Exact Inference of Causal Relations in Dynamical Systems

Zsigmond Benkő[1,2], Ádám Zlatniczki[1,3], Dániel Fabó[4], András Sólyom[4], Loránd Erőss[5,6], András Telcs[1,3,7], and Zoltán Somogyvári[1,8,*]

[1]Department of Computational Sciences, Wigner Research Centre for Physics of the Hungarian Academy of Sciences, H-1121, Hungary
[2]János Szentágothai Doctoral School of Neurosciences, Semmelweis University, H-1085, Hungary
[3]Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, H-1111, Hungary
[4]Epilepsy Center "Juhász Pál", National Institute of Clinical Neurosciences, Budapest, H-1145, Hungary
[5]Department of Functional Neurosurgery, National Institute of Clinical Neurosciences, Budapest, H-1145, Hungary
[6]Faculty of Information Technology and Bionics, Péter Pázmány Catholic University, Budapest, H-1083, Hungary
[7]Department of Quantitative Methods, University of Pannonia, Faculty of Business and Economics, H-8200, Veszprém, Hungary
[8]Neuromicrosystems ltd., Budapest, H-1113, Hungary

September 6, 2018

Abstract

From philosophers of ancient times to modern economists, biologists and other researchers are engaged in revealing causal relations. The most challenging problem is inferring the type of the causal relationship: whether it is uni- or bi-directional or only apparent - implied by a hidden common cause only. Modern technology provides us tools to record data from complex systems such as the ecosystem of our planet or the human brain, but understanding their functioning needs detection and distinction of causal relationships of the system components without interventions. Here we present a new method, which

1

distinguishes and assigns probabilities to the presence of all the possible causal relations between two or more time series from dynamical systems. The new method is validated on synthetic data sets and applied to EEG (electroencephalographic) data recorded in epileptic patients. Given the universality of our method, it may find application in many fields of science.

Causality is one, if not the most fundamental pillar of science. Still, detection of causal relations between deterministic dynamical system, solely based on observations (without interventions) is a challenge. Though several methods have been proposed, non of them provides exact and complete solution.

Predictive causality One branch of standard causality analysis tools grew out of the predictive causality principle phrased first by Norbert Wiener[1]. Consider two time series and a model, which predicts a time series based on its recent past. If the inclusion of the recent past of the other time series enhances the prediction of the first one's, we say that the included time series causes the first one in the Wiener sense. This principle was first implemented by Granger[2], extended later to nonlinear models and non-parametric methods such as transfer entropy[3].

The Granger method, based on the predictive causality principle and designed to analyze stochastic time series, became very popular in many fields of science. At the same time there are several indications that in case of deterministic dynamical systems it can produce false results[4].

Topological causality. An entirely new approach for causality analysis has been introduced by George Sugihara[5] (followed by Tajima[6] and later refined by Harnack[7]).

The basis of Sugihara's convergent cross mapping (CCM) method goes back to Takens' embedding theorem[8]. Takens' theorem states that the topology and all the degrees of freedom of a dynamical system can be reconstructed from a time series observation by time delay embedding. A key idea of the CCM method is that the consequence contains an observation (in Takens' sense, a lower dimensional smooth function of the state) of the cause, thus all the degrees of freedom of the cause can be reconstructed from the consequence.

While Sugihara's CCM method is able to detect circular causal relations, distinguishing a direct causal link from a hidden common cause is still very challenging. Both Sugihara[5] and Harnack[7] note that observed correlation without inferred direct causation indicates a common driver, however, a common driver does not necessarily result in linear correlation between the two driven systems. Thus, the CCM method is able to detect directed and bidi-

rectional causal relationships, but can not properly infer all cases of hidden common cause as we demonstrate it with three nonlinearly coupled logistic maps (Extended Data Fig. 1.) (further elaborations on the applicability can be found in[9–11]).

Recurrence maps are also applied for causality detection. Hirata et al.[12] inferred the existence of a common cause by rejecting both independence and direct dependence based on recurrence maps. The method, however, can not provide quantitative detection of causal relations given that it uses classical hypothesis tests which provide only unidirectional implication. As a result independence and common cause is detected by the method of Hirata[12], but direct causality only on the "can not be rejected" branch of their test. In a consecutive paper Hirata[13] propose majority voting of three causality detection methods (convergence cross mapping[5], recurrence maps[12] and transfer entropy[3;14]) to fill that gap, which seems to be a very promising new direction, although lacks a quantitative framework for the comparison of the different methods.

The method we propose, the Dimensional Causality (DC) method, grows out from Takens' topological theory of dynamical systems[8], as well as information theory and dimension theory[15–18]. In the followings we present our DC method and its applications to time series of different types. We demonstrate that it is novel and unique in its unified nature and that it has the capability to detect and assign probabilities to all types of causal relations, in particular it is able to detect a hidden common cause for dynamical systems.

To our best knowledge the DC method is the first exact one which detects and distinguishes all possible causal relation of deterministic dynamical systems.

Inferring causality from manifold dimensions

Inferring causal relations using manifold dimensions is possible and follows from Takens theorem. Given a time series, we can perform time delay embedding and measure/estimate the dimensionality of the reconstructed dynamics. Takens' theorem ensures that the dimension of the embedded manifold is the same as the dimension of the attractor in the original state-space and invariant against the observation function and dimension of the embedding space (provided it is sufficiently high dimensional).

Several dimensionality notions have been introduced to capture key characteristics of attractors of dynamical systems. These investigations led to the applications of different fractal type dimensionality notations such as Rényi information dimension, correlation dimension and many kinds of intrinsic

dimensions[15;19–22].

As an example of the possible use of manifold dimensions for causal inference, let us consider two simulated, unidirectionally coupled dynamical systems and their reconstructed attractors depicted in Fig. 1 A. One can observe that the dimensions of the two embedded manifolds are different. The dimension of the consequence is greater than the dimension of the cause, the consequence 'contains' the degrees of freedom of the cause. The detection of the difference between the dimensionality of the cause and the consequence provides an opportunity to set a new analysis method which is able to distinguish all forms of causal relations.

The key variable in our framework is the joint dimension D_J of two time series. Given two simultaneous time series one can get D_J by forming the direct product of the two embedded spaces and measure the (intrinsic) dimension in the resulting point-cloud (c.f. SI section 1.1). Asymmetry of manifold dimensions (measured in the subspaces) does not imply causation, but the additional information provided by D_J is sufficient to determine the type of causal relation between the two systems.

Causal and dimensional relations.

In the case of two independent dynamical systems, the joint dimension equals to the sum of the dimension of the two independent systems - however, any interdependence makes the manifold dimensions sub-additive.

As the cause can be reconstructed from the consequence, the information content of the cause is already available in the embedded manifold of the consequence, therefore in case of unidirectional coupling the joint dimension will be equal to the dimension of the consequence.

Note that the dimension of the driven dynamical system (consequence) is always greater than or equal to the dimension of the driver dynamical system (cause) and the dimension of the two manifolds unequivocally determines the direction of the possible causal effect: only the lower dimensional system can have unidirectional causal effect on the higher dimensional one.

There is a special case when the joint dimension is equal to the dimension of both time series: circular causality or full determinism. According to Takens' theorem a circular case means that homeomorphisms exist in both directions, therefore the two manifolds are topologically equivalent, having equal dimensions.

If the joint dimension is less than the sum of the single dimensions, but not equal to either of them, that is the sign of an existing hidden common cause without direct causal effect between the two time series.

In summary, the relation between the dimensions of the two systems and the joint dimension distinguishes the four possible causal scenarios as follows:

Independent case

$$X \perp Y \iff D_X + D_Y = D_J \tag{1}$$

Unidirectional case

$$X \to Y \iff D_X < D_Y = D_J \tag{2}$$

Circular case

$$X \leftrightarrow Y \iff D_X = D_Y = D_J \tag{3}$$

Common cause

$$X \curlyvee Y \iff \max(D_X, D_Y) < D_J < D_X + D_Y \tag{4}$$

The implications in Eqs. (1) - (3) are straightforward from causal relations to dimensional relations, the common cause case and the reasoning in the opposite direction follows from discretizing state variables otherwise further assumptions are required (see SI section 1.1). Thus, in order to infer the causality we estimated these dimensions using the method proposed in Farahmand[22].

Since, we have only estimates for the manifold dimensions based on finite datasets, we can not show exact equalities in Eqs. (1) - (4), but we assign probabilities to these causal cases.

The posterior probabilities of the dimensional relationship of the r.h.s of Eqs. 1-4 are inferred by comparing the estimates of joint dimension (D_J) to the dimensions of the individual embedded time series $(D_X$ and $D_Y)$.

The inference is demonstrated by the following scheme. Let A be the causal relation, which unequivocally determines the relation between dimensions according to Eqs. 1-4, and also let $\bar{d}$ denote the observed dimension vector, formed by the $(\bar{d}_X, \bar{d}_Y, \bar{d}_J, \bar{d}_Z)$ observations. First we apply Bayes' Theorem

$$P(A|\bar{d}) = \frac{p(\bar{d}|A)}{p(\bar{d})} P(A),$$

then assume a non-informative prior over the possible causal relations (A) and finally calculate the conditional likelihood of the observed dimensions as

$$p(\bar{d}|A) = \int p(\bar{d}|w) dP(w|A),$$

where the integrand w runs over the possible dimension combinations, determined by $dP(\cdot|A)$. For details see SI section 1.2.

Dimension estimates are sensitive for parameter choices (e.g. embedding dimension, see SI section 2.1) and affected by unavoidable biases. The consequence of the bias caused by the embedding dimension can be decreased if the manifolds are embedded into spaces with same embedding dimensions. Along this principle, the independent case was represented by a manifold Z with same embedding dimension as J. Z is constructed by joining X and the time-permuted (independalised) Y^* manifolds.

In order to further increase the precision of the dimension comparisons, the embedding dimensions of J and Z were matched to the embedding dimensions of X and Y by introducing new embeddings $J' = aX + Y$ and $Z' = aX + Y^*$ (where a is a properly chosen irrational number) instead of J and Z. Note that J' and Z' are based on generic observation functions of the original systems, thus they are topologically equivalent to the original system as well in almost all cases (for the proof see SI section 1.1).

There are some situations of system interactions that require further thoughts including the case of full determinism, the hidden common driver and the transitivity of causal relations.

In case of full determinism, also called generalized synchrony, the caused time series are fully determined by the cause. In this case there is no point in considering them as different systems, we have only two copies/observations of the same system (maybe with a time delay). Thus, this particular unidirectional coupling can not be distinguished from the circularly coupled case: the reconstructed topologies are equivalent in both cases.

There are particular cases, when the common driver remains hidden. If the common cause coexists with a direct connection at least in one direction, then the direct connection ensures that all the information of the cause is present in the joint, thus the direct connection(s) will be detectable, while the common driver remains hidden.

Theoretically, causality is a transitive relation. Thus, indirect causal relations (through a chain of direct connections) should be identified as direct causal relation as well. However in real world applications the additive noise can result in non-transitive directed causal relations between multiple systems.

Results

We validate our method on data from three simulated dynamical systems where ground truths are known. Following this verification, the method is also applied to EEG data of epileptic patients. The three simulated dynamical systems have fundamentally different dynamics and pose quite different

challenges that causal analysis has to overcome: the coupled logistic maps are discrete-time chaotic dynamical systems with no significant temporal autocorrelation, in contrast the coupled Lorenz-systems are defined in continuous time and show smooth temporal autocorrelation, while the system of Hindmarsh-Rose models exhibit steep spikes and quasi periodic behavior.

