A DOUBLE PENDULUM

Figure 4 presents a graphical representation of a double pendulum with its two masses and two weightless rods. Figure 5 shows examples of trajectories generated by a double pendulum.

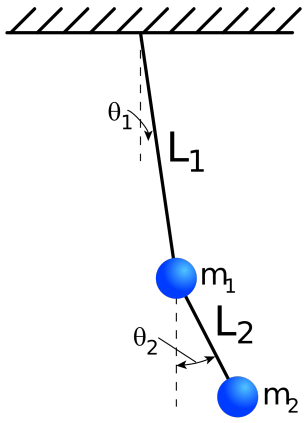


Figure 4: Physical representation of the double pendulum[1]

The double pendulum is a simple physically system that is chaotic and exhibits rich dynamical behavior. The Lagrangian of the double pendulum is

$$\mathcal{L} = \frac{1}{2}(m_1 + m_2)l_1^2\dot{\theta}_1^2 + \frac{1}{2}m_2 l_2^2\dot{\theta}_2^2 + m_2 l_1 l_2 \dot{\theta}_1\dot{\theta}_2\cos(\theta_1 - \theta_2). \tag{4}$$

The corresponding Hamiltonian can be derived using Legendre transform $H = \sum_i \dot{\theta}_i p_i - \mathcal{L}$.

The system evolution can be simulated by integrating the Hamilton equations:

$$\dot{\theta}_i = \frac{\partial H}{\partial p_i}$$
$$\dot{p}_i = -\frac{\partial H}{\partial \theta_i}$$

The Jacobian of the right hand side is

$$J = \begin{bmatrix} \frac{\partial^2 H}{\partial\theta_1\partial p_1} & \frac{\partial^2 H}{\partial^2 p_1} & \frac{\partial^2 H}{\partial p_1\partial\theta_2} & \frac{\partial^2 H}{\partial p_1\partial p_2} \\ -\frac{\partial^2 H}{\partial^2\theta_1} & -\frac{\partial^2 H}{\partial\theta_1\partial p_1} & -\frac{\partial^2 H}{\partial\theta_1\partial\theta_2} & -\frac{\partial^2 H}{\partial\theta_1\partial p_2} \\ \frac{\partial^2 H}{\partial\theta_1\partial p_2} & \frac{\partial^2 H}{\partial p_1\partial p_2} & \frac{\partial^2 H}{\partial\theta_2\partial p_2} & \frac{\partial^2 H}{\partial^2 p_2} \\ -\frac{\partial^2 H}{\partial\theta_1\partial\theta_2} & -\frac{\partial^2 H}{\partial p_1\partial\theta_2} & -\frac{\partial^2 H}{\partial^2\theta_2} & -\frac{\partial^2 H}{\partial\theta_2\partial p_2} \end{bmatrix}.$$

Note that the diagonal elements cancelling in pairs, which results in a trace of zero that indicates the volume-preserving property of the Hamiltonian flow according to Liouville's theorem. This property corresponds to information preservation in nondissipating physical systems. Consequently, a noncoupled double pendulum does not have a proper attractor. However, for a given initial condition, and thus given energy, the possible states still form a densely populated volume in state-space. Applying the nonphysical coupling term, the conservation rule do not hold anymore.

The real part of the eigenvalues of J are called the local Lyapunov exponents.

[1] Source: JabberWok / Wikimedia Commons, CC-by-3.0.

