Deep Koopman operator framework for causal discovery in nonlinear dynamical systems

Juan Nathaniel [*1,2] **Carla Roesch** [*1,2] **Jatan Buch** [1,2] **Derek DeSantis** [3] **Adam Rupe** [4] **Kara Lamb** [1,2] **Pierre Gentine** [1,2]

Abstract

We use a deep **K**oopman operator-theoretic formalism to develop a novel c**ausal** discovery algorithm, `Kausal`. Standard statistical frameworks, such as Granger causality, lack the ability to quantify causal relationships in nonlinear dynamics due to the presence of complex feedback mechanisms, timescale mixing, and nonstationarity. In `Kausal`, we propose to leverage Koopman operators for causal analysis where optimal observables are inferred using deep learning. Our numerical experiments with toy models and real-world phenomena such as El Niño-Southern Oscillation demonstrate `Kausal`'s superior ability in discovering and characterizing causal signals compared to existing approaches. Code is publicly available at `https://github.com/juannat7/kausal`.

1. Introduction

Causal discovery seeks to disentangle cause-effect mechanisms for enhanced scientific understanding, leading to improved modeling of the physical world (Aloisi et al., 2022; Cattry et al., 2025; Roesch et al., 2025; Williams et al., 2025). Since performing interventions to detect causality is infeasible in many domains, there is a great interest to develop observation-driven causal discovery methods (e.g., Granger, 1969; Hyvärinen et al., 2010; Runge, 2020). However, existing methods tend to operate under certain assumptions, such as time-invariance, stationarity, and linearity (Granger, 1969; Runge et al., 2019; Bareinboim et al., 2022; Camps-Valls et al., 2023), which limit their applicability.

[*]Equal contribution [1]Columbia University [2]LEAP NSF Science and Technology Center [3]Los Alamos National Laboratory [4]Pacific Northwest National Laboratory. Correspondence to: Juan Nathaniel <jn2808@columbia.edu>, Carla Roesch <cmr2293@columbia.edu>.

ICML 2025 Workshop on Assessing World Models.

In general, there are two categories of approaches for inferring causal relationships in nonlinear dynamical systems: i) using information flow to identify causal influence among variables and subspaces (Granger, 1969; Liang & Kleeman, 2005; Majda & Harlim, 2007; James et al., 2016; Tank et al., 2021; Rupe & Crutchfield, 2024); and ii) performing a lifting transformation to project the system of interest to a higher-dimensional space where causal effects can be investigated more easily (Lorenz, 1991; Sugihara et al., 2012; Gilpin, 2024; Butler et al., 2024). Thus, we pose our research question: Can we combine these two principles to apply linear causal analysis methods in nonlinear systems?

A natural foundation for us to start answering this question is Koopman theory (Koopman, 1931; Mezić, 2013; Brunton et al., 2022; 2025), which states that finite nonlinear dynamics can be represented with an infinite-dimensional linear operator acting on the space of all possible measurement functions or observables. Recent work by Rupe et al. (2024) has established a theoretical foundation linking Koopman theory to causal analysis for dynamical systems using prescribed embeddings. Simultaneously, efficiently and effectively estimating optimal embeddings has been enabled by rapid advances of deep learning in high-dimensional settings (Brunton et al., 2016; Lusch et al., 2018; Wang et al., 2022). As such, data-driven Koopman operator methods have seen a surge in interest, especially for spatiotemporal analysis in a number of fields (Kutz et al., 2018; Azencot et al., 2020; Rice et al., 2021; Froyland et al., 2021; Lintner et al., 2023; Lamb et al., 2024; Nathaniel & Gentine, 2025).

Main contributions. In this paper, we leverage ideas from deep learning, Koopman theory, and causal inference. Our main contributions can be summarized as follows: (i) We introduce a novel deep **K**oopman operator-based c**ausal** discovery framework, `Kausal`, as illustrated in Figure 1. (ii) We show that `Kausal` infers a better representation of the basis functions through deep learning, as opposed to existing prescribed approaches, which leads to a more accurate and meaningful causal analysis in high-dimensional nonlinear systems. (iii) We successfully apply `Kausal` to both idealized models and real-world phenomena highlighting the algorithm's scalability and generalizability.

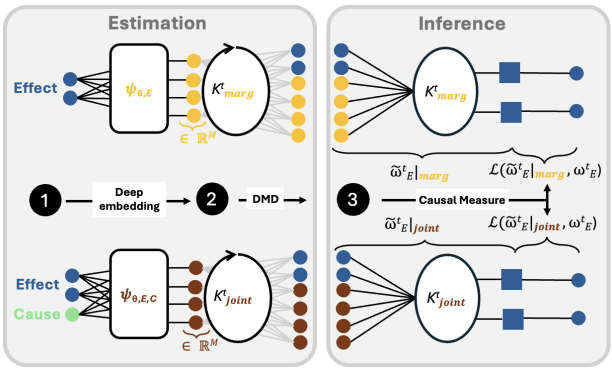


Figure 1. Schematic of the `Kausal` algorithm. We 1) estimate the embeddings using deep learning, and 2) approximate the Koopman operator with dynamic mode decomposition (DMD). Then, 3) we infer causal measures by computing the difference in prediction error between marginal (effect-only) and joint (effect-cause) models.

2. `Kausal`

In this section, we introduce the background for `Kausal` based on the theoretical foundation from Rupe et al. (2024). A more detailed explanation of the framework is given in Appendix A.1 and A.2.

Koopman operator. Koopman theory provides a *global linearization* of high-dimensional nonlinear dynamical systems (Koopman, 1931). Specifically, the nonlinear temporal evolution of system states is represented as a linear flow in the evolution of observables $\boldsymbol{\psi} \in \Psi$ (more details in Appendix A.2). The Koopman operator is then given by:

$$K^t := \Psi_\theta^t \Psi_\theta^\dagger \tag{1}$$

where $\Psi_\theta^\dagger$ is the pseudo-inverse of Ψ_θ. A low-rank approximation is also feasible with matrix projection onto dominant modes using e.g., SVD. Note that K^t is an approximation and converges to the Koopman operator $\mathcal{K}^t$ for $D, M \to \infty$ (Korda & Mezić, 2018).

Structural dynamical causal models. In classical terms, causal mechanisms are described through *structural causal models* (SCM), where a system of d random variables $\boldsymbol{x} = \{x^1, \ldots, x^d\}$ are expressed as an arbitrary function f^k of their direct parents (causes) $\boldsymbol{x}_{\mathrm{PA}_k}$ and an exogenous distribution of noise ϵ^k. For dynamical systems, we can extend SCM for a collection of d ODEs to define *structural dynamical causal models* (SDCM) (Rubenstein et al., 2018; Peters et al., 2022; Boeken & Mooij, 2024):

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega_{k,t} := f^k(\boldsymbol{\omega}_{\mathrm{PA}_{k,t}}, \epsilon^k), \text{ with } \omega_{k,0} = \omega_k(0) \tag{2}$$

where $k \in \{1, \ldots, d\}$.

Causal mechanism. In dynamical systems, the time-evolution of states is defined through a set of functions called flow maps $\{\Phi^t : \Omega \to \Omega\}_{t \in T}$ that carries the phase space back onto itself. For an initial state $\omega \in \Omega$, flow maps evolve ω through time t, given by $\omega(t) = \Phi^t(\omega)$. To define a causal mechanism, we partition the phase space Ω of a system into components:

$$\Omega = \Omega_E \times \Omega_C \times \Omega_R \tag{3}$$

representing the effect, cause, and remainder components.

Causal influence. The existence of a dynamical causal relationship at time t is determined by whether the state of the effect component depends on the initial state of the cause component, i.e., Ω_C *dynamically causally influences* Ω_E at time t. Mathematically we write this as:

$$\Omega_C \to^t \Omega_E \iff \omega_E(t) \text{ is influenced by } \omega_C \text{ and} \tag{4a}$$

$$\Omega_C \not\to^t \Omega_E \iff \omega_E(t) \text{ is } \textit{not} \text{ influenced by } \omega_C, \tag{4b}$$

where Equation 4b implies Ω_C does *not dynamically causally influence* Ω_E.

Koopman causal influence. We can rephrase dynamical causality using the Koopman framework, where it is represented by the linear evolution of observables in reproducing kernel Hilbert space, $\boldsymbol{\psi}_\theta \in \mathcal{F}$. We argue that causal relationships in the phase space Ω also induce distinguishable patterns in the flow of functions. For instance, if $\Omega_C \to^t \Omega_E$, then the application of the Koopman operator on the effect observables $\mathcal{F}_E$ will inherit this dependence which is expressed as $K^t \boldsymbol{\psi}_\theta \in \mathcal{F}_{E,C,R}$. We define this as Ω_C *Koopman causally influences* Ω_E which is denoted by $\Omega_C \to_K^t \Omega_E$. Correspondingly, if Ω_C *does not Koopman causally influence* Ω_E at time t, it is denoted by $\Omega_C \not\to_K^t \Omega_E$ where $K^t \boldsymbol{\psi}_\theta \in \mathcal{F}_{E,R}$. In summary, dynamical causality and Koopman causality at time t are equivalent. Namely, $\Omega_C \to^t \Omega_E$ if and only if $\Omega_C \to_K^t \Omega_E$. For a formal proof, refer to Theorem III.2 in Rupe et al. (2024).

Marginal and joint models. The data-driven measure of Koopman causality leverages the dynamic mode decomposition (DMD) algorithm. Our method involves fitting two distinct DMD models to evolve functions in $\mathcal{F}_E$: i) a *marginal model* restricted to functions from the effect subspace; and ii) a *joint model* which utilizes functions from the subspace incorporating both the effect and cause components.

We define the marginal and joint models, where the addition of identity observables on Ω_E (Brunton et al., 2016), $\Psi_{\mathrm{ID},E} := \boldsymbol{\psi}_{\mathrm{ID}}(\Omega_E) = \Omega_E$, empirically improves the performance of our algorithm:

$$K^t_{\mathrm{marg}} = \Psi^t_{\mathrm{ID},E} \begin{bmatrix} \Psi_{\mathrm{ID},E} \\ \Psi_{\theta,E} \end{bmatrix}^\dagger ; \quad K^t_{\mathrm{joint}} = \Psi^t_{\mathrm{ID},E} \begin{bmatrix} \Psi_{\mathrm{ID},E} \\ \Psi_{\theta,E,C} \end{bmatrix}^\dagger \tag{5}$$

where the matrices $\Psi_{\theta,E}$ and $\Psi_{\theta,E,C}$ are formed from applying the marginal $\boldsymbol{\psi}_{\theta,E}(\omega_E)$ and joint $\boldsymbol{\psi}_{\theta,E,C}(\omega_{E,C})$ on the `training` data in Ω_E and $\Omega_E \times \Omega_C$, respectively.

Causal measure. The *causal measure* $\Delta^{K^t}_{\mathrm{C,E}}$ is defined as,

$$\begin{aligned}\Delta^{K^t}_{\mathrm{C,E}} &:= \Omega_C \xrightarrow[K^t]{} \Omega_E \\ &:= \mathcal{L}(\tilde{\omega}^t_E|_{\mathrm{marg}}, \omega^t_E) - \mathcal{L}(\tilde{\omega}^t_E|_{\mathrm{joint}}, \omega^t_E),\end{aligned} \tag{6}$$

where the difference in errors between the marginal and joint models is calculated over all data points $\{\omega_E, \omega^t_E, \omega_C\}$:

$$\mathcal{L}(\tilde{\omega}^t_E, \omega^t_E) = \frac{1}{N}\sum_N \left\|\tilde{\omega^t_E} - \omega^t_E\right\|^2. \tag{7}$$

Hence, if the joint model has a significantly lower error, it suggests that Ω_C contains causal information about Ω_E. This also implies that if $\Delta^{K^t}_{\mathrm{C,E}} \to 0$, there is minimal causal influence from Ω_C on Ω_E at time t.

Conditional forecasting. Computationally, estimating $\Delta^{K^t}_{\mathrm{C,E}}$ at each time shift t is expensive. Thus, we perform conditional forecasting, which preserves the structure of marginal and joint models for reasonable t. Feeding the predicted outputs $\tilde{\omega}_E(t)$ back as identity observables $\psi_{\mathrm{ID},E}(\tilde{\omega}_E(t))$ enables efficient exploration of causal dynamics over time. However, it is not true forecasting because we supply data from the `test` dataset to construct the non-identity observables. Specifically, starting with an initial value of $\omega_E(t_0)$ from the `test` data we can define the first and following conditional predictions for the marginal model as:

$$\underbrace{K^t_{\mathrm{marg}}\begin{bmatrix}\psi_{\mathrm{ID},E}(\omega_E(t_0))\\ \psi_{\theta,E}(\omega_E(t_0))\end{bmatrix}}_{:=\tilde{w}^t_E|_{\mathrm{marg}}(t_1)} \to \underbrace{K^t_{\mathrm{marg}}\begin{bmatrix}\psi_{\mathrm{ID},E}(\tilde{\omega}_E(t_1))\\ \psi_{\theta,E}(\omega_E(t_1))\end{bmatrix}}_{:=\tilde{w}^t_E|_{\mathrm{marg}}(t_2)} \to \cdots \tag{8}$$

Given an additional initial value for $\omega_C(t_0)$ we follow the same procedure with necessary adjustments to the dictionary functions to obtain conditional forecasts for the joint model.

