

Partially Functional Dynamic Backdoor Diffusion-based Causal Model

Xinwen Liu

YunNan University

liuxinwen@stu.ynu.edu.cn

Lei Qian

Peking University

leiqlian@stu.pku.edu.cn

Song Xi Chen

Tsinghua University

sxchen@tsinghua.edu.cn

Niansheng Tang

YunNan University

nstang@ynu.edu.cn

Abstract

Causal inference in settings involving complex spatio-temporal dependencies, such as environmental epidemiology, is challenging due to the presence of unmeasured confounding. However, a significant gap persists in existing methods: current diffusion-based causal models rely on restrictive assumptions of causal sufficiency or static confounding. To address this limitation, we introduce the Partially Functional Dynamic Backdoor Diffusion-based Causal Model (PFD-BDCM), a generative framework designed to bridge this gap. Our approach uniquely incorporates valid backdoor adjustments into the diffusion sampling mechanism to mitigate bias from unmeasured confounders. Specifically, it captures their intricate dynamics through region-specific structural equations and conditional autoregressive processes, and accommodates multi-resolution variables via functional data techniques. Furthermore, we provide theoretical guarantees by establishing error bounds for counterfactual estimates. Extensive experiments on synthetic data and a real-world air pollution case study confirm that PFD-BDCM outperforms current state-of-the-art methods.

1 Introduction

Causal inference fundamentally addresses interventional (“*What if?*”) and counterfactual (“*Why?*”) questions that go beyond statistical correlations, proving valuable in high-stakes domains like healthcare for treatment effect estimation (Hill, 2011), policy evaluation without randomized trials (LaLonde, 1986). The field’s core challenge stems from the fundamental problem of causal inference: the impossibility of simultaneously observing both an outcome under a treatment and the potential outcomes under the alternative treatment (or control) for the same unit (Imbens and Rubin, 2015), necessitating methods to overcome confounding bias in observational data. Traditional approaches including potential outcomes frameworks (Imbens and Rubin, 2015), propensity scoring (Rosenbaum and Rubin, 1983), instrumental variables (Angrist et al., 1996), and structural causal models (Pearl, 2009) exhibit significant limitations when handling modern complex datasets—they struggle with high-dimensional confounders, ethical constraints of randomized trials, scarcity of valid instruments, and requirement of known causal graphs. These limitations become particularly acute when confounders involve high-dimensional data like medical images or partially unobserved variables (Shalit et al., 2017).

Within the Structural Causal Model (SCM) framework, causal queries can be answered by learning a proxy for the unobserved exogenous noise and the structural equations (Pearl, 2009). This suggests that (conditional) generative models that encode to a latent space could be an option for modeling SCMs, as the latent space serves as proxies for exogenous noises. Recent advances have explored the integration of deep generative models with structural causal models to address these challenges. Chao et al. (2023) proposed the Diffusion-based Causal Model (DCM), which leverages diffusion processes to approximate structural equations and answer causal queries without explicit intervention data. However, DCM assumes causal sufficiency (no unobserved confounders) which rarely holds in practice. Shimizu (2023) extended this line of work with Backdoor Diffusion-based Causal Model (BDCM), incorporating backdoor adjustment to handle certain types of unmeasured confounding. Nevertheless, both methods operate under static assumptions and fail to account for spatio-temporal structure in confounding variables, a critical limitation in real-world systems where confounders often exhibit complex dependencies across space and time.

To address this gap, we propose the Partially Functional Dynamic Backdoor Diffusion-based Causal Model (PFD-BDCM), a generative framework capable of explicitly modeling spatiotemporal correlations among unmeasured confounders while supporting causal inference. First, it integrates valid backdoor adjustments into the diffusion sampling mechanism to mitigate bias from unmeasured confounders. Second, it explicitly models the intricate spatio-temporal dynamics of unmeasured confounders through region-specific structural equations and conditional autoregressive processes. Third, it accommodates variables observed at heterogeneous resolutions via functional data techniques and basis expansions. Furthermore, modeling each node to more accurately address causal inference by leveraging the superior data generation capability of diffusion models (Song et al., 2021). Our approach relaxes the causal sufficiency assumption and captures dynamic confounding through a structured latent representation, allowing for more robust estimation of causal effects in non-stationary environments. We demonstrate empirically that PFD-BDCM outperforms existing diffusion-based causal models across a range of observational, interventional and counterfactual queries, particularly in settings with spatially and temporally varying confounders.

Our Contributions. We propose a Partially Functional Dynamic Backdoor Diffusion-based Causal Model (PFD-BDCM) for modeling partially functional spatio-temporal dynamic causal relationships. Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021) have gained prominence due to their high expressiveness and performance in generative tasks. Our primary contribution is to show how to apply diffusion models to capture partially functional dynamic causal relationship with unmeasured confounders. The core idea is in modeling each node in Partially Functional Spatio-Temporal Dynamic Structural Causal Model (PFST-DSCM) with a diffusion model and cascading the generated samples in the topological order to answer causal queries. For each node, the encoding process of the diffusion model takes the current node variable and its set of backdoor node variables as input to generate latent representations of the unobserved exogenous variables, and then reconstructs the current node variable through a decoding process. Within our framework, the encoding and decoding procedures extend the Denoising Diffusion Implicit Models (DDIMs) (Song et al., 2021) paradigm by integrating backdoor adjustment sets as supplementary covariates. Key contributions of our study include:

(1) [Section 2] The proposed PFD-BDCM provides a unified framework for approximating both interventions (do-operator) and counterfactuals (abduction-action-prediction steps). It has a training procedure requiring only the dynamic causal graph and observational data, and the trained model enables: i) sampling from observational/interventional distributions; ii) precise counterfactual query resolution.

(2) [Section 3] Our theoretical analysis proves that the counterfactual estimates given by PFD-BDCM admit quantifiable error bounds under reasonable assumptions. i) It provides the first formal explanation for the performance gains of encoder-decoder architectures (e.g., diffusion models) in counterfactual querying through error bounds; ii) It extends to the more challenging multivariate case under an additional assumption and to diverse encoder-decoder models.

