ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis

submitted for the degree of

International Bachelor Econometrics and Operations Research

Allocating CO_2 emissions in an emission allocation & pollution routing game

Abstract

In the past years, growing road transport emissions play a big role in the increase in CO_2 emissions for the transport sector. To realise an overall emission reduction, a common methodology to calculate emissions generated in supply chains is necessary. This also benefits initiatives such as carbon footprint labelling that can inform customers on the transport emissions of their orders. This paper builds onto earlier works regarding the allocation of emissions to customer orders on a distribution route. We evaluate a proportional allocation method and four cooperative game theoretical (CGT) allocation methods in (multiple) routing settings using two CGT games from literature: the emission allocation and the pollution routing game. These settings also allow studying the effects that order characteristics have on the routing of vehicles, as they consume the truck capacities to a different degree, influencing the generated emissions on (a) tour(s). The methods are evaluated based on five game-theoretical fairness criteria: stability, consistency, equal treatment of equals, robustness and computation time. We find that the performance of the CGT methods changes in a setting that uses two trucks with limited volume capacity. In this setting, they capture the effect order volume has on the routing of the truck, by allocating more emissions to orders with bigger volumes, but they also allocate more emissions to heavier customer orders. The Star method is unable to capture these effects but performs best or second-best over all considered scenarios in terms of consistency, equal treatment of equals, and robustness, closely followed by one of the CGT methods.

by

Name student: Esmee Tanis
Student ID number: 445666

Supervisor: M.A. van Zon
Second assessor: A.P.M. Wagelmans
Submission Date: July 1, 2021

Contents

1 Introduction

At the beginning of 2021, the World Meteorological Organization (WMO) published a new state of the Global Climate report. Despite lockdown related falls in emissions because of the COVID-19 epidemic, global anthropogenic greenhouse gasses (GHG) continued to increase (WMO, 2020). Transportation is a big contributor to this, accounting for 24% of total global direct CO_2 emissions from fuel combustion (IEA, 2020). Even though emissions in Europe have fallen by 22.5% from 1990 to 2018, emissions in the transport sector did increase by more than 23% (Enzmann & Ringel, 2020). Road transport emissions appear to play the biggest role in this increase, growing nearly 27% over these years, and accounting for almost 95% of all transport emissions in 2018.

To realise the emissions targets set out by the EU in the European Green Deal (2019) and the Paris Agreement (2016), a common methodology to calculate the carbon footprint of supply chains is necessary to be able to compare the environmental performance of logistics suppliers and identify any emission reduction possibilities. In addition, this would benefit emissions reduction initiatives such as carbon footprint labelling, which has developed to be a meaningful instrument (Liu, Wang, & Su, 2016). Designed to gain consumer response and integrate sustainability issues into daily consumer behaviour, carbon footprint labelling can also benefit companies as it allows for communication with their customers and could generate emission savings (Liu et al., 2016).

This paper focuses on the CO_2 emission allocation of transport emissions to customer orders on (a) distribution route(s). For initiatives, such as carbon footprint labelling, it is important that generated CO_2 emissions associated with consumer products are allocated to individual orders fairly and rationally. Transportation emissions contribute to the carbon footprint of products as well. The produced emissions on the distribution tour(s) are calculated based on several factors, including travelled distance by the truck, its payload, speed and weight. We evaluate several emission allocation methods in a (multiple) vehicle routing setting using two cooperative game theoretical games from literature and aim to study the effects that the order characteristics have on the routing of the customers, and hence the generated emissions on the distribution tour(s).

In the past years, research has been done to find a common standard for emissions allocation. One important initiative is the European research project COFRET (Carbon Footprint of Freight Transport), which was started in June 2011 to provide transparency on and to harmonize existing methodologies to calculate the carbon footprint of freight transport (CORFET, 2015). They consider the European Norm EN-16258 by the European Committee for Standardization (CEN), as a promising initiative to facilitate a global standard for emission calculation. The EN-16258 contains a methodology to standardize the allocation of calculated emissions of transport operations over individual shipments (CEN, 2012). The norm recommends allocating emissions based on the product of the weight and the distance travelled to deliver a shipment, but proposes other physical units, instead of weight, for situations in which this might be more appropriate.

The use of simple proportional allocation rules, such as the ones suggested in EN-16258 prevail in practice, because the mathematics required to be able to allocate CO_2 emissions to customers on a distribution tour becomes 'unworkably complex' (Leonardi, McKinnon, & Palmer, 2010). Kellner and Schneiderbauer (2019) proclaim that the EN-16258 norm, although it is considered promising, still contains ambiguities. Yet, no globally accepted standard for the allocation of GHG emissions to shipments in road freight transportation exits (Kellner & Schneiderbauer, 2019).

Even though emission allocation problems are found to be complex, in the last three years, there has been a rapid increase in the number of publications on game theory-based models in green supply chain management (Agi, Faramarzi-Oghani, & Hazır, 2020). These type of methods are more sophisticated than the proportional allocation methods and use concepts from cooperative game theory (CGT). Naber, de Ree, Spliet, and van den Heuvel (2015) and Kellner and Otto (2012), among others, focus on such methods in their papers.

Naber et al. (2015) study the allocation of CO_2 emissions to customer orders on a single distribution route. The distribution route is predefined and heuristically constructed, and the emissions generated on this tour are calculated retrospectively. They introduce the emission allocation game to study the performance of four CGT emission allocation methods and compare these to a common proportional allocation method, which they refer to as the 'Star method'. Defining the emission allocation game requires knowing the emissions generated on any sub-tour for a coalition (subset) of customers. They approximate these sub-tour(s) based on the order in which customers are visited in the total distribution tour. Evaluations of the performance of the allocation methods are done based on game theoretical fairness criteria. In specific, they look at the stability, consistency, robustness and computation time of the allocation methods. A stable emission allocation ensures that no (subset of) customer(s) wants to withdraw its order based on the allocated emissions. Furthermore, when a customer order is larger than another one or located further away, allocated emissions should change accordingly. Robustness should also be guaranteed, as we aim to have similar emission allocations to identical orders over time. They show that the Star method performs worse on the stability and robustness criteria than the CGT allocation methods. However, it does have the best consistency and run time performance.

Kellner and Otto (2012) study 15 allocation methods to allocate CO_2 emission to shipments in road freight transportation. Several of these methods are CGT approaches, while others are ad-hoc, proportional allocation methods. They evaluate these methods in a short distance transport context and evaluate their performance based on several criteria: causality, efficiency, empty core robustness, symmetry, individual rationality, coalition stability, ease of application, and robustness. They find that transport distance is more important in the individual allocation of CO_2 emissions to shipments than the transported shipments mass. Furthermore, their results show that only CGT methods perform reasonably well on the acceptability criteria, but this comes at a cost of increased computational complexity. They similarly define the acceptability criteria as Naber et al. (2015) define the stability criteria.

In the majority of the literature, the objective to generate minimum emissions during performed transport activities is not incorporated into the optimization problem. Instead, the allocation of emissions is often done based on a retrospective calculation for a given routing plan (Guajardo, 2018). This is the case, for example, in the papers of Kellner and Otto (2012), Naber et al. (2015) and Leenders, Velázquez-Martínez, and Fransoo (2017). Kellner and Schneiderbauer (2019) state that their study "contributes to overcome this shortcoming and to fill the corresponding gap in research". They study the impact of shipment characteristics, such as origin, destination, weight, and volume, on the routing decision of logistics providers by incorporating the objective to minimize emissions in the optimization problem to find the optimal tour. The shipment characteristics influence, among others, the amount of capacity they take up in a vehicle, and hence the routing with associated GHG emissions.

In addition, for the application of CGT concepts in their study, Kellner and Schneiderbauer (2019) introduce the 'Pollution Routing Game' (PRG). The routing part of the PRG is connected to the pollution routing problem (PRP). It is an extension of the well-known vehicle routing problem (VRP), but instead of finding the minimum cost or minimum travel time tour, in the PRP the objective is to find a minimum emission tour. For more on the VRP, see Cordeau, Laporte, Savelsbergh, and Vigo (2007). In addition, similar to the vehicle routing game (VRG), the PRG uses concepts from CGT to allocate emissions to shipments on the route. The VRG is extensively studied to do this for cost allocations in a collaborative transport setting (Guajardo, 2018). These cost allocation schemes have been extended to allow for the allocation of GHG emissions in earlier research.

In three transport scenarios Kellner and Schneiderbauer (2019) evaluate the performance of the EN-16258 allocation principles based on game theoretical fairness criteria. The paper combines fairness criteria of earlier works to a comprehensive list of 10 different evaluation measures: efficiency, simplicity, computation time, individual rationality, coalition stability, core performance, routing stability, equal treatments of equal, consistency and robustness. A combination of the evaluation criteria individual rationality, coalition stability, and core performance, is similar to the stability criterion defined by Naber et al. (2015). They find that the allocation unit 'distance' performs best to the CGT allocation benchmark, for which they use the Shapley value. This allocation method is also used by Wick, Klumpp, and Kandel (2011) and Xu, Pan, and Ballot (2012). The Shapley Value, which is the only CGT allocation method they study, performs better on all fairness criteria than the proportional EN-16258 allocation principles, however, it requires exponential computational effort to calculate. This makes its use less pragmatic. Furthermore, their results show that although volume does not directly contribute to the generated CO_2 emission on a route, it does affect the routing of shipments, and hence the generated emissions on the tour(s). This effect proofs to be larger in their studied transport scenario than the effect of shipment weight on the generated emissions.

The work of Kellner and Schneiderbauer (2019) shows that using CGT allocation methods in a PRG requires exponential computational effort, as it requires knowing for each coalition of customers what the emissions are that are generated on an optimal sub-tour visiting this coalition. To define this for the PRG we need to solve an exponential number of PRP's. Özener (2014) study the Shapley Value using a similar type of game and propose a way to deal with this computational complexity. They introduce an approximation of the Shapley Value, in which they reduce the number of coalitions for which the PRP's need to be solved, by only including the coalitions for each customer that includes up to their 5 nearest neighbours. In addition to the approximation of the Shapley Value, they develop a duality-based allocation mechanism for both costs and emissions and study both in a single-vehicle routing setting. They compare these methods to two proportional allocation methods discussed in the literature. They find that the approximated Shapley Value allocation outperformed an individual solution based allocation method, comparable to the proportional allocation method used in Naber et al. (2015).

To conclude our discussion of several relevant papers, we want to emphasize some further research suggestions by Naber et al. (2015) and Kellner and Schneiderbauer (2019). Naber et al. (2015) suggest in their paper that it would be valuable to evaluate the performance of the CGT allocation methods studied in their paper, in a (multiple) vehicle routing setting. Kellner and Schneiderbauer (2019) propose to extend the analysis of the PRG by introducing new specific transport scenarios or by using additional CGT allocation methods, such as the Nucleolus, which is studied by Naber et al. (2015).

This paper aims to respond to these suggestions. We evaluate the performance of the Star method and the four CGT emission allocation methods studied by Naber et al. (2015) in (multiple) vehicle routing scenarios using the emission allocation game they introduce and the pollution routing game Kellner and Schneiderbauer (2019) introduce. These four CGT methods are the Shapley Value, the Lorenz+ Allocation, the EPM+ and the Nucleolus. We discuss them in Section 3.1. When we use the CGT allocations methods in the pollution routing game, we work with an approximation of these methods to deal with the computational effort of using them in this type of game. We measure performance by evaluating five game-theoretical fairness criteria: stability, consistency, equal treatment of equals, robustness and computation time. In addition, we aim to analyse the impact of the order characteristics (volume and weight) on the emission allocations to the customer orders in the considered scenarios. Lastly, we compare the two different emissions functions used by Naber et al. (2015) and Kellner and Schneiderbauer (2019) to study the effect that the chosen emission function has on the performance of the emission allocation methods.

Our main findings are the following. In each scenario, the Star method has the best or second-best performance on the equal treatment of equals, consistency and robustness criteria, closely followed by the EPM+. The Lorenz+ Allocation, EPM+ and Nucleolus have the overall best performance on stability. Moreover, the CGT methods have proportionally different emission allocations when we use them in a pollution routing game with two trucks with limited volume capacity. In either game, they allocate more emissions to heavier orders, as compared to a routing setting where all customer orders can be serviced on a single distribution tour. In addition, in a two truck pollution routing game setting, all methods, except the Star method, capture the effect order volume has on the routing of the vehicles, by allocating more emissions to high volume orders. Regarding the two emissions functions, no remarkable difference in the performance of the allocation methods was found when using either. Lastly, when allocating emissions in a pollution routing game, the computational complexity increases for a large set of customers, making the CGT allocation methods less pragmatic in this case.

The paper is structured as follows. First, we elaborate on the theoretical background of the problem in Section 2. In Section 3, the used methodology is introduced. We describe the different allocation methods and introduce the (pollution) routing model. The setup of our numerical experiments is discussed in Section 4, after which the results of our evaluations are examined. Finally, in Section 5, we share our conclusion, the limitations of our work and suggestions for further research.