3. Experiments

3.1. Toy examples

We perform a range of experiments, including on high-dimensional reaction-diffusion equation, to assess `Kausal`'s ability for causal discovery (see Appendix D). In the following, we show results for the coupled Rössler oscillators, a 6-dimensional system $\Omega \in \mathbb{R}^6$. We estimate causal measures as defined in Equation 7 where the measure for each unique time shift $t \in \mathbb{Z}_+$ is estimated across N data points. We use multilayer perceptrons (MLP) as our deep estimators for the observables, represented as a $M = 32$-dimensional feature vector of the encoder final layer's output. As illustrated in Figure 2, `Kausal` is able to evaluate statistically significant causal signals in the true direction ($\Delta^{K^t}_{C,E}$) when compared with the non-causal case ($\Delta^{K^t}_{E,C}$). For completeness, we also plot the causal measure

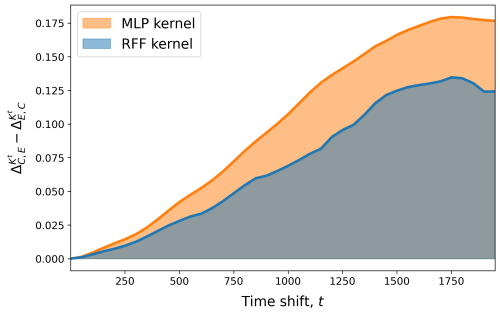
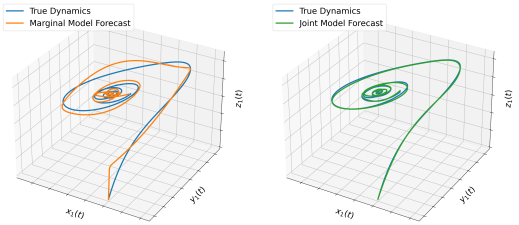


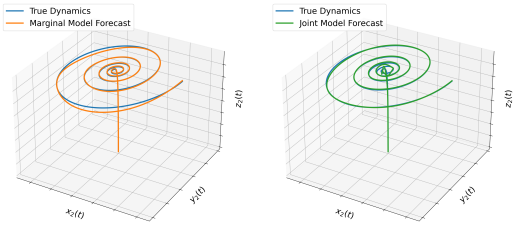
Figure 2. Causal measure estimation of coupled Rössler oscillators by computing the difference between the true causal ($\Delta^{K^t}_{C,E}$) and non-causal direction ($\Delta^{K^t}_{E,C}$) across time shifts t, using different kernels to approximate the observables.

using a random Fourier features (RFF) kernel ($M = 500$; Rupe et al., 2024; Rahimi & Recht, 2007).

Performing conditional forecasting. We also illustrate the concept of conditional forecasting formalized earlier in Equation 8, starting with ω_0. As shown in Figure 3, the inclusion of Ω_C (i.e., the joint model) is crucial in capturing Ω_E dynamics in the *true* causal direction, whereas its exclusion (i.e., the marginal model) leads to significant deviations. Conversely, in the non-causal direction, both the marginal and joint models make no perceptible difference in forecasting Ω_C as $\Omega_E \not\to^t_K \Omega_C$.



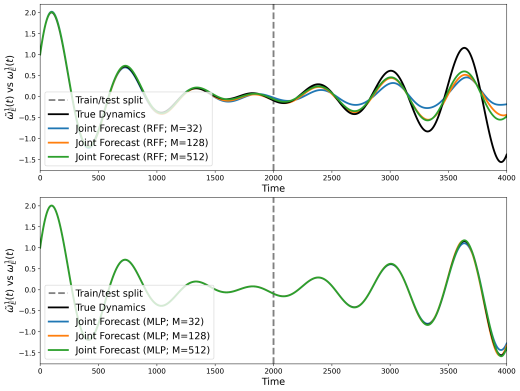
(a) Conditional forecasts in the *true* causal direction $\Omega_C \to^t_K \Omega_E$



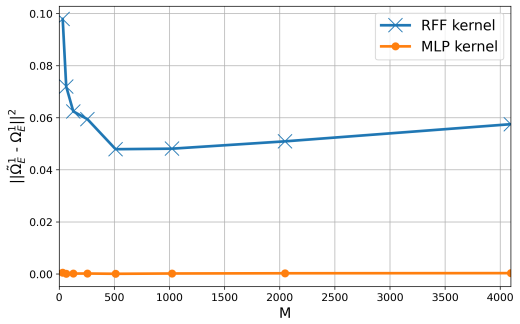
(b) Conditional forecasts in the non-causal direction $\Omega_E \to^t_K \Omega_C$

Figure 3. Conditional forecasts in the (a) *true* and (b) non-causal direction using MLP kernels. In (a), the addition of Ω_C in the joint model improves the forecast of Ω_E relative to the marginal model that excludes it. In (b), however, both marginal and joint models make no qualitative difference as $\Omega_E \not\to^t_K \Omega_C$.

Optimal kernel selection. Robust estimation of observables is crucial for the accurate representation of nonlinear dynamics in $\mathcal{F}$. In order to study this notion, we perform joint forecasts using either a prescribed $\boldsymbol{\psi}_{\text{RFF}}$ or a learnable dictionary of functions, $\boldsymbol{\psi}_\theta$, for varying dimensionality M. As illustrated in Figure 4, joint forecasts utilizing $\boldsymbol{\psi}_\theta$ reproduce the true dynamics more accurately (expected behavior), even beyond the `training` set and for low M. In contrast, joint forecasts using $\boldsymbol{\psi}_{\text{RFF}}$, even with high M, struggle to capture the underlying dynamics, underscoring the limitations of prescribing dictionary functions rather than directly learning them from data.



(a) Ablating optimal observables by performing conditional joint forecasts (for "fixed" $K^{t=1}$) in the $\Omega_C \to_K^t \Omega_E$ direction ($\omega^1(t) := x_1(t) \in \Omega_E$) using different estimators of varying M.



(b) Dimensionality scaling between prescribed $\boldsymbol{\psi}_{\text{RFF}}$ and learnable $\boldsymbol{\psi}_{\text{MLP}}$ comparing the conditional forecasts' MSE across all times.

Figure 4. Performance assessment of MLP and RFF kernels for varying dimensionality (i.e. increasing complexity).

3.2. World model: observed ENSO

Next, we illustrate `Kausal`'s generalization to real-world phenomena using an example from climate science, El Niño–Southern Oscillation (ENSO; Wang & Picaut, 2013). Since ENSO has major societal impacts, understanding its causal drivers is vital for explainable decision making in sectors such as agriculture and disaster preparedness (Glantz et al., 1987; Callahan & Mankin, 2023). We demonstrate the feasibility of `Kausal` in studying ENSO dynamics using ocean reanalysis products from the European Centre for Medium-Range Weather Forecasts (ECMWF; Zuo et al., 2019). Preliminary results as illustrated in Figure 5 are promising. We find that `Kausal` is able to track the emergence of ENSO events as represented by $\Delta_{T,h}^{K^t} > 0$. Specifically, warm-phase El Niño events (red shading in Figure 5) result from an increase in T, weakening the Walker circulation and leading to a decrease in h. Cold-phase La Niña events follow a similar pattern but in the opposite direction (blue shading in Figure 5). See Appendix D.3 for additional details as well as results from a toy ENSO model.

3.3. Additional baselines

In the following, we compare `Kausal` to a range of baseline causal discovery methods across two tasks: (i) causal direction identification, and (ii) causal magnitude estimation. Specifically, we compare against two well-established methods: VARLiGAM (Hyvärinen et al., 2010; Shimizu et al., 2006) and PCMCI+ (Peter Clark Momentary Conditional Independence; Runge, 2020). Further, we evaluate against novel deep learning-based causal discovery methods: cLSTM (Tank et al., 2021) and TSCI (Tangent Space Causal Inference; Butler et al., 2024). See Appendix E for more implementation details.

Table 1. Causal direction identification for (a) Coupled Rössler Oscillators, (b) reaction-diffusion equation. Entries denote the p-value (ρ) for the true (C $\to$ E; $\downarrow$ ρ is better) and false causal direction (E $\to$ C; $\uparrow$ ρ is better).

(a) Coupled Rössler oscillators

METHOD	$\rho_{\text{C}\to\text{E}}$	$\rho_{\text{E}\to\text{C}}$
PCMCI+	0.44	0.39
VARLiNGAM	1.22×10^{-5}	0.00
KAUSAL$_{\text{RFF}}$	**0.01**	**1.00**
KAUSAL$_{\text{MLP}}$	**0.01**	**1.00**

(b) Reaction-diffusion equation

METHOD	$\rho_{\text{C}\to\text{E}}$	$\rho_{\text{E}\to\text{C}}$
PCMCI+	0.74	1.95×10^{-13}
VARLiNGAM	0.00	0.00
KAUSAL$_{\text{RFF}}$	0.03	0.03
KAUSAL$_{\text{CNN}}$	**0.03**	**0.23**

Task 1: Identify causal direction. We first conduct hypothesis testing (see Algorithm 1) to identify the causal direction for the coupled Rössler oscillators (Table 1a) and the reaction-diffusion equation (Table 1b) for which the true directionality is known and clearly distinguishable (more details in Appendix E.1). Because none of the deep learning-

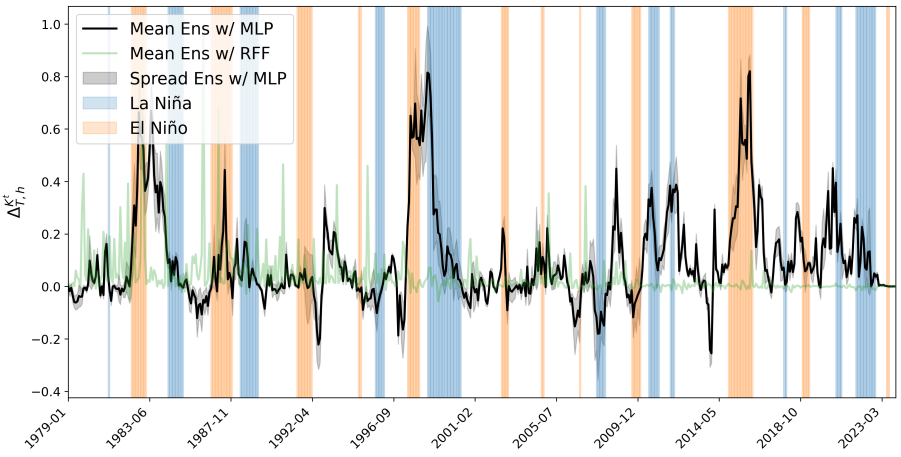


Figure 5. Causal measure of ENSO using real-world data showcasing `Kausal`'s ability (black line) to capture major El Niño (red shading) and La Niña (blue shading) events. Events are estimated using NOAA's Oceanic Niño Index (Glantz & Ramirez, 2020). The causal measure estimated using a prescribed RFF kernel (green line) fails to capture ENSO dynamics. Ensemble is obtained from fitting 5 different randomly initialized kernel.

based methods provide out-of-the-box statistical tests to validate causal direction, we conduct this task with VARLiNGAM and PCMCI+. Our results indicate that PCMCI+ tends to be conservative, often yielding no directionality when there is. In contrast, VARLiNGAM frequently conflates directional cues, inferring bi-directional relationships when there is none. `Kausal`, however, captures the true (and the lack of) causal relationships with greater precision.

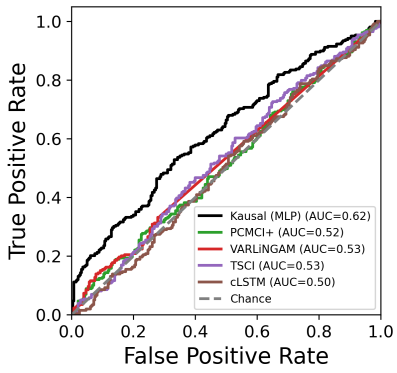


Figure 6. AUROC scores for the observed ENSO (AUROC > 0.5 shows significant skills better than chance).

