Causal Knowledge in Data Fusion: Systematic Evaluation on Quality Prediction and Root Cause Analysis

Jingyi Yu[1,2], Tim Pychynski[1], Karim Said Barsim[1], Marco F. Huber[2,3]
[1]Bosch Center for Artificial Intelligence, Renningen, Germany
[2]Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, Stuttgart, Germany
[3]Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Stuttgart, Germany
Email: {jingyi.yu, tim.pychynski, karim.barsim}@de.bosch.com, marco.huber@ieee.org

***Abstract*—Data fusion deals with combining information from multiple sensors to support decision making. In such settings, machine learning methods, that principally only take correlation into account, have been applied widely due to their strong predictive and computational capabilities. In this paper, we investigate potential benefits of introducing causal knowledge in machine learning-based data fusion to address two common downstream tasks, namely, quality prediction and root cause analysis (RCA). To resemble the complex relationships typically associated with sensor data, we create simulation data with explicit modeling of latent confounding. The results of this study indicate that taking into account true causal knowledge significantly improves the performance of RCA, and leads to prediction models that are more robust to severe distribution shifts in the presence of latent confounding. Furthermore, if causal knowledge needs to be inferred from observational data using existing causal discovery methods, we propose a selection criterion to choose the best causal structure. We show that given a sufficient amount of data, the selected causal structure can be used as reliable input to solve the downstream tasks.**

***Index Terms*—data fusion, causal discovery, latent confounding, quality prediction, root cause analysis (RCA)**

I. Introduction

With the advent of Industry 4.0, the ability to use digitally connected devices and sensors has enabled the collection of a huge amount of data in manufacturing processes [1], providing new opportunities for condition monitoring, optimization and root cause analysis (RCA). To this end, multiple sensors are used to monitor machines, components and the environment, the signals of which need to be combined or fused to support decision making [2]. The objective is usually to deploy as few sensors as possible to minimize costs, leading to numerous unmeasured quantities in manufacturing systems. Unmeasured variables that are common causes of measured variables (a.k.a. *latent confounders*) are known to cause spurious dependencies in the data [3]. Due to the underlying physics, various sensor signals are often related to each other in a convoluted manner, posing a major challenge in sensor selection and refinement of fusion strategies.

A simple example of this challenge is illustrated via the *directed acyclic graph (DAG)* in Fig. 1. The true values of three states X_1, X_2, and X_3 are measured by three sensors S_1, S_2, and S_3, respectively. On the one hand, X_1, X_2, and X_3 are associated due to direct causal relationships. On the other hand, there is an unmeasured common cause L_1, i.e., a latent confounder, which has direct causal effects on measured variables X_2 and X_3. It can be viewed as a causal graph in the sense that the graph structure reflects the underlying cause-and-effect relationships [4].

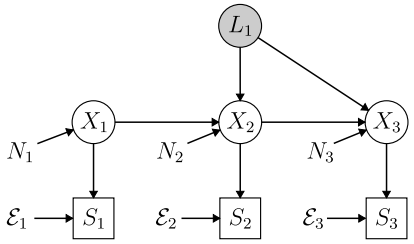


Fig. 1: Causal graph of a three-sensor scenario. White circles, dark circles, and boxes represent measured variables, latent confounders, and sensor measurements, respectively. All other unmeasured causes of each measured variable X_i are represented by an additional noise term N_i. Each sensor is also subject to a measurement error $\mathcal{E}_i$ that is independent of the measured variable X_i.

In this paper, we consider two typical manufacturing tasks, namely *quality prediction* and *RCA*. Assume that X_3 represents a variable that quantifies the quality state of a manufactured product (e.g., a geometric dimension), the accurate measurement of which by sensor S_3 is costly. The product's quality may be affected by other states X_1 and X_2 measured somewhere upstream in the production line. We define the downstream tasks as follows:

1) *Quality prediction:* predict the quality state X_3 based on sensor measurements S_1 and S_2, to reduce the number of costly measurements. This corresponds to a typical soft sensor problem [5].
2) *RCA*: identify the root causes of anomalies at X_3 based on sensor readings from S_1, S_2 and S_3.

There has been a stream of research in manufacturing attempting to address the aforementioned tasks in light of

causation [6]–[10]. In fact, having knowledge of causal relationships rather than statistical associations enables a better understanding of how variables interact with each other. However, due to the complex nature of manufacturing, the full causal knowledge is usually not available. One way of inferring causal relationships among features of interest is to apply *causal discovery* methods. The goal of causal discovery is to establish an estimate of the underlying causal relationships typically from purely observational data. The identified causal relationships are usually represented in a DAG with directed arrows from cause to effect. A common assumption made in causal discovery is, however, the absence of latent confounders, which is also known as causal sufficiency [11].

The aim of this paper is to evaluate the benefits of introducing causal knowledge in machine learning-based data fusion in complex manufacturing systems with the existence of latent confounders. The following research questions will be addressed in the evaluation study:

1) How do data fusion strategies derived from the true causal knowledge affect the performance of quality prediction and RCA?
2) Can causal relationships inferred by causal discovery methods from purely observational data be used as reliable input?

As shown later, our evaluation study is centered around the three-sensor scenario in Fig. 1. The reasons are twofold: first, we know the true causal relationships in such a simulation scenario, while they are often unknown or costly to measure in reality; and second, it is a simple representation of cases where latent confounders can occur during data fusion in real-world scenarios.

