Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 42 (2020) 132–139

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019)

Bypassing Data Issues of a Supply Chain Simulation Model in a Big Data Context

António A. C. Vieira[a,*], Luís Dias[a], Maribel Y. Santos[a], Guilherme A. B. Pereira[a], José Oliveira[a]

[a]*ALGORITMI Research Centre, University of Minho, 4804-533, Portugal*

* Corresponding author. *E-mail address:* antonio.vieira@dps.uminho.pt

Abstract

Supply Chains (SCs) are complex and dynamic networks, where certain events may cause severe problems. To avoid them, simulation can be used, allowing the uncertainty of these systems to be considered. Furthermore, the data that is generated at increasingly high volumes, velocities and varieties by relevant data sources allow, on one hand, the simulation model to capture all the relevant elements. While developing such solution, due to the inherent use of simulation, several data issues were identified and bypassed, so that the incorporated elements comprise a coherent SC simulation model. Thus, the purpose of this paper is to present the main issues that were faced, and discuss how these were bypassed, while working on a SC simulation model in a Big Data context and using real industrial data from an automotive electronics SC. This paper highlights the role of simulation in this task, since it worked as a semantic validator of the data. Moreover, this paper also presents the results that can be obtained from the developed model.

Keywords: Simulation; Supply Chain; Big Data; Data issues; Industry 4.0

1. Introduction

Supply Chains (SCs) are complex and dynamic systems where a proper assessment of their performance is hard to quantify. Simulation can be used as decision-making tools of SC systems, allowing alternative scenarios to be tested, performance measures to be determined, or simply to animate the logistics flows, enhancing the knowledge discovery from raw data. However, SC processes generate huge amounts of data, nowadays referred to as Big Data. Thus, such decision-making tools benefit from Big Data structures, which provide quality and integrated data for SC simulation models [1], [2].

Aligned with the above, such artifact is currently being developed at an organization of the automotive electronics industry sector. The solution integrates a Big Data Warehouse (BDW) [3], which supports the SC simulation model, by extracting raw data from selected data sources, transforming it into quality data and providing it to the simulation model. The research that was conducted to define the data requirements of the BDW was published in [4]. In its turn, a prototype of the simulation model, which was used to validate the set of variables selected for the project was also published in [5]. 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.

Notwithstanding, several data issues were faced when providing the real data, stored in the BDW, to the simulation model. Such issues were verified in the organization hosting this research, despite its technological conditions, such as the advanced Information Systems (IS) and Enterprise Resource Planning (ERP), as well as despite being a flagship in its industry sector with reference business processes.

In fact, the subject of facing data issues while working on simulation projects and using real industrial data is not new, as Bokrantz et al. [6] corroborated. The authors presented a multiple case study within the automotive industry to provide

2351-9789 © 2020 The Authors. Published by Elsevier B.V.

10.1016/j.promfg.2020.02.033

empirical descriptions of data quality problems in simulation projects. As the authors postulated, simulation requires high quality data and, often, extensible transformations to allow its utilization in simulation models, i.e., data issues must be bypassed, in order to produce a coherent simulation model.

In light of the above, the purpose of this paper is to address the data issues that were faced while developing a SC simulation model in a Big Data context, since Big Data concepts and technologies were used in this project. Such issues are identified and the corresponding approaches that were conducted to bypass them are presented. With this work, the authors believe that researchers focusing on similar problems and facing similar difficulties will find the shared approaches and conclusions of this research helpful.

This paper is structured as follows. Section 2 summarizes the related work. Section 3 provides a brief description of the SC system considered in this paper, as well as the main development stages that were conducted in this project. Section 4 details the main data issues that were faced, and which needed to be bypassed, in order to produce a coherent SC simulation model. In its turn, section 5 illustrates examples of results that can be obtained from such simulation model. Finally, the main conclusions and future work are discussed in section 0.

2. Related Work

The need to improve industrial processes is, in fact, one of the main goals of Industry 4.0 as emphasized by Kagermann et al. [7]. Such improvement may involve several methods, with the authors stressing the use of simulation to analyze the behavior of complex systems like SCs, including potential crisis scenarios. The authors also noted the importance of using Big Data in conjunction with such solutions, as it allows data from several data sources to be considered in the model.

