Scientific Conference in Honour of Prof. Emer. Gerasimos A. Athanassoulis
WAVES, PROBABILITIES AND MEMORIES

Monday 4-5 July 2022
NTUA Library Multimedia Amphitheater

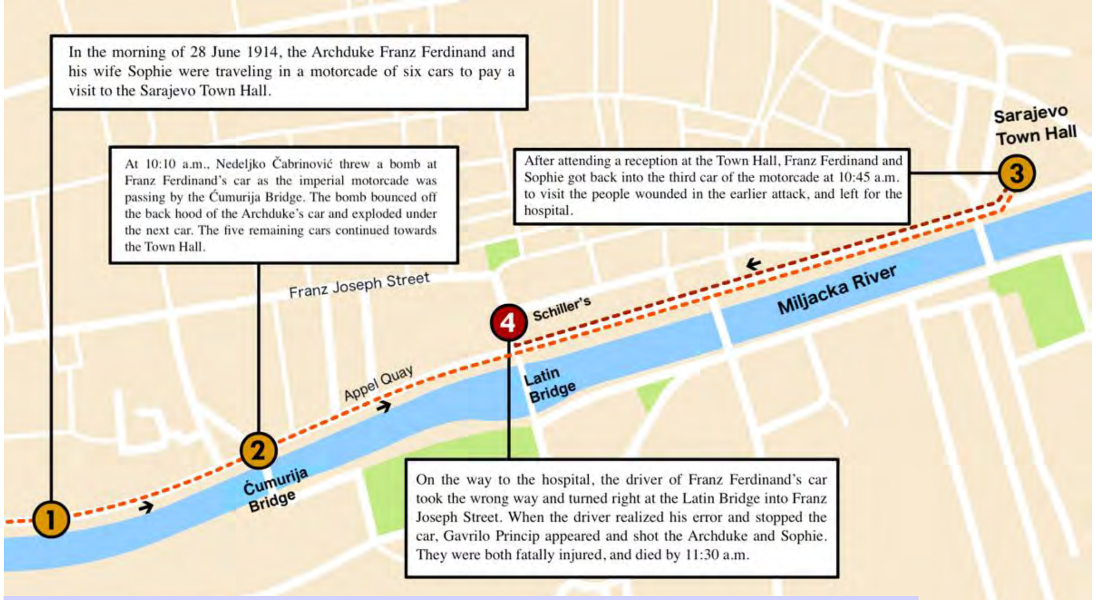
Revisiting causality using stochastics

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering
School of Civil Engineering, National Technical University of Athens
(dk@ntua.gr, http://itia.ntua.gr/dk/)

Available online: http://www.itia.ntua.gr/2216/

A historical example on causal (or acausal?) chains

- The assassination of the Austrian Archduke Franz Ferdinand led Austria-Hungary to declare war on Serbia and triggered World War I.
- The assassination is related to a mistake of Archduke’s driver (see map).
- Would the World War I break out if the driver did not do the mistake?
- Would the World War II occur if World War I did not?
- Would the current war (or special military operation) in Ukraine occur if World War II did not?
- Will war in Ukraine lead to World War III?



https://commons.wikimedia.org/wiki/File:Atentado_de_Sarajevo_en.png

Unanswered causality questions on another current drama

- What caused SARS-CoV-2?
  - A natural process in animals?
  - A laboratory leak of a constructed virus?
  - A planned political/military action?
- What does SARS-CoV-2 cause?
  - A variety of dangerous symptoms?
  - High or low mortality?
- What do the COVID vaccines cause?
  - Protection from COVID?
  - Less severe symptoms in case of COVID infection?
  - Decreased or increased risk for infection, hospitalization and death?
  - Side effects less or more severe than COVID?

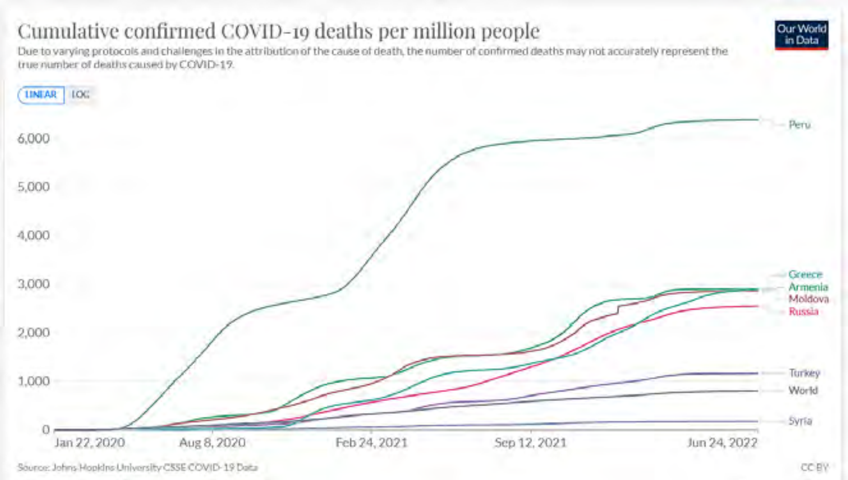
Difficulties in answering the questions

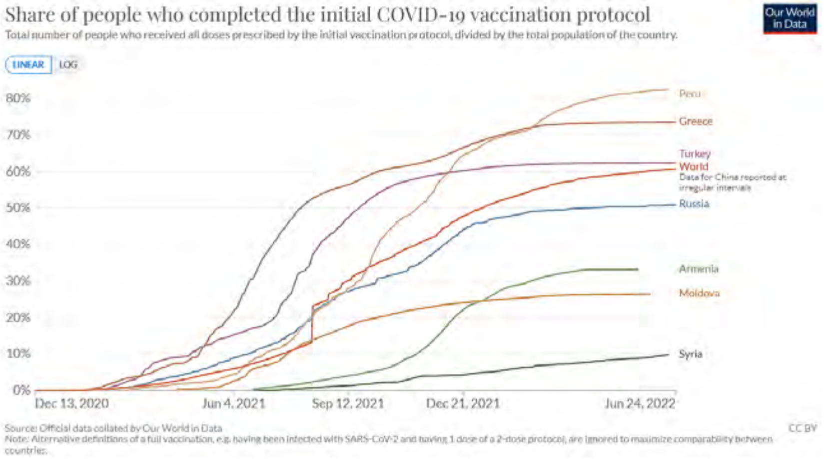
- The graphs show that some countries with larger percentage of vaccinated population against COVID have also larger percentage of COVID deaths.
- For legibility of the graph, data for only a few countries are shown.
- The impression is contrary to expectation.

