A Data Driven Recurrent Event Model for System Degradation with Imperfect Maintenance Actions

Akash Deep, Shiyu Zhou, Dharmaraj Veeramani

Department of Industrial and Systems Engineering, University of Wisconsin Madison, Madison, WI, USA

Abstract

While a large number of degradation models for industrial systems have been proposed by researchers over the past few decades, the modeling of impacts of maintenance actions has been mostly limited to single-component systems. Among multi-component models, past work either ignores the general impact of maintenance, or is limited to studying failure interactions. In this paper, we propose a multivariate imperfect maintenance model which models impacts of maintenance actions across sub-systems while considering continual operation of the unit. Another feature of the proposed model is that the maintenance actions can have any degree of impact on the sub-systems. In other words, we propose a multivariate recurrent event model with stochastic dependence, and for this model we present a two-stage approach which makes estimation scalable, thus practical for large-scale industrial applications. We also derive expressions for Fisher information so as to conduct asymptotic statistical tests for the maintenance impact parameters. We demonstrate the scalability through numerical studies, and derive insights by applying the model on real-world maintenance records obtained from oil rigs.

Keywords: Recurrent event model, Degradation model, Imperfect maintenance, Multivariate stochastic dependence

1 Introduction

In this paper, we focus on an event modeling problem where we consider multiple sequences of events which occur temporally (continuous in time t). The events in each sequence may be of different types and the same events can recur as well. The occurrence of an event influences occurrence of the same and other event-types. In other words, we have multiple stochastic processes which are interlinked. Under such a setting, we develop a statistical model which defines the interrelationships among these multiple event processes. This work is motivated by the opportunities and challenges in analyzing maintenance records of complex engineering systems.

An engineering system often consists of multiple sub-systems/components. Once setup and deployed, a system operates continuously to meet daily demands, and is maintained during the course of its operation to ensure smooth functioning. The maintenance actions considered here are primarily – Preventive Maintenance (PM) and Corrective Maintenance (CM), and are performed at the sub-system level. The PM is often conducted on a periodic basis including inspection, cleaning, and replacement of certain wear and tear components, while the CM is performed to repair the system after a failure occurs. As a result, the history of the CM can be viewed as the history of system failures.

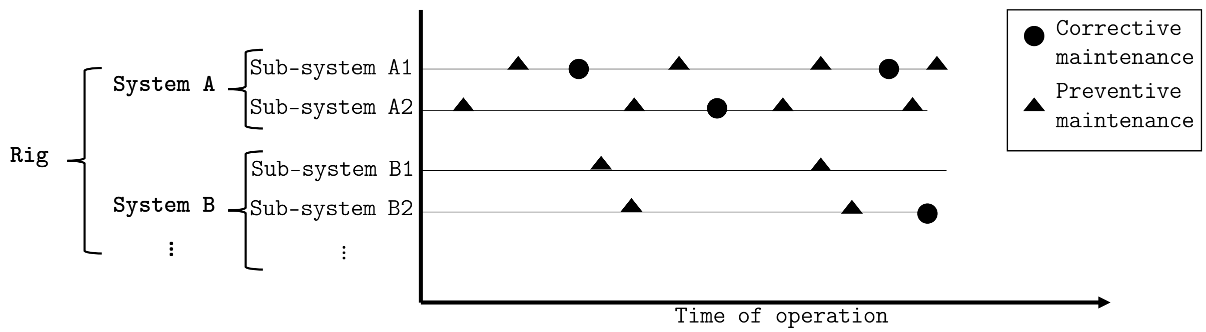

Figure 1: An illustration of maintenance actions conducted on an oil rig.

Figure 1 illustrates the maintenance record of an oil rig. In practice, we may have records from multiple similar rigs. The maintenance records data can provide valuable insights regarding system reliability and the effectiveness of various maintenance actions. From the maintenance records data, we want to build a rigorous statistical model to assess the impact of maintenance actions on the system reliability. Such knowledge will help the equipment manufacturers design better maintenance and warranty policies, thus enhancing customer satisfaction and gaining competitive advantage.

The goal of our work is to establish a general statistical model to model the intensity of recurring failure events (represented by the CM events) on a sub-system with the consideration of impacts of several maintenance actions (termed as *event-types*) on the same and related sub-systems. For example, consider the sub-system A1 in Fig. 1. The failure event occurs twice during the observation period (as there are two CM events). We treat the failure event as a recurrent event, instead of a terminal event. The intensity function of the failure event, which can be viewed as the instantaneous rate of failure occurrence over time, is an important indicator of the reliability of sub-system A1. We want to establish a failure event intensity model based on the maintenance record data, which will not only provide the failure intensity at a given time, but also quantify the impact of various maintenance actions on the failure intensity.

In the reliability literature, many stochastic approaches have been developed to model the terminating failure event (see Gorjian *et al.*, 2010, for a review). On the other hand, for recurrent events, e.g., a Renewal Process, a Non-Homogeneous Point Process (NHPP) (Rigdon & Basu, 2000), or a simple trend-renewal process (Lindqvist *et al.*, 2003) can be employed. However, in reality, a system's life is affected by various maintenance actions. Moreover, these maintenance actions are imperfect, i.e. they can have any degree of impact on the system degradation ranging from worse repair to perfect repair (Pham & Wang, 1996; Doyen & Gaudoin, 2004). However, the above simple recurrent event models cannot capture the impacts of maintenance actions. In the literature, there are data-driven models that consider general impacts of maintenance actions, such as, Kijima (1989); Yanez *et al.* (2002); Doyen & Gaudoin (2004). Further, there are also pre-specified models that are mainly used in the design of optimal maintenance policies (Pham & Wang, 1996, provide a review of such models). How to estimate the model parameters from data are not investigated in these models.

Within the recent literature of recurrent events, Wu (2019) present an exponential smoothing intensity model for series systems, but their work cannot distinguish the event-types. Doyen *et al.* (2017) present a new class of imperfect maintenance models using geometric reduction of age or intensity. Xu *et al.* (2017) present a two-level (system and component) modeling of recurrent events, however their method does not address multiple sub-systems/components on the same level. Unlike their work which models vertical interactions events, in the current work we are interested in horizontal modeling of events. Doyen *et al.* (2019) extend the virtual age models in a generalized framework which incorporates the CM and PM actions conducted on the same system without any distinction of components. The major limitation of the aforementioned literature is that all of them consider the system to be univariate in nature, i.e., even if there are components which operate together or in a close relationship, in principle, they are modeled independently and do not allow for mutual influence which is highly limiting (Meango & Ouali, 2019). Considering multi-component systems, Shaked & Shanthikumar (1986) propose a multiple (p, q) rule, where they essentially study the change in multivariate distribution (of component lifetimes) whenever a component has a perfect or minimal repair. This work is limited by the binary nature of the considered effect of repair. Zhang & Yang (2015) present a reliability model for multi-component systems under dependent competing risks with general repairs. Although they model multiple components, the impacts of a repair are limited only to the respective component which was repaired, and no mutual impacts are considered. While there are studies regarding failure interactions in multi-component systems, they have several limitations, such as, they cater only to simple system architectures, the dependence is mostly unidirectional, and more importantly, the aspect of interventions is not accounted for (see Meango & Ouali, 2019, for a recent review).

Alongside the above literature, several models are constructed from a predictive maintenance point of view, i.e., the basic idea is to anticipate future failures proactively, and to use them to drive maintenance decisions. This body of literature can be broadly classified into two categories – time-based maintenance and condition-based maintenance (see Alaswad & Xiang, 2017, for a recent review on methods). Time-based methods are considered static methods which model failure intensities, and in turn, determine a 'fixed' schedule of future maintenance actions (Wang, 2002, provide a review of such models). On the other hand, condition-based methods monitor the latest available data from system sensors (e.g., continuous degradation signals), and determine the schedule of maintenance (Keizer *et al.*, 2017). The degradation signal based methods have also mostly been developed for single component systems (Alaswad & Xiang, 2017). Although there are several methods developed for multi-component systems, almost all of them are either limited to first time failures or assume that after a failure, the repair resets the system "as good as new" (Nguyen *et al.*, 2014; Shafiee & Finkelstein, 2015; Bian & Gebraeel, 2014; Babishin & Taghipour, 2016; Liu *et al.*, 2017; Arts & Basten, 2018). The impact of imperfect maintenance actions on the system degradation is not considered in these works.

In light of the extant literature, we observe that general repair models are limited to univariate systems, or the multi-component models assume that components' life is reset upon a failure. To the best of our knowledge, there is no method to model stochastic interactions among components (or sub-systems) which caters to the impacts of maintenance actions. In practice, we usually have a multi-component system which has a long period of operation during which it experiences several kinds of PM/CM actions. Furthermore, realistically speaking, the effect of the maintenance actions lie between the two extremes of worse repair and perfect repair. Therefore, in this work, we fill this gap in the research literature by proposing a modeling method which addresses the continual operation of the system/sub-systems/components and simultaneously incorporates the impacts of several maintenance actions. The influence of any maintenance action is experienced by all sub-systems and can have a general degree of impact.

The rest of the paper is organized as follows. In Section 2, we describe the problem setting, state the assumptions and present the model. In Section 3, we present the model likelihood formulation followed by parameter estimation. We also describe a two-stage approach for parameter estimation, and propose an approach to test the significance of these parameters. In Sections 4 and 5, we present several numerical cases and a real-world case study, respectively. Section 6 concludes this paper with suggestions for future research directions.

2 Model Development

2.1 Data description and notations

Consider N identical and independent systems, in which each system consists of the same sub-systems (so we have N counts of each sub-system as well). Let $\mathbf{K}$ denote the set of types of maintenance actions (or event-types; indexed by k) present in the data across all unique sub-systems. There are no restrictions over the types of maintenance, however, generally in the rest of the paper we will only consider CMs and PMs on the sub-system level. The impact of any maintenance action is shared by all sub-systems. Then, with respect to a sub-system, we can define:

- **Self impact:** impacts due to a maintenance on the sub-system itself.
- **Mutual impact:** impacts due to a maintenance on a related sub-system.