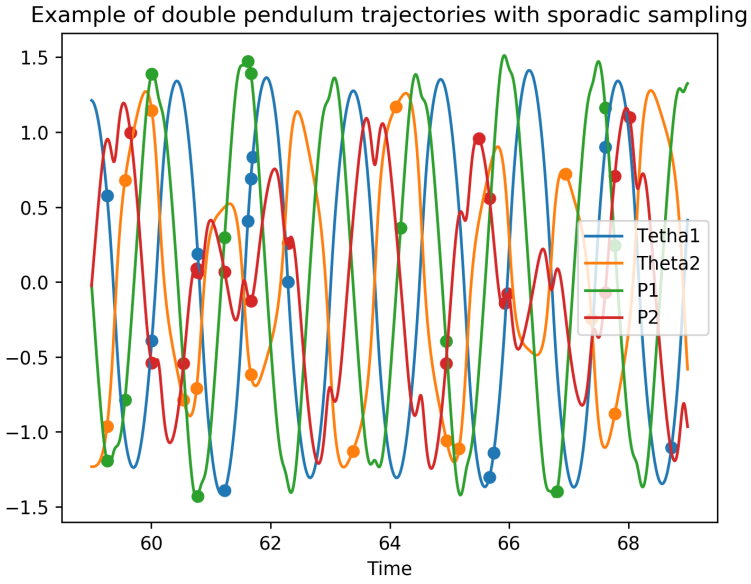


Figure 5: Example of trajectories generated by a double pendulum. The solid lines represent the true process and the dots the sampled measurements.

The direction of the largest expansion evolves as

$$\frac{d\mathbf{q}}{dt} = J\mathbf{q}$$
$$|\mathbf{q}(0)| = 1$$

The largest Lyapunov exponent is given by

$$\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \log |\mathbf{q}(t)|.$$

Note that in stationary processes J is constant, and the differential equation have a closed form solution

$$\mathbf{q}(t) = \mathbf{q}(0)e^{Jt},$$

and the local and global Lyapunov exponents are equal.

The largest Lyapunov exponent can be described intuitively as

$$|\delta(t)| \approx |\delta(0)|e^{\lambda_1 t},$$

where $\delta(t)$ is defined as the difference between two phase-space trajectories, with initial condition infinitesimally close to each other:

$$\mathbf{x}'(t) = \mathbf{x}(t) + \delta(t), t \geq 0$$
$$|\delta(0)| \leq \epsilon.$$

We use numerical integration to compute the largest Lyapunov exponent of the double pendulum, and verify that it is in the chaotic regime.

B INTERACTING NEURON POPULATIONS

The time series is the average membrane potential of two populations of leaky integrate-and-fire neurons with alpha-function shaped synaptic currents (*iaf_psc_alpha*) simulated by NEST-2.20.0 (Fardet

Table 2: Parameters [m, kg] and the largest Lyapunov exponents of the uncoupled pendulums ($\lambda_1 > 0$ indicates chaotic behavior). We report means and their confidence interval over 10 repetitions with initial angles perturbed with $\sigma = 0.05$ normal distributed noise.

SYSTEM		l_1	l_2	m_1	m_2	θ_1	θ_2	λ_1 AND CI (80%)
$X \leftarrow Y$	X	1	0.5	2.0	1.0	1	-0.5	$0.306\,(0.149, 0.468)$
	Y	0.5	1.0	0.5	4.0	1	-0.5	$0.005\,(0.001, 0.010)$
	WHOLE $X \rightarrow Y$ SYSTEM							$0.318\,(0.183, 0.422)$
$X \leftarrow Z \rightarrow Y$	X	0.5	1.0	2.0	1.0	1.0	-0.5	$0.008\,(0.006, 0.009)$
	Y	0.5	1.0	2.0	1.0	1.0	-0.5	$0.008\,(0.006, 0.009)$
	Z	1.0	1.0	1.0	3.0	1.0	-0.5	$0.007\,(0.005, 0.008)$
	WHOLE $X \leftarrow Z \rightarrow Y$ SYSTEM							$0.090\,(0.027, 0.510)$

et al., 2020). Each population contains 100 units with sparse random excitatory synapses inside population, and unidirectionally from population A to population B. A Poisson generator with rate of 40kHz was used to excite the network.

Table 3: Neuron populations. Every non-specified model parameter is left at the default value.

Population	tau_m [ms]	I_e [μA]
A	$\mathcal{U}(15.0, 16.0)$	0.0
B	$\mathcal{N}(15.0, 1.0)$	60.0
C (not obs.)	10.0	0.0

Table 4: Synapses. Parameters have been tuned to achieve stable firing without depolarizing the neuron populations.

