

CAnDOIT: Causal Discovery with Observational and Interventional Data from Time Series

Luca Castri,* Sariah Mghames, Marc Hanheide, and Nicola Bellotto

The study of cause and effect is of the utmost importance in many branches of science, but also for many practical applications of intelligent systems. In particular, identifying causal relationships in situations that include hidden factors is a major challenge for methods that rely solely on observational data for building causal models. This article proposes CAnDOIT, a causal discovery method to reconstruct causal models using both observational and interventional time-series data. The use of interventional data in the causal analysis is crucial for real-world applications, such as robotics, where the scenario is highly complex and observational data alone are often insufficient to uncover the correct causal structure. Validation of the method is performed initially on randomly generated synthetic models and subsequently on a well-known benchmark for causal structure learning in a robotic manipulation environment. The experiments demonstrate that the approach can effectively handle data from interventions and exploit them to enhance the accuracy of the causal analysis. A Python implementation of CAnDOIT is developed and is publicly available on GitHub: <https://github.com/lcastri/causalflow>.

scenarios have been developed.^[6] That is, causal models are generated relying only on passive observations of the surrounding world, without taking into account the effect of hypothetical interventions. Observational data are often insufficient to retrieve the correct causal model in complex scenarios where it is impossible to account for all the variables responsible for the system's evolution. In such cases, hybrid approaches that incorporate first principles into data-driven models can potentially improve the quality of the model.^[7–10] Using first-principles models as constraints might reduce the degrees of freedom in causal discovery algorithms and improve the accuracy of their results. However, this approach requires prior knowledge of the system's type or the fundamental physical laws governing its behavior, which can limit the algorithm's versatility. This potential solution has not


1. Introduction

The increasing complexity of the environments in which future intelligent systems are expected to operate requires an understanding of the cause–effect relations between events, especially those affecting the performance of such systems.^[1] Unsurprisingly, Causal Inference has become a crucial analysis across various research domains, including Earth science,^[2] healthcare,^[3] and robotics.^[4,5] To estimate the causal models that explain the available data, in recent decades, many causal discovery algorithms for observational data in static and dynamic

been widely investigated in the causality community yet. Conversely, an area that has been studied more extensively in the causality community involves using interventional data, that is, data from experiments, to eliminate spurious correlations and enhance the quality of the causal model.^[1,11] For this reason, recent works have led to the development of algorithms for static (i.e., time independent) data capable of constructing the causal model by leveraging both observations and interventions.^[12–18] To our knowledge though, none of the previously mentioned works conduct causal discovery analysis for time-series data using observations and interventions. This represents a significant limitation for many real-world problems, such as sensor readings or dynamic process monitoring. Therefore, although it is obvious that using intervention information significantly enhances the quality of the causal model,^[1,11] our specific goal is to develop a causal discovery algorithm for time-series data capable of achieving this. Indeed, in this article, we extend and improve a recent state-of-the-art causal discovery for time-series data, latent-Peter and Clarke momentary conditional independence (LPCMCi),^[19] enabling it to retrieve the causal model from both observational and interventional data, which increases the accuracy of the causal analysis. We named our solution CAnDOIT, CAusal Discovery with Observational and Interventional data from Time Series. The main contributions of this article are therefore the following: 1) A new causal discovery algorithm for time series capable of integrating interventional and observational data into the causal discovery process. To this end, we significantly enhance the accuracy of causal discovery,

L. Castri, S. Mghames, M. Hanheide, N. Bellotto
L-CAS, School of Engineering & Physical Sciences
University of Lincoln
Lincoln LN6 7TS, UK
E-mail: lcastri@lincoln.ac.uk

N. Bellotto
Department of Information Engineering
University of Padua
35122 Padua, Italy

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202400181>.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202400181

outperforming current state-of-the-art algorithms; and 2) A publicly available implementation of the algorithm, experimentally evaluated on various synthetic models from an ad-hoc random-model generator and a simulated robot manipulation scenario.

The article is structured as follows: related work about causal from observations with and without interventions is presented in Section 2; Section 3 explains the implementation details of our method; Section 4 and 5 present experimental results in randomly generated and robot simulation models, respectively, to validate correctness and performance of our solution; finally, we conclude the article in Section 6 discussing achievements and future work.

2. Related Work

2.1. Observation-Based Causal Discovery

Structural causal models (SCMs) and directed acyclic graphs (DAGs) are at the core of causal inference.^[20] They both represent system variables and their causal dependencies, the former with mathematical expressions, and the latter with nodes and oriented edges. Various methods have been developed to derive causal relationships from observational data, categorized into three main classes:^[6] 1) constraint-based methods, such as PC and fast causal inference (FCI),^[21] use CI tests to recover the causal graph; 2) score-based methods, like greedy equivalence search^[22] and NOTEARS,^[23] assigns scores to DAGs and explore the score space accordingly; and 3) noise-based methods, such as linear non-Gaussian acyclic models (LiNGAM),^[24] identify causal structure within variables, assuming linearity, non-Gaussianity, and acyclicity of the model. However, most of these algorithms are applicable only to static data without temporal information, so they are not suitable for many interesting problems in Earth science,^[2] healthcare,^[3] or robotics.^[4,5]

To address this limitation, various causal discovery algorithm have been developed to deal with time-series data.^[25] For example, within the area of Granger causality, the temporal causal discovery framework^[26] employs deep neural networks to learn complex nonlinear causal relationships among time series. However, it faces challenges due to many hyperparameters and lacks a direct way to set the maximum timelags. Among the noise-based methods, there are time series models with independent noise^[27] and vector autoregressive LiNGAM.^[28] In the score-based methods class, the time-series version of NOTEARS, called DYNOTEARS,^[29] has been introduced. In the constraint-based methods category, variations of the FCI and PC algorithms, namely, time-series FCI (tsFCI)^[30] and PC momentary CI (PCMCI),^[31] were tailored to handle time-series data. PCMCI, in particular, has been applied to many research fields, including climate, healthcare, and robotics.^[3,32,33] Recently, various extensions of this algorithm have been introduced. PCMCI⁺,^[34] for example, allows the discovery of simultaneous causal dependencies. Its further extension, LPCMCI, allows simultaneous causal dependencies and latent confounders.^[19] Filtered-PCMCI,^[35] instead, includes an additional transfer entropy-based feature-selection module to remove unnecessary variables and perform causal discovery only on those responsible for the dynamic

evolution of the observed scenario, resulting in faster model generation. Joint-PCMCI⁺ (J-PCMCI⁺)^[36] enables causal structure learning from multiple observational datasets. Finally, PCMCI_Ω^[37] assumes a finite periodical repetition of causal mechanisms, which is suitable for semistationary time-series data.

Despite dealing with time-series, the above algorithms work only with observational data and cannot handle the theoretical concept of intervention, specifically the so-called do-operator.^[1] This operator overrides the structural equations of the intervention variables, disrupting the original dependencies in the observational setting. The concept of intervention is a fundamental aspect of causality. Studying how a variable reacts in response to a forced modification of another variable provides a deeper understanding of the system compared to what simple observations can offer.

2.2. Observation and Intervention-Based Causal Discovery

Incorporating interventional data into the causal discovery process differs from its use in other causal problems, like measuring the average treatment effect (ATE).^[11] The ATE is a metric employed to evaluate the causal influence of a specific intervention on an outcome, given the presence of a causal model. Conversely, interventional data can be used to enhance the accuracy of the discovery process to find that model. Formally, performing causal discovery on observational data means that the true DAG is only identifiable up to a Markov equivalence class (MEC).^[1] However, by incorporating interventional data, identifiability can be improved up to an interventional class (I-MEC),^[38] which is a subset of MEC. The same consideration applies when dealing with maximal ancestral graphs (MAGs), where the equivalence class is represented by a partial ancestral graph (PAG), capturing the CI relations among observed variables while accounting for latent variables. For this reason, various algorithms have been developed for causal discovery using both observational and interventional datasets.

The interventional greedy sparsest permutation (IGSP)^[38] algorithm has been proposed to learn causal DAGs when both observational and interventional data are available. Its extension, unknown target IGSP (UT-IGSP),^[39] enables DAG learning also when the intervention targets are partially or completely unknown. An extension of the FCI causal discovery method has been proposed to enable causal model learning by combining observations and experiments (i.e., soft interventions).^[16] This extension has been further improved by Ψ -FCI to handle unknown intervention targets.^[17] Recent neural approaches, such as differentiable causal discovery with interventions (DCDI)^[13] and efficient neural causal discovery,^[40] can handle various types of interventional data. An extension of DCDI has also been developed to use latent interventions.^[18] The multiple interventional datasets for efficient global causal structure learning^[15] algorithm learns a causal skeleton using a variant of the PC algorithm adapted to handle multiple interventional datasets simultaneously. It then orients the skeleton by identifying an I-MEC and applies a score function to conduct a greedy search for each causal DAG from each dataset within the remaining search space. Finally, the graphs are merged to obtain the

final causal structure. Recently, a novel framework, called joint causal inference (JCI),^[14] has been proposed to facilitate causal discovery considering different contexts of the same scenario. The authors elaborate on how this approach can be employed to model interventions as context changes. By exploiting this framework, existing observational causal inference algorithms can be enhanced to learn causal structures from interventional data.

All these works, however, conduct causal analysis from observations and interventions on static datasets. As already mentioned, this limitation poses significant challenges for many applications where temporal information cannot be ignored and is crucial for correctly modeling the system. Our approach targets these types of applications and differs from previous solutions as it performs causal discovery on time-series data by incorporating both observations and interventions.

3. Causal Discovery Based on Observational and Interventional Data

The solutions proposed in this article, called CAnDOIT, extends and improves the state-of-the-art causal discovery algorithm for time-series data, LPCMCI,^[19] taking inspiration from the way JCI^[14] handles interventions with known target. The result is a new algorithm that enables precise causal analysis, using both observational and interventional data, with significantly improved accuracy. Specifically, CAnDOIT extends LPCMCI by performing causal discovery with hard interventions on known targets.

3.1. LPCMCI

The original LPCMCI^[19] algorithm is an enhancement of PCMCI⁺,^[34] extending it by removing the causal sufficiency assumption as an initial condition. LPCMCI retains the ability to detect both time-lagged and contemporaneous cause–effect relationships between variables from its predecessors, PCMCI and PCMCI⁺. Additionally, it relaxes the causal sufficiency assumption, allowing for the inclusion of latent confounders, that is, unobserved common causes.

In more detail, LPCMCI is a constraint-based causal discovery algorithm that begins with a fully connected graph G . The algorithm then proceeds through a preliminary phase, which removes many (though not necessarily all) false links and repeatedly applies a set of orientation rules to define the heads and tails of the edges. These rules help identify a subset of the (non-) ancestorships in G . The identified nonancestorships and ancestorships further constrain the conditioning sets for subsequent CI tests. The goal of this iterative process is to accurately determine a subset of the parentships in G . This subset is then used in the final phase of the algorithm, which involves further removal of links, application of orientation rules, and use of identified (non-)ancestorships. Ultimately, LPCMCI outputs a time-series PAG.

