Intervention-Aware Time Series Modeling: Capturing and Evaluating Feature Dependencies

Ibrahim Delibasoglu[1 2], **Sanjay Chakraborty**[1], **Fredrik Heintz**[1], **Mattias Tiger**[1]

[1]ReaL Artificial Intelligence and Integrated Computer Systems (AIICS) Department of Computer and Information Science (IDA), Linköping University Sweden

[2]Department of Software Engineering, Sakarya University, Sakarya, Türkiye

ibrahimdelibasoglu@sakarya.edu.tr, sanjay.chakraborty@liu.se, fredrik.heintz@liu.se, mattias.tiger@liu.se

Abstract

Understanding how localized changes in one variable affect others in multivariate time series is essential for diagnostics and decision-making in complex systems. Existing models often fail to capture realistic inter-feature dynamics when simulating "what-if" scenarios, leading to inaccurate or uncorrelated reconstructions. We propose CFORVAE, a variational autoencoder framework that explicitly addresses this limitation by combining temporal decomposition with frequency-domain feature correlation modeling. Our architecture uses a dual-path encoding of trend and seasonal components, each projected into attention-pooled latent spaces, and applies Fourier Neural Operators (FNO) to capture cross-feature dependencies in the spectral domain. This decomposition-correlation design enables component-specific latent manipulation and ensures that local modifications propagate realistically across correlated variables. Through extensive experiments, we show that CFORVAE outperforms state-of-the-art baselines in preserving temporal and feature-level dependencies, especially under adjustment-based reconstructions, making it a powerful tool for interpretable "what-if" analysis and diagnostics.

Code — https://github.com/mribrahim/cforvae

Introduction

Multivariate time series analysis is vital for understanding complex systems in fields like healthcare, finance, and industrial applications. Recent time series methods have primarily focused on forecasting or imputing missing data, without explicitly revealing the relationships between variables. Recent advances in deep generative models, particularly variational autoencoders (VAEs) and their conditional extensions (Sohn, Lee, and Yan 2015), have enabled the modeling of nonlinear dependencies and the generation of realistic data. However, these models often lack interpretability and control over specific feature interactions (Stepin et al. 2021). The technical challenges in achieving these goals are substantial. The model must simultaneously learn a compressed representation of the time series dynamics while maintaining separable feature dependencies in latent space (Locatello et al. 2019). This work addresses these challenges by introducing an approach that combines several innovations: a structured latent space that disentangles feature-specific and shared temporal patterns, a mechanism to learn pairwise feature dependencies, and a reconstruction process that respects these relationships.

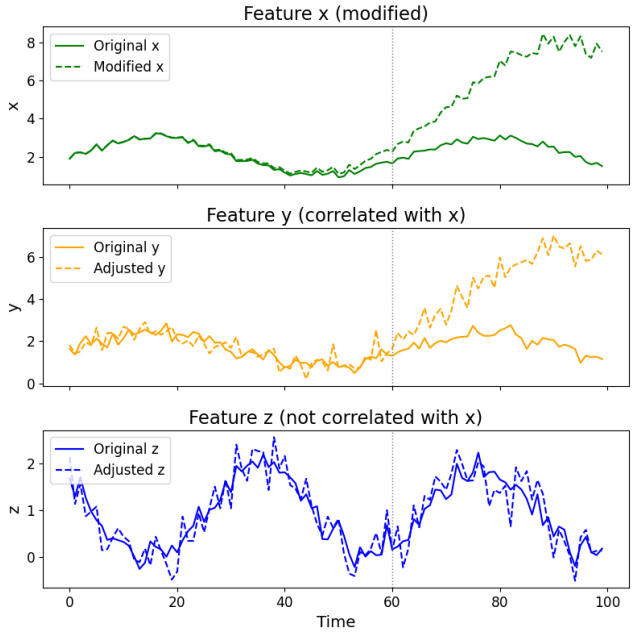

Figure 1: Correlated feature influence under controlled changes. Feature x is modified with a smooth ramp function (dashed). Feature y, which is correlated with x, adjusts accordingly, while feature z, which is uncorrelated with x, remains largely unchanged.

In many real-world applications, particularly in industrial and operational decision-making systems, users often need to manually adjust a single input feature and assess its effect on downstream targets as represented in Figure1. However, such manual interventions typically ignore the inherent correlations between features, potentially leading to implausible or unsafe configurations. While counterfactual methods can optimize inputs for a desired output, they may generate feature combinations that violate real-world dependencies (Mehdiyev, Majlatow, and Fettke 2024; Verma et al. 2024).

Aspect	Standard Generation	Counterfactual Modeling	Intervention-Aware Reconstruction
Goal	Generate realistic sequences from historical data	Estimate alternative outcomes under specified interventions	Estimate time-dependent outcomes under continuous, feature-level perturbations
Input	Full or past time series	Static treatment indicator or binary feature change	Continuous perturbation to one feature + contextual time series
Dependency Modeling	Learns implicit feature/temporal correlations	Based on assumed or learned causal graphs	Learns soft inter-feature dependencies
Intervention Type	Not modeled	Discrete (often binary or categorical)	Continuous, fine-grained adjustments in feature space
Output	Realistic sequences or forecasts	Counterfactual scalar or label prediction	Multivariate time series reconstruction reflecting propagated effects
Application	Forecasting, data augmentation	Fairness analysis, treatment effect estimation	Fine-grained what-if simulation, decision support, diagnostics

Table 1: Comparison of standard time series generation, counterfactual modeling, and intervention-aware reconstruction.