From	To	connection type	parameter
Poisson	A	fixed outdegree	outdegree = 10
Poisson	C	fixed outdegree	outdegree = 10
A	A	fixed indegree	indegree = 67
B	B	fixed indegree	indegree = 20
C	C	fixed indegree	indegree = 67
C	B	fixed outdegree	outdegree = 60

In Figure 6, we plot the reconstruction correlations of the coupled neuron populations obtained with the fully observed time series (10 observations per second) and evaluated with standard CCM. We observe a small convergence in the reconstruction in the non-causal direction, indicating potential synchrony. As presented in Table 1, our approach also captures this small reconstruction signal.

C RESULTS WITH UNIVARIATE GAUSSIAN PROCESSES

In Table 5, we present the results of our experiments with univariate Gaussian Processes (GP). In this case, we only learn a GP on the dimension of interest to compute the delay embeddings. As we can see, results are slightly worse than when using multivariate Gaussian Processes (MVGP).

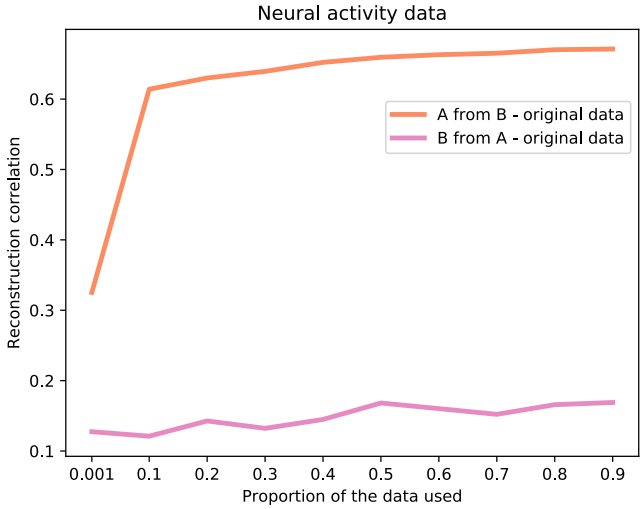


Figure 6: Result of CCM on fully observed neural activity data. Despite a clear signal of A driving B, we observe some positive correlation for the reconstruction in the noncausal direction.

Table 5: Average reconstruction scores $\mathcal{S}_c$ (and their standard deviations) in all directions for the double pendulum and neural activity experiments. Standard deviations are computed using 5 repetitions. Significant correlations compared to noncoupled dynamical systems are in bold ($p < 0.01$). Significance is computed using the Mann-Whitney rank test. Our approach detects the correct causal structure. ✓and ✗ highlight correct and wrong direction detection respectively.

