

Data-driven operations and supply chain management: established research clusters from 2000 to early 2020

Abstract

Scholars and practitioners have long recognised the importance of data-driven operations and supply chain management (OSCM), which typically centres on production and logistics. Given the impressive development of big data analytics (BDA), there are many papers on BDA in OSCM, possibly indicating a shift of focus in OSCM studies. Nevertheless, research finds that firms struggle with BDA adoption (Boldosova 2019; Caesarius and Hohenthal 2018), which suggests the existence of gaps in the literature. Therefore, we conduct this systematic literature review (SLR) of research on data-driven OSCM from 2000 to early 2020 to identify established research clusters and literature lacunae. Using co-citation analysis software and double-checking the results with factor analysis and multidimensional scaling as methodological triangulation, we find six research clusters on data-driven OSCM, whose primary topics are identified by keyword co-occurrence analysis. Overall, manufacturing is commonly-studied in data-driven OSCM scholarship, and quantitative modelling dominates research on data-driven inventory management, demand forecasting and transportation. However, case study, survey, and conceptual modelling are more frequently deployed in research on data-driven SCM and manufacturing system integration. In addition to the insights contributed to the literature, our study is amongst the first efforts to undertake a data-driven methodologically-triangulated SLR of data-driven OSCM.

Keywords: data-driven operations; supply chain management; systematic literature review; co-citation analysis.

Introduction

Through statistics, optimisation (Tiwari, Wee, and Daryanto 2018) and other supply chain analytics (SCA) tools and techniques (Chae, Olson, and Sheu 2014), data have long been exploited in operations and SC management (OSCM), where production and logistics play key roles. The literature has shown positive correlation between the use of data-based tools and

OSCM efficacy (Chavez et al. 2017). In Song et al.'s paper (2018), a large usable survey sample of 309 merchants on an online B2C platform in China run by Alibaba Group was obtained at a high response rate and regression analysis was then used to show the impact of data analytics on the sellers' performance under interaction with product variety and competitive intensity. From a different perspective, Chae et al. (2014), by applying PLS-SEM (partial least squares structural equation modelling) to analyse a global survey sample of 533 manufacturing plants in 15 countries and 14 industries, proved an indirect yet significant impact of advanced analytics on operations via the mediation of such SCM initiatives as Total Quality Management and Just-in-Time. The results also highlighted an indirect effect of data accuracy on operating efficiency through the moderation of SCM initiatives and mediation of manufacturing and planning quality (Chae et al. 2014). Parlaying the same data and methodology, Chae, Olson, and Sheu (2014) corroborated the direct and indirect association between operational effectiveness and three SCA components, i.e., data management, IT-based SC planning and performance management.

In addition to empirical findings, the literature on OSCM discusses frameworks and models to further develop data-driven SCs. For instance, in an attempt to allay overreliance on expert yet possibly subjective judgement (Provost and Fawcett 2013), Cheng et al. (2020) proposed a data-driven technique based on support vector machine (SVM) for supplier evaluation, a fundamental OSCM issue. The simulation-based assessment of the model on a big firm's dataset produced promising results (Cheng et al. 2020). Another example is the decision-making model developed by Long (2018) to tackle the turbulence and intricacy of internal and external factors in SCs. Indeed, that high-dimensional model takes account of data granularity, inputs domain knowledge and employs agent-based simulation for parameter adjustment and verification before finalising the solution (Long 2018). Also aimed at dealing with external uncertainty, Medina-González et al.'s multi-objective data-driven model (2020) exploits

machine learning (ML), robust optimisation and meta-multiparametric programming to handle the stochasticity related to raw materials, demand, and environmental and social impact parameters to optimally manage a multitier SC of bio-energy production. In addition to demand intermittence, the semiconductor industry faces uncertainty from shortening product lifecycle, long lead time, etc. (Uzsoy, Fowler, and Mönch 2018). In response to those challenges, Fu and Chien (2019) proposed a UNISON-based data analytics model, which integrates ML and adopts temporal aggregation-disaggregation mechanisms for demand forecasting, and empirically validated their framework with a global electronics distributor. Even with agri-food, technological advances are transforming its SC into a digital data-driven environment (Kamble, Gunasekaran, and Gawankar 2020).

Given the recognised value of data, more and more firms store and leverage data for decision-making (Brynjolfsson and McElheran 2016). Such data from end-to-end SCM practices have exponentially increased (Tiwari, Wee, and Daryanto 2018) in volumes, at a high velocity, in various formats and from multiple sources, hitting the big data level (Megahed and Jones-Farmer 2015). According to Yu et al. (2019), the orientation and capability to leverage big data (BD) for SCM, conceptualised as data-driven SC orientation, is positively-associated with financial performance when complemented by innovation competencies. Therefore, many scholars and companies are striving to develop BD analytics (BDA) capabilities to realise BD potential (Tiwari, Wee, and Daryanto 2018). Several advantages of BDA in SCM have been reported (Nguyen et al. 2018; Wamba et al. 2015). One example is the cost savings derived from BD-enhanced improvement in detection of and response to disruption (Sheffi 2015), and another is the decision-making improvement in emergency service response of the New South Wales (NSW) State Emergency Service (SES) thanks to real-time information sharing, enhanced transparency and accountability, and BD-based intelligence (Wamba et al. 2015). BDA has been studied and adapted to many OSCM aspects such as BD-driven SCM and

organisational performance (Yu et al. 2018; Gawankar, Gunasekaran, and Kamble 2020), sustainability (Tseng et al. 2019), and scheduling optimisation for condition-based machinery maintenance (Kumar, Shankar, and Thakur 2018).

Nevertheless, BDA adoption remains in early stages in several fields (Baig, Shuib, and Yadegaridehkordi 2019), such as healthcare (Fatt and Ramadas 2018), maritime industry (Zhang and Lam 2019), and logistics and SCM (Lai, Sun, and Ren 2018). In fact, many firms are struggling with BDA adoption (Boldsova 2019; Caesarius and Hohenthal 2018), which suggests research avenues for BDA adoption.

Irfan and Wang (2019) used two constructs, namely IT resources and data assimilation, to operationalise data-driven capabilities. In particular, the former includes IT infrastructure and database, whereas the latter denotes if data are utilised for order management, forecasting or planning. This accords with empirical SCM research predicated on organisational information processing theory (Srinivasan and Swink 2018; Williams et al. 2013; Chen, Preston, and Swink 2015) where information processing capabilities (IPC) moderate/mediate the correlation between organisational performance/competitiveness and SC visibility enabled by the data gained from IT utilisation. Williams et al. (2013) and Srinivasan and Swink (2018) broadly defined IPC as the ability to process and leverage data for a specific purpose whilst Chen, Preston, and Swink (2015) particularly modelled BDA as the unique IPC mediating the relationship conceptualised. In effect, several studies define or refer to data-driven SCs as those that leverage BDA to improve SC competitiveness (Yu et al. 2018), sustainability (Tseng et al. 2019; Kamble, Gunasekaran, and Gawankar 2020), or performance (Yu et al. 2019; Chavez et al. 2017). This might indicate a shift of focus to BDA in research on data-driven OSCM.

Recognising those issues and paradigm shifts, we conduct a systematic literature review (SLR) to identify the knowledge structure of research on data-driven OSCM since 2000. The aims include determining clusters of studies on data-driven OSCM and ascertaining subfields

and lacunae in emergent research topics, e.g., ML and BDA, so that directions and avenues for further research could be elicited. In other words, this SLR is to answer the following research questions:

- i. What is the knowledge structure of research on data-driven OSCM from 2000 to early 2020 (published from 2000 to 2019 or accepted for publication before 2020)?
- ii. Are there any emerging subfields of research on data-driven OSCM, which have constituted a recognisable cluster in the field? Or are there any topics on data-driven OSCM which have recently been heeded by both scholars and practitioners or integrated into an established subfield?

In this literature review, we define *data-driven* or *data-based OSCM* broadly as the use of data for OSCM practices, e.g., forecasting, tracking and scheduling, in accordance with Williams et al. (2013), Srinivasan and Swink (2018) and Irfan and Wang (2019). The selected papers from the Web of Science (WOS) and other databases, which were published from 2000 to 2019 or accepted for publication before 2020, are then analysed, using co-citation analysis to identify research clusters. Factor analysis (FA) and multidimensional scaling (MDS) are then carried out to validate the clustering results. Next, we read the papers retained from the analysis and use keyword co-occurrence analysis to determine each cluster's theme.

There have been SLRs on OSCM using data-driven approaches (Xu et al. 2018) or on a subfield of data-driven OSCM (Kamble, Gunasekaran, and Gawankar 2020), but so far we have seen few SLRs on data-driven OSCM, which employ such data-driven approaches as co-citation analysis. Therefore, our study will be amongst the first attempts to apply this data-driven methodology for a literature review on data-driven OSCM. We would like to emphasise that our SLR focusses on research into data-driven OSCM practices and applications, not on fundamental research which advances data-driven methodology without evident connection to OSCM practices. For instance, if a paper uses BDA to illuminate the characteristics of

multinationals' SCs rather than address an OSCM issue, e.g., supply management and supplier selection, we do not include that article in our sample. Likewise, an optimisation study might be excluded unless data play a role in its model formulation and a practical issue, e.g., demand uncertainty or censored demand, is explicitly targeted. Conversely, our sample incorporates qualitative papers which, for example, discuss the benefits or performance of data-driven OSCM or the management and utilisation of data for OSCM decision-making.

This paper is divided into five sections. Following this introduction, the next section presents our paper's Methodology, whereas the third one illustrates the clustering results. Future research suggestions based on our findings are provided in section 4, and Conclusion is the last section, which also discusses this study's contributions and limitations.

Methodology

This paper is framed in line with Durach, Kembro, and Wieland's six-step guidelines (2017) for SLRs.

- i. First, we define the research questions and justify the timeliness, relevance and expected contribution of our SLR (see Introduction).
- ii. Then, we determine the inclusion and exclusion criteria.
- iii. Next, we retrieve a sample of potentially pertinent literature and specify the search procedure, databases, and keywords.
- iv. By applying the predetermined inclusion and exclusion criteria, we choose the relevant papers.
- v. Given the selected articles, we synthesise the literature, using co-citation analysis, FA, MDS and keyword co-occurrence analysis.
- vi. Finally, we present a descriptive summary of the selected publications and report the thematic findings.

Similar procedures can be found in many recent literature reviews, e.g., Badi and Murtagh

(2019), Díaz, Imitola, and Acosta Amado (2019), Martins and Pato (2019), Rebs, Brandenburg, and Seuring (2019), Mascarenhas, Ferreira, and Marques (2018), Feng, Zhu, and Lai (2017) and Liu et al. (2017). Nonetheless, each step can be modified to support the research. For instance, in the sampling step, some authors utilise cross-referencing or backward snowball search to find additional germane papers in the bibliographies of the selected papers (Rebs, Brandenburg, and Seuring 2019; Kamble, Gunasekaran, and Gawankar 2020; Hosseini, Ivanov, and Dolgui 2019). In forward snowball search, authors seek relevant studies that cite the selected articles (Martins and Pato 2019; Karttunen 2018). Another example is the use of bibliometric software to support co-citation-based clustering analysis (Rebs, Brandenburg, and Seuring 2019; Díaz, Imitola, and Acosta Amado 2019; Feng, Zhu, and Lai 2017; Mascarenhas, Ferreira, and Marques 2018). Since the procedure Durach, Kembro, and Wieland (2017) discussed is commonly-adopted, we frame our research in line with their guidelines to reinforce our SLR rigour. However, the validity and originality of our review are further enhanced by the deployment of multiple co-citation data analysis methods in step (v) for methodological triangulation, which Durach, Kembro, and Wieland (2017) did not explicitly pointed out. We elaborate on the applied procedure in the next subsections, beginning with step 2.

