Causal Representation Learning for Time-Series Forecasting

universität freiburg

Salih Bora Öztürk [1] Hanne Raum [1]

[1]University of Freiburg

TL;DR

CReP: Causal-oriented Representation Learning Predictor

is a self-supervised framework that jointly performs multi-step time-series forecasting and causal analysis by decomposing high-dimensional observations into orthogonal cause, effect, and non-causal representations for a chosen target variable.

Introduction

Importance of the paper

- Enables **simultaneous forecasting and causal analysis** in high-dimensional time-series systems.
- Provides **interpretable causal insights** (causes vs. effects) rather than black-box predictions.

Problem

- Existing approaches often forecast well but remain non-causal, or infer causality but are separate from forecasting.

Objectives

- Given observations X_t and a chosen target $y = x_k$:
  - **Forecast** y for multiple future steps in one shot.
  - **Identify** which variables likely **cause** y and which are likely **effects** of y.
- Achieve this **self-supervised**.

Method

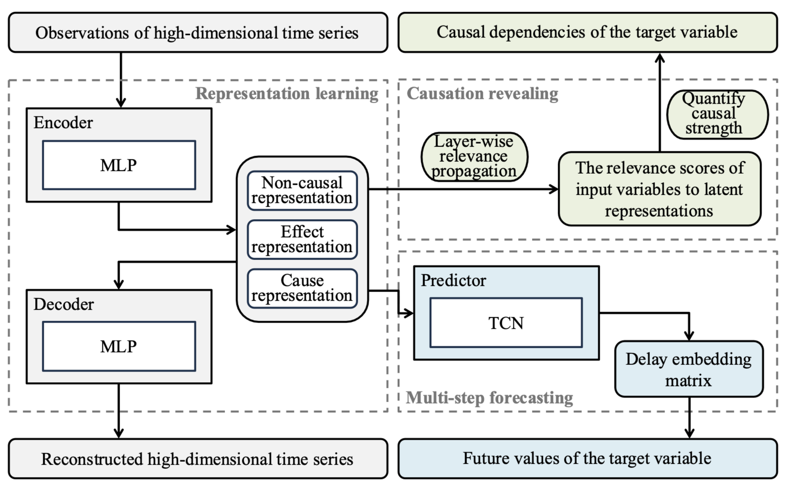

Figure 1. The flowchart of CReP method.

The framework is characterized by three key features:

1. Dynamic causation detection with the STI transformation mechanism (Fig. 2a).
2. Causal-oriented representation learning for multi-step predictions through the CReP (Fig. 2b).
3. Causal analysis of the target variable via $\alpha\beta$-LRP (Fig. 2c).

