Simulation in Produktion
und Logistik 2017
Sigrid Wenzel & Tim Peter (Hrsg.)
kassel university press, Kassel 2017

Comparing Different Distance Metrics for Calculating Distances in Urban Areas with a Supply Chain Simulation Tool

Vergleich verschiedener Distanzmetriken zur Berechnung von Entfernungen in städtischen Gebieten mit einem Supply-Chain-Simulationswerkzeug

Markus Rabe, Astrid Klüter, Simon Tietze, TU Dortmund, Dortmund (Germany),
markus.rabe@tu-dortmund.de, astrid.klueter@tu-dortmund.de,
simon.tietze@tu-dortmund.de

Abstract: This paper presents a modelling approach for the simulation of supply chains including last mile delivery by using approximations. A brief introduction to the most common approximation metrics for distance calculations and their applications within the context of simulation is given. A practical example for the calculation of distances is presented using a discrete-event simulation model. By comparing different distance metrics with the real distances on an exemplary supply chain, it can be determined whether a satisfactory distance metric exists for calculating distances in urban areas, without the integration of an external route planning web service.

1 Introduction

Logistics is the basis for globalization and competitiveness. Without logistics, a successful economy or modern life cannot be shaped. A considerable amount of research has been carried out, aiming to optimize the organizational and operational practices of logistics. However, such attention has been mainly paid to inter-urban transportation, while the last mile is one of the most expensive, least efficient and most polluting sections of the entire logistics chain (Gevaers 2014). The transport volume in urban areas will continue to increase in the coming 25 years. The largest driver of this growth is e-commerce, which has shifted its focus from the B2B to the B2C segment (Joerss et al. 2016). Boyer et al. (2004) state that many e-commerce companies failed, because they could not figure out how to extend the supply chain directly to customers.

Therefore, we have conducted research to extend a supply chain simulation tool with further specific enhancements for the last mile delivery. In this context, a particular challenge is the detailed calculation of routes. If long distances across a variety of countries have to be covered along the supply chain, a rough approximation of the distance is sufficient outside of urban areas. It can be assumed that errors will be compensated during the long journeys. As the road distances are longer than the linear distance, due to the course of the road, a correction factor is used to approximate the linear distance to the road kilometres (Krarup and Pruzan 1980). This factor is between 1.1 and 1.5 and depends on the area. However, it is not evident whether this approximation can also be applied for last mile delivery or whether the error would become very large in this case.

The aim of this paper is to analyze whether a satisfactory distance metric exists for calculating distances in urban areas using a supply chain simulation tool, without the integration of an external route planning web service. There are multiple reasons for this: License issues with respect to the web services, a limited number of free-of-charge queries per day, or the necessity for a stable internet connection and a stable web service. Besides, the communication occurs directly over the Internet which leads to a poor performance. Due to performance reasons, a route planner in offline mode during the simulation is not an option either.

In this paper, we analyze the most common distance metrics and compare them to the real distances in an exemplary urban area, in order to verify whether a suitable distance metric exists. This is one of the main elements to ensure an integrated view of the whole supply chain network including last mile deliveries.

2 Urban Goods Movements and Supply Chain Simulation

Gonzalez-Feliu and Routhier (2012) analyze the different concepts for the construction of models and their further developments in the context of urban goods movement (UGM). In the course of this, a classification of the UGM models is defined, which can be divided into four groups: Models that analyze the current situation; models that simulate a current situation, optimization models, and discrete-event simulation models that create a reliable outcome forecast for the future. As a result, the authors recommend a strategy for creating meta-models by combining the existing models as this generates synergy effects. Kokkinogenis et al. (2011) consider several traffic simulation tools, which they divide into the following four levels: macroscopically, mesoscopically, microscopically, and nanoscopically. The elaborated overview describes seven simulation tools with a focus on simulations and functionalities that could be used to assess urban traffic. Friedrich (2010) stresses the importance of a meso level, to be able to aggregate or disaggregate between the macro and micro level. An example is the transition from transport demand of individual companies in the form of commodity flows (micro level) to the vehicle flows on the network (macro level), which is very complex. He also states that tours are often modelled by urban commercial traffic models, which do not reflect the perspective of a logistic decision maker. The simulation model SYNTRADE is introduced, which first generates a realistic logistic environment within the food industry, in order to determine warehouse structures using an optimization heuristic and to finally simulate the distribution to end customers. However, it seems that the abstraction level is too high to use route planning to end customers within the urban area.

