Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AF6.20250202
Published: April 8, 2025

Exploring Causal Discovery in Manufacturing: Techniques and Applications

Aishwary Pawar
University Decision Support
Southern Methodist University
Dallas, TX 75205
USA
aishwaryp@smu.edu

Abstract

This paper investigates score-based and constraint-based causal discovery methods, emphasizing their application to manufacturing processes. The research focuses on how algorithms such as the PC (Peter and Clark) and FGES (Fast Greedy Equivalence Search) can uncover causal relationships within complex manufacturing systems, which are essential for process optimization, fault detection, and informed decision-making.The study was done in three key phases. First, a thorough literature review established the theoretical foundation for understanding causal discovery methods. These methods were applied to real-world manufacturing data in the second phase, focusing on assembly lines and production processes. The main goal was to find key variables impacting production efficiency, product quality, and equipment failure rates. In the final phase, the performance of the causal discovery algorithms was evaluated through regression analyses, where R-squared values measured how well the identified causal relationships explained observed outcomes. The results revealed that the PC algorithm can identify direct and indirect relationships among variables. Also, the FGES algorithm proved more effective at uncovering complex interactions in high-dimensional datasets.This study suggests that causal discovery techniques can be integrated into automated decision-support systems within manufacturing. Organizations can refine existing systems, optimize operations, and develop predictive models for proactive issue management by better understanding the causal structure of production processes. Future research should expand the application of causal discovery methods across a broader range of manufacturing environments.

Keywords
Causal discovery, Score-based methods, Constraint-based methods, Manufacturing processes, Process optimization.

1. Introduction
The technical advancements in sensors, equipment, and instrumentation over the past few years have resulted in a steady evolution of industrial process monitoring strategies. Modern plant-wide performance monitoring systems can examine many measurable variables simultaneously as these complicated industrial processes can generate a lot of information (Jiang et al. 2019; Yuan et al. 2017).

In the modern era, to understand industrial systems' complicated dynamics, there is a need to develop/implement precise and reliable causal models for aberrant event detection, monitoring, performance evaluation, and maintenance of the machine unit/processes. Causality analysis helps to add information to the already existing knowledge base. The purpose of using causality analysis in manufacturing is to learn more about the production process and improve decision-making based on actionable information. In several instances, data-driven causality analysis techniques used in engineering settings have difficulty modeling controller interactions, considering temporal data, and identifying the sequential relationships between events that affect process performance (Nadim et al. 2022). Therefore, this work aims to demonstrate how various causal discovery algorithms can be utilized to understand manufacturing operations comprehensively and how this data-driven information/knowledge may be connected to the company's business operations.

Causality is a key concept in experimental design and statistics associated with determining its cause and effect. It implies that if we are aware of the causes and contributing variables to a specific issue, it is to our most significant advantage to investigate and evaluate the interactions and connections between these components to forecast the quantitative changes and the consequences on their outcome if some of these components change (Profillidis and Botzoris 2019). Using treatments or randomized trials is a conventional method for determining causal relationships, but these methods are often costly, time-consuming, or both. As a result, causal discovery methods (a process of identifying causes by studying just observational data) have received considerable attention (Glymour et al. 2019). Although causality analysis appears to be a straightforward idea in daily life, establishing causal linkages can be challenging in many situations (Keele 2015). Some assumptions shared by various causal discovery algorithms include acyclicity, Markov property, faithfulness, and sufficiency (Kalainathan et al. 2018).

Two primary causality analysis strategies that have been discussed in this study include score-based methods and constraint-based methods. Score-based methods utilize local search operators to explore the Directed Acyclic Graphs (DAGs) and find the Markov equivalence class of the graph. In contrast, constraint-based methods use conditional independence tests to discover the skeleton of the graph and the v-structures, allocate a score to each, and select a final graph based on the scores. This work includes implementing one well-known method from each category: PC (Peter and Clark), a constraint-based approach, and FGES (Fast Greedy Equivalence Search), a score-based algorithm.

The rest of this article is structured as follows; the associated literature is briefly described in the next section, followed by a detailed explanation of the constraint-based and score-based methods. Next, a case study illustrating the causality algorithms used in this work is discussed. This section covers the description of the data and the entire systematic procedure of the proposed strategy. Also, the findings of the causal discovery algorithms are included in this section, followed by a conclusion section.

1.1 Objectives

The primary objective of this study is to evaluate the effectiveness of score-based and constraint-based causal discovery algorithms, specifically the PC (Peter and Clark) and FGES (Fast Greedy Equivalence Search) methods, in uncovering causal relationships within manufacturing processes. By applying these algorithms to real-world manufacturing data, the research aims to identify key variables that influence production efficiency, product quality, and equipment failure rates. This will help in understanding the critical dependencies and interactions that impact overall performance in industrial systems. Another key objective is to compare the capabilities of the PC and FGES algorithms in modeling complex, high-dimensional datasets. The study seeks to determine which algorithm is better suited to uncovering direct and indirect relationships within manufacturing systems. Additionally, the research will explore how the insights derived from causal discovery can be integrated into automated decision-support systems within manufacturing environments, ultimately optimizing operations, enhancing fault detection, and improving predictive modeling for proactive issue management. Finally, the study aims to provide recommendations for future research, expanding the application of causal discovery techniques to a broader range of manufacturing sectors, and identifying areas where these methods can significantly contribute to process improvements and data-driven decision-making.

2. Literature Review

In recent years, manufacturing has benefited from using computer science methods to improve product quality, fault detection, and process monitoring. Causality relationships enable the analysis of underlying causes, scale faster, and emphasize the critical process factors without overloading expert operators (Menegozzo et al., 2020). Several studies explore and use causality analysis in various fields in existing literature. However, the studies assessing the implementation of causality analysis in manufacturing are scarce. Here, we briefly review the results of the recent empirical studies utilizing causality analysis.

