

<https://doi.org/10.1038/s42005-025-02170-6>

Causal-oriented representation learning for time-series forecasting based on the spatiotemporal information transformation

Sihua Cai^{1,3} , Hao Peng^{1,2,3} , Rui Liu¹ & Pei Chen¹

In real-world high-dimensional systems, both causal dependencies and temporal information of key variables are essential for dissecting the underlying mechanisms governing system dynamics. However, effective approaches to synthesize these two interconnected aspects for deeper insights remain lacking. Here we show a neural network framework, the Causal-oriented Representation Learning Predictor (CReP), which jointly conducts causal analysis and multistep forecasting from a unified perspective. CReP implicitly learns latent causal representations from observed data while simultaneously making multistep predictions, and explicitly interprets the representations to uncover the causes and effects of target variable. The core idea of CReP is to decompose the original space into three orthogonal latent factors, each capturing distinct causal representations: cause-related, effect-related, and non-causal representations of the target variable. The reconstruction-based dynamic causation, generalized through spatiotemporal information (STI) transformation mechanism, provides a theoretical foundation for simultaneously modeling causal interactions via latent representations and predicting future states using the effect representation. Evaluations on three simulation models and two real-world datasets demonstrate CReP's robust forecasting accuracy and reliable causal insights. As a self-supervised-learning approach, CReP shows significant potential for practical applications and provides a unified framework to reveal intrinsic mechanisms in dynamical systems by integrating causal relationships and temporal information.

In a real-world high-dimensional dynamic system, some variables or components are crucial for understanding the system's behavior^{1,2}. It is essential not only to predict their future dynamics but also to uncover the causal relationships of these variables within the system³. However, causal interaction information and temporal information are not independent or defined by a unidirectional dependency; rather, they are intricately interconnected elements that collectively contribute to a deeper understanding of the underlying mechanisms governing system dynamics—an aspect that has not been fully explored in most traditional research^{4–6}. Temporal information provides the foundational materials and basis for discovering causal interactions⁷, while causal information can in turn facilitate the reliable prediction of future temporal state⁸. Meanwhile, most real-world systems are too complex to be accurately described by an explicit model. Therefore, it is imperative to develop a data-driven method aimed at

conducting causal analysis and making future predictions simultaneously from a holistic perspective, addressing a persistent challenge in the fields of data science and deep learning.

Generally, time series forecasting can be broadly categorized into statistical methods and machine learning approaches. Classical statistical models (e.g., autoregressive integrated moving average (ARIMA)⁹, vector autoregression (VAR)¹⁰) are effective but struggle with non-linearity and become computationally inefficient for high-dimensional or multivariate data¹¹. In contrast, machine learning techniques excel at capturing complex patterns and managing multivariate inputs, with representative models including recurrent neural networks (RNNs)¹², gated recurrent units (GRUs)¹³, long short-term memory networks (LSTMs)^{14,15}, long- and short-term time-series networks (LSTNets)¹⁶, neural ordinary differential equations (NODEs)¹⁷, and transformer-

¹School of Mathematics, South China University of Technology, Guangzhou, 510640, China. ²School of Future Technology, South China University of Technology, Guangzhou, 510640, China. ³These authors contributed equally: Sihua Cai, Hao Peng. ✉ e-mail: scliurui@scut.edu.cn; chenpei@scut.edu.cn

based architectures (e.g., LogTrans¹⁸, Informer¹⁹, Reformer²⁰, and iTransformer²¹). These forecasting methods inherently involve representation learning, as they rely on extracting high-level patterns and representations from raw time series data for accurate prediction. However, despite their predictive success, they often fail to fully exploit causal relationships within high-dimensional systems, constraining their generalization under distributional changes and diminishing their ability to elucidate the underlying causal interactions.

To better explore the intrinsic mechanisms of complex system, many studies focus on identifying causal relationships in dynamical systems. Based on the type of measurement data, these methods can be primarily categorized into statistical approaches for cross-sectional data and dynamical methods for time-series data¹¹. While statistical methods, such as the potential outcome model (POM)²² and structural causal model (SCM)^{23–25}, are effective for time-independent or intervention-based data, they struggle with time-series data, feedback loops, and Markov equivalence issues^{11,26–29}. To address these challenges, dynamical methods such as Granger causality (GC)³⁰, transfer entropy (TE), and reconstruction-based technique³¹, leverage time-series data to uncover causal relationships. However, GC is limited by its linear model assumption, and although TE extends GC to nonlinear dynamics, it still struggles with non-separability issues. In contrast, the reconstruction-based technique, derived from delay embedding theorem^{32,33}, has been used to develop the state space reconstruction and widely applied in nonlinear time series causal analysis, including convergent cross mapping (CCM)³⁴, cross map evaluation (CME)³⁵, and cross map smoothness (CMS)³⁶. However, these methods fail to effectively identify and orthogonally separate the causes and effects for key variables in systems and are often treated as an independent module from temporal information prediction, which hinders a comprehensive understanding of the dynamics of complex systems.

To fill the gaps, we introduce a self-supervised approach, i.e., causal-oriented representation learning predictor (CReP), to make multistep-ahead forecasting and conduct causal analysis for any target variable in a high-dimensional system. This framework aims to implicitly learn latent causal representations from observed data while making multistep predictions simultaneously, and explicitly interpret the representations for uncovering the causes and effects for target variable. The core idea of CReP is to decompose the original space into three latent orthogonal factors with causal representations for the target variable: cause representation, effect representation, and non-causal representation. Specifically, the effect representation is then used to reliably predict the future states of the target variable, which is ensured by our reconstruction-based dynamic causation extended through the spatiotemporal information (STI) transformation^{8,26,37,38}.

Our framework is characterized by three key features: 1) dynamic causation detection with the STI transformation mechanism (Fig. 1a), 2) causal-oriented representation learning for multistep predictions through the CReP (Fig. 1b), and 3) causal analysis of the target variable via $\alpha\beta$ -LRP^{39,40} (Fig. 1c). To better puzzle everything together, the flowchart of our framework can be found in Supplementary Fig. 1. Note that the learned latent representations can be elucidated by utilizing the $\alpha\beta$ -LRP interpretation method, mapping them back to the original space variables to uncover the causes and effects of the target variable. From a theoretical explanatory perspective, the generalized dynamic causation grounded in delay embedding theorem^{32,33,37} provides a theoretical foundation for understanding causal interactions and time series forecasting, while deep learning serves as the backbone for computationally implementing space decomposition and future predictions within a nonlinear system. Unlike the recently proposed spatiotemporal information conversion machine (STICM)⁸, which relies on Granger causality (GC), CReP is grounded in dynamic causation theory, inherently accommodating nonlinearity and non-separability. This theoretical advantage enables CReP to effectively handle a broader range of systems beyond those addressable by GC^{26,34}. Moreover, CReP simultaneously identifies both causes and effects of the

target variable within a single run, thus achieving higher computational efficiency than STICM.

In this study, CReP is designed as a refined approach to the long-standing problem of comprehensively capturing system dynamics by integrating causal information mining and temporal information prediction through causal-oriented representation learning. To evaluate the performance of the CReP framework, it was applied on three representative models, i.e., a 60-dimensional Lorenz 96 system⁴¹, a 120-dimensional Kuramoto system⁴² and a 50-dimensional biological system. Additionally, we tested the CReP on two real-world datasets: predicting the daily number of cardiovascular inpatients in major hospitals in Hong Kong^{43,44}, and forecasting the spread of COVID-19 in Japan⁴⁵. The results demonstrate its effectiveness as a powerful tool for analyzing real causal networks and making reliable predictions for key variables based solely on observed data, indicating its potential for practical applications across various fields.

Results

Dynamic causation based on delay embedding scheme

According to dynamical systems theory, a necessary condition for two variables (e.g., a and b) to be causally linked is that they belong to the same dynamic system, meaning that they share a common attractor manifold⁴⁶. Specifically, when the variable a acts as an environmental driver of the variable b , the historical information from a is contained within the time series of b , making it possible for the temporal reconstruction of a using b . To illustrate, using the delay embedding scheme, we transform the original state space to the delay embedding space, represented as $\mathbf{A}^t = (a^t, a^{t+1}, \dots, a^{t+L-1})' \in \mathbb{R}^L$ and $\mathbf{B}^t = (b^t, b^{t+1}, \dots, b^{t+L})' \in \mathbb{R}^{L+1}$, where L is the embedding dimension and the symbol “ $'$ ” is the transpose of a vector. Based on the assumption of steady dynamics, $((\mathbf{A}^t)', (\mathbf{B}^t)')$ is an augmented $2L + 1$ dimensional vector on an attractive manifold with box-counting dimension d . Clearly, \mathbf{B}^t contains both its own temporal information and that of a , whereas \mathbf{A}^t contains only its own dynamic information. Therefore \mathbf{A}^t can be predicted reliably using \mathbf{B}^t , but not vice versa^{47–50}. Specifically, there exists an implicit function with the approximation error $\epsilon^{b,t}$:

