In: Decarbonizing Value Chains. 20th Global Conference on Sustainable Manufacturing (GCSM 2024), 9-11 October 2024, Ho Chi Minh City, Vietnam. Lecture Notes in Mechanical Engineering, pp. 128-136
DOI:10.1007/978-3-031-93891-7_15

The Lean and Green Imperative of Manufacturing Data

John Patsavellas(✉), Yousef Haddad, and Konstantinos Salonitis

Sustainable Manufacturing Systems Centre, Cranfield University, Bedford MK43 0AL, UK
John.Patsavellas@cranfield.ac.uk

Abstract. This study introduces a stochastic model-based framework for the prediction and measurement of the environmental impact of manufacturing systems' digitalization. Utilising a Monte Carlo simulation experimental framework, this paper forecasts the CO_2e emissions from the entire lifecycle of manufacturing data over long-term time horizon under different scenarios. The analysis proceeds to estimate the maximum, average, and minimum potential CO_2e emissions, under different growth models. Findings reveal that, with the current exponential growth of data that exceeds data centres' efficiency improvements and carbon intensity decay rates, the environmental footprint associated with the entire lifecycle of data can have a potential adverse impact on the realisation of net-zero goals. The proposed approach provides a viable pathway for manufacturing enterprises aiming to align their data management practices with environmental sustainability and operational efficiency.

Keywords: Digitalization · Industry 4.0 · Monte Carlo Simulation · Sustainable Data

1 Introduction

The advent of Industry 4.0, with digitization being a founding cornerstone [1], has resulted in an exponential growth of data generation. This increase in the volume of data, resulting from the propagation of advanced information and communication technology (ICT), has far exceeded the growth in population and industrial activity [2]. Indeed, it was estimated in 2019 that 90% of all data back then was generated within the preceding two years [3]. This exponential growth is, however, accompanied by a significant growth in the energy required to generate, transmit, and store these data in data centres [4]. Indeed, the International Energy Agency estimates that 1–1.3% of the worldwide electricity consumption comes from data centres, and 0.6% of total greenhouse gas (GHG) emissions comes from these data centres [5].

In general, carbon intensity, which is the amount of CO_2e per unit energy, of energy mixes worldwide has significantly decreased over the last two decades. For example, China has managed to decrease the carbon intensity of its energy mix by 56% over a 15 year period from 2005 to 2020 [6]. In addition, energy efficiency of data centres has also been improving consistently. For example, the authors in [7] report that an increase

© The Author(s) 2025
H. Kohl et al. (Eds.): GCSM 2024, LNME, pp. 128–136, 2025.
https://doi.org/10.1007/978-3-031-93891-7_15

an excess of fivefold of computing power in data centres has resulted in a mere increase of 6% in energy consumption over an 8 year period from 2010 to 2018. However, this increase might still be insufficient to deal with the exponential growth of data. These improvements in energy efficiency and the carbon intensity of such energy pose the risk of the rebound effect where improvements in efficiency and carbon intensity might be counterpart and overcompensated by an unproportional increase in data usage [8]. The degree of the risk of the rebound effect with regards to the growth of data consumption and its impact on the environment is, however, still unclear. For example, the authors in [6] found that the rebound effect is not significant with regards to the ICT industry, which is tightly related with data (i.e., the use phase of ICT devices), while in another study, the authors in [9] found that digitalization can have a strong rebound effect with regards to energy consumption.

In this paper, a stochastic model-based framework for the prediction of digitalization-related growth in data over its entire lifecycle, along with its environmental footprint, in the manufacturing industry is developed. The framework considers two growth models under a Monte Carlo simulation experimental framework; unbounded (exponential) growth, and bounded (logistic) growth. The rest of the paper is organised as follows: the next section presents the overarching model-based framework, Sect. 3 presents a numerical example, and finally Sect. 4 presents and discusses the concluding remarks.

2 Model-Based Framework

In this section, the overarching approach for the data and CO_2e prediction is presented and discussed. The determination of the growth of data and their subsequent environmental footprint are determined under Monte Carlo simulations to reveal possible future pathways of the carbon footprint of manufacturing digitalization. Monte Carlo simulation is a stochastic modelling approach that involves random sampling of an experiment's stochastic parameters from probability distribution over several independent replications [10].

