

Distribution network design with big data: model and analysis

Gang Wang¹ · Angappa Gunasekaran¹ ·
Eric W. T. Ngai²

Published online: 30 June 2016
© Springer Science+Business Media New York 2016

Abstract This study addresses the problem of locating distribution centers in a single-echelon, capacitated distribution network. Such network consists of several potential distribution centers and various demand points dispersed in different regional markets. The distribution operations of this network generate massive amounts of data. The problem is how to utilize big data generated to identify the right number of distribution centers to open and the right assignment of customers to opened distribution centers while minimizing the total handling and operation costs of distribution centers, transportation, and penalty. Restrictions on both network capacity and single sourcing strategy are also considered. This study formulates this problem as mixed-integer nonlinear program. The effects of different scenarios on distribution-center locations as demand, the operation costs of distribution centers and outbound transportation, and the number of customers are analyzed through simulation on randomly generated big datasets. Empirical results indicate that the model presented is appropriate and robust. The operational value of big data in the distribution network design is revealed through a case study in which several design alternatives are evaluated.

Keywords Big data · Capacitated network design · Service levels · Mixed-integer nonlinear programming · Simulation

1 Introduction

As globalization advances, distribution networks of many manufacturers have evolved into large networks of plants, warehouses and/or distribution centers (DCs), and demand points worldwide. Various products flow through supply chains at a surprisingly fast rate. As such,

✉ Gang Wang
wang.gaby@gmail.com

¹ Department of Decision and Information Sciences, Charlton College of Business, University of Massachusetts Dartmouth, 285 Old Westport Rd, North Dartmouth, MA 02747, USA

² Department of Management and Marketing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

these fixed networks cannot adapt to the constantly changing supply chain operations. Moreover, complex supply chains are generating huge amounts of dynamic data in different formats (i.e., structured and unstructured), which create new challenges for supply chain design and redesign. Traditional network design models and methods are also inappropriate to modern supply chain network design because they no longer deliver business value required to stay competitive in the era of big data. Hence, the need for powerful supply chain network design models is increasing. These models should be able to harness big data to evaluate more variables and scenarios than previous ones. This study aims to present such distribution network design model for achieving a robust supply chain in the context of big data.

The application of big data in supply chain management has received increased attention in the literature. Most studies focus on the conceptual framework of applying big data to supply chain management (Dubey et al. 2016; Rezaee et al. 2015; Wang et al. 2016). Dubey et al. (2016) demonstrated the importance of big data analytics in supporting world-class sustainable manufacturing. They also developed and empirically validated a conceptual framework that incorporates the building block of sustainable manufacturing, which can be extracted from big data. Therefore, the study has identified the important factors that practitioners need to focus on when planning to use big data analytics to improve the economic, social, and environmental performance of their companies. Oliveira et al. (2012) analyzed the effect of the use of business analytics on supply chain performance and discussed the changing information processing needs at different supply chain maturity levels. Through a worldwide sample of 788 companies from different industries, Oliveira et al. (2012) found that the changing effect of business analytics use on performance implied that companies at different maturity levels should focus on different areas. Chae and Olson (2013) proposed a framework of business analytics for supply chain analytics that is IT-enabled and feature analytical dynamic capabilities, namely, data management, analytical supply chain process, and supply chain performance management. Chae and Olson (2013) also presented a dynamic-capability view of SCA and extensively described its three capabilities (i.e., data management, analytical supply chain process, and supply chain performance management) in an integrated manner. Sanders (2014) presented a systematic framework regarding the use of big data analytics in supply chain management and the means to determine business value from big data. Hazen et al. (2014) reviewed research on data quality in the supply chain, defined measures of data quality, introduced a means to control data quality, and suggested theory-based topics for future research. Tan et al. (2015) studied the application of big data in strengthening supply chains and business management in specific areas, such as procurement, supply chains collaboration, end-to-end supply chains execution, and inventory control. Tan et al. (2015) suggested that big data enable companies to gain new product development ideas and understand how different sub-firms or departments could work together for optimizing the manufacturing processes and producing new products in the most cost-effective means. Chae et al. (2014) studied the effect of business analytics resources including accurate manufacturing data and advanced analytics on the operational performance of a firm. Chae et al. (2014) suggested the moderating and mediating role of fact-based SCM initiatives as complementary resources, tested propositions using Global Manufacturing Research Group survey data, and analyzed them using partial least squares/structured equation modeling. Wang and Alexander (2015) introduced the big data concept, its characteristics, and major issues in supply chain management. They then concluded that big data can improve effectiveness and efficiency, produce high-quality outputs, and increase the value-added content of products and services. Chae (2015) proposed a novel, analytical framework (Twitter analytics) for analyzing supply chain tweets, highlighting the current use of Twitter in supply chain contexts, and further developing insights into the potential role of Twitter for supply chain practice

