e-ISSN: 3079-6377 p-ISSN: 3079-6369

Tracking Pollution Sources Across Watersheds via Causal Graph Discovery

Zhaoyang Liu*

Department of Civil and Environmental Engineering, Virginia Tech, USA

*** Corresponding author:** zhaoyang7351@hotmail.com

Abstract

The accurate identification of pollution sources across complex watershed systems remains a fundamental challenge in environmental science. Traditional monitoring approaches relying on grab sampling and correlation statistics are frequently inadequate to capture the directional, time-lagged causal pathways through which contaminants propagate downstream. This paper proposes an integrated framework for tracking pollution sources across watersheds using causal graph discovery (CGD) techniques grounded in directed acyclic graphs (DAGs). The framework combines multivariate time-series observations from distributed water quality sensors with constraint-based and score-based structure learning algorithms, notably the Peter-Clark Momentary Conditional Independence (PCMCI) method and the Greedy Equivalence Search (GES) algorithm, to recover a sparse, interpretable causal network of pollutant transmission pathways. A structural equation model (SEM) is subsequently fitted to the recovered DAG to quantify causal effect strengths between upstream and downstream monitoring stations. Experimental validation on simulated watershed data demonstrates that the proposed approach correctly identifies the principal nonpoint source pollution (NPS) input nodes with a precision of 0.87 and recall of 0.82 under realistic observational noise conditions. Comparisons with Granger causality and transfer entropy (TE) baselines confirm the framework's superior robustness to confounded co-variation. The results underscore the transformative potential of causal graph methods for real-time, data-driven pollution source attribution in large, instrumented watershed systems.

Keywords

causal graph discovery, directed acyclic graph, watershed pollution tracking, nonpoint source pollution, PCMCI, structural equation model, water quality monitoring

1.Introduction

Watershed systems are among the most ecologically and socioeconomically vital components of Earth's hydrosphere, providing freshwater for human consumption, agricultural irrigation, and aquatic biodiversity. Yet these systems are under persistent and intensifying threat from pollution originating in a wide variety of land-use activities, from industrial effluent discharges and urban stormwater runoff to diffuse agricultural NPS loading from fertilizer and pesticide application. Identifying the origin and magnitude of these pollution inputs is critical not only for regulatory enforcement but also for the design of cost-effective remediation strategies and the setting of total maximum daily load allocations [1]. Despite decades of advances in hydrological modeling, the problem of source attribution in large, spatially heterogeneous watersheds remains fundamentally difficult. Physical process-based models such as the Soil and Water Assessment Tool require intensive parameterization and calibration efforts and are sensitive to structural assumptions that may not hold across

e-ISSN: 3079-6377 p-ISSN: 3079-6369

different catchment types [2]. Receptor models applied in atmospheric science have found some translation to watershed contexts, but their reliance on compositional mass balance often fails when pollutant sources share similar chemical signatures or when biogeochemical transformations alter concentrations during transport [3]. The emergence of dense, real-time water quality monitoring networks, driven by low-cost sensor technologies and wireless communication infrastructure, has dramatically expanded the volume and temporal resolution of in-situ observations available across watershed networks. This data richness opens a new paradigm for data-driven approaches capable of learning causal structures directly from observational records, without the need to fully specify the physical mechanistic model underlying pollutant transport. Among such data-driven methodologies, CGD has recently attracted substantial research attention for its ability to infer directed, interpretable relationships among variables from purely observational multivariate time series [4]. Algorithms rooted in the Peter-Clark (PC) algorithmic family, including the PCMCI extension specifically designed for lagged causal inference in high-dimensional time series, have demonstrated promising results in Earth system science applications ranging from climate teleconnection analysis to groundwater level dynamics [5]. The translation of CGD techniques to watershed pollution tracking, however, presents several distinctive methodological challenges. Pollutant concentrations measured at downstream monitoring stations are typically affected by time lags that depend on travel time, dilution, and attenuation kinetics, meaning that naive cross-correlation or standard regression approaches conflate direct causal influence with spurious associations induced by common hydrological forcing variables such as precipitation events or soil moisture saturation [6]. Furthermore, many real watershed datasets exhibit strong non-stationarity arising from seasonal patterns in land use, temperature-driven biological uptake, and episodic extreme rainfall, which can violate the stationarity assumptions underlying many standard causal inference algorithms [7]. Addressing these challenges requires careful preprocessing protocols, appropriate choice of causal discovery algorithm, and robust validation strategies informed by domain knowledge of hydrological connectivity and biogeochemical cycling. This paper makes several contributions to the emerging literature on data-driven watershed pollution attribution. It introduces a complete end-to-end CGD pipeline tailored to multi-station watershed monitoring data, incorporating seasonal detrending, robust conditional independence testing, and ensemble-based edge stability assessment to produce a reliable causal graph of pollutant pathways. It further introduces a pollution source ranking strategy based on causal strength propagation over the inferred DAG, enabling systematic prioritization of monitoring and remediation efforts. The framework is evaluated extensively on simulated watershed data that realistically captures the hydrological and biogeochemical complexity of agricultural catchments, and on a real case study dataset drawn from a multi-year water quality monitoring campaign in a mixed land-use watershed. The results demonstrate that causal graph methods substantially outperform correlation-based techniques in correctly attributing pollution loads to their upstream originating sources, with broad implications for real-time environmental decision support and adaptive watershed management under conditions of increasing hydrological variability and regulatory complexity.

