Semmel, A., Heese, S., & McConnell, B.M. 2025. Evaluating the Implementation of Operational Readiness and Maintenance Policies in US Army Aviation, Accepted at *Journal of Defense Modeling and Simulation*, article in press.

Evaluating the Implementation of Operational Readiness and Maintenance Policies in US Army Aviation

Austin D. Semmel[1,3,5], Hans Sebastian Heese[2], and Brandon M. McConnell[3,4]

[1]**Operations Research Graduate Program, North Carolina State University, adsemmel@ncsu.edu**
[2]**Poole College of Management, North Carolina State University**
[3]**Military Operations Research Group, North Carolina State University, Raleigh, NC**
[4]**Center for Additive Manufacturing and Logistics (CAMAL), North Carolina State University**
[5]**Corresponding Author**

ABSTRACT

This study examines AH-64 Apache dispatch decisions to assess the implementation of Operational Readiness (OR) and maintenance policies in the US Army. Current policies are designed to promote a ready and flexible force that is prepared to respond to global force projection requirements. The Army dictates a 75% OR target for aviation equipment and urges units to utilize aircraft uniformly to distribute maintenance capacity and prevent backlog. Given these objectives, we would expect a reduced OR rating to compel fewer sorties and uniformly distributed flying hours over the phase maintenance horizon. However, using a Generalized Additive Model (GAM), findings indicate that diminished OR does not deter flight operations. Moreover, aircraft are more likely to be grounded when approaching scheduled phase maintenance. Further analysis exposes a significant interaction effect; units place greater weight on an aircraft's hours until phase maintenance in the presence of low OR, highlighting a potential risk aversion in decision-making. Interestingly, control variables (the day of the week and reporting period proximity) highly correlate with flight decisions. The findings suggest that current aviation readiness metrics may have an unintended influence on units' resource allocation. Future research should investigate unit-specific decision-making frameworks to improve aviation maintenance and OR efficiency.

Keywords: Aviation Maintenance, Operational Readiness, Army Aviation Policy, Generalized Additive Model

1 INTRODUCTION AND MOTIVATION

The US Army maintains a globally-deployable ground force, capable of rapid power projection in order to deter, fight, and win its nation's wars. In order to do so, it seeks to efficiently allocate resources in a manner that sustainably maximizes the Army's lethality over time. A lethal force requires a combination of trained soldiers and functional weapon systems. However, an inherent trade-off exists between training and equipment serviceability. The more a unit trains, the more its equipment is used, fails and requires maintenance. On a daily basis, commanders must balance their internal and external training and mission requirements with a maintenance plan designed for sustained operations. While armor and mechanized infantry units certainly boast complex pieces of equipment, few Army units feel the sting of a maintenance mishap or the dread of a supply backlog more than an attack reconnaissance battalion of Apache helicopters. Thus, the Army designs policies and regulations that prescribe minimum levels of expected equipment serviceability and training to aid commanders in navigating these trade-offs.

This paper aims to investigate the behavior of units operating under current policies and uncover decision-making patterns surrounding the practical implementation of their flying hour programs. Army commanders are evaluated based on their ability to effectively train their units while maintaining readiness levels prescribed by the Headquarters, Department of the Army [1]. The Army refers to a unit's equipment serviceability as its "Operational Readiness" (OR) [1, paragraph 5-5]. Broadly speaking, each Active Duty AH-64 Apache pilot is required to fly 70 hours semi-annually with a unit goal of achieving at least 75% OR at all times [2]. Thus, OR and training are naturally competing objectives. Commanders and their representatives are tasked with making daily decisions on aircraft utilization while considering a unit's mission requirements, OR, maintenance schedules, and other factors (such as the weather and budget). This decision-making landscape is the subject of our investigation. We analyze the decision to fly a capable AH-64 Apache helicopter on a given day using a generalized additive model (GAM) to investigate the impact of these factors on a unit's aircraft deployment decisions.

One of the covariates in our model is a unit's current OR rating. As an example, if one helicopter is Fully Mission-Capable (FMC) to a unit for 24 out of 30 days in a reporting period (reported as 576 of 720 hours), then this piece of equipment maintained an 80% OR rating for that specific reporting period. This process is averaged across all equipment items of that type in a given unit. The Army establishes thresholds denoted as R-levels to measure a unit's ability to maintain their on-hand equipment; per Table 5-3 in HQDA [1], a unit achieves level R-1, if its OR rating is above 75%, R-2, if it is above 60%, R-3, if it is at least 50%, and R-4 otherwise.

Strict standards dictate if a helicopter is FMC or assigned to another maintenance category at a given time. A piece of equipment receives the maintenance status FMC if it is fully operational, configured in a safe and proper manner as designated by the US Army, and able to perform its combat mission without endangering the lives of the crew or operators. In contrast, an aircraft designated as Not Mission-Capable (NMC) compromises unit effectiveness by limiting both training and operational capabilities [3]. As such, commanders have a strong incentive to ensure their units maintain an R-1 status when possible.

Due to helicopters' complex maintenance needs, achieving R-1 status is not always feasible. They are technical pieces of equipment that require regular maintenance intervals and also experience individual sub-component failures that demand immediate unplanned maintenance in order to regain FMC status. Planned (or, scheduled) maintenance intervals are called Phase Maintenance Cycles and occur at prescribed times according to the number of flying hours since the last phase maintenance occurred. Phase maintenance is a labor-intensive process that requires the thorough disassembly and inspection of an aircraft. These intervals vary based on the type of helicopter; Apaches must enter into major phase maintenance after the 500th cumulative flying hour. Phase maintenance goals also vary by aircraft with AH-64D/E aircraft being expected to complete their 500-hour phase maintenance in no more than 44 days [4]. Each aircraft type has its own dedicated maintenance team. Based on the data, maintenance timelines are generally more rapid in practice, and the goals in the table functionally serve as upper bounds on phase maintenance timelines. Our goal is to evaluate how aviation units make decisions across different aircraft and fleet conditions.

2 HYPOTHESIS DEVELOPMENT

The objective of our study is to evaluate the effectiveness of existing Army policies on aircraft deployment decisions at the unit level. Specifically, we focus on the requirements regarding Operational Readiness and Phased Maintenance. We review these policies and derive related hypotheses in the following.

2.1 Operational Readiness

OR was first introduced to the US Army in doctrine in 1978 and, in 1985, AR 700-138 established that the "objective of aircraft readiness is to achieve a 75% FMC goal at all times" [5, p. 6]. In 1998, the Office of the Deputy Chief of Staff for Logistics requested that the Army's Operations Research Center conduct a study on the need for an OR reporting system which then branched into a study of the history of OR in the Army [5]. They concluded that the 75% benchmark was created without "any analytical/engineering design criteria" and is not "linked to unit resources or capabilities" [5, p. 3]. Instead, they proposed a new method to monitor OR using control charts.

In the 2020s, the Army commissioned multiple Government Accountability Office (GAO) reports on readiness. Following a study using FY17–19 data, the GAO noted that the "services reported a variety of challenges related to air domain force elements including [...] the effects of trained pilot shortages on the Army's AH-64 attack helicopter" [6, p. 13]. It remains unclear how pervasive this shortage is across various units, and whether the Army employs a prioritization strategy to allocate limited resources, potentially leading to uneven effects on OR across the force. Importantly, newly minted pilots require additional training and can stress a unit's OR upon their initial arrival [7]. The FY17–19 study focused primarily on the human aspect of readiness, while future studies would investigate readiness from a systems and technology perspective [8, 9].

The GAO concluded a report on predictive maintenance for weapon systems that recommends the Army to efficiently improve its weapon system availability by identifying targets of opportunity, such as aircraft that are highly likely to experience failure in the near future [8]. In order to do so, the GAO suggested “reducing unplanned and unneeded maintenance” [8, p. 1] The Army differentiates between *planned* maintenance, which is preventive or predictive, and *unplanned* maintenance, which is reactive. There has been a strategic push towards predictive maintenance with an underlying goal of improving OR. Since 2002, many aviation units have started implementing predictive maintenance with mixed results due to the lack of standardized reporting metrics to properly evaluate its impact [8]. To convert unplanned maintenance occurrences into planned efforts, units now utilize sensors on aircraft that flag abnormal readings. Still, no standardized predictive maintenance doctrine exists, leading to varied implementations and effects [8]. In a recent success story, the 244th Expeditionary Combat Aviation Brigade reports heightened maintenance efficiency with the adoption of the ‘Griffin’ AI-based predictive platform [10].

Given the challenges posed by the absence of a unified predictive maintenance doctrine, it is valuable to explore how recent literature addresses various maintenance strategies and their associated impact on aircraft readiness. Lipina [11] finds Air Force aircraft availability to be closely tied to the capacity of qualified maintenance manpower. MacKenzie et al. [12] demonstrate through simulation how a 10% change in maintenance capacity can significantly impact readiness. Choo et al. [13] further extend this concept, linking sortie generation directly to usage. They argue that for a given level of maintenance capacity, a unit can determine an upper bound on long-run fleet availability, which is highly useful for planning and wargaming purposes.

Despite the advancements in maintenance strategy research, a significant gap exists in the literature regarding overarching policy evaluation. This is evident in studies such as Ritschel et al. [14], which admittedly overlook the impact of maintenance capacity on flight hours. McLean and Reiman [15] illustrate how small adjustments in the spare parts order process can significantly enhance readiness and cost efficiency, pointing towards the potential for policy-level interventions.

Informed by the history, doctrine, and literature above, and because Army commanders are held to the standards we have outlined, we hypothesize that as OR rates fall below R-1 status, units are deterred from continuing operations at the same operational tempo.

Hypothesis 1: *The probability of flying a Fully Mission-Capable (FMC) aircraft positively correlates with increased operational readiness (OR).*

2.2 Phase Maintenance Interval Management

In order to balance OR and utilization, the Army has recommended the usage of *bank hours*, which is simply the number of hours available to be flown, per aircraft, until phase maintenance. The Army advocates for using a bank hour flow chart for efficient maintenance scheduling. Proper phase maintenance interval management is shown in Figure 1 via the *circle* unit. Contrary to this consistent expected flow into maintenance is the backlogged *triangle* unit, which exemplifies poor management practices (three aircraft all simultaneously about to enter phase maintenance). Both example units have eight aircraft (A1-A8). An efficient unit will induct aircraft into phase maintenance over smooth, uniform intervals, which ultimately prevents backlog and surges in spare part requirements while simultaneously offering a greater degree of predictability and capability [4, paragraph 4-59]. Thus, the Commander at the appropriate echelon will specify goals for their unit's OR and total fleet bank time in a training cycle [4, paragraph 1-29]. It further warns commanders that "reporting high OR rates while not supporting high operational requirements" may "mask the ability to regenerate combat power" [4, paragraph 1-28].

