Study of dynamic workload assignment strategies on production performance ⋆

H. Li * **A. K. Parlikad** **

* *Institute for Manufacturing, University of Cambridge, Cambridge, CB3 0FS UK (e-mail: hl433@cam.ac.uk).*
** *Institute for Manufacturing, University of Cambridge, Cambridge, CB3 0FS UK (e-mail: aknp2@cam.ac.uk)*

Abstract: As maintenance has grown to be seen as a prospective tool for production value generation and business performance improvement, it can no longer be considered as isolated from other production activities. Studies have shown that the degradation process of machines is dependent on the operation being performed (e.g., higher workload results in faster degradation). However, the decision-making in maintenance planning with dynamic operation/workload adjustment considerations has not been addressed until recently. Moreover, the existing approaches attempting to tackle this problem have overlooked the fact that dynamics exist in both external production environment and internal production conditions and thus prove to be inefficient to react to unexpected situations arising. This paper has explored the impacts of different workload adjustment strategies on system production performance by a numerical study using agent-based simulation. A detailed discussion is given on the implication of the simulation outcome, based on which some insights into potential future work are also presented.

Keywords: operation-dependent deterioration, workload adjustment, multi-agent systems, condition-based maintenance, simulation

1. INTRODUCTION

The concept of maintenance has evolved over time from a conventional perception where maintenance is considered as a necessary evil that induces excessive costs to the current view where it is seen as a prospective tool for production value generation and business performance improvement (Marais and Saleh, 2009). Maintenance cannot be seen as independent from other production elements as it plays a vital role in generating value for manufacturers. Manufacturing is a complex process involving multiple aspects and steps starting from raw material procurement all the way to product quality control and customer service. These areas used to be treated independently, yielding separate models for each function. It has been widely acknowledged that these models are likely to provide sub-optimal solutions due to the fact that these areas are interrelated (Hadidi et al., 2012). The relationships between maintenance actions and other production elements, such as production scheduling, quality control (Colledani and Tolio, 2012), and human resource management (Bouzidi-Hassini and Benbouzid-Sitayeb, 2013), have been broadly studied, intended for integrated decision-making models.

While various interactions between maintenance scheduling and other aspects of manufacturing process have been introduced into studies seeking system-level performance improvement, one important factor has been overlooked until very recently. Each individual unit in manufacturing is subject to an inevitable degradation process while conducting production tasks and its degradation rate varies with the type of task and the amount of workload assigned to the unit (Celen and Djurdjanovic, 2016). However, most existing literatures on condition-based maintenance build their solutions on the assumption that degradation of machines is a self-evolving process (Hao et al., 2015), overlooking the potential benefits of intervening in machine degradation by adjusting the workload or assigning specific operations. For example, by dynamically controlling the degradation of assets, it is possible to postpone or bring forward a trigger for maintenance to a more preferable time, as opposed to passively waiting for a maintenance request that can be made by any machine at any random time.

In section 2 a more detailed review of existing work on combined maintenance and operation decision-making will be presented to highlight the research gap, and the contribution of this paper is emphasised. Section 3 gives a description on an agent-based simulation model used for numerical studies on how various workload strategies affect system-level performance. The results of the preliminary numerical study are presented in section 4, followed by a discussion on the simulation outcomes. A conclusion on the findings of this paper is given in section 6. The paper is ended with a discussion on potential future work.

2. LITERATURE REVIEW

To the author's knowledge, one of the earliest attempts to supplement traditional maintenance activities with op-

⋆ Acknowledgments to financial support of Cambridge Trust and China Scholarship Council.

eration alternatives is made by Yang et al. (2007), who proposed a model for optimising joint scheduling of maintenance and throughput adjustment operations. The analytical solution for a single-machine is first given and then extended and applied to a simple production system consisting of machines with slow and fast throughput settings, corresponding to a slower and a faster life reduction rate. Generic Algorithm(GA) based optimisation is used to find the optimal joint scheduling of maintenance and throughput adjustment that maximises the expected overall production benefit under constant production targets.