To address this, we focus on a semi-autonomous generative framework that, upon updating one input feature, automatically adjusts other correlated features in a data-consistent way. This allows the operator to explore the effect of plausible interventions while preserving the joint feature distribution. *Our main contribution is a conditional generative model that learns intra-timestep feature correlations from time series data and enables controlled, correlation-aware feature updates—bridging the gap between counterfactual reasoning and practical, interpretable input manipulation.*

We propose a decomposition-driven variational framework that models both temporal dynamics and cross-feature interactions in multivariate time series by separating inputs into seasonal and trend components, each processed through spectral encoders. Leveraging Fourier Neural Operators (FNOs) for global frequency-domain correlation learning and a conditional reconstruction strategy with difference propagation, our method enables feature-level interventions—essential for sensitivity analysis and planning. Unlike prior generative or causal models, our approach explicitly targets intra-timestep conditional feature modeling, generating consistent updates for correlated features when a single feature is changed. We benchmark our model against diverse baselines—CVAE, Causal-VAE, LMSAutoTSF, PatchTST, and MSGNet—chosen for their strengths in sequential modeling, causal inference, channel-independence, and frequency-based correlation learning, ensuring a comprehensive evaluation across paradigms.

Related Work

Long Short-Term Memory (LSTM) networks have been widely used to model temporal dependencies and perform imputation by reconstructing missing values from observed history (Cao et al. 2018). Recent works have explored Variational Autoencoders (VAEs) for time series data, integrating temporal dynamics into structured latent representations. GP-VAE (Fortuin et al. 2020) integrates Gaussian Processes to handle irregular sampling, while TIMEVAE (Desai et al. 2021) introduces interpretable components like trend-seasonal decoders. CR-VAE (Li, Yu, and Principe 2023) advances this further by learning Granger-causal graphs to guide generation, though it requires predefined sparsity constraints. These methods excel at sequence-level tasks but treat features monolithically—our work addresses this by enabling feature-wise interventions through spectral correlation learning and decomposition.

Understanding feature changes relates fundamentally to counterfactual reasoning. Deep structural causal models (Pawlowski, Coelho de Castro, and Glocker 2020) estimate treatment effects but assume known causal graphs, while Time Series Deconfounders (Bica, Alaa, and Van Der Schaar 2020) infer latent confounders for longitudinal data. Such frameworks focus on temporal cause-effect chains (e.g., "how does past treatment affect future outcomes?"), whereas CFOR-VAE targets intra-timestep correlations (e.g., "how does pressure affect temperature now?"). This distinction is critical for industrial settings where immediate cross-feature impacts must be simulated. Some recent models such as FGTI(Yang et al. 2024), RecCSTI(Lai et al. 2024), and PSW-I(Wang et al. 2025) address generative frequency modeling, irregular imputation, and temporal heterogeneity, respectively, for imputation problem but lack support for targeted feature-level interventions, per-feature dependency modeling, or explicit feature disentanglement.

Recent architectures adopt diverse modeling strategies for time series forecasting: PatchTST (Nie et al. 2023) and MSGNet (Cai et al. 2024) utilize Transformer-based attention mechanisms, TimesNet (Wu et al. 2023) leverages temporal convolutional operators, while LMSAutoTSF (Delibasoglu, Chakraborty, and Heintz 2024) is fully based on multilayer perceptrons (MLPs). PatchTST processes univariate patches independently, while LMSAutoTSF combines frequency-domain filtering with autocorrelation encoding. PatchTST (Nie et al. 2023) employs channel-independent processing by treating each feature as an isolated univariate series, thereby neglecting inter-feature correlations—a significant limitation in multivariate settings. In our study, we also utilize PatchTST as a baseline to evaluate reconstruction fidelity without accounting for feature correlations, providing a reference point for correlation-agnostic performance. In contrast, MSGNet (Cai et al. 2024) explicitly models multi-scale inter-series dependencies: it decomposes time series into frequency components, uses self-attention for intra-series patterns, and applies adaptive graph convolution to learn varying inter-feature correlations per time scale. While MSGNet excels in forecasting, its non-generative de-

sign cannot simulate feature interventions. CFOR-VAE addresses this gap by enabling both the analysis and controlled reconstruction of inter-feature dynamics. In contrast to prior approaches, CFOR-VAE explicitly captures intra-timestep feature dependencies through spectral decomposition and Fourier-based correlation modeling, facilitating interpretable local interventions without relying on predefined causal graphs (Table 1).

Methodology

We propose CFOR-VAE, a novel Conditional Feature-Oriented Reconstruction framework that extends traditional CVAEs through three key architectural innovations: (1) a temporal decomposition module that explicitly separates trend and seasonal components, (2) a spectral correlation module employing an adapted version of Fourier Neural Operators (Li et al. 2021) to model global feature dependencies in frequency space while maintaining translation equivariance; and (3) a dual-path variational inference mechanism with component-specific attention pooling, producing disentangled latent representations that preserve both temporal dynamics and cross-feature relationships. This integrated design addresses the critical challenge of maintaining interpretable feature correlations during reconstruction - a capability we strongly validate through difference-based reconstruction analyses.