See original graphs at:

https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&facet=none&uniformYAxis=0&hideControls=true&Metric=Confirmed+deaths&Interval=Cumulative&Relative+to+Population=true&Color+by+test+positivity=false&country=GRC~TUR~OWID_WRL~RUS~MDA~ARM~SYR~PER

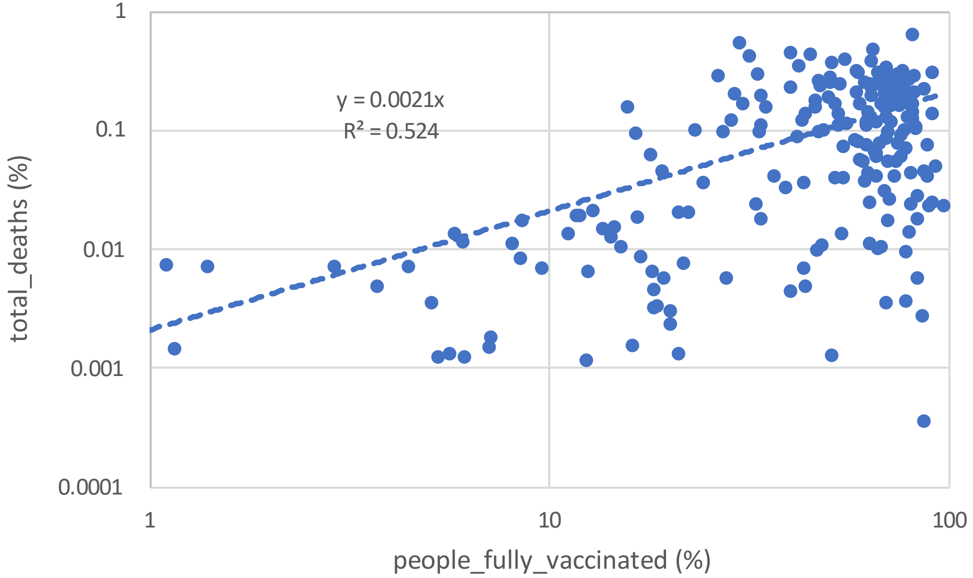
https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&facet=none&uniformYAxis=0&hideControls=true&Metric=People+fully+vaccinated&Interval=Cumulative&Relative+to+Population=true&Color+by+test+positivity=false&country=GRC~TUR~OWID_WRL~RUS~MDA~ARM~SYR~PER





Complete macroscopic picture of death vs. vaxx

- The graph shows all countries that have data in the period from 1 May to 24 June 2022 and population >= 1 M.
- Each point represents one country with the latest available data.
- Notice the positive correlation between the percentages of vaccinated population and COVID deaths.
- Possible interpretations (where "→" means "causes"):
  A. COVID vaccination → COVID death (difficult to support).
  B. COVID death → COVID vaccination. (more plausible as people are frightened by deaths and get vaccinated).
  C. No causality can be detected; data are spurious (the most plausible of the three).

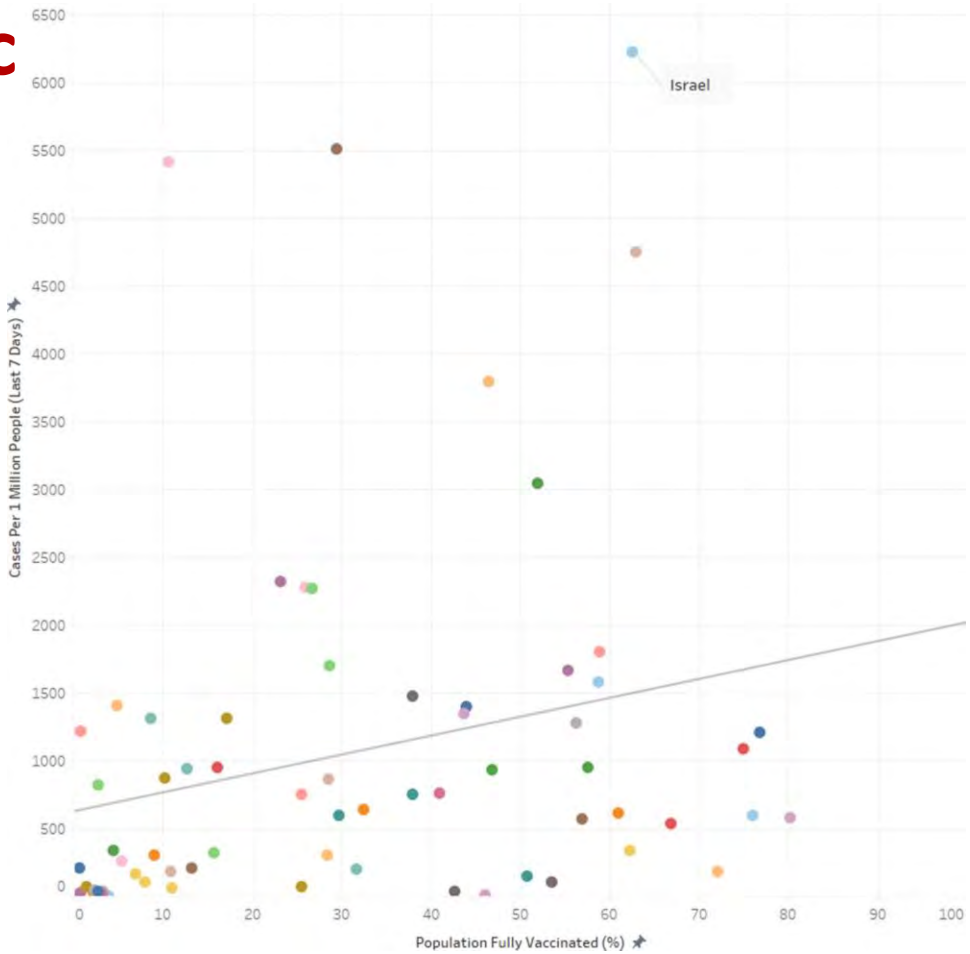


Data source: https://ourworldindata.org/explorers/coronavirus-data-explorer
(Click on DOWNLOAD and then on Full data (CSV) – All countries)

