



HAL
open science

A Deep Learning Algorithm for the Throughput Estimation of a CONWIP Line

Silvestro Vespoli, Andrea Grassi, Guido Guizzi, Valentina Popolo

► **To cite this version:**

Silvestro Vespoli, Andrea Grassi, Guido Guizzi, Valentina Popolo. A Deep Learning Algorithm for the Throughput Estimation of a CONWIP Line. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2021, Nantes, France. pp.143-151, 10.1007/978-3-030-85874-2_15 . hal-04030367

HAL Id: hal-04030367

<https://inria.hal.science/hal-04030367>

Submitted on 15 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

A deep learning algorithm for the throughput estimation of a CONWIP line

Silvestro Vespoli¹, Andrea Grassi¹, Guido Guizzi¹, and Valentina Popolo¹

Univeristy of Naples Federico II, Napoli (NA) 80125, Italy
{silvestro.vespoli, andrea.grassi, guido.guizzi,
valentina.popolo}@unina.it

Abstract. The ability to meet increasingly personalized market demand in a short period of time and at a low cost can be regarded as a fundamental principle for industrialized countries' competitive revival. The aim of Industry 4.0 is to resolve the long-standing conflict between the individuality of on-demand output and the savings realized through economies of scale. Significant progress has been established in the field of Industry 4.0 technologies, but there is still an open gap in the literature regarding methodologies for efficiently manage the available productive resources of a manufacturing system. The CONtrolled Work-In-Progress (CONWIP) production logic, proposed by Spearman et al., allows controlling the Work-In-Progress (WIP) in a production system while monitoring the throughput. However, an affordable estimation tool is still required to deal with the increased variability that enters the current production system. Taking advantage of recent advances in the field of machine learning, this paper contributes to the development of a performance estimation tool for a production line using a deep learning neural network. The results demonstrated that the proposed estimation tool can outperform the current best-known mathematical model by estimating the throughput of a CONWIP Flow-Shop production line with a given variability and WIP value set into the system.

Keywords: Industry 4.0 · Industrial Production System · CONWIP Flow-Shop · Throughput Estimation · Deep Learning

1 Introduction

The continuous transformation of the competitive manufacturing landscape, new challenges spanning innovation, efficiency, expense, and time-to-market all reflect opportunities that a modern company cannot afford to miss. The ability to meet increasingly personalized market demand in a short period of time and at a low cost can be regarded as a fundamental principle for industrialized countries' competitive revival against emerging countries with lower technological development but lower social and labor costs [3]. In this context, it is important to develop the ability to efficiently distribute available resources, as well as the ability to rethink and revolutionize the methods and controlling approaches of production processes in order to react appropriately to new market challenges [5].

The effect of a production paradigm based on increased product customization and shorter time-to-market was significant enough to justify the born of a new industrial

paradigm: the Fourth Industrial Revolution (or Industry 4.0). While previous industrial revolutions were marked primarily by major technological developments, this one pursues logistical/management goals, which are related to the new need for product customization and the consequent versatility needed by a manufacturing plant [1, 12, 15]. Hence, the aim of Industry 4.0 is to resolve the long-standing conflict between the individuality of on-demand output and the savings realized through economies of scale [7, 6]. Significant progress has been established in the field of Industry 4.0 technologies, but to the best of the author's knowledge, there is still an open gap in the literature regarding methodologies for efficiently using the data flow generated by Industry 4.0 Cyber-Physical Systems (CPSs) [14, 18].

Industry 4.0 is intended to enable the transition from a traditional manufacturing model (Mass Production) to a customization-oriented one (Mass Customization). This transformation, however, must be accompanied by a change in the logic of the Manufacturing Planning and Control (MPC) structure. In fact, the main problem in a Mass Customization business scenario is the variability that enters the system as a result of consumer customization requests. And the issue is that, for the time being, this variability can only be dealt with by a traditional MPC scheme, such as Manufacturing Resource Planning (MRP-II, in the following MRP), which act on the manufacturing system with a "push" logic controlling the throughput and monitor the Work-In-Progress (WIP).

