THE ROYAL SOCIETY
PUBLISHING

PROCEEDINGS A

Revisiting causality using stochastics: 1. Theory

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

Article citation details
Proc. R. Soc. A **478**: 20210835.
http://dx.doi.org/10.1098/rspa.2021.0835

Review timeline

Original submission:	3 November 2021
1st revised submission:	16 February 2022
2nd revised submission:	20 April 2022
Final acceptance:	22 April 2022

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPA-2021-0835.R0 (Original submission)

Review form: Referee 1

Is the manuscript an original and important contribution to its field?
Good

Is the paper of sufficient general interest?
Good

Is the overall quality of the paper suitable?
Acceptable

Can the paper be shortened without overall detriment to the main message?
Yes

Do you think some of the material would be more appropriate as an electronic appendix?
No

Do you have any ethical concerns with this paper?
No

Recommendation?
Major revision is needed (please make suggestions in comments)

Comments to the Author(s)
Comments on Koutsoyiannis et al.: Revisiting causality using stochastics: 1. Theory
RSPA-2021-0835

This paper gives a brief summary of the history of the concept of causality before going on to provide an approach to identifying sufficient conditions for causal relationships. That the summary is rather brief and selective is understandable in a research paper concerned more with technique than philosophy but there is quite a lot that does not get mentioned, (eg. Wesley Salmon, David Cox, the most recent edition of Mario Bunge's book, …) and some points that could be contested (e.g. the gross oversimplification of the ideas of Russell on p4.L9). The presentation of this section could also certainly be improved – for example the wording of the mention of Granger (1980) p6,L37-58 is really an aside that does not contribute to the argument, albeit that the later discussion of Granger causality as correlation is useful to repeat.

But concentrating on the technical aspects. The authors focus on the characteristics of temporal asymmetry and irreversibility and suggest that these can be demonstrated empirically for two variables for which data are available by the fitting of a linear impulse-response function (IRF) allowing for stochastic variability under a least squares "optimal" solution (see also comments on Part 2). This approach has some nice features, such as the possibility of identifying anti-causal and hen or egg relationships but also seems to have some important limitations that are rather glossed over in the paper.

The first of these is the linearity assumption (or the assumption that a simple transformation of variables can linearise the relationship) and that g(h) is continuous and always has the same sign. For real world open systems that would seem to be problematic – it is exactly why people build complex models to represent real world systems with causalities built in (though the authors are right to point out that such models cannot be used to test for causality as they intrinsically are built on causal chains). Of course, as shown in the Supplementary Information, we can linearise as an approximation or try differencing or different transforms first, but this will not necessarily result in stochastic terms that can be minimised by least squares as seems to be assumed in both justifications for linearisation in SI2.

This then also creates doubts about accepting only the least squares solution for the IRF, especially when the IRF may be really rather uncertain given noisy real world data (as the authors recognise in their discussion of autocorrelation and cross-correlation which will increase that uncertainty). It is well known that identifying the ordinates of a discrete IRF is an over-parameterised problem in this respect (there is a whole literature on the identification of transfer functions that is not really mentioned here – see below), so the results might depend on the identification method chosen. Hence the imposition of a roughness threshold, and the requirement of positive ordinates (though this would appear to rule out any consideration of oscillatory responses in a nonlinear system).

Finally, the simple two variable analysis of the IRF approach would appear to allow for the possibility of "spurious" causality, where changes in both variables are effected by some unmeasured cause but with different time delays and response functions. The authors discuss this in respect of a simple case of spatial proximity, but there may be cases of spatially extensive observed only at points for some variables where this might be a real issue.

Which ultimately makes me wonder whether this type of simple linear(ised) analysis is really of value in inferring causality. Yes, the approach does allow for the recognition of ambiguity and the authors are quite clear that they are only assessing sufficient rather than necessary conditions but that only then means it is up to the analyst to determine whether the result is spurious or not. In which case, will it only provide a clear inference in obvious cases? Something here, for all the philosophy and maths, does not feel quite right but should perhaps await a consideration of the examples in Part 2.

Other comments

P16. It would be useful to illustrate these types of IRF here.

P22. Although this is an analytical solution it is known to be notoriously sensitive for noisy and correlated data because of the number of effective parameters being identified (relative to the identification of a low parameter functional form). Thus the roughness condition will be very important, but will interact with the form of the output and its uncertainty (as evident for example, rather dramatically in Figure 3 or Part 2). Does this not imply that inference will be somewhat dependent on the roughness condition applied?

P23 L37ff. It is not clear how this differs from past work on using transfer functions for the prediction of bivariate and multivariate stochastic variables. The authors refer only to their own past work here with a suggestion that future work is required in future, but works such as Young (1984, 2011, 2019) is not mentioned. This also deals with the direct estimation of continuous time functions from which differential equations can be inferred (also Young, 2015).

Young, P C, 1984, 2011, Recursive estimation and time series analysis, Springer.

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

Supplementary Information

P11. Define LRD

P12. Why not show the IRF?

Review form: Referee 2

Is the manuscript an original and important contribution to its field?
Acceptable

Is the paper of sufficient general interest?
Good

Is the overall quality of the paper suitable?
Acceptable

Can the paper be shortened without overall detriment to the main message?
No

Do you think some of the material would be more appropriate as an electronic appendix?
No

Do you have any ethical concerns with this paper?
No

Recommendation?
Major revision is needed (please make suggestions in comments)

Comments to the Author(s)
Please see Appendix A.

Decision letter (RSPA-2021-0835.R0)

28-Jan-2022

Dear Professor Koutsoyiannis

The Editor of Proceedings A has now received comments from referees on the above paper and would like you to revise it in accordance with their suggestions which can be found below (not including confidential reports to the Editor).

Please submit a copy of your revised paper within four weeks - if we do not hear from you within this time then it will be assumed that the paper has been withdrawn. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office in advance.

Please note that it is the editorial policy of Proceedings A to offer authors one round of revision in which to address changes requested by referees. If the revisions are not considered satisfactory by the Editor, then the paper will be rejected, and not considered further for publication by the journal. In the event that the author chooses not to address a referee's comments, and no scientific justification is included in their cover letter for this omission, it is at the discretion of the Editor whether to continue considering the manuscript.

To revise your manuscript, log into http://mc.manuscriptcentral.com/prsa and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees" in Step 1: "View and Respond to Decision Letter". Please provide a point-by-point response to the comments raised by the reviewers and the editor(s). A thorough response to these points will help us to assess your revision quickly. You can also upload a 'tracked changes' version either as part of the 'Response to reviews' or as a 'Main document'.

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any unnecessary previous files before uploading your revised version.

When revising your paper please ensure that it remains under 28 pages long. In addition, any pages over 20 will be subject to a charge (£150 + VAT (where applicable) per page). Your paper has been ESTIMATED to be 21 pages.

Open Access

You are invited to opt for open access, our author pays publishing model. Payment of open access fees will enable your article to be made freely available via the Royal Society website as soon as it is ready for publication. For more information about open access please visit https://royalsociety.org/journals/authors/open-access/. The open access fee for this journal is £1700/$2380/€2040 per article. VAT will be charged where applicable. Please note that if the corresponding author is at an institution that is part of a Read and Publishing deal you are required to select this option. See https://royalsociety.org/journals/librarians/purchasing/read-and-publish/read-publish-agreements/ for further details.

Once again, thank you for submitting your manuscript to Proc. R. Soc. A and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Yours sincerely
Raminder Shergill
proceedingsa@royalsociety.org

on behalf of
Professor Graham Hughes
Board Member
Proceedings A

Reviewer(s)' Comments to Author:
Referee: 1
Comments to the Author(s)
Comments on Koutsoyiannis et al.: Revisiting causality using stochastics: 1. Theory
RSPA-2021-0835

This paper gives a brief summary of the history of the concept of causality before going on to provide an approach to identifying sufficient conditions for causal relationships. That the summary is rather brief and selective is understandable in a research paper concerned more with technique than philosophy but there is quite a lot that does not get mentioned, (eg. Wesley Salmon, David Cox, the most recent edition of Mario Bunge's book, …) and some points that could be contested (e.g. the gross oversimplification of the ideas of Russell on p4.L9). The presentation of this section could also certainly be improved – for example the wording of the mention of Granger (1980) p6,L37-58 is really an aside that does not contribute to the argument, albeit that the later discussion of Granger causality as correlation is useful to repeat.

But concentrating on the technical aspects. The authors focus on the characteristics of temporal asymmetry and irreversibility and suggest that these can be demonstrated empirically for two variables for which data are available by the fitting of a linear impulse-response function (IRF) allowing for stochastic variability under a least squares "optimal" solution (see also comments on Part 2). This approach has some nice features, such as the possibility of identifying anti-causal and hen or egg relationships but also seems to have some important limitations that are rather glossed over in the paper.

The first of these is the linearity assumption (or the assumption that a simple transformation of variables can linearise the relationship) and that g(h) is continuous and always has the same sign.

