Prescriptive Maintenance through Workload Allocation for Synchronizing Parallel Machines

Sacha Todoskoff[1] Bruno Castanier[2] Niclas Björsell[3] T. Sunil Kumar[4] Sylvain Verron[5]

[1,2,5] *LARIS, SFR MATHSTIC, University of Angers, France*
sacha.todoskoff@etud.univ-angers.fr
bruno.castanier@univ-angers.fr
sylvain.verron@univ-angers.fr
[1,3,4] *Dept of Electronics, Mathematics and Natural Sciences, University of Gävle, Sweden*
sacha.todoskoff@hig.se
niclas.bjorsell@hig.se
sunilkumar.telagam.setti@hig.se

ABSTRACT

This paper addresses workload allocation for parallel machines subject to stochastic degradation in order to improve maintenance synchronization. In such systems, maintenance actions are often performed simultaneously across machines, which creates a trade off between premature maintenance and unexpected failures. Existing workload allocation strategies mainly focus on reducing failure probability or balancing degradation but do not explicitly aim at synchronizing maintenance conditions.

To address this issue, this paper proposes a framework combining stochastic degradation modeling and workload optimization. Machine degradation is modeled using a Gamma process in which the degradation rate depends on the assigned workload. A Sequential Quadratic Programming optimization is used to dynamically allocate workloads in order to maximize the probability that all machines reach the maintenance window defined by preventive and failure thresholds at the same time. Remaining Useful Life estimates are used to trigger workload reallocations and maintenance decisions.

Monte Carlo simulations compare the proposed strategy with a deterioration based workload allocation method from the literature. The results show that the proposed approach maintains a low proportion of corrective maintenance while achieving a high proportion of preventive maintenance with fewer workload reallocations.

1. INTRODUCTION

Parallel machine systems are widely used in industrial environments where production continuity is critical. In this paper, a parallel machine system refers to a set of machines that can contribute to the same production requirement and among which the workload can be redistributed. The machines are not assumed to be identical in terms of components, degradation parameters, or health evolution. Instead, each machine is represented by its own degradation model, while the workload allocation variable defines the proportion of the total production assigned to each machine. The proposed framework therefore applies to systems in which operations are sufficiently substitutable to make workload redistribution meaningful.

In many sectors, maintenance opportunities are rare and fixed in advance, sometimes occurring only once or twice per year due to operational or contractual constraints. During these limited shutdown periods, all machines must be maintained simultaneously to avoid production disruption and excessive coordination costs.

This context creates a practical dilemma. Machines should not be maintained too early, as this would waste useful life. However, waiting too long increases the risk of unexpected failures, which may reduce system capacity and compromise production targets. The key challenge is therefore not only to decide when to maintain each machine, but also to control their degradation so that maintenance actions can be synchronized without increasing failure risk.

Early work on predictive maintenance in multi-component systems established condition-based and reliability-driven optimization frameworks (Van Horenbeek & Pintelon, 2013; Tian & Liao, 2011). These models improved maintenance timing decisions but generally treated degradation trajectories independently across units.

Stochastic degradation processes such as Wiener and Gamma models have since been introduced to better represent cumulative wear dynamics (Chen & Hao, 2025; Hao, Chen, & Wang, 2025). In parallel, finite-horizon degradation control strategies (Björsell & Dadash, 2021) have shown that workload adjustments can influence degradation evolution when physical models are available.

Joint production–maintenance optimization approaches (Sharifi, Ghaleb, & Taghipour, 2023; Askri, Zied, & Rezg, 2017) have demonstrated the benefits of integrating operational and maintenance decisions. More recently, post-prognostics decision-making strategies (Zuo, Cadet, Li, Berenguer, & Outbib, 2023; Zuo et al., 2025) have shown that dynamic workload allocation can balance degradation across systems. However, the explicit objective of synchronizing maintenance under fixed and infrequent shutdown windows remains largely unexplored.

This paper addresses the challenge of synchronizing maintenance actions under fixed shutdown windows while avoiding premature maintenance and reducing the risk of unexpected failures. A unified framework is proposed that combines stochastic degradation modeling using a Gamma process with dynamic workload reallocation optimized through Sequential Quadratic Programming (SQP). The objective is to reduce dispersion in machine health states over time, so that maintenance actions can be performed simultaneously while minimizing corrective interventions and preserving system reliability.