Logistic maps. We simulated systems of three coupled logistic maps with various connectivity patterns: unidirectional, circular, independent and two uncoupled maps driven by a third unobserved one:

$$x_i[t+1] = rx_i[t](1 - x_i[t] - \sum_{j \neq i} \beta_{i,j} x_j[t]), \tag{5}$$

where $i \in \{1, 2, 3\}$ for the three variables, $r = 3.99$ and $\beta_{i,j}$ are coupling coefficients (values for the different simulated couplings are given in SI section 2.2).

The DC method was able to reveal the original coupling pattern between the observed logistic maps for all cases, in particular it was able to detect the existence of the hidden common cause observing only the two affected logistic maps (Fig. 1).

Lorenz systems We tested our method on differently coupled Lorenz systems[23] (Fig. 2).

We conclude that in each case our method detects the proper causal relationship with high confidence. For further details see SI 2.2.

Hindmarsh-Rose systems We analyze causal relationships between coupled Hindmarsh-Rose systems which were originally proposed to model spiking or bursting neuronal activity[24]. In our simulations we use two electrically coupled neurons where coupling is achieved through the membrane potential, as proposed in[25]. Analysis of such time series is quite difficult due to their sharp dynamical changes.

Fig. 3 shows our results on different Hindmarsh-Rose systems. We conclude that in each case our method detects the proper causal relationship with high confidence. For further details about the Hindmarsh-Rose systems and model parameters we refer the reader to SI 2.2.

Changes of inter-hemispheric connectivity during photo-stimulation

Finally, we set out to test our approach under real-world conditions, where the true dimensionality of the systems and the properties of the noise are unknown. In general, the exact causal relationships between time series in these systems are not known. However, external drive can induce changes in the internal causal relationships that can be detected using our analysis method. In particular, the standard epilepsy-diagnostic photo-stimulation procedure (patients exposed to flashing light at different frequencies in a
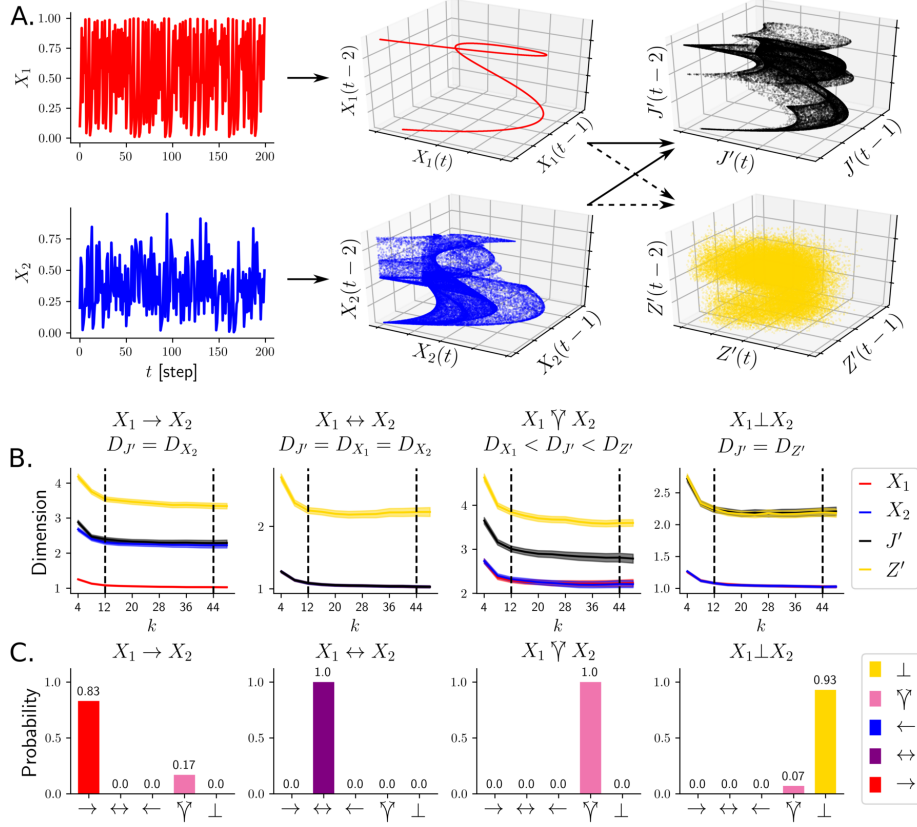
Figure 1: *The workflow and testing of the Dimensional Causality method on coupled logistic map systems (see Eq. 5).* (A) The state spaces of the systems are reconstructed by time delay embedding of the two time series X_1 (red) and X_2 (blue), resulting in the red and blue manifolds. Then joint of the two dataset J' and their time-shuffled version Z' are also embedded, resulting in a reconstruction of the joint state space of the two subsystems (black manifold) and their independent joint (yellow). On (B,C) the test of the DC method on the four simulated examples of five possible causal interactions (one of the unidirectional: $X_1 \rightarrow X_2$, circular: $X_1 \leftrightarrow X_2$, unidirectional - back: $X_1 \leftarrow X_2$, common cause: $X_1 \curlyvee X_2$, independence: $X_1 \perp X_2$) are demonstrated. (B) The intrinsic dimensionality of each manifold is estimated for different neighborhood sizes k. The plateau of dimension-estimates identifies where the estimates can be considered reliable (between dashed lines). Note the match between the actual causal and dimensional relationships: the dimension of the joint manifold (J) relative to the others. (C) Posterior probabilities of the possible causal relationships. The method correctly assigned the highest probability to the actual causal relation in each case.
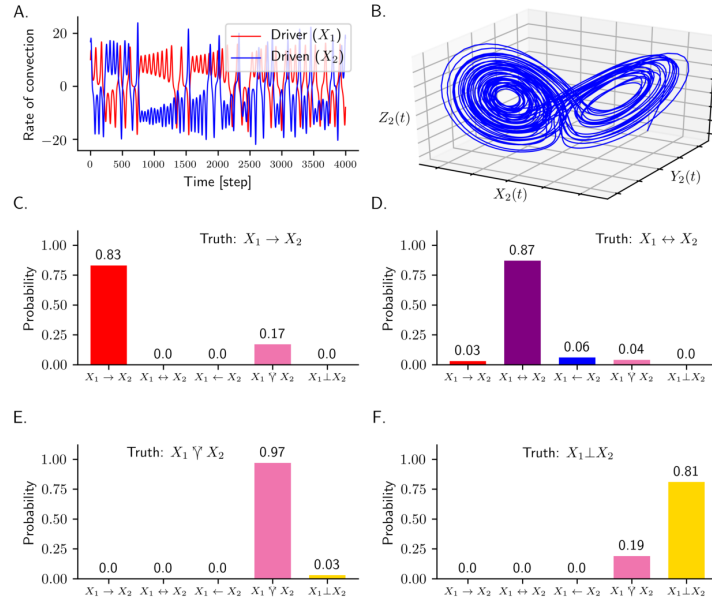
8

Figure 2: *Testing the DC method on coupled Lorenz systems.* (A) The X variables of two Lorenz systems, represented as time series. There is unidirectional coupling from the first system (the driver, red) to the second (the driven, blue). The description of the Lorenz systems is given in the supplementary information. (B) The attractor of the driven system is only slightly perturbed by the driver. (C)-(F) Model probabilities nicely match with the truth for the different couplings. (Color code matches Fig. 1 C.)
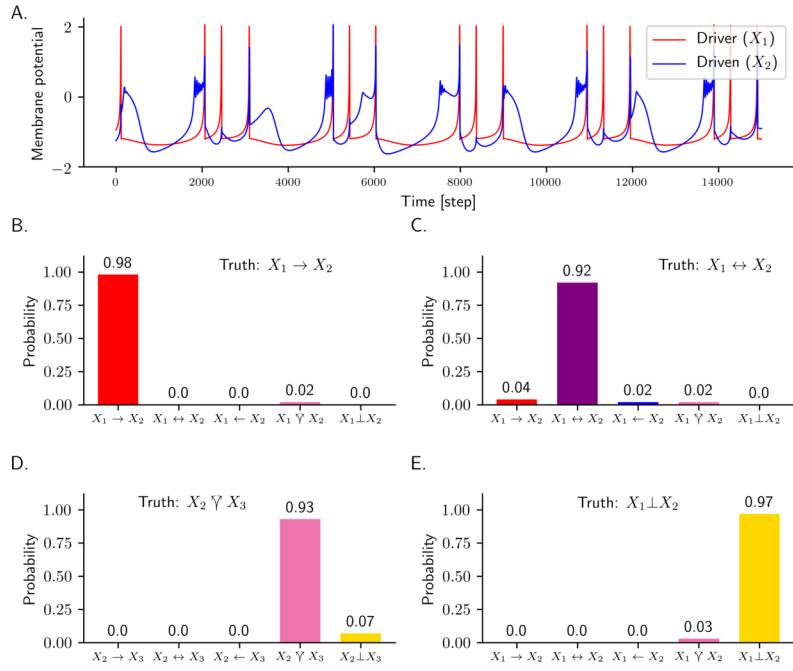
Figure 3: *Testing the DC method on coupled Hindmarsh-Rose systems.* (A) The membrane potentials (X_1, X_2) of unidirectionally coupled Hindmarsh-Rose systems. (B-E) The inferred probabilities for the different types of couplings. (Color code matches with Fig. 1 C.)
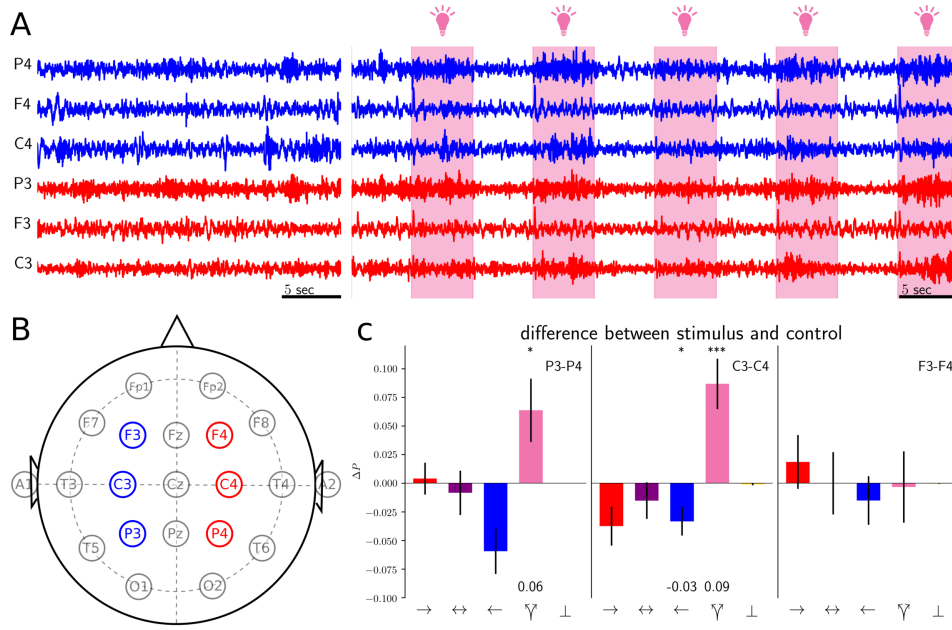
10

Figure 4: *Inter-hemispherical interactions during photo-stimulation.* (A) CSD signal in control condition and photo-stimulation periods (light bulbs) at the six analyzed recording-channels and (B) electrode positions on the scalp. Causal relations were computed between P3-P4, C3-C4 and F3-F4 channel pairs. (C) Difference in probabilities of causal relations between stimulation and control (mean and SE). The probability of the existence of common cause is significantly higher during stimulation periods for P3-P4 ($p = 0.024$) and C3-C4 ($p = 0.0002$) channel-pairs but not for F3-F4 ($n = 87$).

standardized test) is an ideal model for external common cause affecting the two brain hemispheres.