Task 2: Evaluate causal magnitude. Next, we evaluate `Kausal` for causal magnitude estimation. This analysis extends the observed ENSO case (Figure 5) but added a summarized quantitative evaluation in terms of Area Under the Receiver Operating Characteristic (AUROC; Peterson et al., 1954) scores (Figure 6). Since most, if not all existing methods still assume a static causal graph with fixed link magnitudes, we employ the sliding-window protocol introduced by Runge et al. (2019) (see Appendix E.2 for details and results). Beyond the ENSO case, we further apply a similar setup to predict causal magnitudes in other toy problems. For instance, Figures S9 - S11 illustrate the causal magnitude timeseries given a backdrop of extremes to be detected. While Figures S12-S14 highlight their AUROC scores. Overall, `Kausal` consistently achieves the highest score. The Granger causality-based approaches (PCMCI+, cLSTM and VARLiNGAM) perform worse because their fixed time-lagged embeddings cannot capture nonlinearities and multiscale interactions outside the chosen window. Although TSCI leverages cross-correlation maps in an operator-theoretic approach, it still struggles due to the absence of explicit linearity constraints.

4. Conclusion

This paper presents `Kausal`, a novel deep Koopman operator-based algorithm for causal discovery in high-dimensional nonlinear systems, combining the principles of deep learning, Koopman theory, and causal inference. Leveraging deep learning to infer the dictionary functions of observables, our algorithm improves the estimation of causal measures in nonlinear dynamics relative to state-of-the-art methods. `Kausal`'s scalability and generalizability is further highlighted by evaluating its performance for a range of systems of increasing complexity, including coupled Rössler oscillators as well as sea-surface temperature and height observations of ENSO. There are several promising directions for future work: extending the framework to causal effect estimation allowing us to generate explainable physics-informed predictions, discovering causal graphs by developing tractable adjacency testing algorithms, and accounting for stochastic, non-autonomous dynamics.

References

Abarbanel, H. D., Carroll, T., Pecora, L., Sidorowich, J., and Tsimring, L. S. Predicting physical variables in time-delay embedding. *Physical Review E*, 49(3):1840, 1994.

Aloisi, V., Gatto, A., Accarino, G., Donato, F., and Aloisio, G. The effect of known and unknown confounders on the relationship between air pollution and Covid-19 mortality in Italy: A sensitivity analysis of an ecological study based on the E-value. *Environmental research*, 207: 112131, 2022.

Aronszajn, N. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

Azencot, O., Erichson, N. B., Lin, V., and Mahoney, M. W. Forecasting sequential data using consistent Koopman autoencoders. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *ICML'20*, pp. 475–485. JMLR.org, July 2020.

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of Judea Pearl*, pp. 507–556. 2022.

Boeken, P. and Mooij, J. M. Dynamic structural causal models. *arXiv preprint arXiv:2406.01161*, 2024.

Brunton, S. L., Brunton, B. W., Proctor, J. L., and Kutz, J. N. Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control. *PLOS ONE*, 11(2):e0150171, February 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0150171. URL https://dx.plos.org/10.1371/journal.pone.0150171.

Brunton, S. L., Budišić, M., Kaiser, E., and Kutz, J. N. Modern Koopman Theory for Dynamical Systems. *SIAM Review*, 64(2):229–340, May 2022. ISSN 0036-1445. doi: 10.1137/21M1401243. URL https://epubs.siam.org/doi/10.1137/21M1401243.

Brunton, S. L., Zolman, N., Kutz, J. N., and Fasel, U. Machine Learning for Sparse Nonlinear Modeling and Control. *Annual Review of Control, Robotics, and Autonomous Systems*, January 2025. ISSN 2573-5144. doi: 10.1146/annurev-control-030123-015238. URL https://www.annualreviews.org/content/journals/10.1146/annurev-control-030123-015238.

Butler, K., Waxman, D., and Djuric, P. Tangent space causal inference: Leveraging vector fields for causal discovery in dynamical systems. *Advances in Neural Information Processing Systems*, 37:120078–120102, 2024.

Callahan, C. W. and Mankin, J. S. Persistent effect of El Niño on global economic growth. *Science*, 380(6649):1064–1069, June 2023. doi: 10.1126/science.adf2983. URL https://www.science.org/doi/full/10.1126/science.adf2983.

Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., Vinuesa, R., Diaz, E., Zanna, L., and Runge, J. Discovering causal relations and equations from data. *Physics Reports*, 1044:1–68, December 2023. ISSN 0370-1573. doi: 10.1016/j.physrep.2023.10.005. URL https://www.sciencedirect.com/science/article/pii/S0370157323003411.

Caponnetto, A. and De Vito, E. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7:331–368, 2007.

Cattry, M., Zhao, W., Nathaniel, J., Qiu, J., Zhang, Y., and Gentine, P. EcoPro-LSTMv0: A Memory-based Machine Learning Approach to Predicting Ecosystem Dynamics across Time Scales in Mediterranean Environments. *EGUsphere*, 2025:1–37, 2025.

Christiansen, R., Baumann, M., Kuemmerle, T., Mahecha, M. D., and Peters, J. Toward causal inference for spatio-temporal data: Conflict and forest loss in colombia. *Journal of the American Statistical Association*, 117(538): 591–601, 2022.

Froyland, G., Giannakis, D., Lintner, B. R., Pike, M., and Slawinska, J. Spectral analysis of climate dynamics with operator-theoretic approaches. *Nature Communications*, 12(1):6570, November 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-26357-x. URL https://www.nature.com/articles/s41467-021-26357-x.

Gilpin, W. Generative learning for nonlinear dynamics. *Nature Reviews Physics*, 6(3):194–206, March 2024. ISSN 2522-5820. doi: 10.1038/s42254-024-00688-2. URL https://www.nature.com/articles/s42254-024-00688-2.

Glantz, M. H. and Ramirez, I. J. Reviewing the oceanic Niño Index (ONI) to enhance societal readiness for El Niño's impacts. *International Journal of Disaster Risk Science*, 11:394–403, 2020.

Glantz, M. H., Katz, R., and Krentz, Maria and, N. C. f. A. R. The societal impacts associated with the 1982-83 worldwide climate anomalies /: edited by Michael Glantz, Richard Katz, Maria Krenz. National Center for Atmospheric Research,, 1987. URL https://digitallibrary.un.org/record/84611.

Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(5), 2010.

James, R. G., Barnett, N., and Crutchfield, J. P. Information Flows? A Critique of Transfer Entropies. *Physical Review Letters*, 116(23):238701, June 2016. doi: 10.1103/PhysRevLett.116.238701. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.238701.

Jin, F.-F. An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model. *Journal of the Atmospheric Sciences*, 54(7):811–829, April 1997. ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1997)054⟨0811:AEORPF⟩2.0.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.

Koopman, B. O. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

Korda, M. and Mezić, I. On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator. *Journal of Nonlinear Science*, 28(2):687–710, April 2018. ISSN 1432-1467. doi: 10.1007/s00332-017-9423-0. URL https://doi.org/10.1007/s00332-017-9423-0.

Kostic, V., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., and Pontil, M. Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces. *Advances in Neural Information Processing Systems*, 35: 4017–4031, 2022.

Kutz, J. N., Proctor, J. L., and Brunton, S. L. Applied koopman theory for partial differential equations and data-driven modeling of spatio-temporal systems. *Complexity*, 2018(1):6010634, 2018.

Lamb, K. D., van Lier-Walqui, M., Santos, S., and Morrison, H. Reduced-Order Modeling for Linearized Representations of Microphysical Process Rates. *Journal of Advances in Modeling Earth Systems*, 16(7):e2023MS003918, 2024. ISSN 1942-2466. doi: 10.1029/2023MS003918. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003918.

Liang, X. S. and Kleeman, R. Information Transfer between Dynamical System Components. *Physical Review Letters*, 95(24):244101, December 2005. doi: 10.1103/PhysRevLett.95.244101. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.244101.

Lintner, B. R., Giannakis, D., Pike, M., and Slawinska, J. Identification of the Madden–Julian Oscillation With Data-Driven Koopman Spectral Analysis. *Geophysical Research Letters*, 50(10):e2023GL102743, 2023. ISSN 1944-8007. doi: 10.1029/2023GL102743. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2023GL102743.

Lorenz, E. N. Dimension of weather and climate attractors. *Nature*, 353(6341):241–244, 1991.

Lusch, B., Kutz, J. N., and Brunton, S. L. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, November 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-07210-0. URL https://www.nature.com/articles/s41467-018-07210-0.

Majda, A. J. and Harlim, J. Information flow between subspaces of complex dynamical systems. *Proceedings of the National Academy of Sciences*, 104 (23):9558–9563, June 2007. doi: 10.1073/pnas.0703499104. URL https://www.pnas.org/doi/10.1073/pnas.0703499104.

Mezić, I. Analysis of Fluid Flows via Spectral Properties of the Koopman Operator. *Annual Review of Fluid Mechanics*, 45(Volume 45, 2013):357–378, January 2013. ISSN 0066-4189, 1545-4479. doi: 10.1146/annurev-fluid-011212-140652. URL https://www.annualreviews.org/content/journals/10.1146/annurev-fluid-011212-140652.

Nathaniel, J. and Gentine, P. Generative emulation of chaotic dynamics with coherent prior. *arXiv preprint arXiv:2504.14264*, 2025.

Nathaniel, J., Qu, Y., Nguyen, T., Yu, S., Busecke, J., Grover, A., and Gentine, P. Chaosbench: A multi-channel, physics-based benchmark for subseasonal-to-seasonal climate prediction. *arXiv preprint arXiv:2402.00712*, 2024.

Peters, J., Bauer, S., and Pfister, N. Causal models for dynamical systems. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pp. 671–690, 2022.

Peterson, W., Birdsall, T., and Fox, W. The theory of signal detectability. *Transactions of the IRE professional group on information theory*, 4(4):171–212, 1954.

Rahimi, A. and Recht, B. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

Rice, J., Xu, W., and August, A. Analyzing Koopman approaches to physics-informed machine learning for long-term sea-surface temperature forecasting, February 2021. URL http://arxiv.org/abs/2010.00399.

Roesch, C. M., Fons, E., Ballinger, A. P., Runge, J., and Hegerl, G. C. Decreasing aerosols increase the european summer diurnal temperature range. *npj Climate and Atmospheric Science*, 8(1):47, 2025.

Rubenstein, P. K., Bongers, S., Schoelkopf, B., and Mooij, J. M. From deterministic odes to dynamic structural causal models, 2018. URL https://arxiv.org/abs/1608.08028.

Runge, J. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on uncertainty in artificial intelligence*, pp. 1388–1397. Pmlr, 2020.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):2553, 2019.

Rupe, A. and Crutchfield, J. P. On principles of emergent organization. *Physics Reports*, 1071:1–47, 2024.

Rupe, A., DeSantis, D., Bakker, C., Kooloth, P., and Lu, J. Causal Discovery in Nonlinear Dynamical Systems using Koopman Operators, October 2024. URL http://arxiv.org/abs/2410.10103.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., and Jordan, M. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S. Detecting Causality in Complex Ecosystems. *Science*, 338(6106): 496–500, October 2012. doi: 10.1126/science.1227079. URL https://www.science.org/doi/10.1126/science.1227079.

Tank, A., Covert, I., Foti, N., Shojaie, A., and Fox, E. B. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2021.

Wang, C. and Picaut, J. Understanding ENSO Physics-A Review. In Wang, C., Xie, S., and Carton, J. (eds.), *Geophysical Monograph Series*, pp. 21–48. American Geophysical Union, Washington, D. C., March 2013. ISBN 978-1-118-66594-7 978-0-87590-412-2. doi: 10.1029/147GM02. URL https://onlinelibrary.wiley.com/doi/10.1029/147GM02.

Wang, R., Dong, Y., Arik, S. Ö., and Yu, R. Koopman neural forecaster for time series with temporal distribution shifts. *arXiv preprint arXiv:2210.03675*, 2022.

Williams, M., Milodowski, D. T., Smallman, T. L., Dexter, K. G., Hegerl, G. C., McNicol, I. M., O'Sullivan, M., Roesch, C. M., Ryan, C. M., Sitch, S., et al. Precipitation–fire functional interactions control biomass stocks and carbon exchanges across the world's largest savanna. *Biogeosciences*, 22(6):1597–1614, 2025.

Zhao, S., Jin, F.-F., Long, X., and Cane, M. A. On the breakdown of ENSO's relationship with thermocline depth in the central-equatorial pacific. *Geophysical Research Letters*, 48(9):e2020GL092335, 2021.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M. The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. *Ocean science*, 15(3):779–808, 2019.

