Towards a common measure of greenhouse gas related logistics activity using data envelopment analysis

Holden, R., Xu, B., Greening, P., Piecyk, M. and Dadhich, P.

Abstract

Monitoring company emissions from freight transport is essential if future greenhouse gas (GHG) reductions are to be realised. Modern economies are characterised increasingly by lower density freight movements. However, weight-based measures of freight transport activity (tonne-kilometre, tonnes lifted) are not good at describing volume-limited freight. After introducing the need for performance measurement, the problem of benchmarking is outlined in more detail. A context-dependent undesirable output data envelopment analysis (DEA) model, designed to be sensitive to business context, is then tested on a simulated set of fleet profiles. DEA can produce more consistent measures of good-practice, compared to ratio-based key performance indicators (KPI), providing emission reduction targets for companies and an aggregate reporting tool.

Keywords: Carbon Measurement, Logistics Activity, Benchmarking, DEA.

1. Introduction

In response to the climate change challenge, greenhouse gas (GHG) emissions reduction targets are being set at corporate (eg. DB Schenker, Deutsche Post DHL, Tesco), sector (e.g., the UK Freight Transport Association aims for an 8% reduction by 2015, relative to a 2010 baseline, and the European Commission, in its 2011 White Paper, postulates at least 60% reduction in GHG emissions from the transport sector by 2050), or national level (e.g., the UK Government has committed to achieve at least 80% reduction in GHG emissions by 2050, against the 1990 baseline) (Piecyk, 2015). The freight transport sector contributes a significant proportion of total surface transport emissions (McKinnon, 2007), and must therefore incentivise radical changes to achieve substantial improvements in its environmental performance. Various authors investigate measures to reduce supply chain carbon intensity, which reflects a genuine and world-wide motivation within the sector to reduce negative environmental impacts (e.g., Guerrero et al., 2013; Guerrero, 2014; Liimatainen et al., 2012; Li et al., 2013; Piecyk and McKinnon, 2010; Arvidsson et al., 2013; Larson et al., 2013).

For desired changes to be achieved, it is vital that the sector acquires a common language, identifies best practice, and compares companies along common yardsticks—i.e.,benchmarking, grounded in solid research, is needed. This is a common theme in supply chain performance where applications of mathematical models are underutilised (Wong and Wongk, 2008), and the importance of quantitative approaches highlighted (Dullaert and Zamparini, 2013; Hassini et al., 2012; Lättilä et al., 2013). At the same time, a balance must be struck—models must be simple-enough so they can be used in practice, but not so simple they misrepresent the problem in hand.

We contribute by applying a mathematical model to a key freight transport problem—i.e., how to rationalise GHG measurements against the backdrop of highly diverse operating characteristics among different freight transport companies. Traditionally, freight movements are measured with simple ratio indicators (see Section 2 for more detail), e.g. the tonne kilometre (tkm)—most data relating to logistics activity across the globe are still collected and reported by governments in this way (Piecyk and McKinnon, 2009). The proportion of low density / high volume products in the freight mix, however, increases as economies develop. For example, data collected by the UK Department for Transport indicate that low density products increasingly constitute the freight mix and a significant amount of freight is purely volume limited (Department for Transport, 2011).

Consequently, there is a need to incorporate volume-based measures into freight transport performance benchmarking. Unfortunately, finding an appropriate measure of performance is problematic; specialised companies employ a variety of measures—tonne kilometres, tonnes handled, number of drops etc.—in characterising their logistics activity, precisely because freight activity has become so diverse. On the other hand, representative and accurate measures are important tools in the process of change, and the danger of accepting limited tools is to risk slower progress towards GHG reduction targets.

Thus, the development of an adequate freight transport activity measure requires a fresh perspective and a novel approach. Realising an appropriate measure would see poor-performing companies identified relative to best practice and motivate appropriate company-level changes. At a reporting level, the creation of normalised data would also help monitor progress towards national-scale emission reductions. These are all desirable objectives from the point of view of the road freight sector, specifically the various initiatives aimed at reducing emissions, e.g., the Low Carbon Working Group, an industry-led scheme to help freight transport businesses record and report reductions in GHG emissions (FTA, 2013).

In this paper, we develop a benchmarking approach that integrates weight utilisation, volume utilisation, distance travelled, and related GHG emissions. This is achieved by exploiting data envelopment analysis (Cooper et al., 2000, 2007; Thanassoulis, 2001). The result, to the best of our knowledge, is the first attempt to benchmark environmental performance (expressed in terms of GHG emissions) of road freight transport operators with diverse operating characteristics. The rest of the paper is organised into five sections. Section 2 highlights the green logistics background, the problem of characterising freight activity, and the potential usefulness of DEA. The conceptual model, its mathematical formulation, and computational schemes are then covered in Section 3. The approach to data generation is described in Section 4, before results are presented and discussed (see Section 5). Finally, we summarise our findings and conclude by considering future research directions (see Section 6).

2. Background

2.1. Benchmarking emissions from road freight transport

Measuring a carbon footprint can be complicated for a number of reasons, including problems of measurement normalisation (Piecyk, 2015; Carbon Trust, 2007). Guidelines, principles, and the benefits of carbon footprinting have been reviewed (Piecyk, 2015; Piecyk et al., 2015; McKinnon et al., 2015), and conceptual approaches that are specific to road freight transport offered (Pérez-Martínez, 2012; Liimatainen and Pöllänen, 2010). A key question for the freight sector is *how should logistics practice be organised to help drive-down GHG emissions*?

For example, evidence on fuel use per vehicle type shows that fewer heavier vehicles achieve smaller carbon footprints, given the same amount of freight as more numerous smaller vehicles (Piecyk, 2015); load consolidation is clearly an effective way of reducing GHG emissions from this point of view. However, any proposed solution will depend on how the logistics system is conceived.

Generally, freight movement can be considered a system defined by an objective—i.e. the efficient movement of physical entities from origins to destinations. Physical entities are variously referred to as 'products', 'commodities', or simply 'freight', but these classifications are vague. The *commodity* description accommodates a potentially diverse range of products with very different weight and volume characteristics. *Products* are too specific for a generic description of the freight system. The green logistics framework (McKinnon et al., 2015) fails to take account of key factors, such as product handling characteristics (liquid/bulk vs. palletised, for example), which constrain the type of vehicle suitable for the logistics task.

It is the wider context that is most difficult to pin down when developing measures of road freight performance. For example, making multiple drops of relatively light loads in urban areas has a different purpose to hauling heavy goods longer distances. Larger vehicles suited to the latter are more efficient in absolute terms, but more numerous smaller vehicles may be more suitable, e.g. for urban deliveries, than one large vehicle. That is, more numerous vehicles would be required for frequent deliveries, but given the constraints of the urban infrastructure, such vehicles are likely to be much smaller in size and the vehicle-level efficiency obtainable regarding long-haul is simple not attainable. The design of a transport fleet will be influenced by many such operational necessities. Additionally, one company might have a wider set of logistics tasks than a more specialised freight firm. Specialised companies have different niches and companies that have similar logistics tasks perhaps have different fleet characteristics. A general measure must satisfy this range of comparisons and an important stepping-stone towards accurate benchmarking is how to define efficiency in terms of a broad-enough definition of fleet capacity utilisation.

A more general formulation of a logistics system can be developed by adopting a process perspective (e.g. Cooper et al., 2009; Stewart, 2009). In determining the best logistics operation, a service provider/purchaser must first consider what weight/volume of product needs to be moved and consider its handling characteristics. This informs the selection of transport vehicle type, which may be refined by considering vehicle availability and location constraints,

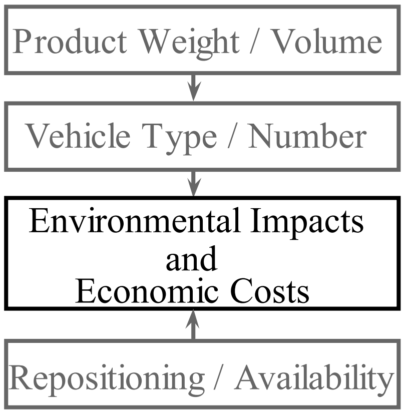

Figure 1: A process perspective on the freight logistics system.

