https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions

Detecting hydrological connectivity using causal inference from time-series: synthetic and real karstic study cases

Damien Delforge[1,2], Olivier de Viron[3], Marnik Vanclooster[1], Michel Van Camp[2], and Arnaud Watlet[4]

[1]Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[2]Royal Observatory of Belgium, Brussels, Belgium
[3]Littoral, Environnement et Sociétés, Université de La Rochelle and CNRS (UMR7266), La Rochelle, France
[4]British Geological Survey, Nottingham, UK

Correspondence: Damien Delforge (damien.delforge@uclouvain.be)

Abstract. We investigate the potential of causal inference methods (CIMs) to reveal hydrological connections from time-series. Four CIMs are selected from two criteria, linear or nonlinear, and bivariate or multivariate. A priori, multivariate and nonlinear CIMs are best suited for revealing hydrological connections because they suit nonlinear processes and deal with confounding factors such as rainfall, evapotranspiration, or seasonality. The four methods are applied to a synthetic case and a real karstic study case. The synthetic experiment indicates that, unlike the other methods, the multivariate nonlinear framework has a low false-positive rate and allows for ruling out a connection between two disconnected reservoirs forced with similar effective precipitation. However, the multivariate nonlinear method appears unstable when it comes to real cases, making the overall meaning of the causal links uncertain. Nevertheless, all CIMs bring valuable insights into the system's dynamics, making them a cost-effective and recommendable tool for exploring data. Still, causal inference remains attached to subjective choices and operational constraints while building the dataset or constraining the analysis. As a result, the robustness of the conclusions that the CIMs can draw deserves to be questioned, especially with real and imperfect data. Therefore, alongside research perspectives, we encourage a flexible, informed, and limit-aware use of CIMs, without omitting any other approach that aims at the causal understanding of a system.

1 Introduction

Causal inference methods (CIMs) aim at identifying causal interactions between variables from variables (Spirtes et al., 2000; Pearl, 2009). When applied to time-series, these empirical methods are built upon the principle of priority of the cause, which goes back to Hume (Hume, 1748). They infer causation from the expected time-dependencies between causes and effects, i.e., the causes must occur before the effects. They have evolved throughout the 20th century to go beyond the well-known correlation, or cross-correlation, between two time-series (see Runge et al., 2019a, for a broad review). Although widely used, the correlation or cross-correlation method is criticized for its inability to identify nonlinear causal relationships. Besides, the correlation cannot discriminate causal links from associations resulting from confounding factors. Indeed, dependencies between variables can be explained either by a direct causal link or through the common cause principle (Reichenbach, 1956; Runge et al., 2019a). The common cause principle tells us that dependencies between the variables could result from a third cause

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

acting on the variables. Nowadays, a plethora of CIMs has been developed, differing in hypotheses or application fields. Some CIMs explicitly deal with, either or both, nonlinear dependencies or confounding factors through multivariate analysis. These new CIMs are of growing interest in Earth, land, and hydrological sciences (Meyfroidt, 2016; Runge et al., 2019a; Goodwell et al., 2020). In hydrology, applications of CIMs remain rare and cover, for example, the potential causal feedbacks between soil moisture and precipitation (Salvucci et al., 2002; Tuttle and Salvucci, 2017), cross-scales rainfall interactions (Molini et al., 2010), the ecohydrological feedback processes (Ruddell and Kumar, 2009), or the study of hydrological connectivity (Sendrowski and Passalacqua, 2017; Rinderer et al., 2018) as in this article.

The study of hydrological connectivity aims at identifying the paths taken or that could be taken by water. There is a priori no causal interaction, i.e., flow, possible between two points in space that do not benefit from a hydrological connection. This link between the concepts of hydrological connectivity and causality motivates the study of one given the other and vice versa. There are different types of connections and various ways to refer to them (Bracken et al., 2013). We refer to the terminology of Rinderer et al. (2018), which is inspired by and borrowed from the field of neurological and brain connectivity (Friston, 2011). There are three types of connectivity: (i) structural, (ii) functional, and (iii) effective connectivity. The structural connectivity is derived from the medium and highlights the potential, static, and time-invariant water flow paths from the geological environment's topography, spatial adjacency, or contiguity. The functional one is dynamic and is retrieved from statistical time-dependencies between local hydrological variables. A statistical association may result from confounding factors, e.g., rainfall acting on two disconnected reservoirs or a shared seasonal pattern. Therefore, dependencies do not necessarily imply factual causation, such as process-based water flows. Then, the functional connectivity is a matter of cross-predictability and still reflects potential rather than actual flow paths for water. CIMs with a multivariate framework address confounding factors. They offer the promises of discriminating functional connectivity from the effective one, which reveals actual flow paths and processes within the system. From the structural to the effective connectivity through the functional one, the search for hydrological connections can be seen as a progressive constraint from the potential paths to the actual paths taken by water.

Because of their hidden and heterogeneous structures, karst systems are regularly studied through time-series analysis or other empirical approaches to derive a causal or functional representation of the system (Bakalowicz, 2005). The application of CIMs to karst is therefore very relevant. However, karst systems present some challenges for causal inference (Bakalowicz, 2005). In particular, the heterogeneity of the fractured geological environment is difficult or impossible to observe, characterize, or hypothesize. This hiddenness jeopardizes the derivation of a reliable map of structural connectivity to guide causal inference, as recommended in Rinderer et al. (2018).

To date, the most commonly used method remains the linear cross-correlation function (CCF), including for karsts (e.g., Angelini, 1997; Larocque et al., 1998; Bailly-Comte et al., 2008; Labat et al., 2000; Mathevet et al., 2004; Watlet et al., 2018). Some study cases are now involving linear multivariate methods (e.g., Salvucci et al., 2002; Tuttle and Salvucci, 2017), including for karst (Kadić et al., 2018). Hydrological systems were first theoretically characterized by linear methods (Dooge, 1973). Indeed, we expect some linearity in mass transfers and positive time-dependent correlations between causes and effects, e.g., precipitation and discharge. This delayed linear transfer allows hydrological signals to be related by the unit hydrograph theory as a simple convolution window (Dooge, 1973). Besides, linear methods also benefit from the computational efficiency

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

of linear algebra, which makes linear CIMs less time-consuming. Despite their monotonic behavior, hydrological systems, especially karsts, are nevertheless sensitive to initial conditions, i.e., nonlinear. Nonlinearity is imputed to inherently nonlinear hydrological processes such as power laws or threshold effects triggering flows (Bakalowicz, 2005; Blöschl and Zehe, 2005). For this reason, nonlinear CIMs may be more suited, although not specifically designed to deal with threshold effects.

To investigate how the CIMs can help to understand the dynamics of the karst system and how the method hypotheses impact this understanding, we compare the connectivity inferred from the four CIMs (Table 1), all operating on the time-domain: the linear and bivariate CCF, the bivariate and nonlinear Convergent Cross Mapping (CCM) method (Sugihara et al., 2012); and two multivariate methods, one linear (ParCorr) and one nonlinear (CMI), both part of the same causal inference framework called PCMCI implemented in the Tigramite Python package (Runge et al., 2019b). First, the four methods are used in a synthetic case study, where two hydrological discharge variables with similar meteorological forcing are either connected or disconnected. Then, the same CIMs are applied to a real-time-series dataset from the Rochefort cave in Southern Belgium. The real dataset include rainfall, potential evapotranspiration data, electrical resistivity patterns in the subsurface obtained from obtained from a geophysical monitoring experiment using time-lapse Electrical Resistivity Tomography (ERT)(Watlet et al., 2018), and drip discharge time-series with distinct dynamical patterns monitoring percolation at three spots within the cave. Time-series also have different numbers of missing values, unevenly distributed over time, allowing a discussion of the impact of missing values on the analysis. To appreciate the results, previous dye tracing tests have revealed fast connected preferential flow between the surface and a particular spot in the cave (Poulain et al., 2018). This prior knowledge can be seen as a reality check on the blind CIMs.

2 Materials and Methods

2.1 Causal Inference Methods (CIMs)

Four CIMs are selected based on two binary criteria: linear or nonlinear and bivariate or multivariate (Table 1). They operate in the time-domain and investigate time-dependencies up to a maximum time delay d_{max}. The methods are presented here in a nutshell; a more detailed description of the methods and their practical implementation is available in the supplementary materials (SM1). The Cross-Correlation Function (CCF) and the Convergent Cross-Mapping (CCM) (Sugihara et al., 2012; Ye et al., 2015) methods respectively check for linear and nonlinear bivariate dependencies.

The two other methods are part of a multivariate causal inference framework implemented in the Tigramite Python package (v.4.1) and based on the PCMCI conditional independence algorithm (see Runge et al., 2019b, and SM1.3). Being multivariate, those methods can cope with confounding variables. The general principle of causal inference using conditional independence is the following: identifying a causal lag between two variables is checking whether or not time-dependencies pertain while removing the effect of the other variables and potential causal lags. Therefore, the analysis framework is stochastic because a background noise must persist to assess the variables' independence after removing deterministic effects. Both methods test the independence between the series either with a linear test, the partial correlation method (ParCorr, section 2.1.3), or a nonlinear one, the Conditional Mutual Information (CMI, section 2.1.4).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions

Table 1. Selected causal inference methods (CIMs)

	Linear	**Nonlinear**
Bivariate	Cross-Correlation Function (CCF)	Convergent Cross-Mapping (CCM)
Multivariate	Partial Correlation (ParCorr)	Conditional Mutual Information (CMI)

In practice, however, one cannot remove the effect of all potentially causal delays without falling into a curse of dimensionality affecting the causal detection power of the approach (Runge et al., 2019b). For this reason, ParCorr (and CMI) are coupled with the PC and MCI algorithms (or PCMCI by concatenation). The first step is the PC algorithm, named after its authors Peter Spirtes and Clark Glymour (1991). Using the independence test, PC iteratively selects a subset of potential causal variables and delays for each variable to avoid the curse of dimensionality (Runge et al., 2019b, SM1.3). The second step, Momentary Conditional Independence (MCI), applies the independence test based on the subsets identified during the PC step. The PC procedure has a tuning hyperparameter named α_{PC}, which controls the number of potential causes. α_{PC} varies from 0 to 1, where 1 is the less restrictive case which implies not pre-selection.

