CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Economic control of next generation aerospace fleets

Felipe Montana[1] ⬤, Will Jacobs[1], Evangelia Pontika[2], Panagiotis Laskaridis[2], Matthew Griffiths[3],
Peter Beecroft[3], Derek Wall[3], Visakan Kadirkamanathan[1] and Andrew R Mills[1]

[1]School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, UK
[2]Centre for Propulsion and Thermal Power Engineering, Cranfield University, Bedford, UK
[3]Rolls-Royce Plc, London, UK
Correspondence author: Felipe Montana; Email: f.montana-gonzalez@sheffield.ac.uk

Abstract

Engineering machines are becoming increasingly complex and possess more control variables, increasing the complexity and versatility of the control systems. Different configurations of the control system, named a policy, can result in similar output behavior but with different resource or component life usage. There is therefore an opportunity to find optimal policies with respect to economic decisions. While many solutions have been proposed to find such economic policy decisions at the asset level, we consider this problem at the fleet level. In this case, the optimal operation of each asset is affected by the state of all other assets in the fleet. Challenges introduced by considering multiple assets include the construction of economic multi-objective optimization criteria, handling rare events such as failures, application of fleet-level constraints, and scalability. The proposed solution presents a framework for economic fleet optimization. The framework is demonstrated for economic criteria relating to resource usage, component lifing, and maintenance scheduling, but is generically extensible. Direct optimization of lifetime distributions is considered in order to avoid the computational burden of discrete event simulation of rare events. Results are provided for a real-world case study targeting the optimal economic operation of a fleet of aerospace gas turbine engines.

Impact statement

There is a significant opportunity to reduce economic costs across a fleet of assets through optimization of the control policies under which assets operate. Reduction of economic costs is particularly important for high-value assets. A framework for the optimization of control policies is presented that considers both asset-level behavior and the fleet-level behavior that arises through the interaction between the assets. The framework is demonstrated for economic criteria relating to resource usage, component lifing, and maintenance scheduling, but is generically extensible. Results from a real-world case study are given that show the reduction in economic costs across a fleet of aerospace gas turbine engines. The framework has wide applicability, particularly in vehicular fleets such as aerospace, marine, and haulage industries as well as military fleets.

1. Introduction

An intelligent approach to the management of fleets can lead to a reduction in both operating costs and emissions while also increasing sustainability. Fleet management is particularly pertinent in the case of fleets containing complex, high-value assets with many degrees of freedom. Such assets tend to be

expensive to operate and generate a large carbon footprint. The fleet management problem has been widely studied (Bielli et al., 2011), examples include; vehicle routing (Bruglieri et al., 2019), fleet composition (Yıldırım and Yıldız, 2021) and maintenance scheduling (Sanchez et al., 2020). In general, it is a challenging task, even in a static context, due to its combinatorial nature, that is combinations of behaviors across assets. The problem becomes more challenging when uncertainties such as future operation conditions, asset faults, resource availability, and uncertain future demands, are introduced.

One approach for fleet management is to use optimization routines to determine the optimal fleet behavior with respect to a set of objectives that represent an economic or environmental (or both) cost. Different problem formulations aim to optimize fleet behavior with respect to different costs, for example; increased fleet availability (Gavranis and Kozanidis, 2015; Vos et al., 2023), reduction of operational costs (List et al., 2003; Militão and Tirachini, 2021), reduction of carbon emission (Sawik et al., 2017; Sheng et al., 2019), etc.

It is often not feasible to construct these optimization problems analytically due to the highly complex nature of modern engineering assets, leading to many works choosing to use simulation-based approaches (Turan et al., 2020) and digital twins of assets (Kraft and Kuntzagk, 2017).

A fleet commonly contains groups of similar assets, but which are unique with respect to their operational and build history, i.e., when, where, and how they have been operated, current health state, original manufacture variability, and so forth each asset has many control variables and operating parameters which influence how the asset operates, see Figure 1. In this paper, it is assumed that every asset has the same control variables. However, the proposed framework can handle assets with different control variables. The chosen configuration of the control variables, named the *policy*, may have a large degree of redundancy with respect to the output of the asset, but have significant implications on resource usage and component lifing– and therefore on the economy of operation. The term lifing refers to the estimation of the remaining useful life of an asset or component. Due to the unique state of each asset and operating parameters, the optimal policy choice is different for each asset and may be constrained by external elements such as commercial contracts, operation costs, acceptable failure rates, asset maintenance, etc.

The problem of finding optimal policies to extend the life of the assets, across these degrees of freedom, has been previously investigated at the asset level (Zheng et al., 2017; Chen and Guo, 2007). However, these solutions do not consider economic benefits that can be made at the fleet level. At the fleet level, the solution needs to consider the current state of all the assets and their interactions. For example, naively extending the life of multiple assets may lead to low availability or high maintenance shop occupancy during a later period. The optimal solution at the asset level may therefore be sub-optimal at the fleet level. Moreover, trade-offs between different objectives at the fleet level might exist. Performing the optimization at the fleet level has the potential to make significant economic and environmental savings but, to the best of the authors' knowledge, has not been addressed in the existing literature.

In this work, we present a computationally tractable framework for fleet-level optimal policy selection. The proposed framework, referred to as an economic-based controller, generates optimal control policies for each
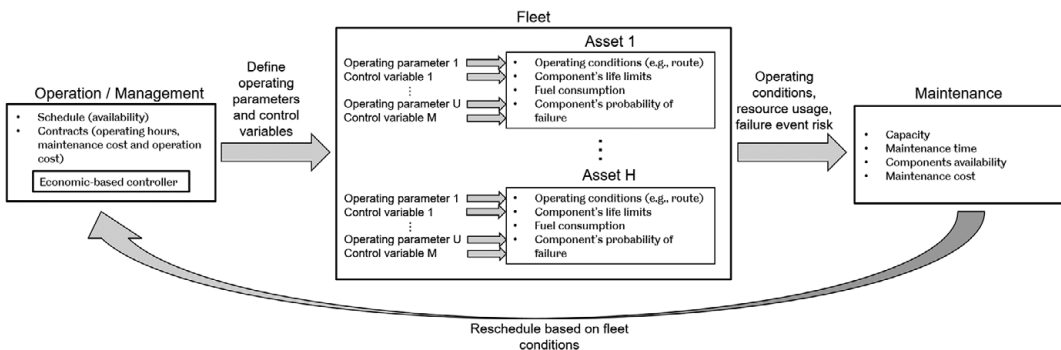

Figure 1. *Interaction between elements in fleet operation.*

of the fleet's assets that improve expected life across the fleet while reducing operating costs and satisfying constraints such as fleet availability and maintenance shop capacity. The developed approach is demonstrated using a real-world case study targeting the optimal operation of a fleet of aerospace gas turbine engines. Asset lifing models are developed to find the relationship between policies and lifing. A fleet-level model is also developed and used to capture fleet-level behaviors. We demonstrate how such models can be used within common optimization frameworks to provide optimal solutions to fleet-level economic control. The main contributions of the proposed approach are: (1) a framework for economic policy optimization at the fleet level, (2) an approach for direct optimization of lifing models that remove the need for discrete event simulation of rare events, and (3) a demonstration of the developed methods in a real-world case study.

2. Fleet optimization

In this paper, we aim to find the optimal control policies across a fleet of complex assets. By finding an optimal policy for each asset, and considering their interactions, economic benefits can be obtained at the fleet level. Optimal policies can be found to increase the availability of the assets, restrict the number of assets simultaneously under maintenance, and so forth For example, for an asset close to planned maintenance, where a subset of its components are to be replaced, a control policy might be selected that can improve the asset performance at the cost of severe component usage–in order to exploit the otherwise wasted remaining life of these components. Alternatively, the fleet-level objective might encourage the selection of a policy that extends the maintenance interval in order to maintain workshop capacity. We are interested in minimizing multiple objectives such as reducing the probability of failure or minimizing the total fuel burn. A multi-objective optimization problem is therefore considered.

2.1. Problem definition

The problem is to find a set of policies for each of the H assets, $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_H\}$, where $\boldsymbol{x}_h = [x_{h,1}, x_{h,2}, \cdots, x_{h,M}]$ is the policy for the h'th asset which possesses M settings of the control system. The control setting $x_{h,m}$ could be, for example, an indexed control scheme or control parameterization, a control objective, or a valve state.

In order to optimize for multiple competing objectives at a different level, a multi-objective optimization problem is considered. The problem is formulated as:

$$
\begin{aligned}
&\min_{\boldsymbol{X}} \ O(\boldsymbol{X}) \\
&\text{where} \ O(\boldsymbol{X}) = (o_1, o_2, \ldots, o_I) \\
&s.t \ g_j(\boldsymbol{X}) \leq 0, \quad j = 1, 2, \ldots, J \\
&\quad d_k(\boldsymbol{x}_h) \leq 0, \quad k = 1, 2, \ldots, K; h \in \{1, \ldots, H\}
\end{aligned}
\tag{1}
$$

where $o_i(\boldsymbol{X})$ is i'th objective function, $g_j(\boldsymbol{X})$ is j'th fleet-level constraint, $d_k(\boldsymbol{x}_h)$ is k'th asset level constraint and I, J, and K are the number of objective functions, fleet-level constraints, and asset-level constraints, respectively.