For real world open systems that would seem to be problematic – it is exactly why people build complex models to represent real world systems with causalities built in (though the authors are right to point out that such models cannot be used to test for causality as they intrinsically are built on causal chains). Of course, as shown in the Supplementary Information, we can linearise as an approximation or try differencing or different transforms first, but this will not necessarily result in stochastic terms that can be minimised by least squares as seems to be assumed in both justifications for linearisation in SI2.

This then also creates doubts about accepting only the least squares solution for the IRF, especially when the IRF may be really rather uncertain given noisy real world data (as the authors recognise in their discussion of autocorrelation and cross-correlation which will increase that uncertainty). It is well known that identifying the ordinates of a discrete IRF is an over-parameterised problem in this respect (there is a whole literature on the identification of transfer functions that is not really mentioned here – see below), so the results might depend on the identification method chosen. Hence the imposition of a roughness threshold, and the requirement of positive ordinates (though this would appear to rule out any consideration of oscillatory responses in a nonlinear system).

Finally, the simple two variable analysis of the IRF approach would appear to allow for the possibility of "spurious" causality, where changes in both variables are effected by some unmeasured cause but with different time delays and response functions. The authors discuss this in respect of a simple case of spatial proximity, but there may be cases of spatially extensive observed only at points for some variables where this might be a real issue.

Which ultimately makes me wonder whether this type of simple linear(ised) analysis is really of value in inferring causality. Yes, the approach does allow for the recognition of ambiguity and the authors are quite clear that they are only assessing sufficient rather than necessary conditions but that only then means it is up to the analyst to determine whether the result is spurious or not. In which case, will it only provide a clear inference in obvious cases? Something here, for all the philosophy and maths, does not feel quite right but should perhaps await a consideration of the examples in Part 2.

Other comments

P16. It would be useful to illustrate these types of IRF here.

P22. Although this is an analytical solution it is known to be notoriously sensitive for noisy and correlated data because of the number of effective parameters being identified (relative to the identification of a low parameter functional form). Thus the roughness condition will be very important, but will interact with the form of the output and its uncertainty (as evident for example, rather dramatically in Figure 3 or Part 2). Does this not imply that inference will be somewhat dependent on the roughness condition applied?

P23 L37ff. It is not clear how this differs from past work on using transfer functions for the prediction of bivariate and multivariate stochastic variables. The authors refer only to their own past work here with a suggestion that future work is required in future, but works such as Young (1984, 2011, 2019) is not mentioned. This also deals with the direct estimation of continuous time functions from which differential equations can be inferred (also Young, 2015).

Young, P C, 1984, 2011, Recursive estimation and time series analysis, Springer.

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

Supplementary Information

P11. Define LRD

P12. Why not show the IRF?

Referee: 2
Comments to the Author(s)
Please see Appendix A.

Board Member:
Comments to Author(s):
The reviewers find this and its companion manuscript make an interesting contribution to the subject and have returned a number of questions and constructive comments for your consideration. Please address these points during the process of revising your manuscript(s).

Author's Response to Decision Letter for (RSPA-2021-0835.R0)

See Appendix B.

RSPA-2021-0835.R1 (Revision)

Review form: Referee 1

Is the manuscript an original and important contribution to its field?
Excellent

Is the paper of sufficient general interest?
Excellent

Is the overall quality of the paper suitable?
Excellent

Can the paper be shortened without overall detriment to the main message?
Yes

Do you think some of the material would be more appropriate as an electronic appendix?
No

Do you have any ethical concerns with this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
P5 L50. Should be t”. not (At)”

P6 L9. (iii’) better? Also (iv’) at L26 and L32

P 6L14 (font problem) “spurious causality”

P16 L4. Delete a,b also P25L18 – do not need to separately reference supplementary information

P19L36 not the latter

P24L23 delete genuine (implies it is an actual causality rather than just assumed as correct)

Review form: Referee 2

Is the manuscript an original and important contribution to its field?
Good

Is the paper of sufficient general interest?
Good

Is the overall quality of the paper suitable?
Acceptable

Can the paper be shortened without overall detriment to the main message?
Yes

Are there details of how to obtain materials and data, including any restrictions that may apply?

Do you have any ethical concerns with this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
Please see Appendix C.

Decision letter (RSPA-2021-0835.R1)

12-Apr-2022

Dear Professor Koutsoyiannis,

On behalf of the Editor, I am pleased to inform you that your Manuscript RSPA-2021-0835.R1 entitled "Revisiting causality using stochastics: 1. Theory" has been accepted for publication subject to minor revisions in Proceedings A. Please find the referees' comments below.

The reviewer(s) have recommended publication, but also suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the reviewer(s)' comments and revise your manuscript. Please note that we have a strict upper limit of 28 pages for each paper. Please endeavour to incorporate any revisions while keeping the paper within journal limits. Please note that page charges are made on all papers longer than 20 pages. If you cannot pay these charges you must reduce your paper to 20 pages before submitting your revision. Your paper has been ESTIMATED to be 24 pages. We cannot proceed with typesetting your paper without your agreement to meet page charges in full should the paper exceed 20 pages when typeset. If you have any questions, please do get in touch.

It is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let me know in advance of the due date.

To revise your manuscript, log into https://mc.manuscriptcentral.com/prsa and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees" in Step 1: "View and Respond to Decision Letter". Please provide a point-by-point response to the comments raised by the reviewers and the editor(s). A thorough response to these points will help us to assess your revision quickly. You can also upload a ‘tracked changes’ version either as part of the ‘Response to reviews’ or as a ‘Main document’.

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission process.

When uploading your revised files, please make sure that you include the following as we cannot proceed without these:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tif, eps or print-quality pdf preferred). The format should be produced directly from original creation package, or original software format.

3) Electronic Supplementary Material (ESM): all supplementary materials accompanying an accepted article will be treated as in their final form. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details where possible (authors, article title, journal name). Supplementary files will be published alongside the paper on the journal website and posted on the online figshare repository (https://figshare.com). The heading and legend provided for each supplementary file during the submission process will be used to create the figshare page, so please ensure these are accurate and informative so that your files can be found in searches. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Alternatively you may upload a zip folder containing all source files for your manuscript as described above with a PDF as your "Main Document". This should be the full paper as it appears when compiled from the individual files supplied in the zip folder.

Article Funder

Please ensure you fill in the Article Funder question on page 2 to ensure the correct data is collected for FundRef (http://www.crossref.org/fundref/).

Media summary

Please ensure you include a short non-technical summary (up to 100 words) of the key findings/importance of your paper. This will be used for to promote your work and marketing purposes (e.g. press releases). The summary should be prepared using the following guidelines:

*Write simple English: this is intended for the general public. Please explain any essential technical terms in a short and simple manner.
*Describe (a) the study (b) its key findings and (c) its implications.
*State why this work is newsworthy, be concise and do not overstate (true 'breakthroughs' are a rarity).
*Ensure that you include valid contact details for the lead author (institutional address, email address, telephone number).

Cover images

We welcome submissions of images for possible use on the cover of Proceedings A. Images should be square in dimension and please ensure that you obtain all relevant copyright permissions before submitting the image to us. If you would like to submit an image for consideration please send your image to proceedingsa@royalsociety.org

Open Access

You are invited to opt for open access, our author pays publishing model. Payment of open access fees will enable your article to be made freely available via the Royal Society website as soon as it is ready for publication. For more information about open access please visit https://royalsociety.org/journals/authors/open-access/. The open access fee for this journal is £1700/$2380/€2040 per article. VAT will be charged where applicable. Please note that if the corresponding author is at an institution that is part of a Read and Publishing deal you are required to select this option. See https://royalsociety.org/journals/librarians/purchasing/read-and-publish/read-publish-agreements/ for further details.

Once again, thank you for submitting your manuscript to Proceedings A and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes
Raminder Shergill
proceedingsa@royalsociety.org
Proceedings A

on behalf of
Professor Graham Hughes
Board Member
Proceedings A

Reviewer(s)' Comments to Author:
Referee: 1
Comments to the Author(s)
P5 L50. Should be t". not (At)"

P6 L9. (iii') better? Also (iv') at L26 and L32

P 6L14 (font problem) "spurious causality"

P16 L4. Delete a,b also P25L18 – do not need to separately reference supplementary information

P19L36 not the latter

P24L23 delete genuine (implies it is an actual causality rather than just assumed as correct)

Referee: 2
Comments to the Author(s)
Please see Appendix C.

Board Member
Comments to Author(s):
Thank you for the revisions of your manuscript. Both reviewers are supportive of publication with some minor changes - please adapt your final version accordingly.

Author's Response to Decision Letter for (RSPA-2021-0835.R1)

See Appendix D.

Decision letter (RSPA-2021-0835.R2)

22-Apr-2022

Dear Professor Koutsoyiannis

I am pleased to inform you that your manuscript entitled "Revisiting causality using stochastics: 1. Theory" has been accepted in its final form for publication in Proceedings A.