For example, a sample of data is presented in Table 1 from a system containing two sub-systems (A & B). There are total four event-types, the set denoted as, $\mathbf{K} = \{1, 2, 3, 4\}$. Event-type 1 & 2 are CM and PM conducted on sub-system A respectively, similarly 3 & 4 are CM and PM conducted on sub-system B respectively. This system experiences its first event at time 11, when a PM is conducted on sub-system A. Although, this maintenance action is conducted upon sub-system A, the impact is experienced by both sub-systems. Continuing, the system experiences its next event at time 23 when a CM is conducted on sub-system B. The CM does not necessarily reset the sub-system and the system continues to operate. Over a period of time, the system experiences many other kinds of events and has its last event recorded at time 112.

Table 1: A sample data illustrating different kinds of events experienced by a system

Event-time (t)	Event-type (k)	Impact on sub-system A (SSA)	Impact on sub-system B (SSB)
11	2	(PM on) Self	Mutually experienced with SSA
23	3	Mutually experienced with SSB	(CM) on Self
46	1	(CM on) Self	Mutually experienced with SSA
..	..	..	..
105	4	Mutually experienced with SSB	(PM) on Self
112	1	(CM on) Self	Mutually experienced with SSA

In the remainder of this paper, the scope of our model development is on the system-level. For a system (say m), the information we have is an ordered sequence of events as: $(\mathcal{T}, \mathcal{K})_m = (t_{m1}, k_{m1}), (t_{m2}, k_{m2}), \cdots, (t_{mn_m}, k_{mn_m})$, where $t_{m1} < t_{m2} < \cdots < t_{mn_m}, k \in \mathbf{K}$. Together, the overall data across all similar systems are denoted as $(\boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{K}})$.

2.2 The proposed model

We consider two assumptions for model development. First, at any instant, no more than one event can occur. Formally, if $N(t)$ is the overall event counting process of the system (i.e., inclusive of all different event-types), then $P(dN(t) \geq 2) = 0$. This is a regularity assumption and is used to avoid identifiability issues of the model. This assumption is well supported in practice in univariate as well as multivariate point processes, e.g., Cox & Lewis (1972); Daley & Jones (2003). Second, each sub-system of the system, in fact, the entire system degrades with time. The baseline intensities of the sub-system represent the degradation of the sub-system. This assumption is considered from natural degradation of a physical system. If in a different context, we do not expect the intensity to represent degradation, then this assumption can be relaxed without any correction to the model. The baseline can be adjusted accordingly.

Using the above assumptions, we are now in a position to introduce the intensity model. The measure of the impact of a maintenance action is arithmetic reduction in intensity (ARI) just before and after the event (as in Doyen & Gaudoin, 2004). The impact depends on the instantaneous value of the intensity which is scaled by the impact parameter. Newby (1992) points out additive hazards models (AHM) by Pijnenburg (1991) is a natural fit for modeling hazard before and after the repair, and ARI model is structurally similar to AHM, therefore suitable for the current context. In summary, we have the baseline intensity which corresponds to natural degradation, and the explanatory effects of maintenance actions are accounted via their impact parameter values.

From a modeling point of view, we see these events as point processes in a temporal space. Corresponding to each event-type (say k), we define its intensity k as, $\lambda_k(t|\mathcal{H}_{t^-}) = P(dN_k(t) = 1|\mathcal{H}_{t^-})/dt$, where $\mathcal{H}_{t^-}$ represents history accumulated until time t. For brevity of notations, we drop the conditioning on history, i.e. let $\lambda_k(t) = \lambda_k(t|\mathcal{H}_{t^-})$.

The stochastic interaction between the processes is introduced in the following manner: let the i^{th} event occur at time t_i, and is of type k_i, then it changes the intensities of every process $k \in \mathbf{K}$. The change for an event-type k is denoted using $\Delta\lambda_k$. Mathematically, $\Delta\lambda_k$ is:

$$\Delta\lambda_k(t_i) = \lambda_k(t_i^+) - \lambda_k(t_i^-) = -\rho_{kk_i}\lambda_k(t_i) \tag{1}$$

where, $\lambda_k(t_i)$ is the intensity of the k^{th} event-type at the time t_i, and ρ_{kk_i} is the impact parameter governing the impact of k_i on k. In other words, the occurrence of an event decreases/increases the intensity of an event-type $k \in \mathbf{K}$ by a certain proportion (ρ_{kk_i}) of their respective instantaneous intensities ($\lambda_k(\cdot)$). For example, in the maintenance context, occurrence of a CM action would improve the system and therefore decrease the chances of future failures (consequently maintenance actions).

This yields that, at any general time t, the proposed process is (see Appendix A for proof):

$$\lambda_k(t|\mathcal{H}_{t^-}) = \lambda_{k0}(t) - \sum_{j=0}^{N_t-1}\left[\rho_{kk_{(N_t-j)}} \cdot \prod_{r=N_t-j+1}^{N_t}\left[(1-\rho_{kk_r})\right] \cdot \lambda_{k0}(t_{N_t-j})\right] \tag{2}$$

where, λ_{k0} is baseline intensity of event-type k, N_t is the succinct form of $N(t)$. In the above, we use convention $(1-\rho_{kk_r}) = 1$ for $r = N_t + 1$.

Please note that, for ease of reading, we have dropped the unit's subscript m in this section. To facilitate interpretation, the intensity of event-type k at any time t is the respective baseline intensity at time t minus the accumulated change (drop or increase) in intensities until time t. The accumulated change can be broken down as the contribution of all N_t, i.e., number of events until time t, and we use the iterator j which ranges from 0 to $N_t - 1$. As j moves 0 to $N_t - 1$, the index $N_t - j$ traverses backwards from the most recent event to the first event. When j is zero, $N_t - j$

corresponds to the most recent event and the change induced by this event is $\rho_{kk_{N_t}} \cdot \lambda_{k0}(t_{N_t})$. Next, when $j = 1$, the change induced is $\rho_{kk_{(N_t-1)}} \cdot (1 - \rho_{kk_{N_t}}) \cdot \lambda_{k0}(t_{N_t-1})$, and so on. We note that, if required, the impact of past events can be truncated by introducing a finite-sized memory term. We use Σ to denote the impact parameter matrix containing all $\rho_{kk'}$, and let $\boldsymbol{\rho}_k$ be the k^{th} row of the impact parameter matrix.

Remark. *If $\rho_{kk'} \leq 1 \quad \forall \{k, k'\} \in \mathbf{K}$, then $\lambda_k \geq 0$ always.*

Remark. *Here, $\rho_{kk'} = 1$ is equivalent to the case of perfect maintenance action, $\rho_{kk'} = 0$ represents minimal impact of maintenance action, and $\rho_{kk'} < 0$ represents that the impact of maintenance action worsens the sub-system (Pham & Wang, 1996).*

Remark. *The density function of the next event time T_b after the latest event time at t_a can be computed as*

$$f_k(t_b|\mathcal{H}_{t_a^-}) = \lambda_k(t_b) \exp\Big(- \int_{t_a}^{t_b} \lambda_k(s) ds \Big) \tag{3}$$

In regard to modeling of the baseline process, there are no restrictions over the choice of the process, as far as it is mathematically a valid process. We tabulate some popular methods in Table 2. For the remainder of the paper, we use Power Law Process as the baseline process which is given as $\lambda_{0k}(t) = \alpha_k \beta_k t^{\beta_k - 1}, \alpha_k > 0, \beta_k > 0$. We denote $\boldsymbol{\theta}_k = \{\alpha_k, \beta_k\}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_K\}$.

Table 2: Considerations for the choice of baseline process

Baseline process	Rate ($\lambda(t)$)	Remarks
Homogeneous Point Process	λ_{exp}	Remains constant over time
Power Law Process (PLP)	$\alpha\beta t^{\beta-1}$	A popular NHPP model
Exponential Law Process	$\exp(\alpha + \beta t)$	Rate is log-linear

In the current formulation, the baseline process, which represents the degradation are independently modeled, and we incorporate the stochastic interaction between intensities through the impact parameters. In a system where we expect that sub-systems also feature degradation dependence, such a dependence can be incorporated through the baseline processes. From a modeling point of view, copula-based methods can be employed to construct such a method in a purely data-driven fashion (see Wang & Pham, 2011).

To assist in the interpretation of the proposed model, we now demonstrate how our proposed multivariate stochastic process evolves over time. For this exercise, we consider a system which experiences three different kinds of events (say three CMs, each corresponding to the three sub-systems in a system). Figure 2 illustrates sample paths from one realization of these three event intensities. For each sub-system, the observation period begins at $t = 0$ and CMs occur over time, with the first event being CM of first sub-system. The immediate resultant intensities are altered (increased/decreased) according to the impact parameters (Σ) – it reduces by factor of 0.5 for sub-system 1, and by factor of 0.2 for the second and the third sub-system. After this shift in intensities, they continue to grow at the same rate as their respective baseline intensities. The second event is the CM on the third sub-system and it reduces the intensity by a factor of 0.5. This repair, however, adversely affects the first sub-system and shifts its intensity upwards by factor of 0.1, and has no effect on the second sub-system. Going further, it can be observed that the interplay of self as well as mutual impact parameters results in different trajectories of intensities. The knowledge of maintenance impacts offers insights to component relationship(s), and can be effectively utilized in modeling their degradation as well as failure prediction.