The problem formulation is generic and accepts different types of objectives and constraints. Here, the objective is chosen with an aim to reduce economic costs through extending the life of the assets while minimizing resource usage under constraints on asset availability. As such, we consider an objective function of the form:

$$
O(\boldsymbol{X}) = (R(\boldsymbol{X}), L(\boldsymbol{X}), P_{\text{fail}}(\boldsymbol{X}), W(\boldsymbol{X})),
\tag{2}
$$

where

$$
R(\boldsymbol{X}) = \frac{1}{H} \sum_{h=1}^{H} r(\boldsymbol{x}_h),
\tag{3}
$$

$$L(\boldsymbol{X}) = -\frac{1}{H}\sum_{h=1}^{H} l(\boldsymbol{x}_h), \tag{4}$$

$$P_{\text{fail}}(\boldsymbol{X}) = \frac{1}{H}\sum_{h=1}^{H} P(T_{\text{fail},h} < T_{L,h}(\boldsymbol{x}_h)), \tag{5}$$

$$W(\boldsymbol{X}) = \frac{1}{H}\sum_{h=1}^{H} w(\boldsymbol{x}_h, s), \tag{6}$$

where $r(\boldsymbol{x}_h)$ is a function that quantifies resource usage of the h'th asset, $l(\boldsymbol{x}_h)$ is a function that quantifies a lifing cost associate with component usage/degradation, $T_{L,h}(\boldsymbol{x}_h)$ is the usage limit (or hard life) of the h'th asset (defined as the time at which the asset must be removed for maintenance as defined by the manufacturer or regulatory body), $T_{\text{fail},h}$ is the failure time of the h'th asset, $w(\boldsymbol{x}_h, s)$ is a function that quantifies economic costs associated with asset maintenance and depends on the state of the maintenance shop s. Maintenance costs could include the cost of replacement components, labor costs, or as demonstrated later, cost associated with occupancy. The objective function, $O(\boldsymbol{X})$ is therefore trying to minimize resource usage through $R(\boldsymbol{X})$, maximize asset life through $L(\boldsymbol{X})$, minimize the the failure probabilities through $P_{\text{fail}}(\boldsymbol{X})$ and reduce maintenance costs through $W(\boldsymbol{X})$.

Note that the objective $W(\boldsymbol{X})$ may introduce a coupling between the assets such that the optimization cannot be performed solely at the asset level.

2.2. Optimization with fleet simulators

In this section, the use of a fleet simulator is introduced as well as an overview of the optimization method used. A fleet simulator is required in order to evaluate the effect of the policies, $\boldsymbol{X}$, on the objective $O(\boldsymbol{X})$. The simulator must therefore be able to estimate resource usage and model lifing effects to estimate asset hard lives and probabilities of failure.

Under competing objectives, it is not always possible to find a solution that minimizes all objectives simultaneously. The variability of the fleet and the external influences, e.g., the current fuel price, may influence the priority of each of the objectives, but these can be difficult to quantify into weights. Therefore, we require an approach that returns a set of non-dominated solutions, called the Pareto optimal set, so the decision maker has the option of selecting a solution according to the context.

Formally, a point $x^* \in X$ is Pareto optimal if there does not exist another point, $x \in X$, such that $o_i(x) \leq o_i(x^*)$ for all $i \in \{1, 2, \ldots, I\}$ and $o_i(x) < o_i(x^*)$ for at least one objective function.

Multiple methods have been proposed to solve the multi-objective optimization problem (Gunantara, 2018). Here, a genetic algorithm (GA) is used since it is a generic approach commonly used for the optimization of complex systems because it does not require an analytical model, which in many cases is impractical to obtain. Moreover, we consider a lightweight objective model (instead of a discrete event simulator) which lowers the benefits of constructing an optimization approach that could exploit additional information such as the objective's gradient.

A GA takes a set, or population, of possible solutions and recombines and mutates them iteratively to produce new solutions. At each iteration, each candidate solution, here, equivalent to a policy set $\boldsymbol{X}$, is assigned a fitness value based on the achieved objective function. The best candidates are given a higher probability of being used in the next iteration. In this manner, solutions iteratively "evolve" to generate better candidates until a stopping criterion, e.g., the maximum number of iterations or minimum improvement, is satisfied.

To find an optimal solution, the GA evaluates the quality of all the candidate solutions (policies) by passing each of the policies to the fleet simulator. Figure 2 shows the interaction between the GA
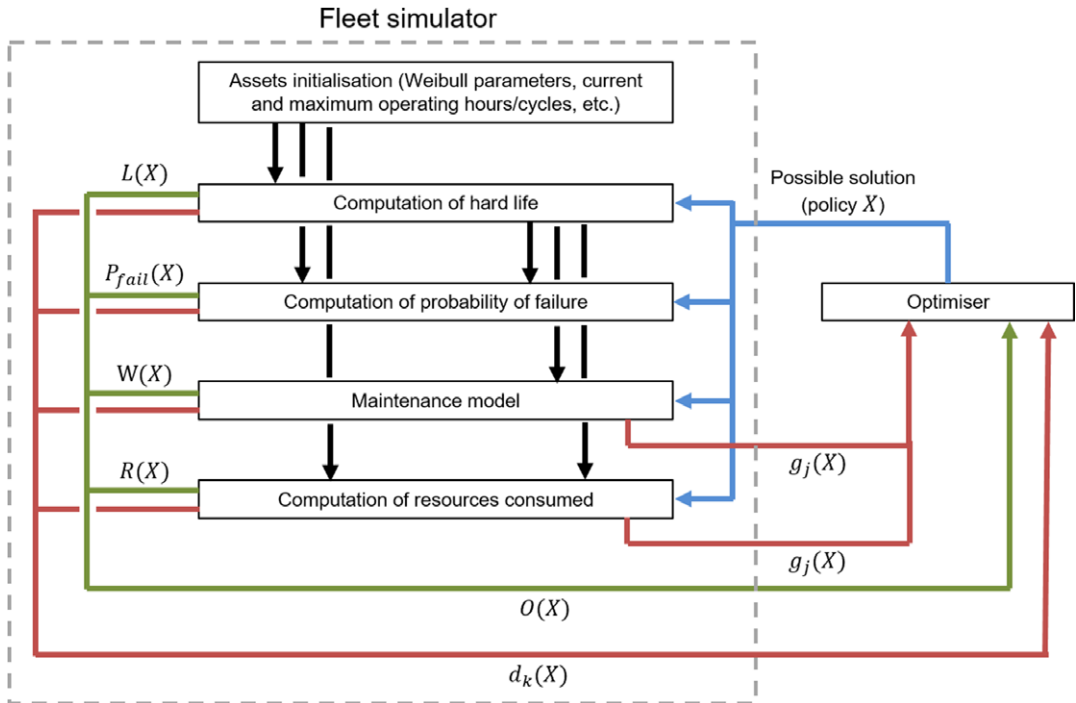
Figure 2. *Interaction between optimizer and fleet simulator. The simulator's modules receive a policy X from the optimizer. The modules return an objective evaluation and constraints that are used by the optimizer to assess the quality of the policy.*

optimizer and the different modules of the fleet simulator. The optimizer passes a candidate solution X to the simulator (blue lines). Each of the modules returns an objective evaluation (green lines) or constraint (red lines) to the optimizer. The optimizer then generates a new set of candidate solutions and the process is repeated. A detailed description of fleet simulator modules is presented in Section 3.

The optimal policies found after the optimization are based on the current condition of the fleet. A re-optimization is therefore required after an event that changes the fleet condition, e.g., asset maintenance or failure. After such an event, a new optimal policy for the whole fleet is computed under the new fleet condition by initializing the assets to reflect the new condition.

3. Asset life modeling

To understand the effects of different policies on the life of the asset, lifing models are required. Two aspects of asset lifing are considered. First, the usage limit of the asset, $T_{L,h}(X)$, which we assume to be a function of the selected policy. Secondly, the asset's probability of failure before its usage limit, $P_{\text{fail}}(X)$, which again is a function of the selected policy. In both cases, the mapping from the policy to the lifing effects is performed using a set of *severity factors*. In the remainder of this section, the estimation of severity factors under different policy choices is first described. This is followed by a description of how the severity factors can be used to evaluate the objectives $L(X)$ and $P_{\text{fail}}(X)$.

