A Modeling Framework for Troubleshooting Automotive Systems

Håkan Warnquist, Jonas Kvarnström and Patrick Doherty

Håkan Warnquist, Jonas Kvarnström and Patrick Doherty, A Modeling Framework for Troubleshooting Automotive Systems, *Applied Artificial Intelligence*, 2016, 30(3), 257-296.

Applied Artificial Intelligence is available online at informaworldTM:
http://dx.doi.org/10.1080/08839514.2016.1156955

A Modeling Framework for Troubleshooting Automotive Systems

Håkan Warnquist, Jonas Kvarnström, and Patrick Doherty

Dept. of Computer and Information Science, Linköping University

Abstract

This paper presents a novel framework for modeling the troubleshooting process for automotive systems such as trucks and buses. We describe how a diagnostic model of the troubleshooting process can be created using event-driven non-stationary dynamic Bayesian networks.

Exact inference in such a model is in general not practically possible. Therefore we evaluate different approximate methods for inference based on the Boyen-Koller algorithm. We identify relevant model classes that have particular structure such that inference can be made with linear time complexity using a novel inference algorithm called the Quickscore Variant. We show how the algorithm can be applied for inference when only a single fault is assumed and when multiple faults are possible. We also show another inference method that can be used when multiple faults are possible but a single fault is most likely.

We also show how models created using expert knowledge can be tuned using statistical data. The proposed learning mechanism can use data that is collected from a heterogeneous fleet of modular vehicles that can consist of different components.

The proposed framework is evaluated both theoretically and experimentally on an application example of a fuel injection system.

1 Introduction

This paper presents a framework for modeling the troubleshooting process for automotive systems such as trucks and buses. When the vehicle is not in a workshop, troubleshooting is done remotely by help-desk personnel who can run remote diagnostic tests or interrogate the driver. In the workshop, troubleshooting is done by a mechanic with access to a wider range of actions for diagnosis and repair. The troubleshooting incurs a repair cost related to both the required materials and the direct costs of actions performed by mechanics and help-desk personnel. Additionally, it incurs a downtime cost which includes indirect costs from disrupting operations and not being able to utilize the vehicle for its purpose. Since trucks and buses are typically used in commercial applications, all of these costs affect the total cost of operation and thereby the profit for the vehicle owner. These costs can be large and should be taken into account when determining which actions to perform. For example, the direct repair and maintenance costs for long-haulage heavy trucks can account for as much as 10000 € per year [7]. When an unexpected breakdown occurs, the economic loss for the vehicle owner of a long-haulage truck can be on average 1000 € each time [11]. To safeguard against unexpected losses, the vehicle owner can buy a repair and maintenance contract from the service provider for a fixed price [20, 34, 40]. Some of these contracts also compensate the owner for downtime and unexpected breakdowns [35, 41]. This gives the service providers reason to offer service solutions that are cost-efficient in regards to both repair and downtime costs as this can improve their margins on these contracts.

1.1 Background

The modeling framework presented here is part of a decision support system for troubleshooting vehicles for which a problem has been detected on the road or in the workshop. In [42] methods for cost modeling and an algorithm for decision making are presented. As presented there, the decision support system is designed to minimize the overall costs of downtime and repair by integrating the planning of the actions that can be performed before a workshop visit with those that can be performed during the workshop visit. Before the workshop visit, remote tests can be performed that contribute to determining how urgently a workshop visit is needed. Also decisions can be made of when and where to visit a workshop, taking into account the different downtime costs of doing so at different times, e.g. aborting the current assignment for an immediate workshop visit or continuing operation until the next already scheduled workshop visit. During the workshop visit, repair actions and workshop tests can be carried out. This system takes as input diagnostic information such as Diagnostic Trouble Codes (DTCs) generated from the on-board diagnostic system [37], driver observations, diagnostic test results, and information about previous repairs. The output consists of explanations of the current state of information in the form of probabilities that a component has a particular fault mode as well as recommendations of possible troubleshooting actions ranked by the expected cost of troubleshooting the vehicle from the current state.

This decision support system has two important subcomponents: a *planner* that computes the recommendations and a *diagnoser* that supports the planner with computations of fault probabilities and test outcome probabilities given current diagnostic information. [42] uses a diagnoser where a Bayesian network [28] models causal dependencies between faults on components and observations. This Bayesian network model is static which means that though the value of a variable may be uncertain, it is assumed that it does not change during the troubleshooting session. Using a static Bayesian network has the advantage that fault and observation probabilities can be computed with sufficient efficiency using a Bayesian inference algorithm such as the Junction Tree algorithm [19]. However, repairs do change the state of variables, e.g. those representing faults. This was handled using the method of [13] where the vehicle is assumed to have a single fault and it is required that each repair is immediately followed by a function control that determines whether the repair solved the vehicle problem or not. If after repairing a component the function control indicates that the vehicle problem is still not solved, this means that the component never was faulty because of the single fault assumption. Then the repair did not change the state of any variable and the static model is still valid. If the function control indicated that the vehicle problem was solved, then we are done and it does not matter that the repair changed the state of a variable. This type of diagnoser is used in other existing decision support systems for troubleshooting, e.g. [13, 18].

The static Bayesian network model can be feasible only for shorter troubleshooting sessions. When troubleshooting remotely, the time between the first observation and the end of the troubleshooting session can be several weeks. In this case it is no longer feasible to assume a static model. New faults could occur and the behavior of faults may change so that previously observed symptoms appear or disappear. Further, it is not always possible or cost efficient to perform a function control after each repair.

1.2 Contributions

In this paper we present how the diagnostic model can instead be created using a dynamic Bayesian model [5] where causal dependencies between component faults and observation are modeled over time. The model captures important aspects of integrated remote and workshop troubleshooting such as how the probabilities of observations and faults are affected by repairing components and operating the vehicle for prolonged periods.

Exact inference in a dynamic Bayesian network is computationally difficult for larger networks [21]. We explore some general methods for approximate inference that are suitable for our application and improve on these with a novel approximate inference algorithm. When the model has a particular structure we can use this algorithm to compute test outcome probabilities and update fault probabilities after observations in time linear with the model size. Typically there is at most a single fault present, but more faults are possible. For this, another inference method is proposed that combines a single fault model with a multiple fault model so that multiple faults are supported, but precision is higher for single faults than when using the basic multiple fault approach.

The task of creating a model is complicated by the fact that modern automotive vehicles are modular in design [27] which means that many unique instantiations of vehicles can be created from a smaller set of modules. Each such instantiation may require a different diagnostic model. It is not feasible to create each such model from scratch as the work effort would be too large. Therefore, in the proposed modeling framework, we show how it is possible to create a model of a complete vehicle from a set of smaller models of individual modules.

To have the models maintainable over time and to be able to learn and adapt to how the vehicles actually behave, we learn model parameters from statistical data. However, we cannot solely rely on this because, especially for newer vehicles, there can be little or no statistical data available. In this case, the experts designing the vehicle can still have an understanding of how often faults occur and how a hypothesized fault in a component would manifest itself in terms of observations. In the proposed method, an initial model is created by experts and then the model parameters are tuned using statistical data. We use a Bayesian learning approach that combines the prior knowledge provided by experts with the statistical data when setting parameter values. The fact that the statistical data is collected from a population of vehicles that are modular and heterogeneous must be taken into consideration when designing the learning mechanism. We show theoretically and experimentally that the proposed learning method works on both synthetic models and a larger real world model.

For evaluation purposes, the modeling framework has been applied to a real world example consisting of a truck fuel injection system. This system is particularly difficult for workshop mechanics to troubleshoot efficiently because many faults are non-deterministic in their behavior and have common symptoms. Being a complex system involving a mix of electronically controlled electrical, mechanical, and hydraulic components, it is particularly challenging to model.

The paper is outlined as follows. First we present the application example as a motivation for the framework. Next we present an overview of the entire decision support system including the parts described in Warnquist et al. [42] and present the problem formulation for the diagnoser. Then we will describe and provide the

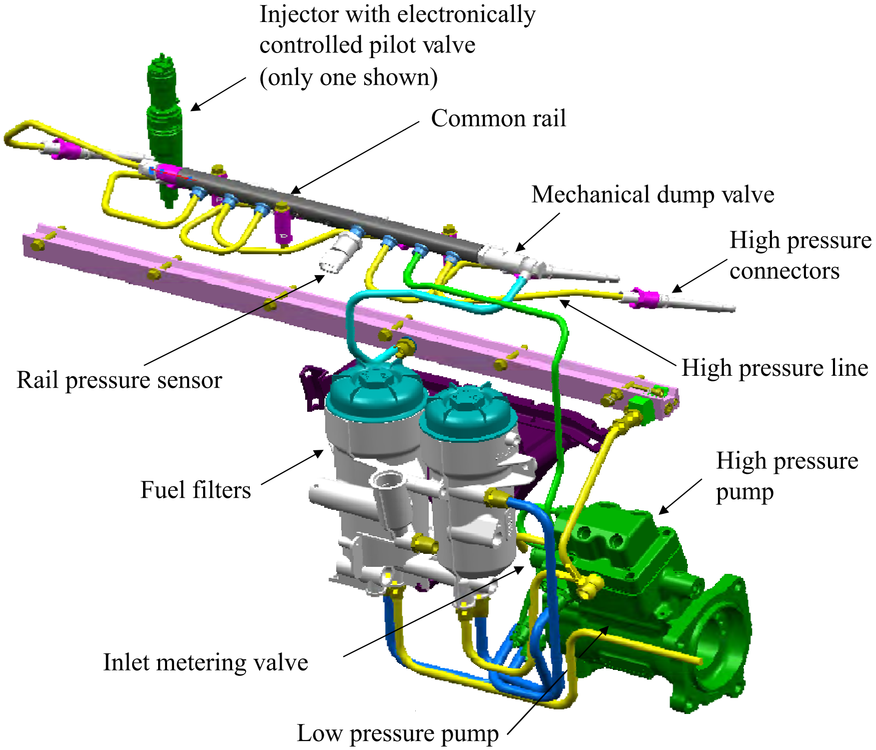

Figure 1: An overview of the XPI system and its main components.

theory for all the parts of the novel diagnostic framework: the model, the inference algorithm, and the learning method. After that we will show an experimental evaluation of all parts of the diagnostic framework. Finally we will discuss related work and conclude.

2 Application Example

As an application example for the diagnostic framework we will model the Scania XPI fuel injection system (e**X**treme high-**P**ressure fuel-**I**njection), which is responsible for dosing the correct amount of diesel fuel into the cylinders. Figure 1 shows an overview of the XPI system and its main components.

The XPI system is a common rail fuel injection system. This means that fuel is injected directly into the cylinder by individually controlled injectors taking fuel from a highly pressurized common fuel rail. This common rail technology makes it possible to inject fuel independently of the camshaft position at much higher pressures than with unit injection systems where the injection pressure and timings are controlled by the camshaft.

The low pressure pump (LPP) circulates fuel from the fuel tank through the fuel filters. The inlet metering valve (IMV) directs some of the fuel to the high pressure pump (HPP) so that the pressure in the common rail, as measured by the pressure sensor, is as commanded by the engine management system (EMS). This pressure typically varies between 800 bars at idle and 2400 bars at maximum engine power output. From the common rail, fuel is directed to the injectors through a high pressure line and a connector. When commanded by the EMS, the injectors spray fuel into the cylinders through a small micrometer-sized valve. Excess fuel from the injectors is returned to the low pressure side through the return fuel rail. There is also a mechanical dump valve that protects the common rail from excessive pressures by releasing fuel into the return fuel rail.

The on-board diagnosis system on the vehicle is very capable of detecting and isolating *electrical* faults on the injectors, the IMV, and the rail pressure sensor. However *mechanical* faults and leakage problems are typically only detectable by observing that it is not possible to gain the commanded pressure in the common rail. If this happens a diagnostic trouble code (DTC) is generated that indicates that the rail pressure is below commanded. Many faults can cause this DTC, e.g. leaking injectors, open IMV, failing HPP or LPP, clogged fuel filters, or a fault in the rail pressure sensor. Faulty behavior is not deterministic, e.g. clogged fuel filters may or may not cause the engine to stall and a leaking injector may or may not cause the output effect of one cylinder to deviate from that of the other cylinders. The EMS may respond to failures that have been detected by the on-board diagnostic system. For example, when a DTC indicates that the rail pressure is below commanded, the EMS limits the engine torque to protect the engine and the driver experiences the vehicle as powerless and sees a warning light that prompts the driver to seek a workshop.

Some of the faults that are indistinguishable by the on-board diagnostic system can be distinguished by additional testing either remotely or at the workshop. For example, a mechanic can measure the return fuel pressure by mounting a pressure meter on the return fuel line or measure the output effect of the cylinders by running a test on the EMS. Other faults can only be distinguished by trial and error. However, the mechanic must take care when troubleshooting the XPI system because many components on the high pressure side must be completely replaced with new components if they are removed because of strict cleanliness requirements. Some of the components of the XPI system are also very expensive which means that the cost of non-efficient troubleshooting can be large.

3 A Decision Support System for Integrated Remote and Workshop Troubleshooting

In this section we will describe the decision support system for integrated remote and workshop troubleshooting proposed in [42] which the modeling framework presented in this paper supports. An overview of the decision support system is shown in Figure 2.

The decision support system uses a state to represent all the information that it needs about the vehicle and its environment. It proposes actions to perform in order to take the vehicle to a desired state with no faults. Whenever an action is performed on the vehicle, this is registered as a sequence of time-stamped events signaling that the state of the vehicle or its environment has changed. A new state can be computed from a previous state and a sequence of new events that have occurred since then. The previous state can be another previously computed state or the initial state when no events have occurred. After each performed action, a new state and recommendation is computed in the decision support system.

Each state consists of two parts: the *environment state* and the *diagnostic state*. The environment state contains information about the vehicle's geographical location, the current time, and the time and location of the next planned workshop occasion. The diagnostic state contains information about the marginalized fault probabilities for each component of the vehicle together with intermediate results from previous computations that can be reused for further computations.

There are six types of actions:

- *Reservation actions* that set the time and place for the next planned workshop visit, creating a *reservation event*.
- *Idling actions* that cause the vehicle to await the time of the next planned workshop visit. While waiting the vehicle can either be operated normally in which case an *operation event* is created containing information of the distance operated, or be put on hold creating a *null event* that only signals the time has changed.
- *Transport actions* that take the vehicle to the location of the planned workshop visit either by driving it there or by towing it there. This creates a *location event* and possibly an operation event if the vehicle is driven.
- *Observing actions* that gain information about the vehicle e.g. by asking the driver questions about the vehicle's behavior or performing tests. These actions can either be possible to perform remotely or require a workshop to be performed. For an observing action one or more *observation events* are created containing information of the outcome of the observation and possibly an operation event if the action requires the vehicle to be operated.

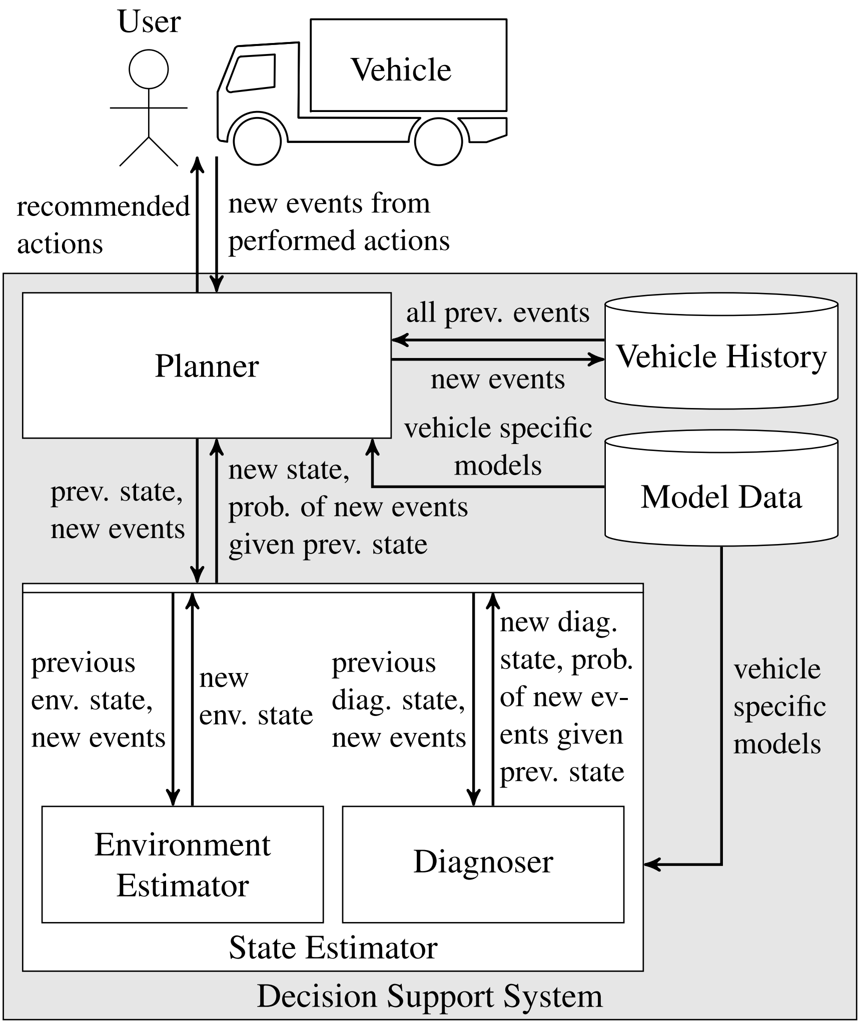

Figure 2: Architecture of decision support system

- *Repair actions* that replace or repair specific components thus removing any fault present on them creating a *repair event*. These actions can only be performed in the workshop.

- The *end session action* causing the ongoing troubleshooting session to end and prompting the user to return the vehicle to normal operation.

The *planner* computes which actions to recommend. The planning algorithm that is used to select actions is described in Warnquist et al. [42]. It is a look-ahead search algorithm that evaluates the expected costs of actions a limited number of steps into the future. The planner uses a model database to determine which actions are applicable for the current vehicle as well as their costs. The results of performed actions are stored in the vehicle history database as events.

The planner is supported by the *state estimator* which, using a previous state and a sequence of new events, computes a new state and the probability that the observation events in the sequence have a particular outcome.

In the state estimator, the *environment estimator* is responsible for updating the environment state after reservation, location, and null events, and the *diagnoser* is responsible for updating the diagnostic state after operation, observation, and repair events. Updating the environment state is a simple constant time operation, but updating the diagnostic state is a complex operation depending on how the diagnoser is implemented. Vehicle-specific models are retrieved from the model database.

Limitations In [42], fault probabilities and action outcome probabilities are computed by making inference in a static Bayesian network model using the Bayesian network tool SMILE [6]. As discussed in the introduction, a drawback with using a static Bayesian network model is that it is incapable of capturing changes in the presence of faults and their behavior. In particular, every repair action must be followed up by a function control of the vehicle to test whether the action did or did not solve the vehicle problem. These are among the issues that are addressed in the new modeling framework presented in this paper.

The remainder of the paper will describe the new modeling framework and how diagnostic models that can be used by the diagnoser are created, used, and maintained. To address the issues with the static Bayesian network this model is based on a dynamic Bayesian network. Care is taken to make the inference and learning of model parameters from statistical data tractable. The framework is applied in the context of troubleshooting trucks for the application example, but can be used more generally for diagnosing any kind of electromechanical systems with similar properties.

4 Problem Formulation

The diagnoser must for every component of the system be able to compute the probability of a fault being present on that component given the sequence of operation, observation, and repair events that have occurred so far. It must also be able to compute the probability of the outcome of a new observation event. Intermediate computations can be stored in the diagnostic state if that removes the need of re-evaluating previous events.

For every recommendation, the planning algorithm needs to compute a new diagnostic state for a number of actions that grows linearly with the number of components. Hundreds of components can be related to a vehicle problem, and therefore the diagnoser must be able to able to compute diagnostic states efficiently for that many components.

The diagnoser will need a diagnostic model that represents the vehicle and its environment. However, vehicles can come in many different configurations, e.g. they have different number of cylinders and use different technologies for fuel injection, leading to a combinatorial explosion. Because of modular design, different configurations share many components, e.g. the same injectors are used for all common rail fuel injection systems regardless the number of cylinders. The diagnostic model must therefore also be modular in the same way. If we know which components a vehicle consists of, then we should be able to assemble a full vehicle model from component level submodels.

For new vehicle types, the knowledge of how faults behave comes primarily from the experts that have engineered the vehicles. However, as the vehicles are used in the field, more knowledge can be gained from statistical data. The diagnostic model must both function for new vehicle types and be able to be automatically improved with statistical data.

5 Diagnostic Model

The diagnostic model is a probabilistic model that describes the behavior of faults with regards to their presence and observations that can be made, and how this behavior is affected by events that occur. Properties of the vehicle that are assessed to be relevant for describing this behavior are modeled with binary discrete random variables. To be able to model how these properties change over time, probabilistic relations between variables over time are modeled with a dynamic Bayesian network (DBN). The DBN models the joint probability distribution of the variables at different discrete time steps. The joint probability distribution over the variables at a single time step given the variables in the previous time step is a *time slice*. The first time slice models the joint probability distribution of the variables when the vehicle is completely new. In our model, each time step corresponds to an event that has occurred since the previous time step. Events do not occur with regular intervals; the time that passes and the number of kilometers the vehicle has traveled between each pair of events can vary. This means that the time slices can differ from each other and the DBN belongs to the class of *non-stationary Dynamic Bayesian Networks* [31].

Example 1 (Example System)**.** When describing the diagnostic framework we will use a small example system consisting of two components from the XPI system: the high pressure pump (HPP) and a single fuel injector. The components can either have a mechanical fault or a leakage fault. Any fault on the HPP can cause the feed pressure to become too low. Any fault on the injector can cause the cylinder effect to become too low for that cylinder. If either component has a leakage fault, this can cause the fuel rail to become unable to hold pressure.

There are three possible observations that can be made. One is a DTC for low fuel pressure that is generated if the rail pressure is below commanded either because the feed pressure is too low or because the fuel rail is unable to hold pressure. Another is the fuel rail's capability to hold pressure which can directly be measured in the workshop. The third is a test that can be run to detect deviations in the cylinder output effect.

5.1 Variables

When describing the diagnostic model we will use upper case letters for random variables. A binary random variable can have two values, a positive value $\top$ and a negative value $\bot$. For a binary random variable X, we will write $X = \top$ or lower case x for positive values, and $X = \bot$ or $\neg x$ for negative values.

The vehicle properties that we model with random variables are described below. Table 1 shows the random variables for the small system described in Example 1.

Repairable Components

A component is a physical part of the vehicle that can have faults and is the minimal repairable unit. Being the minimal repairable unit means that the component is always replaced or repaired as a whole [29].

Let $\mathcal{I}_\mathrm{C}$ be the set of all components in the model. For each component $i \in \mathcal{I}_\mathrm{C}$ and each time step t, there is a *component variable* C_i^t that models whether the component has a fault at time step t. It can have the values *faulty* c_i^t or *not faulty* $\neg c_i^t$.