Our Production Office will be in contact with you in due course. You can expect to receive a proof of your article soon. Please contact the office to let us know if you are likely to be away from e-mail in the near future. If you do not notify us and comments are not received within 5 days of sending the proof, we may publish the paper as it stands.

As a reminder, you have provided the following 'Data accessibility statement' (if applicable). Please remember to make any data sets live prior to publication, and update any links as needed when you receive a proof to check. It is good practice to also add data sets to your reference list.
Statement (if applicable):

Open access
You are invited to opt for open access, our author pays publishing model. Payment of open access fees will enable your article to be made freely available via the Royal Society website as soon as it is ready for publication. For more information about open access please visit https://royalsociety.org/journals/authors/which-journal/open-access/. The open access fee for this journal is £1700/$2380/€2040 per article. VAT will be charged where applicable.

Note that if you have opted for open access then payment will be required before the article is published – payment instructions will follow shortly.

If you wish to opt for open access then please inform the editorial office (proceedingsa@royalsociety.org) as soon as possible.

Your article has been estimated as being 26 pages long. Our Production Office will inform you of the exact length at the proof stage.

Proceedings A levies charges for articles which exceed 20 printed pages. (based upon approximately 540 words or 2 figures per page). Articles exceeding this limit will incur page charges of £150 per page or part page, plus VAT (where applicable).

Under the terms of our licence to publish you may post the author generated postprint (ie. your accepted version not the final typeset version) of your manuscript at any time and this can be made freely available. Postprints can be deposited on a personal or institutional website, or a recognised server/repository. Please note however, that the reporting of postprints is subject to a media embargo, and that the status the manuscript should be made clear. Upon publication of the definitive version on the publisher's site, full details and a link should be added.

You can cite the article in advance of publication using its DOI. The DOI will take the form: 10.1098/rspa.XXXX.YYYY, where XXXX and YYYY are the last 8 digits of your manuscript number (eg. if your manuscript number is RSPA-2017-1234 the DOI would be 10.1098/rspa.2017.1234).

For tips on promoting your accepted paper see our blog post: https://royalsociety.org/blog/2020/07/promoting-your-latest-paper-and-tracking-your-results/

On behalf of the Editor of Proceedings A, we look forward to your continued contributions to the Journal.

Sincerely,
Raminder Shergill
proceedingsa@royalsociety.org

on behalf of
Professor Graham Hughes
Board Member
Proceedings A

Appendix A

Review of "Revisiting causality using stochastics" by D.Koutsoyiannis, C. Onof, A. Christofides, and Z.W. Kundzewicz

I have read through the two manuscripts by Koutsoyiannis et al. on causal inference and its application. I enjoyed reading the comprehensive review and philosophical thoughts in section 2 of the first paper, which lead the authors to a conclusion that a genuine causal relation between two variables cannot be established through algorithms, but only necessary conditions can be proposed to falsify a hypothesis. Though there still exist many issues, this piece of work is in time when causal inference becomes an arena of enormous interest, and should be useful to the community. I myself hence would like to countenance the publication of some pieces of this work, preferably in a shorter form within one paper. But before that, the following issues must be resolved.

General

While I agree that "the big philosophical problem of causality" may "not be resolved by technical tricks", what the authors rely on is the extrapolation-like definition of causality Eq. (6) in their first paper. This definition (from Papoulis 1991), which is claimed in the ms to be "an ideal that we can hardly meet", is actually problematic even if it is met. Here I would not blame its linearity (in fact, linear system is the simplest system which makes a natural starting point). Its simple convolution form of some kernel with the other variable implies that the causal inference boils down to the linear regression, as testified in the later derivations in the ms. This is really problematic, as this is no better than inference of causality from correlation, which has been vehemently criticized by the authors. As a simple example, let us look at the classical problem of crock crowing (written x) versus sunrise (written y). Sun always rises after the crock crows. So y can be rather accurately described by x in a form of Eq. (6) in the ms. By the definition in this study, that means **crockcrow causes sunrise!** This absurdity results from, again, the mistakenly association of correlation to causality, which the authors have criticized. Unfortunately, the approach they propose is fundamentally like that.

Another serious problem with the Papoulis (1991) definition of causality is that, a representation of y using the past history of x does not need to mean a causality from x to y. By Takens' theorem, in a functional space, vectors represented by time-delayed series may not be parallel to the original vector. That is to say, the delayed series of x may NOT be x itself at all!

I however would still like to see the publication of this piece of work in some form. But the above problems must be clearly revealed to the reader, with a discussion of the limitations of the method. Besides, the whole paper can be shortened, with the second part (see a separate review) included.

Specifics

1. The method starts with Eq. (7). By minimizing the variance subject to an inequality restriction on the defined “roughness”, Eq. (41) follows. This is somehow similar to the Kalman filter which has been widely used in data assimilation.

2. p.9, l.19-23, this is for Gaussian only. If not Gaussian, it does not result in a correlation coefficient. Same for the Liang (2016) result. In fact, in Liang (2016), even when a Gaussian process is considered, the theorem asserts that causation implies correlation, but correlation does NOT imply causation; so I have no idea how the authors make their point here. Moreover, as far as I know, Liang defines causality from A to B as the change in information of B due to the existence of A. He opens the possibility of defining “information” with respect to quantities other than entropy. I am not sure whether the authors know the recent advances along that line, of which one being that the resulting causal measure is invariant upon arbitrary nonlinear coordinate transformation (e.g., Liang 2018), implying that it is an intrinsic physical property.

3. p. 10, l. 3, wrt. You’d better write “with respect to” in full here.

4. p.11, l.37-40, Indeed, In the framework of Liang (2016), both “>” and “<” exist, which result in positive and negative causalities.

5. The last paragraph of p.10 – the 1st paragraph of p.11: It would be helpful to add more details about how the absurd result is obtained. From the current description, it is difficult to see it.

6. p.14, ll.11-23, Papoulis’s definition, i.e., Eq. (6), is problematic. It is not appropriate to call it “purely causal”. For example, the simplest method used in geophysics, namely, time-delayed correlation analysis, for causal identification, is actually a particular case of Eq. (6). Starting with such a definition is contradictory to what the authors have strongly criticized in section 2: correlation is not causation.

7. Eq. (7). It is unclear to me why this implies a causality from x$\rightarrow$y. This violates the basic requirement claimed by the authors that cause precedes effect. Perhaps this is for the purpose of including the concept “anticausal’’? In that case, “anticausal” should be clearly defined in advance.

8. p.16, l.17-31, the so-called HOE should be clearly defined in advance. Do you mean a mutual causality?

9. Eq. (7) is essentially about regression. So the authors actually state that regression coefficients determine causality. Indeed this has been widely used, particularly in geophysics. But that is, again, equivalent to correlation analysis.

10. p.23, l.23-35, It should be noted that Liang (2016) considers both continuous time and discrete time; moreover, it is the continuous time formulation that results in the transparent solution therein.

Appendix B

Response to review comments on "Revisiting causality using stochastics: 1. Theory"

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

Summary: Version 1 of our manuscript "Revisiting causality using stochastics: 1. Theory" received two anonymous reviews and an additional assessment by an anonymous Board Member. Based on them, the paper thankfully received a favourable decision with an invitation to revise the manuscript in accordance with the suggestions of the reviewers. As seen below, where all review material is reproduced, the level of criticism we received is appropriate and made with a scientific spirit. Also, the reviewers' suggestions are constructive and thus helped us to improve our paper and clarify our methodology and its limitations, as we explain in detail below. In addition, the reviewers' critiques helped us to enhance our confidence about our method and results, and, in particular, to strengthen our confidence that our work is scientific, posited at the front of knowledge inquiry, where the ground is still under exploration—not completely known. The two reviewers' suggestions are generally in agreement with each other. However, there is one exception clearly distinguished below: While Reviewer's 1 comments imply expansion of our work to address them, Reviewer 2 suggests cutting a part of the companion paper and merging the two papers into one. Given that there was not an editorial directive for such a substantial cutting (see Board Member's Comments to Authors below), we preferred to follow the advice of Reviewer 1 and not to follow the suggestion to merge as, in our view, this would distort, truncate and devalue the entire study. At the same time, as each of the two papers already exceeds the length limitation of the 20 pages (21 pages each according to the Editor's letters), we put a lot of the expansion material into a new Supplementary Information report for part 2, so as to keep the main papers as short as possible. In addition, we cut a few paragraphs of the main paper 1 and moved some others to the Supplementary Information (also we moved a part of the Supplementary Information of part 1 to that of part 2).

Key:

Review comment.

Response.

Quotation from manuscript.

Quotation from other papers or from the reviews.

Note: The list of references contained in the bottom of this report is for the report per se and does not coincide with the references contained in the paper.