When looking at forecasting modelling efforts, different classes of models can be detected, which highlight the urgent need for holistic approaches. In Fischer (2005) and Chow et al. (2010), two classes of freight demand models exist: Logistics models and vehicle touring models. Comi et al. (2012) state that the demand models have been classified in four models: Truck, commodity, delivery and mixed, the last one being the most promising model, as it enables a direct link between the interacting behaviours of commodity consumers and suppliers/shippers/retailers. Nevertheless, in the current urban freight demand modelling literature, the relation between these actors is not sufficiently investigated. In recent years, the attention for freight modelling has been growing, while in new research, logistics is integrated in freight models (Tavasszy 2012). Russo (2013) presents an integrated modelling system that enables a linkage between end customer choices and restocking decisions from retailers.

An overview of simulation-based models that focus on urban freight tours can be found in Holguín-Veras et al. (2013). Goyal et al. (2016) have shown the significant advantages of hyperconnected furniture logistics through a simulation-based analysis. Different scenarios have been created, e.g. to share distribution centres (DCs) or the fleet, or to change the location of some DCs. However, the focus still lies on the last mile delivery.

The field of supply chain simulation is very broad. For an initial overview, surveys on supply chain simulation tools and techniques can be found in Kleijnen (2005) and Terzi and Cavalieri (2004). There are many applications of supply chain simulation. Hence, we are focusing on Discrete-Event Simulation (DES) and the food industry as applied in this work. Hellström and Johnsson (2002) use DES for supply chain planning. The authors van den Vorst et al. (2009) developed a simulation tool called ALADIN™, which takes into account the quality deterioration of perishable goods in the transport chain. The aim is to reduce product waste.

The study of the literature has shown that most publications focus either on the simulation of UGM and tour planning or on the simulation of supply chains. Nevertheless, the holistic view of the entire supply chain is still missing. To our best knowledge, there is no published application that integrates supply chain simulation in the food industry with route planning algorithms to the last mile, except experimental individual case solutions. Quantitative assessments of supply chain efficiency metrics are considered as a main focus and form the basis for objective comparison of different scenarios. The simulation model specifically addresses the objectives and processes of supply chains and urban freight as well as the analyses of practical solutions for collaborative business behaviour, taking into account the specific requirements of food logistics. A first approach for the work described below has been made by Rabe et al. (2016).

In the area of network research, different methods of the graph theory are applied to the simulation in a logistical context, e.g. the Dijkstra algorithm. The problem with respect to simulation is that graph theory methods lead indeed to a fast calculation, but have high storage requirements, or vice versa (Alanis 2014). Thus, it is worth considering methods to estimate travelled distances.

Different approximation metrics for distance calculations are known and practically applied in the area of health service research, as part of health system research in general (Apparicio et al. 2008). Euclidean and Manhattan distances are often sufficiently accurate for research questions in this context. The considered nodes are not merged to a real network, but have instead a link to a special characteristic of the health system, known as star topology. Hence, the research subject cannot be considered as a network comparable to a supply chain network in city logistics.

Topics in other research areas require a more accurate distance approximation for simulation issues. For example, Cleophas and Ehmke (2014) use the weighted Euclidean distance as part of the discrete-event simulation for distance approximation, in order to explore the value-based demand fulfilment. The subject of their research activities is linked to the last mile delivery in a supply chain network. Hoerstebrock (2014) examines the electric mobility in the metropolitan area Bremen/Oldenburg (Germany). He uses a multi-agent system for the simulation of distance travelling and respectively the energy consumption. For distance calculations, the same method and factor from Cleophas and Ehmke (2014) are applied. While Cleophas and Ehmke do not offer an explanation for the choice of the weighting factor, Hoerstebrock (2014) notes that the weighting factor is detected by best-fit through different experiments.

