THE ROYAL SOCIETY PUBLISHING

ROYAL SOCIETY OPEN SCIENCE

Using Fisher information to track stability in multivariate systems

Nasir Ahmad, Sybil Derrible, Tarsha Eason and Heriberto Cabezas

Article citation details
R. Soc. open sci. **3**: 160582.
http://dx.doi.org/10.1098/rsos.160582

Review timeline
Original submission: 8 August 2016
Revised submission: 11 October 2016
Final acceptance: 12 October 2016

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

Review History

RSOS-160582.R0 (Original submission)

Review form: Reviewer 1 (Shweta Singh)

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
YES

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

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

Comments to the Author(s)
This is an important contribution to the study of multivariate complex systems as the challenge of the right metric is still unresolved. The paper presents details on use of Fisher Information as a metric to detect regime shifts in complex systems using a time series data. Availability of a Python module will allow others to use it. Hence, it is a solid contribution to the scientific community
Further, the application of work on long term temperature is a good application example.

Only two points come to my mind that needs to be clarified or corrected.
- It is not clear to an ordinary reader how the quantum mechanics reference is related to equation 3 ? What amplitude they refer to in line 42 on page 2 ?
- In equation 4, I believe the summation sign is I =1 instead of I -1 ?

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

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

Comments to the Author(s)
The manuscript titled, “Using Fisher Information to Track Stability in Multivariate Systems” by Ahmad et al. focuses on the application of information theory to understand stability in multivariate systems. The authors specifically demonstrate the methodology using the concept of Fisher Information. The manuscript is easy to follow and presents the calculation methodology with a dummy example followed by demonstrating the use of FI to track trends in regime shifts

for the global mean temperature over a period of 135 years (1880-2015). In my opinion, the use of information theory for sustainability science problems is highly under appreciated, holds strong promise, and should be encouraged with more applications. I commend the authors for taking on this very important and critical piece of work. Overall, I have a very positive opinion of the manuscript.

Additional clarity in certain sections would further strengthen the manuscript (see my detailed comments).

Specific Comments:

1. I very much appreciate all the equations for Fisher Information and the formulation explicitly utilized in the present manuscript. In going from equation 2 to equation 3, probability is replaced with the amplitude squared, that I believe represents probability density too. While the authors cite some of Cabezas' previous work (Ref. 12), I recommend providing some rationale and interpretation here to inform the readers.
2. Page 5- regarding the discussion on tightening levels (TL), can a preference structure be established in a multivariate system where certain variables get preferential treatment compared to others when detecting states?
3. The reviewer is particularly intrigued by the use of FI in multivariate systems with high multicollinearity. In such systems, while one can calculate FI, there is high redundancy arising from multicollinearity. Could the authors provide some guidance on this since the FI calculated in such cases may be artificially inflated? The authors do mention at the beginning of the manuscript that the choice of variables selected in a system is important.
4. Section 4- Is there any reported study or system that has seen a steady increase in FI?
5. This comment pertains to the choice of case study. While simple, appealing, and a topic for which good quality long-term data exists, I wonder if the application of FI to study patterns in global mean temperatures is the best case study to illustrate the concept of FI. I understand the central highlight of the manuscript is not the case study itself but the demonstration of the FI calculation methodology. Having said that, the case study chosen is such that it is likely to reveal trends in regime shifts (high industrial development and resulting environmental degradation starting in late 1970s and early 1980s).
6. Note that the FI methodology itself has been demonstrated amply and sufficiently in several previous studies by Cabezas and co-workers, several of which are cited in the manuscript. In light of this, I would recommend the authors to explicitly state and rephrase the main contributions of the present manuscript and what it adds to the literature beyond what has done previously. Of course, the authors mention making the python library publicly available (one of the contributions).
7. I recommend a careful proofreading of the entire manuscript to fix minor typos and grammatical errors.

Decision letter (RSOS-160582)

05-Oct-2016

Dear Mr Ahmad

On behalf of the Editors, I am pleased to inform you that your Manuscript RSOS-160582 entitled "Using Fisher Information to Track Stability in Multivariate Systems" has been accepted for publication in Royal Society Open Science subject to minor revision in accordance with the referee suggestions. Please find the referees' comments at the end of this email.

The reviewers and handling editors have recommended publication, but also suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the comments and revise your manuscript.

