Reinforcement learning-based flow management of gas turbine parts under stochastic failures

Michele Compare, Luca Bellani, Enrico Cobelli, Enrico Zio

The International Journal of Advanced Manufacturing Technology (2018) 99:2981–2992
https://doi.org/10.1007/s00170-018-2690-6

ORIGINAL ARTICLE

Reinforcement learning-based flow management of gas turbine parts under stochastic failures

Michele Compare[1,2] · Luca Bellani[2] · Enrico Cobelli[1] · Enrico Zio[1,2,3]

Received: 7 March 2018 / Accepted: 10 September 2018 / Published online: 18 September 2018

Abstract
For maintenance of gas turbines (GTs) in oil and gas applications, capital parts are removed and replaced by parts of the same type taken from the warehouse. When the removed parts are found not broken, they are repaired at the workshop and returned to the warehouse, ready to be used in future maintenance. The management of this flow is of great importance for the profitability of a GT plant. In this paper, we adopt a previously developed formalized framework of the part flow and reinforcement learning (RL) to optimize part flow management. The formal framework and RL algorithm are extended to account for the stochastic failure process of the involved parts. An application to a scaled-down case study derived from an industrial application is illustrated.

Keywords Part flow · Reinforcement learning · Gas turbines · Stochastic failures

1 Introduction

Gas turbines (GTs) are complex systems composed by several expensive capital parts (e.g., buckets, nozzles, and shrouds). Degradation of these parts (e.g., by fracture and fatigue [1–3], fouling [4–6], corrosion [7, 8], and oxidation [9]) can lead the GTs to failure and, thus, to costly forced outages (FOs) for performing corrective maintenance actions, in which the failed parts are scrapped and replaced by parts of the same type selected from those available at the warehouse.

To avoid FOs, GTs undergo periodic maintenance shutdowns (MSs), which restore the capital parts. At every MS, capital parts are removed from the GTs and repaired at the workshop, unless they are scrapped because they have reached their pre-fixed maximum number of working hours. The repaired parts are, then, put back at the warehouse, for use in future maintenance. The parts removed from the GTs are replaced by parts taken from the warehouse, either restored or newly purchased.

This brief description of GT maintenance brings out the complexity of its management, which relies on a specific expertise for performing the intricate procedures for GT disassembling and re-assembling, an efficient logistic organization for spares management (i.e., their ordering, shipping, etc.), a deep knowledge about the degradation processes affecting the parts for their effective repair, etc. (see [10] for an overview). Differently from the GT manufacturing companies, which are usually structured for addressing these issues, their customers are generally not fully qualified to do so. This situation has boosted the diffusion of maintenance service contracts between the GT manufacturers (i.e., the maintenance service providers) and the GT owners (i.e., the recipients of the service) [11–14].

Service contracts yield new business opportunities to GT manufacturers, who can sell the GT production rates, instead of selling the GTs, with consequent added values if they assume portions of the clients' business risks [11, 15]. For the service contract to be profitable, however, GT manufacturers need to develop effective and efficient maintenance strategies and spare part inventory management policies [13, 15–17].

To manage the maintenance events (i.e, MSs and FOs), decisions must be made on both the removed part (send it to the workshop for repair or scrap it?) and the part to be installed on the GT (new part or part taken from

✉ Michele Compare
michele.compare@aramis3d.com

[1] Energy Department, Politecnico di Milano, Milan, Italy

[2] Aramis S.r.l., Milan, Italy

[3] Chair on Systems Science and the Energetic Challenge, Foundation Electricité de France at CentraleSupélec, Paris, France

the warehouse?), which strongly impact on the profitability of the GT maintenance service contract. For example, the decision to repair the removed parts entails, on the one hand, the possibility of re-using the part with consequent reduction in the number of parts to purchase. On the other hand, the repair actions entail both direct workshop costs and indirect costs related to the increased risk of FOs, with consequent penalties to the maintenance service provider for business interruption: repaired parts have a failure probability larger than that of new parts, as the risk of failure generally increases with part age. Furthermore, unnecessary repair actions at the end of the maintenance service contract may lead to the warehouse containing parts ready for installation, whose value is lost by the service provider. On the contrary, scrapping old parts reduces the risk of failure and workshop costs, but increases the number of purchase actions taken by the maintenance service provider.

The parts installed on the GTs are no longer available at the warehouse for replacement at the next MS and when they return to the warehouse, they do so with a reduced number of remaining working cycles. Thus, the decisions at every MS influence the decisions at the next MSs: in this sense, the part flow management (PFM) can be framed as a sequential decision problem (SDP) [18], seeking for the sequence of future maintenance decisions (i.e., the optimal policy) which entails the smallest expected maintenance costs over the duration of maintenance service contract. This requires the decision maker (DM) to consider variables such as the remaining time up to the end of the service contract, the availability of spares, the costs related to the repair actions, etc.

Despite the relevance of PFM for the profitability of the maintenance service contracts, to the authors' best knowledge, systemic approaches to address it are still lacking. Although the literature on maintenance service is very vast [12, 17], it covers issues different from that of optimizing the part flow. For example, methods for setting the optimal price of service contracts are proposed in [11, 12, 14], within the game theory framework. The same issue, i.e., contract pricing optimization, is investigated in [19] in combination with the optimization of logistics (i.e., facility locations, capacities and inventories with given service level), and in combination with the issue of optimally scheduling preventive maintenance in [16, 20]. Other optimization objectives are the minimization of the warehouse costs through the reduction of the average number of parts sojourning therein (e.g., [16]), the identification of the optimal times for performing maintenance actions and ordering parts (e.g., [21, 22]), the level of repair [13, 23], the number of maintenance jobs that can be completed in each maintenance period [24], etc.

The focus of this paper is on the search of the best PFM strategy that minimizes the service contract costs for the GT manufacturer over a finite time horizon. Currently, the management of the part flow is dealt with experience-based rules, such as the most residual cycles (MRC) one: the removed parts are always repaired and the part with the largest residual life among those available at the warehouse are installed on the GT; a new part is purchased only when the warehouse is empty. Although MRC ensures at the smallest failure probability, nonetheless, we have shown in [25] that MRC does not necessarily yield optimal policies on a finite time horizon in which the sequence of MSs is a priori known.

In this work, we extend the modeling and optimization framework developed in [25] to account for part failure stochastic processes and FOs, which change the pre-scheduled sequence of MSs. In particular, we formalize the PFM problem as a SDP in a stochastic environment and propose the use of Reinforcement Learning (RL, [18, 26, 27]) for its solution. RL is a machine learning technique suitable for addressing SDPs in stochastic environments [26] and widely applied to decision-making problems in diverse industrial sectors, such as the electricity market [28, 29], military trucks [30], process industry [31], supply chain and inventory management [21, 32–34], and operations in port container terminal [35, 36], to cite a few.

