MDPI

*Article*

# Causal Algebras on Chain Event Graphs with Informed Missingness for System Failure

Xuewen Yu [1,*] and Jim Q. Smith [1,2]

1   Statistics Department, University of Warwick, Coventry CV4 7AL, UK; j.q.smith@warwick.ac.uk
2   The Alan Turing Institute, London NW1 2DB, UK
*   Correspondence: xuewen.yu@warwick.ac.uk

**Abstract:** Graph-based causal inference has recently been successfully applied to explore system reliability and to predict failures in order to improve systems. One popular causal analysis following Pearl and Spirtes et al. to study causal relationships embedded in a system is to use a Bayesian network (BN). However, certain causal constructions that are particularly pertinent to the study of reliability are difficult to express fully through a BN. Our recent work demonstrated the flexibility of using a Chain Event Graph (CEG) instead to capture causal reasoning embedded within engineers' reports. We demonstrated that an event tree rather than a BN could provide an alternative framework that could capture most of the causal concepts needed within this domain. In particular, a causal calculus for a specific type of intervention, called a remedial intervention, was devised on this tree-like graph. In this paper, we extend the use of this framework to show that not only remedial maintenance interventions but also interventions associated with routine maintenance can be well-defined using this alternative class of graphical model. We also show that the complexity in making inference about the potential relationships between causes and failures in a missing data situation in the domain of system reliability can be elegantly addressed using this new methodology. Causal modelling using a CEG is illustrated through examples drawn from the study of reliability of an energy distribution network.

**Keywords:** Chain Event Graphs; interventions; causal calculus

## 1. Introduction

The use of Bayesian Networks (BN) for the study of reliability has been widely advocated in the literature [1]. However, the asymmetric processes that are common in system reliability can hardly be fully captured by the framework of a BN.

Fortunately, it has been shown that any discrete BN can be embellished into a tree-based graph called a Chain Event Graph (CEG) [2,3]. The CEG is a graphical model that is a function of an underlying event tree and certain context specific conditional independence statements. In particular, the CEG can model and depict the types of structural asymmetries that the BN framework struggles to embody [4]. This can then provide a framework for studying the causal mechanisms behind the failures of a given system. For example, Cowell and Smith [2] developed a dynamic programming algorithm for maximum a posterior (MAP) structural learning for causal discovery within a restricted class of CEGs called stratified CEGs.

Conventional causal algebras have been adapted from Pearl's *do*-calculus for BNs [5] to the singular manipulation on a CEG, and the *back-door* theorem has been generalised to estimate the effect of this manipulation by previous research [6,7]. In a different strand of research, Barclay, Hutton, and Smith [8] developed a class of CEGs suited for incorporating various missing data structures directly through its topology. Unlike BNs, conjugate inference is still well supported by the structure of CEGs even in the presence of missingness [2].

In Section 2, we adapt the MAP structural learning algorithm [2] to search for the best scoring structure of a CEG when some data is informedly missing. The selected model

provides the best explanation of the observed data that has been informedly censored. By assuming that each candidate CEG is causal in the sense formally defined in [6,7], the best scoring CEG is of a CEG in idle mode, and then causal deductions can be made from it.

In our recent work [9], we demonstrated how to embed the causal reasoning underlying engineering reports for CEGs designed specifically for applications in system reliability. The causal calculus we developed there only provided a framework to study and analyse the impact of *remedial interventions*, i.e., interventions designed to rectify the root cause after a failure had been observed.

In Section 3, we extend the use of the CEG causal framework with missingness to express and analyse a different kind of intervention called a *routine intervention*. This new class of intervention is necessary when we are evaluating the impact of interventions within scheduled maintenance regimes. These regimes are prepared in advance and are used to inspect machines with the objective of preventing future failures that might be about to happen. In this context, although the data may be informedly missing, we can still develop algorithms that, under certain stated hypotheses, produce formulae to give quantitative estimates of the impacts of various candidate routine interventions of this type.

In this paper, we can, therefore, show how we can use the underlying CEG model to predict the effect of various types of such interventions. In particular we report a new backdoor criterion—an analogue of Pearl's back-door criterion for BNs [5]. This gives a quick sufficient condition as to whether the effect of such an intervention is identifiable when data is censored in a way that induces informed missingness. This criterion significantly increases the scope of the original causal calculus using CEGs designed for the singular manipulation [6] and the stochastic manipulation established for BNs [5]. It, thus, enables us to transfer causal technologies so that they apply to this graphical family.

In Section 4, we demonstrate how to interpret the causal structures of a best scoring CEG by a simple example of a conservator system. Furthermore, comparative experiments are designed to show that the proposed new causal algebras can embellish the current structural learning algorithm to capture the causal effects of a routine intervention.

The contributions of this paper are threefold. First, we formally derive a method for selecting a CEG providing the framework of a probability model of maintenance regimes, which acknowledges the presence of informed missingness within the fitted data endemic in these applications. Second, we devise new causal algebras for the routine intervention and prove the identifiability of its causal effects in presence of the types of missing data that we might expect from this application. Third, we demonstrate how important this new intervention calculus can be in making valid inferences and how naive inferences that treat the system as uncontrolled and ignore the underlying causal structure within this application can severely mislead the analyst.

## 2. Causal Identifiability on Chain Event Graphs with Informed Missingness

We begin this section by briefly reviewing and then extending the definition of a CEG [2,3,6–8] before providing a systematic approach to embedding information about the context-specific missingness into a CEG customised for the domain of reliability [9,10].

Suppose we have a vector of variables $X = (X_1, X_2..., X_n)$ taking values in a state space $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$, among which we explore various putative causal hypotheses. An *event tree* $\mathcal{T}(X) = (V_{\mathcal{T}}, E_{\mathcal{T}})$ can be constructed to represent relationships embedded in $X$, where $V_{\mathcal{T}}$ denotes the vertex set and $E_{\mathcal{T}}$ denotes the edge set of $\mathcal{T}(X)$. Each non-leaf node is also called a *situation*. Let $S_{\mathcal{T}}$ denote the set of non-leaf nodes. The *floret* of a situation $v \in S_{\mathcal{T}}$ is a subtree of $\mathcal{T}(X)$, denoted by $\mathcal{F}(v) = (V_{\mathcal{F}(v)}, E_{\mathcal{F}(v)})$. The vertex set of $\mathcal{F}(v)$ consists of $v$ and the vertices in $S_{\mathcal{T}}$ connected from $v$ by a directed edge in $E_{\mathcal{T}}$: $V_{\mathcal{F}(v)} = \{v\} \cup \{v' \in V_{\mathcal{T}} | e_{v,v'} \in E_{\mathcal{T}}\}$. The edge set of $\mathcal{F}(v)$ is a subset of $E_{\mathcal{T}}$ satisfying $E_{\mathcal{F}(v)} = \{e_{v,v'} : v' \in V_{\mathcal{T}}, e_{v,v'} \in E_{\mathcal{T}}\}$.

Let $\mathcal{F}_{\mathcal{T}} = \{\mathcal{F}(v)\}_{v \in S_{\mathcal{T}}}$ denote the collection of all florets on the tree $\mathcal{T}$. Let $\mu(v_0, v)$ denote a subpath from the root node $v_0$ to a node $v \in V_{\mathcal{T}}$ on the event tree. Every floret $\mathcal{F}(v)$ represents a random variable conditional on $\mu(v_0, v)$. We denote this conditional

random variable by $X(v) = X|\mu(v_0, v)$ for $X \in \boldsymbol{X}$. Each emanating edge $e_{v,v'}$ of $v$ is labelled by a value $x \in \mathbb{X}(v)$. Thus, every conditional variable $X_i$, $i \in \{1, ..., n\}$, is represented on a set of florets on the event tree, denoted by $\mathcal{F}(v(X_i))$. Previous research [3,4,6,7] has demonstrated the capability of a tree-like structure to encode the asymmetric information. The corresponding event tree $\mathcal{T}$ associated with this description can be asymmetric and non-stratified [2,4] so that the florets representing the same variable can have different distances from the root node $v_0$.

Figure 1 depicts an event tree for a conservator system. Its variables are $\boldsymbol{X} = (X_{cause}, X_{leak}, X_{alarm}, X_{s/b}, X_{fail})$. The categorical variable $X_{cause}$ represents causes of defects and has three levels {temperature, seal/pipe, and breathing system}; $X_{leak}$ is the oil leak indicator; $X_{alarm}$ is the alarm indicator; $X_{s/b}$ is an indicator of whether there is a sight glass defect or a buchholz defect; and $X_{fail}$ is a failure indicator. This tree is constructed under the assumption that the fault caused by low temperature is irrelevant to the sight glass or buchholz defect, labelled as s/b on the tree. The situations of the tree are annotated as $\{v_0, \cdots, v_{37}\}$, and the leaves are the unlabelled vertices. Since the last variable modelled on this tree is $X_{fail}$, the leaves represent the status of the conservator being failed or operational.
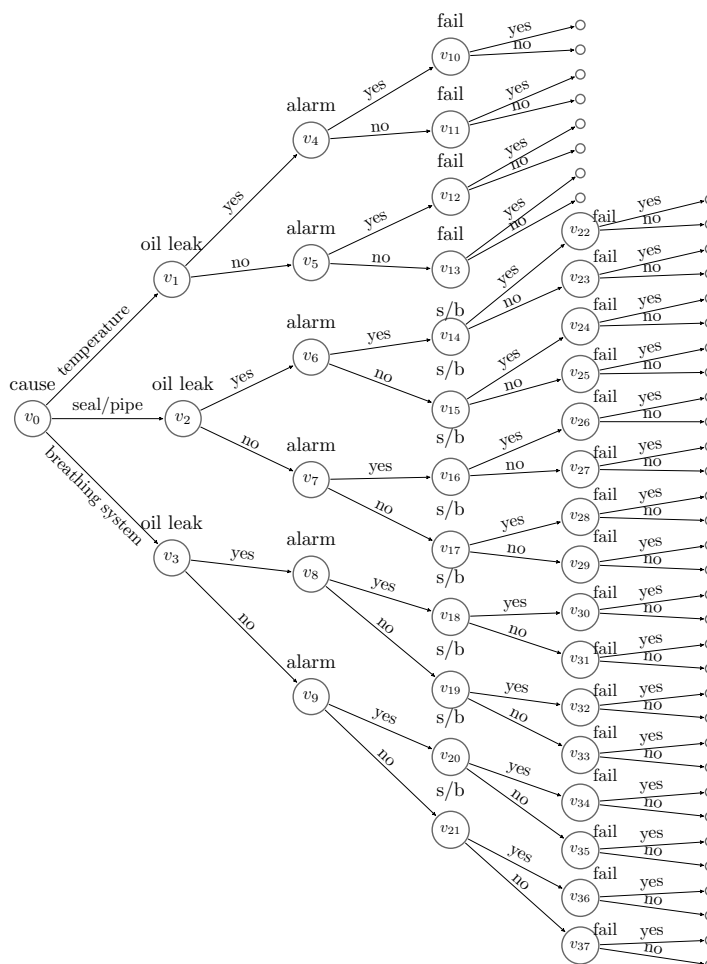


**Figure 1.** An event tree constructed for the conservator system of a transformer.

Let $\Lambda_{\mathcal{T}}$ denote the set of all root-to-leaf paths on the tree and $\lambda(v, v') \in \Lambda_{\mathcal{T}}$ denote the root-to-leaf paths passing through vertices $v, v' \in V_{\mathcal{T}}$. The vector $\boldsymbol{\theta}_v = \mathbb{P}(X(v)) = \mathbb{P}(X|\mu(v_0, v))$ is called the vector of primitive probabilities. Let $\boldsymbol{\theta}_{\mathcal{T}} = (\boldsymbol{\theta}_v)_{v \in V_{\mathcal{T}}}$, which satisfies $\sum_{v' \in ch(v)} \theta_{v,v'} = 1$ and $\theta_{v,v'} \in (0, 1)$ for all $v \in V_{\mathcal{T}}$, where $ch(v) = \{v' \in V_{\mathcal{T}} | e_{v,v'} \in E_{\mathcal{T}}\}$. Then, the pair $(\mathcal{T}(\boldsymbol{X}), \boldsymbol{\theta}_{\mathcal{T}})$ indexes the *probability tree* [2,3] defined over $\boldsymbol{X}$.

