INL/CON-24-76171-Revision-0

VALIDATION, VERIFICATION, AND CALIBRATION THROUGH A CAUSAL LENS

February 2024

Changing the World's Energy Future

Ronald Louis Gonzales, Diego Mandelli, Congjian Wang, Mohammad Gamal M Mostafa Abdo, Paolo Balestra, Sunming Qin, Victor Petov, Zachary Welker, Annalisa Manera

INL/CON-24-76171-Revision-0

VALIDATION, VERIFICATION, AND CALIBRATION THROUGH A CAUSAL LENS

Ronald Louis Gonzales, Diego Mandelli, Congjian Wang, Mohammad Gamal M Mostafa Abdo, Paolo Balestra, Sunming Qin, Victor Petov, Zachary Welker, Annalisa Manera

February 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Proceedings of the ASME 2024
Verification, Validation, and Uncertainty Quantification Symposium
VVUQ2024
May 15-17, 2024, College Station, Texas, USA

VVUQ2024-132343

VERIFICATION, VALIDATION, AND CALIBRATION THROUGH A CAUSAL LENS

Ronald Gonzales[1,†,*], Diego Mandelli[1,†], Congjian Wang[1], Mohammad Abdo[1], Paolo Balestra[1], Sunming Qin[1], Zachary Welker[2], Victor Petrov[2], Annalisa Manera[2],

[1]Idaho National Laboratory, Idaho Falls, ID
[2]University of Michigan, Ann Arbor, MI

ABSTRACT

While typical validation and verification approaches focus on identifying the associations between data elements using statistical and machine learning methods, the novel methods in this paper focus instead on identifying causal relationships between data elements. Statistical and machine-learning-based approaches are strictly data-driven, meaning that they provide quantitative comparison measures between data sets without explicitly considering the hypotheses behind them. This can lead to the erroneous conclusion that, if two data sets are close enough, the models that generated them are similar. In addition, when experimental and simulated data differ to an extent that fails to meet the acceptance criteria, calibration techniques are used to tweak simulation model parameters to reduce the gap between the two types of data. This produces the false expectation that a simulation model will match reality. The methods presented in this paper move away from these strictly data-driven methods for validation and calibration toward more robust, model-driven methods based on causal inference. Causal inference aims to identify the possible mechanisms that might have generated data. Thus, this analysis targets the prediction of the effects when one (or more) of the identified mechanisms are altered. There are many approaches to identify, quantify, and illustrate causal relationships. For the scope of this paper, directed graphs are employed as causal models. If the directed graph lacks cycles, it is known as a directed acyclic graph. A node in such a graph represents an observed data element while a directed edge connecting two nodes represents a causal relationship between two variables. The developed causal methods are designed to extract causal models from simulation models and experimental data. Causal models capture the causal relationships between data elements (e.g., simulated and experimental data). In this context, validation and verification are performed by comparing causal models. The proposed approach does not only inform system analysts on how a simulation model matches real-world data, but also identifies elements of the simulation model that should be revised when discrepancies between simulation and experimental data are observed. Through these causal methods, analysts can identify the portion of the model equation(s) that are behind an edge connecting two variables. Hence, once the structural differences between causal models have been determined, model calibration can occur by changing only those model parameters that impact the identified causal relationships.

Keywords: Causality, Causal Discovery, Causal Inference, Validation, Verification, Calibration

1. INTRODUCTION

Determining whether a simulation model accurately represents the target system poses a significant challenge [1]. Presently, advanced simulation models serve not only in the system design and licensing stages but are also integral to system day-to-day operation and control, such as in autonomous systems operation or within a digital twin situation. These simulation models are carefully designed to closely replicate aspects of the real world to accurately capture the evolution and properties of those aspects. The development of these simulation models starts with defining their underlying laws, expressed through a set of mathematical equations (e.g., conservation laws, equations of motion, thermodynamic laws). At this stage, the causal relationships among variables within the model are temporarily lost due to the inherent nature of the mathematical formalism employed. These causal relationships are later regained when the developed mathematical equations are translated into a simulation model.

After the completion and verification of the coding process, these simulation models undergo a validation phase. This requires testing their ability to model real-world system aspects by comparing simulated results with real-world experimental observations under similar initial or boundary conditions. Current methods for model validation predominantly rely on standard statistical analysis (for measuring statistical differences between

†Joint first authors
*Corresponding author: ronald.gonzales@inl.gov

data populations) or machine learning techniques (for identifying response surfaces from data) [2]. Both approaches are purely data driven, providing quantitative measures for comparing data sets (e.g., simulated vs. measured data) without explicitly considering the hypotheses (e.g., boundary conditions) or the structure of the employed models. This may provide a misleading conclusion that, if two data populations are sufficiently close, the models generating them are similar. Additionally, when simulated and experimental data deviation are beyond acceptable limits, model calibration techniques come into play. These calibration techniques adjust simulation model parameters to narrow the gap between the two types of data, creating the false expectation that a simulation model mirrors reality.

The methods outlined in this paper depart from the current purely data-driven approaches. The presented techniques for validation, verification, and calibration move toward more robust model-driven methods based on causal inference. These methods focus on finding the causal relationships between data elements (e.g., simulated and experimental data) rather than solely examining their associations.

To identify these causal relationships, directed graphs and often directed acyclic graphs (DAGs) are employed [3]. Essentially, a DAG is a graph construct representing a measurable space for the considered data elements (treated as nodes in the graph). In a validation context, the DAGs obtained from simulated and experimental data are compared to quantify the differences between the two data sets. In other words, validation becomes model-based rather than relying solely on data. Once differences between DAGS are identified, model calibration takes place by adjusting only those model parameters that influence the established causal relationships. Once the causality relations are inferred, deviations in system dynamics (e.g., anomaly detection) can be attributed to fewer root causes. This enables the fusion of different data types because the common causality relations serve as the primary controller of the models, rather than relying solely on the data.

This paper now introduces the fundamentals of causal inference and causal discovery along with brief descriptions of some discovery methods. Following this, it presents an analytical comparison of the selected causal discovery methods on test problems. Next, this paper applies the methods to a real-world experiment. The subsequent sections provide the novel model-driven methods for validation, verification, and calibration of simulation models.

