Data-Driven Augmentation of Expert Causal Knowledge in Cyber-Physical Systems

Katrin Ehrenmüller[1,*], Lucas Kook[2], Fajar J. Ekaputra[1] and Marta Sabou[1]

[1]Institute for Data, Process and Knowledge Management, Vienna University of Economics and Business (WU Vienna)
[2]Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU Vienna)

Abstract

As industrial systems (such as smart grids) grow in size and complexity, transparency and explainability become increasingly important for supporting user understanding and trust. Causal models are key components to translate system behaviour into human reasoning. While causal graphs can be created from domain-expert knowledge about system dynamics, this method carries the risk of specifying only partially complete graphs. Yet, attempting to learn the entire causal graph from observational data is also known to be challenging and error-prone. In this paper, we propose a procedure for the data-driven augmentation of existing causal graphs defined by domain-experts. Specifically, we test Granger non-causalities implied by the existing graph on sensor measurement data. Under appropriate statistical and causal assumptions, the test results can then indicate missing edges in the existing graph. We evaluated our approach in a real-life smart charging garage scenario, testing the expert-defined causal graphs with real-world sensor data. The results show inconsistencies between the causal graph defined by experts and the observed data, demonstrating the capability of our approach to reveal missing causal relations. These findings highlight the potential of our approach in combining expert knowledge and data-driven analysis for validating and augmenting causal representations in complex systems, such as smart grids, contributing to the broader topic of transparent and interpretable industrial systems.

Keywords

Causality, Knowledge Completion, Cyber-Physical System, Falsification, Granger Causality

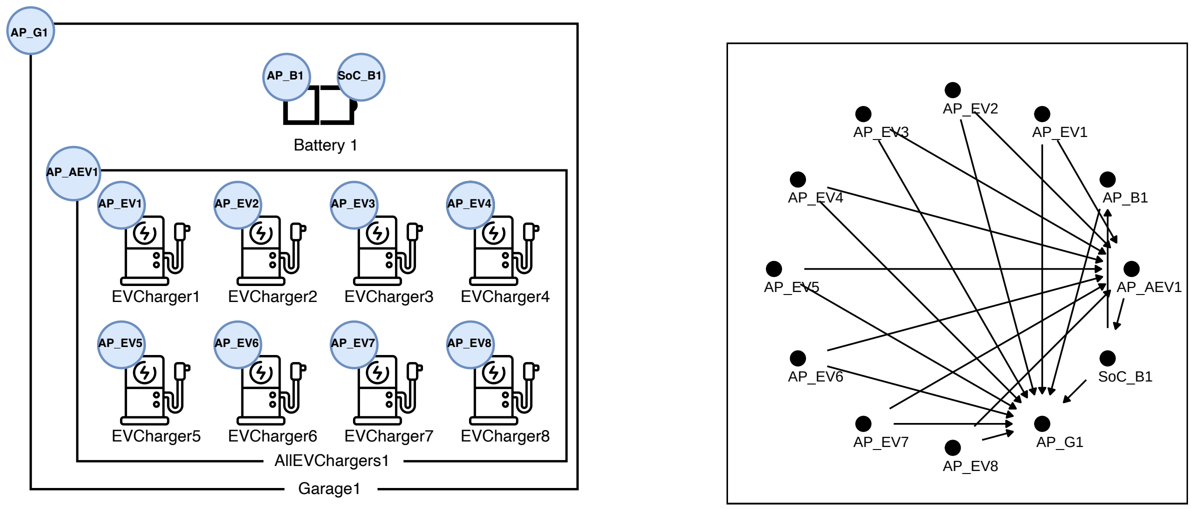
1. Introduction

In cyber-physical systems (CPSs), computational elements tightly interact with physical processes through sensors and actuators. This enables smart monitoring and control of these systems, as well as increased automation. In a smart grid, as an example of a CPS, power flow can be handled efficiently through real-time monitoring. However, it is important to keep system behaviour understandable for users (e.g., grid operators responsible for a stable grid operation), especially when some unexpected events (i.e., anomalies) occur. To ensure transparency of such systems for human stakeholders, causal models of the system can be employed to make system processes more clear, and to trace anomalies back to their root causes. Explanations based on causal knowledge are important to match users' expectations and reasoning patterns [1].