and research. The proposed framework combines three methodologies, descriptive analytics, content analytics integrating text mining and sentiment analysis, and network analytics relying on network visualization and metrics (Chae 2015). Rezaee et al. (2015) proposed an analytical framework for determining the optimal carbon trading credits or allowances under different conditions. The analytical framework could help companies determine the optional material flows and the quantities of carbon credits traded under an environment with high uncertainty (e.g., uncertain carbon price and product demand).

By contrast, supply chain analytics has been applied extensively in supply chain management (Sangwan and Mittal 2015; Souza 2014), especially on supply chain network design (Kılıç and Tuzkaya 2016; Liao et al. 2011; Omar 2013; Shu et al. 2010; Song et al. 2016; Tsao et al. 2016; Wang and Lei 2012, 2015; Wang et al. 2014; Zhang et al. 2013). For example, Sangwan and Mittal (2015) found that supply chain analytics plays an important role in green manufacturing. However, a limited number of studies have focused on the real application of big data to supply chain planning and the development of specific models and/or algorithms in the context of big data, particularly for supply chain network design.

De Matta (2016) proposed an analytic model, that is, a network design optimization model for handling supply disruptions. The analytic model focused on the use of reserve capacity to migrate the production disruptions without affecting the stability in both the prices of products and the profits of the focal business. However, although the models and algorithms were presented, the model was not tested in real-life context but only verified using the achieved data of a company. Gölzera et al. (2015) extended existing methods for decision-making implementation in the design and operation of global manufacturing network. They described big data techniques, highlighted the aspects of decisions tasks, data access patterns, necessary data structures, and handling design scenarios. Finally, they concluded that big data systems can be applied on tactical design decisions in manufacturing networks on the basis of the production master data. Unfortunately, they did not present models, algorithms, and case studies to explore the operational value of big data in designing global manufacturing network.

Given the existing literature, this study attempts to explore the application of big data to solve real supply chain problems by developing powerful models and discussing cases. We consider the location problem of DCs in a single-echelon, capacitated logistics network. The supply chain is composed of DCs to open and geographically dispersed regional markets. DCs with limited capacity have variable and fixed costs for handling one unit of product, each with a desired service level. If DCs cannot achieve desired service levels, penalty costs proportional to unassigned customer quantities are imposed. Furthermore, each customer is assigned to exactly one DC. The problem is how to determine the right number of DCs to open and the optimal assignment of regional markets to DCs, such that the total shipping, handling, and operating costs of DCs and the penalty for unassigned customers are minimized. This study formulates the location problem as a mixed-integer nonlinear program and analyzes different scenarios to verify the proposed location model. Specifically, simulation is performed on randomly generated big datasets in terms of the change in demand, outbound transportation costs, DCs' operations costs, and the number of customers. The results show that the presented model is appropriate and robust. Then, a case study involving big data is provided, where the proposed model evaluates different design alternatives and attempts to determine the best design option. The empirical results reveal the operational value of big data when making location decisions on distribution network design.

The remaining parts of the study are organized as follows. Section 2 presents a big data-driven distribution network design model. Section 3 deals with model validation and analysis of randomly generated test instances in terms of the change in demand patterns, transportation

costs, DCs' operations costs, and the number of customers. Section 4 provides a case study where the proposed model is used to design the distribution network by examining different design options. Finally, Sect. 5 concludes the study and discusses its future extensions.

2 Network design model with big data

The emergence of big data allows us to consider complex factors while designing a supply chain, thereby making the development of a robust supply chain network possible. When faced with huge amounts of data, existing network design models may be intractable. Therefore, the need for powerful network design models based on big data is growing. This section presents an inclusive distribution network design model using big data while considering nonlinear penalty imposed to DCs because of unassigned customers.