In real-world applications, the challenge of latent confounders is almost always present, as it is generally impossible to account for all variables influencing the system's evolution. Consequently, given our interest for real-world applications in

dynamic scenarios, we based our CAnDOIT algorithm on LPCMCI to leverage its capability to handle hidden confounders effectively.

3.2. Interventions Through Context Variables

Combining observational and interventional data in the causal discovery process requires the causal structure related to the intervention variable to adapt to both the observational and interventional cases. Specifically, we need to consider the parents of the intervention variable when dealing with observational data, while we must break all the links affecting the intervention variable when performing the intervention. To address this challenge, we use context nodes to model interventions, taking inspiration from the JCI framework. The latter distinguishes between system variables, which describe the actual system, and context variables, which refer to the observation context. While the system variables are treated as endogenous, the context variables are typically (but not necessarily) exogenous. System and context variables form a new metasystem M , which is used for the causal analysis. Accordingly, our metasystem is defined as follows.

$$\mathcal{M}: \begin{cases} X_i(t) = \tilde{f}(\text{Pa}(X_i), CX_k) & i \in \mathcal{I}, k \in \mathcal{K} \\ CX_k = f_k & k \in \mathcal{K} \end{cases} \quad (1)$$

where \mathcal{I} represents the set of system variables defined as $X = (X_i)_{i \in \mathcal{I}}$, while \mathcal{K} represents the set of context variables defined as $C = (CX_k)_{k \in \mathcal{K}}$. $\text{Pa}(X_i)$ is the parent set of the system variable X_i , instead CX_k is the context variable k . Moreover, the function \tilde{f} models the system variables and can be decomposed as follows.

$$\tilde{f}(\text{Pa}(X_i), CX_k): \begin{cases} f(\text{Pa}(X_i)) & CX_k = 0 \\ CX_k = \xi_k & CX_k = \xi_k \end{cases} \quad (2)$$

where f represents the function that models the evolution of the system variable X_i in the observational case, which depends solely on its parent set $\text{Pa}(X_i)$. Referring again to Equation (1), the function f_k models the context variables and it is defined as follows.

$$f_k: \begin{cases} \xi_k & \text{interventional mode for } k \\ 0 & \text{observational mode for } k \end{cases} \quad (3)$$

The case where $CX_k = 0$ corresponds to no intervention, that is, the observational baseline, while $CX_k = \xi_k$ models the intervention case, where ξ_k is the actual intervention value assigned to the variable X_i through the context variable CX_k . This approach is consistent with the concept of force variables adopted in the literature.^[14,41] In our case, to model hard interventions with known targets as context changes in time-series data, we assume the three assumptions made by the JCI framework, specifically we used the JCI123 version of the framework.^[14] The three assumptions are the following.

JCI 1 Exogeneity: No system variable causes any context variable, i.e.

$$\forall k \in \mathcal{K}, \forall i \in \mathcal{I}: i \rightarrow k \notin \mathcal{G}(\mathcal{M}) \quad (4)$$

JCI 2 Complete randomized context: No context variable is confounded with a system variable, that is

$$\forall k \in \mathcal{K}, \forall i \in \mathcal{I} : i \leftrightarrow k \notin \mathcal{G}(\mathcal{M}) \quad (5)$$

JCI 3 Generic context model: The context graph $\mathcal{G}(\mathcal{M})_{\mathcal{K}}$ is of the following special form.

$$\forall k \neq k' \in \mathcal{K} : k \leftrightarrow k' \in \mathcal{G}(\mathcal{M}) \wedge k \rightarrow k' \notin \mathcal{G}(\mathcal{M}) \quad (6)$$

where $\mathcal{G}(\mathcal{M})$ represents the graph of the metasystem G , and $\mathcal{G}(\mathcal{M})_{\mathcal{K}}$ is the subgraph of $\mathcal{G}(\mathcal{M})$ specifically related to the context variables. The exogeneity assumption (JCI1) is commonly applied when the context encodes interventions on the system that have been predetermined and executed before any measurements are taken. This is the approach we use to model and conduct interventions. Assumption 2 posits that the choice of intervention is independent of any other factors that might influence the system and that the observed context variables fully describe the context. JCI3 is introduced to model a scenario where the context distribution contains no (conditional) independencies. This assumption is made when the goal is not to model causal relationships among the context variables themselves, but rather to use these variables to help model causal relationships among the system variables. Note that JCI3 can only be assumed if both JCI1 and JCI2 hold, which is true in our case. Additionally, since we are dealing with interventions with known targets, we have added a fourth assumption: each context node can only influence a single-system variable, specifically the one undergoing the intervention. This is because the intervention variable is known and provided to the algorithm.

At this point, we can further clarify the concept of context nodes. Essentially, the context node is a dummy exogenous variable (i.e., a “metavariable” that does not exist in the real system) that is used to inject the interventional data into the intervention variable without ignoring its parents. Since the context node is exogenous (by the JCI1 assumption) and models the intervention, the model’s structure does not change when transitioning between observational and interventional cases. Following the JCI framework for causal discovery with both observational and interventional data,^[14] the interventional process in CANDOIT is generated by creating context nodes (e.g., CX_k) that are added as parent of the system variables (e.g., X_k) and govern

their possible values. The context node affects its system variable instantaneously (at the same time step) injecting the interventional data into it and maintaining its value constant for the duration of the intervention. Note that, as the context node does not carry temporal information, that is, its value does not change over time, we modeled it as a unique node in the graph confounding its corresponding system variable at all the time intervals (see **Figure 1** CANDOIT blocks). This way of modeling context node as a parameter rather than a variable has been previously adopted in the literature.^[16,38]

Figure 1a shows an example of a context variable to handle a hard intervention. Initially, a causal structure is obtained using only observational data, Figure 1a (OBS). When intervening on X_2 , all its input dependencies need to be removed, resulting in a different causal structure Figure 1a (INT X_2). CANDOIT instead creates a context node CX_2 connected to X_2 at each time step. Since the intervention is now assigned to the context variable CX_2 , and the latter is not affected by any other variables (JCI1 assumption), there is no need to modify the original causal structure. As a result, we obtain a unified causal structure Figure 1a (CANDOIT) that represents both observational and interventional data. Figure 1b demonstrates the application of CANDOIT with multiple interventions. Figure 1b (OBS) shows the causal structure describing the observational case. Figure 1b (INT X_2) and (INT X_0) depict the two different causal models corresponding to the two interventions, X_2 and X_0 , respectively. Figure 1b (CANDOIT) illustrates the unified causal structure obtained by CANDOIT.

Notably, in the case of multiple interventions, the context nodes are all connected by bidirected links between each pair of context nodes (JCI3). An example of a common situation where context variables are employed to model interventions and the context distribution contains no CIs is what the JCI authors refer to as a diagonal design, that is, only one intervention is active at a time.^[14] Applying this concept to the time-series domain, we model multiple interventions in such a way that only one intervention is active within a specific time interval.

3.3. Faithfulness Assumption

In the context of constraint-based causal discovery, the so-called faithfulness assumption plays a crucial role in ensuring that the

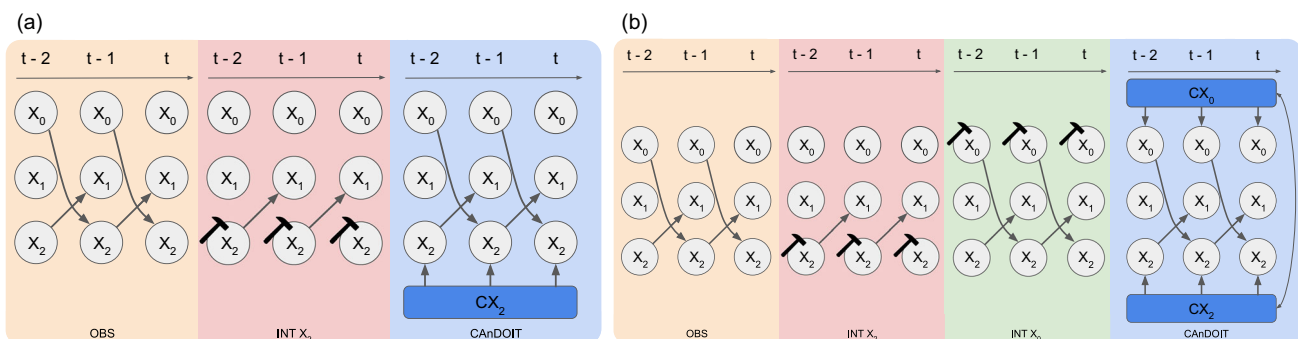


Figure 1. CANDOIT effectively employs context variables to handle observational and interventional data, resulting in a unified causal structure that accommodates both types of data. CANDOIT blocks in a, b) provide examples of this unified causal structure in cases of single and multiple interventions, respectively.

inferred causal relationships from observational data represent accurately the true underlying causal structure. Essentially, this assumption asserts that all CI relationships in the data are encoded in the reconstructed causal graph. This means that if a variable is conditionally independent from another, given a set of variables in the data, then no direct edge should exist between them in the graph.^[42] Like its predecessor LPCMCI, CANDOIT is a constraint-based causal discovery method that requires the faithfulness assumption, even with the introduction of new context variables for modeling interventions. This is crucial for the correctness of the algorithm. Since we combine both observational and interventional data, our output (time-series PAG) must be faithful to the CIs present in both types of data. To ensure this, we test for (CI using only the pooled dataset (observational and interventional) and never consider the two cases separately. Testing for independence using only part of the data might produce results that are unfaithful to the remaining data.

From the example in Figure 1a, we can see how CANDOIT reconstructs the causal model using both observational and interventional data, consistently with the faithfulness assumption. As shown in Figure 1a (CANDOIT), the metasystem M created by CANDOIT includes three system variables (X_0, X_1, X_2) and a context variable CX_2 , which models the intervention on X_2 . Although the hard intervention on the variable X_2 overrides the relationships between X_2 and its parents $Pa(X_2) = X_0$, it does not compromise the faithfulness of the joint distribution $P(CX_2, X_0, X_1, X_2)$ with respect to the joint causal graph. Indeed, even though X_2 and $Pa(X_2) = X_0$ are independent in the intervention setting, they are still dependent in the observational one. Therefore, as long as we do not test for independences in the subset of interventional data separately, but restrict ourselves to testing independences only in the pooled dataset that combines all contexts, the faithfulness assumption remains valid.^[14]

3.4. CANDOIT Algorithm

Figure 2 depicts a detailed flowchart of CANDOIT, explaining each step of the algorithm with an example. In particular, the steps executed by our approach are as follows: 1) CANDOIT takes observational and interventional data as input; 2) Using the

knowledge of the intervention target Z , the context block adds the context node CZ to the set of variables considered in M , plus an instantaneous link $CZ \rightarrow Z$ to the initial causal structure, that is, a fully connected graph that is the starting point of the LPCMCI algorithm; 3) The system variables (X, Y, Z), along with the context node CZ , are injected into the causal discovery block; 4) LPCMCI performs the causal analysis on the metasystem M and then removes both the context variable CZ and the link $CZ \rightarrow Z$ before returning the causal model; and 5) CANDOIT outputs a time-series PAG.