Jin et al. (2018) proposed a causal model-based scheduling method to guide the steel industry's coke oven gas (COG) scheduling operation. The causal relationships developed to lead to a causal diagram. With the development of the training sample using the key variables, the LSSVM approach is used to train the gas tank level prediction model. A modified particle swarm optimization (PSO) technique is used to create and optimize an objective function that considers the scheduling outcome. The findings show that the suggested technique can make practical COG scheduling recommendations.

A recent study compared different variable selection methods based on causality in terms of fault detection performance. The results were compared with several other filter-based, wrapper-based, and embedded-based variable selection methods. The experimental findings demonstrate that using variable selection techniques based on causality for model construction led to models that performed better during the fault detection stage (Clavijo et al., 2021). Wong et al. (2011) deployed a PC algorithm to identify the connectivity relationship between the crucial factors in the supply chain model. This is followed by the implementation of the neural network to quantify the relative significance of various elements in forecasting the essential factors. This approach helped the researchers make subjective decisions, for example, providing a realistic initial path and factor selection for the subsequent predictive modeling. Le et al. (2016) developed a parallel- PC algorithm based on the parallel computing technique and applied it to a variety of synthetic and real-world high-dimensional data. Results show that the parallel-PC algorithm outperformed the original PC algorithm.

The parallel-PC algorithm completed the task within 12 hours with a 4-core CPU computer and six hours with an 8-core computer. Additionally, it enhances both the efficiency and accuracy of the causal inference algorithm. In contrast, the PC algorithm could not generate any results after running for more than 24 hours. A study by Lai and Bessler (2014) utilized two machine learning algorithms (namely PC and LiNGAM) to determine the causal relationships between retail prices, manufacturer prices, and the number of packages sold. Data includes retail sales of carbonated soft drinks in the Chicago region. The findings demonstrated that the PC algorithm could not determine which factor—retail price, manufacturer pricing, or quantity sold—should take precedence over the others in causality. However, the LiNGAM algorithm was successful in doing so in every instance. To identify the specific set of directed pathways that best characterize the dataset, the LiNGAM technique assumes that the data is not normal. The original LiNGAM technique has been enhanced in several ways, including pairwise, parceled, and pooling (Henry and Gates, 2017). Pairwise, LiNGAM is used to orient the edges in an existing undirected graph. In a study, it was found that pairwise LiNGAM properly orients the edges 75% of the time, while for the analysis of concatenated data, the orientation accuracy was 100% (Hyvärinen et al. 2010; Smith 2013).

In a study, Schmidt et al. (2018) addressed an important limitation of the PC algorithm, i.e., its lengthy execution time. Where execution time is exponential w.r.t dataset dimension and a polynomial in case of sparse causal structures, this study created an effective GPU-accelerated solution for Gaussian distribution models by utilizing the parallel processing capability of GPUs to overcome this limitation. Experimental findings show that GPU-accelerated implementation outperforms existing CPU-based versions 700 times. CPC (Conservative PC) algorithm works like PC but employs adjacency search to reduce the number of false orientations (Ramsey et al., 2012). Zhalama et al. (2017) allowed violations of Adjacency-Faithfulness and Orientation-Faithfulness and demonstrated that the CPC algorithm (conservative), a prominent constraint-based method, can be built more robust against unfaithfulness by implementing elements of the GES algorithm; similarly, the GES algorithm errors can be reduced by combining features of the CPC algorithm. GES algorithm has been implemented in various applications. For instance, researchers have incorporated GES to look at the differences in brain functions among people with traumatic brain injury and those with autism spectrum disorder (Hanson et al.2013, Dobryakova et al. 2015). Smith et al. (2011) showed that performing multiple simulation runs of GES on low dimensional issues (such as those with 50 variables and 200 data points) requires significant computational time.

3. Methods

This section describes one well-known method from each category: PC (Peter and Clark), a constraint-based approach, and FGES (Fast Greedy Equivalence Search), a score-based algorithm.

The PC algorithm starts with a comprehensive, undirected graph and removes edges repeatedly based on conditional independence tests, i.e., repetitive conditional independence tests are used to identify the causal links among several variables in a data set. Figure 1 represents different steps in the PC algorithm analysis.

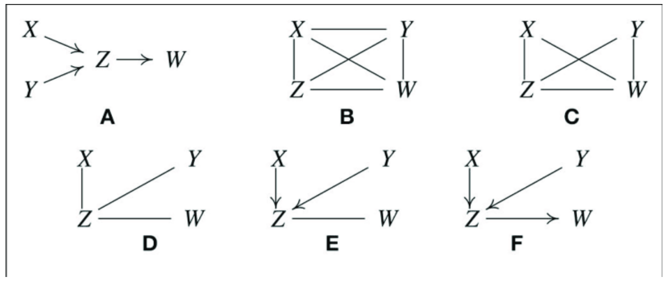

Figure 1. PC Algorithm Fundamentals (Glymour et al., 2019)

As an illustration in the given Figure 1; Step (1A) shows the actual causal structure within a given set of variables, Step (1B) demonstrates the first phase in the PC algorithm, which provides a full undirected graph, Step (1C) eliminates all edges among variables that are unconditionally independent, Step (1D) removes the edge between two variables given a third variable if (Variable 1) ⊥ (Variable 2) | (Variable 3). The algorithm then tests for more conditional independence by expanding the subsets of variables until there are no more adjacent pairs to (Variable 1) or/and (Variable 2). For example: it checks for a condition (Variable 1) ⊥ (Variable 2) | {(Variable 3), (Variable 4)} such that (Variable 3) and (Variable 4) are adjacent to (Variable 1) or/and (Variable 2). Lastly, Steps (1E) and (1F) deal with handling v-structure and orientation propagation, respectively. The term "V-structure" refers to a trio of variables (Variable1, 2, and 3), where Variables 1 and 2 are adjacent to each other, Variables 2 and 3 are adjacent to each other, and Variable 1 and 3 are not adjacent. The algorithm then aligns the edges as Variable 1→Variable 2←Variable 3, where the edge between them was eliminated if Variable 2 was not in the set conditioning on which Variables 1 and 3 became independent. Now, for each triple, Variable 1→Variable 2–Variable 3 such that Variables 1 and 3 are not adjacent, the algorithm performs orientation propagation and aligns the edges as follows: Variable 1→Variable 2→Variable 3 (Glymour et al. 2019).

