European Journal of Operational Research 292 (2021) 579–595

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

A performance-centred approach to optimising maintenance of complex systems

E. Barlow*, T. Bedford, M. Revie, J. Tan, L. Walls

Department of Management Science, University of Strathclyde, Glasgow G4 0GE, Scotland

ARTICLE INFO

Article history:
Received 21 October 2019
Accepted 5 November 2020
Available online 13 November 2020

Keywords:
Maintenance
Performance-centred maintenance
Maintenance optimisation
Markov decision process
Reinforcement learning

ABSTRACT

This paper introduces performance-centred maintenance (PCM) as a novel approach to maintain systems when dual consideration is given to operational performance and degradation condition. We consider situations where performance and condition do not necessarily deteriorate at the same rate typified by, say, an ageing system still achieving good performance or a new system performing poorly. In this problem context, competing interests may arise between different decision-makers, such as operators and maintainers, since alternative strategies may benefit either performance or condition at the expense of the other. To address this challenge we introduce a theoretical framework for the PCM approach and discuss key characteristics of the modelling problem. The general PCM approach is motivated by a real-world industrial system for which maintenance decisions required to be optimised. A specific application is shown for the industry problem which we model by a Markov decision process capable of interrogating decisions over multiple time-scales. We obtain an exact solution using dynamic programming. We also explore a less computationally challenging heuristic using a reinforcement learning algorithm and evaluate its accuracy for the large-scale industry model. We show that optimal maintenance policies from a PCM model can provide decision support to both maintainers and operators taking account of both perspectives of the problem.

1. Introduction

The key idea behind performance centered maintenance (PCM) is to model industrial systems in such a way that it enables informed conversations between operators and maintainers about optimal maintenance. As usage patterns and customer demands become more varied, we argue that there is a need to move to a more sophisticated approach than currently considered in the literature to take account of performance, condition state, and the different types of maintenance action which can focus on a variety of issues including maintaining performance and avoiding critical failure. Our contribution is the novel PCM approach which both provides a general theoretical framework for an important industrial maintenance-operations modelling challenge and is capable of translation to specific model solutions for defined domain problems. We abstract the general problem from our motivating industry case and examine the gaps in relevant maintenance modelling literature to justify the need for PCM.

1.1. Motivating industry problem

Our motivation for developing the PCM approach is based on work with a coal-fired power plant (Alkali, Bedford, Quigley, & Gaw, 2009; Barlow, Revie, Bedford, & Walls, 2013; Bedford & Alkali, 2009; Bedford, Dewan, Meilijson, & Zitrou, 2011). The plant has four generating units each of which have eight grinding mills. The grind quality of the mills influences the amount of energy produced as well as emissions to the environment. We call this the *performance* of the mill. It is influenced by the physical state of the mill but can be "tuned" through interventions. On the other hand, as they wear, the mills can be subject to critical failure, and the *condition* of the mill determines the likelihood of critical failure. Performance and condition are not dependent – a mill can be worn but performing very well, while it may also be in top condition but not performing well – and this is a prototype case for many similar cases in industry where performance and condition are distinct. The key operational research issue motivating this paper is that there are typically two different decision makers involved, one in charge of performance (and motivated by production demands), and one in charge of condition (and motivated by maintenance and longer term asset management

* Corresponding author.
E-mail address: euan.barlow@strath.ac.uk (E. Barlow).

https://doi.org/10.1016/j.ejor.2020.11.005

issues). While details differ from one commercial and technical setting to the next, there is a common problem of managing the conflicts between short term production gains and long term asset maintenance costs. Managing these conflicts requires an approach that we call Performance Centered Maintenance (PCM).

The general approach to PCM, set out in Section 2, has to provide the flexibility to take account of key areas of conflict in any particular case. To illustrate this we go into further detail about our motivating problem. Each unit is considered as a separate generator due to regulatory requirements and its output is sold in advance on an hourly basis. Failure to meet that agreed output forces the owner to replace the lost production at short notice from elsewhere in the market at cost (increasing output at another unit is not allowed under these rules). Hence failure or poor performance of one mill can have an immediate financial impact, and operators will try to run the plant – at the risk of greater wear of remaining mills – in order to meet targets rather than face penalties. We discuss the PCM model for the motivating problem in Section 3.

As well as interdependencies arising through the mode of operation, there are additional complexities arising through maintenance activity. Throughout the life of the mill, there are two types of maintenance action: a one-week service, and a five-week overhaul. In terms of preventive maintenance each action addresses a particular type of failure mode: a service impacts performance shortfalls, whereas an overhaul impacts condition-based failure. Practical restrictions mean it is not possible to maintain more than one mill per unit at a given time, and that a maximum of two maintenance crews are operational across the site at a given time – one crew performing an overhaul on one unit and one crew performing a service on a different unit. An additional layer of complexity is thus added by the sharing of maintenance resource across the units, which from a performance and regulatory perspective are not linked. A key modelling choice is therefore the choice of whether to model at the unit or the whole plant level. These two approaches are considered in Sections 4 and 5.

Our case, while complex and unique, contains features typical of many real-world industrial systems. Important for the general PCM approach is the distinction between performance and condition. By a condition variable, we mean some measurement of the system which is related to a failure mode, for example, wear of a casing. By a performance variable, we mean some measurement of the system that is related to the quality or quantity of its outputs, for example the fineness of the coal powder. Additionally, we might want to consider indicator variables which are prognostic indicators of the system state (used perhaps because the actual system state is not, or only with difficulty, observable), and the value of production might also be variable.

The conflicting operational and maintenance demands of a system need to be accounted for by explicitly considering the operational performance and the degradation condition of an asset as two separate entities. This allows the impact of any maintenance decision upon the performance and condition to be modelled, and provides a method for analysing systems where some trade-off between operation and maintenance is required. For example, maximising the production output of an asset requires both the condition and the performance of the asset to be optimally managed, whereas, maximising the availability of an asset only requires condition modelling. It is this joint consideration of condition and performance which makes the approach distinctive.

The PCM approach enables a practitioner to trade-off long-term asset reliability with short-term operational requirements. Practical applications where this approach will be particularly beneficial include systems where the operational performance of an asset is not perfectly correlated with the degradation condition of the asset throughout its lifetime, or assets where maintenance or build errors can have minimal impact on the condition but a severe impact on the performance.

1.2. Positioning PCM within the literature

Current trends in maintenance tend to focus either on preventive approaches, where faults are identified at an early stage before a critical failure occurs, or predictive approaches where the current state of the asset is used to predict the onset of future faults. For surveys capturing the historical development of maintenance modelling, see Cho and Parlar (1991), Wang (2002), Nicolai and Dekker (2008), Si, Wang, Hu, and Zhou (2011) and de Jonge and Scarf (2020). Predictive approaches rely on obtaining accurate information on the current state of an asset, and the field of condition-monitoring has received much attention for this purpose. Condition-monitoring involves taking regular measurements which can be used to infer the degradation condition throughout the life of an asset. A condition-based maintenance approach utilises this condition-monitoring data to reduce uncertainty about time to failure to optimise the maintenance strategy. Whether this is useful in practice depends on a number of issues which include: the time required to start up the desired maintenance intervention (which could be longer than the predicated time to failure); and the need to balance off equipment use against increased risk of failure. This latter point is rarely considered in the maintenance literature but, depending on the business circumstances and management structure, decisions to run equipment or take it off line for maintenance may be the preserve of a manager outside the maintenance function who may repeatedly also prioritize short term commitments (for example, electricity generation) over maintenance. Under such circumstances the traditional separation between maintenance optimisation models and operational performance models creates a barrier for the maintenance manager. The proposed PCM approach aims to address this industry challenge by supporting an informed conversation between operators and maintainers about the relative costs of maintenance and loss of production against continued production and later maintenance.

The power plant system described in Section 1.1 can be considered a multi-asset system when considered from the point of view of a single boiler and turbine fed by eight coal mills, while the overall plant consisting of four units might also be considered as a portfolio of assets in the language of Petchrompo and Parlikad (2019). Furthermore, since the mills are the least reliable assets, one could also usefully study the portfolio of mill assets. Keizer, Flapper, and Teunter (2017b) and Petchrompo and Parlikad (2019), in their recent reviews of the current state of industrial maintenance for multi-asset systems, suggest traditional maintenance optimisation approaches are limited in their applicability due to the simplicity of the system being considered. Whilst this might have contributed to the limited application and implementation of academic studies by practitioners, it has also contributed to the development of increasingly complex models as researchers strive to represent sufficiently realistic system complexity to obtain usable solutions. See Vu, Do, Barros, and Bérenguer (2014) and Li, Deloux, and Dieulle (2016), amongst others, and Section 3.2 for further examples. We address this gap by developing a methodologically sound, useable model for a real industry system.

Petchrompo and Parlikad (2019) report seven classes of decisions considered in the literature on multi-asset system maintenance. The first class of intervention (maintenance) policies is the key area for application of the PCM approach - that is, dealing with corrective, preventive and condition based maintenance interventions, to which we add performance based interventions. The challenge in solving this problem increases as the asset dependencies are properly accounted for in the optimisation process. Other decision classes are also of interest from the PCM perspective. In particular, intervention scheduling decisions (determining optimal downtime arrangements) is important in relation to the allocation of resources shared across multiple assets or units. Also, asset prioritisation decisions arise when assets make an unequal contribution to the system value and so assets must be prioritised for maintenance under limited resources. This has obvious parallels with our system although under PCM it is further complicated, although, since each asset makes contributions to both condition and performance metrics. While the maintenance optimisation literature reports seven mutually exclusive classes of decision problems, we view the PCM approach as supporting multiple types of decisions thus relating to multiple decision classes, either explicitly or implicitly.

Keizer et al. (2017b), Petchrompo and Parlikad (2019) and de Jonge and Scarf (2020) each provide a useful discussion and categorisation of the relevant literature in terms of the nature of dependencies between the elements of a multi-asset system. However, the generalised PCM approach implies particular dependencies between assets. Petchrompo and Parlikad (2019) focus on performance dependence, stochastic dependence and resource dependence. Here performance dependence arises through the physical reliability structure – for example, whether the system is parallel, series or k out of N; stochastic dependence arises through effects such as failure induced effects or load sharing; resource dependence arises through issues such as workforce constraints. For our case we should add another type of dependence: regulatory dependence. Regulatory dependence occurs when the assets are operated in ways that are shaped by rules of a regulator or market trading mechanism. While all of these dependencies arise in the system we study, the regulatory dependence has a particularly strong impact. The electricity market trading mechanism for power plants in the UK demands that electricity production is sold in one-hour slots by power generation units. Failure to deliver that power is costly, but it can only be delivered by the unit that was contracted. This induces operator behaviour that might lead to stochastic load sharing dependencies amongst the mills within the unit. Failure outside the contracted period can be compensated by other units (as long as they were not at full capacity), leading to load sharing across other units. Furthermore, there are resource dependencies across all four units due to sharing of maintenance crews. Hence our motivating system displays dependencies described by Petchrompo and Parlikad (2019), a type of dependency not described there, and additionally, the system displays switching between these dependencies through time which can be driven by operator behaviour, market conditions, and asset performance.

In addition to the reviews considered above, there are also distinct areas of the literature which relate to specific aspects of the PCM concept. For example, the body of work by Frangopol and colleagues considers the performance-based maintenance of deteriorating structures, with a particular emphasis on deteriorating bridges (see reviews by Frangopol (2011) and Frangopol and Soliman (2016), and references therein on maintenance optimisation). In this setting, the performance of the structure is measured through a condition-state measure and a reliability-based measure, where the condition-state is used to refer to more superficial deterioration (such as cracks and road-surface condition) and the reliability-based measure is used to refer to the structural reliability. Corrective maintenance can improve either one or both of these indexes and maintenance actions are optimised subject to operational constraints. These studies can be considered specific representations of the general PCM modelling framework we present in Section 2 in that both operational and maintenance interests are given consideration, albeit that one of these is essentially fixed in their calculations.

There are some studies within the condition-based maintenance literature adopting a performance-monitoring approach. For example, rather than monitoring vibration or acoustic output from the system, some measure of an asset's operational performance, such as temperature or flow, is monitored instead. Performance-monitoring has been applied in settings such as wind turbines (Jia, Jin, Buzza, Wang, & Lee, 2016), marine renewable energy systems (Mérigaud & Ringwood, 2016), chemical plant compressors (Xenos, Kopanos, Cicciotti, & Thornhill, 2016) and diesel engines (Kökkülünk, Parlak, & Erdem, 2016). However, typically performance-monitoring is employed to address either performance deterioration or as a surrogate measure of the asset degradation, and there is a lack of consideration given to the trade-offs described above when competing interests require that both condition and performance are considered simultaneously as in PCM.

