

# DISSERTATION

Titel der Dissertation

On Bias Correction and Quality Control for Atmospheric in situ  
Observations

verfasst von

Mag.<sup>a</sup> Christina M. Tavorato-Wötzl

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 091 414
Dissertationsgebiet lt. Studienblatt:	Meteorologie und Geophysik
Betreuer:	A. Prof. Dr. Leopold Haimberger



When one admits that nothing is certain  
one must, I think, also add that some things  
are more nearly certain than others.

Bertrand Russell (1872-1970)



## **Abstract**

Meteorological observations of the atmosphere are essential input information to calculate the initial state for every numerical weather prediction model as well as for climatologically used reanalysis. A good analysis of the current state of the atmosphere is the foundation for an accurate forecast. Historic observations, especially, can have large biases that can affect the analysis. These biases can change in time due to a change in the observing system, a change of the observation location and its surroundings, etc. This is highly problematic for climate monitoring purposes and therefore a good bias adjustment is key.

This thesis presents methods to improve the use of conventional observations by adjusting biases with different approaches. The thesis puts a special emphasis on radiosonde data. Due to the availability of radiosonde profiles of the atmosphere over the last century this data set can provide useful information for forecast models and climate analyses. Homogenisation of long term radiosonde time series provides an important input data set for climate studies. Furthermore, comparisons of the homogenised data sets to other upper-air data sets during the last decades are presented and evaluated to show the impact of homogenisation.

Next to bias correction of long term time series for climate analysis, possible methods for quality control within variational data assimilation systems are presented.

A newly developed method for variational quality control within the European Centre of Medium-range Weather Forecast (ECMWF) state of the art four dimensional variational data assimilation system uses a Huber norm distribution to describe the observations rather than a Gaussian plus flat distribution, which leads to more emphasis on observations that differ from the model background to improve the forecast of fast developing small scale weather events. While being quite general in its formulation it is currently applied to conventional data.

Next to this approach to bias correction a more advanced method that estimates the bias during the variational assimilation is presented. This approach has been used to bias correct satellite data as well as surface pressure and aircraft wind data at the ECMWF and is now adapted to specifically adjust radiosonde wind direction bias.

The result of the work shows that bias correction, quality control and homogenisation of observations are necessary to improve weather forecast models as well as reanalysis datasets by improving the data analysis.



## **Zusammenfassung**

Meteorologische Beobachtungen des aktuellen und des vergangenen Zustands der Atmosphäre sind wichtige Informationen, um den Ausgangszustand für jedes numerische Wettervorhersagemodell zu berechnen. Diese Analyse ist sowohl für operationelle Vorhersagemodelle als auch für Reanalysen von großer Bedeutung, da eine qualitativ hochwertige Ausgangsanalyse zur bestmöglichen Vorhersage führt. Vor allem historische Beobachtungen können große systematische Fehler - einen sogenannten Bias - aufweisen, der die Analyse nachhaltig beeinflussen kann. Dieser Bias sollte, um die Daten bestmöglich verwenden zu können, korrigiert werden. Besonders wichtig wird eine Bias-Korrektur, wenn sich dieser im Laufe der Zeit auf Grund eines technischen Fortschritts im Instrument, einer Änderung des genauen Messortes, etc. ändert. Daher sind Qualitätskontrolle und Bias-Korrektur von meteorologischen Beobachtungen wichtige Bestandteile jedes Datenassimilationssystems.

Diese Dissertation präsentiert Methoden zur Qualitätskontrolle und Homogenisierung konventioneller Daten. Dabei liegt das Hauptaugenmerk bei Radiosondendaten.

Durch die Verfügbarkeit dieser Daten über einen großen Zeitraum als vertikale Profile der Atmosphäre enthalten sie wertvolle Informationen für Vorhersagemodelle und Klimareanalysen. Ein homogenisierter Radiosondendatensatz über den gesamten Messzeitraum bildet einen wichtigen Eingangsdatensatz für jede Klimaanalyse. Der homogenisierte Radiosondendatensatz wird weiters in dieser Arbeit mit anderen atmosphärischen Datensätzen der letzten Jahrzehnte verglichen, um die Auswirkungen der Homogenisierung zu zeigen.

Neben der Homogenisierung von Langzeitreihen werden auch Methoden zur Qualitätskontrolle und Bias-Korrektur innerhalb eines Vorhersagemodells vorgestellt.

Einerseits wird die neu entwickelte variationelle Qualitätskontrolle am Europäischen Zentrum für Mittelfristige Wettervorhersage (EZMW), bei der anstelle einer Normalverteilung mit flachen Enden eine Huber-Norm verwendet wird, um die Daten zu beschreiben, vorgestellt. Diese Methode legt ein stärkeres Gewicht auf Beobachtungen, die sich vom Backgroundfeld des Modells unterscheiden und verbessert daher die Vorhersage für sich schnell entwickelnde, kleinskalige Wettersysteme. Obwohl diese Methode sehr allgemein formuliert ist, wird sie momentan nur auf konventionelle Daten angewandt.

Andererseits wird eine Bias-Korrektur vorgestellt, bei welcher der Bias während der variationellen Datenassimilation mitgeschätzt wird. Dieser Ansatz wird am EZMW verwendet, um Satellitendaten sowie Flugzeugtemperatur- und Bodendruckbeobachtungen anzupassen. Die Anwendung auf einen Bias der Radiosondenwindrichtung wird präsentiert.

Die Ergebnisse dieser Arbeit sollen zeigen, dass sowohl Qualitätskontrolle, Bias-Korrektur und Homogenisierung von meteorologischen Beobachtungen wichtige Bestandteile sind, um Wettervorhersagen und Reanalysen durch die Verbesserung der Datenanalyse zu verbessern.





# Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Publications . . . . .	2
1.2.1 Variational Quality Control . . . . .	2
1.2.2 Bias Correction - Homogenisation . . . . .	4
1.2.3 Bias Correction - Data Comparison . . . . .	5
1.2.4 Bias Correction - Variational Approach . . . . .	5
<b>2 Variational Quality Control</b>	<b>7</b>
<b>3 Radiosonde Temperature Homogenisation</b>	<b>23</b>
<b>4 Comparison of Upper-Air Temperature Data</b>	<b>49</b>
<b>5 Variational Bias Correction</b>	<b>65</b>
<b>6 Final Remarks and Conclusions</b>	<b>93</b>
6.1 Final Remarks . . . . .	93
6.1.1 Variational Quality Control . . . . .	93
6.1.2 Bias Correction using Homogenisation Methods . . . . .	94
6.1.3 Comparison of Upper-Air Temperature Data . . . . .	94
6.1.4 Variational Bias Correction . . . . .	95
6.2 Conclusions . . . . .	95
<b>Bibliography</b>	<b>97</b>



# 1 Introduction

## 1.1 Motivation

Meteorological observations are a key factor for any type of numerical weather prediction (NWP) system. Those systems are either used for the current weather forecast or to recalculate the state of the atmosphere in the past with a reanalysis effort. Each observation represents the current state of the atmosphere at the location of the observation. There are two types of observation systems: in situ (conventional) observations (measurements of a parameter on site - e.g. a thermometer measuring temperature at its location) and remote sensing observations (measuring a parameter from space - e.g. satellites measuring radiances which can provide information on temperature, humidity, etc.). This work will concentrate on conventional observations with a special emphasis on radiosonde data.

Due to technical improvements of the observation system, changed setup of meteorological observation sites, etc., the biases of observations change. Those biases carry no useful meteorological information but can input erroneous information into observations and therefore into products that use those observations. To achieve the best possible result in a data assimilation, system observations should be quality controlled and if necessary bias corrected. There are various methods (Hollingsworth et al., 1986; Lorenc and Hammon, 1988; Ingelby and Lorenc, 1993; Andersson and Järvinen, 1999) to bias correct observations. A deeper look into some of the methods and some work to improve quality control and bias correction of meteorological observations are presented here.

The work shown in this thesis spans from a newly developed quality control scheme for conventional observations (Tavolato and Isaksen, 2014) to an analysis of a homogenised radiosonde data set (Haimberger et al., 2008, 2012) and its comparison to other upper air temperature data sets (Ladstädter et al., 2011) to an implementation of variational bias correction of radiosonde wind data (Tavolato-Wötzl, 2015).

This work was completed during a stay as graduated trainee at the European Centre of Medium-range Weather Forecast (ECMWF) as well as during my work for the FWF-Project "Global in situ upper air data for climate change research" (P21772-N22, PI: Leopold Haimberger) at the University of Vienna. As a graduated Trainee at ECMWF I got the chance to work on projects in the data assimilation section (supervised by Lars Isaksen - current head of the data assimilation section at ECMWF) especially introducing a new variational quality control (VarQC) scheme for conventional data using a Huber norm (Huber, 1964) to determine the weights of each observation within the assimilation system. At the University of Vienna, I was part of the radiosonde homogenisation group and especially involved in the work that compared the homogenised data to other upper air data sets such as (A)MSU satellite data (Christy et al., 1998, 2003; Mears and Wentz, 2009) and GPS (Global Positioning System) radio occultation

data (Melbourne et al., 1994; Kursinski et al., 1997).

### 1.2 Publications

This thesis will focus on following aspects of quality control and bias correction highlighted in the publications included, which are:

- Variational Quality Control:

*On the use of a Huber norm for observation quality control in the ECMWF 4D-Var.* Tavolato and Isaksen, 2014.

- Bias Correction - I: Homogenisation:

*Homogenization of the global radiosonde data set through combined comparison with re-analysis background series and neighbouring stations.* Haimberger, Tavolato and Sperka, 2012.

- Comparison of Homogenised Radiosonde Temperature Data

*An assessment of differences in lower stratospheric temperature records from MSU, radiosonde and GPS radio occultation* Ladstädter, Steiner, Foelsche, Haimberger, Tavolato and Kirchengast, 2011.

- Bias Correction - II: Variational bias correction

*Variational Bias Correction for Radiosonde Wind Direction.* Tavolato-Wötzl, 2015.

#### 1.2.1 Variational Quality Control

Any current NWP system consists of two equally important parts. One is the forecast model where physical equations describing the atmosphere are applied in order to calculate its future state. The other part is the data assimilation process which combines current observations and a short range forecast (the background) in a statistically optimal way to determine the best guess of the analysis which describes the atmospheric state at the present time. At ECMWF this is done with a multi-dimensional variational data assimilation system (Courtier et al., 1994).

Within any variational data assimilation system the input data need to run through a set of quality control (QC) checks before contributing to the analysis. In the ECMWF four dimensional variational (4D-Var) data assimilation system the quality control consists of the following steps: first data that is known to be erroneous is blacklisted and removed from the system (the blacklist gets updated constantly with the help of careful data monitoring). After blacklisting observations that differ too much from the background these observations will be first-guess

rejected and as a final quality control the so called variational quality control (VarQC) is performed (Andersson and Järvinen, 1999).

This final step is part of the variational assimilation and determines the weight given to each observation when building the analysis state. Until 2009 the VarQC within the ECMWF 4D-Var assimilation system applied a Gaussian plus flat distribution to the observation departures (as described in Andersson and Järvinen, 1999). Therefore the VarQC weights were either zero or one, with a very steep transition between those values. Observations would either influence the analysis or would be VarQC rejected (this term is used when the VarQC weight drops below 25%). This works fine most of the time, but for certain weather events where there is a fast small scale change in the flow (e.g. fast developing small scale lows), accurate observations can be too far from the background values and therefore will get rejected.

An evaluation of the current data usage at ECMWF (Tavolato and Isaksen, 2011) as well as the changes in data usage over the last years and the differences between an operational model and reanalysis efforts (used in Dee et al., 2011) showed room for improvement in the variational quality control.

The work published in Tavolato and Isaksen (2014) introduces a different approach to the VarQC for conventional data. Instead of the Gaussian plus flat distribution, a Huber norm (Huber, 1964) is used to describe the observation departures. This distribution was found to represent the actual distributions better and allows a relaxation of the first-guess rejection. The main impact can be seen by a wider spread of different VarQC weights. Especially observations with some distance to the background will get small weights instead of zero weight and will contribute to the analysis. This contribution is especially worthwhile if there are more than one observation differing from the background in the same direction and supporting each other. This change should improve the analysis in general and specifically in situations of fast developing weather events that are not picked up by the background field fast enough. One often discussed example are the winter storms in December 1999 (e.g. Dee et al., 2001). In this case the analysis misplaced the storm and showed insufficiently low surface pressures due to the rejection of good observations. Using the VarQC introduced in Tavolato and Isaksen (2014) the storm is well represented in the analysis. Other examples and attempts to verify forecasts for extreme events can be found in Friederichs and Thorarinsdottir (2012).

The change in VarQC helped to improve the analysis in various cases and is especially useful when analysing tropical storms. It was implemented in the ECMWF 4D-Var assimilation system in September 2009 for conventional data and has been applied in the operational forecast system ever since.

### 1.2.2 Bias Correction - Homogenisation

Radiosondes are weather balloons that measure the atmospheric state (wind, temperature, humidity) while the balloon is ascending. Radiosonde ascents are usually performed once to twice a day (at 00UTC and 12UTC) near major cities in each country. Radiosonde observations cover only land areas, the density of their deployment varying depending on the economic situation of the nation operating the radiosonde. Data from radiosondes provides a three dimensional picture of the state of the atmosphere (compared to only two dimensional information obtained by ground based observations). Radiosonde observations go back to the mid 20th-centuries but a large increase of observations can be noticed from 1958, the International Geophysical Year (IGY), onwards. Radiosondes provided upper air information long before satellite observations started and are therefore a valuable long term upper air record for climatological studies.

With any data set that is used in climate studies it is crucial that there are no artificial biases present that could influence the long term trends. Long term radiosonde records can be influenced by technical changes in the observation system as well as relocations of radiosonde stations that kept the same station-ID. Homogenisation methods try to remove these effects from the long observation time series.

The automatic homogenisation system RAOBCORE (RADiosonde OBServation CORrection using REanalysis, see Haimberger, 2007) uses background information from the ERA-40 (Upala et al., 2005) and ERA-Interim (Dee et al., 2011) reanalysis to detect inhomogenities in the radiosonde temperature time series with a statistical test. These break points are thereafter adjusted using the mean of the reference time series before and after the break point. Another system developed at the University of Vienna called RICH (Radiosonde Innovation Composite Homogenisation, Haimberger et al., 2008, 2012) uses the breakpoints detected by RAOBCORE but adjusts the time series using only neighboring time series and therefore is more independent from the reanalysis data sets.

A homogenisation approach provides external bias estimates which thereafter are used in an assimilation system in the following way:

$$J(\vec{x}) = (\vec{x} - \vec{x}_b)^T B^{-1} (\vec{x} - \vec{x}_b) + (\vec{y} + \vec{b} - H(\vec{x}))^T R^{-1} (\vec{y} + \vec{b} - H(\vec{x})) \quad (1.1)$$

$J(\vec{x})$  describes the cost function of the atmospheric state of the model  $\vec{x}$  with  $\vec{x}_b$  as the background field,  $\vec{y}$  are the observations and  $\vec{b}$  the bias adjustment calculated prior to the assimilation with RAOBCORE or RICH.  $H(\vec{x})$  is the observation operator that maps the model onto the observed variables and locations and  $B$  and  $R$  are the background error covariance matrix and the observation error covariance matrix.

### 1.2.3 Bias Correction - Data Comparison

Results of the homogenisations were compared to other upper air data sets and results can be seen in the presented papers Haimberger et al. (2012) and Ladstädter et al. (2011) in section 3 and section 4. These other data sets include other homogenised radiosonde data sets (Seidel et al., 2004) as well as different (A)MSU satellite data sets (Christy et al., 1998, 2003; Mears and Wentz, 2009) and GPS radio occultation data (Melbourne et al., 1994; Kursinski et al., 1997).

The adjusted radiosonde temperatures in Haimberger et al. (2008) were the first to show the warming in the tropical mid troposphere to be as strong as predicted by climate models and as seen in some satellite data sets (Ramella Pralungo and Haimberger, 2014b; Mears et al., 2012). The adjusted radiosonde temperature data set was used as radiosonde data input for the ERA-Interim (Dee et al., 2011) reanalysis. Both data sets have also been used for the annual BAMS (Bulletin of the American Meteorological Society) special edition about the state of the climate (Blunden et al., 2011; Blunden and Arndt, 2012).

Currently, work is done to expand the RAOBCORE and RICH data sets further back beyond 1958. Not only temperature but also radiosonde wind (Gruber and Haimberger, 2008; Ramella Pralungo et al., 2014; Ramella Pralungo and Haimberger, 2014a) and possibly humidity observations are adjusted using the method. The results of this work should lead to improved observations for a future reanalysis at ECMWF.

### 1.2.4 Bias Correction - Variational Approach

Another very powerful approach for bias correction is to calculate a bias estimate within the variational data assimilation (Variational Bias Correction - VarBC, Dee and Uppala, 2009). During the assimilation each bias is estimated with the help of predefined predictors which describe the bias. These predictors range from a constant to different characteristics during the observation (solar elevation, angle of the satellite scan, thickness of a predefined atmospheric layer, etc.). In the simplest case there is only one predictor describing a constant bias. This method introduces a bias term into the variational cost function:

$$J(\vec{x}) = (\vec{x} - \vec{x}_b)^T B^{-1} (\vec{x} - \vec{x}_b) + (\beta_b - \beta)^T B_\beta^{-1} (\beta_b - \beta) + (\vec{y} - H(\vec{x}) - \vec{b}(\vec{x}, \beta))^T R^{-1} (\vec{y} - H(\vec{x}) - \vec{b}(\vec{x}, \beta)) \quad (1.2)$$

Where the bias is described as:

$$\vec{b}(\vec{x}, \beta) = \sum_{i=0}^{N_p} \beta_i \vec{p}_i(\vec{x}) \quad (1.3)$$

using  $p_i$  for the predictors and  $\beta_i$  for the unknown bias parameters. Compared to the approach described before, the contrast is that here the simply used pre calculated bias correction is

changed to a bias that is estimated during the assimilation process.

VarBC was first introduced at ECMWF to account for radiance biases in satellite observations but is nowadays used for a range of satellite observations as well as some conventional observations (aircraft measured temperatures, surface pressure).

This approach should be another option to address biases in radiosondes. To test this hypothesis a radiosonde VarBC was introduced to the ECMWF assimilation system and wind direction was chosen as a first parameter to correct. Gruber and Haimberger (2008) could identify several radiosonde stations with a constant wind direction bias throughout the profile by comparing them to reanalysis data. This bias is artificial as it appears suddenly and is constant thereafter. A wrong north alignment of the radiosonde station is the most likely reason. This bias should be a constant bias throughout a radiosonde wind profile and therefore only needs one predictor in the cost function (eq. 1.2).

Both the variational and the conventional approach have advantages and disadvantages. The advantages of a conventional homogenisation system is that the bias and its development in time is known before the data is used in an assimilation system. On the other hand, the homogenisation step has to be applied to a data set as an extra step before assimilation. Using parameters for bias correction derived from metadata by the observation system manufacturers allows adjustments during assimilation. However, only adjustments known at observation time can be applied in NWP models. Time consuming homogenisation methods of observation time series can only be used for climate studies.

With a variational approach, data is bias corrected during the assimilation. This bias correction scheme is useable in forecast models where new input data gets adjusted straight away. The estimated bias is not known before the assimilation, which is why this approach relies on a good background forecast from the model. Applying VarBC on in situ observations as well as satellite data means more trust is put into the model used. It is uncertain how much anchoring by unbiased observations is needed to prevent the analysed state from drifting into a potentially unrealistic model climate. This is not a big issue for wind direction biases, since wind direction, especially in the vertical mean, is dynamically well constrained.

Tavolato-Wötzl (2015) presented in section 5 describes the implementation and first results of a VarBC for radiosonde wind direction.



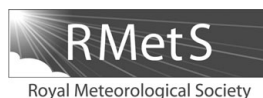
## 2 Variational Quality Control

The following paper was submitted with co-author Lars Isaksen to the Quarterly Journal of the Royal Meteorological Society in March 2014. It was accepted in August 2014 and published online in October 2014. It is currently in press.

As first author of the paper my task was the introduction and implementation of the Huber norm variational quality control to the ECMWF data assimilation system. The idea for such a variational quality control has been in mind at ECMWF but until this work no-one had implemented a quality control using a Huber norm into the operational model. Next to the implementation which was supported and encouraged by the supervisor of this project (the current head of the Data Assimilation department at ECMWF) and co-author of this paper Lars Isaksen, I also performed the long term evaluation of model scores and case studies presented within this publication. While my co-author helped with the historic overview in the introduction the most of the text was written by myself.

Tavolato, C. and Isaksen, L. 2014: *On the use of a Huber norm for observation quality control in the ECMWF 4D-Var* Q. J. Roy. Meteor. Soc., published online.





## On the use of a Huber norm for observation quality control in the ECMWF 4D-Var

Christina Tavalato<sup>a,b</sup> and Lars Isaksen<sup>a\*</sup>

<sup>a</sup>European Centre for Medium-Range Weather Forecasts, Reading, UK

<sup>b</sup>Department of Meteorology and Geophysics, University of Vienna, Austria

\*Correspondence to: L. Isaksen, ECMWF, Shinfield Park, Reading RG2 9AX, UK.

E-mail: lars.isaksen@ecmwf.int

This article describes a number of important aspects that need to be considered when designing and implementing an observation quality control scheme in an NWP data assimilation system. It is shown how careful evaluation of innovation statistics provides valuable knowledge about the observation errors and help in the selection of a suitable observation error model. The focus of the article is on the statistical specification of the typical fat tails of the innovation distributions. In observation error specifications, like the one used previously at ECMWF, it is common to assume outliers to represent gross errors that are independent of the atmospheric state. The investigations in this article show that this is not a good assumption for almost all observing systems used in today's data assimilation systems. It is found that a Huber norm distribution is a very suitable distribution to describe most innovation statistics, after discarding systematically erroneous observations. The Huber norm is a robust method, making it safer to include outlier observations in the analysis step. Therefore the background quality control can safely be relaxed. The Huber norm has been implemented in the ECMWF assimilation system for *in situ* observations. The design, implementation and results from this implementation are described in this article. The general impact of using the Huber norm distribution is positive, compared to the previously used variational quality control method which gave virtually no weight to outliers. Case-studies show how the method improves the use of observations, especially for intense cyclones and other extreme events. It is also discussed how the Huber norm distribution can be used to identify systematic problems with observing systems.

**Key Words:** observation quality control; Huber norm; robust estimation; variational data assimilation

Received 24 March 2014; Revised 16 July 2014; Accepted 19 August 2014; Published online in Wiley Online Library

### 1. Introduction

Quality control (QC) of observations is an important component of any data assimilation system (Lorenc and Hammon, 1988). Observations have measurement errors and sometimes gross errors due to technical errors, human errors or transmission problems. The goal is to ensure that correct observations are used and erroneous observations are discarded from the analysis process. It has long been recognised that a good QC process is required because adding erroneous observations to the assimilation can lead to spurious features in the analysis (Lorenc, 1984).

In data assimilation, the use of departures of observations (*o*) from the short-range (background, *b*) forecast is an integral part of the QC. If observations, evaluated over a long period, systematically or erratically deviate from the background forecast they should be blacklisted, i.e. not taken into account at all in the analysis (Hollingsworth *et al.*, 1986). The remaining observations are, for each analysis cycle, also compared against the background and rejected if the background departures are large. Often departures, normalised by the expected observation error, are assumed to follow a Gaussian distribution. This means outliers

are statistically very unlikely and will unjustly get the same full weight in the analysis as correct observations, increasing the risk of producing an erroneous analysis by using incorrect observations. This is usually resolved by applying fairly tight background departure limits which reject outliers. The background QC limits depend on the specified observation error and background error. For accurate observations and modern high-quality assimilation systems, these are both small, e.g. of the order of 0.5 hPa for automated surface pressure observations. So surface pressure observations will typically be rejected if they differ by more than about 4 hPa from background fields, corresponding to six standard deviations of normalised departures. In most cases this is reasonable, but for extreme events it may well happen that the short-range forecast is wrong by more than 4 hPa near the centre of cyclones. The QC decisions can be improved to some degree by introducing flow-dependent, more accurate, background errors, like the ones recently implemented at the European Centre for Medium-range Weather Forecasts (ECMWF; Isaksen *et al.*, 2010; Bonavita *et al.*, 2012). These errors would typically be larger near the centre of cyclones.

Section 2 of the article investigates the innovation statistics for some of the most important *in situ* observations. This leads to a discussion and description in section 3 of the gross error QC aspects which need to be considered for observations. The special problems that may occur for biases and bias correction of isolated stations is covered in section 4. After this general study of innovation statistics and gross error characteristics, section 5 describes a range of proposed probability distributions with fat tails which are candidates for innovation statistics and observation error specification. Based on the information presented in sections 2–5, it is found that a Huber norm (Huber, 1964, 1972) is the most suitable distribution to use. The method allows the inclusion of outliers in the analysis with reduced weight, because it is a robust estimation method. This is in contrast with a pure Gaussian approach where the analysis can be ruined by a few erroneous outliers. Section 6 covers the aspects which need to be considered when implementing a Huber norm QC in a numerical weather prediction (NWP) system. We describe how this is done in the Integrated Forecast System (IFS) at ECMWF, where it has been used operationally since September 2009. It is also explained how the background QC has been relaxed, and how observation error values have been reduced at the centre of the distribution to consistently reflect the Huber norm distribution. Section 7 presents general impact results and a number of case-studies.

## 2. Distribution of departure statistics for some important *in situ* observations

The main weakness of using background departure statistics for investigations of observation error distributions is that they are a convolution of observation and background information. Further information is required to uniquely determine the observation-related distribution, which is what we really are trying to estimate, as it is needed in the definition of the observation cost function. Despite this weakness innovation statistics are the most common observation-related diagnostics used in data assimilation. Additional research to identify if background errors are non-Gaussian is recommended, but it is outside the scope of this paper. Assuming the background error follows a Gaussian distribution, all non-Gaussian aspects of the innovation distribution can be assigned to the observation error distribution. Evaluation of the tails of innovation distributions is also likely to provide valuable information about the tails of the observation distributions.

The QC aspects are primarily related to small numbers of observations in the tails of the distribution. So to get a sufficiently large sample of relevant departure statistics, 18 months of data assimilation system departure statistics (February 2006 to September 2007) were used for these estimates. This was done for a large number of observation types, to determine the distributions that best represented the normalised departures for each of these sets. The model background fields are from the operational incremental four-dimensional variational (4D-Var) assimilation system (Courtier *et al.*, 1994) at ECMWF, taken at appropriate time ( $\pm 15$  min) and at T799L91 (25 km horizontal grid and 91 levels outer loop) resolution.

Figure 1 shows the departure distributions, normalised by the prescribed observation error, for a number of these observation types. The grey crosses represent the data counts for bins of width 0.1 in the range  $\pm 10$  of normalised departures. Traditionally this distribution would be plotted as a histogram, but crosses were easier to see on the figures. To put the focus on the tails of the distribution, the data are plotted on a semi-logarithmic scale. This means a Gaussian distribution shows up as a quadratic function, and an exponential distribution as a linear function. On the figure the best-fit Gaussian distributions (dashed-dotted line) are included. Figure 1(a) shows temperature data in the 150–250 hPa range for all Vaisala RS92 radiosonde measurements in the Northern Hemisphere (NH) Extratropics. A similar plot for the used data is shown in Figure 1(b) (used data are quality controlled data with more than 25% weight after applying the previously used ‘Gaussian

plus flat’ distribution QC). The ‘Gaussian plus flat’ distribution QC and its implementation at ECMWF is described in Andersson and Järvinen (1999) (abbreviated below as AJ99). Vaisala RS92 radiosondes are known to be of very high quality with very low bias, very few gross errors, and with low random errors. Figure 1(c–f) show normalised departure statistics for other conventional observation types and their data distributions for the extratropical regions. Figure 1(c, d) show two different surface pressure observing systems (land surface pressure in the Southern Hemisphere (SH) Extratropics and ship surface pressure in the NH Extratropics), and Figure 1(e, f) show upper air and surface wind observations (aircraft winds from all levels in the NH Extratropics and winds observed by drifting buoys in the NH Extratropics).

The solid black curves on Figure 1 show the best-fit Huber norm distribution. The Huber norm, a Gaussian distribution with exponential distribution tails, is defined in section 5 (Eq. (1)). Because  $f$  in Eq. (1) is first-order continuous, the Huber norm distribution shows up as a quadratic function which smoothly transforms into a linear function in the tails of the distribution. It is seen that the background departure statistics are well described by a Huber norm distribution, because the data in the tails are in good agreement with the solid black curves. Indeed, these results indicate that the Huber norm distribution fits the data much better than a pure Gaussian distribution (dash-dotted curves). The ‘Gaussian plus flat’ distribution previously used in the operational assimilation system at ECMWF is included on Figure 1(a) as a thick grey curve. It is evident that the ‘Gaussian plus flat’ represents the tails of the normalised departure distribution very poorly for radiosonde observations. This is the case for all the variables shown in Figure 1, and for almost all other observation types that have been investigated (not shown). It is worth mentioning that a sum of Gaussian distributions does not produce a Huber distribution, so this is not the explanation for the fat tails.

It is noted that there is a factor of more than 1000 between the data counts in the tails (at 8–9 normalised departures). At the centre of the distribution (up to 2 normalised departures), departures are close to a Gaussian distribution for most observations. There are no indications of flat-tailed distributions, i.e. no indication of standard gross errors where the observed value is unrelated to the background field. There is rather an indication of an exponential distribution for many observations in the range 2–9 normalised departures. For the used data (Figure 1(b)), the departures to a large extent follow a Gaussian distribution. This is because the departures are from the pre-2009 operational assimilation system at ECMWF which applied the ‘Gaussian plus flat’ QC distribution, resulting in a sharp transition from full Gaussian weight to zero weight, as shown schematically in Figure 5(b) below.

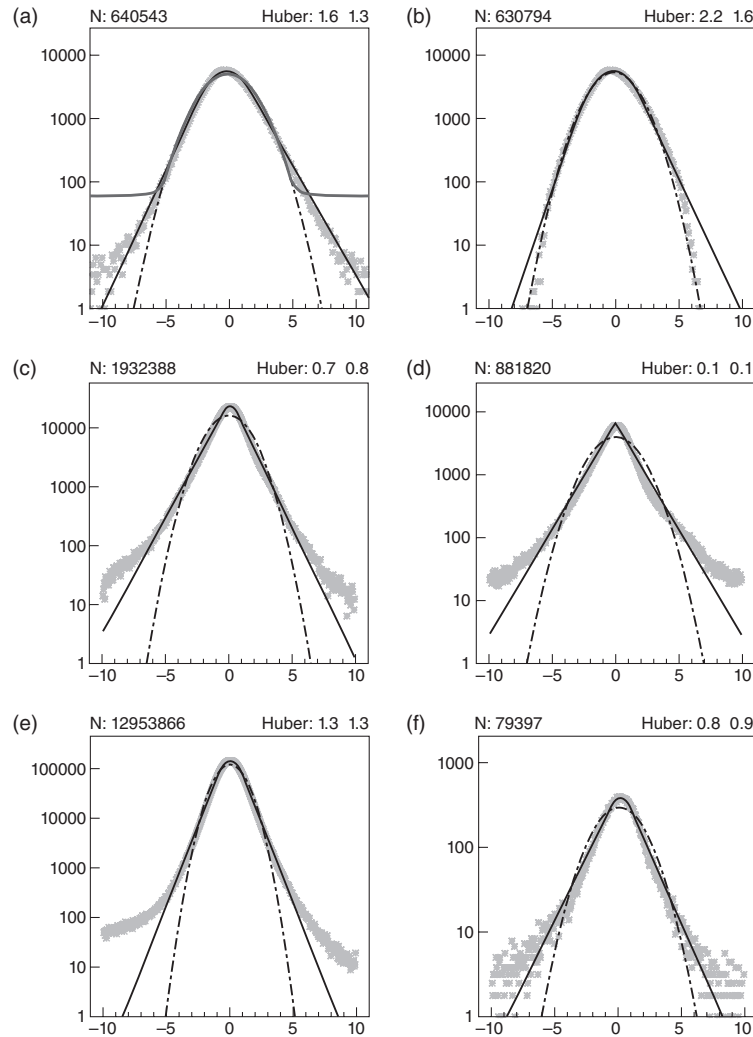
## 3. Gross error QC aspects for observations

Extensive investigation of the normalised background departure statistics for many different observation types and parameters gave a useful insight into gross error aspects. Most distributions have fatter than Gaussian distributions beyond 1–2 normalised departures. The reason why there are only few examples of flat distributions in the tails may well be due to most observing systems now being automated. Automated systems reduce human-related gross errors like swapped latitude/longitude, east/west sign error and swapped digits. If innovation statistics from a station or platform show flat-tail gross error characteristics, it will often be due to a systematic malfunctioning that results in all observations being wrong. It is fairly easy to detect and eliminate (‘blacklist’) these observations via a pre-analysis monitoring procedure. They will then not be part of the observations presented to the analysis. We will now give some examples that highlight these issues.

### 3.1. Chinese aircraft temperature observations

Chinese AMDAR aircraft measurements were reported wrongly from March to May in 2007. Positive (greater than 0 °C)

## Huber Norm Quality Control



**Figure 1.** Innovation statistics, normalised by the prescribed observation error, for (a) all and (b) used Vaisala RS92 radiosonde temperature observations from 150 to 250 hPa in the NH Extratropics. (c) SYNOP surface pressure observations in the SH Extratropics, (d) SHIP surface pressure observations in the NH Extratropics, (e) aircraft wind observations in the NH Extratropics and (f) DRIBU wind observations in the NH Extratropics. The best fit Gaussian distribution (dash-dotted line) and Huber norm distribution (solid line) are included. The best fit Gaussian plus flat distribution is included on (a) as a fat solid grey line. N is the sample size. Huber numbers are the optimal left and right transition point, defined in Eq. (2).

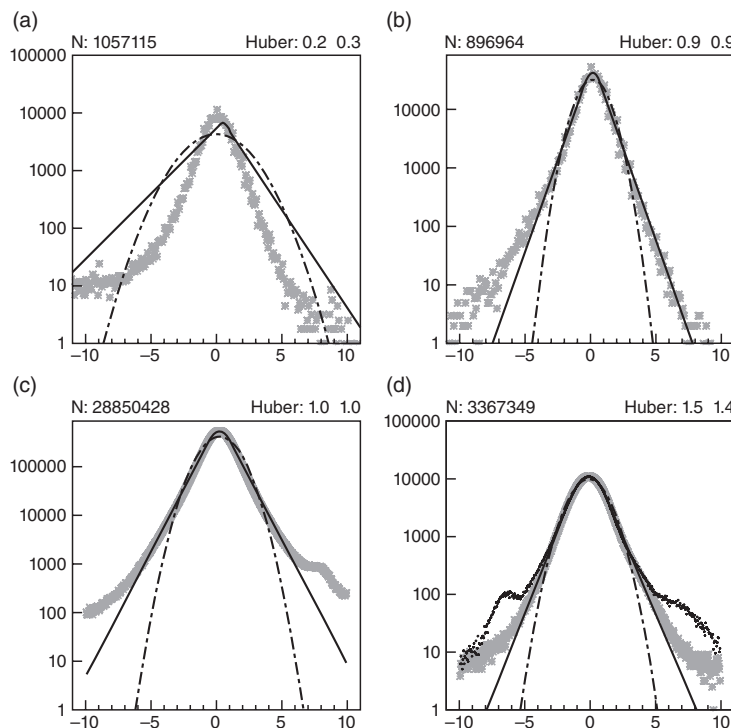
temperatures were reported with the wrong sign as negative Celsius temperatures. This led to a negative tail of gross errors in the innovation statistics. Figure 2(a, b) show the distribution for all AMDAR and ACARS temperature departures for descending aircraft over the NH Extratropics from March to May 2007. The AMDAR data clearly show a large deviation from a Huber distribution for large negative departures which is due to the wrongly reported Chinese measurements. This is one of the few examples of a normalised innovation distribution with an almost flat tail. Over the same period, the ACARS temperature observation departures nicely follow a Huber norm distribution with only a slight misfit for very negative values.

### 3.2. Surface pressure observations

Figure 2(c) shows the distribution of surface pressure departures for NH extratropical synoptic land stations. A hump is clearly identifiable on the positive side of the background departure

distribution. Detailed investigations revealed that this is related to the difference in model orography and station height for some observations. A high percentage of observations with positive background departures between 5 and 10 standard deviations are from stations located in Alpine valleys. The height of these stations tends to be lower than the height according to the forecast model orography since small valleys are not well resolved in the model. Specific QC, like orography difference-related blacklisting, ensures that those observations get rejected so this hump disappears in the distribution of the ‘used’ data.