2 Theoretical framework: Cooperative game theory

Game theoretical concepts have been widely used to study supply chains (Agi et al., 2020). In this section, we introduce the emission allocation game, similar to Naber et al. (2015), the pollution routing game, which is introduced by Kellner and Schneiderbauer (2019), and an important CGT concept, called the core of the game.

We first introduce some standard notations, as used by Guajardo (2018). We have a set $N = \{1, \ldots, n\}$, of customers with associated ordered goods, with i being the index of a single customer. We call N the grand-coalition, and S represents a possible feasible coalition (subset) of N, with $S \subseteq N$.

A coalition S is feasible if the customer orders in the coalition, do not exceed any of the maximum capacities of a delivery truck, either in terms of volume or weight. The customers in a coalition can be visited on one or multiple distribution routes, depending on the set of trucks R, with $r \in R$ we have available and the truck capacities. The customers in the coalition S can be visited in a sequence $\sigma(S)$, a permutation of S. We can break up this sequence into a maximum of $|R|$ tours, that all start and end at a depot, to satisfy the vehicle capacity constraints.

To find the optimal route, $\sigma^*(S)$ we need to solve an optimization problem:

$$\sigma^*(S) \in \underset{\sigma(S)}{\arg\min}\, \{f(\sigma(S))\}. \tag{1}$$

The function $f(\sigma(S)$ can be defined in multiple ways, but we consider it to be either the distance travelled when visiting the customers in coalition S in order $\sigma(S)$, which we denote $d(S)$, or the emissions generated on (a) tour(s) visiting the customers in coalition S in order $\sigma(S)$, which we denote $e(S)$. Here, $e(.)$ is the selected emission function.

2.1 The emission function

We consider two different emission functions. One is a velocity and payload dependent emission function for heavy-duty road freight transportation, developed by Ligterink, Tavasszy, and Lange (2012), and used by Naber et al. (2015). This emission function can be found in Appendix A, and we label it 'Ligterink's emission function'. It takes the impact of the weight of the ordered goods, the distance between the customer orders and the driving speed of a truck into account. The other one is a more general commonly accepted approach, which approximates the generated emissions by averaging the emissions generated by an empty and fully loaded truck (Guajardo, 2018). This emission function is used by Kellner and Schneiderbauer (2019), among others and presented in Appendix B. We refer to it as 'the general emission function'. In addition to the goods' weights and driven distance, this function also incorporates the vehicle fuel consumption per kilometre and an energy conversion factor.

2.2 The emission allocation game

To define the emission allocation game, we need to specify $e(S)$ for each feasible coalition, $S \subseteq N$. This means we need to find (an) optimal sub-route(s), $\sigma^*(S)$, for each coalition S, and compute the generated emission on the route(s). For the emission allocation game, we only find the optimal route $\sigma^*(N)$ for the grand-coalition, similar to Naber et al. (2015). They heuristically compute this grand-tour using a commercial software package, developed by TNO. We consider two cases, where we either find the minimum distance grand-tour or the minimum emission grand-tour. We refer to Section 3.3 for the mixed-integer linear programming (MIP) problem we solve to find these optimal tours. To define $e(S)$

for each coalition, we find the sub-route(s) visiting the customer(s) in this coalition by assuming they are visited in the order of $\sigma^*(N)$, whereby we omit customer(s) not in this coalition. Then, for the sub-route(s), we compute $e(S)$. A benefit of this approach is that defining the game in this way requires low computational effort, although the game might be somewhat less accurately specified.

2.3 The pollution routing game

The pollution routing game is a more specified version of the emission allocation game. Similarly, we need to specify $e(S)$ for each feasible coalition, $S \subseteq N$. This time, we do this by solving the MIP problem (see Section 3.3) for each coalition. In this game, we only specify $f(\sigma(S))$ to be the emissions generated on (a) tour(s), not the distance travelled. A downside of this type of game is that it has exponential computational complexity, as we need to solve $2^n - 1$ MIP problems, for each possible non-empty feasible coalition. In Kellner and Schneiderbauer (2019) it took several hours per test instance with 14 to 16 customers to compute the objective function of a MIP problem, similar to the model introduced in Section 3.3, for each coalition of customers. To reduce the computational effort of defining the pollution routing game for test instances with a similar number of customers, we aim to reduce the number of coalitions for which we need to solve a MIP problem to find $e(S)$.

We adopt the idea of Özener (2014) for the approximation of the Shapely Value in a similar type of game context. For test instances with around 15 customers, we define the pollution routing game for only the coalitions that include up to the 5 nearest neighbours for each customer. For a customer i, we check its distance to customer $j \neq i$, $x_{ij}\delta_{ij}$. We rank these distances from low to high and only include the top five customers j into the coalition associated with customer i. For all possible subsets of these coalitions, we find $e(S)$ by solving the MIP problem. This means that, except for the grand-tour, we never find the minimum emission tour(s) for a set of customers that has a size above six, significantly reducing the computational effort. In Appendix D, we evaluate if this number of nearest neighbours produces a reasonable approximation to the emission allocation of each of the CGT allocation methods in our test scenarios, and we summarize these results in Section 4.1.1.

2.4 The core of the game

When we have specified the emission allocation and the pollution routing game, we can use cooperative game theoretical concepts to find the emission allocations to the customer orders. In the next section, we elaborate on the emission allocation methods we use for this. We wish that these methods produce stable emission allocations. Here, we use a similar definition of stability as used by Naber et al. (2015). An allocation of emissions is defined by $x = (x_i)_{i \in N}$, where $x(S) = \sum_{i \in S} x_i$. Such an allocation is stable if it in the core of the game, which is characterized by

$$core(e) = \{x \in^n : x(N) = e(N); x(S) \leq e(S), \forall S \subset N\}. \quad (2)$$

Note that in this definition, negative allocations are allowed. Another remark is that the core may be empty when a subset of customers has an incentive to leave the grand-coalition because it will be allocated fewer emissions by doing so. In both Naber et al. (2015) and Kellner and Schneiderbauer (2019) empty core games are observed.

In Section 3.2, we discuss other evaluation criteria used to evaluate the allocation methods' performance.

3 Methodology

In Section 3.1, we first elaborate on the emission allocation methods that we evaluate in this paper. Then we introduce the game-theoretical fairness criteria that we use to evaluate the performance of these methods in Section 3.2. Lastly, in Section 3.3, we introduce the MIP problem that we use to find the optimal grand-tour(s) in the emission allocation game and the optimal tour(s) for each coalition in the pollution routing game.

3.1 Allocation methods

Naber et al. (2015) study four CGT emission allocation methods: the Shapley Value, the Lorenz+ Allocation, the Equal Profit Method+ and the Nucleolus. These are compared to the Star method, a benchmark proportional emission allocation method. In the rest of this section, we briefly discuss these methods.

Star method: This methods allocates to every customer order, $i \in N$, the emission amount given by

$$x_i = \frac{e(\{i\})}{\sum_{i \in N} e(\{i\})} e(N).$$

with $e(\{i\})$ being the stand-alone emission in case a customer was visited separately by a truck. Instead of the stand-alone emission, also other characteristics of the order can be used for the proportional allocation. A benefit of this method is that it is easily understandable and implementable, but this comes at a cost, as the allocations generated by the Star method are not guaranteed to be stable, and thus are not necessarily part of the core. This method is independent of the sequence in which customers are visited or the distance between them.

Shapley value: The allocation of emissions in this method, is based on the average marginal emission of a customer entering all possible coalitions (Shapley, 1953). The marginal emission of adding customer i to subset S, $m_i(S)$, is computed by $e(S \cup \{i\}) - e(S)$. The allocated emission amount is then computed by

$$x_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} m_i(S).$$

According to Guajardo and Rönnqvist (2016), this is one of the most used concepts in cooperative games. A benefit of the Shapley Value is that its solution is unique. however, it is not necessarily part of the core.

Nucleolus: This method is concerned with satisfying individual rationality, by producing an emission allocation that minimizes the maximum dissatisfaction of any coalition (Schmeidler, 1969). We introduce the concept of the excess of a coalition $\varepsilon(S, x)$. It is defined by $\varepsilon(S, x) = e(S) - x(S)$. We denote $\theta(x)$ to be the vector of excesses for each coalition of the allocation x, arranged in non-increasing order. The Nucleolus is then defined as the allocation x, which is efficient, and individually rational, in a way that $\theta(x)$ is the lexicographic minimum among all allocations. The lexicographic order is a generalization of the alphabetical order, where for example [1, 12, 4, 3] is lexicographically smaller than Q = [1, 12, 8, 2], and [1, 3, 12, 2] is lexicographically smaller than [1, 4, 1, 1]. The Nucleolus lies in the core if the core is unique and non-empty. As opposed to proportional allocation methods, and the Shapley Value, the computation of the Nucleolus is more complex, and it is commonly mistaken (Guajardo & Jörnsten, 2015). To find the Nucleolus, we use an approach by Engevall, Göthe-Lundgren, and Värbrand (2004), which is similarly adopted by Naber et al. (2015). We solve different linear programming (LP) problems consecutively and using duality theory, we first find an allocation where we maximize the minimum excess. Next, we check the dual variables of this solution, and we fix the excesses for all coalitions that have strictly positive dual variables. Then, given these coalitions with a fixed excess, we repeat the procedure and maximize the minimum excesses of the remaining coalitions. The solution procedure continues until we find a (unique) solution. We implement the LP problem used in Naber et al. (2015) to find the Nucleolus. It can be found in Appendix C.

Lorenz+ Allocation: A variant on the more commonly known "egalitarian solution" of Dutta and Ray (1989), this method produces a core allocation with a minimal difference between the biggest and smallest allocation to any customer order. The method was introduced by Arin Aguirre (2003) as "Leximin". It can be found by solving an LP problem, with f the smallest absolute difference in allocated emissions:

$$\max \quad f, \tag{3}$$

$$\text{s.t.} \quad x_i - x_j \leq f \qquad \forall i, j \in N, \tag{4}$$

$$x(S) \leq e(S) \qquad \forall S \subset N, \tag{5}$$

$$x(N) = e(N). \tag{6}$$

From constraints (5) and (6) it follows that the Lorenz Allocation does not exist in case the core is empty. Hence, Naber et al. (2015) propose to use the Nucleolus method if the core is empty, to preserve the continuity of the allocation method. This results in the Lorenz+ Allocation.

Equal Profit Method+ (EPM+): This allocation method, proposed by Frisk, Göthe-Lundgren, Jörnsten, and Rönnqvist (2010), is similar to the Lorenz Allocation, but in this case we minimize the largest difference between the allocation in the coalition, relative to the stand-alone emission for a customer order. This allocation can be found by solving the following LP problem, where g denotes the largest relative difference:

$$\max \quad g, \tag{7}$$

$$\text{s.t.} \quad \frac{x_i}{e(\{i\})} - \frac{x_j}{e(\{j\})} \leq g \qquad \forall i, j \in N, \tag{8}$$

$$x(S) \leq e(S) \qquad \forall S \subset N, \tag{9}$$

$$x(N) = e(N). \tag{10}$$

Similar to the Lorenz Allocation, the EPM does not provide a solution when the core is empty. We also use the Nucleolus method for this case, and we refer to this method as the EPM+ allocation.

3.2 Evaluation criteria

A trade-off should be considered between the ease of use and understandability of an allocation method, versus its performance in terms of the fairness of the emission allocation relative to each customer order. Hence, we evaluate the emission allocations produced by the allocation methods based on several criteria. We refer to Table 1 for an overview. As opposed to Kellner and Schneiderbauer (2019), we do not use the criteria *Simplicity* and *Routing Stability* in our evaluation. The *Simplicity* criterion is closely linked to the *Duration criterion*, and none of the allocation methods we consider takes the routing sequence into account when distributing the total amount of CO_2 emissions.

Table 1: Evaluation criteria for the allocation methods.

Criteria	**Description**
1) *Stability*	A subset of customers should have no incentive to withdraw from the route based on the allocated emissions.
2) *Consistency*	The allocated CO_2 emissions should change accordingly if the characteristics of an order change.
3) *Equal treatment of equals*	Orders with similar characteristics should be allocated similar amounts of CO_2 emissions.
4) *Robustness*	Customers that receive similar orders periodically, should be allocated similar amounts of CO_2 emissions.
5) *Computation time*	For frequent use of the method, a low computation time is preferred.

In Section 4.2 we elaborate on how to calculate the criteria presented in table above. We report on the more general concept of *Stability* similarly to Naber et al. (2015). Stable allocations are in the core of the game. In Kellner and Schneiderbauer (2019) they decompose stability into *Efficiency*, *Individual Rationality*, and *Coalition Stability*. Where efficiency is defined as, $\sum_{i \in N} x_i = e(N)$, individual rationality is defined as, $x_i \leq e(\{i\}), \forall i \in N$, and coalition stability is defined as, $\sum_{i \in S} x_i \leq e(S), \forall S \subset N$. We note that by definition all methods fulfil the *Efficiency* criterion.

