Diagnosing Systems through Approximated Information

Alexander Diedrich[1], Oliver Niggemann[2]

[1] *Fraunhofer IOSB-INA, Fraunhofer Center for Machine Learning, 32657 Lemgo*
alexander.diedrich@iosb-ina.fraunhofer.de

[2] *Helmut-Schmidt-University, 22043 Hamburg*
oliver.niggemann@hsu-hh.de

ABSTRACT

This article presents a novel approach to diagnose faults in production machinery. A novel data-driven approach is presented to learn an approximation of dependencies between variables using Spearman correlation. It is further shown, how the approximation of the dependencies are used to create propositional logic rules for fault diagnosis. The article presents two novel algorithms: 1) to estimate dependencies from process data and 2) to create propositional logic diagnosis rules from those connections and perform consistency-based fault diagnosis. The presented approach was validated using three experiments. The first two show that the presented approach works well for injection molding machines and a simulation of a four-tank system. The limits of the presented method are shown with the third experiment containing sets of highly correlated signals.

1. INTRODUCTION

Diagnosing faults in physical systems such as production systems is an increasingly important task. In the past, it was common for operators of production machinery to employ people to operate, maintain, and repair their machines. The specific knowledge would be inherent in the minds of the humans employed within the company. Current trends in automation, however, lead to fewer workers on the factory floor, an increase in automation, and a resulting increasing autonomy of each production system.

One major element of autonomous production systems is self-diagnosis in the presence of faults. A fault is defined as some unwanted deviation from normal operating behaviour. Once a fault occurs, production is usually disrupted, which can lead to the destruction of components, loss in revenue, and even harm humans. In the literature, many logic-based approaches to diagnose physical systems have been presented . But only very few approaches are actually usable outside of limited use-cases (Feldman, Provan, & van Gemund, 2009; Feldman, Provan, & Van Gemund, 2010; Stern, Kalech, & Elimelech, 2014; Khorasgani & Biswas, 2017). However, other domains seem to have tackled the problem (Sampath, Sengupta, Lafortune, Sinnamohideen, & Teneketzis, 1995; Leitão, Rosso, Leal, & Zoitl, 2020; M. J. Daigle et al., 2010). Major challenges in real-world systems exist in dealing with changing system parameters, small batch or lot sizes, insufficient instrumentation, and complexity associated with creating manual models of faulty behaviour (Diedrich, Balzereit, & Niggemann, 2019).

Those manual models are often not available for many industrial use-cases, often because of monetary reasons or because process experts were not convinced to give up their knowledge (Bobrow & Whalen, 2002). It is therefore necessary to develop a method for fault diagnosis that works with approximated diagnosis rules.

In this paper we present a data-driven methodology (based on statistics i.e correlation) inspired by the qualitative physics approaches of De Kleer (De Kleer & Brown, 1984) and Forbus (Forbus, 1984a). These are motivated by the analogy to reason about quantitative processes (such as the influence of temperature on pressure within a tank) in a similar way to human thought processes. Humans, instead of calculating the exact amount pressure increase given an increase in temperature, look for some simplification. These are often of a qualitative nature. Qualitative physics captures the qualitative nature of physical processes by describing a process through discrete values such as low ($\perp$), high ($\top$), normal (0). Diedrich et al. (Diedrich, Maier, & Niggemann, 2019) showed how well-known and proven diagnosis algorithms are used in conjunction with such discrete values to perform diagnosis using a logic-based formalism. Here we use the same methodology, but base it on qualitative physics rather than

calculating residual values from quantitative (i.e. control theoretic or machine learning) methods.

To diagnose physical systems three requirements must be met:

1. It must be known which components are used in the system and how those components behave. For example, that a pump moves a liquid with a certain velocity.

2. It must be known how those components are connected with each other and to the outside world

3. The dependency between components must be known.

Instead of attempting to solve and to prove the self-diagnosis problem for all physical systems, we lay our focus on injection molding machines and similar processes. Injection molding machines are common in many industrial areas and are usually used for high-throughput processes. Thus, if a fault occurs, it is of utmost importance to identify its root-cause and find a suitable solution.

Within this article we make the following contributions:

- We adapt a quantitative physics approach to learn weak fault models (only models of normal behaviour) of industrial injection molding machines and similar processes (a tank system and a compounding process).

- We show how propositional logic rules can be approximated from process data alone

- We show how well-known diagnosis algorithms are used to correctly diagnose injection molding machines, given approximate logic rules

The article is structured as follows: The next section will analyze prior art and identify relevant research gaps. Section 3 will show how a typical industrial process is modelled using qualitative physics. Section 4 will use the created model to introduce a novel diagnosis methodology. The following section will present results using real data from an injection molding machine as well as provide a theoretical evaluation of the approach. Section 7 summarizes the findings.

2. STATE OF THE ART

The idea of physical causality was introduced by the works of Forbus (Forbus, 1984a) and De Kleer (De Kleer & Brown, 1984; De Kleer, 1984). With the idea of *envisioning* De Kleer (de Kleer & Brown, 1982) introduced a powerful methodology for diagnosis use cases. In this context time intervals are understood in the way introduced by Allen (Allen, 1983, 1984), which cuts continuous time into very small chunks where a system behaves static.