While the CMI method is the most promising in that it does not assume linearity and accounts for confounding effects, as for the other CIMs, the reliability of the reported causal relationships nevertheless depends on underlying hypotheses (discussed in Runge, 2018a). Perhaps, the most important but the most challenging to verify and conceptualize in practice is the hypothesis of causal sufficiency. Causal sufficiency implies that the analysis should include all potential common causes. This is indeed difficult to verify because (i) potential causes, as variables are generally dictated by data availability, (ii) one is not supposed to know potential causes before the analysis, or (iii) the concept of variable is mathematically abstract and building a finite and parsimonious time-series dataset require a prior or tacit exercise of conceptualization of the continuous system. In this way, one can say that the dataset is a hypothesis in itself, which, if improperly framed, can induce spurious causal links even with the best CIM available.

2.1.1 Cross-Correlation Function (CCF)

The CCF method is the most common to analyze time-dependencies and address causality (e.g., Angelini, 1997; Larocque et al., 1998; Labat et al., 2000; Mathevet et al., 2004; Bailly-Comte et al., 2008; Watlet et al., 2018). A variable X_t is said to be a cause of variable Y_t if the Pearson's correlation coefficient ρ between Y_t and X_{t-d} is significant on their overlapping domain for at least one realistic value of d up to d_{max}.

2.1.2 Convergent Cross-Mapping (CCM)

The CCM method is primarily designed to reveal weak nonlinear interactions between time-series (Sugihara et al., 2012; Ye et al., 2015). CCM tests if dynamic trajectories behave consistently while the system revisits the same states, i.e., dynamic recurrence. The system states are usually unknown; they are approximated by trajectory segments found in a trajectory matrix

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

M_Y given by the Takens' embedding theorem (Takens, 1981): $M_Y = \{Y_t, Y_{t-1}, ..., Y_{t-(m-1)}\}$, where m is the embedding dimension. In this case, with unit lags between time-series, m corresponds to the length of the segments and can be optimized using self-forecasting performance while predicting points in Y_t from their nearest neighbors in M_Y (Sugihara and May, 1990; Delforge et al., 2020a). This length m is set to two days in this study, due to the overall good performance of this value, i.e., $M_Y = \{Y_t, Y_{t-1}\}$. For a causal analysis and to check if X_t is a cause of Y_t, CCM makes forecasts of the points in X_t from points in X_t identified from Y_t. For a single forecast at a time of reference t^*, CCM selects the time-indices of the $m+1$ nearest neighbors of the state Y_{t^*}. The points in X_t mapping to the nearest-neighbors' time-indices are averaged to make a forecast of X_{t^*}. The nearest-neighbor time-indices are selected such that they are at least tw days away from the reference state t^*, where tw is the Theiler window (Theiler, 1986). This window ensures that nearest neighbors are not direct temporal neighbors but recurrent ones remote in time, following the CCM philosophy. The Theiler window tw is set to 10 days in this study such that nearest-neighbor states are likely to be separated from their reference point by at least one rainfall event.

In practice, the time to prediction tp varies between $[-d_{max}, 0]$ to evaluate the time-dependencies between X_{t+tp} and Y_t and infer causality from the principle of priority (Ye et al., 2015). In this case, the time-indices of the nearest neighbors of Y_{t^*} are simply shifted by tp to identify the states in X_t that are averaged into a single prediction of state X_{t^*+tp}. For each tp, we forecast the full vector X_{t+tp} of prediction 100 times based on bootstrapped samples of M_Y (See SM1.2). CCM forecast skills and time-dependencies are appreciated with the mean Pearson's correlation between the 100 estimates and the original time-series. Our implementation of CCM is the one developed by Delforge et al. (2020a).

2.1.3 Partial Correlation Test (ParCorr)

The ParCorr conditional independence test is like Granger causality and relies on linear vector autoregressive models (Granger, 1969). The principle is the following: X_{t-d} causes Y_t if a significant correlation exists between X_{t-d} and Y_t after removing the linear effect of all the potentially causal variables and delays, including their own past. There are recent applications of Granger causality or partial correlations in hydrology (e.g., Salvucci et al., 2002; Tuttle and Salvucci, 2017) or in the case of karst (Kadić et al., 2018). In the case of ParCorr, α_{PC} is automatically optimized between 0.05 and 0.5 for ParCorr based on Akaike's Information criterion of model performance (Akaike, 1974).

2.1.4 Conditional Mutual Information Test (CMI)

CMI is a nonlinear alternative to the ParCorr independence test running with the PCMCI algorithm (SM1.3). CMI is a multivariate extension of the concept of entropy transfer (Schreiber, 2000; Sendrowski and Passalacqua, 2017), i.e., another bivariate nonlinear CIM such as CCM. CMI is evaluated with a nearest-neighbor estimator (Frenzel and Pompe, 2007; Vejmelka and Paluš, 2008) coupled with a shuffling significance test (Runge, 2018b). Several methods exist to estimate mutual information within the Tigramite package. However, the nearest-neighbor estimator is recommended for time-series below 1000 samples, which is the case in this study. Unlike ParCorr, the α_{PC} value is not optimized and has to be set manually. We considered two values, a less restrictive one of 0.2 and a more restrictive one of 0.05.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

CMI, or nonlinear independence tests in general, are very computationally expansive and quickly require high-performance computers or significant computational time depending on the size of the dataset and the hyperparameters of the analysis (see Runge, 2018b). In comparison, the synthetic experiment at section 3 required two weeks of computation with CMI, while the result with ParCorr was almost instantaneous.

2.2 Study Site and Data

The karstic study site is located in the Calestienne, a band of outcropping Devonian limestone crossing southern Belgium. The time-series dataset results from the monitoring of the Lorette cave (Fig. 1.a), one of the largest caves sited next to the city of Rochefort (Watlet et al., 2018; Poulain et al., 2018, and references therein). The dataset is completed by daily potential evapotranspiration data (ET, Fig. 1.c) estimated with the Penman-Monteith FAO-56 method (Allen et al., 1998) and data from a PAMESEB agrometeorological station located 3 km from the study site in Jemelle. Other time-series are obtained from sensors on site (Fig. 1.a). Rainfall data (RF, Fig. 1.a) are daily average from a Lufft tipping bucket rain gauge with a 1 min sample rate located at the surface (elevation ∼225 m AOD). Inside the cave (elevation ∼190 m AOD), two drip discharge monitoring devices (P1, P2, Fig. 1.a) are installed within the main chamber accessible from a sinkhole, which constitutes the cave entrance. In particular, P1 monitors an active dripping point due to a visible fracture on the chamber's ceiling. Based on dye injection at the surface and in-cave tracing, a connection and preferential flow path between the dye injection point (DT, Fig. 1.a) and P1 was identified (Poulain et al., 2018). The breakthrough curve showed an initial arrival time of 3.75 hours, a sustained peak for 80 hours, and a tail lasting up to 120 days. However, sporadic peaks in concentration were observed after every rainfall event, reacting after 1.48 hours, peaking after 7.2 hours, and lasting up to 30 hours on average. P2 monitors a dripping spot draining a porous limestone area. The last one, P3, located in the North gallery, monitors slow discharge from drops falling from one single stalactite below a massive limestone layer. P1, P2, and P3 (Fig. 1.c) are daily means of the percolation rate.

An Electrical Resistivity Tomography (ERT) profile was installed to investigate the hydrology of the subsurface and potential connections above the cave (Watlet et al., 2018). ERT is a geophysical monitoring tool to study various types of hydrological processes (see Slater and Binley, 2021, and references therin). At the study site, the ERT profile is not flat as it starts from the depression of a sinkhole where the entrance to the cave is located (ERT, Fig. 1.a). The ERT experiment allowed collecting ERT datasets daily between 2014 and 2017, which still represents, to the best of the author's knowledge, the longest, high-resolution ERT monitoring experiment conducted in a karst environment (see Watlet et al., 2018, for details). This dataset consists of 1558 spatial cells. Each of them is assigned to a resistivity time-series defined on 465 daily time-steps defining the temporal dimension of the dataset. As a necessary prior step to causal inference, the 1558 time-series were dimensionally reduced to six time-series clusters (R0 to R5, Fig. 1, b, and c). We used hierarchical agglomerative clustering with the Ward linkage method to minimize the squared distance between time-series within clusters. This algorithm is similar to k-means. As clustering was applied on the standardized ERT dataset, clusters represent groups of linearly correlated resistivity dynamics. The optimal number of clusters of 6 was selected using the optimal Silhouette Index as a clustering evaluation metric (Rousseeuw, 1987). The methodological aspects associated with clustering are extensively covered in another issue (Delforge et al., 2020b).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

