Causality Inference for Digital Twins in GPU Data Centers and Smart Grids

Rolando P. Hong Enriquez
Hewlett Packard Labs
Oxfordshire, UK
rhong@hpe.com

Pavana Prakash
Hewlett Packard Labs
Milpitas, CA USA
prakash@hpe.com

Ebad Taheri
Hewlett Packard Labs
Milpitas, CA USA
taheri.ebad@gmail.com

Aditya Dhakal
Hewlett Packard Labs
Milpitas, CA USA
aditya.dhakal@hpe.com

Matthias Maiterth
Oak Ridge National Laboratory
Oak Ridge, TN USA
maiterthm@ornl.gov

Wesley Brewer
Oak Ridge National Laboratory
Oak Ridge, TN USA
brewerwh@ornl.gov

Dejan Milojicic
Hewlett Packard Labs
Milpitas, CA USA
dejan.milojicic@hpe.com

***Abstract*—To the benefit of both technologies, data centers and smart grids are likely to become increasingly integrated in the near future. The downside is that effectively managing those systems will rapidly become burdensome if we neglect adequate preparation. Digital twins that can potentially wrap the benefits of advance analytics and visualizations to manage such complex environments. Yet even today's AI systems lack the proper causal understanding of the data. Here we embark on a journey to collect proper causal data for validating causal inference methods based on three fundamentally different theoretical foundations: causal calculus, information theory, and dynamical system theory. Subsequently, we apply such methods to two target datasets from a smart grid and a GPU data center. We finally analyze the success and failures of applying these methodologies and the indications they offer to create more insightful and energy-efficient prediction strategies for digital twins in support of smart grids and GPU data centers.**

***Index Terms*—causality inference, transfer entropy, causal calculus, convergent cross mapping, dynamical systems, digital twins, smart grids.**

I. Introduction

A. Digital Twins (DTs) for data centers

With a current global revenue of ∼350 billion USD projected to almost double by 2029, the data centers marked is the backbone of modern computing, providing the necessary infrastructure for cloud services, artificial intelligence workloads, and high-performance computing applications [1].Data centers vary by scale, ownership, and purpose, including enterprise (private IT infrastructure), co-location (shared leased space), hyperscale (mostly cloud), and micro/edge (decentralized for low-latency processing) [2].Each type presents distinct operational challenges, often necessitating tailored management strategies with the support of advanced analytics for optimal performance and reliability.

On the other hand, the complexity of modern data centers also brings challenges when handling their energy efficiency. A visionary way to deal with these problems in the long term is by the introduction of smart grids in the technology landscape. Smart grids represent a qualitatively different advance from the now legacy electrical grids. The basic premise is to establish bidirectional flows of both electricity and information with "intelligent devices", understood as those that have the capabilities to operate and interact with similar devices using any sort of network technology. This broad definition accommodates many generic devices. Relevant to our work are computational data centers, which have the possibility to reduce their large energy footprints [3], [4] by participating in smart grid programs [5]. In this sense, bidirectional communication models between data centers and smart grids have long been proposed to optimize the power flow to the data centers [6]. On the other hand, new evolutions of this conceptual framework try to maximize data center services to smart grids [7]. The final goal is a fully dynamical integration between these infrastructures [8]. However, the road to fulfill these aspirations is not without some mayor challenges. For instance, traditional data centers are designed by vendors and their operations within smart grid systems are often inefficient. To mitigate these issues, an intermediate layer of software-enhanced virtualized data centers has been proposed in the past [9]. Alternatively, the use of Digital Twing technologies could also provide interesting solutions to these problems [10]. Eventually, some of these incipient digital twins should widen their scope to manage and control those integrated smart-grid/data-center infrastructures.

B. Data Analytics in DTs

Advanced data analytics is needed when the variety, quantity, and complexity of the data becomes a burden to human understanding. Such is precisely the case for data centers. Here, besides the complexity of the domain specific pipelines running in production, modern data centers must handle ever growing infrastructure management layers [11]. Therefore to be effective, future DTs aiming to go beyond system monitoring will need to use a variety of data analytics methods and techniques. Currently, most data centers use a range of techniques that go from simple filtering, clustering and summarization techniques to more advanced techniques such as support vector machine (SVM) [12], the use of

AI agents [13], generative AI [14] or deep reinforcement learning [15] to name a few. The covered applications of advanced analytics go from the rather foundational problems of energy savings [12], [16] and server consolidation [17] to more challenging cases that require adapting AI techniques for space data centers [18]. However, the application of data analytics by itself does not guarantee understanding of the systems replicated by DTs. More often than not, the task of understanding is left to a human operator with the help of targeted visualizations. In recent years, in a race to match the increases in data quantity and complexity, ever larger AI models have been implemented to improve data center operations. While these models have improved accuracy, they come with undesired collateral liabilities like higher power consumption and diminished transparency and explainability. To be fully trusted and sustainable in the long term, future management strategies for data center management need to deal with these issues. We claim that this can be at least partially accomplished by introducing techniques to uncover causal relationships in data centers at large.

C. Main ideas behind causality studies

Understanding causality has been an informal part of human thinking since time immemorial. The notion that we can make sense of our very existence by understanding which events can be the cause of other events has long be part of religion, philosophy and more recently natural sciences, specially physics and mathematics. We will not concern ourselves with these big questions in this work. Instead we will only use notions of causality than can conceivably have practical implications on the improvement of data-driven data analytics with minimal or none human intervention. Specifically, we will explore three types of methodologies that try to capture the concept of causality in different ways.

- *Causal calculus*. A formal language and framework deeply rooted on probability theory.
- *Transfer entropy*. A notion derived from information theory
- *Convergent Cross Mappings*. A method based on dynamical system theory.

In their respective sections we explain the theoretical basis for these methods followed by their respective experimental sections using different types of causality datasets.

II. Data for causality

Different data-driven approaches to determine causality relationships are both, theoretically and pragmatically incompatible. Therefore here we also include validation data for which we know the ground truth as a way to test the strengths and weaknesses of the methodologies we are reporting. This validation data comes in two flavors: *experimental* and *mathematically generated* data. Causal analysis in digital twins relies on different types of data to establish cause-and-effect relationships. The choice of data affects the accuracy and applicability of causal models, making it essential to use the right data sources for specific problems. In this study we will work with GPU failure on data centers.

A. Validation data - experimental

Experimental benchmarking data for causal studies is relatively scarce. In this work we took data from various sources to created our own testing data suite. Similarly to the mathematically generated data, the suite is divided into *causal pairs* and *general causal structures*. The causal pairs are a collection of files with only two columns of numerical data standing for variables x and y. These data comes with an associated file reporting the ground truth for the causality relationship. For instance, (1) x causes y, ($x \Rightarrow y$), (2) y causes x, ($x \Leftarrow y$), (3) x and y cause each other, ($x \Leftrightarrow y$) and (4) no causality, $x \nLeftrightarrow y$. To the best of our knowledge though, there is no reliable experimental data for case (3) and most sources only report cases (1) and (2). We finally used a curated selection of 93 causal pairs originally published by Mooiji et al. [19]. Although the original work reported 37 datasets from various domains (e.g., finance, biology, engineering), this collection has been increased to more than 100 datasets and is currently maintained online [20].

B. Validation data - mathematically generated

This type of data is created using predefined equations and simulation models that encode known causal relationships. It is particularly useful for testing and validating causal inference methods before applying them to real-world systems. For example, in a digital twin, mathematical models can simulate how CPU usage, cooling efficiency, and power consumption interact over time, allowing researchers to study causal effects without requiring real-world data collection.

Similarly to the experimental validation data, our mathematically generated data suite is also mainly divided into *causal pairs* and *general causal structures*, but given the flexibility and control of this alternative we also generated smaller testing datasets to explore other aspects of causality such as the strength of the causal effects. We describe all our data generation strategies bellow.

Csuite. We started by including causal mathematically generated causal pairs as they were reported in the Csuite [21], [22]. An example of a causal pair in the Csuite is depicted in Figure 1. A full theoretical description of this model can be found under *Example 1* in the original paper [21].

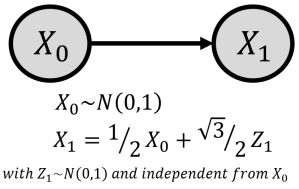


Fig. 1. Structural equations for a causal pair formed by a two node linear gaussian system.

Additional causal pairs were extracted from complex causal models in the Csuite. For instance from a model of a Simpson

paradox depicted in Figure X, we extracted four causal pairs: (X_0, X_1), (X_0, X_2), (X_1, X_2), and (X_2, X_3). Besides using these causal pairs for initial benchmarking, we also used the full causal model in other calculations.

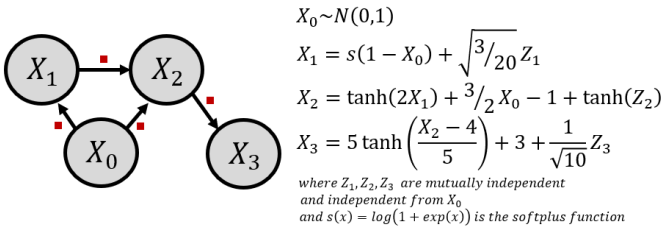


Fig. 2. Extracting causal pairs from complex causal models. In this example we show the structural equations from a case study of the Simpson paradox. Four causal pairs (red squares) were extracted from this model.