In the next section, we formalize the problem statement, followed by the description of our methodology in Section III. Finally, the results of our experiments and conclusions are summarized in Sections IV and V, respectively.

II. Problem Statement

Notation: we use upper-case letters for vectors and bold upper-case letters for matrices.

In causal discovery settings, data are assumed to be generated according to a *functional causal model (FCM)* [12], where each variable is a function of its direct causes and a noise term that represents an aggregate of all other unobserved causes. Motivated by the fact that many concepts and phenomena in causal analysis were first studied in linear FCMs [13] and that the Gaussian distribution has convenient mathematical properties and is usually backed by the central limit theorem, we limit the scope of our study to linear Gaussian FCMs with explicit modeling of latent confounders.

Consider a set of measured variables $X \in \mathbb{R}^m$, latent confounders $L \in \mathbb{R}^p$, and noise variables $N \in \mathbb{R}^m$, the data-generating FCM can be written in a matrix form as

$$X = \mathbf{A}X + \mathbf{B}L + N \tag{1}$$

$$= \mathbf{A}X + \begin{bmatrix}\mathbf{B} & \mathbf{I}_m\end{bmatrix} \cdot \begin{bmatrix} L \\ N \end{bmatrix} , \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{m \times m}$ represents the causal relationships among the measured variables. Given a causal ordering of the measured variables, it could be permuted to strict lower triangularity (lower triangular with all zeros on the diagonal) due to the acyclicity assumption. $\mathbf{B} \in \mathbb{R}^{m \times p}$ represents the causal effects from the latent confounders to the measured variables. Each measured variable X_i is additionally associated with a corresponding noise variable N_i, where $i \in \{1, \ldots, m\}$. $\mathbf{I}_m = \mathrm{diag}(1, 1, \ldots, 1)$ is an identity matrix of dimension m representing the causal effect from the corresponding noise variable to the measured variable. The noise variables $N = (N_1, \ldots, N_m)$ and latent confounders $L = (L_1, \ldots, L_p)$ are assumed to be jointly independent and follow Gaussian distributions. In the remainder of this paper, we refer to the causal graph over $\{X, L, N\}$ underlying the data-generating FCM as the *ground-truth DAG*. The causal graph over X, defined by the non-zero entries in matrix A, is referred to as the *visible DAG*.

Taking additive sensor measurement errors $\mathcal{E} \in \mathbb{R}^m$ into account, sensor readings $S \in \mathbb{R}^m$ can be formalized as

$$S = X + \mathcal{E} \tag{3}$$

$$= \mathbf{A}X + \mathbf{B}L + N + \mathcal{E} , \tag{4}$$

where $\mathcal{E}_i$, $i \in \{1, \ldots, m\}$, is an aggregate representing many small, yet independent, sources of error and thus, by the central limit theorem, can be considered at least approximately Gaussian [14].

We introduce N and $\mathcal{E}$ for two reasons:

- To split the ground-truth sources of noise in sensor measurements. This distinction is necessary because measurement errors $\mathcal{E}$ only affect sensor readings S and should not propagate along the measured states X, while noise terms N affect both [15].
- To conform to the formal definition of an FCM according to [16], even if sensor measurement errors $\mathcal{E}$ are reduced to negligible values. Purely deterministic relationships between the sensor readings S only exist if all noise terms N are set to zero. We avoid determinism because it may introduce extra conditional independence relations among variables [11, Section 3.8], which can be problematic for causal discovery methods that are based on conditional independence tests.

Revisiting the three-sensor scenario in Fig. 1, the measured variables $X \in \mathbb{R}^3$ can be written as

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{A} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} B & \mathbf{I}_3 \end{bmatrix} \cdot \begin{bmatrix} L_1 \\ N_1 \\ N_2 \\ N_3 \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ b_2 & 0 & 1 & 0 \\ b_3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} L_1 \\ N_1 \\ N_2 \\ N_3 \end{bmatrix} , \tag{6}$$

where X_j is a direct cause of X_i if and only if $a_{ij} \neq 0$ and L_1 is a direct cause of X_i if and only if $b_i \neq 0$. For L_1 to be a latent confounder, at least two entries in B are non-zero.

Causal discovery in the presence of measurement error aims to infer causal relationships among the measured variables given observational data on the sensor measurements [14]. For the sake of simplicity, we assume here that the sensor measurement errors $\mathcal{E}$ are negligible, which leads to $S = X$. As a result, quality prediction boils down to predicting X_3 based on X_1 and X_2. Following the counterfactual RCA [17], we identify the root causes of anomalies by quantifying the contribution of each measured variable X_1, X_2, and X_3 to the outlier score at X_3, which requires functional relationships between variables in the form of an FCM. Due to the linear setup, we approach both downstream tasks by estimating functional relationships among the measured states X using linear regression according to

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \underbrace{\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}}_{=\mathbf{C}} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} , \tag{7}$$

where c_{i*} and r_i represent the linear regression coefficients and residual for X_i. Causal knowledge is injected in matrix $\mathbf{C}$ by forcing c_{ij} to be zero if X_j is not a direct cause of X_i in the causal graph and estimating all non-zero entries using linear regression. In addition, matrix $\mathbf{C}$ must be a hollow matrix (diagonal elements are all equal to zero), as a variable cannot be a predictor of itself.