Problem Formulation

Given a multivariate time series $\mathbf{X} \in \mathbb{R}^{B\times T\times N}$ with batch size B, T time steps, and N features, along with a corresponding mask $\mathbf{M} \in \{0,1\}^{B\times T\times N}$ indicating observed ($M_{i,t,n}=1$) and missing ($M_{i,t,n}=0$) values, our objective is to learn a conditional generative model $p_\theta(\mathbf{X}|\mathbf{M})$ that, 1. Accurately reconstructs the complete time series $\hat{\mathbf{X}}$ while preserving temporal dependencies and cross-feature correlations. 2. Enables controlled imputation of masked regions $\mathbf{X}_{\text{masked}}$ conditioned on observed features $\mathbf{X}_{\text{obs}}$. 3. Maintains consistency between modified and adjusted values in case of any feature is modified.

Input Projection

Given an input time series $\mathbf{X} \in \mathbb{R}^{B\times T\times N}$, where B is the batch size, T is the sequence length, and N is the number of input features, we first project the feature dimension to a latent space (L) via a linear transformation:

$$\mathbf{X}_{\text{proj}} = \text{Linear}(\mathbf{X}) \in \mathbb{R}^{B\times T\times D}.$$

Trend-Seasonal Decomposition

To explicitly disentangle structural patterns in the input, we decompose the projected input into *trend* and *seasonal* components using a temporal moving average filter:

$$\mathbf{T} = \text{AvgPool1D}(\mathbf{X}_{\text{proj}}), \quad \mathbf{S} = \mathbf{X}_{\text{proj}} - \mathbf{T},$$

where $\mathbf{T}, \mathbf{S} \in \mathbb{R}^{B\times T\times D}$ denote the trend and seasonal components respectively.

Feature-wise Spectral Modeling via 2D Fourier Neural Operator

To model long-range dependencies and structured feature interactions within the trend component $\mathbf{T} \in \mathbb{R}^{B\times T\times D}$, we propose a spectral operator module, **FNO2D**, which adapts the Fourier Neural Operator (FNO) framework (Li et al. 2020) for multivariate time series. Unlike conventional 1D FNOs, our formulation treats the projected feature axis as a structured spatial-like dimension and jointly reasons over time and feature space via 2D Fourier transforms as represented in Figure 3.

Spectral Projection. The input trend component $\mathbf{T} \in \mathbb{R}^{B\times T\times D}$ is first projected into a latent space via a linear transformation:

$$\mathbf{H}_{\text{trend}} = \mathbf{W}_{\text{in}}\mathbf{T} + \mathbf{b}_{\text{in}} \in \mathbb{R}^{B\times T\times H}.$$

To enable 2D spectral processing, we reshape $\mathbf{H}_{\text{trend}}$ into a tensor representing a 2D grid over time and features:

$$\mathbf{H}_{\text{grid}} \in \mathbb{R}^{B\times 1\times T\times H},$$

where the singleton channel dimension aligns with the structure of conventional spectral operators. We then apply a two-dimensional discrete Fourier transform (2D-DFT) over the temporal and feature axes to map the signal into the frequency domain:

$$\mathcal{F}(\mathbf{H}_{\text{grid}}) = \hat{\mathbf{H}} \in \mathbb{C}^{B\times 1\times T_f\times H_f},$$

where T_f and H_f denote the retained number of frequency modes in the temporal and feature dimensions, respectively.

A learnable complex-valued weight tensor $\mathbf{W}_{\mathbf{c}} \in \mathbb{C}^{1\times T_f\times H_f\times H}$ is used to modulate the spectral coefficients through element-wise complex multiplication:

$$\hat{\mathbf{H}}_{\text{filtered}} = \hat{\mathbf{H}} \odot \mathbf{W}_{\mathbf{c}}.$$

Finally, the filtered frequency map is transformed back into the time-feature domain using the inverse 2D Fourier transform:

$$\mathbf{H}_{\text{filtered}} = \mathcal{F}^{-1}(\hat{\mathbf{H}}_{\text{filtered}}) \in \mathbb{R}^{B\times 1\times T\times H}.$$

The singleton dimension is removed, and the result is projected back to the original feature dimension:

$$\tilde{\mathbf{T}} = \phi(\mathbf{H}_{\text{filtered}})\mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}} \in \mathbb{R}^{B\times T\times D},$$

where ϕ denotes a non-linear activation function (e.g., GELU). Our **FNO2D** module introduces a novel perspective on modeling trend dynamics by:

- Treating time and projected features as a structured 2D domain, enabling spectral reasoning across both axes.
- Learning global correlations between features and across time using low-frequency spectral filters, while maintaining computational efficiency.
- Operating exclusively on the trend component, which benefits from smooth, low-frequency modeling to focus on long-range changes.

