

From Spatiotemporal Feature Alignment to Transferable Causal Discovery: Process Modeling for Interpretable Analysis of Hot-Rolled Surface Quality Mechanisms

Yiquan An^{*†}, Yihang Li[†], Johannes De Smedt[‡], Xi Sun[§], Huaping Chen^{*},  Zhimin Lv^{*¶}

^{*}Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing, Beijing, China

[†]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China

[‡]Research Centre for Information Systems Engineering (LIRIS), KU Leuven, Leuven, Belgium

[§]National Science Library (NSL), Chinese Academy of Sciences, Beijing, China

[¶]Correspondence: lvzhimin@necar.ustb.edu.cn

Abstract—To enable rigorous, interpretable analysis of hot-rolling surface-quality mechanisms, we propose an end-to-end framework for spatiotemporal registration and transferable causal discovery. It targets two central bottlenecks in industrial data-driven modeling: physically-inconsistent representations and regime-sensitive causal discovery under distribution shift. First, a physics-driven nonuniform spatiotemporal warping operator maps heterogeneous time series to a unified spatial domain, resolving spatiotemporal mismatches. Second, the Transferable Industrial Causal Engine (TiCE) formulates a progressive bilevel training paradigm that embeds mechanistic constraints and jointly activates an Independence-Constrained Domain-Invariant regularization with a Hierarchically Adaptive Fusion Weighting mechanism to resist structural drift under mixed regimes. Third, learned causal graphs guide Asymmetric SHAP attributions to preserve causal consistency in mechanistic explanations. Benchmarks and ablations on TE and UF-IMD demonstrate robustness to distributional shift and the necessity of modular synergy. In hot-rolling, the framework reveals a binary antagonistic mechanism for surface defects: temperature-dominated oxidation kinetics versus multi-stage descaling suppression, thereby providing credible decision support for root-cause diagnosis in complex processes.

Index Terms—Spatiotemporal Signal Alignment, Causal Interpretability, Transferable DAGs, Industrial Process Modeling.

I. INTRODUCTION

In modern manufacturing, establishing precise mappings from process parameters to product quality through data-driven methods remains a central challenge for process optimization and intelligent control [1]. However, multi-source time-series signals, acting as key descriptors of process quality regulation, exhibit misalignment, non-stationarity, and multi-scale heterogeneity. These issues severely constrain the accuracy and interpretability of existing models and hinder the discovery of underlying quality mechanisms [2]. Specifically, three challenges are rarely addressed jointly: (i) constructing physically faithful feature representations from heterogeneous

sensors [3]; (ii) decoupling statistical correlations into physical causal structures within this representation [4], [5]; and (iii) deriving mechanism-level explanations that provide actionable guidance for process control. We propose a three-stage framework: registering heterogeneous signals to strip length, learning a regime-stable causal structure, and producing causally consistent attributions for defect diagnosis. In this way, each stage eliminates a specific failure mode of the next.

First, the representation challenge stems from spatiotemporal inconsistencies that make common time-domain fusion physically invalid for length-indexed quality labels [6], [7]. In steel rolling processes, process control variables evolve as time-dependent functions, whereas hot-rolled strip surface quality varies along the product’s length-based spatial coordinates [8]. Due to non-stationary material transport velocities, no direct linear mapping exists between temporal and spatial domains [9]. Naive fusion of such inconsistent data engenders severe feature aliasing, undermining physical coherence in predictive models [10]. Second, the causal mining challenge stems from attribution ambiguity of structural learning [11]. Traditional Causal Discovery (CD) algorithms, facing multi-source heterogeneous, variable operating regimes, and noisy data typical of industrial contexts, encounter dual impasses: statistical test failures and combinatorial explosions in exhaustive searches [12], [13], hindering identification of causal networks consistent with industrial mechanisms. Finally, mainstream interpretive methods, notably conventional SHAP, exhibit biases in highly coupled and strongly heterogeneous process industries, hampering precise quantification of nonlinear feature contributions to defects and thereby limiting the depth and reliability of causal inference [14], [15].

Accordingly, this study proposes an end-to-end modeling framework spanning spatiotemporal feature registration of process data to transferable CD, with three principal contributions:

- **Spatiotemporal registration representation.** Overcoming the limitations of existing linear time–space mappings, we design non-uniform regularization and local aggregation strategies to achieve a physically consistent

This research is supported by the National Science and Technology Major Project of China - Intelligent Manufacturing Systems and Robots (2025ZD1602203). The first author acknowledges financial support from the China Scholarship Council (Grant No. 202506460019) and the Innovative Talent Cultivation Fund of University of Science and Technology Beijing.

- **Transferable causal modeling.** We propose TiCE (Transferable Industrial Causal Engine), a progressive two-stage differentiable DAG learner guided by mechanistic priors. It stabilizes structure learning via two modules: the Independence-Constrained Domain-Invariant (ICDI) module enforces group-level regime invariance through adversarial residual regularization, while Hierarchically Adaptive Fusion Weighting (HAFW) refines sample contributions using reliability-aware reweighting for robust refinement under mixed regimes;
- **Causal interpretability analysis.** Integrating asymmetric SHAP with causal graph guidance to achieve mechanism-level attributions for multi-defect and single-defect samples. This unveils direction-aware and reliable contribution patterns of cross-operational variables.

A. Data Formalization and Problem Definition

B. Deformable Spatiotemporal Signal Alignment

The diagram illustrates the ST-SNL framework for time series forecasting. It shows a multi-scale processing pipeline. At the top, 'Shift Transformation' is applied to the input data. This is followed by 'Rebuilding Features' and 'Edges Map' at the top scale. The process then moves to 'Rebuilding HSB' and 'Feature Extraction' at the next scale. This is followed by 'Transform to Planning Step' and 'Re-use the Table' at the bottom scale. The final output is 'Global Feature Extraction' and 'Global Feature Extraction'. The diagram also shows 'Spatiotemporal Scale Non-linearity' with a 3D visualization of the feature space and 'Feature Domain Histogram' with a 2D histogram. The 'Global Feature Extraction' section includes a 'Global Feature Extraction' table and a 'Global Feature Extraction' table.