The first step is to estimate the amount of data generated by the manufacturing system. This can be done through analysis of historical data and identification of data growth. Although analysis of historical data by itself is insufficient to determine data growth, it can still shed some light on potential growth pathways. This is because data growth depends also on the emergence of new technologies (e.g., sensors, new machine models, advanced software tools etc.…). Therefore, the growth-parameters identification should be conducted carefully by analysing different factors such as the prospects of updating new machines, acquiring software tools etc.… Next, other key parameters such as carbon intensity's decay rate are determined. It is important to point out here that carbon intensity can vary greatly, depending on the data storage location (i.e., local vs data centres). This is because, as mentioned earlier, data centres are highly efficient with regards to energy consumption.

The next step is to calculate the CO_2e emissions over the chosen period. Any growth model can be chosen, however, as discussed earlier, the size of data has demonstrated exponential growth traits. Nevertheless, long-term exponential growth is highly unlikely, even impossible. Indeed, in their seminal work *The Limits to Growth* [11], the authors

discussed how limitations in resources will inevitably inhibit exponential growth and result in a slowdown in growth. Therefore, exponential growth often gradually slows down after a certain period of time elapses, resulting in a sigmoid-shaped growth, most recognizable of which is the logistic curve [12].

After the data growth model has been determined, energy consumption of data lifecycle stages has to be determined next. These stages are: data generation, data transmission, and data storage. These can be calculated as shown in Eq. (1) below:

$$E_{total} = E_{generation} + E_{transmission} + \left(E_{storage} \times t\right) \tag{1}$$

where E_{total} is the total energy over the entire lifecycle of each unit of data (e.g., per bit), $E_{generation}$ is the energy required to generate each unit of data, $E_{transmission}$ is the energy required to transmit each unit of data to the storage location, $E_{storage}$ is the energy required to store each unit of data for a period of t units of time. It should be noted here that the energy required to store data typically constitutes most of the data's lifecycle energy consumption, particularly with regard to cooling requirements [4].

In order to reflect the trend of the ever-decreasing carbon intensity of most grids, carbon intensity is assumed to exponentially decay. Therefore, the amount of CO2e emissions at any year for the logistic and exponential growth models, respectively, is calculated as follows

$$CO_2e_{t+1} = E_t CI_0 (1 - s_t)^t \left[D_t + r_t D_t \left(1 - \frac{D_t}{K} \right) \right], \quad t \in \{0, 1, \ldots, n-1\} \tag{2}$$

$$CO_2e_{total} = (E_t \times CI_0) \times a \prod_{t=1}^{n} ((1 + r_t)(1 - s_t)) \tag{3}$$

where CO_2e_{t+1} is the amount of CO_2e emissions at the end of year t (note that $CO_2e_0 = CI_0 D_0$), K is the carrying capacity (i.e., the maximum amount of data that can be realistically sustained), D_0 is the *current* data generation during time period $t = 0$, r is the growth rate, s is the carbon intensity's decay rate, E_t is the total energy consumption per unit of the data's entire lifecycle, and t is the time period.

3 Numerical Example

In order to test and validate the framework, and to demonstrate its applicability, a numerical example constituting various representative scenarios is presented and discussed in this section. The numerical example consists of 18 scenarios that depict different manufacturing company sizes/ grid's carbon intensity configurations, and growth models as presented in Table 1 below.

The modelled scenarios consist of three manufacturing systems sizes (small, medium, and large) and three geographical locations, representative of different levels of grids' carbon intensities. It should be noted here that one of the primary assumptions of this study is that the amount of data generated by a manufacturing enterprise is proportionally related to its size, which is not always necessarily valid. This assumption follows the logical view that bigger manufacturing firms have more machines, more processes, more people, hence more data associated with the higher number of resources.

Table 1. The examined scenarios

Scenario	D_0 (GB/day)	CI_0 (gCO_2e/ kWh)	Growth model (Unbounded - bounded)	Carrying capacity (GB)
S_1	0.5	258	U	N/A
S_2	0.5	258	B	10^4
S_3	0.5	369	U	N/A
S_4	0.5	369	B	10^4
S_5	0.5	582	U	N/A
S_6	0.5	582	B	104
S_7	5	258	U	N/A
S_8	5	258	B	10^5
S_9	5	369	U	N/A
S_{10}	5	369	B	10^5
S_{11}	5	582	U	N/A
S_{12}	5	582	B	10^5
S_{13}	50	258	U	N/A
S_{14}	50	258	B	10^6
S_{15}	50	369	U	N/A
S_{16}	50	369	B	10^6
S_{17}	50	582	U	N/A
S_{18}	50	582	B	10^6

In addition, big manufacturing facilities are more likely to adopt advanced technologies and wide variety of ICT infrastructure and software tools. It should be acknowledged, however, that there might be exceptions to this rule, especially with the falling cost of data generation means, primarily sensors [3].