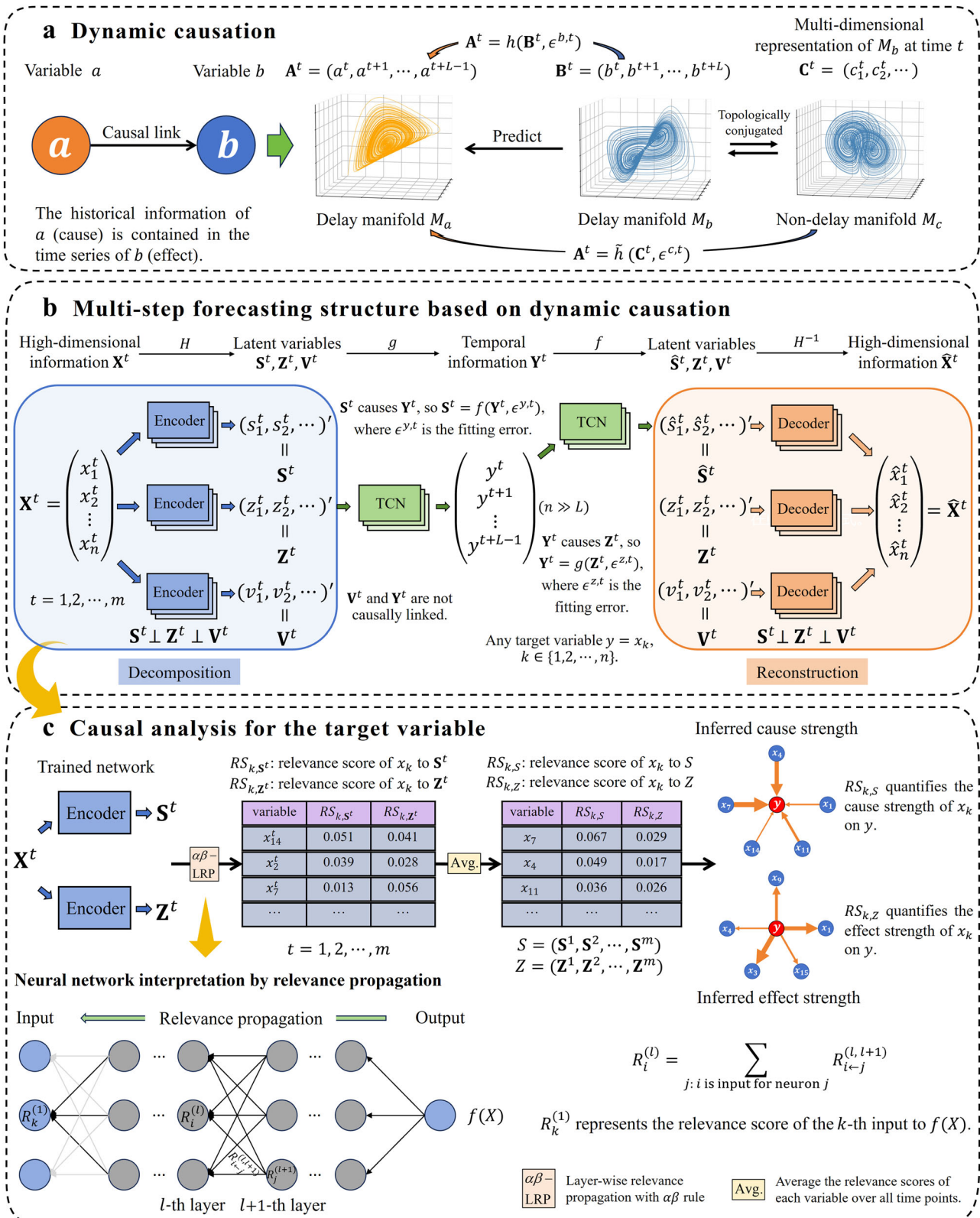
$$\mathbf{A}^t = h(\mathbf{B}^t, \epsilon^{b,t}), \quad (1)$$

where h sufficiently exists when $L > 2d$, and h is the embedding map from the delay manifold $M_b = \{\mathbf{B}^t | t \in \mathbb{R}\}$ to $M_a = \{\mathbf{A}^t | t \in \mathbb{R}\}$ according to Takens' embedding theorem and its stochastic version^{11,26,32,33,51,52}.

More generally, the mapping h in the causal equation Eq. (1) can be extended to map a non-delay manifold to a delay manifold. The generalized embedding theorem³⁷, which expands Takens' embedding theorem³², addresses more generic scenarios. According to the theory, the delay embedding manifold of the response variable b can be topologically conjugated to a non-delay embedding manifold⁵³, which is summarized as the spatiotemporal information (STI) transformation⁸. Let $\mathbf{C}^t = (c_1^t, c_2^t, \dots, c_q^t)$ be the non-delay embedding vector that is topologically conjugated to the delay embedding vector \mathbf{B}^t , where $q > 2d$ and d is the box-counting dimension of M_b . When variable a causes variable b , \mathbf{A}^t can be reliably predicted using \mathbf{B}^t and we can also obtain the implicit function:

$$\mathbf{A}^t = \tilde{h}(\mathbf{C}^t, \epsilon^{c,t}), \quad (2)$$

where $\epsilon^{c,t}$ is the approximation error, and \tilde{h} maps from the non-delay manifold $M_c = \{\mathbf{C}^t | t \in \mathbb{R}\}$ to the delay manifold $M_a = \{\mathbf{A}^t | t \in \mathbb{R}\}$. Therefore, we generalize dynamic causation with the STI transformation, which is illustrated in Fig. 1a. This equation is referred to as the generalized causal equation in this study. Similarly, the delay embedding \mathbf{B}^t can be used to reliably predict a non-delay embedding which is topologically conjugated to



\mathbf{A}^t under certain dimension conditions. More detailed descriptions are provided in Supplementary Note 3.

Multistep forecasting framework with dynamic causation

For each observed state $\mathbf{X}^t = (x_1^t, x_2^t, \dots, x_n^t)'$ with n variables at time t ($t = 1, 2, \dots, m$), the temporal information $\mathbf{Y}^t = (y^t, y^{t+1}, \dots, y^{t+L-1})'$

can be constructed using the delay embedding scheme for any target variable y (e.g., $y^t = x_k^t$) ($k = 1, 2, \dots, n$), where the parameter L is the delay embedding dimension satisfying $L > 2d$. Here, \mathbf{X}^t is a known high-dimensional vector comprising multiple observed variables at time point t , while \mathbf{Y}^t is a temporal vector for the target variable y across multiple time points $t, t + 1, \dots, t + L - 1$. Denote $X = (\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^m)$ as the matrix

Fig. 1 | Schematic illustration of the CReP framework. **a** In a dynamical system where variable a causes variable b , the historical information of a is contained within the time series of b . Dynamic causation, grounded in Takens' embedding theorem ensures reliable predictability from the delay embedding manifold M_b to M_a . By incorporating the spatiotemporal information (STI) transformation, dynamic causation is extended to enable mappings between delay and non-delay manifolds, broadening the scope of causation analysis. **b** CReP is a self-supervised method to make multistep forecasting for a target variable y selected from the high-dimensional observables $\{x_1, x_2, \dots, x_n\}$. The original information \mathbf{X}^t is transformed into temporal information \mathbf{Y}^t by introducing three intermediary latent causal

representations $\mathbf{S}^t, \mathbf{Z}^t, \mathbf{V}^t$ as bridges that capture the relationships with the target variable. The effect representation \mathbf{Z}^t facilitates the future predictions $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$ in \mathbf{Y}^t , while \mathbf{Y}^t helps the temporal reconstruction for cause representation \mathbf{S}^t based on generalized dynamic causation. **c** The latent representations remain too abstract to understand although they are causally informative. To make causal analysis for the target variable y from a practical perspective, $\alpha\beta$ -LRP is employed in the CReP framework to calculate the relevance scores of input variables $\{x_1, x_2, \dots, x_n\}$ to the representations, which undirectly quantify the causal strength of input variables to the target variable y .

of observed states across all time points, and $\mathbf{Y} = (\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^m)$ as the matrix of delay embedding vectors across all time points as follows:

$$\mathbf{Y} = (\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^m) = \begin{pmatrix} y^1 & y^2 & \dots & y^m \\ y^2 & y^3 & \dots & y^{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ y^L & y^{L+1} & \dots & y^{m+L-1} \end{pmatrix}_{L \times m}. \quad (3)$$

Note that $\{y^1, y^2, \dots, y^m\}$ located in the upper-left area is the known information, while $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$ located in the lower-right area is the unknown information that remains to be predicted.

Observations from high-dimensional time series encompass rich information content on causal interactions. Based on their causal relationships with a selected target variable y , the information can be divided into three mutually exclusive and collectively exhaustive categories: cause-related information, effect-related information, and independent information. Inspired by the temporal information of y , the CReP directly decomposes the observed high-dimensional data \mathbf{X}^t into three orthogonal latent representations $\mathbf{S}^t, \mathbf{Z}^t, \mathbf{V}^t$ via a nonlinear function H , capturing three types of causal information for y , respectively, forming the following equations:

$$\begin{cases} H(\mathbf{X}^t) = \begin{pmatrix} \mathbf{S}^t \\ \mathbf{Z}^t \\ \mathbf{V}^t \end{pmatrix}, \\ H^{-1} \begin{pmatrix} \mathbf{S}^t \\ \mathbf{Z}^t \\ \mathbf{V}^t \end{pmatrix} = \mathbf{X}^t, \end{cases} \quad (4)$$

where $\mathbf{S}^t = (s_1^t, s_2^t, \dots, s_q^t)'$, $\mathbf{Z}^t = (z_1^t, z_2^t, \dots, z_q^t)'$ and $\mathbf{V}^t = (v_1^t, v_2^t, \dots, v_q^t)'$. The mapping H^{-1} is the conjugate function for reconstructing the original information. The decomposition and reconstruction processes can be realized by using an autoencoder (AE). According to the generalized causal equation (Eq. (2)), the effect representation \mathbf{Z}^t can be utilized to effectively predict the temporal information of y , and the cause representation \mathbf{S}^t can be reconstructed from y based on dynamic causation. Let $\mathbf{S} = (\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^m)$, $\mathbf{Z} = (\mathbf{Z}^1, \mathbf{Z}^2, \dots, \mathbf{Z}^m)$ and $\mathbf{V} = (\mathbf{V}^1, \mathbf{V}^2, \dots, \mathbf{V}^m)$ denote the matrices of cause, effect, and non-causal representations across all time points, respectively. Therefore, the original spatiotemporal information \mathbf{X} is transformed into three orthogonal causal representations $\mathbf{S}, \mathbf{Z}, \mathbf{V}$, effectively facilitating multistep predictions for the target variable. Specifically, the non-delay embedding vector \mathbf{Z}^t is mapped to the delay embedding vector \mathbf{Y}^t via a nonlinear function g , and \mathbf{Y}^t is mapped to the non-delay embedding vector \mathbf{S}^t via a nonlinear function f , forming the following causal equations with the approximation errors $\epsilon^{z,t}$ and $\epsilon^{y,t}$:

$$\begin{cases} g(\mathbf{Z}^t, \epsilon^{z,t}) = \mathbf{Y}^t, \\ f(\mathbf{Y}^t, \epsilon^{y,t}) = \mathbf{S}^t, \end{cases} \quad (5)$$

where the first equation is to forecast the unknown values for y with $g: \mathbb{R}^q \rightarrow \mathbb{R}^L$, and the second equation is to recover the cause representation

\mathbf{S}^t with $f: \mathbb{R}^L \rightarrow \mathbb{R}^q$. The prediction and reconstruction process can be effectively implemented using Temporal Convolutional Network (TCN)⁵⁴, which possesses multiple advantages including flexible receptive fields, stable gradient propagation and temporal consistency⁸. Let $\mathbf{Z}_{\text{win}} = (\mathbf{Z}^{t-w}, \mathbf{Z}^{t-w+1}, \dots, \mathbf{Z}^t)$ and $\mathbf{Y}_{\text{win}} = (\mathbf{Y}^{t-w}, \mathbf{Y}^{t-w+1}, \dots, \mathbf{Y}^t)$ with window size $w+1$ denote the sliding window matrices from the whole spatiotemporal matrices \mathbf{Z} and \mathbf{Y} , respectively. The causal equations based on the CReP are as follows:

$$\begin{cases} g(\mathbf{Z}_{\text{win}}, \epsilon^{z,t}) = \mathbf{Y}^t, \\ f(\mathbf{Y}_{\text{win}}, \epsilon^{y,t}) = \mathbf{S}^t, \end{cases} \quad (6)$$

where \mathbf{Y}^t represents the temporal information of the target variable at time points $t, t+1, \dots, t+L-1$, and \mathbf{S}^t represents multi-dimensional cause representation at time point t . The implementation details of CReP are illustrated in Supplementary Fig. 2. Given that the number of historical time points is m , we can predict the values of the target variable for the next $L-1$ time steps, i.e., $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$, and reconstruct the cause representation \mathbf{S} using the equation above. Clearly, properly determining the functions g and f are crucial for interpreting the causal relationships of y within the dynamic system and providing reliable multi-step forecasting of the future values. The details of delay embedding theorem and STI equations are provided in Supplementary Note 1 and Supplementary Note 2, respectively.