Dynamical systems. A good source of mathematically generated causal data is dynamical system theory. Our first case study in this area is the couple logistic map, which draws inspiration mainly from population dynamics and represents a discrete dynamical system defined by the following difference equations:

$$X(t+1) = X(t)[r_x - r_x X(t) - \beta_{x,y} Y(t)]$$
$$Y(t+1) = Y(t)[r_y - r_y Y(t) - \beta_{y,x} X(t)]$$

The important thing to notice for our purposes is that the variables X and Y are able to influence (i.e., "cause") each other by the coupling terms $\beta_{x,y}$ and $\beta_{y,x}$. Following Sugihara et al. [23] we used $X_0 = 0.4$ and $Y_0 = 0.2$ as initial conditions , and $r_x = 3.8$ and $r_y = 3.5$ as general parameters. Tuning the coupling terms we can generate relevant causal pairs in the following manner: (1) $x \Rightarrow y$, ($\beta_{x,y} = 0.02$, $\beta_{y,x} = 0.1$), (2) $x \Leftarrow y$, ($\beta_{x,y} = 0.1$,$\beta_{y,x} = 0.02$), (3) $x \Leftrightarrow y$, ($\beta_{x,y} = 0.1$, $\beta_{y,x} = 0.1$), and (4) $x \nLeftrightarrow y$ ($\beta_{x,y} = 0$, $\beta_{y,x} = 0$).

We implemented a second strategy for data generation in this area which involves using a chaotic signal coming from one dynamical system to drive (i.e., "cause") the behavior of another dynamical system. Specifically, we followed the approach from Quyen et al. [24] that involves using the Rössler attractor [25], [26] to modify the behavior of the Lorenz attractor [27]. Our implementation followed the simplified equations bellow:

X - Rössler attractor

$$\dot{x} = -(y + z) \quad (1)$$
$$\dot{y} = x + \alpha y \quad (2)$$
$$\dot{z} = \beta + z(x - \gamma) \quad (3)$$

Y - Lorenz attractor

$$\dot{x} = \delta(y - x) \quad (4)$$
$$\dot{y} = \epsilon x - y - xz + Cy \quad (5)$$
$$\dot{z} = xy - \zeta z \quad (6)$$

The relevant observation is that the factor forcing the normal Lorenz system (in blue) to behave differently is the addition of a factor Cy coming from the Rössler attractor. Cy is correspondingly added to the second equation of the Lorenz system. Here, the parameter C represents the strength of such coupling and as such we use it to create data to test the sensibility of causality methodologies to gradual changes of C. The rest of the parameters for both Lorenz and Rössler attractors were kept constant ($\alpha = 0.2$, $\beta = 0.2$, $\gamma = 5.7$, $\delta = 10$, $\epsilon = 28$, $\zeta = 2.667$). An example of this coupled dynamics for $C = 9.1$ is shown in Figure 3.

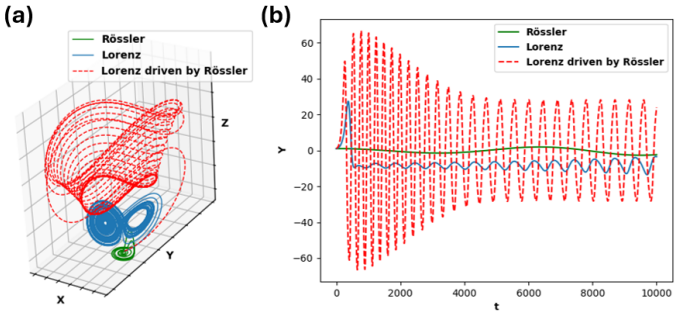


Fig. 3. Evolution of free and coupled chaotic systems. Three systems are shown: the Rössler attractor, the Lorenz attractor, and the evolution of the Lorenz attractor under the influence of the Rössler attractor. (a) 3D view of the phase-space time evolution of the dynamical systems under study (b) Projection of the time evolution on the Y axis which is the initial contact dimension of Rössler's driving effect on the Lorenz system.

Please note that several possibilities for causal data generation are possible using this system . That is, independently or as as a group, all the variables from the Rössler system *causally affect* all the variables from the Lorenz system. However, given that the direct coupling happens in the Y dimensions (Eqn. (5) of the Lorenz system), the strongest causal pairs should result from the projections on the axis Y. For instance Rössler(Y) $\Rightarrow$ Lorenz(Y), Figure 3(b).

In this section we also include other similar strategies for data generation using simplified bio-geothermal models and others. In Table I we resumed the data sources used in this paper and additional details are reported in the Appendix **??**.

TABLE I
DATASETS USED IN THIS WORK.

Dataset	Source	Type	References
Summit[1]	Experimental	Mixed	[28]
Smart Grid[1]	Experimental	Mixed	[29]
inHouse	Experimental	Causal Pairs	[19], [20]
Csuite	Generated	Mixed	[21], [22]
Couple Logistic Maps	Generated	Causal Pairs	[23]
Driven attractors	Generated	Mixed	[24]
BioGeoScience	Generated	DAG	[30]
Health BMI	Generated	Mixed	[31]

[1] Target dataset.

C. Target datasets

1) Failures in data centers: The most reliable form of data for causality comes from real-world experiments, particularly failure events of data centers in our experiments. We use observational data which includes logs of errors and failures. Controlled experiments, such as stress-testing infrastructure under varying conditions, can further enhance causal discovery by actively manipulating variables to observe their effects. Using experimental data, digital twins can refine their models and improve predictive capabilities for proactive failure prevention.

To evaluate our approach, we utilized the Summit dataset [28], which contains node-level data on power, energy, job scheduling, and failures observed over three years, covering 27,648 Tesla V100 GPUs. Specifically, from Summit we curated an extract enriched with failure data, which is quite sparse in the original data. The final data contains 127 entries of the following features: energy (E), maximal core temperature (MCT), maximal memory temperature (MMT), memory temperature (MT), core temperature (CT), power (P), job length (JL), GPU usage (GU), failure (F), core temperature fluctuation (CTF), memory temperature fluctuation (MTF).

2) Operational smart grid: The literature on smart grids has matured over the last few years and currently there is plenty of datasets dedicated to several aspects of smart grids; from load forecasting, customer behavior or anomaly detection (reviewed by Altamini et al. [32]). Although we are particularly interested in smart grids as long as they are related to data centers and ideally we would like to study synchronized recording from both types or operations, this work represents our first step in this direction. Consequently we choose a relatively simple public dataset containing only centralized electric data with no direct connection to data centers [29]. Specifically, the dataset contains 50k entries of the following timestamped numerical features (abbreviation in parenthesis): Voltage (V), current (A), power consumption (Pc), reactive ower (rP), power factor (fP), solar power (sP), wind power (wP), grid supply (gS), voltage fluctuation (Vf), overload condition (Ov), transformer fault (Tf), temperature (Tp), humidity (H), electricity price (Ep), predicted load (Pl).

III. Causality inference

Causal inference is about figuring out the impact of a treatment, policy, intervention, or any change on an outcome. There are several ways to approach it, including causal calculus, information theory, and dynamical system theory, each offering different perspectives on cause-and-effect relationships. These methods help uncover hidden dependencies and guide decision-making in various fields.

A. Causal calculus

1) Theory: Causal calculus, also known as do-calculus, Pearl's Causal Calculus or Calculus of Actions is a formal language introduced by Judea Pearl [33]. This framework makes extensive use of Bayesian networks, which are *Directed Acyclic Graphs* (DAGs) whose nodes are represented by random variables such as:

$$p(x_1, ..., x_d) = \prod_{j=1}^{d} p(x_j | x_{pa(j)})$$

This formulation assumes that a variable or node is independent of previous non-parents nodes given the parents, that is $p(x_j|x_1, ..., x_d) = p(x_j|x_{pa(j)})$. Therefore, this formulation captures dependency between variables or nodes; if there is no link (arrow) between nodes A and B, then they are probabilistically independent $(p(A, B) = p(A) \times p(B))$. Building on top of these foundations, causal calculus provides a formal framework for analyzing cause-and-effect relationships, enabling reasoning about interventions and counterfactuals.

One of the core concepts in causal calculus is the do-calculus, which allows for reasoning about interventions. Unlike traditional statistical approaches that rely on correlation, do-calculus enables the estimation of causal effects by explicitly modeling how changes in one variable influence another under intervention. This is crucial for decision-making scenarios where we aim to predict the outcome of potential actions.

Counterfactual reasoning is another essential component of causal calculus. The counterfactual question—"What would have happened if a different decision had been made?"—is a fundamental problem in causality. This problem is fundamental because counterfactuals cannot be directly observed, making it challenging to validate causal claims. Understanding alternative outcomes is necessary for decision-making, policy evaluation, and scientific discovery. Causal calculus addresses this problem by providing a structured approach to infer counterfactual outcomes based on available data and causal models.