GES is a Bayesian algorithm that iteratively searches the domain of causal Bayesian network (CBN) and delivers the model with the best score it discovers. FGES is an enhanced and parallelized version of a GES algorithm. There are two stages in this algorithm. The first phase of the search begins with an empty graph, i.e., a graph with no edges. It performs a forward search in which edges are added between nodes to increase the Bayesian score (BIC). The BIC score is likelihood penalized for complexity to decrease overfitting (Schwarz, 1978). This process continues until no single edge addition results in a higher score. GES then eliminates edges one at a time using a backward stepping search until no single edge removal can lead to an increase in the score (Shen et al., 2020).

4. Data Collection

For this study, synthetic data was used to replicate the assembly line data from a car manufacturing plant, providing a controlled environment for applying and evaluating score-based and constraint-based causal discovery techniques. The data is designed to mimic the dynamics and complexities observed in a real-world automobile assembly plant. The dataset is structured to represent a series of workstations along an assembly line, with each workstation corresponding to a specific stage in the manufacturing process.

Figure 2 illustrates the layout of these workstations, labeled as clusters A, B, C, etc., where each letter represents a different workstation. For example, A_1, A_2, …A_n refer to various process parameters or variables at stage 1, B_1, B_2, …B_n represent different process parameters at stage 2, and so on, across multiple stages of the assembly line. The direction of the arrows in the figure indicates the order of precedence or flow in the assembly process. The total dataset consists of 110 variables, each representing different process parameters, with 8,198 data points (rows). This large volume of data provides a detailed representation of the processes and interactions at each workstation.

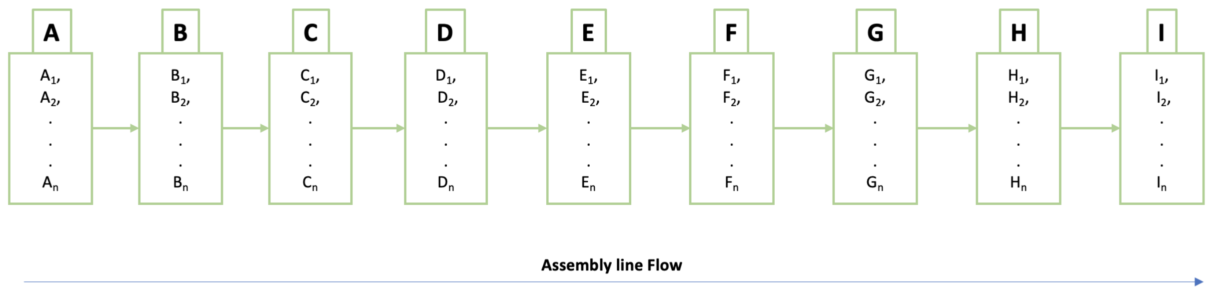

Figure 2. Overview of the workstations and the dataflow in an assembly line

To enhance the robustness and generalizability of the analysis, the dataset was divided into 15 distinct subsets, each containing 500 data points. These subsets were analyzed separately to provide a more comprehensive evaluation of the causal discovery algorithms and their ability to uncover relationships between process variables across different segments of the manufacturing process. By using this synthetic dataset, the study ensures the reproducibility and control needed for testing causal discovery algorithms, while still providing insights applicable to real-world manufacturing environments.

5. Results and Discussion

5.1 Numerical Insights, Graphical Representations, and Statistical Interpretations

In causal discovery, it is crucial to recognize that different causal search algorithms are suited to specific circumstances, and choosing the right algorithm depends on the characteristics of the dataset being analyzed. Furthermore, before applying any algorithm, it is essential to assess whether the assumptions underlying the algorithm can be realistically met for the given data. This study focuses on a synthetic assembly line industrial dataset from an automotive plant and utilizes two distinct causal search algorithms: constraint-based and score-based methods.

Any causal search must begin by ensuring the data is adequately prepared for analysis, i.e., converting the raw data into a clean data set. This includes removing and dealing with unformatted real-world data and checking for irregularities such as missing values, nan values, etc. Next, there is a need to exclude "redundant" variables since causal search algorithms require variables that are "semantically independent" (Spirtes and Scheines, 2004). Multicollinearity is a scenario where two or more explanatory variables are linearly associated. Thus, ensuring the dataset is free of collinearity is one common rule to avoid "matrix inversion" error caused by perfect collinearity in a dataset.

For datasets that include both continuous and categorical variables, causal discovery algorithms must be adapted accordingly. In this study, the dataset consists solely of continuous variables, making it suitable for standard causal discovery methods. Although time series data typically requires additional constraints for causal analysis, the dataset used in this case study does not include time series data. However, the data points are sorted using a timestamp, ensuring that the results are analyzed in a specific order. Finally, incorporating background knowledge to define a logical order of causal relationships can help enhance the accuracy of the analysis. These preprocessing steps are critical to ensuring the validity and robustness of the results produced by the causal discovery algorithms.

The next step in causal search is implementing an appropriate algorithm for a given data set and scenarios. This includes the PC algorithm and FGES algorithm. The PC algorithm assumes that the input data given causal structure is acyclic and that the same latent variable causes no two variables. Additionally, it is expected that the distribution of each variable is normal and that the causal link between two variables is linear (if the input data is continuous, as it is in this instance). The null hypothesis, which is usually an independent or conditional independence hypothesis, is rejected using the alpha value. Thus, in the case of continuous variables, PC utilizes the test of zero correlation or the test of zero partial correlation for continuous variables to determine conditional independence. For discrete or categorical data, PC employs either chi-square or a g-square test of independence or conditional independence. In each scenario, the tests call for an alpha value that the user can change to reject the null hypothesis. In this case, the data is continuous, which requires the PC algorithm to quantify the relationship using Pearson's correlation coefficient or partial correlation and determine the strength of the correlation using Fisher's z-transformation or Fisher's r to z transformation.