Traditionally, causal graphs are constructed by domain experts, who propose a set of causal relations between system components based on their extensive domain knowledge. This process is not only time-consuming, but it is also error-prone. For instance, in the biomedical domain, Kook [2] show that two causal graphs of protein interactions based on domain knowledge and perturbation experiments Sachs et al. [3] are likely to be misspecified. While proposed causal relations from experts are most likely correct, it is easy to miss additional causal relations, especially if they seem too obvious to some experts.

1st Workshop on Leveraging SEmaNtics for Transparency in Industrial Systems (SENTIS), co-located with SEMANTiCS'25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria

*Corresponding author.
katrin.schreiberhuber@wu.ac.at (K. Ehrenmüller); lucas.kook@wu.ac.at (L. Kook); fajar.ekaputra@wu.ac.at (F. J. Ekaputra); marta.sabou@wu.ac.at (M. Sabou)
0000-0003-1815-8167 (K. Ehrenmüller); 0000-0002-7546-7356 (L. Kook); 0000-0003-4569-2496 (F. J. Ekaputra); 0000-0001-9301-8418 (M. Sabou)



(a) The EV charging garage with its installed sensors in blue. AP = Active Power, SoC = State of Charge.

(b) A representation of the causal graph between sensors from expert knowledge.

Figure 1: Input data for experiments - EV charging garage (i.e., CPS) setup on the left, and expert-defined causal graph of system sensors on the right.

In this work, we aim to support the process of causal knowledge acquisition using a data-driven approach. To this end, we investigate Granger non-causality tests to falsify and amend an expert-defined causal graph based on time series data. This method can be used to check if there are discrepancies between a proposed causal graph (e.g., causal knowledge acquired from domain experts) and the time series data, e.g., if there is evidence for additional causal relations in the data that were missed in defining the initial causal graph.

We evaluated this approach on measurement data of an electric vehicle (EV) charging garage from our research project [1]. The garage is equipped with a set of sensors, to measure the power consumption and state of charge of different devices in the system. In Figure 1a, the garage with its devices and sensors is shown. It contains 8 EV charging stations and a battery, which can be used for peak shaving (i.e., reducing peak power consumption of the garage by providing power from the battery). There are multiple sensors installed in the system (shown as blue circles). The garage experiences an overload if the system draws too much power, indicated by `AP_G1` exceeding a certain threshold. A causal graph between sensors is needed to determine potential sources of such an overload. Over multiple workshops and discussions, experts have proposed a candidate causal graph to be used for generating explanations of anomalies. The expert-defined causal graph is represented in Figure 1b [4].

Our evaluation result shows that Granger non-causality tests can be used to identify inconsistencies between the provided causal graph and the observed sensor data. While the statistical tests indicate several possible scenarios of missing causal relations, this information can be presented to domain experts for further analysis and discussion, enabling the development of a more complete and consistent causal graph.

The rest of the paper is structured as follows. In Section 2, we discuss related work on CPS and causal discovery. Then, our proposed methods are explained in Section 3, followed by experimental results in Section 4. Finally, we discuss our contributions, limitations as well as planned future work in Section 5.

2. Related Work

Explainability of CPS is a key factor to make existing systems more transparent for human users. This aligns with the European Commission's vision of Industry 5.0, to create human-centric industrial systems

[1]https://sense-project.net/

beyond efficiency and productivity [5]. Traditional approaches from the CPS research community include Failure Mode and Effect Analysis (FMEA) [6], and Fault Tree Analysis (FTA) [7]. These methods focus on modelling functional dependencies between system components based on expert knowledge. If implemented correctly, these approaches prove to be highly reliable. However, the design and modelling is labour-intensive and relies on fragmented knowledge from various sources of domain knowledge, such as domain experts and user manuals. The fact that this knowledge is mostly represented via natural language descriptions leads to potentially ambiguous, conflicting or incomplete representations of causal relations and data flows [8].