Williams (Williams, 1984) and Raiman (Raiman, 1990) extended Qualitative Process Theory with possible discretization steps in the time domain. Common misconceptions about qualitative physics were answered by Williams and De Kleer in 1991 (Williams & de Kleer, 1991)

Within this article we deal with causal physical systems, meaning systems in which the outputs of one part directly influence the inputs of other system parts (Pearl & Dechter, 2013). Perumalla et al. (Perumalla et al., 2019) have studied how sensor placement and causality in cyber-physical systems can be inferred with data-driven approaches. Guo et al. (Li et al., 2008) have shown how causality graphs can help to understand system behaviour and can be used for diagnosis. Faghraoui et al. (Faghraoui et al., 2014) introduced an entropy-based method for building causality graphs for diagnosis. Kiaei and Lotifard (Kiaei & Lotfifard, 2019) have modelled causality through Petri networks to diagnose power grids.

Many works (P. J. Mosterman & Biswas, 1998; P. Mosterman & Biswas, 1999; Khorasgani & Biswas, 2017; Narasimhan & Biswas, 2007; Roychoudhury, Biswas, & Koutsoukos, 2006; Roychoudhury, Daigle, Biswas, & Koutsoukos, 2011; M. J. Daigle et al., 2010) carry on the idea of flow-like and effort-like propagation through the introduction and qualification of hybrid bond graphs. Matei et al. (Matei et al., 2019) have recently attempted to bridge the gap between qualitative and quantitative approaches by combining parameter tracking with analytical redundancy relations.

Struss (Struss, 1997) published a paper on the fundamentals of model-based diagnosis of dynamic systems. He proposed to capture the temporal and dynamic behaviour of a hybrid system in a set of modes which model the system. He demonstrates his approach on a car's anti-braking system. Struss was also one of the first to describe the introduction of strong-fault models into GDE (Struss & Dressler, 1989). Daigle et al. (M. J. Daigle et al., 2010; M. Daigle et al., 2007) have adapted a discrete event approach to diagnose continuous systems. Grastien et al. and others (Grastien, Haslum, Thiébaux, et al., 2012; Meskin, Khorasani, & Rabbath, 2010) have developed an approach to extend Reiter's diagnosis algorithms which was described for binary circuits to include DES (Discrete Event Systems) and hybrid systems. Their approach is similar to Daigle et al., Struss, and Provan in so far as they transform the continuous parts of a model into qualitative states.

Narasimhan and Biswas (Narasimhan & Biswas, 2007) have proposed an FDI (Fault Detection and Isolation) system for diagnosing the fuel-transfer system for fighter aircraft. In their approach they model the fuel-transfer system with hybrid bond graphs. The model consists of an extended Kalman filter and a state-space representation. For fault identification they compute the Taylor series expansion as the continuous residual signal transient. These residual values are compared to a fault signature generated from the hybrid bond graph. From this they create hypotheses which are used for fault diagnosis.

In another work Khorasgani and Biswas (Khorasgani &

Biswas, 2017) describe a hybrid system model through hybrid minimal structurally overdetermined sets (HMSOs). These are sets of differential equations and (in-) equations which model the behaviour of a hybrid system.

Many diagnosis approaches have been tried out in industrial use cases. For example in rotor systems (Babu Rao & Mallikarjuna Reddy, 2021; Leitão et al., 2020), smart-grids (Jiang, Zhang, Gao, & Wu, 2014), wind turbines (Svärd & Nyberg, 2011), automotive (Struss, 2002; Stein, 2009), power electronics (Poon et al., 2017), embedded systems (Zoeteweij, Pietersma, Abreu, Feldman, & Van Gemund, 2008), process industry (Kallesoe, Cocquempot, & Izadi-Zamanabadi, 2006; Struss & Ertl, 2009), and spacecraft (Bajwa, Sweet, & Korsmeyer, 2003; Balaban, Narasimhan, Cannon, & Brownston, 2007; Biswas et al., 2016b, 2016a).

This work is most similar to works from De Kleer (De Kleer & Williams, 1987) and Pearl (Pearl, 2009). It is certainly less rigorous than Pearls approach or Granger's (Bressler & Seth, 2011), but it facilitates model creation for diagnosis in a statistical, data-driven manner. Therefore we use the traditional approach to diagnosis presented by De Kleer and others (De Kleer & Williams, 1987; Reiter, 1987; Grastien, 2013), but augment the model building with our algorithm to find dependencies between signals. It is therefore a more flexible approach than the above-mentioned expert-knowledge driven methods by Narasimhan (Narasimhan & Brownston, 2007), Khorasgani (Khorasgani & Biswas, 2017), or Roychodhury (Roychoudhury et al., 2006).

3. QUALITATIVE MODELLING OF INDUSTRIAL SYSTEMS

Often injection molding machines operate batch-wise. They produce a product for several hours, while the quality is sampled once during the production time. If the quality is insufficient, either because of faults or because of unforeseen changes in the production environment the production is stopped and the faulty products are destroyed (i.e. shredded) and their material mostly reinserted into production. This can lead to significant losses for companies.