Components $\mathcal{I}_{\mathsf{C}} = \{1,2\}$	
C_1^t	high pressure pump
C_2^t	fuel injector

Faults $\mathcal{I}_{\mathsf{F}} = \{3,4,5,6\}$, $\mathcal{I}_{\mathsf{F}\|1} = \{3,4\}$, $\mathcal{I}_{\mathsf{F}\|2} = \{5,6\}$	
F_3^t	mechanical fault on HPP
F_4^t	leakage fault on HPP
F_5^t	leakage fault on fuel injector
F_6^t	mechanical fault on fuel injector

Symptoms $\mathcal{I}_{\mathsf{S}} = \{7,8,9\}$	
S_7^t	low feed pressure
S_8^t	unable to hold pressure in fuel rail
S_9^t	too low cylinder output effect

Symptom causes $\mathcal{I}_{\mathsf{S}\|3} = \{7\}$, $\mathcal{I}_{\mathsf{S}\|4} = \{7,8\}$, $\mathcal{I}_{\mathsf{S}\|5} = \{8,9\}$, $\mathcal{I}_{\mathsf{S}\|6} = \{9\}$	
$S_{3,7}^t$	mechanical fault on HPP causes low feed pressure
$S_{4,7}^t$	leakage fault on HPP causes low feed pressure
$S_{4,8}^t$	leakage fault on HPP causes unable to hold pressure in fuel rail
$S_{5,8}^t$	leakage fault on fuel inj. causes unable to hold pressure in fuel rail
$S_{5,9}^t$	leakage fault on fuel inj. causes too low cylinder output effect
$S_{6,9}^t$	mechanical fault on fuel inj. causes too low cylinder output effect

Logical dependencies $\mathcal{I}_{\mathsf{L}} = \{10\}$, $\mathcal{I}_{\mathsf{X}\|10} = \{7,8\}$	
L_{10}^t	DTC: "rail pressure is below command" is generated

Observations $\mathcal{I}_{\mathsf{O}} = \{11,12,13\}$, $\mathcal{I}_{\mathsf{X}\|11} = \{10\}$, $\mathcal{I}_{\mathsf{X}\|12} = \{8\}$, $\mathcal{I}_{\mathsf{X}\|13} = \{9\}$	
O_{11}^t	Observed DTC "rail pressure is below commanded" is generated
O_{12}^t	Observed "unable to hold pressure in the fuel rail"
O_{13}^t	Observed "too low cylinder output effect"

Table 1: The random variables in the diagnostic model for the system described in Example 1.

Presence of Faults

A fault is a type of failure on a component that can only be solved by repairing or replacing the component.

Let $\mathcal{I}_{\mathsf{F}|j}$ be the set of all faults that can affect component j and let $\mathcal{I}_{\mathsf{F}} = \bigcup_{j\in\mathcal{I}_{\mathsf{C}}} \mathcal{I}_{\mathsf{F}|j}$ be the set of all faults in the model. A fault only belongs to a single component: For any components $j,k \in \mathcal{I}_{\mathsf{C}}$, $\mathcal{I}_{\mathsf{F}|j}$ and $\mathcal{I}_{\mathsf{F}|k}$ are disjoint if $j \neq k$. For each fault $i \in \mathcal{I}_{\mathsf{F}}$ and each time step t, there is a *fault variable* F_i^t that models whether the fault is present at time step t. It can have the values *present* f_i^t or *not present* $\neg f_i^t$.

Presence of Symptoms

A symptom is a deviation from normal behavior because of a fault on the vehicle. This does not necessarily have to be directly observable by an external observer.

Let $\mathcal{I}_{\mathsf{S}}$ be a set of all symptoms in the model. For each symptom $i \in \mathcal{I}_{\mathsf{S}}$ there is a *symptom variable* S_i^t that models whether the symptom is present at time t. It can have the values *present* s_i^t or *not present* $\neg s_i^t$.

Symptom Causes

A symptom is only present if it is currently being actively caused by one or more present faults. However, a given fault does not *always* cause the same symptoms. We must therefore distinguish between *actively* causing a symptom and having the *potential* to cause a symptom.

Let the $\mathcal{I}_{\mathsf{S}|i} \subseteq \mathcal{I}_{\mathsf{S}}$ be the set of all symptoms that potentially can be caused by fault i. This set does not vary over time. For each fault $i \in \mathcal{I}_{\mathsf{F}}$, each symptom $j \in \mathcal{I}_{\mathsf{S}|i}$, and each time step t, there is a *symptom-cause variable* $S_{i,j}^t$ that models whether fault i is actively causing symptom j at time step t. It can have the value *active* $s_{i,j}^t$ or *not active* $\neg s_{i,j}^t$.

It is possible that several faults are actively causing a symptom. For example, both a leakage in the injector and a leakage in the high pressure pump could be causing the fuel rail to be unable to hold pressure. In this case we will not necessarily be able to hold pressure if we repair only one of the faults.

Results of Observations Made by an External Observer

The variables defined above model what is actually true in the world. For example, in the diagnostic model for the system described in Example 1, the symptom variable S_8^t models whether the system is actually unable to hold pressure in the fuel rail at time t.

An external observer cannot directly see whether this is the case, but must make observations that can be noisy and thereby faulty. All such observations are modeled with *observation variables* as statements about the value of one of the other variables in the model. For example, the observation variable O^t_{12} models whether an observer, through an observation action, *perceived* that the system was unable to hold pressure in the fuel rail. In other words, O^t_{12} models whether the observer perceived S^t_8 to be true.

Let $\mathcal{I}_{\mathrm{O}}$ be the set of all observations that can be made. For each observation i and each time step t, the observation variable O^t_i models whether observation i is true at time step t. It can have the values *true* o^t_i or *false* $\neg o^t_i$. Each observation i is associated with the single other variable j that the statement is about, contained in the set $\mathcal{I}_{\mathrm{X}|i} = \{j\}$.

Dependencies between Observations and Multiple Symptoms

Typically there is a one-to-one relation between observations and symptoms. However, sometimes an observation is dependent on a combination of symptoms. In the example, the DTC for low fuel pressure is generated when either the symptom "low feed pressure" or the symptom "unable to hold pressure in the fuel rail" is present.

We call such a case a *logical dependency*. To each logical dependency i we associate a boolean function of the symptoms ϕ_i and the value of this function at time t is modeled with a *logical variable* L^t_i. E.g. $\phi_{10}(t) = S^t_7 \vee S^t_8$. Let $\mathcal{I}_{\mathrm{L}}$ be the set of all logical dependencies in the model and let the set $\mathcal{I}_{\mathrm{X}|i}$ be all symptoms for dependency i.

5.2 Continuous Quantities

By choosing to model the vehicle with discrete binary variables we risk losing valuable information from the many measurable continuous signals on board the vehicle. However workshop tests and on-board diagnostic tests can be designed to monitor such signals and output a binary result indicating if there is a fault present among the set of faults for which the test is sensitive. This can for example be done using Fault Detection and Isolation (FDI) techniques where a state-space model of the vehicle is used to compare different measurable signals by generating residual signals [15, 16, 39]. If the value of a residual deviates from zero, a fault is detected among the components that generate the signals used by that residual. Such a deviation on a residual can be modeled as a symptom with symptom-cause variables for all faults on the related components. The workshop test for measuring the fuel rail's capability to hold pressure is another such example where a positive test result is reported if the pressure in the fuel rail drops more than 200 bars during one minute after being pressurized.

5.3 Conditional Probability Distributions

Bayesian networks are represented as directed acyclic graphs where the vertices are random variables and conditional dependencies with other variables are specified with edges. For each variable, a conditional probability distribution (CPD) is defined that specifies the probability that the variable has a particular value given the values of all its parent variables. In a DBN, a variable can only depend on variables that are in the same or previous time step. A variable is said to be *persistent* if there are other variables in the next time step that depend on it, otherwise it is *transient* [21]. In the diagnostic model, only fault and symptom-cause variables are persistent.

We want to define the CPDs with few parameters that can be intuitively be understood by the experts modeling the system. The definitions must also be such that the model is modular, the inference is efficient, and it is possible to tune the parameters with statistical data.

The CPDs of component variables are the same in every time slice, regardless of the type of event it is associated with. A component i is faulty if one or more of its faults are present:

$$Pr(c_i^t | F_j^t = v_j \forall j \in \mathcal{I}_{\mathsf{F}|i}) = \begin{cases} 1 & \text{if } \bigvee_{j \in \mathcal{I}_{\mathsf{F}|i}} v_j = \top, \\ 0 & \text{otherwise.} \end{cases}$$

For the other variables, different definitions of the CPDs will be needed for different types of events.

5.4 CPDs for Operation Events

The operation of the vehicle is signaled with an operation event. During operation new faults may appear and the behavior of present faults may change.

CPDs of Fault Variables after Operation Events

The possibility of having faults on completely new vehicles is modeled with a *manufacturing fault probability* parameter $p_{\mathsf{fail},i}$ specifying the probability of fault i being present when the component is new, e.g. a manufacturing fault. At time step 0 the probability of a fault i is

$$\Pr(f_i^0) = p_{\mathsf{fail},i}. \tag{1}$$

The probability that a new fault occurs during a discrete time step increases with the duration of the associated operation event. The rate at which faults occur is called the failure rate. In complex electromechanical systems, the failure rate is often higher in the beginning when the component is new, so called infant mortality, and then settles at a constant level for older components [36].

The manufacturing fault probability can be used to account for infant mortality by approximating and compressing the initial period of time to $t = 0$. After the initial time step, we will assume that the failure rate does not change as the component ages. Then fault occurrence is a *homogenous Poisson-process* where the probability that a fault occurs in a specific time period is given by the cumulative exponential distribution. The distribution is parametrized with the failure rate of the fault measured in units of e.g. expected number of faults per operated kilometer or operated hour. In our application to road vehicles we measure operated time in kilometers as this measure is readily available from the odometer of the vehicle and most often reported in statistical data.

Let there be an operation event that occurs between time steps $t-1$ and t and let Δ_t be the distance in millions of kilometers the vehicle has traveled between those time steps. Let the *failure rate* parameter $\lambda_{\mathsf{fail},i}$ be the expected number of faults of type i per million kilometers of operation. Then the CPD of a fault variable F_i^t of fault i at time step t is dependent on whether there is a fault in the previous step. If fault i is present in the previous time step it will remain:

$$\Pr(f_i^t | f_i^{t-1}) = 1. \tag{2}$$

If there is no fault present in the previous time step, the probability is the cumulative probability of the exponential distribution with parameter $\lambda_{\mathsf{fail},i}$ at Δ_t:

$$\Pr(f_i^t | \neg f_i^{t-1}) = 1 - e^{-\lambda_{\mathsf{fail},i}\Delta_t}. \tag{3}$$

When fault occurrence is modeled as a homogenous Poisson-process we can treat several consecutive operation events as one with the distances summed. If another operation event occurs between time steps t and $t+1$ with duration Δ_{t+1}, then

$$\Pr(f_i^{t+1} | \neg f_i^{t-1}) = 1 - e^{-\lambda_{\mathsf{fail},i}(\Delta_t + \Delta_{t+1})}. \tag{4}$$

This makes it possible to create discrete operation events from the continuous process of operating the vehicle.

We can show that a constant failure rate is a reasonable assumption for the fuel injection system in the application example using data of fault occurrences labeled with the mileage when the failure occurred that has been collected during the warranty period. Using a central moving average we can estimate the failure rate at mileage m as $\hat{\lambda}(m) = n_f(m)/20000 n_v(m)$ where $n_f(m)$ is the number of faults of any type that have occurred on the fuel injection system in the mileage interval $m \pm 10000$ kilometers and $n_v(m)$ is the number of vehicles still contributing to the data at mileage m. Figure 3 shows a comparison of the estimated failure rate with a constant failure rate $\bar{\lambda}$ estimated using the same data set where $\bar{\lambda} = n_{\mathsf{tot}}/m_{\mathsf{tot}}$, n_{tot} is the total number of fault occurrences in the data set, and m_{tot} is

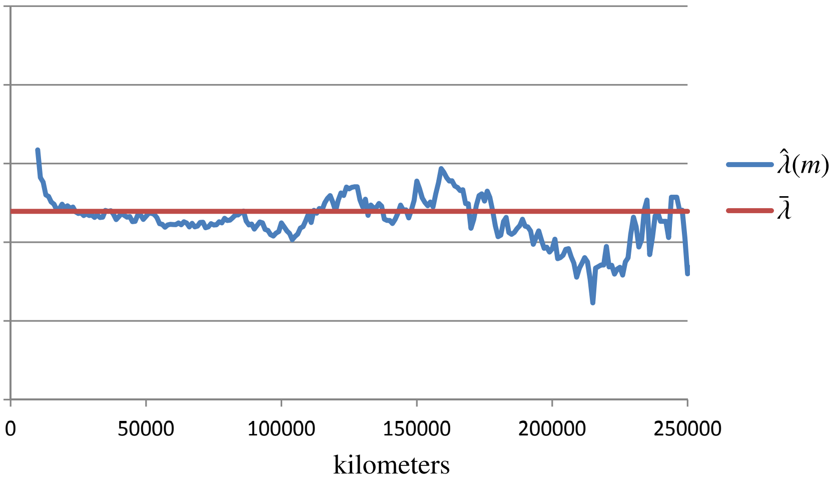

Figure 3: Central moving average estimate of the failure rate compared to a constant failure rate for the fuel injection system.

the total mileage of all vehicles contributing to the data. The uncertainty in the estimate increases with the mileage because the number of vehicles contributing to the data decreases, but we see that there is some infant mortality and then the failure rate remains near the constant value.

CPDs of Symptom-Cause Variables after Operation Events

Faults do not necessarily cause symptoms deterministically, e.g. a leakage in an injector does not always cause too low cylinder output effect. We model the probability that a fault i initially causes a symptom j with a *symptom-cause probability* parameter $p_{i,j}$. If we know that fault i is causing symptom j at one time step it is reasonable to believe that the probability that fault i is causing j is larger than $p_{i,j}$ in the next time step, especially if the distance between the time steps is small. However, when the distance between time steps is large it is unreasonable to believe that fault i must still be causing symptom j only because it did so in the past. We model this rate of change of symptom causes with a *symptom-cause rate* parameter $\lambda_{i,j}$ that can be interpreted as the number of times per million kilometers a change event happens that (just as when the fault initially occurred) sets the symptom cause to active with probability $p_{i,j}$ or to non-active with probability $1-p_{i,j}$. These events are not observable and therefore not modeled directly with their own time steps. Instead the probability of such an event occurring for fault i and symptom j between time steps is modeled using an exponential distribution with parameter $\lambda_{i,j}$.

The symptom-cause variable $S_{i,j}^t$ at time step t for a fault i and one of the symptoms j that it can cause has a CPD that is modeled as follows. If fault i is not present, then no associated symptom-cause can be active:

$$\Pr(s_{i,j}^t | \neg f_j^t) = 0. \tag{5}$$

When the fault occurs sometime between time steps $t-1$ and t, the symptom cause becomes active with probability $p_{i,j}$:

$$\Pr(s_{i,j}^t | \neg f_j^{t-1}, f_j^t) = p_{i,j}. \tag{6}$$

If the fault is present at time step $t-1$ there is a probability of $1-e^{-\lambda_{i,j}\Delta_t}$ that at least one change event has occurred between $t-1$ and t. If the symptom cause is not active at time step $t-1$, then it will become active with probability $p_{i,j}$ if at least one change event occurs:

$$\Pr(s_{i,j}^t | f_j^{t-1}, \neg s_{i,j}^{t-1}) = p_{i,j}(1-e^{-\lambda_{i,j}\Delta_t}). \tag{7}$$

If the symptom cause is already active at time step $t-1$, then we add the probability of having no change event in this time to Eq. (7):

$$\Pr(s_{i,j}^t | f_j^{t-1}, s_{i,j}^{t-1}) = e^{-\lambda_{i,j}\Delta_t} + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_t}). \tag{8}$$

This model for the CPD is chosen so that the transition probabilities are homogenous. This is a realistic assumption as seen in Figure 3 and it is important in order to make consecutive operation events equivalent to a single operation event with combined duration.

Theorem 1 (Symptom-Cause CPD for Consecutive Operation Events)**.** *If two consecutive operation events occur at time steps $t-1$ and t with durations Δ_t and Δ_{t+1}, the conditional probability distributions of a symptom-cause variable $S_{i,j}^{t+1}$ given the values of F_i^{t-1} and $S_{i,j}^{t-1}$ are identical to Eqs. (5)–(8) after a single operation event with duration $\Delta_t + \Delta_{t+1}$:*

$$\Pr(s_{i,j}^{t+1} | \neg f_i^{t+1}) = 0 \tag{9}$$

$$\Pr(s_{i,j}^{t+1} | \neg f_i^{t-1}, f_i^{t+1}) = p_{i,j} \tag{10}$$

$$\Pr(s_{i,j}^{t+1} | f_i^{t-1}, \neg s_i^{t-1}) = p_{i,j}(1-e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})}) \tag{11}$$

$$\Pr(s_{i,j}^{t+1} | f_i^{t-1}, s_i^{t-1}) = e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})} + p_{i,j}(1-e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})}) \tag{12}$$

Proof. See A.1 □

If the symptom-cause behavior of a particular fault is non-intermittent, meaning that when that fault initially causes a symptom it will always keep causing that symptom or when that fault initially does not cause a symptom it will never cause that symptom. Then $\lambda_{i,j} = 0$ and Eqs. (7) and (8) can be simplified:

$$\begin{aligned}\Pr(s^t_{i,j}|f^{t-1}_j, s^{t-1}_{i,j}) &= 1 \\ \Pr(s^t_{i,j}|f^{t-1}_j, \neg s^{t-1}_{i,j}) &= 0.\end{aligned}$$

When $\lambda_{i,j} > 0$ and the duration of the operation event approaches infinity, the probability of having the symptom-cause given the fault approaches $p_{i,j}$ regardless if the symptom-cause is known to be active or not in the previous time step:

$$\lim_{\Delta_t \to \infty} \Pr(s^t_{i,j}|f^{t-1}_j, s^{t-1}_{i,j}) = \lim_{\Delta_t \to \infty} \Pr(s^t_{i,j}|f^{t-1}_j, \neg s^{t-1}_{i,j}) = p_{i,j}.$$

5.5 CPDs for Repair Events

When a repair event occurs for a fault i at time step t, the associated fault variable is set to not present and all of its symptom-cause variables are set to not active, i.e.:

$$\begin{aligned}\Pr(f^t_i) &= 0 && (13) \\ \Pr(s^t_{i,j}) &= 0 \text{ for all } j \in \mathcal{I}_{\mathsf{S}|i}. && (14)\end{aligned}$$

Repair events are treated as being instantaneous because their durations are typically short in comparison to normal operation events. We can therefore assume that no new faults will occur and no symptom-cause variables will have time to change, i.e. for all $k \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}$:

$$\begin{aligned}\Pr(f^t_k|f^{t-1}_k) &= 1 \\ \Pr(f^t_k|\neg f^{t-1}_k) &= 0\end{aligned}$$

and all $j \in \mathcal{I}_{\mathsf{S}|k}$

$$\begin{aligned}\Pr(s^t_{k,j}|s^{t-1}_{k,j}) &= 1 \\ \Pr(s^t_{k,j}|\neg s^{t-1}_{k,j}) &= 0.\end{aligned}$$

When a component is repaired this is treated as multiple repair events occurring in sequence on all of its possible faults. Depending on the inference method used, these events are processed all at once or sequentially.

5.6 CPDs for Observation Events

When an observation event occurs for an observation i at time step t, we get to know the value of the observation variable O^t_i. The value of this variable is often deterministically dependent on the current behavior of the vehicle, i.e. which symptoms are present at the same time. However, it is also possible that the value of O^t_i is non-deterministic given the values of the symptoms, e.g. because of erroneous perception by the external observer due to noisy measurements.

An observation variable has a single parent variable in the network representing the value that the observation attempts to observe. This parent variable can either be a symptom variable or a logical variable. The non-deterministic property of an observation variable is modeled with two parameters, a *false positive probability* parameter $p_{+,i}$ and a *false negative probability* $p_{-,i}$. The probability of observing $O^t_i = \top$ given the value of its parent variable X^t_j is:

$$\Pr(o^t_i | x^t_j) = 1 - p_{-,i} \tag{15}$$

$$\Pr(o^t_i | \neg x^t_j) = p_{+,i}. \tag{16}$$

When X^t_j is a logical variable L^t_j, we need the CPD for L^t_j given its parent symptom variables S^t_k for all $k \in \mathcal{I}_{\mathsf{X}|j}$. Let ϕ_j be the boolean function of the values of the parent symptom variables that describes the logical dependency j. Then

$$\Pr(l^t_j | S^t_k = v_k \, \forall k \in \mathcal{I}_{\mathsf{X}|j}) = \begin{cases} 1 & \text{if } \phi_j((v_k : k \in \mathcal{I}_{\mathsf{X}|j})) = \top \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

We also need the CPD:s for symptom variables given the symptom-cause variables. For a symptom variable S^t_k, the CPD is such that symptom k is present if any symptom cause of a fault that can cause it is active, i.e.:

$$\Pr(s^t_k | S_{j,k} = v_j \, \forall j \in \mathcal{I}_k) = \begin{cases} 1 & \text{if } \bigvee_{j \in \mathcal{I}_k} v_j = \top, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

where $\mathcal{I}_k = \{j \in \mathcal{I}_{\mathsf{F}} : k \in \mathcal{I}_{\mathsf{F}|j}\}$. This definition of the CPD implies that the probability of the symptom being present given the values of all the fault variables is equivalent to a noisy-or distribution [14] where the strength parameters are the probabilities of the symptom-causes given the faults.

An observation event does not change the value of any persistent variable, i.e. for all $k \in \mathcal{I}_{\mathsf{F}}$:

$$\Pr(f^t_k | f^{t-1}_k) = 1$$
$$\Pr(f^t_k | \neg f^{t-1}_k) = 0$$

and all $j \in \mathcal{I}_{\mathsf{S}|k}$

$$\Pr(s^t_{k,j} | s^{t-1}_{k,j}) = 1$$
$$\Pr(s^t_{k,j} | \neg s^{t-1}_{k,j}) = 0.$$

5.7 Example

Figure 4 shows the non-stationary DBN rolled out for the small example for a sequence of four events $e_1, \ldots, e_4$, where e_1 is an operation event of 90000 km, e_2 is a positive observation event for the DTC, e_3 is a positive observation event for the cylinder output test, and e_4 is a repair event for the HPP. Only variables that

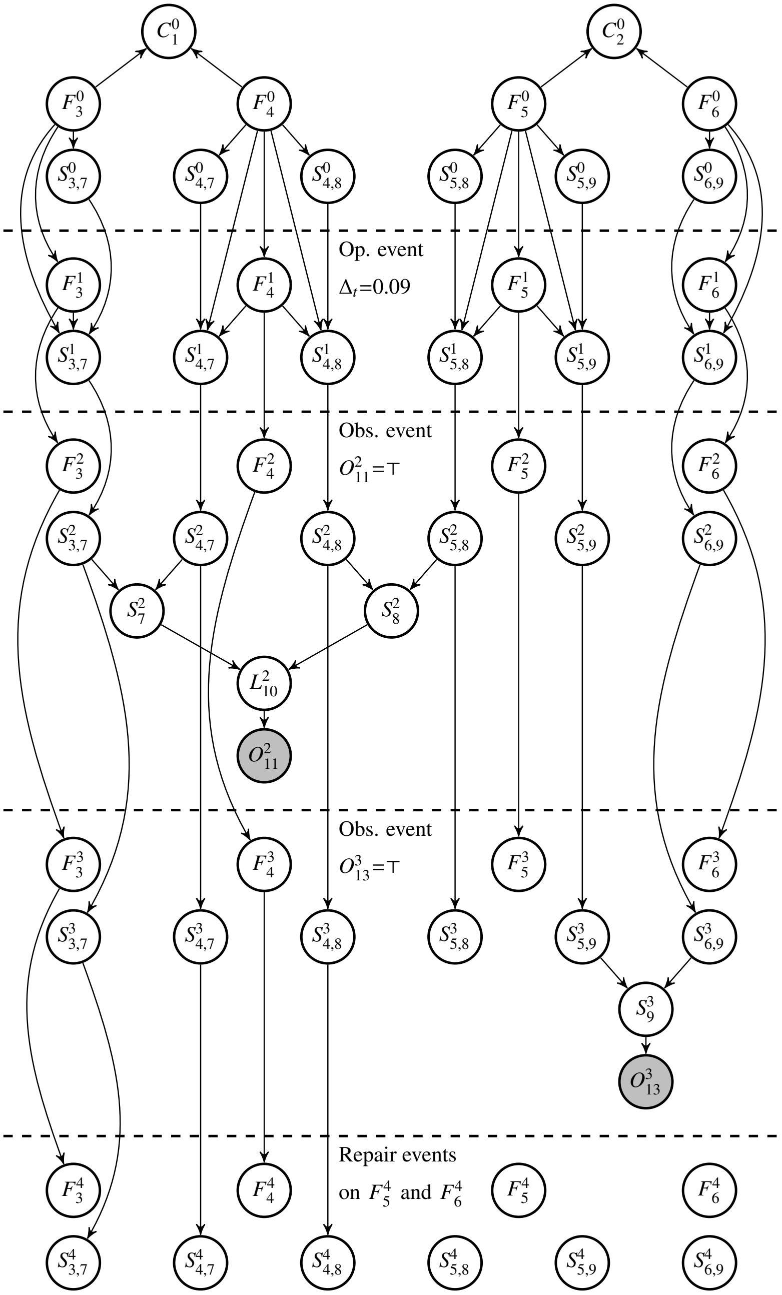

Figure 4: The Bayesian network for the small example rolled out for four events.

t	0	1	2	3	4
$\Pr(f_3^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	2.7×10^{-3}	2.7×10^{-3}
$\Pr(f_4^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.34	2.7×10^{-3}	2.7×10^{-3}
$\Pr(f_5^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	1.0	0
$\Pr(f_6^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	9.1×10^{-4}	5.2×10^{-3}	0
$\Pr(o_{11}^t \vert \mathbf{e}_{1:t})$	2.9×10^{-5}	2.6×10^{-3}	1.0	1.0	5.4×10^{-3}
$\Pr(o_{12}^t \vert \mathbf{e}_{1:t})$	2.9×10^{-5}	1.7×10^{-3}	0.66	1.0	2.7×10^{-3}
$\Pr(o_{13}^t \vert \mathbf{e}_{1:t})$	2.4×10^{-5}	1.3×10^{-3}	0.17	1.0	1.0×10^{-5}

Table 2: The conditional probabilities of selected variables in the small example after events $\mathbf{e}_{1:t} = e_1, \dots, e_t$.

are relevant for inference are shown in the figure. A symptom or logical variable is only relevant for inference if it is an ancestor of a queried observation variable. A component variable is only relevant if its probability is queried by a user. This can occur at any time, but in the figure it is assumed to occur at time 0.