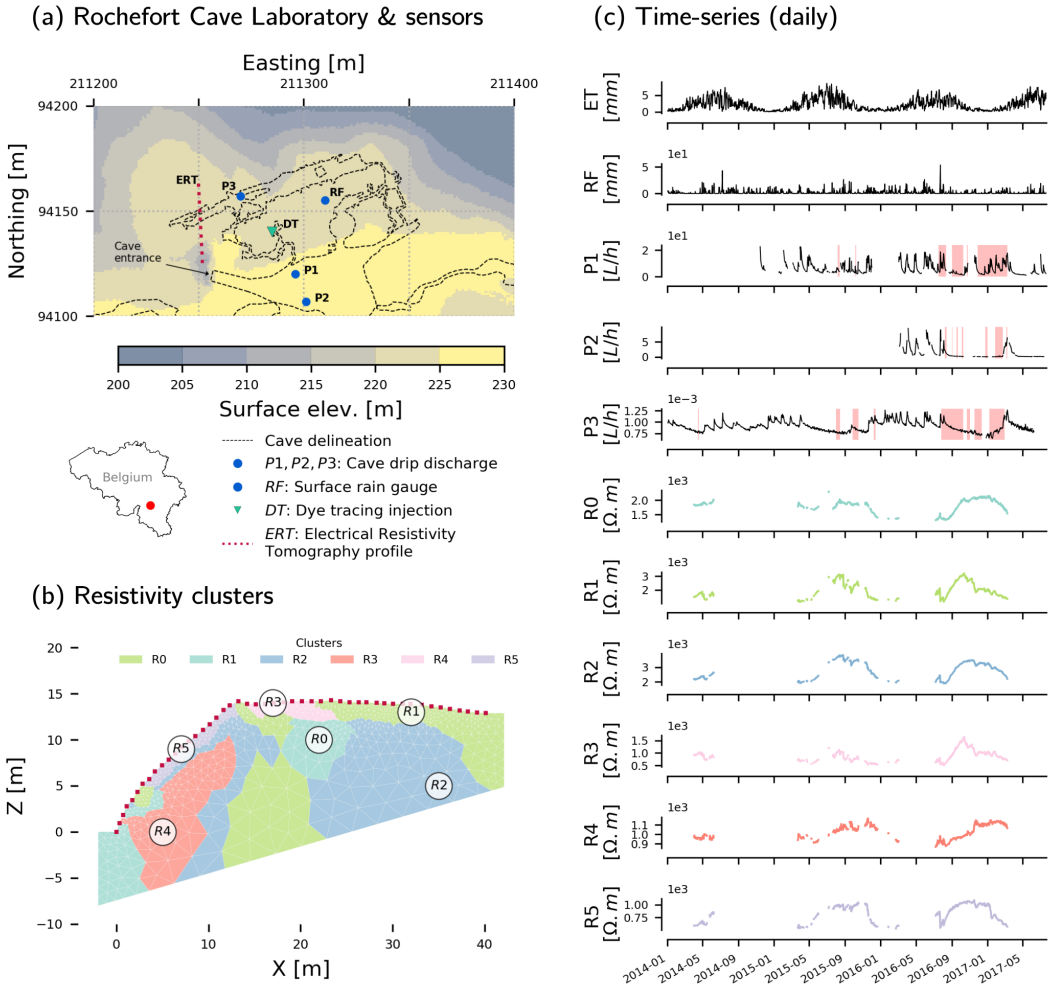

Figure 1. Study site and data: (a) Rochefort cave laboratory and sensors (EPSG: 31370), (b) resistivity clusters obtained from hierarchical agglomerative clustering of standardized resistivity data (Watlet et al., 2018; Delforge et al., 2020b), (c) daily time-series dataset. Resistivity time-series (c, R0 to R5) are the mean resistivity variations per cluster (b). Potential evapotranspiration data (ET) are obtained from an agrometeorological station (PAMESEB) located 3 km from the site. The red areas on (c) show the time-domain resulting from the conditioning on past delays with $d_{max} = 5$ days while considering respectively P1, P2, and P3 only in the causal dataset. A hydrological connection was identified by dye injection and tracing from the surface (DT) to P1 (Poulain et al., 2018). Source: Digital Elevation Model from Service Public de Wallonie, Cave delineation from Watlet et al. (2018).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

Table 2. Summary statistics of the time-series variable.

Statistic	ET [mm]	RF [mm]	P1 [L/h]	P2 [L/h]	P3 [L/h]	R0 [Ω.m]	R1 [Ω.m]	R2 [Ω.m]	R3 [Ω.m]	R4 [Ω.m]	R5 [Ω.m]
Count	1297	1297	718	366	1223	465	465	465	465	465	465
Mean	2.2	2.0	6.13	1.30	9.06E-04	1.80E+03	1.95E+03	2.80E+03	8.78E+02	1.02E+03	8.28E+02
Std dev.	1.8	4.2	4.33	1.92	1.11E-04	2.38E+02	5.68E+02	5.64E+02	2.35E+02	7.75E+01	1.73E+02
Min	0.0	0.0	1.05	0.00	6.35E-04	1.30E+03	1.11E+03	1.88E+03	5.07E+02	8.67E+02	5.33E+02
10%	0.3	0.0	1.91	0.05	7.76E-04	1.37E+03	1.28E+03	2.03E+03	5.92E+02	9.23E+02	5.89E+02
25%	0.7	0.0	3.00	0.09	8.21E-04	1.69E+03	1.48E+03	2.27E+03	7.06E+02	9.59E+02	6.45E+02
50%	1.6	0.1	4.57	0.27	8.99E-04	1.85E+03	1.85E+03	2.83E+03	8.51E+02	1.03E+03	8.52E+02
75%	3.3	2.1	8.46	1.73	9.75E-04	1.96E+03	2.36E+03	3.29E+03	1.01E+03	1.10E+03	9.92E+02
90%	5.0	6.2	12.74	4.20	1.06E-03	2.08E+03	2.86E+03	3.48E+03	1.21E+03	1.13E+03	1.03E+03
Max	8.6	53.8	22.66	9.45	1.26E-03	2.30E+03	3.21E+03	3.83E+03	1.65E+03	1.17E+03	1.09E+03

R0 is associated with a dense limestone area in the model's center (Fig. 1.b, X=22 m, Z=10 m). R1 shows responsive resistivity dynamics at the plateau's surface (Fig. 1.b, X=32 m, Z=13 m), as well as in a fractured area around coordinate X=15 m below the surface patterns of R3. R2 is rather representative of the resistivity patterns of the limestone matrix. R4 is associated with a clayey limestone layer located below the patterns of slope surface R5 (Watlet et al., 2018; Delforge et al., 2020b). We expect causal links to appear primarily between P1 and the near-surface resistivity patterns R1 or R3.

Time-series of Fig. 1.c are the 11 inputs for the four selected CIMs. Table 2 shows their statistics. Bivariate CIMs (CCF and CCM) are applied between each pair of time-series on their overlapping time-domain with a maximum causal delay $d_{max} = 5$ days. This maximum delay allows covering the full time-span of preferential flow peaks lasting up to 80 h (Poulain et al., 2018). For the same reason, we use the same d_{max} for the multivariate methods. In general, it is preferable to use large d_{max} values, because any delay that can be considered as a potential cause must be included in the analysis to satisfy the hypothesis of causal sufficiency. On the other hand, d_{max} cannot take a larger value because of the missing data. Indeed, the PCMCI algorithm dismisses all time slices of samples where missing values occur in any variable and their lags up to $2d_{max}$, which limits the overall overlapping time-domain to 48 days. This low value is mainly imputed to the short time-domain of P2 (Fig. 1.c, Table 2). Concerned that this number would impact the robustness of the analysis, we also applied PCMCI with $d_{max} = 5$ days while considering one percolation data at a time. In this way, P1, P2, P3 are considered separate and independent drainage systems and the overlap domains become larger: 184, 62, and 218 days respectively. These conditioning domains are shown in red for the respective time series considered (Fig. 1.c.).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions

3 Synthetic Study Case

3.1 Conceptual Model

The four CIMs are first applied to test and explore the following theoretical assertion: nonlinear and multivariate CIMs are best suited to detect effective hydrological connectivity. For this purpose, a simple hydrological reservoir model is inspired by the problematic case of the common cause (Fig. 2). The common cause problem is easily conceived through two separate and independent reservoirs (A and B) benefiting from the same meteorological forcing. Without an effective connection, they will nevertheless show a strong temporal dependence. In this case, two reservoirs, A and B, are subject to the same forcing effective precipitation P_{eff} (i.e., precipitation minus evapotranspiration). If A and B are disconnected, the ideal CIM would then reject the effective connection between the reservoirs and their discharges Q_A and Q_B if P_{eff} is included in the multivariate analysis. However, B responds systematically one day later than A to P_{eff}. Hence, with bivariate methods and the priority principle only, Q_A would seemingly cause Q_B. For comparison, we consider a case where Q_A and Q'_B are effectively connected as if they were contributing to the same drainage network, with Q_A upstream of Q'_B. Noteworthily, this experiment does not cover the case of nonlinearities arising from threshold effects and intermittent processes. Instead, we assume a continuous causal relationship over time, as the CIMs do.

The model is forced by real effective precipitation data P_{eff} monitored at the study site (section 2.2). Both reservoirs take as input a net inflow term I_A and I_B resulting from the application of the unit hydrographs H_A and H_B as linear transfer functions convolved forwardly on a noisy precipitation input ($H_R * P_{eff,R}$ with $*$ the convolution operator). Adding some noise is mandatory to check for conditional independence (section 2.1). A multiplicative noise term is preferred as hydrological variables are often characterized by multiplicative noise (e.g., Rodriguez-Iturbe et al., 1991). With R either A or B, $P_{eff,R} = P_{eff} + \varepsilon_R P_{eff}$, with ε_R being randomly generated from a normal distribution with zero mean and standard deviation equal to ε_{lvl} times the standard deviation of P_{eff}. The parameter ε_{lvl} is always identical for A and B, such that ε_A and ε_B have the same distribution. The continuity equation gives the reservoirs storage dynamics: $dS_R/dt = I_R - Q_R$. The outflow Q_R introduces some nonlinearities through a typical nonlinear storage-discharge relationship $Q_R = k_R S_R^{e_R}$, with k_R and e_R the discharge coefficient and the nonlinear exponent. Such power-law formulations are typical in hydrology (Dooge, 1973) and common while modeling karsts as well (Hartmann et al., 2014; Jourde et al., 2015).