The problem formulation and solution framework proposed in this paper is applied to the same case study as that of [25], although here, we take into account the failure of the parts and the FOs. In the case study, it turns out that also when considering the part failures, the solution given by the MRC rule is not optimal, being outperformed by the policy found by RL. Moreover, we compare the optimal policy provided by our RL algorithm in case there are no FOs with that presented in [25].

The original contributions of this paper are:

- The further development of a new problem (i.e., optimization of PFM), which has never appeared in the literature. Given its relevance for maintenance service contract management and profitability, it is expected to give rise to a dedicated line of research.
- The formalization of the PFM problem as a SDP, which allows taking into account the dependency between consecutive decisions and the uncertainty in the part failure.
- The proposal of a RL algorithm to find the optimal PFM policy. The algorithm can be applied to medium-small, real applications and improve the current experience-based practice.

The structure of the paper is as follows. In Section 2, we introduce the extended mathematical formulation of the considered SDP. In Section 3, details about the extended RL algorithm used for optimizing the part flow management are

provided. In Section 4, the case study is discussed. Finally, conclusions are drawn in Section 5.

2 Problem setting

Consider an oil and gas plant in which a number G of GTs are operated (Fig. 1). A scheduled preventive maintenance policy is defined, whereby every GT is maintained every H hours. We assume the maintenance staggering so that two MSs are never performed simultaneously and that the sequence of G MSs is shorter than H hours.

The GTs are operated for T hours each, T being a multiple of H. For generality, the time horizon is made dimensionless through division by H and it is discretized into time channels of length Δt. These are short enough that the probability of having multiple failures in the same channel is negligible.

To formalize the part flow management in a stochastic environment, the model proposed in [25] must be modified to allow decisions to be taken upon FOs, which occur at time instants different from those initially scheduled for the MSs. In fact, any failure event requires a re-scheduling of the maintenance activities, which entails a variability in both the number of events over the GT plant operation horizon and their timing and sequence. To consider this uncertain dynamic aspect of the SDP, the GT plant operation time horizon is partitioned into time channels, which are identified by the instants $t \in \Theta = \{0, \Delta t, 2\Delta t, ..., T/H+1\}$ (Fig. 1 in dashed line), where $t = 0$ corresponds to the first MS, T/H is the time instant of the last MS of the GT maintained at $t = 0$, according to the initial schedule; the last time instant, $t = T/H + 1$, dimensionless, is the upper bound of the time instant of the last scheduled MS, as the maintenance cycle determines the maximum distance between the G MSs. For brevity, we indicate the $\theta - th$ time instant of Θ, in ascending order, by t_θ, $\theta = 1, ..., |\Theta|$, where $|\square|$ indicates the cardinality of its argument set $\square$.

Given the discretization of the time horizon, we assume that if a failure occurs at time $t_\theta + \tau, \tau \in [0, 1]$ on the GT that has been maintained at the $\theta - th$ time instant, t_θ, then the FO is performed at the $\theta^* - th$ time instant $\theta^* = \arg\min_{\eta \in \Theta} \text{abs}[t_\theta + \tau - t_\eta]$ (e.g., Fig. 1, the FO is performed at the time instant $t_{\theta_{k+2}}$).

In case any GT experiences a FO at a time τ after its last MS, $\tau \in [0, 1]$, then all its future MSs are shifted by τ, as maintenance is always intended to allow the GT working continuously for H hours. For example, Fig. 1, bottom, shows the original sequence of MSs of GT G, which is shifted forward by the FO event originating a different MS sequence. Notice that we assume that the end of the plant operational time horizon does not change even when the actual MS and FO sequence change due to random failures.

In regard to the time to repair the parts removed from the GT, we assume that this is negligible with respect to Δt, whereby the parts repaired are immediately available at the next event. Every part is assigned a maximum number of remaining cycles ($MNRC$), indicated by r, which ranges between $r = 0$, in case of parts that must be scrapped and $r = R$, for new parts. The $MNRC$ is reduced by 1 upon the installation of the part on a GT: if the GT is stopped, the part will no longer be able to re-perform the entire started cycle.

The failure times of the parts obey the exponential distribution with failure rate, λ_r, depending on the $MNRC$ value $r \in \{1, ..., R\}$. To have dimensionless time channels, the values of the failure rates are scaled on the duration of the H hours cycle.

The cumulative distribution function (CDF) of the failure time reads:

$$F_r(\tau) = 1 - e^{-\lambda_r \tau} \tag{1}$$

where τ is the time since the installation of the part on the GT. Notice that the choice of describing the part failure behavior by the exponential distribution with failure rate depending on the $MNRC$ value allows modeling the part degradation mechanism as a Markov process. The resulting step-wise, monotonously increasing behavior of the failure rate can be thought of as a rough approximation of a continuously increasing hazard rate [37].

At any shutdown, the DM has to take the following decisions:

- If the maintenance event is a MS, decide whether to repair or scrap the part removed from the maintained GT. $C^{rep}(r)$ is the cost of repairing a part with $r \in \{1, ..., R\}$ remaining cycles, whereas C^{scrap} is the cost of scrapping a part.
- If the maintenance event is a FO, then the part must be scrapped, and a penalty $C^{failure}$ must be paid, which also encodes the extra costs related to the management of an unplanned event.
- To replace the removed part, decide whether to buy a new part or select one from those available at the warehouse, if any. C^{pur} is the cost of purchasing a new part, whereas the cost of selecting a part from the warehouse is zero, as the repair costs have already been accounted for.

To simplify the notation, we define two indicator functions:

$$\mathbf{1}|_\theta^{FO} = \begin{cases} 1 \text{ if a FO occurs at time } t_\theta, \theta = 1, ..., |\Theta| \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

$$\mathbf{1}|_\theta^{MS} = \begin{cases} 1 \text{ if a MS occurs at time } t_\theta, \theta = 1, ..., |\Theta| \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

To keep track of the shutdown temporal sequence, we introduce set $\boldsymbol{\theta} = \{\theta_1, ..., \theta_K\}$ encompassing the indexes of