Table 2 shows the conditional probabilities of faults and observations given the events when the parameters for the small example are as follows:

$$\begin{aligned}
p_{\mathsf{fail},i} &= 10^{-5} && \forall i \in \{3,4,5,6\}\\
\lambda_{\mathsf{fail},i} &= 0.01 && \forall i \in \{3,4,5,6\}\\
p_{i,j} &= 0.95 && \forall (i,j) \in \{(3,7),(4,8),(5,8),(6,9)\}\\
p_{i,j} &= 0.5 && \forall (i,j) \in \{(4,7),(5,9)\}\\
\lambda_{i,j} &= 10^4 && \forall (i,j) \in \{(3,7),(4,7),(4,8),(5,8),(5,9),(6,9)\}\\
\phi_{10}(S_7^t, S_8^t) &= S_7^t \vee S_8^t
\end{aligned}$$

5.8 Modularization

To get a model that is manageable and handles the modularity of the vehicles well, the diagnostic model is partitioned into modules with one module for each component and observation.

A *component module* for a component i specifies which faults the component can have, i.e. the set $\mathcal{I}_{\mathsf{F}|i}$. For each fault $j \in \mathcal{I}_{\mathsf{F}|i}$, the module also specifies the parameters $p_{\mathsf{fail},j}$ and $\lambda_{\mathsf{fail},j}$ and which symptoms this fault can cause, i.e. the set $\mathcal{I}_{\mathsf{S}|j}$. For each symptom $k \in \mathcal{I}_{\mathsf{S}|j}$, the module specifies the parameters $p_{j,k}$ and $\lambda_{j,k}$.

An *observation module* for an observation i specifies the parameters $p_{+,i}$ and $p_{-,i}$ and the set $\mathcal{I}_{\mathsf{X}|i}$. If $\mathcal{I}_{\mathsf{X}|i} = \{j\}$ where j is a logical dependency, then the observation module also specifies this dependency to the symptoms, i.e. the boolean function ϕ_j and the set $\mathcal{I}_{\mathsf{X}|j}$.

For a vehicle v consisting of components $\mathcal{I}_{\mathsf{C}|v} \subseteq \mathcal{I}_{\mathsf{C}}$, a diagnostic model is compiled by joining the component modules of each component $i \in \mathcal{I}_{\mathsf{C}|v}$ and the observation modules of each observation j that contains a symptom that also exists in any of the selected component modules. There can be symptoms in the selected observation modules that are not in the selected component modules. These are treated as if always not present.

6 Diagnostic Inference

To compute probabilities of faults and observation outcomes we need to make inference in a dynamic Bayesian network. For larger diagnostic models, exact methods such as the frontier algorithm [47] or the interface algorithm [21] are not feasible. To get tractable inference, we will have to approximate.

Approximate inference algorithms based on Monte-Carlo methods such as EPIS sampling [45] and particle filtering [22] are problematic for making inference in this diagnostic model because the probability of faults occurring is small. A Monte-Carlo based algorithm would have to use many samples to be able to compute probabilities of these rare events accurately.

Another approximate inference algorithm that is not based on sampling is the Boyen-Koller algorithm [2]. In this section we will describe three different approaches to make inference in the diagnostic model based on the Boyen-Koller algorithm. The first method is a direct application of the Boyen-Koller algorithm. For the second method we make an assumption that at most one fault can be present. The third method is a new method that combines the strengths of the first two methods.

6.1 Diagnoser using Boyen-Koller Algorithm – The Independent Fault Model

The Boyen-Koller algorithm is a frontier algorithm that uses a joint probability distribution over the persistent variables in a time step t to compute the probability distribution over the variables in the next time step $t+1$. In the worst case, the space complexity of exactly representing a joint probability distribution of discrete random variables is exponential in the number of variables. The Boyen-Koller algorithm approximates this distribution by grouping the persistent variables into clusters and assuming the clusters to be independent. To compute the posterior probability distributions of the clusters in the next time step $t+1$ given new evidence in time step $t+1$, the Boyen-Koller algorithm makes inference in a static Bayesian network containing only the variables in time steps t and $t+1$ where the joint probability distribution of the persistent variables in time step t is this approximate distribution. The algorithm can use any standard inference algorithm for static Bayesian networks to do this computation. The clusters can also be selected freely and they do not necessarily have to be completely independent, e.g. it is possible to have conditionally independent clusters [3].

The size of the joint probability distribution grows exponentially with the number of variables included in the largest cluster. When we apply the Boyen-Koller algorithm to the diagnostic model, we form a separate cluster for each fault and a separate cluster for each symptom-cause. We let each symptom-cause cluster be conditioned on the corresponding fault variable to ensure that a symptom cause cannot be present unless the corresponding fault is present at the same time. With this clustering, the size of the joint probability distribution is linear in the number of faults and symptom causes. We call this model the *independent fault model*.

Independent Fault State

The marginalized probabilities for each cluster are stored as intermediate results in the diagnostic state. When computing the posterior distribution after an event, we will define the prior distribution to be the joint probability distribution of the independent clusters. The posterior distribution can then be computed exactly given the prior distribution. However, the prior distribution is an approximation and errors from this approximation are propagated when the posterior becomes the next prior.

Definition 1 (Independent Fault State)**.** An *independent fault state* is a diagnostic state s where the intermediate results consist of the values $p_{\mathsf{f}}(s,i)$ for all $i \in \mathcal{I}_{\mathsf{F}}$ representing the probability of having fault i, and the values $p_{\mathsf{s|f}}(s,i,j)$ for all $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}$ representing the probability that fault i causes symptom j given that fault i is present.

The prior distribution of the persistent variables in the time step t represented by the state s is defined to be:

$$\Pr(F_i^t, S_{i,j}^t \; \forall i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}) = \prod_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(F_i^t) \prod_{j \in \mathcal{I}_{\mathsf{S}|i}} \Pr(S_{i,j}^t | F_i^t)$$

where

$$\begin{aligned} \Pr(f_i^t) &= p_{\mathsf{f}}(s,i), \\ \Pr(s_{i,j}^t | f_i^t) &= p_{\mathsf{s|f}}(s,i,j), \\ \Pr(s_{i,j}^t | \neg f_i^t) &= 0. \end{aligned}$$

Transitions

After an event e at time t, we apply the model as it is described in Section 5 to compute the posterior distribution at time $t+1$ given the prior distribution defined by the previous state s. The values of the next state s' are then set to be:

$$\begin{aligned} p_{\mathsf{f}}(s',i) &= \Pr(f_i^{t+1}), \\ p_{\mathsf{s}|\mathsf{f}}(s',i,j) &= \Pr(s_{i,j}^{t+1}|f_i^{t+1}), \end{aligned}$$

for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$.

Operation Events If e is an operation event with duration Δ we can compute the next state by direct application of Eqs. (2)–(8):

$$p_{\mathsf{f}}(s',i) = 1 - e^{-\lambda_{\mathsf{fail},i}\Delta}(1 - p_{\mathsf{f}}(s,i)), \tag{19}$$

$$p_{\mathsf{s}|\mathsf{f}}(s',i,j) = p_{i,j} + \frac{p_{\mathsf{f}}(s,i)e^{-\lambda_{i,j}\Delta}(p_{\mathsf{s}|\mathsf{f}}(s,i,j) - p_{i,j})}{p_{\mathsf{f}}(s',i)}, \tag{20}$$

for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$.

Repair Events If e is a repair event on fault k, we will apply Eqs. (13) and (14):

$$p_{\mathsf{f}}(s',i) = \begin{cases} 0 & \text{if } i = k, \\ p_{\mathsf{f}}(s,i) & \text{if } i \neq k, \end{cases} \tag{21}$$

$$p_{\mathsf{s}|\mathsf{f}}(s',i,j) = \begin{cases} 0 & \text{if } i = k, \\ p_{\mathsf{s}|\mathsf{f}}(s,i,j) & \text{if } i \neq k, \end{cases}$$

for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$.

Observation Events If e is an observation event on observation k with outcome v, we also need to compute the outcome probability $\Pr(e|s)$. Because observation events do not change the values of persistent variables, we can compute the next state and the outcome probability by making the following queries in a Bayesian network containing only variables at time step t:

$$p_{\mathsf{f}}(s',i) = \Pr(f_i^t|O_k^t = v) \tag{22}$$

$$p_{\mathsf{s}|\mathsf{f}}(s',i,j) = \Pr(s_{i,j}^t|f_i^t, O_k^t = v) \tag{23}$$

$$\Pr(e|s) = \Pr(O_k^t = v). \tag{24}$$

for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$. The CPDs for the persistent variables are defined by the previous state s and the CPDs of for the transient variables are given by Eqs. (15)–(18).

If we make these queries using an exact inference algorithm such as the Junction Tree algorithm [19] the complexity is exponential in the tree-width of the static network. The worst-case tree-width of this network is the number of symptom-cause variables whose value the observation is dependent on. In many cases the Junction Tree algorithm is tractable for our application because the number of symptom-cause variables that a single observation variable depends on typically does not grow with the number of components in the model. However, if the model has certain structure we can make even faster inference.

Exploiting Structure

If the parent of an observation variable is a symptom variable or a logical variable where the boolean function is a disjunction, then for observation events on this variable we can use a different inference algorithm that is linear in the number of faults and symptom-causes. We achieve this by adapting the Quickscore algorithm [12] for our diagnostic model with persistent symptom-causes.

The Quickscore algorithm can make exact inference in two-layer Bayesian networks with binary variables where the CPDs for the variables in the bottom layer are of the type Noisy-Or [14] and the variables in the top layer are independent. Let X be a variable in the bottom layer with parent variables $Y_1, \ldots, Y_n$. The Noisy-Or CPD of X is such that $X = \top$ can independently be caused with probability p_i by any of its parent variables Y_i that has the value $\top$:

$$\Pr(x|Y_1 = v_1, \ldots, Y_n = v_n) = 1 - \prod_{i=1}^{n} \Pr(y_i|Y_1 = v_1, \ldots, Y_n = v_n)(1 - p_i).$$

The Quickscore algorithm computes the probability of X using:

$$\Pr(x) = 1 - \prod_{i=1}^{n} \big(\Pr(y_i)(1 - p_i) + \Pr(\neg y_i)\big) \tag{25}$$

and the probability of a parent variable Y_i given that X has value v using:

$$\Pr(y_i|X = v) = \frac{\Pr(X = v|y_i)\Pr(y_i)}{\Pr(X = v)}$$

where

$$\Pr(x|y_i) = 1 - \Pr(\neg x)\frac{1 - p_i}{\Pr(y_i)(1 - p_i) + \Pr(\neg y_i)}. \tag{26}$$

In a diagnostic model where observation i only has disjunctive dependencies to the symptom-causes, we get a distribution that is similar to a Noisy-Or distribution where the bottom-layer variable X corresponds to the observation variable and the top-layer variables correspond to the fault variables. However, our model differs from the two-layer model described above because we have symptom-cause, symptom, and logic variables in intermediate layers. To overcome this, we must modify the algorithm.

Let state s be an independent fault state corresponding to time step t and let e be an observation event at time step t on observation k having outcome v. Let $\mathcal{I}_{\mathsf{X}|k} = \{l\}$ so that X_l^t is the single other variable associated to observation k (see Section 5.1). If we can compute $\Pr(x_l^t)$ and $\Pr(x_l^t|f_i^t)$ for all faults i, then we can find simple expressions for computing the next state s' and outcome probability (Eqs. (22)–(24)). The outcome probability is:

$$\begin{aligned}\Pr(e|s) &= \Pr(O_k^t = v)\\ &= \Pr(O_k^t = v|\neg x_l^t)\Pr(\neg x_l^t) + \Pr(O_k^t = v|x_l^t)\Pr(x_l^t).\end{aligned} \tag{27}$$

The posterior fault probability for fault i in state s' is:

$$\begin{aligned}p_{\mathsf{f}}(s',i) &= \Pr(f_i^t|O_k^t = v)\\ &= \frac{\Pr(O_k^t = v|f_i^t)\Pr(f_i^t)}{\Pr(O_k^t = v)}\\ &= \frac{\Big(\Pr(O_k^t = v|x_l^t)\Pr(x_l^t|f_i^t) + \Pr(O_k^t = v|\neg x_l^t)\Pr(\neg x_l^t|f_i^t)\Big)\Pr(f_i^t)}{\Pr(O_k^t = v)}.\end{aligned} \tag{28}$$

The posterior symptom-cause probability for fault i and symptom j in state s' is:

$$\begin{aligned}p_{\mathsf{s}|\mathsf{f}}(s',i,j) &= \Pr(s_j^t|f_i^t, O_k^t = v)\\ &= \frac{\Pr(O_k^t = v|s_j^t)\Pr(s_i^t|f_i^t)\Pr(f_i^t)}{\Pr(O_k^t = v)\Pr(f_i^t|O_k^t = v)}\\ &= \frac{\Pr(O_k^t = v|x_l^t)\Pr(s_i^t|f_i^t)\Pr(f_i^t)}{\Pr(O_k^t = v)\Pr(f_i^t|O_k^t = v)}\end{aligned} \tag{29}$$

if $j \in \mathcal{R}_{i|k}$ where

$$\mathcal{R}_{i|k} = \begin{cases}\mathcal{I}_{\mathsf{S}|i} \cap \{l\} & \text{if } X_l^t \text{ is a symptom variable,}\\ \mathcal{I}_{\mathsf{S}|i} \cap \mathcal{I}_{\mathsf{X}|l} & \text{if } X_l^t \text{ is a logic variable,}\end{cases}$$

is the set of symptoms that can be caused by fault i and that have a causal dependency to the observation k. If $j \notin \mathcal{R}_{i|k}$, then $\Pr(s_j^t|f_i^t, O_k^t = v) = \Pr(s_j^t|f_i^t)$.

The probabilities $\Pr(\neg x_l^t)$ and $\Pr(\neg x_l^t|f_i^t)$ can then be computed using analogs to Eqs. (25) and (26) where $(1 - p_i)$ is substituted with the probability

$$q_{\mathsf{s}|i} = \prod_{j \in \mathcal{R}_{i|k}} \Pr(\neg s_{i,j}|f_j) \tag{30}$$

```
 1: procedure observationTransition($s, k, v$)
 2:     for $i \in \mathcal{I}_\mathsf{F}$ do $q_{\mathsf{s}|i} \leftarrow \prod_{j \in \mathcal{R}_{i|k}} (1 - p_{\mathsf{s}|\mathsf{f}}(s, i, j))$
 3:     $q_\mathsf{x} \leftarrow \prod_{i \in \mathcal{I}_\mathsf{F}} (p_\mathsf{f}(s, i) q_{\mathsf{s}|i} + 1 - p_\mathsf{f}(s, i))$
 4:     for $i \in \mathcal{I}_\mathsf{F}$ do $q_{\mathsf{x}|i} \leftarrow q_\mathsf{x} \frac{q_{\mathsf{s}|i}}{p_\mathsf{f}(s,i) q_{\mathsf{s}|i} + 1 - p_\mathsf{f}(s,i)}$
 5:     $p_{\mathsf{o}|\neg\mathsf{x}}, p_{\mathsf{o}|\mathsf{x}} \leftarrow \begin{cases} p_{+,k}, (1 - p_{-,k}) & \text{if } v \\ (1 - p_{+,k}), p_{-,k} & \text{otherwise} \end{cases}$
 6:     $p_\mathsf{o} \leftarrow p_{\mathsf{o}|\neg\mathsf{x}} q_\mathsf{x} + p_{\mathsf{o}|\mathsf{x}} (1 - q_\mathsf{x})$
 7:     for $i \in \mathcal{I}_\mathsf{F}$ do
 8:         $p_\mathsf{f}(s', i) \leftarrow (p_{\mathsf{o}|\mathsf{x}} (1 - q_{\mathsf{x}|i}) + p_{\mathsf{o}|\neg\mathsf{x}} q_{\mathsf{x}|i}) \frac{p_\mathsf{f}(s,i)}{p_\mathsf{o}}$
 9:         for $j \in \mathcal{R}_{i|k}$ do $p_{\mathsf{s}|\mathsf{f}}(s', i, j) \leftarrow \frac{p_{\mathsf{o}|\mathsf{x}} p_{\mathsf{s}|\mathsf{f}}(s,i,j) p_\mathsf{f}(s,i)}{p_\mathsf{o} p_\mathsf{f}(s',i)}$
10:         for $j \in \mathcal{I}_{\mathsf{S}|i} \setminus \mathcal{R}_{i|k}$ do $p_{\mathsf{s}|\mathsf{f}}(s', i, j) \leftarrow p_{\mathsf{s}|\mathsf{f}}(s, i, j)$
11:     end for
12:     return $s', p_\mathsf{o}$
13: end procedure
```

Figure 5: The Quickscore Variant algorithm for computing state transitions for observation events.

that no relevant symptom-cause of fault i is active:

$$\Pr(x_l^t) = 1 - \prod_{i \in \mathcal{I}_\mathsf{F}} \left(\Pr(f_i^t) q_{\mathsf{s}|i} + \Pr(\neg f_i^t) \right) \tag{31}$$

$$\Pr(x_l^t | f_i^t) = 1 - \Pr(\neg x_l^t) \frac{q_{\mathsf{s}|i}}{\Pr(f_i^t) q_{\mathsf{s}|i} + \Pr(\neg f_i^t)}. \tag{32}$$

Figure 5 shows the new algorithm, Quickscore Variant, for computing the new diagnostic state s' and the outcome probability $p_\mathsf{o} = \Pr(e|s)$ after an observation event e that corresponds to making observation k and getting the result v. The marginalized posterior probabilities for the observation, the fault probabilities and symptom-cause probabilities are exact given the prior distribution defined by the previous state. On lines 2–4 we compute Eqs. (30)–(32), on lines 5–6 we compute Eq. (27), and on lines 7–11 we compute Eqs. (28)–(29). The complexity of the algorithm is bounded by the number of symptom-causes in the model.

Even though the inference is exact in one time step, the error introduced by creating independent clusters can be unreasonably large over several time steps.

Example 2. Table 3 shows the marginalized fault and observation probabilities for the small example for the same sequence of events as in the example in Section 5.7. The Quickscore Variant is used to compute the observation probabilities at every time step and the fault probabilities at time steps 2 and 3 which occur after the

t	0	1	2	3	4
$\Pr(f_3^t \mid \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	0.33	0.33
$\Pr(f_4^t \mid \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.34	0.34	0.34
$\Pr(f_5^t \mid \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	1.0	0
$\Pr(f_6^t \mid \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	9.1×10^{-4}	5.2×10^{-3}	0
$\Pr(o_{11}^t \mid \mathbf{e}_{1:t})$	2.9×10^{-5}	2.6×10^{-3}	0.7	1.0	0.56
$\Pr(o_{12}^t \mid \mathbf{e}_{1:t})$	2.9×10^{-5}	1.7×10^{-3}	0.55	1.0	0.33
$\Pr(o_{13}^t \mid \mathbf{e}_{1:t})$	2.4×10^{-5}	1.3×10^{-3}	0.17	1.0	1.0×10^{-5}

Table 3: The estimated conditional probabilities of selected variables in the small example after events $\mathbf{e}_{1:t} = e_1, \ldots, e_t$ computed using the independent fault model.

observation events on $O_{11}^2 = \top$ and $O_{13}^3 = \top$ respectively. At time step 0 the fault probabilities are the manufacturing fault probabilities Eq. (1). At time step 1 after the operation event, the fault probabilities are computed using Eq. (19). At time step 4 after the repair event, they are computed using Eq. (21). We can compare these probabilities with the probabilities in Table 2 that are obtained using the exact inference method. After two events, faults 3, 4, and 5 have approximately the same probability of $1/3$ which is correct as they can all explain the first observation. After the second observation, which can be explained by both fault 5 and the unlikely fault 6, we correctly deduce that fault 5 is the most probable cause. However, because of the independence assumption, faults 3 and 4 remain incorrectly at probability $1/3$ because they are independent of the observation variable O_{13}^3.

To create a diagnostic model for which efficient diagnostic inference can be made, the modeler should avoid using complex non-disjunctive dependencies for observations that depend on many symptoms. In some cases, it is possible to model an observation with non-disjunctive dependencies as such without changing the overall behavior of the model. Assume that in the small example we have an additional symptom, "engine not starting", and that to produce a meaningful result, the observation for detecting deviations in the cylinder output effect requires the engine to be starting. We also assume that the user observes whether the engine starts whenever an attempt to perform the test is made. If we should interpret "engine not starting" as a negative test result then we will detect a deviation if and only if

$$\neg\text{"engine not starting"} \wedge \text{"cylinder output effect is too low"}.$$

If we instead should interpret "engine not starting" as a positive test result then we will detect a deviation if and only if

$$\text{"engine not starting"} \vee \neg\text{"cylinder output effect is too low"}$$

which is a disjunction. Both interpretations of "engine not starting" are equivalent in the sense that we cannot draw any conclusions of "cylinder output effect is too low" when "engine not starting" is true.

6.2 Single Fault Model

Another way to get tractable inference is to assume that at most one fault can be present at one time. This is the *single fault model.* The single fault assumption is reasonable when fault probabilities are small and their prior probabilities are approximately independent. This assumption is also applied in many other diagnostic systems [13, 18, 33]. If there can only be at most one fault, then we can accurately represent the joint probability distribution of all faults with linear space complexity. However, it is not reasonable to assume that there can only be at most one symptom present at one time. Because we do not want to increase the space complexity, we will represent the symptom-cause probabilities as before in the independent fault model by assuming conditional independence given the faults.

Single Fault State

To the model we add a persistent discrete random variable Φ^t where $\Phi^t = 0$ represents the case when there are no faults and $\Phi^t = i$ represents the case when only fault i is present. From here on we will use φ_i^t to denote $\Phi^t = i$ and $\neg\varphi_i^t$ to denote $\Phi^t \neq i$.

Definition 2 (Single Fault State)**.** A *single fault state* is a diagnostic state s where the intermediate results consist of the value $p_{\mathsf{NF}}(s)$ representing the probability of having no fault, the value $p_{\varphi}(s,i)$ for all $i \in \mathcal{I}_{\mathsf{F}}$ representing the probability of having only fault i, and the probability $p_{\mathsf{S}|\varphi}(s,i,j)$ for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$ representing the probability that fault i causes symptom j given that fault i is present.