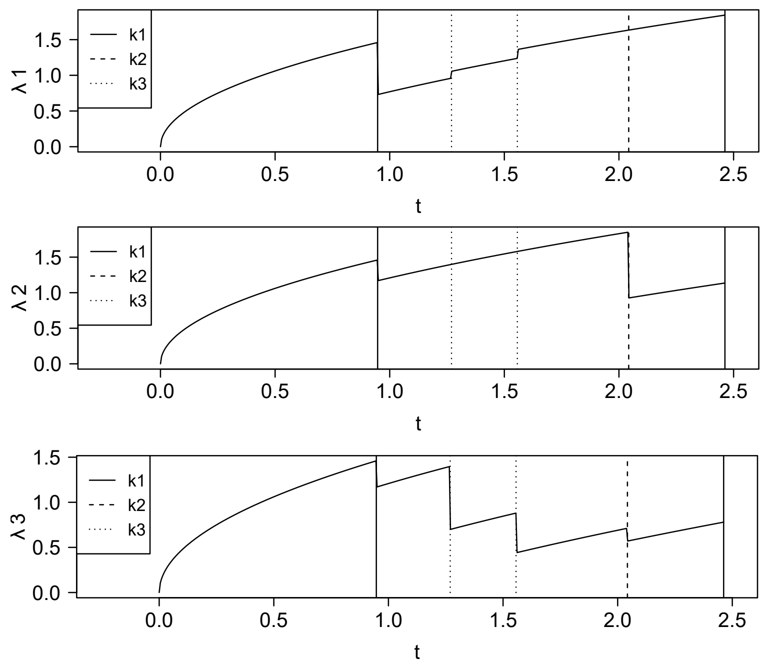

Figure 2: A sample illustration of multivariate imperfect-maintenance model with $K = 3$. The baseline for the three event-types follow PLP with same parameters, $\boldsymbol{\alpha} = 1, \boldsymbol{\beta} = 1.5$, and impact parameters are $\Sigma = \begin{bmatrix} 0.5, 0.0, -0.1 \\ 0.2, 0.5, 0.0 \\ 0.2, 0.2, 0.5 \end{bmatrix}$

3 Model Parameters Estimation

Having established the model, this section is dedicated to estimation of parameters – baseline ($\boldsymbol{\Theta}$) and impact parameters (Σ). We present a likelihood based approach to estimate the parameters.

Proposition 1. *Assuming that systems are i.i.d. distributed, the overall log-likelihood across N systems will be*

$$l = \sum_{m=1}^{N} l_m \tag{4}$$

where, l_m *is log-likelihood for* m^{th} *system which having experienced* n_m *number of events can be expressed as:*

$$l_m(\mathbf{\Theta}, \Sigma|(\mathcal{T}, \mathcal{K})_m) = \sum_{i=1}^{n_m} \Big(\log(\lambda_{k_{mi}}(t_{mi})) \Big) - \sum_{k=1}^{K} \Lambda_k(t_{mn_m}) \tag{5}$$

where, $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$ *is the cumulative intensity function for event-type* k*.*

The schematic proof of the log-likelihood expression is presented in the online supplemental material.

Next, we plug the above formulation in equation 5 which can be maximized using numerical routines to obtain the parameters. It can be observed that, in this model, we simultaneously estimate $K^2 + 2K$ parameters, which can be computationally demanding even for smaller Ks. Moreover, often the log-likelihood has highly non-convex behavior which leads to bad estimates. This motivated us to develop a two-stage procedure which is computationally efficient, parallelizable, and more stable than the general estimation.

3.1 A two-stage approach for parameter estimation

The intuition behind this two-stage approach is that by exploiting the structure of the proposed process, we can split the parameters into two separate sets (baseline and impact parameters) and estimate them sequentially in two steps. In fact, the two-stage method exploits two facts – (i) the impact parameters begin to play their roles only after the occurrence of the first event, and, (ii) the set of first occurring events of a PLP follows a Weibull distribution, which means we can quickly use any standard algorithm. Please note that, depending upon the choice made for the baseline process (from Table 2), the time-to-first-event's distribution would change. The assumptions we make here are – (i) we have sufficient data available in the historical records, and (ii) the set of first event-types across all systems contains all $\mathbf{K}$ events.

The two-stage approach proceeds as follows:

Stage 1: Strip the first occurring event-times and event-types of all systems present in the record. Let this yield a pair of vectors of length N as $(\mathbf{t}_1, \mathbf{k}_1)$. Here, we can observe that the first events do not involve impact parameters. Hence, we can individually estimate the baseline parameters of event-type k denoted by $\boldsymbol{\theta}_k$. From the k^{th} event-type's point of view, the data $(\mathbf{t}_1, \mathbf{k}_1)$ can be bifurcated into instances when the k^{th} event-type is observed or censored. Therefore, the likelihood can be expressed as:

$$l(\boldsymbol{\theta}_k|\mathbf{t}_1, \mathbf{k}_1) = \sum_{m=1}^{N} \mathbb{1}(k_m = k)\Big\{ \log f_k(\mathbf{t}_{1m}; \boldsymbol{\theta}_k)\Big\} + \sum_{m=1}^{N} \mathbb{1}(k_m \neq k)\Big\{ \log S_k(\mathbf{t}_{1m}; \boldsymbol{\theta}_k)\Big\} \tag{6}$$

where, $f_k(\cdot), S_k(\cdot) = 1 - \int f_k(\cdot)$ are the density and survival functions corresponding to Weibull distribution of event-type k. The first term on the right-hand side identifies when the k^{th} event-type occurred, i.e., $\mathbb{1}(k_m = k)$ and the contribution is the density function, whereas, the second term accommodates for censored cases. Upon optimizing the log-likelihood across $\mathbf{K}$, the obtained maximum likelihood estimates (MLE) $\hat{\boldsymbol{\theta}}_k \quad \forall k \in \mathbf{K}$ can be together denoted as $\hat{\boldsymbol{\Theta}}$. We would like to note that the estimates $\hat{\boldsymbol{\Theta}}$ from the restricted dataset are MLE, and therefore, they are asymptotically consistent, efficient as well as follows a joint normal distribution.

Stage 2: In the next step, we plug in $\hat{\boldsymbol{\Theta}}$ in the full log-likelihood (4) to obtain the estimate for Σ as follows:

$$l(\Sigma|\hat{\boldsymbol{\Theta}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{K}}) = \sum_{m=1}^{N} \Big[\sum_{i=1}^{n_m} \Big(\log(\lambda_{k_{mi}}(t_{mi})) \Big) - \sum_{k=1}^{K} \Lambda_k(t_{mn_m}) \Big] \tag{7}$$

Here, the above log-likelihood is still difficult to optimize directly. However, we can observe that the parameterization of the current model allows us to optimize the impact parameters associated with one event-type separately from other event-types (see proof in Appendix B). In other words, the log-likelihood in equation (7) can be decomposed into log-likelihoods of each $\boldsymbol{\rho}_k$, which can computed as follows:

$$l(\boldsymbol{\rho}_k|\hat{\boldsymbol{\Theta}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{K}}) = \sum_{m=1}^{N} \Big[\sum_{i=1}^{n_m} \mathbb{1}(k_{mi} = k) \big(\log(\lambda_{k_{mi}}(t_{mi})) \big) - \Lambda_k(t_{mn_m}) \Big] \tag{8}$$

Through this procedure, the numerical burden of optimization is reduced from one routine of size K^2 to K routines, each of size K, thus offering the opportunity to parallelize the estimation process. Upon optimizing the log-likelihood across $\mathbf{K}$, the obtained estimates $\hat{\boldsymbol{\rho}}_k \quad \forall k \in \mathbf{K}$ can be together denoted as $\hat{\Sigma}$.

At the end of the two-stage procedure, we will have estimated $\hat{\boldsymbol{\Theta}}$ as well as $\hat{\Sigma}$. The estimate $\hat{\Sigma}$ is asymptotically consistent as well as normal. Consistency of $\hat{\Sigma}$ can be argued by considering that if $\boldsymbol{\Theta}$ were known, then $\hat{\Sigma}$ enjoys all results of MLE. Since, $\hat{\boldsymbol{\Theta}}$ converges in probability towards $\boldsymbol{\Theta}$, therefore, $\hat{\Sigma}$ is consistent. Additionally, the estimate is also asymptotically normal owing to the results of the two-step MLE method. Please see the online supplemental material and cited references for more details.

3.2 Asymptotic test of $\rho = 0$ using Observed Fisher Information

Having obtained the model parameter estimates, we now proceed to test the significance of the impact parameters. The approach we adopt is through the observed Fisher information matrix. Since we consider the baseline parameters to be nuisance parameters, our efforts are dedicated to derive the asymptotic variance of impact parameters. We note that other simulation-based methods, such as Monte Carlo or bootstrapping, can also be employed to obtain the variance of the parameters.

Lemma 3.1. *The asymptotic distribution of the impact parameter estimate $\hat{\rho}_{kk'}$ is*

$$\hat{\rho}_{kk'} \sim \mathcal{N}(\rho_{kk'}, \sigma_{\rho_{kk'}}) \tag{9}$$

where, for inference we use, $\hat{\sigma}_{\rho_{kk'}} = \sqrt{-H(\hat{\Sigma})^{-1}_{\{\rho_{kk'}\}}}$. $H(\cdot)$ is Hessian derived in Appendix B, and the expression is in equation (18)*.*

We are mainly interested in testing if the impact parameter is significantly different from zero (meaning no impact). Therefore, the null hypothesis is $H_0 : \rho_{kk'} = 0$, versus the alternative $H_A : \rho_{kk'} \neq 0$ at the critical significance level (α_c) to reject the hypothesis. In an empirical study, we found that for testing $\rho = 0$, normal assumption holds well, and hence, the following test can be used.

Proposition 2. *Since the parameter estimates asymptotically follow a normal distribution* (9)*, under null we can write the distribution as follows*

$$\hat{\rho}_{kk'} \sim \mathcal{N}(0, \sigma_{\rho_{kk'}}) \tag{10}$$

The p-values can be easily computed to test the hypothesis.

4 Numerical Study

We dedicate this section to design and perform numerical experiments. The overall motivation is to demonstrate the data generation and retrieval of original parameters. In section 4.1, we outline the numerical simulation method, describe the several cases considered for experiments, and provide an overview of the experiment procedure. In section 4.2, we present the estimation results and discuss our findings.

4.1 Data generation and numerical experiment settings

The process to generate data of presented multi-event process is outlined in algorithm 1.

We consider two different settings. In setting I, we illustrate the effectiveness of two-stage estimation procedure in the presence of a small number of event-types (i.e. $K = 2$). We use four separate sets of baseline parameters, and within each set, we perform four cases of impact parameters. The first case reflects that there is no impact of other events. The second and third cases introduce mild and high magnitudes of self as well as mutual impacts. Finally, the fourth case only has self impacts and no mutual impacts.