The paper is organized as follows. Section 2 reviews related work in greater detail and clarifies the research gap. Section 3 presents the stochastic degradation model and maintenance thresholds. Section 4 introduces the workload optimization and maintenance decision framework. Section 5 discusses the numerical experiments and the comparison with another workload optimization from an article. Finally, Section 6 concludes the paper and outlines future research directions.

2. LITERATURE REVIEW: FROM DEGRADATION MODELING TO WORKLOAD-CONTROLLED SYNCHRONIZATION

This section reviews the main developments in degradation modeling, maintenance optimization, and workload allocation for parallel systems. The presentation follows a progression from classical maintenance optimization to recent health-aware operational control strategies.

Early research on maintenance optimization focused on multi-component systems and condition-based policies. (Van Horenbeek & Pintelon, 2013) proposed a dynamic predictive maintenance framework for complex systems, integrating degradation information into maintenance decision-making. Similarly, (Tian & Liao, 2011) developed a condition-based maintenance optimization model using a proportional hazards approach to compute conditional failure probabilities and minimize long-term costs.

These contributions provide strong theoretical foundations for maintenance planning under uncertainty. However, they generally consider system components from a reliability perspective and do not explicitly address workload redistribution in parallel machine environments.

A multi-level decision-making perspective was later introduced by Huynh et al. (Huynh, Barros, & Berenguer, 2015), who structured predictive maintenance decisions across strategic and operational levels. While this hierarchical framework enhances coordination, it does not directly consider load balancing as a control variable for degradation synchronization.

To better represent real industrial systems, stochastic degradation processes such as Wiener and Gamma models have been widely adopted.

(Chen & Hao, 2025) proposed a condition-based operation and maintenance strategy for load-sharing systems using a Wiener process to model degradation. Their work shows how workload directly affects degradation evolution and maintenance timing. (Hao et al., 2025) further extended this approach by jointly optimizing operation and maintenance strategies under hybrid load conditions.

These studies clearly demonstrate that workload allocation influences degradation dynamics. However, the primary objective remains reliability improvement or cost reduction, rather than explicitly steering multiple machines toward synchronized maintenance thresholds.

As industrial systems became more integrated, researchers began combining production scheduling and maintenance planning in parallel machine environments. (Sharifi et al., 2023) developed a joint scheduling and maintenance optimization model that accounts for deterioration, unexpected breakdowns, and condition-based maintenance. (Askri et al., 2017) and (Tarek, Zied, & Nidhal, 2017) proposed joint production–maintenance strategies for parallel leased machines, minimizing operational and maintenance-related costs. In a related study, the scheduling problem with deteriorating processing times and periodic maintenance was addressed in (Pérez, Ambati, & Torres, 2018), emphasizing the interaction between job sequencing and maintenance intervals.

These works show that integrating production and maintenance decisions significantly improves overall system performance. However, degradation is often modeled using simplified failure rates or deterministic deterioration, and synchronization of maintenance is typically imposed structurally rather than induced through dynamic health control.

Beyond predictive maintenance, recent research has moved toward prescriptive maintenance strategies, where degradation information is actively used to guide operational decisions. (Esposito, Castanier, & Giorgio, 2022) proposed a prescriptive block replacement policy for degrading systems, highlighting the economic benefits of grouped interventions. However, the control of degradation dispersion across parallel machines remains unaddressed.

Another important development concerns the integration of prognostics into operational decision-making. (Herr, Nicod, & Varnier, 2014) proposed a prognostics-based scheduling approach that adapts task allocation across distributed platforms using Remaining Useful Life (RUL) forecasts. (Sharifi et al., 2023) also incorporated condition monitoring into joint scheduling decisions.

More recently, (Zuo et al., 2023) developed a post-prognostics decision-making strategy for load allocation in stochastically deteriorating multi-stack systems. Their later work (Zuo et al., 2025) extended this approach to jointly optimize load allocation and maintenance scheduling in transportation systems. These studies demonstrate that dynamic load redistribution can balance degradation and improve system reliability.

Although these contributions effectively link health monitoring with operational control, their objective functions focus on reliability, availability, or cost minimization. They do not explicitly aim to reduce degradation dispersion in order to achieve synchronized maintenance under fixed shutdown windows.