Another significant challenge in causal inference is determining causation in purely observational data. Traditional statistical methods often struggle to distinguish between correlation and causation due to confounding factors. Causal calculus provides mathematical tools to disentangle these relationships by leveraging assumptions about the underlying causal structure. This enables more reliable conclusions about the true causal effects, even in the absence of experimental data. We use this approach we performed causal graph nonlinear estimation using the python library *Causality* [34].

2) Experiments: In this section we present the results of using the *causal calculus* methodology on three causal datasets. The first case presented in Figure 4 is the coupled Rössler-Lorenz attractor [24]. Specifically, Figure 4 (a) represents the minimal ground truth for the causality graph as we know from equation 5 that the variable y from the Rössler attractor (x2 in the graph) directly drives variable y from the Lorenz attractor (y2 in the graph). We also know by construction that the variables within any of the individual attractors are causally related (the triangles in the figure). On the other hand, Figure 4 (b) shows the results of applying causal calculus to this dataset. Remarkably, the calculations show a more complex pattern of causal interactions; the variables from

the Rössler attractor (x1,x2,x3) now seem to have a stronger driving (causal) influence over the Lorenz attractor that among themselves. Other tests indicate that the origin of this effect is the use of a relatively high value for the coupling constant C when generating the causal dataset for the coupled attractors (see equation 5).

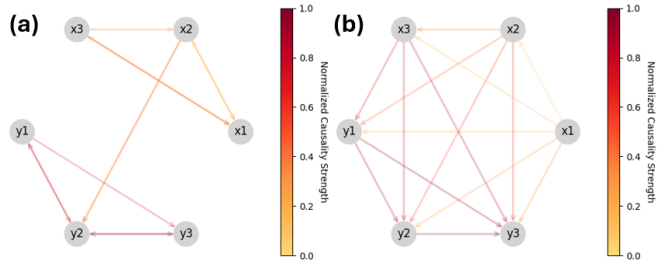


Fig. 4. Performance of causal calculus on the coupled Rossler-Lorenz attractor [24].(a) Ground truth - polynomial graph ($p(x) = x^6$) (b) Generated data -polynomial graph ($p(x) = x^6$). The estimated graph is not isomorphic with the ground truth; their graph edit distance is 13.

In Figure 5 we show the use of causal calculus on the Bio-GeoScience generated dataset (see Appendix A). Specifically, Figure 5 (a) shows the ground truth for this system built from the generating equations and (b) shows the graph discovered using causal calculus. While the calculations recovered some of the direct causal links from the ground truth (e.g., $C \Rightarrow D$), the causality direction was reversed in some cases (e.g., $C \Rightarrow B$ instead of $B \Rightarrow C$) missed completely in other cases (e.g., $A \Rightarrow B$).

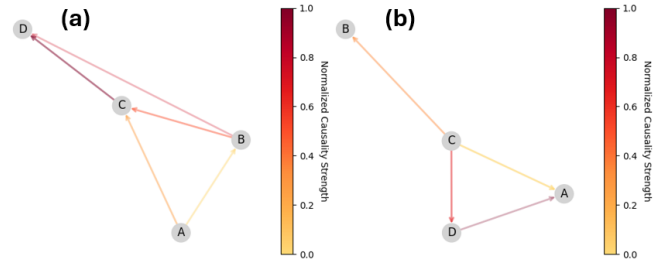


Fig. 5. Performance of causal calculus on the BioGeoScience dataset [30].(a) Ground truth - polynomial graph ($p(x) = x^4$) (b) Generated data -polynomial graph ($p(x) = x^4$). Despite the closeness of the genetated graph with its ground truth, the methods misrepresents the directionality of several causal relationships. The estimated graph is not isomorphic with the ground truth; their edit distance is 3.

In Figure 6 we show the results of applying causal calculus to the experimental fetal ECG dataset [35]. This is a multichannel ECG recording with one channel (D) coming directly from the baby's head at labor and the other channels ($Ab1, ...Ab4$) are indirect measures taken from the abdomen of the mother. An extra channel we certainly know is not functionally related in any way to the measurements is channel E (annotations). Now even though we cannot make the claim that channel D "causes" the others we can certainly assumed that they are related to each other and that the strength of this relationship is probably related to how far are the recording sensors from each other which is an information not provided in the dataset, therefore to the best of our knowledge, the ground truth for this case should be the one represented in Figure 6 (a). The causal calculus estimated inference, remarkably was able to exclude channel E from the causal graph while unveiling most of the other causal links between the ECG channels. Additionally, there are preferential connections to channel D although the directionality of these connections is unexpected ($Abi \Rightarrow D$).

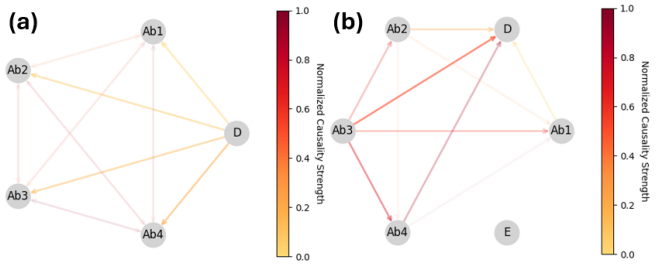


Fig. 6. Performance of causal calculus on the fetal ECG dataset [35].(a) Ground truth - polynomial graph ($p(x) = x^6$) (b) Generated data -polynomial graph ($p(x) = x^5$). This method seems to recognize valid causal relationships and correctly excludes E as a causal variable. However, the estimated graph is not isomorphic with the ground truth; its edit distance from it is 12.

B. *Information theory*

1) Theory: Partially inspired by the work of Kolmogorov [1], Norbert Wiener proposed in 1956 his "Theory of predictions" [37]. In a nutshell, the philosophy behind his approach was that a causal factor has the effect of increasing the predictability of an outcome variable under study. That is, the statement "X causes Y" ($X \Rightarrow Y$) implies that we have better chances of predicting the correct value of variable Y by including the past and present values of X, than by using the past values of Y alone. While Wiener moved ahead using traditional mathematical tools, Schreiber derived a rigorous solution using Information theory [38].

In a truly remarkable and influential paper, Claude Shannon almost single-handily invented Information theory in 1948 [39]. One of the immediate results from his theory is the concept of mutual information ($I_{X,Y}$) which was eventually used to quantify causal relationships. However, causality implies a preferred direction for cause and effect and mutual information was promptly discarded because of it symmetry ($I_{X,Y} = I_{Y,X}$). On the other hand, the time-lagged version of mutual information complies with the directionality requirement; that is, $I_{X,Y}(t) \neq I_{Y,X}(t)$ (see Appendix A). However, Schreiber pointed out that it is possible to incorporate time dynamicity organically by considering transition instead of static probabilities; in doing so he developed the concept of transfer entropy. Assuming that two time series variables $X = x_t$ and $Y = y_t$ can be approximated by Markov processes, we can then write the backbone of Wiener's causality criteria as:

[1]See Mandrekar [36] for more extensive mathematical background

$$p(y_{t+1}|y_t^n, x_t^m) = p(y_{t+1}|y_t^n)$$

Causality can then be measured as the difference between the probability distributions on the right and left part of the equation. A way to quantify this difference is using the Kullback-Leibler divergence between probability distributions. The final result is the *transfer entropy* (TE):

$$TE_{X,Y} = \sum_{y_{t+1}, y_t^n, x_t^m} p(y_{t+1}, y_t^n, x_t^m) log\left(\frac{p(y_{t+1}|y_t^n, x_t^m)}{p(y_{t+1}|y_t^n)}\right)$$

Most of the time, Transfer energy is applied to pairs of measurements (e.g., like time series). Yet, using TE in this bivariate way can potentially underestimate causal effects contributions from other sources. This is analogous to measuring contributions from pairs of nodes in a network of graph; the effects resulting from the interactions [40] of other nodes will also be underestimated. To account for these affects in this section we use a multivariate implementation of the transfer entropy [41], [42].

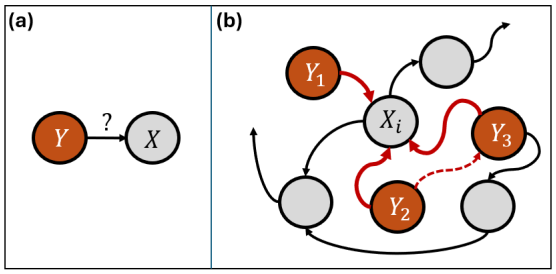


Fig. 7. Transfer entropy. (a) Bivariate case. Aims to determine if there is a causal link between selected nodes. (b) Multivariate case. Aims to find the set of nodes $V_{X_i} = \{Y_1, Y_2, Y_3\}$ that can be used to predict the next state of node X with statistical significance. Repeat and optimized the procedure for all the nodes X_i in the network of potential causal variables

Give that an exhaustive multivariate transfer energy calculation soon becomes impractical due to combinatorial explosion, optimization search strategies are needed. A detailed theoretical description of the method used here was given by Lizier and Rubinovis [40], while an example of its application on physiological data can be found on Vicente et al. [43].