Figure 2(d) shows the importance of not including blacklisted data in the estimation of the most suitable observation departure distribution. This example shows how the tropical METAR surface pressure data fit a Huber distribution well after excluding blacklisted data. It should be noted that the blacklisting is performed as a completely independent task to identify stations of consistently poor quality. This underlines the necessity of a good blacklisting procedure. It also shows the power of Huber norm distribution plots as a diagnostic tool to identify such outliers.



**Figure 2.** Departure statistics for (a) temperature data for all AMDAR (primarily data from European, Chinese and Japanese aircraft) descending over the NH Extratropics; (b) as (a), but for ACARS (primarily data from North American aircraft); (c) SYNOP surface pressure; and (d) METAR surface pressure. In (d) the black dots show the data before blacklisting.

### 3.3. Humidity

Statistical distributions of humidity departures depend a lot on the selected variable. The innovation statistics for specific or relative humidity are far from Gaussian or Huber distributed, even after normalising by the specified observation error. A variable transformation, as the one used operationally at ECMWF (Hólm *et al.*, 2002; Andersson *et al.*, 2005), ensured a better fit. Figure 3(a) shows the distribution of radiosonde relative humidity departures, whereas Figure 3(b) shows statistics for humidity data normalised by the average of analysis and background data values, mimicking the variable transform method used at ECMWF. It is clear that relative humidity departures are poorly fitted by a Huber distribution, whereas the normalised data provide a reasonable fit.

### 3.4. Satellite data

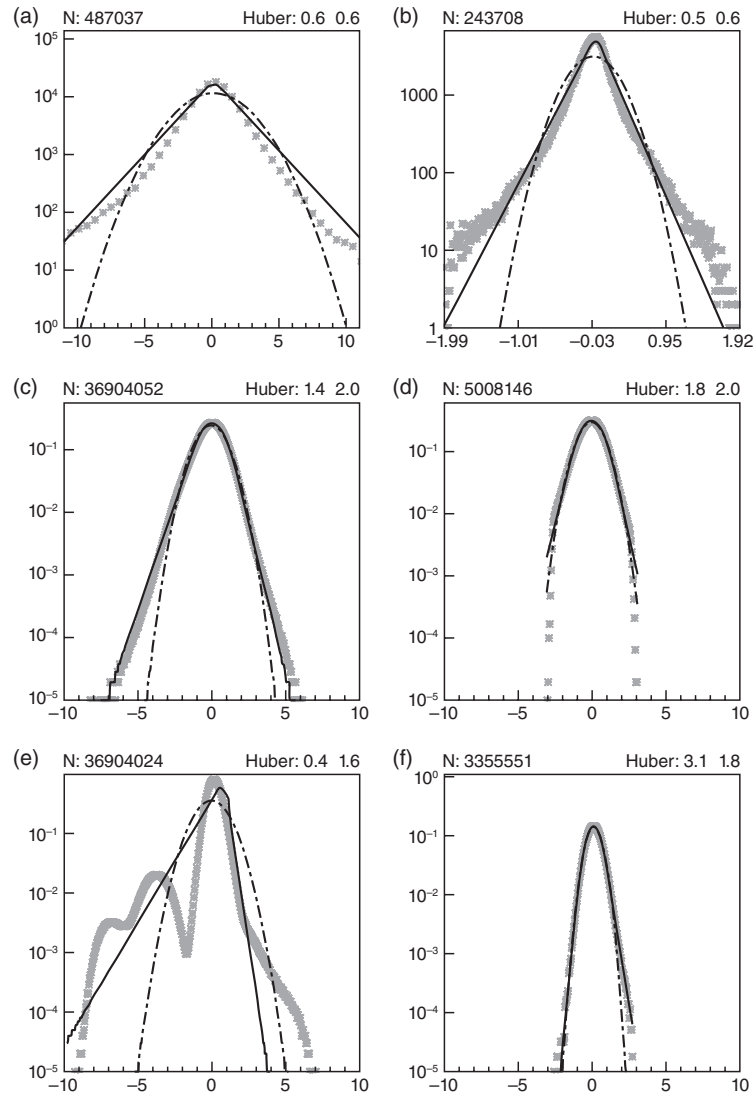
In general it is more difficult to find regular distributions that fit satellite data departures well. Therefore we have not implemented a more relaxed QC for satellite data. We will discuss three reasons for this here. Firstly, most satellite data provide less detailed information than conventional data. The satellite data usually describe the broad features well for the whole swath area. The data seldom pinpoint small-scale weather events, for which a relaxed QC will make the biggest difference. Secondly, even though satellite data departures, for e.g. channels that peak in the stratosphere, typically follow a Huber norm distribution, they are more in accordance with a Gaussian distribution than conventional data. An example is given in Figure 3(c) (all data) and Figure 3(d) (used data) for AMSU-A channel 14, showing that there is a smaller benefit of switching to a Huber distribution. Thirdly, some satellite channels are contaminated by cloud and rain, leading to distributions with large humps, as shown in

Figure 3(e), where all data for mid-troposphere peaking AMSU-A channel 7 are shown. This channel's atmospheric signal is contaminated by cloud and surface returns. Strict QC is applied to eliminate the contaminated tails of the distribution. Figure 3(f) shows the departure statistics for the used data for this channel, with the best Huber and Gaussian distribution included. Note that these plots, with their log-scaling and optimal Gaussian and Huber norm curves, also provide valuable diagnostic information. For example, the two plots for the AMSU-A channel 7 case identify that the cloud clearing is done very well, but it is not perfect for warm departures. Likewise, from comparing Figure 3(c and d), it is evident that the first-guess QC is too strict on AMSU-A channel 14 data. The normalised departure QC limits are 2–3, whereas without problem these could be increased to 6. Further investigations of relaxing background QC and using a Huber norm QC for satellite data will be done in the near future.

### 4. Bias correction problems for isolated biased observations

It is always difficult for an assimilation system to identify problematic isolated observations with large biases. This is a bigger problem when applying an observation QC that allows the use of outliers, because this effectively means relaxing the background QC considerably. This problem was identified at ECMWF in 1998 when hourly SYNOP surface pressure data were assimilated in the first 12 h 4D-Var implementation. Biased isolated stations influenced the analysis negatively. The solution was to identify time-correlated observation errors (Järvinen *et al.*, 1999) to reduce the likelihood of giving frequently reporting biased observations too high a weight. In 2005, ECMWF implemented a station-based bias correction scheme that dynamically corrected surface pressure observation biases (Vasiljevic *et al.*, 2006). This scheme only bias corrected observations in the range  $\pm 15$  hPa, which was

## Huber Norm Quality Control



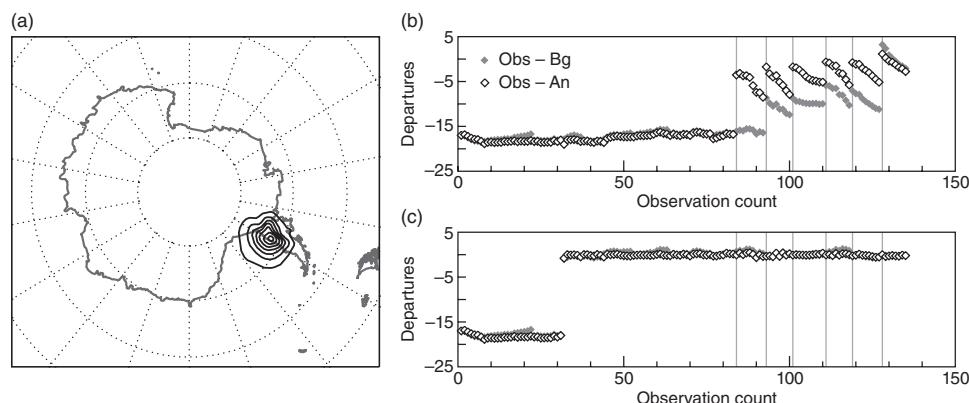
**Figure 3.** Departure statistics for (a) radiosonde relative humidity innovation distributions in the Tropics around 850 hPa. (b) Normalised humidity for radiosonde observations. (c) All brightness temperature departures from METOP-A AMSU-A channel 14 (stratospheric, peaking at 1 hPa) for the SH Extratropics. (d) is as (c) but for used data. (e) is as (c), all data, but for channel 7 (tropospheric, peaking at 250 hPa). (f) is as (e) but for used data.

safe because the then operational background check had much tighter limits than that.

With the introduction of the surface bias correction scheme, there was no longer a need to account for time-correlated observation errors in the assimilation system, so the scheme was abandoned. Relaxing the background check limits with the Huber norm implementation brought the problem back, because data with very large departures were allowed into the assimilation system again, and by mistake the limits for the surface pressure bias correction scheme were not extended accordingly. This provided a very useful reminder that relates to a similar problem to that found by Järvinen *et al.* (1999) related to remote surface stations with very large departures due to observation biases. A similar problem was also recently identified (H. Hersbach, 2014; personal communication) during the testing of the ECMWF ERA-20C surface pressure only reanalysis, where biased measurements from an isolated Pacific island station were spuriously assimilated. Therefore handling of biases is an important aspect to consider

when developing an observation QC scheme. One example encountered during the development of the Huber norm QC scheme was the Antarctic station with WMO station identifier 89622. This station reported surface pressure hourly with an almost constant bias of 18 hPa in this data-sparse area, very likely due to a mis-specified station altitude. Figure 4(b) shows that until 26 December 1999 (observation counts 1–83, marked with the first vertical grey line) almost all the observations from this station were background QC rejected, leading to almost identical background and analysis departures for the first week. At this point the departures are reduced slightly, so some of the data just pass the relaxed background QC and become active observations in the subsequent 12 h analysis (observation counts 84–92 on Figure 4(b)). Each observation initially only got a small weight, but because the biased observations and the observation errors are strongly correlated, the sum of these small weights managed to draw the analysis somewhat towards the biased observations. The background forecast tried to correct the spurious analysis,





**Figure 4.** (a) Surface pressure difference of the Huber norm experiment to ERA-Interim; each black contour is 1 hPa, and solid lines indicate negative differences. Time series of obs minus background (grey diamonds) and obs minus analysis (open diamonds) departures for WMO station 89266 for the Huber norm experiment (b) without and (c) with relaxed surface pressure bias correction limits.

but each subsequent analysis was drawn more and more towards the biased observations. After four additional analyses cycles, identified by the grey vertical lines on Figure 4(b), the background has moved approximately 8 hPa towards the biased observations. The surface pressure bias correction was then activated in the next analysis cycle, after a spin-up period of five cycles, removing the bias between background and observation values (symbols after the last vertical grey line on Figure 4(b)). Because the analysis within those five cycles had already been moved to a biased state by these uncorrected observations, the bias correction applied was 9 hPa. This introduced the spurious (9 hPa too deep) analysis difference seen on Figure 4(a). The difference was maintained for the subsequent period (not shown).

To avoid this problem, the limit for the surface pressure bias correction was extended beyond the values of the relaxed background check. Figure 4(c) shows the departure time series for an assimilation experiment with this extended limit of  $\pm 25$  hPa for the surface pressure bias correction scheme. After a spin-up period (at 30 observation counts on Figure 4(c)), the observations were bias corrected and used in the assimilation system. A bias correction of 17 hPa was performed and resulted in background departures very close to 0 for subsequent analyses, due to the simple almost constant bias pattern for this station. This example shows that careful bias correction is required, especially for remote frequently reporting stations, when very relaxed background limits are used. The problem was identified only because a control analysis was available.

## 5. Potential candidates for distributions with outliers and fat tails

The treatment of outliers in datasets has been discussed for centuries. The simplest method is to assume outliers are gross errors that are then discarded from the analysis of the data. In the 1960s Tukey (1960) and others investigated statistical methods that reduce the problems associated with the large sensitivity to outliers for the estimation of mean and standard deviation of a data sample assumed to follow, e.g. a Gaussian distribution. Tukey (1960) proposed to use a mixture of a one-sigma Gaussian plus a three-sigma Gaussian that represented the effect of fat tails in the distribution. Huber (1964) developed the concept of robust estimation where outliers could be accepted without ruining the estimation process. This method will be discussed further in section 6. Huber (2002) discussed important aspects of robust estimation methods. There are several methods for handling outliers, as the best method depends on the structure of the data and the users' view on the value or risk of outliers. In NWP, outliers occur due to erroneous observations (gross errors),

valuable observations that can help to correct a poor background forecast, or observations that cannot be represented by the forecast model (representativeness errors). It is not always easy to distinguish between these groups of outliers. Robust methods are powerful because they allow the inclusion of outliers, but with some inbuilt safety so that the estimation of mean and standard deviation is less sensitive to the outliers. The most drastic robust method is to eliminate outliers completely. The background error QC described in section 6.6 is an example of such a method. The simplest is then to assume the remaining data are correct and follow e.g. a Gaussian distribution. The 'Gaussian plus flat' distribution is a refinement with a grey zone between correct data and gross error data. A certain small percentage of the data is assumed to be gross errors, without information, which follow a flat distribution. The remaining data are assumed to follow a Gaussian distribution. The variational QC which was used at ECMWF from September 1996 to September 2009 was based on such a formulation. The method and implementation was described in AJ99. The implementation in the variational data assimilation system at ECMWF is technically simple, resulting in only very minor modifications of the nonlinear, tangent linear and adjoint model code. The implementation was based on Dharssi *et al.* (1992) and Ingleby and Lorenc (1993), who argued for the use of a 'Gaussian plus flat' distribution, which assumes all outliers represent gross errors that are completely independent of the background field and therefore provide no useful information for the analysis. The Gaussian distribution and the flat distribution, describing the fraction of outliers, are estimated from large samples of innovation statistics and depend on the quality of each observing system and variable. For small innovations, a Gaussian distribution is typically a good assumption, but for outliers this is often not a correct or safe assumption. It should be noted that robust estimation is used in several areas outside NWP, e.g. finance, noise reduction of images and seismic data analysis (Guitton and Symes, 2003).

## 6. Aspects to consider when implementing a Huber norm QC

### 6.1. Definition and formulation of the Huber norm

Gross errors that are well represented by a flat distribution do exist for some observations, as discussed in section 3.1, but it is evident from Figures 1–3 (and many similar figures not shown) that the flat distribution is usually a poor representation of outliers. There is evidence that the majority of outliers cannot be considered as gross errors, but rather providers of some relevant information. This leads to fat tails in the distributions. In this article, it is



## Huber Norm Quality Control

identified that these fat tails are well represented by a Huber norm.

The Huber norm distribution is defined as a Gaussian distribution in the centre of the distribution and an exponential distribution in the tails. Equations (1) and (2) define the Huber norm distribution as it was introduced by Huber (1972):

$$f(x) = \frac{1}{\sigma_o \sqrt{2\pi}} \exp \left\{ -\frac{\rho(x)}{2} \right\} \quad (1)$$

with

$$\rho(x) = \begin{cases} \frac{x^2}{\sigma_o^2} & \text{for } |x| \leq c, \\ \frac{2c|x| - c^2}{\sigma_o^2} & \text{for } |x| > c, \end{cases} \quad (2)$$

where  $c$  is the transition point, which is the point where the Gaussian part of the distribution ends and the exponential part starts. The definition ensures that  $f$  is continuous and the gradient of  $f$  is continuous. In our implementation, we allow the transition points to differ on the left ( $c_L$ ) and the right ( $c_R$ ) side of the distribution, enabling a better fit to the departure data.

The observation cost function (Lorenc, 1986) for one datum is defined as

$$J_o^{\text{QC}} = -\frac{1}{2} \ln(p^{\text{QC}}) = -\frac{1}{2} \ln[f(x)] = \rho(x) + \text{const.} \quad (3)$$

Note that  $J_o^{\text{QC}}$ , with the Huber norm distribution applied, is an  $L^2$  norm in the centre of the distribution and an  $L^1$  norm in the tails. This is the reason why the Huber norm QC is a robust method which allows the use of observations with large departures. Huber (1972) showed that the Huber norm distribution is the robust estimation that gives most weight to outliers; a higher weight on outliers makes the estimation of statistical moments theoretically unsafe.

Figure 5(a) shows the cost function,  $J_o^{\text{QC}}$ , for the Huber norm distribution, the pure Gaussian distribution, and the 'Gaussian plus flat' distribution. It is clearly seen that the pure Gaussian distribution has large values and large gradients for large normalised departures. The 'Gaussian plus flat' distribution has gradients close to zero for large departures. The Huber norm distribution is a compromise between the two.

Following AJ99, we define the weight applied to an observation as the ratio between the applied  $J_o^{\text{QC}}$  and the pure Gaussian  $J_o$ :

$$W = \frac{J_o^{\text{QC}}}{J_o^{\text{Gaussian}}}. \quad (4)$$

This defines how much the influence of the observation is reduced compared to the influence based on a pure Gaussian assumption. The definition of  $f$  in Eq. (1) ensures that the same weight factor is applicable to the gradient of the cost function, which controls the influence of the observations in the analysis. Figure 5(b) shows  $W$  for the three distributions discussed, as a function of departures normalised by the observation error standard deviation,  $\sigma_o$ . Near the centre of the distribution, both the Huber norm distribution and the 'Gaussian plus flat' distribution follow a Gaussian distribution, i.e.  $W = 1$ .

It can be seen that the 'Gaussian plus flat' distribution has a narrow transition zone of weights from 1 to 0, whereas the Huber norm has a broad transition zone. For medium-sized departures, the Huber norm reduces the weight of the observations and for large departures the weight is significantly higher.

A major benefit of the Huber norm approach is that it enables a significant relaxation of the background QC. With the previous QC implementation, rather strict limits were applied for the background QC, with rejection threshold values of the order

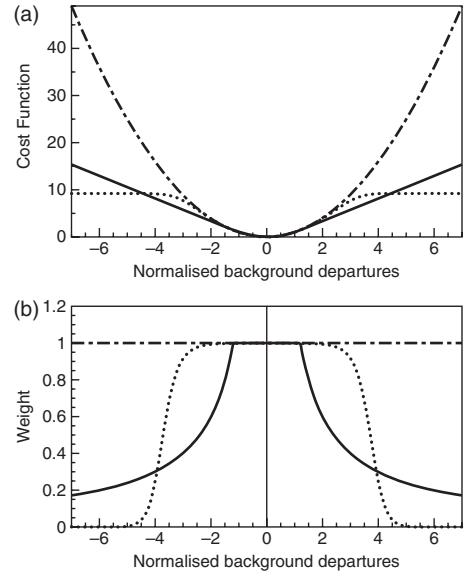


Figure 5. (a) Observation cost functions and (b) the corresponding weights after applying the variational QC; Huber norm distribution (solid), 'Gaussian plus flat' distribution (dashed), and Gaussian distribution (dash-dot).

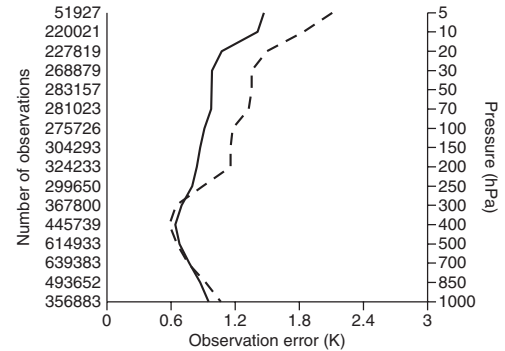
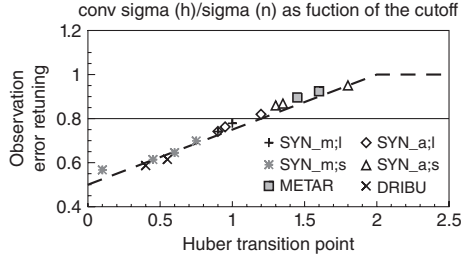


Figure 6. Profile of estimated observation errors (dashed) for Vaisala RS92 radiosonde temperature data over the NH Extratropics compared to the used observation errors (solid). The left axis shows the data count, and the right axis the pressure (hPa).

of 5 standard deviations of the normalised departure values. For the implementation of the Huber norm, this has been relaxed considerably, as discussed in section 6.6. This is especially beneficial for extreme events, e.g. where an intensity difference or a small displacement of the background fields can lead to very large departures. Examples of this will be shown in section 7.

## 6.2. Retuning of observation error

The quality of each observation type is quantified by  $\sigma_o$ , the observation error standard deviation. While implementing the new variational QC scheme, a retuning of  $\sigma_o$  was done with the guidance from estimated observation errors (Desroziers *et al.*, 2005). This led to changes in the observation errors for radiosonde temperature measurements at high altitudes (above 200 hPa; Figure 6) and a retuning of the observation errors used for automatic and manual surface pressure measurements from ships. At the same time, airport surface pressure observation errors were adjusted to be similar to the observation errors



**Figure 7.** Used observation error tuning function (dashed line). The symbols indicate the ratio between the Gaussian and the Huber standard deviation for different kinds of surface pressure observations. SYNOP observations are split into manual (m) and automatic (a) as well as land (l) and ship (s). Every observation type is evaluated for three regions: Tropics and SH and NH Extratropics.

applied to automatic surface stations. The evaluation of the 18 months of departure data clearly supported these adjustments.

A retuning of the observation error was implemented for all data types for which the Huber norm VarQC was applied. This is highly recommendable because the specified observation error in the Huber norm implementation represents the good data in the central Gaussian part of the distribution, whereas it had to represent the whole active dataset in the old method. So theoretically the observation error should be reduced, especially for datasets with a small Gaussian range, i.e. with small Huber transition points.

We examined this for all the observing systems for which a Huber norm distribution was applicable. The symbols on Figure 7 show the ratio of the estimated  $\sigma_o$  for the optimal Huber norm distribution and the optimal value for a Gaussian distribution for a range of surface pressure observing systems. Values are plotted as function of the average Huber left and right transition points ( $c_L$  and  $c_R$ ) for three different areas: NH Extratropics, Tropics and SH Extratropics. The selected observing systems cover a wide range of Huber transition points. It was found that on average the observation error is reduced to 80% of the previously used value. There is an approximately linear relationship between the observation error retuning factor and the Huber norm transition point.

The retuning factor can be estimated well with the simple function

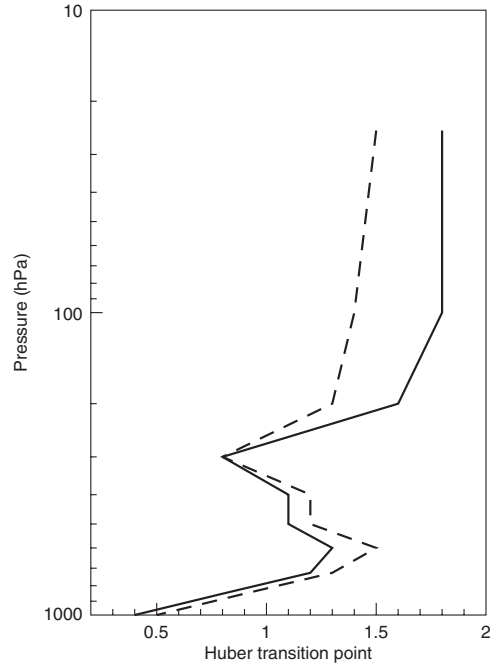
$$T_{\sigma_o} = \min \left\{ 1.0, 0.5 + 0.25 \left( \frac{c_L + c_R}{2} \right) \right\}. \quad (5)$$

Here  $T_{\sigma_o}$  is the retuning value for a certain observation. We choose this simple linear function as it described the relation very accurately (dashed curve on Figure 7). There was no justification for implementing a more complex or statistically based tuning function.

### 6.3. Determination of the optimal Huber distribution and evaluation of the Huber norm transition points

The Huber distributions had to be computed for a large number of observing systems, their associated variables, for various layers for profiling data, and for each channel for satellite data. It was therefore beneficial to develop an objective method to determine the 'optimal' Huber distribution.

The algorithm is described here. First the bias of each data sample is removed, as it is considered an independent problem to address systematic errors. Therefore, as described in section 6.1, the Huber distribution is uniquely defined by  $\sigma_o$ ,  $c_L$  and  $c_R$ . Note that the  $\sigma_o$  described the standard deviation of the central sample data of normalised departures between  $c_L$  and  $c_R$ . So if  $c_L$  and  $c_R$  are very large, the  $\sigma_o$  becomes identical to the value for the 'optimal' Gaussian distribution (shown with dashed-dotted curves on Figures 1–3).



**Figure 8.** Profile of the optimal left (solid) and right (dashed) transition points for Vaisala RS92 radiosonde temperature data for each 100 hPa layer.

The optimal  $c_L$  and  $c_R$  for the Huber norm distribution were determined for each observation type and variable by searching among values in the range 0.0–5.0 in steps of 0.1. The best Huber norm fit was established by least-square-like curve fitting of normalised departures. This was done by computing a cost function, for each  $(c_L, c_R)$  pair, that describes the misfit between Huber distribution and the data sample. The misfit is defined as:

$$\sum_{i=1}^n [p(x_i) \ln\{p(x_i)\} - H(x_i) \ln\{H(x_i)\}]^2, \quad (6)$$

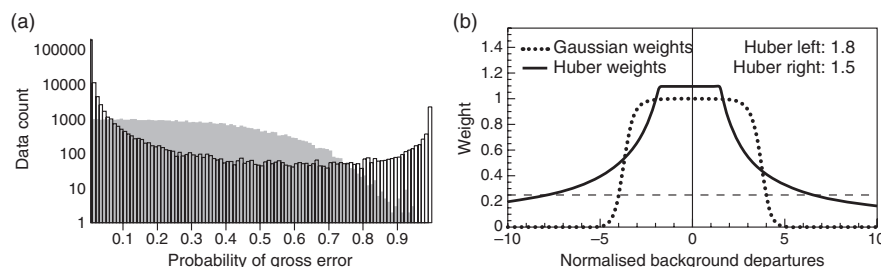
where  $p(x_i)$  is the population in range bin  $i$  and  $H(x_i)$  is the population expected for the specific Huber distribution in the range bin  $i$ .

Typically the selected  $c_L$  and  $c_R$  values are identical or close to each other. It should be noted that different  $c_L$  and  $c_R$  introduce a bias due to the heavier tail to either left or right. This has a very small impact on the bias, because the asymmetry due to this typically represents much less than 1% of the data sample.

The evaluation of Huber transition points in general also confirmed the wide range of values for different observing systems and variables. It would be suboptimal to use a fixed value of (say) 1.0.

For profiling data, the vertical distributions of the Huber left and right transition points were computed for each 100 hPa vertical level. Figure 8 shows an example of this for Vaisala RS92 radiosonde temperatures. Investigations showed that the Huber norm transition points tended to be distinct for three layers in the atmosphere: the stratosphere (observations above 100 hPa), the free troposphere (observations between 100 and 900 hPa), and the boundary layer (observations below 900 hPa). So Huber norm distributions were computed and applied for these three layers for radiosonde, pilot, aircraft and wind profiler data. Notice that the transition points shown in Figure 8 differ in the stratosphere for the left and right transition points. This flexibility in the formulation gives us the opportunity to account for differences in the behaviour of the negative/positive temperature departures. Because we use departure distributions for the

## Huber Norm Quality Control



**Figure 9.** Illustration of the VarQC weights for a Huber norm compared to a Gaussian and flat distribution. (a) Data count as a function of the probability of gross errors, with the previously used ‘Gaussian plus flat’ distribution (black open bars) and the Huber norm distribution (grey solid bars), and (b) the corresponding weight functions for the two distributions. Huber values were taken from the radiosonde temperature observations in the stratosphere ( $\leq 100$  hPa).

evaluation, it is not clear if the observations or the background fields are responsible for asymmetric behaviour in the tails of the distribution. It could be questioned why the left and right transition points for radiosondes should change with height. This could possibly be linked to representativeness errors which in the ECMWF system (and most other assimilation systems) are treated as part of the observation error. But in several cases we are able to link asymmetries of temperature departures to issues with the observing system. Of course it is preferable to correct for systematic observation errors and model errors closer to the source.

#### 6.4. Huber norm VarQC implementation at ECMWF

Further aspects that need to be considered when implementing the Huber norm VarQC in an operational NWP system are discussed here using the ECMWF operational implementation as an example. Because Huber norm VarQC is a robust method, it allows the relaxation of the background QC. This is a very important side benefit of the Huber norm method, because it makes observations with large departures active, so the data get a chance to influence the analysis. The observation errors were also adjusted as discussed in section 6.2.

The weights,  $W$ , are computed based on the high-resolution departures in the nonlinear outer loop of the incremental 4D-Var (Courtier *et al.*, 1994). The weights are kept constant during the minimisation (inner loop), because the Lanczos minimisation algorithm (Fisher *et al.*, 2009) used at ECMWF does not allow the function that is minimised to be modified during the minimisation process. Some minimisation methods are more lenient and would allow the weights to be adjusted slightly for each iteration of the minimisation process. But the benefit of the much faster, but strictly quadratic, Lanczos algorithm outweighs the benefit of a more dynamic QC. The weights are updated at each of the three relinearisation outer loops applied at ECMWF; this makes it possible for the analysis to change the weights during the assimilation cycle.

In this article we concentrate our investigation on conventional observations. As mentioned in section 3.4, it is expected that the Huber norm QC will be most beneficial for conventional data. Of the conventional observing systems used in ECMWF’s assimilation system, it was found that the distributions for the following observation types and variables were very well represented by a Huber norm distribution:

- Radiosonde observations: temperature and wind upper-air data (with special Huber norm distributions fitted to dropsondes);
- Aircraft observations: temperature and wind upper-air data;
- Pilot balloon observations: wind upper-air data;
- Wind profiler observations: wind upper-air data from American, European and Japanese wind profilers;
- Land surface observations: surface pressure data from automatic and manual SYNOP reports;

- Ship observations: surface pressure and wind data from automatic and manual SHIP reports;
- Airport observations: surface pressure data from METAR reports;
- Drifting and moored buoy observations: surface pressure and wind data from drifting and moored buoys.

So these observation types were all included in the operational analysis system update of the variational QC. The remaining observation types and variables kept the ‘Gaussian plus flat’ distribution.

The Huber norm QC is not implemented for humidity in the present implementation at ECMWF due to the difficulties discussed in section 3.3. It is planned to implement this in a forthcoming update.

#### 6.5. Weights for Huber norm VarQC

Following the definition from AJ99, we define the probability of gross error, scaled to the range 0.0–1.0, to be  $1 - W$ . Figure 9(a) shows the distribution of gross error probabilities for the 18 months sample of stratospheric radiosonde temperature data. The transparent black bars are for the previously used ‘Gaussian plus flat’ distribution and the grey shaded bars are for the Huber norm distribution. Note the vertical scale is logarithmic and bars have a width of 0.01. It is seen that more than 99% (100 000) of the observations have gross error probabilities below 0.01. This is the case for both distributions. In the gross error probability range from 0.01 to 0.5 the Huber norm has similar data counts in each bin. For higher values the data counts fall off, because there are so few data values in the extremes of the departure distribution. For the ‘Gaussian plus flat’ distribution bin data counts are reduced between 0.01 and 0.5, and reach a level that is an order of magnitude lower than for the Huber norm distribution at 0.5. For higher gross error probabilities the data counts are increased for the ‘Gaussian plus flat’ distribution – a result of the sharp transition zone for gross error probabilities closer to the centre of the distribution, resulting in more observations with large probability of gross error values. Figure 9(b), similar to Figure 5(b), shows the corresponding weights for the optimal Huber norm distribution and the previously used ‘Gaussian plus flat’ distribution. It gives a qualitative understanding of the different shape of data counts for the gross error probabilities shown in Figure 9(a).

#### 6.6. Relaxation of the background QC

Before the introduction of the Huber norm VarQC, the background QC had rather strict limits. Typical standard deviation values ( $\alpha$ ) would be around 5 (Järvinen and Undén, 1997) for the normalised departures,  $(o - b)^2 < \alpha^2(\sigma_o^2 + \sigma_b^2)$ , where  $\sigma_o$  and  $\sigma_b$  are the observation and background error standard deviations, respectively. For the Huber norm VarQC, this has been relaxed

## C. Tavalato and L. Isaksen

Table 1. Data usage table showing the background QC and the VarQC rejections for 15 November 2008 to 31 December 2008 of operational data (Old) and the Huber norm assimilation experiment (Huber). VarQC rejected is defined by a weight smaller than 25%. The count for all observations is in thousands and is the same for both datasets. More detail is given in the text.

Obstype	Value	Level	All obs *1000	% FG rej		% VarQC rej		BG QC limits		VarQC rej limits		
				Old	Huber	Old	Huber	Old	Huber	Old	Huber	
SYNOP	Ps	surf	5373	0.58	0.19	0.11	0.42	260.0	780.0	200.0	140.0	Pa
SHIP	Ps	surf	360	0.94	0.17	0.97	2.56	280.0	1100.0	200.0	180.0	Pa
SHIP	U/V	surf	350	0.77	0.02	0.43	5.44	11.2	12.7	10.8	5.4	m s <sup>-1</sup>
DRIBU	Ps	surf	1156	1.17	0.55	0.47	0.97	360.0	800.0	200.0	200.0	Pa
DRIBU	U/V	surf	111	4.05	0.77	1.56	6.63	10.7	26.3	7.4	4.3	m s <sup>-1</sup>
METAR	Ps	surf	2070	0.05	0.00	0.09	0.07	1000.0	> 1600	340.0	80.0	Pa
TEMP	T	0-100	693	0.96	0.04	2.03	0.15	5.2	29.0	3.6	6.6	K
TEMP	T	100-900	1614	0.54	0.02	0.66	0.70	3.3	15.8	2.5	2.5	K
TEMP	T	1000-900	188	0.88	0.03	1.45	4.33	5.1	21.8	3.6	2.6	K
TEMP	U/V	0-100	716	0.49	0.09	0.78	0.61	13.9	22.5	10.2	11.5	m s <sup>-1</sup>
TEMP	U/V	100-900	1237	0.35	0.08	0.39	0.79	11.2	23.5	9.1	6.5	m s <sup>-1</sup>
TEMP	U/V	1000-900	189	0.44	0.06	0.48	2.67	11.1	30.9	9.2	6.5	m s <sup>-1</sup>
AIREP	T	100-900	8477	0.08	0.00	0.05	0.19	4.4	15.9	3.8	1.4	K
AIREP	T	1000-900	1529	0.40	0.03	0.09	1.77	6.2	23.9	5.0	1.5	K
AIREP	U/V	100-900	8483	0.09	0.02	0.08	0.28	~15.0	~21.5	12.7	9.1	m s <sup>-1</sup>
AIREP	U/V	1000-900	1483	0.63	0.17	0.11	0.62	~15.0	no data	12.5	8.9	m s <sup>-1</sup>
PILOT	U/V	0-100	238	0.53	0.04	0.81	0.71	14.3	24.6	10.3	11.6	m s <sup>-1</sup>
PILOT	U/V	100-900	536	0.39	0.04	0.61	1.27	11.6	23.4	9.2	6.5	m s <sup>-1</sup>
PILOT	U/V	1000-900	100	0.32	0.03	0.32	2.20	11.5	51.4	9.2	6.5	m s <sup>-1</sup>
profiler	U/V	0-100	73	0.90	0.15	0.52	0.65	15.9	22.0	10.7	12.2	m s <sup>-1</sup>
profiler	U/V	100-900	4061	0.15	0.03	0.10	0.25	12.7	22.2	9.2	6.5	m s <sup>-1</sup>
profiler	U/V	1000-900	346	0.01	0.00	0.02	0.06	13.2	no data	9.2	6.5	m s <sup>-1</sup>
EU-profiler	U/V	0-100	8	0.41	0.00	0.71	0.52	17.3	no data	10.7	12.4	m s <sup>-1</sup>
EU-profiler	U/V	100-900	2036	0.08	0.02	0.06	0.13	12.7	24.2	9.2	6.5	m s <sup>-1</sup>
EU-profiler	U/V	1000-900	246	0.01	0.00	0.02	0.08	13.2	no data	9.2	6.5	m s <sup>-1</sup>
JP-profiler	U/V	100-900	303	0.18	0.01	0.49	0.85	13.2	22.2	9.2	8.4	m s <sup>-1</sup>
US-profiler	U/V	0-100	46	1.36	0.24	0.70	0.93	15.9	22.0	10.7	12.2	m s <sup>-1</sup>
US-profiler	U/V	100-900	1181	0.34	0.07	0.13	0.40	13.4	24.1	9.7	8.9	m s <sup>-1</sup>

to around 15 standard deviations of the normalised innovation departure values. The 'BG QC limits' column in Table 1 shows the background QC values for the 'Gaussian plus flat' distribution (labelled Old) and for the Huber norm QC in details for all the involved observation types and variables. The values shown in Table 1 are absolute values in SI units. It could be argued that a background QC is not necessary any more when a robust estimation in the variational QC is applied, but the relaxed limits are still helpful in rejecting clearly erroneous gross errors, like 0 K temperatures.