In the scientific literature, several control strategies suggesting autonomous and independent control principles have been proposed. Among these, Dolgui et al. pioneered a new research area based on the application of classical Control Theory to scheduling and inventory control systems [4, 16, 10, 9, 11]. Sokolov et al. in [16], in particular, investigated the advantages and limitations of various control approaches and algorithms for the optimal solution of short-term scheduling of a manufacturing system. However, these control strategies requires important change into the production system while the manufacturing firms are still in search of more simple solution, able to controls the WIP level, while monitoring the throughput of the system. One of these solution may be offered from the CONtrolled Work-In-Progress (CONWIP) logic, proposed by Spearman et al [17] in the 1990. They demonstrated that pull systems are more effective than push systems and that pull systems need less WIP to achieve the same value of throughput.

The issue is that, in order to implement a CONWIP production system, the amount of WIP to be allowed in the production system must be defined. The selection of the correct WIP numbers to be admitted is critical for the manufacturing system's performance. In fact, different values of WIP determine different values of the production system's throughput and crossing time. A high level of WIP, in particular, ensures a higher value of throughput at the expense of a higher value of the time required for a job to cross the production system. Low levels of WIP, on the other hand, ensure that jobs have a better crossing time at the expense of the manufacturing system's throughput. As a result, given the current manufacturing paradigm, it is critical to have a production system that allows for dynamically changing the amount of WIP, depending on the required throughput from the market or from the other production line, while keeping the time to cross the production system as short as possible.

In this sense, the aim of this paper is to contribute to the development of a production line performance estimation tool, able to take into account the variability of the processing time that enters the production system and the number of WIP values set. To this extent, a deep learning algorithmic approach for the throughput estimation of a CONWIP Flow Shop production line is proposed. Then, its performance has been assessed against the last-known mathematical model of CONWIP production systems proposed by Spearman et al. in [8] and compared with the results obtained from a simulation model. The remainder of the paper is structured as follow: Section 2 introduces the Problem Statement of the paper; Section 3 introduces the proposed approach, focusing on both the simulation method for generating data and the proposed deep learning algorithm; Section 4 discusses the results, comparing them to the best-known mathematical model; and Section 5 concludes the paper.

2 Problem Statement

Suppose to have a FlowShop CONWIP line with 5 machines where the processing times are generated by a fixed average gamma distribution [Fig 1]. The choice of the gamma distribution is motivated by the desire to introduce controllable variability into the production line. It is, in fact, a continuous probability distribution that includes, among other things, the exponential distributions, and it is used in literature to simulate situations that are far from the *memoryless* characteristic of the exponential distribution. Without loss of generality, we considered different productive scenario in which the average processing time is fixed to 10 minutes and the variability that enters in the system is controlled by a change in the *alpha* value of the above-mentioned gamma distribution.

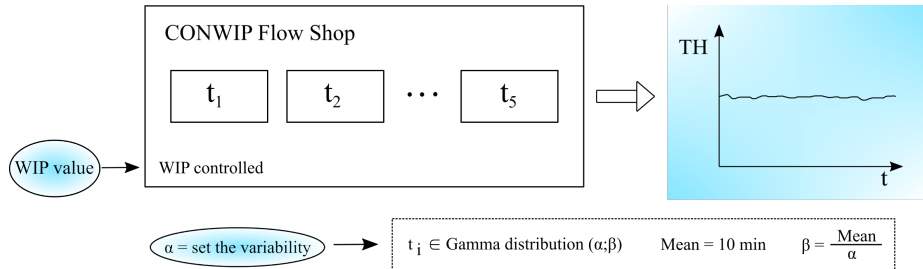


Fig. 1. The considered CONWIP Flow-Shop production line

As said above, in a CONWIP line, the THroughput (TH) is an observable performance parameter that can be controlled by directly acting on the system’s authorized WIP level. Spearman et al in [17, 8] investigated the dynamics of a CONWIP system and developed mathematical models that described its behavior as a function of controlled variables and boundary conditions. They focused on the behavior of the best possible condition (no variability and balanced line, with the same performing times at

all stations) and the worst possible condition. These scenarios are interesting because they represent the best and worst possible performance that a CONWIP production system may show. Furthermore, the researchers assessed the behavior of the same system in a practical case (i.e., the practical worst case, or PWC) in which the performing times of jobs at the stations are exponentially distributed, preserving the balanced line condition in terms of average processing times. Hence, the mathematical relationships that link TH and Cycle Time (CT) to the WIP introduced in the system are known in these cases.