The prior distribution of the persistent variables in the time step t represented by the state s is defined to be:

$$\Pr(F_i^t, S_{i,j}^t \; \forall i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}) \\ = \sum_{k \in \{0\} \cup \mathcal{I}_{\mathsf{F}}} \Pr(\varphi_k^t) \prod_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(F_i^t | \varphi_k^t) \prod_{j \in \mathcal{I}_{\mathsf{S}|i}} \Pr(S_{i,j}^t | \varphi_k^t)$$

where

$$\Pr(f_i^t|\varphi_j^t) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\Pr(\varphi_i^t) = \begin{cases} p_{\mathsf{NF}}(s) & \text{if } i = 0, \\ p_\varphi(s,i) & \text{otherwise,} \end{cases}$$

$$\Pr(s_{i,j}^t|\varphi_k^t) = \begin{cases} p_{\mathsf{s}|\varphi}(s,i,j) & \text{if } i = k, \\ 0 & \text{otherwise.} \end{cases}$$

Transitions

In this model, the fault variables are no longer persistent variables since their values are uniquely determined by the value of Φ^t. Because of the single-fault assumption, we will make exceptions from the model described in Section 5 when computing the posterior distribution after an event. Let e be an event that occurs at time t where the prior distribution is defined by the single fault state s. The values of the next state s' are:

$$\begin{aligned} p_{\mathsf{NF}}(s') &= \Pr(\varphi_0^{t+1}), \\ p_\varphi(s',i) &= \Pr(\varphi_i^{t+1}) \text{ for all } i \in \mathcal{I}_{\mathsf{F}}, \\ p_{\mathsf{s}|\varphi}(s',i,j) &= \Pr(s_{i,j}^{t+1}|\varphi_i^{t+1}) \text{ for all } i \in \mathcal{I}_{\mathsf{F}},\ j \in \mathcal{I}_{\mathsf{S}|i}. \end{aligned}$$

The joint probability distribution of the fault variables in time step 0 differ from that described in Section 5.3:

$$\Pr(\varphi_0^0) = \frac{1}{1 + \sum\limits_{j \in \mathcal{I}_{\mathsf{F}}} \frac{p_{\mathsf{fail},j}}{1 - p_{\mathsf{fail},j}}}$$

$$\Pr(\varphi_i^0) = \frac{\frac{p_{\mathsf{fail},i}}{1 - p_{\mathsf{fail},i}}}{1 + \sum\limits_{j \in \mathcal{I}_{\mathsf{F}}} \frac{p_{\mathsf{fail},j}}{1 - p_{\mathsf{fail},j}}}$$

Operation Events When e is an operation event with duration Δ the posterior is computed differently than before because with the single fault assumption, no more faults can occur during operation once a fault has occurred. When there are no faults we will as before assume that fault occurrence is a homogenous Poisson process. The probability that the next event of n homogenous Poisson processes with parameters $\lambda_1, \ldots, \lambda_n$ occurs in process i is $\lambda_i / \sum_{j=1}^n \lambda_j$. Let $\Lambda_{\mathsf{fail}} = \sum_{i \in \mathcal{I}_{\mathsf{F}}} \lambda_{\mathsf{fail},i}$.

Then the CPD for the variable Φ^{t+1} is:

$$\begin{aligned}
&\Pr(\varphi_0^{t+1}|\varphi_0^t) = e^{-\Lambda_{\mathsf{fail}}\Delta},\\
&\Pr(\varphi_i^{t+1}|\varphi_0^t) = \frac{\lambda_{\mathsf{fail},i}}{\Lambda_{\mathsf{fail}}}(1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \quad \forall i \in \mathcal{I}_{\mathsf{F}},\\
&\Pr(\varphi_i^{t+1}|\varphi_i^t) = 1 \quad \forall i \in \mathcal{I}_{\mathsf{F}},
\end{aligned}$$

where Δ is the duration of the operation event. In the next state s':

$$\begin{aligned}
&p_{\mathsf{NF}}(s') = p_{\mathsf{NF}}(s)e^{-\Lambda_{\mathsf{fail}}\Delta},\\
&p_{\varphi}(s',i) = p_{\varphi}(s,i) + p_{\mathsf{NF}}(s)\frac{\lambda_{\mathsf{fail},i}}{\Lambda_{\mathsf{fail}}}(1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \text{ for all } i \in \mathcal{I}_{\mathsf{F}}.
\end{aligned}$$

For operation events in the single fault model, the CPDs for the symptom-cause variables are as before. We can compute $p_{\mathsf{s}|\varphi}(s',i,j)$ in the same way as we computed $p_{\mathsf{s}|\mathsf{f}}(s',i,j)$ in Eq. (20):

$$p_{\mathsf{s}|\varphi}(s',i,j) = p_{i,j} + \frac{p_{\varphi}(s,i)e^{-\lambda_{i,j}\Delta}(p_{\mathsf{s}|\varphi}(s,i,j)-p_{i,j})}{p_{\varphi}(s',i)}$$

for all $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}$.

Repair Events When e is a repair event on fault i the CPD of Φ^t is:

$$\Pr(\varphi_j^t) = \begin{cases} \Pr(\varphi_0^{t-1}) + \Pr(\varphi_i^{t-1}) & \text{if } j = 0,\\ 0 & \text{if } j = i,\\ \Pr(\varphi_j^{t-1}) & \text{otherwise,} \end{cases}$$

and

$$\begin{aligned}
p_{\mathsf{NF}}(s',j) &= p_{\mathsf{NF}}(s) + p_{\varphi}(s,i),\\
p_{\varphi}(s',j) &= \begin{cases} 0 & \text{if } j = i,\\ p_{\varphi}(s,j) & \text{otherwise.} \end{cases}
\end{aligned}$$

Observation Events If e is an observation event on observation k with outcome v, we make inference in a Bayesian network containing only the variables in time slice t. The query $\Pr(O_k^t = v|\varphi_i)$ can be computed with a worst-case time complexity that is exponential in the number of symptoms $|\mathcal{R}_{i|k}|$ of fault i that O_k^t is dependent on, because all other symptoms are deterministically known to be not present given φ_i. The sought values in the next state are then obtained by applying

t	0	1	2	3	4
$\Pr(f_3^t\|\mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	2.0×10^{-5}	2.0×10^{-5}
$\Pr(f_4^t\|\mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.34	2.1×10^{-5}	2.1×10^{-5}
$\Pr(f_5^t\|\mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	1.0	0
$\Pr(f_6^t\|\mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	3.5×10^{-10}	2.0×10^{-9}	0
$\Pr(o_{11}^t\|\mathbf{e}_{1:t})$	2.9×10^{-5}	2.6×10^{-3}	1.0	1.0	4.0×10^{-5}
$\Pr(o_{12}^t\|\mathbf{e}_{1:t})$	2.9×10^{-5}	1.7×10^{-3}	0.66	1.0	3.0×10^{-5}
$\Pr(o_{13}^t\|\mathbf{e}_{1:t})$	2.4×10^{-5}	1.3×10^{-3}	0.17	1.0	1.0×10^{-5}

Table 4: The estimated conditional probabilities of selected variables in the small example after events $\mathbf{e}_{1:t} = e_1,\dots,e_t$ computed using the single fault model.

Bayes' rule:

$$\Pr(e|s) = \Pr(O_k^t = v) = \sum_{i\in\{0\}\cup\mathcal{I}_{\mathsf{F}}} \Pr(O_k^t = v|\varphi_i^t)\Pr(\varphi_i^t),$$

$$p_{\mathsf{NF}}(s') = \Pr(\varphi_0^t|O_k^t = v) = \frac{\Pr(O_k^t = v|\varphi_0^t)\Pr(\varphi_0^t)}{\Pr(O_k^t = v)},$$

$$p_{\varphi}(s',i) = \Pr(\varphi_i^t|O_k^t = v) = \frac{\Pr(O_k^t = v|\varphi_i^t)\Pr(\varphi_i^t)}{\Pr(O_k^t = v)},$$

$$p_{\mathsf{s}|\varphi}(s',i,j) = \Pr(s_{i,j}^t|\varphi_i^t, O_k^t = v) = \frac{\Pr(O_k^t = v|s_{i,j}^t)\Pr(s_{i,j}^t|\varphi_i^t)}{\Pr(O_k^t = v|\varphi_i^t)}.$$

The size of $\mathcal{R}_{i|k}$ does not depend on the number of faults in the model and therefore the time complexity is linear in the number of faults and symptom-causes.

Example 3. Table 4 shows the marginalized fault and observation probabilities for the small example for the same sequence of events as in the example in Section 5.7. When the single fault model is used, the probabilities for faults 3 and 4 are no longer overestimated when compared to the real values in Table 2. Instead, the fault probabilities of the less likely faults 3, 4, and 6 are underestimated with several orders of magnitude. If the decision support system would decide to end the troubleshooting session after the 4th event, it would estimate the expected cost of doing this to be 100 times smaller than if an exact inference method was used.

6.3 Combined Model

To get around the problems of the single fault assumption and using the independent fault state representation, we propose to combine these methods by representing single faults explicitly and assume the faults to be independent when there are multiple faults.

In this model, the *combined model*, we add several variables to the Bayesian network for representing the prior distribution in a time slice for an event at time t. As before, for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$, the fault variables F_i^t represent whether fault i is present at time t and the symptom-cause variables $S_{i,j}^t$ represent whether fault i is causing j at time t. Similar to the single fault model we let Φ^t be a random variable where $\Phi^t = 0$ represents the case that no faults are present at time t and $\Phi^t = i$ represents the case that only fault i is present at time t. However, we also want Φ^t to represent the case that more than one fault is present at time t by having the value -1.

When $\Phi^t = -1$ we will use the independent fault model to represent the probability distribution over the faults and symptom-causes at time t using variables $\tilde{F}_i^t$ and $\tilde{S}_{i,j}^t$ for all $i \in \mathcal{I}_{\mathsf{F}}$, $j \in \mathcal{I}_{\mathsf{S}|i}$. This distribution is updated after events as described in Section 6.1 independently of how Φ^t is updated. We then combine the single fault model with the independent fault model in the following way. If $\Phi^t \geq 0$, then the value of Φ^t determines the value of the fault variables and symptom-cause variables. Otherwise, if $\Phi^t = -1$, then we let the values of $\tilde{F}_i^t$ and $\tilde{S}_{i,j}^t$ determine the values of F_i^t and $S_{i,j}$ instead.

Definition 3 (Combined State)**.** A *combined state* is a diagnostic state s where the intermediate results consist of the value $p_{\mathsf{NF}}(s)$ representing the probability of having no faults, the values $p_{\varphi}(s,i)$ representing the probability of having only fault i, the value $p_{\mathsf{MF}}(s)$ representing the probability of having multiple faults, the values $p_{\mathsf{s}|\varphi}(s,i,j)$ representing the probability that fault i causes symptom j given that only fault i is present, the values $p_{\mathsf{f}}(s,i)$ representing the probability of having fault i in the independent fault model, and the values $p_{\mathsf{s}|\mathsf{f}}(s,i,j)$ representing the probability that fault i causes symptom j given that fault i is present in the independent fault model. The joint probability distribution of the persistent variables in time slice t is as follows:

$$\Pr(\varphi^t_i) = \begin{cases} p_{\mathsf{MF}}(s) & \text{if } i=-1, \\ p_{\mathsf{NF}}(s) & \text{if } i=0, \\ p_\varphi(s,i) & \text{otherwise,} \end{cases}$$

$$\Pr(\tilde{f}^t_i) = p_{\mathsf{f}}(s,i),$$

$$\Pr(\tilde{s}^t_{i,j}|\tilde{F}^t_i = v) = \begin{cases} p_{\mathsf{s}|\mathsf{f}}(s,i,j) & \text{if } v=\top, \\ 0 & \text{otherwise,} \end{cases}$$

$$\Pr(f^t_i|\varphi^t_j, \tilde{F}^t_i = v) = \begin{cases} 1 & \text{if } j=-1 \text{ and } v=\top, \\ 1 & \text{if } i=j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\Pr(s^t_{i,j}|\varphi^t_k, \tilde{S}^t_{i,j} = v) = \begin{cases} 1 & \text{if } k=-1 \text{ and } v=\top, \\ p_{\mathsf{s}|\varphi}(s,i,j) & \text{if } i=k, \\ 0 & \text{otherwise.} \end{cases}$$

When computing the marginalized probability of a fault or a symptom-cause in a time step t represented by a combined state s we will condition every query with the knowledge that more than one fault is present in the distribution represented by the independent fault variable. We need to do this because we *know* that this model should only be used when there are multiple faults according to Φ^t. Let N^t be a discrete variable representing how many of the variables $\tilde{F}^t_i$ for all $i \in \mathcal{I}_{\mathsf{F}}$ have the value $\top$. Then the marginalized probability of fault i given s is

$$\bar{p}_{\mathsf{f}}(s,i) = \Pr(f^t_i|N^t > 1) = p_\varphi(s,i) + p_{\mathsf{MF}}(s)\Pr(\tilde{f}^t_i|N^t > 1) \tag{33}$$

and the marginalized probability of symptom j being caused by fault i given s is

$$\begin{aligned} \bar{p}_{\mathsf{s}|\mathsf{f}}(s,i,j) &= \Pr(s^t_{i,j}|N^t > 1) \\ &= p_{\mathsf{s}|\varphi}(s,i,j)p_\varphi(s,i) + p_{\mathsf{MF}}(s)\Pr(\tilde{f}^t_i|N^t > 1)p_{\mathsf{s}|\mathsf{f}}(s,i,j). \end{aligned} \tag{34}$$

To compute Eq. (34), we need a computable expression for $\Pr(\tilde{f}^t_i|N^t > 1)$. First we compute the distribution of N^t:

$$\begin{aligned} \Pr(N^t > 1) &= 1 - \Pr(N^t = 0) - \Pr(N^t = 1) \\ \Pr(N^t = 0) &= \prod_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(\neg\tilde{F}^t_i) \ = \prod_{i \in \mathcal{I}_{\mathsf{F}}} (1 - p_{\mathsf{f}}(s,i)) \\ \Pr(N^t = 1) &= \sum_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(\tilde{F}^t_j) \prod_{j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}} \Pr(\neg\tilde{F}^t_j) \ = \ \Pr(N^t = 0) \sum_{i \in \mathcal{I}_{\mathsf{F}}} \frac{p_{\mathsf{f}}(s,i)}{1 - p_{\mathsf{f}}(s,i)}. \end{aligned}$$

Using Bayes' rule, we have

$$\Pr(\tilde{f}^t_i|N^t > 1) = \frac{\Pr(\tilde{f}^t_i)\Pr(N^t > 1|\tilde{f}^t_i)}{\Pr(N^t > 1)}$$

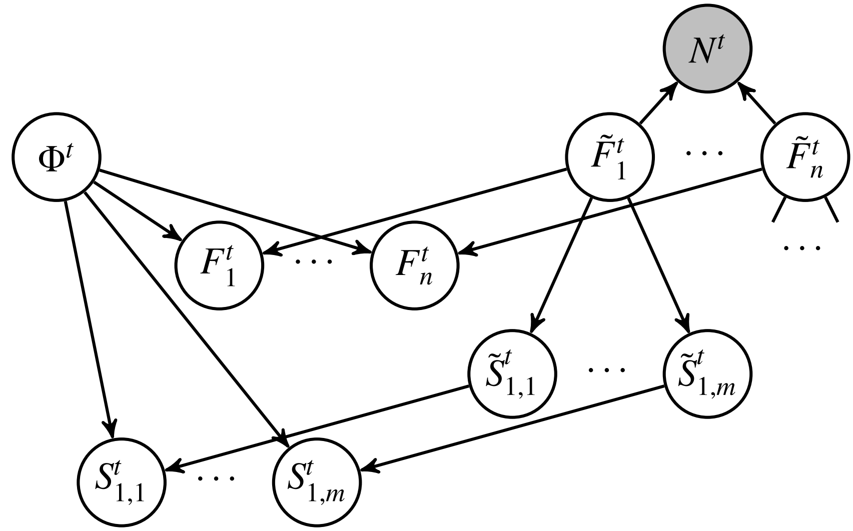

Figure 6: The conditional dependencies of the variables in the combined model.

where we identify that

$$\begin{aligned}
\Pr(N^t > 1|\tilde{f}^t_i) &= 1 - \Pr(N^t \leq 1|\tilde{f}^t_i) = 1 - \Pr(N^t = 1|\tilde{f}^t_i) \\
&= 1 - \frac{\Pr(N^t = 1, \tilde{f}^t_i)}{\Pr(\tilde{f}^t_i)} = 1 - \frac{\Pr(\tilde{f}^t_i) \prod_{j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}} \Pr(\neg \tilde{f}^t_j)}{\Pr(\tilde{f}^t_i)} \\
&= 1 - \frac{\Pr(N^t = 0)}{\Pr(\neg \tilde{f}^t_i)} = 1 - \frac{\Pr(N^t = 0)}{1 - p_{\mathsf{f}}(s,i)}.
\end{aligned}$$

Thus,

$$\Pr(\tilde{f}^t_i|N^t > 1) = p_{\mathsf{f}}(s,i) \frac{1 - \frac{\Pr(N^t = 0)}{1 - p_{\mathsf{f}}(s,i)}}{\Pr(N^t > 1)} \tag{35}$$

If $\Pr(N^t > 1) = 0$ then we substitute $\Pr(\tilde{f}^t_i|N^t > 1)$ with zero in Eqs. (33) and (34). If $p_{\mathsf{f}}(s,i) = 1$ then $\Pr(\tilde{f}^t_i|N^t > 1) = 1$.

The conditional dependencies of the variables in the combined model are shown in Figure 6.

Transitions

Let e be an event that occurs at time t where the prior distribution is defined by the combined state s. When computing the next state s' we treat the independent fault model as a separate process. Therefore, we update the values $p_{\mathsf{f}}(s',i)$ and

$p_{\mathsf{s|f}}(s',i,j)$ as previously described in Section 6.1 for all events. The remaining values are in the next state s':

$$\begin{aligned}
p_{\mathsf{NF}}(s') &= \Pr(\varphi_0^{t+1}|N^t > 1)\\
p_{\varphi}(s',i) &= \Pr(\varphi_i^{t+1}|N^t > 1) \text{ for all } i \in \mathcal{I}_{\mathsf{F}},\\
p_{\mathsf{MF}}(s') &= \Pr(\varphi_{-1}^{t+1}|N^t > 1)\\
p_{\mathsf{s}|\varphi}(s',i,j) &= \Pr(s_{i,j}^{t+1}|\varphi_i^{t+1}, N^t > 1) \text{ for all } i \in \mathcal{I}_{\mathsf{F}},\ j \in \mathcal{I}_{\mathsf{S}|i}.
\end{aligned}$$

We must condition on $N^t > 1$ because we define the prior distribution at time step t such that the independent fault model is only used when there are two or more faults present among $\tilde{F}_i^t$. For certain transitions, the values of Φ^{t+1} and $S_{i,j}^{t+1}$ in the posterior time slice are influenced by the values of $\tilde{F}_i^t$ and $\tilde{S}_{i,j}^t$.

Operation Events When e is an operation event with duration Δ, new faults can be introduced. No faults can become single and multiple faults and single faults can become multiple faults. The probability of having no faults at time $t+1$ is the probability that we had no faults and no new faults occur:

$$p_{\mathsf{NF}}(s') = \Pr(\varphi_0^t) \prod_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(\neg f_i^{t+1}|\neg f_i^t) = p_{\mathsf{NF}}(s) e^{-\Lambda_{\mathsf{fail}}\Delta}$$

where $\Lambda_{\mathsf{fail}} = \sum_{i \in \mathcal{I}_{\mathsf{F}}} \lambda_{\mathsf{fail},i}$.

The probability of having a single fault i at time $t+1$ is the probability that we had no faults and only fault i occurs or that we had fault i and no other faults occur:

$$\begin{aligned}
p_{\varphi}(s',i) &= \Pr(\varphi_0^t)\Pr(f_i^{t+1}|\neg f_i^t) \prod_{j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}} \Pr(\neg f_j^{t+1}|\neg f_j^t)\\
&\quad + \Pr(\varphi_i^t) \prod_{j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}} \Pr(\neg f_j^{t+1}|\neg f_j^t)\\
&= e^{-\Lambda_{\mathsf{fail}}\Delta} \left(\frac{1 - e^{-\lambda_{\mathsf{fail},i}\Delta}}{e^{-\lambda_{\mathsf{fail},i}\Delta}} p_{\mathsf{NF}}(s) + \frac{p_{\varphi}(s,i)}{e^{-\lambda_{\mathsf{fail},i}\Delta}} \right). \qquad (36)
\end{aligned}$$

Because $\sum_{i \in \{-1,0\} \cup \mathcal{I}_{\mathsf{F}}} \Pr(\varphi_i^{t+1}|N^t > 1) = 1$, the probability of having multiple faults at time $t+1$ is

$$p_{\mathsf{MF}}(s') = 1 - p_{\mathsf{NF}}(s') - \sum_{i \in \mathcal{I}_{\mathsf{F}}} p_{\varphi}(s',i). \qquad (37)$$

The symptom-cause probabilities $p_{\mathsf{s}|\varphi}(s,i,j)$ are computed as previously in Eq. (20) for all $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}$:

$$p_{\mathsf{s}|\varphi}(s',i,j) = p_{i,j} + \frac{p_{\varphi}(s,i) e^{-\lambda_{i,j}\Delta}(p_{\mathsf{s}|\varphi}(s,i,j) - p_{i,j})}{p_{\varphi}(s',i)}.$$

In the combined model the transitions for operation events are approximate because the state parameters for the independent fault probabilities are updated independently of Φ^t. However, we can derive a bound on the maximum error on the marginalized posterior fault probabilities after an operation event using the combined model.

Theorem 2 (Error of Operation Event Transitions in the Combined Model)**.** Let e be an operation event with duration Δ that occurs at time step t where the prior distribution is defined by the combined state s. Let $\varepsilon_{\mathsf{NF}}$, ε_i for all $i \in \mathcal{I}_{\mathsf{F}}$, and $\varepsilon_{\mathsf{MF}}$ be the difference between the estimated fault probabilities according to the independent fault model and the single fault model given state s:

$$\varepsilon_{\mathsf{NF}} = p_{\mathsf{NF}}(s) - \Pr(N^t = 0), \tag{38}$$

$$\varepsilon_i = p_{\varphi}(s,i) - \Pr(\tilde{f}_i | N^t = 1) \text{ for all } i \in \mathcal{I}_{\mathsf{F}}, \tag{39}$$

$$\varepsilon_{\mathsf{MF}} = p_{\mathsf{MF}}(s) - \Pr(N^t > 1). \tag{40}$$

Then the error between the marginalized fault probability $\bar{p}_{\mathsf{f}}(s',i)$ for fault i in the next state s' and the exact posterior probability $\Pr(f_i^{t+1} | N^t > 1)$ is:

$$\begin{aligned} \left|\bar{p}_{\mathsf{f}}(s',i) - \Pr(f_i^{t+1}|N^t>1)\right| &= \left|\Pr(f_i^{t+1}|N^{t+1}>1) - \Pr(f_i^{t+1}|N^t>1)\right| \\ &\leq \left|\Pr(\tilde{f}_i^{t-1}|N^{t-1}>1) - \Pr(\tilde{f}_i^t|N^t>1)\right| |\varepsilon_{\mathsf{MF}}| \\ &\quad + (1-e^{-\Lambda_{\mathsf{fail}}\Delta})\Big(2|\varepsilon_{\mathsf{NF}}| + |\varepsilon_{\mathsf{MF}}| + |\varepsilon_i| + \sum_{j\in\mathcal{I}_{\mathsf{F}}} |\varepsilon_j|\Big). \end{aligned} \tag{41}$$

Proof. See A.2. □

The bound is small when the no-fault, single-fault, or multiple-fault probabilities according to the independent fault model are similar to the values $p_{\mathsf{NF}}(s)$, $p_{\varphi}(s,i)$, and $p_{\mathsf{MF}}(s)$. The bound grows with the duration of the event and the probabilities of new faults occurring.