Board Member

Comments to Author(s):

The reviewers find this and its companion manuscript make an interesting contribution to the subject and have returned a number of questions and constructive comments for your consideration. Please address these points during the process of revising your manuscript(s).

We are grateful to the Board Member for the positive assessment. We appreciate the reviewers' comments and we have addressed them in the revised manuscript as we explain in detail in the report that follows.

Reviewer 1

> This paper gives a brief summary of the history of the concept of causality before going on to provide an approach to identifying sufficient conditions for causal relationships. That the summary is rather brief and selective is understandable in a research paper concerned more with technique than philosophy but there is quite a lot that does not get mentioned, (eg. Wesley Salmon, David Cox, the most recent edition of Mario Bunge's book, …) and some points that could be contested (e.g. the gross oversimplification of the ideas of Russell on p4.L9). The presentation of this section could also certainly be improved – for example the wording of the mention of Granger (1980) p6,L37-58 is really an aside that does not contribute to the argument, albeit that the later discussion of Granger causality as correlation is useful to repeat.

We appreciate the Reviewer's stance about our brief summary of the history and philosophy of the concept of causality. The Reviewer is right in that we omitted quite a lot in our summary of the history of causation. She or he also notes that we are oversimplifying the ideas of Russell. This omission is intentional. The amount of literature is staggering, and attempting even to list the important authors would lengthen the text too much, and it would also be largely out of scope. This is why, in the last paragraph of the section, we only mention the four broad categories of theories of causation, without elaborating:

> … interventions (A causes B if by deliberately creating A, B follows), counterfactuals (A causes B if B would not have occurred had A been absent), necessary and sufficient conditions (A causes B if A is necessary and sufficient for B to occur), or probability (A causes B if the presence of A increases the probability of B; see section 2.2). Combinations of these approaches have also been proposed. However, no completely satisfactory characterization has been formulated.

However, the Reviewer is right in that our presentation did not make this choice clear. Therefore, in response, as far as section 2.1 is concerned (now renamed to *Philosophical background*) we have actually removed much text and all references (including Russell's), except those for the three giants: Aristotle, Hume and Kant. We think that this makes the presentation clearer and more effective, since the section is not intended to be a comprehensive overview, but rather a quick introduction for the reader on how tricky the problem of causation is. The length of the section is thus reduced by 40%.

On the other hand, we expanded section 2.2 (now renamed to *Probabilistic theories of causality*) with additional key references, thankfully suggested by the Reviewer. As the Reviewer points out, this is indeed a vast area, but we found it useful in particular to include Wesley Salmon and David Cox's work in our discussion (this also led to including a reference to Brian Skyrms). Specifically, we added the following text in the section 2.2:

Suppes's third criterion conveys the idea that the presence of the cause raises the probability of occurrence of the effect. This idea is arguably better expressed as an inequality between conditional probabilities (Skyrms 1980):

(iii)′ $P(A_t|B_{t'}) > P(A_t|\overline{B}_{t'})$

where $\overline{B}_{t'}$ is the absence (non-occurrence) of event $B_{t'}$. However, using the obvious relationship $P(A_t) = P(A_t B_{t'}) + P(A_t \overline{B}_{t'})$, it can easily be shown that the two versions are equivalent. Cox (1992) points out that such a condition still allows for "spurious causality". The latter could only be eliminated by adding a condition such as:

(iv) there is no event $C_{t''}$ at time $t'' < t' < t$ such that $P(A_t|B_{t'}C_{t''}) = P(A_t|\overline{B}_{t'}C_{t''})$.

A version of this condition was also defined by Salmon (1998) within his statistical-relevance theory of explanation, as the key to distinguishing between statistical and causal relevance which he defines as:

(iv)′ there is no event $C_{t''}$ at time $t'' < t' < t$ which "screens off" $B_{t'}$ from A_t such that $P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''})$.

Salmon's example of statistical relevance which is not causal and therefore defines a spurious correlation is if $A_t, B_{t'}, C_{t''}$ refer respectively to the occurrence of a storm, a barometer drop and an air pressure drop. However, conditions such as (iv) or (iv)′ are pretty much impossible to verify satisfactorily in practice. This places limits upon the practical use of these characterisations of causation.

We also followed the Reviewer's suggestion about the way we refer to Granger (1980). Namely, we replaced the verbatim quotation of his axioms with the following summary.

Further, he provided three axioms, the first of which is equivalent to (i) above and the third highlights the constancy in causality direction throughout time.

We are glad with the Reviewer's statement "the later discussion of Granger causality as correlation is useful to repeat". Indeed, Granger's contributions have been quite popular and influential, forming a "standard" for application, and the reason for our extensive reference to them is to highlight the differences of our method from this "standard".

But concentrating on the technical aspects. The authors focus on the characteristics of temporal asymmetry and irreversibility and suggest that these can be demonstrated empirically for two variables for which data are available by the fitting of a linear impulse-response function (IRF) allowing for stochastic variability under a least squares "optimal" solution (see also comments on part 2). This approach has some nice features, such as the possibility of identifying anti-causal and hen or egg relationships but also seems to have some important limitations that are rather glossed over in the paper.

Again, we are glad that the Reviewer found that our approach has some nice features. As per the limitations, we appreciate the Reviewer's advice and in the revised manuscript we have made explicit reference to them, following the Reviewer's specific suggestions below.

The first of these is the linearity assumption (or the assumption that a simple transformation of variables can linearise the relationship) and that g(h) is continuous and always has the same sign. For real world open systems that would seem to be problematic – it is exactly why people build complex models to represent real world systems with causalities built in (though the authors are right to point out that such models cannot be used to test for causality as they intrinsically are built on causal chains). Of course, as shown in the Supplementary Information, we can linearise as an approximation or try differencing or different transforms first, but this will not necessarily result in stochastic terms that can be minimised by least squares as seems to be assumed in both justifications for linearisation in SI2.

We have now further discussed both issues of linearity and uniqueness of the sign—for the issue of roughness see our reply to next comment. About linearity we have added the following text in section 3.2:

> Further explanations on the motivation for the use of equation (7) as necessary condition for causation are provided in Supplementary Information (Section SI1.2), including a justification for its linear form. The linearity of the equation is kept from the original definition of a causal system by Papoulis (1991) (equation (6)). Certainly, linearity could be regarded by many as a limitation of our approach and possible future nonlinear extensions thereof could be considered. However, it is our opinion that linearity may suffice for most problems, for the following reasons:
>
> - We use a stochastic approach, in which the meaning of linearity vs. nonlinearity is dramatically different from that in deterministic approaches, something not often recognized in literature. In stochastics, linearity is rather a powerful characteristic enabling the study of demanding problems, rather than a limitation. For example, stochastic dynamics need not be nonlinear to produce realistic trajectories and change. Conversely, in a deterministic system with linear dynamics, any perturbation of initial conditions dies off, as does the potential for change—and hence the importance of nonlinearity in deterministic approaches (Koutsoyiannis 2014a, 2021; Koutsoyiannis and Dimitriadis, 2021),
> - In stochastics, linearity is not an (over)simplification of the dynamics but has some sound justification, as indicated by the already mentioned Wold decomposition, in which the stochastic component (the regular process) is linearly equivalent to a white noise process (i.e. a linear combination of white noise terms; Wold, 1938, 1948; Papoulis 1991).
> - In addition, linearity in a stochastic description results from maximum entropy considerations (under plausible conditions; e.g. Papoulis, 1991) and hence it is related to the most powerful mathematical and physical principle of maximum entropy (Jaynes, 1991; Koutsoyiannis et al. 2008; Koutsoyiannis 2014b).
> - In a stochastic approach, a deviation from linearity can be conveniently incorporated through an error term, which is already included in our proposed equation (7), in order to generalize Papoulis' (1991) original equation (6).
> - The fact that linearity is not regarded as a severe limitation in causality assessment is indirectly reflected in the popularity of Granger's (1969) approach, which is also linear (equation (1)).

> In the companion paper (Koutsoyiannis et al., 2022a,b), we show that the linear form of the framework effectively captures the important characteristics of causality, even in cases that the true dynamics is a priori known to be nonlinear.

With respect to the uniqueness of the sign, we have added the following text in section 3.3:

> Here we clarify that the problem of identifying causality is different from that of recovering the full system dynamics, where, clearly, the former and not latter, is the scope of our study. We note that, while there exist oscillatory nonlinear systems, in which the sign of $g(h)$ could alternate, we avoid subsuming them under the causality notion, particularly when causality is inferred from data in an inductive manner. This choice is consistent with Cox's (1992) conditions for causality, according to which the effect *"shows a monotone relation with 'dose'"* of the cause. Here we note that in our framework the "dose" is not regarded as an instantaneous event, but one with some time span (see details in Supplementary Information, section SI1.2).