The research questions proposed in Section I are therefore broken down to:

1) Do the downstream tasks of quality prediction and counterfactual RCA benefit from using the visible DAG to estimate the FCM (i.e., $c_{ij} \neq 0$ if and only if $a_{ij} \neq 0$)?
2) How does it affect the downstream tasks if causal graphs learned by causal discovery methods are imposed on matrix $\mathbf{C}$ (i.e., $c_{ij} \neq 0$ if and only if X_j is estimated to be a direct cause of X_i)?

III. Methodology

In order to answer the aforementioned research questions, we implement an end-to-end pipeline from data generation, causal discovery, FCM estimation to quality prediction and counterfactual RCA, as shown in Fig. 2. Each step is explained in this section.

A. Data Generation

The pipeline starts with the generation of synthetic data using the data-generating FCM defined in (5). We strengthen, weaken, or even deactivate the ground-truth causal relationships among the measured variables X by modifying the non-zero elements in matrix $\mathbf{A}$. The effect of latent confounding on X_2 and X_3 is varied by controlling b_2 and b_3 in B. We describe the ground-truth DAG and varied conditions in Section IV-A in more detail.

B. Causal Discovery

Given the observational data generated from the previous step, causal discovery methods are applied to identify the causal structure among X. We use the structural Hamming distance (SHD) to measure the distance between the learned DAG and visible DAG. It is defined as the total number of edge additions, deletions, or reversals that are required to transform one graph into another.

Without additional assumptions on the functional form of the FCM, the true causal structure can only be identified up to the Markov equivalence class (MEC), namely a set of DAGs that entail the same set of (conditional) independence conditions. An MEC can be represented by a *completed partially directed acyclic graph (CPDAG)*, which is the union graph of all DAGs in the equivalence class it represents. In CPDAGs, an edge is oriented if all DAGs in the MEC contain the edge in that direction. Otherwise, if there is uncertainty about the direction, it is left undirected. When a CPDAG is returned by a causal discovery method, the SHD is averaged over all DAGs it represents.

Restricting the functional form of the data-generating process allows us to distinguish the causal directions. It has been shown that the true causal structure can be uniquely identified from observational data by making further assumptions, such as non-Gaussian noise for linear models [18], additive noise for non-linear models [19], and post-nonlinear models [20]. Linear Gaussian models as in our setup are in general not identifiable from the joint distribution, unless the noise terms have equal variances [21], which is not fulfilled due to the presence of latent confounding. We exclude those FCM-based methods from the present study and choose PC, GES, and GOLEM as representatives of constraint-based, score-based, and gradient-based methods, respectively.

1) PC [11]: The PC algorithm begins with a complete undirected graph and iteratively deletes edges by testing conditional independence involving conditioning sets of increasing cardinality. Then, the second phase of the PC algorithm orients v-structures by re-using the conditional independences found in the first phase. Additional orientations can be inferred via the *Meek orientation rules* [22].

2) GES [23]: GES is a well-known two-phase procedure that performs a greedy search over the space of equivalence classes to find the graph that optimizes the Bayesian information criterion (BIC). In the first phase, it starts with an empty graph and greedily adds edges until a local maximum is reached; in the second stage, edges are greedily removed. The algorithm terminates once a local maximum is reached in the second phase.

3) GOLEM [24]: GOLEM is developed upon NOTEARS [25], formulating the structure learning problem as an optimization task using the log-likelihood objective instead of least squares to learn the weighted adjacency matrix representing the DAG. In particular, GOLEM optimizes within a continuous search space over DAGs rather than a discrete one as in PC and GES.

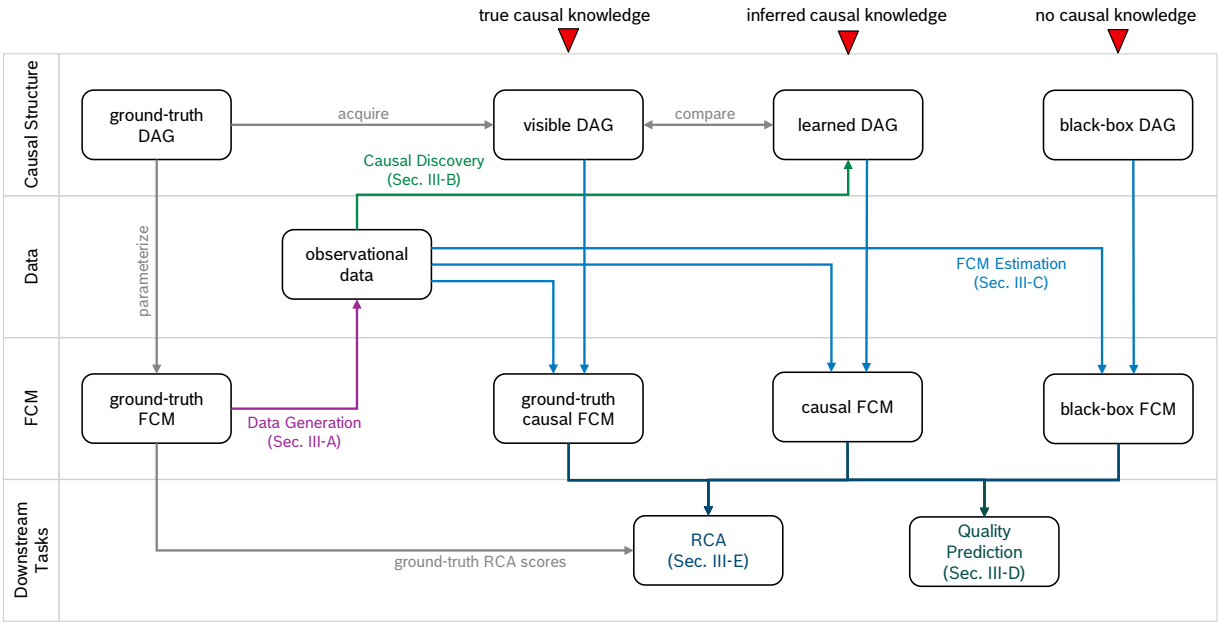


Fig. 2: End-to-end pipeline from data generation, causal discovery, FCM estimation to downstream tasks. The benefits of causal knowledge in quality prediction and RCA are evaluated via FCMs with three distinct levels of causal knowledge.

