Summary of: On Validating Propositional Logic System Descriptions for Fault Diagnosis

Alexander Diedrich[1] , **Lukas Moddemann**[1] , **Oliver Niggemann**[1]
[1]Helmut-Schmidt-University
firstname.lastname@hsu.hamburg

Abstract

This is an extended abstract of the manuscript 'On Validating Propositional Logic System Descriptions for Fault Diagnosis' (Diedrich, Moddemann, and Niggemann, 2026) that was published in the journal Engineering Applications of Artificial Intelligence in January, 2026.

1 Extended Abstract

Model-based fault diagnosis is the task of finding root causes of anomalous system behaviour in technical systems. The root causes of faults are computed by comparing actual observations of a system against a model that is known a-priori. In the article we analyse the models used for model-based fault diagnosis, usually called system descriptions. Since the early days of the research field (De Kleer and Williams, 1987; Reiter, 1987) system descriptions have always been an integral part of model-based fault diagnosis and capture the system's behaviour, as well as the connections between components. However, all diagnosis algorithms were developed with the premise of having a correct model. But what if this is not so?

Our main difference to previous research is that we do not assume the existence of a single propositional logic system description SD (the model). Instead, we assume a propositional logic ground truth model $\mathcal{M}$ that may or may not be known, and a set of system descriptions approximating this ground-truth, such that $SD_i \approx \mathcal{M}$ with $i \in \mathbb{N}$. This is different from previous research, which assumes $SD = \mathcal{M}$, i.e. there is only one ground truth in form of the system description SD. The main reason why SD may deviate from $\mathcal{M}$ is that recent approaches are able to create system descriptions from observational data and other, non-design time external knowledge (see system identification (Ljung, 1998)). This introduces a novel challenge in that usually systems can only be observed at distinct locations and thus are only partially observable (i.e any model created from this data will be limited by the information obtainable from this limited observability).

Therefore, in cases where system descriptions are approximated, we need some way to determine whether some model is suitable for diagnosis. I.e. that the model expresses all the information we need to diagnose most of the components.

In fault diagnosis the term *suitable* in regard to models boils down to two properties: i) how many single-faults can be diagnosed, i.e. how many faults involving only the defect of exactly one component could theoretically be distinguished given the observations, and ii) how well the model describes a real system. While the analysis of diagnosability has seen a lot of work in the past, especially for discrete event systems (Travé-Massuyes, Escobet, and Olive, 2006), determining how well a model fits to the real system was seldomly investigated. The main reason for this is the assumption that a system model is complete (Christopher and Grastien, 2024), i.e. that the model exactly describes the system up to some error margin ($SD = \mathcal{M}$). In this article we deviate from this view. Instead, we assume that a model is only an approximation of the underlying system and thus may be incomplete in many dimensions ($SD \approx \mathcal{M}$). We therefore do not analyse the *system's diagnosability*, but the *model's diagnosability*. To do this, we provide novel methods to measure how well a given model describes the underlying system. Figure 1 describes our setup.

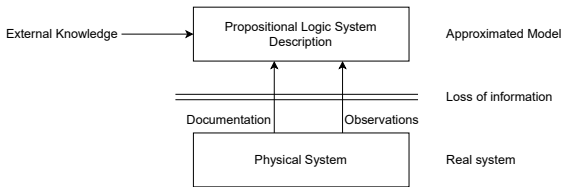

Figure 1: System descriptions may not completely model the underlying system, which requires our analysis of the system description's diagnosability and expressiveness. The model approximation may be learned from data, i.e. from observations, documentation, or through external facts, such as expert knowledge.

Algorithmically, we attempt to solve two problems in this article. For the analysis of a system description's diagnosability we solve a decision problem, as we have to decide for each possible set of observations, whether a component is diagnosable as a single-fault. Since we assume that the system description is (at least slightly) different from the real system and cannot make any assumptions about which observations will realistically occur, we must exponentially iterate through each set of observations. Therefore, we require a novel algorithm that takes a system description as its input and computes all possible single-faults from the complete

set of observations. The second problem we solve is the comparison of at least two system descriptions along a number of dimensions. We therefore develop a novel method to solve the graph alignment and graph matching problems for propositional logic system descriptions for the specified dimensions.

We present two main contributions:

1. A new algorithm CHECKDIAG to validate the diagnosability of approximated system descriptions.
2. The novel algorithms CHECKEXPRESS and CALCOSIM to compare known system descriptions, for example, ones created from expert knowledge, to approximated system descriptions.