From a broader systems perspective, (Domingo & Aguado, 2015) introduced the Overall Environmental Equipment Effectiveness (OEEE) metric, emphasizing the importance of integrating operational efficiency and sustainability. While not directly focused on degradation control, this work highlights the need for coordinated decision-making across production and maintenance dimensions.

Overall, the literature provides valuable building blocks: stochastic degradation modeling (Chen & Hao, 2025; Hao et al., 2025), predictive maintenance optimization (Van Horenbeek & Pintelon, 2013; Tian & Liao, 2011), joint production and maintenance scheduling (Sharifi et al., 2023; Askri et al., 2017), and dynamic load allocation based on prognostics (Zuo et al., 2023, 2025).

However, these elements are rarely combined with the explicit objective of actively reducing health dispersion among parallel machines so that they reach maintenance thresholds at approximately the same time. In systems where maintenance windows are fixed, this synchronization objective becomes critical.

The present paper addresses this gap by integrating Gamma-process degradation modeling with dynamic workload optimization. The goal is not only to preserve reliability but also to deliberately steer degradation trajectories toward convergence, enabling simultaneous maintenance while minimizing corrective interventions.

3. DEGRADATION MODEL AND MAINTENANCE THRESHOLDS

The degradation of each machine i is modeled as a Gamma process with shape parameter α_i and scale parameter β_i. The degradation variable $D_i(t)$ is defined at the machine level. It can be interpreted either as the degradation of a critical component or as an aggregated health indicator summarizing the condition of several components. The degradation increment $\Delta D_i(\tau)$ at time step $k \cdot \tau$ is given by:

$$\Delta D_i(\tau) \sim gamma(\cdot; \alpha_i \cdot L_i(k \cdot \tau) \cdot \tau, \beta_i), \qquad (1)$$

where $L_i(k \cdot \tau)$ is the workload of machine i at time step $k \cdot \tau$, and τ is the time interval. The goal of the approach is to set this load at the optimal value. The mean and variance of the degradation increment are:

$$\text{Mean: } \mu = \frac{\alpha_i}{\beta_i} L_i(k \cdot \tau) \cdot \tau, \quad \text{Variance: } \sigma^2 = \frac{\alpha_i}{\beta_i^2} L_i(k \cdot \tau) \cdot \tau. \qquad (2)$$

The degradation level at the next time step can therefore be expressed as:

$$D_i((k+1) \cdot \tau) = D_i(k \cdot \tau) + \Delta D_i(\tau) \qquad (3)$$

Existing studies, such as those by (Zuo et al., 2023, 2025), often assume predefined maintenance thresholds (s_m) and failure thresholds (s_f) to determine an optimal interval within which maintenance should be performed. If a machine is maintained before reaching s_m, part of its useful lifetime is wasted, leading to unnecessary maintenance. Conversely, if its degradation level exceeds s_f, a failure occurs, thereby reducing the system's production capacity. The ideal strategy, is to dynamically adjust the workload ($L_i(k \cdot \tau)$) assigned to each machine so that their degradation levels converge to the interval $[s_m, s_f]$ at the time of simultaneous maintenance.

Maintenance actions are modeled at the machine level. When a machine is maintained, it is assumed to be restored to an as-good-as-new state with respect to the degradation indicator considered in the model. This assumption can represent either a full overhaul of the machine or the replacement of the critical component responsible for the machine-level degradation. The objective of this paper is not to optimize component-level replacement policies, but to study how workload allocation can synchronize machine-level maintenance decisions.

Figure 1 illustrates the degradation evolution of three machines modeled by Gamma processes with respective parameters α_i and β_i, operating under an equal load of $1/3$. The green and magenta dotted lines denote the preventive and failure thresholds s_m and s_f. Degradation is observed at eight time steps with interval $\tau = 10$. At time 70, machine 1 exceeds the failure threshold and enters the red zone, corresponding to failure and therefore to be strictly avoided. Machine 2 lies between s_m and s_f, within the desirable range for preventive maintenance (green zone). In contrast, machine 3 remains below s_m, performing maintenance at this stage would lead to over-maintenance, corresponding to the yellow zone.