Table 1: Logistics parameters and constraints

Freight Params.	**Vehicle Params.**	**Constraints**
Weight	Capacity (Cube/Weight)	Delivery/pick-up windows
Volume	Configuration	Repositioning
Handling Characteristics		Origin / destination

such as restricted time windows. Furthermore, the pool from which a vehicle is selected will determine the distance it travels to reach the origin of a journey (i.e., repositioning). As a whole, this process has environmental impact and economic cost, as illustrated in Figure 1.

A process perspective provides a general specification, expressed in terms of freight and vehicle parameters, along with related constraints, all focused primarily on individual journeys and vehicles (see Table 1). With this specification, logistics is constituted by products defined by weight, cube and character that are moved by vehicles, selected on the basis of suitability and capacity, to travel a given distance between origin and destination. Numerous factors determine the efficiency of logistics operation: vehicle characteristics (engine size, tare weight, aerodynamic features); available vehicle fleet; type of road and properties of gradient; weather conditions etc., all impact on performance.

In such circumstances the most efficient operations are those that have the best fit of resources for the specified logistics task. In other words, the most parsimonious description of road freight movement must account for the weight, cube and character of the freight, and the energy consumed during product movement. The energy consumed is a reflection of the constraints imposed on any particular operation.

Thus, as a consequence of the logistics context, benchmarking freight companies is problematic. It is highly questionable as to whether existing measures have the right pedigree. The use of simple KPIs (e.g. McKinnon, 2009), summarised in Table 2, is only meaningful if applied to similar organizations, or in longitudinal studies of one firm. This has relevance for GHG emissions, clearly, where understanding future projections is a key goal (e.g. Liimatainen et al., 2012; Li et al., 2013; Piecyk and McKinnon, 2010; Arvidsson et al., 2013; European Comission, 2011). Environmental benchmarking often utilises KPIs (Mckinnon, 2015), of the form ($P = \frac{a}{b}$), examples including: $P = \frac{CO_2^{kg}}{tkm}$ (Léonardi and Baumgartner, 2004), which could be used for weight-limited loads, and by analogy, $P = \frac{CO_2^{kg}}{m^3 km}$ might be applied for volume-limited loads. Here the denominator, b, is often referred to as a normaliser and is chosen to allow comparison across companies. Other examples of normalisers include *turnover*, *tonnes lifted*, *volume carried*, *km travelled*, etc., which have relevance to freight transport operators (FTA, 2010, 2012).

Table 2: Simple key performance indicators

KPI	**Description**
Vehicle Loading	Weight pallet numbers and average pallet height
Empty Running	Distance travelled empty
Fuel use	By vehicle type
Vehicle time use	Running, on the road but stationary, being loaded/unloaded, preloaded awaiting departure, delayed, maintenance, and stationary.
Schedule deviations	Problem at collection, problem at delivery, company actions, congestion and breakdown.

Simple KPI have an advantage. They are attractive to analysts in the sense that they are simple to understand and easily applied. One the other hand, their methodological limitations must be considered. The simple ratio form, for example, can undermine the very objective of allowing a comparison in the first place and knowing *which* normaliser is the most representative across firms is highly problematic—i.e., a firm might have strong results for some ratios, but perform poorly in others, making it difficult to judge whether the firm is efficient or inefficient.

Therefore, whilst the application of simple KPI to a single organisation can be useful in determining company performance over time (assuming no changes in operational drivers), they have limited use for comparing to other businesses; any differences can be argued to reflect different operational characteristics. In short, ratios typically examine parts of company activities, failing to provide sufficient performance information to reflect a firms multidimensional nature.

The question of how road freight transport operations can be compared cannot be answered by a set of single variables, or ratio measures. At a vehicle level, it is clearly optimal to fully utilise capacity (either weight or cube). However, as suggested, the vehicle *has* to be selected on the basis of its suitability to the logistics task. The combination of multiple KPIs mapped onto a system defined by complex multi-layered interactions makes comparisons between business units extremely difficult because common tasks, vehicles, and measurements are extremely unlikely.

2.2. Relevance of DEA, and business as usual

This complexity makes clear the desirability of a means by which comparisons between organizations can be simplified and made relevant. DEA models assume an input-output form, where generally applicable resources are used to produce generalised outputs. Furthermore, by creating a multi-dimensional space, the performance of the system can be reduced to a singular measure of system efficiency

(Olesen, 1995; Tone and Sahoo, 2003). Thus, in contrast to overly simple KPIs, multi-dimensional optimisation avoids the issue of choosing 'the correct' normaliser (Bogetoft and Otto, 2011). This is because non-parametric mathematical models, like DEA, are designed to capture *multiple* inputs and outputs and have mathematical mechanisms that express several denominators within an integrated optimisation space (Bogetoft and Otto, 2011).

We are not the first to recognise the relevance of DEA to environmental benchmarking and transportation, and it has found recent application in a number of relevant areas, as follows: individual truck performance (Odeck and Hjalmarsson, 1996); many modes of transportation, including sea (Odeck and Bråthen, 2012; Tongzon, 2001), air (Merkert and Hensher, 2011; Merkert and Mangia, 2014) and rail (Yu, 2008; Jain et al., 2008); transportation routing (Chiou et al., 2012; Zhao et al., 2011); national-scale environmental performance (Zhou et al., 2008; Ramanathan, 2005, 2006); logistics networks and green supply chains (Lau, 2013; Mirhedayatian et al., 2014; Azadi et al., 2014) and other company scale problems (Xu et al., 2009; Kao, 2010; Homburg, 2001; Tone and Sahoo, 2003; Lee and Saen, 2012; Lewis and Sexton, 2004), where the objective might be to minimise unwanted outputs alongside the maximisation of intended business objectives (Charles et al., 2012).

DEA is based on seminal ideas from the 1950s (Farrell, 1957), which were then developed in subsequent decades from the simpler idea of Farell efficiency to that of frontier analysis (Charnes et al., 1978; Färe and Grosskopf, 1983; Färe, 1986; Banker et al., 1984; Bogetoft and Otto, 2011). These ideas seek to define a set of efficient decision-making units (DMUs), firstly, then optimise relatively inefficient DMUs against them. This is achieved by creating a convex outer-envelope, derived from a non-dominated subset of data points. By exploiting linear programming (LP), and associated mathematical search techniques, a measure of the relative efficiency of dominated firms can be found.

Efficiency for all firms is thus defined relative to the outer-envelope that wraps-around efficient DMUs as a continuum, also known as the *efficiency frontier* (e.g. Thanassoulis, 2001). Any number of inputs can be isolated from any number of outputs in creating the efficiency front[1].

In the context of the current application, consider three freight companies (*Firms A, B* and *C*) with different logistics task. *Firm A* is a food-service wholesale distributor, *Firm B* is an independent lubricants company and *Firm C* provides distribution services to the newspaper and magazine supply chain. Given the above arguments, successfully applying DEA would mean that the fleet profiles of each company can be compared, resulting in a ranking of relative efficiency in terms of weight/volume usage against GHG outputs. In addition, if *Firm A* and *Firm B* outperform *Firm C*, but not each other, then they are said to dominate *Firm C*. The goal of *Firm C* is to step to the envelope constructed from the other two firms; DEA can provide the *direction* and *size* of the step required. In other words, DEA can establish benchmark companies, while acknowledging that logistics operations occupy multiple dimensions.

Given the intended practical application of DEA, the idea of a step-size is very important. When developing practical computational tools, computational modellers should be mindful of business risk. Risk is implicated in the size of the steps required, and whether this (or that) amount of change disrupts business as usual. For example, a large change might incur greater cost, improve the performance of the company, but at high risk. Arguments over the superiority of certain strategies over other are not compelling when divorced from business context (Normann, 2001); e.g., how different market environments impact industry in various ways can be seen through the application of cladistics (Rose-Anderssen et al., 2009).

Although business strategy is not a central topic of this paper, it is clearly beneficial if tools provide a range

[1]The quantities used are constrained in practice by their impact on the structure of the search space, which will influence search effectiveness.