We consider a retailer's logistics network with a group of geographically dispersed DCs and many demand points (e.g., over 2,000 stores). Each DC has limited capacity to serve demand points and a set desired service level. Failure to achieve the desired service level results in penalty costs that are proportional to unassigned customer quantities. Any shipment from DCs to demand points incurs variable and fixed costs. Customers adopt single sourcing strategies to take advantage of low costs and guarantee product availability. Given the big data with regard to customer demand, warehouse operations, and transportation, the problem is how to determine the right number of DCs to open and the optimal assignment of regional markets to DCs, such that the total shipping, handling, and operating costs of DCs and penalty for unassigned customers are minimized.

2.1 Data collection

In the big data era, decision makers acquire huge amounts of data from many heterogeneous sources and input the most up-to-date parameters from operational databases directly to those models. Big data involved in a typical network design problem include historical data recorded in databases and updated behavioral data collected from social media (e.g., LinkedIn, Facebook, Twitter, and Google+), web clicks, comments or reviews, or complaints. Our model assumes that behavioral data have been analyzed using marketing intelligence tools, thereby obtaining accurate demand information and identifying a set of candidate DCs. Table 1 lists the type of data used in designing distribution network (i.e., historical and behavioral data; Waller and Fawcett 2013).

Table 1 Historical and behavioral data used in supply chain network design models

Type of data	Considerations
Historical data	The number of DCs and customers
	Transportation costs, including fixed and variable costs
	Labor, handling, fixed operating, and order processing costs of DCs
	The capacity of DCs
Behavioral data	The forecasted demand by customer location
	data on decision and purchasing behavior, including items browsed and purchased, frequency, dollar value, and timing
	Profiling data on buyer identification and emotion detection; eye-tracking data; customer sentiment data on purchased products or locations

2.2 Mathematical model

For convenience of further analysis, the notations used later are summarized as follows:

Sets and indices

\mathcal{J} = set of customers, i.e., $j \in \mathcal{J} = \{1, 2, \dots, J\}$;
 \mathcal{M} = set of processing centers, i.e., $m \in \mathcal{M} = \{1, 2, \dots, M\}$.

Parameters:

C_m = capacity of PC_m ;
 d_j = order quantity of customer j ;
 b_{mj} = shipping rate (\$) from PC_m to customer j ;
 π_{mj} = fixed shipping cost (\$) from PC_m to customer j ;
 h_m = unit handling cost at PC_m ;
 f_m = fixed operating cost at PC_m ;
 α_m = service level that the PC_m desires;
 p_m = unit penalty cost (\$) imposed to PC_m for unassigned customers.

Decision variables:

$$y_{mj} = \begin{cases} 1, & \text{if customer } j \text{ is assigned;} \\ 0, & \text{otherwise.} \end{cases}$$

$$x_m = \begin{cases} 1, & \text{if } PC_m \text{ is to be opened;} \\ 0, & \text{otherwise.} \end{cases}$$

Let $a_{mj} = (b_{mj} + h_m) d_j + \pi_{mj}$, and distribution network design problem can be formulated as a mixed integer nonlinear program as follows:

$$\begin{aligned} \text{Minimize} \quad & \sum_{m \in \mathcal{M}} f_m x_m + \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}} a_{mj} y_{mj} \\ & + \sum_{m \in \mathcal{M}} p_m \left(\max \left\{ 0, \alpha_m \sum_{j \in \mathcal{J}} d_j - \sum_{j \in \mathcal{J}} d_j y_{mj} \right\} \right) \end{aligned}$$

Subject to

(1) PCs' capacity constraints:

$$\sum_{j \in \mathcal{J}} d_j y_{mj} \leq x_m C_m, \quad m \in \mathcal{M},$$

(2) Single sourcing constraints:

$$\sum_{m \in \mathcal{M}} y_{mj} \leq 1, \quad j \in \mathcal{J},$$

(3) Integrality and non-negativity:

$$x_m, y_{mj} \in \{0, 1\}, \quad m \in \mathcal{M}, j \in \mathcal{J}.$$

The first term in the objective function represents the total fixed operations costs of DCs, whereas the second term represents the total outbound transportation and handling costs of DCs. The third term refers to the total penalty costs imposed on DCs because of unfulfilled customer orders. Constraint (1) ensures that if DC_m is chosen to open, the total order quantities that DC_m satisfy cannot exceed the processing capacity of DC_m . Constraint (2) states that customer j is assigned to only one DC (i.e., single sourcing). Constraint (3) enforces integrality and non-negativity of relevant decision variables.