However, their mathematical model is based on the strong assumption of an exponential processing time distribution case, that is not of great help when the variability is different from the exponential case. To this extent, Hopp and Spearman in [8] have then proposed an iterative model based on average value considerations, taking the advantage of the Mean Value Analysis, able to correctly estimate the TH of a CONWIP line with different level of variability, selecting as input for the model the value of WIP admitted in production and the quadratic coefficient of variation of the processing time distribution. However, as already showed by [2] their estimation suffers from an approximation error when forced to work with variability far from the exponential ones, due to the fact that the model's assumptions are not properly verified in these cases.

Therefore, the goal of this paper is to use a machine learning algorithm, based on Deep Learning Neural Network architecture to estimate the Throughput of the production line based on the same input considered by Hopp and Spearman in their last work (i.e., the WIP to be admitted in production and the quadratic coefficient of variation of the processing time distribution). The proposed model's limitations must be found in the requirement for a consistent amount of data that is wide and complete, as well as the fact that the trained model is specific to the considered production system architecture. The proposed method, on the other hand, has the advantage of having a data-driven model architecture that can be trained with data from a simulator or a physical production line and is easily replicable for any type of production system.

3 The proposed deep learning tool

As above analysed, the problem of estimating the TH of a CONWIP line is critical for the proper sizing of the production line's WIP. After analyzing the limitations of the previous modeling in the literature, the goal here is to develop a TH estimation tool for a CONWIP production line using a machine learning algorithm based on Deep Learning Neural Network architecture. However, in order to begin training of this kind of model, a large amount of data must be collected. To that end, a simulation model using the Discrete-Event and Agent-Based techniques with the help of the Anylogic simulation software has been created.

The model's Main agent is depicted in the [Fig 2]. As can be seen, the simulation tool is very simple, consisting of five different services blocks (that simulate the processing machines) and other blocks presented for statistical purposes. To collect all the data required, the model was run for different values of variability (i.e., varying the α value of the processing time distribution at step of 0.1 for value between 0.5 and 3) and of WIP (i.e. varying it between 1 and 20), for a simulation time of 4 years (in order to

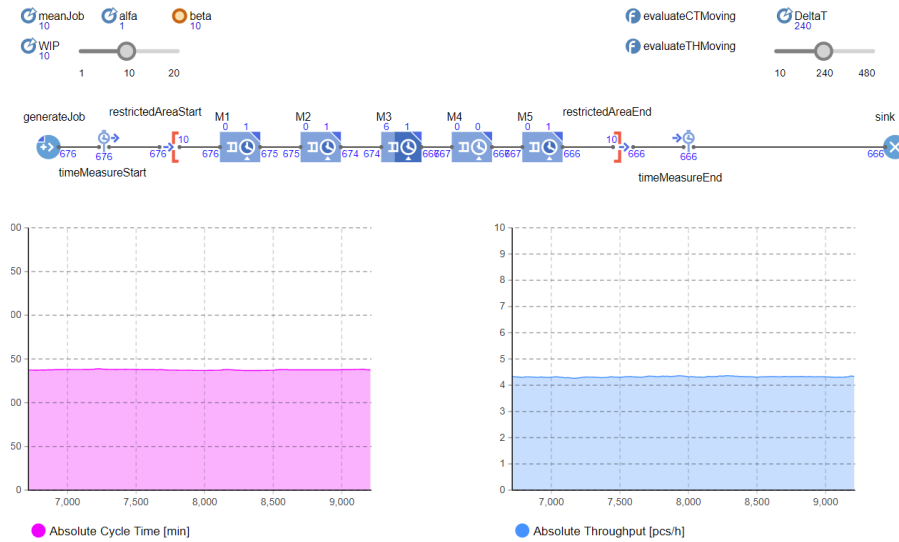


Fig. 2. The main agent of the simulation tool

be sure that the steady state is reached), replicated 40 times for each combination of parameter. Then, the TH and CT of the production line were collected and saved in an Excel spreadsheet after each simulation run.

