Atmospheric
Measurement
Techniques
Open Access
Discussions

Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles

A. Fassò[1], R. Ignaccolo[2], F. Madonna[3], and B. B. Demoz[4]

[1]University of Bergamo, Dalmine, BG, Italy
[2]University of Turin, Torino, TO, Italy
[3]CNR-IMAA, Tito Scalo, PZ, Italy
[4]Howard University, Washington, D.C., USA

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD

6, 7505–7533, 2013

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

A. Fassò et al.

Abstract

The uncertainty of important atmospheric parameters is a key factor for assessing the uncertainty of global change estimates given by numerical prediction models. One of the critical points of the uncertainty budget is related to the collocation mismatch in space and time among different observations. This is particularly important for vertical atmospheric profiles obtained by radiosondes or LIDAR.

In this paper we consider a statistical modelling approach to understand at which extent collocation uncertainty is related to environmental factors, height and distance between the trajectories. To do this we introduce a new statistical approach, based on the heteroskedastic functional regression (HFR) model which extends the standard functional regression approach and allows us a natural definition of uncertainty profiles. Moreover, using this modelling approach, a five-folded uncertainty decomposition is proposed. Eventually, the HFR approach is illustrated by the collocation uncertainty analysis of relative humidity from two stations involved in GCOS reference upper-air network (GRUAN).

1 Introduction

While global availability of profiling measurements of atmospheric parameters is increasing, full exploitation of these measurements is still far from being achieved. In fact the lack of an extensive effort, at a global scale, aimed at coordinating the operation of available measurement stations towards harmonized and traceable observations, uncertainty included, has hampered exploitation of the data. GRUAN (GCOS Reference Upper-Air Network, www.gruan.org) is a network aiming at rectifying this issue in order to provide traceable measurements of Essential Climate Variables (ECVs), namely pressure, temperature, water vapour, wind and aerosol, with their uncertainty, over long term period of time (GCOS-112, 2007). The quantification of the uncertainty budget is one of the key priorities for GRUAN (Seidel et al., 2011).

Instrumental contribution to the error budget (random and systematic uncertainties) have been investigated for various sensors, e.g. Raman lidars (e.g. Whiteman et al., 2001) or weather radars (e.g. O'Connor et al., 2005). On the other hand, one of the critical contributions to the uncertainty budget is related to the collocation mismatch in space and time among pairs of sensors. Although these different measurements are assumed nominally collocated there is a real physical separation between their actual measurement locations and timings, either they are both from ground based or one from ground based and one from satellite observation platforms.

Estimates of the representativeness error resulting from the effects of small scale turbulence have been performed in many cases, for example, for rawinsonde wind measurements (e.g. Frehlich and Sharman, 2004) or high-resolution radiosonde wind shear (Houchi et al., 2010).

However, there is a need for flexible statistical modelling assessing jointly the dynamic impact of both the imperfect collocation of atmospheric observations and environmental forcing factors on collocation uncertainty. The approach should be flexible enough to cover atmospheric processes characterized by regimes ranging from quasi-linearity (e.g. horizontally homogenous atmosphere) to non-linearity.

Radiosondes provide one of the primary data sources for vertical atmospheric profiles (Immler et al., 2010), but they may be affected by uncontrolled drift once they are launched (Seidel et al., 2011). Radiosonde data have been extensively used for a wide range of applications including intercomparison with ground-based and space-based remote sensing systems, atmospheric model evaluations, and studies of atmospheric variability. However, these studies have mainly assumed the radiosonde measurement to represent the atmospheric conditions over a wider area, as if they came from a fixed measurement location and neglecting the impact of their uncontrolled drift. In case of satellite validation, it is usually assumed that radiosondes are spatially collocated with the satellite field of view. Representativeness error can be minimized only if the validation is performed in homogeneous conditions. But quite often, the uncertainty introduced by representativeness dominates the error budget of the validation experi-

ment (e.g. Frehlich and Sharman, 2004). Uncontrolled radiosonde drift might also affect the evaluation of model data where the problem is the representativeness of observations, i.e. these data should represent the range of conditions influencing the model prediction.

5   Spatial collocation mismatch does not seem to play a big role in the radiance matching, due to large footprint characterizing these measurements. On the contrary, temporal collocation and time interpolation are critical to achieve these results due to the related vertical thermodynamic factors (Tobin et al., 2006).

The satellite validation community considers, as a priority, the availability of robust
10   collocation criteria that would increase the matches by a significant amount at an affordable cost due to data synergy. Appropriate collocation criteria are strongly required to combine different measurements, to potentially reduce the overall uncertainty on the atmospheric profile measurement (Tobin et al., 2006; Calbet et al., 2010). For example, in the former paper, temporal collocation and time interpolation are critical to
15   achieve good results, although collocation does not seem to play a big role in radiance matching.

In this study, we aim at two objectives. The first is introducing a general statistical modelling approach to understand the vertical profiles of collocation uncertainty for any climate variable, in relation to environmental factors, altitude of measurement and
20   distance between trajectories. The second objective is developing a illustrative example based on relative humidity data from ground rawinsonde measurements, which are made from two different locations at almost the same time. The case study is important because humidity is known to have forecast errors with large components at small scales. To do this, we rely on the statistical methods known as functional data
25   analysis, which dates back to the eighties, see e.g. Ramsay and Dalzell (1991) or the primer of Ramsay and Silverman (2005). In the last decade these methods have been increasingly developed and used in various scientific areas and especially in life and environment observation. For example, in climatological studies Ruiz-Medina and Espejo (2012) proposed spatial interpolation of functional ocean surface temperature

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

and, in environmental ruling, Ignaccolo et al. (2013) proposed zoning according to functional air quality data. Moreover, Sangalli et al. (2012) proposed functional regression for complex spatial configurations which are important, for example, in the study of hemodynamic forces, see Ettinger et al. (2013). In this paper developing the idea of Ignaccolo (2013), we propose the heteroskedastic functional regression (HFR) model, which extends the standard functional regression approach to cover for non constant functional conditional variance, as an effective approach to understand and decompose the uncertainty of the atmospheric thermodynamic profiles.

The rest of the paper is organized as follows: in Sect. 2 a new statistical approach, based on the HFR, is presented for modelling atmospheric thermodynamic profiles. Using this general modelling approach, conditional and global uncertainty profiles are defined and the corresponding total uncertainties computed in Sect. 3. In Sect. 4, collocation uncertainty is embedded in the HFR approach allowing a five-folded uncertainty decomposition. Section 5 illustrates the method using data from two North American stations involved in GRUAN, and focus on collocation uncertainty of relative humidity, which is interesting to model because is characterized by high vertical variability. Section 6 gives concluding remarks.