The FGES algorithm uses a BIC score to determine which CBN structure is the most likely. It requires samples in the data to be independent and identically distributed. Other assumptions include the Markov condition, causal faithfulness condition, no hidden confounders, and selection bias. Lastly, there are no feedback cycles among the measured variables. Both the algorithms assume that direct causal influences are linear. Results of the PC and FGES analyses are displayed in the next section. These algorithms are executed in tetrad software, a Java program for causal discovery developed by researchers from the Center for Causal Discovery (CCD) (Scheines et al., 1998; Cooper et al., 2015). The first step for causality analysis in Tetrad Software is to add "Data" and "Search box" to the GUI and connect them with an arrow (as shown in Figure 3). The user may then load the data into TETRAD by double-clicking on the "Data" button and choosing the file to load. This leads to a new GUI, which shows data preview and gives some options to set data attributes. This includes missing data, first-row variable names, Data file type, etc. (as shown in Figure 4).

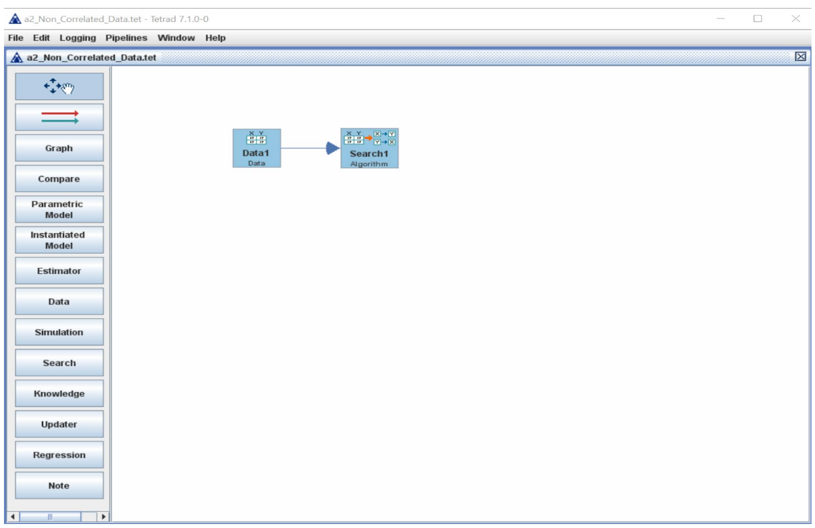

Figure 3. Loading Data in TETRAD

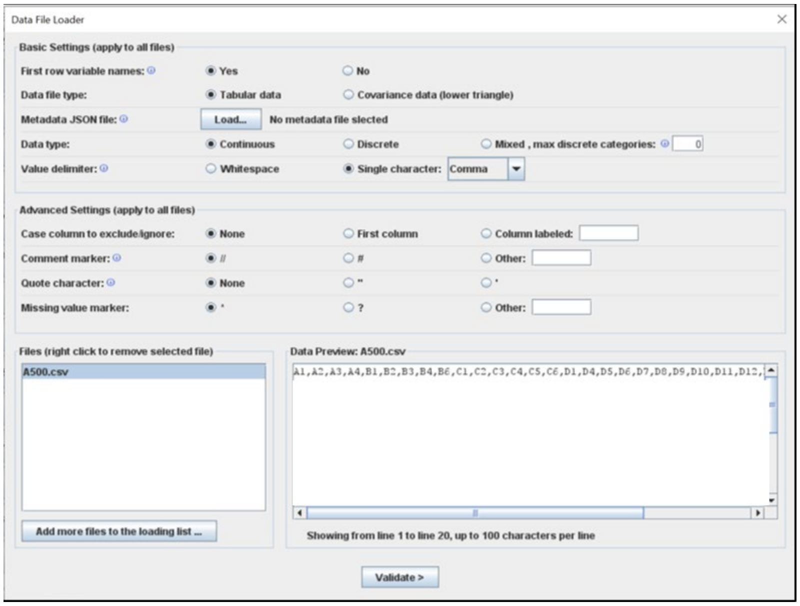

Figure 4. Data Preview and features

Thus, this step involves data validation, and finally, data is loaded for further analysis (if validation is passed). The next step to performing a causal search requires adding a "Search" icon and "Knowledge box" (as shown in Figure 5). Figure 6 demonstrates the functionality of the Knowledge Box, which allows users to organize the data points and variables into specific tiers, ensuring proper temporal ordering. Additionally, it enables the inclusion of forbidden and required groups, which contribute to defining constraints and relationships, ultimately leading to more robust and accurate results.