An explicit mathematical model to tie maintenance, production rate, and product quality within the Partially Observable Markovian Decision Process (POMDP) is proposed (AlDurgam and Duffuaa, 2012). This model also takes the basic assumption that higher operation speed results in higher machine deterioration rate, adding an extra link that relates a heavier machine deterioration to a larger likelihood of faulty products. However, the decision-making model proposed by AlDurgam and Duffuaa (2012) is targeted at single-machine systems yet in reality it requires the cooperation of multiple machines to achieve production goals. Moreover, no constraint is considered to obtain the optimal solution while constraints such as production demands are likely to exist in reality.

Another work on maintenance and operation-related actions is conducted by Zhou et al. (2007) where reconfiguration is considered as a means of mitigating production loss caused by machine degradation and failure. A framework is proposed for the integrated decision making of reconfiguration and age-based preventive maintenance for a general single-product two-stage parallel-serial system with reconfigurable capabilities to transfer operations between the two stages, giving guidance on how to couple the reconfiguration action of operation transfer with an age-based preventive or a corrective maintenance. However, it assumes the constraint that the system throughput has to be put at its maximum at every decision epoch.

Several papers published recently have been motivated by the use of clustering tools in semiconductor industry able to perform more than one operations degrading the tools at various rates. Celen and Djurdjanovic (2012) looked into the interactions between degradation, operations, and product quality, and devised a combined operational and maintenance decision-making policy. One unique feature of this model is that it acknowledges the fact that preventive maintenance triggering states are operation-dependent. The goal is to optimise a customisable objective function while attempting to meet the production demand where certain numbers of different products need to be produced within time T. In a later work Celen and Djurdjanovic (2016), the model is expanded on the condition that maintenance conducted at 'less busy shifts' costs less than that at normal working shifts. In both papers, it requires the next task to be assigned to the least degraded machine.

Also targeted at operation-dependent deteriorating systems, Jin (2015) took an analytical approach to explore the structural properties of the objective function, aiming to identify the conditions that can limit the optimal solution to a set of monotone procedures.

Hao et al. (2015) developed a decision support model to actively control the residual life of parallel units to prevent overlapping unit failures by dynamically adjusting the workload assigned to each unit. Specifically, higher workloads are assigned to less healthy units so that the more degraded units can fail even sooner, thus separating its expected failure time from the other units. It is assumed that the instantaneous degradation rate is proportional to the workload by a degradation coefficient β_m. A simulation-based numerical example of dynamic work load adjustment among five identical units is presented and discussed. However, demands are assumed to be constant in this work, whereas in reality demands can be constantly changing. Moreover, it is not specified how the decision-making interval is determined, which has a large influence on the performance of the model since with a small decision interval, some constraints can be released and different solutions might be obtained.

In conclusion, existing research attempts considering the mutual influence between operation and maintenance activities suffer from at least one of the following drawbacks:

(1) targeted at single-component systems and lacks generalisation for a multi-component system;
(2) targeted at pre-planned preventive maintenance instead of condition-based maintenance;
(3) simplified and case-specific assumptions or constraints on workload assignment rules.
(4) not focused on the fact that production is a versatile process with random and sudden changes

Maintenance planning systems with alternative operation considerations found in the aforementioned literature are not designed to be able to fast come up with acceptable solutions to react to sudden changes since the solution would need to be recalculated globally that requires centrally available and continuously updated knowledge about all activities and resources (Mes et al., 2007). However, under certain predefined levels of guidance on perturbations, the previously obtained globally optimised plan can be refined in real time according to local constraints to cope the situation arising (Pach et al., 2014). As a distribution of intelligence and decision-making is obvious in this scenario, centralised decision-making structure is no longer suitable.

This paper is focused on the third gap identified in the previous paragraphs. A preliminary agent-based simulation model is built to study numerically the impacts of different workload assignment strategies on production system performance. The reason for choosing an multi-agent structure is that MAS has have been applied to various aspects in asset management (Muller et al., 2008; Khelifati and Benbouzid-Sitayeb, 2013; Cerrada et al., 2007), and proved to be feasible and widely accepted in terms of decentralised decision-making, which is a potential starting point for addressing the fourth gap mentioned previously.