• Ethics statement
If your study uses humans or animals please include details of the ethical approval received, including the name of the committee that granted approval. For human studies please also detail whether informed consent was obtained. For field studies on animals please include details of all permissions, licences and/or approvals granted to carry out the fieldwork.

• Data accessibility
It is a condition of publication that all supporting data are made available either as supplementary information or preferably in a suitable permanent repository. The data accessibility section should state where the article's supporting data can be accessed. This section should also include details, where possible of where to access other relevant research materials such as statistical tools, protocols, software etc can be accessed. If the data has been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that has been made publicly available. Data sets that have been deposited in an external repository and have a DOI should also be appropriately cited in the manuscript and included in the reference list.

If you wish to submit your supporting data or code to Dryad (http://datadryad.org/), or modify your current submission to dryad, please use the following link:
http://datadryad.org/submit?journalID=RSOS&manu=RSOS-160582

• Competing interests
Please declare any financial or non-financial competing interests, or state that you have no competing interests.
• Authors' contributions
All submissions, other than those with a single author, must include an Authors' Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

All contributors who do not meet all of these criteria should be included in the acknowledgements.

We suggest the following format:
AB carried out the molecular lab work, participated in data analysis, carried out sequence alignments, participated in the design of the study and drafted the manuscript; CD carried out the statistical analyses; EF collected field data; GH conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

• Acknowledgements
Please acknowledge anyone who contributed to the study but did not meet the authorship criteria.

• Funding statement
Please list the source of funding for each author.

Please note that we cannot publish your manuscript without these end statements included. We have included a screenshot example of the end statements for reference. If you feel that a given heading is not relevant to your paper, please nevertheless include the heading and explicitly state that it is not relevant to your work.

Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days (i.e. by the 14-Oct-2016). If you do not think you will be able to meet this date please let me know immediately.

To revise your manuscript, log into https://mc.manuscriptcentral.com/rsos and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions". Under "Actions," click on "Create 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 referees and upload a file "Response to Referees" in "Section 6 - File Upload". You can use this to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the referees.

When uploading your revised files please make sure that you have:

1) A text file of the manuscript (tex, txt, rtf, docx or doc), references, tables (including captions) and figure captions. Do not upload a PDF as your "Main Document".
2) A separate electronic file of each figure (EPS or print-quality PDF preferred (either format should be produced directly from original creation package), or original software format)
3) Included a 100 word media summary of your paper when requested at submission. Please ensure you have entered correct contact details (email, institution and telephone) in your user account
4) Included the raw data to support the claims made in your paper. You can either include your data as electronic supplementary material or upload to a repository and include the relevant doi within your manuscript
5) All supplementary materials accompanying an accepted article will be treated as in their final form. Note that the Royal Society will neither edit nor 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.

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

Kind regards,
Alice Power
Editorial Coordinator

Royal Society Open Science
openscience@royalsociety.org

on behalf of Mark Chaplain
Subject Editor, Royal Society Open Science
openscience@royalsociety.org

Reviewer comments to Author:
Reviewer: 1

Comments to the Author(s)
This is an important contribution to the study of multivariate complex systems as the challenge of the right metric is still unresolved. The paper presents details on use of Fisher Information as a metric to detect regime shifts in complex systems using a time series data. Availability of a Python module will allow others to use it. Hence, it is a solid contribution to the scientific community
Further, the application of work on long term temperature is a good application example.

Only two points come to my mind that needs to be clarified or corrected.
- It is not clear to an ordinary reader how the quantum mechanics reference is related to equation 3 ? What amplitude they refer to in line 42 on page 2 ?
- In equation 4, I believe the summation sign is I =1 instead of I -1 ?

Reviewer: 2

Comments to the Author(s)
The manuscript titled, “Using Fisher Information to Track Stability in Multivariate Systems” by Ahmad et al. focuses on the application of information theory to understand stability in multivariate systems. The authors specifically demonstrate the methodology using the concept of Fisher Information. The manuscript is easy to follow and presents the calculation methodology with a dummy example followed by demonstrating the use of FI to track trends in regime shifts for the global mean temperature over a period of 135 years (1880-2015). In my opinion, the use of information theory for sustainability science problems is highly under appreciated, holds strong promise, and should be encouraged with more applications. I commend the authors for taking on this very important and critical piece of work. Overall, I have a very positive opinion of the manuscript.