Explicit causation revealing based on layer-wise relevance propagation

As mentioned above, the original spatiotemporal information $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^m)$ is causally decomposed into three orthogonal latent representations that suggest different causal interactions with the target variable. However, these causal representations, e.g., $\mathbf{S} = (\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^m)$, $\mathbf{Z} = (\mathbf{Z}^1, \mathbf{Z}^2, \dots, \mathbf{Z}^m)$, are abstract and thus difficult to interpret in practical terms. To enhance the interpretability of these factors from a practical perspective, we focus on explaining the learned function H that performs the decomposition of the original space.

Layer-wise relevance propagation (LRP)^{39,40} is an effective method for explaining deep neural networks and has demonstrated its efficiency and accuracy for both classification and regression tasks⁵⁵⁻⁵⁷. Specifically, given a trained neural network, such as a Multilayer Perceptron (MLP)⁵⁸, the observed n variables are processed through the network to produce a prediction output $f(\mathbf{X})$. LRP then redistributes the output back to the input variables layer by layer and computes the relevance score $R_k^{(l)}$ ($k = 1, 2, \dots, n$) to assess the importance of the k -th input variable in the network's first layer to the output. To better understand the causal representations learned by CReP, the $\alpha\beta$ -rule-based LRP method ($\alpha\beta$ -LRP) is applied to interpret the decomposition function H by calculating the relevance scores of each input variable x_k to \mathbf{S} and \mathbf{Z} (see Fig. 1c), denoted by $RS_{k,S}$ and $RS_{k,Z}$, respectively. These scores are obtained by aggregating the relevance scores to the vector outputs \mathbf{S}^t and \mathbf{Z}^t , represented by RS_{k,S^t} and RS_{k,Z^t} , over all time points. $RS_{k,S}$ quantifies the contribution of variable x_k to the cause representation, thus indirectly used to infer the cause strength on y . $RS_{k,Z}$ measures the contribution of variable x_k to the effect representation, thus indirectly employed to infer the effect strength on y . A higher value of

$RS_{k,S}$ indicates a stronger causal link from x_k to y , while a higher value of $RS_{k,Z}$ indicates a stronger causal link from y to x_k . Consequently, the causal strength networks of y can be inferred by interpreting the latent causal factors S and Z from a practical perspective, which facilitates the identification of the causes and effects of y . The detailed procedure for calculating the relevance scores is provided in Methods, Supplementary Fig. 3 and Supplementary Note 5.

Performance of the framework on simulation models

To illustrate the application of the CReP framework, simulated time series datasets were generated using the Lorenz 96⁴¹, Kuramoto models⁴² and gene regulatory network. Detailed results concerning the forecasting performance and causal inference are demonstrated in Supplementary Fig. 4 and Supplementary Fig. 5. The known data for each dataset is shown in Supplementary Table 1. The specific parameter settings of CReP are provided in Supplementary Table 2 and Supplementary Table 3. More experimental runs of different target variables are provided in Supplementary Table 5. In addition, we compared the results with other baseline methods (details provided in Supplementary Note 8) as shown in Supplementary Table 4.

Lorenz96 model

The Lorenz 96 system, introduced by Edward Lorenz in 1996⁴¹, is a simplified mathematical model used to study atmospheric dynamics and chaotic behavior in weather prediction. The system consists of a set of ordinary differential equations for n variables that describe how a series of variables evolve over time with the following equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad (7)$$

where $i = 1, 2, \dots, n$ and the parameter F is an external forcing term. Note that the causal link from x_j to x_i exists if x_j is one of the independent variables of x_i , which is a standard criterion to identify dynamical causality in ordinary differential equations^{59–61}.

We first apply the CReP to a Lorenz 96 system with $n = 60$, taking a time series of 50 steps ($m = 50$) as known information and performing a 15-step ahead prediction ($L - 1 = 15$) for a randomly selected target variable y_1 . Note that the forecasted values for the target variable were obtained through a single prediction run, demonstrating the method's capability for effective multi-step forecasting. As shown in Fig. 2a and Fig. 2b, the CReP framework shows satisfactory performance in terms of the Pearson correlation coefficient (PCC) and the root mean square error (RMSE) values. Specifically, the RMSEs for the randomly selected samples were as low as 0.0141 and 0.0299. The average RMSE and PCC across all samples are shown in Table 1. In addition, as presented in Fig. 2c and Fig. 2d, the relevance scores of variables to causal representations were computed to quantify the causal strength of variables on the target variable. The inferred cause and effect strength networks constructed from the relevance scores, shown in Fig. 2f and Fig. 2g, align well with the causal relationships derived from the original equations (see Fig. 2e). For the target variable x_2 , the direct causes in the system are determined as x_{60}, x_3, x_1 , while the direct effects in the system are determined as x_4, x_3 and x_1 (Fig. 2e). Based on the CReP, x_{60}, x_3 and x_1 were identified as the variables most likely to cause the target variable (Fig. 2f, inferred cause strength network with only the top six variables), while x_4, x_3 and x_1 were identified as being among those most likely to be caused by the target variable (Fig. 2g, inferred effect strength network with only the top six variables). The results validated the reliability of the causal inference conducted in this study.

Power grid system

The Kuramoto system is a mathematical model used to study synchronization phenomena in systems of coupled oscillators. It was introduced by Yoshiki Kuramoto in 1975⁴². The model explores how a large number of oscillators, each with its own natural frequency, can synchronize when coupled together. We applied the CReP framework to a power grid system modeled by Kuramoto equations involving 120 units, which can be

represented as follows:

$$\frac{d\theta_i}{dt} = \omega_i + \gamma \sum_{j=1}^N A_{ij} \sin(\theta_j - \theta_i), \quad (8)$$

where θ_i and ω_i are the phase and natural frequency of the i -th oscillator respectively, γ is the coupling strength, while pairwise interactions are encoded in the adjacency matrix A ⁶².

For the 120-dimensional system ($n = 120$), we used 30 known time points ($m = 30$) and predicted 9 future time points ($L - 1 = 9$). The target variable y_1 was randomly selected from $\{\theta_1, \theta_2, \dots, \theta_{120}\}$. The forecasting performance for the target variable θ_{33} using the CReP framework is illustrated in Fig. 3a,b, and the average RMSE and PCC across all samples are demonstrated in Table 1. Additionally, causal inference for θ_{33} was conducted as shown in Fig. 3f,g, based on the computed relevance scores depicted in Fig. 3c,d. A higher relevance score to the cause representation indicates a stronger cause strength to the target variable, while a higher relevance score to the effect representation reflects a stronger effect strength to the target variable. For the true cause variables θ_{32} and θ_{86} (Fig. 3e), their cause strengths were the strongest, while the true effect variables θ_{32} and θ_{30} (Fig. 3e) exhibited prominent effect strengths compared to other variables. This fact demonstrates that causal analysis of the target variable can be effectively performed using the causal strength networks through the interpretation method $\alpha\beta$ -LRP.

Gene regulatory network

The DREAM4 (Dialogue for Reverse Engineering Assessments and Methods) dataset^{63–65} is a benchmark for evaluating algorithms in gene regulatory network inference, featuring simulated gene expression time series data generated from known biological networks. This dataset, produced using GeneNetWeaver (GNW)⁶⁶, includes various challenges, each characterized by distinct network structures, noise levels, and the number of genes. Numerous researchers utilize DREAM4 to assess the performance of causal inference methods and enhance the understanding of underlying biological mechanisms. By integrating GNW's capabilities, DREAM4 provides a robust framework for testing and refining methodologies in systems biology.

To validate the performance of CReP, we utilize the DREAM in silico dataset with 50 nodes ($n = 50$), which contains known gene interactions and incorporates noise to reflect biological variability. We employed the information from 40 known time points ($m = 40$) to forecast the future values of 8 time points ($L - 1 = 8$). To better demonstrate the performance of CReP, the target gene y_1 was selected as G18, which exhibits complex connections with other genes. The forecasting results for the target variable using CReP are visualized in Fig. 4a,b, with average RMSE of 0.0984 and PCC of 0.8894 shown in Table 1. Additionally, the true gene regulatory network for the target variable G18 is depicted in Fig. 4e, suggesting that G16, G22 and G23 are the primary causes of the dynamic change in G18, while G19, G20 and G28 are the variables mainly affected by G18. Derived from the relevance scores in Fig. 4c,d, the inferred causal strength networks of top 8 variables for the target variable are illustrated in Fig. 4f,g, highlighting that G16 and G22 exhibit prominent cause strength on G18, while G19 and G28 demonstrate significant effect strength on G18, which aligns well with the gold standard in Fig. 4e.

The application of the framework on real-world datasets

To better illustrate the generalizability of the CReP framework, we applied it to two real-world datasets of complex natural dynamic systems. The specific descriptions of each dataset are presented in Supplementary Note 6.