2 Modelling vertical profiles

Let y denote the measurements of a physical quantity, e.g. an ECV, along a trajectory through the atmosphere. A measurement at spatial point s and time t is denoted here by $y(s, t)$, where $s = (\text{lat}, \text{lon}, h)$, $h \geq h_0$ is the measurement height, $t \geq t_0$ is measurement time, while t_0 and h_0 are launch time and height. The space-time vertical trajectory can then be described by the parametric representation $h \rightarrow (s_h, t_h)$ for $h_0 \leq h \leq h_1$. In other words we represent data $y(s, t)$ as vertical profiles or functions $y(\cdot) = y(h)$. Using the functional data analysis (FDA) approach described e.g. by Ramsay and Silverman (2005), we consider a vertical profile as a single object described by a smooth function:

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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

According to standard measurement error decomposition, an observation profile, labeled by launch place and time (s_j, t_j), $j = 1, \ldots, n$, is given by a random function

$$y_j(\cdot) = \mu_j(\cdot) + b_j(\cdot) + \varepsilon_j(\cdot)$$

5 where $\mu(\cdot)$ is the "true" smooth profile, $b(\cdot)$ is the instrumental bias profile and $\varepsilon(\cdot)$ is the measurement error with zero mean, $E(\varepsilon(\cdot)) = 0$, and variance given by $\text{Var}(\varepsilon(\cdot)) = \sigma_\varepsilon^2(\cdot)$.

If $\mu_{s,t}(\cdot)$ or $\varepsilon_{s,t}(\cdot)$ are correlated over launching space (s) and time (t), we have space-time functional data models. In this paper we assume that measurements, conditionally on a set of forcing factors denoted by $x(\cdot)$, are independent random functions and we
10 focus on modelling their conditional mean and variance. In particular we are interested in modelling the effect of (functional) environmental factors $x(\cdot)$ by means of a functional trend model given by

$$\mu(\cdot) = \beta(\cdot)' x(\cdot) + \omega(\cdot) \tag{1}$$

where prime denotes matrix transposition. In Eq. (1) we assume that the trend is locally
15 linearly related to x but the global relation is not assumed linear.

Moreover the error $\omega(\cdot)$ is assumed to be an heteroskedastic component. The term heteroskedasticity derived from the ancient Greek language, means varying variance and, in this paper, we assume conditional heteroskedasticity, that is the conditional variance, namely $\sigma_\omega^2(\cdot|x) = \text{Var}(\omega(\cdot)|x)$, is assumed to be a linear or log-linear function
20 of x. The former case is given by

$$\sigma_\omega^2(\cdot|x) = \gamma(\cdot)' x(\cdot) \tag{2}$$

and the latter is given by

$$\sigma_\omega^2(\cdot|x) = \exp(\gamma(\cdot)' x(\cdot)). \tag{3}$$

The choice among these skedastic models is based on the data and both describe the uncertainty unaccounted by the local linear component. So Eqs. (1) and (2) or (3) define an heteroskedastic functional regression or HFR model.

In the sequel, when clear from the context, we will omit the function notation and, for example, we will denote the functional model of Eq. (1) by $\mu = \beta' x + \varpi$. Moreover, symbols $\widehat{\beta}$ and $\widehat{\gamma}$ denote estimates of β and γ based on historical data. In this paper we will use an estimation algorithm based on penalized regression splines where the smoothing parameters are chosen by restricted maximum likelihood estimation (REML) as implemented in the R package mgcv (Wood, 2004, 2011). In particular, Ignaccolo (2013) shows how HFR can be represented as a heteroskedastic semiparametric generalized additive model (GAM), and how the corresponding two stages estimation is obtained by iteratively using the mentioned penalized regression splines with REML smoothing parameters.

3 Uncertainty profiles

Assuming no instrumental bias, $b = 0$, and constant measurement error variance, we consider here the conditional uncertainty profile given by the following conditional mean squared error:

$$u(\cdot|x) = E((y - \mu)^2|x) = \sigma_\varpi^2(\cdot|x) + \sigma_\varepsilon^2. \tag{4}$$

This equation gives the uncertainty profile for each value of the forcing factors set x at height h, which is given by the conditional variance function $\sigma_\varpi^2(\cdot|x)$ plus the measurement error σ_ε^2 which is constant.

Summarizing the effects of environmental factors x, we get the global uncertainty profile $U(\cdot) = \mathrm{Var}(y(\cdot))$. Its decomposition extends the heteroskedastic uncertainty decomposition of Fassò et al. (2003) and is given by

$$U(\cdot) = U_\mu(\cdot) + \sigma_\varepsilon^2 \tag{5}$$

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

where, $U_\mu(\cdot) = \mathrm{Var}(\mu(\cdot))$ defines the natural variability profile. Moreover, as discussed in Appendix A1, the latter quantity is decomposed in three terms:

$$U_\mu(\cdot) = U_x(\cdot) + U_\omega(\cdot) + U_{\widehat{\beta}}(\cdot). \tag{6}$$

In this formula $U_x + U_\omega$ defines the environmental error. In particular U_x is the drift uncertainty and is given by

$$U_x(\cdot) = \widehat{\beta}'(\cdot)\Sigma_x(\cdot)\widehat{\beta}(\cdot), \tag{7}$$

and $\Sigma_x(\cdot) = \mathrm{Var}(x(\cdot))$ is the functional variance covariance matrix of $x(\cdot)$. In practice, this error could be reduced by observing the forcing factors $x(\cdot)$. The second component of the environmental error, namely U_ω, cannot be reduced by observing $x(\cdot)$. This is the average of the estimated skedastic function Eqs. (2) or (3), namely

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

with computation details shown in the Appendix A2. Finally the estimation uncertainty or sampling error is given by

$$U_{\widehat{\beta}}(\cdot) = E(x(\cdot)'\Sigma_{\widehat{\beta}}(\cdot)x(\cdot))$$

where $\Sigma_{\widehat{\beta}}(\cdot) = \mathrm{Var}(\widehat{\beta}(\cdot)|x)$ is the estimation functional variance covariance matrix of $\widehat{\beta}(\cdot)$. When the historical information used for estimation is large, $\Sigma_{\widehat{\beta}}$ is small and $U_{\widehat{\beta}}$ can be neglected, otherwise this type of random error is irreducible.