This then also creates doubts about accepting only the least squares solution for the IRF, especially when the IRF may be really rather uncertain given noisy real world data (as the authors recognise in their discussion of autocorrelation and cross-correlation which will increase that uncertainty). It is well known that identifying the ordinates of a discrete IRF is an over-parameterised problem in this respect (there is a whole literature on the identification of transfer functions that is not really mentioned here – see below), so the results might depend on the identification method chosen. Hence the imposition of a roughness threshold, and the requirement of positive ordinates (though this would appear to rule out any consideration of oscillatory responses in a nonlinear system).

We agree that identifying the ordinates of a discrete IRF is an over-parameterized problem. Actually, we have tried to highlight that problem in the original manuscript. That is why we introduced the roughness constraint—and we are very glad that the Reviewer recognizes that. In the revised manuscript we have made more specific reference to this problem, by including the following text:

> As already mentioned, the above least-squares-based determination of the ordinates g_j is not the only technique for the identification of the IRF; additional techniques can be found in Young (2011, 2015 and references therein). A well-known weakness of determining numerous ordinates is that it is an over-parameterized problem, which is typically addressed by assuming a parametric model (such as a Box-Jenkins model or an autoregressive moving average exogenous—ARMAX—model; Young, 2011, 2015). Here we prefer to use a nonparametric approach and we tackle the over-parameterization problem by imposing the roughness threshold, as discussed above. 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 of the companion paper (Koutsoyiannis et al., 2022b; sections SI2.3 and SI2.4).

In addition, we have provided additional material in the new Supplementary Information report for the part 2, where we show that:

- The uncertainty is very large if we do not use the roughness constraint, but is substantially reduced after imposing this constraint. (Figure SI2.1).
- Even in the former case (without the roughness constraint), the key characteristics of causality (characteristic times, explained variance ratio) are estimated almost with certainty (Figure SI2.2).
- The methodology is flexible enough and can become as parsimonious as wanted (it can even become under-parameterized) by using a single parametric expression—and without a substantial cost in terms of the key characteristics of causality (Table SI2.1 and Figure SI2.3).

Finally, we agree with the Reviewer that positive ordinates would appear to rule out any consideration of oscillatory responses in a nonlinear system. However, in our view, this is a desideratum in order to avoid statistical artifacts, and has some theoretical/philosophical background as we tried to show in the Supplementary Information of part 1 (cf. our example with the tennis ball). While we recognize that there exist oscillatory nonlinear systems, we would avoid subsuming them under the causality notion, particularly when causality is inferred from data in an inductive manner. Further explanations about this have already given in our reply to the previous comment, where we also include the new text added to the manuscript.

> Finally, the simple two variable analysis of the IRF approach would appear to allow for the possibility of "spurious" causality, where changes in both variables are effected by some unmeasured cause but with different time delays and response functions. The authors discuss this in respect of a simple case of spatial proximity, but there may be cases of spatially extensive observed only at points for some variables where this might be a real issue.

We agree with the Reviewer and we have added the following text in section 3.2:

> We further note that our proposed bivariate approach to causality could allow for the possibility of "spurious" causality, where changes in both variables are affected by another cause, possibly with different time delays and response functions. This is not a drawback insofar as our framework of detecting necessary, rather than sufficient, conditions. But further, the inclusion of an error term in it allows for such more remote causes to be represented in the framework. Additional clarifications on multiple causes are provided in Supplementary Information (Section SI1.2),

> Which ultimately makes me wonder whether this type of simple linear(ised) analysis is really of value in inferring causality. Yes, the approach does allow for the recognition of ambiguity and the authors are quite clear that they are only assessing sufficient rather than necessary conditions but that only then means it is up to the analyst to determine whether the result is spurious or not. In which case, will it only provide a clear inference in obvious cases? Something here, for all the philosophy and maths, does not feel quite right but should perhaps await a consideration of the examples in Part 2.

We believe that what the Reviewer meant in her or his phrase "they are only assessing sufficient rather than necessary conditions" is the opposite, i.e., "they are only assessing necessary rather than sufficient conditions", because actually this is what we are doing. Furthermore, we strongly believe that our "analysis is really of value in inferring causality".

Indeed, it is noteworthy that the past attempts at characterising causality that we reviewed have in effect only been able to produce necessary conditions—even though this may have not been explicitly stated. But this does not entail that the decision about whether the causality is spurious or not is largely up to the analyst as the Reviewer suggests: some necessary conditions are more useful than others in weeding out cases of spurious causality. The second paper is precisely designed to show how well our proposal performs in identifying the presence of real causality.

We briefly state what the value of our analysis is in our following statement in the end of section 2.2:

> Therefore, here we focus on simpler problems, such as falsifying an assumed genuine causality and adding statistical evidence, in an inductive context, for potential causality and its direction.

A similar statement appears in the Conclusions:

> The methodological framework proposed herein features substantial differences from existing methods, such as those discussed in section 2.2. A first difference is in its epistemological background which leads to a less ambitious objective, that of seeking necessary conditions of causality rather than sufficient ones. The usefulness of this objective lies in its ability to falsify an assumed genuine causality and to add statistical evidence, in an inductive context, for potential causality and its direction.

Other comments
P16. It would be useful to illustrate these types of IRF here.

We thankfully followed the suggestion. In the revised version we have included the new Figure 1, also reproduced here.

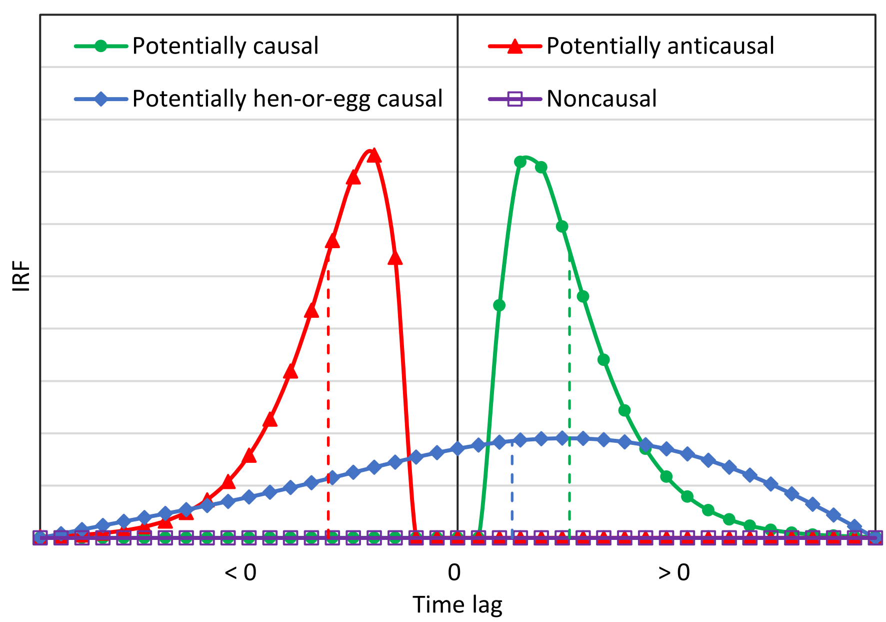

Figure 1 Explanatory sketch for the definition of the different potential causality types. For each graph the mean μ_h is also plotted with dashed line.

P22. Although this is an analytical solution it is known to be notoriously sensitive for noisy and correlated data because of the number of effective parameters being identified (relative to the identification of a low parameter functional form). Thus the roughness condition will be very important, but will interact with the form of the output and its uncertainty (as evident for example, rather dramatically in Figure 3 or Part 2). Does this not imply that inference will be somewhat dependent on the roughness condition applied?

We agree. Yet, our analytical solution (equation (41)) includes the roughness term $\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}$, not appearing in standard solutions and this makes the difference—as the Reviewer points out, "Thus the roughness condition will be very important". Actually, Figure 3 of Part 2 precisely serves this purpose: as no constraints were imposed to infer IRF, the estimate deviates from the true IRF, and this illustrates how important the constraints are. And yes, the inference of the IRF per se depends on the constraints. On the other hand, as shown in part 2 and its new Supplementary Information report, already mentioned above, the key characteristics (characteristic times, explained variance) depend only slightly on the constraints and this makes inference robust enough.

P23 L37ff. It is not clear how this differs from past work on using transfer functions for the prediction of bivariate and multivariate stochastic variables. The authors refer only to their own past work here with a suggestion that future work is required in future, but works such as Young (1984, 2011, 2019) is not mentioned. This also deals with the direct estimation of continuous time functions from which differential equations can be inferred (also Young, 2015).

Young, P C, 1984, 2011, Recursive estimation and time series analysis, Springer.