Deep Koopman operator framework for causal discovery in nonlinear dynamical systems

Supplementary Material

A. Background

A.1. Dynamical systems

In this section we provide a brief overview of the necessary mathematical background and define the required notations to introduce our novel causal algorithm. The causal framework and notations follow Rupe et al. (2024).

Dynamical systems. Dynamical systems theory describes the evolution of a system over time, with the system state represented as a point in a phase space, $\omega \in \Omega$, which we assume to be in the Euclidean space $\mathbb{R}^N$. Throughout this paper variables are treated as vectors unless explicitly stated otherwise.

Flow maps. Time-evolution of states is defined through a set of functions called *flow maps* $\{\Phi^t : \Omega \to \Omega\}_{t\in T}$ that carries the phase space back onto itself. For an initial state $\omega \in \Omega$, flow maps evolve ω through time *t*, given by $\omega(t) = \Phi^t(\omega)$. The *orbit* of a state ω is the set $\{\omega(t) = \Phi^t(\omega)\}_{t\in T}$, representing the trajectory of the system over the time domain T. Although we denote the initial state as ω, writing $\omega(t, \omega_0) = \Phi^t(\omega_0)$ with the initial state ω_0 would be equivalent. For simplicity, we also restrict our analysis to *autonomous* dynamical systems characterized by time-independent flow maps.

Ordinary differential equations. In a deterministic setting, the evolution of $\omega(t)$ over time t can be described through *ordinary differential equations* (ODEs) of the form:

$$\frac{d\omega}{dt} = f(t, \omega) \tag{9}$$

where f is an arbitrary function and the solution $\omega(t)$ depends on the initial condition $\omega(t_0) = \omega_0$ at time t_0.

Components. To formalize the idea of causality in a dynamical system, we need to define different *components* of the system which can impact each other. For an N-dimensional phase space (i.e. N degrees of freedom) as described above, we partition the space into disjoint system component subspaces as $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_L$ with $\omega_i \in \Omega_i$ and $\omega = [\omega_1\ \omega_2\ \ldots\ \omega_L]$, where $L \leq N$.

Reproducing Kernel Hilbert Spaces (RKHS). Koopman operators act on functions rather than directly on system states, mapping one function to another by evolving it under the system dynamics. To ensure the operations of the Koopman operator are mathematically consistent and computationally feasible, we therefore require a well-defined function space. For data-driven analysis, a natural choice of a function space is reproducing kernel Hilbert spaces (RKHS).

A RKHS is a Hilbert space of functions $\mathcal{F}$ such that the evaluation $f \mapsto f(\omega)$ is continuous for each point $\omega \in \Omega$. This requires the existence of a kernel function $k : \Omega \times \Omega \to \mathbb{C}$ such that $k(\cdot, \omega') \in \mathcal{F}$ and $f(\omega') = \langle f, k(\cdot, \omega')\rangle$. The Moore-Aronszajn theorem (Aronszajn, 1950) implies that a positive definite kernel function defines a unique RKHS, so we use the notation $\mathcal{F} = \mathcal{H}(k)$ unambiguously. We will express the kernel k in terms of an associated feature map $\boldsymbol{\psi} : \Omega \to \mathcal{F}$, namely $k(\omega, \omega') = \langle \boldsymbol{\psi}(\omega), \boldsymbol{\psi}(\omega')\rangle_{\mathcal{H}}$. The kernel k is constructed as the tensor product of the kernels k_i, which are the kernel functions for each component Ω_i of the partitioned phase space $\Omega = \Omega_1 \times \cdots \times \Omega_L$. For a subspace Ω_X given $X \subset \{1, 2, ..., L\}$ we define the component observable function subspace $\mathcal{F}_X$ which is itself a RKHS. Thus, $\mathcal{F}_X = \mathcal{H}(k_X)$ is given through the kernel:

$$\begin{aligned} k_X(\omega, \omega') &:= \prod_{j\in X} k_j(\boldsymbol{\psi}(\omega_j), \boldsymbol{\psi}(\omega'_j)) \\ &:= \langle \boldsymbol{\psi}(\omega_j), \boldsymbol{\psi}(\omega'_j)\rangle_{\mathcal{H}} \end{aligned} \tag{10}$$

for all $\omega, \omega' \in \Omega$. We summarize key features of RKHS relevant to the Koopman framework in Table S1.

A.2. Koopman theory

Koopman operator. Koopman theory provides a *global linearization* of high-dimensional nonlinear dynamical systems (Koopman, 1931; Brunton et al., 2022). Specifically, the nonlinear temporal evolution of system states is represented as

a linear flow in the evolution of system observables. The map from system states to observables is given by a function, $\boldsymbol{\psi} : \Omega \to \mathcal{F}$.

The *Koopman operator* $\mathcal{K}^t : \mathcal{F} \to \mathcal{F}$ therefore acts as a time shift t on an observable $\boldsymbol{\psi}(\omega)$:

$$\boldsymbol{\psi}^t(\omega) = [\mathcal{K}^t \boldsymbol{\psi}](\omega) := [\boldsymbol{\psi} \circ \Phi^t](\omega) = \boldsymbol{\psi}(\omega(t)). \tag{11}$$

We need Koopman operators acting on $\mathcal{F}$ to be bounded to ensure the continuity of the time-evolution of $\boldsymbol{\psi}$, i.e. $\mathcal{K}^t\boldsymbol{\psi}$. This property is given by defining the Koopman operator over a reproducing kernel Hilbert space (RKHS) (Kostic et al., 2022; Rupe et al., 2024).

Observables. Obtaining observables in which the nonlinear dynamics appear approximately linear is an open challenge (Brunton et al., 2022). Existing works often rely on a set of M-dimensional dictionary functions $\boldsymbol{\psi} = [\psi_1\, \psi_2\, \cdots\, \psi_M]^\intercal$ spanning a finite-dimensional Hilbert space $\mathcal{F}_{\boldsymbol{\psi}} \subset \mathcal{F}$, such as Gaussian, and Random Fourier features (RFF; Rahimi & Recht, 2007), Polynomial, Time Delay (TDF; Abarbanel et al., 1994). Note that these dictionary functions are the previously defined feature maps. Often these functions may not be sufficiently expressive to represent complex dynamics (Kostic et al., 2022). Nonetheless, deep learning provides a natural framework for inferring the optimal high-dimensional embedding, which constitutes one of the goals of this paper.

In this paper, we compute a finite-rank approximation of the Koopman operator with dictionary functions $\boldsymbol{\psi}_\theta : \Omega \to \mathcal{H}$ that are learned using different neural network-based encoder-decoder architectures. Since we use appropriate regularizations and activations, namely the bounded sigmoid function, we ensure that the observables are finite. As a result, the kernel function is also bounded, i.e., $k_\theta(\omega, \omega') = \langle \boldsymbol{\psi}_{\boldsymbol{\theta}}(\omega), \boldsymbol{\psi}_{\boldsymbol{\theta}}(\omega')\rangle_{\mathcal{H}} < \infty$ for all $\omega, \omega' \in \Omega$ (Aronszajn, 1950; Caponnetto & De Vito, 2007; Kostic et al., 2022). The size of our encoder final layer's output vector $M \in \mathbb{Z}_+$ determines the dimensionality of $\mathcal{H}$, which is the number of learned dictionary functions.

Dynamic mode decomposition. To estimate the Koopman operator $\mathcal{K}^t$ we use *dynamic mode decomposition* (DMD), which solves a least-squares regression problem to identify a finite-dimensional linear representation of the system dynamics. Using dictionary functions defined as,

$$\boldsymbol{\psi}_{\boldsymbol{\theta}}(\omega(n))) = [\psi_{\theta,1}(\omega(n)) \ldots \psi_{\theta,M}(\omega(n))]^\intercal \in \mathbb{R}^{M\times 1}, \tag{12}$$

we construct data matrices consisting of observables for $D \in \mathbb{Z}_+$ pairs of data points $\{(\omega(n), \omega^t(n)\}_D$, denoted as Ψ_θ and Ψ_θ^t (t time shifted):

$$\Psi_\theta = \begin{bmatrix} | & & | \\ \boldsymbol{\psi}_\theta(\omega(n_1)) & \ldots & \boldsymbol{\psi}_\theta(\omega(n_D)) \\ | & & | \end{bmatrix}, \tag{13}$$

$$\Psi_\theta^t = \begin{bmatrix} | & & | \\ \boldsymbol{\psi}_\theta^t(\omega(n_1)) & \ldots & \boldsymbol{\psi}_\theta^t(\omega(n_D)) \\ | & & | \end{bmatrix}. \tag{14}$$

Assuming there exists a unique linear operator $K^t : \mathcal{F}_{\Psi_\theta} \to \mathcal{F}_{\Psi_\theta}$ that satisfies $\langle \boldsymbol{\psi}'_\theta, \mathcal{K}^t\boldsymbol{\psi}_\theta\rangle_{\mathcal{H}} = \langle \boldsymbol{\psi}'_\theta, K^t\boldsymbol{\psi}_\theta\rangle_{\mathcal{H}}$, where $\boldsymbol{\psi}_\theta, \boldsymbol{\psi}'_\theta \in \Psi_\theta$, we can define the least-squares regression problem as:

$$\begin{aligned} &\sum_{n=1}^{D} \left\| [\mathcal{K}^t\Psi_\theta](\omega(n)) - K^t\Psi_\theta(\omega(n)) \right\|^2 \\ &= \sum_{n=1}^{D} \left\| [\Psi_\theta \circ \Phi^t](\omega(n)) - K^t\Psi_\theta(\omega(n)) \right\|^2 \\ &= \sum_{n=1}^{D} \left\| \Psi_\theta(\omega^t(n)) - K^t\Psi_\theta(\omega(n)) \right\|^2. \end{aligned} \tag{15}$$

The full rank least-squares solution is given as:

$$K^t := \Psi_\theta^t \Psi_\theta^\dagger \tag{16}$$

where $\Psi_\theta^\dagger$ is the pseudo-inverse of Ψ_θ. A low-rank approximation is also feasible with matrix projection onto dominant modes using e.g., SVD. Note that K^t is an approximation and converges to the Koopman operator $\mathcal{K}^t$ for $D, M \to \infty$ (Korda & Mezić, 2018).

B. Relevant aspects of reproducing kernel Hilbert spaces

Table S1. Aspects of RKHS relevant to the Koopman Operator

Aspect	**Relevance to Koopman Operator**	**Mathematical Representation**
Function Space	Koopman operator maps functions (observables); RKHS provides structured function spaces.	$\mathcal{F} \subseteq \mathcal{H}, \quad \mathcal{K}^t : \mathcal{F} \to \mathcal{F}$
Discrete Data	Handles point evaluations of functions, key for real-world data-driven analysis.	$\boldsymbol{\psi}(\omega) = \langle \boldsymbol{\psi}, k(\cdot, \omega) \rangle$ (reproducing property)
Kernel Function	Efficiently constructs feature spaces for function representation.	$k(\omega, \omega') = \langle \boldsymbol{\psi}(\omega), \boldsymbol{\psi}(\omega') \rangle_{\mathcal{H}}$
Product Spaces	Represents multi-dimensional interactions between variables.	$\mathcal{H}_{\text{product}} = \mathcal{H}_E \otimes \mathcal{H}_C$
Theoretical Simplicity	Ensures boundedness, completeness, and well-defined operator properties.	$\|\mathcal{K}\boldsymbol{\psi}\| \leq C\|\boldsymbol{\psi}\|$ (bounded linear operator)
Data-Driven Approximations	Enables finite-dimensional approximations for computational feasibility.	$\mathcal{K}^t \approx K^t := \Psi^t \Psi^\dagger$ (this paper)

C. `Kausal`: A Differentiable Deep Koopman Causal Analysis Module

In this section, we highlight code snippets applying the `Kausal` algorithm.