A detailed pseudocode explanation of our approach is presented in Algorithm 1. A Python implementation of CANDOIT is also publicly available (<https://github.com/lcastricausalflow>).

Being based on LPCMCI, our CANDOIT inherits its necessary conditions for proper functioning: causal Markov condition, faithfulness, acyclicity. Furthermore, like its predecessor, CANDOIT can adapt to any type of data, including linear and nonlinear relationships, multiple timelags, various types of noise, and it cannot detect cyclical relationships. It retains the output format of a time-series PAG. Specifically, CANDOIT produces a time-series PAG with a number of layers determined by the algorithm parameters τ_{\min} and τ_{\max} (see Algorithm 3 inputs). By default, τ_{\min} is set to 0 to account for the instantaneous links created for the context variables. On the other hand, τ_{\max} represents the maximum time delay considered when the algorithm performs CI tests between variables across different time steps. Consequently, the time-series PAG consists of $\tau_{\max} + 1$ layers, corresponding to the time steps $t - \tau_{\max}, t - (\tau_{\max} - 1), \dots, t$.

PAGs are used to represent the MEC of MAGs. The latter extend the DAGs representation by including also the bidirected link (\leftrightarrow) to represent variables confounded by a latent confounder. PAGs further generalize MAGs by incorporating additional edge types, specifically $\circ \rightarrow$ and $\circ \circ$, to handle uncertainties in edge orientations. For example, in a PAG, a link $\circ \rightarrow$ corresponds to two possible MAGs: $X \rightarrow Y$ (where X is an ancestor of Y) or $X \leftrightarrow Y$ (where X and Y are confounded by a latent variable). Similarly, a link $\circ \circ$ in a PAG represents two possible MAGs: $X \rightarrow Y$ (where X is an ancestor of Y) or $Y \rightarrow X$ (where Y is an ancestor of X).

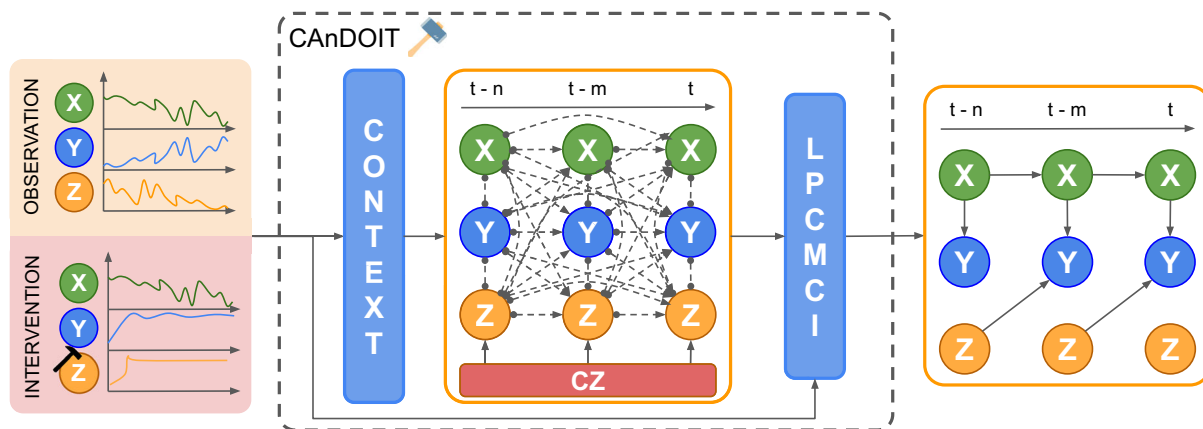


Figure 2. CANDOIT's block scheme representation. CANDOIT processes observational and interventional data; the context block adds context variables (CZ) linked to the actual intervention variable (Z) with an instantaneous link ($CZ \rightarrow Z$); Finally, the LPCMCI block finalizes the causal discovery process.

Algorithm 1. CAnDOIT.

Require: obs. D_{obs} and int. D_{int} ts data, int. target variables X_i , significance level α , min and max timelag τ_{min} , τ_{max} .

1. $CM_0 \leftarrow$ fully connected PAG with $\circ \rightarrow$ for lagged dependencies and $\circ \circ$ for contemporaneous dependencies ← LPCMCI starting point
2. $\mathcal{M} \leftarrow \mathcal{F}$ add the set of system variables $X = (X_i)_{i \in \mathcal{F}}$ to the meta-system \mathcal{M}
3. **for each** int. target variables X_i **do**
4. $CX_k \leftarrow$ create the context variable CX_k associated to the intervention system variable X_i
5. $\mathcal{M} \leftarrow CX_k$ add CX_k to the meta-system \mathcal{M}
6. $CM_0 \leftarrow$ add CX_k to the LPCMCI initial condition CM_0
7. **for each** τ in range(τ_{min} , τ_{max}) **do**
8. $CM_0 \leftarrow$ add the link $CX_k \rightarrow X_i(t-\tau)$ to CM_0
9. $D_s \leftarrow [D_{obs}, D_{int}]$
10. $CM =$ LPCMCI ($D_s, \alpha, \tau_{min}, \tau_{max}, CM_0$)
11. $CM \leftarrow$ remove context variables CX_k and related links
12. **return** CM ← time-series PAG

4. Evaluation on Random Synthetic Models

We designed experiments to assess the accuracy and effectiveness of our approach in handling interventional data and to examine its impact on the causal discovery process. These experiments aimed to evaluate our method’s performance in various scenarios: linear and nonlinear systems, with and without latent confounders, and involving single or multiple interventions.

In this section, we evaluated the correctness and performance of CAnDOIT on a large set of randomly generated synthetic models. Later, in Section 5, we will also test it on a simulated robotic scenario.

4.1. Random Synthetic Models

To evaluate our approach’s effectiveness in handling interventional data and its impact on the causal structure, we devised five testing strategies, denoted as $S_1, S_2, S_3, S_4,$ and S_5 . In the first strategy (S_1), we assessed our approach’s performance with linear systems while varying the number of observable variables and without hidden confounders. Strategy S_2 extends S_1 by introducing hidden confounders while retaining the system as linear and maintaining the same range of the number of variables. In both S_1 and S_2 , only a single intervention is conducted. In strategy S_3 , we evaluated how CAnDOIT performs with linear systems and hidden confounders when multiple interventions are applied, while keeping the number of observable variables fixed. Strategies S_4 and S_5 mirror S_2 and S_3 , respectively, but focus on nonlinear systems.

To facilitate the aforementioned evaluations, we developed a synthetic model generator capable of creating random systems of equations with hidden confounders. This tool offers various adjustable parameters, including time-series’ length, number of observable variables, observable parents per variable (link density), hidden confounders, and confounded variables per hidden confounder. Moreover, it includes also noise configuration, minimum and maximum time delay to consider in the equations, coefficient range, plus functional forms, and operators used to link various equation’s terms. With this generator, we can create

ground-truth causal models to test our algorithm and generate observational and interventional data based on the generated causal structure. This enables us to simulate different scenarios and thoroughly analyze the behavior of our approach under various conditions. A detailed explanation of the random-model generator is presented in Appendix A, Supporting Information. Examples of randomly-generated causal models for each specific evaluation strategy are shown in **Figure 3**.

Some model generators and causal discovery benchmarks have been introduced in the literature, for example, CauseMe,^[32] a collection of synthetic, hybrid, and real observational data mostly for climate and weather scenarios. Other random model generators have been proposed.^[43] However none of them can generate entirely random causal structures along with both observational and interventional data like the one here developed.

Our synthetic model evaluation is crucial to validate CAnDOIT’s performance in retrieving more accurate causal models compared to other methods. Opportunely tuned, our random model generator is not constrained by any assumptions regarding model linearity, effectively excluding from the evaluation noise-based discovery methods.^[27,28] As CAnDOIT is based on the LPCMCI causal discovery algorithm, the latter was used as the benchmark for this evaluation.

4.2. Evaluation Setting

We evaluated CAnDOIT using the random model generator mentioned in Section 4.1. Five evaluation strategies were devised. For each strategy, we first conducted causal discovery analysis with LPCMCI. Based on its result, we then selected which variable(s) to intervene on. The choice of intervention variable(s) was determined by the observable variable(s) identified by LPCMCI as having ambiguous links ($\circ \rightarrow$ and $\circ \circ$). The goal of these interventions was to (partially) resolve the ambiguities in these links. In cases where LPCMCI returned a causal model with ambiguous links originating from multiple nodes, we performed an intervention for each of these nodes. In the evaluation results presented in Section 4.4, we report two curves for

