

BIG DATA IN EMISSION PRODUCING MANUFACTURING INDUSTRIES – AN EXPLORATIVE LITERATURE REVIEW

E. Hämäläinen¹, T. Inkinen^{1,*}

¹ Brahea-Centre, University of Turku, 20014-Finland - (esa.hamalainen, tommy.inkinen)@utu.fi

KEY WORDS: Big data, Emission tracking, Industry, Production

ABSTRACT:

Data management and intelligent systems provide new possibilities and trajectories for environmentally robust industrial production. Global climate change has evoked a number of fresh studies on data management, openness and environmental innovation. This literature review approaches current academic research focusing on big data, industry, and emissions. The paper is based on key word searches that returned publications from high-class scientific journals dedicated to this particular topic. The article reading illustrates that big data is utilised in various industries, and explores a large variety of polluting substances. The authors argue that innovative and insightful new ways of using big data provide tools for emission monitoring and clean technology utilisation. The diversity in the analysed articles proves the complexity of market operations and corporate responsibility. The paper concludes in addressing the need of using and combining data resources, particularly at the industry unit level, in order to develop more efficient tools for environmental monitoring and decision making.

1. INTRODUCTION

Technological advancements and the use of extensive (big) data sources provide research-based tools help in controlling and reducing industrial emission. The climate change report IPCC 2018 shows that knowledge-based action is needed (at all organisational levels), in order to steer societies and industrial cultures towards low emission driven mindsets in coming decades. This paper focuses on big data utilisation in industrial (environmental) research. It examines responsible production and emission control. Additionally, there is a vast array of technologies measuring and managing emission responsibility. This is also a requirement, since the majority of industries producing large emissions continuously adhere to the 24/7 principle. This also means that data volumes obtained from production (concerning inputs, processes, and outputs) are extensive. These data sources are quite often difficult to combine without considerable modifications.

Big data usually contains numerous content blocks from different fields of operations. Big data may serve as a basis for unveiling complicated links within a single production facility. Exposing all significant pollution sources helps finding solutions for responsible production. This should be a global environmental objective in all the fields of manufacturing industries. Low emission development and enhancement of cleaner production requires understanding a multitude of data sources, as well as enhanced information processing. Combining big data, emission measurement, and active (real-time) monitoring is the basis of generic corporate responsibility.

Steering the manufacturing industry towards sustainability requires well-structured, reliable, and valid data from the production environment (being researched). Otherwise, the path is 'blurry' and there will be a lack of vision. Data transferability and interoperability are crucial, the risk of partial optimisation increases significantly. This easily leads to

environmental projects with good intentions that do not manage to produce real input-output impacts (or interdependencies) in the actual production process. This forms the foundation of this paper using a carefully selected set of articles published in academic journals (Journal of Cleaner Production; Energy). The selected articles focus on the use of big data, intelligent systems, and reporting tools in emission control and the management of environmental impacts.

2. BIG DATA AS RESEARCH PLATFORM IN EMISSION SOURCE TRACKING

The foundations of big data applications can be traced back to the extensive increase of digital understanding in the 1980s. The development of diversified and versatile programming languages, improved programming know-how, increased usage of computers, logarithmically growing internet, advanced sensor measurement, development of scalable databases, emergence of global cloud services, continuous increase of computer efficiency, and finally understanding what digital applications can offer an organisation, inevitably led to the development of large digital data storages. Collectively these are called 'big data'. Presumably, Larson (1989) was among the first ones to introduce the term 'big data' in his article containing the following two sentences: "The keepers of big data say they do it for the consumer's benefit. But data have a way of being used for purposes other than originally intended." It is justifiable to claim that the origin of big data use is strongly associated with consumer behaviour analysis. A major point in Larson's comments was that data could be used in very different ways compared to the original intent. This data could also be combined with other databases, in order to create big data.

The public sector also has started to pay more attention to data management, security, and processing: For example, the EU (2016) observed that large amounts of data are being rapidly

* Corresponding author

produced and that the increasing understanding of big data requires recognition beyond metadata (properties), including data management and processing. This is augmented through functionality and the practice of interactive user interface design, cloud storage management, and data query properties. It finally results in decision support tools (systems). In this regard, Wang et al. (2016) point out that decisions require methodology, technology (hardware), and tools (software) to identify the best solution based on the assumption that the decision maker is rational or bounded rational. In this process, big data can offer a platform for exhaustive information management, but it requires refinement and adequate relation identification. This refined data functions as a base for reliable decision making in quickly changing operational environments.