The model's global parameters (i.e., those that are not scenario-specific), were chosen to reflect the current, and the forecasted future state of data growth and carbon intensity decay as displayed in Table 2 below.

The data storage period is assumed to be 25 years since it is unlikely that companies would want to store *all* data for more than that period, especially since most equipment and software tools have a lifetime of less than 25 years. As for the carbon intensity figures, they were chosen to reflect regional discrepancies. These carbon intensity values correspond with average grids' carbon intensity in the European Union [13] and the USA and China [14]. These carbon intensity values represent a wide range of carbon intensities in areas with active manufacturing industries.

Table 2. The model's global parameters

Parameter	Value	Unit
Simulation period	25	Year
Growth rate (μ, σ)	(0.2, 0.04)	-
Decay rate U[λ_{min}, λ_{max}]	[0.01–0.03]	-
E_{total}	1	kWh/GB/year
Number of independent replications	1000	Replication

3.1 Results and Discussion

When the model was run for a total of 1000 independent replications, using random number seeds for each replication, the following results, displayed in Fig. 1 below, were obtained.

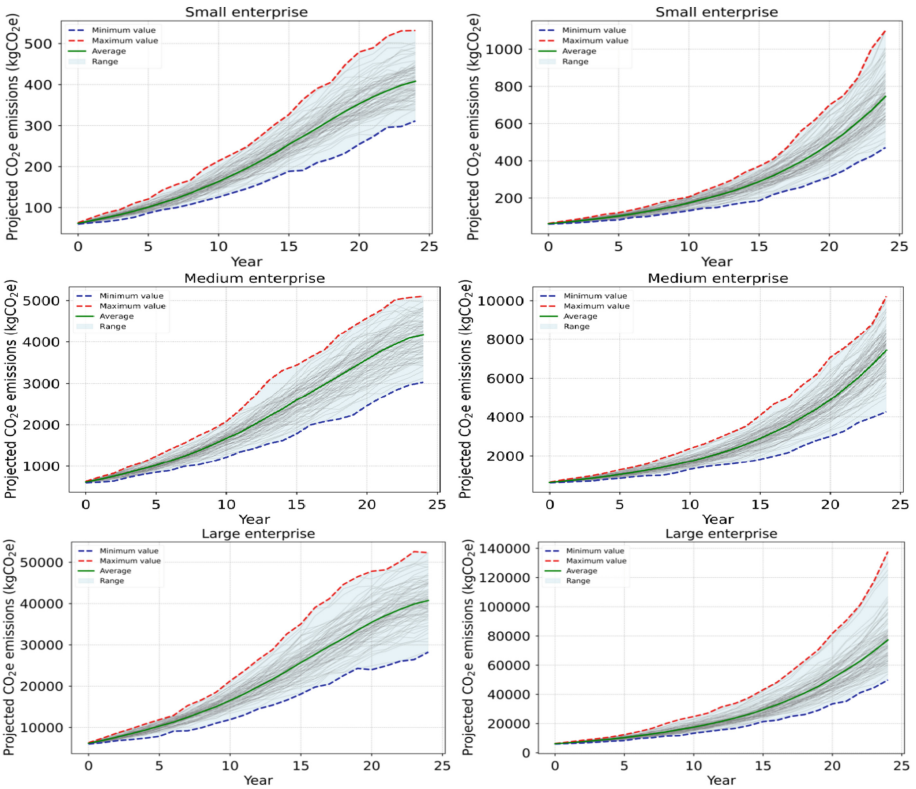

Fig. 1. Total CO_2e emissions for the small, medium and large manufacturing enterprises for the logistic (left) and exponential (right) growth models when carbon intensity is medium

It could be noticed from Fig. 1 that although the carbon intensity of the energy mix decreases exponentially, and although relatively low carrying capacities were assumed, substantial amount of CO_2e emissions were emitted from each of the presumed enterprises. Another important remark, is that the uncertainty with regards to data generation growth, and the decay rate of carbon intensity, can have a significant impact on the prediction of total CO_2e emissions, differing by as much as 50%.

In order to further examine the behaviour of CO_2e emissions and volumes of data at individual years, Fig. 2 below displays a heatmap, at logarithmic scale, for CO_2e emissions over the years.