Vieira et al. [8] reviewed simulation studies closely related with the concept of Industry 4.0, in order to identify the boiling research directions for simulation in this industrial revolutionary movement. According to the authors, such studies include the use of Big Data technologies applied to SC problems, due to the possibility of capturing the detail of processes that Big Data allows, along with the ability to consider alternative scenarios that simulation offers.

Zhong et al. [2] outlined the current movements on the application of Big Data for Supply Chain Management (SCM). According to the authors, the increasing volume of data in the several SC sectors is a challenge, which requires tools to make full use of the data, with Big Data emerging as a discipline capable of providing solutions for analysis, knowledge extraction, and advanced decision-making.

According to Tiwari et al. [1], analytics in SCs, including simulation, is not new. However, the advent of Big Data presents as an opportunity for its use with such analytics methods (e.g. simulation). The authors stressed the importance of such duo in predictive and prescriptive analytics, with simulation being used in the former to predict future events and in the latter to enhance the decision-making process.

As the cited works suggest, and to the best of the authors' knowledge, a gap can be identified in literature, which concerns the existence of Big Data structures to store and integrate data from several sources, with the end goal of providing such data to a SC simulation model. As such solution is currently being developed by the authors, this paper builds on the identified gap, by contributing with the approaches conducted to bypass the data issues found while developing such solution.

3. Materials and Methods

This section starts with a description of the SC in analysis, to convey the complexity associated to the problem. Thereafter, second subsection describes the development stages of the project, to provide the approaches used for the problem.

3.1. System Characterization

This project is being developed at a plant of the automotive industry sector, which produces electronic components. This subsection describes the SC at hand, to give a perspective of the scale and complexity of the network in analysis.

The plant considers roughly 7 000 different types of materials which are actively being supplied by around 500 different suppliers, which are located in more than 30 countries from around the world. After analyzing the obtained data, and during the time frame considered in such data, it was possible to observe that most suppliers were from Europe and Asia, with Germany and Netherlands having more suppliers and shipments from Europe, and Malasya, Taiwan, China, Hong Kong and Singapore having more shipments from Asia. All suppliers shipped more than 200 000 deliveries.

3.2. Methods

It is widely accepted that SCs generate huge amounts of data, leveraging the need for Big Data technologies [1], [2], [8], which were used in this project. It should also be noted that the Big Data cluster of the organization is also being used.

SC activities may differ between companies, geographic locations, businesses and industry sectors. Thus, this work started by studying the processes associated to the SC at hand. For this purpose, interviews and trainings with process specialists were helpful to gain insights from them. Internal documents of the organization were also analyzed at this stage.

To develop a BDW, despite being in a Big Data context, where data models are not usually a main concern in terms of providing an overall and integrated view of the data, it is still important to start by analyzing the data requirements of the BDW system, namely its elicitation. This was done by applying user-, goal- and data-driven approaches in conjunction, in order to have all the relevant perspectives. It should be noted that this was continuously done throughout the development phase in successive iterations, as described in detail in [4]. By doing this step, the following major benefits were achieved, as corroborated by previous studies [9], [10]: better understanding of the data, organizational processes and relevant variables to include in the BDW and in the simulation model; making sure that no important data is excluded; helped in the definition of the BDW model, namely the Hive tables to use.

After selecting the relevant variables, the data profiling is conducted, where its quality is assessed, to determine the necessary transformations. Such data profiling techniques allow to verify, for instance, the existence of null values, the distribution of values, and the quality of categorical values.

Thereafter, ETL (Extract, Transform, Load) jobs are developed, which extract data from data sources, compute eventual transformations that are identified after assessing the data quality, and send the transformed data to HDFS (Hadoop Distributed File System) of the Hadoop ecosystem. The next step consists in defining the schema of the BDW and load the necessary data, so that it can provide the required data to the simulation model. See [4] and [11] for more details regarding the Big Data concepts and tools that were used.

The last step of the project consists in using the data in the simulation model, which was developed in SIMIO [12]. However, despite the data profiling phase that was conducted, the authors experienced several data issues, which needed to be handled, to maintain the coherence of the simulation model. Such data issues are addressed in next section.