There are numerous challenges connected to big data. One way of identifying them is to use V-challenges referring to ‘Volume’ (too large), ‘Variety’ (too extensive), ‘Velocity’ (too fast), ‘Veracity’ (too uncertain), and ‘Value’ (too difficult to evaluate) (see e.g. Gil and Song, 2016; Zhong et al., 2016). The validity of manufacturing process redesign, based on big data, is often difficult to measure. Especially in the case of long production and supply chains, adjustments may result in emission enhancements in some part (of the chain), but it is very difficult to evaluate the overall impact of this sub-process. This is a question of whether or not industrial production is viewed from the micro (bottom-up) or macro (top-down) perspective.

Big data opens new alternatives for increasing knowledge management of industrial production chains. According to Wong et al. (2016), utilisation of big data most likely increases the accuracy of decision making. Ala-Harja and Helo, 2015 note that there is no (real) conflict between sustainability and economic growth – the traditional trade-off in industrial production. Especially in complicated and polluting manufacturing processes, access to detailed process information offers a way to mitigate harmful emissions by selecting a new composition of raw materials, chemicals, and other substances. Hämäläinen and Inkinen (2017) note that, in a large industrial unit (such as a paper mill), big data offers an opportunity to combine production, economic, and emission data helping production and steering it into a more responsible direction.

2.1 Querying research papers

Big data applications are outcomes and manifestations of continuous digitalisation. This has allowed technological interdependencies involving embedded networks between different (technology) architectures, creating so called Internet of Things (IoT) applications and solutions for machine-to-machine interaction. These also have tremendous potential in emission measurement, monitoring, and indicator development. Correspondingly, after running initial queries in the ‘Science Direct’ database using separate key words (OR operator) ‘big data’, ‘emissions’ and ‘industry’, the results yielded thousands of published articles in numerous journals. After careful consideration, the authors decided to use only two key journals in this field: ‘Journal of Cleaner Production’ and ‘Energy’. They both particularly focus on the topics of emissions and industry. A selection criterion was that the amount of papers published in these journals was feasible and that their reputation within the field of science was high. This

is assumed in the Impact Factors (Clarivate analytics): ‘Energy’ Cite Score was 5.60 and Impact Factor 4.968 (2018) and the Journal of Cleaner Production’s Cite Score was 5.79 and Impact Factor 5.651 (2018).

Publication types were limited to include only ‘research’ and ‘review’ articles in order to maintain a homogenous scientific standard. The above selection ruled out extensive research. However, it increases focus and narrows down the point-of-view of the articles analysed. After reviewing a total of 1324 papers from the years 2016–2019, the authors used an AND operator with the key words producing an adequate number (29) of papers for the content analysis (Table 1).

Search words	Year	Article types
'Big data' 'Emissions' 'Industry'	2019 (50)	Review articles (73)
	2018 (529)	Research articles (1251)
	2017 (434)	
	2016 (311)	

Table 1. Search words, years, article types and publication titles.

The small segment of 29 papers is considered representative particularly in terms of big data related manufacturing production studies. The selection process of the final papers (representing only approximately 2.2% of all search word results) is fundamentally based on our qualitative consideration of the most suitable articles. The selected amount is also manageable to be treated qualitatively content wise as has been done here. Nonetheless, it is has to be stated that significant aspects may have been left out. In this regard, the papers studies are regarded as an insight into these complex phenomena. As an example, the utilisation of big data requires specialised algorithms enabling the reception of continuously valid and reliable data from the data storage. Furthermore, it is difficult to obtain access to real-life industrial big data storages that cover all the necessary information, such as emissions, manufacturing information, and raw materials (also see Hämäläinen and Inkinen, 2017), as they provide information of corporate core operations.

Table 1 indicates that the overall amount of publications is increasing annually. However, the number of papers explicitly concerning big data applications was rather low in comparison to emission-focused papers. Big data repositories and storages are considered as the fundamental resources and platforms also enabling new possibilities for complicated environmental research endeavours. This notion calls forth, especially in environmental research, large data (quantity) analyses that are suitable for discovering deep knowledge threads, weak signals, and interrelations existing in industrial processes.

Table 2 shows the studied articles sorted according to their publishing year including titles, key words, and abstracts. The search criteria returned articles covering an extensive selection of industry sectors including tourism, mining, farming, and manufacturing. The papers included 119 different key words, a majority of which concerning cleaner environment. Commonly the root key word ‘big data’ has extension forms or continuations such as ‘big data mining’ or ‘big data analytics’. In terms of terminology, ‘big data’ was also found in all 29

abstracts (but not as a separate key word). In total, 'big data' and its derivatives were present 25 times in key word listings. Similarly, 'sustainability' and its derivatives (such as 'sustainable operations' and 'sustainable production') were present in 12 key word listings. The papers also applied varying terminologies describing 'sustainability', which made straightforward queries difficult. The alternatives included terms such as 'environmental protection', 'environmental management' and 'healthy environment'. However, their (interpreted) semiotic contents are similar, making it justified to consider them as synonyms.