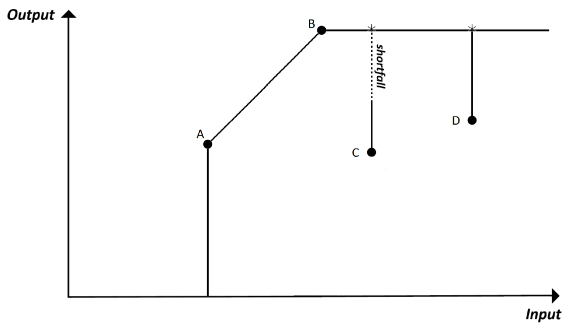

Figure 2: Two non-dominated firms define an envelope. The envelope is used to determine desirable outputs for the dominated firms, one of which (Firm C) has a shortfall. Please see in-text explanation.

of information relevant for a range of possible business contexts. Certainly, DEA assumes that an operationally altered *Firm C* can improve, and with the right changes perhaps to the point where it performs as well as *Firm A* and *Firm B*. Initially, *Firm C* might either be 'far from' efficient or 'close to' efficient and move more or less within the defined technology set to improve, but whether a firm is able to make the required step, of course, depends on an ability to do so. This situation is illustrated in Figure 2. Here, two non-dominated firms (*Firms A* and *B*) help define the efficient frontier, whereas two inefficient firms (*Firms C* and *D*) are required to increase their output. *Firm D* is able to do this, although *Firm C* is unable to make the required business changes to achieve, resulting in a shortfall of 'movement' required to reach the efficient frontier. The following method addresses this issue by adopting a context sensitive approach to allow firms with less flexibility to still improve.

3. A Context-dependent DEA framework for assessing GHG emissions from road freight transport activities

In this paper, we propose a context-dependent undesirable outputs DEA model. The model is designed to assess the relative performance of competing freight transport companies in terms of GHG emissions from freight activities. The key advantages of a DEA approach reside in its ability to handle many inputs and outputs simultaneously, and it does not require any specific knowledge of the production function. Furthermore, DEA identifies the efficient frontier (also referred to as the best-practice production frontier) along with a reference set for each decision making unit (DMU), as well as different types of targets to aim for. In addition, DEA works very well with small samples, which is relevant to our application (Maudos et al., 2002; Pasiouras et al., 2008).

Hereafter, we first present the basic concepts and models of DEA before discussing how one might adapt them to benchmark freight transport companies performances.

3.1. Basic concepts and models

DEA is a mathematical programming-based approach for assessing the relative performance of a set of n peer DMUs $\{\text{DMU}_j, j = 1, 2, \ldots n\}$, where each DMU is viewed as a system and defined by its inputs $x_{ij}(i = 1, 2, \ldots, m)$, and its outputs $y_{ij}(i = 1, 2, \ldots, s)$. The basic optimization problem addressed by DEA may be stated as follows: maximize the performance of a given DMU as measured by the ratio of a weighted linear combination of outputs to a weighted linear combination of inputs under the constraints that such a ratio is less than or equal to one for each DMU and the weights are non-negative. Using the Charnes-Cooper transformation (Charnes and Cooper, 1962), the fractional programming formulation of this optimization problem is transformed into a linear program and therefore is easy to solve. The mathematical formulation of the basic DEA input- and output- oriented analyses (Charnes et al., 1978), often referred to as CCR envelopment models after the authors, are presented in Table 3, where the variable θ_k^* (ϕ_k^*) is the technical efficiency ratio of DMU_k under evaluation. If the optimal value of θ_k^* (respectively, ϕ_k^*) is equal to 1, then the DMU_k under evaluation is efficient, otherwise $\theta_k^* < 1$ (respectively, $\phi_k^* > 1$) indicates that DMUk is inefficient and the current

Table 3: Input- (top) and output- (bottom) orientated CCR Models.

$$\min_{\theta,\lambda} \quad z = \theta_k - \varepsilon \left(\sum_{i=1}^{m} \mathbf{s}_i^- + \sum_{r=1}^{s} \mathbf{s}_r^+ \right)$$

$$\begin{aligned} \textbf{s.t.:} \quad & \sum_{j=1}^{n} \lambda_j \mathbf{x}_{i,j} + \mathbf{s}_i^- = \theta_k \cdot \mathbf{x}_{i,k}, \ \forall i \\ & \sum_{j=1}^{n} \lambda_j \mathbf{y}_{r,j} + \mathbf{s}_r^+ = \mathbf{y}_{r,k}, \ \forall r \\ & \lambda_j \ \geq 0, \ \forall j; \ \ \mathbf{s}_i^- \ \geq 0, \ \forall i; \ \ \mathbf{s}_i^+ \ \geq 0, \ \forall r \end{aligned}$$

$$\max_{\phi,\lambda} \quad z = \phi_k + \varepsilon \left(\sum_{i=1}^{m} \mathbf{s}_i^- + \sum_{r=1}^{s} \mathbf{s}_r^+ \right)$$

$$\begin{aligned} \textbf{s.t.:} \quad & \sum_{j=1}^{n} \lambda_j \mathbf{x}_{i,j} + \mathbf{s}_i^- = \mathbf{x}_{i,k}, \quad \forall i \\ & \sum_{j=1}^{n} \lambda_j \mathbf{y}_{r,j} - \mathbf{s}_r^+ = \phi_k \cdot \mathbf{y}_{r,k}, \quad \forall r \\ & \lambda_j \ \geq 0, \ \forall j; \ \ \mathbf{s}_i^- \ \geq 0, \ \forall i; \ \ \mathbf{s}_i^+ \ \geq 0, \ \forall r \end{aligned}$$

level of inputs (respectively, outputs) should be decreased (respectively, increased). Slack and surplus values are denoted s_i^- and s_i^+ respectively, and ϵ is a non-Archimedean element. The left-hand-side of the envelopment models is called the *reference set*, and any non-zero optimal λ_j^* is the weight assigned to DMUjs inputs and outputs in constructing the ideal benchmark of DMU_k (i.e., the projection of DMU_k on the efficient frontier). As to the interpretation of the constraints of CCR envelopment models, for example, the input-oriented envelopment model, the first set of constraints state that, for each input i, the amount used by k^{th} DMU's *ideal* benchmark plus the amount of slack, if any, must be equal to the *revised* amount used by DMU_k (i.e., amount adjusted for the degree of technical efficiency of DMU_k), and the second set of constraints state that, for each output r, the amount produced by the k^{th} DMUs *ideal* benchmark, minus the surplus, if any, must be equal to the amount produced by DMU_k.

The Charnes, Cooper and Rhodes (CCR) models, presented in Table 3, are also known as Constant-Returns-to-Scale (CRS) models; that is, a change in inputs does not achieve more or less than a proportional change in outputs. In practice, for the road freight industry, this assumption can be quite restrictive as not all firms will be operating at an optimal scale, and assuming CRS will result in measures of technical efficiencies (q_k/f_k) are confounded by scale efficiencies. The BCC model (named so after Banker, Charnes and Copper), under the variable-returns-to-scale (VRS), was developed to estimate the pure technical efficiency of decision making units with reference to the efficient frontier. Note that BCC models are formulated similarly to CCR models with just one additional constraint, namely $\sum_j^n \lambda_j = 1$, to be added into the CCR models in Table 3, such that they become VRS. In this study, we have opted for VRS for our proposed context-dependent undesirable output DEA models, which thus include the $\sum_j^n \lambda_j = 1$ constraint (see Eq 1 and 2).

3.2. Adaptation of DEA framework to model evaluation in GHG emissions

DEA is a generic framework and as such its implementation for our specific evaluation of relative performance requires a number of key decisions to be made.

First, what are the units to be assessed or DMUs? In this paper, there are 100 DMUs each of which has business inputs and outputs that are relevant to road freight transportation in fulfilment of the logistics tasks.

Second, what are the inputs and outputs? The inputs and outputs are measures of the relevant criteria for assessing the performance of a freight firm on their logistic task. A freight firm produces GHGs as an output from logistic activities, and the activity itself i.e., the use of vehicle fleets include constraints relating to weight and volume. That is, the fleet has a maximum capacity in terms of both weight and volume of goods it can carry. The goods themselves are carried over distances travelled by various vehicles to fulfil the freight tasks in hand. Therefore, our inputs relate to fleet-wise weight and volume utilisation and distance travelled by the fleet. The output is the GHG produced from these activities. Note that in the conventional DEA models (e.g., CCR, BCC envelopment models), it is assumed that outputs should be increased and the inputs should be decreased to improve the performance and/or to reach the best-practice frontier. However, our output (GHG) is clearly an undesirable output; it needs to be decreased in order to improve efficiency. However, if one treats the undesirable outputs as inputs so that the bad outputs can be reduced, the resulting DEA model does not reflect the true production process (Seiford and Zhu, 2002).