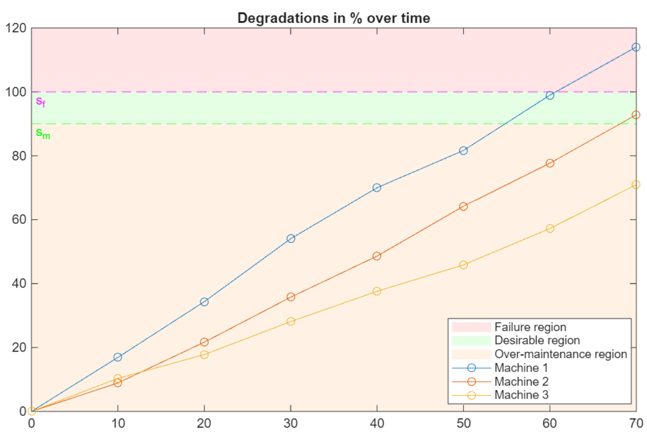

Figure 1. Degradation over time for 3 machines

A classical quantity in the maintenance optimization framework is the RUL. We introduce the a–RUL percentile as the time RUL_i defines as:

$$P(D_i(a\text{–}RUL_i) < s_f) = a \tag{4}$$

The probability $P(D_i(a\text{–}RUL_i) < s_f)$ is computed using the cumulative distribution function (CDF) of the Gamma distribution and like in (4) it's equal to a:

$$F\left(s_f - D_i(t + a\text{–}RUL_i), \alpha_i \cdot L_i(t) \cdot (a\text{–}RUL_i), \beta_i\right) = a \tag{5}$$

where $F\left(s_f - D_i(t + a\text{–}RUL_i), \alpha_i \cdot L_i(t) \cdot (a\text{–}RUL_i), \beta_i\right)$ is the CDF of the Gamma distribution with shape $k = \alpha_i \cdot L_i(t) \cdot (T_M - t)$ with T_M, the maintenance time and $x = s_f - D_i(a\text{–}RUL_i)$:

$$F(x, k, \beta_i) = \frac{\beta_i^k}{\Gamma(k)} \int_0^x t^{k-1} e^{-\beta_i \cdot t}\, dt \tag{6}$$

This value is used as a decision parameter to determine whether workload reallocation should be performed and whether maintenance is required. If the minimum a–RUL_i across all machines is less than the observation interval τ, the system stops to avoid failure. If the difference between the maximum and minimum RUL values exceeds a threshold s_{RUL}, workloads are reallocated to balance the degradation rates across machines.

4. WORKLOAD OPTIMIZATION AND COMPARISON METHODOLOGY

4.1. Workload Optimization Problem

The proposed approach uses workload allocation as a control variable to synchronize machine-level maintenance decisions. Maintenance resources are assumed to be available when a simultaneous maintenance action is triggered. Maintenance costs are not explicitly modeled in the current objective function, which focuses on maximizing the probability of synchronized maintenance in the interval $[s_m, s_f]$ at the maintenance time T_M:

$$P_{sys} = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} P(s_m \le D_i(T_M) < s_f) \tag{7}$$

The probability P_i that machine i has a degradation level within the interval $[s_m, s_f]$ at the maintenance time T_M is given by:

$$P_i = P(D_i(T_M) < s_f) - P(D_i(T_M) < s_m) \tag{8}$$

The probability P_i is computed using the CDF of the Gamma distribution:

$$\begin{aligned} P_i = {} & F\left(s_f - D_i(t), \alpha_i \cdot L_i(t) \cdot (T_M - t), \beta_i\right) \\ & - F\left(s_m - D_i(t), \alpha_i \cdot L_i(t) \cdot (T_M - t), \beta_i\right) \end{aligned} \tag{9}$$

The computation of P_i assumes that the workload selected at the current inspection time remains constant over $[t, T_M]$. This is a local prediction assumption. Since the optimization could be repeated at each inspection time, future workload reallocations can still occur when new degradation observations become available.