Setting II is dedicated to illustrate the scalability of the model considering a large number of event-types. We also use different patterns of self and mutual impacts so as to demonstrate that the

Algorithm 1 Simulation algorithm for multivariate imperfect-maintenance process

Event-sequence generation:

Specify number of units N, number of event-types K, baseline parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, impact parameters (Σ), and a stopping criteria. In our simulations we use max-time (T_{max}) as the stopping criterion.

for m in $1 : N$ **do**

Initialize $(\mathcal{T}, \mathcal{K})_m = \varnothing, t = 0, n_m = 0$

while $t < T_{max}$ **do**

if $\mathcal{T}_m = \varnothing$ **then**

generate **K** candidate event times $s_1, s_2, \cdots, s_K$ using $s_k \sim Weib(\alpha_k, \beta_k)$, corresponding to each event type.

$\mathcal{T}_m \leftarrow \min(s_1, s_2, \cdots, s_K)$

$\mathcal{K}_m \leftarrow (k : \min(s_1, s_2, \cdots, s_K))$

$n_m \leftarrow n_m + 1$

else

generate $\bar{\lambda} = \sum_{k=1}^{K} \lambda_k^*(\max(\mathcal{T}))$ using equation 2

generate $u \sim \mathcal{U}[0, 1]$

generate $w \leftarrow -\log(u)/\bar{\lambda}$

$t \leftarrow t + w$

$\mathcal{T}_m \leftarrow \mathcal{T} \bigcup t$

generate $d \sim \mathcal{U}[0, 1]$

$l : \sum_{k=1}^{l-1} \lambda_l^*(t) < d\bar{\lambda} \leq \sum_{k=1}^{l} \lambda_l^*(t)$

$\mathcal{K}_m \leftarrow \mathcal{K} \bigcup l$

$n_m \leftarrow n_m + 1$

end if

end while

return $\mathcal{T}_m \backslash \mathcal{T}_{mn_m}, \mathcal{K}_m \backslash \mathcal{K}_{mn_m}$

end for

proposed strategy can well recover any structure between events. In this setting, during estimation, we assume that the baseline parameters are known, hence the focus solely lies in estimation of the impact parameters. We present in total 5 cases, where two cases have $K = 4$ but different pattern of impacts, then we increase K to 8, 10 and 12 with different patterns of impacts.

The procedure we adopt to carry out the experiment is as follows:

Algorithm 2 Procedure for numerical experiments

Specify number of runs B, number of units N (same for each run), number of event-types K, baseline parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, impact parameters (Σ), and a stopping criteria.
for b in $1 : B$ **do**
    Use algorithm 1 to generate N sequences
    Employ two-stage approach provided in section 3.1
    Save $\hat{\Theta}_b$ and $\hat{\Sigma}_b$
end for
Report means of estimates

4.2 Results and Discussion

The estimated parameters from setting I are in Tables 3 and 7 (in Appendix C). For setting II Fig. 3 is shown below and other cases are presented in figs. 5 to 8 in Appendix C. From the results we have the following observations:

For setting I, we find that in general, the estimation procedure yields good estimates of the original parameters. The estimates are also normally distributed. However, for the cases where the impact parameter is 0.9, i.e., near to the boundary, the estimated distribution is not well-approximated, and the estimated mean is far from the true parameter. In this case, we would require a large number of samples so that the MLE assumptions hold well near the boundary.

We also studied the mean squared error (MSE) of parameter estimates and its decomposition. It is a well-known result that MSE can be decomposed into the squared-bias and variance of the estimator (James *et al.*, 2013). Two main observations from this study were – first, the decomposition remain valid and they show that bias always remains very small compared to the variance. Second, as we increase the number of sequences in the sample, MSE and its components decrease.

In setting II, the estimates are very good even when small number of sequences are considered. We observe that the estimated parameters are quite close to the original parameters. In some cases, such as $\rho_{1,11}$ for $K = 12$, the estimated parameters are a bit far from the true parameters. To tackle this, we have to increase the length sequence; however, that comes at a greater cost of computation time. It is worth mentioning that knowledge of baseline parameter as accurate possible is important for better stage 2 estimation.

From a computational point of view, by increasing the values of K, we observe the time to conduct the entire Algorithm 2 increases substantially. Since we expect to have larger sequences for bigger K such that all event-types show up in the sequence, the stopping criterion has to

Table 3: True and estimated parameters for the numerical experiments considered under setting I.

$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	Case (Σ)	N	T_{max}	$\hat{\alpha_1}$	$\hat{\beta_1}$	$\hat{\alpha_2}$	$\hat{\beta_2}$	$\hat{\rho_{11}}$	$\hat{\rho_{12}}$	$\hat{\rho_{21}}$	$\hat{\rho_{22}}$
1	1	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	100	20	1.014 (0.172)	1.025 (0.110)	0.988 (0.138)	1.045 (0.112)	0.003 (0.019)	0.004 (0.017)	0.004 (0.017)	0.005 (0.016)
		$\begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix}$	100	20	1.003 (0.159)	1.022 (0.113)	1.021 (0.164)	1.016 (0.110)	0.424 (0.184)	0.381 (0.179)	0.375 (0.173)	0.416 (0.188)
		$\begin{bmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{bmatrix}$	100	20	1.012 (0.148)	1.025 (0.118)	1.013 (0.159)	0.999 (0.116)	0.882 (0.189)	0.719 (0.257)	0.658 (0.255)	0.875 (0.163)
		$\begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}$	100	20	0.991 (0.144)	1.013 (0.112)	1.027 (0.154)	1.040 (0.106)	0.762 (0.250)	0.175 (0.257)	0.212 (0.282)	0.822 (0.213)
	1.5	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	100	5	1.006 (0.103)	1.531 (0.183)	1.014 (0.109)	1.535 (0.165)	0.007 (0.044)	-0.006 (0.029)	0.011 (0.043)	0.001 (0.029)
		$\begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix}$	100	5	0.988 (0.074)	1.544 (0.175)	1.009 (0.104)	1.514 (0.177)	0.383 (0.162)	0.350 (0.153)	0.369 (0.161)	0.474 (0.133)
		$\begin{bmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{bmatrix}$	100	5	1.013 (0.104)	1.521 (0.198)	1.001 (0.092)	1.503 (0.164)	0.727 (0.223)	0.796 (0.194)	0.733 (0.201)	0.983 (0.044)
		$\begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}$	100	5	0.989 (0.093)	1.553 (0.182)	0.991 (0.097)	1.523 (0.182)	0.827 (0.152)	0.037 (0.234)	0.045 (0.257)	0.911 (0.079)

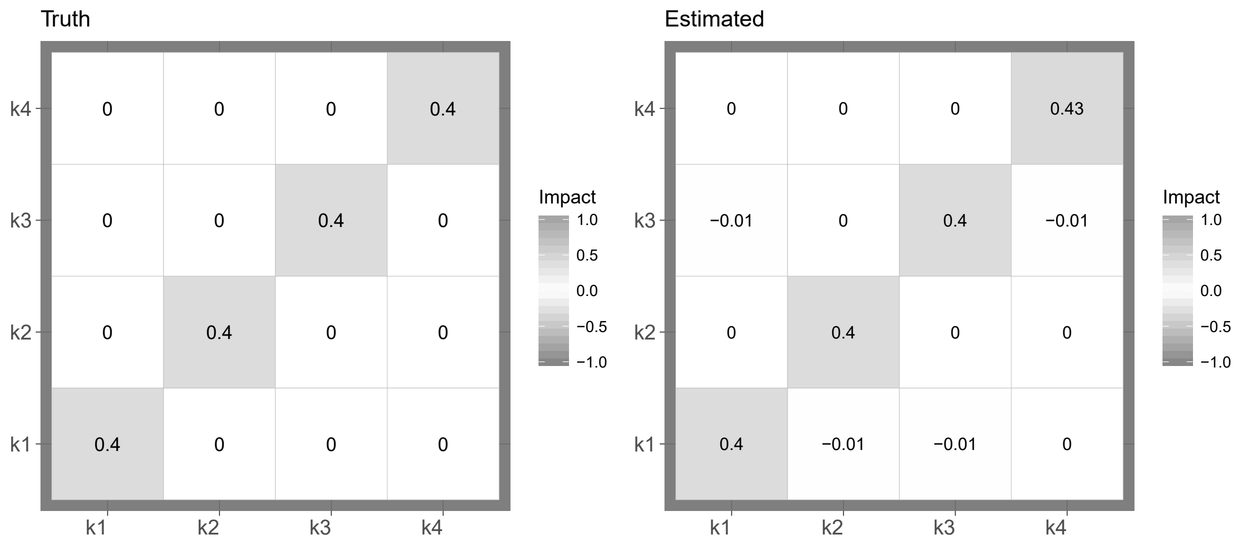

Figure 3: True and estimated impact parameters for the numerical experiments considered under setting II. Number of runs: 200.

be increased, thus leading to more time required for sequence generation. During estimation, although the dimensionality of K linearly increases the number of parameters to be estimated, the non-convexity of the likelihood function results in longer time for convergence of the optimization procedure.

In a data-rich environment, the proposed estimation procedure can be effectively employed. And, in the cases where limited data is available, say early stages of a system, the explicit expression of provided Hessian can be used to obtain the variance of parameters (assuming asymptotic normality), which can guide the uncertainty associated with the estimation. Besides, we can also resort to simulation-based methods as a mathematical approach which can provide the uncertainty quantification. However, simulation-based methods can be computationally intensive.

We also investigated the impact of data contamination on parameter estimation. We expect that, in real-life, it may happen that certain records may be populated incorrectly, such as, a maintenance conducted on one sub-system has been mistakenly entered as an event on another sub-system. From a mathematical viewpoint, this results in the recorded event-type sequence ($\mathcal{K}$) to be different from the actual sequence. Please note that the recorded event-time sequence ($\mathcal{T}$) remains the same. Through some additional experiments, we found that if the percentage of data contamination is mild (around 10%), then the estimated parameters stay close to the true parameters. The deviations increase with further contamination. In practice, however, we have dedicated management systems to record maintenance actions and, therefore, we expect that such data-entry errors to be infrequent. The details of the study have been provided in the online supplementary material.