Third, what is the choice of the model orientation? One of the main objectives of a DEA benchmarking exercise is to project the inefficient DMUs onto the production frontiers. One can choose either input-oriented or output-oriented DEA models. To be more specific, input-oriented analysis minimizes input amounts for fixed amounts of output, while output-oriented analysis maximizes outputs for fixed amounts of inputs. In our empirical analysis, we opt for an output-oriented analysis because we aim minimise undesirable GHG output while performing the same freight activities.

Fourth, what is the appropriate DEA mathematical program to solve? Although all DEA models could be used to classify DMUs into efficient and non-efficient ones, and provide suggestions on reducing the current level of inputs and/or augmenting the current level of outputs to the best performing ones, the need to benchmark performance should not result in information that cannot in practice be used. The ability of a firm to adapt will vary from firm to firm, and over time. Therefore, it is important to provide a degree of choice for poor-performing DMUs, so that companies without a current capacity to improve *as far* as the most optimal efficiency frontier might still be able to identify better-if-not-best practices that are *reachable*. In this paper, we propose to a context-dependent undesirable outputs DEA model to assess the relative performance of freight firms' logistic activities and allow step-by-step

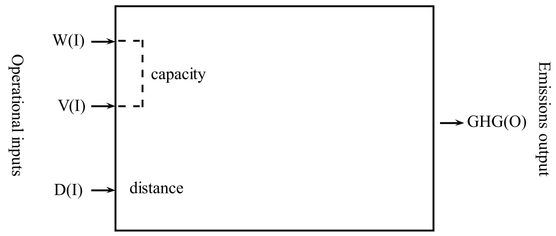

Figure 3: The decision-making unit must minimise the output, here GHG(O), according to fixed inputs, or maximise the outputs, i.e. W(I) , V(I) , D(I), according to a fixed input.

improvements. The proposed framework is a three-stage process: 1) Select the relevant DEA model; 2) Classify the DMUs and; 3) Compute context dependent process scores.

3.2.1. Select the relevant DEA model:

As suggested in the previous section, rather than maximising an output, the problem is to reduce GHG outputs from a road freight transport businesses. This concept is illustrated at the top of Figure 3. The output (O) is GHG. The fleet capacity (in terms of weight W and volume V), in addition to distance D, are specified as input (I) resources. This makes sense according to the business process; a fleet of vehicles, together, have an overall capacity in terms of weight and volume limitation, and the distance a vehicle travels can also be thought of as a kind of resource used in the delivery of goods. Naturally, GHG is an undesirable by-product of different input combinations. More specific mathematically defined inputs and outputs are described below (see Equations 3 to 6).

In our application, the relevant DEA model to use is the undesirable output model under VRS. Let y^G_{rj} denote the desirable outputs and y^B_{rj} undesirable ones. Specifically, we wish to decrease y^B_{rj} to improve the performance. In order to decrease the undesirable outputs, as described, with scope later on to increase y^G_{rj} we apply an existing approach (Seiford and Zhu, 2002). That is, we introduce a linear monotone decreasing transformation for undesirable output, as follows by: multiplying undesirable outputs by -1 and; finding a proper value v_r to let all negative undesirable outputs to be positive, where $v_r = \max_j$. The transformed undesirable outputs are: $\bar{y}^B_{rj} = -y^B_{rj} + v_r > 0$. The undesirable output model is summarized as:

$$
\begin{aligned}
\textbf{max} \quad & z = \phi_k + \varepsilon \left(\sum_{i=1}^{m} \mathbf{s}_i^- + \sum_{r=1}^{s} \mathbf{s}_r^{G+} + \sum_{r=1}^{s} \mathbf{s}_r^{B+} \right) \\
\textbf{s.t.:} \quad & \sum_{j=1}^{n} \lambda_j \mathbf{x}_{i,j} + \mathbf{s}_i^- \ = \mathbf{x}_{i,k}, \quad \forall i \\
& \sum_{j=1}^{n} \lambda_j \mathbf{y}^G_{r,j} - \mathbf{s}_r^{G+} \ = \phi_k \cdot \mathbf{y}^G_{r,k}, \quad \forall^G_r, \quad \forall r \\
& \sum_{j=1}^{n} \lambda_j \bar{\mathbf{y}}^B_{r,j} - \mathbf{s}_r^{B+} \ = \phi_k \cdot \mathbf{y}^B_{r,k}, \quad \forall^B_r, \quad \forall r \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \lambda_j \ \geq 0, \ \forall j; \ \ \mathbf{s}_i^- \ \geq 0, \ \forall i \\
& \mathbf{s}_r^{G+} \& \mathbf{s}_r^{B+} \ \geq 0, \ \forall r
\end{aligned}
\tag{1}
$$

where the i^{th} input, r_G^{th} output and r_B^{th} outputs of $\text{DMU}_j (j = 1, \ldots, n)$ are denoted by $x_{i,j} (i = 1, \ldots, m)$, $y^G_{r,j} (r_G = 1, \ldots, s)$ and $y^B_{r,j} (r_B = 1, \ldots, s)$, respectively, λ_j is the weight assigned to DMU_j in constructing it's *ideal* benchmark, s_i^-, s_r^{G+} and s_r^{B+} are slack variables associated with the first, the second and the third sets of constraints, respectively, and ϕ_k denotes the efficiency score of DMU_k. If the optimal value of $\phi_k^\ell = 1$, then DMU_k is classified as efficient; otherwise DMU_k is classified as inefficient and the efficient target

$$
\text{for DMU}_k := \begin{cases} \hat{\mathbf{x}}_{i,k} = \mathbf{x}_{i,k} - \mathbf{s}_i^{-*} \\ \hat{\mathbf{y}}^G_{r,k} = \phi_k^* \cdot \mathbf{y}^G_{r,k} + \mathbf{s}_r^{G+*} \\ \hat{\mathbf{y}}^B_{r,k} = v_r - (\phi_k^* \cdot \bar{\mathbf{y}}^B_{r,k} + \mathbf{s}_r^{B+*}) \end{cases}
$$

3.2.2. Classification of DMUs

We use the following algorithm (see also Table 4) to further partition the set of DMUs into several levels of best-practice frontiers or evaluation contexts, say L, so that the 1st-level efficient frontier DMUs have a better performance than the 2nd-level efficient frontier DMUs and so on, until no DMU is left.

Table 4: Iterative procedure: levels of best practice

Initialization step
Initialize the performance level counter ℓ to 1 and the set of DMU's to evaluate at level ℓ, say J^ℓ, to $\{\text{DMU}_k, k = 1, \ldots, n\}$. Use the undesirable output DEA model to evaluate J^ℓ and set the l^{th} best-practice frontier $E^\ell = \{k \in J^\ell : \text{Efficiency Score} \rho_k^\ell = 1\}$. Exclude the current performance level best-practice frontier E^ℓ from the set of DMUs to evaluate next—i.e., set $J^{\ell+1} = J^\ell - E^\ell$, increment ℓ by 1 and proceed to the iterative step.

Iterative step
While $J^\ell \neq \emptyset$ **Do**
{
  Use the *undesirable output DEA model* to evaluate J^ℓ
  Set the ℓ^{th} best-practice frontier E^ℓ accordingly
  Set $J^{\ell+1} = J^l - E^\ell$
  Increment $\ell = \ell + 1$
}

Once DMUs have been partitioned into ℓ best-practice frontiers with different levels of performance, one could select the suitable level of efficient frontiers to serve as the evaluation context for measuring the relative progress of worse performing DMUs.