(3) [Section 4] We evaluated the performance of PFD-BDCM on three synthetic datasets of varying scales involving spatiotemporal dynamic structural equations and three types of causal queries. Experimental results demonstrate that PFD-BDCM consistently outperforms existing state-of-the-art methods (Chao et al., 2023; Shimizu, 2023) as well as PFD-DCM. Furthermore, we also demonstrate the strong performance of PFD-BDCM on a real-world atmospheric pollution dataset.

2 Methodology

We first introduce some useful notations and concepts.

Notations: To distinguish between the nodes in the causal graph and diffusion random variables, we use subscripts to denote graph nodes. Let $[n] := \{1, \dots, n\}$ and $\dim(x)$ represents the dimension of x ; let Z_k^t be the exogenous noise at diffusion step t in the forward process, with $Z_k := Z_k^T$, and \hat{X}_k^t the endogenous variable at step t in the reverse process, where $\hat{X}_k := \hat{X}_k^0$.

In causal inference, a confounder denotes a variable that causally influences both a treatment variable X and an outcome variable Y , thereby inducing a non-causal association between them. (Pearl, 2009). **Observable Confounders** refer to confounders that can be measured, which permit adjustment through statistical methods such as stratification, matching or regression (Pearl, 2009). **Unobservable Confounders** denote latent variables that fulfill confounding criteria but resist direct measurement, which can potentially bias causal estimates when unaccounted for.

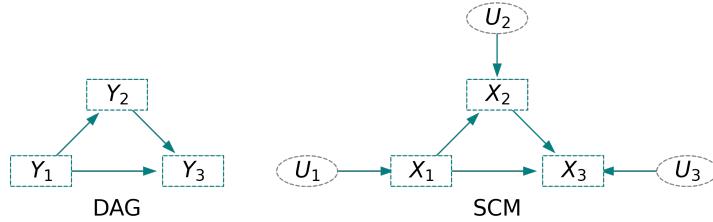


Figure 1: DAG with three nodes (left) and SCM with three exogenous and endogenous nodes (right)

Unobserved explanatory (explained) nodes are unobserved confounder nodes that have no unobserved confounder nodes as its parent (descendant).

Structural Causal Models. Consider a directed acyclic graph (DAG) (such as Fig. 1) \mathcal{G} with nodes $[K]$ in a topologically sorted order¹, where a node k is built with an endogenous random variable X_k defined on a space $\mathcal{X}_k \subset \mathbb{R}^{d_k}$, which has a random exogenous input U_k . Let PA_k be the set of parent nodes of node k in \mathcal{G} and let $X_{\text{PA}_k} := \{X_l\}_{l \in \text{PA}_k}$ be the set of parent random variables of PA_k . A structural causal model (SCM) \mathcal{M} characterizes the relationship between the endogenous variable X_k of a node K with the endogenous variables of its parents X_{PA_k} and its own exogenous variable U_k . Formally, we define $\mathcal{M} := (\mathbf{F}(\mathbf{X}, \mathbf{U}), p_{\mathbf{U}})$, where $\mathbf{F}(\mathbf{X}, \mathbf{U})$ specifies how entire endogenous variables $\mathbf{X} := \{X_1, \dots, X_K\}$ are generated from the set of exogenous random variables $\mathbf{U} := \{U_1, \dots, U_K\}$ with a prior distribution $p_{\mathbf{U}}$. The structural mechanism is governed by $\mathbf{F}(\mathbf{X}, \mathbf{U}) := (f(X_{\text{PA}_1}, U_1), \dots, f(X_{\text{PA}_K}, U_K))$, where each $X_k := f(X_{\text{PA}_k}, U_k)$ for $k \in [K]$ (Pearl, 2009).

Recent advances in deep learning-based causal inference (Chao et al., 2023; Shimizu, 2023), particularly a class of methods integrating structural causal models (Pearl, 2009) with generative models, have demonstrated the effectiveness of diffusion models for answering causal queries. However, spatial heterogeneity and temporal dependencies in unmeasured confounders undermine the validity of existing DCMs and BDCMs for genuine causal inference. For example, in environmental studies, pollution levels may vary significantly across regions (spatial heterogeneity) and exhibit serial correlation over time (temporal dependence), which standard causal models often fail to capture adequately. To overcome this fundamental limitation, we propose a Spatio-Temporal Dynamic Structural Causal Model (ST-DSCM) based on the Backdoor Criterion. We commence by establishing several essential definitions. Due to the complexity of the model, some symbols are mixed up in the article, but the overall readability remains unaffected.

2.1 Partially Functional Spatio-Temporal Dynamic Structural Causal Model

Suppose there is a spatio-temporal dataset $\mathbf{X} := \{X_k\}_{k \in [K]}$ containing K variables across n regions over J time points. Using i to index regions and j for time points, $X_k = (X_{ijk})_{n \times J}$ represent the value of the k -th variable X_k over the time points and regions. We use a DAG to characterize the causal relationships among $\{X_k\}_{k \in [K]}$.

Consider a DAG \mathcal{G} with nodes $[K]$ and a topologically sorted order such that each node k has the X_k as the random variable. Let \mathcal{C}_1 and $\mathcal{C}_2 \subseteq [K]$ be two distinct sets of nodes with unobserved confounders, where \mathcal{C}_1 designates a set of unobserved explanatory nodes and \mathcal{C}_2 denotes a set of unobserved explained nodes. For $h = 1, 2$, let $\mathbf{X}_{\mathcal{C}_h} := \{\mathbf{X}_{\mathcal{C}_h ij}\}_{i \in [n], j \in [J]} := \{X_{qij}\}_{i \in [n], j \in [J], q \in \mathcal{C}_h}$, $\mathbf{U}_{\mathcal{C}_h} := \{\mathbf{U}_{\mathcal{C}_h ij}\}_{i \in [n], j \in [J]} := \{U_{qij}\}_{i \in [n], j \in [J], q \in \mathcal{C}_h}$.