Table 2 Values of the parameters used in test problems

Parameters	C_m	h_m	f_m	α_m	d_j	p_j	b_{mj}	π_{mj}
Range	$U [50, 100]$	$U [0.2, 0.5]$	$U [1, 5]$	$U [0.5, 1]$	$U [5, 10]$	$U [10, 15]$	$U [0.5, 0.8]$	$U [1, 4]$

3 Model analysis: simulation

This section carries out sensitivity analysis of the number of DCs to open and the overall service level on parameters by producing randomly generated big datasets. The primary factors considered in sensitivity analysis include variations on demand, outbound transportation costs, DCs' operations costs, and the number of customers. The computational results validate our model in accordance with supply chain practices.

This study considers not only the location of DCs but also the overall service level, which is a nonlinear function. We are interested in big data because they facilitate the what-if analysis with different, complex scenarios. Table 2 provides all parameters used in test problems, which are generated randomly using uniform distributions. We solve all problems associated with our model by CPLEX 12.6 on a Lenovo X1 Carbon dual-core 2.7GHz laptop with 1 GB RAM. The following figures report the results of sensitivity analysis on the proposed model in terms of the aforementioned factors. Orange lines represent the effect of parameter variations on the overall service level that all processing centers achieve, whereas blue lines represent the effect of parameter variations on the number of DCs to be opened.

3.1 Effect of demand variability

Demand variability has been identified as the largest challenge for supply chain management (How to Meet the Challenge of Demand Variability in Your Supply Chain, by Narayan Laksham accessed from <http://info.ultriva.com/ultriva-blog/how-to-meet-the-challenge-of-demand-variability-in-your-supply-chain> on 3/16/16 1:19 PM) because it has a significant effect on firm profitability. Hence, demand variability must be considered when designing distribution networks. Demand variability also determines the type of distribution network, either centralized or decentralized system. Figure 1 illustrates the trends of the number of DCs and the overall service level as demand variation increases from 0 to 0.8. The number of opened DCs decreases from 8 to 6 as demand variations rise up to 0.8. When the variation on demand is large, a small number of opened DCs indicate a centralized supply chain system, which can achieve risk pooling, thereby reducing the demand variability and bullwhip effect. Conversely, the overall service level increases as demand variations grow, which also results from risk pooling. Forecasting accuracy on demand is enhanced significantly because a centralized distribution system reduces demand variations. The trends on both the number of DCs and the overall service level verify the validity of the proposed model because it can capture demand variability by achieving a centralized distribution system.

3.2 Effect of outbound transportation costs

Outbound transportation costs play a vital role in distribution network design because they directly affect the location and number of DCs. Figure 2 depicts how variations of outbound transportation costs affect the number of DCs to be opened and the overall service level. When the variations of outbound transportation costs increase from 0 to 0.83, the number of opened DCs increases and the overall service level rises accordingly, which is consistent with