DATA	SPATIAL CCM	GP	MVGP	LATENT-CCM
CASE 1				
$X \to Y$	-0.017 $\pm$ 0.037	0.002 $\pm$ 0.004	0.009 $\pm$ 0.014	0.001 $\pm$ 0.013
$\mathbf{X \leftarrow Y}$	0.018 $\pm$ 0.0056	-0.001 $\pm$ 0.003	-0.001 $\pm$ 0.021	0.055 $\pm$ 0.001
$AUC_{1\to2}$	0.2 (P=0.98)✓	0.6 (P=0.22) ✓	0.66 (P=0.11) ✓	0.55 (P=0.35) ✓
$AUC_{2\to1}$	0.6 (P=0.23)✗	0.44 (P=0.67) ✗	0.44 (P=0.67) ✗	**1 (P<0.001)** ✓
CASE 2				
$X \to Y$	0.488 $\pm$ 0.074	-0.01 $\pm$ 0.008	-0.005 $\pm$ 0.009	0.001 $\pm$ 0.005
$X \leftarrow Y$	0.181 $\pm$ 0.119	-0.000 $\pm$ 0.003	-0.007 $\pm$ 0.012	0.009 $\pm$ 0.014
$X \to Z$	0.054 $\pm$ 0.021	-0.003 $\pm$ 0.003	-0.002 $\pm$ 0.014	0.035 $\pm$ 0.019
$\mathbf{Z \to X}$	0.324 $\pm$ 0.197	0.061 $\pm$ 0.004	0.012 $\pm$ 0.014	0.657 $\pm$ 0.105
$Y \to Z$	-0.071 $\pm$ 0.078	-0.003 $\pm$ 0.003	-0.003 $\pm$ 0.023	0.005 $\pm$ 0.011
$\mathbf{Z \to Y}$	0.101 $\pm$ 0.052	0.039 $\pm$ 0.008	-0.003 $\pm$ 0.016	0.555 $\pm$ 0.109
$AUC_{1\to2}$	**1.00 (P<0.001)**✗	0.21 (P=0.98) ✓	0.31 (P=0.91) ✓	0.78 (P=0.02) ✓
$AUC_{2\to1}$	**1.00 (P<0.001)**✗	0.49 (P=0.53)✓	0.31 (P=0.92) ✓	0.67 (P<0.09) ✓
$AUC_{1\to3}$	**0.98 (P<0.001)**✗	0.35 (P=0.87) ✓	0.61 (P=0.19) ✓	0.79 (P=0.02) ✓
$AUC_{3\to1}$	**0.93 (P<0.001)**✓	0.74 (P=0.03)✗	**0.81 (P=0.01)** ✓	**1.00 (P<0.001)** ✓
$AUC_{2\to3}$	0.26 (P=0.97)✓	0.36 (P=0.85) ✓	0.45 (P=0.63) ✓	0.46 (P=0.62) ✓
$AUC_{3\to2}$	0.79 (P=0.02)✓	0.58 (0.26) ✗	0.43 (P=0.69) ✗	**1.00 (P<0.001)** ✓
COUPLED				
$\mathbf{A \to B}$	0.267 $\pm$ 0.001	0.028 $\pm$ 0.006	0.028 $\pm$ 0.006	0.295 $\pm$ 0.012
$A \leftarrow B$	0.055 $\pm$ 0.003	0.026 $\pm$ 0.010	0.026 $\pm$ 0.010	0.033 $\pm$ 0.012
$AUC_{A\to B}$	**1.00 (P=0.006)**✓	**1.00 (P=0.006)** ✓	**1.00 (P=0.006)** ✓	**1.00 (P=0.006)** ✓
$AUC_{B\to A}$	**1.00(P=0.006)** ✗✗	**1(P=0.006)** ✗	**1.00(P=0.006)** ✗	**1.00(P=0.006)** ✗
INDEPENDENT				
$X \to Y$	-0.012 $\pm$ 0.001	-0.002 $\pm$ 0.008	-0.002 $\pm$ 0.008	-0.006 $\pm$ 0.007
$X \leftarrow Y$	-0.001 $\pm$ 0.001	-0.003 $\pm$ 0.005	-0.003 $\pm$ 0.005	-0.002 $\pm$ 0.008

D FAILURE OF THE GRANGER CAUSALITY FRAMEWORK

To show how the Granger causality framework fails in the general nonlinear dynamical systems case, we consider the following coupled dynamical system:

$$X[t+1] = X[t](a - bX[t] - cY[t])$$
$$Y[t+1] = Y[t](d - eY[t])$$

Following Granger causality, including values of Y for predicting $X[t+1]$ should increase the prediction accuracy, and thus hint towards a causal effect of Y on X. However, dynamics of $X[t]$ can be rearranged such that all information about $Y[t]$ is contained in $X[t]$ already. Indeed,

$$Y[t] = \frac{-1}{c}(\frac{X[t]}{X[t-1]} - a + b)(d + \frac{e}{c}(\frac{X[t]}{X[t-1]} - a + b).$$

Conditioning on $Y[t]$ would not bring additional information and Granger causality would then fail to uncover the right causal structure.

E PROOF OF LEMMA 3.1

Proof. We first write the map $\Phi^k_{g(\phi_H),\alpha_H}(H(t))$ in its full form:

$$\begin{aligned}
&\Phi^k_{g(\phi_H),\alpha_H}(H(t)) : \mathcal{H} \to \mathbb{R}^k \text{ s.t.} \\
&\Phi^k_{g(\phi_H),\alpha_H}(H(t)) = (\alpha_H(g(\phi_{H,0}(H(t)))), \alpha_H(g(\phi_{H,-\tau}(H(t)))), ..., \alpha_H(g(\phi_{H,-k\tau}(H(t))))) \\
&\qquad = (\alpha_H(g(H(t))), \alpha_H(g(H(t-\tau))), ..., \alpha_H(g((H(t-k\tau))))) \\
&\qquad = (\alpha_H(X[t]), \alpha_H(X[t-\tau]), ..., \alpha_H(X[t-k\tau])),
\end{aligned}$$

where the last line follows from the definition of the dynamical system. Φ then maps the latent process to the delay embedding of X obtained with observation function α_H.

