ROOT CAUSE ANALYSIS USING ANOMALY DETECTION AND TEMPORAL INFORMED CAUSAL GRAPHS

Josephine Rehak[1], **Shahenda Youssef**[1], **and Jürgen Beyerer**[1,2]
[1]Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, Karlsruhe
[2]Fraunhofer IOSB, Fraunhoferstraße 1, Karlsruhe

ABSTRACT

In industrial processes, anomalies in the production equipment may lead to expensive failures. To avoid and avert such failures, the identification of the right root cause is crucial. Ideally, the search for a root cause is backed by causal information such as causal graphs. We have extended a framework that fuses causal graphs with anomaly detection to infer likely root causes. In this work, we add the use of temporal information to draw temporal valid conclusions about the potential propagation of anomalous information in causal graphs. The use of the framework is demonstrated on a robotic gripping process.

Keywords causal graphs · anomaly detection · multivariate timeseries · root cause analysis

1 Introduction

The investigation of root causes has bugged humans since the beginning of engineering [9, 12]. Since then, a variety of approaches have been published to infer root causes [7]. Only a minority of works are based on causal research. Consider that the spread of anomalous information throughout any system or process has to adhere to the laws of causality. Causal research inspects in-depth the interplay between the variables in given data and allows the identification of causes and effects [13]. Thus, it may support the identification of a fault's root cause and the factors contributing to it [8, 4].

We show in this publication how to draw conclusions about the root cause of anomalies and failures by considering their fundamental temporal laws. Thereby, we add to the causal reasoning framework introduced by [14]. The fundamental setup of the framework is to first detect the anomalies, to determine the features that made a data entry anomalous, to retrace their potential causes in the corresponding provided causal graph, and finally to evaluate the match between the potential anomalies of a root cause according to the graph and the actual detected anomalies. In this publication, we make use of coarse temporal information in the form of process steps to allow temporal sound deductions about the potential root causes in the graph. It is important to note that the anomaly

detection step is not part of this contribution. In this matter, we refer to existing work as summarized by [6].

We have structured our work in the following manner. First in Section 2, we briefly introduce the required fundamentals to then show in Section 3 the related work in the causal research domain. Section 4 is dedicated to the adaptation of the root cause algorithm to integrate coarse temporal information in its deductions. Section 5 shows in an example robotic gripping scenario how the deductions of the algorithms improve. Finally, Section 6 concludes the work.

2 Fundamentals

Causal Graphs
A common way to represent causal relations is in the form of graphs $G = (\mathcal{V}, \mathcal{E})$. These graphs commonly consist of a set of nodes representing variables $\mathcal{V}$ and a set of edges $\mathcal{E}$. The latter represents direct causal relations between variables [13]. For timeseries data, the use of summary graphs is common [3]. In them, each variable is associated with a timeseries. An edge $a \to b$ exists between the variables $a \in \mathcal{V}$ and $b \in \mathcal{V}$ if there exists any sample $a_i \in a$ causing an effect in a sample of $b_j \in b$ if $a \neq b$, such that a_i precedes or is equitemporal $i \leq j$ to b_j in time or if $a = b$ that a_i precedes b_j in time $i < j$. We call it a causal path $a \to^* c$ from a to c if a sequence of directed edges from a_i to c_j is present [2].

Temporal Law of Causation
It is known that the effect of an event may never precede its cause [5]. Under the assumption that no other influence is present and if it is known that two variables are causal related $a \to b$, then b can never occur before a. Likewise, if $a \to b \to c$ is given, c may only occur after a and b.

Jaccard Similarity
The Jaccard similarity [10] is a similarity measure often used for recommender systems. It divides the intersection of two sets $\mathcal{A} \cap \mathcal{B}$ by their union $\mathcal{A} \cup \mathcal{B}$.

$$\text{Jaccard} = \frac{\mathcal{A} \cap \mathcal{B}}{\mathcal{A} \cup \mathcal{B}} \tag{1}$$

A result of one indicates a perfect match, while a result of zero indicates two different sets.

3 Related Work

3.1 Causal Research for Root Cause Analysis

Causal Methods and Anomaly detection methods have been combined before in the literature for various application areas.

Yang et al. [15] published a general framework to detect anomalies and to identify root causes in multivariate time series. They learned causal structure from data, identified modules in the investigated system, and performed anomaly detection on each separate module. The location of the anomalies is supposed to represent the root causes.