Cardiovascular inpatients forecasting

The first real-world dataset comprises a series of records on cardiovascular inpatients from major hospitals in Hong Kong, along with data on air pollutants^{43,44}. The pollutants include the daily concentrations of nitrogen dioxide (NO₂), sulfur dioxide (SO₂), ozone

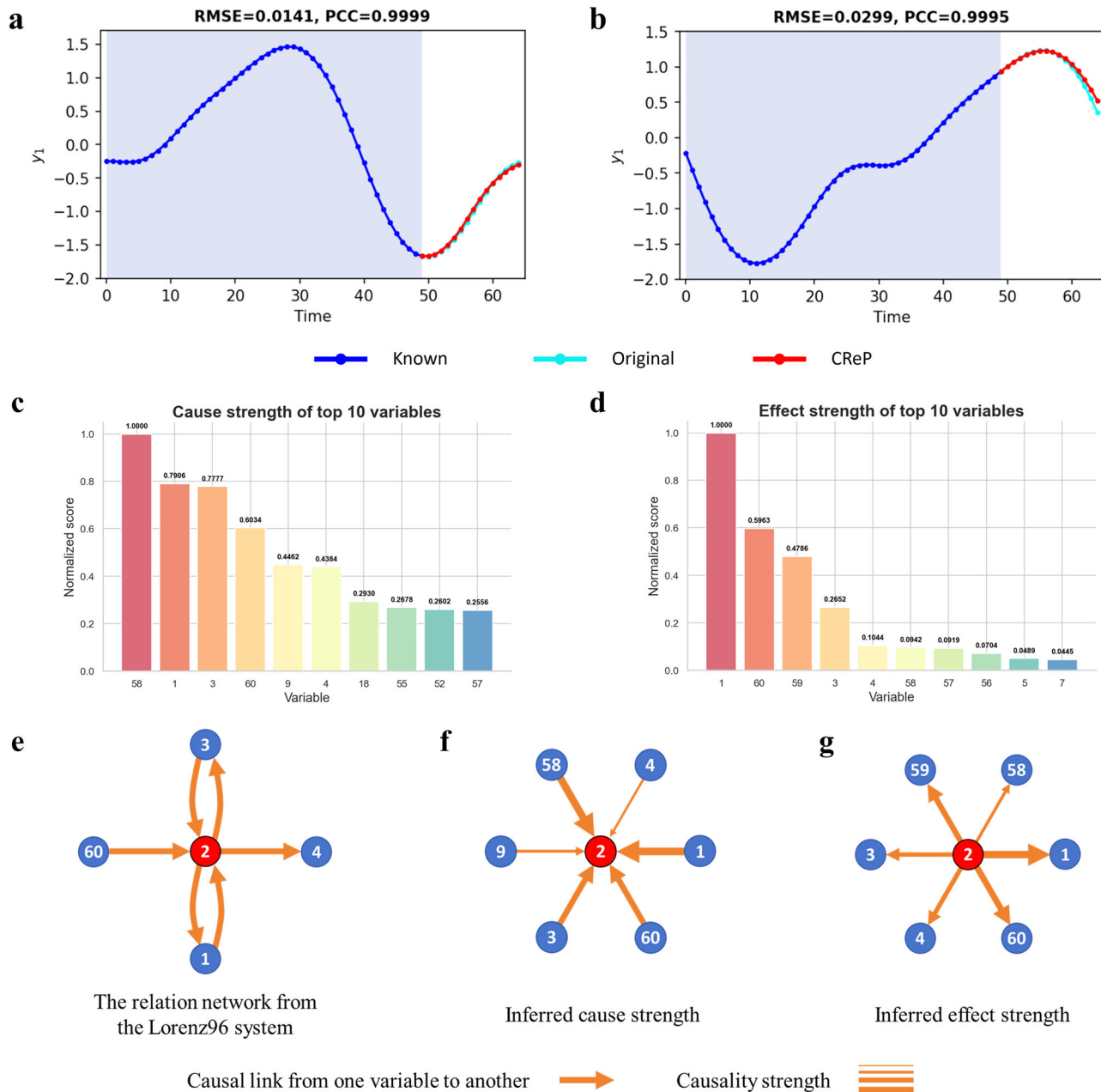


Fig. 2 | The forecasting performance and causal analysis of the CReP on the Lorenz 96 system. The time-series data were generated on the basis of the 60D Lorenz 96 system (Eq. (7)). The target variable was randomly selected among variables $\{x_1, x_2, \dots, x_{60}\}$. Taking a time series of 50 steps ($m = 50$) as known information, the CReP was applied to perform a 15-step-ahead prediction ($L - 1 = 15$) for the target variable. The prediction performance of CReP is shown over two periods in (a) and (b). The root mean square errors (RMSEs) between the CReP predictions and the original values are 0.0141 in (a) and 0.0299 in (b), while the

Pearson correlation coefficients (PCCs) are 0.9999 in (a) and 0.9995 in (b). By interpreting the causal representations via $\alpha\beta$ -LRP, the relevance scores for variables $\{x_1, x_2, \dots, x_{60}\}$ with regard to S and Z were calculated to quantify their causal strength on y , and the top 10 variables are shown in (c) and (d). Based on the CReP, the inferred causal strength networks (f) and (g) were constructed to show the responsible causes and effects of the target variable, which align well with the gold standard in (e).

(O_3), respirable suspended particulate (Rspar), mean daily temperature, relative humidity, and other factors. The data were obtained from air monitoring stations in Hong Kong between the years 1994 and 1997^{43,44}. Based on prior research demonstrating the association between air pollutants and cardiovascular admissions⁶⁷, the CReP was applied to predict daily cardiovascular disease admissions using the air pollutants data. For the 14-dimensional system ($n = 14$), the 70 determined-state time points were used ($m = 70$), and the 25 future values remained to be predicted ($L - 1 = 25$). The forecasting performance using the CReP is demonstrated in Fig. 5a,b. Table 1 shows the average RMSE and PCC across all samples.

Considering the real-world causal implications, we focus on inferring the cause strength for the target variable, i.e., cardiovascular diseases, to identify the most significant factors causing the number of cardiovascular inpatients. According to the inferred cause strength network shown in Fig. 5f, respirable suspended particulates and air pollutants (e.g., NO_2 and SO_2), were identified as primary contributors to cardiovascular diseases. This conclusion aligns well with the causal links in Fig. 5e supported by relevant literatures^{11,68–70} that have studied the relationship between air pollutants and cardiovascular inpatients. The results enable us to understand the causal mechanisms behind the evolution process, thereby providing

Table 1 | The average RMSEs and PCCs of each dataset across all samples

Dataset	RMSE	PCC
Lorenz 96	0.1086	0.9908
Power grid	0.1127	0.9868
Gene regulatory network	0.0984	0.8894
COVID19	0.2916	0.7432
HK	0.2285	0.9695

Our method is evaluated on three simulated datasets (Lorenz 96, power grid system and gene regulatory network) and two real-world datasets (COVID19 transmission and Hong Kong cardiovascular dataset). The forecasting performance metrics include the root mean square error (RMSE) and the Pearson correlation coefficient (PCC).

direction and guidance for implementing environmental protection and management to reduce the number of cardiovascular inpatients.

Japan Covid-19 transmission forecasting

The worldwide emergence of the coronavirus disease 2019 (COVID-19) pandemic has constituted a grave challenge to global public health. It is crucial to forecast the dissemination of this infectious disease in order to aid public health departments in their strategic planning endeavors. The second dataset contains a series of number representing COVID-19 daily new cases of 47 districts ($n = 47$) in Japan. The CReP in the study helps to predict the dynamic trends of daily new cases, thereby supporting more informed decision-making. Focusing on Tokyo with severe epidemic^{45,71}, the CReP provided a 14-step-ahead prediction ($L - 1 = 14$) of COVID-19 case numbers, based on 30 known time points ($m = 30$), as shown in Fig. 5c,d. The average forecasting performance in terms of RMSE and PCC across all samples is illustrated in Table 1.

Additionally, we analyzed the dynamic causal relationships among all districts by calculating relevance scores. The inferred cause and effect strength networks (Fig. 4g,h) indicate that Kyoto, Chiba, and Kanagawa were the primary cities influencing the dynamic changes in daily new cases in Tokyo, while Kanagawa, Saitama and Chiba were the main cities impacted by the spread of the disease in Tokyo. To better understand these causal relationships, the geographic locations around Tokyo are depicted in Fig. 5i. The proximity of Chiba and Kanagawa to Tokyo supports the rationale that these cities significantly influenced Tokyo's COVID-19 case numbers. Although Kyoto is geographically farther from Tokyo, the frequent population flow from it to Tokyo made it a significant factor in affecting the infection rates in Tokyo. Similarly, Kanagawa, Saitama and Chiba are geographically close to Tokyo, thus explaining their susceptibility to the case dynamics originating in Tokyo. The results help to better understand the disease transmission between regions and allows for timely implementation of isolation, control, and other measures before the pandemic spreads across regions, thus effectively containing its expansion.

Ablation study on loss function

To further demonstrate the effectiveness of CReP framework, we conduct an ablation study by eliminating different loss terms from our training loss. The experiment results on the simulation datasets are presented in Table 2. The “without \mathcal{L}_{REC} ”, “without \mathcal{L}_{FC} ” and “without $\mathcal{L}_{\text{ORTH}}$ ” columns indicate that the reconstruction loss, the future-consistency loss and the orthogonal loss are eliminated from loss function during training CReP, respectively. Removing \mathcal{L}_{REC} prevents CReP from guaranteeing the recovery of original high-dimensional information. Without \mathcal{L}_{FC} , the temporally self-constrained conditions are not enforced, and the removal of $\mathcal{L}_{\text{ORTH}}$ means that CReP is prone to extracting overlapping or insufficient causal information from observed data.

In this ablation study, the model performance is evaluated from two perspectives: time series forecasting and causal analysis. Forecasting

performance is assessed using the root mean square error (RMSE), while causal performance requires further quantification of inferred causal strengths to enable accuracy as a metric. For the inferred causal strengths obtained through relevance scores, P-value hypothesis testing is applied to determine an appropriate threshold, enabling the identification of causes and effects associated with the target variable (Supplementary Fig. 6). This identification allows for the construction of a directed causal graph centered on the target variable, without distinctions in causal strength. By comparing the inferred causal graph to the gold standard, metrics such as accuracy and recall are computed to assess causal performance in the ablation study. The overall process to evaluate the inferred causal results are summarized as follows:

1. Determine an appropriate threshold using P-value hypothesis testing.
2. Identify the causes and effects related to the target variable.
3. Construct an inferred causal graph centered on the target variable.
4. Calculate accuracy and recall against the gold standard.

Through the evaluation process outlined above, a comparison is conducted between the predictive performance and causal analysis capability of the model after removing each component of the loss function. As presented in the Table 2, the model trained with the full loss achieves the best overall performance, indicating that all the constrains in the CReP framework contribute to producing accurate time-series forecasting and robust causal analysis. Further details regarding the ablation study are provided in Supplementary Note 7.

