

Towards a Trustworthy Causal Decision-Support Framework for Wastewater Treatment

ZhiJun Cheng^{1,*}, Måns Kjellander^{2,†} and Esteban Guerrero^{1,†}

¹Department of Computing Science, Umeå University, Sweden

²Umeå Energi AB, Sweden

Abstract

Understanding how process variables influence one another is central to effective wastewater treatment, yet extracting directed causal relationships from operational sensor data remains challenging. This paper presents a framework that learns causal structure from plant sensor time series using VARLiNGAM, which exploits non-Gaussianity to infer edge directions without requiring expert-specified network topology. The discovered causal graph is then used to parameterize a Dynamic Bayesian Network (DBN) and is integrated into an interactive web-based visualization system. Applied to 19 sensor variables collected over two years from a municipal treatment plant, the approach yields a directed graph whose principal pathways appear consistent with expected process behavior. The interactive interface enables operators to trace causal connections over time, aiming to support anomaly diagnosis and deeper process understanding. As future work, we outline a possibilistic reasoning extension that would propagate epistemic states to a declarative Answer Set Programming (ASP) knowledge base for rule-based decision support. Overall, this work provides an initial demonstration of how causal discovery and interactive visualization could support exploratory monitoring in wastewater treatment.

Keywords

Causal Discovery, Dynamic Bayesian Networks, VARLiNGAM, Trustworthy AI, Answer Set Programming, Wastewater Treatment

1. Introduction

Modern wastewater treatment plants (WWTPs) routinely record dozens of physical, chemical, and biological parameters at sub-hourly intervals, generating thousands of multivariate observations per day. Yet this wealth of sensor data does not automatically translate into process understanding: operators still struggle to determine how variables influence one another, especially when tracing the root cause of an operational anomaly [1].

Bayesian networks (BNs) and their temporal extension, Dynamic Bayesian Networks (DBNs) [2], have been widely adopted for WWTP modeling because they represent probabilistic dependencies in an interpretable directed graph. Prior work has applied BNs and DBNs to predict treatment performance [3], forecast effluent quality [4], monitor plant-wide processes [5], and diagnose fault propagation [6]. A common thread across these studies is the use of probabilistic graphical models to capture process dynamics while retaining interpretability—yet the way the network structure is obtained varies significantly and has important implications for the reliability of the resulting model.

However, in most of these previous approaches the DBN structure was either specified by domain experts or learned through score-based and correlation-based methods that do not guarantee causal interpretability. Expert-designed structures are prone to subjective bias and may overlook latent or indirect dependencies, while data-driven scoring methods find a statistically well-fitting graph but cannot distinguish cause from effect without additional assumptions [7]. Moreover, existing systems [3, 4]

RuleML+RR'26: Companion Proceedings of the 10th International Joint Conference on Rules and Reasoning, August 24–26, 2026, Vilnius, Lithuania

*Corresponding author.

[†]This project is supported by VINNOVA under the project: “A trustworthy decision-support system for energy management at Umeå Energi”.

✉ chengzhijun1024@gmail.com (Z. Cheng); esteban.guerrero@umu.se (E. Guerrero)

id 0009-0005-8381-4927 (Z. Cheng); 0000-0002-6035-800X (E. Guerrero)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

have generally focused on prediction accuracy rather than providing operators with tools to explore the learned causal relationships.

To address these limitations, this paper contributes an end-to-end pipeline that produces an interpretable causal model directly from plant data, without requiring experts to specify the network structure. The pipeline integrates three stages (Fig. 1): (1) data-driven causal discovery with VAR-LiNGAM, which exploits temporal ordering and non-Gaussianity to identify directed effects from sensor time-series; (2) DBN inference over the discovered structure to compute conditional state probabilities; and (3) an interactive web-based interface through which operators can explore the causal model and investigate the factors behind past observations. In addition, we outline a possibilistic reasoning extension [8] that would derive possibility and necessity measures to distinguish feasibility, certainty, conflict, and ignorance—distinctions that a single probability distribution cannot capture—and propagate the resulting epistemic states to a declarative Answer Set Programming (ASP) knowledge base for rule-based decision support.