2. CAUSAL INFERENCE AND DISCOVERY

The concept of causation provides more information than simply a correlation between variables, it suggests a cause and effect relationship in which a change in one variable induces a change in another. Here, we follow the definition of causality provided by [4] which relies on the concept of intervention: the forced action on a particular system element (i.e., the cause) leads to an alteration of other system elements (i.e., the consequences). This is possible when "information" is carried through from the causes to the effects [5]. Here, the process of sharing information is considered in physical sense where energy, momentum, mass, or force are exchanged between elements of the system. Causal inference is an emergent field of study that aims to discover and make use of cause and effect relationships. The focus of this work is to discover the cause and effect relationships between variables and utilize this information for validation, verification, and calibration. This section provides a brief introduction to the field of causal inference and common terminology within it and briefly describes some methods to discover causal relationships from observational or experimental data.

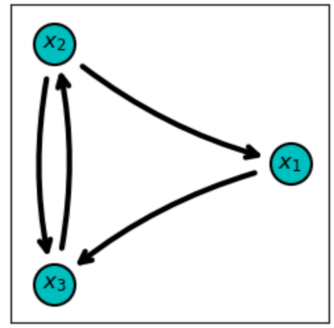

FIGURE 1: EXAMPLE OF A DIRECTED GRAPH DESCRIBED BY THREE VARIABLES (I.E., x_1, x_2, AND x_3) THAT CONTAINS CYCLES

There are many methods to identify, quantify, and illustrate causal relationships [6]. For the scope of this work, directed graphs are employed as causal models. A directed graph illustrates the existence of cause and effect relationships between a set of variables and the strengths of those relationships. A node in these graphs represents an observed variable while an edge connecting two nodes represents a causal relationship between the two variables. If it is known which variable is the cause and which is the effect, the edge is directed, represented as an arrow from the parent to the affected variable. As an example, assume that two variables x_1 and x_2 are linked by an edge, $x_2 \rightarrow x_1$; this implies that x_2 causes x_1 or, similarly, x_2 is a *parent* of x_1. A graph with multiple directed edges can contain cycles (see Figure 1). When dealing with physics-based data, mutual interactions between physical elements are common; these kinds of interactions are translated into cycles within the graph architecture. If a directed graph lacks cycles, it is known as a DAG.

The process of identifying and quantifying causal relationships from observational or experimental data is known as causal discovery. Many causal discovery methods are graph based; consider the following graph terminology. A pair of nodes is said to be *adjacent* if there is an edge between them in the graph, either directed or nondirected [3]. An *immorality* refers to a triple of nodes in a directed graph that have the structure $x_1 \rightarrow x_3 \leftarrow x_2$, where x_1 and x_2 are not adjacent. *Skeleton* graphs contain only the edges with causal dependency between the nodes, but direction has yet to be determined. Some discovery methods are unable to direct all edges in a graph but can discover the Markov equivalence class of that DAG. Two DAGs are *Markov equivalent* if and only if they have the same skeleton and immoralities [3].

A fundamental assumption in causal discovery is that all variables that can influence a system have been observed. That is, a set of variables X is said to be *causally sufficient* if $\mathrm{pa}(x) \subseteq X$ for all x in X, where $\mathrm{pa}(x)$ is the set of parents of x. There are discovery methods that work to drop causal sufficiency [7–9];

however, in the context of this work, it is assumed all influential variables are observed. If the causal relationships and noise distributions are invariant with respect to time, the causal model is said to be *causally stationary* [10]. This work assumes that there are no contemporaneous causal effects (i.e., causes occur strictly before effects). There is work addressing cases where contemporaneous causal effects are needed, see [11–13].

Some discovery methods focus specifically on DAG discovery. In these methods, assumptions on the graph and probability distribution over the nodes are needed. Consider a DAG $\mathcal{G}$ with nodes x_i where $i \in \mathbb{N}$. A set S is said to *block* [3] the path from x_1 to x_m, where $x_1, x_m \notin S$ if either (1) or (2):

1. There exists a node $x_k \in S$ that satisfies
$$\begin{aligned} & x_{k-1} \rightarrow x_k \rightarrow x_{k+1} \\ \text{or } & x_{k-1} \leftarrow x_k \leftarrow x_{k+1} \\ \text{or } & x_{k-1} \leftarrow x_k \rightarrow x_{k+1} \end{aligned}$$
2. $(\{x_k\} \cup \text{des}(x_k)) \cap S = \emptyset$ and $x_{k-1} \rightarrow x_k \leftarrow x_{k+1}$, where $\text{des}(x_k)$ is the set of descendants of x_k.

Consider disjoint sets of nodes A, B, and C in $\mathcal{G}$. A and B are said to be *d-separated* [3] by C, denoted $A \perp\!\!\!\perp_{\mathcal{G}} B | C$, if each path from x_a to x_b is blocked by C for all $x_a \in A$ and $x_b \in B$. Let $P_{N(\mathcal{G})}$ be the joint probability distribution over $N(\mathcal{G})$, that is, the nodes of $\mathcal{G}$. The *Markov condition* holds if

$$A \perp\!\!\!\perp_{\mathcal{G}} B | C \implies A \perp\!\!\!\perp B | C$$

for all disjoint A, B, $C \subset N(\mathcal{G})$ [3]. Also, $P_{N(\mathcal{G})}$ is said to be *faithful* to $\mathcal{G}$ if

$$A \perp\!\!\!\perp B | C \implies A \perp\!\!\!\perp_{\mathcal{G}} B | C$$

for all disjoint A, B, $C \subset N(\mathcal{G})$ [3]. The Markov condition and faithfulness assumptions provide $x_i \perp\!\!\!\perp_{\mathcal{G}} x_j | S \iff x_i \perp\!\!\!\perp x_j | S$ where $i \neq j$ and $x_i, x_j \notin S \subset N(\mathcal{G})$. That is, two nodes are d-separated in a DAG $\mathcal{G}$ by a set S if and only if they are independent with respect to the probability distribution $P_{N(\mathcal{G})}$ when conditioned on S. The following subsections provide a brief description of a select set of causal discovery methods.

2.1 PC

The PC algorithm is an independence-based DAG discovery method named after its authors Peter Spirtes and Clark Glymour. Under the PC algorithm, causal sufficiency, acyclicity, the Markov condition, and faithfulness are assumed. Under these assumptions, d-separation in the DAG and conditional independence in the probability distribution over the nodes of the graph are equivalent. Causal stationarity is also assumed for time series causal discovery. The PC algorithm utilizes conditional independence (CI) testing and requires the assumptions for the specified CI test.