Interestingly, the word 'responsibility' is absent from the key word listings. Overall, it is argued here that in industrial management 'responsibility' is a clearly more suitable term (in comparison to 'sustainability'). They are directing and supporting lower and zero-level emission objectives in production firms. Another significant finding is that the word 'solution' was not present in the articles. This is also an indication of the fact that not all 29 papers are oriented toward problem solving to begin with. The articles did, however, yield more than pure descriptions of the existing production conditions. This could be interpreted as potential for improving data management impact on emission control in different types of manufacturing environments.

Various methodologies were used in the big data treatment and analyses. The most common ones included, for example, Delphi, Data envelopment analysis (DEA), Decoupling analysis, Dual model, Logarithmic mean Divisia index (LMDI), and the slacks-based measure (SBM) model. These various analysis methods showed that there are several ways to process big data in the environmental context. These processing methods are applicable in data-based decision making. In addition to methodologies, the geographical location of the published papers is also relevant and interesting: more than 70% of the included research papers focus on pollution challenges in China. These papers also applied big data sets in their analyses evidencing the significance of cleaner production in the major cities of China where air pollution is an acute risk.

3. RESULTS: DISCUSSION AND INTERPRETATION

The included papers are relatively new, which is why they have not been cited very often. The most cited paper (as of December 27 2018) was composed by Zhang et al. (2017). This paper links big data analytics with cleaner manufacturing and maintenance processes. Cleaner Production (CP) is considered one of the most important means for manufacturing enterprises to achieve sustainable production and improve their sustainable competitive advantage. Zhang et al. (2017) observed that the implementation of the CP strategy was facing obstacles, such as the lack of complete data and valuable knowledge. With extensive use of smart sensing devices in product lifecycle management, a large amount of real-time and multi-source lifecycle big data can now be collected (also Hämäläinen & Inkinen, 2017; Hämäläinen et al. 2017).

Big data exposes the complexity of interlinking between different data producing units needed in the pursuit of sustainable low-level emission production. The analysed big data research does not, however, directly tackle the question of how to reduce emissions significantly. A comprehensive review

of big data analytics throughout the product lifecycle is essential in the support of sustainable smart manufacturing. This requires recognising the following segments of the analytical process: framework building; challenge definition within the framework; and an analytical vision for the future.

The idea of 'smart manufacturing' is assessed by Ren et al. (2019) and incorporated with big data analytics. They highlight the significance of hidden relations that are not easily observed through straightforward description. These relations include interactions between lifecycle decisions and process parameters supporting industrial management. These relations also bring forth the decision support system approach necessary in complex management systems. According to the literature, big data analytics and smart manufacturing are individually researched in academia and industry. Increasing the co-operation with industry and academia could support even better decision making in industrial lifecycle-based studies (Ren et al. 2019). Similarly, Kumar et al. (2018) designed a sustainable robust layout using the big data approach, which they call a key towards 'industry 4.0'. This layout is based on large-scale (big) data, and the writers argue that it is more efficient and effective in today's competitive market, since the robust and sustainable industry layout design is capable of handling the variations. On the contrary, inferior layout designs decrease the precision and accuracy of product manufacturing.

The analysed articles also recognise societal aspects. For example, Song et al. (2017) examined how big data could support societal development and environmental sustainability. The writers argue that there should be increased collaboration between researchers from different backgrounds and contexts. On the other hand, Roman et al. (2018) researched, in Brazil, business opportunities and challenges as two sides of climate change, and how to utilise big data in management on the path to a low carbon society. Climate change represents a long-term risk to society, and affects several industrial sectors.

In his review, Seele (2018) addresses how to transfer the potential of big-data-driven 'predictive policing' to corporate sustainability management. Big data (and data mining algorithms) also have the potential to promote sustainable development. This would be based on rigorous data management leading toward the predictive identification of potentially unsustainable developments or events. A major driver in enhancing responsible and low emission production in industries is the public's opinion and the negative public image projected by non-environmentally friendly polluting production.

An interesting topic from China is presented by Zhang and Chen (2017) who examine regional sustainability characteristics, including environmental efficiency. Their results indicated that several regions in the 'Poyang Lake Eco-Economics Zone' are not performing well in terms of environmental efficiency, and that they have a potential for improving (31.9% on average). According to their environmental efficiency analysis of China's regional industry, Chen and Jia (2017) also came to the non-surprising conclusion that the environmental efficiency of the Chinese industry is generally low. Their research is based on industrial big data from 2008 to 2012.

