Bayesian estimation of the mean time between failures of subsystems with different causes using interval-censored system maintenance data

David Han[1]*, James D. Brownlow[2], Jesse Thompson[2], and Ralph G. Brooks[2]

[1] *Department of Management Science and Statistics, the University of Texas at San Antonio, Texas, USA*

[2] *Air Force Test Center at 412th Test Wing, Edwards Air Force Base, California, USA*

ABSTRACT

Ensuring an acceptable level of reliability stands as a primary imperative for any mission-focused operation since it serves as a critical determinant of success. Inadequate reliability can lead to severe repercussions, including substantial expenses for repairs and replacements, missed opportunities, service disruptions, and, in the worst cases, safety violations and human casualties. Within national defense organizations such as the USAF, the precise assessment and maintenance of system reliability play a pivotal role in ensuring the success of mission-critical operations. In this research, our primary objective is to model the reliability of repairable subsystems within the framework of competing and complementary risks. Subsequently, we construct the overall reliability of the entire repairable system, utilizing day-to-day group-censored maintenance data from two identical aircraft systems. Assuming that the lifetimes of subsystems follow non-identical exponential distributions, it is theoretically justified that the system reliability can be modeled by homogeneous Poisson processes even though the number of subsystems of any particular type is unknown and the temporal order of multiple subsystem failures within a given time interval is uncertain due to interval censoring. Using the proposed model, we formulate the likelihood function for the mean time between failures of subsystems with different causes, and subsequently establish an inferential procedure for the model parameters. Given a considerable number of parameters to estimate, we explore the efficacy of a Bayesian approach, treating the contractor-supplied estimates as the hyperparameters of prior distributions. This approach mitigates potential model uncertainty as well as the practical limitation of a frequentist-based approach. It also

---

*Corresponding author. *phone*: +1-210-458-7895; *eMail*: david.han@utsa.edu

facilitates continuous updates of the estimates as new maintenance data become available. Finally, the entire inferential procedures were implemented in Microsoft Excel so that it is easy for any reliability practitioner to use without the need to learn sophisticated programming languages. Thus, this research supports an ongoing, real-time assessment of the overall mission reliability and helps early detection of any subsystem whose reliability is below the threshold level.

Keywords: Bayesian analysis, Competing risks, Interval monitoring, Mean time between failures, Poisson process, Preventive maintenance

1 Introduction

In the realm of industrial and manufacturing engineering, quality is defined as the suitability for a specific purpose, while reliability represents the enduring quality, gauged by a device's performance over time. Sustaining an acceptable level of reliability stands as a critical responsibility for any industrial or mission-driven organization, as it profoundly impacts the outcome of missions or operations. Inadequate reliability can result in severe repercussions such as substantial repair and replacement expenses, missed business opportunities, service interruptions, safety violations, and even human casualties. An illustrative example of this is the 2013 Chrysler recall of 2.7 million Jeep vehicles following 51 fatalities caused by rear-end collisions; see Eis (2013). To safeguard the gas tanks, they had to retrofit trailer hitches. A few years later, Chrysler recalled 4.8 million vehicles in the United States alone, an astounding six-fold increase from its entire 2013 recalls; see Limer (2018). More recently, Nidumolu (2023) reported that Tesla, a leading electric vehicle manufacturer, recalled 55,000 of its Model X vehicles due to a potential brake fluid safety issue, which would increase the risk of a crash. In the same year, it also recalled more than one million vehicles after discovering issues in the acceleration and braking systems; see Goksedef (2024). This vividly demonstrates the criticality of reliability in the automotive industry. Similarly, in the semiconductor manufacturing sector, equipment failures can bring an entire factory to an abrupt halt, incurring associated costs that can escalate to a staggering $150,000 per hour; see Chan (2020). To put this into perspective, it is worth noting that in 2003, services and spare parts constituted 8% of the entire U.S. GDP, as reported by Lengu, et al. (2014). Moreover, in 2006, estimates by the US Bancorp revealed that expenditures on spare parts in the United States alone, excluding maintenance, services, and downtime costs, reached a staggering $700 billion. These figures underscore the economic significance of reliability in various industrial domains. A recent survey of executives conducted by Henke, et al. (2016) underscores the paramount importance of asset management and the associated risks. Executives widely identified the failure of critical assets as a paramount concern. Among the best-in-class companies, it was found that they leverage real-time data to optimize their decision-making processes, emphasizing the evolving role of data in ensuring reliability. For national defense organizations, including the United States Air Force (USAF), the precise estimation and maintenance of system reliability is absolutely indispensable for the successful execution of mission-critical operations. The significance of reliability in this context cannot be overstated, as lives and national security often hinge on the dependable performance of critical assets.

In this work, we aim to model the reliability of repairable subsystems within a system under the competing and complementary risk frameworks, and subsequently construct the overall reliability of the entire repairable system, using the day-to-day group-censored maintenance data of two identical aircraft systems in operation. Conducting a reliability assessment is a standard requirement in defense test and maintenance programs. This requirement is typically fulfilled using a frequentist approach, where the mean time between failures (MTBF) is determined by dividing the total test duration by the number of failures. The decision on the system's suitability is made by comparing the lower confidence bound of MTBF to the MTBF threshold. However, this method encounters challenges when dealing with small data sets, limited test durations, or no recorded failures. Confidence intervals become wider with small data sets, and estimating MTBF becomes impossible when there are no failures. Even if one assumes a high MTBF due to a lack of failures, the lower bound calculation using an exponential distribution can yield information of minimal value for decision-makers. To address these issues and provide a more comprehensive assessment with intuitive interpretations of results, the Bayesian analysis is considered a more appropriate approach; see Berger (2006). In the field of reliability engineering, researchers have explored the application of Bayesian estimation methods to determine the system reliability affected by various failure causes as well as for handling interval-censored data. For example, Sen, et al. (2010) introduced a semiparametric Bayesian approach that allows for a variety of prior specifications while Xu and Tang (2011) considered a nonparametric Bayesian approach to series systems with partially masked competing risks. More recently, Pan, et al. (2020) carried out a study on the Bayesian approach for analyzing partially interval-censored data within a proportional hazards model. Song, et al. (2021) utilized a Bayesian network based on copula method and probability-box for reliability assessment of failure-dependent and uncertain systems. Using a Bayesian approach with a conjugate continuous prior, Ahmadi and Doostparast (2022) developed and evaluated the lifetime performance index of products under progressive Type-

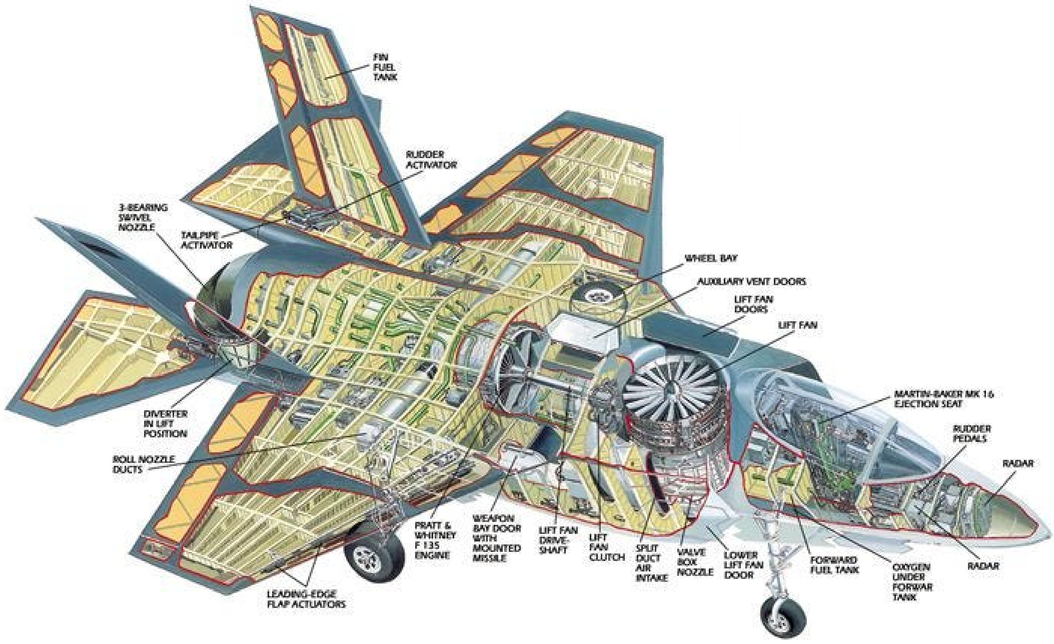

Figure 1: Specifications of the Lockheed Martin F-35 Lightning II for illustration of various subsystems of an aircraft; adapted from Hollings (2023)

II censoring. Dutta and Kayal (2022) developed Bayesian and non-Bayesian inferential methods for Weibull lifetimes based on partially observed competing risks data under hybrid censoring scheme. Bayesian analysis of designed reliability improvement experiments was considered by Wang, et al. (2024) and they proposed adaptive test termination by estimating and comparing model coefficients under different experiment times. To choose suitable D-optimal designs in regular and irregular design regions, Taylor, et al. (2024) explored the Bayesian approach on various lifetime regression models with censoring. Now, the following describes an aircraft systems maintenance dataset, which motivated the research presented in this work.

The USAF Test Center provided a dataset which contains the day-to-day scheduled and unscheduled maintenance activities for two identical repairable aircraft systems along with the daily flight hours each system has completed. The maintenance activities for the system #1 run from November 15, 2016 up until September 8, 2020 while they run from May 8, 2017 up until June 13, 2022 for the system #2. Each system consists of a number of subsystems; see Figure 1 for graphical illustration. It is known that there are 29 types of non-identical subsystems although the exact number of subsystems for each type is unknown for this analysis. A failure of each subsystem is discovered during the maintenance period after a flight mission or when a failure state is suspected or notified by a pilot after the flight. The failure is diagnosed as one of the 3 known and mutually exclusive failure types, which are

- **Type 1:** inherent/natural failure;
- **Type 2:** induced failure, caused by an outside/external influence;
- **Type 6:** often associated with the CND (could not duplicate) failures, which occur when a failure is indicated or suspected but when the maintenance crews attempt to verify the failure, the failure cannot be duplicated during troubleshooting and hence, no repair or replacement actions can be taken; Type 6 failure also refers to a situation where the maintenance crews had to remove a part due to a suspected failure but later a failure root cause analysis could not confirm the failure. This is commonly known as *no-defect* failure in the maintenance practice.