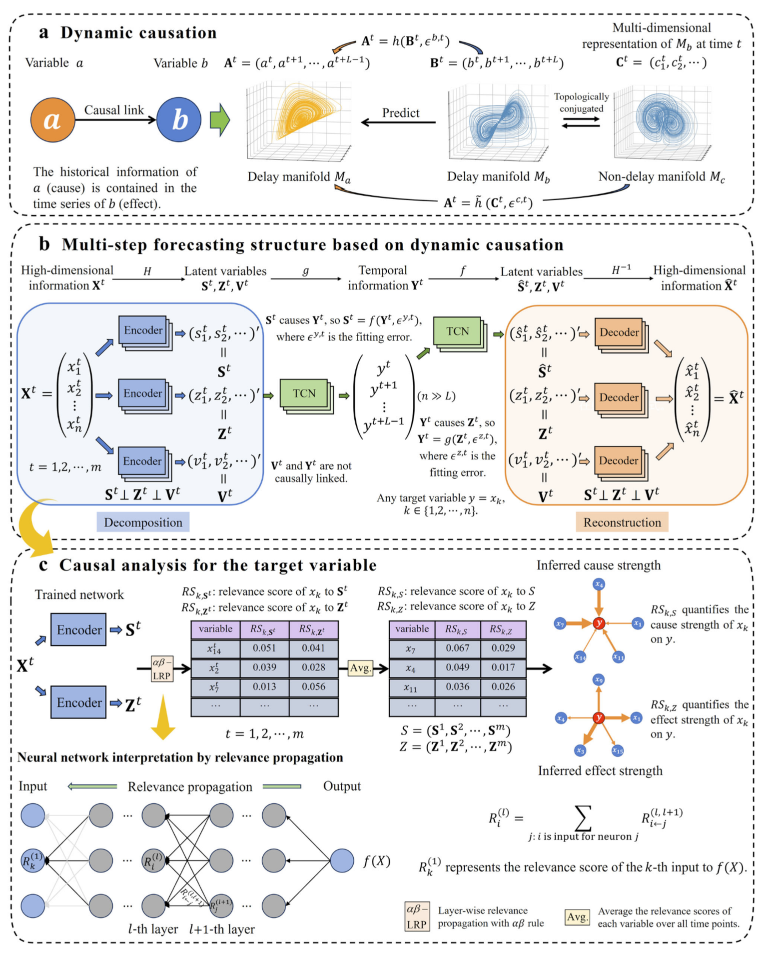

Figure 2. Schematic illustration of the CReP framework.

Background Theory

Dynamic causation (reconstruction principle)

If variable a drives b in a dynamical system, then information about a is embedded in the time series of b. Using delay embeddings there exists an implicit mapping (Takens-style):

$$A_t = (a_t, \dots, a_{t+L-1})^\top, \quad B_t = (b_t, \dots, b_{t+L})^\top, \quad A_t \approx h(B_t)$$

STI transformation (delay $\leftrightarrow$ non-delay)

Spatiotemporal information transformation allows a learned non-delay embedding C_t to be topologically conjugate to B_t.

$$A_t \approx h(C_t),$$

This motivates learning a latent Z_t that plays the role of a non-delay embedding from which Y_t can be predicted:

$$Y_t \approx g(Z_t),$$

Results

- Simulation systems:
  - Lorenz 96 (50x60 - 15) (Fig. 3 left)
  - Kuramoto power grid (30x120 - 9) (Fig. 3 middle)
  - Gene regulatory network (Dream4) (40x50 - 8) (Fig. 3 right)
- Real datasets:
  - Hong Kong cardiovascular inpatients (70x14 - 25) (Fig. 4)
  - Japan COVID-19 transmission (30x47 - 14) (Fig. 4)
- Metrics:
  - RMSE (lower is better)
  - PCC (higher is better)

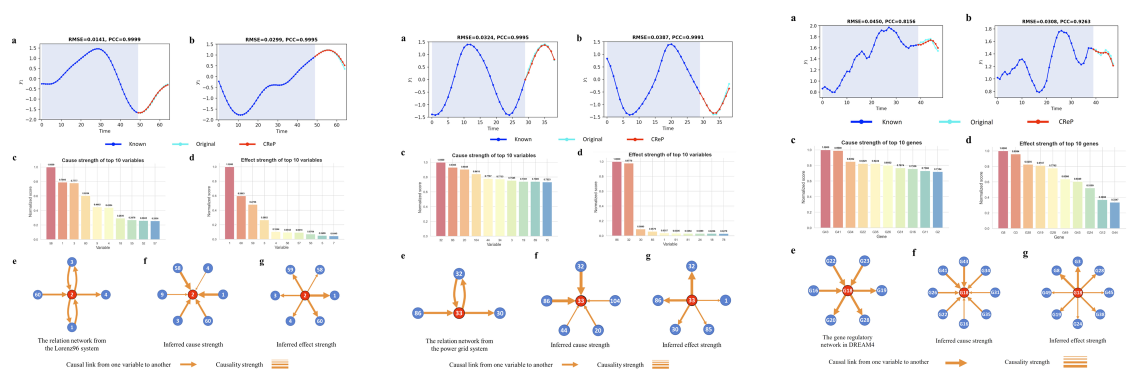

Figure 3. Performance of the CReP on the three simulation models.