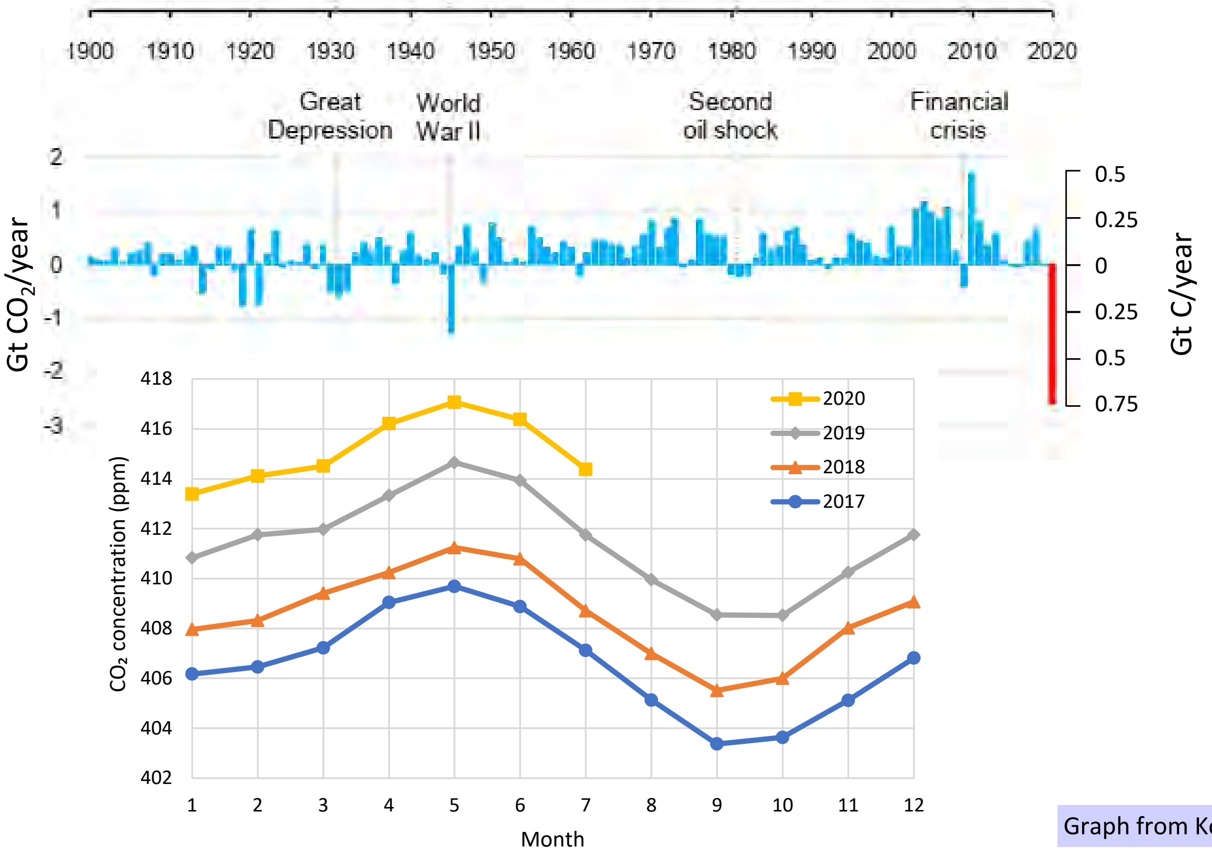
Additional macroscopic picture—infection vs. vaxx

- Positive correlation is also seen between the number of COVID cases vs. percentage of COVID vaccination.
- The graph shows the relationship between cases per 1 million people (last 7 days) and percentage of population fully vaccinated across 68 countries as of September 3, 2021
- It is from a peer-reviewed paper.

Subramanian, S.V. and Kumar, A., 2021.