Repair Events If e is a repair event on fault i, then a multiple fault consisting of fault i and exactly one other fault will become a single fault. Therefore the appropriate amount of probability mass needs to be transferred from the multiple fault value $p_{\mathsf{MF}}(s)$ to the single-fault values $p_{\varphi}(s',j)$ of all other faults j. The proportion of $p_{\mathsf{MF}}(s)$ that should be transferred to $p_{\varphi}(s',j)$ is given by the independent fault state. Let probability $d_{i,j}$ be the probability that exactly two faults, fault i and fault j, are present at time t given the distribution defined by the combined state s where $N^t > 1$. When there are exactly two faults in this distribution, then $\Phi^t = -1$

and $N^t = 2$.

$$\begin{aligned}
d_{i,j} &= \Pr(f_i^t, f_j^t, \varphi_{-1}^t, N^t = 2 | N^t > 1) \\
&= \Pr(\varphi_{-1}^t | N^t > 1) \Pr(\tilde{f}_i^t, \tilde{f}_j^t, N^t = 2 | N^t > 1) \\
&= \Pr(\varphi_{-1}^t | N^t > 1) \frac{\Pr(\tilde{f}_i^t) \Pr(\tilde{f}_j^t) \prod_{k \in \mathcal{I}_{\mathsf{F}} \setminus \{i,j\}} \Pr(\neg \tilde{f}_k^t)}{\Pr(N^t > 1)} \\
&= p_{\mathsf{MF}}(s) \frac{p_{\mathsf{f}}(s,i) p_{\mathsf{f}}(s,j) \Pr(N^t = 0)}{(1 - p_{\mathsf{f}}(s,i))(1 - p_{\mathsf{f}}(s,j)) \Pr(N^t > 1)}.
\end{aligned}$$

This probability is then added to the single fault probability of fault j in the next state

$$p_\varphi(s', j) = \begin{cases} 0 & \text{if } j = i \\ p_\varphi(s,j) + d_{i,j} & \text{if } j \neq i \end{cases}$$

and removed from the multiple fault probability:

$$p_{\mathsf{MF}}(s') = p_{\mathsf{MF}}(s) - \sum_{j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}} d_{i,j}.$$

If no fault or only fault i is present at time t then we have no faults at time $t+1$:

$$p_{\mathsf{NF}}(s') = p_{\mathsf{NF}}(s) + p_\varphi(s,i).$$

The single fault symptom-cause probability $p_{\mathsf{s}|\varphi}(s', j, k)$ for each unrepaired fault $j \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}$ and symptom $k \in \mathcal{I}_{\mathsf{S}|j}$ also needs to be updated. This probability is the probability of having the symptom cause and the single fault j plus the probability of having the symptom cause and the double fault i and j in the previous time step:

$$\begin{aligned}
\Pr(s_{j,k}^{t+1}, \varphi_j^{t+1} | N^t > 1) = {} & \Pr(s_{j,k}^t | \varphi_j^t) \Pr(\varphi_j^t) \\
& + \Pr(\tilde{s}_{j,k}^t | \tilde{f}_j^t) \Pr(f_i^t, f_j^t, \varphi_{-1}^t, N^t = 2 | N^t > 1)
\end{aligned}$$

thus

$$p_{\mathsf{s}|\varphi}(s', j, k) = \frac{p_{\mathsf{s}|\varphi}(s,j,k) p_\varphi(s,j) + p_{\mathsf{s}|\mathsf{f}}(s,j,k) d_{i,j}}{p_\varphi(s,j) + d_{i,j}}.$$

Probability mass is properly transferred from the multiple fault case to the single fault case which makes the repair transition exact in the combined model.

Observation Events When e is an observation event on observation k with outcome v we need to know $\Pr(O_k^t = v | \varphi_i, N^t > 0)$ for all $i \in \mathcal{I}_{\mathsf{F}} \cup \{-1, 0\}$ and $\Pr(O_k^t = v, s_{i,j}^t | \varphi_i, N^t > 0)$ for all $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}$ in order to compute

$$\begin{aligned}
\Pr(e|s) &= \Pr(O_k^t = v | N^t > 1) \\
&= \sum_{l \in \mathcal{I}_{\mathsf{F}} \cup \{-1,0\}} \Pr(O_k^t = v | \varphi_l^t, N^t > 1) \Pr(\varphi_l^t | N^t > 1) \\
p_{\mathsf{NF}}(s') &= \Pr(\varphi_0^t | O_k^t = v, N^t > 1) \\
&= \frac{\Pr(O_k^t = v | \varphi_0^t, N^t > 1) \Pr(\varphi_0^t | N^t > 1)}{\Pr(O_k^t = v | N^t > 1)} \\
p_{\varphi}(s', i) &= \Pr(\varphi_i^t | O_k^t = v, N^t > 1) \\
&= \frac{\Pr(O_k^t = v | \varphi_i^t, N^t > 1) \Pr(\varphi_i^t | N^t > 1)}{\Pr(O_k^t = v | N^t > 1)} \\
p_{\mathsf{MF}}(s') &= \Pr(\varphi_{-1}^t | O_k^t = v, N^t > 1) \\
&= \frac{\Pr(O_k^t = v | \varphi_{-1}^t, N^t > 1) \Pr(\varphi_{-1}^t | N^t > 1)}{\Pr(O_k^t = v | N^t > 1)} \\
p_{\mathsf{S}|\varphi}(s', i, j) &= \Pr(s_{i,j}^t | \varphi_i^t, O_k^t = v, N^t > 1) \\
&= \frac{\Pr(O_k^t = v, s_{i,j}^t | \varphi_i^t, N^t > 1)}{\Pr(O_k^t = v | \varphi_i^t, N^t > 1)}
\end{aligned}$$

for all $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|i}$.

For the no fault and single fault case (when $i \geq 0$), the desired probabilities

$$\begin{aligned}
\Pr(O_k^t = v | \varphi_i, N^t > 0) &= \Pr(O_k = v | \varphi_i) \\
\Pr(O_k^t = v, s_{i,j}^t | \varphi_i, N^t > 0) &= \Pr(O_k = v, s_{i,j} | \varphi_i)
\end{aligned}$$

can be computed as previously described for the single fault model.

For the multiple fault case we need to formulate the probability $\Pr(O_k^t = v | \varphi_{-1}^t, N^t > 1)$ in terms of the probabilities $\Pr(O_k^t = v | \varphi_{-1}^t, N^t = 0)$, $\Pr(O_k^t = v | f_i, \varphi_{-1}^t, N^t = 1)$, and $\Pr(O_k^t = v | \varphi_{-1}^t)$ which can be computed as before using the single fault and independent fault models respectively:

$$\begin{aligned}
\Pr(O_k^t = v | \varphi_{-1}^t, N^t > 1) &= \frac{\Pr(N^t > 1 | O_k^t = v, \varphi_{-1}^t) \Pr(O_k^t = v | \varphi_{-1}^t)}{\Pr(N^t > 1 | \varphi_{-1}^t)} \\
&= \frac{\Pr(O_k^t = v | \varphi_{-1}^t) - \Pr(O_k^t = v, N^t = 0 | \varphi_{-1}^t) - \Pr(O_k^t = v, N^t = 1 | \varphi_{-1}^t)}{\Pr(N^t > 1)}
\end{aligned}$$

t	0	1	2	3	4
$\Pr(f_3^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	2.4×10^{-3}	2.4×10^{-3}
$\Pr(f_4^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.34	2.4×10^{-3}	2.4×10^{-3}
$\Pr(f_5^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	0.33	1.0	0
$\Pr(f_6^t \vert \mathbf{e}_{1:t})$	1.0×10^{-5}	9.1×10^{-4}	4.6×10^{-6}	3.8×10^{-5}	0
$\Pr(o_{11}^t \vert \mathbf{e}_{1:t})$	2.9×10^{-5}	2.6×10^{-3}	1.0	1.0	4.0×10^{-3}
$\Pr(o_{12}^t \vert \mathbf{e}_{1:t})$	2.9×10^{-5}	1.7×10^{-3}	0.66	1.0	2.4×10^{-5}
$\Pr(o_{13}^t \vert \mathbf{e}_{1:t})$	2.4×10^{-5}	1.3×10^{-3}	0.17	1.0	1.0×10^{-5}

Table 5: The estimated conditional probabilities of selected variables in the small example after events $\mathbf{e}_{1:t} = e_1, \ldots, e_t$ computed using the combined model.

where

$$\begin{aligned}
&\Pr(O_k^t = v, N^t = 0 \mid \varphi_{-1}^t) = \Pr(O_k^t = v \mid \varphi_{-1}^t, N^t = 0)\Pr(N^t = 0)\\
&\Pr(O_k^t = v, N^t = 1 \mid \varphi_{-1}^t)\\
&\qquad = \sum_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(O_k^t = v \mid f_i^t, \varphi_{-1}^t, N^t = 1)\Pr(f_i^t, N^t = 1 \mid \varphi_{-1}^t)\\
&\qquad = \sum_{i \in \mathcal{I}_{\mathsf{F}}} \Pr(O_k^t = v \mid f_i^t, \varphi_{-1}^t, N^t = 1)\frac{\Pr(N^t = 0)\Pr(f_i^t \mid \varphi_{-1}^t)}{\Pr(\neg f_i^t \mid \varphi_{-1}^t)}
\end{aligned}$$

Given the distribution modeled by s, the observation transition in the combined model is exact with regard to the observation outcome probability and the marginals of the fault and symptom-cause probabilities. It can be computed in a linear number of no fault and single fault inferences and one independent fault inference which, when the Quickscore Variant algorithm is applicable, is linear in the number of faults.

Example 4. Table 5 shows the marginalized fault and observation probabilities for the small example for the same sequence of events as in the example in Section 5.7. When the combined model is used, the error on faults 3 and 4 in the independent model is avoided. Also, unlike the single fault model, the underestimation of the less likely faults is avoided.

7 Parameter Learning

We use a Bayesian learning approach to learn the model parameters θ which are considered to be random variables. The expert knowledge defines a prior distribution $p(\theta)$, and then we estimate the expected value of the parameters given data $\mathbf{y}$

collected from the troubleshooting process:

$$E(\theta|\mathbf{y}) = \int \theta \Pr(\mathbf{y}|\theta)p(\theta)\,d\theta.$$

The more data that is available, the less important the prior distribution will be for the posterior distribution. We want the learning process to be tractable and modularizable so that data from different vehicles can be used to learn parameters for shared components. We must also tolerate that the data can be incomplete. The parameters we will learn are the failure rates $\lambda_{\mathsf{fail},i}$ and manufacturing fault probabilities $p_{\mathsf{fail},i}$ for all faults $i \in \mathcal{I}_{\mathsf{F}}$, and the symptom-cause probabilities $p_{i,j}$ for all symptoms $j \in \mathcal{I}_{\mathsf{S}|i}$.

7.1 Prior Distributions

We will use Gamma distributions to model priors for the failure rates and Beta distributions for the other parameters because these have the same support as the domains of the parameters we want to learn and have a physical interpretation. Let

$$\begin{aligned}
\lambda_{\mathsf{fail},i} &\sim \mathrm{Gamma}(\alpha_{\mathsf{fail},i}, \beta_{\mathsf{fail},i}) \\
p_{\mathsf{fail},i} &\sim \mathrm{Beta}(\alpha_{\mathsf{man},i}, \beta_{\mathsf{man},i}) \\
p_{i,j} &\sim \mathrm{Beta}(\alpha_{i,j}, \beta_{i,j})
\end{aligned}$$

for $i \in \mathcal{I}_{\mathsf{F}}, j \in \mathcal{I}_{\mathsf{S}|j}$ where the hyperparameters can be interpreted as follows: on average $\alpha_{\mathsf{fail},i}$ faults of type i will occur in $\beta_{\mathsf{fail},i}$ million kilometers of operation, on average $\alpha_{\mathsf{man},i}$ out of $\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i}$ components have a manufacturing fault of type i, on average $\alpha_{i,j}$ out of $\alpha_{i,j} + \beta_{i,j}$ faults of type i have symptom j. The confidence in the expert knowledge is higher the higher the β-values are.

7.2 Data

Data for learning parameters can be generated from using the decision support system or be collected from other sources. In this work we will use warranty data to parametrize the models. The warranty data is collected from the claims that workshops write in order to get a refund for repairs made during the warranty period. In the claim, the workshop describes the fault and which observations were made at the time. The claim is submitted to the vehicle manufacturer together with the faulty component. This component may then be analyzed by the manufacturer in a laboratory to determine which fault the component has and why it occurred. However, the laboratory analyzes are not always performed and the reports of observations are sometimes incomplete.

Each data sample $y^{[i]} \in \mathbf{y}$ consists of a vehicle identifier $v_{\mathsf{id}}^{[i]}$, the mileage when the fault was reported $m^{[i]}$, an identifier $c_{\mathsf{id}}^{[i]}$ for the faulty component, an identifier $f_{\mathsf{id}}^{[i]}$ for the fault on the component (if the information is missing, $f_{\mathsf{id}}^{[i]} = 0$), a flag $f_{\mathsf{man}}^{[i]}$ specifying whether the claimed fault is a manufacturing fault (only used if the fault is known), and a list of observations $\mathbf{o}^{[i]}$ made at the time (fault codes and other symptoms). Also associated with the data is the mileage of each vehicle contributing to the data at the end of the warranty period and which components each of these vehicles has.

For each fault $i \in \mathcal{I}_{\mathsf{F}|j}$ of component $j \in \mathcal{I}_{\mathsf{C}}$ we extract the number of data samples $n_{\mathsf{man},i}$ where the fault occurred as a manufacturing fault, the number of data samples $n_{\mathsf{fail},i}$ where the fault occurred as a non-manufacturing fault, the number of data samples $n_{\mathsf{miss},j}$ for which fault information is missing, the number of vehicles $n_{\mathsf{comp},j}$ that have the component in the population from which data is collected, and the total mileage $n_{\mathsf{mil},j}$ of all vehicles that have the component in this population.

7.3 Learning Fault Probabilities and Failure Rates

In reality a fault is not necessarily detected as soon as it occurs and may therefore not be reported in the data. However, we will assume that if a vehicle in the population has a manufacturing fault it will be represented in the data and that if a non-manufacturing fault occurs on a vehicle during the warranty period it will also be represented in the data. The data format does not support multiple faults. In the rare case this happens, we would need to insert one claim for each fault in the data. If the assumptions are true and the data is complete, i.e. all claims are analyzed, then the posterior distribution of the fault probabilities and failure rates can be computed in closed form independently for each fault because the priors and posteriors are conjugate.

Lemma 1 (Fault Probability Posterior - All Claims Analyzed)**.** Let $\mathbf{y}^*$ be data where the cause of all claims are analyzed. Then for each fault $i \in \mathcal{I}_{\mathsf{F}|j}$ of component $j \in \mathcal{I}_{\mathsf{C}}$ the expected values of $\lambda_{\mathsf{fail},i}$ and $p_{\mathsf{fail},i}$ given $\mathbf{y}^*$ is:

$$E(p_{\mathsf{fail},i}|\mathbf{y}^*) = \frac{\alpha_{\mathsf{man},i} + n^*_{\mathsf{man},i}}{\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i} + n_{\mathsf{comp},j}} \tag{42}$$

$$E(\lambda_{\mathsf{fail},i}|\mathbf{y}^*) = \frac{\alpha_{\mathsf{fail},i} + n^*_{\mathsf{fail},i}}{\beta_{\mathsf{fail},i} + n_{\mathsf{mil},j}} \tag{43}$$

where $n^*_{\mathsf{man},i}$ and $n^*_{\mathsf{fail},i}$ are the actual number of manufacturing and non-manufacturing faults of type $i \in \mathcal{I}_{\mathsf{F}}$ as seen in the laboratory.

Proof. See A.3. □

Typically not all claims are analyzed and therefore we cannot use Lemma 1 directly. However, to obtain the expected value of the posteriors it is sufficient to know the expected values of $n^*_{\mathsf{man},i}$ and $n^*_{\mathsf{fail},i}$ given the data.

Lemma 2 (Fault Probability Posterior - Incomplete Analysis)**.** For each $i \in \mathcal{I}_{\mathsf{F}|j}$, $j \in \mathcal{I}_{\mathsf{C}}$, the expected values

$$
\begin{aligned}
E(p_{\mathsf{fail},i}|\mathbf{y}) &= \frac{\alpha_{\mathsf{fail},i} + E(n^*_{\mathsf{man},i}|\mathbf{y})}{\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i} + n_{\mathsf{comp},j}} \\
E(\lambda_{\mathsf{fail},i}|\mathbf{y}) &= \frac{\alpha_{\mathsf{fail},i} + E(n^*_{\mathsf{fail},i}|\mathbf{y})}{\beta_{\mathsf{fail},i} + n_{\mathsf{mil},j}}.
\end{aligned}
$$

Proof. The result follows directly from Lemma 1 because Eqs. (42) and (43) are linear with regard to $n^*_{\mathsf{man},i}$ and $n^*_{\mathsf{fail},i}$. □

Unfortunately, there is no simple and modular closed-form expression for the expectations $E(n^*_{\mathsf{man},i}|\mathbf{y})$ and $E(n^*_{\mathsf{fail},i}|\mathbf{y})$ because the probability that an unanalyzed sample is of a certain fault is not independent of the observations made at the same time which causes a dependency to the symptom-cause probabilities and observation parameters.

We could estimate the parameters based solely on the analyzed cases. Then we must reduce the total mileage and number of components proportionally to the fraction of analyzed components to compensate for faults that did not occur, but if they would have, they would not have been analyzed anyway. We estimate the probability of analyzing a claim on component j as:

$$
p_{\mathsf{an},j} = \frac{\sum_{i \in \mathcal{I}_{\mathsf{F}|j}} n_{\mathsf{man},i} + n_{\mathsf{fail},i}}{n_{\mathsf{miss},j} + \sum_{i \in \mathcal{I}_{\mathsf{F}|j}} n_{\mathsf{man},i} + n_{\mathsf{fail},i}}.
$$

Then the compensated number of vehicles in the population with component j is $\tilde{n}_{\mathsf{comp},j} = n_{\mathsf{comp},j} p_{\mathsf{an},j}$ and the compensated total mileage of the vehicles in the population with component j is $\tilde{n}_{\mathsf{mil},j} = n_{\mathsf{mil},j} p_{\mathsf{an},j}$. Let $\tilde{\mathbf{y}} = \{y^{[i]} \in \mathbf{y} : f^{[i]}_{\mathsf{id}} \neq 0\}$ be the subset of $\mathbf{y}$ where all samples are analyzed. Then we can use Lemma 1 to get estimates based only on the analyzed cases:

$$
\tilde{p}_{\mathsf{fail},i} = E(p_{\mathsf{fail},i}|\tilde{\mathbf{y}}) = \frac{\alpha_{\mathsf{man},i} + n_{\mathsf{man},i}}{\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i} + n_{\mathsf{comp},j} p_{\mathsf{an},j}} \tag{44}
$$

$$
\tilde{\lambda}_{\mathsf{fail},i} = E(p_{\mathsf{fail},i}|\tilde{\mathbf{y}}) = \frac{\alpha_{\mathsf{fail},i} + n_{\mathsf{fail},i}}{\beta_{\mathsf{fail},i} + n_{\mathsf{mil},j} p_{\mathsf{an},j}}. \tag{45}
$$

However, we can get better estimates if we use knowledge of the unanalyzed claims on a component to learn the fault probabilities of that component in relation to other components. This can for example be done by using the learned parameters

$\tilde{p}_{\mathsf{fail},i}$ and $\tilde{\lambda}_{\mathsf{fail},i}$ to estimate the expected values of $n_{\mathsf{man},i}$ and $n_{\mathsf{fail},i}$ for all $i \in \mathcal{I}_{\mathsf{F}|j}$ and then compute the posterior expected values using Lemma 2. Let the estimates of the expected values $E(n^*_{\mathsf{man},i}|\mathbf{y})$ and $E(n^*_{\mathsf{fail},i}|\mathbf{y})$ be

$$\hat{n}_{\mathsf{man},i} = n_{\mathsf{man},i} + n_{\mathsf{miss},j} \frac{n_{\mathsf{comp},j}\tilde{p}_{\mathsf{fail},i}}{\sum\limits_{k \in \mathcal{I}_{\mathsf{F}|j}} n_{\mathsf{comp},j}\tilde{p}_{\mathsf{fail},k} + n_{\mathsf{mil},j}\tilde{\lambda}_{\mathsf{fail},k}}$$

$$\hat{n}_{\mathsf{fail},i} = n_{\mathsf{fail},i} + n_{\mathsf{miss},j} \frac{n_{\mathsf{mil},j}\tilde{\lambda}_{\mathsf{fail},i}}{\sum\limits_{k \in \mathcal{I}_{\mathsf{F}|j}} n_{\mathsf{comp},j}\tilde{p}_{\mathsf{fail},k} + n_{\mathsf{mil},j}\tilde{\lambda}_{\mathsf{fail},k}}$$

then using Lemma 2, we can estimate the posteriors as

$$\hat{p}_{\mathsf{fail},i} = \frac{\alpha_{\mathsf{fail},i} + \hat{n}_{\mathsf{man},i}}{\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i} + n_{\mathsf{comp},j}} \tag{46}$$

$$\hat{\lambda}_{\mathsf{fail},i} = \frac{\alpha_{\mathsf{fail},i} + \hat{n}_{\mathsf{fail},i}}{\beta_{\mathsf{fail},i} + n_{\mathsf{mil},j}}. \tag{47}$$

When there are few analyzed faults we assume that the unanalyzed faults are distributed like the prior and when there are many analyzed faults we rely more on the data. In difference to Eqs. (44) and (45), the estimates Eqs. (46) and (47) are capable of learning from unanalyzed samples. A data sample that has not been analyzed contributes to learning the relative failure probabilities between faults on different components, but it does not affect the relative failure probabilities of the faults on the same component. Also, as the size of the data grows, these approximate parameter posteriors will converge toward the true parameter values.

Theorem 3 (Convergence of Approximate Fault Probability Posteriors)**.** For all $j \in \mathcal{I}_\mathsf{C}, i \in \mathcal{I}_{\mathsf{F}|j}$, if $p^*_{\mathsf{fail},i}$ and $\lambda^*_{\mathsf{fail},i}$ are the true manufacturing fault probabilities and failure rates, then as the number of vehicles having the component j that are contributing to the data grows, the estimates $\hat{p}_{\mathsf{fail},i}$ and $\hat{\lambda}_{\mathsf{fail},i}$ as given by Eqs. (46) and (47) converge to $p^*_{\mathsf{fail},i}$ and $\lambda^*_{\mathsf{fail},i}$:

$$\lim_{n_{\mathsf{comp},j}\to\infty} \hat{p}_{\mathsf{fail},i} = p^*_{\mathsf{fail},i},$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} \hat{\lambda}_{\mathsf{fail},i} = \lambda^*_{\mathsf{fail},i}.$$

Proof. See A.4 □

Example 5 (Learning Fault Probabilities)**.** Consider the example from Section 5.7. Let the priors be

$$\alpha_{\mathsf{man},i} = 0.1 \qquad \beta_{\mathsf{man},i} = 9999.9 \qquad \alpha_{\mathsf{fail},i} = 20 \qquad \beta_{\mathsf{fail},i} = 2000$$

for $i = 3,4,5,6$. Suppose we have data $\mathbf{y}$ according to the following tables for components and faults respectively:

i	$n_{\mathsf{comp},i}$	$n_{\mathsf{mil},i}$	$n_{\mathsf{miss},i}$
1	100000	20000	0
2	100000	20000	2727

i	$n_{\mathsf{man},i}$	$n_{\mathsf{fail},i}$
3	20	1000
4	10	2000
5	2	100
6	1	200

The only difference between the priors and data for component 1, with faults 3 and 4, and those for component 2, with faults 5 and 6, is that for component 2 only 10 % of all samples are analyzed (303 out of $303+2727$). Proportionally, the data for both components is identical. Regardless which method is used, the posterior parameter values for the faults on component 1 are identical:

$$\begin{array}{rclcrcl}
\tilde{p}_{\mathsf{fail},3} & = & 1.8\times 10^{-4} & \qquad & \hat{p}_{\mathsf{fail},3} & = & 1.8\times 10^{-4} \\
\tilde{p}_{\mathsf{fail},4} & = & 9.2\times 10^{-4} & & \hat{p}_{\mathsf{fail},4} & = & 9.2\times 10^{-5} \\
\tilde{\lambda}_{\mathsf{fail},3} & = & 0.046\times 10^{-4} & & \hat{\lambda}_{\mathsf{fail},3} & = & 0.046\times 10^{-4} \\
\tilde{\lambda}_{\mathsf{fail},4} & = & 0.092\times 10^{-4} & & \hat{\lambda}_{\mathsf{fail},4} & = & 0.092\times 10^{-5}
\end{array}$$

However when only 10 % of the samples are analyzed, the posterior using the second method is closer to the posterior when all samples are analyzed:

$$\begin{array}{rclcrcl}
\tilde{p}_{\mathsf{fail},5} & = & 1.1\times 10^{-4} & \qquad & \hat{p}_{\mathsf{fail},5} & = & 1.7\times 10^{-4} \\
\tilde{p}_{\mathsf{fail},6} & = & 5.5\times 10^{-5} & & \hat{p}_{\mathsf{fail},6} & = & 8.9\times 10^{-5} \\
\tilde{\lambda}_{\mathsf{fail},5} & = & 0.030\times 10^{-4} & & \hat{\lambda}_{\mathsf{fail},5} & = & 0.049\times 10^{-4} \\
\tilde{\lambda}_{\mathsf{fail},6} & = & 0.055\times 10^{-5} & & \hat{\lambda}_{\mathsf{fail},6} & = & 0.089\times 10^{-5}
\end{array}$$

7.4 Symptom-Cause Probabilities

There is no simple closed-form expression for the expected value of the symptom-cause probabilities because the symptom variables are hidden. If we fix the observation parameters $p_{+,k}, p_{-,k}$ for all $k \in \mathcal{I}_{\mathsf{O}}$ we can express the posterior distribution of the analyzed data given the symptom-cause probabilities of a fault $i \in \mathcal{I}_{\mathsf{F}}$ that is independent of all other faults.