Another important insight from the results is that although carbon intensity of the grid is assumed to decrease exponentially (and its decrease indeed resembles exponential decay), this did not have a determining impact on the resulting amounts of CO_2e emissions. This might be attributed to data generation growth rate exceeding the growth rate of the renewables' share in most energy grids.

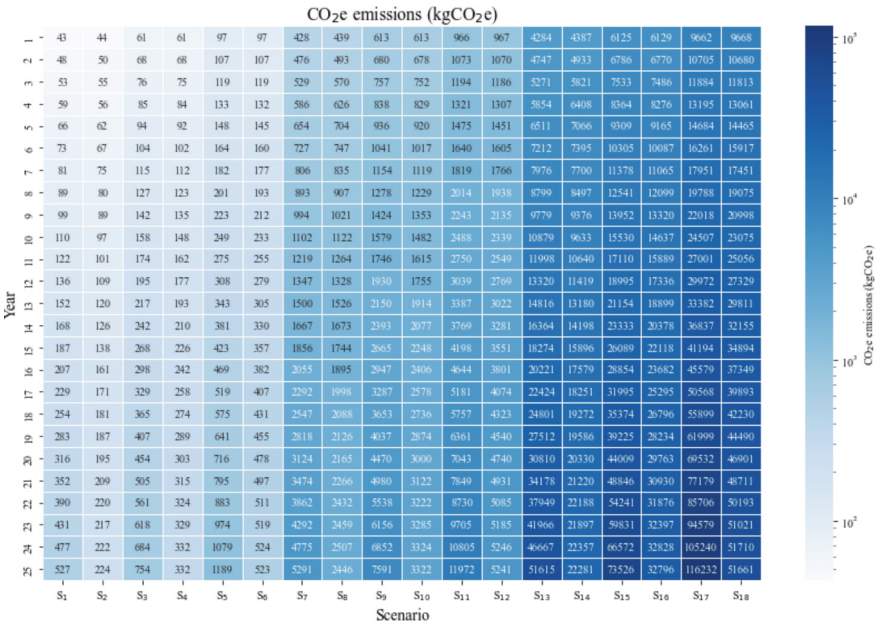

Fig. 2. Heatmap (logarithmic scale) displaying total amount of CO_2e emissions for all scenarios during all years

Figure 2 shows that scenarios 13 to 18 generate the largest amount of CO_2e emissions. This is an intuitive outcome since these scenarios are associated with large enterprises. However, it is important to emphasize here that the limit to growth does not have a substantial impact on mitigating the environmental impact of data. To explain more, although the scenarios where growth followed logistic distribution had significantly lower data volumes, and subsequently CO_2e emissions, these projected emissions still constitute a significant burden on the environment. Considering the thousands of large manufacturing

enterprises worldwide, the total amount of CO_2e emissions from data alone will be several million metric tons, which can have an adverse effect on the prospects of achieving net-zero emissions by 2050. Indeed, this result is in line with [9] where the authors found that ICT (which is the cornerstone of digitalization) can have a strong rebound effect, hence rendering efficiency and carbon intensity improvements insufficient to match the growth in data volumes.

Since there is a lack in studies that model data growth over the long-term, direct comparison of the results with previous studies is difficult. However, in order to validate the results obtained from the numerical example, input data were validated by two subject matter experts [15], who are senior managers at small manufacturing enterprises in the Middle East and the UK. The rest of the input data were mostly obtained from official sources (e.g., the IEA and the EEA, amongst others). Finally, the mathematical formulations are well-established and widely used in population dynamics modelling, therefore, they have high validity. One point, however, that requires further exploration is the amount of energy required per GB per year. There is very little agreement about this in the literature. For example, estimations of the energy required for data transmission has been reported to vary between 0.004–135 kWh/GB [16], while estimations with regards to data storage varied between 0.0006–0.04 kWh/GB/year [17]. Therefore, it is recommended for future studies to conduct sensitivity analysis experiments, in particular with respect to energy consumption.

The aforementioned results indicate that with the current trend of exponential increase of data, improvements in energy efficiency and carbon intensity are unlikely to compensate for the environmental impact of these data. Therefore, manufacturing enterprises should carefully how to prioritize their data requirements, particularly with the propagation of cloud storage, big data, large language models, and other data-intensive tools.

