Supply Chain Simulation in a Big Data Context: Risks and Uncertainty Analysis

António A. C. Vieira(✉), Luís M. S. Dias, Maribel Y. Santos, Guilherme A. B. Pereira, and José A. Oliveira

ALGORITMI Research Centre, University of Minho, Braga, Portugal
{antonio.vieira, lsd, gui, zan}@dps.uminho.pt,
maribel@dsi.uminho.pt

Abstract. Due to their complex and dynamic nature, Supply Chains are prone to risks that may occur at any time and place. To tackle this problem, simulation can be used. However, such models should use Big Data technologies, in order to provide the level of data and detail contained in the data sources associated to the business processes. In this regard, this paper considered a real case of an automotive electronics Supply chain. Hence, the purpose of this paper is to propose a simulation tool, which uses real industrial data, provided by a Big Data Warehouse, and use such decision-support artifact to test different types of risks. More concretely, risks in the supply and demand end of the network are analyzed. The presented results also demonstrate the possible benefits that can be achieved by using simulation in the analysis of risks in a Supply Chain.

Keywords: Simulation · Supply chain · Big data · Risk management

1 Introduction

Supply Chains (SC) operate under 2 dimensions: uncertainty and complexity [1]. This nature has been further enhanced by currently adopted industrial practices (e.g., just-in-time, shorter product life cycle). On one hand, this makes modern SC leaner costlier, greener, with fewer buffers and stored materials; on the other hand, this also exposes organizations to disruptions when certain events occur [2]. These events result in unanticipated and unpredictable consequences, which affect the performance of individual entities of a SC and other entities with relationships with the affected one, possibly including several SC [3]. As Sodhi et al. [4] assert, consequences can come in revenue, reputation and other types of losses. In fact, several examples exist in literature which report on such cases, e.g. [1, 3, 5–7]. The impact of such events may be enormously negative in several ways. Thus, companies require tools to allow them to mitigate such consequences. Despite this, few companies have succeeded in taking actions to smooth these negative impacts [2, 3, 8].

The aforementioned events are characterized by being very rare, unpredictable and of great impact on the performance of the SC [3]. Due to these characteristics, organizations struggle to deal with them, as decision-makers lack information about the SC network and environment, hence being unable to predict where, when and how much impact can a given event deliver. The uncertainty surrounding the possible outcomes of

S. Misra et al. (Eds.): ICCSA 2019, LNCS 11619, pp. 817–829, 2019.
https://doi.org/10.1007/978-3-030-24289-3_60

the aforementioned events, portrays an exposure to what literature refers to as risks [9]. Thus, risks can be interpreted as events that may occur at any part of the SC (uncertainty), causing negative impact on it, in many possible ways.

Ho et al. [5] argued that these risks can be classified in different ways, including: internal or external; operational or disruption; and others. The authors also provided their own framework for classifications of SC risks, included:

- Manufacturing or internal risks: occur within a plant;
- Supply: occur in the supplier's side of the SC;
- Demand: occur in the customer's side of the SC;
- External: events very rare with severe consequences to SC, generally consisting in risks of natural order (e.g., weather-related or earthquakes) or man-made (e.g., wars, terrorist attacks, political-related), which have origin outside the SC.

Following the author's risks classification, this paper proposes a SC simulation model, developed in SIMIO [10], which assesses the impact of supply and demand risks in a SC of the automotive industry, by using real industrial data form a plant of the Bosch organization. The data is originated from several data sources and a Big Data Warehouse (BDW) was implemented to store, integrate and provide such data to the simulation model. Thus, the presented SC simulation model uses Big Data to test the system's performance under supply and demand risks.

The research that was conducted to develop the BDW has already been published [11], as well as a prototype of the simulation model [12]. This previously presented model did not use real industrial data, but rather used the typically available approach in simulation, consisting in using random distributions, to validate the data model, i.e., the variables selected for the project. Having validated such data model, the next step in the project was to complement the simulation model, so that it is capable of using data provided by the BDW and assess certain types of risks.

This paper is organized as follows. Next section analyzes related works in literature. Section 3 provides a characterization of the SC under study. Section 4 briefly describes the development approach that was followed, focusing on the data modeling in a Big Data context and in the data-driven simulation approach. Section 5 presents the obtained results by using the simulation model to test the performance of the SC under supply and demand risks. Finally, conclusions and future research directions are provided in the last section.

2 Related Work

The literature of SC simulation studies is vast. However, to the best of the authors' knowledge, it can be argued that such solutions, in what regards the use of Big Data technologies, are limited. This is corroborated by several studies [13–15]. With the lack of such studies, this section focuses on reviewing the SC studies that applied simulation using some type of external data storage (e.g. relational databases) that provides data to the proposed simulation model.