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

The differences of the proposed methodology from existing methods are discussed in the last five paragraphs of section 4, *Discussion and conclusions*. However, the Reviewer is right that we have omitted those key references. In the revised manuscript we have cited them. Thus, in the beginning of section 3.4 we write:

> The literature offers several methods for estimating an IRF in terms of auto- and cross-correlations (Young, 2011, 2015) or their Fourier transforms, i.e., power spectra and cross-spectra (e.g. Papoulis, 1991).

and below equation (31) we added the following text:

> As already mentioned, the above least-squares-based determination of the ordinates g_j is not the only technique for the identification of the IRF; additional techniques can be found in Young (2011, 2015 and references therein). A well-known weakness of determining numerous ordinates is that it is an over-parameterized problem, which is typically addressed by assuming a parametric model (such as a Box-Jenkins model or an autoregressive moving average exogenous—ARMAX—model; Young, 2011, 2015). Here we prefer to use a nonparametric approach and we tackle the over-parameterization problem by imposing the roughness threshold, as discussed above. 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 of the companion paper (Koutsoyiannis et al., 2022b; sections SI2.3 and SI2.4).

Supplementary Information
P11. Define LRD

Done; please notice that in the revised submission this section of the Supplementary Information was moved to that of part 2.

P12. Why not show the IRF?

We have included the IRF in the revised manuscript by adding the new Figure SI2.4, also reproduced here below.

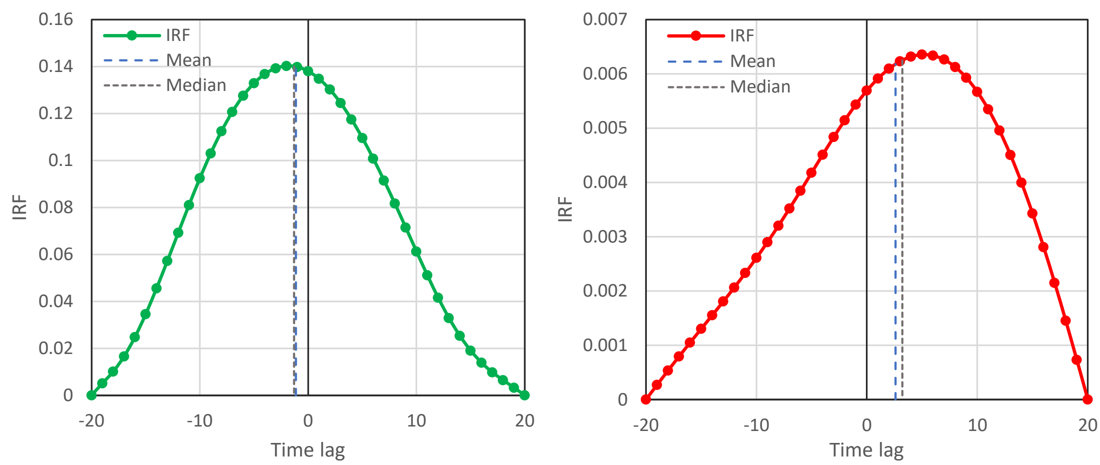

Figure SI2.4 IRFs for the synthetic example of spurious IRF estimation due to high autocorrelation of Figure SI1.1 for causality directions (left) $\underline{x} \rightarrow \underline{y}$ and (right) $\underline{y} \rightarrow \underline{x}$. For the estimated IRF the number of weights is $2J + 1$ with $J = 20$.

Reviewer 2

I have read through the two manuscripts by Koutsoyiannis et al. on causal inference and its application. I enjoyed reading the comprehensive review and philosophical thoughts in section 2 of the first paper, which lead the authors to a conclusion that a genuine causal relation between two variables cannot be established through algorithms, but only necessary conditions can be proposed to falsify a hypothesis. Though there still exist many issues, this piece of work is in time when causal inference becomes an arena of enormous interest, and should be useful to the community. I myself hence would like to countenance the publication of some pieces of this work, preferably in a shorter form within one paper. But before that, the following issues must be resolved.

We are grateful to the anonymous Reviewer for the scientific spirit of her or his review. We are glad that the Reviewer enjoyed the comprehensive review and the philosophical part. About the issues announced by the Reviewer, please refer to our responses to the next comments.

As per the Reviewer's suggestion that she or he would like "preferably in a shorter form within one paper", we believe that this is not possible for the following reasons:

- Reviewer 1 makes several points that require an expansion, rather than shortening of the work.
- Even Reviewer 2 includes comments that require expansion (see below), while she or he explicitly recognizes that "causal inference becomes an arena of enormous interest". On the other hand, the Reviewer suggests cutting the real-world case studies in the part 2, which however may be related to this "enormous interest". The extent of these case studies is less than 10 pages. Thus, even if we adopted this suggestion, the final outcome, after merging the two papers, would substantially exceed the length restrictions set by the journal (https://royalsocietypublishing.org/rspa/for-authors#question2).
- In the decision letter and the Board Member's Comments to Authors (see above) there was no editorial directive for such a radically severe cutting and merging.
- We strongly disagree to erase the real-world case studies, which were our inspiration in developing the methodology and preparing this work. We, all four authors, strongly believe that the real value of any methodology is assessed by confronting it with real-world data. While we certainly appreciate the philosophical and mathematical aspects related to causality—and this is reflected in the paper and also recognized by the reviewers—we are convinced that the removal of real-world examples would severely distort, truncate and devalue the entire study.
- The two parts could raise interest of two separate audiences. Part 1 could be of interest to theorists, statisticians, time-series experts, philosophers and system scientists, while part 2 to applied geophysicists, hydrologists and climatologists.
- In our reply to part 2, we provide additional information about our disagreement with the particular review comments for each of the case studies.

Therefore, to address the entire set of review comments, we were obliged to expand the study. On the other hand, to keep the "formal" part of the study as short as possible (the part that would hopefully be accepted and published in journal pages), we have put a lot of the expansion material into a second Supplementary Information report (new in part 2 of the study).

General
While I agree that "the big philosophical problem of causality" may "not be resolved by technical tricks", what the authors rely on is the extrapolation-like definition of causality Eq. (6) in their first paper. This definition (from Papoulis 1991), which is claimed in the ms to be "an ideal that we can hardly meet", is actually problematic even if it is met. Here I would not blame its linearity (in fact, linear system is the simplest system which makes a natural starting point). Its simple convolution form of some kernel with the other variable implies that the causal inference boils down to the linear regression, as testified in the later derivations in the ms. This is really problematic, as this is no better than inference of causality from correlation, which has been vehemently criticized by the authors. As a simple example, let us look at the classical problem of crock crowing (written x) versus sunrise (written y). Sun always rises after the crock crows. So y can be rather accurately described by x in a form of Eq. (6) in the ms. By the definition in this study, that means **crockcrow causes sunrise**! This absurdity results from,

again, the mistakenly association of correlation to causality, which the authors have criticized. Unfortunately, the approach they propose is fundamentally like that.

In terms of the definition by Papoulis, please refer to our reply to the next comment.

As the Reviewer correctly stated in the beginning of her or his review, we stress the fact that "only necessary conditions can be proposed" (the Reviewer may refer to the entire list of necessary conditions listed below equation (23)). In principle, the fact that the conditions we propose are necessary can save us from paradoxical or ridiculous results, which would emerge if we regarded the conditions sufficient. But even the necessary condition of explained variance does not hold in this particular case: If one sets x_τ for the time of the cockcrow and y_τ for the sunrise time, one would recognize that the former would be represented as a stochastic process (the time of the crow is never the same) while the latter is known precisely (deterministically) with zero variance. Thus, in our equation (41) the term $(\boldsymbol{y} - \boldsymbol{\mu})$ is zero by identity, and hence the resulting vector of IRF ordinates would be $\boldsymbol{g} = \boldsymbol{0}$. Consequently, speaking about the concept of explained variance is meaningless. It is also meaningless to think that a process that is known deterministically can be an effect of a stochastic process.

Furthermore, without being experts on biological issues, our perception is that the circadian rhythms of roosters are adapted to the Earth's motion, as experienced by them in the previous days. Let us not forget that the process brought about by the Reviewer is periodic, which means, for example, that an hour before today's sunrise is better thought of as 23 hours after yesterday's sunrise; the latter may better explain what happens with roosters' circadian rhythm adaptation to the Earth's motion.

Another serious problem with the Papoulis (1991) definition of causality is that, a representation of y using the past history of x does not need to mean a causality from x to y. By Takens' theorem, in a functional space, vectors represented by time-delayed series may not be parallel to the original vector. That is to say, the delayed series of x may NOT be x itself at all!

We have not stated that Papoulis (1991) gave a definition of causality. Rather, we state that he defines a *causal system*, i.e.:

> We recall from stochastics (e.g. Papoulis, 1991, pp. 405, 508) that the two processes form a *causal system* …
>
> But it was perhaps the book by Papoulis (1991…), that disseminated the concept of a "causal system".

Now we have further clarified it by adding the sentence:

> Noticeably, Papoulis did not provide a definition of causality per se, but used the concept of a causal system, defined through equation (6).

Other authors (Suppes, 1970; Granger, 1980) indeed gave definitions of causality and these are included in our paper.

As for Takens's theorem, this treats time series (series of numbers) rather than stochastic processes (series of stochastic variables) and therefore we do not see its relevance to our paper.