3. SIMULATION MODEL

A numerical study using the agent-based modelling and simulation environment Netlogo (Wilensky, 1999) is conducted to investigate the impact of different workload assignment strategies on production performance. The ex-

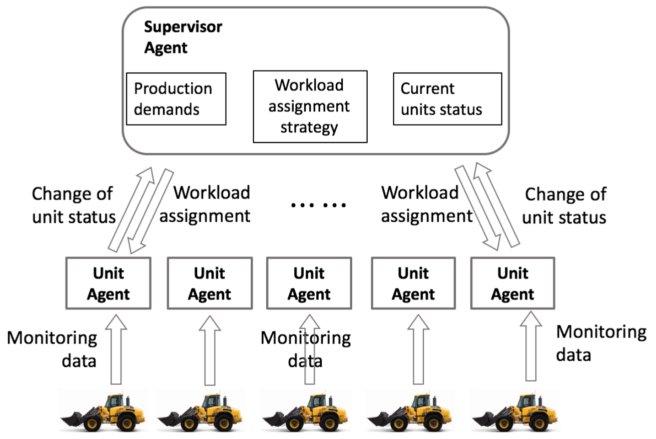

Fig. 1. Agent-based simulation architecture and information flow

tension package developed by Sakellariou et al. (2008) to equip NetLogo with BDI agent is imported to enhance agent communication capability. The simulation considers a system consisting of five identical machines working in parallel to meet constant production demands. This setting is inspired by the work of Chen (2006) where the degradation parameters have been obtained through experiments, and the work of Hao et al. (2015) that also conducts a numerical study to evaluate the performance of the workload adjustment strategy proposed in their paper.

3.1 Multi-agent architecture

The model consists of five machine/unit agents (UAs) and one supervisor agent (SA). The architecture and information flow of the system is presented in Fig. 1. The flowchart of the simulation is shown in Fig. 2

- The UAs continously monitor their own degradation status and at each decision epoch make decisions on whether a preventive maintenance should be conducted according to the level of degradation. They are also obliged to inform the SA of any change in their status.
- The job of the SA is to keep track of the number of UAs that are operating, idling, or undergoing maintenance actions and the current workload assigned to each unit. It is also due to the SA to calculate and reassign the workload of each unit at each decision epoch according to the production demand, number of units available, and the workload assignment strategy chosen.

3.2 Degradation, failure, and maintenance

The assumptions made in the simulation concerning degradation, failure, and maintenance are listed as follows.

(1) Degradation is inevitable as long as a unit is in operation. The instantaneous degradation rate is dependent on the work load assigned to the unit at time t with positive correlation. Degradation signals of a single unit m are generated using the model proposed by Chen (2006), written as