5 Case study on oil rig data

This case study is based on real-world from maintenance records collected from oil rigs. An oil rig comprises several large systems, such as Derrick Hoisting System (DRHS), Riser Tensioner System (RTS) etc., and each system comprises several sub-systems which themselves can be a collection of components, such as, valve, pump, compensator etc. in DRHS, and recoil valve, sheave assembly etc. in RTS. The rig operates continuously to meet daily demands of petroleum production requirements. Maintenance actions are performed *in-situ* on a regular basis (PM), and on event of an unanticipated failure, the system is shut-down until resources are available for corrective maintenance (CM). The maintenance records are generated through a management system where associated with each record a time-stamp is available. The data-entry consists of two parts – first, the structured part where a technician enters information, such as, system, sub-system, maintenance code, failure etc. from a drop-down list. Second, the unstructured part, where the technician types-in the textual information, such as, instruction, description etc. For each record, a time-stamp is generated which remains fixed (see Table 4 for sample). We analyzed maintenance data from 10 rigs which can be considered identical with respect to structure of systems and their functionality. We assume the downtime to be negligible in comparison to the running time. Due to high reliability of the equipment, the records we have are mostly dominated by PM actions.

In the following, we independently study two systems (DRHS and RTS), both comprising of two sub-systems with each sub-system having PM and CM records available. First, we adjust the calendar time to operation time of system. Then, considering a system we have four different event-types named as – $\{1a, 1b, 2a, 2b\}$ corresponding to CM, PM for sub-system 1, and CM, PM for sub-system 2 respectively. Therefore, intensity functions $\lambda_{1a}\&\lambda_{2a}$ denote failure intensities of sub-system and 1 and 2, respectively, and they are the intensities of interest. Occurrences of these maintenance actions have the potential to impact any sub-system.

We employ the two-stage estimation method to obtain baseline parameters, followed by impact parameters. Since we are only interested in failure intensities of sub-systems, we only estimate baseline and impact parameters for event-type $1a$ and $2a$. In other words, studying ρ_{1a1a} would provide the impact of a CM on sub-system 1, and parameter ρ_{1a1b} would correspond to impact of PM of sub-system 1 on itself. The parameters ρ_{1a2a} & ρ_{1a2b} would convey the mutual effects on sub-system 1 due to the events conducted on the sub-system 2. The remaining parameters would have similar meanings with respect to sub-system 2.

Considering PLP as the baselines, we compute the baseline and impact parameters. Please note that baseline parameters are not reported since they can potentially convey sensitive information

Table 4: A snapshot of sample data obtained from an off-shore rig.

System	Sub-System	Type	Time (adj.)	Action	Failure Mech	Instructions
RTS	1	CM	116	Other	Leakage	Inspection of valves with in a X years period.
RTS	1	CM	405	Other	Clearance	Inspection of valves with in a X years period.
RTS	2	CM	541	None	General	Reqd AA to top up BB
RTS	2	PM	648			1. tighten bolts. 2. Examine for pitting - corrosion - scoring - flaking
DRHS	1	CM	135	Repair	Breakage	
DRHS	1	CM	212	Repair	Deformation	
DRHS	1	PM	412			1. Examine for wear, corrosion, damage, and security of all fasteners 2. Grease components
DRHS	2	PM	526			Examine for damage - dirt - debris - rust - dirt - leaks
DRHS	2	PM	763			Examine for damage - dirt - debris - rust - dirt - leaks

regarding their operation. As an illustration, in Fig. 4 we provide sample event intensities path from the DRHS system in a range of approximately 100 time-units. During the time-period presented in the figure, there are two CMs carried out for sub-system 1, and only one CM for sub-system 2, whereas there are several PM actions for both sub-systems. Regarding the impact parameters, the asymptotic statistical test presented in section 3.2 is also carried out. The null hypothesis is $H_0 : \rho_{kk'} = 0$, vs $\rho_{kk'} \neq 0$. We choose $\alpha_c = 0.05$ as the critical significance level. We would like to note that we also used empirical bootstrapping, and found the resultant distribution to appear approximately normal. We adopt the same procedure for the second system, and the results are tabulated together in Table 5. Below, we provide a discussion on the parameter estimates, their statistical test results, and model performance.

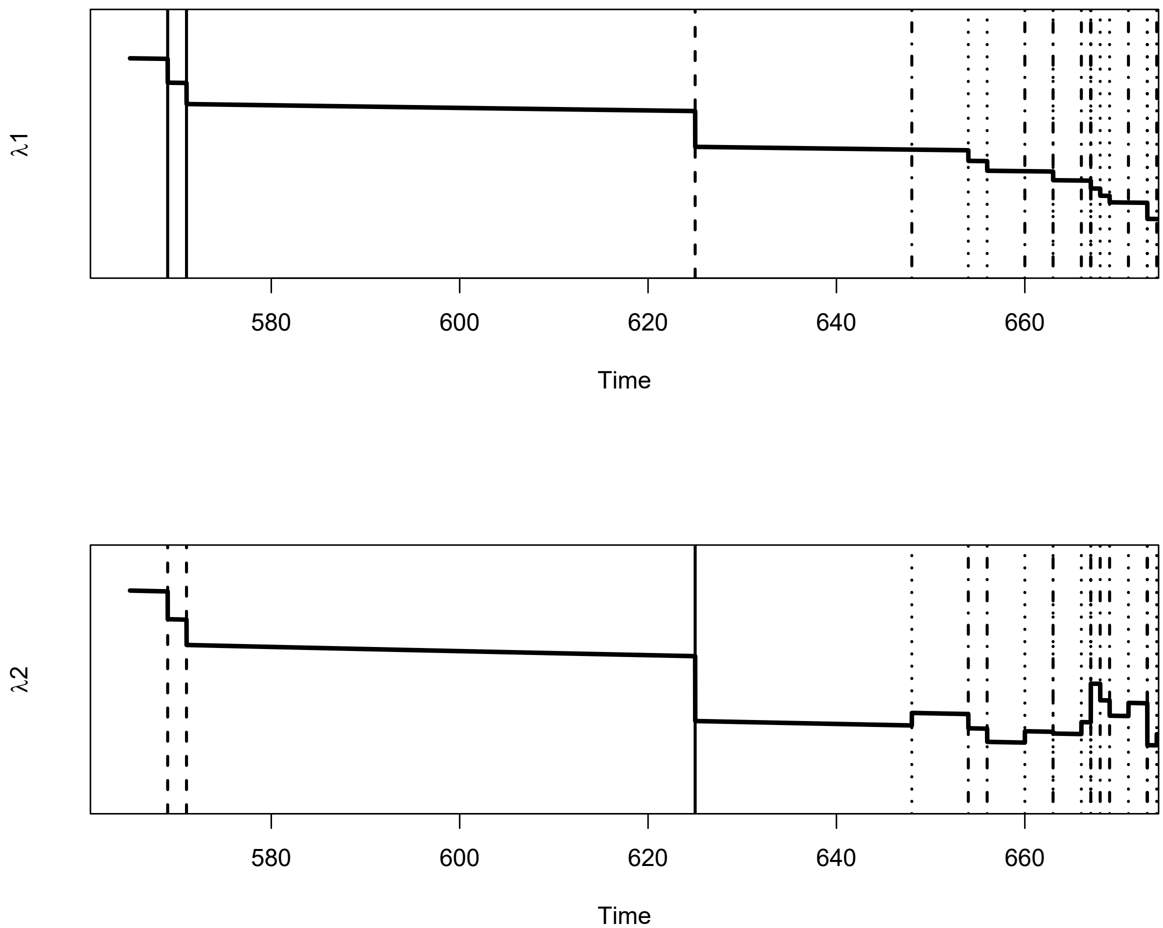

Figure 4: Sample intensity paths of two sub-systems present in DRHS system in the same time-period. Solid vertical lines – CMs on the self, Dashed vertical lines – CMs on the other sub-sytem, Dotted vertical lines – PMs on self, Dot-dashed vertical lines – PMs on the other sub-system.

First, we observe that, for both of the systems, the self impacts of CM actions are higher than those of PMs. Individually, these impact values are not substantially high (say close to 1), and this observation can be ascribed to the fact that although the sub-system is replaced/repaired, this action does not improve the effects of overall degradation of the system and the working environment. Hence, the repair is not perfect (a realistic observation).

Second, although there are significant mutual impacts of CM actions, not surprisingly, the effect of a CM on self is higher than on the other sub-system. The findings show that there are substantial mutual effects shared between the sub-systems. Such observations are realistically appealing since these sub-systems operate physically together, such as valve and pipe, and sheave and tensioner. Therefore, the effect of a repair or replacement of one sub-system percolates into the other sub-

Table 5: Obtained impact parameters and results from asymptotic test applied on the real-world case study data. Based on chosen critical level of significance, $\alpha_c = 0.05$, we reject the null hypothesis in all of the cases where the test statistic were computable.

System	Sub-system	Impact parameters			
DRHS	1	ρ_{1a1a}	ρ_{1a1b}	ρ_{1a2a}	ρ_{1a2b}
		0.1138	0.0898	0.2278	-0.00012
	p-value	$< 1e-4$	$< 1e-4$	$< 1e-4$	$< 1e-4$
	2	ρ_{2a1a}	ρ_{2a1b}	ρ_{2a2a}	ρ_{2a2b}
		0.0870	0.0725	0.2551	-0.0677
	p-value	$< 1e-4$	$< 1e-4$	$< 1e-4$	$< 1e-4$
RTS	1	ρ_{1a1a}	ρ_{1a1b}	ρ_{1a2a}	ρ_{1a2b}
		0.292	0.0043	0.160	0.0445
	p-value	$< 1e-4$	–	–	–
	2	ρ_{2a1a}	ρ_{2a1b}	ρ_{2a2a}	ρ_{2a2b}
		0.0678	0.01696	0.2286	0.0564
	p-value	$< 1e-4$	$< 1e-4$	$< 1e-4$	$< 1e-4$

system as well.