Typically, independence-based discovery methods prune a fully connected and undirected graph down to its skeleton by performing CI tests between each pair of nodes. Some methods determine the existence of an edge between nodes x_1 and x_2 in a graph $\mathcal{G}$ by conditioning on all subsets of $N(\mathcal{G}) \setminus \{x_1, x_2\}$. The PC algorithm reduces the number of CI tests needed, particularly in the case of sparse DAGs, by conditioning on $(\text{pa}(x_1) \setminus \{x_2\}) \cup (\text{pa}(x_2) \setminus \{x_1\})$, a sufficient conditioning set for determining the existence of an edge between x_1 and x_2 [3]. Clearly, $\text{pa}(x_1)$ and $\text{pa}(x_2)$ are unknown, so PC conditions on all subsets of $(\text{adj}(x_1) \setminus \{x_2\}) \cup (\text{adj}(x_2) \setminus \{x_1\})$ as $(\text{pa}(x_1) \setminus \{x_2\}) \cup (\text{pa}(x_2) \setminus \{x_1\}) \subseteq (\text{adj}(x_1) \setminus \{x_2\}) \cup (\text{adj}(x_2) \setminus \{x_1\}) \subseteq N(\mathcal{G}) \setminus \{x_1, x_2\}$. Once the skeleton graph is discovered, the edges need to be directed. This work focuses on time series data and makes the assumption that there are no contemporaneous effects, that is, their causes must occur strictly before their effects. This implies the directions of the edges in the skeleton graph.

2.2 PCMCI

PCMCI is an independence-based causal discovery method tailored for time series data that utilizes an abridged version of PC combined with the momentary conditional independence (MCI) [14]. Under the PCMCI algorithm, causal sufficiency, acyclicity, the Markov condition, faithfulness, causal stationarity, and no contemporaneous effects are assumed. This DAG discovery method aims to provide statistically significant causal links by producing a sufficient conditioning set while reducing the computational complexity required by the PC algorithm. As described above, CI testing is used to determine d-separation between nodes in a graph.

PCMCI consists of two stages. The first stage utilizes a reduced version of PC to identify a suitable conditioning set, $\hat{\text{pa}}(\vec{x}_i^{(0)})$, for each $\vec{x}_i^{(0)}$. The variable $\vec{x}_i^{(0)}$ represents a time series with 0 lag. These conditioning sets are first assumed to contain all possible parents of $\vec{x}_i^{(0)}$ and then pruned by performing iterative CI tests with a significance level α_{pc} to remove insignificant causal relationships. The sets $\hat{\text{pa}}(\vec{x}_i^{(0)})$ are sorted by the magnitude of their test statistic at each iteration and the CI tests condition on nodes with the strongest relationships [14].

The second stage determines the time series DAG by performing the momentary conditional independence (MCI) test on the pair of nodes $(\vec{x}_j^{(l)}, \vec{x}_i^{(0)})$ for $i, j \in \{1, \ldots, N\}$ and $l \in \{1, \ldots, \tau\}$ where τ is the maximum lag considered. The MCI test is a CI test that conditions on the union of the estimated parents from both nodes, that is $\hat{\text{pa}}(\vec{x}_j^{(l)}) \cup \hat{\text{pa}}(\vec{x}_i^{(0)}) \setminus \{\vec{x}_j^{(l)}\}$, where $\hat{\text{pa}}(\vec{x}_j^{(l)})$ and $\hat{\text{pa}}(\vec{x}_i^{(0)})$ are determined by the first stage. The edge $\vec{x}_j^{(l)} \rightarrow \vec{x}_i^{(0)}$ is included in the DAG if $\vec{x}_j^{(l)} \not\perp\!\!\!\perp \vec{x}_i^{(0)} | \hat{\text{pa}}(\vec{x}_j^{(l)}) \cup \hat{\text{pa}}_k(\vec{x}_i^{(0)}) \setminus \{\vec{x}_j^{(l)}\}$ at a significance level of α_{mci}.

2.3 SINDy

The primary purpose of the widely used Sparse Identification on Nonlinear Dynamics (SINDy) algorithm is to reveal the governing physics of a black-box sampled data set. SINDy aims to discover the underlying dynamics of a complex system by identifying sparse representations of its governing equations from observational data. This algorithm is based on the assumption that there are only a few relevant terms that govern a system's dynamics [15]. SINDy uses sparse regression to find a linear combination of basis functions that best capture the dynamic behavior of the physical system (i.e., to discover the system of differential equations that produced the data). This is accomplished by estimating Ξ in $\frac{dX}{dt} \approx \theta(X)\Xi$, where

$$X = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & & \vdots \\ x_1(t_m) & x_2(t_m) & \cdots & x_n(t_m) \end{bmatrix}$$

$$\dot{X} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \cdots & \dot{x}_n(t_m) \end{bmatrix}$$

$$\theta(X) = \begin{bmatrix} \theta_1(X) & \theta_2(X) & \dots & \theta_l(X) \end{bmatrix}$$

$$\Xi = \begin{bmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{bmatrix}$$

here X is the matrix of time series data, $\{\theta_1, \theta_2, \dots, \theta_l\}$ is a set of basis functions, and Ξ is a presumably sparse coefficient matrix.

This regression process basically allows the analyst to determine the dynamic behavior of each variable x_n by estimating its analytical expression $dx_n dt$ From a causal point of view, note that ordinary and partial differential equations contain cause-effect information as they describe how a variable's rate of change is affected by the other variables. For example, $\frac{dx_1}{dt} = f(x_2, x_3)$ describes the temporal evolution of x_1, which is "caused" by x_2 and x_3 where the causal relationship between x_1 and x_2, x_3, is quantified by f. This information provides the parents for each variable, giving the necessary elements for a DAG. Through sensitivity analysis, the causal relationships (edges) can be quantified.

2.4 MMD

Rather than relying on a pure statistical analysis of the data, an intervention-based Maximum Mean Discrepancy (MMD) testing method to automatically identify causal structure of a control system is proposed in [16]. The core idea behind the proposed causal analysis is to test the physical influence when manipulating the input variables, as stated by the authors "manipulating causes changes the effects". Experimental designs or system interventions have been introduced to manipulate input variables, while the physical influence, i.e. causal relation, has been identified via kernel-based statistical test based on MMD [17]. The authors have demonstrated the proposed method for several control examples, such as kinematic control of a robot and the quadruple tank process model, with accurate causal structure identification and enhanced generalization. In this work, we propose to adapt this intervention-based MMD method for simulation model validation, calibration and uncertainty quantification. As illustrated in Figure 2, MMD can be used to identify causal structure in simulation models, while knowledge-based causal structure can be constructed by utilizing design requirements, experimental data, expert judgements, physics governing equations, etc. Once both causal structures have been constructed, we can perform model validation via comparing the differences in causal graphs and the strength of causal connections. In this way, we can validate whether the structure knowledge has been properly implemented inside the simulation models or not, and we think this process can enhance the generalization of the validation process. If there are any differences identified between the two causal structures, we can perform model calibration to enhance the agreements between the causal structures, and quantify the model uncertainties through the calibrated parameters.