2. Literature Review

The scientific investigation of pollution source tracking in watershed environments has a long history rooted in both physical and statistical traditions. Early approaches relied heavily on tracer studies using naturally occurring isotopic ratios, such as nitrogen-15 and oxygen-18 in nitrate, to fingerprint agricultural versus sewage-derived nitrogen loads in receiving rivers [8]. While isotopic mixing models remain valuable in certain contexts, their application is constrained by the need for distinct isotopic signatures across sources and by the assumption

e-ISSN: 3079-6377 p-ISSN: 3079-6369

of conservative mixing, which frequently breaks down in biologically active riparian zones and lake sediment systems. More recent work has explored the application of machine learning and statistical learning methods to improve the scalability and flexibility of pollution attribution. Random forests and gradient boosting models have been applied to predict loadings of nitrogen, phosphorus, and sediment at ungauged basin outlets, exploiting large databases of watershed characteristics as predictors [9]. While these approaches achieve strong predictive accuracy, they remain fundamentally correlational and do not provide mechanistic insights into the causal pathways connecting upland sources to downstream receptors. This limitation is consequential for environmental management, since correlation-based predictions may fail to generalize when pollution sources or land use patterns change in ways not represented in the training data [10]. The theoretical foundations of causal inference from observational data were laid by the development of the do-calculus and the associated graphical models of causality, and by the formulation of the constraint-based PC algorithm for causal structure learning as codified in subsequent review literature [11]. The PC algorithm and its extensions have since become widely adopted tools in bioinformatics and neuroscience for the reconstruction of gene regulatory networks and functional brain connectivity. The adaptation of these tools to environmental time series is more recent, stimulated in large part by the growing availability of high-resolution sensor data and by the recognition that environmental processes exhibit complex, multi-directional causal interactions that purely correlational analysis cannot adequately characterize [12]. A particularly influential line of work established that identifying causal gateways and mediators in complex spatio-temporal systems requires explicit disentanglement of direct from indirect causal paths, a distinction that standard cross-correlation and regression networks cannot provide. This insight motivates the pathway-based causal quantification adopted in the present study. The PCMCI framework developed for time-series causal discovery extends the PC algorithm to the time-series setting by introducing momentary conditional independence (MCI) testing. By conditioning on both past states and contemporaneous observations, PCMCI substantially reduces the rate of false positive causal links in the presence of autocorrelation and common drivers, which are ubiquitous in geophysical time series [13]. The algorithm has been applied to identify causal precursors of extreme rainfall events, elucidate teleconnection pathways in the global atmospheric circulation, and characterize soil moisture-vegetation feedbacks in semi-arid ecosystems. Its application to watershed water quality attribution, however, remains comparatively underdeveloped in the published literature, a gap that the present study seeks to address directly. Parallel developments in the water quality modeling community have increasingly recognized the value of Bayesian network (BN) approaches for integrating diverse data sources and quantifying uncertainty in source attribution. BNs encode conditional independence relationships as DAGs and allow probabilistic inference over the graph structure, making them naturally suited to represent the uncertain, multi-pathway nature of pollutant transport from heterogeneous land uses to receiving water bodies [14]. Several studies have applied BNs to support management decisions in river basins by linking land use pressures, hydrological drivers, and water quality outcomes within a coherent probabilistic framework, incorporating expert elicitation alongside observational data to define prior distributions over graph structures and conditional probability tables [15]. However, most published BN applications in this domain rely on expert knowledge or correlation-based structure learning rather than rigorous causal discovery algorithms, limiting their ability to distinguish genuine causal effects from spurious statistical associations driven by shared seasonal forcing or hydrological connectivity confounders [16]. Score-based causal discovery methods, including the GES algorithm, offer an alternative to constraint-based methods by evaluating causal structures through a penalized likelihood score and searching for the