Cheng et al. [16] used GBSE (General Business Simulation Environment) to propose a simulation model to help making tactical level decisions in a SC. Their simulation model offered 2D visualization and a direct connection to an external database. The authors tested the response of holding cost, transportation cost and order delivery time to different forecast accuracy levels, production strategies (namely, make-to-order, make-to-stock and postponement) and special transportation costs.

Fornasiero et al. [17] and Macchion et al. [18] proposed a SIMIO simulation model, which assessed the impact of orders size in the SC performance. In [17], Fornasiero et al. based their experiments on order size, lead time variation and supplier scrap rate to achieve findings regarding the impact on the delivery time to customer, customer order quality and inventory costs. Later, Macchion et al. [18] assessed the impact of order size, inventory management policy, supplier's lead time and quality on inventory level and order lead time. Fornasiero et al. [17] applied their model to a fashion industry comprised by 60 suppliers and 1 manufacturer, whilst Macchion et al. [18] applied it to a SC of the footwear industry comprised by 4 suppliers, 1 warehouse, 1 manufacturer, 1 distributor and 2 customers. In both studies, the authors reported that their simulation models are able to retrieve data from the ERP system.

Sahoo and Mani [19] presented a simulation model in ExtendSim to model a SC of the biomass industry. The modelled SC comprised producer and farmer of biomass and suppliers which transported the raw materials to the plant, which could store them or process them for later bioenergy production, in order to deliver heat and electricity to customers. The simulation model used a direct connection to a database, which, among other operational data, stored weather data for long time periods. This data was used by the authors to test the exposure of the SC to KPIs (Key Performance Indicators) such as: quantity produced, inventory levels, transportation quantity, transportation costs and handling cost. Lastly, Ponte et al. [20] evaluated the impact that inventory management and different forecast methods have on the demand variation propagation upstream the SC.

3 Supply Chain Characterization

This project is being developed at a plant of the Bosch Group, which concerns with producing automotive electronic components. This section first briefly describes the SC at hand, to give a perspective of the scale and complexity of the network in analysis. Figure 1 shows the countries and the number of suppliers per country, which supply materials to the plant.

The numbers in each country represent the number of suppliers from that country. The lines placed between these countries and the plant represent the number of material shipments; the width is proportional to the number of shipments. Finally, the color scale of each country is associated to the number of different types of materials that suppliers from that country provide to the plant. According with the considered data, around 7 000 different types of materials are actively being supplied by roughly 500 different suppliers, located in more than 30 countries. Moreover, as the figure shows, Germany, Netherlands, Switzerland, Spain, China, Taiwan and Malasya are the countries that supply more types of materials. Also suggested by Fig. 1, the plant received materials from more than 400 suppliers, especially from Europe and Asia, with Germany (209 suppliers) and Netherlands (10 suppliers) having more suppliers and shipments from Europe, and Malasya (16 suppliers), Taiwan (13 suppliers), China (12 suppliers), Hong Kong (11 suppliers) and Singapore (7 suppliers) having more shipments from Asia. All these suppliers shipped more than 200 000 deliveries, during the last year. Figure 2 illustrates a summary of the main material and information flows of this SC system.

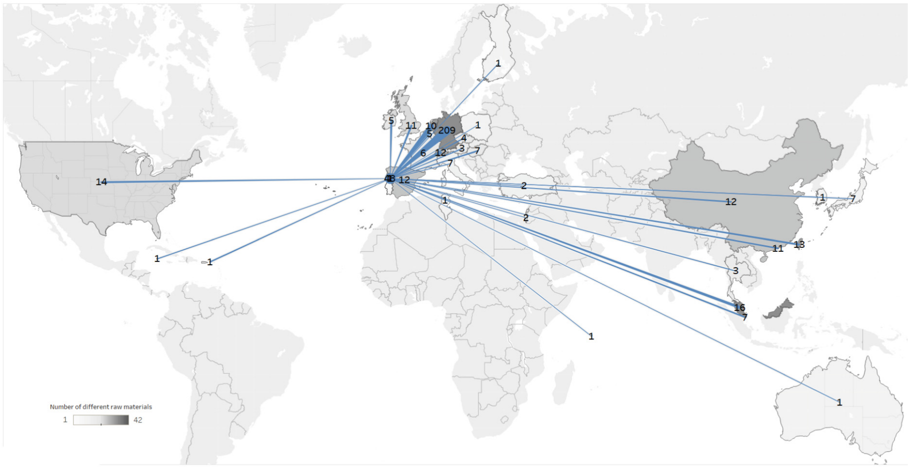

Fig. 1. Geographic location, number of suppliers, different raw materials provided and number of shipments per suppliers.