For the purpose of modeling, it is assumed that any reported failure is fully repaired in order to bring the system back to the state of *as-good-as-new* before the next flight assignment. This means that any failure recorded on a day must have occurred during the flight hours on the same day. If there is no flight operation on the day a failure is corrected, the actual failure must have occurred during the previous flight hours on the day closest to the day a failure is corrected. That is, potential damage or wear-out of any subsystem is assumed to take place only during the flight operation but not during the standby period on land. Since the exact failure time is unknown in any case, it is interval-censored during the cumulative flight hours of a day on which the failure is suspected to have occurred.

Initially, under the classical homogeneous Poisson process for the failure of a repairable system with perfect maintenance, it was assumed that these subsystems are connected in series such that the lifetime of a whole system is the minimum among all the subsystem lifetimes (*i.e.*, a competing risk structure), which still follows an exponential distribution. However, the dataset features occasional multiple subsystem failures which occurred after the same flight operation, and the temporal order of these multiple subsystem failures is not known due to interval censoring. This indicates that a subsystem failure may not be *hard* or catastrophic. It could mean operation under a degraded or sub-optimal state. As this violates the assumption of a series connection of the subsystems, an alternative model for the system reliability had to be sought. Since the whole system could still fail physically by a single (hard) failure of a critical (but unknown) subsystem, the reliability of the whole system could not be modeled according to a k-out-n system either. It is noted that not all subsystem failures are critical as some are more important than others but this information is not known for analysis. Thus, a parallel connection of the subsystems is assumed (*i.e.*, a complementary risk structure) although the whole system could fail before all the subsystems fail. This approach seems reasonable in this case as no catastrophic system failure is observed from the dataset with frequent maintenance activities to fix any subsystem failure and bring the entire system back to a good working order. In other words, all the observed subsystem failures are from non-critical subsystems, which could be relatively less reliable since the critical subsystems need to be more reliable and they are unlikely to fail during the system lifetime.

With the objective to model the system reliability from the subsystem level, the rest of the paper is organized as follows. Section 2 presents the general model formulation under interval/group censoring. Modeling the subsystem lifetime under the competing risk structure of different failure types, theoretical justification of the final model is provided via asymptotics. In Section 3, the likelihood function is constructed and it is used to estimate the MTBF for each subsystem type by each failure type, based on the maximum likelihood principle. Realizing its shortcomings, Bayesian approach is then considered with informative conjugate priors for the parameter estimation and the credible intervals. The Bayes factor is introduced in Section 4 for examining whether the contractor-supplied MTBF overestimates the actual MTBF based on the field maintenance data. Section 5 illustrates an application of the procedures developed in this article to the maintenance activity data of the aircraft systems described in Section 1. Concluding remarks are provided in Section 6 along with certain limitations of the methodology as well as the direction of future research. For simplicity, no notational distinction is made in this article between the random variables and their corresponding realizations. Also, we adopt the usual conventions that $\sum_{j=m}^{m-1} a_j \equiv 0$ and $\prod_{j=m}^{m-1} a_j \equiv 1$.

2 Model Formulation

For modeling the total system reliability, let n denote the number of s-independent, identical, and repairable systems while n_{sub} denotes the number of different subsystem types of a single system and n_f denotes the number of different failure types for a subsystem (*e.g.*, $n = 2$, $n_{sub} = 29$, and $n_f = 3$). For $i = 1, 2, \ldots, n$ and $l = 1, 2, \ldots$, let $\tau_{li} \geq 0$ be the total flight hours completed by the system i on the l-th day, and let $\mathscr{S}_{li}$ denote the random index set of failed subsystem type(s) of the system i on the l-th day such that $\mathscr{S}_{li} \subseteq \{1, 2, \ldots, n_{sub}\}$. Furthermore, let $n_{li} = |\mathscr{S}_{li}|$ define the cardinality of the index set $\mathscr{S}_{li}$, or the number of different subsystem types that must have failed during τ_{li} flight hours of the system i on the l-th day. When there is no flight completed (*viz.*, $\tau_{li} = 0$) or there are no subsystem failures, $\mathscr{S}_{li}$ is empty and $n_{li} = 0$.

The maintenance dataset also features multiple failures of the same subsystem type after a flight operation either with different or same failure types. This indicates that there are a (unknown) number of subsystems of the same subsystem type. Let N_{ij} denote the number of subsystems of type j within the system i. Then, for $j = 1, 2, \ldots, N_{ij}$, let $\mathscr{S}_{lij}$ denote the random multiset (or bag) of failure types of the failed subsystems of type j within the system i on the l-th day. It should be noted that a multiset $\mathscr{S}_{lij}$ allows for multiple instances for each of its elements unlike the index set $\mathscr{S}_{li}$ which is a proper set or a multiset having multiplicity of 1 only. For instance, $\mathscr{S}_{315} = [1, 1, 2, 6]$ means that on the third day, 4 subsystems of type 5 in the system 1 failed. The failure type of the two of them is 1 while the other two have different failure types, 2 and 6. Furthermore, let $n_{lij} = |\mathscr{S}_{lij}|$ define the cardinality of the multiset $\mathscr{S}_{lij}$, or the number of failures of the subsystem type j that must have occurred during τ_{li} flight hours of the system i on the l-th day. Also, let n_{lijk} define the multiplicity of the failure type k in the multiset $\mathscr{S}_{lij}$, or the number of failure type k for the subsystem type j that must have failed during τ_{li} flight hours of the system i on the l-th day. For instance, the multiset $\mathscr{S}_{315} = [1, 1, 2, 6]$ in the previous example provides $n_{3151} = 2$, $n_{3152} = 1$, and $n_{3156} = 1$ such that $n_{315} = 4$. Hence, it is obvious that $n_{lij} = \sum_{k=1,2,6} n_{lijk}$, and when there is no flight completed (*viz.*, $\tau_{li} = 0$) or there are no failures of the subsystem type j, $\mathscr{S}_{lij}$ is empty and $n_{lij} = n_{lijk} = 0$ for all $k \in \{1, 2, 6\}$. Thus, if $\mathscr{S}_{li}$ is empty, $\mathscr{S}_{lij}$ is also empty for all $j = 1, 2, \ldots, n_{sub}$.

Now, let us define the probability of observing this day-to-day failure dataset for n independent and identical repairable aircraft systems with perfect maintenance. Since any reported failure is repaired in order to bring the system back to the state of *as-good-as-new* before the next flight assignment, the daily flight hours for each system are determined without the need to consider the historic maintenance activities or the prior failure records collected daily. This means that for the modeling purpose, the daily flight hours for each system can be treated as pre-fixed or deterministic quantities without any relevance to the stochastic failure data. Denoting n_d to be the index for the last recorded day of this maintenance dataset, the probability of observing this day-to-day failure dataset composed of $\mathscr{S}_{li}$'s and $\mathscr{S}_{lij}$'s can be expressed as

$$
\begin{aligned}
& P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l = 1, 2, \ldots, n_d,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n_{sub}\right) \\
= & \prod_{i=1}^{n} P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l = 1, 2, \ldots, n_d,\ j = 1, 2, \ldots, n_{sub}\right) \\
= & \prod_{i=1}^{n} P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l = 1, 2, \ldots, n_d,\ j \in \mathscr{S}_{li}\right) \\
= & \prod_{l=1}^{n_d} \prod_{i=1}^{n} P\left(\mathscr{S}_{li}, \mathscr{S}_{lij},\ j \in \mathscr{S}_{li};\ \tau_{li} \,\middle|\, \mathscr{S}_{l'i}, \mathscr{S}_{l'ij'},\ j' \in \mathscr{S}_{l'i};\ \tau_{l'i},\ l' = 1, 2, \ldots, l-1\right)
\end{aligned} \tag{1}
$$

given τ_{li} for $l = 1, 2, \ldots, n_d$, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, n_{sub}$. The first equality is due to the s-independence of the systems, and the second equality is due to the fact that $\mathscr{S}_{lij}$ is null for $j \notin \mathscr{S}_{li}$.

The third equality is based on the generalized multiplication rule of conditional probabilities.

In order to concretize the abstract probability model in (1), the failure time of a subsystem by each failure type/mode is assumed to follow non-identical exponential distribution with an individual rate parameter to specify each distribution. Formally, let $T_{j;k}$ be the lifetime of a subsystem of type j by the k-th failure type only (in the absence of other failure types/modes) for $j = 1, 2, \ldots, n_{sub}$ and $k = 1, 2, 6$. Please note that for brevity of expressions, we are ignoring the system index i and the flight time index l for now. This is reasonable since the systems are independent and their physical parameters are identical when manufactured and put on service. Also, the distribution model being constructed here does not depend on the flight duration yet. Then, the probability density function (PDF) and cumulative distribution function (CDF) of $T_{j;k}$ are defined by

$$\begin{aligned} f_{j;k}(t) &= \lambda_{jk} \exp\left(-\lambda_{jk} t\right), & 0 < t < \infty; \\ F_{j;k}(t) &= 1 - S_{j;k}(t) = 1 - \exp\left(-\lambda_{jk} t\right), & 0 < t < \infty, \end{aligned}$$

respectively, where $\lambda_{jk} > 0$ is the failure rate parameter for the subsystem type j and the failure type k. Then, the mean time to failure (MTTF) of the subsystem type j by the failure type k only (in the absence of other failure types/modes) is simply its reciprocal, $1/\lambda_{jk}$.

Since any observed failure is associated with only one of the failure types under consideration, the competing risk framework can be used to model the actual failure time of a subsystem by each failure type/mode in the presence of other failure types/modes. Assuming that $T_{j;k}$'s are s-independent, let T_{jk} be the lifetime of a subsystem of type j by the failure type k (in the presence of other failure types/modes) for $j = 1, 2, \ldots, n_{sub}$ and $k = 1, 2, 6$. Then, using the PDF and CDF of $T_{j;k}$, the PDF and CDF of T_{jk} are derived as

$$\begin{aligned} f_{jk}(t) &= f_{j;k}(t) \prod_{k' \neq k} S_{j;k'}(t) = \lambda_{jk} \exp\left(-\lambda_j t\right), & 0 < t < \infty; \\ F_{jk}(t) &= \pi_{jk} F_j(t) = \pi_{jk}\big(1 - \exp\left(-\lambda_j t\right)\big), & 0 < t < \infty, \end{aligned} \tag{2}$$

where $\lambda_j = \sum_{k=1,2,6} \lambda_{jk}$ is the total failure rate for the subsystem type j, and

$$F_j(t) = 1 - S_j(t) = 1 - \exp\left(-\lambda_j t\right), \qquad 0 < t < \infty \tag{3}$$

is the CDF of $T_j = \min_{k=1,2,6}\{T_{j;k}\}$, the lifetime of a subsystem of type j regardless of the failure types under the competing risk framework. Here, $\pi_{jk} = \lambda_{jk}/\lambda_j \in (0, 1)$ is known as the *relative risk* of failure of the subsystem type j caused by the failure type k (in the presence of other failure types/modes); see Crowder (2001). It should be noted that π_{jk} is a constant in this case, not relying on the failure time

t due to the property of the underlying risk factors following exponential distributions with constant hazard rates.