As the observation function $\alpha_H \in \mathcal{C}^2$, the flow ϕ_H and the function $g(\cdot)$ are all continuous, this implies that the map Φ is also continuous in $H(t)$. It is also surjective as all delay embeddings (or points in the state-space) will have at least one latent value generating this delay embedding. Indeed, if we write $\mathcal{M}'_{\alpha_H}$ as the shadow manifold of the delay embeddings of X with observation function α_H, we have that

$$\forall m \in \mathcal{M}'_{\alpha_H}, \exists h \in \mathcal{H} \text{ s.t. } \Phi^k_{g(\phi_H),\alpha_H}(h) = m.$$

Let us now assume that there exists a specific observation function α^*_H such that Φ is injective. The map $\Phi_{\alpha^*_H}$ is then bijective. Furthermore, as both $\mathcal{H}$ and $\mathcal{M}'_{\alpha^*_H}$ are endowed with a metric, the map $\Phi_{\alpha^*_H}$ is a homeomorphism between $\mathcal{H}$ and $\mathcal{M}'_{\alpha^*_H}$.

We now show that Φ is a homeomorphism for any observation function. From Takens' theorem, any delay embedding with valid observation function α, dimension k, and delay τ is a valid embedding of the strange attractor of the dynamical system. There must then exists a homeomorphic map Ψ between any two valid delay embeddings with different observation functions:

$$\forall \alpha, \beta \in \mathcal{C}^2, \mathbb{R} \to \mathbb{R}\ , \exists \text{ homeomorphism } \Psi_{\alpha,\beta} : \mathcal{M}'_\alpha \to \mathcal{M}'_\beta \text{ s.t.}$$
$$\forall m_\alpha \in \mathcal{M}'_\alpha, m_\beta \in \mathcal{M}'_\beta, \Psi_{\alpha,\beta}(m_\alpha) = m_\beta.$$

By transitivity, there is now a homeorphism between $\mathcal{H}$ and any valid delay embedding defined with observation function α_H defined as $\Psi_{\alpha^*_H,\alpha_H} \circ \Phi^k_{g(\phi_H),\alpha^*_H}(H(t))$. By Takens' theorem, $\mathcal{H}$ is thus an embedding of the strange attractor of the dynamical system containing $X[t]$.

□

F COMPARISON WITH PCMCI AND VARLINGAM

F.1 PCMCI

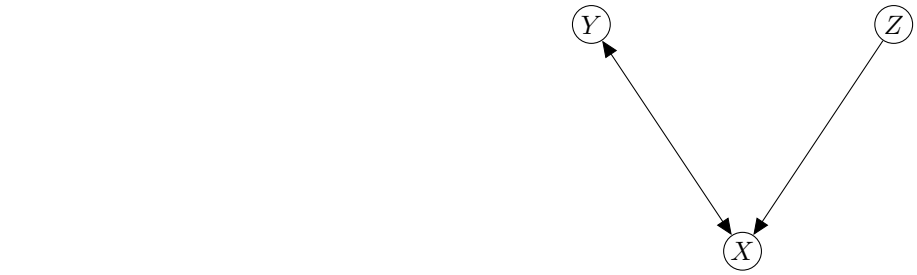
We compared our approach with PCMCI (Runge et al., 2019), a recently introduced method to estimate causal networks from large-scale time series datasets. The method uses independence tests at various time lags to infer causal links between time series. The method does not allow for sporadic time series as a constant time lag is required for the conditional independence tests. Furthermore, the method does not support a way to handle a collection of short time series from a common dynamical system. We then used the method on a less challenging variant of our data where the observations are sampled at a constant rate corresponding to the sampling rate used for generating the sporadic time series. We only feed a very long time series without interruption. We used a maximum time lag of 10 seconds and report the results of cases 1 and 2 of the double pendulum for various significance thresholds in Table 6. We used the implementation of the method provided by the authors at `https://github.com/jakobrunge/tigramite/`. Because the method infers causality with a different score than ours, we report the inferred configuration at each repeat and for each case. The different configurations are shown on Figure 7. For case 1, we observe that PCMCI recovers the true causal graph 2 times out of 5 when the significance is set to $p < 0.001$. For case 2, at all levels of confidence, PCMCI infers a fully connected graph (configuration Δ). We suggest this results form the large number of time series for this configuration (12) as well as the existence of complex coupling dynamics of the chaotic dynamical system, make the causal inference challenging. An example of inferred causal network with PCMCI (case 1 with significance level of $p < 0.001$) is presented on Figure 8.