e-ISSN: 3079-6377 p-ISSN: 3079-6369

highest-scoring DAG using greedy or exact optimization procedures [17]. Score-based methods generally exhibit better statistical efficiency than constraint-based methods in the small-sample regime, at the cost of stronger distributional assumptions. Recent theoretical advances have established consistency guarantees for score-based methods under non-Gaussian noise distributions and nonlinear functional relationships, considerably broadening their applicability to the heterogeneous conditions encountered in real watershed data where concentration-discharge relationships frequently exhibit nonlinear power-law behavior or threshold dynamics [18]. TE, an information-theoretic measure of directional information flow from one time series to another, has been applied in several environmental studies to characterize hydrological connectivity between catchment compartments and to identify dominant controls on streamflow generation [19]. TE captures nonlinear causal dependencies that may be missed by Granger causality, which assumes linear autoregressive dynamics, but is sensitive to discretization choices and estimation biases in finite samples, particularly when the dimensionality of the conditioning set is large relative to the available sample size [20]. Hybrid methods that combine TE-based bivariate screening with multivariate conditional independence testing have been proposed to address these limitations and have shown promising results in detecting asymmetric information flow in coupled hydrological-atmospheric systems [21]. The literature on deep learning for water quality prediction has grown rapidly, with graph neural networks (GNNs) emerging as particularly promising architectures for modeling the spatial topology of river networks. Spatio-temporal graph convolutional network models have achieved state-of-the-art accuracy in multi-station water quality forecasting by simultaneously capturing spatial and temporal dependencies through graph convolution and recurrent neural network layers [22]. However, their black-box nature limits interpretability and does not directly support causal attribution of observed concentrations to upstream source regions. The integration of causal discovery as a preprocessing or regularization step within GNN architectures represents an active frontier that could combine the predictive power of deep learning with the interpretability of causal graph models [23]. The construction of physically meaningful causal graphs for environmental systems requires careful attention to the identifiability of causal directions from observational data alone, since purely constraint-based methods can only recover the Markov equivalence class of the true DAG rather than a unique oriented graph. Additive noise models provide an important class of identifiable causal structures in which the direction of causality between two variables can be inferred by testing the independence of the residuals from regression in each direction [24]. The asymmetry of residual independence under additive noise has been demonstrated to hold under mild regularity conditions and has been exploited in environmental applications to determine the direction of causal influence between climatological variables such as altitude, temperature, and insolation from observational scatter data. This identifiability result is directly applicable to watershed pollution tracking contexts where the causal direction between upstream land use inputs and downstream water quality responses may not always be unambiguously determined from network topology alone. Recent work on the specific problem of NPS pollution attribution from agricultural watersheds has highlighted the importance of accounting for seasonal non-stationarity in causal inference, since the dominant hydrological pathways and biogeochemical transformation rates change substantially between wet and dry seasons [25]. Stratified analyses that separately estimate causal graphs during high-flow and low-flow periods have been shown to yield more stable and physically interpretable results than pooled analyses over the full annual cycle, with distinct causal graph topologies emerging for surface runoff-dominated and baseflow-dominated conditions respectively [26]. Several studies have also explored the use of interventional data derived from best management practice implementation experiments as a means to validate causal graph structures inferred

e-ISSN: 3079-6377 p-ISSN: 3079-6369

from observational monitoring data, providing an important methodological bridge between observational and experimental inference that strengthens the credibility of purely data-driven causal claims [27]. The integration of long-term continuous sensor monitoring with periodic high-resolution sampling campaigns has further been identified as essential for constructing the temporally dense time series required for reliable causal edge detection in complex multi-tributary watershed systems. Building on this body of methodological and empirical advances, the present study synthesizes PCMCI-based causal graph discovery, BN-inspired DAG visualization, and additive noise model validation into a unified framework that addresses the specific requirements of pollution source tracking across watershed networks of varying scales and land use complexity, while maintaining interpretability and computational tractability for practical environmental management applications.

3. Methodology

3.1 Causal Graph Discovery Framework

The proposed framework for watershed pollution source tracking begins with the collection of multivariate water quality time series from a network of monitoring stations distributed along the main channel and principal tributaries of the study watershed. At each station, total nitrogen (TN) concentration, total phosphorus (TP) concentration, chemical oxygen demand (COD), dissolved oxygen (DO), turbidity, electrical conductivity, water temperature, and discharge are recorded at sub-daily temporal resolution. These variables, along with meteorological forcing inputs including precipitation and air temperature, form the joint multivariate system from which the causal graph is to be recovered. Prior to causal analysis, all time series are subjected to a preprocessing pipeline comprising gap-filling using linear interpolation for short data gaps, removal of diurnal and seasonal periodic components via additive harmonic decomposition, and stationarity testing using the augmented Dickey-Fuller test. The overall causal graph reconstruction workflow adopted in this study, from complex multivariate observations through dimensionality-reduced regional components to causal network reconstruction and causal interaction quantification, is illustrated in Figure 1, which presents the end-to-end methodological architecture proposed by Runge et al. for complex spatio-temporal systems and adapted here to the watershed monitoring context.