The connectivity between these brain regions suggests that visual information arrives first to (O1, O2 electrodes) and spreads to the parietal (P3-P4) and central (C3-C4) area (Fig. 4 B) afterwards. Hence, it is expected that the DC method is suitable to detect the visual stimulus as a common cause between the affected electrode sites. Thus, DC method was applied to Current Source Density (CSD, see SI 2.2) calculated from EEG recordings of 87 patients participating in the photo-stimulation task (Fig. 4 A).

The probability of common cause (γ) was significantly increased during visual stimulation periods for P3-P4 and C3-C4 channel-pairs relative to the resting state but not for the frontal areas (F3-F4) (Fig.4 C, SI section 2.2).

Causal connections during epileptic seizure

Our next example shows how the identification of the causal connection between the investigated areas can contribute to the diagnosis and surgical treatment of epileptic patients. Our investigations led to self-consistent results, supporting our assumption that the DC method is applicable in such complex situations as well.

The EEG data from a 20-year-old patient, suffering from drug resistant epilepsy with frequent seizures, was analyzed. As part of pre-surgical examination, a subdural grid and 2 strip electrodes were placed onto the surface of the brain, which allowed the identification of the brain areas participating in the seizure activity with high spatial precision.

The seizures showed a variable and complex picture, where most of high frequency seizure activity were observable on the fronto-basal (Fb) and the frontal (Fl1) region (Fig. 5 C). The fronto-lateral (Fl2) region was impacted only moderately by the seizures and the infero-parietal (iP) region showed irregular high amplitude spiking activity during both interictal and ictal periods (Fig. 5 A).

We investigated all the causal relations between the four brain regions and found variable connectivity. The DC analysis showed two main connectivity types during seizures (from analysis of $n = 18$ seizures, Ext. Dat. Fig. 3, see SI 2.2). In the first main type iP was found to be the driver ($n = 6$), while the existence of a common cause between Fb, Fl1 and Fl2 got the highest probability during the pairwise analysis (Fig. 5 B, E, F). In the second main type of seizures hidden common cause was detected between all the observed channels ($n = 10$, Fig. 5 C), which indicates the possible existence of a driver which was out of the region of investigation. In contrast, Fb was found to be dominant driver node during normal interictal activity (Fig. 5 D, Ext. Dat. Fig. 4, SI section 2.2).

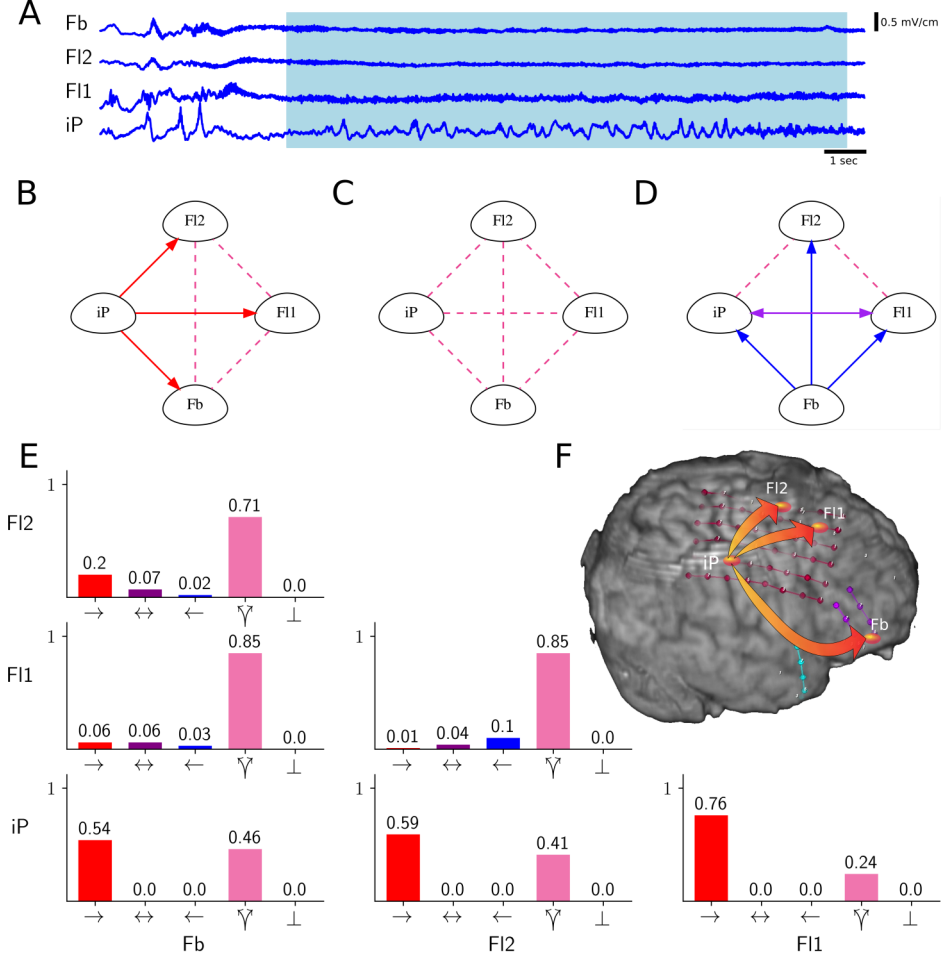Given the difficult accessibility of the iP region, the medical panel came

Figure 5: *Cortical connectivity during epileptic seizure* (A) CSD signal at fronto-basal, frontal (Fl1) fronto-lateral (Fl2) and infero-parietal (iP) areas, the blue selection shows the analyzed time period of the seizure. Basically two types of connectivity were detected for seizures and a third type of connectivity for interictal conditions (B-D). (B) Maximal aposteriory probability (MAP) causal connection structure for the example seizure in A. Red arrows mark unidirectional relations and the pink dashed lines mark detected common cause relations. (seizure type 1, $n = 6$) (C) MAP Causal connection structure for seizure type 2 ($n = 10$) (D) MAP mean causal connection structure interictal sections ($n = 16$). Blue arrows mark unidirectional relations and purple arrows denotes circular causal relations. (E) Causal relation probabilities of the seizure showed on A and B. (F) The inferred driver and driven areas represented on the brain surface for the same seizure as on A, B, E.

13

to the conclusion to resect the frontal and the fronto-basal region and the less active areas were left intact. The patient was seizure free for 1 year, before reinstated.

The DC analysis highlighted the variability of the seizures as well as the existence of a possible hidden common cause behind the observed seizure activity, which could form a basis for the cortical reorganization and the recovery of the epileptic activity in case of the investigated patient. However, systematic analysis on large population of patients will be necessary to clarify the underlying causal structures during epileptic network activity.

Conclusions

We presented the Dimensional Causality method, which is the first unified way to detect all types of causal relations in dynamical systems . In particular, it can reveal the existence of hidden common cause and quantifies the probability of each causal relationship type. The DC method was validated on synthetic data and self-consistency and practical applicability have been demonstrated on human EEG data from epileptic patients.

The precision of our inference is based on the reliability of dimension estimations, which is affected by the choice of embedding parameters and a couple of signal properties such as autocorrelation, sample size, noise level and manifold geometry, especially curvature.

We proposed a systematic method to find good parameter settings (SI section 2.1, and see also[20;21]).

The longer the signal's temporal autocorrelation, the more data points are required to correctly sample the manifold's surface (in the reconstructed state-space) in order to infer causal relations. Our simulations indicate that the proper determination of dimensions required typically at least a few thousand samples, which is more than the data CCM needs to detect directed causal relations under optimal conditions. While being operational at smaller sample sizes, CCM is less reliable in the identification of hidden common cause, and lacks assigned probabilities of causal cases. At the price of being more data intensive, DC has the capability to solidly detect hidden common cause and assigns easily interpretable probabilities to each possible causal relation.

In general, the presence of observational or dynamic noise decreases the reliability of the inference as it is demonstrated in[13]. Additionally, the curvature of delay-reconstructed manifold causes the overestimation of local dimensionality (for a finite sample) and also sets limit on maximal noise level, which allows meaningful state space reconstruction[38].

We should mention that correlation dimension and mutual information dimension have been applied to evaluate connection between dynamical systems, but none of them led to exhausting description of bivariate causal relationships due to their symmetric nature[17;27].

Future directions. In the present paper we focused on systems with deterministic dynamical components, however, we hope that our unified framework could be extended to pure stochastic systems, and systems of mixed character, having deterministic and stochastic components as well.

In the present work we assumed no delay in the causal effect. Similarly to[28] this assumption can be relaxed and causality can be determined with different effect delays, thus not only existence of the connection, but the delay of the effect can be determined.

The pairwise determination of the causal relationships used here could be extended to multivariate comparisons to reveal more complex network structures.

The Dimensional Causality method is a unified method to quantify all types of causal relationships. Based on the presented examples and tests we believe that this new method will be applicable in many different scientific areas, although it is clear that there are many open questions and further directions to work out.

References

1. Wiener, N. The Theory of Prediction. In *Modern Mathematics for Engineers*, vol. 58, 323–329 (McGraw-Hill Book Company, Inc., New York, 1956).

2. Granger, C. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* **37**, 424–438 (1969).

3. Schreiber, T. Measuring Information Transfer. *Physical Review Letters* **85**, 461–464 (2000).

4. Lusch, B., Maia, P. D. & Kutz, J. N. Inferring connectivity in networked dynamical systems: Challenges using Granger causality. *Physical Review E* **94** (2016).

5. Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science (New York, N.Y.)* **338**, 496–500 (2012).

6. Tajima, S., Yanagawa, T., Fujii, N. & Toyoizumi, T. Untangling brain-wide dynamics in consciousness by cross-embedding. *PLoS computational biology* **11**. URL https://doi.org/10.1371/journal.pcbi.1004537 (2015).

7. Harnack, D., Laminski, E., Schünemann, M. & Pawelzik, K. R. Topological causality in dynamical systems. *Physical Review Letters* **119**, 098301 (2017).

8. Takens, F. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence* **898**, 366–381 (1981).

9. McCracken, J. M. & Weigel, R. S. Convergent cross-mapping and pairwise asymmetric inference. *Physical Review E* **90**, 062903 (2014).

10. Cobey, S. & Baskerville, E. B. Limits to causal inference with state-space reconstruction for infectious disease. *PloS one* **11**, e0169050 (2016).

11. Mønster, D., Fusaroli, R., Tylén, K., Roepstorff, A. & Sherson, J. F. Causal inference from noisy time-series data—testing the convergent cross-mapping algorithm in the presence of noise and external influence. *Future Generation Computer Systems* **73**, 52–62 (2017).

12. Hirata, Y. & Aihara, K. Identifying hidden common causes from bivariate time series: A method using recurrence plots. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **81** (2010).

13. Hirata, Y. *et al.* Detecting causality by combined use of multiple methods: Climate and brain examples. *PloS one* **11**, e0158572. URL https://doi.org/10.1371/journal.pone.0158572 (2016).

14. Vicente, R., Wibral, M., Lindner, M. & Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience* **30**, 45–67 (2011).

15. Rényi, A. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica* **10**, 193–215 (1959).

16. Pincus, S. M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* **88**, 2297–2301 (1991).

17. Romano, S., Chelly, O., Nguyen, V., Bailey, J. & Houle, M. E. Measuring dependency via intrinsic dimensionality. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, 1207–1212 (IEEE, 2016).