C. FCM Estimation

Due to the linear setup, linear regression is used to estimate the functional relationships between measured variables X, resulting in (7). More specifically, given a specific causal structure over X, a linear regression model is trained for each dependent node in the graph using its parents as predictors. If a variable is represented as a node without any incoming edges in the graph, solving linear regression is reduced to calculating the constant term in the equation, which is estimated by using the mean in the training data.

To assess to what extent having causal knowledge helps solve the downstream tasks, FCMs with distinct levels of causal knowledge are learned. The first level resembles a black-box machine learning model without any causal information, representing a fusion strategy that takes into account all available information. This level is referred to as *black-box* in the following. The second level corresponds to utilizing the inferred causal knowledge from data using causal discovery methods. The third level uses the visible DAG to estimate the FCM and is referred to as *ground-truth causal*.

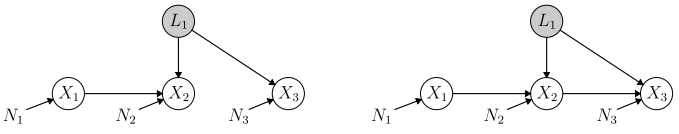
Following the best practice of data splitting in machine learning, we split the observational data used in causal discovery into training and validation data. Training data are used to estimate the FCM. If a causal discovery method is only able to recover up to the MEC, validation data are used to choose the DAG which leads to the most accurate causal FCM. More specifically, we estimate a causal FCM for each member in the MEC and compare the accuracy by averaging the mean squared error (MSE) over all nodes on validation data. We show in Section IV-B that the selected DAG is always the one closest to the visible DAG in every evaluated case.

D. Quality Prediction

Given every estimated FCM in the form of (7), the quality state X_3 is predicted as $c_{31}X_1 + c_{32}X_2 + r_3$. We compare the performance of FCMs learned with the described levels of causal knowledge on independent and identically distributed (i.i.d.) test data using the MSE of X_3. In addition, we generate out-of-distribution (OOD) test data with distribution shifts to showcase the robustness of the ground-truth causal FCM in the presence of latent confounding, compared to the black-box FCM without any causal insights.

E. RCA

The same set of i.i.d. test data are used to conduct RCA. Counterfactual RCA quantifies the contribution of each variable to the target outlier score by asking the counterfactual question: would the event not have been an outlier had we assigned rather "normal" causal mechanisms at the node instead of the existing causal mechanism associated with the outlier. To get counterfactuals, an FCM describing functional relationships between variables is required. We follow the work by Budhathoki et al. [17] and calculate the RCA scores for X_1, X_2 and X_3 by randomizing the causal mechanism (essentially the associated noise term) at each node and measuring the contribution of each noise term to the anomaly at X_3 using Shapley values [26] from cooperative game theory. RCA results calculated from the data-generating FCM serve as the ground-truth. RCA is done instance-wise for each sample in the test data. We use the root mean squared error (RMSE) between the predicted RCA scores and the ground-truth RCA scores averaged over X_1, X_2, and X_3 across all samples to measure the ability of FCMs with distinct levels of causal knowledge to solve RCA.



(a) Independent confounding (b) Dependent confounding

Fig. 3: Ground-truth DAGs underlying the data-generating FCMs. We differentiate between two cases based on whether X_2 is a direct cause of X_3 or not.

IV. Experiments

A. Experimental Setup

We consider two different cases of latent confounding based on whether the confounded variables are directly dependent or not. Fig. 3a presents a DAG where the ground-truth causal relationship between X_2 and X_3 is eliminated by setting a_{32} to 0 in (6). We refer to it as *independent confounding*. In this case, causal discovery methods tend to draw a false positive edge between X_2 and X_3 due to observed correlation in the data. In the case where X_2 is a direct cause of X_3, referred to as *dependent confounding*, as shown in Fig. 3b, a common mistake is to infer an edge between X_1 and X_3.

We generate data under controlled conditions of latent confounding by setting all non-zero entries in matrix $\mathbf{A}$ to 1 and varying the values of b_2 and b_3 according to

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ b_2 & 0 & 1 & 0 \\ b_3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} L_1 \\ N_1 \\ N_2 \\ N_3 \end{bmatrix} \quad (8)$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ b_2 & 0 & 1 & 0 \\ b_3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} L_1 \\ N_1 \\ N_2 \\ N_3 \end{bmatrix} \quad (9)$$

for independent confounding (8) and dependent confounding (9), respectively. Here, latent confounder L_1 and noise variables N_1, N_2 and N_3 are jointly independent and follow a standard Gaussian distribution. Assuming that the strength of latent confounding is less prominent than direct causal relationships among measured variables, we generate 100 combinations of b_2 and b_3 over a unit square grid, where both b_2 and b_3 range from 0 to 1. For each data-generating FCM using a specific combination of b_2 and b_3, we sample observational data of different sizes 100, $1,000$ and $3,000$ to further study the impact of sample size. The observational data is split into a $80:20$ ratio for training and validation during the step of FCM estimation. We generate i.i.d. test data consisting of 300 samples for the downstream tasks. To compare the robustness of prediction models to distribution shifts in the presence of latent confounding, we create additional OOD test data by varying a_{21}, i.e., the weight of the edge between X_1 and X_2, while setting $b_2 = b_3 \in \{0, 0.5, 1\}$.