The Semantic Web research community proposed a different approach which focuses on modelling only relations between sensors and variables that are interesting to system users, thus reducing the efforts to model a fully engineered system [4]. Ontology-based representations can help to reduce ambiguity [9, 10]. However, the exploration of potential causal relations still requires substantial manual work from highly-skilled domain experts.

Learning causal relationships from observational time series data (e.g., sensor data in CPS) can help experts in the process of creating a semantic causal representation of their system [11], in a process referred to as *causal discovery*. Traditionally, there are two approaches for causal discovery: *Constraint-based methods* apply statistical tests of conditional independence to infer an underlying causal graph that is supported by the data. *Score-based methods* use scoring methods from machine learning, such as BIC, or likelihood-based scores [12]. Recently, the use of LLMs to provide potential causal knowledge has been investigated as well [13]. However, learning the entire causal graph of a complex system is often too ambitious and available algorithms typically output a graph without uncertainty estimates. Furthermore, experts commonly already have some understanding of the causal relationships in their system which can be encoded in a candidate causal graph. Instead of learning the causal graph from scratch, tests that are used for constraint-based methods can be used to falsify and amend a candidate causal graph [2].

Our work is positioned at the intersection of knowledge acquisition and statistical causality as solution domains, while we focus on CPS as a potential problem domain for our proposed methods. Specifically, we investigate the use of Granger non-causality tests as a method to falsify and amend a causal graph that was provided by domain experts. Thus, domain knowledge and expert assumptions are tested against existing sensor data to check if it is consistent with observed values. Furthermore, this method can also provide suggestions for further causal relations that have been overlooked by experts initially.

3. Methods

We propose a procedure to find inconsistencies between an existing causal graph, derived (e.g., through expert knowledge) and sensor measurements. As shown in Figure 2, the initial causal graph implies assumptions about Granger causalities in the system which can be tested against data from system sensors. If statistical tests of Granger non-causality indicate discrepancies between the graph and the data, the graph is considered falsified. The results of these tests can suggest missing edges in the graph, which the domain expert can use to amend the existing graph and create a graph that is more consistent with the data.

Expert knowledge about causal relationships in a CPS can be encoded in a causal graph, in which edges indicate the direct influence of the history of a process on the future of another process (such as in Figure 1b). Assuming the graph is induced by a structural causal model (as in Equation 1), the graph encodes causal assumptions in the form of graphical separations, called μ-separations. By the causal Markov property, these separations imply conditional independencies (or Granger non-causalities) in the observed data distribution [14, 15]. By using data from the CPS, testing those implications allows us to falsify and amend the given causal graph. We briefly introduce the necessary tools from graphical modelling and Granger causality, before describing the proposed falsification procedure in detail.

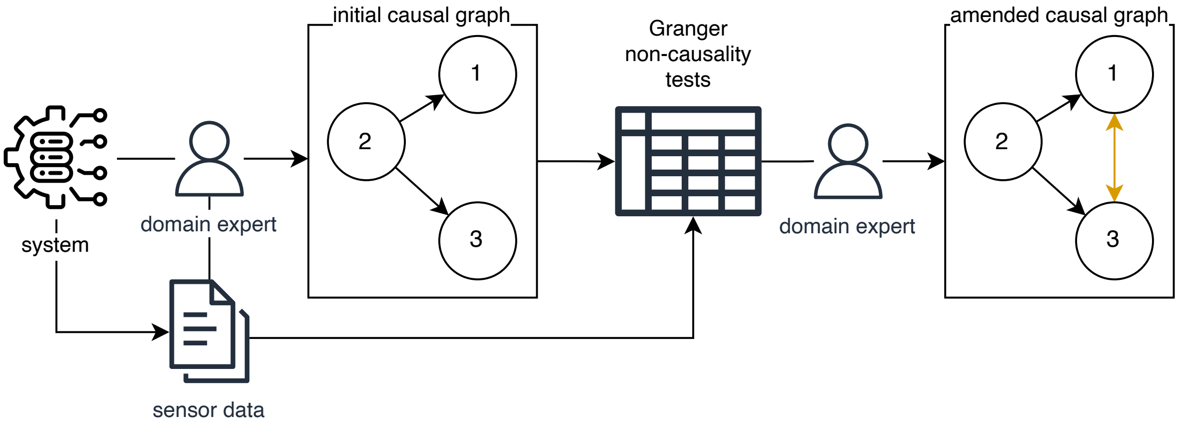
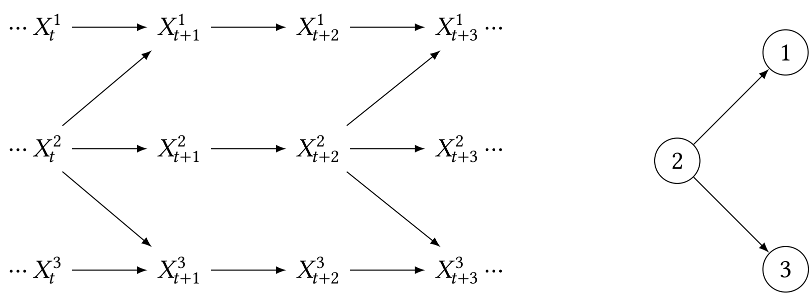