As stated before, the exact failure time is unobservable but any failure observed is assumed to have occurred during the flight hours on the same day. That is, the failure times are interval-censored or group-censored, allowing only the failure count data to be available for inference. Given that an aircraft system i has already flown for the time duration $\tau > 0$ in total, let X_{ijk} denote the number of failures of the subsystem type j by the failure type k (in the presence of other failure types/modes) during the additional flight time Δ (*viz.*, between τ and $\tau + \Delta$). Since it is assumed that any reported failure is fully repaired before the next flight assignment, there are N_{ij} number of subsystems of type j at risk of failure in the beginning of flight at time τ (ignoring any time spent on land between flights). Assuming that the failures of the subsystem type j are s-independent and identically distributed, it is easy to see that $(X_{ijk})_{k=1,2,6}$ follows a multinomial distribution with the parameters N_{ij}, p_{ijk} for $k = 1, 2, 6$, where $p_{ijk} \in [0, 1]$ denotes the probability of failure of the subsystem type j by the failure type k during the additional flight time between τ and $\tau + \Delta$, given that the system i has already flown for τ. Formally, using (2) and (3), it can be expressed as

$$\begin{aligned} p_{ijk} &= P(T_{jk} \leq \tau + \Delta \mid T_{jk'} > \tau, \text{ for } k' = 1, 2, 6) \\ &= P(T_{jk} \leq \tau + \Delta \mid T_j > \tau) = \frac{P(\tau < T_{jk} \leq \tau + \Delta)}{P(T_j > \tau)} \\ &= \frac{F_{jk}(\tau + \Delta) - F_{jk}(\tau)}{S_j(\tau)} = \frac{\pi_{jk}\big(F_j(\tau + \Delta) - F_j(\tau)\big)}{S_j(\tau)} \\ &= \pi_{jk}\left(1 - e^{-\lambda_j \Delta}\right) = \pi_{jk} F_j(\Delta) \end{aligned}$$

so that $\sum_{k=1,2,6} p_{ijk} = F_j(\Delta)$. Because of the memoryless property of the exponential distribution, it is noted that p_{ijk} does not depend on τ, the starting time of the next flight assignment. This has a very important implication in the modeling process. The above definition for p_{ijk} assumes that the subsystem of type j has been operating/surviving without any failure record by time τ. However, if the subsystem has a failure history before τ, the definition of p_{ijk} needs to change. Since any reported failure is repaired in order to bring the system back to the state of *as-good-as-new* before the next flight assignment, this resets τ to zero for such subsystems. Since p_{ijk} does not depend on τ (but on Δ), the definition of p_{ijk} remains the same even in that case, which means that the failures of the subsystem type j are indeed identically distributed and in turn, it validates the multinomial distribution deduced above. To emphasize the dependence of p_{ijk} on the flight duration Δ, let us re-express p_{ijk} as $p_{ijk}(\Delta)$, a function of Δ.

Returning to (1), the probability of observing the day-to-day failure dataset composed of $\mathscr{K}_{li}$'s and $\mathscr{S}_{lij}$'s can be further derived as

$$
\begin{aligned}
& P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l=1,2,\ldots,n_d,\ i=1,2,\ldots,n,\ j=1,2,\ldots,n_{sub}\right) \\
=& \prod_{l=1}^{n_d}\prod_{i=1}^{n}\prod_{j=1}^{n_{sub}} \binom{N_{ij}}{n_{lij1}, n_{lij2}, n_{lij6}} \left[\prod_{k=1,2,6} p_{ijk}^{n_{lijk}}(\tau_{li})\right] \left[1-\sum_{k=1,2,6} p_{ijk}(\tau_{li})\right]^{N_{ij}-n_{lij}} \\
=& \prod_{l=1}^{n_d}\prod_{i=1}^{n}\prod_{j=1}^{n_{sub}} \binom{N_{ij}}{n_{lij1}, n_{lij2}, n_{lij6}} \left[\prod_{k=1,2,6} p_{ijk}^{n_{lijk}}(\tau_{li})\right] \left[S_j(\tau_{li})\right]^{N_{ij}-n_{lij}}
\end{aligned}
\tag{4}
$$

given the flight durations τ_{li}'s for $l = 1, 2, \ldots, n_d$, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, n_{sub}$. The first equality is based on the assumption of the s-independence of the subsystems within a system. Going further, it is reasonable to assume that N_{ij}, the number of subsystems of type j within the system i is substantially large even if it is unknown. In that case, Arenbaev (1976) has shown that a multinomial random vector asymptotically behaves like a vector of independent and non-identically distributed Poisson random variables, whose mean parameters are determined by the marginal means of the multinomial random vector. Applying this result, (4) can be approximated as

$$
\begin{aligned}
& P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l=1,2,\ldots,n_d,\ i=1,2,\ldots,n,\ j=1,2,\ldots,n_{sub}\right) \\
\approx & \prod_{l=1}^{n_d}\prod_{i=1}^{n}\prod_{j=1}^{n_{sub}}\prod_{k=1,2,6} \frac{\mu_{lijk}^{n_{lijk}}}{n_{lijk}!}\exp\left(-\mu_{lijk}\right),
\end{aligned}
\tag{5}
$$

where $\mu_{lijk} = N_{ij}\ p_{ijk}(\tau_{li})$. For achieving the mission reliability, τ_{li}, a daily flight duration of the aircraft system i, is considered much smaller than the MTTF of any subsystem by any failure type (*i.e.*, $\tau_{li} << 1/\lambda_{jk}$ for all j and k at any given l and i). As a matter of fact, the maximum value of τ_{li} in the dataset was found to be less than 5% of the contractor-supplied mean time estimate of any subsystem by any failure type. This ensures that $p_{ijk}(\tau_{li})$ stays very small in magnitude and so, it can be further simplified using the first order approximation. That is, $p_{ijk}(\tau_{li}) = \pi_{jk}\left(1-e^{-\lambda_j\tau_{li}}\right) \approx \pi_{jk}\lambda_j\tau_{li} = \lambda_{jk}\tau_{li}$, and using this, (5) can be shown to be

$$
\begin{aligned}
& P\left(\mathscr{S}_{li}, \mathscr{S}_{lij};\ \tau_{li},\ l=1,2,\ldots,n_d,\ i=1,2,\ldots,n,\ j=1,2,\ldots,n_{sub}\right) \\
\propto & \prod_{j=1}^{n_{sub}}\prod_{k=1,2,6} \lambda_{jk}^{\sum_{i=1}^{n}\sum_{l=1}^{n_d} n_{lijk}} \exp\left(-\lambda_{jk}\sum_{i=1}^{n} N_{ij}\sum_{l=1}^{n_d}\tau_{li}\right)
\end{aligned}
\tag{6}
$$

3 Inference for MTBF

3.1 Maximum Likelihood Estimation

Treating (6) as the likelihood function of λ_{jk}'s, the log-likelihood function of $\boldsymbol{\lambda}$, a vector of λ_{jk}'s, can be expressed as

$$\log L(\boldsymbol{\lambda}) = \sum_{j=1}^{n_{sub}} \sum_{k=1,2,6} \left[\left(\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk} \right) \log \lambda_{jk} - \lambda_{jk} \sum_{i=1}^{n} N_{ij} \sum_{l=1}^{n_d} \tau_{li} \right]$$

and by differentiating $\log L(\boldsymbol{\lambda})$ with respect to each λ_{jk} and setting it to zero, the maximum likelihood estimator (MLE) of λ_{jk} is obtained as

$$\widehat{\lambda}_{jk} = \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk}}{\sum_{i=1}^{n} N_{ij} \sum_{l=1}^{n_d} \tau_{li}}$$

It is basically the total number of failures recorded for the subsystem type j by the failure type k across all the aircraft systems divided by the total flight duration of all the subsystems of type j across all the aircraft systems. Another important finding from (5) is that n_{lijk}, the number of failures of the subsystem type j by the failure type k during τ_{li} flight hours of the system i behaves asymptotically as an independent and non-identically distributed Poisson random variable, whose mean parameter is $\mu_{lijk} \approx N_{ij}\ \lambda_{jk}\tau_{li}$. This implies that the time between any two consecutive failures of the subsystem type j by the failure type k is exponentially distributed with the rate/intensity parameter $\eta_{ijk} = N_{ij}\ \lambda_{jk}$. The MTBF of the subsystem type j by the failure type k is then its reciprocal, $\theta_{ijk} = 1/\eta_{ijk}$. Thanks to the invariance property of MLE, the MLE of η_{ijk} is obtained as

$$\widehat{\eta}_{ijk} = N_{ij}\ \widehat{\lambda}_{jk} = N_{ij}\ \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk}}{\sum_{i=1}^{n} N_{ij} \sum_{l=1}^{n_d} \tau_{li}}$$

Since the aircraft systems under this study are identical, N_{ij}'s are all equal for a given subsystem type j. This further reduces the MLE of η_{ijk} (or equivalently, η_{jk} by dropping the system index i) to $\widehat{\eta}_{jk} = \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk}}{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \tau_{li}}$. It is simply the total number of failures recorded for the subsystem type j by the failure type k across all the aircraft systems divided by the total flight duration of all the aircraft systems. Subsequently, the MLE of the MTBF of the subsystem type j by the failure type k is obtained as $\widehat{\theta}_{jk} = \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \tau_{li}}{\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk}}$.