Around 3GB of data were considered, corresponding to a year of data and roughly 8 000 000 rows of data. This volume only considered the data integrated in the BDW and no other data that could not be included after analyzing its quality.

4. Data Issues: Simulation Model Coherence

Traditional data profiling techniques aim to verify the quality of data. However, this is done at a syntactic level, by evaluating aspects such as checking for null values or errors. In a simulation project, such data profiling techniques are required, albeit limited, as a simulation model needs not only quality data, but also coherent data, so that the result is an equally coherent model. Thus, in some cases, data needs to be estimated; on other occasions, data sources of certain relevant business processes simply do not exist but need to be incorporated in the simulation model in some way. Hence, in such situations, it is important to involve process experts, as well as query the available data, in order to identify reasonable approaches to bypass the identified data issues.

The purpose of this section is to present the data issues that were faced, mentioning the importance of handling them, while also emphasizing the approaches that were adopted to bypass them. The following provides such discussion.

4.1. Customers' orders

To assess the impact of customers' orders variability, the quantity of each finished good ordered by the end customers, as well as the respective delivered quantity, are crucial. This is an information that could not be provided by the Logistics Department hosting this research, as it is considered sensitive purchasing information, since the Department is solely responsible for providing the raw material needs to the Production. To bypass this lack of data, such material needs can be considered. However, this approach is limited in some ways. First, as field observations and the interviewed managers suggested, it is common for raw materials to be required by Production, only to remain stored in supermarkets, waiting hours or even days to be consumed. Second, the consequences of customers' order variability cannot be efficiently measured, as it is different to change the order quantity of a certain raw material than changing the order quantity of a finished good, which is comprised by critic and not-so-critic raw materials. Third, a shortage of a material that is used in a relevant number of finished goods should have a greater impact than the shortage of a material that is used in fewer finished goods. Such difference is hardly felt if the scope only considers the raw materials. However, the mentioned cons, such data is crucial since the end goal of a SC is to fulfill its customers' needs. Thus, to bypass this lack of data, the Production's orders were used as the demand that stimulates the simulated SC to operate.

4.2. Data of orders to suppliers

In the data profiling phase, it was not possible to obtain the date of roughly 27,8% of the orders to suppliers, and roughly 0,5% of the order dates had to be altered, since they had an arrival date prior to the order date, which would bring several problems for the simulation model. After discussing this issue with process experts, this was solved by subtracting a constant value to the arrival date, which corresponds to the estimated lead time of suppliers operating in similar circumstances.

4.3. Suppliers' locations

Some values of the city of suppliers cannot be used, due to data problems that were identified in the data profiling phase. This was observed even though the values are provided by SAP. Thus, whenever it was possible to use the city coordinates, these were used. However, the simulation requires a location for every supplier, even for those without a city in SAP. In these cases, the country's geographic location was used. Moreover, the geographic coordinates of cities and countries had to be generated, since these are not contained in any of the organization's Information Systems.

Finally, to visualize all orders sent to the same location, a small deviation to their coordinates was applied. This ensures that all entities, in their respective locations, can be visualized when viewing the model running. These location changes and the above discussed estimation do not affect the simulation results, as the respective distance and travel duration of the associated entities, between the source and destination locations, are considered when calculating the lead time.

4.4. Travel mode

The travel model for some orders could not be obtained, because the supplier does not exist in the data source. Thus, with the help of process experts, some rules were implemented. Nevertheless, these rules only set the symbol of the entity, not affecting the transportation durations, neither the results.

4.5. Transit and lead times

Some transit time durations preclude the arrival of orders at the date specified in the data. In these situations, both transit and lead time were estimated. Moreover, these problems were handled in conjunction, since transit time can be considered part of total lead time. In light of this, first, it is verified if there is a transit time specified for a given entity. If not, it is estimated based on the transit times of other suppliers from the same country. Afterwards, it is verified if the transit time allows the associated entities to arrive to the plant at the date in the data. If the durations are not adequate, the lead time and transit time values are adjusted to allow the entity to arrive to the plant at the arrival date specified in the data. This approach does not influence the results, as the total lead time remains the same.

