A model for collocation uncertainty of atmospheric profiles

Maria Franco-Villoria*
University of Torino, Torino, Italy - maria.francovilloria@unito.it

Rosaria Ignaccolo
University of Torino, Torino, Italy - rosaria.ignaccolo@unito.it

Alessandro Fassò
University of Bergamo, Bergamo, Italy - alessandro.fasso@unibg.it

Abstract

Understanding collocation mismatch is particularly relevant for atmospheric profiles obtained by radiosondes, as the balloons containing the measuring instruments tend to drift uncontrollably from their initial launch position. We propose a point based formulation of an heteroskedastic functional regression model that includes a trivariate smooth function to account for time and space mismatch. Results from a case study where we model collocation error of atmospheric pressure show that model fitting is improved once heteroskedasticity is taken into account.

Keywords: Functional linear model; heteroskedasticity; generalized additive models; mixed models.

1. Introduction

Uncertainty of atmospheric thermodynamic variables is a key factor in assessing uncertainty of global change estimates given by numerical prediction models (Thorne et al. (2013)). Data, e.g. atmospheric pressure, temperature or water vapour, are gathered by high technology remote instruments such as radiosondes. An important source of uncertainty is related to the collocation mismatch in space and time among different observations; this is nothing but the difference between the measurements obtained from two instruments that are meant to measure the same environmental variable. It is important then to understand collocation mismatch and how this may depend on potential covariates. Data, recorded at different values of height as the radiosonde goes up into the atmosphere, can be considered as functional observations. This kind of functional data are usually modelled as functions depending only on height. The radiosonde balloons, however, drift away in the atmosphere resulting in not necessarily vertical but three-dimensional (3D) trajectories. On the other hand, little reference is made in the literature to heteroskedasticity in a functional data context; the latter is important as it allows adjusting mean estimates for non-constant variability, on top of the fact that modelling the variance function itself is of interest to understand which covariates significantly affect it.

2. Methods

The work presented at the conference is based on the paper by Ignaccolo et al (2015) (see the published version for further details), that introduces a "point based" formulation of an heteroskedastic functional regression model by extending the work done in Fassò et al. (2014), where a unidimensional model for collocation uncertainty was considered. The proposed model includes a trivariate smooth function to account for time and space mismatch, along with potential covariates. Functional coefficients of both the conditional mean and variance are estimated by reformulating the model as a standard generalized additive model and subsequently as a mixed model. This reformulation leads to a double mixed model whose parameters are fitted using an iterative algorithm (following Ruppert et al. (2003)) that allows to adjust for heteroskedasticity. As a result, covariates estimates can be adjusted for non-constant variability and estimation of the functional mean is improved. Simultaneously the conditional variance is explicitly modelled, allowing to identify significant covariates.

3. Case study

The dataset used in this work is the same as in Fassò et al. (2014) where the interested reader can find further details. It consists of radiosounding profiles of atmospheric thermodynamic variables measured at two locations: the Howard University research site in Beltsville, Meryland, USA (39.054°, -76.877°, 88 m a.s.l.), which is also a GRUAN site (GCOS Reference Upper-Air Network, see *www.gruan.org* and Thorne et al. (2013)), and the U.S. National Weather Service operational site at Sterling, Virginia, USA (38.98°, -77.47°, 53 m a.s.l.). The two sites are sufficiently close (52 km line distance) to consider collocation mismatch between them, and represent a similar climate regime. Data were converted to functional observations using penalized cubic B-splines with knots regularly spaced every 50 m and penalty parameter $\lambda = 1$. An illustration of the data can be seen in Figure 1. We model collocation error of atmospheric pressure in terms of space (longitude, latitude), time mismatch (calendar time, flight duration difference) and a number of meteorological covariates: temperature, relative humidity, water vapor mixing ratio and orthogonal wind components from both collocated radiosondes. Results show that model fitting is improved once heteroskedasticity is taken into account; the 95% confidence bands for the estimated functional coefficients become generally narrower and the functional coefficients associated to meteorogical covariates change in shape and magnitude. AIC criteria indicates that a model with trivariate functions that take into account the interaction among longitude, latitude and height and as well among distance and height is preferred to a model where all the components act additively.

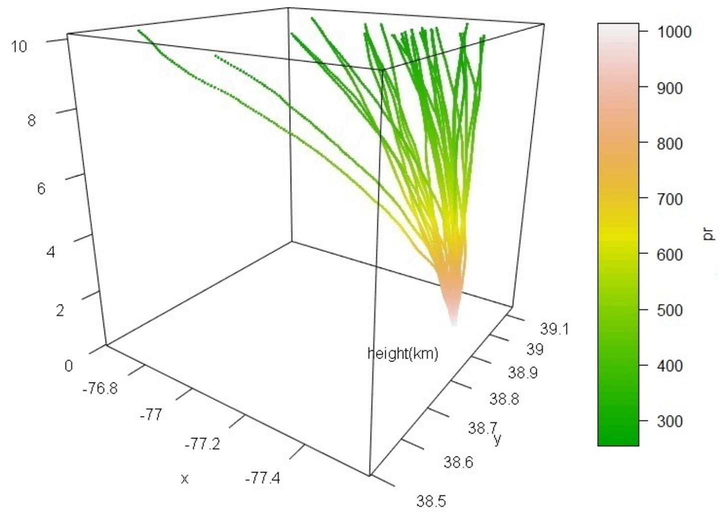

Figure 1: 3D pressure (pr) atmospheric profiles. Each curve represents a different launch at the Sterling site

4. Conclusions

The modelling strategy describes both conditional mean and variance as a sum of a 3D functional term and some unidimensional functional regression components. This results in great flexibility as seen in the application to collocation uncertainty of atmospheric termodynamic profiles. The reformulation of the model as a double mixed model, with the implementation of an iterative algorithm, allows to handle the impact of covariates on conditional uncertainty by means of functional heteroskedasticity. The new 3D component is shown to improve model fitting with respect to the purely undimensional model previously considered by Fassò et al. (2014) when modelling collocation mismatch of atmospheric pressure. The resulting model includes a number of terms that take into account time and space for the two collocated measurements. These effects are not linear but they smoothly change in shape along vertical direction and horizontal distance. In addition, the small unexplained collocation uncertainty changes in magnitude as explained by the heteroskedastic 3D component.

References

Fassò, A., Ignaccolo, R., Madonna, F., Demoz, B. & Franco-Villoria, M. (2014). Statistical modelling of collocation uncertainty in atmospheric thermodynamic proles. Atmospheric Measurement Techniques, 7, pp.1803-1816, DOI: 10.5194/amt-7-1803-2014.

Ignaccolo, R., Franco-Villoria, M. & Fassò, A. (2015). Modelling collocation uncertainty of 3D atmospheric profiles. Stochastic Environmental Research and Risk Assessment, 29(2), 417–429.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). Semiparametric Regression. Cambridge University Press, New York

Thorne, P. W., Vömel, H., Bodeker, G. et al. (2013). GCOS reference upper air network (GRUAN): Steps towards assuring future climate records. AIP Conference Proceedings 1552: 1042–1047.
doi: http://dx.doi.org/10.1063/1.4821421