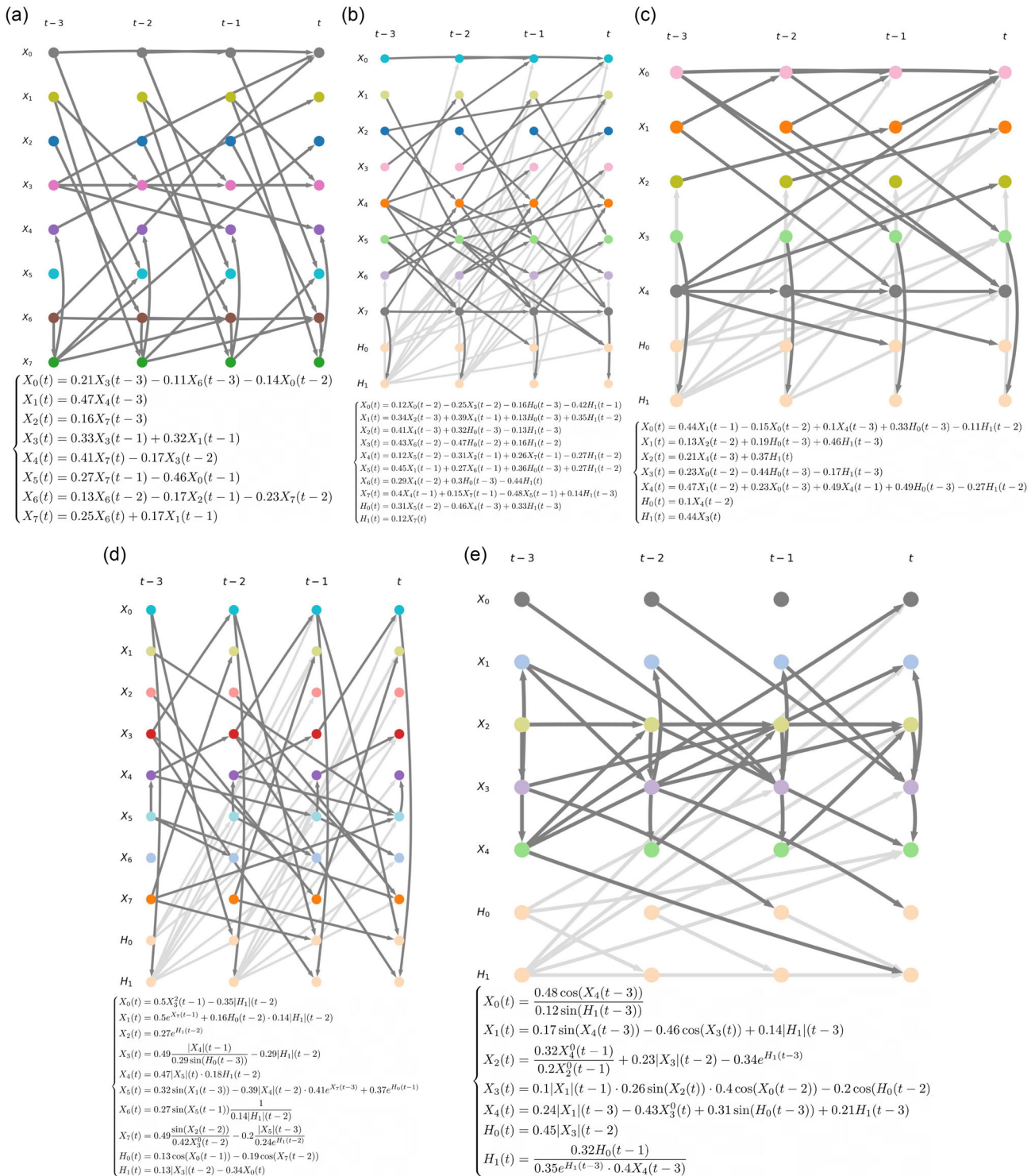


Figure 3. Causal models randomly generated with their corresponding systems of equations on the bottom. a) An example of a linear system for the S_1 evaluation strategy, which has no hidden confounders. b) A random linear system for the S_2 evaluation strategy, including two hidden confounders (H_0 , H_1). c) An example of a five-variable random linear system with two hidden confounders (H_0 , H_1) for the S_3 evaluation strategy. d,e) Nonlinear counterparts of (b,c), respectively. For ease of reading, we present only examples with a maximum timelag of 3, 8 observable variables and 2 hidden confounders. However, in the S_1 , S_2 , S_4 evaluation strategies, the maximum number of variables was 12, and the maximum number of hidden confounders was 3.

each metric, both describing CAnDOIT's performance: CAnDOIT_mean, which depicts the average for the analyzed metric across all interventions, and CAnDOIT_best, which shows the result that produced the most accurate (highest F_1 -score) causal model.

The S_1 evaluation strategy analyzes how CAnDOIT behaves with linear systems, where the number of observable variables ranges from 5 to 12 and there are no latent confounders. S_2 extends S_1 by introducing a random number of latent confounders (from 1 to 3) that affect a specific number of observable variables. In both S_1 and S_2 , we perform a single intervention following the strategy explained above. In the S_3 evaluation strategy, we analyze how CAnDOIT performs with an increasing number of interventions while keeping the number of observable variables fixed. We focus again on linear systems, this time composed of five observable variables plus a random number of hidden confounders (from 1 to 3). The number of interventions ranges from 1 to 3. As mentioned earlier, the S_4 and S_5 evaluation strategies are the nonlinear counterparts of S_2 and S_3 , respectively.

The two algorithms in the comparison, LPCMCI and CAnDOIT, used observational time-series data. For CAnDOIT, in case of S_1 , S_2 , S_4 , we also conducted a single intervention on the variable-originating ambiguous links, while for S_3 and S_5 , we conducted up to three interventions. To manage the computational cost and ensure a fair comparison, both algorithms receive the same amount of data. Specifically, LPCMCI used 1300 observational data samples, while CAnDOIT used 1000 samples of observational data and 300 samples of interventional data. The 300 samples of interventional data were equally divided among the number of interventions. Moreover, the minimum and the maximum timelags considered in the causal analysis were set to 0 and 3, respectively. Finally, the link density parameter, which corresponds to the number of observable parent per variable, was set to 3.

Inspired by the evaluation analyses done in the literature for PCMCI, PCMCI⁺, and LPCMCI,^[19,31,34] we used similar noise configurations (uniform, Gaussian, Weibull) and the same value range for the coefficients that interconnect different terms within the equation, that is, [0.1, 0.5]. This choice of parameters allows to create random systems without encountering divergences or singularities, which commonly arise when attempting to generate time-series equations, especially when a broad range of parameter values and autocorrelation terms is considered. Moreover, differently from PCMCI⁺ and LPCMCI's evaluation strategies, which used a predetermined nonlinear functional form, we employed several functional forms, both linear and nonlinear, randomly chosen. This approach increased the sparsity and diversified the experiments to explore a broader range of scenarios. The complete set of parameters used in our evaluation strategies is listed and discussed in detail in Table S1 and S2 of Appendix A, Supporting Information.

4.3. Evaluation Metrics

The evaluation metrics were categorized into two main groups. The first one comprises metrics to assess CAnDOIT's ability in removing ambiguous links and includes mean false positive rate (FPR), mean number of ambiguous links (uncertainty), and mean number of equivalent MAGs that the PAG output by

the algorithms can represent (PAG size) across all tests. The second category instead evaluates CAnDOIT's overall performance in recovering the correct causal model and its execution time. These last metrics include mean structural hamming distance (SHD), mean F_1 -score, and mean execution time across all tests. All the means were computed based on 25 test runs with different random systems for each configuration. Note that for the FPR, SHD, and F_1 -score metrics, we adopted the same approach used in LPCMCI's evaluation metrics.^[19] Specifically, we not only considered the existence of a link between two nodes but also the tail and head markers of the links. Thus, in calculating these metrics, we measured both the adjacency (presence of a specific link) and the orientation (tail and head markers of a specific link).

For sake of clarity, since less is known in the literature, we explain uncertainty and PAG size metrics a little more in detail. The uncertainty metric quantifies the number of ambiguous links present in the discovered causal model ($\circ \rightarrow$ and $\circ \circ$). This metric is utilized to assess CAnDOIT's capability in removing ambiguities. Consequently, a lower value indicates better performance. The PAG size metric quantifies the number of MAGs equivalent to the discovered one. It calculates this number based on the count of ambiguous links in the estimated causal model. As already explained in Section 3.4, each ambiguous link can represent two possible links: $\circ \rightarrow$ can represent either \rightarrow or \leftrightarrow , while $\circ \circ$ can represent either \rightarrow or \leftarrow . Consequently, for each ambiguous link present in the final causal model, two equivalent MAGs are generated. Finally, the PAG size metric is defined as follows.

$$\text{PAG Size} = 2^{\text{Uncertainty}} \quad (7)$$

This metric is employed to evaluate CAnDOIT's effectiveness in enhancing identifiability by using interventional data.

4.4. Results on Synthetic Models

Figure 4–8 compare the causal discovery results obtained with LPCMCI (red dotted lines), CAnDOIT_mean (green dashed lines), and CAnDOIT_best (blue lines). The different markers in the graphs represent the mean scores computed across 25 run tests for each configuration, while the error bars show the confidence levels determined by 1000 bootstraps over the 25 results. Additionally, a linear mixed model (LMM) was used to assess the statistical validity and robustness of the analysis,^[44] detailed results of which are provided by Table S3 and S4 in Appendix B, Supporting Information.

The S_1 , S_2 , S_3 , S_4 , and S_5 analyses are presented in Figure 4–8, respectively, demonstrating the superior performance of CAnDOIT compared to LPCMCI across all scenarios. Note that, for all the analyses, the PAG size score is shown on a logarithmic scale to improve readability. In S_1 (Figure 4), CAnDOIT (both CAnDOIT_mean and CAnDOIT_best) performs remarkably well in all metrics. Figure 4a,c,e shows FPR, uncertainty, and PAG size, respectively. It consistently maintains lower FPR and uncertainty scores compared to LPCMCI, leading to improved identifiability across all S_1 cases, largely due to its use of interventional data. The superiority of CAnDOIT is evident also in the SHD and F_1 -score comparisons shown in Figure 4b,d, respectively, where CAnDOIT continues to

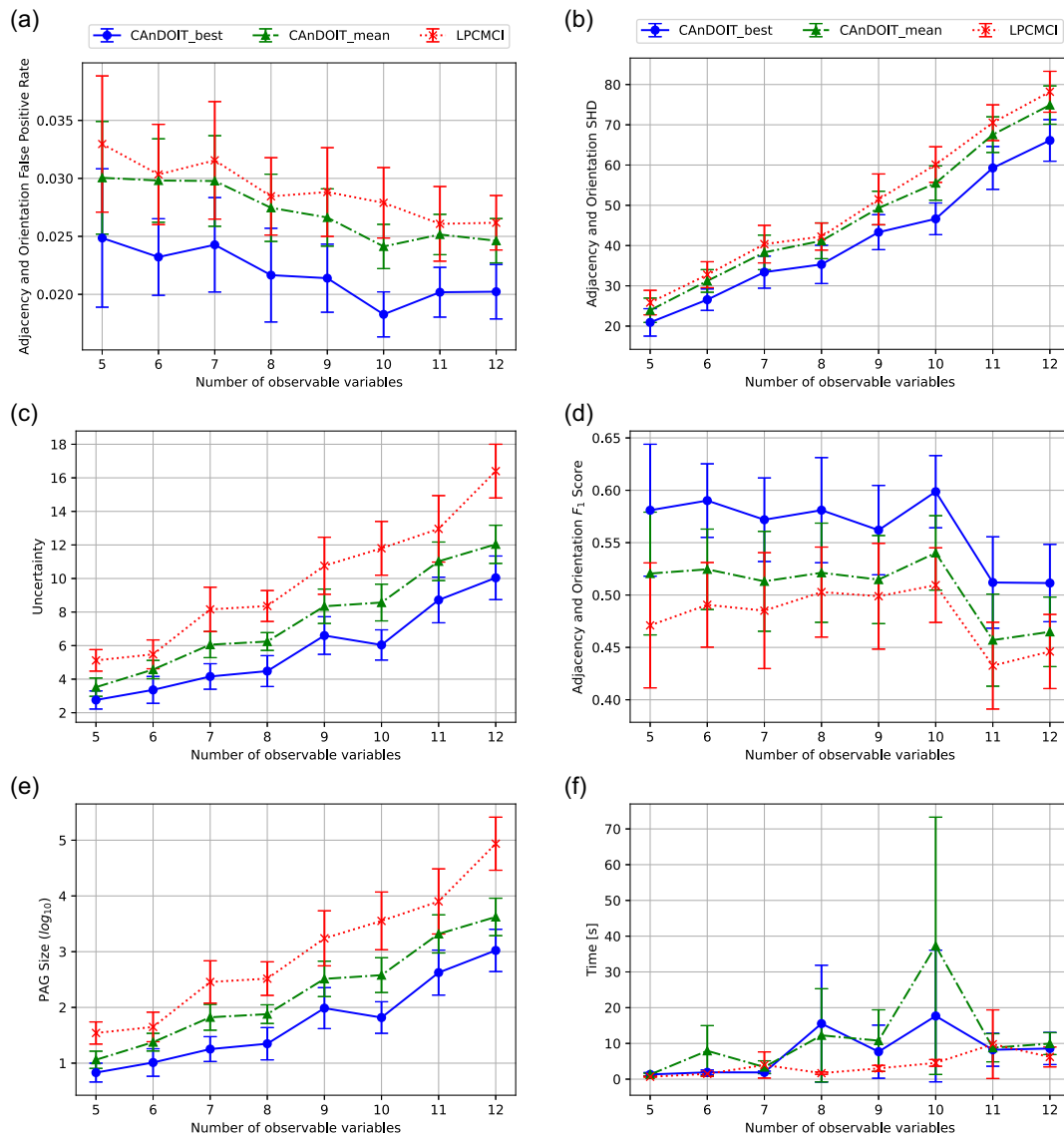


Figure 4. LPCMCI (red dotted line), CAnDOIT_mean (green dashed line), and CAnDOIT_best (blue) in S_1 analysis: linear systems with a number of observable variables ranging from 5 to 12 and no hidden confounders. a) FPR; b) SHD; c) uncertainty; d) F_1 -score; e) PAG size (reported in logarithmic scale); f) time (expressed in seconds).