Third, regarding impacts of PMs, we observe that they have relatively less impact on improving the failure intensity of sub-systems. As noted in Table 4, more often these actions are like cleaning which do not substantially contribute to the health of the sub-system. However, the statistical tests do indicate that these improvements are significant. Thus, based on the current study on this dataset, a take-away is that although the PM event effects are small, they are effective.

Fourth, we note here that the sequences at hand are long (over 50 events per sub-system). This yields that if we use the Fisher information approach to compute variance, then the computed values of Hessian are large owing to numerically large values of time. Upon taking the inverse, we find that the variance values are extremely small in magnitude. Thus, we end up rejecting tests even when the parameter estimates seem very close to zero. Finally, we note that, in some cases, we were unable to calculate the test statistic. In these cases, inverting the information matrix yielded negative variance values. To overcome this challenge, one can use generalized inverses to create pseudo-variance matrices which are well-conditioned (Gill & King, 2004).

Model performance comparison

We use two information criteria to assess our model's performance. As information criteria are relative measures to compare the models, we compare the proposed method against the same formulation, however, the impact parameters are ignored. In other words, the competitive model does not have information of past events and it is simply the baseline process. The two criteria are as below:

Akaike Information Criterion (AIC): This criterion offers us a relative numerical comparison of model. The AIC of a model can be computed as follows (Akaike, 1998):

$$AIC = 2q - 2\log(\hat{L}) \tag{11}$$

where, q is the number of estimated parameters in the model, and $\hat{L}$ is likelihood value from the fitted model. Bayesian Information Criterion (BIC): This criterion is related to AIC, and penalizes number of parameters based on sample size (Schwarz *et al.*, 1978).

$$BIC = q * \log(n_N) - 2\log(\hat{L}) \tag{12}$$

where, $n_N = \sum_{m=1}^{N} n_m$ is the number of observations in the model.

System	Model	AIC value	BIC value
DRHS	Without considering impact parameters	1083876	1083893
	Considering impact parameters	508886.7	508920.4
RTS	Without considering impact parameters	2466030	2466047
	Considering impact parameters	858816.9	858850.6

Table 6: Comparison of AIC and BIC values obtained from two models for both systems

For both systems, we observe that when we consider impact parameters, they offer a lower value of AIC as well as BIC, thus indicating that considering impact parameters yields a better model. We also observe that the difference between the two models is substantial, and this observation can be attributed to the likelihood part of these criteria. We conclude that considering maintenance impacts indeed offers a better model.

6 Conclusion and Future Directions

In this paper, we have presented a multivariate stochastic process to address the impact of maintenance actions which is experienced by multiple components present in a system. To estimate the parameters of the model, we developed a two-stage approach which offers good scalability in practice. Furthermore, we also used a asymptotic testing procedure to test for the significance of impact of parameters. Practitioners should account for estimates and asymptotic test results when taking managerial decisions. Development of better tests for finite-sequences is one of our aims for future work.

We believe that there are several possibilities to extend the proposed model not just theoretically, but also from an application point of view. An immediate application of the proposed model is maintenance policy optimization of multi-component systems with stochastic interaction. Two main challenges are – first, the intensity is history-dependent, and to tackle this we have to identify the mean-rate of this stochastic process. And, second, the intensities can be seen as convolution of impacts from other stochastic processes (i.e., maintenance actions), therefore, computing the mean-rate is not very straightforward and might require mathematical approximations. Another extension of current formulation is to model the cumulative impact of simultaneous/tied events. This scenario arises when a system is brought into repair – it may happen that several other parts might be repaired or replaced. Another assumption regarding the failure mechanism provides us the opportunity to develop failure-mode specific impact parameters. In our work above, we interchangeably used events to denote a component always failing under same failure mode, however, it will be more realistic to model components and failure modes as different stochastic dimensions, probably as a marked point-process. Another interesting direction is to develop methods for online estimation in fast-moving applications.

The method can also be cast in related industrial settings where event sequences are available such as, in warranty claims processing. From a unit's point of view, these actions can be viewed as sequence of events during its lifetime. In customer service processes such as in the insurance industry, claims are processed over a series of touch-points. The evolution of a claim from initiation to final decision is also an event sequence. Thus, the methods developed in our research can find broad applicability in industry.

7 Supplemental Information

In the online supplemental material, we provide the following: (i) sketch of proof for likelihood, (ii) convergence analysis, (iii) contamination analysis, and (iv) a set of R codes to implement the current method.

References

Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. *Pages 199–213 of: Selected Papers of Hirotugu Akaike*. Springer.

Alaswad, S., & Xiang, Y. (2017) A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, **157**, 54–63.

Arts, J., & Basten, R. (2018) Design of multi-component periodic maintenance programs with single-component models. *IISE Transactions*, **50**(7), 606–615.

Babishin, V., & Taghipour, S. (2016) Optimal maintenance policy for multicomponent systems with periodic and opportunistic inspections and preventive replacements. *Applied Mathematical Modelling*, **40**(23-24), 10480–10505.

Bian, L., & Gebraeel, N. (2014) Stochastic modeling and real-time prognostics for multi-component systems with degradation rate interactions. *IIE Transactions*, **46**(5), 470–482.

Cox, D. R., & Lewis, P. A. W. (1972) Multivariate point processes. *Pages 401–448 of: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3.

Daley, D. J., & Jones, D. V. (2003) *An Introduction to the Theory of Point Processes: Elementary Theory of Point Processes*. Springer.

Doyen, L., & Gaudoin, O. (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety*, **84**(1), 45–56.

Doyen, L., Gaudoin, O., & Syamsundar, A. (2017) On geometric reduction of age or intensity models for imperfect maintenance. *Reliability Engineering & System Safety*, **168**, 40–52.

Doyen, L., Drouilhet, R., & Brenière, L. (2019) A generic framework for generalized virtual age models. *IEEE Transactions on Reliability*.

Gill, J., & King, G. (2004) What to do when your Hessian is not invertible: Alternatives to model respecification in nonlinear estimation. *Sociological methods & Research*, **33**(1), 54–87.

Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., & Sun, Y. (2010) A review on reliability models with covariates. *Pages 385–397 of: Engineering Asset Lifecycle Management*. Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013) *An introduction to statistical learning*. Vol. 112. Springer.

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

Kijima, M. (1989) Some results for repairable systems with general repair. *Journal of Applied Probability*, **26**(1), 89–102.

Lindqvist, B. H., Elvebakk, G., & Heggland, K. (2003) The trend-renewal process for statistical analysis of repairable systems. *Technometrics*, **45**(1), 31–44.

Liu, B., Yeh, R.-H., Xie, M., & Kuo, W. (2017) Maintenance scheduling for multicomponent systems with hidden failures. *IEEE Transactions on Reliability*, **66**(4), 1280–1292.

Meango, T. J.-M., & Ouali, M.-S. (2019) Failure interaction models for multicomponent systems: a comparative study. *SN Applied Sciences*, **1**(1), 66.

Newby, M. (1992) A critical look at some point-process models for repairable systems. *IMA Journal of Management Mathematics*, **4**(4), 375–394.

Nguyen, K.-A., Do, P., & Grall, A. (2014) Condition-based maintenance for multi-component systems using importance measure and predictive information. *International Journal of Systems Science: Operations & Logistics*, **1**(4), 228–245.

Pham, H., & Wang, H. (1996) Imperfect maintenance. *European Journal of Operational Research*, **94**(3), 425–438.

Pijnenburg, M. (1991) Additive hazards models in repairable systems reliability. *Reliability Engineering & System Safety*, **31**(3), 369–390.

Rigdon, S. E., & Basu, A. P. (2000) *Statistical methods for the reliability of repairable systems.* Wiley New York.

Schwarz, G., *et al.* (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), 461–464.

Shafiee, M., & Finkelstein, M. (2015) An optimal age-based group maintenance policy for multi-unit degrading systems. *Reliability Engineering & System Safety*, **134**, 230–238.

Shaked, M., & Shanthikumar, J. G. (1986) Multivariate imperfect repair. *Operations Research*, **34**(3), 437–448.

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

Wang, Y., & Pham, H. (2011) Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas. *IEEE Transactions on Reliability*, **61**(1), 13–22.

Wu, S. (2019) A failure process model with the exponential smoothing of intensity functions. *European Journal of Operational Research*, **275**(2), 502–513.

Xu, Z., Hong, Y., Meeker, W. Q., Osborn, B. E., & Illouz, K. (2017) A multi-level trend-renewal process for modeling systems with recurrence data. *Technometrics*, **59**(2), 225–236.

Yanez, M., Joglar, F., & Modarres, M. (2002) Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliability Engineering & System Safety*, **77**(2), 167–180.

Zhang, N., & Yang, Q. (2015) Optimal maintenance planning for repairable multi-component systems subject to dependent competing risks. *IIE Transactions*, **47**(5), 521–532.

Appendix A Proof of intensity function by recurrence

Owing to the structure of intensity specification, we prove equation (2) by recurrence.