As a side note, it is also easy to see from (4) that n_{lij}, the number of failures of the subsystem type j by any failure type during τ_{li} flight hours of the system i, has a binomial distribution with the parameters N_{ij} and $\sum_{k=1,2,6} p_{ijk}(\tau_{li}) = F_j(\tau_{li})$ through convolution of n_{lijk}'s. Applying the result of Arenbaev (1976) along with the linear approximation, n_{lij} also behaves asymptotically as a Poisson random variable, whose mean parameter is $\mu_{lij} = \sum_{k=1,2,6} \mu_{lijk} = N_{ij}F_j(\tau_{li}) \approx N_{ij}\ \lambda_j\tau_{li}$. This implies that the time between any two consecutive failures (regardless of the failure types) is exponentially distributed with the rate/intensity parameter $\eta_{ij} = N_{ij}\lambda_j = \sum_{k=1,2,6} \eta_{ijk}$ since $\eta_{ijk} = N_{ij}\lambda_{jk}$. Since N_{ij}'s are all equal for a given subsystem type j, η_{ijk} was re-expressed as η_{jk} by dropping the system index

i. Thus, η_{ij} can be also re-expressed as $\eta_j = \sum_{k=1,2,6} \eta_{jk}$ by dropping the system index i. The mean time between unscheduled maintenances (MTBUM) of the subsystem type j is then its reciprocal, $\theta_j = 1/\eta_j$. Owing to the invariance property of MLE, the MLE of η_j is obtained as

$$\widehat{\eta}_j = \sum_{k=1,2,6} \widehat{\eta}_{jk} = \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \sum_{k=1,2,6} n_{lijk}}{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \tau_{li}}$$

It is the total number of failures recorded for the subsystem type j regardless of the failure types across all the aircraft systems divided by the total flight duration of all the aircraft systems. Subsequently, the MLE of the MTBUM of the subsystem type j is obtained as $\widehat{\theta}_j = \frac{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \tau_{li}}{\sum_{i=1}^{n} \sum_{l=1}^{n_d} \sum_{k=1,2,6} n_{lijk}}$.

This formulation of the MLE imposes a certain practical issue as n_{lijk}'s are often zero with short τ_{li}'s, leading to $\widehat{\eta}_{jk} = 0$ and unbounded MTBF estimate (*viz.*, $\widehat{\theta}_{jk} = \infty$). Moreover, the frequentist approach to parameter estimation makes the sequential inferential procedure more cumbersome when the additional maintenance data come in on a regular basis to update the parameter estimates. In order to tackle this, it is proposed to employ the popular Bayesian framework for estimating the MTBF of each subsystem type j by each failure type k.

3.2 Bayesian Estimation of MTBF

For mathematical simplicity, here we construct the posterior distribution of η_{jk} although the focus of inference is on $\theta_{jk} = 1/\eta_{jk}$. Then, (6) can be reparametrized into the likelihood of $\boldsymbol{\eta}$, a vector of η_{jk}'s as

$$L(\boldsymbol{\eta}) = \prod_{j=1}^{n_{sub}} \prod_{k=1,2,6} L(\eta_{jk}), \tag{7}$$

where $L(\eta_{jk}) = \eta_{jk}^{\sum_{i=1}^{n} \sum_{l=1}^{n_d} n_{lijk}} \exp\left(-\eta_{jk} \sum_{i=1}^{n} \sum_{l=1}^{n_d} \tau_{li}\right)$ is the likelihood of η_{jk}. In addition to the likelihood function, we need to elicit the prior historical information, the experts' opinions or beliefs about the MTBF through a distribution function of choice, denoted by $g_0(\boldsymbol{\eta}) = g_0(\boldsymbol{\eta}; \boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0$ being a vector of hyperparameters $\theta_{0;jk}$'s. For the sake of simplicity without loss of generality, it is considered that the priors of η_{jk}'s are mutually independent so that $g_0(\boldsymbol{\eta})$ can be decomposed into a product form. That is, $g_0(\boldsymbol{\eta}) = \prod_{j=1}^{n_{sub}} \prod_{k=1,2,6} g_0\left(\eta_{jk}; \theta_{0;jk}\right)$, where $g_0\left(\eta_{jk}; \theta_{0;jk}\right)$ is the prior distribution of η_{jk} with the hyperparameter $\theta_{0;jk}$. Then, by the Bayes' theorem, the posterior distribution function of $\boldsymbol{\eta}$ given the maintenance data is expressed as

$$g(\boldsymbol{\eta}) = \frac{L(\boldsymbol{\eta})\ g_0(\boldsymbol{\eta})}{\int_{\boldsymbol{\eta}} L(\boldsymbol{\eta})\ g_0(\boldsymbol{\eta})\ d\boldsymbol{\eta}} = \prod_{j=1}^{n_{sub}} \prod_{k=1,2,6} g\left(\eta_{jk}; \theta_{0;jk}\right), \tag{8}$$

where $g\Big(\eta_{jk};\theta_{0;jk}\Big) = \dfrac{L(\eta_{jk})\ g_0\Big(\eta_{jk};\theta_{0;jk}\Big)}{\int_{\eta_{jk}} L(\eta_{jk})\ g_0\Big(\eta_{jk};\theta_{0;jk}\Big) d\eta_{jk}}$ is the marginal posterior distribution of η_{jk}, and its denominator is called the marginal likelihood.

As shown above, the posterior distribution is a combination of the prior (determined by the researcher/practitioner) and the likelihood (determined by the maintenance data). The contribution of these two quantities to the posterior is not equal though. With more maintenance data, the likelihood is given much more relative weight in calculating the posterior for $\boldsymbol{\eta}$ (and in turn, for the MTBF). When the information contained in the likelihood is relatively small due to a limited amount of the maintenance data, the prior will play a key role in the posterior for estimating the MTBF and each additional piece of information will have a pronounced impact. Thus, it is necessary to include external information in the form of informative (or subjective) priors; see Gelman (2006). For instance, one might need to consult application-specific experts, meta-analyses, or review studies in the area of interest to obtain informative, accurate priors that can meaningfully contribute to the posterior distribution of $\boldsymbol{\eta}$ and in turn, $\boldsymbol{\theta}$, a vector of the MTBF θ_{jk}'s. Specifying the priors of $\boldsymbol{\eta}$ based on expert opinions or previous studies can potentially improve the inferential performance since it allows to base results on more information than what is strictly provided in the maintenance data, which is especially helpful with small data sizes. Since the contractor-supplied estimates of the MTBF θ_{jk}'s are available for this analysis, here we treat these estimates as the hyperparameters $\theta_{0;jk}$'s for η_{jk}'s. To enjoy additional simplicity of derivations and computations, the prior distribution of each η_{jk} is proposed to be exponential, conjugate to (7), with its mean equal to the reciprocal of $\theta_{0;jk}$. That is, $g_0(\eta;\theta_0) = \theta_0 \exp(-\theta_0\eta)$ for $\eta > 0$ in general.

Once the posterior is specified, the inference for $\boldsymbol{\eta}$ can be performed using the decision-theoretic approach with a loss function of choice; see van de Schoot, et al. (2014). The usual quantity of interest is the measurement of central tendency or location of the posterior distribution. Then, Bayes estimator (or action) is an estimator or decision rule that minimizes the posterior expected loss (*viz.*, Bayes risk). Equivalently, it maximizes a posterior utility function of choice. Depending on the property of the posterior, one could utilize the posterior mean, which minimizes the expected loss with respect to the quadratic error loss function while the posterior median is the robust estimator and minimizes the expected loss under the absolute difference loss function in a univariate case. If a value with the greatest posterior probability is desired, the maximum a posteriori (MAP) estimate (*i.e.*, posterior mode(s)) can be used as the measurement of center, and it also minimizes the expected loss with respect to the 0-1 loss function. If one wants to quantify the uncertainty about the posterior center in addition to a point estimate, the HPD (highest posterior density) credible interval or set can be derived, which is analogous to the frequentist confidence interval but has a more intuitive interpretation; see Jackman (2009). Since the posterior distribution of $\boldsymbol{\eta}$ in (8) is expressed as a product of the marginal posterior distributions of η_{jk}'s, the inference for each η_{jk} can be carried out separately. This dimension reduction dramatically speeds up the computation of the inferential results. It is also observed that

$$g\Big(\eta_{jk};\theta_{0;jk}\Big) \;\propto\; L(\eta_{jk})\; g_0\Big(\eta_{jk};\theta_{0;jk}\Big) \;\propto\; \eta_{jk}^{n_{jk}^*-1}\exp\left(-\eta_{jk}\tau_{jk}^*\right),$$

where

$$n_{jk}^* \;=\; 1+\sum_{i=1}^{n}\sum_{l=1}^{n_d} n_{lijk}, \tag{9}$$

$$\tau_{jk}^* \;=\; \theta_{0;jk}+\sum_{i=1}^{n}\sum_{l=1}^{n_d}\tau_{li} \tag{10}$$

This indicates that the posterior distribution of η_{jk} follows a gamma distribution with the shape parameter n_{jk}^* and the rate parameter τ_{jk}^*. If the posterior mean is chosen as the Bayes estimator, it is then given as

$$\widehat{\eta}_{jk}^B = \frac{n_{jk}^*}{\tau_{jk}^*} = \frac{1+\sum_{i=1}^{n}\sum_{l=1}^{n_d} n_{lijk}}{\theta_{0;jk}+\sum_{i=1}^{n}\sum_{l=1}^{n_d}\tau_{li}}$$

while the MAP estimate is given as

$$\widehat{\eta}_{jk}^M = \frac{n_{jk}^*-1}{\tau_{jk}^*} = \frac{\sum_{i=1}^{n}\sum_{l=1}^{n_d} n_{lijk}}{\theta_{0;jk}+\sum_{i=1}^{n}\sum_{l=1}^{n_d}\tau_{li}}$$

It is noted that $\widehat{\eta}_{jk}^M < \widehat{\eta}_{jk}$, the MLE of η_{jk}, and hence, $\widehat{\theta}_{jk}^M = 1/\widehat{\eta}_{jk}^M$ is larger than $\widehat{\theta}_{jk}$, the MLE of the MTBF. On the other hand, $\widehat{\theta}_{jk}^B = 1/\widehat{\eta}_{jk}^B = \dfrac{\theta_{0;jk}+\sum_{i=1}^{n}\sum_{l=1}^{n_d}\tau_{li}}{1+\sum_{i=1}^{n}\sum_{l=1}^{n_d} n_{lijk}}$, the plugin estimator of the MTBF stays strictly finite unlike $\widehat{\theta}_{jk}$. Also, its estimate space is tighter than that of the MLE. When the maintenance data is not available yet, it coincides with the hyperparameter $\theta_{0;jk}$, which is the prior MTBF of the subsystem type j by the failure type k.