consistently outperform LPCMCI. In terms of execution time, depicted in Figure 4f, CAnDOIT appears slower compared to LPCMCI due to the introduction of the context node used to model the single intervention. The decreasing trend in the FPR score observed in Figure 4a is associated with the increasing number of variables. This decrease is a result of the higher true negative score, which grows with the number of variables. Furthermore, it is important to specify that the different performances between CAnDOIT_mean and CAnDOIT_best are due to the fact that CAnDOIT_best (the CAnDOIT result with the single intervention yielding the highest F_1 -score) is often associated with an intervention on a variable that is the origin of multiple ambiguous links. Intervening on such a variable is more likely to yield better performance compared to intervening on a variable that originates only a single ambiguous link. On the other hand,

CAnDOIT_mean considers all conducted interventions and averages the results.

In the S_2 analysis (Figure 5), which is analogous to S_1 but includes a random number of hidden confounders, CAnDOIT continues to outperform LPCMCI. As shown in Figure 5a,c, CAnDOIT consistently maintains low FPR and uncertainty scores, surpassing LPCMCI. The S_2 analysis reaffirms the benefits of using interventional data to enhance the identifiability of the causal graph. In particular, the PAG size metric in Figure 5e shows how CAnDOIT, utilizing interventional data, is able to keep the number of equivalent MAGs consistently lower than LPCMCI. Unlike in S_1 , the difference between CAnDOIT and LPCMCI is less pronounced in the SHD score, as shown in Figure 5b, though the advantage of using interventional data remains evident in the F_1 -score (Figure 5d). Similar to S_1 , the

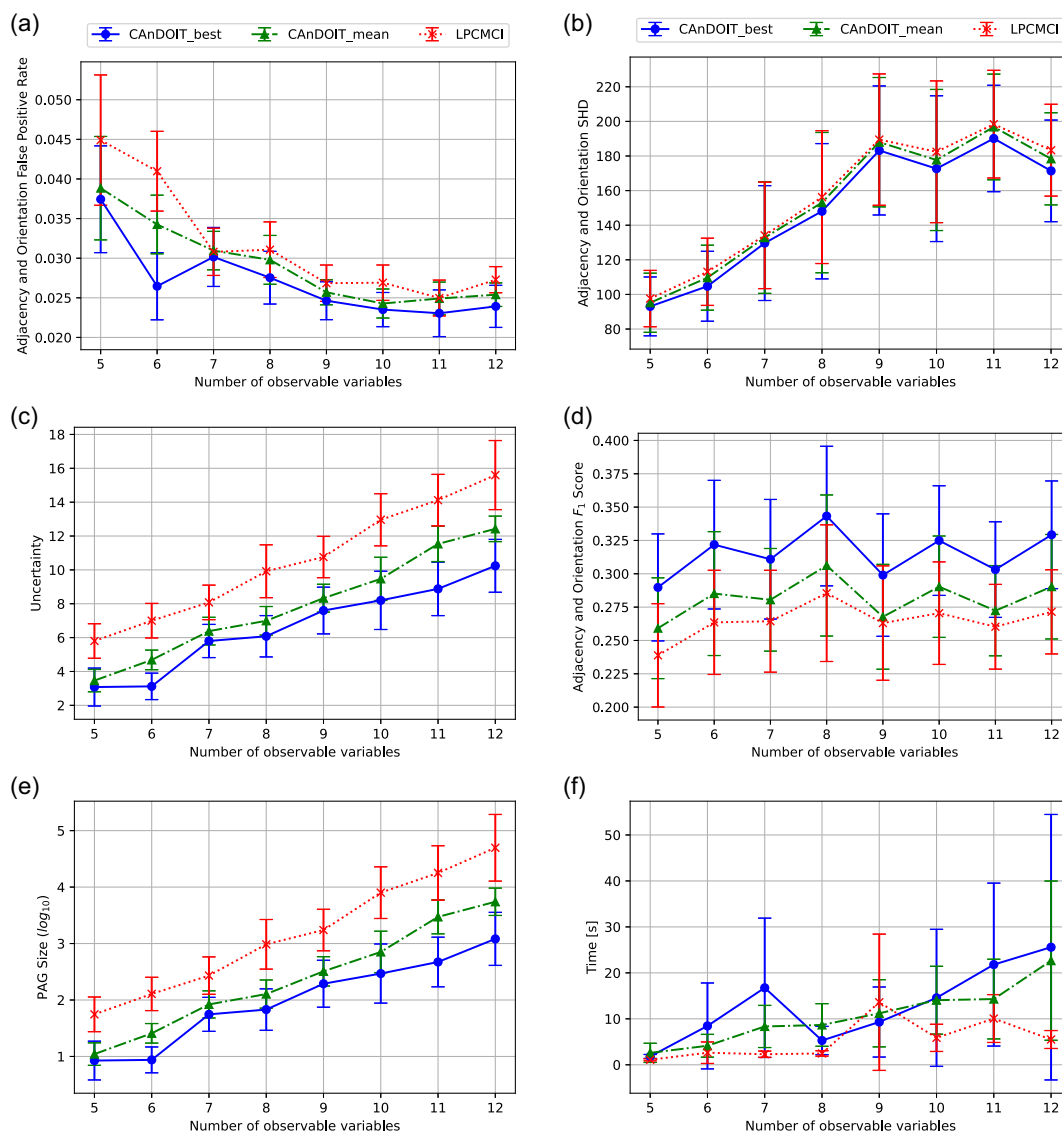


Figure 5. LPCMCI (red dotted line), CAnDOIT_mean (green dashed line), and CAnDOIT_best (blue) in S_2 analysis: linear systems with a number of observable variables ranging from 5 to 12 and a random number of hidden confounders (from 1 to 3). a) FPR; b) SHD; c) uncertainty; d) F_1 -score; e) PAG size (reported in logarithmic scale); f) time (expressed in seconds).

employment of context variables to model interventions makes CAnDOIT slower compared to LPCMCI, as shown in Figure 5f.

In the S_3 analysis (Figure 6), we fixed the number of observable variables and varied the number of interventions to evaluate how our algorithm handles multiple interventions and whether they contribute to better performance compared to the single intervention cases analyzed in S_1 and S_2 . CAnDOIT continues to significantly outperform LPCMCI. As shown in Figure 6a,c, e, CAnDOIT yields lower FPR, uncertainty, and PAG size scores than LPCMCI, demonstrating how an increasing number of interventions helps reduce the number of uncertain links and, consequently, the PAG size. Figure 6b,d shows an improvement in the SHD and F_1 scores when transitioning from one to two interventions, though there is a slight deterioration in

performance with three interventions. This can be attributed to the number of samples associated with each intervention. As explained in Section 4.2, to ensure a fair comparison, LPCMCI and CAnDOIT always use the same amount of data. In the case of multiple interventions, the 300 interventional samples are equally divided among the number of interventions, meaning that with three interventions, each has 100 samples. This reduced sample size for each intervention is not sufficient to further improve the SHD and F_1 scores compared to the two interventions cases, explaining the deterioration in these metrics in the three interventions case. As in previous analyses, CAnDOIT appears slower compared to LPCMCI, as shown in Figure 6f.

The S_4 analysis (Figure 7) is the nonlinear counterpart to S_2 . Specifically, it analyzes nonlinear systems with a number of