Within information theory, information processing as initially envisioned by Turing [44] should be able to be decomposed into information transfer (accounted for by the transfer entropy just described), information modification (for which, to the best of our knowledge we are yet to find a proper formalization) and information storage, which we will cover briefly now. Specifically, in this section we use the concept of Active Information Storage (AIS) [45], [46]. The main idea is finding a measures of how much the past of a time series can be use to make predictions about the future and here we distinguish two cases: (a) Making predictions *at some point in the future*, which is capture by the concept of total information storage and can me measure in practice by the the *excess entropy*. We will not use this measure here but for details please see Lizier et al. [45]. (b) Making predictions about the *next state* of a variable or process, which is done by determining the amount of past information that is *currently* in use by such variable. This can be measured by the *Active Storage Information*. Effectively, for a current proces process in state i with a memory of k steps or states in the past, we measure AIS as the mutual information between those k states and the future state $i+1$ and we repeat this for every state i in the time series. Formally, for the x_t realizations of a variable X_t with finite memory k, the mutual information between the past values in memory $X_{t-1}^{k-} = (X_{t-1}, ...X_{t-1-k})$ and the next state X_t is:

$$A_{X_t} = I(X_{t-1}^{k-}, X_t) = \left\langle log\frac{p_t(x_t|x_{t-1}^{k-})}{p_t(x_t)} \right\rangle$$

2) Experiments: In Figure 8 we show the multivariate transfer energy causal estimations (mTE) on the BioGeoScience dataset [30]. An important thing to notice is that mTE brings time causality lags as a new type of information, this feature is not provided my other methods like causal calculus. Additionally, mTE correctly recognizes the presence and directionality of some of the causal relationships in the model (e.g., $A \Rightarrow B$ and $A \Rightarrow B$) missed one true causal relationship ($B \Rightarrow D$) and unveils a new, probably indirect causal relationship with time lag 5 ($A \Rightarrow D$). This later link is not part of the original ground truth, but it might as well be included as it is reasonably to assume that such causal relationship should exist although maybe with less strength (as indicated by a higher time lag). In this figure we also use the relative size of each node to map the values of their Active Information Storage (AIS). The results indicate that the main causal driver node in this model, node A, has the highest AIS values. This indicates A contains within itself most of the information to predict it future state. On the contrary, the main causally driven variable in the model, node D has the lowest AIS values. This indicates that most of the information needed to predict the future states of D does not come from itself, comes from other nodes.

The results in Figure 9 show the causality discovery process using multivariate transfer entropy (mTE) on a fetal ElectroCardioGram (EGC) dataset [35]. The experimental dataset represents a multichannel ECG recording of woman in labor. Five signals are taken from the abdomen and are labeled $Ab_1, ..., Ab_5$ and another signal comes directly from the baby's head which is labeled D. An additional signal E, contains causally unrelated numerical annotations.

C. Dynamical system theory

1) Theory: In this section we use causality inference methods based on dynamical system theory which in general aims to describe the behavior complex systems by using differential or difference equations. This area represents both an interesting and challenging use case for causality analysis for several reasons. First, it was a disturbing fact for both

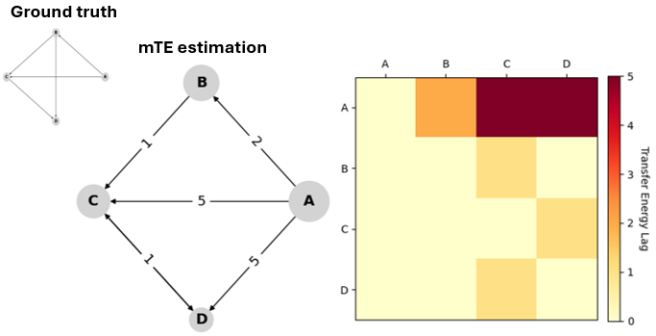


Fig. 8. Performance of Multivariate Transfer Energy (mTE) on the BioGeo-Science dataset [30].The ground truth for this model is shown in the inset. Although mTE recognizes valid causal relationships the graph polynomial extracted from the estimated model ($p(x) = x^4 - x^2$) is different from the one extracted from the ground truth graph ($p(x) = x^4$). The estimated graph is also not isomorphic with the ground truth; its edit distance from it is 3.

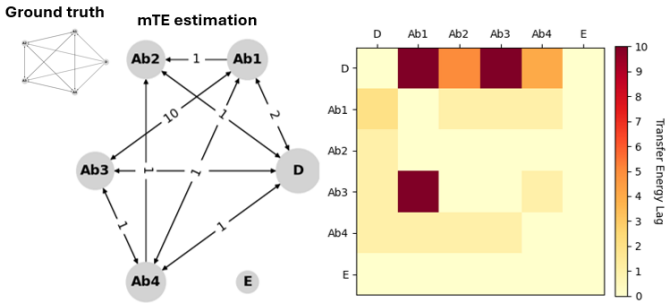


Fig. 9. Performance of Multivariate Transfer Energy (mTE) on the fetal ECG [35].The ground truth for this model is shown in the inset. The method recognizes valid causal relationships, particularly it leaves out the only node that does not have causal relationships with the others. The graph polynomial extracted from the estimated model ($p(x) = x^6 + 7x^4 - 10x^3 - 4x^2$) is different from the one extracted from the ground truth graph ($p(x) = x^5$). The estimated graph is also not isomorphic with the ground truth; its edit distance from it is 7.

physicist and mathematicians learning that the phase spaces of even very simple nonlinear dynamical systems can settle down on specific regions known as *strange attractors*, which implies that the long-term evolution of the system is difficult to predict. That is, two nearby points on the attractor can soon diverge and be mapped to larger distances. Because strange attractor are non-periodic; even when the evolution of the system is restricted to a relatively small region, the strange attractor's trajectory never closes on itself. Prototypical examples are the Lorenz and Rössler systems shown in Figure 3. Although still potentially problematic, we can still study the phase space of simple low-dimensional dynamic systems like a pendulum with direct experiments. But most dynamical systems of interest (e.g., weather, ecosystems) are high-dimensional. Direct phase space experimentation in such systems implies monitoring, perturbing or controlling a large set of not clearly defined variables. So how could we proceed? Here is where embedding theorems can help, we are specially interest in Taken's theorem [47]. In its original formulation and later extensions and generalizations [48]–[50], the theorem is described in detail. However, to fully appreciate these sources, prospective readers need a deep knowledge of differential topology. Here we only highlight the theorem's basic results and practical implications for the study of dynamical systems and causality inference.

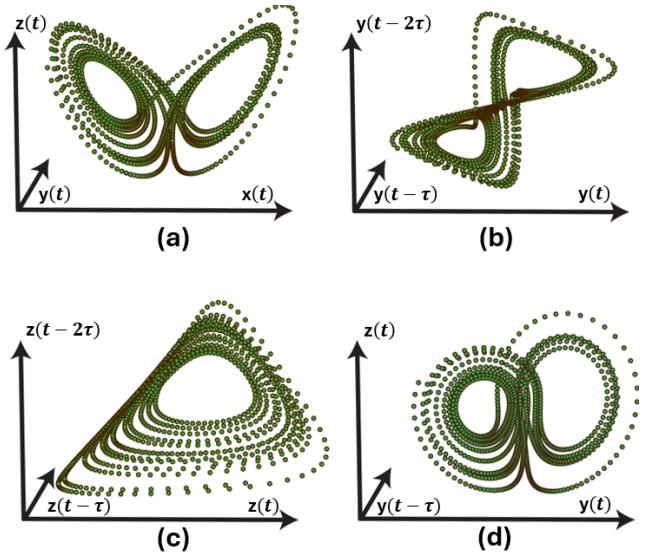


Fig. 10. Lorenz attractor with three shadow manifolds.(a) Lorenz attractor in its original phase space, (b) Univariate transformation using time lags in the y-coordinate ($\Phi = f(y(t), y(t-\tau), y(t-2\tau))$). Topologically equivalent to the original phase space, validating Takens' theorem. (c) Univariate transformation using time lags in the z-coordinate ($\Phi = f(z(t), z(t-\tau), z(t-2\tau))$). Not topologically equivalent to the original phase space, technically not a valid embedding. (d) Multivariate transformation using time lags in the y- and z-coordinates ($\Phi = f(y(t), y(t-\tau), z(t))$). Topologically equivalent to the original phase space but built following generalized versions of the Takens' theorem [50] instead of the original one [47] (images adapted from Deile and Sigihara [50]).