Koutroulis et al. [11] combined causal graphs and anomaly detection methods to robustly detect anomalous cyber security attacks in the control system of a water treatment testbed. They trained structural causal models (SCMs) offline with attack free data. In online scenarios, the SCMs allowed the computing of prediction errors to identify critical

dependencies in the data. Each error sequence was assessed with the average treatment effect and hypothesis tests. Root cause candidates are indicated by providing the set of sensors associated with the attacks.

Agrawal et al. [1] used one Bayesian Network (BN) for each type of root cause to diagnose root causes in a coal-fired power plant. Each was provided by experts in advance. Based on fixed thresholds, the variables were detected to be either normal or anomalous. The anomalies were propagated in the BN and the magnitude and the type of root cause were indicated.

The framework RootCLAM by Han et al. [8] learns a Variational Causal Graph Autoencoder from normal data to then identify for anomalous data the root cause features where in comparison to normal data, the presence of a foreign influence is evident.

None of the shown approaches provided the possibility to argue over potential root causes. Also, none of them assessed the relation between the root cause and the sensor measurements themselves.

3.2 Temporal Information for Root Cause Analysis

The procedure EasyRCA [4] uses a given acyclic summary causal graph (ASCL) to identify root causes in multivariate timeseries data. According to their work, root causes may be variables with anomalies, whose first appearing anomaly might not have been propagated by any other causing variable under the assumption the anomaly may not precede its cause in time in the ASCL. Additionally, root cause variables are identified by comparing the adapted total effect estimations on direct relations of a normal and an anomalous regime to detect variables affected by foreign influences.

Other than in our work, the summary graph is required to be acyclic and root causes are not known beforehand.

3.3 Preceding Work

In [14], we published the first algorithm to infer the probability of a root cause to be the cause for observed anomalies. It makes use of the Jaccard Similarity Root Cause Score (JRCS) to evaluate the fit between the set of potentially affected measured variables $\mathcal{V}_{G,r}$ for a root cause r and the set of discovered anomalous variables $\mathcal{V}_A$.

$$\text{JRCS}(r) = \frac{\mathcal{V}_A \cap \mathcal{V}_{G,r}}{\mathcal{V}_A \cup \mathcal{V}_{G,r}} \tag{2}$$

This version of the algorithm only uses variables for inference on the root cause without considering temporal aspects.

4 Temporal Informed Algorithm for Root Cause Analysis

4.1 Prerequisites

To perform the updated RCA procedure, we split the process into process steps $p \in \mathcal{P}$, given $\mathcal{P} = \{1, 2, ..., |\mathcal{P}|\}$. Also, we require the order of assignment of process steps p to correspond to the temporal and causal order.

We will work with an adaptation of the summary graphs $G = (\mathcal{V}, \mathcal{E})$ introduced before. In them, we may not access the timeseries of a subset of variables $\mathcal{V}_L \subset \mathcal{V}$, as they are latent or unmeasured. Integrating these latent variables is optional. They support human

understanding, but may be omitted by introducing direct causal relations between its causing and caused non-latent variables.

The adapted summary graphs need to contain a set of unary variables $\mathcal{V}_R \subset \mathcal{V}$ as root cause variables annotated by an expert. Also, we require a set of measured variables $\mathcal{V}_M = \mathcal{V}/(\mathcal{V}_R \cup \mathcal{V}_L)$.

Additionally, it needs annotations of process steps $\mathcal{P}_{a\to b} = \{p \in \mathbb{N}\}$ to be provided for all $a, b \in \mathcal{V}$ if $a \neq b$ and $a \to b$ holds. To create such annotations as $\mathcal{P}_{a\to b}$ manually, one should list all process steps in which an anomaly in a would show effects in b.

As the anomaly detection should be performed in advance and is not part of this contribution, we require it to provide a tuple set $\mathcal{T}_A = \{(m, p), ...\}$ given $m \in \mathcal{V}_M$, $p \in \mathbb{N}$, which contains pairs of the discovered anomalies and the process steps they occurred in.

4.2 Graph Preprocessing

In the preprocessing step, we can calculate $\mathcal{P}_{r\to^* m}$ for each path $r \to^* m$. It represents the maximum possible set of process steps in which $r \in \mathcal{V}_R$ could cause anomalies in $m \in \mathcal{V}_M$ for the specific path. The algorithm to calculate a unique set of $\mathcal{P}_{\{r\to^* m\}}$ is shown in Algorithm 1. Since multiple paths $r \to^* m$ may be possible, we express with $\mathcal{P}_{\{r\to^* m\}}$ the maximum possible set of process steps for a set of paths. To exclude double entries, we use in the following the $unique$ function.