3.1. Component life and severity factors

The set of severity factors associated with the h'th asset is defined as $c_h = [c_{1,h}, c_{2,h}, \ldots, c_{N,h}]$, where N is the number of considered components belonging to the asset. The severity factors provide a mechanism to model changes in the life of the assets under different control policies. An overview of the procedure to
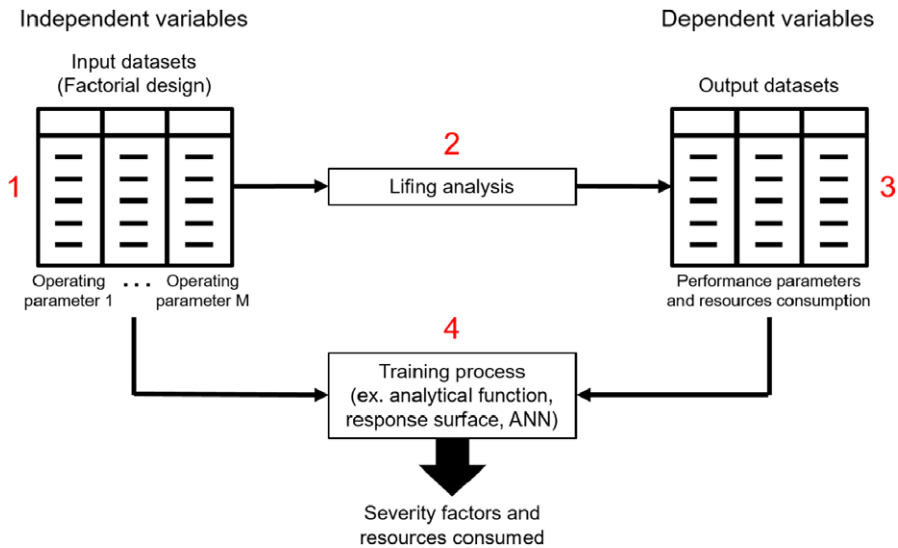
Figure 3. *Computation of analytical models.*

compute the severity factors for different policies is presented below. It outlines a holistic framework that involves the asset performance calculation, models of the asset's components, and computation of the oxidation, creep, and fatigue life.

Models of asset performance are used to simulate the asset performance under different operating conditions and policies. As shown in Figure 3, data from performance simulations under variable operating parameters are combined with thermal and mechanical damage models of individual components to perform a lifing analysis (1 and 2 in Figure 3). The analysis returns the damage suffered by each of the components (3 in Figure 3) which is compared to a reference mission to obtain a severity factor based on the policy used during the asset simulation. Coupling the performance and the lifing analysis process with an optimizer is complex and computationally expensive, if not infeasible in practice. For this reason, a Design of Experiment is performed with the policies as the independent variables, and the output performance and life consumption parameters are fitted to the analytical functions of the independent variables (4 in Figure 3). The data sets generated from the lifing analysis are used within a simple, yet novel, framework for the probabilistic modeling of system failure rates under different policies.

3.2. Time to maintenance model

The lifing characteristics of an asset or its sub-components can be described with a lifetime distribution. This approach is used in reliability engineering to describe the probability of failure, and other derived statistics, over a component's life. The most commonly used lifetime distribution is the Weibull distribution due to its versatility and mathematical tractability (Life Data Analysis, 2024). These distributions can be obtained by fitting a statistical distribution to data collected during the asset's operation under a nominal control policy. Here, Weibull analysis is extended to describe an asset comprised of a system of components with varying initial lives. The concept of component 'severity' is then introduced and it is shown how this acts on each component of the system through the parameters of the Weibull distribution.

As presented in Section 2.1, one optimization objective is to extend the life of the asset. That is, to increase the amount of operating hours available before a maintenance action is required. Maintenance is required in order to replace a subset of usage-limited components that reach the end of their life. Such

components are termed life-limited components, and their life is defined in terms of a certain number of standard operating cycles or hours. The usage limit of the h'th asset must therefore be the minimum component usage limit for the components possessed by the assets, $T_{L,h}^0 = \min\{T_{1,h}^0, T_{2,h}^0, \ldots, T_{N,h}^0\}$, where $T_{i,h}^0$ is the usage limit of the i'th component of the h'th asset. The severity factors are estimated such that they act as a multiplicative constant on the component usage limit, i.e., $T_i(c_{i,h}) = c_{i,h}T_i^0$. The asset usage limit is then given by;

$$T_{L,h}(\mathbf{c}_h) = \min\{c_{1,h}T_{1,h}^0, c_{2,h}T_{2,h}^0, \ldots, c_{N,h}T_{N,h}^0\} \tag{7}$$

3.3. Asset failure model

In this section, an expression for $P_{\text{fail}}(\mathbf{X})$ is derived. Crucially, this formulation does not require Monte-Carlo simulation of a discrete event model in order to evaluate it, rather the probability is computed analytically allowing it to be directly minimized. The lifing characteristics of an asset or sub-component can be described with a lifetime distribution. This approach is used in reliability engineering to describe the probability of failure, and other derived statistics, over a component's life. Weibull distributions are usually obtained by fitting to data collected during the asset's operation or lifing analysis of individual components. Here, Weibull analysis is extended to describe an asset comprised of a system of components with varying initial lives. We make the assumption that component expected life changes depending on the policy and that this can be approximated using severity factors that act on each component through the parameters of the Weibull distribution.

The probability density function of the Weibull distribution for the i'th component in the system is given by:

$$f_i(t) = f_i(t|k_i, \lambda_i) = \frac{k_i}{\lambda_i}\left(\frac{t}{\lambda_i}\right)^{k_i-1} e^{-(t/\lambda_i)_i^k}, \tag{8}$$

for $t \geq 0$, and 0 otherwise, where λ_i and k_i are the scale and shape parameters respectively. t is the time at which the Weibull is evaluated. This could be, for instance, the component life in hours or cycles.

The cumulative distribution function for the i'th components in the system are

$$F_i(z) = F_i(z|k_i, \lambda_i) = \int_0^z f_i(t)\,dt \tag{9}$$

$$= 1 - e^{-(z/\lambda_i)^{k_i}}. \tag{10}$$

It is also useful to define the survival function, $S_i(z) = 1 - F_i(z)$, which is the probability that the component i survives beyond time z.

To consider multiple components with non-zero initial lives, denoting the initial life of the ith component as $z_{i,0}$, then, from Equation 9:

$$F_i^*(z) = \int_{z_{i,0}}^{z+z_{i,0}} \frac{f_i(t)}{\int_{z_{i,0}}^{\inf} f_i(z)\,dz}\,dt \tag{11}$$

$$= 1 - e^{-\left(\frac{z+z_{i,0}}{\lambda_i}\right)^{k_i} + \left(\frac{z_{i,0}}{\lambda_i}\right)^{k_i}}. \tag{12}$$

See Appendix F for a derivation of Equation 12.

Considering an asset comprised of a system of N components with varying initial lives, $z_{i,0}$. The survival function of the system is:

$$S_{sys}(z) = \prod_{i=1}^{N} [1 - F_i(z)] \tag{13}$$

$$= 1 - F_{sys}(z). \tag{14}$$

From Equations 12 and 14 the cumulative distribution of the system is described by:

$$F_{sys}(z) = F_{sys}(z|\boldsymbol{k},\boldsymbol{\lambda}) \tag{15}$$

$$= 1 - e^{\sum_{i=1}^{N} -\left(\frac{z+z_{i,0}}{\lambda_i}\right)^{k_i} + \left(\frac{z_{i,0}}{\lambda_i}\right)^{k_i}}, \tag{16}$$

where $\boldsymbol{k} = [k_1, k_2, \ldots, k_N]$ and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_N]$. The Weibull distribution has the useful property that, for a constant c_i:

$$F_i(z|k_i, \lambda_i/c_i) = F_i(c_i z|k_i, \lambda_i), \tag{17}$$

and this can be trivially extended to the system's cumulative distribution function

$$F_{sys}(z|\boldsymbol{c},\boldsymbol{k},\boldsymbol{\lambda}) = 1 - e^{\sum_{i=1}^{N} -\left(c_i\frac{z+z_{i,0}}{\lambda_i}\right)^{k_i} + \left(c_i\frac{z_{i,0}}{\lambda_i}\right)^{k_i}}, \tag{18}$$

This formulation of the cumulative distribution function provides a useful mechanism to model changes in the life of the asset under operating conditions and control policies that have different effects on the lifing characteristics of individual components. As can be seen from Equation 18, the severity factor acts as a multiplicative constant on the component's expected failure time. It is used as the mechanism to modify the lifetime distributions based on the policy choice. The mapping between the policy and the severity factor is acquired by using first principles physics models of the component and is discussed in the next section.

By modifying the Weibull distribution based on a policy, i.e., applying different severity factors $\boldsymbol{c}$, the probability of failure for different policies can be computed. When rare and stochastic events associated with a low probability of failure are considered as objectives in an optimization problem, evaluating the effects of different policies can be challenging and might require thousands of simulations (Chen and Guo, 2007). The computation of the probability of failure presented in this section is crucial to reducing the time required to solve the fleet optimization problem.