Another area of work which shares similarities with the PCM approach is the problem of jointly optimising maintenance planning and production scheduling. Corrective and preventive maintenance actions each incur a loss of production capacity, and with a stated product demand to be satisfied this therefore gives rise to competing interests from operational and maintenance perspectives. Aghezzaf, Jamali, and Ait-Kadi (2007) present early work on this problem. The joint maintenance and production problem is typically considered from a production scheduling perspective, with maintenance actions activated by age-based or condition-based triggers, and more complex production capability and requirements. Recent examples of this broad approach include Fakher, Nourelfath, and Gendreau (2018), Ghaleb, Taghipour, Sharifi, and Zolfagharinia (2020), Cheng and Li (2020) and Wang, Lu, and Ren (2020). There is, however, a growing body of work taking a more maintenance focused approach to the problem, where the decision of when to conduct a maintenance activity is included as a decision variable to the optimisation problem. Early examples of this approach include the works by AlDurgam and Duffuaa (2013) and Bajestani, Banjevic, and Beck (2014), and more recent examples include Paprocka (2018), Ouaret, Kenné, and Gharbi (2018), Ekin (2018), Kang and Subramaniam (2018), Alimian, Saidi-Mehrabad, and Jabbarzadeh (2019) and Ao, Zhang, and Wang (2019). This latter approach has more commonality with the PCM framework; however, these problems typically do not exhibit the complex failure structure and interaction which we shall define through the PCM framework.

1.3. Summary and overview

The premise of this paper is that the actions taken by operators and maintainers impact on each other's roles, and that by considering these together it is possible to gain insights that lead to improvements in the overall performance of the system. The roles of operator and maintainer are, typically, distinct. The operator is focused on ensuring that production meets quality and quantity targets, possibly focusing on cost reduction (e.g. energy and production efficiency), while the maintainer is focused on ensuring that the equipment is kept in a state where the operator can carry out their function. Typically the role of an operator will focus on short term issues, while those of the maintainer will more naturally look at the longer term state of the system. This distinction, which is clear in the system we study, is one that depends on the production method requiring relatively "fast" intervention by operators to keep production at an optimum level, and relatively "slow" intervention by maintainers to keep the physical infrastructure operational.

A common approach in the literature is to consider operators and maintainers separately, and indeed in practice there will be constraints placed on these roles that encourage such separation, for example: limits on the permitted operating state of equipment (e.g. speed, temperature, pressure etc); conditions under which maintenance is carried out (e.g. time based, condition based etc);

the separation of budgets between different functions; the separation of reporting lines between different functions. These factors all mitigate against an overall optimisation of the system which would take into account aspects such as potentially varying costs and value of production, of outages and of costs and benefits of maintenance. For example, in the system we study, the value of power generated can vary considerably as can the costs of outages. Hence it can be of interest from an economic point of view to push the equipment beyond its usual operating limits (without violating safety limits) even if this increases degradation. Without an understanding of the overall impact on the financials such action, or the prevention of such action, is likely to cause friction between operators and maintainers. Hence a modelling approach is needed that allows for such issues to be evaluated. We propose PCM as that approach and describe a theoretical framework in Section 2.

Many variants of modelling approaches are possible within the broad PCM framework and the details of the Markov decision process model selected for the industry case has been chosen because of its particular features, as discussed in Section 3. We do not claim that it is always necessary to explicitly model the actions of the two decision-makers within one model, and indeed in the model developed for our industry problem we have not done that. However our view is that by starting with a conceptual model in which the two roles are explicitly represented, we can then construct models which support the adoption of new approaches to identify the boundary between operator and maintainer action.

For the industry problem, we approach the maintenance optimisation at two levels. First we take the unit of analysis to be the set of eight grinding mills within a generating unit, as discussed in Section 4, since this is where the main dependencies between operations and maintenance occur. Then we jointly optimise the maintenance at the plant and unit levels as explained in Section 5. For the unit level model we develop both exact and heuristic solutions for the mill-unit maintenance optimisation to provide flexibility for model use contexts where different computational resources might be available. We compare the relative performance of these optimisation solutions based on both computational accuracy and implementation considerations. The additional size and complexity of the problem at the plant level means that we adopt a heuristic approach for maintenance optimisation. Section 6 presents our conclusions and suggests further work.

2. Framework for performance-centered maintenance

In general, we consider that each asset in a system can be described by various aspects of its operational performance or degradation condition. Further, observable information about the state of the system may be available to the decision-makers – typically reflected in measured sensing data or judgemental operator data. Each measure of degradation condition is associated with a failure mode and a threshold level, and a degradation failure occurs if a condition measure deteriorates to its threshold level (note that this includes the possibility of critical failure where the threshold level simply represents the critical failure). Multiple failure modes can be simultaneously considered for a single asset through a competing risks reliability model (Bedford & Alkali, 2009; Bedford et al., 2011), where an asset is deemed to be failed at the first instance any failure mode process crosses its associated threshold. Each measure of performance provides information about the operation of the asset and there may also be an associated threshold level where substandard performance leads to unacceptable operation and hence to functional failure. The observable information may be dependent on the underlying system state and used to provide inference on the degradation and performance measures if these are not directly observable. Additionally there is a time dependent demand for performance which the system operator seeks to achieve; the asset may therefore be operated in a way that speeds up degradation in order to meet the required performance.

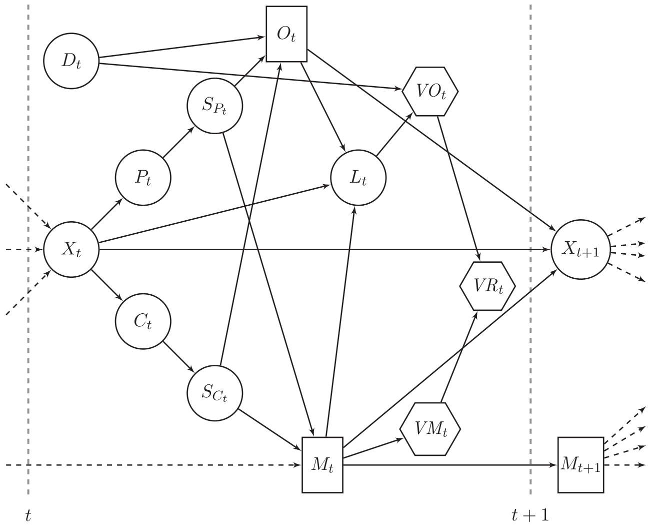

Fig. 1. Schematic of the PCM modelling framework and influences at times t and $t+1$, where at time t: X_t is the underlying asset state, P_t and C_t are the performance and condition components of this true state, respectively, S_{P_t} and S_{C_t} are the performance and condition states revealed by the sensor data, respectively, L_t is the level of production, D_t is the demand, O_t is the operator decision, M_t is the maintainer decision, VO_t is the operational return, VM_t is the cost of maintenance, and VR_t is the overall return.

For this type of system an optimised approach to operation and maintenance sets out a strategy for maintenance and also a strategy for operation that specifies in what ways the asset can be used when its condition is degraded. Given the different business objectives of the operator and maintainer, a PCM model aims to provide a basis for discussion between these stakeholders by identifying linked decision strategies delivering the required performance consistently over time with low maintenance costs.

In mathematical notation (and where all the variables may be multivariate), we think of an asset as having an underlying state $X(t)$ at time t, where X is a stochastic process. For simplicity we write $X_t = X(t)$, and will use the t-subscript to denote the value of a variable at time t. The asset state is revealed by sensor data S_{P_t} and S_{C_t} from which the asset performance state, P_t, and condition state, C_t, are inferred, respectively. The operator will attempt to meet an exogenous demand D_t which yields a return VO_t through the level of production L_t, and will influence the future state of the asset X_{t+1}. The maintainer will consider carrying out maintenance on the asset at a cost of $-VM_t$, and thus influencing the future state of the asset X_{t+1}. The net return from the system, VR_t is a combination of the operational return and the maintenance cost. Note that the maintenance options will each address a particular set of condition and performance-based variables, and each action represents a trade-off between immediate costs and future rewards. The maintenance actions take place over a longer timescale than operator decisions and there are resource constraints which limit the availability of each potential maintenance action.

The conceptual relationship between the variables is illustrated in Fig. 1, which shows the operator decision O_t and maintainer decision M_t as well as the other variables discussed above. If we just consider this from the perspective of the maintainer then we could replace the operator decision node by a random variable node – in this case the O_t node in Fig. 1 would change from a square decision-node to a circle random process-node. The maintainer understands what kinds of decisions the operator is likely to take, and builds these into the model. This model would be capa-

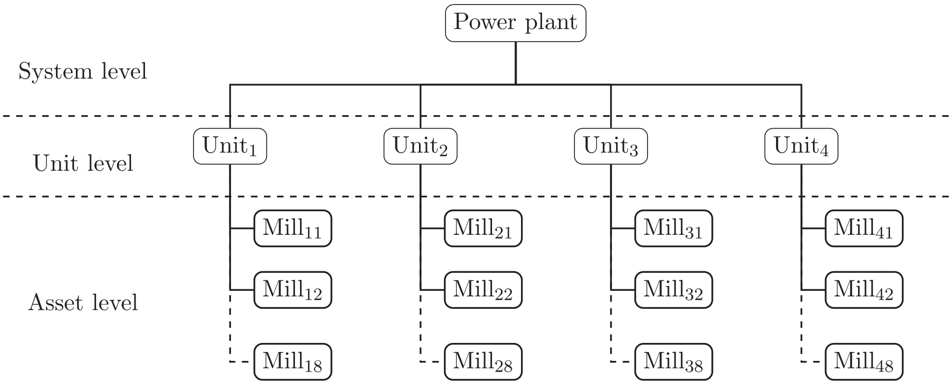

Fig. 2. Schematic illustrating the hierarchical structure of the different levels at which the power plant maintenance and operation can be considered: at the asset level a single mill is considered, at the unit level all eight mills within that unit are considered concurrently, and at the system level all four units and their corresponding mills are considered concurrently.

ble of supporting conversations between operators and maintainers about their respective strategies. Because the operator decision timescales are shorter than those of the maintainer planning horizons, it makes sense for the maintainer to build operator behaviour into the model and then assess how different operator behaviours would affect costs and benefits.

Furthermore, it is possible that the maintenance capability and demand are shared across different assets within a system, in which case the generated value and maintenance costs are also determined by the whole system and by the decisions to switch across demand and maintenance resources between assets. For example, this is true for the system we study since it is possible to switch demand between different parallel mills within a generating unit, each of which has shared maintenance resources. It is also possible to switch production between different generating units to some extent, as well as sharing maintenance resources across their mills. The same conceptual model as above is applicable, by taking parallel assets shared between the same decision-makers.

In addition to these factors, the overall optimisation is also affected by the choice of the unit of analysis. For the system we study, the eight mills contributing to the performance of a single generating unit are more-or-less interchangeable. Furthermore there are four different generating units within the plant. From an operational point of view, electricity contracts specify which unit will supply the power. So, if a unit fails between agreeing the contract and the time of supply, then it cannot be replaced by one of the other units. However, if it fails before the contract is made, then there is no issue with supplying the market with any of the other working units. From a maintenance point of view there is a limited group of maintainers but they can be used across different mills and different units.

Modelling can be performed at different levels within the overall system, and different simplifying assumptions would have to be made at these different levels in order to be realistic. While the best approach might be to build a model at the highest level, such an approach might be infeasible in practice because of the computational complexity involved. Therefore a mix of exact and heuristic approaches, applied at different levels within the system and using appropriate simplifying approaches, can be required. We aim to explicate these issues in the specific model development for our industry problem.

3. Building a PCM model for the industry problem

Fig. 2 illustrates the physical structure of the coal-fired power station described in Section 1.1 in a form we reference when model-building. We can approach modelling of this system at different levels. The same broad principles are applicable at each level, but the modelling complexity increases from asset level to system level. The value of models at each level depend on the form of additional interaction they can capture. As discussed in Sections 1 and 3.1, the most significant interactions between operations and maintenance in this problem occur at unit level, where the choice of maintenance action can represent a trade-off between short-term gain and long-term operability.

3.1. Structuring the model

We elicited the expertise of plant operators and maintainers, and sought to use this expertise and problem-knowledge to drive our modelling choices such that the problem structuring represents the real-world with acceptable accuracy and provides tractable decision-making support.

The degradation condition failure modes of the critical mill components are measured in terms of the volume of processed coal. In order to view maintenance and operational considerations in the same frame of reference, we need a temporal re-expression of volume. This is accomplished by considering the average cumulative volume of coal processed during the operating life of a mill before a given component fails and then defining the earliest such volume of coal (the first failure threshold) to be the average lifetime of a mill. This transition from volume to time assumes that a mill is operated in a consistent manner; while not true in every case, the main differences in operating conditions arise during maintenance periods and through changes to the operating workrate, each of which are accounted for explicitly in the modelling discussed below. As such, the engineering feedback was that this time-based representation provided a reasonably accurate representation of the average life-time of a typical mill.

Typically, the degradation condition of a mill, C_t, is believed to deteriorate at a steady rate due to the gradual wear of the mill elements through use, and so a simple model of the degradation condition was agreed upon. The deterioration rate (and hence the failure rate) is dependent on the state of the unit (i.e. the state of all eight mills in the unit). If all eight mills are fully operational and under normal operating conditions then they are assumed to degrade monotonically following the same process. However, if one of the mills has a reduced level of performance and there is a requirement to meet demand, then the remaining mills will be worked harder with a consequential increase in degradation rate. Therefore the transition probabilities between states for a given mill are also dependent on the current state of the unit. To capture this behaviour the impact of the different operating regimes is reviewed. Two work-rates (operating modes) for full performance are typically deployed: normal operation and increased work-rate. First the average duration of operating life is established for a mill under normal operation and frequent services. The effect of increased work-rate on the normal-operation life-time duration is then established, and the engineering feedback was that the life-time duration would decrease linearly with increasing work-rate. This process allows the average duration of operating life to be identified under each mode of operation. The life-time duration is then evenly partitioned into the number of condition states being considered; each condition state therefore represents a particular range in time during a mill's operating life. It is assumed that a mill will deteriorate gracefully with time in that the condition will only degrade by up to one level in a single time-step. Thus the probability of transition for C_t between subsequent degradation condition states is expressed as the rate of change between states, derived from the number of weeks typically spent in each degradation condition state for a particular work-rate level.