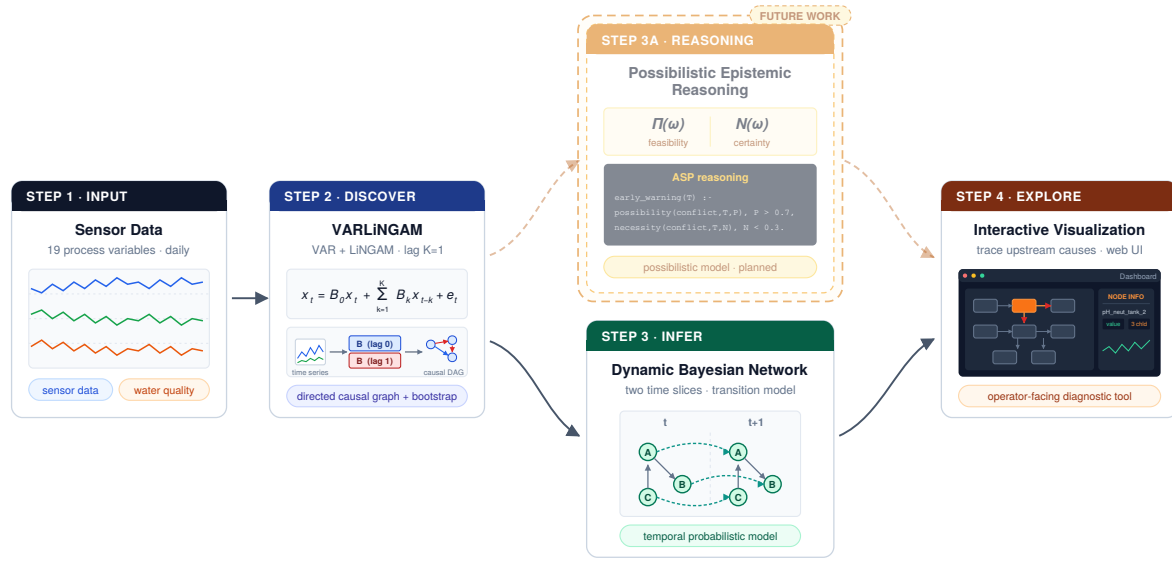


Figure 1: Pipeline of the proposed framework. Sensor telemetry is first processed by VARLiNGAM to identify causal dependencies between process variables. The discovered causal structure supports two complementary reasoning branches. The implemented branch constructs a Dynamic Bayesian Network (DBN) for temporal probabilistic inference and visualization. The proposed future extension (dashed) introduces a possibilistic epistemic reasoning layer that derives possibility and necessity measures, identifies conflict and ignorance states, and propagates these states to an ASP knowledge base for rule-based reasoning. Both branches are designed to contribute diagnostic information to the interactive visualization system.

2. Methodology

Wastewater treatment plant data. The data were obtained from a municipal wastewater treatment plant and span the years 2023–2024. The raw dataset contained a large number of sensor channels, from which 19 process variables were selected based on equipment identifiers provided by the plant operator. The 19 channels were resampled to daily means and standard deviations (38 features); after removing six near-constant features, adding 14 laboratory water-quality parameters, and filtering highly correlated variables (Pearson > 0.95), 40 variables remained for analysis. Finally, Shapiro–Wilk [9] and Jarque–Bera [10] tests indicated that all retained variables deviate from normality at the 0.05 significance level (Fig. 2), satisfying the assumptions of the causal discovery method described next.