F.2 VARLINGAM

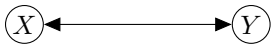
We also compared the Latent CCM method with VARLinGAM (Hyvärinen et al., 2010). VAR-LinGAM detects causal links between longitudinal variables by learning a directed acyclic graph of interactions of the variables and their time lags. In particular, VARLinGAM derives the best acyclic graph with the LinGAM method (Shimizu et al., 2006). We infer causality between the group of variables X and another Y by checking the existence of causal edges between individual variables of X and Y. We used the implementation of VARLinGAM available at `https://github.com/cdt15/lingam` and use a maximum time lag of 10 seconds (same as for PCMCI) and a minimum causal weight of 0.01. As for PCMCI, the score provided to infer a graph is different than ours and we provide the results of the learnt causal graphs for cases 1 and 2 of the double pendulum in Table 6. The method only recovers the true graph of case 1 in 60% of the time. For case 2, the method fails to recover the causal generative model for all the repeats.

Table 6: Inferred causal configurations for double pendulum cases with PCMCI and Var-LinGAM. Details of the configuration codes used are given on Figure 7. Sign. Level stands for significance level.

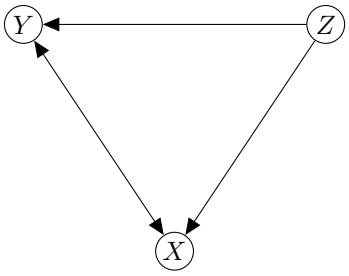
MODEL	CASE	SIGN. LEVEL	REPEAT 1	REPEAT 2	REPEAT 3	REPEAT 4	REPEAT 5
PCMCI	CASE 1	$p < 0.01$	A (✗)	A (✗)	A (✗)	A (✗)	A (✗)
		$p < 0.001$	B (✗)	Γ (✓)	B (✗)	B (✗)	Γ (✓)
	CASE 2	$p < 0.01$	Δ (✗)	Δ (✗)	Δ (✗)	Δ (✗)	Δ (✗)
		$p < 0.001$	Δ (✗)	Δ (✗)	Δ (✗)	Δ (✗)	Δ (✗)
VARLINGAM	CASE 1	$p < 0.01$	A (✗)	A (✗)	A (✗)	A (✗)	A (✗)
		$p < 0.001$	A (✗)	A (✗)	Γ (✓)	Γ (✓)	Γ (✓)
	CASE 2	$p < 0.01$	A (✗)	B (✗)	B (✗)	B (✗)	Γ (✗)
		$p < 0.001$	Γ (✗)	Γ (✗)	A (✗)	Γ (✗)	Γ (✗)



A configuration

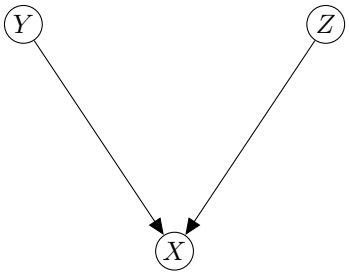


A configuration

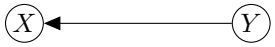


B configuration

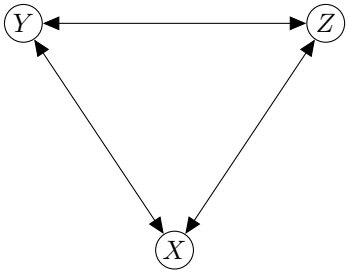
B configuration



Γ configuration



Γ configuration



Δ configuration

(a) Case 1

(b) Case 2

Figure 7: Different configurations inferred by PCMCI in both double pendulum cases.