In terms of operational performance, P_t, energy production from the plant is produced (and measured) at a unit level. The contribution of a single mill to the overall unit output is classed into three performance levels. A two stage process was adopted to define the classes together with the maintainers. First to create ordinal levels of performance for each of the amount of coal currently being processed (the throughput-rate) and the coal grind quality. Then, secondly, to form overall mill performance levels for paired combinations of the through-put and grind quality classes. The overall mill classes are defined to be full (satisfactory), reduced (unsatisfactory) and offline (grossly unsatisfactory) performance. The transitions of P_t between each performance level were then assessed for each condition level under normal operating conditions. The dependence of the transition rates between performance levels on the mill work-rate is captured indirectly by recognising that the rates of transition between performance levels are dependent on the mill condition.

It is assumed that the sensor information available to a maintainer directly represents the true performance and condition states of a mill, so that the sensor nodes S_{P_t} and S_{C_t} are incorporated into the state nodes P_t and C_t.

The profits accrued over a given time-step, VR_t, come from the direct costs of the maintenance (VM_t – calculated from the lost generating capacity) and the production gained from operational mills, VO_t. An offline mill has zero production. A fully operational mill will contribute more to production if working at an increased rate to compensate for other mills in reduced production.

The two types of action available to a maintainer (represented by M_t) are distinct, and as noted in Section 1 each action is deployed to address different types of failure mode. A service is an interim measure performed several times between overhauls. In terms of a preventive action, a service is carried-out to improve the mill's operational performance. In comparison, a preventive overhaul is carried-out before a severe failure occurs to return the mill to a good-as-new state. After a service the performance of a mill, P_t, is restored to full, although there is no effect on the overall long-term condition of the mill, C_t. On the other hand, an overhaul always restores both condition and performance to the highest level. Both types of maintenance activities can also be carried out as reactive maintenance actions (in response to minor – repairable – faults or severe failures, respectively), and as periodic maintenance actions (related to the cumulative volume of coal processed by a mill); however, the focus here is on modelling preventive maintenance actions. The role of the operator decision-maker, O_t, is simplified by assuming the modes of operation as described above, for standard and increased work-rates. That is, all mills will work at a standard rate unless any mill is offline, in which case all mills at full performance will work at the increased work-rate. Therefore the role of the demand variable, D_t, is essentially removed, and the operator node becomes a variable node influenced by the current state of the mills.

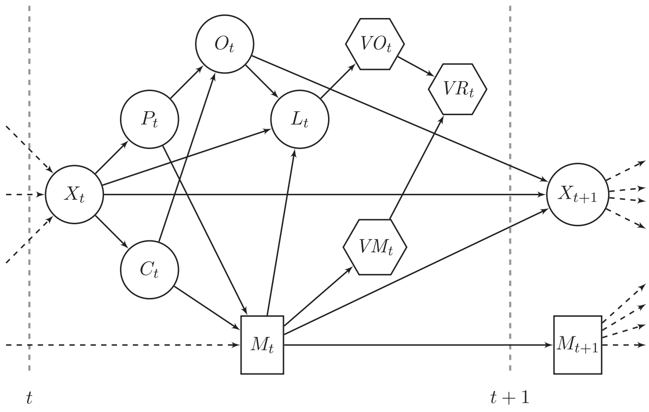

Fig. 3. Simplified schematic of the PCM modelling framework and influences at times t and $t+1$ for the power plant maintenance problem, where the notation is consistent with Fig. 1.

Through the above modelling choices, we can simplify the general PCM modelling framework shown in Fig. 1 for this particular problem; the reduced schematic is shown in Fig. 3.

3.2. Maintenance optimisation model selection

Several assumptions are now formalised to focus the scope of the solution approach.

The individual mills are assumed identical in all respects – for example states, (conditional) degradation rates, associated maintenance actions, maintenance costs, and (conditional) production capacity. Therefore each mill is considered to be a particular instance of a generic mill. However, as discussed above there are dependencies between the mill processes. By a similar extension, all four generating units are considered to be identical. The state for a unit at time t is represented by $X(t) = (P(t), C(t))$, and for different mills within the unit these are distinguished as X_j, P_j and C_j, for mills $j = 1, \ldots, 8$. When consideration is given to the broader problem of maintaining all four units, the corresponding quantities are denoted as in $X_{i,j}$, where i represents the unit and j represents the mill. Time is assumed to be discrete, with a time-step corresponding to one week. Technicians are assumed to carry out maintenance actions over a period of several consecutive shifts, and so it is assumed that maintenance decisions are taken at discrete intervals, such as weekly. Additionally, it is assumed that performance $P(t)$ and condition $C(t)$ have discrete states since they are each measured against ordinal classes. The value gained from operation of a mill is assumed to be dependent on its operational performance measures, and the operating value gained from a unit is therefore defined as $v(P_1(t), \ldots, P_8(t))$.

Under these assumptions, a natural modelling paradigm is to describe the state of the unit, $X(t)$, as a Markov process. Since maintenance decisions are required to be taken at discrete intervals, even if they are not possible at every time-step, we have a problem of sequential decision making under uncertainty – evaluating the best decision to make at each time-step, where the system evolves stochastically with dependence on the decision taken as well as the state of the system itself. A Markov decision process (MDP) is a stochastic dynamic programming method providing a natural framework for this problem.

A detailed analysis on MDPs is provided by Howard (1960), and these models have been applied to maintenance optimisation problems for many years; see, for example, Hontelez, Burger, and Wijnmalen (1996), Gürler and Kaya (2002), Moustafa, Maksoud, and Sadek (2004). Papakonstantinou and Shinozuka (2016) present a contemporary overview of the application of MDPs and their variants specifically to structural maintenance problems. Recent developments to the MDP framework include Sun, Ye, and Chen (2018) who consider a parallel multi-component system where degradation of each component is modelled using a Wiener process. Keizer, Teunter, Veldman, and Babai (2018) consider a parallel system subject to both failure dependence and economic dependence. Deng, Santos, and Curran (2020) present a practical application of MDP to schedule maintenance operations for an aircraft fleet, incorporating several strategies into their problem formulation in order to reduce the action-state space of this large-scale problem. Several authors have applied MDPs to problems with competing interests, such as Ao et al. (2019) to jointly optimise maintenance planning and production scheduling (discussed in Section 1). Other recent examples include joint optimisation of maintenance and inspection planning for a non-repairable multi-component series system (Ba, Cholette, Borghesani, Ma, & Kent, 2020), joint optimisation of maintenance planning and budget allocation for a multi-facility network (Shi, Xiang, Xiao, & Xing, 2021), joint optimisation of maintenance planning and inventory control of spares for single asset (Eruguz, Tan, & van Houtum, 2018) and multi-asset systems (Keizer, Teunter, & Veldman, 2017a; Liu, Yang, Pei, Liao, & Pohl, 2019; Wang & Zhu, 2020), and joint optimisation of maintenance planning and operation of multi-asset systems (Compare, Marelli, Baraldi, & Zio, 2018; Eruguz, Tan, & van Houtum, 2017).

An MDP is defined by the state space X, the set of actions (or decisions) A and the actions permitted from the multi-dimensioned system state $\mathbf{x} \in X$, $A(\mathbf{x})$, the reward function $R(\mathbf{x}, a)$ which gives the reward associated with state $\mathbf{x}$ under action a, and the probability of transition from state $\mathbf{x} \in X$ to state $\mathbf{y} \in X$ when action a is taken, $P(\mathbf{y}|\mathbf{x}, a)$. In MDP terminology, a policy defines the action to be taken in any given state, and the goal when applying an MDP is to find the optimal policy such that the expected reward available from each state over the planning horizon is maximised. Mathematically, this is written as

$$V^*(\mathbf{x})=V^{\pi^*}(\mathbf{x})=\max_{\pi\in\Omega}\left\{V^{\pi}(\mathbf{x})=E\left[\sum_{t=0}^{T-1}\gamma^t R(\mathbf{x}_t,\pi_t(\mathbf{x}_t))|\mathbf{x}_0=\mathbf{x}\right]\right\}, \tag{1}$$

where the operator $E[.]$ indicates the expected value, T is the number of time-steps over the current planning horizon, $\mathbf{x}_t$ is the state at the beginning of the $t+1$th time-step, π_t is the policy employed over the $t+1$th time-step and π defines the policy to be employed across all time-steps in the planning horizon. The optimal policy is denoted by π^*, and the resulting value under this policy is $V^*(\mathbf{x})$; this is the maximum value which can be gained over T time-steps through any available action. MDPs provide a user with the optimal strategy to apply for each potential system state, along with the expected value of this state over the time-horizon considered. A discount factor, γ, is included, where $0 < \gamma \le 1$ ensures that current and future values are equitable. This enables both finite- and infinite-horizon problems to be considered, with $T \to \infty$ in the latter case.

We develop our maintenance optimisation model as an MDP firstly for a single unit in Section 4 since this is where the key interactions between maintenance and operations occur, before considering a plant level model in Section 5.

4. Unit level maintenance optimisation model

We begin by making additional assumptions about the maintenance scheduling and resource availability in order to decouple a single unit from the power system in a manner which has relevance to the real-world problem. Then we develop an exact solution for the MDP model using traditional dynamic programming. Since the complexity of the unit level model is already sufficiently high, obtaining exact solutions is computationally challenging. Indeed we used high performance computing facilities to obtain the exact solution. Therefore we also explore a heuristic approach which would allow industry users to generate solutions using more standard computational resources. We examine the relative accuracy of the heuristic to our exact solution so that we might assess usefulness for future application.

4.1. Maintenance resource availability

Each of the four units operates and deteriorates independently of the other three; however, the restrictions on the number of maintenance jobs which can be performed simultaneously apply across all four units. The choice of maintenance action taken for one unit will therefore influence the maintenance options available for the other three units.

A preliminary attempt to solve this unit level problem is presented in Barlow et al. (2013), where the restrictions on maintenance crew availability are incorporated into the decision problem by structuring this as a multi-level MDP. At a given time-step all available maintenance actions on the unit are evaluated; specifically these are: perform a service on any of the eight mills, perform an overhaul on any of the eight mills, or do no maintenance at this time-step. Considering the actions to either do no maintenance or to perform a service on one of the mills, these are evaluated in a standard iteration of the MDP algorithm. The expected value of each option is evaluated by considering the possible state-transitions over one week and the corresponding values associated with these transitions. To evaluate actions that overhaul one of the mills, Barlow et al. (2013) formulate a sub-problem which imposes the restriction that an overhaul can only be completed once per 20 weeks. A five-week overhaul is followed by a 15-week period where the only available actions are to perform a service or do no maintenance. This ensures that the policies considered involve realistic utilisation of the option to perform overhauls. For modelling simplicity, no such restrictions are placed on the utilisation of services. At a given iteration of the MDP, an overhaul action is therefore evaluated by considering the potential evolution of the state during a period where a five week overhaul is followed by an optimised combination of no maintenance and services over 15 weeks. Due to available computing power at the time, Barlow et al. (2013) reduced the problem to consider only six mills within a unit.

We present here an alternative approach to incorporating the restrictions on maintenance crew availability into the decision problem. Two key drivers in formulating this alternative model are to also incorporate realistic restrictions on the availability of performing services on a particular unit, and ultimately to build upon this to generate a realistic model for the plant level problem.

