
Probabilistic Convergent Cross Mapping with Signature Kernels

Kim Bente

School of Computer Science
The University of Sydney
Sydney, Australia
kim.bente@sydney.edu.au

Roman Marchant

Human Technology Institute
University of Technology Sydney
Sydney, Australia
roman.marchant@uts.edu.au

Fabio Ramos

NVIDIA
Seattle, USA
School of Computer Science
The University of Sydney
Sydney, Australia
fabio.ramos@sydney.edu.au

Abstract

Inferring causation directly from time series data of chaotic dynamical systems, such as the climate system, presents considerable challenges but holds substantial scientific value. While many causal inference methods struggle with non-linear and non-separable systems, convergent cross mapping (CCM) has proven to be a practical data-driven technique for such cases, however lacking robustness under deceptive conditions like synchrony, and also missing a probabilistic treatment. We introduce elaborate time series kernel methods into the CCM framework, to leverage their ability to detect intricate relationships among time series segments. With SIGCCM we propose a novel, non-parametric method for time series causal discovery (TSCD), that particularly suits data-sparse problems. SIGCCM leverages the power of signature kernels, to represent similarities and implicit uncertainty among response signatures. We evaluate SIGCCM’s ability to infer underlying causal mechanisms in challenging experiments on simulated and hybrid data, and we further demonstrate its merits on real-world data of past climate from Antarctic ice cores. SIGCCM shows promising capability in uncovering true causal structures, even in the presence of synchrony and confounding.

1 Introduction

Understanding the causal mechanisms that govern our physical world, lies at the core of the scientific pursuit [1], [2]. Our knowledge about causal relationships among variables from complex real-world systems like climate, ecology, neuroscience or economics, is still incomplete, yet it is imperative to inform the design of effective policy via targeted interventions. Beyond scientific discovery, causal inference also has a symbiotic relationship with predictive modelling, as new knowledge about the causal structures driving the data-generating process can lead to more robust models, for example yielding better extrapolation capabilities, which are instrumental for projecting future behaviour of drifting systems like the warming climate system. Causal methods therefore present an important **complementary analytical tool** [2] that can directly aid physics-based models, hybrid models like emulators or digital twins, but also data-driven machine learning models, which rely on few but

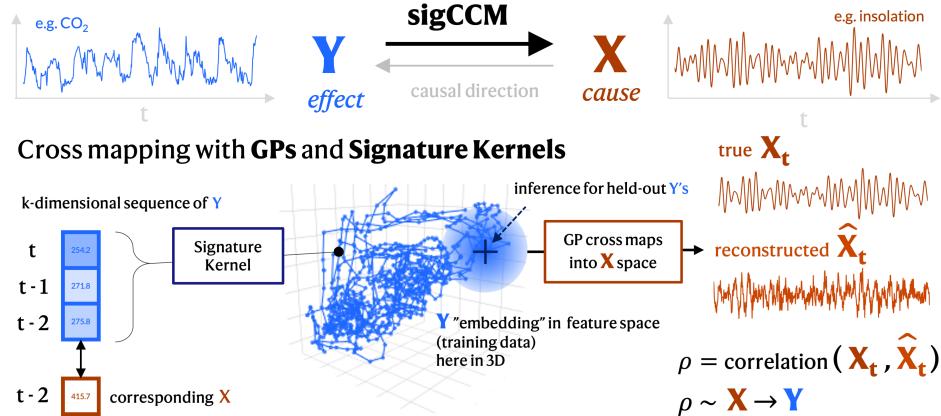


Figure 1: Overview of causal inference with SIGCCM

critical assumptions reflected in the choice of model family and architecture, which shall be aligned to reflect process understanding regarding, for instance, dependence and independence among inputs, or the temporal delays between cause and effect.

While interventions, like in the form of *Randomised Control Trials* (RCTs), are the gold standard of causal inference in the experimental setting, in practice, experiments are often not feasible, thus requiring data-driven approaches that enable **causal inference from time series observations**. Particularly natural, large-scale dynamical systems like the earth system rely on *Time-Series Causal Discovery* (TSCD) [2], since “we only have one earth”. Nonetheless, also in other domains such as finance, medicine or ecology, time series data from simulation or observation are abundant, while experiments incur additional costs and potential ethical issues.

Although the notion that ‘*correlation does not imply causation*’ is widely instilled in graduate students, linear correlation measures like the Pearson correlation coefficient remain the default, generic tool to reason about pairwise relationships between variables. Notwithstanding, simple examples by Sugihara, May, Ye, *et al.* [3] demonstrate the ineptitude of Pearson correlation to infer causal structure in non-linear and coupled dynamical systems. A prominent concept that many TSCD methods are based on [4], [5] is **Granger Causality (GC)** [6]. GC builds on the connection between forward predictability and causation: If X causes Y , Granger expects Y to be more predictable if we have knowledge about preceding X values, compared to when these are removed from such a predictive model. Therein, it presumes separability of X and Y , i.e. that removing X removes all information contained in X . If there also is a causal link from $X \rightarrow Y$ however, information about X will inherently be contained in Y itself. Consequently, GC’s **separability** requirement is not satisfied in many real-world systems with coupled dynamics [3], i.e. when variables are functions of each other, which limits its application scope.

In recent years, the application of machine learning techniques to the various paradigms of data-driven causal inference has advanced greatly: *PCMCI* [4] and subsequent *PCMCI+* [7], which extends the method to contemporaneous links, utilise conditional independence tests in combination with the PC algorithm and were designed specifically with large-scale multi-variate data in mind. *Neural Granger Causality (NGC)* [8] incorporates MLPs and RNNs into GC, while *CUTS* [9] builds upon this idea but extends it to irregular data using GNNs. *CUTS+* [10] presents an extension of CUTS to higher dimensional time series. We recommend [5] and [2] for a more complete overview. In addition to newly proposed methods, competitions like the *Causality for Climate Competition* at NeurIPS 2020 [2], [11] have drawn attention to causal analysis in particular domains like the earth sciences, and have spurred the development of appropriate benchmark datasets [5], [12] to close the gap between simulation-based experiments and causal inference in the wild.

Convergent Cross Mapping (CCM) Sugihara, May, Ye, *et al.* [3] was developed to fill the need for causal inference in coupled chaotic dynamical systems (more details about CCM will follow in Section 2). CCM reverses the GC concept of *forward* predictability into *backward* reconstruction, by determining causal associations based on how well state-space embeddings of the response variable

enable reconstruction of the prior causal variable (the *causal X* is reconstructed from succeeding time series segments of the *affected Y*).