When computing products of probabilities, numerical instability may arise due to the accumulation of extremely small values. During the optimization process, small improvements or deteriorations in the objective value must be reliably detected. However, multiplying many probabilities can generate values close to the limits of floating-point precision. To mitigate this issue, the base-10 logarithm of the product is used:

Algorithm 1

```
Input: Current degradation levels D_k, workloads L_i(k · τ), thresholds s_m, s_f, τ, s_RUL
Output: Updated workloads L_i((k + 1) · τ) or maintenance decision
Compute a–RUL_i for all machines i = 1, . . . , N
if min(a–RUL_i) > τ then
    if max(a–RUL_i) − min(a–RUL_i) > s_RUL then
        Reallocate workloads using SQP to maximize log(P_sys)
        Update L_i(k · τ) with optimized workloads
    else
        Continue without workload reallocation
    end if
    Update degradation: D_i((k + 1) · τ) = D_i(k · τ) + ΔD_i(τ)
else
    Reallocate workloads using SQP to maximize log(P_sys)
    Update L_i(k · τ) with optimized workloads
    if min(a–RUL_i) < τ then
        Stop production for maintenance
    else
        Update degradation: D_i((k+1)·τ) = D_i(k·τ) + ΔD_i(τ)
    end if
end if
```

$$\log(P_{sys}) = \sum_{i=1}^{N} \log P(s_m \leq D_i(T_M) < s_f). \tag{10}$$

The total workload allocated to all machines is constrained to be equal to 1. The workload variable $L_i(t)$ is continuous and represents the proportion of the total production demand assigned to machine i. A value of 0 indicates that the machine is temporarily stopped due to excessive degradation. So, the workload allocation is subject to these constraints:

$$\sum_{i=1}^{N} L_i(t) = 1, \quad 0 \leq L_i(t) \leq 1. \tag{11}$$

As shown in equation (10), the maintenance time T_M is a key parameter for evaluating the system probability. In the proposed approach, several candidate maintenance times are evaluated. It's possible to perform a maintenance at each observation so we could try the optimization for a large set of maintenance time.

To reduce the computational cost, the predicted remaining useful life $a\text{–}RUL_i$ is used to define a relevant search interval. Only candidate maintenance times between the minimum and the maximum values of $a\text{–}RUL_i$ across all machines are considered. The overall decision process of the proposed approach is summarized in Algorithm 1.

4.2. Comparison Methodology

The proposed approach is compared with the deterioration-based strategy from (Zuo et al., 2023), which uses a deterioration objective function to minimize:

$$\begin{aligned} F_{det} = \omega_1 \cdot \frac{\sum_{i=1}^{n} (D_i(k \cdot \tau) \cdot P(D_i((k+1) \cdot \tau) > s_f)}{\sum_{i=1}^{n} D_i(k \cdot \tau)} \\ + \omega_2 \cdot Var(D_i((k+1) \cdot \tau)), \end{aligned} \tag{12}$$

where $P(D_i((k+1) \cdot \tau) > s_f)$ is the probability of failure at the next inspection. $Var(D_i((k+1) \cdot \tau))$ is the Variance of the degradation of all machines at the next inspection. The optimization follows the same constraints as it is in the previous optimization and is detailed in (11).

The comparison focuses on the ability to balance degradation rates across machines, the impact on the distribution of machines across corrective, preventive, and over-maintenance categories, and the sensitivity to workload reallocation intervals and maintenance thresholds. Two weighting parameters, ω_1 and ω_2, are used to balance the influence of the degradation level, the probability of failure, and the degradation variability at the next step. In this paper, we do not attempt to determine the optimal weights for our problem. Instead, we assume that both parameters have equal influence and therefore set $\omega_1 = \omega_2 = 0.5$.

The workload allocation is optimized using SQP. In the proposed decision process, workload reallocation is only triggered if the RUL dispersion exceeds the threshold s_{RUL}. This trigger can be seen in Algorithm 1 in the second If, the dispersion is the difference between the highest and the lowest value of $a\text{–}RUL_i$. In the reference study, workload reallocation is performed at every time step.

Another major difference concerns the maintenance decision policy. In (Zuo et al., 2025) there is no predictive policy. They propose a preventive maintenance when the degradation is above the maintenance threshold and a corrective maintenance when the degradation is above the failure threshold. Whenever one machine exceeds the maintenance threshold, simultaneous maintenance is performed for all machines.

To isolate the effect of the workload reallocation strategy from the maintenance policy, a mixed approach is also considered. In this configuration, workload reallocation is performed by minimizing $F_{deterioration}$, while simultaneous maintenance is triggered when the minimum value of $a\text{–}RUL_i$ across all machines becomes smaller than the observation interval τ.

5. SIMULATION RESULTS AND SENSITIVITY ANALYSIS

The performance of the proposed maintenance strategy is evaluated through Monte Carlo simulations and compared with the strategy proposed by (Zuo et al., 2023) and a mixed approach. The analysis focuses on the distribution of machines among three maintenance categories: Corrective Maintenance (CM), Preventive Maintenance (PM), and Over-Maintenance (OM). Those distribution are evaluated in percentage.