B. Results

1) Causal FCM: Before investigating the results of downstream tasks, it is important to know what causal structures are learned by standard causal discovery methods under varied conditions of latent confounding, and which DAGs are selected among those structures to learn causal FCMs.

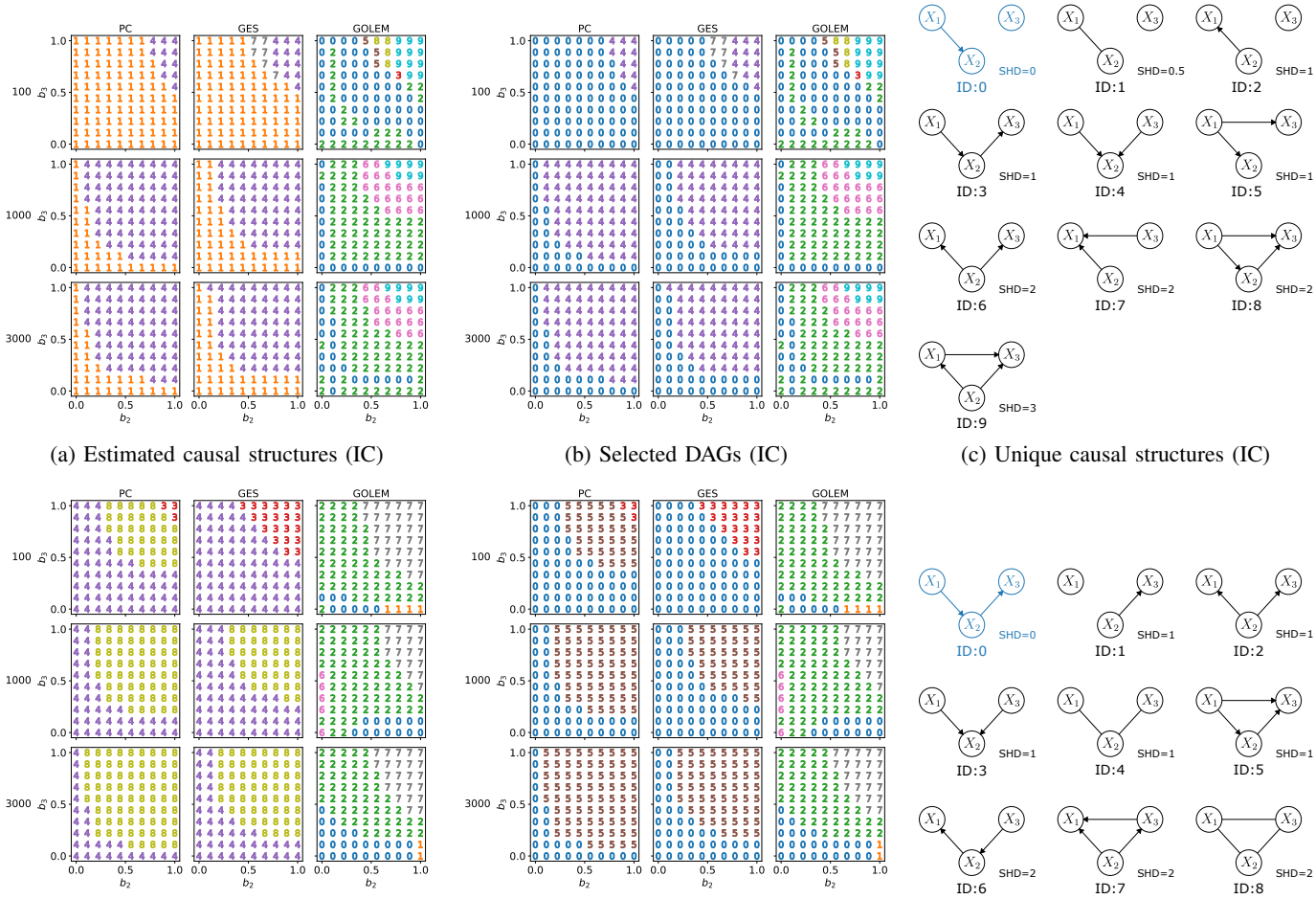
The first column in Fig. 4 presents the estimated causal structures using causal discovery methods in the case of independent confounding (top) and dependent confounding (bottom). It is a clear trend that the performance of all causal discovery methods drops with an increasing effect of latent confounding, i.e., increasing values of b_2 and b_3. Since PC and GES are only able to recover a DAG up to its MEC, CPDAGs are returned for both methods. In special cases, the estimated CPDAG represents a single DAG when there is only one DAG in the MEC it represents, e.g., ID 4 and 7 in Fig. 4a and ID 3 in Fig. 4d. A further inspect of the estimated causal structures reveals that PC and GES give consistent and stable results—predicting the false positive edge between X_2 and X_3 in the case of independent confounding and between X_1 and X_3 in the case of dependent confounding. In contrast, GOLEM gives less stable results and tends to wrongly infer the directions of true positive edges before gradually picking up false positive edges when the effect of latent confounding increases.