4 Conclusion

In this paper, a stochastic model-based framework for the prediction of data growth and the associated environmental footprint that result from digitalization of the manufacturing industry is developed. Several scenarios constituting various manufacturing systems' sizes, geographical locations and growth models were stochastically examined under a Monte Carlo simulation framework. Results reveal that despite the exponentially decreasing carbon intensity, and the limitations posed on data growth (the carrying capacity), the environmental footprint of data over the long-term is significantly high. Another interesting insight from the results is that although uncertainty was significantly high in the long-term, the lower bound of emissions within all scenarios for the medium and large manufacturing enterprises was significantly high.

This research can be extended in a number of directions. First, more scenarios can be examined, particularly with regards to long-term energy-efficiency. Another possible direction is to carry out sensitivity analysis experiments on key parameters, particularly regarding energy consumption and carrying capacity for the logistic growth model. Finally, it is worthwhile to investigate the viability of optimization models that can determine the optimal amounts of data that manufacturing companies can sustain over the long-term with minimal environmental footprint.

References

1. Ghobakhloo, M.: Industry 4.0, digitization, and opportunities for sustainability. J. Clean. Prod. **252**, 119869 (2020)
2. Belkhir, L., Elmeligi, A.: Assessing ICT global emissions footprint: trends to 2040 & recommendations. J. Clean. Prod. **177**, 448–463 (2018)
3. Patsavellas, J., Salonitis, K.: The carbon footprint of manufacturing digitalization: critical literature review and future research agenda. Procedia CIRP **81**(March), 1354–1359 (2019)
4. Zhang, Q., Clements-croome, D.: Innovative approaches for deep decarbonization of data centers and building space heating networks: modeling and comparison of novel waste heat recovery systems for liquid cooling systems. Appl. Energy **357**(December 2023) (2024)
5. Data Centres and Data Transmission Networks. International Energy Agency (2023)
6. Sun, H., Kim, G.: Technology in society the composite impact of ICT industry on lowering carbon intensity: from the perspective of regional heterogeneity. Technol. Soc. **66**(March) (2021)
7. Farfan, J., Lohrmann, A.: Gone with the clouds: estimating the electricity and water footprint of digital data services in Europe. Energy Convers. Manag. **290**(June), 117225 (2023)
8. Pohl, J., Hilty, L.M., Finkbeiner, M.: How LCA contributes to the environmental assessment of higher order effects of ICT application: a review of different approaches. J. Clean. Prod. **219**, 698–712 (2019)
9. Lange, S., Pohl, J., Santarius, T.: Digitalization and energy consumption. Does ICT reduce energy demand? Ecol. Econ. **176**(June), 106760 (2020)
10. Raychaudhuri, S.: Introduction to Monte Carlo simulation. In: Proceedings of the 2008 Winter Simulation Conference, pp. 91–100. IEEE (2008)
11. Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.W.: The Limits to Growth. Universe Books, New York (1972)
12. Tsoularis, A., Wallace, J.: Analysis of logistic growth models. Math. Biosci. **179**, 21–55 (2002)
13. Greenhouse gas emission intensity of electricity generation in Europe. European Environment Agency (2024). https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1?activeAccordion=546a7c35-9188-4d23-94ee-005d97c26f2b
14. Statistical Review of World Energy. Ember; Energy Institute (2024). https://ourworldindata.org/grapher/carbon-intensity-electricity
15. Law, A.M.: How to build valid and credible simulation models. In: Proceedings - Winter Simulation Conference, pp. 24–33. IEEE (2019)
16. Aslan, J., Mayers, K., Koomey, J.G., France, C.: Electricity intensity of internet data transmission untangling the estimates. J. Ind. Ecol. **22**(4), 785–798 (2018)
17. Al Kez, D., Foley, A.M., Laverty, D., Del Rio, D.F., Sovacool, B.: Exploring the sustainability challenges facing digitalization and internet data centers. J. Clean. Prod. **371**(August), 133633 (2022)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

https://dspace.org/

The lean and green imperative of manufacturing data

Patsavellas, John; Haddad, Yousef; Salonitis, Konstantinos

2024-10-09
Springer
https://doi.org/10.1007/978-3-031-93891-7_15
Patsavellas J, Haddad Y, Salonitis K. (2025) The lean and green imperative of manufacturing data. In: Decarbonizing Value Chains. 20th Global Conference on Sustainable Manufacturing (GCSM 2024), 9-11 October 2024, Ho Chi Minh City, Vietnam. Lecture Notes in Mechanical Engineering, pp. 128-136

Downloaded from a DSpace repository. https://dspace.org/