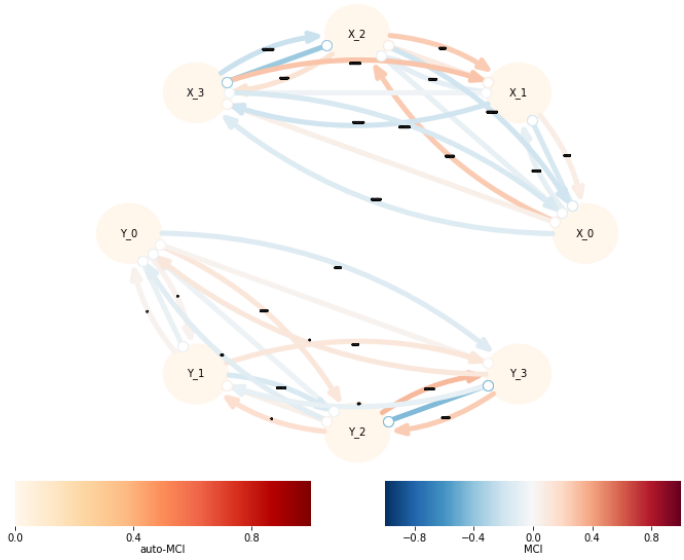


Figure 8: Example of causal network inferred by the PCMCI method for case 1 (repeat 1) of the double pendulum data. Significance level : $p < 0.001$.

G MODEL SELECTION PROCEDURE

When training GRU-ODE-Bayes models, we use 80% of the available time series samples as training set and the remaining 20% as validation set. For training, we use the reconstruction loss proposed by the authors (De Brouwer et al., 2019). For validation, we feed the time series until half of the available horizons (*e.g.* for time series of length 10 seconds, we feed the 5 first seconds) and compute the MSE on the reconstruction of the subsequent available samples. We choose the model hyperparameters that minimize the MSE over the validation set. Note that we do not need test set as our ultimate goal resides in causal direction inference and not in accurate forecasting of the time series. Importantly, models for each time series are learnt independently and no information about causal direction is available at any time in the process.

H SPORADIC DATA WITH MISSINGNESS NOT AT RANDOM

Experiments and results presented in Section 4.5 consider a a random sporadic sampling of the data. The data is thus missing at random (MAR). In practice, however, the sampling of a process is usually not fully random but rather depends on the value of the process itself. As a simple example, doctors measure the blood pressure of patients more often when it's high or likely to be high. The sampling pattern then gives information about the value of the process we want to model.

In order to account for this bias occurring in practice, we consider also a variant of the double pendulum dataset (case 1) where the missingness is not at random (MNAR) and with noise standard deviation of 0.01. The sampling pattern we consider is the following. If the absolute value of the angle of the first rod θ_1 is larger than $\frac{\pi}{4}$, we sample the process with a probability two times larger than if the angle is smaller than $\frac{\pi}{4}$. The sampling probability of an observation $p_s(X_t)$ is then :

$$p_s(X_t) = \begin{cases} p & \text{if } \theta_1(X_t) \leq \frac{\pi}{4} \\ 2p & \text{if } \theta_1(X_t) > \frac{\pi}{4} \end{cases}$$

Note that the total number of observations is still kept constant with respect to the MAR case.

Figure 9 shows the results of latent CCM on the MNAR double pendulum data (case 1). We observe that latent CCM is still inferring the correct causal directions, despite the sampling bias.

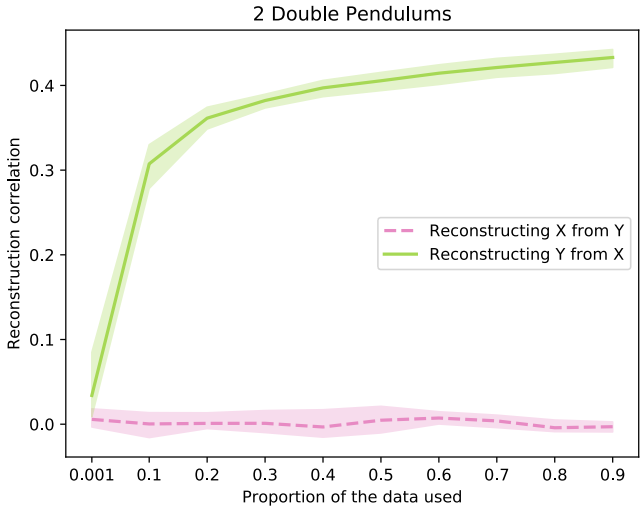


Figure 9: Result of latent CCM on the MNAR double pendulum data (case 1 with low noise). The correct causal directions are inferred.