Note that, in Eq. (5), the functional object $U(\cdot)$ gives the uncertainty profile at height h irrespective of the particular value assumed by the forcing factors set x and corresponds to the usual variance estimate given by

$$S(h)^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y(t_j, h) - \overline{y}(h))^2$$

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

where $\overline{y}(h)$ is the mean of observations at height h and is usually computed grouping the data in vertical bins $h \pm \Delta h$, see e.g. Immler et al. (2010).

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

$$\overline{U} = \overline{U}_x + \overline{U}_\omega + \overline{U}_{\widehat{\beta}} + \sigma_\varepsilon^2 \tag{8}$$

where $\overline{U}$ is the profile average, id est

$$\overline{U} = \frac{1}{h_1 - h_0} \int\limits_{h_0}^{h_1} U(h)dh$$

and $\overline{U}_x$, $\overline{U}_{\widehat{\beta}}$ and $\overline{U}_\omega$ are similarly defined as integrals over the vertical dimension. Equation (8) gives a scalar uncertainty decomposition which is related to the usual concept of total uncertainty.

4 Collocation model

Suppose we are comparing two instruments, e.g. radiosondes, at the same height and giving measurements y and y^0 respectively. Following the stochastic model approach of the previous section, we have

$$\Delta y(\cdot) = y(\cdot) - y^0(\cdot) = \Delta\mu(\cdot) + \Delta\varepsilon(\cdot). \tag{9}$$

In this formula $\Delta\mu = \mu - \mu^0$ is the collocation drift and $\Delta\varepsilon = \varepsilon - \varepsilon^0$ is the collocation measurement error, with $\mathrm{Var}(\Delta\varepsilon) = \sigma_\varepsilon^2 + \sigma_{\varepsilon^0}^2 = 2\sigma_\varepsilon^2$.

From a practical point of view the observed collocated profiles $y(\cdot)$ and $y^0(\cdot)$ are not observed exactly at the same heights h, while $\mu(\cdot)$ and $\mu^0(\cdot)$ are continuous functions

and $\Delta\mu(h)$ may be easily computed for every height h. Moreover we assume that the collocation drift is an heteroskedastic functional regression model

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

where β_0 is the constant component of the collocation bias, and it is a identifiable coefficient since no instrumental bias has been assumed in the previous section. In particular, using a centered design, with $E(x(\cdot)) = 0$, the collocation bias is given by $E(\Delta y) = E(\Delta\mu) = \beta_0$ and the natural collocation uncertainty is properly defined and decomposed as follows:

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

Note that the first term β_0^2 can be eliminated by simple collocation calibration as shown in the following application section.

5 Beltsville case study

Data used in this work consist of radiosounding profiles of pressure, temperature, humidity and wind measured at the Howard University research site in Beltsville, Meryland, USA (39.054°, −76.877°, 88 m a.s.l.), which is also a GRUAN site, and the US National Weather Service (NWS) operational site located in Sterling, Virginia, USA (38.98°, −77.47°, 53 m a.s.l.).

These two sites, being separated by about 50 km only, are selected because of their relatively close proximity representing a similar climate regime. Moreover, they would serve a good example of using one GRUAN research site to understand a non-GRUAN site, where knowledge can be transferred to a larger network represented by NWS.

Beltsville soundings are based on RS92-SGP sondes, manufactured by Vaisala Inc., while Sterling uses the Radiosonde Replacement System (RRS), build by Lockheed Martin Sippican and referred hereafter as LMS6 sonde. Differences in the vertical

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

sounding of the atmosphere between the two sensor types are known. In dry regions of the troposphere, it has been reported amply that relative humidity derived from Mark IIA (LMS6 sonde) shows substantial errors. This limitation has been reported by Blackmore and Taubvurtze (1999) to be a result of errors in calibration, sensor hysteresis, and sensor response time. They also report that at low temperatures, the time response slows significantly resulting in large relative humidity errors. A lidar-based study of these variations was reported by Adam et al. (2010) that reached similar conclusions. During the latest WMO intercomparison of high quality radiosonde systems (Nash et al., 2010), the RS92 version tested has shown systematic errors of less than 2 % in relative humidity and random errors of about 5 % from the surface to the lower stratosphere, whereas the LMS6 exhibited significant biases in the upper troposphere and layers above. Moreover, the LMS6 sensor did reveal a day-night difference but significant only in the upper troposphere, see Miloshevich et al. (2006).

Most of the sonde-to-sonde or otherwise comparisons reported in the literature are a result of multi-payload sonde launches where two or more sondes are tethered to a single balloon to minimize the atmospheric variability (commonly assumed to be zero). These types of comparisons come mainly from a coordinated and an organized and intensive field campaign, see Miloshevich et al. (2006). These intensive operations tend to be expensive, are held less frequently, and do not offer a climatically representative data across seasons and climates. As a result, the data and opportunity for building statistics is usually limited and cross-instrument and cross network knowledge transfer is limited.

The Beltsville–Sterling radiosonde flights are launched on separate balloon payloads, are operated by different operators, are different instruments and are sampling the atmospheric profile with some variability. Quantifying the latter is a major issue. Traditionally, a simple averaged ensemble comparison as shown in Fig. 1 is done. The time-height matched, difference between two data pairs from the Beltsville–Sterling flights is averaged to show temperature comparisons. The temperature profile difference of sondes launched from these two sites within a 3 h window did not show that

large of a difference. As can be seen from the figure, the temperature difference (standard deviation) was well within about a percent. As expected, the water vapor mixing ratio ($g\,kg^{-1}$) varied greatly for the same sondes throughout the tropopause, above about 2 km. Mid-tropospheric mean differences of about fifty percent were recorded.

Alternatively, comparison of the column integrated water vapor amounts between these two stations revealed correlation coefficients of 0.95 or better. The difference in these comparative plots is a result of the measurement location mismatch, instrument quality, and statistical sampling of the atmospheric variability. These types of "standard" statistics plots, while important in understanding the overall characteristics of the atmospheric state variables, cannot be used to do quantitative contribution of the different error components. The statistical "tool-kit" described here has quantitative description and separation of the different error components as its goal. Note also that despite the performance limitations of the RRS, we proceed with using the Beltsville–Sterling data in our study here to demonstrate the efficacy of the statistics developed.