Algorithm 1: Compute valid process steps for path sets

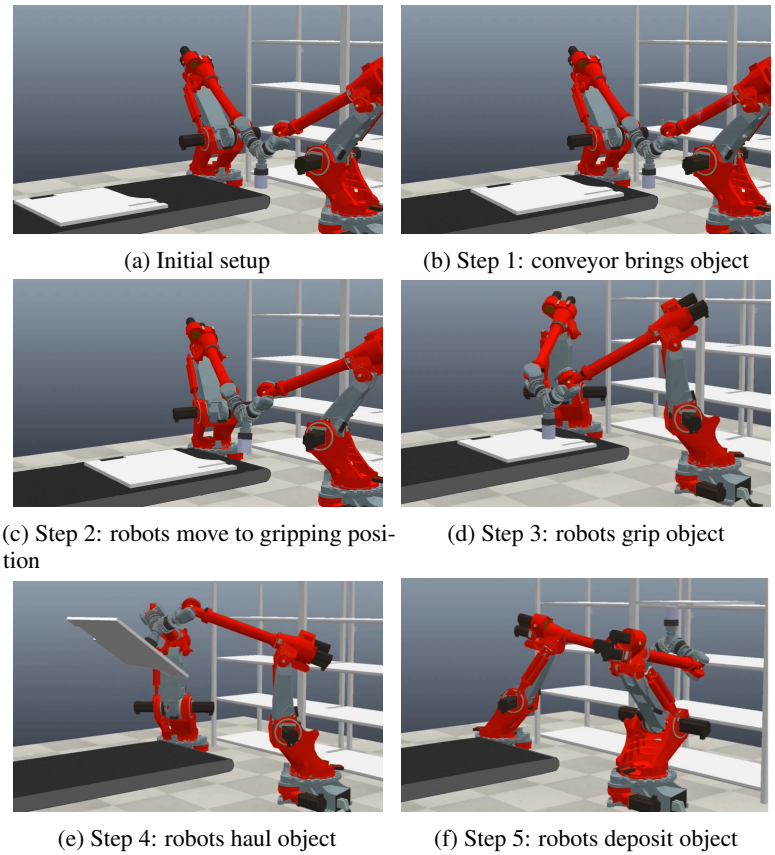
Input : Set of edges $\{r \to^* m\}_G$ from r to m in graph G,
set of assigned process steps $\mathcal{P}_{a\to b}$ for any edge $a \to b$
Output : Set of process steps $\mathcal{P}_{\{r\to^* m\}}$ for each path $r \to^* m$

$\mathcal{P}_{\{r\to^* m\}} \leftarrow \emptyset$
for $r \to^* m \in \{r \to^* m\}_G$ **do**
  $i \leftarrow 0$
  $\mathcal{P}_{\text{temp}} \leftarrow \emptyset$
  for $a \to b \in r \to^* m$ **do**
    $\mathcal{P}_{\text{temp}} \leftarrow \{\mathcal{P}_{a\to b} \geq i\}$
    $i \leftarrow \min(\mathcal{P}_{\text{temp}})$
  $\mathcal{P}_{\{r\to^* m\}} \leftarrow \mathcal{P}_{\text{temp}} \cap \mathcal{P}_{\{r\to^* m\}}$
return unique$(\mathcal{P}_{\{r\to^* m\}})$

As each edge of a path can occur in each process step of its corresponding set and thus allows a variety of process step sequences, we need to ensure the valid temporal sequence of edges for each $r \to^* m$ as is implied by the temporal law. We consider only possible process step sequences which allow the sequential and equitemporal spread of anomalous causal information. The latter is due to the coarse-grained temporal categorization of edges into process steps, which does not allow detailed deductions about the temporal sequence.