This section shows how a data-driven method can be used to remedy some of the drawbacks of finding faults in batch-wise production with physical systems. In the past, qualitative physics (Forbus, 1984b) and consistency-based diagnosis algorithms (De Kleer & Williams, 1987; de Kleer & Brown, n.d.) were used. However, the amount of expert knowledge and time required is prohibitive for many companies. Especially for the small and medium sized enterprises (SMEs), which often use smaller production systems such as injection molding machines. Therefore, we propose our data-driven method. By relying only on the available process data we expect our method to perform worse than many traditional methods. However, it may provide a first approach to make
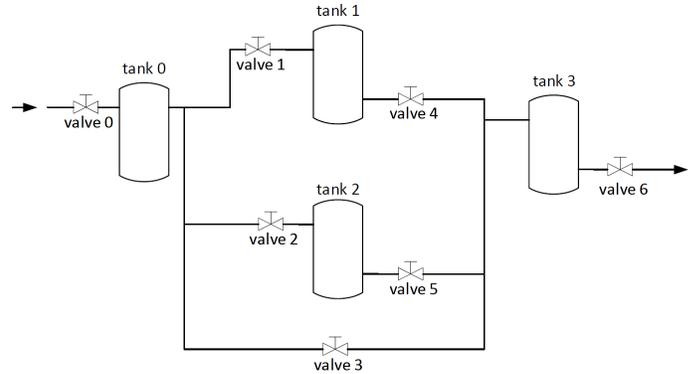
Figure 1. A four tank system system without pumps

the benefits of fault diagnosis available to smaller companies with no resources for modelling.

3.1. Overview

To be helpful for SMEs an automated method is needed which can be deployed on SME production floors, but requires little to no attention from experts. It needs to learn a diagnosis model on its own and inform experts only when faults occur. Therefore, our method is based on automatically calculating the Spearman correlation between signals in historical process data. We define process data as two time series S and M, where S contains data such as setpoints, machine parameters, process parameters etc. and is of the form $((t_0, x_0), ..., (t_n, x_n))$, with t being a timestamp and some value $x \in \mathcal{R}$. M contains measured sensor data and quality data and is of the form $((t_0, x_0), ..., (t_n, x_n))$.

We assume that for physical systems a sufficient correlation between two signals implies a certain amount of temporal dependency. What follows can to some degree be likened to Pearl's causal graphs (though with the limitation that we assume a somewhat accurate causation through high correlation). Pearl has shown that for physical systems one can draw a causal graph $G = (V, E)$ (Pearl, 1995) with nodes V and edges E, where each node describes a signal and an edge denotes causation. Taking this approach we change the definite causation to mean the existence of a correlation above a threshold τ.

3.2. Running Example

We will use the four-tank system (Diedrich & Niggemann, 2018) depicted in figure 1 as a running example. The figure shows four tanks connected though pipes and valves. The input to each tank is protected by a valve, as well as the output of the rightmost tank. We chose a tank system as these are common in the consistency-based diagnosis field and can provide some practical insights into the methods described below. It must be noted that the example has been created

using OpenModelica 1.13 on Windows in 64-bit. The dataset of the simulation contains 1939 variables of which only a few have been selected for this running example by filtering the signal names to only those which were associated with temperature, flow, and water level

3.3. Processing correlation

We use Spearman correlation to describe the nonlinear and distribution independent relationships between variables. Spearman correlation is defined as

$$\rho = \frac{cov(rX, rY))}{\sigma(rX) \cdot \sigma(rY)} \tag{1}$$

, where rX and rY are the ranks of two signals, σ is the standard deviation, and cov is the covariance. Hallin and Peri (Hallin & Puri, 1991) have shown that using rank-based methods work well for time series data.

Calculating ρ consisting of the signals $S \cup M$ tells us how far different signals are correlated. We only take the lower triangular matrix from ρ to remove redundant information.

We aim to reformulate the matrix ρ such that one signal $m \in M$ is associated to signals $s_i \in S$. The reformulation from matrix to tuple is done by selecting one signal $m_j \in M$ and then selecting $\forall s_i \in S$, with $|s_i| > \tau$ for all signals $m \in M$. This ensures that only those signals are represented within the tuple whose Spearman correlation is greater than threshold τ and thus have a meaningful contribution to the sensor value m_j (i.e. the assumed causation). The result is a set of tuples with variable length of the form $(s_0, s_1, ...s_i, m_j)$. The set of all tuples is denoted as T.

The set consists of tuples of the form

$$
\begin{aligned}
&(s_0, s_1, ...s_x, m_0) \\
&(s_0, s_1, ...s_y, m_1) \\
&(s_0, s_1, ...s_z, m_2) \\
&\qquad\qquad \vdots
\end{aligned}
\tag{2}
$$

, for example. I.e. a subset of sensor values s is associated with one measurement variable $m \in M$. For the running example this translates to the tuples

$$
\begin{aligned}
&(t1level, t1medT, valve4flow) \\
&(t1level, t1medT, t2level, t2medT, valve5flow) \\
&(t1medT, valve6flow) \\
&(t1level, t1medT, t2t4flow) \\
&(t1level, t1medT, t2level, t2medT, t3_t4flow) \\
&(t1medT, t4sinkflow)
\end{aligned}
\tag{3}
$$

The tuples have been identified on a threshold $\tau = 0.3$. The rightmost signals are the measurement variables $m \in M$. The

rules establish a clear relationship between sensors at the system input to sensors at the system output. The prevalence of the medium temperature *tXmedT* is due to the simulation being executed in a very stable process with low variance.