In this work, intervention-based MMD method plays a critical role in the identification of causal structure. In the following, we provide a brief introduction of MMD method. It is a kernel-based two-sample test method that is proposed by [17] to measure the distance between the kernel mean embeddings of two random variables in the reproducing kernel Hilbert space (RKHS). As defined in [17],

$$MMD[\Phi, P, Q] := \sup_{\phi \in \Phi} (E_{x\sim P}[\phi(x)] - E_{y\sim Q}[\phi(y)])$$

where Φ is a class of functions ϕ. Random variables x and y have distributions P and Q, respectively. Finally, $E_{x\sim P}[\phi(x)]$ and $E_{y\sim Q}[\phi(y)]$ are the kernel mean embeddings for x and y, respectively. Then $P = Q$ if and only if $E_{x\sim P}[\phi(x)] = E_{y\sim Q}[\phi(y)]$ or $MMD[\Phi, P, Q] = 0$ for all $\phi \in \Phi$. In practice, an unbiased empirical estimate for the squared population MMD is provided in [17], and summarized in the following:

$$\begin{aligned} MMD_u^2[\Phi, P, Q] = & \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j\neq i}^{m} k(x_i, x_j) \\ & + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} k(y_i, y_j) \\ & - \frac{2}{mn} \sum_{i=1}^{n} \sum_{j=i}^{n} k(x_i, y_j) \end{aligned}$$

where $k(\cdot,\cdot)$ is a continuous kernel function with $k(x, y) = \langle \phi(x), \phi(y) \rangle$ defined in the RKHS. A bootstrap method to compute the MMD_u^2 and test the null hypothesis, i.e., $H_0 : P = Q$ was proposed in [17]. Additionally, [17] proposed moment matching to Pearson curves approach and asymptotic approach using the linear time statistic to accelerate the testing process. In [16], the authors proposed to utilize Chebyshev's inequality combined with surrogate models to further accelerate the statistical hypothesis testing. The executions of surrogate models can be much more efficient compared to the physical models. However, there is an intrinsic difficulty in the selection, training, validation, and generalization of surrogate models. In this work, we rely on the bootstrap method from [17] to perform the hypothesis testing for its simplicity. We have adapted the method to test several dynamic systems, such as the harmonic oscillator, quadruple tank system, and double DAG model.

3. ANALYTICAL COMPARISON

Prior to implementing causal inference for validation, verification, and calibration, the performance of the methods presented in Section 2 is tested. Initial results of directed graph construction from time series data from multiple analytical test cases are presented. Robustness testing against violations of causal discovery method assumptions is necessary to determine the method's performance under real-world conditions where idealized assumptions may not hold. The analytical test cases implemented have

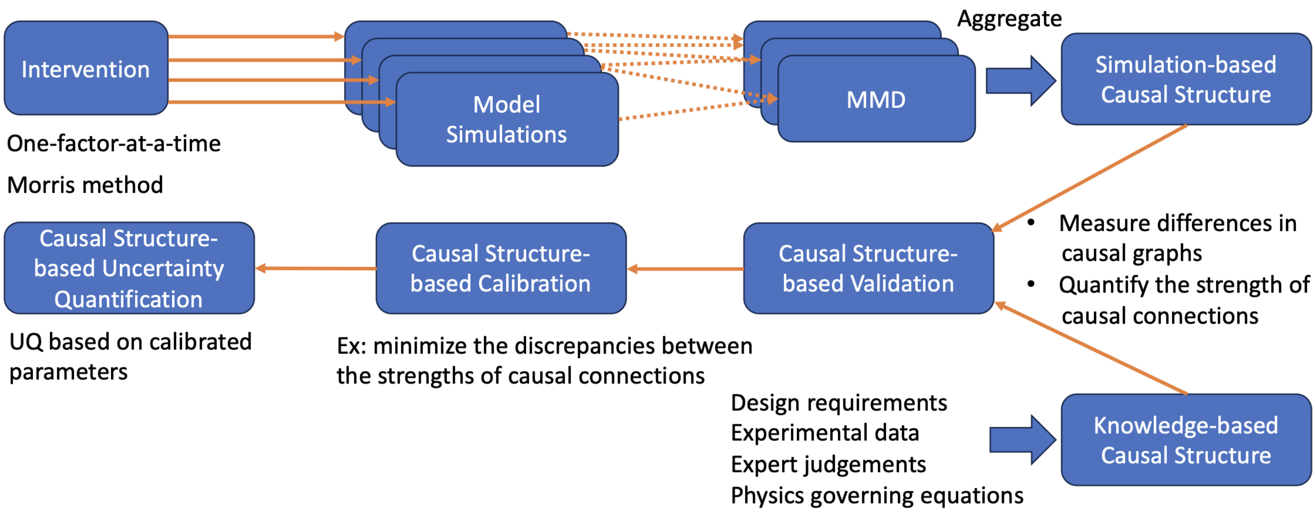

FIGURE 2: INTERVENTION-BASED MMD FOR MODEL VALIDATION, CALIBRATION AND UNCERTAINTY QUANTIFICATION

known directed graphs to quantify the accuracy of causal discovery methods by examining false positive and false negative rates. Causal relationships may be obscured by noise or variations in observational data; therefore, the robustness of the discovery methods is tested in multiple cases with both deterministic and stochastic time series data. For clarification, the stochastic cases are based on the deterministic case (where the system set of ordinary partial differential equations (ODEs) are numerically solved) with the only difference that white noise terms $\eta(t)$ are added. Given a generic ODE in the form $\frac{dx_n}{dt} = f_n(x_1, \ldots, x_N)$, a noise can be added internally, in the form of $\frac{dx_n}{dt} = f_n(x_1, \ldots, x_N) + \eta_n(t)$, or externally where the noise term is added after the behavior of $x_n(t)$ has been determined numerically (i.e., the new variable to be analyzed becomes $x_n(t) + \eta_n(t)$). The first situation aims to emulate physical phenomena that have an impact on the temporal dynamics but are not being explicitly considered in the causal analysis since the observed variables associated with these phenomena are unavailable. The second situation emulates noise that is added by the available monitoring system. A brief description of three test cases follows.