3.2.3. Compute context-dependent progress scores

This stage is designed to measure the progress of inefficient DMUs with respect to a particular evaluation context. For each $\text{DMU}_k \in E^\ell (\ell = 2, \ldots L)$, we can compute the relative progress scores with respect to a better evaluation contexts, say $P_k^*(d)$ for values of the evaluation context index d ranging from 1 to $\ell - 1$, provide potential incremental improvements in performance, where the projection of DMU_k onto the 1st-level efficient frontier is referred to as the *global target*. In practice, it is likely that inefficient DMUs are unable to improve their performance immediately onto the 1st-level efficient frontier due to certain constraints (e.g., unavailable resources). Thus, it is more desirable to set up the intermediate targets for inefficient DMUs and move the DMUs step by step onto an attainable best-practice frontier. These progress scores are often referred to as d-degree progress and are determined by solving the following model for values of d ranging from 1 to $\ell - 1$:

$$
\begin{aligned}
\textbf{max} \quad & P_k^*(d) = \phi_k(d) + \varepsilon \left(\sum_{i=1}^{m} \mathbf{s}_i^- + \sum_{r=1}^{s} \mathbf{s}_r^{G+} + \sum_{r=1}^{s} \mathbf{s}_r^{B+} \right) \\
\textbf{s.t.:} \quad & \sum_{j=1}^{n} \lambda_j \mathbf{x}_{i,j} + \mathbf{s}_i^- \ = \mathbf{x}_{i,k}, \quad \forall i \\
& \sum_{j=1}^{n} \lambda_j \mathbf{y}_{r,j}^G - \mathbf{s}_r^{G+} \ = \phi_k \cdot \mathbf{y}_{r,k}^G, \ \ \forall_r^G, \ \ \forall r \\
& \sum_{j=1}^{n} \lambda_j \bar{\mathbf{y}}_{r,j}^B - \mathbf{s}_r^{B+} \ = \phi_k \cdot \mathbf{y}_{r,k}^B, \ \ \forall_r^B, \ \ \forall r \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \lambda_j \ \geq 0, \ \forall j; \ \ \mathbf{s}_i^- \ \geq 0, \ \forall i \\
& \mathbf{s}_r^{G+} \& \mathbf{s}_r^{B+} \ \geq 0, \ \forall r
\end{aligned} \tag{2}
$$

Note that the larger value of $P_k^*(d)$, the more progress is expected for DMU_k; thus, a smaller value of $P_k^*(d)$ is preferred.

4. Data

Most road freight operators do not currently collect freight volume data. Indeed, one of the purposes of the current work is to motivate companies to do so by demonstrating the usefulness of DEA as a performance measurement tool. It is as a result necessary to simulate data as a means of demonstrating how the proposed model can be applied. However, this must be achieved in a way that relates to the profile of a real fleet of vehicles. Generating fleet data, for example, based on assumptions about the *overall* statistics of a fleet would be the most simple approach, but would be inadequate to demonstrate how results from DEA can be related to a fleet profile—i.e., the level at which a decision-maker might implement changes in order to improve performance. We thus adopt a process perspective on the freight logistics system, outlined above by defining a unique fleet of vehicles with unique journeys to fulfil the delivery of orders, and where each vehicle has its own weight and volume utilisation.

4.1. Generative model

Consider a fleet of vehicles which is tasked with fulfilling the delivery of a total load $L^f(W^f, V^f)$, which consists of a total weight W^f and volume V^f that requires delivery by a fleet over a given distance (D^f).

A vehicle fleet of size N^f is parametrised randomly in the range $N^f = [20, 30]$ such that each company has between 20 and 30 vehicles. Each vehicle v will have a number of properties. The vehicle is determined randomly to be either of two types $T \in \{T^1, T^2\}$, with corresponding maximum weight $W_v^{\max} \in \{25200 \text{ kg}, 39750 \text{ kg}\}$ and maximum volume $V_v^{\max} \in \{12200 \text{ L}, 36000 \text{ L}\}$. Vehicle weight, vehicle volume, and the distance a vehicle travels are assigned such that $W_v \leq W_v^{\max}$, $V_v \leq V_v^{\max}$ and $D_v \leq 3553528$, respectively. That is, no vehicle can carry more than its capacity allows and no vehicle undertakes a job with a distance greater than the total. Distance travelled D is processed by a fuel conversion factor $F \in \{0.731606, 0.707793\}$ for the respective vehicle types so that GHG emissions can be calculated by vehicle types. Table 5 provides a summary of the parameters used for the data generation exercise. The values for the logistics tasks per firm, based on the description above, are derived from realistic ranges, being taken from an example data file used within Optrak®, a real-world software program dedicated to vehicle routing problems. The values for fuel conversion to GHG emissions are derived from the UK Government (Department for Environment, Food and Rural Affairs (Defra, 2012).

Using the procedure outlined in Algorithm 1, we generate 100 fleets with weights, volumes and distances allocated per vehicle, and which sum to the total weights and volumes used to define the DEA model inputs (I) and outputs (O) defined below (see Equations 3 to 6).

Each fleet thus represents a separate logistics firm, or *DMU* in the terminology introduced above. For presentation to the DEA model (see Equations 1 and 2) the inputs and output are aggregated across each fleet as follows:

Table 5: Parameters for data generation

	description	Ranges / Values
W	weight to drop (job)	$[331013, 367792]$ kg
V	volume to drop (job)	$[398833, 443147]$ L^3
D	travel distance (job)	$[3198176, 3553528]$ km
N^f	fleet size (job)	$[20, 30]$
$W_v^{\max}$	max vehicle weight	$\in \{25200, 39750\}$ kg
$V_v^{\max}$	max vehicle volume	$\in \{12200, 36000\}$ L^3
D_v	Distance travelled	$[0, 3553528]$ km
F_v	fuel factor (Defra)	$\in \{0.731606, 0.707793\}$

Algorithm 1 Data generation pseudo code

procedure R(a, b) **return** random integer $\in \{[a, b]\}$
end procedure
procedure DistributeToFleet(am, Z)
  while ac $\neq$ am **do**
    $v \leftarrow R(0, N)$
    if am - ac $< Z_v^{\max}$ **then** max = am - ac
    end if
    if am - ac $\geq Z_v^{\max}$ **then** max = $Z_v^{\max}$
    end if
    $load \leftarrow \text{increment}(v, Z, R(0, max))$
    $ac \leftarrow ac + load$
  end while
end procedure
procedure Weights(n)
  for n **do**
    $w_n = R(1, 10)$
  end for
return $w_n \leftarrow \frac{w_n}{\sum_i^n w_i}$
end procedure
procedure Main(am, Z)
  fleetSize $\leftarrow R(N_{lo}^f, N_{hi}^f)$
  fleetWeights $\leftarrow$ DistributeToFleet$(W, Z = W)$
  fleetVols $\leftarrow$ DistributeToFleet$(V, Z = V)$
  fleetDistances $\leftarrow D\times$ Weights(N)$_v$
end procedure

$$W(I) = \frac{\sum_v^{N^f} W_v}{\sum_v^{N^f} W_v^{\max}} \tag{3}$$

$$V(I) = \frac{\sum_v^{N^f} V_v}{\sum_v^{N^f} V_v^{\max}} \tag{4}$$

$$D(I) = \sum_v^{N^f} D_v \tag{5}$$

$$GHG(O) = \sum_v^{N^f} F_v D_v \tag{6}$$

The result of the data generation process is the creation of the input $\mathbf{x}$ and output $\mathbf{y}$ vectors for the DEA models described above (see Section 3).

5. Results and discussion

This section is separated into two parts. Firstly, we present results from the application of a number of different ratio-based measures to our data. This demonstrates the problem with existing KPI measures in the road freight transportation sector. Secondly, we present results of the application of the DEA model, and discuss how this approach offers a more satisfactory alternative to current practice.

5.1. Ratio-based KPI as performance measures

McKinnon (McKinnon, 2009; Mckinnon, 2015) lists a number of freight transportation key performance indicators, including those listed in Table 2 and others related to *tkm*, the typical payload-distance measure used in the sector (see Section 2). As suggested, one of the main problems with ratio-based measures is their lack of ability to provide consistent performance evaluation. For example, in road freight transportation, *vehicle loading* is a frequently used KPI, but has different definitions. These differences represent variations in fleet constraints across firms whose definition of 'vehicle fill' will be specific to the operation (McKinnon, 2009). For instance, vehicles employed to move heavy-duty commodities will weight-out, whereas in 'retailing' and 'non-food' sectors space (area and volume) is the key vehicle constraint.