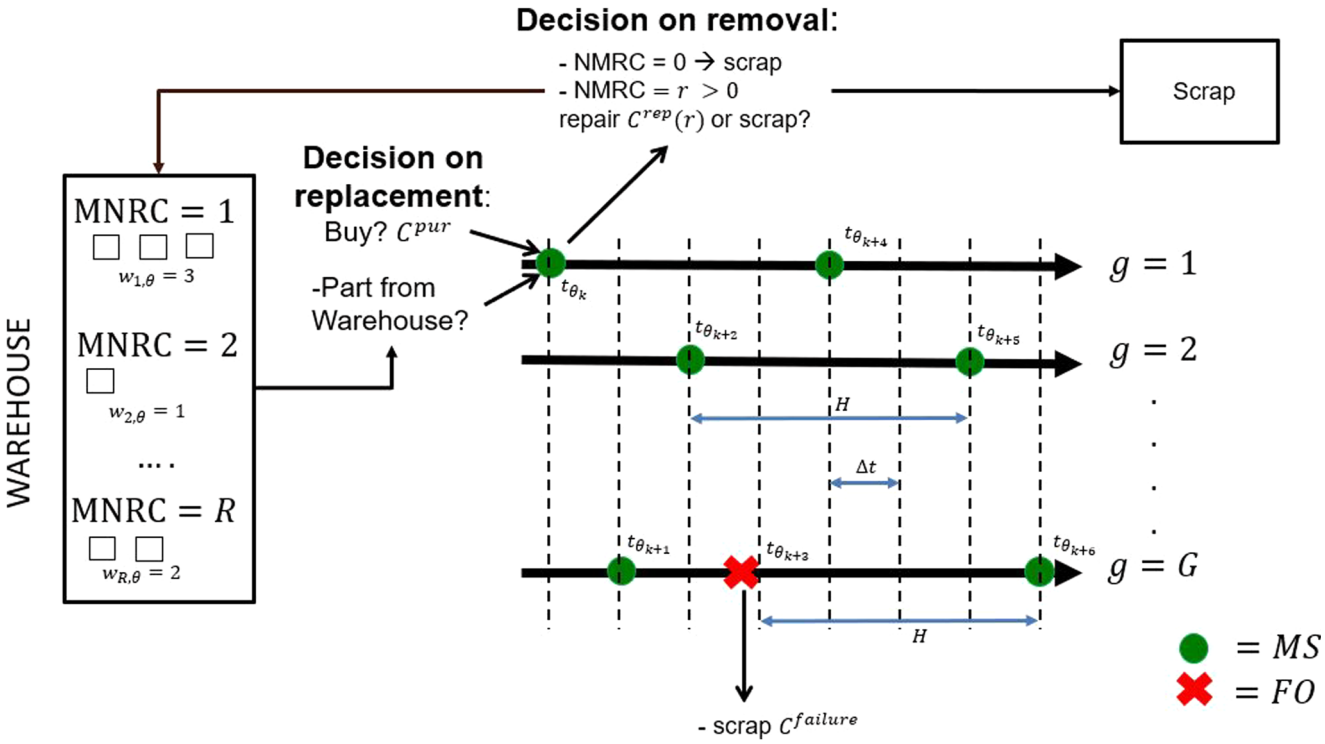

Fig. 1 Summary of the model setting

the time instants $t_\theta \in \Theta$ at which a shutdown event occurs, where K is a random variable indicating the last shutdown within the operational time horizon: if there are no failures, then $K = T/H \cdot G$. Notice that the time index $\theta \in \boldsymbol{\theta}$ if $\mathbf{1}|_\theta^{FO} + \mathbf{1}|_\theta^{MS} = 1, \theta = 1, ..., |\Theta|$.

We also introduce the integer variables $d_{g,\theta}$ and $w_{r,\theta}$ to indicate the $MNRC$ of the capital part on the g-th GT at the $\theta - th$ time instant and the number of parts with $MNRC$ equal to r available at the warehouse at the $\theta - th$ time instant, respectively, $\theta = 1, ..., |\Theta|$, $g \in \{1, ..., G\}$, $r \in \{1, ..., R\}$, $w \in \{0, ..., W\}$, where W is the maximum number of parts that can be stored in the warehouse for each $MNRC$ value (Fig. 1). The index of the GT maintained at the k-th shutdown, i.e., at time t_{θ_k} is traced by $\delta_{\theta_k} \in \{1, ..., G\}$.

Finally, the Boolean variable $a_{\theta_k,\rho} \in \{0, 1\}$ indicates the action $\rho \in \{0, ..., 2R + 1\}$ taken at the $k - th$ maintenance event at time $t_{\theta_k} \in \boldsymbol{\Theta}$, $\theta_k \in \boldsymbol{\theta}$:

- $a_{\theta_k,0} = 1$ when a new part is purchased and installed and the removed part is scrapped.
- $a_{\theta_k,\rho} = 1$, $\rho \in \{1, ..., R\}$, when a part with $MNRC = \rho$ is installed and the removed part is scrapped.
- $a_{\theta_k,R+1} = 1$ when a new part is purchased and installed and the removed part is repaired.
- $a_{\theta_k,\rho} = 1$, $\rho \in \{R + 2, ..., 2R + 1\}$, when a part with $MNRC = \rho - R - 1$ is installed and the removed part is repaired.

The boolean variable $a_{\theta_k,\rho}$ is such that only one action can be taken at the k-th shutdown:

$$\sum_{\rho=0}^{2R+1} a_{\theta_k,\rho} = 1 \tag{4}$$

From the above, the cost incurred at the k-th shutdown is:

$$C_k = (a_{\theta_k,0} + a_{\theta_k,R+1}) \cdot C^{pur} + \sum_{\rho=0}^{R} a_{\theta_k,\rho} \cdot C^{Scrap} + \sum_{\rho=R+1}^{2R+1} a_{\theta_k,\rho} \cdot C^{Rep}(d_{g,\theta_k}) + C^{failure} \cdot \mathbf{1}|_{\theta_k}^{FO} \tag{5}$$

The objective is to minimize the value of the expected maintenance expenditures incurred in the whole time horizon, which is given by the sum of the costs of all shutdowns within the time horizon:

$$V = \mathbb{E}\left[\sum_{k=1}^{K} C_k\right] \tag{6}$$

Notice that the total maintenance expenditures within the operational time horizon is a random variable depending on both the number of failures in the time horizon and their occurrence time. For simplicity, this sum is considered undiscounted.

Notice also that in real industrial applications, the failures of the capital parts mounted on the same GT are dependent on each other, as failures can originate cascading effects.

Nonetheless, we track a single capital part, only. The object of future research work will be the extension of the developed framework to applications in which the flows of different capital parts are considered as a whole for a global optimization.

3 Algorithm

In this section, we provide some insights about the RL algorithm developed to address the part flow management issue. To develop the RL algorithm, we need to define the environment state, the actions available at each state and the corresponding rewards [18].

The state at the shutdown occurring at time t_{θ_k} is defined by vector $\mathbf{S}_{\theta_k} \in \mathbb{N}^{R+G+2}$, whose j-th element is:

$$\mathbf{S}_{\theta_k,j} = \begin{cases} w_{j,\theta_k} & \text{if} \quad j \in \{1, ..., R\} \\ d_{j-R,\theta_k} & \text{if} \quad j \in \{R+1, ..., R+G\} \\ \delta_{\theta_k} & \text{if} \quad j = R+G+1 \\ \theta_k & \text{if} \quad j = R+G+2 \end{cases} \tag{7}$$

In words, the first R entries of the state vector define the number of parts with the different $MNRC$ values available at the warehouse; the G entries from $R+1$ to $R+G$ indicate the $MNRC$ of the parts installed on the GTs at their corresponding last MS; the $(R+G+1)$-th entry points to the GT maintained at time instant t_{θ_k}; the last entry encodes the time of the shutdown. This definition of the environment state entails that its size is equal to $(W+1)^R \cdot R^G \cdot G \cdot (T/H+1) \cdot (H/\Delta t)$. For a medium scale problem in the oil and gas industry, with $G=6$, $R=6$, $W=6$, $H=10 \cdot \Delta t$ and $T=8 \cdot H$, this corresponds to $3 \cdot 10^{12}$ states.