3.1 Harmonic Oscillator

The first analytical test problem is that of a harmonic oscillator. That is, four masses are arranged in a linear fashion linked by springs as seen in Figure 3. Each mass is subject to a force proportional to its displacement from the equilibrium position. From Hooke's Law, the governing equations for n masses can be written as

$$\frac{dP_i}{dt} = k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i}P_i$$
$$\frac{dQ_i}{dt} = \frac{P_i}{m_i}$$

where m_i are the masses, P_i are the momenta, Q_i are the positions, l_i are the lengths of the springs, k_i are the spring constants, and b_i are the friction coefficients [18]. Note that L is Figure 3 is the sum of l_i for $i \in \{1, 2, 3, 4\}$. This example is considered strictly in the deterministic case. The true directed graph of this system contains cycles as seen in Figure 4 [18], therefore violating the underlying assumptions of both the PC and PCMCI discovery methods.

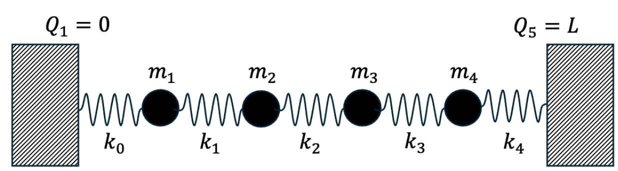

FIGURE 3: FOUR MASS SYSTEM

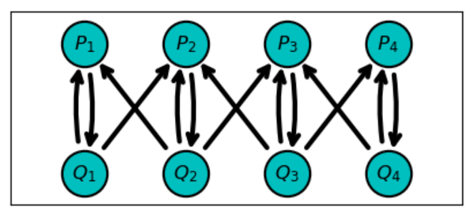

FIGURE 4: GRAPHICAL REPRESENTATION OF THE HARMONIC OSCILLATOR

3.2 Double DAG Model

The double DAG test problem produces time series data for two triples of variables. The pair of triples are independent as seen in the following system of ODEs and as shown in Figure 5. The objective of such a test problem is to evaluate the ability of the algorithms to disassociate causal relationships of a subsystem inside of a larger system.

$$\frac{dA}{dt} = -k_A A \qquad \frac{dC}{dt} = -k_C C$$
$$\frac{dB}{dt} = -k_B B \qquad \frac{dD}{dt} = -k_D D$$
$$\frac{d\alpha}{dt} = k_A A + k_B B \qquad \frac{d\beta}{dt} = k_C C + k_D D$$

This example is considered strictly in the deterministic case, and the causal relationships have no cycles, hence the causal graph is a DAG.

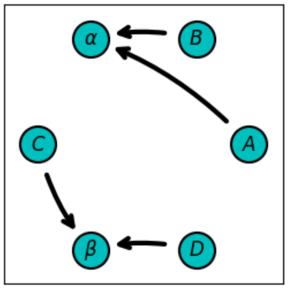

FIGURE 5: GRAPHICAL REPRESENTATION OF THE DOUBLE DAG MODEL

3.3 Quadruple Tank System

The quadruple tank process [19] is illustrated in Figure 6. This model consists of two pumps that provide fluid to four tanks. Two valves are designed to partition the flow rate generated by a pump to two corresponding tanks, that is, the flow generated by the first pump is partitioned by the first valve between the first and fourth tanks. The amount of water from the pumps to the tanks is controlled by two parameters: the mass flow rates u_1 and u_2 generated by the pumps and the flow partition γ_1 and γ_2.

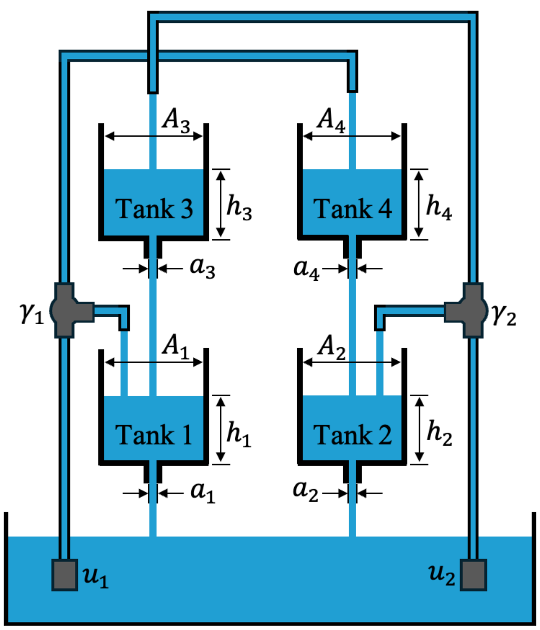

FIGURE 6: GRAPHICAL REPRESENTATION OF THE QUADRUPLE SYSTEM

The water level in each tank is indicated with h_i for $i \in \{1, \ldots, 4\}$; thus, using basic mass conservation laws, the dynamic laws for this system are:

$$\begin{aligned}
\frac{dh_1}{dt} &= -\frac{a_1}{A_1}\sqrt{2gh_1} + \frac{a_3}{A_1}\sqrt{2gh_3} + \frac{\gamma_1 k_1}{A_1}u_1 \\
\frac{dh_2}{dt} &= -\frac{a_2}{A_2}\sqrt{2gh_2} + \frac{a_4}{A_2}\sqrt{2gh_4} + \frac{\gamma_2 k_2}{A_2}u_2 \\
\frac{dh_3}{dt} &= -\frac{a_3}{A_3}\sqrt{2gh_3} + \frac{(1-\gamma_2)k_2}{A_3}u_2 \\
\frac{dh_4}{dt} &= -\frac{a_4}{A_4}\sqrt{2gh_4} + \frac{(1-\gamma_1)k_1}{A_4}u_1
\end{aligned}$$

where g is the acceleration due to gravity. For this analysis, the parameters u_1 and u_2 vary using a sinusoidal behavior and the remaining parameters are set as constant. Both deterministic and stochastic cases are considered. The additional white noise is added in for two scenarios, that is, external (variation from measurement) and internal noise (variation propagating through the system). The causal relationships in this model have no cycles, hence the causal graph is the DAG shown in Figure 7.

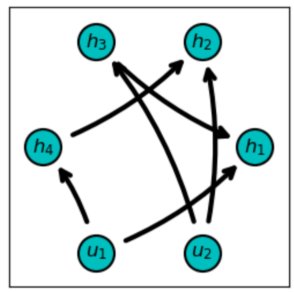

FIGURE 7: DAG MODEL ASSOCIATED WITH THE QUADRUPLE SYSTEM

3.4 Results

The results obtained from applying the causal discovery methods presented in Section 2 on the test problems described above offer valuable insight into the efficacy of each approach. This analysis examines the false positive and false negative (i.e., presence or absence of causal edges in the DAGs) rates of these results when compared to the analytically computed directed graphs. The following outlines the advantages and limitations of the different methods.