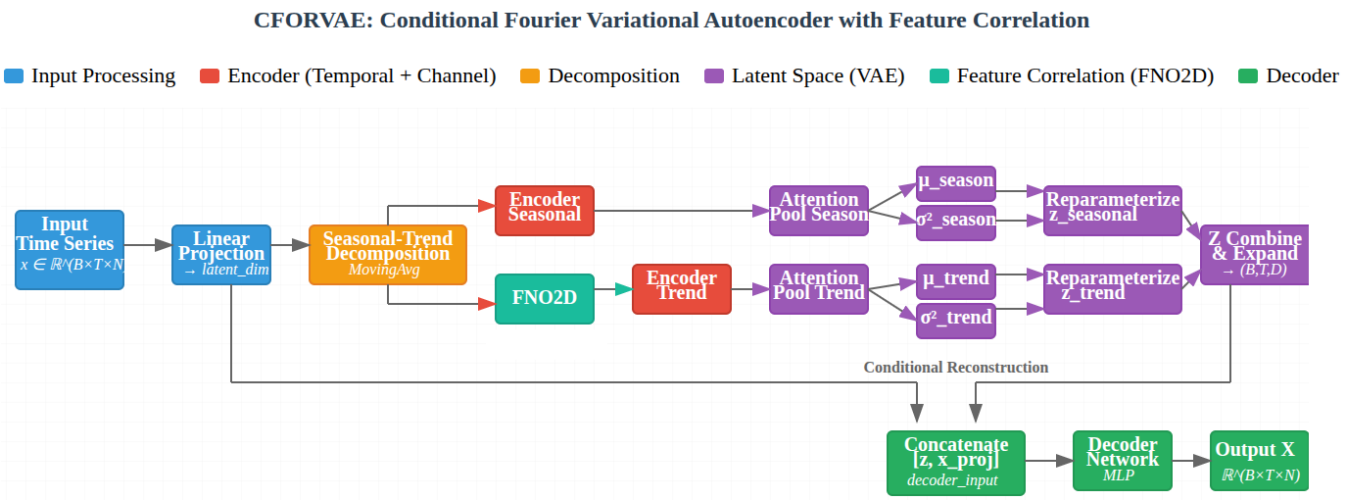

Figure 2: Proposed CFOR-VAE model workflow

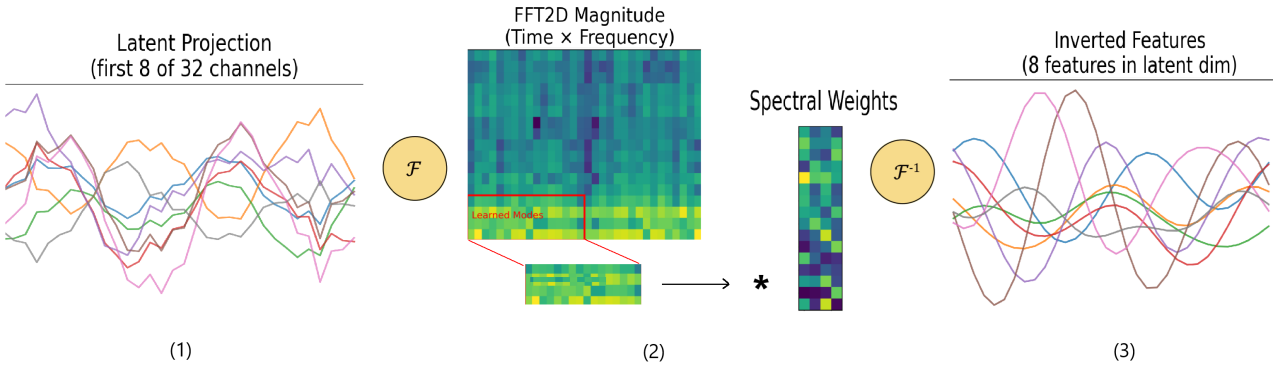

Figure 3: FNO2D Pipeline: (1) Input time series are projected to latent space (32 channels). (2) 2D FFT transforms the data to frequency domain (time × channel frequencies). Only low-frequency modes (red box, controlled by modes1×modes2) are preserved and multiplied with learned weights. (3) Inverse FFT reconstructs the output. The spectral convolution learns global filters while preserving temporal dynamics and feature correlations.

Temporal-Channel Encoding

Both $\tilde{\mathbf{T}}$ and $\mathbf{S}$ are passed through a shared dual-path encoder that sequentially captures temporal and channel-wise patterns:

$$\mathbf{H}_{\text{trend}} = \text{Encoder}(\tilde{\mathbf{T}}), \quad \mathbf{H}_{\text{seasonal}} = \text{Encoder}(\mathbf{S}).$$

Each encoder comprises stacked fully connected layers applied first over the temporal axis and then over the feature/channel axis.

Variational Latent Encoding

We pool the encoded trend and seasonal sequences via temporal attention to produce condensed latent representations:

$$\mathbf{z}_T = \text{AttnPool}(\tilde{\mathbf{H}}_{\text{trend}}), \quad \mathbf{z}_S = \text{AttnPool}(\mathbf{H}_{\text{seasonal}}).$$

Separate μ and $\log \sigma^2$ projections are computed for each:

$$\mu_T, \log \sigma_T^2 = f_T(\mathbf{z}_T), \quad \mu_S, \log \sigma_S^2 = f_S(\mathbf{z}_S).$$

Latent variables are sampled via the reparameterization trick:

$$\mathbf{z}_T \sim \mathcal{N}(\mu_T, \sigma_T^2), \quad \mathbf{z}_S \sim \mathcal{N}(\mu_S, \sigma_S^2).$$

The combined latent is formed as:

$$\mathbf{z} = \mathbf{z}_T + \mathbf{z}_S, \quad \mu = \frac{1}{2}(\mu_T + \mu_S)$$

$$\log \sigma^2 = \frac{1}{2}(\log \sigma_T^2 + \log \sigma_S^2).$$

Decoding

To encourage latent utilization and prevent posterior collapse, Gaussian noise is added to the projected input during training:

$$\mathbf{X}'_{\text{proj}} = \mathbf{X}_{\text{proj}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

The decoder receives the concatenation of the expanded latent $\mathbf{z}$ ($\mathbf{z}_{\text{expanded}}$: repeated over T) and the noisy input, and reconstructs the original sequence:

$$\hat{\mathbf{X}} = \text{Decoder}([\mathbf{z}_{\text{expanded}}, \mathbf{X}'_{\text{proj}}]).$$

where Decoder$(\cdot)$ applies Linear projection: $2L \rightarrow L$, GELU activation, Dropout ($p = 0.1$), and Final projection: $L \rightarrow D$.