Furthermore, it is not clear to us what the phrase "the delayed series of x may NOT be x itself at all!" means.

I however would still like to see the publication of this piece of work in some form. But the above problems must be clearly revealed to the reader, with a discussion of the limitations of the method. Besides, the whole paper can be shortened, with the second part (see a separate review) included.

We appreciate the Reviewer's overall stance to "… like to see the publication of this piece of work …". As regards shortening, please refer to our detailed reply to Reviewer's 2 first comment.

Specifics

1. The method starts with Eq. (7). By minimizing the variance subject to an inequality restriction on the defined "roughness", Eq. (41) follows. This is somehow similar to the Kalman filter which has been widely used in data assimilation.

Apparently, there are similarities with other methods used in systems analysis and stochastics. Kalman filtering also provides options for smoothing. However, our algorithm is simpler and more intuitive, uses a different smoothing approach based on a roughness constraint, where the roughness is defined in terms of second derivatives, and instead of using recursive/iterative procedures, it determines the IRF by nonlinear optimization—numerically implemented. Thanks to its simplicity, the entire algorithm can be easily set up and run (actually, was run) in an Excel spreadsheet without difficulties. We see no reason to refer to something more complex (Kalman filter) to explain something simpler (our method).

2. p.9, l.19-23, this is for Gaussian only. If not Gaussian, it does not result in a correlation coefficient. Same for the Liang (2016) result. In fact, in Liang (2016), even when a Gaussian process is considered, the theorem asserts that causation implies correlation, but correlation does NOT imply causation; so I have no idea how the authors make their point here. Moreover, as far as I know, Liang defines causality from A to B as the change in information of B due to the existence of A. He opens the possibility of defining "information" with respect to quantities other than entropy. I am not sure whether the authors know the recent advances along that line, of which one being that the resulting causal measure is invariant upon arbitrary nonlinear coordinate transformation (e.g., Liang 2018), implying that it is an intrinsic physical property.

We agree that "causation implies correlation, but correlation does NOT imply causation". On the other hand, we do not view the concept of correlation as identical with the *correlation coefficient*. There can be many metrics of correlation (more generally, statistical association) including entropy- based (or information-based) ones. The mutual information, which is quoted (from Koutsoyiannis and Kundzewicz, 2020), in our opinion is also a measure of correlation (also the information flow). In the case of the Gaussian distribution, the correlation coefficient explicitly appears in the mutual information expression (and information flow), as seen in equation (3). Even when it does not appear explicitly (in distributions other than Gaussian) the essence that it is a measure of correlation/ statistical association and not one of causation, does not change. Nb. in logic just one counterexample (in our cast the Gaussian distribution) suffices to falsify a claim such as that information flow is (qualitatively) different from correlation and can rigorously determine causation per se.

That is the reason we tried to "make [our] point" and this is also the reason why we included the philosophical background about causality—to show that causality is not a technicality but a theoretical concept and quality.

The reference to Liang (2018) has been included in the revised manuscript as follows:

> For example, Liang (2016) used the so-called information flow (or information transfer) between two processes, while in later works he authored or co-authored this method has been called "*Liang causality*" (Stips et al., 2016). He asserted that "*causality actually can be rigorously derived in terms of information flow from first principles*" (Liang, 2018). However, the situation does not change if one uses information-based (equivalently, entropic-based) measures of correlation or statistical association instead of the standard correlation coefficients.

3. p. 10, l. 3, wrt. You'd better write "with respect to" in full here.

The phrase is within quotation marks, so we kept it intact in order not to modify the original.

4. p.11, l.37-40, Indeed, In the framework of Liang (2016), both ">" and "<" exist, which result in positive and negative causalities.

We thank the Reviewer for the confirmation.

5. The last paragraph of p.10 – the 1st paragraph of p.11: It would be helpful to add more details about how the absurd result is obtained. From the current description, it is difficult to see it.

We have extended the description as follows:

> Koutsoyiannis and Kundzewicz (2020) used the two-valued stochastic variables $\underline{x}, \underline{y}, \underline{z}$ to model the states of temperature, clothes weight and sweat, respectively, and assumed a hypothetical "artificial intelligence entity" (AIE) which decides on causality based upon the probability rules of Hannart et al. (2016). After assigning plausible values to the conditional probabilities of high sweat for the four conditions of cold/hot and heavy/light clothes, and following detailed numerical calculations of PN and PS, they obtained the absurd result that the AIE will decide that there is all necessary and sufficient evidence that light clothes cause high sweat.

Besides, the analytical calculations can be found in the reference cited.

6. p.14, ll.11-23, Papoulis's definition, i.e., Eq. (6), is problematic. It is not appropriate to call it "purely causal". For example, the simplest method used in geophysics, namely, time-delayed correlation analysis, for causal identification, is actually a particular case of Eq. (6). Starting with such a definition is contradictory to what the authors have strongly criticized in section 2: correlation is not causation.

We thank the Reviewer for the remark, but we would not characterize Papoulis definition problematic. We have explained that in our reply to Reviewer's 2 second general comment. On the other hand, the Reviewer may be right that the adverb "purely" may not express the indented meaning. Therefore, in the revised submission we have replaced the term "purely" with "classic" throughout both companion papers.

7. Eq. (7). It is unclear to me why this implies a causality from x◎y. This violates the basic requirement claimed by the authors that cause precedes effect. Perhaps this is for the purpose of including the concept "anticausal"? In that case, "anticausal" should be clearly defined in advance.

We do not say that equation (7) implies a causality direction $x \rightarrow y$. Actually, it does not. It is a general equation, valid for any pair of processes $\underline{x}(t)$ and $\underline{y}(t)$. The conditions that specify the different direction of causality are found later, in the numbered list 1-4 and now also illustrated in Figure 1. The term "anticausal" is also defined in this list (by the way, the term is also used by Papoulis).

8. p.16, l.17-31, the so-called HOE should be clearly defined in advance. Do you mean a mutual causality?

This term is also defined in the numbered list 1-4, illustrated in the new Figure 1 and explained in the next paragraph as follows:

> In this respect, in a HOE causal system, earlier realizations of $\underline{x}(t)$ affect the current realization of $\underline{y}(t)$, but also earlier realizations of $\underline{y}(t)$ affect the current realization of $\underline{x}(t)$. Thus, each one of the processes $\underline{x}(t)$ and $\underline{y}(t)$ is correlated to both the past and the future of the other one. This may seem paradoxical in terms of a conventional way of thinking about causality, but it is not more paradoxical than the expression "hen-or-egg", first used by Plutarch (*Moralia, Quaestiones convivales*, B, Question III). Clearly, Plutarch (and subsequent users of this expression) did not mean one particular hen and one particular egg; in this case the existence or not of a causal relationship would be easy to tell. Rather, he meant the sequences of all hens and all eggs, something similar with what the abstract term "process" used here represents.

Different authors may use different terms (perhaps including mutual causality), but a more common term is referred to in the line before the above paragraph, i.e.,

> In other texts, the potentially HOE causal systems are treated as causal systems with feedback.

9. Eq. (7) is essentially about regression. So the authors actually state that regression coefficients determine causality. Indeed this has been widely used, particularly in geophysics. But that is, again, equivalent to correlation analysis.

We never stated that. We clarify in many parts that we look for necessary conditions. Please refer to our replies to other comments above.

10. p.23, l.23-35, It should be noted that Liang (2016) considers both continuous time and discrete time; moreover, it is the continuous time formulation that results in the transparent solution therein.

We thank the Reviewer for the correction. We have now modified this statement to read:

> A fourth difference of our method from many other methods lies in the recognition that natural time is continuous rather than discrete (nb., some methods, e.g. Liang, 2016, also use continuous time).

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.

Granger, C.W. 1980. Testing for causality: a personal viewpoint. *J. Econ. Dynamics and Control*, 2, 329-352.

Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P. & Ghil, M. 2016 Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Amer. Met. Soc.* **97**(1), 99-110.

Jaynes, E.T. 1957 Information theory and statistical mechanics. *Physical Review*, 106 (4), 620-630

Koutsoyiannis, D. 2014a Random musings on stochastics (Lorenz Lecture). *AGU 2014 Fall Meeting*, American Geophysical Union, San Francisco, USA , doi: 10.13140/RG.2.1.2852.8804.

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

Koutsoyiannis, D. 2021 *Stochastics of hydroclimatic extremes - A cool look at risk.* Athens: Kallipos, ISBN: 978-618-85370-0-2, 333 pp; http://www.itia.ntua.gr/2000/.

Koutsoyiannis, D. & Dimitriadis, P. 2021 Towards generic simulation for demanding stochastic processes, Sci, 3, 34, doi: 10.3390/sci3030034.

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. & Kundzewicz, Z. W. 2022a Revisiting causality using stochastics: 2. Applications, *Proceedings of the Royal Society A*, in review.