The application of PC and PCMCI to the time series generated by the models described above produced varying results; however, they appear to have similar advantages and disadvantages when applied to the test problems. In the case of the quadruple tank system, PC performed perfectly while PCMCI missed a causal connection between the third and first tank as seen in Figure 8. Additionally, PC and PCMCI are limited in application to deterministic data and stochastic data with external noise. These methods outperformed the others in the stochastic case with internal noise.

Using the time series, the causal discovery with SINDy excels in the deterministic case with all three test problems. SINDy does not perform well in the stochastic setting, particularly with internal noise. However, with external noise, SINDy does provide

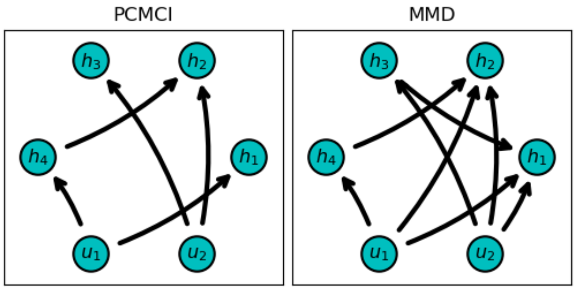

FIGURE 8: QUADRUPLE TANK MMD AND PCMCI RESULTS

favorable results after the data is pre-processed with smoothing via rolling average.

The simulation models of the test problems were employed to test MMD performances. In all case studies, the MMD approach can identify the correct causal structures without any missing connections. However, there are a few weak spurious causal connections, seen in Figure 8, that MMD could not eliminate. In future work, we will conduct further investigation to enhance the MMD testing by post-processing the MMD causal model using conditional independence testing similar to that in PC. In addition, compared to [16] testing approach, the bootstrap method can identify more accurate causal structures with less time series data.

The choice between these methods should be guided by the nature of the system under consideration and by the available data. While PC and PCMCI excel in DAG discovery with stochastic data from a system with internal noise, SINDy proves advantageous when faced with deterministic data and systems that may contain cycles. These tests show it is crucial to consider the specific characteristics of the data and underlying real-world system under consideration.

4. ANALYSIS OF HAIRE EXPERIMENTAL SETUP

Applying causal discovery methods to experimental data from a real-world system can provide further insight into the robustness of the different methods. Knowledge of the advantages and limitations of the causal discovery methods explored in the previous section can inform their applications to a real-world data set. This section will first introduce the real-world system and experiment.

The Helium Air Ingress gas Reactor Experimental (HAIRE) facility at the University of Michigan, seen in Figure 9, was used for this real-world system. It consists of a pair of hollow stainless-steel vessels, one within the other. Tank 1, seen in Figure 9, depressurizes into Tank 2. Tank 2 can also be vented into the atmosphere. The inner tank, Tank 1, can be equipped with many different sized cracks to study depressurization and air ingress phases of depressurized loss of forced cooling and air ingress accidents in high-temperature gas reactors.

In the performed experiment, Tank 1 is filled with helium at approximately 40.0 psig. The crack size in Tank 1 is set at 1.07 mm ± 0.02 mm. Tank 2 is filled with air at atmospheric pressure and is not vented. Tanks 1 and 2 have volumes 0.04204 m^3 ± 0.003809 m^3 and 0.1214 m^3 ± 0.0011 m^3,

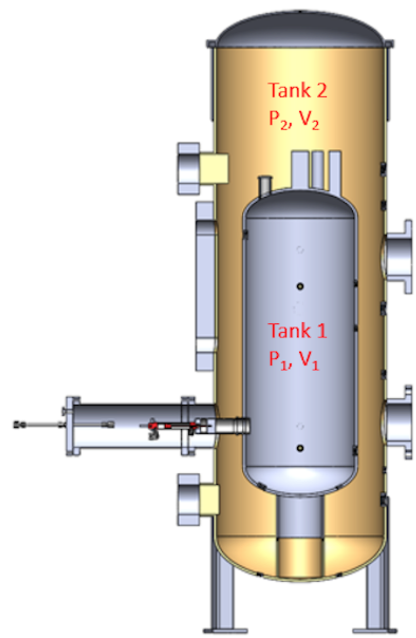

FIGURE 9: SCHEMATIC OF THE HAIRE FACILITY

respectively. A graphical representation of the temporal profile of experimental pressure (psig), filtered pressure (psig), temperature (K), and mass flow rate (kg/s) values are shown in Figure 10.

Given the nature of the transient under consideration, the depressurization of the first tank causes a pressurization of the second through a transfer of mass from the first to second tank. The rates of pressurization and depressurization for the two tanks are proportional to the pressure difference between the tanks; hence, there is a mutual cause and effect relationship between the pressure variables.

Figure 10 shows significant noise variables ΔP and P_2, the difference in the tank pressures and the pressure of Tank 2, respectively. On the scale of the figure it is not visible, but the mass flow, $\dot{m}$ also has noise. This implies that causal discovery with PC or PCMCI may provide favorable results. To apply SINDy, the time series for $\dot{m}$ needs to be smoothed, but the filtered pressures ($F\Delta P$ and P_2) and temperatures (T_1 and T_2) are already smooth.

Preliminary results of the PC, PCMCI, and SINDy methods as applied to the HAIRE experiment show their respective strengths and weaknesses. While PC and PCMCI benefit from the presence of internal noise they struggle with external noise. In an experimental setting, it may be difficult to determine the source of the noise when dealing strictly with data and no knowledge of the system. However, in the HAIRE experiment it is known that the noise is instrumental, i.e. external. An additional challenge for these two methods appeared in the analysis of this experiment. Variables with different time steps can hide causal relationships as these methods depend on the noise propagating from one variable to another to reveal these relationships. Further analysis is being performed to see if PC and PCMCI can perform better in the case external noise case.