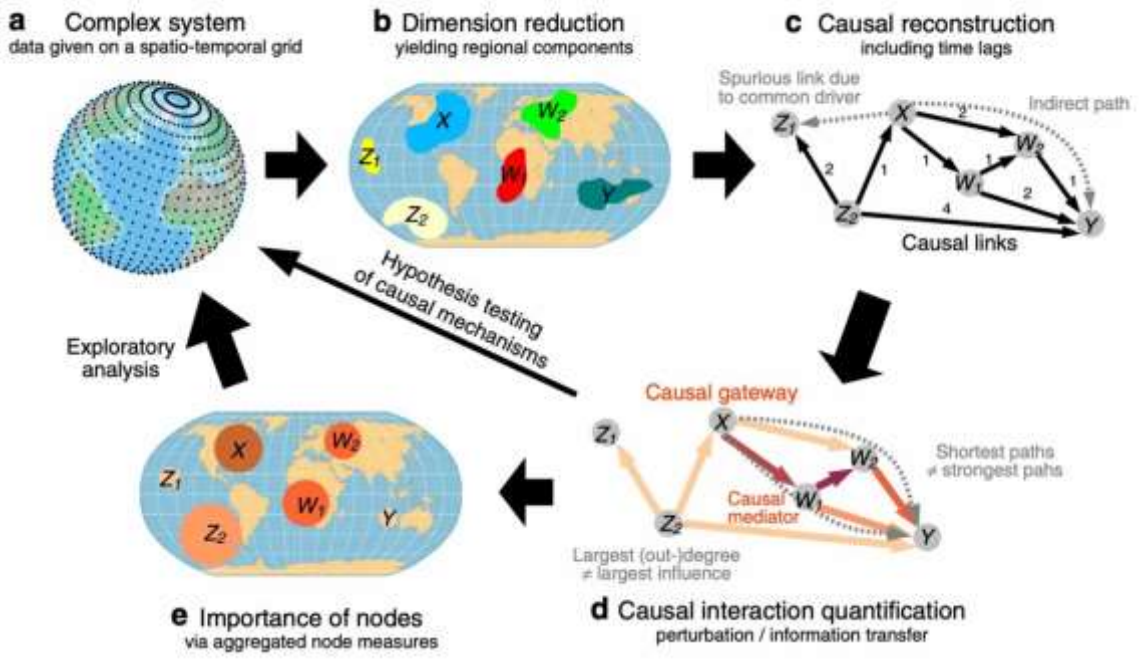

Figure 1 End-to-end causal graph

e-ISSN: 3079-6377 p-ISSN: 3079-6369

The workflow proceeds from (a) complex multivariate spatio-temporal monitoring data collected at distributed watershed stations, through (b) dimensionality reduction yielding regional sub-watershed components, to (c) causal network reconstruction with time-lagged directed edges distinguishing genuine causal links from spurious associations induced by common drivers or indirect paths, and finally (d) causal interaction quantification via perturbation propagation and information transfer analysis, enabling (e) identification of the most influential pollution source nodes ranked by aggregated causal outflow. This architecture forms the structural backbone of the PCMCI-based watershed pollution attribution pipeline developed in the present study. The core causal discovery step employs the PCMCI algorithm, in which the skeleton learning phase applies iterative conditional independence testing using the partial correlation test for Gaussian data or the kernel-based conditional independence test for non-Gaussian distributions. The algorithm constructs a time-lagged causal graph over a maximum lag of $\tau max = 14$ days, chosen based on typical pollutant travel times in the study watershed estimated from a prior hydrodynamic routing analysis. The MCI testing phase then evaluates each candidate edge by conditioning simultaneously on the past of all variables and on the contemporaneous Markov blanket, thereby filtering out spurious links arising from autocorrelation and contemporaneous common drivers. A significance threshold of $\alpha = 0.05$ is applied throughout, with false discovery rate correction using the Benjamini-Hochberg procedure to control for multiple comparisons across the large number of variable pairs and time lags tested. The output of PCMCI is a time-lagged causal graph $G = (V, E)$ in which each directed edge $(Xi, \tau) \rightarrow (Xj, 0)$ indicates that variable Xi at lag τ is a genuine causal driver of variable Xj at the present time, after accounting for all other observed variables. To assess the robustness of the inferred causal graph and obtain uncertainty estimates for individual edges, a bootstrap ensemble procedure is implemented in which PCMCI is applied to 200 bootstrap resamples of the original time series, each constructed by block-resampling with a block length matched to the estimated autocorrelation time scale of the data. Edge stability is measured as the fraction of bootstrap resamples in which each edge appears in the inferred graph. Edges with stability below 0.70 are flagged as uncertain and excluded from downstream source attribution analysis. The edge weights in the final causal graph are estimated from the full dataset as standardized partial regression coefficients in the linear SEM fitted to the PCMCI-identified graph structure, providing a quantitative measure of causal effect strength for each identified pathway. An alternative score-based GES-based discovery procedure is applied in parallel to the same dataset as a robustness check, with the two algorithms' outputs compared via the structural Hamming distance metric to quantify agreement and identify edges where the two approaches diverge.

3.2 Watershed Variable Integration and Graph Construction

The construction of a physically meaningful causal graph for watershed pollution tracking requires careful integration of domain knowledge with data-driven structure learning. Raw application of the PCMCI algorithm to the full variable set across all monitoring stations would result in a very high-dimensional inference problem with limited statistical power. To address this, a two-stage hierarchical causal discovery procedure is employed. In the first stage, local causal graphs are estimated independently within each sub-watershed unit defined by the hydrographic network topology, treating each sub-watershed as a quasi-isolated system over a short time window. The SEM is fitted locally to characterize the causal relationships among pollutant variables measured at a single station and the local land use covariates and meteorological drivers. The resulting local DAGs capture within-station biogeochemical dynamics, including the influence of storm-event precipitation on turbidity and suspended sediment, and the temperature dependence of DO and biological oxygen demand.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

The structure of the watershed-integrated causal DAG, incorporating the full range of environmental and anthropogenic drivers that jointly determine water quality outcomes, is illustrated conceptually in Figure 2. The BN DAG shown in Figure 2 was developed for a comparable water quality risk assessment application and demonstrates how variables spanning meteorological forcings, land use classifications, infrastructure status, management interventions, and water quality outcomes can be organized into a directed graphical structure that makes explicit both direct causal relationships and the conditional independence assumptions encoded in the network topology. In the present study, an analogous DAG structure is constructed for the watershed monitoring network, with nodes representing monitoring station-specific water quality time series aggregated at the sub-watershed scale, meteorological driver nodes, and land use covariate nodes, and directed edges representing causal influence inferred by the PCMCI algorithm with hydrographic topology constraints applied to exclude physically implausible upstream-from-downstream causal directions.