Main content category/ year	Title, Authors, Country and Citation index (as 18.2.2019)	Main objectives of the articles (retrieved from the abstracts)
Business 2017	A big data study on emitting companies' performance in the first two phases of the European Union Emission Trading Scheme. Liu et al. (2017), China, Cit. Index 8	As CO2 emissions are quantified by allowances and traded in markets, wise trading strategies will bring emitting companies higher profits or lower costs. Based on the big data of Community Independent Transaction Log (CITL), this article hereby presents a micro study on the emitting companies' efforts in increasing profits and saving costs during the allowances trading in the first two phases of the European Union Emission Trading Scheme (EU ETS).
Business 2017	Predictive Sustainability Control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management. Seele, P. (2018), Switzerland, Cit. Index 4	The article reviews first efforts to utilise big data for promoting sustainability and for reducing harm. By analogy, the concept of "predictive policing" is identified to be transferred to a concept called "Predictive Sustainability Control". The question of governance, data protection is discussed critically and future research avenues for theory advancement and empirical testing are presented. Big data and algorithms have also for promoting sustainable development based on rigorous data management leading to the predictive identification of likely unsustainable events.
Business 2017	Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for big data management towards a low carbon society. Roman et al. (2018), Brazil, Cit. Index 5	It is important to understand the implications that big data will have for managing corporate responses to the opportunities and challenges presented by climate change. The review showed that the opportunities emerging from the climate crisis tend to improve performance (environmental, financial and operational) as well as generating new business, while the challenges are related to the lack, or inefficiency, of governmental and regulatory support and an increase in costs.
Business 2018	Examining the role of big data and predictive analytics on collaborative performance in context to sustainable consumption and production behaviour. Dubey et al. (2018), France, Cit. Index 1	The organisations engaged in sustainable development programs are increasingly paying serious attention towards synergetic relationships between focal firms and their partners to achieve the goal of sustainable consumption and production (SCP) via big data and predictive analytics (BDPA). The study examines the role of BDPA in collaborative performance (CP) among the partners engaged in sustainable development program to achieve the goal of SCP.
Decision support 2017	A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products, Zhang et al. (2017), China, cit. ind. 60	Implementation of the Cleaner production (CP) strategy was facing barriers, such as the lack of complete data and valuable knowledge. With the wide use of smart sensing devices in product lifecycle management, a large amount of real-time and multi-source lifecycle big data can now be collected.
Decision support 2017	Allocation of carbon dioxide emission permits with the minimum cost for Chinese provinces in big data environment. An et al. (2017), China, Cit. Index 21	A new data envelopment analysis approach is proposed to evaluate the efficiency of decision making units in a big data environment and set the carbon dioxide emission (CDE) permits for each decision making unit with the minimum costs. The results show that the approach can well allocate the CDE permits with the minimum cost, and reveal the relationship between the marginal cost of reducing CDE permits and the adjustment on them.
Decision support 2017	A framework for Big Data driven product lifecycle management. Zhang et al. (2017), China, Cit. Index 21	Manufacturing enterprises need a more advanced analysis approach to develop a solution on a strategic level from using such lifecycle Big Data. The application of Big Data in lifecycle faces challenges, such as the lack of reliable data and valuable knowledge. The key findings were summarised as managerial implications, which can guide manufacturers to ensure improvements in energy saving and fault diagnosis related decisions in the whole lifecycle.
Decision support 2017	Management of urban land expansion in China through intensity assessment: A big data perspective, Zeng et al. (2017), China, Cit. Index 6	Big data is a platform for multi-source data fusion that helps to create spatially explicit decisions in regulating urban land expansion. In this study, we use big data to assess the intensity of urban land use in the metropolitan areas of China. Results show that Guangzhou has the most efficient urban land use system, followed by Shanghai and Shenzhen, and that Suzhou has the lowest urban land intensity.
Decision support 2018	A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions. Ren et al. (2019), China, Cit. Index 0	In smart manufacturing, big data analytics can uncover hidden knowledge and other useful information like relations between lifecycle decisions and process parameters helping industrial leaders to make more-informed business decisions in complex management environments. According to the literature, big data analytics and smart manufacturing were individually researched in academia and industry. After a comprehensive overview of big data in smart manufacturing, a conceptual framework was proposed from the perspective of product lifecycle.