Table 3. Model parameters for the synthetic cases

Model	H_A	k_A	e_A	H_B	k_B	e_B
1	[0.7, 0.2, 0.1]	0.1	1	[0.1, 0.8, 0.1]	0.1	1
2	[0.7, 0.2, 0.1]	0.1	1	[0.1, 0.8, 0.1]	0.01	1.5
3	[0.7, 0.2, 0.1]	0.01	1.5	[0.1, 0.8, 0.1]	0.1	1
4	[0.7, 0.2, 0.1]	0.01	1.5	[0.1, 0.8, 0.1]	0.01	1.5

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions

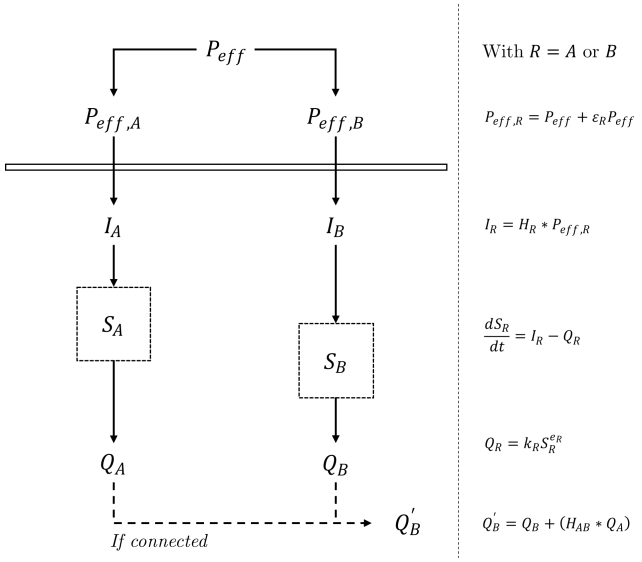

Figure 2. The conceptual and mathematical model for the synthetic study case. Two reservoirs A and B are forced by inflows I_A or I_B resulting from the unit hydrograph H_A and H_B forward convolution on noisy variant $P_{eff,A}$ and $P_{eff,B}$ of the same effective precipitation P_{eff}. The storage S_A and S_B dynamics follows a typical continuity equation $dS/dt = I - Q$, where the discharge Q_A and Q_B follow a nonlinear power law $Q = kS^e$ with two parameters k and e. The reservoir B responds mainly 1 day after A, which introduced a systematic time-dependency between the reservoir suggesting causation. Q'_B is a flow downstream of Q_B draining Q_A and transferred considering the unit hydrograph H_{AB}. The causal analysis involves either the disconnected case Q_A and Q_B or the connected case Q_A and Q'_B. Multivariate methods includes P_{eff} within the analysis.

For the synthetic cases, we derived four models based on the parameters presented in Table 3. The unit hydrographs H_A and H_B are constant, with their maxima differing by one daily time-step. The lag introduces the desired constant time-dependencies between the two reservoirs despite the absence of connection. The recession parameters allow generating distinct dynamic patterns with various degrees of nonlinearity thanks to ε_R. In addition, we also considered 14 stochastic noise level $\varepsilon_{lvl} \in \{0.05, 0.1, \ldots, 0.65, 0.70\}$. With the four combinations of Table 3 and the 14 noise levels, 56 datasets were generated from four years of effective precipitation data P_{eff} (2014-2018) and initial storages S_A and S_B equal to 30 mm. Only the last year of the three variables P_{eff}, Q_A, Q_B were considered for the causal inference experiment. The data generation is repeated to produce 56 additional datasets with an effective connection. Q_A and Q_B are causally related by overwriting Q_B such that $Q'_B = Q_B + (H_{AB} * Q_A)$ where $H_AB = [0.1, 0.8, 0.1]$ is a linear transfer function convolved forwardly on Q_A. Finally, the whole synthesis process is repeated to generate first-order differenced datasets, making a total of 224 datasets with 4 model combinations, 14 noise levels, connected or not, and differenced or not. The primary purpose of the difference is to create a case where the shared seasonality that affects the bivariate dependencies between the data is eliminated.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

3.2 Result

Figure 3 shows the results of the experiment. It depicts the average and interquartile range of Q_A-Q_b time-dependencies, or Q_A-Q'_B if connected, obtained with the four CIMs on the datasets synthesized from the numerous model combinations (Table 3) and noise levels with a maximum delay $d_{max} = 5$ days. The multivariate analysis includes P_{eff}. We distinguish between cases where the reservoirs are connected or not and where the data are differenced or not. Regarding the bivariate methods, CCF and CCM both exhibit sustained time-dependencies when the data is not differenced due to the auto-correlation in the series and seasonality. From differenced data, the results better screen the expected peak at lag one. However, it is also the case while the reservoirs are disconnected because of the confounding effect related to the common forcing of the two reservoirs. This is not an effective connection but a functional and apparent one resulting from the delayed responses of the two reservoirs. The sustained time-dependency of CCM over the lag of 2 days is an artifact of the embedding dimension defining the length of trajectory segments, which is two days in this case (see SM1.2).

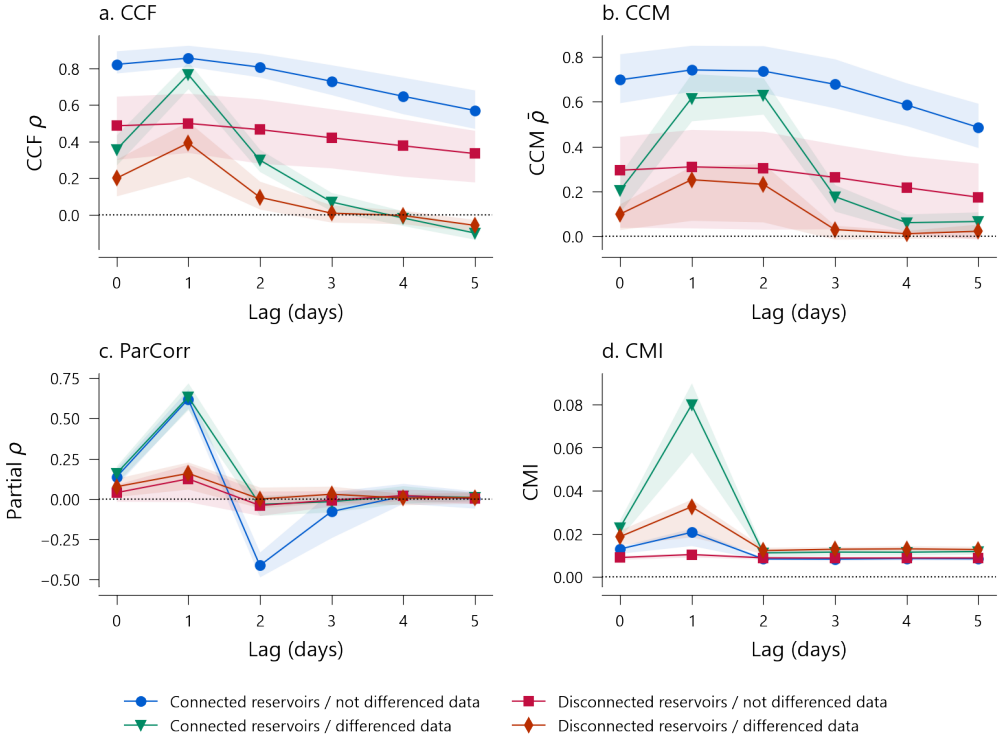

Figure 3. Patterns of statistical time-dependencies between Q_A and Q_b (Q'_B if connected) for the four CIMs (a. to d.). Lines are the average statistics for 56 synthetic datasets obtained from four different model structures (Table 3) and 14 distinct noise levels. The envelope represents the interquartile range of the statistics. In general, connected reservoirs show a $Q_A \rightarrow Q'_B$ causal dependencies at a lag of one day. However, except for CMI on the not differenced data, disconnected reservoirs show a non-causal, yet, significant statistical dependencies $Q_A \rightarrow Q_B$ at lag one day, since reservoir A mostly reacts one day before B to the effective precipitation.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

Regarding the multivariate CIMs patterns, the linear ParCorr method seems to discriminate the connected case from the disconnected case. Still, it always shows a peak at lag 1, whatever the cases, which could be misinterpreted as an effective connection. Only the nonlinear CMI method applied on the not differenced data seems to reject the idea of connection when it is effectively absent. This finding supports our theoretical assertion: the multivariate nonlinear method is the best suited to address effective hydrological connectivity. Furthermore, the method appears to perform better if seasonality is left present in the time-series. Still, Fig. 3 shows the pattern of the statistics, not the result of a causality test and its p-value.

Since we know for each simulation whether or not there is an actual causal link between A and B, Table 4 assesses the performance of the statistical test from true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for the problematic lag of 1 day. We consider the multivariate PCMCI methods ParCorr and CMI, with the latter having two different α_{PC} values for pre-selection of potential causes (PC stage). Two levels of significance are considered at 99% and 95% based on the p-values obtained by the tests. Table 4 shows that, for a similar level of accuracy, CMI for not differenced data has higher precision and a lower false-positive rate, meaning that positive tests are likely to detect actual causal relations. This is particularly contrasting with other methods and provides a valuable piece of information. However, the high precision comes at the cost of a low recall: CMI misses about half of the actual causal links. On the contrary, ParCorr misses none but has a bad precision, i.e., many false positives. This analysis thus provides an overview of the contrasts between methods. Of course, this virtual three-variable configuration is far from representative of the great variety of natural hydrological systems and their spatiotemporal organizations.