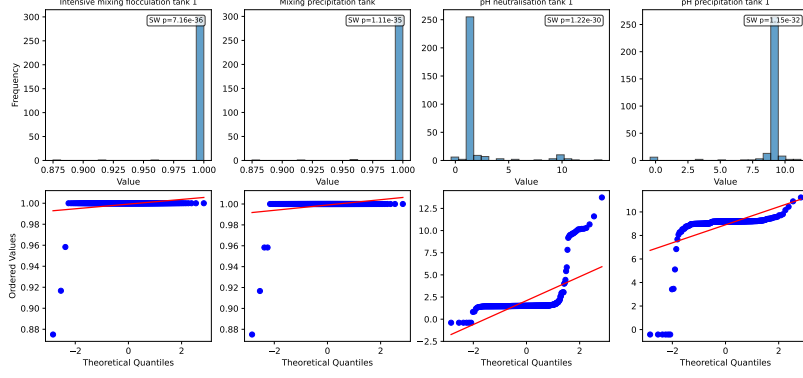


Figure 2: Histograms (top) and Q-Q plots (bottom) for four representative process variables. All variables reject the Shapiro-Wilk null hypothesis of normality at $p < 0.05$.

VARLiNGAM for causal discovery. The causal structure among process variables was identified using VARLiNGAM [11], which combines Vector Autoregression (VAR) with the Linear Non-Gaussian Acyclic Model (LiNGAM) [12]. Constraint-based methods such as the PC algorithm [7] rely on conditional independence tests, which can only recover an equivalence class of graphs. VARLiNGAM instead exploits the non-Gaussianity of residuals to determine edge directions uniquely. Because the sensor data in this study exhibits non-Gaussian characteristics, the identifiability assumption of VARLiNGAM is naturally satisfied.

Given a set of p observed variables $\mathbf{x}_t$ at time t , VARLiNGAM assumes:

$$\mathbf{x}_t = B_0 \mathbf{x}_t + \sum_{k=1}^K B_k \mathbf{x}_{t-k} + \mathbf{e}_t \quad (1)$$

where B_0 captures instantaneous causal effects (with zeros on the diagonal), B_k for $k \geq 1$ captures lagged effects, and $\mathbf{e}_t$ is a vector of mutually independent non-Gaussian disturbances [11]. The maximum lag was set to $K = 1$, as the data was resampled to daily resolution and the limited sample size makes higher-order lags prone to overfitting. Under the non-Gaussianity assumption, the model identifies the full causal ordering rather than an equivalence class.

To assess robustness, bootstrap resampling was performed with 100 iterations. Each bootstrap sample was fitted independently, and an edge was counted as present if its coefficient was non-negligible. An edge from the main fit was retained only if (1) it appeared in more than 50% of bootstrap samples, ensuring statistical stability, and (2) its absolute coefficient in the main fit fell between 0.2 and 1.0, where the lower bound excludes effects weaker than a moderate standardized effect size and the upper bound removes coefficients that may reflect collinearity rather than genuine causal influence.

Dynamic Bayesian network. A Dynamic Bayesian Network (DBN) extends standard Bayesian networks to model temporal processes by relating variables across adjacent time slices [2]. A DBN is defined by an initial distribution $P(\mathbf{X}_0)$ and a transition model $P(\mathbf{X}_{t+1} | \mathbf{X}_t)$, both encoded as directed acyclic graphs with associated conditional probability tables. The causal graph from VARLiNGAM was used to define the DBN structure. Each edge $(x_i^{(t)}, x_j^{(t+k)})$, where $k \in \{0, 1\}$, became a directed edge in the DBN. Temporal self-edges $(x_i^{(0)}, x_i^{(1)})$ were added for all variables, capturing temporal persistence—the expectation that each variable’s current state is influenced by its own previous state. Before parameter estimation, continuous sensor readings were discretized into three levels—*Low*, *Medium*, and *High*—using quantile-based binning so that each bin contained approximately the same number of observations. For variables with insufficient variation to support three bins, the number was reduced to two or one. Conditional probability distributions were then learned from the discretized historical data using Maximum Likelihood Estimation (MLE) via the pgmpy library [13].