As the below figure depicts, to comply with final customers' orders, the plant places orders to its suppliers, which later culminate in arrivals. Most of these arrivals occur within the scheduled date. However, some suppliers provide the orders before the scheduled data. In these situations, these materials are stored in a special warehouse until the scheduled date is met. This way, the plant is not responsible for storage costs of these materials. On the other hand, supplier deliveries may be delayed, potentially resulting in orders arriving after the scheduled data. Whilst the previous situation (materials arriving before the scheduled date) may originate high warehousing costs, this last situation may originate material shortages, which can jeopardize the production and hence customers' orders. To bypass this, special freights are scheduled, which are usually faster but also have considerably higher associated costs.

When materials arrive to the plant, the contents are examined to assess their quality and if the order requirements were met. Afterwards, the materials are put in the respective storage unit, e.g. boxes or pallets, in the respective quantities, to be stored in the warehouse. The warehouse is divided in 2 main locations, both storing raw materials, but storage location 1 is being used for electronic components, while the other for bulkier materials. In its turn, each storage location is divided in multiple bins, wherein each one stores a type of material and a storage unit (e.g., box), at the same time. When materials are sent to the warehouse, they are always allocated to an empty storage bin.

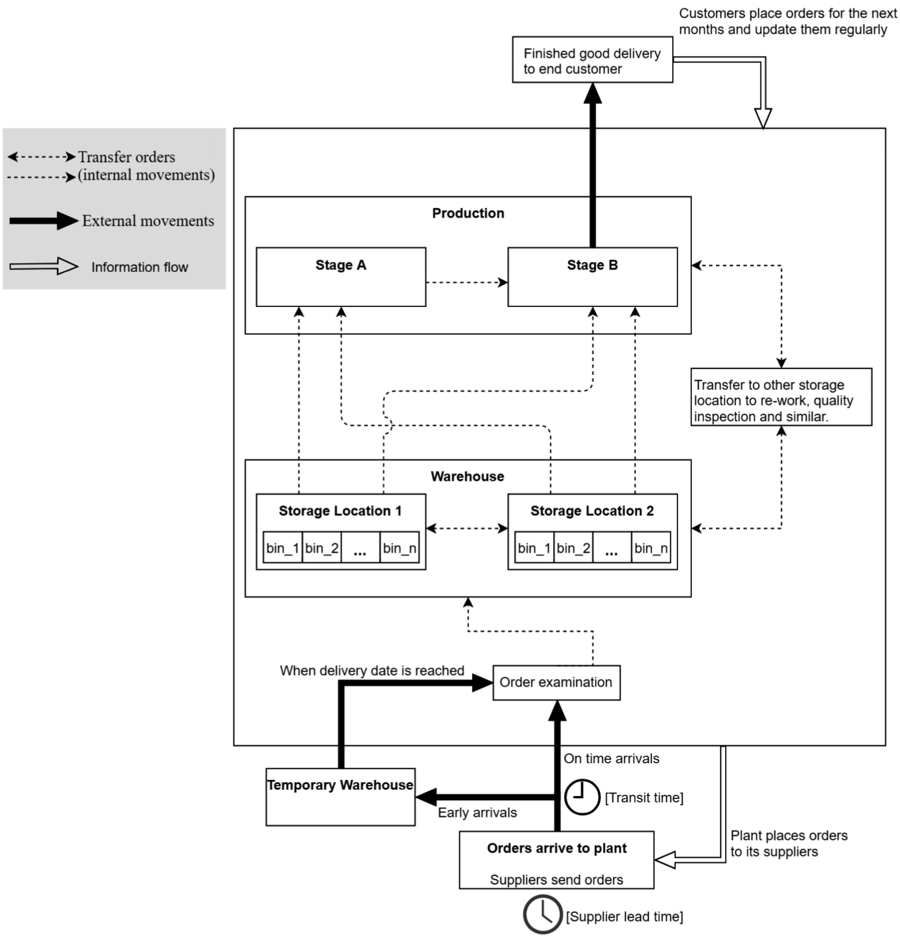

Fig. 2. Summary of material and information flows of the SC system.

The production is divided in 2 main areas: Stage A, where the main electronic components are produced; and Stage B, where the finished goods are produced, using electronic raw materials stored in the warehouse (form either storage location) and also other components produced in Stage A. Finally, when production in Stage B finishes, materials are ready to be delivered to final customers.

Figure 2 also shows that besides the main movements, others exist, which represent movements for quality inspection, re-works and other similar activities required to ensure the quality of the final product. It is important to consider these movements, as they represent materials temporarily not available to be transferred to production. As can be observed, these movements are represented with bidirectional arrows, as the materials are later transferred back to another bin of the warehouse, or back to the production, depending on the situation. The movements depicted in Fig. 2 represent the main ones that occur in the plant. However, there are many others, most of them used for quality inspection, re-work and similar tasks.