Figure 2: Overview of the proposed knowledge acquisition process using data-driven causality tests. Granger non-causality tests can suggest potentially missing causal relations between nodes, which can be checked by domain experts.



(a) Full time causal graph. (b) Summary causal graph.

Figure 3: Example of a full-time and the corresponding summary causal graph.

Structural causal models and causal graphs. We begin by introducing key concepts of structural causal models for time series [16, Chapter 10.1]. Let $(X_t)_{t\in\mathcal{T}}$ denote a multivariate time series, in which each coordinate of $X_t \in \mathbb{R}^d$ corresponds to a measurement of a sensor in the CPS and $\mathcal{T} \subseteq \mathbb{Z}$ denotes the discrete time domain. We assume the time series is generated by a structural causal model (SCM), such that, for all $t \in \mathcal{T}$,

$$X_t = F(\overline{X}_t, \epsilon_t), \tag{1}$$

where F denotes the causal mechanism (assumed to be constant over time), $\overline{X}_t^j$ denotes the history of the process before time point t and ϵ_t are independent and identically distributed noise terms. The SCM induces a so called *full-time graph* (Figure 3a) in which there is an edge from X_t^i to $X_{t'}^j$ if X_t^i appears in the structural equation of $X_{t'}^j$. Based on the full-time graph, we define the *summary graph* (henceforth referred to as the causal graph) with node set $\{1, \dots, d\}$ and in which there is an edge from i to j, if and only if there is at least one edge from X_t^i to $X_{t'}^j$ (for arbitrary time points $t < t'$) in the full-time graph (Figure 3b).

We next give the necessary prerequisites on graphical models based on the more detailed expositions in [15, 17]. The causal graph $G = (V, E)$, with node set $V = \{1, \dots, d\}$ and edge set $E \subseteq V \times V$, encodes the dependence structure of the coordinates of X_t using directed edges ($\leftarrow, \rightarrow$). A *walk*, $i_0 \sim_1 i_1 \sim_2 \dots \sim_n i_{n+1}$, is an alternating sequence of nodes and edges in G, where $\sim$ is a placeholder for any edge. A node i_k on a walk is a collider, if $i_{k-1} \rightarrow i_k \leftarrow i_{k+1}$ and a non-collider otherwise. Consider a walk ω between nodes i and j and let $C \subseteq \{1, \dots, d\}$. We call ω a *μ-connecting walk from i to j given C*, if (i) $i \notin C$, (ii) every non-collider on ω is not in C, (iii) every collider on ω is among the ancestors of C, and (iv) ω has a head at j. Finally, for $A, B, C \subseteq V$, we say *B is μ-separated from A given C* if there exists no μ-connecting walk from any $i \in A$ to any $j \in B$ given C, and write $A \not\to B \mid C$.