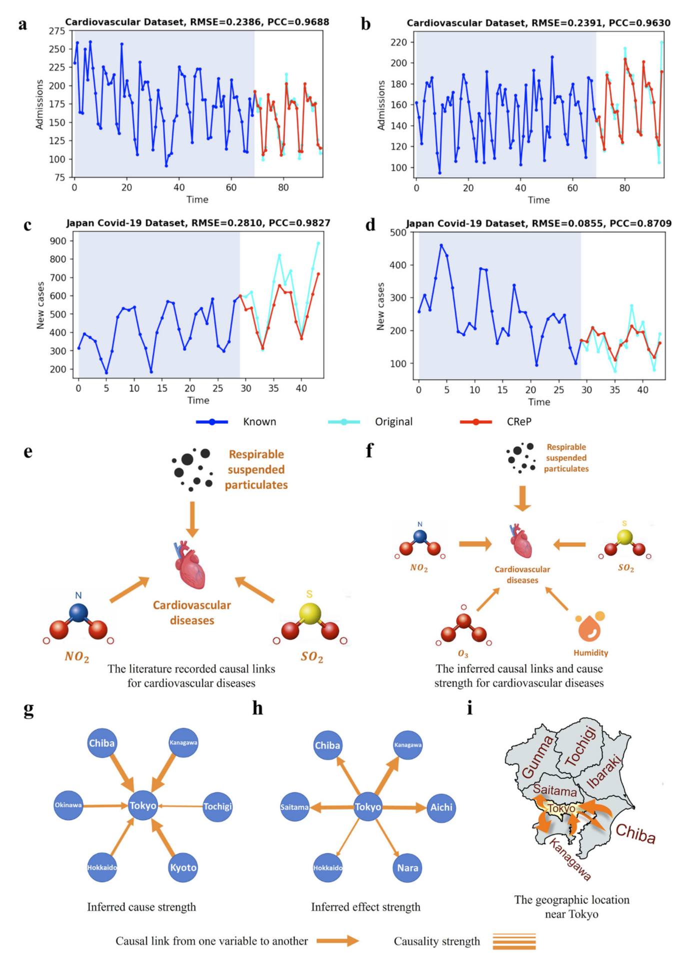

Figure 4. Performance of the CReP on the two real-world datasets.

	Lorenz 96		Power grid		Dream4	
Method	RMSE	PCC	RMSE	PCC	RMSE	PCC
CReP	**0.1086**	**0.9908**	**0.1127**	**0.9868**	**0.0984**	0.8894
ARNN	0.2328	0.8206	0.1501	0.9784	0.2281	0.8167
STICM	0.1822	0.9752	0.1530	0.9854	0.1080	**0.9928**
LSTM	0.4883	0.8303	0.3916	0.9659	0.8063	0.0615
ARIMA	0.4204	0.6156	0.2558	0.9583	0.2650	0.1218
SVR	0.8311	0.1618	0.4856	0.9540	0.6779	0.0566
Informer	0.1543	0.9005	0.1163	0.9837	0.4716	0.0489

Table 1. Comparison to other forecasting methods.