Takens' embedding theorem bridges the gap between theory and experiments specially in high dimensional dynamical system. For instance, if we want to study the population of a rabbits (x), we might initially write a model like the classic Lotka-Volterra [51], [52] to account for the presence of their natural predators in that ecosystem, for instance, foxes(y):

$$dx/dt = ax - \beta xy$$
$$dy/dt = -\gamma y + \delta xy$$

But what if we want to consider all the possible factors that might affect the population of rabbits? Then we are confronted with a system that has many more variables, some that we might be abe to measure (e.g., temperature) and other variables that we cannot measure or that we are not sure if they have an effect on the population of rabbits ($x = f(y, z, ...)$). How can we access the phase space of such a system? Takens' theorem provides a partial solution. It states that if a variable x has enough information about its own system Φ then it is possible to create a new system Φ' using time delay embedding of x, such as Φ and Φ' are *topologically equivalent* (e.g., like a doughnut and a mug). The general form of Φ' is then:

$$\Phi'(x) = (\alpha(x), \alpha(f(x)), ..., \alpha(f^{k-1}(x)))$$

Where α is some generic observation function of x, and k is the embedded dimension of the reconstructed attractor (which is different from the actual dimension of the original and maybe unknown dynamical system). In the related literature, sometimes the attractor in the original dynamical system is referred to as the manifold $\mathcal{M}$ and a reconstructed one using Takens' theorem is called a shadow manifold $\mathcal{M}'$. We illustrate these ideas in Figure 10 that shows the the original manifold for the Lorenz system and the reconstruction of three shadow manifolds. This result has remarkable practical applications. Coming back to our example it implies that we might be able study some features of a high-dimensional complex system (e.g., an ecosystem) by measuring only a single relevant variable (e.g., the number of rabbits).

Of greater interest for this work are the practical implications of these theoretical results for the study of causality. Specifically, here we briefly describe the main ideas behind *Convergent Cross Mapping* (CCM), a method of causal inference inspired by the relationships between original and shadow manifolds [23]. The basic ideas are relatively simple. First, two variables X and Y are causally related if they belong to the same dynamical system, which means that they share the same manifold $\mathcal{M}$. Therefore, we can recover the state of one variable, by looking at the past values of the other (e.g., the current population of foxes can be recovered from the past values of the Rabbit's population). Additionally, we should be able to quantify the extend to which X is causing Y by measuring the reliability of predicting Y using past values of X. CCM performs such measures on the original and shadow manifolds. For instance, if X and Y share manifold $\mathcal{M}$ and using Takens' theorem we create shadow manifolds $\mathcal{M}_X$ and $\mathcal{M}_Y$ for X and Y respectively, then a collection of nearby points in $\mathcal{M}_X$ should mapped to a similar collection of points on $\mathcal{M}$ and the same can be said about variable Y. Therefore $\mathcal{M}_X$ and $\mathcal{M}_Y$ are in a way "connected" through $\mathcal{M}$ and indirectly, a collection of nearby points in $\mathcal{M}_X$ should mapped to a similarly collection of collection of nearby points in $\mathcal{M}_Y$. This would not be true if the variables X and Y are not causally related (i.e., don't belong to the same dynamical system with manifold $\mathcal{M}$). One last noticeable feature of CCM is its convergence property. This means that the causal estimations improve with the length of the time series X and Y; the reason is that longer time series will produce more dense attractors and therefore more consistency when selecting a collection of nearby points in the manifolds. The drawback is that the convergence property increases the time to obtain accurate calculation. On the other hand, convergence is a property that discriminates causation from simple correlations. Additional details of the method can be found in the supplementary material of the original paper and a related implementation [23], [53].

2) Experiments: In this section we present the experimental results of applying Convergent Cross Mapping (CCM) to several datasets. First, in Figure 11, we show the results on the coupled Rössler-Lorenz attractor [24]. (a) represents the minimal ground truth for the causality graph as we know from the equations of the coupled Rössler- Lorenz system (see section II-B). (b) shows the results of applying CCM to this dataset. Overall the results for this dataset are similar than those obtained with causal calculus (Figure 4). Justifiably, CCM shows a more complex pattern of causal interactions than the ground truth; the variables from the Rössler attractor (x1,x2,x3) now seem to have a stronger driving (causal) influence over the Lorenz attractor that among themselves due to the relatively high value for the coupling constant C when generating the causal dataset for the coupled attractors (see equation 5).

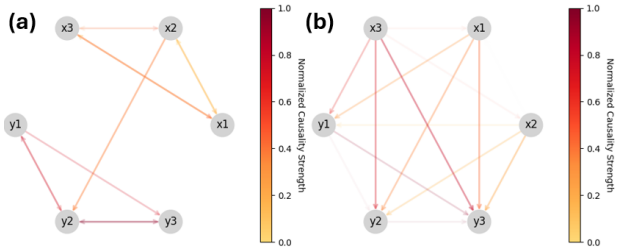


Fig. 11. Performance of Convergent Cross Mapping (CCM) on the coupled Rossler-Lorenz attractor [24].(a) Ground truth (b) CCM estimation. Besides finding all the causal relationships from the ground model, CCM also unveils other indirect causal relationships in the coupled attractor. The graph polynomial extracted from the estimated model and the ground truth are the same ($p(x) = x^4$), so the models share some core invariants. However, the estimated graph is not isomorphic with the ground truth; their edit distance from it is 13.

In Figure 12 we show the use of CCM on the BioGeoScience generated dataset (see Appendix A). Figure 5 (a) shows the ground truth for this system built from the generating equations and (b) shows the graph discovered using causal calculus. The results from CCM here are essentially identical to those obtained with causal calculus 5.

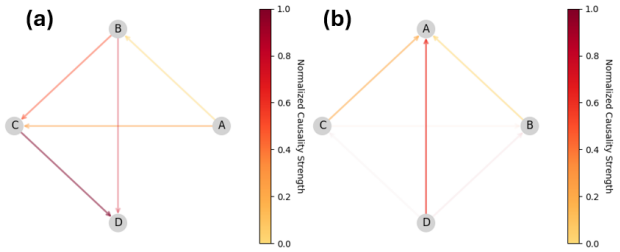


Fig. 12. Performance of Convergent Cross Mapping (CCM) on the BioGeoScience dataset [30].(a) Ground truth (b) CCM estimation. CCM recognized valid causal relationships; the graph polynomial extracted from the estimated model and the ground truth are the same ($p(x) = x^4$). Although close to be, the estimated graph is not isomorphic with the ground truth; its edit distance from it is 1.

In Figure 13 we show the results of applying CCM to the experimental fetal ECG dataset [35]. This is a multichannel ECG recording with one channel (D) coming directly from the baby's head at labor and the other channels ($Ab1, ...Ab4$) are

indirect measures taken from the abdomen of the mother. An extra channel we certainly know is not functionally related in any way to the measurements is channel E (annotations). The results are similar to those using causal calculus 6. However, those results, CCM failed to exclude channel E as causally unrelated.

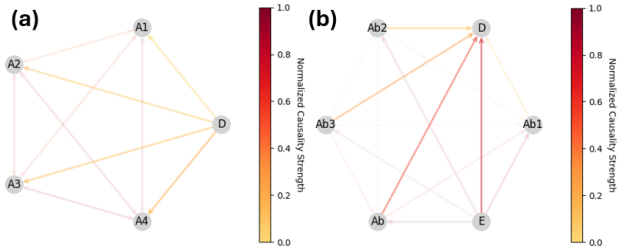


Fig. 13. Performance of Convergent Cross Mapping (CCM) on the fetal ECG [35].(a) Ground truth bas build based on several observations. Node *E* are non-causally related annotations. Node *D* is measured on the baby's head directly. The other measures are performed on the mother's abdomen (b) CCM estimation based on recorded data. Although CCM fails to exclude *E* from the graph, it does seems to recognize valid causal relationships. The graph polynomial extracted from the estimated model ($p(x) = x^6$) is different from the one extracted from the ground truth graph ($p(x) = x^5$). The estimated graph is also not isomorphic with the ground truth; its edit distance from it is 12.

D. Calculations on the target datasets.

In this section we present results for the available target datasets. Specifically, (1) a dataset containing operating data from a smart grid [29], and (2) the Summit dataset [28], containing operating GPU data from the Summit Supercomputer. The target datasets were chosen due to their direct relationships to data centers and also because of their challenging characteristics. Specifically, our initial ideas were related to the study of data center reliability as in the datasets there is information about GPU failure. In Figure 16 we show causality inference results on the Summit dataset using Causal Calculus and Convergent Cross Mapping (CCM). Interestingly enough both methods diverge in their results and seem to uncover different perspectives on how to see these data. Causal calculus (Figure 16 (a)) seems to indicate that the failures have a causal influence on all the other features except power. On the other hand, CCM (Figure 16 (b)) seem to indicate that all the features in the dataset have an impact on GPU usage.

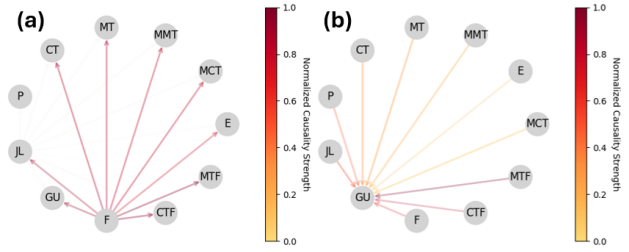


Fig. 14. Causality inference of the Summit target dataset. [28] using:(a) Causal calculus, and (b) Convergent Cross Mapping (CCM).