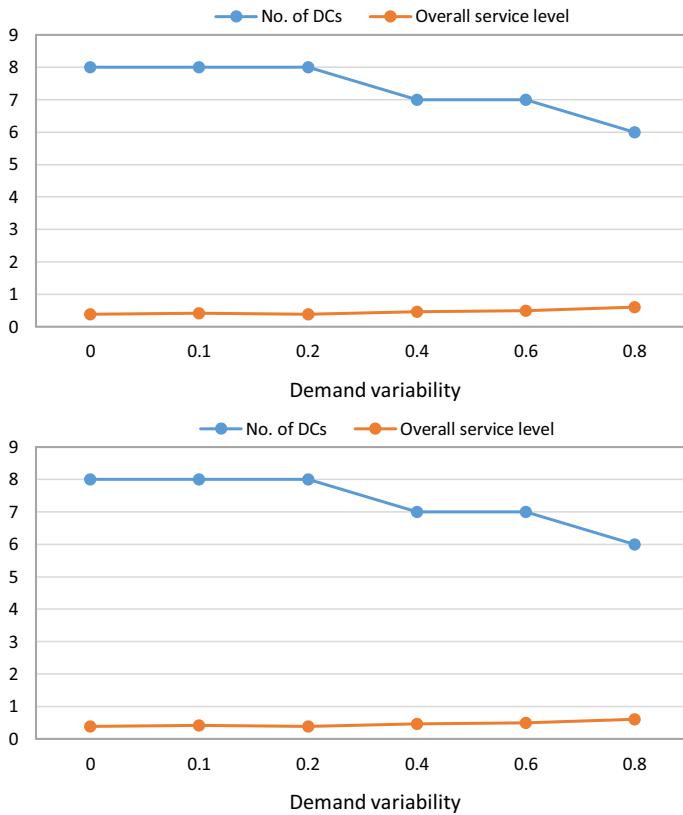


Fig. 1 Effect of demand variability on the number of opened DCs and overall service level

a decentralized distribution system. High variability in outbound transportation costs lead to increased overall transportation costs. By establishing a decentralized distribution system with the increased number of opened DCs, the overall transportation costs can decrease significantly because of the short distance for deliveries. Furthermore, delivery frequency grows because DCs are close to markets; the reduced overall transportation costs enable the transfer of added value to customers. Both scenarios lead to improved service level.

3.3 Effect of DCs’ operations costs

Given the high throughput at DCs on daily basis, the effect of operations costs on the number of DCs to be opened and overall service level should be considered. In our model, operations costs include fixed and handling costs. Potential DCs will be eliminated from consideration if their total operation cost is extremely high, which can lead to the highest total distribution costs. Figure 3 demonstrates the trends in the change in the number of opened DCs and overall service level as the variation of DCs’ operations costs increases from 0 to 0.83. Both the number of opened DCs and overall service level decrease. Highly variable operations costs result in the inflation of average total operations costs, thereby increasing the total distribution costs. Hence, as variability on DCs’ operations costs increases, the number of potential DCs and the overall service level declines accordingly.

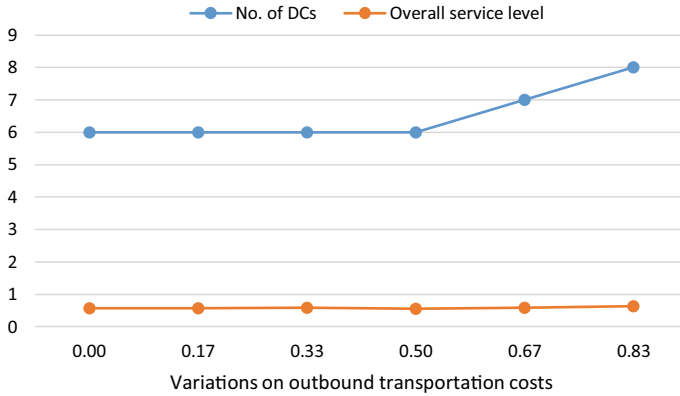


Fig. 2 Effect of transportation-cost variability on the number of opened DCs and overall service level