4 Real Study Case

4.1 Bivariate Methods

Figure 4 shows pairwise dependencies between first-order differenced time-series, for a better screening of time-dependencies using bivariate methods (Fig. 3). Detailed time-dependencies are also reported in the Supplementary Materials (SM2.1, SM2.2). The links displayed are those for which the correlation value is significantly different from zero following a Student's t-test. The CCF method (Fig. 4.a) is reporting many potential linear causal associations. If causality is hard to infer from such a diagram, the results make sense in general. A typical pattern is that the sign of time-dependencies tends to flip after a few delays due to RF's forcing and the fact that dry periods come after the rain. Considering low delays, ET is positively related to the resistivity patterns, mostly at the surface (R1, R3). ET is negatively correlated with P1 only, which is known to drain fast flow from the surface through the epikarst. All variables are dependent on RF, the main confounding factor, but R0, associated with a dense limestone area, depends to a lesser extent. R0 and R4 put apart, the quartet R1, R2, R3, and R5 exhibit strong positive and contemporaneous correlations together. P1, P2, and P3 also are instantaneously related. P1 and P2 have strong dependencies with all resistivity patterns, but inconsistent and positive correlations are reported between the anomalous resistivity series R4 representative of the clayey limestone. P3 seems rather dependent on R5 (slope) and R2 (mostly limestone matrix), makings sense since P3 most likely drains the matrix's delayed flow.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions

Table 4. Causal test statistics for the synthetic cases at lag 1 day

Method	**ParCorr**		**CMI** $\alpha_{PC} = 0.05$		**CMI** $\alpha_{PC} = 0.2$	
Datasets	Not diff.	Diff.	Not diff.	Diff.	Not diff.	Diff.
Data Statistics						
Total count	112	112	112	112	112	112
Actual +	56	56	56	56	56	56
Actual -	56	56	56	56	56	56
Test results						
TP at 99%	56	56	28	56	15	56
(95%)	(56)	(56)	(34)	(56)	(27)	(56)
FP at 99%	29	29	3	28	1	28
(95%)	(35)	(32)	(4)	(42)	(2)	(35)
TN at 99%	27	27	53	28	55	28
(95%)	(21)	(24)	(52)	(14)	(54)	(21)
FN at 99%	0	0	28	0	39	0
(95%)	(0)	(0)	(22)	(0)	(29)	(0)
Test metrics						
Accuracy[1] 99%	0.74	0.74	0.72	0.75	0.64	0.75
(95%)	(0.69)	(0.71)	(0.77)	(0.62)	(0.72)	(0.69)
Precision[2] 99%	0.66	0.66	0.90	0.67	0.94	0.67
(95%)	(0.62)	(0.64)	(0.89)	(0.57)	(0.93)	(0.62)
Recall[3] 99%	1.00	1.00	0.5	1.00	0.3	1.00
(95%)	(1.00)	(1.00)	(0.61)	(1.00)	(0.48)	(1.00)
FP rate[4] 99%	0.52	0.52	0.05	0.5	0.02	0.5
(95%)	(0.63)	(0.57)	(0.07)	(0.75)	(0.04)	(0.63)

[1](TP+TN)/(TP+FP+FN+TN); [2]TP/(TP+FP); [3]TP/(TP+FN); [4]FP/(FP+TN)

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

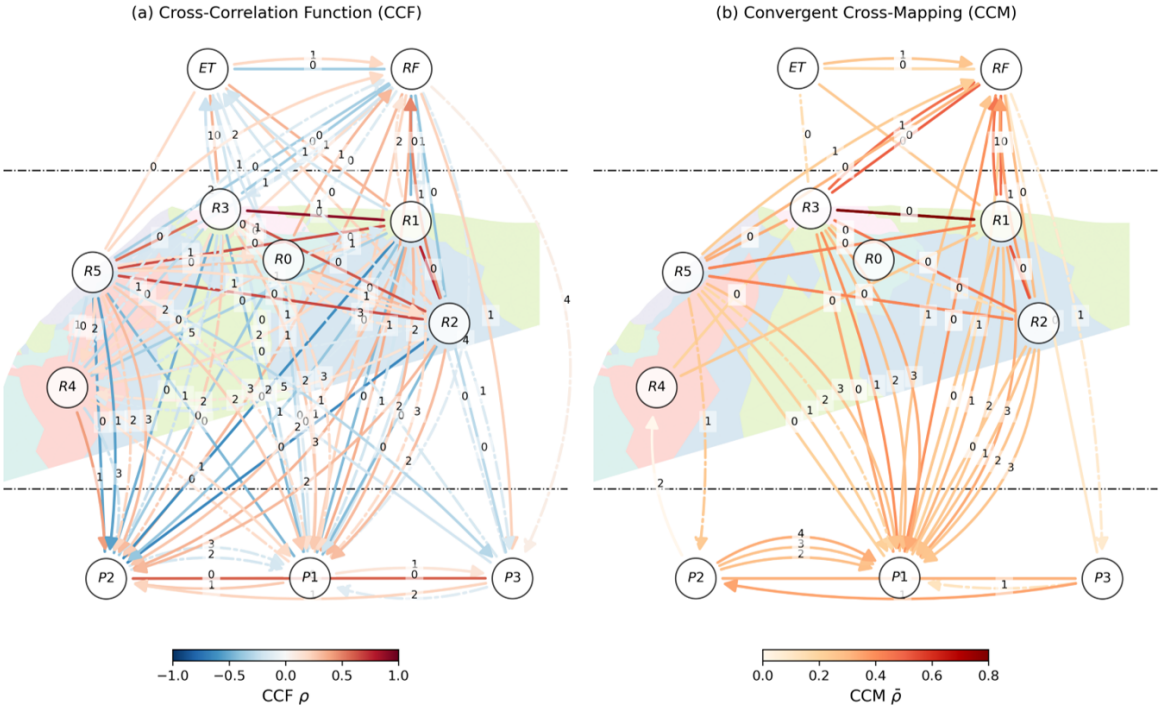

Figure 4. Graph of pairwise cross-dependencies: (a) with the linear Cross-Correlation Function (CCF),(b) with the nonlinear Convergent Cross-Mapping (CCM) method. An undirected line represents contemporaneous dependencies. Delayed dependencies are shown using directed curved arrows. All corresponding delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to the strength of dependencies. Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01.

Compared to CCF, the CCM results (Fig. 4.b) are more intelligible since fewer significant links are reported. However, CCM as a nonlinear method gives no clue about the nature of the underlying dynamics or the dominant sign of the dependencies. P2 is now exclusively related to R5, and P3 has no dependencies on resistivity patterns. P1 is CCM-related to the surface resistivity patterns R1, R3, and R5, but also R2, which is representative of the limestone matrix resistivity. Compared to CCF, CCM supports the particular conclusion of connected and preferential flows occurring between the surface and P1.

4.2 Multivariate methods

For multivariate methods, we chose to report causal graphs for the raw (not differenced) data since differencing did not impact ParCorr but reduced the precision of CMI in the virtual experiment (Table 3). Hence, Fig. 5 shows linear conditional dependencies (ParCorr) obtained from the raw time-series for the full dataset (All data, Fig. 5.a) and considering the discharge series one by one (P1, P2, and P3, Fig. 5, b to d). The P1, P2, and P3 datasets allow the analysis to be performed over larger time domains (Fig. 1.c). Except for R4, the dominant relationships between resistivity and meteorological variables are maintained between

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

the graphs, demonstrating stability in the ParCorr results despite differences in the considered time-domain. As for CCM (Fig. 4.b), P1 is associated with R1, R2, R3, and R5 (Fig. 5, a and b). The rainfall RF remains significantly related to P1, suggesting that resistivity patterns are not sufficient causes of P1. Regarding P2, similarly to CCM (Fig. 4.b), ParCorr on all data (Fig. 5.a) reveals a significant link between R5 and P2. However, the link is absent in Fig. 5.c, and two direct links from RF appear and bypass the resistivity patterns. Yet, R3 and R4 seem to influence P2 at lag 2, but the relationship is positive, which cannot be interpreted as a linear mass transfer. We also denote two upward links to R4 and ET. These links seem physically unrealistic and potentially problematic since the effect of P2 is be removed from these variables, which may alter the whole causal graph. P3, i.e., the low rate stalactite drip discharge, remains unrelated (Fig. 5, a and d).

Regarding the nonlinear analysis (CMI), we found unstable results on all datasets: All data, P1, P2, and P3 (see SM2.3). The causal graphs varied substantially when we repeated the analysis with the same parameters due to the stochastic nature of the independence test (Runge, 2018b). Consequently, we developed a sensitivity analysis by varying the hyperparameters of the method, hoping to isolate more stable configurations. Whatever the configurations, the results remained unstable. The instability is attributed to the lack of data and the fact that CMI, the most sensitive method, is applied to highly correlated data from a smooth inverted electrical resistivity model. This creates anomalies in the representation of causality between resistivity variables potentially due to the overly deterministic relationships that link these series. This problem is illustrated and further detailed in the supplementary materials (SM2.3). Consequently, to achieve a causal representation of the system with the nonlinear multivariate method, we had to adopt another logic, that of the consensus brought by the set of models from the sensitivity analysis. We considered all simulations done for the sensitive analysis as an ensemble of models rendering each a causal graph. Figure 6 reports the links achieving majority (50% considering a p-value of 0.05) among the ensemble of causal graphs from the sensitivity analysis. Following the consensus logic, the results again suggest a preferential connection with P1, while the two other drip discharge series remain unrelated.