In Figure 15 we show causality inference results on the Smart Grid dataset using Causal Calculus and Convergent Cross Mapping (CCM). Also in this case the methods diverge in their results pointing to different perspectives on how to see these data. Causal calculus (Figure 15 (a)) seems to correctly indicate a causal relationship between the reactive power (rP) and the current (A) and yet it seems to miss the causal relationship with the voltage (assuming that they are measurung reactive power as $rP = V \times A \times sin(\phi)$, where ϕ is the phase between V and A). Causal calculus also unveils related and yet probably spurious links between reactive power (rP) and power (P) and some intesting connection with the predictive load (Pl) which, if true could be very useful predicting the computing load with few features. On the other hand, CCM (Figure 15 (b)) seem to indicate that that the most features causally contribute to few variables in the dataset such as current (A), wind power (wP) and solar power (sP). Although interesting and potentially useful to create energy efficient prediction models for these variables one wonders if a higher threshold for causality detection could make these results clearer by eliminated potentially spurious causal relationships. It is also worth noting that although the CCM results are shown in a graph, the calculations were made between each pair of values independently. To the best of our knowledge, there is no multivariate versions of CCM or even other similar methods based on dynamical systems, which should introduce errors when estimating these network-like causal relationships.

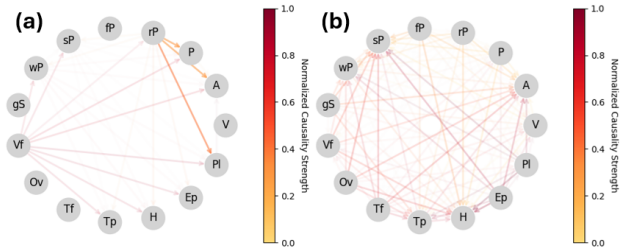


Fig. 15. Causality inference of the smart grid target dataset. [29] using:(a) Causal calculus, and (b) Convergent Cross Mapping (CCM).

Finally, we would like to point out that the correlation heatmaps of the features in both, the Summit and the smart grid datasets fail to offer any hints regarding our causality results. This intuitively confirms that correlation does not indicate causation, or in this case, that lack of correlation does not preclude causation. Unfortunately, multivariate transfer entropy (mTE) did not produced significant results for any of these datasets. This could mean that the causal relationships are spurious, but it can also be affected by the length and quality of the data. For instance, we only have 127 time series entries for every feature, which is not ideal. This is also known to affect CCMas we already mentions that this method (and other similar ones based on dynamical system theory) are affected by the length of the data. For instance, with longer time series, we should have better odds of convergence and therefore more reliable CCM results.

Finally, in table II we resumed the success (■) or failure (□) of using the causal inference methods to generate causal graphs or DAGs for selected datasets. By failure here we

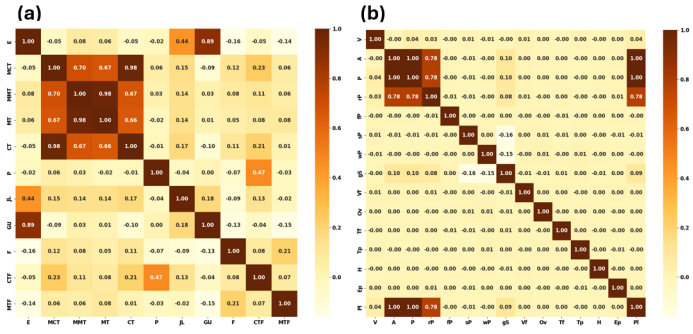


Fig. 16. Correlation Heatmaps for the target datasets.(a) Summit [28] (b) Smart Grid [29]

mean that, even though the implementation of the method ran without error, it was not able to produce results of significant value. That was the case of multivariate transfer entropy (mTE) on both target datasets, but more surprisingly, for a generated dataset (Driven attractors). Although in the case of the Summit dataset we could assume that the cause of the failure was insufficient data length (only 127 records) for the generated dataset we haven't been able to find a reasonable explanation.

TABLE II
CAUSAL GRAPH ESTIMATION ON SELECTED DATASETS.[1]

	Causal Calc.	mTE	CCM
Fetal ECG	■	■	■
BioGeoScience	■	■	■
Driven Attractors	■	□	■
Smart Grid	■	□	■
Summit	■	□	■

[1] Other results are shown in the Appendix.

IV. DISCUSSION

On the practical use within DTs A robust Digital Twin augmented with a causal analyzer can not only simulate and monitor data center operations but also uncover the underlying causes of failures and optimize for performance. By integrating causal metrics like the Average Treatment Effect (ATE), our Digital Twin offers actionable insights to enhance failure prediction models and optimize data center reliability. For instance, if we know that fluctuations in CPU temperature may have a greater effect on failures than simply keeping the temperature below a certain threshold, this insight can guide proactive strategies to prevent future failures.

There are many opportunities in using causal analysis as the engine of a digital twin for data center maintenance and optimization. Key applications include root cause analysis to identify the true sources of failures, workload scheduling optimization to improve performance and energy efficiency, and preventing performance degradation by pinpointing the causes of temperature fluctuations. Additionally, anomaly detection can be refined using causal inference to reduce false positives and missed failures, while energy efficiency and power optimization benefit from identifying the true drivers of power surges. Finally, digital twin-based what-if analysis allows operators to simulate interventions before implementation, minimizing risks and ensuring effective decision-making.

On the methods

In this work, we explore methods that are mostly related to causality inference regarded as a data driven approach to estimate the presence and strength of causal links between variables. A related and yet conceptually different approach uncovered here is the area of equation discovery, technically known as symbolic regression (see Radwan et al. [54]). Although, as we reported, we can get a glimpse of these ideas by reporting the invariant polynomials for the estimated DAGs; symbolic repression will bring insights and understanding to new qualitative levels.

Regarding dynamical system theory, in this work we limited ourselves to work with the original Convergent Cross Mapping (CCM) method [23]. However, other extensions for this method has been proposed that deal with spatiotemporal data [55] and/or correct some of problems in the original method. We are currently testing some of these methods, specially CCM implemented with pairwise asymmetric inference (PAI) [53], [56] and Tangent Space Causal Inference (TSCI) [57], [58]. These variants will be included in future versions of this work.

One observation that comes from our own experiments here and seems to be confirmed by our admittedly informal literature search, is the scarcity of *proper* multivariate methods and/or implementations for causal discovery. By proper here we mean mathematically derived multivariate versions of the method in full consistency with the original theories in which said methods were initially introduced. This has clearly an impact when creating DAGs or causal graphs as the effects of non-immediate causal links can be ignored resulting in the visualization of spurious causal links and probably the dismissal of relevant ones. In the case of CCM, the authors did suggested ways to deal with confounders variables. Yet, an integrally derived multivariate implementation was not offered then and, to the best of our knowledge, has not been offered by other researchers in this area ever since.

Data-driven approach underpinned by machine learning based techniques such as Bayesian networks, neural networks can significantly enhance causal inference. These methods enable the automation of causal relationships discovery [59], facilitate the determination of the structure of DAGs and support counterfactual predictions, which are crucial for determining causal effects. We intend to explore and expand upon these in our future works.

V. CONCLUSIONS

Future data centers are bound to be tightly integrated with smart grids for their mutual benefit. Managing such large infrastructures will require a combination of both: advance analytics and visualizations. Digital twins strikes like a perfect vehicle for delivering such services. However, even the most advanced analytics and AI methods today fall short in providing real causal understanding of the data which degrades our

trust in their decision making capabilities. This work marks our first step in evaluating and eventually providing causal inference capabilities to digital twins. Specifically here we: (1) Curated and tested datasets for causality inference, which is an important requirement given the scarcity of those datasets. (2) Using both generated and experimental datasets with known ground truth, we Validated three types of causality inference methods respectively based on causal calculus, information theory and dynamical system theory, and (3) We applied the lessons learned during method validation to two target datasets containing information about GPU data centers and smart grids. Causal insights were clearly beyond standard analytics showing interesting causal relationships even in the absence of data correlations. The gathered causal insights suggest new ways to performs ever more meaningful and energy-efficient predictions supportixng future digital twin solutions.

VI. Acknowledgement

This research was sponsored by and used resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility at the Oak Ridge National Laboratory (ORNL) supported by the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Appendix A
Supplementary Materials

A. Data Generation

BioGeoScience model. One of the strategies for causal data generation used in this work implements BioGeoscience model that studies the causal effect of air temperature on ecosystem respiration. The system of equations for this model is:

$$\begin{aligned} A_t &= |280sin(t\pi/365)^2 + 50|sin(t\pi/365)|\eta_t^A| \\ B_t &= 0.8B_{t-1} + 0.02A_t + 5\eta_t^B \\ C_t &= 0.2C_{t-1} + 0.002A_tB_t + 3\eta_t^C \\ D_t &= 0.3D_{t-1} + 0.9C_t0.8^{0.12(B_t-15)} + 2\eta_t^D \end{aligned}$$

Where: A_t is the shortwave radiation, B_t is the air temperature. C_t is the gross primary production, D_t is the ecosystem respiration, and η_t are noise terms. The ground truths for this model are experimentally well known. A snapshot of the data can be seen in Figure S1. Additional domain knowledge implications of the model can be found in the reference work by Runge et al. [30]

Health-BMI dataset. Of the handful of available libraries to generate causal Directed Acyclic Graphs (DAGs) we choose simDAG [31] due to its ability to generate arbitrarily complex causal models with bespoke statistical properties. In this work we used the intuitive Health-BMI causal model they report in their documentation. The DAG and structural generating equations for this model Figure S2.