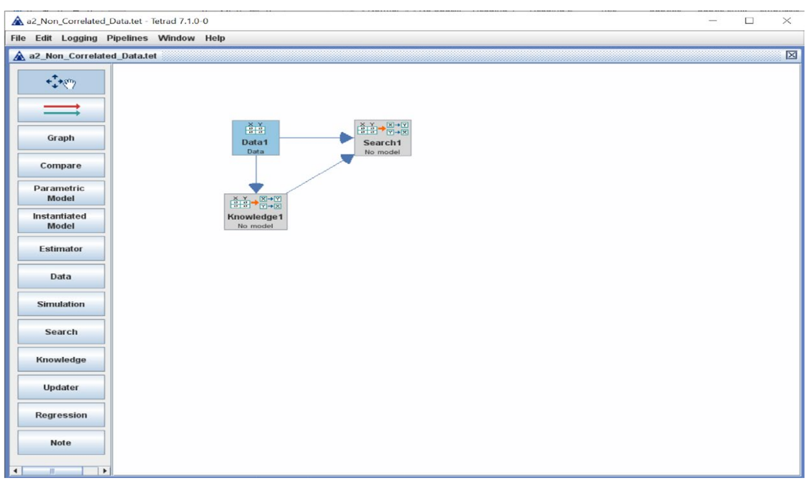

Figure 5. Search Icon and Knowledge Box in TETRAD

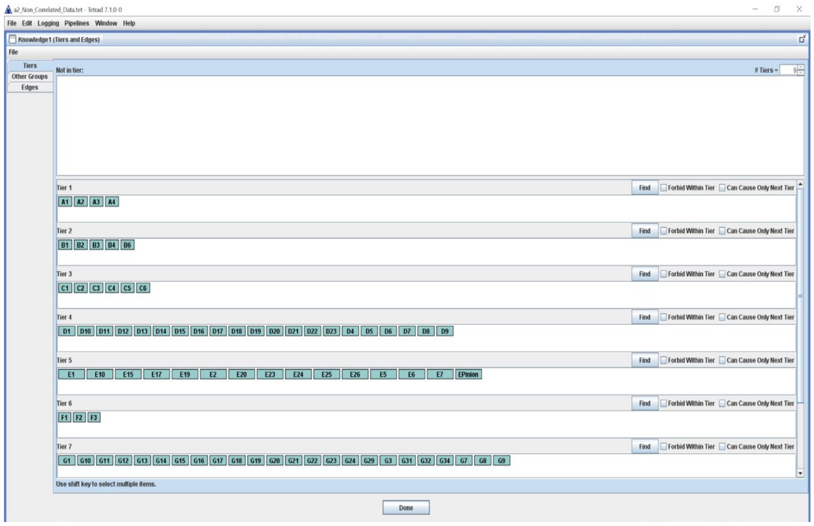

Figure 6. Knowledge Box Tier information

The "Search" icon enables users to choose a specific algorithm for causality analysis. In this study, both the PC algorithm and the FGES algorithm are selected simultaneously. Figure 7 illustrates the algorithm selection process for the PC algorithm. The graphical user interface (GUI) also includes filters for algorithm configuration, such as the option to "search for Markov blankets," and statistical tests, including conditional independence tests like the Fisher Z Test. Once an algorithm is selected, users can define search parameters, including maximal conditioning sets, p-value thresholds, and bootstrapping settings, as shown in Figure 8. These customizable settings allow for more tailored and precise causal analysis. Execution of this step gives the PC-based causal relationships (as shown in Figure 9). This procedure is again repeated for finding causal relationships based on the FGES algorithm