The BN is capable of handling the missing data whenever this applies to all values of a pre-assigned set of variables by assigning a missingness indicator to each unobservable variable within that set. It is, therefore, possible to use the BN as a framework for identifying when causal hypotheses are identifiable in this rather restricted setting. The associated analyses use various graphically stated criteria—such as the front-door and the back-door criteria—see e.g., [11–13]. However, unfortunately, the types of missingness that routinely occur in reliability—and, in particular, those associated with the data we collect when performing routine maintenance—are rarely missing across the original random vector associated with the system in this sort of symmetric way. This is because we only learn about those parts of a system that we have chosen to inspect.

In contrast, the probability tree provides a natural and more flexible way to visualise and model the context-specific missingness, where the unobservability of the variable partially depends on which path it lies on the tree. Here, we import the informed missingness into the event tree by defining the *floret-dependent missingness* [14]. Thus, consider a floret $\mathcal{F}(v)$, if the value of the corresponding variable $X(v)$ is unobservable, then we classify this floret into $\mathcal{F}(v) \in \mathcal{F}^M$.

On the other hand, if conditioned on $\mu(v_0, v)$, the value of the variable $X(v)$ is always observed, and then the corresponding floret is classified into $\mathcal{F}(v) \in \mathcal{F}^O$. Accordingly, we have two subsets of florets, $\mathcal{F}^M$ and $\mathcal{F}^O$, representing unobservable florets and fully observed florets, respectively. Then, $\mathcal{F}^M \cap \mathcal{F}^O = \varnothing$ and $\mathcal{F}^M \cup \mathcal{F}^O = \mathcal{F}_\mathcal{T}$. For every unobservable floret $\mathcal{F}(v_j) \in \mathcal{F}^M$, we define a *missing floret indicator* as:

$$B_{\mathcal{F}(v_j)} = \begin{cases} 1 & \text{if } x(v_j) \text{ is missing,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Then, $B_{\mathcal{F}(v_j)}$ represents the conditional missingness and

$$\mathbb{P}(B_{\mathcal{F}(v_j)} = 1) = \mathbb{P}(X(v_j) \text{ missing} | \mu(v_0, v_j)). \tag{2}$$

When $\mathbb{P}(B_{\mathcal{F}(v_j)} = 1) \in (0, 1)$, we construct a floret representing this indicator, denote this by $\mathcal{F}(v(B_{\mathcal{F}(v_j)}))$, and call it a *missing indicator floret*. We then reconstruct an event tree by importing the missing indicator florets on to $\mathcal{T}$. We call this a *missingness event tree* (m-tree). Here, we assume that $B_{\mathcal{F}(v_j)}$ precedes $X(v_j)$, denoted by $B_{\mathcal{F}(v_j)} \prec X(v_j)$. In particular $\mathcal{F}(v_j)$ is appended to the edge emanating from $v(B_{\mathcal{F}(v_j)})$ labelled by $B_{\mathcal{F}(v_j)} = 0$. This artificially introduced ordering has already been shown to be useful for interpreting an event tree constructed with informed missingness [8]. The m-tree then has a new class of florets $\mathcal{F}^{MI} = \mathcal{F}(v(\boldsymbol{B}))$ for $\boldsymbol{B} = \{B_{\mathcal{F}}\}_{\mathcal{F} \in \mathcal{F}^M}$, which is the set of missing indicator florets. The variables associated to the m-tree are expanded to $(\boldsymbol{X}, \boldsymbol{B})$. We denote the topology of the m-tree by $\mathcal{T}(\boldsymbol{X}, \boldsymbol{B})$. An example of the missingness event tree is shown in Figure 2.

Having a missingness event tree, we further elicit a *missingness staged tree* from $\mathcal{T}(\boldsymbol{X}, \boldsymbol{B})$. For two situations $v$ and $w$, if $\mathcal{F}(v)$ and $\mathcal{F}(w)$ represent the same variable, then these two situations are in the same stage whenever $\boldsymbol{\theta}_v = \boldsymbol{\theta}_w$ [3], and the emanating edge $e_{v,v'}$ is labelled the same value of $X$ as $e_{w,w'}$ when $\theta_{v,v'} = \theta_{w,w'}$. Here, we relax the restrictions for a *stratified staged tree* where two situations in the same stage have the same distance from the root node [2,4]. For example, $v_{18}$ can be in the same stage as $v_{38}$ in the missingness event tree in Figure 2, similar example see [8].

Here, we assume that situations along the same root-to-leaf path cannot be in the same stage. This is the *square-free* condition defined by Collazo et al. [3]. Vertices in the same stage are assigned a unique colour, and the edges emanating from the same stage with the same label are assigned the same colour. Such a coloured tree that embeds context-specific conditional independence relations is a missingness staged tree. Let $U = \{u_1, .., u_l\}$ denote the set of stages in the m-tree. Let $u(X_i)$ represent the set of stages associated with variable $X_i$ and $U(\boldsymbol{X}) = \{u(X_1), \cdots, u(X_n)\}$. Let $U(\boldsymbol{B}) = U/U(\boldsymbol{X})$ denote the set of

stages associated to the missing floret indicators. An example of a missingness staged tree of the m-tree in Figure 2 is depicted in Figure 3.

Two situations $v_j, v_k \in u_i \in U$ in the same stage are in the same *position* $w$ if the rooted subtrees $\mathcal{T}_{v_j}(X, B)$ and $\mathcal{T}_{v_k}(X, B)$ are isomorphic. This clustering gives a finer partition of vertices than $U$, denoted by $W = \{w_1, ..., w_m\}$. A *missingness chain event graph* (MCEG) $\mathcal{C}(X, B) = (V_{\mathcal{C}}, E_{\mathcal{C}})$ can be constructed from a missingness staged tree as follows. A sink node $w_\infty$ is created by merging all the leaves of $\mathcal{T}(X, B)$. Then, the vertex set is $V_{\mathcal{C}} = W \bigcup w_\infty$.

For any two $w, w' \in V_{\mathcal{C}}$, we create an edge for every $v \in w$ and the child node $v' \in ch(v) \in V_{\mathcal{T}}$, which belongs to the position $w'$, where the annotating edge probability is the same as that of $e_{v,v'} \in E_{\mathcal{T}}$ and is inherited from the original tree. The colours of the vertices and edges of the MCEG are the same as the corresponding stages and edges in the missingness staged tree [15].

Note that the events on the event tree are chronologically ordered. By definition, a cause comes before its effects. We can be reasonably confident in providing $X$ with a plausible order. For example, the trajectory of the events that lead to a machine's failure always starts with a cause, followed by symptoms, and terminates with a failure. Therefore, we can construct event trees for analysing system failures following this order. In this case, having failed or not is always modelled on the leaves of the tree. Examples are shown in Figures 1 and 2.

It follows that, for this special application of CEGs in system reliability, it is convenient to adapt the semantics and to replace the sink node $w_\infty$ defined above by $w_\infty^f$ and $w_\infty^n$. In this way, $w_\infty^f$ is the receiving node of the edges labelled by a failure, while $w_\infty^n$ is the receiving node of the edges labelled by an operational condition.

Thus, we can classify the root-to-sink paths into two categories: failure paths and deteriorating paths. The former terminate in $w_\infty^f$, while the latter terminate in $w_\infty^n$. Figure 4 gives an example of such a MCEG derived from Figure 3.

It is possible to perform conjugate inference on an idle MCEG even when the data is informed censored [8,16]. This enables us to greatly speed up the search for good explanatory models. The simplest prior to set up in this context assumes each stage vector $\theta_u = (\theta_{u_1}, ..., \theta_{u_l})$ is independently Dirichlet with parameters $(\alpha_{u1}, ..., \alpha_{um_u})$ [3,8]. This is identical to the case when there are no missingness indicators:

$$f(\theta | \mathcal{C}(X, B)) = \prod_{u \in U} \frac{\Gamma(\sum_{j=1}^{m_u} \alpha_{uj})}{\prod_{j=1}^{m_u} \Gamma(\alpha_{uj})} \prod_{j=1}^{m_u} \theta_{uj}^{\alpha_{uj}}. \tag{3}$$

Let $\alpha_u = \sum_{j=1}^{m_u} \alpha_{uj}$ so that, in particular, the equivalent sample size is $\alpha = \sum_{u \in U} \sum_{j=1}^{m_u} \alpha_{uj}$.

Then, given a set of observations $D$, the posterior can be computed in a closed form due to Dirichlet-multinomial conjugacy. Thus,

$$f(\theta | D, \mathcal{C}(X, B)) = \prod_{u \in U} f(\theta_u | D, \mathcal{C}(X, B))$$
$$= \prod_{u \in U} \frac{\Gamma(\sum_{j=1}^{m_u} \alpha_{uj+})}{\prod_{j=1}^{m_u} \Gamma(\alpha_{uj+})} \prod_{j=1}^{m_u} \theta_{uj}^{\alpha_{uj+}}, \tag{4}$$

where $\alpha_{uj+} = \alpha_{uj} + n_{uj}$, and $\alpha_{u+} = \alpha_u + n_u$ is the updated parameter vector.

The log-likelihood score for a MCEG $\mathcal{C}(X, B)$ can be decomposed into local scores associated with the variables $X$ and the missingness indicators $B$.

$$\log Q(\mathcal{C}(X, B)) = \log f_{\mathcal{C}(X,B)}(D) = \sum_{i=1}^{n} \log Q_{u(X_i)}(\mathcal{C}(X, B)) + \sum_{u \in U(B)} \log Q_u(\mathcal{C}(X, B)) \tag{5}$$

We can explicitly compute the log-likelihood in a closed form:

$$\log Q(\mathcal{C}(\boldsymbol{X}, \boldsymbol{B})) = \sum_{u \in U} \left( \log \Gamma(\alpha_u) - \log \Gamma(\alpha_{u+}) - \sum_{j=1}^{m_u} (\log \Gamma(\alpha_{uj}) - \log \Gamma(\alpha_{uj+})) \right). \quad (6)$$

To elicit a best scoring CEG from an event tree, it is necessary to search over all possible orderings over the variables modelled by the tree when the total order over the variables is unknown. The event tree is defined to be built with respect to $\boldsymbol{X}$, and the associated missingness event tree is built as a function of $\mathcal{T}(\boldsymbol{X})$ with appropriate hypotheses of missingness. Therefore, even when the dataset has missing values, we still only search over permutations over $\boldsymbol{X}$ to find an appropriate ordering that best explains the observed process.

Let $\Pi$ denote an ordering of $\boldsymbol{X}$. This could be a set of partial orderings. All variables represented on the m-tree can automatically be ordered given $\Pi$. We denote the m-tree with a specified ordering $\Pi$ by $\mathcal{T}(\boldsymbol{X}(\Pi), \boldsymbol{B}) = (V_\mathcal{T}, E_\mathcal{T})$.