Despite its profound impact and practical relevance, CCM struggles under various scenarios, e.g. to discern strong uni-directional causality from bi-directional coupling in the case of *generalized synchrony* [3], [13]. Furthermore, CCM uses *Simplex Projection* (SP) [14], which is based on a k-nearest-neighbour algorithm with a fixed number of neighbours, to map from the response embedding space to the causative variable. Such projections fail to incorporate, for instance, the density of reference (training) points in the embedding space. A principled representation of this uncertainty, reflecting if similar patterns have been observed in the reference data, is not only important to assure robust results if the length of available time series is limited, but also in the case of system drift, for instance due to anthropogenic warming of the climate. We address these limitations by introducing SIGCCM, which translates CCM into a Gaussian Process framework [15], using a signature kernel to define covariances between time series segments, see Figure 1. Our contributions include:

- We propose **SIGCCM**, a non-parametric, probabilistic causal inference method, that combines CCM with expressive **path signature kernels** in a **Gaussian Process framework**.
- We demonstrate SIGCCM’s remarkable capabilities on simulated data, testing for **CONFOUNDING**, and on hybrid data, inferring causal direction and time lags under **SYNCHRONY**.
- Using real-world climate data from the Vostok ice core we showcase how SIGCCM provides interpretable insights relating to e.g. uncertainty, and delays in response.

The paper is structured as follows: Section 2 provides an overview over CCM, with an emphasis on machine learning causal inference extensions. Section 3 introduces our proposed method, SIGCCM. Experiments on simulated, hybrid and real-world data, as well as respective results are presented in Section 4 and subsequently discussed in Section 5. Discussions extend to limitations and future work, and Section 6 provides a brief conclusion of our work.

2 Background – Convergent Cross Mapping (CCM)

In this section we start by reviewing the history and evolution of Convergent Cross Mapping (CCM) and emerging methods, particularly machine learning extensions to CCM. We briefly position our work in their context, and we further provide a brief overview of recent machine learning advances in causal inference from time series data more broadly.

CCM and Taken’s theorem. Sugihara, May, Ye, *et al.* [3] developed Convergent Cross Mapping (CCM) as a causality test that suits coupled dynamical systems, where separability - a requirement for Granger causality (GC) - is violated (as discussed in Section 1). CCM is grounded in **Taken’s delay embedding theorem** [16] for chaotic dynamical systems, which postulates that the attractor manifold of a dynamical system can be fully recovered by the delay embedding of any variable that belongs to that dynamical system. CCM hence exploits the idea that information about the cause is contained in the ‘signature’ of the effect: If the causal variable (the driver or the ‘causal X’) can be reconstructed from delay embeddings of the time series of the response variable (the effect or the ‘causal Y’), and reconstruction quality, also referred to as the *cross mapping skill*, converges with the size of training data, causation in the $X \rightarrow Y$ direction is inferred. By repeated pair-wise assessment of a system’s variables, causality, encompassing direct and indirect causality through transitivity, is established. CCM has since been applied to diverse domains including climate [17], finance [18], ecology [3], neuroscience [19], and the geoscience [20].

Extended CCM. One of the initial drawbacks of CCM was its challenge to infer correct causality in the presence of synchrony [3], [13]: In what is known as ‘generalized synchrony’, a strong uni-directional forcing makes the response variable move approximately in lockstep with the driving variable. Synchrony is hard to distinguish from bi-directional coupling, where both variables are a function of the other [3]. Whereas it uses the same algorithm as CCM, *Extended CCM* (ECCM) [13] extends the inference process by the explicit consideration of response delays, to help distinguish strong uni-directional forcing from coupling. To do so, ‘shift plots’, visualising cross mapping skill over a range of shift values, are used to determine at which shift peak cross mapping skill occurs. Joined analysis of $X \rightarrow Y$ and $Y \rightarrow X$ enable reasoning about the true causal structure: *Causal precedence* requires the logical temporal ordering of ‘cause precedes effect’ to infer a valid causal link (and given the causal response is not instantaneous). Another practical CCM extension was

introduced by Benkő, Zlatniczki, Stippinger, *et al.* [19], who propose a Bayesian approach that outputs a probability distributions over causal configurations.

Machine learning extensions to CCM. Recent works have also extended the CCM framework with machine learning techniques, to facilitate CCM causal inference under specific, challenging circumstances: *Reservoir Computing Causality* (RCC) [21] builds on CCM but replaces the embedding-based prediction model with a reservoir computing prediction model, that directly maps from the response variable to the reconstructed, causative variable. RCC demonstrates robustness to noise, as well as computational efficiency during inference, scaling to high-dimensional scenarios. Nonetheless, RCC is a parametric approach, requiring training, and thus also access to sufficient training data prior to inference. Brouwer, Arany, Simm, *et al.* [22] proposed *Latent CCM* (LCCM), another direct approach, which, extends the CCM framework to challenging scenarios characterised by high degrees of missing data, where a number of short segments of time series are observed (addressed previously by *Multi-Spatial CCM* [23]), and additionally, segments consist of irregular, sporadic observations. Like RCC, Latent CCM bypasses the computation of fixed-dimensional delay embeddings, by learning a Neural ODE latent process from the sporadic data to emulate state-space dynamics, and subsequently evaluating cross-mapping skill between system variables in the latent space. Alongside CUTS+ [10], LCCM achieved overall leading performance on the recently published CausalTime benchmark [5]. In contrast to both of these novel, parametric machine learning CCM methods, we designed a non-parametric approach, that does not require training, and that we thus deem particularly useful for data constrained causal inference.

Throughout this paper we will use causal X, Y notation, consistent with [3] and the main text in [17], where X represents the causal variable and Y represent the response variable. In CCM we test the existence and strength of the causal $X \rightarrow Y$ relationship through predictability of cause $x(t)$ given the corresponding $y(t)$.