Theorem 4 (Symptom-Cause Posterior)**.** Let $\tilde{\mathbf{y}} = \{y^{[i]} \in \mathbf{y} : f_{\mathsf{id}}^{[i]} \neq 0\}$ be a subset of $\mathbf{y}$ where all samples are analyzed, and let $\mathcal{I}_{\mathbf{y}}(i)$ be the indexes of analyzed samples in $\mathbf{y}$ where fault i is reported. Then the posterior distribution of $p_{i,j}$ for all symptoms $j \in \mathcal{I}_{\mathsf{S}|i}$ given $\tilde{\mathbf{y}}$ is

$$\Pr(p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i} | \tilde{\mathbf{y}}) \propto \prod_{m \in \mathcal{I}_{\mathbf{y}}(i)} \Pr(\mathbf{o}^{[m]} | f_i^t, \neg f_n^t \forall n \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}, p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) \prod_{k \in \mathcal{I}_{\mathsf{S}|i}} p(p_{i,k}).$$

Proof. See A.5 □

If we can compute the density of the posterior distribution we can estimate the expected values of $p_{i,j}$ for all $i \in \mathcal{I}_\mathsf{F}, j \in \mathcal{I}_{\mathsf{S}|i}$ by drawing samples from the distribution using Monte Carlo methods or finding its mode using a numerical optimizer to compute the maximum à posteriori (MAP). Regardless of which method we choose we must be able to efficiently compute the posterior density. The density of the prior has a closed form expression

$$p(p_{i,j}) \propto p_{i,j}^{\alpha_{i,j}-1}(1-p_{i,j})^{\beta_{i,j}-1}$$

and we can efficiently compute an approximation of the likelihood

$$\Pr(\mathbf{o}^{[m]}|f_i^t, \neg f_n^t \forall n \in \mathcal{I}_\mathsf{F} \backslash \{i\}, p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i})$$

using the single fault method described in Section 6.2. Computing the MAP is computationally easier than using Monte-Carlo methods, but there is a risk of numerical instability when $\alpha_{i,j} < 1$ or $\beta_{i,j} < 1$ for some $i \in \mathcal{I}_\mathsf{F}, j \in \mathcal{I}_{\mathsf{S}|i}$.

In general, convergence is not guaranteed as the data set grows. The symptom-cause probabilities of symptoms that that are indistinguishable given the observations cannot be learned, for example when a fault can cause two symptoms that are related to the same set of observations. Such a situation can be avoided by modeling such symptoms as the same since they are indistinguishable to an external observer.

8 Diagnostic Model for Application Example

We have developed the diagnostic model for the application example from Section 2 together with domain experts. It has 19 components, 47 faults, 86 symptoms, and 89 observations. There are 527 symptom cause variables and 15 logical variables. Figure 7 shows the topology of the model for the rail pressure sensor containing 32 of the 783 variables.

We have little prior knowledge of manufacturing fault probabilities and failure rates of the individual faults in the model. These are therefore all given identical parameter values typical for any type of fault. The manufacturing fault probabilities are set to 1 manufacturing fault of each type per million parts and the failure rates are set to 1 failure of each type per 100 million km.

The rates of change for symptom causes are also difficult for an expert to estimate and therefore these are the same for all symptom-causes. With a rate of change of 1 change per 100 km, the probability that a symptom cause with probability parameter p follows a previously observed distribution rather than Bernoulli(p) is reduced by 50 % for each 70 km. This means that an observation of a symptom is

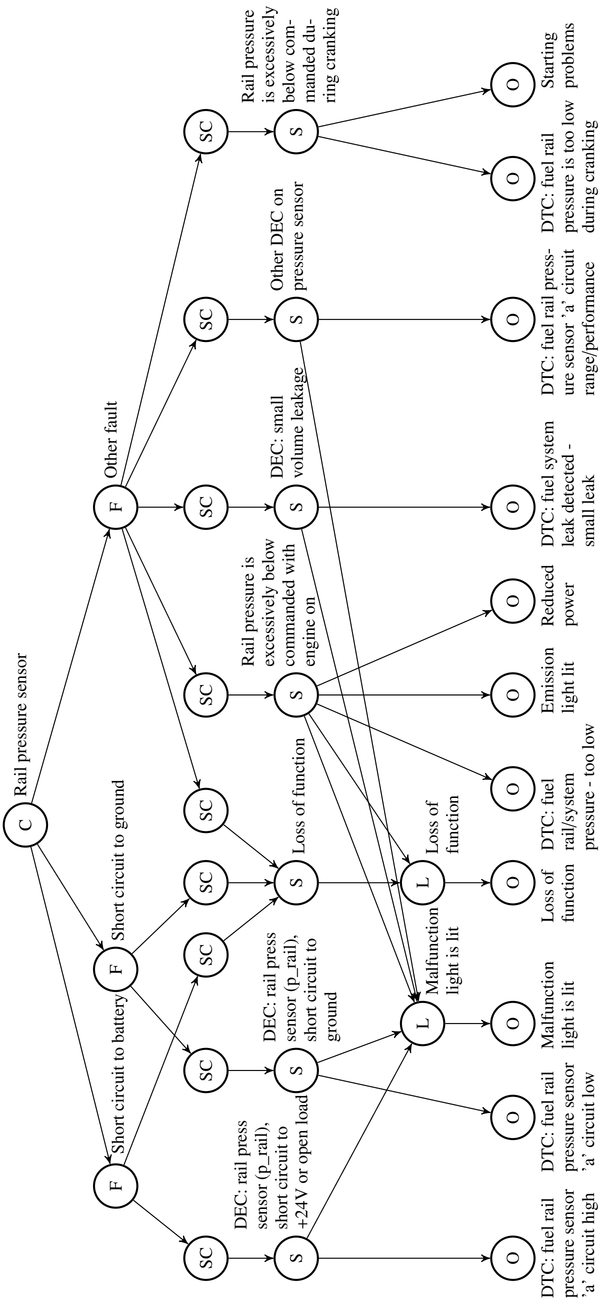
C
Rail pressure sensor
F
Short circuit to battery
F
Short circuit to ground
F
Other fault
SC
S
L
O
DEC: rail press sensor (p_rail), short circuit to +24V or open load
DEC: rail press sensor (p_rail), short circuit to ground
Loss of function
Rail pressure is excessively below commanded with engine on
DEC: small volume leakage
Other DEC on pressure sensor
Rail pressure is excessively below commanded during cranking
Malfunction light is lit
Loss of function
DTC: fuel rail pressure sensor 'a' circuit high
DTC: fuel rail pressure sensor 'a' circuit low
Malfunction light is lit
Loss of function
DTC: fuel rail/system pressure - too low
Emission light lit
Reduced power
DTC: fuel system leak detected - small leak
DTC: fuel rail pressure sensor 'a' circuit range/performance
DTC: fuel rail pressure is too low during cranking
Starting problems

Figure 7: The model topology for the rail pressure sensor in the application example.

highly dependent of other observations of that symptom that have been made during the same troubleshooting session, but almost completely independent of such observations that have been made during for example the previous vehicle service 50000 km ago.

Of the observations, 36 variables correspond to DTCs that can be generated by the on-board diagnostic system. A DTC is generated by the on-board diagnostic system when one or more internal tests are triggered. Each such internal test is modeled with a symptom variable and joined together with a logical variable. A symptom-cause variable is added for each fault the test is sensitive for. DTCs for electrical faults are very reliable and are given a symptom-cause probability of 0.99999. For other DTCs, the symptom-cause probabilities are given a value between 0.1 and 0.9 by an expert. By design, the false positive probabilities for DTCs are rare and therefore the false positive probability parameters have very small values (10^{-9}–10^{-5}). The false negative probability is accounted for by the symptom cause probabilities so these values are also set low (10^{-9}–10^{-5}). Most DTCs are sensitive for very few faults. However some, like the DTC for low fuel rail pressure, are sensitive for almost every fault on the fuel injection system.

There are 9 observations for symptoms that a driver can experience. Some of these are vague and can be caused by many faults, e.g. "high fuel consumption" and "loss of function" which makes it difficult for an expert to correctly set parameters. However, these symptoms are often reported in the warranty claims and can be learned from that source instead. Other symptoms corresponding to warning lights being lit and torque limitations being active are consequences of the actions of the on-board diagnostic system. For these symptoms, the parameters can be set with a much higher confidence.

The remaining 44 observations are for tests that can be run in the workshop. The parameters for these cannot be learned from the warranty data, but since the workshop tests are developed by the experts themselves they can set parameter values with great confidence.

The model was validated by simulating the instructions from the troubleshooting guide in the workshop manual used by the mechanics and comparing whether the output of the diagnoser agrees with the information found in the manual. For each possible sequence of tests proposed in the manual, the diagnoser assigns the highest probabilities to the components listed as possible explanations in the manual.

9 Experimental Evaluation

To evaluate the performance of the diagnostic framework three series of experiments have been conducted. The first series of experiments has the purpose of evaluating the different inference algorithms in terms of computation time and precision. The second series of experiments evaluates how well we can learn the model parameters from warranty data. The third series of experiments evaluates how sensitive the framework is to errors in the model parameters.

All implementations were made in Java and run on an Intel Core i7-4800MQ processor with 16 GB of RAM.

9.1 Diagnostic Inference

To evaluate which of the described inference algorithms is the best for our application we have conducted a series of four experiments using random models.

The random models are generated using parameters n, m, and p as following. First n fault variables and m symptom variables are created. For each fault-symptom pair, a symptom-cause variable is created with probability p. For each symptom variable, an observation variable is created. For each fault i, symptom j, and observation k, parameter values are drawn as $p_{\mathsf{fail},i} \sim \text{Beta}(2,10000)$, $\lambda_{\mathsf{fail},i} \sim$ Gamma(2,1000), $p_{i,j}{\sim}$Beta(10,2), $\lambda_{i,j}{\sim}$Gamma(10000,2), $p_{+,k}{\sim}$Beta(2,1000), and $p_{-,k} \sim \text{Beta}(2,1000)$.

Four different inference methods are evaluated: *exact* inference, Boyen-Koller (*BK*) using the Quickscore Variant algorithm, the *single* fault model using the method described in Section 6.2, and the *combined* model using the method described in Section 6.3.

Small Models – Single Fault

In the first experiment we measure the average absolute error ε in the logit of the marginal fault probability distributions after a sequence of events $\mathbf{e}$:

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left|\ln\frac{p_i}{1-p_i} - \ln\frac{p_i^*}{1-p_i^*}\right|$$

where $p_1,\dots,p_n$ are the marginal fault probabilities of an approximate inference method and $p_1^*,\dots,p_n^*$ are the marginal fault probabilities of the exact inference. By transforming the marginal probabilities into logits, a given absolute error on a value close to 0 or 1 has a larger impact on ε than the same absolute error on a value close to 0.5: Erroneously estimating a fault probability to be 1 % when it should be 11 % is considered more severe than estimating it to be 60 % when it should be 70 %. We also measure the average absolute error of the untransformed marginal fault probabilities:

$$\delta = \frac{1}{n}\sum_{i=1}^{n}|p_i - p_i^*|$$

n	t_{exact}	t_{single}	t_{BK}	t_{comb}	ε_{single}	ε_{BK}	ε_{comb}	δ_{single}	δ_{BK}	δ_{comb}
2	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	1.8	3.6	0.030	0.11	0.28	0.0040
3	0.043	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	4.9	1.9	0.68	0.052	0.11	0.022
4	4.3	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	6.6	1.3	0.94	0.026	0.056	0.013
5	520	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	7.5	0.75	0.85	0.016	0.034	0.0090

Table 6: Comparing algorithms with computation time and error in posterior fault probabilities for the first experiment with a single fault.

The sequence of events is set to mimic a typical troubleshooting session which begins with observation actions with the purpose of isolating the fault, continues with repair actions, and concludes with further observations to verify that the problem is solved. To generate the sequence of events, we first draw a random state s with at least one fault after 10000 km of operation. Then two observation variables are randomly selected with uniform probability and outcomes for these observations are drawn by simulating the system response in state s. Then an observation of a fault present on s is drawn followed by a repair event on that fault and one more random observation. Between each pair of events a 100-km operation event is inserted. We set $m = n$ and $p = 2/n$ and for each $n = 2,3,4,5$ we repeated the experiment 100 times.

Averages of the computation time t_i and errors ε_i, δ_i are reported for each method $i \in \{exact, single, BK, combined\}$ in Table 6. We note that the computation time for exact inference grows exponentially while the computation times for approximate inference is small. The logit error is large for the single fault method. This is because it grossly underestimates the probability of other faults after the first repair. Both the BK method and the combined method have smaller logit error, but the BK method is weaker in this aspect for smaller model sizes. For the absolute errors, it is the single fault and combined methods that are the strongest. This is because the problem with the independence assumptions as demonstrated in Example 2 can be avoided.

Small Models – Multiple Faults

The second experiment is set up in the same way as the first except the random state s is drawn so that two faults are present. The averages of the computation time t_i and errors ε_i, δ_i are reported for each method $i \in \{exact, single, BK, combined\}$ in Table 7. Contrary to intuition, the logit error decreases for the single-fault method compared to the case where there was only one fault. This is because it explains anomalous observations from the secondary fault with a new fault occurring after the first repair. However, except for when $n = 2$, the logit error is still significantly larger than for the other methods. Compared to the combined method, the BK-

n	t_{exact}	t_{single}	t_{BK}	t_{comb}	ε_{single}	ε_{BK}	ε_{comb}	δ_{single}	δ_{BK}	δ_{comb}
2	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.86	4.3	0.030	0.17	0.32	0.0040
3	0.043	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	3.1	2.2	1.2	0.16	0.16	0.081
4	4.3	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	4.9	1.7	1.3	0.11	0.11	0.050
5	510	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	5.7	1.2	1.4	0.072	0.065	0.044

Table 7: Comparing algorithms with computation time and error in posterior fault probabilities for the second experiment with two faults.

method has both large logit and absolute errors for the smaller model sizes, but the effect seems to diminish with model size. The logit and absolute errors of the combined method are never far from the best of the single fault method and BK-method. Like the BK-method it is able to recognize the possibility of multiple faults, but after the repair, double faults that used to include the repaired fault become single faults for which it can update the probabilities more precisely than the BK-method.

Larger Models – Single Fault

In the third experiment we use larger models with up to 1000 fault variables. We simulate the vehicle with a randomly drawn single fault. We compare each inference method i by its estimates of the probability γ_i of having the correct fault after making several observations and its estimates of the probability η_i of having no faults after inspecting and repairing the fault and making further observations. Because we cannot make exact inference on models of this size, we do not know what the correct values of the estimates are. However, we can expect that after the first series of observations the correct fault shows symptoms that makes it a probable candidate with a significant probability (>0.2) and that after the fault is inspected and repaired the true probability of having no faults should be close to 1. We also want to see that the methods scale well in computation time.

The experiment is set up as follows. After 10000 km of operation a random state with a single fault is selected to simulate the responses of subsequent observations. Then 3n/8 observation variables are drawn uniformly and observed in intervals of 100 km. After making these observations we compute and record γ_i. Then the fault is inspected and repaired and $n/8$ more random observations are made in intervals of 100 km before the estimate η_i is computed and recorded. We repeated this procedure 100 times using models generated with $n = m = 10, 20, 50, 100, 200, 500, 1000$ and $p = 5/n$ and using the XPI model.

The averages of the computation time t_i and the estimated probabilities γ_i and η_i are reported for each method i in Table 8. All methods point out the correct fault before repair with approximately equal probability. The single fault method gives

n	t_{single}	t_{BK}	t_{comb}	γ_{single}	γ_{BK}	γ_{comb}	η_{single}	η_{BK}	η_{comb}
10	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.27	0.27	0.27	1.0	0.88	1.0
20	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.40	0.37	0.41	1.0	0.88	1.0
50	0.0018	0.0022	0.0049	0.46	0.45	0.46	1.0	0.87	1.0
100	0.0074	0.0096	0.019	0.44	0.44	0.45	1.0	0.90	0.97
200	0.031	0.039	0.080	0.49	0.46	0.49	1.0	0.88	0.96
500	0.22	0.28	0.57	0.46	0.46	0.47	0.99	0.80	0.89
1000	0.90	1.1	2.3	0.41	0.43	0.45	0.92	0.69	0.74
XPI	0.0055	0.0065	0.013	0.36	0.36	0.36	1.0	0.78	0.99

Table 8: Comparing algorithms for larger models.

a high estimate of the no-fault probability for all cases. The other methods estimate this probability to be lower which is assumed to be incorrect. The combined model behaves more like the BK method the more faults there are in the model because the probability of multiple faults increases. The behavior of the random models is consistent with that of the XPI model. The computation time grows quadratically with n because the computation time of each event is linear in n and the number of events grows linearly with n. Most importantly, for large n, the computation time is still small. The combined method has a computation time that is about twice that of the other methods.

Larger Models – Multiple Faults

The fourth experiment is set up in the same way as the third except that we select a state with two faults just as we did in the second experiment. After the first series of observations we record the probability γ_i of having *either* of the two faults. Then we inspect and repair any one of the two faults and make further observations and record the probability η_i of having the *other* unrepaired fault. Again we do not know the exact values but we expect that the two faults show symptoms that make them probable candidates and that the unrepaired fault should remain probable after the first fault is repaired.

The averages of the computation time t_i and probabilities γ_i and η_i are reported for each method i in Table 9. In this experiment, the single fault method estimates the probability of the second fault to be much lower than the other methods which is assumed to be incorrect. The observations are inconsistent with a single fault and therefore the combined method gives estimates η_{comb} that lie close to η_{BK}.

n	t_{single}	t_{BK}	t_{comb}	γ_{single}	γ_{BK}	γ_{comb}	η_{single}	η_{BK}	η_{comb}
10	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.36	0.43	0.34	0.044	0.39	0.32
20	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.50	0.53	0.52	0.038	0.40	0.33
50	0.0016	0.0022	0.0042	0.61	0.64	0.63	0.051	0.51	0.44
100	0.0067	0.0084	0.017	0.68	0.72	0.71	0.060	0.60	0.55
200	0.028	0.035	0.072	0.65	0.67	0.67	0.047	0.53	0.51
500	0.2	0.26	0.54	0.63	0.63	0.64	0.062	0.51	0.50
1000	0.93	1.2	2.4	0.75	0.76	0.77	0.063	0.48	0.48
XPI	0.0057	0.0069	0.014	0.51	0.63	0.59	0.14	0.51	0.50

Table 9: Comparing algorithms for larger models with double fault.

9.2 Parameter Learning

To evaluate the framework's ability to learn parameters from warranty data we have conducted a series of experiments where data generated from a reference model is used to learn a model with offset prior parameters. The data is generated for vehicle populations of different sizes and the average differences between the posterior parameter values and the parameters in the reference model are recorded.

In this experiment, the reference model represents the true behavior of faults on the vehicles. It has 100 fault variables and 34 component variables and it is generated randomly using the same parameters as for the experiment described in Section 9.1. The reference model is used to generate sets of warranty data for simulated vehicle populations of $100, 200, 500, 1000, \ldots, 10^6$ vehicles. When a component has been in use a couple of years, the number of vehicles contributing to the data for this component is typically in the order of 100000 vehicles. For new or rare components this can be smaller, but a vehicle population greater than 1000000 vehicles is unlikely. The mileage in millions of km during the warranty year of each vehicle is simulated by drawing it from the exponential distribution with mean 0.2. A mileage of 0.2 million km per year is a typical value for trucks operating in long-haulage.

The prior model represents the model created by the experts before any parameter learning using data has been performed. To simulate possible errors introduced by the experts, the prior model is created from a copy of the reference model where the parameter values are randomly offset. For parameter values θ that are probabilities, a small difference close to 0 or 1 has a greater impact on the characteristics of the model than the same difference near 0.5. Because of this, such a parameter value in the reference model is first transformed using the logit transformation $\text{logit}(\theta) = \ln(\theta/(1-\theta))$ before it is offset with a random value drawn from a normal distribution. Then it transformed back into a probability using the inverse of the transformation. For parameter values θ that are rates, a small difference close to zero has a greater impact than at larger values. These are first transformed using the logarithm $\ln(\theta)$ before offsetting them with a random values drawn from a normal distribution. For all parameters except manufacturing fault probabilities and failure rates a normal distribution with variance 4 and mean 0 is used. For the manufacturing fault probabilities and failure rates noise is added twice. First a normal distribution with variance 3 and mean 0 is used to draw an offset value for each component that is added to the transformed value of the parameters associated to that same component. Then another normally distributed offset with variance 1 and mean 0 is added independently to the transformed values of each parameter. This is done so that the probability of any fault on a particular component in the prior model will deviate from that in the reference model. The error in the prior model is fairly large: The average difference between the transformed parameter values in the prior model and the reference model will be $\sqrt{8/\pi} \approx 1.6$. For 20 % of the parameters the difference will be larger than 2.5. E.g. an error of +2.5 in the transformed values implies an overestimation of a failure rate or a small probability with a factor 12 and a probability near 0.5 as 0.92.

The priors in the prior model are set as following:

$$\begin{aligned}
\alpha_{\mathsf{man},i} &= 10^4 p_{\mathsf{fail},i} \\
\beta_{\mathsf{man},i} &= 10^4(1 - p_{\mathsf{fail},i}) \\
\alpha_{\mathsf{fail},i} &= 2000\lambda_{\mathsf{fail},i} \\
\beta_{\mathsf{fail},i} &= 2000 \\
\alpha_{i,j} &= 100 p_{i,j} \\
\beta_{i,j} &= 100(1 - p_{i,j})
\end{aligned}$$

where $p_{\mathsf{fail},i}$, $\lambda_{\mathsf{fail},i}$, and $p_{i,j}$ are the parameter values in the prior model. We can interpret the meaning of these priors as corresponding to the following "hypothetical data": We have seen 10000 vehicles and $\alpha_{\mathsf{man},i}$ of those had fault i initially. The vehicles have rolled 2000 million km (10000 vehicles driving 0.2 million km per year) and there were $\alpha_{\mathsf{fail},i}$ occurrences of fault i. In 100 occurrences of fault i, symptom j was present $\alpha_{i,j}$ times.

For each vehicle population size, we learn a posterior model from the prior model and the corresponding data set by applying the learning method described in Section 7. The symptom-cause probabilities are learned by finding the maximum posterior using the optimizer BOBYQA as implemented in the Apache Commons Math library version 3.1.1 [1]. In Figure 8 we compare the posterior models with the reference model by showing the mean errors on the transformed parameter values when vehicle population size n is varied from 100 to 1000000. At $n = 100$ the parameter values have not yet started to converge and the mean errors of the transformed values are approximately 1.6. This is a fairly large error. E.g. an error of +1.6 means for the untransformed values that a small probability or a failure rate differ between the models by a factor 5 and a probability of 0.5 in the posterior model is 0.8 in the reference model. When $n = 1000000$, the mean error is below 0.5 which is much smaller. E.g. an error of +0.5 means that a small probability or a failure rate differ between the models by only a factor 1.6 and a probability of 0.5 in the posterior model is 0.6 in the reference model. In Section 9.3 we will study how errors of different sizes affect the performance of diagnosis.