3.3 Mixed-integer linear programming problem: (Pollution) Routing model

We now present an adaptation of the Cumulative Capacitated VRP with Pick-ups and Deliveries as introduced by Kellner and Schneiderbauer (2019). We refer to this MIP problem as the '(pollution) routing model' and it is used to find the optimal grand-tour(s) in the emission allocation game and the sub-tours for (a set of) feasible coalitions $S \subseteq N$ in the pollution routing game.

We define a graph, $G = (V, A)$. The set of vertices, V contains the depot, 0, and the set of customer $1, \ldots, n$, i.e. $V_1 = \{1, 2, \ldots n\}$ and $V = \{0\} \cup V_1$. We introduce a set of trucks R, with $r \in R$, to allow for a routing setting with multiple trucks. In contrast to Kellner and Schneiderbauer (2019), we do not incorporate the daily operating time constraint in our model, as we consider a relatively small customer service area (see Section 4.1). Furthermore, we also neglect the time for unloading activities.

The following decision variables are used in the model. A binary variable x_{ijr} to indicate if arc (i, j) is traversed by truck r, or not ($x_{ijr} = 0$). Another binary variable, y_{ir}, to indicate if a customer i is visited by truck r, or not ($y_{ir} = 0$). Lastly, we use two continuous variables dw_{ijr} and dv_{ijr} to keep track of the weight-based and volume-based delivery load transported on arc (i, j) by truck r.

Furthermore, the model contains several parameters. We have the weight-based and volume-based delivery request of customer i, respectively ω_i and μ_i. For the truck, we have a maximum order weight, W, and a maximum order volume, V that it can transport. Finally, δ_{ij} defines the distance between location i and j.

As mentioned in Section 2.2, for the emission allocation game, we want to either find (a) minimum distance tour(s), or (a) minimum emission tour(s). Hence, we introduce the model with two objective functions. The distance travelled, $d(N)$, on (a) minimum distance grand-tour(s) can be found by solving the model with the following objective function:

$$\min \quad \sum_{i,j \in V} x_{ij} \delta_{ij} \tag{11}$$

The minimum emission tour(s) can be found by solving the model using the general emission function (see Appendix B) in the objective function:

$$\min \quad \sum_{i \in V} \sum_{j \in V} \sum_{r \in R} x_{ijr} \frac{\delta_{ij}}{100} (FC_{empty} + (FC_{full} - FC_{empty} \frac{dw_{ijr}}{W}) ECF \tag{12}$$

We note here that in case we want to use Ligterink's emission function in the objective (see Appendix A), we need to add the following constraint:

$$KWt_{ijr} = \frac{131.25}{5 + 0.01 dw_{ijr}} \forall i, j \in V, r \in R \tag{13}$$

Unfortunately, this constraint is non-linear, which makes this emission function unsuitable to use.

The (pollution) routing model is further defined with the following constraints:

$$\text{s.t.} \quad \sum_{j \in V} x_{ijr} = y_{ir} \qquad \forall i \in V, r \in R, \tag{14}$$

$$\sum_{j \in V} x_{jir} = y_{ir} \qquad \forall i \in V, r \in R, \tag{15}$$

$$\sum_{r \in R} y_{0r} \leq |R| \tag{16}$$

$$\sum_{r \in R} y_{ir} = 1 \qquad \forall i \in V_1, \tag{17}$$

$$\sum_{j \in V} dw_{jir} - \sum_{j \in V} dw_{ijr} = \omega_i y_{ir} \qquad \forall i \in V_1, r \in R, \tag{18}$$

$$\sum_{j \in V} dv_{jir} - \sum_{j \in V} dv_{ijr} = \mu_i y_{ir}, \qquad \forall i \in V_1, r \in R, \tag{19}$$

$$\omega_j x_{ijr} \leq dw_{ijr} \leq (W - \omega_i) x_{ijr} \qquad \forall i, j \in V, r \in R, \tag{20}$$

$$\mu_j x_{ijr} \leq dv_{ijr} \leq (V - \mu_i) x_{ijr} \qquad \forall i, j \in V, r \in R, \tag{21}$$

$$\omega_0 = 0, \mu_0 = 0, \tag{22}$$

$$x_{ijr} \in \{0, 1\}, \tag{23}$$

$$dw_{ijr}, dv_{ijr} \geq 0. \tag{24}$$

Constraints (14) and (15) ensure that a truck enters and leaves each visited customer. The number of routes that can be started is limited by constraint (16) to at most $|R|$, and restriction (17) ensures each customer is visited by at most one truck. Constraints (18) and (19) are flow conservation constraints. They respectively ensure that the weight- and volume-based delivery load in a truck is increased on the arc leaving a visited customer, by the amount of weight- and volume-based delivery request of the specific customer. These constraints also ensure that no sub-tours can occur in the distribution route. Constraints (20) and (21) restrict the cumulative weight- and volume-based delivery load transported in a truck over an arc. The weight- and volume-based delivery request of the depot is set to zero in restriction (22). Lastly, constraints (23) and (24) are binary and non-negativity constraints.

Before concluding this section, we want to remark that instead of solving the model with the above-presented constraints using the minimum distance objective function (11), we can also leave out constraints (18) - (22), and (24), and instead add sub-tour elimination constraints, according to the well-known Miller-Tucker-Zemlin (MTZ) formulation (Miller, Tucker, & Zemlin, 1960).

4 Numerical experiments

Section 4.1 of this chapter elaborates on the four scenarios we use to evaluate the performance of the emission allocation methods using two sets of randomly generated test instances. Furthermore, we elaborate on the parameter settings of these sets of test instances and the target customers we study. Next, in Section 4.2 we clarify the computation of the evaluation criteria. Finally, we present the results of our numerical experiments in Section 4.3.

4.1 Set-up of the numerical experiments

In Table 2 we first give an overview of the characteristics of the four scenarios in which we test two sets of random test instances. For each scenario, we make a choice regarding the set of test instances we evaluate, the type of emission function we use and the parameter settings for the (pollution) routing model (introduced in Section 3.3). We elaborate on these choices in the remainder of this section. We evaluate all the five allocation methods in each scenario unless mentioned otherwise in the column 'Additional features'. We note that the Star method allocates similar emissions to all customer orders in Scenario 2 and 3, and therefore we only report the output of this method for Scenario 2.

Table 2: Summary of the characteristics for the considered scenarios.

Scenario	Type of game	Evaluated with set of test instances with ... customers	Emission function, $e(.)$	Optimization method to find the optimal tour	Objective for the optimal tour	# of trucks, $\|R\|$	Additional features
1	Emission allocation	10	Ligterink's	MIP Model	Min. distance (11)	1	
	Emission allocation	10 & 14	General	MIP Model	Min. distance (11)	1	
2	Emission allocation	10	Ligterink's	Exact	Min. emission	1	
	Emission allocation	10 & 14	General	MIP Model	Min. emission (12)	1	
3	Pollution Routing	10 & 14	General	MIP Model	Min. emission (12)	1	With 14 customers we use the approximation as described in Section 2.3 using 6 nearest neighbours & exclude the Nucleolus from evaluation
4	Pollution Routing	10	General	MIP Model	Min. emission (12)	2	Volume of all orders doubled

For the MIP problem, see Section 3.3.
The numbers in brackets indicate the number of the objective function in Section 3.3.

4.1.1 The number of customers: two sets of test instances

Based on preliminary research we find that we can perform the emission allocation using all allocation methods in each scenario in reasonable run-time with up to 10 customers. With this small number of customers, the computational effort to define the pollution routing game did not cause any problems. Hence, there was no need to use an approximation to reduce the number of coalitions in the game, as described in Section 2.3. We also evaluate a set of test instances with a larger customer base, where we chose 14 customers based on conducted experiments. From preliminary results, we find that with this number of customers, the computational complexity to define the pollution routing game creates issues with regards to the run-time and out of memory errors. Therefore, there is a need to reduce the number of coalitions for which we need to solve the (pollution) routing model in this case. We use the approximation described in Section 2.3 using 6 nearest neighbours in Scenario 3, but exclude the Nucleolus from our evaluation as the approximation has bad performance for this method. In addition, we do not evaluate the set of test instances with 14 customers in Scenario 4, as the approximations did not work for any of the methods in this scenario. We refer to Appendix D for a more elaborate evaluation of the quality of the used approximation with regards to the performance of the CGT emission allocation methods in Scenario 3 and 4 using a varying number of nearest neighbours. We note that in Scenario 1 and 2, there is no need to use such an approximation, as in these scenarios we use the emission allocation game, for which we do not need to solve an exponential number of (pollution) routing problems, but only a single one for the grand-tour.

Another result from preliminary research is that it requires too much memory space to find the minimum emission distribution tour visiting 14 customers in an exact manner using Ligterink's emission function. Therefore, we only use the general emission function to evaluate the emission allocation methods for the set of test instances with 14 customers.

4.1.2 An elaboration on the four scenarios

In Scenario 1 we find the minimum distance tour for the grand-coalition. For the set of problems instances with 10 customers, we compute and allocate the emissions generated on this tour using both emission functions to compare their performance. In this scenario, we assume that the weight of the orders shipped by the truck never exceeds its capacity. In addition, as emissions are not incorporated in the optimization problem, this scenario does not take into account the influence of the order characteristics on the routing of the truck.

In Scenario 2, we also analyze the performance of both emission functions for the set of test instances with 10 customers. As stated in Section 3.3, we cannot use Ligterink's emission function in the (pollution) routing model to find the minimum emission grand-tour. Hence, we find the grand-tour for this emission function in an exact manner, where we enumerate all possible orders in which we can route the customers on a distribution tour and check the generated emissions for each tour. We select the tour with the lowest emissions, to find $e(N)$ for the grand-coalition.

Both Scenario 2 and 3 allow us to study the influence of the order characteristic weight on the routing of the truck to a certain degree. This is the case because the payload of the truck influences the generated emissions on the tour, and every single order consumes the truck weight capacity to a different degree. This influences the order in which the customers are visited in the minimum emission grand-tour. We note that in these scenarios, we pick the truck order weight and volume capacity in a way that we can always create one tour visiting all customers. In Scenario 3, where we solve the (pollution) routing model for each coalition, the order characteristics also influence the routing of the sub-tours.

Scenario 4 is similar to Scenario 3, but now we double the order volumes for all customer orders for each test instance and pick the truck volume capacity in a way that we always need to use two trucks to visit all customers in the grand-coalition. We set the truck volume capacity in each test instance to a value that is slightly more than half of the total volume consumed by the customer orders. This allows us to study the effect that the characteristics of a single order might have on the CO_2 emission of another route, as a single order now consumes the truck volume capacity to a different degree.

4.1.3 The parameter settings for the test instances

Next, we discuss the settings of the parameters in our (pollution) routing model. These vary over the different test instances but are the same in each scenario. Table 3 summarizes these possible settings. In a simulation, we generate 30 test instances for the set of test instances with 14 customers and 15 for the set of test instances with 10 customers. A smaller set of test instances was used for the set of test instances with 10 customers, as it takes more time per test instance to define the pollution routing game when no approximation to the total number of coalitions is used. In the simulation, we study the average effects on the allocated emissions to three target customers (see Section 4.1.4) in each scenario for the different allocation methods.

Table 3: Parameter settings for the test instances.

Parameters	Settings
Depot	Number: 1
	Location: x, y-coordinate $(0, 0)$
Customers	Number: 10 or 14
	Location: randomly generated x, y-coordinates between 5 and 15 (uniformly distributed)
Order weight *(in loading units* = 0.01 *ton)*	Low (1.4), medium (7.1) or high (35.7)
Order volume *(in whole pallets)*	Low (0.06), medium (0.3) or high (1.5)
Truck	Truck weight: 5 ton
	Maximum capacities: 507 loading units & 21 pallets*
	Driving speed: 35 km/h**

Naber et al. (2015) use similar parameter settings for the order weight, truck weight, maximum loading units and truck velocity.
* In Scenario 4, the volume capacity is chosen to be slightly more than half of the total order volume for both trucks.
** The truck driving speed is an input variable in Ligterink's emission function, see Appendix A.

The location coordinates are chosen to represent a small city setting, where the city covers an area of 10 by 10 kilometres and the depot is located 5 kilometres outside of the city, at coordinate (0,0). For all customers, the location coordinates are picked randomly. In Scenario 1, 2 and 3 one truck with medium size container leave the depot, and in Scenario 4, two similar-sized trucks leave the depot, but these have limited volume capacity. From preliminary results, it follows that the service area set-up has some influence on the way the emission allocation methods incorporate the effect of the order characteristics in their emission allocation. However, the performance of the methods with respect to the evaluation criteria does not vary much when comparing two different service area set-ups. We refer to Appendix E for a more elaborate analysis of this performance comparison between two service area set-ups.

4.1.4 The (target) customers

In our evaluation, we have three target customers for which the order characteristics remain fixed over all randomly generated test instances: 1) a customer with low order weight (14 kg), low order volume (0.06 pallet) and a fixed distance to the depot of 15 kilometres, 2) a customer with low order weight, high order volume (1.5 pallets), and a 15-kilometre distance to the depot, and 3) a customer with high order weight (357 kg), low order volume and a 15-kilometre distance to the depot. These target customers lie on a circle with a radius of 15-kilometre with respect to the depot. Furthermore, for the evaluation criteria *Consistency*, we fix some of the characteristics of an additional customer in all randomly generated test instances. This customer (4) has low order weight and volume, and varying distance to the depot. For the other customers, the order weight and volume category are picked randomly.