4. DIAGNOSING QUALITATIVE MODELS

Within this section we show how to reformulate the set of tuples T into propositional logic rules, which can be used for consistency-based diagnosis and thus leads to the identification of faults in physical systems. We focus on the formulation of weak fault models (Stern et al., 2014). Such models only describe the system's normal behaviour.

In the research field of consistency-based diagnosis (De Kleer & Williams, 1987; Reiter, 1987; De Kleer & Kurien, 2003) faults are located by attempting to satisfy the set of propositional logic expressions

$$SD \cap \alpha \nvdash \bot \tag{4}$$

, where SD is the system description and contains proposition or predicate logic formulae and α are observations.

The general idea in traditional consistency-based diagnosis, which was defined on binary circuits is as follows: Analysing binary circuits SD contains formulae describing the behaviour of logic gates such as AND, OR etc. And α are actual values from the system inputs, outputs, and intermediate values.

In this article we adapt this approach to the field of production machinery. To adapt the formalisation from traditional consistency-based diagnosis to physical systems we redefine the system model SD and the observations α. Therefore, SD is defined as

Definition 1 (System Description - SD). *Given a set of variables $COMPS$ and a set of measurements M, SD is a conjunction $\phi = \bigwedge_{c \in COMPS} ok(s_i) \to ok(m_i)$, with $m_i \in M$ that relates the variables in $COMPS$ to one measurement each. The ok-literals are interpreted in McCarthy's way (McCarthy, 1989).*

Within the running example of the four-tank system the definition of SD translates to

$$
\begin{aligned}
&t1level \wedge t1medT \to valve4flow \\
&t1level \wedge t1medT \wedge t2level \wedge t2medT \to valve5flow \\
&t1medT \to valve6flow \\
&t1level \wedge t1medT \to t2t4flow \\
&t1level \wedge t1medT \wedge t2level \wedge t2medT \to t3t4flow \\
&t1medT \to t4sinkflow
\end{aligned}
$$

$$\tag{5}$$

in propositional logic.

Consequently observations are defined as a grounding of the

measurement variables by assigning a truth value to them. Formally, an observation is defined as

Definition 2 (Observation - α). *Given a sensor value $\tilde{\alpha}$ and a discretization function $d : \mathcal{R} \rightarrow \{0,1\}$ the observation $\alpha = d(\tilde{\alpha}, \tau)$ is a binary discretization of the sensor value, with parameters τ.*

We limit the projection of the discretization function $d(\cdot)$ to binary values. Alternative formulations include ternary logic ($\{-1,0,1\}$), and the calculation of residual values (Khorasgani & Biswas, 2017). By interpreting the number 0 as false ($\bot$) and 1 as true ($\top$), we can establish a direct relationship between sensor values and logic statements.

Taking the set of tuples T introduced in section 3 we can use a bijective projection to assign the sensor values s_i to component symbols in $COMPS$. It must be noted that in traditional consistency-based diagnosis the symbols in $COMPS$ represent discrete components within a system (for example pumps, transformers, pipes, resisters etc.). Since we are using a purely data-driven approach, we must relax this constraint in that the symbols in $COMPS$ stand for sensor names. It is the task of the process expert to find the suitable components in case of faults.

Beside assigning the sensors of each tuple to the $COMPS$ in each rule, we can also assign the measurement from the tuple to the m_i within the rule. Consequently, we create the rules set $\Phi = \{\phi_0, \phi_1, ...\}$. This rule set is the basis to use traditional diagnosis algorithms such as GDE (De Kleer & Williams, 1987), Reiter's diagnosis lattice (Reiter, 1987), or SAFARI (Feldman et al., 2010).

The assignment of the discretized sensor values to the variables in the running example is done through adding the propositional logic constraints

$$
\begin{aligned}
valve4flow &\Leftrightarrow \top \\
valve5flow &\Leftrightarrow \top \\
valve6flow &\Leftrightarrow \top \\
t2t4flow &\Leftrightarrow \top \\
t3t4flow &\Leftrightarrow \top \\
t4sinkflow &\Leftrightarrow \top
\end{aligned}
\tag{6}
$$

, if no fault exists within the system. $\top$ denotes a true value and is interpreted such that no fault exists. According to McCarthy's AB-literals (McCarthy, 1989) this would, for example, be written as $ok(valve4flow) = \top$.

The goal of diagnosis algorithms is to calculate a set of diagnoses

Definition 3 (Diagnosis). *Given a fault-augmented model SD with fault variables $COMPS$ and an observation α, a diagnosis ω is defined as an assignment to all fault variables in $COMPS$ such that $\omega \models SD \wedge \alpha$.*

, where the assignment to fault variables is done by

Definition 4 (Health Assignment). *A health assignment is a binary assignment to all elements in $COMPS$, such that $SD \cap \alpha \nvdash \bot$*

A diagnosis algorithm outputs a diagnosis for each possible assignment to $COMPS$ given the observations α. In practice, when diagnosing large systems with hundreds of sensor values, the amount of possible diagnoses can become quite large. Therefore, it is common to introduce the assumption that always the least amount of components may fail. So when a fault occurs it is sensible to first look at those diagnoses containing the least amount of possible faulty components. This set is defined as the minimal cardinality diagnosis

Definition 5 (Minimal-Cardinality Diagnosis). *A diagnosis δ_{min} is minimal, if no diagnosis $\tilde{\delta}_{min} \subseteq \delta_{min}$ exists that is also a diagnosis.*