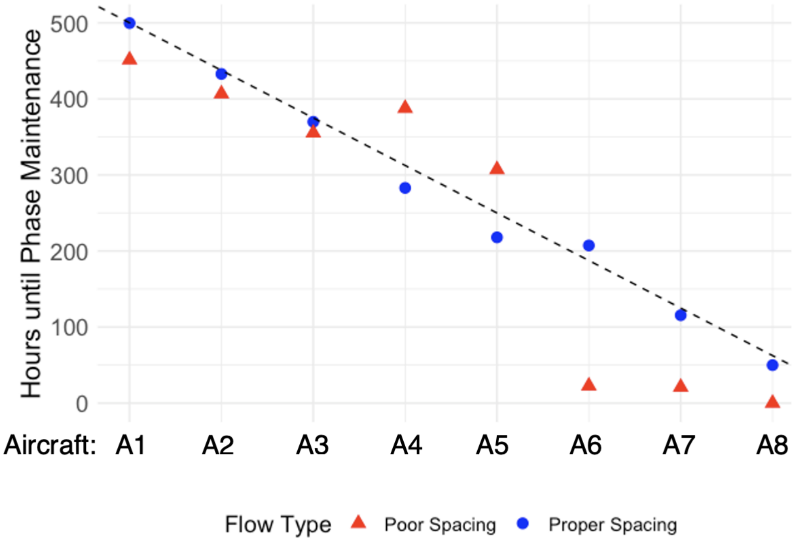

Figure 1. (color online) Examples of Proper (blue circle) vs Poor (red triangle) Phase Maintenance Interval Management (adapted from [4], ATP 3-04.7, Figures 4-2 and 4-3).

The shift towards predictive and preventive maintenance strategies is also central to current research. Gavranis and Kozanidis [16] introduce the concept of *residual flight time*—akin to what we refer to here as *hours until phase maintenance*—proposing algorithms to maximize fleet availability. This idea of strategic scheduling is further developed by Barde et al. [17], who focus on minimizing equipment downtime through a sequential reinforcement learning model. Öhman et al. [18] discuss *frontlog-scheduling*, emphasizing the optimization of maintenance schedules. These studies collectively highlight the importance of systematic maintenance planning in enhancing military aviation readiness.

A study by Colonel (retired) Bradley Pippin on *Allocating Flight Hours to Army Helicopters* reveals how manual flight hour allocation at the battalion level leads to decreased deployability, particularly when units hold back aircraft with low hours until phase maintenance [19]. Pippin suggests a policy adjustment in Army aviation reporting metrics by introducing a new availability metric that accounts for the deployability of aircraft in relation to maintenance schedules. He points out that "an aircraft with only one flight hour remaining until phase maintenance may be FMC, but is not available [for deployment]" [19, p.37].

Doctrine on maintenance interval management in ATP 3-04.7 guides units in preventing maintenance bottlenecks by encouraging an even distribution of aircraft usage throughout each maintenance cycle. This guidance seeks to avoid the worst-case scenario, identified in Pippin [19], of multiple aircraft with low hours until phase maintenance. Given the requirement to employ an aircraft evenly throughout its maintenance cycle, we hypothesize that the decision to deploy an aircraft should not be affected by its remaining hours to phase maintenance.

Hypothesis 2: *The probability of flying a FMC aircraft is independent of the remaining number of hours until required phase maintenance.*

3 DATA AND METHODOLOGY

We model the decision to fly a fully mission-capable aircraft via a Generalized Additive Model (GAM) given a set of covariates relevant to a decision-maker.

3.1 Data Origin and Filtering Criteria

We commence our analysis with a dataset from the Army's Engineering Research and Development Center. This dataset includes information such as the date, aircraft model, aircraft serial number, unit hierarchy, a detailed breakdown of maintenance status by hour, and recorded flying time in hours.

First, we filter our data to include only AH-64 Apache airframes, as these helicopters are generally assigned to Attack and Reconnaissance Battalions with a specific mission set. Per FM 3-04, their primary missions of attack, reconnaissance, and security do not include general support operations apart from aerial escort duties [20]. As a result, flight hours on these airframes are most likely to reflect either training or mission-specific usage. This reduces the impact of confounding factors associated with the usage of other airframes that perform transportation and logistical roles.

We further apply filtering criteria to analyze data from 1 October 2019 to 30 May 2022. In order to avoid crossing an additional fiscal year (which begins on 1 October each year), we exclude three months of data from the summer of 2019. The dataset does not continue past June 2022. This time frame comprises 973 days, encompassing 642,232 daily status reports from 797 unique aircraft as recorded on Department of the Army Form 1352. We focus on Active Duty, non-training units, reducing the dataset to 417,678 observations across 61 companies within 21 battalions. The final filtered data includes 423 unique aircraft serial numbers, each of which, coupled with a date, creates a single observation that we refer to as an *aircraft day*. We concentrate on the decision to deploy a FMC aircraft, and so filter out any observations without at least one FMC hour in the aircraft day (8.2% of data), reducing the dataset to 314,575 observations. We also impute a zero for units that do not report any flying hours for a given day. The unit with the highest observed missing flight data is missing 41% of its fleet's total potential observations. This unit had a seven-months-long period from August 2021 until February 2022 in which they did not report at least seven of their 24 aircraft. However, this time period is known to correspond to a deployment for that unit. Removing observations that contain incomplete data, such as the unit described above, reduces the dataset further. This step involved the removal of six companies and two battalions, one of which did not have available data until 2021. Flight occurs on 16.7% of the days with an average sortie length of 3.3 hours. Lastly, there are 7,165 instances in which an aircraft flew on a given day in which no FMC hours are observed—these hours were originally presumed to be comprised mainly of partially mission-capable hours. Surprisingly, there were also 1,665 occurrences (2.9% of flying days) in which an aircraft logged 24 hours of NMC time in a day and also flew. Of these occurrences, 83%

spent all 24 hours in a FIELD state, suggesting that field maintenance teams routinely conduct flight tests while the aircraft are in their hands. These are all fringe cases and are ultimately dropped from the modeling portion of the paper as we are only interested in the decision to fly when an aircraft observes FMC hours in a day.

3.2 Variables: Description, Operationalization, and Controls

Based on our hypothesized effects, the key independent variables in our model influencing the decision to fly a mission-capable aircraft are unit operational readiness and the remaining flying hours until phase maintenance is required. The day of the week plus the remaining number of days in the reporting period are controls that could plausibly affect the decision to fly. We further account for the year, month, and battalion to control for seasonality and training cycles. The response variable is binary: whether a FMC aircraft flies on a given day or not. First, we compare the bivariate relationships between each of the covariates and our response.

Operational Readiness OR is a unit's monthly average equipment availability rating, measured as its percentage of time in FMC status. The reporting period begins on the 16^{th} of the current month and closes on the 15^{th} of the following month. From Figure 2, the probability of aircraft flying on a given day decreases roughly linearly from 60% to 100%. There appears to be an inflection point at about 60% OR, suggesting that the overall relationship might not be linear. Only 15.4% of observations fall into R-3 or R-4; accordingly, the confidence interval as we approach the tail end of OR widens substantially.

Hours until Phase Maintenance Throughout this study, when referencing phase maintenance, we are referring to *major phase maintenance*, which Apaches are required to undergo after every 500 flying hours. For example, if an aircraft has flown 50 hours since its last phase maintenance, it would have 450 flying hours left until its next one. Per Figure 3, the probability of flying is lowest just before (after) aircraft enter (exit) phase maintenance.

We impute the number of hours that an aircraft has until phase maintenance by searching for each aircraft's top five longest consecutive streaks of NMC hours using Algorithm 1 in Appendix B. The logic of this algorithm is the following: first, we calculate the total flying hours over the dataset's duration and estimate the number of 500-hour phase maintenance cycles. We then focus on identifying the top five longest periods of NMC status for each aircraft to search for candidate phase maintenance cycles.

For each identified NMC period, we assess the flying hours accumulated prior to the period. This step is crucial in estimating the likelihood of the NMC period in question representing a phase maintenance cycle. We consider both the duration of downtime and the flying hours leading up to the NMC period. In making this estimation, it is important to recognize that extremely short durations of NMC, such as three days, are highly unlikely to correspond to a full phase maintenance period due to the insufficiency of time for all required maintenance tasks. Finally, we allocate the maintenance gaps to the most probable maintenance blocks following approximately 500 flying hours. In doing so, we also account for the possibility of early phasing by units as recommended in Paragraph 4-46 of HQDA [4], acknowledging the operational variability in maintenance

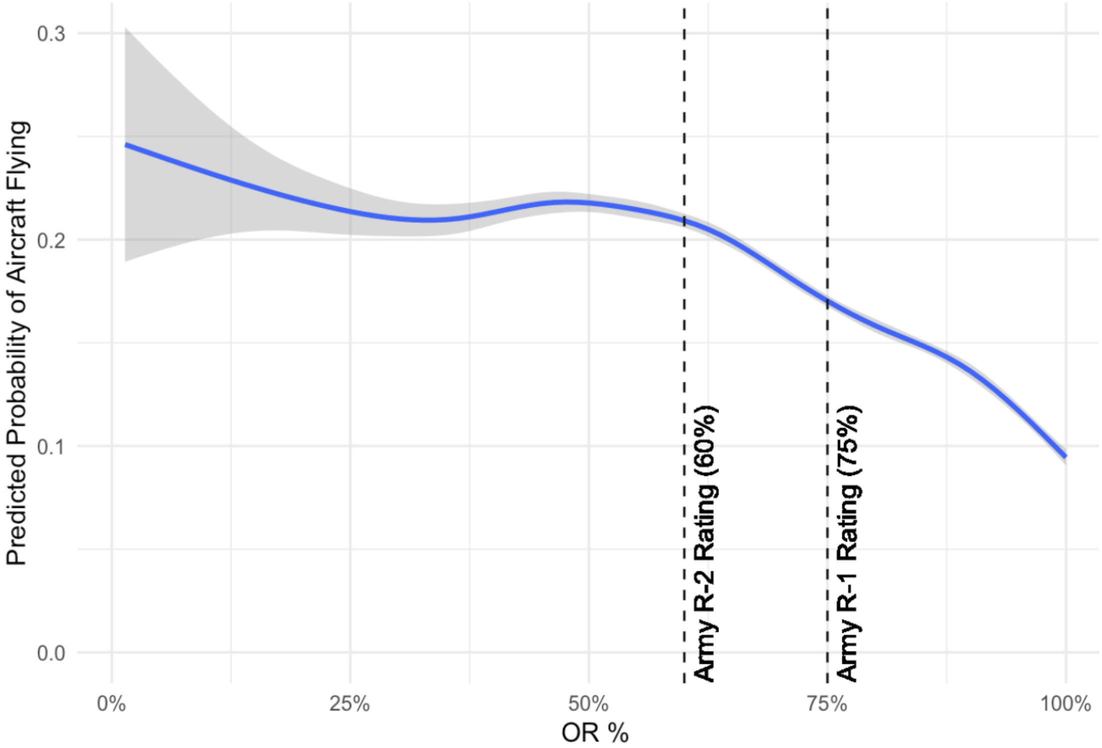

Figure 2. (color online) Fitted GLM of OR vs flying; units tend to fly more when OR is reduced. The 95% confidence interval widens below 30% OR due to limited observations in this range.

scheduling across different battalions. Once we determine the date on which an aircraft enters into phase maintenance, we decrement the number of bank hours remaining on the aircraft each time it flies to keep a running tally of the number of hours until phase maintenance is required.