5 Discussion

We introduced CMI as the most relevant for effective connection detection. Our results showed that CMI differs from the other CIMs in its low false-positive rate for the virtual case. However, CMI remains imperfect and missed a significant amount of effective connections. Moreover, the method proved to be unstable in our real case study, forcing us to adopt a logic initially absent from causal inference methods, that of a consensus graph (Fig. 6). The instability is possibly due to a large amount of missing data. Another hypothesis is that the ERT time-series are too smooth and deterministically related, which could contribute to the instability of the method (Runge, 2018a). The supplementary materials indeed show a strange behavior in the links with the CMI method, especially on the reduced time domains due to missing data (see SM2.3, Fig. SM.5 P2, and Fig. SM.6, All data).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

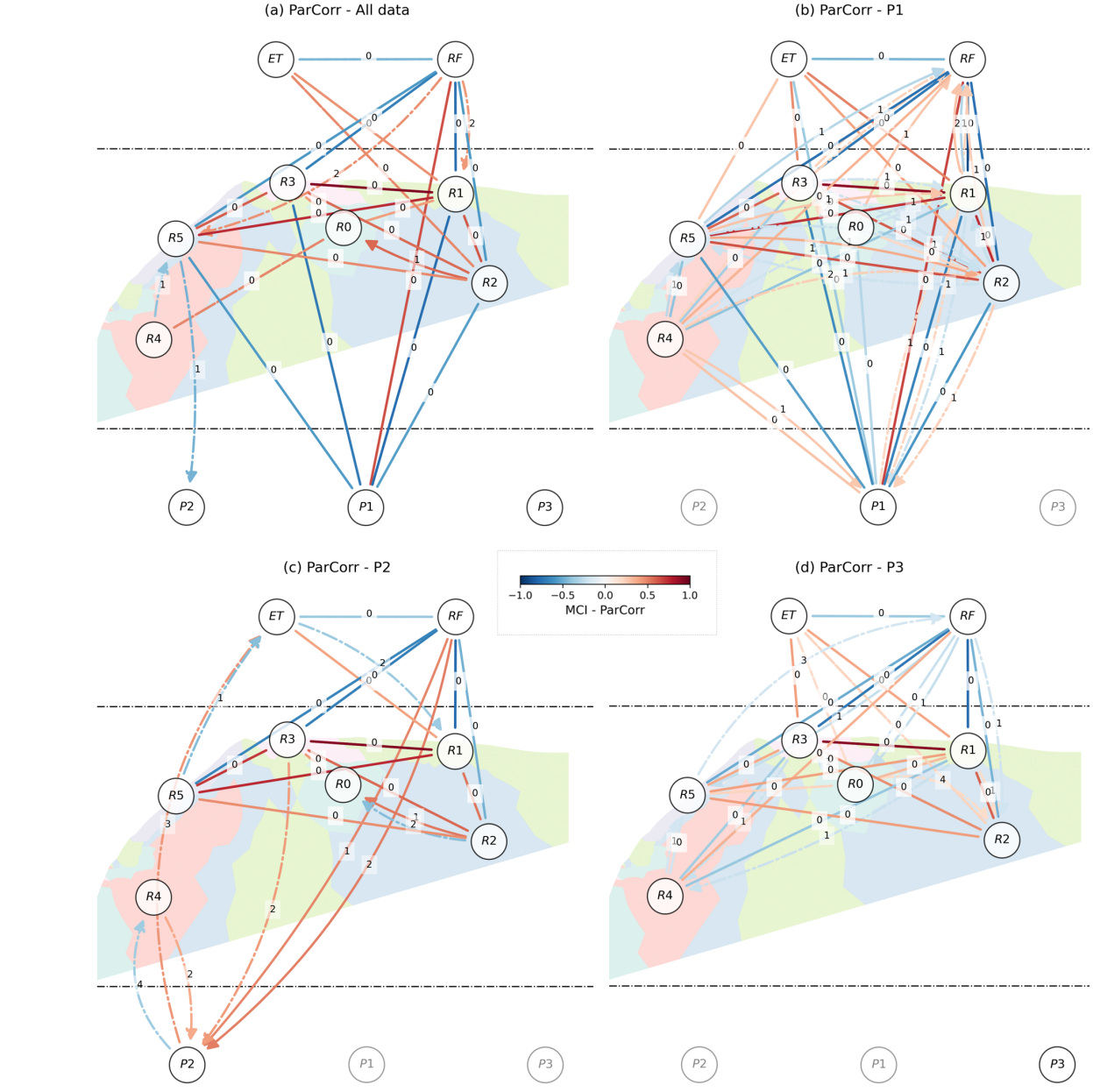

Figure 5. Graph of ParCorr cross-dependencies: considering (a) all data or one unique discharge series (b) P1, (c) P2, or (d) P3. An undirected line represents contemporaneous dependencies. Delayed dependencies are shown using directed curved arrows. All corresponding delays d are displayed in the middle of its corresponding arrow. The color of arrows maps to the strength of dependencies. Solid and dash-dotted arrows represent respectively significant dependencies with p-value < 0.001 and < 0.01. For each graph, the size of the overlapping time-domain between the variables changes as follows: 48 days (a), 184 days (b), 62 days (c), and 218 days (d).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

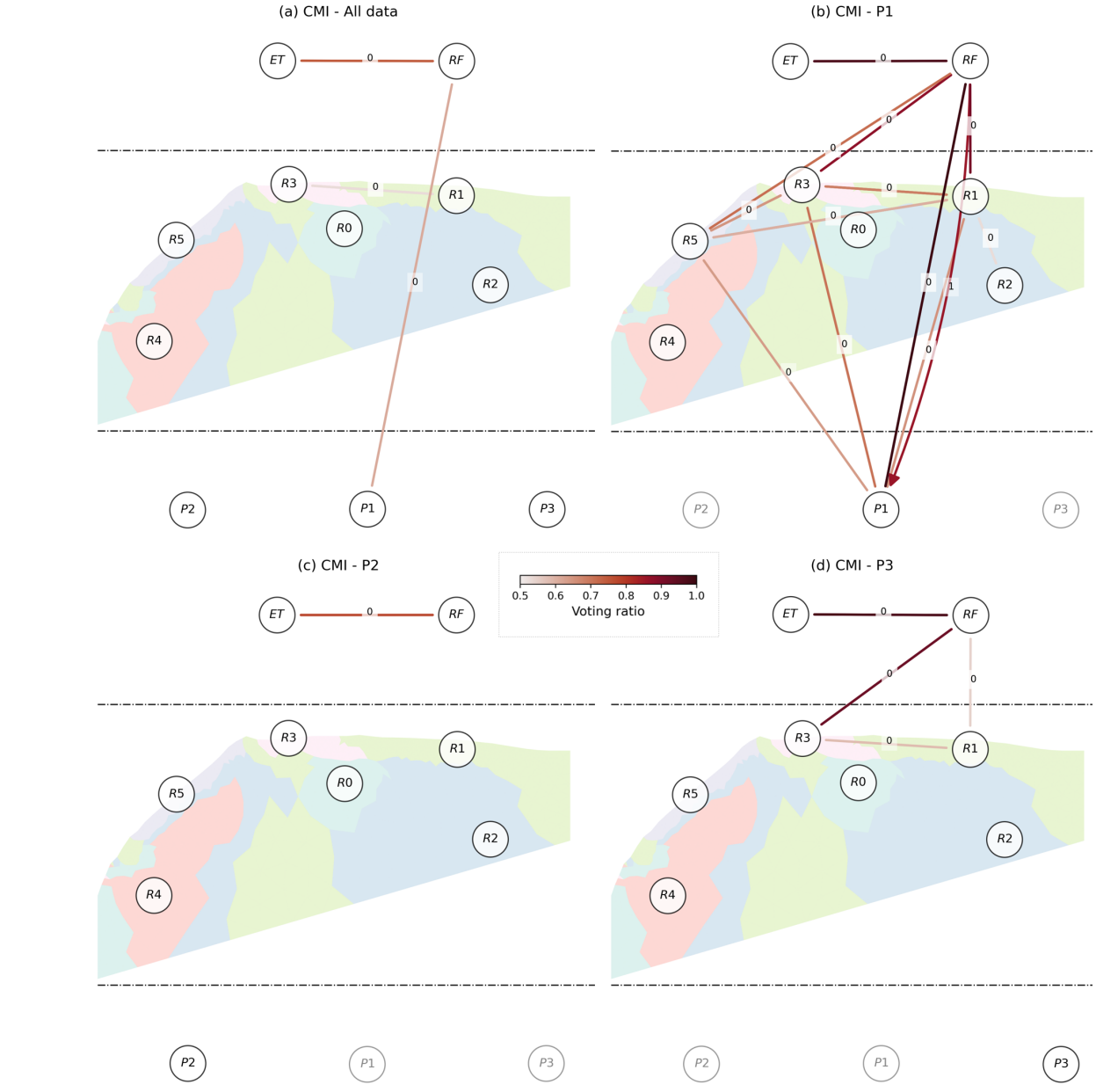

Figure 6. Consensual graph of CMI cross-dependencies obtained from the ensemble of simulations performed in the sensitivity analysis: (considering (a) all data or one unique discharge series (b) P1, (c) P2, or (d) P3. An undirected line represents the Contemporaneous dependencies. Delayed dependencies are shown using directed curved arrows. All corresponding delays d are displayed in the middle of its corresponding arrow. The links displayed are those reaching the majority (50%) for all causal graphs where the connections are established with a p-value of significance at 0.05. The color of arrows maps to the strength of the voting ratio. For each graph, the size of the overlapping time-domain between the variables changes as follows: 48 days (a), 184 days (b), 62 days (c), and 218 days (d).

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

5.1 On the Practical Use of CIMs

Optimistically, we note that CIMs tend to reveal the preferential connection between P1 and the surface, as expected from the dye tracing test (Poulain et al., 2018, and Fig. 1.a). Yet, no statistical test provides the actual probability of a causal relationship, but rather that of an association under the assumptions of the CIM's underlying model and the adequacy of the dataset, notably, i.e., through the causal sufficiency hypothesis (Runge, 2018a). In other words, claiming that a link inferred from CIMs is a causal or effective connection is a delicate step. On this front, the CIMs are all equivalent, and they only differ on the types of connections that they can or cannot detect with a fair rate of false alarm. We demonstrate that the CMI method has a low false-positive rate compared to the others. Still, bivariate methods are helpful if only to select potential links and check which ones are dismissed while using multivariate CIMs. The value of CCM compare to CCF is the opportunity to reveal weak nonlinear interactions for cases where such relations are suspected. Besides, linear methods are faster, more interpretable, and explainable to an audience, i.e., many reasons that could favor their practical use.