The DAGs selected to estimate the causal FCMs are shown in the second column of Fig. 4. By comparing them to the estimated causal structures shown in the first column, it can be seen that when a CPDAG with undirected edge(s) is estimated by PC and GES, e.g., ID 1 in Fig. 4a and ID 4 and 8 in Fig. 4f, our selection criterion is able to select the DAG in which the undirected edge(s) is oriented in such a way that the resulting DAG has the least SHD to the visible DAG. As a result, causal FCMs are estimated using the optimum DAGs inferred by causal discovery methods.

It should be mentioned that repeating the simulations with different random seeds leads to minor variations in the learned graphs in the border region between the correctly inferred causal structure and the incorrect causal structure. Such variations are expected due to the slight deviations in the joint distributions of observational data with limited sample sizes. While the fuzziness of the border region increases with decreasing sample size, the general behavior described previously remains the same.

2) Quality Prediction: In Fig. 5, we show the prediction performance of FCMs with distinct levels of causal knowledge on i.i.d. test data. We observe that the prediction quality drops with an increasing b_3, as the effect of latent confounder L_1 on the target X_3 is unobserved and cannot be learned by any prediction models. An improvement in prediction quality is commonly observed with an increasing amount of training data.

To assess if it is beneficial to take into account the true causal knowledge during quality prediction, we first compare the performance between the black-box FCM and ground-truth causal FCM. Without leveraging any causal information, the



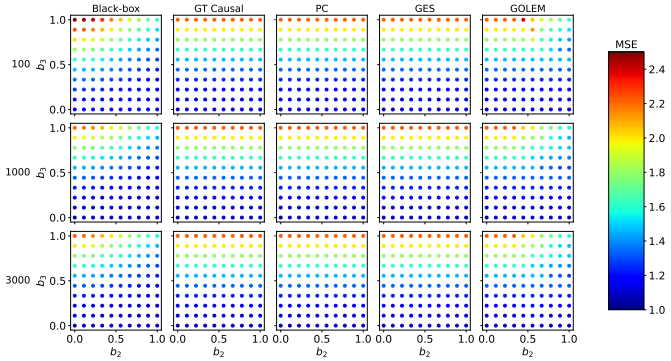
(a) Estimated causal structures (IC) (b) Selected DAGs (IC) (c) Unique causal structures (IC)

(d) Estimated causal structures (DC) (e) Selected DAGs (DC) (f) Unique causal structures (DC)

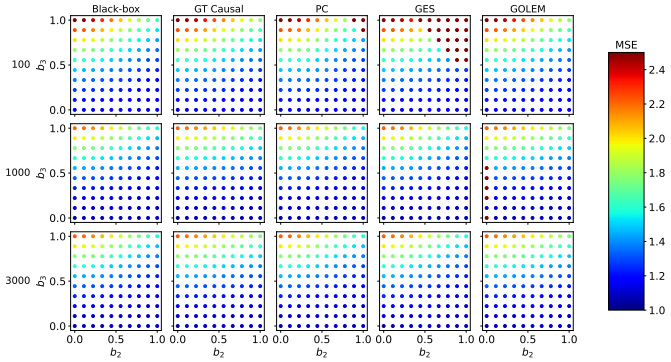
Fig. 4: Inferred causal knowledge under varied conditions of latent confounding controlled by b_2 on the x-axis and b_3 on the y-axis in the case of independent confounding (IC) and dependent confounding (DC). The three rows in (a), (b), (d), and (e) from top to bottom correspond to the causal structures obtained from observational data of size 100, 1000, and 3000, respectively. Unique causal structures are assigned unique IDs and sorted according to the SHD in the ascending order. The DAG with ID 0 corresponds to the visible DAG in each case. The numbers in (a) and (b) coincide with the IDs in (c). Same holds for (d) and (e) with respect to (f).

black-box prediction model always takes X_1 and X_2 as input to predict the value of X_3, while the ground-truth causal model leverages the mean of X_3 in the training data in the case of independent confounding, and X_2 in the case of dependent confounding. We observe that in the latter case, the ground-truth causal FCM and black-box FCM achieves comparable results, regardless of the effect of latent confounding (see Fig. 5b). In the case of independent confounding, as shown in Fig. 5a, the black-box model outperforms the ground-truth causal model, especially when both b_2 and b_3 are large. It can be explained by the ability of the black-box model to capture the correlation between X_2 and X_3 when the effect of latent confounding is stronger and thus predict X_3 better. However, the pattern learned from such spurious correlation caused by latent confounding does not generalize well to distribution shifts in OOD test data caused by a shifted relationship between X_1 and X_2, as we demonstrate in Fig. 6. In the case of independent confounding, the black-box model becomes less stable when the effect of latent confounding is stronger, while the ground-truth causal model is always stable, as it does not rely on the value of X_2 to predict X_3. A similar conclusion can also be made in Fig. 6b, where the ground-truth causal model is more robust to severe distribution shifts in the presence of latent confounding.