It is non-trivial to identify causal structure from a finite observational dataset. However, the idle model first needs to be estimated before any causal relations can be explored. Many advances have been made in casting the causal discovery as a Bayesian model selection problem [2,17,18]. The MAP structural learning algorithm is a popular and well-developed tool for selecting a best topology of CEGs that best explains the data.

Under the hypothesis that there are no unobserved confounders [2], we render the best scoring structure selected by the MAP algorithm causal and assume it is the model of the idle system when there is no intervention imported. This enables us to further perform causal analysis. Given such a causal graph, we can derive causal hypotheses from the structure and estimate causal effects under different hypothesised underlying causal mechanisms.

Sometimes there is only a putative partial order rather than total order on the variables $\boldsymbol{X}$ whose causal relationship needs to be explored. However, in this setting we can still perform the search over candidate CEGs for the best fitting model, providing that the missing variables only extend to later nodes of the tree.

Cowell and Smith [2] and Collazo et al. [3] presented a recursive algorithm to find the best sink variable for every subset of $\boldsymbol{X}$ ordered by increasing size. This algorithm can be simply adapted for the tree built for the informedly missing data. Let $\boldsymbol{X}_j \subseteq \boldsymbol{X}$ denote the subset of variables whose ordering is needed to be learned and $\Pi_{\boldsymbol{X}_j}$ denote the best partial ordering over $\boldsymbol{X}_j$.

Then, through applying the algorithm designed by [2,3] on every $\boldsymbol{X}_j$, we can find the best ordering over the variables defined on the tree, where $\Pi = \{\Pi_{\boldsymbol{X}_j}\}_j$. Here, we search over subspaces $\mathbb{X}_{j_1} \times \cdots \times \mathbb{X}_{j_k}$ for $X_{j_1}, ..., X_{j_k} \in \boldsymbol{X}_j$ and compute the local scores with respect to the corresponding $\boldsymbol{Y}$. In particular, for every $\boldsymbol{X}_j^{(l)} = \{X_{j_1}, ..., X_{j_l}\}$, where $l \in \{1, ..., k-1\}$, we find a best sink variable $X' \in \boldsymbol{X}_j^{(l)}$ for every $\boldsymbol{X}_j^{(k-1)} \subseteq \boldsymbol{X}_j^{(k)}$ that has been ordered appropriately. The best sink variable $X'$ is found by computing the local score of the best subtree spanned by $\boldsymbol{X}_j^{(k-1)} \bigcup \{X_s\}$ for every $X_s \in \boldsymbol{X}_j^{(k)}$ together with the corresponding missingness indicators.

The MAP score can be evaluated directly from the local scores that have been computed because the total score is the sum of local scores as shown in Equation (5). Two MCEGs $\mathcal{C}_1$ and $\mathcal{C}_2$ with respect to the same data set can be compared by the log-posterior Bayes factor. Suppose both trees have Dirichlet priors whose hyperparameters are $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. The Bayes factor, then, has a closed form [3]:

$$lpBF(\mathcal{C}_1, \mathcal{C}_2) = \log q(\mathcal{C}_1) - \log q(\mathcal{C}_2) + \log Q(\mathcal{C}_1) - \log Q(\mathcal{C}_2), \quad (7)$$

where $\log q(\mathcal{C}_i)$ denotes the log prior. Different priors over models can be chosen given expert judgement on different missingness mechanisms and conditional dependencies.

When using a uniform prior, $\log q(\mathcal{C}_1) - \log q(\mathcal{C}_2) = \frac{1}{N_C} - \frac{1}{N_C} = 0$, where $N_C$ denotes the total number of models.
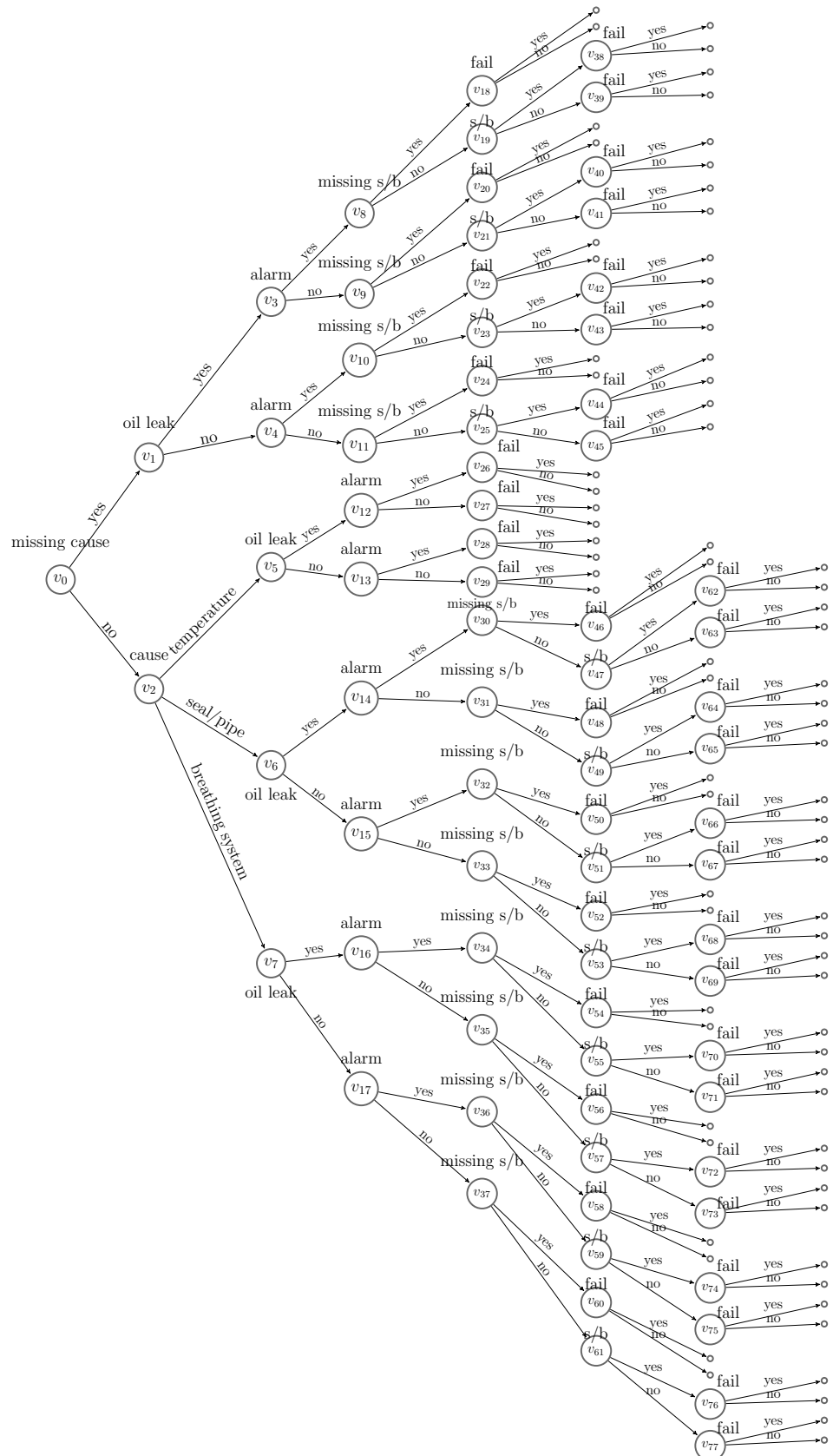


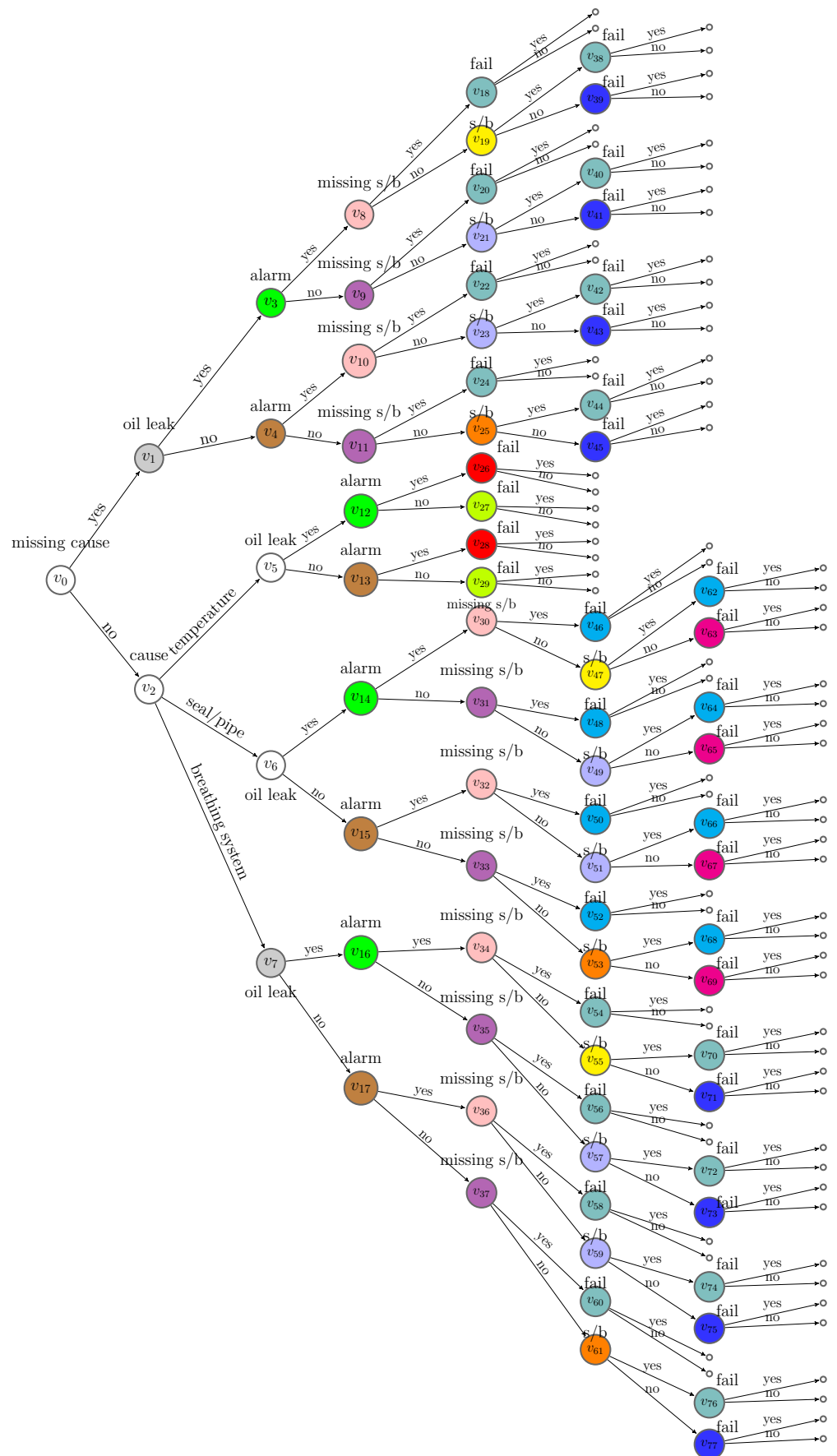**Figure 2.** A missingness event tree constructed from Figure 1.