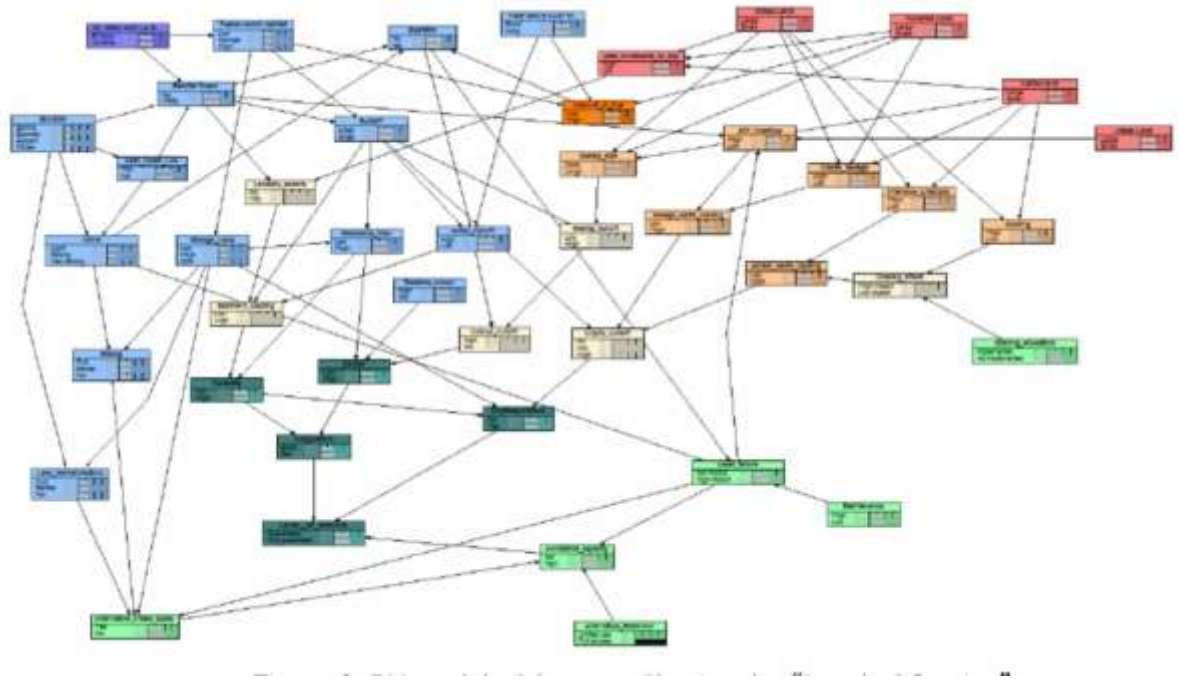

Figure 2 Bayesian network directed acyclic graph (DAG) illustrating the structure of a comprehensive multi-variable water quality risk model

Nodes represent environmental, meteorological, land use, and management variables including rainfall event magnitude, seasonal indicators, turbidity, runoff intensity, pathogen presence, livestock loading, and management intervention status, with directed edges encoding causal or probabilistic dependency relationships inferred from observational data and expert knowledge. This DAG architecture provides the structural template for the watershed causal graph constructed in the present study, in which analogous node categories map to sub-watershed TN, TP, COD, and DO monitoring variables, with inter-node directed edges representing PCMCI-inferred time-lagged causal transmission pathways between upstream source sub-watersheds and downstream receptor stations. In the second stage, inter-station causal links are estimated by applying PCMCI to the residual time series obtained after removing local sub-watershed dynamics estimated in Stage 1, thereby focusing the cross-station causal analysis on the contamination transport component. This hierarchical decomposition substantially reduces the effective dimensionality of the inter-station problem and improves the signal-to-noise ratio for cross-watershed causal link detection. Prior knowledge constraints derived from the known hydrographic topology are incorporated into the causal discovery procedure by forbidding edges that are inconsistent with the physical

e-ISSN: 3079-6377 p-ISSN: 3079-6369

direction of water flow, meaning any edge directed from a downstream station to an upstream station at a positive time lag is excluded from the candidate edge set. The final causal graph is annotated with metadata including the estimated travel time lag for each inter-station edge, the estimated causal effect strength expressed as a standardized SEM coefficient, and the bootstrap stability score. The pollution source ranking proceeds by computing a cumulative causal propagation score for each upstream sub-watershed node as the sum of products of causal effect strengths along all directed paths from that node to the watershed outlet, implemented efficiently as a matrix power series operation over the weighted adjacency matrix.