A more precise distance approximation method is the Minkowski distance. The Minkowski distance can be seen as a generalization of the Euclidean and the Manhattan distance. The following calculation method specifies the Minkowski distance between two nodes a and b in $\mathbb{R}^n$:

$$d_{ab} = \left(\sum_{i=1}^{n} |a_i - b_i|^p \right)^{\frac{1}{p}} \tag{1}$$

$$d = distance$$

$$a, b \in \mathbb{R}^n$$

$$p = variable\ order$$

For the two-dimensional Euclidean space, the complex calculation method can be simplified to the following calculation scheme:

$$d_{ab} = \sqrt[p]{\left(a_i - a_j\right)^p + \left(b_i - b_j\right)^p} \tag{2}$$

For $p = 2$, it is the Euclidean distance; for $p = 1$, it is the Manhattan distance. Finally, for the values between 1 and 2, the Minkowski metric results in a wide range of distances.

Besides the Euclidean and Manhattan distance, the weighted Euclidean distance and the Minkowski distance, the exponential Euclidean distance is proposed for distance approximation in road distance calculation (Morris and Love 1972).

In the following, the conventional approximation methods like Manhattan or Euclidean distance are compared with more exact approximation methods like Minkowski distance as well as the weighted and exponential Euclidean distance. Figure 1 illustrates the different approximations.

In order to measure the quality of the different methods, two common comparison criteria in statistical geographic research (referred to as AD and SD) are applied. In this application case, AD sums up the absolute failures between real and approximated distance. SD represents the weighted sum of squared absolute failures between real and approximated distance (Morris and Love 1972). The criterion AD is useful, because long distances have to be approximated more precisely than short distances. The criterion SD does not focus on the difference between long and short distances like AD, but measures the goodness of fit of all considered distances. Additionally, some statistical key figures are applied for the comparison of the approximation metrics.

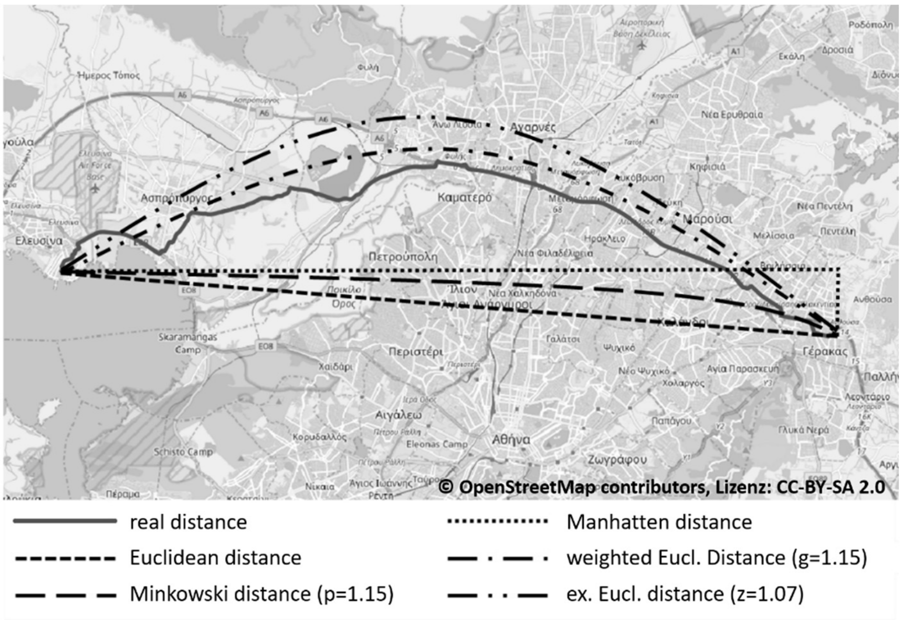

Figure 1: *Example of a delivery route in the metropolitan area of Athens*

3 Solution Approach

When looking at delivery practices, the cause-effect relationships within the whole supply chain should be considered. Therefore, we are using the simulation tool SimChain (Gutenschwager and Alicke 2004). It is a discrete-event supply chain simulation tool, which has been developed as a class library for the simulation tool Plant Simulation. SimChain consists of three major parts: A graphical user interface used for model configuration, a database in which all configuration data and simulation results are stored, and a DES Supply Chain Simulation Framework based on Plant Simulation. While the toolset offers features for order policy, stock and volume logistics or different means of transport, it lacks specific details in the consideration of city logistics and food transportation. Regarding the distance calculation, the simulation tool has no interface to a web service, but calculates the distances from the geo coordinates.

