Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-342-RC2, 2020

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Dynamic Complex Network Analysis of $PM_{2.5}$ Concentrations in the UK using Hierarchical Directed Graphs (V1.0.0)" *by* Parya Broomandi et al.

Anonymous Referee #2

Received and published: 24 April 2020

The authors have used two statistical techniques to look at how in particulate matter (PM) concentration in different cites over the UK are linked. The methods they use are an undirected correlation method and a Granger causality network. The core point they argue to explore is the groupings of cities within the data whether a time-series in one could predict another. The data they use is a 1-year subset of freely available UK PM data for 15 sites. This has been considered with emissions data from the UK National Atmospheric Emission Inventory (NAEI).

I have concerns about the possible artefacts in the analysis and even the scientific validity of the results due to: (1) the small subset of data used (e.g. 15 sites in 14 cities),

Printer-friendly version

Discussion paper

(2) the small temporal period considered (just 1 year), (3) small about of parameters considered (solely PM observations and a national inventory of emissions), (4) the lack of consideration of known variables that influence PM (e.g. meteorology), (5) the insufficient of placing this work in the context of existing related publications

Importantly, I am not convinced that authors have shown the efficacy of this method and how it could be used for the policy goals they motivate the study with. I have written some specific points below. However, this study would need to be greatly expanded, all points addressed, an general presentation/readability improved to be suitable for publication in GMD. I cannot recommend publication of this manuscript.

Many years of data are freely available for hundreds of sites across the UK. Why have only 15 sites from 14 locations been considered here? Why were these 15 sites (listed in the SI) chosen above the hundreds of others? Are the authors saying they are representative of the sites? Why only 15 sites? Is it due to limitations in a computer resource or does this approach not scale well? If this is the case could a representative average be taken of the sites in a city? Why has a mixture of types been used (e.g. Traffic urban, Background suburban, Background Urban, and Industrial Urban)? What impact could the choice of data set have on the conclusions? All of these issues need to be explored through sensitivity testing. This analysis must be expanded in breadth (# sites) and depth (e.g. years of data) if it is to be considered a case study or proof of concept is.

How many air quality events happened in this year? What does this approach show about these events? Often pollution events in the UK are synoptic in scale, so would affect multiple cities. It would be interesting to see what information could be drawn out from this analysis.

The discussion needs to be expanded to explain what the pros and cons of this approach are over the generally taken approaches (e.g. chemical transport models). Also consider this approach against other non-explicit approaches taken in this field

(e.g. Nowack et al 2020, Keller et al 2019). ML methods that consider much more input data (e.g. meteorology, emissions, chemistry) have already been demonstrated in a much more thorough way within this journal (e.g. Keller et al 2019). Also make it clear what new information that this approach will provide and why this would be advantageous above these other approaches or complementary to them.

Line 102 - This is sentence is far too strong. The authors have not demonstrated that this technique could be this useful.

“This is a timely study as strategic investments in national and local air quality monitoring networks require an evaluation on the usefulness, or not, of network design”

Line 112 - See earlier point about the limitations of data used.

Line 117 - Was this emissions inventory used for the same year? Given more details here.

Line 163 - What is the link to the broader picture here? If we accept that a background urban site in Manchester can be used to predict future concentrations in Preston, is the suggestion that something like this could be done for other sites and a national level of prediction gained at a computational cheap cost?

Line 424 - There is no supporting evidence given for this predictive capacity

Technical points

Title / other text - Some lines are highlighted blue. Why?

Table 2 - typo in units? (Kelvin metres instead of kilometres?)

Figure 1 - resolution needs to be improved. Most of UK coastline should be shown for context. Which map layer was used?

Figure 2 - resolution needs to be improved.

Figure 3 - resolution needs to be substantially improved. Axis labels are not large

enough or readable.

Line 9 - Specify the region over which the deaths occur. Europe? Global?

Line 32 - Have people not already been successful in doing this? (e.g. the climate science community - Nowack et al 2020)

Line 34 - “12 month”, “52 week” are used interchangeably in the manuscript. . . would a single phrasing (e.g. year-long) be easier for the reader?

Line 35 - Sentence does not scan. superfluous “two” and “a”?

Line 80 - Globally? The focus of the article is the UK, so you need to be more specific.

Line 112 - This time-split would be better placed in a table and in the SI. It is not a very readable way of presenting this information.

Line 116 - Was data downloaded for the same period. This sentence does not need to start with ”also”.

Line 393 - formating error of “y” in word primary.

Data availability - Add a data availability section. Include specific details on which version of data was used and how to access this.

References - remove full stops at starts of lines.

S1 list - please update the information to be in a table format. The current format is unwieldy, not read-friendly.

Citations

Keller, C. A. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geosci. Model Dev., 12, 1209–1225, https://doi.org/10.5194/gmd-12-1209-2019, 2019.

Nowack, P., Runge, J., Eyring, V. and Haigh, J.D., 2020. Causal networks for climate

model evaluation and constrained projections. Nature communications, 11(1), pp.1-11.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-342, 2020.

Printer-friendly version

Discussion paper