GATNDOR Topic Collocation Uncertainty in Vertical Profiles: Statistical Functional Regression Approach

Alessandro Fassò*

Joint work with: B. Demoz, R. Ignaccolo, F.Madonna.

-

* University of Bergamo - DIIMM - alessandro.fasso@unibg.it

De Bilt, **ICM-5**, 26 February 2013

RegioneLombardia

Scientific questions

We would like to answer the following questions:

1. is the co-location uncertainty related to environmental factors?
2. is the co-location uncertainty related to the paired trajectories distance?
3. Is the co-location uncertainty related to height?
4. Is uncertainty a static or dynamic concept?
5. Are above points valid for all ECV?

Scientific questions

We would like to answer the following questions:

1. is the co-location uncertainty related to environmental factors?
2. is the co-location uncertainty related to the paired trajectories distance?
3. Is the co-location uncertainty related to height?
4. Is uncertainty a static or dynamic concept?
5. Are above points valid for all ECV?

Scientific questions

We would like to answer the following questions:

1. is the co-location uncertainty related to environmental factors?
2. is the co-location uncertainty related to the paired trajectories distance?
3. Is the co-location uncertainty related to height?
4. Is uncertainty a static or dynamic concept?
5. Are above points valid for all ECV?

Scientific questions

We would like to answer the following questions:

1. is the co-location uncertainty related to environmental factors?
2. is the co-location uncertainty related to the paired trajectories distance?
3. Is the co-location uncertainty related to height?
4. Is uncertainty a static or dynamic concept?
5. Are above points valid for all ECV?

Scientific questions

We would like to answer the following questions:

1. is the co-location uncertainty related to environmental factors?
2. is the co-location uncertainty related to the paired trajectories distance?
3. Is the co-location uncertainty related to height?
4. Is uncertainty a static or dynamic concept?
5. Are above points valid for all ECV?

Introduction

"... The primary goals of GRUAN are to **provide vertical profiles** of reference measurements suitable for reliably detecting changes in global and regional climate on decadal time scales" (GRUAN Manual V7).

- We consider here an empirical approach to uncertainty analysis with no reference to a priori metadata but relying on data by means of a statistical modelling approach.
- Stochastic structure of vertical profiles is important from the statistical point of view not only for facing the co-location problem but also for uncertainty analysis and decomposition of vertical profiles in general.

Introduction

"... The primary goals of GRUAN are to **provide vertical profiles** of reference measurements suitable for reliably detecting changes in global and regional climate on decadal time scales" (GRUAN Manual V7).

- We consider here an empirical approach to uncertainty analysis with no reference to a priori metadata but relying on data by means of a statistical modelling approach.
- Stochastic structure of vertical profiles is important from the statistical point of view not only for facing the co-location problem but also for uncertainty analysis and decomposition of vertical profiles in general.

Outline

1. Heteroskedastic functional regression co-location model
2. Application to relative humidity from Beltsville-Sterling radiosonde

The Co-location Model

Let y denote the measurement of a physical quantity, e.g. an ECV, along a trajectory through the atmosphere.

$$y\left(s,t\right), s=\left(lat,lon,h\right), \quad h\geq h_0, \quad t\geq t_0$$

The space-time vertical trajectory can then be described by the parametric representation $h\longrightarrow\left(s_h,t_h\right)$ for $h_0\leq h\leq h_1$.

$$y\left(\cdot\right)=y\left(h\right)$$

We consider a vertical profile as a single object described by a smooth function:

$$\mu\left(h\right), \quad h \geq h_0.$$

and the standard measuerment error decomposition of observation profile launched at time t_j from place s_j, $j = 1, ..., n$, is given by a random function

$$y_j\left(\cdot\right) = \mu_j\left(\cdot\right) + \varepsilon_j\left(\cdot\right)$$

where $\mu\left(\cdot\right)$ is the "true" smooth profile and $\varepsilon\left(\cdot\right)$ is the zero mean measurement error with variance $\sigma_\varepsilon^2\left(\cdot\right)$.

Suppose we are comparing two instruments, e.g. radiosondes, at the same height and giving measurements y and y^0 respectively. Hence we have

$$\Delta y(\cdot) = y(\cdot) - y^0(\cdot) = \Delta\mu(\cdot) + \Delta\varepsilon(\cdot)$$

where

$$\Delta\mu = \mu - \mu^0$$

is the co-location drift and

$$\Delta\varepsilon = \varepsilon - \varepsilon^0$$

is the co-location measurement error, with

$$Var(\Delta\varepsilon) = \sigma_\varepsilon^2 + \sigma_{\varepsilon^0}^2 = 2\sigma_\varepsilon^2.$$

The Heteroskedastic Functional Regression Model for Co-location

$$\Delta\mu(\cdot) = \beta_0 + \beta(\cdot)' x(\cdot) + \omega(\cdot)$$

where

- β_0 is the inter-site constant co-location bias.
- $x(\cdot)$ is a set of functional environmental factors.
- $\omega(\cdot)$ is an heteroskedastic error component: namely

$$\sigma^2_\omega(\cdot|x) = Var(\omega(\cdot)|x)$$

For example a log-linear component

$$\sigma^2_\omega(\cdot|x) = \exp\left(\gamma(\cdot)' x(\cdot)\right).$$

The Global Uncertainty Profile

$$U(\cdot) = Var(\Delta y(\cdot))$$

Its decomposition extends the heteroskedastic uncertainty decomposition of Fassò et al. (2003) and is given by

$$Var(\Delta y(\cdot)) = U(\cdot) = U_{\Delta\mu}(\cdot) + \sigma_{\varepsilon}^{2}$$

where, $U_{\mu}(\cdot) = Var(\mu(\cdot))$ defines the apportionable co-location error.