4.2 Computation of the evaluation criteria

For the computation of the evaluation criteria, we look at the work of Kellner and Schneiderbauer (2019). For the *Stability* criterion, we check if the emission allocations produced by the various allocation methods are part of the core of the game, see Section 2.4. We note here that when we use an approximation to the number of coalitions in the pollution routing game this affects the interpretation of the *Stability* criterion. The core is no longer properly defined, as we reduce the set $S \subset N$, which results in a reduction of the number of $x(S) \leq e(S)$ constraints, that are part of the definition of the core. This enlarges the feasible region. In addition, such an approximation affects the Shapley Value, which loses its theoretical property of efficiency. This also influences the stability performance of this method. To ensure the Shapley Value still produces efficient emission allocations, we multiply the allocated emissions by a correction factor based on the deviation between the allocated emissions and the generated emissions on each sub-tour for each coalition. This ensures that all emissions generated on the grand-tour(s) are allocated over the customer orders.

To evaluate the criterion *Equal treatment of equals*, we create within each test instance two orders with the same characteristics and see if they get allocated the same proportion of total CO_2 emissions. We use a margin of 0.01 here to avoid this criterion being rejected based on a small difference in the many numbers behind the decimal when we check the equality of allocated emissions. The characteristics of the two orders can vary over different test instances but have to be the same within a specific test instance.

The *Consistency* criterion is examined by tracking the emissions allocated to target customers 1, 3 and 4 in each test instance over all test instances. Target customer 3 has an order with a heavier weight as compared to target customer 1, and is expected to be allocated more CO_2 emissions. Moreover, target customers 4 and 1 have similar order weight and volume, however, we expect target customer 4 's order to be allocated more (less) emissions than target customer 1's when it has a larger (smaller) distance to the depot. By separately comparing the emissions allocated to these three target customers, we can also independently evaluate if an emission allocation is consistent in terms of weight or distance to the depot. If the emission allocation of a method is both consistent with order weight and distance to the depot, we consider the method's performance 'consistent' for the specific test instance. We also look at the effect of order volume on the emission allocation, by checking the proportion of emission allocated to target customer 2's order. We expect this effect to become more apparent in Scenario 4, because of the tighter constraints on the available volume capacity in each truck.

To evaluate the *Robustness* criterion, we study the allocation of emissions to target customer 1's order over multiple test instances. The order characteristics of the non-target customers change, while they remain the same for the target customers. We would expect a similar CO_2 emission allocation to target customer 1 if the order characteristics remain identical. Robustness is then measured as the standard deviation divided by the average allocated emission, which is referred to as the CoV.

4.3 Results

This section presents the results of our computational study. All results are obtained using an Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz with 8 GB of RAM. The MIP and LP problems are solved in Eclipse IDE Version 2020-12 (4.18.0), using CPLEX 12.1 with Java as the programming language.

4.3.1 Characteristics of the four scenarios for both emission functions

Table 4 presents characteristics of the four scenarios for the two sets of test instances. First, the core turns out to be non-empty for both sets of test instances in Scenario 1, 2 and 3. As opposed to the other scenarios, Scenario 4 has 5 test instances with an empty core. In this scenario, two trucks with limited volume capacity are used to service all customer orders. We see that the average combined length of the resulting two tours is longer than in Scenario 1 and 2, with associated higher average emissions per tour. In addition, the average emissions per kilometre computed using the general emission function in Scenario 2 (0.41), is lower than in Scenario 4 (0.45). These two shorter tours are possibly less efficient in terms of CO_2 emitted, as compared to servicing all customers on a single tour. This could cause the average emissions allocated to a customer order when cooperating in these grand-tours in Scenario 4 to be higher than when it would not cooperate, making the allocation unstable, and certain coalitions possibly more efficient. We expect this to increase the number of test instances with an empty core.

Regarding the set of test instances with 10 customers, we see a larger computation time to define the pollution routing game (in Scenario 3 and 4), as compared to the emission allocation game (in Scenario 1 and 2). This is what we would expect, as we need to solve an exponential number of pollution routing problems to define the pollution routing game and only one for the emission allocation game. When we use the approximation discussed in Section 2.3 to reduce the number of coalitions for which we need to solve a pollution routing problem in Scenario 3 for the set of test instances with 14 customers, we observe a lower computation time to define the pollution routing game, despite the increase in the number of customers.

Table 4: Characteristics of the scenarios for both sets of test instances.

Scenario	Random instances with	Emission function type	% of empty core games	Avg. comp. time to define the game (s)	Avg. emissions per tour (kg)	Avg. total length of tour(s) (km)
1	*10 cust.*	1	0.00	1.00	19.00	45.96
		2	0.00	0.67	20.73	45.96
	14 cust.	2	0.00	12.60	23.01	50.25
2	*10 cust.*	1	0.00	1.07	18.91	45.96
		2	0.00	0.53	20.67	45.96
	14 cust.	2	0.00	11.63	22.89	50.27
3	*10 cust.*	2	0.00	30.91	18.91	45.96
	*14 cust.**	2	0.00**	25.35	22.89	50.27
4	*10 cust.*	2	33.33	94.42	32.22	72.28

The reported values for the test instances with 10 customers are averages over 15 random test instances, and for the test instances with 14 customers over 30 random test instances.
* The reported values for Scenario 3 for the set of test instances with 14 customers are computed using an approximation to the total number of coalitions including only the 6 nearest neighbours for each customer, see Section 2.3.

Lastly, we want to note the very slight reduction in average emissions per tour in Scenario 2, as compared to Scenario 1. For Ligterink's emission function this reduction is 0.48% and for the general emission function for the sets of test instances with 10 and 14 customers, respectively 0.29% and 0.52%. For the set of test instances with 14 customers, we also see a slight increase of 0.04% in the average total length of the tour in Scenario 2, as compared to Scenario 1. This is what we would expect to see, as in Scenario 2 we find the minimum emission tour, while in Scenario 1 we find the minimum distance tour for the grand-coalition. However, the difference is only minor. This suggests that for both emission functions, the distance travelled by the truck has the most influence on the average generated emissions of a tour, while the payload of the truck has only a minor effect. We expect the difference in the input parameters and the chosen values for these parameters in both emission functions to explain the differences in the average emissions generated on a tour. In Appendix E we briefly elaborate on this by evaluating, among others, the sensitivity of our results concerning these chosen parameter values.

Before we present the results of the performance of the allocation methods for the two sets of test instances in the four scenarios, we want to note that we separately checked the evaluation criterion *Individual Rationality*, part of the *Stability* criterion. Kellner and Schneiderbauer (2019) argue that performance on this criterion is more important for an allocation method than producing allocations that are part of the core because it is easier to verify for a customer if it would benefit to cooperate and get lower or similar emissions than their stand-alone emissions. We find that all allocation methods satisfy this criterion for our evaluated test instances in each scenario.

4.3.2 Performance of the allocation methods in the four scenarios with 10 customers

We start with the results for the set of test instances with 10 customers, in case we use the general emission function. Table 5 contains these results. We first discuss the performance of the allocation methods with respect to the evaluation criteria and then make some remarks regarding the average emission allocations to the target customers by the various emission allocation methods in the different scenarios.

The stability performance : The performance on this criterion is related to the number of empty core instances. When the core is non-empty, the Lorenz+ Allocation, EPM+ and Nucleolus guarantee a solution in the core, and their allocations are stable by definition. Kellner and Schneiderbauer (2019) found that although the Shapley Value does not incorporate any core criteria in its solution, it was in the core in almost all (98.7%) of their instances. For the Star method, they found that this was the case in only 34.9% of their test instances. The values in Table 5 for Scenario 1, 2 and 3 confirm their result for Shapley Value. In addition, we find a reasonably high percentage of test instances for which the Star method has stable allocations. Only in Scenario 4, the emission allocations of the Shapley Value and the Star method are not part of the core of the game for any of the test instances. For the Star method, this could be the case because it allocates emissions in the same way in each scenario, proportional to the stand-alone emissions of a customer, so in Scenario 4 it might not be able to capture the dynamics of the problem. The Shapley Value allocates emission to each customer equal to their average marginal emission over all coalitions. However, in Scenario 4 the marginal contribution to the emissions by adding a customer to a route could be very high if a new route has to be started to serve all these customers in the coalition. This might influence its *Stability* performance. Based on these criteria, these two methods seem to not be very suitable to the problem dynamics of Scenario 4.

Table 5: Performance of the five allocation methods with respect to the evaluation criteria in the four scenarios, where we use the general emission function and no approximation to the total number of coalitions in Scenario 3.

Sce-nario	Alloc. method	Average emission allocated to: (kg)			% of alloc. in core	% of alloc. consistent	% of alloc. eq. treatment of eq.	Robust-ness (CoV)	Avg. comp. time (s)
		Target cust. 1	Target cust. 2	Target cust. 3					
1	*Star method*	2.12 (0.12)	2.12 (0.12)	2.14 (0.12)	73.33	100.00	100.00	0.06	0.00
	Shapley	1.99 (0.36)	1.90 (0.45)	2.02 (0.33)	100.00	60.00	0.00	0.18	0.00
	Lorenz+	2.05 (0.17)	2.05 (0.17)	2.05 (0.17)	100.00	0.00	93.33	0.08	0.00
	EPM+	2.10 (0.12)	2.10 (0.12)	2.12 (0.12)	100.00	100.00	93.33	0.06	0.00
	Nucleolus	1.81 (0.47)	2.57 (2.89)	1.81 (0.48)	100.00	33.33	0.00	0.26	0.50
2	*Star method*	2.12 (0.12)	2.12 (0.12)	2.13 (0.12)	73.33	100.00	100.00	0.06	0.00
	Shapley	1.99 (0.36)	1.90 (0.45)	2.00 (0.34)	100.00	60.00	0.00	0.18	0.00
	Lorenz+	2.04 (0.17)	2.04 (0.17)	2.04 (0.17)	100.00	0.00	93.33	0.08	0.00
	EPM+	2.10 (0.12)	2.10 (0.12)	2.11 (0.12)	100.00	100.00	93.33	0.06	0.00
	Nucleolus	1.81 (0.47)	2.57 (2.89)	1.78 (0.48)	100.00	33.33	0.00	0.26	0.44
3	*Shapley*	1.99 (0.38)	1.91 (0.45)	2.00 (0.35)	100.00	60.00	0.00	0.19	0.00
	Lorenz+	2.04 (0.16)	2.04 (0.16)	2.04 (0.16)	100.00	0.00	93.33	0.08	0.01
	EPM+	2.09 (0.12)	2.09 (0.12)	2.11 (0.12)	100.00	100.00	93.33	0.06	0.01
	Nucleolus	1.53 (1.78)	2.99 (0.55)	1.78 (1.20)	100.00	33.33	0.00	0.36	0.63
4	*Star method*	3.29 (0.15)	3.29 (0.15)	3.32 (0.15)	0.00	100.00	100.00	0.05	0.00
	Shapley	2.10 (0.32)	5.87 (1.55)	2.11 (0.42)	0.00	60.00	0.00	0.15	0.00
	Lorenz+	0.80 (0.74)	8.00 (3.18)	1.38 (1.14)	66.67	46.67	26.67	0.93	0.19
	EPM+	0.90 (0.77)	7.91 (3.17)	1.29 (1.09)	66.67	53.33	20.00	0.86	0.19
	Nucleolus	0.88 (1.22)	8.19 (3.25)	1.22 (1.17)	66.67	33.33	0.00	0.83	0.60

The reported values are computed for a set of 10 customers over 15 random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
Robustness is measured for the target customer 1, see Section 4.1.
The computation time is for emission allocation only. See Table 4 for the run time to define the emission allocation/pollution routing game.

The Computation Time performance: We see that the Star method has the lowest computation time, reflecting the polynomial computational complexity of the method. The other methods have exponential time complexity, with the Nucleolus having the highest computation time, as it solves multiple LP's to find a solution. Because of the low number of customers, the Lorenz+ Allocation and EPM+ also have very fast computation times in Scenario 1, 2 and 3. In Scenario 4, we see that the computation time of the EPM+ and Lorenz+ Allocation is higher because in the case that test instances have an empty core, the EPM+ and Lorenz+ Allocation coincides with the Nucleolus method. These results are consistent with the work of Naber et al. (2015).

The Robustness performance: When we look at the *Robustness* evaluation criterion, we observe rather low values for all methods in Scenario 1, 2 and 3, and significantly larger values for the Lorenz+ Allocation, EPM+ and Nucleolus in Scenario 4. We conclude that the test instances with an empty core negatively influence the robustness of these three methods, when we look at the results of the allocation methods in Scenario 4 for the test instances with a non-empty core in Table 6. Over all test instances and scenarios, we see that the Star method has the lowest value for the CoV, making it the most robust method. This result is in contrast with what Naber et al. (2015) found for their set of test instances. They consider the Nucleolus method to be the best performing, although in our evaluation this is the worst-performing method in terms of robustness. Also, as opposed to what Naber et al. (2015) found, we observe the EPM+ to have similar performance on the *Robustness* evaluation criterion as the Star method, with the Lorenz+ Allocation performing well, too, deviating only 0.02 from their CoV values. We expect the performance of the allocation methods on the *Robustness* criterion to be affected by the service area set-up of our test instances. This shows to some degree in our evaluation for different service area set-ups in Appendix E, although we still conclude that the Star method and the EPM+ are the most robust methods.