Discussion

Time series forecasting, which predicts future trends based on historical data, can incorporate spatial information and causal interactions to improve predictive accuracy in complex systems. In this work, we proposed the CReP framework to learn causal representations from observed data, thereby performing robust multi-step-ahead predictions. The CReP framework causally transforms high-dimensional information into temporal information of the target variable by introducing three intermediary latent representations—cause, effect, and non-causal—as bridges that capture the causal relationships with the target variable. Specifically, an autoencoder is utilized to decompose the spatiotemporal information X into the cause representation S , the effect representation Z and the non-causal representation V by H , and then recover these causal representations to the original information by H^{-1} . According to dynamic causation incorporating the STI transformation, the effect representation Z^t is mapped to the temporal vector \mathbf{Y}^t by g , thereby forecasting multiple future values $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$, and \mathbf{Y}^t is employed to recover the cause representation \mathbf{S}^t by f . The nonlinear functions H, g, f, H^{-1} are fit simultaneously in a self-supervised manner. In addition, the CReP framework implicitly uncovers underlying causal interactions by learning these causal representations, facilitating robust predictions based on causal equations. Moreover, explicit causal inference is performed through the interpretation of learned causal representations S and Z using $\alpha\beta$ -LRP, which calculates the relevance scores of each variable to the outputs S and Z . This facilitates the identification of causes and effects of the target variable. Therefore, the CReP not only achieves multi-step-ahead forecasting but also infers the causal network for the target variable, offering a comprehensive understanding of the system's intrinsic dynamics. The applications of CReP to both simulation models and real-world datasets demonstrate its effectiveness in forecasting and causal analysis.

The CReP framework offers several advantages. Firstly, with a self-supervised training scheme, the CReP integrates multi-step time-series forecasting with implicit causal discovery by utilizing high-dimensional but short-term time series data. Secondly, the use of $\alpha\beta$ -LRP to interpret the learned causal representations reliably reveals the causal relationships within the dynamical system, which is depicted by causal strength networks. Third, built on a solid theoretical foundation of dynamic causation

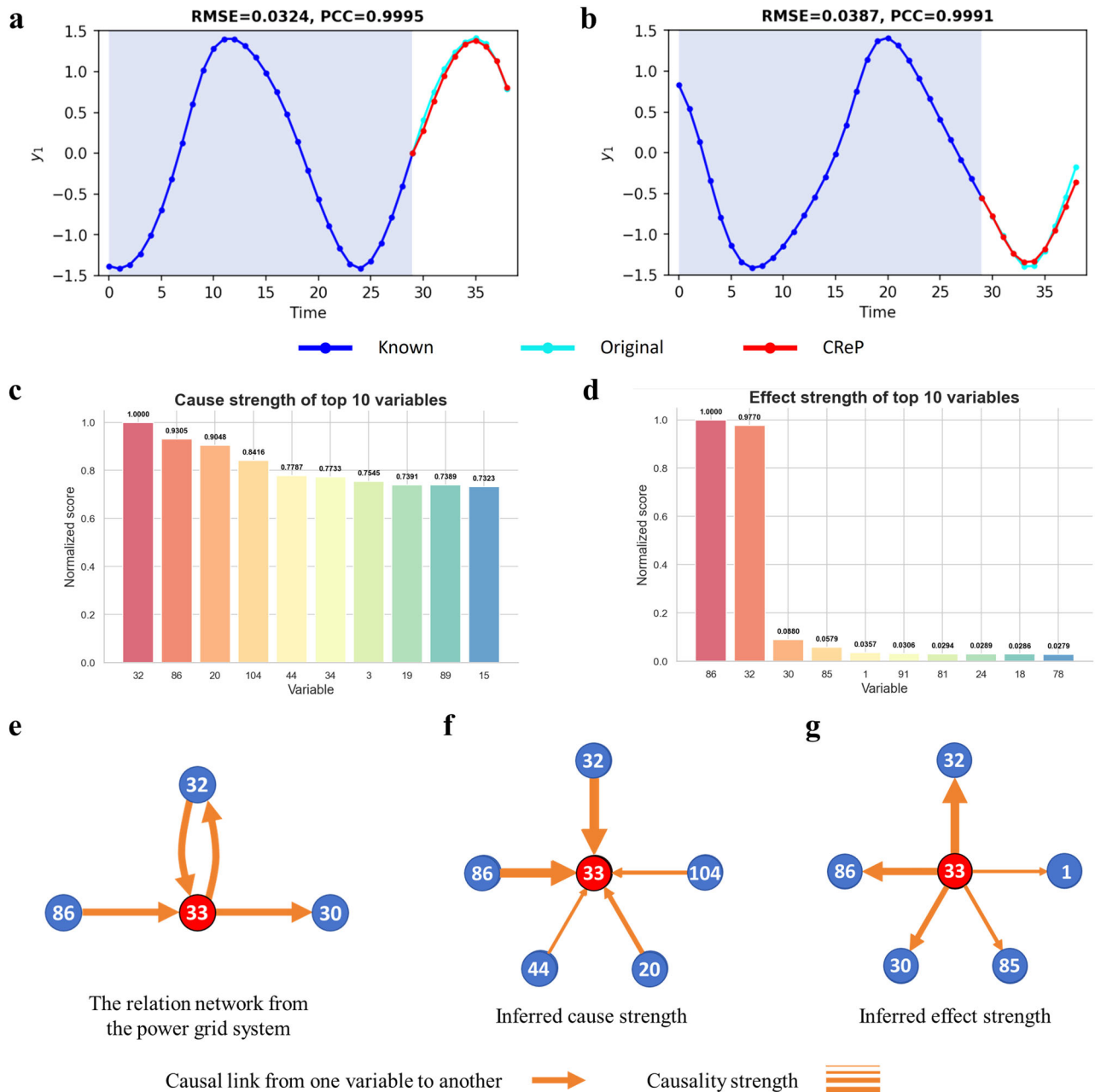


Fig. 3 | The forecasting performance and causal analysis of the CReP on the power grid system. The time-series data were generated using the 120D Kuramoto model (Eq. (8)). We randomly selected the target variable among variables $\{\theta_1, \theta_2, \dots, \theta_{120}\}$. By applying the CReP framework, we used 30 known time points ($m = 30$) and predicted 9 future time points ($L - 1 = 9$) for the target variable. The forecasting results of CReP on the system are demonstrated in (a) and (b).

Additionally, (c) and (d) show the relevance scores of top 10 variables to the cause representation S and the effect representation Z , respectively. As presented in the inferred causal networks in (f) and (g), the responsible causes and effects for the target variable are highlighted based on the calculated relevance scores in (c) and (d). The true causal network for the target variable is shown in (e).

and the STI transformation, the CReP presents an innovative approach for exploring spatiotemporal information in high-dimensional time series from a causal perspective, enhancing temporal prediction accuracy and showing strong potential for practical applications. However, there are still limitations in causal-oriented representation learning within CReP, particularly in finding universally applicable explanatory methods for data across all domains. In the future, we will focus on integrating different causal learning methods to enhance applicability across diverse domains. In addition, we will further explore causal detection methods through active intervention rather than mere passive observations, as they offer more intuitive causal interpretations and identify causal directions with high accuracy as well⁷².

Methods

Dynamic causation and generalized embedding theorem

According to the principle of dynamic causation, when the variable a acts as a cause of the variable b within a dynamical system, the temporal information of a can be reliably predicted using b , as shown in Fig. 1a. Even in the presence of noise, dynamic causation is still applicable according to the stochastic version of Takens' embedding theorem^{51,52} (see Supplementary Note 1). However, the mapping h in Eq. (1) is constructed within delay embedding spaces, which limits its broader applicability for detecting causality. The generalized embedding theorem³⁷ provides a robust theoretical foundation for extending the applicability of the mapping h . According to the theorem:

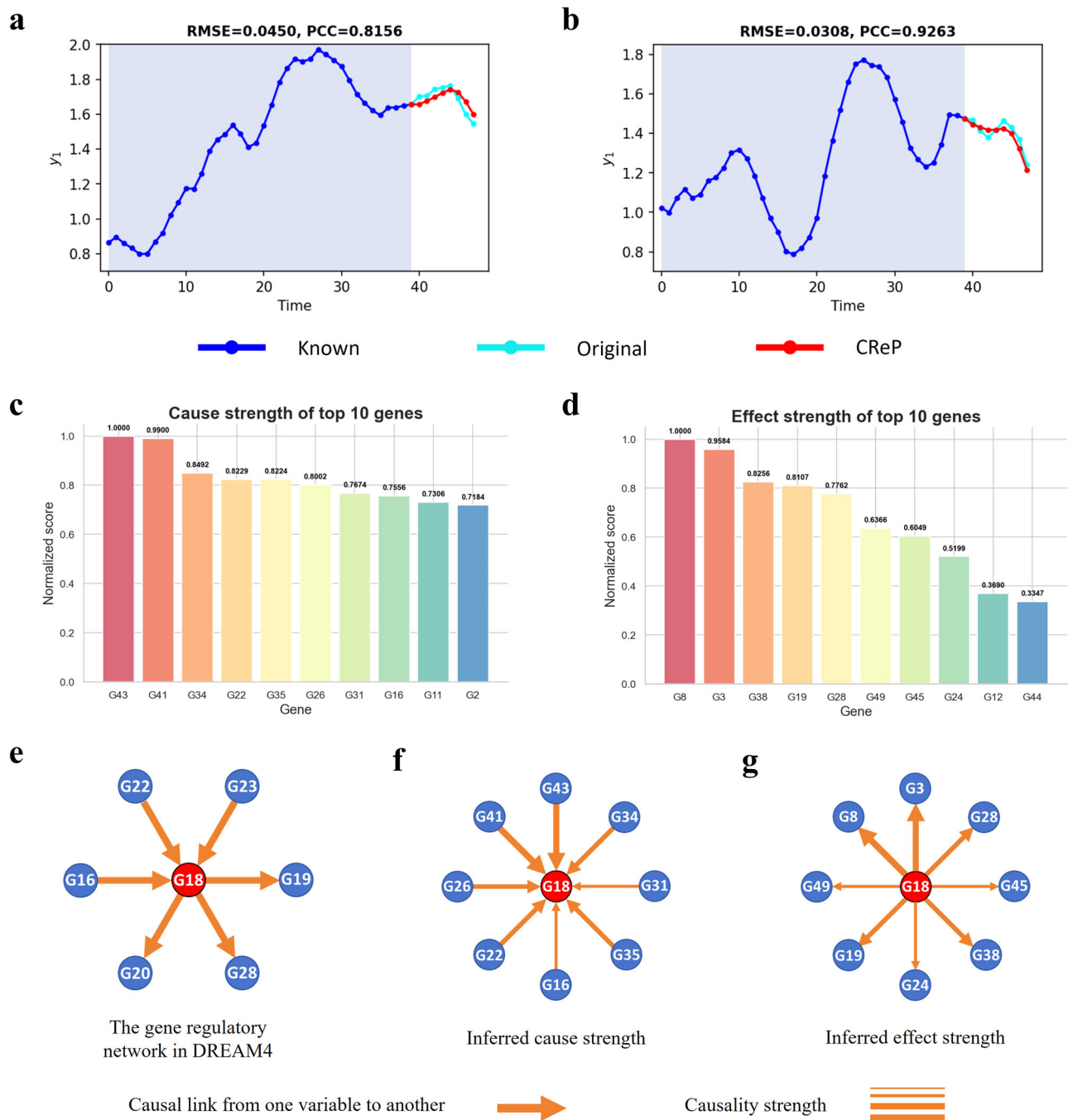


Fig. 4 | The forecasting performance and causal analysis of the CReP on the gene regulatory network. We utilized the DREAM in silico dataset with 50 nodes ($n = 50$) and applied a simple smoothing technique to reduce the noise in data. The target variable was selected due to its complex connections. Using the CReP with parameter $m = 40$ (i.e., the length of the input series is 40), we performed 8-step-ahead forecasting ($L - 1 = 8$) for the target variable, as shown in (a) and (b).