Moreover, a simple transformation tells us that $2\eta_{jk}\tau_{jk}^*$ follows a χ^2 distribution with $2n_{jk}^*$ degrees of freedom. This can be used to construct $100(1-\alpha)\%$ credible interval for η_{jk}. For instance, an equi-tailed two-sided $100(1-\alpha)\%$ credible interval for η_{jk} is obtained as

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2};\;2n_{jk}^*}}{2\tau_{jk}^*},\;\frac{\chi^2_{\frac{\alpha}{2};\;2n_{jk}^*}}{2\tau_{jk}^*}\right),$$

where $\chi^2_{\alpha;\nu}$ is the upper $100\alpha^{\text{th}}$-percentile of a χ^2 distribution with ν degrees of freedom. By taking the reciprocal, the two-sided $100(1-\alpha)\%$ credible interval for θ_{jk} is obtained as

$$\left(\frac{2\tau_{jk}^*}{\chi^2_{\frac{\alpha}{2};\;2n_{jk}^*}},\;\frac{2\tau_{jk}^*}{\chi^2_{1-\frac{\alpha}{2};\;2n_{jk}^*}}\right)$$

It should be noted that with a choice of non-conjugate and/or objective priors, these point and interval estimators cannot be obtained analytically. Then, one should utilize a popular stochastic simulation-based approach such as the Markov chain Monte Carlo (MCMC) sampling method implementing the Metropolis-Hastings algorithm with the Gibbs sampler in order to elicit the posterior distribution of $\boldsymbol{\eta}$ computationally. Models that are difficult or inestimable with frequentist methods can be fit straightforwardly under the Bayesian framework; see Levy (2009), Muthen and Asparouhov (2012). Since Bayesian methods do not rely on large sample asymptotics, they are also better equipped to handle small sample situations but the estimates can be sensitive to the specification of the prior.

For practitioners, it is also important to understand that confidence intervals and credible intervals serve a similar purpose of estimating population parameters, like averages or proportions, but they differ in their interpretation and methodology due to how they handle uncertainty. Confidence intervals, rooted in frequentist statistics, offer a fixed probability that the parameter lies within the interval, disregarding prior beliefs or knowledge. They tell us that if we were to take many samples and calculate intervals the same way, a certain percentage (*e.g.*, 95%) of those intervals would contain the true parameter. However, for a single interval, we do not know if it actually contains the true value. On the other hand, credible intervals, based on Bayesian statistics, provide a range of plausible parameter values by incorporating both the current observed data and prior beliefs or knowledge about the parameter. These intervals provide a probability distribution of where the parameter is likely to be, and this probability varies depending on the data and prior information. The probability that the parameter falls within this interval is not fixed but varies according to the available data and prior beliefs. The choice between these two approaches hinges on the analytical method, the nature of data, and the significance of prior knowledge in the analysis. In practical terms, credible intervals are becoming more popular because they allow us to flexibly combine prior knowledge with new data, making them especially valuable in complex real-world situations where understanding uncertainty is crucial.

4 Bayesian Testing

4.1 Bayesian Factor for Testing

A principled approach to Bayesian hypothesis testing involves the Bayes factor (BF). It is a ratio between two competing statistical hypotheses or models, each represented by their respective evidence; see Jeffreys (1998). It serves to quantify the favorability of one hypothesis (say, H_1) over the other (say, H_0). Essentially, it measures how well two rival hypotheses perform in predicting outcomes and

Table 1: Heuristic classification scheme for the Bayes factor, BF by Kass and Raftery (1995)

$\log_{10} BF$	BF	Strength of Evidence
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

signifies the extent to which data warrant a shift in beliefs regarding the relative credibility of these hypotheses. In comparison to traditional p-values, Bayes factors offer an informative alternative for reporting the results of hypothesis tests. They provide direct measures of the extent to which data support one hypothesis over another and have the ability to quantify support for true null hypotheses. Mathematically, the Bayes factor is defined as the ratio of two marginal likelihoods, or the odds ratio of posterior to prior, given by

$$BF = \frac{P(D|H_1)}{P(D|H_0)} = \frac{P(H_1|D)}{P(H_0|D)} \bigg/ \frac{P(H_1)}{P(H_0)}, \tag{11}$$

where D denotes the observed data. In the second equality of (11), the second ratio indicates the prior odds, quantifying the relative credibility of the competing hypotheses before observing the data while the first ratio indicates the posterior odds, quantifying the relative credibility of the competing hypotheses after observing the data. The Bayes factor then measures the degree of evidence provided by the data for one hypothesis against the other. Its value ranges from 0 to ∞, and when the prior probabilities between two rival hypotheses are equal $\big(viz., P(H_0) = P(H_1)\big)$, the prior odds takes the value of one and the Bayes factor simply becomes the posterior odds or the ratio of the posterior probabilities. The Bayes factor of one indicates that both hypotheses have equal predictive performance. When $BF > 1$, the alternative hypothesis or model is more strongly supported by the data under consideration than the null. For interpretation of the Bayes factor, a widely cited table in Kass and Raftery (1995) is reproduced in Table 1; also, see Jeffreys (1998).

The Bayes factor serves as a Bayesian counterpart to the likelihood ratio test (LRT) in the frequentist paradigm. It relies on the marginal likelihood (integrated over prior) instead of the maximized likelihood for each statistical model. Unlike the classical LRT, this Bayesian model comparison does not rely on a single set of parameters. It integrates across all the parameters in each competing model, taking into account their respective priors. It is important to note that the Bayesian test of hypotheses does not necessitate a distinction between the null and alternative hypotheses as defined in the frequentist universe. In classical hypothesis testing, the null hypothesis H_0 is favored, and evidence is collected and quantified against it. Frequentist tests yield a p-value, which is used to decide whether to reject H_0 in favor of the alternative H_1. The interpretation of the p-value assumes a binary problem formulation of either the null or alternative hypothesis being true. In the Bayesian context, however, the outcome is a probability statement about the hypothesis or model under consideration, free from deterministic assumptions. It does not involve determining whether a value of the test statistic is significant or rejecting H_0. Instead, Bayesian hypothesis testing provides an explicit probability statement about the hypothesis being tested. Thus, in contrast to the frequentist's significance testing, the Bayes factor facilitates the assessment of evidence in support of H_0, rather than just determining whether to reject or not to reject it.

While the concept is straightforward, the actual calculation of the Bayes factor in (11) can be computationally demanding, especially when dealing with complex models and hypotheses. Because closed-form expressions for the marginal likelihood are typically unavailable, numerical approximations relying on MCMC samples have been proposed as a solution. For models where an explicit form of the likelihood is either unavailable or computationally expensive to assess, approximate Bayesian computation (ABC) also offers a way to perform model selection within a Bayesian framework; see Beaumont, et al. (2009), Turner and van Zandt (2012), Rabhi, et al. (2023). However, it is important to note that estimates of the Bayes factors obtained through approximate Bayesian methods often carry bias. One of the benefits of employing the Bayes factor is that it naturally incorporates a penalty for having excessive model complexity, effectively guarding against overfitting. There exists an approximation known as the Bayesian information criterion (BIC), which is derived by applying Laplace's approximation to the marginal likelihoods. In situations involving a large amount of data, the Bayes factor in (11) tends to converge toward the BIC as the influence of prior beliefs diminishes. In the case of small datasets, the choice of priors can be critical, and improper priors should be avoided because the Bayes factor becomes undefined if either of the two integrals in its ratio is not finite.

4.2 Bayesian Monitoring of MTBF

In this study, it is desired to investigate whether the contractor-supplied MTBF of a subsystem is credible and supported by the continuously collected maintenance data. In other words, it is desired to assess whether the contractor-supplied MTBF overestimates the actual MTBF based on the field maintenance data. Developing a Bayesian method for this enables an early detection of any subsystem whose reliability goes below the threshold level of acceptance so that the maintenance engineers can take timely actions to mitigate the issue and improve the overall mission reliability. The use of the Bayesian framework in this regard is particularly attractive as it makes the real-time sequential testing simpler and automatic when new maintenance data become available on a daily basis and update the estimates of interest. On the contrary, the frequentist-based procedures can be computationally burdensome. Let us denote $\Theta_{jk} = \{\theta_{jk} : \ \theta_{jk} > (1-\beta)\theta_{0;jk}\}$ be the set of acceptable values of the MTBF of the subsystem type j by the failure type k based on the contractor-supplied estimate $\theta_{0;jk}$. Here, $\beta \in [0, 1)$ is a pre-specified value that determines the threshold level of acceptable MTBF. For example, $\beta = 0.1$ indicates that the administration is prepared to tolerate up to 10% reduction in the contractor-provided MTBF. The complement of this set is denoted by Θ^c_{jk}, which lists unacceptable MTBF values for the subsystem type j by the failure type k. Then, the two competing hypotheses for testing can be formulated as $H_0 : \ \theta_{jk} \in \Theta_{jk}$ vs. $H_1 : \ \theta_{jk} \notin \Theta_{jk}$. The alternative can be expressed equivalently as $H_1 : \ \theta_{jk} \in \Theta^c_{jk}$. It should be noted that when the parameter of interest (*e.g.*, MTBF θ_{jk}) is continuous, the Bayesian test of hypotheses does not allow formulation of H_0 as a simple hypothesis (*i.e.*, a single null value) but rather as a composite hypothesis with an interval of values. Otherwise, $P(H_0|D) = 0$ and $P(H_0) = 0$ for assuming a single null value, and the Bayes factor cannot be calculated as a result.

To calculate (11), it is first obtained that

$$\begin{aligned} P(H_0) &= P(\theta_{jk} \in \Theta_{jk}) = P\big(\theta_{jk} > (1-\beta)\theta_{0;jk}\big) \\ &= P\left(\eta_{jk} < \frac{1}{(1-\beta)\theta_{0;jk}}\right) = \int_0^{\frac{1}{(1-\beta)\theta_{0;jk}}} g_0\Big(\eta_{jk}; \theta_{0;jk}\Big)\, d\eta_{jk} \\ &= 1 - e^{-1/(1-\beta)} \end{aligned}$$

and $P(H_1) = 1 - P(H_0) = e^{-1/(1-\beta)}$ since $g_0(\eta; \theta_0) = \theta_0 \exp(-\theta_0 \eta)$ for $\eta > 0$. Also, we have

$$\begin{aligned} P(H_0|D) &= P(\theta_{jk} \in \Theta_{jk} \mid D) = P\big(\theta_{jk} > (1-\beta)\theta_{0;jk} \mid D\big) \\ &= P\left(\eta_{jk} < \frac{1}{(1-\beta)\theta_{0;jk}} \,\middle|\, D\right) = \int_0^{\frac{1}{(1-\beta)\theta_{0;jk}}} g\Big(\eta_{jk}; \theta_{0;jk}\Big)\, d\eta_{jk} \\ &= \frac{1}{\Gamma(n^*_{jk})} I_g\left(\frac{\tau^*_{jk}}{(1-\beta)\theta_{0;jk}};\ n^*_{jk}\right) \\ &= \frac{1}{(n^*_{jk}-1)!} I_g\left(\frac{\tau^*_{jk}}{(1-\beta)\theta_{0;jk}};\ n^*_{jk}\right) \end{aligned}$$