Koutsoyiannis, D., Onof, C. Christofides, A. & Kundzewicz, Z. W. 2022b Revisiting causality using stochastics: 2. Applications – Supplementary Information, *Proceedings of the Royal Society A*, in review.

Koutsoyiannis, D., Yao, H. & Georgakakos, A. 2008 Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods, *Hydrol. Sci. J.*, 53 (1), 142–164, doi:10.1623/hysj.53.1.142.

Liang, X.S., 2016. Information flow and causality as rigorous notions ab initio. *Phys. Rev. E*, 94, 052201.

Liang, X.S., 2018. Causation and information flow with respect to relative entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7), 075311.

Papoulis, A., 1991. *Probability, Random Variables and Stochastic Processes*, 3rd ed.; New York, NY, USA, McGraw-Hill (1st edition 1965).

Salmon, W. 1998 *Causality and Explanation*, New York: Oxford University Press.

Skyrms, B., 1980. *Causal Necessity: A Pragmatic Investigation of The Necessity of Laws*. New Haven and London: Yale University Press.

Stips, A., Macias, D., Coughlan, C., Garcia-Gorriz, E., Liang, X. S. 2016 On the causal structure between CO_2 and global temperature. *Sci. Rep.* **6**, 21691, doi:10.1038/srep21691.

Suppes, P., 1970. *A Probabilistic Theory of Causality*. Amsterdam, The Netherlands: North-Holland Publishing.

Wold, H.O. 1938 A Study in the Analysis of Stationary Time-Series. Ph.D. Thesis, Almquist and Wicksell, Uppsala, Sweden.

Wold, H. O. A. 1948 On prediction in stationary time series. *Ann. Math. Statist.* **19**(4), 558 – 567, doi: 10.1214/aoms/1177730151.

Young, P.C., 1984, 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.

Appendix C

I thank the authors for the thoughtful response. I particularly like the explanation on roosters' circadian rhythm. A similar remark is also seen in Liang (2014), where the causal inference approach using time-delayed correlation analysis (which has been extensively used in geophysics), is criticized, as in a periodic process there is no way to distinguish a phase lag, say $\pi/2$, from a phase advance $3\pi/2$.

The following are some new issues raised.

l. 50, "while in later works he authored or co-authored this method has been called '*Liang causality*'".
As far as I know, Liang himself never called it "Liang causality"; he usually calls it "information flow" or "information transfer". Does this appear in the paper as cited? If so, he was the last coauthor who might have overlooked that. So here the authors should be careful by writing this. The words "he authored or co-authored" should be taken out, in order not to mislead the reader.

l. 53-56: "However, the situation does not change if one uses information-based (equivalently, entropic-based) measures of correlation or statistical association instead of the standard correlation coefficients."

No. The authors are wrong here.
Liang (2016)'s formalism is by NO means about using information-based measures of correlation/association to infer causality (like mutual information which may be viewed as a nonlinear extension of correlation). Rather, it is something which ultimately results in a (closed-form) formula (see below) telling that correlation only makes a necessary condition for causation, just as the authors claimed in the present manuscript.

To clarify this, and to avoid further misunderstanding, here what Liang really stated in Liang (2016), and many other papers, is written down:

The causality from X_2 to X_1 within a stochastic system for $\mathbf{X} = (X_1, X_2, X_3, \ldots, X_n)$

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(t;\mathbf{X}) + \mathbf{B}(t;\mathbf{X})\dot{\mathbf{w}}$$

is quantitatively given by

$$T_{2\to 1} = -\int_{\mathbb{R}^n} \rho_{2|1} \frac{\partial (F_1 \rho_{\backslash 2})}{\partial x_1} dx + \frac{1}{2} \int_{\mathbb{R}^n} \rho_{2|1} \frac{\partial^2 (g_{11} \rho_{\backslash 2})}{\partial x_1^2} dx,$$

where $\rho_{\backslash 2} = \int \rho dx_2$, and $\rho_{2|1}$ is the probability density function of X_2 conditioned on X_1.

I would like to caution that, just on the contrary to what the authors criticized, Liang

(2016) took a stance similar to what the authors here are taking in this ms. In fact, Liang (2016) proved that the above formalism results in a mathematical expression which asserts that “causation implies correlation, but not vice versa”. This is equivalent to the authors’ opinion that correlation/association only makes a necessary condition.

This statement in l.53-56 will leave the reader a wrong impression, and hence should be deleted. Besides, the above fact should be mentioned (1) to avoid misleading the reader, and perhaps (2) to substantiate the authors’ claim that only necessary conditions can be found for causal inference.

Appendix D

Response to round 2 review comments on "Revisiting causality using stochastics: 1. Theory"

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

Summary: Version 2 of our manuscript "Revisiting causality using stochastics: 1. Theory" received two anonymous reviews and an additional assessment by the Board Member Graham Hughes. Based on them, the paper thankfully received a favourable decision of acceptance after minor revisions. All reviewers' suggestions for minor changes have been implemented in the revised Version 3.

Key:

> Review comment.

Response.

Board Member

> Comments to Author(s):
> Thank you for the revisions of your manuscript. Both reviewers are supportive of publication with some minor changes - please adapt your final version accordingly.

We are grateful to the Board Member for the positive assessment. We appreciate the reviewers' comments and we have addressed them in Version 3 as we explain in detail below.

Reviewer 1

> P5 L50. Should be t''. not (At)''

These were not primes (′ ′) but closing quotation marks ("). To avoid possible confusion, we removed both opening and closing quotation marks.

> P6 L9. (iii′) better? Also (iv′) at L26 and L32

Modified as suggested

> P 6L14 (font problem) "spurious causality"

Corrected.

> P16 L4. Delete a,b also P25L18 – do not need to separately reference supplementary information

Done.

> P19L36 not the latter

Done.

> P24L23 delete genuine (implies it is an actual causality rather than just assumed as correct)

Done.

We are grateful for all these detailed suggestions and we thank the Reviewer for the impressively attentive reading.

Reviewer 2

> I thank the authors for the thoughtful response. I particularly like the explanation on roosters' circadian rhythm. A similar remark is also seen in Liang (2014), where the causal inference approach using time-delayed correlation analysis (which has been extensively used in geophysics), is criticized, as in a periodic process there is no way to distinguish a phase lag, say $\pi/2$, from a phase advance $3\pi/2$.

We are grateful to the anonymous Reviewer for her or his help and we are glad that she or he found our response thoughtful (as well as being in agreement with Liang (2014) as regards phase lags in periodic processes).

> The following are some new issues raised.
>
> l. 50, "while in later works he authored or co-authored this method has been called '*Liang causality*'".
> As far as I know, Liang himself never called it "Liang causality"; he usually calls it "information flow" or "information transfer". Does this appear in the paper as cited? If so, he was the last coauthor who might have overlooked that. So here the authors should be careful by writing this. The words "he authored or co-authored" should be taken out, in order not to mislead the reader.

Done: The words "he authored or co-authored" have been deleted.

> l. 53-56: "However, the situation does not change if one uses information-based (equivalently, entropic-based) measures of correlation or statistical association instead of the standard correlation coefficients."
>
> No. The authors are wrong here.
> Liang (2016)'s formalism is by NO means about using information-based measures of correlation/association to infer causality (like mutual information which may be viewed as a nonlinear extension of correlation). Rather, it is something which ultimately results in a (closed-form) formula (see below) telling that correlation only makes a necessary condition for causation, just as the authors claimed in the present manuscript.

To clarify this, and to avoid further misunderstanding, here what Liang really stated in Liang (2016), and many other papers, is written down:

The causality from X_2 to X_1 within a stochastic system for $\mathbf{X} = (X_1, X_2, X_3, ..., X_n)$

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(t;\mathbf{X}) + \mathbf{B}(t;\mathbf{X})\dot{\mathbf{w}}$$

is quantitatively given by

$$T_{2\to 1} = -\int_{\mathbb{R}^n} \rho_{2|1} \frac{\partial (F_1 \rho_{\backslash 2})}{\partial x_1} dx + \frac{1}{2}\int_{\mathbb{R}^n} \rho_{2|1} \frac{\partial^2 (g_{11} \rho_{\backslash 2})}{\partial x_1^2} dx,$$

where $\rho_{\backslash 2} = \int \rho dx_2$, and $\rho_{2|1}$ is the probability density function of X_2 conditioned on X_1.

I would like to caution that, just on the contrary to what the authors criticized, Liang (2016) took a stance similar to what the authors here are taking in this ms. In fact, Liang (2016) proved that the above formalism results in a mathematical expression which asserts that "causation implies correlation, but not vice versa". This is equivalent to the authors' opinion that correlation/association only makes a necessary condition.

This statement in l.53-56 will leave the reader a wrong impression, and hence should be deleted. Besides, the above fact should be mentioned (1) to avoid misleading the reader, and perhaps (2) to substantiate the authors' claim that only necessary conditions can be found for causal inference.

Done. The indicated lines have been deleted and the quotation "*causation implies correlation, but not vice versa*" has been included.