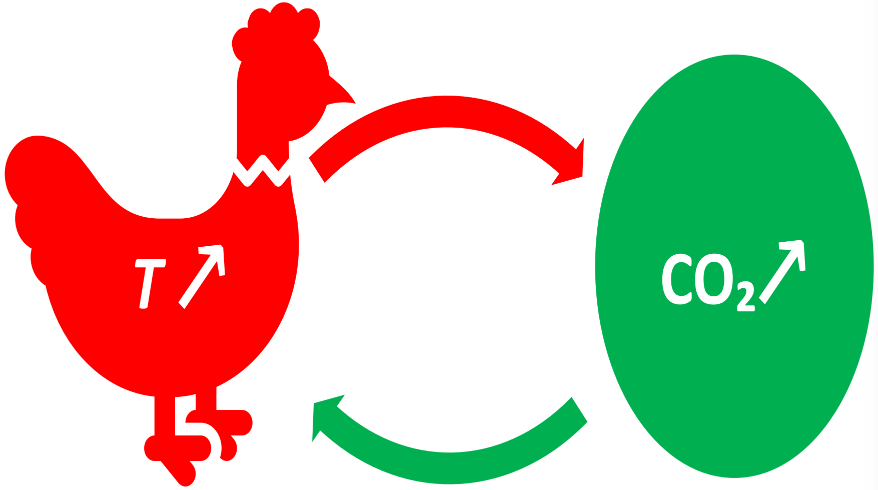
COVID and an unfortunate experiment

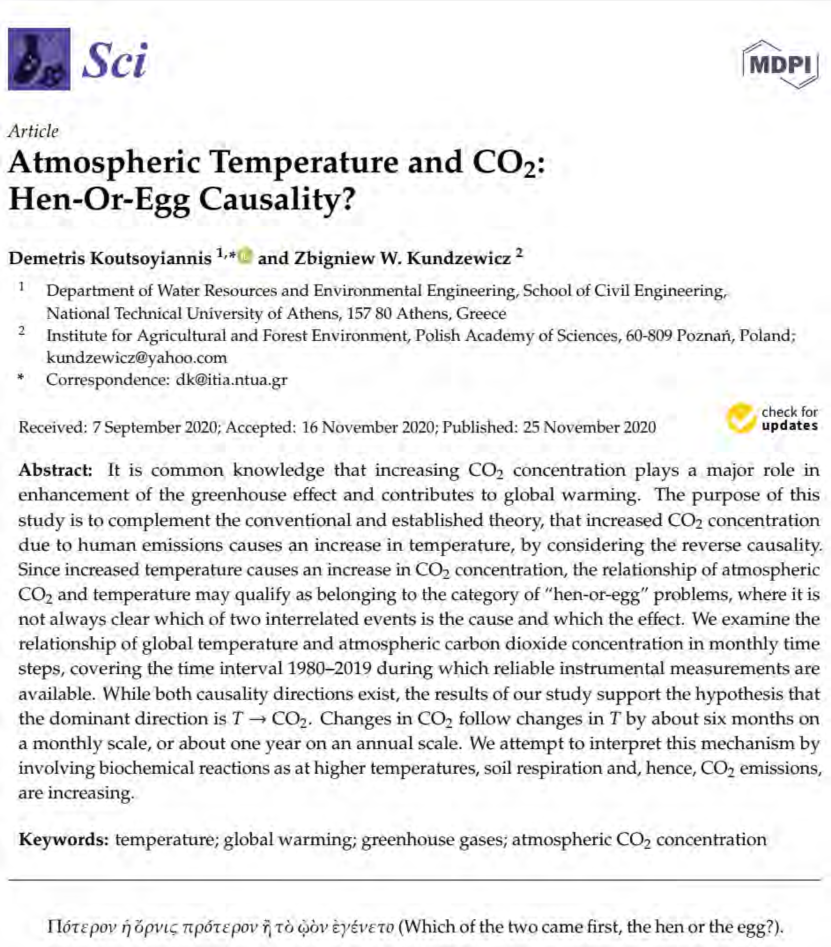


- COVID-caused lockdowns caused the greatest in history decrease of CO_2 emissions.
- The global CO_2 emissions were over 5% lower in the first quarter of 2020 than in that of 2019 (IEA, 2020).
- However, the increasing pattern of atmospheric CO2 concentration, as measured in Mauna Loa, did not change.

Graph from Koutsoyiannis and Kundzewicz (2020); see next page.

Causal relationship between CO_2 & temperature: “ὄρνις ἢ ᾠὸν;” (“hen or egg?”)





Sci

MDPI

Article

Atmospheric Temperature and CO_2: Hen-Or-Egg Causality?

Demetris Koutsoyiannis [1,*] and Zbigniew W. Kundzewicz [2]

1 Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, 157 80 Athens, Greece
2 Institute for Agricultural and Forest Environment, Polish Academy of Sciences, 60-809 Poznań, Poland; kundzewicz@yahoo.com
* Correspondence: dk@itia.ntua.gr

check for updates

Received: 7 September 2020; Accepted: 16 November 2020; Published: 25 November 2020

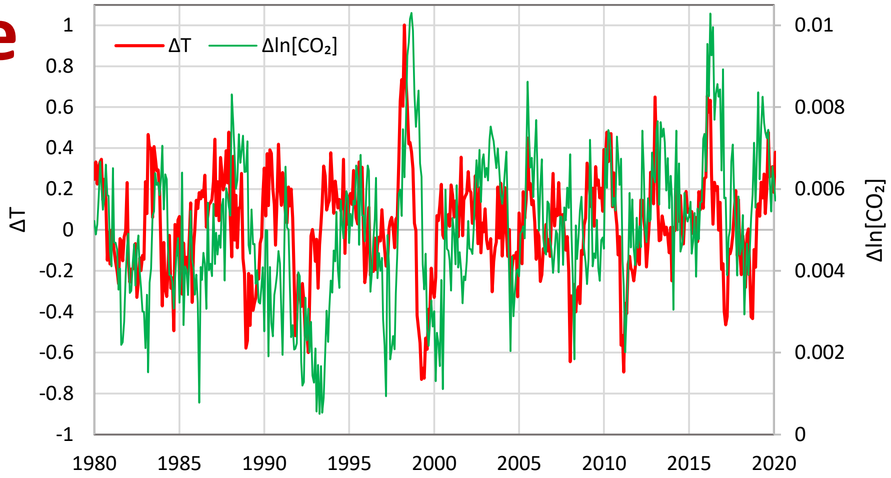
Abstract: It is common knowledge that increasing CO_2 concentration plays a major role in enhancement of the greenhouse effect and contributes to global warming. The purpose of this study is to complement the conventional and established theory, that increased CO_2 concentration due to human emissions causes an increase in temperature, by considering the reverse causality. Since increased temperature causes an increase in CO_2 concentration, the relationship of atmospheric CO_2 and temperature may qualify as belonging to the category of "hen-or-egg" problems, where it is not always clear which of two interrelated events is the cause and which the effect. We examine the relationship of global temperature and atmospheric carbon dioxide concentration in monthly time steps, covering the time interval 1980–2019 during which reliable instrumental measurements are available. While both causality directions exist, the results of our study support the hypothesis that the dominant direction is $T \rightarrow CO_2$. Changes in CO_2 follow changes in T by about six months on a monthly scale, or about one year on an annual scale. We attempt to interpret this mechanism by involving biochemical reactions as at higher temperatures, soil respiration and, hence, CO_2 emissions, are increasing.

Keywords: temperature; global warming; greenhouse gases; atmospheric CO_2 concentration

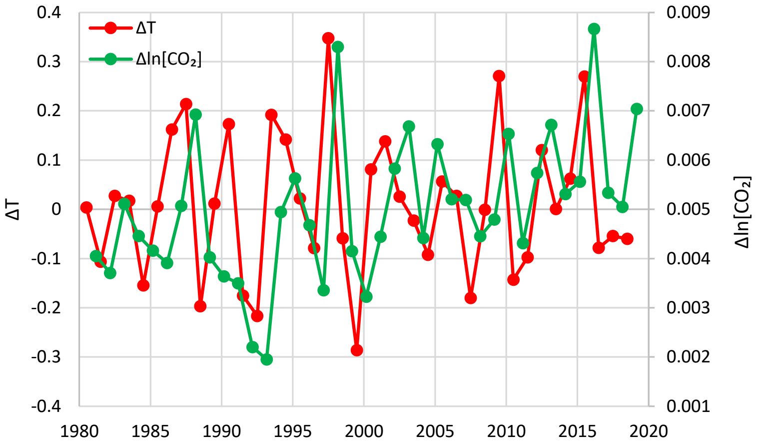
Πότερον ἡ ὄρνις πρότερον ἢ τὸ ᾠὸν ἐγένετο (Which of the two came first, the hen or the egg?).