and $P(H_1|D) = 1 - P(H_0|D) = 1 - \dfrac{1}{(n^*_{jk}-1)!} I_g\left(\dfrac{\tau^*_{jk}}{(1-\beta)\theta_{0;jk}};\ n^*_{jk}\right)$, where n^*_{jk} and τ^*_{jk} are as defined in (9) and (10), respectively. Here, $I_g(z;\ \gamma) = \displaystyle\int_0^z x^{\gamma-1} e^{-x}\, dx$ is the lower incomplete gamma function for $0 < z < \infty$ with the shape parameter $\gamma > 0$, and $\Gamma(\gamma) = I_g(\infty;\ \gamma)$ is the (complete) gamma function, which reduces to $(\gamma - 1)!$ when γ is a positive integer. Hence, $P(H_0|D)$ in this case is the regularized lower incomplete gamma function (or the lower-tail gamma probability) while $P(H_1|D)$ is the regularized upper incomplete gamma function (or the upper-tail gamma probability). As a result, the Bayes factor is obtained as

$$\begin{aligned} BF &= \frac{P(H_1|D)}{P(H_0|D)} \Big/ \frac{P(H_1)}{P(H_0)} \\ &= \frac{1 - \frac{1}{(n^*_{jk}-1)!} I_g\left(\frac{\tau^*_{jk}}{(1-\beta)\theta_{0;jk}};\ n^*_{jk}\right)}{\frac{1}{(n^*_{jk}-1)!} I_g\left(\frac{\tau^*_{jk}}{(1-\beta)\theta_{0;jk}};\ n^*_{jk}\right)} \Big/ \frac{e^{-1/(1-\beta)}}{1-e^{-1/(1-\beta)}} \end{aligned}$$

and one should refer to Table 1 for its interpretation. It is interesting to note that the prior probabilities $P(H_0)$ and $P(H_1)$ do not depend on the hyperparameter value $\theta_{0;jk}$ in this case but only on the tolerable reduction proportion β. At $\beta = 0.1$, the prior odds is computed to be $P(H_1)/P(H_0) = 0.4907$, which means that the probability of claiming no overestimation of the MTBF by the contractor is about twice the probability of claiming overestimation even without any maintenance data or evidence. Even if no reduction is allowed (*viz.*, $\beta = 0$), the prior odds is $P(H_1)/P(H_0) = 0.5820$, indicating that the probability of claiming no overestimation of the MTBF is still 72% higher than the probability of claiming overestimation without any maintenance data (*i.e.*, a conservative test). In the next section, we demonstrate the application of the Bayesian inferential procedures developed in this section using the USAF maintenance data of the aircraft systems described in Section 1.

5 Aircraft Maintenance Application

Here we apply the inferential procedures derived in the previous sections to the daily maintenance activity data of the aircraft systems described in Section 1, which motivated the research presented in this work. The daily flight hours of the aircraft system 1 ranged between 0.02 and 3.78 with the average and standard deviation given by 1.55 ± 0.78 while those of the aircraft system 2 ranged between 0.02 and 4.60 with the average and standard deviation given by 1.66 ± 0.96. Together, the average day-to-day flight hours were 1.61 ± 0.89. Table 2 lists the contractor-supplied MTBF (in hours) for each subsystem by each failure type as described in Section 1. Type 1 indicates the inherent/natural failure, Type 2 indicates the induced failure caused by an outside/external influence, and Type 6 indicates no defect, often associated with the CND situation where the environment/field condition cannot be emulated during troubleshooting. Some of the values are not available (NA) as they could not be estimated by the contractor. As these values represent the historical information, the experts' opinions or beliefs about the MTBF, they were used as the hyperparameters $\theta_{0;jk}$'s of the prior distributions for Bayesian analysis. Based on the formulation in Section 3, the Bayes estimates (*viz.*, the posterior means) of the MTBF (in hours) for each subsystem type by each failure type were calculated, and they are provided in Table 3 along with the corresponding equi-tailed two-sided 95% credible intervals of the MTBF (in hours) in Table 4. In order to examine potential discrepancies between the contractor-provided MTBF and the actual MTBF of each subsystem based on the field data, the differences between the Bayes estimates and the contractor-supplied MTBF (in hours) for each subsystem type by each failure type were computed, and they are provided in Table 5. The numbers in brackets indicate that the contractor-supplied MTBF overestimates the actual MTBF estimated based on the field maintenance data, and it is noted that the magnitudes of overestimation vary substantially from subsystem to subsystem and from failure type to failure type.

As the confirmatory analysis, the Bayes factor for each subsystem type by each failure type was calculated based on the formulation in Section 4, using $\beta = 0.1$ to tolerate up to 10% reduction from the contractor-provided MTBF as acceptable. They are tabulated in Table 6, and it is discovered that a number of subsystems have their BF more than 10, providing strong evidence for overestimation by the contractor according to Table 1. These subsystem types and failure types include

- the subsystem types 2,6,9, and 21 for the failure type 1 only;
- the subsystem types 3 and 19 for the failure type 2 only;
- the subsystem type 10 for the failure types 2 and 6;
- the subsystem type 14 for all failure types;
- the subsystem type 15 for the failure types 1 and 2.

Among these, the subsystem type 9 for the failure type 1 (inherent), the subsystem type 10 for the failure type 6 (no defect), and the subsystem type 14 for the failure type 2 (induced) exhibit the BF values overwhelmingly greater than 100, suggesting decisive evidence of overestimation, according to Table 1. This observation definitely calls for further systemic investigations by the reliability engineers and maintenance staff. Figure 2 visualizes the temporal evolution of the Bayesian estimates of the MTBF of these subsystems for the corresponding failure types along with their 95% credible bands. For easier detection of the discrepancy, the contractor-provided MTBF is overlaid as a dotted horizontal line in each subplot. It is noticeable that none of these credible intervals captures the contractor-supplied estimates of the MTBF after all. Supported by the Bayes factors, here we were able to detect certain subsystems whose estimated MTBF significantly dropped below the initial MTBF estimate at the induction. To make the inferential procedures developed in this work more user-friendly, we implemented these in Microsoft Excel so that it does not require reliability practitioners to learn high-level programming languages such as R and Python. The Bayesian approach indeed enables a

Table 2: Contractor-supplied MTBF (in hours) for each subsystem by each failure type as described in Section 1 (Type 1: inherent failure, Type 2: induced failure, Type 6: no defect)

Subsystem Type	MTBF by Type 1	MTBF by Type 2	MTBF by Type 6
1	202.8	113.9	984.7
2	350.1	873.3	3500.8
3	433.6	701.1	2838.8
4	4521.7	2255.8	27107.7
5	870.9	272.7	2594.8
6	417.1	189.6	1818.3
7	824.0	802.4	4849.3
8	115013.8	38344.6	410737.9
9	168.4	124.1	969.3
10	81.6	230.6	330.1
11	227.4	497.8	2158.0
12	111.9	178.1	893.3
13	345.9	1161.2	1773.3
14	438.7	996.9	4279.4
15	343.9	150.8	566.6
16	997.6	1103.4	1936.6
17	1005.6	1063.9	8576.0
18	717.0	903.0	5647.2
19	716.4	243.4	1391.0
20	462.0	NA	NA
21	1489.3	433679.9	953130.5
22	1402.6	58669.4	93120.2
23	2158.0	NA	NA
24	1024.2	NA	NA
25	2029.2	16439.1	56662.1
26	569.1	NA	NA
27	33878.9	NA	NA
28	4426.8	4076.7	30279.6
29	1132.6	20464.9	38123.3

Table 3: Bayes estimates of the MTBF (in hours) for each subsystem by each failure type as described in Section 1 (Type 1: inherent failure, Type 2: induced failure, Type 6: no defect)

Subsystem Type	MTBF by Type 1	MTBF by Type 2	MTBF by Type 6
1	132.3	220.1	437.0
2	112.7	1199.7	3827.2
3	380.0	205.5	3165.2
4	2424.0	2582.2	27434.1
5	1197.3	199.7	2921.2
6	185.9	172.0	2144.7
7	575.2	376.3	2587.8
8	115340.2	38671.0	411064.3
9	26.0	112.6	431.9
10	81.6	111.4	43.8
11	138.4	824.2	2484.4
12	87.7	504.5	1219.7
13	672.2	743.8	1049.9
14	191.3	220.5	1535.3
15	134.0	59.7	446.5
16	1324.0	1429.7	2263.0
17	666.0	695.1	8902.4
18	521.7	409.8	5973.6
19	347.6	95.0	1717.4
20	788.4	NA	NA
21	605.2	217003.1	476728.4
22	1728.9	29497.9	93446.6
23	2484.4	NA	NA
24	1350.6	NA	NA
25	2355.5	16765.5	56988.5
26	895.5	NA	NA
27	34205.2	NA	NA
28	4753.1	4403.1	30606.0
29	1459.0	10395.6	19224.8

Table 4: Equi-tailed two-sided 95% credible intervals of the MTBF (in hours) for each subsystem by each failure type as described in Section 1 (Type 1: inherent failure, Type 2: induced failure, Type 6: no defect)

Subsystem Type	MTBF by Type 1	MTBF by Type 2	MTBF by Type 6
1	(60.4, 485.6)	(79.0, 1817.8)	(181.5, 2119.2)
2	(58.0, 307.2)	(325.2, 47384.7)	(1037.5, 151164.6)
3	(136.4, 3137.7)	(100.3, 632.9)	(858.0, 125017.1)
4	(870.1, 20016.2)	(700.0, 101990.8)	(7437.0, 1083588.9)
5	(324.6, 47291.8)	(82.9, 968.3)	(791.9, 115380.4)
6	(84.8, 682.2)	(71.4, 834.0)	(581.4, 84711.9)
7	(206.5, 4749.4)	(156.2, 1824.5)	(928.9, 21368.6)
8	(31267.0, 4555693.9)	(10483.1, 1527422.3)	(111433.4, 16236172.4)
9	(17.4, 43.3)	(51.4, 413.3)	(179.3, 2094.3)
10	(39.8, 251.3)	(54.4, 343.1)	(27.9, 78.2)
11	(63.2, 508.1)	(223.4, 32552.8)	(673.5, 98127.0)
12	(42.8, 270.0)	(136.8, 19927.5)	(330.6, 48174.9)
13	(182.2, 26552.1)	(267.0, 6141.8)	(376.9, 8669.0)
14	(87.3, 702.0)	(113.4, 601.0)	(637.5, 7444.6)
15	(65.4, 412.8)	(33.1, 138.2)	(160.3, 3686.9)
16	(358.9, 52293.8)	(387.6, 56471.9)	(613.5, 89384.0)
17	(239.1, 5499.1)	(249.5, 5740.0)	(2413.3, 351626.3)
18	(187.3, 4307.8)	(170.2, 1987.2)	(1619.4, 235944.3)
19	(144.3, 1685.5)	(48.8, 258.8)	(465.5, 67831.9)
20	(213.7, 31138.7)	NA	NA
21	(251.3, 2934.8)	(77895.6, 1791864.6)	(171126.7, 3936500.2)
22	(468.7, 68289.8)	(10588.6, 243573.6)	(25332.0, 3690941.9)
23	(673.5, 98128.8)	NA	NA
24	(366.1, 53344.7)	NA	NA
25	(638.5, 93038.5)	(4544.9, 662201.7)	(15448.7, 2250925.7)
26	(242.7, 35369.0)	NA	NA
27	(9272.5, 1351035.1)	NA	NA
28	(1288.5, 187739.1)	(1193.6, 173912.3)	(8296.8, 1208873.3)
29	(395.5, 57626.3)	(3731.6, 85840.0)	(6901.0, 158745.7)

Table 5: Difference between the Bayes estimate and the contractor-supplied MTBF (in hours) for each subsystem by each failure type as described in Section 1; the numbers in brackets indicate negative values.