## 7. General impact and case-studies

### 7.1. A general summary of QC decisions for the Huber norm implementation

The overall impact of the Huber norm implementation was evaluated over a three month data assimilation period in 2008, and for a number of intense weather events where the Huber norm implementation would be expected to make the biggest difference. For all the experiments presented in this Section the only difference between the control assimilations and Huber norm assimilations are the quality control and error distribution differences described in this article.

Table 1 shows the QC statistics for control (old) and Huber norm assimilations for all conventional observations that use the Huber norm QC. Upper air observation statistics are split up into three vertical bins, as described in section 6.3. The data were evaluated for the period from 15 November 2008 to 31 December 2008. The change in percentage of background rejected (FG rej) data is clear for all observation types, with significantly fewer rejections for the Huber norm assimilation experiment. The percentages of data with very low variational QC weight (less than 25%) are called VarQC rejected data, even though the data are not fully rejected. The data are still active data and influence the analysis according to their reduced weight. As discussed in

section 6.5, the percentage of VarQC rejected data are generally larger for the Huber norm because this is the percentage of a much larger sample that pass the background QC. It is also related to the shape of the probability of gross error distributions, as shown in Figure 9. The final four columns show the approximate limits used by the different QC decisions. The term 'no data' means that no data were background rejected for this data type during the six-week period evaluated. The VarQC limits show the range for which the weights get below 25%.

This change in variational QC was implemented into the operational forecasting system at ECMWF in September 2009 (Tavalato and Isaksen, 2010) and has proved to have a positive impact on the use of conventional observations within the assimilation system.

A number of impact studies and general investigations have been performed to evaluate the impact of the Huber norm QC. Assimilation experiments over a period of three months in 2008 showed a small positive impact over Europe and the NH Extratropics in general, and neutral scores for the SH Extratropics.

During the last week of December 1999, two small-scale lows affected Europe with intense gusts and storm damage. These storms are ideal case-studies due to the high-density, high-quality synoptic land station surface pressure network over France and Germany. These surface pressure observations captured the intensity and location of the storms very well, and neighbouring stations consistently supported each other. However, the strength of these storms was poorly represented in both the operational ECMWF analysis and the ERA-Interim (Dee *et al.*, 2011). Both assimilation systems used the old ('Gaussian plus flat' distribution) QC method.

Several case-studies were performed to investigate the assimilation impact of applying the Huber norm VarQC in the analysis system. The Huber norm experiments were run with the same model version as ERA-Interim, and for most experiments at the same resolution.

## Huber Norm Quality Control

7.2. *Lothar*; 26 December 1999

The first of the December 1999 storms that hit Europe on 26 December 1999 is known as *Lothar* (Ulbrich *et al.*, 2001). It followed a path from the Atlantic to France, moving eastwards into Germany. The position of this storm was well predicted in both analyses (ERA-Interim as well as the Huber norm experiment), but the intensity was not captured well in ERA-Interim. Indeed, the SYNOP observations reporting the lowest surface pressure were background rejected in ERA-Interim. The Huber norm experiment showed a reduced central pressure compared to the reanalysis because many more observations were assimilated. However, the analysis was still significantly above the lowest observed surface pressure. One of the reasons is that the analysis is not able to capture the small scale of this event well enough at the reanalysis resolution.

To evaluate the influence of the resolution, several Huber norm experiments with different resolutions (inner and outer loop) were carried out and the results are shown in Table 2. It shows that increased resolution is beneficial.

7.3. *Martin*, 27 December 1999

The second storm was the very intense *Martin* which reached the French coast on 27 December 1999 (Ulbrich *et al.*, 2001). It was poorly predicted, being too weak and misplaced in the operational ECMWF analysis; ERA-Interim produced similarly poor results. Most surface pressure observations near the cyclone centre were rejected by the background QC (shown as filled triangles on Figure 10(a)) even though a hand analysis showed that all the observations from France were correct. This led

Table 2. The lowest observed and analysed surface pressures on 26 December 1999, 0600 UTC. The Huber norm assimilation experiment deepens the low compared to ERA-Interim. A further improvement is found when the resolution (outer/inner loop) is increased.

System	Resolution	Lowest $P_s$ (hPa)
ERA-Interim	T255 / T95,T159	978.0
Huber expt.	T255 / T95,T159	976.9
Huber expt.	T319 / T95,T159	976.4
Huber expt.	T319 / T95,T255	975.6
Huber expt.	T511 / T95,T255	974.3
Observation	—	962.4

to an analysis with the storm centre further to the east than surface pressure observations would suggest. The lowest surface pressure observation at 1800 UTC on 27 December 1999 reported 963.5 hPa. It was one of the background QC rejected observations in ERA-Interim.

Figure 10(b) shows rejections and observation weights from the Huber norm assimilation experiment. The numbers show the QC weight associated with each surface pressure observation: they are 16% or higher for all stations. More observations get higher QC weights than in the reanalysis due to the Huber norm. The centre of the low has correctly moved further to the west in good agreement with the observations. Furthermore, the minimum surface pressure is reduced significantly.

The analysis and the observation rejections for the December 1999 storm cases have also been discussed by Dee *et al.* (2001). They use an adaptive buddy check QC approach with the same effect as the Huber norm method to analyse this case.

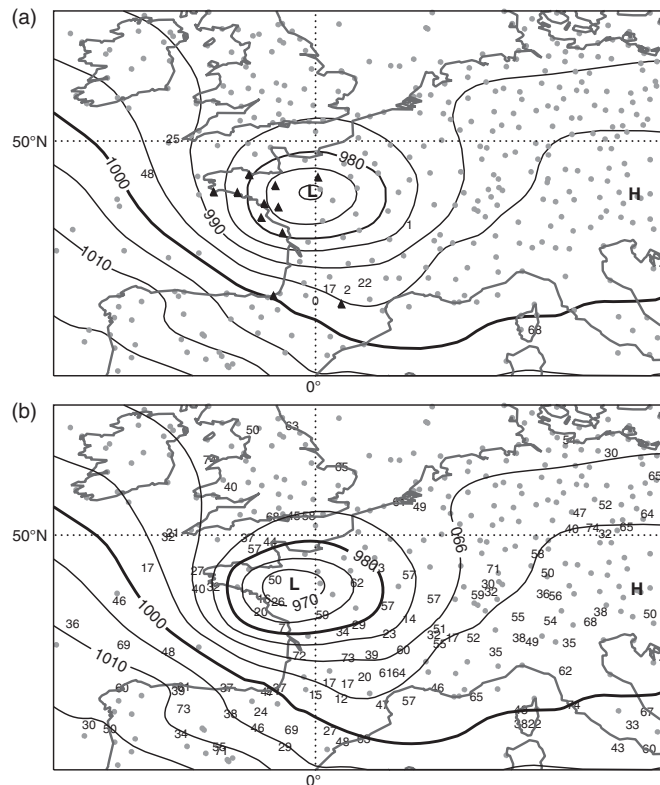
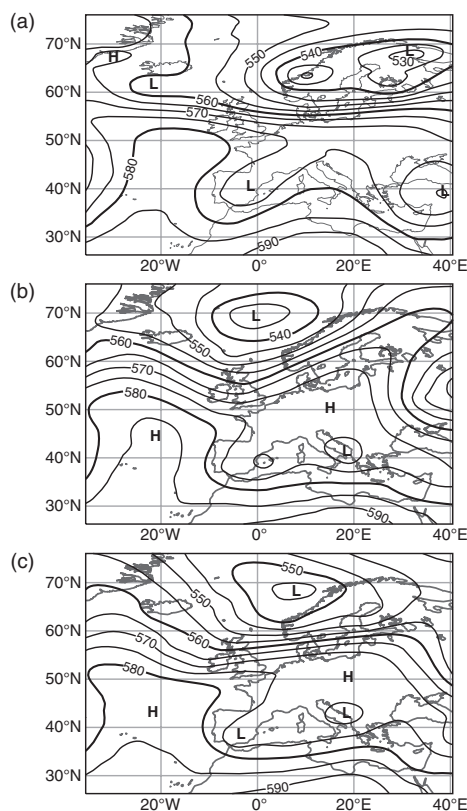


Figure 10. Rejections on 27 December 1999, 1800 UTC, for (a) ERA-Interim, (b) Huber norm experiment. The contours show the analysed surface pressure field for each experiment. Black triangles indicate background rejected observations, and numbers the effective VarQC weights for quality controlled stations. Grey circles indicate observations with weights higher than 75%.





**Figure 11.** Analysis of 500 hPa geopotential height for (a) 11 June 2008, and the 5-day forecast valid at the same time from (b) the operational ECMWF system, and (c) the Huber norm experiment.

However, the Huber norm method is simpler to implement in the IFS.

#### 7.4. June 2008 extratropical event

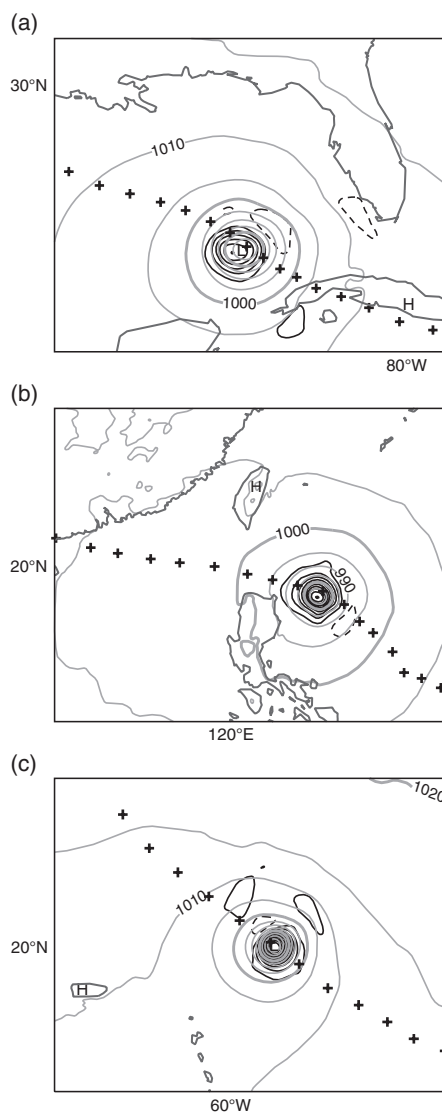
At the beginning of June 2008, exceptionally low forecast scores were seen for the 5-day 500 hPa geopotential height forecast over Europe (anomaly correlation errors for 500 hPa geopotential height were below 0) in several NWP models (not shown).

In the operational ECMWF system, this drop in performance was linked to the rejection (mainly background rejection) of radiosonde and aircraft observations around 200 Pa over North America. Most of the background rejected data had relatively small background departures, just outside the QC limits. Applying the Huber norm VarQC had the effect that all these observations were used and the 5-day forecast improved drastically. Figure 11 shows the verifying analysis over Europe on 11 June 2008 and the two 5-day forecasts. The westerly flow over Europe is predicted much better in the Huber norm VarQC experiment.

#### 7.5. Tropical cyclones

The Huber norm QC and relaxation of rejection limits are also applied for dropsonde wind and temperature observations, resulting typically in more correctly analysed tropical cyclones.

Results for hurricane *Ike*, hurricane *Bill* and typhoon *Hagupit* from September 2008 (*Ike*, *Hagupit*) and August 2009 (*Bill*) are discussed here. The two Atlantic hurricanes were well observed by dropsondes. Usage statistics for this period confirms that more

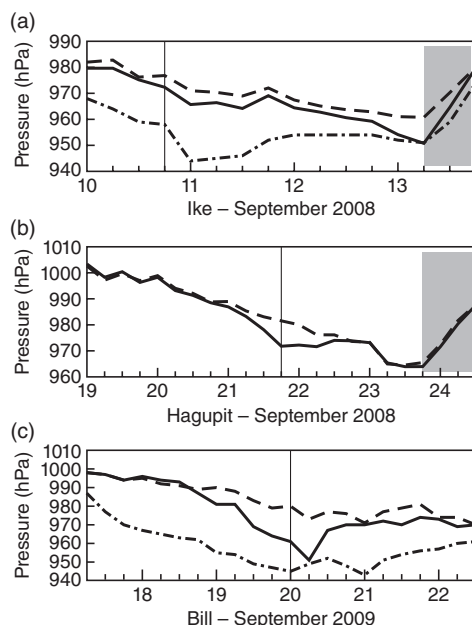


**Figure 12.** Improvement in tropical cyclone analyses due to Huber norm QC. (a) Hurricane *Ike* on the 10 September 2008 1800 UTC in the Gulf of Mexico approaching Texas. (b) Typhoon 0814 *Hagupit* on the 21 September 2008 1800 UTC in the Pacific approaching the Chinese coast. (c) Hurricane *Bill* on the 20 September 2009 0000 UTC in the Caribbean Sea. The grey contours (at 5 hPa intervals) show the MSL pressure analysis, and black crosses indicate the cyclone centre's position every 6 h. Solid/dashed black contours indicate pressure reduction/increase in the surface pressure analysis compared with the control; the contour interval is 1 hPa in (a) and (b) and 2 hPa in (c).

dropsonde wind and temperature data were used in the Huber norm experiment than in the operational system.

Figure 12 shows the observed cyclone track, marked with crosses for every 6 h, and the analysis of surface pressure for the three tropical cyclones at one selected analysis time during the most intense cyclone phase. It is evident that all three tropical cyclones have been intensified very significantly by using the revised observation QC. Figure 13 shows the time series of core surface pressure every 6 h. These results indicate that the use of the Huber norm intensified the core pressure compared with the analysis using the 'Gaussian plus flat' distribution in the QC

## Huber Norm Quality Control



**Figure 13.** Time series of tropical cyclone centre surface pressure (hPa) for the storms (a) *Ike*, (b) *Hagupit* and (c) *Bill*. The solid line shows the surface pressure analysis from the Huber norm assimilation experiment, and the dashed curve is the control experiment. The dash-dotted line shows the observed surface pressure if available. The grey shaded area indicates the time after the landfall of the cyclone and the vertical line marks the date and time used in Figure 12.

for many analysis cycles during the intense phase of the tropical cyclones.

For the Atlantic hurricanes *Ike* and *Bill*, measurements of the core MSL pressure are available. For storm *Hagupit* no core MSL pressure observations are available, but the intensity estimates indicates it developed into a typhoon from a tropical storm on 20 September 2008. This means it was too weak in both analyses. So all three time series show that the Huber norm experiment improved the surface pressure analysis of the tropical cyclones.

It is clear that, when extensive dropsonde data are available, as with hurricane *Bill*, deeper and more accurate analyses were obtained when the Huber norm QC was applied.

## 8. Conclusions

The article describes a number of aspects that are important to consider for QC of observations used in data assimilation systems. Observations have measurement errors, representativeness errors and sometimes gross errors. In data assimilation, innovations are used extensively for observation QC and provide generally very valuable information (Hollingsworth *et al.*, 1986). But the background forecasts also have errors, sometimes very large ones, so it may be difficult to determine if an observation or the equivalent model value is the outlier. Monitoring time series for individual *in situ* stations and satellite channels provide a powerful method for detecting poorly performing or erratic platforms. It is shown how semi-logarithmic plots of normalised departures also are able to identify groups of outliers. Studying these outlier samples often makes it possible to identify problems with observations used in the data assimilation system, e.g. due to representativeness errors in mountainous areas. For humidity data, normalisation is required to obtain Gaussian-like innovations. For satellite data, clouds, rain and surface emissivity may contaminate the atmospheric signal. The semi-logarithmic plots of normalised departures also provide useful guidance in detecting this. A comprehensive evaluation was performed of the

departure statistics for every observing system and every variable used in the ECMWF assimilation system over an 18-month period. After filtering out systematic outliers ('blacklisting' of stations and satellite channels) there is very little evidence of gross errors for the observations used in the departure distributions. This is likely because most observations are now automated and therefore either are of nominal quality or all display gross errors which are easy to detect.

The difficulties related to assimilation of isolated frequently reporting stations with biases is discussed. It is shown how important it is to do bias correction of isolated stations, especially when observation QC limits are relaxed. It is important to consider this when an observation QC scheme is developed.

Various QC methods for handling outliers are presented. The main issue is how much weight to assign to outliers. The 'Gaussian plus flat' distribution method (Andersson and Järvinen, 1999), used at ECMWF from 1999 to 2009, had fairly strict background QC limits and so there was a sharp transition from observations being active to being given virtually no weight in the analysis. The article describes the introduction of a Huber norm QC which makes it safe to use observations with large departures in the analysis. This is because it is a robust method where the moments of the distribution are affected very little by a few erroneous outliers.

Evaluating the 18-month sample of innovation data from ECMWF showed that almost all departure distributions were well described by the Huber norm distribution, after removing systematically erroneous data. The fit was much better than for the pure Gaussian distribution or a 'Gaussian plus flat' distribution. It was also shown to be beneficial to introduce the flexibility of allowing different left and right transition points from the Gaussian to exponential part of the Huber distribution. It is acknowledged that a Huber distribution fit for the normalised innovations does not prove that the observation-error distribution follows a Huber distribution. Innovation distributions are a convolution of observation- and background-error distributions. For the background QC, it is theoretically correct to use the innovation statistics, but for the observation cost function term it is not. But it is shown that it is more beneficial to relax the observation QC and allow outliers to influence the analysis under the assumption that observation errors follow the robust Huber distribution. Several case-studies show how the Huber norm QC deserves the credit for improved analyses and forecasts of extreme events such as extratropical storms and tropical cyclones. The examples show the strength of the robust Huber norm approach which enables the analysis to benefit from observation outliers in the situations when several observations deviate significantly and consistently from the model background. The previously used QC method would reject such observations.

The Huber norm QC has been implemented successfully at ECMWF in September 2009 for wind, temperature and surface pressure measurements from all conventional observations available. In the future this will be extended to humidity and some satellite data.

This work has also shown that refined QC and observation error tuning can be an important method to help extract more information from observations. It is an area of data assimilation where there is potential for further improvements.

## Acknowledgement

The authors want to thank Hans Hersbach (ECMWF) for providing the software to compute the best fit Huber distribution. Elias Hölm and Erik Andersson, both ECMWF staff, provided helpful comments. We thank Rob Hine and Anabel Bowen, ECMWF staff, for improving our figures significantly. Part of this work was funded by project P21772-N22 of the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF).

## References

- Andersson E, Järvinen H. 1999. Variational quality control. *Q. J. R. Meteorol. Soc.* **125**: 697–722.
- Andersson E, Bauer P, Beljaars ACM, Chevallier F, Hólm EV, Janisková M, Källberg P, Kelly G, Lopez P, McNally AP, Moreau E, Simmons AJ, Thépaut J-N, Tompkins AM. 2005. Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bull. Am. Meteorol. Soc.* **86**: 387–403.
- Bonavita M, Isaksen L, Hólm EV. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.* **138**: 1540–1559, doi: 10.1002/qj.1899.
- Courtier P, Thépaut J-N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1388.
- Courtier P, Andersson E, Heckley W, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M, Pailleux J. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.* **124**: 1783–1807.
- Dee DP, Rukhovets L, Todling R, da Silva AM, Larson JW. 2001. An adaptive buddy check for observational quality control. *Q. J. R. Meteorol. Soc.* **127**: 2451–2471.
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Källberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavalato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.
- Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation-, background- and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396.
- Dharssi I, Lorenc AC, Ingleby NB. 1992. Treatment of gross errors using probability theory. *Q. J. R. Meteorol. Soc.* **118**: 1017–1036.
- Fisher M, Nocedal J, Trémolet Y, Wright SJ. 2009. Data assimilation in weather forecasting: a case study in PDE-constrained optimization. *Optim. Eng.* **10**: 409–426.
- Guitton A, Symes WW. 2003. Robust inversion of seismic data using the Huber norm. *Geophysics* **68**: 1310–1319.
- Hollingsworth A, Shaw DB, Lönnberg P, Illari L, Arpe K, Simmons AJ. 1986. Monitoring of observations and analysis quality by a data assimilation system. *Mon. Weather Rev.* **114**: 861–879.
- Hólm EV, Andersson E, Beljaars ACM, Lopez P, Mahfouf J-F, Simmons AJ, Thépaut J-N. 2002. *Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans*, ECMWF Technical Memorandum 383. ECMWF: Reading, UK.
- Huber PJ. 1964. Robust estimates of a location parameter. *Ann. Math. Stat.* **35**: 73–101.
- Huber PJ. 1972. Robust statistics: A review. *Ann. Math. Stat.* **43**: 1041–1067.
- Huber PJ. 2002. John W. Tukey's contribution to robust statistics. *Ann. Math. Stat.* **30**: 1640–1648.
- Ingleby NB, Lorenc AC. 1993. Bayesian quality control using multivariate normal distributions. *Q. J. R. Meteorol. Soc.* **119**: 1195–1225.
- Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. *Ensemble of Data Assimilations at ECMWF*, ECMWF Technical Memorandum 636. ECMWF: Reading, UK.
- Järvinen H, Undén P. 1997. *Observation Screening and Background Quality Control in the ECMWF 3D-Var Data Assimilation System*. ECMWF Technical Memorandum 236. ECMWF: Reading, UK.
- Järvinen H, Andersson E, Bouttier F. 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus* **51A**: 469–488.
- Lorenc AC. 1984. 'Analysis methods for the quality control of observations'. In *Proceedings of ECMWF Workshop on the Use and Quality Control of Meteorological Observations for Numerical Weather Prediction*, Reading, UK, 6–9 November 1984, 397–428.
- Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.
- Lorenc AC, Hammon O. 1988. Objective quality control of observations using Bayesian methods – Theory, and practical implementation. *Q. J. R. Meteorol. Soc.* **114**: 515–543.
- Tavalato C, Isaksen L. 2010. Huber norm quality control in the IFS. *ECMWF Newsletter* **122**: 27–31.
- Tukey JW. 1960. A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, Olkin I, Ghurye S, Hoefding W, Madow W, Mann H. (eds.): 448–485. Stanford University Press: Stanford, CA.
- Ulbrich U, Fink AH, Klawe M, Pinto JG. 2001. Three extreme storms over Europe in December 1999. *Weather* **56**: 70–80.
- Vasiljevic D, Andersson E, Isaksen L, Garcia-Mendez A. 2006. Surface pressure bias correction in data assimilation. *ECMWF Newsletter* **108**: 20–27.



### 3 Radiosonde Temperature Homogenisation

The efforts of comparing the homogenised radiosonde temperature data sets RAOBCORE and RICH to other upper air data sets are published together with Leopold Haimberger and Stefan Sperka in the following paper in the Journal of Climate.

My contribution to Haimberger et al. (2012) was the data evaluation in MSU equivalents which included the calculation of the layer equivalent temperatures for the four MSU layers as well as the calculation of zonal and global statistics of the different data sets. This evaluation allows intercomparison of the unadjusted as well as the homogenised data to satellite data. Furthermore trends and time series at MSU equivalents are widely used to compare different upper-air data in the community and therefore are essential. Providing new upper-air data sets in such a format allows them to be included in data comparison studies such as Blunden et al. (2011); Blunden and Arndt (2012).

Haimberger, L. and Tavalato, C. and Sperka, S., 2012: *Homogenization of the global radiosonde data set through combined comparison with reanalysis background series and neighbouring stations*. J. Climate, 25: 8108-8131.



## Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations

LEOPOLD HAIMBERGER, CHRISTINA TAVOLATO, AND STEFAN SPERKA

*Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria*

(Manuscript received 16 November 2011, in final form 30 May 2012)

### ABSTRACT

This article describes progress in the homogenization of global radiosonde temperatures with updated versions of the Radiosonde Observation Correction Using Reanalyses (RAOBCORE) and Radiosonde Innovation Composite Homogenization (RICH) software packages. These are automated methods to homogenize the global radiosonde temperature dataset back to 1958. The break dates are determined from analysis of time series of differences between radiosonde temperatures (obs) and background forecasts (bg) of climate data assimilation systems used for the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) and the ongoing interim ECMWF Re-Analysis (ERA-Interim).

RAOBCORE uses the obs–bg time series also for estimating the break sizes. RICH determines the break sizes either by comparing the observations of a tested time series with observations of neighboring radiosonde time series (RICH-obs) or by comparing their background departures (RICH- $\tau$ ). Consequently RAOBCORE results may be influenced by inhomogeneities in the bg, whereas break size estimation with RICH-obs is independent of the bg. The adjustment quality of RICH-obs, on the other hand, may suffer from large interpolation errors at remote stations. RICH- $\tau$  is a compromise that substantially reduces interpolation errors at the cost of slight dependence on the bg.

Adjustment uncertainty is estimated by comparing the three methods and also by varying parameters in RICH. The adjusted radiosonde time series are compared with recent temperature datasets based on (Advanced) Microwave Sounding Unit [(A)MSU] radiances. The overall spatiotemporal consistency of the homogenized dataset has improved compared to earlier versions, particularly in the presatellite era. Vertical profiles of temperature trends are more consistent with satellite data as well.

### 1. Introduction

The radiosonde network is a central part of the global upper air observing system. Since many stations have operated since the late 1950s or longer, radiosonde temperature records have been extensively used also for climate studies. It became clear that these records need to be homogenized before their trends and low-frequency variability can be interpreted (Parker et al. 1997). Various homogenization approaches for the global radiosonde network have been put forward (Luers and Eskridge 1995; Lanzante et al. 2003; Thorne et al. 2005b; Free

et al. 2005; Sherwood et al. 2008; McCarthy et al. 2008; Titchner et al. 2009) but none of them could explain and remove the apparent pervasive cooling bias in the radiosonde temperature data compared to satellite data. In particular, the vertical trend profiles in the tropics did not show the enhanced upper tropospheric amplification as predicted by climate models, with the exception of the ensemble described in Thorne et al. (2011a) that included this possibility in its uncertainty bounds. Only temperature trend assessments based on changes in the zonal mean thermal wind structure support enhanced upper tropospheric warming in the tropics (Allen and Sherwood 2008).

Discrepancies between layer mean atmospheric temperatures derived from radiances of the Microwave Sounding Unit (MSU) and MSU-equivalent temperatures calculated from radiosonde data can be attributed at least partly to inhomogeneities in the raw radiosonde data. Any remaining inconsistency would be caused by inhomogeneities and uncertainties in the MSU brightness temperatures as described, for example, by Thorne

 Denotes Open Access content.

*Corresponding author address:* Leopold Haimberger, Department of Meteorology and Geophysics, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria.  
E-mail: leopold.haimberger@univie.ac.at

DOI: 10.1175/JCLI-D-11-00668.1

© 2012 American Meteorological Society

et al. (2011b) or Mears et al. (2011). Temperature trends from raw radiosonde data are also inconsistent with climate models, which project an upper tropospheric warming maximum, especially in the tropics (Santer et al. 2005; Trenberth et al. 2007; Santer et al. 2008).

Haimberger (2007) introduced a new homogenization method [Radiosonde Observation Correction Using Reanalyses (RAOBCORE)] that analyzed not only 0000–1200 UTC difference time series and station history information but also time series of background departures from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; Uppala et al. 2005). These background departures, also referred to as innovations, are the differences between observations  $\mathbf{y}$  and the state vector  $\mathbf{x}_b$  of the background forecast by an assimilating model. Here  $\mathbf{x}_b$  is mapped to the observation location by the observation operator  $H$ . Following the notation in data assimilation literature (see, e.g., Courtier et al. 1998; Lewis et al. 2005) we write  $\tau = \mathbf{y} - H\mathbf{x}_b$ . The background state  $H\mathbf{x}_b$  is considered independent of the radiosonde observations  $\mathbf{y}$ , which is generally a good assumption except that persistent common biases at several neighboring radiosonde stations may have a noticeable effect on the background state (Haimberger 2007).

Analysis of daily innovation time series proved highly efficient for break detection, but break size estimation from reanalysis innovations is complex because of a few inhomogeneities of the ERA-40 background forecast and analysis time series (Uppala et al. 2005; Haimberger 2007; Grant et al. 2008; Screen and Simmonds 2011). RAOBCORE, which uses innovation series for both break detection and innovation, has also been criticized for being not independent of satellite data and of the assumptions made in the assimilating model. This aspect certainly limits the value of radiosonde data homogenized by RAOBCORE for comparison with satellite datasets. Nevertheless, RAOBCORE adjustments have been used as radiosonde bias correction in the interim ECMWF Re-Analysis (ERA-Interim; Dee et al. 2011b; Haimberger and Andrae 2011), in the Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011), and in the Japanese 55-yr reanalysis (Ebita et al. 2011).

To overcome the dependence problem, Haimberger et al. (2008) developed a method called Radiosonde Innovation Composite Homogenization (RICH), which uses the breakpoint date information from RAOBCORE but calculates the break size estimates by comparison with neighboring radiosonde temperature records. Since these reference records are independent of satellite data, satellite data can affect RICH estimates only through the breakpoint dates provided by RAOBCORE. RICH

worked quite well for homogenizing the time series of the satellite period 1979 onward. Also it revealed an upper tropospheric warming maximum in the much debated period 1979–99 (Santer et al. 2008).

The present paper is motivated by the fact that Haimberger et al. (2008) provided only rudimentary documentation of RICH, by the availability of very much enhanced background forecasts due to the advent of ERA-Interim (Dee and Uppala 2009; Dee et al. 2011a,b), and by substantial extensions and improvements made to the RICH algorithm itself. Another motivation was the desire to quantify the uncertainties in the homogenization approach through the use of ensemble techniques. This has been pioneered by McCarthy et al. (2008) and Thorne et al. (2011a) for radiosonde temperatures. Ensembles of MSU brightness temperatures have become available recently as well (Mears et al. 2011) and are also employed for assessing uncertainties in SST data (Kennedy et al. 2011a,b). They allow for much better assessment of whether differences between various datasets are statistically significant.

The next section describes the RICH method and its parameters in some detail. Section 3 describes input data for RICH as well as the datasets used for comparison and validation. Section 4 explains the methodology used to quantify parametric and to a certain extent also structural adjustment uncertainties. Section 5 shows results for selected stations and large-scale means. Implications of the results for a general climate data improvement strategy are discussed in the conclusions.

## 2. Description of the RICH adjustment method

The basic idea of RICH—homogenization of tested time series through comparison with neighboring reference series of the same observation type—is not new. A large variety of upper air data and surface data homogenization methods work with this idea (see, e.g., Thorne et al. 2005b; Sherwood et al. 2005; Venema et al. 2012).

The main novelty of RICH is that it tries to make optimal use of the output from a dynamical climate data assimilation system for break detection and adjustments. For this purpose it uses the break detection part of the RAOBCORE algorithm, which analyzes 0000 and 1200 UTC daily time series of background departures  $\tau = \mathbf{y} - H\mathbf{x}_b$  from reanalyses.

For radiosonde temperature measurements  $H$  is specified as a simple interpolation from the ECMWF model grid to the observation location. The innovations are an important standard diagnostic for time series models or data assimilation systems. Haimberger (2007) demonstrated that statistical analysis of daily innovation time series with homogeneity tests is quite efficient in

finding breaks in these time series because of their small variance. The RAOBCORE algorithm, which does this analysis, has not changed appreciably since publication in 2007. However, the background forecast data used as reference have improved (see section 3).

RAOBCORE yields the date of potential breaks in the radiosonde time series. These dates are input for the Radiosonde Innovation Composite Homogenization algorithm, where the word “innovation” in the acronym should remind the user that the used break dates have been calculated by analyzing innovation time series.

*a. Definition of averaging operators and break size estimates*

For break estimation, RICH compares either neighboring observation (obs) time series (RICH-obs) or neighboring  $\tau$  time series (RICH- $\tau$ ). The obs time series are independent of satellite or other nonradiosonde observing systems. RICH-obs assumes that a sufficiently long homogeneous reference time series can be constructed from neighboring radiosonde temperature time series for each break date in a tested station time series so that the size of the shift can be accurately estimated.

While the essence of break size estimation is rather simple (comparing means), we formally define the estimates to make the details of the break size estimation process transparent. In this paper we denote differences between tested and reference stations as follows:

$${}^i\nabla_x^{jk} = {}^i x^j - {}^i x^k, \quad (1)$$

where  ${}^i x^j$  denotes either obs or bg temperature values or their difference  $\tau = \text{obs} - \text{bg}$  at station  $j$ , and  $i$  is the launch index that stands for the date and time when the  $x$  values were calculated or measured. Also,  ${}^i\nabla_x^{jk}$  denotes the difference of  $x$  values between stations  $j$  and  $k$  at the same observation time  $i$ . Since the stations  $j$  and  $k$  are spatially separated, it seems suitable to use the gradient symbol for this difference.

In general we do not want to compare individual values but averages over a time period. We define the time average  $\bar{x}^j(a)$ :

$$\frac{1}{n} \sum_{i=1}^n {}^i x^j = \bar{x}^j(a). \quad (2)$$

The choice of the interval  $a$  depends on the dates of other breaks in the tested or the reference series as well as on data availability. Its length is 0.5–8 yr. The sample size  $n$  is the number of launches (130–2920) at a specific time of day (0000 or 1200 UTC);  $n$  depends on the length of  $a$  and data availability at station  $j$  in this interval. While smaller sample sizes are possible they have not been found advantageous. Details for the choice of the

intervals used for break size estimation are given in section 2b below. Note also that  $\bar{x}^j(a)$  is seasonally invariant, as are all means below. Therefore seasonal variations of biases, especially in polar regions, are not taken into account. Andrae et al. (2004) and Haimberger and Andrae (2011) discuss methods to estimate the radiation error as a function of solar elevation. They can complement the seasonally invariant adjustments calculated in this paper but this is not pursued further here.

1) MEAN DIFFERENCES OF OBSERVATIONS AND BACKGROUND

The observation difference is the mean temperature difference between observations from two neighboring radiosonde stations and is defined as

$$\bar{\nabla}_{\text{obs}}^{jk}(a) = \bar{\text{obs}}^j(a) - \bar{\text{obs}}^k(a), \quad (3)$$

where obs are the measured temperatures at stations  $j$  and  $k$  in the interval  $a$ , which contains  $n$  pairs of observations;  $\bar{\text{obs}}^j(a)$  and  $\bar{\text{obs}}^k(a)$  are mean values estimated at stations  $j$  and  $k$  in interval  $a$ . It is important that only those data are counted where values at both stations  $j$  and  $k$  are available. This reduces the risk of unrealistic differences due to unequal sampling at stations  $j$  and  $k$ . The same quantity can be defined for the bg:

$$\bar{\nabla}_{\text{bg}}^{jk}(a) = \bar{\text{bg}}^j(a) - \bar{\text{bg}}^k(a). \quad (4)$$

Let us now consider the situation where an interval  $a$  with  $n$  numbers of observation pairs and an interval  $b$  with  $m$  numbers of observation pairs are separated by a break in the time series of station  $k$ . Then we expect that the observation difference to station  $j$  in the two intervals is different:  $\bar{\nabla}_{\text{obs}}^{jk}(a) \neq \bar{\nabla}_{\text{obs}}^{jk}(b)$ . The size of the change between intervals  $a$  and  $b$  can be written as

$$\begin{aligned} \Delta_{\text{obs}}^{jk}(a, b) &= [\bar{\text{obs}}^j(b) - \bar{\text{obs}}^k(b)] - [\bar{\text{obs}}^j(a) - \bar{\text{obs}}^k(a)] \\ &= \bar{\nabla}_{\text{obs}}^{jk}(b) - \bar{\nabla}_{\text{obs}}^{jk}(a). \end{aligned} \quad (5)$$

If the stations  $j$  and  $k$  are close to each other, and if the reference time series at station  $j$  is homogeneous,  $\Delta_{\text{obs}}^{jk}(a, b)$  already represents an unbiased estimate for the size of the break at station  $k$  occurring between intervals  $a$  and  $b$ . As explained in section 2b below, RICH-obs uses weighted means of  $\Delta_{\text{obs}}^{jk}$  from several neighboring stations for estimating the break sizes at station  $k$ . The black profiles in the left panels of Fig. 1 are  $\Delta_{\text{obs}}^{jk}(a, b)$  estimates for different reference stations  $j$  and different pressure levels.