Control Variables We control for *Days until Reporting Period Closure*. Units are required to report their monthly average OR rating on the 15th of each month–the process then resets on the following day. Thus, on any given day, we determine the number of days until the reporting period ends. We did not discretize this covariate as there is no distinct block in which doing so made practical sense. It is treated as continuous from zero to thirty days. We also control for the *Day of the Week*, which we treat as a categorical variable, with Wednesday serving as the reference value. We further include *Month* (reference value: January) and *Year* (reference value: 2019) as categorical variables to control for seasonality, and an aircraft's *Battalion* as a random effect.

3.3 Generalized Additive Model

We employ a generalized additive model (GAM) to investigate the determinants influencing the log-likelihood of a FMC aircraft being dispatched for flight on a given day. Let Y_{ij} represent the binary outcome of flight for aircraft j on date i, where $Y_{ij} = 1$ indicates flight and $Y_{ij} = 0$ indicates grounding of aircraft j on date i; then, the log-likelihood of

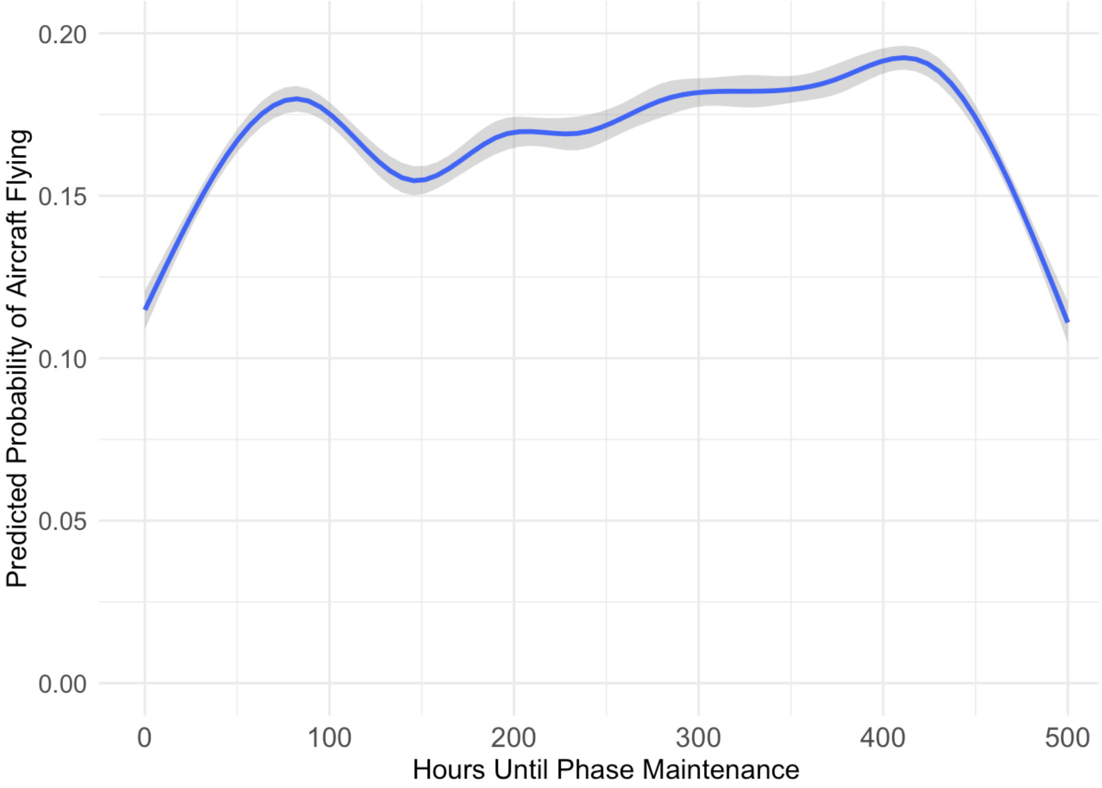

Figure 3. (color online) Fitted GLM of phase maintenance vs flying; flight likelihood is lowest immediately before and after phase maintenance.

flight can be expressed as

$$\begin{aligned}\log\left(\frac{\hat{P}(Y_{ij}=1)}{1-\hat{P}(Y_{ij}=1)}\right) &= \beta_0 + f_1 \times \text{Operational Readiness}_{ij} \\ &\quad + f_2 \times \text{Hours until phase maintenance}_{ij} \\ &\quad + f_3 \times \text{Days remaining in reporting period}_i \\ &\quad + f_4 \times \text{Battalion}_j \text{ (Random Effect)} \\ &\quad + \beta_1^{\text{d}} \times \text{Day of the week}_i \\ &\quad + \beta_2^{\text{m}} \times \text{Month}_i + \beta_3^{\text{z}} \times \text{Year}_i + \varepsilon_{ij}.\end{aligned} \tag{1}$$

The smooth functions f_1, f_2, and f_3 represent continuous covariates and are estimated non-parametrically. Each function can be articulated as

$$f_i(x) = \sum_{k=1}^{K_i} a_{ik} b_{ik}(x), \tag{2}$$

where the $b_{ik}(x)$ denote the basis functions, and the a_{ik} are the coefficients estimated for the i-th smooth function. K_i represents the number of basis functions utilized for the i-th smooth term. The function f_4 is associated with the smoothing spline for the random effect related to an aircraft's battalion. In this framework, the β coefficients resemble those in a logistic regression model, signifying fixed effects for each day of the week d, each month m, and each year z (2019–2022), plus an intercept. In contrast, the a_{ik} coefficients correspond to the non-linear effects captured by the smooth functions. The term ε_{ij} encapsulates the random error or the unexplained variation in the data for aircraft j on date i.

The spline fitting process minimizes a penalized likelihood criterion, where the smoothing parameter applied to reduce overfitting is adjusted based on cross-validation. Credible intervals for the fitted splines are calculated using the estimated covariance matrix of the coefficients, assuming a Gaussian distribution for these coefficients, as detailed in Wood [21].

4 RESULTS AND DISCUSSION

GAMs are flexible, non-parametric models that allow non-linear fits to a response if such fits are warranted. Splines and their credible intervals in Figures 6–8 show how OR, maintenance requirements, and reporting periods relate to flight decisions. These illustrations, combined with model results in Table 1, allow us to assess the effect of each covariate on the decision to fly a FMC aircraft. The continuous nature and visualization of the splines aid in model interpretability and allow us to identify trends in the data without relying strictly on fixed effects.

By selecting the optimal prediction acceptance threshold via cross-validation, we are able to achieve a maximum F1 score of 0.38 and 63.5% accuracy, as shown in Figure 4. F1 gradually decreases as the model moves towards a standard 50-50 weighting on flying and non-flying days. F1, which balances sensitivity and precision, is the conventional metric for classification models and is well-suited for evaluating imbalanced data such as ours. Explained deviance is adapted from Faraway [22, p. 32]; F1, sensitivity, and precision are adapted from Dalianis [23]. The number of knots chosen for each spline is selected via cross-validation. Table 10 in Appendix C shows the adequacy of fit at the given knot level of four, five, and five for OR, hours until phase, and days until report, respectively.

4.1 Analysis and Test of Hypotheses

We summarize the results of the regression in Table 1. The final model (Model D) achieves an Adjusted R^2 of 0.11 and explained deviance of 9.7%.

We evaluate Hypotheses 1 and 2 using a combination of the results presented in Table 1 and a component-wise visual inspection of the smoothing splines and their respective significance regions in Figures 6, 7 and 8. Here, we use the term *credible interval* to refer to these significance regions, which is the Bayesian analog to frequentist *confidence intervals* [24]. We follow conventional practice by using credible intervals of the smoothing splines to perform inference (cf., Hastie and Tibshirani [25] and Wood [26], sections 6.8 and 6.10). Building on the work from Marra and Wood [27], who derive credible intervals for individual smoothing splines using the covariance matrix V_f of each spline's coefficients, Wood [28] shows these intervals have Wald-like test statistics with frequentist coverage properties, allowing for the creation of posterior $1-\alpha$ credible intervals. Following the suggestion by Ruppert et al. [29, section 6.8], we use visual inspections of the first derivative to identify areas with higher rates of change between flying and our covariates.