18. Geiger, B. C. & Koch, T. On the information dimension of stochastic processes. *arXiv preprint arXiv:1702.00645* URL `https://arxiv.org/abs/1702.00645` (2017).

19. Mandelbrot, B. Fractals: form, chance and dimension. *Fractals: form, chance and dimension., by Mandelbrot, BB. San Francisco (CA, USA): WH Freeman & Co., 16+ 365 p.* (1979).

20. Grassberger, P. & Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett* **50**, 346 (1983).

21. Levina, E. & Bickel, P. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information ...* **17**, 777–784 (2004).

22. Massoud Farahmand, A., Szepesvári, C. & Audibert, J.-Y. Manifold-adaptive dimension estimation. In *Proceedings of the 24th international conference on Machine learning - ICML '07*, 265–272 (ACM Press, New York, New York, USA, 2007).

23. Lorenz, E. N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**, 130–141 (1963).

24. Hindmarsh, J. L. & Rose, R. M. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London B: Biological Sciences* **221**, 87–102 (1984).

25. Xia, S. & Qi-Shao, L. Firing patterns and complete synchronization of coupled Hindmarsh–Rose neurons. *Chinese Physics* **14**, 77–85 (2005).

26. Casdagli, M., Eubank, S., Farmer, J. & Gibson, J. State space reconstruction in the presence of noise. *Physica D: Nonlinear Phenomena* **51**, 52–98 (1991).

27. Sugiyama, M. & Borgwardt, K. M. Measuring statistical dependence via the mutual information dimension. *IJCAI International Joint Conference on Artificial Intelligence* 1692–1698 (2013).

28. Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific reports* **5**, 14750 (2015).

Author Contribution Z.B., A.Z., Z.S., A.T. worked out the DC method and wrote the manuscript. Z.B. and A.Z. ran the analysis on simulated and EEG data. Z.B. and A.Z. are equally contributed to the work. A.Z. implemented the algorithm. A.S. recorded the photostimulation EEG data. L.E. and D.F. done the surgery and recorded intracranial LFP data. Z.S. and A.T. are equally contributed to the work. All authors approved the final version of the manuscript.

Competing interests statement The authors of the paper state that they do not have any competing interest.

Author Information Correspondence and requests for materials should be addressed to somogyvari.zoltan@wigner.mta.hu.
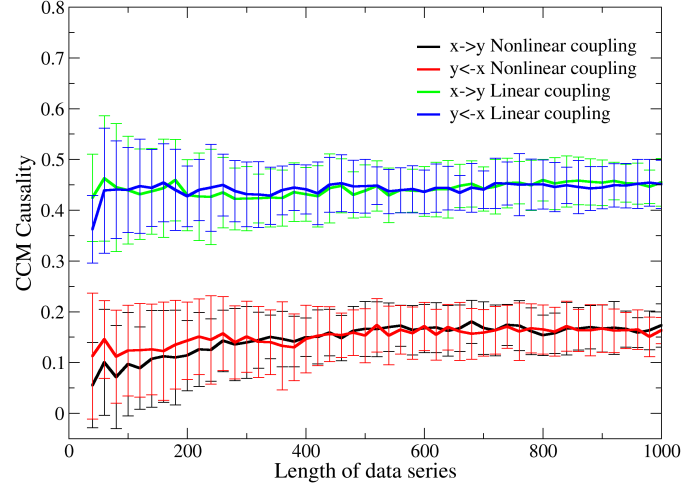
Extended Data Figure 1: *CCM causality between two observed logistic maps driven by a hidden common cause* The CCM causality measure dependence on the length of the data series are shown. The observed time series (x and y) were not directly coupled, but they were driven by a third logistic map as a common cause. The causality measure shows high values in both directions independently from the data length only if the hidden common cause was linearly coupled to the x and y, thus they were correlated, with $r = 0.46$ linear correlation coefficient (blue and green lines). In the case of nonlinear coupling, the CCM increases with the increasing data length but converges to lower values, which again corresponds to the linear correlation between x and y $r = 0.15$ (black and red lines). Thus, from the data length dependence of CCM, the existence of a hidden common cause can not be inferred reliably.
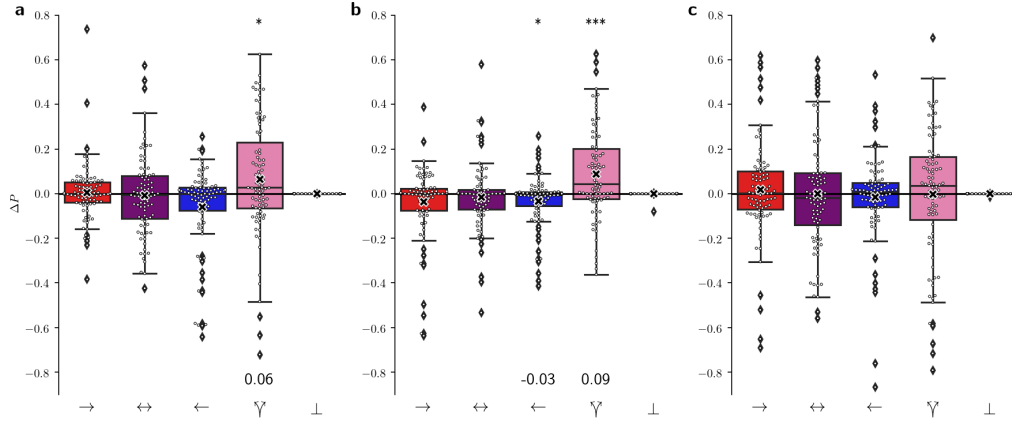
Extended Data Figure 2: *Differences in probability for causal connection types: stimulus minus control.* The difference in probabilities of causal types are depicted for P3-P4 (**a**), C3-C4 (**b**), F3-F4 (**c**) channel pairs ($n = 87$). Boxes denote inter-quartile (IQ) ranges, whiskers are at 1.5 IQ range from the quartiles and the points outside the whiskers are marked as outliers by diamond symbols. Sample median and mean are indicated by horizontal line and **x** marker respectively. Asterisks mark significant differences of sample mean (*: $p < 0.05$, ***: $p < 0.001$). **a** The probability of common cause significantly increased on the P3-P4 channel pair ($p = 0.026$). **b** On C3-C4 channel pair, the probability of common cause shows significant increase ($p = 0.0006$) and the directed causal link from C4 to C3 is significantly decreased ($p = 0.011$). **c** No significant changes in probabilities are detected on the F3-F4 channel pair.
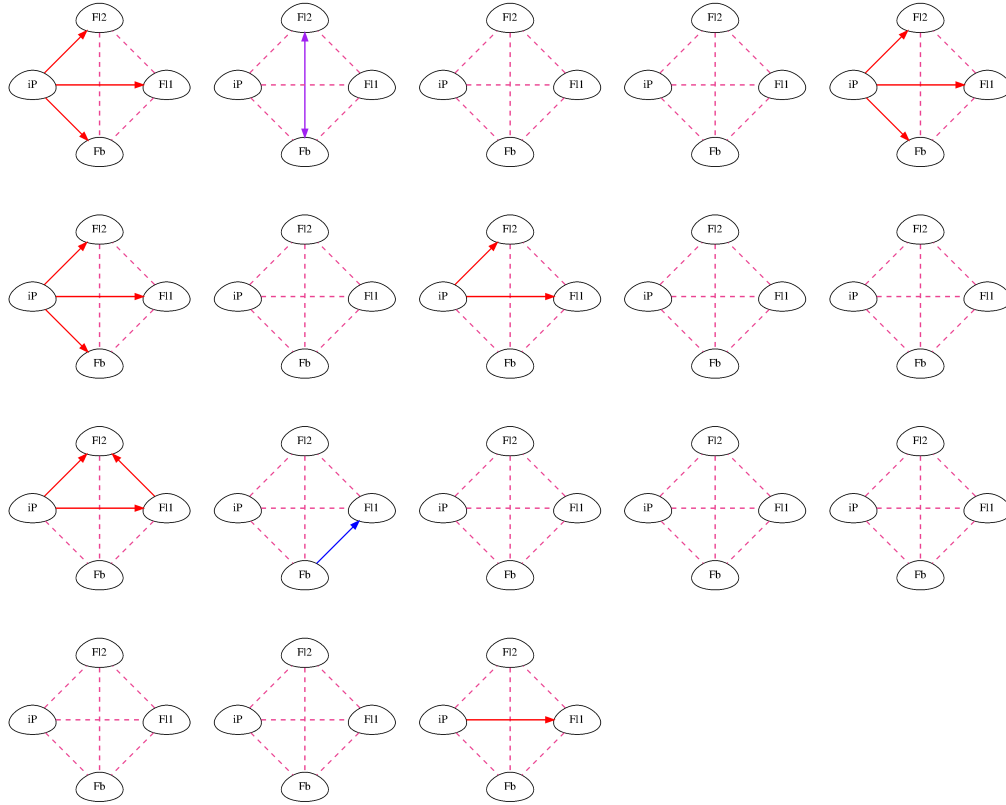
Extended Data Figure 3: *Inferred causal connectivity during seizures.* Two main types of MAP connectivity pattern were observed: in 10 cases the existence of a hidden common cause were inferred, while in 6 cases the infero-Parietal (iP) area was found as a driver.
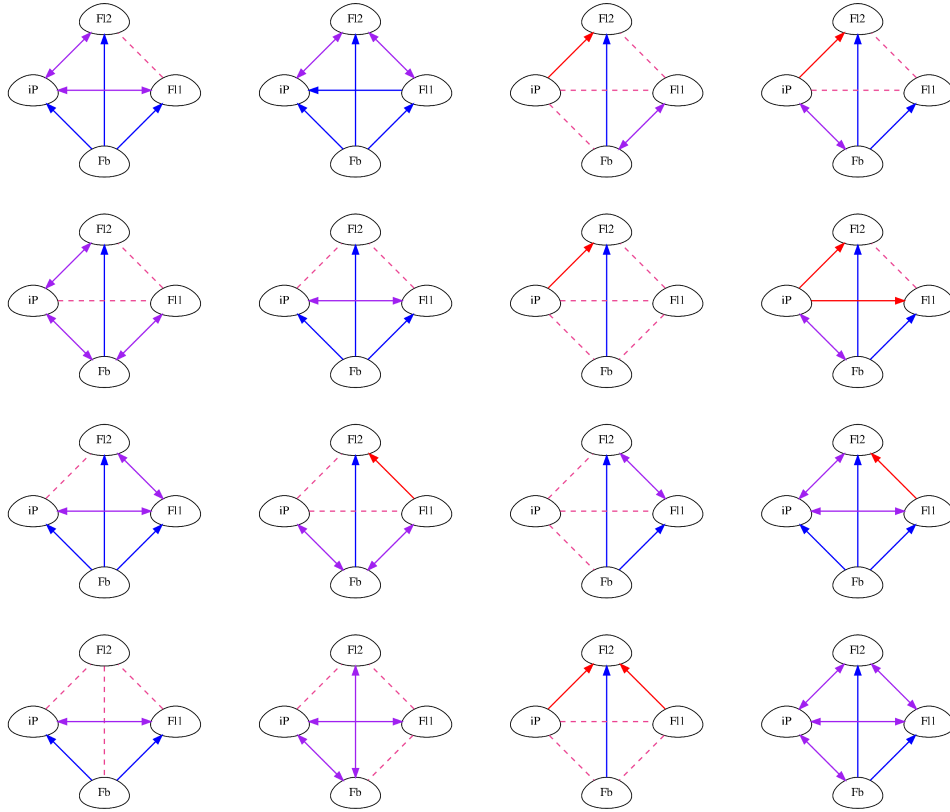
Extended Data Figure 4: *Causal relations for interictal periods.* MAP connectivity of the 16 interictal periods show Fronto-basal (Fb) area as a main driver.