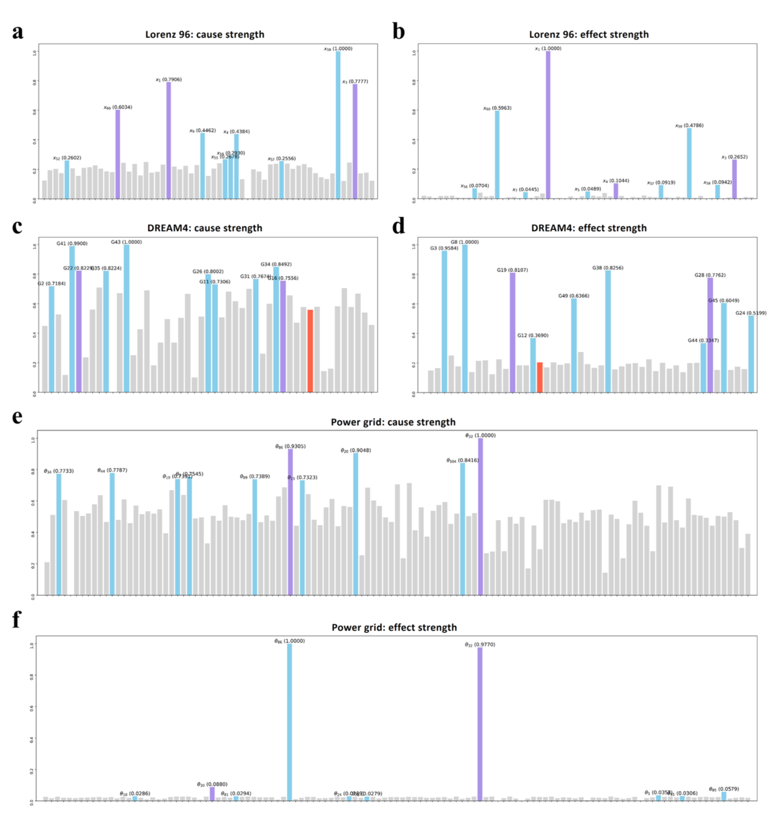

Figure 5. The normalized causal results of CReP on simulation datasets. Blue bars denote the top 10 variables by causal strength, purple bars highlight true causes or effects among them, and red bars indicate true causes or effects that were missed.

Critical discussion.

While CReP correctly ranks many true causes and effects among the top variables, Fig. 5 also reveals **false positives** and **missed true causes**. Moreover, the forecasting comparison in Table 1 is limited to relatively **basic baselines** (e.g., ARIMA, LSTM, SVR) and does not include more recent state-of-the-art foundation time-series models. In addition, history length choice in datasets, prediction length, and hyperparameter choices differ between datasets, raising questions about the fairness of the evaluations. A more controlled evaluation with matched forecasting horizons and stronger baselines would be required to fully assess CReP's performance.

Loss

$$\mathcal{L} = \lambda_1 \mathcal{L}_{DS} + \lambda_2 \mathcal{L}_{FC} + \lambda_3 \mathcal{L}_{REC} + \lambda_4 \mathcal{L}_{ORTH}$$

- Determined-State Loss: RMSE on known historical y.
- Future-Consistency Loss: RMSE between overlapping future estimates.
- Reconstruction Loss: assesses information recovery
  - $\mathcal{L}_{REC\_X}$: spatiotemporal information X from (S, Z, V)
  - $\mathcal{L}_{REC\_S}$: latent cause representation S recovered by Y
- Orthogonality Loss: enforces orthogonality among S, Z, and V.

Ablation Study

Full loss achieves best overall performance; removing any term reduces causal identification reliability (accuracy/F1/recall), and often worsens forecasting RMSE.

Supplementary Details

CReP uses an autoencoder to decompose inputs into cause, effect, and non-causal representations, and TCNs to perform stable multi-step forecasting (Fig. 6). The learned abstract representations are interpreted using $\alpha\beta$-LRP to map causal relevance back to the original variables and identify causes and effects (Fig. 7).