CCM defines delay embeddings in the form of $\{y(t), y(t-\tau), \dots, y(t-(E-1)\times\tau)\}$, where E is the effective embedding dimension, and τ is a fixed step size. This formulation of CCM constraints delay embedding to regular lags. We define k as the span of the embedded sequence, where $k = E \times \tau - 1$. CCM uses Simplex Projection (SP), based on the nearest-neighbour algorithm, to "cross map" from the state-space embedding to reconstruct $x(t+s)$, where s is the shift or offset parameter. We usually set $s = -k + 1$ (or smaller) so that x corresponds to the earliest embedding position, and to thereby avoid information leaks that would be inconsistent with causal precedence. As discussed in Section 2, in ECCM we purposefully look at a range of s values to determine $\text{argmax}_s(\rho)$. SP suffers from ties arising from non-unique data and is also rather slow when not computed as a tensorised version.

3 Signature Kernel CCM (SIGCCM)

We formulate CCM within a Gaussian Process (GP) framework [15], where the GP replaces the Simplex Projection (SP) to achieve a fully probabilistic cross mapping, reconstructing y from x , as visualised in Figure 1. A GP prior over this mapping incorporates the principles behind non-parametric probabilistic modelling, which are highly flexible, yet capable of controlling over-fitting and quantifying uncertainty. Gaussian Processes make convergence more robust by exploiting patterns in the embedding space and generalising even in cases with a low number of data points. Intuitively, as the space becomes denser and cross mapping skill increases.

We go one step further by using **path signature kernels** [24], [25], a powerful time series kernel method with strong theoretical guarantees, that directly takes as input segments of the time series rather than delay embeddings. Signature kernels extract informative features from time series, including moments and invariant properties, and represent them in a higher m -dimensional latent space. Given a time series segment $\{y_t, y_{t-1}, \dots, y_{t-k}\}$, its signature up to depth m is defined by Equation 1:

$$\text{Sig}^m(y_{t:t-k}) = \left(1, y_t, y_{t-1}, \dots, y_{t-k}, \int y_t dy_{t-1}, \dots, \int \dots \int y_t dy_{t-1} \dots dy_{t-k} \right), \quad (1)$$

where $\int \dots \int y_t dy_{t-1} \dots dy_{t-k}$ represents iterated integrals up to depth m . The static kernel used to encompass the signature kernel:

$$k_{\text{sig}}(y, y') = \exp \left(- \frac{\|\text{Sig}^m(y) - \text{Sig}^m(y')\|^2}{2 \sigma_{\text{rbf}}^2} \right) \quad (2)$$

only has a single hyperparameter, σ_{rbf} which can be fitted by minimising the negative log likelihood [15]. Covariances between sequences are based on their representation in this latent space and act as a similarity measure within the GP framework. Like LCCM [22] and RCC [21], SIGCCM thus bypasses the explicit state-space embedding step to directly map from time series sequences of the response variable $\{y_t, y_{t-1}, \dots, y_{t-k}\}$ back to the causal variable x_{t+s} .

Metric. Although the output of the Gaussian Process mapping is a posterior predictive distribution over x , and thereby enables evaluation of the likelihood, we consider the predictive mean as our deterministic reconstruction \hat{x} to allow comparison with non-probabilistic methods like CCM. The *cross correlation skill* ρ is subsequently defined as the Pearson correlation coefficient between the mean value of the reconstructed time series \hat{x} and the ground truth time series x , as defined in Equation 3 and visualised in Figure 1.

$$\rho = \text{corr}(\hat{x}, x) \quad (3)$$

Experimental design. We determine cross mapping skill ρ with a training data sequence length of $n_{\text{train}} = 100$ across experiments. We use k-fold cross-validation with $k \approx N$, so given our consistent time series length of $N = 400$, we evaluate over approximately 400 folds, with the start position of the uninterrupted training sequence iterating over the time series. Each fold thereby has $n_{\text{test}} \approx 300$ test data points. Variation is span k and shift s used for the embedding step, appear as minor deviations in the resulting n_{test} and the number of folds, as only complete pairs of delay embedding and target variables can be used both for training and inference. Thus ρ typically refers to the mean cross mapping skill over all folds.

Statistical significance. Although original CCM infers causality based on convergence with increasing size of n_{train} , we adopt the approach based on an ‘independence model’ presented in [21] to test for statistical significance. Therein, we generate surrogates of x with the iterative amplitude adjusted Fourier transform (IAAFT) [26], which preserves the amplitude distribution and power spectrum of the original time series, but eradicates the causal link $X \rightarrow Y$. Again, we perform k-fold cross-validation, but we generate a new surrogate time series x_{ind} for every fold. The 95th percentile of the distribution of ρ_{ind} is reported as an upper bound, $\rho_{\text{ind}} p95$. If $\rho_{\text{ind}} p95 > \rho$, the link $X \rightarrow Y$ is considered insignificant, whereas $\rho_{\text{ind}} p95 < \rho$ leads to the conclusion that $X \rightarrow Y$ is significant.

4 Experiments

To rigorously evaluate SIGCCM we construct three experiments that examine various particularly challenging test cases, which have previously been discussed in the literature [3], [13], [22]. Firstly, we test on simulated data, to ensure that we have access to the data generating processes and underlying causal relationships. Secondly, we test SIGCCM in a hybrid setting, where we use a real-world time series, within the Earth science domain, as the causal driver of a dependent, simulated time series. Thereby, we maintain the knowledge of the ground truth while exposing SIGCCM to the complexity and characteristics of natural temporal phenomena [5]. Lastly, we apply SIGCCM to real-world data from the Vostok ice core.

Because a key motivation behind the design of our proposed non-parametric method is to preserve efficacy in data-sparse scenarios, all experiments are conducted on relatively short time series ($N = 400$), approximately equivalent to the length of the time series reconstructed from the Vostok ice cores. Recent CCM-based methods, like LCCM [22] and RCC [21], discussed in Section 2, are parametric approaches which require parameter training and large quantities of data (although LCCM takes sporadic data as input, the total length of the data is high with $t = 100,000$). We restrict our comparison to the less data-hungry, original CCM [3], as well as the methodologies introduced in ECCM [13]. In the remaining of this section, we first present all three experiments, including the intent of the experiment, baselines and data-generation/data, and then we elaborate on results obtained for each setting.