To ensure that resources are distributed evenly across all four units, each maintenance crew is assumed to operate on a rotating basis subject to the requirement that two maintenance actions cannot be performed simultaneously on a given unit. Over a 20-week period each unit will receive one five-week overhaul followed by 15 weeks where no overhaul is carried out as the overhaul crew rotates to the other three units. During this 15-week period a unit will receive a one-week service no more than once every three weeks as the service crew rotates between the three units which are not currently being overhauled, and in-between services a unit will receive no maintenance. The available maintenance actions for a given unit will then be a repeating 20-week sequence which consists of a five-week overhaul and five repetitions

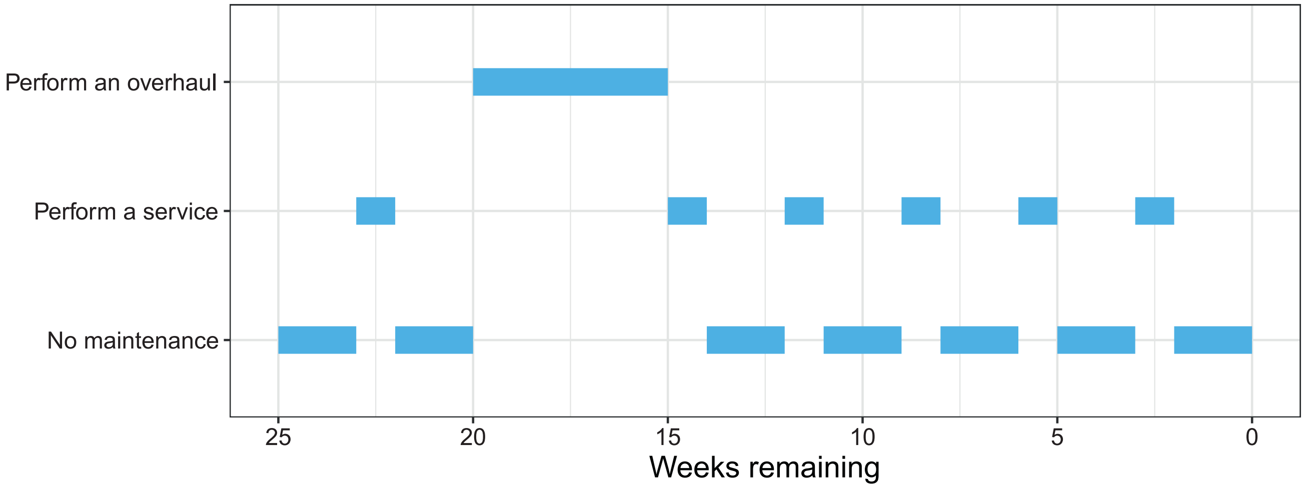

Fig. 4. Illustrative sequence of maintenance actions available to a unit, with a given number of weeks remaining in the planning horizon.

of (some combination of) *perform a service, do no maintenance, do no maintenance*. An example of this maintenance scheduling is illustrated in Fig. 4. By assuming that resources are evenly distributed between units in this way, the maintenance of each can be considered independently of the other three units. Hence the maintenance of a single generic unit can be considered since it is assumed to be representative of each of the four units.

4.2. Dynamic programming exact solution

Implementing the above maintenance sequence within a decision framework produces an MDP where the state-value calculations are periodically varied to reflect the nature of the 20-week maintenance sequence. A finite-horizon approach is considered here as the units ultimately have a finite lifetime which more generally, depending on the system, may be in the short- medium- or long-term future. As performance and condition are both defined as discrete states, the problem is therefore structured as a finite-space discrete-time MDP, which is well-suited to the standard MDP solution methods (Sutton & Barto, 1998).

The common approach to find the optimal policy π^* which satisfies Eq. (1) is to apply dynamic programming algorithms, of which there are two prominent variants: value iteration and policy iteration. The approach taken here is value iteration (VI), which seeks to find the maximum value to be gained from each state by identifying the expected value to be gained in one time-step under each action, and iteratively building this value throughout all time-steps in the planning horizon, given how the unit state could probabilistically evolve throughout this period. Hence following a VI framework, the expected optimal value of a unit state $\mathbf{x}$ at week i (that is, the ith iteration of the MDP) is defined as

$$V_i^*(\mathbf{x}) = \begin{cases} R(\mathbf{x}, NM) + \gamma \sum\limits_{\mathbf{y}\in X} P(\mathbf{y}|\mathbf{x}, NM) V_{i-1}^*(\mathbf{y}), & \text{for } i \in I_{NM}, \quad (2) \\ \max\limits_{a\in A_{SM}(\mathbf{x})} \left\{ R(\mathbf{x}, a) + \gamma \sum\limits_{\mathbf{y}\in X} P(\mathbf{y}|\mathbf{x}, a) V_{i-1}^*(\mathbf{y}) \right\}, & \text{for } i \in I_{SM}, \quad (3) \\ \max\limits_{a\in A_{OM}(\mathbf{x})} \left\{ R_5(\mathbf{x}, a) + \gamma \sum\limits_{\mathbf{y}\in X} P^5(\mathbf{y}|\mathbf{x}, a) V_{i-5}^*(\mathbf{y}) \right\}, & \text{for } i \in I_{OM}, \quad (4) \end{cases}$$

where I_{NM}, I_{SM} and I_{OM} represent the sets containing weeks where the only available actions are: no maintenance, a service or no maintenance, and to begin an overhaul or perform no maintenance for five weeks, respectively, and NM, $A_{SM}(\mathbf{x})$ and $A_{OM}(\mathbf{x})$ are the corresponding subsets of actions. For the maintenance sequence illustrated in Fig. 4, the sets I_{NM}, I_{SM} and I_{OM} are each highlighted by those weeks on which activity is scheduled on the respective rows. Eq. (4) is structured to reflect that, for weeks in the set I_{OM}, a decision made persists for the five subsequent weeks. This is modelled through $R_5(\mathbf{x}, a)$ and $P^5(\mathbf{y}|\mathbf{x}, a)$, where the latter represents the state-transitions over the five-weeks of the overhaul and is defined as element $\{\mathbf{x}, \mathbf{y}\}$ of the fifth power of the one-week transition matrix under action a. The value $R_5(\mathbf{x}, a)$ is defined recursively via

$$R_k(\mathbf{x}, a) = R(\mathbf{x}, a) + \gamma \sum_{\mathbf{y}\in X} P(\mathbf{y}|\mathbf{x}, a) R_{k-1}(\mathbf{x}, a), \quad \text{for } k = 1, \ldots, 5, \quad (5)$$

and represents the expected value accumulated during the five weeks of overhaul a.

The right-hand side of Eqs. (2) and (3) each consist of the immediate reward gained over the next time-step by taking action a in state $\mathbf{x}$, combined with the discounted expected value over the remaining $(i-1)$ time-steps, with only a single action (no maintenance) considered in Eq. (2). Eq. (3) has a form which most resembles the standard format of the VI definition, and the subtle differences between Eqs. (3) and (4) are required to model decision-making over the different time-scales of the two types of maintenance action. To solve Eqs. (2)–(4) it is necessary to define the value returned from each state under each action with no time-steps remaining. A typical assumption is that there is no value to be gained from any state at the end of the planning horizon (that is, $V_0^*(\mathbf{x}) = 0 \,\forall\, \mathbf{x} \in X$); applying this assumption here implies that there is no production value to be returned from operation beyond the end of the planning horizon, which seems reasonable given that the planning horizon is the operational life of a unit. Eqs. (2)–(4) are then solved for $i = 1$ and iterated until $i = T$, with the decision which maximises the value being selected at each time-step.

More generally, for a system of N_m assets, where the number of possible condition and performance states for each asset are given by N_c and N_p, respectively, the total size of the state-space is $(N_c N_p)^{N_m}$. Traditional solution approaches such as VI iteratively build optimal policy decisions for increasing time-horizons; however, to do so requires that the full state-space is evaluated at each iteration. Solving even simple problems can therefore be computationally challenging, in terms of memory requirements and computation time (Chang, Hu, Fu, & Marcus, 2013; Deng et al., 2020). With eight mills, 4 levels of condition and 3 levels of performance the total number of states is 12^8 (which is approximately 4.3×10^8) and working with this number of states is intractable. To address this challenge a state-space aggregation which exploits the structure of the problem is used to reduce the number of states. The eight mills within a unit are assumed to be identical in terms of their production value in any state, their probability of transitioning between any states, and the effect of the available actions in any state. The critical features of a unit state are therefore the number of mills within the unit which are in each mill state (that is, each combination of the performance and condition levels). The numbering of mills 1,2,...,8 within a unit can therefore be based on their ranking in terms of performance and condition levels (rather than, for example, their physical location within the plant). Taking this numbering approach, it becomes clear that many of the 12^8 potential states are essentially reordered versions of an equivalent unit state with mills numbered by performance and condition ranking, and can be fully represented by the unique combinations of mill states within a unit. This equivalency under re-ordering is exploited to reduce the state space from $(N_cN_p)^{N_m}$ to $\binom{N_cN_p+N_m-1}{N_m}$, which is just over 75,000 states for this problem – providing a substantial reduction without any loss of accuracy.

Even after application of the state-space aggregation, implementing a traditional VI approach to this problem is not straightforward, requiring a state-transition matrix of 75,000×75,000 to be stored and updated at each iteration. High-performance computing facilities were utilised to conduct these computations. Eight cores (each with 4 gigabyte RAM and 2.66 gigahertz CPU) were used to implement a parallelised version of the VI system defined in Eqs. (2)–(4) by partitioning the state-transition matrix, and this enabled a solution to be found in under 20 minutes.

4.3. Heuristic solution

4.3.1. Choice of reinforcement learning method

A large body of research exists on approximations to the MDP, which aim to improve the computational burden and to deliver tractable solution approaches. They include state-space aggregation (Zhou, Guo, Lin, & Ma, 2018; Zhou, Lin, Sun, & Ma, 2016), simulation-based methods (Chang et al., 2013; Ohno, Boh, Nakade, & Tamura, 2016), and reinforcement learning (RL) – sometimes referred to as neuro-dynamic programming (Barde, Yacout, & Shin, 2019; Compare et al., 2020). RL is a field of machine learning, which aims to iteratively develop an approximation to a true MDP model by repeatedly simulating potential trajectories of state-transitions through the planning horizon, and revising the value of a trajectory as the iterations proceed. Sutton and Barto (1998) and Gosavi (2003) discuss various RL methods. A recent approach is deep reinforcement learning - combining RL with deep learning - that has been applied by several authors to large-scale maintenance problems (see for example (Huang, Chang, & Arinez, 2020; Liu, Chen, & Jiang, 2020; Zhang & Si, 2020)). RL methods are typically applied to large-scale MDP, as well to as semi-MDP problems. Watkins and Dayan (1992) and Singh, Jaakkola, Littman, and Szepesvári (2000) demonstrate that, under certain conditions, RL methods such as Q-learning are guaranteed to converge to accurate value estimations and optimal policies.

Applications of MDP and semi-MDP to maintenance problems often obtain near-optimal results; however, there have been few studies which tackle complex systems with multiple degradation-dependent components. Instead, studies typically assume a simple degradation process (Das, Gosavi, Mahadevan, & Marchalleck, 1999; Xia, Zhao, & Jia, 2008) or consider small problem scales (Gosavi, 2004; 2014). Furthermore, for large-scale problems RL is typically only benchmarked against other heuristics or simplified optimisation problems, as exact solutions are not available.

In contrast to dynamic programming approaches such as VI, RL methods progress forwards in time through the planning horizon, repeatedly updating the value estimations. Therefore, the standard form of the VI definition (e.g. as in Eq. (3)) is rewritten as

$$V_t^a(\mathbf{x}) = R(\mathbf{x}, a) + \gamma \sum_{\mathbf{y}\in X} P(\mathbf{y}|\mathbf{x}, a) V^*_{t+1}(\mathbf{y}). \tag{6}$$

A single state is selected and this state is propagated through the planning horizon, with the trajectory of the state defined by selecting an action at each time-step and a corresponding state-transition under this action. In an application, the initial starting state may be selected to be the system state which is observed in practice. The action selected at each time-step is then selected randomly with a preference towards those actions which have a higher estimated expected value, and the state-transition at each time-step is selected randomly with a preference towards those states which have a higher estimated transition probability – based on the transition distributions of the governing random variables. For our case, the governing random variables are the mill condition- and performance-levels. It is these features which deliver the computational benefits of RL methods in comparison with VI – rather than simulating every transition and action, preferential selection focuses efforts on those which are more likely to be observed or selected, and the memory burden is substantially reduced by requiring only the transition distribution of the governing variables. All expected values are initially estimated as zero, and at each iteration this is updated for the relevant state-action pair at the current time-step by applying a temporal difference adjustment (Sutton & Barto, 1998), which is typically a function of the immediate reward and future estimated expected value (given the action and state-transition which have been simulated). The starting state continues to progress along the simulated state-action trajectory until the end of the planning horizon, at which point the state reverts to the initial starting-state and a new planning horizon is simulated, with the expected value estimations continually updated as the planning horizons are restarted. This process repeats until a predefined number of iterations is reached, at which point the optimal action is selected as that action which returns the highest estimated expected value from the initial starting state.

The RL method we apply is Q-learning (Watkins, 1989), which is one of the most widely investigated. In Q-learning, the estimation of expected value $V_t^a(\mathbf{x})$ (for state $\mathbf{x}$ under action a at time-step t) is denoted as $Q(\mathbf{x}, a, t)$, defined as

$$Q(\mathbf{x}, a, t) = \sum_{\mathbf{y}\in X} P(\mathbf{y}|\mathbf{x}, a)[r(\mathbf{x}, a, \mathbf{y}) + \gamma \max_{b\in A(\mathbf{y})} Q(\mathbf{y}, b, t+1)], \ \forall(\mathbf{x}, a), \tag{7}$$

where $r(\mathbf{x}, a, \mathbf{y})$ is the immediate reward returned with a transition from state $\mathbf{x}$ to state $\mathbf{y}$ under action a.

If the simulated state-trajectory results in a transition to state $\mathbf{y} \in X$ at time-step $t+1$, then the incremental adjustment to the expected value estimate is defined in terms of an updating algorithm as

$$Q(\mathbf{x}, a, t) \leftarrow (1-\alpha)Q(\mathbf{x}, a, t) + \alpha(r(\mathbf{x}, a, \mathbf{y}) + \gamma \max_{b\in A(\mathbf{y})} Q(\mathbf{y}, b, t+1)), \ \forall(\mathbf{x}, a), \tag{8}$$

where $0 < \alpha < 1$ is a controlling parameter, and ← indicates that the expression on the left-hand side is updated by that on the right. The updated estimation therefore retains a proportion of the current expected value estimation, and includes an adjustment by a weighted temporal difference. The temporal difference in this case approximates Eq. (6), while utilising the maximum estimated expected value which could potentially be returned from the next state, $\mathbf{y}$, as a proxy for the expected maximum value over the remaining $(T-t)$ time-steps.