4.6. Internal material movements

One of the benefits of including all material movements is to model the storage strategy followed at the plant and, hence, measure the stock level. In the considered plant, such strategy implies that materials are stored in an empty storage location. Hence, such strategy implies the following two premises:

- When a material is moved to a bin, the bin is empty before this movement occurs;
- When a material is moved out of a bin, the material needs to have been previously stored in the same bin.

The organization measures its stock level by assessing the percentage of occupied bins. However, while running the simulation model, several cases in which these premises were not verified could be observed. Thus, since simulation allowed to discover these problems in data, it would also be interesting to use it to understand the scale of these problems. Thus, the simulation model was used to record the percentage of movements that do not follow the storage strategy of the plant and the obtained results can be seen in Fig. 1. These percentages were registered after the first move to each bin.

As the figure shows, the percentage in the movements out of the warehouse maintains the same level throughout the year, with the exception of one day at the end of the year. Conversely, the number of failed movements to the warehouse, on average, is higher throughout the year.

After trying to understand this problem with process experts, two main justifications arose. The first is that not all movements are registered. The second is that movements are registered with a wrong date. For instance, a material may be consumed, but its consumption register is not immediately created (or is created with a wrong date), hence movements appear in the wrong order. This problem, in fact, demanded a change in the approach to model the warehouse. While the ideal approach would be to have a data structure comprised by a position for each bin of the warehouse, this was not possible due to this data issue. Thus, the solution was to measure the variation of the total quantity of material in the plant.

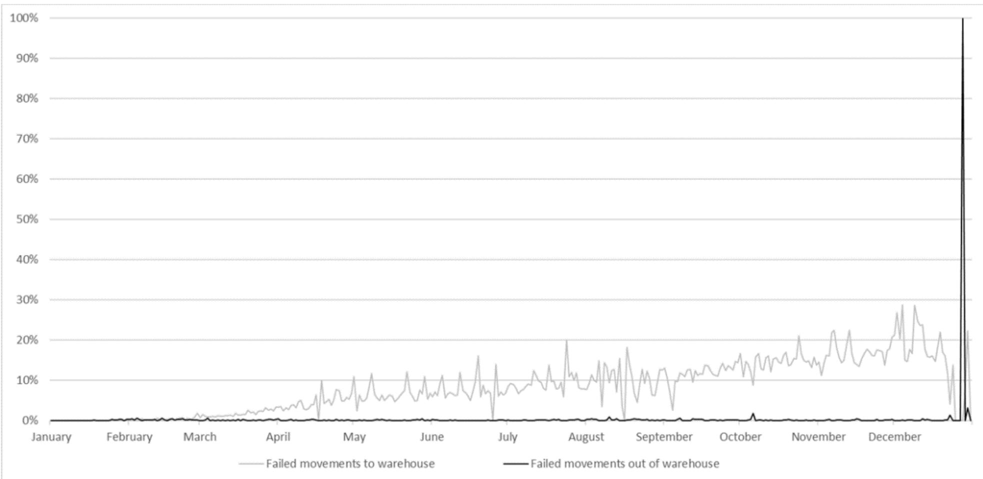

Fig. 1. Percentage of movements not consistent with the storage strategy followed at the plant.

4.7. Initial stock

The stock level at the beginning of the simulation should correspond to the one verified at the day corresponding to the start of the simulation. However, it was not possible to obtain historical stock data, as the ERP only displays the current level.

To bypass this lack of data, the simulation model was run without considering any stock method, and the average, standard deviation and other aggregation values were stored. In a new run, several expressions and approaches were considered, which use the previously calculated aggregation values, in order to obtain the quantity for each material at the beginning of the simulation. This way, the simulation is "learning" the stock level to use. The following expressions were used to calculate the initial stock:

$$Q_{consumed} * TB_{suppliers} \quad (1)$$

$$Q_{consumed} * LT \quad (2)$$

$$Q_{consumed} * SS_{time} \quad (3)$$

$$SF(SL) * \sqrt{LT * SD(Q_{consumed})^2 + Q_{consumed}^2 * SD(LT)^2} \quad (4)$$

with $Q_{consumed}$, average consumed quantity; $TB_{suppliers}$, average time between orders to suppliers; LT, average lead time; SS_{time}, safety stock in time; $SF(SL)$, safety factor for service level, which in this case was considered to be 99,9%, based on a normally distributed demand, thus obtaining the value 3,9; $SD(Q_{consumed})$, standard deviation for consumed quantity; $SD(LT)$, standard deviation for lead time.