4 Proposed Approach

This section describes the modeling approach applied in this project. The first step consisted in identifying the relevant business processes to include in the simulation model. Such processes entail data sources which are used by managers of the plant. Therefore, those data sources must be carefully studied in order to identify the relevant variables to include in the project. In fact, roughly 2 000 variables were analyzed, which culminated in the inclusion of around 200 in the BDW. After determining the relevant variables, further interviews and data analysis are required in order to identify possible treatments that have to be performed on the data to store in the BDW. Such data treatments are required, for several reasons, such as empty fields or wrong values. Having defined such treatments, the traditional Extract-Transform-Load (ETL) process was conducted, which ended with storing the data in the BDW.

To store such data in the BDW, despite being in a Big Data context, data schemas should be defined, as it revealed the following benefits: (1) better understanding of the data, organizational processes and relevant KPIs to include in the BDW; (2) ensures the inclusion of the all relevant data, making sure that no important attributes were excluded; (3) helps in the definition of the Hive tables to use [11]. Due to these reasons, the next step in the project comprised, in fact, the dimensional model, proposed by Kimball (see [21]). However, Costa et al. [22] analyzed the performance of Hive-based DWs in Big Data environments (i.e., a BDW), having concluded that, in Big Data environments, fully denormalized tables outperform tables in the dimensional format, following the star schema design pattern. Furthermore, the authors also concluded that adequate data partitioning strategies data showed a clear reduction in the query execution time. This was analyzed in more detail in [23].

Therefore, after the multidimensional model, which provided the above discussed benefits, the tables were denormalized, hence creating the BDW tables. Figure 3 shows a simplified version of the dimensional model that was defined for this project, and the consequent BDW tables which were thereafter created.

As the figure shows, five BDW tables were defined. One stores data of orders placed to suppliers and the posterior receipts (OrdersSentAndReceived). The remaining four represent movements that occur within the plant. With these tables, the simulation model was connected to the Big Data cluster to retrieve the data. Afterwards, the model was adapted to reflect this stored data.

The model runs in a built-in 3D environment with only one physical object, which sets the location of the plant. Thus, entities travel without any links (since there are no physical objects), with their movements being specified by processes which model the behavior of the entities, according with the data in the BDW. Due to this data-driven approach, the simulation model is able to automatically adapt to data changes in the BDW. Figure 4 shows an example of such processes, which models the lead time of the suppliers, and the posterior material shipment to the plant. In its turn, Fig. 5 illustrates the simulation model reproducing the historical data stored in the BDW.