In general, the CIMs allow us to focus on a limited number of relationships: the strongest, the most cross-predictable, the most robust, or consensual if the approaches or parameters are varied. For this reason, the high number of connections found by the CCF method (Fig. 4.a) should act as a warning light. Nonetheless, and this is true for all methods, we can adjust the number of causal connections retrieved and limit the results to the strongest dependencies by selecting a proper method parameterization: lowering the p-values, using more restrictive tests, or confronting the statistical dependencies with those obtained from surrogate data sets (Schreiber and Schmitz, 2000). The predictive potential of the most significant links gives them value, and their robustness should prevent futile discussions on an evanescent singularity. Altogether, it appears difficult to rely on any particular link obtained with a single method and in a unique configuration. Indeed, this study showed how volatile links CIMs could be. Therefore, CIMs'output, like many others in statistics, should be taken with caution. Their robustness should be assessed by varying the methods, the datasets, or the time domains to address the robustness of the links. On the other hand, we cannot exclude that an elusive but actual causal connection would be only detectable by one or the other method, using the right parameterization.

By being labeled as causal, we may have more expectations of CIMs than other types of experiments or investigation methods, hoping to get simple answers to complex problems from them. Nevertheless, all types of methods can contribute to our causal understanding of environmental systems (e.g., dye tracing tests or spatially detailed inverse resistivity models). However, the sensitivity of CIMs to their assumptions and parameterization makes them hazardous to use alone. Despite their limitations, CIMs complement other methods, and they could be combined to narrow the range of possible causal representations of the system under study.

5.2 Research Perspectives

Today, CIMs benefit from a growing interest; they evolve and progress. CIMs pursue the ideal of causality solely inferred by data and machines. They are very generic methods and should be studied and appropriated for the specificity of hydrological data. In general, Klemeš (1982) was particularly critical of letting the data speak for itself. For the multivariate methods, we

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

have chosen to let the causal graphs be formed from the data. We have not prescribed any constraint on the conditioning of variables. This means that variables can be conditioned on potentially aberrant links, negatively impacting the whole causal graph. Studying how to constraint causal inference is an exciting research perspective. It should be done with caution to prevent us from constraining the analysis on perceptually biased hypotheses on the system's functioning. In particular, this could be done by reintroducing some physical concerns or spatial dimensions into the analysis. Rinderer et al. (2018) already proposed to constrain CIMs with structural connectivity, which is the first level of potential connectivity allowed by the geophysical environment. Currently, distances nor the length of flow path or energy potentials do not matter in CIMs. This spatial dimension is initially present in Hume's contiguity principle in time and space (Hume, 1748). To Schrodinger, the spatial continuum is also a causal paradigm in physics (Schrödinger, 1954). Then, we recommend research avenues on reconciling CIMs with space and physics.

Other research perspectives are more fundamentally related to the nature of the hydrological data.

1. Long time-scale. The methods proposed here work on the temporal domain. However, the assumption of causal sufficiency is questionable since hydrological connections can be spread over much longer time-scales. Thus, methods that work in the frequency domain or that couple the frequency domain with the time-domain deserve a particular interest (e.g., Molini et al., 2010).

2. Intermittent connectivity. We did not address the nonlinearity associated with threshold effects or intermittent hydrological connections (Blöschl and Zehe, 2005; Bracken et al., 2013). The CIMs assume constant interactions, even if modulable in the case of nonlinear methods. Since a hydrograph at the output of a hydrological system is a good representation of the hydrological state of the system, applying the CIMs on a segmentation of the hydrograph (e.g., high or low flows) is another interesting application. Segmentation would, however, reduce the available temporal domain for conditioning the variables, similarly to missing values in this study.

3. Contemporaneous dependencies. Most hydrological dependencies are contemporaneous, preventing the inference of causality from the principle of priority. Some CIMs are not built upon the principle of priority (e.g., Spirtes et al., 2000; Pearl, 2009; Runge et al., 2019a) and could be investigated. In particular, the Tigramite package has been recently updated to include an improved and faster PCMCI+ algorithm that deals with contemporaneous links and strong auto-correlation in series, with the promises of stronger recall and well-controlled false positive (Runge, 2020).

4. Statistical models and their parametrization. The CMI method deserves to be studied in more detail for its potentiality revealed in the virtual experiment. There are other estimators of the CMI that could be more adapted to longer series or other time-scales (Runge et al., 2019b). Furthermore, the parameterization of estimators could be studied in more detail, supported by the current interest in hydrology about the information theory (see Goodwell et al., 2020, and other debaters). Virtual experiments also provide an opportunity to study in more detail how to parameterize CIMs.

5. Inversion or processing artifacts. Many hydrological datasets are not directly measured but inverted, modeled, or processed so that artificial dependencies between the data may exist. These dependencies should not be confused with

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

causal relationships, and further study could be conducted to assess the sensitivity of CIMs to these artifacts and propose strategies to mitigate the problem.

6 Conclusions

We applied four causal inference methods for detecting hydrological connections. These CIMs study the temporal dependencies between variables. They are either linear or nonlinear, and bivariate (pairwise dependencies) or multivariate (conditional pairwise dependencies). The four CIMs are applied to both a synthetic and a real case in a karstic study site. The synthetic data are effective precipitation and modeled discharge from two reservoirs, either parallel and disconnected or contributing in series to the same drainage channel. The real karstic dataset involves precipitation and evapotranspiration data, subsurface resistivity series clustered from an inverted electrical resistivity model associated with a time-lapse ERT dataset, and three drip percolation discharge series in the cave below the surface. In this case, we know that a preferential flow exists between the surface and the one spot of drip discharge in the cave.

For the synthetic case, bivariate methods cannot discriminate causal dependencies from those arising from confounding by the meteorological forcing of the model. Bivariate methods generally only account for potential connections, while multivariate methods attempt to extricate actual ones. In accordance with our theoretical expectations, we found that the nonlinear and multivariate CIM based on conditional mutual information (CMI) is indeed the most precise: the identified connections are likely to be actual connections. Yet, this method has a bad recall: it misses many connections. Moreover, this same method proved to be unstable for the real case. We believe that the temporal domain, further reduced by missing data and the conditioning process, was too small to converge. Possibly, the instability is also due to the electrical resistivity model and the clustered series being too smooth, correlated, and therefore too deterministically related to establishing causality based on the concept of conditional independence. To overcome this issue, we built a consensual causal graph from multiple iterations of the method and its parameters, with the outcome in phase with our perceptual understanding of the system. Yet, consensus introduces a new logic in the causal inference process that is not part of the initial theory. The causal graph of the linear bivariate method, the cross-correlation function (CCF), shows ubiquitous dependencies between variables. Not much can be learned about connectivity and preferential flow paths without considering more restrictive tests. The three other methods support the idea of a preferential connection where it is suspected.

Nonetheless, any multivariate CIM will assess causality under the hypothesis of causal sufficiency, meaning that all the potential causes are monitored alongside other hypotheses. The sufficiency of the data is hardly verified in practice, as well as the complete adequacy of one particular CIM and its parameters. As a result, different causal links may appear with different configurations. Accordingly, it is delicate to interpret a causality test as the probability of an actual hydrological connection. While caution and flexibility are always warranted, CIMs are interesting data mining methods, undoubtedly capable of making causal discoveries and constantly improving. We propose suggestions for research perspectives considering the specificity of hydrological dynamics. In the meantime, we promote non-exclusive approaches while using CIMs by varying the methods, their parameters, the time domain, and the data. They participate in a causal understanding of hydrological systems, but causal

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021


Hydrology and Earth System Sciences
Discussions

understanding is not limited to them. They synergize with other leading approaches such as physically-based modeling, other empirical data analysis approaches, or field investigations.

Code and data availability. CCF and the Student's t-test are computed using the Scipy Python package (Virtanen et al., 2020). The CCM python implementation is available from Delforge et al. (2020a). The official R implementation is available from the CRAN repository: https://CRAN.R-project.org/package=rEDM. PCMCI and independence tests are implemented within the Tigramite (v.4.1 in this case) Python package: https://jakobrunge.github.io/tigramite/. Evapotranspiration data were obtained from the agrometeorological PAMESEB network for the station of Jemelle: https://agromet.be. All other environmental time-series can be obtained from Watlet et al. (2018) and the related repository: https://zenodo.org/record/1158631. Resistivity clustered time-series can be reconstructed following Delforge et al. (2020b) and the example available from the repository: http://dx.doi.org/10.17632/zh5b88vn78.2

.

Author contributions. Conceptualization, D. Delforge; methodology, D. Delforge and O. de Viron.; formal analysis, D. Delforge.; investigation, A. Watlet.; data curation, A. Watlet; writing—original draft preparation, D. Delforge.; writing—review and editing, O. de Viron, M. Vanclooster, M. Van Camp, A. Watlet; supervision, M. Vanclooster, M. Van Camp. All authors have read and agreed to the published version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is part of a Ph.D. supported by a FRIA grant from the Belgian Fund for Scientific Research (FSR-FNRS). The publication in an open access journal has been supported by the sector of science and technology of UCLouvain. A. Watlet publishes with the permission of the Executive Director, British Geological Survey (UKRI-NERC). We are thankful to M. Van Ruymbeke who designed and constructed drip discharge sensors utilized in this study and to O. Kaufmann who also constructed drip discharge sensors, maintained the underground monitoring system and conceptualized the time-lapse ERT experiment.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

References

Akaike, H.: A new look at the statistical model identification, IEEE Transactions on Automatic Control, 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements, no. 56 in FAO irrigation and drainage paper, Food and Agriculture Organization of the United Nations, Rome, 1998.

Angelini, P.: Correlation and spectral analysis of two hydrogeological systems in Central Italy, Hydrological Sciences Journal, 42, 425–438, https://doi.org/10.1080/02626669709492038, 1997.

Bailly-Comte, V., Jourde, H., Roesch, A., Pistre, S., and Batiot-Guilhe, C.: Time series analyses for Karst/River interactions assessment: Case of the Coulazou river (southern France), Journal of Hydrology, 349, 98–114, https://doi.org/10.1016/j.jhydrol.2007.10.028, 2008.

Bakalowicz, M.: Karst groundwater: a challenge for new resources, Hydrogeology Journal, 13, 148–160, https://doi.org/10.1007/s10040-004-0402-9, 2005.