Following the outcome of the mentioned WMO intercomparison, for this study, we selected 32 pairs of vertical profiles in the range 100–10 000 m launched between July 2006 and September 2009. A flight from Beltsville was matched to Sterling if launch time was within 3 h. For each flight, we consider data profiles of relative humidity (rh, in %), water vapor mixing ratio (mr), pressure (p), temperature (t), measurement calendar time, flight duration (in sec), wind vector (u, v), height a.s.l. (h) and coordinates (lat, lon). The different natural variability and collocation mismatch may be appreciated in Fig. 2, where water vapor mixing ratio, relative humidity, wind and temperature data are plotted. The collocation mismatch, plotted as the time-height matched difference-data, for relative humidity and pressure is plotted in Fig. 3, which shows quite a strong variability for humidity at all altitude levels without an apparent simple pattern. Hence it is challenging for HFR model trying to explain this strong variability, which is assumed to be a cumulative contribution of errors from the instrument overall performance and water vapor spatial variability arising from the drifts shown in Fig. 4.

Note that, the depicted distances between the collocated profile trajectories result to be between 45 and 95 km.

In the rest of this section we model the data about collocation uncertainty of relative humidity in Beltsville radiosondes as a function of corresponding rh level in Sterling as well as: water vapor mixing ratio, temperature, measurement calendar time, flight duration, wind and coordinates from both collocated radiosondes which are considered as explanatory factors. The resulting collocation error analysis corresponds to forecasting the single rh sensor rather than all radiosonde ECV's. In particular, since we use also the collocation error of mixing ratio, Δmr, the heteroskedastic component (Eq. 3) of this HFR model describes the variability of relative humidity for fixed water vapour content in dry air mass.

In Ignaccolo (2013) it is described in some details the statistical process of model identification and validation which involved fitting and testing a large number of alternative models for various combinations of different covariates. Note that, to avoid scale effects and facilitate interpretation, the functional covariates $x(\cdot)$ have been standardized so that the total profile average is zero and the total profile variance is unity.

The final estimated model for the collocation error includes relative humidity and vapour from Sterling radiosondes (rh^0 and mr^0) and the difference in vapour (Δmr) while the other covariates related to the rest of the atmospheric information as well as time, space and distance were excluded for this dataset. This is given by

$$\Delta rh(\cdot) = \underset{(0.15)}{3.37} + \widehat{\beta}_1(\cdot)rh^0(\cdot) + \widehat{\beta}_2(\cdot)mr^0(\cdot) + \widehat{\beta}_3(\cdot)\Delta mr(\cdot) + \underset{(\sigma_\omega)}{\widehat{\omega}(\cdot)} + \underset{(0.90)}{\Delta\varepsilon(\cdot)} \tag{10}$$

where the standard deviations of the corresponding quantities are given in the bracketed subscripts. The beta functions are plotted in Fig. 5 with 95 % confidence bands and show the stable influence of rh^0 on the collocation drift, which hints at an approximately linear relation, ceteris paribus. Moreover, the increasing behaviour of $\widehat{\beta}_3$ compensates for the sharp decrease of mr shown in Fig. 2. It is worth observing that, after accounting for the above covariates, the collocation drift does not depend on the distance of the paired trajectories. The intercept term $\widehat{\beta}_0 = 3.37$ characterizes the constant collocation

bias between Beltsville and Sterling radiosoundings, and could be used for adjusting collocated measurements in practice.

With an adjusted determination coefficient $R^2 = 0.885$, this model misses only 11.5 % of the collocation uncertainty which is covered by $\sigma_{\bar{\omega}}^2(\cdot)$. The latter is estimated by the functional regression model applied to the squares of first order model functional errors $\widehat{\omega}^2 = (\mu - \widehat{\beta}'x)^2$. In doing this we find that the collocation 2nd order uncertainty of relative humidity depends on pressure (p^0), temperature (t^0) and relative humidity (rh^0) in Sterling radiosondes, difference in pressure (Δp), longitude (Δlon), water vapour (Δmr) and wind ($\Delta u, \Delta v$). This gives the functional log-linear model given in Eq. (11) whose γ-functions are given in Figs. 6 and 7. Note that the Est-West distance Δlon is important especially at lower altitudes, say below 3000 m, that is inside the boundary layer, where the variability of water vapour is higher. North-South direction did not enter the model and this is consistent with both the marked trajectory anisotropy shown by Fig. 4 and prevailing wind direction.

$$\log \omega^2(\cdot) = \underset{(0.26)}{1.6} + \widehat{\gamma}_1(\cdot)p^0(\cdot) + \widehat{\gamma}_2(\cdot)t^0(\cdot) + \widehat{\gamma}_3(\cdot)rh^0(\cdot) + \widehat{\gamma}_4(\cdot)\Delta p(\cdot) + \widehat{\gamma}_5(\cdot)\Delta lon(\cdot)$$
$$+ \widehat{\gamma}_6(\cdot)\Delta mr(\cdot) + \widehat{\gamma}_7(\cdot)\Delta u(\cdot) + \widehat{\gamma}_8(\cdot)\Delta v(\cdot) \tag{11}$$

The model given by Eqs. (10) and (11) is used to compute the uncertainty budget of Eqs. (5) and (8). In particular, Fig. 8 clearly shows that the major part of the uncertainty is related to the observed atmospheric conditions as described by Eq. (10). Moreover we can see that the environmental error $U_{\bar{\omega}}$ is smaller inside the boundary layer so the environmental trend μ has a greater explanation capability at these altitudes.

It is worth noting that, after fixing the atmospheric conditions as in Eq. (10), the collocation drift does not depend on the distance between paired trajectories. Nevertheless the conditional uncertainty of Eq. (11) depends on the distance along parallels as mentioned above. In other words, after adjusting for the other environmental factors, the distance cannot be used as a correction factor for the relative humidity collocation error but it is a determinant of the collocation uncertainty. Moreover, for these stations,

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

the distance along meridians is not a key factor for relative humidity collocation uncertainty.

Taking averages of uncertainty profiles, Table 1 shows the role of the various components in the uncertainty budget for these data. The major component is given by the reducible environmental error which is related to water vapour in Sterling and to the collocation difference of water vapour content in dry air mass as shown by Eq. (10) and Fig. 5. Observing that this collocation difference of water vapour content in dry air mass is a combination of the short term variability of water vapor and the sensor response to that variability, it is instructive to find this known fact through purely statistical model formulations.

The second component is the irreducible environmental error, which has been shown to partly depend on distance along parallels. Moreover it can be observed that the "measurement error" $\sigma_\varepsilon = 0.9$ and estimation uncertainty, $\sqrt{\overline{U}_{\hat{\beta}}} = 0.7$ are quite small. Last but not least, a simple constant bias correction for these data would reduce the global collocation uncertainty by about 4 %.

6 Conclusions

In this paper we presented a new and general statistical method for defining and computing a detailed uncertainty budget of the collocation error in atmospheric vertical profiles. It is based on an extension of the classical functional regression model able to cover for hereoskedasticity. This model based approach allows us to decompose global uncertainty in five different components, namely constant (adjustable) bias, reducibile and irreducible environmental errors, sampling error and measurement error. Moreover the conditional uncertainty may be computed for any environmental conditions enabling a deeper insight in the problem.