$$dX_m(s_m) = \beta_m ds_m + dW_m(s_m), \tag{1}$$

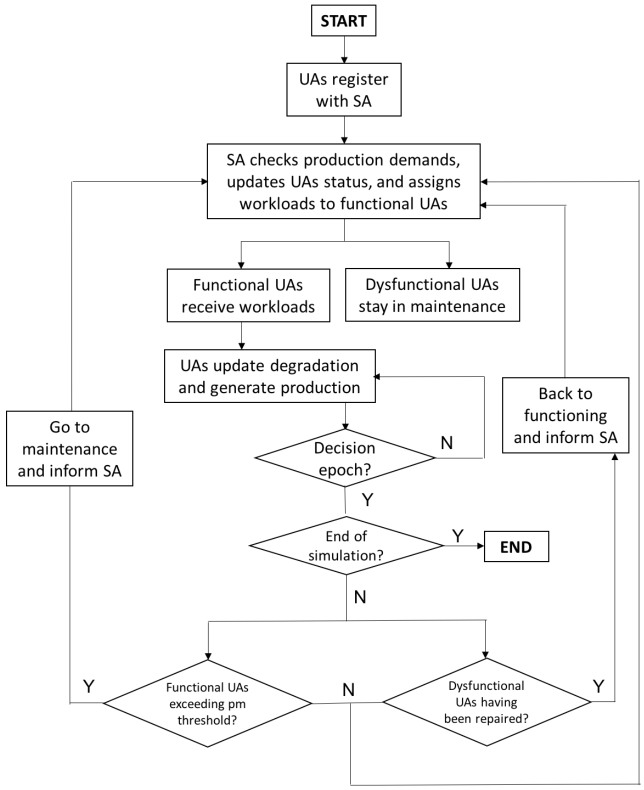

Fig. 2. Flowchart of the agent-based simulation

where $X_m(s_m)$ is the amplitude of the degradation signal of unit m, s_m is the age (the time since its last resetting) of unit m, $E[W_m(s_m)] = 0$, $Var[W_m(s_m)] = \sigma_m^2 s_m$, and β_m is a known constant that captures how fast the degradation of unit m increases with age, which we call the degradation rate coefficient of unit m. Here we admit the fact that due to factors like material inhomogeneity, β_m varies from unit to unit but is generated using a known distribution as shown in equation (2). σ_m is also a known constant.

$$\beta_m \sim N(\kappa, \tau^2) \tag{2}$$

where $\tau/\kappa = cv$, and cv is the coefficient of variation. Following (Chen, 2006), the production time is measured in terms of the number of products fabricated $ds_m(t) = u_m(t)dt$, leading to

$$dX_m(t) = \beta_m u_m(t)dt + dW_m(t), \tag{3}$$

where $u_m(t)$ is the workload assigned to unit m at time t.

(2) Though there may exist other types of failure modes, only unit failure due to degradation is taken into consideration in this work. A unit is considered to have failed once its degradation level exceeds a predefined threshold H and will be put to maintenance immediately.

(3) Unmet production is permanently lost.

(4) Maintenance tasks take a constant period of time and always restore the unit to the as-good-as-new state.

(5) Maintenance resource is infinite, which means multiple machines can undergo repairs simultaneously.

(6) The degradation level of a unit does not affect the quality of the operation performed by the unit.

3.3 Workload assignment strategies

In order to demonstrate and evaluate the effects of workload assignment on system-level production performance, four different strategies are tested in the simulation.

(1) Workloads are assigned uniformly to all functional units.
(2) Workloads are assigned randomly to all functional units. Specifically, all feasible solutions are generated at each decision epoch and one solution is drawn randomly from the option pool. In this study, for simplicity, it is assumed that the workload distributed to each unit can only be multiples of 80 instead of the entire integer space.
(3) Workloads are assigned in order to ensure that the overall system degradation rate at the k^{th} decision epoch t_k, $R(t_k)$ as defined in equation (4) is minimised. In other words, full workload will be assigned to the unit with the smallest β among all functional units until all workloads are distributed.

$$R(t_k) = \sum_{m=1}^{M(t_k)} \beta_m u_m(t_k) \tag{4}$$

where $M(t_k)$ is the number of functional units at t_k.
(4) This strategy assigns the most work to the unit with the largest β. Workloads are assigned according to the ratio of degradation rate coefficients of all functional units as calculated using equation (5). Note that if the calculated workload of the unit with the largest β exceeds the single unit capacity, the surplus workload will be assigned to the unit with the second largest β and so forth until all workloads are distributed;

$$u_m(t_k) = D\frac{\beta_m}{\sum_{i=1}^{M(t_k)} \beta_i} \tag{5}$$

where D is the production demand.

In all the strategies, if the production demand exceeds the sum of capacity of all functional units, every unit will be assigned its full workload U_m.

3.4 Simulation procedure and parameter settings

The parameters used in the simulation are given here in table 1.

Table 1. Parameter settings

U_m	κ	cv	H
1440/d	5.97E-8	0.1	0.004

In this study, the percentage loss of production is chosen as the key performance index (KPI) for performance evaluation of different strategies. In order to evaluate the four strategies in various scenarios, two control variables are used.

(1) Production demand D: the production demand is assumed to be constant over the simulation period and takes one of the the two values (5040 and 6480 parts/d) in each scenario. As the unit capacity has been set to be 1440 parts/d, it requires at least 4 units to be functional for a 5040 parts/d demand, and all 5 units for a demand of 6480 parts/d.

