Condition-based Selective Maintenance Optimization for Large-scale Systems Consisting of Many Homogeneous Units

Euihwan Lee, Jaehyoung Ju, Eunshin Byon, *Member, IEEE,* and Young Myoung Ko, *Member, IEEE*

***Abstract*—This study considers a maintenance optimization problem for a large-scale system comprising many homogeneous units. Each unit degrades by transitioning on predefined health states, where transition times follow general distributions from one state to another. We consider a selective maintenance scheme where one can repair or replace units under specific health states at each maintenance activity. The system, therefore, does not seem to have a regenerative nature because some units still remain in their degraded states after maintenance. This selective nature challenges the analysis and maintenance decision-making. We first formulate such a process as a fluid system to characterize the degradation pattern at the system level. Based on the fluid model, we devise a condition-based maintenance policy that triggers a repair action when the weighted sum of the fraction of units in specific health states hits a predetermined threshold. Our implementation finds that the fluid system shows periodic behavior as time goes by; the system becomes asymptotically regenerative. Hinging upon this observation, we develop an algorithm to find a threshold triggering a maintenance action that minimizes the long-run average cost. Numerical experiments show that the proposed fluid model accurately approximates the dynamics of system degradation, and the condition-based selective maintenance scheme is cost-effective against the alternative periodic maintenance strategy.**

***Index Terms*—Fluid system, Multi-unit systems optimization, Non-Markovian system, Non-renewal process, Reliability.**

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) under Grant 2021R1A4A1031019 and 2021R1A2C1094699, in part by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008691, HRD Program for Industrial Innovation), and in part by the U.S. National Science Foundation, Division of Information and Intelligent Systems, under Grant IIS-1741166.

Euihwan Lee is with the Hyundai Motor Company., 12, Heolleung-ro, Seocho-gu, Seoul, 06797 Republic of Korea (e-mail: euihwan.lee@lstlab.org)

Jaehyoung Ju is with the Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk, 37673 Korea (e-mail: idgogojjh@postech.ac.kr).

Eunshin Byon is with the Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: ebyon@umich.edu).

Young Myoung Ko is with the Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk, 37673 Korea (corresponding author, e-mail: youngko@postech.ac.kr).

Notation

N	Number of units
M	Number of states
$F_i(t)$	Transition time distribution from state i to $i+1$
$f_i(t)$	Density of $F_i(t)$
$h_i(t)$	Hazard function of $F_i(t)$
$\bar{F}_i(t)$	$1 - F_i(t)$
$\mathbf{X}(t)$	$(X_1(t), \ldots, X_M(t))^T$, status of the system at time t
$X_i(t)$	Fraction of units in state i at time t
γ	Threshold for triggering a maintenance activity
S	Last state where the units are not repaired or replaced under selective maintenance policy
τ_n	n^{th} hitting time of the threshold
α_i	weight of state i in triggering maintenance activity
P	Setup cost per visit
C_i	Repair (or replacement) cost of a unit in state i
L	Revenue loss/unit time that a failed unit (state M) incurs
$\mathbf{X}(t, y)$	$(X_1(t, y), \ldots, X_M(t, y))^T$, two-parameter status of the system at time t with staying in states less than or equal to y amount of time.
$X_i(t, y)$	Fraction of units staying in state i less than or equal to y amount of time at time t
$x_i(t, y)$	Density of $X_i(t, y)$
$r_i(t)$	Actual flow rate from state i to $i+1$

I. Introduction

We study a maintenance optimization problem for a large-scale system consisting of massive homogeneous units. Consider a large-scale wind farm with hundreds or thousands of wind turbines [1]. If the wind turbines have the same specification and same age, their degradation pattern would follow the same process, even though there would be unit-to-unit differences due to the stochastic nature of degradation. Another example that exhibits similar properties is a solar park with a large number of photovoltaic panels. These energy systems contain a large number of small generators, unlike the traditional power plants such as nuclear, thermal and hydroelectric power plants that comprise a few big generators. For the system consisting of many units, system operators may not be able to repair all units at once in a single trip. Instead, they would prefer to choose units in certain degraded states (e.g. alarm or failure states) and conduct repair operations for those selected units only, while doing nothing for the rest of units and allowing to continuously operate until the next maintenance action. Such maintenance type is called selective maintenance [2]. Selective maintenance is a realistic operational type adopted in many real-world applications [3]–[6] where maintenance resources, such as maintenance crew, equipment and cost, are limited.

There are several challenges in devising a cost-effective selective maintenance strategy for these systems. First, in several studies [7], [8], a Markovian degradation process has been assumed. However, in many practical systems, the lifetime distributions of units could be non-Markovian. For example, studies in [9]–[11] employ the Weibull distribution. The degradation processes based on non-exponential distributions exhibit a non-Markovian behavior, which challenges in tracking the system-level dynamics of the system. Here, the system-level dynamics implies the progression of degradation states of entire units in the system, e.g., how many units are in normal, alert, alarm and failure states at any time point and how the numbers evolve over time. The selective maintenance strategy for large-scale systems should be devised, based on the accurate estimation of the system dynamics considering the system as a whole.

Next, for these large-scale systems, we easily face a scalability issue. Many prior studies on maintenance optimization consider a single unit or a small number of units [12]–[14], and their approaches may not be scalable enough to address several hundreds or thousands of units due to the combinatorial nature of problems. To avoid the combinatorial approach, we need to characterize the system-level dynamics mentioned above, however, the selective maintenance strategy adds another layer of challenge. In the simultaneous maintenance policies where all units are repaired together and the system becomes as good as new after each maintenance operation, the system dynamics follow a regenerative process [15]–[17], however, with selective maintenance policy one should not simply assume the regenerative process.

To address the challenges previously mentioned and to overcome the limitations of prior research, this paper offers the following contributions:

- We introduce a new fluid model to capture the dynamics of systems comprised of *hundreds, thousands, or even more units*, implicitly leveraging the law of large numbers. This model describes the transition of units into progressively degraded states. The results from the fluid model give a collective view of the system's health, specifying the number of units in each degradation state and forecasting the system's future trajectory. Consequently, the fluid model can serve as a overarching solution for large-scale systems, complementing the multi-unit (or multi-component) maintenance optimization methods present in existing literature.
- The proposed fluid model is versatile, designed to accommodate any transition distributions with density by directly using the distribution function. This indicates that our model can analyze *non-Markovian system dynamics* without requiring approximations of distribution functions.
- We present a threshold-type *selective condition-based maintenance (CBM)* scheduling problem and outline a solution method. Numerical experiments show that the initial vector of the system state right after each maintenance operation converges to a constant vector, leading the fluid system to become *asymptotically regenerative* at maintenance intervals. This regenerative nature is proven for a three-state continuous-time Markov chain scenario. We believe such asymptotic regenerative nature holds for other general non-Markovian systems. Based on this conjecture, we provide a cost-effective selective maintenance strategy that minimizes the long-run average cost.

The remainder of the paper is organized as follows. Section II reviews previous related studies. Section III describes our problem settings and formulates the optimization problem. Section IV introduces the fluid model and shows the accuracy of the model approximating the system dynamics. Section V proposes an algorithm to solve the optimization problem using the fluid model. Section VI conducts numerical studies, and Section VII wraps up this study and discusses future works.

II. Literature Review

This section provides literature reviews on maintenance optimization problems. The scope of the maintenance optimization is quite broad. As the system we consider in this paper is different from those studied in the previous studies that handled single-unit systems or multi-component systems (e.g., k-out-of-n systems), we omit reviews of those studies. Instead we focus on studies closely relevant to the proposed approach.

Petchrompo and Parlikad [18] provided a comprehensive literature review on asset management of multi-asset systems. They distinguished multi-component systems with multi-asset systems, while both are called multi-unit systems in the literature. Several studies on multi-component systems account for three dependency types: economic, stochastic and structural dependence [19]. Olde Keizer et al. [20] further added resource dependence to these categories for multi-component systems. Salari and Makis [14] proposed two condition-based maintenance policies for a multi-unit system (identical units) considering production and demand rates when determining control limits that trigger maintenance operations. They formulated the problem using a semi-Markov decision process and optimized the long-run average cost. In their study the number of states of the Markov process is $O(N^2)$, where N is the number of units. Therefore, their approach encounters scalability issues as the number of units in the system gets larger.

Castanier et al. [21] used the semi-regenerative process property to calculate the long-run average maintenance cost of two-unit systems. Zhang and Zeng [22] used the semi-regenerative process property to calculate the long-run average maintenance cost of multi-unit systems. However, these studies considered only two- or three-unit systems so they are not applicable to large-scale systems.

Cao et al. [2] carried out a literature review specifically focusing on selective maintenance. They introduced three aspects for selective maintenance decision-making: system, maintenance, and mission profile characteristics. System characteristics include system structure (e.g., series, parallel), lifetime distributions, states, and dependence. Maintenance characteristics represent maintenance degree (perfect, imperfect, and minimal) and resource consumption (negligible, constant, and random). Mission profile characteristics determine mission objectives, planning horizon, available resources, mission types, and working conditions. Maaroufi et al. [5] considered a system conducting multiple missions where each mission requires a minimum level of reliability. Their goal is to select components to be renewed after each mission with minimum cost.

Selective maintenance strategy has been studied in mission-critical applications. Jiang and Liu [3] proposed a selective maintenance strategy for systems that perform multiple missions. They considered three types of maintenance actions: perfect maintenance, minimal repair (MR), and imperfect maintenance (IM). They assumed that the IM action reduces the physical age of the component. They constructed a max-min optimization problem to allocate a maintenance budget so that the minimum of the probabilities of the successful missions can be maximized. Due to the complexity of the problem, they used a simulated annealing-based genetic algorithm to solve the problem. Jiang and Liu [23] considered the uncertainty arising from imperfect observations and formulated a multi-objective optimization problem that maximizes the expectation while at the same time, minimizing the variance of the probability of successful missions. Liu et al. [24] considered a situation where inspections are inaccurate and formulated a finite-horizon Mixed Observability Markov Decision Process (MOMDP) model to address this. Ghorbani et al. [25] assumed that the system undergoes uncertain condition scenarios and adopted a stochastic programming approach to tackle them. Maillart et al. [8] consider a corrective selective maintenance model in a series-parallel system to identify which components to replace in the finitely long periods of time. Ahadi and Sullivan [26] developed an approximate dynamic programming (ADP) algorithm to address the scalability issue of the Markov decision process in the selective maintenance problem for series-parallel systems. Xu et al. [27] proposed a hybrid algorithm, combining the deep Q-network and the discrete differential evolution algorithm, to overcome the same scalability challenge of MDP.