Table 1. Consolidated Model Summary

	Model			
Term	A	B	C	D
Spline EDF (Chi-Squared Test Stat.)				
OR	—	2.993	—	2.993
	—	**(6045.7)**	—	**(5961)**
Hours until Phase	—	—	3.994	3.993
	—	—	**(1269.2)**	**(1180.4)**
Days until Report	3.989	3.989	3.97	3.989
	(502.2)	**(507.6)**	**(504.3)**	**(508.5)**
Battalion†	17.918	17.882	17.914	17.876
	(3993.4)	**(2848.8)**	**(3791.8)**	**(2719.9)**
Coefficient Estimates (Std. Error)				
Sunday	**-1.72**	**-1.61**	**-1.72**	**-1.61**
	(0.014)	(0.014)	(0.014)	(0.014)
Monday	**-0.31**	**-0.24**	**-0.31**	**-0.24**
	(0.011)	(0.012)	(0.011)	(0.011)
Tuesday	-0.007	0.008	0.016	0.019
	0.014	(0.011)	(0.011)	(0.011)
Thursday	**-0.23**	**-0.22**	**-0.23**	**-0.22**
	(0.011)	(0.011)	(0.11)	(0.011)
Friday	**-1.05**	**-0.99**	**-1.05**	**-0.99**
	(0.012)	(0.013)	(0.012)	(0.013)
Saturday	**-1.73**	**-1.62**	**-1.73**	**-1.62**
	(0.014)	(0.014)	(0.014)	(0.014)
Month	✓	✓	✓	✓
Year	✓	✓	✓	✓
Model Metrics				
Adjusted R^2	0.097	0.11	0.0997	0.112
Explained Deviance	8.3%	9.49%	8.56%	9.71%
Count				265,472

Note: **Bold** indicates $p-Value < 0.001$; EDF: Estimated Degrees of Freedom; † random effect; ✓control variable included but not reported

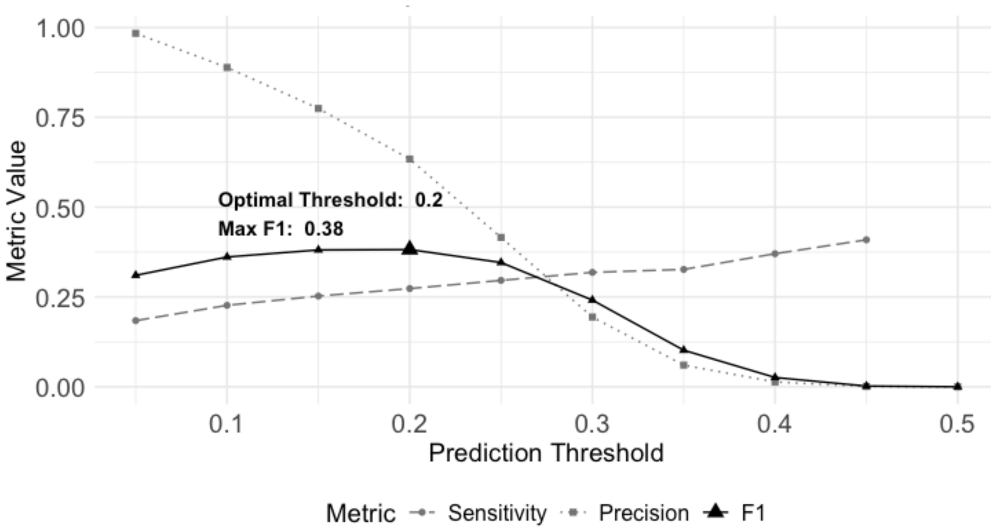

Figure 4. Final Model (D) performance metrics across prediction acceptance thresholds using 100 iterations of 5-fold cross-validation.

Table 6 in Appendix A presents odds ratios for each covariate at different levels, with all other variables held at their medians. The odds ratios provide context to the models and illustrate the practical implications of the model. The results indicate that units are almost 9% less likely to fly an aircraft with 100 hours until phase compared to 400 hours, so they are unlikely to be following the spirit of the phase interval management doctrine.

Interestingly, as shown in Figure 5, the day of the week has a strong impact on the decision to fly a FMC aircraft. It is the strongest determinant, in terms of magnitude, of the odds of flight on a given day in the model. While we observe a statistically significant drop in the odds of flight on the weekend of approximately 80.0%, we also find a significant decrease in the odds on Mondays (21.3%), Thursdays (19.4%), and Fridays (62.9%) compared to Wednesdays, all else constant. As expected, units are more likely to fly in the spring and fall months, which corresponds with traditional training cycles. Units are least likely to fly in 2020 and gradually increase sortie occurrences in both 2021 and 2022 but still do not return to 2019 levels.

The battalion random effect is highly statistically significant, pointing towards location-specific idiosyncrasies in determining flight operations. The number of days remaining in the reporting period has a statistically significant yet practically slight effect on flight operations. Specifically, when comparing 1 day to 15 days until the reporting period closure, the odds ratio increases from 0.465 to 0.478, reflecting a 2.8% increase. Further extending the timeframe to 25 days elevates the odds ratio to 0.523, which constitutes a 9.4% rise from 15 days and a 12.5% increase from the 1-day scenario.

Hypothesis 1: Operational Readiness Table 1 and Figure 7 provide convincing evidence against **Hypothesis 1**. In fact, units are generally *more* likely to fly when their OR rating is in the R-2, R-3, or R-4 levels compared to R-1, as evidenced by the sharp first derivative between OR levels of 0.8 to 1.0. The spline for OR is statistically significant, with an effective degrees of freedom (EDF) of 2.997, a Chi-squared statistic of 5992.7, and a p-value below 0.001, indicating a strong non-linear effect on the

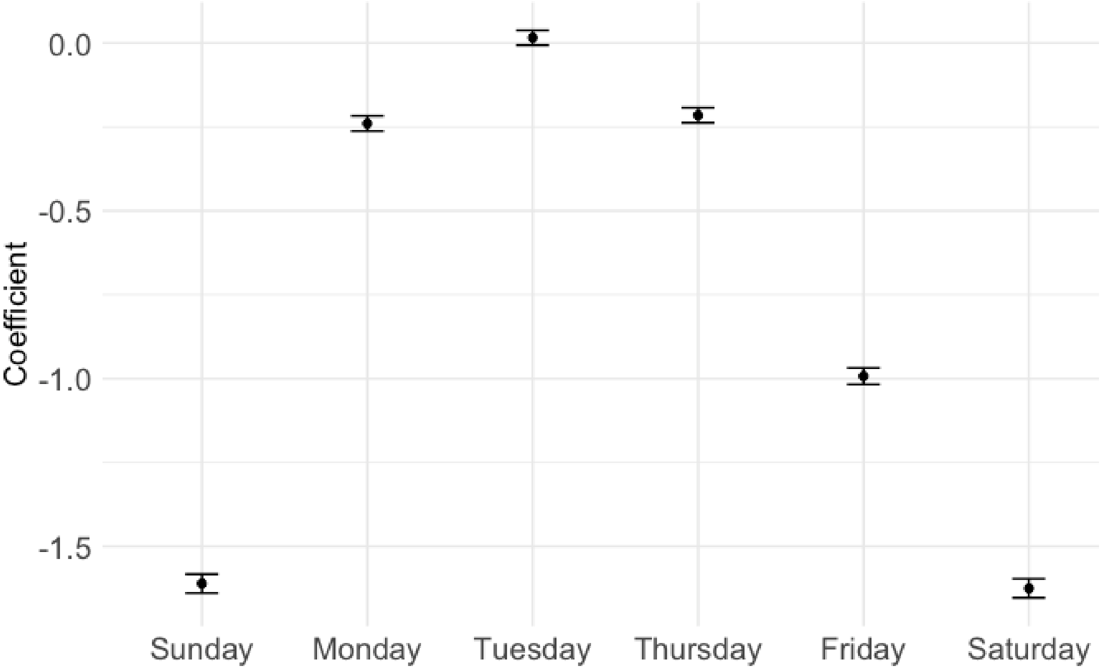

Figure 5. Estimated fixed effect of day of the week on flying (Full Model D) with 95% confidence intervals relative to the reference level *Wednesday*. Units tend to normally distribute their flying hours around the center of the work week.

response variable. Our initial hypothesis that a diminished OR rating acts as a deterrent to flying is thus rejected. In Table 6, the odds ratio for a flight occurrence at 60% OR is 0.563, which is a 64.1% increase from the odds ratio at 90% OR (0.343). Similarly, there is a 14.4% increase from the odds ratio at 75% OR (0.492), with all other variables held constant at their median values.

Hypothesis 2: Hours until Phase Maintenance We also reject **Hypothesis 2**. Evidence in Figure 8 and Table 1 suggests that units are significantly less likely to fly a FMC aircraft that is close to phase maintenance (either recently departed or soon to enter). The first derivative is visibly increased in the 0-50 hours and the 450-500 hours intervals, indicating a potential change in deployment decisions for aircraft in these intervals. Holding all other covariates at their median levels, an isolated increase in *Hours until Phase* from 10 to 400 hours corresponds to a 19.9% rise in the odds ratio, from 0.442 to 0.530. Additionally, a change from 250 to 400 hours yields a 7.3% increase in the odds ratio, from 0.494 to 0.530. Units clearly do not place equal emphasis on flying each aircraft regardless of its remaining hours until major phase maintenance.

Given that the Army explicitly states in its doctrine that the goal for equipment readiness is 75% FMC [3, 1-19] and that units should seek well-spaced flow into phase maintenance (such as in Figure 1), our findings suggest a disconnect between policy objectives and their practical implementation.

Contingency tables outlining the frequency of observations in various OR and hours until phase levels can be found in Tables 3 and 4 in Appendix A. From the contingency tables, only 45.3% of observations meeting our filtering criteria fall into R-1 status–the

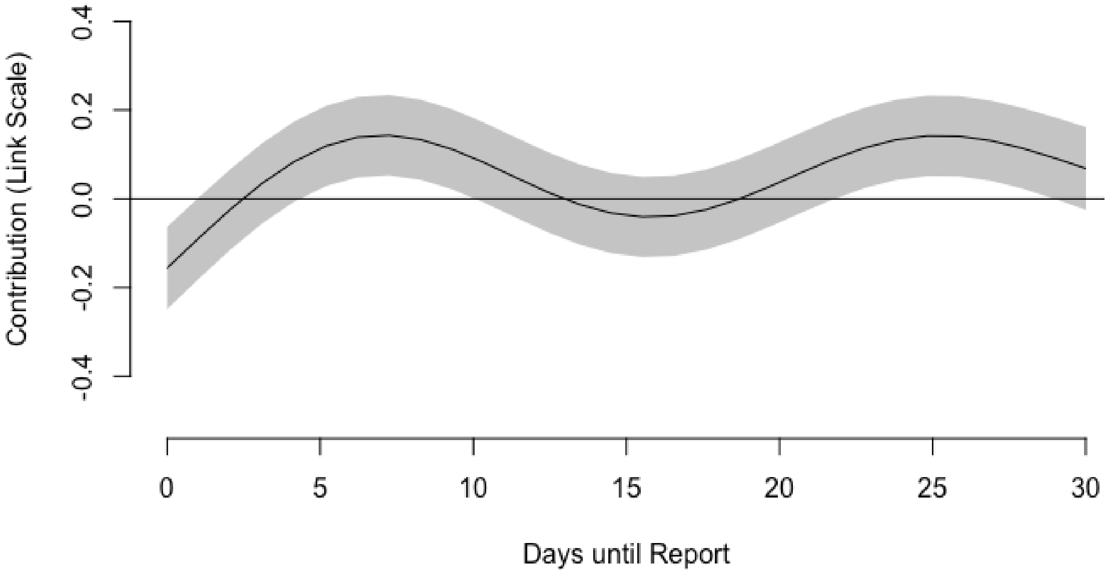

Figure 6. Days until Report spline with 95% credible interval; flight likelihood is lowest in the final 3 days of the reporting period (Full Model D).

majority of them fall below it. Looking at unfiltered, aggregated battalion-level data in Table 4, only 41.3% of battalion days achieve R-1 status. Again, units are more likely to fly in R-2 or R-3/R-4 compared to R-1 (cf. Figure 2. One potential reason behind this could be cyclical training patterns in which high flying levels and, thus, reduced OR, correlate for periods of time. Falling below R-1 status does not preclude units from further operations. Certainly, the mission clearly does (and should) take precedence, but it brings into question the efficacy of an OR-centric policy. If a unit is meeting its current mission requirements, what levels of OR might be acceptable? Currently, in doctrine, there is no official link between operational intensity and OR.