Expression 4 was obtained from the literature and, as suggested by Ruiz-Torres and Mahmoodi [13], it is one of the most commonly used methods for the safety stock calculation. On this note, the use of this expression assumes that the safety stock of each material can be used as the stock to start the simulation. The same method was also analyzed by Schmidt et al. [14] in their review of safety stock calculation methods. This problem, in fact, remains as one of the hottest and more complex research topics in the field [13], [14]. Besides the above approaches, the following were also considered:

- A: Sum of all consumptions;
- B: Quantity difference between all consumptions and all arrivals;
- C: Sum of all consumptions until the first arrival of each material;
- No initial stock.

Fig. 2 shows the evolution of the stock for each implemented method. The graph shows the stock approaches corresponding to expressions 1 to 4 with dashed or dotted lines, and the remaining four approaches with continuous lines.

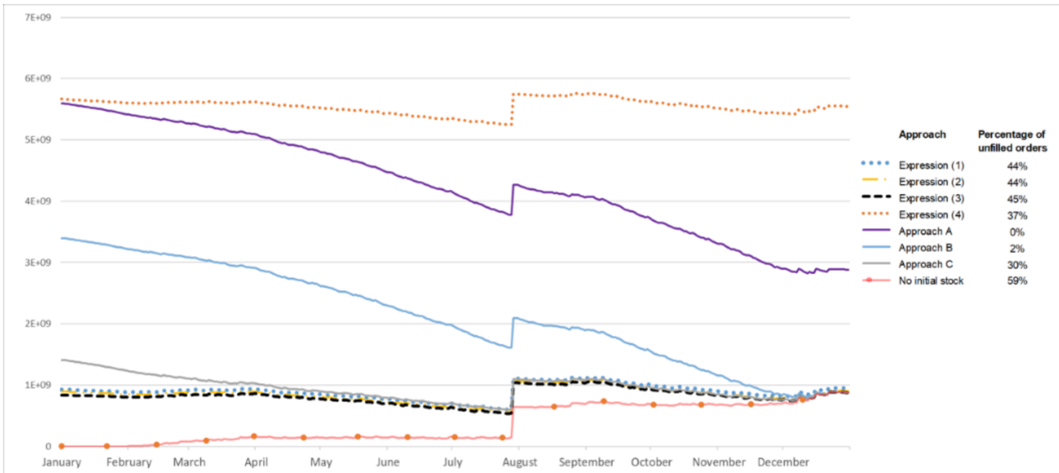

Fig. 2. Evolution of the stock level using different safety stock approaches.

Regarding this later set, it can be seen that approach A results in a high stock level, which is related with the nature of the approach, starting the simulation with all the quantity of materials that will be consumed throughout the year, already in stock. Conversely, approach B is the result of the difference between all consumptions and all arrivals. However, as the graph shows, the stock indeed decreased, albeit with the cost of some unfilled orders (2%), which can be justified by the arrival of some materials later than expected (volatile demand or lead time). Approach C shows that it is not enough to consider the quantity consumed until the first arrival, as the unfilled orders considerably increases, in comparison to the previous approaches. Lastly, the graph also includes a scenario without initial stock with 59% of unfilled orders.

In their turn, expressions 1 to 3 returned considerably lower initial stock, albeit with high unfilled orders percentage (respectively 44%, 44% and 45%). It is interesting to note that all approaches, except for expression 4 and approach A, tend to the same stock level, although all, except approach B, obtained the highest percentage of unfilled orders.

In sum, obtaining a method or an expression to calculate the optimum safety stock is a very complex task, as corroborated by Ruiz-Torres and Mahmoodi [13] and Schmidt et al. [14]. Thus, with all the pros and cons above discussed, it is certainly an arguable decision, however approach B and C and expression 4 can be emphasized. The former resulted in the second lowest unfilled orders percentage, albeit approach A cannot be selected for disruption scenarios, since it would never result in unfilled orders, as it starts with the exact stock required during the simulation. In its turn, expression 4 is one of the most adopted calculation methods in literature [13], [14] and resulted in less unfilled percentage than the remaining calculation methods. Hence, as the analysis suggests, there is no solution unarguably better than the others.