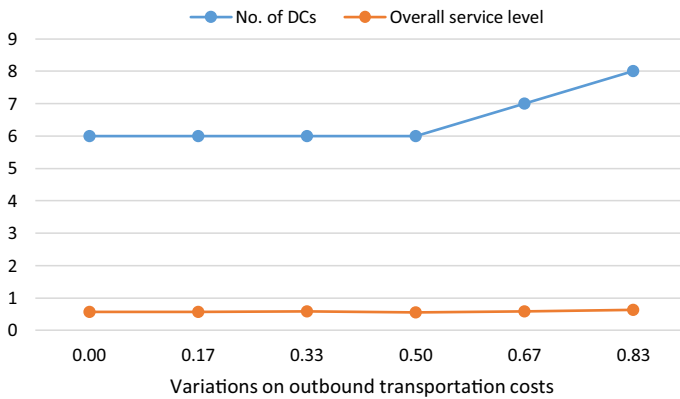


Fig. 3 Effect of operations costs on the number of DCs and overall service level

3.4 Effect of the number of customers

The ultimate goal of distribution network design is to satisfy customer demand responsively with the lowest distribution costs. The number of customers that need to be satisfied significantly affects the distribution network design. Figure 4 depicts the effect of the number of customers on the number of DCs to be opened and overall service level. Before the number of customers increases to 1,200, the number of opened DCs increases quickly, but the overall service level decreases gradually. The increase in the number of customers dominates the growth of the number of opened DCs. Although the number of opened DCs increases, the fast increasing number of customers creates high variability in demand, and the opened DCs cannot satisfy the increasing customer demand and compensate for high variations that a larger number of customers may bring. Hence, the overall service level decreases over time. However, when the number of customers is greater than 1,200, the number of opened DCs is nearly constant around seven, which signifies that satisfying many customers is beyond the overall capacity of current potential DCs. The overall service level continues to decrease given the large number of unassigned customers.

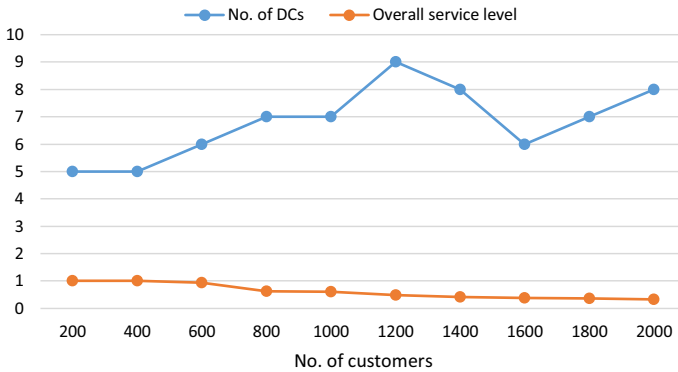


Fig. 4 Effect of customer size on the number of DCs and overall service level

4 Case study: location problem in retail distribution network design

This section provides a case study on the distribution center location problem over a retailer's distribution network. Huge amounts of data collected from its current network configuration are employed to evaluate many scenarios to reconstruct the extant network configuration.

We consider the distribution network of a retailer who sells sports products and leases a small DC in St. Louis to manage a large amount of products being sold. A third-party logistics company packs and ships customer orders to over 20,000 stores located all over the United States. The retailer's management team anticipates a demand growth of approximately 70% in the next 5 years. The team is concerned that operations costs will grow faster than revenues if the supply chain network is not redesigned because demand continues to increase. They analyzed the current network's performance and decided to redesign current network to best cope with the rapid anticipated growth over the next five years. The management team decided to lease more DCs to expand the capacity of the current logistics network because the expansion at the current small DC in St. Louis could not satisfy the demand growth. The realistic alternative that the management team proposes is to add DCs all over the country to complement the existing distribution system. Leasing a DC involved fixed and variable costs, depending on the size of the DC and the quantity shipped through the DC, respectively. Four potential locations are identified in Denver, Seattle, Atlanta, and Philadelphia. The retailer charged a flat fee of \$3 per shipment sent to a customer. The retailer also contracted UPS to handle all its outbound shipments. UPS charges were based on both the origin and destination of the shipment. The fixed and variable costs (\$/unit), capacity (units/year), service level, and penalty costs (\$/unit) of small and large DCs in different locations are provided in Table 3.

Location decisions in distribution network have a significant effect on the competitive advantage of organizations. Locations near suppliers can result in low supply costs, whereas locations close to markets can provide convenience, leading to low transportation costs and quick delivery times. Therefore, a tradeoff has to be made in distribution center locations in terms of operations costs and convenience. To identify the best network configuration, three scenarios are investigated in terms of the aforementioned report from the management team. The first scenario is that only small DCs are considered. The second is to consider only large DCs. The last involves the combination of small and large DCs. The optimal solutions, minimum total costs, and overall service levels for three different scenarios are summarized in Table 4.