Although the method is quite general and data driven, for the considered collocation data from two stations 52 km faraway, we found out that the mismatch on relative

humidity has an adjustable constant bias amounting to 4 % of the global collocation uncertainty. Moreover it turn out that the collocation error is related to physical quantities and, in principle, it could be reduced by auxiliary information.

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 to assess the size of the sampling error with respect to the other uncertainty components.

Appendix A

This appendix considers some mathematical details about uncertainty computations of Sect. 3 and the uncertainty decomposition of Eqs. (5) and (6). To do this, we simplify the functional notation $\mu = \mu(\cdot)$ and observe that, in practice, μ and ω are estimated on suitable historical data by $\widehat{\mu}$ and $\widehat{\omega}$ respectively.

A1 Mean estimation and global uncertainty

In this section we focus on μ and the heteroskeadstic structure of ω is ignored. Writing $\widehat{\mu} = \widehat{\beta}' x$ and $\mu = \widehat{\beta}' x + e$, we observe that, under suitable conditions, approximate orthogonality and unbiasedness hold, that is $E(\widehat{\mu}e) \cong 0$ and $E(\widehat{\mu}|x) = \beta' x$ respectively. Hence the well known two stage variance formula entails

$$\mathrm{Var}(\mu) \cong \mathrm{Var}(\widehat{\mu}) + \mathrm{Var}(e)$$
$$\cong E(\mathrm{Var}(\widehat{\mu}|x)) + \mathrm{Var}(E(\widehat{\mu}|x)) + \mathrm{Var}(e)$$

now, for the first term of the right hand side, we have

$$E(\mathrm{Var}(\widehat{\mu}|x)) = E(\mathrm{Var}(\widehat{\beta}' x|x)) = E(x'\mathrm{Var}(\widehat{\beta}|x)x) = U_{\widehat{\beta}}$$

while, for the second term, we have that

$$\mathrm{Var}(E(\widehat{\mu}|x)|\beta) = \mathrm{Var}(\beta' x|\beta) = U_x$$

which in practice is computed for $\beta = \widehat{\beta}$ and Σ_x in Eq. (7) is suitably estimated on the historical data. For example, under LS estimation, we have

$$U_\beta = \widehat{\sigma}_\omega^2 x'(\mathbf{X}'\mathbf{X})^{-1}x$$

where $\mathbf{X}$ is the functional design matrix.

A2   Conditional variance estimation and global uncertainty

In the case of linear heteroskedastic model (Eq. 2), according to conditional heteroskedastic literature (see e.g. Chatfield, 1995), we write

$$\omega^2 = \eta^2(\gamma'x + \zeta)$$

where η is a zero mean, unit variance and symmetric random variable independent on x and ζ. This motivates the assumptions we made in this paper on ω, namely $E(\omega|x) = 0$ and $E(\omega^2|x) = \gamma'x$. Now by simple algebra, we write

$$\omega^2 = \widehat{\gamma}'x + (\eta^2\gamma - \widehat{\gamma})'x + \zeta.$$

Using this expression and approximate unbiasdness of $\widehat{\gamma}$ it follows that

$$E(\omega^2|x) \cong \gamma'x$$

and

$$\mathrm{Var}(\omega) = \gamma'E(x)$$

which in practice is computed for $\gamma = \widehat{\gamma}$ and $E(x)$ as the historical mean f the covariates.
   Similarly in the log-linear case, we have

$$\mathrm{Var}(\omega|x, \gamma) = E(\exp(\gamma'x))$$

and using a second order expansion of $\exp((\gamma - \widehat{\gamma})'x)$ we get

$$\mathrm{Var}(\omega|\gamma) \cong \frac{1}{2} E(\exp(\gamma'x) \cdot x' \mathrm{Var}(\widehat{\gamma}|x)x)$$

which in practice is computed for $\gamma = \widehat{\gamma}$.

References

Blackmore, W. and Taubvurtzel, B.: Environmental chamber tests of NWS radiosonde relative humidity sensors, in: 15th International Conference on Interactive Information and Processing Systems, available at: http://www.ua.nws.noaa.gov/paper-1.htm, Am. Meteorol. Soc., Dallas, TX, , 2010.

Blackwell, K. G. and McGuirk, J. P.: Tropical upper-tropospheric dry regions from TOVS and Rawinsondes, J. Appl. Meteorol., 25, 464–481, 1996.

Calbet, X., Kivi, R., Tjemkes, S., Montagner, F., and Stuhlmann, R.: Matching radiative transfer models and radiosonde data from the EPS/Metop Sodankylä campaign to IASI measurements, Atmos. Meas. Tech., 4, 1177–1189, doi:10.5194/amt-4-1177-2011, 2011.

Chatfield, C.: The Analysis of Time Series, Chapman and Hall, London, 1995.

Ettinger, B., Perotto, S., and Sangalli, L. M.: Studying hemodynamic forces via spatial regression models over non-planar domains, available at: http://meetings.sis-statistica.org/index.php/sis2013/ALV/paper/viewFile/2626/307, in: Proceedings of SIS 2013, 19–21 June 2013, Advances in Latent Variables, Brescia, 2013.

Fassò, A., Esposito, E., Porcu, E., Reverberi, A. P., and Vegliò, F.: Statistical sensitivity analysis of packed column reactors for contaminated wastewater, Environmetrics, 14, 743–759, 2003.

Frehlich, R. and Sharman, R.: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation, Mon. Weather Rev., 132, 2308–2324, 2004.

GCOS Reference Upper-Air Network (GRUAN): Justification, requirements, siting and instrumentation options, GCOS–112, WMO Tech. Doc. 1379, WMO, http://www.wmo.int/pages/prog/gcos/Publications/gcos-112.pdf (last access: August 2013), 2007.

Houchi, K., Stoffelen, A., Marseille, G. J., and de Kloe, J.: Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and te ECMWF model, J. Geophys. Res. D, 115, 22123, doi:10.1029/2009JD013196, 2010.

Ignaccolo, R.: Functional Data Modeling in climatology, available at: http://meetings.sis-statistica.org/index.php/sis2013/ALV/paper/viewFile/2693/486, in: Proceedings of SIS 2013, Advances in Latent Variables, 19–21 June 2013, Brescia, 2013.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

Miloshevich, L. M., Vömel, H., Whiteman, D., Lesht, B., Schmidlin, F. J., and Russo, F.: Absolute accuracy of water vapor measurements from six operational radiosondes types launched during AWEX-G and implications for AIRS validation, J. Geophys. Res., 111, 1–25, doi:10.1029/2005JD006083, 2006.

