# Revisiting causality using stochastics: 2. Applications

Demetris Koutsoyiannis[1], Christian Onof[2], Antonis Christofidis[1] and
Zbigniew W. Kundzewicz[3]

[1] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National
Technical University of Athens (dk@itia.ntua.gr)

[2] Department of Civil and Environmental Engineering, Faculty of Engineering, Imperial College London

[3] Meteorology Lab, Department of Construction and Geoengineering, Faculty of Environmental
Engineering and Mechanical Engineering, Poznan University of Life Sciences, Poznań, Poland

**Abstract** In a companion paper, we develop the theoretical background of a stochastic
approach to causality with the objective of formulating necessary conditions that are
operationally useful in identifying or falsifying causality claims. Starting from the idea of
stochastic causal systems, the approach extends it to the more general concept of hen-or-
egg causality, which includes as special cases the classic causal, and the potentially causal
and anticausal systems. The framework developed is applicable to large-scale open
systems, which are neither controllable nor repeatable. In this paper we illustrate and
showcase the proposed framework in a number of case studies. Some of them are
controlled synthetic examples and are conducted as a proof of applicability of the
theoretical concept, to test the methodology with a priori known system properties. Other
are real-world studies on interesting scientific problems in geophysics, and in particular
hydrology and climatology.

疑惑深 智慧深 疑惑浅 智慧浅
(*Deep doubts, deep wisdom; shallow doubts, shallow wisdom* – Chinese proverb)

## 1 Introduction

The companion paper (Koutsoyiannis et al., 2022) studies theoretically the identification
and characterization of a causal link between two stochastic processes $\underline{x}(t)$ and $\underline{y}(t)$,
which represent two natural processes. Our proposal relies on the relationship:

$$\underline{y}(t) = \int_{-\infty}^{\infty} g(h)\underline{x}(t - h)\mathrm{d}h + \underline{v}(t) \tag{1}$$

where $\underline{v}(t)$ is another stochastic process (not necessarily white noise), assumed
uncorrelated to $\underline{x}(t)$, which represents the part of the process that is not explained by the

causal link. The function $g(h)$ is the impulse response function (IRF) of the system $(\underline{x}(t), \underline{y}(t))$; to see this, we set $\underline{v}(t) \equiv 0$ and $\underline{x}(t) = \delta(t)$ (the Dirac delta function, representing an impulse of infinite amplitude at $t = 0$ and attaining the value of 0 for $t \neq 0$), and we readily get $\underline{y}(t) = g(t)$.

For any two processes $\underline{x}(t)$ and $\underline{y}(t)$, equation (1) has infinitely many solutions in terms of the function $g(h)$ and the process $\underline{v}(t)$. An obvious and trivial one is $g(h) \equiv 0, \underline{y}(t) \equiv \underline{v}(t)$. The sought solution is the one that corresponds to the minimum variance of $\underline{v}(t)$, called the least-squares solution. Assuming that this has been determined for the system $(\underline{x}(t), \underline{y}(t))$, we call that system:

1. *potentially causal* if $g(h) = 0$ for any $h < 0$, while the explained variance is non negligible;
2. *potentially anticausal* if $g(h) = 0$ for any $h > 0$, while the explained variance is non negligible (this means that the system $(\underline{y}(t), \underline{x}(t))$ is potentially causal);
3. *potentially hen-or-egg* (*HOE*) *causal* if $g(h) \neq 0$ for some $h > 0$ and some $h < 0$, while the explained variance is non negligible;
4. *noncausal* if the explained variance is negligible.

In the above we use the adverb "potentially" to highlight the fact that the conditions tested provide necessary but not sufficient conditions for causality. In a potentially causal (or anticausal) system the time order is explicitly reflected in the definition. In a potentially HOE causal system the time order needs to be clarified by defining the principal direction. The most natural indices for this are: (a) the time lag $h_\mathrm{c}$ maximizing the absolute value of cross-covariance; (b) the mean (time average) of the function $g(h)$; and (c) the median $h_{1/2}$ of the function $g(h)$.

Assuming that processes $\underline{x}(t)$ and $\underline{y}(t)$ are positively correlated (i.e. an increase in $\underline{x}(t)$ would result in an increase in $\underline{y}(t)$; if not we multiply one of the two by $-1$), we seek an optimal solution for the IRF by minimising the variance of $\underline{v}(t)$. We also set additional desiderata for

(a)   an adequate time span $\mathbb{h}$ of $h$;
(b)   a nonnegative $g(h) \geq 0$ for all $h \in \mathbb{h}$; and
(c)   a smooth $g(h)$ with the smoothness expressed as a constraint $E \leq E_0$, where $E$ is determined in terms of the second derivative of $g(h)$ and $E_0$ is a real number.

The proposed theoretical framework is formulated in terms of natural, i.e., continuous time. On the other hand, as the estimation of the IRF relies on data in an inductive manner, and data are only available in discrete time, conversion of the continuous- to a discrete-time framework is necessary for the application. This results in

$$\underline{y}_\tau = \sum_{j=-\infty}^{\infty} g_j \underline{x}_{\tau-j} + \underline{v}_\tau \tag{2}$$

where the sequence $g_j$ is determined from the function $g(h)$ in the companion paper. Furthermore, any data set is finite and allows only a finite number of $g_j$ terms to be estimated. Therefore, in the applications the summation limits $\pm\infty$ in equation (2) are replaced by $\pm J$, apparently assuming that $g_j = 0$ for $|j| > J$, where, in order to identify $g_j$ from data, $J$ should be chosen much lower than the length of the dataset.

If we exclude the nonnegativity constraint, then the problem of identifying the IRF has an analytical solution (equation (37) in the companion paper). However, a numerical solution is always possible, simple and fast. The theoretical framework is illustrated, tested and explored in a number of case studies which are presented below. The majority of these are synthetic, with a priori known system properties, and they serve to test the methodology developed. The remaining are real-world case studies that serve to illustrate the usefulness of the method.

Apparently, the estimation of the IRF from data involves uncertainty. As explained in the companion paper (Koutsoyiannis et al., 2022), the method of choice for the uncertainty assessment is Monte Carlo simulation, because the complexity of the calculations for optimizing the IRF fitting do not allow analytical solutions. To make the study short, this task was kept out of its scope. However, a preliminary investigation and some first results are provided in the Supplementary Information, section SI2.1, where it is shown that (i) the uncertainty in the IRF ordinates per se can be large if no constraints are used to determine it, (ii) the use of constraints decreases the uncertainty, and (iii) the uncertainty in the key characteristics related to causality (time directionality, time lags, explained variance) is small, irrespective of the constraints used.

Increased estimation uncertainty may also result in spurious causality claims. As high autocorrelation increases uncertainty in the long term, this could be a major case leading to false identification of causality. This is illustrated in the Supplementary Information (Section SI2.2) by means of a synthetic example, in which the processes $\underline{x}_\tau$ and $\underline{y}_\tau$ are, by construction, independent of one another, but with high autocorrelation. Techniques to handle such situations and avoid false conclusions are also discussed there.

As also explained in the companion paper (Koutsoyiannis et al., 2022), the proposed method for the determination of the ordinates $g_j$ based on the minimization of the variance of the error process, is, by its construction, nonparametric. A well-known weakness of determining numerous ordinates is that it is an over-parameterized problem. Alternative techniques may overcome this problem using a parametric model (such as a Box-Jenkins model or an autoregressive moving average exogenous—ARMAX—model; Young, 2011, 2015). In our nonparametric approach the over-parameterization problem

can also be tackled—and here lies the usefulness of imposing constraints (a)-(c) discussed above. For comparison, an additional parametric method, formulated in terms of parameterizing the IRF per se in continuous time is also discussed and compared to the proposed non-parametric method in the Supplementary Information, section SI2.3. This method is also applied in one of the case studies in the Supplementary Information, section SI2.4.

## 2  Case studies

Thirty case studies have been conducted, whose results are summarized in Table 1. In all of them we started by assuming a potentially HOE causal model with a rather small number of weights $g_j$, namely 41 (i.e. $J = 20$). Depending on the results of the estimation procedure, the system is deemed potentially HOE causal if we have $g_j > 0$ for both some positive and some negative lags $j$, and potentially causal if $g_j = 0$ for all $j < 0$ (or anticausal if this happens for all $j > 0$). If the explained variance ratio is close to 0, the system would be deemed noncausal, but this case did not appear in the case studies. Furthermore, if it happens that at the edge of the window the IRF vanishes off ($g_{\pm J} = 0$), we have a strong indication that the chosen $J = 20$ (defining the window size) is sufficient to recover the system dynamics in terms of causality. The opposite case would mean that a larger time window with a greater $J$ is required. The optimal $J$ in this case is not investigated as our scope here is not to construct an optimal model for a specific system, but only to test the proposed methodology and seek some insights within it.

The majority of the case studies are synthetic (#1 to #18) and are conducted as a proof of concept, i.e., to test the methodology developed with a priori known system properties. The remaining are real-world case studies divided in two subcategories. Namely, studies #19 to #22 deal with the precipitation – runoff system which is well understood in hydrology. Here we try to investigate whether our framework is consistent with the known fact that precipitation at a specific location is the cause and runoff at the associated location the effect. In other words, here we know a priori that precipitation causes runoff and we test our methodology (whether it can capture this known fact) rather than try to find the actual causality direction. On the contrary, studies #23 – #30 deal with systems that are much more complex and not well understood. Namely, in studies #23 – #28 we investigate the links between atmospheric temperature and $CO_2$ concentration (cf. Koutsoyiannis and Kundzewicz, 2020) and in studies #29 – #30 we investigate the links between atmospheric temperature and El-Niño Southern Oscillation (ENSO; cf. Kundzewicz et al., 2020).

**Table 1** Summary indices for the results of all case studies elaborated. The $h_c$ is the time lag maximizing the cross-covariance $c_{yx}(h)$, or equivalently the cross-correlation $r_{yx}(h) := c_{yx}(h)/\sqrt{c_{xx}(0)c_{yy}(0)}$; $\mu_h$ is the mean (time average) of the function $g(h)$; $h_{1/2}$ is the median of the function $g(h)$; $e$ is the explained variance ratio; and $\varepsilon$ is the roughness ratio. In parentheses are the true values for the cases that they are known.