We can immediately see that the expression holds for $N_t = 0$ or $t \le t_1$, $\lambda_k(t) = \lambda_{k0}(t)$. Assuming it also holds for $N_t \le i$ or $t < t_{i+1}$, then we prove for $N_t = i+1$ or $t_{i+1} \le t < t_{i+2}$. We use the relation in equation (1) to obtain

$$\begin{aligned}
\lambda_k(t_{i+1}^{+}) &= (1-\rho_{kk_{i+1}})\lambda_k(t_{i+1}^{-}) \\
&= (1-\rho_{kk_{i+1}})\left(\lambda_{k0}(t_{i+1}) - \sum_{j=0}^{i-1}\left[\rho_{kk_{(i-j)}} \cdot \prod_{r=i-j+1}^{i}\big[(1-\rho_{kk_r})\big]\cdot \lambda_{k0}(t_{i-j})\right]\right) \\
&= \lambda_{k0}(t_{i+1}) - \rho_{kk_{i+1}}\lambda_{0k}(t_{i+1}) - (1-\rho_{kk_{i+1}})\sum_{j=0}^{i-1}\left[\rho_{kk_{(i-j)}} \cdot \prod_{r=i-j+1}^{i}\big[(1-\rho_{kk_r})\big]\cdot \lambda_{k0}(t_{i-j})\right] \\
&= \lambda_{k0}(t_{i+1}) - \rho_{kk_{i+1}}\lambda_{0k}(t_{i+1}) - \sum_{j=0}^{i-1}\left[\rho_{kk_{(i-j)}} \cdot \prod_{r=i-j+1}^{i+1}\big[(1-\rho_{kk_r})\big]\cdot \lambda_{k0}(t_{i-j})\right] \\
&= \lambda_{k0}(t_{i+1}) - \sum_{j=0}^{i}\left[\rho_{kk_{(i+1-j)}} \cdot \prod_{r=i-j+2}^{i+1}\big[(1-\rho_{kk_r})\big]\cdot \lambda_{k0}(t_{i+1-j})\right]
\end{aligned}$$

Since the accumulated change (drop/increase) will remain same in range $t_{i+1} < t < t_{i+2}$, equation (2) remains valid. This concludes the proof. □

Appendix B Score and Hessian of impact parameters

Since the likelihood form uses intensities at the time of event-occurrences, we present another formulation of intensity in equation 2 below. The expression below is easier to develop in a software. Also, for neatness, we drop the subscript m.

$$\lambda_k(t_i|\mathcal{H}_t^-) = \lambda_{k0}(t_i) - A_{k,t_i} \tag{13}$$

where, A_{k,t_i} provides the change in intensity due to past, and is given as:

$$A_{k,t_i} = \rho_{kk_{i-1}}\lambda_{k0}(t_{i-1}) + \sum_{j=1}^{i-2}\rho_{kk_{i-j-1}}\prod_{r=i-j}^{i-1}(1-\rho_{kk_r})\lambda_{k0}(t_{i-j-1}) \tag{14}$$

Score:

$$\frac{\partial l}{\partial \rho_{kk'}} = \sum_{i=1}^{n}\frac{-A'_{k_i,t_i}}{\lambda_{k_i}(t_i)} + \sum_{k=1}^{K}\sum_{i=2}^{n}[A'_{k,t_i}(t_i - t_{i-1})] \tag{15}$$

where, $A_{k,t_1} = 0$ and for $i > 1$:

$$A'_{k,t_i} = \frac{\partial A_{k,t_i}}{\partial \rho_{kk'}} = \lambda_{k0}(t_{i-1})\mathbb{1}(k' = k_{i-1}) + \sum_{j=1}^{i-2} \lambda_{k0}(t_{i-j-1}) \frac{\partial}{\partial \rho_{kk'}} \Big\{ \rho_{kk_{i-j-1}} \prod_{r=i-j}^{i-1} (1 - \rho_{kk_r}) \Big\} \tag{16}$$

First, let us introduce $\mathbf{R}$ as set of all events between $i - j$ and $i - 1$, and set $\mathbf{Q} \subseteq \mathbf{R}$ represents events not equal to k'. Then, the second term in Eq. 16 we will be as following:

$$\mathbb{1}(k' = k_{i-j-1}) \Big[\prod_{r \in \mathbf{R}} (1 - \rho_{kk_r}) \Big] + \rho_{kk_{i-j-1}} \cdot \Big[\prod_{q \in \mathbf{Q}} (1 - \rho_{kk_q}) \Big] \cdot \Big[(-1)(|\mathbf{R}| - |\mathbf{Q}|)(1 - \rho_{kk'})^{|\mathbf{R}| - |\mathbf{Q}| - 1} \Big] \tag{17}$$

Before proceeding, let $\mathcal{Q} = \Big[\prod_{q \in \mathbf{Q}} (1 - \rho_{kk_q}) \Big]$, and $\mathcal{R} - \mathcal{Q} = -(|\mathbf{R}| - |\mathbf{Q}|)(1 - \rho_{kk'})^{|\mathbf{R}| - |\mathbf{Q}| - 1}$ **Hessian:** Here, for a given K, we have total K^2 parameters, therefore, the Hessian will have K^4 terms which with the help of 4 iterators can be written as follows:

$$\frac{\partial^2 l}{\partial \rho_{kk'} \rho_{k''k'''}} = \begin{cases} -\sum_{i=1}^{n} \Big[\frac{(A'_{k_i,t_i})^2}{\lambda_{k_i}^2(t_i)} + \frac{A''_{k_i,t_i}}{\lambda_{k_i}(t_i)} \Big] + \sum_{k=1}^{K} \sum_{i=2}^{n} A''_{k,t_i}(t_i - t_{i-1}), k = k'' \\ 0, k \neq k'' \end{cases} \tag{18}$$

The above equation indicates that if k is not equal to k'', then the Hessian term is zero. *In other words, this implies that impact of an event (say $l \neq k''$) on k has nothing to do with event k''.* This yields that Hessian is a block diagonal matrix containing K blocks, each of size $K \times K$.

We also have $A_{k,t_i} = 0$ for $i = \{1, 2\}$, and for $i > 2, j > 0$:

$$A''_{k,t_i} = \sum_{j=1}^{i-2} \lambda_{k0}(t_{i-j-1}) \frac{\partial^2}{\partial \rho_{kk'} \rho_{kk'''}} \Big\{ \rho_{kk_{i-j-1}} \prod_{r=i-j}^{i-1} (1 - \rho_{kk_r}) \Big\} \tag{19}$$

Below, for clear representation we drop the subscript k, and from equation 17 we make the following cases:

$$\frac{\partial^2}{\partial \rho_{k'} \rho_{k'''}} \Big\{ \rho_{k_{i-j-1}} \prod_{r=i-j}^{i-1} (1 - \rho_{k_r}) \Big\} = \frac{\partial}{\partial \rho_{k'''}} \bigg(\mathbb{1}(k' = k_{i-j-1}) \Big[\prod_{r \in \mathbf{R}} (1 - \rho_{k_r}) \Big] + \rho_{k_{i-j-1}} \cdot \mathcal{Q} \cdot (\mathcal{R} - \mathcal{Q}) \bigg) \tag{20}$$

$$= \text{Case I} + \text{Case II} \tag{21}$$

Again, let us split the set $\mathbf{R}$ in sets $\mathbf{U_R}$ and $\mathbf{V_R}$, where set $\mathbf{U_R}$ represents events of type k''', and remaining events are in set $\mathbf{V_R}$. Similarly, let $\mathbf{U_Q}$ and $\mathbf{V_Q}$ represent splits of set $\mathbf{Q}$.

$$\text{Case I} = \prod_{\mathbf{V_R}}(1-\rho_{k_{v_r}}) \cdot \Big[-|\mathbf{U_R}|(1-\rho_{k_{u_r}})^{|\mathbf{U_R}-1|}\Big] \tag{22}$$

$$+\mathbb{1}(k''' = k_{i-j-1})\Bigg\{\mathcal{Q} \cdot \Big[(\mathcal{R}-\mathcal{Q}) + \rho_{k_{i-j-1}}\Big((|\mathbf{R}|-|\mathbf{Q}|)(|\mathbf{R}|-|\mathbf{Q}|-1)(1-\rho_{k'})^{|\mathbf{R}|-|\mathbf{Q}|-2}\Big)\Big]\Bigg\} \tag{23}$$

$$+\mathbb{1}(k''' \neq k_{i-j-1})\Bigg\{\rho_{k_{i-j-1}} \prod_{\mathbf{V_Q}}(1-\rho_{k_{v_q}}) \cdot \Big(-|\mathbf{U_Q}|(1-\rho_{k_{u_q}})^{|\mathbf{U_Q}-1|}\Big)(\mathcal{R}-\mathcal{Q})\Bigg\} \tag{24}$$

And

$$\text{Case II} = \mathbb{1}(k''' = k_{i-j-1})\Bigg\{(\mathcal{R}-\mathcal{Q}) \cdot \Big[\mathcal{Q} + \rho_{k_{i-j-1}} \prod_{\mathbf{V_Q}}(1-\rho_{k_{v_q}}) \cdot \Big(-|\mathbf{U_Q}|(1-\rho_{k_{u_q}})^{|\mathbf{U_Q}-1|}\Big)\Big]\Bigg\} \tag{25}$$

$$+\mathbb{1}(k''' \neq k_{i-j-1})\Bigg\{\rho_{k_{i-j-1}}\Bigg[\mathbb{1}(k' = k''')\{\mathcal{Q}\}\{(|\mathbf{R}|-|\mathbf{Q}|)(|\mathbf{R}|-|\mathbf{Q}|-1)(1-\rho_{k'})^{|\mathbf{R}|-|\mathbf{Q}|-2}\} \tag{26}$$

$$+\mathbb{1}(k' \neq k''')\{\mathcal{R}-\mathcal{Q}\}\Big\{\prod_{\mathbf{V_Q}}(1-\rho_{k_{v_q}}) \cdot \Big(-|\mathbf{U_Q}|(1-\rho_{k_{u_q}})^{|\mathbf{U_Q}-1|}\Big)\Big\}\Bigg]\Bigg\} \tag{27}$$

Appendix C Numerical study cases

C.1 Setting I

Table 7: True and estimated parameters for the numerical experiments considered under setting I (continued)