Within the algorithms we compute the novel metrics in depicted in Table 1 which provide quantitative information about the quality of approximated system descriptions.

Factor	Description
q_{comps}	Quotient of the number of components in two system descriptions. $q_{\text{comps}} = \frac{\lvert SD_1 \rvert}{\lvert SD_2 \rvert}$.
q_{cosim}	Quotient of co-similarity measuring structural overlap between two system descriptions. It captures how closely the generated model matches the reference model with respect to shared components.
q_F	Quotient of diagnosable versus non-diagnosable faults. It evaluates diagnostic quality by relating uniquely diagnosable or isolable faults to non-diagnosable ones. $q_F = \frac{\#Isol.}{\#Non-isol.}$.
q_{obs}	Quotient of observable variables. It captures the relationship between the number of observations produced by a model and those available in the reference system. $q_{\text{obs}} = \frac{\lvert OBS_1 \rvert}{\lvert OBS_2 \rvert}$.

Table 1: Validation factors used for comparing automatically generated system descriptions with reference and expert models.

Overall, three usages of our algorithms arise: i) Some single system description exists and we use CHECKDIAG to determine its diagnosability. ii) Several alternative system descriptions exist (generated through system identification algorithms) where we want to determine the most suitable one. Then we can compare the system descriptions among each other using CHECKEXPRESS. iii) We have one or more approximated system descriptions, and a known ground truth, and want to compare which system description is closest to the ground truth. In this case we also use CHECKEXPRESS.

We evaluate our approach on one real system (data from the international Space Station's COLUMBUS module), the well-known ISCAS-85 Benchmark, a simulation of the Tennessee Eastman Process, the ADAPT Benchmark of spacecraft power distribution, and two sets of process industry benchmarks. Overall this shows the broad applicability of our method from real-valued systems over discrete systems to Boolean circuits. We demonstrate that well-designed systems such as Boolean circuits, the Tennessee Eastman Process, or the COLUMBUS module have a low number of non-diagnosable components. While systems created to test algorithms handling partial observability, such as the two benchmarks of process industry use-cases, indicate a higher number of non-diagnosable components.

To the best of our knowledge this article is the first attempt to present a comprehensive method to validate propositional logic system descriptions in the context of model-based fault diagnosis. Meaning in particular for abductive models (Peischl, Pill, and Wotawa, 2016), which reason from non-normal observations back towards possible causes. Of course, this line of research is closely associated with the disciplines of knowledge-base verification (Preece and Shinghal, 1994) and the works on diagnosability of discrete-event systems (Bittner et al., 2022), albeit the area of application is somewhat different. The benefits of our approach are that it is straightforward to apply to any propositional logic system description used for abductive fault diagnosis (Diedrich et al., 2025). Thus, our contribution can hopefully serve as a basis to validate system descriptions generated by any process that performs system identification and approximates propositional logic expressions. Further, it can serve as a validation mechanism for experts who have created system descriptions manually.

References

Bittner, B.; Bozzano, M.; Cimatti, A.; Gario, M.; Tonetta, S.; and Vozarova, V. 2022. Diagnosability of fair transition systems. *Artificial Intelligence* 309:103725.

Christopher, C. J., and Grastien, A. 2024. Critical observations in model-based diagnosis. *Artificial Intelligence* 104116.

De Kleer, J., and Williams, B. C. 1987. Diagnosing multiple faults. *Artificial intelligence* 32(1):97–130.

Diedrich, A.; Krysander, M.; Heesch, R.; and Niggemann, O. 2025. Diagnosis driven anomaly detection for cyber-physical systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Diedrich, A.; Moddemann, L.; and Niggemann, O. 2026. On validating propositional logic system descriptions for fault diagnosis. *Engineering Applications of Artificial Intelligence* 165:113379.

Ljung, L. 1998. System identification. In *Signal analysis and prediction*. Springer. 163–173.

Peischl, B.; Pill, I.; and Wotawa, F. 2016. Abductive diagnosis based on modelica models. In *27th International Workshop on Principles of Diagnosis (DX)*.

Preece, A. D., and Shinghal, R. 1994. Foundation and application of knowledge base verification. *International journal of intelligent Systems* 9(8):683–701.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial intelligence* 32(1):57–95.

Travé-Massuyes, L.; Escobet, T.; and Olive, X. 2006. Diagnosability analysis based on component-supported analytical redundancy relations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 36(6):1146–1160.