GRASPA 2015
BARI 15-16 JUNE

Collocation uncertainty in climate monitoring

M. Franco-Villoria[1,*], R. Ignaccolo[1], A. Fassò[2], F. Madonna[3] and B.B. Demoz[4]

[1] *Department of Economics and Statistics, University of Torino, Italy; maria.francovilloria@unito.it, rosaria.ignaccolo@unito.it*
[2] *Department of Management, Information and Production Engineering, University of Bergamo, Italy; alessandro.fasso@unibg.it*
[3] *CNR-IMAA, Tito Scalo, PZ, Italy; madonna@imaa.cnr.it*
[4] *Department of Physics and Astronomy, Howard University, Washington, DC, USA; bbdemoz@Howard.edu*
[*] *Corresponding author*

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 heteroskedastic functional regression model capable of explaining the relationship between collocation uncertainty and a set of environmental factors, height and distance between imperfectly collocated trajectories. Along this line, a five-fold decomposition of the total collocation uncertainty is proposed, giving both a profile budget and an integrated column budget. Considering the profiles as three-dimensional trajectories, we extend the model to include a trivariate smooth function that accounts for time and space mismatch. Results from a case study where we model collocation error of relative humidity and atmospheric pressure show that model fitting is improved once heteroskedasticity is taken into account.*

Keywords. *Functional linear model; Heteroskedasticity; Generalized additive models; Mixed models; Uncertainty budget*

---

1 Introduction

Uncertainty of atmospheric thermodynamic variables is a key factor in assessing uncertainty of global climate change estimates given by numerical prediction models [4]. 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; i.e. the difference between the measurements obtained from two instruments (at two nearby places) 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 can be modelled as functions depending only on height. Alternatively, they can be considered as three-dimensional (3D) trajectories, as the radiosonde balloons drift away in the atmosphere resulting in not necessarily vertical profiles. 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 papers by Fassò et al. (2014) [1] and Ignaccolo et al. (2015) [2]. In the first part, an heteroskedastic functional regression approach is proposed to model the vertical profiles of collocation uncertainty for a climate variable, in relation to environmental factors, altitude of measurement and distance between trajectories. The error variance is not assumed to be constant but a function of some variables. To better understand random and systematic uncertainty (i.e. the uncertainty budget), a five-fold decomposition of the total collocation uncertainty is proposed, namely constant bias, reducible and irreducible environmental errors, sampling error and measurement error. This allows to obtain both a profile budget and an integrated column budget. In the second part, a "point based" formulation of the heteroskedastic functional regression model mentioned above is proposed. This 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 [3]) 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 consists of 32 pairs 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 [4]), 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.

On one hand, we focus on collocation of relative humidity and explain its profile uncertainty using time, altitude, coordinates, and the following environmental factors: water vapour mixing ratio, pressure, temperature and wind vector. This particular variable is of interest because humidity is known to have large forecast errors, even on small time and space scales. In this case, 85% of the total collocation uncertainty is ascribed to reducible environmental error, 11% to irreducible environmental error, 3.4% to adjustable bias, 0.1% to sampling error and 0.2% to measurement error. Moreover, the collocation error has an adjustable constant bias amounting to 3.4% of the total collocation uncertainty. The model performs better below 3000m of altitude and, globally, it misses only 11.4% of the collocation uncertainty for relative humidity.

The model used for relative humidity [1] considers profiles as being vertical, i.e. as functions of height only. Instead, for modelling atmospheric pressure, we use a point based version of the heteroskedastic functional regression model [2], where profiles are considered as functions with three-dimensional domain (longitude, latitude and height). In this case, 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 vapour 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.

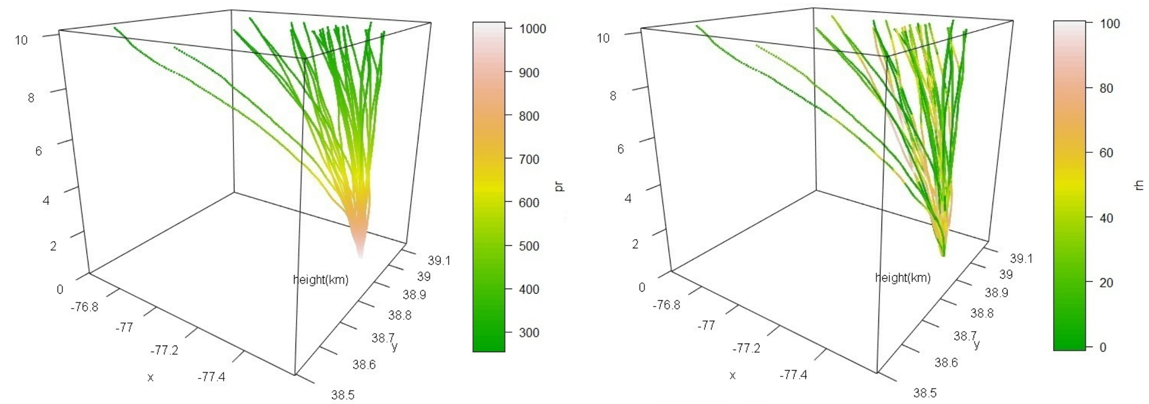

Figure 1: 3D pressure (left) and relative humidity (right) atmospheric profiles. Each curve represents a different launch at the Sterling site.

4 Conclusions

This paper presents general statistical methodology for modelling collocation uncertainty of atmospheric thermodynamic variables as introduced in Fassò et al. (2014) [1] and Ignaccolo et al. (2015) [2]. The functional regression model proposed for vertical profiles takes into account heteroskedasticity and allows the decomposition of total uncertainty budget up to five different components, namely constant bias, reducible and irreducible environmental errors, sampling error and measurement error. Moreover, the conditional uncertainty may be computed for any set of environmental conditions, providing, inter alia, more information about the factors determining the collocation uncertainty. The proposed method is self-assessing, in the sense that it is able to consider the information content of the data for the model and evaluate the size of the sampling error with respect to the other uncertainty components. In the case study considered, the collocation drift for relative humidity was found to be strongly dependent on the direction of air mass advection and not on the distance between the paired trajectories. It can be concluded that the collocation uncertainty of relative humidity is related to physical quantities and, in principle, could be reduced by inclusion of auxiliary information. The extended point based model considers the profiles as functions with three-dimensional domain, describing 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) [1] 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

[1] Fassò, A., Ignaccolo, R., Madonna, F., Demoz, B. & Franco-Villoria, M. (2014). Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles. *Atmospheric Measurement Techniques* **7**, 1803–1816.

[2] 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.

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

[4] 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