Table 6: Performance of the five allocation methods with respect to the evaluation criteria in Scenario 4 for test instances with non-empty core, where we use the general emission.

Scenario	Alloc. method	Average emission allocated to: (kg)			# of alloc. in core	# of alloc. consistent	# of alloc. eq. treatment of eq.	Robustness (CoV)	Avg. comp. time (s)
		Target cust. 1	Target cust. 2	Target cust. 3					
4	*Star method*	3.27 (0.17)	3.27 (0.17)	3.30 (0.18)	0	10	10	0.05	0.00
	Shapley	2.09 (0.38)	6.32 (1.43)	1.95 (0.22)	0	4	0	0.18	0.00
	Lorenz+	0.72 (0.40)	9.19 (2.81)	1.02 (0.63)	10	4	4	0.56	0.19
	EPM+	0.87 (0.50)	9.06 (2.85)	0.89 (0.34)	10	5	3	0.58	0.19
	Nucleolus	0.85 (0.39)	9.48 (2.80)	0.78 (0.47)	10	2	0	0.47	0.60

The reported values are computed for a set of 10 customers over 10 non-empty core random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
Robustness is measured for the target customer 1, see Section 4.1.
The computation time is for emission allocation only. See Table 4 for the run time to define the emission allocation/pollution routing game.

The consistency performance: When we look at the overall performance on the *Consistency* criterion, we see that the Star method performs best. This result is in accordance with Naber et al. (2015). We also similarly find that the Nucleolus performs relatively poor on this criterion for this set of test instances in Scenario 1, 2 and 3. The Shapley Value has average performance, but we would not consider its performance poor, a conclusion reached by Naber et al. (2015). The EPM+ has the overall second-best performance, which we expect as the method allocates emissions by minimizing the largest difference between the allocation relative to the stand-alone emission for each customer order, so this ensures a certain consistency in its allocations. As the stand-alone emissions for each customer are mostly determined by their distance to the depot, which is the same for the three target customers, we are, however, not surprised to see similar average emissions allocated over these customers in Scenario 1, 2 and 3. We also observe this for the Lorenz+ Allocation, which we can similarly expect, as this method tries to minimize the difference between the largest and smallest allocation to any customer. The Lorenz+ Allocation, on the contrary, does not seem to incorporate the effect of order weight in its emission allocation in Scenario 1, 2 and 3, which is reflected in its bad performance on the *Consistency* criterion. The Star method, EPM+ and Shapley Value, although they only allocate 0.01 to 0.02 kg more emission to target customer 3's order, do seem to incorporate order weight into their emission allocation in these scenarios. These results are in accordance with Naber et al. (2015), as they also find that the weight of the customer orders had little effect on the emission allocation in case the customer can be visited on a single distribution tour.

We shortly elaborate on the *Consistency* performance of the Lorenz+ Allocation and the EPM+ in Scenario 4 at the end of the section where we discuss the average emission allocations to the target customers in the different scenarios. Furthermore, we want to make a quick note here on the number of test instances we used and the set-up of the test instances. In Appendix E, we evaluate the performance of the emission allocation method in only Scenario 1 and 2, for two service area set-ups over 30 test instances, of which one is the same as we use in our main results. For both service area set-ups, when we use this larger set of test instances, we see that especially the Shapley Value, Lorenz+ Allocation and EPM+ allocate relatively more emissions to the order of target customer 3, as compared to the values we observe in Table 5. In addition, from the output results presented in this Appendix, we also conclude that the Nucleolus method has the least consistent emission allocation, and is most sensitive to a change in the service area set-up. Hence, the service area set-up and the rather small set of test instances we consider for the set of 10 customers has some influence on the conclusions we draw in this section.

The Equal Treatment of Equals performance: Kellner and Schneiderbauer (2019), finds that the proportional allocation methods achieve better results on the *Equal Treatment of Equals* evaluation criterion than the Shapley Value. We indeed see that the Star method's emission allocation always satisfies this criterion and the Shapley Value's emission allocations never. Also, the Lorenz+ Allocation and EPM+ both have a good performance on the *Equal treatment of Equals* criterion in Scenario 1, 2 and 3. Both methods try to minimize differences between allocations to customer orders, and when the effect of the order weight influences the emission allocation to a small degree, these methods seem to ensure that orders with similar characteristics get allocated similar amounts of emission.

The Nucleolus has the largest relative differences in the average emissions allocated over the three target customers. Although target customers 1 and 2 have similar order characteristics, except for the order volume, average allocated emissions differ 41.99% in Scenario 1 and 2, and even 95.42% in Scenario 3. For the Star method, Lorenz+ Allocation and EPM+ we observe no such difference, and only a small difference is observed for Shapley Value, of between 4% to 5%. It is thus not surprising that the Nucleolus performs very badly on the *Equal treatment of Equals* criterion, and also on the *Consistency* criterion, its performance is the worst of all allocation methods in each scenario.

Comparison of the average emission allocations in the different scenarios: In Table 5 we see that the average emission allocations in Scenario 1 and 2, almost coincide, confirming that the minimum emission tours found in Scenario 2, are very similar to the minimum distance tours found in Scenario 1. Interestingly, if we compare the average emission allocations to the target customers' orders in Scenario 2 and 3, we see similar allocations for the Shapley Value, Lorenz+ Allocation, and EPM+. This implies that for these methods, it does not seem to have significant added value to use a computationally more complex pollution routing game, instead of a simpler emission allocation game. Only the Nucleolus allocates significantly different amounts of emission in Scenario 2 and 3, with an 18.30% difference for target customer 1's order and 16.34% difference for target customer 2's. As this method tries to minimize the maximum disaffection of any coalition, using it in the pollution routing game, which better specifies the minimum emission tours by solving a pollution routing problem for each coalition, seems to have added value. We note here, that the small difference between the average emission allocation to the target customer's orders for the other methods in Scenario 2 and 3, could be related to the small customer sample considered. Larger differences might only become apparent for a bigger customer set.

The problem dynamics of Scenario 4 differ from the other three scenarios. In this scenario, some of the coalitions of customers need to be serviced using two trucks, which increases the emissions generated on such a sub-tour significantly, as compared to a case where all coalitions can be serviced using a single truck. This affects the CGT allocation methods. As the order volume is now influencing the routing of the customer orders, the marginal emissions of adding a customer to a tour seem to have changed, reflected by the higher emission allocation to target customer 2's order for the Shapley Value.

As the Lorenz+ Allocation, EPM+ and Nucleolus coincide in case the core is empty for a specific test instance, we also look at the results in Table 6. The Lorenz+ Allocation, EPM+ and Nucleolus each explicitly incorporate the emissions generated on a sub-tour for each coalition in their LP problems. In Scenario 4, these emissions changed because of the influence of the order volume characteristic on the routing of the two trucks, which is reflected by a change in the proportion of average emissions allocated over the orders of the target customers. Each CGT allocation method seems to capture the effect that order volume has on the routing of the customers, and hence on the total generated emissions for both tours. In the conclusion of the paper of Naber et al. (2015) they share that they expect the demand of a customer order to have a greater effect on the allocated emissions in the case that total demand exceeds the limit of the truck. We can confirm this expectation with our results. In addition, Kellner and Schneiderbauer (2019) conclude that volume has a greater effect on fuel consumption, and hence generated emissions, than weight, because of its influence on the routing of trucks. From the values in Tables 5 and 6, we can conclude this as well.

We furthermore see that in Scenario 4, we have the lowest average emissions allocated to the orders of target customers 1 and 3 using the Lorenz+ Allocation, EPM+ and Nucleolus, as compared to the other scenarios. This reflects that the order weight might now even have a smaller effect on the emission allocation done by the allocation methods, although the difference in emissions allocated to target customer 1's and 3's orders has increased. Lastly, we note that the different problem dynamics of Scenario 4, make that the Lorenz+ Allocation seems to more consistently allocate emissions to the orders of the three target customers. In contrast, the EPM+ has worse performance on the *Consistency* criterion, as compared to the other scenarios.

4.3.3 Comparison of the allocation method performance in Scenario 1 and 2 with 10 customers using both emission functions

In this section, we compare the allocation methods' performance when we use Ligterink's emission function and the general emission function. Table 7 presents the emission percentages allocated to the orders of the target customers for the set of test instances with 10 customers, using both emission functions for Scenario 1 and 2.

We see that the Star method, the Lorenz+ Allocation and the EPM+ allocate very similar percentages of emissions to the target customers' orders using either of the two emission functions in both scenarios. However, the Shapley Value seems to better capture the effect of order weight in its emission allocation using Ligterink's emission function. It allocates 6.70% more emissions to the order of target customer 3, as compared to target customer 1 in Scenario 1, and 5.32% in Scenario 2. Using the general emission function, this difference is 1.24% and 0.5%, respectively. For both emission functions, these percentage differences do decrease in Scenario 2, as compared to Scenario 1.

The most striking difference between the use of the two emission functions is observed when we look at the performance of the Nucleolus. When we look at its emission allocation in Scenario

Table 7: Comparison of the two emission functions with respect to the emission allocation to the target customers in Scenario 1 and 2.

Scenario	Alloc. method	Avg. emission alloc. Ligterink's emission function (%)			Avg. emission alloc. General emission function (%)		
		Target cust. 1	Target cust. 2	Target cust. 3	Target cust. 1	Target cust. 2	Target cust. 3
1	*Star method*	10.24	10.24	10.37	10.26	10.26	10.33
	Shapley	9.55	9.12	10.19	9.70	9.28	9.82
	Lorenz+	9.88	9.88	9.88	9.88	9.88	9.88
	EPM+	10.16	10.16	10.29	10.18	10.18	10.25
	Nucleolus	9.13	8.62	9.51	8.82	12.22	8.74
2	*Shapley*	9.59	9.15	10.10	9.72	9.29	9.77
	Lorenz+	9.88	9.88	9.88	9.89	9.89	9.89
	EPM+	10.16	10.16	10.29	10.18	10.18	10.25
	Nucleolus	8.77	9.05	9.34	8.85	12.25	8.68

The reported values are computed for a set of 10 customers over 15 random test instances.
The target customers is fixed in all test instances, see Section 4.1.

1 using Ligterink's emission function, target customer 2's order gets allocated 5.92% less per cent of emissions than target customer 1's, while if we use the general emission function, target customer 2's order gets allocated 38.55% more. In Scenario 2, we do not observe this. Furthermore, when we use the Nucleolus with Ligterink's emission function, we see a higher percentage of emissions allocated to target customer 3's order than to target customer 1's in both scenarios, while we see an opposite result for the general emission function, although the difference is smaller. It seems that the emission allocation using the Nucleolus is most sensitive to a change in emission function. In Appendix E we reach a similar conclusion in case we change the parameters of a specific emission function. Also then, the Nucleolus is most affected.

In Appendix F we have included a table with results of the emission allocation methods' performance with respect to the evaluation criteria for Ligterink's emission function in Scenario 1 and 2. They do not perform very differently with respect to the evaluation criteria when using either emission function. Hence, similar conclusions can be drawn from these results, as from the results in Table 5.

4.3.4 Performance of the allocation methods in the four scenarios with 14 customers

We now look at the results in Table 8 for the set of test instances with 14 customers. These results are comparable to the results in Table 5 for the set of 10 customers. We do notice some small differences in the relative CoV values for the Lorenz+ Allocation and the EPM+, and around a 40% higher value for the CoV for the Shapley Value, and a 60% higher value for the Nucleolus in both Scenario 1 and 2 for the set of test instances with 14 customers. However, the conclusion that the Shapley Value has the second-worst performance on the *Robustness* criterion and the Nucleolus the worst, remain the same.

Table 8: Performance of the five allocation methods with respect to the evaluation criteria in the four scenarios, where we use the general emission function and an approximation to the total number of coalitions in Scenario 3.