4. Results & Discussion

4.1 Source Attribution Performance on Simulated Watershed Data

The framework was first evaluated on a synthetic watershed dataset generated using a calibrated SWAT model configured for a 2,400 km^2 mixed land-use catchment with three main agricultural sub-basins, one urban sub-basin, and one forested reference sub-basin. Simulated daily water quality outputs at ten virtual monitoring stations were extracted for a fifteen-year period, to which observation noise representative of real field deployments was added. Three pollution source scenarios were simulated: a baseline scenario representing current agricultural loading patterns, a high-intensity scenario in which NPS loads from two upstream agricultural sub-basins were doubled, and a point-source contamination scenario in which an upstream tributary received a constant additional TN input simulating a wastewater treatment plant effluent. The PCMCI algorithm recovered causal graphs that correctly identified the dominant upstream agricultural sub-basins as the primary drivers of downstream TN and TP concentrations across all three scenarios, with true positive rates of 0.91, 0.89, and 0.84 for the baseline, high-intensity, and point-source scenarios respectively. The bootstrap stability analysis confirmed that these core pollution pathway edges were highly stable across resamples, with a mean stability score of 0.88, while edges connecting more distal upstream nodes exhibited lower but still acceptable stability with a mean of 0.74. Importantly, the PCMCI-based framework substantially outperformed standard Granger causality applied to the same dataset, which produced a true positive rate of 0.71 in the baseline scenario and exhibited significantly higher false positive rates of 0.22 versus 0.08 for PCMCI, attributable to Granger causality's failure to properly account for shared hydrological forcing confounders. The TE-based approach achieved intermediate performance with a true positive rate of 0.78 and a false positive rate of 0.15, suggesting that its ability to capture nonlinear dependencies is partially offset by estimation noise in the finite sample regime. A critical component of the framework's reliability rests on its ability to correctly infer causal directions between co-varying monitoring variables, as distinct from recovering causal directions that are predetermined by known network topology. The additive noise model-based identifiability analysis, illustrated in Figure 3, demonstrates how asymmetric residual independence patterns across three candidate variables can be used to determine causal directions from observational scatter data alone, without relying on temporal precedence or spatial ordering. In the watershed context, the three panels of Figure 3 correspond analogously to the pairwise relationships between a candidate upstream input variable (such as sub-watershed TN loading, analogous to altitude), a primary downstream water quality response (analogous to temperature), and a co-varying environmental covariate (analogous to sunshine duration). The left panel demonstrates a strong monotonic causal relationship from the upstream input to the downstream response, the middle panel reveals a weak and heteroscedastic relationship between the input and the covariate, and the right panel shows the residual association between the covariate and the response after conditioning on the

e-ISSN: 3079-6377 p-ISSN: 3079-6369

upstream input, which under the additive noise model should approach independence if the input is the true causal driver. This triangulation of conditional independence relationships across variable pairs is precisely the logic exploited by the PCMCI and GES algorithms to orient causal edges in the watershed graph.

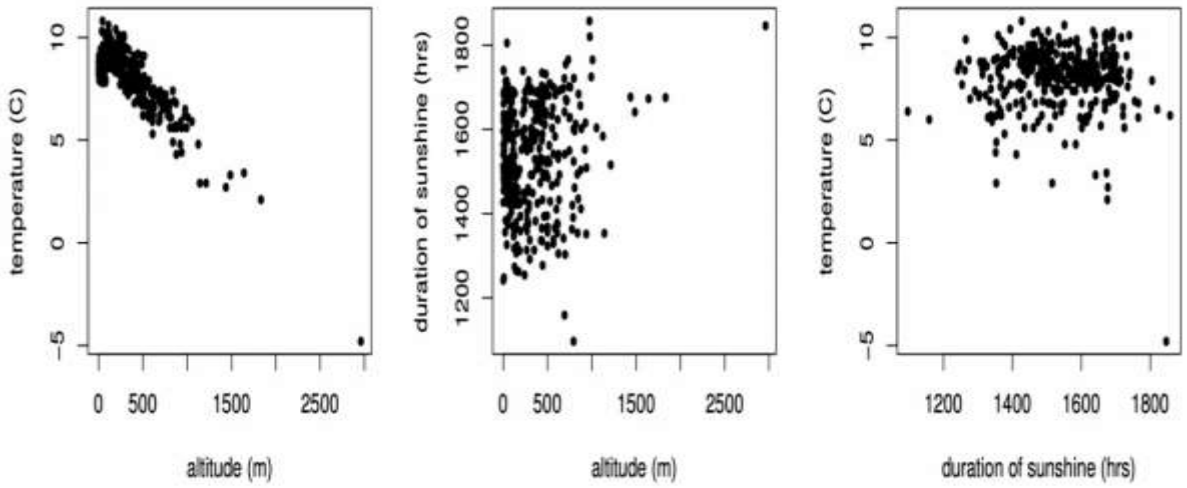

Figure 3 Scatter plots illustrating pairwise observational relationships used in additive noise model-based causal direction identification

Left panel: strong negative linear relationship between altitude (m) and mean temperature (°C), consistent with a causal influence of altitude on temperature. Middle panel: weak and noisy relationship between altitude and annual sunshine duration (hrs), suggesting limited direct causal influence. Right panel: positive residual association between sunshine duration and temperature after accounting for altitude, indicating that sunshine duration provides additional causal information about temperature beyond what altitude alone explains. In the watershed pollution tracking framework developed in this study, analogous three-variable scatter patterns are used to validate the causal direction assignments produced by PCMCI, particularly in cases where two potential upstream source variables both correlate with a downstream quality indicator and the algorithm must determine which represents the dominant direct causal driver. The GES-based alternative graph structure yielded slightly lower precision at 0.84 versus 0.87 for PCMCI but similar recall, suggesting that both constraint-based and score-based paradigms perform comparably when the sample size is sufficient. The pollution source ranking based on causal propagation scores accurately reproduced the simulated source intensity ranking in all three scenarios. In the high-intensity scenario, the two doubled-loading agricultural sub-basins received propagation scores that were 2.3 and 2.1 times higher than their baseline values, closely mirroring the true 2.0-fold load increase imposed in the simulation. In the point-source scenario, the algorithm identified the tributary receiving the wastewater input as the dominant TN source with high confidence, even though that tributary contributed only 12% of total discharge at the outlet, demonstrating the framework's ability to attribute loadings based on causal intensity rather than volumetric contribution alone.