Inclusion and exclusion criteria

We adopt the inclusion criteria utilised in Glock et al.'s SLR (2019), which are in conformity with prior studies as follows:

- Language: English.
- Time span: from 2000 to 2019. Papers with publication time after 2019 which were accepted for publication and became available online before 2020 are also included in the sample.
- Article type: Academic (peer-reviewed) journal article. In accordance with prior literature reviews (e.g. He et al. 2018; Xu et al. 2018), we only include research or review papers published in peer-reviewed journals in our study because such publications are considered

verified knowledge (Ramos-Rodríguez and Ruíz-Navarro 2004).

- At least one keyword mentioned in the title, abstract or keywords of the paper (cf. Badi and Murtagh 2019). Words with similar or related meaning are accepted, e.g., manufacturing and production, purchasing and procurement.

Regarding the exclusion criteria, we eliminate the papers which include the keyword(s) but do not pertain directly to OSCM, e.g., research on trading and price prediction of financial asset. Since we follow the broad conceptualisation of data-driven OSCM (cf. Irfan and Wang 2019), the research papers selected can focus on data assimilation or data management resources (e.g. IT), but must explicitly mention the purpose(s) for which the data (recorded internally, obtained from external sources or shared with partners) are used in practice. In other words, the data in the selected articles must be employed to address a practical issue, e.g., supplier selection and inventory management, rather than test a conceptual model unless it is about the relationship between data-driven OSCM and organisational performance. We peruse the abstract and research question(s) of each paper to assess its relevance to our SLR focus.

Databases and keywords

To retrieve germane literature, we primarily use the WOS, which provides access to over ‘*100 million references from 33,000 journals*’ (Martins and Pato 2019). According to Chadegani et al. (2013), this database covers many publication outlets in English from such publishers as Elsevier, Springer, Taylor and Francis, etc. (Kamble, Gunasekaran, and Gawankar 2020). This platform has been used by many scholars for literature review, e.g., Kamble, Gunasekaran, and Gawankar (2020), Díaz, Imitola, and Acosta Amado (2019) and Rebs, Brandenburg, and Seuring (2019). Nonetheless, other researchers, e.g., Badi and Murtagh (2019) and Martins and Pato (2019), searched other databases in addition to the WOS to ensure holistic literature retrieval. Indeed, although the WOS covered a large proportion of the papers identified in Liu et al.’s literature review (2017), other databases still contained a nontrivial number of unique

publications that were not included on the WOS. Therefore, we also search other databases, namely Emerald, INFORMS, ProQuest, Sage, Science Direct, Springer, Taylor & Francis, and Wiley, and reformat their metadata as per the WOS template.

Table 1. Some keywords used in recent literature reviews on OSCM

With regard to the keywords, we use ‘*data-driven*’ or ‘*data-based*’ with those that have been utilised in previous literature reviews or deemed to be important OSCM themes (**Table 1**). Although these keywords cannot cover all OSCM topics, we believe that such generic keywords as SC and operations help find publications on uncommon subfields. To avoid overlooking important papers, we adopt cross-referencing to identify commonly-cited articles in our sample. Specifically, if a peer-reviewed journal paper is cited at least four times by our selected studies and satisfies the inclusion criteria aforesaid, we add that publication to our sample. The threshold of four citations is proposed by Feng, Zhu, and Lai (2017).

Literature synthesis

To determine clusters of research in the selected publications, we deploy co-citation analysis, which was put forth by Small (1973). When two papers are cited together in another study, we regard that as one co-citation, whose frequency indicates the similarity of the papers’ research topics (Small 1973; Batistič, Černe, and Vogel 2017). Co-citation analysis is primarily to identify the main research topics in the literature (Feng, Zhu, and Lai 2017). This method is preferred to bibliographic coupling, which assesses two papers’ similarity based on their bibliographies, because co-citation-based similarity reflects many other scholars’ judgement and can vary over time, whereas its counterpart predicated on bibliographic coupling is fixed and reflective of a limited number of researchers only (Wang, Liang, et al. 2016; Zupic and Čater 2015).

We follow the threshold of four co-citations recommended by Feng, Zhu, and Lai (2017) to ensure adequacy and tractability of the citation data for analysis (Zupic and Čater 2015). This

methodology has been adopted in previous literature reviews (Batistič, Černe, and Vogel 2017; Feng, Zhu, and Lai 2017; Wang, Liang, et al. 2016). To analyse the co-citation data, we use VOSviewer open-source software (version 1.6.10), which was developed by van Eck and Waltman (2010) and has been employed in co-citation-based literature reviews, e.g., Wang, Liang, et al. (2016) and Zhao, Zhang, and Kwon (2018).

As van Eck et al. (2010) explained, VOSviewer determines the locations of n items on a map based on their co-citation similarity and then clusters them based on their distances (van Eck et al. 2010). Papers assigned to the same cluster by VOSviewer are those that are often co-cited. In their comparison study, van Eck et al. (2010) found that VOSviewer visualisation is more effective than that of MDS, which, by minimising a stress function, projects items into a low-dimensional space so that the distance between any two items can best reflect their relatedness or similarity. To check robustness of VOSviewer results, we utilise MDS (followed by k-means clustering) in scikit-learn (Pedregosa et al. 2011) and FA (with SEM) in STATA 15.1 for methodological triangulation. With FA, we can find the high-order level factors that capture most of the correlation space in the co-citation matrix. Items captured by the same factor are highly correlated with each other and associated with the same latent variable, or in other words, an overarching topic. These analyses were performed by Zhao, Zhang, and Kwon (2018) and Wang, Liang, et al. (2016), but our paper is amongst the first to use all three tools for a data-driven SLR of data-driven OSCM.

We use BibExcel (Persson, Danell, and Schneider 2009) to adapt the input data for analysis in different software (cf. Xu et al. 2018; Zhao, Zhang, and Kwon 2018; Feng, Zhu, and Lai 2017; Hosseini, Ivanov, and Dolgui 2019). We then apply the three clustering techniques to the pre-processed data and retain the papers that are consistently clustered by all the methods. Next, we read the papers in each cluster to determine its theme with the support of VOSviewer keyword co-occurrence analysis (cf. Zupic and Čater 2015; Ikeziri et al. 2019). In addition to

the author-assigned and WOS-provided keywords in the metadata, we employ Wang, Liu, and Wang's PageRank-based algorithm (2007) to extract important tokens in the abstract and title of each paper. If these tokens are not already included in the keyword sections of the article, we add them to the database before carrying out VOSviewer keyword co-occurrence analysis. With this, we can take advantage of the abstract content in the metadata and make additional contributions to our rigorous analysis. Details on our implementation of PageRank-based keyword extraction are given in the online companion.

Result report

The clustering results are reported for papers published from 2000 to 2019 or accepted for publication before 2020. Given that the WOS metadata give credit to first authors only, we focus our analysis on the papers rather than their researchers. As citation is commonly-used to measure a study's impact (Feng, Zhu, and Lai 2017), we use four indices, namely, global citation index, WOS citation index, in-sample citation index and PageRank, to identify influential papers in our sample. We utilise Google Scholar citation index as a proxy for our global citation index since Google Scholar covers many academic platforms including Scopus and WOS (Feng, Zhu, and Lai 2017). The in-sample citation index shows how many papers in our sample cite a given article. To obtain a comprehensive picture of research impact, we use PageRank which indicates the degree to which a paper is cited by highly-cited studies (Xu et al. 2018). Descriptive details of the selected papers can be found in the supplemental material.

Results and discussion

The preceding sections describe the first two steps in Durach, Kembro, and Wieland's procedure (2017). We discuss the remaining steps, namely literature retrieval and selection, literature synthesis and thematic report, in the next subsections.

Literature retrieval and selection

By setting the English language and academic/scholarly journal as search parameters, and using

the aforementioned keywords, we retrieve a total of 2341 search results from all the databases. Then, we apply the inclusion and exclusion criteria and remove duplications to obtain a sample of 398 pertinent papers, 365 of which were published from 2011 onwards. This accords with prior findings that there has been mounting growth of interest in BD from both scholars and practitioners since 2011 (Wamba et al. 2015; Nguyen et al. 2018; Tiwari, Wee, and Daryanto 2018) and that approximately 90% of the world's data have been generated ever since (Megahed and Jones-Farmer 2015).

Figure 1. Literature Retrieval and Selection

As we focus on the WOS, papers which were found in both WOS and another database are credited to the WOS only (26 September 2020). In other words, the figures reported for each non-WOS database after duplication check in **Figure 1** (those remarked with *) refer to the number of unique publications retrieved from that platform only. As can be inferred from **Figure 1**, 94.22% of the shortlisted articles are included on the WOS. This ratio is 95.07% for publications after 2010 and 84.85% of those published before. The difference between these proportions is marginally significant at the 5% significance level, implying that the WOS publication coverage for data-driven OSCM can be deemed consistent between the two periods.

With an initially shortlisted sample of 398 papers, we conduct cross-referencing to identify highly-cited and relevant references in these publications. Journal papers that are cited at least four times and related to data-driven OSCM are added to the sample and we repeat the process until no new highly-cited reference can be found. In line with the broad conceptualisation of data-driven OSCM discussed, the papers considered relevant can focus on data assimilation or IT/data infrastructure but must explicitly mention how data are utilised for OSCM. After six iterations, we find another 182 relevant studies. In total, 580 papers are considered relevant to data-driven OSCM and their bibliographies constitute the metadata for our clustering analysis. We note that this sample update has no statistically significant impact on the aforementioned

distribution of publications before and since 2011 nor between the WOS and other platforms.

The decrease in the WOS coverage proportion of papers in 2018 and 2019 in **Figure 2** is because some studies were accepted for publication and made available on the publisher's database but have not been updated on the WOS. Another reason is that with the growth of attention to data-driven OSCM from both scholars and practitioners, relevant studies are published in various journals, some of which are not included on the WOS. One interesting insight from **Figure 2** is the high correlation between the research volume of data-driven OSCM and Google Trends for '*Big Data*' and '*Industry 4.0*.' Indeed, the observed upsurge in Google user search for '*Big Data*' after 2010 was followed by the increase in popularity of '*Industry 4.0*' search queries and the growth in studies on data-driven OSCM some years later. This partially reflects the attention of OSCM scholarship to practitioners' needs.