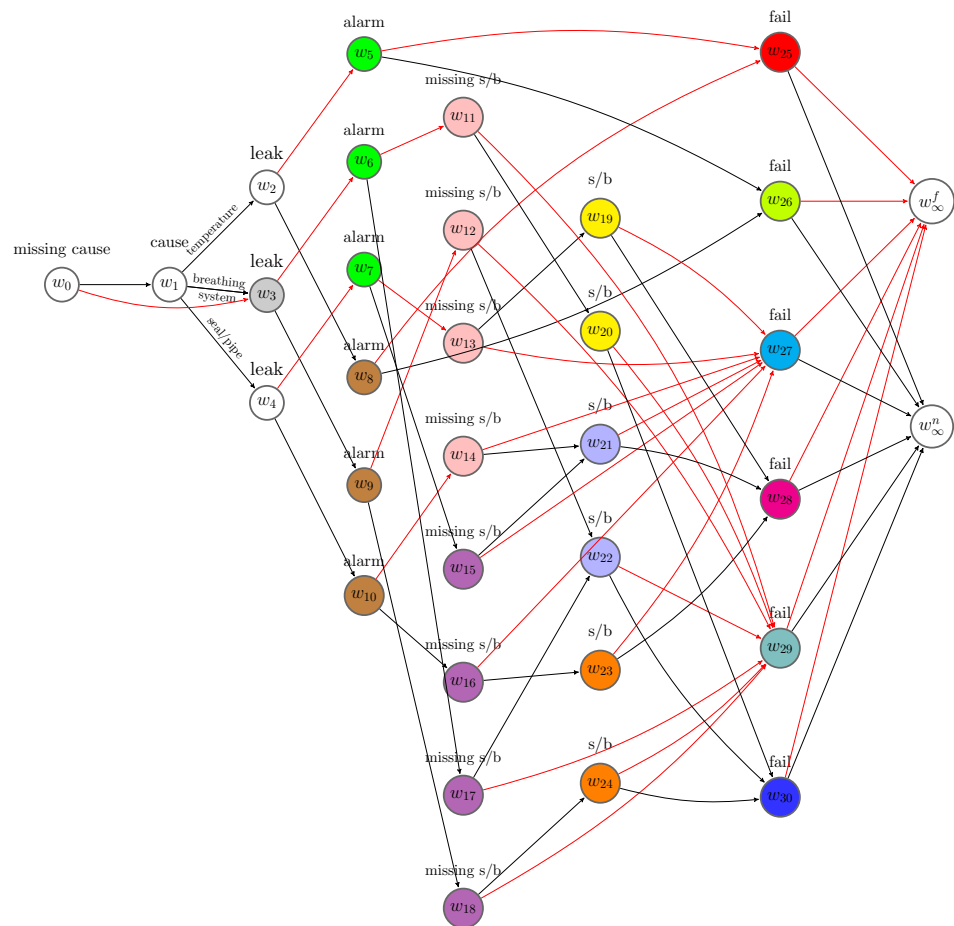**Figure 3.** A missingness staged tree of the m-tree in Figure 2.

**Figure 4.** A MCEG derived from Figure 3. For simplicity, the edges labelled "no" are coloured in red.

## 3. Causal Algebras for Routine Maintenance

By assuming the best scoring CEG causal and treating it as the idle system, one can always design experiments to collect data under the influence of interventions, and thus we can estimate the causal effects from the partially observed system. By controlling certain events on the tree, the semantics of a causal CEG allow us to explore its effect on the events that lie downstream of the controlled events along the root-to-sink paths. For a reliability analysis, it is extremely useful to trace and discover the potential causes of abnormal conditions or failures. By designing causal algebras for different interventions, we can make predictive inferences about the effects of a variety of types of maintenance and thus improve the prediction of system failures.

Having defined the remedial intervention on the CEG for the reliability system in [9], here, we investigate a new class of intervention regime. In the reliability literature, there are two main categories of maintenance: *corrective maintenance* (CM) and *preventive maintenance* (PM) [19]. CM takes place after a failure, while PM often refers to a scheduled maintenance that helps to identify and prevent problems during inspections before a failure occurs [20]. In this section, we carefully customise causal algebras for the intervention in light of the latter case, calling this a *routine intervention*. A routine intervention not only has an impact on the lifetime of the maintained equipment but also affects the likelihood of different defects that may occur in the equipment.

### 3.1. Effects on Lifetime

In the context of reliability, the interventions largely consist of replacing failed components of the system. This type of intervention—unusual in most causal analyses—requires special attention, especially as there are some very well-known effects of such interventions that need to be incorporated before it is possible to realistically model the effects of interventions. In particular, when describing the failure of equipment, the bathtub effect [20] is widely applicable. This divides the lifetime of an equipment into three periods: the early life of a new component has a decreasing failure rate; this is followed by a period with a constant failure rate; the failure rate rises during the wear-out period [21]. A Weibull distribution whose density is given by

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} \tag{8}$$

is often used by reliability engineers to model this varying hazard [20], where the scale parameter is $\eta > 0$, and the shape parameter is $\beta > 0$. The survival function takes the form:

$$1 - F(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}. \tag{9}$$

Let $\Lambda_{\mathcal{C}}$ denote the set of all root-to-sink paths on the MCEG $\mathcal{C}(X(\Pi), B)$. Then, the lifetime of the repaired equipment can be modelled on the associated root-to-sink paths, denoted by $\tilde{\Lambda} \subseteq \Lambda_{\mathcal{C}}$. For $\lambda \in \tilde{\Lambda}$, let $T(\lambda)$ represent the total lifetime of the equipment when the failure trajectory is modelled on the path $\lambda$.

For a repairable system, the PM prolongs the life of the component [22–24]. By adopting the Arithmetic Reduction of Age (ARA) model, which assumes the life of the equipment is shortened up to proportionality [23], we now establish methods to evaluate the effect of the scheduled PM on the equipment's lifetime.

Let $Z_s^{\lambda}$ represent the failure time of an equipment with observed age $s$ given a failure process that is modelled on the path $\lambda$. Then, the survival function is

$$\mathbb{P}(Z_s^{\lambda} > t) = \frac{1 - F_{\lambda}(s+t)}{1 - F_{\lambda}(s)} = e^{-\left(\frac{s+t}{\eta_{\lambda}}\right)^{\beta_{\lambda}} + \left(\frac{s}{\eta_{\lambda}}\right)^{\beta_{\lambda}}}, \tag{10}$$

where $F_{\lambda}(\cdot)$ denotes the reliability distribution for failure trajectory $\lambda$.

In an idle system, for $\lambda \in \tilde{\Lambda}$, $T(\lambda)$ has the same distribution as $Z_0^{\lambda}$, i.e., $T(\lambda) \overset{d}{=} Z_0^{\lambda}$. Thus,

$$\mathbb{P}(T(\lambda) > t) = \mathbb{P}(Z_0^{\lambda} > t) = 1 - F_{\lambda}(t). \tag{11}$$

Preventive maintenance can be scheduled periodically. However, for simplicity, we only demonstrate the effect of a single time routine maintenance in this paper. We suppose that an equipment is diagnosed during a routine maintenance and is repaired at age $\tau$. Kijima [24] and Guessoum and Aupiedy [23] introduced a parameter representing the degree of repair, denoted by $A \in [0, 1]$. When $A = 0$, the repair is *perfect* and restores the maintained part to *as good as new* (AGAN). On the other hand, $A = 1$ corresponds to a *minimal repair*, after which the maintained part is functioning as it was just prior to the repair.

Since the repaired equipment is rejuvenated, the virtual age [23,24] after maintenance is then $A\tau$. Let $T^*(\lambda)$ denote the post-intervention time to failure. Then, after a routine intervention, the residual lifetime of the maintained equipment has the same distribution as $Z_{A\tau}^{\lambda}$. Therefore,

$$\mathbb{P}(T^*(\lambda) > t) = \mathbb{P}(Z_{A\tau}^{\lambda} > t) = \frac{1 - F_{\lambda}(t + A\tau)}{1 - F_{\lambda}(A\tau)}. \tag{12}$$

### 3.2. Manipulations on the MCEG

If $X_i \in \boldsymbol{X}$ takes value $x_{ij}$, let $e(x_{ij}) \in E_{\mathcal{C}}$ denote the edges labelled by this value that emanate from $w(X_i)$. The set of vertices receiving $e(x_{ij})$ are represented by $w(x_{ij})$. The path related probability, denoted by $\pi(\lambda)$ for $\lambda \in \Lambda_{\mathcal{C}}$, can then be factorised as:

$$\pi(\lambda) = \prod_{e \in E_\lambda} \theta_e, \tag{13}$$

where $E_\lambda$ represents a collection of edges lying along the path $\lambda$.

When there is a routine intervention, we are only interested in the process portrayed by the deteriorating paths. We denote this set of paths by $\Lambda_{x_{fail,0}} = \Lambda(e(x_{fail,0}))$, where $x_{fail,0}$ represents $X_{fail} = 0$. Whatever this preventive action is, an analogue of the *do*-operation $do(X_{fail} = 0)$ is imported into the idle MCEG. Thus, we force $e(x_{fail,0})$ to have probability 1 and $e(x_{fail,1})$ to have probability 0, or, equivalently, we manipulate $\Lambda_{x_{fail,0}}$. Therefore, we always have the post-intervened path probability:

$$\hat{\pi}^{\Lambda_{x_{fail,0}}}(\lambda) = \begin{cases} \dfrac{\prod_{e \in E_\lambda} \theta_e}{\theta_{e(x_{fail,0})}} & \text{if } \lambda \in \Lambda_{x_{fail,0}}; \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

This is a singular manipulation on the MCEG and yields a manipulated MCEG with respect to $\Lambda_{x_{fail,0}}$. We denote this by $\hat{\mathcal{C}}^{\Lambda_{x_{fail,0}}}$.

Depending on the preventive action taken, other manipulations can also be imported into the MCEG in addition to the singular manipulation on $\Lambda_{x_{fail,0}}$. We next demonstrate two scenarios of composite manipulations.

#### 3.2.1. Composite Singular Manipulations under Routine Intervention

In this section, we discuss the situation where the preventive maintenance perfectly repaired a problem, and, as a consequence of this repair, an event $x_r$ is forced to occur. The event $x_r$ is labelled on a set of edges $e(x_r)$ whose receiving nodes are $w(x_r)$ and emanating nodes are $pa(w(x_r))$. In this case, the unit will be forced to pass through every edge $e \in e(x_r)$ with probability 1. We, therefore, have a composition of singular manipulations, and the manipulated events are $\boldsymbol{x} = \{x_{fail,0}, x_r\}$. On an MCEG, the controlled event is represented by

$$\Lambda_{\boldsymbol{x}} = \Lambda(e(\boldsymbol{x})) = \Lambda(e(x_{fail,0})) \cap \Lambda(e(x_r)). \tag{15}$$

Let $\mathcal{F}(e(\boldsymbol{x}))$ denote the set of florets that the edges $e(\boldsymbol{x})$ lie in.

If we are interested in the effect of the routine maintenance on event $y$, then, on the MCEG, we represent it by $\Lambda_y = \Lambda(e(y))$. The set of florets that $e(y)$ lies in is denoted by $\mathcal{F}(e(y))$.

Given a CEG, let $\pi(\Lambda_y || \Lambda_{\boldsymbol{x}})$ denote the probability of observing event $y$ given a manipulation that forces the events $\boldsymbol{x}$ to occur. We aim to estimate this probability from the observed data and to demonstrate that the effects of a routine intervention are identifiable. We have shown in [9] that causal effects from a singular manipulation are estimable, also called *recoverable*, by adapting the back-door theorem [5]. Here, we simply extend our previous results [9] so that it now also applies to the types of composite manipulations that we discuss here.