Given the need to benchmark performance, this situation is currently problematic. In Table 6 we present six different ratio-based measures. These are arranged from tonne-kilometre to GHG-per-kilometre along the top row. In order to include a vehicle-loading measure for both weight *and* volume, we include $\frac{t}{t_{max}}$ and $\frac{m^3}{m^3_{max}}$, measures that attribute performance according to capacity utilisation for weight (tonne) *or* volume (meter-cube), respec-

Table 6: Simple KPI: Rankings

	tkm	$\frac{t}{t_{max}}$	$\frac{m^3}{m^3_{max}}$	$\frac{\text{GHG}t_{max}}{t}$	$\frac{\text{GHG}m^3_{max}}{V}$	$\frac{\text{GHG}}{km}$
1	51	<u>36</u>	0	<u>36</u>	40	49
2	47	80	49	41	98	14
3	48	<u>91</u>	14	<u>91</u>	63	58
4	50	<u>86</u>	<u>84</u>	<u>86</u>	36	<u>84</u>
5	45	63	99	80	34	1
6	46	41	70	40	86	27
7	49	25	72	98	54	5
8	52	40	<u>30</u>	89	91	<u>30</u>
9	10	15	59	63	17	41
10	6	10	9	3	92	66

tively. We also include $\frac{\text{GHG}t_{max}}{t}$ and $\frac{\text{GHG}m^3_{max}}{V}$, which can be thought of as the GHG-per-utilisation of weight and GHG-per-utilisation of volume, respectively. These equations, perhaps, appeal more to intuition if we express them differently—i.e., $\frac{\text{GHG}}{(t/t_{max})}$ and $\frac{\text{GHG}}{m^3/m^3_{max}}$—rather than as they are tabulated. We also include GHG-per-kilometre.

At the left-hand side of the table are the top-ten ranked firms out of the set of DMUs considered in this study. For example, DMU_{51}, DMU_{47}, and DMU_{48} are ranked 1^{st}, 2^{nd} and 3^{rd} according to the tkm KPI. That is, assuming that tonnes and kilometres are the only business inputs that matter, we can compare firms along the tkm yardstick. Of course, the whole motivation for this paper is based on a recognition that single inputs do not capture a range of operational constraints and that using a number of ratio-based KPI cannot easily allow consistent rankings. To illustrate further, when comparing tkm ranks to $\frac{t}{t_{max}}$ ranks, we notice that each ranking position for the respective measure is occupied by a different DMU index. Furthermore, only a single firm (DMU_{10}) appears in both columns; 90% of the top ranked performers are therefore different across these measures. In other words, tkm and $\frac{t}{t^{max}}$ produce highly inconsistent results. Although this extreme level of inconsistency is not common over the entire table, a high degree of inconsistency is prevalent. We indicate this by underlining consistent results—i.e., if a DMU index appears in the same position for different KPI—which occur only $10/60 \approx 17\%$ of the time!

The only reachable conclusion from the above analysis is that ratio-based KPI, for ranking performance across different road freight transport companies, is inadequate. This problem is known in practice, in the road freight transport sector, and is one that is shared in the business modelling literature (Bogetoft and Otto, 2011). Furthermore, ratio-based KPI do not aggregate particularly well; it can been shown that even if two separate KPI produce the same ranking, it is possible that when combined the rankings change (Bogetoft and Otto, 2011; Fox, 1999), a proof known as Fox's paradox. In light of the view that performance measurements are often most effective when holistic (Hanman, 1997), this paradox is another troubling feature of simple, ratio-based KPI.

5.2. The DEA model as a performance measure

Our survey of the literature revealed that current practice tends to use any number of criteria to evaluate road freight transport activities. However, when firms are compared to each other using a simple KPI, we have shown that one cannot consistently determine which firm performs best. In order to overcome this methodological issue, we propose a multidimensional framework based on a context-dependent undesirable output DEA model, to rank road freight transport firm efficiency. All DEA models could be used to classify firms into efficient and non-efficient ones, and provide suggestions on improving the current level of inputs and/or outputs of the business. However, the need to benchmark performance *should not* result in information that *cannot* be used in practice. Therefore, our proposed methodology allows decision makers to select the suitable evaluation context to benchmark the poor-performing firms, providing intermediate targets; firms without a current capacity to improve 'as far' as the most optimal efficiency frontier might still as a result be able to identify better-if-not-best practices that are 'reachable'.

Results of the DEA analysis return efficiency scores of 1, if the given DMU is efficient, otherwise the DMU is inefficient. Results reveal a technically efficient set $E_1 = \{0, 28, 97, 99\}$ whose slack values are all 0. For example, while minimising the output and identifying the optimum, there is no scope for improvement in the use of business inputs for these firms. Clearly, without using a context sensitive approach, one practical problem for companies who require large changes to be made in order to meet targets is the potential amount of associated business upheaval. Data envelopment analysis itself relies on a feasible space that separates firms along a number of dimensions. When the distance from the efficient frontier is large, then the assumption is that the amount of change required to achieve similar performance is also large.

Results from a context dependent DEA run are presented in Table 7. The set of DMU's are partitioned into 18 levels of best-practice frontiers, where E_1 is better practice than E_2, E_2 is better practice than E_3 and so on, down to the worst performing $\text{DMU}_{25} = E_{18}$. Apart from the very best practice frontier, E_1, each other frontier has an adjacent frontier that contains other targets when considering a smaller potential improvement. Therefore, for each DMU in a given efficient frontier $E_{i>1}$ the set of 'slightly better' DMUs are E_{i-1}. In this way, by identifying inefficient DMUs we also consider their evaluation context in defining their potential target levels.

However, a firm within a given efficiency frontier may not necessarily need to target each and every frontier E_{i-1} to progress in manageable steps. For example, we have a number of DMUs that are located on the 7^{th} efficient frontier and it might be realistic for these firms to progress immediately to the 5^{th} efficient frontier, then in the medium-term to the 3^{rd}, and ultimately in the long-term the 1^{st}, the 'best-practice' frontier. In Table 8, we present the progress context-dependent scores for each DMU in set E_7, per row, associated with these two "intermediate" frontiers (E_5 and E_3), and the best efficiency frontier (E_1). Larger progress scores means more progress is expected from a DMU, this is what we have observed in Table 8, as the efficient frontier decreases, the progress score increases for each DMUs in E_7.

Table 7: Efficiency frontiers from best to worst

Context sensitive efficiency frontiers
$E_1 = \{0, 28, 97, 99\}$
$E_2 = \{18, 41, 43, 49, 72, 84, 95, 96\}$
$E_3 = \{14, 30, 36, 42, 59, 70, 73, 87, 90, 94\}$
$E_4 = \{9, 37, 39, 61, 83, 98\}$
$E_5 = \{13, 33, 35, 38, 44, 58, 66, 71, 79\}$
$E_6 = \{8, 23, 29, 40, 46, 53, 60, 65, 85, 88, 93\}$
$E_7 = \{4, 6, 16, 31, 45, 67, 86, 89, 92\}$
$E_8 = \{5, 20, 21, 32, 81, 82, 91\}$
$E_9 = \{1, 2, 3, 12, 34, 51, 56, 68, 74\}$
$E_{10} = \{11, 50, 55, 57, 77\}$
$E_{11} = \{17, 24, 54, 64, 69, 78\}$
$E_{12} = \{15, 26, 75\}$
$E_{13} = \{7, 19, 27, 62, 76\}$
$E_{14} = \{47, 63, 80\}$
$E_{15} = \{10, 48\}$
$E_{16} = \{52\}$
$E_{17} = \{22\}$
$E_{18} = \{25\}$

Table 8: E_7 Evaluation Context: Context scores for target frontiers

Context DMUs $\in E_7$	Short-term (E_5)	Mid-term (E_3)	Long-term (E_1)
4	1.25	1.51	1.57
6	5.27	8.12	9.17
16	1.66	2.22	2.31
31	1.36	1.78	1.86
45	5.74	7.42	8.75
67	1.74	2.42	2.52
86	1.21	1.25	1.28
89	1.21	1.25	1.29
92	1.30	1.33	1.38

One of the key features of DEA models is to suggest how to move inefficient firms towards efficiency i.e., the changes/percentage changes required for projecting a DMU to the respective target levels. In Table 9, we present an example for DMU_6. The table is arranged by row according to the current values for DMU_6, then the percentage changes required (Δ_j) to reach the corresponding efficient front (E_j) in the following row, for three efficient front examples $j \in \{5, 3, 1\}$. For example, in the short-term, it needs to reduce 4.22% of its current travel distance, and 4.93% of its current CO2 emissions to reach the 5th best practice frontier. Short-term, this might im-

Table 9: Percentage changes required to achieve target improvement values for DMU_6.

