A Framework For Assessing Diagnostics Model Fidelity

Gregory Provan[1] and **Alex Feldman**[2]
[1]Computer Science Department, University College Cork, Cork, Ireland
e-mail: g.provan@cs.ucc.ie
[2]PARC Inc., Palo Alto, CA 94304, USA
e-mail: afeldman@parc.com

Abstract

"All models are wrong but some are useful" [1]. We address the problem of identifying which diagnosis models are more useful than others. Models are critical to diagnostics inference, yet little work exists to be able to compare models. We define the role of models in diagnostics inference, propose metrics for models, and apply these metrics to a tank benchmark system. Given the many approaches possible for model metrics, we argue that only information-theoretic methods address how well a model mimics real-world data. We focus on some well-known information-theoretic modelling metrics, demonstrating the trade-offs that can be made on different models for a tank benchmark system.

1 Introduction

A core goal of Model-Based Diagnostics (MBD) is to accurately diagnose a range of systems in real-world applications. There has been significant progress in developing algorithms for systems of increasing complexity. A key area where further work is needed is scaling-up to real-world models, as multiple-fault diagnostics algorithms are currently limited by the size and complexity of the models to which they can be applied. In addition, there is still a great need for defining metrics to measure diagnostics accuracy, and to measure the computational complexity of inference and of the models' contribution to inference complexity.

This article addresses the modeling side of MBD: we focus on methods for measuring the size and complexity of MBD models. We explore the role that diagnostics model fidelity can play in being able to generate accurate diagnostics. We characterise model fidelity and examine the trade-offs of fidelity and inference complexity within the overall MBD inference task.

Model fidelity is a crucial issue in diagnostics [2]: models that are too simple can be inaccurate, yet highly detailed and complex models are expensive to create, have many parameters that require significant amounts of data to estimate, and are computationally intensive to perform inference on. There is an urgent need to incorporate inference complexity within modelling, since even relatively simple models, such as some of the combinational ISCAS-85 benchmark models, pose computational challenges to even the most advanced solvers for multiple-fault tasks. In addition, higher-fidelity models can actually perform worse than lower-fidelity models on real-world data, as can be explained using over-fitting arguments within a machine learning framework.

To our knowledge, there is no theory within Model-Based Diagnostics that relates notions of model complexity, model accuracy, and inference complexity. To address these issues, we explore several of the factors that contribute to model complexity, as well as a theoretically sound approach for selecting models based on their complexity and diagnostics performance, i.e., their accuracy in diagnosing faults.

Our contributions are as follows:

- We characterise the task of selecting a diagnosis model of appropriate fidelity as an information-theoretic model selection task.
- We propose several metrics for assessing the quality of a diagnosis model, and derive approximation versions of a subset of these metrics.
- We use a dynamical systems benchmark model to demonstrate our compare how the metrics assess models relative to the accuracy of diagnostics output based on using the models.

2 Related Work

This section reviews work related to our proposed approach.

Model-Based Diagnostics: There is some seminal work on modelling principles within the Model-Based Diagnosis (MBD) community, e.g., [2; 3]; this early work adopts an approach based on logic or qualitative physics for model specification. However, this work provides no means for comparing models in terms of diagnostics accuracy. More recent work ([4]) provides a logic-based specification of model fidelity. There is also work specifying metrics for diagnostics accuracy, e.g., [5].

However, none of this work defines precise metrics for computing both diagnostics accuracy and model complexity, and their trade-offs. This article adopts a theoretically well-founded approach for integrating multiple MBD metrics.

Multiple Fidelity Modeling There is limited work describing the use of models of multiple levels of fidelity. Examples of such work includes [6; 7; 8]. In this article we focus on methods for evaluating multi-fidelity models and their impact on diagnostics accuracy, as opposed to developing methodoligies for modelling at multiple levels of fidelity.

Multiple-Mode Modeling One approach to MBD is to use a separate model for every failure mode, rather than to define a model containing all failure modes. Examples of this approach include [9; 10; 11; 12]. Note that this work does not specify metrics for computing *both* diagnostics accuracy and model complexity, or their trade-offs.

Model- Selection The metrics that we adopt and extend have been used extensively to compare different models, e.g., [13]. The metrics are used to compare *simulation performance* of models only. In contrast, we extend this framework to examine *diagnostics performance*. In the process, we explore the use of multiple loss functions for penalising models, in addition to the standard penalty functions based on number of model parameters.

Model-Order Reduction Model-Order reduction [14] aims to reduce the complexity of a model with an aim to limit the performance losses of the reduced model. The reduction methods are theoretically well-founded, although they are highly domain-specific. In contrast to this approach, we assume a model-composition approach from a component library containing hand-constructed models of multiple levels of fidelity.

3 Diagnostics Modeling and Inference

This section formalises the notion of diagnostics model within the process of diagnostics inference. We first introduce the task, and then define it more precisely.

3.1 Diagnosis Task

Assume that we have a system $\mathcal{S}$ that can operate in a nominal state, ξ_N, or a faulty state, ξ_F, where Ξ is the set of possible states of $\mathcal{S}$. We further assume that we have a discrete vector of measurements, $\tilde{\boldsymbol{Y}} = \{\tilde{y}_1, ..., \tilde{y}_n\}$ observed at times $t = \{1, ..., n\}$ that summarizes the response of the system $\mathcal{S}$ to control variables $\boldsymbol{U} = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$. Let $\boldsymbol{Y}_\phi = \{y_1, ..., y_n\}$ denote the corresponding predictions from a dynamic (nonlinear) model, ϕ, with parameter values θ: this can be represented by $\boldsymbol{Y}_\phi = \phi(x_0, \theta, \xi, \tilde{\boldsymbol{U}})$, where x_0 signifies the initial states of the system at t_0.