Additional clarity in certain sections would further strengthen the manuscript (see my detailed comments).

Specific Comments:

1. I very much appreciate all the equations for Fisher Information and the formulation explicitly utilized in the present manuscript. In going from equation 2 to equation 3, probability is replaced with the amplitude squared, that I believe represents probability density too. While the authors cite some of Cabezas' previous work (Ref. 12), I recommend providing some rationale and interpretation here to inform the readers.
2. Page 5- regarding the discussion on tightening levels (TL), can a preference structure be established in a multivariate system where certain variables get preferential treatment compared to others when detecting states?

3. The reviewer is particularly intrigued by the use of FI in multivariate systems with high multicollinearity. In such systems, while one can calculate FI, there is high redundancy arising from multicollinearity. Could the authors provide some guidance on this since the FI calculated in such cases may be artificially inflated? The authors do mention at the beginning of the manuscript that the choice of variables selected in a system is important.
4. Section 4- Is there any reported study or system that has seen a steady increase in FI?
5. This comment pertains to the choice of case study. While simple, appealing, and a topic for which good quality long-term data exists, I wonder if the application of FI to study patterns in global mean temperatures is the best case study to illustrate the concept of FI. I understand the central highlight of the manuscript is not the case study itself but the demonstration of the FI calculation methodology. Having said that, the case study chosen is such that it is likely to reveal trends in regime shifts (high industrial development and resulting environmental degradation starting in late 1970s and early 1980s).
6. Note that the FI methodology itself has been demonstrated amply and sufficiently in several previous studies by Cabezas and co-workers, several of which are cited in the manuscript. In light of this, I would recommend the authors to explicitly state and rephrase the main contributions of the present manuscript and what it adds to the literature beyond what has done previously. Of course, the authors mention making the python library publicly available (one of the contributions).
7. I recommend a careful proofreading of the entire manuscript to fix minor typos and grammatical errors.

Author's Response to Decision Letter for (RSOS-160582)

See Appendix A.

Decision letter (RSOS-160582.R1)

12-Oct-2016

Dear Mr Ahmad,

I am pleased to inform you that your manuscript entitled "Using Fisher Information to Track Stability in Multivariate Systems" is now accepted for publication in Royal Society Open Science.

You can expect to receive a proof of your article in the near future. Please contact the editorial office (openscience_proofs@royalsociety.org and openscience@royalsociety.org) to let us know if you are likely to be away from e-mail contact. Due to rapid publication and an extremely tight schedule, if comments are not received, your paper may experience a delay in publication.

Royal Society Open Science operates under a continuous publication model (http://bit.ly/cpFAQ). Your article will be published straight into the next open issue and this will be the final version of the paper. As such, it can be cited immediately by other researchers. As the issue version of your paper will be the only version to be published I would advise you to check your proofs thoroughly as changes cannot be made once the paper is published.

In order to raise the profile of your paper once it is published, we can send through a PDF of your paper to selected colleagues. If you wish to take advantage of this, please reply to this email with the name and email addresses of up to 10 people who you feel would wish to read your article.

On behalf of the Editors of Royal Society Open Science, we look forward to your continued contributions to the Journal.

Best wishes,
Andrew Dunn
Senior Publishing Editor
Royal Society Open Science
openscience@royalsociety.org

Appendix A

Dear Editor,

We have attached the revised manuscript, "Using Fisher Information to Track Stability in Multivariate Systems" (RSOS-160582). The reviewers' comments were very helpful in revising the manuscript and our responses are summarized below.

Reviewer 1

1. **It is not clear to an ordinary reader how the quantum mechanics reference is related to equation 3? What amplitude they refer to in line 42 on page 2?**

   Agreed. To make it less complicated, we removed the quantum mechanics statement. The amplitude of the probability density $p(s)$ is $q^2(s)$ and was used to eliminate the need for handing small probability densities in the denominator of Equation 2. Some text was added to clarify the relationship.

2. **In equation 4, I believe the summation sign is I =1 instead of I -1?**

   Yes, thank you. We corrected it.