The MCEG provides flexible choices of events $\boldsymbol{z}$ to be the back-door partition so that $\Lambda_{\boldsymbol{z}}$ partitions $\Lambda_{\mathcal{C}}$ [6,7]. We first impose a constraint on $\boldsymbol{z}$ that $\mathcal{F}(e(\boldsymbol{z})) \nsubseteq \mathcal{F}^{MI}$, i.e., that cannot be a missingness indicator. This is to ensure that $\pi(\Lambda_y || \Lambda_{\boldsymbol{x}})$ can be estimated from the partially observed data [9]. Note that any of $\mathcal{F}(e(\boldsymbol{x})), \mathcal{F}(e(y)), \mathcal{F}(e(\boldsymbol{z}))$ might be unobservable. Let

$$\mathcal{F}_{\boldsymbol{x} \cup y \cup \boldsymbol{z}} = \{\mathcal{F} : \mathcal{F} \in \mathcal{F}(e(\boldsymbol{x})) \cup \mathcal{F}(e(y)) \cup \mathcal{F}(e(\boldsymbol{z})) \text{ and } \mathcal{F} \notin \mathcal{F}^{MI}\}. \tag{16}$$

We define the *manifest paths* to be the largest set of root-to-sink paths on the MCEG passing along edges labelled by $x$, $y$ and $z$. We let $\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{x})),0} = \{b_{\mathcal{F},0}\}_{\mathcal{F} \in \mathcal{F}(e(\boldsymbol{x}))}$ denote the set of missingness indicators of florets $\mathcal{F}(e(\boldsymbol{x}))$ taking value 0, i.e., values of the corresponding floret variables are observed. Then, the manifest paths are

$$\Lambda(w(\boldsymbol{b}_{\mathcal{F}_{x \cup y \cup z},0})) = \Lambda(w(\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{x})),0})) \cap \Lambda(w(\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{y})),0})) \cap \Lambda(w(\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{z})),0})). \tag{17}$$

We can construct a sub-MCEG $\mathcal{C}^M$ using the manifest paths. Let the collection of the root-to-sink paths of this subgraph be $\Lambda_{\mathcal{C}^M} = \Lambda(w(\boldsymbol{b}_{\mathcal{F}_{x \cup y \cup z},0}))$. We call this sub-MCEG the *manifest MCEG*. This construction ensures that there is no edge in the manifest MCEG associated with a controlled event, effect, or partition event being missing.

We next reconstruct $\pi(\Lambda_y || \Lambda_x)$ from the manifest MCEG. Let $\pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_y || \Lambda_x)$ denote the probability of observing an event $y$ given a manipulation forcing $x$ to happen within the manifest MCEG. Note that the manipulated MCEG is a subgraph of the manifest MCEG. For a singular manipulation on $\Lambda_x$, the manipulated paths on the manifest MCEG are

$$\Lambda_* = \Lambda(w(\boldsymbol{b}_{\mathcal{F}_{x \cup y \cup z},0})) \cap \Lambda_x. \tag{18}$$

The manipulated MCEG with respect to $\Lambda_*$ is then denoted by $\hat{\mathcal{C}}^{\Lambda_*}$ and satisfies $\Lambda_{\hat{\mathcal{C}}^{\Lambda_*}} = \Lambda_*$.

**Theorem 1** (The m-back-door criterion for composite singular manipulations). *When a dataset has missing values, the effect of a singular manipulation on $\boldsymbol{x}$ on $y$ is identifiable on the MCEG if we can find a partition $\Lambda_z$ of $\Lambda_{\mathcal{C}^M}$ such that*

$$\pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_y || \Lambda_x) = \sum_z \pi(\Lambda_y | \Lambda_x, \Lambda_z, \Lambda(w(\boldsymbol{b}_{\mathcal{F}_{x \cup y \cup z},0}))) \pi(\Lambda_z | \Lambda(w(\boldsymbol{b}_{\mathcal{F}_{x \cup y \cup z},0}))). \tag{19}$$

For the proof of this theorem, see [9,14].

**Example 1.** *Given the causal MCEG in Figure 4 of a conservator system, we demonstrate how the formulae defined above works for a specific routine maintenance that successfully prevents an oil leak. This is equivalent to importing a combination of $do(X_{fail} = 0)$ and $do(X_{leak} = 0)$ operations to the idle MCEG. The controlled events are $\boldsymbol{x} = \{x_{fail,0}, x_{leak,0}\}$. From Figure 4, we next identify the associated root-to-sink paths. In particular,*

$$\Lambda_{x_{fail,0}} = \bigcup_{w \in \{w_{25}, \cdots, w_{30}\}} \Lambda(e_{w, w_\infty^n}), \tag{20}$$

$$\Lambda_{x_{leak,0}} = \Lambda(e_{w_2, w_8}) \cup \Lambda(e_{w_3, w_9}) \cup \Lambda(e_{w_4, w_{10}}), \tag{21}$$

*and $\Lambda_{x_{fail,0}, x_{leak,0}} = \Lambda_{x_{fail,0}} \cap \Lambda_{x_{leak,0}}$.*

*To next focus on alarm, the effect event is $x_{alarm,1}$. The associated set of paths is $\Lambda_{x_{alarm,1}} = \Lambda(e_{w_5, w_{25}}) \cup \Lambda(e_{w_6, w_{11}}) \cup \Lambda(e_{w_7, w_{13}}) \cup \Lambda(e_{w_8, w_{25}}) \cup \Lambda(e_{w_9, w_{12}}) \cup \Lambda(e_{w_{10}, w_{14}})$. The causal query with respect to $\boldsymbol{x}$ is identifiable whenever $\pi(\Lambda_{x_{alarm,1}} || \Lambda_{x_{fail,0}, x_{leak,0}})$ can be recovered from the MCEG by estimating it from the dataset with missing entries. There are a variety of possible choices for the partition events $\boldsymbol{z}$. Here, we simply let $\boldsymbol{z}$ be $X_{cause}$ whose corresponding positions lie upstream of the controlled events $x_{leak,0}$ on the tree. The corresponding floret is, then, $\mathcal{F}(e(\boldsymbol{z})) = \mathcal{F}(w_1)$.*

*We now construct the manifest MCEG and the manipulated MCEG in order to identify the effects of the intervention. Notice that the controlled events and the effect events are always observable in our example. Thus,*

$$\Lambda(w(\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{x})),0})) = (\bigcup_{w \in \{w_{25}, \cdots, w_{30}\}} \Lambda(w)) \cap (\bigcup_{w \in \{w_2, w_3, w_4\}} \Lambda(w)) = \Lambda_{\mathcal{C}}, \tag{22}$$

$$\Lambda(w(\boldsymbol{b}_{\mathcal{F}(e(\boldsymbol{y})),0})) = \bigcup_{w \in \{w_5, \cdots, w_{10}\}} = \Lambda_{\mathcal{C}}, \tag{23}$$

*However, the back-door partition events might be missing. The collection of paths along which **z** are observed is*

$$\Lambda(w(b_{\mathcal{F}(e(\mathbf{z})),0})) = \Lambda(w_1).\tag{24}$$

*Following Equation (17), the manifest paths are $\Lambda(w(\mathbf{b}_{\mathcal{F}_{\mathbf{x}\cup y\cup \mathbf{z},0}})) = \Lambda_{\mathcal{C}} \cap \Lambda(w_1) = \Lambda(w_1)$. Thus, to investigate this, we construct the manifest MCEG with respect to $\Lambda(w_1)$. This is a subgraph of the idle MCEG in Figure 4 obtained by simply removing the edge $e_{w_0,w_3}$, which represents the causes that are missing. We further elicit the manipulated MCEG from the manifest MCEG. By the definition of the manipulated paths given in Equation (18), we select the manipulated paths from the manifest paths: $\Lambda_* = \Lambda(w_1) \cap \Lambda_{x_{fail,0},x_{leak,0}}$. Since the intervention forces $x_{fail,0}$ and $x_{leak,0}$ to happen, the events $x_{fail,1}$ and $x_{leak,1}$ should never be observed. Thus, the probability of a manipulated path passing along the edges $e(x_{fail,1})$ and $e(x_{leak,1})$ is 0.*

*Equivalently, the positions $w(x_{fail,1}) = w_\infty^f$ and $w(e(x_{leak,1})) = \{w_5, w_6, w_7\}$ should never be passed through by any path in the manipulated graph. Then, by removing the nodes and edges that are not traversed by the manipulated paths in the manifest MCEG, we can derive the manipulated MCEG with respect to $\Lambda_*$, see Figure 5. We can then estimate the causal effects on alarm using the formula given in the m-back-door theorem defined above. The conditional path probabilities in Equation (19) can simply be evaluated using the factorisation of the corresponding primitive probabilities in the manipulated MCEG.*
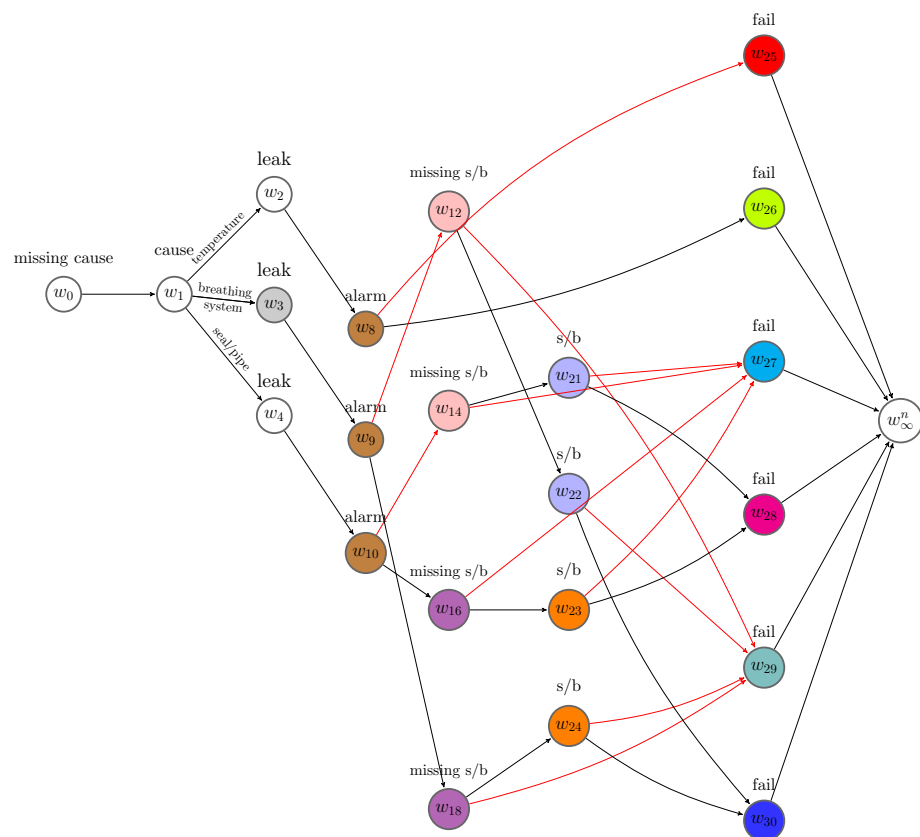


**Figure 5.** The manipulated MCEG when controlling $x_{leak,0}$ and $x_{fail,0}$

### 3.2.2. Composite Singular and Stochastic Manipulations under Routine Intervention

During routine inspections, the field engineers may clean the components, check the oil level and leakage, replace some units, and so on [25]. Since there are different types of repair and because the degree of this repair varies, the manipulations enacted by the routine intervention could be more complicated than forcing a specific event to happen. In

fact, repairing or replacing an equipment could affect multiple units or multiple defects of a unit.