Experimental Methodology

We assess each model's ability to capture feature dependencies through controlled reconstruction tests. For each input sequence $\mathbf{x} \in \mathbb{R}^{T \times N}$:

1. **Baseline Reconstruction**: Generate $\hat{\mathbf{x}}$ from the original input with no modifications
2. **Adjusted Reconstruction**:
   - Select one feature x_i to modify
   - Apply a parametric transformation (linear/exponential/logistic) with random magnitude ($0.5\times$ to $1.5\times$)
   - Generate $\hat{\mathbf{x}}^*$ from this perturbed input

We assess model performance through the reconstruction strategy: **Difference Propagation** isolates the model's change-specific behavior through a three-step process: (1) computing the difference between original and modified values for the target feature, (2) feeding only this difference vector (with other features zeroed), and (3) applying the reconstructed differences to original values. This method specifically tests the model's ability to propagate targeted changes while maintaining feature relationships.

Evaluation Metrics

We use three complementary metrics detailed in supplementary file to quantitatively assess the quality of the reconstruction for adjusted features. Distance Correlation (DC) effectively captures whether the statistical dependence structure is maintained. Normalized Cross Correlation (NCC) verifies the preservation of linear temporal relationships. Autocorrelation Error (AE) ensures the intrinsic dynamics remain consistent.

Evaluation

We evaluate our method on six diverse datasets: Beijing PM2.5 (Chen 2015), Pulp-and-Paper (Ranjan et al. 2018), ETTh1 (Zhou et al. 2021), Energy (Candanedo, Feldheim, and Deramaix 2017), AirQuality (De Vito et al. 2008), and a synthetic dataset (SYN1) with controlled correlations. Table 3 presents the mean Moving Average Correlation (MA Corr, window = 5) to assess trend similarity between original, reconstructed, and adjusted features. MSGNet assigns high similarity scores to correlated features, yet underperforms on adjusted metrics, revealing its limitations for adjusted reconstruction. Tables 4 and 5 report reconstruction fidelity using DC, NCC, and AC Error across three conditions: Modified vs Reconstructed (MvsR), Baseline, and Adjusted. Higher DC/NCC and lower AC Error indicate better performance. CFOR-VAE consistently outperforms others on Adjusted metrics across datasets demonstrating its strength in reconstructing correlated features under perturbation. All mean results are aggregated over different seed values, and detailed breakdowns are available in the supplementary material. Figures 4 and 5 illustrate CFOR-VAE's response to input modifications (e.g., DEWP), reflecting its ability to model feature-level causal dependencies in predictions.

Method	Metric	Energy	PM2.5
w/o decomposition and FNO2D	DC	0.7208	0.7036
	NCC	0.6293	0.6272
	AC Error	0.0575	0.1345
+Decomposition	DC	0.7216	0.7004
	NCC	0.6309	0.6234
	AC Error	0.0579	0.1360
+FNO2D	DC	**0.7287**	**0.7201**
	NCC	**0.6377**	**0.6439**
	AC Error	**0.0559**	**0.1295**

Table 2: Ablation Study Across Datasets (Adjusted Reconstruction Scores)

Ablation

We performed an ablation study on the Energy and PM2.5 datasets using the adjusted reconstruction metric to assess the contribution of each component in CFOR-VAE. As shown in Table 2, the complete model—combining temporal decomposition and the FNO2D-based feature correlation module—achieves the best results across all metrics, particularly in maintaining dynamic correlations (DC, NCC) and reducing adjusted correlation error (AC Error). The baseline model excludes both modules, while adding only decomposition improves results for the Energy dataset, indicating that decomposition aids in disentangling temporal patterns but remains insufficient on its own. Incorporating FNO2D—which learns inter-feature dependencies in the frequency domain—yields a performance gain over the baseline and decomposition-only variants. These results validate the effectiveness of both modules, with FNO2D providing the most significant contribution to correlation-preserving reconstruction.

Conclusion

While modern models like LMSAutoTSF and MSGNet excel at capturing temporal patterns, they often fail to preserve the critical correlations between features when reconstructing or manipulating data. This limits their use for realistic simulations and "what-if" analysis. VAEs offer a better foundation for this, as their structured latent space can inherently model joint feature dependencies.

We build on this strength by introducing CFOR-VAE, a novel framework that integrates Fourier-based spectral encoding into a VAE. This approach explicitly separates and models the trend and seasonal components of time-series data. The result is a model that not only captures cyclical temporal patterns but also maintains the essential correlations across different features during reconstruction and intervention.