The application of SINDy requires a set of basis functions. Designating the correct basis set of functions without prior knowl-

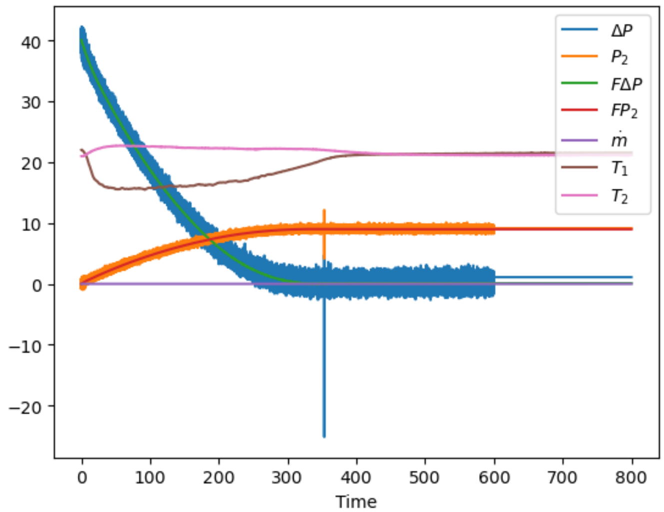

FIGURE 10: SCHEMATIC OF THE HAIRE FACILITY

edge of the underlying physics of the system is difficult, meaning the resulting DAG may be misleading or erroneous. In the HAIRE experiment many of the governing physics are known, but further analysis of the system is ongoing. Currently, we are applying various data pre-processing methods to improve SINDy's ability to detect the expected DAG.

In conclusion, the test problems from sections 2 and 3 have revealed valuable insights into the applicability and robustness of the causal discovery methods from Section 2. PC and PCMCI struggle with inconsistent time steps between variables and deterministic data. SINDy performs well with deterministic data but requires knowledge of the system under consideration to form a sufficient set of basis functions. These tests show it is crucial to consider the characteristics of the data and systems that generate it. Additionally, the tests show that a hybrid of discovery methods that leverages the individuals strengths could benefit causal discovery.

5. GENERATION OF CAUSAL MODELS

In the context of this paper, validation and verification are performed by comparing the causal models generated by the simulation model against experimental measurements and the mathematical model, respectively. Sections 5.1, 5.2, and 5.3 provide details on how the algorithms described in Section 2 are combined to create causal models for three different situations (i.e., simulation model, experimental data, and mathematical model).

5.1 Generation of Causal Models from Simulation Models

It is here assumed that a simulation model is available and it can be described by the generic formulation $\frac{d\mathbf{X}}{dt} = \Phi(\mathbf{X}, t, \mathbf{\Theta})$, where $\mathbf{X} = [x_1, \ldots, x_N]$ represents the set of state variables, $\mathbf{\Theta}$ represents the set of internal parameters (including initial conditions), and Φ represents the function that describes the dynamic behavior of $\mathbf{X}$.

The causal analysis is performed by combining MMD, PC and SINDy as follows:

1. Construct the initial DAG by applying MMD on the simulation model
2. Test conditional independence for the edges obtained in Step 1 using PC to remove spurious causal associations
3. Generate a SINDy model using the provided simulation model and informed by the DAG obtained in Step 2

The obtained causal model consists of the DAG and the SINDy model obtained in Step 2 and 3, respectively.

5.2 Generation of Causal Models from Experimental Data

It is here assumed that only actual measurements of an experiment transient is available in form of a time series; it is here assumed that one single transient is available. As indicated in Section 2, the process of causal discovery starts from pure data elements in the form of $\mathbf{X}(t) = [x_1(t), \ldots, x_N(t)]$. Here we also have additional information about the experimental setup that can be used to improve the limitations of the methods described in Section 2 for causal discovery. Such information can be translated into a nondirected graph.

The causal analysis is performed by combining PC and SINDy as follows:

1. Generate the causal DAG by applying PC on the time series data set $\mathbf{X}(t)$ for the edges originally indicated by the nondirected graph derived from the experimental setup
2. Generate a SINDy model given the time series data set $\mathbf{X}(t)$ and informed by the DAG generated in Step 1.

The obtained causal model consists of the DAG and the SINDy model obtained in Step 1 and 2, respectively.

5.3 Generation of Causal Models from Mathematical Models

Since we are dealing mostly with dynamic systems (i.e., systems that are described by a time-dependent behaviour), the set of governing equations is typically written in the form of ordinary or partial differential equations coupled with equations of state. Without a loss of generality, let's consider a system composed of a set of N variables X_n with $n = 1, \ldots, N$. The dynamic behavior of the system is described by a set of ordinary and partial differential equations (in the form of $\frac{dx_n}{dt} = f_n(x_1, \ldots, x_N)$) and by a set of equations of state (in the form of $x_n = g_n(x_1, \ldots, x_N)$, as indicated in [18]. Often, these equations are derived from energy, mass, and momentum conservation laws, and thus, they contain already causal information. More specifically, the set of variables whose behavior is described by ordinary and partial differential equations are caused by the variables explicitly mentioned in the corresponding function f_n. Thus, the analytical causal DAG can be constructed by the set of equations of the considered system.

6. VALIDATION AND VERIFICATION

In the context of this paper, verification is the process designed to check that the developed simulation model correctly implements the set of governing mathematical equations of the considered system. Similarly, validation is the process designed to test that the developed simulation model matches a transient measured from an experimental setup. As indicated in Section 5, the causal models obtained from simulation models or experimental setups consists of two entities: a DAG and mathematical expression that describes each variable X_n dynamics. Thus, here verification and verification are performed by comparing causal models.

In this respect, code verification is performed by comparing:

- The analytical causal DAG with the causal DAG generated from the simulation model
- The SINDy model generated from the simulation model against the original set of dynamic equations of the considered system.

Similarly, code validation is performed by comparing:

- The causal DAGs generated from the simulation model and from the experimental data set
- The SINDy models generated from the simulation model and from the experimental data set.

Comparison between DAGs simply is performed by checking the discrepancies between graph structures (i.e., missing or additional edges). Comparison between SINDy models directly targets the comparison of the mathematical expression derived by SINDy term by term.

7. CAUSAL MODEL ASSISTED CALIBRATION

Model calibration occurs when discrepancies between simulation and experimental results are observed. If such discrepancies are observed, model calibration takes place by adjusting only those model parameters that influence the established causal relationships.

As indicated in Section 6, discrepancies can be observed between DAGs and SINDy models. A discrepancy between DAGs simply implies that a cause-effect relation between two variables is either non existent or negligible. A discrepancy between SINDy models implies that the quantitative measure of the cause-effect relation between two variables is different. As a basic example, let's consider two SINDy models; in Model 1 the ordinary differential equation associated with variable X_1 is in the form of $\frac{dx_1}{dt} = 3.1x_2 + 2.7x_3$, while Model 2 provides a slightly different formulation, $\frac{dx_1}{dt} = 3.1x_2 + 2.3x_3$. The analytical comparison of these two expressions highlights the different causal contribution of x_3 toward x_1 (i.e., 2.3 vs. 2.7). Such quantitative comparison provides indication on the simulation model parameters to adjust such that newly obtained expressions are closer.