Therefore, depending on the repaired subcomponent and the degree of repair, multiple florets can be influenced separately and simultaneously. Thus, a routine intervention could introduce more uncertainty to the probability distributions over these relevant florets. Therefore, the distributions of some of the primitive probabilities may need to be reassigned. This manipulation is then called a stochastic manipulation on the MCEG.

Unlike a remedial intervention [9], a stochastic manipulation induced by a routine intervention is not restricted to root causes. Consider a floret $\mathcal{F}$ whose distribution is manipulated by a routine intervention. The events represented by this floret could be defects or symptoms of the maintained equipment.

Let $\boldsymbol{x}_r$ denote the controlled events of a routine intervention. Suppose we can find the edges labelled by these events, denoted by $e(\boldsymbol{x}_r)$, then $\mathcal{F}(e(\boldsymbol{x}_r))$ is the set of florets whose distribution are manipulated under the routine intervention. Let $\boldsymbol{w}^* = pa(w(\boldsymbol{x}_r))$ denote the set of emanating nodes of edges $e(\boldsymbol{x}_r)$. We can then conclude that $\mathcal{F}(\boldsymbol{w}^*) = \mathcal{F}(e(\boldsymbol{x}_r))$.

For $w \in \boldsymbol{w}^*$, we update the probability distribution after a routine intervention via the transformation:

$$\hat{q}(\boldsymbol{\theta}_w) = G[q(\boldsymbol{\theta}_w)] \tag{25}$$

where $\hat{q}(\cdot)$ represents the post-intervened distribution. The transformation $G$ preserves the properties of the transition probabilities so that $\sum_{e \in E(w)} \theta_e = 1$ and $\theta_e > 0$.

Motivated by the steady model [26,27], one straightforward option is to map distributions to distributions through non-linear state space models. A possible transformation to increase uncertainty in a distribution is the *power steady transformation* [26,28], which can be characterised by information loss after the intervention takes.

$$\hat{q}(\boldsymbol{\theta}_w) \propto q(\boldsymbol{\theta}_w)^{\phi}, \tag{26}$$

where $\phi \in (0, 1]$. Assume that the value of $\phi$ can be assessed and informed by the domain experts. Then, a power steady evolution assumes that such information loss is linear and proportional to $\phi$ so that:

$$\mathbb{E}[\log \hat{q}(\boldsymbol{\theta}_w)] = \phi \mathbb{E}[\log q(\boldsymbol{\theta}_w)] + c, \tag{27}$$

for some constant $c$.

For a Dirichlet prior $\boldsymbol{\theta}_w \sim Dirichlet(\boldsymbol{\alpha}_w)$ with concentration parameters $\boldsymbol{\alpha}_w = (\alpha_{w1}, \cdots, \alpha_{wm_w})$, following [29], we can transform it to $Dirichlet(\hat{\boldsymbol{\alpha}}_w)$, where $\hat{\boldsymbol{\alpha}}_w = (\hat{\alpha}_{w1}, \cdots, \hat{\alpha}_{wm_w})$ and $\hat{\alpha}_{wj} - 1 = \phi(\alpha_{wj} - 1)$, for $j \in \{1, \cdots, m_w\}$. By this transformation, the mode remains the same. We can consider such manipulations when searching for the best scoring MCEG for causal discovery. This is explained in Section 4.

Having updated the transition probabilities, the path probabilities under the stochastic manipulation given a routine intervention can be re-evaluated. Let $\Lambda(\boldsymbol{w}^*)$ denote the set of root-to-sink paths on the MCEG passing through any position $w \in \boldsymbol{w}^*$. Let $\overline{\Lambda}(\boldsymbol{w}^*) = \Lambda_{\mathcal{C}}/\Lambda(\boldsymbol{w}^*)$. Then, the probabilities of paths in $\Lambda_{x_{fail,0}} \cap \Lambda(\boldsymbol{w}^*)$ are affected by both the singular manipulation on $x_{fail,0}$ and the stochastic manipulation on $\mathcal{F}(\boldsymbol{w}^*)$. The probabilities of paths in $\Lambda_{x_{fail,0}} \cap \overline{\Lambda}(\boldsymbol{w}^*)$ are affected by the singluar manipulation on $x_{fail,0}$. Therefore, the post-intervened path probabilities on the MCEG are:

$$\hat{\pi}(\lambda) = \begin{cases} \frac{\prod_{e \in E_\lambda} \theta_e}{\theta_{e(x_{fail,0})} \prod_{e' \in E(\boldsymbol{w}^*) \cap E_\lambda} \theta_{e'}} \times \prod_{e' \in E(\boldsymbol{w}^*) \cap E_\lambda} \hat{\theta}_{e'} & \text{if } \lambda \in \Lambda_{x_{fail,0}} \cap \Lambda(\boldsymbol{w}^*), \\ \frac{\prod_{e \in E_\lambda} \theta_e}{\theta_{e(x_{f,0})}} & \text{if } \lambda \in \Lambda_{x_{fail,0}} \cap \overline{\Lambda}(\boldsymbol{w}^*), \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

Let $\boldsymbol{x}^*$ denote the set of all events represented on $\mathcal{F}(\boldsymbol{w}^*)$ and let $\boldsymbol{x} = x_{fail,0} \cap \boldsymbol{x}^*$ denote the set of events that are manipulated. Then, the set of florets associated with the manipulated events, the effect event and the partition events is

$$\mathcal{F}_{\boldsymbol{x} \cup y \cup z} = \{\mathcal{F} : \mathcal{F} \in \mathcal{F}(e(x_{fail,0})) \cup \mathcal{F}(\boldsymbol{w}^*) \cup \mathcal{F}(e(y)) \cup \mathcal{F}(e(z)) \text{ and } \mathcal{F} \notin \mathcal{F}^{MI}\}. \quad (29)$$

The manifest paths are defined analogously to Equation (17) so that no event of interest, i.e., $\boldsymbol{x}, y,$ and $z$, is missing in this restricted class of paths.

$$\Lambda(w(\boldsymbol{b}_{\mathcal{F}_{\boldsymbol{x} \cup y \cup z},0})) = \Lambda(w(b_{\mathcal{F}(e(x_{fail,0})),0})) \cap \Lambda(\boldsymbol{w}^*) \cap \Lambda(w(b_{\mathcal{F}(e(y)),0})) \cap \Lambda(w(b_{\mathcal{F}(e(z)),0})). \quad (30)$$

We next show the identifiability of the effects by adapting the back-door criterion for stochastic manipulation [9]. More specifically, this is possible whenever we need to identify a $\Lambda_z$ that partitions the root-to-sink paths of the manifest MCEG $\mathcal{C}^M$ so that

$$\pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_y || \Lambda_{x_{fail,0}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{w}^*}) = \sum_{x \in \boldsymbol{x}} \sum_z \pi(\Lambda_y | \Lambda_x, \Lambda_z, \Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z},0}))) \pi(\Lambda_z | \Lambda_x, \Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z},0})))$$
$$\times \hat{\pi}(\Lambda_x | \Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z},0}))), \quad (31)$$

where

$$\hat{\pi}(\Lambda_x | \Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z},0}))) = \frac{\hat{\pi}(\Lambda_x, \Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z}})))}{\hat{\pi}(\Lambda(w(\boldsymbol{b}_{F_{\boldsymbol{x} \cup y \cup z},0})))}. \quad (32)$$

The numerator and denominator are the post-intervened path probabilities. Note that these can be computed using Equation (28). Assuming that a stochastic manipulation on $\hat{\boldsymbol{\theta}}_{w^*}$ is equivalent to forcing each $x$ with probability $\pi(\Lambda_x || \hat{\boldsymbol{\theta}}_{w^*})$ for every $x \in \boldsymbol{x}^*$ [5], we can obtain Equation (31) by expressing the causal query as

$$\pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_y || \Lambda_{x_{fail,0}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{w}^*}) = \sum_{x \in \boldsymbol{x}^*} \pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_y || \Lambda_{x_{fail,0}}, x) \pi^{\Lambda_{\mathcal{C}^M}}(\Lambda_x | \Lambda_{x_{fail,0}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{w}^*}). \quad (33)$$

The first component on the right hand side of the equation can be evaluated by applying the results in Equation (19), and the second component can be simplified to Equation (32). By doing this, we have the expression in Equation (31).

**Example 2.** *Given the idle system in Figure 4, suppose routine maintenance involved in checking the oil level, cleaning the leakage, and topping up the oil, but this did not fully prevent the oil leak. The manipulations imported to the idle system under this intervention are then different from the one we discussed in Example 1. Suppose florets $\mathcal{F}(w_2), \mathcal{F}(w_3), \mathcal{F}(w_4)$ are directly affected in response to the maintenance. Then, these florets are stochastically manipulated, and $\boldsymbol{w}^* = \{w_2, w_3, w_4\}$. This gives the same $\Lambda(w(b_{\mathcal{F}(e(\boldsymbol{x})),0}))$ as in Example 1. If we are interested in how the sight glass or buchholz defect is affected by this intervention, then the effect event is $x_{s/b,1}$. Note that this event is unobservable and $\Lambda(w(b_{\mathcal{F}(e(x_{s/b,1})),0})) = \bigcup_{w \in \{w_{19}, \cdots, w_{24}\}} \Lambda(w)$.*

*Here, we can choose $X_{alarm}$ as the partition events $z$, and these are always observable. Next the manifest MCEG is constructed from the idle MCEG by removing the paths that do not traverse any position in $\{w_{19}, \cdots, w_{24}\}$. The manipulated MCEG is obtained by further deleting the paths that terminate in $w_\infty^f$ from the manifest MCEG, see Figure 6. If the post-intervention probabilities $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}^*}$ are known, then we can evaluate the path probabilities in the manipulated MCEG following the factorisations we specified in Equation (28). Then, conditional on the manifest paths, each probability in Equation (31) can be computed to estimate the effects of the observed maintenance on the sight glass or the buchholz.*
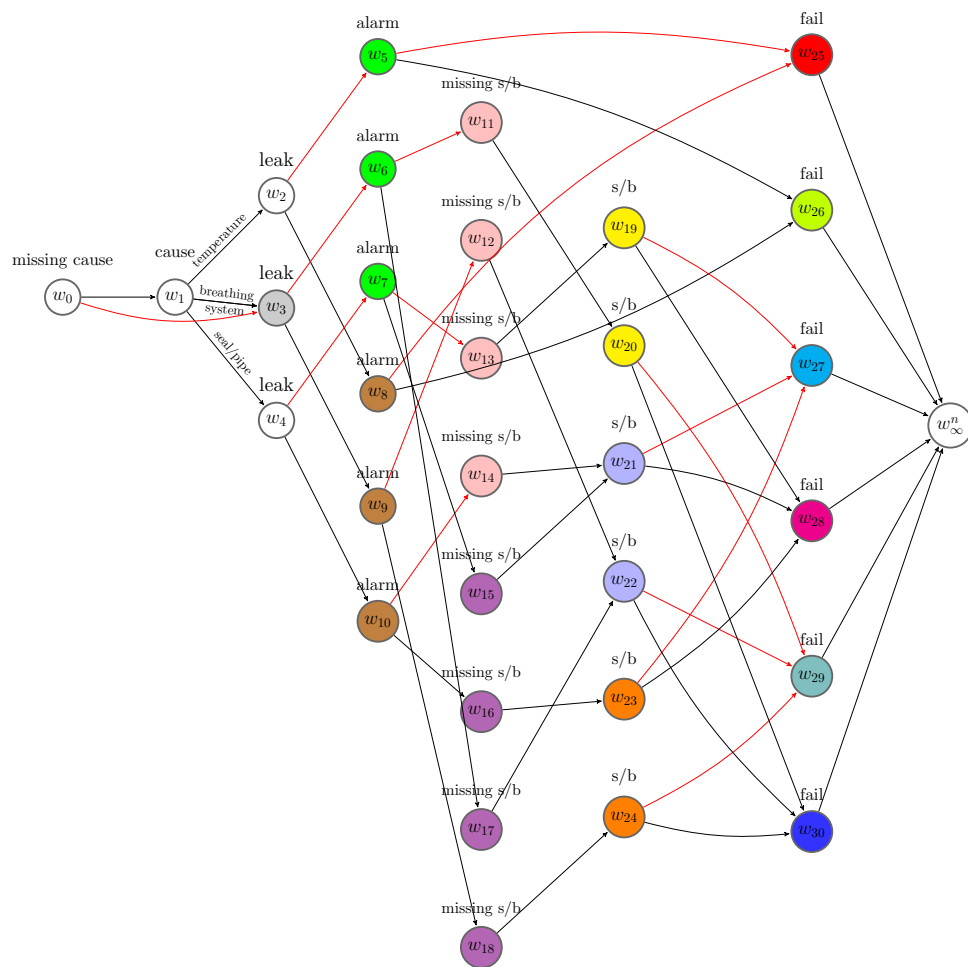
**Figure 6.** The manipulated MCEG for Example 2.