(DTW) [16], which warps time to match signal similarity, our mapping enforces physical spatial correspondence by explicitly integrating the measured instantaneous velocity $v_i(t)$. In continuous form, the operator is an integral mapping:

Under discrete observations, given the velocity sequence $\{v_{i,k}\}$ at timestamps $\{t_{i,k}\}$, we approximate the integral by the trapezoidal rule to obtain a discrete operator $\mathcal{W}$. The spatial position $l_{i,k}$ associated with sampling time $t_{i,k}$ is iteratively updated as

Applying $\tilde{W}$ to the raw process-signal set $\tilde{X}_i$ yields a spatially registered point set $S_i = \{(x_{i,k}, l_{i,k})\}_{k=1}^{K_i}$. In S_i , each process observation $x_{i,k}$ is endowed with a precise physical spatial coordinate $l_{i,k}$, thereby resolving the spatiotemporal inconsistency.

We construct for each analysis unit a fixed-dimensional feature vector via a Localized Feature Generation operator L_f . We first partition the spatial domain L along the length direction with stride δ_l and window size Δ_l to obtain N_{win} contiguous spatial analysis windows $\{\text{ROI}_j\}_{j=1}^{N_{\text{win}}}$. For each ROI_j , two parallel steps are performed: 1) Label assignment. Based on overlap between the quality-event set Y_i and ROI_j , assign a label y_j according to the following rule:

2) Dynamic feature extraction. Select all samples in S_i whose spatial coordinates fall within ROI_j to form $S_{i,j} = \{(x, l) \in S_i \mid l \in \text{ROI}_j\}$. Apply an aggregation operator p to obtain $f_j^{\text{dyn}} = p(S_{i,j}) = [\text{mean}(\{x\}_{S_{i,j}}), \text{var}(\{x\}_{S_{i,j}}), \dots]^\top$. To fuse the high-dimensional static attribute vector z_i (encoding intrinsic product/process properties), we introduce the Static Embedding operator $\mathcal{C}$. Leveraging similarity in chemistry and specifications among homologous products, we apply K-Means [17] on $\{z_i\}_{i=1}^N$ to obtain a cluster-index embedding $Cl_s: \mathbb{R}^q \rightarrow \{1, 2, \dots, K_{\text{clusters}}\}$. For product i , the static feature $f_i^{\text{static}} = Cl_s(z_i)$ serves as a low-dimensional categorical descriptor of its material/specification archetype. Finally, for the j -th spatial window of product i , the complete feature vector is

$f_{ij} = f_j^{\text{dyn}} \oplus f_i^{\text{static}}$. Through these operators, the raw, heterogeneous, spatiotemporally misaligned data are transformed into a structured, physically grounded, high-fidelity feature dataset $\{F, y\}$ directly usable by machine-learning models.

III. TRANSFERABLE CAUSALITY-AWARE INTERPRETABLE MODELING

A. Fundamental Paradigm of Differentiable Causal Discovery

Grounded in Functional Causal Models (FCMs), differentiable CD methods formulate Directed Acyclic Graph (DAG) learning as a continuous optimization task to enable scalable search on high-dimensional industrial data [13], expressed as:

$$\min_G \mathcal{M}_{\text{Base}} = \mathcal{L}_{\text{rec}}(X; \Theta; G) + \alpha T_{\text{sparse}}(G) + \beta P_{\text{DAG}}(G). \quad (4)$$

The scoring function $\mathcal{M}_{\text{Base}}$ encompasses three objectives: (1) $\mathcal{L}_{\text{rec}}(X; \Theta; G) = \frac{1}{n} \sum_{i=1}^n l(X_i, F(X_i; \Theta_i))$, quantifying the capacity of graph G and learnable model F to reconstruct data X 's generative process, typically fitted via predefined neural networks where $\Theta_k = (W^{(1)}, \dots, W^{(l_k)})$ denotes parameters of a k -layer fully connected structure; (2) $T_{\text{sparse}}(G) = \|A\|_1$, a sparsity regularizer for graph structure, with adjacency matrix A as G 's numerical representation; (3) $P_{\text{DAG}}(G) = \lambda h(A) + \frac{\mu}{2} \|h(A)\|^2$, an acyclicity penalty pivotal for converting discrete optimization to a continuous constrained form, with the mainstream DAGness constraint $h(A) = \text{Tr}[(I + \kappa(A \odot A)^s)] - s$, where $\kappa > 0$ is a Lagrange multiplier, and the DAG structure holds if $h(A) = 0$. However, industrial settings often feature regime-varying noise and reliability disparities, breaking i.i.d. assumptions in differentiable CD. Thus, we model the process as a mixture distribution $p(x) = \sum_{k=1}^K p(c = k)p(x | c = k)$ across K regimes, aiming to learn a regime-stable causal graph. To this end, the proposed Transferable Industrial Causal Engine (TiCE) for industrial processes is illustrated in Fig. 2.

B. Robust Causal Modeling under Regime Shifts

TiCE extends the differentiable DAG learner with staged robustness: ICDI suppresses regime-dependent residual artifacts to improve identifiability, and HAFW downweights unreliable samples under mixed regimes to mitigate spurious edges.

Embedding prior knowledge. From an optimization perspective, mechanistic priors restrict the feasible search space, thereby mitigating Markov-equivalent ambiguity and suppressing false edges without requiring costly interventional data. By prescribing or forbidding edges using priors, TiCE reduces the search space and helps mitigate spurious edges under regime shifts. Let $\Theta = \{\bar{\Theta}, \hat{\Theta}\}$ denote the learnable parameters, where $\bar{\Theta}$ is free and $\hat{\Theta}$ is constrained. Based on domain priors, we impose bound constraints $\hat{\Theta}_l \leq \hat{\Theta} \leq \hat{\Theta}_u$. The prior-guided objective becomes $\min_{G; \bar{\Theta}, \hat{\Theta}_l \leq \hat{\Theta} \leq \hat{\Theta}_u} \mathcal{M}(X; \Theta; G)$.