Decision support 2018	An optimal renewable energy management strategy with and without hydropower using a factor weighted multi-criteria decision analysis and nationwide big data. Ifaei et al. (2018), Iran, Cit. Index 0	In the present study, an optimal renewable energy management strategy (REMS) is proposed as a sustainable policy for non-fossil fuel-based centralistic energy policies. Accordingly, the demand load data and weather big datasets were analysed using neural networks and analysis of variance, respectively. A factor analysis (FA) was performed using a big matrix consisting of 13 social, economic, environmental, and technical variables in all of Iran's provinces.
Energy efficiency 2018	A big data driven analytical framework for energy-intensive manufacturing industries. Zhang et al. (2017), China, Cit. Index 3	Energy-intensive industries account for almost 51% of energy consumption in China. There is a lack of manufacturing data due to the difficult implementation of sensors in harsh production environment. Two key technologies, namely energy big data acquisition and energy big data mining, are utilised to implement energy big data analytics. The results show that the energy consumption and energy costs are reduced by 3% and 4% respectively.
Energy efficiency 2018	Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. Liu et al. (2018) China, Cit. Index 2	This paper reviews the research progress on big data analysis and industrial energy efficiency evaluation and focuses on the energy efficiency evaluation methods based on energy consumption process analysis and big data mining approach. Multi-dimensional association rules are used to find the correlation of processes, companies and energy efficiency to guide the energy conservation in regional energy monitor. The results generated from advance analytics techniques applied specifically to solve regional energy efficiency problems.
Methodology 2017	Allocation of emission permits in large data sets: a robust multi-criteria approach. Ji et al. (2017), China, Cit. Index 19	This paper addressed the issue of the allocation of emission permits (AEP) in large data sets, with the goal of providing government strategies to practically operate the AEP in a group of organisations, and realise economic, social and environmental goals at the same time. The authors propose a robust multi-criteria AEP approach, together with its tractable algorithm, by extending the classical theory of data envelopment analysis (DEA) for large data sets.
Methodology 2017	Environmental efficiency analysis of China's regional industry: a data envelopment analysis (DEA) based approach, Chen, L. and Jia, G. (2017), China, Cit. In.43	Data envelopment analysis (DEA) method is incorporated and the slacks-based measure (SBM) model considering undesirable outputs is introduced to measure the environmental efficiency of different regions. Big data theory is introduced in the collection and selection of the input and output data of the regions. Further, the paper evaluates the environmental efficiencies of China's industry using data from 2008 to 2012. The results show that the environmental efficiencies of China's industry are generally low.
Resources 2017	China's regional natural resource allocation and utilisation - DEA-based approach in a big data environment. Zhu et al. (2017), China, Cit. Index 21	Emerged in 1980s, "big data" has developed very quickly in recent years, bringing new perspectives and opportunities for all kinds of academic fields, especially in DEA since it is a research highly based on data. Our empirical study shows that most provincial regions perform well in terms of efficiency in utilising natural resources, but that is based on huge consumption of natural resources.
Smart city 2018	Smart city with Chinese characteristics against the background of big data: Idea, action and risk. Wu et al. (2017), China, Cit. Index 4	Chinese urbanisation has generated urban problems, e.g., environmental pollution, resources shortage and traffic jam. This paper aims to put forward a development framework of smart city with Chinese characteristics against the background of big data. Paper also investigates the risks embedded in development of smart city with Chinese characteristics, e.g., information safety, weak emergency responding capacity and poor independent R&D capacity of core technology.
Supply chain 2017	Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties, Wu et al. (2017), Taiwan, Cit. Index 35	Rapid market changes aimed at sustainability have led to supply chain risks and uncertainties in the Taiwanese light-emitting diode industry. This study aggregates the various data to undertake an extensive investigation of supply chain risks. Specifically, this study proposes using the fuzzy and grey Delphi methods to identify a set of reliable attributes and, based on these attributes, transforming big data to a manageable scale.
Supply chain 2017	Pricing policies of green supply chain considering targeted advertising and product green degree in the Big Data environment. Liu, P. and Yi, S. (2017), China, Cit. Index 10	In the Big Data era, targeted advertising develops to a more precise direction, and to improve their marketing accuracy and success rate, many enterprises begin to use it. For a green supply chain, products green degree is also an important factor in influencing sale. Results indicated that the optimal retail price and the wholesale price had a negative correlation with the green degree and the input level of targeted advertising.
Supply chain 2018	Big data cloud computing framework for low carbon supplier selection in the beef supply chain. Singh et a. (2018) United Kingdom, Cit. Index 0	Economic development of nations across the globe, there is proportionate increment in corresponding carbon footprint. They have neglecting significant contributor of global emissions-agricultural industry and majority of its emission are generated in beef farms. There is need to raise awareness among farmers and thereby incorporate carbon footprint as a major cattle supplier selection attribute by abattoir and integrate it as a standard practice in procurement.