Blöschl, G. and Zehe, E.: On hydrological predictability, Hydrological Processes, 19, 3923–3929, https://doi.org/10.1002/hyp.6075, 2005.

Bracken, L. J., Wainwright, J., Ali, G. A., Tetzlaff, D., Smith, M. W., Reaney, S. M., and Roy, A. G.: Concepts of hydrological connectivity: Research approaches, pathways and future agendas, Earth-Science Reviews, 119, 17–34, https://doi.org/10.1016/j.earscirev.2013.02.001, 2013.

Delforge, D., Muñoz-Carpena, R., Van Camp, M., and Vanclooster, M.: A Parsimonious Empirical Approach to Streamflow Recession Analysis and Forecasting, Water Resources Research, 56, e2019WR025 771, https://doi.org/10.1029/2019WR025771, 2020a.

Delforge, D., Watlet, A., Kaufmann, O., Van Camp, M., and Vanclooster, M.: Time-series clustering approaches for subsurface zonation and hydrofacies detection using a real time-lapse electrical resistivity dataset, Journal of Applied Geophysics, p. 104203, https://doi.org/10.1016/j.jappgeo.2020.104203, 2020b.

Dooge, J.: Linear Theory of Hydrologic Systems, Agricultural Research Service, U.S. Department of Agriculture, 1973.

Frenzel, S. and Pompe, B.: Partial mutual information for coupling analysis of multivariate time series, Physical Review Letters, 99, 204 101, https://doi.org/10.1103/PhysRevLett.99.204101, 2007.

Friston, K. J.: Functional and Effective Connectivity: A Review, Brain Connectivity, 1, 13–36, https://doi.org/10.1089/brain.2011.0008, 2011.

Goodwell, A. E., Jiang, P., Ruddell, B. L., and Kumar, P.: Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback, Water Resources Research, 56, https://doi.org/10.1029/2019WR024940, 2020.

Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica, 37, 424–438, https://doi.org/10.2307/1912791, 1969.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M.: Karst water resources in a changing world: Review of hydrological modeling approaches, Reviews of Geophysics, 52, 2013RG000 443, https://doi.org/10.1002/2013RG000443, 2014.

Hume, D.: Philosophical Essays Concerning Human Understanding, A. Millar, 1748.

Jourde, H., Mazzilli, N., Lecoq, N., Arfib, B., and Bertin, D.: KARSTMOD: A Generic Modular Reservoir Model Dedicated to Spring Discharge Modeling and Hydrodynamic Analysis in Karst, in: Hydrogeological and Environmental Investigations in Karst Systems, edited by Andreo, B., Carrasco, F., Durán, J. J., Jiménez, P., and LaMoreaux, J. W., no. 1 in Environmental Earth Sciences, pp. 339–344, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-17435-3_38, 2015.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Kadić, A., Denić-Jukić, V., and Jukić, D.: Revealing hydrological relations of adjacent karst springs by partial correlation analysis, Hydrology Research, 49, 616–633, https://doi.org/10.2166/nh.2017.064, 2018.

Klemeš, V.: Empirical and causal models in hydrology, in: Scientific Basis of Water-Resource Management, Washinton D.C., http://www.itia.ntua.gr/en/docinfo/1075/, 1982.

Labat, D., Ababou, R., and Mangin, A.: Rainfall–runoff relations for karstic springs. Part I: convolution and spectral analyses, Journal of Hydrology, 238, 123–148, https://doi.org/10.1016/S0022-1694(00)00321-8, 2000.

Larocque, M., Mangin, A., Razack, M., and Banton, O.: Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France), Journal of Hydrology, 205, 217–231, https://doi.org/10.1016/S0022-1694(97)00155-8, 1998.

Mathevet, T., Lepiller, M. l., and Mangin, A.: Application of time-series analyses to the hydrological functioning of an Alpine karstic system: the case of Bange-L'Eau-Morte, Hydrology and Earth System Sciences, 8, 1051–1064, https://doi.org/https://doi.org/10.5194/hess-8-1051-2004, 2004.

Meyfroidt, P.: Approaches and terminology for causal analysis in land systems science, Journal of Land Use Science, 11, 501–522, https://doi.org/10.1080/1747423X.2015.1117530, 2016.

Molini, A., Katul, G. G., and Porporato, A.: Causality across rainfall time scales revealed by continuous wavelet transforms, Journal of Geophysical Research, 115, D14 123, https://doi.org/10.1029/2009JD013016, 2010.

Pearl, J.: Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, 2 edn., https://doi.org/10.1017/CBO9780511803161, 2009.

Poulain, A., Watlet, A., Kaufmann, O., Van Camp, M., Jourde, H., Mazzilli, N., Rochez, G., Deleu, R., Quinif, Y., and Hallet, V.: Assessment of groundwater recharge processes through karst vadose zone by cave percolation monitoring, Hydrological Processes, 32, 2069–2083, https://doi.org/10.1002/hyp.13138, 2018.

Reichenbach, H.: The Direction of Time, California library reprint series, University of California Press, https://books.google.be/books?id=f6kNAQAAIAAJ, 1956.

Rinderer, M., Ali, G., and Larsen, L. G.: Assessing structural, functional and effective hydrologic connectivity with brain neuroscience methods: State-of-the-art and research directions, Earth-Science Reviews, 178, 29–47, https://doi.org/10.1016/j.earscirev.2018.01.009, 2018.

Rodriguez-Iturbe, I., Entekhabi, D., and Bras, R. L.: Nonlinear Dynamics of Soil Moisture at Climate Scales: 1. Stochastic Analysis, Water Resources Research, 27, 1899–1906, https://doi.org/10.1029/91WR01035, 1991.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Ruddell, B. L. and Kumar, P.: Ecohydrologic process networks: 1. Identification, Water Resources Research, 45, https://doi.org/10.1029/2008WR007279, 2009.

Runge, J.: Causal network reconstruction from time series: From theoretical assumptions to practical estimation, Chaos: An Interdisciplinary Journal of Nonlinear Science, 28, 075 310, https://doi.org/10.1063/1.5025050, 2018a.

Runge, J.: Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information, in: International Conference on Artificial Intelligence and Statistics, pp. 938–947, http://proceedings.mlr.press/v84/runge18a.html, 2018b.

Runge, J.: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, in: Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), edited by Peters, J. and Sontag, D., vol. 124 of *Proceedings of Machine Learning Research*, pp. 1388–1397, PMLR, https://proceedings.mlr.press/v124/runge20a.html, 2020.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., Nes, E. H. v., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J.: Inferring causation from time series in Earth system sciences, Nature Communications, 10, 2553, https://doi.org/10.1038/s41467-019-10105-3, 2019a.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets, Science Advances, 5, eaau4996, https://doi.org/10.1126/sciadv.aau4996, 2019b.

Salvucci, G. D., Saleem, J. A., and Kaufmann, R.: Investigating soil moisture feedbacks on precipitation with tests of Granger causality, Advances in Water Resources, 25, 1305–1312, https://doi.org/10.1016/S0309-1708(02)00057-X, 2002.

Schreiber, T.: Measuring Information Transfer, Physical Review Letters, 85, 461–464, https://doi.org/10.1103/PhysRevLett.85.461, 2000.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D: Nonlinear Phenomena, 142, 346–382, https://doi.org/10.1016/S0167-2789(00)00043-9, 2000.

Schrödinger, E.: Nature and the Greeks, Shearman lectures, 1948, University Press, https://books.google.be/books?id=H7sAAAAAMAAJ, 1954.

Sendrowski, A. and Passalacqua, P.: Process connectivity in a naturally prograding river delta, Water Resources Research, 53, 1841–1863, https://doi.org/10.1002/2016WR019768, 2017.

Slater, L. and Binley, A.: Advancing hydrological process understanding from long-term resistivity monitoring systems, WIREs Water, 8, https://doi.org/10.1002/wat2.1513, 2021.

Spirtes, P. and Glymour, C.: An Algorithm for Fast Recovery of Sparse Causal Graphs, Social Science Computer Review, 9, 62–72, https://doi.org/10.1177/089443939100900106, 1991.

Spirtes, P., Glymour, C. N., and Scheines, R.: Causation, prediction, and search, Adaptive computation and machine learning, MIT Press, Cambridge, Mass, 2nd ed edn., 2000.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344, 734–741, https://doi.org/10.1038/344734a0, 1990.

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

Takens, F.: Detecting strange attractors in turbulence, Lecture Notes in Mathematics, Berlin Springer Verlag, 898, 366, https://doi.org/10.1007/BFb0091924, 1981.

Theiler, J.: Spurious dimension from correlation algorithms applied to limited time-series data, Physical Review. A, General Physics, 34, 2427–2432, 1986.

Tuttle, S. E. and Salvucci, G. D.: Confounding factors in determining causal soil moisture-precipitation feedback, Water Resources Research, 53, 5531–5544, https://doi.org/10.1002/2016WR019869, 2017.

Vejmelka, M. and Paluš, M.: Inferring the directionality of coupling with conditional mutual information, Physical Review E, 77, 026 214, https://doi.org/10.1103/PhysRevE.77.026214, 2008.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

https://doi.org/10.5194/hess-2021-445
Preprint. Discussion started: 6 September 2021

Watlet, A., Kaufmann, O., Triantafyllou, A., Poulain, A., Chambers, J. E., Meldrum, P. I., Wilkinson, P. B., Hallet, V., Quinif, Y., Ruymbeke, M. V., and Camp, M. V.: Imaging groundwater infiltration dynamics in the karst vadose zone with long-term ERT monitoring, Hydrology and Earth System Sciences, 22, 1563–1592, https://doi.org/https://doi.org/10.5194/hess-22-1563-2018, 2018.

555 Ye, H., Deyle, E. R., Gilarranz, L. J., and Sugihara, G.: Distinguishing time-delayed causal interactions using convergent cross mapping, Scientific Reports, 5, 14 750, https://doi.org/10.1038/srep14750, 2015.