Instrumental temperature and CO_2 data in search of causality

Differenced monthly time series of global temperature (UAH) and logarithm of CO_2 concentration (Mauna Loa)



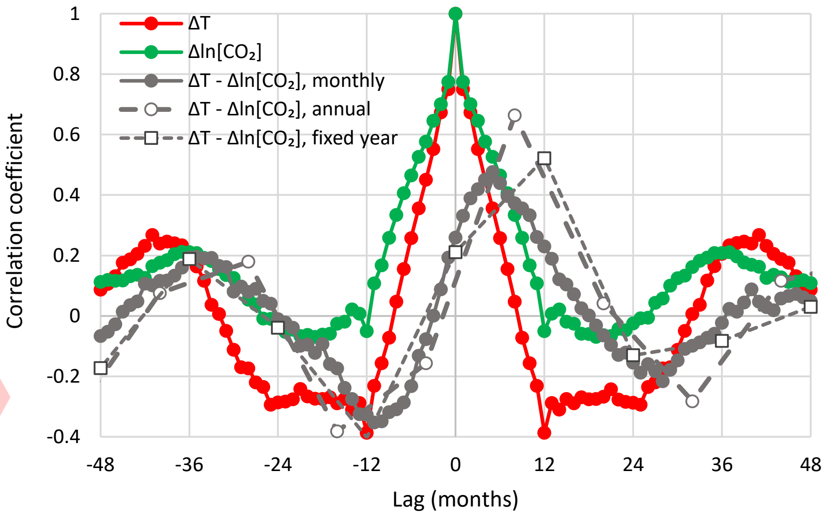
Annually averaged time series of differenced temperatures (UAH) and logarithm of CO_2 concentration (Mauna Loa). Each dot represents the average of a one-year duration ending at the time of its abscissa.



Which is the cause and which the effect?

Graphs from Koutsoyiannis and Kundzewicz (2020). Notice that logarithms of CO_2 concentration are used for linear equivalence with temperature. The differenced processes represent changes in the original processes.

Changes in CO_2 follow changes in global temperature



Auto- and cross-correlograms of the differenced time series of temperature (UAH) and logarithm of CO_2 concentration (Mauna Loa)

Which is the cause and which the effect?

Maximum cross-correlation coefficient (MCCC) and corresponding time lag in months						
	Monthly time series		Annual time series – sliding annual window		Annual time series – fixed annual window	
Temperature - CO_2 series	MCCC	Lag	MCCC	Lag	MCCC	Lag
UAH – Mauna Loa	0.47	5	0.66	8	0.52	12
UAH – Barrow	0.31	11	0.70	14	0.59	12
UAH – South Pole	0.37	6	0.54	10	0.38	12
UAH – Global	0.47	6	0.60	11	0.60	12
CRUTEM4 – Mauna Loa	0.31	5	0.55	10	0.52	12
CRUTEM4 – Global	0.33	9	0.55	12	0.55	12

Graph and table from Koutsoyiannis and Kundzewicz (2020).

Development and application of a theoretical framework

THE ROYAL SOCIETY PUBLISHING | All Journals

PROCEEDINGS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Research articles

Revisiting causality using stochastics: 1. Theory

Demetris Koutsoyiannis, Christian Onof, Antonis Christofides and Zbigniew W. Kundzewicz

Published: 25 May 2022 https://doi.org/10.1098/rspa.2021.0835

Review history

Abstract

Causality is a central concept in science, in philosophy and in life. However, reviewing various approaches to it over the entire knowledge tree, from philosophy to science and to scientific and technological applications, we locate several problems, which prevent these approaches from defining sufficient conditions for the existence of causal links. We thus choose to determine necessary conditions that are operationally useful in identifying or falsifying causality claims. Our proposed approach is based on stochastics, in which events are replaced by processes. Starting from the idea of stochastic causal systems, we extend it to the more general concept of hen-or-egg causality, which includes as special cases the classic causal, and the potentially causal and anti-causal systems. Theoretical considerations allow the development of an effective algorithm, applicable to large-scale open systems, which are neither controllable nor repeatable. The derivation and details of the algorithm are described in this paper, while in a companion paper we illustrate and showcase the proposed framework with a number of case studies, some of which are controlled synthetic examples and others real-world ones arising from interesting scientific problems.

THE ROYAL SOCIETY PUBLISHING | All Journals

PROCEEDINGS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Research articles

Revisiting causality using stochastics: 2. Applications

Demetris Koutsoyiannis, Christian Onof, Antonis Christofidis and Zbigniew W. Kundzewicz

Published: 25 May 2022 https://doi.org/10.1098/rspa.2021.0836

Review history

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 anti-causal 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. Others are real-world studies on interesting scientific problems in geophysics, and in particular hydrology and climatology.

Philosophical reflections—acknowledging difficulties

Aristotle (384 – 322 BC; Image source: Visconti, 1817):

> *that which when present is the cause of something, when absent we sometimes consider to be the cause of the contrary*.

Plutarch (AD 46 –119; Greek Middle Platonist philosopher):

> First posed the *hen or the egg* type of causality as a philosophical problem: *"Πότερον ἡ ὄρνις πρότερον ἢ τὸ ᾠὸν ἐγένετο"* (Πλούταρχος, Ηθικά, Συμποσιακὰ Β, Πρόβλημα Γ).

David Hume (1711– 1776; Scottish Enlightenment philosopher):

> the concept of a cause is merely a way we use to describe regularities.

Immanuel Kant (1724–1804, German Enlightenment philosopher):

(a) causality is understood in terms of rule-governedness;

(b) the temporal causal order is irreversible.

Theoretical probabilistic approaches to causality

Patrick Suppes (1922 –2014; American philosopher—Stanford Univ.)

Definition: *An event* $B_{t'}$ [occurring at time t'] *is a* prima facie *cause of the event* A_t [occurring at time t] *if and only if* (i) $t' < t$, (ii) $P(B_{t'}) > 0$, (iii) $P(A_t|B_{t'}) > P(A_t)$ Suppes (1970)

Our note: The definition is not very useful as it almost identifies causality with dependence: In fact it says that any two events that are neither synchronous nor independent establish a (*prima facie*) causal relationship.

David Cox (1924 –2022; British statistician—Oxford)

To the above three conditions of the **definition** he added a fourth: (iv) *there is no event* $C_{t''}$ *at time* $t'' < t' < t$ *such that* $P(A_t|B_{t'}C_{t''}) = P(A_t|\overline{B}_{t'}C_{t''})$. Cox (1992)

Our note: While this addition is certainly a theoretical advance, it is impractical: One cannot enumerate all events that happened before time t' and calculate their related conditional probabilities.

Applied probabilistic approaches to causality

Clive Granger (1934 – 2009; British-American econometrician—Univ. Nottingham and Univ. California, San Diego; Nobel in Economics, 2003)

Mostly known for the so-called "**Granger causality test**", based on the linear regression equation $\underline{y}_\tau = \sum_{j=1}^{\eta} a_j \underline{y}_{\tau-j} + \sum_{j=1}^{\eta} b_j \underline{x}_{\tau-j} + \underline{\varepsilon}_\tau$. If the coefficients b_j are nonzero, the interpretation is that the process $\underline{x}_\tau$ causes $\underline{y}_\tau$. Granger (1969)

Our notes: We find the framework problematic, both formally and logically:

- Formally testing hypotheses in geophysics can be inaccurate (by orders of magnitude) due to time dependence.
- The test is about prediction, which is fundamentally different from causality.