The situation becomes more complicated, however, if the distance between stations is large. In this case the true

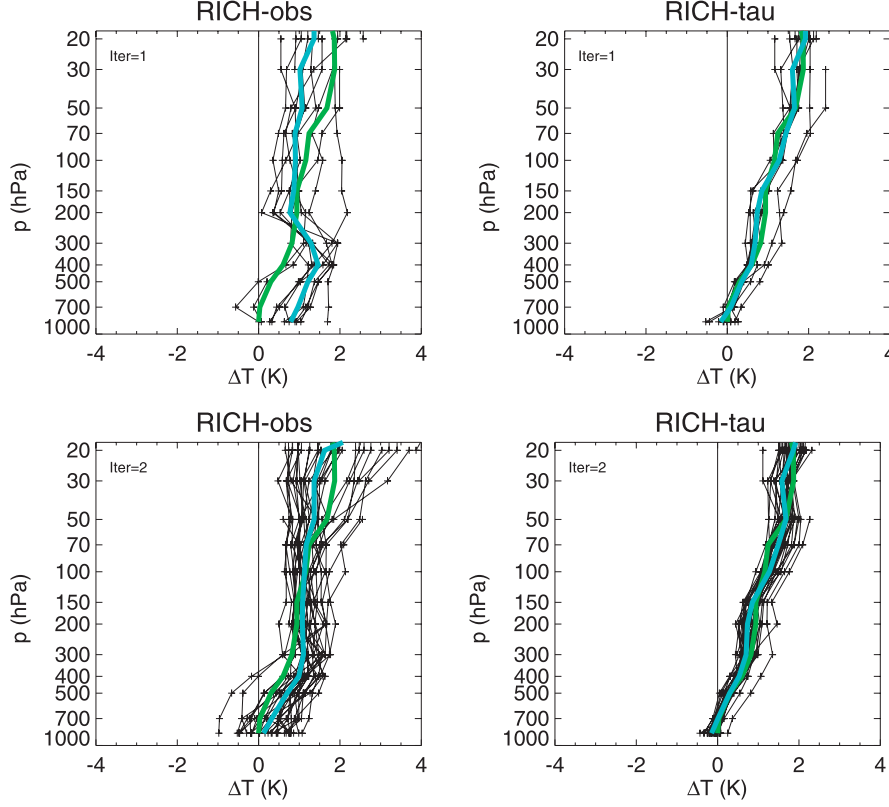


FIG. 1. RICH break size estimation: black curves are individual (left)  $\Delta_{\text{obs}}^{jk}$  and (right)  $\Delta_{\tau}^{jk}$  profiles for a break in the radiosonde record of Bethel (Alaska, 70219) in July 1989. Thick green profile (identical in all four panels) is the RAOBCORE estimate, thick blue profiles are (left) RICH-obs and (right) RICH- $\tau$  estimates, respectively. (top) Profiles from first iteration of RICH (see Fig. 3; 10 neighbors used); (bottom) profiles from second iteration with 30 neighbors.

mean temperatures at stations  $j$  and  $k$  may evolve differently due to regional climate anomalies. In such situations  $\Delta_{\text{obs}}^{jk}(a, b)$  could be different from zero even if both records at stations  $j$  and  $k$  were free of artificial jumps. Interpreting  $\Delta_{\text{obs}}^{jk}(a, b)$  as break size estimate would lead to false adjustments in this case and must be avoided.

The problem can be circumvented if the short-term climate anomalies can be realistically represented by an independent dataset. ERA-40/ERA-Interim background fields (bg) are generally, with known exceptions in polar regions (e.g., Grant et al. 2008), of sufficient quality that they can represent regional climate anomalies well. If this is true, the difference  $\Delta_{\text{bg}}^{jk}$ , defined as

$$\begin{aligned} \Delta_{\text{bg}}^{jk}(a, b) &= [\overline{\text{bg}}^j(b) - \overline{\text{bg}}^k(b)] - [\overline{\text{bg}}^j(a) - \overline{\text{bg}}^k(a)] \\ &= \overline{\nabla_{\text{bg}}^{jk}}(b) - \overline{\nabla_{\text{bg}}^{jk}}(a), \end{aligned} \quad (6)$$

describes a climate change in the temperature gradient between stations  $j$  and  $k$ , which needs to be subtracted from the break size estimate  $\Delta_{\text{obs}}^{jk}(a, b)$  gained from comparison of the observations.

## 2) INNOVATION DIFFERENCES BETWEEN INTERVALS AND STATIONS

If one considers the bg temperature gradient to be a true estimate of the state of the atmosphere, the innovations, or obs – bg differences must contain information about the systematic difference of the radiosondes that were used. For a single station ( $k$ ) and one interval ( $a$ ) the obs – bg difference is defined as

$$\tau^k(a) = \text{obs}^k(a) - \text{bg}^k(a). \quad (7)$$

If the background provides a spatially and temporally homogeneous field, a change in the systematic bias of radiosonde data, caused by a change of observation systems, must change  $\tau^k(a)$ . This value can be estimated as

$$\begin{aligned}\Delta_\tau^k(a, b) &= [\overline{\text{obs}}^k(a) - \overline{\text{bg}}^k(a)] - [\overline{\text{obs}}^k(b) - \overline{\text{bg}}^k(b)] \\ &= \overline{\tau}^k(a) - \overline{\tau}^k(b).\end{aligned}\quad (8)$$

This estimate is used to calculate the adjustments performed by RAOBCORE (Haimberger 2007). As one can see, no reference station  $j$  is necessary to get a break size estimate in this case. However, evaluation of Eq. (8) puts high demands on the homogeneity of the bg. The green curve in all panels of Fig. 1 is the profile of  $\Delta_\tau^k(a, b)$  for a specific break in the time series of Bethel, Alaska.

The RICH- $\tau$  method makes different use of the bg. It requires only that the gradient of the bg between stations  $j$  and  $k$  be realistic. The absolute value of the bg does not need to be unbiased or at least homogeneous, as is required for RAOBCORE. It uses the information provided by both the background and the time series of a neighboring station for break size estimation. An innovation difference  $\nabla_\tau^{jk}(a)$  is defined as

$$\begin{aligned}\nabla_\tau^{jk}(a) &= [\overline{\text{obs}}^j(a) - \overline{\text{bg}}^j(a)] - [\overline{\text{obs}}^k(a) - \overline{\text{bg}}^k(a)] \\ &= \overline{\tau}^j(a) - \overline{\tau}^k(a).\end{aligned}\quad (9)$$

It thus combines observations and background information of two neighboring radiosondes, and is a modification of  $\nabla_{\text{obs}}^{jk}$ . If only station  $k$  changes its upper air observation system, and the background provides a homogeneous field, the difference of  $\nabla_\tau^{jk}$  before and after the artificial break must contain information about the systematic error in the observations:

$$\begin{aligned}\Delta_\tau^{jk}(a, b) &= [\overline{\text{obs}}^j(b) - \overline{\text{bg}}^j(b)] - [\overline{\text{obs}}^k(b) - \overline{\text{bg}}^k(b)] \\ &\quad - [\overline{\text{obs}}^j(a) - \overline{\text{bg}}^j(a)] + [\overline{\text{obs}}^k(a) - \overline{\text{bg}}^k(a)] \\ &= \overline{\nabla}_\tau^{jk}(b) - \overline{\nabla}_\tau^{jk}(a).\end{aligned}\quad (10)$$

In contrast to  $\Delta_{\text{obs}}^{jk}(a, b)$ , this difference takes a possible regional climate anomaly into account, provided the bg gradients are realistic. Using definitions (5) and (6),  $\Delta_\tau^{jk}(a, b)$  can also be written as

$$\Delta_\tau^{jk}(a, b) = \Delta_{\text{obs}}^{jk}(a, b) - \Delta_{\text{bg}}^{jk}(a, b).\quad (11)$$

RICH- $\tau$  uses weighted means of  $\Delta_\tau^{jk}$  from several neighboring stations to estimate the break sizes at

station  $k$ . The black profiles in the right panels of Fig. 1 are  $\Delta_\tau^{jk}(a, b)$  estimates for different reference stations  $j$  and different pressure levels.

### 3) INTERPRETATION

If the bg is correct at all times and places and the reference radiosonde station time series is homogeneous throughout the combined interval  $[a, b]$ , the break size is given as the value of  $\Delta_\tau^{jk}$ , which is related to  $\Delta_{\text{obs}}^{jk}$  and  $\Delta_{\text{bg}}^{jk}$  as

$$\Delta_\tau^{jk}(a, b) = \Delta_{\text{obs}}^{jk}(a, b) - \Delta_{\text{bg}}^{jk}(a, b) = \Delta_\tau^k(a, b) - \Delta_\tau^j(a, b).\quad (12)$$

In the ideal case, if both bg and neighboring radiosonde record  $j$  are accurate and homogeneous,  $\Delta_\tau^k$  should be equal to  $\Delta_\tau^{jk}$ . For an ideal reference radiosonde station  $\Delta_\tau^j$  is zero, if the bg is free of artificial shifts, and an optimal bg would have zero  $\tau^j$  in case of true observations.

If  $\Delta_\tau^k$  and  $\Delta_\tau^{jk}$  are different, however, it is difficult to tell where the discrepancies are coming from. If the background is biased and the reference sonde  $j$  is homogeneous, one can speculate that even a biased background would deliver a realistic  $\nabla_{\text{bg}}^{jk}$ , making  $\Delta_\tau^{jk}$  a better estimate for the systematic bias between radiosondes than  $\Delta_\tau^k$ . If the reference record is inhomogeneous, it may be better to use  $\Delta_\tau^k$  for break estimation since this difference is not affected by inhomogeneities in record  $j$ .

It is instructive to compare the standard errors of  $\Delta_\tau^k$ ,  $\Delta_\tau^{jk}$ , and  $\Delta_{\text{obs}}^{jk}$ . These are calculated for each of these  $\Delta_m$  as

$$\sigma_m = \sqrt{\frac{1}{n_a + n_b} [n_a \sigma_x(a)^2 + n_b \sigma_x(b)^2]},\quad (13)$$

where  $\sigma_x(a)$  is the standard deviation of

$$\begin{aligned}\overline{\tau}^k(a) &\quad \text{for } m = \text{RAOBCORE}, \\ \overline{\text{obs}}^k(a) - \overline{\text{obs}}^j(a) &\quad \text{for } m = \text{RICH-obs}, \\ \overline{\tau}^k(a) - \overline{\tau}^j(a) &\quad \text{for } m = \text{RICH-}\tau.\end{aligned}$$

The same definition applies to  $\sigma_x(b)$ . Note that the size of a possible break between intervals  $a, b$  has no effect on the combined standard deviation  $\sigma_m$ . As is shown in Fig. 2,  $\sigma_m$  is smaller for RAOBCORE than for RICH- $\tau$  and much smaller than for RICH-obs. The main reason is the smallness of the individual background departures ( $\tau^k$ ) compared to background departure differences ( $\tau^k - \tau^j$ ) and to temperature differences ( $\text{obs}^k - \text{obs}^j$ ). A second reason is the smaller sample sizes in RICH due to missing data at individual reference stations.

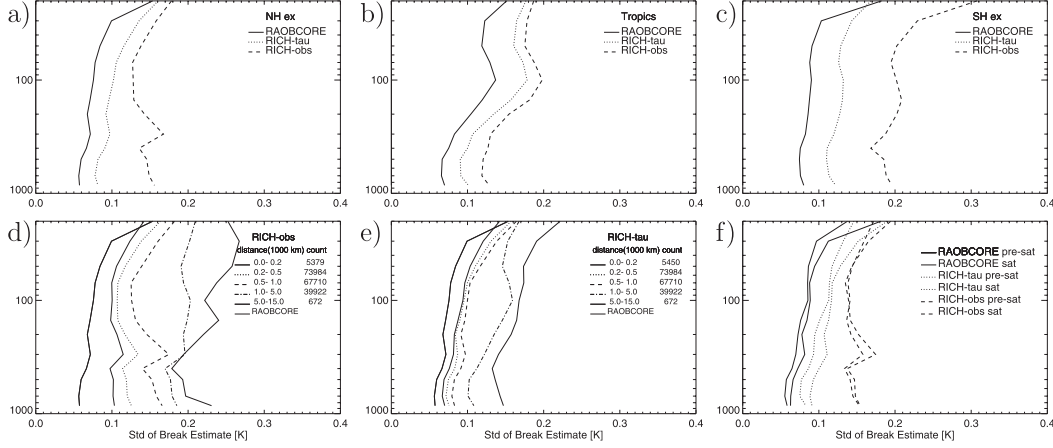


FIG. 2. Log-pressure profiles of sample error standard deviation of break size estimates as defined in Eq. (13) for RAOBCORE (solid), RICH-obs (dashed), and RICH- $\tau$  (dotted). First row shows dependency on region: (a) 20°–90°N, (b) 20°S–20°N, and (c) 90°–20°S. Second row shows dependency on (d), (e) distance between stations and epoch [(f) presatellite (1958–78) and satellite (1979–present) periods].

Figure 2 shows also to what extent standard errors are smaller in the NH extratropics than in the tropics and in the SH extratropics. For RICH there is a strong dependence of standard errors on the distance to the neighboring station. There is relatively weak dependence on the period considered. Estimates in the satellite era are only slightly more accurate than in the presatellite era. The vertical profile of the standard errors is interesting in that it does not increase monotonically with height. Especially for RICH-obs the tropopause region is the most challenging, since at those levels near jet streams it can happen that data in the troposphere at one station are compared with data at a neighboring station at the same pressure level that is already in the stratosphere.

It should be noted that only  $\Delta_r^k$  immediately yields the break size, whereas in RICH values of  $\Delta_r^{jk}$  or  $\Delta_{\text{obs}}^{jk}$  have to be averaged over several stations  $j$  to yield the break size estimate. This reduces the standard error to a certain extent, depending on the distribution of neighboring stations, but not much since the  $i_{\text{obs}}^k - i_{\text{obs}}^j$  differences are not independent of, for example, the  $i_{\text{obs}}^k - i_{\text{obs}}^{j+1}$  differences.

#### b. Sampling strategy for constructing the reference series

The expressions above describe break estimates from comparison of a test station  $k$  with a single reference station  $j$ . To reduce noise it is necessary to consult several reference stations and to build a weighted average over the break estimates. For RICH-obs, the break estimate at a known breakpoint date  $t$  at pressure level  $p$  is

$$\hat{\Delta}_{\text{obs}}^k(p, t) = \frac{1}{\left(\sum_{j=1}^J w^{jk}\right)} \sum_{j=1}^J \Delta_{\text{obs}}^{jk}(a^{jk}, b^{jk}, p, t) w^{jk}. \quad (14)$$

The hat symbol denotes averaging over the break size estimates from comparison with a sample of neighboring stations  $j$ . Note that the intervals  $a^{jk}, b^{jk}$  are dependent on the tested station  $k$  and on reference station  $j$ . This dependency exists also in the formulas above but has not been made explicit to keep the notation simple. The location of breakpoints at reference station  $j$  as well as at the tested station  $k$  near the breakpoint to be adjusted determines the length of intervals  $a^{jk}, b^{jk}$ . The break estimates from the individual comparisons with reference series are then averaged using weights  $w^{jk}$ . For RICH- $\tau$ , we use the same averaging procedure.

While several weighting procedures are possible, we chose weights decreasing exponentially with distance:

$$w^{jk} = \exp(-d^{jk}/1500 \text{ km}). \quad (15)$$

The parameter  $d$  is not the usual spherical distance but has been chosen as

$$d^{jk} = r_E \Delta\phi^{jk} + r_E (0.1 \Delta\lambda^{jk}), \quad (16)$$

where  $r_E$  is Earth's radius and  $\Delta\phi^{jk}, \Delta\lambda^{jk}$  are latitude and longitude differences between stations  $j$  and  $k$ . Note that the meridional distance is weighted much higher than the east–west distance, especially at low latitudes. This choice reflects the fact that climate zones depend mostly



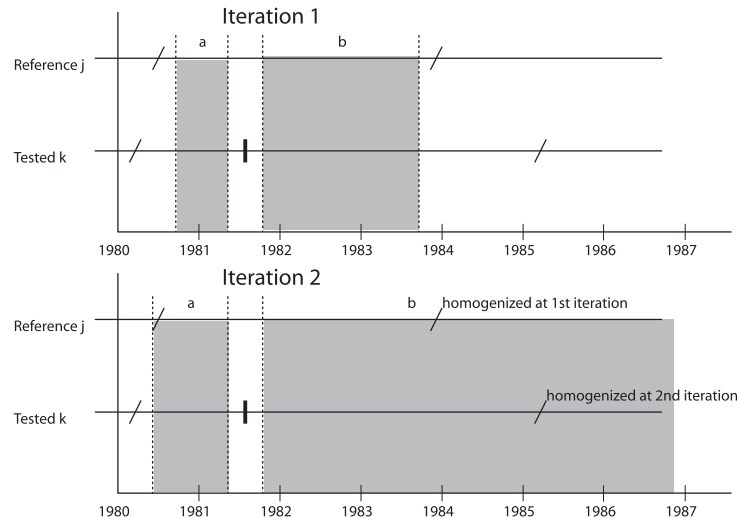


FIG. 3. Illustration of data selection when adjusting break (thick vertical line) at tested series. Slashes are other breakpoints at tested station and at reference station. In first iteration, time series must be considered inhomogeneous at all breakpoints. Gray areas indicate intervals used for break estimation. Data near breakpoints are not used to account for uncertainty in break date detection. In the second iteration, the reference series from the first iteration are considered homogeneous. Also the tested series is considered homogeneous after the breakpoint just tested, since the adjustment procedure works backward, starting from the most recent breakpoint. Thus the interval  $b$  is often much larger in second iteration. Interval  $a$  is bounded only by earlier breakpoints in the tested series but not in the reference series and thus may also be longer.

on latitude, at least in zeroth order (Kottek et al. 2006; Rougier 2007) and is consistent with findings of Wallis (1998) and McCarthy (2008). It also reduces the risk to make composites of stations where the tropopause height is different. The parameter  $d$  is used not only for weighting the composites, but also for choosing the stations used in the composites. Neighboring stations are sorted according to  $d^{jk}$ . Only the nearest stations that have enough data for break size estimation at station  $j$  are used for the composite. The number of neighbors varies between 3 and 30 (see section 4b below). Depending on the region, the most distant neighbor considered is between 300 km (over densely populated areas) and 10 000 km (in the tropical Pacific or the Southern Ocean) away.

Other choices for  $w^{jk}$  such as linear correlation of temperature series from reanalyses are also possible (Thorne et al. 2005b). However, correlation is rather height dependent. The linear temperature correlation patterns in the lower troposphere look very different to the correlation patterns at the tropopause level. We chose to avoid different weights at different levels.

The validity of the break size estimates relies on the assumption that the difference time series for estimating

$\Delta_{\text{obs}}^{jk}$  or  $\Delta_{\tau}^{jk}$  are temporally homogeneous before and after a breakpoint at the test station. In other words, it is assumed that each difference in the series belongs to the same statistical population. To guarantee this, it is essential to know the dates of breakpoints of the tested stations and of the reference stations and to choose  $a^{jk}$  and  $b^{jk}$  accordingly. Only then averaging over inhomogeneities in the records can be avoided. Figure 3 shows how the averaging intervals  $a$  and  $b$  are influenced by the dates of neighboring breakpoints at the tested station and the reference station.

In the case of RICH, the dates of breakpoints have been determined with RAOBCORE. As described by Haimberger (2007), RAOBCORE analyzes  $i_{\tau}^k$  (innovation) time series using a version of the standard normal homogeneity test (Alexandersson and Moberg 1997) that has been adapted to work well with difference series that have an annual cycle and data gaps, a common case for radiosonde temperatures especially at high levels. About 8000 breaks could be detected between 1958 and 2011 (an average of about seven breaks per station). With more liberal parameter settings, even more breaks could be identified, but too many breakpoints

restrict the length of the time series available for break size estimation and could lead to unnecessarily high noise levels in the break size estimates.

### c. Adjustment procedure

The break sizes at station  $k$  are estimated, beginning with the most recent break. The record is then adjusted at this breakpoint. Then the next earliest break is adjusted, working backward in time. This procedure is performed for every station  $k$  on the globe with at least two years of data.

It has proven useful to perform the adjustment of the global radiosonde temperature dataset in two steps. In the first iteration, averaging over known breakpoints has been strictly avoided (see Fig. 3a). This idea is consistent with the concept of pairwise intercomparison together with using only homogeneous subperiods for estimating break sizes (Della-Marta and Wanner 2006; Sperka 2007; Caussinus and Mestre 2004; Menne and Williams 2009). Also, the number of reference series has been limited to a very small number (3 or 10) to reduce the risk of using of neighboring stations that are far away. However, the spatial noise in trend estimates after the first iteration is relatively large because of the small number of reference stations and the sometimes short time intervals permitted for break size estimation.

When the first homogenization has been applied at all radiosonde stations, one could finish. It proved highly advantageous, however, to perform a second adjustment step where the tested series are compared with neighboring series that have been homogenized in the first iteration. In the second iteration, it is assumed that the pervasive biases in the reference stations have been removed in the first iteration. If this is the case, it is allowed to average over breakpoint dates in the reference stations (see Fig. 3b). If it is known that this is not the case, because the adjustment failed in the first iteration, the RICH algorithm avoids averaging over those breakpoint dates in the reference stations. The main reason for failure in the first iteration at some pressure level is lack of reference data. In many cases only the uppermost pressure levels are affected. The default averaging interval of 8 years in the second iteration is reduced only by breaks in the test station occurring earlier than the break just estimated, and by data gaps. Since the conditions on break dates are less strict in the second iteration, more stations can be used as reference stations without the need to interpolate over large distances. Thus the number of stations used for break size estimation is set 3 times higher in the second iteration.

The two-step approach minimizes the risk of averaging over inhomogeneous samples. Nevertheless, the following additional measures have been taken to avoid

averaging over breakpoints and thus noisy break estimates or even divergence of the iterative procedure:

- A neighboring station is used as reference only if it has no breakpoint 180 days before and 180 days after the breakpoint diagnosed by RAOBCORE at the tested station.
- At least 130–330 good values must be available for comparison before/after the breakpoint and the next breakpoint at the tested or the reference station.
- At least 30 days of data next to the breakpoints at the tested station and at the reference station are discarded to avoid inhomogeneities due to inaccurately detected breakpoints. If the averaging interval is long enough, 180 days of data are discarded. If the sample size is small (i.e., close to the minimum number specified above), the number of discarded data is reduced to 120, then 60, then 30.
- Care is taken to avoid unequal sampling of the annual cycle in the intervals before/after the breakpoint of the tested station, as described in Haimberger (2007). If, for example, no January values are available in interval  $a^{jk}$ , January values in interval  $b^{jk}$  are deleted.
- When calculating the break size estimate  $\hat{\Delta}_{\text{obs}}^k(p, t)$  from the sample of neighboring stations, the maximum and minimum break size estimate are discarded if the sample size is larger than three. This trimming of the mean leads to more robust break size estimates. We refrained from using the even more robust median or the interquartile range average, however, since we found that spatiotemporal consistency is better when using the mean or only slightly trimmed means (not shown).

All these measures are designed to ensure that the members of the compared samples in intervals  $a^{jk}$  and  $b^{jk}$  belong to the same populations so that the estimated means are meaningful. It may seem trivial to mention this, but we think that clean separation between homogeneous and inhomogeneous parts of the analyzed records is the key to successful homogenization. The good performance of the pairwise intercomparison methods of Caussinus and Mestre (2004) and Menne and Williams (2009) in a recent intercomparison of surface data homogenization methods (Venema et al. 2012) supports this.

It should also be noted that the above measures are only successful if the vast majority of breaks have been detected by RAOBCORE. Undetected breaks still may contaminate the break size estimates. However, the settings of RAOBCORE have been quite liberal and the sensitivity experiments below indicate that there are only few undetected breaks.

The consequent avoidance of averaging over inhomogeneities is most likely a key advantage compared to

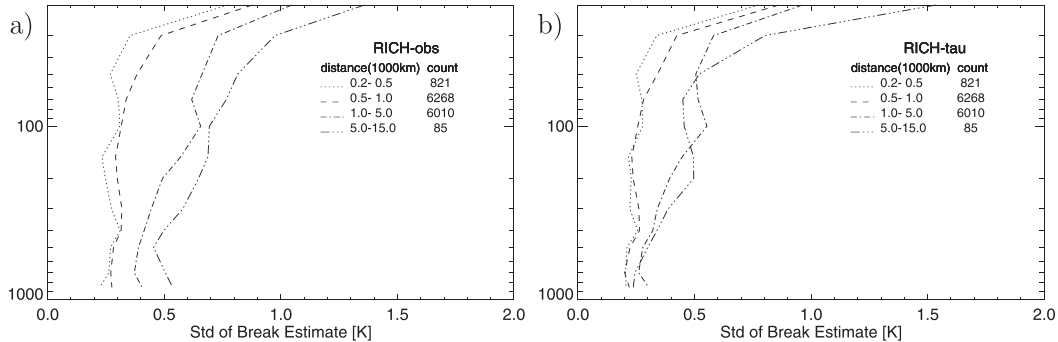


FIG. 4. Log-pressure profiles of rms difference between break size estimates from RAOBCORE and (a) RICH-obs and (b) RICH- $\tau$  for different maximum distances between stations.

Thorne et al. (2005b) and McCarthy et al. (2008) and to Sherwood et al. (2008), who have also tried to avoid it but used monthly data.

The use of daily data leaves much more choices in data selection and allows us to cope with breaks separated by only a few months. It also helps to avoid unjustified adjustment of sampling biases that may occur, for example, if balloons burst earlier if the stratosphere is colder. While daily data could be used to make more sophisticated adjustments that modify the probability distribution functions of the observations (e.g., Dai et al. 2011; Della-Marta and Wanner 2006), we restricted ourselves to modifying the mean, since this is challenging enough, as is clearly shown in Figs. 2 and 4.

#### ADJUSTMENT OF THE CLIMATOLOGY

As already discussed by Haimberger (2007), the availability of background departure time series allows efficient adjustment of the mean of short time series that cannot be analyzed for breakpoints and of time series that are known to have sizeable biases even in their most recent part. The basic assumption is that the mean background departure of the most recent part (at most 8 years, with a minimum 1/2 year for very short records) should be equal to the background departures of neighboring stations. The mean difference between the background departures  $\nabla_{\tau}^{jk}(a)$  of stations  $j, k$  has already been defined in Eq. (9). The time interval  $a$  is now the most recent part of the series.

The adjustment of the most recent part is now calculated as

$$\hat{\Delta}_{mr}^k(p, t) = \frac{1}{\left(\sum_{j=1}^J w^{jk}\right)} \sum_{j=1}^J \nabla_{\tau}^{jk}(a^{jk}, p, t) w^{jk}. \quad (17)$$

This is formally quite similar to the break size calculation used in RICH- $\tau$ . The weights have the same

meaning as above and of course only one time interval must be considered per intercomparison. Since the adjustment of the climatology is performed after the homogeneity adjustments, the involved time series are considered homogeneous.

While the adjustment of the climatology appears formally simpler, it works only if the neighboring stations are unbiased. This strong assumption is justifiable only for a few of the most recent radiosonde types in use. We chose stations using Vaisala RS90, RS92, Meisei, and current Sippican radiosondes (type codes 37, 52, 53, 60, 61, 62, 63, 66, 67, 71, 72, 73, 74, 78, 79, 80, 81, 82, 83, 47, 55, 56, 26, 76, 85, 86, and 87 according to the World Meteorological Organization (WMO) Binary Universal Form for data Representation (BUFR) common code table C-2) as reference types stations (about 400). These are also fairly well distributed over Earth.

The adjustment of the climatologies does not affect trend analysis. However, it allows to bias-correct even very short records that are normally not considered for climatological purposes but are certainly useful for re-analysis efforts, and it reduces the rejection rate of records during assimilation.

#### 3. Input data and comparison strategy

The following input data have been used for this intercomparison:

- Daily (0000 and 1200 UTC) radiosonde data on 16 standard pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa) from the Integrated Global Radiosonde Archive (IGRA; Durre et al. 2006) and the ERA-40/ERA-Interim radiosonde archives from 1958 onward. Metadata information is mainly based on the IGRA metadata file, whose origins go back to Gaffen (1996) and which

is constantly updated (<http://www1.ncdc.noaa.gov/pub/data/igra/igra-metadata.txt>). This valuable source of information reports 8715 events regarding the radiosonde type alone. Additional metadata information is retrieved from ERA-40/ERA-Interim radiosonde records, which often include the radiosonde type as transmitted by the Global Telecommunication System from about 1990 onward.

- Background forecasts from the ERA-40/ERA-Interim reanalyses interpolated to the station locations. The properties of the forecasts are quite different between ERA-40 (Uppala et al. 2005; used for the period 1958–78) and ERA-Interim (Dee et al. 2011a,b; for the period 1979–2010). Most notably, ERA-40 uses 6-hourly cycling and a three-dimensional variational data assimilation (3D-VAR) system, whereas ERA-Interim uses a considerably more advanced 4D-VAR assimilation system with a better forecast model, 12-h cycling, and variational bias correction of satellite data. No bias correction has been applied to radiosonde temperatures in ERA-40 during the period 1958–79. In ERA-Interim, the radiosondes have been adjusted to remove both the annual mean bias, using RAOBCORE v1.3 adjustments, and the annual cycle of the radiation error. For a more detailed description, see Haimberger and Andrae (2011).

To avoid a shift in the reference at the change point from ERA-40 to ERA-Interim in 1979, zonal mean background departures between reanalysis and radiosonde observations have been calculated for the years 1978 and 1980. The difference in these zonal mean departures has been taken as estimate for the shift between ERA-40 and ERA-Interim temperatures. 1979 has been avoided since it had atypical data coverage due to the Global Weather Experiment in 1979. The shift has been subtracted from the ERA-40 background departures. While this procedure may reduce the break detection power of RAOBCORE during these years, it has been found to yield the smoothest transition from ERA-40 to ERA-Interim. Alternative approaches, such as using twentieth-century reanalysis (Compo et al. 2011) data or using regional instead of zonal means, left some shifts in the time series of zonal mean background departures after the adjustment and also led to less spatiotemporal consistency of trend estimates in time intervals that include the year 1979. Further investigations of the uncertainties introduced through this transition will nevertheless be necessary in the future.

Solely for intercomparison, we use the following datasets:

- Remote Sensing Systems (RSS) MSU v3.3 brightness temperatures (Mears and Wentz 2009) on the standard  $2.5^\circ \times 2.5^\circ$  lat/lon grid at lower stratospheric (LS) and upper tropospheric (TS) and midtropospheric (MT) layers. An ensemble of realizations of this dataset using a Monte Carlo technique is available as well (Mears et al. 2011). From this ensemble we use the mean and the 5% and 95% percentiles, where the ensemble members have been sorted after their global mean trend.
- University of Alabama at Huntsville (UAH) MSU v5.4 data (Christy et al. 2003) on the standard  $2.5^\circ \times 2.5^\circ$  lat/lon grid at the LS and MT layers and National Environmental Satellite, Data and Information Service (NESDIS) Center for Satellite Applications and Research (STAR) (Zou et al. 2009) version 2.0 data at LS, TS, and MT layers. The RSS, UAH, and STAR temperature datasets have been calculated from the same raw radiance datasets, though with quite different analysis methods.
- The updated Hadley Centre Atmospheric Temperature (HadAT2) dataset (Thorne et al. 2005a; McCarthy et al. 2008). It is a well-known homogeneity-adjusted pure radiosonde dataset containing 676 stations. The input data resolution is monthly and the adjustments are provided with seasonal resolution. An ensemble of realizations of this dataset is described by Thorne et al. (2011a). Minima, 25%/50%/75%, and maxima of zonal belt mean trends 1979–2003 are available. Some of the variations in this ensemble appear very strong (e.g., a removal of signal experiment). Since the trend interval is also shorter than the one analyzed in the present paper, the uncertainties are likely smaller than the values given for this ensemble. Therefore we preferred to use the interquartile range of this ensemble as uncertainty estimate in the comparisons below.

For comparison with satellite data, the radiosonde and reanalysis temperature profiles have been converted into brightness temperatures using the Radiative Transfer for TOVS version 10 (RTTOV v10) software package (Saunders et al. 2011). This is an improvement to earlier studies that used static weighting functions that leads to up to 20% less variance in the difference series between brightness temperatures from adjusted radiosonde data and MSU data, especially in the lower stratosphere.

We did comparisons also with Global Positioning System radio occultation data for the period 2001–10, but these are published elsewhere (Ladstätter et al. 2011; Steiner et al. 2011).

#### 4. Estimation of adjustment uncertainty

Homogenization is useful if the adjustment uncertainties are smaller than the time variation of biases that need to be estimated. That has to be shown for the

adjustments of individual breaks as well as for the global mean adjustments that affect estimates of the global climate change signal.

For this purpose we need to understand the uncertainties in the adjustments that are calculated by RICH/RAOBCORE. We reduce the uncertainty of the low-frequency variability estimates only if we can show that the adjustment uncertainties are appreciably smaller than the estimated time-varying biases. This section tries to highlight some manifestations of the adjustment errors due to sampling of neighbor stations and due to setting of parameters in the adjustment methods. Differences between RICH-obs/RICH- $\tau$  and RAOBCORE may be seen as manifestation of structural uncertainties (Thorne et al. 2005a), whereas differences within the RICH-ensembles are interpreted as parametric uncertainties.

About 8000 breaks are found by RAOBCORE in the period 1958–2011, and for each of these breaks, samples of neighboring records to adjust this break have to be determined. Individual break profiles tell much about the uncertainties involved in estimating the adjustments. Figure 1 shows the estimated profiles of a break caused by transition from VIZ-A to Space Data radiosondes in 1989 at station Bethel (70219) in Alaska. The thin lines in the upper-left panel of Fig. 1 are individual break size estimates from comparison of neighboring anomaly records ( $\Delta_{\text{obs}}^{jk}$ ). The thick green line is the RAOBCORE estimate for this break, and the thick blue line is the distance weighted mean of the  $\Delta_{\text{obs}}^{jk}$  profiles, which comprises the RICH-obs break size estimate. The break size estimates are all clearly nonzero, and one can say that the adjustment uncertainties are much smaller in this case than the estimated shift. While the agreement between RAOBCORE and RICH-obs estimates is reasonable, there is considerable spread of the  $\Delta_{\text{obs}}^{jk}$ .

When estimating the same break by comparing neighboring innovation records ( $\Delta_{\tau}^{jk}$ ), the spread is much smaller. For this plot it appears clearly advantageous to use innovations instead of observations of neighboring stations for comparison. In particular, the reduced spread indicates that regional climate anomalies have substantial influence on the break size estimates if they are not accounted for. This effect is largest for remote stations where one has to compare stations that are far apart. In the case of this specific break at Bethel the maximum distance to a reference station is 3070 km.

Note also that no vertical smoothing has been applied at any stage of break size estimation, which is an important feature of the present adjustment system. Earlier experiments with vertical smoothing generally yielded unsatisfactory results, since the vertical profiles of the biases can be rather complex. At some occasions

there are individual profiles that are clear outliers. This may indicate that the reference station record is not as homogeneous as thought, or it may come from a very different region than the other reference profiles. Of course, there may also be issues with data density in the time series that may cause spread. As mentioned above, the maximum and minimum individual estimates are removed to increase the robustness of the estimate.

While Fig. 2 shows the sampling errors involved in estimating the break sizes in individual profiles, it is instructive to see also how much RAOBCORE and RICH break size estimates differ. In the example in Fig. 1 this difference between the thick blue and green curves is on the order of 0.5 K. Figure 4 depicts the rms difference between RAOBCORE and RICH-obs adjustment estimates averaged over stations with different “remoteness,” measured as the distance between the tested station and the most distant station used for comparison. As long as this distance is less than 1000 km the rms difference is quite small for both RICH-obs and RICH- $\tau$ . If it is larger the difference between RAOBCORE and RICH estimates can become relatively large, especially for RICH-obs. The difference becomes also large for the highest levels (<50 hPa), mainly because reference stations with enough data are hard to find there.

#### a. Spatiotemporal consistency of adjusted datasets

Inhomogeneities in an observation dataset often manifest themselves in large trend discrepancies at nearby stations. Trends from homogenized records should therefore be spatially more consistent than the unadjusted datasets. Figure 5 indeed shows improved spatiotemporal consistency of trend estimates for the period 1979–2006 for the MSU-equivalent lower stratospheric layer for all three homogenization methods. The adjusted datasets are more consistent than those shown in Fig. 1 of Haimberger et al. (2008) with parameter settings similar to those used in this paper. The improvements mainly come from the improved background and changes in the neighbor station selection. The use of RTTOV for calculating brightness temperatures also contributes to the better consistency, which has been quantified with a cost function<sup>1</sup> introduced by Haimberger (2007).