From the 580 journal papers selected for literature synthesis, those which were cited four times or more by other selected studies are retained in co-citation analysis as Feng, Zhu, and Lai (2017) recommended. The sample in clustering then includes 196 journal articles¹, but the metadata for PageRank computation and co-citation database are from the 580-paper pool.

**Figure 2. Number of publications in data-driven OSCM from 2000 to 2019
and Google Trends for '*Big Data*' and '*Industry 4.0*'**

The next subsection synthesises those publications to attain an overall picture of the data-driven OSCM literature before cluster analysis is performed.

Literature synthesis results

Overall, the 580 papers selected were published in 229 peer-reviewed journals, amongst which the journals with the largest numbers of selected publications are presented in **Table 2**. Of a particular note is that there is no huge difference between the numbers of papers selected from

¹ These 196 papers are summarised in the online appendix. Details on the 580 papers selected are provided upon request.

these publication outlets. In fact, no journal accounts for a majority of research articles on data-driven OSCM in our sample. This is consistent with the broadening interdisciplinary and integrative scope of OM (Narasimhan 2014) and SCM as in the extant literature cited by Halldórsson, Hsuan, and Kotzab (2015). In addition to journals specialising in transportation and production, two subfields of OSCM, we can see, in **Table 2**, journals in operations research and engineering, which also relate to OSCM. We can see the diverse yet mostly high impact factors of these journals, signifying that there are many high-quality studies on data-driven OSCM.

Table 2. Journals with the most papers included in the sample

Looking at some of the most highly-cited papers in our sample (see **Table 3**), we can observe a similar pattern between Google Scholar (proxy for global citation) and WOS citation indices, which is not surprising since Google Scholar citation index includes that of the WOS. Indeed, the correlation between global (Google Scholar) and WOS citation indices is 0.9320 for the 580 papers selected. However, the correlation between in-sample citation and Google Scholar citation (WOS citation) is only 0.5689 (0.6021). As shown in **Table 3**, only one paper belongs to both groups of top-ten globally and in-sample cited research. According to Feng, Zhu, and Lai (2017), this disparity can be explicated by the varying degree of attention from scholars in different fields. This means that some globally highly-cited research is less heeded or deemed less pertinent by academics in data-driven OSCM and vice versa.

Table 3. Most highly-cited papers in the sample with respect to global and in-sample citation

With respect to Google Scholar citation, the top-ten papers study a variety of OSCM subfields, e.g., OSCM overall (Wamba et al. 2015), inventory and sales (Elmaghraby and Keskinocak 2003; Ben-Tal et al. 2004), supply management (Ho, Xu, and Dey 2010), healthcare operations (Raghupathi and Raghupathi 2014), production (Xu 2012), and

demand/transportation forecasting (Lv et al. 2015; Zhao and Magoulès 2012; Hippert, Pedreira, and Souza 2001; Wu, Ho, and Lee 2004). Nonetheless, most of them are literature reviews (Elmaghraby and Keskinocak 2003; Ho, Xu, and Dey 2010; Zhao and Magoulès 2012; Hippert, Pedreira, and Souza 2001; Raghupathi and Raghupathi 2014). Modelling-based studies include those of Lv et al. (2015), Ben-Tal et al. (2004), and Wu, Ho, and Lee (2004), whereas Wamba et al.'s (2015) and Xu's (2012) articles are empirical and conceptual, respectively. It is interesting that although modelling (normative, descriptive, and predictive) dominates research on data-driven OSCM in general, literature reviews account for the largest proportion of the ten globally most cited papers in our sample. Explicably, literature reviews give an overview of a certain research topic and thus can be cited by many papers.

On the other hand, except for one paper on BD-driven logistics in manufacturing (Zhong et al. 2015), the top-ten most cited papers in our sample focus on BD (Analytics) in SCM in general (Hazen et al. 2014; Trkman et al. 2010; Wang, Gunasekaran, et al. 2016; Wamba et al. 2015) and forecasting in transportation (Smith, Williams, and Oswald 2002; Vlahogianni, Golias, and Karlaftis 2004; Vlahogianni, Karlaftis, and Golias 2005; Lv et al. 2015; Yin et al. 2002). They adopt diverse research methods, namely, descriptive/normative modelling (Smith, Williams, and Oswald 2002; Stathopoulos and Karlaftis 2003; Vlahogianni, Karlaftis, and Golias 2005; Yin et al. 2002; Zhong et al. 2015), conceptual modelling (Hazen et al. 2014; Wang, Gunasekaran, et al. 2016), survey (Trkman et al. 2010), case study (Wamba et al. 2015; Hazen et al. 2014), and literature review (Wang, Gunasekaran, et al. 2016; Vlahogianni, Golias, and Karlaftis 2004).

As indicated in **Table 3**, none of the top-ten papers in terms of Google Scholar citation is amongst the ten most influential papers according to PageRank score computed on our selected research on data-driven OSCM. Out of the top in-sample cited research, only two belong to the group of ten most influential articles as per this index. The correlation between in-sample

citation and PageRank index in our sample is 0.5886 whilst the figure is 0.5689 for global citation and PageRank index. This result is expected as highly-cited research is not necessarily influential because PageRank-based research impact is determined by the extent to which a study is cited by high-impact papers (Xu et al. 2018; Brin and Page 1998). Hence, the nominal value of citations cannot fully reflect a paper's influence as indicated by PageRank index. Although the 580-paper metadata are not exhaustive and our paper is not focussed on PageRank algorithm development, our results imply that more care should be taken when we assess research impact.

The next subsection will discuss the knowledge structure or, in other words, research clusters on data-driven OSCM studied since 2000.

Thematic report

We load the 580-publication metadata used for PageRank computation into VOSviewer, but only references being cited at least four times are retained as vertices. We then identify seven clusters of research on data-driven OSCM as depicted in **Figure 3**.

Figure 3. Research clusters identified by VOSviewer software

We check the robustness of VOSviewer results by performing FA in STATA and MDS in scikit-learn for methodological triangulation.

For FA, we utilise the eigenvalues and factor rotation (Yong and Pearce 2013) to select the high-order level factors that capture most correlation in the co-citation space. Each factor here can be interpreted as a cluster of closely-related studies and each item in a factor is a paper from our retained sample. There are 26 factors with eigenvalues greater than 1, which altogether explain 100% of the variation, but only nine of them have more than three items with loadings greater than or equal to 0.7 each. To test the significance of the results, we run SEM to perform convergent validity analysis (Sethi and King 1994) and discriminant validity analysis (Fornell and Larcker 1981).

Table 4. Factor analysis of co-citation matrix

As illustrated in **Table 4**, the Cronbach's alpha of each factor exceeds the 0.7 threshold, indicating consistency amongst the items included therein (Dunn, Baguley, and Brunsten 2014). Whilst the correlations between factors is below the recommended 0.85 threshold (Yu et al. 2018), which indicates good discriminant validity, the Average Variance Extracted (AVE) of each factor being higher than its squared correlations with other factors is another discriminant validity indicator (Song et al. 2018). All factor loadings are of acceptable magnitude (greater than 0.7) and statistically significant, implying good convergent validity (Sethi and King 1994). The Composite Reliability and AVE are respectively above the threshold of 0.7 and 0.5 (Fornell and Larcker 1981; Yu et al. 2018), further confirming the convergent validity. Thus, these nine factors can be considered robust.

Table 5. Result matches between VOSviewer and Factor Analysis

Table 5 demonstrates that most factors identified in FA fit entirely in the clusters appearing in VOSviewer result. Although the number of clusters/factors differs, the membership stays consistent. With these results, six clusters and 120 papers are retained.

According to van Eck et al. (2010), the similarity index used in MDS stress function can be the correlation between two items or their cosine. We run MDS with these similarity indices in scikit-learn (Pedregosa et al. 2011) and select the results of lowest dimensionality whose stress level is below the recommended threshold of 0.1 (Zhao, Zhang, and Kwon 2018). Afterwards, we perform the modified (Bagirov 2008) and fuzzy (Khan et al. 2020) k-means algorithms to cluster the cited papers. Our implementation and detailed results of MDS and k-means clustering are provided in the e-companion.

Table 6. Multidimensional Scaling preliminary results with different similarity indices

As can be seen from **Table 6**, despite changes in the number of clusters identified, the 120 papers retained in **Table 5** fit neatly in the k-means clusters appearing in MDS outputs. Still,

as observed by van Eck et al. (2010), MDS visualisation is less effective than its VOSviewer counterpart. Overall, out of the seven VOSviewer-assigned clusters, only six survive all the three clustering methods and we have a final sample of 120 papers for thematic interpretation.

In the next step, we read the 120 peer-reviewed journal articles retained from the analyses and use keyword co-occurrences to identify the themes. The lower right corner of each figure in the next subsections shows the average publication year of the papers mentioning the coloured keyword.

Big data (analytics) in OSCM

This cluster corresponds to Factor 2 with 25 papers retained. At **Figure 4**'s centre are '*big data*' and '*supply chain*,' the two keywords co-occurring most often with other keywords in this cluster. Indeed, most papers therein discussed the application or benefits of (B)DA in OSCM via literature reviews (Addo-Tenkorang and Helo 2016; Hazen et al. 2018; Wang, Gunasekaran, et al. 2016; Zhong et al. 2016), surveys and interviews (Chen, Preston, and Swink 2015; Schoenherr and Speier-Pero 2015; Yu et al. 2018; Gunasekaran et al. 2017), case studies (Hazen et al. 2014; Tan et al. 2015; Zhao et al. 2017), conceptual framework proposition (Chae 2015; Hazen et al. 2014; Giannakis and Louis 2016), or simulation (Hofmann 2017). For instance, Hazen et al. (2014) used a brief case study on jet engine remanufacturing to demonstrate how the data quality problem could be addressed to enhance data-driven SCM. In another example, Chae (2015) used data from Twitter related to companies in manufacturing, logistics, news, and IT to demonstrate his proposed analytical framework and the value of social media data in SCM. Schoenherr and Speier-Pero (2015) interviewed experts from several firms, including providers of professional services (e.g., consulting and analytics), and found that the desired skills for SCM professionals entailed data analytics.

Figure 4. Keyword co-occurrences in Cluster 1

Some papers discuss the application or benefits of BDA in a specific OSCM subfield, e.g.,

Wamba et al. (2015) presented a case study on data-driven operations of the NSW SES. Chong et al. (2016) and Cui et al. (2018) developed predictive models for retailing. As a hybrid model of service provision and manufacturing, Opresnik and Taisch (2015) proposed a conceptual framework for BD strategy in servitisation. Nonetheless, the sector discussed most often by this cluster's papers is manufacturing. To leverage BD for fault detection in manufacturing, Kumar et al. (2016) proposed a MapReduce framework, which can handle imbalanced data. Carrying out empirical research on the manufacturing industry in India, Dutta and Bose (2015) and Dubey et al. (2016) investigated the influence of BD via a survey and case study, respectively. In another Asian country's context (China), Tan et al. (2015) and Zhao et al. (2017) developed a BD-based model and illustrated it in a case study, whereas Yu et al. (2018) used survey data from manufacturing firms to validate the relationship between data-driven SCs and firm performance. In addition to the empirical research mentioned in the preceding paragraph, other examples related to the manufacturing sector are the modelling-based papers of Huang and Van Mieghem (2014) and Wu et al. (2017). This reflects a finding in the extant literature that the manufacturing sector accounts for a large proportion of OSCM research.