Granger causality. Granger causality is a predictive notion of causality in time series [18, 16, Chapter 10.3.3]. Informally, a process X_t is Granger causal for another process Y_t given another process Z_t, if $\overline{X}_t$ improves the prediction of Y_t when already accounting for $\overline{Z}_t$ [19]. Formally, we define the negation of Granger causality, i.e., *Granger non-causality*, as the following conditional independence,

$$\text{for all } t, \quad Y_t \perp\!\!\!\perp \overline{X}_t \mid \overline{Z}_t.$$

Under mild technical conditions, there is a close connection between Granger causality and μ-separations in causal graphs: If j is μ-separated from i given $C \cup \{j\}$, then X_t^i is Granger non-causal for X_t^j given $X_t^{C \cup \{j\}}$ [14, 15]. The Granger non-causalities implied by the μ-separations in a candidate causal graph are testable conditional independencies in the observed data distribution.

Falsification procedure. We list all μ-separations of the form $i \not\to j \mid C \cup \{j\}$ implied by the candidate causal graph for $|C| \leq k$. Here, k is a user-defined constant that can be chosen based on computational tractability, as a larger k results in a larger number of tests. As outlined above, each μ-separation implies a testable Granger non-causality,

$$H_0(i, j, C) \ : \ \text{for all } t, \quad \overline{X}_t^i \perp\!\!\!\perp X_t^j \mid \overline{X}_t^{C \cup \{j\}}.$$

Assuming linear relationships between $\Delta_t^j = X_t^j - X_{t-1}^j$ and $\overline{X}_t^i, \overline{X}_t^{C \cup \{j\}}$ and stationarity of Δ_t^i, we perform Granger non-causality tests as a paired t-test comparing the out-of-sample mean squared prediction error contributions of two autoregressive linear (AR(p)) models in which one model includes $\overline{X}_t^i$ while the other does not. The chance of falsely rejecting at least one null hypothesis is exacerbated when multiple tests are performed. Therefore, we apply a Bonferroni-Holm correction to control the family-wise error rate across all tested Granger non-causalities. Small adjusted p-values indicate discrepancies between the data and the candidate causal graph and indicate missing causal connections. The graph can then be amended by an expert based on the results of the Granger non-causality tests.

4. Results

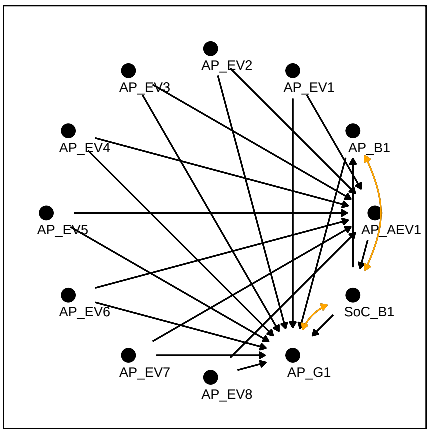
We applied the proposed falsification procedure on a real-life dataset of a smart charging garage. In total, a set of 39 sensors is installed at the facility, collecting data in 1 minute intervals. The sensors measure different variables, including active power, voltage, state of charge, energy and temperature. In our experiments, we included 12 sensors (measuring active power and state of charge), which have been identified by domain experts to be relevant for potentially identifying root causes of a power overload (see the garage setup in Figure 1a).

We used the candidate causal graph from Figure 1b and conditioning sets C of size at most $k = 1$ to limit the number of tests, considering computational tractability and power of the procedure. In total, the graph implies 1025 μ-separations at the chosen k. We applied the corresponding tests of Granger non-causality using AR(10) models across four disjoint time periods of at least 8000 minutes and combined the p-values for each hypothesis by taking the smallest p-value after a Bonferroni-Holm correction over all four time periods. We then adjusted the resulting 1025 p-values again using a Bonferroni-Holm correction. The results for all μ-separations rejected at the 5% level are shown in Table 4a.