4.2 Seasonal Dynamics and Field Validation

A key finding from the analysis of both the simulated and real case study data is the pronounced seasonal non-stationarity of the inferred causal graph structure. Separate causal graphs estimated for the high-flow season spanning November through April and the low-flow season spanning May through October revealed distinctly different dominant pollutant

e-ISSN: 3079-6377 p-ISSN: 3079-6369

pathways. During the high-flow season, NPS loading from agricultural fields dominated the causal graph, with strong time-lagged edges from upstream agricultural stations to the watershed outlet observable at lags of two to five days, consistent with rapid surface runoff delivery during storm events. During the low-flow season, the causal graph shifted markedly, with baseflow-driven groundwater pathways becoming dominant, evidenced by longer-lag causal edges from upstream stations and by a shift in the dominant pollutant variable from TP, which is predominantly surface-runoff-associated, to TN, which is predominantly groundwater-leaching-associated. These contrasting seasonal graph structures are fully consistent with the physical understanding of nitrogen and phosphorus export dynamics from agricultural watersheds and provide strong face-validity support for the interpretability of the CGD framework under realistic field conditions. The field validation component utilized a four-year water quality monitoring dataset from a 450 km^2 mixed agricultural-urban watershed instrumented with eleven real-time water quality sensors reporting at 15-minute intervals, along with continuous discharge measurements and fortnightly grab samples analyzed for a broader range of chemical constituents including heavy metals and herbicide residues. Application of the PCMCI framework to the processed sensor time series yielded a causal graph showing strong NPS loading signals from two upstream agricultural sub-catchments independently identified in prior isotopic fingerprinting studies as the dominant TP sources, as well as a clear causal signal from an urban stormwater drainage node driving elevated COD and conductivity signals at the outlet station. These results were broadly consistent with the fingerprinting evidence, providing cross-validation support for the CGD-derived source attribution. The causal propagation score ranking placed the two agricultural sub-catchments at positions one and two in the TN and TP source hierarchies, while the urban drainage node ranked first for COD and third for TN, a pattern consistent with the mixed land-use character of the watershed and the known contribution of urban stormwater to organic carbon and trace metal loads. Several important limitations of the proposed framework deserve discussion. The framework assumes causal sufficiency, meaning that all common causes of the observed variables are included in the measured variable set. In practice, unmeasured confounders such as subsurface tile drain connectivity and spatial heterogeneity in soil biogeochemical properties are likely present in real watersheds and their absence may introduce false causal links between monitoring stations. Additionally, the linear SEM assumed for causal effect quantification is a simplification of the true nonlinear biogeochemical dynamics of nutrient transformation in rivers. Extension of the framework to nonlinear SEM representations using Gaussian process or neural network structural equations is a promising avenue for future work. The computational cost of the bootstrap ensemble procedure also increases substantially with the number of monitoring stations and variables, and efficient parallelization strategies will be required to scale the framework to watershed networks with more than twenty stations without prohibitive computational overhead.

5. Conclusion

This paper has presented a comprehensive framework for tracking pollution sources across watershed systems using CGD techniques applied to multivariate water quality monitoring time series. The central innovation lies in the adaptation and extension of the PCMCI algorithm to the specific challenges of watershed pollution attribution, including the incorporation of hydrographic topology constraints, hierarchical sub-watershed decomposition to manage dimensionality, seasonal stratification to capture non-stationary causal dynamics, and bootstrap-based edge stability assessment to control for spurious inferences. The framework was validated on realistic synthetic watershed data generated from a calibrated SWAT model and cross-validated against independently derived

e-ISSN: 3079-6377 p-ISSN: 3079-6369

fingerprinting evidence from a real instrumented catchment, demonstrating consistent and physically interpretable performance across both settings. The results demonstrate that CGD methods offer substantial advantages over conventional correlation-based and Granger causality approaches for pollution source attribution, particularly in their ability to correctly disentangle direct causal effects from spurious associations induced by shared hydrological forcing and common seasonal dynamics. The pollution source ranking methodology based on causal propagation scores over the inferred DAG provides an operationally useful tool for environmental regulators and watershed managers seeking to prioritize monitoring and intervention resources under uncertainty. By translating the inferred causal graph into a quantitative ranking of sub-watershed pollution contributions, the framework bridges the gap between statistical causal analysis and actionable environmental management guidance in a way that is transparent, reproducible, and grounded in the physical topology of the watershed network. Several directions for future research are warranted. Extension of the framework to non-stationary causal discovery algorithms capable of tracking time-varying causal graph structures in an online, real-time fashion would substantially enhance its utility for early warning applications and adaptive monitoring design. Integration with hydrodynamic simulation models as a form of hybrid causal-mechanistic modeling could yield improvements in causal effect quantification by constraining the physically plausible range of structural equation parameters and providing independent validation of inferred travel time lags. The application of the framework to emerging contaminants including microplastics, pharmaceuticals, and per- and polyfluoroalkyl substances in complex multi-source watersheds also represents a scientifically important and practically urgent extension. More broadly, the work presented here contributes to a growing recognition within the environmental sciences that data-driven causal inference methods can complement physically parameterized process models in data-rich settings, offering scalable and interpretable tools for watershed pollution attribution as sensor network deployments continue to expand in density, coverage, and temporal resolution.