Additionally, the relevance scores of each gene to the causal representations were calculated and the top 10 variables are displayed in (c) and (d). The network in (e) illustrates the true causal relations of the target variable within the dataset. The cause strength network in (f) is constructed from the relevance scores in (c), while the effect strength network in (g) is derived from the relevance scores in (d). Genes with higher relevance scores indicate greater causal influence on the target variable.

If M is a compact manifold with box-counting dimension d , for a diffeomorphism $\phi : M \rightarrow M$ and L smooth observation functions $y_k : M \rightarrow \mathbb{R}$, then it is a generic property that the mapping described by

$$\Phi_{(\phi, \langle y_k \rangle)} = (y_1(X), y_2(X), \dots, y_L(X)), \quad (9)$$

is an embedding as long as $L > 2d$.

According to the theory, the delay embedding manifold of the response variable b can be topologically conjugated to a non-delay embedding manifold⁵³, which is summarized as the spatiotemporal information (STI)

transformation⁸. Therefore, the mapping h in Eq. (1) can be extended to map a non-delay manifold to a delay manifold and we can obtain the generalized causal equation Eq. (2), which provides a promising perspective for causal analysis and multistep forecasting of high-dimensional but short-term time series in CReP.

Unified causal and temporal modeling with CReP

We assume that the high-dimensional observations X are generated by latent variables that support causal statements^{73,74}. Specifically, given the observed target variable y , we make the following assumptions: a) the data is

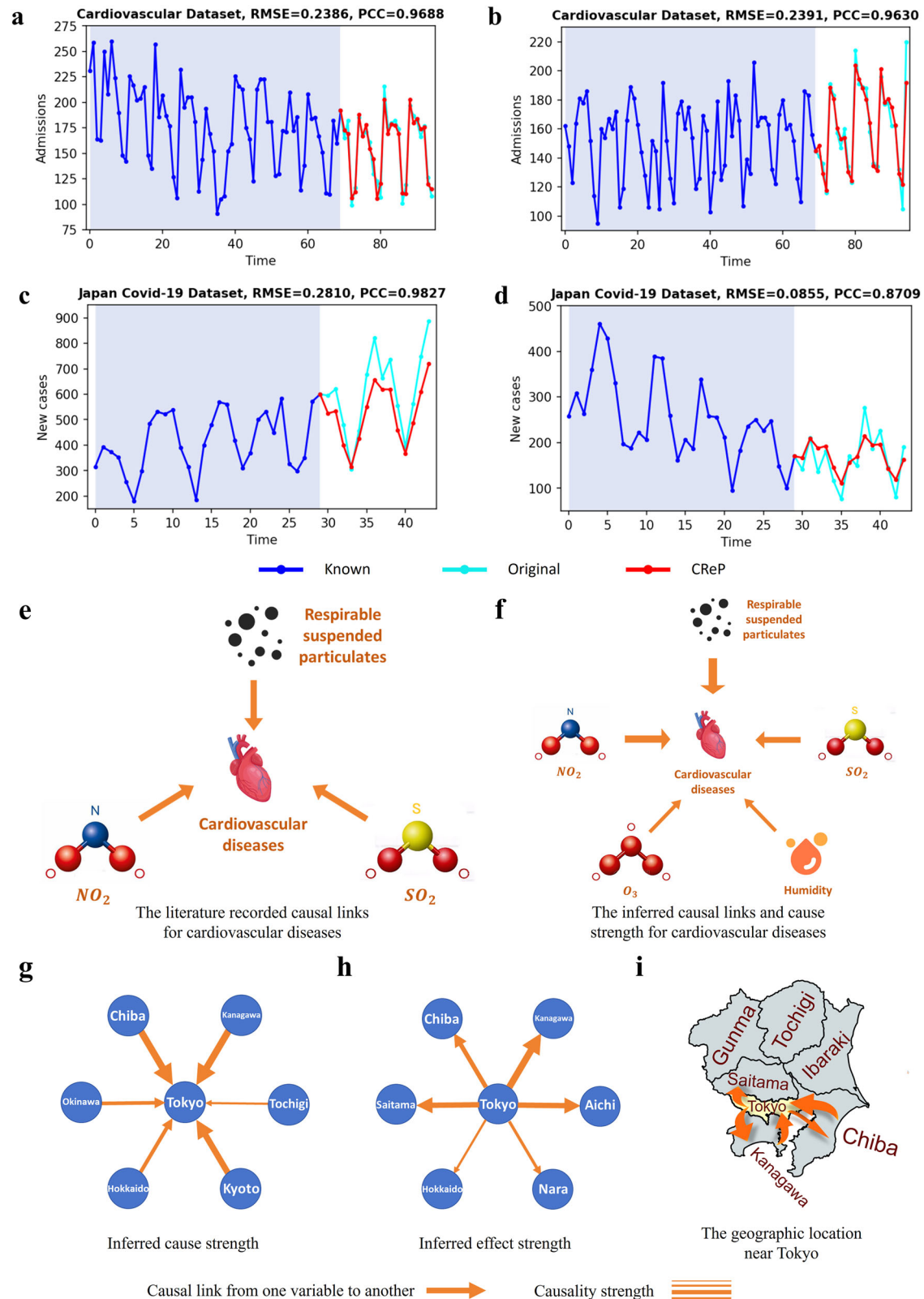


Fig. 5 | Performance of the CReP on the two real-world datasets. For two periods (a) and (b), the CReP predicted the number of cardiovascular admissions using the high-dimensional time series of air pollutant indices, with an input length of $m = 70$ and a forecasting horizon of $L - 1 = 25$. Based on the time series of COVID-19 new cases of 47 districts, the CReP predicted the number of future new cases for Tokyo over two periods in (c) and (d), with 30 steps of known information ($m = 30$) and

14 steps of future information ($L - 1 = 14$). Additionally, the causes of cardiovascular admissions inferred by CReP, shown in (f), closely align well with the true drivers in (e). For COVID-19 transmission in Tokyo, the inferred causal strength networks are demonstrated in (g) and (h), constructed from the relevance scores. As illustrated in (i), the dominant causes and effects inferred by CReP are geographically near Tokyo, supporting the reliability of causal analysis using CReP.

Table 2 | The ablation study results of CReP on three simulated datasets

Dataset		Lorenz 96	Power grid	Gene regulatory network
Original	RMSE	0.1089	0.1127	0.0984
	Accuracy	0.95	0.82	0.86
	F1-score	0.67	0.15	0.36
	Recall	1.0	1.0	0.67
Without $\mathcal{L}_{\text{ORTH}}$	RMSE	0.1352	0.1023	0.1029
	Accuracy	0.92	0.80	0.82
	F1-score	0.44	0.08	0.18
	Recall	0.67	0.5	0.33
Without \mathcal{L}_{REC}	RMSE	0.1270	0.1326	0.1018
	Accuracy	0.93	0.82	0.80
	F1-score	0.56	0.15	0.09
	Recall	0.83	1.0	0.17
Without \mathcal{L}_{FC}	RMSE	0.1142	0.1206	0.1169
	Accuracy	0.93	0.82	0.78
	F1-score	0.56	0.15	0.0
	Recall	0.83	1.0	0.0

The performance of the model trained with the full loss is compared with that of models trained with each individual loss term removed. The root mean square error (RMSE) is used to evaluate forecasting performance, while accuracy, F1-score, and recall assess the reliability of causal identification.

generated from three type of latent variables, corresponding to the cause factors, effect factors and non-causal factors of y ; b) these three types of latent variables are mutually independent; and c) the latent variables are unobserved⁷⁵. Accordingly, CReP employs an autoencoder to decompose X into three mutually orthogonal causal representations implicitly based on their causal relationship with y via the nonlinear function H , and recover the original information using the causal representations via the conjugate mapping H^{-1} (Eq. (4)). Furthermore, cause factors and effect factors are useful representations of the data for disentangling causal dependencies of y in the latent space based on dynamic causation⁷⁶.

Assuming $g = (g_1, g_2, \dots, g_L)'$ and $f = (f_1, f_2, \dots, f_q)'$, we can establish the following causal equations for the target variable y according to their implicit causal information based on the TCN scheme:

$$\begin{pmatrix} g_1((Z^1), \epsilon^{z,t}) & g_1((Z^1, Z^2), \epsilon^{z,t}) & \dots & g_1((Z^{m-w}, \dots, Z^m), \epsilon^{z,t}) \\ g_2((Z^1), \epsilon^{z,t}) & g_2((Z^1, Z^2), \epsilon^{z,t}) & \dots & g_2((Z^{m-w}, \dots, Z^m), \epsilon^{z,t}) \\ \vdots & \vdots & \ddots & \vdots \\ g_L((Z^1), \epsilon^{z,t}) & g_L((Z^1, Z^2), \epsilon^{z,t}) & \dots & g_L((Z^{m-w}, \dots, Z^m), \epsilon^{z,t}) \end{pmatrix} = \begin{pmatrix} y^1 & y^2 & \dots & y^m \\ y^2 & y^3 & \dots & y^{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ y^L & y^{L+1} & \dots & y^{m+L-1} \end{pmatrix}, \quad (10)$$

$$\begin{pmatrix} f_1((Y^1), \epsilon^{y,t}) & f_1((Y^1, Y^2), \epsilon^{y,t}) & \dots & f_1((Y^{m-w}, \dots, Y^m), \epsilon^{y,t}) \\ f_2((Y^1), \epsilon^{y,t}) & f_2((Y^1, Y^2), \epsilon^{y,t}) & \dots & f_2((Y^{m-w}, \dots, Y^m), \epsilon^{y,t}) \\ \vdots & \vdots & \ddots & \vdots \\ f_q((Y^1), \epsilon^{y,t}) & f_q((Y^1, Y^2), \epsilon^{y,t}) & \dots & f_q((Y^{m-w}, \dots, Y^m), \epsilon^{y,t}) \end{pmatrix} = \begin{pmatrix} s_1^1 & s_1^2 & \dots & s_1^m \\ s_2^1 & s_2^2 & \dots & s_2^m \\ \vdots & \vdots & \ddots & \vdots \\ s_q^1 & s_q^2 & \dots & s_q^m \end{pmatrix}. \quad (11)$$

By leveraging dynamic causation with y , the CReP provides the multi-step forecasting, i.e., $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$, simultaneously uncovering the causal interactions for y through analyzing and interpreting the causal representations.