	Weight	Vol	Distance	GHG
DMU_6	0.39	0.62	3.52	2540
Projection_5	0%	0%	-4.22%	-4.93%
E_5	0.39	0.62	3.37	2414.69
Projection_3	0%	0%	-8.03%	-8.23%
E_3	0.39	0.62	3.24	2330.84
Projection_1	-5.78%	0%	-9.02%	-9.44%
E_1	0.37	0.62	3.2	2300.01

ply better route planning. In the medium-term, it needs to reduce 8.03% of its current travel distances, and 8.23% of its current GHG emissions to reach to the 3$^{\text{rd}}$ best practice frontier. In the long-term, this firm needs to improve its weight utilisation by 5.86%, reduce 9.02% of its current travel distance, and reduce 9.44% of its current CO2 emissions to reach the 1$^{\text{st}}$ best practice frontier. Medium to long-term changes might involve changes to the fleet, or more strategic alterations, such as the repositioning of supply-chain facilities etc.

From the point of view of a given firm, especially a poor-performing one, context sensitivity provides results that are conducive to business implementation. While implementing percentage changes, a useful approach would be to be able to compare the DMU under consideration to specific firms in the target efficiency frontier. For example, if it is known that one firm is underperforming as a result of DEA analysis, and that improvements can be made in % changes, then the question is 'how, practically can this be acheived?'. A freight operator may reason, based on experience, that few improvements can be made and being able to compare, qualitatively, with benchmark DMU's would be highly desirable. For example, consider again DMU_6 in a case where the business is able to invest in changes to improve, from its current frontier (E_7) to the best practice frontier (E_1). In order to know how to invest, the company could compare to a subset of firms in E_1 (say, DMU_0 and DMU_{28}). When we look at the fleet profiles of these two best practice examples, we notice that they have characteristics similar to DMU_6, in terms of number of vehicles and proportion of type 1 and type 2 vehicles, which relate, recall, to the fuel conversion factor (see Section 4). From this we can say that the logistics task is similar in terms of the characteristics and amount of commodities moved. However, it is clear that DMU_6, completes its logistics tasks by travelling significantly further. Therefore, DMU_6 should focus on business strategies that reduce the distance travelled, if possible—for example, investing in a vehicle routing software, possibly with 'green' routing algorithms (Erdoğan and Miller-Hooks, 2012; Bektaş and Laporte, 2011).

While discussing business strategy, it it important to note that the use of standard conversion factors in the current simulations employ an activity-based approach where only distance and weight is available per vehicle type with resulting GHG being derived from these. However, in a real world application a fuel-based approach would be more appropriate i.e., to use data on the amount of fuel consumed per DMU. This is because the activity-based approach, being based simply on distance and weight, cannot represent the contribution to fuel efficiency interventions—e.g. driver training, reconfiguration of the fleet in terms of vehicle technology, such as lightweighting etc. Furthermore, any changes to the number of vehicles used might also reflect a more strategic supply chain reconfiguration. The assumptions we have made in the simulated data, of course, cannot capture the diversity of a real logistics context that would otherwise be accessible through real-world data from companies.

6. Summary and conclusions

Existing measures of freight transport activity are best suited for describing weight-limited freight and fail to accurately reflect the level of spatial utilisation of available vehicle capacity. For benchmarking, which is needed in the face of environmental pressures, a different approach to measuring the performance of freight operations must be employed. We proposed DEA as a suitable alternative to existing measures and demonstrated how such a multi-dimensional approach might be exploited to capture weight *and* volume limitations. We also demonstrated the limitation of simple ratio-based KPI typically used in the road freight transport sector, which drives the need for a more novel application.

A key practical issue is that, although volume information is important, few companies actually collect such data. Nevertheless, in order to motivate data collection and demonstrate the use of the proposed model, we have created a scheme that simulates the kind of data required. The subsequent DEA has demonstrated how minimisation of GHG can be used to interpret the relative performance of DMUs, and an example DUM was chosen to illustrate this. However, in the end, improvements must be achieved by individual firms whose business context should be taken account of. We presented a context-dependent undesirable output DEA model whose context sensitivity provides a potential means to enrich the analysis and encourage attainable, company-level adaptation.

Considering that DEA is a data-driven modelling approach, the lack of real data is a limitation of the current work, especially considering the data-driven model employed. However, we cannot overemphasise the motivational role that such a study can play in helping bootstrap the process of data collection. The authors are currently in collaboration with a leading freight transport body whose task is to encourage its members to reduce emissions and to report sectoral-level emissions to government. The use of DEA is relevant at this level—not necessarily in the context of individual companies—where comparisons *across* firms have the aim of informing policy. The fact that data collection can be an expensive task is a reason why preliminary demonstrations of newly proposed measures, such as the one presented here, can be persuasive, if shown to have potential. In other words, it is highly unlikely that companies will collect data without a business case to support this. Our aim in the future is to continue to engage with stakeholders. The desired outcome is that real data will become available and that the framework proposed here can be employed to produce real fleet-to-fleet comparisons.

References

Arvidsson, N., Woxenius, J., Lammgård, C., Feb. 2013. Review of road hauliers' measures for increasing transport efficiency and sustainability in urban freight distribution. Transport Reviews 33 (1), 107–127.

Azadi, M., Shabani, A., Khodakarami, M., Saen, R. F., 2014. Planning in feasible region by two-stage target-setting {DEA} methods: An application in green supply chain management of public transportation service providers. Transportation Research Part E: Logistics and Transportation Review 70 (0), 324 – 338.

Banker, R. D., Charnes, A., Cooper, W. W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092.

Bektaş, T., Laporte, G., 2011. The pollution-routing problem. Transportation Research Part B: Methodological 45 (8), 1232 – 1250.

Bogetoft, P., Otto, L., 2011. Benchmarking with DEA, SFA, and R. Springer, New York.

Carbon Trust, 2007. Carbon footprinting: An introduction for organisations. Tech. rep., Carbon Trust, London.

Charles, V., Kumar, M., Kavitha, S. I., 2012. Measuring the efficiency of assembled printed circuit boards with undesirable outputs using data envelopment analysis. International Journal of Production Economics 136 (1), 194 – 206.

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

Charnes, A., Cooper, W. W., 1962. Programming with linear fractional functionals. Naval Research Logistics Quarterly 9 (3-4), 181–186.

Chiou, Y.-C., Lan, L. W., Yen, B. T., 2012. Route-based data envelopment analysis models. Transportation Research Part E: Logistics and Transportation Review 48 (2), 415 – 425.

Cooper, M. C., Lambert, D. M., Pagh, J. D., Nov. 2009. Supply chain management: More than a new name for logistics. International Journal of Logistics Management 8 (1), 1–14.

Cooper, W. W., Seiford, L. M., Tone, K., 2000. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and Dea-solver Software. Klewer, Massachusetts.

Cooper, W. W., Seiford, L. M., Tone, K., 2007. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and Dea-solver Software, 2nd Edition. Springer, New York.

Defra, 2012. Guidelines to Defra / DECC's GHG Conversion Factors for Company Reporting: Version 1. Defra, London.

Department for Transport, 2011. Road freight statistics 20011.

Dullaert, W., Zamparini, L., 2013. The impact of lead time reliability in freight transport: A logistics assessment of transport economics findings. Transportation Research Part E: Logistics and Transportation Review 49 (1), 190 – 200.

Erdoğan, S., Miller-Hooks, E., 2012. A green vehicle routing problem. Transportation Research Part E: Logistics and Transportation Review 48 (1), 100 – 114.

European Comission, 2011. Roadmap to a single european transport area towards a competitive and resource efficient transport system s. Tech. rep., European Comission, Brusses.

Färe, R., 1986. Addition and efficiency. The Quarterly Journal of Economics 101 (4), pp. 861–866.

Färe, R., Grosskopf, S., 1983. Measuring output efficiency. European Journal of Operational Research 13 (2), 173 – 179.

Farrell, M. J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General) 120 (3), 253–290.

Fox, K. J., 1999. Efficiency at different levels of aggregation: public vs. private sector firms. Economics Letters 65 (2), 173 – 176.

FTA, 2010. Logistics carbon reduction scheme, first annual report. Tech. rep., FTA, Tunbridge Wells.

FTA, 2012. Logistics carbon reduction scheme, second annual report. Tech. rep., FTA, Tunbridge Wells.