A further investigation into the performance of causal FCMs in Fig. 5 reveals that in the case of independent confounding, PC and GES produce the same results as the ground-truth causal FCM, while the performance of GOLEM is comparable to that of the black-box FCM. This observation is consistent with the DAGs underlying the causal FCMs shown in Fig. 4b—the selected DAGs for PC and GES always correctly represent X_3 as a node without any incoming edges, while GOLEM tend to infer X_1 and/or X_2 as direct causes of X_3 when the effect of latent confounding increases. In the

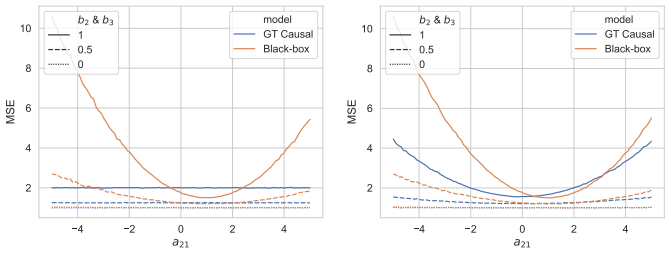


(a) Independent confounding



(b) Dependent confounding

Fig. 5: MSE of prediction models on i.i.d. test data when there is (a) independent confounding and (b) dependent confounding.



(a) Independent confounding (b) Dependent confounding

Fig. 6: MSE of the ground-truth causal model and the black-box model on OOD test data under different strengths of latent confounding. The results are averaged over 100 runs. In each run, both models are trained on $1,000$ samples with $a_{21} = 1$ and evaluated on OOD data consisting of $1,000$ samples generated from shifted a_{21}.

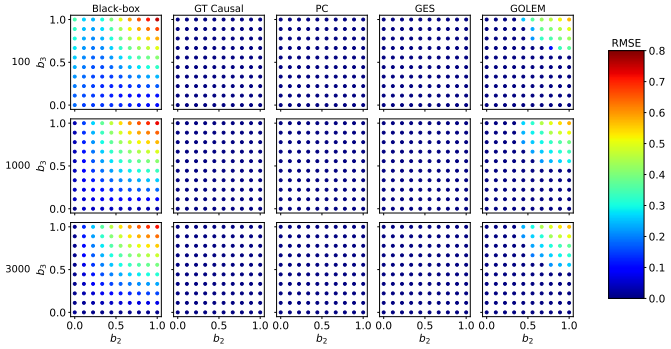
case of dependent confounding, all three methods are able to achieve comparable performance to that of the ground-truth causal FCM most of the time, with the exception being the scenario when the size of observational data is small, leading to wrongly inferred causal structures where X_3 is estimated not to have direct causes, as shown in Fig. 4e. It calls for caution when applying inferred causal relationships if the amount of training data is limited.

3) RCA: In Fig. 7, we show the performance of FCMs with distinct levels of causal knowledge in identifying the root causes of the anomaly at the quality state X_3. Similarly, we first compare the performance of the black-box FCM and ground-truth causal FCM. In both cases of latent confounding, utilizing the true causal knowledge significantly improves the RCA results. When X_2 is not a direct cause of X_3, Fig. 7a indicates that the ground-truth causal FCM is able to learn the RCA scores correctly despite the effect of latent confounding. Its underlying causal structure (i.e., the visible DAG) enables it to correctly attribute the anomaly at X_3 completely to itself (essentially the noise it receives), as there is no causal information flow from X_1 and X_2 to X_3. In contrast, the RCA scores predicted by the black-box FCM deviate a lot from the ground-truth RCA scores, because X_1 and X_2 are also considered as potential root causes of the anomaly. When X_2 is a direct cause of X_3, Fig. 7b shows that the RCA scores predicted by the ground-truth causal FCM are no longer error-free, as the existence of the latent confounder makes the estimation of the functional relationship between X_2 and X_3 biased. Nevertheless, the ground-truth causal FCM still outperforms the black-box FCM by a large margin.

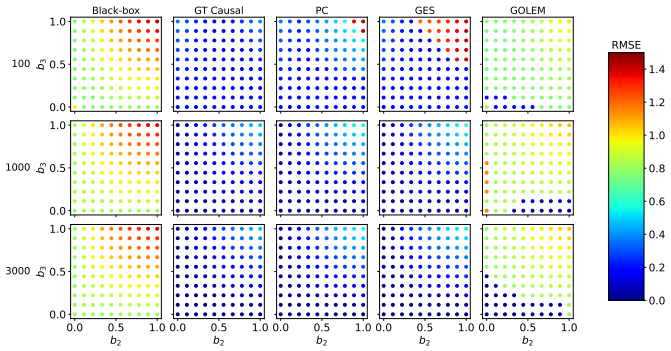
To analyze the reliability of inferred causal knowledge in solving RCA, the performance of the estimated causal FCMs and ground-truth causal FCM is compared. In the case of independent confounding, the causal FCMs estimated by the PC and GES algorithms achieve the same performance as the ground-truth causal FCM, while GOLEM leads to slightly worse RCA results when the effect of latent confounding is strong. This observation is again consistent with the DAGs underlying the causal FCMs shown in Fig. 4c – PC and GES only attribute the anomaly at X_3 to the noise on X_3 while GOLEM tends to attribute it to the noise on X_1 and/or X_2 when the effect of latent confounding is strong. Fig. 7b indicates that the task of RCA is very sensitive to the underlying causal structures. Models assuming wrong causal relationships could harm the performance of RCA to a large extent. Nevertheless, given enough training data, using inferred causal knowledge in both cases of latent confounding improves the performance of RCA, in comparison to not using causal knowledge at all.