3.4. Maintenance model

A model of the maintenance process is required to estimate which components are replaced during maintenance based on the asset's state. This procedure is entirely problem-specific and will depend on the cost of replacing components, the remaining life of all other components, the cost associated with workshop occupancy, and so forth The maintenance model also estimates the time required to repair the asset based on the state of the asset. This model is required to generate maintenance-based constraints as presented in Figure 2 and demonstrated in the next section.

4. Case study and results

The proposed approach is demonstrated using a real-world case study in which economic optimization for a fleet of aerospace gas turbine engines is performed. Gas turbine engines have many control

variables and configurations that can be modified to affect their performance. They are subject to many constraints and external stimuli at the asset and fleet level, e.g., flight routes, contractually obliged maintenance, engine availability, and so forth By finding optimal control policies, that are specific to fleet operating conditions, we show that we can improve asset availability, i.e., reducing the time that assets are not available for use, while simultaneously reducing fuel burn.

The high-pressure turbine (HPT) of the modeled aerospace gas turbine engine (GTE) in this case study is temperature-regulated with independent cooling air valves. The setting of the cooling air valves are control variables that modulate the HPT blade cooling and, therefore, the HPT blade temperature. Increasing the level of cooling air decreases the blade temperature which reduces degradation through effects such as oxidation and creep. However, it requires air to be diverted from the main gas path causing an increase in gas path temperature and fuel burn to achieve the required thrust and therefore a decrease in the fuel efficiency. In the engine performance model, the cooling flow is simulated as high-pressure compressor (HPC) bleed air being mixed at the HPT inlet (for external cooling) and at the HPT outlet (for internal blade cooling) as shown in Figure 4. The amount of the modulated HPT internal blade cooling is then used in the blade temperature calculations and life consumption estimations as described by Pontika et al. (2023). The aim of this study is to find an optimal policy that defines a setting for the cooling air control variables across a fleet of engines. The level of cooling will affect the lives of the components of the HPT and the intermediate pressure turbine (IPT) differently depending on their location, material composition, etc., as well as the flight phase, e.g., take-off, climb, and cruise. Other components without active cooling, e.g., IPT blades, will be adversely affected by the rise in gas path temperature caused by the additional fuel burn. Therefore, a component-level model is important to assess the lifing impact of control policies with optimization results depending on the relative initial lives of all components. The optimizer finds a policy, for each engine, that sets control variables differently for each phase. Hence, $X \in \mathbb{Z}^{M \times H}$ with $M = 3$. For this case, discrete decision variables are considered. A case with continuous decision variables is presented later.
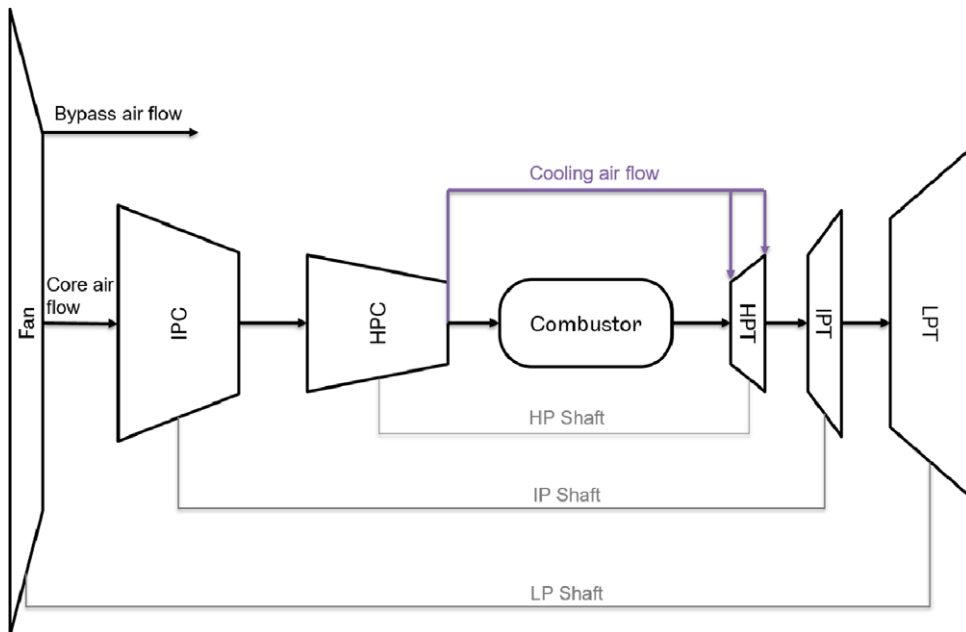

Figure 4. *Illustration of cooling flow applied to the high-pressure turbine in the engine performance model. Air from the high-pressure compressor (HPC) is diverted to the high-pressure turbine (HPT).*

4.1. GTE modeling

The engine performance modeling under variable policies is performed in Turbomatch (Giannakakis et al., 2015; Macmillan, 1974), an in-house component-based gas turbine performance simulation tool at Cranfield University. The aircraft performance and flight analysis, including the mission fuel burn, is performed in Hermes (Laskaridis et al., 2005), another in-house aircraft performance tool, which is coupled with Turbomatch to obtain the engine performance analysis. A generic three-spool turbofan engine model and a typical long-range wide-body aircraft model are considered. The Larson-Miller Parameter (Schneider et al., 1992) and Basquin/Coffing-Manson (Fuchs, 1981) models together with the Arrhenius equation (Meier et al., 1991) are used to calculate the effects of policies on the engine components' lives and estimate the severity factors. An overview of the process is shown in Figure 5A and the interested reader is referred to Pontika et al. (2023) for a detailed description. The result of the modeling is a mapping between the policy and the severity factor and fuel burn in the form of a look-up table, graphed in Figure 5B. This mapping shows the sensitivity of the fuel burn and HPT/IPT severities to the different control variables. For example, it can be seen that the fuel consumption is highly sensitive to the cooling flow during cruise and relatively insensitive during take-off and climb.

4.2. Life of components and fuel trade-off

There is a trade-off between fuel consumption and the component life usage. Therefore, a Pareto front can be found from which a user can select a desirable policy for each of the engines. A fleet-level objective function is chosen which consists of the mean probability of failure (before maintenance) and mean fuel consumption across the fleet:

$$R(\boldsymbol{X}) = \frac{1}{H}\sum_{h=1}^{H}\sum_{m=1}^{M} FC_m(\boldsymbol{c}_h), \tag{19}$$

$$L(\boldsymbol{X}) = 0, \tag{20}$$

$$P_{\text{fail}}(\boldsymbol{X}) = \frac{1}{H}\sum_{h=1}^{H} F_{sys}\big(T_{L,h}(\boldsymbol{c}_h)|\boldsymbol{c}_h,\boldsymbol{k}_h,\boldsymbol{\lambda}_h\big) \tag{21}$$

$$W(\boldsymbol{X}) = 0, \tag{22}$$

where $H = 50$ and the function FC_m is a function that estimates the fuel consumption over the m'th flight phase under policy $\boldsymbol{x}_h$ obtained through the framework presented in Section 3.1 and with the relationship shown in Figure 5B. The severity factors $\boldsymbol{c}_h$ and hence the asset's hard lives $T_{L,h}(\boldsymbol{c}_h)$ are implicitly a function of the policy $\boldsymbol{x}_h$. The probability of failure $P_{\text{fail}}(\boldsymbol{X})$ is derived from Equation 18. Reducing the probability of engine maintenance through $P_{\text{fail}}(\boldsymbol{X})$ also reduces the cost associated with wasted life components replaced during maintenance. For simplicity, we therefore choose to exclude $L(\boldsymbol{X})$ from the cost. Objectives based on maintenance are considered in the next section. These objectives are computed by using the fleet simulator presented in Section 2.2. The fleet simulator was validated against a proprietary agent-based discrete event simulator provided by our industrial collaborator.