The procedure suggests that the μ-separations (i) `AP_B1` $\not\to$ `SoC_B1` given any of the EV chargers and (ii) `AP_G1` $\not\to$ `SoC_B1` given the sum of all EV chargers are not in line with the data at the 5% level. Therefore, there is strong evidence that either `AP_B1` and `AP_G1` cause `SoC_B1` directly, or that there are unobserved processes that influence `AP_B1` and `SoC_B1` or `AP_G1` and `SoC_B1`. The amended graph in

i	j	C	Adjusted p-value
AP_B1	SoC_B1		< 0.001
AP_B1	SoC_B1	AP_AEV1	< 0.001
AP_B1	SoC_B1	AP_EV1	< 0.001
AP_B1	SoC_B1	AP_EV2	< 0.001
AP_B1	SoC_B1	AP_EV3	< 0.001
AP_B1	SoC_B1	AP_EV4	< 0.001
AP_B1	SoC_B1	AP_EV5	< 0.001
AP_B1	SoC_B1	AP_EV6	< 0.001
AP_B1	SoC_B1	AP_EV7	< 0.001
AP_B1	SoC_B1	AP_EV8	< 0.001
AP_G1	SoC_B1	AP_AEV1	< 0.001

(a) Results of the Granger non-causality tests.



(b) Amended causal graph.

Figure 4: Test results for the rejected μ-separations and the corresponding amended causal graph, in which a bidirected edge (orange) was added between i and j.

Figure 4b includes bidirected edges between the two pairs of nodes to emphasize this ambiguity. While the ambiguity cannot be resolved based on the given data, a domain expert may be able to judge the plausibility of these suggested edges.

From a domain-perspective (i) suggests that active power from the battery (AP_B1) could have an effect on the state of charge of the battery (SoC_B1) according to observational data. Since power consumption of the battery is used to either discharge, or charge the battery, this effect is plausible.

Furthermore, (ii) would suggest that power consumption of the garage (AP_G1) is related to the state of charge of the battery (SoC_B1), when controlling for EV charging overall (AP_AEV1). The state of charge of the battery usually changes either by discharging (usually through providing power to EV chargers), or by charging (power is consumed from the external grid). Thus, this yet undefined relation between (AP_G1) and (SoC_B1) should be further investigated by domain experts.

5. Conclusion

Causal relations are a key component for human users to be able to interpret CPS behaviour. In this paper, we have addressed the challenge of constructing a causal graph for such systems. While causal graphs that are informed by domain experts are grounded in extensive domain knowledge, they can be incomplete as experts might miss certain causal relations. Data-driven approaches can uncover such relations. However, learning entire causal graphs from observational data alone can be difficult and unreliable.

To bridge this gap, we propose testing implied conditional independencies that are created by a causal graph from system experts on observational sensor data. Specifically, we apply the proposed tests to a real-life use case of a smart charging garage, where active power and state of charge sensor measurements are available that correspond to the proposed causal graph. Granger non-causality tests are applied to check if the causal graph aligns with existing sensor data.

In our experiments, we have successfully identified some inconsistencies between the provided causal graph and sensor data. While the statistical tests suggest multiple scenarios of which causal relations are missing, this knowledge can be presented to domain experts for further discussion, to define a more complete and consistent causal graph.

Currently, we have applied our approach only to observational data, testing for statistical independence. While we can, in this way, investigate predictive relationships between variables, causal relations can be investigated with more confidence and in more detail by actively manipulating the system (e.g. disconnecting an EV charger, or starting fast-charging sessions). Obtaining data from different interventional settings could allow the use of causal discovery methods that rely on different data sources [e.g., 20] and thereby greatly improve the causal understanding of the system. Additionally, the Granger non-causality tests used in this work rely on parametric modelling assumptions. Under model misspecification, the test results may be unreliable and reflect inflated type I error rates, rendering expert-based evaluation of the results invaluable.

In future work, we will investigate how the proposed falsification procedure performs on data from the system under interventions. Furthermore, we will investigate the potential of expert validation as a more integrated causal discovery process, potentially integrating multiple causal discovery methods from experts, data and LLMs. Finally, nonparametric tests of Granger non-causality can be considered to alleviate the problem of misspecification.

Acknowledgments

This work was supported by the FFG research project SENSE (project Nr. FO999894802), and the Austrian Science Fund (FWF) Bilateral AI project (Grant Nr. 10.55776/COE12).