Similarly, Liu et al. [4] developed a selective maintenance model for multi-state systems to maximize the mission-completion probability. They considered the randomness of break duration which imposes an computational burden due to multiple integration. To alleviate the computational complexity, they used the saddle point approximation and used the ant colony algorithm to obtain the solution. To validate their approach they used a coal transportation system as a numerical experiment. Liu et al. [28] addressed a finite-horizon selective maintenance problem for multi-state systems considering imperfect maintenance. They formulated the problem using a discrete-time finite-horizon Markov decision process. For a solution approach, they proposed a deep reinforcement learning method customized for their problem. Hesabi et al. [29] employed both deep learning and mathematical programming to minimize the total selective maintenance cost under intermission break time constraints.

Yang et al. [6] suggested a heuristic sequential game approach for a fleet-level selective maintenance problem with a phased mission scheme. They considered a fleet consisting of many pieces of equipment that is further composed of subsystems. The fleet goes through a phased mission with short breaks between phases. Due to difficulty in developing a unified game framework for the selective maintenance of such systems, they proposed a heuristic game approach with state backtracking. They conducted an aircraft fleet case study to validated their approach.

Khatab et al. [30] formulated a joint non-linear programming optimization problem for the integration of selective maintenance and repair channel assignment when multiple repair channels are available. Yin et al. [31] considered heterogeneous repair channels and addressed the selective maintenance problem using the ant colony metaheuristic algorithm. Ma et al. [32] also took a metaheuristic approach, specifically the cooperative co-evolutionary genetic algorithm, to solve a complicated optimization problem. Chaabane et al. [33] also integrated selective maintenance and repair-person assignment for multiple missions into an unified framework. Due to the combinatorial nature of the problem, they developed a heuristic algorithm based on the genetic algorithm.

Most studies, including aforementioned ones, limited their analysis to specific lifetime distributions such as exponential, Gamma, and Weibull distributions and handle a small number of units. Ko and Byon [10] proposed a method to keep track of the dynamics of the system's degradataion state when each unit degrades following a general lifetime distribution. They approximated general lifetime distributions with phase-type distributions using the denseness of phase-type distributions. Ko and Byon [15] further devised a condition-based optimization method, building on the results in [10]. Their work was one of the first attempts to obtaining an optimal maintenance scheduling for large-scale systems. However, there are a few limitations that hinder their approach applicable in practice. First, the accuracy of the phase-type approximation depends on the number of phases: in general, the more phases the more accurate. However, using more than 10 phases lead to very complicated equations which are not traceable. Most importantly, their maintenance policy is limited to simultaneous repair for all units where all units become as good as new after a maintenance activity. Such policy would not be feasible for large-size systems when repairing each unit requires substantial resources.

III. Problem description

We consider a system consisting of N homogeneous units and each unit deteriorates independently following the same degradation process. We assume each unit's degradation status can be classified into M states, e.g., normal, alert, alarm and failure states. We allow the transition time from state i to the next state $i+1$ to obey any arbitrary distribution, e.g., Exponential, Weibull, Lognormal, Gamma distribution, etc.,

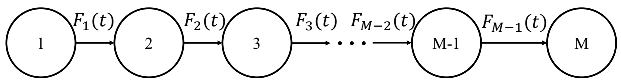


Fig. 1. Degradation process of each unit in [15]

with a known cumulative distribution function (cdf) denoted by $F_i(\cdot)$. Since transitions are unidirectional, states with higher numbers represent more deteriorated health conditions. When a unit reaches the last state, i.e. state M, can be regarded as failure. Figure 1 describes the degradation process of each unit.

One of the effective approaches to characterize the system's health state is to investigate the number of units, or proportion of units, in each degradation state and its progression [10]. Specifically, let $\mathbf{X}(t) = (X_1(t), \cdots, X_M(t))^T$ denote the status of the system at time t where $X_i(t)$ is the fraction of units in state i at time t, $\sum_{i=1}^{M} X_i(t) = 1,\ \forall t \geq 0$. Initially at time $t = 0$, the system starts with a brand new status, i.e., $\mathbf{X}(0) = (1, 0, \cdots, 0)$. Since we allow the transition times between states to follow any distributions, the stochastic degradation process of each unit is non-Markovian and so is the system.

The units in the system degrade over time by transitioning from one state to the next. When a unit reaches state the last (failure) state, M, it incurs a revenue loss L per unit time. The system operator who continuously monitors the system decides to visit the system site and perform maintenance operations based on the fraction of units in state $S+1$ through M, e.g., alarm and failure states. When a maintenance activity is triggered, the maintenance engineers only repair (or replace) units in state $S+1$ through M and leave other units in state 1 through S untouched. Once a maintenance operation is completed, the units in state $S+1$ through M become *as good as new* and the rest of the units stay in their states. The duration for completing a maintenance activity is assumed to be negligible, compared to the maintenance interval [15]. Maintenance activities incur a setup cost P per visit and a repair (or replacement) cost C_i for each unit in state i.

As we consider the selective maintenance to repair degraded units only, an effective strategy would be to trigger maintenance when the proportion of units $S+1$ through M becomes above a given threshold. In particular, considering the significance of each state, we use a weighted fraction of degraded units as a measure for how strongly it is desired to execute maintenance and trigger maintenance when it hits (or exceeds) a threshold, denoted by γ, for $0 < \gamma \leq 1$. That is, maintenance is prompted when the system state satisfies $\sum_{s=S+1}^{M} \alpha_s X_s(t) \geq \gamma$. Here, α_i is the weight for state i, satisfying $\alpha_{S+1} \leq \alpha_{S+2} \leq \cdots \leq \alpha_M$.

We can consider S and α_i as hyperparameters that are dependent on the context. They are influenced by factors beyond just cost. The value of S can be chosen based on the imminence of states and the repair capacity during a single trip. Meanwhile, the values of α_i signify the significance of each state. One might choose these values in proportion to the repair cost of each state, factoring in the availability and delivery time of parts.

Nevertheless, it is possible to optimize both S and α_i in terms of cost. A detailed discussion on this optimization can be found in Section VI-E.

Let τ_n denote the n^{th} hitting time of the threshold as follows:

$$\tau_n = \inf\left\{t > \tau_{n-1} : \sum_{s=S+1}^{M} \alpha_s X_s(t) \geq \gamma\right\},$$

where $\tau_0 = 0$.

The objective of the selective maintenance optimization problem in this study is to find a threshold that minimizes the long-run average maintenance cost as

$$\underset{\gamma}{\text{minimize}} \quad f_0(\gamma) = \lim_{t\to\infty} \frac{1}{t} \left(n_t P + \sum_{j=1}^{n_t} \sum_{i=S+1}^{M} C_i N X_i(\tau_j) + \sum_{j=1}^{n_t} \int_{\tau_{j-1}}^{\tau_j} L N X_M(t) dt \right), \quad (1)$$

where $n_t = \sup\{n : \tau_n \leq t\}$.

The optimization problem in (1) is not directly solvable because we do not have closed form expressions of $\mathbf{X}(t)$ and τ_n. Furthermore, optimization on an infinite horizon is not trivial unless the problem has a special structure. To tackle this issue, we take an approach that treats the transitions of units as *flows of fluid* when the number of units in the system is sufficiently large [10]. This perspective allows us to keep track of the time-varying fraction of units in each state.

Studies on queueing theory have developed useful mathematical tools to modeling and analysis of large-scale systems. Especially, exploring asymptotic behavior via the law of large numbers (fluid model) and the central limit theorem (diffusion model) provides tractable ways to analyze queueing systems that show non-Markovian and time-varying behavior. Details of limit theorems on queueing systems are summarized in [34], [35]. In particular, the fluid model is often used to approximate the mean behavior of the system. Whitt [36] developed fluid models for non-Markovian multiserver queues with abandonment introducing two-parameter deterministic limit processes. We derive a fluid model with two-parameter processes for describing the degradation process with non-Markovian transition times between degradation states in Section IV. Then Section V proposes a solution method for solving the problem in (1).

IV. System State Characterization with Fluid Model

We first explain the basic fluid model that describes the degradation process without maintenance activities in Section IV-A. Then Section IV-B explains the fluid model with selective maintenance.

A. Basic model

To describe the degradation process of a non-Markovian system, we should characterize how the system's health state will evolve, as well as the current health state. Therefore, it is not sufficient to keep the fraction of units (or the number of units) in each state alone. We formulate the system-level degradation process using a two-parameter fluid process [36], [37]. Define $\mathbf{X}(t,y) = (X_1(t,y), \dots, X_M(t,y))^T$ be a vector, where $X_i(t,y)$ denotes the fluid fraction of units in state i that has been in state i less than, or equal to, y amount of time. Then, we have

$$\mathbf{X}(t) = \lim_{y\to\infty} \mathbf{X}(t,y).$$

We assume that $X_i(t,y)$ has a smooth density $x_i(t,y)$ with respect to y such that

$$X_i(t,y) = \int_0^y x_i(t,u)du.$$

Suppose that the transition distribution $F_i(\cdot)$ has a pdf $f_i(\cdot)$ and a hazard rate $h_i(\cdot)$, and we define $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$. The actual rate of fluid from state i to $i+1$, denoted by $r_i(t)$, can be obtained by

$$r_i(t) = \int_0^t x_i(t,y)h_i(y)dy.$$

Then we have the following fluid model for state 1 of the system that starts with brand new units at time 0:

$$X_1(t) = X_1(0)\bar{F}_1(t), \tag{2}$$

$$r_1(t) = -\frac{d}{dt}X_1(t) = X_1(0)f_1(t). \tag{3}$$

Here, state 1 does not have in-flow from other states, so $X_1(t)$ can be easily obtained by multiplying $X_1(0)$ and $\bar{F}_1(t)$ as in (2). The rate of out-flow from state 1 in (3) is nothing but the negative derivative of $X_1(t)$ with respect to t.