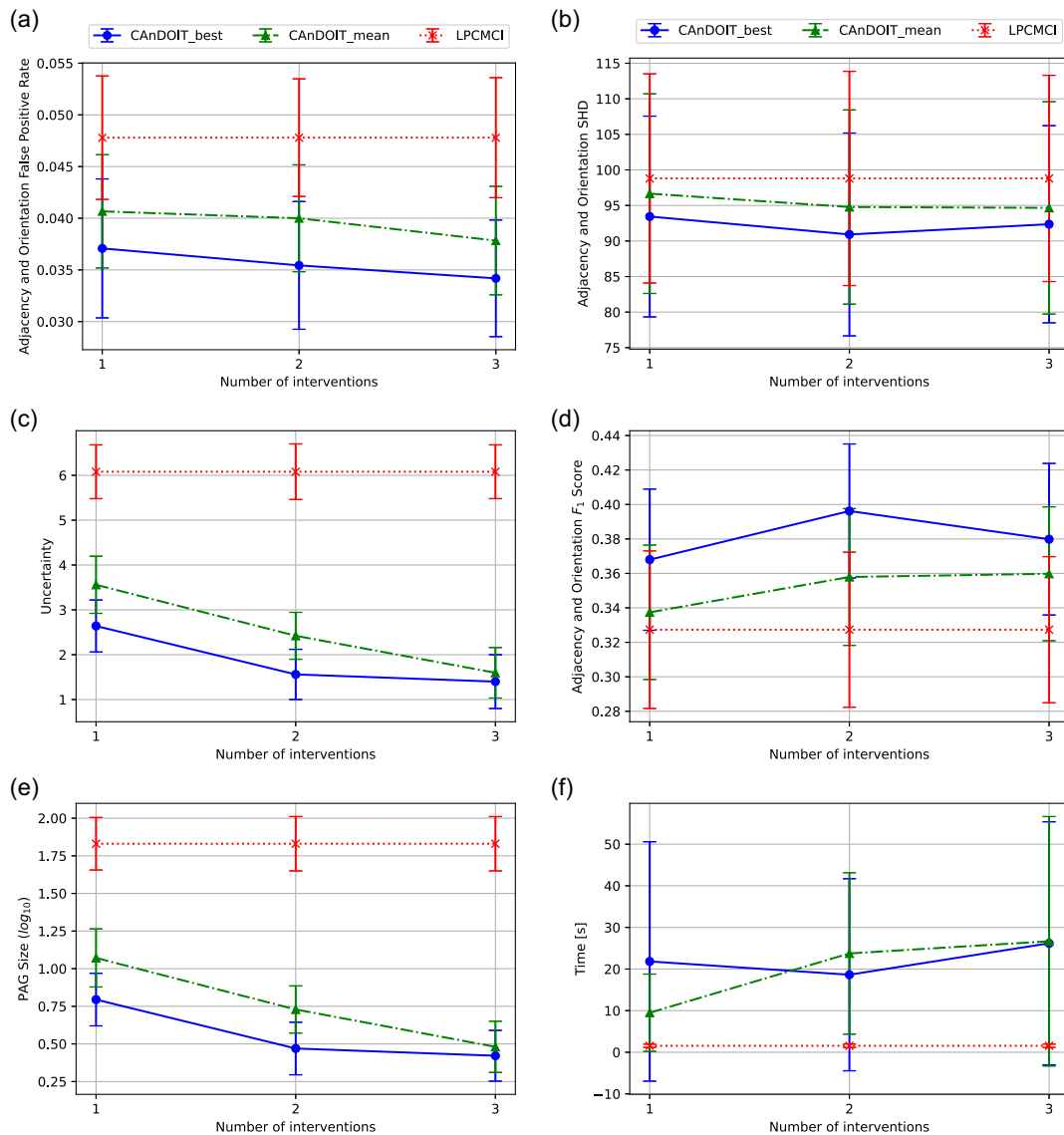


Figure 6. LPCMCI (red dotted line), CAnDOIT_mean (green dashed line), and CAnDOIT_best (blue) in S_3 analysis: linear systems with five observable variables, a random number of hidden confounders (from 1 to 3) and an increasing number of interventions, ranging from 1 to 3. a) FPR; b) SHD; c) uncertainty; d) F_1 -score; e) PAG size (reported in logarithmic scale); f) time (expressed in seconds).

observable variables ranging from 5 to 12 and a random number of hidden confounders (from 1 to 3). Like S_2 , this analysis helps to understand how CAnDOIT scales with an increasing number of variables in the system, but in this case with nonlinear relationships. In this analysis, CAnDOIT's FPR and SHD scores, depicted in Figure 7a,b respectively, remain comparable to LPCMCI. However, the benefit of using interventional data is evident in the uncertainty and consequently the PAG size, as shown in Figure 7c,e respectively. Additionally, in terms of F_1 -score (Figure 7d), CAnDOIT consistently outperforms LPCMCI, with both CAnDOIT_mean and CAnDOIT_best. The difference in terms of execution time between LPCMCI and CAnDOIT (Figure 7f) is more pronounced for the nonlinear case. This is due to CI test used in this case, namely Gaussian process and distance correlation, which is slower compared to the

partial correlation test used in the linear case. In CAnDOIT, the introduction of context variables to model interventions further accentuates this aspect.

Finally, in the S_5 analysis (Figure 8), similar to S_3 , we fixed the number of observable variables and varied the number of interventions to assess CAnDOIT's ability to handle multiple interventions and their contribution to performance enhancement compared to a single intervention. In this case, the FPR scores for LPCMCI and CAnDOIT are comparable across the three scenarios, as shown in Figure 8a. However, using interventional data, CAnDOIT outperforms LPCMCI in the remaining metrics, demonstrating its capability to remove ambiguities: it achieves lower uncertainty and consequently lower PAG size scores (Figure 8c,e). Performing multiple interventions results in improvements in CAnDOIT's SHD and F_1 scores (Figure 8b,d)

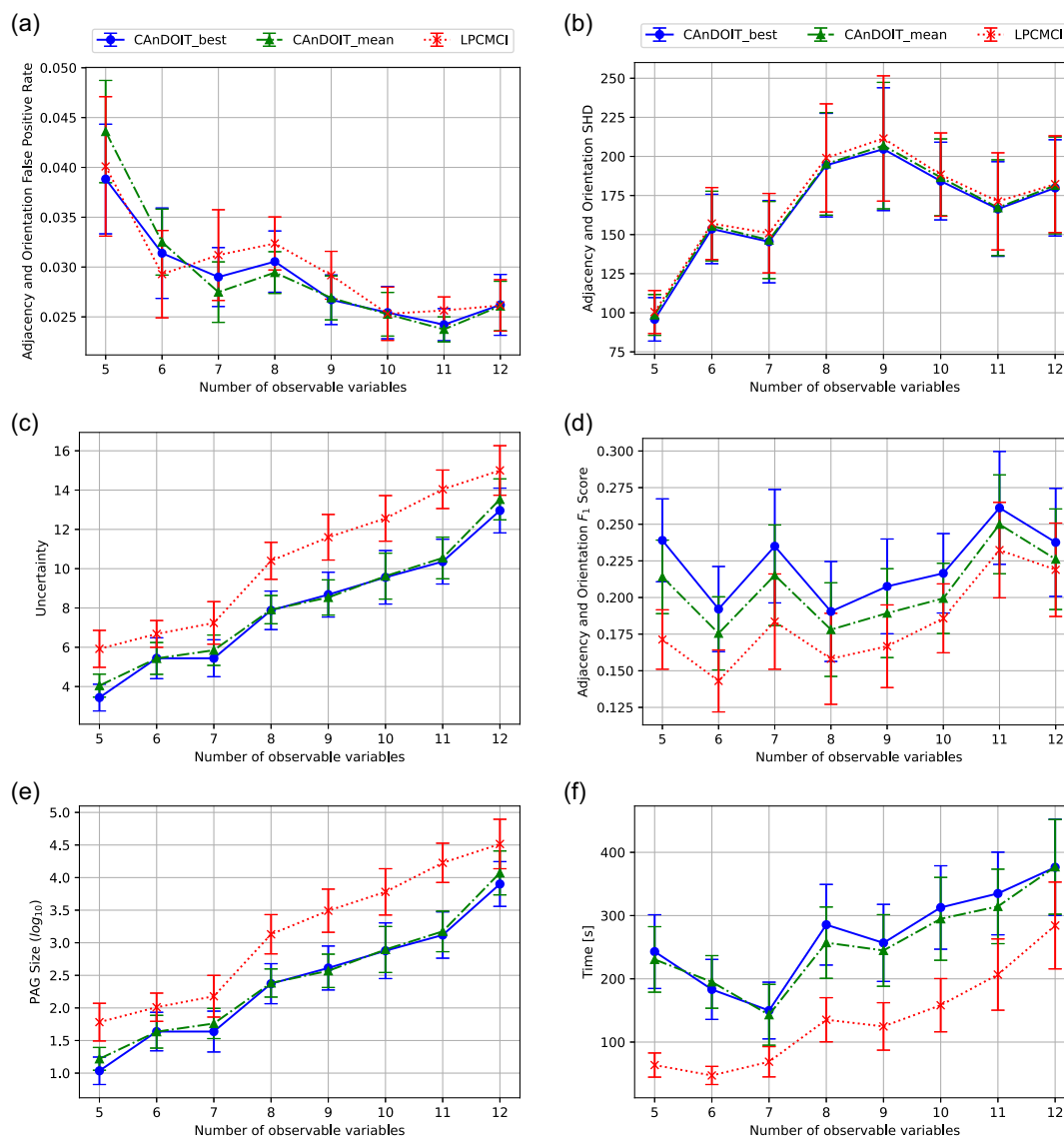


Figure 7. LPCMCI (red dotted line), CAnDOIT_mean (green dashed line), and CAnDOIT_best (blue) in S_4 analysis: nonlinear systems with a number of observable variables ranging from 5 to 12 and a random number of hidden confounders (from 1 to 3). a) FPR; b) SHD; c) uncertainty; d) F_1 -score; e) PAG size (reported in logarithmic scale); f) time (expressed in seconds).

when transitioning from one to two interventions. However, as seen in S_3 , CAnDOIT is not able to further enhance its performance with three interventions. In this case, CAnDOIT's execution time is longer compared to LPCMCI, and Figure 8f shows that this increase in time directly depends on the number of interventions.

It is important to highlight the significance of these analyses. Although in all these evaluation strategies, CAnDOIT utilized a subset of observational data alongside interventional data specifically aimed at testing links affected by ambiguities, this still yielded a consistent improvement in overall accuracy of the causal analysis. This improvement is evident in Figure 4d, 5d, 6d, 7d, and 8d, which represent the F_1 -scores for S_1 , S_2 , S_3 , S_4 , and S_5 , respectively. Additionally, there is a significant enhancement in the identifiability of the causal graph, as shown

in Figure 4e, 5e, 6e, 7e, and 8e, which illustrate the PAG size. These positive results are even more noteworthy considering that CAnDOIT operates with 300 fewer observational data samples compared to LPCMCI.

Overall, these evaluation strategies demonstrate CAnDOIT's effectiveness in handling interventional data and its superiority over the current state-of-the-art LPCMCI causal discovery algorithm in producing more accurate causal models, decreasing the uncertainty and the PAG size. The LMM analysis confirmed the statistical significance of our evaluation strategies, as indicated by the F-statistics and p -values for all scores shown in Table S3 and S4 of Appendix B, Supporting Information. Moreover, the table provides further useful statistical information, confirming the importance of the algorithm choice (LPCMCI or CAnDOIT).