To show how to perform diagnosis on the running example we assume Φ to be grounded with a random health assignment

$$
\begin{aligned}
t1level \wedge t1medT &\rightarrow valve4flow \\
t1level \wedge t1medT \wedge t2level \wedge t2medT &\rightarrow valve5flow \\
t1medT &\rightarrow valve6flow \\
t1level \wedge t1medT &\rightarrow t2t4flow \\
t1level \wedge t1medT \wedge t2level \wedge t2medT &\rightarrow t3_t4flow \\
t1medT &\rightarrow t4sinkflow \\
valve4flow &\Leftrightarrow \top \\
valve5flow &\Leftrightarrow \top \\
valve6flow &\Leftrightarrow \bot \\
t2t4flow &\Leftrightarrow \bot \\
t3t4flow &\Leftrightarrow \top \\
t4sinkflow &\Leftrightarrow \top
\end{aligned}
\tag{7}
$$

in this case the diagnosis would compute the set of diagnoses $\{\{t1level, t1medT\}, \{t1medT\}\}$ through a hitting set calculation. Identifying the minimum cardinality diagnosis would mean to take the smallest set, thus returning $\{t1medT\}$ as the correct diagnosis.

4.1. An algorithm for data-driven diagnosis of production systems

Given the above definitions and the set of tuples T defined in section 3, it is possible to state the algorithm in listing 1.

The algorithm DDRC takes a time series as its input, where signals are divided into two groups. The first group S are signals that influence the production process and the second group M are signals from quality control measurements. The Spearman correlation coefficient is calculated and converted into a lower triangular matrix. The matrix is then traversed and evaluated according to threshold τ. The result from this evaluation is a set of edges, describing which signals are highly correlated (above the threshold). The set of edges is

Algorithm 1: DDRC: data-driven diagnosis rule creation

Data: $X = ((t_0, x_0), ..., (t_n, x_n)), \tau$
Result: T
1   $\rho \leftarrow Spearman(X);$ `// Eq.1`
2   $\rho_l \leftarrow toLowerTriangular(\rho);$
3   $edges \leftarrow \emptyset;$
4   **foreach** $row, column \in \rho_l$ **do**
5   **if** $\rho(row, column) > \tau$ **then**
6       $edges \leftarrow edges \cup (row, column);$
7   **else if** $\rho(row, column) < -\tau$ **then**
8       $edges \leftarrow edges \cup (row, column);$
9   $T \leftarrow \emptyset;$
10   **foreach** $m \in M$ **do**
11     $S_m \leftarrow select(edges, m);$
12     $t \leftarrow (S_m, m);$
13     $T \leftarrow T \cup t;$

---

then converted into the set of tuples T, where one measurement m is related to a subset of sensors. The algorithm returns the set T.

The diagnosis algorithm DDA-IM uses a set of tuples T and actual observations α to compute minimum cardinality diagnoses δ_{min}. As described above, the set T is translated into propositional logic and used for diagnosis.

Algorithm 2: DDA-IM: data-driven diagnosis algorithm for physical systems

---

Data: $T, \tilde{\alpha}$
Result: δ_{min}
1   $\Phi \leftarrow toSymbolic(T);$ `// Def.1`
2   $\alpha \leftarrow d(\tilde{\alpha});$ `// Def.2`
3   $\omega \leftarrow diagnose(\Phi, \alpha);$ `// Def.3`
4   $\delta_{min} \leftarrow min(\omega);$ `// Def.5`

The algorithm DDA-IM (Listing 2) takes the output from algorithm DDRC and creates symbolic rules for diagnosis out of the set of tuples T. The function $toSymbolic(T)$ inserts a conjunction ($\wedge$) between the sensors of set S and creates an implication to each variable m. Therefore, the result is of the form: $\phi_i : \bigwedge_{s_i} s \rightarrow m$. All symbolic rules comprise the set Φ. Function $d()$ discretises the sensor measurements $\tilde{\alpha}$ into a binary representation. Function $diagnose()$ assigns the actual values α to variables $m \in \Phi$ and then runs the GDE diagnosis algorithm. The result is a set of diagnosis candidates ω. DDA-IM returns only the smallest diagnosis candidate δ_{min}, according to the minimal cardinality assumption.

5. EMPIRICAL EVALUATION

We have evaluated our approach on several systems. System 1 is an injection molding machine. System 2 is a simulation of a four-tank system, which has extensively been used in consistency-based diagnosis research (Diedrich & Niggemann, 2018; Diedrich, Maier, & Niggemann, 2019). System

Table 1. The results for System 1

Threshold: τ	0.1	0.3	0.5	0.7	0.8	0.9
Sensible rules achieved	x	x	x	x	x	x
Diagnosis correct	x	x	x	x	x	x

3 is a compounding process for rubber pre-products. Using three different processes we can show that our approach is able to generalise to other kinds of systems.

We executed all experiments with Python 3.7 on a 64-bit Windows computer with 16GB of RAM and Intel i7-9750H processor. To set up the diagnosis rules Φ we used only data from normal operating conditions without any anomalies to create the weak fault models. Faults were simulated by assigning a random value ($\{\top, \bot\}$) to all variables $m \in M$. This results in a varying amount of faulty components for each experiment run. Thus, calculating and tracking symptoms of the systems was omitted.

The goal of the evaluation is to prove the following:

1. The diagnosis rules Φ shall be approximate correct to perform diagnosis. I.e. they should differ among each other and make sense to an expert.