For $2 \le i \le M-1$, state i has in-flow from state $i-1$. Therefore, state i has a different form of fluid model from state 1:

$$x_i(t,y) = \bar{F}_i(y)r_{i-1}(t-y), \tag{4}$$

$$r_i(t) = \int_0^t x_i(t,y)h_i(y)dy, \tag{5}$$

$$X_i(t) = \int_0^\infty \bar{F}_i(y)r_{i-1}(t-y)dy = \int_0^t \bar{F}_i(y)r_{i-1}(t-y)dy. \tag{6}$$

Equation (4) represents the density of $X_i(t,y)$, obtained by multiplying the input rate from the previous state at $(t-y)$, i.e., $r_{i-1}(t-y)$, and the probability (fraction of rates in fluid model) of staying more than y amount of time. We can obtain the rate of out-flow from state i by integrating the multiplication of the density ($x_i(t,y)$) and the failure rate ($h_i(y)$) over y (equation (5)). The fraction of units in state i is then calculated by integrating the density over y (equation (4)).

State M is the final state, so $X_M(t)$ can be obtained by subtracting the sum of $X_i(t)$, $i \in \{1,\dots,M-1\}$, from 1:

$$X_M(t) = 1 - \sum_{k=1}^{M-1} X_k(t). \tag{7}$$

The fluid model described here is used to capture the mean behavior of a stochastic system. We can obtain the fluid model by numerically solving the integral equations in (2)-(7). To confirm the accuracy of the fluid model, we compare the fluid model with simulation. We consider a four-state system with Weibull transition times: the transition distribution F_i ($i = 1,2,3$) with shape parameter 3.05 for all i and scale parameter $300, 200$ and 168 for $i = 1, 2$ and 3, respectively. The number of units is 100. The simulation mean is obtained by averaging the result of 1,000 independent runs with the time horizon $[0, 1000]$. Figure 2 illustrates the number of units in each state over time. As seen in the figure, the fluid model shows great accuracy, almost coinciding with the simulation result.

It is worthwhile to mention that the proposed modeling approach provides several advantages over the existing fluid model, e.g., in [10]. Ko and Byon [10] approximate a general lifetime distribution using phase-type distributions. On the contrary, we directly utilize lifetime distribution functions without the intermediate step of phase-type approximations. Doing so considerably simplifies the mathematical representation with a much smaller number of equations, hence, it is easier to solve. Moreover, it does not rely on an additional step required in the approach in [10]—finding a phase-type distribution approximating the state transition distribution, which further simplifies the procedure. Most importantly, the model in [10] is designed for the simultaneous maintenance which renews every unit at once upon each maintenance activity. Conversely, the new fluid model presented in this section enables us to devise the fluid model under the selective maintenance policy, which is more general than that in [10].

While the new fluid model offers several advantages over the one presented in [10], it does introduce some computational challenges. Specifically, the fluid model accommodates non-Markovian transition times, necessitating the tracking of all previous state trajectories. This leads to the requirement of evaluating multiple integrals for a single time point. The specific integral equations for $X_n(t)$ are as follows: For state 1,

$$X_1(t) = X_1(0)\bar{F}_1(t)$$
$$r_1(t) = X_1(0)f_1(t).$$

For state 2,

$$x_2(t,y) = \bar{F}_2(y)X_1(0)f_1(t-y)$$
$$r_2(t) = X_1(0)\int_0^t f_1(t-y)f_2(y)dy$$
$$X_2(t) = \int_0^t \bar{F}_2(y)X_1(0)f_1(t-y)dy.$$

For state 3,

$$x_3(t,y) = \bar{F}_3(y)X_1(0)\int_0^{t-y} f_1(t-y-y_1)f_2(y_1)dy_1$$
$$r_3(t) = X_1(0)\int_0^t\int_0^{t-y} f_1(t-y-y_1)f_2(y_1)dy_1 f_3(y)dy$$
$$X_3(t) = \int_0^t \bar{F}_3(y)X_1(0)\int_0^{t-y} f_1(t-y-y_1)f_2(y_1)dy_1 dy.$$

For state 4,

$$x_4(t,y) = \bar{F}_4(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_2} f_1(t-y-y_1-y_2)f_2(y_1) dy_1 f_3(y_2)dy_2$$

$$r_4(t) = X_1(0)\int_0^t\int_0^{t-y}\int_0^{t-y-y_2} f_1(t-y-y_1-y_2)f_2(y_1) dy_1 f_3(y_2)dy_2 f_4(y)dy$$

$$X_4(t) = \int_0^t \bar{F}_4(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_2} f_1(t-y-y_1-y_2) f_2(y_1)dy_1 f_3(y_2)dy_2 dy.$$

For state 5,

$$x_5(t,y) = \bar{F}_5(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_3}\int_0^{t-y-y_2-y_3} f_1(t-y-y_1-y_2-y_3)f_2(y_1)dy_1 f_3(y_2)dy_2 f_4(y_3)dy_3$$

$$r_5(t) = X_1(0)\int_0^t\int_0^{t-y}\int_0^{t-y-y_3}\int_0^{t-y-y_2-y_3} f_1(t-y-y_1-y_2-y_3)f_2(y_1)dy_1 f_3(y_2)dy_2 f_4(y_3)dy_3 f_5(y)dy$$

$$X_5(t,y) = \int_0^t \bar{F}_5(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_3}\int_0^{t-y-y_2-y_3} f_1(t-y-y_1-y_2-y_3)f_2(y_1)dy_1 f_3(y_2)dy_2 f_4(y_3)dy_3 dy.$$

For state n,

$$x_n(t,y) = \bar{F}_n(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_{n-2}}\cdots\int_0^{t-y-\sum_{i=2}^{n-2} y_i} f_1\left(t-y-\sum_{i=1}^{n-2} y_i\right) f_2(y_1)dy_1\cdots f_{n-1}(y_{n-2})dy_{n-2}$$

$$r_n(t) = X_1(0)\int_0^t\int_0^{t-y}\int_0^{t-y-y_{n-2}}\cdots\int_0^{t-y-\sum_{i=2}^{n-2} y_i} f_1\left(t-y-\sum_{i=1}^{n-2} y_i\right) f_2(y_1)dy_1\cdots f_{n-1}(y_{n-2})dy_{n-2} f_n(y)dy$$

$$X_n(t) = \int_0^t \bar{F}_n(y)X_1(0)\int_0^{t-y}\int_0^{t-y-y_{n-2}}\cdots\int_0^{t-y-\sum_{i=2}^{n-2} y_i} f_1\left(t-y-\sum_{i=1}^{n-2} y_i\right) f_2(y_1)dy_1 \cdots f_{n-1}(y_{n-2})dy_{n-2}dy.$$

For a fixed t_*, the number of integrals required to calculate $X_n(t_*)$ is $n-1$. Fortunately, many research studies on maintenance optimization problems, including our previous work, utilize two to four states. Two states are used in [38], three states in [39]–[41], and four states in [10], [13], [15], [42]–[44]. For $M=4$, double integrals need to be solved as $X_M(t) = \sum_{i=1}^{M-1} X_i(t)$. These double integrals can be accurately evaluated using existing integration algorithms. We note that the model presented in [10] struggles due to the large number of differential equations. Each phase used for approximating the distribution necessitates solving one differential equation.

The next section will present the fluid model for selective maintenance in detail.

B. Fluid model for selective maintenance

Since the maintenance activities change the state of the system, we should adjust the fluid model discussed in the previous section. The fluid model reflecting maintenance activities is as follows:

$$X_1(t) = X_1(0)\bar{F}_1(t) + \sum_{j=1}^{n_t}\sum_{i=S+1}^{M} X_i(\tau_j)\bar{F}_1(t-\tau_j), \quad (8)$$

$$r_1(t) = -\frac{d}{dt}X_1(t) = X_1(0)f_1(t) + \sum_{j=1}^{n_t}\sum_{i=S+1}^{M} X_i(\tau_j)f_1(t-\tau_j). \quad (9)$$

When the n^{th} maintenance activity is completed, the units from state $S+1$ through state M are repaired (replaced) and become as good as new. The second term of the right-hand side of (8) represents the newly added fraction of units whenever the maintenance activity is completed. After being added, the fraction of units decreases according to the distribution $F_1(\cdot)$. The rate, $r_1(t)$ in (9), at which $X_1(t)$ is decreasing, is simply a negative time-derivative of $X_1(t)$.

For $2 \le i \le S$, the fluid model is the same as that in (2)-(6), i.e.,

$$x_i(t,y) = \bar{F}_i(y)r_{i-1}(t-y), \quad (10)$$

$$r_i(t) = \int_0^t x_i(t,y)h_i(y)dy, \quad (11)$$

$$X_i(t) = \int_0^t x_i(t,y)dy. \quad (12)$$

For $S+1 \le i \le M-1$, the units are *renewed* after the maintenance activities. The fluid model reflecting the renewal is given by

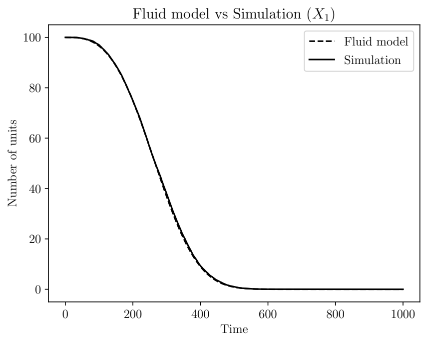
$$x_i(t,y) = \bar{F}_i(y)r_{i-1}(t-y),$$

$$r_i(t) = \int_0^{t-\tau_{n_t}} x_i(t,y)h_i(y)dy, \quad (13)$$

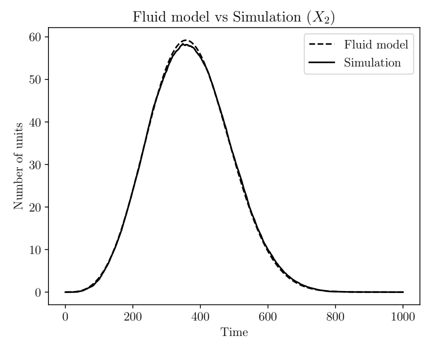
$$X_i(t) = \int_0^{t-\tau_{n_t}} x_i(t,y)dy, \quad (14)$$

where the rate of flow, $r_i(\cdot)$, and the fraction of units, $X_i(\cdot)$, are reset after each maintenance activity, so the second parameter, y, which represents the length of stay in state i, should be restarted as in (13)-(14).