V. Conclusions

The aim of the current study was to evaluate the benefits of using causal knowledge in machine learning-based data fusion to solve two typical manufacturing tasks, namely quality prediction and RCA. We conducted the evaluation in an end-to-end fashion from data generation, causal discovery, FCM estimation to downstream tasks. The results revealed that using true causal knowledge can prevent prediction models from learning spurious correlations caused by latent confounding, and thus make them more robust to severe distribution shifts in the data. Moreover, RCA benefits significantly from using causal knowledge during data modelling. Although the presence of latent confounding makes it challenging for causal discovery methods to infer the correct causal structure, we



(a) Independent confounding



(b) Dependent confounding

Fig. 7: RMSE of RCA scores calculated using FCMs with distinct levels of causal knowledge when there is (a) independent confounding and (b) dependent confounding.

showed that the selected DAGs according to the proposed selection criterion can be used as reliable input to solve downstream tasks when the amount of data is sufficient. A limitation of this study is that only three sensors are considered. However, we believe that having this simple scenario is necessary to ensure the interpretability of the results and we compensated the limitation by studying varied conditions of latent confounding in two different cases: independent confounding and dependent confounding. The findings in this study are subject to linear Gaussian systems without sensor measurement errors. A natural progression of this work is to include sensor measurement errors in the end-to-end pipeline described here. Another possible direction for future research is to evaluate the benefits of causal knowledge in non-linear and/or non-Gaussian systems. Moreover, it would also be interesting to repeat the study in larger simulation models, and eventually on real scenarios, to investigate the generalizability of the findings.

REFERENCES

[1] K. Marazopoulou, R. Ghosh, P. Lade, and D. Jensen, "Causal discovery for manufacturing domains," *arXiv preprint arXiv:1605.04056*, 2016.

[2] H. Boström, S. F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. Van Laere, L. Niklasson, M. Nilsson, A. Persson, and T. Ziemke, "On the definition of information fusion as a field of research," University of Skövde, School of Humanities and Informatics, Tech. Rep. HS-IKI-TR-07-006, 2007.

[3] H. Reichenbach, *The direction of time*. Univ of California Press, 1991, vol. 65.

[4] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[5] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *Journal of Process Control*, vol. 24, no. 3, pp. 223–233, 2014.

[6] L. Yao and Z. Ge, "Causal variable selection for industrial process quality prediction via attention-based gru network," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105658, 2023.

[7] J.-H. Hu, Y.-N. Sun, H.-W. Xu, Z.-L. Zhang, W. Qin, and X.-Y. Li, "Causality-based prediction method for the diesel engine assembly line system," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2022, pp. 199–204.

[8] H. Zhang, K. Peng, and L. Ma, "A systematic nonstationary causality analysis framework for root cause diagnosis of faults in manufacturing processes," *Control Engineering Practice*, vol. 131, p. 105404, 2023.

[9] E. E. Oliveira, V. L. Miguéis, and J. L. Borges, "Understanding overlap in automatic root cause analysis in manufacturing using causal inference," *IEEE Access*, vol. 10, pp. 191–201, 2021.

[10] R. Hattori, Y. Ota, T. Fujii, and H. Nakajima, "Anomaly ranking of failure causes in manufacturing process using causal model," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2021, pp. 34–39.

[11] P. Spirtes, C. N. Glymour, and R. Scheines, *Causation, prediction, and search*, 2nd ed., ser. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2000.

[12] M. Dhanakshirur, F. Laumann, J. Park, and M. Barahona, "A continuous structural intervention distance to compare causal graphs," *arXiv preprint arXiv:2307.16452*, 2023.

[13] J. Pearl, "Linear models: A useful "microscope" for causal analysis," *Journal of Causal Inference*, vol. 1, no. 1, pp. 155–170, 2013.

[14] R. Scheines and J. Ramsey, "Measurement Error and Causal Discovery," *CEUR workshop proceedings*, vol. 1792, pp. 1–7, Jun. 2016.

[15] T. Blom, A. Klimovskaia, S. Magliacane, and J. M. Mooij, "An upper bound for RandomMeasurement error in causal discovery," in *Uncertainty inArtificial Intelligence: proceedings of the Thirty-Fourth Concerence*, A. Globerson and R. Silva, Eds. Monterey, California, USA: AUAI Press, 2018, pp. 570–579.

[16] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. USA: Cambridge University Press, 2009.

[17] K. Budhathoki, L. Minorics, P. Blöbaum, and D. Janzing, "Causal structure-based root cause analysis of outliers," in *International Conference on Machine Learning*. PMLR, 2022, pp. 2357–2369.

[18] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan, "A linear non-gaussian acyclic model for causal discovery." *Journal of Machine Learning Research*, vol. 7, no. 10, 2006.

[19] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf, "Nonlinear causal discovery with additive noise models," *Advances in neural information processing systems*, vol. 21, 2008.

[20] K. Zhang and A. Hyvärinen, "On the identifiability of the post-nonlinear causal model," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 2009, pp. 647–655.

[21] J. Peters and P. Buhlmann, "Identifiability of gaussian structural equation models with equal error variances," *Biometrika*, vol. 101, no. 1, pp. 219–228, 2014.

[22] C. Meek, "Causal inference and causal explanation with background knowledge," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 403–410.

[23] D. M. Chickering, "Optimal structure identification with greedy search," *Journal of machine learning research*, vol. 3, no. Nov, pp. 507–554, 2002.

[24] I. Ng, A. Ghassami, and K. Zhang, "On the role of sparsity and dag constraints for learning linear dags," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 943–17 954, 2020.

[25] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," *Advances in neural information processing systems*, vol. 31, 2018.

[26] L. Shapley, "17. a value for n-person games," in *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, 2016, pp. 307–318.