The controlling parameter α is referred to as the step-size parameter. This should decay as the Q-learning progresses and

Gosavi (2003) recommends definitions such as $\alpha_i = M/(N+i)$, where M and N are large positive constants such that $M < N$ and i denotes the i^{th} iteration of the algorithm. Multiple approaches to govern the selection-preference of the available action options have been investigated previously. These include pure exploration policies (where the selection preference is uniformly-distributed between the available options) and general selection rules such as ϵ-greedy policy and softmax policies such as Boltzmann selection rule (Gosavi, 2003). The pure exploration approach is straightforward to implement and can avoid convergence to locally-optimal policies, but the uniform nature of the search can result in slow convergence of the Q values; in contrast the selection rules can typically demonstrate faster convergence, but require additional tuning parameters to be defined for each application.

4.3.2. Q-learning application

The Q-learning method is applied to the unit level problem in order to examine the effectiveness of this approximation technique on a large-scale MDP problem with a complex degradation structure. To the best of our knowledge, this is the first comparison of Q-learning accuracy on this scale of problem with an exact solution (rather than alternative heuristic approaches or solutions to simplified optimisation problems).

To explore the efficacy of the Q-learning algorithm, 10 states are selected from the state-space for investigation. These 10 states are chosen to represent 10 realisations of the physical status of the mill-unit ranging from “good” to “bad” (in terms of the combinations of condition and performance levels), and with each state having a different optimal action. Each of these representative states is used in turn to initialise an application of the Q-learning algorithm, executed in Matlab on a standard PC (8 gigabyte RAM and a single processor with 3.40 gigahertz CPU). For each of the 10 representative starting states one million Q-learning iterations are carried out, taking approximately 2.5 hours of computation time. The step-size parameter α is defined with $M = 85,500,000$ and $N = 90,000,000$. The discount factor is set as $\gamma = 0.9988$. Action selection is controlled by a mixed scheme: for the first 500,000 iterations a pure exploration rule is used, and the Boltzmann selection rule is used subsequently, with the “temperature” (Gosavi (2003)) set as 100,000,000. All random numbers utilised within the action selection and state selection processes are generated from a uniform distribution, and the random number generator is seeded with the default Matlab values. These parameter definitions have been selected based on the engineering domain knowledge of this problem and numerical experimentation on a small-scale version of the problem.

Diagnostics from the 10 Q-learning applications are displayed in Table 1. V_{opt} represents the exact optimal expected value of a state as determined from the VI analysis, and V_{est} represents the estimated optimal expected value of the state as returned from the Q-learning. The corresponding exact and estimated optimal actions are represented by action$_{opt}$ and action$_{est}$, respectively. The Q-learning algorithm converges for V_{est} from each of the 10 initial states. The accuracy of the converged values are evidenced by the results for the absolute relative error between the estimated and exact optimal expected value, with the Q-learning algorithm estimating the optimal expected value within an error of 0.05. The behaviour of the value convergence is typified in Fig. 5, which shows the accuracy per iteration for each of the 10 representative starting states. The Q-learning estimate achieves an accuracy of over 90% in under 200,000 iterations (that is, during the pure exploration phase), and then shows marginal improvements to this accuracy over the remaining iterations.

Table 1 shows that for most initial states, the Q-learning algorithm converges to a single estimate of the optimal action. Exceptions are initial states 2, 3, 5 and 6, where the estimates fluctuate between two or three alternatives as the iterations progress. Even in these cases, the optimal action from the exact solution is one of the estimates. To provide additional context to this convergence behaviour, the exact expected value which would be returned from taking a particular estimate of the optimal action (denoted as $V_{\text{action}_{est}}$) is compared with V_{opt} (the exact expected value which is returned from taking the optimal action). Consider initial state 2, as the iterations progress the Q-learning algorithm fluctuates between returning actions 1 and 16 as the estimated optimal action. The exact optimal solution is action 16. The exact expected value returned by taking action 1 from initial state 2 is 99.85% of the optimal expected value which can be returned for this state (and would be achieved by taking action 16). The results in Table 1 show that in situations where Q-learning does not converge to a single estimate of the optimal action, then the sub-optimal alternative actions would still return an expected value within 1.7% of the optimal expected value. Furthermore, additional analysis conducted, but not shown, reveals that for initial states 7-10 the next-best alternative to the optimal action would return an expected value less than 96% of the optimal expected value, and in each of these cases the Q-learning algorithm successfully returns a single optimal action estimate.

Table 1
Comparison between Q-learning and Value Iteration calculations of discounted value for 10 starting states, using Eqs. (8) and (2)–(5), respectively. The errors are defined as $|V_{est} - V_{opt}|/|V_{opt}|$.

Initial state	Error	action$_{opt}$	action(s)$_{est}$	$V_{\text{action}_{est}}/V_{opt}$
1	0.0019	1	1	100.00%
2	0.0088	16	1	99.85%
			16	100.00%
3	0.0043	15	1	98.82%
			15	100.00%
			16	99.47%
4	0.0248	14	14	100.00%
5	0.0157	13	13	100.00%
			14	99.42%
6	0.0428	12	12	100.00%
			13	98.96%
			14	98.30%
7	0.0027	11	11	100.00%
8	0.0140	10	10	100.00%
9	0.0092	9	9	100.00%
10	0.0143	8	8	100.00%

As an extension to the above comparison, we also briefly compare the performance of the Value Iteration and Q-learning algorithms for average (rather than discounted) value. Eqs. (2)–(8) feature the discount factor γ. This is a mechanism commonly used in industry to equate decisions over long timescales, where decision outcomes could be realised immediately or in several years time. For values of γ close to zero, the effect on the model will be that more expensive decisions are delayed into the future (where the financial impact of these decisions are reduced due to the scaling by γ). However, a γ value close to one provides a more equitable comparison between immediate and future decisions. In the investigations presented in Table 1 and Fig. 5, $\gamma = 0.9988$. A natural extension to these investigations is therefore to consider an average-value formulation of the problem, setting $\gamma = 1$. Average-value MDPs have been widely studied in the literature, in both finite- and infinite-horizon formulations. Employing a Q-learning formulation for an average-value MDP based on Gosavi (2004), the Q-learning and Value Iteration performance on calculating the average-value is presented in Table 2. All other parameters are the same as in the previous investigations. It is clear that the performance of the Q-learning algorithm is more varied on the average-value problem, with absolute relative errors ranging from 0.0034-0.2682. The same features as discussed for Table 1 are also evident,

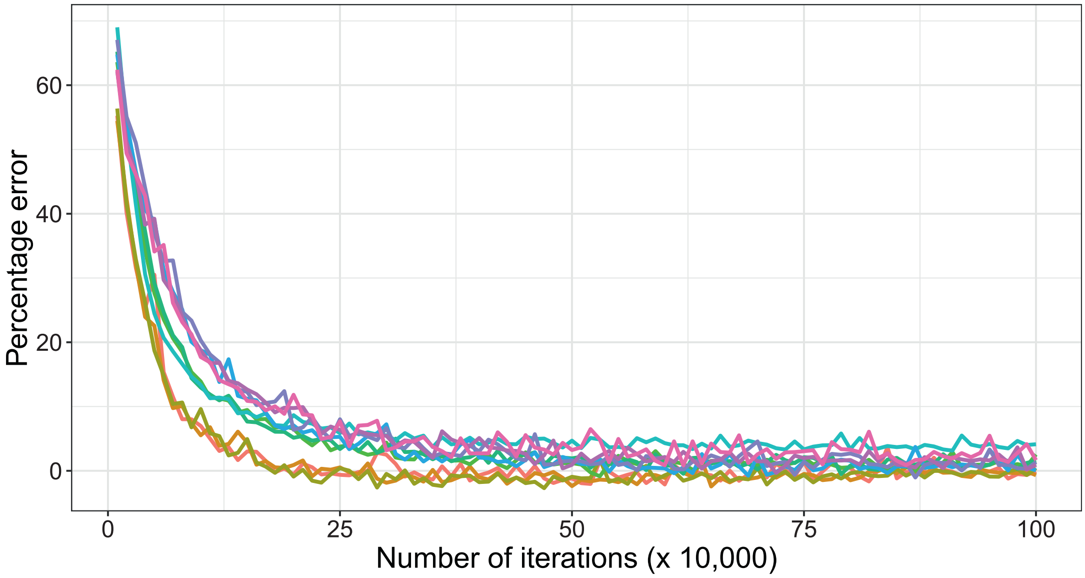

Fig. 5. Error (as % of the optimal expected value) in every $10,000$ iterations for each representative starting state.

Table 2
Comparison between Q-learning and Value Iteration calculations of average-value for 10 starting states. The errors are defined as $|V_{est} - V_{opt}|/|V_{opt}|$.

Initial state	Error	action_{opt}	action(s)_{est}	$V_{\text{action}_{est}}/V_{opt}$
1	0.2682	1	1	100.00%
2	0.2355	16	16	100.00%
3	0.0461	15	15	100.00%
			16	99.51%
4	0.0093	14	15	98.32%
			16	97.73%
5	0.0045	13	13	100.00%
			14	99.44%
			15	97.04%
6	0.0436	12	12	100.00%
7	0.2444	11	11	100.00%
			16	95.91%
8	0.1687	10	10	100.00%
			16	95.97%
9	0.0344	9	9	100.00%
			15	95.96%
			16	95.65%
10	0.0034	8	8	100.00%

with the algorithm fluctuating between several actions for some starting states. In each case, however, the actions proposed by the Q-learning algorithm would provide an average value within 5% of the optimal value.

4.4. Findings from unit level analysis

With such a large state-space it is challenging to meaningfully display the optimal service and overhaul policies through time for the unit model. At a high-level both service and overhaul optimal policies can be categorised into three groups: *logical* – the mill with the poorest performance or condition state, respectively, is always maintained; *consistent* – the maintained mill is consistent through time, but has better performance or condition state, respectively, than another mill in the unit; or *varying* – the mill selected for maintenance transitions between the *logical* and *consistent* selections as the planning horizon increases. From the perspective of the PCM modelling framework, the most interesting aspect of these optimal policies are the trade-offs being made between (short-term) performance and (long-term) condition. To illustrate transitions in the balance of these trade-offs between different states, consider Fig. 6 which focuses on four alternative unit states that each comprise of mills with similar performance levels and identical condition levels. The four unit states have the same high level of total unit condition (summing the individual condition of the eight mills), and have slightly different degrees of total unit performance (summing the individual performance of the eight mills). Within each unit, five of the mills are identical and these are depicted in Fig. 6(a). Three mills are offline and the other two are operating at full performance but with reduced condition levels. The remaining three mills are depicted in Fig. 6(b). These remaining mills all have the highest level of condition, but their performance is either at reduced or full levels.

When all three mills have full performance (bottom row) then the total unit performance is not so critical, and the *logical* optimal policy is pursued – to maintain the mill shown in Fig. 6(a) with poor condition but full performance. We contrast this with the top row of Fig. 6(b) where the three distinguishing mills all have reduced performance. For this unit, the *consistent* optimal policy is to always maintain the mill shown in Fig. 6(a) with good condition but offline performance, leaving the mill with poor condition online in order to sustain the full performance that this mill can provide. When the three distinguishing mills have a combination of full and reduced performance (middle two rows), then as the planning horizon progresses the optimal policy transitions from the *logical* to the *consistent* approach, as the long-term advantages from enhanced condition become less valuable than sustaining performance in the short term.

This visibility and exploration of the optimal service and overhaul policies at the unit level is a key benefit to maintainers at the plant. By implicitly modelling the operator actions, the complex balance between operation and maintenance throughout the planning horizon is evaluated. Situations can then be identified where the intuitively logical maintenance approach is sub-optimal, and where an alternative trade-off between condition and performance (evolving through time as appropriate) will probabilistically provide greater returns.

5. Plant level maintenance optimisation

We now extend the unit MDP model to the joint optimisation at the plant and unit levels to support the plant manager develop a comprehensive maintenance strategy.

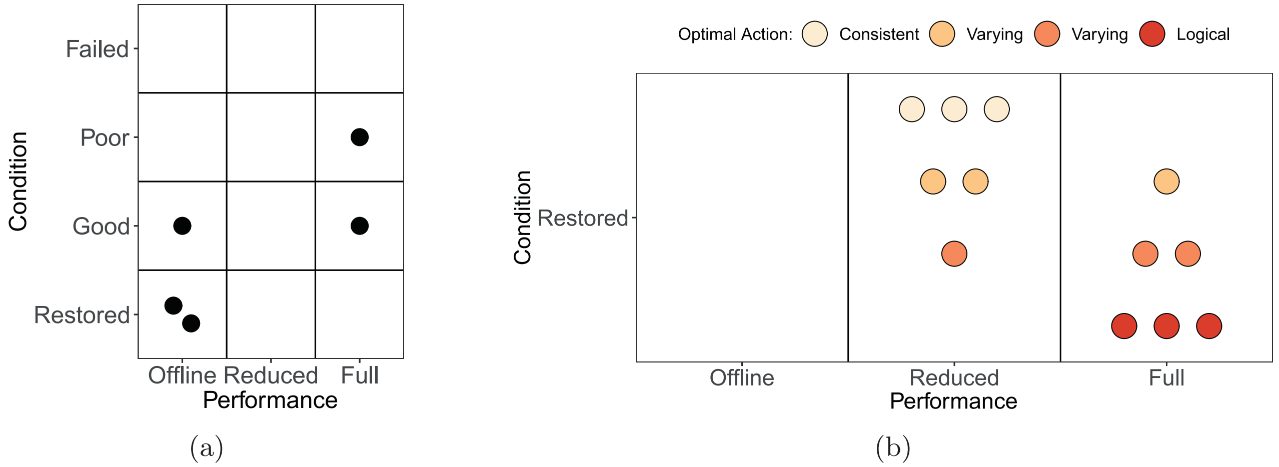

Fig. 6. Selection of four similar unit-states to illustrate the variation in optimal overhaul actions. Five mills are constant across the four unit-states and these are displayed in (a). The states of the remaining three mills are shown in a separate row of (b) for each unit-state, shaded with respect to the corresponding optimal overhaul policy.