## 4. Experiments

Due to commercial sensitivity, we cannot disclose the real maintenance data from the energy distribution company and examine our methodology on it. Here, we show experimentally, using synthetic data, how the structural learning algorithm over a class of MCEGs can be used to provide useful causal inferences. We, then, perform a comparative study to demonstrate how the predictions are improved when incorporating the causal algebras we specified in previous section into the algorithm for the synthetic experimental data.

### 4.1. Causal Discovery with the Structural Learning Algorithm

Assume a ground truth missingness staged tree in Figure 3 and a corresponding MCEG in Figure 4 are valid. Assume the causal ordering here is $\Pi_1 = X_{cause} \prec X_{leak} \prec X_{alarm} \prec X_{s/b} \prec X_{fail}$. The oil leak, alarm, and sight glass or buchholz defect are faults that may appear before a failure or routine maintenance. Thus, the oil leak could be a potential cause of alarm and the defect in buchholz or sight glass. We assume that, for any floret, the parameters of primitive probability vector are independent, and the vectors of primitive probabilities associated with each stage are mutually independent.

This ensures a model search based on product of independent Dirichlet priors over the model parameters and a closed-form conjugate analysis [30]. Based on these assumptions, we now generate observation data $D_1$ of size 5000 from the ground truth MCEG with the corresponding hypothesized transition probabilities. This emulates the dataset in a situation when there has been no intervention to the system.

To begin to learn a best model for $D_1$ given the event tree in Figure 2, we specify the Dirichlet hyperparameters. We use established methods and treat each $\alpha_{uj}$ as the number of phantom units [3], which is believed to arrive at $j^{th}$ child of stage $u$. We let the total phantom units entering the root vertex $v_0$ be 1 and denote this by $\alpha = 1$.

By performing the MAP algorithm, the best scoring MCEG is shown in Figure 7. In this MCEG, denoted by $\mathcal{C}(\boldsymbol{X}(\Pi_1), \boldsymbol{B})$, the positions representing the same variable $X_i \in \boldsymbol{X}$ are vertically aligned in descending order with respect to $\mathbb{P}(X_i(w) = 1|D_1, \mu(w_0, w))$. For transparency, the edges that are supposed to have a label "yes" have been coloured red for clarity.

The posterior means for each stage are summarised in Table 1. The score of this selected model is $-20{,}389.83$. The stages for $X_{leak}$, $X_{alarm}$ and the missing indicator of s/b defect in this tree are accurately learned by the algorithm when these are compared with the stages in the ground truth MCEG. In terms of the stages for $X_{s/b}$, the stage assigned to $v_{23}$ is wrong. There are 15 misclassifications appearing for $X_{fail}$. One possible reason is that the dataset is not sufficiently large to provide sufficient information on the last event modelled on the tree.

The best scoring MCEG in Figure 7 has a complex topology because many stages for the last variable modelled on the tree are misspecified. However, we can still summarise some causal explanations from it when assuming it is causal. We read the causal relationships from the semantics of a causal CEG in an analogous way to a causal BN [3,6]. For example, from Figure 7, we see that all the edges representing oil leak point to the stage $u' = \{w_6, \cdots, w_9\}$, which is coloured in green, while the edges representing no leak point to the stage $u'' = \{w_{10}, \cdots, w_{13}\}$, which is coloured in brown. The stage $u'$ is located above $u''$ on the tree, meaning the mean posterior probability of alarm at this stage is higher than that at $u''$.

Therefore, the oil leak gives rise to the likelihood of alarm. Root causes also lie upstream of alarm on the tree and can affect the possibility of alarm. However, from Figure 7, whether the cause is missing and which cause is observed appear to have no influence on alarm given an oil leak. Thus, given the oil leak, the alarm is independent of the root causes we specified for this model. We could say that the oil leak is the main cause of alarm given the hypothesised causal ordering $\Pi_1$. One causal implication of this discovery is that we could prevent an alarm by fixing or preventing the oil leak. For positions associated with failure indicators, $w_{37}$ is aligned at the lowest position. This means that the probability $\mathbb{P}(X_{fail} = 1|\mu(w_0, w_{37}), D_1)$ is the lowest compared with the probability of failure conditional on the position $w_{34}$ or $w_{35}$ or $w_{36}$. There are eight edges pointing to $w_{37}$ labelled by no s/b defect and only one edge pointing to it labelled by a s/b defect. Thus, to increase the reliability of the machine, we can schedule the preventive maintenance for the sight glass or the buchholz.

*4.2. A Comparison Study*

Now, we assume the routine intervention described in Example 2 has occurred, and Figure 4 portrays the real causal structure. We, then, simulate synthetic data $D_2$ of size 5000 from this intervened model to emulate an experimental dataset by the following setups. First, we assume the 5000 pieces of equipment here have been intervened in the same way by the same routine maintenance. Second, a complete and unique root-to-sink path on the tree can be identified for each case in $D_2$. Third, assume we have the estimated posteriors from the past failure data before conduction of routine maintenance, and these are now used as priors to generate the data that would be observed after the routine maintenance.

Here, the prior independence assumptions are still assumed to be valid so that conjugate sampling can be characterised. To simulate from the intervened system instead of the idle system, the florets $\mathcal{F}(w_2), \mathcal{F}(w_3), \mathcal{F}(w_4)$ are stochastically manipulated in response to the routine maintenance, and we adjust the corresponding Dirichlet hyperparameters as described in the previous section.
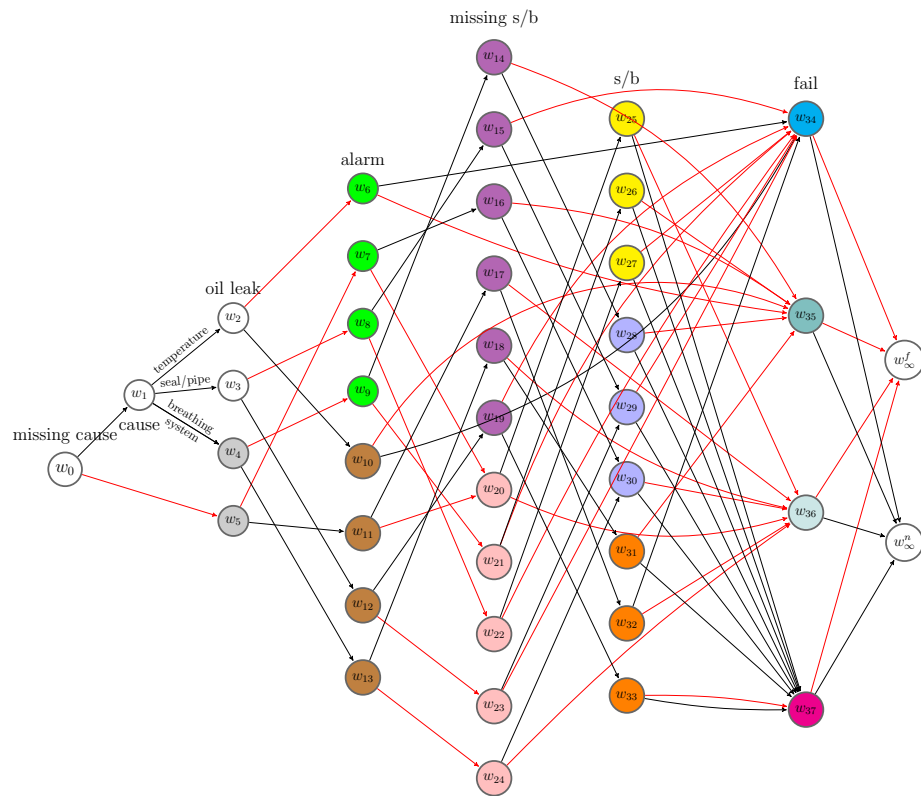
**Figure 7.** The best scoring MCEG selected for $D_1$ with hypothesised causal ordering $\Pi_1$.

**Table 1.** Mean posterior probabilities $\mathbb{P}(X = 1|stage, D_1)$.

| | $X_{leak} = 1$ | | | $X_{alarm} = 1$ | | $B_{s/b} = 1$ | | $X_{s/b} = 1$ | | | $X_{fail} = 1$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| stage | $w_2$ | $w_3$ | $w_4, w_5$ | $w_6, \cdots, w_9$ | $w_{10}, \cdots, w_{13}$ | $w_{14}, \cdots, w_{19}$ | $w_{20}, \cdots, w_{24}$ | $w_{25}, w_{26}, w_{27}$ | $w_{28}, w_{29}, w_{30}$ | $w_{31}, w_{32}, w_{33}$ | $w_{34}$ | $w_{35}$ | $w_{36}$ | $w_{37}$ |
| estimate | 0.77 | 0.69 | 0.50 | 0.69 | 0.49 | 0.51 | 0.29 | 0.80 | 0.67 | 0.51 | 0.78 | 0.70 | 0.59 | 0.45 |

It is possible to embody the effects of this intervention when learning the causal structure by incorporating the stochastic manipulations we developed in the previous section into the MAP algorithm. We can check whether this improves the causal structure learning and parameter estimations. On the corresponding missingness event tree, see Figure 2, we accordingly revise the Dirichlet hyperparameters of florets $\mathcal{F}(v_1), \mathcal{F}(v_5), \mathcal{F}(v_6)$ and $\mathcal{F}(v_7)$ using the method we proposed in Section 3.2.2.

We defined $\phi$ in Equation (27) to add uncertainties to the intervened floret distributions. In this study, we aim to compare the estimates learned from the best scoring model selected by the algorithm when no distributions are manipulated, i.e., $\phi = 1$, with the estimates learned from the best scoring model selected by the algorithm when inputting $\phi < 1$. In particular, we consider six different cases here: $\phi = 0.1, \phi = 0.3, \phi = 0.5, \phi = 0.7, \phi = 0.9$, and $\phi = 1$.