While we think it is clear that a homogenization algorithm should improve spatiotemporal consistency, this parameter is not necessarily an indicator of improved temporal homogeneity of large scale means. For example, MSU brightness temperatures are spatiotemporally

<sup>1</sup>  $\text{Cost}[1/N(N-1)] \sum_i \sum_{j=i+1}^N \Delta_{ij} \exp^{-d_{ij}/1000\text{km}}$ , where  $i$  and  $j$  are radiosonde stations,  $d_{ij}$  is distance in km and  $\Delta_{ij}$  is the trend difference in K decade<sup>-1</sup>.



1 DECEMBER 2012

HAIMBERGER ET AL.

8119

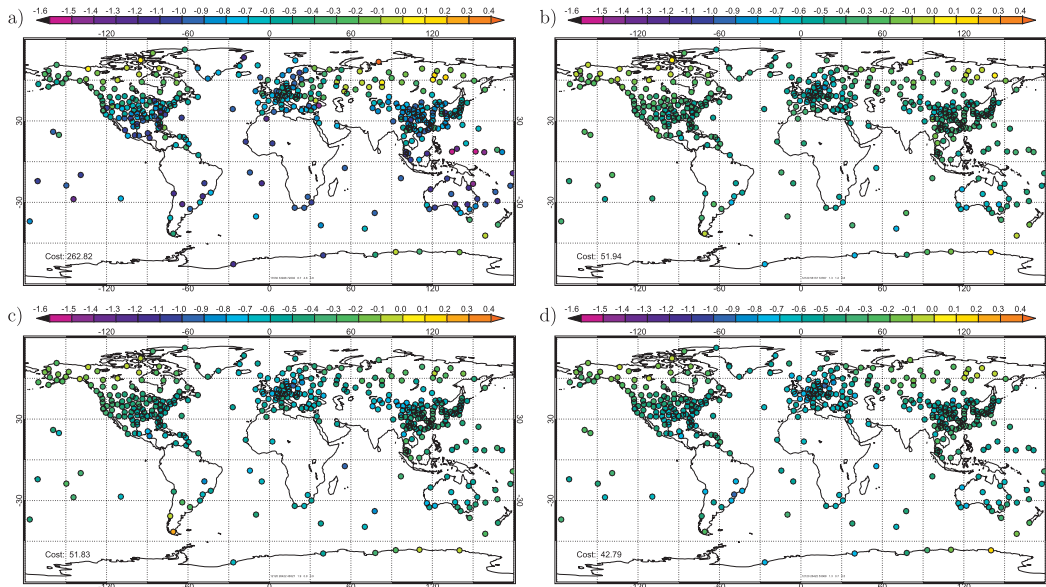


FIG. 5. Daily mean TLS trends 1979–2006: (a) unadjusted, (b) adjusted with RAOBCORE, (c) adjusted with RICH-obs ensemble mean, and (d) adjusted with RICH- $\tau$  ensemble mean. Reduced cost values (in lower left corners of panels) indicate improvement compared to Fig. 1 of Haimberger et al. (2008). Only for this plot the 30-hPa level has been omitted in the brightness temperature calculation to allow a better comparison.

very consistent, but nevertheless can have significant inhomogeneities or at least uncertainties (Mears et al. 2011). The temporal homogeneity can only be guaranteed with measurement practices that allow traceability to SI standards (Seidel et al. 2009). These are lacking for historic radiosonde observations. One can, however, constrain the uncertainties by comparison with independent data and with sensitivity experiments where uncertain parameters of the adjustment system are varied.

#### b. Sensitivity experiments and ensembles

In RICH, the following parameters have been varied to estimate their influence on adjustment results when the breakpoints from RAOBCORE are fixed:

- Using either RICH-obs or RICH- $\tau$  for break adjustments. Both versions have their merits as discussed above, but they can yield quite different results particularly at remote locations.
- The test station and a reference station must have a minimum number of data points in both intervals  $a$  and  $b$  (see Fig. 3) to deliver a valid break size estimate at one level. This minimum data threshold for adjustment is varied between 130 and 330 values for interval  $a$  and  $\sim 200$ –900 values for interval  $b$ . The minimum number is dependent on the pressure level.

For the higher levels fewer data are required. A smaller number increases the chance to find nearby neighbors for adjustment calculation but produces higher noise and makes the estimates more susceptible to effects of short-term climate anomalies, especially for RICH-obs.

- The number of neighbors used for calculating the composites is another important parameter. It is varied between 3 and 10 in the first iteration and between 9 and 30 in the second iteration (see Fig. 3). A small number reduces the risk that information from very distant stations can enter the break size estimates but leads to noisier estimates due to the small sample of profiles.
- We used inverse distance weighting of the neighbors. The  $1/e$  decay of the weighting function defined in (15) was varied between 3000 and 5000 km.
- When calculating adjustments, one can start with the uppermost or the lowermost pressure level. When one starts with the uppermost level, suitable neighbors are probably farther away but those likely have a full profile. The same set of stations is used for all levels of an adjustment profile in this case.

When one starts with the lowermost level, nearby stations are more likely to be used, even if they have few data in the uppermost levels. Data for the uppermost levels are then taken from more distant but complete

records. This means that for one adjustment profile, different sets of stations may be used for different levels.

- Treatment of the Phillips-RS4-MKIII to Vaisala-RS80 transition over Australia/Malaysia and the surrounding islands. This transition during the years 1987–89 was a major change in the radiosonde observing system in the tropics and the Southern Hemisphere. Because of the sparseness of reference stations and due to the fact that a major ENSO event (1987 El Niño followed by a strong La Niña in 1988) caused large regional anomalies over the tropics, there is much uncertainty involved in adjusting with neighbor composites. We used two strategies:

- a) Adjust the Phillips-Vaisala breaks using a profile estimated from an international radiosonde intercomparison (Nash and Schmidlin 1987). In particular, RICH-obs has problems getting reliable break size estimates when two breakpoints are relatively close to each other. As an additional measure, some (about 50) breakpoints at remote stations have been deleted to get longer intervals for break size estimation.
- b) Let RICH work as usual (i.e., strictly use neighbor intercomparison for break size estimation; do not delete any breakpoints detected in RAOBCORE).

In total, two choices have been tried for six parameters, yielding an ensemble with  $2^6 = 64$  members. This approach is less exhaustive than, for example, employing a Monte Carlo technique but nevertheless allows us to explore at least part of the parameter space of the present adjustment methods.

The combination of all these parameter settings yields moderate spread in the RICH adjusted trend profiles. While there is not enough space to show the effect of each particular parameter setting, it can be said that all contribute appreciably to the spread. The effect of using either RICH-obs or RICH- $\tau$  is indicated by the gray shades. For the interval shown in the tropics, RICH-obs leads to vertically smoother trend profiles than RICH- $\tau$ , which yields trend profiles that are roughly between RAOBCORE and RICH-obs. Below we often split the ensemble into 32-member RICH-obs and RICH- $\tau$  ensembles because of their fundamental difference concerning data dependency. Separate ensemble means of the RICH-obs and RICH- $\tau$  members have been used in Figs. 6–10 below.

Since RICH results are also dependent upon the breakpoint dataset provided by RAOBCORE, we conducted sensitivity experiments with breakpoints from different RAOBCORE versions:

- 1) breakpoints from RAOBCORE v1.5 (these are used for most plots in the present paper);
- 2) breakpoints from RAOBCORE v1.5 where no prior adjustment of ERA-40 background between 1971 and 1978 has been applied, similar to RAOBCORE v1.3 in Haimberger et al. (2008);
- 3) breakpoints from RAOBCORE v1.5 where no metadata from documented changes of equipment are taken into account;
- 4) breakpoints from RAOBCORE v1.5 with neither background adjustment nor metadata; and
- 5) RAOBCORE v1.4 breakpoints, as in Haimberger et al. (2008). This is for comparison with older findings. The main differences between RAOBCORE v1.4 and RAOBCORE v1.5 are the absence of breaks after 2005 in RAOBCORE v1.4 and the use of ERA-Interim as reference from 1979 onward in RAOBCORE v1.5.

### 5. Selected results

Adjustment results are documented online (<http://www.univie.ac.at/theoret-met/research/raobcore/>) with thousands of adjustment profile plots, trend maps, and time series. Only a few are highlighted here. Figure 5 shows how RAOBCORE and the two RICH versions improve the spatial consistency of TLS equivalent trends. This figure can be compared with Fig. 1 of Haimberger et al. (2008). A clear improvement is noticeable compared to these earlier versions of RICH-obs and RAOBCORE.

Much improved spatiotemporal consistency can be found also in the presatellite period. This period was characterized by several quite stable national observing systems, like in the United States, but also by observing systems with extreme changes. Most notably the stratospheric temperature biases in early MESURAL radiosondes shifted by up to 10 degrees K over Europe as well as at several stations in the Pacific. There were also strong changes over the former Soviet Union, and these shifts lead to the rather noisy trend pattern in Fig. 6a. The mean RICH-obs adjustments are able to remove most breaks, yielding a surprisingly consistent map of trend estimates even at 100 hPa, as shown in Fig. 6b. Chinese radiosondes are available but are missing in this plot since early Chinese ascents did not reach higher than 300 hPa. The trend map (trend cost value 221) is even more consistent than RAOBCORE adjusted trends (cost 355) and trends from ERA-40/ERA-Interim (cost 248). If we restrict the ensemble to members that use 30 neighbors, the cost is reduced even to 185.

One should not ignore, however, that there are a few remote stations whose trends still look unrealistic after adjustment. We could have removed them but we found it instructive to see the limitations of the method as well. It is also important to recognize that we show only the ensemble mean. Many of these station trends that look

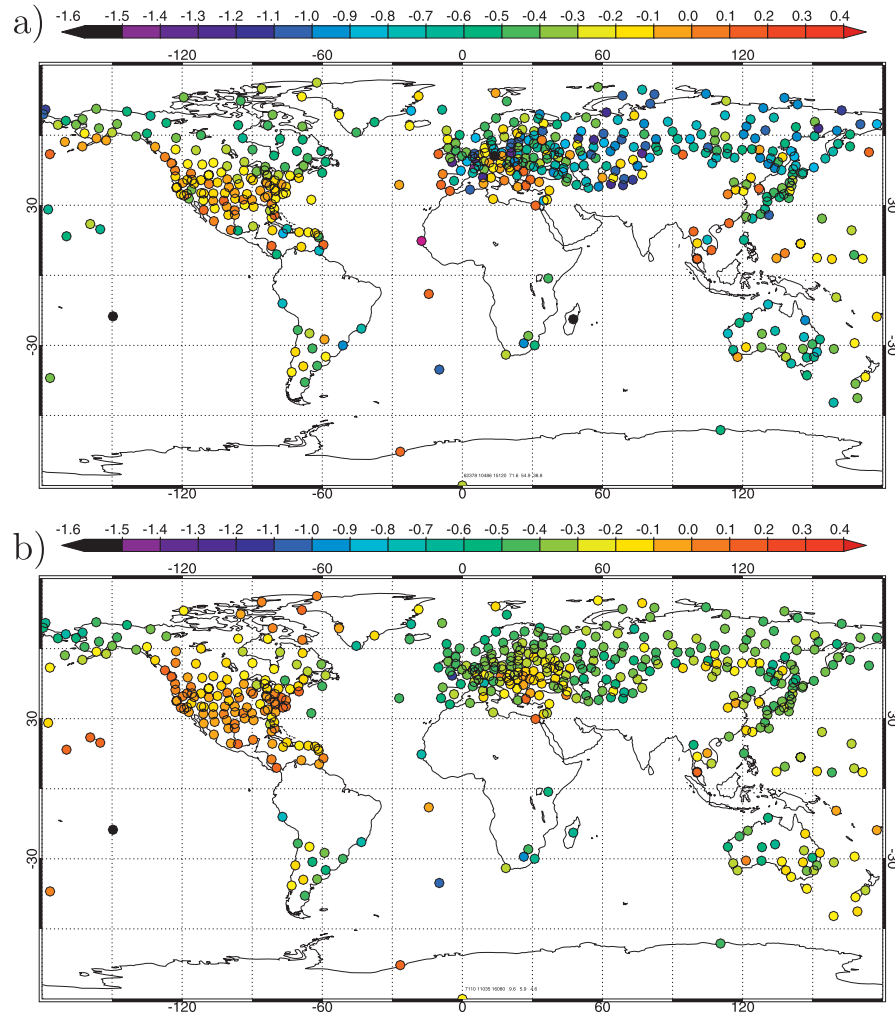


FIG. 6. Daily mean 100-hPa trends 1964–84 in units  $\text{K decade}^{-1}$ : (a) unadjusted, trend cost 1564, (b) adjusted with ensemble mean RICH-obs, trend cost 221. Especially in the early period, RICH has been improved compared to Haimberger et al. (2008). The RICH cost value with this past version of RICH was 508.

suspicious in the ensemble mean are consistent with their neighbors in some ensemble members. Certainly these stations need to be worked on in future releases. It is also interesting to see that trends over America tend to be close to zero or even positive whereas trends over Europe/Asia tend to be negative. It is quite possible that the breaks that lead to strong negative trends in the unadjusted data over the former Soviet Union have not been fully removed.

Since there are practically no other upper air data to compare with, the internal consistency of the adjusted

dataset, the spread of adjustment ensembles, and comparison with other adjusted radiosonde datasets, such as Thorne et al. (2011a) must be the main quality measures. The sensitivity experiments below and comparison with HadAT are some steps in this direction, although more efforts in this direction are needed.

#### a. Comparison with satellite data

For the satellite era from 1979 onward, there are more possibilities for intercomparison. Figures 7 and 8 show Hovmöller plots of zonal mean temperature anomaly



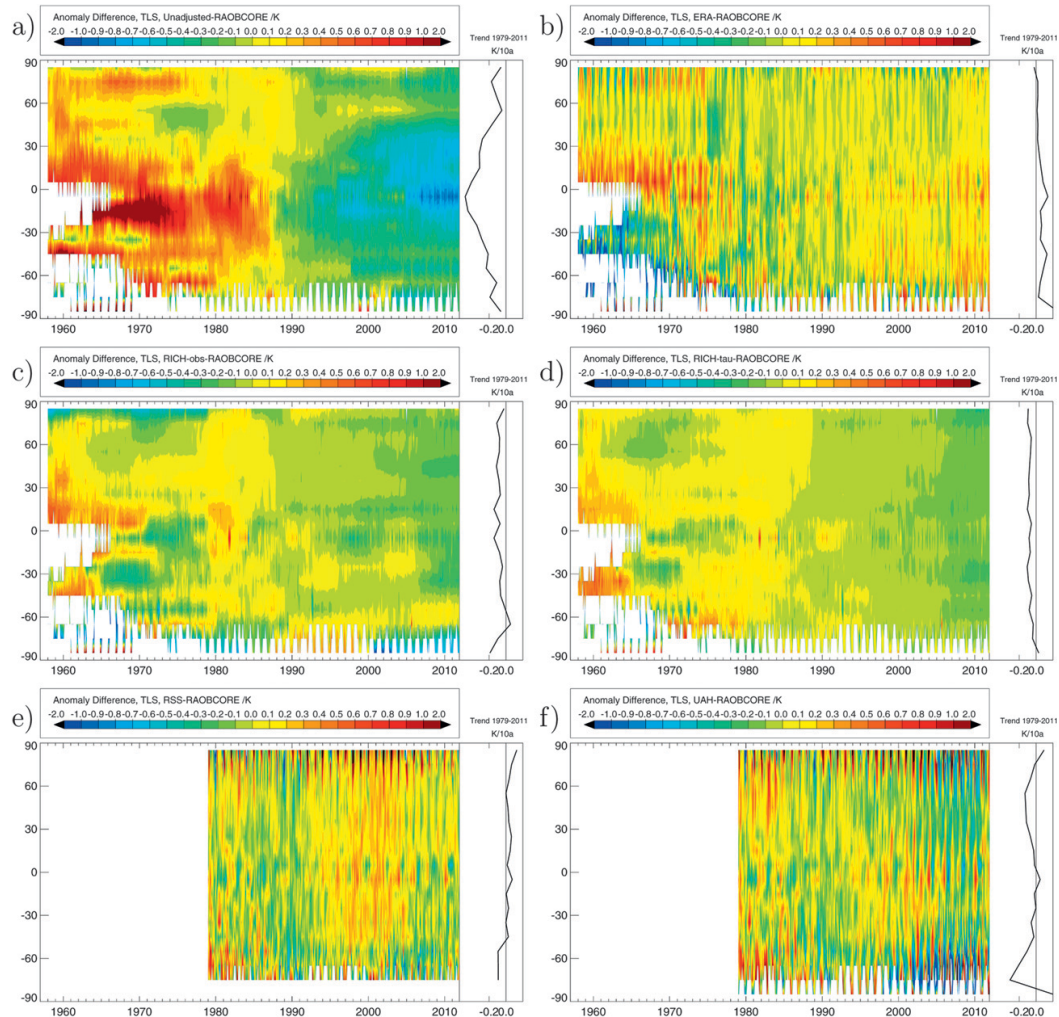


FIG. 7. Hovmöller plots of zonal mean temperature anomaly differences (anomalies relative to 2001–10 climatology, all differences relative to RAOBCORE v1.5 adjusted temperatures) for the MSU LS layer. Black lines at right of plots show trend difference 1979–2010. Differences to (a) unadjusted radiosonde data, (b) to ERA-40/ERA-Interim reanalyses, (c) mean of RICH-obs ensemble, (d) mean of RICH- $\tau$  ensemble, (e) RSS brightness temperatures, and (f) UAH brightness temperatures.

differences at the lower stratospheric and midtropospheric layers. RAOBCORE v1.5 adjusted temperature anomalies are taken as reference.<sup>2</sup> Figure 7a shows how substantial the temperature adjustments are in the lower stratosphere, particularly in the tropics. The absolute lower stratospheric cooling trends are essentially halved

by the adjustments. The strong warm bias of the unadjusted radiosondes in the tropics in the presatellite era mainly stems from many stations using MESURAL equipment in the tropical Pacific. Another important period of changes was 1987–89 when Australia changed the radiosonde type and the early to mid-1990s when several former French colonies and the United States changed the radiosonde type in the tropical Pacific. Relatively small adjustments are made at polar regions, at least in the zonal mean.

<sup>2</sup> This does by no means imply that the RAOBCORE-adjusted records are the most reliable dataset.

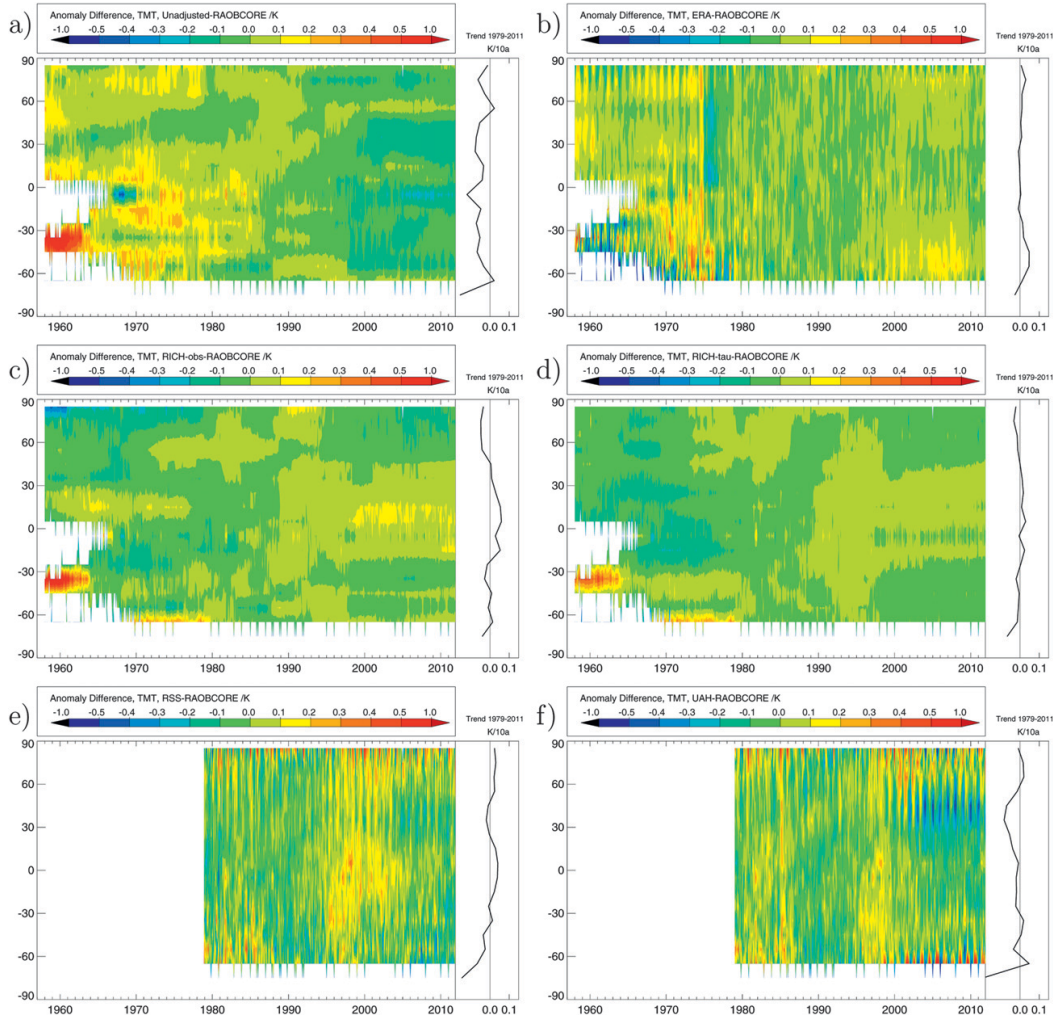


FIG. 8. As in Fig. 7, but for the MSU MT layer. Note different contour interval and scaling.

The ERA-Interim analysis shows less cooling than RAOBCORE in the LS in the satellite era, particularly in the tropics and the Southern Hemisphere. Warming relative to RAOBCORE occurs in 2007, when data from the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) have been introduced in ERA-Interim (Poli et al. 2010). Both RICH versions (mean RICH-obs and mean RICH- $\tau$ ) show more cooling than RAOBCORE at all latitudes, the difference being on the order of  $0.1 \text{ K decade}^{-1}$ .

RSS and STAR (not shown) data yield less cooling than RAOBCORE in the LS, although RSS temperatures

have recently (since ca. 2002) cooled compared to RAOBCORE. UAH shows about the same cooling as RAOBCORE/RICH. Cooling from UAH in the extratropics is rather strong compared to the other datasets.

The biases in the MT layer (Fig. 8) are much weaker but still the unadjusted data show cooling relative to RAOBCORE and the satellite datasets. Agreement between RAOBCORE, RICH- $\tau$ , and ERA-Interim reanalysis is excellent in the satellite era. In the pre-satellite era, RAOBCORE anomalies are cooler than ERA-40, and known homogeneity problems in ERA-40 in the 1970s (Uppala et al. 2005) are evident. There is

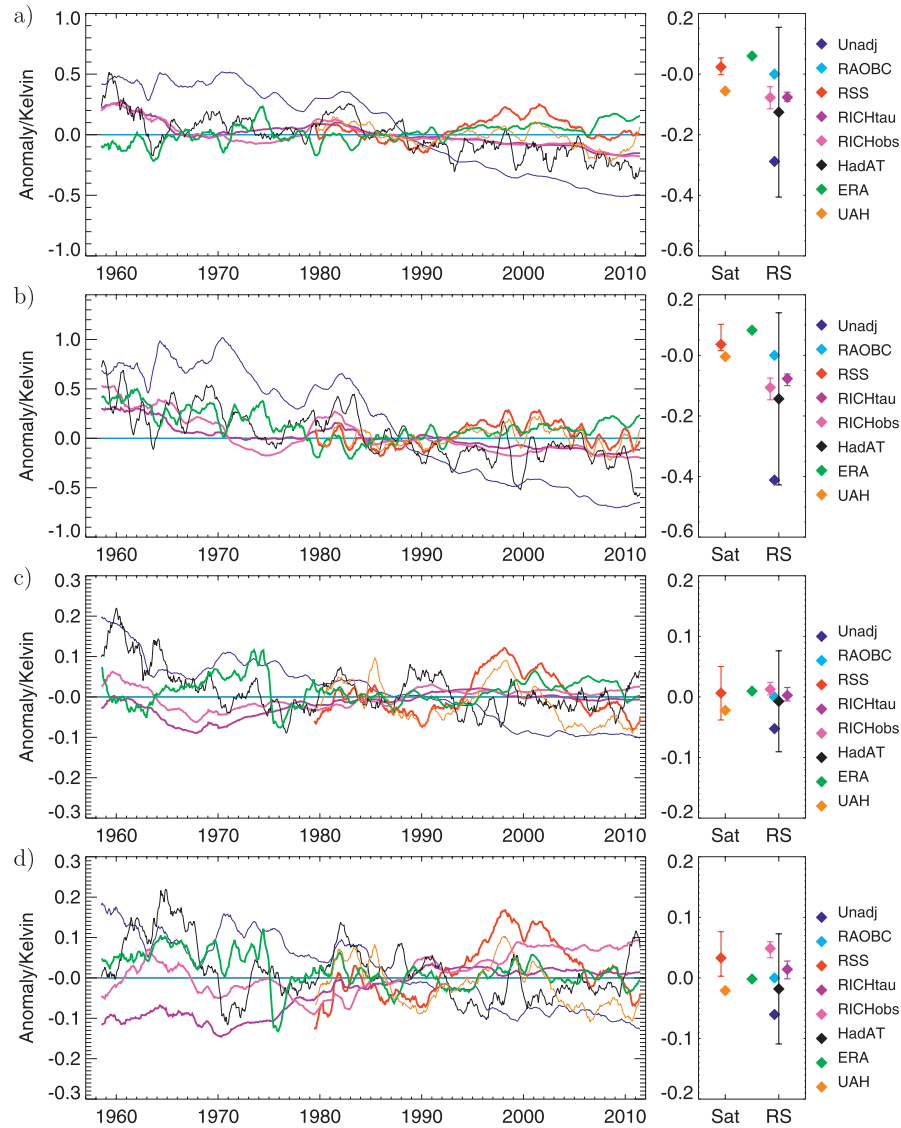


FIG. 9. MSU equivalent temperature difference time series relative to RAOBCORE v1.5. Trend difference for period 1979–2011 in  $\text{K decade}^{-1}$  is indicated in right panels. (a) LS global, (b) LS tropics ( $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ), (c) MT global, (d) MT tropics. Note scaling differences. Uncertainty bars for RSS, HadAT, RICH-obs, and RICH- $\tau$  are 5% and 95% percentiles. Original HadAT trend spread valid for 1979–2003 has been reduced by a factor of 1.32 for the longer interval 1979–2011. Absolute trends of RAOBCORE v1.5 can be found in Fig. 13.

better consistency with HadAT (not shown), although there is more warming in RAOBCORE/RICH than in HadAT. The RSS and UAH satellite data show generally stronger warming than the radiosonde datasets and ERA-Interim until 2002 and substantial cooling compared

to these datasets thereafter, especially in midlatitude regions. This cooling in the past decade is noticeable not only compared to radiosondes but also compared to GPS radio occultation (GPS-RO) data (Ladstätter et al. 2011; Steiner et al. 2011). Those authors argue that residual

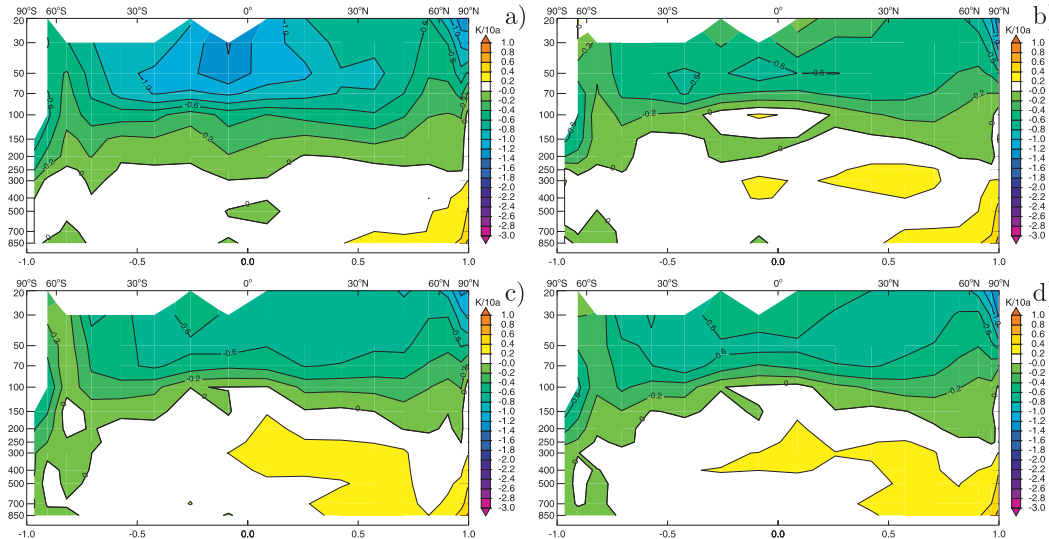


FIG. 10. Zonal mean trends 1979–2010 from (a) unadjusted, (b) RAOBCORE adjusted, (c) mean RICH-obs adjusted, and (d) mean RICH- $\tau$ -adjusted radiosondes.

inhomogeneities in the MSU datasets are the most likely cause given the excellent agreement between GPS-RO data from various platforms and also given the good agreement between radiosondes and GPS-RO after 2002.

Figure 9 shows global belt mean MSU (equivalent) brightness temperature anomaly differences with respect to RAOBCORE v1.5. RAOBCORE and RICH show better agreement with satellite data in the LS than HadAT in the satellite era. In the presatellite era, there is no upper air dataset independent of radiosondes. We can only compare with other radiosonde datasets and reanalyses. During this period RAOBCORE adjusted data have least cooling in the tropics, whereas they show good agreement with the ERA-40 reanalysis when averaged over the whole globe. In the tropics, the ERA-40 analysis may be biased warm in the LS due to the effect of unadjusted MESURAL observations before 1975.

The lower panels show the comparison in the MT layer. Note that the temperature scale is very fine: biases of 0.1 K appear large in this plot. There is generally excellent agreement of between RICH/RAOBCORE and the ERA-40/ERA-Interim analysis. Only in the tropics RICH-obs shows more warming in the satellite era, most likely because its profiles do not exhibit the warming minimum in the tropics around 700 hPa that is evident in ERA-Interim and RAOBCORE adjusted data. The transient warming feature of the UAH and RSS MT temperatures relative to RAOBCORE with a peak in 1998 that has been noted above for the LS is

very evident also in the MT. It could be related to enhanced uncertainty in the MSU record at this time (Mears et al. 2011).

The zonal mean trend spread within the RICH-obs and RICH- $\tau$  ensembles is rather small in the satellite era compared to the spread given by the HadAT and RSS ensembles. Larger spread may be achievable by varying the breakpoints, but Figs. 11c and 11d suggest that this also has limited impact. We found that too strong variation of parameters inevitably led to strongly reduced spatiotemporal consistency of trend estimates and is therefore not justifiable. If one considers the differences among RAOBCORE, RICH-obs, and RICH- $\tau$  as crude uncertainty estimates, it is comparable to the uncertainties estimated from the RSS and HadAT ensembles.

In the presatellite era, Fig. 9 shows more spread between the radiosonde datasets. It should be noted, however, that observation density is quite limited such that different treatment of one breakpoint at one station in the tropics can have quite an effect on the curves shown in Fig. 9. RICH- $\tau$  yields the coolest MT temperatures. The difficulties in ERA-40 analyses due to assimilation of Vertical Temperature Profile Radiometer (VTPR) temperatures (Uppala et al. 2005) are evident in the analysis time series despite measures to adjust the global mean background forecasts between 1971 and 1978.

Despite some improvements in spatiotemporal consistency also for the time intervals 1973–2006 and 1958–2006



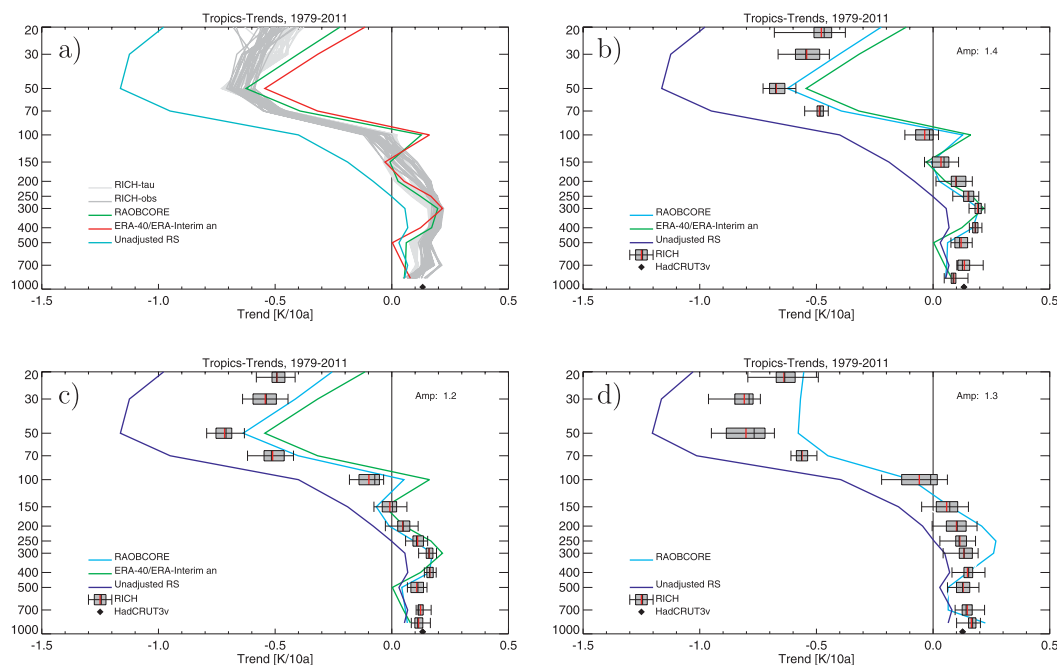


FIG. 11. Plots of tropical belt mean trends, using different RAOBCORE versions for break detection, using (a) RAOBCORE v1.5, (b) the same data as in (a) but ensembles are depicted with whisker quantile plots [markers in whiskers are minimum, maximum, 25%, 50%, and 75% quantiles, and mean (red)], (c) RAOBCORE 1.5 but without background correction between 1971 and 1978 and without metadata, and (d) the earlier RAOBCORE version v1.4.

(not shown) at lower levels in the tropics compared to RAOBCORE 1.4, one should therefore still be cautious when interpreting time series from single stations or small regions. One should also note that the adjustments only affect temperatures, not dewpoints. This is important for the analysis of radiosonde humidity measures such as dewpoint depression time series (see, e.g., Dai et al. 2011). For example, the difference between the various radiosonde datasets is 0.2–0.3 K in the 1960s in the tropical MT layer in Fig. 9d). A dewpoint depression error of 0.2 K converts to about 1% uncertainty in relative humidity.

#### b. Vertical structure of temperature changes

One strength of radiosonde data is their vertical resolution, which allows us to resolve the sharp transition from tropospheric heating to stratospheric cooling much better than MSU data can.

Figure 10 shows that all three homogenization methods yield a warming maximum in the tropical midtroposphere in the period 1979–2010 in contrast to unadjusted data, which show cooling in the deep tropics' troposphere. RAOBCORE yields the weakest maximum in the tropical

midtroposphere; however, it has an additional warming maximum at the 100-hPa level that appears to be inherited from the ERA-Interim bg (not shown) that has a similar spike. The medians of the RICH- $\tau$  and RICH-obs ensembles yield vertically smoother trend profiles and quite similar zonal mean trends, with RICH-obs showing slightly less stratospheric cooling in the southern extratropics. The RICH estimates in the Southern Hemisphere, also near Antarctica, look much more reasonable than they did in the first version of RICH and RAOBCORE, where Haimberger et al. (2008) found that RICH showed only slightly more warming than unadjusted data and RAOBCORE adjusted trends had unrealistic vertical variations.

Figure 11a shows tropical belt mean vertical trend profiles of unadjusted data, ERA-Interim analysis data, RAOBCORE adjusted temperatures, and individual profiles from the RICH-obs and RICH- $\tau$  ensembles. In Fig. 11b the RICH data are presented as quantile plots that contain both RICH-obs and RICH- $\tau$  ensembles. The belt means for the tropics show a robust upper tropospheric warming maximum that is gentler and looks more plausible than the RAOBCORE maximum.