Setting up. First, you can define multivariate timeseries and assign them as cause and effect variables by which a causal analysis is going to be conducted.

```
import torch
from kausal.koopman import Kausal

# Define your cause/effect variables
cause = torch.tensor(...)
effect = torch.tensor(...)

# Initialize `Kausal` object
model = Kausal(cause = cause, effect = effect)
```

Performing causal analysis. You can then perform causal analysis by providing `time_shift` parameter (defaults: 1).

```
# Estimate causal effect
causal_effect = model.evaluate(time_shift = 1)
```

Defining custom observable functions. You can then specify specific observable functions, e.g., `MLP` or `CNN`. By default, it will use `RFF` with dictionary of size $M = 500$. If you choose a learnable dictionary, you can fit them given hyperparameters.

```
from kausal.observables import MLPFeatures

model = Kausal(
    marginal_observable = MLPFeatures(...),
    joint_observable = MLPFeatures(...),
    cause = torch.tensor(...),
    effect = torch.tensor(...)
)

# Fit both marginal and joint observables if using learnable dictionaries.
marginal_loss, joint_loss = model.fit(
    n_train = n_train,
    epochs = epochs,
    lr = lr,
    batch_size = n_train,
    **kwargs
)
```

Changing decomposition method. Several regression techniques to estimate the Koopman operator are also provided, e.g., full-rank pseudo-inverse (`pinv`) or low-rank mode decomposition (`DMD`).

```
from kausal.regressors import DMD

model = Kausal(
    regressor = DMD(svd_rank = 4),
    cause = torch.tensor(...),
    effect = torch.tensor(...)
)
```

D. Experiments

D.1. Coupled Rössler oscillation

We begin our exposition with the coupled Rössler oscillators, a 6-dimensional system $\Omega \in \mathbb{R}^6$ described as:

$$\begin{aligned} \dot{x}_1 &= -\varphi_1 y_1 - z_1, \\ \dot{y}_1 &= \varphi_1 x_1 + a y_1 + c_1(y_2 - y_1), \\ \dot{z}_1 &= b + z_1(x_1 - d). \end{aligned} \tag{17a}$$

$$\begin{aligned} \dot{x}_2 &= -\varphi_2 y_2 - z_2, \\ \dot{y}_2 &= \varphi_2 x_2 + a y_2 + c_2(y_1 - y_2), \\ \dot{z}_2 &= b + z_2(x_2 - d) \end{aligned} \tag{17b}$$

where $a, b, d, \varphi_i \in \mathbb{R}$ are prescribed parameters and $c_i \in \mathbb{R}$ the coupling terms. Setting $c_2 = 0$ naturally partitions the variables into $\Omega_C = [x_2, y_2, z_2]^\intercal \to_K^t \Omega_E = [x_1, y_1, z_1]^\intercal$. Here, we use multilayer perceptrons (MLP) as our deep estimators for the observables, represented as a $M = 32$-dimensional feature vector of the encoder final layer's output.

Experimental setup. The system is solved with Dormand Prince 5th-Order scheme with $t \in [0, 10]$ and $\Delta t = 10^{-2}$. The parameters used here are $a = 0.2, b = 0.2, d = 5.7, \varphi_1 = 1.0, \varphi_2 = 1.0, c_1 = 0.5, c_2 = 0.0$. In the case of the dictionary learning setting, we parameterize the lifting functions ψ_θ of the marginal and joint models with a 2-layer Multi-Layer Perceptron (MLP) with hidden channels of $[16, 32]$, activated by `sigmoid`, and optimized with `AdamW` using a learning rate of 10^{-2} over 500 epochs. Otherwise, in the prescribed dictionary setting, we use Random Fourier Feature (RFF) of size $M = 500$ as the default. Figure S1 illustrates a sample trajectory when varying coupling strength c_1 while setting $c_2 = 0$.

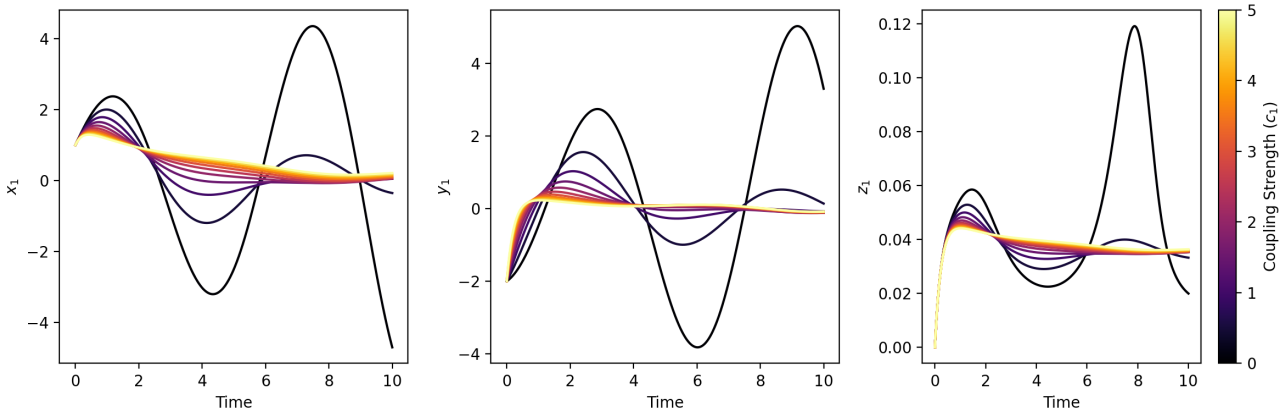


Figure S1. Coupled Rössler Oscillation with varying coupling strength c_1, when $c_2 = 0$.

D.2. Reaction-diffusion equation

We now showcase `Kausal`'s scalability to high-dimensional settings and underscore the notion of uncertainty quantification of causal measures using a 2D nonlinearly coupled reaction-diffusion process described as:

$$\begin{aligned} \dot{u} &= D_u \nabla^2 u - u(u-a)(u-1) + \beta v, \\ \dot{v} &= D_v \nabla^2 v - v(v-b)(v-1) + \gamma u \end{aligned} \tag{18}$$

where $D_u, D_v, a, b \in \mathbb{R}$ are prescribed parameters, $\nabla^2 : \mathbb{R}^{n_x, n_y} \to \mathbb{R}^{n_x, n_y}$ the 2D Laplace operator ($n_x = n_y = 16$), and $\beta, \gamma \in \mathbb{R}$ the coupling terms. The states (u, v) include components of the horizontal velocity. Here, we set $\gamma = 0$, such that the variables are naturally partitioned into $\Omega_C = [v]^\intercal \to^t_K \Omega_E = [u]^\intercal$. In our setup, we use convolutional neural networks (CNN) as our deep estimators for the observables, represented as a 128-dimensional feature map of the encoder final layer's output.

Experimental setup. The system is solved over a 16×16 grid with Dormand Prince 5th-Order scheme with $t \in [0, 10]$ and $\Delta t = 10^3$. The parameters used here are $D_u = D_v = 0.1$ as the diffusion coefficients, $a = b = 0.3$ as the reaction parameters. As described in the main text, the coupling terms are set as $\beta = 10$ and $\gamma = 0$. In the case of the dictionary learning setting, we parameterize the lifting functions ψ_θ of the marginal and joint models with a convolution-based encoder-decoder symmetric structure (CNN) with hidden channels of $[16, 32, 64, 128]$, activated by $sigmoid$, and optimized with $AdamW$ using a learning rate of 10^{-4} over 50 epochs. Otherwise, in the prescribed dictionary setting, we use Random Fourier Feature (RFF) of size $M = 500$ as the default.

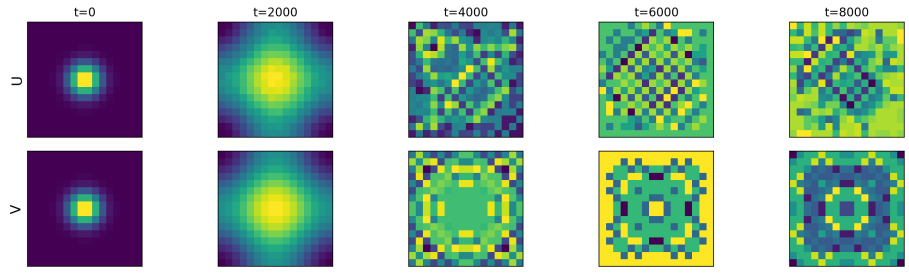


Figure S2. Nonlinear and coupled reaction-diffusion equation sample realization for different time steps.

Uncertainty quantification of causal measures. Our finite approximation of linear flow in $\mathcal{F}$ is imperfect, and therefore, capturing its representation uncertainty is crucial. To this end, we fit a number of CNN kernels to form an ensemble, where we maintain the dimensionality ($M = 128$) but randomly initialize the learnable weights. In Figure S3, `Kausal` is able to extract statistically significant causal signals in the true direction ($\Omega_C \to^t_K \Omega_E$) when compared with the non-causal case ($\Omega_E \to^t_K \Omega_C$) as $t \gg 0$. The ensemble spread is represented as the shaded region around the mean solid line. Note that due to small state values, we scale the causal measure to $[-1, 1]$ and negative values of the causal measure are due to insignificant causal signals.

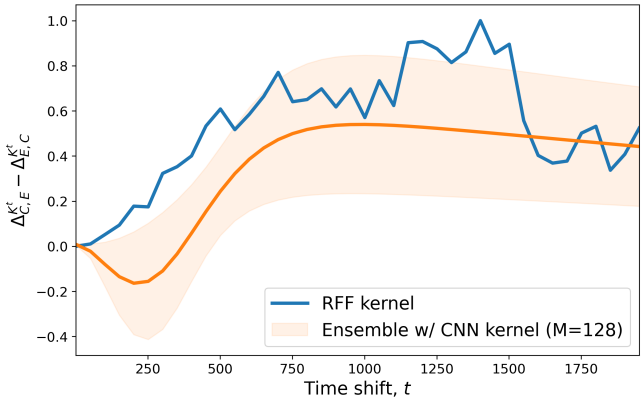


Figure S3. Causal measure of reaction-diffusion equation by computing the difference between the true causal ($\Delta^{K^t}_{C,E}$) and non-causal direction ($\Delta^{K^t}_{E,C}$) across time shifts t, using an ensemble of randomly initialized CNN kernels ($M = 128$) to estimate the observables, versus the default RFF.

D.3. El Niño–Southern Oscillation

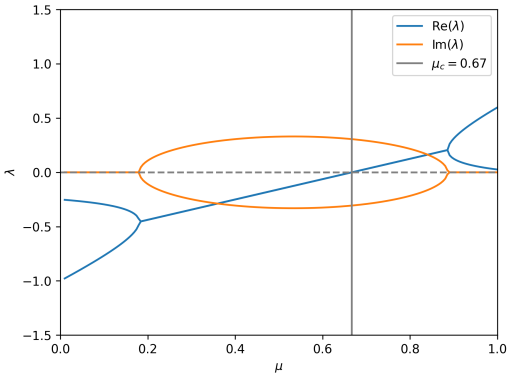
We analyze both simulated and real-world datasets of El Niño–Southern Oscillation (ENSO). We first describe the simulation setup before elaborating on the experiments. We first apply `Kausal` to analyze a well-known ENSO physics-based dynamical model (Jin, 1997) defined as:

$$\begin{aligned} \dot{T} &= -rT - \mu\alpha b_0 h - \epsilon T^3, \\ \dot{h} &= \gamma T + (\gamma\mu b_0 - c)h \end{aligned} \tag{19}$$

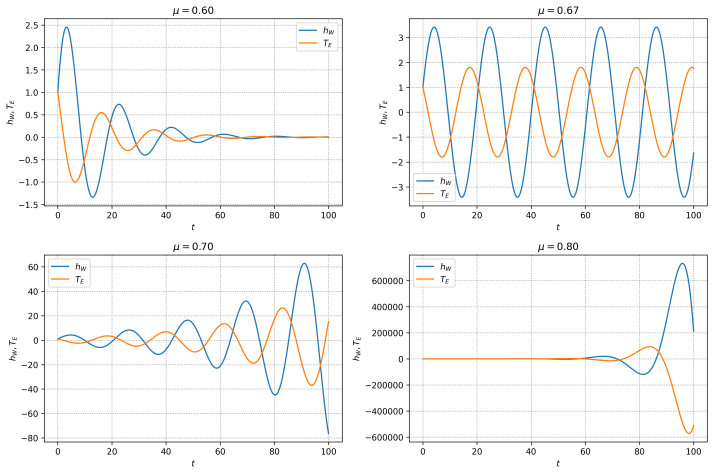
where $r, \alpha, b_0, c \in \mathbb{R}$ are prescribed parameters, $\epsilon \in \mathbb{R}$ is a damping coefficient regulating the strength of nonlinearity, and $\gamma, \mu \in \mathbb{R}$ are the coupling terms. The states include anomalies in sea surface temperature T and thermocline height h in the eastern and western equatorial Pacific, respectively. For notational uniformity, we partition $\Omega_C = [T]^\intercal$ and $\Omega_E = [h]^\intercal$. However, this does not imply $\Omega_E \not\to^t_K \Omega_C$ as long as the coupling terms are non-zero, which is the case throughout our exposition. In our setup, we use MLP as our deep estimator for the observables, represented as a 32-dimensional vector of the encoder final layer's output.

Experimental setup. The system is solved with Dormand Prince 5th-Order scheme with $t \in [0, 100]$ and $\Delta t = 10^{-2}$. The parameters used here are $r = 0.25, \alpha = 0.125, b_0 = 2.5, c = 1.0, \gamma = 0.75$. As described in the main text, we vary both the coupling term μ and nonlinearity parameter ϵ. In the case of the dictionary learning setting, we parameterize the lifting functions $\boldsymbol{\psi}_\theta$ of the marginal and joint models with a 2-layer Multi-Layer Perceptron (MLP) with hidden channels of $[16, 32]$, activated by `sigmoid`, and optimized with `AdamW` using a learning rate of 10^{-2} over 500 epochs. Otherwise, in the prescribed dictionary setting, we use Random Fourier Feature (RFF) of size $M = 500$ as the default.