Now, we run the algorithm for $\alpha = 0.001, \alpha = 0.01, \alpha = 0.1, \alpha = 1, \alpha = 3, \alpha = 5$, where $\alpha$ is the prior parameter representing the number of phantom units entering the root node. We assess the resulted models in terms of situational errors [31] and MAP scores. The situational error (The total situational error of a tree is evaluated as $\gamma(\mathcal{T}) = \sum_{v \in V_{\mathcal{T}}} ||\boldsymbol{\theta}_v^* - \tilde{\boldsymbol{\theta}}_v||_2$) for a situation $v$ measures the Euclidean distance between the true conditional probabilities $\boldsymbol{\theta}_v^*$ and the mean posterior probabilities $\tilde{\boldsymbol{\theta}}_v$ estimated on the best scoring model.

The results are shown in Figure 8. The upper panel of each plot displays the total situational errors, while the lower panel displays the MAP scores for the best scoring models for different values of $\phi$. For any prior parameter $\alpha$ we choose, we observe that the best scoring model is selected from the algorithm by setting $\phi = 0.1$, which gives the smallest

situational error and the highest MAP score. In particular, the situational error rises when $\phi$ increases towards 1. Thus, the posterior parameters are better estimated by incorporating the manipulations into the learning algorithm when modelling the experimental data for an intervened system.

When $\phi = 1$ (i.e., the distributions are not manipulated), the MAP score in each plot of Figure 8 is much lower than that for $\phi = 0.1$. This means the best structure selected with $\phi = 0.1$ is more consistent with the dataset $D_2$ than the best model selected by the algorithm without importing stochastic manipulations.



(**a**) $\alpha = 0.001$

(**b**) $\alpha = 0.01$

(**c**) $\alpha = 0.1$

(**d**) $\alpha = 1$

(**e**) $\alpha = 3$

(**f**) $\alpha = 5$
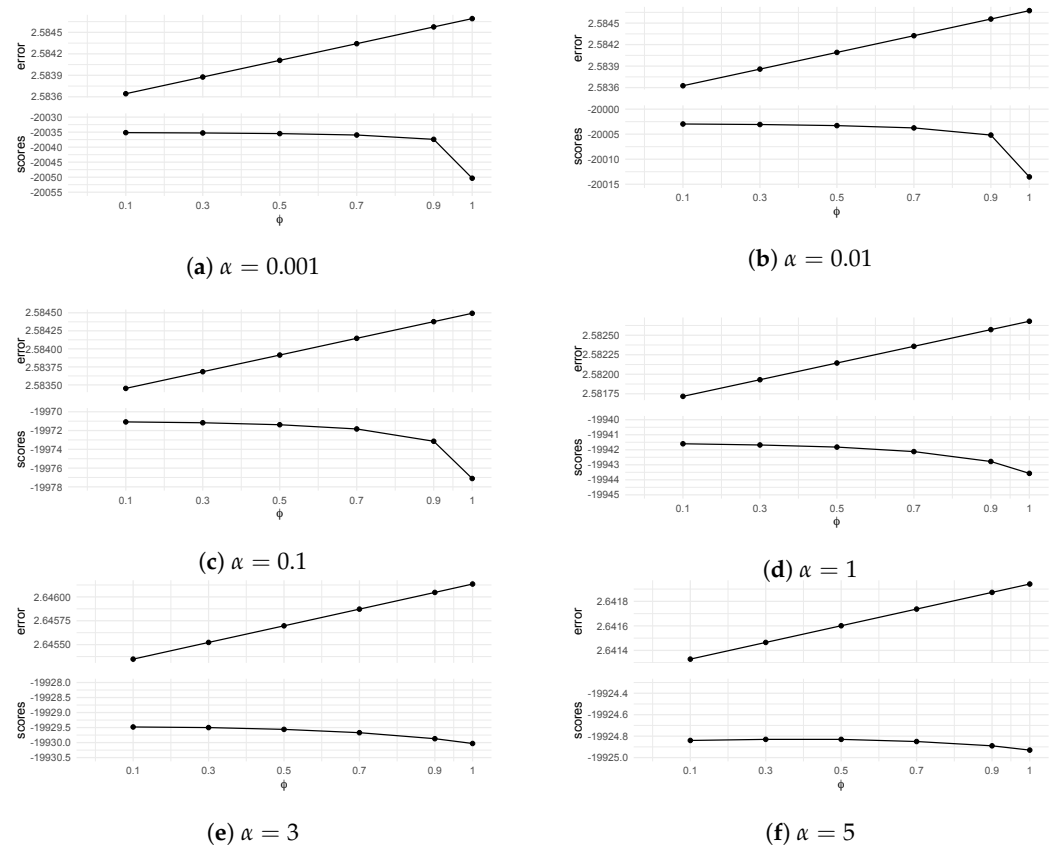
**Figure 8.** Comparing situational errors and MAP scores for the best scoring models selected to fit $D_2$. The x-axis of each plot is labelled by different values of $\phi$, where $\phi = 1$ refers to the case when no manipulation is imported to the prior. Each plot displays results for a specified total phantom number $\alpha$.

## 5. Discussion

Thus far, we demonstrated how the context-specific CEG is a compelling graphical tool for analysing system failure data. This happens not only because of its ability to represent structural asymmetries but also its flexibility in being able to perform the necessary analyses in a straightforward way even in the presence of censored data that are informedly missing; causal analyses can be performed through simple MAP structural learning algorithms. We developed bespoke causal algebras for the routine intervention and extended the back-door theorems for identifying its causal effects on the MCEG. The results from our designed experiments confirm the usefulness of these bespoke causal algebras in structural learning to improve the predictions needed for system reliability.

One concern of the study is that the model classes containing the best explanation can become huge when the systems are very large. However, the established methodology allows us to scale up the search space for more complex models with up to 20 variables [32]. Furthermore, these challenges associated with scalability are generic ones and are currently being actively researched. Each new development can be simply translated into causal analyses of reliability systems using the technologies we described above.

## References

1. Langseth, H.; Portinale, L. Bayesian networks in reliability. *Reliab. Eng. Syst. Saf.* **2007**, *92*, 92–108. [CrossRef]
2. Cowell, R.G.; Smith, J.Q. Causal discovery through MAP selection of stratified chain event graphs. *Electron. J. Stat.* **2014**, *8*, 965–997. [CrossRef]
3. Collazo, R.A.; Görgen, C.; Smith, J.Q. *Chain Event Graphs*; CRC Press: Boca Raton, FL, USA, 2018.
4. Görgen, C.; Smith, J.Q. A differential approach to causality in staged trees. In Proceedings of the Conference on Probabilistic Graphical Models, Lugano, Switzerland, 6–9 September 2016.
5. Pearl, J. *Causality: Models, Reasoning and Inference*; MIT press: Cambridge, MA, USA, 2000; Volume 29.
6. Thwaites, P.; Smith, J.Q.; Riccomagno, E. Causal analysis with chain event graphs. *Artif. Intell.* **2010**, *174*, 889–909. [CrossRef]
7. Thwaites, P. Causal identifiability via chain event graphs. *Artif. Intell.* **2013**, *195*, 291–315. [CrossRef]
8. Barclay, L.M.; Hutton, J.L.; Smith, J.Q. Chain event graphs for informed missingness. *Bayesian Anal.* **2014**, *9*, 53–76. [CrossRef]
9. Yu, X.; Smith, J.Q. Hierarchical Causal Analysis of Natural Languages on a Chain Event Graph. *arXiv* **2021**, arXiv:2110.01129.
10. Yu X; Smith, J.Q.; Nichols, L. Bayesian Learning of Causal Relationships for System Reliability. In Proceedings of the 7th International Symposium on Reliability Engineering and Risk Management, Beijing, China, 12–14 November 2020.
11. Mohan, K.; Pearl, J. Graphical models for recovering probabilistic and causal queries from missing data. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1520–1528.
12. Mohan, K.; Pearl, J. Graphical models for processing missing data. *arXiv* **2018**, arXiv:1801.03583.
13. Saadati, M.; Tian, J. Adjustment criteria for recovering causal effects from missing data. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Würzburg, Germany, 16–20 September 2019; pp. 561–577.
14. Yu, X. Hierarchical Causal Analysis on Chain Event Graphs. Ph.D. Thesis, University of Warwick, Coventry, UK, Unpublished.
15. Barclay, L.M.; Collazo, R.A.; Smith, J.Q.; Thwaites, P.A.; Nicholson, A.E. The dynamic chain event graph. *Electron. J. Stat.* **2015**, *9*, 2130–2169. [CrossRef]
16. Freeman, G.; Smith, J.Q. Dynamic staged trees for discrete multivariate time series: Forecasting, model selection and causal analysis. *Bayesian Anal.* **2011**, *6*, 279–305. [CrossRef]
17. Pensar, J.; Talvitie, T.; Hyttinen, A.; Koivisto, M. A Bayesian approach for estimating causal effects from observational data. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 5395-5402.
18. Cooper, G.F.; Yoo, C. Causal discovery from a mixture of experimental and observational data. *arXiv* **2013**, arXiv:1301.6686.
19. 9 types of Maintenance: From Preventive Maintenance to Corrective Maintenance and Everything in Between. Available online: https://www.roadtoreliability.com/types-of-maintenance/ (accessed on 19 February 2021).
20. Bedford, T.; Cooke, R. *Probabilistic Risk Analysis: Foundations and Methods*; Cambridge University Press: Cambridge UK, 2001.
21. Lienig, J.; Bruemmer, H. Reliability Analysis. In *Fundamentals of Electronic Systems Design*; Springer: Cham, Switzerland, 2017; pp. 45–73.
22. Bicen, Y. Monitoring of critical substation equipment. In Proceedings of the 3rd International Istanbul Smart Grid Congress and Fair (ICSG), Istanbul, Turkey, 29–30 April 2015; pp. 1–4.
23. Guessoum, Y.; Aupiedy J. Modeling the impact of preventive maintenance over the lifetime of equipments. In Proceedings of the International Conference on Electricity Distribution, Lyon, France, 7–8 June 2010.
24. Kijima, M. Some results for repairable systems with general repair. *J. Appl. Probab.* **1989**, *26*, 89–102. [CrossRef]
25. Preventive Maintenance Checklist for Transformer. *Learn Electrician*. Available online: https://learnelectrician.com/preventive-maintenance-checklist-for-transformer/(accessed on 1 June 2021).
26. Smith, J.Q. Non-linear state space models with partially specified distributions on states. *J. Forecast.* **1990**,*9*, 137–149. [CrossRef]
27. Smith, J.Q. A comparison of the characteristics of some Bayesian forecasting models. *Int. Stat. Rev. Int. Stat.* **1992**, *60*, 75–87. [CrossRef]
28. Smith, J.Q. The multiparameter steady model. *J. R. Stat. Soc. Ser. (Methodological)* **1981**, *43*, 256–260. [CrossRef]

29. Smith, J.Q. A generalization of the Bayesian steady forecasting model. *J. R. Stat. Soc. Ser. (Methodological)* **1979**, *41*, 375–387. [CrossRef]
30. Freeman, G.; Smith, J.Q. Bayesian MAP model selection of chain event graphs. *J. Multivar. Anal.* **2011**, *102*, 1152–1165. [CrossRef]
31. Collazo, R.A.; Smith, J.Q. A new family of non-local priors for chain event graph model selection. *Bayesian Anal.* **2016**, *11*, 1165–1201. [CrossRef]
32. Carli, F.; Leonelli, M.; Riccomagno, E.; Varando, G. The R package stagedtrees for structural learning of stratified staged trees. *arXiv* **2020**, arXiv:2004.06459.