The agent-based discrete event simulator is tuned heuristically by subject matter experts with real data from gas turbine engines. A high-level description is now presented. The simulator starts by initializing every asset of the fleet. This initialization considers the type of components, e.g., model, of each asset and their initial life. After the initialization, information about the control policy to be applied and the conditions in which each asset is going to operate, e.g., route, is used to compute the maximum number of hours before maintanance (hard life). Then, given an operating schedule for each asset, e.g., number of
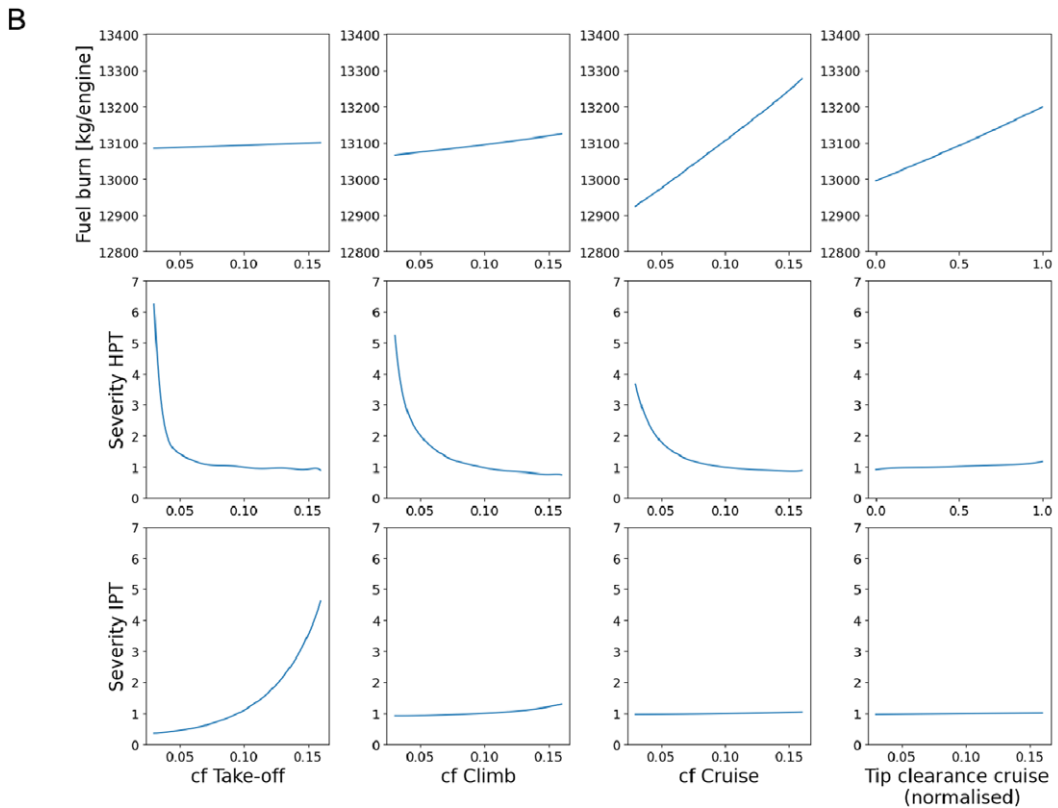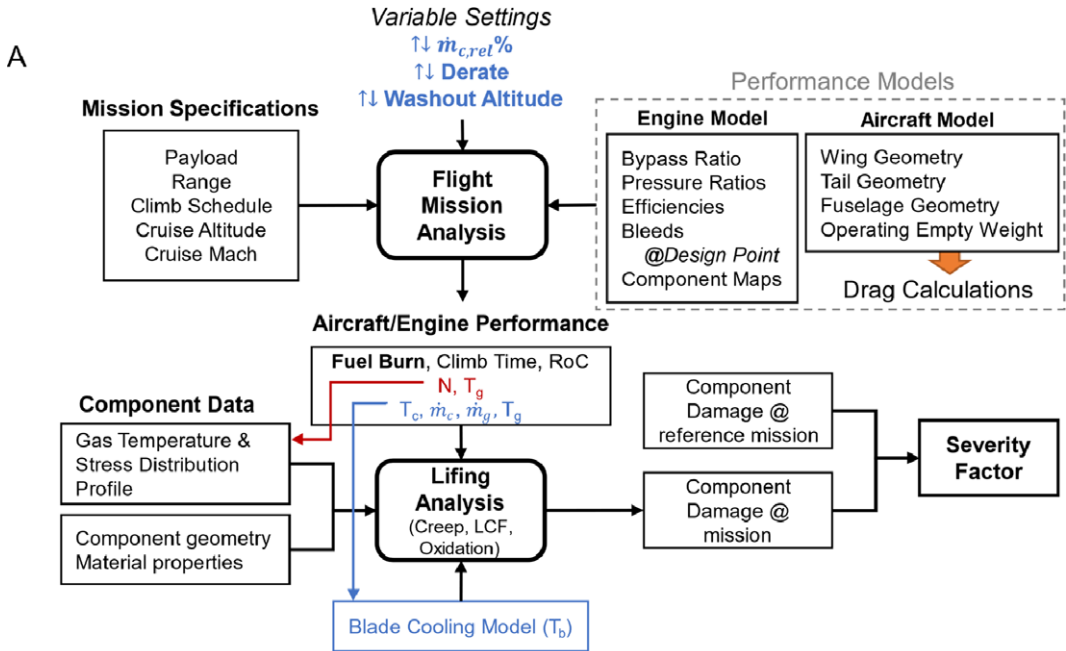
Figure 5. *Estimation of severity factors. (A) Framework for performance and lifing analysis of future advanced aerospace engines. (B) Effects of cooling flow (cf) and tip clearance change at take-off, climb, and cruise.*

hours to be operated, the life of the assets is simulated until maintance is required. The resources consumed until that point and the time required to repair the asset are computed. During maintenance, the model of the components can change based on an scheduled campaign. Therefore, the simulator updates the information of each repaired asset before repeating the process described above.

Our simulator, used for optimization problems in this section, runs 600 times faster than the proprietary software, allowing optimization within a reasonable time frame. The Python toolbox Pymoo (Blank and Deb, 2020) was used to perform the optimization. Specifically, the non-dominated sorting GA NSGA-III (Deb and Jain, 2013), which is efficient for many-objective optimization problems, was used. The solution is found in approximately 340 seconds on a laptop with a 2.30 GHz processor i7 and 16 GB of RAM. The stopping criterion is based on a threshold for objective improvement.

In the obtained Pareto front, each point represents a policy set for the whole fleet, i.e., a possibly different policy for each engine and each flight phase (take-off, climb, and cruise), see Figure 6A. In the Pareto front, the left-most and right-most policy sets are the solutions to the optimization problems with a single objective.

A user can select a single policy set, based on conditions such as fuel price, from the Pareto optimal policies returned by the optimizer. For instance, a policy that reduces the probability of failure at the cost of fuel consumption can be chosen when the fuel price is low. To demonstrate the effects of applying different policies, two policy sets are selected from the Pareto front. Policy set 1 is selected to reduce the fuel consumption across the fleet while policy set 2 is selected to reduce the probability of failure. The cooling flow levels for the selected policies are shown in Figure 6B. It is possible to see that, in general, higher levels of cooling are applied with policy set 2 to reduce the probability of failure. For example, for engine number 1, cooling flow is applied for only two phases (two bars) while for policy set 2 cooling flow is applied during every phase. The additional cooling flow has a negative effect on the fuel consumption, i.e., higher fuel consumption. To see how the selected policies affect the life of the engines, the policies are applied to the fleet in a Monte-Carlo simulation under the initial conditions used for the the optimization. The distribution of the number of engines that fail before their usage limit for each policy set is shown in
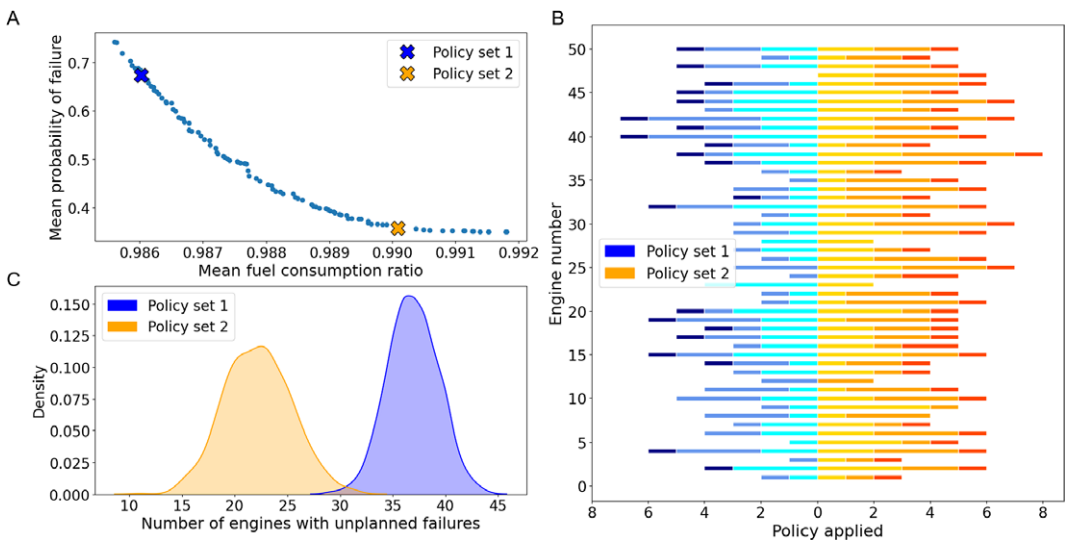

Figure 6. *(A) Pareto front returned by the optimizer. A single solution can be selected by a user based on the current conditions. Policy set 1 can be selected to save fuel while policy set 2 reduces the probability of failure at the cost of higher fuel consumption. (B) Cooling flow levels applied to each engine. The stacked bars show the cooling flow level applied to each of the flight phases, i.e., take-off (light blue/orange), climb (blue/orange), and cruise (dark blue/orange). C) Distribution of unplanned failures after applying policies 1 and 2.*

Figure 6C. As expected, applying policy 2 reduces the number of unplanned failures at the cost of higher fuel consumption.