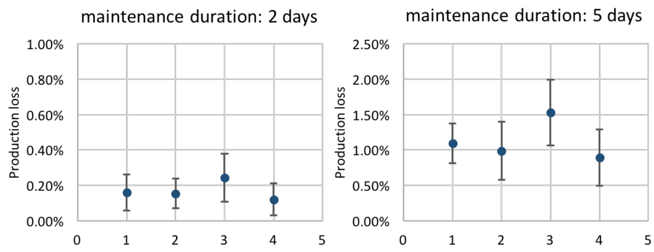

Fig. 3. Production loss of different strategies for a production demand of 5040 parts/d

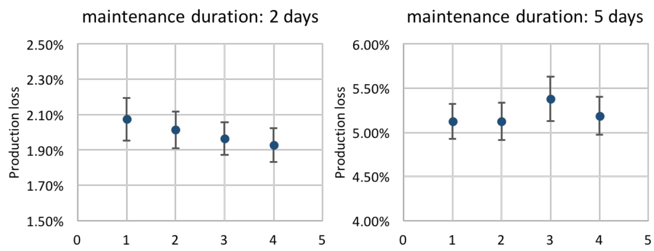

Fig. 4. Production loss of different strategies for a production demand of 6480 parts/d

(2) Preventive maintenance duration pm: the amount of time required for repair takes one of the two values (2, and 5 days) in each scenario.

The length of the simulation period is set to be one year (365 days) and the decision epoch is one day. For each strategy under each of the 4 conditions, we ran 30 repetitions of simulation.

4. RESULTS

The simulated results for different levels of production demands are plotted in Fig. 3, and 4, where the error bars represent the standard deviation.

The following observations can be made out of the plots:

(1) Under the same production demand, all four strategies perform worse in scenarios with a 5-day maintenance duration than they do in those with a 2-day maintenance duration, since more machine downtime results in more unmet production. Similarly, under the same maintenance duration, higher production loss is observed with higher demand. It can also be seen that the gap between the same strategy under different maintenance durations gets larger with increasing production demands, which can be explained by the fact that a higher production demand is more sensitive to machine downtime since it requires more machines to be functional simultaneously.
(2) There is no dominant strategy that yields the best system performance under all conditions, nor is there any strategy that always lags behind
  - Specifically, strategy (4) gives the least production loss in most scenarios, but is out performed by strategy (1) when the maintenance lasts 5 days and the production demand is set as 6480 parts/d.

- Strategy (3) yields the most production loss in three out of four settings, except for when the maintenance duration is 2 days and the production demand is set to be 6480 parts/d where it is the second best strategy.

5. DISCUSSION

From the results presented above, some counterintuitive facts can be observed. First, strategy (3) is considered to be the most cost-effective strategy due to the characteristic of this strategy that overall the production system degrades at the minimum rate implied by equation (4). However, it can be noticed that it has never outperformed the other three strategies in any of the scenarios studies. Second, strategy (4), which yields the highest overall degradation rate, has led to the least production loss under three out of four production settings.

This section gives a detailed discussion on two possible factors that have contributed to strategy (4) outperforming strategy(3) in most scenarios as shown in Fig. 3 and 4. Here, the scenario with 6480 parts/d production demand is taken as an example, since the case where demand is 5040 parts/d can be explained using the same approach.

Table 2. Degradation rate coefficients of units

Unit id	1	2	3	4	5
$\beta(10^{-8})$	5.2523	5.2745	6.2552	5.8443	5.3468

Table 3. Workload distribution (parts/d)

Unit id	1	2	3	4	5
strategy (3)	1440	1440	720	1440	1440
strategy (4)	1217	1222	1440	1363	1238

Table 4. Total production of units (parts)

Unit id		1	2	3	4	5
pm 2d	(3)	508320	508320	288192	505536	508320
	(4)	441359	442961	505296	482613	447370
pm 5d	(3)	482400	482400	309696	481104	482400
	(4)	432695	433999	475056	462650	438111

Table 5. Times of preventive maintenance

Unit id		1	2	3	4	5
pm 2d	(3)	6.00	6.00	4.00	6.93	6.00
	(4)	5.00	5.00	6.97	6.60	5.30
pm 5d	(3)	6.00	6.00	4.00	6.33	5.97
	(4)	5.00	5.00	6.97	6.00	5.03