Supplementary Information to Exact Inference of Causal Relations in Dynamical Systems

Zsigmond Benkő[1,2], Ádám Zlatniczki[1,3], Dániel Fabó[4], András Sólyom[4], Loránd Erőss[5,6], András Telcs[1,3,7], Zoltán Somogyvári[1,8,*]

[1]*Department of Computational Sciences, Wigner Research Centre for Physics of the Hungarian Academy of Sciences, H-1121, Hungary*
[2]*János Szentágothai Doctoral School of Neurosciences, Semmelweis University, H-1085, Hungary*
[3] *Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, H-1111, Hungary*
[4] *Epilepsy Center "Juhász Pál", National Institute of Clinical Neurosciences, Budapest, H-1145, Hungary*
[5] *Department of Functional Neurosurgery, National Institute of Clinical Neurosciences, Budapest, H-1145, Hungary*
[6] *Faculty of Information Technology and Bionics, Péter Pázmány Catholic University, Budapest, H-1083, Hungary*
[7] *Department of Quantitative Methods, University of Pannonia, Faculty of Business and Economics, H-8200, Veszprém, Hungary*
[8] *Neuromicrosystems ltd., Budapest, H-1113, Hungary*
*Electronic address: somogyvari.zoltan@wigner.mta.hu

1 Supplementary Notes

1.1 Intrinsic dimension and causality

In this section we lay down the mathematical background of the relationship between causal connections and relationships between the dimension of the state space of systems. We begin with some general remarks on the possibilities and limitations of the analysis of the connection of dynamical systems via an observation, a time series. It is standard to assume that the system is in steady state and consequently the observed time series is stationary. We will investigate the stationary distribution via its information dimension. Let us recall first the definition of the information dimension introduced by Rényi[15].

$$d_X = \lim_{N \to \infty} \frac{1}{\log N} H([X]_N),$$

23

where $X \in \mathbb{R}^m$ is a continuous vector valued random variable, $[X]_N = \frac{\lfloor NX \rfloor}{N}$ is its N-quantization (a discrete variable) and $H(.)$ stands for the Shannon entropy. The finer the resolution $r = 1/N$ the more accurate the approximation of the information content is, but the normalization in the limit eliminates the contribution of the discrete part of the distribution (if it exists). That means that the estimate the information dimension has a trade-off and also a limitation given the finiteness of the sample. It is suggested that if the variable lives in a D dimensional space then a proper estimate needs at least $10^D - 30^D$ sample points[29]. Following Pincus' ideas[16;30;31] we introduce the approximate information dimension

$$d_{X,1/N} = \frac{1}{\log N} H([X]_N).$$

or in short $d_{X,r}$, for arbitrary partition with box size r.

Assumption 1 *The investigated time series are stationary.*

Dimensions. The notion of dimension has many definitions depending on the context and methods. Takens' work was confined to topological dimensionality, but the embedding theorem was extended to fractal dimensions as well[32]. In our case the attractor, the support of the stationary distribution, can be a fractal. We already recalled the Rényi information dimension. On the other hand the local intrinsic dimension is defined as follows. Let $X \in \mathbb{R}^m$ be the investigated random variable

$$D_X(x) = \lim_{r \to 0} \frac{1}{\log r} \log P(X \in B(x,r))$$

where $B(x,r)$ is the hypercube in $\mathbb{R}^m$ with lower corner x. Then the intrinsic dimension is $D_X = E(D_X(X))$ and

$$d_X = D_X$$

(see the work of Camastra and Staiano[33], and theorems 1 and 2 of Romano et al.[17]). ID estimation has a waste literature, several excellent reviews[34;33] help to find the best performing one. Reviews and benchmark tests indicate that the k-Nearest Neighbour based methods have several advantages, including the rigorous derivation of the estimate presented by Levina and Bickel[21] and the very good convergence properties of a variant[22]. We follow the latter one, in particular we do so since it provides a hint for the choice of scaling, which is one of the crucial points in the ID estimation procedure.

Assumption 2 *The embedded manifolds are homogeneous with respect to (the existing) dimension.*

Based on the previous introduction our intrinsic dimension estimates are

$$\widehat{D}\left(x\right)_r = \frac{1}{\log r} \log\left|N\left(x,r\right)\right|$$

and

$$\widehat{D}_{X,r} = \frac{1}{n} \sum_{i=1}^{n} \widehat{D}\left(x_i\right)_r,$$

where n is the sample size, $\{X_i\}_{i=1}^n$ is the set of sample points on the manifold and $N(x,r) = \{X_i : X_i \in B\left(x,r\right)\}$.

We shall use a bit reversed logic (following Farahmand, Szepesvári and Audibert[22]) and calculate the dimension from the distance of k-th nearest neighbour $r(x,k) = d(x, X^k(x))$ where $X^k(x)$ is the k-th closest point to x in our sample series. In this setting the resolution is given by

$$r^D \approx \frac{k}{n} \tag{6}$$

where n is the sample size.

Causal relations. Let us introduce the box partitioning of $\mathbb{R}^m$ using $r\mathbb{Z}^m$ and denote the boxes by $B(x,r)$, where x is the lower left corner. Also we index the partition element $B(x,r)$ by x. The quantized variables are defined as follows.

$$X^r = x \text{ if } X \in B(x,r) \tag{7}$$

Definition 3 *We say that X causes/drives Y at resolution r (denoted by $X \rightarrow_r Y$) if there is a mapping f s.t. $X_t^r = f\left(Y_t^r\right)$ for all t.*

In case of a time delayed causal relation between the systems with lag τ we have $X_{t-\tau}^r = f\left(Y_t^r\right)$, therefore the proper time-shift should be applied as a preprocessing step.

Definition 4 *We say that X and Y are in circular causal relations at resolution r (denoted by $X \leftrightarrow_r Y$), if both drives the other at resolution r.*

Definition 5 *We say that the x_t and y_t observations are basically identical (also called generalized synchrony) if there is a bijective mapping f s.t. $x_t = f\left(y_t\right)$ for all t. From a dynamical systems point of view, generalized synchrony is a special case of circular causality.*

One should note that we defined causality (definition 3) counter-intuitively in a kind of reversed direction. X can be reconstructed from the information contained in Y given that exactly that is what conveyed from X to Y in the action of driving. If just partial reconstruction is possible, say $X = (X', Y'')$ and $Y = (Y', Y'')$, where all the components are independent, then Y'' is a common cause of X and Y, and X does not drive Y or vice versa.

Definition 6 *We say that X and Y are independent at resolution r (denoted by $X \perp_r Y$ if X^r independent of Y^r.*

Definition 7 *We say that $\{Y\}$, $\{X\}$ have a common cause at resolution r (denoted by $X \curlyvee_r Y$) if they are, at resolution r, not independent and there is no driving connection between then again at resolution r.*

Link between causal and dimensional relations. First we recall some elementary facts which will be useful. Let $\mathbb{J} = (X, Y)$ be the joint embedding. The intrinsic dimension D coincides with the information dimension d and the same applies for the quantized versions[17].

Lemma 8 *The elementary properties of the discrete Shannon-entropy imply that for all r*

$$\max\{H(X^r), H(Y^r)\} \leq H(X^r, Y^r) \leq H(X^r) + H(Y^r) \qquad (8)$$

and

$$\max\{D_{X,r}, D_{Y,r}\} \leq D_{(X,Y),r} \leq D_{X,r} + D_{Y,r}, \qquad (9)$$

where equality on the r.h.s. holds if and only if the variables are independent at resolution r $(X \perp_r Y)$.

Corollary 9 $X \perp_r Y$ *iff.*

$$D_{(X,Y),r} = D_{X,r} + D_{Y,r}, \qquad (10)$$

Proof. The statement is immediate from the additivity of the Shannon entropy of independent variables. ∎

Corollary 10 *If $X \to_r Y$ then*

$$D_{X,r} \leq D_{Y,r} = D_{(X,Y),r}$$

and if $X \leftrightarrow_r Y$ then

$$D_{Y,r} = D_{X,r} = D_{(X,Y),r}.$$

26

Proof. Both statements follow from the fact that if $X^r = f(Y^r)$ then $H(X^r|Y^r) = H(f(Y^r)|Y^r) = 0$:

$$
\begin{aligned}
H(X^r, Y^r) &= H(X^r|Y^r) + H(Y^r) \\
&= H(f(Y^r)|Y^r) + H(Y^r) \\
&= H(Y^r).
\end{aligned}
\tag{11}
$$

∎

Theorem 11 *If $D_{(X,Y),r} = D_{Y,r}$ then X drives Y at resolution r.*

Proof. It is immediate from the condition that

$$H(X^r|Y^r) = 0$$

but that implies that there is a mapping f s.t. $X^r = f(Y^r)$.

∎

Theorem 12 *X^r and Y^r have a common cause if and only if*

$$\max\{D_{X,r}, D_{Y,r}\} < D_{(X,Y),r} < D_{X,r} + D_{Y,r}.$$

Proof. The statement follows from the combination of Corollary 9 and Theorem 11. ∎

If we combine all our results from Lemma 8 to Theorem 12 then we have the following full table of implications.

$$
\begin{bmatrix}
X \to_r Y & \iff & D_{X,r} < D_{Y,r} = D_{(X,Y),r} \\
Y \to_r X & \iff & D_{Y,r} < D_{X,r} = D_{(X,Y),r} \\
X \leftrightarrow_r Y & \iff & D_{X,r} = D_{Y,r} = D_{(X,Y),r} \\
X \; \gamma_r \; Y & \iff & \max\{D_{X,r}D_{Y,r}\} < D_{(X,Y),r} < D_{X,r} + D_{Y,r} \\
X \perp_r Y & \iff & D_{(X,Y),r} = D_{X,r} + D_{Y,r}
\end{bmatrix}
$$

Let us note that the first implication remains true as $r \to 0$ if it is ensured that the conditional random variable $X|Y$ has no discrete part. A bit more indirect extra condition is needed for the validity of the common cause in the limit, though. If Z is the common driver, it should satisfy that $Z|X$ and $Z|Y$ have no discrete part. Finally, for independence in the limit, it is necessary that there is no discrete connection between the variables (the formal definition is omitted for brevity). Note that if these implications hold as $r \to 0$, then there is causal relationship between the systems in classical dynamical system theory sense.

Causality detection using additive observation of the series. The genericity of Takens' theorem allows us to choose an observation function (mapping) almost freely. As demonstrated in eq. 9 the direct product of the embedding of the series has such good properties. On the other hand it increases the embedding dimension, which is unfortunate in practice, so we may look for other mappings. The simplest, linear one, is a natural choice. Let us imagine $W = X + Y$. In general it is again a good mapping but if in particular $X = X_1 - Y$, for independent X_1 and Y the result would be the drop of dimension.

There is a simple resolution of that problem. If we choose a random number $a \in [0, 1], a \neq 0$ uniformly and consider $U = aX + Y$ than, with probability one U will be a proper observation function and can replace the direct product J. Also if we know that $D_X, D_Y < m + 1$ and we chose $a_i \in [-1, 1], i = 1...m + 1$ random number, and set $U_i = a_i X - Y$ at least one should be proper observation function and the wrong once can be identified by the drop of the dimension.