4.8. Production time, capacity and utilization

The simulation model must consider the production capacity of the plant, albeit it is hard to obtain such metric. Hence, simulation was used to estimate it. The plant's production is divided in two Departments, dedicated to different production phases: automatic insertion and final assembly. Thus, the number of capacity units of these production units was set to infinity and the results were plot in Fig. 3. Nevertheless, besides recording the number of units in the production, it was also necessary to establish a production time, which was also not available. Thus, with the help of process experts and some field observations to measure average production times, a generic normal distribution was applied to all materials. Note that the customers' orders were replaced by Production orders (as previously discussed in this section), which reduced the scope of the SC system. Thus, a considerable impact of this expression in the performance of the system is not expected.

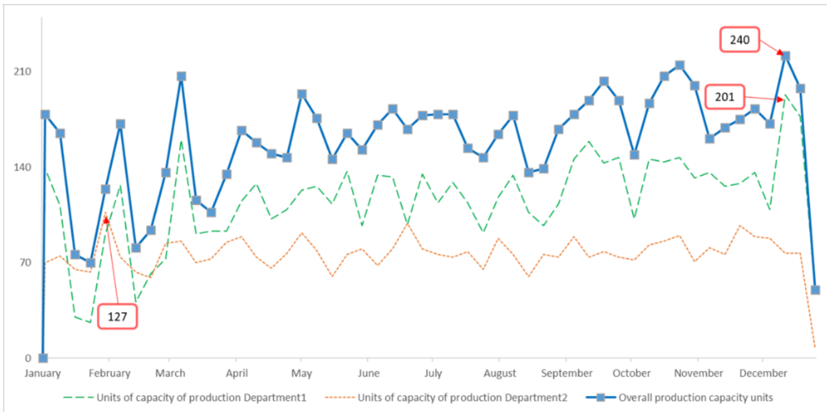

Fig. 3. Utilized capacity units of both production Departments per week.

As the figure shows, the maximum capacity of both production Departments can be determined. This is the required number of capacity units in order to fulfill all the orders registered in the data. The figure also shows the required capacity units for the overall production is 240.

The data issues discussed in this section, allowed to understand that, despite the Big Data that organizations already have, it is arguable if their data models are complete and consistent. In fact, this section showed that in the plant considered in this case study this is not the case. Hence, the solution was to apply the approaches described in this section to bypass such issues, which was done in an iterative way. Next section shows the types of results that can be obtained from the simulation model, after bypassing the dissed data issues.

5. Results

In this section, the main results that can be retrieved form the developed and coherent SC simulation model are addressed. In this regard, Fig. 4 shows a picture of the model during a simulation run.

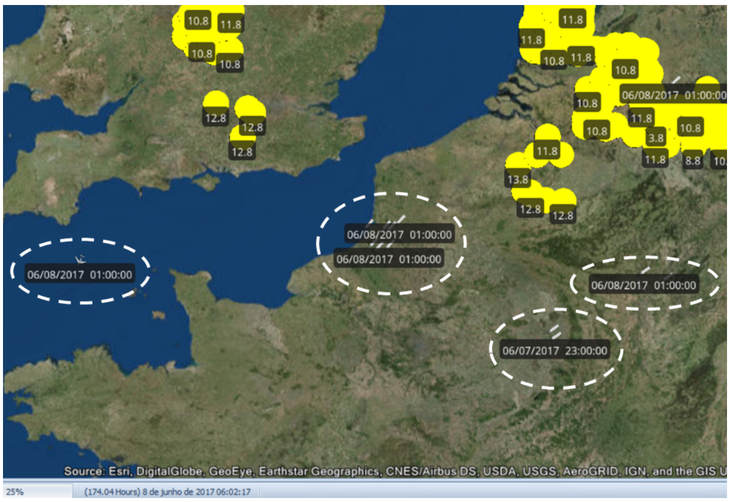

Fig. 4. Orders being sent to the plant.