The results are obtained from 1000 simulated lifecycles. Each lifecycle starts from the beginning of the machines' operation and ends when a simultaneous maintenance action is triggered. All simulations are initialized with $D_i(0) = 0$ and a uniform initial workload allocation $L_i(0) = 1/N$. The failure threshold is fixed to $s_f = 100$. Each Monte Carlo lifecycle is generated using an independent random seed, and the same seeds are used across the compared strategies.

Table 1. Gamma degradation parameters used in the numerical experiments

Machine	α_i	β_i	α_i/β_i
1	6	3	2
2	16	6	2.67
3	4.5	2	2.25

Table 1 reports the Gamma parameters used in the numerical experiments. The parameters are chosen so that the expected degradation coefficient remains between 2 and 3 for all machines under unit workload, while preserving heterogeneous degradation behaviors.

Each machine follows a stochastic degradation process characterized by different Gamma shape and scale parameters in order to represent heterogeneous degradation behaviors. Since the machines operate in parallel and are expected to experience similar operating conditions, the mean degradation increment was constrained between 2 and 3 so that the degradation trajectories remain relatively close.

The experiments were conducted for systems composed of multiple machines operating in parallel. In order to clearly illustrate the behavior of the maintenance strategies, the following analysis focuses on a system composed of three machines. This configuration allows a clear comparison between the different approaches while keeping the computational effort reasonable.

Different inspection intervals τ were considered to analyze the effect of observation and reallocation frequency on maintenance decisions. Table 2 presents the percentage of machines in each maintenance category for three representative values of τ. The maintenance threshold is fixed to $s_m = 90$ and the system contains three machines. And the percentile used for the a–RUL is 0.9, so it's the 0.9–RUL.

The results show that increasing the inspection interval τ leads to a noticeable increase in corrective maintenance for the strategy proposed by (Zuo et al., 2023). This behavior is expected since less frequent inspections increase the probability that machines exceed the failure threshold between two observations.

Table 2. Impact of the inspection interval τ on maintenance categories ($s_m = 90$, 3 machines, 1000 simulations, 0.9–RUL)

τ	CM	PM	OM
	(Zuo et al., 2023)		
10	0.005	0.820	0.175
15	0.064	0.765	0.171
20	0.173	0.621	0.206
	Mixed approach		
10	0.001	0.917	0.073
15	0.005	0.590	0.405
20	0.004	0.539	0.457
	Proposed approach		
10	0.013	0.917	0.070
15	0.007	0.623	0.370
20	0.004	0.573	0.422

In contrast, both the mixed strategy and the proposed SQP-based approach show a much smaller increase in corrective maintenance. Instead, a moderate increase in over-maintenance can be observed as τ increases. This indicates that the proposed policy is more conservative, prioritizing early preventive interventions over the risk of unexpected failures. From a practical perspective, this behavior is desirable since the cost of early maintenance is typically significantly lower than the cost associated with machine failure.

To better understand the trends observed in Table 2, the distribution of the final degradation levels of all machines was analyzed over the 1000 simulations and for the inspection interval $\tau = 15$ and can be seen in Figure 2.

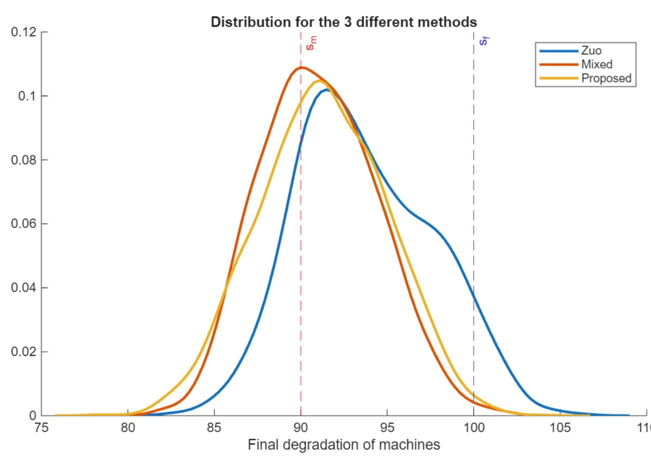

Figure 2. Distribution of the degradation of the machines at maintenance time

The degradation distributions confirm the observations made in the previous table. When the inspection interval increases, the degradation levels at the maintenance decision time become more dispersed. As a consequence, the probability that some machines exceed the failure threshold increases.