Once the experimental data-set has been acquired, we proceeded to build the deep learning neural network for the estimation of the TH. These networks are structured of multiple layers of neurons: we have a first layer known as the input layer, which in our case will be made up of two inputs, the WIP of the production line and the quadratic coefficient of variation within the line; different hidden layers with variable number of neurons, which are responsible for numerical interpolation; a final layer that represents the output of the network (i.e. the TH in the considered case) [Fig 3].

Once the network's architecture in terms of input and output has been established, the network's hyper-parameters, such as the number of neurons inside the hidden layer, activation functions of the neurons, loss function, and so on, must be determined. Unfortunately, there is no objective method for selecting these parameters [13]. For this reason, the network has been scaled with a "Trial and Error" approach and the results shown in the following paragraph.

4 The deep learning algorithm - Experimental and Validation Scenario

The model was developed on Google Colaboratory using the Tensor Flow Keras library. The dataset used for model training was made up of 20800 data points extracted from the simulator and divided into three sections: training (75%), validation (20%), and test (5%).

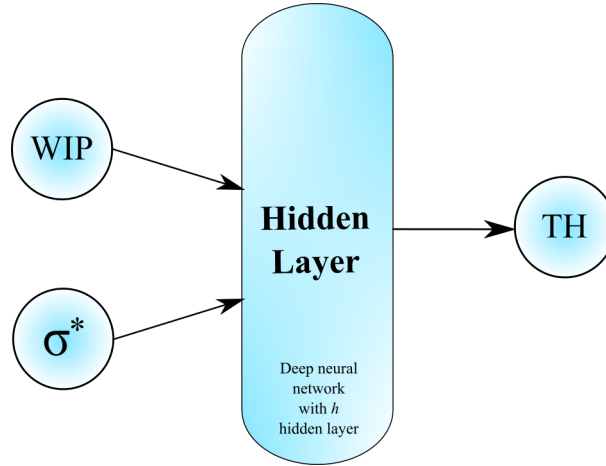


Fig. 3. The structure of the proposed deep learning neural network

As previously stated, several experiments has been carried out in order to find the best combination of hyper-parameters through a trial and error procedure. To ensure an effective comparison of the results, we fixed the seed so that we could see how the model responds to changes in hyper-parameters. After experimenting with different configurations, we concluded that the best results were obtained with a neural network composed of the following layers: an input layer that accepts the WIP values in the production system as well as the quadratic coefficient of variation; a hidden normalization layer; two dense hidden layers of 12 neurons each with *elu* as activation function; an output layer formed by a single neuron.

It should be noted that, prior to arriving at this network configuration, several experiments, which were omitted for space reasons has been carried out, which involved reduced neural network structures (single hidden layer, different number of neurons, different combination of activation function, etc.). For the loss function, we preferred the use of the classic Mean Squared Error (MSE), adopting the ADAM optimizer with a learning rate of 0.01. Also in this case, experiments with different learning rate has been carried out and this value resulted as the ones that performed with good value of MSE also in the Test phase. Regarding the epoch number, we experienced the best performances to a value around 20 epochs for the proposed structure. The results of the deep learning neural network model are shown in the [Fig 4].

Finally, in order to assess the performances of the proposed model, fixed the variability of the system (i.e., α), we compared the predicted values of the TH to the average values obtained from the simulation and to the values obtained from the iterative Spearman's model [8]. As illustrated in [Fig 5 in which the scenario with $\alpha = 3$ is drawn, the proposed neural network model is capable of forecasting the TH of the productive line with high accuracy (i.e., with a a MSE value in the order of 10^{-2}) whereas the Hopp and Spearman's iterative models deviate consistently as the value of the WIP in the line increases. To summarize, the Hopp and Spearman iterative model is extremely precise