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4.5-turbo and Perplexity AI in order to: paraphrase and reword, drafting content. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Pearl, An introduction to causal inference, The International Journal of Biostatistics 6 (2010).
[2] L. Kook, Falsifying causal models via nonparametric conditional independence testing, in: D. O'Sullivan, J. Sweeney, S. Moghaddam, K. Burke (Eds.), Proceedings of the 39th International Workshop on Statistical Modelling (IWSM), 2025, pp. 229–233.
[3] K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, G. P. Nolan, Causal protein-signaling networks derived from multiparameter single-cell data, Science 308 (2005) 523–529. doi:10.1126/science.1105809.
[4] K. Schreiberhuber, F. J. Ekaputra, M. Sabou, D. Hauer, K. Diwold, T. Frühwirth, G. Steindl, T. Schwarzinger, Towards a state explanation framework in cyber-physical systems, in: DACH+ Energy Informatics 2024, Lugano, Switzerland, Proceedings, ACM SIG Energy, 2024, pp. 1–13.
[5] E. Commission, D.-G. for Research, Innovation, A. Renda, S. Schwaag Serger, D. Tataj, A. Morlet, D. Isaksson, F. Martins, M. Mir Roca, C. Hidalgo, A. Huang, S. Dixson-Declève, P. Balland, F. Bria, C. Charveriat, K. Dunlop, E. Giovannini, Industry 5.0, a transformative vision for Europe – Governing systemic transformations towards a sustainable industry, Publications Office of the European Union, 2021. doi:doi/10.2777/17322.
[6] M. Ben-Daya, Failure mode and effect analysis, in: Handbook of maintenance management and engineering, Springer, 2009, pp. 75–90.
[7] C. A. Ericson, Fault tree analysis primer, CreateSpace Incorporated, 2011.
[8] L. Dittmann, T. Rademacher, S. Zelewski, Performing fmea using ontologies, in: 18th International Workshop on Qualitative Reasoning. Evanston USA, 2004, pp. 209–216.
[9] Z. Rehman, C. V. Kifor, An ontology to support semantic management of fmea knowledge, International Journal of Computers Communications & Control 11 (2016) 507–521.
[10] A. Zhou, D. Yu, W. Zhang, A research on intelligent fault diagnosis of wind turbines based on ontology and fmeca, Advanced Engineering Informatics 29 (2015) 115–125.
[11] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, J. Zscheischler, Inferring causation

from time series in earth system sciences, Nature Communications 10 (2019). doi:10.1038/s41467-019-10105-3.
[12] P. Spirtes, C. N. Glymour, R. Scheines, Causation, Prediction, and Search, MIT press, 2000. doi:10.7551/mitpress/1754.001.0001.
[13] C. K. Assaad, E. Devijver, E. Gaussier, Survey and evaluation of causal discovery methods for time series, Journal of Artificial Intelligence Research 73 (2022) 767–819.
[14] M. Eichler, Graphical modelling of multivariate time series, Probability Theory and Related Fields 153 (2011) 233–268. doi:10.1007/s00440-011-0345-8.
[15] S. W. Mogensen, N. R. Hansen, Markov equivalence of marginalized local independence graphs, Annals of Statistics 48 (2020) 539–559. doi:10.1214/19-AOS1821.
[16] J. Peters, D. Janzing, B. Schölkopf, Elements of Causal Inference: Foundations and Learning Algorithms, The MIT Press, 2017.
[17] S. W. Mogensen, Weak equivalence of local independence graphs, Bernoulli (2025). doi:10.48550/arXiv.2302.12541, to appear.
[18] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (1969) 424–438. URL: http://www.jstor.org/stable/1912791.
[19] A. Shojaie, E. B. Fox, Granger causality: A review and recent advances, Annual Review of Statistics and Its Application 9 (2022) 289–319. doi:10.1146/annurev-statistics-040120-010930.
[20] L. Kook, S. Saengkyongam, A. R. Lundborg, T. Hothorn, J. Peters, Model-based causal feature selection for general response types, Journal of the American Statistical Association 120 (2025) 1090–1101. doi:10.1080/01621459.2024.2395588.