In contrast, the proposed and the mixed approach maintains most machines below the failure threshold due to its conservative maintenance policy, which explains the lower corrective maintenance rates.

The maintenance threshold s_m directly controls the width of the preventive maintenance window. To analyze its impact, simulations were conducted with $s_m = 90$ and $s_m = 85$, while keeping the inspection interval fixed at $\tau = 15$ and considering three machines. And the percentile used for the a–RUL is 0.9, so it's the 0.9–RUL.

Table 3. Impact of the maintenance threshold ($\tau = 15$, 3 machines, 1000 simulations, 0.9–RUL)

s_m	CM	PM	OM
	(Zuo et al., 2023)		
90	0.064	0.765	0.171
85	0.002	0.904	0.094
	Mixed approach		
90	0.005	0.590	0.405
85	0.005	0.97	0.025
	Proposed approach		
90	0.007	0.623	0.370
85	0.006	0.957	0.037

The results show that decreasing the threshold from $s_m = 90$ to $s_m = 85$ reduces the proportion of over-maintenance while increasing the proportion of preventive maintenance. In practice, several machines that were previously classified as over-maintained when $s_m = 90$ fall into the preventive maintenance category when $s_m = 85$.

However, strategies that trigger maintenance as soon as at least one machine exceeds s_m tend to initiate maintenance actions too early, which increases the risk of unnecessary maintenance operations. That's why the value of PM are less than the 2 other methods.

Finally, the influence of the RUL threshold s_{RUL} was investigated for the proposed strategy. This parameter determines when load reallocations are triggered during the simulations. Increasing the RUL threshold reduces the number of reallocations performed during the simulations.

Table 4. Impact of the RUL threshold ($\tau = 15$, $s_m = 90$, 3 machines, 1000 simulations, 0.9–RUL))

s_{RUL}	CM	PM	OM
Low threshold	0.005	0.965	0.030
Medium threshold	0.006	0.950	0.044
High threshold	0.008	0.935	0.057

The results indicate that even with a limited number of reallocations, the proposed approach still achieves effective maintenance optimization. This is particularly relevant from an industrial perspective, since load reallocations or production adjustments may introduce additional operational costs. The proposed strategy therefore enables a compromise between maintenance performance and operational stability.

To further analyze this trade-off, an additional indicator based on load variability could be introduced in future work in order to quantify the stability of the production system under the proposed maintenance policy.

6. CONCLUSION AND PERSPECTIVES

This paper proposed an SQP-based maintenance optimization strategy for systems composed of multiple degrading machines operating in parallel, aiming to maximize the probability of simultaneous maintenance while limiting unnecessary interventions. The approach was evaluated through Monte Carlo simulations and compared with the strategy introduced by (Zuo et al., 2023) as well as with a mixed approach. The results show that the proposed method achieves performance comparable to the reference strategy, with a low proportion of corrective maintenance and a high proportion of preventive actions. The main difference lies in the load reallocation philosophy: while (Zuo et al., 2023) focuses on short-term convergence of degradation trajectories, the proposed method promotes long-term synchronization of maintenance events. As a result, it does not require systematic load adjustments at each observation step, which is particularly beneficial in industrial contexts where frequent adjustments may induce operational costs or disrupt production. Overall, the proposed strategy provides a practical compromise between maintenance performance and operational stability.

Future work could focus on improving the observation policy used in the maintenance framework. In the current formulation, inspections are performed at fixed time intervals defined by the parameter τ. A promising direction would be to replace this fixed time step by an adaptive observation policy. In such a framework, the next inspection time could be dynamically determined based on the predicted degradation evolution and the expected benefit of a potential load reallocation. This would allow the system to trigger observations only when they are likely to improve the maintenance decision, potentially reducing monitoring costs while preserving maintenance performance.

ACKNOWLEDGMENT

The author gratefully acknowledges the financial support provided by the European Commission within the European Regional Development Fund, the Agence nationale de la recherche (France 2030), the Swedish Agency for Economic and Regional Growth, the Region Pays de La Loire, Angers Loire Métropole, Region Gävleborg, as well as the Universities of Angers and Gävle. This research was made possible in the

frame of the EU GREEN Alliance.