In this model, α, β, γ, and δ are adjustable parameters and η is a noise factor. Datasets to explore the strength of causal effects can therefore be produced by fine-running the variables and/or parameters of this model. For instance, generating different data files using fixed settings for all the parameters except for δ that can be allowed to take equally spaced values in the interval $[0, 1]$, can shed light on the causal effect $bmi \Rightarrow death$.

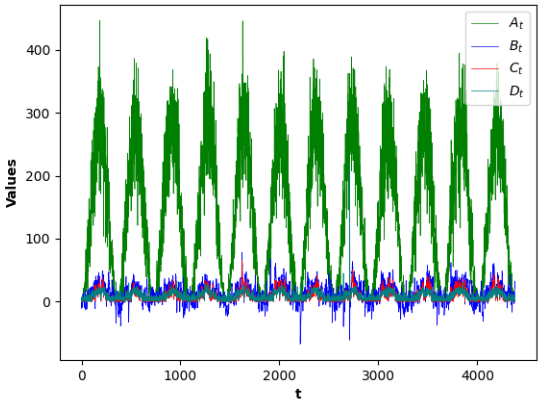
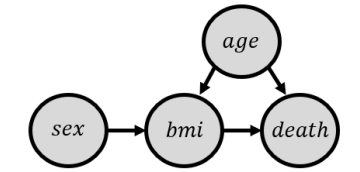


Fig. S1. Time series causal data generated using the BioGeoScience model.



$$bmi = \alpha_1 + \beta_1 \times sex + \gamma_1 \times age + \eta$$
$$p(death) = \alpha_2 + \gamma_2 \times age + \delta \times bmi$$
$$with\ p(x) = \frac{1}{1 + e^{-x}}$$

Fig. S2. Health-BMI DAG with its generating structural equations.

B. Essential quantities from information theory.

Entropy. Expected value of the amount of information needed to predict the next state x_{n+1} of a process:

$$H_X = \langle -\log_2 p(x_{n+1}) \rangle_n$$

Joint entropy. For variables X and Y, is the uncertainty of their joint distributions:

$$H_{X,Y} = \langle \log_2 p(x_n, y_n) \rangle_n$$

Conditional entropy. The conditional entropy for a variable X given Y is the average amount of uncertainty that remains of the state x_n once we know the state y_n:

$$H_{X|Y} = \langle \log_2 p(x_n|y_n) \rangle_n$$

Mutual information. The mutual information for variables X and Y is the average reduction in the uncertainty of the state

x_n once we know the state y_n or vice versa (this measure is symmetric respect to X and Y).

$$I_{X,Y} = H_X - H_{X|Y} = \langle \log_2 p(x_n|y_n)/p(x_n) \rangle_n$$
$$I_{Y,X} = H_Y - H_{Y|X} = \langle \log_2 p(y_n|x_n)/p(y_n) \rangle_n$$
$$I_{X,Y} = I_{YX}$$

Timed-lagged mutual information. With a small change in the definition of mutual information we can also measure this quantity for a state lagged a time step tf. Unlike mutual information, lagged mutual information is not symmetric and therefore can (in theory) be considered as a measure for causality. Furthermore, by incorporating time (t), this metric implicitly accounts for some dynamicity:

$$I_{X,Y}(t) = H_X(t) - H_{X|Y}(t) = \langle \log_2 p(x_{n+t}|y_n)/p(x_{n+1}) \rangle_n$$
$$I_{Y,X}(t) = H_Y(t) - H_{Y|X}(t) = \langle \log_2 p(y_{n+1}|x_n)/p(y_{n+1}) \rangle_n$$
$$I_{X,Y}(t) \neq I_{YX}(t)$$

C. Extra figures

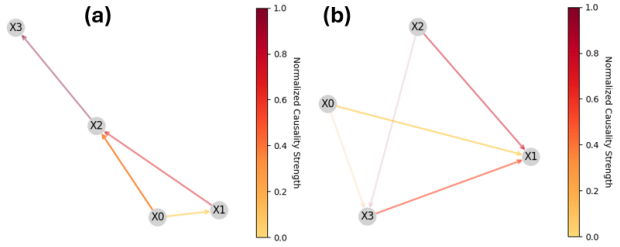


Fig. S3. Performance of causal calculus on the Simpson paradox model from from the Csuite [21], [22].(a) Ground truth, (b) Generated data. Both results show the same polynomial graphs ($p = x^4$), accounting for the preservation of some core graph invariants. However, the two graphs are not isomorphic. Their graph edit distance is 3.

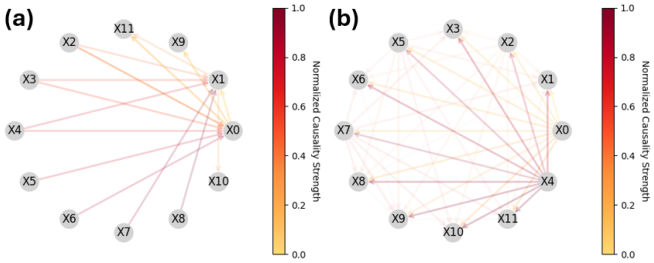


Fig. S4. Performance of causal calculus on the mixed confounding model from from the Csuite [21], [22].(a) Ground truth, (b) Generated data. Both results show the same polynomial graphs ($p(x) = x^{12}$), accounting for the preservation of some core graph invariants. However, the two graphs are not isomorphic. Their graph edit distance is 37.

REFERENCES

[1] P. Taylor, "Data centers - statistics and facts," 2025, https://www.statista.com/topics/6165/data-centers/ [Accessed: 30/03/2025].

[2] V. Fogarty and S. Flucker, "What drives the need and the various types of data centres," in *Data Centre Essentials: Design, Construction, and Operation of Data Centres for the Non-expert*. Wiley-Blackwell, 2023, pp. 5–22.

[3] M. A. B. Siddik, A. Shehabi, and L. Marston, "The environmental footprint of data centers in the united states," *Environmental Research Letters*, vol. 16, no. 6, p. 064017, may 2021. [Online]. Available: https://dx.doi.org/10.1088/1748-9326/abfba1

[4] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data center: a survey on hardware technologies," *Cluster Computing*, vol. 25, no. 1, pp. 675–705, Feb 2022. [Online]. Available: https://doi.org/10.1007/s10586-021-03431-z

[5] H. Chen, M. C. Caramanis, and A. K. Coskun, "Reducing the data center electricity costs through participation in smart grid programs," in *International Green Computing Conference*, 2014, pp. 1–10.

[6] S. M. Ali, I. Hussain, Z. Ullah, I. Sami, R. Asghar, U. Farid, B. Khan, C. A. Mehmood, and A. Haider, "Need for mutual services interaction between smart grid and cloud data centers," in *2018 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET)*, 2018, pp. 1–5.

[7] S. Tayeb, M. Mirnabibaboli, L. Chato, and S. Latifi, "Minimizing energy consumption of smart grid data centers using cloud computing," in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, 2017, pp. 1–5.

[8] Z. Ullah, F. Mehmood, S. M. Ali, T. Mujahid, I. Khan, B. Khan, and M. Jawad, "Smart grid (sg) and data center (dc) integration: A new conceptual framework," in *2019 15th International Conference on Emerging Technologies (ICET)*, 2019, pp. 1–6.

[9] S. Jeong and S. Shin, "Efficient network administration for smart grid data center," in *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, 2020, pp. 48–51.

[10] Z. Zhang, Y. Zeng, H. Liu, C. Zhao, F. Wang, and Y. Chen, "Smart dc: An ai and digital twin-based energy-saving solution for data centers," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, 2022, pp. 1–6.

[11] D. Huang, "Data center infrastructure management." in *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021, pp. 627–644.

[12] N. Shaik, B. V. Kumar, A. Bhati, S. Kumar, J. Singh, and D. P. Sahu, "Integrating random forest and support vector regression models for optimized energy consumption evaluation in cloud computing data centers," in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2023, pp. 451–456.

[13] R. Dharaniya, S. V S, S. B. N, and S. A. Zaid, "Ai agents at different data centers to minimize the energy spending," in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 2023, pp. 813–818.

[14] Y. Zheng, L. Jiao, Y. Xu, B. An, X. Wang, and Z. Li, "Scheduling generative-ai job dags with model serving in data centers," in *2024 IEEE/ACM 32nd International Symposium on Quality of Service (IWQoS)*, 2024, pp. 1–6.

[15] N. Hogade and S. Pasricha, "Game-theoretic deep reinforcement learning to minimize carbon emissions and energy costs for ai inference workloads in geo-distributed data centers," *IEEE Transactions on Sustainable Computing*, pp. 1–14, 2024.