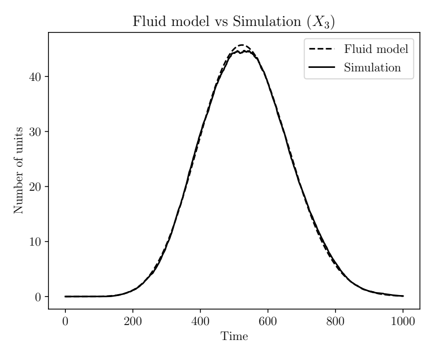
To illustrate, let us consider a four-state fluid system where the transition time follows the Weibull distribution. Figure 3 shows how the fraction of units in each state evolves over time when the threshold γ is 0.3 with weights for states 3 and 4 given by 0.5 and 1.0, respectively. We observe spikes when $0.5X_3(t) + X_4(t)$ hits 0.3, which means the units in states 3 and 4 are repaired (or replaced) immediately and become brand-new. Here, we notice an interesting phenomenon that the dynamics of the fluid system presents a repeated pattern after



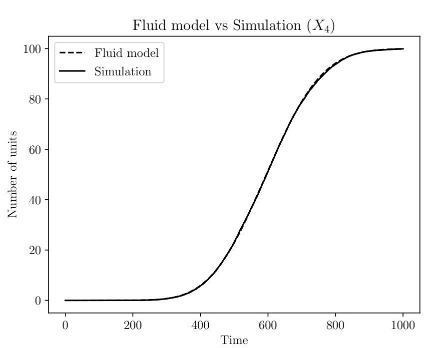
(a) $X_1(t)$



(b) $X_2(t)$



(c) $X_3(t)$



(d) $X_4(t)$

Fig. 2. Comparison between fluid model outputs and simulation outputs

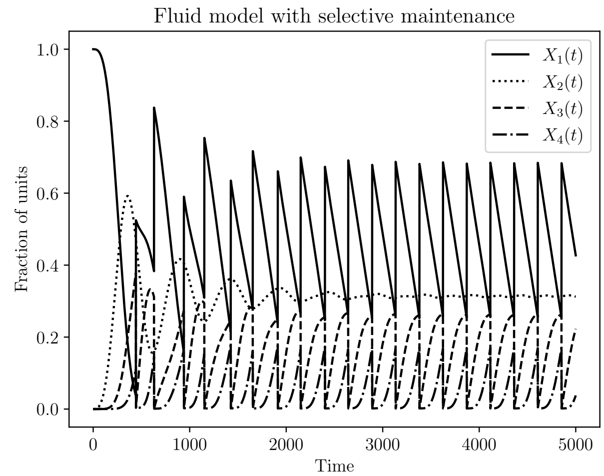


Fig. 3. Example of fluid model results for a four-state system with selective maintenance for $\gamma = 0.3$, $\alpha_3 = 0.5, \alpha_4 = 1.0$ and the Weibull transition distribution with shape parameter 3.05 for all i and scale parameter $300, 200$ and 168 for $i = 1, 2$ and 3, respectively.

some initial period passes by. It appears that each maintenance cycle, given any fixed threshold γ, tends to become periodic, providing an important implication that we can optimize the average cost for an infinite-horizon selective maintenance problem in (1) by optimizing the average cost of one cycle. The next section will explain how we develop a method for solving the optimization problem in (1) based on this observation.

V. Solution method for selective maintenance optimization

With the fluid model discussed in Section IV, this section solves the optimization problem in (1) to find an optimal threshold, γ^*. Since the maintenance activities are selective, only a subset of units are repaired or replaced. Therefore, the process is not regenerated and solving such a problem with a non-Markovian setting becomes non-trivial. However, the cyclic pattern we observe in Figure 3 after some warm-up (or burn-in) periods provides a clue to solve the problem. In various numerical settings with a wide range of different settings of transition distributions, weights and threshold values, we observe similar patterns that the fraction of units in each state when the threshold is hit converges, as maintenance activities are repeated over time. This observation leads to the following conjecture.

Conjecture V.1 (Convergence of states)**.** *Given a threshold γ, let $\mathbf{X}(\tau_n+)$ be the state of the fluid model right after the threshold is touched and the n^{th} maintenance activity is done. Then, $\mathbf{X}(\tau_n+)$ converges to a constant vector $\mathbf{X}$ as $n \to \infty$. Furthermore, $\tau_n - \tau_{n-1}$ also converges to a constant τ as $n \to \infty$.*

We first provide a proof for a special case—a three-state continuous time Markov chain. The proof is, nevertheless, not trivial at all. The detailed proof of Theorems V.2 and

V.3 below is available in Appendix. Let us consider a three-state degradation process, i.e., $M = 3$. State 1 represents a brand new status, and state 3 indicates failure. Transition times from state i to state $i+1$ follow the exponential distribution with parameter λ_i. We assume that $\lambda_1 \neq \lambda_2$. Due to the memoryless property of the exponential distribution, we do not have integral equations of convolution forms and instead have a system of ordinary differential equations. Then, we obtain a closed form solution when we do not consider a maintenance operation as follows:

$$\begin{aligned}
X_1(t) &= e^{-\lambda_1 t} X_1(0) \\
X_2(t) &= \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) e^{-\lambda_1 t} \\
&\quad + \left(X_2(0) - \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) \right) e^{-\lambda_2 t} \\
X_3(t) &= 1 - X_1(t) - X_2(t).
\end{aligned}$$

Suppose we apply a threshold $\gamma \in (0, 1)$ to state 3 for triggering a maintenance operation. When a maintenance operation is triggered and completed, the starting point of a new maintenance period changes by adding the value of state 3 to that of state 1 and setting the value of state 3 to be zero. We can consider that this operation resets the clock with a new initial value.

We consider a set $\mathbb{I} = \{(x_1, x_2, x_3) \in \mathbb{R}^3_+ + : x_1 + x_2 + x_3 = 1\}$. Let $\mathbf{I}^{(k)} \in \mathbb{I}$ denote the initial point immediately after the k^{th} maintenance; $\mathbf{I}^{(k)} = (X_1(\tau_k+), X_2(\tau_k+), 0)$. The initial point at the beginning of the horizon (i.e. $t = 0$) is given by $\mathbf{I}^{(0)} = (1, 0, 0)$. Then, we have the following theorems to show the periodic behavior of the system.

Theorem V.2 (The existence and uniqueness of a periodic solution)**.** *Consider a three-state Markovian system where a maintenance activity is triggered when $X_3(t) \geq \gamma$ for $\gamma > 0$. There exists a unique $\mathbf{I}^* = (x_1^*, x_2^*, 0) \in \mathbb{I}$ satisfying $\mathbf{I}^{(k)} = \mathbf{I}^*$ for all $k \in \mathbb{Z}+$.*

Proof. See Appendix □

Theorem V.3 (Global attractiveness)**.** *Consider a three-state Markovian system where a maintenance activity is triggered when $X_3(t) \geq \gamma$ for $\gamma > 0$. Then, $\mathbf{I}^{(k)} \longrightarrow \mathbf{I}^*$ as $k \to \infty$. Furthermore, the result holds for any initial point $\mathbf{I}^{(0)} = (x_1, x_2, 0) \in \mathbb{I}$.*

Proof. See Appendix □

The results in Theorems V.2 and V.3 imply that the proportion of units in each degradation state after each maintenance action remains the same under the selective maintenance policy where only units in state 3 are repaired. While these results are obtained for a specific process, we believe the results can be extended to other general non-Markovian processes, based on our observations in Figure 3 and in other settings. We note that proving this regenerative nature in a general setting requires in-depth analysis, e.g., differential (or integral) equations with discontinuity such as the theory of the impulsive differential equations as discussed in [45]. Such extensive theoretical analysis in general settings of non-Markovian systems with an arbitrary number of states is beyond the scope of this study, as we instead focus on solving an engineering problem. Thus, we leave the proof of Conjecture V.1 on the regenerative process in general settings for future work.

Suppose Conjecture V.1 holds for general non-Markovian cases. Then the maintenance process becomes *asymptotically* a regenerative process, which allows us to convert the optimization problem in (1) into the following problem:

$$\begin{aligned}
\underset{\gamma}{\text{minimize}} \quad f_0(\gamma) = \frac{1}{\tau} \Bigg(P + \sum_{i=S+1}^{M} C_i N X_i(\tau) \\
+ \int_0^{\tau} L N X_M(t) dt \Bigg), \qquad (15)
\end{aligned}$$

for $\mathbf{X}(0) = \mathbf{X}$ where $\mathbf{X}$ denotes the asymptotic initial state immediately after each maintenance activity.

Now, we have a simpler form of the optimization problem. One good feature is that we have a bounded single variable $\gamma \in [0, 1]$; full enumeration by discretizing the interval is possible. A plausible way to discretize the space is to use the number of units. For example, if there are N units in the system, we can use $1/N$ to increment the γ value so that we can directly get the number of units corresponding to the threshold value γ. However, the challenge comes from the fact that the asymptotic initial state $\mathbf{X}$ depends on the decision variable γ.

To address the challenge, we take a two-step approach. First, we obtain the asymptotic initial value $\mathbf{X}$ using the so-called *warm-up* period. The warm-up period is required to acquire the periodicity of the system. The initial values of $\mathbf{X}(t)$ just after each maintenance activity, i.e., $\mathbf{X}(\tau_n+)$, are unknown except at time 0; $\mathbf{X}(\tau_n+)$ is indeed determined by the choice of threshold γ. Hence, given γ, we numerically calculate $\mathbf{X}(t)$ over time. When it is τ_n, we set a new initial value, $\mathbf{X}(\tau_n+)$. We keep calculating $\mathbf{X}(\tau_n+)$ until $||\mathbf{X}(\tau_n+) - \mathbf{X}(\tau_{n-1}+)|| \leq \varepsilon$ for a predefined value $\varepsilon > 0$ where $|| \cdot ||$ is the supremum norm. Then, we set $\mathbf{X} = \mathbf{X}(\tau_n+)$ and $\tau = \tau_n - \tau_{n-1}$. Second, once $\mathbf{X}$ and τ are set, we can calculate the objective value in (15) using numerical integration. Algorithm 1 describes the solution procedure.

VI. NUMERICAL EXPERIMENTS

This section conducts numerical experiments to demonstrate the accuracy of our approach. We also perform a sensitivity analysis in a range of different parameters. Finally, we benchmark the proposed threshold-based condition-based maintenance against the alternative periodic maintenance strategy.