2. The diagnosis algorithm shall be able to identify the injected fault(s)

Our experiments show that a completely uninformed data-driven approach is sufficient to perform some rudimentary diagnosis. This provides a baseline to augment the algorithms DDRC and DDA-IM with some expert knowledge to improve diagnosis in real industrial use-cases.

5.1. System 1: Injection Molding Machine

The injection molding machine produces plastic casings for Raspberry PI systems. For our method we split the time series data from the machine into the subsets S and M. In set S we captured signals such as temperatures, cycle times, pressure values, and speed of various components. In set M we represented measurements from an optical quality control system. The size of the datasets was 11 signals for S and 10 signals for M. Faults were injected through randomly setting some signals in set M to false, which corresponds to wrong size measurements of the produced parts in the real world.

Table 1 contains the results for system 1 over different thresholds. The first row shows, whether the generated rules were suitable for diagnosis. The second row shows whether the diagnosis algorithm was able to determine the injected faults. This shows that our methods works well for the intended use-case of injection molding machines. Only few signals are highly correlated such that causation can be assumed. The correct diagnosis was found in each experiment and the generated diagnosis rules are approximately correct, given the coarse data-driven approach.

Table 2. The results for System 2

Threshold: τ	0.1	0.3	0.5	0.7	0.8	0.9
Sensible rules achieved	-	x	x	x	x	x
Diagnosis correct	x	x	x	x	x	x

Table 3. The results for System 3

Threshold: τ	0.1	0.3	0.5	0.7	0.8	0.9
Sensible rules achieved	x	x	-	-	-	-
Diagnosis correct	x	x	-	-	-	-

5.2. System 2: Four-tank system simulation

The simulation of the four-tank system created with OpenModelica 1.13 contained 1939 signals including all Modelica-internal variables. Of these, 18 were identified as measurement variables, by filtering the signal names to only those which were associated with temperature, flow, and water level. Given the size of the dataset we argue that our approach shows good results as can be validated on the running example and by choosing a reasonable value for τ. Table 2 contains the results for system 2 over different thresholds. The first row shows, whether the generated rules were suitable for diagnosis. The second row shows whether the diagnosis algorithm was able to determine the injected faults. Again, faults were injected through randomly setting some signals in set M to false, which simulates wrongly closed valves or a leaky tank. This evaluation shows that apart from injection molding machines, which perform very well for this approach, the standard four-tank systems established in the literature perform quite well given this uninformed approach.

5.3. System 3: Compounding process

The production of rubber products requires a batch-wise compounding process in which different ingredients are mixed according to a defined recipe. For this experiment we used time series of single batches encompassing 6506 rows of 87 signals. Of these 35 were identified to be quality control signals from laboratory data, thus forming set M and the rest being contained in set S. Table 2 contains the results for system 2 over different thresholds. The first row shows, whether the generated rules were suitable for diagnosis. The second row shows whether the diagnosis algorithm was able to determine the injected faults. In contrast to systems 1 and 2 signals within the compounding process are not highly correlated. This is partly due to some signals being aggregated over time. For example the laboratory data is only available once a batch is finished and has been stored for some time. As a result only few distinct rules exist such that Φ is comparatively small (about 8 rules with $\tau = 0.3$. Above a threshold $\tau = 0.5$ only one rule exists, which is insufficient for diagnosis. The injected faults were simulated through randomly setting some signals in set M to false. This simulates wrong measurements of ingredients, viscosity, or other material properties.

6. DISCUSSION

The running example (section 3) and the evaluation section 5 show that creating diagnosis rules (i.e weak fault models) in a completely statistical, data-driven manner for injection molding machines is possible, but with some drawbacks. Injection molding machines produce time series with only few highly correlated signals. These highly correlated signals lead to rule sets Φ that facilitate the usage of hitting set algorithms such as GDE (De Kleer & Williams, 1987). Some rules may be surprising (such as the reliance on the medium temperature in the four-tank system), but are due to artifacts within the simulation. However, for both systems diagnosable rules could be obtained.

The experiment with system 3 has shown that our proposed method breaks down for processes with many highly correlated signals. In this case, the resulting rules contain mostly the same symbols and thus do not lead to helpful diagnoses.

The diagnosis rules and the resulting diagnosis sets help process experts to quickly locate faults within injection molding machines. The diagnosis output is a set of signals causing the faulty behaviour. The process expert can use this information and act accordingly.

Future work should extend this method with more expert knowledge to create better diagnosis rules. This could be through integrating the results of existing Failure Modes Effects Analyses (FMEA) or other methods to capture expert knowledge. Further, to remedy the difficulties encountered with system 3, we plan to use more elaborate methods such as Granger Causality and Transfer Entropy (Bressler & Seth, 2011).

7. CONCLUSION

Within this article we have presented a novel method for the data-driven creation of diagnosis rules in propositional logic. To obtain these rules we presented an algorithm to estimate a qualitative physics model using Spearman correlation. We have also introduced an algorithm which uses the qualitative physics model to create propositional logic rules, merge them with disretized observations, and compute a diagnosis.

The evaluation has shown that the estimated rules work well for injection molding machines and the four-tank system, but break down for systems with highly correlated signals.

REFERENCES

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843.

Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, *23*(2), 123–154.