Notice that the state definition in Eq. 7 differs from that given in [25], which encodes only two variables: the number of parts with the different $MNRC$ values available at the warehouse and the index of the MS (in place of the index of the time channels). As shown in [25] for the deterministic environment (DE) case, i.e., without stochastic failures, the other variables entering (7) contain redundant information and, thus, can be neglected: this strongly reduces the dimension of the state space and the computational burden.

Notice also that the definition of the environment state in Eq. 7 does not fully satisfy the Markov property [18], as the state vector does not include the time up to the next MS. This time interval determines the probability of moving from one state to another, as parts have higher chances of failing when operated for longer time periods. Then, omitting the information about the remaining time up to the next scheduled event undermines the knowledge about the probabilistic behavior of the future evolution of the state. However, the state definition completely satisfying the Markov property turns out into a very large state space, thus requiring a much larger computational effort. Then, our state definition seems the best compromise between an accurate description of the environment and a computationally manageable number of states.

The action taken at the shutdown occurring at time t_{θ_k} is indicated as:

$$A_{\theta_k} = \sum_{\rho=0}^{2R+1} (a_{\theta_k,\rho} \cdot \rho) \tag{8}$$

The base reward of the shutdown at time t_{θ_k} is the opposite of the maintenance cost $-C_k$, as RL is usually addressed as a maximization task (i.e., minimizing cost is equal to maximizing its opposite). In the RL framework, each state-action pair is described by $Q_\pi(\mathbf{S}_{\theta_k}, A_{\theta_k})$, which measures the expected return starting from state $\mathbf{S}_{\theta_k}$, taking action A_{θ_k} and thereafter following the policy π [18]:

$$Q_\pi(\mathbf{S}_{\theta_k}, A_{\theta_k}) = \mathbb{E}_\pi \left[\sum_{k=k^*}^{K} (-C_k) | \mathbf{S}_{\theta_k}, A_{\theta_k} \right] \tag{9}$$

where $\pi = \pi(\epsilon)$ is the ϵ-greedy policy [18], which selects with probability ϵ an action uniformly among the available ones; with probability $1-\epsilon$, the action with largest expected return is selected on each state, i.e. $A_{\theta_k} = \arg\max_{A \in \{A_0, ... A_{2 \cdot R+1}\}} Q_\pi(\mathbf{S}_{\theta_k}, A)$.

Note that $\epsilon = \epsilon_n$ decreases at each episode, as the first episodes require a large exploration rate to rapidly move from the initial values assigned to the state-action function. As the simulation proceeds, more state-action pairs are visited, whereby the values of $Q_\pi(\mathbf{S}_{\theta_k}, A_{\theta_k})$ become more accurate. This allows selecting the optimal action in every state, as the epsilon-greedy exploration policy converges to the optimal (greedy) policy (for further details, see exploration-exploitation dilemma, e.g., [18]). To properly set the value of the exploration rate and its evolution over time, we have applied a trial-and-error procedure.

In this work, we use the SARSA(λ) algorithm (e.g., [18, 25, 38]) to find the best approximation of the values of $Q_\pi(\mathbf{S}_{\theta_k}, A_{\theta_k})$, which relies on the following updating formula:

$$\begin{aligned} Q(\mathbf{S}_{\theta_z}, A_{\theta_z}) \longleftarrow\ & Q(\mathbf{S}_{\theta_z}, A_{\theta_z}) + (\lambda)^{(k-z)} \alpha_n \\ & \cdot [-C_k + Q(\mathbf{S}_{\theta_{k+1}}, A_{\theta_{k+1}}) - Q(\mathbf{S}_{\theta_k}, A_{\theta_k})] \end{aligned} \tag{10}$$

where $z \in \{1, ..., k\}$ is the MS counter, k is the actual MS, $\lambda \in [0, 1]$ is the eligibility trace and $\alpha_n \in [0, 1]$ is the learning rate at the n-th episode. According to [27], we have applied a trial-and-error procedure to set the value of $\lambda = 0.8$.

Notice that the eligibility trace λ is different from the failure rate, λ_r (i.e., with subscript), although we indicate them with the same letter. This is due to the large use of this letter in the respective fields.

Table 1 Initial scenario and model parameters

G	H	T	Δt	W	R	C^{Scrap}	C^{rep} $(r=1)$	C^{rep} $(r=2)$	C^{pur}	$C^{failure}$	$\lambda_{r=1}$	$\lambda_{r=2}$	$\lambda_{r=3}$
2	24 000	216 000	0.1	3	3	0	50	50	100	200	0.06	0.03	0.01

The choice of using SARSA(λ) among the available RL algorithms (e.g., [27]) is justified by the fact that within the family of value-based RL algorithms, SARSA(λ) has been shown to be a very effective on-policy method [27]. This makes it simpler to extend it to the eligibility trace paradigm, which guarantees fast and robust convergence, especially in case of finite time horizon SDPs [18, 38]. On the contrary, off-policy RL algorithms such as Q(λ) need to be finely set to avoid biased estimations of the state-action values. Further research work will focus on the comparison of SARSA(λ) with policy-based and actor-critic RL algorithms [27].

Other optimization algorithms such as dynamic programming algorithms [18] could also be used for the specific setting considered in this work. However, in this respect the choice of RL has a twofold justification. On one side, RL algorithms allow encoding the aleatory uncertainty in the failure times of the GT parts more easily than the other algorithms. On the other side, although here not considered, the complexity of the real industrial applications requires that the SDP encode many additional GT operational aspects, such as the possibility of inspecting the parts without performing maintenance (i.e., condition-based maintenance), the different duration of the maintenance intervals for parts of different technologies, the constraints on the sharability of the parts on GTs with different operation temperatures, and the long repair durations that make the parts not readily available for the next MS. Accounting for these GT operational aspects requires encoding constraints about the actions that can be taken in each state, which are really difficult to set in model-based frameworks such as dynamic programming. On the contrary, RL acts on the simulation of the decision process and, thus, selects actions from those feasible, only. This makes RL easily integrable with part flow simulators.