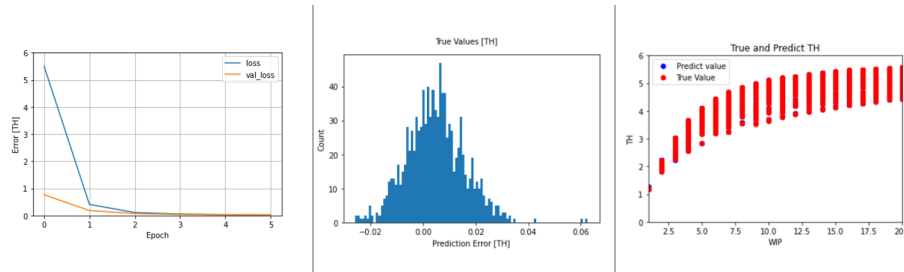


Fig. 4. Results of the deep learning neural network model.

for values of variability close to the exponential (i.e. alpha equal to 1), but it begins to earn systemic errors when the variability that enters the system deviates from this hypothesis. Instead, the proposed model proved to be accurate even when different levels of variability were introduced into the system. .

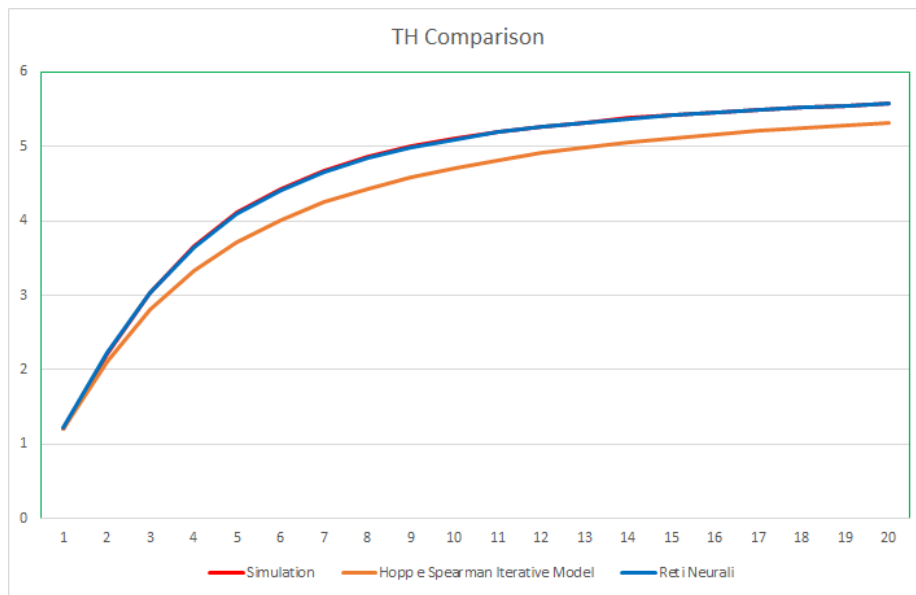


Fig. 5. Comparison of the proposed deep learning neural network model with the iterative *Spearman et al.* [8] model and the data from the simulation tool in a scenario with $\alpha = 3$.

5 Conclusion

In today’s production scenario, where time to market has a significant added value, even when compared to simple productivity, it is critical for production lines to have

performance controllable dynamically. Considering a flow-shop that works in a CON-WIP configuration, this control knob may be the WIP value that can be controlled and modulated appropriately to meet different production constraints. At the same time, as a result of the need to create added value through extreme product customization, the variability introduced within the production system is constantly increasing. The coexistence of these causes complicates the dimensioning of the WIP within the productive system, when a specific value of performances is needed (e.g. a specific throughput).

To that end, we proposed a deep neural network estimation tool able to estimate the throughput value of a production line once the variability and WIP enters the system are known. The current work demonstrated that this tool can make very accurate predictions of production performance, allowing a much finer sizing of WIP than was previously possible even in scenarios with non-exponential variability. The current model's limitation is that it was tested on an a-priori known production line model with a fixed average processing time and number of machines. However, the same methodology here proposed may be applied with a larger dataset that will allow a greater generalisation of the problem. The authors hope that future research will focus on extending this work's reasoning to more general neural network architectures that can take into account different average production times and, in particular, different number of machines in the production line.

