

Review of Broomandi et al.

Broomandi et al. present a graph theoretical analysis of causality and PM2.5. They demonstrate the utility of network analysis for understanding complex behavior of atmospheric pollutants. The work applies a novel technique in the analysis of PM2.5 data across the United Kingdom, with results consistent with domain knowledge. While the results are potentially scientifically interesting, the presentation of the manuscript is difficult to read, and it is ultimately challenging to assess the quality of this work. This is particularly the case as it relates to the assessment of the impact of meteorology on PM2.5 using these causal networks. I cannot recommend this manuscript for publication in GMD at this time.

More specific comments are below:

References and Introductory Motivation

It is not clear from the manuscript what problem the authors are intending to solve with this work. They mention the need to “overcome the high dimensionality challenge” on line 70, but there are plenty of existing tools for studying the high dimensionality in atmospheric composition, chiefly among them chemical transport models. What does this work provide that more traditional simulation experiments cannot?

Is it instead a demonstration of a new technique in atmospheric pollution research? If so, the authors should state this and demonstrate its utility in relationship to existing knowledge.

The references are quite sparse for this manuscript. Additional background and motivational clarity should include more details of previous applications of causality in the geosciences (e.g. Ebert-Uphoff, Imme, and Yi Deng. “Causal Discovery for Climate Research Using Graphical Models.” *Journal of Climate*).

Materials and Methods

Section 2.1

The description of the data in this section was inadequate for assessing the quality of the research in this work. How often were measurements taken? How was averaging done? What instruments were used? Was data quality assessed in any way? Is there a DOI or citation appropriate for any of the data used?

Section 2.2

The software used (e.g. PAST, EVIEW) should be appropriately cited.

Line 139: What is this threshold for, how is it selected, and how is it calculated?

Section 2.3

This section outlines a set of methods unfamiliar to the majority of the geoscientific modelling community. I suggest the authors include more details and relevant citations for the broader community should they be inclined to dig deeper into this sort of analysis. At the very least, this section should be carefully edited for clarity. The large number of parenthetical elements

throughout the section make it challenging to parse what is actually being stated. The three sentences within a parenthetical statement on lines 150-155 are emblematic of this.

Line 149: Detrending a dataset does not always correct for non-stationarity. Please cite a reference for its appropriateness here.

Line 157: What is a lag used for in this case?

Lines 345-348: It isn't clear in the text why this basic description of a directed graph is in the middle of the section on Granger Causality?

Section 2.4

The language regarding trophic coherence and sources/sinks is unreferenced and thus assumed to be an innovation of this work. All cities in this region are known to be atmospheric sources of PM2.5 due to anthropogenic and natural processes. The discussion here and elsewhere in the manuscript of only some cities classified as sources of PM2.5 is at odds with reality and should be clarified further.

While trophic coherence has been shown to provide interesting results for the analysis of food webs, it's not clear from the text that this is the most appropriate method for assessing important vertices in a causal graph for PM2.5. Why was it used here?

On line 39 in the abstract, the authors claim that winter is the most coherent of the seasons. They attribute this to meteorological features like wind speeds and inversions. Table 2 shows that summer has nearly the same incoherence factor as winter, and yet this is not discussed at all in the manuscript. Given the vastly different meteorological features in summertime, does this influence the interpretation that meteorology is a driving factor?

Discussion

In general, this discussion seems incomplete and largely conjecture without appropriate referencing or analysis presented herein.

Lines 378-381: What in the previous analysis indicates that meteorological conditions and diurnal emissions from regional sources dominate? I acknowledge that this is known the be case throughout the field a priori, but it's not clear how this analysis in this manuscript leads to that conclusion.

Lines 399-409: How are these relationships known without detailed trajectory modeling? The fact that the causal networks are consistent with known transport mechanisms is interesting but does not provide evidence for these sweeping assessments of pollutant transport.

Lines 410-415: There is not evidence presented in this manuscript that this transport mechanism is attributable to PM2.5 variability in the Northern UK.

Conclusions

Line 423-425: The authors haven't shown any results related to predicting future PM2.5 between cities.

Line 439-431: This work is not the first to demonstrate that meteorology drives much of the variability in PM2.5. It is not clear from this manuscript how connections beyond conjecture can be made to meteorological variability and PM2.5.

Minor Comments

Line 78: "Atmospheric particulate matters" should be "Atmospheric particulate matter"

Figure captions should contain more detail regarding the content of the figured. For example in Figure 2, what do edge thicknesses correspond to?

Line 232: "%50" should be "50%".

Line 393: "Primar_y" should be "Primary".