Uncertainty Decomposition

$$U_{\Delta\mu}(\cdot) = E\left((\Delta\mu(\cdot))^2\right) = \beta_0^2 + U_x(\cdot) + U_\omega(\cdot) + U_{\hat{\beta}}(\cdot) .$$

where

- β_0^2 is the inter-site error
- $U_x + U_\omega$ is the environmental error
- $U_{\hat{\beta}}$ is the sampling error.

- In particular U_x is the drift uncertainty and is given by

$$U_x\left(\cdot\right) = \hat{\beta}'\left(\cdot\right)\Sigma_x\left(\cdot\right)\hat{\beta}\left(\cdot\right)$$

- The second component of the environmental error, namely U_ω, cannot be reduced by observing $x\left(\cdot\right)$ and is given by

$$U_\omega\left(\cdot\right) = E_x\left(\hat{\sigma}^2_\omega\left(\cdot|x\right)\right)$$

- The sampling error is given by

$$U_{\hat{\beta}}\left(\cdot\right) = E\left(x\left(\cdot\right)'\Sigma_{\hat{\beta}}\left(\cdot\right)x\left(\cdot\right)\right)$$

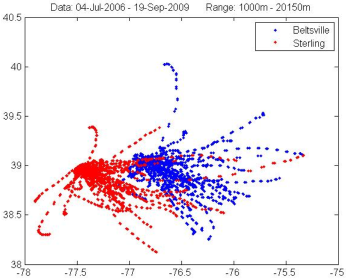
where $\Sigma_{\hat{\beta}}\left(\cdot\right) = Var\left(\hat{\beta}\left(\cdot\right)|x\right)$ is the estimation functional variance covariance matrix of $\hat{\beta}\left(\cdot\right)$.

Average Uncertainty

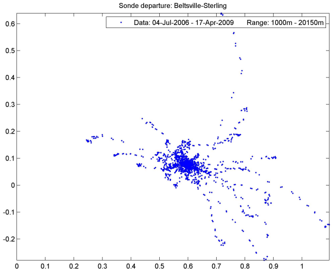
In order to get a simple global uncertainty decomposition, we sumarize the above uncertainty profiles by integrating over the vertical profile. This is given by

$$\bar{U} = \beta_0^2 + \bar{U}_x + \bar{U}_\omega + \bar{U}_{\hat{\beta}} + \sigma_\varepsilon^2$$

The Beltsville Case Study



n=49 trajectories



Mutual departure trajectories

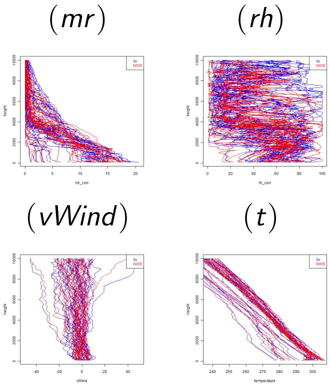


Figure: Data profiles

Modelling of relative humidity

- The response variable used here is relative humidity and all the data from the other radiosonda and all but relative humidity from the "co-located" radiosonda are used as explanatory factors.
- The resulting co-location error analysis corresponds understanding the variability of relative humidity for fixed water vapour content in dry air mass.
- Information on location, distance and ECV's have been used as regressors for both μ and σ^2 and selected using suitable statistical model selection techniques.

Modelling of relative humidity

- The response variable used here is relative humidity and all the data from the other radiosonda and all but relative humidity from the "co-located" radiosonda are used as explanatory factors.
- The resulting co-location error analysis corresponds understanding the variability of relative humidity for fixed water vapour content in dry air mass.
- Information on location, distance and ECV's have been used as regressors for both μ and σ^2 and selected using suitable statistical model selection techniques.

Modelling of relative humidity

- The response variable used here is relative humidity and all the data from the other radiosonda and all but relative humidity from the "co-located" radiosonda are used as explanatory factors.
- The resulting co-location error analysis corresponds understanding the variability of relative humidity for fixed water vapour content in dry air mass.
- Information on location, distance and ECV's have been used as regressors for both μ and σ^2 and selected using suitable statistical model selection techniques.