In addition, we would expect maintenance officers to adjust flying hours on aircraft judiciously across a fleet to ensure the unit achieves the appropriate balance of ‘bank hours’ that evenly spaces aircraft into phase maintenance. It seems, however, that units are more likely to resort to adjusting flight patterns of aircraft that have either just left phase maintenance or are about to enter it. This behavior, while perhaps rational under certain circumstances, suggests a lack of long-range planning endemic to the entire force.

4.2 Post Hoc Analysis of an Interaction Effect

Our observations in relation to Hypotheses 1 and 2 suggest that other factors, possibly behavioral, may influence flight operations. A unit’s OR status plausibly affects its flight patterns. Despite the extensive research on optimizing maintenance scheduling and spare parts logistics in military supply networks [30–34], one aspect often remains overlooked: the human element in executing these systems. Behavioral factors play a role in the effectiveness of any operational strategy, influencing outcomes in sometimes unpredictable ways. In order to further test flight operation decision-making at the unit level, we develop a model that takes into account potential differences between units

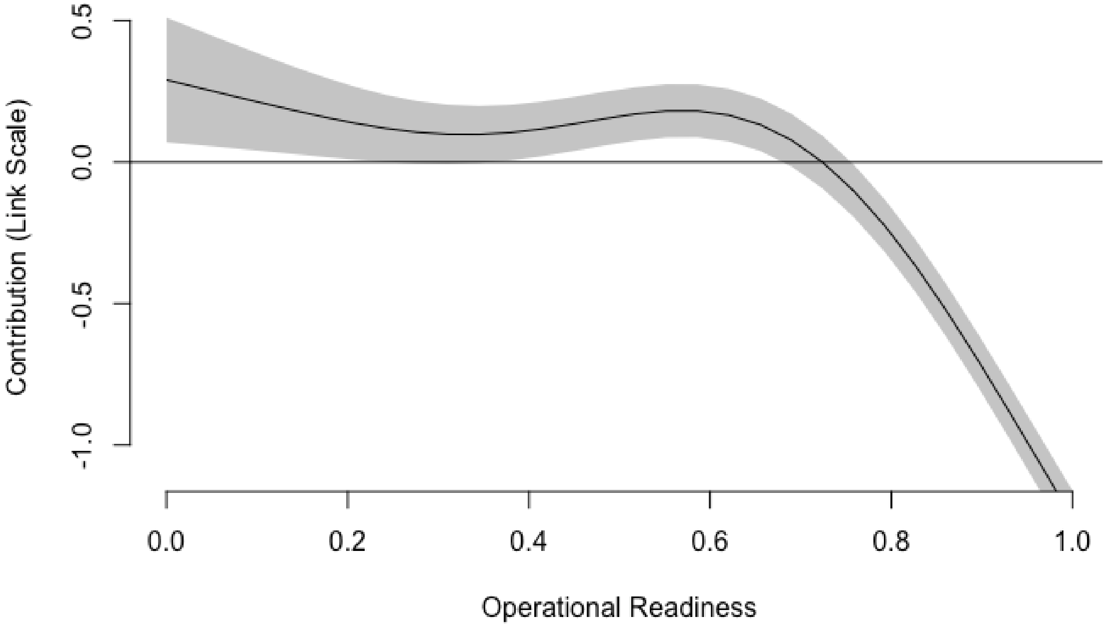

Figure 7. OR spline with 95% credible interval; units are more likely to fly at lower OR levels (Full Model D).

operating at different OR levels. We fit a tensor product as the interaction between OR and hours until phase. An aircraft's battalion is still treated as a random effect, serving as a proxy for latent location-specific factors like weather, mission requirements, geography, budget, and command preferences.

Again, let Y_{ij} represent the binary outcome of flight for aircraft j on date i. The formulation is consistent with Equation 1, with the addition of the term f_5, which represents the tensor product spline $f_1 \otimes f_2$ and models the interaction between OR and hours until phase maintenance.

A Chi-squared test of significance from an ANOVA comparing Model D and Model E indicates a statistically significant improvement in fit. Full model results are presented in Table 7 in Appendix A; splines for hours until phase, OR, and days until reporting period closure are illustrated in Figures 9-11 in Appendix B. The optimal knot selection remains the same as in Model 1 from Figure 13. The adjusted R^2 increases to 0.113 and explained deviance to 9.81%. While interpretability suffers from the inherent additional complexity due to the inclusion of a tensor product, the interaction between hours until phase and OR is statistically and practically significant. Rather than compare odds ratios from spline coefficients, we instead can directly view predicted probabilities of flight across various scenarios, as shown in Table 2.

The non-linear, disproportionate decrease in the probability of flight occurrence across OR levels indicates a potential interaction effect present in the decision-making process. While the overall impact of low OR still dominates—a unit at 90% OR is not more likely to fly an aircraft than at 75% OR, regardless of the hours until phase maintenance remaining—the interaction can be summarized succinctly from Table 2. We observe that, as hours to phase maintenance decrease from 400 to 100, for an average unit, there is a 9.7% decrease in the probability of flying an aircraft if the OR rating is at

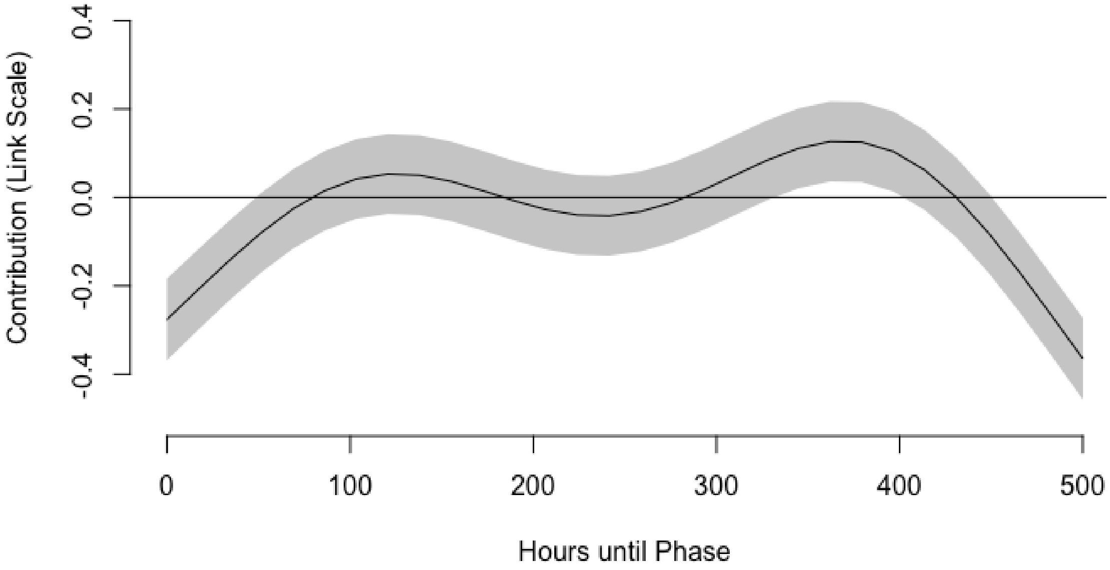

Figure 8. Hours until Phase Maintenance spline with 95% Credible Interval; flight likelihood decreases near phase maintenance thresholds (Full Model D).

Table 2. Predicted probability of flying under varying circumstances of OR and hours until phase.

Hours until Phase	OR		
	60%	75%	90%
100 hours	0.536	0.489	0.330
250 hours	0.642	0.567	0.450
400 hours	0.635	0.586	0.394

75% compared to a 6.4% decrease at 90% OR. Similarly, as hours to phase maintenance decreases from 250 hours to 100 hours, there is a 7.7% decrease in the probability of flying an aircraft at 75% OR versus a 12.0% decrease at 90% OR. We conclude that the marginal effect of hours until phase on aircraft deployment decisions clearly is not independent of OR. Our observations suggest the possible presence of risk-averse behavior: units striving to meet readiness ratings could opt for a conservative approach, holding back aircraft close to phase maintenance in order to reduce risk. The results are consistent with the notion from Lehman et al. [35] that low-performing organizations may avoid risky behavior.

Interestingly, the effect of days until the reporting period—as shown in Appendix C, Table 11—is diminished and is modeled as a nearly linear fit with an EDF of 1.54. In the second model, holding all other covariates at their median levels, a unit with 30 days remaining in the reporting period exhibits a 1.8% increase in the predicted probability of flying compared to when there is just one day remaining. The effect of the day of the week is still also statistically and practically significant. The findings related to Hypotheses 1 and 2 are consistent with the previous model. A deeper discussion of the post hoc model assumptions and diagnostics, specifically with regard to concurvity, can be found in Appendix C.

5 CONCLUSION

Evidence from this study suggests a disconnect between policy objectives and their practical implementation. Given the doctrine surrounding Army aviation, we would expect to find flight operations reduce in a unit when OR falls below R-1 and that the time until phase maintenance for an aircraft does not correlate with usage. Instead, we find that diminished OR levels do not seem to preclude units from continued flight operations, and the remaining number of hours until phase maintenance has a significant impact on a unit's decision to dispatch a FMC Apache.

Our results suggest that units flying according to mission requirements may exercise limited discretion in their day-to-day operations, aiming instead to meet the overarching needs of the Command. Further, units with reduced OR are significantly less likely to fly an aircraft with low time until phase maintenance, possibly indicating risk-averse behavior and a potential lack of long-range planning in accordance with the recommended phase maintenance interval management doctrine. Control variables such as the day of the week and the number of remaining days in the reporting period are significant across all models. These behavioral factors, which on the surface may not seem linked to flight decisions, show an unexpectedly high correlation with aircraft flight occurrences. This study thus provides potential evidence that current policies overly focus on OR and consequently do not place enough emphasis on flying (and thus training).

Future research should focus on the decision-making frameworks of different units, as such patterns could dictate outcomes that vary in efficiency and effectiveness. These findings suggest that OR alone may not be a sufficient statistic to assess aviation readiness or unit performance. A more holistic approach that also accounts for the human elements present in the system could provide the Army and aviation community with a deeper understanding of a unit's true potential for sustained and reliable combat power projection.