Given $\mathcal{P}_{r\to^* m}$, the tuple set $\mathcal{T}_G(r)$ is generated for each r by calculating tuples of measurement variables and their potentially affected process steps. For that, we make use of $\{r \to^* m\}_G$ denoting the set of all possible paths from r to m in G.

$$\mathcal{T}_G(r) = \{(m, p) | \forall\, r \to^* m \in \{r \to^* m\}_G, p \in \mathcal{P}_{\{r\to^* m\}}\} \tag{3}$$



(a) Initial setup

(b) Step 1: conveyor brings object

(c) Step 2: robots move to gripping position

(d) Step 3: robots grip object

(e) Step 4: robots haul object

(f) Step 5: robots deposit object

Figure 1: Depicted is each process step of the robotic gripping process

4.3 Scoring of Potential Root Causes

In this step, the JRCS scores are calculated for each root cause by the use of an updated formula. This time, its resulting score rates the fit between two tuple sets on a scale from zero to one. We assume this similarity rating to reflect the likelihood of a root cause to have caused the anomalies.

$$\text{JRCS}(r) = \frac{|\mathcal{T}_A \cap \mathcal{T}_G(r)|}{|\mathcal{T}_A \cup \mathcal{T}_G(r)|} \tag{4}$$

The higher the JRCS result, the better the inspected root cause $r \in \mathcal{V}_R$ fits the anomalous observation $\mathcal{T}_A$.

5 Application

5.1 Application Environment

We apply the extended algorithm to a simple robotic gripping process. In this application, we have identified five process steps. Each step is depicted in Figure 1 and is described in the following.

In the first step, a conveyor moves the prepared object into the gripping position. A proximity sensor is located at the gripping position and measures the distance from the sensor to the gripping position using a laser. The distance decreases significantly if the object is present. In the second step, the coupled robots will move from their default position to the gripping position at the conveyor in unison. Then in the third process step, the magnetic gripper attached to the coupling piece of the robots is powered to attract and grip the workpiece. In step four, the robots move the gripped object from the gripping position to the deposit position. Due to the weight of the gripped object, the torque measured by the robots significantly increases during movement. The sensor back at the gripping position does not indicate the presence of the object anymore. Instead, a similar proximity sensor at the deposit position starts indicating its presence upon arrival. In the fifth and final process step, the robots disengage the magnetic gripper to place the object in the final deposit position.

We simulated a fault-free scenario and also two additional faulty scenarios. One root cause is a shifted object on the conveyor belt. The other root cause is an object in the gripping position, where the magnetic force of the gripper is not sufficient to grab it. For each, we generated sensor measurements corresponding to $\mathcal{V}_M$.

5.2 Adapted Summary Graph

We manually constructed an adapted summary graph for the described scenario and depicted it in Figure 2. The corresponding variable sets $\mathcal{V}_M$ and $\mathcal{V}_R$ are indicated by the variables' horizontal placement. The required set of process steps of a causal relation is annotated close to its edge. If edges are bidirected, the annotation for edges is the closest one along the edge. We simplified the causal graph by summarizing the robots' axis angle and angle speed measurements as one variable each. Commonly, each of the seven robot axes has individual measurements. To allow the easy depiction of the graph, some variables with identical names occur several times.

5.3 Application of the RCA Algorithm

We performed the algorithm described in Section 4 on the two anomalous datasets described in Section 5.1. The graph in Figure 2 was preprocessed as described, and the results are depicted in Table 1.

For the anomaly detection, we used a supervised version of the Kth Nearest Neighbor algorithm. Within each process step, it calculated the Euclidean distance for each inspected time step to the features of its closest normal time step. A sum was formed over all such distances of a process step and divided by the number of the process step's entries. A fixed threshold decided if the inspected process step was detected as normal or anomalous. The set $\mathcal{T}_A$ was formed by retrieving from each process step the measurements, which contributed to the Euclidean distance. Table 2 shows the anomalous process steps and the corresponding measurement variables that we discovered as anomalous.

Given the detected anomalies ($\mathcal{T}_A$) shown in Table 2 and the potential anomalous measurements per root cause ($\mathcal{T}_G(r)$) in Table 1, we calculated the JRCS for each potential root cause for both inspected faulty simulation scenarios. In Table 3, we show the JRCS result for our new temporal extension and the original algorithm. It becomes visible that without considering the temporal information, the actual root causes cannot be distinguished. This is the case as in both simulated scenarios, the same variables are affected.