We assume that we have a prior probability distribution $P(\Xi)$ over the states Ξ of the system. This distribution denotes the likelihood of the failure states of the system.

We define a residual vector $\mathcal{R}(\tilde{\boldsymbol{Y}}, \boldsymbol{Y}_\phi)$ to capture the difference between the actual and model-simulated system behaviour. An example of a residual vector is the mean-squared-error (MSE). We assume a fixed diagnosis task $\mathcal{T}$ throughout this article, e.g., computing the most likely diagnosis, or a deterministic multiple-fault diagnosis.

The classical definition of diagnosis is as a state estimation task, whose objective is to identify the system state that minimises the residual vector:

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmin}}\, \mathcal{R}(\tilde{\boldsymbol{Y}}, \boldsymbol{Y}_\phi) \tag{1}$$

Since this is a minimisation task, we typically need to run multiple simulations over the space of parameters and modes to compute ξ^*. We can abstract this process as performing model-inversion, i.e., computing some $\xi^* = \phi^{-1}(x_0, \theta, \xi, \tilde{\boldsymbol{U}})$ that minimises $\mathcal{R}(\tilde{\boldsymbol{Y}}, \boldsymbol{Y}_\phi)$.

During this diagnostics inference task, a model ϕ can play two roles: (a) simulating a behaviour to estimate $\mathcal{R}(\tilde{\boldsymbol{Y}}, \boldsymbol{Y}_\phi)$; (b) enabling the computation of $\xi^* = \phi^{-1}(x_0, \theta, \xi, \tilde{\boldsymbol{U}})$. It is clear that diagnostics inference requires a model that has good fidelity and is computationally efficient for performing these two roles.

We generalise that notion to incorporate inference efficiency as well as accuracy. We can define an inference complexity measure as $\mathcal{C}(\tilde{\boldsymbol{Y}}, \phi)$. We can then define our diagnosis task as jointly minimising a function g that incorporates the accuracy (based on the residual function) and the inference complexity:

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmin}}\, g\left(\mathcal{R}(\tilde{\boldsymbol{Y}}, \boldsymbol{Y}_\phi), \mathcal{C}(\tilde{\boldsymbol{Y}}, \phi)\right). \tag{2}$$

Here g specifies a loss or penalty function that induces a non-negative real-valued penalty based on the lack of accuracy and computational cost.

In forward simulation, a model ϕ, with parameters $\boldsymbol{\theta}$, can generate multiple observations $\tilde{\boldsymbol{Y}} = \{\tilde{y}_1, ..., \tilde{y}_n\}$. The diagnostics task involves performing the inverse operation on these observations. Our objective thus involves optimising the state estimation task over a future set of observations, $\tilde{\boldsymbol{Y}} = \{\tilde{\boldsymbol{Y}}_1, ..., \tilde{\boldsymbol{Y}}_n\}$. Our model ϕ and inference algorithm $\mathcal{A}$ have different performance based on $\tilde{\boldsymbol{Y}}_i, i = 1, ..., n$: for example, [15] shows that both inference-accuracy and -time vary based on the fault cardinality . As a consequence, to compute ξ^* we want to optimise the *mean* performance over future observations. This notion of *mean* performance optimisation has been characterised using the Bayesian model selection approach, which we examine in the following section.

3.2 Diagnosis Model

We specify a diagnosis model as follows:

Definition 1 (Diagnosis Model)**.** *We characterise a Diagnosis Model ϕ using the tuple $\langle \boldsymbol{V}, \boldsymbol{\theta}, \Xi, \mathcal{E} \rangle$, where*

- *$\boldsymbol{V}$ is a set of variables, consisting of variables denoting the system state ($\boldsymbol{X}$), control ($\boldsymbol{U}$), and observations ($\boldsymbol{Y}$).*
- *$\boldsymbol{\theta}$ is a set of parameters.*
- *Ξ is a set of system modes.*
- *$\mathcal{E}$ is a set of equations, with a subset $E_\xi \subseteq \mathcal{E}$ for each mode $\xi \in \Xi$.*

We will assume that we can use a physics-based approach to hand-generate a set $\mathcal{E}$ of equations to specify a model. Obtaining good diagnostics accuracy, given a fixed $\mathcal{E}$, entails estimating the parameters $\boldsymbol{\theta}$ to optimise that accuracy.

3.3 Running Example: Three-Tank Benchmark

In this paper, we use the three-tank system shown in Fig. 1 to illustrate our approach. The three tanks are denoted as T_1, T_2, and T_3. Each tank has the same area $A_1 = A_2 = A_3$. For $i = 1, 2, 3$, tank T_i has height h_i, a pressure sensor p_i, and a valve V_i, $i = 1, 2, 3$ that controls the flow of liquid out of T_i. We assume that gravity $g = 10$ and the liquid has density $\rho = 1$.