Judea Pearl (born 1936; Israeli-American computer scientist and philosopher)

He proposed a framework for causality combining probability with **graph theory**. Pearl (2009); Pearl et al. (2016)

Our notes: We find the framework problematic, both formally and logically:

- In using conditional probability, the chain rule is used inappropriately.
- It is based on the assumption that we already have a causal graph—a way of identifying causes.

Premises of the developed methodology

- Our framework is for **open systems** (in particular, **geophysical** systems), in which:
  - **External influences** cannot be controlled or excluded.
  - Only a **single realization** is possible.
  - There is **dependence** in time.
- Our framework is not formulated on the basis of events, but of **stochastic processes**.
- It is understood that only **necessary conditions of causality** can be investigated using stochastics. The usefulness of this objective lies in its ability:
  - to **falsify** an assumed causality and
  - to add statistical evidence, in an **inductive** context, for **potential causality** and its direction.
- The only “hard” requirement kept from previous studies is the **time precedence** of the cause from the effect.

Mathematical representation

- Any two stochastic processes $\underline{x}(t)$ and $\underline{y}(t)$ can be related by

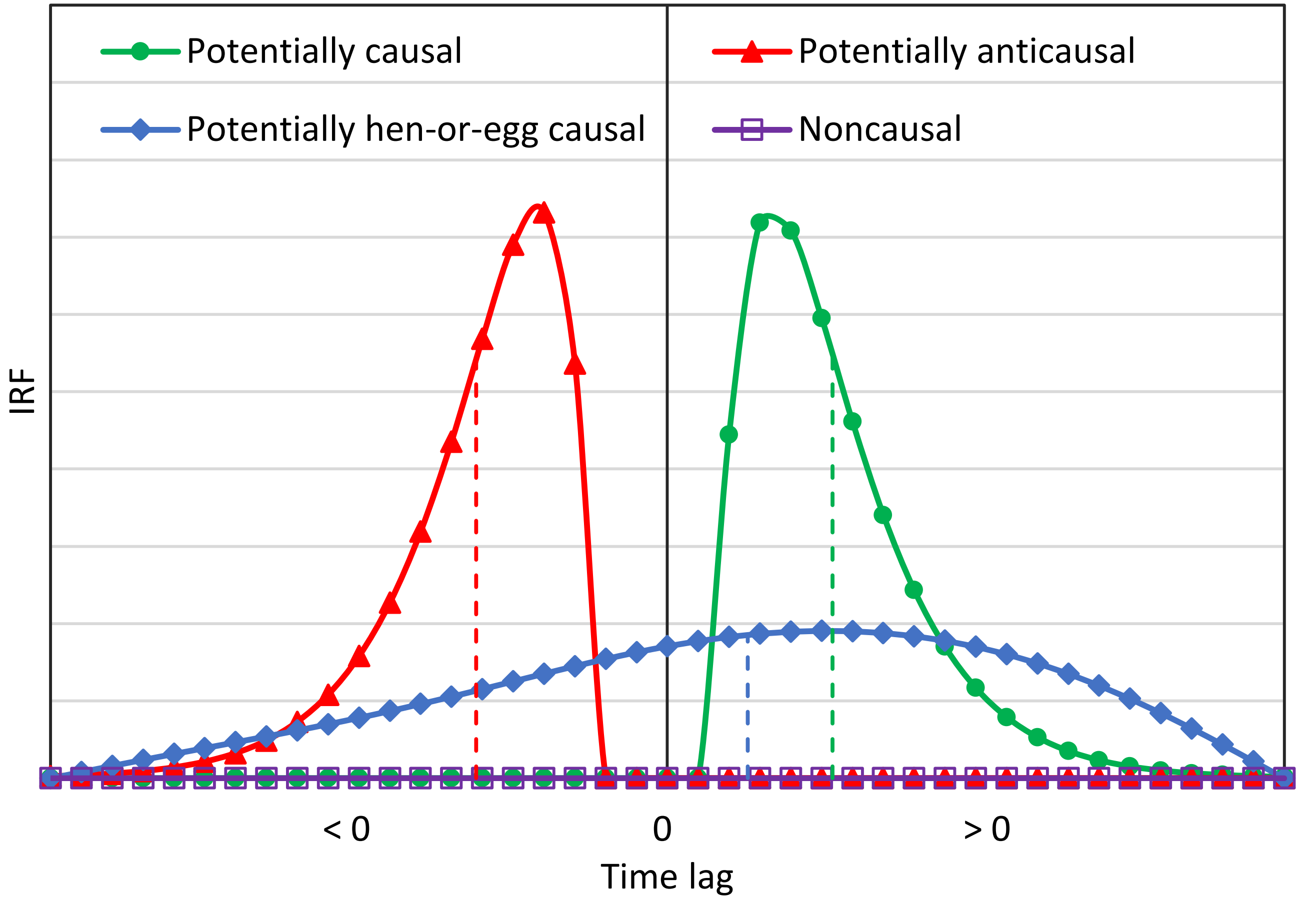
$$\underline{y}(t) = \int_{-\infty}^{\infty} g(h)\underline{x}(t-h)\mathrm{d}h + \underline{v}(t)$$

  where $g(h)$ is the **Impulse Response Function** (IRF) and $\underline{v}(t)$ is another process uncorrelated to $\underline{x}(t)$.
- There exist infinitely many pairs $(g(h), \underline{v}(t))$ of which we find the least squares solution—LSS: that resulting in the min $\mathrm{var}[\underline{v}(t)]$, or the max explained variance $e \coloneqq 1 - \mathrm{var}[\underline{v}(t)]/\mathrm{var}[\underline{y}(t)]$.
- Assuming that the LSS $g(h)$ has been determined, the system $(\underline{x}(t), \underline{y}(t))$ is:

  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.

- The framework of causality identification is constructed for case 3, with all other three cases resulting as special cases.