Table 3 Fixed and variable costs, capacity, service level, and penalty of potential DCs

Location	Small DC					Large DC				
	Fixed cost (\$/year)	Variable cost (\$/Unit)	Capacity (units)	Service level (%)	Penalty (\$/Unit)	Fixed cost (\$/year)	Variable cost (\$/Unit)	Capacity (units)	Service level (%)	Penalty (\$/unit)
Seattle	300,000	0.2	6,660,225	90	8	500,000	0.5	13,320,450	72	15
Denver	250,000	0.1	9,341,155	88	5	420,000	0.2	18,682,309	68	17
St. Louis	220,000	0.4	8,427,703	90	7	375,000	0.3	16,855,406	70	12
Atlanta	220,000	0.2	3,733,691	95	6	375,000	0.4	7,467,381	85	15
Philadelphia	240,000	0.3	10,040,226	92	8	400,000	0.2	20,080,452	90	18

Table 4 Optimal solutions and costs as well as overall service levels for three scenarios

Location/ scenarios	Seattle	Denver	St. Louis	Atlanta	PA	Total cost	Overall service level (%)
Small DCs	1	1	1	0	1	\$2,045,752,274	31
Large DCs	1	1	1	1	1	\$3,317,782,974	100
Combination							
Small DCs	1	0	0	0	0	\$5,444,216,137	100
Large DCs	1	1	1	1	1		

If only small DCs all over the county are considered, the retailer keeps the small DC in St. Louis open and needs to lease three other small DCs in Seattle, Denver, and Pennsylvania. In this case, although the retailer can significantly reduce the total operations costs to \$2,045,752,274, which is relatively low, the overall service level it can achieve is only 31%. When retailers prefer to improve operations efficiency rather than customer satisfaction, leasing only small DCs is a good choice. However, in supply chain practices, organizations always seek the best tradeoff between cost reduction and customer service. A plausible alternative is to lease a combination of small and large DCs. Surprisingly, we obtain a counter-intuition result, which prioritizes leasing large DCs rather than the combination of small and large DCs (see Table 4). The low utilization of five large DCs in addition to one small DC to satisfy the remaining demand of unassigned customers creates a decentralized distribution system and increases total operations costs. Hence, leasing five large DCs in five potential locations is the best alternative because it incurs the least operations costs of \$3,317,782,974 and achieves 100% overall service level.

Additionally, to see operational value of Big Data, we apply the same datasets in this case to solve the well-known facility location problem as shown below:

$$\begin{aligned}
 &\min \sum_{m \in \mathcal{M}} f_m x_m + \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}} a_{mj} y_{mj} \\
 &\text{s.t.} \\
 &\quad \sum_{j \in \mathcal{J}} d_j y_{mj} \leq x_m C_m, \quad m \in \mathcal{M}, \\
 &\quad \sum_{m \in \mathcal{M}} y_{mj} = 1, \quad j \in \mathcal{J}, \\
 &\quad x_m, y_{mj} \in \{0, 1\}, \quad m \in \mathcal{M}, \quad j \in \mathcal{J}.
 \end{aligned}$$

In addition, to determine the operational value of big data, we apply the same datasets in this case to solve the well-known facility location problem. Unfortunately, we do not obtain an integer solution because the large number of stores involved exceeds the total capacity of all DCs. The aforementioned facility location problem considers a single objective, that of operations efficiency under the unrealistic assumption that customer satisfaction is fully achieved.

In comparing the optimal solutions between our model and the traditional location problem, additional information on service levels and penalty costs allows us to determine the tradeoff between operations efficiency and customer satisfaction, thereby leading to an enhanced solution to distribution network design problem. We can conclude that big data challenge the existing distribution network design models, but they create opportunities for

designing complex distribution network by providing huge amounts of various data for realistic scenarios (e.g., data for service levels and penalty costs).

5 Conclusions and future extensions

In the context of big data, our research explored the location problem of DCs in the distribution network design and examined the effects of parameter variations on network configuration and operations performance. Our model differs from existing models related to distribution network design because a nonlinear penalty function associated with service levels is introduced. Furthermore, simulation was performed to validate the proposed model and assess its sensitivity in terms of the variations on demand, outbound transportation costs, DCs' operations costs, and the number of customers. Our simulated results contributed to the understanding of the implications of parameter variability on distribution network design. From our analyses, variability on demand, outbound transportation costs, and the number of customers strongly affect the distribution network design and service level requirements. However, DCs' operations costs have a slight effect on the number of DCs and service levels. Hence, variability on demand, outbound transportation costs, and the number of customers are highly relevant in designing an efficient and effective distribution network.