Linear stability analysis. Here, we provide more details on our linear stability analysis, including a bifurcation diagram and sample trajectories around the critical point $\mu_c = 0.67$ (see Figure S4a).



(a) Bifurcation diagram with varying μ. The critical point is reached when $\mu_c = 0.67$. Meanwhile, $\mu < \mu_c$ and $\mu > \mu_c$ represent dissipative and chaotic dynamics respectively.



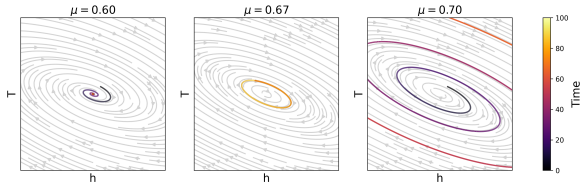
(b) Sample trajectories given identical initial condition but different coupling terms μ. We only analyze the coupled oscillatory regime because the decoupled mechanism is physically unrealistic (i.e., $\text{Im}(\lambda) \neq 0$).

Figure S4. Linear stability analysis of ENSO model by Jin (1997) showcasing in a) the bifurcation diagram that identifies fixed point structures, and in b) sample trajectories given different coupling strength μ that produce dissipative, stable, or chaotic oscillatory realizations (i.e., $\text{Im}(\lambda) \neq 0$).

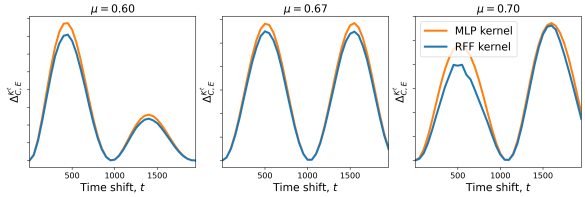
Setting $\epsilon = 0$ in Equation 19 linearizes the dynamics such that several evaluations could be performed, including stability analysis. This is crucial for identifying bifurcation, or tipping points of the system. The full linear stability analysis, including the bifurcation diagram, are described and illustrated in Figure S4. Given a prescribed set of parameters, the critical point $\mu_c = 0.67$ corresponds to a stable periodic attractor. Meanwhile, $\mu < \mu_c$ and $\mu > \mu_c$ represent dissipative and diverging oscillations, respectively. Here, we only consider physically realistic oscillatory dynamics defined in μ-regimes where $\text{Im}(\lambda) \neq 0$.

In Figure S5, `Kausal` reveals accurate linear analysis across μ-regimes that induces dissipative, stable, or diverging oscillatory ENSO dynamics. For instance, in the critical delta region where $\mu_c = 0.67$, our causal measures show identical

peaks with consistent periodicity. Whereas in the dissipative (or diverging) dynamics, our causal measures reveal decreasing (or increasing) influence in the $\Omega_C \to^t_K \Omega_E$ direction. This is physically consistent with the model as both T and h are exerting coupled feedback, and partly governed by μ.



(a) Phase space of ENSO dynamics in different μ-regimes.



(b) Causal measure estimation of ENSO ($\Delta^{K^t}_{T,h}$) across μ-regimes using different estimators for the observables.

Figure S5. `Kausal` reveals the underlying ENSO dynamics exhibiting (*left*) dissipative, (*center*) stable, and (*right*) diverging oscillatory patterns shown in (a) the phase space, and (b) causal measure.

Capturing nonlinearity strength. We perform similar analysis in the stable region of $\mu_c = 0.67$, but vary nonlinearity by setting $\epsilon \neq 0$. As illustrated in Figure S6, we note two physically consistent observations in the causal measure: (a) *oscillatory dynamics* (insofar $\text{Im}(\lambda) \neq 0$), and (b) a *dampening pattern* as nonlinearity strengthens, as evidenced by the slower rate of change in $\Delta^{K^t}_{C,E}$ and smaller magnitude of peaks. Overall, a robust identification of weakly or strongly coupled variables within a system, as demonstrated by `Kausal`, is crucial for better understanding, effective control, and accurate forecasting of dynamics.

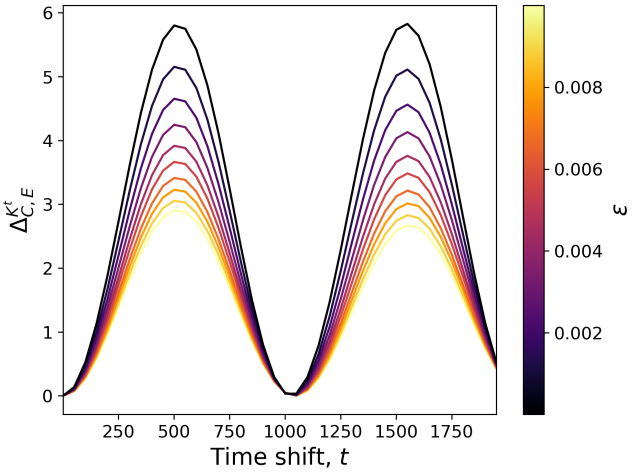


Figure S6. Causal measure of ENSO for varying nonlinearity strength ϵ using MLP kernels to approximate the observables. Our `Kausal` framework is able to meaningfully differentiate and characterize nonlinearity in the underlying ENSO dynamics.

Application to real-world data. Lastly, we move beyond simulation and to a real-world application, using a long-term ocean reanalysis product.

As mentioned briefly in the main text, we use the ECMWF monthly $1.5°$ ocean reanalysis product (Zuo et al., 2019) as preprocessed in Nathaniel et al. (2024). In particular, we use anomalies of `sosstsst` (sea surface temperature) for T and `sossheig` (sea surface height) for h in Equation 19. Note, that while in the ENSO model described above h resembles the thermocline height, we use sea surface height in this observation-based study. Both variables are interchangeable when studying the feedback mechanisms of ENSO (as the sea level height decreases the thermocline shallows; Zhao et al., 2021). We define the equatorial band as $5°$S-$5°$N and then estimate T over the Nino-3 region ($150°$E-$90°$W) and h over the Nino-4 region ($150°$E-$150°$W). The background SST anomaly is computed along the Nino3.4 region ($170°$W-$120°$W). Lastly, we deseasonalize the variables by subtracting the 30-year monthly climatology. Following NOAA's Oceanic Niño Index (ONI; Glantz & Ramirez, 2020), El Niño events are defined when at least five consecutive 3-month running means of SST anomalies in the Nino3.4 region show $T \geq 0.5°$C, and La Niña events when $T \leq -0.5°$C (see Figure S7).

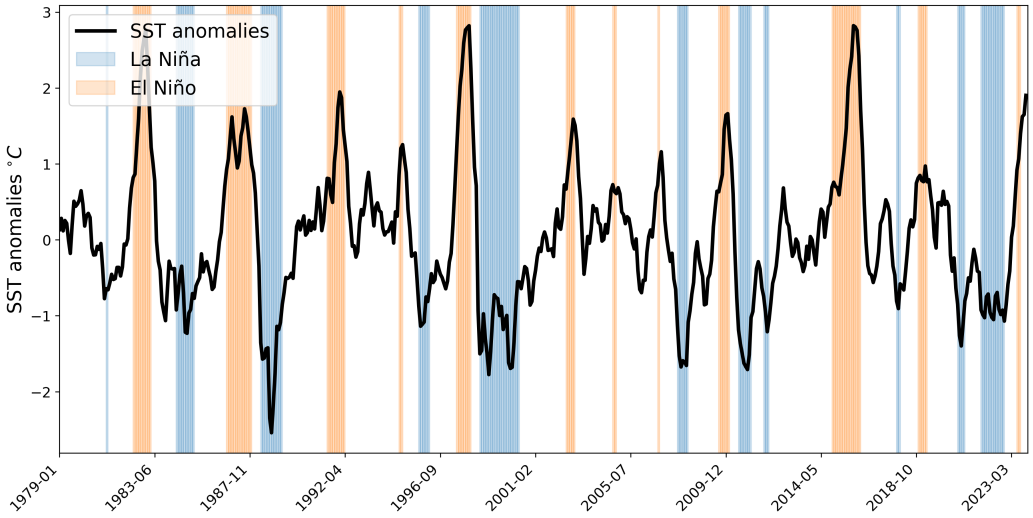


Figure S7. Sea surface temperature (SST) anomalies (black line) in Nino3.4 region to identify major El Niño and La Niña events (shading) according to NOAA's ONI index (Glantz & Ramirez, 2020).

Instead of computing an area average, we use each gridcell as its own independent dimensionality to increase the number of features. The marginal and joint models are estimated using MLP kernels with 2 hidden-channels of size [512, 1024], activated by `sigmoid`, and optimized with `AdamW` using a learning rate of 10^{-2} over 500 epochs. Causal measures for the displayed time periods in Figure 5 and Figure S8 are estimated by iteratively increasing time shifts t in Equation 7 for fixed initial conditions ω_0, which is given by the respective first displayed date.

In Figure S8, an additional analysis of the observed ENSO highlights the limitations of prescribed observables through the failure of RFF to capture causal signals in major ENSO events .

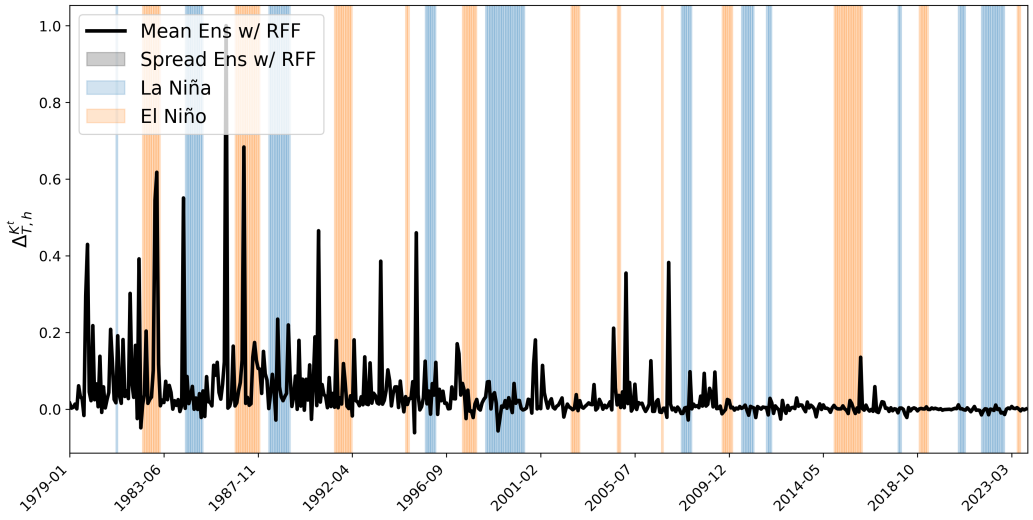


Figure S8. Causal measure of ENSO using real-world data showcasing the use of RFF-derived observables results in failure to detect meaningful causal signals around major ENSO events.

E. Baselines

We run additional ablation studies to benchmark `Kausal` with state-of-the-art causal discovery for dynamical systems.

E.1. Task 1: Causal direction identification

We perform hypothesis testing to identify $\Omega_C \to \Omega_E$. For this work, we use the Coupled Rössler oscillators and reaction-diffusion equation where the true directionality is known and clearly distinguishable (no feedback mechanism i.e., $\Omega_C \to \Omega_E$ and $\Omega_E \not\to \Omega_C$). For `Kausal`, we propose a statistical test to check $\Omega_C \to \Omega_E$, given by Algorithm 1, drawing inspiration from Christiansen et al. (2022). In plain English, the algorithm checks if there is distinguishable causal effect in the direction of $\Omega_C \to \Omega_E$ when compared with an N-times randomly shuffled timeseries for a given time shift t. By default, we use $N = 100$ and $t = 1$ and the significance is measured with a one-sided test.