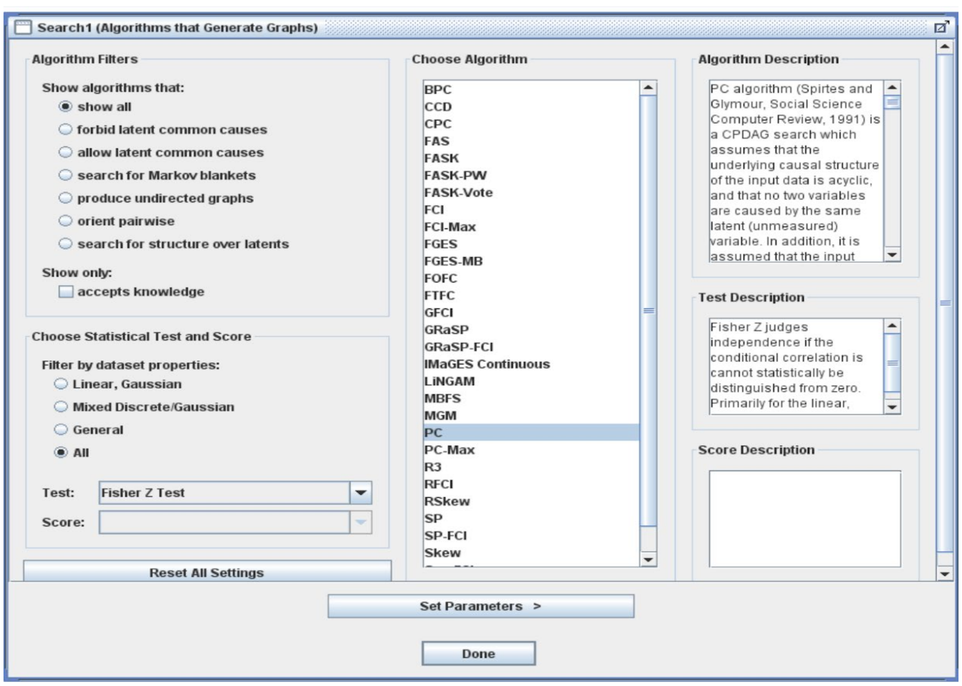

Figure 7. Algorithm and test selection

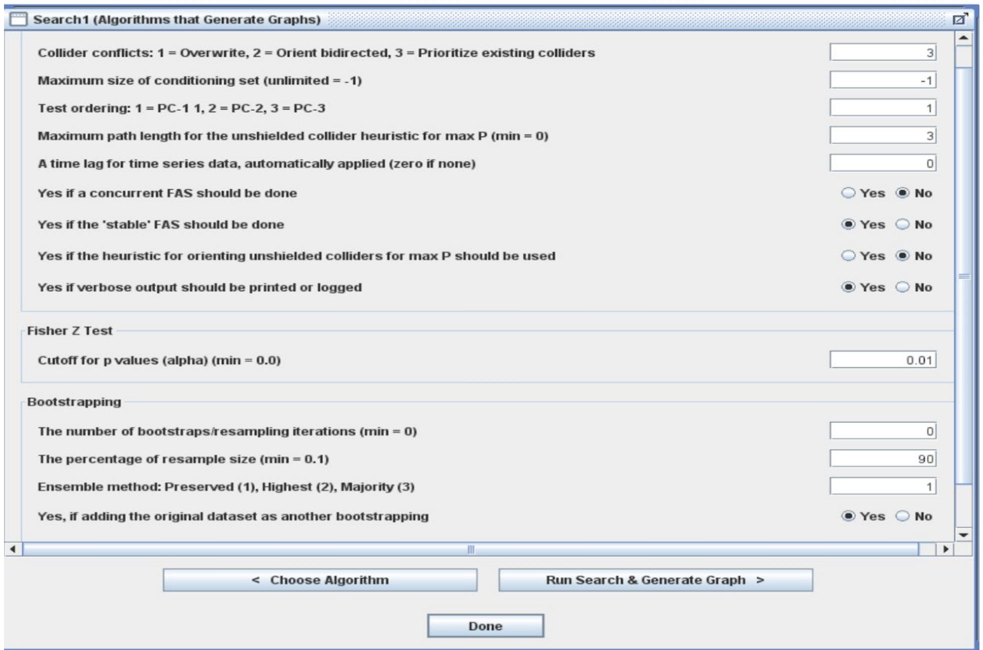

Figure 8. Algorithm Search Parameters

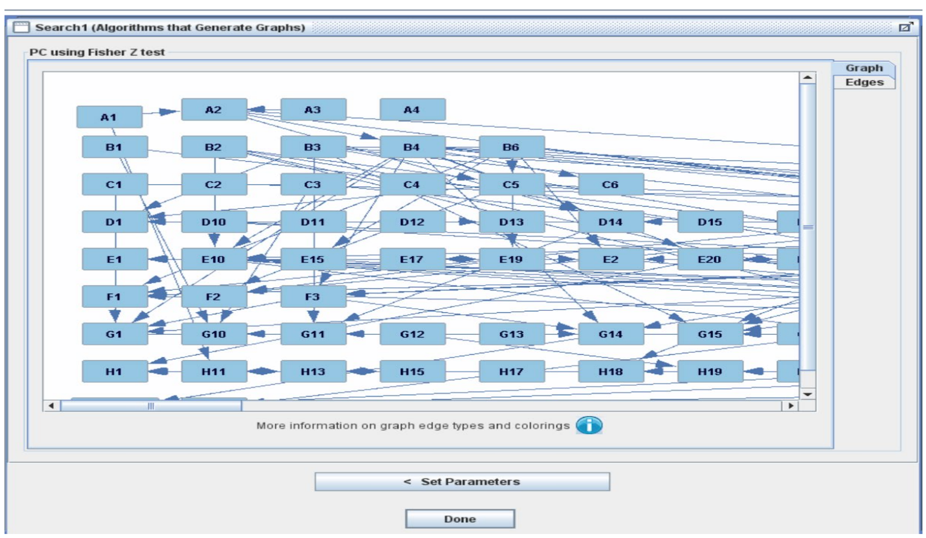

Figure 9. PC based Causal Relationship Graph

5.2 Validation

To evaluate the performance of the causal algorithms, the identified causal relationships are used to construct a linear regression model. Figure 10 and Figure 11 shows a snapshot of PC-based causal relationships and FGES-based causal relationships respectively.

```
'E7': ['B3'],
'E5': ['D21'],
'E25': ['B3', 'D10', 'D18'],
'E24': ['D19'],
'E23': ['E1'],
'E20': ['B4', 'E25'],
'E2': ['B6', 'E17', 'E5'],
'E19': ['B6', 'E1'],
'E17': ['E19'],
'E15': ['B6'],
'E10': ['B2', 'B4', 'D9', 'E20'],
'E1': ['E17', 'E5'],
'D9': ['B4', 'D5', 'D6'],
'D8': ['A2', 'D15'],
'D7': ['C2'],
'D5': ['D6', 'D8'],
'D4': ['B4', 'D6'],
'D23': ['B3'],
'D22': ['B4', 'C5'],
'D21': ['B6', 'D20'],
'D20': ['B6'],
'D18': ['C5'],
'D17': ['D13'],
'D16': ['D17', 'D18'],
'D14': ['A2', 'B4', 'D18'],
'D13': ['D1', 'D12'],
'D1': ['B3', 'C5', 'D17'],
'C6': ['B4'],
'C5': ['B2', 'B6'],
'C3': ['C1'],
'B6': ['B3'],
'B4': ['A2'],
'A2': ['A1', 'A3']}
```

Figure 10. PC based Causal Relationship Dictionary

```
'E7': ['B3', 'D9', 'E10', 'E25'],
'E5': ['B4', 'B6', 'D1', 'D21', 'E1', 'E17', 'E2', 'E20', 'E7', 'EPinion'],
'E25': ['B3', 'D10'],
'E23': ['E1'],
'E20': ['B4', 'D9', 'E25'],
'E2': ['B3', 'B6', 'D18', 'E1'],
'E19': ['B3', 'B6', 'C5', 'D1', 'D17', 'D21', 'E1', 'E2', 'E7', 'EPinion'],
'E17': ['B3', 'B6', 'D9', 'E1', 'E10', 'E2'],
'E15': ['B6'],
'E10': ['A2', 'B2', 'B4', 'C5', 'D14', 'D20', 'D22', 'D9', 'E20'],
'E1': ['D21', 'E10'],
'D9': ['B3', 'B4', 'D6'],
'D8': ['D5', 'D6'],
'D6': ['D4'],
'D5': ['B2', 'D6', 'D9'],
'D4': ['B4'],
'D22': ['C5'],
'D21': ['B6', 'C5'],
'D20': ['B6', 'D21'],
'D18': ['C5', 'D17', 'D5'],
'D17': ['D1', 'D13', 'D9'],
'D16': ['D17', 'D18'],
'D15': ['D8'],
'D14': ['B4', 'D18'],
'D13': ['D1', 'D12', 'D9'],
'D12': ['C5'],
'D1': ['B3', 'C5'],
'C1': ['C3'],
'B4': ['A1', 'A2'],
'B3': ['B6'],
'A3': ['A2'],
'A2': ['A1']}
```

Figure 11. FGES based Causal Relationship Dictionary

The efficacy of the model is then analyzed by comparing the R-Squared values between two scenarios: one based on causal relationships (using the algorithms) and one based on the inclusion of all variables from the preceding workstation, without considering causality. In a regression model, R-Squared (also known as R2 or the coefficient of determination) is a statistical metric that measures the percentage of the dependent variable's variation that can be accounted for by the independent variable. Figure 12 shows the R-Squared value for each variable without considering causal relationships. In this case, the independent variables include all the variables from the preceding stage for a given dependent variable. For instance, If the dependent variable is C1, then independent variables will include A_1, A_2, $\ldots A_n$, B_1, B_2, $\ldots B_n$.