Table 7: Results of latent CCM on the MNAR double pendulum data (case 1 with low noise).

	$X \to Y$	$Y \to X$
$\mathcal{S}_c$	-0.009 ± 0.008	0.399 ± 0.029
AUC	0.49 (p=0.52)	**1 (p<0.001)**

I USING A RNN INSTEAD OF A NEURAL-ODE

As Equation 3 in the paper suggests, we consider our observations are generated from a continuous latent process $H(t)$. Different techniques could be used to infer this process $H(t)$ in our latent CCM approach. Among those techniques, neural-ODE models such as [1] or [2] embody the assumptions of Equation 3 and are thus a natural choice for the inference of informative latent vectors. Another choice could be to learn those dynamics with a non-continuous recurrent neural network approach. In this section, we compare Neural-ODE methods with using a standard recurrent neural network (GRU) for learning the dynamics of the process. Because the processing of missing data across dimensions is not well defined for GRU, we use a version of the data as in Appendix F, namely the observations are sampled at a constant rate (similar to the one used for sporadic data, taking into account the sampling across dimensions) and importantly, no missing dimensions are allowed. We then use the learnt latent process to infer causal direction for case 2 of the double pendulum data. Results are presented in Table 8. We observe that incorrect causal directions are inferred (from X to Z and from Y to Z).

Some theoretical properties of Neural-ODEs can help explain this result. Because of their continuous resolution, Neural-ODEs allow to have a denser coverage of the attractor we want to reconstruct. This feature is further strengthened by the fact that different integrators can be used to recover the latent process, therefore allowing to tune the resolution of the learnt latent process. In the case of physical systems, a symplectic integrator can also be used, to ensure conservation of energy and more accurate learning of the dynamics.

Table 8: Results of latent CCM on the double pendulum data with latents learn from a GRU. Data is constantly sampled with no missing values and simulated as for the case 2 of the double pendulum.

	$X \to Y$	$Y \to X$	$X \to Z$	$Z \to X$	$Y \to Z$	$Z \to Y$
$\mathcal{S}_c$	0.013 ± 0.020	0.022 ± 0.019	0.108 ± 0.049	1.064 ± 0.057	0.053 ± 0.011	0.513 ± 0.018
AUC	0.34 (p=0.884)✓	0.41 (p=0.753)✓	**1.00 (p<0.001)✗**	**1.00 (p<0.001)✓**	**1.00 (p<0.001)✗**	**1 (p<0.001) ✓**

J COMPLEXITY OF THE METHOD AND BASELINES

As all methods require a k-nearest neighbors step for each pair of time series, the difference in computation arises in the computation of the embeddings fed to the kNN. We then report the complexity of computing the embeddings to be used for the cross-mapping in Table 9. Computing delay embeddings scales linearly in the number of embedding dimensions (H) and the number of samples in each time series (M). When using Gaussian Processes, one has first to infer the latent process at all time points and invert a covariance matrix of size $D \times D$ which requires an additional $\mathcal{O}(D^3 \times M)$. For latent-CCM, we avoid the computation of the delay embeddings but we require to train a Neural-ODE model which requires $\mathcal{O}(H^2)$ for each time step and at each observation.

Table 9: Time complexity for computing the embeddings in the different methods. Complexities depend on the embedding dimension (H), the number of samples observed per time series (M), the length of each time series (T) and the number of features in the time series (D).

Method	Time complexity
Multi-spatial CCM	$\mathcal{O}(H \times M)$
GP	$\mathcal{O}(H \times T + M \times D^3)$
MVGP	$\mathcal{O}(H \times T + M^3 \times D^3)$
Latent-CCM	$\mathcal{O}(H^2 \times T + D \times H \times M + H^2 \times M)$