Reviewer 2

1. **I very much appreciate all the equations for Fisher Information and the formulation explicitly utilized in the present manuscript. In going from equation 2 to equation 3, probability is replaced with the amplitude squared, that I believe represents probability density too. While the authors cite some of Cabezas' previous work (Ref. 12), I recommend providing some rationale and interpretation here to inform the readers.**

   Moving from equation 2 to 3, the probability density $p(s)$ is replaced by the probability amplitude function, using $q^2(s) = p(s)$, and was used to eliminate the need for handing small probability densities in the denominator of equation 2. We updated the text to clarify this relationship.

2. **Page 5- regarding the discussion on tightening levels (TL), can a preference structure be established in a multivariate system where certain variables get preferential treatment compared to others when detecting states?**

   At this time, the tightening level (TL) algorithm is "blind" in that there is no preference at all. Essentially the TL starts at 100% indicating that all variables must meet the size of states criteria in order to bin (group) points into the same state. Then, the TL decreases from 1 to 99% and the algorithm checks to see if 99% of the variables meet the size of state criteria, and so on. This process is done iteratively but without identifying the variables that do or do not meet the criteria. Indeed, a preference structure could be established which would be useful in identifying the variables that have the greatest impact on changes in FI. This is something that we have considered and definitely plan to explore in the future, but it is not part of the algorithm at this moment.

3. **The reviewer is particularly intrigued by the use of FI in multivariate systems with high multicollinearity. In such systems, while one can calculate FI, there is high redundancy arising from multicollinearity. Could the authors provide some guidance on this since the FI calculated in such cases may be artificially inflated? The authors do mention at the beginning of the manuscript that the choice of variables selected in a system is important.**

   The variables used need to properly characterize system condition. Therefore, their selection is critical to ensure that the dynamic behaviour of the system is properly captured. Since variable trends are typically unique to a particular system, it is quite possible that some variables may reflect similar patterns. However, given that we are not developing a predictive model, the "correlation structure is not critical" to computing FI (Eason and Cabezas 2012). The FI computation entails monitoring system dynamics by assessing changes in each individual variable and not in comparison to each other. With that said, we highly recommend that care be taken when selecting variables and make sure not to use both specific and summative variables to describe a particular system characteristic (e.g. summative: meat consumption, specific: bovine consumption, sheep consumption). If desired, principle component analysis (PCA) can be used to examine the correlation structure and create a sub-set of uncorrelated variables (US EPA 2010). However, one has to understand that the presence or absence of correlation between variables as determined from a statistical procedure is valid only for the specific values in the particular data set which characterize a certain set of conditions. The system under different conditions may or may not have these correlations. Simply put, it is the responsibility of the user to ensure that the right data is selected (otherwise the computer saying "garbage in – garbage out" applies).

4. **Section 4- Is there any reported study or system that has seen a steady increase in FI?**

   Yes, this behaviour is often seen particularly when studying regime shifts. Typically, a system becomes less stable as it moves toward a threshold. In many systems, this instability is reflected as a declining FI. As the system passes the threshold, if there are attractors that draw the system into a new steady state regime, the system will begin to reorganize into a new set of conditions resulting in a steadily increasing FI for a period and then plateauing as the system stabilizes into a new regime. As an example, we explored shifting dynamics in a shallow lake by modelling the inflow and removal of phosphorous. Initially, the phosphorous concentration was stable and then slowly increased. During the initial transition, there was an abrupt decline, followed by a prompt increase and stabilization in FI. This was followed by a more prolonged period of phosphorous loading characterized by a rapid decline and then a shallow, yet steadily rising FI, preceding a more rapid FI increase as the system reached a tipping point and phosphorous level overwhelmed the phosphorous removal efforts (e.g. sedimentation, sequestration) causing the system to transition from oligotrophic to eutrophic (Eason et al. 2016). Other examples include a predator-prey ecosystem model, regime shift in the pacific ocean, long term instability in Foy Lake (Montana, USA) diatom community dynamics, social-economic conditions in urban systems, US air quality and a drought impacted Spanish watershed (Karunanithi et al. 2008, Gonzalez-Mejia et al. 2012, Eason et al. 2014; Gonzalez-Mejia et al. 2014; Spanbauer et al. 2014; Gonzalez-Mejia et al. 2015; Eason et al. 2016).

As noted in Section 4, FI trends are very much dependent on the dynamics of the underlying system variables (Eason et al. 2014; Gonzalez-Mejia et al. 2015).