| Case system | Direction | $h_c$ | $\mu_h$ | $h_{1/2}$ | $r_{yx}(h_c)$ | $e$ | $\varepsilon$ | # |
|---|---|---|---|---|---|---|---|---|
| ***Synthetic cases*** | | | | | | | | |
| Pure HOE causal (symmetric), 41 weights, no constraints | $x \to y$ | 0 (0) | 0 (0) | 0 (0) | 0.954 | 1 | 3.7×10⁻⁵ * | 1 |
| | $y \to x$ | 0 (0) | 0 (0) | – § | 0.954 | 0.99 | 1.22* | 2 |
| As #2 but with 21 weights | $y \to x$ | 0 (0) | 0.02 (0) | – § | 0.954 | 0.99 | 3.2×10⁻⁵ * | 3 |
| As #3 with roughness constr. | $y \to x$ | 0 (0) | 0.11 (0) | – § | 0.954 | 0.98 | 2.4×10⁻⁵ * | 4 |
| As #2 but with nonnegativity constraint (or both) | $y \to x$ | 0 (0) | 0 (0) | 0.2 (0) | 0.954 | 0.91 | 0* | 5 |
| Causal, 21 weights, no constraints | $x \to y$ | 6 (6) | 6.86 (6.84) | – § (5.42) | 0.947 | 0.94 | 1.32 | 6 |
| | $y \to x$ | –6 (–6) | –6.50 | – § | 0.947 | 0.97 | 1.4×10⁻³ | 7 |
| Causal, 21 weights, no roughness constraint | $x \to y$ | 6 (6) | 6.79 (6.84) | 5.49 (5.42) | 0.947 | 0.94 | 0.633 | 8 |
| | $y \to x$ | –6 (–6) | –5.73 | –5.62 | 0.947 | 0.94 | 0.0053 | 9 |
| Causal, 21 weights, both constraints | $x \to y$ | 6 (6) | 6.86 (6.84) | 5.34 (5.42) | 0.947 | 0.94 | 1.9×10⁻⁵ * | 10 |
| | $y \to x$ | –6 (–6) | –5.15 | –4.49 | 0.947 | 0.93 | 3.2×10⁻⁵ * | 11 |
| Synthetic, causal, 21 weights exponentiated, both constraints | $x \to y$ | 6 | 6.64 | 5.89 | 0.554 | 0.32 | 1.2×10⁻⁶ * | 12 |
| | $y \to x$ | –6 | –3.34 | –4.62 | 0.554 | 0.43 | 9.8×10⁻⁷ | 13 |
| Causal, 1025 weights, both constraints | $x \to y$ | 8 (8) | 9.65 (298.3) | 8.87 (184.9) | 0.704 | 0.57 | 4.6×10⁻⁵ * | 14 |
| | $y \to x$ | –8 (–8) | –8.38 | –8.31 | 0.704 | 0.50 | 1.4×10⁻³ * | 15 |
| HOE causal (symmetric), 2049 weights, both constraints | $x \to y$ | 0 (0) | 0.06 (0) | 0.06 (0) | 0.758 | 0.71 | 1.7×10⁻⁵ * | 16 |
| | $x \to y$ | 0 (0) | 0.06(0) | 0.08 (0) | 0.758 | 0.71 | 5.5×10⁻⁶ | 17 |
| | $y \to x$ | 0 (0) | –0.28 | –0.26 | 0.758 | 0.57 | 4.7×10⁻⁴ | 18 |
| ***Geophysical cases*** | | | | | | | | |
| Precipitation – runoff (time step: 3 h) | $P \to R$ | 6 | 10.24 | 9.83 | 0.198 | 0.17 | 1.4×10⁻⁴ * | 19 |
| | $R \to P$ | –6 | –6.14 | –6.04 | 0.198 | 0.04 | 0.768 * | 20 |
| Precipitation – runoff transformed by eqn. (9) (time step: 3 h) | $P' \to R'$ | 8 | 10.84 | 10.72 | 0.173 | 0.68 | 1.5×10⁻⁴ * | 21 |
| | $R' \to P'$ | –8 | –8.24 | –7.89 | 0.173 | 0.03 | 0.293 * | 22 |
| Modern temperature and CO$_2$ data (time step: month) | $\Delta T \to \Delta \ln[CO_2]$ | 5 | 7.70 | 6.35 | 0.480 | 0.31 | 1.3×10⁻⁵ * | 23 |
| | $\Delta \ln[CO_2] \to \Delta T$ | –5 | –5.67 | –5.49 | 0.480 | 0.23 | 7.3×10⁻⁴ * | 24 |
| Paleoclimatic temperature and CO$_2$ data (time step: millennium) | $\Delta T \to \Delta \ln[CO_2]$ | 0 | 0.79 | 1.11 | 0.404 | 0.17 | 9.0×10⁻⁴ | 25 |
| | $\Delta \ln[CO_2] \to \Delta T$ | 0 | –0.56 | –0.82 | 0.404 | 0.17 | 1.6×10⁻³ | 26 |
| As #25 - #26 but cumulative (time step: millennium) | $T \to \ln[CO_2]$ | 1 | 1.74 | 1.87 | 0.875 | 0.86 | 7.9×10⁻⁵ | 27 |
| | $\ln[CO_2] \to T$ | –1 | –1.10 | –1.03 | 0.875 | 0.77 | 0.018 | 28 |
| Modern temperature and SOI data (time step: month) | $\Delta SOI \to \Delta T$ | 7 | 6.31 | 5.84 | 0.525 | 0.39 | 4.2×10⁻⁵ * | 29 |
| | $\Delta T \to \Delta SOI$ | –7 | –5.03 | –5.64 | 0.525 | 0.30 | 5.4×10⁻⁵ * | 30 |

* The roughness was calculated without considering the second derivative at zero.

§ We have not defined the median in cases that include negative values of $g_j$.

Notice that for each case, each of the directions $x \rightarrow y$ and $y \rightarrow x$ are investigated separately as there is no symmetry (or antisymmetry) in the produced IRFs in the two directions and, hence in the quantified measures, which are summarized in Table 1. When we refer to direction $y \rightarrow x$ we mean that we interchange the time series $x$ and $y$ and still estimate the IRF in the same way, as described in our equations (e.g. equation (1)), in which the direction $x \rightarrow y$ is assumed.

The details of the case studies are given in the following subsections.

## 2.1   Synthetic examples

All synthetic examples are based on the same input series $x_\tau$, which was constructed by the methodology in Koutsoyiannis (2020a) whose software is available online as Supplementary Information of that paper.

In particular, to generate $\underline{x}_\tau$ we use the moving average scheme:

$$\underline{x}_\tau = \sum_{i=-I}^{I} a_i \underline{w}_{\tau-i} \tag{3}$$

where $I = 1024$, $\underline{w}_\tau$ is white noise, assumed standard Gaussian, and the sequence of coefficients $a_i$ is assumed time symmetric, i.e., $a_{-i} = a_i$, and is determined assuming a Filtered Hurst-Kolmogorov process with a generalized Cauchy-type climacogram (FHK-C; Koutsoyiannis 2016, 2017):

$$\gamma(k) = \lambda^2 \left( 1 + \left( \frac{k}{\alpha} \right)^{2M} \right)^{\frac{H-1}{M}} \tag{4}$$

The term climacogram denotes the variance $\gamma(k)$ of a stochastic process averaged at time scale $k$, as a function of $k$; by specifying it, all second order properties of the process (autocovariance, power spectrum, variogram) are also uniquely specified. In the last equation $\alpha$ and $\lambda$ are scale parameters with dimensions of $[t]$ and $[x]$, respectively, while $M$ (fractal parameter) and $H$ (Hurst parameter) are dimensionless parameters determining the dependence structure at a local level (smoothness or fractality) and the global level (long-range dependence). The parameter values are chosen as $\lambda = 1, \alpha = 20, H = M = 0.85$. The chosen values of $H$ and $M$ suggest respectively (long-term) persistence and (short-term) smoothness of the process $\underline{x}_\tau$. We deliberately choose a high value of Hurst parameter $H$, corresponding to a process with long-range dependence, in order to make the case study more challenging and insightful, also noting that a high $H$ implies high uncertainty, also in the estimation process. The resulting autocorrelations are high, but not prohibitively high in the sense described in Section SI2.2 of the Supplementary Information of this paper.

We recall (Koutsoyiannis, 2010, 2016) that the discrete time autocovariance $c_\eta$ for integer time lag $\eta$ is the second discrete derivative of the climacogram multiplied by the

square of the time scale. Specifically, assuming a discretization time step $D = 1$, the autocovariance is

$$c_\eta = \frac{1}{2}\left((\eta - 1)^2 \gamma(|\eta - 1|) - 2\eta^2 \gamma(|\eta|) + (\eta + 1)^2 \gamma(|\eta + 1|)\right) \tag{5}$$

The series of coefficients $a_i$ is calculated from that of $c_\eta$ as described in detail in Koutsoyiannis (2020a).

The system output $\underline{y}_\tau$ is calculated in our synthetic case studies from an equation similar to (3) but now replacing the white noise $\underline{w}_\tau$ with the input $\underline{x}_\tau$, and the latter with the output $\underline{y}_\tau$, while also adding some noise $\underline{u}_\tau$:

$$\underline{y}_\tau = \sum_{i=-I_L}^{I_H} a_i \underline{x}_{\tau-i} + \underline{u}_\tau \tag{6}$$

where the integers $I_L$ and $I_H$ differ in the various applications as specified below. The noise $\underline{u}_\tau$ is assumed Gaussian with standard deviation 0.5, except in one case marked as "pure" (applications #1 and #2) where no noise is added. The length of the generated series is 8000 in all synthetic case studies. Here we assume that only these synthetic series are known, while the generation equation (6) is unknown, and we will try to recover (or approximate) it from the data by using equation (2).

We note that, since the sequence of $a_i$ was determined assuming (long-term) persistence ($H = 0.85$), all $a_i$ are nonnegative. Hence in case where $I_L, I_H \le J$, if our method works well, we expect to find that the sequence of $g_i$ (which we assume unknown and try to estimate by our method) is identical to the sequence of $a_i$. This, however, will apparently not be the case if $I_L, I_H > J$.

In case studies #1 and #2 we use $I_L = I_H = 20$ without noise, thus building a symmetric HOE causal system without an error term. In this case, assuming $J = 20$ (41 weights) without using the constraints for roughness and nonnegativity (i.e. using the analytical solution of equation (37) in the companion paper, with $\lambda = 0$), we fully recover the system dynamics in the direction $x \to y$ (case study #1), as shown in Figure 1 (left). As seen in Table 1, the variance is fully explained by that dynamics ($e = 1$).

If we reverse the direction, i.e. $y \to x$, without using any constraint (case study #2), the explained variance ratio remains very high, $e = 0.99$ (Table 1), but the IRF, depicted in Figure 1 (right), becomes very rough. Clearly, the time symmetry is captured, but apart from this, the shape of the IFR looks like representing a random pattern alternating between positive and negative parts. Does such an alternating random pattern suggest causality?

One may claim that the number of weights (41) is too high and regard the random shape as an artefact of non-parsimonious modelling (even though the data size is quite large, 8000, which should support the estimation of 41 parameters). For this reason, we

have also tried to recover the dynamics with fewer (21) weights ($J = 10$). The resulting IRF is also plotted in Figure 1 (right) and again has a similar shape, with a symmetric, yet random pattern alternating between positive and negative parts. The explained variance ratio remains almost equally high, $e = 0.99$ (Table 1, case study #3), which certainly suggests that the solution with 41 weights could become more parsimonious without sacrificing predictability. But does this high predictability of $\underline{x}_\tau$ from $\underline{y}_\tau$ suggest causality, given the rough and alternating pattern of weights?
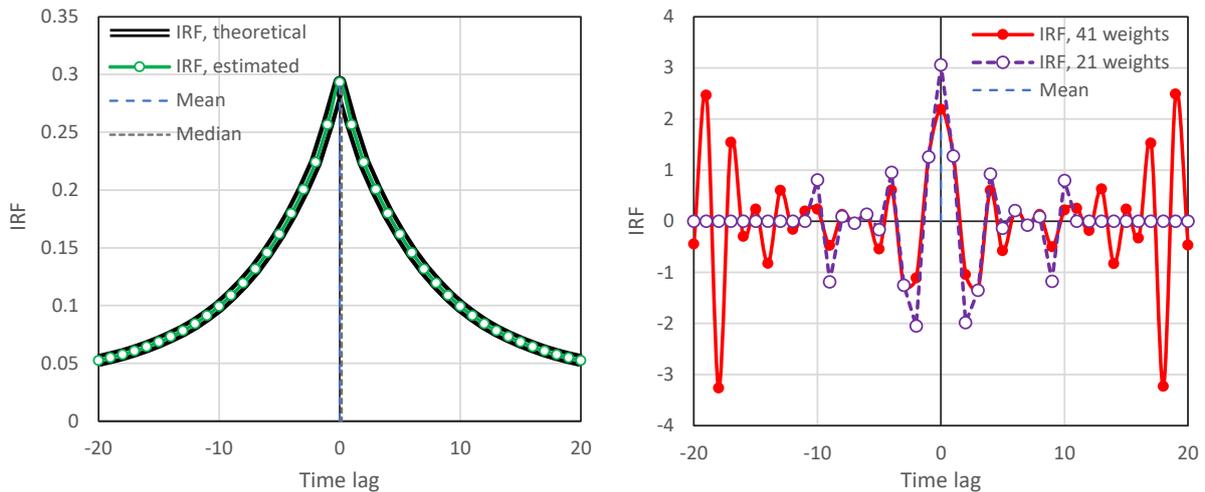


**Figure 1** IRFs for synthetic applications #1 (left) and #2 – #3 (right) representing a HOE causal (symmetric) system without an error term. By construction, the true IRF has 41 nonzero weights ($I_L = I_H = 20$) and the system dynamics does not contain a random term. No constraints were used in the IRF estimation. For the estimated IRF the number of weights is $2J + 1$ with $J = 20$ (applications #1 – #2) or $J = 10$ (application #3).
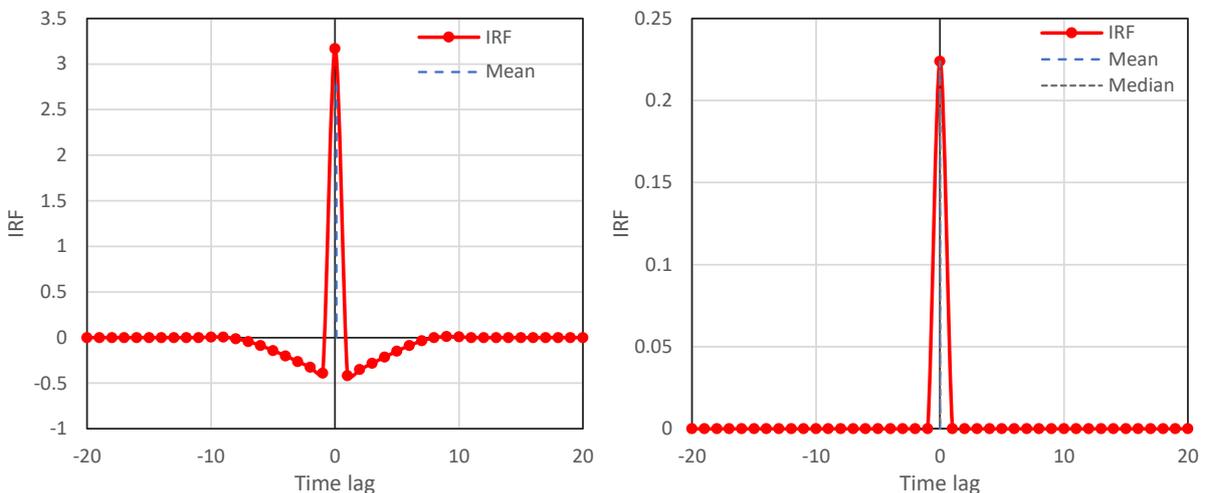


**Figure 2** IRFs for synthetic applications #4 (left; $J = 10$ roughness constraint) and #5 (right; $J = 20$, nonnegativity constraint) representing a HOE causal (symmetric) system without an error term, the same as in Figure 1.