(1) The degradation rate coefficients for unit 1 to 5 are given in table 2, which shows that unit 1 has the smallest β, hereto referred to as 'the best unit', and that unit 3 has the largest β, here to referred to as 'the worst unit'. Following the rules defined in strategies (3) and (4), while all units are functional, the workload distribution for a demand of 6480 parts/d is given in table 3. The simulation results of the total production of each unit within a one-year period under the two strategies are presented in table 4. Ideally, the ratio of the total production r^p_{ij} of unit i and unit j should be approximately the same as that of the workload assigned r^w_{ij} to these two units, which is not the case here as it can be calculated that in strategy (3), r^p_{31}(0.5669 for pm = 2d and 0.6420 for pm = 5d) is larger than r^w_{31} = 0.5, indicating that the best unit is being used less frequently and the worst unit more frequently than expected. This is due to the fact that much more workload is assigned to the best unit, incurring more degradation thus more failures of the best unit despite of its smaller degradation rate coefficient, which can also be seen from the corresponding times of failure unit 1 and 3 where unit 1 have gone through more preventive maintenance than unit 3 for both pm durations, as shown in table 5. In the case of strategy (4), opposite results can be observed.

(2) Compared with strategy (4), strategy (3) tends to give rise to more overlapping machine failures (more than one machine undergoing maintenance at the same time): strategy (3) yields overlapping failures equivalent to 6.67 and 35.70 days of two simultaneous machine failures for pm =2d and pm =5d, respectively, while strategy(4) yields overlapping failures equivalent to 5.73 and 30.20 days of two simultaneous machine failures for pm = 2d and pm = 5d, respectively. When the production demand is 6480 parts/d and the unit capacity is 1440 parts/d, a single machine failure causes a production loss of $6480 - 1440 \times 4 = 720$ parts/d whereas two machines failing simultaneously results in a loss of $6480 - 1440 \times 4 =$ 2160 parts/d, which is more than twice the amount of production loss caused by as single machine failure.

6. CONCLUSION

The following conclusions can be drawn from the preliminary numerical study presented above.

(1) The performance of a production system comprised of identical units working in parallel, measured here as production loss within a specific time period, is largely affected by the workload assignment strategy applied.

(2) Under various production settings (e.g. customer demands, maintenance duration, etc.), different maintenance strategies result in different production yields, which implies that there is no predefined universal workload assignment rule that has the optimal performance under any circumstance. Since production is a versatile process with random and sudden changes, the workload assignment needs to be adjusted accordingly to prevent system performance from straying away from the optimal level.

7. FUTURE WORK

(1) The only performance matrix used to evaluate the workload assignment strategies is the production loss, as no additional data is needed to calculate its value. However, other performance matrices, such as total cost of ownership, system reliability, etc., are also being widely adopted by the industry. It is worth exploring the performance of the strategies using other performance matrices.

(2) It has been assumed in the simulation that maintenance resource is unlimited and that all maintenance

is perfect. This assumption can be released for a closer representation of the reality.

(3) The numerical study has justified the necessity of a decision-support system for workload adjustment to counter varying conditions over the production phase, which can be further broken down into three subtasks:
   - It has been proposed that the multi-agent system is a feasible solution for the decentralisation of intelligence and responsibility. Thus further research is needed to develop a multi-agent structure which is both part of the decision-making strategy as well as the skeleton that carries forward the decision-making algorithms.
   - This paper has discussed a specific case where the degradation rate is proportional to the workload, however, a general formulation of the operation-dependent/workload-dependent degradation process is needed;
   - The optimisation approach for combined maintenance and dynamic workload adjustment strategies.

(4) In the proposed methodology, maintenance threshold is assumed to be constant for all units and the intention is to investigate the role of workload distribution in shaping maintenance planning and production performance. However, it is worth exploring how maintenance threshold settings affect the suitability of different workload adjustment strategies under various conditions.

ACKNOWLEDGEMENTS

This research work was performed within the context of Sustain-Owner ("Sustainable Design and Management of Industrial Assets through Total Value and Cost of Ownership"), a project sponsored by the EU Framework Programme Horizon 2020, MSCA-RISE-2014: Marie Skodowska-Curie Research and Innovation Staff Exchange (RISE) (grant agreement number 645733-Sustain-Owner-H2020-MSCA-RISE-2014).