Subsystem Type	MTBF by Type 1	MTBF by Type 2	MTBF by Type 6
1	(70.5)	106.2	(547.7)
2	(237.4)	326.4	326.4
3	(53.6)	(495.6)	326.4
4	(2097.7)	326.4	326.4
5	326.4	(73.0)	326.4
6	(231.3)	(17.6)	326.4
7	(248.8)	(426.1)	(2261.5)
8	326.4	326.4	326.4
9	(142.4)	(11.5)	(537.4)
10	(0.0)	(119.2)	(286.3)
11	(88.9)	326.4	326.4
12	(24.2)	326.4	326.4
13	326.4	(417.4)	(723.5)
14	(247.5)	(776.3)	(2744.1)
15	(209.8)	(91.2)	(120.1)
16	326.4	326.4	326.4
17	(339.6)	(368.8)	326.4
18	(195.3)	(493.2)	326.4
19	(368.8)	(148.4)	326.4
20	326.4	NA	NA
21	(884.1)	(216676.7)	(476402.1)
22	326.4	(29171.5)	326.4
23	326.4	NA	NA
24	326.4	NA	NA
25	326.4	326.4	326.4
26	326.4	NA	NA
27	326.4	NA	NA
28	326.4	326.4	326.4
29	326.4	(10069.2)	(18898.5)

Table 6: Bayes factors for each subsystem by each failure type with $\beta = 0.1$, indicating tolerance of up to 10% reduction in the contractor-provided MTBF (in hours) as acceptable

Subsystem Type	BF for Type 1	BF for Type 2	BF for Type 6
1	4.1	0.2	8.9
2	88.7	0.6	0.9
3	1.5	78.6	0.8
4	4.1	0.8	1.0
5	0.6	2.6	0.8
6	12.6	1.5	0.8
7	2.4	7.8	4.1
8	1.0	1.0	1.0
9	9.3×10^8	1.5	8.8
10	1.1	13.1	1.4×10^8
11	5.1	0.4	0.8
12	2.6	0.1	0.6
13	0.3	2.9	3.3
14	13.4	493.6	15.0
15	27.6	72.5	1.9
16	0.6	0.6	0.8
17	2.7	2.7	0.9
18	2.2	8.4	0.9
19	7.2	39.5	0.7
20	0.4	NA	NA
21	11.1	4.6	4.6
22	0.7	4.6	1.0
23	0.8	NA	NA
24	0.6	NA	NA
25	0.8	1.0	1.0
26	0.4	NA	NA
27	1.0	NA	NA
28	0.9	0.9	1.0
29	0.6	4.5	4.6

real-time continuous update of the parameter estimates as new maintenance data come in on a daily basis. Upon recording new maintenance data, the spreadsheet algorithm automatically updates the Bayes estimates, the corresponding credible intervals, and the Bayes factors as well as the posterior distribution of the model parameters. As a result, we can predict the overall mission reliability (or a survival probability within a given flight duration) in a continuous and convenient manner. This demonstrates that the Bayesian method, if carefully implemented, can be advantageous for accurate assessment and prediction of the mission reliability of critical air force operations.

6 Discussion & Conclusion

In this work, the reliability of repairable subsystems within a system was modeled under the competing and complementary risks framework, assuming that the lifetime of each subsystem type by each failure type follows non-identical exponential distribution. Subsequently, we constructed the overall reliability of the entire system based on the multinomial distribution as the temporal order of multiple subsystem failures within a given time interval is uncertain due to interval/group-inspection. A relatively short flight duration each day also allowed the linear approximation of the failure probability of each subsystem type. Since the number of subsystems of each type is unknown but substantial, the multinomial distribution could be transformed to the Poisson distribution, which theoretically justifies the use of homogeneous Poisson processes for modeling the subsystem reliability in an asymptotic manner. Using the proposed model, the likelihood function was formulated and we developed the inferential procedures for point and interval estimations of the MTBF of subsystems with different failure causes. Given a considerable number of parameters to estimate, we explored the efficacy of a Bayesian method, treating the contractor-supplied MTBF as the hyperparameters of conjugate priors. This approach mitigates potential model uncertainty as well as the practical limitation of a frequentist-based approach. It also facilitates continuous updates of the parameter estimates as new maintenance data become available. Under this setup, we were able to obtain simple and exact Bayes estimators ($viz.$, the posterior means) as well as a tractable posterior distribution for constructing exact credible intervals for the model parameters.

Additionally, the Bayes factor was discussed to investigate potential discrepancies between the contractor-provided MTBF and the actual MTBF of each subsystem based on the field data. The inferential procedures developed in this work were demonstrated utilizing the day-to-day maintenance data of identical aircraft systems. Evolution of the Bayesian estimates of the MTBF of each subsystem by each failure type was visualized along with the 95% credible bands over time. Aided by the Bayes

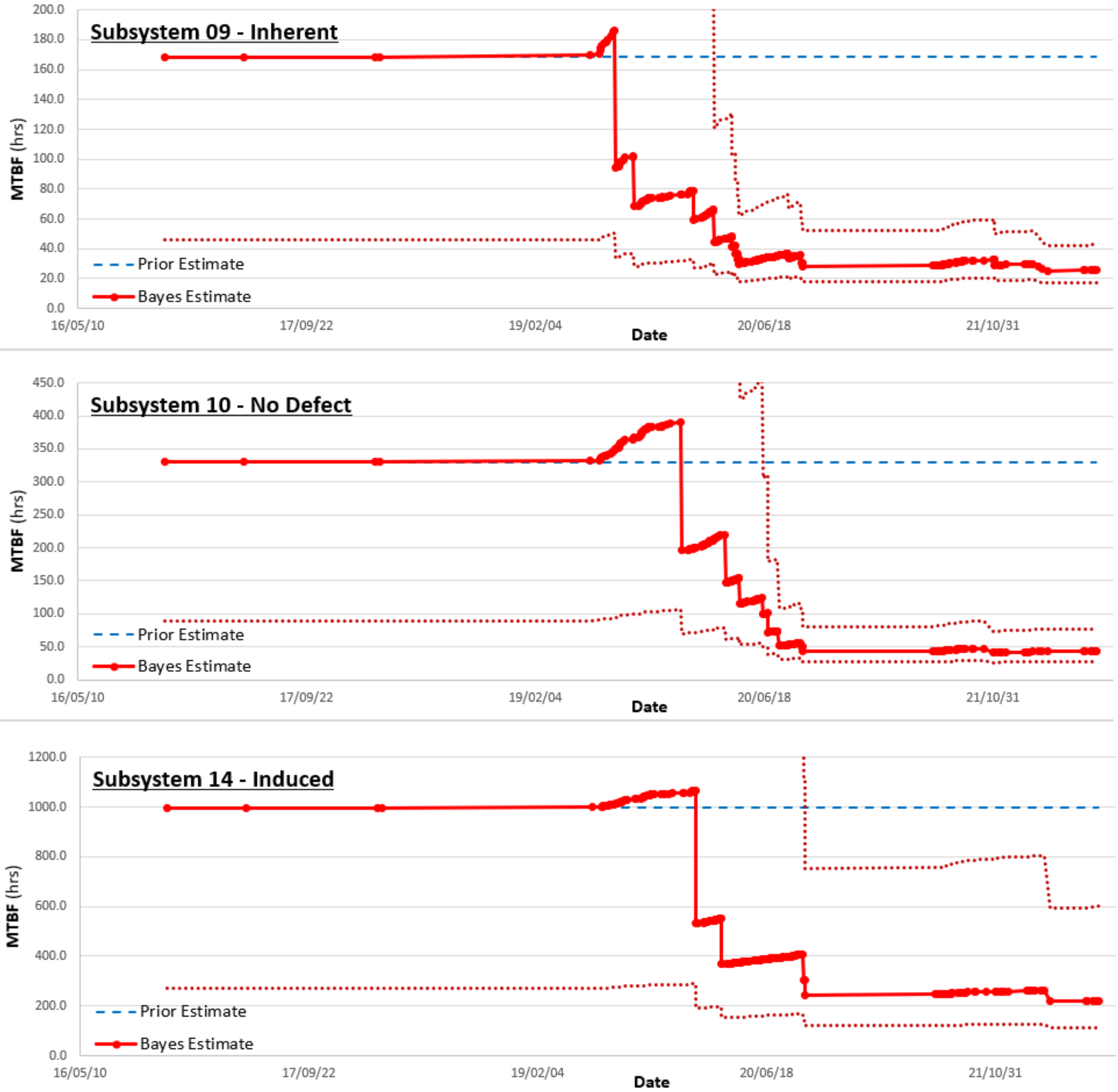

Figure 2: Evolution of the Bayesian estimates of the MTBF (in hours) along with the 95% credible bands over calendar days for selected subsystems, whose estimated MTBF by a particular failure type dropped below the contractor-provided MTBF (*i.e.*, the prior estimate) with the Bayes factor over 100

factors, we were able to detect certain subsystems whose MTBF significantly dropped below the initial MTBF estimate since the induction. It is noteworthy that the inferential procedures developed in this work were implemented in Microsoft Excel so that it is easy for any reliability practitioner to use without the need to learn sophisticated programming languages such as R or Python. Thus, this research supports an ongoing, real-time assessment of the overall mission reliability and helps early detection of any subsystem whose reliability performance goes below the expected level.