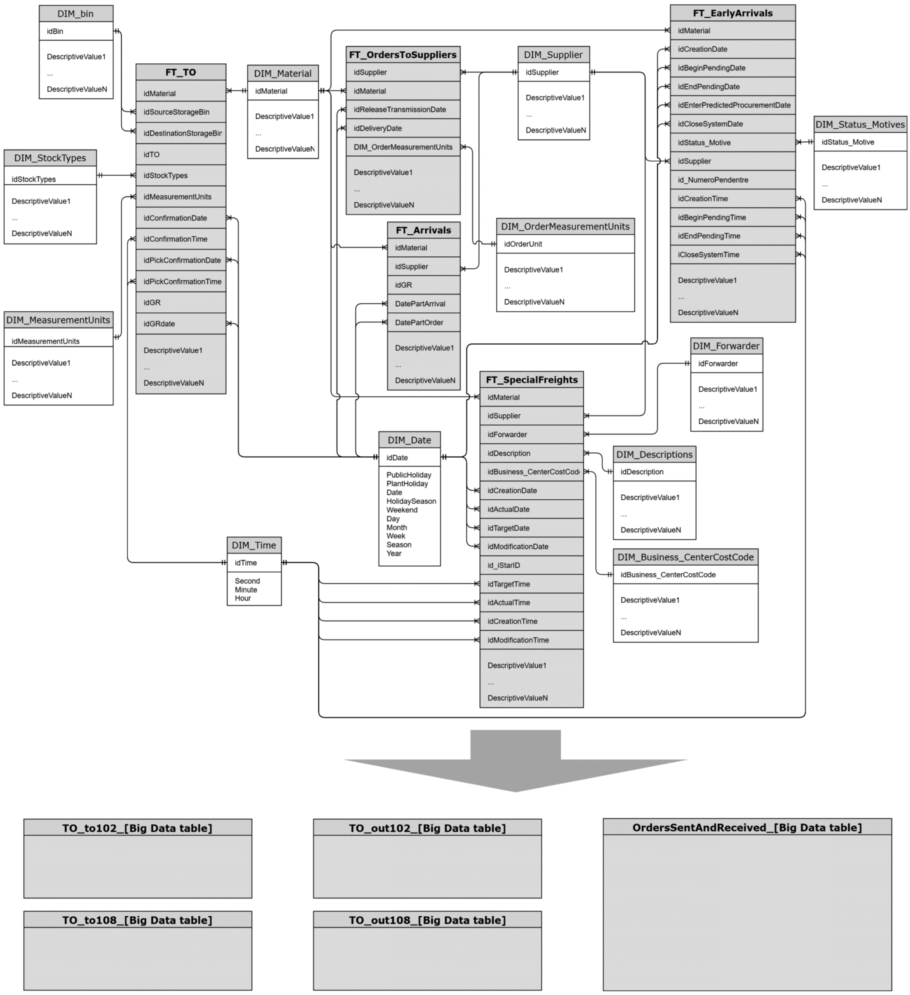

Fig. 3. Traditional dimensional model (top part of the figure) and created BDW tables (bottom part of the figure).

The figure shows yellow circles, which represent orders in production in the respective geographic supplier location; the location of these entities represents the exact location of the supplier, as stored in the BDW, however, as a supplier may have multiple orders at the same time, a small deviation in the location of each order is applied, so that it is possible to see all entities. Finally, the number presented above each yellow entity represents the number of days remaining for the order to be shipped to the plant. This number decreases as the simulation clock advances in time. When it is time to ship the order, the symbol of the orders change to the respective transport type and their speed is also adjusted to the transportation lead time represented in the BDW.

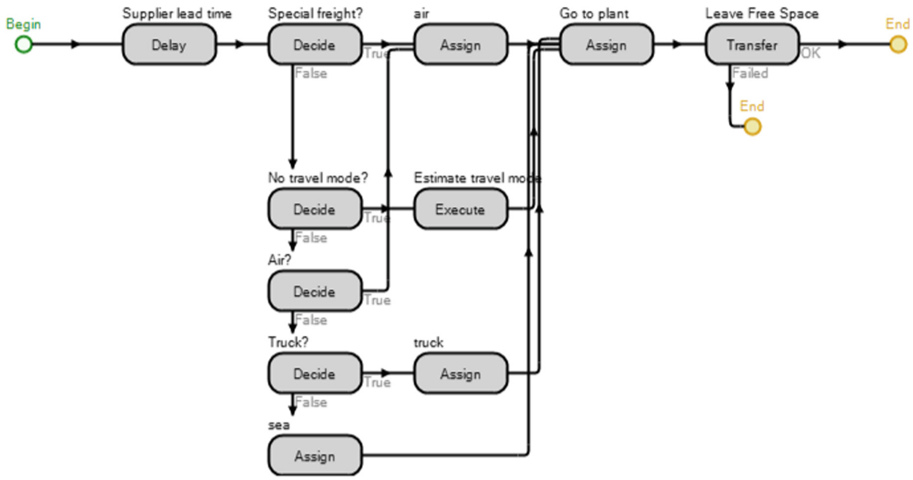

Fig. 4. Process executed to model the production and shipment of orders to the plant.

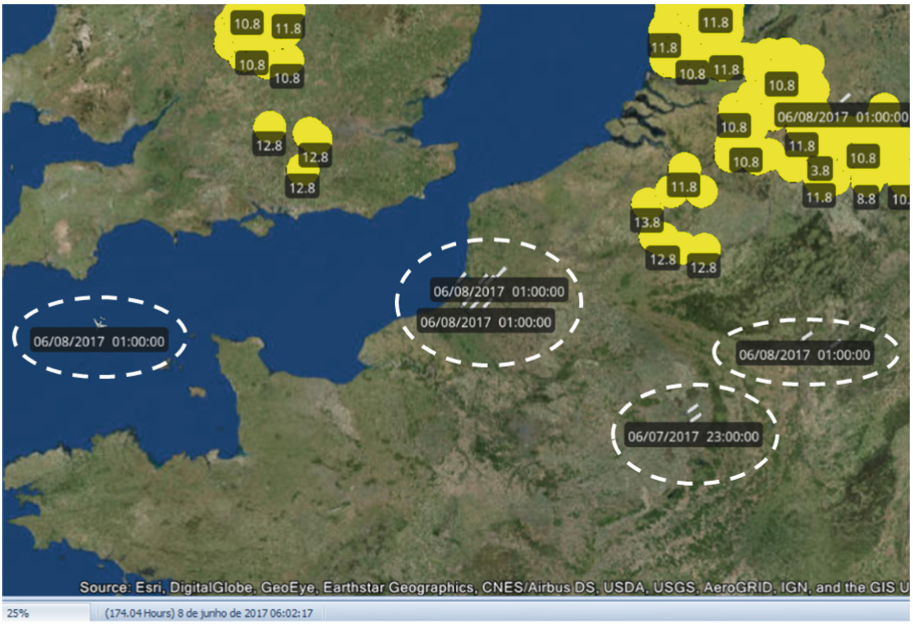

Fig. 5. Orders being sent to the plant. (Color figure online)

Figure 5 shows some of these entities highlighted. The date time values below each entity represent the instant when those deliveries were shipped to the plant. When orders arrive to the plant, they are stocked, so that they can be managed by the plant's internal material movements.