Interactive visualization system. To allow users to explore the learned causal model, a web-based visualization system was developed. It renders the causal graph as an interactive directed graph, in which users can switch between temporal views, highlight causal paths, inspect individual node states, and consult the historical sensor readings behind each node.

3. Results

Causal discovery results. The full causal graph is dense, with 186 directed edges over 40 variables, and is difficult to interpret as a whole. Figure 3 therefore shows a representative 12-node, 11-edge subgraph revealing two distinct pathways: a mixing–precipitation chain and a pH treatment chain.

In the mixing–precipitation pathway, the mean of intensive mixing tank 1 was strongly associated with that of the mixing–precipitation tank (coefficient = 0.97). Its variability further propagated downstream, influencing both intensive mixing tank 2 (0.89) and the mixing–precipitation tank (0.92). In addition, variability in CaCO_3 dosing was linked to pH in the neutralisation tanks (0.20), consistent with its role as a neutralizing agent and its expected impact on downstream pH.

In the pH pathway, variability in neutralisation tank 2 acted as a hub, driving downstream variability in three locations: the precipitation tank 2 (0.34), neutralisation tank 1 (0.73), and precipitation tank 1 (0.78). This pattern suggests that pH fluctuations originating in the neutralisation stage propagate through subsequent precipitation units.

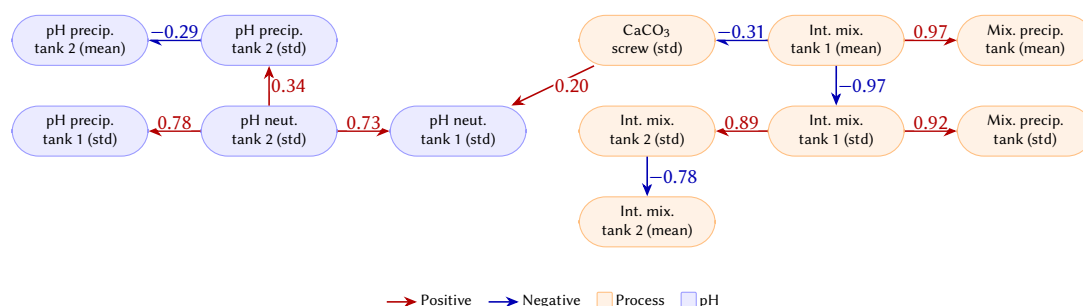


Figure 3: Causal graph considering 12 variables of the water treatment plant dataset. Results indicate that there are two distinct pathways, a mixing–precipitation chain and a pH treatment chain. The arrows indicate positive (red), and negative (blue) influence of the variables.

System demonstration. Figure 4 illustrates the web-based visualization interface. The selected node (highlighted in blue) displays its causal parents and children, while the right-hand panel presents its historical sensor values and associated causal connections.

4. Discussion and Conclusions

This paper presented a framework for integrating VARLiNGAM-based causal discovery, Dynamic Bayesian Network inference, and interactive visualization for wastewater treatment modeling, and outlined a possibilistic epistemic reasoning extension as future work. Applied to 19 sensor variables from a municipal plant over two years of daily data, VARLiNGAM yields a directed graph—learned purely from data, without a predefined topology—in which two representative pathways emerge: a mixing–precipitation chain and a pH treatment cascade, both appearing consistent with expected process behavior.

A practical observation is that exposing the learned causal structure, rather than predictions alone, changes how users can interact with the system. Where existing tools flag an abnormal reading but leave its cause to operator experience, the causal graph lets them trace upstream along directed edges—for an unusual pH value, to the responsible dosing or mixing stage—turning prediction into diagnosis.