Illustration of the four different cases of potential causality



Additional mathematical considerations

- We also set additional desiderata for
  - (a) an **adequate time span** $\mathbb{h}$ of h (the **causal action is not instant**);
  - (b) a **nonnegative** $g(h) \geq 0$ for all $h \in \mathbb{h}$ (replacing $\underline{x}(t)$ with $-\underline{x}(t)$ for negative correlation);
  - (c) a **smooth** $g(h)$ assured by a constraint $E \leq E_0$, where E is determined in terms of the second derivative of $g(h)$ ($E \coloneqq \int_{-\infty}^{\infty} \left(g''(h)\right)^2 \mathrm{d}h$) and E_0 is a positive number.
- Although the theoretical framework is formulated in terms of **natural (continuous) time**, 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 results in

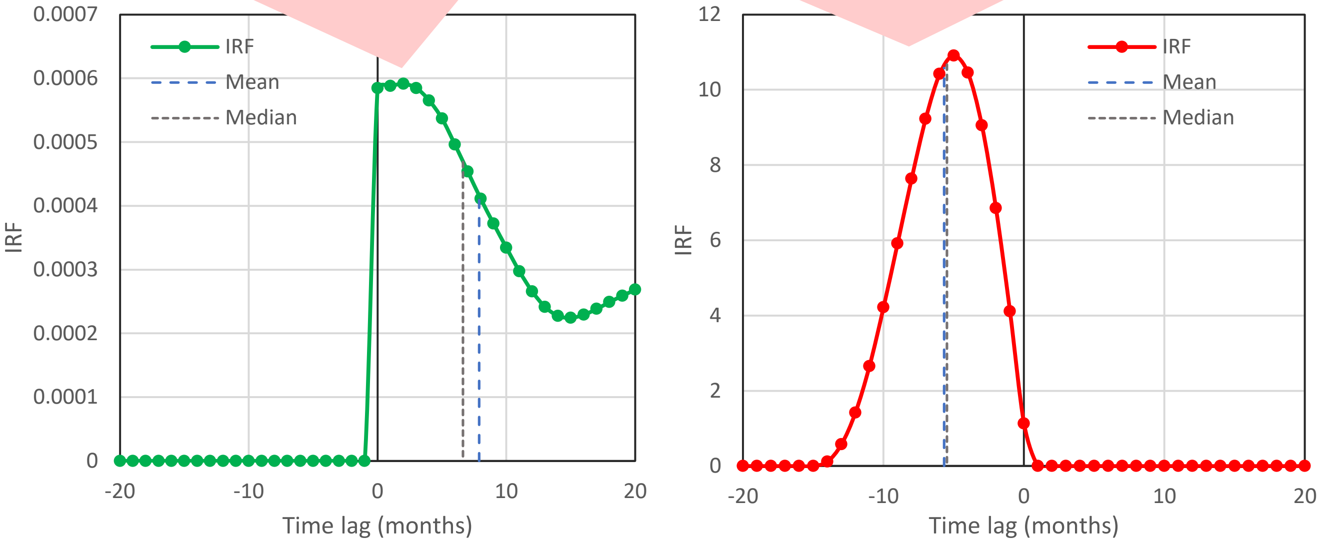
$$\underline{y}_\tau = \sum_{j=-\infty}^{\infty} g_j \underline{x}_{\tau-j} + \underline{v}_\tau$$

  where the sequence g_j can be determined accurately from the function $g(h)$.
- 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 $\pm\infty$ are replaced by $\pm J$, assuming that $g_j = 0$ for $|j| > J$, where, J should be chosen much lower than the length of the dataset.
- A **solver** can be used to resolve the **constrained optimization problem**: The determination of g_j is based on the minimization of $\mathrm{var}\left[\underline{v}(t)\right]$ subject to the constraints.

Application to the temperature – [CO_2] problem

Treating the system (T,[CO_2]) as potentially HOE causal, we conclude that it is potentially causal (mono-directional) with explained variance 31%

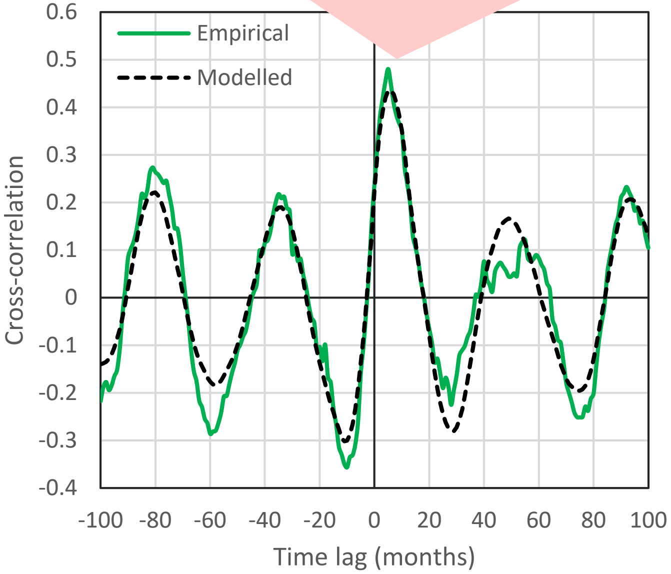
Treating the system ([CO_2], T) as potentially HOE causal, we conclude that it is potentially anticausal (counter-directional) with explained variance 23%



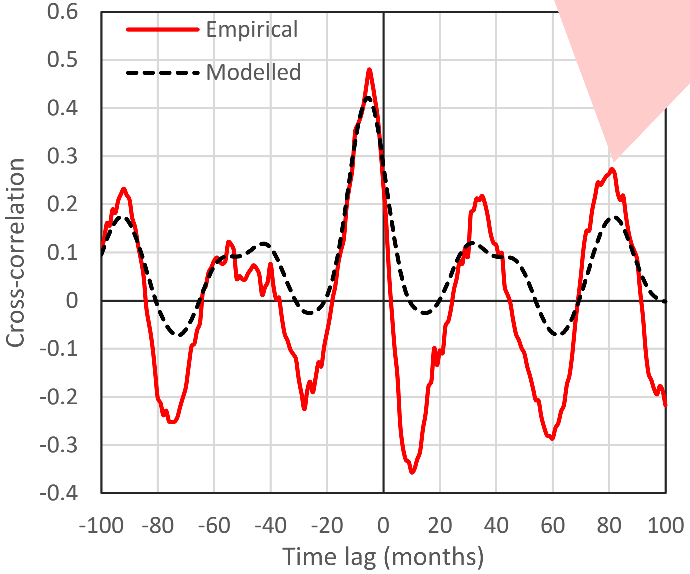
Conclusion: The common perception that increasing [CO_2] causes increased T **can be excluded** as it violates the necessary condition for this causality direction.
In contrast, the causality direction $T \rightarrow$ [CO_2] is plausible.