4.1 Simulated data: CONFOUNDING

We construct a simulated experiment to evaluate the capacity of SIGCCM to identify confounding, i.e. whether it can distinguish that two variables a and b which have a common cause c , the confounder, but are not causing each other. According to this principle, we generate data from a three-variable system consisting of 3 deterministic logistic difference equations, defined in Equation 4.

$$\begin{aligned} c(t+1) &= c(t) [3.9 - 3.9 c(t)] \\ a(t+1) &= a(t) [3.8 - 3.8 a(t) - 0.3 c(t)] \\ b(t+1) &= b(t) [3.6 - 3.6 b(t) - 0.3 c(t)] \end{aligned} \quad (4)$$

We initialise $c(1) = 0.2$, $a(1) = 0.3$ and $b(1) = 0.4$, and then autoregressively generate the deterministic time series of length $N = 400$. Despite the forcing, the linear Pearson correlation coefficient between c and a is 0.218, between c and b is only -0.004 and between confounded a and b is -0.108. For SIGCCM we set the span of the effect embedding window to $k = 3$, and as a hyperparameter we use $\sigma_{\text{rbf}} = 0.3$. As a baseline for this experiment we use CCM, and we set $E = 3$ and $\tau = 1$ for a fair comparison. We set the common shift parameter to $s = -k + 1$, so that the earliest position of the embedding is at the same time as the cause eventuates, which effectively embeds around the true effect lag of 1, while adhering to the causal precedence constraint. We then evaluate the inferred causal graphs of SIGCCM and CCM, obtained based on significance test comparing to an ‘independence model’ described above, in section 3.

4.2 Hybrid data: SYNCHRONY and delay analysis

Next, we construct a hybrid test case, combining real and simulated data, to test causal inference in the presence of synchrony and to determine if true effect lags can be detected by SIGCCM or the ECCM baseline. Discerning a true, strong uni-directional forcing from bi-directional coupling between two time series that move in synchrony, is known to be a challenging task [3], [13], [27]. Thus, we mimic a uni-directional forcing $d \rightarrow e$ where variable d dictates variable e , resulting in an inverse synchronous relationship, defined in Equation 5 and visualised in Figure 5. Additionally, we delay the effect of d on e by 2. For time series d we use atmospheric CO₂ recovered from the Vostok ice core [28], which may be interpreted as a stochastic external forcing, that drives the temporal process of e . The Pearson correlation coefficient between d and e is -0.78, reflecting the inverse synchronous relationship. We describe preprocessing of the ice core data in more detail in subsection 4.3 and in the Appendix A.3.

$$\begin{aligned} d(t) &:= \text{CO}_2(t) \\ e(t+1) &= e(t) [1.4 - 1.9 e(t) - 0.3 d(t-2)] \end{aligned} \quad (5)$$

We initialise the first time steps of $d = [0.2, 0.1, 0.1]$ and again we set $n_{\text{train}} = 100$. To infer causal relationships in the presence of synchrony we follow the method proposed in ECCM [13], which considers the cross mapping skills achieved at various shifts s , typically presented as ‘shift plots’, see Figure 3. Interpreting at what value of s optimal cross mapping skill can be observed, i.e. $\text{argmax}_s(\rho)$, in combination with significance test discussed in subsection 3, is taken as an indication of the true causal structure and true delay between cause and effect. Only peaks at shifts that conform with causal precedence are interpreted as $X \rightarrow Y$, while peaks in the overlap region may lead to the conclusion of instantaneous effects, or that the temporal resolution of the data is not sufficient to infer causal order.

4.3 Real-world data: Application to CLIMATE data

We further test our proposed method on historical climate records from the Vostok ice core [28], from East Antarctica, spanning $\sim 400,000$ years (400 ky) of past climate. The data represents a test case where we do not have full knowledge about the data-generating mechanisms, although prior scientific knowledge of course exists, see [17]. We replicate experiments from Nes, Scheffer, Brovkin, *et al.* [17], however with a strictly enforced causal precedence constraint, and derive at alike conclusions about causal relationships between CO₂, CH₄, temperature, insolation. Please refer to the Appendix A.3 for more information on the data and preprocessing. In the absence of ground truth

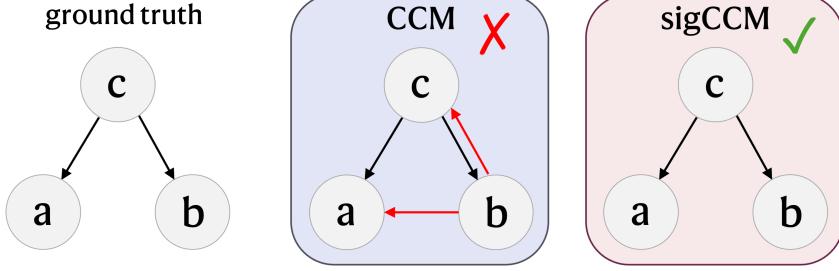


Figure 2: Ground truth and inferred causal graphs of the **CONFOUNDING** experiment on simulated data, see 4.1. Note that SIGCCM correctly reconstructs the graph from data while CCM does not.

Table 1: Results of the **CONFOUNDING** experiment on simulated data, see subsection 4.1 and Figure 2. gt stands for ground truth, ρ represents the cross mapping skill and $\rho_{\text{ind}} p95$ represents the upper bound of the independent model to determine significance.

$X \rightarrow Y$	gt	CCM			sigCCM		
		ρ	$\rho_{\text{ind}} p95$	correctness	ρ	$\rho_{\text{ind}} p95$	correctness
$c \rightarrow a$	✓	0.819	0.138	✓	0.721	0.098	✓
$c \rightarrow b$	✓	0.904	0.149	✓	0.759	0.096	✓
$a \rightarrow b$	✗	0.091	0.134	✓	-0.006	0.099	✓
$a \rightarrow c$	✗	0.052	0.134	✓	-0.017	0.104	✓
$b \rightarrow a$	✗	0.159	0.116	✗	0.073	0.104	✓
$b \rightarrow c$	✗	0.156	0.124	✗	0.098	0.118	✓

to evaluate our findings, we apply SIGCCM to demonstrate some valuable properties of our method, focusing on CO₂ and temperature time series. In this demonstration we set $n_{\text{train}} = 200$ and we select the first 200 ky as our training data. For the remaining $n_{\text{test}} = 200$ test points, we reconstruct temperature from CO₂, testing for a feedback in the direction of temperature → CO₂, in addition to the well established link from CO₂ → temperature. As SIGCCM produces posterior distributions for temperature, we show samples of the posterior that reflect the reconstruction uncertainty given (the lack of) similarities with sequences contained in the training data. We set $k = 3$, $s = -4$ and $\sigma_{\text{rbf}} = 0.4$.