To decrease the roughness, we have included case study #4 with direction $y \rightarrow x$, 21 weights and a roughness constraint. The solution is depicted in Figure 2 (left), where a more logical pattern of the IRF is seen. This was achieved almost with a negligible cost in predictability ($e = 0.98$ in case study #4 vs. 0.99 in case study #3; Table 1). Yet again we

have an alternation of negative and positive parts. Notably, $g_0 = 3.2$, while $g_{\pm 1} = -0.4$. The alternation of signs for a change of lag of just 1 is fully understandable in terms of improving predictability, but does it have any meaning in establishing causality?

Our final application with the same data set and direction $y \rightarrow x$ is case study #5 in which we enable the nonnegativity constraint (with or without the roughness constraint). Here we get a reasonable solution (Figure 2, right), with only one nonzero weight at time lag zero. The resulting explained variance ratio is somewhat smaller, $e = 0.91$, and is fully due to the high lag-zero cross-correlation of $\underline{x}_\tau$ and $\underline{y}_\tau$ ($r_{yx}(0) = 0.94$). All characteristic time lags are equal to 0, reflecting the full temporal symmetry, as also happens with the direction $x \rightarrow y$ (case study #1). Yet in this fully symmetric case, even if we did not know that $\underline{y}_\tau$ was constructed from $\underline{x}_\tau$, we would conclude that there is a preferential causality direction, $x \rightarrow y$, as this results in higher explained variance and a more consistent and prolonged IRF (a bigger time span $\mathbb{h}$).

In case studies #6 and #7 we use $I_L = 0, I_H = 20$, thus making a typical causal system (including an error term). If we do not use any constraint, we get the solutions shown in Figure 3 (left for $x \rightarrow y$, right for $y \rightarrow x$). In both cases, the time directionality $x \rightarrow y$ is captured (see Table 1) but the rough shape in the direction $x \rightarrow y$ (case study 6) and the alternating positive and negative IRF values in both directions do not have any relationship with the true system dynamics. Interestingly, the explained variance ratio is higher in the direction $y \rightarrow x$ ($e = 0.97$) than in $x \rightarrow y$ ($e = 0.94$), which could not be expected. While causality can be inferred from characteristic time lags, the IRFs do not help in understanding causality. At this stage the results seem rather puzzling but things will become clearer with the next case studies.
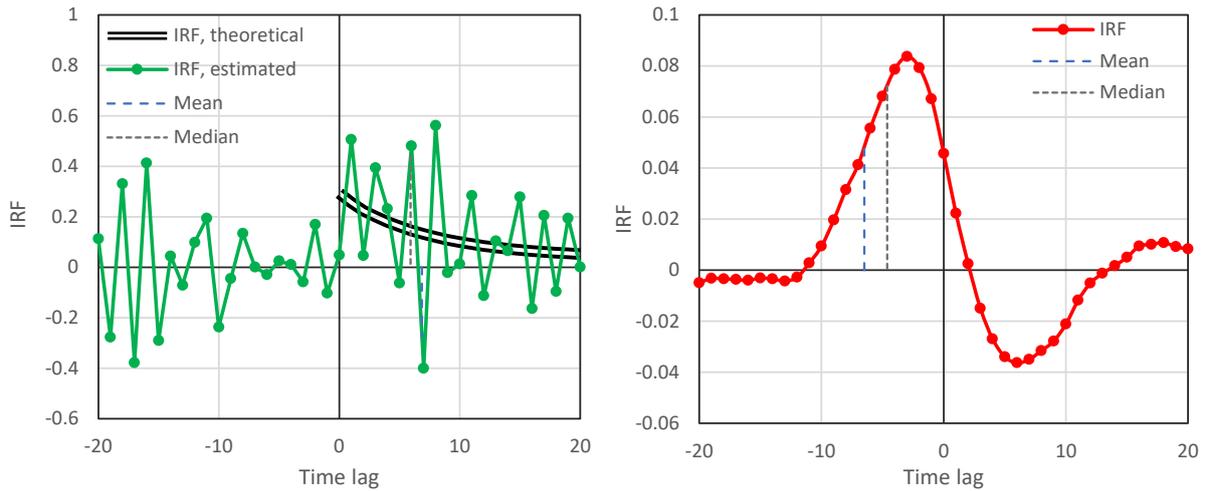


**Figure 3** IRFs for synthetic applications #6 (left) and #7 (right) representing a causal system. By construction, the IRF has 21 nonzero weights ($I_L = 0, I_H = 20$) and a random term in the system dynamics. No constraints were used in the estimation of IRF.

In case studies #8 and #9 we use the same time series $x_\tau$ and $y_\tau$ as in applications #6 and #7. Here we additionally enrol the nonnegativity constraint for the IRF estimation,

but not the roughness one. As seen in Figure 4 (left) and Table 1, the estimated IRF reproduces the zero $g_j$ for time lags $j < 0$, thus capturing the direction and the correct time lags, but the IRF is far from the true system dynamics because it is very rough. If we reverse the direction, i.e. $y \to x$, the system becomes anticausal (zero $g_j$ for time lags $j > 0$), thus again recovering the correct causality direction. In both cases the explained variance ratio is very high, $e = 0.94$, while, as seen in Figure 4 (right), case #9 results in a rather smooth IRF. Yet the clearly positive lags for $x \to y$ and negative ones for $y \to x$ do not leave any doubt that the causality direction is the former.
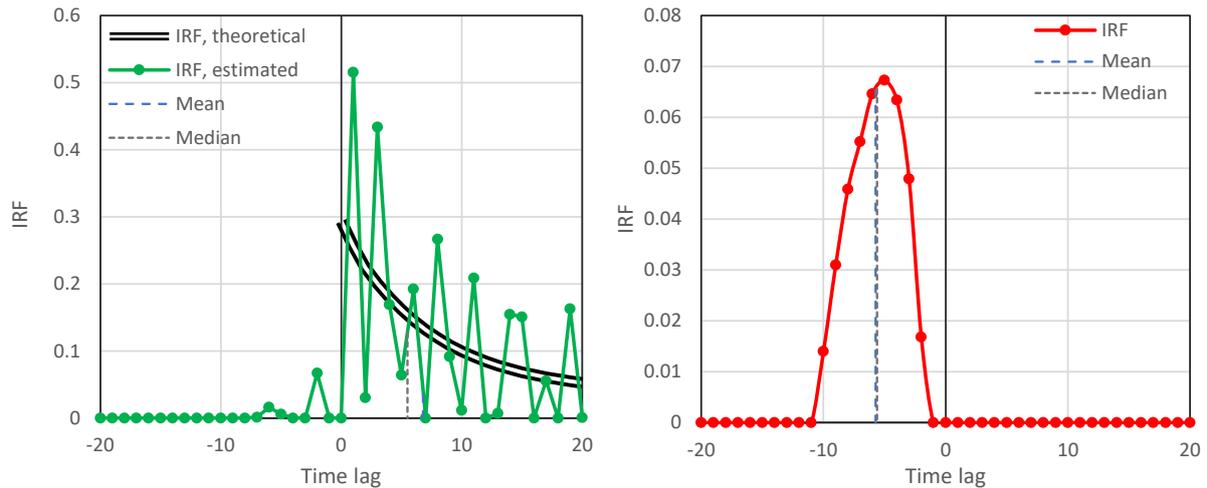


**Figure 4** IRFs for synthetic applications #8 (left) and #9 (right) representing a causal system, the same as that of Figure 3. Only the nonnegativity constraint was used in the estimation of IRF.
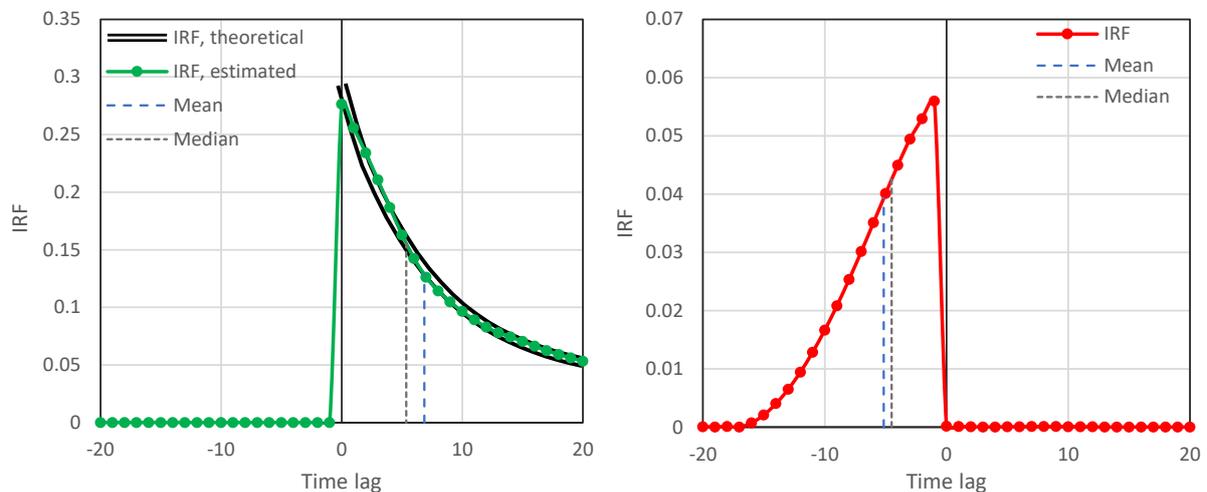


**Figure 5** IRFs for synthetic applications #10 (left) and #11 (right) representing the same causal system as that in Figure 4, but using both the nonnegativity and the roughness constraint in the IRF estimation.

Case studies #10 and #11 again use the same time series $x_\tau$ and $y_\tau$ as in #6 – #9, representing the same typical causal system. The difference is that we now use both constraints, nonnegativity and small roughness. As seen in Figure 5 (left), now the system dynamics is fully recovered, while, as seen in Table 1, this has been done at virtually no cost in terms of explained variance ratio, which is as high as in case #8 ($e = 0.94$). If we

investigate the reverse direction, $y \rightarrow x$ (Figure 5 right), the system correctly appears as anti-causal (zero $g_j$ for time lags $j > 0$), thus again recovering the correct causality direction.

The above illustrations of the IRF behaviour with and without the constraints prove that both constraints are essential in studying causality. Therefore, in what follows, we always use both of them.