A. Experiment setting

We first set the number of degradation states. A too large M would be unnecessary and require excessive computations, while a too small M cannot classify each unit's health condition correctly. Following the literature [10], [15], [39], [42], [43], [46], we implement four degradation states, e.g., normal, alert, alarm and failure states, and only repair (or replace) units in alarm or failure state (i.e. $M = 4,\ S = 2$). For the transition distributions, we use the data describing bearing degradation

Algorithm 1: Finding optimal threshold and long-run average cost

```
Data: N, S, M, ε, P, L, α_i, C_i, i = S+1, ..., M
Result: γ*, OptimalCost
γ ← 1/N
OptimalCost = ∞
while γ < 1.0 do
    while True do
        Solve integral equations (8)-(14) until
          Σ_{s=S+1}^{M} α_i X_i(t) hits γ
        τ_n ← t
        τ ← τ_n − τ_{n−1}
        if ||X(τ_{n−1}) − X(τ_n)|| < ε then
            break
        else
            X_1(τ_n+) ← X_1(τ_n) + Σ_{i=S+1}^{M} X_i(τ_n)
            X_i(τ_n+) ← 0.0 for i ∈ {S+1, ..., M}
            n ← n + 1
    Calculate Cost
    if OptimalCost < Cost then
        OptimalCost ← Cost
        γ* = γ
    γ ← γ + 1/N
```

in [47] which is also adopted in [10], [15]. Tian and Liao [47] employ a Weibull distribution with a shape parameter of 3.046 and a scale parameter of 667.6 to characterize the lifetime distribution of a single unit (a two-state model) in their bearing degradation case study. We use their data for illustrating maintenance decision-making. Unlike their model, ours is a four-state version comprising three transitions. As such, we divide the scale parameter 667.6 (rounded to approximately 668) into three segments: 300, 200, and 168 for each transition, retaining the same shape parameter (rounded to approximately 3.05). This implies we use the Weibull distribution as the transition distribution F_i $(i = 1, 2, 3)$ with a consistent shape parameter of 3.05 for all i. The scale parameters for $i = 1, 2,$ and 3 are $300, 200,$ and 168, respectively.

The long-run average costs from the proposed approach are obtained using Algorithm 1 with $\varepsilon = 0.01$. Costs from simulation are calculated by averaging the costs from 100 independent runs with time horizon of $60,000$. Code is written in Julia 1.7.1 and run under a regular PC environment.

B. Implementation results

We conduct experiments to investigate the approximation quality of the proposed approach. We borrow the cost structure for repairing (or replacing) units in states 3 (alert) and 4 (failure), denoted by C_3 and C_4, respectively, from [15]—$C_3 = 1,800, C_4 = 16,300$—. Note that failure state incurs a much larger repairing cost than alarm state does. Further, we consider the revenue loss, $L = 10$ per unit time, while the unit is in state 4 (the failure state). We implement the approach under different settings. Specifically, we consider six different weight vectors ($\alpha_3 = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$) Without loss of generality, we set $\alpha_4 = 1.0$ because the relative weight (α_3/α_4) determines the optimal cost. We further consider six different setup costs ($P = 3600, 9000, 18000, 36000, 54000, 90000$), in line with the setup cost ratio detailed in [15]. This ratio signifies the relationship of the setup cost to C_3. In our setup, when $C_3 = 1,800$: $3,600$ corresponds to a setup cost ratio of 2, $9,000$ to 5, $18,000$ to 10, $36,000$ to 20, $54,000$ to 30, and $90,000$ to 50.

TABLE I
COMPARISON OF LONG-RUN AVERAGE COSTS BETWEEN THE PROPOSED ANALYTICAL APPROACH AND SIMULATION OVER DIFFERENT SETTINGS (NOTE: $N = 100, \alpha_4 = 1.0$)

α_3	Setup cost	Proposed	Simulation[a]	Gap[b]
0.0	3,600	559.67	559.06 (2.50)	0.1%
	9,000	615.92	617.37 (3.14)	0.2%
	18,000	709.67	716.13 (4.18)	0.9%
	36,000	897.17	911.68 (6.78)	1.6%
	54,000	1,065.52	1,067.68 (6.29)	0.2%
	90,000	1,345.56	1,357.40 (8.05)	0.8%
0.2	3,600	521.92	479.79 (3.48)	8.8%
	9,000	595.85	575.61 (4.50)	3.5%
	18,000	703.00	677.58 (4.54)	3.8%
	36,000	889.29	890.57 (6.07)	0.1%
	54,000	1,044.47	1,043.84 (7.20)	0.1%
	90,000	1,331.72	1,337.85 (8.05)	0.5%
0.4	3,600	506.95	502.91 (3.79)	0.8%
	9,000	570.48	567.54 (4.45)	0.5%
	18,000	676.36	673.56 (4.99)	0.4%
	36,000	888.12	885.10 (5.20)	0.3%
	54,000	1,057.09	1,045.10 (7.83)	1.1%
	90,000	1,343.88	1,340.03 (7.48)	0.3%
0.6	3,600	490.23	484.12 (1.88)	1.3%
	9,000	578.33	561.66 (4.52)	3.0%
	18,000	682.55	675.33 (4.97)	1.1%
	36,000	878.21	875.01 (5.74)	0.4%
	54,000	1,055.42	1,051.41 (7.14)	0.4%
	90,000	1,332.34	1,330.40 (8.46)	0.2%
0.8	3,600	508.81	483.48 (4.03)	5.2%
	9,000	581.78	557.30 (3.74)	4.4%
	18,000	693.98	674.05 (4.77)	3.0%
	36,000	882.77	870.68 (6.25)	1.4%
	54,000	1,048.05	1,038.92 (6.23)	0.9%
	90,000	1,340.19	1,333.08 (8.07)	0.5%
1.0	3,600	483.43	483.34 (4.01)	0.0%
	9,000	553.56	557.77 (3.90)	0.8%
	18,000	670.45	683.05 (3.97)	1.8%
	36,000	880.53	868.96 (6.15)	1.3%
	54,000	1,045.87	1,036.73 (6.82)	0.9%
	90,000	1,335.36	1,330.37 (8.24)	0.4%

[a] : numbers in parentheses are the standard deviation.
[b] : Gaps are calculated using the formula, $\frac{|\text{Simulation}-\text{Proposed}|\times 100}{\text{Simulation}}$.

To evaluate the approximation quality of the proposed analytical approach, we compare the long-run average costs between the proposed approach (third column) and simulation (fourth column) when $N = 100$ in Table I. We observe that the proposed approach achieves good accuracy. Though not shown in Table I due to space restrictions, we observe similar results when the number of units increases, i.e., $N = 1000, 10000$. We scaled the setup costs proportional to the number of units, and the long-run average costs are proportional to the number of units. For example, increasing the number of units from 100 to 1000 (10 times) incurs ten times higher maintenance costs. The gaps between the proposed method and simulation lie between 0.1% and 5.0%, which is similar to the case of $N = 100$ and implies $N = 100$ is large enough for achieving good accuracy.

The large setup cost tends to cause less frequent maintenance operations, which in turn increases the threshold value. Figure 4 illustrates the optimal thresholds triggering maintenance operations with different setup costs. Clearly, larger setup costs delay the maintenance activity; i.e., γ value increases as P gets larger. We also observe that γ increases in most cases, as α_3 increases. Larger α_3 implies that the weight for repairing units in the alarm state gets similar to that in the failure state. Thus, the selective maintenance policy waits until more units in the alarm state transit to the failure state.

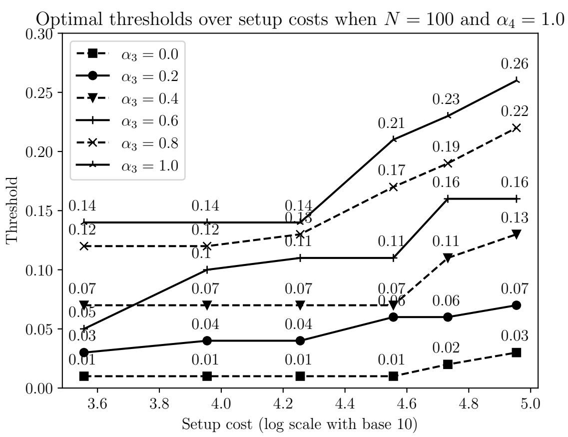


Fig. 4. Optimal thresholds with different setup costs and weights ($N = 100$)

C. Sensitivity analysis

We conduct additional experiments over a range of cost parameters to examine their impacts on the maintenance policy. With $N = 100$, $C_3 = 1,800$, $C_4 = 16,300$, $L = 10$ and $P = 70,000$ as a baseline, we vary the repair costs for units in alarm (C_3) and failure (C_4) states and revenue loss (L). In general, failure cost C_4 is much larger than maintenance cost C_3. We fix $\alpha_3 = 0.6, \alpha_4 = 1.0$ for experiments.

Table II reports the results of the experiments. First, we observe that when C_3 are 5,400 and 7,200, γ^* is 1.0. In those cases, the maintenance operations are triggered after all units fail. It is due to high repair costs of units in state 3. Similar to what we observed with high α_3 in Fig. 4, larger C_3 (with fixed C_4 at 16,300) delays maintenance by increasing γ^*, suggesting the system operator to wait until more units fail. This can be explained by the smaller gap between C_3 and C_4; as the repair cost C_3 for an unit in the alarm state gets closer to the repair cost C_4 for a failed unit, system operators permit more units to transit to the fail state. On the other hand, larger C_4 (with fixed C_3 at 1,800) expedites the maintenance activity to reduce the number of units in the failure state; see decreasing γ^* as C_4 increases. Lastly, larger revenue loss L tend to shorten the maintenance interval via smaller γ^*.

TABLE II
LONG-RUN AVERAGE COSTS OVER DIFFERENT COST FACTORS

		γ^*	Maintenance cost
Baseline		0.16	1,178.49
C_3	300	0.16	923.29
	2,700	0.16	1,331.62
	5,400	1.0	1,686.29
	7,200	1.0	1,686.29
C_4	4,075	0.73	735.52
	8,150	0.2	998.00
	32,600	0.11	1,401.95
	48,900	0.11	1,556.13
L	30	0.16	1,192.83
	50	0.16	1,207.17
	100	0.16	1,243.02
	200	0.13	1,292.07

D. Comparison with scheduled maintenance

Finally, we compare our approach with the periodic maintenance strategy. We consider two periodic maintenance schedules, referred to as Scheduled_{236} and Scheduled_{447}, based on the mean times to alert (236 days) and alarm states (447 days) obtained from the state transition distributions $F_i(\cdot)$, $i \in \{1, 2, 3\}$ explained in Section VI-A. When conducting a maintenance activity in the proposed approach, only units in state $S + 1$ through M (i.e., alarm and failure states in this case) are repaired (or replaced).

Table III shows experimental results. We observe that between the two periodic maintenance policies, Scheduled_{236} outperforms Scheduled_{447}. Increasing the setup cost makes the gap between two polices smaller because the cost increases faster under Scheduled_{236} than under Scheduled_{447}. However, our proposed policy dominates both periodic policies in all cases with considerably smaller long-run average costs.