On the other hand, the proposed RL solution suffers from some limitations that can still prevent its full application to the industrial practice. First, in complex problems the state-space becomes very large, whereby the tabular representation of the state-action value function is not doable. For this, action-value approximation techniques can be used instead of the tabular approach hereby presented. This issue will be tackled in future works. Yet, although the time required to run a single part flow simulation episode is very small (in the order of milliseconds), nonetheless, RL requires performing a very large number of simulated episodes. This can undermine the application of RL to contexts in which decisions must be taken

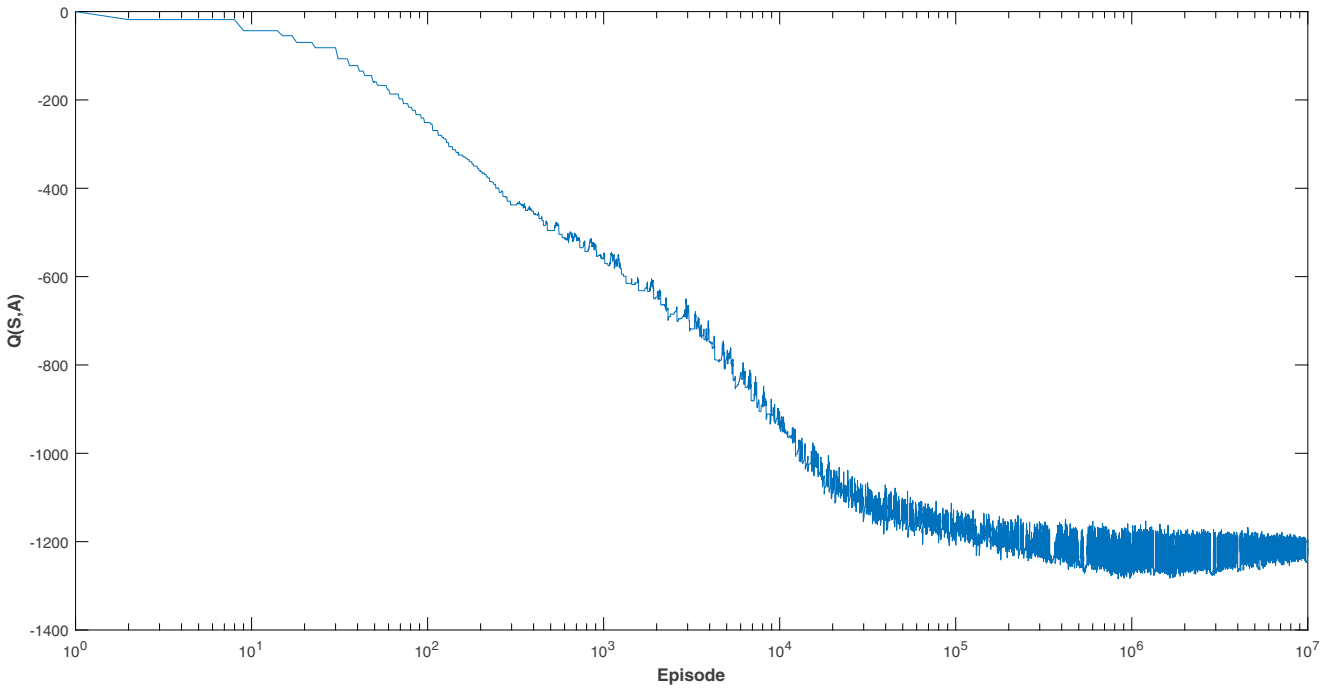

Fig. 2 $Q(\mathbf{S}_1, A_1)$ over the 10^7 simulated episodes

Table 2 Comparison of MRC and RL policies

Number of FOs	MRC	RL	Average maintenance costs for MRC	Average maintenance costs for RL
0	0.519 ± 0.00044	0.5324 ± 0.00049	1150	1050
1	0.3476 ± 0.00047	0.3446 ± 0.00047	1361	1291
2	0.1085 ± 0.00031	0.1021 ± 0.00030	1586	1549
3	0.0215 ± 0.00014	0.0183 ± 0.00013	1804	1803
4 or more	0.0034 ± 0.00005	0.0025 ± 0.00005	2078	2079
–	–	Average total maintenance costs	1288	1200

readily. In any case, the proposed RL framework is not meant to be used in a real-time setting. Rather, it is conceived to be run either at the first decision time or when unforeseen events such as a change in the warehouse configuration due to external reasons, modify the environment and pose a new optimization problem. In these cases, however, we usually have plenty of time to take decisions.

4 Case study

In this section, we extend the case study proposed in [25], which derives from an industrial application, to include consideration of the parts failure stochastic process. The main characteristics are summarized in Table 1.

In the considered oil and gas plant there are $G = 2$ GTs (first column in Table 1), each one maintained every $H = 24\,000$ hours (second column) over a time horizon of $T = 216\,000$ hours (third column). The time step is set to $\Delta t = 0.1$, dimensionless (fourth column). The maximum part $MNRC$, R, and the maximum number of available parts in the warehouse for each $MNRC$ value, W, are both set to 3 (fifth and sixth columns in Table 1, respectively). The cost values are reported from the 7-th to the 11-th columns of Table 1, in arbitrary units. Finally, the failure rates $\lambda_r, r = 1, 2, 3$, are reported in the last three columns, dimensionless. These values are for illustration, only.

Notice that the failures we are referring to do not entail the complete loss of the entire GT. Rather, we consider as failure the degradation of the functional performance to a level which requires the GT control system to command the

Table 3 MRC policy, scenario with no FO

k	t_{θ_k}	θ_k	w_{1,θ_k}	w_{2,θ_k}	w_{3,θ_k}	$MNRC$@ GT $g = 1$	$MNRC$@ GT $g = 2$	$MNRC$ Installed	Repair	Purchase	C_k
1	0	1	3	1	0	**2**	0	2	Y	N	50
2	0.5	6	3	1	0	1	**0**	2	N	N	0
3	1	11	3	0	0	**1**	1	1	Y	N	50
4	1.5	16	3	0	0	0	**1**	1	Y	N	50
5	2	21	3	0	0	**0**	0	1	N	N	0
6	2.5	26	2	0	0	0	**0**	1	N	N	0
7	3	31	1	0	0	**0**	0	1	N	N	0
8	3.5	36	0	0	0	0	**0**	3	N	Y	100
9	4	41	0	0	0	**0**	2	3	N	Y	100
10	4.5	46	0	0	0	2	**2**	3	Y	Y	150
11	5	51	0	1	0	**2**	2	2	Y	N	50
12	5.5	56	0	1	0	1	**2**	2	Y	N	50
13	6	61	0	1	0	**1**	1	2	Y	N	50
14	6.5	66	1	0	0	1	**1**	1	Y	N	50
15	7	71	1	0	0	**1**	0	1	Y	N	50
16	7.5	76	1	0	0	0	**0**	1	N	N	0
17	8	81	0	0	0	**0**	2	3	N	Y	100
18	8.5	86	0	0	0	2	**2**	3	N	Y	100
19	9	91	0	0	0	**2**	2	3	Y	Y	150
20	9.5	96	0	1	0	2	**2**	2	Y	N	50
–	–	–	0	1	0	**2**	1	–	–	TOT	1150