$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	Case (Σ)	N	T_{max}	$\hat{\alpha_1}$	$\hat{\beta_1}$	$\hat{\alpha_2}$	$\hat{\beta_2}$	$\hat{\rho_{11}}$	$\hat{\rho_{12}}$	$\hat{\rho_{21}}$	$\hat{\rho_{22}}$
2	1	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	100	10	1.990 (0.278)	1.019 (0.126)	2.006 (0.351)	1.014 (0.104)	0.002 (0.015)	0.000 (0.015)	-0.001 (0.015)	0.000 (0.014)
		$\begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix}$	100	20	2.054 (0.299)	1.017 (0.120)	2.027 (0.285)	1.005 (0.114)	0.412 (0.207)	0.401 (0.199)	0.380 (0.186)	0.392 (0.193)
		$\begin{bmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{bmatrix}$	100	20	2.008 (0.329)	1.013 (0.121)	2.001 (0.247)	1.025 (0.120)	0.826 (0.217)	0.695 (0.253)	0.712 (0.241)	0.843 (0.202)
		$\begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}$	100	20	1.979 (0.241)	1.021 (0.110)	2.049 (0.325)	1.039 (0.126)	0.760 (0.261)	0.232 (0.245)	0.286 (0.270)	0.780 (0.278)
	1.5	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	100	5	2.010 (0.192)	1.550 (0.166)	1.989 (0.187)	1.533 (0.169)	0.005 (0.020)	0.003 (0.017)	0.003 (0.021)	0.003 (0.020)
		$\begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix}$	100	10	2.016 (0.194)	1.506 (0.159)	2.026 (0.202)	1.516 (0.183)	0.409 (0.172)	0.404 (0.151)	0.472 (0.186)	0.494 (0.163)
		$\begin{bmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{bmatrix}$	100	10	2.037 (0.209)	1.527 (0.187)	2.007 (0.173)	1.550 (0.174)	0.870 (0.145)	0.931 (0.092)	0.962 (0.063)	0.991 (0.019)
		$\begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}$	100	10	2.021 (0.184)	1.533 (0.175)	2.050 (0.208)	1.545 (0.180)	0.875 (0.129)	0.153 (0.311)	0.148 (0.369)	0.964 (0.050)

C.2 Setting II

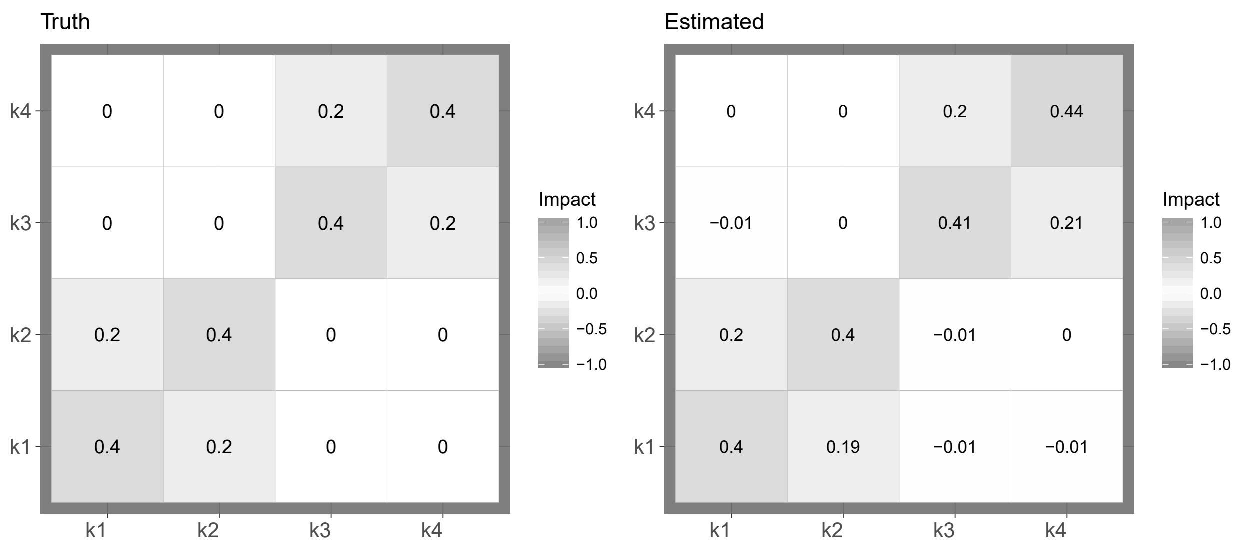

Figure 5: Number of runs: 200

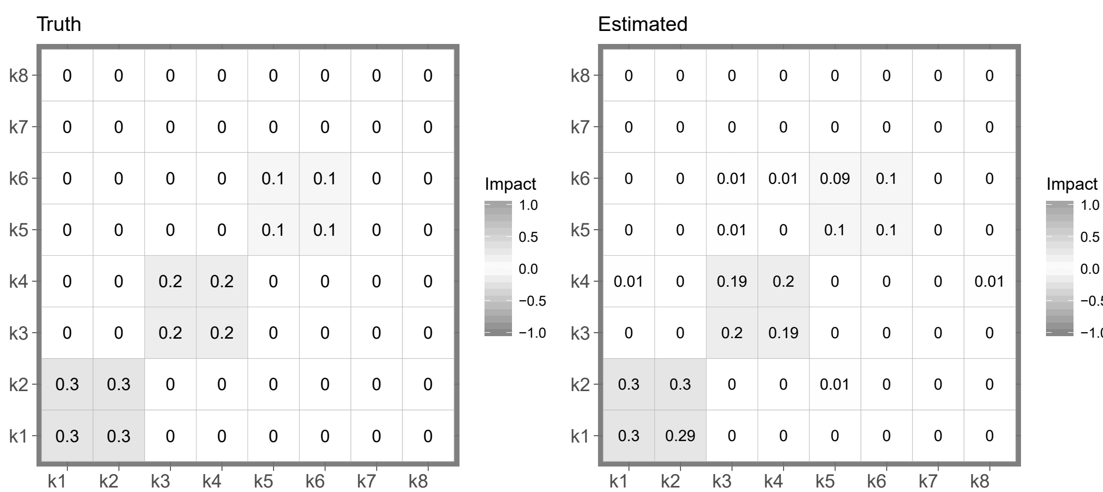

Figure 6: Number of runs: 100

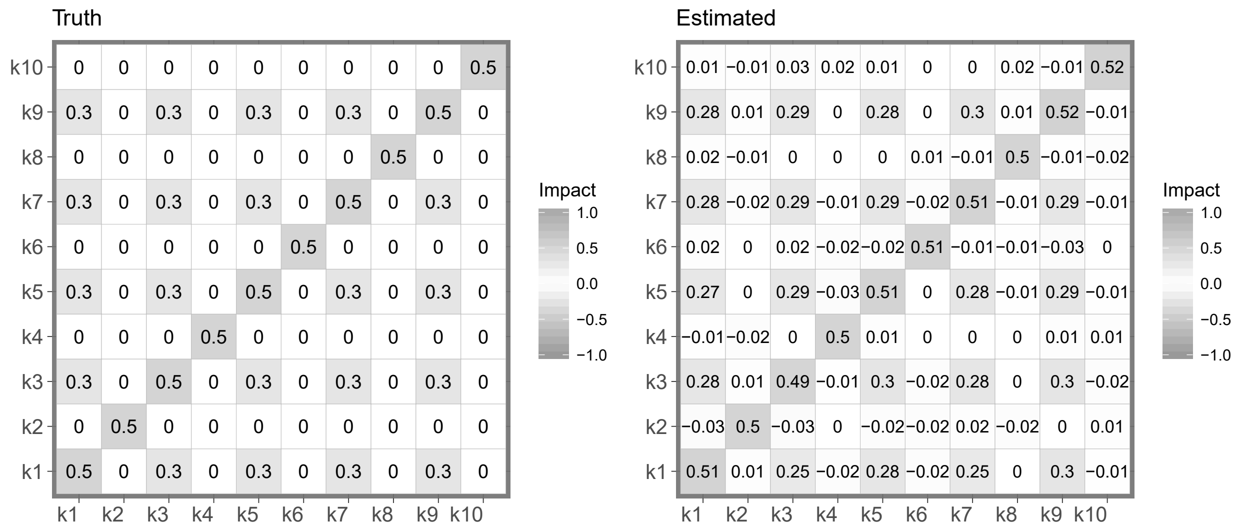

Figure 7: Number of runs: 50

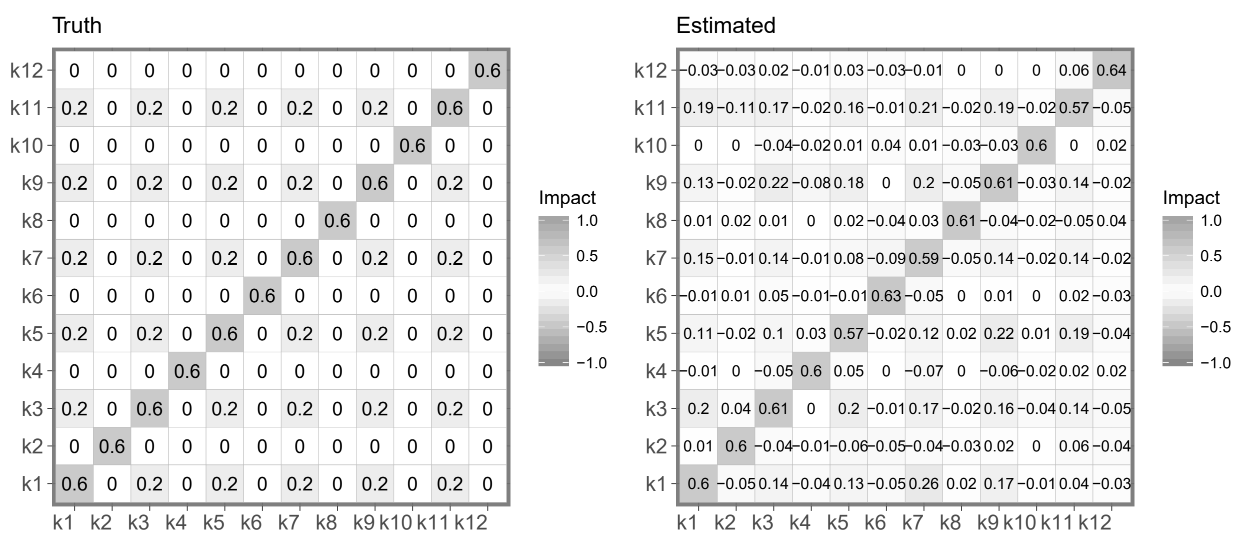

Figure 8: Number of runs: 50