The RICH profiles also look more plausible than those from ERA-Interim, which shows a strong warming maximum at 100 hPa. This maximum is likely related to a problem in the assimilation of satellite data near this level in the early 1980s since the maximum disappears when analyzing trends from periods starting five years later.

The impact of using breakpoints from different RAOBCORE versions on RICH is relatively small, as can be seen from Figs. 11c and 11d. Figure 11c shows the profiles of RAOBCORE and RICH if no metadata and no background adjustment in 1971–78 are applied. There is slightly stronger cooling in the stratosphere and a less pronounced warming maximum in the troposphere.

The trend profiles from RAOBCORE v1.4 had quite different shapes compared to RAOBCORE v1.5 (cf. Figs. 11b and 11d), with more warming in the upper troposphere and less cooling in the stratosphere. The different breakpoints have some influence on the RICH trends as well but these show less variation than is the difference between RAOBCORE versions 1.4 and 1.5. This underlines the robustness of both RICH-obs and RICH- $\tau$  to moderate changes in the location of breakpoints.

The inset parameter Amp is the ratio between surface trend and the maximum trend of the ensemble mean in the troposphere. For the RAOBCORE v1.5 based profiles it is 1.4 whereas for the RAOBCORE v1.4 based profiles it is 1.3. Both values are within the uncertainty range of climate models (see Santer et al. 2005) although these often predict even stronger amplification. They are also in accord with evaluations at ECMWF (A. J. Simmons 2011, personal communication) showing that the amplification factor tropical surface temperature variability (not trends) for 300-hPa temperatures in the ERA-Interim reanalysis for this period is 2.2. Haimberger et al. (2012, manuscript submitted to *Meteor. Z.*) demonstrate that this warming amplification seen here in the interval 1979–2011 is persistent and often stronger in practically all 21-yr periods since 1960. This is new evidence that amplification of surface trends in the tropics, which has been subject of debate for 20 years (Thorne et al. 2011b; Douglass et al. 2008; Santer et al. 2008), is real. An earlier analysis of 21-yr periods only a few years ago (Thorne et al. 2007) yielded clearly smaller amplification factors. Note also, however, that this amplification factor is highly dependent also on uncertainties in the surface datasets. For example, Kennedy et al. (2011b) specify about  $0.05 \text{ K decade}^{-1}$  global mean marine temperature trend uncertainty for 30-yr trends in recent periods.

The vertical trend profiles 1979–2011 in other than tropical regions, now using again RAOBCORE v1.5 and the corresponding RICH ensembles, are shown in Fig. 12. RICH adjusted data show more warming than

ERA-Interim at the lowest levels, generally fitting excellently to the Hadley Centre–Climatic Research Unit temperature dataset (HadCRUT3v; Brohan et al. 2006) surface trends. They show more cooling than ERA-Interim above 100 hPa, which is attributed to still unadjusted breaks in RICH and to the introduction of COSMIC data in ERA-Interim from 2007 onward. This event has a clear effect in the LS temperatures of ERA-Interim, as shown in Figs. 7 and 9.

Figure 13 gives an overview of trends for four MSU layers in the tropics and for the globe. The values can be compared with those in Blunden et al. (2011), although these are valid for 1979–2010. The RICH ensembles fit very well with ERA-Interim at the MT layer and show more warming than other radiosonde datasets and the reanalysis in the LT layer. In general, RICH shows more warming than UAH and fits very well to RSS in the LT and MT. Trends from STAR v2.0 show the most pronounced warming of all datasets at the MT and TS layers. Trends for the TS layer are not available over the 1979–2011 period from RSS and UAH.

In the LS the RICH estimates show more cooling than reanalysis and satellite datasets, but less cooling than HadAT data. The originally large gap between unadjusted radiosonde data and satellite data has been reduced considerably. Only RAOBCORE estimates lie within the uncertainty bounds given by RSS and fit well to the estimates by UAH. This agreement should not be overinterpreted, however, since RAOBCORE estimates are not independent of satellite data.

The spread of the RICH ensemble is rather small, particularly in the global mean. It is likely that the ensemble generated from the sensitivity experiments underestimates the true uncertainty. Little spread has been generated particularly in the Northern Hemisphere, as can be seen also in Fig. 12b). Nevertheless it is encouraging that at least three datasets now provide improved uncertainty guidance through ensemble methods.

## 6. Discussion and conclusions

This paper described improvements on radiosonde temperature homogenization made with the RAOBCORE and RICH homogenization methods, where RICH has been described in some detail. Both methods utilize background departure statistics available from climate data assimilation systems such as ERA-40 (Uppala et al. 2005). The method used for break detection (RAOBCORE) has already been described by Haimberger (2007). While RAOBCORE uses the background forecasts also for break size estimation, RICH estimates the breaks by comparison with reference series generated from surrounding radiosonde

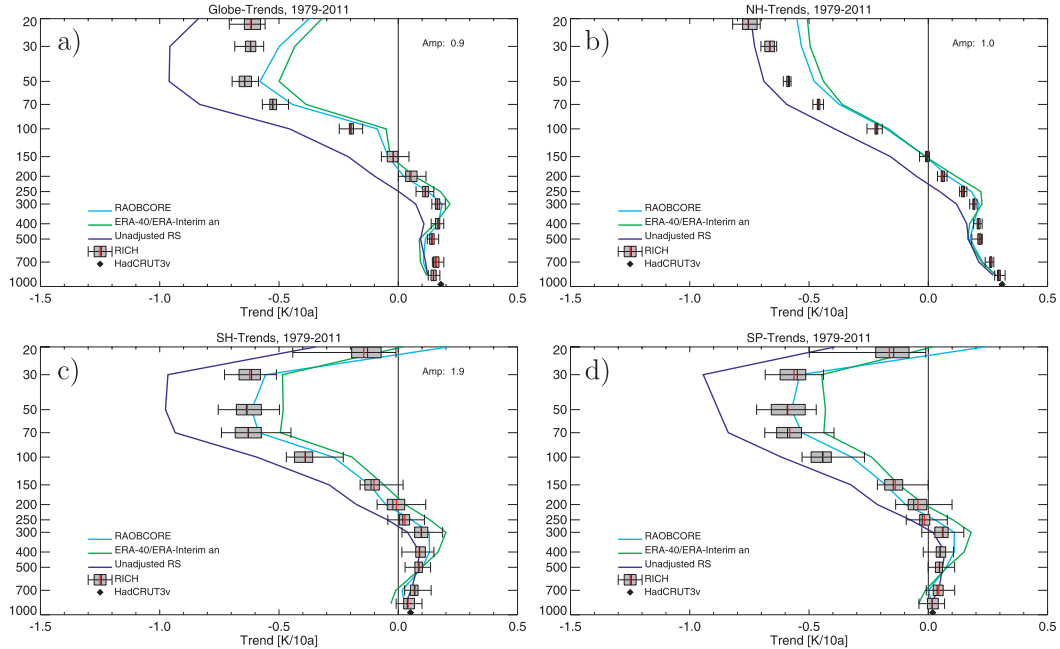


FIG. 12. Whisker plots of global belt mean trends: (a) globe, (b) extratropical Northern Hemisphere ( $>20^{\circ}\text{N}$ ), (c) extratropical Northern Hemisphere ( $>20^{\circ}\text{S}$ ), and (d) southern polar region ( $>60^{\circ}\text{S}$ ).

stations. Reference series are generated either with (i) radiosonde observations (RICH-obs) or (ii) background departures of neighboring radiosonde stations (RICH- $\tau$ ). While RICH-obs estimates are independent of the background and thus of satellite information, RICH- $\tau$  seems to have advantages for individual break size estimation. RICH- $\tau$  achieves the best overall

spatiotemporal consistency of trend estimates in the satellite era. In the presatellite era, RICH-obs has the best spatiotemporal consistency. In the zonal belt means it is hard to tell at the present stage whether RICH-obs or RICH- $\tau$  yield more accurate results. Thus, using RICH-obs seems advantageous for now since the break size estimation process is independent of satellite data.

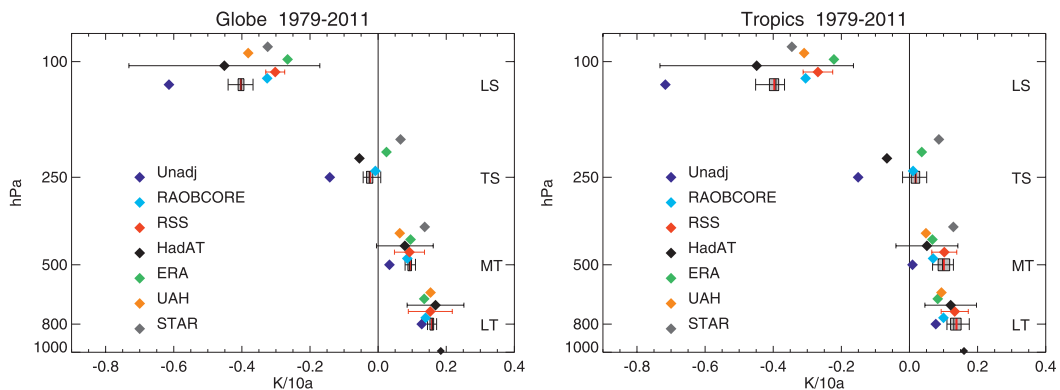


FIG. 13. Global and tropical belt mean MSU equivalent trends 1979–2011 from all datasets used in this study. HadCRUT3v surface trend is  $0.12 \text{ K decade}^{-1}$ . The gray whiskers depict the combined RICH-obs/RICH-tau v1.5 ensemble. Vertical shifts between markers are solely for better readability. RSS, UAH, and HadAT ensemble trends are not available for TS layer. MSU-STAR trends are not available for LT.

It was shown that the radiosonde temperatures adjusted by either RICH-obs or RICH- $\tau$  are more consistent with independent satellite estimates than other homogenized radiosonde temperature datasets. In sensitivity experiments and a few illustrative examples it has been shown that

- 1) RICH-adjusted radiosonde data reveal the upper tropospheric warming maximum projected by climate models and in most cases lead to spatially more consistent trend patterns than RAOBCORE in both the presatellite and satellite period.
- 2) In the satellite era adjusted trends still show less warming/more cooling than RSS and STAR satellite datasets, especially in the period 1979–99 and in the LS layer. In the period 1999–2011 they show more warming/less cooling than RSS/UAH.
- 3) In the presatellite era, the uncertainties become larger but all three adjustment methods do a good job in making the radiosonde series more consistent. It can be expected that rejection rates of the adjusted radiosonde data will be much smaller when assimilated with a climate data assimilation system.
- 4) RICH- $\tau$  is only slightly more sensitive to errors in the ERA-40 background than RICH-obs. In most cases it could be seen that RICH- $\tau$  time series and trends are between RICH-obs and RAOBCORE.
- 5) It is essential to have a good breakpoint database and to avoid averaging across breakpoints.

The performance of the RAOBCORE/RICH adjustment system suggests that the use of background departures from climate data assimilation systems helps to improve the original observation datasets. The next candidate is radiosonde wind, where earlier studies showed promising results (Gruber and Haimberger 2008); others are tropospheric humidity and surface parameters in remote areas. It is expected that the presented methodology will also help improving pre-1958 upper air data, as they are collected in the ongoing ERA-CLIM project (<http://www.era-clim.eu>). Thousands of plots and data as well as additional documentation can be found online at <http://www.univie.ac.at/theoret-met/research/raobcore/>.

**Acknowledgments.** This work has been funded by projects P18120-N10 and P21772-N22 of the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF), as well as by the EU 7th Framework Programme collaborative project ERA-CLIM (Grant 265229). The authors thank the anonymous reviewers for their constructive comments. RTTOVS is software developed within the NWP SAF. The HadCRUT3v dataset is produced by the University of East Anglia; HadAT is provided by the Met Office. MSU comparison data are

produced by Remote Sensing Systems Inc., NESDIS, and University of Huntsville, Alabama, sponsored by the NOAA Climate and Global Change Programme.

#### REFERENCES

- Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol.*, **17**, 25–34.
- Allen, R. J., and S. C. Sherwood, 2008: Warming maximum in the tropical upper troposphere deduced from thermal winds. *Nat. Geosci.*, **1**, 399–403.
- Andrae, U., N. Sökka, and K. Onogi, 2004: The radiosonde temperature bias correction in ERA-40. ERA-40 Project Report Series, Vol. 15, 34 pp.
- Blunden, J., D. S. Arndt, and M. O. Baringer, 2011: State of the Climate in 2010. *Bull. Amer. Meteor. Soc.*, **92**, S1–S266.
- Brohan, P., J. J. Kennedy, I. Harris, Tett S. F. B., and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.
- Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *J. Roy. Stat. Soc.*, **53C**, 405–425.
- Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker, 2003: Error estimates of version 5.0 of MSU–AMSU bulk atmospheric temperatures. *J. Atmos. Oceanic Technol.*, **20**, 613–629.
- Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137A**, 1–28, doi:10.1002/qj.776.
- Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1807.
- Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data. *J. Climate*, **24**, 965–991.
- Dee, D. P., and S. M. Uppala, 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.*, **135**, 1830–1841.
- , P. Poli, and A. J. Simmons, 2011a: Extension of the ERA-Interim reanalysis to 1979. *ECMWF Newsletter*, No. 128, ECMWF, Reading, United Kingdom 7. [Available online at <http://www.ecmwf.int/publications/newsletters/pdf/128.pdf>.]
- , and Coauthors, 2011b: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Della-Marta, P., and H. Wanner, 2006: A method for homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179–4197.
- Douglass, D. H., J. R. Christy, B. D. Pearson, and S. F. Singer, 2008: A comparison of tropical temperature trends with model predictions. *Int. J. Climatol.*, **28**, 1693–1701, doi:10.1002/joc.1651.
- Durre, I., R. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53–68.
- Ebita, A., and Coauthors, 2011: The Japanese 55-year reanalysis “JRA-55”: An interim report. *SOLA*, **7**, 149–152, doi:10.2151/sola.2011-038.
- Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde atmospheric temperature products for assessing climate (RATPAC): A new data set of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi:10.1029/2005JD006169.



- Gaffen, D. J., 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL ARL-211, 37 pp.
- Grant, A., S. Brönnimann, and L. Haimberger, 2008: Recent arctic warming vertical structure contested. *Nature*, **455**, E2–E3, doi:10.1038/nature07257.
- Gruber, C., and L. Haimberger, 2008: On the homogeneity of radiosonde wind time series. *Meteor. Z.*, **17**, 631–643.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403.
- , and U. Andrae, 2011: Radiosonde temperature bias correction in ERA-Interim. ERA Rep. Series, Vol. 8, 17 pp.
- , C. Tavolato, and S. Sperka, 2008: Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data. *J. Climate*, **21**, 4587–4606.
- Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.*, **116**, D14103, doi:10.1029/2010JD015218.
- , —, —, —, and —, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, doi:10.1029/2010JD015220.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Köppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263.
- Ladstätter, F., A. K. Steiner, U. Foelsche, L. Haimberger, C. Tavolato, and G. Kirchengast, 2011: An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes and GPS radio occultation. *Atmos. Meas. Tech.*, **4**, 1965–1977.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate*, **16**, 241–262.
- Lewis, J. M., S. Lakshminarayanan, and S. Dhall, 2005: *Dynamic Data Assimilation*. Cambridge University Press, 816 pp.
- Luers, J. K., and R. E. Eskridge, 1995: Temperature corrections for the VIZ and Vaisala radiosondes. *J. Appl. Meteor.*, **34**, 1241–1253.
- McCarthy, M. P., 2008: Spatial sampling requirements for monitoring upper-air climate change with radiosondes. *Int. J. Climatol.*, **28**, 985–993, doi:10.1002/joc.1611.
- , H. Titchner, P. Thorne, L. Haimberger, and D. E. Parker, 2008: Assessing bias and uncertainty in the HadAT adjusted radiosonde climate record. *J. Climate*, **21**, 817–832.
- Mears, C. A., and F. J. Wentz, 2009: Construction of the Remote Sensing Systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders. *J. Oceanic Atmos. Technol.*, **26**, 1040–1056.
- , —, P. W. Thorne, and D. Bernie, 2011: Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte Carlo estimation technique. *J. Geophys. Res.*, **116**, D08112, doi:10.1029/2010JD014954.
- Menne, M. J., and C. N. J. Williams, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717.
- Nash, J., and F. J. Schmidlin, 1987: WMO international radiosonde intercomparison: Final report. WMO/TD 195, 103 pp.
- Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.*, **24**, 1499–1502.
- Poli, P., S. B. Healy, and D. P. Dee, 2010: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.*, **136**, 1972–1990.
- Rienecker, M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648.
- Rougier, J., and D. M. H. Sexton, 2007: Inference in ensemble experiments. *Philos. Trans. Roy. Soc. London*, **365A**, 2133–2143, doi:10.1098/rsta.2007.2071.
- Santer, B. D., and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556.
- , and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, doi:10.1002/joc.1756.
- Saunders, R., and Coauthors, 2011: RTTOV-10 science and validation report. EUMETSAT, 31 pp. [Available online at [http://research.metoffice.gov.uk/research/interproj/nwpsal/rtm/docs\\_rttov10/rttov10\\_svr\\_1.11.pdf](http://research.metoffice.gov.uk/research/interproj/nwpsal/rtm/docs_rttov10/rttov10_svr_1.11.pdf).]
- Screen, J. A., and I. Simmonds, 2011: Erroneous arctic temperature trends in the ERA-40 reanalysis: A closer look. *J. Climate*, **24**, 2620–2627.
- Seidel, D. J., and Coauthors, 2009: Reference upper-air observations for climate: Rationale, progress, and plans. *Bull. Amer. Meteor. Soc.*, **90**, 361–369.
- Sherwood, S. C., J. Lanzante, and C. Meyer, 2005: Radiosonde daytime biases and late 20th century warming. *Science*, **309**, 1556–1559, doi:10.1126/science.1115640309.
- , C. L. Meyer, R. J. Allen, and H. A. Titchner, 2008: Robust tropospheric warming as revealed by iteratively homogenized radiosonde data. *J. Climate*, **21**, 5336–5352.
- Sperka, S., 2007: Homogeneity adjustments of radiosonde temperature time series using composites of innovations from ERA-40. M.S. thesis, Dept. of Meteorology and Geophysics, University of Vienna, 60 pp. [Available online at [http://www.univie.ac.at/img-wien/dipldiss/dipl/DA\\_Sperka.pdf](http://www.univie.ac.at/img-wien/dipldiss/dipl/DA_Sperka.pdf).]
- Steiner, A. K., B. C. Lackner, F. Ladstätter, B. Scherllin-Pirscher, U. Foelsche, and G. Kirchengast, 2011: GPS radio occultation for climate monitoring and change detection. *Radio Sci.*, **46**, RS0D24, doi:10.1029/2010RS004614.
- Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears, 2005a: Uncertainties in climate trends: Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.*, **86**, 1437–1442.
- , —, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005b: Revisiting radiosonde upper-air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, doi:10.1029/2004JD005753.
- , and Coauthors, 2007: Tropical vertical temperature trends: A real discrepancy? *Geophys. Res. Lett.*, **34**, L16702, doi:10.1029/2007GL029875.
- , and Coauthors, 2011a: A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *J. Geophys. Res.*, **116**, D12116, doi:10.1029/2010JD015487.
- , J. R. Lanzante, and T. C. Peterson, D. J. Seidel, and K. P. Shine, 2011b: Tropospheric temperature trends: History of an ongoing controversy. *WIREs Climate Change*, **2**, 66–88, doi:10.1002/wcc.80.
- Titchner, H., M. McCarthy, P. W. Thorne, S. F. B. Tett, L. Haimberger, and D. E. Parker, 2009: Critically reassessing tropospheric

1 DECEMBER 2012

HAIMBERGER ET AL.

8131

- temperature trends from radiosondes using realistic validation experiments. *J. Climate*, **22**, 465–485.
- Trenberth, K. E., and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis*, Cambridge University Press, 235–336.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Venema, V. K. C., and Coauthors, 2012: Benchmarking monthly homogenization algorithms. *Climate Past*, **8**, 89–115.
- Wallis, T., 1998: A subset of core stations from the Comprehensive Aerological Reference Dataset (CARDS). *J. Climate*, **11**, 272–282.
- Zou, C. Z., M. Gao, and M. D. Goldberg, 2009: Error structure and atmospheric temperature trends in observations from the Microwave Sounding Unit. *J. Climate*, **22**, 1661–1681.

## 4 Comparison of Upper-Air Temperature Data

The efforts of comparing the homogenised radiosonde temperature data sets to other upper air data sets with special emphasis on a comparison to GPS radio occultation data are published together with Florian Ladstädter and co-authors in the following paper in Atmospheric Measurement Techniques.

My contribution to Ladstädter et al. (2011) was providing the homogenised radiosonde data in MSU equivalents for the comparison to the other upper air data sets. This task included the calculation of the layer equivalent temperatures for the four MSU layers as well as the calculation of zonal and global statistics of the different data sets. The provided data was evaluated and compared against satellite data and temperature data from GPS radio occultation in the lower stratosphere.

Ladstädter, F., Steiner, A. K., Foelsche, U., Haimberger, L., Tavorato, C. and Kirchengast, G., 2011: *An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes and GPS radio occultation*. Atmos. Meas. Tech., 4: 1965-1977.



Atmos. Meas. Tech., 4, 1965–1977, 2011  
 www.atmos-meas-tech.net/4/1965/2011/  
 doi:10.5194/amt-4-1965-2011  
 © Author(s) 2011. CC Attribution 3.0 License.



## An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes, and GPS radio occultation

F. Ladstädter<sup>1</sup>, A. K. Steiner<sup>1</sup>, U. Foelsche<sup>1</sup>, L. Haimberger<sup>2</sup>, C. Tavalato<sup>2</sup>, and G. Kirchengast<sup>1</sup>

<sup>1</sup>Wegener Center for Climate and Global Change (WEGC) and Institute for Geophysics, Astrophysics, and Meteorology/Inst. of Physics (IGAM/IP), University of Graz, Graz, Austria

<sup>2</sup>Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

Received: 1 March 2011 – Published in Atmos. Meas. Tech. Discuss.: 7 April 2011

Revised: 15 August 2011 – Accepted: 18 September 2011 – Published: 21 September 2011

**Abstract.** Uncertainties for upper-air trend patterns are still substantial. Observations from the radio occultation (RO) technique offer new opportunities to assess the existing observational records there. Long-term time series are available from radiosondes and from the (Advanced) Microwave Sounding Unit (A)MSU. None of them were originally intended to deliver data for climate applications. Demanding intercalibration and homogenization procedures are required to account for changes in instrumentation and observation techniques. In this comparative study three (A)MSU anomaly time series and two homogenized radiosonde records are compared to RO data from the CHAMP, SAC-C, GRACE-A and F3C missions for September 2001 to December 2010. Differences of monthly anomalies are examined to assess the differences in the datasets due to structural uncertainties. The difference of anomalies of the (A)MSU datasets relative to RO shows a statistically significant trend within about  $(-0.2 \pm 0.1)$  K/10 yr (95 % confidence interval) at all latitudes. This signals a systematic deviation of the two datasets over time. The radiosonde network has known deficiencies in its global coverage, with sparse representation of most of the Southern Hemisphere, the tropics and the oceans. In this study the error that results from sparse sampling is estimated and accounted for by subtracting it from radiosonde and RO datasets. Surprisingly the sampling error correction is also important in the Northern Hemisphere (NH), where the radiosonde network is dense over the continents but does not capture large atmospheric variations in NH winter. Considering the sampling error, the consistency of radiosonde and RO anomalies is improving substantially; the trend in the anomaly differences is generally very small. Regarding (A)MSU, its poor vertical

resolution poses another problem by missing important features of the vertical atmospheric structure. This points to the advantage of homogeneously distributed measurements with high vertical resolution.

### 1 Introduction

The upper troposphere-lower stratosphere (UTLS) region is known to react sensitively to climate change (Baldwin et al., 2007). High-quality observations are crucial to assess the anthropogenic influence on the climate system in the UTLS. It is well known that the temperature trend patterns in the troposphere and stratosphere can provide valuable information on the mechanisms of climate change (Karl et al., 2006; Solomon et al., 2007; Thompson and Solomon, 2005). Until now observational data exist primarily from radiosondes (since 1958) and from the (Advanced) Microwave Sounding Unit (A)MSU instrument flying on US National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites (since 1979). However, none of these existing long-term measurement systems for the upper-air were originally intended to be used for climate monitoring purposes. While surface temperature trends are in accordance amongst different groups (Solomon et al., 2007), the uncertainties regarding trend values for the upper-air are still substantial (Randel et al., 2009; Randall and Herman, 2008; Titchner et al., 2009). The main reasons for these uncertainties derive from demanding intercalibration and homogenization procedures. These *structural uncertainties* have been results of changing instrumentation and observation practice over the decades (Karl et al., 2006; Thorne et al., 2005). This is true for both main sources of upper-air temperature data. The radiosonde time series has specifically experienced numerous changes in their stations, types of sensors, and changes in



Correspondence to: F. Ladstädter  
 (florian.ladstaedter@uni-graz.at)

data processing systems. Using advanced homogenization techniques, these artificial data discontinuities are reduced (Haimberger, 2007; Haimberger et al., 2008). The sparse spatial sampling is causing further uncertainties in the global radiosonde stations' network (Free and Seidel, 2005). Unlike radiosondes, (A)MSU data provide very good global coverage. The instrumentation biases introduced in the chain of NOAA satellites (most recent being NOAA-19) still need to be accounted for. Further errors affecting (A)MSU data include shifts in the diurnal sampling, orbit variations and calibration changes (Karl et al., 2006). Many of these issues are addressed by calibrated datasets produced by different groups (Christy et al., 2007; Mears and Wentz, 2009a; Zou et al., 2009).

There have been significant efforts in the past to create reliable climate records despite these obstacles (Mears and Wentz, 2009a; Christy et al., 2003; Haimberger et al., 2008; Zou and Wang, 2010). It has been argued that the uncertainties in upper-air temperature trends are inevitable due to structural uncertainties involved in the methodology (Thorne et al., 2005). Increasing the number of independent datasets decreases the structural uncertainty (Seidel et al., 2004). The need for new upper-air measurement systems has already been stated by the implementation plan for the Global Observing System for Climate (GCOS, 2010). One already existing relatively new system is GPS radio occultation (RO) that can be considered as of potential benchmark quality (Steiner et al., 2009b). RO uses Global Positioning System (GPS) radio signals in limb sounding geometry to deliver observations in the UTLS region with high accuracy, global coverage, and high vertical resolution (Melbourne et al., 1994; Kursinski et al., 1997; Steiner et al., 2001; Hajj et al., 2002). Additionally it is self-calibrating, thus avoiding error-prone intercalibration procedures. These properties make the technique well qualified to be used for climate applications, as has been shown in a considerable number of publications (e.g., Scherllin-Pirscher et al., 2011b; Steiner et al., 2009b; Foelsche et al., 2009; Ho et al., 2009b; Leroy et al., 2006). Therefore RO can be considered a good choice to assess the adequacy of the observational data mentioned above for climate applications. This has been done in several previous studies for (A)MSU (Schröder et al., 2003; Ho et al., 2007; Steiner et al., 2007, 2009a). Regarding radiosondes, Kuo et al. (2005), He et al. (2009), and Sun et al. (2010) concluded that RO soundings are of sufficient quality to differentiate between different types of radiosondes. Steiner et al. (2007, 2009a), and Ho et al. (2007) found significant differences between RO and (A)MSU records. Ho et al. (2009a) suggested to use RO data for calibration of (A)MSU temperatures.

This study advances previous work (Steiner et al., 2007), using the most recent datasets for RO, (A)MSU and radiosondes, and substantially longer records. It furthermore improves on previous work by analysing error characteristics of RO and radiosondes resulting from sparse spatial and

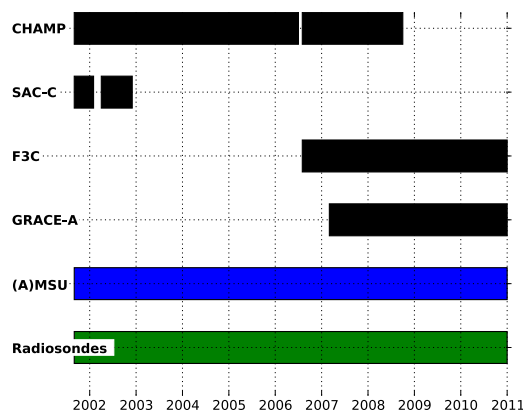


Fig. 1. Time frames of datasets used (black, GPS RO datasets).

temporal sampling. The data used in this study are briefly introduced in Sect. 2, the method of comparison and assessing sampling error characteristics is described in Sect. 3, the results are discussed in Sect. 4, followed by a summary of the results and conclusions of this comparative study.

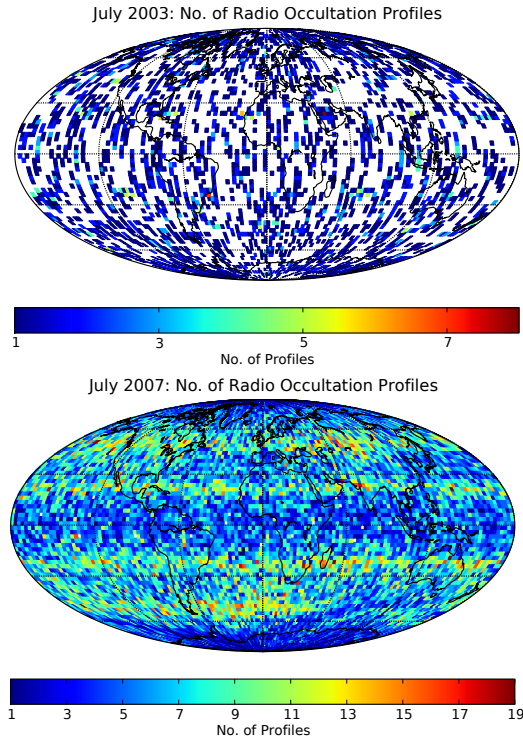
## 2 Data

The comparison time range is limited by the availability of continuous RO data. The CHAMP satellite (Wickert et al., 2001) delivered data from September 2001 to September 2008. Data from the FORMOSAT-3/COSMIC (F3C) mission (Anthes et al., 2008) are used starting from August 2006 until December 2010. Available data from SAC-C (2001, 2002) (Hajj et al., 2004) and GRACE-A (2007 to 2010) (Beyerle et al., 2005) are also used. The study time frame is therefore September 2001 to December 2010 (Fig. 1).

### 2.1 GPS Radio Occultation

We use CHAMP, SAC-C, GRACE-A, and F3C profiles from September 2001 to December 2010 as processed by the Wegener Center for Climate and Global Change (WEGC). We applied the current processing scheme OPSv5.4 (Occultation Processing System, version 5.4) to excess phase data and precise orbit information provided by the University Corporation for Atmospheric Research (UCAR) (Pirscher, 2010). The data of the various instruments can be combined to a consistent single climate record as long as the processing chain is the same for all sources (Pirscher, 2010; Foelsche et al., 2011). Only high-quality profiles are provided and can be downloaded from the global climate monitoring website<sup>1</sup>.

<sup>1</sup>[www.wegcenter.at/globclim](http://www.wegcenter.at/globclim)



**Fig. 2.** Global monthly coverage of RO profiles for July 2003 (top) single-satellite (CHAMP) and for July 2007 (bottom) multi-satellite data (CHAMP, COSMIC, GRACE-A). Number of profiles in  $2.5^\circ \times 2.5^\circ$  bins are shown.

We use dry-temperature profiles in an altitude range of 4 km to 35 km at a vertical resolution of 0.1 km. The RO specific dry-temperature is essentially the same as actual temperature at altitudes above 10 km where moisture is negligible (Scherllin-Pirscher et al., 2011a). Therefore it can be directly used to study the (A)MSU lower stratosphere channel of interest here (Steiner et al., 2007, 2009a). The number of profiles ranges from about 120 to 160 per day (single-satellite) up to about 2000 per day (multi-satellite); see the representative example months in Fig. 2. In both cases the observations are distributed approximately uniformly in space and time within each month.

## 2.2 (Advanced) Microwave Sounding Unit

The (Advanced) Microwave Sounding Unit ([A]MSU) instruments provide satellite-based nadir measurements of layer-average brightness temperatures. The instruments fly on board of the NOAA series of polar orbiting satellites.

We use calibrated post-processed data from three different groups, all of them provided at  $2.5^\circ \times 2.5^\circ$  horizontal resolution. The AMSU instruments are in orbit since 1998, while the last NOAA satellite with an MSU instrument aboard was decommissioned in 2007 (NOAA-14), with decreasing data continuity after 2004. Therefore, during this overlap time contained in the study time frame, the (A)MSU datasets include data from both instrument types.

The bulk temperature of the lower stratosphere region (TLS) corresponds to MSU channel 4 and AMSU channel 9, respectively. These two channels match each other fairly well in order to ensure continuation of the temperature time series. Remaining differences between the MSU and the AMSU instrument are accounted for in the merging procedures of the processing groups so that the time series matches the MSU data (Mears and Wentz, 2009b).

The layer between 150 hPa and 30 hPa ( $\approx 13$  km to 25 km) contributes most to the TLS layer mean temperature, peaking at around 90 hPa ( $\approx 18$  km) (Christy et al., 2003). The poor vertical resolution results in considerable influence of the troposphere to the TLS in the tropics.

TLS brightness temperatures were retrieved from the University of Alabama at Huntsville (UAH) (Christy et al., 2003) in version UAHv5.4;<sup>2</sup> from Remote Sensing Systems (RSS) (Mears and Wentz, 2009a) in version RSSv3.3;<sup>3</sup> and from the National Environmental Satellite, Data and Information Service (NESDIS) Center for Satellite Applications and Research (STAR) (Zou et al., 2009) in version STARv2.0.<sup>4</sup>

## 2.3 Radiosondes

For this comparison, we use the latest homogenized radiosonde datasets: The Radiosonde Observation using Reanalysis (RAOBCORE) dataset (Haimberger, 2007) and the Radiosonde Innovation Composite Homogenization (RICH) dataset (Haimberger et al., 2008), both in version 1.5. Compared to the already published version 1.4, the new version uses raw radiosonde data and background forecast data from ERA-Interim (Dee et al., 2011) instead of operational ECMWF analyses. More than 1000 stations are used. 00:00 UTC and 12:00 UTC launches are kept separately. Figure 3 shows the global coverage of these archives and indicates the launch times. The homogenization procedure works on daily data, which enables very effective breakpoint detection.

RAOBCORE uses time series of the ERA-Interim background forecasts as reference for homogenization. RAOBCORE is therefore, strictly speaking, not independent of satellite data, because ERA-Interim contains (A)MSU information. RICH uses the breakpoints detected by RAOBCORE, but relies only on up to 30 neighboring stations for

<sup>2</sup><http://vortex.nsstc.uah.edu/data/msu/>

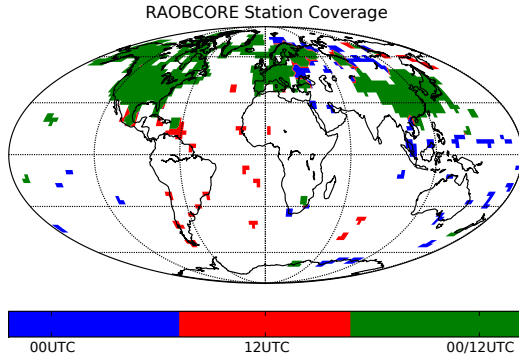
<sup>3</sup>[http://www.remss.com/msu/msu\\_browse.html](http://www.remss.com/msu/msu_browse.html)

<sup>4</sup><ftp://ftp.orbit.nesdis.noaa.gov/pub/smcd/emb/mscat/data/v2.0/>



1968

F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO



**Fig. 3.** Global coverage of radiosonde launches used in the RAOB-CORE and RICH datasets. The color code shows whether there are launches at 00UTC (blue), 12:00 UTC (red), or at both times (green), in the corresponding  $2.5^\circ \times 2.5^\circ$  bin.

the actual homogenization. It is therefore a completely independent dataset (Haimberger et al., 2008).