Sce-nario	Alloc. method	Average emission allocated to: (kg)			% of alloc. in core	% of alloc. consistent	% of alloc. eq. treatment of eq.	Robust-ness (CoV)	Avg. comp. time (s)
		Target cust. 1	Target cust. 2	Target cust. 3					
1	*Star method*	1.71 (0.13)	1.71 (0.13)	1.72 (0.13)	66.67	100.00	100.00	0.07	0.00
	Shapley	1.60 (0.41)	1.58 (0.39)	1.66 (0.35)	100.00	53.33	10.00	0.26	0.07
	Lorenz+	1.61 (0.12)	1.61 (0.12)	1.61 (0.12)	100.00	3.33	90.00	0.07	0.16
	EPM+	1.69 (0.13)	1.69 (0.13)	1.70 (0.13)	100.00	100.00	96.67	0.08	0.16
	Nucleolus	1.44 (0.61)	1.44 (0.43)	1.44 (0.49)	100.00	33.33	6.67	0.42	10.76
2	*Star method*	1.70 (0.12)	1.70 (0.12)	1.71 (0.12)	66.67	100.00	100.00	0.07	0.00
	Shapley	1.60 (0.40)	1.58 (0.39)	1.66 (0.35)	100.00	50.00	6.67	0.25	0.06
	Lorenz+	1.60 (0.12)	1.60 (0.12)	1.61 (0.11)	100.00	3.33	90.00	0.07	0.14
	EPM+	1.69 (0.12)	1.69 (0.12)	1.70 (0.13)	100.00	100.00	96.67	0.07	0.14
	Nucleolus	1.26 (0.58)	2.08 (2.82)	1.34 (0.57)	100.00	30.00	3.33	0.46	9.63
3*	*Shapley*	1.58 (0.21)	1.62 (0.27)	1.63 (0.21)	0.00	60.00	13.33	0.13	0.00
	*Lorenz+***	1.63 (0.12)	1.63 (0.12)	1.63 (0.12)	100.00	13.33	100.00	0.07	0.02
	EPM+	1.70 (0.12)	1.70 (0.12)	1.71 (0.13)	100.00	100.00	100.00	0.07	0.02

The reported values are computed for a set of 14 customers over 30 random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
Robustness is measured for the target customer 1, see Section 4.1.
The computation time is for emission allocation only. See Table 4 for the run time to define the emission allocation/pollution routing game.
The reported values for Scenario 3 are computed using an approximation to the total number of coalitions including only the 6 nearest neighbours, see Section 2.3.

A more noticeable difference is discovered when we look at the emission allocation to the three target customers' orders for the Nucleolus method in both Scenario 1 and 2. Similar to the set of test instances with 10 customers, the Nucleolus method allocates the least emissions to the orders of target customers 1 and 2 in both scenarios. However, for the set of test instances with 10 customers, the Nucleolus allocates around 42% more emissions to target customer 2's order as compared to target customer 1's in Scenario 1, and around 2% fewer emissions to target customer 3's order as compared to target customer 1's in Scenario 2. Table 8 present contrasting results. The emission allocations of the Nucleolus in Scenario 1 for the set of test instances with 14 customers, however, is more aligned with the emission allocations of the other methods. In addition, this time we see in Scenario 2 that the Nucleolus method allocates more emission to the order of target customer 3, as we would expect because this customer has a bigger order weight.

Another observation is the increase in computation time of the allocation methods, associated with the increase in the number of customers. This is what we would expect to see. In addition, similar to what Naber et al. (2015) found, we see that increasing the number of customers on a route, decreases the emissions allocated to orders of the target customers.

Lastly, we look at the results of the emission allocation methods' performance in Scenario 3. For this set of test instances, we used an approximation to the total number of coalitions to reduce the computational effort to define the pollution routing game. The emission allocations of the Shapley Value, Lorenz+ Allocation and EPM+ in Scenario 3 using this approximation, are comparable to the emissions allocations of these methods in Scenario 2. We have also seen this for the set of test instances with 10 customers. If we compare the average emission allocations to the target customers in Scenario 3 for both sets of test instances, the most notable difference is that the Shapley Value allocates slightly more emissions to target customer 2's order than to target customer 1's, in the set of test instances with 14 customers, while we see an opposite result for the set of test instances with 10 customers where no approximation is used. Other than that, we do not observe striking differences. Furthermore, the used approximation affects the emission allocations' standard deviation for the Shapley Value, which results in a lower CoV value, as compared to the set of test instances with 10 customers.

The performance of the approximation seems reasonable, however, for the set of test instances with 10 customers, we already commented that using the pollution routing game (in Scenario 3), might not be worth the extra computational efforts for all types of emission allocation methods. We can reach a similar conclusion for this set of test instances.

5 Conclusion & Discussion

In this paper, we evaluate the performance of the Star method, a proportional allocation method, and four cooperative game theory (CGT) allocation methods, the Shapley Value, the Lorenz+ Allocation, the EPM+ and the Nucleolus, in the context of an emission allocation game and a pollution routing game. The Star method is easily applicable and understandable and it allocates emissions proportional to the emissions allocated when a customer would be served on a separate tour. In the emission allocation game, we first determine either a minimum distance or minimum emission route visiting all customer orders by solving a mixed-integer linear programming (MIP) problem. Based on this grand-tour, for each coalition of customers, we approximate the emissions generated on the sub-tours visiting the customers in the coalition, to define the game. In the pollution routing game, we find (a) minimum emission grand-tour(s) and for each coalition, we also solve a MIP problem to find the associated minimum emission sub-tours. As this requires solving an exponential number of MIP problems when we consider test instances with 14 customers, we use an approximation to the total number of coalitions, where we only consider the coalitions for each customer that includes up to their nearest 6 neighbours. We consider the pollution routing game in a setting with one truck that can always fit all customer orders, and a setting with two trucks, with each having a limited volume capacity. The five allocation methods are all evaluated based on stability, consistency, robustness, equal treatment of equals and computation time. In addition, we evaluate the impact of the order characteristics, volume and weight, on the allocated emission and compare the emission function of Ligterink et al. (2012) and a general emission function used in Kellner and Schneiderbauer (2019).

In all our test instances and considered scenarios for the emission allocation and pollution routing game, the Star method is best performing in terms of equally treating equal customer orders, and consistently allocating emissions when customer order characteristics change. In addition, it also has the best or second-best performance on the robustness criteria. Furthermore, it has the lowest computation time. However, the Star method does not capture the effect order volume has on the routing of the customer's orders, and hence the total produced emissions of a tour, in the pollution routing game where we use two trucks with small volume capacity. Regarding the stability criterion, the Lorenz+ Allocation, the EPM+ and the Nucleolus all have equally good performance. Depending on what criteria are most important to customers, either the Star method or the EPM+ can be considered the best performing method over all scenarios.

Regarding the effect that the order characteristics have on the emission allocation of each method, we see that the CGT allocations methods allocate more emissions to customers with orders of larger volume in a pollution routing game context where we use two trucks with small volume capacity. This is not observed when we use one truck with enough volume and weight capacity to transport all orders, suggesting that a customer in one route also affects the emission generated on other routes. In general, in the pollution routing game where we needed two trucks to visit all customers, the CGT allocation methods seem to better capture the effect of the order characteristics, both volume and weight. This suggests that the dynamics of the problem change the performance of the allocation methods.

Regarding the comparison of the performance of the allocation methods using different emission functions, we find no striking differences. Hence, we would prefer to use the simpler emission function as used by Kellner and Schneiderbauer (2019) as it is easily applicable, also in a vehicle routing context. Furthermore, we find that the method most sensitive to changes in the emission function or their parameters was the Nucleolus.

Lastly, we want to note that we did not observe a big difference in the performance of the allocation methods in the emission allocation game versus the pollution routing game in case we use one truck that can serve all customers in both games. Only the Nucleolus method had a significantly different performance. This suggests that it might not benefit the performance of all the CGT emission allocation methods in all routing settings when they are used in the more specified, but computationally more complex pollution routing game. A benefit of using the methods in the emission allocation game is that it is more easily implementable and has fast run times.

Limitations: We evaluate consistency by comparing three customer orders with different characteristics (distance, volume and weight) and check if they get allocated emissions in a way we would expect. Naber et al. (2015) evaluate consistency by performing an ordinary least squares regression with the allocated emission as dependent variables and the distance to the depot, the average distance to other customers and the order characteristic as independent variables. We believe this gives more insight into the consistency of the allocation methods than the approach we use.

Furthermore, we want to remark that in the implementation of the Nucleolus method we did not explicitly implement a stop criterion, but always performed a maximum of 50 iterations to ensure the method found a solution, which we assumed to be optimal. This however comes at a cost, as it increased the computation time of the allocation method.

Lastly, we remark that we use a small service area, and hence do not incorporate the maximum daily operating time into our routing model. This reduces the effect that the distance to the depot for each customer order has on the proportion of emissions allocated to a customer. Moreover, not only the service area we define but also the choice for each order weight, order volume and the truck capacities influence our results.

Future research: We see in our experiment that the pollution routing game becomes computational more complex when we increase the number of customers. This makes the CGT allocation methods hard to implement in such a game if we aim to achieve solutions within reasonable run times. Only a proportional allocation method seems to be suitable in this case. Future research could be done to better specify the benefits of using the emission allocation methods in a pollution routing game. In addition, new research could look into which (CGT) allocation methods perform well in various routing scenarios, where different service areas, constraints on the truck capacities and the daily operating time, and number of trucks are used. Furthermore, it would be useful to find good approximations to the CGT allocation methods in a pollution routing game context with a large number of customers.

Moreover, to use the allocation methods in a more realistic scenario, with maybe over 50 customers, heuristics need to be used to find the grand- and sub-tours for the coalitions of customers. It could be interesting to investigate which common heuristics used in vehicle routing problems could be adapted to incorporate emissions into the objective, in addition to costs and/or travel time.

References

Agi, M. A., Faramarzi-Oghani, S., & Hazır, Ö. (2020). Game theory-based models in green supply chain management: a review of the literature. *International Journal of Production Research*. doi: 10.1080/00207543.2020.1770893

Arin Aguirre, F. J. (2003, January). *Egalitarian distributions in coalitional models: The Lorenz criterion* (Tech. Rep.). Universidad del País Vasco - Departamento de Fundamentos del Análisis Económico I.

CEN (European Committee for Standardization). (2012). *EN 16258:2012 - Methodology for calculation and declaration of energy consumption and GHG emissions of transport services (freight and passengers).* Retrieved May 4 2021, from https://infostore.saiglobal.com/preview/98699884974.pdf?sku=876833_SAIG_NSAI_NSAI_2084000

COFRET (Carbon Footprint of Freight Transport). (2015). *Results Report.* Retrieved from https://cordis.europa.eu/docs/results/265/265879/periodic2-cofret-results-report-final.pdf

Communication from the commission to the European Parliament, the European Council, the European Economic and Social Committee, and the Committee of the Regions. (2019, December). *The European Green Deal.* Brussels. Retrieved May 6 2021, from https://ec.europa.eu/info/sites/default/files/european-green-deal-communication_en.pdf

Cordeau, J.-F., Laporte, G., Savelsbergh, M., & Vigo, D. (2007, 01). Vehicle routing. In (Vol. 14, p. 195-224).

Council of the European Union. (2016). *Council decision (eu) 2016/1841 of 5 october 2016 on the conclusion, on behalf of the european union, of the paris agreement adopted under the united nations framework convention on climate change.* Retrieved May 6 2021, from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016D1841&from=ES

DBEI (Department for Business, Energy Industrial Strategy). (2021). *Government conversion factors for company reporting of greenhouse gas emissions - Conversion Factors 2021.* Retrieved May 18 2021, from https://www.gov.uk/government/collections/government-conversion-factors-for-company-reporting

Dutta, B., & Ray, D. (1989). A Concept of Egalitarianism Under Participation Constraints. *Econometrica*, *57*(3), 615–635. doi: 10.2307/1911055

Engevall, S., Göthe-Lundgren, M., & Värbrand, P. (2004). The Heterogeneous Vehicle-Routing Game.