Tank T_1 gets filled from a pipe, with measured flow q_0. Using Torricelli's law, the model can be described by the following non-linear equations:

$$\frac{dh_1}{dt} = \frac{1}{A_1}\left[-\kappa_1\sqrt{h_1 - h_2} + q_0\right], \tag{3}$$

$$\frac{dh_2}{dt} = \frac{1}{A_2}\left[\kappa_1\sqrt{h_1 - h_2} - \kappa_2\sqrt{h_2 - h_3}\right], \tag{4}$$

$$\frac{dh_3}{dt} = \frac{1}{A_3}\left[\kappa_2\sqrt{h_2 - h_3} - \kappa_3\sqrt{h_3}\right]. \tag{5}$$

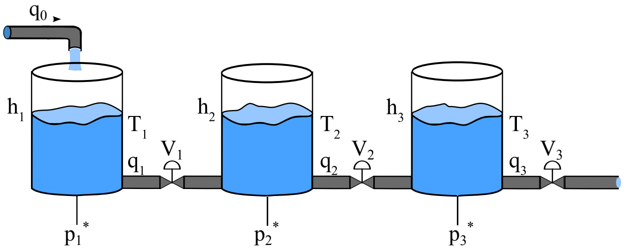

Figure 1: Diagram of the three-tank system.

In eq. 3, the coefficient κ_1 denotes a parameter that captures the product of the cross-sectional area of the tank A_1, the area of the drainage hole, a gravity-based constant ($\sqrt{2g}$), and the friction/contraction factor of the hole. κ_2 and κ_3 can be defined analogously.

Finally, the pressure at the bottom of each tank is obtained from the height: $p_i = g\,h_i$, where i is the tank index ($i \in \{1, 2, 3\}$).

We emphasize the use of the $\kappa_i,\ i = 1, 2, 3$ because we will use these parameter-values as a means for "diagnosing" our system in term of changes in $\kappa_i,\ i = 1, 2, 3$. Consider a physical valve R_1 between T_1 and T_2 that constraints the flow between the two tanks. We can say that the valve changes proportionally the cross-sectional drainage area of q_1 and hence κ_1. The diagnostic task will be to compute the true value of κ_1, given p_1, and from κ_1 we can compute the actual position of the valve R_1.

We now characterise our nominal model in terms of Definition 1:

- variables $\boldsymbol{V}$ consist of variables denoting the system state ($\boldsymbol{X} = \{h_1, h_2, h_3\}$), control ($\boldsymbol{U} = \{q_0, V_1, V_2, V_3\}$), and observations ($\boldsymbol{Y} = \{p_1, p_2, p_3\}$).
- $\boldsymbol{\theta} = \{\{A_1,\ A_2,\ A_3\}, \{\kappa_1,\ \kappa_2,\ \kappa_3\}\}$ is the set of parameters.
- Ξ consists of a single nominal mode.
- $\mathcal{E}$ is a set of equations, given by equations 3 through 5.

Note that this model has a total of 6 parameters.

Fault Model In this article we focus on valve faults, where a valve can have a blockage or a leak. We model this class of faults by including in equations 3 to 5 an additive parameter β, which is applied to the parameter κ, i.e., as $\kappa_i(1+\beta_i)$, $i = 1, 2, 3$, where $-1 \leq \beta_i \leq \frac{1}{\kappa_i} - 1$, $i = 1, 2, 3$. $\beta > 0$ corresponds to a leak, such that $\beta \in (0, 1/\kappa - 1]$; $\beta < 0$ corresponds to a blockage, such that $\beta \in [-1, 0)$. The fault equations can be written as:

$$\begin{aligned}
\frac{dh_1}{dt} &= \frac{1}{A_1}\left[-\kappa_1(1+\beta_1)\sqrt{h_1 - h_2} + q_0\right], \qquad (6)\\
\frac{dh_2}{dt} &= \frac{1}{A_2}\Big[\kappa_1(1+\beta_1)\sqrt{h_1 - h_2}\\
&\qquad - \kappa_2(1+\beta_2)\sqrt{h_2 - h_3}\Big],\\
\frac{dh_3}{dt} &= \frac{1}{A_3}\left[\kappa_2(1+\beta_2)\sqrt{h_2 - h_3} - \kappa_3(1+\beta_3)\sqrt{h_3}\right].
\end{aligned}$$

The fault equations allow faults for any combination of the valves $\{V_1, V_2, V_3\}$, resulting in system modes $\Xi = \{\xi_N, \xi_1, \xi_2, \xi_3, \xi_{12}, \xi_{13}, \xi_{23}, \xi_{123}\}$, where ξ_N is the nominal mode, and $\xi_\cdot$ is the mode where $\cdot$ denotes the combination of valves (taken from a combination of $\{1, 2, 3\}$) which are faulty. This fault model has 9 parameters.

4 Modelling Metrics

This section describes the metrics that can be applied to estimate properties of a diagnosis model. We describe two types of metrics, dealing with accuracy (fidelity) and complexity.

4.1 Model Accuracy

Model accuracy concerns the ability of a model to mimic a real system. From a diagnostics perspective, this translates to the use of a model to simulate behaviours that distinguish nominal and faulty behaviours sufficiently well that appropriate fault isolation algorithms can identify the correct type of fault when it occurs. As such, a diagnostics model needs to be able to simulate behaviours for multiple modes with "appropriate" fidelity.

Note that we distinguish model accuracy from diagnosis inference accuracy. As noted above, model accuracy concerns the ability of a model to mimic a real system through simulation, and to assist in diagnostics isolation. Diagnosis inference accuracy concerns being able to isolate the true fault given an observation and the simulation output of a model.