We see in the plot that the posterior error does not begin to converge until the population size is greater than 10000. This is because the priors are selected to correspond to 10000 vehicles of "hypothetical data". This appears to be a good trade-off. Stronger priors would have made convergence even slower. A weaker prior can cause the posterior parameter values to be even more erroneous than the prior parameter values when the population size is small. E.g. when the expected number of a fault occurrences of a particular type in the data is smaller than 1 it is likely that the fault does not appear at all in the data with the consequence that the failure rate of this fault becomes underestimated. If an unlikely fault instead do appear in the data, its failure rate becomes overestimated. Also, symptom-cause probabilities that are close to 1 tend to be overestimated because the data only consists of observations supporting the positive case. Likewise holds for symptom-cause probabilities close to 0 which tend to be underestimated. We see this effect in Figure 8 for the symptom-cause probabilities when the vehicle population size is small.

In this experiment the failure rates and manufacturing fault probabilities are the fastest to converge, while the symptom cause probabilities converge slower. Therefore it is important that the experts put more effort in setting the priors for these parameters.

The same experiment with generated data has also been conducted for the XPI model. The results shown in Figure 9 are consistent with those in Figure 8. The generated data had proportionally very few cases of manufacturing faults hence the slow learning for those parameters.

To see the effect of not determining the exact fault on a component, we varied the experiment by simulating that only one tenth of all faults are analyzed. The results shown in Figure 10 indicate that we can still be able to learn the parameters but the learning is slower. For symptom-cause probabilities, we require data from a ten times larger population to achieve the same error because unanalyzed data entries do not contribute to learning. For manufacturing fault probabilities and failure rates, the convergence is initially unaffected by the missing data (error greater than one), but full convergence is slower. This is because we can learn the mean fault probabilities and failure rates of all faults on the component using the unanalyzed data, but we require an equal amount of analyzed data to learn the individual fault probabilities and failure rates of each fault.

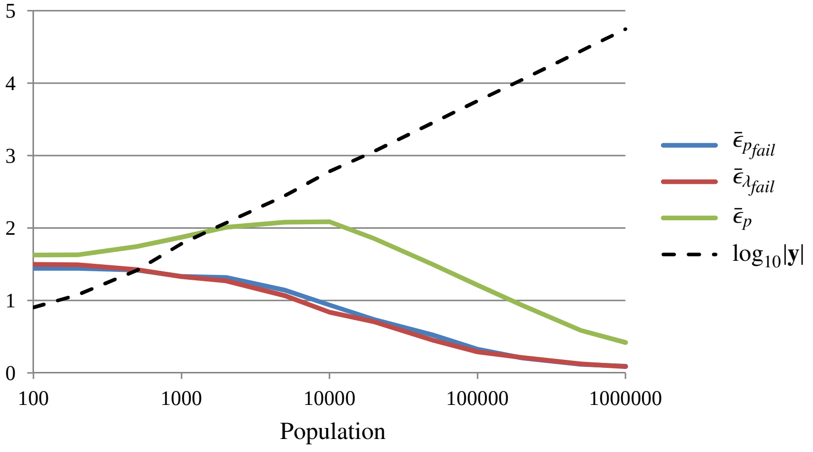

Figure 8: Learning of parameters using simulated warranty data on a random model with 100 faults.

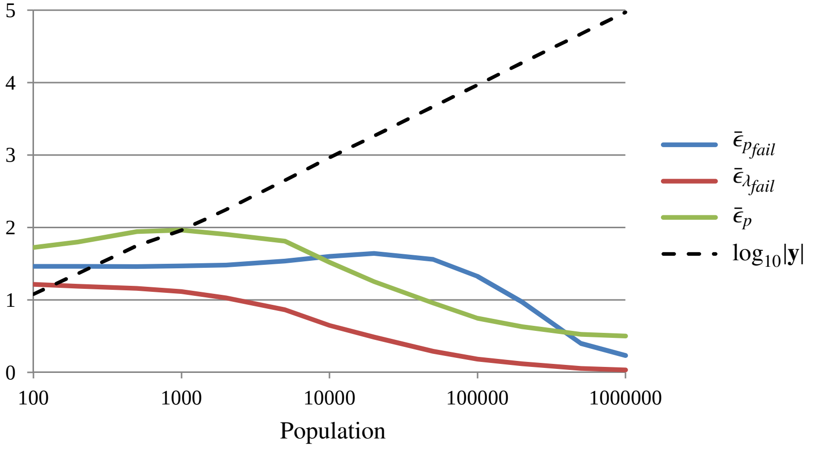

Figure 9: Learning of parameters using simulated warranty data on the XPI model.

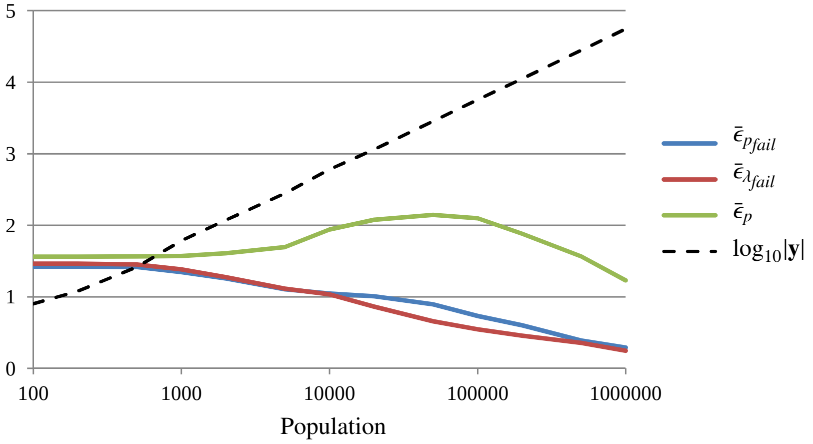

Figure 10: Learning of parameters using simulated warranty data on a random model with 100 faults when only one tenth of all faults are analyzed.

9.3 Modeling Errors

To evaluate the effect of modeling errors on the parameters, an experiment is conducted where random noise is added to the parameter values of random models and the difference in the posterior marginal distribution is measured. In this experiment we have varied the variance v of the random noise. For each compared variance level v, 100 reference models with 100 fault variables are generated in the same way as in the previous experiment. Then 7 noisy models are generated for each of these reference models in the same way as we generated the prior model from the reference model in the previous experiment. One noisy model has noise added to all parameters $\bar{\varepsilon}_{all}$, and six have noise added to only one parameter type at a time: the manufacturing fault probabilities $\bar{\varepsilon}_{p_0}$, failure rates $\bar{\varepsilon}_{\lambda_f}$, symptom-cause probabilities $\bar{\varepsilon}_{p_s}$, symptom-cause rates $\bar{\varepsilon}_{\lambda_s}$, false positive probabilities $\bar{\varepsilon}_{p^+}$, and false negative probabilities $\bar{\varepsilon}_{p^-}$.

Table 10 shows a comparison of the average errors in the logit transformations of the posterior marginalized fault probabilities. We see that the average error is approximately proportional to the standard deviation of the noise $\sqrt{v}$. The largest sensitivity to errors is on the fault probability parameters and symptom-cause probabilities. Fortunately these parameters are also those parameters for which our initial estimates can best be improved using statistical data. Hence we become less sensitive to errors made by the experts in the prior estimates.

Table 11 shows the results for the same experiment when we instead look at the averages of the largest absolute errors in the posterior marginal fault probabilities for each model. Only the errors in the more likely faults in the posterior have

v	$\sqrt{v}$	$\bar{\varepsilon}_{all}$	$\bar{\varepsilon}_{p_0}$	$\bar{\varepsilon}_{\lambda_f}$	$\bar{\varepsilon}_{p_s}$	$\bar{\varepsilon}_{\lambda_s}$	$\bar{\varepsilon}_{p^+}$	$\bar{\varepsilon}_{p^-}$
0.1	0.31	0.28	0.11	0.16	0.11	0.005	0.017	0.001
0.2	0.45	0.39	0.16	0.23	0.16	0.007	0.024	0.002
0.50	0.71	0.63	0.26	0.37	0.24	0.011	0.037	0.003
1	1	0.90	0.38	0.52	0.34	0.017	0.051	0.005
2	1.4	1.3	0.56	0.74	0.46	0.026	0.070	0.009
5	2.2	2.0	0.93	1.2	0.67	0.043	0.11	0.030
10	3.2	2.8	1.3	1.8	0.85	0.056	0.16	0.084
20	4.5	3.9	1.9	2.6	1.0	0.069	0.22	0.21

Table 10: Average logit error in the posterior marginal fault probabilities when normal noise with variance v and mean 0 is added to 100 random models with 100 faults.

v	$\sqrt{v}$	$\bar{\varepsilon}_{all}$	$\bar{\varepsilon}_{p_0}$	$\bar{\varepsilon}_{\lambda_f}$	$\bar{\varepsilon}_{p_s}$	$\bar{\varepsilon}_{\lambda_s}$	$\bar{\varepsilon}_{p^+}$	$\bar{\varepsilon}_{p^-}$
0.1	0.31	0.011	0.007	0.001	0.003	0.0	0.002	3.2×10^{-5}
0.2	0.45	0.016	0.01	0.002	0.005	0.001	0.003	4.8×10^{-5}
0.50	0.71	0.027	0.017	0.003	0.007	0.001	0.005	8.8×10^{-5}
1	1	0.041	0.027	0.004	0.01	0.001	0.006	0.0
2	1.4	0.068	0.045	0.009	0.015	0.002	0.008	0.0
5	2.2	0.15	0.11	0.031	0.028	0.003	0.012	0.001
10	3.2	0.4	0.3	0.15	0.045	0.003	0.015	0.004
20	4.5	0.87	0.77	0.87	0.077	0.003	0.02	0.011

Table 11: Average largest absolute error in the posterior marginal fault probabilities when normal noise with variance v and mean 0 is added to 100 random models with 100 faults.

impact on this measure. When the noise on the transformed parameter values has a variance that is smaller than 5 the average largest absolute error is small. As a comparison, the variance of the difference between the the transformed values of the parameters in the reference model and the prior model used in Section 9.2 was 4. For the largest noise levels, the differences are however much larger because the faults with the highest posterior probabilities are unlikely to be the same.

10 Related Work

The troubleshooting system which this diagnostic modeling framework is a part of is unique in the way it integrates remote and workshop troubleshooting. Other existing approaches for constructing decision support systems that emphasize minimizing the expected cost of troubleshooting typically consider only the direct costs of the actions performed in the workshop. Many of these approaches use diagnostic models that are static Bayesian networks, e.g. [13, 18, 23, 25, 26, 46]. Others use consistency-based diagnostic models augmented with probabilities, see e.g. [4, 9, 38].

There is work in the area of condition-based and predictive maintenance that considers both the expected repair and downtime costs over an extended period of time. However, these methods do not consider the problem of selecting diagnostic tests that provide indirect information of the health of the components. For example Dieulle et al. [8] and Grall et al. [10] use statistical models to schedule the best times for inspecting and repairing deteriorating components. Weber and Jouffe [43] use dynamic Bayesian networks to model a deteriorating system. These are Object Oriented Bayesian networks [17] which allows efficient modeling of large modular systems. Neil and Marquez [24] use hybrid Bayesian networks supporting both continuous and discrete variables where an approximate inference algorithm is used where continuous variables are discretized dynamically.

In the area of model-based diagnosis there are methods where DBN:s are used. Weber et al. [44] combines models of component deterioration with a model describing the relations between faults and symptoms using probabilities of having false positive and false negative test outcomes. In difference to the method presented in this paper, this method assumes that the symptom probabilities are non-persistent and independent between time steps. The computational issues of making inference in larger systems are also not considered. Roychoudhury et al. [32] use DBN:s generated from a bond-graph model describing the physical behavior to create diagnostic models for online diagnosis. To make the inference tractable, a factoring method is used to create independent clusters of variables by replacing certain state variables with functions of the observed variables. It differs from our offline event-based approach in that it is an online diagnostic method where the time between time steps is constant and all observable variables are observed at each time step. Pernestål et al. [29] and Pernestål et al. [30] use nsDBNs to model an event-driven troubleshooting process similar to the process in this paper. Repair, observation, and operation events are supported and symptom behavior can be both persistent and non-persistent. Tractability of inference is achieved by transforming the nsDBN into a static BN that changes structure with each event. However the method is intended for troubleshooting in the workshop only and the presence of faults is assumed to be constant unless repaired. To achieve the model transformation it is also required that there is at least one operation event between consecutive repair events.

11 Conclusion

In this paper we have presented a novel modeling framework for an off-board diagnostic troubleshooting system for heavy vehicles such as trucks and buses. It is intended to be used with a planner that decides which actions to perform to troubleshoot a vehicle that has problems. It selects actions with regard to both repair and vehicle downtime costs. To function, the planner needs to be able to know the probability of possible diagnoses given the current state of information and the probabilities of outcomes of tentative tests. This is done by the diagnoser.

The diagnoser uses a statistical model represent dependencies between faults and events including operation, repair of components, and observations. The observations are discrete and can be symptoms experienced by a driver or mechanic, diagnostic test results, and output from the On-Board Diagnostic system. In the domain, labeled data correlating faults with observations is scarce. Therefore the diagnostic model is created manually by domain experts and then parametrized using statistical data in the form of warranty data. The modeling framework allows the modeler to define variables for components, faults, symptoms, and observations and the causal dependencies between these in order to get a dynamic Bayesian network model of the troubleshooting process. We showed a model of a fuel injection system that has been created in the modeling framework.

The diagnoser functions by making Bayesian inference in this network. It is not possible to make exact inference in a model of reasonable size in reasonable time. We demonstrated how the Boyen-Koller approximation can be used for the modeling framework. This makes the time complexity of inference sufficiently small, but in certain cases this approximation introduces intolerable errors. Because the probability of having a single fault is many times greater than the probability of having multiple faults, we proposed a novel inference method that can be applied to this framework that is more precise without adding additional time complexity. The novel inference method combines a single fault model with a multiple fault model based on the Boyen-Koller approximation to be able to handle both single and multiple fault cases. We showed this both theoretically and experimentally on random diagnostic models and a model of a diesel fuel injection system.

Even though data is scarce, we want to be able to use the data we do have to tune the model parameters provided by the expert using data. We demonstrated a Bayesian learning method to learn key model parameters in the framework using warranty data. We showed both theoretically and experimentally how the parameters can be learned. We saw that data from realistic vehicle population sizes can correct some errors in the model parameters. We also showed how errors in the model parameters affect the output of the diagnoser and that the error in this output is small for small parameter errors that can remain after learning.

Acknowledgements

This work is supported in part by Scania CV AB and grants from FFI - Strategic Vehicle Research and Innovation, the Excellence Center at Linköping and Lund in Information Technology (ELLIIT), and the Research Council (VR) Linnaeus Center CADICS.

Appendix A Proofs

A.1 Proof of Theorem 1

We need to prove Eqs. (9)–(12). We can identify Eq. (9) directly using Eq. (5). For Eq. (10), by summing over F_i^t and $S_{i,j}^t$ and using Eqs. (2) and (5), we get:

$$
\begin{aligned}
&\Pr(s_{i,j}^{t+1} | f_i^{t+1}, \neg f_i^{t-1}) = \Pr(s_{i,j}^{t+1}, f_i^{t+1} | \neg f_i^{t-1}) / \Pr(f_i^{t+1} | \neg f_i^{t-1}) \\
&= \frac{1}{\Pr(f_i^{t+1} | \neg f_i^{t-1})} \Big(\Pr(f_i^{t+1} | \neg f_i^t) \Pr(s_{i,j}^{t+1} | \neg f_i^t, f_i^{t+1}) \Pr(\neg f_i^t | \neg f_i^{t-1}) \\
&\quad + \Pr(s_{i,j}^{t+1} | f_i^t, s_{i,j}^t) \Pr(s_{i,j}^t | f_i^t, \neg f_i^{t-1}) \Pr(f_i^t | \neg f_i^{t-1}) \\
&\quad + \Pr(s_{i,j}^{t+1} | f_i^t, \neg s_{i,j}^t) \Pr(\neg s_{i,j}^t | f_i^t, \neg f_i^{t-1}) \Pr(f_i^t | \neg f_i^{t-1}) \Big)
\end{aligned}
$$

and inserting Eqs. (2), (4), (6)–(7):

$$
\begin{aligned}
&= \frac{1}{1 - e^{-\lambda_{\mathsf{fail},i}(\Delta_t + \Delta_{t+1})}} \Big((1 - e^{-\lambda_{\mathsf{fail},i}\Delta_{t+1}}) p_{i,j} e^{-\lambda_{\mathsf{fail},i}\Delta_t} \\
&\quad + \big(e^{-\lambda_{i,j}\Delta_{t+1}} + p_{i,j}(1 - e^{-\lambda_{i,j}\Delta_{t+1}})\big) p_{i,j} (1 - e^{-\lambda_{\mathsf{fail},i}\Delta_t}) \\
&\quad + p_{i,j}(1 - e^{-\lambda_{i,j}\Delta_{t+1}})(1 - p_{i,j})(1 - e^{-\lambda_{\mathsf{fail},i}\Delta_t}) \Big)
\end{aligned}
$$

then gather and cancel out terms:

$$
\begin{aligned}
&= \frac{p_{i,j}}{1 - e^{-\lambda_{\mathsf{fail},i}(\Delta_t + \Delta_{t+1})}} \Big(\cancel{e^{-\lambda_{\mathsf{fail},i}\Delta_t}} - e^{-\lambda_{\mathsf{fail},i}(\Delta_t + \Delta_{t+1})} + \cancel{e^{-\lambda_{i,j}\Delta_{t+1}}} \\
&\quad - \cancel{e^{-\lambda_{i,j}\Delta_{t+1} - \lambda_{i,j}\Delta_t}} + \cancel{p_{i,j}} - \cancel{p_{i,j} e^{-\lambda_{\mathsf{fail},i}\Delta_t}} - \cancel{p_{i,j} e^{-\lambda_{i,j}\Delta_{t+1}}} \\
&\quad + \cancel{p_{i,j} e^{-\lambda_{i,j}\Delta_{t+1} - \lambda_{i,j}\Delta_t}} + 1 - \cancel{e^{-\lambda_{i,j}\Delta_{t+1}}} - \cancel{p_{i,j}} + \cancel{p_{i,j} e^{-\lambda_{i,j}\Delta_{t+1}}} \\
&\quad - \cancel{e^{-\lambda_{\mathsf{fail},i}\Delta_t}} + \cancel{e^{-\lambda_{i,j}\Delta_{t+1} - \lambda_{i,j}\Delta_t}} + \cancel{p_{i,j} e^{-\lambda_{\mathsf{fail},i}\Delta_t}} - \cancel{p_{i,j} e^{-\lambda_{i,j}\Delta_{t+1} - \lambda_{i,j}\Delta_t}} \Big) \\
&= p_{i,j}.
\end{aligned}
$$

For Eq. (12), using Eqs. (8) and (7), we get:

$$\begin{aligned}
\Pr(s^{t+1}_{i,j}|f^{t-1}_i,s^{t-1}_i) &= \Pr(s^{t+1}_{i,j}|f^{t}_i,s^{t}_i)\Pr(s^{t}_{i,j}|f^{t-1}_i,s^{t-1}_i) \\
&\quad + \Pr(s^{t+1}_{i,j}|f^{t}_i,\neg s^{t}_i)\Pr(\neg s^{t}_{i,j}|f^{t-1}_i,s^{t-1}_i) \\
&= \big(e^{-\lambda_{i,j}\Delta_{t+1}} + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t+1}})\big)\big(e^{-\lambda_{i,j}\Delta_{t}} + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t}})\big) \\
&\quad + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t+1}})\Big(1-\big(e^{-\lambda_{i,j}\Delta_{t}} + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t}})\big)\Big) \\
&= e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})} + p_{i,j}\big(1-e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})}\big).
\end{aligned}$$

For Eq. (11), using Eqs. (8) and (7), we get:

$$\begin{aligned}
\Pr(s^{t+1}_{i,j}|f^{t-1}_i,\neg s^{t-1}_i) &= \Pr(s^{t+1}_{i,j}|f^{t}_i,s^{t}_i)\Pr(s^{t}_{i,j}|f^{t-1}_i,\neg s^{t-1}_i) \\
&\quad + \Pr(s^{t+1}_{i,j}|f^{t}_i,\neg s^{t}_i)\Pr(\neg s^{t}_{i,j}|f^{t-1}_i,\neg s^{t-1}_i) \\
&= \big(e^{-\lambda_{i,j}\Delta_{t+1}} + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t+1}})\big)\big(p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t}})\big) \\
&\quad + p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t+1}})\big(1-p_{i,j}(1-e^{-\lambda_{i,j}\Delta_{t}})\big) \\
&= p_{i,j}\big(1-e^{-\lambda_{i,j}(\Delta_t+\Delta_{t+1})}\big).
\end{aligned}$$

A.2 Proof of Theorem 2

By inserting Eqs. (38)–(40) into Eqs. (36) and (37) we get

$$p_{\varphi}(s',i) = \alpha_i + \beta_i \tag{48}$$

$$p_{\mathsf{MF}}(s') = \alpha_{\mathsf{MF}} + \beta_{\mathsf{MF}} \tag{49}$$

where

$$\begin{aligned}
\alpha_i &= e^{-\Lambda_{\mathsf{fail}}\Delta}\Pr(N^t=0)\frac{1-e^{-\lambda_{\mathsf{fail},i}\Delta}(1-p_{\mathsf{f}}(s,i))}{e^{-\lambda_{\mathsf{fail},i}\Delta}(1-p_{\mathsf{f}}(s,i))} \\
\beta_i &= e^{-\Lambda_{\mathsf{fail}}\Delta}\left(\frac{\varepsilon_i}{e^{-\lambda_{\mathsf{fail},i}\Delta}} + \frac{1-e^{-\lambda_{\mathsf{fail},i}\Delta}}{e^{-\lambda_{\mathsf{fail},i}\Delta}}\varepsilon_{\mathsf{NF}}\right) \\
\alpha_{\mathsf{MF}} &= 1-e^{-\Lambda_{\mathsf{fail}}\Delta}\Pr(N^t=0)\left(1+\sum_{j\in\mathcal{I}_{\mathsf{F}}}\frac{1-e^{-\lambda_{\mathsf{fail},j}\Delta}(1-p_{\mathsf{f}}(s,j))}{e^{-\lambda_{\mathsf{fail},j}\Delta}(1-p_{\mathsf{f}}(s,j))}\right) \\
\beta_{\mathsf{MF}} &= -e^{-\Lambda_{\mathsf{fail}}\Delta}\left(\varepsilon_{\mathsf{NF}} + \sum_{j\in\mathcal{I}_{\mathsf{F}}}\left(\frac{1-e^{-\lambda_{\mathsf{fail},j}\Delta}}{e^{-\lambda_{\mathsf{fail},j}\Delta}}\varepsilon_{\mathsf{NF}} + \frac{\varepsilon_j}{e^{-\lambda_{\mathsf{fail},j}\Delta}}\right)\right)
\end{aligned}$$

From Eq. (35) we get

$$\begin{aligned}
\Pr(\tilde{f}^{t+1}_i|N^{t+1}>1) &= p_{\mathsf{f}}(s',i)\frac{1-\frac{\Pr(N^{t+1}=0)}{1-p_{\mathsf{f}}(s',i)}}{\Pr(N^{t+1}>1)} \\
&= \frac{1-e^{-\lambda_{\mathsf{fail},i}\Delta}\big(1-p_{\mathsf{f}}(s,i)\big)-\alpha_i}{\alpha_{\mathsf{MF}}}
\end{aligned} \tag{50}$$