The model runs in a 3D world map view. The figure also shows some circles placed at north of Europe. The location of these entities represents the location of the supplier. The number presented under each yellow entity is 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, with the figure showing some of these entities highlighted. The date time values associated to each entity represent the instant when those deliveries were shipped to the plant. Apart from graphical results, it is also possible to retrieve analytical results from the tool, with Fig. 5 showing the total quantity of materials ordered, consumed and arrived to the plant during the years of data stored in the BDW.

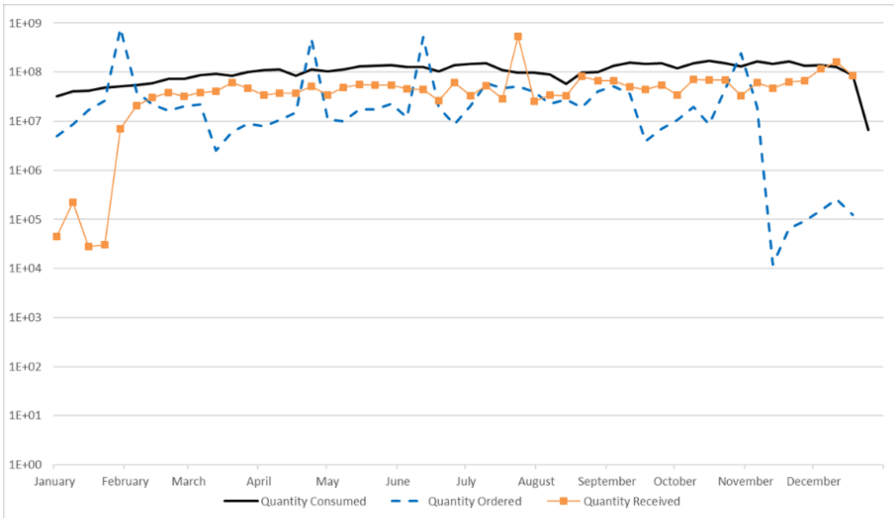

Fig. 5. Total quantity of materials ordered, received and consumed per week.

The adopted approach allowed to achieve a simulation model that is coherent and consistent with the system being modelled, in the sense that the main elements stored in the BDW are reflected in the simulations. Hence, managers from the plant can use such tool to aid them in the analysis of uncertain and alternative scenarios. Nevertheless, the achieved results also show that simulation can be used as a data validation technique, further extending traditional data profiling ones. In fact, simulation allowed data issues to be identified, by evaluating the semantics of data, and also allowed certain missing data to be estimated.

6. Conclusions

SC systems generate huge amounts of data, due to the several data sources that are used to manage the associated business processes. Furthermore, SCs are complex systems, being useful to use both Big Data and Simulation to model SC problems. With these, it would be possible to test uncertainty scenarios using simulation, as well as to consider the detail provided by Big Data. In this paper, an industrial project using real data from an automotive electronics SC was presented, which is associated to a plant of the automotive electronics industry sector. In such highly dynamic environments, it is common for data issues to be verified. Thus, this paper aimed to present the most relevant data issues that were faced while developing the SC simulation model in a Big Data context, while also discussing their impact on the solution and the measures that were taken to bypass them.

Indeed, some data issues can be handled by traditional data profiling techniques. However, such techniques only allow the assessment of the data quality to a syntactic level (e.g., null value verification), which is not enough for simulation, where this verification needs to be taken to a different level of exigency. In fact, in simulation there is an obligation to integrate data in such way that it must originate a coherent simulation model (in order to accurately mimic a process, all its elements must be present and coherent). In this work, the authors argue that simulation can be used as a semantic validator of the data model, advancing traditional data profiling techniques, in the sense that it allowed additional data issues to be identified and missing data to be estimated.

The identified issues and the respective approaches that were implemented to bypass them, allowed to better understand both the data sources and the associated business processes, hence helping in the development of the simulation model. In fact, the obtained results (both graphical and numerical) were the result of bypassing the identified issues, while still maintaining the overall coherence of the model.