Table 4 RL policy without FO

k	t_{θ_k}	θ_k	w_{1,θ_k}	w_{2,θ_k}	w_{3,θ_k}	$MNRC$@ GT $g=1$	$MNRC$@ GT $g=2$	$MNRC$ Installed	Repair	Purchase	C_k
1	0	1	3	1	0	**2**	0	1	Y	N	50
2	0.5	6	2	2	0	0	**0**	3	N	Y	100
3	1	11	2	2	0	**0**	2	1	N	N	0
4	1.5	16	1	2	0	0	**2**	3	Y	Y	150
5	2	21	1	3	0	**0**	2	2	N	N	0
6	2.5	26	1	2	0	1	**2**	2	Y	N	50
7	3	31	1	2	0	**1**	1	2	Y	N	50
8	3.5	36	2	1	0	1	**1**	1	Y	N	50
9	4	41	2	1	0	**1**	0	3	Y	Y	150
10	4.5	46	3	1	0	2	**0**	3	N	Y	100
11	5	51	3	1	0	**2**	2	3	Y	Y	150
12	5.5	56	3	2	0	2	**2**	2	Y	N	50
13	6	61	3	2	0	**2**	1	2	Y	N	50
14	6.5	66	2	3	0	1	**1**	1	Y	N	50
15	7	71	2	3	0	**1**	0	1	N	N	0
16	7.5	76	1	3	0	0	**0**	1	N	N	0
17	8	81	0	3	0	**0**	0	2	N	N	0
18	8.5	86	0	2	0	1	**0**	2	N	N	0
19	9	91	0	1	0	**1**	1	2	Y	N	50
20	9.5	96	1	0	0	1	**1**	1	N	N	0
–		–	0	0	0	**0**	0	0 –	–	TOT	1050

stop of the GT for removing the degraded part. The major costs associated to this event are those related to business interruption and to the loss of the part, which is scrapped.

The total number of possible states is $(W+1)^R \cdot R^G \cdot G \cdot (T/H+1) \cdot (10) = 126\,720$ and the total number of state-action pairs is $(W+1)^R \cdot R^G \cdot G \cdot (T/H+1) \cdot (10) \cdot (2R+2) = 1\,013\,760$.

The SARSA(λ) algorithm has been run for 10^7 episodes, which took 25 200 seconds on a 2.20GHz CPU, 4GB Ram computer. The convergence path is reported in Fig. 2, which shows the values of $Q(\mathbf{S}_1, A_1)$, where A_1 is the optimal action at $t_{\theta_k} = 0$ (i.e., in this case $A_1 = 5$). To verify that SARSA algorithm converged to the optimal solution, we considered the oscillating behavior at the end of the episodes and checked that this is coherent with the stochastic nature of the considered SDP, which entails that $Q(\mathbf{S}_\theta, A_\theta)$ oscillates around its average value, for any $\theta = 1, ..., |\mathbf{\Theta}|$ [27].

To fairly compare the optimal policy found by RL with that provided by MRC, these are tested for 10^6 Monte Carlo (MC) episodes, in which the GT parts fail according to the exponential distributions introduced above. Table 2 summarizes the results of these simulations. In particular, the last row reports the average total maintenance expenditures, independently on the number of failures leading to FOs. From these values, we can see that RL outperforms MRC for managing the part flow.

To understand this result, in the next sub-sections, we investigate the MC simulation outcomes summarized in Table 2: the first column shows the possible number of FOs occurring over the time horizon; for every number of FO, the second and third columns report the corresponding

Table 5 Comparison between MRC and RL policies in case of no FO, and RL in the deterministic environment

	Number of purchasing	Repairs of parts with r=2	Repairs of parts with r=1	Scrap of parts with r>0	Scrap of parts with r=0
RL	5	6	5	2	7
MRC	6	6	5	0	9
RL Det	5	6	5	1	8

Table 6 Number of episodes vs V, single FO scenario

Total maintenance expenditures V	% MRC	% RL
1250	0	41.32
1300	2.17	36.11
1350	74.17	20.41
1400	21.67	2.16
1450	1.99	0

average portion of MC episodes for MRC and RL policies, respectively, with related 68% confidence bounds, whereas the two last columns report the mean total maintenance expenditures for MRC and RL, respectively.

4.1 Scenario with no FO

The second row of Table 2 shows that 53.24% of the simulated episodes do not experience FOs if we apply the RL policy, against 51.90% obtained by MRC, with no overlap of the confidence intervals of these estimates. This leads us to conclude that for a significant portion of the possible stochastic evolutions of the part flow, the RL policy yields a large number of episodes without FOs and, thus, small costs.

To investigate this result, we can refer to Table 3, which shows the part flow policy derived by the application of the MRC rule. Namely, the first three columns report the index of the shutdown, k, the MS time instant, t_{θ_k} and the corresponding time index θ_k. The following three columns report the corresponding situation of the warehouse. For example, at the beginning of the considered time horizon, i.e., at t_1=0, there are three parts with one remaining cycle, $w_{1,1} = 3$, one part with two remaining cycles, $w_{2,1} = 1$, and no new parts, $w_{3,1} = 0$.

Table 7 Number of episodes vs V, double FOs case-scenario

V	% MRC	% RL
1450	0	10.21
1500	2.33	18.55
1550	36.93	41.55
1600	47.22	20.12
1650	12.45	7.71
1700	1.03	1.40
1750	0.05	0.35
1800	0.00	0.06
1850	0.00	0.03
1900	0.00	0.00
1950	0.00	0.01
2000	0.00	0.02

Table 8 Repair costs, DE setting

C^{rep} $(r = 2)$	C^{rep} $(r = 1)$
56	62

The $MNRC$ values of the parts installed on GTs $g = 1$ and $g = 2$ are reported in the seventh and eighth columns, respectively, where the maintained GT is indicated in bold. For example, the part on the GT undergoing maintenance at $\theta_1 = 1$, i.e., $g = 1$, has $d_{1,1} = 2$ remaining cycles, whereas the GT $g = 2$ has been equipped with a part with one remaining cycle at the last MS.

The next three columns detail the action taken at the k-th shutdown. For example, at the first MS, the $MNRC$ of the part installed on GT $g = 1$ is $r = 2$ (ninth column) and the removed part is repaired (tenth column), with no purchase of new parts (eleventh column), i.e., $A_1 = 6$. Finally, the last column reports the maintenance cost, C_k, at the k-th shutdown.

To further detail the updating dynamics of Table 3, we can see that at the second MS $w_{2,6} = 1$, because the part removed from GT $g = 1$ is now available at the warehouse for installation on GT $g = 2$. The removed part must be scrapped, as it has no remaining cycles, $d_{2,6} = 0$. This gives a maintenance cost $C_2 = 0$.