In contrast, Figures 13 and 14 show the R-Squared values based on causal relationships derived from the PC and FGES algorithms, respectively. In these cases, the independent variables are determined by the causal relationships identified through each algorithm (refer to Figures 10 and 11). For instance, for the dependent variable E7, the independent variable would be B3 based on the PC-based causal relationship.

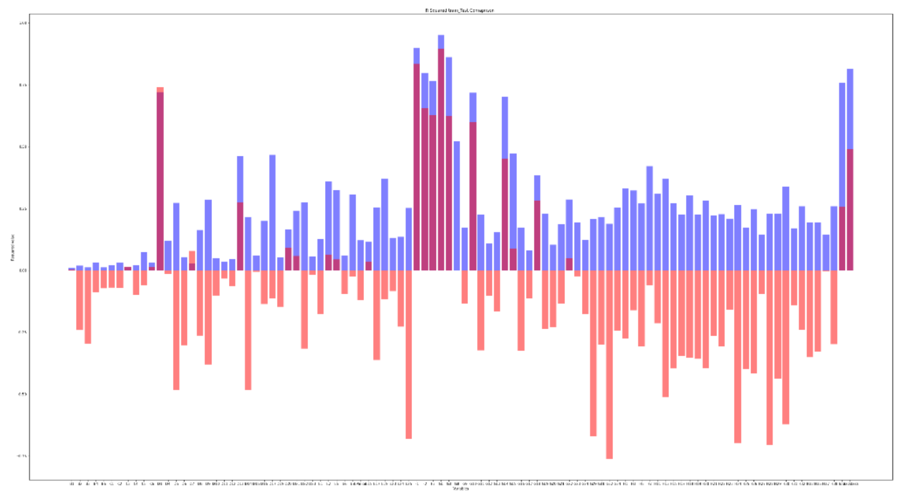

Figure 12. R-Squared value based on Full Regression Model

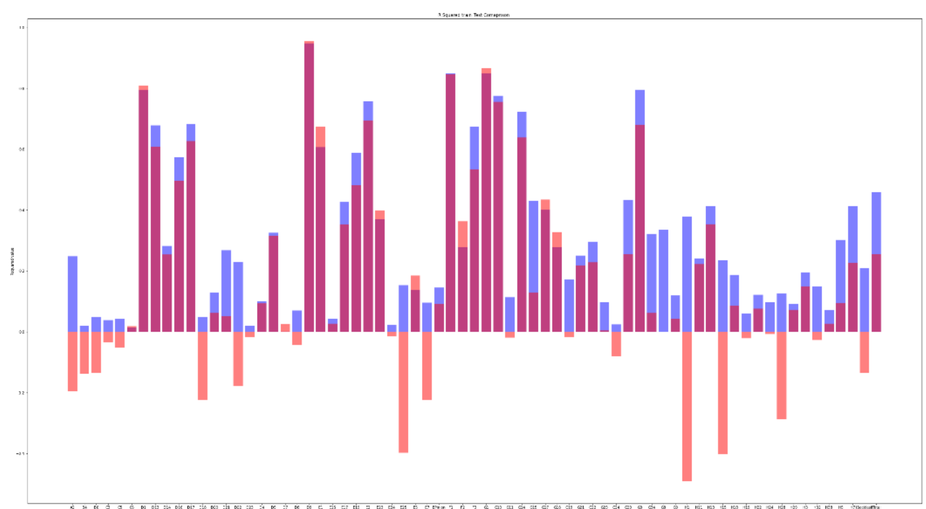

Figure 13. R-Squared value based on PC based Causal Relationship

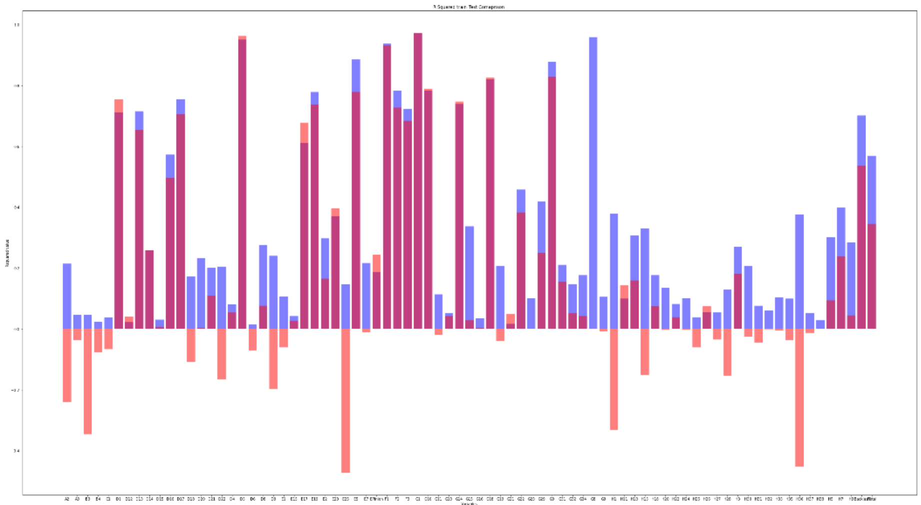

Figure 14. R-Squared value based on FGES based on Causal Relationship

6. Conclusion

In conclusion, this study highlights the potential of score-based (FGES) and constraint-based (PC) causal discovery algorithms in uncovering causal relationships within manufacturing processes, a tool not widely recognized in this domain. The research demonstrated that these algorithms could effectively identify the causal structure between variables, ultimately improving the understanding of key dependencies that drive production efficiency, product quality, and equipment failure rates. By evaluating and comparing the effectiveness of the PC and FGES algorithms, this work contributes to determining the best-suited approach for modeling complex, high-dimensional datasets typical of manufacturing environments. These findings suggest that causal discovery can enhance decision-making in industrial systems by providing clearer insights into how different factors influence outcomes, enabling more accurate predictions and proactive fault detection.