[16] Y. Zhang, F. Hu, Y. Han, W. Meng, Z. Guo, and C. Li, "Ai-based energy-saving for fog computing-empowered data centers," in *2023 International Conference on Mobile Internet, Cloud Computing and Information Security (MICCIS)*, 2023, pp. 16–21.

[17] M. G. Brahmam and R. Vijay Anand, "An investigation of consolidating virtual servers and data centers based on energy consumptions using various algorithms," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, 2023, pp. 1521–1528.

[18] T. Eda, M. Busto, T. Udagawa, N. Ishihama, K. Tabata, Y. Matsuo, and I. Yamasaki, "Technical challenges for ai in space data centers," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 1733–1735.

[19] J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf, "Distinguishing cause from effect using observational data: Methods and benchmarks," *Journal of Machine Learning Research*, vol. 17, no. 32, pp. 1–102, 2016. [Online]. Available: http://jmlr.org/papers/v17/14-518.html

[20] Mooij, Peters, Janzing, Zscheischler, and Schölkopf, "Database with cause-effect pairs," 2016, https://webdav.tuebingen.mpg.de/cause-effect/ [Accessed: 21/03/2025].

[21] T. Geffner, J. Antoran, A. Foster, W. Gong, C. Ma, E. Kiciman, A. Sharma, A. Lamb, M. Kukla, N. Pawlowski, M. Allamanis, and C. Zhang, "Deep end-to-end causal inference," 2022. [Online]. Available: https://arxiv.org/abs/2202.02195

[22] Geffner, Antoran, Foster, Gong, Ma, Kiciman, Sharma, Lamb, Kukla, Pawlowski, Allamanis, and Zhang, "Csuite: A suite of benchmark datasets for causality," 2022, https://github.com/microsoft/csuite [Accessed: 22/03/2025].

[23] G. Sugihara, R. May, H. Ye, C. hao Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1227079

[24] M. L. V. Quyen, J. Martinerie, C. Adam, and F. J. Varela, "Nonlinear analyses of interictal eeg map the brain interdependences in human focal epilepsy," *Physica D: Nonlinear Phenomena*, vol. 127, no. 3, pp. 250–266, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278998002589

[25] O. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0375960176901018

[26] O. Rossler, "An equation for hyperchaos," *Physics Letters A*, vol. 71, no. 2, pp. 155–157, 1979. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0375960179901506

[27] E. Lorenz, "The statistical prediction of solutions of dynamic equations," *Proceedings of the International Symposium on Numerical Weather Prediction in Tokyo*, pp. 629–635, 1960. [Online]. Available: https://web.archive.org/web/20190523190103/http://eaps4.mit.edu/research/Lorenz/The_Statistical_Prediction_of_Solutions_1962.pdf

[28] W. Shin, V. Oles, A. Schmedding, G. Ostrouchov, E. Smirni, C. Engelmann, and F. Wang, "Olcf summit supercomputer gpu snapshots during double-bit errors and normal operations," Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States). Oak Ridge ..., Tech. Rep., 2023.

[29] Public-Domain, "Smart grid real-time load monitoring dataset," 2025, https://www.kaggle.com/datasets/ziya07/smart-grid-real-time-load-monitoring-dataset [Accessed: 28/03/2025].

[30] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls, "Causal inference for time series," *Nature Reviews Earth & Environment*, vol. 4, no. 7, pp. 487–505, Jul 2023. [Online]. Available: https://doi.org/10.1038/s43017-023-00431-y

[31] R. Denz and K. Meiszl, "simdag: An r package to simulate simple and complex (longitudinal) data from a dag and associated node information," 2023. [Online]. Available: https://github.com/RobinDenz1/simDAG

[32] E. Altamimi, A. Al-Ali, Q. M. Malluhi, and A. K. Al-Ali, "Smart grid public datasets: Characteristics and associated applications," *IET Smart Grid*, vol. 7, no. 5, pp. 503–530, 2024. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/stg2.12161

[33] J. Pearl, *Causality*, 2nd ed. Cambridge University Press, 2009.

[34] A. Kelleher, S. Fuhrmann, and J. Attenberg, "Causality: Tools for causal analysis using observational (rather than experimental) datasets." 2024, https://github.com/akelleh/causality/ [Accessed: 03/04/2025].

[35] J. J, M. A, K. T, R. D, and C. R, "Abdominal & direct fetal ecg database(csv format)," 2022. [Online]. Available: https://www.kaggle.com/dsv/4776543

[36] V. Mandrekar, "The work of wiener and masani on prediction theory and harmonic analysis," in *Connected at Infinity II: A Selection of Mathematics by Indians*, R. Bhatia, C. S. Rajan, and A. I. Singh, Eds. Gurgaon: Hindustan Book Agency, 2013, pp. 173–184. [Online]. Available: https://doi.org/10.1007/978-93-86279-56-9_8

[37] N. Wiener, "Theory of predictions," in *Modern Mathematics for the Engineer: First Series.*, ser. Dover books on engineering. Dover Publications, Incorporated, 2013. [Online]. Available: https://books.google.co.uk/books?id=WY8ihfzpimMC

[38] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, Jul 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.85.461

[39] C. E. Shannon, N. J. A. Sloane, and A. D. Wyner, *Claude Elwood Shannon: collected papers*. IEEE Press, 1993.

[40] J. T. Lizier and M. Rubinov, "Multivariate construction of effective computational networks from observational data," 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:15897400

[41] P. Wollstadt, J. T. Lizier, R. Vicente, C. Finn, M. Martinez-Zarzuela, P. Mediano, L. Novelli, and M. Wibral, "Information dynamics toolkit xl (idtxl): a comprehensive software package for efficient inference of networks and their node dynamics from multivariate time series data using information theory," *Journal of Open Source Software*, vol. 4, no. 34, p. 1081, 2019. [Online]. Available: https://doi.org/10.21105/joss.01081

[42] ——, "Database with cause-effect pairs," 2019, https://github.com/pwollstadt/IDTxl [Accessed: 31/03/2025].

[43] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy–a model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, p. 45–67, Feb. 2011. [Online]. Available: https://doi.org/10.1007/s10827-010-0262-3

[44] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, vol. 2, no. 42, pp. 230–265, 1936. [Online]. Available: http://www.cs.helsinki.fi/u/gionis/cc05/OnComputableNumbers.pdf

[45] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, "Local measures of information storage in complex distributed computation," *Information Sciences*, vol. 208, pp. 39–54, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025512002800

[46] M. Wibral, J. Lizier, S. Vögler, V. Priesemann, and R. Galuske, "Local active information storage as a tool to understand distributed neural information processing," *Frontiers in Neuroinformatics*, vol. 8, 2014. [Online]. Available: https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2014.00001

[47] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381.

[48] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3, pp. 579–616, Nov 1991. [Online]. Available: https://doi.org/10.1007/BF01053745

[49] J. Stark, D. Broomhead, M. Davies, and J. Huke, "Takens embedding theorems for forced and stochastic systems," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 30, no. 8, pp. 5303–5314, 1997, proceedings of the Second World Congress of Nonlinear Analysts. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0362546X96001496

[50] E. R. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *PLOS ONE*, vol. 6, no. 3, pp. 1–8, 03 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0018295

[51] A. J. Lotka, "Contribution to the theory of periodic reactions," *The Journal of Physical Chemistry*, vol. 14, no. 3, pp. 271–274, Mar 1910. [Online]. Available: https://doi.org/10.1021/j150111a004

[52] V. Volterra, *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. Società anonima tipografica" Leonardo da Vinci", 1926.

[53] P. J. E. Javier, "causal-ccm a Python implementation of Convergent Cross Mapping," 6 2021.

[54] Y. A. Radwan, G. Kronberger, and S. Winkler, "A comparison of recent algorithms for symbolic regression to genetic programming," 2024. [Online]. Available: https://arxiv.org/abs/2406.03585

[55] B. Gao, J. Yang, Z. Chen, G. Sugihara, M. Li, A. Stein, M.-P. Kwan, and J. Wang, "Causal inference from cross-sectional earth system data with geographical convergent cross mapping," *Nature Communications*, vol. 14, no. 1, p. 5875, Sep 2023. [Online]. Available: https://doi.org/10.1038/s41467-023-41619-6

[56] J. M. McCracken and R. S. Weigel, "Convergent cross-mapping and pairwise asymmetric inference," *Physical Review E*, vol. 90, no. 6, p. 062903, 2014.

[57] K. Butler, D. Waxman, and P. M. Djurić, "Tangent space causal inference: Leveraging vector fields for causal discovery in dynamical systems," in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, Eds., vol. 37. Curran Associates, Inc., 2024, pp. 120 078–120 102. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/file/d9253fba38ed8a140f86fa22d89344ec-Paper-Conference.pdf

[58] D. W. Kurt Butler, "Tangent space causal inference: Leveraging vector fields for causal discovery in dynamical systems," 2024, https://github.com/KurtButler/tangentspaces [Accessed: 03/04/2025].

[59] B. Schölkopf, "Causality for machine learning," in *Probabilistic and causal inference: The works of Judea Pearl*, 2022, pp. 765–804.