Sustainability 2017	A comparative study on decoupling relationship and influence factors between China's regional economic development and industrial energy-related carbon emissions. Zhou et al. (2017), China, Cit. Index 36	To study the relationship between China's regional economic development and industrial carbon emission, reveal the major influencing factors and mechanism of carbon emission change in all regions in eight major regions of China between 1996 and 2012. The overall level of carbon emission reduction technology in China was low and backwards, and limited contributions have been made to economic growth and industry energy carbon emission decoupling.
Sustainability 2017	Sustainability characteristics of China's Poyang Lake Eco-Economics Zone in the big data environment. Zhang, N. and Chen, Z. (2017), China, Cit. Index 16	In this study, we aim to examine its sustainability characteristics, including environmental efficiency, the shadow price of pollutants, and substitutability among inputs and outputs, so that to improve the environmental regulation in this area. Results show that many regions in the Poyang Lake Eco-Economics Zone are not performing well on environmental efficiency and have potential to improve by 31.9% on average.
Sustainability 2017	How would big data support societal development and environmental sustainability? Insights and practices. Song et al. (2017), China, Cit. Index 26	The authorship reflected growing collaboration between researchers from many different countries or universities. Natural resources have been exploited regardless of the consequences. New approaches to bridge the gaps between the economic development and environmental protection were increasingly dominant. Empirically, many of the papers provided case studies of regions in China and other regions.
Sustainability 2017	Spatiotemporal dynamic analysis of forest ecosystem services using "big data": A case study of Anhui province, central-eastern China. Li et al. (2017), China, Cit. Index 6	Few papers have concentrated on accounting the spatiotemporal dynamic changes of forest ecosystem services in a region. To make a more accurate accounting, this paper selects intense data sets that assess the physical quantity and value of forest ecosystem services. The results demonstrate that the total value of forest ecosystem services is beyond the previous estimation while using spatiotemporal visualisation. The results address that forest ecosystem services should be measured more accurately with the help of big data in the future.
Sustainability 2017	The effects of carbon dioxide, methane and nitrous oxide emission taxes: An empirical study in China. Liu, L. and Wu, G. (2018), China, Cit. Index 3	China has become the largest producer of carbon dioxide (CO ₂) and non-CO ₂ greenhouse gas emissions in the world. Current Chinese policies are emphasising the control of CO ₂ emissions only, and no goals have been established for non-CO ₂ greenhouse gas emission reductions. Results indicate that the tax reducing CH ₄ and N ₂ O emissions will cause lower impacts. Chinese government should also pay more attention to CH ₄ and N ₂ O emission reduction.
Sustainability 2018	The effects of public attention on the environmental performance of high-polluting firms: Based on big data from web search in China. Cheng, J. and Liu, Y. (2018), China, Cit. Index 1	Rapid economic development in China has led to a growing public awareness of the importance of environmental protection. Using public web search volume data, this study explored the effects of public attention on the environmental performance of firms in high-polluting industries. We found that firms exposed to higher levels of public attention had better environmental performance, especially state-owned firms.
Sustainability 2018	Using big data for insights into sustainable energy consumption in industrial and mining sectors. Herman et al. (2018), South Africa, Cit. Index 0	This study investigates sustainable energy consumption in the industrial and mining sectors. These sectors are experiencing exponential data growth; their adoption of big data technologies is limited. Hence, big data methods were applied to design a big data system that makes novel use of micro services and containers for modularity and extensibility. MongoDB served as the primary data store. Sustained reductions in energy consumption result in cleaner production and will aid in addressing the energy challenges confronting the economy.
Sustainability 2018	Sustainable robust layout using Big Data approach: A key towards industry 4.0. Kumar et al. (2018), India, Cit. Index 0	A layout design based on large-scale (Big) data is more efficient and effective in today's competitive market. A robust and sustainable industry layout design is capable to handle the variations and is seen as first step towards Industry 4.0 to keep precise and accurate manufacturing of products in given due time. Poor layout design decreases the precision and accuracy in manufacturing of products and increases the production time.
Transport 2017	Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China. Luo et al. (2017), China, Cit. Index 19	Air pollutions from transportation sector have become a serious urban environmental problem, especially in developing countries with expending urbanisation. Cleaner technologies advancement and optimal regulation on the transporting behaviours and related design in infrastructures is critical to address above issue. Paper analysed the taxi's energy consumption, emissions, and their spatial-temporal distribution in Shanghai, applying big data.

Transport 2017	Carbon emission flow from self-driving tours and its spatial relationship with scenic spots - A traffic-related big data method. Huang et al. (2017), China, Cit. Index 8	Low-carbon tourism is an important means of saving energy and reducing emissions. With the advent of big data mining technology, it is urgent for academia to discuss the measurement of self-driving tour carbon emission flows and its spatial relationship with scenic spots based on big data on traffic. The results address that high carbon emission flows from self-driving tours may happen to both high-grade and low-grade scenic spots.
-------------------	---	---

Table 2. Main categories, article information, and main description of the selected articles.