REFERENCES

AlDurgam, M.M. and Duffuaa, S.O. (2012). Optimal joint maintenance and operation policies to maximise overall systems effectiveness. *International Journal of Production Research*, 51(5), 1–12.

Bouzidi-Hassini, S. and Benbouzid-Sitayeb, F. (2013). Multi-agent based joint production and maintenance scheduling considering human resources. *2013 5th International Conference on Modeling, Simulation and Applied Optimization, ICMSAO 2013*, 0–4.

Celen, M. and Djurdjanovic, D. (2016). Integrated Maintenance Decision-Making and Product Sequencing in Flexible Manufacturing Systems. *accepted for publication in the ASME Journal of Manufacturing Science and Engineering*, 137(August 2015), 1–15.

Celen, M. and Djurdjanovic, D. (2012). Operation-dependent maintenance scheduling in flexible manufacturing systems. *CIRP Journal of Manufacturing Science and Technology*, 5(4), 296–308.

Cerrada, M., Cardillo, J., Aguilar, J., and Faneite, R. (2007). Agents-based design for fault management systems in industrial processes. *Computers in Industry*, 58(4), 313–328.

Chen, Y. (2006). Quality-Oriented-Maintenance for Multiple Interactive Tooling Components in Discrete Manufacturing Processes. *IEEE Transactions on Reliability*, 55(1), 123–134.

Colledani, M. and Tolio, T. (2012). Integrated quality, production logistics and maintenance analysis of multistage asynchronous manufacturing systems with degrading machines. *CIRP Annals - Manufacturing Technology*, 61(1), 455–458.

Hadidi, L.A., Turki, U.M.A., and Rahim, A. (2012). Integrated models in production planning and scheduling, maintenance and quality: a review. *International Journal of Industrial and Systems Engineering*, 10(1), 21.

Hao, L., Liu, K., Gebraeel, N., and Shi, J. (2015). Controlling the Residual Life Distribution of Parallel Unit Systems Through Workload Adjustment. *IEEE Transactions on Automation Science and Engineering*, (October), 1–11.

Jin, L. (2015). Optimal decision procedure for an operation-dependent deteriorating system. *Applied Stochastic Models in Business and Industry*, 31(3), 394–404.

Khelifati, S.L. and Benbouzid-Sitayeb, F. (2013). A multi-agent scheduling approach for the joint scheduling of jobs and maintenance operations in the flow shop sequencing problem. *International Journal of Intelligent Engineering Informatics*, 2(1), 47–70.

Marais, K.B. and Saleh, J.H. (2009). Beyond its cost, the value of maintenance: An analytical framework for capturing its net present value. *Reliability Engineering and System Safety*, 94(2), 644–657.

Mes, M., van der Heijden, M., and van Harten, A. (2007). Comparison of agent-based scheduling to look-ahead heuristics for real-time transportation problems. *European Journal of Operational Research*, 181(1), 59–75.

Muller, A., Crespo Marquez, A., and Iung, B. (2008). On the concept of e-maintenance: Review and current research. *Reliability Engineering and System Safety*, 93(8), 1165–1187.

Pach, C., Berger, T., Bonte, T., and Trentesaux, D. (2014). ORCA-FMS: A dynamic architecture for the optimized and reactive control of flexible manufacturing scheduling. *Computers in Industry*, 65(4), 706–720.

Sakellariou, I., Kefalas, P., and Stamatopoulou, I. (2008). Enhancing NetLogo to Simulate. In *Proceedings - Artificial Intelligence: Theories, Models, and applications, 5th Hellenic Conference on AI, SETN 2008*, 263–275. Syros, Greece.

Wilensky, U. (1999). Netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. URL `http://ccl.northwestern.edu/netlogo/`.

Yang, Z., Djurdjanovic, D., and Ni, J. (2007). Maintenance scheduling for a manufacturing system of machines with adjustable throughput. *IIE Transactions*, 39(12), 1111–1125.

Zhou, J., Djurdjanovic, D., Ivy, J., and Ni, J. (2007). Integrated reconfiguration and age-based preventive maintenance decision making. *IIE Transactions*, 39(12), 1085–1102.