1.2 Assigning probabilities to causal relations

In the whole sequel we omit marking the resolution r where it does not cause confusion. Given the constraints $\max \{D_X, D_Y\} \leq D_J \leq D_Z \leq D_X + D_Y$ we have the next complete partition of the event space of possible causal relationships

	direct drive	$\curlyvee$	$\perp$
$D_X < D_Y$	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$
$D_X > D_Y$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$
$D_X = D_Y$	$A_{3,1}$	$A_{3,2}$	$A_{3,3}$

and $\mathcal{A} = \{A_{i,j} : i, j = 1, 2, 3\}$ (in the main text, $A \in \mathcal{A}$ in the Bayesian argument). In the main text the causal relations referred as $X \to Y$, corresponds here to $A_{1,1}$, $X \leftarrow Y$ to $A_{2,1}$, $X \leftrightarrow Y$ to $A_{3,1}$, while $X \curlyvee Y = \sum_{i=1}^{3} A_{i,2}$ and $X \perp Y = \sum_{i=1}^{3} A_{i,3}$.

Based on the work of Romano et al.[17] we consider the expected value of local dimensions to be the global dimension - therefore, the mean of the local dimension estimates yields our estimate of the global dimension. Consequently, the sampling distribution of the global dimension estimates is multivariate normal with true dimensions D being its mean and covariance matrix Σ.

Let $\bar{D}$ denote the observed global dimension that depends on D and Σ. Note however that D and Σ are uncertain parameters themselves, since they

depend on the causal model A. The dependence of D on A is trivial by the links derived in the previous sections between causality and dimensions, but A also has an effect on the noise model Σ: if for example the causal relationship is $X \to Y$ ($A = A_{1,1}$), then $D_Y = D_J$ in theory, therefore D_Y and D_J should be correlated.

Let U represent the parameters of our method (e.g. k, τ, embedding dimension m). Our method's parameters also affect dimension estimation, therefore influence the noise model (but not the true dimensions D).

Finally we can conclude that $\bar{D}$ depends on D and Σ, D depends only on A, while Σ depends on both A and U, where U can be considered hyperparameters.

In this setting the realization of $\bar{D}$ is a single (4-dimensional) data point, let's denote it as $\bar{d}$. The likelihood of the data can be written as

$$p_{\bar{D}}(\bar{d}) = \sum_{A_{ij} \in \mathcal{A}} p_{\bar{D}|A}(\bar{d}|A_{ij}) \, P_A(A_{ij}).$$

We are interested in the probabilities of each A_{ij} given the data, therefore we apply Bayes' theorem and get

$$P_{A|\bar{D}}(A_{ij}|\bar{d}) = \frac{p_{\bar{D}|A}(\bar{d}|A_{ij})}{p_{\bar{D}}(\bar{d})} P_A(A_{ij}).$$

Note that $p_{\bar{D}}(\bar{d})$ is only a normalizing term, therefore it is enough to calculate $p_{\bar{D}|A}(\bar{d}|A_{ij})P_A(A_{ij})$ for all A_{ij}. We assume a non-informative prior over the possible causal relations, $P_A(A_{ij}) = \frac{1}{9} \, \forall i, j$. Taking the dependence structure of the random variables into consideration we can write the likelihood as

$$p_{\bar{D}|A}(\bar{d}|A_{ij}) = \int p_{\bar{D}|D,\Sigma}(\bar{d}|w, s) \, dP_{D,\Sigma|A}(w, s|A_{ij}).$$

If we extend this with U we get

$$\begin{aligned} p_{\bar{D}|A}(\bar{d}|A_{ij}) &= \int p_{\bar{D}|A,U}(\bar{d}|A_{ij}, u) \, dP_U(u) \\ &= \int \int p_{\bar{D}|D,\Sigma}(\bar{d}|w, s) \, dP_{D,\Sigma|A,U}(w, s|A_{ij}, u) \, dP_U(u). \end{aligned}$$

We know that $\bar{d}$ comes from a 4-variate normal distribution with expected value vector w and covariance matrix s, therefore

$$
\begin{aligned}
p_{\bar{D}|A}(\bar{d}|A_{ij}) &= \int\int \varphi_{w,s}(\bar{d}) \, dP_{D,\Sigma|A,U}(w,s|A_{ij},u) \, dP_U(u) \\
&= \int\int \varphi_{\bar{d},s}(w) \, dP_{D,\Sigma|A,U}(w,s|A_{ij},u) \, dP_U(u) \\
&= \int\int\int \varphi_{\bar{d},s}(w) \, dP_{D|A,U}(w|A_{ij},u) \, dP_{\Sigma|A,U}(s|A_{ij},u) \, dP_U(u) \\
&= \int\int\int \varphi_{\bar{d},s}(w) \, dP_{D|A}(w|A_{ij}) \, dP_{\Sigma|A,U}(s|A_{ij},u) \, dP_U(u),
\end{aligned}
$$

where $\varphi_{w,s}$ denotes the multivariate normal PDF with expected value vector w and covariance matrix s. We used the conditional independence of D and Σ, the independence of D and U, and that in the probability density function of the normal distribution the expected value and the data can be exchanged, since $(\bar{d} - w)^T s^{-1} (\bar{d} - w) = (w - \bar{d})^T s^{-1} (w - \bar{d})$.

Let us now consider $dP_{D|A}(w|A_{ij})$. Every A_{ij} induces a set $S_{ij} \subset \mathbb{R}_+^4$ such that each element of S_{ij} satisfies the conditions given by A_{ij}. For example, $S_{1,1} = \{v : v \in \mathbb{R}_+^4, v_1 < v_2 = v_3 < v_4\}$. It is easy to verify that S_{ij} is a convex cone with (algebraic) dimension $q_{ij} \in \{2,3,4\}$, simply embedded into a 4-dimensional Euclidean space. Let $C_{ij} \in \mathbb{R}_+^{q_{ij}}$ denote the convex cone in its original, lower dimensional space. For example $C_{1,1} = \{w : w \in \mathbb{R}_+^3, w_1 < w_2 < w_3\}$. For any A_{ij} there exists a simple linear bijective mapping M_{ij} such that $M_{ij}w \in S_{ij} \; \forall w \in C_{ij}$, for example

$$
M_{1,1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
$$

We assume a non-informative uniform prior on the true dimensions D. Since the prior is conditioned on A_{ij}, C_{ij} becomes its support, resulting in an improper prior (due to the cone being infinite). It is obvious that this is not an actual distribution, but specifies a prior with correct proportions (granting equal weight to each element of the support). Putting these together we get

$$
p_{\bar{D}|A}(\bar{d}|A_{ij}) = \int\int\int_{C_{ij}} \varphi_{\bar{d},s}(M_{ij}w) \, dw \, dP_{\Sigma|A,U}(s|A_{ij},u) \, dP_U(u).
$$

Let us now focus on marginalizing Σ. The 4×4 covariance matrix has 10 elements in the upper-triangular part - we would have to marginalize each of them. This comes with computational issues and lack of knowledge about the conditional distribution of Σ, both of which we wish to avoid. Instead,

30

we calculate a sample covariance matrix $\hat{s}$, which is a maximum-likelihood estimate, and assume that the distribution is very peaked at this estimate.

Note however that the neighbourhoods of our sample points are overlapping and the calculated local dimensions are not independent. Therefore, the covariance matrix must be calculated by taking the correlation of local dimension estimates into consideration. If the samples were independently drawn, then the covariance of the means is given by the covariance of the local dimensions divided by the number of samples. If the samples are correlated though, then we must divide by the effective sample size instead. Loosely speaking, the effective sample size of an estimator of the population mean is the number with the property that our estimator has the same variance as the estimator achieved by sampling the same amount of independent individuals. In our case, two local dimension estimates are independent if their k-neighbourhoods do not intersect. Therefore we can (approximately) sample $\frac{n}{2k}$ independent elements from them, and that is our effective sample size.

While a specific A_{ij} explicitly excludes certain dimension combinations, it can not exclude a covariance matrix - $\hat{s}$ has positive likelihood given any causal relationship. We assume that the distribution of Σ is very peaked at the maximum-likelihood estimate, so much that we regard it to be a Dirac delta, $dP_{\Sigma|A,U}(s|A_{ij}, u) = dH(s - \hat{s})$, where $H(\cdot)$ is a multivariate unit step function. Note that $\hat{s}$ is a function of U as well.

After incorporating this into the integral and marginalizing we get

$$p_{\bar{D}|A}(\bar{d}|A_{ij}) = \int \int_{C_{ij}} \varphi_{\bar{d},\hat{s}}(M_{ij}w) \, dw \, dP_U(u).$$

In the $A_{ij} = A_{1,1} = \{X < Y = J < Z\}$ case the above would result in

$$p_{\bar{D}|A}(\bar{d}|A_{1,1}) = \int \left(\int_0^\infty \int_{v_1}^\infty \int_{v_2}^\infty \varphi_{\bar{d},\hat{s}}(w_1, w_2, w_2, w_3) dw_3 \, dw_2 \, dw_1 \right) dP_U(u).$$

In the current implementation U consists only of k (the neighbourhood size, which is discrete uniform), therefore integrating by U practically boils down to averaging. The other $A_{i,j} \in \mathcal{A}$ cases can be treated similarly.

2 Supplementary Methods

2.1 Analysis work-flow

Our proposed causality analysis method starts with two time series. First both time series must undergo data cleaning: one has to make sure that the

data is stationary (which can be checked for example with an augmented Dickey-Fuller unit root test) and that the observational noise is dealt with (for example with a filter). Transforming the series to stationary ones is non-trivial and usually depends on the scientific field, requiring much attention and field expertise - by choosing a wrong transformation precious information may be lost or artefacts are introduced, biasing the final results. For example differentiating the time series may remove too much information, using moving averages to smooth the series introduces higher autocorrelation, or calculating current source density with non-disjoint sets of signals to remove correlation (which only indicates first-order relationship) may introduce higher-order dependence. Another important phase of preprocessing is normalization: if the two time series have different scale (or magnitude), then the results of the dimension estimation will become biased. This is easy to see, for example if values of X are around 1 while values of Y are around 100, then the dimension estimation of X and Y will be correct, but since $J = (X, Y)$, or $J = aX + Y$, the k-NN distances will be the same in J and Y, dominating the effects of X. There are several ways one can normalize the data, e.g. [0,1]-scaling, z-scores, quantile normalization (or rank normalization), etc.

The preprocessed time series are embedded into an m dimensional space with lag τ, according to Takens' theorem. Both m and τ are parameters of our model that have to be specified. Takens showed that $m = 2d + 1$ (where d is the true dimension) is a sufficient choice for the embedding dimension in order to reconstruct dynamics, but unfortunately the true dimension is usually unknown - one may have to consult with a field expert who has sufficient knowledge about the dynamical system at hand and can give a proper estimation of the true dimension. If this is not possible, one can use several methods proposed in the literature for the selection of m. The correct value of m can be determined by searching for a plateau (saturation effect) in invariant quantities (such as intrinsic dimensionality) or by using the false nearest neighbour method[36;37]. We applied an iterative process, starting with a high embedding dimension and decreasing it, checking the estimated manifold dimensions after each decrease and selecting the lowest possible m which did not reduce estimated d value sharply. Takens says that $m = 2d+1$ is a sufficient choice, but that does not mean that smaller $m < 2d+1$ can not be acceptable, this depends on the system. In general[38] the self-intersections do not alter the estimated manifold dimension if $m > d$. In finite samples it is common that dimension estimates increase as the embedding dimension increases, therefore in general it is better to select m as low as possible.