Transportation sector air pollution has become a serious urban environmental problem, especially in developing countries that are facing expending urbanisation. Luo et al. (2017) explored taxi emissions by applying big-data-based travel patterns in Shanghai. They examined energy consumption and emission rates, and their spatiotemporal distributions in the city. The clear message is that there are large variations in daytime pollution accumulation. Respectively, Wu et al. (2017) focused on big-data-driven analysis of Chinese urbanisation that has generated well-known problems such as air pollution, resource shortages, and congestion. The paper presents a development framework for smart city development applying big data.

Big data is a substantial topic in the studies focusing on forestry, farming, and land use. For example, Singh et al. (2018) argue that researchers have neglected a significant contributor to global emissions: The agricultural industry and especially the emissions generated in meat production. Their study assessed the big data cloud computing framework in the context of responsible (low carbon producing) supplier selection in a livestock farming supply chain. According to Singh et al. (2018), there is a need to raise awareness among farmers and to start using the carbon footprint as an important selection criterion in the selection of cattle suppliers. In addition, forests are extremely important CO₂ sinks both globally and locally.

Spatiotemporal (dynamic) changes of forest ecosystem services are not a widely studied topic. This provides an interesting insight into forest ecosystems and big data applications. Li et al. (2017) conducted a spatiotemporal dynamic case analysis of forest ecosystem services using big data in the Anhui province located in central-eastern China. The paper argues that forest ecosystem services should (in the future) be measured more accurately with the help of big data. Especially in populous countries (such as China), urban land expansion management requires rigid measurement and coordination. Urban land is here considered a ‘limited resource’ that should be managed and planned based on accurate data. Similarly, Zeng et al. (2017) applied a big data point-of-view that can be considered a platform for multi-source data fusion. This aids the creation of spatially explicit decision tools for regulating urban land expansion. The paper used big data for assessing the intensity of urban land use in the metropolitan areas of China, and its results prove that the urban land use system is highly efficient in some areas including areas with low (urban) land intensity.

4. CONCLUSIONS

Global climate is expected to change more rapidly than suggested by the IPCC (2018) report. Knowledge-and science-based actions are needed globally on every level in order to make societies and industrial production environmentally sustainable. Big data analytics offer improved ways to study emissions, since manufacturing industries produce an extensive

amount of data that is (in most cases) automatically monitored and recorded. These data sources provide a good opportunity to uncover complicated links in production value chains. The articles analysed indicate that production chains are too often ignoring environmental responsibilities. Including emission data in decision support models would provide one solution for tackling the responsibility dilemma. Thus, big data opens new perspectives in environmental research. However, big data itself seems to be a complex topic. This complexity is a result of the various methods used to collect the data. It includes questions relating to storing intervals, the reliability of the data source, the level of automation (e.g. IoT data storing), and the logical interrelations within the data.

Proposals (or recommendations) for action to solve these negative impacts of different industries were, largely, absent from the papers that were studied. All these papers exhibited and addressed the existing situation, but did not so much indicate how to reach the desired (lower) emission rates. It is understandable that lowering the number of polluting particles and climate warming potentials is usually not voluntary in companies (e.g. the case of maritime scrubbers, see Makkonen & Inkinen 2018). It could be stated as a general conclusion that extensive international co-operation is needed in terms of legislation and environmental agreements.

It is evident that building zero-emission industrial production (and society) needs accurate and continuously updated data. The use of big data allows a data-driven way to (re)consider how to achieve sustainable production. As pointed out, the reviewed articles focus more on the existing situation than recommended solutions for environmental efficiency. However, the pollution sources discovered may help in finding solutions in different production environments. Several papers state that environmental protection actions are still rather rare in many areas, particularly in China. Thus, there are significant research challenges ahead. There is a need for a comprehensive review of big data analytics throughout product lifecycles. Its focus could be on the support tools of sustainable smart manufacturing. This requires reviewing the current analytical frameworks and classifications that are applied in emission control tools.

A concluding observation, based on the articles, is that actions reducing the rate of climate warming urgently require mandatory actions, such as the creation of strict internationally binding environmental laws (resulting in penalties). An alternative, and perhaps an easier road, would require international co-operation and increasing environmental actions significantly through sufficient funding. However, the latter seems unlikely when taking into account the political decisions made during the observation period (2016–2018) e.g. in the US. Additionally, more research and familiarisation concerning big data is necessary in the development of decision support systems. This also requires co-operation between

academia and industries. Big data offers vast possibilities in structured information storage and retrieval for the needs of environmental management.