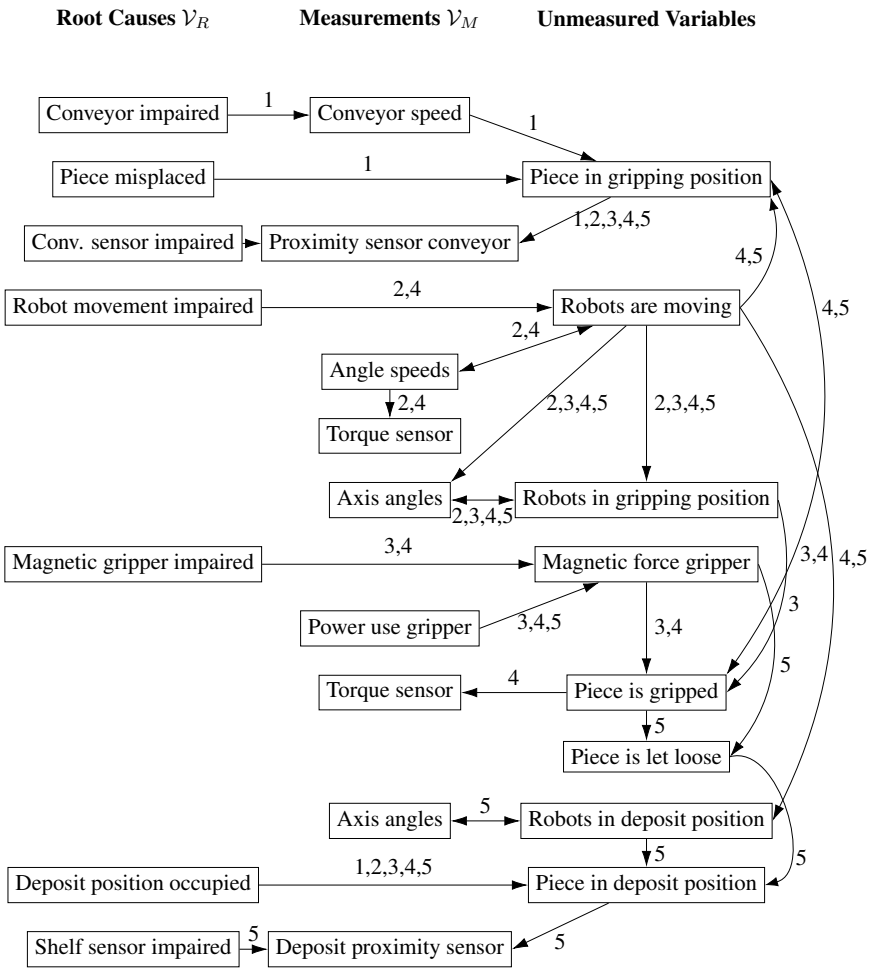


Figure 2: Shown is the adapted summary graph of the robotic setup. The horizontal orientation of the variables indicates their affiliation with the variable sets shown above. The numbers along the edges indicate in which process step an anomaly in the causing variable would be observed in the caused variable. Variables appearing several times are identical.

Table 1: Overview of the potential anomalous measurements affected by the root causes in the shown process steps. Each row corresponds to $\mathcal{T}_G(r)$ for its respective root cause.

		Potential Anomalous Measurements $\mathcal{V}_M$						
		Conveyor Speed	Prox. Sensor Conveyor	Angle Speeds	Torque Sensor	Axis Angles	Power Use gripper	Deposit Prox. Sensor
Pot. root causes $\mathcal{V}_R$	Conveyor impaired	1	1,2,3,4,5	-	4	-	-	5
	Piece misplaced	-	1,2,3,4,5	-	4	-	-	5
	Conv. Prox. Sensor impaired	-	1,2,3,4,5	-	-	-	-	-
	Robot mov. impaired	-	4,5	2,4	2,4	2,3,4,5	-	5
	Magnetic gripper impaired	-	4,5	-	4	-	-	5
	Shelf Prox. Sensor impaired	-	-	-	-	-	-	1,2,3,4,5
	Deposit position occupied	-	-	-	-	-	-	1,2,3,4,5

Table 2: Overview of the discovered anomalous measurements and their corresponding process steps. Each row corresponds to $\mathcal{T}_A$ for its respective scenario.

	Detected Anomalous Measurements $\mathcal{V}_M$						
	Conveyor Speed	Prox. Sensor Conveyor	Angle Speeds	Torque Sensor	Axis Angles	Power Use gripper	Deposit Prox. Sensor
Faulty Scenario I	-	1,2,3,4,5	-	4	-	-	5
Faulty Scenario II	-	4,5	-	4	-	-	5

Table 3: The JRCS for both simulated anomalous scenarios and each root cause if the temporal law is considered or not.