Another interesting question to investigate is this. What happens if the system is nonlinear in nature, while our entire framework is based on a linear relationship? To study this question, we exponentiate the time series $y_\tau$ of case studies #8 – #11, while we leave $x_\tau$ unchanged and we apply the linear framework again. In other words, now the actual system is

$$\underline{y}_\tau = \exp\left(\sum_{i=0}^{I_H} a_i \underline{x}_{\tau-i}\right)\exp(\underline{u}_\tau) \tag{7}$$

and we try to approximate it in a linear fashion and estimate it as

$$\hat{\underline{y}}_\tau = \sum_{j=-J}^{J} g_j \underline{x}_{\tau-j} + \mu_v \tag{8}$$

Now we should expect a larger variance of $\underline{v}_\tau$ because of the disagreement between the actual system and the model used for causality detection.

As seen in Figure 6 (left), referring to the direction $x \rightarrow y$, our framework correctly detected that we have a potentially causal system. The characteristic time lags do not show a noteworthy change in comparison with case #10, despite the fact that, as seen in Table 1, the explained variance ratio has been substantially reduced from 0.94 to 0.32. If we change the direction to $y \rightarrow x$ (Figure 6, right), the ratio increases to 0.43, and the causality appears as HOE type, but with the correct principal direction, $x \rightarrow y$ (anticausal at $y \rightarrow x$). Therefore, here we locate a potential problem of the methodology, as the actual causality is not HOE.

It is not too difficult to resolve this ambiguity: If we produce a scatter plot of $y_\tau$ vs. $x_\tau$ (Figure 7, right), it becomes evident that the relationship of $\underline{x}_\tau$ and $\underline{y}_\tau$ is not linear; for comparison, Figure 7 (left) shows a similar plot for case studies #8 – #11, which is typical for linear relationships. Once we are aware of the nonlinearity, we will perform a nonlinear transform on $\underline{x}_\tau$, $\underline{y}_\tau$ or both, and reapply the methodology on the transformed series. In this particular case, if we apply the logarithmic transformation on $y_\tau$, we will switch to cases #8 – #11 and we will fully recover the system dynamics. Additional details on the application of this technique are given in section 2.2.
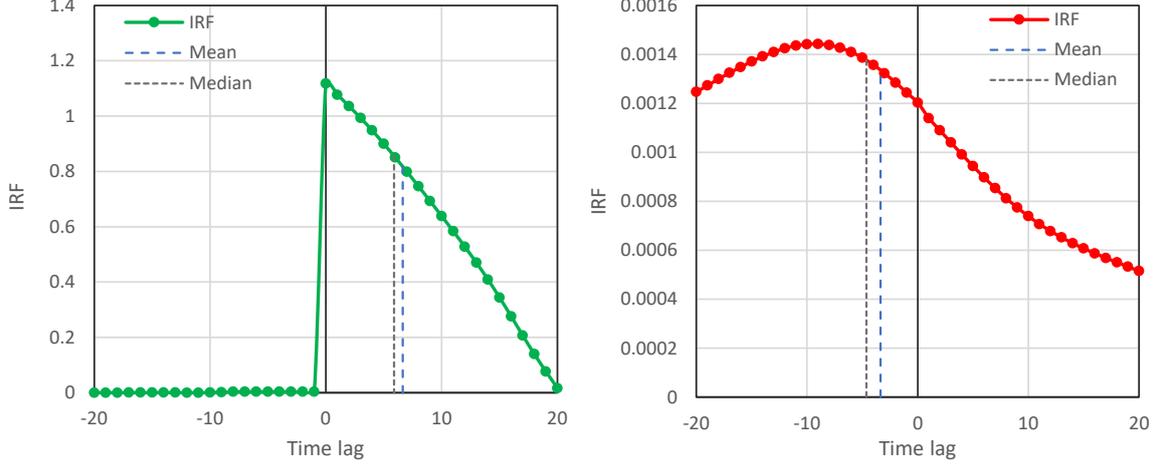
**Figure 6** IRFs for synthetic applications #12 (left) and #13 (right) representing the causal system of Figure 5, but with exponentiating the output of the latter.
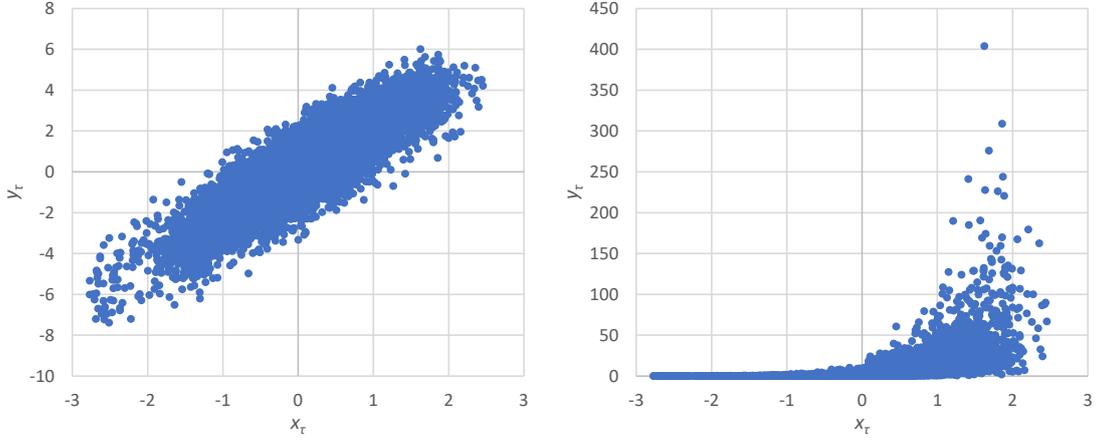


**Figure 7** Scatter plots of synchronous $y_\tau$ vs. $x_\tau$ for applications #8 – #11 (left) and #12 – #13 (right).

In the next case studies, #14 and #15, we have a causal model with long-range cross-dependence, constructed by using $I_L = 0, I_H = 1024$ in equation (6). However, we keep using $J = 20$ in our causality detection framework and thus we do not expect to fully recover the true long-term system dynamics. We only test whether we can correctly detect causality and its direction. Using only the roughness constraint we find the IRF shown in Figure 8 (left). Notice that the true (theoretical) IRF curve exceeds the horizontal axis span by far, going up to lag 1024, while the plotted area coincides with the time window of the sought IRF estimate (i.e., it goes up to lag 20 only). As seen in Table 1 the true characteristic lags $\mu_h$, $h_{1/2}$ are of the order of 200 while with our chosen time window we can estimate lags up to 20—and most likely of the order of 10. Indeed, the methodology captured the mono-directional causality yet the estimated characteristic time lags are, as expected, too small compared to the true ones of Table 1. Interestingly, Figure 8 (left) shows increasing IRF magnitude beyond the estimated average lag time. This may seem inconsistent, as the true IRF does not contain an increasing branch. However, this is a reasonable behaviour reflecting the fact that the chosen time window ($J = 20$) is too narrow to contain the true IRF. It is thus reasonable for our framework to

increase the power of the most distant components (closer to lag 20) as a replacement of the power of even more distant ones.

It is most important that the estimated IRF reproduces the zero $g_j$ for time lags $j < 0$, thus capturing the correct causality direction. If we reverse the direction, i.e. $y \rightarrow x$, the system becomes anti-causal (zero $g_j$ for time lags $j > 0$), thus again recovering the correct causality direction. In both cases the explained variance ratio is high, $e = 0.57$ and $0.50$, for $x \rightarrow y$ and $y \rightarrow x$, respectively. Notably, in the latter case if we considered only the correlation with just one term, with time lag –8, the explained variance ratio would be virtually the same. Overall, the results do not leave any doubt that the true causality direction is $x \rightarrow y$.
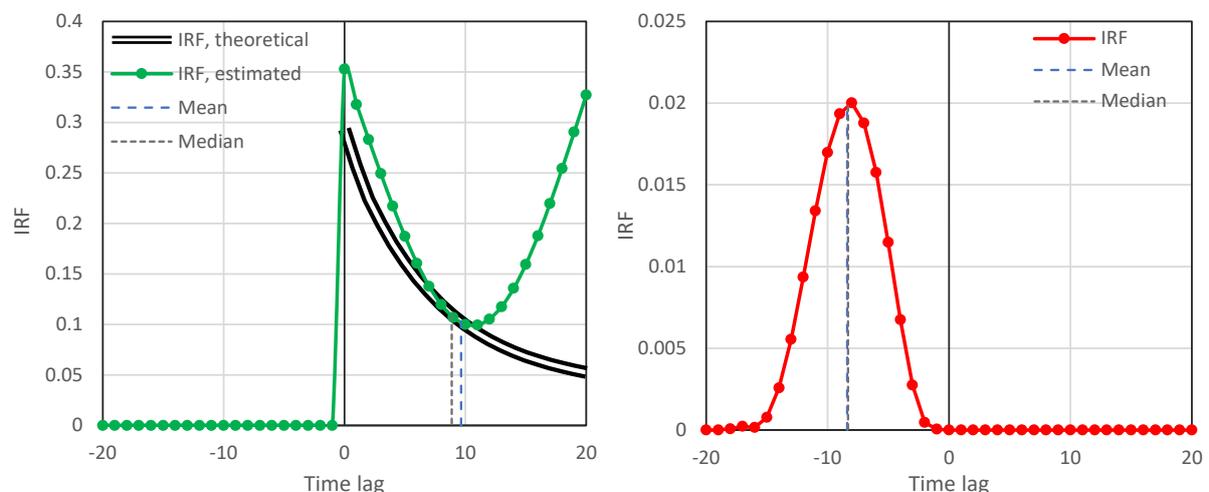


**Figure 8** IRFs for synthetic applications #14 (left) and #15 (right) representing a causal system. By construction, the IRF has 1025 nonzero weights ($I_H = 1024$) and a random term in the system dynamics.

In our final synthetic applications, #16 – #18, we consider a symmetric HOE causal system. This is similar to that in the first applications #1 and #2 , except that we now use long-range cross-dependence with $I_L = I_H = 1024$ and include a random term $\underline{u}_\tau$. Applications #16 and #17, depicted in Figure 9 (left), examine the causal direction $x \rightarrow y$ and they only differ in the roughness term in optimization of IRF. Namely, in application #16 the roughness was calculated without considering the second derivative at zero, in agreement with the fact that in the true system the second derivative is not defined at zero due to discontinuity of the first derivative. In application #17 all roughness terms are included and, as a result, a smoother IRF curve is produced. In both cases, our methodology captures the time symmetry of the system. The increasing $g_j$ as the $j$ approaches ±20 give an alert that our time window is too narrow to capture the long-range cross-dependence—something similar with what we observed in application #14. If we reverse the direction, i.e. $y \rightarrow x$ (application #18), again the time symmetry of the system is confirmed, but now the explained variance ratio (Table 1) decreases from $e = 0.71$ (for $x \rightarrow y$) to $0.57$ (for $y \rightarrow x$). Notably, in the latter case, if we considered only the synchronous (lag 0) cross-correlation, the explained variance ratio would be virtually the

same. For this reason, we will again choose as preferential the direction $x \rightarrow y$ and try a wider time window to determine the IRF.
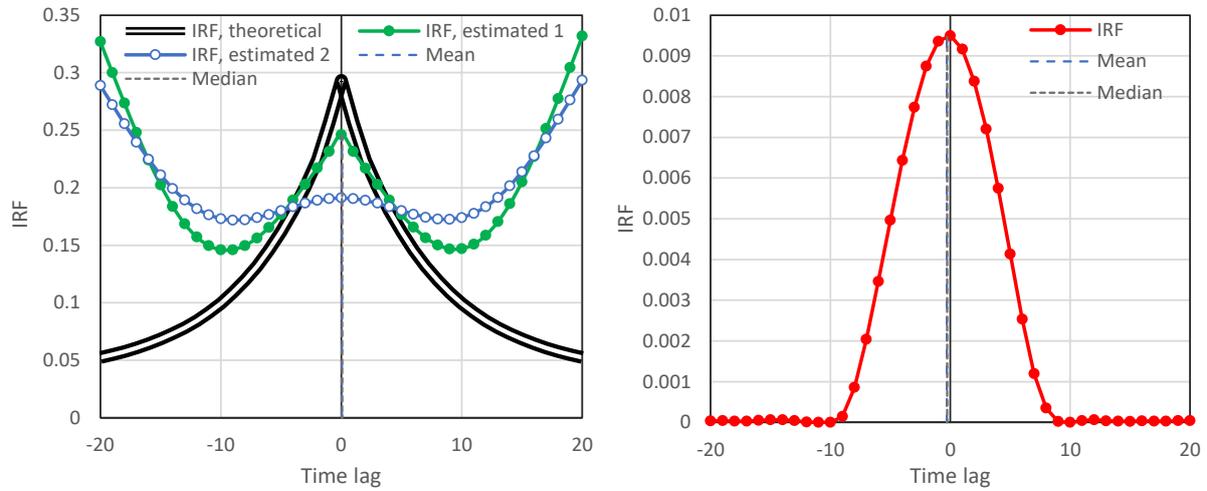


**Figure 9** IRFs for synthetic applications #16 and #17 (left), and #18 (right) representing a symmetric HOE causal system. By construction, the IRF has 2049 nonzero weights ($I_L = I_H = 1024$). In "estimation 1" the roughness was calculated without considering the second derivative at zero (application #16), while in "estimation 2" all roughness terms are included (application #17).