Babu Rao, K., & Mallikarjuna Reddy, D. (2021). Fault detection in rotor system by discrete wavelet neural network algorithm. *Journal of Vibration and Control*, 10775463211030754.

Bajwa, A., Sweet, A., & Korsmeyer, D. (2003). The livingstone model of a main propulsion system. *RIACS Technical Report*.

Balaban, E., Narasimhan, S., Cannon, H. N., & Brownston, L. S. (2007). Model-based fault detection and diagnosis system for nasa mars subsurface drill prototype. In *2007 ieee aerospace conference* (pp. 1–13).

Biswas, G., Khorasgani, H., Stanje, G., Dubey, A., Deb, S., & Ghoshal, S. (2016a). An application of data driven anomaly identification to spacecraft telemetry data. In *Prognostics and health management conference.*

Biswas, G., Khorasgani, H., Stanje, G., Dubey, A., Deb, S., & Ghoshal, S. (2016b). An approach to mode and anomaly detection with spacecraft telemetry data. *Int. J. Prognostics Health Manage*, *7*, 1–18.

Bobrow, D. G., & Whalen, J. (2002). Community knowledge sharing in practice: the eureka story. *Reflections: The SoL Journal*, *4*(2), 47–59.

Bressler, S. L., & Seth, A. K. (2011). Wiener–granger causality: a well established methodology. *Neuroimage*, *58*(2), 323–329.

Daigle, M., Koutsoukos, X., & Biswas, G. (2007). A discrete event approach to diagnosis of continuous systems. In *Proceedings of the 18th international workshop on principles of diagnosis* (pp. 259–266).

Daigle, M. J., Roychoudhury, I., Biswas, G., Koutsoukos, X. D., Patterson-Hine, A., & Poll, S. (2010). A comprehensive diagnosis methodology for complex hybrid systems: A case study on spacecraft power distribution systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(5), 917–931.

De Kleer, J. (1984). How circuits work. *Artificial intelligence*, *24*(1-3), 205–280.

de Kleer, J., & Brown, J. S. (n.d.). Theories of Causal Ordering. , 29.

de Kleer, J., & Brown, J. S. (1982). Foundations of envisioning. In *Aaai* (pp. 434–437).

De Kleer, J., & Brown, J. S. (1984). A qualitative physics based on confluences. *Artificial intelligence*, *24*(1-3), 7–83.

De Kleer, J., & Kurien, J. (2003). Fundamentals of model-based diagnosis. *IFAC Proceedings Volumes*, *36*(5), 25–36.

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

Diedrich, A., Balzereit, K., & Niggemann, O. (2019). First approaches to automatically diagnose and reconfigure hybrid cyber-physical systems. *ML4CPS 2020 – Machine Learning for Cyber Physical Systems Conference.*

Diedrich, A., Maier, A., & Niggemann, O. (2019). Model-based diagnosis of hybrid systems using satisfiability modulo theory. In *Proceedings of the thirty-third aaai conference on artificial intelligence (aaai-19).* Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19).

Diedrich, A., & Niggemann, O. (2018). Diagnosing hybrid cyber-physical systems using state-space models and satisfiability modulo theory. *29th International Workshop on the Principles of Diagnosis.*

Faghraoui, A., Kabadi, M. G., Sauter, D., Boukhobza, T., & Aubrun, C. (2014, October). Data-driven causality digraph modeling of large-scale complex system based on transfer entropy. In *2014 IEEE Conference on Control Applications (CCA)* (pp. 705–710). (ISSN: 1085-1992) doi: 10.1109/CCA.2014.6981423

Feldman, A., Provan, G., & van Gemund, A. (2009). The lydia approach to combinational model-based diagnosis. *Proc. DX*, *9*, 403–408.

Feldman, A., Provan, G., & Van Gemund, A. (2010). Approximate model-based diagnosis using greedy stochastic search. *Journal of Artificial Intelligence Research*, *38*, 371–413.

Forbus, K. D. (1984a). Qualitative process theory. *Artificial intelligence*, *24*(1-3), 85–168.

Forbus, K. D. (1984b). Qualitative Process Theory. *Artificial Intelligence*, 187.

Grastien, A. (2013). Diagnosis of hybrid systems by consistency testing. In *24th international workshop on principles of diagnosis (dx-13)* (pp. 9–14).

Grastien, A., Haslum, P., Thiébaux, S., et al. (2012). Conflict-based diagnosis of discrete event systems: Theory and practice. In *Kr.*

Hallin, M., & Puri, M. L. (1991). Rank tests for time series analysis a survey.

Jiang, H., Zhang, J. J., Gao, W., & Wu, Z. (2014). Fault detection, identification, and location in smart grid based on data-driven computational methods. *IEEE Transactions on Smart Grid*, *5*(6), 2947–2956.

Kallesoe, C. S., Cocquempot, V., & Izadi-Zamanabadi, R. (2006). Model based fault detection in a centrifugal pump application. *IEEE transactions on control systems technology*, *14*(2), 204–215.

Khorasgani, H., & Biswas, G. (2017). Structural fault detection and isolation in hybrid systems. *IEEE Transactions on Automation Science and Engineering.*

Kiaei, I., & Lotfifard, S. (2019, July). A Two-Stage Fault Location Identification Method in Multiarea Power Grids Using Heterogeneous Types of Data. *IEEE Transactions on Industrial Informatics*, *15*(7), 4010–4020. doi: 10.1109/TII.2018.2885320

Leitão, H. A., Rosso, R. S., Leal, A. B., & Zoitl, A. (2020). Fault handling in discrete event systems applied to iec 61499. In *2020 25th ieee international conference on emerging technologies and factory automation (etfa)* (Vol. 1, pp. 1039–1042).