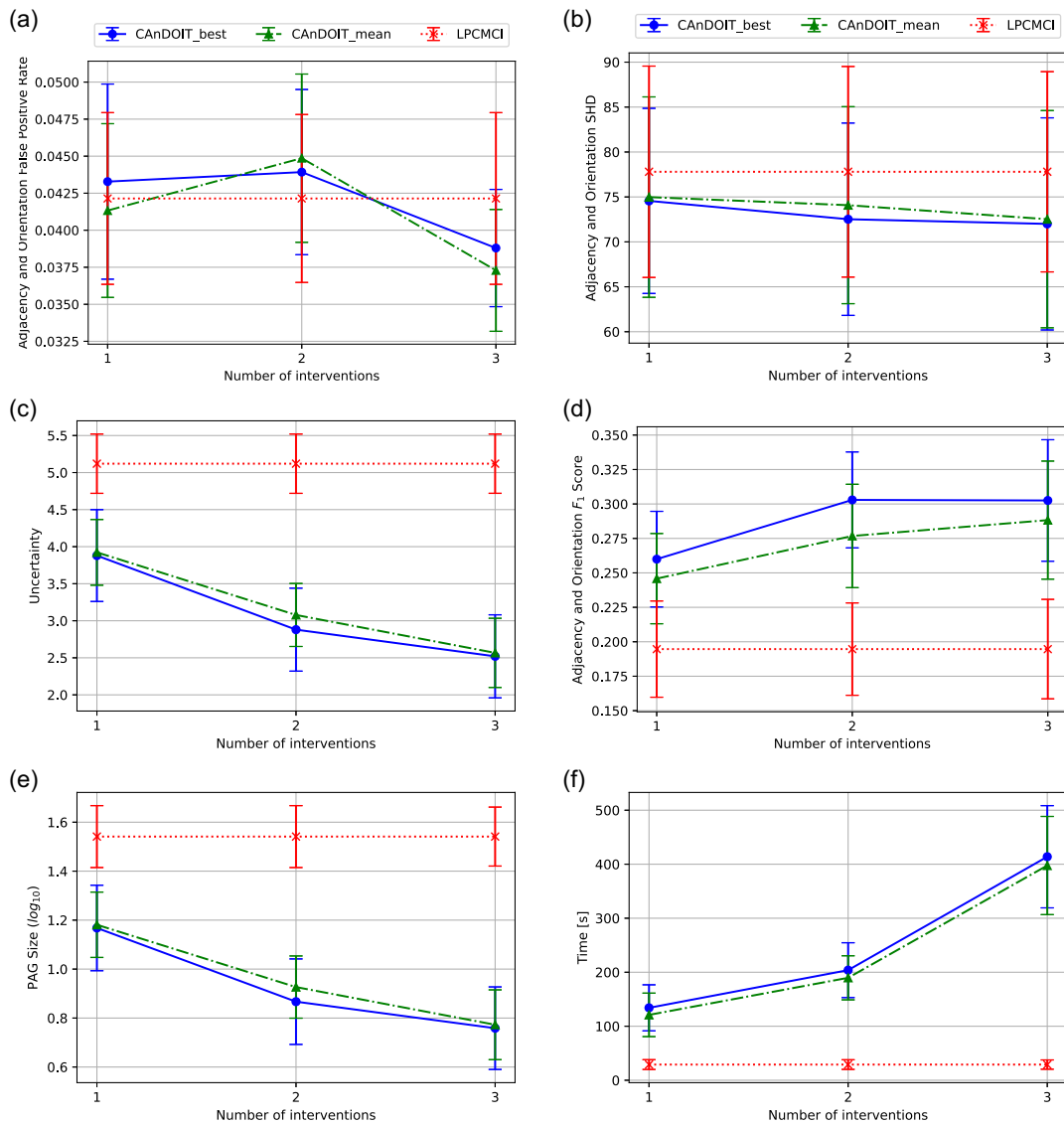


Figure 8. LPCMCI (red dotted line), CAnDOIT_mean (green dashed line), and CAnDOIT_best (blue) in S_5 analysis: nonlinear systems with five observable variables, a random number of hidden confounders (from 1 to 3) and an increasing number of interventions, ranging from 1 to 3. a) FPR; b) SHD; c) uncertainty; d) F_1 -score; e) PAG size (reported in logarithmic scale); f) time (expressed in seconds).

5. Evaluation on Robotic Scenario

Once established, through the evaluation strategies presented in Section 4, that our approach works correctly, we used it for modeling a robotic scenario in a simulated environment. Our strategy was first to extract time-series data from the simulator and then use it for causal discovery in the presence of a hidden confounder.

5.1. Causal World for Robot Camera Modeling

We designed an experiment to learn the causal model in a hypothetical robot arm application equipped with a camera. Our focus was on estimating the causal relationship between the color's brightness of objects as captured by the camera and various

factors, including camera-to-object distance. For this evaluation, we utilized the well-known benchmark Causal World,^[45] which is designed for causal structure learning in a robotic manipulation environment. The environment consists of a TriFinger robot, a floor, and a stage. It allows for the inclusion of objects with various shapes, for example, cubes. This simulator is widely used in the causality community due to its ability to support diverse manipulation tasks and interventions,^[46–48] including changing the objects' color or mass.

For simplicity, we focused on a specific scenario using only one finger of the robot, where the finger's end-effector was equipped with a camera. The scenario (shown in **Figure 9**) consists of a cube placed at the center of the floor, surrounded by a white stage. The color's brightness (b) of the cube and the floor is modeled as follows.

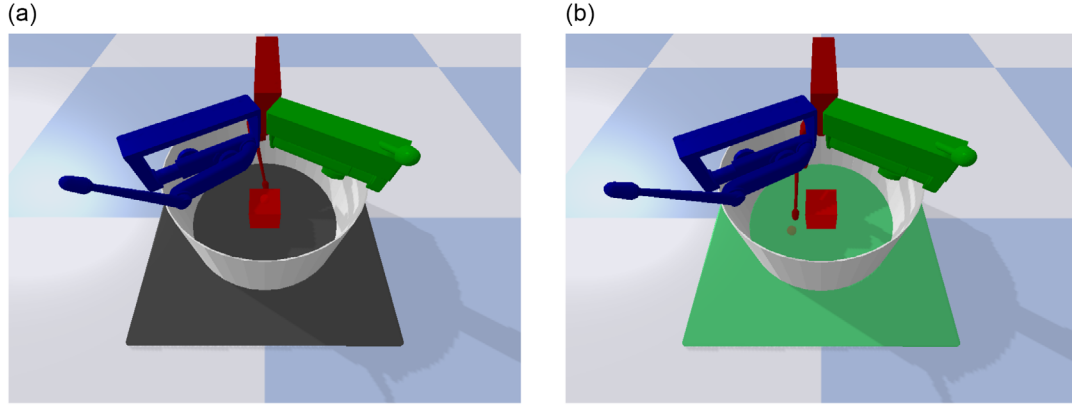


Figure 9. Causal World:^[45] a robotic manipulation environment. a) observational experiment; b) experiment with intervention performed on the floor's color.

$$b = K_h \frac{H}{H_{\max}} + K_v \left(1 - \frac{v}{v_{\max}}\right) + K_d \frac{d_c}{d_{c_{\max}}} \quad (8)$$

where H is the end-effector height, v its absolute velocity, and d_c the distance between the end-effector and the cube. K_h , k_v , k_d are the gains associated to each factors, while H_{\max} , v_{\max} , and $d_{c_{\max}}$ are the maximum values for H , v , and d_c , respectively. This model captures the shading and blurring effects on the cube due to the height of the end-effector, its velocity, and its distance from the cube. On the other hand, the floor, being darker and larger than the cube, is only affected by the end-effector's height.

The data collected from the scenario therefore includes the floor (F_c) and the cube (C_c) colors, as well as the height (H), the absolute velocity (v) of the end-effector, and its distance from the cube (d_c). The ground-truth SCM for the variables F_c and C_c is expressed as follows.

$$\begin{cases} F_c(t) = b(H(t-1)) \\ C_c(t) = b(H(t-1), v(t-1), d_c(t-1)) \end{cases} \quad (9)$$

Note that H , v , and d_c are obtained directly from the simulator and not explicitly modeled.

5.2. Results on Robotic Scenario

The evaluation involved three main steps. 1) We generated observational data containing all the variables in the system (F_c , C_c , H , v , d_c), as shown in Figure 9a, and performed the causal analysis using LPCMCI. 2) We intentionally hid the variable H , representing the height of the end-effector, to create a hidden confounder and a spurious relationship between C_c and F_c . Again, we used LPCMCI for the causal analysis. 3) We conducted an intervention on the floor's color, setting it to green (Figure 9b), and collected data from the simulator. Then we used CANDOIT for the causal analysis with both observational and interventional, accounting for the hidden confounder H . The observational time series had a length of 600 samples, while the interventional time series consisted of 125 samples. Both were recorded at a sampling rate of 10 Hz. Also in this case, to ensure a fair analysis, LPCMCI and CANDOIT used exactly the same amount of data.

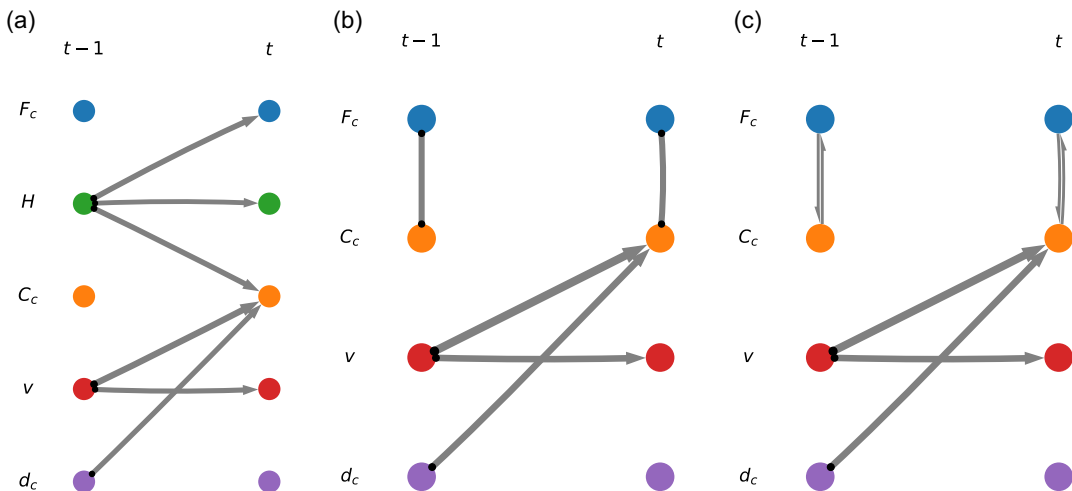


Figure 10. Causal model of the robot camera in Causal World. a) LPCMCI's result with all the variables being observable; b) LPCMCI's result with hidden H ; c) CANDOIT's causal model with hidden H .

Consequently, LPCMCI received the complete set of observational data, whereas for CAnDOIT part, the observational data was replaced by interventions, specifically 475 observational samples and 125 interventional ones.

Figure 10 shows the results for each specific step: (10a) causal model using LPCMCI with observable variables only; (10b) LPCMCI's result with hidden H . Figure 10c shows the causal model retrieved by CAnDOIT that, by leveraging both observational and interventional data, successfully identify the bidirected relation between C_c and F_c , representing the presence of a latent confounder (H).

Also in this experiment, we can see the benefit of using intervention data alongside the observations. In Figure 10b, LPCMCI is not able to orient the contemporaneous (spurious) link between F_c and C_c due to the hidden confounder H . This yields the ambiguous link $\circ\leftrightarrow$, which does not encode the correct link \leftrightarrow (the $\circ\leftrightarrow$ represents either \rightarrow or \leftarrow). Instead CAnDOIT, using interventional data, correctly identifies the bidirected link $F_c \leftrightarrow C_c$, decreasing once again the uncertainty level and increasing the accuracy of the reconstructed causal model.