## 2.2 Precipitation – runoff

At the *global scale,* the hydrological cycle is obviously a family of processes that act in a cyclical manner (Koutsoyiannis, 2020b), precipitation – runoff – precipitation – …, and therefore can be thought of as a hen-or-egg case of causality. However, if we specify a particular location on Earth, the situation is different and it is well known that at a *local scale* runoff is caused by past precipitation upstream in the drainage basin in a mono-directional fashion. We thus expect that precipitation ($P$) and runoff ($R$) data should reflect this pattern.[1]

To explore this, we use rainfall and streamflow data from the database of the U.S. Geological Survey for the site USGS 01603000 North Branch Potomac River Near Cumberland, MD (39°37'18.5"N, 78°46'24.3"W, catchment area 2271 km[2]). The data series are for the period 2013-10-01 to 2021-02-25 for a time step of 15 min (a part of these data was also used for a similar purpose in Koutsoyiannis, 2019). The discharge data were converted to metric units (from cubic feet per second to m[3]/s and the precipitation data from inches to mm). Both series have a small percentage of missing values, which

---

[1] There is a common perception in the hydrological community that the precipitation – runoff transformation can be modelled in a deductive way. Epistemologically, this cannot hold, as a river basin is a complex geophysical system (cf. Bode and Shannon, 1950, quoted in the companion paper; Koutsoyiannis et al., 2022). It is true that some of the mechanisms of the transformation are described by differential equations as dynamical systems. However, the modelling of the entire system cannot be reliably made without data and without moving from a deterministic to a stochastic description (cf. Montanari and Koutsoyiannis, 2012; Koutsoyiannis and Montanari, 2022). Therefore, induction is absolutely necessary. This remains a demanding problem as reflected in the notion of equifinality (Beven, 2019).

were left unfilled. The series were aggregated to a time scale of 3 h in order for our time window of length 40 to represent a period of a few days (120 h or 5 d).

Indeed, Figure 10 (left) and Table 1 (case study #19) suggest a causal system with direction $P \rightarrow R$, a time span > 20 time units (60 h) and characteristic time lags > 10 time units (30 h). The increasing $g_j$ beyond the time lag $j = 17$, is clearly an artefact produced by the fact that the time span of the causal relationship exceeds the size of our time window. Had the latter been large enough, we would expect to see a shape like a unit hydrograph, with a monotonically decreasing limb after the peak at time lag 6 (see below). If we reverse the direction, i.e. $R \rightarrow P$ (Figure 10, right, and Table 1, case study #20) we just confirm that the true causality direction is the opposite, i.e. $P \rightarrow R$. The explained variance ratio in the direction $R \rightarrow P$ is very low, $e = 0.04$, not greater than implied by merely the maximum cross-correlation value. In the correct direction, $P \rightarrow R$, this increases by a factor of 4 ($e = 0.17$). Yet the latter value would still be too low if our aim were to capture the system dynamics. One reason for this low value is the narrow time window, as already discussed. A second reason becomes obvious from the scatter plot of runoff vs. precipitation in Figure 11 (left). The scatter plot does not suggest linearity and we may expect a larger $e$ if we transform the two variables.
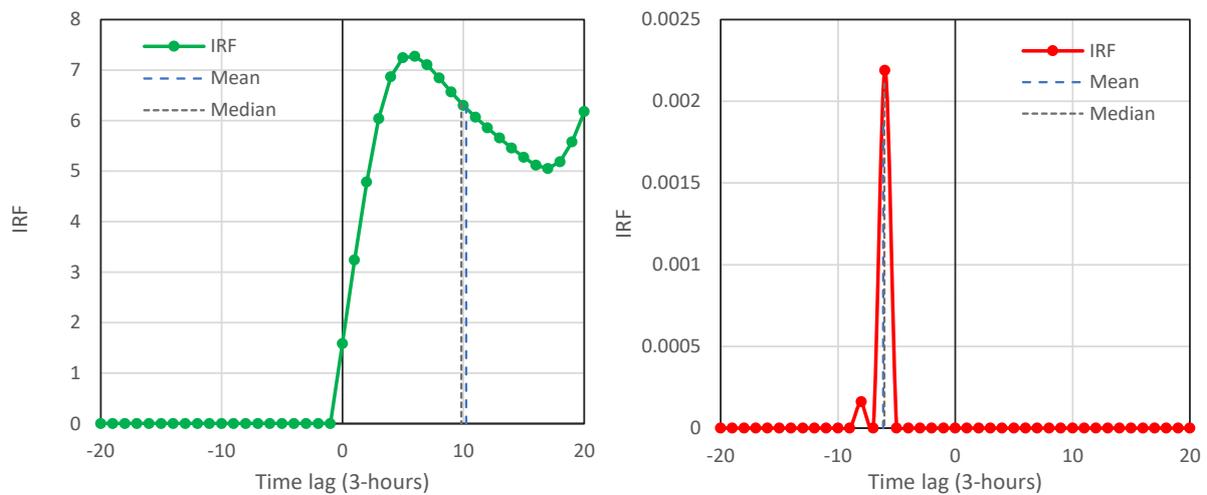


**Figure 10** IRFs for precipitation – runoff case studies #19 (left) and #20 (right).

We choose to apply to each of the variables a single-parameter ($c$) transformation suggested in an entropy maximization framework (Koutsoyiannis, 2014), namely

$$P' = c_{\mathrm{P}} \ln\left(1 + \frac{P}{c_{\mathrm{P}}}\right), \qquad R' = c_{\mathrm{R}} \ln\left(1 + \frac{R}{c_{\mathrm{R}}}\right) \tag{9}$$

where the two parameters were optimized simultaneously with the IRF and were found to be $c_{\mathrm{P}} = 0.0392 \text{ mm/h}$, $c_{\mathrm{R}} = 0.001 \text{ m}^3/\text{s}$. Practically, the low values of the parameters mean that the transformations made are almost pure logarithmic. Yet the transformation in equation (9) is more advantageous than the pure logarithmic, because of its property of keeping the smallest values for $P$ or $R < c$ virtually unchanged and the zero values

precisely unchanged. The scatter plot of the transformed variables is presented in Figure 11 (right). While a linear arrangement of the points is more visible in the transformed variables rather than in the untransformed ones (Figure 11, left), there was no improvement (actually there was slight worsening) in the achieved maximum cross-correlation (Figure 11 and Table 1). Yet the transformation resulted in a substantial improvement in the explained variance ratio: from $e = 0.17$ it increased fourfold, to $e = 0.68$ for the direction $P' \rightarrow R'$. In the opposite direction there was no change. This means that a cross-correlation at a single time lag is a very poor representation of (potential) causality and justifies our framework which is based on an appropriate range of time lags.

The IRFs of the transformed processes are shown in Figure 12, which is not essentially different from Figure 10. This means that, while the transformation offers explanatory power, our methodology, even applied to the untransformed processes, is able to reveal the causal relationship (recall also case studies #12 – #13 and see theoretical explanation in Section SI1.2 of the Supplementary Information of the companion paper—Koutsoyiannis et al., 2022).
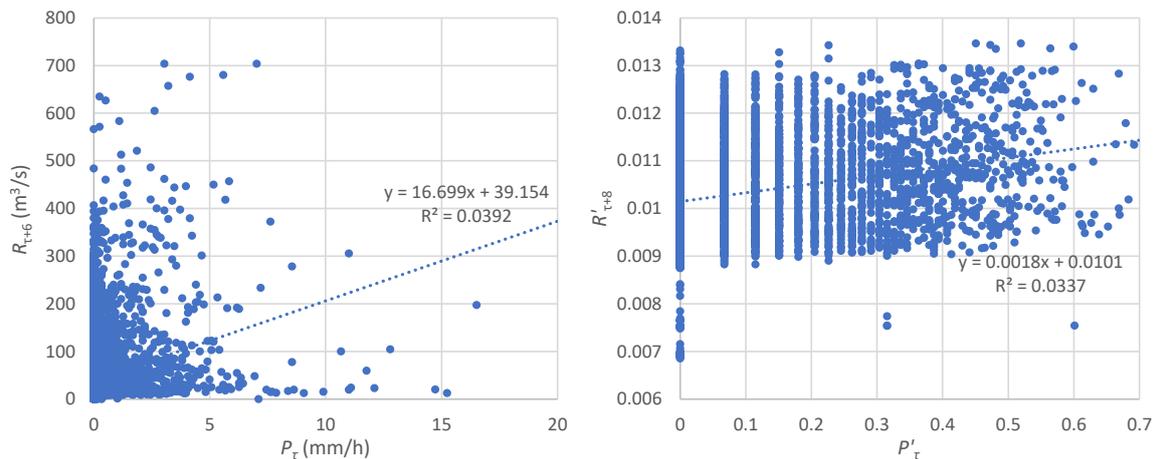


**Figure 11** Scatter plots of lagged runoff ($R$) vs. precipitation ($P$), where the time lag is the one that maximizes cross-correlation. Left: untransformed variables, applications #19 – #20; right: transformed variables, applications #21 – #22.
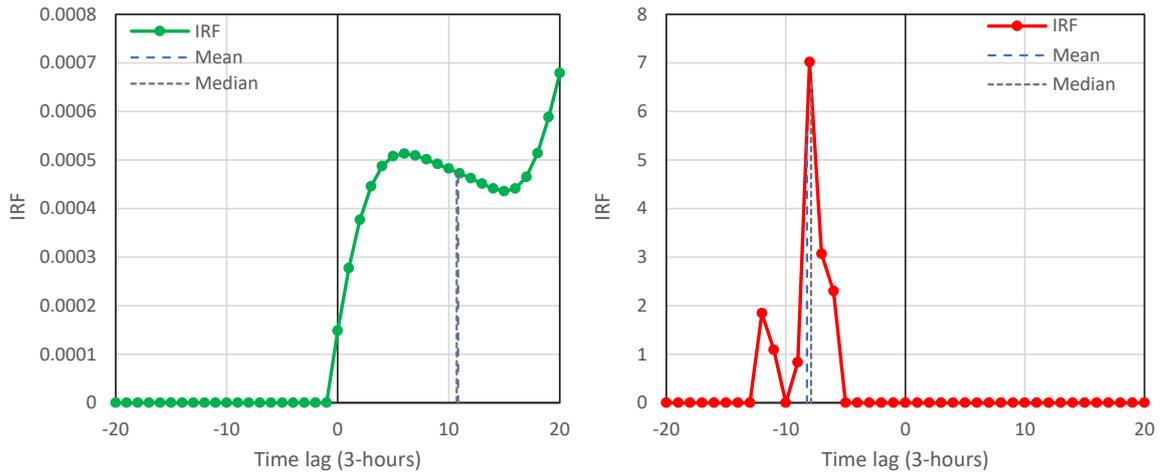
**Figure 12** IRFs for transformed precipitation – runoff case studies #21 (left) and #22 (right); for the IRFs of the original data set see Figure 10.