Our work deals with supply chain simulation including last mile deliveries for fresh food products in a historical European city and analyses the distance calculation of final deliveries to supermarkets in the urban area. The metropolitan area of Athens, Greece, has been selected for our case study, because it is the most populous and largest city in Greece and because there is immense road traffic in the narrowly developed centre of Athens. The area can be specified by the ZIP codes from 10xxx to 19xxx. Both the area with ZIP codes beginning with 18xxx and the islands of the Attica are excluded for further exploration, because we are focusing on city logistics. This led to a sample size of 6,779 routes. Figure 2 illustrates the area under investigation.

Figure 2: *Representation of the area under investigation*

At first, different approximation metrics (conventional and more newly reported) for distance calculations are selected from literature. The approximation metrics are applied to Attica as the area of this study. The results are explained through statistical key figures, in order to be able to compare distance approximation metrics, as shown in Table 1. Special attention is given to the weighted Euclidean distance with $g = 1.3$ as this is the approximation which is currently used for the simulation. Here, the distance for each trip is calculated as the Euclidean distance between origin and destination, multiplied with the factor 1.3.

The overall conclusion is that the different distance metrics achieve surprisingly good results according to the real distances.

At first, the maximum and minimum of each metric spread out a range similar to the one of the real distance with small differences, only. Furthermore, the standard deviation indicates whether the mean is representative for the whole distribution. The standard deviation of the approximation methods and the distribution of the real distances are very high and consequently not representative for the distribution itself. But, in this case, it is not necessary that the mean is representative for the whole distribution. It is more important that the approximation metrics generate standard deviations and means that are alike to the values of the distribution of real distances. The optimized approximation metrics are generating values in mean and standard deviation that are similar to the distribution of real distances. Only the Euclidean distance generates a lower range, a lower mean and a lower standard deviation in contrast to the real distances.

Table 1: *Key figures of different distance metrics (R = Range; σ = Standard Deviation; wg. = weighted; euc. = Euclidean; ex. = exponentiated).*

Metric	Min	Max	R	Mean	σ	AD	SD
Real distance	0.4	78.3	77.9	8.9	9.1		
Euclidean distance	0.2	59.4	59.2	7.1	7.3	1,1980.9	2,735.6
Manhattan	0.2	83.9	83.7	9.0	9.3	7,137.8	1,241.6
Minkowski distance ($p = 1.15$)	0.2	76.7	76.5	8.4	8.6	6,017.5	995.0
wg. euc. distance ($g = 1.225$)	0.2	72.4	72.2	8.7	8.9	4,334.0	540.8
wg. euc. distance ($g = 1.25$)	0.2	74.2	74.0	8.9	9.1	4,413.1	539.6
wg. euc. distance ($g= 1.3$)	0.3	77.2	76.9	9.2	9.5	5,184.8	682.7
ex. euc. distance ($z = 1.07$)	0.2	79.0	78.8	8.4	9.3	5,532.9	910.5
ex. euc. distance ($z = 1.08$)	0.2	82.3	82.1	8.6	9.6	5,589.5	896.3

Based on the criteria AD and SD, the conventional approximation methods result in weaker outcomes compared to the optimized forms of the newer approximation methods. The outcomes for AD and SD for the Euclidean distance are often twice as high as the outcomes for the newer approximation methods. Remarkable is the fact that the values for the criteria AD and SD for the Manhattan distance are higher than the other approximation metrics, but generate a surprisingly good distribution towards the real distances. Summing up the observations of AD and SD, the Euclidean and Manhattan distances are really weak for distance approximation. The focus should be on the newer approximation methods.

Even for the newer approximation metrics, large differences exist between the Minkowski distance and the weighted as well as the exponentiated Euclidean distance, taking into consideration the criteria AD and SD. As can be seen in Figure 1, in respect of the good results of the weighted Euclidean distance ($g = 1.25$), the real route is illustrated well. In relation to the criteria AD and SD, the weighted and exponentiated Euclidean distances generate better results than the optimized Minkowski distances (Tab. 1). The Euclidean distance weighted with ($g = 1.3$) produces the largest arithmetic mean and the second largest standard deviation.

The optimized p values regarding a specific distance metric are very near for both criteria AD and SD. Hence, it is not possible to approximate long distances better than short distances. This might be an indicator for a very dispersive structure of the road network in the considered research area. Therefore, the weighted or exponentiated Euclidean distances are more applicable than the Minkowski distance.