Potential Root Cause	**With Temporal Law**		**Without Temporal Law**	
	Scenario I	Scenario II	Scenario I	Scenario II
Conveyor impaired	0.875	0.500	0.750	0.750
Piece misplaced	**1**	0.571	**1**	**1**
Conv. Prox. Sensor impaired	0.625	0.286	0.333	0.333
Robot mov. impaired	0.286	0.364	0.600	0.600
Magnetic gripper impaired	0.400	**1**	**1**	**1**
Shelf Prox. Sensor impaired	0.111	0.125	0.333	0.333
Deposit position occupied	0.111	0.125	0.333	0.333

We can observe that the original algorithm, which considers only the variables to compute the JRCS, cannot distinguish between two root causes because they affect the same set of variables. In comparison, the extended algorithm rates the actual root causes as the likeliest. The temporal algorithm scores other root causes, which affect similar measurement variables and process steps as depicted in Table 1, as likelier. Root causes with identical $\mathcal{T}_G(r)$, like the last two root cause entries in the table, are rated as equally likely. This shows that the algorithm does not guarantee a clear distinction between root causes. Instead, it must result from the placement of sensors in the process.

6 Summary

In this work, we adapted a causal root cause analysis algorithm to consider temporal process step information. This allows us to draw temporal valid deductions about potential root causes for prior discovered anomalies.

For this purpose, the algorithm uses an adapted summary graph provided by an expert which includes annotations for root cause variables, measurable variables and latent variables. Additionally, the edge annotations contain knowledge about the temporal spread of anomalous information. This causes the algorithm to rely heavily on provided information by an expert but may contribute to preserving expert knowledge.

In our application on a robotic gripper setup, we demonstrate how the discovery of the actual root causes improves by considering coarse process step information. Future work may tackle the issue of adding detailed anomaly types and patterns to the matching process.

Acknowledgements

This work greatly benefited from the Dagstuhl seminar 24031 "Fusing Causality, Reasoning and Learning for Fault Management and Diagnosis".

References

[1] V. Agrawal, B. K. Panigrahi, and P. Subbarao. Intelligent decision support system for detection and root cause analysis of faults in coal mills. *IEEE Transactions on Fuzzy Systems*, 25(4):934–944, 2016.

[2] C. K. Assaad, E. Devijver, and E. Gaussier. Entropy-based discovery of summary causal graphs in time series. *Entropy*, 24(8):1156, 2022.

[3] C. K. Assaad, E. Devijver, and E. Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73:767–819, 2022.

[4] C. K. Assaad, I. Ez-Zejjari, and L. Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In *International Conference on Artificial Intelligence and Statistics*, pages 8395–8404. PMLR, 2023.

[5] H. Beebee. *Hume and the Problem of Causation*. Oxford Handbooks, 1981.

[6] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[7] Z. Gao, C. Cecati, and S. X. Ding. A survey of fault diagnosis and fault-tolerant techniques—part i: Fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3757–3767, 2015.

[8] X. Han, L. Zhang, Y. Wu, and S. Yuan. On root cause localization and anomaly mitigation through causal inference. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 699–708, 2023.

[9] K. Ishikawa and K. Ishikawa. *Guide to quality control*, volume 2. Asian Productivity Organization Tokyo, 1982.

[10] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.

[11] G. Koutroulis, B. Mutlu, and R. Kern. A causality-inspired approach for anomaly detection in a water treatment testbed. *Sensors*, 23(1):257, 2022.

[12] T. Ohno. *Toyota production system: beyond large-scale production*. Productivity press, 2019.

[13] J. Pearl. Causal inference. In I. Guyon, D. Janzing, and B. Schölkopf, editors, *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, volume 6 of *Proceedings of Machine Learning Research*, pages 39–58, Whistler, Canada, 12 Dec 2010. PMLR.

[14] J. Rehak, A. Sommer, M. Becker, J. Pfrommer, and J. Beyerer. Counterfactual root cause analysis via anomaly detection and causal graphs. In *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*, pages 1–7. IEEE, 2023.

[15] W. Yang, K. Zhang, and S. C. Hoi. Causality-based multivariate time series anomaly detection. *arXiv preprint arXiv:2206.15033*, 2022.