To show that the increasing limb of $g_j$ beyond the time lag $j = 17$ in applications #19 and #21 is an artefact of the small time window, we repeated the calculations with a time window twice as long, i.e. $J = 40$ instead of $J = 20$. The results are shown in Figure 13. Where there were increasing limbs in applications #19 and #21, now there is continuation of the decreasing limbs, as expected. As a result, the characteristic lags increased: in 3-h time units, the mean $\mu_h$ became 17.35 and 19.10 for untransformed and transformed variables, respectively (from 10.24 and 10.84, respectively); and the median $h_{1/2}$ increased to 15.64 and 18.54 (from 9.83 and 10.72, respectively). The explained variance increased to 0.26 and 0.71 for untransformed and transformed variables, respectively (from 0.17 and 0.68, respectively). However, again there appear (smaller) increasing limbs close to the right end of the wider window. This means that even $J = 40$ is still not enough to recover the dynamics and we should choose an even higher value of $J$. But as we have repeatedly stated, the scope of the study is not to provide a model of the process but to explore causality.
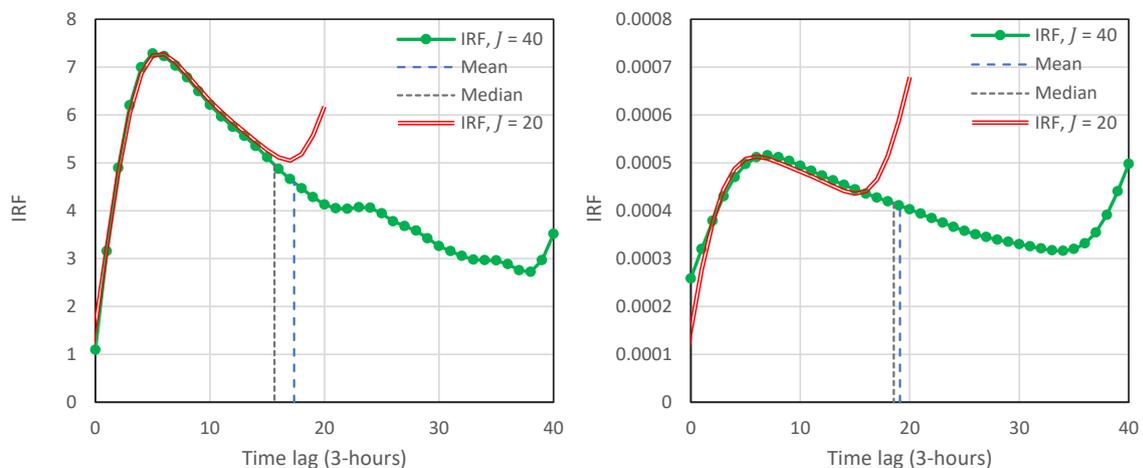


**Figure 13** IRFs for precipitation – runoff case studies similar to #19 (left; untransformed variables) and #21 (right; transformed variables) but with a with a time window twice as long, i.e. $J = 40$ instead of $J = 20$; results are plotted only for nonnegative lags.

## 2.3 Atmospheric temperature and $CO_2$ concentration

The problem related to the causal relationship between atmospheric temperature ($T$) and concentration of carbon dioxide ($[CO_2]$) is regarded by many as part of a "settled science" yet it remains challenging and still debated. For example, the study by Koutsoyiannis and Kundzewicz (2020) concluded, making use of the hen-or-egg causality concept and based on the analysis of modern measurements of $T$ and $CO_2$, that the principal causality direction is $T \rightarrow [CO_2]$, despite the common conviction that the opposite is true. In addition, using palaeoclimatic proxy data from Vostok ice cores, Koutsoyiannis (2019) and Koutsoyiannis and Kundzewicz (2020) found a time lag of $[CO_2]$ from $T$ of a thousand years. Here we re-examine both modern and paleo data sets with our proposed causality detection methodology.

As modern observations for global temperature we use the satellite dataset developed at the University of Alabama in Huntsville (UAH). The temperature of three broad levels of the troposphere is inferred from satellite measurements of the oxygen radiance in the microwave band, using advanced (passive) microwave sounding units on NOAA and NASA satellites (Spencer and Christy, 1990; Christy et al., 2007). The dataset begins in 1979 continues to date. It is publicly available at a monthly scale in the form of time series of "anomalies" (defined as differences from long-term means) for several parts of the Earth, as well as in map form. Here we use only the global average (noting that Liang, 2022, disputes the use of averages as he claims that generally they conceal regional patterns of change) on the monthly scale for the lowest level, referred to as the lower troposphere. For the $CO_2$ concentration we use the most famous dataset, that of the Mauna Loa Observatory (Keeling et al., 1976). The Observatory, located on the north flank of Mauna Loa Volcano, on the Big Island of Hawaii, USA, at an elevation of 3397 m above sea level, is a premier atmospheric research facility that has been continuously monitoring and collecting data related to the atmosphere since 1958. Here we examine the common 42-year period of the two datasets (1979 to May 2021) at a monthly scale.

Both data sets were also used by Koutsoyiannis and Kundzewicz (2020) who also examined a ground-based temperature data series (CRUTEM.4.6.0.0 global T2m land temperature) and three additional $CO_2$ series (Barrow, Alaska, USA; South Pole; global average). The results did not substantially differ for the different pairs of $T – [CO_2]$ series and therefore here we limit our current analysis to one pair. We additionally note that we also examined the temperature data of the other two satellite levels for the troposphere and the results were very similar to those reported here for the lower troposphere.

Furthermore, in our analysis we follow the data pre-processing justified in Koutsoyiannis and Kundzewicz (2020); namely:

1. We take the logarithms of $[CO_2]$, based on Arrhenius's (1896) rule that when $[CO_2]$ increases in geometric progression, $T$ will increase nearly in arithmetic progression.

2. We take the difference of each monthly value from that of the same month of the previous year. This diminishes the effect of seasonality while eliminating possible artificial effects of the convention of giving the temperature data as "anomalies" (departures from changing monthly means) rather than actual values.

We note that, since we are differencing the process, taking the average of 12 consecutive monthly differences (which is a more established process) is equivalent to taking a difference for a time step of 12 months (notice that $x_2 - x_1 + x_3 - x_2 + \cdots x_{13} - x_{12} = x_{13} - x_1$). We further note that Koutsoyiannis and Kundzewicz also examined the non-differenced series and showed that they give spurious results due to the continuously rising $[CO_2]$ in the time window of modern observations, which results in autocorrelation values that are virtually 1 for all lags. Here the case of spurious results due to very high autocorrelation is explained more thoroughly—see Section SI2.2 in Supplementary information, and notice the similarity of Figure SI2.3 with Figure 9 in Koutsoyiannis and Kundzewicz (2020).

Hence the examined processes are $\Delta T$ and $\Delta \ln[CO_2]$. Figure 14 gives the obtained IRFs in the directions $\Delta T \rightarrow \Delta \ln[CO_2]$ (left panel) and $\Delta \ln[CO_2] \rightarrow \Delta T$ (right panel). Impressively, the results are not different from those in the precipitation – runoff case study. Clearly, the results in Figure 14 suggest a (mono-directional) potentially causal system with $T$ as the cause and $[CO_2]$ as the effect. Hence the common perception that increasing $[CO_2]$ causes increased $T$ can be excluded as it violates the necessary condition for this causality direction. Even the possibility of hen-or-egg causality, supported by Koutsoyiannis and Kundzewicz (2020), is not confirmed by the new methodological framework. The causality direction $\Delta T \rightarrow \Delta \ln[CO_2]$ is further supported by all numerical indices given in Table 1, as well as the graphical comparison of modelled and empirical cross-correlation functions depicted in Figure 15. Namely:

- All characteristic time lags ($h_c, \mu_h, h_{1/2}$) are positive in the direction $\Delta T \rightarrow \Delta \ln[CO_2]$ (ranging from 5 to about 8 months), and negative in the direction $\Delta \ln[CO_2] \rightarrow \Delta T$.
- The explained variance ratio is greater in the direction $\Delta T \rightarrow \Delta \ln[CO_2]$ ($e = 0.31$) than in the direction $\Delta \ln[CO_2] \rightarrow \Delta T$ ($e = 0.23$).
- In the direction $\Delta T \rightarrow \Delta \ln[CO_2]$, the cross-correlation function, reconstructed from the IRF and the autocorrelation function using the discretized version of equation (12) in the companion paper; Koutsoyiannis et al., 2022), agrees impressively well with the empirical cross-correlation function. In the direction $\Delta \ln[CO_2] \rightarrow \Delta T$ the proximity of the two is much lower.

Remarkably, however, the explained variance ratio of $e = 0.31$ is low and suggests that the two processes have a behaviour that is much more complex and affected by additional geophysical processes. However, insofar the relationship of these two processes is concerned, this explained variance ratio is adequate to detect the main characteristics, i.e. direction and time lags. Indications for this adequacy are provided by the precipitation – runoff case study (section 2.2), in which $e = 0.17$, as well in the additional controlled (synthetic) case study that is provided in Supplementary Information (section SI2.1, $e = 1/3$).

Having gathered strong indications that in the recent decades the increase of temperature is potentially the cause of the increased $CO_2$ concentration, while the opposite is not probable, it is interesting to examine if this is also the case for longer periods. To this aim we use the datasets from the Vostok ice cores (Jouzel et al., 1987; Petit et al., 1999; Caillon et al., 2003) which were originally given for an irregular time step and were regularized in the study of Koutsoyiannis (2019) for a time resolution of 1000 years.
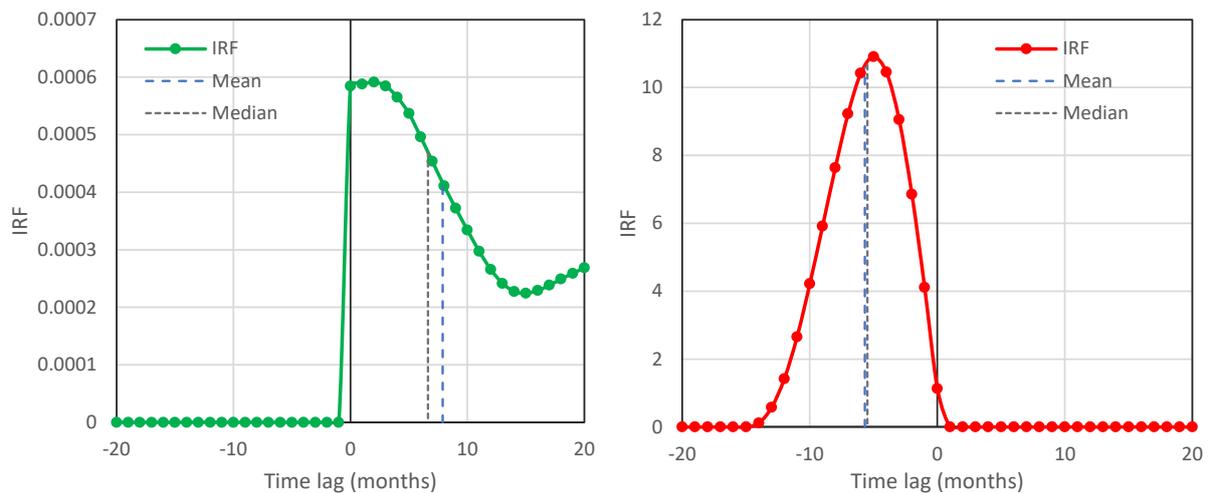