Nash, J., Oakley, T., Vömel, H., and Wei, L.: WMO Intercomparison of high Quality Radiosonde Systems Yangjiang, China, 12 July–3 August 2010; WMO report reference number IOM 107 (TD 1580), available at: http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html (last access: August 2013), 2010.

O'Connor, E. J., Hogan, R. J., and Illingworth, A. J.: Retrieving stratocumulus drizzle parameters using Doppler radar and lidar, J. Appl. Meteorol., 44, 14–27, 2005.

Ramsay, J. O. and Dalzell, C.: Some tools for functional data analysis (with discussion), J. Roy. Stat. Soc. B, 53, 539–572, 1991.

Ramsay, J. O. and Silverman, B. W.: Functional Data Analysis, Springer, Dordrecht, 2005.

Ruiz-Medina, M. D. and Espejo, R. M.: Spatial autoregressive functional plug-in prediction of ocean surface temperature, Stoch. Environ. Res. Risk A., 26, 335–344, 2012.

Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O.: Spatial spline regression models, J. Roy. Stat. Soc. B, 75, 1–23, 2013.

Seidel, D. J., Sun, B., Pettey, M., and Reale, A.: Global radiosonde balloon drift statistics, J. Geophys. Res., 116, 102, 1–8, doi:10.1029/2010JD014891, 2011.

Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Lesht, B. M., Strow, L. L., Hannon, S. E., Feltz, W. F., Moy, L. A., Fetzer, E. J., and Cress, T. S.: Atmospheric Radiation Measurement

Collocation uncertainty in atmospheric profiles

A. Fassò et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, J. Geophys. Res., 111, D09S14, doi:10.1029/2005JD006103, 2006.

Whiteman, D. N., Schwemmer, G., Berkoff, T., Plotkin, H., Ramos-Izquierdo, L., and Pappalardo, G.: Performance modeling of an airborne Raman water-vapor lidar, Appl. Optics, 40, 375–390, 2001.

Wood, S. N.: Stable and efficient multiple smoothing parameter estimation for generalized additive models, J. Am. Stat. Assoc., 99, 673–686, 2004.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, J. Roy. Stat. Soc. B, 73, 3–36, 2011.

Table 1. Collocation Uncertainty budget for relative humidity (Δrh) in Beltsville–Sterling, using HFR model (Eqs. 10–11).

Source of uncertainty		$\overline{U}$	$\overline{U}\%$	$\sqrt{\overline{U}}$
Natural variability	$\Delta\mu$	318.7		17.9
Bias (adjustable)	β_0^2	11.6	3.6 %	3.4
Environmental Error (reducible)	x	268.6	84.3 %	16.4
Environmental Error (irreducible)	ω^2	37.6	11.8 %	6.1
Sampling error	$\widehat{\beta}$	0.2	0.1 %	0.7
Measurement error	$\Delta\varepsilon$	0.8	0.3 %	0.9
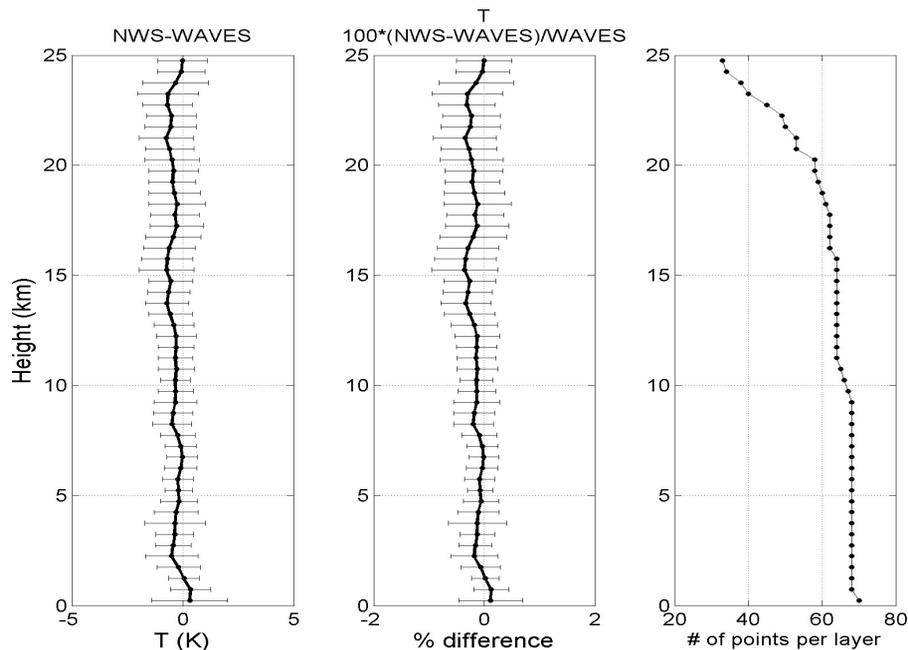
Fig. 1. Comparison of radiosonde flights made at Beltsville and NWS-Sterling. Data pairs were matched if they were within 3 h. The temperature difference and the standard deviation of the difference are shown as absolute difference and in percent as well as the number of data pairs used at each layer.
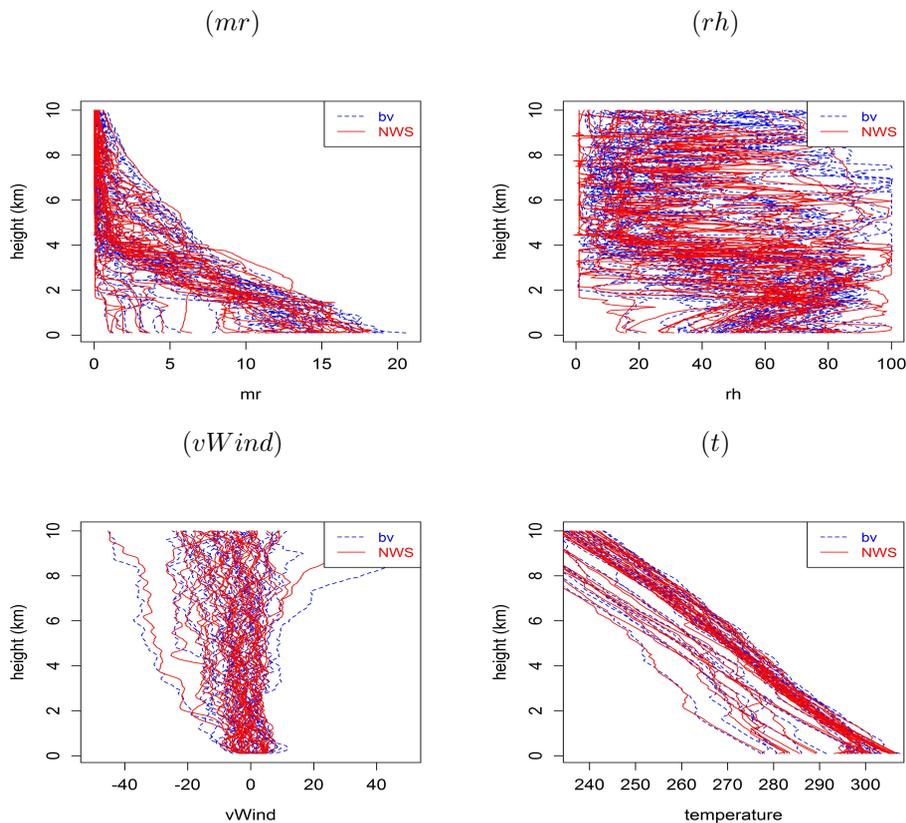
Fig. 2. Data profiles in Beltsville (bv) and Sterling (NWS). Top left panel: mixing ratio (mr); top right panel: relative humidity (rh); bottom left panel: East-West wind component (vWind); bottom right panel: temperature (*t*).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