4.4 Results

We present our results for the **CONFOUNDING experiment** on simulated data, described in subsection 4.1 in Table 1. We also visualise the causal graphs, corresponding to the inference presented in Table 1 in Figure 2. Notably, SIGCCM correctly infers the ground truth, i.e. it correctly infers all 6 pairwise relationships, and thereby identifies the confounding relationship. Ordinary CCM misclassifies 2 pairwise relationship and thus predicts an incorrect causal graph.

‘Shift plots’ for the **SYNCHRONY experiment** on hybrid data, introduced in subsection 4.2, are shown in Figure 3. Note that the peak of SIGCCM for $d \rightarrow e$ corresponds to the true lag, shown in the dashed green vertical line, while ECCM has a less clear signal around the true lag, but attains high cross mapping skills across various negative shift values. This shows the superior sensitivity of SIGCCM which enables it to identify true lags. Importantly, higher scores do not indicate better model performance, as causal inference is an analytical rather than a predictive tool. Additionally, ρ values must be interpreted in relation to the upper bound of the independent model: Figure 3 clearly shows considerably higher signal for the independent model in both directions. For $e \rightarrow d$, ECCM falsely identifies a significant causal links at shifts that represent causal precedence, whereas SIGCCM correctly has non-significant ρ values across all shift values that conform to ‘cause precedes effect’.

We showcase samples from the temperature reconstruction posterior distribution for the **CLIMATE experiment** in Figure 4. Note how certain regions have lower associated variance, representing more

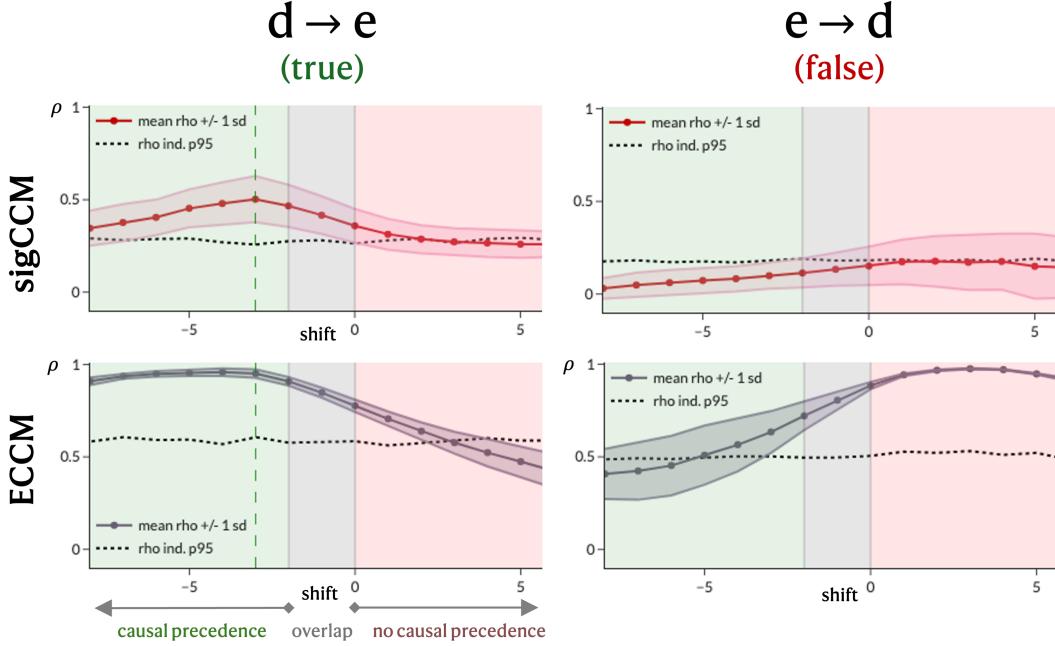


Figure 3: Resulting ‘shift plots’ for the SYNCHRONY experiment, where s is presented on the x-axis and ρ is represented on the y-axis. Mean ρ values $+$ / $-$ one standard deviation are shown in relation to the upper bound of the independence model, $\rho_{\text{ind}} \text{ p95}$, presented by the dotted line. The dotted green vertical line represents the true lag for $d \rightarrow e$. The green area represents causal precedence shift values, i.e. where the reconstructed x occurs simultaneously or prior to the earliest position of the embedded y sequence.

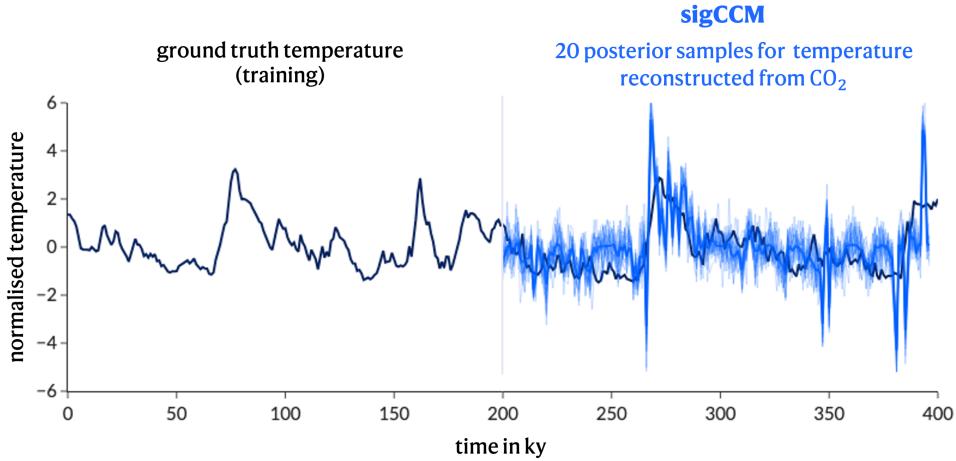


Figure 4: Probabilistic temperature reconstruction from CO₂ data for the **CLIMATE experiment**

similar sequences in the training data, whereas other regions exhibit high variance. This may be useful to scientists to identify unusual behaviour, or to further interpret explainable and less explainable system behaviour. Fitted hyperparameters are interpretable and offer additional insight: We find the highest ρ at a shift of $s = -5$, indicating a lag of 3ky - 5ky in the response of CO₂ to temperature, which corroborates with previous findings [13], [17]. We further note that posterior samples are not smooth, like in GP time series predictions, because time is not used as an input of the kernel function, but only time series segments are used in this CCM framework.