The radiosonde data are available on 12 pressure levels from 850 hPa to 30 hPa.<sup>5</sup> For both homogenized radiosonde time series, the University of Vienna used the Radiative Transfer for TOVS (RTTOV) model (Saunders, 2008) to construct MSU-equivalent brightness temperature (TLS) anomalies. The TLS values were then binned into a  $2.5^\circ \times 2.5^\circ$  horizontal grid.<sup>6</sup>

#### 2.4 ECMWF

As reference dataset in the estimation of sampling error characteristics of RO and radiosondes (see method description in Sect. 3), we use the ECMWF operational analysis fields. For each RO profile, OPSv5.4 extracts a collocated profile from the global ECMWF field at T42 resolution, comparable to the horizontal resolution of RO (Scherllin-Pirscher et al., 2011b). The analysis fields are available for four time layers, 00:00 UTC, 06:00 UTC, 12:00 UTC, and 18:00 UTC. The 00:00 UTC and 12:00 UTC time layers correspond to the radiosonde launch times and are used in  $2.5^\circ \times 2.5^\circ$  horizontal resolution on 25 pressure levels (from 1 hPa to 1000 hPa) as collocated fields to radiosonde data at station locations. The vertical resolution was increased in 2007, but additional levels after 2007 were introduced below 500 hPa only, which has no effect on the TLS. The averaged field over all time layers is used as reference for the radiosondes and RO, as described in the next section.

<sup>5</sup> 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 250 hPa, 200 hPa, 150 hPa, 100 hPa, 70 hPa, 50 hPa, 30 hPa

<sup>6</sup> <http://www.univie.ac.at/theoret-met/research/raobcore/>

### 3 Method

The different comparisons in this study are based on TLS layer-average brightness temperatures (“MSU-equivalent”). We compare monthly and zonal means for regularly-spaced  $20^\circ$  bands and for four regions, tropics ( $20^\circ$  S to  $20^\circ$  N), extra-tropics ( $70^\circ$  S to  $30^\circ$  S and  $30^\circ$  N to  $70^\circ$  N), and quasi-global ( $70^\circ$  S to  $70^\circ$  N).

#### 3.1 Setup of comparable data

Consistent with the computational procedure applied for radiosonde profiles (see Sect. 2.3), we use RTTOV to compute layer-average TLS from RO and collocated ECMWF temperature profiles. RTTOV uses internally 43 vertical levels from 0.1 hPa to 1013 hPa. The input profiles are interpolated to these levels. To match the horizontal and temporal resolutions of the other datasets, we then bin the resulting TLS field into a  $2.5^\circ \times 2.5^\circ$  grid (monthly means). Averaging involves weighting by the cosine of the latitude, which accounts for area changes between meridians of different latitudes (Foelsche et al., 2008a). This is only a minor effect at this resolution though. We do not distinguish between the various RO missions, all available RO profiles are incorporated into the respective monthly mean. As noted above, this procedure is justified given that the processing chain is the same for all sources (up to negligible differences in raw processing) and that the inter-satellite consistency is thus very high (Foelsche et al., 2011).

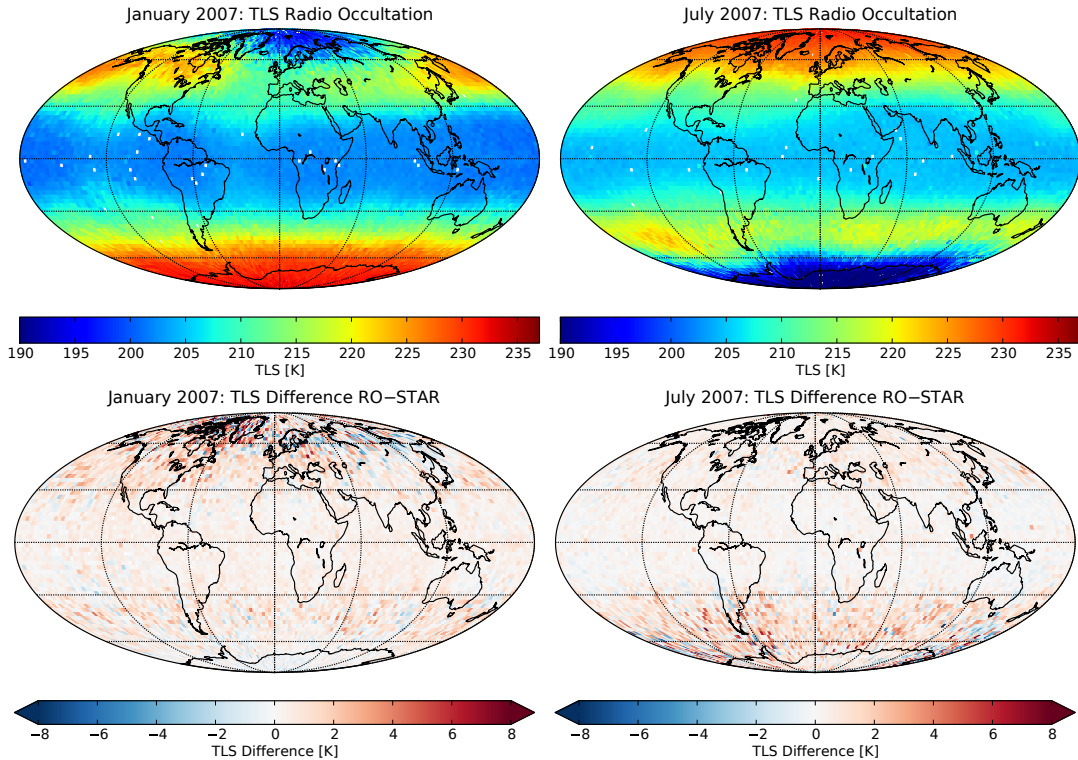
The ECMWF analysis field at  $2.5^\circ \times 2.5^\circ$  resolution is also processed by RTTOV separately for all four available time layers. As a result, all datasets involved in this comparison are now available at the same monthly-means,  $2.5^\circ \times 2.5^\circ$  resolution and in MSU-equivalent TLS. In Fig. 4 we show representative TLS fields for RO and differences of RO to STAR for two months (Northern Hemisphere (NH) winter and summer). TLS temperatures of RO and STAR show larger deviations at higher latitudes, but are generally in very good agreement, especially on a zonal mean scale as used below.

In the next step, we create latitudinal bands by simply averaging over all bins at each respective latitude. Then we aggregate those to larger bands. Here we apply weighting with the surface area of the bands involved. This approach accounts for the decreasing area of latitude bands of equal width (Foelsche et al., 2011).

#### 3.2 Sampling error estimation

All observational datasets inherently differ from reality because of their finite sampling of the atmosphere. Depending on the sampling density and the variability of the atmosphere, it often is essential to account for this difference. A decent approach to estimate the magnitude of error made by discrete sampling is to compare atmospheric fields to a





**Fig. 4.** Brightness temperatures (TLS) for two monthly means at  $2.5^\circ \times 2.5^\circ$  resolution. (left) January 2007, (right) July 2007, (top) Radio occultation synthetic TLS, (bottom) Difference of RO synthetic TLS to AMSU TLS (STAR).

“true” reference field (Foelsche et al., 2008a). In this study, the sampling error estimation for RO and radiosondes is performed consistently. We do not consider sampling error for (A)MSU because we can assume that the error reaches virtually zero due to high horizontal resolution of the dataset. Potential temporal (A)MSU sampling errors caused by diurnal cycle drifts are already accounted for in the homogenized (A)MSU datasets (Christy et al., 1998; Mears and Wentz, 2009b).

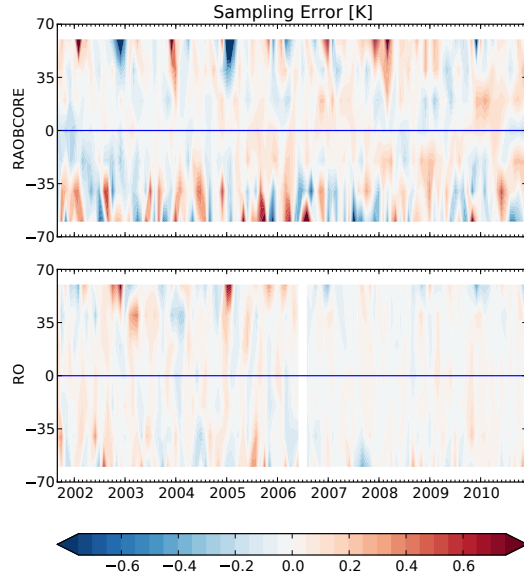
We use ECMWF analysis fields for all four time layers assuming that they are valid approximations of the “true” global field. The methodology for estimating the sampling error of RO is described in detail elsewhere (Pirscher, 2010; Foelsche et al., 2008a). In short, the collocated ECMWF profiles are averaged to latitudinal bands and monthly means as described above. They represent the atmospheric state at the times and locations of RO measurements as seen by the reference field. We then subtract the full reference field, representing the “true” atmospheric state. We define this dif-

ference as *sampling error* of RO for the respective month and latitudinal band. We finally subtract the estimated de-seasonalized sampling error from RO anomalies. This substantially improves the quality of RO climatological fields as has been shown in several studies (Foelsche et al., 2011; Scherllin-Pirscher et al., 2011a). The actual data is thus not used for estimating the sampling error.

In contrast to satellite measurements, the global coverage of radiosondes is not uniform. Most notable, the Southern Hemisphere (SH), the tropics, and the oceans are sparsely represented. In other regions, especially over the NH continents, the coverage is good. Free and Seidel (2005) stated that the concentration of stations in those regions does not necessarily improve the dataset because it oversamples those continental areas while under-representing the oceans. At most of the stations in the SH, radiosonde launches occur only once a day, see Fig. 3. Using an equivalent approach as for RO we estimate the sampling error for radiosondes. We take the ECMWF analysis fields for 00:00 UTC and

1970

F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO



**Fig. 5.** Sampling error of (top) radiosondes and (bottom) RO. Shown are latitudinal bands at 20° resolution.

12:00 UTC separately, and sub-sample the  $2.5^\circ \times 2.5^\circ$  fields to bins where we have radiosonde data for the respective time. This results in a temporally and spatially collocated reference field, analogous to the method above described for RO. After averaging to latitudinal bands we subtract the full reference field containing all four time layers to get the sampling error for radiosondes. Finally we subtract the sampling error from the radiosonde data as we did for RO.

### 3.3 Computation of TLS anomalies and anomaly differences

For RO and (A)MSU data, we calculate monthly TLS anomalies relative to the period 2002 to 2010 to de-seasonalize the data. The radiosonde time series are already provided in anomaly space for the same reference time period. After subtracting the respective de-seasonalized sampling error from RO and radiosonde anomalies (as described above), we compute differences of these anomaly time series. Thereby the climatological variability common to both datasets is removed. Then remaining are the differences due to structural uncertainties. We then compute the linear trends in the anomalies and anomaly differences and their statistical significance to assess deviations between the datasets. In particular, a statistically significant trend of the anomaly differences indicates that both datasets involved behave differently in their time evolution.

**Table 1.** Trends of anomalies for the period of Sep. 2001 to Dec. 2010. The  $\pm$  value defines the 95 % confidence intervals for the trends. Trend values which are significantly different from 0 at the 90 % and 95 % level are marked by a single and double asterisk, respectively.

Dataset	Trend (K/10 years)	StdDev <sub>Residuals</sub> (K)
70° S to 70° N		
RO	$+0.057 \pm 0.133$	0.19
RAOBCORE	$+0.137 \pm 0.128^{**}$	0.19
RICH	$+0.045 \pm 0.130$	0.19
UAH	$-0.140 \pm 0.130^{**}$	0.19
RSS	$-0.149 \pm 0.132^{**}$	0.19
STAR	$-0.162 \pm 0.126^{**}$	0.18
20° S to 20° N		
RO	$+0.351 \pm 0.404^*$	0.59
RAOBCORE	$+0.342 \pm 0.443$	0.64
RICH	$+0.225 \pm 0.439$	0.64
UAH	$+0.137 \pm 0.374$	0.54
RSS	$+0.138 \pm 0.388$	0.56
STAR	$+0.111 \pm 0.364$	0.53
30° N to 70° N		
RO	$+0.222 \pm 0.344$	0.50
RAOBCORE	$+0.237 \pm 0.333$	0.48
RICH	$+0.098 \pm 0.333$	0.48
UAH	$-0.033 \pm 0.330$	0.48
RSS	$-0.028 \pm 0.336$	0.49
STAR	$-0.062 \pm 0.321$	0.47
70° S to 30° S		
RO	$-0.743 \pm 0.451^{**}$	0.66
RAOBCORE	$-0.752 \pm 0.474^{**}$	0.69
RICH	$-0.712 \pm 0.470^{**}$	0.68
UAH	$-0.864 \pm 0.449^{**}$	0.65
RSS	$-0.905 \pm 0.456^{**}$	0.66
STAR	$-0.870 \pm 0.444^{**}$	0.65

## 4 Results

### 4.1 Sampling error

Only by considering the sampling error for both RO and radiosonde records, a consistent comparison is possible. In Fig. 5 the resulting sampling error for radiosondes and RO is shown for 20° zonal bands from 70° S to 70° N.

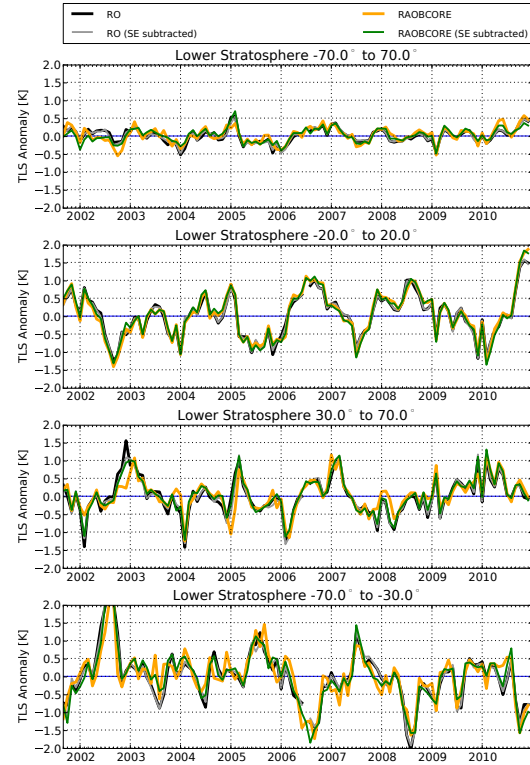
For RO, the sampling error is generally very small ( $< 0.2$  K), except at high latitudes, where it becomes increasingly difficult to capture atmospheric variability (Scherllin-Pirscher et al., 2011a). The additional F3C multi-satellite data as of 2006 provide only moderate reduction of the RO sampling error. For the monthly and zonal means considered in this study, the essential atmospheric variability is already

**Table 2.** Trends of anomaly differences for the period of Sep. 2001 to Dec. 2010. The  $\pm$  value defines the 95 % confidence intervals for the trends. Trend values which are significantly different from 0 at the 90 % and 95 % level are marked by a single and double asterisk, respectively.

Datasets	Trend (K/10 years)	StdDevResiduals(K)
70° S to 70° N		
RAOBCORE-RO	$+0.080 \pm 0.061^{**}$	0.09
RICH-RO	$-0.013 \pm 0.065$	0.10
UAH-RO	$-0.198 \pm 0.042^{**}$	0.06
RSS-RO	$-0.206 \pm 0.043^{**}$	0.06
STAR-RO	$-0.220 \pm 0.045^{**}$	0.07
20° S to 20° N		
RAOBCORE-RO	$-0.012 \pm 0.083$	0.12
RICH-RO	$-0.129 \pm 0.089^{**}$	0.13
UAH-RO	$-0.216 \pm 0.061^{**}$	0.09
RSS-RO	$-0.215 \pm 0.058^{**}$	0.08
STAR-RO	$-0.242 \pm 0.071^{**}$	0.10
30° N to 70° N		
RAOBCORE-RO	$+0.014 \pm 0.061$	0.09
RICH-RO	$-0.125 \pm 0.063^{**}$	0.09
UAH-RO	$-0.256 \pm 0.051^{**}$	0.07
RSS-RO	$-0.252 \pm 0.045^{**}$	0.06
STAR-RO	$-0.286 \pm 0.054^{**}$	0.08
70° S to 30° S		
RAOBCORE-RO	$-0.006 \pm 0.139$	0.20
RICH-RO	$+0.034 \pm 0.139$	0.20
UAH-RO	$-0.118 \pm 0.055^{**}$	0.08
RSS-RO	$-0.159 \pm 0.052^{**}$	0.08
STAR-RO	$-0.124 \pm 0.056^{**}$	0.08

captured by a single satellite (Pirscher et al., 2007; Foelsche et al., 2008b, 2009).

For radiosondes (cf. Fig. 5, top), the sampling error is rather small ( $< 0.3$  K) between about 50° S to 50° N. For higher latitudes the sampling error becomes large. We attribute this to greater variability of the atmosphere at higher latitudes and to the small number of stations in the SH. The sampling density in the tropics is also small but seems to be sufficient to capture the main features of atmospheric variability there. The patterns in southern and northern high latitudes differ substantially: While in the SH temporal evolution of the sampling error seems to be a rather random effect related to sparse sampling, the pattern in the NH shows a clear relation to the NH winter. Every NH winter the sampling error reaches a maximum. Comparing with Fig. 4 (top left), showing the TLS pattern in January, implies that the radiosonde network misses the large atmospheric variability in winter. This results in a larger sampling error.



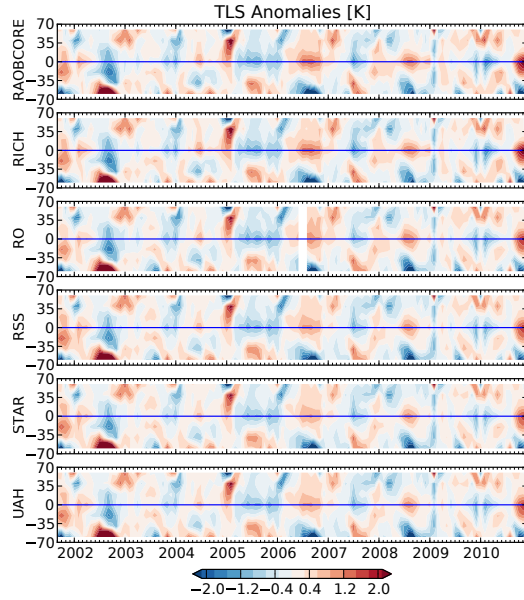
**Fig. 6.** TLS anomalies before/after subtracting the sampling error for RO (black/grey) and for RAOBCORE (orange/green). Shown for quasi-global region, tropics, and for NH/SH extratropics (top to bottom).

Temporal sampling of radiosondes (00:00 UTC and 12:00 UTC) seems to basically capture the diurnal variation up to the semi-diurnal cycle. This was investigated by using only 00:00 UTC and 12:00 UTC time layers of the reference field for calculating the sampling error, instead of the “full” field of four time layers. Comparing the sampling error based on 00:00 UTC and 12:00 UTC time layers with that based on the “full” field showed very small differences only.

The effect of subtracting the respective sampling error from RO and radiosonde anomalies is shown in Fig. 6 for the large-scale zonal bands defined above. It is especially pronounced in NH and SH extratropics. The distinct influence of the sampling error correction in NH winter is clearly visible, as well as the all-year random effect in the SH extratropics. Generally, the radiosonde data get significantly closer to the RO time series after removing the sampling error. In the following, the RO and radiosonde datasets are

1972

F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO

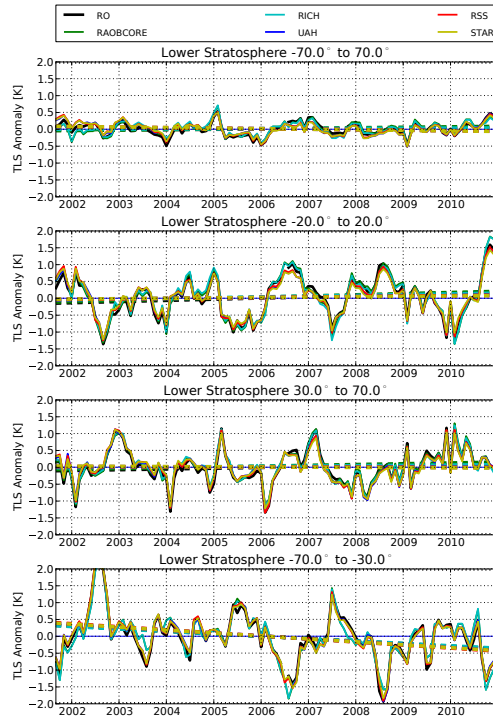


**Fig. 7.** Evolution of TLS anomalies for radiosondes (RAOBCORE, RICH), RO, and (A)MSU (RSS, STAR, UAH) (top to bottom), shown at 20° resolution.

always being used in the corrected form of having their respective sampling errors subtracted. We focus on 70° S to 70° N to avoid sampling problems at polar latitudes.

#### 4.2 TLS anomalies and anomaly differences

The TLS anomalies of all datasets are shown in Fig. 7 at 20° latitudinal resolution for 70° S to 70° N. Overall, the anomaly patterns of the various datasets are consistent. Figure 8 shows TLS anomaly time series for the investigated large-scale zonal bands. The anomalies show good agreement over the whole time range. The anomaly trend values are summarized in Table 1. We observe statistically significant (at 95 % significance level) negative TLS trends in the global mean for all (A)MSU datasets. These negative trends mostly stem from the SH extratropics, where the trend values are in rough agreement with Randel et al. (2009). In the tropics the trend values are positive for the TLS brightness temperature anomalies for all datasets, though statistically not significant (except for RO, showing low significance). This probably is a result from the coarse vertical resolution of TLS MSU-equivalents, where TLS derives from integrating over upper troposphere/lower stratosphere parts of the tropics (Randel et al., 2009). As shown by Schmidt et al. (2010), RO detects a positive trend signal in the tropics around the

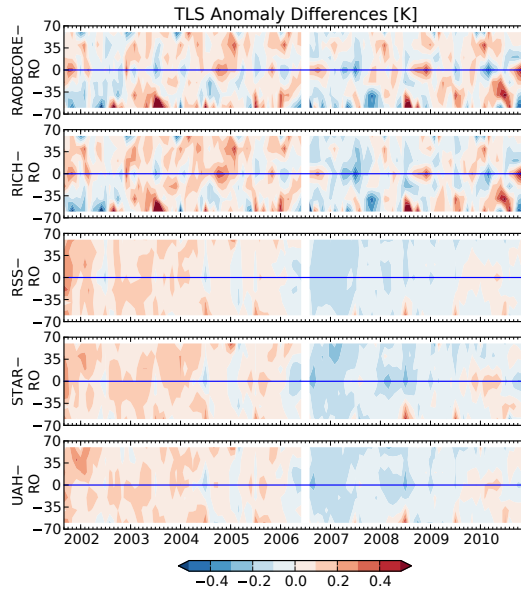


**Fig. 8.** TLS anomaly time series for all datasets, shown for quasi-global, tropical, and NH/SH extratropical zonal bands (top to bottom). The linear regression lines are shown as dashed lines.

tropical tropopause, most probably strongly influencing the integral TLS. We do not further enter here into a climatological interpretation of the trends (which is difficult because of the short time period involved) but focus below on the structural differences of the datasets.

The differences of radiosonde and (A)MSU anomalies to RO anomalies are shown in Fig. 9 at 20° latitudinal resolution and in Fig. 10 for the large-scale zonal regions. The anomaly difference trend values are summarized in Table 2. RAOBCORE and RICH show nearly negligible trends in their difference to RO, ( $0.08 \pm 0.06$ ) K/10 yr and ( $-0.01 \pm 0.07$ ) K/10 yr globally, which indicates that they do not diverge in time relative to RO. A moderate exception of this can be observed for the RICH dataset in the tropics and NH, with difference trend values of ( $-0.13 \pm 0.09$ ) K/10 yr and ( $-0.13 \pm 0.06$ ) K/10 yr. The above mentioned problem of the radiosonde network to correctly capture NH winter atmospheric variations is visible in the NH and quasi-global latitudinal bands. These differences are much more pronounced





**Fig. 9.** Evolution of TLS anomaly differences of radiosonde (RAOBCORE, RICH) and (A)MSU (RSS, STAR, UAH) datasets to RO at 20° resolution (top to bottom).

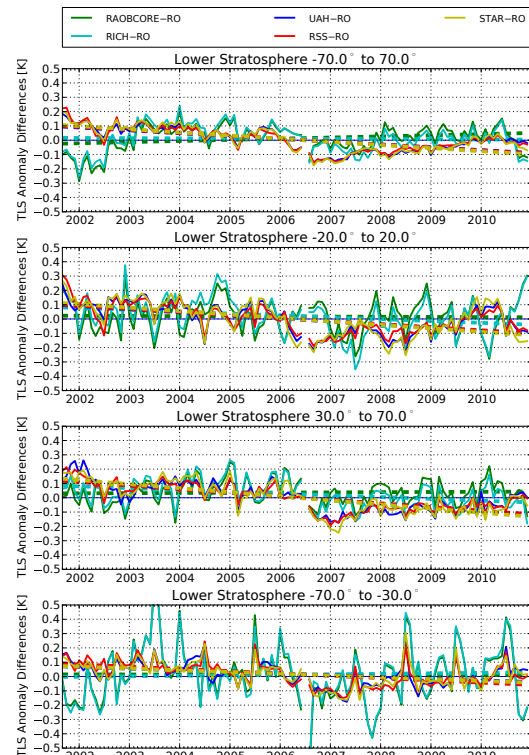
if the radiosonde datasets are not corrected for their sampling error (not shown; cf. Fig. 6).

The TLS anomaly difference trend of (A)MSU relative to RO is within about  $(-0.2 \pm 0.1)$  K/10 yr throughout all latitude ranges, and statistically significant everywhere and for all three (A)MSU datasets (with relatively smallest values of about  $(-0.14 \pm 0.05)$  K/10 yr in the SH and relatively largest ones of about  $(-0.26 \pm 0.05)$  K/10 yr in the NH).

These results are visually summarized in Fig. 11, and also include the respective difference of the radiosonde datasets to a representative (A)MSU dataset (STAR) and the difference of RAOBCORE to RICH, all with their 95 % confidence interval.

## 5 Summary and conclusions

This study focused on comparing (A)MSU data and radiosonde data to radio occultation data, which are well qualified as reference dataset for climate applications. We included RO data from CHAMP, SAC-C, GRACE-A, and F3C satellites for the time period September 2001 to December 2010. All RO profiles were transformed to MSU-equivalent layer-average brightness temperatures (TLS) using a radiative transfer model (RTTOV). Using inter-satellite consistency, the RO data were combined to form a single TLS

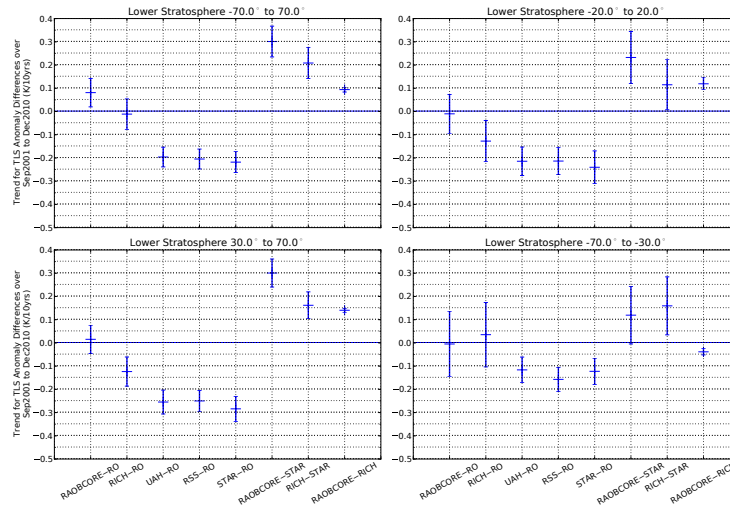


**Fig. 10.** TLS anomaly difference time series for all datasets, shown for quasi-global, tropical, and NH/SH extratropical zonal bands (top to bottom). The linear regression lines are shown as dashed lines.

RO dataset. This dataset was compared to (A)MSU datasets (UAH, RSS, STAR) and recent homogenized radiosonde datasets (RAOBCORE, RICH).

We estimated the spatiotemporal sampling error of radiosonde and RO data. Comparing the RO reference anomalies with radiosonde anomalies, we showed the importance of taking into account these error characteristics also for radiosondes. The consistency of radiosondes and RO was improved substantially by subtracting their respective sampling errors. We thus compared radiosonde and RO datasets in corrected form, i.e., with their sampling errors subtracted. The resulting anomaly time series for TLS showed good agreement of radiosonde data with RO.

Rather surprisingly, we found that it is also important to take into account the sampling error for radiosondes in the Northern Hemisphere (NH) extratropics where radiosonde station coverage is generally very good. We conclude that this results from the radiosonde network missing



**Fig. 11.** Trend values of anomaly differences with 95 % confidence interval for quasi-global, tropics, and NH/SH extratropics regions (top left to bottom right).

the atmospheric variability, particularly in NH winter. The advantage of homogeneously distributed measurements is thus clearly visible. In the tropics the deviations of radiosonde TLS from RO TLS are relatively small. This implies that despite the small number of stations in this region the sampling of radiosondes seems to be sufficient to largely capture the relatively homogeneous atmosphere in the tropics. RAOBCORE showed small to insignificant differences compared to RO in all of the three subregions. RICH also showed insignificant differences in the global mean. While RICH shows more cooling than RO in the tropics and NH, its anomalies still agree clearly better with RO than (A)MSU data. Generally radiosonde data showed larger errors in SH than elsewhere because the station coverage is very sparse there. Trends in TLS anomaly differences of radiosondes compared to RO were found to be small to insignificant in the global mean,  $(0.08 \pm 0.06)$  K/10 yr for RAOBCORE and  $(-0.01 \pm 0.07)$  K/10 yr for RICH.

(A)MSU data do not need sampling error correction because they provide very dense horizontal sampling. We found statistically significant trend values within about  $(-0.2 \pm 0.1)$  K/10 yr for the anomaly differences relative to RO in all large-scale zonal regions for all three (A)MSU datasets. This latitudinally consistent result somewhat deviates from the results of Steiner et al. (2007), who showed significant difference trends mainly in the tropics for the time period 2001 to 2006. The difference trend values are smaller compared to Steiner et al. (2007), which we attribute to the longer time series, the most recent versions of RO and

(A)MSU data, and to a change in the difference trend characteristics of the (A)MSU record in 2006 where the negative trend tendency ceased to continue. The overall conclusion of Steiner et al. (2007) that multiple independent datasets are needed for detecting weaknesses in climate records remains valid.

The good vertical resolution of the RO and radiosonde data (compared to the layer-average TLS of the (A)MSU instrument) will be of advantage to further analyze and understand the differences. We expect the remaining differences to be easiest to explain in the tropics (which we will analyze in a future study). The high quality of RO measurements and the good agreement of radiosonde and RO anomalies indicate that the detected differences mainly stem from the (A)MSU data.

**Acknowledgements.** We are grateful to UCAR/CDAAC (USA) and WEGC (A) GPS RO operational team members for provision of RO data, at WEGC especially to J. Fritzer and B. Scherllin-Pirscher for their contributions to OPS system development and operations. We thank B. C. Lackner for fruitful discussions. UAH, RSS and NESDIS/STAR (all USA) are acknowledged for providing (A)MSU records. Furthermore we thank ECMWF (UK) for access to their atmospheric analysis data. This work was funded by the Austrian Science Fund (FWF; BENCHCLIM project P22293-N21, TRENDEVAL project P21642-N21, C. Tavalato and L. Haimberger supported by project P21772-N10). WEGC OPS development was co-funded by ESA/ESTEC Noordwijk, ESA/ESRIN Frascati, and FFG/ALR Austria.