Li, G., Gao, J., & Chen, F. (2008, January). Construction of causality diagram model for diagnostics. In *2008 Annual Reliability and Maintainability Symposium* (pp. 97–102). (ISSN: 0149-144X) doi: 10.1109/RAMS.2008.4925777

Matei, I., Zhenirovskyy, M., de Kleer, J., & Feldman, A. (2019). A hybrid qualitative and quantitative diagnosis approach. In *Annual conference of the phm society* (Vol. 11).

McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In *Philosophical logic and artificial intelligence* (pp. 161–190). Springer.

Meskin, N., Khorasani, K., & Rabbath, C. A. (2010, November). A Hybrid Fault Detection and Isolation Strategy for a Network of Unmanned Vehicles in Presence of Large Environmental Disturbances. *IEEE Transactions on Control Systems Technology*, *18*(6), 1422–1429. (Conference Name: IEEE Transactions on Control Systems Technology) doi: 10.1109/TCST.2009.2038066

Mosterman, P., & Biswas, G. (1999, November). Diagnosis of continuous valued systems in transient operating regions. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *29*(6), 554–565. doi: 10.1109/3468.798059

Mosterman, P. J., & Biswas, G. (1998). A theory of discontinuities in physical system models. *Journal of the Franklin Institute*, *335*(3), 401–439.

Narasimhan, S., & Biswas, G. (2007). Model-based diagnosis of hybrid systems. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, *37*(3), 348–361.

Narasimhan, S., & Brownston, L. (2007). Hyde-a general framework for stochastic and hybrid modelbased diagnosis. *Proc. DX*, *7*, 162–169.

Pearl, J. (1995). Causal Diagrams for Empirical Research. *California Digital Library*, 36.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J., & Dechter, R. (2013). Identifying independencies in causal graphs with feedback. *arXiv preprint arXiv:1302.3595*.

Perumalla, K., Yoginath, S., & Lopez, J. (2019, November). Detecting Sensors and Inferring their Relations at Level-0 in Industrial Cyber-Physical Systems. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1–5). doi: 10.1109/HST47167.2019.9032891

Poon, J., Jain, P., Konstantakopoulos, I. C., Spanos, C., Panda, S. K., & Sanders, S. R. (2017). Model-based fault detection and identification for switching power converters. *IEEE Transactions on Power Electronics*, *32*(2), 1419–1430.

Raiman, O. (1990). Order of magnitude reasoning. In *Readings in qualitative reasoning about physical systems* (pp. 318–322). Elsevier.

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

Roychoudhury, I., Biswas, G., & Koutsoukos, X. (2006). A bayesian approach to efficient diagnosis of incipient faults. In *Proceedings of the 17th international workshop on principles of diagnosis (dx-06)* (pp. 243–250).

Roychoudhury, I., Daigle, M. J., Biswas, G., & Koutsoukos, X. (2011). Efficient simulation of hybrid systems: A hybrid bond graph approach. *Simulation*, *87*(6), 467–498.

Sampath, M., Sengupta, R., Lafortune, S., Sinnamohideen, K., & Teneketzis, D. (1995). Diagnosability of discrete-event systems. *IEEE Transactions on automatic control*, *40*(9), 1555–1575.

Stein, B. (2009). Using models for dynamic system diagnosis a case study in automotive engineering. In *Dagstuhl-workshop mbees: Modellbasierte entwicklung eingebetteter systeme* (p. 46).

Stern, R., Kalech, M., & Elimelech, O. (2014). Hierarchical diagnosis in strong fault models. In *Twenty fifth international workshop on principles of diagnosis.*

Struss, P. (1997). Fundamentals of model-based diagnosis of dynamic systems. In *Ijcai* (pp. 480–485).

Struss, P. (2002). Model-based on-board diagnosis and tools for the developer of on-board systems-vmbd and idd: Two projects of the european car industries. *MONET Newsletter Issue 2 - Automotive Supplement*.

Struss, P., & Dressler, O. (1989). " physical negation" integrating fault models into the general diagnostic engine. In *Ijcai* (Vol. 89, pp. 1318–1323).

Struss, P., & Ertl, B. (2009). Diagnosis of bottling plants–first success and challenges. In *20th international workshop on principles of diagnosis* (pp. 83–90).

Svärd, C., & Nyberg, M. (2011). Automated design of an fdi-system for the wind turbine benchmark. *IFAC Proceedings Volumes*, *44*(1), 8307–8315.

Williams, B. C. (1984). 1984 - The Use of Continuity in a Qualitative Physics. *Association for the Advancement of Artificial Intelligence*, 5.

Williams, B. C., & de Kleer, J. (1991). Qualitative reasoning about physical systems: a return to roots. *Artificial Intelligence*, *5*(1), 1–9.

Zoeteweij, P., Pietersma, J., Abreu, R., Feldman, A., &

Van Gemund, A. J. (2008). Automated fault diagnosis in embedded systems. In *2008 second international conference on secure system integration and reliability improvement* (pp. 103–110).