Unsupervised regime sensing. Since regime labels are unavailable, heterogeneous fitting difficulties manifest as distributional shifts in predictive residuals; hence, ε_i serves as a self-supervised proxy to identify the latent regime c_i . We determine K regime centers via a quantile scheme: compute

the Empirical Cumulative Distribution Function (ECDF) [18] of the residual-energy set E ; place $K + 1$ equally spaced quantile levels on $[0, 1]$; with the quantile function $Q(\cdot)$, define centers $c_k = \frac{1}{2}(Q((k-1)/K) + Q(k/K))$. To handle fuzzy regime boundaries, a distance-based soft assignment is employed to learn a K -dimensional membership vector $\pi_i = [\pi_{i1}, \pi_{i2}, \dots, \pi_{iK}]$:

$$\pi_{ik} \triangleq p(c_i = k | E_i \triangleq \|\varepsilon_i\|_2^2) \approx \frac{\exp(-|E_i - c_k|/T)}{\sum_{j=1}^K \exp(-|E_i - c_j|/T)}. \quad (5)$$

where $\sum_{k=1}^K \pi_{ik} = 1$ and $T > 0$ is a temperature controlling the softness. This self-consistent regime sensing updates soft memberships based on current residuals, avoiding hard partitioning and stabilizing training under mixed regimes.

Noise independence constraint. Residual independence tightens Additive Noise Model (ANM) identifiability and discourages shortcut edges when reconstruction dominates the score, preventing regime-specific spurious links in neural FCMs. The proposed model's noise independence module quantifies and reconstructs residual mutual independence as an optimization regularizer. Drawing on multivariate independence testing theory [19], [20], an integrated statistic $I(\varepsilon) = \sum_{i=1}^s \left\| \frac{\text{Cov}[u_i(\varepsilon_i), v_i(\varepsilon_{\text{all}-i})]}{\sqrt{\text{Var}[u_i(\varepsilon_i)]} \sqrt{\text{Var}[v_i(\varepsilon_{\text{all}-i})]}} \right\|$ measures interdependence within the residual set $\varepsilon = \{\varepsilon_i\}_{i=1}^s$; non-zero values denote correlations, with nonlinear mappings $u_i \in L_{\varepsilon_i}^2$ and $v_i \in L_{\varepsilon_{\text{all}-i}}^2$. For continuous optimization, the ℓ_2 -norm is employed to formulate the regularizer $\max_{\xi} \|I(\varepsilon)\|_2^2$, where u_i and v_i are parameterized by an auxiliary learner F^ξ that mirrors the base architecture and is defined by parameters ξ_i .

Condition-domain invariant representation. To strengthen generalization against regime drift, we introduce a regularization framework for learning condition-invariant representations that adopts unsupervised domain adaptation: a condition discriminator F^φ plays a min-max game with the residual learner F^ξ , aligning the conditional residual distributions $P(\varepsilon | c_k)$ across regimes. To stabilize and expedite adversarial training, define a nonparametric mapping $f_{\text{stat}} : \mathbb{R}^s \rightarrow \mathbb{R}^d$ ($d \ll s$) that projects the high-dimensional residual vector onto a low-dimensional space of interpretable moment statistics, thereby capturing essential distributional traits, where $f_{\text{stat}}(\varepsilon_i) = (\|\varepsilon_i\|_2^2/s, \|\varepsilon_i\|_1/s, \|\varepsilon_i\|_\infty)^\top \in \mathbb{R}^3$. We then construct a condition discriminator that predicts $\hat{\pi}_i \in \mathbb{R}^K$ from the residual statistics, using the previously obtained soft regime memberships π_i as pseudo-labels. Its objective minimizes the cross-entropy between predictions and pseudo-labels:

$$\min_{\varphi} \mathcal{L}_{\text{Disc}} = -\mathbb{E}_{x_i \sim p(x)} \left[\sum_{k=1}^K \pi_{ik} \log \left(F^\varphi(f_{\text{stat}}(F^\xi(\varepsilon_i)))_k \right) \right]. \quad (6)$$

Meanwhile, the generator F^ξ maximizes the same loss, i.e., $\max_{\xi} \mathcal{L}_{\text{Disc}}$, by iteratively updating ξ and $f_{\text{stat}}(F^\xi(\varepsilon_i))$ so that F^φ cannot distinguish regimes. In sum, coupling adversarial independence constraints with condition-domain

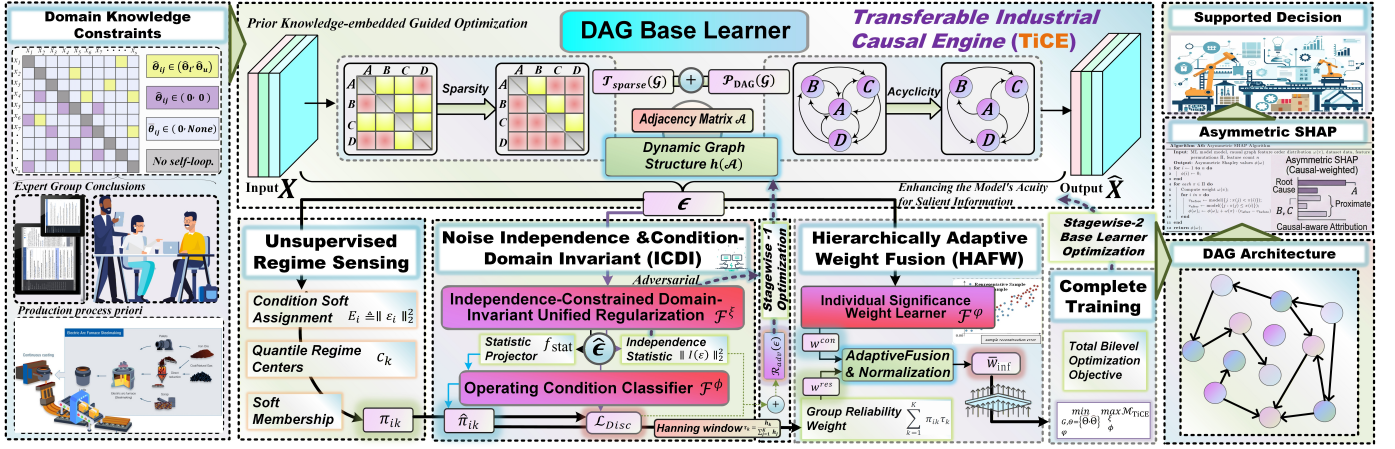


Fig. 2. Model structure of the proposed TiCE framework.

invariance yields the cascaded framework Independence-Constrained Domain-Invariant (ICDI), a robust identifiability regularizer, whose full objective is:

$$\arg \max_{\xi} \min_{\varphi} \mathcal{R}_{\text{adv}}(\epsilon) = \|I(\epsilon)\|_2^2 + \mathcal{L}_{\text{Disc}}. \quad (7)$$

Hierarchically adaptive weight fusion. To address sample-level reliability under mixed regimes, we propose HAFW (Hierarchically Adaptive Fusion Weighting), which assigns each sample a weight w_i derived from salience and regime reliability, thereby complementing ICDI's distribution-level alignment to promote the structural precision of causal edges. Individual salience weights are produced by an auxiliary Adaptive Weighting Network F^ϕ , yielding $w_i^{\text{res}} = \{w_i^{\text{res}} \in C(\zeta), \zeta \in (0, 1) \mid 0 < \zeta/n \leq w_1, \dots, w_n \leq 1/(\zeta n), \sum_{i=1}^n w_i^{\text{res}} = 1\}$, where ζ bounds deviation from uniform weighting and shifts attention toward boundary-informative salient samples. Group reliability weights rely on a prior regime-reliability vector $\tau \in \mathbb{R}^K$ and are defined as the expectation of τ under the soft membership π_i : $w_i^{\text{con}} = \mathbb{E}_{\pi_i}[\tau] = \sum_{k=1}^K \pi_{ik} \tau_k$. To meet boundary-instability and smooth-transition requirements, $\tau_k = h_k / (\sum_{j=1}^K h_j)$ is constructed via a Hanning window h_k [21]. Finally, with a fusion coefficient $\eta \in [0, 1]$, the two components are convexly combined to yield the preliminary fused weight $w'_i = \eta \tilde{w}_{i,\text{res}} + (1 - \eta) \tilde{w}_{i,\text{con}}$. Prior to fusion, to ensure scale invariance, each component is mini-batch normalized as $\tilde{w}_{i,\text{inf}} = w_{i,\text{inf}} / (\sum_{j=1}^{N_{\text{batch}}} w_{j,\text{inf}} + \epsilon)$, with $\text{inf} \in \{\text{res}, \text{con}\}$. To further suppress high-frequency noise induced by mini-batch stochasticity, an Exponential Moving Average (EMA) serves as a low-pass filter to smooth the fused weights; at training step t , $\bar{w}_i^{(t)} = \Xi \bar{w}_i^{(t-1)} + (1 - \Xi) w'_i$. Accordingly, the fused-weighted goodness-of-fit term is $\mathcal{L}_{\text{FW-rec}}(X; \Theta; \phi; G) = \mathbb{E}_{c \in \mathbb{R}^K} [\frac{1}{n} \sum_{i=1}^n \bar{w}_i \|X_i - F(X_i; \Theta; \phi; G)\|_2^2]$.

Complete training framework. Formally, TiCE's combinatorial optimization objective distills into a nested bilevel progressive problem with inner updates refining structure:

$$\begin{aligned} & \min_{G, \Theta = \{\hat{\Theta}, \hat{\Theta}\}} \max_{\xi, \phi} \mathcal{M}_{\text{TiCE}}(X; \Theta; \xi; \phi; G) \\ & = \begin{cases} \mathcal{L}_{\text{rec}} + \alpha T_{\text{sparse}} + \beta P_{\text{DAG}} + \gamma \mathcal{R}_{\text{adv}}, & \text{Stage ICDI} \\ \mathcal{L}_{\text{FW-rec}} + \alpha T_{\text{sparse}} + \beta P_{\text{DAG}} + \gamma \mathcal{R}_{\text{adv}}, & \text{Stage HAFW} \end{cases} \quad (8) \end{aligned}$$

Specifically, Equation (8) encompasses two sequential bilevel loops: first, the inner loop of objective one alternates updates between base F^Θ and ICDI components F^ξ , F^φ to iterate $\mathcal{R}_{\text{adv}}$ over steps, while the outer minimizes overall ICDI-stage loss $\mathcal{M}_{\text{ICDI}}$ for interim graph G optimization; second, objective two's inner loop adaptively learns individual weights via the HAFW component F^ϕ , fusing regime reliability to minimize loss and optimize $\bar{w}_{\text{inf}}$, with the outer minimizing HAFW-stage loss $\mathcal{M}_{\text{HAFW}}$ for final G refinement. Algorithm 1 summarizes the two-stage progressive training procedure (ICDI $\rightarrow$ HAFW) for solving Equation. (8).

C. Interpretable Analysis Based on Asymmetric SHAP

Shapley Additive exPlanations (SHAP) decomposes a prediction into additive feature attributions via Shapley values; however, under causal dependence, standard SHAP enforces the symmetry axiom by averaging over all feature permutations, ignoring the graph structure. This can down-weight distal causes and inflate proximal proxies [22]. Let Π be the set of all permutations of s features and let $w_{\text{SHAP}} \in \Delta(\Pi)$ denote a distribution over Π . Asymmetric SHAP [23] reweights permutations and computes $\sigma_v^{(w)}(i) = \sum_{\pi \in \Pi} w(\pi) [v(\{j : \pi(j) \leq \pi(i)\}) - v(\{j : \pi(j) < \pi(i)\})]$, where $v(\cdot)$ is the valuation function. Weights are induced by a DAG G . Distal weighting places mass only on permutations that respect every ancestor $\rightarrow$ descendant relation, i.e., $w_{\text{distal}}(\pi) \propto \mathbf{1}\{\pi(i) < \pi(j) \text{ for all } i \rightarrow j\}$, concentrating attribution on root causes and suppressing redundant credit to descendants. Proximate weighting $w_{\text{proximate}}(\pi)$ reverses this order; ancestor scores capture their effects, while descendant scores quantify incremental effects conditional on ancestors. In the absence of causal information, uniform weights $w(\pi) = 1/s!$ recover symmetric SHAP. Thus, asymmetric SHAP adjusts permutation weights using the causal graph, yielding causally coherent attributions in multistage processes.

IV. EXPERIMENTS

A. Spatiotemporal Data Fusion and Dataset Construction

We address spatial heterogeneity in hot-rolled strip defects by introducing a physics-driven nonuniform spatiotemporal

Algorithm 1 TiCE: ICDI $\rightarrow$ HAFW

Input: Data $X \in \mathbb{R}^{n \times s}$; regimes K ; FCM $F(\cdot; \Theta, A)$; prior bounds $[\hat{\Theta}_l, \hat{\Theta}_u]$ on constrained parameters $\hat{\Theta}$; hyperparams $(\alpha, \beta, \gamma, \eta, \Xi, \zeta, T)$; iters $(M_{\text{ICDI}}, M_{\text{HAFW}})$; inner steps (R_{adv}, R_ϕ) ; constraint set $C(\zeta)$.
Output: Adjacency matrix A^* and the extracted DAG G^* .

- 1: **Init:** initialize $A, \Theta = \{\hat{\Theta}, \hat{\Theta}\}$; ICDI modules (ξ, φ) ; weight net ϕ ; EMA weights $\bar{w}_i \leftarrow 1/n$.
- 2: $\hat{\Theta} \leftarrow \Pi_{[\hat{\Theta}_l, \hat{\Theta}_u]}(\hat{\Theta})$.
- 3: **Regime sensing (refreshed with current residuals):**
- 4: $\varepsilon_i \leftarrow X_i - F(X_i; \Theta, A), E_i \leftarrow \|\varepsilon_i\|_2^2$.
- 5: Quantile centers: $c_k = \frac{1}{2} \left(Q(\frac{k-1}{K}) + Q(\frac{k}{K}) \right)$.
- 6: Soft membership: $\pi_{ik} \leftarrow \frac{\exp(-|E_i - c_k|/T)}{\sum_{j=1}^K \exp(-|E_i - c_j|/T)}$.
- 7: Regime reliability $\tau \in \mathbb{R}^K$: normalized Hanning window, $\sum_k \tau_k = 1$.
- 8: **Stage I: ICDI (structure learning under regime invariance)**
- 9: **for** $m = 1$ **to** M_{ICDI} **do**
- 10: $\varepsilon \leftarrow X - F(X; \Theta, A)$.
- 11: Refresh π using current ε (as above).
- 12: **for** $r = 1$ **to** R_{adv} **do**
- 13: $\varphi \leftarrow \arg \min_{\varphi} \mathcal{L}_{\text{Disc}}(\varphi; f_{\text{stat}}(F^\xi(\varepsilon)), \pi)$.
- 14: $\xi \leftarrow \arg \max_{\xi} (\|I(F^\xi(\varepsilon))\|_2^2 + \mathcal{L}_{\text{Disc}})$.
- 15: **end for**
- 16: $(\Theta, A) \leftarrow \arg \min_{\Theta, A} \mathcal{L}_{\text{rec}} + \alpha \|A\|_1 + \beta P_{\text{DAG}}(A) + \gamma (\|I(F^\xi(\varepsilon))\|_2^2 + \mathcal{L}_{\text{Disc}})$.
- 17: $\hat{\Theta} \leftarrow \Pi_{[\hat{\Theta}_l, \hat{\Theta}_u]}(\hat{\Theta})$.
- 18: **end for**
- 19: **Stage II: HAFW (sample reweighting for robust refinement)**
- 20: **Keep ICDI active:** continue updating (ξ, φ) while enabling HAFW.
- 21: **for** $m = 1$ **to** M_{HAFW} **do**
- 22: $\varepsilon \leftarrow X - F(X; \Theta, A)$, Refresh π using current ε (as above).
- 23: **for** $r = 1$ **to** R_ϕ **do**
- 24: $w^{\text{res}} \leftarrow \Pi_{C(\zeta)}(F^\phi(\varepsilon \odot \varepsilon)); w_i^{\text{con}} \leftarrow \sum_{k=1}^K \pi_{ik} \tau_k$.
- 25: Normalize: $\bar{w}^{\text{res}}, \bar{w}^{\text{con}}; \text{fuse } w_i^f = \eta \bar{w}_i^{\text{res}} + (1 - \eta) \bar{w}_i^{\text{con}}$.
- 26: EMA: $\bar{w}_i \leftarrow \Xi \bar{w}_i + (1 - \Xi) w_i^f$.
- 27: $\phi \leftarrow \arg \max_{\phi} \mathcal{L}_{\text{FW-rec}}(\Theta, A; \bar{w}(\phi))$.
- 28: **end for**
- 29: **for** $r = 1$ **to** R_{adv} **do**
- 30: $\varphi \leftarrow \arg \min_{\varphi} \mathcal{L}_{\text{Disc}}(\varphi; f_{\text{stat}}(F^\xi(\varepsilon)), \pi)$.
- 31: $\xi \leftarrow \arg \max_{\xi} (\|I(F^\xi(\varepsilon))\|_2^2 + \mathcal{L}_{\text{Disc}})$.
- 32: **end for**
- 33: $(\Theta, A) \leftarrow \arg \min_{\Theta, A} \mathcal{L}_{\text{FW-rec}}(\Theta, A; \bar{w}) + \alpha \|A\|_1 + \beta P_{\text{DAG}}(A) + \gamma (\|I(F^\xi(\varepsilon))\|_2^2 + \mathcal{L}_{\text{Disc}})$.
- 34: $\hat{\Theta} \leftarrow \Pi_{[\hat{\Theta}_l, \hat{\Theta}_u]}(\hat{\Theta})$.
- 35: **end for**
- 36: **Return:** $A^* \leftarrow A$ (final optimized A in $F(\cdot; \Theta, A)$), $G^* \leftarrow \text{DAGify}(\text{threshold}(A^*))$.

warping operator within the spatiotemporal registration framework. Using the strip-speed profile in Fig. 3a and the integral transform in Equation. 1, we reparameterize each time-series process signal $\tilde{X}_i$ from the time domain to the strip-length domain. The transformation yields a spatially registered set $S_i = \{(x_{i,k}, l_{i,k})\}_{k=1}^{K_i}$, where each observation is indexed by the spatial coordinate $l_{i,k}$. This representation preserves position-dependent process dynamics under the nonstationary motion shown in Fig. 3b and provides a unified spatial reference for downstream quality mapping.

We optimize defect spatial encoding by deploying a localized feature generation operator L_f with a 50, mm initial scale and an adaptive stride, which partitions the spatial domain into analysis windows $\{\text{ROI}_j\}_{j=1}^{N_{\text{win}}}$. Based on the defect encoding in Fig. 4a, the labeling function in Equation. 3 assigns a single

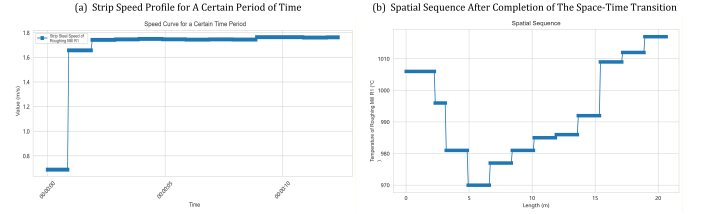


Fig. 3. Nonuniform Time-to-Space Registration for Physically Consistent Process Representation.

quality label to each ROI_j , as shown in Fig. 4b. For each window, we extract the registered subset $S_{i,j}$ and apply an aggregation operator to obtain ROI-level dynamic descriptors. In parallel, we embed static covariates into a vector and derive the cluster-index feature f_i^{static} through a low-dimensional clustering projection, with alternative unsupervised clustering methods compared in Fig. 4c. Finally, we integrate key process values across stages to construct the process-to-defect mapping dataset with 10,222 samples, as summarized in Fig. 4d.

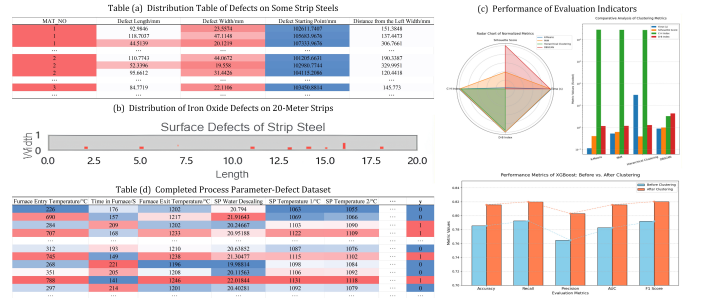


Fig. 4. High-Fidelity Process-Defect Dataset Construction via ROI-Based Defect Encoding and Feature Aggregation.

B. Benchmark Evaluation

We evaluate TiCE on TE and UF-IMD, two process-industry causal datasets [24], reporting Precision ($\uparrow$), False Discovery Rate (FDR, $\downarrow$), True Positive Rate (TPR, $\uparrow$), False Positive Rate (FPR, $\downarrow$), Structural Hamming Distance (SHD, $\downarrow$), and F1 Score ($\uparrow$). Figure 5 benchmarks TiCE against CD baselines spanning multiple paradigms, including constraint-based methods (FCI [25], MCSL [26], PC [27]), score-based methods (GES [28], GOLEM [29]), functional approaches (CORL [30], DirectLiNGAM [31], ICALiNGAM [32]), and differentiable formulations (NOTEARS [33], NOTEARS-MLP [34]). On TE, without domain priors, concurrent gains in TPR and F1 Score demonstrate that TiCE recovers true causal relations by leveraging invariance and reweighting against regime-dependent noise. On UF-IMD, superior Precision with lower FDR and FPR demonstrates that TiCE suppresses spurious edges driven by varying sample reliability across regimes.

C. Ablation Study

Ablations in Fig. 6 progressively add TiCE components to Equation. (4): +IC adds $\|I(\varepsilon)\|_2^2$, +ICDI further adds $\mathcal{L}_{\text{Disc}}$, +RES enables residual-based reweighting w^{res} , and +HAFW completes the hierarchical fusion with regime-reliability weighting w^{con} (EMA-smoothed). On UF-IMD, TiCE leads in F1 and Precision; moving from +RES to

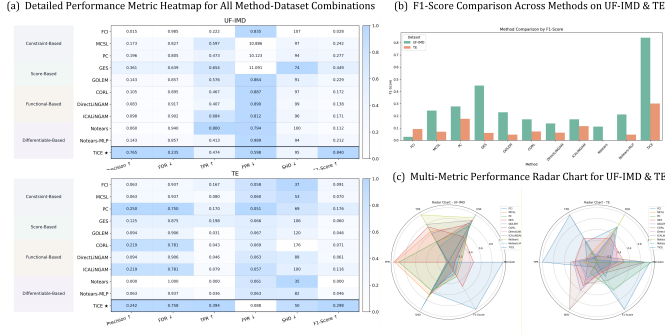


Fig. 5. Comparison of TiCE with other multi-category methods on UF-IMD & TE dataset.

+HAFW boosts TPR but can inflate FPR, indicating that hierarchical fusion improves sensitivity yet requires careful false-positive control; +RES and +ICDI slightly reduce SHD but at the expense of Precision and TPR. On TE, TiCE retains the highest F1; +RES raises TPR yet degrades Precision and FDR, revealing low-SNR vulnerability; +IC/+ICDI marginally lower FPR but increase SHD, showing that invariance constraints without sample-level adaptivity underfit complex causal structure. Overall, the results confirm that each module targets a distinct failure mode, rendering the full TiCE framework essential for robust causal discovery.

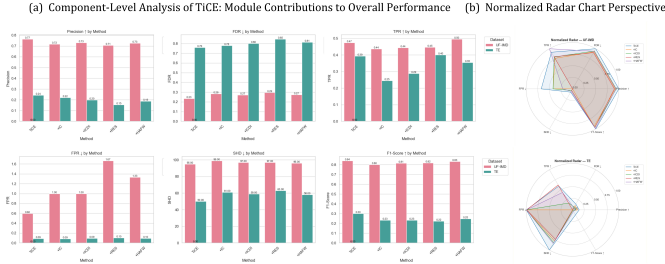


Fig. 6. Experimental Comparison of the Components within the TiCE Framework.

D. Causal Modeling of Hot-Rolling Quality with Interpretable Defect Analysis

TiCE’s learned DAG in Fig. 7 reveals a hierarchical causal mechanism of mill-scale defects in the dataset described in Sec. IV-A. It reflects strict temporal ordering and path dependence, where upstream temperature states act as boundary conditions for downstream thermal dynamics. The inferred root cause follows a binary antagonistic mechanism between oxide-growth kinetics driven by the furnace-exit temperature and Finishing Delivery Temperature (FDT), and multistage suppression induced by water descaling operations. The graph further localizes nodes where generation and removal effects counteract, while edges from cluster labels to defects support defect formation as an emergent property of coordinated process-parameter regimes.

Applying asymmetric SHAP to the DAG, with directional weighting to curb attribution leakage, yields across 1,000 labeled defect samples a process-wide “temperature backbone + descaling modulation” map (Fig. 8a). The attributions form a temperature-dominated backbone with descaling modulation,

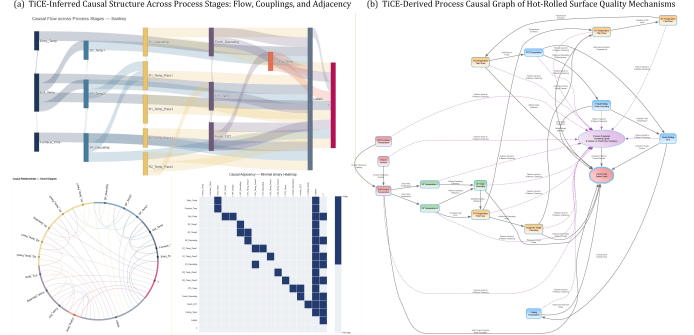


Fig. 7. Causal Structure Mining in Hot-Rolling Production via Knowledge-Integrated TiCE Algorithm.

where dominant contributors (contribution > 0.025) include the furnace-exit temperature, $R1$ temperatures (passes 1 and 3), $R2$ temperature (pass 3), furnace residence time, $FT0$, and FDT , while SP and finishing-stage water descaling show negative inhibitory effects. Compared with standard SHAP (Fig. 8b), data concordance improves by avoiding inflated importance for the entry temperature, and flagged features exhibit defect means above the normal-sample third quartile Q_3 . Labels receive stronger negative effects consistent with labels $\uparrow \Rightarrow$ defect rate $\downarrow$, and single-sample explanations align in direction with box-plot tests (Fig. 8c). These results support multistage coupling characterization, inhibitory mechanism identification, and intervention-oriented interpretation for oxide-scale defect diagnosis and parameter optimization.

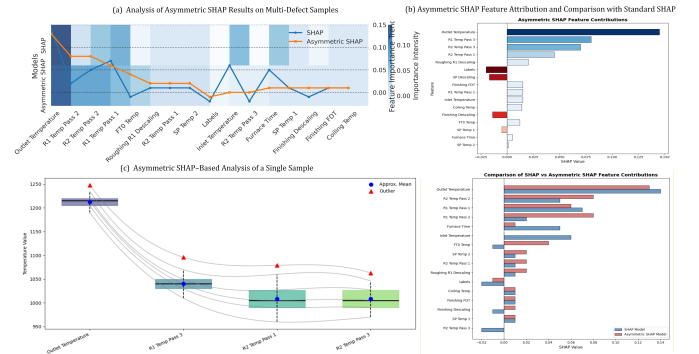


Fig. 8. Causal-Explainable Analysis Outcomes Using Asymmetric SHAP.

V. CONCLUSION

We present a closed-loop paradigm that unifies high-fidelity spatiotemporal representation, transferable causal discovery, and causally coherent interpretation. We introduce velocity-driven nonuniform warping with localized aggregation and static embeddings to align time with space and yield a physically coherent dataset that links process parameters to defect outcomes. We develop TiCE and its industrial-oriented modules to encode mechanistic priors, employ ICDI to enforce adversarial residual independence and condition invariance, and use HAFW to learn regime-invariant DAGs under mixed operating regimes. Benchmarks demonstrate strong distributional robustness, and ablations confirm the necessity and complementarity of the proposed modules. Building on the

learned graph, we integrate asymmetric SHAP with direction-aware weighting, which reveals a binary antagonism in which a temperature backbone comprising furnace exit, $FT0$, and FDT drives oxidation, whereas SP and finishing descaling suppress it. It further shows that the directed relation from process variables to defects is consistent with a multiparameter emergent phenomenon. The pipeline operationalizes heterogeneous data alignment, robust causal structure learning, and intervention-oriented explanations, providing reliable guidance for mill scale diagnosis and cross-regime optimization.

REFERENCES

- [1] Y. Cheng, H. Zhan, J. Yu, and R. Wang, "Data-driven online prediction and control method for injection molding product quality," *Journal of Manufacturing Processes*, vol. 145, pp. 252–273, Jul. 2025.
- [2] J. Cação, J. Santos, and M. Antunes, "Explainable AI for industrial fault diagnosis: A systematic review," *Journal of Industrial Information Integration*, vol. 47, p. 100905, Sep. 2025.
- [3] Y. Wu, B. Sicard, and S. A. Gadsden, "Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring," *Expert Systems with Applications*, vol. 255, p. 124678, Dec. 2024.
- [4] G. Camps-Valls, A. Gerhardus, U. Ninad, G. Varando, G. Martius, E. Balaguer-Ballester, R. Vinuesa, E. Diaz, L. Zanna, and J. Runge, "Discovering causal relations and equations from data," *Physics Reports*, vol. 1044, pp. 1–68, Dec. 2023.
- [5] Y. Zhou, H. Wang, C. Gu, and H. Yang, "Detecting causalities between strongly coupled dynamical systems," *Physica A: Statistical Mechanics and its Applications*, vol. 653, p. 130074, Nov. 2024.
- [6] B. Li, P. Song, C. Zhao, and M. Xie, "Facing spatiotemporal heterogeneity: A unified federated continual learning framework with self-challenge rehearsal for industrial monitoring tasks," *Knowledge-Based Systems*, vol. 289, p. 111491, Apr. 2024.
- [7] Y. An, N. Hou, and Z. Lv, "A systematic review of transfer learning research from the perspective of knowledge graph: Visual analysis based on the bibliometric method," *WIREs Data Mining and Knowledge Discovery*, vol. 16, no. 2, p. e70084, Apr. 2026.
- [8] S. Chen, S. Jiang, X. Wang, P. Sun, C. Hua, and J. Sun, "An efficient detector for detecting surface defects on cold-rolled steel strips," *Engineering Applications of Artificial Intelligence*, vol. 138, p. 109325, Dec. 2024.
- [9] Y. Yan, J. Hou, Z. Song, and E. E. Kuruoglu, "Signal processing over time-varying graphs: A systematic review," Nov. 2024, arXiv:2412.00462. [Online]. Available: <https://arxiv.org/abs/2412.00462>
- [10] H. Zhang, Q. Qi, and F. Tao, "A consistency evaluation method for digital twin models," *Journal of Manufacturing Systems*, vol. 65, pp. 158–168, Oct. 2022.
- [11] L. Jiao, Y. Wang, X. Liu, L. Li, F. Liu, W. Ma, Y. Guo, P. Chen, S. Yang, and B. Hou, "Causal inference meets deep learning: A comprehensive survey," *Research*, vol. 7, p. 0467, Sep. 2024.
- [12] W. Niu, Z. Gao, L. Song, and L. Li, "Comprehensive review and empirical evaluation of causal discovery algorithms for numerical data," Sep. 2024, arXiv:2407.13054. [Online]. Available: <https://arxiv.org/abs/2407.13054>
- [13] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 2, p. e1449, 2022.
- [14] H. Hwang, A. Bell, J. Fonseca, V. Pliatsika, J. Stoyanovich, and S. E. Whang, "Shap-based explanations are sensitive to feature representation," in *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. Athens, Greece: ACM, Jun. 2025, pp. 1588–1601. [Online]. Available: <https://dl.acm.org/doi/10.1145/3715275.3732105>
- [15] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: SHAP and LIME," *Advanced Intelligent Systems*, vol. 7, no. 1, p. 2400304, 2025.
- [16] M. Herrmann and G. I. Webb, "Amercing: an intuitive and effective constraint for dynamic time warping," *Pattern Recognition*, vol. 137, p. 109333, May 2023.
- [17] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023.
- [18] W. L. Dunn and J. K. Shultis, "The basis of Monte Carlo," in *Exploring Monte Carlo Methods*, W. L. Dunn and J. K. Shultis, Eds. Amsterdam: Elsevier, Jan. 2012, pp. 21–46. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780444515759000026>
- [19] Y. Fan, P. L. De Micheaux, S. Penev, and D. Salopek, "Multivariate nonparametric test of independence," *Journal of Multivariate Analysis*, vol. 153, pp. 189–210, Jan. 2017.
- [20] Y. He, P. Cui, Z. Shen, R. Xu, F. Liu, and Y. Jiang, "Daring: Differentiable causal discovery with residual independence," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Virtual Event, Singapore: ACM, Aug. 2021, pp. 596–605. [Online]. Available: <https://dl.acm.org/doi/10.1145/3447548.3467439>
- [21] M. Parker, "Windowing," in *Digital Signal Processing 101 (Second Edition)*, M. Parker, Ed. Newnes, Jan. 2017, pp. 59–63.
- [22] K. Aas, M. Jullum, and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to shapley values," *Artificial Intelligence*, vol. 298, p. 103502, Sep. 2021.
- [23] C. Frye, C. Rowat, and I. Feige, "Asymmetric shapley values: Incorporating causal knowledge into model-agnostic explainability," Dec. 2021, arXiv:1910.06358. [Online]. Available: <https://arxiv.org/abs/1910.06358>
- [24] G. Menegazzo, D. Dall'Alba, and P. Fiorini, "CIPCaD-Bench: Continuous industrial process datasets for benchmarking causal discovery methods," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. Mexico City, Mexico: IEEE, Aug. 2022, pp. 2124–2131. [Online]. Available: <https://doi.org/10.1109/CASE49997.2022.9926420>
- [25] P. L. Spirtes, C. Meek, and T. S. Richardson, "Causal inference in the presence of latent variables and selection bias," Feb. 2013, arXiv:1302.4983 [cs.AI]. [Online]. Available: <https://arxiv.org/abs/1302.4983>
- [26] I. Ng, S. Zhu, Z. Fang, H. Li, Z. Chen, and J. Wang, "Masked gradient-based causal structure learning," Jan. 2022, arXiv:1910.08527 [cs.LG]. [Online]. Available: <https://arxiv.org/abs/1910.08527>
- [27] M. Kalisch and P. Bühlmann, "Robustification of the PC-algorithm for directed acyclic graphs," *Journal of Computational and Graphical Statistics*, vol. 17, no. 4, pp. 773–789, Dec. 2008.
- [28] D. M. Chickering, "Optimal structure identification with greedy search," *The Journal of Machine Learning Research*, vol. 3, pp. 507–554, Jan. 2003.
- [29] I. Ng, A. Ghassami, and K. Zhang, "On the role of sparsity and DAG constraints for learning linear DAGs," in *Advances in Neural Information Processing Systems*. Vancouver, BC, Canada: Curran Associates, Inc., 2020, pp. 17943–17954. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/d04d42cdf14579cd294e5079e0745411-Abstract.html>
- [30] X. Wang, Y. Du, S. Zhu, L. Ke, Z. Chen, J. Hao, and J. Wang, "Ordering-based causal discovery with reinforcement learning," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. Montreal, Canada: International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 3566–3573. [Online]. Available: <https://www.ijcai.org/proceedings/2021/491>
- [31] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen, "Directlingam: A direct method for learning a linear non-gaussian structural equation model," *The Journal of Machine Learning Research*, vol. 12, pp. 1225–1248, Jan. 2011.
- [32] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen, "A linear non-gaussian acyclic model for causal discovery," *The Journal of Machine Learning Research*, vol. 7, pp. 2003–2030, Jan. 2006.
- [33] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing, "DAGs with NO TEARS: Continuous optimization for structure learning," in *Advances in Neural Information Processing Systems*, vol. 31. Montréal, Canada: Curran Associates, Inc., 2018, pp. 9492–9503.
- [34] X. Zheng, C. Dan, B. Aragam, P. Ravikumar, and E. Xing, "Learning sparse nonparametric DAGs," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 108. PMLR, Jun. 2020, pp. 3414–3425.