5. **This comment pertains to the choice of case study. While simple, appealing, and a topic for which good quality long-term data exists, I wonder if the application of FI to study patterns in global mean temperatures is the best case study to illustrate the concept of FI. I understand the central highlight of the manuscript is not the case study itself but the demonstration of the FI calculation methodology. Having said that, the case study chosen is such that it is likely to reveal trends in regime shifts (high industrial development and resulting environmental degradation starting in late 1970s and early 1980s).**

   Indeed, we thought it best to present a simple case study to make it easy for the reader to grasp the basic concepts of computing FI. Our hope is that this approach entices the reader to use the python scripts to explore their own systems and use some of more advanced studies as inspiration. As the reviewer pointed, the evolution of global mean temperatures may not be the ideal data set, but we still prefer to use it since it simply relates to many other disciplines, and most researchers will have a "feel" for it when reading the article and trying out the scripts.

6. **Note that the FI methodology itself has been demonstrated amply and sufficiently in several previous studies by Cabezas and co-workers, several of which are cited in the manuscript. In light of this, I would recommend the authors to explicitly state and rephrase the main contributions of the present manuscript and what it adds to the literature beyond what has done previously. Of course, the authors mention making the python library publicly available (one of the contributions).**

   In line with the reviewers' assessments, the goal of the work is to expand the use of FI for sustainability science problems and extend it to big data applications. The method has been used in multiple arenas but the reach of the approach has been limited by only having a Matlab version of the algorithm which is only available either by capturing the code from the US EPA study or contacting the code developer directly to access the executable graphical user interface (US EPA 2010). This manuscript serves as a means of presenting the FI methodology, demonstrating the utility of the approach with a simple example, and introducing the openly accessible python scripts. It is much more inline with a Nature Scientific Data article (hence the "official" publication of the code that will now be citable by citing this article). Text has been added to clarify the contributions.

7. **I recommend a careful proofreading of the entire manuscript to fix minor typos and grammatical errors.**

   Reviewed and revised as necessary. Thank you very much for your comments.

References

- Eason, T., Garmestani, A. S. & Cabezas, H. 2014 Managing for resilience: early detection of regime shifts in complex systems. Clean Technol. Environ. Policy 16, 773–783. (doi:10.1007/s10098-013-0687-2)

- Eason, T., Garmestani, A. S., Stow, C. A., Rojo, C., Alvarez-Cobelas, M. & Cabezas, H. 2016. Managing for resilience: an information theory-based approach to assessing ecosystems. J. Appl. Ecol. , n/a-n/a. (doi:10.1111/1365-2664.12597)

- Gonzalez-Mejía, A. M., Eason, T. N., Cabezas, H. & Suidan, M. T. 2012 Assessing Sustainability in Real Urban Systems: The Greater Cincinnati Metropolitan Area in Ohio, Kentucky, and Indiana. Environ. Sci. Technol. 46, 9620–9629. (doi:10.1021/es3007904)

- González-Mejía, A. M., Eason, T. N., Cabezas, H. & Suidan, M. T. 2014 Social and economic sustainability of urban systems: comparative analysis of metropolitan statistical areas in Ohio, USA. *Sustain. Sci.* **9**, 217–228. (doi:10.1007/s11625-013-0227-3)

- Gonzalez-Mejia, A. M., Vance, L., Eason, T. & Cabezas, H. 2015 Recent developments in the application of Fisher information to sustainable environmental management. In Assessing and Measuring Environmental Impact and Sustainability (ed J. J. Klemes), pp. 25–72. Butterworth-Heinemann.

- Karunanithi, A. T., Cabezas, H., Frieden, B. R. & Pawlowski, C. W. 2008 Detection and assessment of ecosystem regime shifts from Fisher information. Ecol. Soc. 13(1): 22.

- Spanbauer, T. L., Allen, C. R., Angeler, D. G., Eason, T., Fritz, S. C., Garmestani, A. S., Nash, K. L. & Stone, J. R. 2014 Prolonged instability prior to a regime shift. PloS ONE 9(10): e108936. doi:10.1371/journal.pone.0108936.

We appreciate your consideration of this revised manuscript.

Sincerely,

Nasir Ahmad

Sybil Derrible

Tarsha Eason

Heriberto Cabezas