Despite the huge amounts of available data (around 3 GB of data), this work showed that the data model of organizations is still incomplete, in the sense that it still does not allow complete mimics of their SC systems to be reproduced. This suggests that, despite using many software packages, spreadsheets, IS and others, organizations are still lacking data that is relevant, in order to allow the creation of accurate simulations of their SCs. Such issues included data sources which could not be obtained and data that did not reflect a given business strategy followed at the plant, indicating that the data was incomplete, or not registered in the correct order or with the correct date. Some of these issues may be related with the top management view that often disregards the existence of low-level data (e.g., material movements), which is necessary in order to produce a coherent simulation model. Notwithstanding, this barrier should be bypassed when the Industry 4.0 revolution is completely materialized, which will allow some of this data to be automatically generated, stored and integrated – without eventual errors related with manual interactions - to allow analytical methods (e.g., simulation) to be employed.

In terms of future work, the following directions are highlighted. In what concerns the issue of missing historical data, the BDW can be used to maintain it, however, these will only be accessible in the mid- to long-term. The remaining missing data sources have to be covered with solutions aligned with the organization. Furthermore, despite the identified data issues, this paper also showed that it is possible to retrieve results from a coherent simulation model, hence allowing several types of SC risks to be analyzed. Thus, future work should also concern in performing such risks analysis.

Acknowledgements

This work has been supported by national funds through 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] Tiwari S, Wee HM, Daryanto Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*; 2018. 115, (Jan. 2018), 319–330.

[2] Zhong RY, Newman ST, Huang GQ, Lan S. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers and Industrial Engineering*; 2016. 101, 572–591.

[3] Costa E, Costa C, Santos M. Evaluating partitioning and bucketing strategies for Hive-based Big Data Warehousing systems. *Journal of Big Data*; 2019. 6, 1 (Dec. 2019), 34.

[4] Vieira AC, Pedro L, Santos MY, Fernandes JM, Dias LS. Data Requirements Elicitation in Big Data Warehousing. *European, Mediterranean, and Middle Eastern Conference on Information Systems, EMCIS, Lecture Notes in Business Information Processing*; 2019. 106–113.

[5] Vieira AC, Dias LS, Santos MY, Pereira GB, Oliveira JA. Simulation of an Automotive Supply Chain in Simio: Data Model Validation. *30th European Modeling and Simulation Symposium, EMSS*; 2018. 294–301.

[6] Bokrantz J, Skoogh A, Lämkull D, Hanna A, Perera T. Data quality problems in discrete event simulation of manufacturing operations. *Simulation*; 2018. 94, 11 (Nov. 2018), 1009–1025.

[7] Kagermann H, Helbig J, Hellinger A, Wahlster. Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry ; Final Report of the Industrie 4.0 Working Group. Forschungsunion. 2013.

[8] Vieira AC, Dias LS, Santos MY, Pereira GB, Oliveira JA. Setting an industry 4.0 research and development agenda for simulation – A literature review. *International Journal of Simulation Modelling*; 2018. 17, 3, 377–390.

[9] Costa E, Costa C, Santos M. Efficient big data modelling and organization for hadoop hive-based data warehouses. *European, Mediterranean, and Middle Eastern Conference on Information Systems, EMCIS, Lecture Notes in Business Information Processing*; 2017. 3–16.

[10] Santos MY, Costa C. Data Models in NoSQL Databases for Big Data Contexts. *International Conference of Data Mining and Big Data, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 2016. 475–485.

[11] Vieira AC, Dias LS, Santos MY, Pereira GB, Oliveira JA. Supply chain hybrid simulation: From Big Data to distributions and approaches comparison. *Simulation Modelling Practice and Theory*; 2019. 97, (Dec. 2019), 101956.

[12] Vieira AC, Dias LS, Pereira GB, Oliveira J, Carvalho MC, Martins P. Automatic simulation models generation of warehouses with milk runs and pickers. *28th European Modeling and Simulation Symposium*; 2016. 231–241.

[13] Ruiz-Torres AJ, Mahmoodi F. Safety stock determination based on parametric lead time and demand information. *International Journal of Production Research*; 2010. 48, 10, 2841–2857.

[14] Schmidt M, Hartmann W, Nyhuis P. Simulation based comparison of safety-stock calculation methods. *CIRP Annals - Manufacturing Technology*; 2012. 61, 1, 403–406.