The part flow solution given by the application of the MRC rule yields a total maintenance cost of 1150 (in arbitrary units), as reported in the last row of Table 3.

The application of the RL policy yields the part flow summarized in Table 4, which follows the same scheme as Table 3, whereas Table 5 summarizes the main differences between the two policies. From this Table, we can see that RL is able to find a more efficient part flow policy in the case of no FO, because it scraps two parts with $MNRC > 0$ (second row, fifth column), with one less purchase.

4.2 Scenario with single FO

The percentage of episodes with one FO is almost the same for RL and MRC (i.e., 34.46% and 34.76%, respectively). However, if we look at the average total maintenance expenditures (third row in Table 2), these are significantly different: 1291 for RL and 1361 for MRC, both in arbitrary units. This result can be explained by looking at Table 6, where the first column reports the values of all the possible maintenance expenditures that are encountered in case there is one FO, whereas the second and third columns show the corresponding percentage of time in which these are encountered in case of application of RL and MRC policies, respectively. We can see that around 77% of the episodes corresponding to the RL policy end with maintenance expenditures smaller or equal to 1300, whereas

Table 9 RL policy, DE setting [25]

k	t_{θ_k}	θ_k	w_{1,θ_k}	w_{2,θ_k}	w_{3,θ_k}	$MNRC$@ GT $g=1$	$MNRC$@ GT $g=2$	$MNRC$ Installed	Repair	Purchase	C_k
1	0	1	3	1	0	**2**	0	1	Y	N	56
2	0.5	6	2	2	0	0	**0**	3	N	Y	100
3	1	11	2	2	0	**0**	2	2	N	N	0
4	1.5	16	2	1	0	1	**2**	3	Y	Y	156
5	2	21	2	2	0	**1**	2	3	Y	Y	162
6	2.5	26	3	2	0	2	**2**	2	Y	N	56
7	3	31	3	2	0	**2**	1	2	Y	N	56
8	3.5	36	3	2	0	1	**1**	1	Y	N	62
9	4	41	3	2	0	**1**	0	1	Y	N	62
10	4.5	46	3	2	0	0	**0**	1	N	N	0
11	5	51	2	2	0	**0**	0	3	N	Y	100
12	5.5	56	2	2	0	2	**0**	1	N	N	0
13	6	61	1	2	0	**2**	0	2	Y	N	56
14	6.5	66	1	2	0	1	**0**	1	N	N	0
15	7	71	0	2	0	**1**	0	2	Y	N	62
16	7.5	76	1	1	0	1	**0**	2	N	N	0
17	8	81	1	0	0	**1**	1	3	N	Y	100
18	8.5	86	1	0	0	2	**1**	1	Y	N	62
19	9	91	1	0	0	**2**	0	1	Y	N	56
20	9.5	96	0	1	0	0	**0**	2	N	N	0
–	–	–	0	0	0	**0**	1	–	–	TOT	1146

MRC accounts only for 2.17%. Thus, also if the number of FOs is the same, still RL is capable of managing well the unplanned event, better than MRC.

4.3 Scenario with multiple FOs

The reasoning for single FO also applies to the case of two FOs, which occur almost the same number of times in both RL and MRC policies (i.e.,10.21% for RL and 10.84% for MRC, fourth row in Table 2), but with sensibly different average maintenance expenditures (i.e., 1549 for RL and 1586 for MRC, in arbitrary units, Table 2). Similarly to Tables 6, 7 reports the possible values of maintenance expenditures and the corresponding percentage of time in which these are encountered in case of application of RL and MRC policies. We notice that 70.31% of times RL total maintenance expenditures are smaller than or equal to 1550, whereas the percentage reduces to 39.26% of times for MRC. However, for larger cost values the difference between their percentages decreases. For instance, consider the total maintenance expenditures smaller than or equal to 1600; then, RL accounts for 90.43% of times, whereas MRC for 86.48%. This explains why the difference between the average maintenance expenditures of RL and MRC in case of two FOs is smaller than that of the single FO scenario. With respect to Table 2, finally notice that the MRC policy has a large percentage of episodes in which the FOs are larger than 2, although with similar values of expenditures. However, these scenarios account only for roughly 2% of times (i.e., they are quite rare events).

4.4 Comparison with RL in deterministic environment

A final comment seems in order about the comparison of the RL solution found in the stochastic environment with that of the DE setting [25] where it is assumed that parts cannot fail during operation. To fairly compare the policies, we assume that in the DE setting, the risk of failure before the end of the cycle is factored into the MS maintenance cost: the repair costs reported in Table 1 are summed to the expected value of the failure cost (i.e., the product between $C^{failure}$ and the failure probability within H hours, Table 8) to have a rough estimate of the total cost of the random failure. This DE setting is a simplification of the proposed framework. On the one hand, this allows a large reduction in the dimension of the state space. On the other hand, the DE setting neither considers the increase in the total number of maintenance events over the time horizon nor provides a policy in the event of a FO during operation, which changes the environment because the failed part is

no longer available for the next maintenance events. The part flow solution found by RL in the DE setting is shown in Table 9 [25], whereas Table 5, last row, summarizes the policy characteristics. The number of purchase and repair actions is the same as that of the RL in case of stochastic environment and no FO, the only difference being that in the DE setting, one additional part is scrapped with $r = 0$ and one less with $r = 1$. Consequently, the RL solution for DE requires installing a part with $r = 1$ on the last maintained GT, whereas, in the stochastic environment with no FO, a part with $r = 0$ is set on the GT at the same maintenance event (see last rows of Tables 4 and 9). This implies that the policy found by RL in DE would increase the total risk of failure if it were adopted in the stochastic environment, as one part with $r = 0$ would be put on the GT during the early maintenance events, instead of a part with $r = 1$. To sum up, the solution under the stochastic environment overcomes that found in the DE setting because it takes into account the decrease of risk of failure provided by setting newer parts on the GTs during the maintenance period rather than at the last maintenance event.

Moreover, the final maintenance expenditure in DE is 1146 in arbitrary units, which is larger than that of the RL policy in case of stochastic environment with no FO, 1050, but smaller than the average value of the RL policy, 1200 (Table 2). This difference is obviously due to the fact that the policy found by RL in the DE, does not take into account that the occurrence of a FO entails not only a failure expenditure, which is encoded in the cost of the DE setting, but also the loss of the failed part, which requires re-scheduling the MSs. This confirms that the policies found in the DE are not optimal in the stochastic one and, also, that the estimations of the maintenance expenditures provided by the RL algorithm in the DE setting are not correct, even if they encode the average cost of failure.