A significant challenge for a diagnosis model is the need to simulate behaviours for multiple modes. Two approaches that have been taken are to use a single model with multiple modes explicitly defined (a multi-mode approach), or to use multiple models [9; 16; 17], each of which is optimised for a single or small set of modes (a multi-mode**l** approach).

The AI-based MBD approach typically uses a single model ϕ with multiple modes explicitly defined [18], or a single model with just nominal behaviour [19]. From a diagnostics perspective, accuracy must be defined with respect to the task $\mathcal{T}$. We adopt here the task of computing the most-likely diagnosis.

Given evidence suggesting that model fidelity for a multi-mode approach varies depending on the mode, it is important to explicitly consider the *mean performance* of ϕ over the entire observation space $\mathcal{Y}$ (the space of possible observations of the system).

In this article we adopt the expected residual approach, i.e., given a space $\mathcal{Y} = \{\tilde{\boldsymbol{Y}}_1, ..., \tilde{\boldsymbol{Y}}_n\}$ of observations, the expected residual is the average over the n observations, e.g., as given by: $\bar{\mathcal{R}} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}(\tilde{\boldsymbol{Y}}_i, \boldsymbol{Y}_\phi)$.

4.2 Model Complexity

At present, there is no commonly-accepted definition of model complexity, whether the model is used purely for simulation or if it is used for diagnostics or control. Defining the complexity of a model is inherently tricky, due to the number of factors involved.

Less complex models are often preferred either due to their low computational simulation costs [20], or to minimise model over-fitting given observed data [21; 22]. Given the task of simulating a variable of interest conditioned by certain future values of input (control) variables, overfitting can lead to high uncertainty in creating accurate simulations. Overfitting is especially severe when we have limited observation variables for generating a model representing the underlying process dynamics. In contrast, models with low parameter dimensionality (i.e. fewer parameters) are considered less complex and hence are associated with low prediction uncertainty [23].

Several approaches have been used, based on issues like (a) number of variables [24], (b) model structure [25], (c) number of free parameters [23], (d) number of parameters that the data can constrain [26], (e) a notion of model weight [27], or (f) *type* and *order* of equations for a non-linear dynamical model [14], where type corresponds to non-linear, linear, etc.; e.g., order for a non-linear model is such that a k-th order system has k-th derivates in $\mathcal{E}$.

Factors that contribute to the true cost of a model include: (a) model-generation; (b) parameter estimation; and (c) simulation complexity, i.e., the computational expense (in terms of CPU-time and memory) needed to simulate the model given a set of initial conditions Rather than try to formulate this notion in terms of the number of model variables or parameters, or a notion of model structural complexity, we specify model complexity in terms of a measure based on parameter estimation, and inference complexity, assuming a construction cost of zero.

A thorough analysis of model complexity will need to take into consideration the model equation class, since model complexity is class-specific. For example, for nonlinear dynamical models, complexity is governed by the *type* and *order* of equations [14]. In contrast, for linear dynamical models, which have only matrices and variables in equations (no derivatives), it is the order of the matrices that determines complexity. In this article, we assume that models are of appropriate complexity, and hence do not address Model order reduction techniques [14], which aim to generate lower-dimensional systems that trade off fidelity for reduced model complexity.

4.3 Diagnostics Model Selection Task

The model in this model selection problem corresponds to a system with a single mode. Given a space Φ of possible models, we can define this model selection task as follows:

$$\phi^* = \operatorname*{argmin}_{\phi\in\Phi} g_1\left(\mathcal{R}(\tilde{\boldsymbol{Y}},\boldsymbol{Y}_\phi)\right) + g_2\left(\mathcal{C}(\tilde{\boldsymbol{Y}},\phi)\right), \qquad (7)$$

adopting the simplifying assumption that our loss function g is additively decomposable.

4.4 Information-Theoretic Model Complexity

The Information-Theoretic (or Bayesian) model complexity approach, which is based on the model likelihood, measures whether the increased “complexity" of a model with more parameters is justified by the data. The Information-Theoretic approach chooses a model (and a model structure) from a set of competing models (from the set of corresponding model structures, respectively) such that the value of a Bayesian criterion is maximized (or prediction uncertainty in choosing a model structure is minimized).

The Information-Theoretic approach addresses prediction uncertainty by specifying an appropriate likelihood function. In other words, it specifies the probability with which the observed values of a variable of interest are generated by a model. The marginal likelihood of a model structure, which represents a class of models capturing the same processes (and hence have the same parameter dimensionality), is obtained by integrating over the prior distribution of model parameters; this measures the prediction uncertainty of the model structure [28].

Statistical model selection is commonly based on Occam's parsimony principle (ca.1320), namely that hypotheses should be kept as simple as possible. In statistical terms, this is a trade-off between bias (distance between the average estimate and truth) and variance (spread of the estimates around the truth).

The idea is that by adding parameters to a model we obtain improvement in fit, but at the expense of making parameter estimates “worse"’ because we have less data (i.e., information) per parameter. In addition, the computations typically require more time. So the key question is how to identify how complex a model works best for a given problem.

If the goal is to compute the likelihood of a given model $\phi(x_0, \theta, \xi, \boldsymbol{U})$, then $\boldsymbol{\theta}$ and $\boldsymbol{U}$ are nuisance parameters. These parameters affect the likelihood calculation but are not what we want to infer. Consequently, these parameters should be eliminated from the inference. We can remove nuisance parameters by assigning them prior probabilities and integrating them out to obtain the marginal probability of the data given only the model, that is, the model likelihood (also called integrative, marginal, or predictive likelihood). In equational form, this looks like: $P(\boldsymbol{Y}|\phi) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{U}} P(\phi|\boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{U}) P(\boldsymbol{\theta}, \boldsymbol{U}|\phi) d\boldsymbol{\theta} d\boldsymbol{U}$. However, this multidimensional integral can be very difficult to compute, and it is typically approximated using computationally intensive techniques like Markov chain Monte Carlo (MCMC).