CFOR-VAE bridges a key gap: it enables both high-fidelity reconstruction and interpretable, controlled manip-

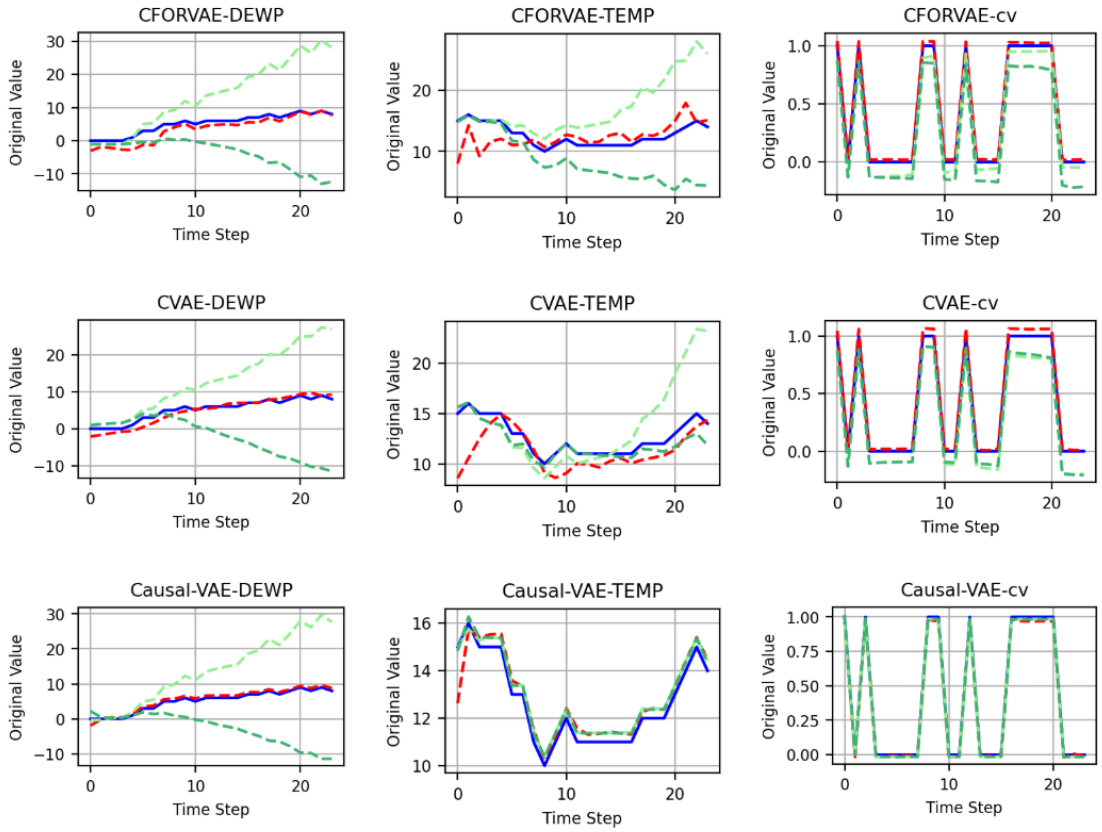

Figure 4: Feature-Level Response to Targeted Interventions on the PM2.5 Dataset. The second column shows a feature that is correlated with the adjusted one (first column), where changes are expected to propagate. The third column displays an uncorrelated feature, where little to no response is expected. Red plot indicates reconstructed feature, blue is original and green ones are adjustments applied on original features.

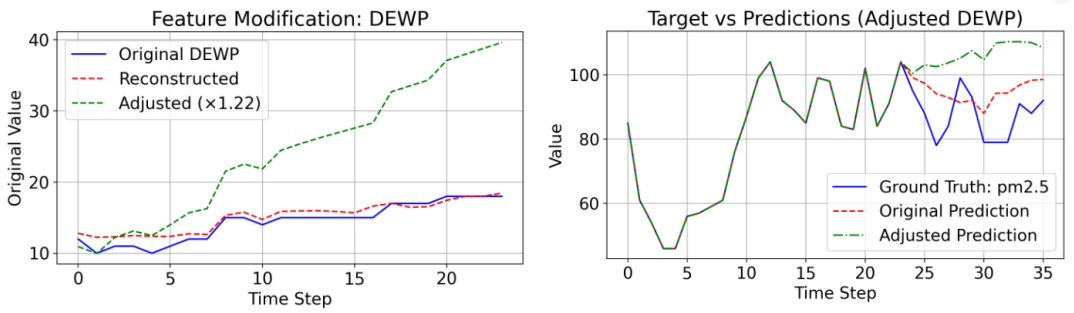

Figure 5: Effect of modifying a input feature DEWP on the model's prediction. "Adjusted Prediction" illustrates how the target output responds to the modification, highlighting the model's sensitivity to feature-level changes.

Dataset	CVAE			CausalVAE			MSGNet			**CFORVAE**		
	Orig.	Base.	Adj.	Orig.	Base.	Adj.	Orig.	Base.	Adj.	Orig.	Base.	Adj.
syn1	0.7659	0.9231	0.8670	0.7551	0.9138	0.7419	0.7504	1.000	0.7358	0.7522	0.9009	**0.8889**
pm25	0.5530	0.6003	0.6266	0.5500	0.5871	0.5948	0.5542	1.000	0.5903	0.5558	0.6093	**0.7194**
pulp	0.4148	0.6141	0.7255	0.4080	0.5219	0.7013	0.4092	1.000	0.5506	0.4070	0.6035	**0.7756**
air	0.6657	0.6976	0.7157	0.6623	0.6810	0.7019	0.6591	1.000	0.6973	0.6597	0.7586	**0.7432**
etth1	0.9740	0.9777	0.9707	0.9744	0.9775	0.942	0.9750	1.000	0.9663	0.9749	0.9793	**0.9737**
energy	0.5981	0.6076	0.6335	0.5957	0.5984	0.6157	0.5978	1.000	0.5842	0.5948	0.6935	**0.6473**

Table 3: Mean Moving Average Correlation scores across models, measuring trend similarity between original, reconstructed (Baseline) and adjusted features. Red (bold) indicates the best adjusted values, while blue represents the second best values.