The above impressions are supported by the view on the boxplot diagram (Fig. 3). It visualizes the location and dispersion of failures between the real distance and the approximation of it. The mean of the weighted Euclidean distance ($g = 1.25$) is 0, which at the same time matches the median. Thus, the mean is not influenced by strong outliers in contrast to the weighted Euclidean distance ($g = 1.3$) and the Minkowski distance ($p = 1.15$). The mean of the weighted Euclidean distance ($g = 1.3$) is distorted downwards and for the Minkowski distance ($p = 1.15$) upwards.

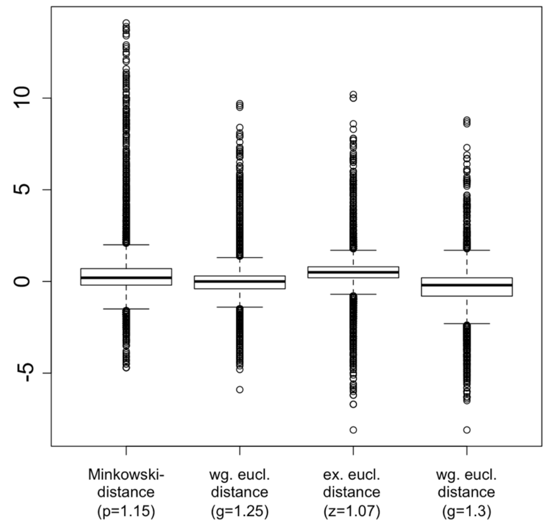

Figure 3: *Boxplot of several distance metrics*

Overall, it can be summarized that the distance approximation based on the weighted Euclidean distance with $g = 1.25$ is producing the best results compared to the other optimized distance metrics and the weighted Euclidean distance with $g = 1.3$ taken from the literature review. We can conclude that general parameters from the literature review do not necessarily fit properly to approximate the distance. Therefore, it is advisable to investigate the parameters for every approximation metric taking the specific properties of the network into consideration. With regard to the problem of expensive and least efficient delivery on the last mile in urban areas, it seems important to simulate the network and distances in order to ensure precise results.

4 Conclusion and Outlook

In this paper, we presented a possibility to integrate last mile deliveries to end customers in urban areas in supply chain simulation. For this purpose, we analysed whether a satisfactory distance metric exists for the calculation of distances, without the integration of an external route planning web service. We can conclude that the common approximations achieve excellent results compared to the real distances.

As a next step, we will test our concept for another city with a different road network. For example, the cityscape of an American city is very homogeneous and uniform, especially due to the chessboard layout. This is the exact opposite of medieval cities with their labyrinth of many small streets and alleys.

If it is not possible to find a good approximation, we have to manually calculate the real distances within the framework of data pre-processing, before running the simulation experiments. This means that we then need to find a way to provide this information, when generating the data model for the simulation.

References

Alanis, R.: Efficient storage of transport network routes for simulation models. In: Tolk, A.; Diallo, S.Y.; Ryzhov, I.O., Yilmaz, L.; Buckley, S.; Miller, J.A.: (Eds.): Proceedings of the 2014 Winter Simulation Conference (WSC), Savannah (USA) 2014, pp. 1931-1942.

Apparicio, P.; Abdelmajid, M.; Riva, M.; Shearmur, R.: Comparing alternative approaches to measuring the geographical accessibility of urban health services: distance types and aggregation-error issues. International Journal of Health Geographics 7 (2008) 7, pp. 7-21.

Boyer, K.K.; Frohlich, M.T.; Hult, T.G.: Extending the supply chain: how cutting edge companies bridge the critical last mile into customers' homes. New York: American Management Association 2004.

Chow, J.Y.J.; Yang, C.H.; Regan, A.C.: State-of-the art of freight forecast modeling: lessons learned and the road ahead. Transportation 37 (2010) 6, pp. 1011–1030.

Cleophas, C.; Ehmke, J.F.: When are deliveries profitable? - Considering order value and transport capacity in demand fulfillment for last-mile deliveries in metropolitan areas. Business & Information Systems Engineering 6 (2014) 3, pp. 153-163.

Comi, A.; Delle Site, P.; Filippi, F.; Nuzzolo, A.: Urban freight transport demand modelling: a state of the art. European Transport/Trasporti Europei 51 (2012) 7, pp. 1-17.