Although this study focuses on AH-64 Apache helicopters, the methods and ideas used can serve as a basis for future research on similar patterns in other aircraft and even ground vehicles. An important question becomes whether units exhibiting these behaviors tend to perform better or worse overall. This calls for further study on how unit decisions affect readiness and operational success. Understanding these connections could help improve readiness goals and ensure policies both support and encourage effective real-world practices.

Disclaimer The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the United States Army, the Department of Defense, or the United States Government.

The first author completed this work as part of his official duties as an employee of the United States Government. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements A special thank you to LTC Thomas Dirienzo from the US Army's Artificial Intelligence Integration Center; Dr. Kyle Miller from Carnegie Mellon

University; Mr. Gregory Jinks and Mr. John Holdcraft from the US Army's Program Executive Office–Aviation; Ms. LaKenya Walker from the US Army Engineer Research and Development Center; and Dr. Shawn Mankad from North Carolina State University.

REFERENCES

[1] HQDA. Army Regulation 220 – 1: Army Unit Status Reporting and Force Registration–Consolidated Policies, 8 2022. Headquarters, Department of the Army, `https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN31982-AR_220-1-000-WEB-1.pdf`.

[2] GAO. National Guard Helicopters: Additional Actions Needed to Prevent Accidents and Improve Safety. Report to the Committee on Armed Services, House of Representatives GAO-23-105219, United States Government Accountability Office, 3 2023. `https://www.gao.gov/assets/gao-23-105219.pdf`.

[3] HQDA. Army Regulation 700-138: Army Logistics Readiness and Sustainability, 4 2018. Headquarters, Department of the Army, `https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN7663_AR700-138_Web_FINAL.pdf`.

[4] HQDA. Army Techniques Publication 3-04.7: Army Aviation Maintenance, 10 2020. Headquarters, Department of the Army, `https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN31028-ATP_3-04.7-000-WEB-1.pdf`.

[5] COL James L. Kays, LTC William B. Carlton, MAJ Mark M. Lee, and CPT William L. Ratliff Jr. Analysis of Operational Readiness Rates. Technical report, United States Military Academy, Department of Systems Engineering and Operations Research Center, West Point, New York, July 1998. `https://apps.dtic.mil/sti/pdfs/ADA361030.pdf`.

[6] GAO. MILITARY READINESS: Department of Defense Domain Readiness Varied from Fiscal Year 2017 through Fiscal Year 2019. Report to Congressional Committees GAO-21-279, United States Government Accountability Office, 4 2021. `https://www.gao.gov/assets/gao-21-279.pdf`.

[7] Kyle McDermott. Conversation on AH-64 Apache company command experiences. Personal communication, 2024. Major Kyle McDermott, U.S. Army, conversation on 21 February 2024.

[8] GAO. MILITARY READINESS: Actions Needed to Further Implement Predictive Maintenance on Weapon Systems. Report to the Committee on Armed Services, House of Representatives GAO-23-105556, United States Government Accountability Office, 12 2022. `https://www.gao.gov/assets/gao-23-105556.pdf`.

[9] GAO. Weapon System Sustainment: Aircraft Mission Capable Goals Were Generally Not Met and Sustainment Costs Varied by Aircraft. Report to

Congressional Committees GAO-23-106217, United States Government Accountability Office, November 2022. `https://www.gao.gov/assets/gao-23-106217.pdf`.

[10] Jonathan Barbee. Griffin Program is a “Game Changer” for Maintenance. *Army Aviation Magazine*, 73(1):22, January 2024. Army Aviation Association of America (AAAA).

[11] A.J. Lipina. Identifying Critical Factors Affecting Combat Mission Ready Status Among USAF Europe’s Aircrew. Graduate Research Project, School of Engineering and Management, Air Force Institute of Technology, 6 2009.

[12] Adam MacKenzie, J.O. Miller, Raymond R. Hill, and Stephen P. Chambal. Application of Agent-based Modelling to Aircraft Maintenance Manning and Sortie Generation. *Simulation Modelling Practice and Theory*, 20(1):89–98, 2012.

[13] Nathaniel Choo, Darryl Ahner, and Lance Champagne. The Effects of Aircraft Use and Available Repair Spares on Aircraft Sortie Generation: a Long-duration Logistical Wargaming Simulation Tool. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 20:154851292110407, 09 2021.

[14] J. Ritschel, T. Ritschel, and N. York. Providing a Piece of the Puzzle: Insights into the Aircraft Availability Conundrum. *Journal of Defense Analytics and Logistics*, 3 (1):29–40, 2019.

[15] V. McLean and A. D. Reiman. Transportation Service Level Impact on Aircraft Availability. *Journal of Defense Analytics and Logistics*, 6(1):46–58, 2022.

[16] Andreas Gavranis and George Kozanidis. An Exact Solution Algorithm for Maximizing the Fleet Availability of an Aircraft Unit Subject to Flight and Maintenance Requirements. *European Journal of Operational Research*, 242:631–643, 04 2015. doi: 10.1016/j.ejor.2014.10.016.

[17] Stephane Barde, Soumaya Yacout, and Hayong Shin. Optimal Preventive Maintenance Policy Based on Reinforcement Learning of a Fleet of Military Trucks. *Journal of Intelligent Manufacturing*, 30(1):147–161, 01 2019. doi: 10.1007/s10845-016-1237-7.

[18] Mikael Öhman, Markus Hiltunen, Kai Virtanen, and Jan Holmström. Frontlog Scheduling in Aircraft Line Maintenance: From Explorative Solution Design to Theoretical Insight into Buffer Management. *Journal of Operations Management*, 67(2):120–151, 2021.

[19] Bradley W. Pippin. Allocating Flight Hours to Army Helicopters. Master’s thesis, Naval Postgraduate School, Monterey, California, 6 1998. `https://apps.dtic.mil/sti/pdfs/ADA350027.pdf`.

[20] HQDA. Field Manual 3-04: Army Aviation, 4 2020. Headquarters, Department of the Army, `https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN21797_FM_3-04_FINAL_WEB_wfix.pdf`.

[21] Simon N. Wood. Inference and computation with generalized additive models and their extensions. *TEST*, 29:307–339, April 2020. `https://doi.org/10.1007/s11749-020-00711-5`.

[22] Julian J. Faraway. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2 edition, 2016. ISBN 9781498720984.

[23] Hercules Dalianis. *Clinical Text Mining*, chapter 6. Springer, 2018. ISBN 978-3-319-78502-8.

[24] Gavin L. Simpson. Modelling Palaeoecological Time Series Using Generalised Additive Models. *Frontiers in Ecology and Evolution*, 6, 2018. `https://www.frontiersin.org/articles/10.3389/fevo.2018.00149`.

[25] Trevor Hastie and Robert Tibshirani. Generalized Additive Models: Some Applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987. ISSN 01621459. `http://www.jstor.org/stable/2289439`.

[26] Simon N. Wood. *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman and Hall/CRC, 2 edition, 2017. `https://doi.org/10.1201/9781315370279`.

[27] Giampiero Marra and Simon N. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39 (1):53–74, 2012. `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2011.00760.x`.

[28] Simon N. Wood. On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1):221–228”, March 2013. ISSN 0006-3444.

[29] D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003. ISBN 9780521785167. `https://books.google.com/books?id=Y4uEvXFP2voC`.

[30] Blake E. Schwartz. *Performance Evaluation and Improvement of Expeditionary Military Sustainment Network Models*. PhD thesis, NC State University, Raleigh, NC, 2020. `http://www.lib.ncsu.edu/resolver/1840.20/37489`.

[31] Brandon M. McConnell, T.J. Hodgson, M.G. Kay, R.E. King, Y. Liu, G.H. Parlier, K. Thoney-Barletta, and J.R. Wilson. Assessing Uncertainty and Risk in an Expeditionary Military Logistics Network. *Journal of Defense Modeling and Simulation*, 18(2):135–156, 2021. `https://doi.org/10.1177/1548512919860595`.

[32] Matthew B. Rogers, Brandon M. McConnell, Thom J. Hodgson, Michael G. Kay, Russell E. King, Greg H. Parlier, and Kristin Thoney-Barletta. A Military Logistics Network Planning System. *Military Operations Research*, 23(4):5–24, 2018. `https://repository.lib.ncsu.edu/items/76464071-5a3f-4bf5-aa18-123b72cd4a3a`.

[33] Thomas R. O'Neal. *Sortie-based Aircraft Component Demand Rate to Predict Requirements*. PhD thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, 2020. `https://scholar.afit.edu/etd/3199`.

[34] M. Verhoeff, W.J.C. Verhagen, and R. Curran. Maximizing Operational Readiness in Military Aviation by Optimizing Flight and Maintenance Planning. *Transportation Research Procedia*, 18:941–950, 2015. doi: 10.1016/j.trpro.2015.09.048. 18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015. `https://www.sciencedirect.com/science/article/pii/S2352146515002941`.

[35] David W. Lehman, Jungpil Hahn, Rangaraj Ramanujam, and Bradley J. Alge. The Dynamics of the Performance-Risk Relationship Within a Performance Period: The Moderating Role of Deadline Proximity. *Organization Science*, 22(6):1613–1630, 2011. ISSN 10477039, 15265455.

[36] Eva Cantoni and Trevor Hastie. Degrees of Freedom Tests for Smoothing Splines. *Biometrika*, 89, 01 2001.

[37] Simon N. Wood. mgcv: Mixed gam computation vehicle with automatic smoothness estimation, 2023. R package version 1.8-40, `https://CRAN.R-project.org/package=mgcv`.

[38] Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, 2023. R package version 3.4.4, `https://ggplot2.tidyverse.org`.

[39] Jacolien Van Rij, Martijn Wieling, R. Harald Baayen, and Hedderik Van Rijn. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs, 2022. R package version 2.4.1.

ORCID ID

Austin D. Semmel https://orcid.org/0000-0001-8679-9295
H. Sebastian Heese https://orcid.org/0000-0002-4897-9655
Brandon M. McConnell https://orcid.org/0000-0003-0091-215X

A BACKGROUND DATA & MODEL RESULTS

Table 3. OR and Hours until Phase (Model Data).

	OR			
Hours until Phase	**High**	**Medium**	**Low**	**Total**
High	25,522	25,953	10,527	62,002
	(9.6%)	(9.8%)	(4.0%)	(23.4%)
Medium	65,069	55,881	22,286	143,236
	(24.5%)	(21.0%)	(8.4%)	(54.0%)
Low	29,698	22,407	8,129	60,234
	(11.2%)	(8.4%)	(3.1%)	(22.7%)
Total	120,289	104,241	40,942	265,472
	(45.3%)	(39.3%)	(15.4%)	(100%)

Table 4. OR and Bank Hour Percentage (Battalion Aggregated Data).