The application of causal discovery models, particularly those based on the PC and FGES algorithms, opens up new opportunities for refining manufacturing operations. While previous studies have not explored the use of these models to estimate heterogeneous causal effects in manufacturing, this research fills that gap and offers a foundational framework for future investigations. By demonstrating the potential for causal analysis in industrial settings, the study advocates for the integration of causal discovery techniques into automated decision-support systems. This integration can optimize operations, improve predictive modeling, and help organizations better manage issues before they escalate. Looking ahead, further research can expand the application of these methods across various manufacturing sectors, providing deeper insights into process improvements and advancing data-driven decision-making.

References

Clavijo, N., Melo, A., Soares, R. M., Campos, L. F. d. O., Lemos, T., Câmara, M. M., . . . Pinto, J. C., Variable Selection for Fault Detection Based on Causal Discovery Methods: Analysis of an Actual Industrial Case, *Processes*, vol. 9, no. 3, pp. 544, 2021.

Dobryakova, E., Boukrina, O. and Wylie, G. R., Investigation of information flow during a novel working memory task in individuals with traumatic brain injury, *Brain Connectivity*, vol. 5, no. 7, pp. 433-441, 2015.

Glymour, C., Zhang, K. and Spirtes, P., Review of causal discovery methods based on graphical models, *Frontiers in Genetics*, vol. 10, pp. 524, 2019.

Hanson, C., Hanson, S. J., Ramsey, J. and Glymour, C., Atypical effective connectivity of social brain networks in individuals with autism, *Brain Connectivity*, vol. 3, no. 6, pp. 578-589, 2013.

Henry, T. and Gates, K., Causal search procedures for fMRI: review and suggestions, *Behaviormetrika*, vol. 44, no. 1, pp. 193-225, 2017.

Hyvarinen, A., Pairwise measures of causal direction in linear non-gaussian acyclic models, *Proceedings of 2nd Asian Conference on Machine Learning*, 2010.

Hyvärinen, A. and Smith, S. M., Pairwise likelihood ratios for estimation of non-Gaussian structural equation models, *Journal of Machine Learning Research: JMLR*, vol. 14, pp. 111, 2013.

Jiang, Q., Yan, X. and Huang, B., Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes, *Industrial & Engineering Chemistry Research*, vol. 58, no. 29, pp. 12899-12912, 2019.
Jin, F., Lv, Z., Li, M., Mou, L., Zhao, J. and Wang, W., A causal model-based scheduling approach for coke oven gas system in steel industry, *IFAC-PapersOnLine*, vol. 51, no. 21, pp. 7-12, 2018.
Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D. and Sebag, M., Structural agnostic modeling: Adversarial learning of causal graphs, *arXiv preprint arXiv:1803.04929*, 2018.
Keele, L., The statistics of causal inference: A view from political methodology, *Political Analysis*, vol. 23, no. 3, pp. 313-335, 2015.
Lai, P.-C. and Bessler, D. A., Price discovery between carbonated soft drink manufacturers and retailers: a disaggregate analysis with PC and LiNGAM algorithms, *Journal of Applied Economics*, vol. 18, no. 1, pp. 173-197, 2015.
Le, T. D., Hoang, T., Li, J., Liu, L., Liu, H. and Hu, S., A fast PC algorithm for high dimensional causal discovery with multi-core PCs, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 5, pp. 1483-1495, 2016.
Menegozzo, G., Dall'Alba, D. and Fiorini, P., Causal interaction modeling on ultra-processed food manufacturing, *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, 2020.
Nadim, K., Ragab, A. and Ouali, M.-S., Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining, *Journal of Intelligent Manufacturing*, pp. 1-27, 2022.
Profillidis, V. A. and Botzoris, G. N., Modeling of transport demand: Analyzing, calculating, and forecasting transport demand, *Elsevier*, 2018.
Ramsey, J., Zhang, J. and Spirtes, P. L., Adjacency-faithfulness and conservative causal inference, *arXiv preprint arXiv:1206.6843*, 2012.
Schmidt, C., Huegle, J. and Uflacker, M., Order-independent constraint-based causal structure learning for gaussian distribution models using GPUs, *Proceedings of the 30th International Conference on Scientific and Statistical Database Management*, 2018.
Schwarz, G., Estimating the dimension of a model, *The Annals of Statistics*, pp. 461-464, 1978.
Shen, X., Ma, S., Vemuri, P. and Simon, G., Challenges and opportunities with causal discovery algorithms: application to Alzheimer's pathophysiology, *Scientific Reports*, vol. 10, no. 1, pp. 1-12, 2020.
Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., . . . Woolrich, M. W., Network modelling methods for fMRI, *Neuroimage*, vol. 54, no. 2, pp. 875-891, 2011.
Spirtes, P. and Scheines, R., Causal inference of ambiguous manipulations, *Philosophy of Science*, vol. 71, no. 5, pp. 833-845, 2004.
Wong, T. C., Law, K. M., Yau, H. K. and Ngan, S.-C., Analyzing supply chain operation models with the PC-algorithm and the neural network, *Expert Systems with Applications*, vol. 38, no. 6, pp. 7526-7534, 2011.
Yuan, Z., Qin, W. and Zhao, J., Smart manufacturing for the oil refining and petrochemical industry, *Engineering*, vol. 3, no. 2, pp. 179-182, 2017.
Zhang, J. and Mayer, W., Weakening faithfulness: some heuristic causal discovery algorithms, *International Journal of Data Science and Analytics*, vol. 3, no. 2, pp. 93-104, 2017.

Biography

Aishwary Pawar is a statistician, and educator at Southern Methodist University. He holds a PhD in Industrial Engineering with a focus on Operations and Research from the University of Michigan. He teaches Introduction to Data Science and applies advanced data analytics to support academic research and institutional initiatives. His expertise spans data-driven decision-making, student success analytics, and educational program evaluation. His research focuses on leveraging data science to improve learning outcomes and access to higher education. He has worked on projects analyzing first-year retention, financial aid impact, and enrollment trends. Beyond academia, he is passionate about Python programming and mentoring aspiring developers. He actively contributes to the Python community by guiding students and early-career professionals through coding, data analysis, and problem-solving. His teaching emphasizes hands-on learning, equipping students with practical skills for real-world applications.