TABLE III
COMPARISON OF LONG-RUN AVERAGE COSTS WITH SCHEDULED MAINTENANCE ($N = 100, C_3 = 1,800, C_4 = 16,300, L = 10$)

P[a]	Proposed		Scheduled_{236}[b]	Scheduled_{447}[b]
	γ^*	Cost		
3.6	0.05	490.23	1,458 (11.82)	1,737.72 (8.88)
9	0.10	578.33	1,477.59 (12.62)	1,749.74 (8.72)
18	0.11	682.55	1,508.26 (12.87)	1,770.50 (7.82)
36	0.11	878.21	1,577.59 (11.05)	1,811.42 (8.58)
54	0.16	1,055.42	1,643.25 (11.96)	1,850.43 (8.69)
90	0.16	1,332.34	1,777.27 (10.51)	1,932.09 (9.22)

[a] Actual setup cost P is multiplied by 1,000.
[b] Subscripts 236 and 447 are maintenance intervals. Numbers in parentheses are the standard deviation.

E. Optimizing α_i and S

As mentioned in Section III, the values of α_i and S can be determined depending on the problem context. However, they can be optimized in terms of cost. In our current setting ($M = 4$), we can simply enumerate possible cases by discretizing α_3 values. Table IV shows the optimal α_3^* by enumerating discretized values when $N = 100$, $\alpha_4 = 1.0$, $P = 18000$, $C_3 = 1800$, $C_4 = 16300$, and $L = 10$.

We observe that the optimal threshold γ^* tends to increase as α_3 increases. However, the optimal cost does not exhibit

TABLE IV
OPTIMIZING S AND α_3 BY ENUMERATION

$S=2$		
α_3	cost	γ^*
0.00	709.67	0.01
0.05	700.00	0.02
0.10	686.43	0.02
0.15	685.72	0.03
0.20	703.00	0.04
0.25	692.42	0.05
0.30	701.18	0.06
0.35	700.66	0.06
0.40	676.36	0.07
0.45	700.70	0.08
0.50	709.68	0.09
0.55	**656.58**	**0.08**
0.60	682.55	0.11
0.65	662.37	0.11
0.70	663.36	0.10
0.75	692.73	0.15
0.80	693.98	0.13
0.85	663.20	0.12
0.90	684.53	0.16
0.95	699.42	0.16
1.00	670.45	0.14
$S=3$		
N/A	1,801.20	0.12

monotonic behavior. In Table IV, we see that the optimal α_3 is 0.55 with $\gamma^* = 0.08$ when $S = 2$. Since the proposed algorithm relies on numerical iterations, the cost values are slightly bumpy over α_3. Improving the algorithm's stability is an avenue for future work.

Due to the nature of our algorithm, obtaining the optimal cost values for given α_i and S is only possible through numerical evaluation. Thus, the cost function can be regarded as a black-box function. For larger values of M, where the enumeration approach is impractical, techniques designed for hyperparameter optimization can be employed. One potential method is Bayesian optimization, which has recently gained significant attention from researchers for tuning hyperparameters [48].

VII. CONCLUSION

This study derives a new fluid degradation model for large-scale systems consisting of many independent homogeneous units. We show that the fluid model can accurately describe the dynamics of the system-level health condition over time. Based on the fluid model, we devise a cost-effective selective maintenance strategy that minimizes a long-run average cost by triggering a maintenance activity when hitting a threshold value. Finding the optimal threshold is not trivial since the optimization problem involves integral equations. Further, the selective nature of maintenance operations prevents us from using a renewal theory. We, however, observe that the degradation process with selective maintenance becomes asymptotically regenerative. Using this finding, we reformulate the problem and devise a new algorithm.

From the numerical experiments, we confirm that the proposed fluid-model-based strategy estimates the system dynamics and maintenance costs accurately. We benchmark the proposed strategy against the two periodic maintenance policies where the periods are obtained from the average transition time to alert and alarm states. Notably, the proposed strategy dominates both periodic policies.

There can be a few directions for future work extending the proposed approach. First, we plan to take several types of dependencies among units into consideration. This paper accounts for only economic dependence among units. Considering other dependence such as stochastic and structural dependence will make the model more realistic, although it would result in a much complicated mathematical model. Second, this paper provides the proof for the conjecture on the regenerative process for a limited three-state Markovian case. We plan to extend it and conduct comprehensive analysis for more general non-Markovian settings. We believe doing so will involve in-depth theoretical analysis related to the impulsive ordinary differential equation.

APPENDIX

Proof of Theorems V.2 and V.3 Recall the solution of the system of the ordinary differential equations.

$$
\begin{aligned}
X_1(t) &= e^{-\lambda_1 t} X_1(0) \\
X_2(t) &= \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) e^{-\lambda_1 t} \\
&\quad + \left(X_2(0) - \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) \right) e^{-\lambda_2 t} \\
X_3(t) &= 1 - X_1(t) - X_2(t).
\end{aligned} \tag{16}
$$

We consider a flow ϕ associated with $\mathbf{X}$. Let $\mathbf{I}$ be the initial point at time 0. Let $\mathbf{H}$ and τ be the point and the time when ϕ hits the threshold, γ, respectively; $\mathbf{H} = \phi(\mathbf{I}, \tau)$. The maintenance operation moves the value of state 3 and keeps the value of state 2 untouched. As such, if there exist $\mathbf{I}$ and $\mathbf{H}$ such that $\mathbf{I}|_{x_2} = \mathbf{H}|_{x_2}$, the flow $\phi(\mathbf{I}, t)$ forms a periodic cycle, where $\mathbf{X}|_{x_i}$ denote the x_i-coordinate of $\mathbf{X}$.

Lemma A.1. *There exist* $\mathbf{I}$ *and* $\mathbf{H}$ *such that* $\mathbf{I}|_{x_2} = \mathbf{H}|_{x_2}$.

Proof. Now we will prove the existence of such $\mathbf{I}$ and $\mathbf{H}$. In our maintenance problem, $\mathbf{X}(0) = (1, 0, 0)$. The starting points after successive maintenance operations lie in the set $\mathcal{I} = \{\mathbf{X} = (X_1, X_2, X_3) \in \mathbb{R}^3_+ \mid X_1 + X_2 = 1, X_3 = 0\}$. Consider an arbitrary initial point $\mathbf{I} \in \mathcal{I}$ and its corresponding hitting point $\mathbf{H}$ with hitting time τ. Define

$$
\begin{aligned}
f(\mathbf{I}) &:= \mathbf{I}|_{x_2} - \mathbf{H}|_{x_2} \\
&= X_2(0) - X_2(\tau).
\end{aligned}
$$

Then, using (16) and the fact that $X_1(0) + X_2(0) = 1$, we have

$$
\begin{aligned}
f(\mathbf{I}) &= X_2(0) - \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) e^{-\lambda_1 \tau} \\
&\quad - \left(X_2(0) - \frac{\lambda_1}{\lambda_2 - \lambda_1} X_1(0) \right) e^{-\lambda_2 \tau} \\
&= \frac{1}{\lambda_2 - \lambda_1} \left(\lambda_1 e^{-\lambda_1 \tau} - \lambda_2 e^{-\lambda_2 \tau} + (\lambda_2 - \lambda_1) \right) X_2(0) \\
&\quad - \frac{\lambda_1}{\lambda_2 - \lambda_1} \left(e^{-\lambda_1 \tau} - e^{-\lambda_2 \tau} \right).
\end{aligned} \tag{17}
$$

Consider the boundary points of $\mathcal{I}$: $\mathbf{I}_0 = (1,0,0)$ and $\mathbf{I}_1 = (0,1,0)$. The corresponding hitting times are denoted by $\tau_{\mathbf{I}_0}$ and $\tau_{\mathbf{I}_1}$, respectively. Then, we have

$$f(\mathbf{I}_0) = -\frac{\lambda_1}{\lambda_2-\lambda_1}\left(e^{-\lambda_1\tau_{\mathbf{I}_0}} - e^{-\lambda_2\tau_{\mathbf{I}_0}}\right) < 0 \tag{18}$$

$$\begin{aligned} f(\mathbf{I}_1) &= \frac{1}{\lambda_2-\lambda_1}\left(\lambda_1 e^{-\lambda_1\tau_{\mathbf{I}_1}} - \lambda_2 e^{-\lambda_2\tau_{\mathbf{I}_1}} + (\lambda_2-\lambda_1)\right) \\ &\quad - \frac{\lambda_1}{\lambda_2-\lambda_1}\left(e^{-\lambda_1\tau_{\mathbf{I}_1}} - e^{-\lambda_2\tau_{\mathbf{I}_1}}\right) \\ &= 1 - e^{-\lambda_2\tau_{\mathbf{I}_1}} > 0. \end{aligned} \tag{19}$$

Therefore, by the intermediate value theorem, there exists $\mathbf{I}^* \in \mathcal{I}$ such that $f(\mathbf{I}^*) = 0$. □

To extend Lemma A.1 to the uniqueness of such a cycle, Lemmas A.2 and A.3 show that $f(\mathbf{I})$ is an increasing function in $X_2(0)$.