Algorithm 1 `Kausal` identifiability test ($\Omega_C \to \Omega_E$)

input: Ω_C, Ω_E, N, t
require: $N > 1$ — Number of random permutation
$\Delta_{C \to E} \leftarrow \texttt{Kausal}\left(\Omega_C \xrightarrow{t}_K \Omega_E\right)$ — Apply Equation 6
$\Delta_{C \not\to E} \leftarrow \emptyset$ — Initialize causal measure placeholder for randomized timeseries
for $n \leftarrow 1$ to N **do**
  $\hat{\Omega}_C \leftarrow \texttt{RandomTemporalPermute}(\Omega_C)$
  $\hat{\Omega}_E \leftarrow \texttt{RandomTemporalPermute}(\Omega_E)$
  $\Delta_{C \not\to E}[n] \leftarrow \texttt{Kausal}\left(\hat{\Omega}_C \xrightarrow{t}_K \hat{\Omega}_E\right)$ — Apply Equation 6
end for
$\rho \leftarrow \texttt{PTest}(\Delta_{C \to E} > \Delta_{C \not\to E})$ — Perform one-sided p-test
return: ρ

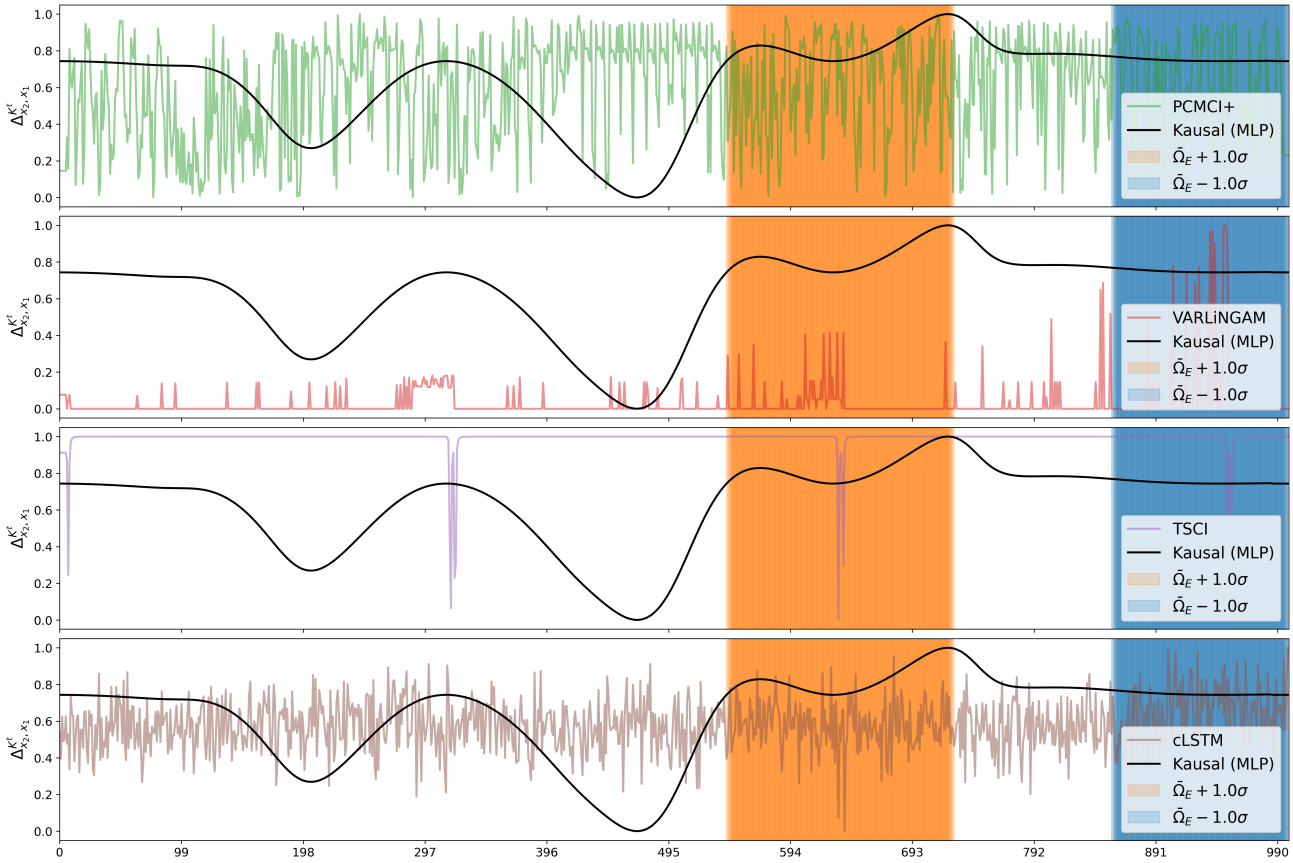
We use PCMCI+ (Runge, 2020) and VARLiNGAM (Hyvärinen et al., 2010) as our baselines to check the true and false causal directions as they provide statistical tests out-of-the-box. The baseline setups are as follow:

- PCMCI+: $\tau_{max} = 1$ to recover at most $t - 1$ time lag with partial correlation (ParCorr) conditional test,
- VARLiNGAM: $\tau_{max} = 1$ to recover at most $t - 1$ time lag.

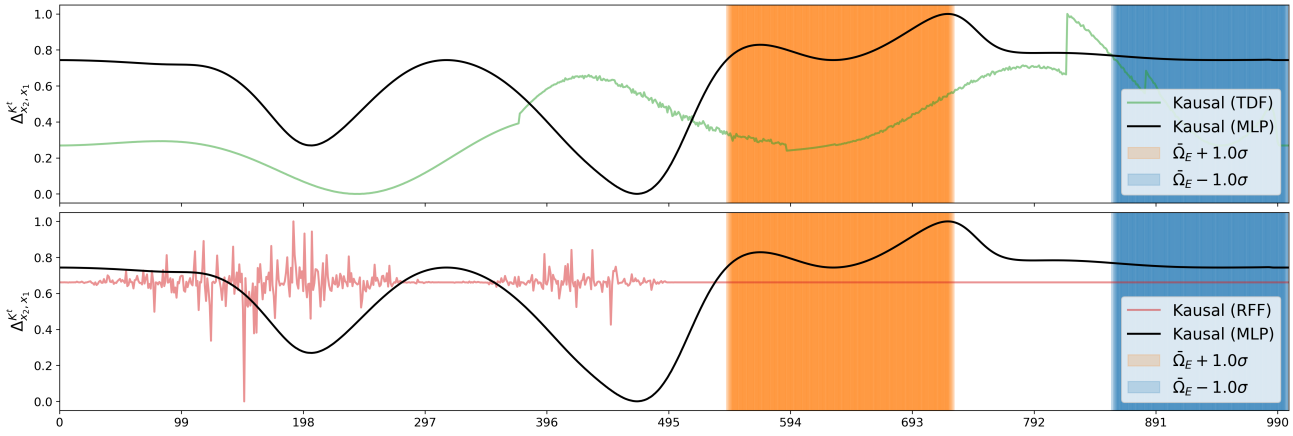
E.2. Task 2: Causal magnitude estimation

Since most baselines (PCMCI+ (Runge, 2020), VARLiNGAM (Hyvärinen et al., 2010), cLSTM (Tank et al., 2021), and TSCI (Butler et al., 2024)) aggregate causal measures given a timeseries (i.e., estimated causal graph is assumed to be static), we employ an identical sliding window strategy proposed in `Tigramite` tutorial for ENSO identification analysis (Runge et al., 2019). In short, we define a sliding window of size 10 and step size of 1. For each time instance $t \geq 1$, we extract $\Omega_C(t-1) \to \Omega_E(t)$ causal effect measure. In this experiment, we compute the absolute causal magnitude estimate (i.e., unsigned) and evaluate it using Area Under the Receiver Operating Characteristic (AUROC) (Peterson et al., 1954) where the binary classification include both positive and negative extrema about a pre-defined $\pm\gamma\sigma$, where $\gamma \in [0, 1]$ and σ is the deviation about a mean state (e.g., in our real ENSO example, the extrema refer to El Niño and La Niña, and the mean state is defined as the climatology). For toy problems, we prescribe $\gamma_{\texttt{Coupled-Rossler}} := 1.00$ and $\gamma_{\texttt{reaction-diffusion}} := 0.50$ to sufficiently capture meaningful extremes. Note that in the reaction-diffusion experiment, we run baselines with a spatially-averaged timeseries as most baseline algorithms do not scale in high-dimensional, multivariate nodes. We use the default hyperparameters settings across models. Finally, we extract the final $L = 1000$ time samples in the Coupled Rössler oscillations, and further subsample at every $S = 50$ steps in the reaction-diffusion trajectory, to mimic how dynamics are typically observed in real-world (i.e., steady-state behaviors and sparse sampling, respectively).

Figures S9 - S11 illustrate the causal magnitude estimation as timeseries given a backdrop of extremes to be detected. While Figures S12-S14 highlight the AUROC curve and scores in the causal detection of extreme signals. Across experiments, we find the superiority of `Kausal` in detecting extreme signals, even in the most challenging real-world ENSO dynamics.

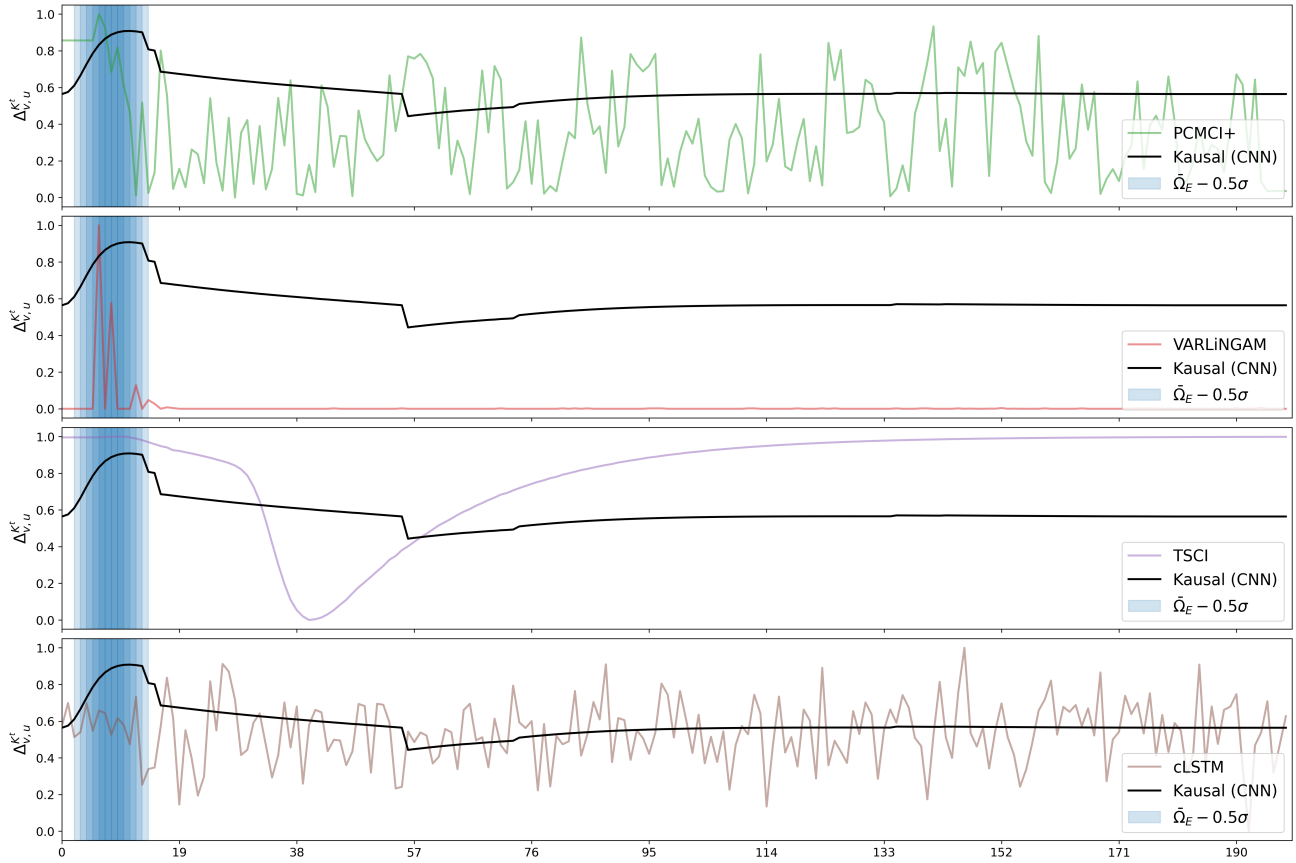


(a) `Kausal` (MLP) and baselines

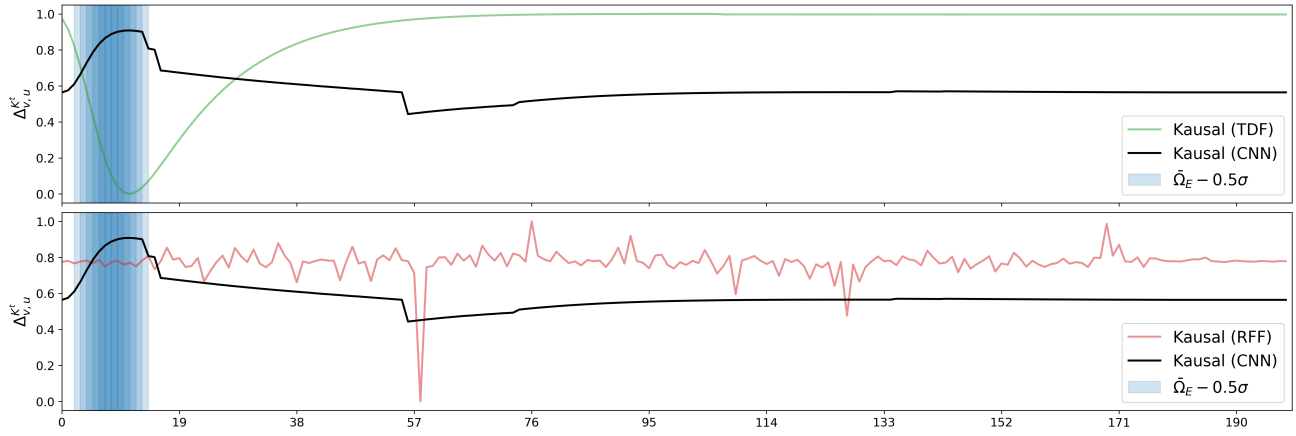


(b) `Kausal` variants

Figure S9. Causal magnitude estimation for Coupled Rössler oscillators given a backdrop of extremes to be detected.