Note that because there are no constraints, the mean of the solutions obtained by solving the multi-objective optimization of each engine individually lies in the Pareto front presented in 6A. However, this is not true once constraints that affect the fleet are added to the problem as in the next section.

4.3. Maintenance scheduling

The results presented in the previous section do not consider any constraints. Besides constraints on the asset-level probability of failure and fuel consumption, constraints on the number of engines in the workshop can be considered with the proposed approach. These constraints can be used, for instance, to limit the simultaneous number of engines in the workshop (to conform to workshop capacity) or to increase the availability of engines during busy periods.

Recall that an engine is sent to the workshop when it reaches its usage limit, i.e., a maximum number of cycles. As presented in Section 3.2, some policies such as the cooling air can affect the component usage limit. Hence, it is possible to find policies where the number of engines in the workshop, within a period of time, is constrained. To satisfy such a constraint, the date when each engine would reach its usage limit, based on a policy, is computed by considering the flight schedules, i.e., the number of planned operating hours. The workshop model predicts the time required to repair the engine. This computation is performed during the fleet optimization cycle presented in Section 2.2. For this example, we aim to set the maximum number of engines in the workshop during July and August to zero, i.e., $g_1(X) = S(X) = 0$, where $S(X)$ is a function that returns the number of engines in a shop in this period. However, a better approach is to treat this constraint as an additional objective during the optimization to increase the number of feasible solutions. That is, $W(X) = S(X)$. This allows more solutions to be considered by relaxing the workshop requirement and increases the likelihood of finding feasible solutions. In addition, the probability of failure during the busy period can also be added as an objective to reduce the probability of unplanned maintenance during these months, $\widehat{P}_{\text{fail}} = P(T_1 < T_{\text{fail},h} < T_2)$, where T_1 and T_2 are the time points at the start and end of this period, i.e., start July and end August. Hence, in this example, we minimize $O(X) = \left(R, P_{\text{fail}}, W, \widehat{P}_{\text{fail}}\right)$. Note that for this computation, the initial operating hours are considered from the beginning of the considered period, i.e., July.

To present the evaluated objectives for all the policies found during the constrained optimization, we use a parallel coordinate plot, see Figure 7A. In this plot, each color represents a policy set, i.e., a point in the 2D Pareto presented in Figure 6A. As in the previous case, the trade-off between probabilities of failure and fuel consumption can be observed. The result shows that it is not always possible to find optimal policies with zero maintenance during the busy period. Given these policy sets, the decision maker can select a set with an acceptable level of assets in maintenance if necessary.

Figure 7B, C shows the number of engines in the workshop on different dates for two selected policies with similar fuel consumption, one for the unconstrained solution and the other for the constrained one. In the constrained solution, the optimizer returns policies that avoid the maintenance of engines during summer, see Figure 7D.

It is not always possible to find optimal policies with zero planned maintenance during a specific period, see Figure 7. To improve the flexibility of the obtained policies, an extra decision variable, x_e, representing early maintenance can be added to the optimization problem. By sending an engine to the workshop before its scheduled maintenance, based on the new decision variable, policies that satisfy tighter constraints can be found. Sending an engine to the workshop before its scheduled maintenance incurs an economic cost through, for example, wasted component life. For simplicity, we do not consider a specific cost, instead, the mean time before maintenance (in months) is considered in the cost function. This objective can be defined as $l(x_h) = x_{h,e}$. Because the additional decision variable increases the number of policies where the number of engines in shop during the busy period is zero, we remove the objective $W(X)$ and add the constraint $g_1(X) = S(X) = 0$. Hence, the objectives considered in this example are $O(X) = \left(R, L, P_{\text{fail}}, \widehat{P}_{\text{fail}}\right)$. Figure 8 shows the parallel coordinate of each solution obtained
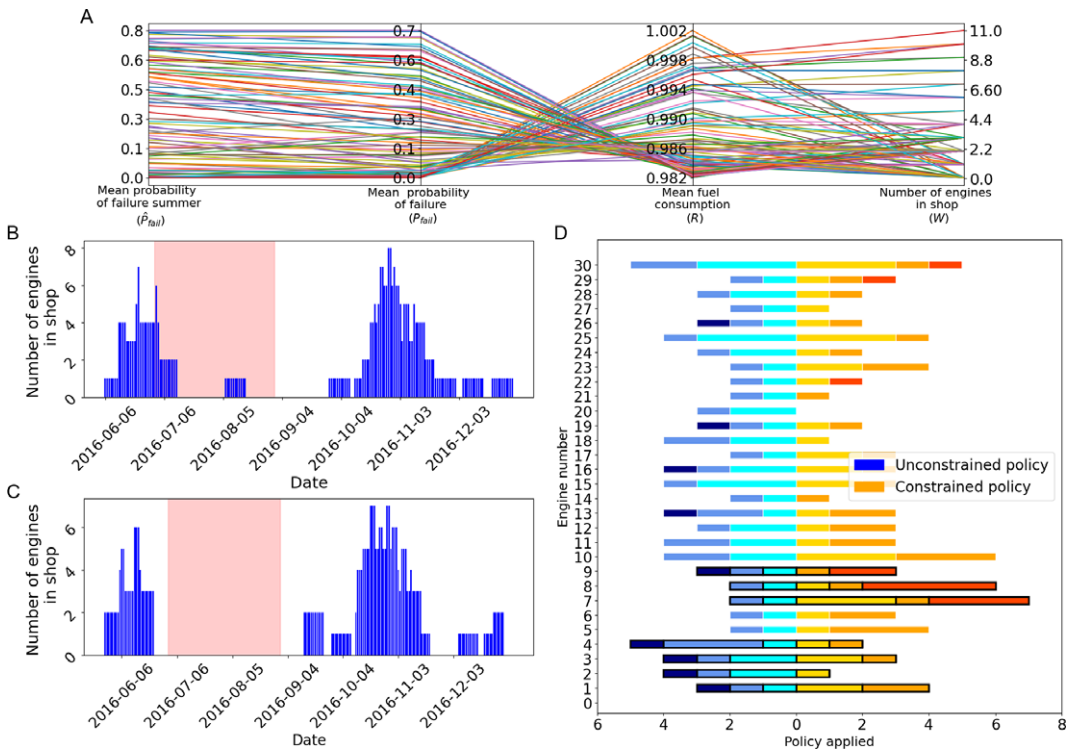
Figure 7. *(A) Parallel coordinate plot of Pareto front. (B) Maintenance dates for an optimal policy without constraints. (C) Maintenance of engines with a constraint on the number of engines in the workshop. Maintenance during the busy period (highlighted region) is avoided by repairing four engines early and three engines after this period. (D) Cooling flow level applied to each engine. The policies of the engines that are repaired during the busy period in the unconstrained problem are highlighted.*

after the optimization. It is possible to see that, in contrast to the previous results, a richer set of policies that meet the condition of zero maintenance during the busy period is obtained.

The result in this section demonstrates that the proposed approach can be used for maintenance scheduling in two manners. Firstly, to find optimal policies given a predefined maintenance preference as demonstrated above. Secondly, to plan the maintenance after selecting a policy obtained from an unconstrained optimization.

4.4. Tip clearance control

This section presents how the proposed framework can be used for different control variables. Specifically, tip clearance is considered in addition to the cooling flow presented in the previous sections. Tip clearance is the radial distance between the blade tip of a turbine and the casing, see Figure 9. Increasing the tip clearance directly decreases the efficiency of the engine Holeski and Futral (1969) resulting in higher fuel consumption. This is because the power that a turbine can provide depends on the airflow flow through the area where the blade is located. In the presented case study, the relative change in tip clearance is translated into a relative change in component efficiency using the empirical correlation provided by Kurzke and Halliwell (2018) for shrouded turbine blades.