Model	Metric	SYN1			PM2.5			PULP		
		M*vs*R	Baseline	Adj.	M*vs*R	Baseline	Adj.	M*vs*R	Baseline	Adj.
CVAE	DC	0.9387	0.8705	0.7024	0.8048	0.5810	0.6372	0.8099	0.6345	0.7239
	NCC	0.9452	0.8841	0.7145	0.7550	0.4847	0.5261	0.7539	0.5452	0.6417
	AC Error	0.0939	0.0845	0.1048	0.1482	0.1659	0.1645	0.1522	0.1622	0.1701
Causal-VAE	DC	0.9478	0.8279	0.5886	0.8059	0.5508	0.5853	0.7864	0.5657	0.6851
	NCC	0.9518	0.8486	0.5956	0.7787	0.4473	0.4806	0.7441	0.4461	0.5920
	AC Error	0.0779	0.0931	0.2046	0.1282	0.1757	0.1791	0.1655	0.2213	0.2224
LMSAutoTSF	DC	0.9140	0.7426	0.4783	0.9141	0.6117	0.5823	0.8740	0.5262	0.5627
	NCC	0.9179	0.7623	0.4621	0.8987	0.5217	0.4869	0.8498	0.4178	0.4584
	AC Error	0.1122	0.1114	0.3066	0.0594	0.1502	0.1919	0.1129	0.2855	0.3323
PatchTST	DC	0.9221	0.6442	0.4633	0.9168	0.5874	0.5698	0.8728	0.5130	0.5489
	NCC	0.9306	0.6565	0.4499	0.9038	0.5075	0.4758	0.8501	0.4009	0.4458
	AC Error	0.1094	0.1178	0.3218	0.0583	0.1640	0.1927	0.1098	0.3064	0.3387
MSGNet	DC	0.9739	1.0000	0.5687	0.9026	1.0000	0.5637	0.7341	1.0000	0.4975
	NCC	0.9787	1.0000	0.5825	0.8838	1.0000	0.4659	0.6869	1.0000	0.3762
	AC Error	0.0466	0.0000	0.1161	0.0598	0.0000	0.1739	0.2003	0.0000	0.2727
CFOR-VAE	DC	0.9233	0.8393	**0.7418**	0.9147	0.5837	**0.7198**	0.8956	0.5879	**0.7574**
	NCC	0.9269	0.8588	**0.7538**	0.8977	0.4892	**0.6444**	0.8724	0.5038	**0.7007**
	AC Error	0.0991	0.0591	**0.1004**	0.0520	0.1807	**0.1292**	0.0843	0.1741	**0.1483**

Table 4: *Reconstruction Fidelity on Correlated Features across two runs. "Modified vs Reconstructed" (MvsR)* compares the intended (manually perturbed) input feature to its reconstruction by the model. *"Baseline"* denotes reconstruction from original inputs. *"Adj."* refers to the reconstruction of other correlated features conditioned on the modified input. Red indicates the best adjusted values, while blue represents the second best values.(Correlation Threshold=0.5)

Model	Metric	Air			ETTh1			Energy		
		M*vs*R	Baseline	Adj.	M*vs*R	Baseline	Adjç	M*vs*R	Baseline	Adj.
CVAE	DC	0.8836	0.7129	0.7383	0.9603	0.9656	0.9451	0.9197	0.6740	0.7143
	NCC	0.8540	0.6242	0.6527	0.9645	0.9725	0.9552	0.9005	0.5851	0.6177
	AC Error	0.0961	0.1513	0.1479	0.0331	0.0233	0.0282	0.0284	0.0782	0.0637
Causal-VAE	DC	0.8682	0.7123	0.7259	0.9786	0.9561	0.9237	0.9089	0.6823	0.7030
	NCC	0.8475	0.6031	0.6357	0.9793	0.9642	0.9288	0.8840	0.5762	0.5996
	AC Error	0.0906	0.1630	0.1549	0.0201	0.0245	0.0380	0.0272	0.0883	0.0628
LMSAutoTSF	DC	0.9501	0.7610	0.7057	0.9729	0.9812	0.9153	0.9109	0.6992	0.7063
	NCC	0.9420	0.6896	0.6174	0.9764	0.9844	0.9223	0.8874	0.6067	0.6070
	AC Error	0.0429	0.1463	0.1724	0.0315	0.0199	0.0441	0.0269	0.0494	0.0604
PatchTST	DC	0.9339	0.7440	0.6737	0.9791	0.9625	0.9201	0.9196	0.6870	0.7035
	NCC	0.9228	0.6694	0.5761	0.9817	0.9681	0.9264	0.9000	0.5914	0.6027
	AC Error	0.0468	0.1371	0.1899	0.0237	0.0242	0.0395	0.0258	0.0505	0.0601
MSGNet	DC	0.9487	1.0000	0.7099	0.9945	1.0000	0.9434	0.9216	1.0000	0.6505
	NCC	0.9433	1.0000	0.6261	0.9948	1.0000	0.9521	0.9168	1.0000	0.5591
	AC Error	0.0438	0.0000	0.1409	0.0115	0.0000	0.0293	0.0447	0.0000	0.0971
CFOR-VAE	DC	0.9595	0.7673	**0.7635**	0.9745	0.9707	**0.9498**	0.9391	0.7414	**0.7260**
	NCC	0.9517	0.7057	**0.6909**	0.9746	0.9774	**0.9586**	0.9253	0.6779	**0.6345**
	AC Error	0.0296	0.1318	**0.1292**	0.0232	0.0220	**0.0266**	0.0215	0.0641	**0.0564**