Nonetheless, there are some potential limitations of the methodology. Due to the limited amount of information on the data collection procedures, potential human errors by the maintenance engineers and staff on record keeping, unknown shifts in the maintenance regime and scope over time, etc., some critical assumptions had to be made in order to manage the derivation of the final reliability model. With a high degree of data aggregation (*i.e.*, lack of data granularity) as well as unknown system signature/architecture (*i.e.*, physical connections of components), all the subsystems were assumed *operationally* s-independent and *functionally* connected in parallel although they could be *structurally* connected in series (to manifest critical failures), and/or in parallel, or both, giving rise to complex interoperability. Although these assumptions are made pretty routinely for mathematical tractability in reliability analyses, it has certain consequences on the inferential outcomes, especially when they are too simplistic or unrealistic. To mitigate this issue, we would like to explore the dependent competing risks structure for the subsystem failures as well as the operational dependency of the subsystems through copulas. This could have some impact on the inferential results. Moreover, it is desired to investigate the implication of selecting non-exponential lifetime distributions for the competing risk factors (*viz.*, subsystem failure types) such as the popular Weibull, lognormal or gamma distributions. It is well understood that the complexity and property of these distributions will not be able to accommodate many nice properties of the Poisson process, which means that the development of inferential procedures becomes computationally expensive although these distributions provide more flexibility; see Escobar, et al. (2021).

In this work, we purposely employed the proper and informative prior for Bayesian analysis by utilizing the contractor-provided information about the MTBF of all the subsystems. If one has very little or no prior information, noninformative or objective priors such as Jeffreys prior, which is proportional to the square root of the Fisher information, or a reference prior, which maximizes the expected Kullback-Leibler divergence (*i.e.*, mutual information), could be employed to conduct the objective Bayesian inference. Gelman (2006) also recommended using half-Cauchy priors for variance components. Often, a closed form expression of the posterior distribution is not available by taking this route. Thus, it becomes necessary to implement simulation-based or computational methods such as the MCMC sampling with the Metropolis-Hastings algorithm and the Gibbs sampler to elicit the posterior distribution in order to obtain the parameter estimates, the credible intervals, and the Bayes factors; see Levy (2009).

As future works, we would also like to tackle some critical issues, including imperfect inspection and imperfect maintenance. Here we had to assume that the failure reporting is thorough and the maintenance action is immediate and precise so that the system is always brought back to the state of *as-good-as-new* before the next flight assignment. However, the reality is quite different in this regard. Due to the potential human errors and resource constraints/availability in addition to the aging of the aircraft system with varying rates of subsystem degradation over its usage and environmental conditions, it is not possible to restore the system state to a perfect condition. Taking this fact into account can, however, overly complicate the modeling stage. Nonetheless, it still needs to be addressed to produce a more realistic model for proper assessments of the system reliability. For instance, the failure rate of each subsystem could be changing over time due to the aging and usage in different operational conditions. In that case, non-homogeneous Poisson processes (NHPP) and/or regression techniques via the generalized linear model (GLM) could be considered; see Shafiq, et al. (2024). Moreover, the impact of subsystem dependencies and/or changing operational conditions over time could be modeled by adopting the concept of cumulative exposures often used in the context of accelerated life testing (ALT); see Han (2017, 2019, 2020a,b), Smit, et al. (2024). By developing the overall impact index of a subsystem failure on the reliability of other subsystems, the Markovian property can be attained when modeling the system reliability although this requires the simplex integration and permutations even with simple exponential lifetimes for the subsystems.

Another important and practical aspect to consider in modeling the system and subsystem reliability is the inspection and maintenance lead times. Although they are regarded negligible (*i.e.*, immediate and instantaneous repairs/replacements) to avoid the derivational complications, these quantities can never be ignored in practice as they are influenced by many practical resource constraints such as budget, parts and labor. Consequently, they could have some serious effects on the overall system health management (SHM); see Wiedner and Han (2021). Thus, taking these into account appropriately when modeling the system and subsystem reliability is ideal, and it can also help formulating the optimal maintenance scheduling with more efficient resource utilization. Eventually, it is desired to conduct a sensitivity analysis to compare and contrast the sensitivities of the inferential results over various models discussed in this section. The research works in this direction are currently under progress and we hope to report our findings in the future papers.

Acknowledgments

The authors are grateful to the associate editor and anonymous reviewers for their critical comments and valuable suggestions for improving the previous version of the manuscript.

References

Ahmadi, M.V. and Doostparast, M. (2022). “Evaluating the lifetime performance index of products based on progressively Type-II censored Pareto samples: a new Bayesian approach,” *Quality and Reliability Engineering International*, **38:** 1612–1634.

Arenbaev, N.K. (1976). “Asymptotic behavior of the multinomial distribution,” *Theory of Probability and Its Applications*, **4:** 805–810.

Beaumont, M.A., Cornuet, J., Marin, J., and Robert, C.P. (2009). “Adaptive approximate Bayesian computation,” *Biometrika*, **96:** 983–990.

Berger, J. (2006). “The case for objective Bayesian analysis,” *Bayesian Analysis*, **1:** 385–402.

Chan, R. (2020). “Machine downtime by the numbers: what maintenance teams need to know,” *UpKeep* (https://upkeep.com/blog/machine-downtime-costs)

Crowder, M.J. (2001). *Classical Competing Risks.* Boca Raton, FL: Chapman & Hall.

Dutta, S. and Kayal, S. (2022). “Bayesian and non-Bayesian inference of Weibull lifetime model based on partially observed competing risks data under unified hybrid censoring scheme,” *Quality and Reliability Engineering International*, **38:** 3867–3891.

Eis, P.A. (2013). “Chrysler agrees to recall of 2.7 million Jeep vehicles,” *NBC News* (https://www.cnbc.com/id/100824803)

Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models,” *Bayesian Analysis*, **1:** 515–534.

Goksedef, E. (2024). “Tesla recalls more than 1.6 million cars in China over steering software issues,” *BBC News* (https://www.bbc.com/news/technology-67891080)

Han, D. (2017). “Optimal accelerated life tests under a cost constraint with non-uniform stress durations,” *Quality Engineering*, **29:** 409–430.

Han, D. (2019). "Optimal design of a simple step-stress accelerated life test under progressive Type-I censoring with nonuniform durations for exponential lifetimes," *Quality and Reliability Engineering International*, **35:** 1297–1312.

Han, D. (2020a). "On the existence of the optimal step-stress accelerated life tests under progressive Type-I censoring," *IEEE Transactions on Reliability*, **69:** 903–915.

Han, D. (2020b). "Expected termination times of progressively Type-I censored step-stress accelerated life tests under continuous and interval inspections," *Statistica Neerlandica*, **74:** 112–124.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., and Sethupathy, G. (2016). "The age of analytics: competing in a data-driven world," *McKinsey Global Institute* (https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world)

Hollings, A. (2023). "Why the F-35 fighter jet is such a badass plane," *Popular Mechanics* (https://www.popularmechanics.com/military/aviation/a35281371/f-35-history)

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* New York, NY: Wiley & Sons.

Jeffreys, H. (1998). *The Theory of Probability.* Oxford, UK: Oxford University Press.

Kass, R.E. and Raftery, A.E. (1995). "Bayes factors," *Journal of the American Statistical Association*, **90:** 773–795.

Lengu, D., Syntetos, A.A., and Babai, M.Z. (2014). "Spare parts management: linking distributional assumptions to demand classification," *European Journal of Operational Research*, **235:** 624–635.

Levy, R. (2009). "The rise of Markov chain Monte Carlo estimation for psychometric modeling," *Journal of Probability and Statistics*, 1–18.

Limer, E. (2018). "Fiat Chrysler recalls 4.8 million vehicles over terrifying cruise control flaw," *Popular Mechanics* (https://www.popularmechanics.com/cars/a20916003/fiat-chrysler-recall-cruise-control-may-2018)

Escobar, L.A., Pascual, F.G., and Meeker, W.Q. (2021). *Statistical Methods for Reliability Data.* New York, NY: Wiley & Sons.

Muthen, B. and Asparouhov, T. (2012). “Bayesian structural equation modeling: a more flexible representation of substantive theory,” *Psychological Methods*, **17:** 313–335.

Nidumolu, J. (2023). “Tesla to recall nearly 55,000 Model X vehicles, auto regulator says,” *Reuters* (https://www.reuters.com/business/autos-transportation/tesla-recalls-nearly-55000-model-x-vehicles-auto-regulator-says-2023-10-17)

Pan, C., Cai, B., and Wang, L. (2020). “A Bayesian approach for analyzing partly interval-censored data under the proportional hazards model,” *Statistical Methods in Medical Research*, **29:** 3192–3204.

Rabhi, M., Abdessalem, A.B., Saintis, L., Castanier, B., and Sohoin, R. (2023). “Discrimination between accelerated life models via approximate Bayesian computation,” *Quality and Reliability Engineering International*, **39:** 1058–1082.

Sen, A., Banerjee, M., Li, Y., and Noone, A.M. (2010). “A Bayesian approach to competing risks analysis with masked cause of death,” *Statistics in Medicine*, **29:** 1681–1695.

Shafiq, A., Sindhu, T.N., Lone, S.A., Abushal, T.A., and Hassan, M.K.H. (2024). “An updated software reliability model using the Shanker model and failure data,” *Quality and Reliability Engineering International*, **40:** 2078–2095.

Smit, N., Raubenheimer, L., Mazzuchi, T., and Soyer, R. (2024). “A Bayesian generalized Eyring-Weibull accelerated life testing model,” *Quality and Reliability Engineering International*, **40:** 1110–1125.

Song, Y., Mi, J., Cheng, Y., Bai, L., and Chen, K. (2021). “Reliability assessment for failure-dependent and uncertain systems: a Bayesian network based on copula method and probability-box,” *Quality and Reliability Engineering International*, **37:** 1894–1921.

Taylor, D., Rigdon, S.E., Pan, R., and Montgomery, D.C. (2024). “Bayesian D-optimal design for life testing with censoring,” *Quality and Reliability Engineering International*, **40:** 71–90.

Turner, B.M. and van Zandt, T. (2012). “A tutorial on approximate Bayesian computation,” *Journal of Mathematical Psychology*, **56:** 69–85.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J., and Aken, M.A. (2014). “A gentle introduction to Bayesian analysis: applications to developmental research,” *Child Development*, **85:** 842–860.

Wang, G., Shao, M., Li, J., Lv, S., Kong, X., and Zhou, Y. (2024). "Bayesian analysis of designed reliability improvement experiments with application to adaptive termination," *Quality and Reliability Engineering International*, **40:** 57–70.

Wiedner, C. and Han, D. (2021). "Bayesian inference for the simple step-stress accelerated life tests under order-restriction," *Procedia Manufacturing*, **55:** 147–153.

Xu, A. and Tang, Y. (2011). "Nonparametric Bayesian analysis of competing risks problem with masked data," *Communications in Statistics – Theory and Methods*, **40:** 2326–2336.