Rather than try to solve such a computationally challenging task, we adopt an approximation to the multidimensional integral. In the statistics literature several decomposable approximations have been proposed.

Spiegelhalter et al. [26] have proposed a well-known such decomposable framework, termed the Deviance Information Criterion (DIC), which measures the number of model parameters that the data can constrain.: $DIC = \overline{D} + p_D$, where $\overline{D}$ is a measure of fit (expected deviance), and p_D is a complexity measure, the *effective* number of parameters. The Akaike Information Criterion (AIC) [29; 30] is another well-known measure: $AIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + 2k$, where $\hat{\boldsymbol{\theta}}$ is the Maximum Likelihood Estimate (MLE) of $\boldsymbol{\theta}$ and k is the number of parameters.

To compensate for small sample size n, a variant of AIC, termed AIC$_c$, is typically used:

$$AIC_c = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + 2k + \frac{2k(k+1)}{(n-k-1)} \qquad (8)$$

Another computationally more tractable approach is the Bayesian Information Criterion (BIC) [31]: $BIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + klogn$, where k is the number of *estimable* parameters, and *n* is the sample size (number of observations). BIC was developed as an approximation to the log marginal likelihood of a model, and therefore, the difference between two BIC estimates may be a good approximation to the natural log of the Bayes factor. Given equal priors for all competing models, choosing the model with the smallest BIC is equivalent to selecting the model with the maximum posterior probability. BIC assumes that the (parameters') prior is the unit information prior (i.e., a multivariate normal prior with mean at the maximum likelihood estimate and variance equal to the expected information matrix for one observation).

Wagenmakers [32] shows that one can convert the BIC metric to

$$BIC = n\,log\frac{SSE}{SS_{total}} + k\,logn,$$

where SSE is the sum of squares for the error term. In our experiments, we assume that the non-linear model is the “correct" model (or the null hypothesis H_0), and either the linear or qualitative models are the competing model (or alternative hypothesis H_1). Hence what we do is use BIC to compare the non-linear to each of the competing models.

Suppose that we obtain the BIC values for the alternative and the correct models, using the relevant SS terms. When computing $\Delta_{BIC} = BIC(H_1) - BIC(H_0)$, note that both the null (H_0) and the alternative hypothesis (H_1) models share the same SS_{total} term (both models attempt to explain the same collection of scores), although they differ with respect to SSE. The SS_{total} term common to both BIC values cancels out in computing Δ_{BIC}, producing

$$\Delta_{BIC} = n\,log\frac{SSE_1}{SSE_0} + (k_1 - k_0)logn, \qquad (9)$$

where SSE_1 and SSE_0 are the sum of squares for the error terms in the alternative and the null hypothesis models, respectively.

5 Experimental Design

This section compares three tank benchmark models according to various model-selection measures. We adopt as our “correct" model the non-linear model. We will examine the fidelity and complexity tradeoffs of two simpler models over a selection of failure scenarios.

The diagnostic task will be to compute the fault state of the system, given an injected fault, which is one of (ξ_N, ξ_B, ξ_P), denoting nominal blocked and passing valves, respectively. This translates to different tasks given the different models.

non-linear model estimate the true value of κ_1 given p_1, which corresponds to a most-likely failure mode assignment of one of (ξ_N, ξ_B, ξ_P).

linear model estimate the true value of κ_1 given p_1, which corresponds to a most-likely failure mode assignment of one of (ξ_N, ξ_B, ξ_P).

qualitative model estimate the failure mode assignment of one of (ξ_N, ξ_B, ξ_P).

5.1 Alternative Models

This section describes the two alternative models that we compare to the non-linear model, a linear and a qualitative model.

Linear Model

We compare the non-linear model with a linearised version. We can perform this linearised process in a variety of ways [33]. In this simple tank example, we can perform the linearisation directly through replacement of non-linear and linear operators, as shown below.

Nominal Model We can linearise the the non-linear 3-tank model by replacing the non-linear sub-function $\sqrt{h_i - h_j}$ with the linear sub-function $\gamma_{ij}(h_i - h_j)$, where γ_{ij} is a parameter (to be estimated) governing the flow between tanks i and j. The linear model has 4 parameters, $\gamma_{12},\ \gamma_{12},\ \gamma_{23},\ \gamma_3$.

Fault Model The fault model introduces a parameter β_i associated with κ_i, i.e., we replace κ_i with $\kappa_i(1+\beta_i),\ i = 1,2,3$, where $-1 \leq \beta_i \leq \frac{1}{\kappa_i} - 1,\ i = 1,2,3$. This model has 7 parameters, adding parameters $\beta_1,\ \beta_2,\ \beta_3$.

Qualitative Model

Nominal Model For the model we replace the non-linear sub-function $\sqrt{h_i - h_j}$ with the qualitative sub-function $M^+(h_i - h_j)$, where M^+ is the set of reasonable functions f such that $f' > 0$ on the interior of its domain [34].