An example based on the model shown in Section 3.3 is here described. A simulation model for this system was developed and the casual model (i.e., the DAG and the corresponding SINDy model) for the simulation case was generated using the method described in Section 5.1.

To emulate virtual experimental conditions, the same simulation model has been modified by adding white noise terms to each equation that describe system dynamic behavior. In addition, a single parameter has been modified which had an impact on the temporal response of the system. Then, the casual model for the two cases were generated from a single transient of the model by following the method described in Section 5.2.

In both situations, the generated DAGs were matching initial expectations given by the mathematical formulation of the system. In addition, the corresponding SINDy models were able to capture the discrepancy introduced in the generation of the experimental data set as indicated below.

SINDy model from experimental data set

(h_1)' = -1.122 f0(h_1) + 1.122 f0(h_3) + 0.714 f1(v_1)
(h_2)' = -0.788 f0(h_2) + 0.786 f0(h_4) + 0.419 f1(v_2)
(h_3)' = -1.123 f0(h_3) + 0.718 f1(v_2)
(h_4)' = -0.788 f0(h_4) + 0.416 f1(v_1)

SINDy model from simulation model

(h_1)' = -1.120 f0(h_1) + 1.115 f0(h_3) + 0.714 f1(u_1)
(h_2)' = -0.787 f0(h_2) + 0.783 f0(h_4) + 0.838 f1(u_2)
(h_3)' = -1.126 f0(h_3) + 0.240 f1(u_2)
(h_4)' = -0.788 f0(h_4) + 0.416 f1(u_1)

This allows the analyst to pinpoint the likely cause of such highlighted discrepancies to the term $\frac{\gamma_2 k_2}{A_2} u_2$ as indicated in the formulation of $\frac{dk_2}{dt}$ shown Section 3.3.

8. CONCLUSION

This paper has presented the initial evaluation of an approach to perform code verification, validation, and calibration that is relying on causal inference methods. The rationale is to move away from classical statistical methods, which are designed to measure the statistical differences between data sets (e.g., simulated data vs. experimental data) toward methods designed to estimate the actual causes that are the source of the statistical differences between data sets. We have presented the causal inference methods that, in our opinion, better fit our scope and their limitations when performing causal inference and discovery on a few analytical test cases. We have highlighted in particular how such limitations can be addressed by coupling qualitative models of the process under consideration with the considered causal inference methods. The analysis of the performance of causal methods have been applied to a few analytical test cases and to a more complex experimental case.

ACKNOWLEDGMENTS

The material presented herein was based on work supported through the INL Laboratory Directed Research & Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517 for LDRD-22A1059-091FP. This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and

the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

REFERENCES

[1] Pang, Tao. *An Introduction to Computational Physics*. Cambridge University Press (2006).

[2] Oberkampf, William and Roy, Christopher. *Verification and Validation in Scientific Computing*. Cambridge University Press (2010).

[3] Peters, Jonas, Janzing, Dominik and Schölkopf, Bernhard. *Elements of causal inference: foundations and learning algorithms*. The MIT Press (2017).

[4] Pearl, Judea, Glymour, Madelyn and Jewell, Nicholas P. *Causal inference in statistics: A primer*. John Wiley & Sons (2016).

[5] Ismael, Jenann. "Causation, Free Will, and Naturalism." *Scientific Metaphysics*. Oxford University Press (2013): pp. 208–235. URL https://doi.org/10.1093/acprof:oso/9780199696499.003.0010.

[6] Nogueira, Ana Rita, Pugnana, Andrea, Ruggieri, Salvatore, Pedreschi, Dino and Gama, João. "Methods and tools for causal discovery and causal inference." *Wiley interdisciplinary reviews: data mining and knowledge discovery* Vol. 12 No. 2 (2022): p. e1449.

[7] Zhang, Jiji. "On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias." *Artificial Intelligence* Vol. 172 No. 16-17 (2008): pp. 1873–1896.

[8] Forré, Patrick and Mooij, Joris M. "Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders." *arXiv preprint arXiv:1807.03024* (2018).

[9] Gerhardus, Andreas and Runge, Jakob. "High-recall causal discovery for autocorrelated time series with latent confounders." *Advances in Neural Information Processing Systems* Vol. 33 (2020): pp. 12615–12625.

[10] Runge, Jakob, Gerhardus, Andreas, Varando, Gherardo, Eyring, Veronika and Camps-Valls, Gustau. "Causal inference for time series." *Nature Reviews Earth & Environment* Vol. 4 No. 7 (2023): pp. 487–505. DOI 10.1038/s43017-023-00431-y. URL https://doi.org/10.1038/s43017-023-00431-y.

[11] Runge, Jakob. "Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets." *Conference on Uncertainty in Artificial Intelligence*: pp. 1388–1397. 2020. PMLR.

[12] Assaad, Charles K, Devijver, Emilie and Gaussier, Eric. "Discovery of extended summary graphs in time series." *Uncertainty in Artificial Intelligence*: pp. 96–106. 2022. PMLR.

[13] Rohekar, Raanan Yehezkel, Nisimov, Shami, Gurwicz, Yaniv and Novik, Gal. "From Temporal to Contemporaneous Iterative Causal Discovery in the Presence of Latent Confounders." *International Conference on Machine Learning*: pp. 39939–39950. 2023. PMLR.

[14] Runge, Jakob, Nowack, Peer, Kretschmer, Marlene, Flaxman, Seth and Sejdinovic, Dino. "Detecting and quantifying causal associations in large nonlinear time series datasets." *Science advances* Vol. 5 No. 11 (2019): p. eaau4996.

[15] Brunton, Steven L, Proctor, Joshua L and Kutz, J Nathan. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." *Proceedings of the national academy of sciences* Vol. 113 No. 15 (2016): pp. 3932–3937.

[16] Baumann, Dominik, Solowjow, Friedrich, Johansson, Karl H and Trimpe, Sebastian. "Identifying causal structure in dynamical systems." *arXiv preprint arXiv:2006.03906* (2020).

[17] Gretton, Arthur, Borgwardt, Karsten M, Rasch, Malte J, Schölkopf, Bernhard and Smola, Alexander. "A kernel two-sample test." *The Journal of Machine Learning Research* Vol. 13 No. 1 (2012): pp. 723–773.

[18] Mooij, J., Janzing, D. and Schölkopf, B. "From ordinary differential equations to structural causal models: the deterministic case." *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence*: pp. 440–448. 2013.

[19] Johansson, Karl Henrik. "The quadruple-tank process: A multivariable laboratory process with an adjustable zero." *IEEE Transactions on control systems technology* Vol. 8 No. 3 (2000): pp. 456–465.