5 Discussion

We introduce SIGCCM, an extension to CCM, harnessing signature kernels for causal inference in dynamical systems. Importantly, our proposed method does not require training like recently proposed machine learning CCM extensions [21], [22]. Experimental results of SIGCCM demonstrate high skill in determining challenging true causal mechanisms from limited data, consistently outperforming our baselines. Thus SIGCCM may bring high practical value to use cases where the acquisition of every data point is costly, as in the case of ice core data that incurs logistical obstacles as well as high processing burden.

Since SIGCCM is based on CCM (although not strictly representing the state-space), assumptions such as causal sufficiency, and limitations, such as the restriction of the CCM methods to chaotic dynamical systems [3], also apply to our method. Furthermore, evaluation of time-series causal discovery methods remains a challenge, trading-off access to causal ground truth with real-world complexity in the data [2], [5], [12], making it hard to determine a universally reliable technique. Additionally, SIGCCM should be tested in high-dimensional scenarios, given contemporaneous links, and also in the presence of noise.

Results from our experiments on simulated data, see Table 3, as well as hybrid data, see Figure 1, highlight the greater selective tendency in SIGCCM compared to CCM. Even for significant causal links, SIGCCM's ρ values are typically lower than those of CCM. It is important at this point to reiterate, that different to predictive machine learning, cross mapping skill itself is not the objective, but it is the means to an end, which is to make reliable inference about causal relationships. This higher selective power appears to enable SIGCCM to confidently detect confounding and uni-directional relationships from the data. The consistently higher cross mapping skills by CCM inferred from the independent model, reinforce the observation that CCM picks up more spurious relationships, whereas SIGCCM provides a more nuanced representation of "the signature". In addition we demonstrate the merits of the probabilistic GP approach, that increases robustness compared to Simplex Projection and provides interpretable insights to experts.

A neat theoretical property of the signature kernel, that we currently don't fully exploit, is its ability to handle **irregular and multi-variate time series**. This may enable SigCCM to be directly applied to data that other methods are unable to handle. Future experiments are needed to evaluate how SigCCM performs on irregular and also sporadic data and how it compares to Latent CCM [22], which was developed specifically for such requirements. Generally, the GP CCM framework paves the way for future research to explore the use of other elaborate kernel functions, particularly those designed with time series data in mind, like Random Fourier Signature Features [29] or the Global Alignment Kernel [30], or to move to a fully probabilistic approach that outputs posterior distributions over delay embedding positions.

6 Conclusion

We propose SIGCCM, a novel, non-parametric, and probabilistic kernel method for causal inference in dynamical systems. Two methodological enhancements to CCM, the introduction of signature kernels, and using a Gaussian Processes framework, demonstrate considerable progress towards robust causal inference. We show in experiments on simulated, hybrid and real data, that the path signature kernel has stronger capacity to discriminate causal relationships than ordinary CCM. Our method is therefore more robust at identifying complex causal relationships directly from limited time series data. Particularly, we demonstrate the superiority of SIGCCM with respect to CCM and ECCM, who fail to identify causal relationships in the following scenarios: i) distinguish uni-directional from bi-directional influence among synchronous time series, and ii) decipher confounding from direct causation.

While extended benchmarking experiments are needed to determine if SIGCCM's promising results transfers beyond dynamical systems to broader cases of TSCD [5], formulating CCM within a modular Gaussian Process framework opens opportunities to explore how other elaborate time series kernels, e.g. Random Fourier Signature Features [29] or the Global Alignment Kernel [30], may aid causal inference. We believe that particularly interpretable probabilistic causal inference approaches provide a powerful tool to scientists, that can benefit machine learning for scientific discovery.

Acknowledgments and Disclosure of Funding

This research was supported by an Australian Government Research Training Program (RTP) Scholarship, as well as through the Australian Research Council's Industrial Transformation Training Centre in Data Analytics for Resources and Environments (DARE) (project ID IC190100031).

References

- [1] J. Pearl, *Causality: models, reasoning, and inference*, eng, 2nd ed. Cambridge ; Cambridge University Press, 2009, ISBN: 978-0-521-89560-6.
- [2] J. Runge, S. Bathiany, E. Bollett, *et al.*, “Inferring causation from time series in Earth system sciences,” *Nature Communications*, vol. 10, Dec. 2019. DOI: 10.1038/s41467-019-10105-3.
- [3] G. Sugihara, R. May, H. Ye, *et al.*, “Detecting Causality in Complex Ecosystems,” *Science*, vol. 338, no. 6106, pp. 496–500, Oct. 2012, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.1227079. [Online]. Available: <https://www.science.org/doi/10.1126/science.1227079> (visited on 03/22/2024).
- [4] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, “Detecting and quantifying causal associations in large nonlinear time series datasets,” *Science Advances*, vol. 5, no. 11, eaau4996, Nov. 2019, Publisher: American Association for the Advancement of Science. DOI: 10.1126/sciadv.aau4996. [Online]. Available: <https://www.science.org/doi/10.1126/sciadv.aau4996> (visited on 04/11/2024).
- [5] Y. Cheng, Z. Wang, T. Xiao, Q. Zhong, J. Suo, and K. He, “CausalTime: Realistically Generated Time-series for Benchmarking of Causal Discovery,” en, Oct. 2023. [Online]. Available: [https://openreview.net/forum?id=iad1yyyGme&referrer=%5Bthe%20profile%20of%20Yuxiao%20Cheng%5D\(%2Fprofile%3Fid%3D~Yuxiao_Cheng1\)](https://openreview.net/forum?id=iad1yyyGme&referrer=%5Bthe%20profile%20of%20Yuxiao%20Cheng%5D(%2Fprofile%3Fid%3D~Yuxiao_Cheng1)) (visited on 05/08/2024).
- [6] C. W. J. Granger, “Investigating Causal Relations by Econometric Models and Cross-spectral Methods,” *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969, Publisher: [Wiley, Econometric Society], ISSN: 0012-9682. DOI: 10.2307/1912791. [Online]. Available: <https://www.jstor.org/stable/1912791> (visited on 04/11/2024).
- [7] J. Runge, “Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets,” en, in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, ISSN: 2640-3498, PMLR, Aug. 2020, pp. 1388–1397. [Online]. Available: <https://proceedings.mlr.press/v124/runge20a.html> (visited on 05/22/2024).
- [8] A. Tank, I. Covert, N. Foti, A. Shojaiie, and E. Fox, “Neural Granger Causality,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021, arXiv:1802.05842 [stat], ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2021.3065601. [Online]. Available: <http://arxiv.org/abs/1802.05842> (visited on 05/18/2024).
- [9] Y. Cheng, R. Yang, T. Xiao, *et al.*, *CUTS: Neural Causal Discovery from Irregular Time-Series Data*, arXiv:2302.07458 [cs, stat], Feb. 2023. DOI: 10.48550/arXiv.2302.07458. [Online]. Available: <http://arxiv.org/abs/2302.07458> (visited on 04/11/2024).
- [10] Y. Cheng, L. Li, T. Xiao, *et al.*, “CUTS+: High-Dimensional Causal Discovery from Irregular Time-Series,” en, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 10, pp. 11 525–11 533, Mar. 2024, Number: 10, ISSN: 2374-3468. DOI: 10.1609/aaai.v38i10.29034. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/29034> (visited on 04/11/2024).
- [11] J. Runge, X.-A. Tibau, M. Bruhns, J. Muñoz-Marí, and G. Camps-Valls, “The Causality for Climate Competition,” en, in *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, ISSN: 2640-3498, PMLR, Aug. 2020, pp. 110–120. [Online]. Available: <https://proceedings.mlr.press/v123/runge20a.html> (visited on 05/09/2024).
- [12] A. R. Lawrence, M. Kaiser, R. Sampaio, and M. Sipos, *Data Generating Process to Evaluate Causal Discovery Techniques for Time Series Data*, arXiv:2104.08043 [cs, stat], Apr. 2021. DOI: 10.48550/arXiv.2104.08043. [Online]. Available: <http://arxiv.org/abs/2104.08043> (visited on 05/09/2024).