Transportation Science, *38*(1), 71–85. doi: 10.1287/trsc.1030.0035

Enzmann, J., & Ringel, M. (2020). Reducing road transport emissions in Europe: Investigating a demand side driven approach. *Sustainability*, *12*(18), 1-31. doi: 10.3390/su12187594

Frisk, M., Göthe-Lundgren, M., Jörnsten, K., & Rönnqvist, M. (2010). Cost allocation in collaborative forest transportation. *European Journal of Operational Research*, *205*(September (2)), 448–458. doi: 10.1016/j.ejor.2010.01.015

Guajardo, M. (2018). Environmental benefits of collaboration and allocation of emissions in road freight transportation. In V. Zeimpekis, E. Aktas, M. Bourlakis, & I. Minis (Eds.), *Sustainable freight transport: Theory, models, and case studies* (pp. 79–98). Cham: Springer International Publishing. doi: 10.1007/978-3-319-62917-9_6

Guajardo, M., & Jörnsten, K. (2015). Common mistakes in computing the nucleolus. *European Journal of Operational Research*, *241*(3), 931-935. doi: 10.1016/j.ejor.2014.10.037

Guajardo, M., & Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International Transactions in Operational Research*, *23*(3), 371–392. doi: 10.1111/itor.12205

IEA. (2020). *Tracking Transport 2020.* Paris. Retrieved May 7 2021, from `https://www.iea.org/reports/tracking-transport-2020`

INFRAS. (2011). *HBEFA 3.1: Handbuch Emissionsfaktoren des Straßen- verkehrs, Version 3.1.* Retrieved May 18 2021, from `www.hbefa.net`

Kellner, F. (2016). Allocating greenhouse gas emissions to shipments in road freight transportation: Suggestions for a global carbon accounting standard. *Energy Policy*, *98*, 565–575. doi: 10.1016/j.enpol.2016.09.030

Kellner, F., & Otto, A. (2012). Allocating CO2 emissions to shipments in road freight transportation. *Journal of Management Control*, *22*(4), 451–479. doi: 10.1007/s00187-011-0143-6

Kellner, F., & Schneiderbauer, M. (2019). Further insights into the allocation of greenhouse gas emissions to shipments in road freight transportation: The pollution routing game. *European Journal of Operational Research*, *278*(1), 296–313. doi: 10.1016/j.ejor.2019.04.007

Kranke, A., Schmied, M., & Schön, A. (2011). *CO_2-Berechnung in der Logistik: Datenquellen, Formeln, Standards.* München: Verlag Heinrich Vogel. Retrieved May 18 2021, from `https://www.heinrich-vogel-shop.de/img/asset/readings/23726\_0\_reading.pdf`

Leenders, B. P., Velázquez-Martínez, J. C., & Fransoo, J. C. (2017). Emissions allocation in transportation routes. *Transportation Research Part D: Transport and Environment*, *57*(September), 39–51.

doi: 10.1016/j.trd.2017.08.016

Leonardi, J., McKinnon, A., & Palmer, A. (2010). Guidance on measuring and reporting Greenhouse Gas (GHG) emissions from freight transport operations. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/218574/ghg-freight-guide.pdf

Ligterink, N. E., Tavasszy, L. A., & Lange, R. D. (2012). A velocity and payload dependent emission model for heavy-duty road freight transportation. *Transportation Research. Part D: Transport Environment*, *17*(6), 487–491. doi: 10.1016/j.trd.2012.05.009

Liu, T., Wang, Q., & Su, B. (2016). A review of carbon labeling : Standards , implementation , and impact. *Renewable and Sustainable Energy Reviews*, *53*, 68–79. doi: 10.1016/j.rser.2015.08.050

Miller, C. E., Tucker, A. W., & Zemlin, R. A. (1960, October). Integer programming formulation of traveling salesman problems. *J. ACM*, *7*(4), 326–329. doi: 10.1145/321043.321046

Naber, S. K., de Ree, D. A., Spliet, R., & van den Heuvel, W. (2015). Allocating CO2 emission to customers on a distribution route. *Omega (United Kingdom)*, *54*, 191–199. doi: 10.1016/j.omega.2015.01.017

Özener, O. Ö. (2014). Developing a Collaborative Planning Framework for Sustainable Transportation. *Mathematical Problems in Engineering*, *2014*. doi: 10.1155/2014/107102

Schmeidler, D. (1969). The Nucleolus of a Characteristic Function Game. *SIAM Journal on Applied Mathematics*, *17*(November (6)), 1163–1170. doi: 10.1137/0117107

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games (am-28), volume ii* (pp. 307–318). Princeton University Press. doi: 10.1515/9781400881970-018

Wick, C., Klumpp, M., & Kandel, C. (2011, 01). Shapley value simulation for allocating ghg emission savings due to logistics pooling within ecr cooperations. *Proceedings of the European Simulation and Modelling Conference*, 240-244. Retrieved from https://www.pim.wiwi.uni-due.de/uploads/tx_itochairt3/publications/Wick_Klumpp_Kandel_-_Shapley_Value_Simulation_for_Allocating_GHG_Emission_Savings_due_to_Logistics_Pooling_within_ECR_Cooperations.pdf

WMO (World Meteorological Organisation). (2020). *State of the global climate 2020 : provisional report* (No. 1264). Retrieved May 6 2021, from https://gcos.wmo.int/en/global-climate-indicators\%0Ahttps://www.ncdc.noaa.gov/sotc/global/202008

Xu, X., Pan, S., & Ballot, E. (2012). Allocation of transportation cost co2 emission in pooled supply chains using cooperative game theory. *IFAC Proceedings Volumes*, *45*(6), 547-553. doi: 10.3182/20120523-3-RO-2023.00078

6 Appendix

A Ligterink's emission function

We follow the reasoning of Naber et al. (2015) and use Eqs. (2) and (3) of Ligterink et al. (2012), and the values found for CO_2 emission in Table 1 to compute emissions on a distribution route. Each truck is assumed to have a mass of 5 ton, each loading unit weighs 0.01 ton, and the truck has a maximum capacity of 507 loading units. Let d be the total number of loading units that are transported. The specific power, KWt, is the engine power per ton gross vehicle weight expressed in kW/ton. In our case, this is defined as, $KWt = 131.25/(5 + 0.01d)$. In addition, we let V be the velocity of the truck in km/h. Similar to Naber et al. (2015) we assume the velocity is 35 km/h between two locations in the considered region, see Section 4.1. Now we introduced the relevant parameters, we can compute the emissions of a truck transporting d loading units and travelling one kilometre with velocity V by

$$\begin{aligned} \frac{465.390 + 48.143KWt}{V} &+ 32.389 + 0.8931KWt - (0.4771 + 0.02559KWt)V \\ &+ (0.0008889 + 0.0004055KWt)V^2. \end{aligned} \tag{25}$$

We want to make a critical remark here. In Naber et al. (2015), the emission function is presented as above, without mention of the units of the produced emissions. If we look at the original paper of Ligterink et al. (2012), we see that the emissions produced by the above function are in grams per ton transported and kilometre driven. Hence, it is of importance to multiply the above number with the travelled distance and the gross vehicle weight (GVW) of the truck per segment of the distribution route driven to calculate the right amount of produced emissions. If in the above function, the value is not multiplied with the GVW, we see that CO_2 emissions will increase when the transported truck weight decreases, as KWt will increase. This result is inconsistent with what we would expect in practice and will lead to a wrong allocation of emissions to customer orders.

B The general emission function

In Kellner and Schneiderbauer (2019), and Kellner (2016) emissions are calculated by combining the emissions generated by an empty truck, and a fully loaded truck, in the following way:

$$(FC_{empty} + (FC_{full} - FC_{empty} * \frac{to}{Cap})/100 * distance * ECF \tag{26}$$

The function calculates emissions, in kg CO_2, by multiplying the vehicle fuel consumption (FC) per kilometre, the driven distance, and an energy conversion factor (ECF). For this ECF, we use 2.67 kg CO_2 per liter 100% mineral diesel (DBEI, 2021). The technical data to be used in the emissions function is presented for different heavy good vehicle in Kellner and Schneiderbauer (2019). They derive this data from INFRAS (2011) and Kranke, Schmied, and Schön (2011). We use the values for FC_{empty}, and FC_{full}, for a truck with gross vehicle weight rating $7.5 - 12$ ton with a payload of 6.0, as this best compares to the truck characteristics used in the emission function in the paper of Naber et al. (2015). The values are $FC_{empty} = 16.5$, and $FC_{full} = 19.9$ per 100 km.

In the above formula, $\frac{to}{Cap}$ is the weight-based load factor. This value will be different on each segment of the distribution route. For the truck capacity, we similarly as in Ligterink's emission function use a value of 5.07 ton.

C Linear programming model to find the Nucleolus

The following procedure to compute the Nucleolus is used in Naber et al. (2015) and Engevall et al. (2004). We denote, ε, to be the excess of an allocation x. Furthermore, we let ε_l be the optimal value of ε in iteration l. All coalitions for which the excess is fixed after solving the below linear programming (LP) problem in iteration l, are part of the set F_l, with $F_1 = \emptyset$.

The subsequent LP problem is solved successively until a unique solution to the LP problem is found. We represent the LP for iteration l:

$$\varepsilon_l = max \quad \varepsilon, \tag{27}$$

$$\text{s.t.} \quad x(\{i\}) \leq e(\{i\}) \qquad \forall i \in N, \tag{28}$$

$$x(S) + \varepsilon \leq e(S) \qquad \forall S \subset N, S \notin (\cup_{m<l} F_m), \tag{29}$$

$$x(S) + \varepsilon_m = e(S) \qquad \forall m < l, S \in F_m, \tag{30}$$

$$x(N) = e(N). \tag{31}$$

In the first iteration, constraint (30) is not included and constraint (29) holds for all coalitions, $S \subseteq N$. From duality theory, when the optimal value of a dual variable is positive, the inequality constraint that is associated with this variable holds with equality at an optimal solution. We denote y_l to be the vector dual variable for iteration l for the coalition with non-fixed excess. For a given solution of the above LP at iteration l, we can find set F_l by including in the set all coalitions for which the associated dual variable is strictly positive. For these coalition, we fix the excess to ϵ_l, because constraint (29) will hold with equality for the coalition with $y_l(S) > 0$. More precisely, $F_l = \{S \in N | y_l(S) > 0\}$. The algorithm stops when the constraint matrix of constraints (30) and (31) of the corresponding fixed coalitions has rank $|N|$.

D Evaluation of the quality of the approximations of the number of coalitions in the pollution routing game for Scenario 3 and 4

In Section 2.3 we propose the idea to use an approximation to reduce the number of coalitions considered in the pollution routing game (that we use in Scenario 3 and 4) by only considering for each customer the coalitions including its 5 nearest neighbours. Before running our test instances with such an approximation, we check if these approximations produce accurate emission allocations as compared to the case where such an approximation is not used.

Table A.1 provides the results of the evaluation of the performance of the allocations methods when using the discussed approximation with a varying number of nearest neighbours in Scenario 3 and 4, as compared to the case where this approximation is not used. We use a set of test instances with 10 customers in our evaluation.

Table A.1: Comparison of the emission allocation to target customer 1 in Scenario 3 and 4, calculated with and without an approximation to the total number of coalitions using a varying number of nearest neighbours.

Scenario	Alloc. method	# of nearest neighbours	Avg. emission alloc. No Approx. (kg)	Avg. emission alloc. Approx. (kg)	Avg. difference alloc. (%)
3	Shapley	4	2.09 (0.44)	2.06 (0.21)	2.74
		5		2.06 (0.26)	2.79
		6		2.08 (0.33)	0.88
	Lorenz+	4	2.03 (0.17)	2.06 (0.18)	3.04
		5		2.06 (0.18)	3.04
		6		2.06 (0.18)	3.04
	EPM+	4	2.11 (0.11)	2.14 (0.10)	2.59
		5		2.14 (0.10)	2.59
		6		2.14 (0.10)	2.59
	Nucleolus	4	1.37 (0.60)	2.07 (1.33)	69.93
		5		1.59 (1.06)	21.74
		6		2.80 (2.74)	143.13
4	Shapley	4	2.15 (0.33)	2.78 (0.50)	63.62
		5		2.50 (0.38)	35.25
		6		2.40 (0.30)	25.28
	Lorenz+	4	0.78 (0.44)	3.11 (0.36)	232.48
		5		2.98 (0.44)	219.56
		6		3.23 (1.88)	244.82
	EPM+	4	0.91 (0.54)	3.13 (0.21)	222.55
		5		2.84 (0.69)	193.39
		6		2.85 (1.97)	194.66
	Nucleolus	4	0.85 (0.49)	2.59 (1.80)	174.27
		5		2.54 (1.76)	169.47
		6		2.75 (3.20)	190.51

The reported values are computed for a set of 10 customers over 10 random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
Target customer 1 is fixed in all test instances, see Section 4.1.

We find a deviation in the emissions allocated to target customer 1's order, of more than 25% for all allocation methods in Scenario 4 using either 4, 5 or 6 nearest neighbours in our approximations, as compared to the emission allocation using these methods without an approximation to the total number of coalitions. It appears that reducing the number of coalitions used in the allocation methods makes them unable to capture the dynamics of the pollution routing problem in Scenario 4. This is presumably the case because coalitions of customers that would need to be served with two trucks are excluded from the set of coalitions when using the approximation. In Scenario 3, we also find that the approximation performs bad and inconsistent for the Nucleolus method using either of these number of nearest neighbours, with a deviation larger than 20%. As the Nucleolus method tries to minimize the maximum disaffection of any coalition, it is not surprising that when we leave out certain coalitions when using the approximation, this method does not perform accurately and consistently anymore. However, for the Shapley Value, the Lorenz+ Allocation and EPM+, we find a deviation below 5%, suggesting the approximation can be reliably used for these methods in Scenario 3. The approximation has the best performance for the Shapley value when using 6 nearest neighbours in Scenario 3.

E Sensitivity of the results with respect to the set-up of the test instances and the parameters of the emission functions

We want to check our results with regards to the sensitivity they might have to the set-up of the test instances and the choice of the input values for the emission functions.

Effect of the input parameters in the general emission function: This emission function (see Appendix B) has parameters for the vehicle fuel consumption (FC) per kilometre, and the energy conversion factor (ECF). DBEI (2021) also provide values for the CO_2 emissions (kg) per kilometre driven for a truck with gross vehicle weight rating 7.512 ton. These values do not take the FC explicitly into account and can be considered more general. For an empty truck, 0.53196 kg CO_2 are emitted per kilometre, and for a full truck 0.69111 kg CO_2 per kilometre. We label this 'settings 2', while the input parameters discussed in Appendix B, which are used to generate our main results, are labelled 'settings 1'.