6. Conclusions and Future Work

In this article, we proposed CAnDOIT, a new state-of-the-art algorithm that enables causal discovery using both observational and interventional data via context variables. We validated our approach experimentally on random synthetic models and tested on a robotic simulator for causal discovery, focusing on the significance of interventional data. Our results confirmed that CAnDOIT significantly improves previous causal discovery methods, offering enhanced accuracy, model identifiability. They also highlight its capability to handle interventional data effectively and its potential for real-world robot applications. The proposed method lays the foundation for new observations- and interventions-based causal discovery methods on time-series data, with numerous opportunities for future research. Firstly, to the best of our knowledge, various causal discovery algorithms combine observational data and hard/soft interventions with known/UT, but only in the static domain. We present, for the first time, a causal discovery algorithm that combines observational and interventional data in a time-series setting. We acknowledge that the current version of CAnDOIT only deals with hard interventions on known targets. However, this challenge has not yet been investigated in the time-series domain. Furthermore, this solution is useful in many applications where intervention variables are known (e.g., robotics). We believe that CAnDOIT is a significant step in the right direction and future extensions could accommodate soft interventions and UT.

An interesting analysis would be to test how the ratio between the lengths of observational and interventional time-series data influences the algorithm's performance. Since our application targets are intelligent robot applications, particularly in industrial and intralogistics settings,^[49] it would be helpful to determining the optimal amount of observational and interventional data required to establish an accurate causal model, within a reasonable time.

Scalability is another aspect to be further investigated. In this article, we partially explored the scalability of CAnDOIT through

the S_1 , S_2 , S_4 analyses, studying how the algorithm behaves with an increasing number of variables. In the future we would like to extend this analysis to check also how CAnDOIT scales with varying lengths of observational and interventional time-series data.

Finally, inspired by recent advances in the literature,^[50] we plan to enhance CAnDOIT by automating the selection of intervention variables and optimizing the causal model under constraints. Also, since the current version of the algorithm first processes observational data, and then interventional one, a future improvement would be to inject opportune interventions between observation sequences. Such a functionality would help CAnDOIT better adapt to real-world applications, where a robot observes the scenario, intervenes, and then observes again.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101017274 (DARKO).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Luca Castri: Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing—original draft (lead). **Sariah Mghames:** Supervision (supporting); Validation (supporting); Visualization (supporting); Writing—review & editing (supporting). **Marc Hanheide:** Funding acquisition (supporting); Project administration (supporting); Resources (supporting); Supervision (supporting). **Nicola Bellotto:** Conceptualization (supporting); Funding acquisition (lead); Project administration (lead); Resources (lead); Supervision (lead); Writing—review & editing (lead).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

causal robotics, observations and interventions-based causal discoveries, time series

Received: March 8, 2024

Revised: August 9, 2024

Published online:

- [1] J. Pearl, *Causality*, Cambridge University Press, New York, NY **2009**.
- [2] J. Runge, A. Gerhardus, G. Varando, V. Eyring, G. Camps-Valls, *Nat. Rev. Earth Environ.* **2023**, *4*, 487.
- [3] S. Saetia, N. Yoshimura, Y. Koike, *Front. Neuroinf.* **2021**, *15*, 619557.
- [4] T. Hellström, *Paladyn, J. Behav. Rob.* **2021**, *12*, 238.
- [5] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, Y. Bengio, *Proc. IEEE* **2021**, *109*, 612.
- [6] C. Glymour, K. Zhang, P. Spirtes, *Front. Genet.* **2019**, *10*, 524.
- [7] Y. Zhu, N. Zabararas, P.-S. Koutsourelakis, P. Perdikaris, *J. Comput. Phys.* **2019**, *394*, 56.
- [8] M. Raissi, P. Perdikaris, G. E. Karniadakis, *J. Comput. Phys.* **2019**, *378*, 686.
- [9] G. Gössler, O. Sokolsky, J.-B. Stefani, in *2nd Int. Workshop on Causal Reasoning for Embedded and safety-critical Systems Technologies (CREST 2017)*, Vol. 259, Uppsala, Sweden, April **2017**, pp. 47–53.
- [10] A. Sivaram, V. Venkatasubramanian, *AIChE J.* **2022**, *68*, e17687.
- [11] D. B. Rubin, *J. Educ. Psychol.* **1974**, *66*, 688.
- [12] A. Hauser, P. Bühlmann, *J. R. Stat. Soc. B* **2015**, *77*, 291.
- [13] P. Brouillard, S. Lachapelle, A. Lacoste, S. Lacoste-Julien, A. Drouin, *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21865.
- [14] J. M. Mooij, S. Magliacane, T. Claassen, *J. Mach. Learn. Res.* **2020**, *21*, 3919.
- [15] Y. Wang, F. Cao, K. Yu, J. Liang, in *Proc. of the AAAI Conf. on Artificial Intelligence*, (Online) February **2022**, Vol. 36, pp. 8584–8593.
- [16] M. Kocaoglu, A. Jaber, K. Shanmugam, E. Bareinboim, *Adv. Neural Inf. Proc. Syst.* **2019**, *32*, 14369.
- [17] A. Jaber, M. Kocaoglu, K. Shanmugam, E. Bareinboim, *Adv. Neural Inf. Proc. Syst.* **2020**, *33*, 9551.
- [18] G. R. A. Faria, A. Martins, M. A. Figueiredo, in *Conf. on Causal Learning and Reasoning*, PMLR, April **2022**, pp. 253–274.
- [19] A. Gerhardus, J. Runge, *Adv. Neural Inf. Proc. Syst.* **2020**, *33*, 12615.
- [20] J. Pearl, M. Glymour, N. P. Jewell, *Causal Inference in Statistics: A Primer*, John Wiley & Sons, Chichester, West Sussex, United Kingdom **2016**.
- [21] P. Spirtes, C. N. Glymour, R. Scheines, *Causation, Prediction, and Search*, MIT press, Cambridge, MA **2000**.
- [22] D. M. Chickering, *J. Mach. Learn. Res.* **2003**, *3*, 507.
- [23] X. Zheng, B. Aragam, P. K. Ravikumar, E. P. Xing, *Adv. Neural Inf. Proc. Syst.* **2018**, *31*, 9472.
- [24] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, M. Jordan, *J. Mach. Learn. Res.* **2006**, *7*, 10.
- [25] C. K. Assaad, E. Devijver, E. Gaussier, *J. Artif. Intell. Res.* **2022**, *73*, 767.
- [26] M. Nauta, D. Bucur, C. Seifert, *Mach. Learn. Knowle. Extr.* **2019**, *1*, 19.
- [27] J. Peters, D. Janzing, B. Schölkopf, *Adv. Neural Inf. Proc. Syst.* **2013**, *26*, 154.
- [28] A. Hyvärinen, K. Zhang, S. Shimizu, P. O. Hoyer, *J. Mach. Learn. Res.* **2010**, *11*, 5.
- [29] R. Pamfil, N. Sriwattanaworachai, S. Desai, P. Pilgerstorfer, K. Georgatzis, P. Beaumont, B. Aragam, in *Int. Conf. on Artificial Intelligence and Statistics* (Online), PMLR, August **2020**, pp. 1595–1605.
- [30] D. Entner, P. O. Hoyer, in *Proc. of the 5th European Workshop on Probabilistic Graphical Models*, Helsinki Institute for Information Technology HIIT, Finland **2010**, *16*, pp. 121–128.
- [31] J. Runge, *Chaos* **2018**, *28*, 7.
- [32] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, D. Sejdinovic, *Sci. Adv.* **2019**, *5*, eaau4996.
- [33] L. Castri, S. Mghames, M. Hanheide, N. Bellotto, in *Int. Conf. on Social Robotics*, Springer, Florence, Italy, December **2022**, pp. 154–164.
- [34] J. Runge, in *Conf. on Uncertainty in Artificial Intelligence* (Online), PMLR, August **2020**, pp. 1388–1397.
- [35] L. Castri, S. Mghames, M. Hanheide, N. Bellotto, in *2nd Conf. on Causal Learning and Reasoning*, Tübingen, Germany, April **2023**.
- [36] W. Günther, U. Ninad, J. Runge, in *Proc. of the Thirty-Ninth Conf. on Uncertainty in Artificial Intelligence*, volume 216 of *Proc. of Machine Learning Research* (Eds: R. J. Evans, I. Shpitser), PMLR, Pittsburgh, PA July **2023**, pp. 766–776 <https://proceedings.mlr.press/v216/gunther23a.html>
- [37] S. Gao, R. Addanki, T. Yu, R. A. Rossi, M. Kocaoglu, in *Thirty-Seventh Conf. on Neural Information Processing Systems*, December **2023**.
- [38] K. Yang, A. Katcoff, C. Uhler, in *Int. Conf. on Machine Learning*, PMLR, Stockholm, Sweden, July **2018**, pp. 5541–5550.
- [39] C. Squires, Y. Wang, C. Uhler, in *Conf. on Uncertainty in Artificial Intelligence* (Online), PMLR August **2020**, pp. 1039–1048.
- [40] P. Lippe, T. Cohen, E. Gavves, in *Int. Conf. on Learning Representations*, (Online) May **2021**.
- [41] J. Pearl, *Stat. Sci.* **1993**, *8*, 266.
- [42] J. Peters, D. Janzing, B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*, The MIT Press, Cambridge, MA, **2017**.
- [43] A. R. Lawrence, M. Kaiser, R. Sampaio, M. Sipos, in *Causal Discovery and Causality-Inspired Machine Learning Workshop at Neural Information Processing Systems 2020*.
- [44] J. Jiang, T. Nguyen, *Linear and Generalized Linear Mixed Models and their Applications*, Vol. 1, Springer, New York, NY **2007**.
- [45] O. Ahmed, F. Träuble, A. Goyal, A. Neitz, M. Wüthrich, Y. Bengio, B. Schölkopf, S. Bauer, in *Int. Conf. on Learning Representations*, (Online) May **2021**.
- [46] S. Lachapelle, P. Rodriguez, Y. Sharma, K. E. Everett, R. Le Priol, A. Lacoste, S. Lacoste-Julien, in *Conf. on Causal Learning and Reasoning*, PMLR, Eureka, CA, April **2022**, pp. 428–484.
- [47] W. Yao, Y. Sun, A. Ho, C. Sun, K. Zhang, in *Int. Conf. on Learning Representations*, May **2021**.
- [48] P. Lippe, S. Magliacane, S. Löwe, Y. M. Asano, T. Cohen, E. Gavves, in *The 39th Conf. on Uncertainty in Artificial Intelligence*, Pittsburgh, PA, July **2023**, <https://openreview.net/forum?id=VS7Dn31xuB>.
- [49] S. Ghidoni, M. Terreran, D. Evangelista, E. Menegatti, C. Eitzinger, E. Villagrossi, N. Pedrocchi, N. Castaman, M. Malecha, S. Mghames, L. Castri, M. Hanheide, N. Bellotto, in *Ital-IA Convegno Nazionale sull'Intelligenza Artificiale*, (Online) February **2022**, <https://www.ital-ia2022.it/>.
- [50] V. Aglietti, A. Malek, I. Ktena, S. Chiappa, in *Int. Conf. on Machine Learning*, Honolulu, HI, July **2023**.