The tank-heights are constrained to be non-negative, as are the parameters κ_i. As a consequence, we can discretize the h_i to take on values $\{+, 0\}$, which means that $M^+(h_i - h_j)$ can take on values $\{+, 0, -\}$. The domain for $\frac{dh_1}{dt}$ must be $\{+, 0, -\}$, since the qualitative version of q_0, $\mathcal{Q}$ is non-negative (domain of $\{+, 0\}$) and each $M^+(h_i - h_j)$ can take on values $\{+, 0, -\}$. We see that this model has no parameters to estimate.

Fault Model

The qualitative fault model has different M^+ functions for the modes where the valve is passing and blocked. We derive these functions as follows. From a qualitative perspective, the domain of β_i is {0,+} for a passing valve, and {-,0} for a blocked valve. To create a new M^+ function for the cases of passing and blocked valve, we qualitatively apply these corresponding domains to the standard M^+ function with domain {-,0,+} to obtain fault-based M^+ functions : $M_P^+(h_i - h_j)$ denotes the M^+ function when the valve is passing, and $M_B^+(h_i - h_j)$ denotes the M^+ function when the valve is blocked.

5.2 Simulation Results

We have compared the simulation performance of the models under nominal and faulty conditions, considering faults to individual valves $V_1,\ V_2$ and V_3, as well as double-fault combinations of the valves. In the following we present some plots for simulations of faults and fault-isolation for different model types.

Figure 2 shows the results from a single-fault scenario, where valve V_1 is stuck at 50%) at $t = 250$, based on the non-linear model. The plot from this simulation show that at the time of the fault injection, the water level in tank T_1 starts increasing while the water level at tanks T_2 and T_3 start decreasing due to the lower inflow.

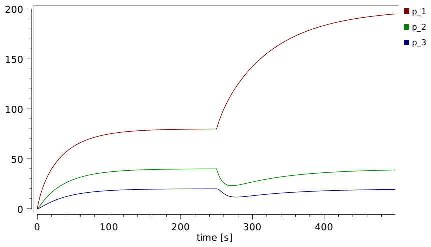

Figure 2: Simulation with non-linear model for the scenario of a fault in valve 1 at $t = 250$ s

Table 1 shows the simulation error-difference between the non-linear and linear models, for the nominal case and the faulty case (where valve 1 is faulted). Given that we measure the pressure levels for $p_1,\ p_2$ and p_3 every second, we use the difference in these outputs to identify the sum-of-squared-error (SSE) values for the simulations.

	p_1	p_2	p_3	Total
Nominal	2600.3	316.2	118.1	3034.6
V_1-fault	2583.1	347.5	137.2	3067.8

Table 1: Data for SSE values for simulations using Non-linear and Linear representations, given two scenarios: nominal and faulty (valve V_1 at 50% after 250 s)

Figure 3 shows the results for diagnosing the V_1-fault using the non-linear model. We can see that the diagnostic accuracy is high, as $P(V_1)$ converges to almost 1 with little time lag.

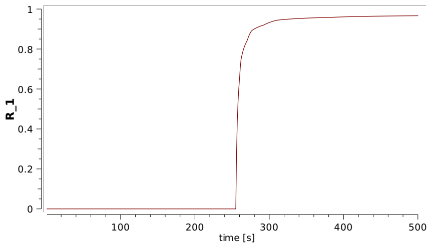

Figure 3: Simulation of fault isolation of fault in valve 1 with non-linear model. The figure depicts the probability of valve 1 being faulty.

In contrast, Figure 4 shows the diagnostic accuracy and isolation time with a linear model. First, note that there is a false-positive identified early in the simulation, and the model incorrectly identifies both valves 2 and 3 as being faulty. This linear model thus delivers both poor diagnostic accuracy (classification errors) and poor isolation time (there is a lag between when the fault occurs and when the model identifies the fault). After the fault injection at $t = 250$ [s], the predictive accuracy improves and the correct fault becomes the most likely fault.

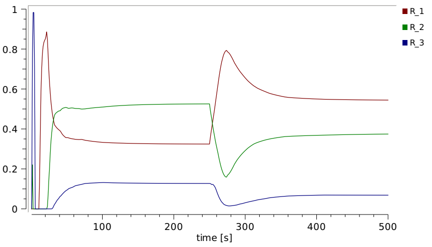

Figure 4: Simulation of fault isolation of fault in valve 1 with linear model.The figure depicts the probability of valves 1, 2 and 3 being faulty.

Figure 5 depicts the diagnostic performance with a mixed linear/non-linear model (T_1 is non-linear, while T_2 and T_3 are linear). The diagnostic accuracy is almost the same as that of the non-linear model (cf. Figure 3), except for a false-positive detection at the beginning of the scenario.

6 Experimental Results

This section describes our experimental results, summarising the data first and then discussing the implications of the results.

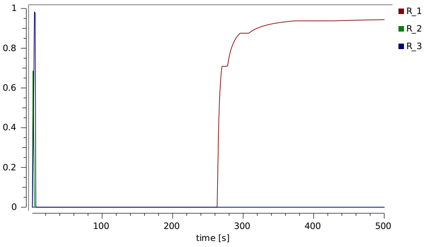

Figure 5: Simulation of fault isolation of fault in valve 1 with mixed non-linear/linear model (T_1 non-linear and both T_2 and T_3 linear). The figure depicts the probability of valves 1, 2 and 3 being faulty.