A small tip clearance is desirable to improve the engine efficiency but increases the risk of the turbine blade making contact with, and abrading, the case lining (shroud), named a tip rub event. The variable conditions under which the engine operates, e.g., variable temperatures, pressures, and vibration change the gap between the blade tip and the shroud (clearance) which changes the probability of a tip rub
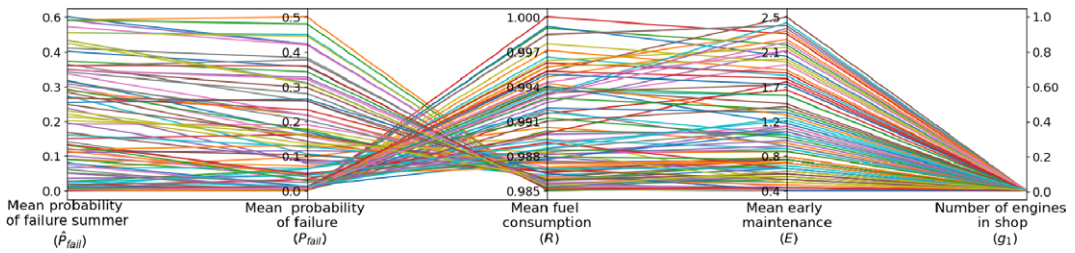
Figure 8. *Parallel coordinate plot of Pareto front obtained with an extra decision variable representing early maintenance. A rich set of policies that meet the condition of zero maintenance during the busy period are obtained.*
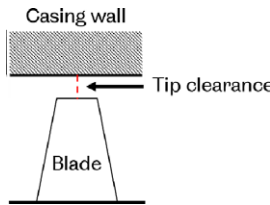

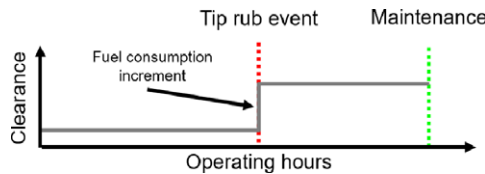***Figure 9.*** *Illustration of the blade–casing clearance.*

Figure 10. *Illustration of the effect of tip rub event. The gap between the blades and the casing increases after a tip rub until the engine is repaired. This results in an increment in fuel consumption.*

occurring. The abrasion of the shroud after the tip rub event permanently increases the tip clearance until the engine is repaired, see Figure 10. Hence, an engine will have a higher fuel consumption following a tip rub event. This suggests that there may be a benefit in applying a smaller tip clearance when the engine is nearing maintenance as the future loss of fuel efficiency is less. Tip clearance can be modulated by cooling the case, again using cooling air which is controlled using the turbine case cooling valve. Cooling the case causes it to contract which decreases the tip clearance. The tip clearance is controlled in a closed loop, the target clearance is a further control variable that can be added to the policy.

In addition to affecting fuel consumption, the tip clearance also affects the probability of failure. If an engine runs less efficiently, more power, and as a consequence, more heat is generated to reach the required operation point. Therefore, the life of the engine components is reduced. The gas turbine modeling presented in Section 4.1 was used to estimate the effects of the tip clearance in cruise and cooling flow in the fuel consumption and severity factors of the HPT and IPT. Given this model, three different problems were solved:

1. The decision variable of the h'th engine, x_h, is defined as the change or delta in the tip clearance (compared to the tip clearance without control).
2. The decision variable $\boldsymbol{x}_h = [x_{h,1}, x_{h,2}, x_{h,3}]$ is defined as the cooling flow for each of the flight phases. This is equivalent to the problem solved in Section 4.2 but with continuous decision variables and the severity factor models obtained by modeling the effects of tip clearance and cooling flow.

3. The decision variable $x_h = [x_{h,1}, x_{h,2}, \cdots, x_{h,4}]$ is defined as the delta of the tip clearance and the cooling flow for each of the flight phases.

For these problems, a fleet of 20 engines was considered. Moreover, no constraints were considered and the objection function presented in Section 4.2 was minimized.

   The solution to the first problem is presented in Figure 11. As shown in Figure 11A, the tip clearance delta applied to the engine has a direct correlation with the time to maintenance. As presented above, this result is expected because the effect of a tip rub is minimized when the engine is closed to be repaired. Figure 11B shows how the fuel consumption is reduced as the tip clearance delta increases until it reaches a point where a tip rub happens. Because the tip clearance increases after this point, the efficiency of the engine is affected. This results in a higher fuel consumption. Because of the correlation between tip clearance and hours to maintenance, for this problem, a single optimal solution can be found.
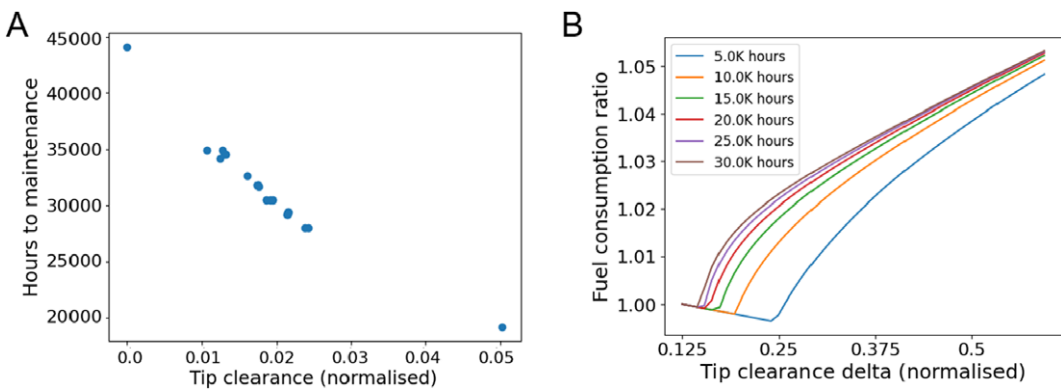

Figure 11. *A: Tip clearance delta applied to the engines. B: Total fuel consumption before maintenance for different tip clearance policies.*
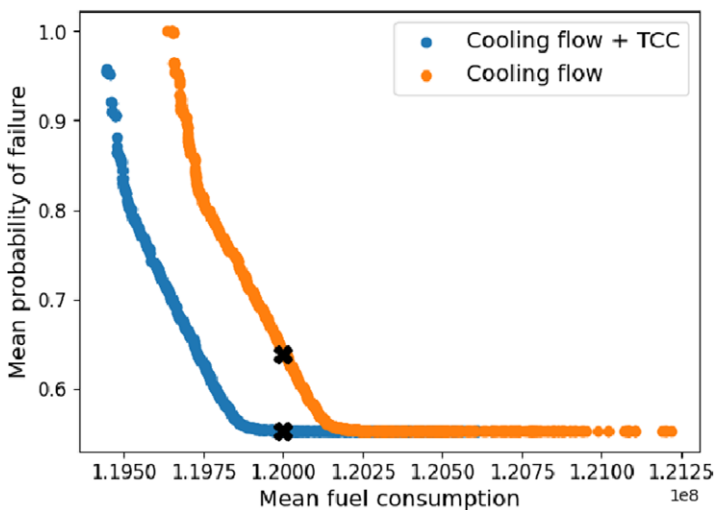

Figure 12. *Pareto front returned by optimizer after solving the problem with tip clearance delta as decision variable and tip clearance with cooling flow, respectively. Two policy sets with similar fuel consumption are selected for comparison.*
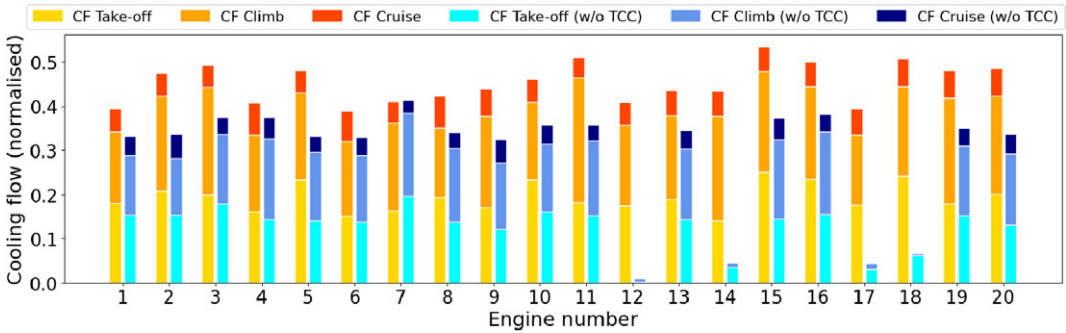
Figure 13. *Cooling flow policies applied to engines based on the the policy sets selected in 12.*
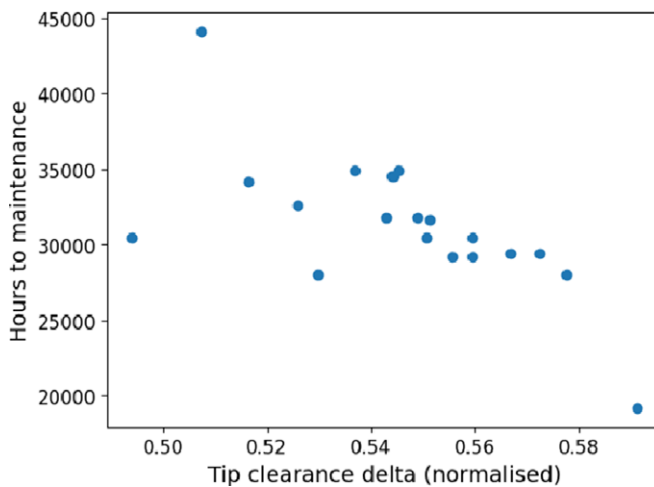

Figure 14. *Tip clearance delta applied to the engines.*

The Pareto fronts for the second and third problems are presented in Figure 12. It is possible to see that the addition of an extra degree of freedom improves the solution obtained, i.e., a lower probability of failure and fuel consumption can be achieved.