Additional evidence

Cross-correlation function of the causal system (T,[CO_2]) obtained from its IRF and the autocorrelation function of T.



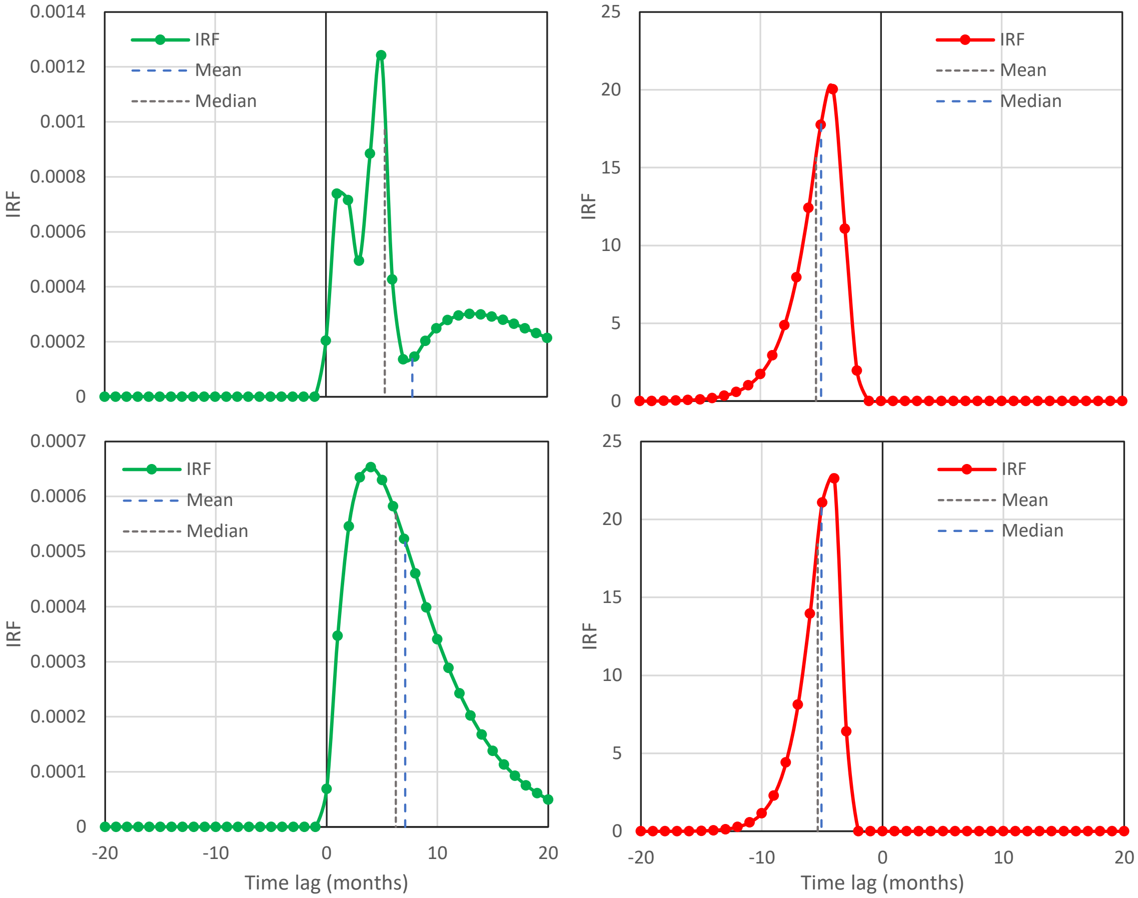
Cross-correlation function of the anticausal system ([CO_2], T) obtained from its IRF and the autocorrelation function of [CO_2].



Conclusion: The causal system (T,[CO_2]) is more consistent to reality than the anticausal system ([CO_2], T). This adds evidence that the actual causality direction is $T \rightarrow [CO_2]$.

More additional evidence

- For those fearing that our algorithm may produce incorrect results, a different algorithm was additionally used, whose results are shown in the graphs on the right.
- Namely a parametric IRF was constructed based on alpha basis functions (4 in upper graph, just one in lower graph).
- These results confirm that (T, [CO_2]) is potentially causal and ([CO_2], T) potentially anticausal.
- This adds evidence that the causality direction is $T \rightarrow [CO_2]$.



Conclusions

- Causality is a central concept in **science**, **philosophy** and **life**, with very high **economic** importance.
- Recently causal inference has become an **arena of enormous interest**.
- Yet our review of various approaches to causality over the entire knowledge tree, from philosophy to science and to scientific, technological and socio-political application, locates **major problems that are unsolved**.
- Our method posits a more modest objective: To determine **necessary** (not sufficient) **conditions** that are operationally useful in identifying or falsifying causality claims.
- It also replaces events with **stochastic processes**.
- Application of the method to climate shows that **it is the increase of temperature that caused increased CO_2 concentration**, despite the common perception for the opposite causality direction.

References

- Cox, D.R., 1992. Causality: Some statistical aspects. *J. Roy. Stat. Soc. A* 155(2), 291-301.
- Granger, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**(3), 424-438.
- Koutsoyiannis, D., and Kundzewicz, Z.W., 2020. Atmospheric temperature and CO_2: Hen-or-egg causality?, *Sci*, 2 (4), 83, doi: 10.3390/sci2040083.
- Koutsoyiannis, D., Onof, C., Christofides, A., and Kundzewicz, Z.W., 2022a. Revisiting causality using stochastics: 1.Theory, *Proceedings of The Royal Society A*, 478 (2261), 20210835, doi: 10.1098/rspa.2021.0835.
- Koutsoyiannis, D., Onof, C., Christofides, A., and Kundzewicz, Z.W., 2022b. Revisiting causality using stochastics: 2. Applications, *Proceedings of The Royal Society A*, 478 (2261), 20210836, doi: 10.1098/rspa.2021.0836.
- Pearl, J. 2009. Causal inference in statistics: An overview. *Statistics Surveys* 3, 96-146, doi: 10.1214/09-SS057.
- Pearl, J., Glymour, M. & Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer*. Chichester, UK: Wiley.
- Subramanian, S.V. and Kumar, A., 2021. Increases in COVID-19 are unrelated to levels of vaccination across 68 countries and 2947 counties in the United States. European Journal of Epidemiology, 36 (12), 1237-1240, doi: 10.1007/s10654-021-00808-7.
- Suppes, P. 1970 *A Probabilistic Theory of Causality.* Amsterdam, The Netherlands: North-Holland Publishing.