	OR			
Bank Hour Category	**High**	**Medium**	**Low**	**Total**
High	964	1,577	1,129	3,670
($>$ 65%)	(3.2%)	(8.5%)	(6.1%)	(19.8%)
Medium	4262	4295	2,583	11,140
(45% $\leq$ x $\leq$ 65%)	(23.1%)	(23.2%)	(14.0%)	(60.3%)
Low	2,424	1,033	220	3,677
($<$ 45%)	(13.1%)	(5.6%)	(1.2%)	(19.9%)
Total	7,650	6,905	3,932	18,487
	(41.4%)	(37.3%)	(21.3%)	(100%)

Table 5. Outlier Analysis by Battalion.

Battalion	**Outliers**	**Total Obs**	**Outlier Percentage**
A	441	17123	2.58%
B	363	13002	2.79%
C	250	12483	2.00%
D	364	14494	2.51%
E	389	14391	2.70%
F	6	11930	0.05%
G	418	6341	6.59%
H	257	16851	1.53%
I	226	7502	3.01%
J	199	17079	1.17%
K	346	10450	3.31%
L	306	10768	2.84%
M	380	13561	2.80%
N	401	14984	2.68%
O	352	14669	2.40%
P	317	24648	1.29%
Q	333	16680	2.00%
R	417	13924	2.99%
S	366	12392	2.95%

Note: Outliers found via Cook's Distance.

Table 6. Spline Contribution and Odds Ratio Comparisons (Full Model D).

Measure	Value	Contribution	Odds Ratio	Difference from Max
OR	0.9	-0.651	0.343	-0.22
	0.75	-0.0885	0.492	-0.071
	0.6	0.252	0.563	–
Hours until Phase	10	-0.233	0.442	-0.088
	250	-0.022	0.494	-0.036
	400	0.119	0.530	–
Days to Report	1	-0.144	0.465	-0.058
	15	-0.085	0.478	0.045
	25	0.091	0.523	–

Table 7. GAM with Tensor Product Spline Model Output Summary.

Term	**Estimate**	**Std. Error**	**p-Value**
(Intercept)	0.529	0.047	<0.001
Year 2020	-0.324	0.015	<0.001
Year 2021	-0.311	0.015	<0.001
Year 2022	-0.241	0.018	<0.001
February	0.106	0.016	<0.001
March	0.110	0.016	<0.001
April	0.098	0.016	<0.001
May	0.039	0.016	0.013
June	-0.028	0.018	0.113
July	-0.037	0.018	0.037
August	0.142	0.018	<0.001
September	0.135	0.018	<0.001
October	-0.023	0.017	0.173
November	-0.088	0.017	<0.001
December	-0.498	0.018	<0.001
Sunday	-1.614	0.014	<0.001
Monday	-0.243	0.012	<0.001
Tuesday	0.014	0.011	0.228
Thursday	-0.213	0.011	<0.001
Friday	-0.993	0.013	<0.001
Saturday	-1.628	0.014	<0.001
Approximate Sig. of Smooth Terms	**EDF**	**Ref. df**	**p-Value**
OR	3.688	3.750	<0.001
Hours until Phase	3.994	4.000	<0.001
Days Until Report	1.540	1.788	<0.001
Hours until Phase $\otimes$ OR	18.851	19.935	<0.001
Battalion	17.873	18.000	<0.001

Adjusted R^2 = 0.113; Explained Deviance = 9.81%; n = 265,472

B SUPPLEMENTARY MATERIAL

Algorithm 1 Impute Aircraft Hours Until Phase Maintenance

```
for each aircraft $j$ do
    Calculate total observed flying hours over dataset duration
    Estimate number of 500-hour phase maintenance cycles
    Identify top five longest consecutive NMC days
    for each NMC period do
        Calculate flying hours leading up to the NMC period
        Determine the probability of the NMC being a phase maintenance cycle,
        considering both flying hours and duration of downtime
        if probability indicates likely phase maintenance then
            Allocate as a phase maintenance cycle
            Reset phase maintenance cycle to 500 hours
        end if
    end for
    for each subsequent flight of aircraft $j$ do
        Decrement hours remaining in phase maintenance cycle based on flight duration
    end for
end for
```

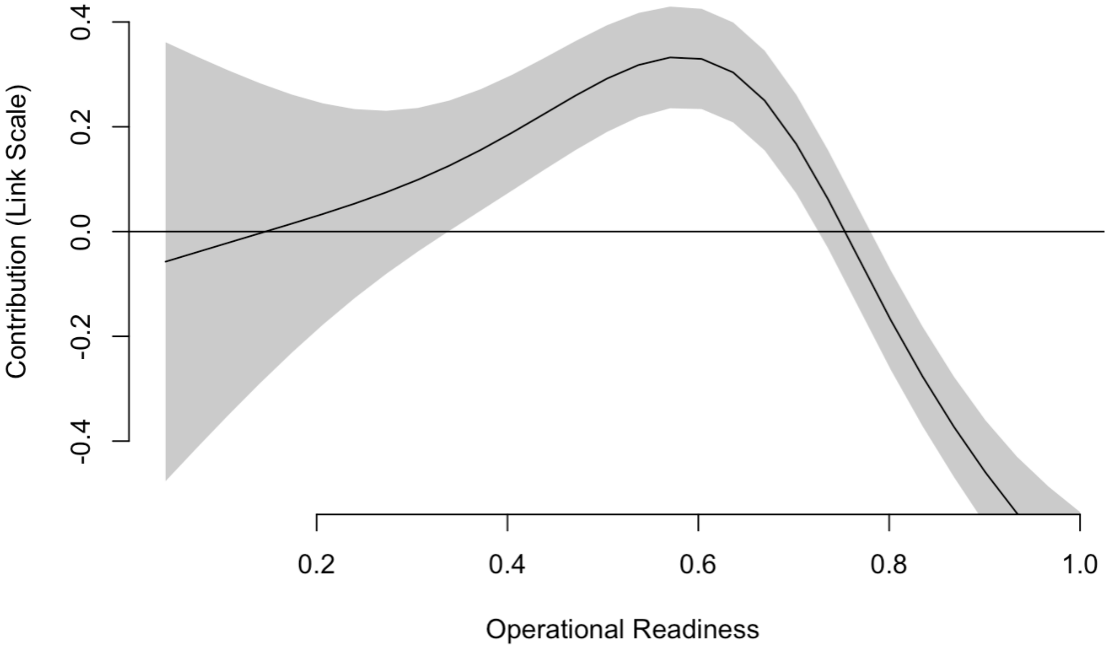

Figure 9. Spline for OR (Tensor Model) with 95% Credible Interval.

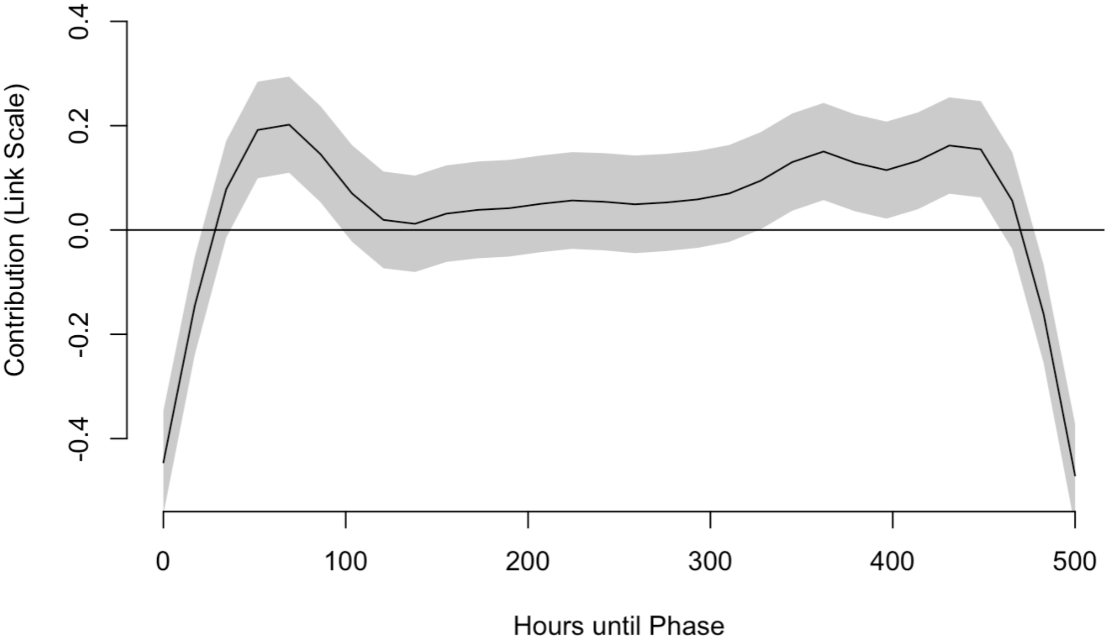

Figure 10. Spline for Hours until Phase (Tensor Model) with 95% Credible Interval.

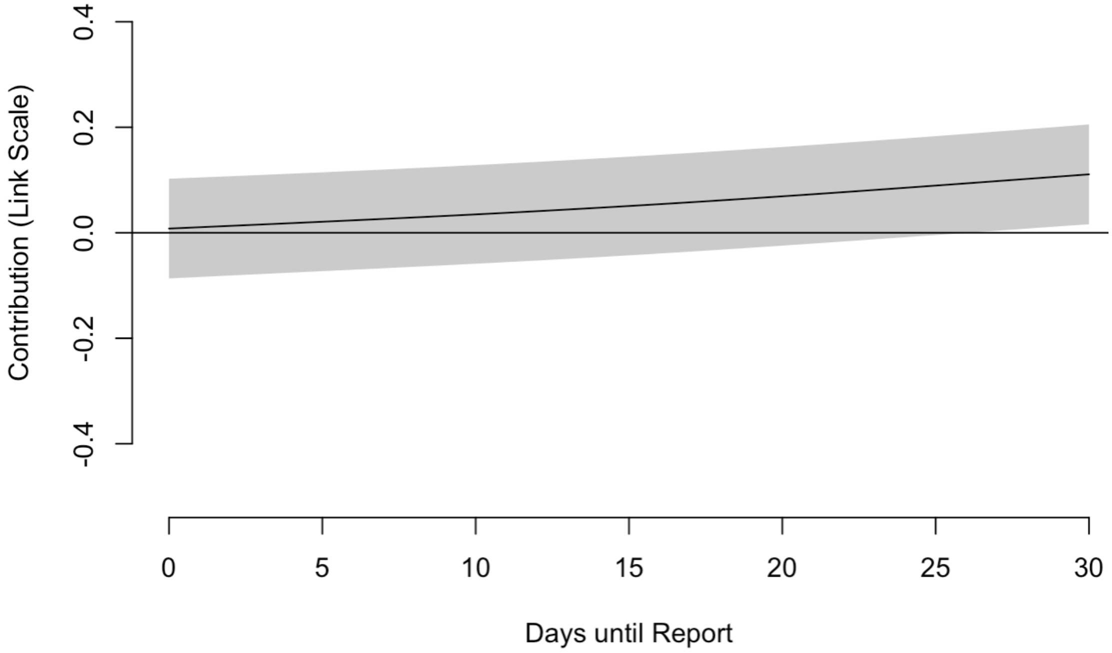

Figure 11. Spline for Days until Report (Tensor Model) with 95% Credible Interval.