The optimal value of τ can be determined from the first zero point of the autocorrelation function or from the first minima of automutual-information

function[39]. Additionally, one can optimize for m and τ at once by applying differential geometric arguments or using the statistical properties of dimension estimates calculated on embedded data[40;41], or determine an optimal embedding window[42] $(m-1)\tau$. In order to find optimal τ we analysed the partial autocorrelation function (PACF) of the time series and selected the largest significant lag (which is how one would fit an autoregressive model). The first insignificant lag in the autocorrelation function (ACF) could be used as well, but we found PACF more effective in practice. The largest significant lag of PACF as embedding delay results in relatively independent (moderately redundant) coordinates, but still not too independent (irrelevance) to reconstruct dynamics. The ACF may diminish very slowly, resulting in very large τ, or for example if it is monotonic and we select the first insignificant lag we exclude the linear and their induced non-linear relationships. On the other hand, the largest significant lag in PACF tells us which is the largest lag where there is no direct linear relationship, but non-linear relationships induced by these linear ones remain (think of the classical partial correlation example when there is direct linear relationship between t_1 and t_2, t_2 and t_3, and therefore a quadratic between t_1 and t_3).

Given the two m dimensional embedded manifolds of the series, X and Y, the joint J and independent Z manifolds are created as well. For this purpose one can either choose direct products or additive observation function $(aX + Y)$. We advise the usage of additive observation function, since (as we have shown) the two result in equivalent relationships between manifold dimensions, yet direct products produce higher dimensional product spaces where dimension estimation becomes more-and-more unreliable. Our experience showed that $a = \sqrt{\frac{29}{31}}$ is a good choice in general, because it ensures that the scale of the series do not change too much.

In some cases the embedded manifolds may have to undergo further transformations. For example continuous dynamics (like a Lorenz system) evolve rather slowly and in a thread-like manner - if one does not choose a sufficiently large neighbourhood when estimating the local dimension at a point the manifold may seem one-dimensional, because the nearest points will all lie on the same thread. This effect is similar to sampling a slow process with too high frequency. This may be handled by down-sampling the manifold, which means re-sampling it with a lower frequency.

The (post-processed) manifolds are used to estimate local dimensions, for which we use an estimator proposed by Farahmand, Szepesvári and Audibert[22]. Their estimator has one parameter, k, the size of the neighbourhood around a point in which the dimension will be estimated. Depending on the system this estimator may be sensitive to this parameter, therefore we try

several values and aggregate the results (as described in the previous section).

As stated before, dimension estimation is rather hard and local estimates may be quite far from the actual dimension. To handle this one should remove the outliers from the estimates. We advise to trim the dimension estimates, simply by dropping those that belong to the upper or lower α-percent tail.

At this point all the data is available for estimating model probabilities, as described in the previous section. We repeat this process for a range of k-s, then assuming a uniform prior over them we get the final model probabilities by practically averaging the probability of each model for each k out.

2.2 Empirical results

Logistic maps. For testing purposes we applied the DC method on systems of three-coupled logistic maps with various connectivity patterns. The logistic map is a simple non-linear discrete-time dynamical system, which serves as a model for various economic and ecosystems and can (re)produce chaotic behaviour even in one dimension. It is defined as

$$x_j[t+1] = rx_j[t](1 - \sum \beta_{jl} x_l[t]), \tag{12}$$

where $r = 3.99$, $j, l \in \{1, 2, 3\}$ is the index for the three variables and β_{jl} are the elements of the coupling matrix ($\underline{\underline{B}}$) according to an actual coupling scenario:

(a) Direct coupling:
$$\underline{\underline{B}} = \begin{bmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) Circular coupling:
$$\underline{\underline{B}} = \begin{bmatrix} 1 & \beta_{12} & 0 \\ \beta_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(c) Common cause case:
$$\underline{\underline{B}} = \begin{bmatrix} 1 & 0 & \beta_{13} \\ 0 & 1 & \beta_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

(d) Independent case:
$$\underline{\underline{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

34

where $\beta_{12} = \beta_{21} = \beta_{13} = \beta_{23} = 0.5$ in the example shown in the main text.

We simulated unidirectional, circularly causal, independent and hidden common cause connection patterns (Fig. 1) with $N = 10000$ time series length. Only two subsystems were observed, the activity of the third subsystem was hidden.

We preprocessed time series data by applying a rank normalization.

We set model-parameters and applied the DC method on logistic map datasets. We set embedding delay to $\tau = 1$ and we found that embedding dimension $m = 4$ was big enough in all cases. The probabilities were averaged over the neighbourhood sizes $k = [12, 44]$, where the dimension-estimates were constant.

DC was able to reconstruct the original coupling pattern between the observed logistic maps for all test cases, curiously it was able to detect hidden common cause between the two observed logistic maps.

In order to test the effect of a hidden common drive to Sugihara's CCM method, two types of coupling were applied: The case of non-linear coupling corresponds to equation (12), however the form for linear coupling is slightly different:

$$x_j[t+1] = rx_j[t](1 - x_j[t]) + \sum \beta_{jl} x_l[t] \qquad (13)$$

Sugihara et al.[5] stated that high values of CCM measure which is independent of the data length is a sign of hidden common cause. High CCM for short data series are typical for highly correlated data series. Similarly, Harnack et al.[7] apply the existence of the correlation without detectable direct causality to reveal hidden common cause. Our simulation results show that the above described properties of CCM hold only for those cases in which the hidden common cause is linearly coupled to the observed variables, thus it implies higher linear correlation ($r = 0.46$) among them (Extended Data Fig. 1, blue and green lines). If the hidden common driver is non-linearly coupled to the observed time series, thus the implied correlation is low ($r = 0.15$), the CCM increases with the increasing data length but still remains low (Extended Data Fig. 1, black and red lines). We conclude that the data length dependence of the CCM or the presence of correlation without causality does not allow us to reveal or distinguish the existence of a hidden common cause reliably, the CCM presumably follows only the linear correlation between the two driven variable in this case. Note that the DC method was tested and works well on the non-linearly coupled coupled case as well.

Lorenz systems. There are several ways to couple three Lorenz-systems -

we implement the coupling through their X coordinate such that

$$\dot{X}_1 = \sigma(Y_1 - X_1) + c_{2 \to 1}(Y_1 - X_2) + c_{3 \to 1}(Y_1 - X_3),$$
$$\dot{Y}_1 = X_1(\rho - Z_1) - Y_1,$$
$$\dot{Z}_1 = X_1 Y_1 - \beta Z_1,$$

$$\dot{X}_2 = \sigma(Y_2 - X_2) + c_{1 \to 2}(Y_2 - X_1) + c_{3 \to 2}(Y_2 - X_3),$$
$$\dot{Y}_2 = X_2(\rho - Z_2) - Y_2,$$
$$\dot{Z}_2 = X_2 Y_2 - \beta Z_2,$$

$$\dot{X}_3 = \sigma(Y_3 - X_3),$$
$$\dot{Y}_3 = X_3(\rho - Z_3) - Y_3,$$
$$\dot{Z}_3 = X_3 Y_3 - \beta Z_3,$$

where (X_i, Y_i, Z_i) are the three coordinates of the ith system, σ, ρ and β are model parameters, and $c_{i \to j}$ denotes the strength of coupling from the ith system to the jth system.

For the simulations we set $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, $\Delta t = 0.01$. The initial conditions are $X_1(0) = 10$, $Y_1(0) = 15$, $Z_1(0) = 21.1$, $X_2(0) = 17$, $Y_2(0) = 12$, $Z_2(0) = 14.2$, $X_3(0) = 3$, $Y_3(0) = 8$, $Z_3(0) = 12.4$, and we take 200,000 samples. The coupling coefficients are chosen for the four cases as follows:

(a) Direct cause: $c_{1 \to 2} = 3.5$, $c_{2 \to 1} = c_{3 \to 1} = c_{3 \to 2} = 0$,

(b) Circular cause: $c_{1 \to 2} = c_{2 \to 1} = 3.5$, $c_{3 \to 1} = c_{3 \to 2} = 0$,

(c) Hidden common cause: $c_{1 \to 2} = c_{2 \to 1} = 0$, $c_{3 \to 1} = c_{3 \to 2} = 3.5$,

(d) Independence: $c_{1 \to 2} = c_{2 \to 1} = c_{3 \to 1} = c_{3 \to 2} = 0$.

Our method requires the specification of the following parameters: the embedding dimension m, time-delay τ and a set of integers for different k-s for the $k-NN$ search. Takens found that if a manifold has dimension D then $2D + 1$ is a good candidate for the embedding dimension. The dimension of the joint manifold is at most 6 (the number of state variables). We try to minimize the embedding dimension: we start with $2D + 1$ and decrease it as long as the dimension estimates are not forced down. This way we found the following embedding dimensions for the four causal cases: (a) 7, (b) 7, (c) 7 and (d) 5. We select $\tau = 25$ based on the partial autocorrelation function (PACF) of the time series and $k \in [10, 38]$. Due to the threadlike phase space we also apply down-sampling on the reconstructed manifold by keeping only

every fourth point. We drop the outlying dimension estimates that belong to the lower or upper 5% tail.

Hindmarsh-Rose systems. We use three electrically coupled Hindmarsh-Rose neurons where coupling is achieved through their membrane potential (X), described by the differential equations

$$\dot{X}_1 = Y_1 - aX_1^3 + bX_1^2 - Z_1 + I_1 + c_{2\to1}(X_2 - X_1),$$
$$\dot{Y}_1 = c - dX_1^2 - Y_1,$$
$$\dot{Z}_1 = r_1\left(s\left(X_1 - \chi\right) - Z_1\right),$$

$$\dot{X}_2 = Y_2 - aX_2^3 + bX_2^2 - Z_2 + I_2 + c_{1\to2}(X_1 - X_2),$$
$$\dot{Y}_2 = c - dX_2^2 - Y_2,$$
$$\dot{Z}_2 = r_2\left(s\left(X_2 - \chi\right) - Z_2\right),$$

$$\dot{X}_3 = Y_3 - aX_3^3 + bX_3^2 - Z_3 + I_3 + c_{1\to3}(X_1 - X_3),$$
$$\dot{Y}_3 = c - dX_3^2 - Y_3,$$
$$\dot{Z}_3 = r_3\left(s\left(X_3 - \chi\right) - Z_3\right),$$

where a, b, c, d, χ, r_j, s, and I_j are model parameters and $c_{i\to j}$ denotes the strength of coupling from the ith neuron to the jth. Note that the first neuron is coupled into the remaining two, while the second neuron is coupled into the first one, meaning that the second and third neurons are not coupled directly, only through a hidden common driver. Common choices for the parameters are $a = 1$, $b = 3$, $c = 1$, $d = 5$, $s = 4$, $\chi = -1.6$. It is also important that r_i should be in the magnitude of 10^{-3} and $I_i \in [-10, 10]$ - we choose $r_1 = 0.001$, $r_2 = r_3 = 0.004$, $I_1 = 2.0$, $I_2 = 2.7$ and $I_3 = 2.4$. The initial conditions are $X_1(0) = 0$, $Y_1(0) = 0$, $Z_1(0) = 0$, $X_2(0) = -0.3$, $Y_2(0) = -0.3$, $Z_2(0) = -0.3$, $X_3(0) = 0.3$, $Y_3(0) = 0.3$ and $Z_3(0) = 0.3$. We take 70,000 samples with time steps $\Delta t = 0.1$ and drop the first 10,000 where the system is still in a burn-in period. The coupling coefficients are chosen for the four cases as follows:

(a) Direct cause: $c_{1\to2} = 0.615$, $c_{2\to1} = c_{1\to3} = 0$,

(b) Circular cause: $c_{1\to2} = c_{2\to1} = 0.615$, $c_{1\to3} = 0$,

(c) Hidden common cause: $c_{1\to2} = c_{1\to3} = 0.615$, $c_{2\to1} = 0$,

(d) Independence: $c_{1\to2} = c_{2\to1} = c_{1\to3} = 0$.