To incorporate spatio-temporal dynamic structures among unmeasured confounders, and to account for spatial heterogeneity, we posit a relationship such that

$$\mathbf{X}_{\mathcal{C}_2 ij} = \boldsymbol{\Gamma}_i \mathbf{G}(\mathbf{X}_{\mathcal{C}_1 ij}) + \mathbf{U}_{\mathcal{C}_2 ij}, \quad (1)$$

where $\boldsymbol{\Gamma}_i$ is a $\dim(\mathbf{X}_{\mathcal{C}_2 ij}) \times \dim(\mathbf{X}_{\mathcal{C}_1 ij})$ structural coefficient matrix, which are permitted to vary across regions; $\mathbf{G}(\cdot) = (g_1(\cdot), \dots, g_q(\cdot))^{\top}$ is $q \times 1$ nonzero vector-valued function with differentiable functions g_1, \dots, g_q , and $q \geq \dim(\mathbf{X}_{\mathcal{C}_1 ij})$. And to capture the temporal dependence, we let $\mathbf{X}_{\mathcal{C}_1 i} = (\mathbf{X}_{\mathcal{C}_1 i 1}^{\top}, \dots, \mathbf{X}_{\mathcal{C}_1 i J}^{\top})^{\top}$ and $\mathbf{U}_{\mathcal{C}_2 i} = (\mathbf{U}_{\mathcal{C}_2 i 1}^{\top}, \dots, \mathbf{U}_{\mathcal{C}_2 i J}^{\top})^{\top}$, then, the covariance matrices $\text{Var}(\mathbf{X}_{\mathcal{C}_1 i})$ and $\text{Var}(\mathbf{U}_{\mathcal{C}_2 i})$ of $\mathbf{X}_{\mathcal{C}_1 i}$ and $\mathbf{U}_{\mathcal{C}_2 i}$ can be expressed as

$$\text{Var}(\mathbf{X}_{\mathcal{C}_1 i}) = \mathbf{D}_{\mathcal{C}_1 i} \otimes \text{Var}(\mathbf{X}_{\mathcal{C}_1}), \quad \text{Var}(\mathbf{U}_{\mathcal{C}_2 i}) = \mathbf{D}_{\mathcal{C}_2 i} \otimes \text{Var}(\mathbf{U}_{\mathcal{C}_2}), \quad (2)$$

where $\mathbf{D}_{\mathcal{C}_h i} (h \in \{1, 2\})$ are the $J \times J$ adjacent time covariance matrices, $\text{Var}(\cdot)$ represent between-variable covariances. In order to establish a rigorous framework for the temporal adjacency structure, the conditional autoregressive (CAR) model (Besag, 1974) is adopted with the forms

$$\mathbf{D}_{\mathcal{C}_h i} = (\mathbf{I}_J - \rho_{\mathcal{C}_h i} \mathbf{H}_{\mathcal{C}_h i})^{-1}, \quad (3)$$

¹A topological order is a linear arrangement of variables where a variable appears after all its direct causes (parents) (Pearl, 2009)

where $\rho_{C_h i}(h \in \{1, 2\})$ are adjacent time association parameters, and $\mathbf{H}_{C_h i}(h \in \{1, 2\})$ are $J \times J$ adjacency matrices in which the element $h_{ji} = 1$ implies that time l is adjacent to time j and otherwise $h_{jl} = 0$.

However, real-world applications often present significant challenges in elucidating causal relationships among variables observed under heterogeneity. To address these complexities, we extend the ST-DSCM by incorporating functional random variables $\mathbf{X}_k(t)$, leading to a Partially Functional Spatio-Temporal Dynamic Structural Causal Model (PFST-DSCM)². We adopt a basis expansion framework to achieve dimensionality reduction in the functional space via a set of orthogonal basis functions $\{\mathbf{b}_1, \dots, \mathbf{b}_{K_n}\} \in \mathbb{R}^{T \times K_n}$. Let

$$X_{ijm} = \int \mathbf{b}_m(t) X_{ij}(t) dt, \quad (4)$$

for $i \in [n], j \in [J], m \in [K_n]$, which are used as nodes within the ST-DSCM, and detailed mathematical derivations are provided in Appendix A.

Notably, our model operates in settings where both observational data and causal structures are available, enabling it to answer observational, interventional, and counterfactual queries. All parameters in the structural model are assumed to be known in this context; in practice, they can be estimated using methods from Song et al. (2012) and Tang et al. (2017).

2.2 Partially Functional Dynamic Backdoor Diffusion-based Causal Model

We now present the PFD-BDCM, a model designed to handle causal queries under the PFST-DSCM framework, which explicitly accounts for unmeasured confounders with spatial heterogeneity and temporal dependence. The model employs an encoder-decoder architecture to enable causal reasoning across multi-resolution variables.

Let \mathcal{B}_k be the set of backdoor nodes³ of k , and $\mathbf{X}_{\mathcal{B}_k} := \{X_l\}_{l \in \mathcal{B}_k}$ represent the variables on \mathcal{B}_k . Here, the set of parent nodes PA_k satisfies the definition of backdoor nodes, that is, $\text{PA}_k \subseteq \mathcal{B}_k$, we will use \mathcal{B}_k instead of PA_k . We assume that the unobserved random variables are jointly independent (Markovian SCM), and the Partially Functional Spatio-Temporal Dynamic Directed Acyclic Graph (PFST-DDAG) \mathcal{G} is the graph induced by PFST-DSCM \mathcal{M} . Every PFST-DSCM \mathcal{M} entails a unique joint observational distribution satisfying the causal Markov assumption: $p(\mathbf{X}) = \prod_{k=1}^K p(X_k | X_{\mathcal{B}_k})$.