**Figure 14** IRFs for of temperature – $CO_2$ concentration based on the modern time series. (left) Case study #23 ($\Delta T \rightarrow \Delta\ln[CO_2]$); (right) case study #24 ($\Delta\ln[CO_2] \rightarrow \Delta T$).
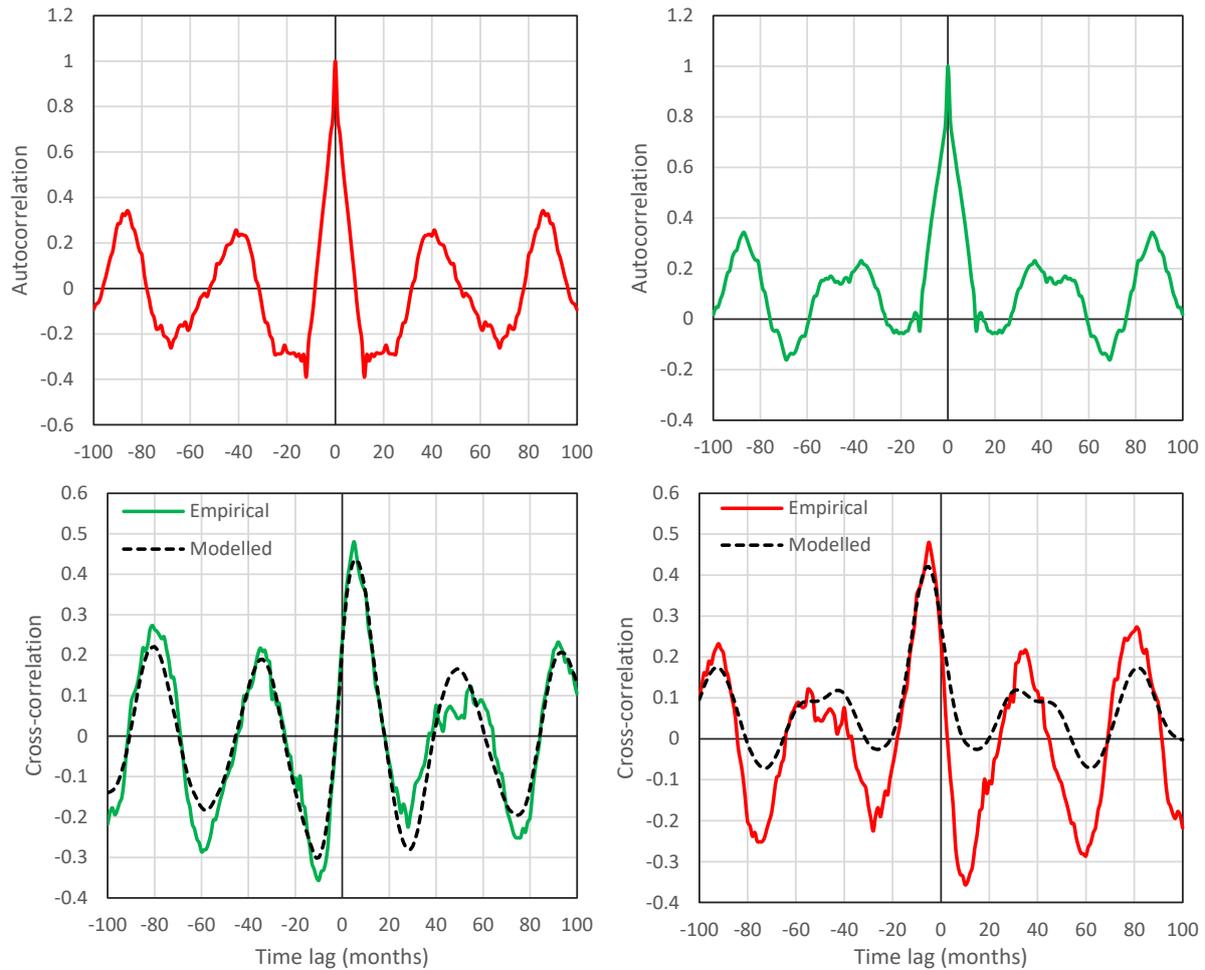
**Figure 15** (**upper**) Autocorrelation function of (**left**) $\Delta T$ and (**right**) $\Delta\ln[CO_2]$. (**lower**) Cross-correlation functions, empirical and reconstructed (marked as "modelled") from the IRF and the autocorrelation functions in the upper panels, using the discretized version of equation (12) in the companion paper, for case studies (left) #23 and (right) #24.

We study again the processes $\Delta T$ and $\Delta\ln[CO_2]$, where the differences are taken for 1 time step (1000 years). Figure 16 gives the obtained IRFs in the directions $\Delta T \rightarrow \Delta\ln[CO_2]$ (left panel) and $\Delta\ln[CO_2] \rightarrow \Delta T$ (right panel). Here the results support a HOE causality. Nonetheless, again the principal direction is $\Delta T \rightarrow \Delta\ln[CO_2]$ with a time lag of the order of 1000 years (0.79 to 1.11 time steps for $\Delta T \rightarrow \Delta\ln[CO_2]$; –0.56 to –0.82 time steps for $\Delta\ln[CO_2] \rightarrow \Delta T$; see Table 1, case studies #25 – #26). The explained variance ratio is small, $e = 0.17$ in both directions.
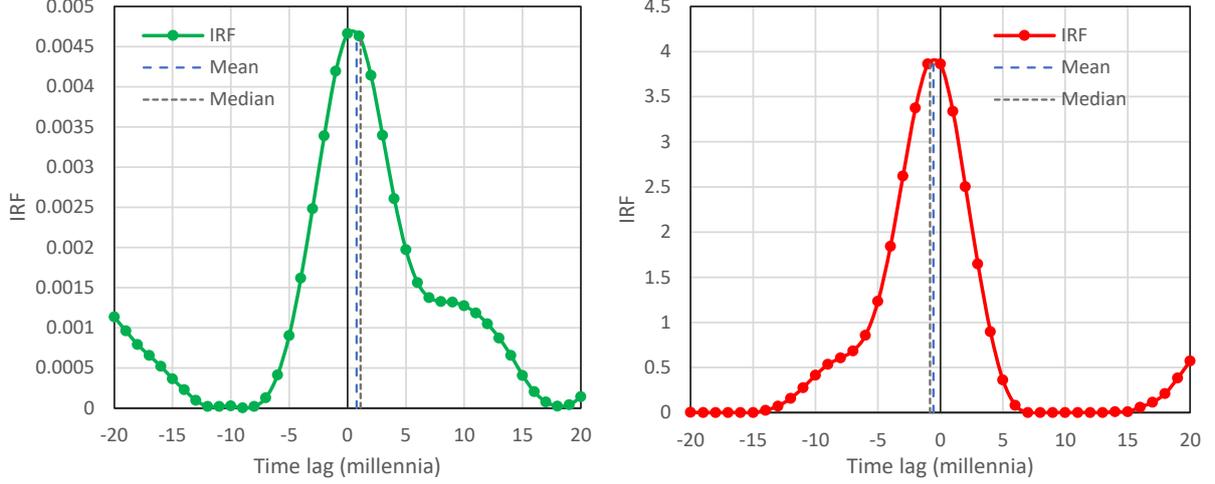
**Figure 16** IRFs for of temperature – $CO_2$ concentration based on the proxy time series from the Vostok ice cores. (left) Case study #25 ($\Delta T \rightarrow \Delta\ln[CO_2]$); (right) case study #26 ($\Delta\ln[CO_2] \rightarrow \Delta T$).

As the proxy data sets are free of monotonic trends and produce reasonable empirical autocorrelation functions (see Koutsoyiannis, 2019), here we could also apply our framework for the non-differenced processes. We initially note that, if $\underline{x}_\tau$ is a differenced process (where the differences are taken for 1 time step) and $\underline{X}_\tau$ the non-differenced (original) one, then the two are related by

$$\underline{X}_\tau = \sum_{j=1}^{\tau} \underline{x}_j \tag{10}$$

and likewise for $\underline{Y}(t)$ and $\underline{V}(t)$. In this case from equation (2) it follows

$$\underline{Y}_\tau = \sum_{j=-\infty}^{\infty} g_j \underline{X}_{\tau-j} + \underline{V}_\tau \tag{11}$$

This is identical to equation (2) which means that the original and differenced processes are related with the same potential causality equation involving the same IRF. Furthermore, as a result of aggregation, and the resulting smoothing of the noise $\underline{V}_\tau$, it is expected that the explained variance ratio should be higher after aggregation.

We test if this happens in the case of the proxy $T$ and $[CO_2]$ data. Figure 17 gives the obtained IRFs in the directions $T \rightarrow \ln[CO_2]$ (left panel) and $\ln[CO_2] \rightarrow T$ (right panel). The shapes of the curves are quite similar as those of Figure 16. Again, we find HOE causality with principal direction is $T \rightarrow \ln[CO_2]$ and with a time lag of the order of 1000 years (1.74 to 1.87 time steps for $\Delta T \rightarrow \Delta\ln[CO_2]$; –1.03 to –1.10 time steps for $\ln[CO_2] \rightarrow T$; see Table 1, case studies #27 – #28). The explained variance ratio increased substantially, from $e = 0.17$ to $e = 0.86$ and is greater in the direction $T \rightarrow \ln[CO_2]$, thus confirming the latter as the principal causality direction.
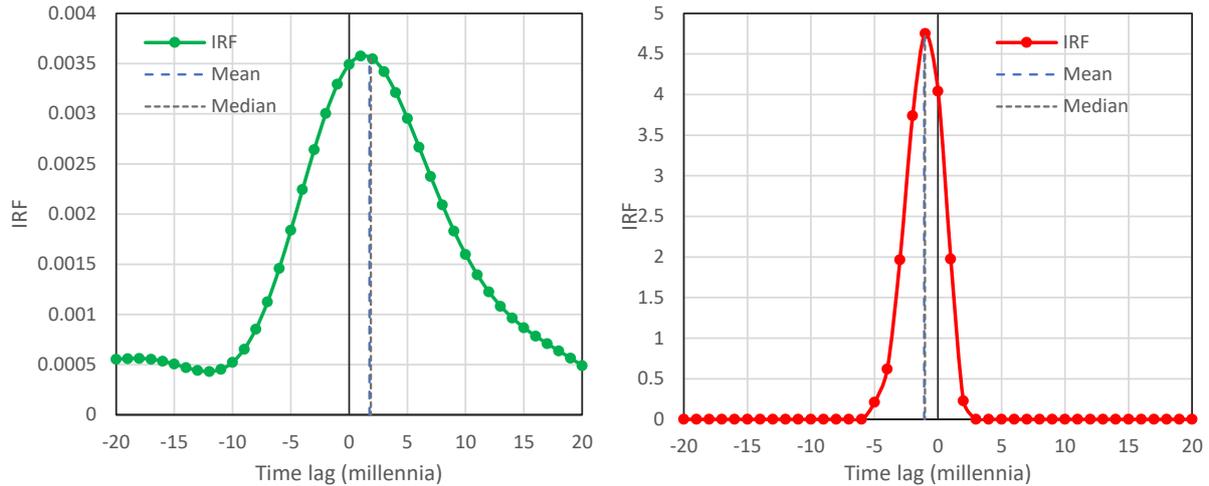
**Figure 17** IRFs for of temperature – $CO_2$ concentration based on the proxy time series from the Vostok ice cores as in Figure 16 but for the aggregated (non-differenced) processes. (left) Case study #27 ($T \rightarrow$ $\ln[CO_2]$); (right) case study #28 ($\ln[CO_2] \rightarrow T$).

## 2.4 Atmospheric temperature and ENSO

For a second case study related to climate, we examine the El Niño–Southern Oscillation (ENSO) in relation to the global temperature. The ENSO is associated with irregular, anti-persistent (at the time scale of a few years, else known as quasi-periodic), variation of sea surface temperature and air pressure over the tropical Pacific Ocean. It is broadly recognized as the principal climate variability mode (McPhaden et al., 2006). There exists a plethora of indices related to ENSO, among which here we use the Southern Oscillation Index (SOI) of the USA's National Oceanic and Atmospheric Administration (NOAA) (see also Kaplan, 2011).

In a recent paper, Kundzewicz et al. (2020) examined several ENSO indices, as well as indices of similar phenomena in the Pacific and in the Atlantic, and found that they meaningfully influence the global mean annual temperature. Here we examine a single pair of processes, the UAH temperature, also used in case studies #23 – #28 (section 2.3), and the SOI. While Kundzewicz et al. (2020) processed the residuals from 5-year running averages to remove the influence of large-scale fluctuations, here we use the data without averaging. However, in both time series we take the difference of each monthly value from that of the same month of the previous year, a technique already described in section 2.3.