Based on the publication time (**Figure 4**'s lower right corner), we can see that this is a recent cluster of research given the recent rise of BD research (Wamba et al. 2015; Nguyen et al. 2018; Tiwari, Wee, and Daryanto 2018), but three of the top-ten in-sample cited papers in our sample belong to this cluster and one of them also appears in the top-ten globally cited papers (**Table 3**). This implies the importance of this research cluster on data-driven OSCM. Hazen et al. (2014) are amongst the highly-cited authors with multiple publications in this cluster. Even though modelling is a widely-used methodology in data-driven OSCM, empirical research dominates Cluster 1. Since empirical studies constitute a critical part of OM research (Gattiker and Parente 2007), the formation of a cluster of highly-co-cited research dominated by this methodology can be expected. An interesting insight from this cluster is that we did not

use ‘*big data*’ or ‘*data analytics*’ as search keywords, but their appearance in our keyword co-occurrence analysis along with ‘*data science*’ heralds the upward trend of (B)DA in data-driven OSCM.

Transportation and traffic flow prediction

Figure 5 shows that the central keywords in this cluster are ‘*traffic*’ and ‘*model*’ or, more specifically, traffic-flow forecasting models, which are deemed, by all research in this cluster, vital in (operating) advanced/intelligent transportation systems (ITS). This cluster’s primary methodology is modelling, which includes both traditional (statistical) methods, e.g., autoregressive integrated moving average (ARIMA) and econometric regression (Stathopoulos and Karlaftis 2003; Min and Wynter 2011; Ghosh, Basu, and O’Mahony 2009; Ishak and Al-Deek 2002), and ML schemes, e.g., k-nearest neighbours (Chang et al. 2012; Smith and Oswald 2003), SVM (Castro-Neto et al. 2009; Wu, Ho, and Lee 2004), and artificial neural networks (ANN) (Boto-Giralda et al. 2010; Dia 2001; Vlahogianni, Karlaftis, and Golias 2005). Some authors, e.g., Chen et al. (2001), Tan et al. (2009), Zheng, Lee, and Shi (2006), and Chan et al. (2012), combined both approaches in their models, but there are overall twice as many papers adopting ML algorithms as conventional method-based articles in this cluster. Nonetheless, the publication time (**Figure 5**) shows that these two research streams have continued to develop in parallel. Overviews and comparisons of both statistical and ML techniques can be found in the reviews of Vlahogianni, Golias, and Karlaftis (2004) and Karlaftis and Vlahogianni (2011), whereas Zhang, Wang, et al.’s survey (2011) focusses exclusively on ML. Overall, there is an increase in complexity of the models studied, but Karlaftis and Vlahogianni (2011) claim that simple and complex models can produce equally good results.

Figure 5. Keyword co-occurrences in Cluster 2

Revisiting **Table 3**, we find four article in Cluster 2 amongst the top-ten in-sample cited papers, which partially reflects this cluster’s popularity. Stathopoulos and Karlaftis (2003) and

Vlahogianni, Karlaftis, and Golias (2005) are examples of authors with multiple publications in this cluster. Overall, this is an established cluster that contributes to the methodological landscape of data-driven OSCM, where modelling dominates.

Nearly 85% of the papers in this cluster mentioned their data sources, and around three quarters of them collected data from western countries, including Australia. Even in Asia, the countries from which the research data came have developed infrastructure, e.g., Taiwan, Singapore and South Korea. This is not surprising given the infrastructure necessitated for ITS, where data can be recorded in large volumes by detector stations (Chan et al. 2012; Dia 2001) or loop detectors (Stathopoulos and Karlaftis 2003; Min and Wynter 2011). With noisy or corrupt data recorded in such systems, Chen et al. (2001) and Quek, Pasquier, and Lim (2006) investigated the impact of missing data and noise tolerance, respectively, of ANN-based traffic-flow prediction, whereas others developed models to address this issue, which include data imputation (Chen et al. 2012; Qu et al. 2009) or denoising (Jiang and Adeli 2004). Thanks to the increasing number of data-recording devices in ITS, voluminous traffic data have been collected (Zhang, Wang, et al. 2011), but we note that the retained papers did not explicitly discuss BD.

Demand forecasting

This VOSviewer-assigned cluster comprises two FA factors, one related to water demand forecasting (Factor 4) and the other to energy consumption prediction (Factor 9). We can infer from the three keywords at **Figure 6**'s centre that this cluster's topic is '*demand forecasting*.' Like Cluster 2, over 90% of this cluster's research deployed predictive modelling. The publication time (**Figure 6**) of this cluster's papers illustrates that this is an established research cluster, which is consistent with the vital role played by forecasting in OSCM (Ren and Choi 2016) in general, and in power system planning (Taylor 2003) and water distribution (Adamowski 2008) in particular.

Figure 6. Keyword co-occurrences in Cluster 3

Both traditional statistical methods, e.g., regression, exponential smoothing and ARIMA, and ML schemes, i.e., ANN, were deployed almost equally in this cluster, but those published in the early 2000s focussed more on the former (e.g., Taylor and Majithia 2000; Zhou et al. 2000, 2002; Taylor 2003; Taylor and Buizza 2003). The predictive power of both modelling approaches were compared in more recent papers, of which typical examples include the research of Adamowski et al. (e.g., Adamowski et al. 2012; Adamowski and Karapataki 2010; Adamowski 2008). According to Donkor et al. (2014), ML algorithms are often used for short-term prediction, whereas traditional models, especially regression, are for long-term decision-making. This partly explains why both approaches have been studied in parallel, but we predict that this cluster will see growth in publications combining both schemes like Cluster 2.

Of a particular note is that over 85% of this cluster's studies used data from western countries, including Australia. Since that those are industrialised and urbanised nations whose electricity and water infrastructures were built long ago and are now subject to ageing and deterioration, there is a compelling need for research into those distribution systems given that electricity and water are deemed vital in the economy and urban life (Adamowski et al. 2012; Hong and Fan 2016). This partly explains why we obtain an established research cluster on those topics. Another reason is the use of smart meters in the system, which help track demand more accurately (Adamowski et al. 2012; Hong and Fan 2016).

System integration in manufacturing

This is a recent cluster of research given that the average publication year was 2011 (**Figure 7**) with a standard deviation of 2.50. This cluster is in fact composed of Factors 5 and 6 in FA. Whilst Factor 5's articles study manufacturing in the context of radio frequency identification (RFID) (e.g., Huang et al. 2008; Huang, Zhang, and Jiang 2007, 2008; Zhang et al. 2010; Zhang, Jiang, and Huang 2008; Zhang, Huang, et al. 2011; Zhang, Qu, et al. 2011), their

counterparts in Factor 6 deal with the Internet of Things (IoT) context (e.g., Tao, Cheng, et al. 2014; Tao, Zuo, Xu, and Zhang 2014; Tao, Zuo, Xu, Lv, et al. 2014; Xu 2011; Xu, He, and Li 2014). Except for the literature reviews (i.e., Bi, Xu, and Wang 2014; Xu 2011; Xu, He, and Li 2014), the other studies in Cluster 4 are based on conceptual modelling, where conceptual frameworks/architectures are proposed to apply the IoT or RFID to modern, wireless or cloud manufacturing.

Figure 7. Keyword co-occurrences in Cluster 4

We can notice that the concept behind these papers is system integration, which is required to provide timely information for decision-making in manufacturing (Huang, Zhang, and Jiang 2008). As can be inferred from Jun et al.'s (2009) discussion about RFID applications in product lifecycle (PLC) management, data need to be shared and analysed at each PLC phase to facilitate decision-making and augment efficiency, e.g., manufacturing, maintenance and reverse logistics. With regard to cloud manufacturing, a recent model where production resources can be shared and operated in the cloud amongst multiple enterprises (Tao, Cheng, et al. 2014; Tao, Zuo, Xu, and Zhang 2014), real-time data sharing and analysis are clearly entailed for system monitoring. Obviously, such wireless technologies as the IoT and RFID now play an integral role in collecting and synchronising information in manufacturing systems involving multiple processes, levels and resources (Bi, Xu, and Wang 2014; Huang, Zhang, and Jiang 2007). Given the content commonality amongst this cluster's publications, the keyword '*system*' is explicably located at **Figure 7**'s centre, connecting all other vertices. Therefore, we name this cluster '*System integration in manufacturing.*' Like Cluster 1, this cluster is not dominated by normative, predictive or descriptive modelling as the primary methodology, but is anticipated to see more highly-cited case studies, action research and prototyping-based papers examining the empirical performance of those frameworks and architectures (e.g., Fernández-Caramés et al. 2019). The publication time (**Figure 7**) illustrates

that more recent research shifts to the IoT, which leverages RFID at an advanced level (Xu, He, and Li 2014).

Data mining in manufacturing

As can be seen at **Figure 8**'s centre, this cluster's articles study data mining in manufacturing. Indeed, manufacturing systems record and accumulate a large volume of data, which can in turn provide meaningful insight for decision-making if properly-analysed (Chien, Wang, and Cheng 2007). Therefore, research on knowledge extraction from manufacturing data has long been carried out by scholars, amongst whom the highly-cited authors with multiple publications in this cluster include Kusiak et al. (e.g., Agard and Kusiak 2004; Kusiak 2000, 2001; Kusiak and Kurasek 2001).

Figure 8. Keyword co-occurrences in Cluster 5

An overview of data-mining applications in manufacturing, including design, production operations, quality control and maintenance, can be found in Harding et al.'s review (2005). For example, Kusiak (2001) proposed a rule-structuring algorithm based on rough set theory for data-driven knowledge discovery in semiconductor manufacturing. For yield enhancement in that industry, Chien, Wang, and Cheng (2007) proposed a fault diagnosis framework whilst Braha and Shmilovici (2002) experimented with data-mining methods to improve processes. Via case studies, Kusiak (2000) and Kusiak and Kurasek (2001) illustrated data-mining applications to fault detection in wafer manufacturing and electronics assembly, respectively. An example of data-mining applications to product design is the framework Agard and Kusiak (2004) proposed to handle product families. Overall, this is an established cluster with diverse research methods adopted, but the retained studies did not specifically target BDA. Nonetheless, they indubitably laid foundations for recent research on BDA in OSCM (e.g., Dutta and Bose 2015).

Inventory management with censored demand

The keyphrases linking all other vertices in **Figure 9** include ‘*inventory*,’ ‘*lost sales*,’ ‘*newsvendor*,’ ‘*censored*,’ ‘*demand*’ and ‘*distribution*.’ In effect, this cluster’s articles develop models to address inventory-management problems where recorded sales data represent only a sample of the unknown demand distribution owing to censored demand (unrecorded lost sales).