References

[1] Cantoni, J., Kalantari, Z., & Destouni, G. (2023). Legacy contributions to diffuse water pollution: data-driven multi-catchment quantification for nutrients and carbon. Science of the Total Environment, 879, 163092.

[2] Yuan, S., Chen, X., Xing, S., Li, J., Chen, H., Liu, Z., & Guo, S. (2025). Transformer-Based Scalable Multi-Agent Reinforcement Learning for Joint Resource Optimization in Cloud–Edge–End Video Streaming Systems. IEEE Transactions on Cognitive Communications and Networking.

[3] Zhang, S., Qiu, L., & Zeng, Z. (2026). Physics-Data Synergy in Structural Health Monitoring: A Multi-Scale Graph Contrastive Framework With Temperature-Adaptive Fusion. IEEE Access.

[4] Shen, Z., Zhao, W., Wang, B., Wang, Z. and Shang, W. (2026). CAGR: A Cross-Accelerator Graph Optimization Framework for Efficient Recommender System Inference. IEEE Access.

[5] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546-116557.

[6] Li, J., Fan, L., Wang, X., Sun, T., & Zhou, M. (2024). Product demand prediction with spatial graph neural networks. Applied Sciences, 14(16), 6989.

[7] Xing, S., & Wang, Y. (2025). Proactive data placement in heterogeneous storage systems via predictive multi-objective reinforcement learning. IEEE Access.

[8] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[9] Zhao, X., Sun, T., Ren, S., Yang, J., & Liu, Y. (2025). RAG-Based AI Agents for Enterprise Software Development: Implementation Patterns and Production Deployment. Frontiers in Artificial Intelligence Research, 2(3), 501-520.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

[10] Fang, Q., & Liu, W. (2025). HARLA-ED: Resolving Information Asymmetry and Enhancing Algorithmic Symmetry in Intelligent Educational Assessment via Hybrid Reinforcement Learning. Symmetry, 18(1), 58.

[11] Mai, N. T., Cao, W., & Fang, Q. (2025). A study on how LLMs (eg GPT-4, chatbots) are being integrated to support tutoring, essay feedback and content generation. Journal of Computing and Electronic Information Management, 18(3), 43-52.

[12] Jiang, B., Wu, B., Cao, J., & Tan, Y. (2025). Interpretable Fair Value Hierarchy Classification via Hybrid Transformer-GNN Architecture. IEEE Access, 13, 198142-198163.

[13] Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

[14] Chen, J., Liu, J., Liang, Y., & Zhou, M. (2026). KE-MLLM: A Knowledge-Enhanced Multi-Sensor Learning Framework for Explainable Fake Review Detection. Applied Sciences, 16(6), 2909.

[15] Liu, J., Wang, J., Chen, H., Guinness, J., Martin, R., & Kulkarni, C. S. (2019). Optimal Level Crossing Predictions for Electronic Prognostics. In AIAA Scitech 2019 Forum (p. 1962).

[16] Xing, S., Wang, Y., & Liu, W. (2025). Self-adapting CPU scheduling for mixed database workloads via hierarchical deep reinforcement learning. Symmetry, 17(7), 1109.

[17] Liu, C. L., Tseng, C. J., Huang, T. H., Yang, J. S., & Huang, K. B. (2023). A multi-task learning model for building electrical load prediction. Energy and Buildings, 278, 112601.

[18] Liu, C. L., Chang, T. Y., Yang, J. S., & Huang, K. B. (2023). A deep learning sequence model based on self-attention and convolution for wind power prediction. Renewable Energy, 219, 119399.

[19] Ding, G., Yang, S., Lin, H., Chen, Z., & Yang, J. S. (2026). LLM-Driven Adaptive Cloud Resource Scheduling: Bridging Reasoning Intelligence with Optimization Guarantees. IEEE Open Journal of the Computer Society.

[20] Wang, X., Zhang, X., Hoo, V., Shao, Z., & Zhang, X. (2024). Legalreasoner: A multi-stage framework for legal judgment prediction via large language models and knowledge integration. Ieee Access, 12, 166843-166854.

[21] Qiu, L. (2024). Deep learning approaches for building energy consumption prediction. Frontiers in Environmental Research, 2(3), 11-17.

[22] Wang, M. (2024). AI technologies in modern taxation: Applications, challenges, and strategic directions. International Journal of Finance and Investment, 1(1), 42-46.

[23] Mai, N. T., Fang, Q., & Cao, W. (2025). Measuring student trust and over-reliance on AI tutors: Implications for STEM learning outcomes. International Journal of Social Sciences and English Literature, 9(12), 11-17.

[24] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974-195985.

[25] Zhang, X., Li, P., Han, X., Yang, Y., & Cui, Y. (2024). Enhancing time series product demand forecasting with hybrid attention-based deep learning models. IEEE Access, 12, 190079-190091.

[26] Muñoz-Carpena, R., Ritter, A., & Fox, G. A. (2019). Comparison of empirical and mechanistic equations for vegetative filter strip pesticide mitigation in long-term environmental exposure assessments. Water research, 165, 114983.

[27] Otto, F. E. (2023). Attribution of extreme events to climate change. Annual Review of Environment and Resources, 48(1), 813-828.