The results are depicted in Figure 18, which suggests a clear case of a potentially causal system in the direction SOI $\rightarrow T$. As seen in Table 1, the characteristic lags are close to 6 months and the explained variance is 0.39. We note that the technique used by Kundzewicz et al. (2020) (residuals from 5-year running averages) explained a higher percentage of the variance (>0.60), which means that there is a margin for improvement of the results obtained here. However, as already mentioned, the scope of the current study is the exploration of potential causality, rather than the building of a reliable model.
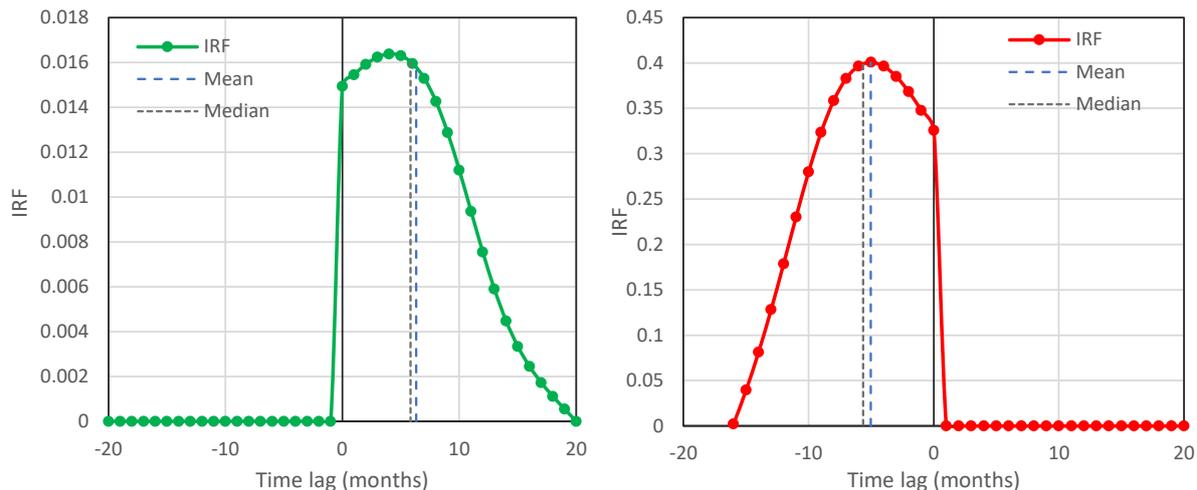
**Figure 18** IRFs for of South Oscillation Index (SOI) and temperature ($T$). (left) Case study #29 ($\Delta$SOI $\rightarrow$ $\Delta T$); (right) case study #30 ($\Delta T \rightarrow \Delta$SOI).

## 3 Discussion and conclusions

The theoretical methodology proposed in the companion paper (Koutsoyiannis et al., 2022) was applied to a range of case studies, both synthetic (artificial) and real world. Since the system dynamics in the artificial cases is known, this enables the method to be tested and validated. We showed that the method does not introduce any spurious potential causation when none is present in the system dynamics, except in cases of very high autocorrelation (see Section SI2.2 in Supplementary Information), which, nonetheless, is easily handled by studying the changes in the time series (differenced time series) instead of the original time series. Further, the requirement that the IRF coefficients be nonnegative is, on its own, sufficient to enable the correct direction of causality to be inferred. (Recall from the companion paper—Koutsoyiannis et al., 2022— that we do not subsume oscillatory nonlinear systems, in which the sign of $g(h)$ could alternate, under the causality notion, which accords with Cox's (1992) conditions for causality and in particular the monotonic relation of the cause with the effect.) When the roughness condition is added, this enables the correct system dynamics to be recovered.

We showed how to deal with the presence of nonlinearity (in which case a nonlinear transform of the potential cause should be carried out) and of long-term persistence in the time-series (which causes rising limbs in the IRF at the edges of its domain).

In addition to causality studies of synthetic examples, the theoretical framework was also applied to three real-world geophysical examples. In the first case, the causing of runoff by rainfall over a catchment is well established. The proposed framework indicates the correct direction of causation, even while suggesting the need for a nonlinear transformation and the need to extend the size of the domain of the IRF. The framework was further validated in its ability to detect the clear causal connection between Southern Oscillation Index and global mean annual temperature.

The remaining real-world case study led to an important side product of the current research. This is the surprising finding that, while in general the causal relationship of atmospheric $T$ and $CO_2$ concentration, as obtained by proxy data, appears to be of hen-or-egg type with principal direction $T \rightarrow [CO_2]$, in the recent decades the more accurate modern data support a conclusion that this principal direction has become exclusive. In other words, it is the increase of temperature that caused increased $CO_2$ concentration. Though this conclusion may sound counterintuitive at first glance, because it contradicts common perception (and for this reason we have assessed the case with an alternative parametric methodology in the Supplementary Information, section SI2.4, with results confirming those presented here), in fact it is reasonable. The temperature increase began at the end of the Little Ice Period, in the early 19th century, when human $CO_2$ emissions were negligible; hence other factors, such as the solar activity (measured by sunspot numbers), as well as internal long-range mechanisms of the complex climatic systems had to play their roles.

A possible physical mechanism for the $[CO_2]$ increase, as a result of temperature increase, was proposed by Koutsoyiannis and Kundzewicz (2020) and involves biochemical reactions, as, at higher temperatures, soil respiration, and hence natural $CO_2$ emissions, are increasing. In addition, as pointed out by Liu et al. (2017) the influence of El Niño on climate is accompanied by large changes to the carbon cycle, with the pantropical biosphere releasing much more carbon into the atmosphere during large El Niño occurrences. Noticeably, in a very recent paper, Goulet Coulombe and Göbel (2021) seem to confirm the finding by Koutsoyiannis and Kundzewicz (2020), yet they deem it an "*apparently counterintuitive finding that GMTA* [global mean surface temperature anomalies] *explains a larger portion of the forecast error variance of $CO_2$ than vice versa*". To "*resolve*" it, they "*explore a last avenue, that of using annual $CO_2$ emissions*". However, using anthropogenic $CO_2$ emissions, which contribute only a small portion (3.8%) to the global carbon cycle (Koutsoyiannis, 2021), as a principal variable is definitely less meaningful than using the atmospheric $CO_2$ concentration.

We believe that counterintuitive results, such as those about the causal link between temperature and $CO_2$ concentration conveyed in this paper, can indeed open up avenues of research. However, these avenues of research might not resolve the issue in a way compatible with what intuition dictates. In the history of science, such avenues were often created when established ideas were overturned by new findings.

By letting the geophysical records speak for themselves, with the help of our original methodology, we discovered a regularity that apparently contradicts common opinion. Our innovative findings should be given considerable attention as well as careful and critical scrutiny in the form of public discussion by the scientific community, which will undoubtedly improve understanding. If the methodology we proposed in the companion

paper (Koutsoyiannis et al., 2022) stands up to scrutiny, then our novel, high-impact results, i.e. those of cases #23 – #28 in the present paper, will have to be taken seriously and interpreted. Further research on the regularities of the causal behaviour of the climate system reported herein, being of considerable importance and relevance, is urgently needed.

**Conflicts of Interest:** We declare no conflict of interest.

**Data availability:** All data used in the paper are available online for free. The rainfall and runoff data were obtained from this U.S. Geological Survey database, publicly available at https://nwis.waterdata.usgs.gov/md/nwis/uv/?site_no=01603000&agency_cd=USGS (accessed June 2021). The temperature time series and the Mauna Loa $CO_2$ time series are readily available on monthly scale from http://climexp.knmi.nl (accessed June 2021). The palaeoclimatic data of Vostok $CO_2$ were retrieved from http://cdiac.ess-dive.lbl.gov/ftp/trends/co2/vostok.icecore.co2 (dated January 2003, accessed September 2018) and the temperature data from http://cdiac.ess-dive.lbl.gov/ftp/trends/temp/vostok/vostok.1999.temp.dat (dated January 2000, accessed September 2018). The Southern Oscillation Index (SOI) data are provided by National Centers for Environmental Information (NCEI) of the USA's National Oceanic and Atmospheric Administration (NOAA) https://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/.

# References

Arrhenius, S. 1896 On the influence of carbonic acid in the air upon the temperature of the ground. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **41**, 237–276, doi:10.1080/14786449608620846.

Beven, K., 2019 Validation and equifinality. In *Computer Simulation Validation Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, Beisbart, C. & Saam, N. J. (eds), 791-809, Springer, Cham.

Bode, H. W. & Shannon, C. E. 1950 A simplified derivation of linear least square smoothing and prediction theory. *Proc. of the IRE*, **38**(4), 417-425.

Caillon, N.; Severinghaus, J. P.; Jouzel, J.; Barnola, J. M.; Kang, J. & Lipenkov, V. Y. 2003 Timing of atmospheric $CO_2$ and Antarctic temperature changes across Termination III. *Science*, **299**, 1728–1731.

Christy, J. R., Norris, W. B., Spencer, R. W. & Hnilo, J. J. 2007 Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J. Geophys. Res.* **112**, D06102, doi:10.1029/2005JD006881.

Cox, D.R., 199. Causality: Some statistical aspects. *J. Roy. Stat. Soc. A* **155**(2), 291-301.

Goulet Coulombe, P. & Göbel, M. 2021 On spurious causality, CO2, and global temperature. *Econometrics* **9**(3), 33, https://doi.org/10.3390/econometrics9030033

Jouzel, J., Lorius, C., Petit, J. R., Genthon, C., Barkov, N. I., Kotlyakov, V. M. & Petrov, V. M. 1987 Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years). *Nature* **329**, 403–408.

Kaplan, A. 2011 Patterns and indices of climate variability. In: State of the Climate in 2010. *Bull. Amer. Meteor. Soc.* **92**(6), S20-S26.

Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Ekdahl, C. A., Guenther, P. R. & Waterman, L. S. 1976 Atmospheric carbon dioxide variations at Mauna Loa observatory. Hawaii. *Tellus* **28**, 538–551.

Koutsoyiannis, D. 2010 A random walk on water. *Hydrol. and Earth System Sci.* **14**, 585–601, doi:10.5194/hess-14-585-2010,.

Koutsoyiannis, D. 2014 Entropy: from thermodynamics to hydrology. *Entropy* **16** (3), 1287–1314, doi:10.3390/e16031287.

Koutsoyiannis, D. 2016 Generic and parsimonious stochastic modelling for hydrology and beyond. *Hydrol. Sci. J.* **61**(2), 225–244, doi:10.1080/02626667.2015.1016950.

Koutsoyiannis, D. 2017 Entropy production in stochastics. *Entropy* **19**(11), 581, doi:10.3390/e19110581.

Koutsoyiannis, D. 2019 Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes. *Hydrol. Sci. J.* **64** (9), 1013–1037, doi:10.1080/02626667.2019.1600700.

Koutsoyiannis, D. 2020a Simple stochastic simulation of time irreversible and reversible processes. *Hydrol. Sci. J.* **65**(4), 536–551, doi:10.1080/02626667.2019.1705302.

Koutsoyiannis, D. 2020b Revisiting the global hydrological cycle: is it intensifying? *Hydrol. and Earth System Sci.* **24**, 3899–3932, doi:10.5194/hess-24-3899-2020.

Koutsoyiannis, D. 2021 Rethinking climate, climate change, and their relationship with water. *Water* **13**(6), 849, doi:10.3390/w13060849.

Koutsoyiannis, D. & Kundzewicz, Z. W. 2020 Atmospheric temperature and $CO_2$: Hen-or-egg causality? *Sci* **2**(4), 83, doi:10.3390/sci2040083.

Koutsoyiannis, D. & Montanari, A. 2022 Bluecat: A local uncertainty estimator for deterministic simulations and predictions. *Water Resour. Res.,* 58(1), e2021WR031215, doi: 10.1029/2021WR031215.

Koutsoyiannis, D., Onof, C. Christofidis, A., & Kundzewicz, Z. W. 2022 Revisiting causality using stochastics: 1. Theory, *Proceedings of the Royal Society A*, this issue.

Kundzewicz, Z. W., Pińskwar, I. & Koutsoyiannis, D. 2020 Variability of global mean annual temperature is significantly influenced by the rhythm of ocean-atmosphere oscillations. *Science of the Total Environment* **747**, 141256, doi:10.1016/j.scitotenv.2020.141256.

Liang, X.S. 2022 The causal interaction between complex subsystems. *Entropy* 24(1), 3, doi: 10.3390/e24010003.

Liu, J., Bowman, K.W., Schimel, D.S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y. & O'Dell, C. W., 2017. Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science* **358**(6360), doi: 10.1126/science.aam5690.

McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. 2006 ENSO as an Integrating Concept in Earth Science. *Science* **314**, 1740-1745.

Montanari, A. & Koutsoyiannis, D. 2012 A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, 48, W09555, doi:10.1029/2011WR011412.

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M. et al. 1999 Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, **399**, 429–436.

Spencer, R. W. & Christy, J. R. 1990 Precise monitoring of global temperature trends from satellites. *Science* **247**, 1558–1562.

Young, P.C. 2011. *Recursive Estimation and Time Series Analysis*, Berlin, Heidelberg: Springer-Verlag.

Young, P.C., 2015. Refined instrumental variable estimation: Maximum likelihood optimization of a unified Box-Jenkins model. *Automatica*, **52**, 35–46.