Fischer, M.; Outwater, M.; Cheng, L.; Ahanotu, D.; Calix, R.: Innovative framework for modeling freight transportation in Los Angeles county, California. Transportation Research Record: Journal of the Transportation Research Board, No. 1906. Washington, D.C.: Transportation Research Board of the National Academies 2005, pp. 105-112.

Hellström, D.; Johnsson, M.: Using discrete event simulation in supply chain planning. In: Solem, O. (Eds.): Proceedings of the 14th annual conference for Nordic researchers in logistics. Trondheim: NOFOMA 2002, pp. 13-28.

Holguín-Veras, J., Thorson, E.; Wang, Q.; Xu, N.; González-Calderón, C.; Sánchez-Díaz, I.; Mitchell, J.: Urban freight tour models: state of the art and practice. In: Ben-Akiva, M.; Meersman, H.; Van de Voorde, E.: Freight Transport Modelling. Bingley: Emerald Group Publishing Limited 2013, pp. 335-352.

Gevaers, R.; van de Voorde, E.; Vanelslander, T.: Cost modelling and simulation of last-mile characteristics in an innovative B2C supply chain environment with implications on urban areas and cities. In: Procedia – Social and Behavioral Sciences, 125 (2014), pp. 398-411.

Goyal M.; Cook, J.; Montreuil, B.; Kim, N.; Lafrance, C.: Hyperconnected city logistics for furniture and large appliance industry: Simulation-based Exploratory Investigation, 3rd International Physical Internet Conference. July 2016. http://www.pi.events/IPIC2016/sites/default/files/Workshop_TA1_Hyperconnected_City_Logistics_for_Furniture_Large_Appliance_Industry.pdf. Last access: 12.05.2017.

Gutenschwager, K.; Alicke, K.: Supply Chain Simulation mit ICON-SimChain. In: Spengler, T.; Voß, S.; Kopfer, H. (Eds.): Logistik Management. Heidelberg: Physica 2004, pp. 161-178.

Hoerstebrock, T.: Strategische Analyse der Elektromobilität in der Metropolregion Bremen/Oldenburg – Multi-Agenten basierte Simulation alternativer Antriebssysteme. Diss. Oldenburg: Carl von Ossietzky Universität Oldenburg, 2014.

Kleijnen, J.P.C.: Supply chain simulation tools and techniques: A survey. International Journal of Simulation and Process Modelling 1 (2005) 1-2, pp. 82-89.

Joerss, M.; Schröder, J.; Neuhaus, F.; Klink, C.; Mann, F.: Parcel delivery: The future of last mile. McKinsey&Company study. Sept. 2016. https://www.mckinsey.de/files/2016_the_future_of_last_mile.pdf. Last access 10.02.2017.

Krarup, J.; Pruzan, P. M.: The impact of distance on location problems. European Journal of Operational Research, 4 (1980) 4, pp. 256-269.

Morris, J.G.; Love, R.F.: Modelling inter-city road distance by mathematical functions. Journal of the Operational Research Society 23 (1972) 1, pp. 61-71.

Rabe, M.; Klueter, A.; Clausen, U.; Pöting, M.: An approach for modeling collaborative route planning in supply chain simulation. In: Roeder, T.M.K.; Frazier, P.I. ; Szechtman, R.; Zhou, E.; Huschka, T.; Chick, S.E.: Proceedings of the 2016 Winter Simulation Conference, Arlington (VA). Piscataway: IEEE 2016, pp. 2228-2238.

Russo, F.: Modeling behavioral aspects of urban freight movements. In: Ben-Akiva, M.; Meersman, H.; Van de Voorde, E.: Freight Transport Modelling. Bingley: Emerald Group Publishing Limited 2013, pp. 353-375.

Tavasszy, L.A.; Ruijgrok, K.; Davydenko, I.: Incorporating logistics in freight transport demand models: state-of-the-art and research opportunities. Transport Reviews 32 (2012) 2, pp. 203-219.

Terzi, S.; Cavalieri, S.: Simulation in the supply chain context: a survey. Computers in Industry 53 (2004) 1, pp. 3-16.

van der Vorst, J.G.; Tromp, S.-O.; van der Zee, D.-J.: Simulation modelling for food supply chain redesign; integrated decision making on product quality, sustainability and logistics. International Journal of Production Research 47 (2009) 23, pp. 6611-6631.