We compare the performance of the emission allocation methods in Scenario 1 and 2 using these two settings in the general emission function. The results of their performance can be found in Table A.2 and Table A.3. The emission function finds the same optimal tours using both settings, but we see in Table A.2 that for parameter settings 2, we observe higher average emissions generated on a tour. This is reflected in the emission allocations using settings 2, as we see in Table A.3. The only noticeable difference is that using settings 2, the emission allocation of the Nucleolus to the order of target customer 2 is lower, than using settings 1, despite the overall higher emissions generated on an average tour. We observe this as well when we compare the general emission function using settings 1 with Ligterink's emission function (see Section 4.3.3). This suggests that the Nucleolus appears to be most sensitive to changes in the emission function and its parameter settings.

Table A.2: Characteristics of Scenario 1 and 2 using two parameter settings for the general emission function.

Scenario	Random instances with	% of empty core games	Avg. emissions per tour (kg)	Avg. total length of tour(s) (km)
1	*Settings 1*	0.00	21.17	46.90
	Settings 2	0.00	25.79	46.90
2	*Settings 1*	0.00	21.11	46.91
	Settings 2	0.00	25.70	46.91

The reported values are computed for a set of 10 customers over 30 random test instances.

Table A.3: Average emission allocations to the target customers in Scenario 1 and 2 using two parameter settings for the general emission function.

Scenario	Alloc. method	Settings 1 Average emissions allocated to (kg):			Settings 2 Average emissions allocated to (kg):		
		Target cust. 1	Target cust. 2	Target cust. 3	Target cust. 1	Target cust. 2	Target cust. 3
1	*Star method*	2.14 (0.14)	2.14 (0.14)	2.16 (0.14)	2.61 (0.17)	2.61 (0.17)	2.64 (0.17)
	Shapley	1.93 (0.35)	1.90 (0.46)	2.10 (0.57)	2.34 (0.42)	2.30 (0.56)	2.60 (0.70)
	Lorenz+	2.09 (0.15)	2.09 (0.15)	2.15 (0.35)	2.54 (0.18)	2.54 (0.18)	2.62 (0.44)
	EPM+	2.12 (0.13)	2.12 (0.13)	2.19 (0.34)	2.58 (0.16)	2.58 (0.16)	2.68 (0.42)
	Nucleolus	1.68 (0.50)	3.37 (3.79)	1.92 (0.83)	1.87 (0.81)	2.91 (3.27)	2.42 (1.05)
2	*Star method*	2.14 (0.14)	2.14 (0.14)	2.15 (0.14)	2.60 (0.17)	2.60 (0.17)	2.63 (0.17)
	Shapley	1.93 (0.35)	1.90 (0.46)	2.08 (0.57)	2.34 (0.42)	2.30 (0.56)	2.57 (0.69)
	Lorenz+	2.09 (0.15)	2.09 (0.15)	2.14 (0.34)	2.54 (0.18)	2.54 (0.18)	2.61 (0.43)
	EPM+	2.12 (0.13)	2.12 (0.13)	2.19 (0.33)	2.58 (0.16)	2.58 (0.16)	2.67 (0.41)
	Nucleolus	1.74 (0.38)	2.99 (3.34)	1.94 (0.83)	2.06 (0.71)	2.17 (0.93)	2.48 (0.96)

The reported values are computed for a set of 10 customers over 30 random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
The target customers are fixed in all test instances, see Section 4.1.

Effect of velocity in Ligterink's emission function: Naber et al. (2015) chose to use a velocity of 35 km/h in Ligterink's emission function. In Table A.4 we share some percentual output differences in case we choose a different velocity. The average emission per tour changes significantly when we change the velocity parameter in the emission function. When we decrease (increase) the velocity, we observe higher (lower) emissions. In general, the average emissions allocated to target customer 1's order change proportional to this increase/decrease. This is especially the case for the Star method and the Lorenz+ Allocation. The Shapley Value, EPM+ and Nucleolus do show some changes in the emission allocation to target customer 1's order when different velocities are used, which are not directly proportional to the change in overall emissions. Again, the Nucleolus method seems to be most sensitive to changes in the parameter settings.

Table A.4: Percentual output differences as compared to Ligterink's emission function with truck velocity of 35 km/h.

Output variable		Differences (%)		
	Velocity (km/h)	*25*	*50*	*65*
Avg. emission on tour (kg)		-30.78	22.41	29.96
Avg. alloc. to target cust. 1 (kg):	*Star method*	-30.75	22.37	29.84
	Shapley	-30.78	22.04	28.95
	Lorenz+	-30.78	22.39	29.90
	EPM+	-26.79	22.35	29.78
	Nucleolus	-30.78	24.37	32.72

The reported values are computed for a set of 10 customers over 30 random test instances.
Target customer 1 is fixed in all test instances, see Section 4.1.

Effect of a change in the service area set-up: When we defined the location coordinate area and the depot location, we also considered a situation where the depot was located at coordinates $(0, 0)$, target customers 1, 2 and 3 had a fixed distance to the depot of 6 kilometres, and the x, y-coordinates of the other customers were randomly chosen between 0 and 15 kilometres. We label this service area, 'set-up 2', and the set-up discussed in Section 4.1.3, 'set-up 1'. We evaluate the emission allocation methods in both service area set-ups, to see if it results in any difference in the performance of these methods. The results of the average emission allocation to the target customers for both set-ups can be found in Table A.5, and Table A.6 presents the performance of the emission allocations methods with respect to the evaluation criteria when we use service area set-up 2 in a set of test instances with 10 customers.

From Table A.5 we see that the percentages allocated to the target customers in service area set-up 2 are overall lower. This seems reasonable as for all customers in service area set-up 1, the distance to the depot is at least 5 kilometres, and presumably larger than in service area set-up 2. Furthermore, when service area set-up 2 is used in Scenario 1, all CGT allocation methods allocate more emissions to orders with a bigger volume, such as the order of target customer 2, as compared to when we use service area set-up 1. In Scenario 2, this is also the case, except for the EPM+, although for all emission allocation methods the difference between the emissions allocated to the orders of target customers 1 and 2 are smaller. This is a difference we would not have predicted, and hence it seems that the set-up of the service area influences the conclusions we make with respect to the average emission allocations over the target customers, and the degree to which the allocation methods incorporate the effect of the order characteristics into their emission allocation. Only the Star method performs exactly similar in both service area set-ups, and also the EPM+ has a very similar performance. The Shapley Value, Lorenz+ Allocation and Nucleolus method seem to be most affected, as the proportion of emissions allocated to the target customers in both service area set-ups differs inconsistently.

Table A.5: Average emission allocations to the target customers using two different service area set-ups in Scenario 1 and 2.

Scenario	Alloc. method	Set-up 1 Average emissions allocated to (%):			Set-up 2 Average emissions allocated to (%):		
		Target cust. 1	Target cust. 2	Target cust. 3	Target cust. 1	Target cust. 2	Target cust. 3
1	*Star method*	10.15 (0.62)	10.15 (0.62)	10.22 (0.61)	6.20 (1.18)	6.20(1.18)	6.24 (1.19)
	Shapley	9.19 (1.86)	9.03 (2.30)	9.95 (2.68)	4.22 (1.02)	4.58 (1.04)	5.02 (1.01)
	Lorenz+	9.87 (0.25)	9.87 (0.25)	10.14 (1.35)	8.51 (1.49)	8.74 (1.68)	8.49 (1.52)
	EPM+	10.05 (0.61)	10.05 (0.61)	10.36 (1.43)	6.02 (1.03)	6.08 (1.02)	6.12 (1.12)
	Nucleolus	7.98 (2.60)	15.75 (17.44)	9.08 (3.77)	6.77 (1.51)	7.27 (1.58)	7.97 (1.38)
2	*Star method*	10.15 (0.62)	10.15 (0.62)	10.22 (0.61)	6.19 (1.18)	6.19 (1.18)	6.24 (1.19)
	Shapley	9.20 (1.86)	9.05 (2.30)	9.90 (2.70)	4.33 (1.06)	4.56 (0.97)	4.73 (1.09)
	Lorenz+	9.88 (0.25)	9.88 (0.25)	10.14 (1.33)	8.48 (1.48)	8.52 (1.34)	8.91 (1.84)
	EPM+	10.06 (0.61)	10.06 (0.61)	10.37 (1.42)	6.05 (1.03)	6.05 (1.03)	6.13 (1.12)
	Nucleolus	8.29 (2.11)	13.92 (14.94)	9.19 (3.77)	6.91 (1.52)	7.82 (2.89)	7.55 (1.70)

The reported values are computed for sets of 10 customers over 30 random test instances using the general emission function.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
The target customers are fixed in all test instances, see Section 4.1.

When we look at the results in Table A.6 we see that the relative performance of the allocation methods on the evaluation criteria in Scenario 1 and 2 is similar to their performance when service area set-up 1 is used. We do see some differences, for example, the Star method has somewhat worse stability performance, the Nucleolus method performs better on the *Consistency* criterion, and both the Lorenz+ Allocation and the EPM+ perform worse on the *Equal Treatment of Equals* criterion. In addition, we see higher values for the CoV, but still, the Shapley Value and Nucleolus methods have the worst performance, while the Star method and EPM+ perform best in both scenarios. The Lorenz+ Allocation performance on the *Robustness* criterion is worse for service area set-up 2, slightly below the Shapley Value CoV. Nevertheless, a similar overall conclusion can be drawn when test instances are used with service area set-up 1 and 2, namely, that the Star Method has the best performance on the *Consistency* and *Equal Treatment of Equals* criteria, and good performance on the *Robustness* criteria. The CGT allocation methods perform best on the *Stability* criterion, and when looking at the performance over all criteria, the EPM+ performs best.

It seems that the service area set-up mostly affects how the allocation methods capture the effect of the order characteristics in their emission allocation. This effect should become even more visible when we would evaluate the allocation methods in Scenario 3 and 4, as in Scenario 1 and 2, we would not yet expect the emission allocation methods to allocate more emissions to target customer 2's order. Further research could be conducted to get more insight into these dynamics.

Table A.6: Performance of the five allocation methods with respect to the evaluation criteria in Scenario 1 and 2 for service area set-up 2.

Sce-nario	Alloc. method	% of alloc. in core	% of alloc. consistent	% of alloc. eq. treatment of eq.	Robust-ness (CoV)	Avg. comp. time (s)
1	*Star method*	46.67	100.00	100.00	0.14	0.00
	Shapley	93.33	76.67	0.00	0.20	0.00
	Lorenz+	100.00	10.00	60.00	0.19	0.00
	EPM+	100.00	100.00	76.67	0.12	0.00
	Nucleolus	100.00	56.67	0.00	0.21	0.52
2	*Star method*	56.67	100.00	100.00	0.14	0.00
	Shapley	100.00	63.33	0.00	0.22	0.00
	Lorenz+	100.00	16.67	66.67	0.19	0.00
	EPM+	100.00	100.00	83.33	0.13	0.00
	Nucleolus	100.00	60.00	0.00	0.21	0.46

The reported values are computed for a set of 10 customers over 30 random test instances using the general emission function.
The computation time is for the allocation of emissions only.

F Performance of the five allocation methods with respect to the evaluation criteria in Scenario 1 and 2, using Ligterink's emission function

Table A.7: Performance of the five allocation methods with respect to the evaluation criteria in Scenario 1 and 2, using Ligterink's emission function.

Scenario	Alloc. method	Settings 1 Average emissions allocated to (%):			% of alloc. in core	% of alloc. consistent	% of alloc. eq. treatment of eq.	Robust-ness (CoV)	Avg. comp. time (s)
		Target cust. 1	Target cust. 2	Target cust. 3					
1	*Star method*	1.94 (0.11)	1.94 (0.11)	1.97 (0.12)	73.33	100.00	100.00	0.06	0.00
	Shapley	1.80 (0.33)	1.71 (0.41)	1.92 (0.30)	100.00	73.33	0.00	0.18	0.00
	Lorenz+	1.88 (0.16)	1.88 (0.16)	1.88 (0.16)	100.00	0.00	86.67	0.08	0.00
	EPM+	1.92 (0.11)	1.92 (0.11)	1.95 (0.11)	100.00	100.00	93.33	0.06	0.00
	Nucleolus	1.72 (0.37)	1.63 (0.33)	1.80 (0.34)	100.00	40.00	0.00	0.22	0.59
2	*Star method*	1.93 (0.11)	1.93 (0.11)	1.96 (0.11)	80.00	100.00	100.00	0.06	0.00
	Shapley	1.80 (0.33)	1.71 (0.41)	1.89 (0.31)	100.00	60.00	0.00	0.18	0.00
	Lorenz+	1.87 (0.15)	1.87 (0.15)	1.87 (0.15)	100.00	0.00	86.67	0.08	0.00
	EPM+	1.92 (0.11)	1.92 (0.11)	1.94 (0.11)	100.00	100.00	93.33	0.06	0.00
	Nucleolus	1.65 (0.55)	1.70 (0.40)	1.76 (0.33)	100.00	40.00	0.00	0.33	0.60

The reported values are computed for a set of 10 customers over 15 random test instances.
For the average emissions allocated to the target customers, the standard deviations are reported in the brackets.
Robustness is measured for the target customer 1, see Section 4.1.
The computation time is for the allocation of emissions only.