was supported by the Thailand Research Fund and the Office of the Higher Education Commission [grant numbers MRG6080084].

References

1. L.M.D.F. Ferreira, C. Silva, and S.G. Azevedo, An environmental balanced scorecard for supply chain performance measurement (Env_BSC_4_SCPM), Benchmarking: An International Journal. **23(6)**, 1398-1422 (2016)
2. C.-N. Wang, H.-X.T. Ho, S.-H. Luo, and T.-F. Lin, An integrated approach to evaluating and selecting green logistics providers for sustainable development, Sustainability. **9(2)**, 218 (2017)
3. P. Oberhofer and M. Dieplinger, Sustainability in the transport and logistics sector: lacking environmental measures, Business Strategy and the Environment. **23(4)**, 236-253 (2014)
4. M. Björklund and H. Forslund, The purpose and focus of environmental performance measurement systems in logistics, International Journal of Productivity and Performance Management. **62(3)**, 230-249 (2013)
5. M.I. Piecyk and M. Björklund, Logistics service providers and corporate social responsibility: sustainability reporting in the logistics industry, International Journal of Physical Distribution & Logistics Management. **45(5)**, 459-485 (2015)

6. J.F. Henri and M. Journeault, Environmental performance indicators: an empirical study of Canadian manufacturing firms, *Journal of Environmental Management*. **87**, 165-176 (2008)
7. A. Scipioni, A. Mazzi, F. Zuliani, and M. Mason, The ISO 14031 standard to guide the urban sustainability measurement process: an Italian experience, *Journal of Cleaner Production*. **16**, 1247-1257 (2008)
8. C. Jasch, Environmental performance evaluation and indicators, *Journal of Cleaner Production*. **8(1)**, 79-88 (2000)
9. H. Walker, L. Di Sisto, and D. McBain, Drivers and barriers to environmental supply chain management practices: lessons from the public and private sectors, *Journal of Purchasing and Supply Management*. **14(1)**, 69–85 (2008)
10. D. Nawrocka, Environmental supply chain management, ISO 14001 and RoHS. How are small companies in the electronics sector managing?, *Corporate Social Responsibility and Environmental Management*. **15(6)**, 349–360 (2008)
11. R.L. Keeney and H. Raiffa, *Decision with multiple objectives: preferences and value tradeoffs* (John Wiley & Sons, New York, 1976)
12. W.L. Winston, *Operations research: applications and algorithms* (Thomson Learning, Toronto, 2004)
13. P. Zhou, B.W. Ang, and K.L. Poh, Comparing aggregating methods for constructing the composite environmental index: an objective measure, *Ecological Economics*. **59(3)**, 305-311 (2006)
14. W.W. Cooper, L.M. Seiford, and K. Tone, *Data Envelopment Analysis: a comprehensive text with models, applications, reference and DEA-solver software* (Springer, New York, 2007)
15. K.P. Yoon and C.L. Hwang, *Multiple attribute decision making: an introduction* (SAGE Publications, Thousand Oaks, California, 1995)
16. P. Sureeyatanapas, J.B. Yang, and D. Bamford, Evaluation of corporate sustainability, *Frontiers of Engineering Management*. **1(2)**, 176-194 (2014)
17. C.M. Tam, V.W.Y. Tam, and S.X. Zeng, Environmental performance evaluation (EPE) for construction, *Building Research & Information*. **30(5)**, 349-361 (2002)
18. International Standard ISO 14031: Environmental management - Environmental performance evaluation - Guidelines (International Organization for Standardization, 2013)
19. J.B. Yang, B.G. Dale, and C.H.R. Siow, Self-assessment of excellence: an application of the evidential reasoning approach, *International Journal of Production Research*. **39(16)**, 3789-3812 (2001)
20. K.S. Chin, J.B. Yang, M. Guo, and J.P.K. Lam, An evidential-reasoning-interval-based method for new product design assessment, *IEEE Transactions on Engineering Management*. **56(1)**, 142-156 (2009)
21. C. Fu and S. Yang, The combination of dependence-based interval-valued evidential reasoning approach with balanced scorecard for performance assessment, *Expert Systems with Applications*. **39(3)**, 3717-3730 (2012)
22. F. Liu, W.-d. Zhu, Y.-w. Chen, D.-l. Xu, and J.-b. Yang, Evaluation, ranking and selection of R&D projects by multiple experts: an evidential reasoning rule based approach, *Scientometrics*. **111(3)**, 1501–1519 (2017)
23. H. Sellak, B. Ouhbi, and B. Frikh, Energy planning under uncertain decision-making environment: an evidential reasoning approach to prioritize renewable energy sources, *Inteligencia Artificial*. **20(59)**, 21-31 (2017)
24. J.B. Yang and M.G. Singh, An evidential reasoning approach for multiple attribute decision making with uncertainty, *IEEE Transactions on Systems, Man, and Cybernetics*. **24(1)**, 1-18 (1994)
25. Y.M. Wang, J.B. Yang, and D.L. Xu, Environmental impact assessment using the evidential reasoning approach, *European Journal of Operational Research*. **174(3)**, 1885-1913 (2006)
26. J.B. Yang, Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty, *European Journal of Operational Research*. **131(1)**, 31-61 (2001)
27. Y.M. Wang, J.B. Yang, D.L. Xu, and K.S. Chin, The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees, *European Journal of Operational Research*. **175(1)**, 35-66 (2006)
28. M. Guo, J.B. Yang, K.S. Chin, and H.W. Wang, Evidential reasoning based preference programming for multiple attribute decision analysis under uncertainty, *European Journal of Operational Research*. **182(3)**, 1294-1312 (2007)
29. P.A. Bottomley, J.R. Doyle, and R.H. Green, Testing the reliability of weight elicitation methods: direct rating versus point allocation, *Journal of Marketing Research*. **37(4)**, 508-513 (2000)
30. P. Sureeyatanapas and S. Pathumnakul, Impacts of using relative weights in multiple criteria decision making: a comparative study between independent and overlapping-criteria decision problems, *International Journal of Applied Decision Sciences*. **10(2)**, 101-117 (2017)
31. P. Sen and J.B. Yang, *Multiple criteria decision support in engineering design* (Springer, London, 1998)
32. P. Sureeyatanapas, J.B. Yang, and D. Bamford, The sweet spot in sustainability: a framework for corporate assessment in sugar manufacturing, *Production Planning & Control*. **26(13)**, 1128-1144 (2015)