(Δrh)

(Δp)
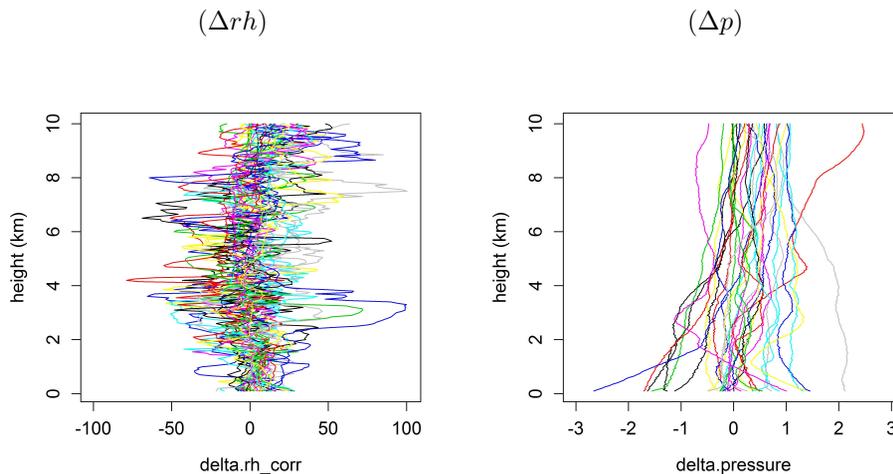
Fig. 3. Collocation mismatch profiles given by differences BV-NWS. Left panel: relative humidity (Δrh); right panel: pressure (Δ*p*).