The PFD-BDCM's data-generating process is formalized as: $\{\mathbf{X}_k : X_{ijk} = f_{ij}(X_{\mathcal{B}_k}, \mathbf{U}_k)\}_{k \in [K]}$. The encoder g maps $(\mathbf{X}_k, \mathbf{X}_{\mathcal{B}_k})$ to a latent variable $\mathbf{Z}_k := g(\mathbf{X}_k, \mathbf{X}_{\mathcal{B}_k})$, which captures information of the unmeasured confounders \mathbf{U}_k . The decoder h reconstructs \mathbf{X}_k as $\hat{\mathbf{X}}_k = h(\mathbf{Z}_k, \mathbf{X}_{\mathcal{B}_k})$. Perfect reconstruction $\hat{\mathbf{X}}_k = \mathbf{X}_k$ implies that h approximates the true structural function f . We next detail the architecture and training of PFD-BDCM, and then describe its use in causal query answering.

Our approach leverages the generative power of diffusion models to learn the complex functional relationships inherent in a Structural Causal Model. The core idea is to represent the structural equation for each endogenous variable X_k with a dedicated conditional diffusion model. This model learns the distribution $p(X_k | X_{\text{PA}_k}, \mathbf{U}_k)$. Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) approximate a target data distribution $q(x^0)$ via a two-stage process. First, a fixed **forward process** gradually injects Gaussian noise into the data x^0 over T steps. The distribution of the noisy data x^t at step t is a Gaussian: $q(x^t | x^0) = \varphi(x^t; \sqrt{\alpha_t}x^0, (1 - \alpha_t)I)$, where $\varphi(x; \mu, \Sigma)$ denote the Gaussian density with mean μ and covariance Σ . Here, $\alpha_t := \prod_{s=1}^t (1 - \beta_s)$ where β_s is a predefined noise schedule at each time step s . As $t \rightarrow T$, x^T converges to a standard Gaussian distribution. Second, a learnable **reverse process**, parameterized by θ , is trained to denoise the data. This is achieved by training a neural network ϵ_θ to predict the added noise ϵ using the available observational data. For learning the distribution $p(X_k | X_{\text{PA}_k}, \mathbf{U}_k)$, the network ϵ_θ is trained on samples where X_k serves as the target variable x^0 and X_{PA_k} provides the conditioning context c . Here θ represents learnable parameters of the neural network, conditioned on the noisy data x^t , the step t , and the parental variables contextual information c . The objective function is

$$\mathbb{E}_{t, x^0, c, \epsilon} [\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}x^0 + \sqrt{1 - \alpha_t}\epsilon, c, t)\|^2]. \quad (5)$$

In our framework, the conditioning context c for a variable X_k is its set of parent variables X_{PA_k} (or $X_{\mathcal{B}_k}$).

²For example, Figure 2 is a PFST-DSCM with 33 exogenous and endogenous nodes, where nodes X_{28}, X_{29} and X_{30} are unmeasured confounders with spatial heterogeneity and temporal dependencies, $Y_1(t), Y_2(t), Y_3(t)$ are functional nodes, and X_4, \dots, X_{21} are the corresponding base expansion nodes.

³A set of node \mathcal{B} satisfies backdoor criterion (Pearl et al., 2016) for tuple (X, Y) in DAG \mathcal{G} if no node in \mathcal{B} is a descendant of X and \mathcal{B} blocks all paths between X (cause) and Y (outcome) which contains an arrow into X .

Although diffusion models excel at data generation, causal inference, especially counterfactual reasoning, requires a deterministic mapping between observations and latent codes. Denoising Diffusion Implicit Models (DDIMs) (Song et al., 2021) provide such a deterministic non-Markovian reverse process, a property essential for identifiability. In PFD-BDCM, we extend DDIM by incorporating backdoor adjustment sets as additional covariates. The resulting diffusion model for node k is denoted $\epsilon_\theta^k(X_k, \mathbf{X}_{\mathcal{B}_k}, t)$.

Formally, for each node $k \in [K]$, the latent variable $Z_k := Z_k^T$ is generated through the forward implicit diffusion process

$$Z_k^{t+1} := \sqrt{\alpha_{t+1}/\alpha_t} Z_k^t + \epsilon_\theta^k(Z_k^t, \mathbf{X}_{\mathcal{B}_k}, t) (\sqrt{1 - \alpha_{t+1}} - \sqrt{\alpha_{t+1}(1 - \alpha_t)/\alpha_t}), \quad (6)$$

for $t = 0, \dots, T - 1$, initialized with $Z_k^0 := X_k$. This latent representation Z_k serves as a proxy for the exogenous noise U_k . The reconstruction $\hat{X}_k := \hat{X}_k^0$ is obtained via the reverse implicit diffusion process

$$\hat{X}_k^{t-1} := \sqrt{\alpha_{t-1}/\alpha_t} \hat{X}_k^t - \epsilon_\theta^k(\hat{X}_k^t, \mathbf{X}_{\mathcal{B}_k}, t) \left(\sqrt{\alpha_{t-1}(1 - \alpha_t)/\alpha_t} - \sqrt{1 - \alpha_{t-1}} \right), \quad (7)$$

for $t = T, \dots, 1$, initialized with $\hat{X}_k^T := Z_k$. The encoding and decoding functions for node k are denoted as $\text{Enc}_k(X_k, \mathbf{X}_{\mathcal{B}_k})$ (Eq. (6)) and $\text{Dec}_k(Z_k, \mathbf{X}_{\mathcal{B}_k})$ (Eq. (7)) respectively, and the pseudocodes are provided in Appendix B.1.

Training PFD-BDCMs. The comprehensive training methodology (Appendix B.2 Algorithm 3) incorporates backdoor adjustment sets as covariates while training distinct diffusion models per node. Crucially, generative models for endogenous nodes exhibit mutual independence during training, thereby enabling parallelized optimization. This parallelism is feasible since each diffusion model necessitates only its target node's values and corresponding backdoor adjustment set values. The final PFD-BDCM architecture integrates these K trained diffusion models $\{\epsilon_\theta^k\}_{k \in [K]}$.

We now elucidate the methodology for leveraging trained PFD-BDCMs to approximate diverse causal queries. Resolution of observational and interventional queries necessitates sampling from their respective observational and interventional distributions. Counterfactual queries, however, operate at unit granularity by modifying structural equation assignments while preserving the latent exogenous noise variables consistent with empirical observations.

Generating samples for observational/interventional queries. To generate samples approximating the interventional distribution $p(\mathbf{X}|\text{do}(X_{\mathcal{L}} := \gamma))$ using a trained PFD-BDCM model, we implement the following procedure: i) For intervened nodes $l \in \mathcal{L}$, we set $\hat{X}_l := \gamma_l$ deterministically; ii) For root nodes k , sample \hat{X}_k from empirical training distributions; iii) For non-intervened nodes $k \notin \mathcal{L}$, sample latent vectors $Z_k \sim \mathcal{N}(\mathbf{0}, I_{d_k})$, where $d_k = \dim(X_k)$, and subsequently compute $\hat{X}_k := \text{Dec}_k(Z_k, \hat{\mathbf{X}}_{\mathcal{B}_k})$ utilizing inductively generated backdoor variable values $\hat{\mathbf{X}}_{\mathcal{B}_k}$. Generated values propagate to child nodes as backdoor inputs. Observational sampling ($p(\mathbf{X})$) corresponds to $\mathcal{L} = \emptyset$, with pseudocode formalized in Appendix B.3 (Algorithm 4).

Counterfactual Queries. To compute counterfactual estimates $\hat{\mathbf{x}}^{\text{CF}}$ within the PFD-BDCM framework, given factual observation $\mathbf{x}^{\text{F}} := (x_1^{\text{F}}, \dots, x_K^{\text{F}})$ and intervention set \mathcal{L} with values γ , we implement the following systematic procedure: i) For intervened nodes $l \in \mathcal{L}$, assign $\hat{x}_l^{\text{CF}} := \gamma_l$ deterministically; ii) For non-intervened descendant nodes k , using inductively generated backdoor estimates $\hat{\mathbf{x}}_{\mathcal{B}_k}^{\text{CF}}$, we compute $\hat{x}_k^{\text{CF}} := \text{Dec}_k(\text{Enc}_k(x_k^{\text{F}}, \mathbf{x}_{\mathcal{B}_k}^{\text{F}}), \hat{\mathbf{x}}_{\mathcal{B}_k}^{\text{CF}})$, where factual noise is implicitly encoded. The complete formalization appears in Appendix B.3 (Algorithm 5).

3 Counterfactual Error Bounds

In this section, we establish the theoretical guarantees for the counterfactual estimation accuracy of the PFD-BDCM framework. The primary contribution is the derivation of an error bound that formally links the reconstruction fidelity of the encoder-decoder architecture to the precision of its counterfactual predictions. Notably, these theoretical guarantees accommodate higher-dimensional settings through additional structural assumptions. Formal proofs are presented in Appendix C.

Consider an endogenous variable X_k governed by structural equation $X_k := f_{ij}(X_{\mathcal{B}_k}, U_k)$ with backdoor adjustment set $X_{\mathcal{B}_k}$ and exogenous noise U_k . We analyze a single node without loss of generality (by permutation invariance

of nodes), henceforth denoting the target variable as $X \in \mathcal{X} \subseteq \mathbb{R}$, its backdoors as $X_{\mathcal{B}} \in \mathcal{X}_{\mathcal{B}} \subseteq \mathbb{R}^K$, and exogenous noise as U . The encoder-decoder architecture comprises

$$g : \mathcal{X} \times \mathcal{X}_{\mathcal{B}} \rightarrow \mathcal{Z} \quad (\text{encoding function}); \quad h : \mathcal{Z} \times \mathcal{X}_{\mathcal{B}} \rightarrow \mathcal{X} \quad (\text{decoding function}),$$

where \mathcal{Z} denotes the latent space. Within PFD-BDCM, g and h correspond to the Enc and Dec operators, respectively.

Our theoretical results rely on a set of assumptions regarding the structural equation and the encoder-decoder model. These conditions are essential for ensuring that the latent variable learned by the encoder can uniquely recover the unobserved exogenous noise, which is the cornerstone of accurate counterfactual estimation (Lu et al., 2020; Nasr-Esfahany and Kiciman, 2023; Nasr-Esfahany et al., 2023). For a variable $X \in \mathcal{X} \subset \mathbb{R}$ with structural equation $X := f_{ij}(X_{\mathcal{B}}, U)$ where exogenous noise $U \sim \mathcal{N}(0, \psi)$ and $U \perp\!\!\!\perp X_{\mathcal{B}}$, we have the following assumptions:

Assumption 1. The encoded latent variable is independent of the backdoor variables, $g(X, X_{\mathcal{B}}) \perp\!\!\!\perp X_{\mathcal{B}}$.

Assumption 2. The structural equation f_{ij} is differentiable and strictly increasing with respect to U for all values of the backdoor variables $x_{\mathcal{B}} \in \mathcal{X}_{\mathcal{B}}$.

Assumption 3. The encoding function g is invertible and differentiable with respect to its first argument X for all $x_{\mathcal{B}} \in \mathcal{X}_{\mathcal{B}}$.

These assumptions, while formal, are well-motivated in the context of causal inference and deep generative models. Assumption 1 ensures that the encoder learns a representation of the exogenous noise that is not contaminated by information from the backdoor variables. This is naturally satisfied in settings like additive noise models with $f_{ij}(X_{\mathcal{B}}, U) = f'_{ij}(X_{\mathcal{B}}) + U$ where $X_{\mathcal{B}}$ and U is independent. If the fitted model $\hat{f}_{ij} \equiv f'_{ij}$, then $g(X, X_{\mathcal{B}}) = U$. Assumption 2 is satisfied by major identifiable model classes, including additive noise, post-nonlinear, and heteroscedastic formulations (Strobl and Lasko, 2023) while concurrently resolving symmetric noise ambiguities characteristic of observational data. This assumption further aligns with contemporary identifiability frameworks (Nasr-Esfahany and Kiciman, 2023) and intrinsically precludes non-identifiable structural equations. The encoder invertibility condition (Assumption 3) is intrinsically satisfied by the bijective properties of deterministic diffusion architectures (Song et al., 2021), guaranteeing uniqueness in latent representations while preserving compatibility with standard implementations.

Under these assumptions, we can prove that the encoder successfully isolates the exogenous noise up to an invertible transformation.

Theorem 1. *Under Assumptions 1, 2, and 3, the encoded latent variable $Z = g(X, X_{\mathcal{B}})$ is an invertible transformation of the true exogenous noise U . That is, there exists an invertible function \tilde{q} such that $Z = \tilde{q}(U)$.*

Theorem 1 provides the foundation for assessing the accuracy of counterfactuals given by the PFD-BDCM. It implies that the abduction step, $\text{Enc}(X, X_{\mathcal{B}})$, correctly captures the essence of the unobserved confounder u that generated the factual observation. We now explore the direct consequences of this result.

In an oracle scenario where the model achieves perfect reconstruction which means $h(g(X, X_{\mathcal{B}}), X_{\mathcal{B}}) = X$ holding almost surely (a.s.), the counterfactual estimate will be “perfect”. This precise case implies the satisfaction of Theorem 1, yielding the relationship $h(\tilde{q}(U), X_{\mathcal{B}}) = f_{ij}(X_{\mathcal{B}}, U)$. Consequently, when making a counterfactual prediction for a new intervention $X_{\mathcal{B}} := \gamma$, the model computes $h(\tilde{q}(\gamma), \gamma)$, which a.s. equates to the true counterfactual $f_{ij}(\gamma, u)$.

Corollary 1. *Assume Assumptions 1, 2, and 3 hold and the encoder-decoder model pair (g, h) achieves perfect reconstruction, i.e., $h(g(X, X_{\mathcal{B}}), X_{\mathcal{B}}) = X$ a.s.. For a factual observation pair $(x, x_{\mathcal{B}})$ generated by $x = f_{ij}(x_{\mathcal{B}}, u)$ and a counterfactual intervention $\text{do}(X_{\mathcal{B}} := \gamma)$, the estimated counterfactual $\hat{x}^{\text{CF}} := h(g(x, x_{\mathcal{B}}), \gamma)$ is a.s. identical to the true counterfactual $x^{\text{CF}} := f_{ij}(\gamma, u)$.*

More practically, models are not perfect. Corollary 2 allows us to bound the counterfactual error by the model’s reconstruction error. This is a powerful result, as it connects a measurable property of the model (how well it auto-encodes data) to its performance on a causal task.

Corollary 2. *Under Assumptions 1, 2, and 3, if the model’s reconstruction error is bounded by τ under a metric d , such that $d(h(g(X, X_{\mathcal{B}}), X_{\mathcal{B}}), X) \leq \tau$, then the counterfactual estimation error is also bounded by τ . For a factual observation $(x, x_{\mathcal{B}})$ and intervention $\text{do}(X_{\mathcal{B}} := \gamma)$, we have $d(\hat{x}^{\text{CF}}, x^{\text{CF}}) \leq \tau$.*

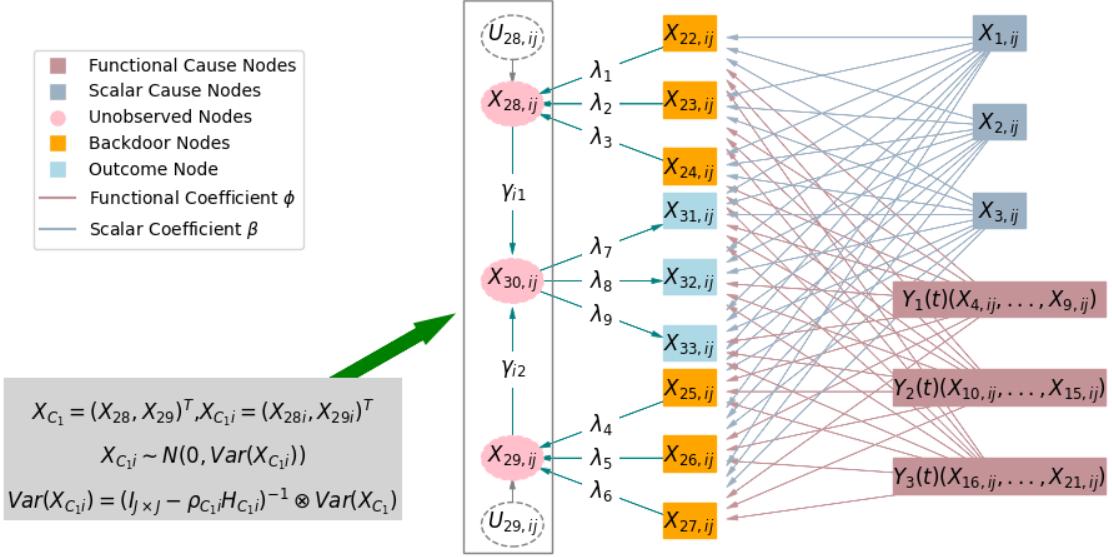


Figure 2: PFST-DSCM with 33 exogenous and endogenous nodes (where nodes X_{28} , X_{29} and X_{30} are unmeasured confounders with spatial heterogeneity and temporal dependencies, $Y_1(t)$, $Y_2(t)$, $Y_3(t)$ are functional nodes, and X_4, \dots, X_{21} are the corresponding base expansion nodes)

This corollary formally establishes that minimizing the reconstruction loss during training directly optimizes the model for better counterfactual prediction. The framework extends to the more general multivariate setting where $X \in \mathbb{R}^d$ (Theorem 2). This requires a stronger assumption on the encoder's Jacobian to ensure that information is not lost in the higher-dimensional space. We present it in Appendix C.

4 Experimental Evaluation

We empirically evaluated the efficacy of PFD-BDCM in addressing causal queries across synthetic and real-world datasets. To demonstrate that PFD-BDCM faithfully samples from the target interventional distribution, we designed scenarios where causal sufficiency was deliberately violated within Partially functional structural models.

4.1 Simulation Study

Figure 2 presented the instantiated PFST-DSCM, where causal sufficiency was compromised. Consider $\{X_{28}, X_{29}\}$ as unobserved explanatory variables and X_{30} as an unobserved explained variable, we assumed exhibit pronounced spatial heterogeneity coupled with temporal dependence between $\{X_{28}, X_{29}\}$ and X_{30} . Let X_1, X_2, X_3 represent endogenous cause variables; X_{31}, X_{32}, X_{33} denote outcome variables; and $\mathbf{X}_B = \{X_{22}, \dots, X_{27}\}$ constitute backdoor adjustment sets. For visual clarity, exogenous noise terms \mathbf{U} were omitted.

The partially functional dynamic structural equations were defined as: $X_{ijk} = f_{ijk}(X_{Bk}, U_k)$ (Eq. (37) in Appendix D). The structural equations governing the Partially Functional Dynamic Diffusion-based Causal Model (PFD-DCM) and PFD-BDCM were instantiated with additive noise models (ANM) (Peters et al., 2013) furnishing elementary baselines.

Our objective was to accurately sample from the post-interventional distribution $q(X_k | \text{do}(X_l = \gamma_l))$, where $k \in \{31, 32, 33\}$ indexes outcomes and $l \in \{1, 2, 3\}$ indexes causes. During intervention, X_l is fixed to γ_l , while root variables X_h ($h = 1, \dots, 21$) were sampled from their empirical marginals E_h . For outcome X_k , PFD-DCM (DCM) employed $\text{Dec}_k(Z_k, \hat{X}_h)$, whereas PFD-BDCM (BDCM) utilized $\text{Dec}_k(Z_k, \hat{X}_{Bk}, \hat{X}_h)$, thereby leveraging backdoor adjustments. Comprehensive simulation setting was detailed in the Appendix D.

Table 1 summarizes aggregated performance metrics-observational (Obs.), interventional (Int.), and counterfactual (CF.)-averaged over nine independent random initializations. Comprehensive diagnostics (boxplots (Fig. 4 and 5) and kernel density estimates (Fig. 3)) were provided in Appendix D. PFD-DCM and PFD-BDCM achieved compelling statistical fidelity across all query types, evidenced by MMD for observational/interventional query and MSE

Table 1: Mean \pm standard deviation of $\text{MMD}^2 (\times 10^{-3})$, MSE and Time (seconds) of PFD-BDCM, PFD-DCM, BDCM and DCM compared to the true target distribution (simulation)

Causal query	$J = 6$	PFD-BDCM	PFD-DCM	BDCM	DCM
$\downarrow \text{MMD}$ (Obs.)	$n = 30$	3.616 ± 4.368	4.054 ± 3.585	3.739 ± 3.709	4.934 ± 4.249
	$n = 80$	1.494 ± 1.412	1.560 ± 1.300	3.376 ± 4.936	4.037 ± 5.191
	$n = 200$	0.533 ± 0.486	0.737 ± 0.672	3.032 ± 4.471	3.474 ± 4.658
$\downarrow \text{Time}$ (Obs.)	$n = 30$	2.665	2.705	2.660	2.697
	$n = 80$	8.837	8.570	8.830	8.566
	$n = 200$	15.671	15.309	15.665	15.306
$\downarrow \text{MMD}$ (Int.)	$n = 30$	3.580 ± 3.013	4.282 ± 3.952	3.990 ± 3.962	3.922 ± 3.247
	$n = 80$	1.408 ± 1.204	1.511 ± 1.268	2.009 ± 2.052	2.079 ± 2.003
	$n = 200$	0.595 ± 0.504	0.592 ± 0.582	2.803 ± 3.620	3.187 ± 3.163
$\downarrow \text{Time}$ (Int.)	$n = 30$	2.291	2.171	2.286	2.165
	$n = 80$	5.997	5.972	5.992	5.969
	$n = 200$	15.584	15.565	15.579	15.560
$\downarrow \text{MSE}$ (CF.)	$n = 30$	0.835 ± 0.259	0.636 ± 0.236	1.936 ± 0.085	1.947 ± 0.075
	$n = 80$	0.645 ± 0.130	0.212 ± 0.055	1.982 ± 0.024	1.980 ± 0.027
	$n = 200$	0.601 ± 0.089	0.081 ± 0.020	1.990 ± 0.010	1.992 ± 0.013
$\downarrow \text{Time}$ (CF.)	$n = 30$	1.135	1.037	1.136	1.042
	$n = 80$	3.625	3.432	3.812	3.751
	$n = 200$	8.616	8.621	8.173	8.023

for counterfactual query. PFD-BDCM consistently outperforms baselines, demonstrating superior MMD metrics in observational queries and enhanced stability in counterfactual queries under varying data scales. This advantage stems from its principled integration of backdoor adjustment with spatiotemporal modeling of unobserved confounders, which mitigates information loss in latent nodes and corrects confounding-induced biases, significantly improving causal query performance.

4.2 Empirical Application

This investigation employed the PFD-BDCM framework to examine spatio-temporal dynamic structural causal relationships among air pollutant indicators and their determinants. Our analysis encompassed 30 provincial-level administrative divisions across Chinese mainland during the period January 2015 to December 2020. The study integrates China's provincial CO₂ emission inventories from the China Emission Accounts and Datasets (CEADs) (Guan et al., 2021; Xu et al., 2024) and emissions data for nine atmospheric pollutants from the Multi-scale Emission Inventory of China (MEIC) (Li et al., 2019; Geng et al., 2024) as response variables for air pollutant emissions.

Building upon prior research (Ozcan, 2013; Zhu et al., 2021) and incorporating domain-specific characteristics of regional emissions, we systematically collected foundational determinants across ten conceptual dimensions. The comprehensive dataset comprised 118 indicator variables, through collinearity diagnostics and random forest-based feature selection, we retained 49 statistically robust indicators for subsequent modeling (detailed indicators shown in Appendix D.2 Table 8. Complete experimental specifications and supplementary materials were documented in Appendix D.2, with observational query results presented in Table 2.

Table 2: Mean \pm standard deviation of $MMD^2(\times 10^{-2})$ of PFD-DCM and PFD-BDCM compared to the true target distribution(Observation query)

Variable	PFD-DCM	PFD-BDCM	Variable	PFD-DCM	PFD-BDCM
SO_2	0.495 ± 0.419	0.512 ± 0.476	PM_{10}	0.487 ± 0.426	0.485 ± 0.457
NO_x	0.511 ± 0.449	0.473 ± 0.424	$PM_{2.5}$	0.400 ± 0.356	0.435 ± 0.400
CO	0.564 ± 0.531	0.524 ± 0.450	BC	0.417 ± 0.406	0.350 ± 0.339
VOC	0.489 ± 0.464	0.480 ± 0.403	OC	0.487 ± 0.462	0.398 ± 0.413
NH_3	0.519 ± 0.538	0.494 ± 0.446	CO_2	0.336 ± 0.257	0.357 ± 0.304

5 Concluding Remarks

We propose the Partially Functional Dynamic Backdoor Diffusion-based Causal Model (PFD-BDCM), a methodological framework crafted for robust causal inference amidst spatial heterogeneity, temporal dependencies, and unmeasured confounding. Our contributions are threefold.

Model Innovation: PFD-BDCM synergistically integrates functional basis expansions with diffusion-based causal modeling, facilitating simultaneous resolution of: i) Multi-resolution variables through partially functional representations; ii) Spatio-temporal dynamics via regionally parameterized structural equations; iii) Unmeasured confounder bias utilizing backdoor adjustment sets.

Theoretical Foundation: We establish pioneering error bounds formally connecting counterfactual estimation accuracy to encoder-decoder reconstruction fidelity under: i) Monotonic structural functional constraints; ii) Invertible encoding operators; iii) Multivariate generalizations with supplementary structural assumptions.

Empirical Validation: Comprehensive experiments on synthetic and real-world data demonstrate that PFD-BDCM significantly outperforms existing methods in answering observational, interventional, and counterfactual queries.

While PFD-BDCM advances causal inference in complex settings, future work should address: i) Scalability enhancements for ultra-high-dimensional functional data via tensor decomposition; ii) Automated backdoor set identification through causal discovery algorithms; iii) Temporal graph neural network integration for non-stationary processes; iv) Real-time deployment in environmental policy decision support systems.

The proposed framework opens new avenues for causal inference in environmental science, epidemiology, and econometrics, where functional data and unmeasured confounders are prevalent.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the royal statistical society series b-methodological*, 36:192–225.

Chao, P., Blöbaum, P., and Kasiviswanathan, S. P. (2023). Interventional and counterfactual inference with diffusion models.

Geng, G., Liu, Y., Liu, Y., and et al (2024). Efficacy of china’s clean air actions to tackle pm2.5 pollution between 2013 and 2020. *Nature Geoscience*, 17(10).

Guan, Y., Shan, Y., Huang, Q., and et al (2021). Assessment to china’s recent emission pattern shifts. *Earth’s Future*, 9(11).

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, pages 604–620.

Li, M., Zhang, Q., Zheng, B., and et al (2019). Persistent growth of anthropogenic nmvoc emissions in china during 1990-2017: Dynamics, speciation, and ozone formation potentials. *Atmospheric Chemistry and Physics*, 19(13):8897–8913.

Lu, C., Huang, B., Wang, K., and et al (2020). Sample-efficient reinforcement learning via counterfactual-based data augmentation.

Nasr-Esfahany, A., Alizadeh, M., and Shah, D. (2023). Counterfactual identifiability of bijective causal models.

Nasr-Esfahany, A. and Kiciman, E. (2023). Counterfactual (non-) identifiability of learned structural causal models.

Ozcan, B. (2013). The nexus between carbon emissions, energy consumption and economic growth in middle east countries: A panel data analysis. *Energy Policy*, 62:1138–1147.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146.

Pearl, J., Glymour, M., and Jewell, N. (2016). Causal inference in statistics: A primer. *Wiley*.

Peters, J., Mooij, J., Janzing, D., and et al (2013). Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Shalit, U., Johansson, F. D., and Sontag, D. (2017). Estimating individual treatment effect: Generalization bounds and algorithms. *International Conference on Machine Learning*, pages 3076–3085.

Shimizu, T. (2023). Diffusion model in causal inference with unmeasured confounders. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 683–688.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and et al (2015). Deep unsupervised learning using nonequilibrium thermodynamics. *Proceedings of Machine Learning Research*, 37:2256–2265.

Song, J., Meng, C., and Ermon, S. (2021). Denoising diffusion implicit models. *International Conference on Learning Representations*.

Song, X., Tang, N., and Chow, S. (2012). A bayesian approach for generalized random coefficient structural equation models for longitudinal data with adjacent time effects. *Computational Statistics & Data Analysis*, 56(12):4190–4203.

Strobl, E. V. and Lasko, T. A. (2023). Identifying patient-specific root causes with the heteroscedastic noise model. *Journal of Computational Science*, 72:102099.

Tang, N., Chow, S.-M., Ibrahim, J. G., and Zhu, H. (2017). Bayesian sensitivity analysis of a nonlinear dynamic factor analysis model with nonparametric prior and possible nonignorable missingness. *Psychometrika*, 82(4):875–903.

Xu, J., Guan, Y., Oldfield, J., Guan, D., and Shan, Y. (2024). China carbon emission accounts 2020-2021. *Applied Energy*, 360.

Zhu, Y., Liang, Y., and Chen, S. X. (2021). Assessing local emission for air pollution via data experiments. *Atmospheric Environment*, 252:118323.