Figure 9. Keyword co-occurrences in Cluster 6

In particular, Burnetas and Smith (2000) employed a multiarmed bandit framework and a stochastic approximation procedure respectively to propose an adaptive pricing and ordering mechanism for perishable items, where the policy is updated as new information arrives. Meanwhile, Godfrey and Powell (2001) leveraged the concave adaptive value estimation algorithm to directly arrive at an optimal decision in response to a given level of remaining inventory and emphasised that they need not estimate demand distribution. Likewise, under the assumption of no information on demand distribution or its family, ordering decisions can be optimised using only sales data with adaptive inventory policies which are developed based on stochastic gradient descent and online convex programming (Huh and Rusmevichientong 2009) or the Kaplan-Meier (KM) Estimator (Huh et al. 2011). In a different research vein, Liyanage and Shanthikumar (2005) proposed operational statistics, where demand estimation and inventory optimisation could be carried out together in one single step, directly estimating the optimal order quantity from the data under the assumption that demand distribution although unknown belongs to a given family. Seeing that the data utilised for inventory management under censored demand are just samples of the true demand distribution, Levi, Roundy, and Shmoys (2007) theoretically analysed the number of samples and sample size required to guarantee that the decision taken would result in a total cost within a predefined confidence level given that the SAA (sample average approximation) is deployed for the single-period problem and (shadow/approximate) dynamic programming for the multi-period

problem. Later, Levi, Perakis, and Uichanco (2015) derived a tighter bound for the SAA applied to data-driven inventory management under censored demand. Overall, the impact of censored demand on inventory management can be found in the theoretical analysis of Besbes and Muharremoglu (2013), and Ding, Puterman, and Bisi (2002).

This is an established research cluster on data-driven OSCM (publication time in **Figure 9**) with the primary methodology being modelling. We can see, in recent publications citing this cluster's research (e.g., Ban and Rudin 2019; Bertsimas and Kallus 2020), a rising trend of adopting ML and BDA to optimise inventory-management decisions with one-step algorithms instead of the traditional two-step approach where demand (distribution) must be estimated and then inputted into prescriptive analytics.

Future research suggestions

Our thematic discussion highlights that manufacturing, demand prediction, inventory planning and transportation/traffic forecasting are commonly-studied in data-driven OSCM. Several highly-cited models and frameworks have been developed to utilise data, be they voluminous, noisy, or incomplete (as in censored demand). Practitioners in production can refer to these established clusters to guide their decision-making but need to consider if the research context and assumptions are relevant to their enterprises. It should be noted that simple and complex models can perform equally (Karlaftis and Vlahogianni 2011), so managers should assess the fit between a model/framework and their system/practice and carefully plan the change process before deciding on adoption.

Given the broad scope of OSCM (Narasimhan 2014; Halldórsson, Hsuan, and Kotzab 2015), the small number of established research clusters on data-driven OSCM might partly account for the reasons why firms, e.g., those in retailing, maritime shipping and logistics (Boldsova 2019; Caesarius and Hohenthal 2018; Lai, Sun, and Ren 2018), face difficulty adopting BDA. Indeed, ML, BDA, and related technologies, i.e., IoT and RFID, were not explicitly discussed

in all these research clusters. Nonetheless, this opens numerous research directions.

First, despite the recognised importance of service operations in OSCM scholarship and practices (Heineke and Davis 2007), there has not been an established cluster for that subfield in research on data-driven OSCM. Most Cluster 1's papers have been recently published (after 2010), but the focus is on manufacturing, which is directly supported by Clusters 4 and 5. With servitisation being facilitated by today's BD (Opresnik and Taisch 2015), OSCM scholars should attend to data-driven service operations and servitisation to ensure practical relevance of their research endeavours.

With regard to studies on data-driven transportation, the main focus was on prediction and description of flows in transportation systems. This does not mean that there are no journal-published papers on prescriptive models for data-driven transportation. Examples include the ML-based vehicle-routing research (Mao and Shen 2018; Tang et al. 2019; Lee et al. 2020). Nonetheless, with so few articles in the literature, there are obviously plentiful opportunities for further research, e.g., empirical evaluation of such data-driven vehicle-routing models. Future research can investigate how the improved traffic-flow forecast can benefit delivery planners and logistics managers and which forecast horizon optimally facilitates their operations planning.

Turning next to demand-forecasting studies, we can see articles on prediction of demand for electricity and water. This implies that research on demand forecasting for other products and services has plenty of scope for further exploration, e.g., how to leverage data analytics for demand forecasting in e-commerce (see Ferreira, Lee, and Simchi-Levi 2016), and how to integrate traditional and ML approaches into a firm's system if their combination improves demand planning. Case studies and action research can boost practitioners' confidence in the model's efficacy and provide them with implementation guidelines, which should illuminate the necessitated investment, extent of change and change management process.

As regards research on data-driven inventory management, its cluster is dominated by modelling papers. Thus, empirical studies on those models' real-life performance are essential additions to this research topic. Moreover, the data-driven inventory-management literature can be enriched by applying ML schemes and comparing or combining traditional optimisation approaches and ML algorithms (e.g., Ferreira, Lee, and Simchi-Levi 2016; Bertsimas and Kallus 2020). There are similar papers in our 580-paper pool (e.g., Sachs and Minner 2014; Ban and Rudin 2019), but their in-sample citations are currently below the inclusion threshold in our co-citation analysis.

Finally, there are several articles on other subfields of OSCM such as supply management (Cheng et al. 2020), warehousing (Fernández-Caramés et al. 2019), distribution planning (Chen et al. 2017), retail management (Bernstein, Modaresi, and Sauré 2018; Ozgormus and Smith 2020), and maintenance (Kumar, Shankar, and Thakur 2018), some of which are highly-cited and included in our sample, e.g., supply management (Ho, Xu, and Dey 2010; Cheng et al. 2020), but have not established a distinct cluster in data-driven OSCM. Thus, the need for further research is inevitable. However, to avoid compromising our SLR rigour, we limit the scope of this section to the data-based results of our analyses.

Conclusion

In this paper, we used the keywords and databases commonly-used in the OSCM literature to seek pertinent publications and applied co-citation analysis to determine the knowledge structure of data-driven OSCM since 2000. Our co-citation analysis was performed by VOSviewer open-source software and its robustness was checked by methodological triangulation with FA in STATA and scikit-learn MDS-based enhanced k-means clustering (Bagirov 2008; Khan et al. 2020). We read the 120 highly-co-cited peer-reviewed journal articles retained in clustering and collated our academic judgement against VOSviewer keyword co-occurrence analysis to identify each cluster's theme. There are prior reviews on

data-driven OSCM and co-citation-based SLRs, but our paper is amongst the first endeavours to conduct a data-driven methodologically-triangulated SLR of data-driven OSCM. This is the originality of our study.

There are six clusters of research appearing in our analysis results, namely *Big data (analytics) in OSCM*, *Transportation and traffic flow prediction*, *Demand forecasting*, *System integration in manufacturing*, *Data mining in manufacturing*, and *Inventory management with censored demand*. Amongst these clusters, ML and BDA have been widely-undertaken in the literature on SCM, production, traffic prediction and demand forecasting. Whilst empirical research dominates the research cluster on BDA in OSCM, the primary methodology adopted by the papers on transportation and forecasting is modelling. Traditional statistical and econometric approaches remain commonly-deployed in forecasting, but ML programs and BDA are becoming popular.

With the broad scope of OSCM in today's BD context, data-driven OSCM has been studied across geography, but there is a call for more research in other contexts and subfields of OSCM. Indeed, given the interdisciplinary nature of OSCM (Narasimhan 2014; Halldórsson, Hsuan, and Kotzab 2015), the topics of the research clusters identified here are slightly limited. This partly impounds on why firms are struggling with BDA adoption. Therefore, there exists a need for research into other data-based OSCM subfields such as warehousing, supply management, service operations and retailing, where data-driven tools, especially ML and BDA, can be useful or may be in demand. With respect to the established modelling-based clusters, more empirical research is needed, e.g., traffic-flow forecasting in developing countries and inventory management in real-life contexts.

Albeit we carry out a systematic literature review, which has allowed finding answers to our research questions, our study has limitations. First, the clusters identified may be confined to the keywords utilised for database search. However, since such generic keywords as SCM, OM

and logistics were used and that research on other OSCM subfields are highly cited (see **Table 3**), the clusters identified here may well reflect the current state of knowledge of the OSCM literature published since 2000. Second, one inherent drawback of co-citation analysis is its inability to identify emerging research areas (Fahimnia et al. 2019) as lately published articles are less likely to be included given their insufficient time to accumulate citations. Nonetheless, this indirectly confirms the proposed research opportunities.

References

- Adamowski, Jan, Hiu Fung Chan, Shiv O. Prasher, Bogdan Ozga-Zielinski, and Anna Sliusarieva. 2012. "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada." *Water Resources Research* 48 (1):W01528.
- Adamowski, Jan Franklin. 2008. "Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks." *Journal of Water Resources Planning and Management* 134 (2):119-28.
- Adamowski, Jan, and Christina Karapataki. 2010. "Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms." *Journal of Hydrologic Engineering* 15 (10):729-43.
- Addo-Tenkorang, Richard, and Petri T. Helo. 2016. "Big data applications in operations/supply-chain management: A literature review." *Computers & Industrial Engineering* 101:528-43.
- Agard, B., and A. Kusiak. 2004. "Data-mining-based methodology for the design of product families." *International Journal of Production Research* 42 (15):2955-69.
- Badi, Sulafa, and Niamh Murtagh. 2019. "Green supply chain management in construction: A systematic literature review and future research agenda." *Journal of Cleaner Production* 223:312-22.
- Bagirov, Adil M. 2008. "Modified global k-means algorithm for minimum sum-of-squares clustering problems." *Pattern Recognition* 41 (10):3192-9.
- Baig, Maria Ijaz, Liyana Shuib, and Elaheh Yadegaridehkordi. 2019. "Big data adoption: State of the art and research challenges." *Information Processing & Management* 56 (6):102095.
- Ban, Gah-Yi, and Cynthia Rudin. 2019. "The Big Data Newsvendor: Practical Insights from Machine

- Learning." *Operations Research* 67 (1):90-108.
- Batistič, Saša, Matej Černe, and Bernd Vogel. 2017. "Just how multi-level is leadership research? A document co-citation analysis 1980–2013 on leadership constructs and outcomes." *The Leadership Quarterly* 28 (1):86-103.
- Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski. 2004. "Adjustable robust solutions of uncertain linear programs." *Mathematical Programming* 99 (2):351-76.
- Bernstein, Fernando, Sajad Modaresi, and Denis Sauré. 2018. "A Dynamic Clustering Approach to Data-Driven Assortment Personalization." *Management Science* 65 (5):2095-115.
- Bertsimas, Dimitris, and Nathan Kallus. 2020. "From Predictive to Prescriptive Analytics." *Management Science* 66 (3):1025-44.
- Besbes, Omar, and Alp Muharremoglu. 2013. "On Implications of Demand Censoring in the Newsvendor Problem." *Management Science* 59 (6):1407-24.
- Bi, Zhuming, Li Da Xu, and Chengen Wang. 2014. "Internet of Things for Enterprise Systems of Modern Manufacturing." *IEEE Transactions on Industrial Informatics* 10 (2):1537-46.
- Boldosova, Valeriia. 2019. "Deliberate storytelling in big data analytics adoption." *Information Systems Journal* 29 (6):1126-52.
- Boto-Giralda, Daniel, Francisco J. Díaz-Pernas, David González-Ortega, José F. Díez-Higuera, Míriam Antón-Rodríguez, Mario Martínez-Zarzuela, and Isabel Torre-Díez. 2010. "Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks." *Computer-Aided Civil and Infrastructure Engineering* 25 (7):530-45.
- Braha, Dan, and Armin Shmilovici. 2002. "Data mining for improving a cleaning process in the semiconductor industry." *IEEE Transactions on Semiconductor Manufacturing* 15 (1):91-101.
- Brin, Sergey, and Lawrence Page. 1998. "The anatomy of a large-scale hypertextual web search engine." *Computer networks and ISDN systems* 30 (1-7):107-17.
- Brynjolfsson, Erik, and Kristina McElheran. 2016. "The Rapid Adoption of Data-Driven Decision-Making." *The American Economic Review* 106 (5):133-9.
- Burnetas, Apostolos N., and Craig E. Smith. 2000. "Adaptive Ordering and Pricing for Perishable Products." *Operations Research* 48 (3):436-43.

- Caesarius, Leon Michael, and Jukka Hohenthal. 2018. "Searching for big data: How incumbents explore a possible adoption of big data technologies." *Scandinavian Journal of Management* 34 (2):129-40.
- Castro-Neto, Manoel, Young-Seon Jeong, Myong-Kee Jeong, and Lee D. Han. 2009. "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions." *Expert Systems With Applications* 36 (3, Part 2):6164-73.
- Chadegani, Arezoo Aghaei, Hadi Salehi, Melor Yunus, Hadi Farhadi, Masood Fooladi, Maryam Farhadi, and Nader Ale Ebrahim. 2013. "A comparison between two main academic literature collections: Web of Science and Scopus databases." *Asian Social Science* 9 (5):18-26.
- Chae, Bongsug. 2015. "Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research." *International Journal of Production Economics* 165:247-59.
- Chae, Bongsug, David Olson, and Chwen Sheu. 2014. "The impact of supply chain analytics on operational performance: a resource-based view." *International Journal of Production Research* 52 (16):4695-710.
- Chae, Bongsug, Chenlung Yang, David Olson, and Chwen Sheu. 2014. "The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective." *Decision Support Systems* 59:119-26.
- Chan, Kit Yan, Tharam S Dillon, Jaipal Singh, and Elizabeth Chang. 2012. "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm." *IEEE Transactions on Intelligent Transportation Systems* 13 (2):644-54.
- Chang, H, Youngjoo Lee, B Yoon, and Sanghoon Baek. 2012. "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences." *IET Intelligent Transport Systems* 6 (3):292-305.
- Chavez, Roberto, Wantao Yu, Mark A. Jacobs, and Mengying Feng. 2017. "Data-driven supply chains, manufacturing capability and customer satisfaction." *Production Planning & Control* 28 (11-12):906-18.
- Chen, Chenyi, Yin Wang, Li Li, Jianming Hu, and Zuo Zhang. 2012. "The retrieval of intra-day trend and its influence on traffic prediction." *Transportation Research Part C: Emerging Technologies*

22:103-18.

- Chen, Daniel Q., David S. Preston, and Morgan Swink. 2015. "How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management." *Journal of Management Information Systems* 32 (4):4-39.
- Chen, Haibo, Susan Grant-Muller, Lorenzo Mussone, and Frank Montgomery. 2001. "A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting." *Neural Computing & Applications* 10 (3):277-86.
- Chen, Xin, Wenchuan Wu, Boming Zhang, and Chenhui Lin. 2017. "Data-Driven DG Capacity Assessment Method for Active Distribution Networks." *IEEE Transactions on Power Systems* 32 (5):3946-57.
- Cheng, Yijun, Jun Peng, Xin Gu, Xiaoyong Zhang, Weirong Liu, Zhuofu Zhou, Yingze Yang, and Zhiwu Huang. 2020. "An intelligent supplier evaluation model based on data-driven support vector regression in global supply chain." *Computers & Industrial Engineering* 139:105834.
- Chien, Chen-Fu, Wen-Chih Wang, and Jen-Chieh Cheng. 2007. "Data mining for yield enhancement in semiconductor manufacturing and an empirical study." *Expert Systems With Applications* 33 (1):192-8.
- Chong, Alain Yee-Loong, Boying Li, Eric W. T. Ngai, Eugene Ch'ng, and Filbert Lee. 2016. "Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach." *International Journal of Operations & Production Management* 36 (4):358-83.
- Cui, Ruomeng, Santiago Gallino, Antonio Moreno, and Dennis J. Zhang. 2018. "The Operational Value of Social Media Information." *Production and Operations Management* 27 (10):1749-69.
- Dia, Hussein. 2001. "An object-oriented neural network approach to short-term traffic forecasting." *European Journal of Operational Research* 131 (2):253-61.
- Díaz, Henry Lamos, Karin Aguilar Imitola, and Rolando José Acosta Amado. 2019. "OR/MS research perspectives in disaster operations management: a literature review." *Revista Facultad de Ingeniería Universidad de Antioquia* (91):43-59.
- Ding, Xiaomei, Martin L. Puterman, and Arnab Bisi. 2002. "The Censored Newsvendor and the Optimal

- Acquisition of Information." *Operations Research* 50 (3):517-27.
- Donkor, Emmanuel A., Thomas A. Mazzuchi, Refik Soyer, and J. Alan Roberson. 2014. "Urban Water Demand Forecasting: Review of Methods and Models." *Journal of Water Resources Planning and Management* 140 (2):146-59.
- Dubey, Rameshwar, Angappa Gunasekaran, Stephen J. Childe, Samuel Fosso Wamba, and Thanos Papadopoulos. 2016. "The impact of big data on world-class sustainable manufacturing." *The International Journal of Advanced Manufacturing Technology* 84 (1):631-45.
- Dunn, Thomas J., Thom Baguley, and Vivienne Brunsden. 2014. "From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation." *British Journal of Psychology* 105 (3):399-412.
- Durach, Christian F., Joakim Kembro, and Andreas Wieland. 2017. "A New Paradigm for Systematic Literature Reviews in Supply Chain Management." *Journal of Supply Chain Management* 53 (4):67-85.
- Dutta, Debprotim, and Indranil Bose. 2015. "Managing a Big Data project: The case of Ramco Cements Limited." *International Journal of Production Economics* 165:293-306.
- Elmaghraby, Wedad, and Pinar Keskinocak. 2003. "Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions." *Management Science* 49 (10):1287-309.
- Fahimnia, Behnam, Mehrdokht Pournader, Enno Siemsen, Elliot Bendoly, and Charles Wang. 2019. "Behavioral Operations and Supply Chain Management—A Review and Literature Mapping." *Decision Sciences* 50 (6):1127-83.
- Fatt, Q, and Amutha Ramadas. 2018. "The Usefulness and Challenges of Big Data in Healthcare." *Journal of Healthcare Communications* 3 (2):21-5.
- Feng, Yunting, Qinghua Zhu, and Kee-Hung Lai. 2017. "Corporate social responsibility for supply chain management: A literature review and bibliometric analysis." *Journal of Cleaner Production* 158:296-307.
- Fernández-Caramés, Tiago M. , Oscar Blanco-Novoa, Iván Froiz-Miguez, and Paula Fraga-Lamas. 2019. "Towards an Autonomous Industry 4.0 Warehouse: A UAV and Blockchain-Based System for

- Inventory and Traceability Applications in Big Data-Driven Supply Chain Management." *Sensors* 19 (10):2394.
- Ferreira, Kris Johnson, Bin Hong Alex Lee, and David Simchi-Levi. 2016. "Analytics for an Online Retailer: Demand Forecasting and Price Optimization." *Manufacturing & Service Operations Management* 18 (1):69-88.
- Fornell, Claes, and David F. Larcker. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18 (1):39-50.
- Fu, Wenhan, and Chen-Fu Chien. 2019. "UNISON data-driven intermittent demand forecast framework to empower supply chain resilience and an empirical study in electronics distribution." *Computers & Industrial Engineering* 135:940-9.
- Gattiker, Thomas F., and Diane H. Parente. 2007. "Introduction to the special issue on innovative data sources for empirically building and validating theories in Operations Management." *Journal of Operations Management* 25 (5):957-61.
- Gawankar, Shradha A., Angappa Gunasekaran, and Sachin Kamble. 2020. "A study on investments in the big data-driven supply chain, performance measures and organisational performance in Indian retail 4.0 context." *International Journal of Production Research* 58 (5):1574-93.
- Ghosh, Bidisha, Biswajit Basu, and Margaret O'Mahony. 2009. "Multivariate short-term traffic flow forecasting using time-series analysis." *IEEE Transactions on Intelligent Transportation Systems* 10 (2):246-54.
- Giannakis, Mihalis, and Michalis Louis. 2016. "A multi-agent based system with big data processing for enhanced supply chain agility." *Journal of Enterprise Information Management* 29 (5):706-27.
- Glock, Christoph H., Eric H. Grosse, Mohamad Y. Jaber, and Timothy L. Smunt. 2019. "Applications of learning curves in production and operations management: A systematic literature review." *Computers & Industrial Engineering* 131:422-41.
- Godfrey, Gregory A., and Warren B. Powell. 2001. "An Adaptive, Distribution-Free Algorithm for the Newsvendor Problem with Censored Demands, with Applications to Inventory and Distribution." *Management Science* 47 (8):1101-12.
- Gunasekaran, Angappa, Thanos Papadopoulos, Rameshwar Dubey, Samuel Fosso Wamba, Stephen J.

- Childe, Benjamin Hazen, and Shahriar Akter. 2017. "Big data and predictive analytics for supply chain and organizational performance." *Journal of Business Research* 70:308-17.
- Halldórsson, Árni, Juliana Hsuan, and Herbert Kotzab. 2015. "Complementary theories to supply chain management revisited – from borrowing theories to theorizing." *Supply Chain Management: An International Journal* 20 (6):574-86.
- Harding, J. A., M. Shahbaz, Srinivas, and A. Kusiak. 2005. "Data Mining in Manufacturing: A Review." *Journal of Manufacturing Science and Engineering* 128 (4):969-76.
- Hazen, Benjamin T., Christopher A. Boone, Jeremy D. Ezell, and L. Allison Jones-Farmer. 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.
- Hazen, Benjamin T., Joseph B. Skipper, Christopher A. Boone, and Raymond R. Hill. 2018. "Back in business: operations research in support of big data analytics for operations and supply chain management." *Annals of Operations Research* 270 (1):201-11.
- He, Yong, Hongfu Huang, Dong Li, Chunming Shi, and Sarah J. Wu. 2018. "Quality and Operations Management in Food Supply Chains: A Literature Review." *Journal of Food Quality* 2018:7279491.
- Heineke, Janelle, and Mark M. Davis. 2007. "The emergence of service operations management as an academic discipline." *Journal of Operations Management* 25 (2):364-74.
- Hippert, Henrique Steinherz, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. 2001. "Neural networks for short-term load forecasting: A review and evaluation." *IEEE Transactions on Power Systems* 16 (1):44-55.
- Ho, William, Xiaowei Xu, and Prasanta K. Dey. 2010. "Multi-criteria decision making approaches for supplier evaluation and selection: A literature review." *European Journal of Operational Research* 202 (1):16-24.
- Hofmann, Erik. 2017. "Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect." *International Journal of Production Research* 55 (17):5108-26.
- Hong, Tao, and Shu Fan. 2016. "Probabilistic electric load forecasting: A tutorial review." *International Journal of Forecasting* 32 (3):914-38.