REFERENCES

Askri, T., Zied, H., & Rezg, N. (2017, April). Jointly Production and Correlated Maintenance Optimization for Parallel Leased Machines. *Applied Sciences*, *7*, 461. doi: 10.3390/app7050461

Björsell, N., & Dadash, A. H. (2021, January). Finite Horizon Degradation Control of Complex Interconnected Systems∗. *IFAC-PapersOnLine*, *54*(1), 319–324. doi: 10.1016/j.ifacol.2021.08.036

Chen, W., & Hao, S. (2025, September). Condition-Based Operation and Maintenance Strategy for Load-Sharing Systems Based on Wiener Process. *IEEE Transactions on Reliability*, *74*(3), 4402–4416. doi: 10.1109/TR.2025.3545037

Domingo, R., & Aguado, S. (2015, July). Overall Environmental Equipment Effectiveness as a Metric of a Lean and Green Manufacturing System. *Sustainability*, *7*(7), 9031–9047. doi: 10.3390/su7079031

Esposito, N., Castanier, B., & Giorgio, M. (2022, January). Prescriptive block replacement policy for production degrading systems. *IFAC-PapersOnLine*, *55*(10), 1974–1979. doi: 10.1016/j.ifacol.2022.09.688

Hao, S., Chen, W., & Wang, J. (2025, October). Joint optimization of condition-based operation and maintenance strategy for load-sharing systems subject to hybrid continuous and discrete loads. *Reliability Engineering & System Safety*, *262*, 111166. doi: 10.1016/j.ress.2025.111166

Herr, N., Nicod, J.-M., & Varnier, C. (2014, August). Prognostics-based scheduling in a distributed platform: Model, complexity and resolution. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 1054–1059). (ISSN: 2161-8089) doi: 10.1109/CoASE.2014.6899455

Huynh, K. T., Barros, A., & Berenguer, C. (2015, March). Multi-Level Decision-Making for The Predictive Maintenance of -Out-of- :F Deteriorating Systems. *Reliability, IEEE Transactions on*, *64*, 94–117. doi: 10.1109/TR.2014.2337791

Pérez, E., Ambati, R. R., & Torres, A. J. R. (2018). Maximising the number of on-time jobs on parallel servers with sequence dependent deteriorating processing times and periodic maintenance. *International Journal of Operational Research*, *32*(3), 267. doi: 10.1504/IJOR.2018.092735

Sharifi, M., Ghaleb, M., & Taghipour, S. (2023, December). Joint parallel-machine scheduling and maintenance planning optimisation with deterioration, unexpected breakdowns, and condition-based maintenance. *International Journal of Systems Science: Operations & Logistics*, *10*(1), 2200888. (_eprint: https://doi.org/10.1080/23302674.2023.2200888) doi: 10.1080/23302674.2023.2200888

Tarek, A., Zied, H., & Nidhal, R. (2017, April). Improved maintenance strategy for a parallel leased machines system. In *2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO)* (pp. 1–5). doi: 10.1109/ICMSAO.2017.7934893

Tian, Z., & Liao, H. (2011, May). Condition based maintenance optimization for multi-component systems using proportional hazards model. *Reliability Engineering & System Safety*, *96*(5), 581–589. doi: 10.1016/j.ress.2010.12.023

Van Horenbeek, A., & Pintelon, L. (2013, December). A dynamic predictive maintenance policy for complex multi-component systems. *Reliability Engineering & System Safety*, *120*, 39–50. doi: 10.1016/j.ress.2013.02.029

Zuo, J., Cadet, C., Li, Z., Berenguer, C., & Outbib, R. (2023, February). Post-prognostics decision-making strategy for load allocation on a stochastically deteriorating multi-stack fuel cell system. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *237*(1), 40–57. doi: 10.1177/1748006X221086381

Zuo, J., Yousfi Steiner, N., Li, Z., Cadet, C., Bérenguer, C., & Hissel, D. (2025, February). Optimal Post-prognostics Decision-Making for Multistack Fuel Cells in Transportation: Toward Joint Load Allocation and Maintenance Scheduling. *IEEE Transactions on Transportation Electrification*, *11*(1), 2456–2466. doi: 10.1109/TTE.2024.3423404