Optimization of CReP

Grounded in the causal equations above, the CReP effectively leverages the observed spatiotemporal information via causal decomposition, thereby improving the accuracy and robustness of multi-step forecasting. The

determination of H, g, f, H^{-1} relies on a self-supervised learning scheme. In this study, each layer of the nonlinear mappings is followed by the LeakyReLU activation function. The optimization of our network uses a “consistently self-constraint scheme”⁸ which focuses on the preservation of temporal consistency for the target variable. For the delay embedding matrix Y , there are m known historical values $\{y^1, y^2, \dots, y^m\}$ and $L - 1$ to-be-predicted future values $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$. The estimated delay embedding matrix \hat{Y} in each iteration is obtained through Eq. (10):

$$\hat{Y} = \begin{pmatrix} (\hat{y}^1)_1 & (\hat{y}^2)_1 & \dots & (\hat{y}^m)_1 \\ (\hat{y}^2)_2 & (\hat{y}^3)_2 & \dots & (\hat{y}^{m+1})_2 \\ \vdots & \vdots & \ddots & \vdots \\ (\hat{y}^L)_L & (\hat{y}^{L+1})_L & \dots & (\hat{y}^{m+L-1})_L \end{pmatrix}. \quad (12)$$

We require that the estimates $(\hat{y}^j)_j$ obtained from different sub-mapping functions $g_j (j \in \{1, 2, \dots, L\})$ at the same time point be consistent for both historical and future information, forming $m - 1$ and $L - 2$ self-constrained conditions, respectively. In total, these $m + L - 3$ conditions regulate the network training process by enforcing temporal sequence consistency across samples.

Through the auto perception procedure, CReP is optimized through minimizing the following loss function, which comprises four weighted components:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{DS}} + \lambda_2 \mathcal{L}_{\text{FC}} + \lambda_3 \mathcal{L}_{\text{REC}} + \lambda_4 \mathcal{L}_{\text{ORTH}}. \quad (13)$$

The determined-state loss \mathcal{L}_{DS} for the target variable y over time points $\{1, 2, \dots, m\}$ is calculated using the root mean squared error (RMSE), quantifying the estimation error of all historical values of y . The second part \mathcal{L}_{FC} is the future consistency loss for y over time points $\{m + 1, m + 2, \dots, m + L - 1\}$ to measure the deviation between multiple estimates of y for the same future time point using RMSE. The third part \mathcal{L}_{REC} is the reconstruction loss for latent cause representation S and the original information X , designed to assess the degree of information recovery. The last part $\mathcal{L}_{\text{ORTH}}$ in Eq. (13) is the orthogonal loss between implicit causal representations for the efficient utilization of spatiotemporal information. The detailed mathematical explanations of these loss terms, along with the ablation study, are provided in Supplementary Note 7.

Multistep forecasting and causal analysis by CReP

The CReP framework is carried out to predict the future values $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$ of the target variable y , which is selected from $\{x_1, x_2, \dots, x_n\}$, and explore the causal relationships for y . After training the network by minimizing the loss function in a self-supervised manner, the $L - 1$ future values $\{y^{m+1}, y^{m+2}, \dots, y^{m+L-1}\}$ can be determined from the estimated delay embedding matrix \hat{Y} as follows:

$$y^{m+i} = \text{mean}(\hat{y}^{m+i}) = \frac{1}{L-i} \sum_{j=i+1}^L (\hat{y}^{m+i})_j, \quad (14)$$

where $i = 1, 2, \dots, L - 1$.

While generalized dynamic causation ensures that the learned representations can effectively model causal interactions, they remain too abstract to interpret directly in a practical context. To uncover the causal links associated with y —the causes and effects— $\alpha\beta$ -LRP is employed to explain the transformation mechanism H from X to S and Z . Specifically, for a multi-layer neural network, let c_i and c_j represent neurons in adjacent layers l and $l + 1$, respectively, where i and j denote the indices of the neurons. A common mapping from one layer to the next one can be expressed in the following form:

$$c_j = \sigma \left(\sum_{i:i \in K} z_{ij} \right), \quad (15)$$

where $z_{ij} = w_{ij}c_i$, σ is the activation function, and w_{ij} is the weight between neurons c_i and c_j . The set K contains all neurons in the l -th layer. LRP calculates the relevance scores of all neurons from the output $f(X)$ backward to the input variables. Let $R_i^{(l)}$ and $R_j^{(l+1)}$ denote the relevance scores of neurons c_i and c_j in layers l and $l+1$, respectively. The amount of relevance score $R_j^{(l+1)}$ that spread to neuron c_i from neuron c_j is denoted as $R_{i \leftarrow j}^{(l+1)}$. By posing the following constraint:

$$R_i^{(l)} = \sum_{j: c_j \text{ is input for neuron } c_i} R_{i \leftarrow j}^{(l+1)}, \quad (16)$$

the conservation property for the relevance scores would hold between layers globally⁵⁶:

$$\sum_{k=1}^n R_k^{(1)} = \dots = \sum_{i=1}^{n_l} R_i^{(l)} = \sum_{j=1}^{n_{l+1}} R_j^{(l+1)} = \dots = f(X), \quad (17)$$

where n is the number of input variables, n_l and n_{l+1} are the number of neurons in layer l and $l+1$, and the output value $f(X)$ is the relevance score of the output neuron. The key process for LRP is to calculate the message $R_{i \leftarrow j}^{(l+1)}$, for which various rules have been proposed^{39,40}. The adopted propagation method in this work is the $\alpha\beta$ -rule^{39,40} defined as

$$R_{i \leftarrow j}^{(l+1)} = \left(\alpha \frac{z_{ij}^+}{z_j^+} + \beta \frac{z_{ij}^-}{z_j^-} \right) R_j^{(l+1)}, \quad (18)$$

where $z_{ij}^+ = \max\{z_{ij}, 0\}$, $z_{ij}^- = \min\{z_{ij}, 0\}$, $z_j^+ = \sum_i z_{ij}^+$, and $z_j^- = \sum_i z_{ij}^-$. The parameters α and β need to satisfy the constraint $\alpha + \beta = 1$. By propagating relevance layer by layer, the relevance scores $R_k^{(1)}$ ($k = 1, 2, \dots, n$) are obtained to evaluate the contribution of each input variable to the network's output. Higher relevance scores indicate greater contributions of the input variables to the output.

In CReP, the relevance scores of each input variable x_k to causal factors S and Z , denoted by $RS_{k,S}$ and $RS_{k,Z}$, are calculated to reveal the explicit causal relationships of the target variable. A comprehensive description of $\alpha\beta$ -LRP in CReP can be found in Supplementary Note 4. $RS_{k,S}$ quantifies the contribution of variable x_k to the cause representation, while $RS_{k,Z}$ measures the contribution of x_k to the effect representation. Based on the quantified contributions of all input variables to the causal representations, the inferred causal networks centered on the target variable can be constructed, where the thickness of the edges represents the strength of causal interactions. By employing the network interpretation method, the learned implicit and abstract causal information is transformed into explicit and interpretable causal relationships, thereby enhancing the understanding of the internal mechanisms of the system.

Data availability

The data that support the findings of this study are available at <https://github.com/csh-cut/causal>.

Code availability

All relevant code used in this study is available at <https://github.com/csh-cut/causal>.

Received: 13 December 2024; Accepted: 2 June 2025;

Published online: 10 June 2025

References

- Sussillo, D. & Barak, O. Opening the Black Box: Low-Dimensional Dynamics in High-Dimensional Recurrent Neural Networks. *Neural Comput* **25**, 626–649 (2013).
- Csete, M. E. & Doyle, J. C. Reverse Engineering of Biological Complexity. *Science* **295**, 1664–1669 (2002).
- Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms* (The MIT press, Cambridge, Mass, 2017).
- Duan, Z., Xu, H., Huang, Y., Feng, J. & Wang, Y. Multivariate Time Series Forecasting with Transfer Entropy Graph. *Tsinghua Sci. Technol.* **28**, 141–149 (2023).
- Sun, Y. et al. Using causal discovery for feature selection in multivariate numerical time series. *Mach. Learn.* **101**, 377–395 (2015).
- Oliveira, D. C., Lu, Y., Lin, X., Cucuringu, M. & Fujita, A. Causality-Inspired Models for Financial Time Series Forecasting. *arXiv preprint arXiv:2408.09960* (2024).
- Löwe, S., Madras, D., Zemel, R. S. & Welling, M. Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data. In *1st Conference on Causal Learning and Reasoning, CLeaR 2022, Sequoia Conference Center, Eureka, CA, USA, 11–13 April, 2022* (eds Schölkopf, B., Uhler, C. & Zhang, K.) vol. 177 509–525 (PMLR, 2022).
- Peng, H., Chen, P., Liu, R. & Chen, L. Spatiotemporal information conversion machine for time-series forecasting. *Fundam. Res.* S2667325822004538. <https://doi.org/10.1016/j.fmr.2022.12.009> (2022).
- Box, G. E. P. & Pierce, D. A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *J. Am. Stat. Assoc.* **65**, 1509–1526 (1970).
- Chandra, S. R. & Al-Deek, H. Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. *J. Intell. Transp. Syst.* **13**, 53–72 (2009).
- Yan, J. et al. Dynamical causality under invisible confounders. *arXiv preprint arXiv:2408.05584* (2024).
- Jiang, J. & Lai, Y.-C. Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. *Phys. Rev. Res.* **1**, 033056 (2019).
- Cho, K. et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25–29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL* (eds Moschitti, A., Pang, B. & Daelemans, W.) 1724–1734 (ACL, 2014). <https://doi.org/10.3115/V1/D14-1179>.
- Karevan, Z. & Suykens, J. A. K. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Netw.* **125**, 1–9 (2020).
- Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput* **9**, 1735–1780 (1997).
- Lai, G., Chang, W.-C., Yang, Y. & Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08–12, 2018* (eds Collins-Thompson, K., Mei, Q., Davison, B. D., Liu, Y. & Yilmaz, E.) 95–104 <https://doi.org/10.1145/3209978.3210006> (ACM, 2018).
- Chen, T. Q. et al. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3–8, 2018, Montréal, Canada* (eds Bengio, S. et al.) 6572–6583 (2018).
- Li, S. et al. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8–14, 2019, Vancouver, BC, Canada* (eds Wallach, H. M. et al.) 5244–5254 (2019).
- Zhou, H. et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **35**, 11106–11115 (2021).

