

Geometric and Statistical Methods for Aggregating Predictors in Pollutant Risk Models

Wesley S. Burr – [Trent University](#)



[wsburr](#)



[@wsburr](#)



wesleyburr@trentu.ca

Hwashin H. Shin – EHSRB, Health Canada & [Queen's University](#)

2018-08-27

Slides available at <http://bit.ly/isesisee18>

Multiple Pollutant Models

Problem: Exposure assessment is challenging because "air pollution" is a complex mixture, not a static construct

- historical models have either used single pollutants, or handwaved vigorously at the inclusion of multiple pollutants
- obvious dependencies among pollutants, up to and including significant correlations and coherences
- topic of active research among a number of research groups (e.g., HEI-funded projects in the USA, LSH&TM, etc.)

Approach #1: Geometric

Genesis:

- air pollutants are correlated
- the correlated structure interferes with the linearity of the model framework and estimation
- why not try to leverage the correlation?

Sketch of Algorithm

- begin by estimating two acute risk models, one for each pollutant, accounting for the usual confounding variables, obtaining risks $\langle \phi_{01} \rangle$ and $\langle \phi_{02} \rangle$
- estimate correlation (angle) between 1st and 2nd pollutant, $\langle \phi_{12} \rangle$

Then ... in $\langle \mathbb{R}^3 \rangle$:

- set pollutant risk vector $\langle \vec{x}_1 = \vec{i} + \tan(\phi_{01})\vec{z} \rangle$
- set pollutant risk vector $\langle \vec{x}_2 = \cos(\phi_{12})\vec{i} + \langle \sin(\phi_{12})\vec{j} + \tan(\phi_{02})\vec{z} \rangle \rangle$

Functionally ...

This approach is just treating time series realizations as vectors

- overall covariances interpreted as inner products, giving the structural issue consideration as a rotational angle
- magnitudes of individual models taken as vector magnitudes

So how do we extract a singular representative of risk?

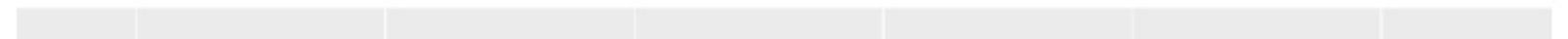
Aggregating

This approach gives a planar surface in (\mathbb{R}^3) for each pair of pollutants, and standardized model.

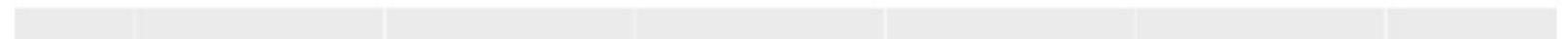
Options:

- standardize a direction, take slope of line in plane in that direction
- bisect the two pollutants
- choose a direction based on a secondary weighting scheme?

Results: Toronto, Ontario - 1984-2012 (NO₂ and O₃)



Results: National Risks (Canada) - 1984-2012 (N02 and 03)



Comments

This is very much still a work-in-progress.

- How to interpret the combined risk: an increase of N units of combined concentration?
- "User degrees-of-freedom": distinct choice of aggregate extraction
- Crazy idea?

Synthetic Lag

Recent development of a univariate distributed lag-like methodology for estimation of risk called synthetic lag naturally extends to multiple pollutants.

- Identify coherent (frequency domain analogue of correlated) frequency bands
- Phase align those bands of the pollutant individually with the phases of the health effect
- Equivalent to decomposing pollutant series and lagging individual series

Burr, Shin and Takahara (2018) Synthetically lagged models. Statistics & Probability Letters. DOI:10.1016/j.spl.2018.07.008

Extending Synthetic Lag

- Methodology naturally extends to arbitrary series
- Identify bands from pollutant + health effect series, but then ...
 - use the pollutants as inputs
 - de-correlate pollutants by selectively aligning or masking bands, respectively
- requires user choice of "primary" and "secondary" pollutants
- allows traditional naive multiple pollutant models (e.g., GAM with two pollutant series)
- interpretation remains complex

Interpretation?

So how might we interpret this approach?

- primary pollutant can be interpreted as traditionally (albeit with the phase alignment; so in a distributed lag sense)
- secondary pollutant - "after accounting for"?

Results: Toronto, Ontario - 1984-2012 (Ozone)

Results: Toronto, Ontario - 1984-2012 (N02)

Future Work

- methodologically, both approaches "work"
- interpretation remains complicated!
- better, worse, comparison to other current work?
- common components - worth extracting? (other talks in this session)



If you're interested, let me know - I'd love to talk more about our work.

- Contact me: [Email](#) or [Twitter](#)
- Slides created via the R package [xaringan](#) by Yihui Xie
- Slides at <http://bit.ly/isesisee18>

This research was partially funded under a CESI/CARA grant via the EHSRB of Health Canada.