REFERENCES

- Ala-Harja, H. & P. Helo (2015). Green supply chain decisions – Case-based performance analysis from the food industry. *Transportation Research Part E: Logistics and Transportation Review* 74: 11–21.
- Chen, L. & G. Jia (2017). Environmental efficiency analysis of China's regional industry: a data envelopment analysis (DEA) based approach. *Journal of Cleaner Production* 142, Part 2: 846–853.
- Cheng, J. & Y. Liu (2018). The effects of public attention on the environmental performance of high-polluting firms: Based on big data from web search in China. *Journal of Cleaner Production* 186: 335–341.
- EU (2016). Digital single market Big Data. Accessed 15.1.2019 at: <https://ec.europa.eu/digital-single-market/en/big-data/>
- Gil, D. & I.-Y. Song (2016). Modelling and management of Big Data; Challenges and opportunities. *Future Generation Computer Systems* 63: 96–99.
- Hämäläinen, E., E. Twrdy & T. Inkinen (2017). Cost aggregation in export logistics chain. *Journal of Open Innovation* 3:26, 1–14.
- Hämäläinen, E. & T. Inkinen (2017). How to generate economic and sustainability reports from Big Data? Qualifications of process industry. *Processes* 5: 4, 64.
- IPCC (2018). Special Report 1.5. Accessed 10.2.2019 at <https://www.ipcc.ch/sr15/>
- Kumar, R., S.P. Singh & K. Lamba (2018). Sustainable robust layout using Big Data approach: A key towards industry 4.0. *Journal of Cleaner Production* 204: 643–659.
- Larson, E. (1989). What sort of car-rt-sort am I? Junk mail and the search for self Harper's Magazine. Accessed 15.1.2019 at: <http://harpers.org/archive/1989/07/what-sort-of-car-rt-sort-am-i-junk-mail-and-the-search-for-self/>
- Li, T., Y. Cui & A. Liu (2017). Spatiotemporal dynamic analysis of forest ecosystem services using "big data": A case study of Anhui province, central-eastern China. *Journal of Cleaner Production* 142, Part 2: 589–599.
- Luo, X., L. Dong Y. Dou, N. Zhang, J. Renh, Y. Li, L. Sun & S. Yao (2017). Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China. *Journal of Cleaner Production* 142, Part 2: 926–935.
- Makkonen, T. & T. Inkinen (2018). Sectoral and technological systems of environmental innovation: The case of marine scrubber systems. *Journal of Cleaner Production* 200: 110–121.
- Ren, S., Y. Zhang, Y. Liu, T. Sakao, D. Huisingh & C.M.V.B. Almeida (2019). A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions. *Journal of Cleaner Production* 210: 1343–1365.
- Roman, B.M., A.B.L. de Sousa Jabbour, C.J.C. Jabbour, P de Camargo Fiorini, Y. Mohd-Yusoff & A.M. T. Thomé (2018). Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for big data management towards a low carbon society. *Journal of Cleaner Production* 189: 763–774
- Seele, P. (2018). Predictive Sustainability Control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management. *Journal of Cleaner Production* 153: 673–686.
- Singh, A., S. Kumari, H. Malekpoor & N. Mishra (2018). Big data cloud computing framework for low carbon supplier selection in the beef supply chain. *Journal of Cleaner Production* 202: 139–149.
- Song, M., L. Cen, Z. Zheng, R. Fisher, X. Liang, Y. Wang & D. Huisingh (2017). How would big data support societal development and environmental sustainability? Insights and practices. *Journal of Cleaner Production* 142, Part 2: 489–500.
- Wang, G., A. Gunasekaran, E.W.T. Ngai, & T. 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.
- Wu, K.-J., C.-J. Liao, M.-L. Tseng, M.K. Lim, J. Hu & K. Tan (2017). Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties. *Journal of Cleaner Production* 142, Part 2: 663–676.
- Zeng, C., L. Yang & J. Dong (2017). Management of urban land expansion in China through intensity assessment: A big data perspective. *Journal of Cleaner Production* 153: 637–647.
- Zhang, N. & Z. Chen (2017). Sustainability characteristics of China's Poyang Lake Eco-Economics Zone in the big data environment. *Journal of Cleaner Production* 142, Part 2: 642–653.
- Zhang, Y., S. Ren, Y. Liu & S. Si (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *Journal of Cleaner Production* 142, Part 2: 626–641.
- Zhong, R.Y., S.T. Newman, G.Q. Huang & S. Lan (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering* 101: 572–591.
- Zhu, Q., J. Wu, X. Li & B. Xiong (2017). China's regional natural resource allocation and utilization–DEA-based approach in a big data environment. *Journal of Cleaner Production* 142, Part 2: 809–818.