Lemma A.2. *τ is strictly increasing in $X_1(0)$.*

Proof. Consider two initial points, $\mathbf{I}_1 = (x_1, x_2, 0)$ and $\mathbf{I}_2 = (x_1', x_2', 0)$ with $x_1 < x_1'$. Let $X_i^{\mathbf{I}_j}(t)$ denote the $X_i(t)$ with initial point $\mathbf{I}_j$ for $i \in \{1,2,3\}$ and $j \in \{1,2\}$. Then, for all $t > 0$,

$$\begin{aligned} X_3^{\mathbf{I}_1}(t) - X_3^{\mathbf{I}_2}(t) &= X_1^{\mathbf{I}_2}(t) - X_1^{\mathbf{I}_1}(t) + X_2^{\mathbf{I}_2}(t) - X_2^{\mathbf{I}_1}(t) \\ &= -\lambda_2(x_1' - x_1)\frac{e^{-\lambda_2 t} - e^{-\lambda_1 t}}{\lambda_2-\lambda_1} > 0. \end{aligned} \tag{20}$$

From (20), we see that $X_3(t)$ hits the threshold γ early when $X_1(0) = x_1$ is small. □

Lemma A.3. *$f(\mathbf{I})$ is a strictly increasing function in $X_2(0)$*

Proof. Define $g(\mathbf{I}) := \mathbf{I}|_{x_1} - \mathbf{H}|_{x_1}$. Then, we have $g(\mathbf{I}) = \gamma - f(\mathbf{I})$. If we show that $g(\mathbf{I})$ is a strictly increasing function in $X_1(0)$, $f(\mathbf{I})$ is a strictly decreasing function in $X_1(0)$. Then, by $X_2(0) = 1 - X_1(0)$, $f(\mathbf{I})$ is a strictly increasing function in $X_2(0)$.

$$\begin{aligned} g(\mathbf{I}) &= X_1(0) - X_1(\tau) = X_1(0)(1 - e^{-\lambda_1\tau}) \\ \frac{\partial g(\mathbf{I})}{\partial X_1(0)} &= 1 - e^{-\lambda_1\tau} + X_1(0)\lambda_1 e^{-\lambda_1\tau}\frac{\partial\tau}{\partial X_1(0)} \\ &= 1 - e^{-\lambda_1\tau}\left(1 - X_1(0)\lambda_1\frac{\partial\tau}{\partial X_1(0)}\right) \\ &> 1 - e^{-\lambda_1\tau} \\ &> 0 \end{aligned}$$

□

Theorem V.2. *Let $\mathbf{I}^{(k)} \in \mathbb{I}$ denote the initial point right after the k^{th} maintenance where $\mathbf{I}^{(0)}$ denotes the initial point at the beginning of the horizon (i.e., $t = 0$). Then, there exists a unique $\mathbf{I}^* \in \mathbb{I}$ satisfying $\mathbf{I}^{(k)} = \mathbf{I}^*$ for all $k \in \mathbb{Z}+$.*

Proof. Lemma A.1 shows the existence of $\mathbf{I}^*$, and Lemma A.3 proves the uniqueness of $\mathbf{I}^*$. □

Next, we prove the global attractiveness of the solution.

Lemma A.4.

$$\frac{\partial\tau}{\partial X_1(0)} \le \frac{1}{\lambda_1 X_1(0)}. \tag{21}$$

Proof.

$$\begin{aligned} X_3(\tau) = \gamma &= 1 - X_1(\tau) - X_2(\tau) \\ &= 1 - e^{-\lambda_1\tau}X_1(0) - \frac{\lambda_1}{\lambda_2-\lambda_1}X_1(0)e^{-\lambda_1\tau} \\ &\quad - \left(X_2(0) - \frac{\lambda_1}{\lambda_2-\lambda_1}X_1(0)\right)e^{-\lambda_2\tau} \\ &= 1 - \frac{\lambda_2}{\lambda_2-\lambda_1}e^{-\lambda_1\tau}X_1(0) \\ &\quad - \left(1 - \frac{\lambda_2}{\lambda_2-\lambda_1}X_1(0)\right)e^{-\lambda_2\tau}. \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial\gamma}{\partial X_1(0)} \\ &= -\frac{\lambda_2}{\lambda_2-\lambda_1}\left(e^{-\lambda_1\tau} - e^{-\lambda_2\tau}\right) + \left(\frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}e^{-\lambda_1\tau}X_1(0)\right. \\ &\quad \left. + \left(\lambda_2 - \frac{\lambda_2^2}{\lambda_2-\lambda_1}X_1(0)\right)e^{-\lambda_2\tau}\right)\frac{\partial\tau}{\partial X_1(0)}. \end{aligned}$$

What we need to show is

$$\begin{aligned} \frac{\partial\tau}{\partial X_1(0)} &= \frac{\lambda_2\left(e^{-\lambda_1\tau} - e^{-\lambda_2\tau}\right)}{\lambda_1\lambda_2 e^{-\lambda_1\tau}X_1(0) + \left(\lambda_2^2 - \lambda_1\lambda_2 - \lambda_2^2 X_1(0)\right)e^{-\lambda_2\tau}} \\ &\le \frac{1}{\lambda_1 X_1(0)} \end{aligned} \tag{22}$$

Consider the following two cases:
Case 1: For $\lambda_1 > \lambda_2$, then $e^{-\lambda_1\tau} - e^{\lambda_2\tau} < 0$ and $\lambda_1\lambda_2 e^{-\lambda_1\tau}X_1(0) + \left(\lambda_2^2 - \lambda_1\lambda_2 - \lambda_2^2 X_1(0)\right)e^{-\lambda_2\tau} < 0$. S Showing (22) is equivalent to the following

$$\begin{aligned} &\lambda_2\left(e^{-\lambda_1\tau} - e^{\lambda_2\tau}\right)\lambda_1 X_1(0) \\ &\qquad \ge \lambda_1\lambda_2 e^{-\lambda_1\tau}X_1(0) + \left(\lambda_2^2 - \lambda_1\lambda_2 - \lambda_2^2 X_1(0)\right)e^{-\lambda_2\tau} \\ &\iff -\lambda_1\lambda_2 X_1(0) \ge \lambda_2^2 - \lambda_1\lambda_2 - \lambda_2^2 X_1(0) \\ &\iff \lambda_2(\lambda_2-\lambda_1)(X_1(0) - 1) \ge 0. \end{aligned}$$

Case 2: For $\lambda_1 < \lambda_2$, similar to Case 1, we can show (22) by showing $\lambda_2(\lambda_2-\lambda_1)(X_1(0)-1) \le 0$. Therefore, we have

$$\frac{\partial\tau}{\partial X_1(0)} \le \frac{1}{\lambda_1 X_1(0)}.$$

□

Corollary A.5. *Define $g(\mathbf{I}) := \mathbf{I}|_{x_1} - \mathbf{H}|_{x_1}$. Then $0 < \frac{\partial g(\mathbf{I})}{\partial X_1(0)} \le 1$*

Proof. From the proof of Lemma A.3, we have

$$\begin{aligned} \frac{\partial g(\mathbf{I})}{\partial X_1(0)} &= 1 - e^{-\lambda_1\tau} + X_1(0)\lambda_1 e^{-\lambda_1\tau}\frac{\partial\tau}{\partial X_1(0)} \\ &= 1 - e^{-\lambda_1\tau}\left(1 - X_1(0)\lambda_1\frac{\partial\tau}{\partial X_1(0)}\right) \le 1, \end{aligned}$$

which completes the proof. □

Theorem V.3. *For any $\mathbf{I} = \mathbf{I}^{(0)} \in \mathbb{I}$, $\mathbf{I}^{(k)} \longrightarrow \mathbf{I}^*$ as $k \to \infty$.*

Proof. Define $X_1^* = \mathbf{I}^*|_{x_1}$. By Corollary A.5,

$$0 < \frac{g(I) - g\left(I^*\right)}{X_1(0) - X_1^*} = \frac{X_1(0) - X_1(\tau) - \gamma}{X_1(0) - X_1^*} \leq 1.$$

Consider the following cases:
Case 1: For $X_1(0) = X_1^*$, by Theorem V.2, we have

$$\|\mathbf{I}|_{x_1} - X_1^*\| - \|\mathbf{I}^{(1)}|_{x_1} - X_1^*\| = 0.$$

Case 2: For $X_1(0) < X_1^*$, it holds

$$\begin{aligned} &X_1(0) - X_1(\tau) - \gamma < 0, \quad 0 \geq X_1(\tau) + \gamma - X_1^* \\ &\left\|\mathbf{I}|_{x_1} - X_1^*\right\| - \|\mathbf{I}^{(1)}|_{x_1} - X_1^*\| \\ &\qquad = \|X_1(0) - X_1^*\| - \|X_1(\tau) + \gamma - X_1^*\| \\ &\qquad = (X_1^* - X_1(0)) - (-X_1(\tau) - \gamma + X_1^*) \\ &\qquad = X_1(\tau) + \gamma - X_1(0) \\ &\qquad > 0 \end{aligned}$$

Case 3: For $X_1(0) > X_1^*$, we similarly obtain

$$\begin{aligned} &X_1(0) - X_1(\tau) - \gamma > 0, \quad 0 \leq X_1(\tau) + \gamma - X_1^* \\ &\left\|\mathbf{I}|_{x_1} - X_1^*\right\| - \|\mathbf{I}^{(1)}|_{x_1} - X_1^*\| \\ &\qquad = \|X_1(0) - X_1^*\| - \|X_1(\tau) + \gamma - X_1^*\| \\ &\qquad = (X_1(0) - X_1^*) - (X_1(\tau) + \gamma - X_1^*) \\ &\qquad = -X_1(\tau) - \gamma + X_1(0) \\ &\qquad > 0 \end{aligned}$$

The sequence $\left\{\|\mathbf{I}^{(k)}|_{x_1} - X_1^*\|, k \in \mathbb{N}\right\}$ in $\mathbb{R}$ is monotonically decreasing and bounded below with lower bound 0 . By the monotone convergence theorem, this sequence converges to $\inf \left\|\mathbf{I}^{(k)}|_{x_1} - X_1^*\right\| = 0$, i.e., $\mathbf{I}^{(k)}$ converges to $\mathbf{I}^*$ as $k \to \infty$. □

References

[1] M. You, E. Byon, J. Jin, and G. Lee, "When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms," *IISE Trans.*, vol. 49, no. 1, pp. 84–95, 2017.

[2] W. Cao, X. Jia, Q. Hu, J. Zhao, and Y. Wu, "A literature review on selective maintenance for multi-unit systems," *Qual. Reliab. Eng. Int.*, vol. 34, no. 5, pp. 824–845, 2018.

[3] T. Jiang and Y. Liu, "Selective maintenance strategy for systems executing multiple consecutive missions with uncertainty," *Reliab. Eng. Syst. Safe.*, vol. 193, p. 106632, 2020.

[4] Y. Liu, Y. Chen, and T. Jiang, "On sequence planning for selective maintenance of multi-state systems under stochastic maintenance durations," *Eur. J. Oper. Res.*, vol. 268, no. 1, pp. 113–127, 2018.

[5] G. Maaroufi, A. Chelbi, and N. Rezg, "Optimal selective renewal policy for systems subject to propagated failures with global effect and failure isolation phenomena," *Reliab. Eng. Syst. Safe.*, vol. 114, no. 1, pp. 61–70, 2013.

[6] D. Yang, H. Wang, Q. Feng, Y. Ren, B. Sun, and Z. Wang, "Fleet-level selective maintenance problem under a phased mission scheme with short breaks: A heuristic sequential game approach," *Computers and Industrial Engineering*, vol. 119, pp. 404–415, 2018.