- Hosseini, Seyedmohsen, Dmitry Ivanov, and Alexandre Dolgui. 2019. "Review of quantitative methods for supply chain resilience analysis." *Transportation Research Part E: Logistics and Transportation Review* 125:285-307.
- Huang, George Q., Y. F. Zhang, X. Chen, and Stephen T. Newman. 2008. "RFID-enabled real-time wireless manufacturing for adaptive assembly planning and control." *Journal of Intelligent Manufacturing* 19 (6):701-13.
- Huang, George Q., Y. F. Zhang, and P. Y. Jiang. 2007. "RFID-based wireless manufacturing for walking-worker assembly islands with fixed-position layouts." *Robotics and Computer-Integrated Manufacturing* 23 (4):469-77.
- . 2008. "RFID-based wireless manufacturing for real-time management of job shop WIP inventories." *The International Journal of Advanced Manufacturing Technology* 36 (7):752-64.
- Huang, Tingliang, and Jan A. Van Mieghem. 2014. "Clickstream Data and Inventory Management: Model and Empirical Analysis." *Production and Operations Management* 23 (3):333-47.
- Huh, Woonghee Tim, Retsef Levi, Paat Rusmevichientong, and James B. Orlin. 2011. "Adaptive Data-Driven Inventory Control with Censored Demand Based on Kaplan-Meier Estimator." *Operations Research* 59 (4):929-41.
- Huh, Woonghee Tim, and Paat Rusmevichientong. 2009. "A Nonparametric Asymptotic Analysis of Inventory Planning with Censored Demand." *Mathematics of Operations Research* 34 (1):103-23.
- Ikeziri, Lucas Martins, Fernando Bernardi de Souza, Mahesh C. Gupta, and Paula de Camargo Fiorini. 2019. "Theory of constraints: review and bibliometric analysis." *International Journal of Production Research* 57 (15-16):5068-102.
- Irfan, Muhammad, and Mingzheng Wang. 2019. "Data-driven capabilities, supply chain integration and competitive performance." *British Food Journal* 121 (11):2708-29.
- Ishak, Sherif, and Haitham Al-Deek. 2002. "Performance Evaluation of Short-Term Time-Series Traffic Prediction Model." *Journal of Transportation Engineering* 128 (6):490-8.
- Jiang, Xiaomo, and Hojjat Adeli. 2004. "Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis." *Computer-Aided Civil and Infrastructure Engineering* 19 (5):324-37.
- Jun, H. B., J. H. Shin, Y. S. Kim, D. Kiritsis, and P. Xirouchakis. 2009. "A framework for RFID

- applications in product lifecycle management." *International Journal of Computer Integrated Manufacturing* 22 (7):595-615.
- Kamble, Sachin S., Angappa Gunasekaran, and Shradha A. Gawankar. 2020. "Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications." *International Journal of Production Economics* 219:179-94.
- Karlaftis, M. G., and E. I. Vlahogianni. 2011. "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights." *Transportation Research Part C: Emerging Technologies* 19 (3):387-99.
- Karttunen, Elina. 2018. "Purchasing and supply management skills revisited: an extensive literature review." *Benchmarking: An International Journal* 25 (9):3906-34.
- Khan, Imran, Zongwei Luo, Joshua Zhexue Huang, and Waseem Shahzad. 2020. "Variable Weighting in Fuzzy k-Means Clustering to Determine the Number of Clusters." *IEEE Transactions on Knowledge and Data Engineering* 32 (9):1838-53.
- Kumar, Ajay, Ravi Shankar, Alok Choudhary, and Lakshman S. Thakur. 2016. "A big data MapReduce framework for fault diagnosis in cloud-based manufacturing." *International Journal of Production Research* 54 (23):7060-73.
- Kumar, Ajay, Ravi Shankar, and Lakshman S. Thakur. 2018. "A big data driven sustainable manufacturing framework for condition-based maintenance prediction." *Journal of Computational Science* 27:428-39.
- Kusiak, Andrew. 2000. "Decomposition in data mining: an industrial case study." *IEEE Transactions on Electronics Packaging Manufacturing* 23 (4):345-53.
- . 2001. "Rough set theory: a data mining tool for semiconductor manufacturing." *IEEE Transactions on Electronics Packaging Manufacturing* 24 (1):44-50.
- Kusiak, Andrew, and Christian Kurasek. 2001. "Data mining of printed-circuit board defects." *IEEE Transactions on Robotics and Automation* 17 (2):191-6.
- Lai, Yuanyuan, Huifen Sun, and Jifan Ren. 2018. "Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation." *The International Journal of Logistics Management* 29 (2):676-703.

- Lee, Sangmin, Younghoon Kim, Hyungu Kahng, Soon-Kyo Lee, Seokhyun Chung, Taesu Cheong, Keeyong Shin, Jeehyuk Park, and Seoung Bum Kim. 2020. "Intelligent traffic control for autonomous vehicle systems based on machine learning." *Expert Systems With Applications* 144:113074.
- Levi, Retsef, Georgia Perakis, and Joline Uichanco. 2015. "The Data-Driven Newsvendor Problem: New Bounds and Insights." *Operations Research* 63 (6):1294-306.
- Levi, Retsef, Robin O. Roundy, and David B. Shmoys. 2007. "Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models." *Mathematics of Operations Research* 32 (4):821-39.
- Liu, Weihua, Enze Bai, Liwei Liu, and Wanying Wei. 2017. "A Framework of Sustainable Service Supply Chain Management: A Literature Review and Research Agenda." *Sustainability* 9 (3):421.
- Liyanage, Liwan H., and J. George Shanthikumar. 2005. "A practical inventory control policy using operational statistics." *Operations Research Letters* 33 (4):341-8.
- Long, Qingqi. 2018. "Data-driven decision making for supply chain networks with agent-based computational experiment." *Knowledge-Based Systems* 141:55-66.
- Lv, Yisheng, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. 2015. "Traffic Flow Prediction With Big Data: A Deep Learning Approach." *IEEE Transactions on Intelligent Transportation Systems* 16 (2):865-73.
- Mao, Chao, and Zuojun Shen. 2018. "A reinforcement learning framework for the adaptive routing problem in stochastic time-dependent network." *Transportation Research Part C: Emerging Technologies* 93:179-97.
- Martins, C. L., and M. V. Pato. 2019. "Supply chain sustainability: A tertiary literature review." *Journal of Cleaner Production* 225:995-1016.
- Mascarenhas, Carla, João J. Ferreira, and Carla Marques. 2018. "University–industry cooperation: A systematic literature review and research agenda." *Science and Public Policy* 45 (5):708-18.
- Medina-González, Sergio, Ahmed Shokry, Javier Silvente, Gicela Lupera, and Antonio España. 2020. "Optimal management of bio-based energy supply chains under parametric uncertainty through a data-driven decision-support framework." *Computers & Industrial Engineering* 139:105561.
- Megahed, Fadel M., and L. Allison Jones-Farmer. 2015. "Statistical Perspectives on “Big Data”." In

- Frontiers in Statistical Quality Control 11*, edited by Sven Knoth and Wolfgang Schmid, 29-47. Cham: Springer International Publishing.
- Min, Wanli, and Laura Wynter. 2011. "Real-time road traffic prediction with spatio-temporal correlations." *Transportation Research Part C: Emerging Technologies* 19 (4):606-16.
- Narasimhan, Ram. 2014. "Theory Development in Operations Management: Extending the Frontiers of a Mature Discipline via Qualitative Research." *Decision Sciences* 45 (2):209-27.
- Nguyen, Truong, Li Zhou, Virginia Spiegler, Petros Ieromonachou, and Yong Lin. 2018. "Big data analytics in supply chain management: A state-of-the-art literature review." *Computers and Operations Research* 98:254-64.
- Opresnik, David, and Marco Taisch. 2015. "The value of Big Data in servitization." *International Journal of Production Economics* 165:174-84.
- Ozgormus, Elif, and Alice E. Smith. 2020. "A data-driven approach to grocery store block layout." *Computers & Industrial Engineering* 139:105562.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825-30.
- Persson, Olle, Rickard Danell, and J Wiborg Schneider. 2009. "How to use Bibexcel for various types of bibliometric analysis." *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday* 5:9-24.
- Provost, Foster, and Tom Fawcett. 2013. "Data Science and its Relationship to Big Data and Data-Driven Decision Making." *Big Data* 1 (1):51-9.
- Qu, Li, Jianming Hu, Li Li, and Yi Zhang. 2009. "PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach." *IEEE Transactions on Intelligent Transportation Systems* 10 (3):512-22.
- Quek, Chai, Michel Pasquier, and Bernard Boon Seng Lim. 2006. "POP-TRAFFIC: a novel fuzzy neural approach to road traffic analysis and prediction." *IEEE Transactions on Intelligent Transportation Systems* 7 (2):133-46.
- Raghupathi, Wullianallur, and Viju Raghupathi. 2014. "Big data analytics in healthcare: promise and

- potential." *Health Information Science and Systems* 2 (1):3.
- Ramos-Rodríguez, Antonio-Rafael, and José Ruíz-Navarro. 2004. "Changes in the intellectual structure of strategic management research: A bibliometric study of the strategic management journal, 1980-2000." *Strategic Management Journal* 25 (10):981-1004.
- Rebs, Tobias, Marcus Brandenburg, and Stefan Seuring. 2019. "System dynamics modeling for sustainable supply chain management: A literature review and systems thinking approach." *Journal of Cleaner Production* 208:1265-80.
- Ren, Shuyun, and Tsan-Ming Choi. 2016. "Selection and industrial applications of panel data based demand forecasting models." *Industrial Management & Data Systems* 116 (6):1131-59.
- Sachs, Anna-Lena, and Stefan Minner. 2014. "The data-driven newsvendor with censored demand observations." *International Journal of Production Economics* 149:28-36.
- Schoenherr, Tobias, and Cheri Speier-Pero. 2015. "Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential." *Journal of Business Logistics* 36 (1):120-32.
- Sethi, Vijay, and William R. King. 1994. "Development of Measures to Assess the Extent to Which an Information Technology Application Provides Competitive Advantage." *Management Science* 40 (12):1601-27.
- Sheffi, Yossi. 2015. "Preparing for Disruptions Through Early Detection." *MIT Sloan Management Review* 57 (1):31-42.
- Small, Henry. 1973. "Co-citation in the scientific literature: A new measure of the relationship between two documents." *Journal of the American Society for Information Science* 24 (4):265-9.
- Smith, Brian L., and R. Keith Oswald. 2003. "Meeting Real-Time Traffic Flow Forecasting Requirements with Imprecise Computations." *Computer-Aided Civil and Infrastructure Engineering* 18 (3):201-13.
- Smith, Brian L., Billy M. Williams, and R. Keith Oswald. 2002. "Comparison of parametric and nonparametric models for traffic flow forecasting." *Transportation Research Part C: Emerging Technologies* 10 (4):303-21.
- Song, Peijian, Chengde Zheng, Cheng Zhang, and Xiaofeng Yu. 2018. "Data analytics and firm