6.1 Model Comparisons

We have empirically compared the diagnostics performance of several multi-tank models. In our first set of experiments, we ran a simulation over 500 seconds, and induced a fault (valve V_1 at 50%) after 250 s. The model combinations involved a non-linear (NL) model, a model (denoted M) with tank T_1 being linear (and other tanks non-linear), a fully linear model (denoted L), and a Qualitative model (denoted Q).

To compare the relative performance of the models, we compute a measure of diagnostics error (or loss), using the difference between the true fault (which is known for each simulation) and the computed fault. We denote the true fault existing at time t using the pair (ω, t); the computed fault at time t is denoted using the pair $(\hat{\omega}, \hat{t})$. The inference system that we use, LNG [35], computes an uncertainty measure associated with each computed fault, denoted $P(\hat{\omega})$. Hence, we define a measure of diagnostics error over a time window $[0, T]$ using

$$\gamma_1^D = \sum_{t=0}^{T} \sum_{\xi \in \Xi} |P(\hat{\omega_t}) - \omega_t|, \tag{10}$$

where Ξ is the set of failure modes for the model, and ω_t denotes ω at time t.

Our second metric covers the fault latency, i.e., how quickly the model identifies the true fault (ω, t): $\gamma_2 = t - \hat{t}$.

Table 2 summarises our results. The first columns compare the number of parameters for the different models, followed by comparisons of the error (γ_1) and the CPU-time (γ_2). The data show that the error (γ_1) does not grow very much as we increase model size, but it increases as we decrease model fidelity from non-linear through to qualitative models. In contrast, the CPU-time (a) increases as we increase model size, and (b) is proportional to model fidelity, i.e., it decreases as we decrease model fidelity from non-linear through to qualitative models.

In a second set of experiments, we focused on multiple model types for a 3-tank system, with simulations running over 50s, and we induced a fault (valve V_1 at 50%) after 25 s. The model combinations involved a non-linear (NL) model, a model with tank 3 linear (and other tanks non-linear), a model with tanks 2 and 3 linear and tank 1 non-linear, a fully linear model, and a qualitative model. Table 3 summarises our results.

The data show that, as model fidelity decreases, the error γ_1 increases significantly and the inference times γ_2 decrease modestly. If we examine the outputs from AIC_c, we see that the best model is the mixed model (T_3-linear). BIC

Tanks		2	3	4
# Parameters	NL	7	9	11
	M	6	8	10
	L	5	7	9
	Q	2	3	4
γ_1	NL	242	242	242
	M	997	1076	1192
	L	1236	1288	1342
	Q	3859	3994	4261
γ_2	NL	10.59	23.7	39.5
	M	8.52	17.96	34.6
	L	6.11	10.57	32.0
	Q	4.64	7.31	26.4

Table 2: Data for 2-, 3-, and 4-tank models using Non-linear (NL), Mixed (M), Linear (L) and Qualitative (Q) representations

indicates the qualitative model as the best; it is worth noting that BIC typically will choose the simplest model.

	γ_1	γ_2	AIC_c	BIC
Non-Linear	0.97	23.7	29.45	43.7
T_3-linear	3.12	17.96	26.77	42.9
T_2, T_3-linear	21.96	13.21	31.12	39.56
Linear	77.43	10.57	35.76	37.55
Qualitative	304.41	9.74	43.01	29.13

Table 3: Data for 3-tank model, using Non-linear, Mixed, Linear and Qualitative representations, given a fault (valve V_1 at 50%) after 25 s

6.2 Discussion

Our results show that MBD is a complex task with several conflicting factors.

- The diagnosis error γ_1 is inversely proportional to model fidelity, given a fixed diagnosis task.
- The error γ_1 increases with fault cardinality.
- The CPU-time γ_2 increases with model size (i.e., number of tanks).

This article has introduced a framework that can be used to trade off the different factors governing MBD “accuracy". We have shown how one can extend a set of information-theoretic metrics to combine these competing factors in diagnostics model selection. Further work is necessary to identify how best to extend the existing information-theoretic metrics to suit the needs of different diagnostics applications, as it is likely that the “best" model may be domain- and task-specific.

It is important to note that we conducted experiments with un-calibrated models, and we have ignored the cost of calibration in this article. The literature suggests that linear models can be calibrated to achieve good performance, although performance inferior to that of calibrated non-linear models. This class of qualitative models does not possess calibration factors, so calibration will not improve their performance.

7 Conclusions

This article has presented a framework for evaluating the competing properties of models, namely fidelity and computational complexity. We have argued that model performance needs to be evaluated over a range of future observations, and hence we need a framework that considers the *expected performance*. As such, information-theoretic methods are well suited.

We have proposed some information-theoretic metrics for MBD model evaluation, and conducted some preliminary experiments to show how these metrics may be applied. This work thus constitutes *a start* to a full analysis of model performance. Our intention is to initiate a more formal analysis of modeling and model evaluation, since there is no framework in existence for this task. Further, the experiments are only preliminary, and are meant to demonstrate how a framework can be applied to model comparison and evaluation.

Significant work remains to be done, on a range of fronts. In particular, a thorough empirical investigation is needs on diagnostics modeling. Second, the real-world utility of our proposed framework needs to be determined. Third, a theoretical study of the issues of mode-based parameter estimation and its use for MBD is necessary.

References

[1] George EP Box. Statistics and science. *J Am Stat Assoc*, 71:791–799, 1976.

[2] Peter Struss. What’s in SD? Towards a theory of modeling for diagnosis. *Readings in model-based diagnosis*, pages 419–449, 1992.