5 Supply Chain Simulation in a Big Data Context

This section comprises both supply and demand risks, each one analyzed in one of the two following subsections. Hence, the experiments conducted in this section consist in using the simulation model to reproduce the historical data stored in the BDW, while also incorporating risks, through the utilization of random distributions. Thus, these experiments must consider a given number of replications, which will use different random seeds, thus attenuating the differences provided by such distributions. In this regard, 10 replications were executed for these scenarios.

5.1 Supply Risks

In this experiment, a variable lead time was applied to all orders placed to suppliers, by considering a triangular distribution with minimum of 0 days (no delay), mode of 2 days and maximum of 5 days. This way, all orders will either arrive at the date specified in the BDW, or later, allowing the impact of such delay to be analyzed. The opposite to this would be to analyze the orders that arrive before the schedule date. However, as explained in Sect. 3, these cases do not have a significant impact on the plant, as the plant does not incur in excessive warehousing costs of storing these materials in the dedicated warehouse for early arrivals. Figure 6 shows the number of arrivals per day, illustrated with a greed-red color scale for supplier delays of 0 to 5 days, respectively.

The below figure shows that the number of orders that arrived on-time is minimum and that the orders that arrive with 2 days of delay were the most common case, which is in accordance to the distribution used to set the supplier's delay time. To analyze the impact that such delays could have on the performance of the plant, Fig. 7 shows the

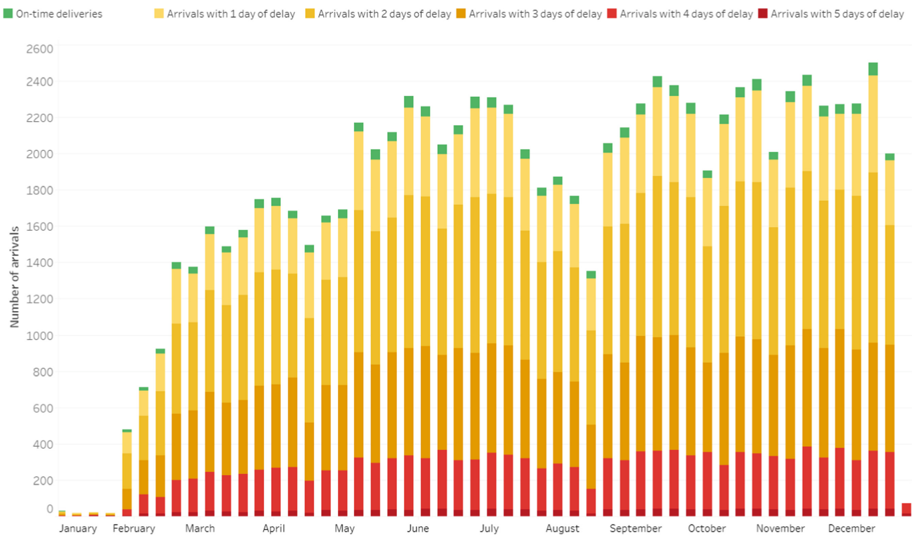

Fig. 6. Number of arrivals per week and the respective arrival delay time

total special freights costs and the total number of unfilled orders that these delays originated. As can be seen, the maximum number of unfilled orders per week was 350, which totalized more than 60 000 000 € with special freights during the year.

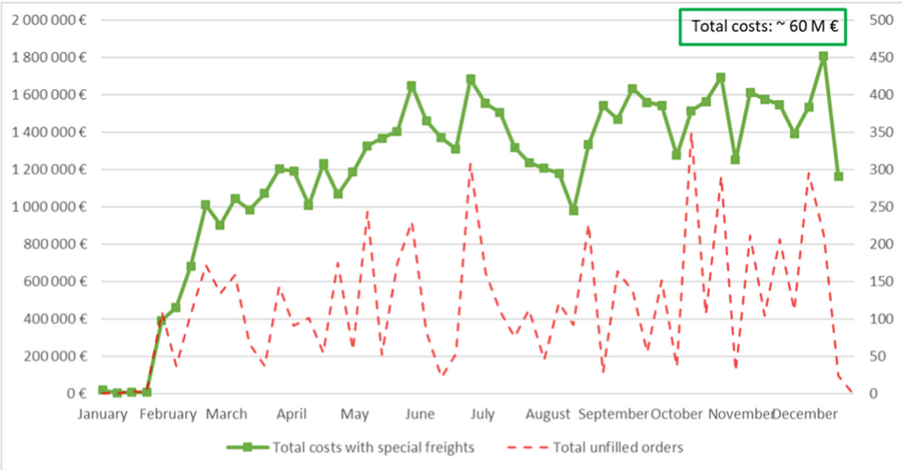

Fig. 7. Total special freights costs and number of unfilled orders per week due to the supplier's delay time.