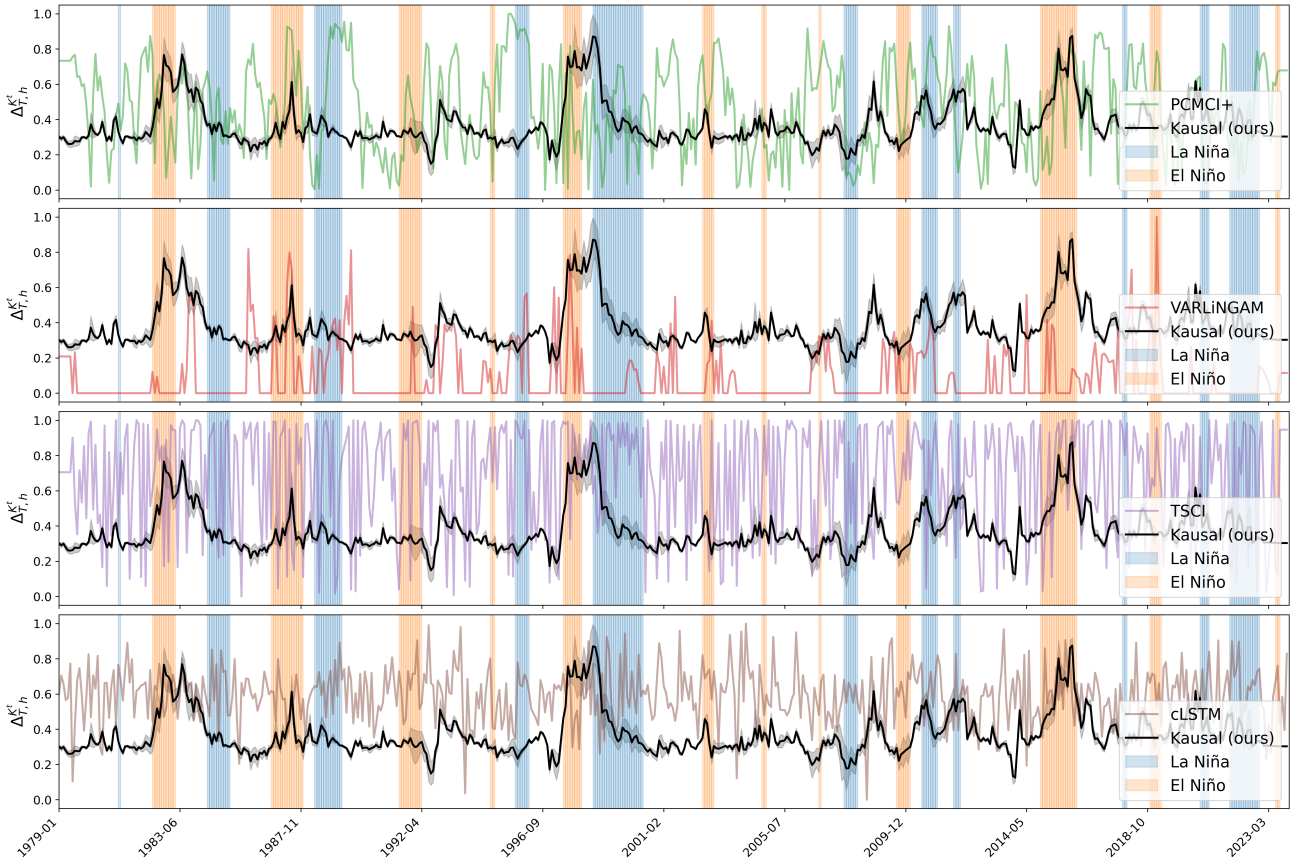


(a) `Kausal` (CNN) and baselines

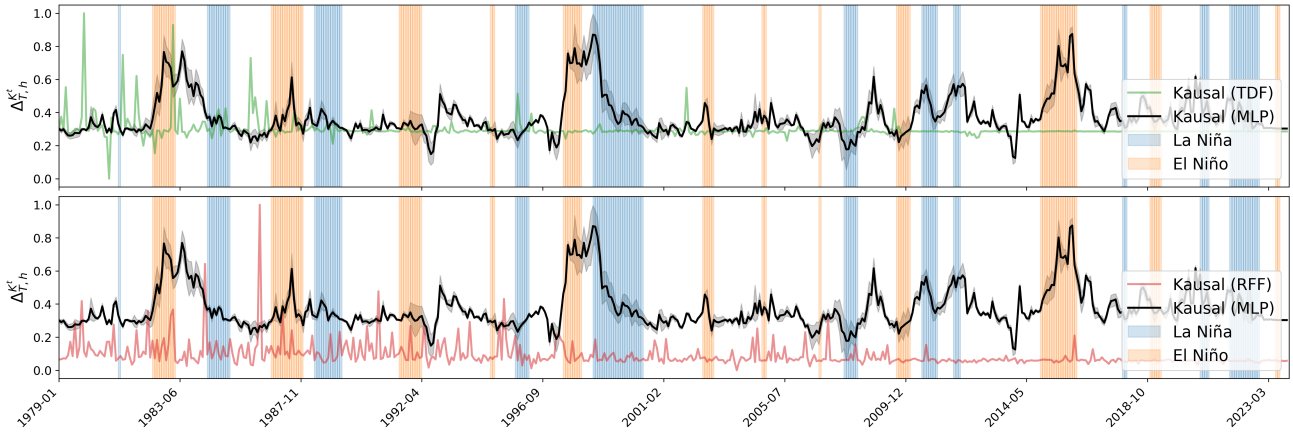


(b) `Kausal` variants

Figure S10. Causal magnitude estimation for reaction-diffusion equation given a backdrop of extremes to be detected.

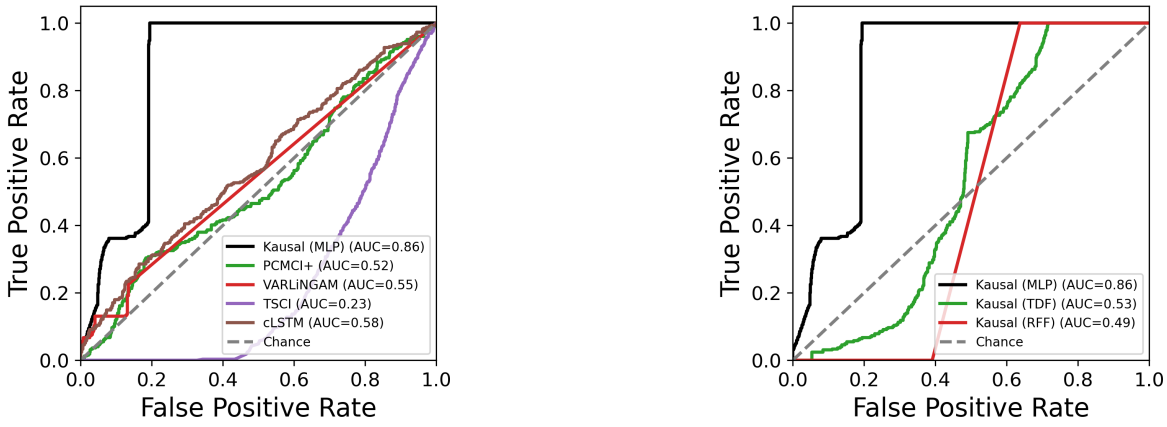


(a) `Kausal` (MLP) and baselines



(b) `Kausal` variants

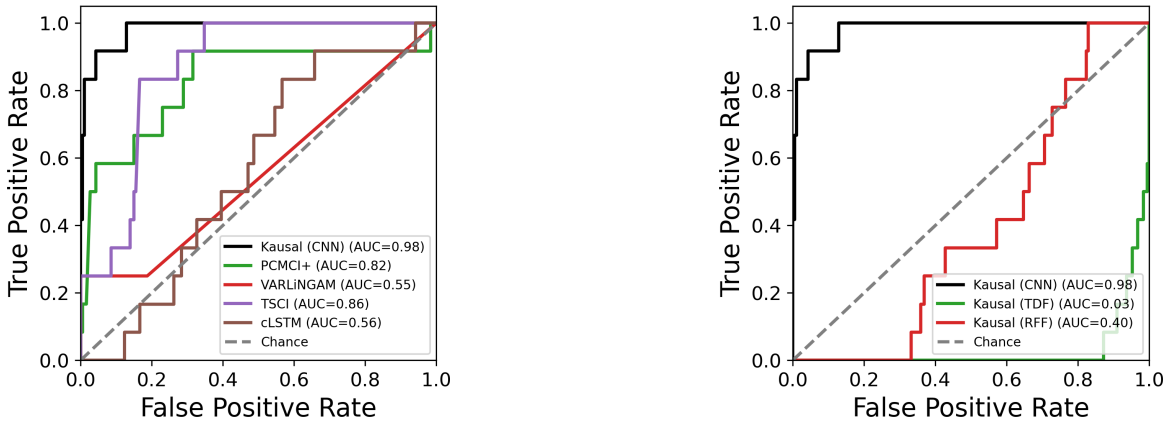
Figure S11. Causal magnitude estimation for real ENSO observations given a backdrop of extremes to be detected.



(a) `Kausal` (MLP) and baselines

(b) `Kausal` variants

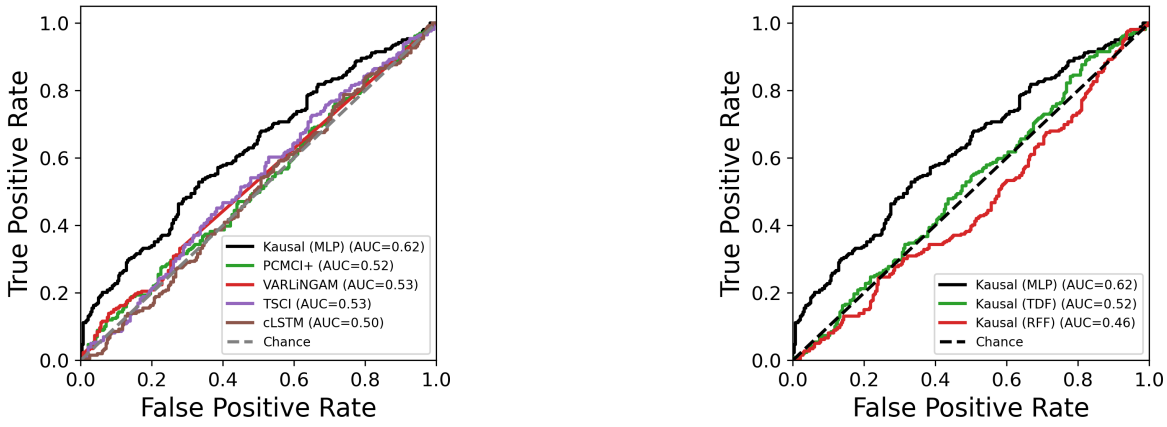
Figure S12. AUROC (↑ is better) for Coupled Rössler oscillators in the causal detection of extreme signals.



(a) `Kausal` (CNN) and baselines

(b) `Kausal` variants

Figure S13. AUROC (↑ is better) for reaction-diffusion equation in the causal detection of extreme signals.



(a) `Kausal` (MLP) and baselines

(b) `Kausal` variants

Figure S14. AUROC (↑ is better) for real ENSO observations in the causal detection of extreme signals.