References

1. Beier, G., Ullrich, A., Niehoff, S., Reißig, M., Habich, M.: Industry 4.0: How it is defined from a sociotechnical perspective and how much sustainability it includes – A literature review. *Journal of Cleaner Production* **259** (2020). <https://doi.org/10.1016/j.jclepro.2020.120856>
2. Buzacott, J.A., Shanthikumar, J.G.: *Stochastic Models of Manufacturing Systems*. Prentice Hall, first edit edn. (1993)
3. Culot, G., Nassimbeni, G., Orzes, G., Sartor, M.: Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics* **226**(October 2019), 107617 (2020). <https://doi.org/10.1016/j.ijpe.2020.107617>
4. Dolgui, A., Ivanov, D., Sethi, S.P., Sokolov, B.: Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art and applications. *International Journal of Production Research* **57**(2), 411–432 (2019). <https://doi.org/10.1080/00207543.2018.1442948>
5. Fogliatto, F.S., Da Silveira, G.J., Borenstein, D.: The mass customization decade: An updated review of the literature. *International Journal of Production Economics* **138**(1), 14–25 (2012). <https://doi.org/10.1016/j.ijpe.2012.03.002>
6. Grassi, A., Guizzi, G., Santillo, L.C., Vespoli, S.: The manufacturing planning and control system: A journey towards the new perspectives in industry 4.0 architectures. In: *Scheduling in Industry 4.0 and Cloud Manufacturing*, pp. 193–216. Springer (2020)
7. Hermann, M.; Pentek, T.: Design principles for industrie 4.0 scenarios: a literature review **1**(1), 15 (2015). <https://doi.org/10.1109/HICSS.2016.488>
8. Hopp, W.J., Spearman, M.L.: *Factory Physics*. Inc, Waveland Pr, third edit edn. (2011)
9. Ivanov, D., Dolgui, A., Sokolov, B.: Scheduling of recovery actions in the supply chain with resilience analysis considerations. *International Journal of Production Research* **56**(19), 6473–6490 (2018). <https://doi.org/10.1080/00207543.2017.1401747>

10. Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., Ivanova, M.: A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *International Journal of Production Research* **54**(2), 386–402 (2016). <https://doi.org/10.1080/00207543.2014.999958>
11. Ivanov, D., Sethi, S., Dolgui, A., Sokolov, B.: A survey on control theory applications to operational systems, supply chain management, and Industry 4.0. *Annual Reviews in Control* **46**, 134–147 (2018). <https://doi.org/10.1016/j.arcontrol.2018.10.014>
12. Oztemel, E., Gursev, S.: Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing* **31**(1), 127–182 (2020). <https://doi.org/10.1007/s10845-018-1433-8>
13. Patterson, J., Gibson, A.: *Deep Learning: A Practitioner’s Approach*. O’Reilly, first edit edn. (2017)
14. Riedl, M., Zipper, H., Meier, M., Diedrich, C.: Cyber-physical systems alter automation architectures. *Annual Reviews in Control* **38**(1), 123–133 (2014). <https://doi.org/10.1016/j.arcontrol.2014.03.012>
15. Shan, S., Wen, X., Wei, Y., Wang, Z., Chen, Y.: Intelligent manufacturing in industry 4.0: A case study of Sany heavy industry. *Systems Research and Behavioral Science* **37**(4), 679–690 (2020). <https://doi.org/10.1002/sres.2709>
16. Sokolov, B., Dolgui, A., Ivanov, D.: Optimal control algorithms and their analysis for short-term scheduling in manufacturing systems. *Algorithms* **11**(5) (2018). <https://doi.org/10.3390/a11050057>
17. Spearman, M.L., Woodruff, D.L., Hopp, W.J.: CONWIP: A pull alternative to kanban. *International Journal of Production Research* **28**(5), 879–894 (1990). <https://doi.org/10.1080/00207549008942761>
18. Vespoli, S., Grassi, A., Guizzi, G., Santillo, L.C.: Evaluating the advantages of a novel decentralised scheduling approach in the Industry 4.0 and Cloud Manufacturing era. In: *IFAC-PapersOnLine*. vol. 52, pp. 2170–2176. Elsevier B.V. (sep 2019). <https://doi.org/10.1016/j.ifacol.2019.11.527>