Displacement of co-located trajectories
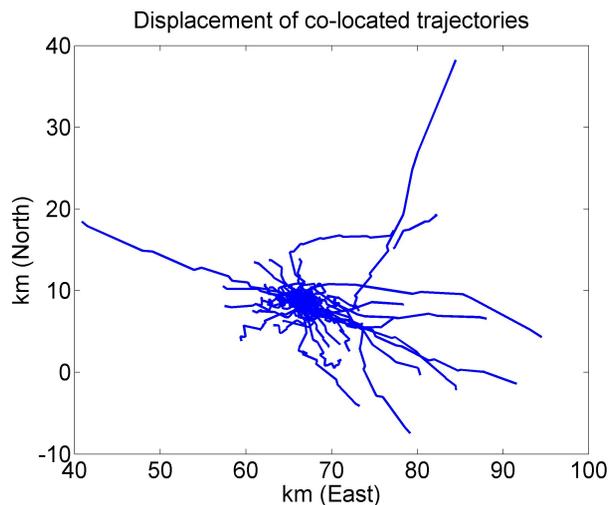
Fig. 4. The displacement of collocated trajectories is given by the distance among the positions of the two collocated instruments at same altitude. *x* axis is distance along parallels and *y* axis along meridians. Distance range is 45–95 km.
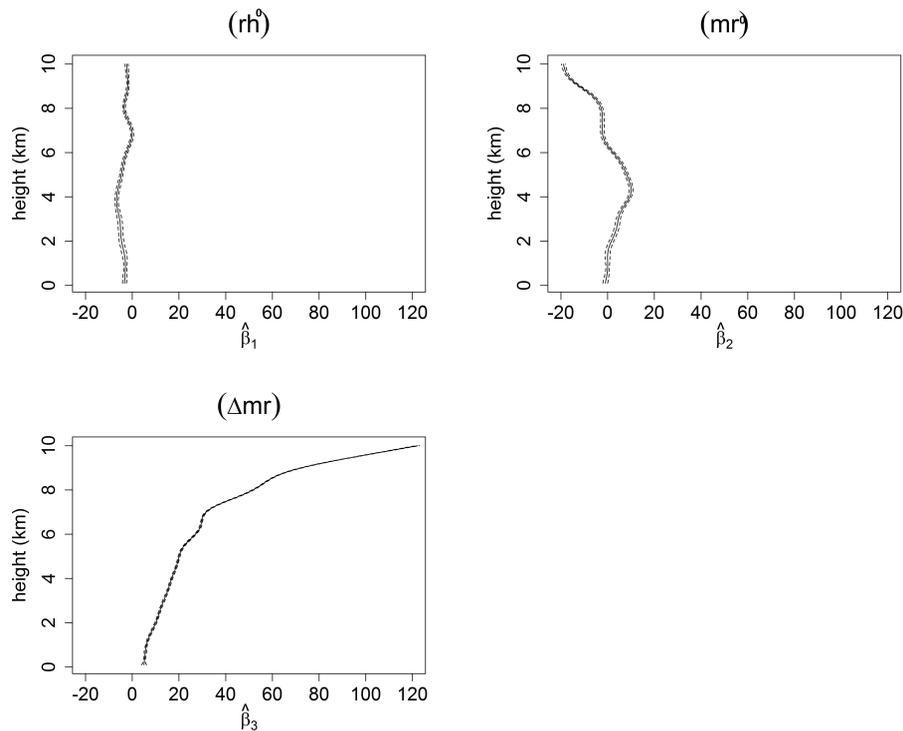
Fig. 5. Beta functions for Beltsville collocation drift model of relative humidity (rh) given in Eq. (10). Top left panel: Sterling relative humidity (rh^0); top right panel: Sterling mixing ratio (mr^0); bottom left panel: difference in mixing ratio (Δmr).
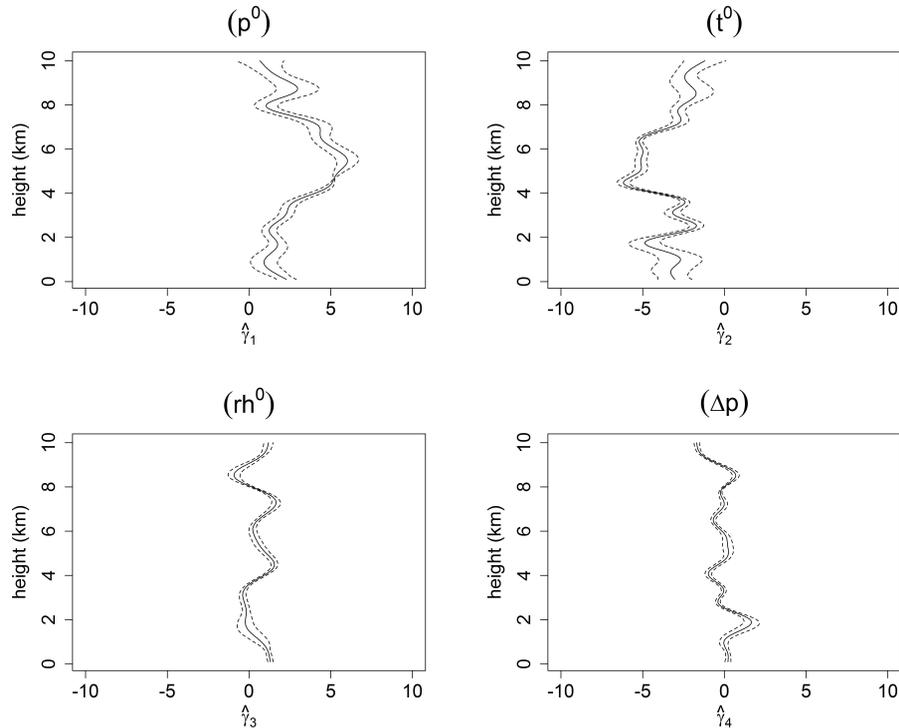
Fig. 6. Gamma functions of collocation error ϖ^2 for relative humidity in Beltsville using model given in Eq. (11). Top left panel: Sterling pressure (p^0); top right panel: Sterling temperature (t^0); bottom left panel: Sterling relative humidity (rh^0); bottom right panel: difference in pressure (Δp).
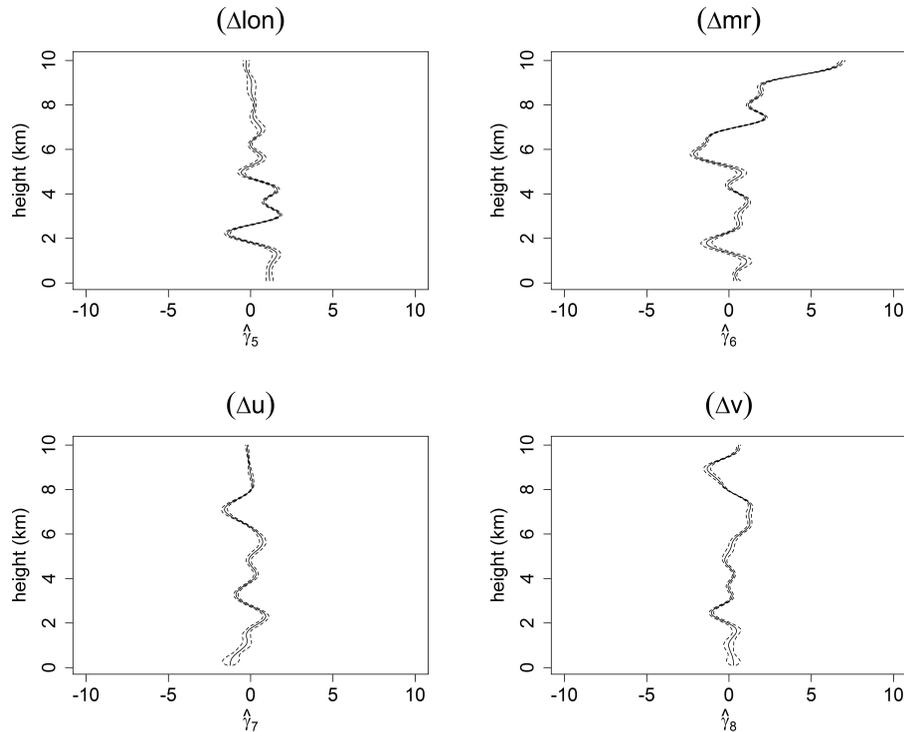
Fig. 7. Gamma functions of collocation error ω^2 for relative humidity in Beltsville using model given in Eq. (11). Top left panel: difference in longitude (Δlon); top right panel: difference in mixing ratio (Δmr); bottom left panel: difference in wind, u-direction (Δu); bottom right panel: difference in wind, v-direction (Δv).
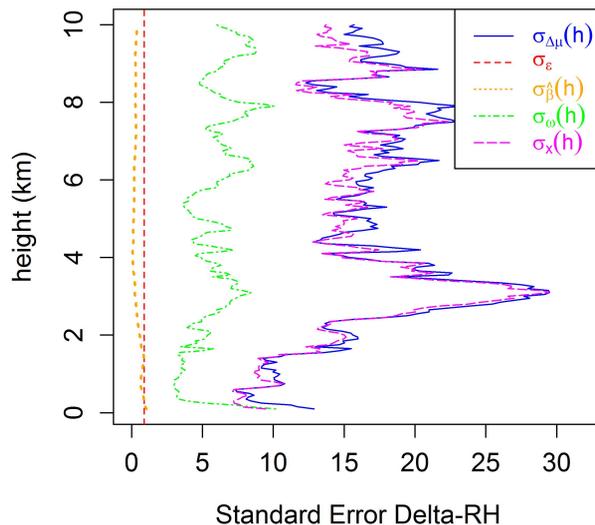
Fig. 8. Squareroot uncertainty budget ($\sigma = \sqrt{U}$) for relative humidity collocation mismatch. Natural variability ($\sigma_{\Delta\mu}$), measurement error (σ_{ε}), sampling error: ($\sigma_{\hat{\beta}}$), irreducible environmental error (σ_{ω}) and reducible environmental error (σ_x).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion