Functional Data Analysis Model for Air Quality and its Impact on Health

Prof. Jeanne Houwing-Duistermaat (SoM), Dr Haiyan Liu (SoM), Dr Georgios Aivaliotis (SoM), Dr Leonid Bogachev (SoM).

Project partner(s): Prof. Sujit Sahu (University of Southampton)

Contact email: H.Liu1@leeds.ac.uk

During the last decade, more evidence of the detrimental effects of air pollution on human health has been evolving. Therefore, modelling the air pollution concentrations (e.g. NO2, O3, PM10, and PM2.5), in conjunction with other risk factors on health, is of great importance. Functional data analysis (FDA) is a powerful recent approach to modelling temporal observations, which is complementary to the usual time series techniques. The purpose of this project is to develop FDA methodology in the air pollution context, with a particular focus on the impact of air pollutants on health.

When building models for health outcomes using multiple temporal and static predictors, it is important to ensure a correct interpretation of the results to understand the correlation among the predictors. The relationship between various factors (e.g. urban/rural, summer/winter, day/night, etc.) and the pollutant concentrations will be studied by extending the functional linear model from binary to categorical covariates. Methodology for building the optimal model will be developed by adapting and extending functional linear models by including multiple temporal and static predictors.

Air quality and health are two of the most pressing issues facing our society. We are in a unique position at Leeds to answer important unresolved questions about the relationships between pollutants, the distribution of various pollutants across regions, and the influence of pollutants on health outcomes. The research topic has immediate policy-relevant potential, and the prospective PhD students may be expected to produce several papers with at least one being suitable for submission to a high impact journal.

Objectives:

The PhD candidate will work on the development of new FDA models such as functional dynamical correlation models and functional linear regression models, with the aim to investigate correlation between pollutants, the relationship between factors (such as regionality and weather conditions), and joint impact of pollutants and other factors on health.

Potential for high impact outcome

Air quality and climate change are two of the most pressing issues facing our industrialised society. We are in a unique position at Leeds to answer important unresolved questions about how

expected aerosol composition changes will influence climate and air quality. The research topic has immediate policy-relevant findings, and we therefore anticipate the project generating several papers with at least one being suitable for submission to a high impact journal.

Training

The successful PhD student will work under the supervision within the Department of Statistics and the Leeds Institute for Data Analytics (LIDA). This project provides a high-level specialist training in (i) development and use of FDA models; (ii) analysis and interpretation of inference from statistical research of air quality data, including impact on health; (iii) methods of variable/model selection. Supervision will involve weekly meetings between supervisors and the student as well as regular meetings with other partners (such as the Leeds City Council). The PhD candidate will have access to a broad spectrum of training workshops put on by the Faculty that includes an extensive range of training in theory development, numerical modelling, and data analysis.

Student profile:

The successful PhD candidate should have a strong interest in and a flair for statistical modelling of health and environmental problems; a solid background in mathematics and statistics; and familiarity and experience with programming in R and statistical computing in general.

CASE Partner

Negotiations with the Leeds City Council (LCC) as a potential CASE partner are in progress.

References

- Gertheiss, J., Maity, A., & Staicu, A.M. (2013). Variable selection in generalized functional linear models. *Stat* **2**, 86-101, https://doi.org/10.1002/sta4.20
- Liu, S., Zhou, Y., & Palumbo, R. (2016). Dynamical correlation: A new method for quantifying synchrony with multivariate intensive longitudinal data. *American Psychological Association* **21**, 291-308, https://doi.org/10.1037/met0000071
- Ramsay, J.O. & Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer.
- Sahu, S.K. (2018). Air pollution estimates for England and Wales during 2007-2011. Webinar hosted by the Royal Statistical Society on February 21, 2018. Available online at: http://www.soton.ac.uk/~sks/pollution_estimates/