5.2 Demand Risks

To analyze this type of risks, a triangular distribution with a minimum of 0%, a mode of 30% and a maximum of 100% added to the original production's orders quantity was considered. Figure 8 shows the impact on the stock by applying this variability to the production orders, as well as the associated total number of unfilled orders per week.

The figure shows the stock level decreasing as expected, and the number of unfilled orders increasing since the beginning of the simulation, which is explained by the fact that, in some situations, there were enough material buffers to fulfill the increased demand. In their turn, the materials that did not have enough buffers in stock result in unfilled orders, later in the simulation.

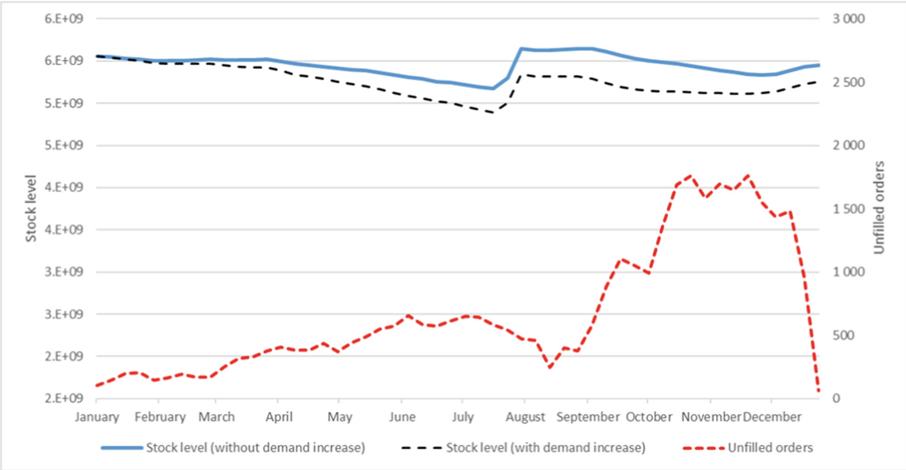

Fig. 8. Stock level and total number of unfilled orders per week for (1) normal demand and (2) increased demand scenarios.

6 Conclusions

The main contributions from this paper derives from the presented simulation model, which is able to reproduce the historical Big Data stored in the BDW. In fact, while some studies exist which reported the use of data storage and analysis with simulation, no study was found coupling the benefits of Big Data technologies with the latter. The simulation model also allows certain types of variability (e.g., delays and quantity required by production) to be considered. This way, apart from reproducing the historical data from the BDW, the simulation model also incorporates disruption scenarios.

The solution's animation feature and its integration with a Google Maps view should also be stressed because, from the analyzed literature, the lack of studies investing in animation features of their SC systems was noteworthy. Thus, the simulation runs on a 3D world map view, which enriches its visualization, with all the associated benefits.

While out of the scope of this paper, other features of the presented solution can also be emphasized, such as its ability to automatically retrieve data from the BDW and adapt to changes or new data, hence matching the main characteristics of a real-time simulation. In fact, none of the suppliers' locations, as well as the entities' movements, were manually modelled, since the model is able to automatically adjust its elements to the data it gets from the BDW; i.e., the model is drawn by the data and not by the users. In addition, the user may fire disruptive events at any given time, duration and location (e.g., disrupt suppliers in a geographic location) and use the simulation model to assess its impact in the performance of the plant.

Lastly, regarding the area of applicability, this paper targeted a SC larger (in number of suppliers from different countries and number of materials) than the most commonly analyzed in literature using simulation methods. The fact that this is the case of an automotive electronics industry also adds complexity to the problem, as these SC are typically characterized by having single sourced materials, with suppliers typically providing multiple materials, thereby exposing the entire SC to risks.

Regarding the future research in this project, the following items can be highlighted: ensure that the BDW is able to refresh its data in real time; use the simulation model to project future scenarios, rather than reproducing the historical data stored in the BDW; and, use the simulating model to fire notifications when certain unusual behaviors are detected, e.g., a given supplier is taking more time than usual to produce a given order.

Acknowledgments. This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019 and by the Doctoral scholarship PDE/BDE/114566/2016 funded by FCT, the Portuguese Ministry of Science, Technology and Higher Education, through national funds, and co-financed by the European Social Fund (ESF) through the Operational Programme for Human Capital (POCH).

References

1. Colicchia, C., Strozzi, F.: Supply chain risk management: a new methodology for a systematic literature review. Supply Chain Manag. **17**(4), 403–418 (2012)
2. Ghadge, A., Dani, S., Kalawsky, R.: Supply chain risk management: present and future scope. Int. J. Logist. Manag. **23**(3), 313–339 (2012)
3. Fahimnia, B., Tang, C.S., Davarzani, H., Sarkis, J.: Quantitative models for managing supply chain risks: a review. Eur. J. Oper. Res. **247**(1), 1–15 (2015)
4. Sodhi, M.S., Son, B.-G., Tang, C.S.: Researchers' perspectives on supply chain risk management. Prod. Oper. Manag. **21**(1), 1–13 (2012)
5. Ho, W., Zheng, T., Yildiz, H., Talluri, S.: Supply chain risk management: a literature review. Int. J. Prod. Res. **53**(16), 5031–5069 (2015)
6. Sáenz, M.J., Revilla, E.: Creating more resilient supply chains. MIT Sloan Manag. Rev. **55**(4), 22–24 (2014)
7. Rao, S., Goldsby, T.J.: Supply chain risks: a review and typology. Int. J. Logistics Manag. **20**(1), 97–123 (2009)
8. Fan, Y., Stevenson, M.: A review of supply chain risk management: definition, theory, and research agenda. Int. J. Phys. Distrib. Logist. Manag. **48**(3), 205–230 (2018)
9. Kilubi, I.: The strategies of supply chain risk management – a synthesis and classification. Int. J. Logist. Res. Appl. **19**(6), 604–629 (2016)
10. Dias, L.M.S., Vieira, A.A.C., Pereira, G.A.B., Oliveira, J.A.: Discrete simulation software ranking-a top list of the worldwide most popular and used tools. In: Proceedings - Winter Simulation Conference (2017)
11. Vieira, A.A., Pedro, L., Santos, M.Y., Fernandes, J.M., Dias, L.M.: Data requirements elicitation in big data warehousing. In: European, Mediterranean and Middle Eastern Conference on Information Systems (EMCIS) (2018)
12. Vieira, A.A., Dias, L.M., Santos, M.Y., Pereira, G.A., Oliveira, J.A.: Simulation of an automotive supply chain in Simio: data model validation. In: European Modeling and Simulation Symposium (EMSS) - Part of International Multidisciplinary Modeling & Simulation Multiconference (I3 M) (2018)

13. Tiwari, S., Wee, H.M., Daryanto, Y.: Big data analytics in supply chain management between 2010 and 2016: insights to industries. Comput. Ind. Eng. **115**, 319–330 (2018)
14. Zhong, R.Y., Newman, S.T., Huang, G.Q., Lan, S.: Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. Comput. Ind. Eng. **101**, 572–591 (2016)
15. Vieira, A.A., Dias, L.M., Santos, M.Y., Pereira, G.A., Oliveira, J.A.: Setting an industry 4.0 research and development agenda for simulation – a literature review. Int. J. Simul. Model. **17**(3), 377–390 (2018)
16. Cheng, F., Lee, Y.M., Ding, H.W., Wang, W., Stephens, S.: Simulating order fulfillment and supply planning for a vertically aligned industry solution business. In: Proceedings of the 40th Conference on Winter Simulation, pp. 2609–2615 (2008)
17. Fornasiero, R., Macchion, L., Vinelli, A.: Supply chain configuration towards customization: a comparison between small and large series production. IFAC-PapersOnLine **28**(3), 1428–1433 (2015)
18. Macchion, L., Fornasiero, R., Vinelli, A.: Supply chain configurations: a model to evaluate performance in customised productions. Int. J. Prod. Res. **55**(5), 1386–1399 (2017)
19. Sahoo, K., Mani, S.: GIS based discrete event modeling and simulation of biomass supply chain. In: Proceedings - Winter Simulation Conference, vol. 2016-February, pp. 967–978 (2016)
20. Ponte, B., Sierra, E., de la Fuente, D., Lozano, J.: Exploring the interaction of inventory policies across the supply chain: an agent-based approach. Comput. Oper. Res. **78**, 335–348 (2017)
21. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley (2011)
22. Costa, E., Costa, C., Santos, M.Y.: Efficient big data modelling and organization for Hadoop hive-based data warehouses. In: Themistocleous, M., Morabito, V. (eds.) EMCIS 2017. LNBIP, vol. 299, pp. 3–16. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65930-5_1
23. Costa, E., Costa, C., Santos, M.Y.: Evaluating partitioning and bucketing strategies for Hive-based Big Data Warehousing systems. J. Big Data **6**(1), 34 (2019)