In each case we use the observations of the first two neurons, except for the hidden common cause case where we use the second and third neurons.

We parametrize the method based on the ideas pointed out in the Lorenz case. We set $k \in [10, 98]$, select $\tau = 5$ based on the PACF of the time-series, and the embedding dimension to be (a) 4, (b) 5, (c) 5, and (d) 3. We drop the lower and upper 5% of the dimension estimates.

Changes of inter-hemispheric connectivity during photo-stimulation.
EEG recordings and photo-stimulation. Monopolar recordings were taken with standard electrode arrangement of the 10-20 system by a Brain Vision LLC (Morrisville, NC 27560, USA) EEG device with 500 Hz sampling frequency from $n = 87$ patients and the data was stored by the Vision Recorder software with a $0.1 - 1000$ Hz bandpass filter enabled (No notch filter). Flashing light stimulation were carried out in a $0.1 - 29$ Hz range.
Preprocessing. We computed 2D current source density (CSD) of the EEG signal at the six selected electrode positions ($P3, P4, C3, C4, F3, F4$) to filter out linear mixing between channels[43]. CSD is roughly proportional to the negative second derivative of the electric potential, so we can represent CSD by the discrete Laplace of the original EEG signal up to a constant factor by the simplified formula:

$$CSD(x) \propto \sum_{i=1}^{N} n_i^{(x)} - Nx \tag{14}$$

where x is the EEG signal measured at an electrode position, $n_i^{(x)}$ is the signal measured at a neighbouring contact point, N is the number of neighbours and $CSD(x)$ is the CSD at x. We have to mention that equation 14 overlooks the non-uniform spatial distances of neighbours. The following neighbourhoods were used for the EEG channels:

- P3: {C3, T5, O1}

- P4: {C4, T6, O2}

- C3: {T3, F3, P3}

- C4: {T4, F4, P4}

- F3: {Fp1, F7, C3}

- F4: {Fp2, F8, C4}

Medial electrode points (common neighbours) were omitted from the neighbourhoods to avoid the artificial introduction of common cause relations.

Moreover $1 - 30$ Hz band-pass filtering (4th order Butterworth filter) was applied to the CSD signal. From stimulation periods, we used only the $21, 24.5, 26, 29$ Hz) stimulation segments in the analysis, which induced the greatest evoked EEG activity on recording channels. The resulting time series segments were $46760 - 84907$ samples long for stimulations. Artifactless segments with matching sample sizes were taken as paired control from stimulation-free periods.

Model parameters. We determined model parameters embedding delay and embedding dimension empirically as pointed out e.g. in the Lorenz case. Embedding delay was set according to the first insignificant value of PACF on the first dataset, and also a rough screening was made where we applied the method with unrealistically big delays and compared the results with smaller delay values. We assumed that choosing unrealistically big τ yields erroneous results, and a smaller τ which yields different result may be a valid choice of the embedding parameter. Both approaches showed that $\tau = 5$ was a reasonable value for embedding delay.

Embedding dimension was set according to a screening from $D = 5$ to $D = 14$, $D = 7$ as the smallest dimension required for the embedding. The joint manifold (J') was created with $a = 1$.

Causal connections during epileptic seizure.

Electrode implantation and recording. Patients had medically intractable seizures and were referred for epilepsy surgical evaluation. After detailed non invasive EEG, video-EEG and various imaging studies including MRI and PET, he underwent phase 2 invasive video-EEG examination. We used intracranial subdural electrodes (grids and strips) for the localization of epileptogenic zone. In the present study we used the original data obtained along the clinical investigation without any influence by the study. All the patients provided informed consent for clinical investigation and surgeries along institutional review board guidelines, according to the Declaration of Helsinki.

The patient underwent subdural strip and grid electrode implantation (AD-TECH Medical Instrument Corp., Racine, WI). Subdural electrodes (10 mm intercontact spacing) were implanted with the aid of neuronavigation and fluoroscopy to maximize accuracy[44] via craniotomy with targets defined by clinical grounds. Video-EEG monitoring was performed using

Brain-Quick System Evolution (Micromed, Mogliano Veneto, Italy). All signals were recorded with reference to the skull or mastoid at 1024 Hz sampling rate. To identify the electrode locations, the patient received an anatomical T1-weighted MPRAGE MRI before electrode implantation as well as a full head CT scan (1mm slices) after electrode implantation. Electrode locations were identified on the postimplantation CT scan using the software BioImage Suite[45] (http://www.bioimagesuite.org). These locations were then mapped to the preimplant MRI via an affine transformation derived from coregistering the preimplant and postimplant MRIs and postimplant MRI and CT scans using FLIRT[46] and the skull-stripping BET 2 algorithm[47], both part of the Oxford Centre for Functional MRI of the Brain (FMRIB) software library (FSL: www.fmrib.ox.ac.uk/fsl). The reconstructed pial surface was computed from the preimplantation MRI using FreeSurfer[48] (http://surfer.-nmr.mgh.harvard.edu/) and the electrode coordinates projected to the pial surface[49] to correct for possible brain shift caused by electrode implantation and surgery. Intraoperative photographs and ESM were used to corroborate this registration method. This pial surface projection method has been shown to produce results that are compatible with the electrode locations in intraoperative photographs (median disagreement of approx. 3 mm,[49]).

Preprocessing. Extracellular local field potential recordings were preprocessed: CSD was computed (at Fl1, Fl2, iP and Fb), 1-30 Hz Fourier filtering (4th order Butterworth filter) and subsequent rank normalization was carried out on the recordings.

Model parameters. Embedding delay was set to $\tau = 11$ samples (according to PACF) and embedding dimension was set to $D = 7$. The joint manifold (J') was created with $a = 1$.

Results on multiple seizures and multiple control sections. We examined the causal connectivity with the DC method between the fronto-basal (Fb), frontal(Fl1), frontolateral (Fl2) and inferior-parietal (iP) areas during seizure and interictal control conditions.

Our results showed that during epileptic seizures iP was found to be a dominant driver in the network and also the existence of non-observed drivers could be inferred (Extended Data Fig. 2). Connectivities calculated on 18 seizures showed various connection patterns: in 6 cases iP was a driver in the network, in 1 case Fb was a driving Fl1, in 1 case Fb and Fl2 were circularly coupled and the remaining 10 cases the nodes were driven by unobserved common cause.

During interictal control conditions Fb was found to be the dominant node (Extended Data Fig. 3). Connectivities were calculated on 16 control periods showed that Fb was playing a driver role in the network. Additionally in 5 cases iP also drove the other nodes, in 2 cases Fl1 was a driver of Fl2,

and in many cases circular connections were present between the nodes.

References

5. Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. *Science*, **Vol. 338**, No. 6106, pp. 496-500.

7. Harnack, D., Laminski, E., Schünemann, M., Pawelzik, K. R., 2017. Topological Causality in Dynamical Systems. *Physical Review Letters*, **Vol. 119**, No. 9, pp. 098301.

15. Rényi, A., 1959. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, **Vol. 10**, No. 1-2, pp. 193-215.

16. Pincus, S. M., 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, **Vol. 88**, No. 6, pp. 2297-2301.

17. Romano, S., Chelly, O., Nguyen, V., Bailey, J., Houle, M. E., 2016. Measuring dependency via intrinsic dimensionality. *23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016*, pp. 1207-1212.

21. Levina, E., Bickel, P. J., 2004. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, **Vol. 17**, pp. 777-784.

22. Farahmand, A. M., Szepesvári, Cs., Audibert, J.-Y., 2007. Manifold-adaptive dimension estimation. *Proceedings of the 24th international conference on Machine learning - ICML '07*, **Vol. 88**, No. 6, pp. 265-272, ACM Press, New York.

29. Wolf, A., Swift, J. B., Swinney, H. L. and Vastano, J. A., 1985. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, **Vol. 16**, No. 3, pp. 285-317.

30. Pincus, S. M., Gladstone, I. M., Ehrenkranz, R. A., 1991. A regularity statistic for medical data analysis. *Journal of clinical monitoring*, **Vol. 7**, No. 4, pp. 335-345.

31. Pincus, S. M., 1995. Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **Vol. 5**, No. 1, pp. 110-117.

32. Sauer, T., Yorke, J. A., Casdagli, M., 1991. Embedology. *Journal of Statistical Physics*, **Vol. 65**, No. 3-4, pp. 579-616.

33. Camastra, F., Staiano, A., 2016. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, **Vol. 328**, pp. 26-41.

34. Campadelli, P., Casiraghi, E., Ceruti, C., Rozza, A., 2015. Intrinsic dimension estimation: Relevant techniques and a benchmark framework. *Mathematical Problems in Engineering*, **Vol. 2015**, pp. 1-21.

35. Apostol, T.M., 1990. *Modular Functions and Dirichlet Series in Number Theory*, 2nd ed., Springer-Verlag, New York.

36. Cao, L., 1997. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, **Vol. 110**, No. 1-2, pp. 43-50.

37. Rhodes, C., Morari, M., 1997. The false nearest neighbors algorithm: An overview. *Computers & Chemical Engineering*, **Vol. 21**, Supplement, pp. S1149-S1154.

38. Casdagli, M., Eubank, S., Farmer, J. D., Gibson, J., 1991. *Physica D: Nonlinear Phenomena*, **Vol. 51**, No. 1-3, pp. 52-98.

39. Fraser, A. M., Swinney, H. L., 1986. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, **Vol. 33**, No. 2, pp. 1134–1140.

40. Nichkawde, C., 2013. Optimal state-space reconstruction using derivatives on projected manifold. *Physical Review. E*, **Vol. 87**, No. 2, pp. 022905.

41. Tamma, A., Khubchandani, B. L., 2016. Accurate determination of time delay and embedding dimension for state space reconstruction from a scalar time series. Preprint at http://arxiv.org/abs/1605.01571.

42. Small, M., Tse, C. K., 2004. Optimal embedding parameters: A modelling paradigm. *Physica D: Nonlinear Phenomena*, **Vol. 194**, No. 3, pp. 283 - 296.

43. Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C. E., Ding, M., 2016. Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. *Frontiers in Systems Neuroscience*, **Vol. 9**, pp. 189.

44. Erőss, L., Bagó, A. G., Entz, L., Fabó, D., Halász, P., Balogh, A., Fedorcsák, I., 2009. Neuronavigation and fluoroscopy-assisted subdural strip electrode positioning: a simple method to increase intraoperative accuracy of strip localization in epilepsy surgery. *Journal of neurosurgery*, **Vol. 110**, No. 2, pp. 327-331.

45. Duncan, J. S., Papademetris, X., Yang, J., Jackowski, M., Zeng, X., Staib, L. H., 2004. Geometric strategies for neuroanatomic analysis from MRI. *NeuroImage*, **Vol. 23**, Suppl. 1. S34-45.

46. Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, **Vol. 5**, No. 2, pp. 143-156.

47. Smith, S. M., 2002. Fast robust automated brain extraction. *Human Brain Mapping*, **Vol. 17**, No. 3, pp. 143-155.

48. Dale, A. M., Fischl, B., Sereno, M. I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, **Vol. 9**, No. 2, pp. 179-194.

49. Dykstra, A. R., Chan, A. M., Quinn, B. T., Zepeda, R., Keller, C. J., Cormier, J., Madsen, J. R., Eskandar, E. N., Cash, S. S., 2012. Individualized localization and cortical surface-based registration of intracranial electrodes. *NeuroImage*, **Vol. 59**, No. 4, pp. 3563-3570.