By inserting Eqs. (48)–(50) into Eq. (33) we get

$$\begin{aligned}\bar{p}_{\mathsf{f}}(s',i) &= p_{\varphi}(s',i) + p_{\mathsf{MF}}(s',i)\Pr(\tilde{f}_i^{t+1}|N^{t+1}{>}1)\\ &= \alpha_i + \beta_i + \alpha_{\mathsf{MF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1}{>}1) + \beta_{\mathsf{MF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1}{>}1)\\ &= 1 - e^{-\lambda_{\mathsf{fail},i}\Delta}\big(1 - p_{\mathsf{f}}(s,i)\big) + \beta_i + \beta_{\mathsf{MF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1}{>}1) \qquad (51)\end{aligned}$$

Using Eqs. (2), (3), and (38)–(40), identify that

$$\begin{aligned}\Pr(f_i^{t+1}|N^t > 1) &= 1 - e^{-\lambda_{\mathsf{fail},i}\Delta}\big(1 - \Pr(f_i^t|N^t > 1)\big)\\ &= 1 - e^{-\lambda_{\mathsf{fail},i}\Delta}\big(1 - p_{\mathsf{f}}(s,i)\big)\\ &\quad + e^{-\lambda_{\mathsf{fail},i}\Delta}\big(\varepsilon_i + \varepsilon_{\mathsf{MF}}\Pr(\tilde{f}_i^t|N^t > 1)\big) \qquad (52)\end{aligned}$$

In the case that $\Pr(f_i^{t+1}|N^t > 1) \geq \bar{p}_{\mathsf{f}}(s',i)$ we get by inserting Eqs. (51) and (52) into Eq. (41):

$$\begin{aligned}\big|\bar{p}_{\mathsf{f}}(s',i) - \Pr(f_i^{t+1}|N^t{>}1)\big| &= -\beta_i - \beta_{\mathsf{MF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1)\\ &\quad + e^{-\lambda_{\mathsf{fail},i}\Delta}\big(\varepsilon_i + \varepsilon_{\mathsf{MF}}\Pr(\tilde{f}_i^t|N^t > 1)\big) \qquad (53)\\ &= \gamma_{\mathsf{NF}} + \gamma_{\mathsf{MF}} + \sum_{j\in\mathcal{I}_{\mathsf{F}}}\gamma_j \qquad (54)\end{aligned}$$

where γ_{NF}, γ_{MF}, and γ_j are obtained by expanding β_i and β_{MF} in Eq. (53) and collecting the terms for $\varepsilon_{\mathsf{NF}}$, $\varepsilon_{\mathsf{MF}}$, and ε_j:

$$\gamma_{\mathsf{NF}} = \varepsilon_{\mathsf{NF}} e^{-\Lambda_{\mathsf{fail}}\Delta}\left(\left(1 + \sum_{j\in\mathcal{I}_{\mathsf{F}}}\frac{1 - e^{-\lambda_{\mathsf{fail},j}\Delta}}{e^{-\lambda_{\mathsf{fail},j}\Delta}}\right)\Pr(\tilde{f}_i^{t+1}|N^{t+1}{>}1) - \frac{1 - e^{-\lambda_{\mathsf{fail},i}\Delta}}{e^{-\lambda_{\mathsf{fail},i}\Delta}}\right),$$

$$\gamma_{\mathsf{MF}} = \varepsilon_{\mathsf{MF}} e^{-\lambda_{\mathsf{fail},i}\Delta}\Pr(\tilde{f}_i^t|N^t > 1),$$

$$\gamma_j = \begin{cases}\varepsilon_i\left(e^{-\lambda_{\mathsf{fail},i}\Delta} - \frac{e^{-\Lambda_{\mathsf{fail}}\Delta}}{e^{-\lambda_{\mathsf{fail},i}\Delta}}\big(1 - \Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1)\big)\right) & \text{if } j = i,\\ \varepsilon_j \frac{e^{-\Lambda_{\mathsf{fail}}\Delta}}{e^{-\lambda_{\mathsf{fail},j}\Delta}}\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1) & \text{if } j \neq i.\end{cases}$$

By identifying the following inequalities:

$$\begin{array}{ccccccc} 0 & \leq & e^{-\Lambda_{\mathsf{fail}}\Delta} & \leq & e^{-\lambda_{\mathsf{fail},j}\Delta} & \leq & 1\\ e^{-\Lambda_{\mathsf{fail}}\Delta} & \leq & \multicolumn{3}{c}{e^{-\Lambda_{\mathsf{fail}}\Delta}\left(1 + \sum_{j\in\mathcal{I}_{\mathsf{F}}}\frac{1-e^{-\lambda_{\mathsf{fail},j}\Delta}}{e^{-\lambda_{\mathsf{fail},j}\Delta}}\right)} & \leq & 1\\ 0 & \leq & \multicolumn{3}{c}{\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1)} & \leq & 1\\ 0 & \leq & \multicolumn{3}{c}{\Pr(\tilde{f}_i^t|N^t > 1)} & \leq & 1\end{array}$$

we get

$$\begin{aligned}\gamma_{\mathsf{NF}} &\leq \begin{cases}\varepsilon_{\mathsf{NF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1) & \text{if } \varepsilon_{\mathsf{NF}} \geq 0\\ \varepsilon_{\mathsf{NF}}\big(\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1) - 2(1 - e^{-\Lambda_{\mathsf{fail}}\Delta})\big) & \text{if } \varepsilon_{\mathsf{NF}} \leq 0\end{cases}\\ &= \varepsilon_{\mathsf{NF}}\Pr(\tilde{f}_i^{t+1}|N^{t+1} > 1) + 2(1 - e^{-\Lambda_{\mathsf{fail}}\Delta})|\varepsilon_{\mathsf{NF}}| \qquad (55)\end{aligned}$$

$$\begin{aligned}
\gamma_{\mathsf{MF}} \leq &\begin{cases}
\varepsilon_{\mathsf{MF}} \Pr(f_i^{t+1}|N^{t+1}>1) \\
\qquad\qquad\qquad \text{if } \varepsilon_{\mathsf{NF}} \geq 0, \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \geq \Pr(\tilde{f}_i^{t}|N^{t}>1) \\
\varepsilon_{\mathsf{MF}} \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) + \Pr(\tilde{f}_i^{t}|N^{t}>1) - \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \big) \\
\qquad\qquad\qquad \text{if } \varepsilon_{\mathsf{NF}} \geq 0, \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \leq \Pr(\tilde{f}_i^{t}|N^{t}>1) \\
\varepsilon_{\mathsf{MF}} \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) - (1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \\
\qquad + \Pr(\tilde{f}_i^{t}|N^{t}>1) - \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \big) \\
\qquad\qquad\qquad \text{if } \varepsilon_{\mathsf{NF}} \leq 0, \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \geq \Pr(\tilde{f}_i^{t}|N^{t}>1) \\
\varepsilon_{\mathsf{MF}} \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) - (1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \big) \\
\qquad\qquad\qquad \text{if } \varepsilon_{\mathsf{NF}} \leq 0, \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) \leq \Pr(\tilde{f}_i^{t}|N^{t}>1)
\end{cases} \\
&= \varepsilon_{\mathsf{MF}} \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) + |\varepsilon_{\mathsf{MF}}|(1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \\
&\quad + |\varepsilon_{\mathsf{MF}}||\Pr(\tilde{f}_i^{t}|N^{t}>1) - \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1)| \qquad (56)
\end{aligned}$$

$$\begin{aligned}
\gamma_j &\leq \begin{cases}
\varepsilon_j \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) + (1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \big) & \text{if } \varepsilon_j \geq 0, j=i \\
\varepsilon_j \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) - 2(1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \big) & \text{if } \varepsilon_j \leq 0, j=i \\
\varepsilon_j \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) & \text{if } \varepsilon_j \geq 0, j \neq i \\
\varepsilon_j \big(\Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) - (1-e^{-\Lambda_{\mathsf{fail}}\Delta}) \big) & \text{if } \varepsilon_j \leq 0, j \neq i
\end{cases} \\
&= \begin{cases}
\varepsilon_j \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) + 2(1-e^{-\Lambda_{\mathsf{fail}}\Delta})|\varepsilon_j| & \text{if } j=i \\
\varepsilon_j \Pr(\tilde{f}_i^{t+1}|N^{t+1}>1) + (1-e^{-\Lambda_{\mathsf{fail}}\Delta})|\varepsilon_j| & \text{if } j \neq i
\end{cases} \qquad (57)
\end{aligned}$$

Because $p_{\mathsf{NF}}(s) + p_{\mathsf{MF}}(s) \sum_{j \in \mathcal{I}_{\mathsf{F}}} p_{\varphi}(s,j) = 1$,

$$\varepsilon_{\mathsf{MF}} + \varepsilon_{\mathsf{NF}} + \sum_{j \in \mathcal{I}_{\mathsf{F}}} \varepsilon_j = 0. \qquad (58)$$

We get the final result Eq. (41) by using Eqs. (55)–(58) on Eq. (54). The bound is the same when $\Pr(f_i^{t+1}|N^t > 1) \leq \bar{p}_{\mathsf{f}}(s',i)$ because $|\bar{p}_{\mathsf{f}}(s',i) - \Pr(f_i^{t+1}|N^t > 1)|$ is an even function with regard to $\varepsilon_{\mathsf{NF}}$, $\varepsilon_{\mathsf{MF}}$, and ε_j and that $\varepsilon_{\mathsf{NF}}, \varepsilon_{\mathsf{MF}}, \varepsilon_j \in [-1,1]$.

A.3 Proof of Lemma 1

Given $p_{\mathsf{fail},i}$ and $n_{\mathsf{comp},j}$, the number of manufacturing faults

$$n^*_{\mathsf{man},i} | p_{\mathsf{fail},i}, n_{\mathsf{comp},j} \sim \text{Bernoulli}(p_{\mathsf{fail},i}).$$

The prior $p_{\mathsf{fail},i} \sim \text{Beta}(\alpha_{\mathsf{man},i}, \beta_{\mathsf{man},i})$ which is a conjugate prior to the Bernoulli distribution, therefore the posterior

$$p_{\mathsf{fail},i} | n^*_{\mathsf{man},i}, n_{\mathsf{comp},j} \sim \text{Beta}(\alpha_{\mathsf{man},i} + n^*_{\mathsf{man},i}, \beta_{\mathsf{man},i} + n_{\mathsf{comp},j} - n^*_{\mathsf{man},i})$$

where the expected value

$$E(p_{\mathsf{fail},i}|n^*_{\mathsf{man},i}, n_{\mathsf{comp},j}) = \frac{\alpha_{\mathsf{man},i} + n^*_{\mathsf{man},i}}{\alpha_{\mathsf{man},i} + \beta_{\mathsf{man},i} + n_{\mathsf{comp},j}}.$$

When faults are assumed to be reported and repaired immediately when they occur, the number of failures

$$n^*_{\mathsf{fail},i}|\lambda_{\mathsf{fail},i}, n_{\mathsf{mil},j} \sim \text{Poisson}(\lambda_{\mathsf{fail},i} n_{\mathsf{mil},j}).$$

The prior $\lambda_{\mathsf{fail},i} \sim \text{Gamma}(\alpha_{\mathsf{fail},i}, \beta_{\mathsf{fail},i})$ which is conjugate prior to the Poisson distribution, therefore the posterior

$$\lambda_{\mathsf{fail},|} n^*_{\mathsf{fail},i}, n_{\mathsf{mil},j} \sim \text{Gamma}(\alpha_{\mathsf{fail},i} + n^*_{\mathsf{fail},i}, \beta_{\mathsf{fail},i} + n_{\mathsf{mil},j})$$

where the expected value

$$E(\lambda_{\mathsf{fail},i}|n^*_{\mathsf{fail},i}, n_{\mathsf{mil},j}) = \frac{\alpha_{\mathsf{fail},i} + n^*_{\mathsf{fail},i}}{\beta_{\mathsf{fail},i} + n_{\mathsf{mil},j}}.$$

A.4 Proof of Theorem 3

Let $p^*_{\mathsf{an},j}$ be the true probability that a fault on component j is analyzed and let $\bar{m}_j = n_{\mathsf{mil},j}/n_{\mathsf{comp},j}$ be the mean mileage at the end of the warranty period for vehicles with component j. Then for all $i \in \mathcal{I}_{\mathsf{F}|j}$, the values of $n_{\mathsf{mil},j}$, $n_{\mathsf{man},i}$, $n_{\mathsf{fail},i}$, and $n_{\mathsf{miss},j}$ divided by $n_{\mathsf{comp},j}$, and $p_{\mathsf{an},j}$ converges in probability as follows:

$$\lim_{n_{\mathsf{comp},j}\to\infty} \frac{n_{\mathsf{mil},j}}{n_{\mathsf{comp},j}} = \bar{m}_j \tag{59}$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} \frac{n_{\mathsf{man},i}}{n_{\mathsf{comp},j}} = p_{\mathsf{an},j} p^*_{\mathsf{fail},i} \tag{60}$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} \frac{n_{\mathsf{fail},i}}{n_{\mathsf{comp},j}} = p_{\mathsf{an},j} \lambda^*_{\mathsf{fail},i} \bar{m}_j \tag{61}$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} \frac{n_{\mathsf{miss},i}}{n_{\mathsf{comp},j}} = (1 - p^*_{\mathsf{an},j})\Big(\sum_{i\in\mathcal{I}_{\mathsf{F}|j}} p^*_{\mathsf{fail},i} + \lambda^*_{\mathsf{fail},i}\bar{m}_j\Big) \tag{62}$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} p_{\mathsf{an},i} = p^*_{\mathsf{an},j}. \tag{63}$$

By applying Eqs. (59)–(61) and (63) on Eqs. (44) and (45) we get:

$$\lim_{n_{\mathsf{comp},j}\to\infty} \tilde{p}_{\mathsf{fail},i} = p^*_{\mathsf{fail},i} \tag{64}$$

$$\lim_{n_{\mathsf{comp},j}\to\infty} \tilde{\lambda}_{\mathsf{fail},i} = \lambda^*_{\mathsf{fail},i} \tag{65}$$

By applying Eqs. (59)–(65) on Eqs. (46) and (47) we get the final result.

A.5 Proof of Theorem 4

Because each data sample is independent, the posterior can be formulated as the following product using Bayes' rule:

$$\Pr(p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i} | \tilde{\mathbf{y}}) \propto \prod_{m \in \mathcal{I}_{\mathbf{y}}(n), n \in \mathcal{I}_{\mathsf{F}}} \Pr(y^{[m]} | p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) \prod_{k \in \mathcal{I}_{\mathsf{S}|i}} p(p_{i,k}). \tag{66}$$

For each sample $y^{[m]}$ we know that $f_{\mathsf{id}}^{[m]}$ was present as the assumed only fault and $\mathbf{o}^{[m]}$ was observed. Unless $f_{\mathsf{id}}^{[m]} = i$ the observation probability is independent of the symptom-cause probabilities $p_{i,j}$ for all symptoms $j \in \mathcal{I}_{\mathsf{S}|i}$. We know nothing of what has happened or been observed during the time before the fault was discovered, so if $f_{\mathsf{id}}^{[m]} = i$, then

$$\begin{aligned} \Pr(y^{[m]} | p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) &= \Pr(\mathbf{o}^{[m]} | v_{\mathsf{id}}^{[m]}, m^{[m]}, c_{\mathsf{id}}^{[m]}, f_{\mathsf{id}}^{[m]}, f_{\mathsf{man}}^{[m]}, p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) \\ &\qquad \Pr(v_{\mathsf{id}}^{[m]}, m^{[m]}, c_{\mathsf{id}}^{[m]}, f_{\mathsf{id}}^{[m]}, f_{\mathsf{man}}^{[m]} | p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) \\ &\propto \Pr(\mathbf{o}^{[m]} | f_{\mathsf{id}}^{[m]}, p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}) \\ &\propto \Pr(\mathbf{o}^{[m]} | f_i^t, \neg f_n^t \forall n \in \mathcal{I}_{\mathsf{F}} \setminus \{i\}, p_{i,j} \forall j \in \mathcal{I}_{\mathsf{S}|i}). \end{aligned} \tag{67}$$

We get the final result by inserting Eq. (67) into Eq. (66).

References

[1] Apache Commons. Commons Math: The Apache Commons Mathematics Library. http://commons.apache.org/proper/commons-math/ (accessed September 30, 2014), 2014.

[2] Xavier Boyen and Daphne Koller. Tractable inference for complex stochastic processes. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 33–42, 1998.

[3] Xavier Boyen and Daphne Koller. Exploiting the architecture of dynamic systems. In *Proceedings of the 11th Innovative Applications of Artificial Intelligence Conference on Artificial Intelligence (IAAI-99)*, pages 313–320, 1999.

[4] Johan de Kleer and Brian C. Williams. Diagnosing Multiple Faults. *Artificial Intelligence*, 32(1):97–130, 1987.

[5] Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3):142–150, 1990.

[6] Decision Systems Laboratory of University of Pittsburg. SMILE reasoning engine and GeNie modeling environment. http://genie.sis.pitt.edu (accessed May 6, 2014), 2014.

[7] DFF International. RHA Cost Tables 2014. http://dffintl.co.uk/Cost_Tables_2014.pdf (accessed May 6, 2014), 2014.

[8] Laurence Dieulle, Cristophe Bérenguer, Antoine Grall, and Michel Roussignol. Sequential condition-based maintenance scheduling for a deteriorating system. *European Journal of Operational Research*, 150(2):451 – 461, 2003.

[9] Alexander Feldman, Gregory Provan, and Arjan Van Gemund. FRACTAL: Efficient fault isolation using active testing. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'09)*, pages 778–784, 2009.

[10] Antoine Grall, Cristophe Bérenguer, and Laurence Dieulle. A condition-based maintenance policy for stochastically deteriorating systems . *Reliability Engineering & System Safety*, 76(2):167 – 180, 2002.

[11] Anton Gustafsson and Sebastian Wassberg. Ekonomiska konsekvenser till följd av oplanerade stillestånd – En multipel fallstudie av företag i den svenska åkeribranschen. Master's thesis, Linköping University, 2013.

[12] David Heckerman. A Tractable Inference Algorithm for Diagnosing Multiple Diseases. In *Proceedings of the 5th Annual Conference on Uncertainty in Artificial Intelligence*, pages 163–172, 1990.

[13] David Heckerman, John S. Breese, and Koos Rommelse. Decision-Theoretic Troubleshooting. *Communications of the ACM*, 38(3):49–57, 1995.

[14] Max Henrion. Some Practical Issues in Constructing Belief Networks. In *Proceedings of the 3rd Conference on Uncertainty in Artificial Intelligence (UAI'87)*, 1987.

[15] Rolf Isermann. Model-based fault detection and diagnosis: status and applications. In *Proceedings of the 16th IFAC Symposium on Automatic Control in Aerospace (ACA'04)*, 2004.

[16] Frank Kimmich, Anselm Schwarte, and Rolf Isermann. Fault detection for modern Diesel engines using signal-and process model-based methods. *Control Engineering Practice*, 13(2):189–203, 2005.

[17] Daphne Koller and Avi Pfeffer. Object-oriented Bayesian networks. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI'97)*, pages 302–313, 1997.

[18] Helge Langseth and Finn V. Jensen. Decision theoretic troubleshooting of coherent systems. *Reliability Engineering & System Safety*, 80(1):49–62, 2002.

[19] Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224, 1988.

[20] MAN Truck International. MAN service agreements. http://www.truck.man.eu/global/en/services-and-parts/maintenance-and-parts/service-contracts/Service-contracts.html (accessed May 6, 2014), 2014.

[21] Kevin Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, UC Berkeley, USA, July 2002.

[22] Kevin Murphy and Stuart Russell. Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Sequential Monte Carlo methods in practice*, pages 499–515. 2001.

[23] Peter Nabende and Tom Wanyama. An expert system for diagnosing heavy-duty diesel engine faults. In *Advances in Computer and Information Sciences and Engineering*, pages 384–389. 2008.

[24] Martin Neil and David Marquez. Availability modelling of repairable systems using bayesian networks. *Engineering Applications of Artificial Intelligence*, 25(4):698 – 704, 2012.

[25] Thorsten J. Ottosen and Finn V. Jensen. The Cost of Troubleshooting Cost Clusters with Inside Information. In *26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, 2010.

[26] Thorsten J. Ottosen and Finn V. Jensen. When to test? Troubleshooting with postponed system test. *Expert Systems with Applications*, 38(10):12142 – 12150, 2011.

[27] John Pandremenos, John Paralikas, Konstantinos Salonitis, and George Chryssolouris. Modularity concepts for the automotive industry: A critical review. *CIRP Journal of Manufacturing Science and Technology*, 1(3): 148 – 152, 2009.

[28] Judea Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, 1985.

[29] Anna Pernestål, Håkan Warnquist, and Mattias Nyberg. Modeling and Troubleshooting with Interventions Applied to an Auxiliary Truck Braking System. In *Proceedings of the 2nd IFAC Workshop on Dependable Control of Discrete Systems (DCDS'09)*, 2009.

[30] Anna Pernestål, Håkan Warnquist, and Mattias Nyberg. Modeling and inference for troubleshooting with interventions applied to a heavy truck auxiliary braking system. *Engineering Applications of Artificial Intelligence*, 25 (4):705 – 719, 2012.

[31] Joshua W. Robinson and Alexander J. Hartemink. Non-stationary dynamic Bayesian networks. In *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS'08)*, pages 1369–1376, 2008.

[32] Indranil Roychoudhury, Gautam Biswas, and Xenofon Koutsoukos. Distributed Diagnosis of Dynamic Systems Using Dynamic Bayesian Networks. In *Proceedings of the 20th International Workshop on the Principles of Diagnosis (DX'09)*, 2009.

[33] Indranil Roychoudhury, Gautam Biswas, and Xenofon Koutsoukos. Designing distributed diagnosers for complex continuous systems. *Automation Science and Engineering, IEEE Transactions on*, 6(2):277–290, 2009.

[34] Scania CV. Scania Repair & Maintenance contract. http://www.scania.com/products-services/services/workshop-services/ (accessed May 6, 2014), 2014.

[35] Scania CV. Scania Repair & Maintenance contract. http://www.scania.co.uk/services/support-programmes/repair-maintenance/ (accessed May 6, 2014), 2014.

[36] Anthony M. Smith. *Reliability-centered maintenance*. McGraw-Hill, 1993.

[37] SAE Standards. *J2012: Diagnostic Trouble Code definitions*. Number J2012-1307. SAE International, 2013.

[38] Ying Sun and Daniel S. Weld. A framework for model-based repair. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI'93)*, 1993.

[39] Carl Svärd and Mattias Nyberg. Residual generators for fault diagnosis using computation sequences with mixed causality applied to automotive systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(6):1310–1328, 2010.

[40] Volvo Trucks. Service and maintenance agreements. Retrieved from http://www.volvotrucks.com/trucks/global/en-gb/trucks/services (accessed May 6, 2014), 2014.

[41] Volvo Trucks. Volvo Uptime Assurance. Retrieved from http://www.volvotrucks.com/trucks/global/en-gb/trucks/services/VAS/Pages/Volvo_uptime_assurance.aspx (accessed May 6, 2014), 2014.

[42] Håkan Warnquist, Mattias Nyberg, and Jonas Biteus. Guided Integrated Remote and Workshop Troubleshooting of Heavy Trucks. *SAE International Journal of Commercial Vehicles*, 7(1):25–36, 2014.

[43] Philippe Weber and Lionel Jouffe. Complex system reliability modelling with Dynamic Object Oriented Bayesian Networks (DOOBN). *Reliability Engineering & System Safety*, 91(2):149 – 162, 2006.

[44] Philippe Weber, Didier Theilliol, Christophe Aubrun, and Alexandre Evsukoff. Increasing effectiveness of model-based fault diagnosis: A dynamic Bayesian network design for decision making. In *Proceedings of the Conference on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS'07)*, 2007.

[45] Changhe Yuan and Marek J. Druzdzel. An importance sampling algorithm based on evidence pre-propagation. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI'02)*, pages 624–631, 2002.

[46] Alice X. Zheng, Irina Rish, and Alina Beygelzimer. Efficient test selection in active diagnosis via entropy approximation. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI'05)*, pages 675–682, 2005.

[47] Geoffrey Zweig. A forward-backward algorithm for inference in Bayesian networks and an empirical comparison with HMMs. Master's thesis, Deptartment of Computer Science, UC Berkeley, 1996.