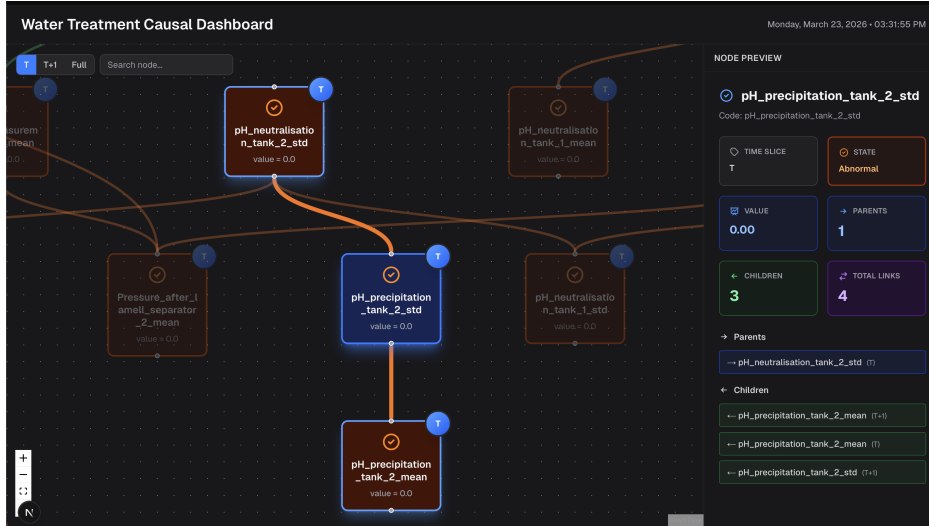


Figure 4: Interactive visualization interface. The selected node (highlighted in blue) shows its causal parents and children. The right panel displays the node’s sensor value and connection details.

Overall, combining data-driven causal discovery with interactive visualization keeps the reasoning behind each suggested cause explicit and open to inspection, a precondition for trustworthy decision support in wastewater treatment.

Limitations. Two limitations remain. VARLiNGAM assumes linearity, so nonlinear effects (e.g., saturation in biological stages) are missed; nonlinear methods such as neural Granger causality [14] are a natural extension. We also report no quantitative comparison with alternative structure learning methods or predictive accuracy metrics; systematic benchmarking [15] is planned next, in particular against DYNOTEARS [16], a more recent method that learns DBN structures directly from time series.

Future work. While the DBN models temporal probabilistic dependencies between process variables, *possibility theory* [8] provides an alternative representation of uncertainty focused on feasibility, certainty, and ignorance. We envision a possibilistic reasoning layer operating in parallel with the DBN, allowing epistemic states to be derived from the causal model and subsequently propagated to a symbolic reasoning component. Given the set $\Omega = \{\omega_1, \dots, \omega_n\}$ of operational states inferred from the DBN and the posterior probability $P(\omega_i | X_t)$ of state ω_i given the telemetry window X_t observed at time t , this layer would derive a possibility measure $\Pi(\omega_i) = P(\omega_i | X_t)$, expressing the compatibility of a state with the observed telemetry, and a necessity measure $N(\omega_i) = 1 - \max_{j \neq i} P(\omega_j | X_t)$, quantifying the degree to which competing alternatives are excluded. Two configurable thresholds $\tau_\Pi, \tau_N \in (0, 1]$ would control the strictness of epistemic state identification: a *conflict* state arises when $\Pi(\omega_i) \geq \tau_\Pi$ and $N(\omega_i) < \tau_N$, indicating that the system operates under several competing explanations, whereas an *ignorance* state arises when $\sup_{\omega_i \in \Omega} \Pi(\omega_i) < \tau_\Pi$, indicating that the observed telemetry is insufficiently compatible with all known operational states. The two thresholds thus act as epistemic sensitivity parameters, controlling how conservative the framework is when identifying ambiguous or previously unobserved operational conditions. The symbolic reasoning component would be realized using ASP [17], encoding expert operational knowledge provided by plant engineers. The possibilistic outputs would be translated into symbolic facts of the form `possibility(state, time, value)` and `necessity(state, time, value)`, and operational alerts would be inferred through declarative rules such as:

```
early_warning(T) :- possibility(conflict,T,P), P>0.7, necessity(conflict,T,N), N<0.3.
```

The resulting symbolic alerts and epistemic states would be propagated to the interactive visualization system. We plan to implement this layer and validate it on operational plant data.