[7] E. Byon, L. Ntaimo, and Y. Ding, "Optimal maintenance strategies for wind turbine systems under stochastic weather conditions," *IEEE Trans. Rel.*, vol. 59, no. 2, pp. 393–404, 2010.

[8] L. M. Maillart, C. R. Cassady, C. Rainwater, and K. Schneider, "Selective Maintenance Decision-Making Over Extended Planning Horizons," *IEEE Trans. Rel.*, vol. 58, no. 3, pp. 462–469, 2009.

[9] Z. Tian and H. Liao, "Condition based maintenance optimization for multi-component systems using proportional hazards model," *Reliab. Eng. Syst. Safe.*, vol. 96, no. 5, pp. 581–589, 2011.

[10] Y. M. Ko and E. Byon, "Reliability evaluation of large-scale systems with identical units," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 420–434, 2015.

[11] J. A. Andrawus, J. Watson, and M. Kishk, "Modelling system failures to optimise wind turbine maintenance," *Wind Eng.*, vol. 31, no. 6, pp. 503–522, 2007.

[12] B. Castanier, A. Grall, and C. Berenguer, "A condition-based maintenance policy with non-periodic inspections for a two-unit series system," *Reliab. Eng. Syst. Safe.*, vol. 87, no. 1, pp. 109–120, 2005.

[13] C. van Oosterom, H. Peng, and G.-J. van Houtum, "Maintenance optimization for a Markovian deteriorating system with population heterogeneity," *IISE Trans.*, vol. 49, no. 1, pp. 96–109, 2017.

[14] N. Salari and V. Makis, "Comparison of two maintenance policies for a multi-unit system considering production and demand rates," *Int. J. Prod. Econ.*, vol. 193, pp. 381–391, 2017.

[15] Y. M. Ko and E. Byon, "Condition-based joint maintenance optimization for a large-scale system with homogeneous units," *IISE Trans.*, vol. 49, no. 5, pp. 493–504, 2017.

[16] Y. Wang and H. Pham, "A multi-objective optimization of imperfect preventive maintenance policy for dependent competing risk systems with hidden failure," *IEEE Trans. Rel.*, vol. 60, no. 4, pp. 770–781, Dec 2011.

[17] L. Jiang, Q. Feng, and D. Coit, "Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds," *IEEE Trans. Rel.*, vol. 61, no. 4, pp. 932–948, 2012.

[18] S. Petchrompo and A. K. Parlikad, "A review of asset management literature on multi-asset systems," *Reliab. Eng. Syst. Safe.*, vol. 181, pp. 181–201, 2019.

[19] R. P. Nicolai and R. Dekker, *Optimal Maintenance of Multi-component Systems: A Review*. London: Springer London, 2008, pp. 263–286.

[20] M. C. Olde Keizer, S. D. P. Flapper, and R. H. Teunter, "Condition-based maintenance policies for systems with multiple dependent components: A review," *Eur. J. Oper. Res.*, vol. 261, no. 2, pp. 405–420, 2017.

[21] B. Castanier, A. Grall, and C. Bérenguer, "A condition-based maintenance policy with non-periodic inspections for a two-unit series system," *Reliab. Eng. Syst. Safe.*, vol. 87, no. 1, pp. 109–120, 2005.

[22] X. Zhang and J. Zeng, "Joint optimization of condition-based opportunistic maintenance and spare parts provisioning policy in multiunit systems," *Eur. J. Oper. Res.*, vol. 262, no. 2, pp. 479–498, 2017.

[23] T. Jiang and Y. Liu, "Robust selective maintenance strategy under imperfect observations: A multi-objective perspective," *IISE Trans.*, vol. 52, no. 7, pp. 751–768, 2020.

[24] Y. Liu, J. Gao, T. Jiang, and Z. Zeng, "Selective maintenance and inspection optimization for partially observable systems: An interactively sequential decision framework," *IISE Trans.*, vol. 55, pp. 463–479, 2023.

[25] M. Ghorbani, M. Nourelfath, and M. Gendreau, "A two-stage stochastic programming model for selective maintenance optimization," *Reliab. Eng. Syst. Safe.*, vol. 223, p. 108480, 7 2022.

[26] K. Ahadi and K. M. Sullivan, "Approximate Dynamic Programming for Selective Maintenance in Series-Parallel Systems," *IEEE Trans. Rel.*, vol. 69, pp. 1147–1164, 9 2020.

[27] Y. Xu, D. Pi, Z. Wu, J. Chen, and E. Zio, "Hybrid Discrete Differential Evolution and Deep Q-Network for Multimission Selective Maintenance," *IEEE Trans. Rel.*, vol. 71, pp. 1501–1512, 12 2022.

[28] Y. Liu, Y. Chen, and T. Jiang, "Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach," *Eur. J. Oper. Res.*, vol. 283, no. 1, pp. 166–181, 2020.

[29] H. Hesabi, M. Nourelfath, and A. Hajji, "A deep learning predictive model for selective maintenance optimization," *Reliab. Eng. Syst. Safe.*, vol. 219, p. 108191, 3 2022.

[30] A. Khatab, C. Diallo, U. Venkatadri, Z. Liu, and E. H. Aghezzaf, "Optimization of the joint selective maintenance and repairperson assignment problem under imperfect maintenance," *Computers and Industrial Engineering*, vol. 125, pp. 413–422, 2018.

[31] M. Yin, Y. Liu, S. Liu, Y. Chen, and Y. Yan, "Scheduling heterogeneous repair channels in selective maintenance of multi-state systems with maintenance duration uncertainty," *Reliab. Eng. Syst. Safe.*, vol. 231, p. 108977, 3 2023.

[32] W. Ma, Q. Zhang, T. Xiahou, Y. Liu, and X. Jia, "Integrated selective maintenance and task assignment optimization for multi-state systems executing multiple missions," *Reliab. Eng. Syst. Safe.*, vol. 237, p. 109330, 9 2023.

[33] K. Chaabane, A. Khatab, C. Diallo, E. H. Aghezzaf, and U. Venkatadri, "Integrated imperfect multimission selective maintenance and repairpersons assignment problem," *Reliab. Eng. Syst. Safe.*, vol. 199, p. 106895, 2020.

[34] H. Chen and D. D. Yao, *Fundamentals of queueing networks: performance, asymptotics, and optimization*, 1st ed. Springer-Verlag, New York, 2001.
[35] W. Whitt, *Stochastic Process Limits*, 1st ed. Springer-Verlag, New York, 2002.
[36] ——, “Fluid Models for Multiserver Queues with Abandonments,” *Oper. Res.*, vol. 54, no. 1, pp. 37–54, 2006.
[37] Y. Liu and W. Whitt, “The $G_t/GI/s_t + GI$ many-server fluid queue,” *Queueing Syst.*, vol. 71, no. 4, pp. 405–444, 2012.
[38] C. v. Oosterom, L. M. Maillart, and J. P. Kharoufeh, “Optimal maintenance policies for a safety-critical system and its deteriorating sensor,” *Nav. Res. Log.*, vol. 64, no. 5, pp. 399–417, 2017.
[39] L. M. Maillart and L. Zheltova, “Structured maintenance policies on interior sample paths,” *Nav. Res. Log.*, vol. 54, no. 6, pp. 645–655, 2007.
[40] M. J. Kim and V. Makis, “Joint Optimization of Sampling and Control of Partially Observable Failing Systems,” *Oper. Res.*, vol. 61, no. 2, pp. 777–790, 2013.
[41] H. Zhang and W. Zhang, “Analytical Solution to a Partially Observable Machine Maintenance Problem with Obvious Failures,” *Manage. Sci.*, vol. 69, no. 7, pp. 3993–4015, 2023.
[42] L. M. Maillart, “Maintenance policies for systems with condition monitoring and obvious failures,” *IIE Trans.*, vol. 38, no. 6, pp. 463–475, 2006.
[43] E. Byon, “Wind turbine operations and maintenance: A tractable approximation of dynamic decision-making,” *IIE Trans.*, vol. 45, no. 11, pp. 1188–1201, 2013.
[44] X. Liu, Q. Sun, Z.-S. Ye, and M. Yildirim, “Optimal multi-type inspection policy for systems with imperfect online monitoring,” *Reliab. Eng. Syst. Safe.*, vol. 207, p. 107335, 2021.
[45] D. Bainov and P. Simeonov, *Impulsive Differential Equations: Periodic Solutions and Applications*, 1st ed. Chapman and Hall/CRC, 1993.
[46] E. Byon, E. Perez, L. Ntaimo, and Y. Ding, “Simulation of wind farm operations and maintenance using devs,” *Simulation*, vol. 87, no. 12, pp. 1093–1117, 2011.
[47] Z. Tian and H. Liao, “Condition based maintenance optimization for multi-component systems using proportional hazards model,” *Reliab. Eng. Syst. Safe.*, vol. 96, no. 5, pp. 581–589, 2011.
[48] R. Garnett, *Bayesian Optimization*, 1st ed. Cambridge University Press, 2023.

Euihwan Lee is a manager of Global Data Analysis Team of Hyundai Motor Company. He received his B.S. and M.S. degrees in Mathematics and Industrial and Management Engineering from Pohang University of Science and Technology (POSTECH) in 2016 and 2018, respectively. His research focuses on the optimization of stochastic systems and data analysis.

Jaehyoung Ju is a Ph. D. student at Pohang University of Science and Technology (POSTECH). He received his B.S. degree in Mathematics from POSTECH in 2021. His main interests include distributionally robust optimization, simulation, and stochastic approximation.

Eunshin Byon is an associate professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her B.S. and M.S. in Industrial and Systems Engineering from the Korea Advanced Institute of Science and Technology (KAIST) and Ph.D. in Industrial and Systems Engineering from Texas A&M University. Her research interests include optimizing operations of renewable systems, data science, quality and reliability engineering, and sustainability. She is a member of IISE, INFORMS, and IEEE.

Young Myoung Ko received the B.S. and M.S. degrees in industrial engineering from Seoul National University, Seoul, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree in industrial engineering from Texas A&M University, College Station, TX, USA, in 2011.

He is currently an Associate Professor in the Department of Industrial and Management Engineering with the Pohang University of Science and Technology (POSTECH), South Korea. He serves as a Secretary of INFORMS Telecommunications and Network Analytics Section and also an Associate Editor of *IISE Transactions*. His research focuses on the reliability and maintenance optimization of large-scale stochastic systems such as service systems, telecommunication networks, ICT infrastructure, and renewable energy systems.