- performance: An empirical study in an online B2C platform." *Information & Management* 55 (5):633-42.
- Srinivasan, Ravi, and Morgan Swink. 2018. "An Investigation of Visibility and Flexibility as Complements to Supply Chain Analytics: An Organizational Information Processing Theory Perspective." *Production and Operations Management* 27 (10):1849-67.
- Stathopoulos, Anthony, and Matthew G. Karlaftis. 2003. "A multivariate state space approach for urban traffic flow modeling and prediction." *Transportation Research Part C: Emerging Technologies* 11 (2):121-35.
- Tan, Kim Hua, YuanZhu Zhan, Guojun Ji, Fei Ye, and Chingter Chang. 2015. "Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph." *International Journal of Production Economics* 165:223-33.
- Tan, Man-Chun, Sze Chun Wong, Jian-Min Xu, Zhan-Rong Guan, and Peng Zhang. 2009. "An aggregation approach to short-term traffic flow prediction." *IEEE Transactions on Intelligent Transportation Systems* 10 (1):60-9.
- Tang, Yujie, Nan Cheng, Wen Wu, Miao Wang, Yanpeng Dai, and Xuemin Shen. 2019. "Delay-Minimization Routing for Heterogeneous VANETs With Machine Learning Based Mobility Prediction." *IEEE Transactions on Vehicular Technology* 68 (4):3967-79.
- Tao, Fei, Ying Cheng, Li Da Xu, Lin Zhang, and Bo Hu Li. 2014. "CCIoT-CMfg: cloud computing and internet of things-based cloud manufacturing service system." *IEEE Transactions on Industrial Informatics* 10 (2):1435-42.
- Tao, Fei, Ying Zuo, Li Da Xu, Lin Lv, and Lin Zhang. 2014. "Internet of Things and BOM-Based Life Cycle Assessment of Energy-Saving and Emission-Reduction of Products." *IEEE Transactions on Industrial Informatics* 10 (2):1252-61.
- Tao, Fei, Ying Zuo, Li Da Xu, and Lin Zhang. 2014. "IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing." *IEEE Transactions on Industrial Informatics* 10 (2):1547-57.
- Taylor, James W. 2003. "Short-term electricity demand forecasting using double seasonal exponential smoothing." *Journal of the Operational Research Society* 54 (8):799-805.

- Taylor, James W., and Roberto Buizza. 2003. "Using weather ensemble predictions in electricity demand forecasting." *International Journal of Forecasting* 19 (1):57-70.
- Taylor, James W., and S. Majithia. 2000. "Using combined forecasts with changing weights for electricity demand profiling." *Journal of the Operational Research Society* 51 (1):72-82.
- Tiwari, Sunil, H. M. Wee, and Yosef Daryanto. 2018. "Big data analytics in supply chain management between 2010 and 2016: Insights to industries." *Computers & Industrial Engineering* 115:319-30.
- Trkman, Peter, Kevin McCormack, Marcos Paulo Valadares de Oliveira, and Marcelo Bronzo Ladeira. 2010. "The impact of business analytics on supply chain performance." *Decision Support Systems* 49 (3):318-27.
- Tseng, Ming-Lang, Kuo-Jui Wu, Ming K. Lim, and Wai-Peng Wong. 2019. "Data-driven sustainable supply chain management performance: A hierarchical structure assessment under uncertainties." *Journal of Cleaner Production* 227:760-71.
- Uzsoy, Reha, John W. Fowler, and Lars Mönch. 2018. "A survey of semiconductor supply chain models Part II: demand planning, inventory management, and capacity planning." *International Journal of Production Research* 56 (13):4546-64.
- van Eck, Nees Jan, and Ludo Waltman. 2010. "Software survey: VOSviewer, a computer program for bibliometric mapping." *Scientometrics: an international journal for all quantitative aspects of the science of science, communication in science and science policy* 84 (2):523-38.
- van Eck, Nees Jan, Ludo Waltman, Rommert Dekker, and Jan van den Berg. 2010. "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS." *Journal of the American Society for Information Science and Technology* 61 (12):2405-16.
- Vlahogianni, Eleni I., John C. Golias, and Matthew G. Karlaftis. 2004. "Short-term traffic forecasting: Overview of objectives and methods." *Transport Reviews* 24 (5):533-57.
- Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. 2005. "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach." *Transportation Research Part C: Emerging Technologies* 13 (3):211-34.
- Wamba, Samuel Fosso, Shahriar Akter, Andrew Edwards, Geoffrey Chopin, and Denis Gnanzou. 2015. "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case

- study." *International Journal of Production Economics* 165:234-46.
- Wang, Gang, Angappa Gunasekaran, Eric W. T. Ngai, and Thanos Papadopoulos. 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, Jinghua, Jianyi Liu, and Cong Wang. 2007. Keyword extraction based on pagerank. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- Wang, Nianxin, Huigang Liang, Yu Jia, Shilun Ge, Yajiong Xue, and Zhining Wang. 2016. "Cloud computing research in the IS discipline: A citation/co-citation analysis." *Decision Support Systems* 86:35-47.
- Williams, Brent D., Joseph Roh, Travis Tokar, and Morgan Swink. 2013. "Leveraging supply chain visibility for responsiveness: The moderating role of internal integration." *Journal of Operations Management* 31 (7):543-54.
- Wu, Chun-Hsin, Jan-Ming Ho, and D. T. Lee. 2004. "Travel-time prediction with support vector regression." *IEEE Transactions on Intelligent Transportation Systems* 5 (4):276-81.
- Wu, Kuo-Jui, Ching-Jong Liao, Ming-Lang Tseng, Ming K. Lim, Jiayao Hu, and Kimhua Tan. 2017. "Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties." *Journal of Cleaner Production* 142:663-76.
- Xu, Li Da. 2011. "Information architecture for supply chain quality management." *International Journal of Production Research* 49 (1):183-98.
- Xu, Li Da, Wu He, and Shancang Li. 2014. "Internet of Things in Industries: A Survey." *IEEE Transactions on Industrial Informatics* 10 (4):2233-43.
- Xu, Xinhan, Xiangfeng Chen, Fu Jia, Steve Brown, Yu Gong, and Yifan Xu. 2018. "Supply chain finance: A systematic literature review and bibliometric analysis." *International Journal of Production Economics* 204:160-73.
- Xu, Xun. 2012. "From cloud computing to cloud manufacturing." *Robotics and Computer-Integrated Manufacturing* 28 (1):75-86.
- Yin, Hongbin, S. C. Wong, Jianmin Xu, and C. K. Wong. 2002. "Urban traffic flow prediction using a fuzzy-neural approach." *Transportation Research Part C: Emerging Technologies* 10 (2):85-98.

- Yong, An Gie, and Sean Pearce. 2013. "A Beginners Guide to Factor Analysis: Focusing on Exploratory Factor Analysis." *Tutorials in Quantitative Methods for Psychology* 9 (2):79-94.
- Yu, Wantao, Roberto Chavez, Mark A. Jacobs, and Mengying Feng. 2018. "Data-driven supply chain capabilities and performance: A resource-based view." *Transportation Research Part E: Logistics and Transportation Review* 114:371-85.
- Yu, Wantao, Mark A. Jacobs, Roberto Chavez, and Mengying Feng. 2019. "Data-Driven Supply Chain Orientation and Financial Performance: The Moderating Effect of Innovation-Focused Complementary Assets." *British Journal of Management* 30 (2):299-314.
- Zhang, Junping, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. 2011. "Data-Driven Intelligent Transportation Systems: A Survey." *IEEE Transactions on Intelligent Transportation Systems* 12 (4):1624-39.
- Zhang, Xiunian, and Jasmine Siu Lee Lam. 2019. "A fuzzy Delphi-AHP-TOPSIS framework to identify barriers in big data analytics adoption: case of maritime organizations." *Maritime Policy & Management* 46 (7):781-801.
- Zhang, Yingfeng, George Q. Huang, T. Qu, Oscar K. Ho, and Shudong Sun. 2011. "Agent-based smart objects management system for real-time ubiquitous manufacturing." *Robotics and Computer-Integrated Manufacturing* 27 (3):538-49.
- Zhang, YingFeng, George Q. Huang, Ting Qu, and Oscar Ho. 2010. "Agent-based workflow management for RFID-enabled real-time reconfigurable manufacturing." *International Journal of Computer Integrated Manufacturing* 23 (2):101-12.
- Zhang, Yingfeng, Pingyu Jiang, and George Huang. 2008. "RFID-based smart kanbans for just-in-time manufacturing." *International Journal of Materials and Product Technology* 33 (1-2):170-84.
- Zhang, Yingfeng, T. Qu, Oscar K. Ho, and George Q. Huang. 2011. "Agent-based Smart Gateway for RFID-enabled real-time wireless manufacturing." *International Journal of Production Research* 49 (5):1337-52.
- Zhao, Hai-xiang, and Frédéric Magoulès. 2012. "A review on the prediction of building energy consumption." *Renewable and Sustainable Energy Reviews* 16 (6):3586-92.
- Zhao, Hongyan, Fenghua Zhang, and Jongwook Kwon. 2018. "Corporate social responsibility research

- in international business journals: An author co-citation analysis." *International Business Review* 27 (2):389-400.
- Zhao, Rui, Yiyun Liu, Ning Zhang, and Tao Huang. 2017. "An optimization model for green supply chain management by using a big data analytic approach." *Journal of Cleaner Production* 142:1085-97.
- Zheng, Weizhong, Der-Horng Lee, and Qixin Shi. 2006. "Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach." *Journal of Transportation Engineering* 132 (2):114-21.
- Zhong, Ray Y., George Q. Huang, Shulin Lan, Q. Y. Dai, Xu Chen, and T. Zhang. 2015. "A big data approach for logistics trajectory discovery from RFID-enabled production data." *International Journal of Production Economics* 165:260-72.
- Zhong, Ray Y., Stephen T. Newman, George Q. Huang, and Shulin Lan. 2016. "Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives." *Computers & Industrial Engineering* 101:572-91.
- Zhou, S. L., T. A. McMahon, A. Walton, and J. Lewis. 2000. "Forecasting daily urban water demand: a case study of Melbourne." *Journal of Hydrology* 236 (3):153-64.
- . 2002. "Forecasting operational demand for an urban water supply zone." *Journal of Hydrology* 259 (1):189-202.
- Zupic, Ivan, and Tomaž Čater. 2015. "Bibliometric Methods in Management and Organization." *Organizational Research Methods* 18 (3):429-72.