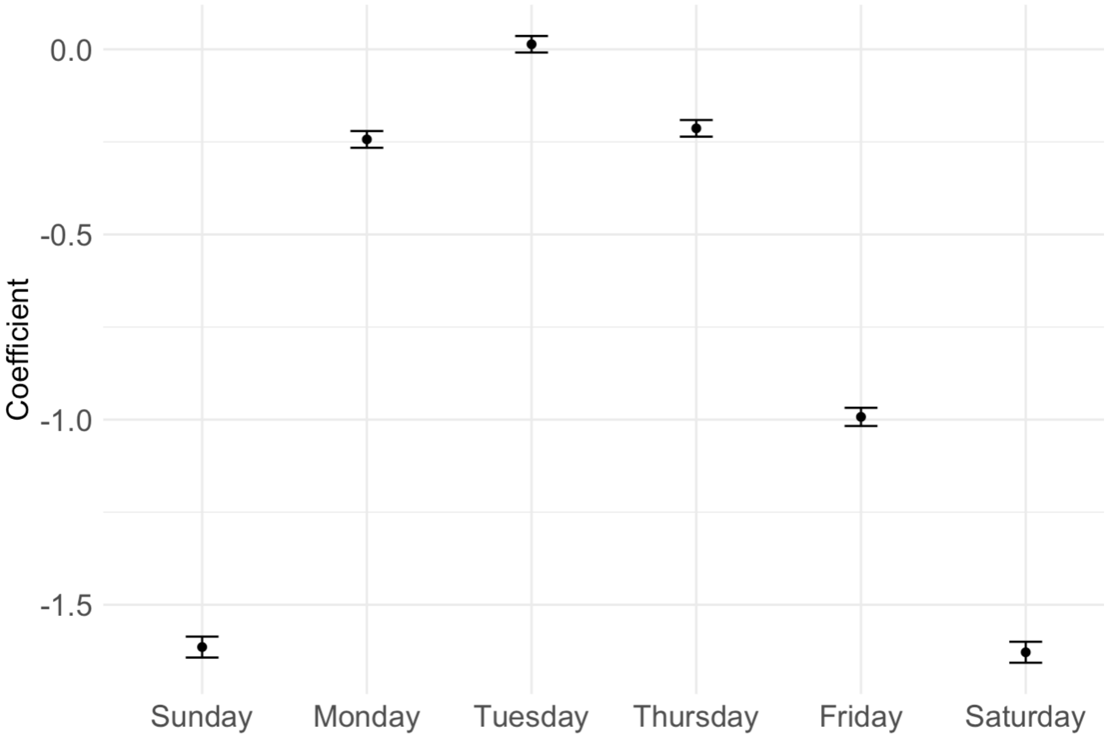

Figure 12. Estimated Fixed Effect of Day of the Week (Tensor Model) with 95% confidence intervals.

C MODEL ASSUMPTIONS AND DIAGNOSTICS

C.1 Model Assumptions and Diagnostics

In modeling both GAMs, we make necessary model assumptions and conduct diagnostics to ensure their validity.

Tests for multicollinearity among covariates indicated no significant issues. Table 8 shows the Spearman and Pearson correlation coefficients for each two-way combination of covariates. Numbers approaching ± 1 indicate near-perfect collinearity in the data—we observe no values greater than 0.05, indicating a lack of linear relationship between the variables. *Day of the week* versus *Days until report* is intentionally omitted as it is completely determined by a calendar alone and provides no meaningful interpretation. The linearity of covariates and the impact of outliers are verified through diagnostic plots and Cook's distance, respectively, detailed in Table 5 in Appendix A. We do not omit outliers at any point during the modeling stage.

Table 8. Correlation Coefficients between Independent Variables.

Variable 1	**Variable 2**	**Pearson**	**Spearman**
OR	Day of the Week	0.0004	0.0018
OR	Hours until Phase Maint.	-0.0484	-0.0477
OR	Days until Report	0.0119	0.0116
Hours until Phase Maint.	Day of the Week	0.004	0.0041
Hours until Phase Maint.	Days until Report	0.0041	0.0041

An additional test for concurvity, the non-linear analog to multicollinearity, indicates a lack of correlation between the smoothing spline terms. Table 9 below illustrates how removing the Battalion random effect term eliminates concurvity concerns, as many units tend to exhibit distinct flying personalities (never allowing OR to drop below 75%).

Again, low values indicate a lack of a non-linear relationship between covariates. In certain units, low OR rarely or never occurs and can exhibit increased uncertainty in these outlier scenarios.

We select the optimal number of knots for each smoothing term by first examining the BIC of each model over different k values according to standard methodologies from Cantoni and Hastie [36] and implemented in Wood [37]. We do so for each smoothing term individually and select K = 4 knots for OR; K = 5 knots for hours until phase; and K = 5 knots for day of the week. Next, we perform a k-Index test to verify if additional complexity is required. Results for this process are collectively found in Table 10 and Figure 13.

High concurvity here is to be expected given that OR and hours until phase main effect splines are still included. Model interpretability and generalization suffer if they are removed, however. Further, as seen in Model 1, the presence of the random effect induces a high degree of concurvity.

All analyses were performed using R Statistical Software (v4.3.1; R Core Team 2023). Data visualization from `ggplot2` and `itsadug` R packages [38, 39], while statistical modeling and cross-validation were conducted with the `mgcv` R package [37].

Table 9. Concurvity Measures for GAM Models with and without Battalion Random Effect (Equation 1).

Term	With Random Effect	Without Random Effect
	Worst	
Hours until Phase	0.10	0.09
OR	0.87	0.09
Days until Report	0.02	0.01
Battalion	1.00	-
	Observed	
Hours until Phase	0.03	0.01
OR	0.19	0.08
Days until Report	0.00	0.00
Battalion	0.02	-
	Estimate	
Hours until Phase	0.08	0.06
OR	0.68	0.06
Days until Report	0.00	0.00
Battalion	0.07	-

Table 10. k-Index Test for knot complexity in GAM spline terms (Section 3.3, Equation 1) indicates an increased quantity of knots, k' is not warranted.

Term	k'	edf	k-index	p-value
OR	4.0	3.0	0.89	0.55
Hours Until Phase	5.0	4.0	0.88	0.26
Days Until Report	5.0	4.0	0.88	0.17

Table 11. Concurvity Measures for Tensor Product GAM.

Term	Worst	Observed	Estimate
Days Until Report	1.0	1.0	1.0
Hours Until Phase	0.0052	0.0041	0.0038
OR	0.999	0.999	0.999
Hours Until Phase $\otimes$ OR	0.999	0.999	0.998
Battalion (Random Effect)	1.0	0.0291	0.081

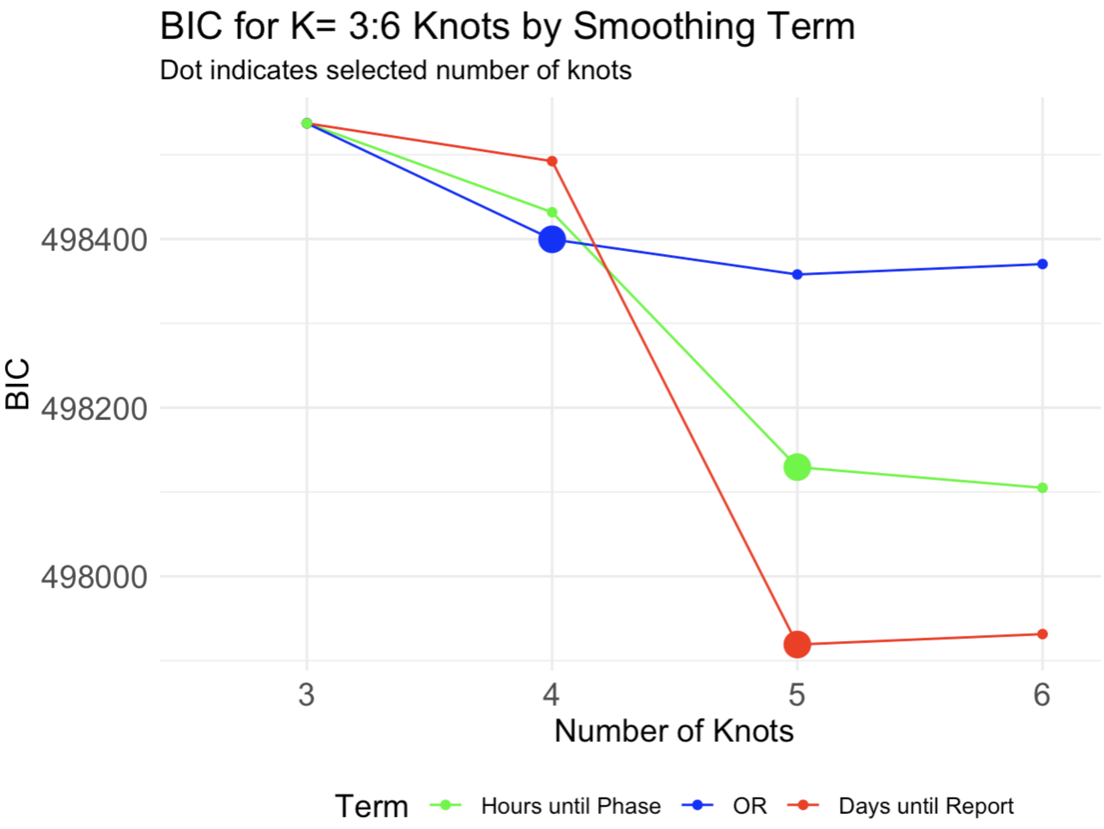

Figure 13. (color online) Optimal number of knots chosen via elbow method using model Bayesian Information Criterion (Equation 1).