Nevertheless, our research has certain limitations. We only focused on the analysis of what-if scenarios, although we presented a novel distribution network design model in the paper. Hence, we suggest a few extensions for further research. The first is to introduce a case with big data to test our model and then use it as a benchmark. The second is to investigate different design options by using of our validated model as benchmark, thereby identifying the best design alternative. Because big data exist and CPLEX is slow in determining an optimal solution when the size of distribution network is large, the third extension should be to develop efficient algorithms to solve our model when big data are involved.

Acknowledgements The authors wish to thank the anonymous reviewers for their constructive comments.

References

- Chae, B. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247–259.
- Chae, B., & Olson, D. (2013). Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology and Decision Making*, 12(1), 9–26.
- Chae, B., Yang, C., Olson, D., & Sheu, C. (2014). The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. *Decision Support Systems*, 59, 119–126.
- de Matta, R. (2016). Contingency planning during the formation of a supply chain. *Annals of Operations Research*, 236, 1–31.
- Dubey, R., Gunasekaran, A., Childe, S., Wamba, S., & Papadopoulos, T. (2016). The impact of big data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology*, 84(1), 631–645.
- Gölzera, P., Simon, L., Cato, P., & Amberg, M. (2015). Designing global manufacturing networks using Big Data. *Procedia CIRP*, 33, 191–196.
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80.
- Kılıç, Y., & Tuzkaya, U. (2016). A two-stage stochastic mixed-integer programming approach to physical distribution network design. *International Journal of Production Research*, 53(4), 1291–1306.

- Liao, H., Hsieh, L., & Lin, S. (2011). A multi-objective evolutionary optimization approach for an integrated location-inventory distribution network problem under vendor-managed inventory systems. *Annals of Operations Research*, 186, 213–229.
- Oliveira, M., McCormack, K., & Trkman, P. (2012). Business analytics in supply chains—The contingent effect of business process maturity. *Expert Systems with Applications*, 39(5), 5488–5498.
- Omar, B. A. (2013). Parcel distribution network design problem. *Operational Research*, 13(2), 211–232.
- Rezaee, A., Dehghanian, F., Fahimnia, B., & Beamon, B. (2015). Green supply chain network design with stochastic demand and carbon price. *Annals of Operations Research*, 1–23. doi:10.1007/10479-015-1936-z.
- Sanders, N. (2014). *Big data driven supply chain management: A framework for implementing analytics and turning information into intelligence*. Pearson FT Press.
- Sangwan, K. S., & Mittal, V. K. (2015). A bibliometric analysis of green manufacturing and similar frameworks. *Management of Environmental Quality: An International Journal*, 26(4), 566–587.
- Shu, J., Ma, Q., & Li, S. (2010). Integrated location and two-echelon inventory network design under uncertainty. *Annals of Operations Research*, 181, 233–247.
- Song, M., Fisher, R., Wang, J., & Cui, L. (2016). Environmental performance evaluation with big data: theories and methods. *Annals of Operations Research*, 238, 1–14.
- Souza, C. G. (2014). Supply chain analytics. *Business Horizons*, 57, 595–605.
- Tan, K. H., Zhan, Y. Z., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, 165, 223–233.
- Tsao, C., Zhang, Q., & Chen, T. (2016). Multi-item distribution network design problems under volume discount on transportation cost. *International Journal of Production Research*, 54(2), 426–443.
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84.
- Wang, G., Gunasekaran, A., Ngai, E., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98–110.
- Wang, G., & Lei, L. (2012). Polynomial-time solvable cases of the capacitated multi-echelon shipping network scheduling problem with delivery deadlines. *International Journal of Production Economics*, 2(137), 263–271.
- Wang, G., & Lei, L. (2015). Integrated operations scheduling with delivery deadlines. *Computers & Industrial Engineering*, 85, 177185.
- Wang, G., Lei, L., & Lee, K. (2014). Supply chain scheduling with receiving deadlines and a non-linear penalty. *Journal of the Operational Research Society*, 66(3), 380–391.
- Wang, L., & Alexander, C. A. (2015). Big data driven supply chain management and business administration. *American Journal of Economics and Business Administration*, 7(2), 60–67.
- Zhang, K., Wang, G., & Shu, J. (2013). Logistics distribution network design with two commodity categories. *Journal of the Operational Research Society*, 64(9), 1400.