Co-location model (Beltsville - Sterling)

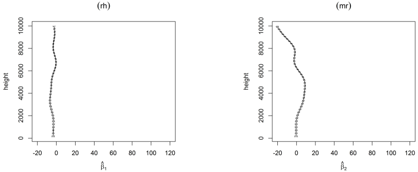
Mean component $R^2 = 88\%$

$$\Delta rh(\cdot) = 3.37 + \hat{\beta}_1(\cdot)\, rh^0(\cdot) + \hat{\beta}_2(\cdot)\, mr^0(\cdot) + \hat{\beta}_3(\cdot)\, \Delta mr(\cdot) + \hat{\omega}(\cdot) + \Delta\varepsilon(\cdot)$$

Skedastic component

$$\log \sigma^2_\omega(\cdot|x) = \underset{(0.26)}{1.6} + \hat{\gamma}_1(\cdot)\, p^0(\cdot) + \hat{\gamma}_2(\cdot)\, t^0(\cdot) + \hat{\gamma}_3(\cdot)\, rh^0(\cdot) + {} + \hat{\gamma}_4(\cdot)\, \Delta p(\cdot) + \hat{\gamma}_5(\cdot)\, \Delta lon(\cdot) + \hat{\gamma}_6(\cdot)\, \Delta mr(\cdot) + {} + \hat{\gamma}_7(\cdot)\, \Delta u(\cdot) + \hat{\gamma}_8(\cdot)\, \Delta v(\cdot)$$

(rh) (mr)



(Δmr)

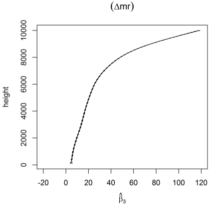


Figure: Beta functions for relative humidity co-location drift.

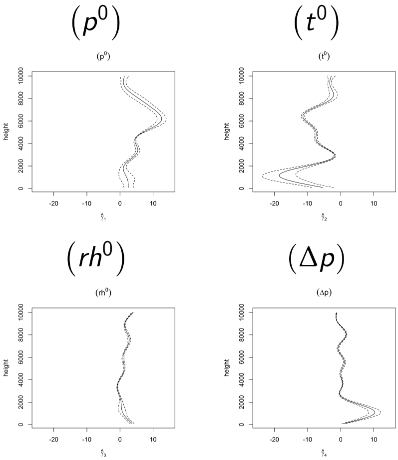
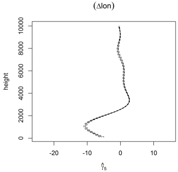
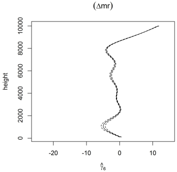


Figure: Gamma functions for relative humidity co-location error ω^2, Figure 1 of 2.

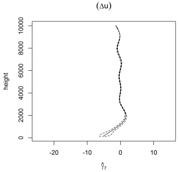
(Δlon)



(Δmr)



(Δu)



(Δv)

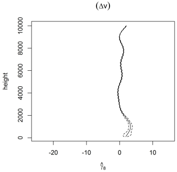


Figure: Gamma functions for relative humidity co-location error ω^2, Figure 2 of 2.

Uncertainty Component Profiles

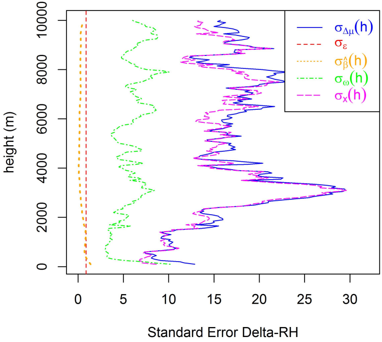


Figure: Square-root uncertainty budget $\left(\sqrt{U}\right)$ for relative humidity.

Average Uncertainty Components

Source		$\bar{U}$	$\bar{U}\%$	$\sqrt{\bar{U}}$
Natural variability	$\Delta\mu$	321		17.9
Inter-site error	β_0^2	14	4.3%	3.37
Environmental Error (reducible)	x	269	83.6%	16.4
Environmental Error (irreducible)	ω^2	38	11.7%	6.1
Sampling error	$\hat{\beta}$	0.5	0.1%	0.7
Measurement error	$\Delta\varepsilon$	0.8	0.2%	0.9

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance? For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height? According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept? Yes and the skedastic function $\sigma_\omega^2(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance? For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height? According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept? Yes and the skedastic function $\sigma_{\omega}^{2}(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance?
     For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height?
     According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept?
     Yes and the skedastic function $\sigma_{\omega}^{2}(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance? For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height? According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept? Yes and the skedastic function $\sigma_{\omega}^{2}\left(\cdot|x\right)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance? For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height? According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept? Yes and the skedastic function $\sigma_{\omega}^{2}(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance?
     For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height?
     According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept?
     Yes and the skedastic function $\sigma^2_\omega(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Conclusions

- This approach is a viable method for detaild uncertainty decomposition
- For example in our case study
  1. is the co-location uncertainty related to environmental factors? These data show strong nonlinear correlation of *rh* on other variables
  2. is the co-location uncertainty related to the paired trajectories distance?
     For these (limited spatial coverage and anisotropic) data, we found an EST-WEST effect on the skedastic component.
  3. Is the co-location uncertainty related to height? According to data the larger uncertainty is about 3'000m
  4. Is uncertainty a static or dynamic concept? Yes and the skedastic function $\sigma_\omega^2(\cdot|x)$ may account for its variation
  5. Are above points valid for all ECV?

Thanks for your attention