[3] Peter Struss. Qualitative modeling of physical systems in AI research. In *Artificial Intelligence and Symbolic Mathematical Computing*, pages 20–49. Springer, 1993.

[4] Nuno Belard, Yannick Pencolé, and Michel Combacau. Defining and exploring properties in diagnostic systems. *System*, 1:R2, 2010.

[5] Alexander Feldman, Tolga Kurtoglu, Sriram Narasimhan, Scott Poll, and David Garcia. Empirical evaluation of diagnostic algorithm performance using a generic framework. *International Journal of Prognostics and Health Management*, 1:24, 2010.

[6] Steven D Eppinger, Nitin R Joglekar, Alison Olechowski, and Terence Teo. Improving the systems engineering process with multilevel analysis of interactions. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 28(04):323–337, 2014.

[7] Sanjay S Joshi and Gregory W Neat. Lessons learned from multiple fidelity modeling of ground interferometer testbeds. In *Astronomical Telescopes & Instrumentation*, pages 128–138. International Society for Optics and Photonics, 1998.

[8] Roxanne A Moore, David A Romero, and Christiaan JJ Paredis. A rational design approach to gaussian process modeling for variable fidelity models. In *ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 727–740. American Society of Mechanical Engineers, 2011.

[9] Peter D Hanlon and Peter S Maybeck. Multiple-model adaptive estimation using a residual correlation Kalman filter bank. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(2):393–406, 2000.

[10] Redouane Hallouzi, Michel Verhaegen, Robert Babuška, and Stoyan Kanev. Model weight and state estimation for multiple model systems applied to fault detection and identification. In *IFAC Symposium on System Identification (SYSID), Newcastle, Australia*, 2006.

[11] Amardeep Singh, Afshin Izadian, and Sohel Anwar. Fault diagnosis of Li-Ion batteries using multiple-model adaptive estimation. In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pages 3524–3529. IEEE, 2013.

[12] Amardeep Singh Sidhu, Afshin Izadian, and Sohel Anwar. Nonlinear Model Based Fault Detection of Lithium Ion Battery Using Multiple Model Adaptive Estimation. In *World Congress*, volume 19, pages 8546–8551, 2014.

[13] Aki Vehtari, Janne Ojanen, et al. A survey of bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6:142–228, 2012.

[14] Athanasios C Antoulas, Danny C Sorensen, and Serkan Gugercin. A survey of model reduction methods for large-scale systems. *Contemporary mathematics*, 280:193–220, 2001.

[15] Alexander Feldman, Gregory M Provan, and Arjan JC van Gemund. Computing observation vectors for max-fault min-cardinality diagnoses. In *AAAI*, pages 919–924, 2008.

[16] Amardeep Singh, Afshin Izadian, and Sohel Anwar. Nonlinear model based fault detection of lithium ion battery using multiple model adaptive estimation. In *19th IFAC World Congress, Cape Town, South Africa*, 2014.

[17] Youmin Zhan and Jin Jiang. An interacting multiple-model based fault detection, diagnosis and fault-tolerant control approach. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 4, pages 3593–3598. IEEE, 1999.

[18] Peter Struss and Oskar Dressler. " physical negation" integrating fault models into the general diagnostic engine. In *IJCAI*, volume 89, pages 1318–1323, 1989.

[19] Johan De Kleer, Alan K Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2):197–222, 1992.

[20] Elizabeth H Keating, John Doherty, Jasper A Vrugt, and Qinjun Kang. Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research*, 46(10), 2010.

[21] Saket Pande, Mac McKee, and Luis A Bastidas. Complexity-based robust hydrologic prediction. *Water resources research*, 45(10), 2009.

[22] G Schoups, NC Van de Giesen, and HHG Savenije. Model complexity control for hydrologic prediction. *Water Resources Research*, 44(12), 2008.

[23] S Pande, L Arkesteijn, HHG Savenije, and LA Bastidas. Hydrological model parameter dimensionality is a weak measure of prediction uncertainty. *Natural Hazards and Earth System Sciences Discusions, 11, 2014*, 2014.

[24] Martin Kunz, Roberto Trotta, and David R Parkinson. Measuring the effective complexity of cosmological models. *Physical Review D*, 74(2):023503, 2006.

[25] Gregory M Provan and Jun Wang. Automated benchmark model generators for model-based diagnostic inference. In *IJCAI*, pages 513–518, 2007.

[26] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[27] Jing Du. The “weight" of models and complexity. *Complexity*, 2014.

[28] Jasper A Vrugt and Bruce A Robinson. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and bayesian model averaging. *Water Resources Research*, 43(1), 2007.

[29] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

[30] Hirotugu Akaike. Likelihood of a model and information criteria. *Journal of econometrics*, 16(1):3–14, 1981.

[31] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–466, 1978.

[32] Eric-Jan Wagenmakers. A practical solution to the pervasive problems ofp values. *Psychonomic bulletin & review*, 14(5):779–804, 2007.

[33] Pol D Spanos. *Linearization techniques for non-linear dynamical systems*. PhD thesis, California Institute of Technology, 1977.

[34] Benjamin Kuipers and Karl Åström. The composition and validation of heterogeneous control laws. *Automatica*, 30(2):233–249, 1994.

[35] Alexander Feldman, Helena Vicente de Castro, Arjan van Gemund, and Gregory Provan. Model-based diagnostic decision-support system for satellites. In *Proceedings of the IEEE Aerospace Conference, Big Sky, Montana, USA*, pages 1–14, March 2013.