- [13] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara, “Distinguishing time-delayed causal interactions using convergent cross mapping,” en, *Scientific Reports*, vol. 5, no. 1, p. 14 750, Oct. 2015, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/srep14750. [Online]. Available: <https://www.nature.com/articles/srep14750> (visited on 05/18/2024).
- [14] G. Sugihara and R. M. May, “Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series,” en, *Nature*, vol. 344, no. 6268, pp. 734–741, Apr. 1990, Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/344734a0. [Online]. Available: <https://www.nature.com/articles/344734a0> (visited on 03/22/2024).
- [15] C. E. Rasmussen, *Gaussian Processes in Machine Learning*, en, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds. Berlin, Heidelberg: Springer, 2004, ISBN: 978-3-540-28650-9. [Online]. Available: https://doi.org/10.1007/978-3-540-28650-9_4 (visited on 10/01/2023).
- [16] F. Takens, “Detecting strange attractors in turbulence,” en, in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds., vol. 898, Series Title: Lecture Notes in Mathematics, Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381, ISBN: 978-3-540-11171-9 978-3-540-38945-3. DOI: 10.1007/BFb0091924. [Online]. Available: <http://link.springer.com/10.1007/BFb0091924> (visited on 04/11/2024).
- [17] E. H. van Nes, M. Scheffer, V. Brovkin, et al., “Causal feedbacks in climate change,” en, *Nature Climate Change*, vol. 5, no. 5, pp. 445–448, May 2015, ISSN: 1758-6798. DOI: 10.1038/nclimate2568. [Online]. Available: <https://www.nature.com/articles/nclimate2568> (visited on 07/27/2022).
- [18] S. K. Stavroglou, A. A. Pantelous, H. E. Stanley, and K. M. Zuev, “Hidden interactions in financial markets,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 22, pp. 10 646–10 651, May 2019, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.1819449116. [Online]. Available: <https://www.pnas.org/doi/10.1073/pnas.1819449116> (visited on 05/18/2024).
- [19] Z. Benkő, Zlatniczki, M. Stipplinger, et al., *Complete Inference of Causal Relations between Dynamical Systems*, en, Aug. 2018. [Online]. Available: <https://arxiv.org/abs/1808.10806v4> (visited on 05/18/2024).
- [20] A. A. Tsonis, E. R. Deyle, H. Ye, and G. Sugihara, “Convergent Cross Mapping: Theory and an Example,” in *Advances in Nonlinear Geosciences*, A. A. Tsonis, Ed., Cham: Springer International Publishing, 2018, pp. 587–600, ISBN: 978-3-319-58895-7. DOI: 10.1007/978-3-319-58895-7_27. [Online]. Available: https://doi.org/10.1007/978-3-319-58895-7_27.
- [21] Y. Huang, Z. Fu, and C. L. E. Franzke, “Detecting causality from time series in a machine learning framework,” eng, *Chaos (Woodbury, N.Y.)*, vol. 30, no. 6, p. 063 116, Jun. 2020, ISSN: 1089-7682. DOI: 10.1063/5.0007670.
- [22] E. D. Brouwer, A. Arany, J. Simm, and Y. Moreau, “Latent Convergent Cross Mapping,” en, Oct. 2020. [Online]. Available: <https://openreview.net/forum?id=4TSi0TkKe5P> (visited on 04/10/2024).
- [23] A. T. Clark, H. Ye, F. Isbell, et al., “Spatial convergent cross mapping to detect causal relationships from short time series,” en, *Ecology*, vol. 96, no. 5, pp. 1174–1181, 2015, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1890/14-1479.1>, ISSN: 1939-9170. DOI: 10.1890/14-1479.1. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1890/14-1479.1> (visited on 04/11/2024).
- [24] C. Salvi, T. Cass, J. Foster, T. Lyons, and W. Yang, “The Signature Kernel is the solution of a Goursat PDE,” en, *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 3, pp. 873–899, Jan. 2021, arXiv:2006.14794 [cs, math], ISSN: 2577-0187. DOI: 10.1137/20M1366794. [Online]. Available: [http://arxiv.org/abs/2006.14794](https://arxiv.org/abs/2006.14794) (visited on 05/03/2024).
- [25] I. Chevyrev and A. Kormilitzin, *A Primer on the Signature Method in Machine Learning*, en, arXiv:1603.03788 [cs, stat], Mar. 2016. [Online]. Available: [http://arxiv.org/abs/1603.03788](https://arxiv.org/abs/1603.03788) (visited on 05/03/2024).
- [26] V. Venema, F. Ament, and C. Simmer, “A Stochastic Iterative Amplitude Adjusted Fourier Transform algorithm with improved accuracy,” English, *Nonlinear Processes in Geophysics*, vol. 13, no. 3, pp. 321–328, Jul. 2006, Publisher: Copernicus GmbH, ISSN: 1023-5809. DOI: 10.5194/npg-13-321-2006. [Online]. Available: <https://npg.copernicus.org/articles/13/321/2006/> (visited on 05/20/2024).