From these Pareto fronts, two policy sets are selected for comparison. This comparison can be seen in Figure 13. This figure shows that, in general, more cooling flow is applied to the engines when the tip clearance is also controlled. From Figure 5, it can be seen that as the cooling flow applied to the engine increases the fuel consumption also increases. On the other hand, the fuel consumption is reduced when a small tip clearance is applied. Therefore, by running an engine more efficiently (and reducing fuel consumption) by applying a tip clearance policy, more cooling flow can be applied without having high levels of fuel consumption. Therefore, more cooling flow can be applied compared to the case where the tip clearance is not changed.

When cooling flow and tip clearance are considered as a control policy, the clear correlation between hours to maintenance and tip clearance disappears, see Figure 14. Therefore, selecting an optimal policy without solving a multi-objective optimization problem is not trivial. Understanding what factors, e.g., route and component's model, determine the optimal policies is future work.

5. Conclusions

Fleet management has been widely studied in different settings due to its potential to reduce both operating costs and emissions while also increasing sustainability. This paper presents a novel framework

for multi-objective fleet optimization to find optimal control policies with the objective of increasing the life of the fleet's assets and/or minimizing the resources required for the fleet operation. The proposed framework consists of engine lifing models for mapping control policies to the life of engine components and a fleet model developed to reduce the number of simulations required to evaluate candidate solutions. Results show that based on the selected policy, the probability of unplanned maintenance or fuel consumption across the fleet can be reduced. Moreover, they show how, by considering constraints such as the maximum number of assets in maintenance during busy periods, the framework can be used to schedule the maintenance of the fleet assets. Different types of control policies as well as different optimization scenarios and constraints are considered in the examples to demonstrate the framework flexibility.

References

Biell M, **Bielli A and Rossi R** (2011) Trends in models and algorithms for fleet management. *Procedia-Social and Behavioral Sciences 20*, 4–18.

Blank J and Deb K (2020) pymoo: Multi-objective optimization in python. *IEEE Access 8*, 89497–89509.

Bruglieri M, **Mancini S**, **Pezzella F and Pisacane O** (2019) A path-based solution approach for the green vehicle routing problem. *Computers & Operations Research 103*, 109–122.

Chen CP and Guo T-H (2007) Design of intelligent acceleration schedules for extending the life of aircraft engines. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37*(5), 1005–1015.

Deb K and Jain H (2013) An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints. *IEEE Transactions on Evolutionary Computation 18*(4), 577–601.

Fuchs HO (1981) *Metal Fatigue in Engineering*. Wiley, *103*.

Gavranis A and Kozanidis G (2015) An exact solution algorithm for maximizing the fleet availability of a unit of aircraft subject to flight and maintenance requirements. *European Journal of Operational Research 242*(2), 631–643.

Giannakakis P, **Laskaridis P**, **Nikolaidis T and Kalfas AI** (2015) Toward a scalable propeller performance map. *Journal of Propulsion and Power 31*(4), 1073–1082.

Gunantara N (2018) A review of multi-objective optimization: Methods and its applications. *Cogent Engineering 5*(1), 1502242.

Holeski DE and Futral SM (1969) *Effect of Rotor Tip Clearance on the Performance of A 5-Inch Single Stage Axial Flow Turbine*. National Aeronautics and Space Administration.

Kraft J and Kuntzagk S (2017) Engine fleet-management: The use of digital twins from a mro perspective. In *Turbo Expo: Power for Land, Sea, and Air*, vol. *50770*. American Society of Mechanical Engineers, pp. V001T01A007.

Kurzke J and Halliwell I (2018) *Propulsion and Power: An Exploration of Gas Turbine Performance Modeling*. Springer.

Laskaridis P, **Pilidis P and Kotsiopoulos P** (2005) An integrated engine-aircraft performance platform for assessing new technologies in aeronautics. In *International Society for air-Breathing Engines Conference*.

Life Data Analysis (2024). Available at https://www.weibull.com/basics/lifedata.htm.(accessed 17 January 2024).

List GF, **Wood B**, **Nozick LK**, **Turnquist MA**, **Jones DA**, **Kjeldgaard EA and Lawton CR** (2003) Robust optimization for fleet planning under uncertainty. *Transportation Research Part E: Logistics and Transportation Review 39*(3), 209–227.

Macmillan W (1974) *Development of a Modular-Type Computer Program for the Calculation of Gas Turbine Off-Design Performance*. PhD Thesis, Cranfield University.

Meier SM, **Nissley DM**, **Sheffler KD and Cruse TA** (1991) Thermal barrier coating life prediction model development. In *Turbo Expo: Power for Land, Sea, and Air*, vol. *79023*. American Society of Mechanical Engineers, p. V005T13A003.

Militão AM and Tirachini A (2021) Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. *Transportation Research Part A: Policy and Practice 151*, 52–80.

Pontika E, **Laskaridis P**, **Montana Gonzalez F**, **Jacobs W and Mills A** (2023) Mapping the effect of variable HPT blade cooling on fuel burn, engine life and emissions for fleet optimization using active control. In *AIAA SCITECH 2023 Forum*, p. 0706.

Sanchez DT, **Boyacı B and Zografos KG** (2020) An optimisation framework for airline fleet maintenance scheduling with tail assignment considerations. *Transportation Research Part B: Methodological 133*, 142–164.

Sawik B, **Faulin J and Pérez-Bernabeu E** (2017) Multi-criteria optimization for fleet size with environmental aspects. *Transportation Research Procedia 27*, 61–68.

Schneider W, **Mughrabi H and Hammer J** (1992) Creep deformation and rupture behavior of the monocrystalline superalloy cmsx-4–a comparison with the alloy srr 99. *Superalloys 26*(8).

Sheng D, **Meng Q and Li Z-C** (2019) Optimal vessel speed and fleet size for industrial shipping services under the emission control area regulation. *Transportation Research Part C: Emerging Technologies 105*, 37–53.

Turan HH, **Elsawah S and Ryan MJ** (2020) A long-term fleet renewal problem under uncertainty: A simulation-based optimization approach. *Expert Systems with Applications 145*, 113158.

Vos K, **Peng Z**, **Lee E and Wang W** (2023) Aircraft fleet availability optimisation: a reinforcement learning approach. *The Aeronautical Journal 127*(1318), 2204–2218.

Yıldırım Ş and Yıldız B (2021) Electric bus fleet composition and scheduling. *Transportation Research Part C: Emerging Technologies 129*, 103197.

Zheng Q, **Zhang H**, **Miao L and Sun F** (2017) On-board real-time optimization control for turbo-fan engine life extending. *International Journal of Turbo & Jet-Engines 34*(4), 321–332.

Appendix 1

Derivation of Equation 12

$$F_0 = \int_{z_{i,0}}^{\infty} f_i(z)\, dz$$
$$= \int_{z_{i,0}}^{\infty} \left(\frac{z}{\lambda}\right)^{k-1} e^{-\left(\frac{z}{\lambda}\right)^k} dz$$

making the substitution

$$u = \left(\frac{z}{\lambda}\right)^k, \quad \frac{du}{dz} = \frac{k}{\lambda^k} z^{k-1} = \frac{k}{\lambda}\left(\frac{z}{\lambda}\right)^{k-1}$$
$$F_0 = \int_{\left(\frac{z_{i,0}}{\lambda}\right)^k}^{\infty} e^{-u}\, du \tag{23}$$
$$= \left[-e^{-u}\right]_{\left(\frac{z_{i,0}}{\lambda}\right)^k}^{\infty}$$
$$= e^{-\left(\frac{z_{i,0}}{\lambda}\right)^k}$$

$$F_i^*(z) = \int_{z_{i,0}}^{z+z_{i,0}} \frac{f_i(t)}{\int_{z_{i,0}}^{\infty} f_i(z)\, dz}\, dt$$
$$= \frac{1}{F_0} \int_{z_{i,0}}^{z+z_{i,0}} f_i(t)\, dt$$

making the same substitution given in Equation 23

$$F_i^*(z) = e^{\left(\frac{z_{i,0}}{\lambda}\right)^k} \left[-e^{-u}\right]_{\left(\frac{z_{i,0}}{\lambda}\right)^k}^{\left(\frac{z+z_{i,0}}{\lambda}\right)^k}$$
$$= e^{\left(\frac{z_{i,0}}{\lambda}\right)^k} \left[-e^{-\left(\frac{z+z_{i,0}}{\lambda}\right)^k} + e^{-\left(\frac{+z_{i,0}}{\lambda}\right)^k}\right]$$
$$= 1 - e^{-\left(\frac{z+z_{0,i}}{\lambda_i}\right)^{k_i} + \left(\frac{z_{0,i}}{\lambda_i}\right)^{k_i}}$$