Then, the size of the state is much smaller than that of a real case study (i.e., $3 \cdot 10^{12}$, see Section 3). Future research work will address the issue of extending the methodology to large state spaces, which requires substituting the tabular representation of the state-action space by a suitable value-function approximation method [18, 27].

5 Conclusions

This work extends the formalization of the GT part flow management in the oil and gas industry as a SDP by considering the part failure stochastic process. RL is used as solving technique. The results of a case study inspired by a real industrial application show that RL finds a more efficient part flow policy, which increases the GT reliability, as the percentage of episodes with no forced outages is increased, but even in case of one or two forced outages, the policy found by RL results more efficient, leading to lower total maintenance expenditures.

The application of the proposed framework to a case study in which the number of turbines is larger and with larger $MNRC$ values of the parts would require large computational efforts to explore the search space for finding the optimal solution. Future research work will, then, focus on extending the proposed modeling and optimization framework and the RL algorithms for its solution.

References

1. Yang K, He C, Huang Q, Huang ZY, Wang C, Wang Q, Liu YJ, Zhong B (2016) Very high cycle fatigue behaviors of a turbine engine blade alloy at various stress ratios. International Journal of Fatigue
2. Morini M, Pinelli M, Spina PR, Venturini M (2010) Influence of blade deterioration on compressor and turbine performance. J Eng Gas Turb Power 132:3
3. Boyce BL, Ritchie RO (2001) Effect of load ratio and maximum stress intensity on the fatigue threshold in Ti–6Al–4V. Eng Fract Mech 68(2):129–147
4. Bodrov AI, Stalder JP (1998) An analysis of axial compressor fouling and a blade cleaning method. J Turbomach 120:256–61
5. Kurz R, Brun K (2012) Fouling mechanisms in axial compressors. J Eng Gas Turb Power 134:3
6. Peters JO, Ritchie RO (2000) Influence of foreign-object damage on crack initiation and early crack growth during high-cycle fatigue of Ti–6Al–4V. Eng Fract Mech 67(3):193–207
7. Eliaz N, Shemesh G, Latanision RM (2002) Hot corrosion in gas turbine components. J Eng Gas Turb Power 9(1):31–43
8. Goward GW (1998) Progress in coatings for gas turbine airfoils. Surf Coat Technol 108:73–79
9. Compare M, Martini F, Mattafirri S, Carlevaro F, Zio E (2016) Semi-Markov model for the oxidation degradation mechanism in gas turbine nozzles. IEEE Trans Reliab 65(2):574–581
10. Ng I, Parry G, Wild P, McFarlane D, Tasker P (2011) Complex engineering service systems: concepts and research. Springer, London
11. Wang W (2010) A model for maintenance service contract design, negotiation and optimization. Eur J Oper Res 201:239–246
12. Murthy DNP, Asgharizadeh E (1999) Optimal decision making in a maintenance service operation. Eur J Oper Res 116:259–273
13. Godoy DR, Pascual R, Knights P (2014) A decision-making framework to integrate maintenance contract conditions with critical spares management. Reliab Eng Syst Safety 131:102–108
14. Jin T, Tian Z, Xie M (2015) A game-theoretical approach for optimizing maintenance, spares and service capacity in performance contracting. Int J Prod Econ 161:31–43
15. Jackson C, Pascual R, Knights P (2008) Optimal maintenance service contract negotiation with ageing equipment. Eur J Oper Res 189(2):387–398
16. Bollapragada S, Gupta A, Lawsirirat C. (2007) Managing a portfolio of long term service agreements. Eur J Oper Res 182:1399–1411
17. Hu Q, Boylan JE, Chen H, Labib A (2018) in spare parts management: a review. Eur J Oper Res 266:395–414

18. Sutton R, Barto A (1998) Introduction to reinforcement learning, vol 135. MIT Press, Cambridge
19. Lieckens KT, Colen PJ, Lambrecht MR (2015) Network and contract optimization for maintenance services with remanufacturing. Comput Oper Res 54:232–244
20. Kurz J (2016) Capacity planning for a maintenance service provider with advanced information. Eur J Oper Res 251:466–477
21. Olde Keizer MCA, Teunter RH, Veldman J (2017) Joint condition-based maintenance and inventory optimization for systems with multiple components. Eur J Oper Res 257:209–222
22. Van Horenbeek A, Buré J, Cattrysse D, Pintelon L, Vansteenwegen P (2013) Joint maintenance and inventory optimization systems: a review. Int J Prod Econ 143:499–508
23. Jaturonnatee J, Murthy DNP, Boondiskulchok R (2006) Optimal preventive maintenance of leased equipment with corrective minimal repairs. Eur J Oper Res 174:201–215
24. Zanjani MK, Nourelfath M (2014) Integrated spare parts logistics and operations planning for maintenance service providers. Int J Prod Econ 158:44–53
25. Compare M, Bellani L, Cobelli E, Zio E, Annunziata F, Sepe M, Carlevaro F A Reinforcement Learning approach to optimal part flow management for gas turbine maintenance, submitted for publication to European Journal of Operational Research
26. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285
27. Szepesvári Cs (2010) Algorithms for reinforcement learning. Morgan and Claypool
28. Kuznetsova E, Li YF, Ruiz C, Zio E, Ault G, Bell K (2013) Reinforcement learning for microgrid energy management. Energy Elsevier 59:133–146
29. Rahimiyan M, Mashhadi HR (2010) An adaptive Q-learning algorithm developed for agent-based computational modeling of electricity market. IEEE Trans Syst Man Cybrn Part C, Appl Rev 40(5):547
30. Barde S, Yacout S, Shin H (2016) Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. J Intell Manuf, 1–15
31. Aissani N, Beldjilali B, Trentesaux D (2009) Dynamic scheduling of maintenance tasks in the petroleum industry: a reinforcement approach. Eng Appl Artif Intell 22(7):1089–1103
32. Pontrandolfo P, Gosavi A, Okogbaa OG, Das T (2002) Global supply chain management: a reinforcement learning approach. Int J Prod Res 40(6):1299–1317
33. Giannoccaro I, Pontrandolfo P (2002) Inventory management in supply chains: a reinforcement learning approach. Int J Prod Econ 78(2):153–161
34. Kim CO, Jun J, Baek JK, Smith RL, Kim YD (2005) Adaptive inventory control models for supply chain management. Int J Adv Manuf Technol 26(9–10):1184–1192
35. Ehleiter A, Jaehn F (2016) Housekeeping: foresightful container repositioning. Int J Prod Econ 179:203–211
36. Kim KH, Lee KM, Hwang H (2003) Receiving operations for yard cranes in port container terminals sequencing delivery. Int J Prod Econ 84:283–292
37. Zio E (2007) An introduction to the basics of reliability and risk analysis, vol 13. World Scientific
38. Wang YH, Li TH, Lin CJ (2013) Backward Q-learning: the combination of Sarsa algorithm and Q-learning. Eng Appl Artif Intel 26:2184–2193