Table 5: *Reconstruction Fidelity on Correlated Features across two runs. "Modified vs Reconstructed" (MvsR)* compares the intended (manually perturbed) input feature to its reconstruction by the model. *"Baseline"* denotes reconstruction from unmodified inputs. *"Adj."* refers to the reconstruction of other correlated features conditioned on the modified input. Red indicates the best adjusted values, while blue represents the second best values.(Correlation Threshold=0.5)

ulation of inputs. This creates a powerful practical tool for industrial decision-makers, who can interactively simulate the effects of potential interventions—adjusting one feature while realistically observing its correlated impact on others—all while preserving the underlying statistical integrity of the system. By providing these correlation-aware feature updates, CFOR-VAE advances interpretable temporal modeling for applications like process optimization, financial scenario testing, predictive maintenance, and supply chain demand forecasting.

Acknowledgements

This work is partially supported by the 'Resurssmarta Processor (RSP)', the Wallenberg AI, Autonomous Systems and Software Program (WASP), and the Wallenberg Initiative Materials Science for Sustainability (WISE), all funded by the Knut and Alice Wallenberg Foundation. We would also like to thank Vinnova and BioInnovation program.

References

Bica, I.; Alaa, A.; and Van Der Schaar, M. 2020. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International conference on machine learning*, 884–895. PMLR.

Cai, W.; Liang, Y.; Liu, X.; Feng, J.; and Wu, Y. 2024. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 11141–11149.

Candanedo, L. M.; Feldheim, V.; and Deramaix, D. 2017. Data driven prediction models of energy use of appliances in a low-energy house. *Energy and buildings*, 140: 81–97.

Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; and Li, Y. 2018. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31.

Chen, S. 2015. Beijing PM2.5. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5JS49.

De Vito, S.; Massera, E.; Piga, M.; Martinotto, L.; and Di Francia, G. 2008. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2): 750–757.

Delibasoglu, I.; Chakraborty, S.; and Heintz, F. 2024. LMS-AutoTSF: Learnable Multi-Scale Decomposition and Integrated Autocorrelation for Time Series Forecasting. *arXiv preprint arXiv:2412.06866*.

Desai, A.; Freeman, C.; Wang, Z.; and Beaver, I. 2021. Timevae: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*.

Fortuin, V.; Baranchuk, D.; Rätsch, G.; and Mandt, S. 2020. Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, 1651–1661. PMLR.

Lai, Z.; Zhang, D.; Li, H.; Zhang, D.; Lu, H.; and Jensen, C. S. 2024. ReCTSi: resource-efficient correlated time series imputation via decoupled pattern learning and completeness-aware attentions. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1474–1483.

Li, H.; Yu, S.; and Principe, J. 2023. Causal recurrent variational autoencoder for medical time series generation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 8562–8570.

Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; and Anandkumar, A. 2020. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

Li, Z.; Kovachki, N. B.; Azizzadenesheli, K.; Bhattacharya, K.; Stuart, A.; Anandkumar, A.; et al. 2021. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations*.

Locatello, F.; Bauer, S.; Lucic, M.; Raetsch, G.; Gelly, S.; Schölkopf, B.; and Bachem, O. 2019. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, 4114–4124. PMLR.

Mehdiyev, N.; Majlatow, M.; and Fettke, P. 2024. Counterfactual explanations in the big picture: An approach for process prediction-driven job-shop scheduling optimization. *Cognitive Computation*, 16(5): 2674–2700.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Pawlowski, N.; Coelho de Castro, D.; and Glocker, B. 2020. Deep structural causal models for tractable counterfactual inference. *Advances in neural information processing systems*, 33: 857–869.

Ranjan, C.; Reddy, M.; Mustonen, M.; Paynabar, K.; and Pourak, K. 2018. Dataset: rare event classification in multivariate time series. *arXiv preprint arXiv:1809.10717*.

Sohn, K.; Lee, H.; and Yan, X. 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28.

Stepin, I.; Alonso, J. M.; Catala, A.; and Pereira-Fariña, M. 2021. A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence. *Ieee Access*, 9: 11974–12001.

Verma, S.; Boonsanong, V.; Hoang, M.; Hines, K.; Dickerson, J.; and Shah, C. 2024. Counterfactual explanations and algorithmic recourses for machine learning: A review. *ACM Computing Surveys*, 56(12): 1–42.

Wang, H.; Li, H.; Chen, X.; Gong, M.; Chen, Z.; et al. 2025. Optimal transport for time series imputation. In *The Thirteenth International Conference on Learning Representations*.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *The Eleventh International Conference on Learning Representations*.

Yang, X.; Sun, Y.; Chen, X.; et al. 2024. Frequency-aware generative models for multivariate time series imputation. *Advances in Neural Information Processing Systems*, 37: 52595–52623.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.