Acknowledgments The authors thank Joris Leipus, Elias Niko, and Hugo Mater for their contributions to this work.

Declaration on Generative AI

During the preparation of this work, the authors used Claude in order to: Grammar and spelling check, Improve writing style, Citation management, Formatting assistance, and Peer review simulation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] J. Loy-Benitez, S. Tariq, H. T. Nguyen, S. Heo, C. Yoo, Sludge bulking monitoring in industrial wastewater treatment plants through graphical methods: A dynamic graph embedding and Bayesian networks approach, *Journal of Environmental Management* 344 (2023) 118804.
- [2] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning* [PhD dissertation], University of California, Berkeley, 2002.
- [3] D. Li, H. Yang, X. Liang, Prediction analysis of a wastewater treatment system using a Bayesian network, *Environmental Modelling & Software* 40 (2013) 140–150.
- [4] Y. Liu, Y. Pan, Z. Sun, D. Huang, Modeling of wastewater treatment processes using dynamic Bayesian networks based on fuzzy PLS, *IEEE Access* 8 (2020) 92129–92140.
- [5] Q. Chen, J. Zhang, Y. Zheng, A systematic monitoring framework utilizing dynamic Bayesian network in plant-wide processes, *IEEE Transactions on Instrumentation and Measurement* 73 (2024) 1–13.
- [6] J. Yu, M. M. Rashid, A novel dynamic Bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis, *AIChE Journal* 59 (2013) 2348–2365.
- [7] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction, and Search*, 2nd ed., MIT Press, 2000.
- [8] D. Dubois, H. Prade, *Possibility Theory - An Approach to Computerized Processing of Uncertainty*, Springer, 1988. URL: <https://doi.org/10.1007/978-1-4684-5287-7>. doi:10.1007/978-1-4684-5287-7.
- [9] S. S. Shapiro, M. B. Wilk, An analysis of variance test for normality (complete samples), *Biometrika* 52 (1965) 591–611.
- [10] C. M. Jarque, A. K. Bera, A test for normality of observations and regression residuals, *International Statistical Review* 55 (1987) 163–172.
- [11] A. Hyvärinen, K. Zhang, S. Shimizu, P. O. Hoyer, Estimation of a structural vector autoregression model using non-Gaussianity, *Journal of Machine Learning Research* 11 (2010) 1709–1731.
- [12] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, A linear non-Gaussian acyclic model for causal discovery, *Journal of Machine Learning Research* 7 (2006) 2003–2030.
- [13] A. Ankan, A. Panda, pgmpy: Probabilistic graphical models using Python, in: *Proceedings of the 14th Python in Science Conference (SciPy)*, 2015, pp. 6–11.
- [14] A. Tank, I. Covert, N. Foti, A. Shojaie, E. B. Fox, Neural Granger causality, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (2022) 4267–4279.
- [15] M. Scutari, C. E. Graafland, J. M. Gutiérrez, Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms, *International Journal of Approximate Reasoning* 115 (2019) 235–253.
- [16] R. Pamfil, N. Sriwattanaworachai, S. Desai, P. Pilgerstorfer, K. Georgatzis, P. Beaumont, B. Aragam, DYNOTEARS: Structure learning from time-series data, in: *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 1595–1605.
- [17] M. Gelfond, V. Lifschitz, The stable model semantics for logic programming., in: *ICLP/SLP*, volume 88, Cambridge, MA, 1988, pp. 1070–1080.