5.1. MDP model for set of generating units

The decision choices are made over two timescales with the overhaul decisions taken on a five-week basis, and service decisions taken on a weekly basis. The logistical constraint that a unit cannot receive two maintenance actions simultaneously allows this decision problem to be framed as a multiple time-scale MDP, where the decision of which unit to overhaul considers the expected value of the optimal service actions which could be performed at the same time. Full details of an MDP with decisions made over multiple time-scales are given in Chang, Fard, Marcus, and Shayman (2003). As discussed earlier, the effect of any maintenance decision on the state of the plant is dependent on the states of the individual units, therefore maintenance decisions are made at plant level. The complexity of this problem can be somewhat reduced by recognising that the operation and degradation of a mill is dependent only on the other mills in the same unit, rather than all mills at the plant. The transitions of a single unit can be considered as independent from the other units and can be described as before. Let the state of the plant be defined as $\mathbf{x}_P = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$, the combined states of the four units. Then let $\mathbf{x}_b$ denote the state of the unit to be overhauled and $\mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e$ denote the states of the other three units, such that the set $\{b, c, d, e\}$ is a particular ordering of the set $\{1, 2, 3, 4\}$. This decision problem can intuitively be expressed as: decide which unit the overhaul crew should maintain; for each week during the overhaul decide which of the other three units the service crew should maintain. Overhaul and service decisions are each valued by taking account of the value gained from the maintained unit as well as the value gained from not maintaining the remaining units. At the beginning of an overhaul period the optimal action is therefore identified by giving full consideration to the impact of an overhaul action on all four units and the associated values. For the overhauled unit, there will be a period of five weeks where no other maintenance is performed, and for the remaining three units there will be a period of five weeks where services are optimally allocated between these three units according to the respective values of each service decision. At each time-step, the optimal expected value of a service is calculated for each unit along with the value accrued under no maintenance. The expected value for a generic unit b during an overhaul, and a generic unit c during a service and no maintenance are defined respectively from Eqs. (4), (2) and (3) as

$$V_{i_O}^{UO}(\mathbf{x}_b|\mathbf{x}_P) = V_{i_O}^{*}(\mathbf{x}_b|\mathbf{x}_P), \tag{9}$$

$$V_{i_S}^{US}(\mathbf{x}_c|\mathbf{x}_P) = V_{i_S}^{*}(\mathbf{x}_c|\mathbf{x}_P), \tag{10}$$

$$V_{i_S}^{UN}(\mathbf{x}_c|\mathbf{x}_P) = V_{i_S}^{*}(\mathbf{x}_c|\mathbf{x}_P). \tag{11}$$

where $\mathbf{x}_b|\mathbf{x}_P$ denotes the unit state $\mathbf{x}_b$ given that the state of the plant is $\mathbf{x}_P$, the subscript i_O signifies this decision is made over the five-week time-scale of an overhaul and the subscript i_S signifies this decision is made over the one-week time-scale of a service action.

The optimal service decision is then identified by maximising the return from performing a service on one of the three units, and no maintenance on the other two units. That is,

$$V_{i_S}^{UT}(\mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e|\mathbf{x}_P) = \max \begin{cases} V_{i_S}^{US}(\mathbf{x}_c|\mathbf{x}_P) + V_{i_S}^{UN}(\mathbf{x}_d|\mathbf{x}_P) + V_{i_S}^{UN}(\mathbf{x}_e|\mathbf{x}_P), \\ V_{i_S}^{UN}(\mathbf{x}_c|\mathbf{x}_P) + V_{i_S}^{US}(\mathbf{x}_d|\mathbf{x}_P) + V_{i_S}^{UN}(\mathbf{x}_e|\mathbf{x}_P), \\ V_{i_S}^{UN}(\mathbf{x}_c|\mathbf{x}_P) + V_{i_S}^{UN}(\mathbf{x}_d|\mathbf{x}_P) + V_{i_S}^{US}(\mathbf{x}_e|\mathbf{x}_P). \end{cases} \tag{12}$$

This value calculation is repeated for each week during the overhaul period to give $V_{i_O}^{UT}(\mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e|\mathbf{x}_P)$. The expected value from performing an overhaul on a unit b is then defined as

$$V_{i_O}^{U_b}(\mathbf{x}_P) = V_{i_O}^{UO}(\mathbf{x}_b|\mathbf{x}_P) + V_{i_O}^{UT}(\mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e|\mathbf{x}_P). \tag{13}$$

This entire process is repeated as the expected value of an overhaul on each unit is calculated and compared to identify the maximum expected value which can be obtained from all four units. The optimal overhaul and service actions at time-step i_O are then found by calculating

$$V_{i_O}^{U^*}(\mathbf{x}_P) = \max\{V_{i_O}^{U_1}(\mathbf{x}_P), V_{i_O}^{U_2}(\mathbf{x}_P), V_{i_O}^{U_3}(\mathbf{x}_P), V_{i_O}^{U_4}(\mathbf{x}_P)\}. \tag{14}$$

Solving Eq. (14) would generate a state-space which is the cross-product of each individual unit state-space, defined as $\left(\binom{N_cN_p+N_m-1}{N_m}\right)^4$ when the aggregation method is applied. For four units of eight mills each with four levels of condition and three levels of performance, the state space required would be over 3 × 10^{19}; simply enumerating each state would require a vast amount of memory and so implementing an MDP of this size is infeasible. Therefore it is necessary to simplify the model described above, and this is achieved here by removing the impact from a maintenance decision upon units which are not maintained by this action. That is, the dependence on the plant state is removed from the right-hand side of Eqs. (9)–(13). By assuming that each unit will be maintained in the same manner regardless of the states of the other units, there is no need to consider the full state-space of the plant model and the maintenance of each unit can be considered separately by a unit level maintenance model. The unit level model is particularly appropriate for this purpose as this model additionally gives consideration to all four units through the assumption that resources are evenly distributed throughout the plant, generating a maintenance sequence which allows for all four units to be maintained. Applying this model at a plant level results in the maintenance of each unit being described by a staggered version of the single-unit model such that the periods of overhaul and service for each unit do not coincide with maintenance periods on any of

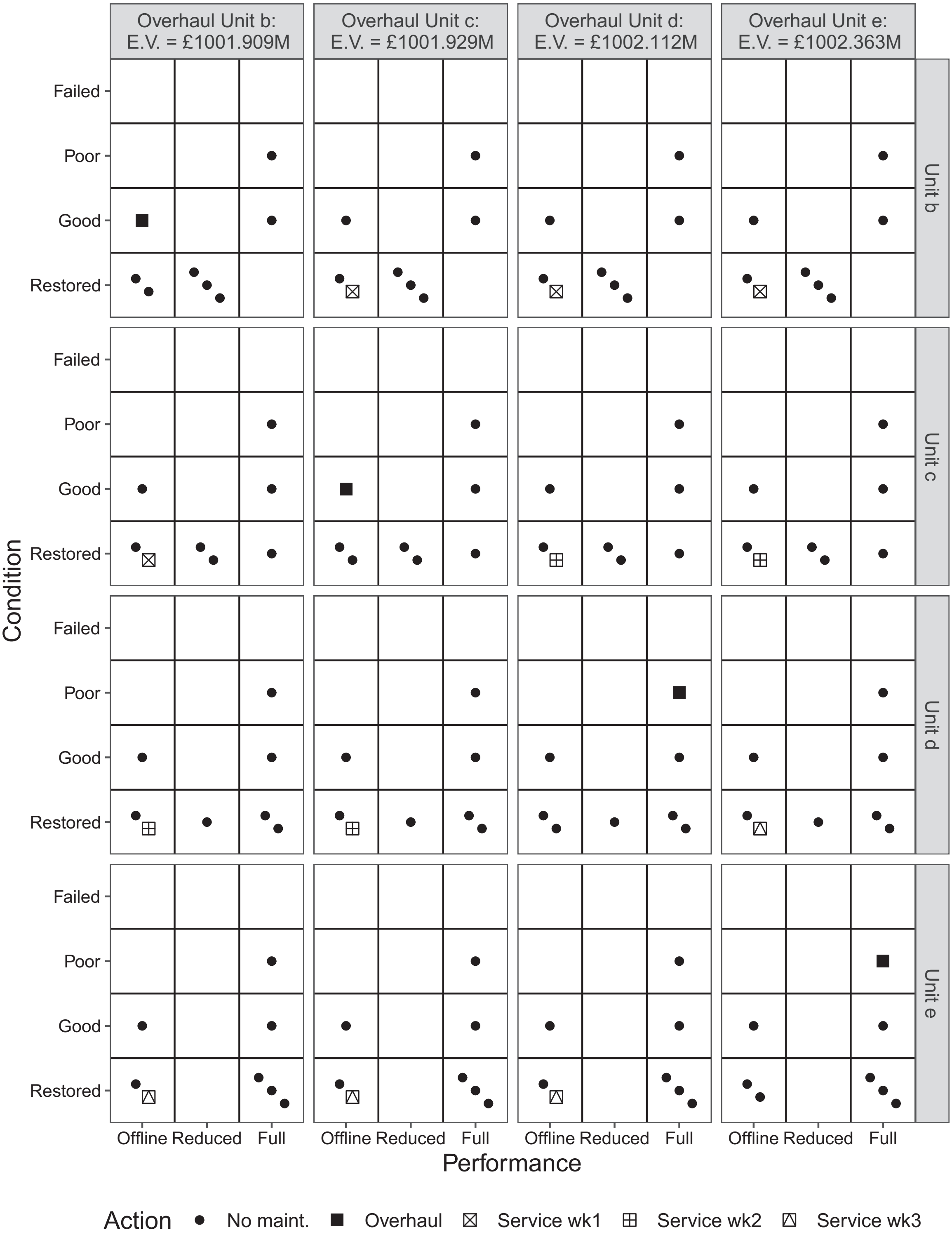

Fig. 7. Visualisation of the four high-level maintenance options (overhaul of each alternative unit), each represented by the different columns, along with the corresponding sequence of service actions over a three-week period. The rows show the maintenance actions that are carried out on each mill within a unit, under the four high-level options. Each mill is represented in terms of its condition and performance states, and the shape associated with the mill signifies the corresponding maintenance action. The value of each overhaul decision is calculated from Eq. (16) and the sequencing of service actions is determined from Eq. (15).

the other units. To achieve this, the single-unit model is combined here with a heuristic framework which enables all four units to be considered simultaneously for minimal increase in computational requirements. Appendix A provides further details, discussion and justification of this heuristic framework.

5.2. Findings from plant level analysis

An example of the plant level decision support provided to maintainers is illustrated in Fig. 7. A visual comparison of the four overhaul options is given where the value of each overhaul decision is calculated from Eq. (16) and the sequencing of service actions determined from Eq. (15). Each option informs a maintainer which unit and mill would be overhauled at a particular week, and which unit and mill would be serviced this week and for each of the next two weeks. The four unit states considered in this example are the same states considered in Section 4.4 and Fig. 6, and so there are small but significant differences in the mills that comprise each unit. The decision is modelled with approximately three years of operating life remaining; the timing of the decision was selected in order to provide a balance between the states of the mills which would be overhauled under each option.

Fig. 7 shows the optimal policy in the right-hand column: overhaul unit *e* and service in the order unit *b*, unit *c* then unit *d*. Intuitively this seems sensible. The comparatively poorer performance of units *b*, *c* and *d* is improved through services, and these are prioritised in terms of the poorest performance. Unit *e* performs best between the four units, and the condition of the unit is enhanced by overhauling the mill with the poorest condition. Comparable information can be extracted for any unit states at any stage in the planning horizon, to provide maintainers with a holistic approach to scheduling the plant maintenance operations.

The plant level maintenance model is based on the default assumption that the maintenance of each unit follows a defined sequence such that maintenance actions are distributed evenly between the four units. Any decision which results in a deviation from this sequence of actions indicates that the model is making a trade-off between long-term maintenance value and short-term operational value. For example, if a particular unit is identified to receive subsequent overhauls then, for this unit, the model has determined that short-term value should be sacrificed for long-term gain – performing a second overhaul will lose value over the duration of the overhaul, however, this has been outweighed by the long-term value expected to be gained by improving the state of the overhauled mill.