- [27] E. H. van Nes, B. M. S. Arani, A. Staal, *et al.*, “What Do You Mean, ‘Tipping Point’?” en, *Trends in Ecology & Evolution*, vol. 31, no. 12, pp. 902–904, Dec. 2016, ISSN: 0169-5347. DOI: 10.1016/j.tree.2016.09.011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169534716301835> (visited on 11/07/2022).
- [28] J. Petit, J. Jouzel, D. Raynaud, *et al.*, “Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica,” English, *Nature*, vol. 399, no. 6735, pp. 429–436, 1999, ISSN: 0028-0836. DOI: 10.1038/20859.
- [29] C. Toth, H. Oberhauser, and Z. Szabo, *Random Fourier Signature Features*, en, arXiv:2311.12214 [cs, stat], Nov. 2023. [Online]. Available: <http://arxiv.org/abs/2311.12214> (visited on 05/19/2024).
- [30] M. Cuturi, J.-P. Vert, O. Birkenes, and T. Matsui, “A kernel for time series based on global alignments,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, arXiv:cs/0610033, Apr. 2007, pp. II–413–II–416. DOI: 10.1109/ICASSP.2007.366260. [Online]. Available: <http://arxiv.org/abs/cs/0610033> (visited on 05/22/2024).
- [31] A. Berger and M. F. Loutre, ‘Insolation values for the climate of the last 10 million years,’ *Quaternary Science Reviews*, vol. 10, no. 4, pp. 297–317, Jan. 1991, ISSN: 0277-3791. DOI: 10.1016/0277-3791(91)90033-Q. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/027737919190033Q> (visited on 05/18/2024).

A Appendix

A.1 Computing resources

All experiments were run on a NVIDIA GeForce RTX 4090 GPU with 24 GB memory. We include reproducible PyTorch code in our submission, including a fast, tensorised implementation of CCM as well as an implementation of SIGCCM. Our implementation relies on the sigkernel package <https://github.com/crispitagorico/sigkernel/tree/master> and as well as python code for IAAFT from <https://github.com/mlcs/iaaft/blob/main/iaaft.py>.

A.2 Additional information relating to SYNCHRONY experiment on hybrid data

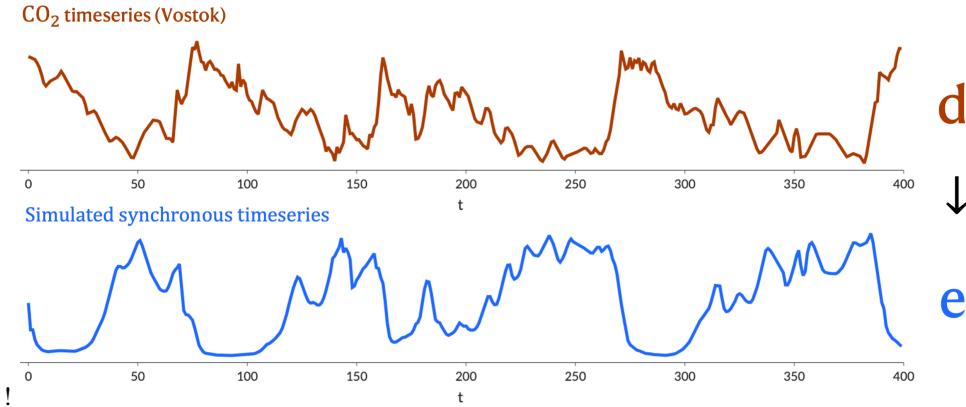


Figure 5: Normalised real data d and simulated time series e used in the SYNCHRONY experiment. Note minor delay in effect e and inverse synchronous relationship between d and e

A.3 Vostok ice core data

We use reconstructions of atmospheric CO₂ concentrations and temperature [28], as well as modelled insolation [31] for the corresponding time period. Our experimental design is largely based on Nes, Scheffer, Brovkin, *et al.* [17]. The insolation time series was obtained from the ORBIT91 dataset [31], which models the incoming solar energy based on earth's eccentricity, obliquity and precession, and reflects mid-month insolation at 65N for July in W/m^2 .

Following [17] we linearly interpolate the CO₂ and temperature data to generate regular time series with 1000 year (1 ky) time steps, correspond to the insolation data, resulting in $N = 401$ data points. All time series are standardised. We evaluate akin to N-cross-validation, however with $n_{\text{train}} < n_{\text{test}}$, periodically shifting the start point of the training data sequence through the time series, and all data points not contained in the training set are used for testing.

All real-world data were accessed through the NCEI Paleo data search tool, <https://www.nci.noaa.gov/access/paleo-search/>, and were selected in accordance with experiments in Nes, Scheffer, Brovkin, *et al.* [17].

The CO₂ data from the Vostok ice core [28] can be directly accessed through the following link: <https://www.nci.noaa.gov/pub/data/paleo/icecore/antarctica/vostok/co2nat-noaa.txt>. The file includes meta data as well as the CO₂ time series in one text file.

The temperature data text file [28] can be accessed through this link: <https://www.nci.noaa.gov/pub/data/paleo/icecore/antarctica/vostok/deutnat-noaa.txt>

The insolation data [31] and can directly be accessed through the following link: https://www.nci.noaa.gov/pub/data/paleo/climate_forcing/orbital_variations/insolation/orbit91. We select column '65NJJul' and all years since 410 BC. The data corresponds to the sixth column ('mid-month insolation 65N for July in W/m^2 ') of file ORBIT91 described in the following meta data file: https://www.nci.noaa.gov/pub/data/paleo/climate_forcing/orbital_variations/insolation/readme_insolation.txt

A.4 Hyper parameters

Continuous hyperparameters like σ_{rbf} in SIGCCM can be optimised by minimising the negative log likelihood [15], where as other parameters like s are determined through processes described in the paper.

SIGCCM

- k : span of sequence (int)
- σ_{rbf} : output variance of the rbf kernel
- s : shift/offset parameter
- n_{train} : length of training sequence

(E)CCM

- f : filter
 - E : (implicit in f)
 - τ : spacing between (regular) active positions of f (implicit in f)
 - k : span of filter (int) (implicit in f)
- s : shift/offset parameter
- n_{train} : length of training sequence