FTA, May 2013. Logistics carbon review. Tech. rep., FTA, Tunbridge Wells.

Guerrero, S. E., 2014. Modeling fuel saving investments and fleet management in the trucking industry: The impact of shipment performance on {GHG} emissions. Transportation Research Part E: Logistics and Transportation Review 68 (0), 178 – 196.

Guerrero, S. E., Madanat, S. M., Leachman, R. C., 2013. The trucking sector optimization model: A tool for predicting carrier and shipper responses to policies aiming to reduce ghg emissions. Transportation Research Part E: Logistics and Transportation Review 59 (0), 85 – 107.

Hanman, S., 1997. Efficiency at different levels of aggregation: public vs. private sector firms. Logistics Managemen 8 (2), 1 – 18.

Hassini, E., Surti, C., Searcy, C., 2012. A literature review and a case study of sustainable supply chains with a focus on metrics. International Journal of Production Economics 140 (1), 69 – 82.

Homburg, C., 2001. Using data envelopment analysis to benchmark activities. International Journal of Production Economics 73 (1), 51 – 58.

Jain, P., Cullinane, S., Cullinane, K., 2008. The impact of governance development models on urban rail efficiency. Transportation Research Part A: Policy and Practice 42 (9), 1238 – 1250.

Kao, C., 2010. Congestion measurement and elimination under the framework of data envelopment analysis. International Journal of Production Economics 123 (2), 257 – 265.

Larson, P., Elias, A., Viafara, J., 2013. Towards sustainable trucking: reducing emissions and fule consumption. Transportation Journal 52 (1), 108–120.

Lättilä, L., Henttu, V., Hilmola, O.-P., 2013. Hinterland operations of sea ports do matter: Dry port usage effects on transportation costs and {CO2} emissions. Transportation Research Part E: Logistics and Transportation Review 55 (0), 23 – 42.

Lau, K. H., 2013. Measuring distribution efficiency of a retail network through data envelopment analysis. International Journal of Production Economics 146 (2), 598 – 611.

Lee, K.-H., Saen, R. F., 2012. Measuring corporate sustainability management: A data envelopment analysis approach. International Journal of Production Economics 140 (1), 219 – 226.

Léonardi, J., Baumgartner, M., 2004. Co2 efficiency in road freight transportation: Status quo, measures and potential. Transportation Research Part D: Transport and Environment 9 (6), 451 – 464.

Lewis, H. F., Sexton, T. R., Aug. 2004. Network dea: efficiency analysis of organizations with complex internal structure. Comput. Oper. Res. 31 (9), 1365–1410.

Li, H., Lu, Y., Zhang, J., Wang, T., Jul. 2013. Trends in road freight transportation carbon dioxide emissions and policies in china. Energy Policy 57, 99–106.

Liimatainen, H., Pöllänen, M., 2010. Trends of energy efficiency in finnish road freight transport 1995–2009 and forecast to 2016. Energy Policy 38 (12), 7676–7686.

Liimatainen, H., Stenholm, P., Tapio, P., McKinnon, A., Dec. 2012. Energy efficiency practices among road freight hauliers. Energy Policy 50, 833–842.

Maudos, J., Pastorb, J., Péreza, F., Quesadaa, J., 2002. Cost and profit efficiency in european banks. Journal of International Financial Markets, Institutions and Money 12 (1), 33 – 58.

Mckinnon, A., 2015. Green Logistics, Ch. 11. Vol. 3 of McKinnon et al. (2015), pp. 237–251.

McKinnon, A., Browne, M., Piecyk, M., Whiting, A. (Eds.), 2015. Green Logistics. Vol. 3. Kogan Page, London.

McKinnon, A. C., 2007. Co2 emissions from freight transport in the uk. Report, Climate Change Working Group of the Commission for Integrated Transport, London.

McKinnon, A. C., Nov. 2009. Benchmarking road freight transport: Review of a government-sponsored programme. Benchmarking: An International Journal 16 (5), 640 – 656.

Merkert, R., Hensher, D. A., 2011. The impact of strategic management and fleet planning on airline efficiency a random effects tobit model based on dea efficiency scores. Transportation Research

Part A: Policy and Practice 45 (7), 686 – 695.

Merkert, R., Mangia, L., 2014. Efficiency of italian and norwegian airports: A matter of management or of the level of competition in remote regions? Transportation Research Part A: Policy and Practice 62 (0), 30 – 38.

Mirhedayatian, S. M., Azadi, M., Saen, R. F., 2014. A novel network data envelopment analysis model for evaluating green supply chain management. International Journal of Production Economics 147 (B), 544–554.

Normann, R., 2001. Reframing Business: When the Map Changes the Landscape. John Wiley & Sons, Ltd., Chichester.

Odeck, J., Bråthen, S., 2012. A meta-analysis of dea and sfa studies of the technical efficiency of seaports: A comparison of fixed and random-effects regression models. Transportation Research Part A: Policy and Practice 46 (10), 1574 – 1585.

Odeck, J., Hjalmarsson, L., 1996. The performance of trucks—an evaluation using data envelopment analysis. Transportation Planning and Technology 20 (1), 49–66.

Olesen, O., 1995. Some unsolved problems in data envelopment analysis: A survey. International Journal of Production Economics 39 (12), 5 – 36.

Pasiouras, F., Gaganis, C., Zopounidis, C., 2008. International evidence on the impact of regulations and supervision on banks' technical efficiency: an application of two-stage data envelopment analysis. Review of Quantitative Finance and Accounting 30 (2), 187–224.

Pérez-Martínez, P. J., 2012. Energy consumption and emissions from the road transport in spain: a conceptual approach. Transport 27 (4), 383–396.

Piecyk, M. I., 2015. Green Logistics, Ch. 3. Vol. 3 of McKinnon et al. (2015), pp. 55–79.

Piecyk, M. I., McKinnon, A. C., 2009. Analysing Global Energy Trends in Road Freight Transport. Proceedings of the 2009 Logistics Research Network Conference, Cardiff University, UK.

Piecyk, M. I., McKinnon, A. C., Dec. 2010. Forecasting the carbon footprint of road freight transport in 2020. International Journal of Production Economics 128 (1), 31–42.

Piecyk, M. I., McKinnon, A. C., Allen, J., 2015. Green Logistics, Ch. 4. Vol. 3 of McKinnon et al. (2015), pp. 80–103.

Ramanathan, R., 2005. An analysis of energy consumption and carbon dioxide emissions in countries of the Middle East and North Africa. Energy 30 (15), 2831–2842.

Ramanathan, R., Jun. 2006. A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions. Technological Forecasting and Social Change 73 (5), 483–494.

Rose-Anderssen, C., Baldwin, J., Ridgway, K., Allen, P., Varga, L., Strathern, M., 2009. A cladistic classification of commercial aerospace supply chain evolution. Journal of Manufacturing Technology Management 2 (2), 235 – 257.

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

Stewart, G., Nov. 2009. Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply-chain management. Logistics Information Management 10 (2), 62 – 67.

Thanassoulis, E., 2001. Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software. Klewer, Massachusetts.

Tone, K., Sahoo, B. K., 2003. Scale, indivisibilities and production function in data envelopment analysis. International Journal of Production Economics 84 (2), 165 – 192.

Tongzon, J., 2001. Efficiency measurement of selected australian and other international ports using data envelopment analysis. Transportation Research Part A: Policy and Practice 35 (2), 107 – 122.

Wong, W. P., Wongk, K. Y., 2008. A review on benchmarking of supply chain performance measures. Benchmarking 15 (6), 25–51.

Xu, J., Li, B., Wu, D., 2009. Rough data envelopment analysis and its application to supply chain performance evaluation. International Journal of Production Economics 122 (2), 628 – 638.

Yu, M.-M., 2008. Assessing the technical efficiency, service effectiveness, and technical effectiveness of the worlds railways through ndea analysis. Transportation Research Part A: Policy and Practice 42 (10), 1283 – 1294.

Zhao, Y., Triantis, K., Murray-Tuite, P., Edara, P., 2011. Performance measurement of a transportation network with a downtown space reservation system: A network-dea approach. Transportation Research Part E: Logistics and Transportation Review 47 (6), 1140 – 1159.

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

Acknowledgements

The authors gratefully acknowledge the EPSRC for funding this work through its financial support of The Centre for Sustainable Road Freight (EPSRC Reference: EP/K00915X/1).