6. Conclusions and further work

Motivated by the industry example and in light of the existing literature, we have developed a novel theoretical framework for performance centered maintenance (PCM) which explicitly considers the operational performance and the condition of an asset. We explain the core concepts of PCM and articulate the reasoning underpinning the general modelling framework. The framing of the model to provide decision support to both maintainers and operators is key since the intent is to generate optimal policies that take account of both perspectives.

Through the industry problem we show how a PCM model might be developed in detail. We explain how we have abstracted the key characteristics of our system to translate the engineering descriptions and experiential understanding into statements of assumptions that allow us to represent the problem mathematically. Key modelling choices are discussed, including issues such as the level of modelling within the system hierarchy to appropriately capture the condition and performance relationship. Given the characteristics of our particular case, we have selected an MDP as the means of modelling sequential decisions under uncertainty. We developed both heuristic and exact approaches to maintenance optimisation giving two options for computation to a model user. Through experimentation we have shown new results about the relative accuracy of a *Q*-learning algorithm for an industrial scale maintenance problem. We acknowledge that future numerical experiments of a wider set of algorithms/PCM models would allow a fuller examination of computational solutions. We also discuss the key insights gained using our model and the implications for the maintainers at the power plant by providing an overview of findings from analysis based on the real data.

Our detailed modelling of the power plant not only shows how the PCM concept can be translated to a real problem, but it also allows us to gain new scientific knowledge about both the model performance and the accuracy of methods used to find optimal policies. Of course, our model is based upon the assumptions made for this particular application. Naturally other modelling choices could have been made, and these would arrive at an alternative structuring of the power plant maintenance problem. There is no requirement that the mathematical model be constrained to an MDP within the PCM framework.

Future work will focus on extending the existing modelling using the broad MDP framework while relaxing key modelling assumptions. Within our modelling, we have assumed that the machine state is continuously monitored and degrades in discrete time steps, and that monitoring information can accurately reveal the machine state. More realistically, the transitions are likely to be dependent on the duration for which a machine has been in a particular state, and the monitoring information will provide imperfect information. As such, we could extend the MDP used to a Partially Observable Semi Markov Decision Process, similar to that used in Zhang and Revie (2017). This would allow more accurate transition rates to be used and for uncertainty in the machine condition and performance to be modelled, so that the generated optimal maintenance policy is more closely tailored to the real application.

The current model assumes that operators always try to maximise demand; in reality, there are two ways this aspect could be extended. First, demand will fluctuate throughout the year to reflect the rise and fall in the use of renewable energy. Forecasting models could be used to extract the underlying trends from historical data, and this demand-series could then be incorporated as an input to the maintenance modelling. Second, instead of treating demand as fixed or random, we could consider demand as a controllable variable. On occasions, given the operating conditions and requirements on revenue generation, it is conceivable that restricting power production for a sustained period would reduce the risk of a critical failure. Incorporating such operator decisions into the model in addition to the maintenance decisions would provide a more holistic approach to managing assets, and would therefore enhance the true value of decision making support provided by the modelling. One approach to achieve this would be to include a third decision-making timescale to the model, where the operator-based production decisions are made at a higher frequency than both types of maintenance decision. This inclusion would substantially compound the dimensionality challenges of the current model, and would further drive the development of computationally tractable solution approaches.

Acknowledgements

This project was part of a Pathways to Impact grant funded by EPSRC, Scottish Power and Scottish Water, with additional funding through EPSRC grant no. EP/C53526X/1. Results were obtained using the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk), EPSRC grant no. EP/K000586/1.

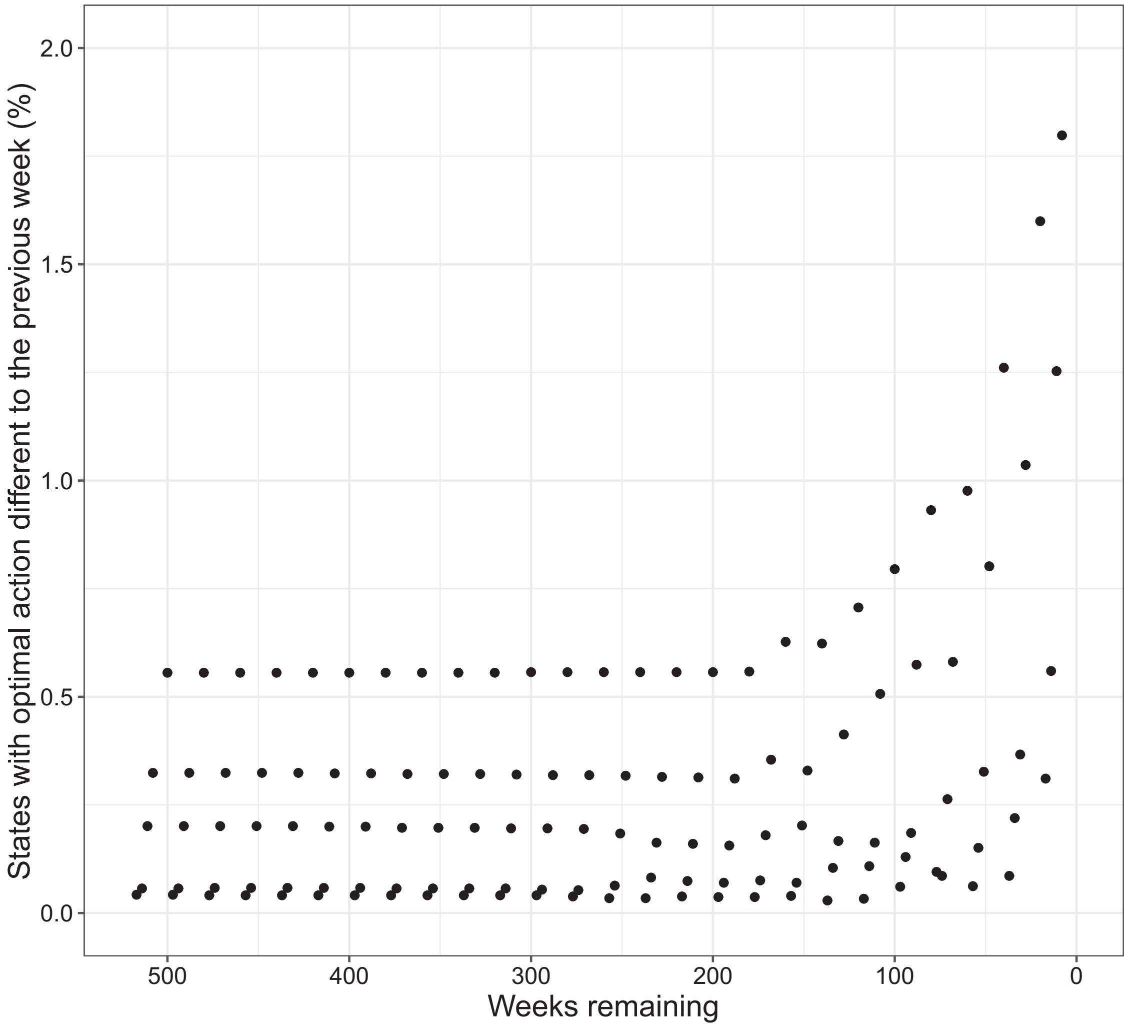

Fig. 8. Variation in optimal service maintenance actions between consecutive weeks, throughout a 10-year planning horizon.

Appendix A. Plant level heuristic framework

The heuristic framework for the plant level model introduced in Section 5 provides sufficient simplification of the problem to enable a tractable solution to be computed. This model is based upon a fundamental assumption which is demonstrated in Fig. 8.

Fig. 8 displays the percentage of states for which the optimal action for a service is different to the optimal action at the previous iteration. It is clear that for the vast majority of states the optimal service action does not vary from one week to the next (less than 2% of states varying at any given week). During the first few iterations (that is, the final weeks of the planning horizon), for some states there is no service action that can yield a reward, and for these states the option of doing no maintenance is preferable. As the planning horizon increases, however, the benefits of a maintenance action are more likely to be realised. This is the main reason for the higher levels of variation shown over the final weeks in Fig. 8. In subsequent iterations the optimal service will typically target the mill which can yield the greatest improvement in production over the remainder of the planning horizon, however, for some states alternative mills are periodically being selected for the service action. The effects of this are the non-zero levels of settled variation as the planning horizon increases. Investigations have revealed that in cases where this variation is evident throughout the planning horizon, the optimal service policy for a state cycles between servicing two alternative mills, and that these cycles are structured periodically around the weeks where an overhaul decision is available. This indicates that for a small number of states, the optimal choice of which mill to service consistently varies in the period between overhaul decisions. Although different variants of the cyclic pattern are observed, the pattern will typically be consistent throughout the majority of the planning horizon for a given state. This explains the relatively consistent patterns of variation that are observed in Fig. 8 from week to week.

The heuristic framework implemented here capitalises on this quasi-steady-state maintenance policy by making the fundamental assumption that the optimal service maintenance decision will not vary from one decision time-step to the next for all mill-unit states. Under this assumption, if the sequence of maintenance actions outlined in Section 4.1 is slightly perturbed the effect on future maintenance decisions will be minimal. For example, a service action which is identified as optimal under the sequence (*perform a service, do no maintenance, do no maintenance*) is assumed to be similarly optimal under the sequences (*do no maintenance, perform a service, do no maintenance*) and (*do no maintenance, do no maintenance, perform a service*). The optimal service actions found for each state at a given iteration from the generic single-unit model can therefore be applied to each of the three mill-units which are not currently being overhauled, as the optimal service for a given unit-state is assumed to remain constant regardless of whether the service is performed in the first, second or third week. This allows the maintenance of all three units to be considered simultaneously for a fraction of the computational demand that would be required to consider each mill-unit explicitly.

As in Eq. (12), the problem of identifying the optimal unit to service requires consideration of the impact of this choice on the serviced unit as well as the impact upon the remaining two units. From consideration of Eq. (12), the value gained from servicing a particular unit can be expressed as the difference in value between doing no maintenance or completing the service; however, calculation of this value gain would require storage of an additional matrix to record specifically the value under no maintenance which make the computational burden even more challenging. Given the maintenance sequence associated with the single-unit model, an alternative approach is to consider the value gain as quantification of the urgency of servicing each unit – the value gained from performing a service on a given unit immediately, and the value gained from performing this service in the second or third week. For simplicity these values are approximated here by the value gained through performing a service on a particular state from one week to the next. For each state this value gained can be expressed as

$$g(\mathbf{x}) = V_i^*(\mathbf{x}) - V_{i-1}^*(\mathbf{x}), \quad \text{for } i \in I_{SM}, \tag{15}$$

and the optimal order for services over a three week rotation is then obtained through a greedy heuristic where the order of units serviced is c, d, e such that $g(\mathbf{x}_c) \geq g(\mathbf{x}_d) \geq g(\mathbf{x}_e)$. A more refined approach could perform a separate run of the single-unit model where the sequence of maintenance used is instead (*do no maintenance, do no maintenance, perform a service*). This would give an accurate representation of the value obtained from delaying a given service till the end of the three week period and could therefore be used in conjunction with the value obtained from the initial run of the model to provide the optimal order for services. Over a three week period, however, these values can be expected to be extremely similar with any differences arising due to the increased benefit of performing a service as soon as possible. As these benefits will be maximised for units with a relatively low value prior to receiving a service, the underlying variation in these values is captured through the gain measure defined in Eq. (15). The extra computational effort required to perform a second run of the model is therefore unlikely to be merited by the additional accuracy gained through this run.

Taking a similar approach to define the optimal sequence of overhaul actions could be problematic in that the value of a unit can be expected to vary substantially over a 20-week period. The optimal overhaul action at a given time is therefore identified by calculating the expected value of performing an overhaul on each unit in terms of the impact of this action on all units. The value of this action is essentially a combination of the value obtained from overhauling a particular unit and the value obtained from not overhauling the remaining three units. For this purpose the single-unit model described in Section 4.1 is evaluated under two sequences of twenty-week maintenance decisions: {*perform an overhaul*, followed by five repetitions of (*perform a service, do no maintenance, do no maintenance*)} and {five repetitions of (*perform a service, do no maintenance, do no maintenance*), followed by *perform an overhaul*}. These sequences represent respectively, $V_i^{UO}(\mathbf{x})$ defined by Eq. (9), the expected value of choosing to do an overhaul at this decision, and $V_i^{U\bar{O}}(\mathbf{x})$, the expected value from the worst case scenario when an overhaul is not chosen at this decision and it is necessary to wait 15 weeks to perform the overhaul. The expected value obtained from choosing to overhaul unit b is then defined similarly to Eq. (13) as

$$V_i^{UX_b}(\mathbf{x}_P) = V_i^{UO}(\mathbf{x}_b) + V_i^{U\bar{O}}(\mathbf{x}_c) + V_i^{U\bar{O}}(\mathbf{x}_d) + V_i^{U\bar{O}}(\mathbf{x}_e), \quad \text{for } i \in I_{OM}. \tag{16}$$

By considering the value of the three units not overhauled as if these were all overhauled in fifteen weeks time, the true value obtained by performing an overhaul in five, ten, and fifteen weeks time is guaranteed to be greater than the expected value calculated above. In a similar vein to Eq. (14) the optimal overhaul action is then identified by calculating

$$V_i^{UX^*}(\mathbf{x}_P) = \max\{V_i^{UX_1}(\mathbf{x}_P), V_i^{UX_2}(\mathbf{x}_P), V_i^{UX_3}(\mathbf{x}_P), V_i^{UX_4}(\mathbf{x}_P)\} \text{ for } i \in I_{OM}. \tag{17}$$

In addition to identifying the optimal overhaul to perform at this decision-time, Eq. (16) can be used to generate a probable sequence of overhauls over the next twenty-week period as units b, c, d, e such that $V_i^{UX_b}(\mathbf{x}_P) \geq V_i^{UX_c}(\mathbf{x}_P) \geq V_i^{UX_d}(\mathbf{x}_P) \geq V_i^{UX_e}(\mathbf{x}_P)$. When it is desirable to explicitly identify the optimal overhaul action at the beginning of each five-week overhaul period, performing repeated runs of the single-unit model under appropriate manipulations of the two maintenance sequences outlined above will enable a similar approach to be taken.