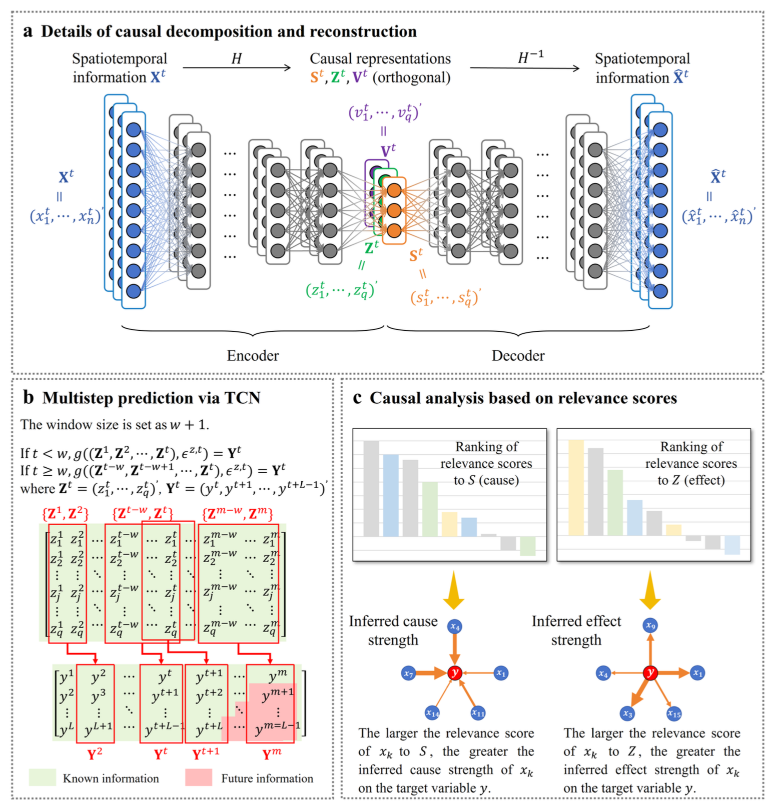

Figure 6. Implementation details of CReP framework.

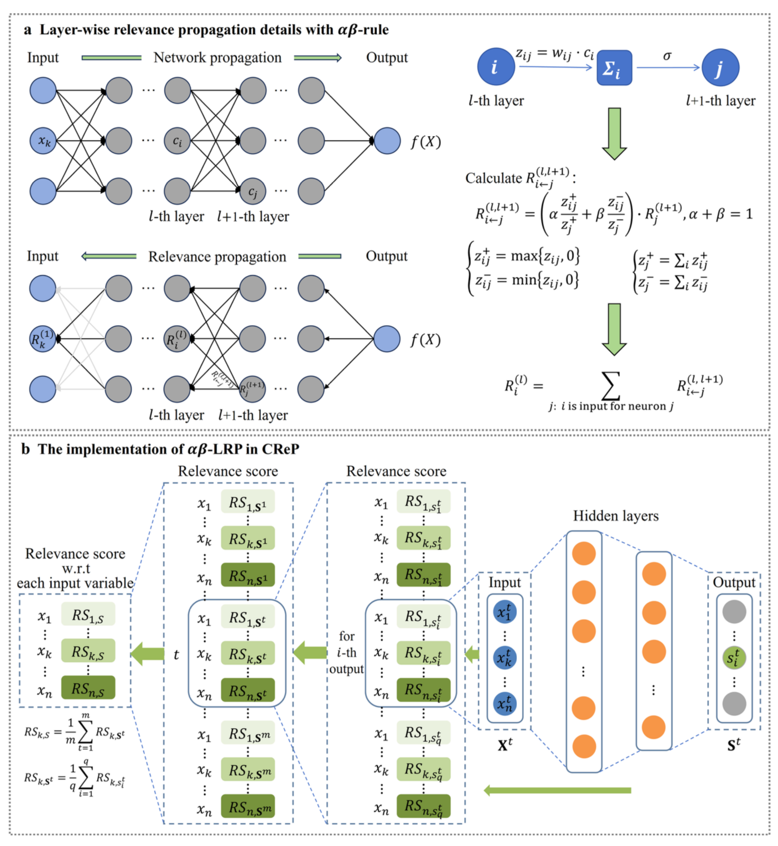

Figure 7. The causal analysis process through interpretation method $\alpha\beta$-LRP.

Conclusion

Key Points

- Causes leave **detectable traces** in the dynamics of what they influence.
- CReP learns to **extract these traces** from high-dimensional time series.
- These representations are used to **predict the future** and **identify causes and effects**.

Future Work

- Reducing false positives in causal discovery.
- Integrating different causal learning methods to enhance applicability across diverse domains.
- Exploring causal detection methods through active intervention rather than mere passive observations.

References

[1] Sihua Cai, Hao Peng, Rui Liu, and Pei Chen. Causal-oriented representation learning for time-series forecasting based on the spatiotemporal information transformation. *Communications Physics*, 8(1):242, 2025.