20. Kitaev, N., Kaiser, L. & Levskaya, A. Reformer: The Efficient Transformer. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020* (OpenReview.net, 2020).
21. Liu, Y. et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7–11, 2024* (OpenReview.net, 2024).
22. Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688–701 (1974).
23. Pearl, J. Causal diagrams for empirical research. *Biometrika* **82**, 669–688 (1995).
24. Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009).
25. Goldberg, L. R. The Book of Why: The New Science of Cause and Effect: by Judea Pearl and Dana Mackenzie, Basic Books (2018). ISBN: 978-0465097609. *Quant. Financ.* **19**, 1945–1949 (2019).
26. Shi, J., Chen, L. & Aihara, K. Embedding entropy: a nonlinear measure of dynamical causality. *J. R. Soc. Interface* **19**, 20210766 (2022).
27. Pearl, J. *Causality: Models, Reasoning, and Inference*. (Cambridge University Press, Cambridge New York, NY Port Melbourne New Delhi Singapore, 2022).
28. Tao, P. et al. Detecting dynamical causality by intersection cardinal concavity. *Fundam. Res.* S2667325823000122. <https://doi.org/10.1016/j.fmre.2023.01.007> (2023).
29. He, Y.-B., Geng, Z. & Liang, X. Learning Causal Structures Based on Markov Equivalence Class. In *Algorithmic Learning Theory* (eds Jain, S., Simon, H. U. & Tomita, E.) vol. 3734, 92–106 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005).
30. Granger, C. W. J. Some properties of time series data and their use in econometric model specification. *J. Econom.* **16**, 121–130 (1981).
31. Arnhold, J., Grassberger, P., Lehnertz, K. & Elger, C. E. A robust method for detecting interdependencies: application to intracranially recorded EEG. *Phys. Nonlinear Phenom.* **134**, 419–430 (1999).
32. Takens, F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980* (eds Rand, D. & Young, L.-S.) vol. 898, 366–381 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1981).
33. Sauer, T., Yorke, J. A. & Casdagli, M. Embedology. *J. Stat. Phys.* **65**, 579–616 (1991).
34. Sugihara, G. et al. Detecting Causality in Complex Ecosystems. *Science* **338**, 496–500 (2012).
35. Ma, H. et al. Detection of time delays and directional interactions based on time series from complex dynamical systems. *Phys. Rev. E* **96**, 012221 (2017).
36. Ma, H., Aihara, K. & Chen, L. Detecting Causality from Nonlinear Dynamics with Short-term Time Series. *Sci. Rep.* **4**, 7464 (2014).
37. Deyle, E. R. & Sugihara, G. Generalized Theorems for Nonlinear State Space Reconstruction. *PLoS ONE* **6**, e18295 (2011).
38. Chen, P., Liu, R., Aihara, K. & Chen, L. Autoreervoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nat. Commun.* **11**, 4568 (2020).
39. Bach, S. et al. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE* **10**, e0130140 (2015).
40. Achtabat, R. et al. From attribution maps to human-understandable explanations through Concept Relevance Propagation. *Nat. Mach. Intell.* **5**, 1006–1019 (2023).
41. Lorenz, E. N. Predictability: A Problem Partly Solved. In *Proc. of the Seminar on Predictability* vol. 1 1–18 (European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK, 1996).
42. Kuramoto, Y. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics* (ed. Araki, H.) vol. 39, 420–422 (Springer-Verlag, Berlin/Heidelberg, 1975).
43. Wong, T. W. et al. Air pollution and hospital admissions for respiratory and cardiovascular diseases in Hong Kong. *Occup. Environ. Med.* **56**, 679–683 (1999).
44. Fan, J. & Zhang, W. Statistical estimation in varying coefficient models. *Ann. Stat.* **27**, 1491–518 (1999).
45. Wahltinez, O., Cheung, A., Alcantara, R. et al. COVID-19 Open-Data a global-scale spatially granular meta-dataset for coronavirus disease. *Sci Data* **9**, 162 (2022).
46. Zhang, B., Li, W., Shi, Y., Liu, X. & Chen, L. Detecting causality from short time-series data based on prediction of topologically equivalent attractors. *BMC Syst. Biol.* **11**, 128 (2017).
47. Ma, H., Leng, S. & Chen, L. Data-based prediction and causality inference of nonlinear dynamics. *Sci. China Math.* **61**, 403–420 (2018).
48. Butler, K., Feng, G. & Djurić, P. M. On Causal Discovery With Convergent Cross Mapping. *IEEE Trans. Signal Process.* **71**, 2595–2607 (2023).
49. Gao, B. et al. Causal inference from cross-sectional earth system data with geographical convergent cross mapping. *Nat. Commun.* **14**, 5875 (2023).
50. Runge, J. et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**, 2553 (2019).
51. Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. Takens embedding theorems for forced and stochastic systems. *Nonlinear Anal. Theory Methods Appl.* **30**, 5303–5314 (1997).
52. Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. Delay Embeddings for Forced Systems. II. Stochastic Forcing. *J. Nonlinear Sci.* **13**, 519–577 (2003).
53. Ma, H., Leng, S., Aihara, K., Lin, W. & Chen, L. Randomly distributed embedding making short-term high-dimensional data predictable. *Proc. Natl. Acad. Sci.* **115**, E9994–E10002 (2018).
54. Bai, S., Kolter, J. Z. & Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv preprint arXiv:1803.01271* (2018).
55. Berrone, S., Della Santa, F., Mastropietro, A., Pieraccini, S. & Vaccaro, F. Layer-wise relevance propagation for backbone identification in discrete fracture networks. *J. Comput. Sci.* **55**, 101458 (2021).
56. Sun, J., Zhou, S. & Veeramani, D. A neural network-based control chart for monitoring and interpreting autocorrelated multivariate processes using layer-wise relevance propagation. *Qual. Eng.* **35**, 33–47 (2023).
57. Kim, D. et al. Untangling the contribution of input parameters to an artificial intelligence PM2.5 forecast model using the layer-wise relevance propagation method. *Atmos. Environ.* **276**, 119034 (2022).
58. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
59. Yang, L., Lin, W. & Leng, S. Conditional cross-map-based technique: From pairwise dynamical causality to causal network reconstruction. *Chaos Interdiscip. J. Nonlinear Sci.* **33**, 063101 (2023).
60. Ma, H., Haluszczynski, A., Prosperino, D. & R  th, C. Identifying causality drivers and deriving governing equations of nonlinear complex systems. *Chaos Interdiscip. J. Nonlinear Sci.* **32**, 103128 (2022).
61. Krakovsk  , A., Jakub  k, J., Bud  cov  , H. & Holecov  , M. Causality studied in reconstructed state space. Examples of uni-directionally connected chaotic systems. *arXiv preprint arXiv:1511.00505* (2015).
62. Li, X. et al. Higher-order Granger reservoir computing: simultaneously achieving scalable complex structures inference and accurate dynamics prediction. *Nat. Commun.* **15**, 2506 (2024).
63. Marbach, D., Schaffter, T., Mattiussi, C. & Floreano, D. Generating Realistic In Silico Gene Networks for Performance Assessment of Reverse Engineering Methods. *J. Comput. Biol.* **16**, 229–239 (2009).
64. Stolovitzky, G., Prill, R. J. & Califano, A. Lessons from the DREAM2 Challenges: A Community Effort to Assess Biological Network Inference. *Ann. N. Y. Acad. Sci.* **1158**, 159–195 (2009).

65. Stolovitzky, G., Monroe, D. & Califano, A. Dialogue on Reverse-Engineering Assessment and Methods: The DREAM of High-Throughput Pathway Inference. *Ann. N. Y. Acad. Sci.* **1115**, 1–22 (2007).
66. Schaffter, T., Marbach, D. & Floreano, D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27**, 2263–2270 (2011).
67. Xia, Y. & Härdle, W. Semi-parametric estimation of partially linear single-index models. *J. Multivar. Anal.* **97**, 1162–1184 (2006).
68. Devlin, R. B. et al. Controlled Exposure of Healthy Young Volunteers to Ozone Causes Cardiovascular Effects. *Circulation* **126**, 104–111 (2012).
69. Lee, B.-J., Kim, B. & Lee, K. Air Pollution Exposure and Cardiovascular Disease. *Toxicol. Res.* **30**, 71–75 (2014).
70. Luo, K. et al. Acute Effects of Nitrogen Dioxide on Cardiovascular Mortality in Beijing: An Exploration of Spatial Heterogeneity and the District-specific Predictors. *Sci. Rep.* **6**, 38328 (2016).
71. Liu, R. et al. Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy. *Sci. Bull.* **66**, 2265–2270 (2021).
72. Zhao, J. et al. Detecting dynamical causality via intervened reservoir computing. *Commun. Phys.* **7**, 232 (2024).
73. Schölkopf, B. et al. Towards causal representation learning. *Proc. IEEE* **109**, 612–634 (2021).
74. Ahuja, K., Mahajan, D., Wang, Y. & Bengio, Y. *Interventional Causal Representation Learning*, vol. 202, 372–407 (PMLR, 2023).
75. Zhang, K., Xie, S., Ng, I. & Zheng, Y. *Causal representation learning from multiple distributions: A General Setting*, vol. 235, 60057–60075 (PMLR, 2024).
76. Mitrovic, J., McWilliams, B., Walker, J. C., Buesing, L. H. & Blundell, C. Representation Learning via Invariant Causal Mechanisms. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021* (OpenReview.net, 2021).

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Nos. 42450084, T2341022, 12322119, and 12271180), Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011797). We thank Dr. Jifan Shi for his insightful advice and encouragement during this study.

Author contributions

S. Cai, H. Peng, R. Liu, and P. Chen jointly conceived and conceptualized the project. S. Cai conducted the experimental research and performed the

analysis. H. Peng, R. Liu, and P. Chen prepared the data and reviewed the related literature. S. Cai, H. Peng, R. Liu, and P. Chen prepared the figures and wrote the manuscript. R. Liu and P. Chen supervised the research. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-025-02170-6>.

Correspondence and requests for materials should be addressed to Rui Liu or Pei Chen.

Peer review information *Communications Physics* thanks Jonas Wahl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025