The heuristic for the plant level maintenance then comprises a two-level algorithm which operates on the output from the single-unit model at a given decision-epoch – the expected value and the optimal maintenance policy for each mill-unit state. The first level identifies the optimal unit to be overhauled at this time as identified by Eq. (17); the second level then identifies through Eq. (15) which order the remaining three units should be serviced in to maximise the expected value.

References

Aghezzaf, E. H., Jamali, M. A., & Ait-Kadi, D. (2007). An integrated production and preventive maintenance planning model. *European Journal of Operational Research, 181*(2), 679–685.

AlDurgam, M. M., & Duffuaa, S. O. (2013). Optimal joint maintenance and operation policies to maximise overall systems effectiveness. *International Journal of Production Research, 51*(5), 1319–1330.

Alimian, M., Saidi-Mehrabad, M., & Jabbarzadeh, A. (2019). A robust integrated production and preventive maintenance planning model for multi-state systems with uncertain demand and common cause failures. *Journal of Manufacturing Systems, 50*, 263–277.

Alkali, B. M., Bedford, T., Quigley, J., & Gaw, J. (2009). Failure and maintenance data extraction from power plant maintenance management databases. *Journal of Statistical Planning and Inference, 139*(5), 1766–1776.

Ao, Y., Zhang, H., & Wang, C. (2019). Research of an integrated decision model for production scheduling and maintenance planning with economic objective. *Computers & Industrial Engineering, 137*, 106092.

Ba, H. T., Cholette, M. E., Borghesani, P., Ma, L., & Kent, G. (2020). Condition-based inspection policies for boiler heat exchangers. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2020.09.030.

Bajestani, M. A., Banjevic, D., & Beck, J. C. (2014). Integrated maintenance planning and production scheduling with Markovian deteriorating machine conditions. *International Journal of Production Research, 52*(24), 7377–7400.

Barde, S. R. A., Yacout, S., & Shin, H. (2019). Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. *Journal of Intelligent Manufacturing, 30*(1), 147–161. https://doi.org/10.1007/s10845-016-1237-7.

Barlow, E., Revie, M., Bedford, T., & Walls, L. (2013). Trading off asset performance and condition to model strategic maintenance decisions. In *Safety, reliability and risk analysis: Beyond the horizon* (pp. 723–731). CRC Press, Boca Raton.

Bedford, T., & Alkali, B. (2009). Competing risks and opportunistic informative maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 223*(4), 363–372.

Bedford, T., Dewan, I., Meilijson, I., & Zitrou, A. (2011). The signal model: A model for competing risks of opportunistic maintenance. *European Journal of Operational Research, 214*(3), 665–673.

Chang, H. S., Fard, P. J., Marcus, S. I., & Shayman, M. (2003). Multitime scale Markov decision processes. *IEEE Transactions on Automatic Control, 48*(6), 976–987.

Chang, H. S., Hu, J., Fu, M. C., & Marcus, S. I. (2013). *Simulation-based algorithms for Markov decision processes*. London: Springer.

Cheng, G., & Li, L. (2020). Joint optimization of production, quality control and maintenance for serial-parallel multistage production systems. *Reliability Engineering & System Safety, 204*, 107146.

Cho, D. I., & Parlar, M. (1991). A survey of maintenance models for multi-unit systems. *European Journal of Operational Research, 51*(1), 1–23.

Compare, M., Bellani, L., Cobelli, E., Zio, E., Annunziata, F., Carlevaro, F., & Sepe, M. (2020). A reinforcement learning approach to optimal part flow management for gas turbine maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 234*(1), 52–62.

Compare, M., Marelli, P., Baraldi, P., & Zio, E. (2018). A Markov decision process framework for optimal operation of monitored multi-state systems. *Journal of Risk and Reliability, 232*(6), 677–689.

Das, T. K., Gosavi, A., Mahadevan, S., & Marchalleck, N. (1999). Solving semi-Markov decision problems using average reward reinforcement learning. *Management Science, 45*(4), 560–574.

Deng, Q., Santos, B. F., & Curran, R. (2020). A practical dynamic programming based methodology for aircraft maintenance check scheduling optimization. *European Journal of Operational Research, 281*(2), 256–273.

Ekin, T. (2018). Integrated maintenance and production planning with endogenous uncertain yield. *Reliability Engineering & System Safety, 179*, 52–61.

Eruguz, A. S., Tan, T., & van Houtum, G. (2017). Optimizing usage and maintenance decisions for k-out-of-n systems of moving assets. *Naval Research Logistics (NRL), 64*(5), 418–434.

Eruguz, A. S., Tan, T., & van Houtum, G.-J. (2018). Integrated maintenance and spare part optimization for moving assets. *IISE Transactions, 50*(3), 230–245.

Fakher, H. B., Nourelfath, M., & Gendreau, M. (2018). Integrating production, maintenance and quality: A multi-period multi-product profit-maximization model. *Reliability Engineering & System Safety, 170*, 191–201.

Frangopol, D. M. (2011). Life-cycle performance, management, and optimisation of structural systems under uncertainty: Accomplishments and challenges. *Structure and Infrastructure Engineering, 7*(6), 389–413.

Frangopol, D. M., & Soliman, M. (2016). Life-cycle of structural systems: Recent achievements and future directions. *Structure and Infrastructure Engineering, 12*(1), 1–20.

Ghaleb, M., Taghipour, S., Sharifi, M., & Zolfagharinia, H. (2020). Integrated production and maintenance scheduling in a single degrading machine with deterioration-based failures. *Computers & Industrial Engineering, 143*, 106432.

Gosavi, A. (2003). *Simulation-based optimization*. New York: Springer.

Gosavi, A. (2004). Reinforcement learning for long-run average cost. *European Journal of Operational Research, 155*(3), 654–674.

Gosavi, A. (2014). Variance-penalized Markov decision processes: dynamic programming and reinforcement learning techniques. *International Journal of General Systems, 43*(6), 649–669.

Gürler, Ü., & Kaya, A. (2002). A maintenance policy for a system with multi-state components: An approximate solution. *Reliability Engineering & System Safety, 76*(2), 117–127.

Hontelez, J. A., Burger, H. H., & Wijnmalen, D. J. (1996). Optimum condition-based maintenance policies for deteriorating systems with partial information. *Reliability Engineering & System Safety, 51*(3), 267–274.

Howard, R. A. (1960). *Dynamic programming and Markov processes*. Cambridge MA: MIT Press.

Huang, J., Chang, Q., & Arinez, J. (2020). Deep reinforcement learning based preventive maintenance policy for serial production lines. *Expert Systems with Applications, 160*, 113701.

Jia, X., Jin, C., Buzza, M., Wang, W., & Lee, J. (2016). Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves. *Renewable Energy, 99*, 1191–1201.

de Jonge, B., & Scarf, P. A. (2020). A review on maintenance optimization. *European Journal of Operational Research, 285*, 805–824.

Kang, K., & Subramaniam, V. (2018). Integrated control policy of production and preventive maintenance for a deteriorating manufacturing system. *Computers & Industrial Engineering, 118*, 266–277.

Keizer, M. C. A. O., Teunter, R. H., & Veldman, J. (2017a). Joint condition-based maintenance and inventory optimization for systems with multiple components. *European Journal of Operational Research, 257*(1), 209–222.

Keizer, M. C. O., Flapper, S. D. P., & Teunter, R. H. (2017b). Condition-based maintenance policies for systems with multiple dependent components: A review. *European Journal of Operational Research, 261*(2), 405–420.

Keizer, M. C. O., Teunter, R. H., Veldman, J., & Babai, M. Z. (2018). Condition-based maintenance for systems with economic dependence and load sharing. *International Journal of Production Economics, 195*, 319–327.

Kökkülünk, G., Parlak, A., & Erdem, H. H. (2016). Determination of performance degradation of a marine diesel engine by using curve based approach. *Applied Thermal Engineering, 108*, 1136–1146.

Li, H., Deloux, E., & Dieulle, L. (2016). A condition-based maintenance policy for multi-component systems with Lévy copulas dependence. *Reliability Engineering & System Safety, 149*, 44–55.

Liu, X., Yang, T., Pei, J., Liao, H., & Pohl, E. A. (2019). Replacement and inventory control for a multi-customer product service system with decreasing replacement costs. *European Journal of Operational Research, 273*(2), 561–574.

Liu, Y., Chen, Y., & Jiang, T. (2020). Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. *European Journal of Operational Research, 283*(1), 166–181.

Mérigaud, A., & Ringwood, J. V. (2016). Condition-based maintenance methods for marine renewable energy. *Renewable and Sustainable Energy Reviews, 66*, 53–78.

Moustafa, M., Maksoud, E., & Sadek, S. (2004). Optimal major and minimal maintenance policies for deteriorating systems. *Reliability Engineering & System Safety, 83*(3), 363–368.

Nicolai, R. P., & Dekker, R. (2008). Optimal maintenance of multi-component systems: A review. In *Complex system maintenance handbook*. *Springer series in reliability engineering*. London: Springer.

Ohno, K., Boh, T., Nakade, K., & Tamura, T. (2016). New approximate dynamic programming algorithms for large-scale undiscounted Markov decision processes and their application to optimize a production and distribution system. *European Journal of Operational Research, 249*(1), 22–31.

Ouaret, S., Kenné, J.-P., & Gharbi, A. (2018). Production and replacement policies for a deteriorating manufacturing system under random demand and quality. *European Journal of Operational Research, 264*(2), 623–636.

Papakonstantinou, K. G., & Shinozuka, M. (2016). Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. *Reliability Engineering & System Safety, 130*, 202–213.

Paprocka, I. (2018). The model of maintenance planning and production scheduling for maximising robustness. *International Journal of Production Research*, 1–22.

Petchrompo, S., & Parlikad, A. K. (2019). A review of asset management literature on multi-asset systems. *Reliability Engineering & System Safety, 181*, 181–201.

Shi, Y., Xiang, Y., Xiao, H., & Xing, L. (2021). Joint optimization of budget allocation and maintenance planning of multi-facility transportation infrastructure systems. *European Journal of Operational Research, 288*, 382–393.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation: A review on the statistical data driven approaches. *European Journal of Operational Research, 213*(1), 1–14.

Singh, S., Jaakkola, T., Littman, M. L., & Szepesvári, C. (2000). Convergence results for single-step on-policy reinforcement-learning algorithms. *Machine Learning, 38*(3), 287–308.

Sun, Q., Ye, Z.-S., & Chen, N. (2018). Optimal inspection and replacement policies for multi-unit systems subject to degradation. *IEEE Transactions on Reliability*, *67*(1), 401–413.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*: 1. Cambridge, MA: MIT Press.

Vu, H. C., Do, P., Barros, A., & Bérenguer, C. (2014). Maintenance grouping strategy for multi-component systems with dynamic contexts. *Reliability Engineering & System Safety, 132*, 233–249.

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research, 139*(3), 469–489.

Wang, J., & Zhu, X. (2020). Joint optimization of condition-based maintenance and inventory control for a k-out-of-n: F system of multi-state degrading components. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2020.08.016.

Wang, L., Lu, Z., & Ren, Y. (2020). Joint production control and maintenance policy for a serial system with quality deterioration and stochastic demand. *Reliability Engineering & System Safety, 199*, 106918.

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3-4), 279–292.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. University of Cambridge England Ph.D. thesis..

Xenos, D. P., Kopanos, G. M., Cicciotti, M., & Thornhill, N. F. (2016). Operational optimization of networks of compressors considering condition-based maintenance. *Computers & Chemical Engineering, 84*, 117–131.

Xia, L., Zhao, Q., & Jia, Q.-S. (2008). A structure property of optimal policies for maintenance problems with safety-critical components. *IEEE Transactions on Automation Science and Engineering, 5*(3), 519–531.

Zhang, M., & Revie, M. (2017). Continuous-observation partially observable semi–Markov decision processes for machine maintenance. *IEEE Transactions on Reliability, 66*(1), 202–218.

Zhang, N., & Si, W. (2020). Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. *Reliability Engineering & System Safety, 203*, 107094.

Zhou, Y., Guo, Y., Lin, T. R., & Ma, L. (2018). Maintenance optimisation of a series production system with intermediate buffers using a multi-agent FMDP. *Reliability Engineering & System Safety, 180*, 39–48.

Zhou, Y., Lin, T. R., Sun, Y., & Ma, L. (2016). Maintenance optimisation of a parallel-series system with stochastic and economic dependence under limited maintenance capacity. *Reliability Engineering & System Safety, 155*, 137–146.