F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO

1975

Edited by: K. B. Lauritsen

## References

- Anthes, R. A., Ector, D., Hunt, D. C., Kuo, Y.-H., Rocken, C., Schreiner, W. S., Sokolovskiy, S. V., Syndergaard, S., Wee, T.-K., Zeng, Z., Bernhardt, P. A., Dymond, K. F., Chen, Y., Liu, H., Manning, K., Randel, W. J., Trenberth, K. E., Cucurull, L., Healy, S. B., Ho, S.-P., McCormick, C., Meehan, T. K., Thompson, D. C., and Yen, N. L.: The COSMIC/FORMOSAT-3 mission: Early results, *B. Am. Meteorol. Soc.*, 89, 313–333, doi:10.1175/BAMS-89-3-313, 2008.
- Baldwin, M. P., Dameris, M., and Shepherd, T. G.: How will the stratosphere affect climate change?, *Science*, 316, 1576–1577, doi:10.1126/science.1144303, 2007.
- Beyerle, G., Schmidt, T., Michalak, G., Heise, S., Wickert, J., and Reigber, C.: GPS radio occultation with GRACE: Atmospheric profiling utilizing the zero difference technique, *Geophys. Res. Lett.*, 32, L13806, doi:10.1029/2005GL023109, 2005.
- Christy, J. R., Spencer, R. W., and Lobl, E. S.: Analysis of the merging procedure for the MSU daily temperature time series, *J. Climate*, 11, 2016–2041, doi:10.1175/1520-0442-11.8.2016, 1998.
- Christy, J. R., Spencer, R. W., Norris, W. B., Braswell, W. D., and Parker, D. E.: Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures, *J. Atmos. Ocean. Tech.*, 20, 613–629, doi:10.1175/1520-0426(2003)20<613:EEOVOM>2.0.CO;2, 2003.
- Christy, J. R., Norris, W. B., Spencer, R. W., and Hnilo, J. J.: Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements, *J. Geophys. Res.*, 112, D06102, doi:10.1029/2005JD006881, 2007.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hlm, E. V., Isaksen, I., Kilberg, P., Khlér, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thpaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Foelsche, U., Borsche, M., Steiner, A., Gobiet, A., Pirscher, B., Kirchengast, G., Wickert, J., and Schmidt, T.: Observing upper troposphere-lower stratosphere climate with radio occultation data from the CHAMP satellite, *Clim. Dynam.*, 31, 49–65, doi:10.1007/s00382-007-0337-7, 2008a.
- Foelsche, U., Kirchengast, G., Steiner, A. K., Kornblüh, L., Manzini, E., and Bengtsson, L.: An observing system simulation experiment for climate monitoring with GNSS radio occultation data: setup and testbed study, *J. Geophys. Res.*, 113, D11108, doi:10.1029/2007JD009231, 2008b.
- Foelsche, U., Pirscher, B., Borsche, M., Kirchengast, G., and Wickert, J.: Assessing the climate monitoring utility of radio occultation data: From CHAMP to FORMOSAT-3/COSMIC, *Terr. Atmos. Ocean. Sci.*, 20, 155–170, doi:10.3319/TAO.2008.01.14.01(F3C), 2009.
- Foelsche, U., Scherllin-Pirscher, B., Ladstädter, F., Steiner, A. K., and Kirchengast, G.: Refractivity and temperature climate records from multiple radio occultation satellites consistent within 0.05%, *Atmos. Meas. Tech. Discuss.*, 4, 1593–1615, doi:10.5194/amtd-4-1593-2011, 2011.
- Free, M. and Seidel, D. J.: Causes of differing temperature trends in radiosonde upper air data sets, *J. Geophys. Res.*, 110, D07101, doi:10.1029/2004JD005481, 2005.
- GCOS: Implementation plan for the Global Observing System for Climate in support of the UNFCCC, (2010 Update), WMO-TD/No. 1523 GCOS-138 (GOOS-184, GTOS-76), WMO, 2010.
- Haimberger, L.: Homogenization of radiosonde temperature time series using innovation statistics, *J. Climate*, 20, 1377–1403, doi:10.1175/JCLI4050.1, 2007.
- Haimberger, L., Tavolato, C., and Sperka, S.: Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data, *J. Climate*, 21, 4587–4606, doi:10.1175/2008JCLI1929.1, 2008.
- Hajj, G. A., Kursinski, E. R., Romans, L. J., Bertiger, W. I., and Leroy, S. S.: A technical description of atmospheric sounding by GPS occultation, *J. Atmos. Sol-Terr. Phys.*, 64, 451–469, doi:10.1016/S1364-6826(01)00114-6, 2002.
- Hajj, G. A., Ao, C. O., Iijima, B. A., Kuang, D., Kursinski, E. R., Mannucci, A. J., Meehan, T. K., Romans, L. J., de la Torre Juárez, M., and Yunc, T. P.: CHAMP and SAC-C atmospheric occultation results and intercomparisons, *J. Geophys. Res.*, 109, D06109, doi:10.1029/2003JD003909, 2004.
- He, W., Ho, S.-p., Chen, H., Zhou, X., Hunt, D., and Kuo, Y.-H.: Assessment of radiosonde temperature measurements in the upper troposphere and lower stratosphere using COSMIC radio occultation data, *Geophys. Res. Lett.*, 36, L17807, doi:10.1029/2009GL038712, 2009.
- Ho, S.-P., Kuo, Y.-H., Zeng, Z., and Peterson, T.: A comparison of lower stratosphere temperature from microwave measurements with CHAMP GPS RO data, *Geophys. Res. Lett.*, 34, L15701, doi:10.1029/2007GL030202, 2007.
- Ho, S. P., Goldberg, M., Kuo, Y. H., Zou, C. Z., and Schreiner, W.: Calibration of temperature in the lower stratosphere from microwave measurements using COSMIC Radio Occultation data: preliminary results, *Terr. Atmos. Ocean. Sci.*, 20, 87–100, doi:10.3319/TAO.2007.12.06.01(F3C), 2009a.
- Ho, S.-P., Kirchengast, G., Leroy, S., Wickert, J., Mannucci, T., Steiner, A., Hunt, D., Schreiner, W., Sokolovskiy, S., Ao, C., Borsche, M., von Engeln, A., Foelsche, U., Heise, S., Iijima, B., Kuo, Y.-H., Kursinski, E., Lackner, B., Pirscher, B., Ringer, M., Rocken, C., and Schmidt, T.: Estimating the uncertainty of using GPS radio occultation data for climate monitoring: Intercomparison of CHAMP refractivity climate records from 2002 to 2006 from different data centers, *J. Geophys. Res.*, 114, D23107, doi:10.1029/2009JD011969, 2009b.
- Karl, T. R., Hassol, S. J., Miller, C. D., and Murray, W. L. (eds.): Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences, U.S. Climate Change Science Program/Subcommittee on Global Change Research, 2006.
- Kuo, Y.-H., Schreiner, W. S., Wang, J., Rossiter, D. L., and Zhang, Y.: Comparison of GPS radio occultation

www.atmos-meas-tech.net/4/1965/2011/

Atmos. Meas. Tech., 4, 1965–1977, 2011

1976

F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO

- soundings with radiosondes, *Geophys. Res. Lett.*, 32, L05817, doi:10.1029/2004GL021443, 2005.
- Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R.: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, *J. Geophys. Res.*, 102, 23429–23465, 1997.
- Leroy, S., Dykema, J., and Anderson, J.: Climate Benchmarking Using GNSS Occultation, in: *Atmosphere and Climate: Studies by Occultation Methods*, edited by Foelsche, U., Kirchengast, G., and Steiner, A., 287–301, doi:10.1007/3-540-34121-8\_24, Springer Berlin Heidelberg, 2006.
- Mears, C. A. and Wentz, F. J.: Construction of the RSS V3.2 lower-tropospheric temperature dataset from the MSU and AMSU microwave sounders, *J. Atmos. Ocean. Tech.*, 26, 1493–1509, doi:10.1175/2009JTECHA1237.1, 2009a.
- Mears, C. A. and Wentz, F. J.: Construction of the remote sensing systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders, *J. Atmos. Ocean. Tech.*, 26, 1040–1056, doi:10.1175/2008JTECHA1176.1, 2009b.
- Melbourne, W. G., Davis, E. S., Duncan, C. B., Hajj, G. A., Hardy, K. R., Kursinski, E. R., Meehan, T. K., Young, L. E., and Yunck, T. P.: The application of spaceborne GPS to atmospheric limb sounding and global change monitoring, JPL Publication, 94-18, 147, 1994.
- Pirscher, B., Foelsche, U., Lackner, B., and Kirchengast, G.: Local time influence in single-satellite radio occultation climatologies from Sun-synchronous and non-Sun-synchronous satellites, *J. Geophys. Res.*, 112, D11119, doi:10.1029/2006JD007934, 2007.
- Pirscher, B.: Multi-satellite climatologies of fundamental atmospheric variables from radio occultation and their validation (Ph.D. thesis), Sci. Rep. No. 33, Wegener Center Verlag Graz, Austria, 2010.
- Randall, R. M. and Herman, B. M.: Using limited time period trends as a means to determine attribution of discrepancies in microwave sounding unit-derived tropospheric temperature time series, *J. Geophys. Res.*, 113, D05105+, doi:10.1029/2007JD008864, 2008.
- Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillett, N. P., Keckhut, P., Langematz, U., Lin, R., Long, C., Mears, C., Miller, A., Nash, J., Seidel, D. J., Thompson, D. W. J., Wu, F., and Yoden, S.: An update of observed stratospheric temperature trends, *J. Geophys. Res.*, 114, D02107, doi:10.1029/2008JD010421, 2009.
- Saunders, R.: RTTOV-9 Science and Validation Report, NWP SAF NWPSAF-MO-TV-020, EUMETSAT, 2008.
- Scherllin-Pirscher, B., Kirchengast, G., Steiner, A. K., Kuo, Y.-H., and Foelsche, U.: Quantifying uncertainty in climatological fields from GPS radio occultation: an empirical-analytical error model, *Atmos. Meas. Tech. Discuss.*, 4, 2749–2788, doi:10.5194/amtd-4-2749-2011, 2011.
- Scherllin-Pirscher, B., Steiner, A., Kirchengast, G., Kuo, Y.-H., and Foelsche, U.: Empirical analysis and modeling of errors of atmospheric profiles from GPS radio occultation, *Atmos. Meas. Tech. Discuss.*, 4, 2599–2633, doi:10.5194/amtd-4-2599-2011, 2011b.
- Schmidt, T., Wickert, J., and Haser, A.: Variability of the upper troposphere and lower stratosphere observed with GPS radio occultation bending angles and temperatures, *J. Adv. Space Res.*, 46, 150–161, doi:10.1016/j.asr.2010.01.021, 2010.
- Schröder, T., Leroy, S., Stendel, M., and Kaas, E.: Validating the microwave sounding unit stratospheric record using GPS occultation, *Geophys. Res. Lett.*, 30, 1734, doi:10.1029/2003GL017588, 2003.
- Seidel, D. J., Angell, J. K., Christy, J., Free, M., Klein, S. A., Lanzante, J. R., Mears, C., Parker, D., Schabel, M., Spencer, R., Sterin, A., Thorne, P., and Wentz, F.: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, *J. Climate*, 17, 2225–2240, doi:10.1175/1520-0442(2004)017<2225:UISOLC>2.0.CO;2, 2004.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., eds.: *IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Steiner, A., Kirchengast, K., Foelsche, U., Kornblüeh, L., Manzini, E., and Bengtsson, L.: GNSS occultation sounding for climate monitoring, *Phys. Chem. Earth*, 26, 113–124, doi:10.1016/S1464-1895(01)00034-5, 2001.
- Steiner, A., Kirchengast, G., Borsche, M., Foelsche, U., and Schoengassner, T.: A multi-year comparison of lower stratospheric temperatures from CHAMP radio occultation data with MSU/AMSU records, *J. Geophys. Res.*, 112, D22110, doi:10.1029/2006JD008283, 2007.
- Steiner, A., Kirchengast, G., Borsche, M., and Foelsche, U.: Lower stratospheric temperatures from CHAMP RO compared to MSU/AMSU records: An analysis of error sources, in: *New Horizons in Occultation Research: Studies in Atmosphere and Climate*, edited by: Steiner, A., Pirscher, B., Foelsche, U., and Kirchengast, G., Springer, Berlin Heidelberg, doi:10.1007/978-3-642-00321-9\_18, 2009a.
- Steiner, A., Kirchengast, G., Lackner, B., Pirscher, B., Borsche, M., and Foelsche, U.: Atmospheric temperature change detection with GPS radio occultation 1995 to 2008, *Geophys. Res. Lett.*, 36, L18702, doi:10.1029/2009GL039777, 2009b.
- Sun, B., Reale, A., Seidel, D. J., and Hunt, D. C.: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, *J. Geophys. Res.*, 115, D23104, doi:10.1029/2010JD014457, 2010.
- Thompson, D. W. J. and Solomon, S.: Recent stratospheric climate trends as evidenced in radiosonde data: Global structure and tropospheric linkages, *J. Climate*, 18, 4785–4795, doi:10.1175/JCLI3585.1, 2005.
- Thorne, P. W., Parker, D. E., Christy, J. R., and Mears, C. A.: Uncertainties in climate trends: Lessons from upper-air temperature records, *B. Am. Meteorol. Soc.*, 86, 1437–1442, doi:10.1175/BAMS-86-10-1437, 2005.
- Titchner, H. A., Thorne, P. W., McCarthy, M. P., Tett, S. F. B., Haimberger, L., and Parker, D. E.: Critically reassessing tropospheric temperature trends from radiosondes using realistic validation experiments, *J. Climate*, 22, 465–485, doi:10.1175/2008JCLI2419.1, 2009.
- Wickert, J., Reigber, C., Beyerle, G., König, R., Marquardt, C., Schmidt, T., Grunwaldt, L., Galas, R., Meehan, T. K., Melbourne, W. G., and Hocke, K.: Atmosphere sounding by GPS radio occultation: First results from CHAMP, *Geophys. Res. Lett.*,



- F. Ladstädter et al.: Assessing temperature record differences of (A)MSU, radiosondes, and GPS-RO 1977  
28, 3263–3266, doi:10.1029/2001GL013117, 2001.
- Zou, C.-Z. and Wang, W.: Stability of the MSU-derived atmospheric temperature trend, *J. Atmos. Ocean. Tech.*, 27, 1960–1971, doi:10.1175/2009JTECHA1333.1, 2010.
- Zou, C.-Z., Gao, M., and Goldberg, M. D.: Error structure and atmospheric temperature trends in observations from the Microwave Sounding Unit, *J. Climate*, 22, 1661–1681, doi:10.1175/2008JCLI2233.1, 2009.



## 5 Variational Bias Correction

The work on building a Variational Bias Correction (VarBC) module for radiosonde wind direction biases is documented in the following report. It describes the first steps of building a VarBC module in the ECMWF forecasting system. This report is planned to be published as an ECMWF research department document.

As part of a research project my work was to adapt the existing VarBC system for satellite data and apply it to radiosonde wind direction data as a first step in testing VarBC for conventional data. As author of the report I built the system with help from ECMWF and tested first case studies. These first steps are described in the following report.

Tavolato-Wötzl, C. 2015: *Variational Bias Correction for Radiosonde Wind Direction*, technical report, ECMWF.



**VARIATIONAL BIAS CORRECTION FOR RADIOSONDE WIND DIRECTION**

ECMWF technical RD report  
Author: Christina Tavalato-Wötzl  
Published in: 2015

---

**Abstract**

Variational bias correction is a modern tool to handle observation biases in data assimilation. It was originally developed for satellite data and has just recently been introduced to conventional observations. This report discusses the use of variational bias correction for radiosonde wind data - with a first attempt on radiosonde wind direction data. Known biases can be located during the variational analysis and the quality control is done online while assimilating the data. This system should prove useful to operational data assimilation as well as reanalysis efforts.

**1 Introduction**

Variational bias correction (VarBC) was introduced at the European Centre of Medium-Range Weather Forecast (ECMWF) for satellite data (Dee and Uppala, 2009) in the operational forecast system and in the reanalysis system for the production of ERA-Interim (Dee et al., 2011). During the last years more efforts have been made to also include bias correction for conventional data into the system. VarBC is already applied to aircraft temperature (Isaksen et al., 2012) and there are efforts for a similar system to adjust surface pressure data (Poli et al., 2013).

With VarBC a well known bias dependence (e.g. constant biases for an observation location, dependence on the scan angle, dependence on the solar elevation at observation time, ...) can be described and adjusted within the four dimensional variational (4D-Var) assimilation system.

This report will describe the introduction of VarBC for radiosonde and pilot balloon wind direction data in the assimilation system at the ECMWF. Such a bias correction is particularly important in the context of climate reanalyses, because wind direction biases were widespread in the mid-1950s. They occur, however, also in more recent times (Ramella Pralungo et al., 2014; Ramella Pralungo and Haimberger, 2014; Hollingsworth et al., 1986; Gruber and Haimberger, 2008). A short introduction to VarBC in section 2 is followed by details on the implementation of VarBC for radiosonde wind direction data in section 3. Section 4 shows first results before section 5 summarizes and concludes.

**2 The idea behind Variational Bias Correction**

In variational data assimilation the prototype bias-blind cost function can be written as:

$$J(\vec{x}) = (\vec{x} - \vec{x}_b)^T B^{-1} (\vec{x} - \vec{x}_b) + (\vec{y} - H(\vec{x}))^T R^{-1} (\vec{y} - H(\vec{x})) \quad (2.1)$$

where the cost function uses the common notation with  $\vec{x}$  as the model state,  $\vec{y}$  as the observations,  $H$  is the observation operator and  $B$  and  $R$  are the background and model error covariance matrixes (Andersson and Järvinen, 1999).

With VarBC a bias term is introduced to the cost function of the assimilation system. In the simplest form the bias term in the assimilation consists of a constant bias parameter. Further dependencies of the bias can be described by using a set of bias parameters (they always include the constant bias parameter and then add parameters depending on observed quantities such as the solar elevation at the time of measurement for example). Each bias correction is estimated

for a group of observations. Satellite data groups tend to be formed by satellite and satellite channel (therefore all observations of the same satellite channel measured from the same satellite will receive the same bias correction). The grouping is one of the crucial parts when setting up the system and needs to be done wisely. Grouping data from different populations need to be strictly avoided. Grouping for radiosonde data will be discussed in detail in section 3.2.

Expanding the cost function minimized in the assimilation with the bias term leads to the following VarBC cost function:

$$J(\vec{x}_{\vec{\beta}}) = (\vec{x} - \vec{x}_b)^T B^{-1} (\vec{x} - \vec{x}_b) + \underbrace{(\vec{\beta}_b - \vec{\beta})^T B_{\beta}^{-1} (\vec{\beta}_b - \vec{\beta})}_{\text{bias term}} + \underbrace{(\vec{y} - H(\vec{x}) - \underbrace{\vec{b}(\vec{x}, \vec{\beta})}_{\text{bias term}})^T R^{-1} (\vec{y} - H(\vec{x}) - \underbrace{\vec{b}(\vec{x}, \vec{\beta})}_{\text{bias term}})}_{\text{bias term}} \quad (2.2)$$

In the expanded cost function the bias parameter  $\vec{b}$  is described by a linear model as:

$$\vec{b}(\vec{x}, \vec{\beta}) = \sum_{i=0}^{N_p} \beta_i \vec{p}_i(\vec{x}) \quad (2.3)$$

using  $p_i$  for the predictors and  $\beta_i$  for the unknown bias parameters.

The bias can therefore be dependent on a set number of predictors with the simplest option of a constant bias parameter. Other possible predictors could be the solar angle at observation time, satellite settings, etc. as mentioned above. Each bias parameter is addressed to one chosen predictor but the bias parameters are not dependent on them.

While minimizing the cost function the bias parameters  $\vec{\beta}$  get adjusted to fit the model in the best possible way therefore resulting in a bias correction for each specified observation group.

### 3 Radiosonde wind direction bias correction

When looking at conventional data the main changes in the system will be in the grouping process. While satellites gain multiple observations at different locations with the same observing system over a long time period, certain conventional observations tend to be at a fixed location. In the case of radiosonde observations the observing system (the radiosonde) in principle changes each launch - although it is reasonable to assume that the instrument setup will be constant over time when the same radiosonde type is used. The grouping depends not only on the kind of the observation system but also on the expected bias. Some biases will be dependent on the quality of the observing instrument (the radiosonde type) and therefore grouping by observation type (if the information is available) or country (as within a country mostly the same observation type is used - this is especially an option when the observation type is not documented) can be sensible. For other biases that depend more on the setup of the observation system than on the observation itself there is no need to have more than a single radiosonde profile in a group. So careful planned grouping will help when implementing the bias correction to achieve the best possible results.

#### 3.1 Motivation for wind direction bias correction

After successfully adjusting satellite biases at ECMWF with VarBC (Dee and Uppala, 2009) the search for suitable data sets for first attempts with in situ data started. During the homogenization project RAOBCORE (Haimberger, 2007) first analysis of radiosonde wind data showed abrupt changes in the difference between the observations and the reanalysis data set at certain stations in the wind direction time series (Gruber and Haimberger, 2008). As an example at the station Marion Island in the Indian Ocean (WMO-station-id: 68994) a change in wind direction was found in 1978 that introduced a 10 degree direction bias throughout the whole profile and a

similar change in the opposite direction was found again in 1982 (not shown here - presented in Gruber and Haimberger, 2008). Those biases are constant for a whole radiosonde wind profile and therefore seem to be unnatural and introduced by mistakes in the north alignment of the radiosonde station.

These kind of artificial biases are well suited for VarBC since they are constant throughout a whole radiosonde profile and do not depend on the general circulation. Therefore these biases will only need a simple VarBC setting with one constant VarBC parameter for each radiosonde profile.

The first analysis by Gruber and Haimberger (2008) identified that there are several stations where such a bias could be detected. Those stations are wide spread and an improvement of adjusting this bias will be gained especially from isolated stations such as islands or groups of biased stations (which were found in Asia). An even more severe case of pervasive wind direction biases has been documented by Ramella Pralungo and Haimberger (2014).

### 3.2 Grouping

As mentioned before grouping the data for VarBC is essential and depends on the observation system as well as on the bias. With radiosonde wind direction a constant bias throughout a whole radiosonde profile is expected and therefore the whole profile needs to be in one group to allow the same bias correction for the whole profile. Furthermore since the bias does not depend on the sonde type or country but is caused by a wrong station setup there is no need to group different radiosonde stations together. Therefore each radiosonde station profile is its own group for this specific kind of bias.

### 3.3 Implementation

Implementation of a VarBC for radiosonde wind direction bias only used one bias parameter to describe the expected bias. This constant wind direction bias can be described using a single predictor for the whole radiosonde profile. This reduces the bias term for one station (introduced in equation 2.3) to the following:

$$\vec{b}(\vec{x}, \vec{\beta}) = \beta_o \quad (3.4)$$

Therefore for each station the linear parameter  $\beta_o$  (a wind direction bias) has to be independently estimated. No covariance between bias parameters at different stations is expected.

Next to the bias parameter some other numbers are set to control the bias correction. Values defining a minimum of observations that is needed to start the bias correction scheme, the way of treating the first observations when introducing the system and a maximum departure value for observations to be considered by the system are set. These settings are done in a VarBC module in the code for details see appendix A.

One of the more difficult things during the implementation was the fact that wind direction is not a model variable - only the wind components  $u$  and  $v$  are. Therefore a transformation into wind direction had to be implemented before calculating the bias and the result from the bias model needs to be transformed back into wind components. A transformed bias term for both wind components had to be calculated and its tangent linear and adjoint had to be added to the system.

#### 3.3.1 Transformation from wind components to wind direction

In the forecast model the wind vector is described with its two components  $u$  and  $v$  rather than through wind direction  $dd$  and wind speed  $f$ . However observations detected a possible synthetic

wind bias in wind direction rather than in its components as described in section 3. Therefore the model values have to be transformed to apply the bias correction.

$$\text{for } v < 0 : \quad dd = 270 - \left( \arctan \left[ \frac{u}{v} \right] \right) \cdot \frac{180}{\pi} \quad (3.5)$$

$$\text{for } v \geq 0 : \quad dd = 90 - \left( \arctan \left[ \frac{u}{v} \right] \right) \cdot \frac{180}{\pi} \quad (3.6)$$

Firstly the wind components are used to calculate a wind direction departure which is used for the variational bias correction. A linear bias term in wind direction is then transformed backwards to generate bias increments for both wind components. This is done by applying a rotation matrix to the bias in wind direction.

$$u_{corr} = u \cdot \cos(270 - dd_{corr} \cdot \frac{\pi}{180}) - v \cdot \sin(270 - dd_{corr} \cdot \frac{\pi}{180}) \quad (3.7)$$

$$v_{corr} = u \cdot \sin(270 - dd_{corr} \cdot \frac{\pi}{180}) + v \cdot \cos(270 - dd_{corr} \cdot \frac{\pi}{180}) \quad (3.8)$$

This leads to a non-linear bias correction for both wind components. This might cause underestimation in the bias correction when the transformation for the rotation is used rather than its linearization. For a closer evaluation of this problem we take a look at a simple rotation matrix  $R$ :

A general transformation matrix  $R$  for such a rotation would be:

$$R = \begin{pmatrix} \cos - 1 & -\sin \\ \sin & \cos - 1 \end{pmatrix} \quad (3.9)$$

This rotation of the evaluated bias  $\beta$  is then multiplied with the wind components to get the bias for each component:

$$B(\beta) = R(\beta) \cdot \vec{v} \quad \text{using} \quad \vec{v} = \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.10)$$

This however transforms a linear bias in wind direction into a non-linear bias for the wind components which are the model variables. Therefore one has to be cautious and might use the linearization of the rotation rather than the rotation itself.

The linearization of this rotation leads to the following:

$$R_L = \begin{pmatrix} -\sin & -\cos \\ \cos & -\sin \end{pmatrix} \quad (3.11)$$

Using the rotation  $R$  rather than its linearization  $R_L$  in the minimization should be fine with small wind direction biases but could cause an underestimation of the bias in the model variables. An approximation of  $R$  and  $R_L$  shows that using  $R$  can lead to a factor of  $\sin(\beta)$  with small wind direction biases. Using some basic mathematics and an approximation of small  $\beta$  leads to:

$$\begin{aligned} \cos^2 + \sin^2 &= 1 \\ \cos^2 - 1 &= -\sin^2 \\ (\cos + 1)(\cos - 1) &= -\sin^2 \\ \cos - 1 &= \frac{-\sin^2}{\cos + 1} \\ \cos - 1 &\approx \frac{-\sin^2}{2} \quad \text{using} \quad \cos\beta \approx 1 \end{aligned}$$



This leads to a approximation of the rotation which looks like:

$$R \approx \sin \begin{pmatrix} \frac{-\sin}{2} & -1 \\ 1 & \frac{-\sin}{2} \end{pmatrix} \quad (3.12)$$

Whereas an approximation of the linearization leads to the following (for a small bias  $\beta$ ):

$$R_L \approx \begin{pmatrix} -\sin & -1 \\ 1 & -\sin \end{pmatrix} \quad (3.13)$$

In the current system the rotation rather than the linearization is used. This might lead to a slower convergence of the bias correction than expected and will need to be investigated further when applying this bias correction to data in the future as stated above. Preconditioning or rescaling the bias might be other options to assess the problem of a slow convergence in the future (H. Hersbach personal communication). A close look on the currently used transformation in the model code is given in appendix B.

### 3.3.2 Changes in the IFS code

First a VarBC module similar to the already existing VarBC modules was built for radiosonde wind data (a copy of this module can be found in appendix A). The module called *varbc.rsw.F90* is introduced in the same way as the other VarBC modules to the system. It includes the setup and grouping for the added bias correction. A call to this module was added in the VarBC setup routine and the number of predictors was defined. SQL requests that identify data for this bias correction are introduced and the list of bias parameters is set to include only the first (constant) bias parameter.

Besides the made VarBC module other changes had to be made in the routines describing the observation operator of the cost function. The bias term had to be introduced when handling radiosonde wind data and after the minimisation the resulting bias term needs to be added to the previous value so that the bias is accounted for. Next to the bias a VarBC index is stored in the ODB (Observational Data Base) to indicate observations that belong to the same group. In following cycles this previously stored VarBC index is read and the previous bias estimates are used as a first guess for the new bias correction.

Due to the transformation from the wind components to wind direction also the adjoint and tangent linear of the observation operator needed to be adjusted to provide wind direction bias. This bias finally has to be retransformed to bias increments for both wind components.

A complete list of the changed and new routines and the newly developed module for the radiosonde wind direction bias correction can be found in the appendix. Where appendix A introduces the new VarBC module and appendix B describes the changes in existing routines.

## 4 First Results

Experiments were carried out in cycle 38 release 2 (CY38R2) and run over a series of months during the summer 2012. Next to long term experiments with this specific experimental setting a regular experiment run with the VarBC module turned off was run alongside. Experiments with different parameter settings were also performed.

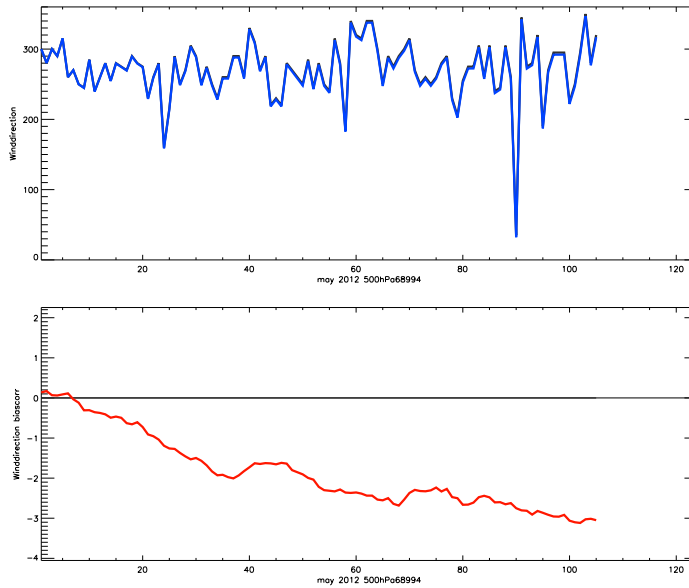
After analyzing the results of the different experiments the main evaluation concentrated on two radiosonde stations:

- Marion Island, South Africa - 68994 - as one of the test stations were an artificial bias of 15 degrees was introduced. This station also gives valuable information since it is located on a remote island.

- Vienna, Austria - 11035 - as a control station which is located in central Europe surrounded with a lot of other observations and has not changed during the experiments

#### 4.1 Long term experiments

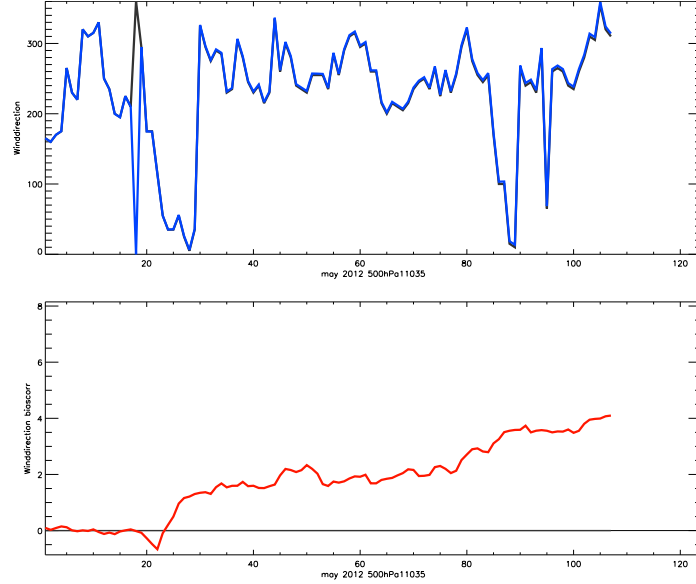
Long term experiments were set up to start from May 2012 to August 2012. Due to a change in the ECMWF model only runs until the 15th of August 2012 were possible. This allows to evaluate time series of around 100 days. A control run (with the VarBC module turned off) was run with the same settings and resolution over the same period for comparison. The resolution was T159 (the same resolution as used in ERA-40 reanalysis).



**Figure 4.1:** Marion Island long term test over 100 days: Upper panel wind direction time series of the experimental and the long run (black and blue) - they show almost similar curves. Lower panel: bias correction applied by the model. After an initial shift of two degrees the bias correction only includes little variations. Accounting the  $360^\circ$  scaling in the upper panel a shift of two degrees is not visible therefore the two curves in this panel cannot be distinguished.

The evaluations show wind direction in the 500hPa level and the applied bias correction (which is constant throughout the whole wind profile by definition). The upper panel of Figure 4.1 shows that at Marion Island an overall constant westerly wind direction occurs during the evaluated period. Even though there is no bias expected the VarBC gradually adjusts the data to a bias of 3 degrees in wind direction during the first half of the period. This is visible in the lower panel of Figure 4.1. This bias only shows little variations which leads to the conclusion that this is the bias of the station. These variations could be due to the fact that radiosonde wind direction is reported in 5 degree steps.

Compared to this isolated island station we take a closer look at Vienna (Figure 4.2) where a different situation was found. Here there is barely any bias correction for the first 20 days thereafter a correction of 2 degrees is applied. After 80 days we see another shift in the bias correction to 4 degrees where the correction flattens at the end of the 100 day period. Again mainly westerly winds were observed even though the wind direction changes from time to time



**Figure 4.2:** Vienna long term test: same as Figure 4.1. No difference in the wind direction time series is clearly identifiable. The bias shows that this station gets adjusted twice over the period of 100 days.

to a easterly to northeasterly direction.

#### 4.1.1 Introducing an artificial wind direction bias

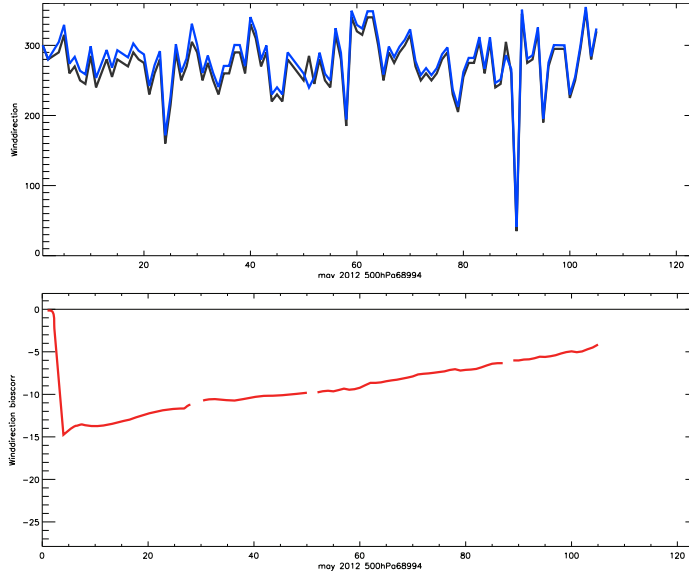
As first step after introducing the system a test setup for an experiment with an artificial introduced large bias at 4 radiosonde stations was built (details to the introduction of the artificial bias can be found in appendix B). 4 different radiosonde stations got an additional wind direction bias of 15 degrees throughout the whole profile from the beginning of the experimental setting onwards. The selected stations were:

- Bethel, Alaska, USA - 70219
- Athens, Greece - 16716
- Marion Island, South Africa - 68994
- Aktyubinsk, Russia - 35229

Once again the evaluation will concentrate on Marion Island and Vienna (as a control station which was not artificially modified during the test). The artificial bias of 15 degrees was introduced on the 4th day of observation so the main time period is with the additional wind direction bias.

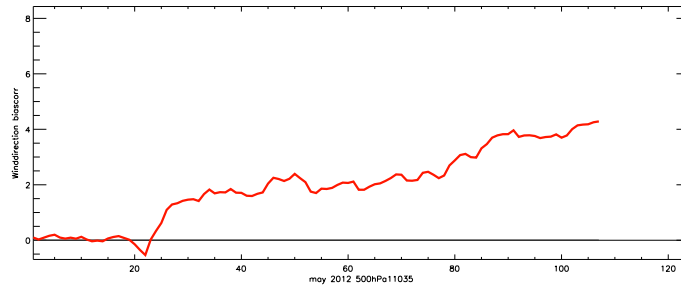
Compared to Figure 4.1 in the upper panel of Figure 4.3 the artificially introduced wind bias is clearly visible compared to the wind direction time series of the control run. The lower panel shows the bias correction in relation to the artificial bias were after the introduction on day 4 the system starts to adjust the bias and within the 100 day experimental setup it resolves more than 10 degrees.

With artificial changes in the model it is important to see if there are any changes in other radiosonde observations. Therefore the bias time series of Vienna is compared to the one from



**Figure 4.3:** Marion Island long term test with artificial 15 degree bias from day 4 onwards. Upper panel: wind direction time series: black curve shows the control run, blue curve the experimental run. Lower panel: The applied bias correction in relation to the artificial introduced bias. The gaps in the bias correction time series are due to data lost by quality checks performed before the bias correction.

the long term experiment without an artificial bias. To compare Figure 4.4 shows the same curve as the lower panel of Figure 4.2. No differences are visible therefore the artificial introduced bias at certain stations does not interfere with other observations. This is very sensible and necessary for a stable system.



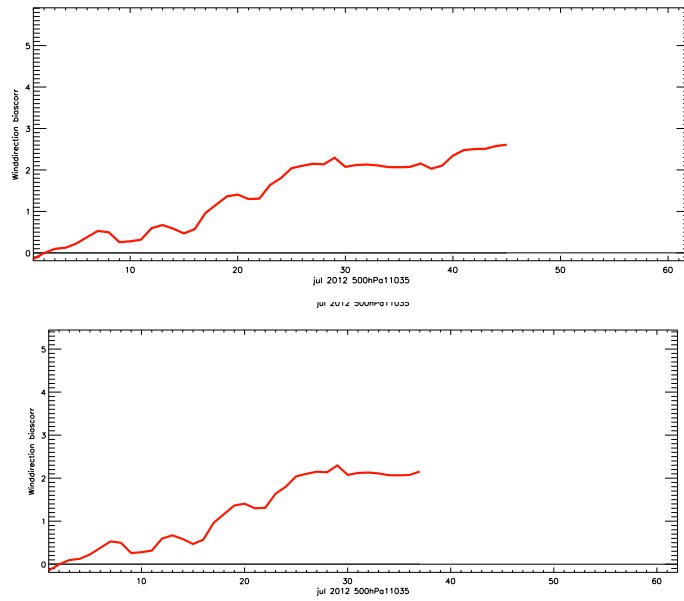
**Figure 4.4:** Vienna long term test in the experimental run with 15 degree artificial bias on certain stations. No clear change is visible in unchanged stations such as Vienna.

## 4.2 Adjustment of the settings

To test the settings of the VarBC there is one key number that can be adjusted. One of the variables in the system determines how many observations are needed before VarBC starts operating. In this sensitivity tests the number used varied from 10 to 100. The sensitivity tests were carried

out over a period of 45 days from July 2012 to mid August 2012 in the same settings as the long term experiments (see section 4.1).

The results are presented in Figure 4.5 and show that in 2012 radiosonde profiles already are stable enough and observe plenty of data within one profile so barely any difference could be observed changing the values. After the 45 days of the test run exactly the same bias was adjusted. With a state of the art radiosonde during the first soundings enough data is collected that the difference in settings is not visible and both runs lead to the same bias correction.



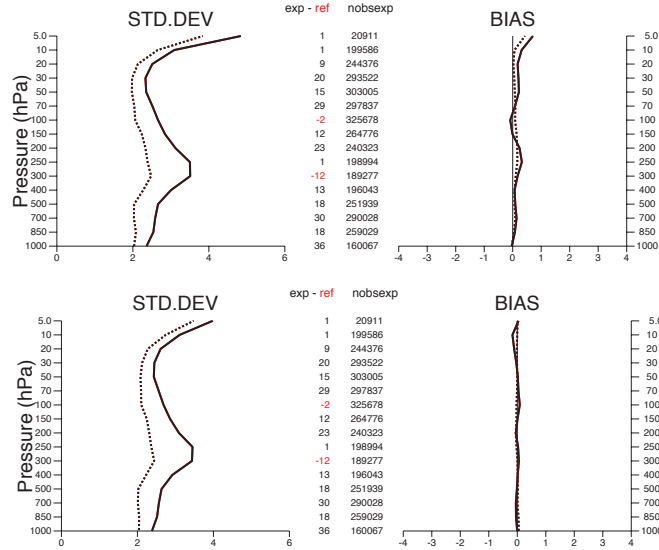
**Figure 4.5:** Sensitivity tests for Vienna. Upper panel: Sensitivity test bias correction in place with 10 observations per group. Lower panel: Sensitivity test bias correction in place with 100 observations per group.

### 4.3 Global Statistics

To evaluate the impact of this bias correction on a wider scale statistics of used radiosonde data were plotted for both wind components and three different areas: Northern Hemisphere (NH), Tropics (Tr) and the Southern Hemisphere (SH). The statistics show the standard deviation and the bias of the used parameter (in this case U and V wind components) as well as the data count of the used data in the experiment and the difference in data count between the experiment and a control run. Two different scenarios were evaluated. In both cases the 100 day period (as described at the beginning of section 4.1) was evaluated.

- A straight forward comparison of the operational run (as control run) to an experiment where the VarBC for radiosonde wind direction was turned on.
- A comparison of two runs with the artificial bias added at 4 stations (see section 4.1.1). One run with and one run without VarBC for radiosonde wind direction.

The main focus of this evaluation is on the Northern Hemisphere. There is a lot of available data across the globe with Canada, the US, Europe, Russia and Japan providing a dense network



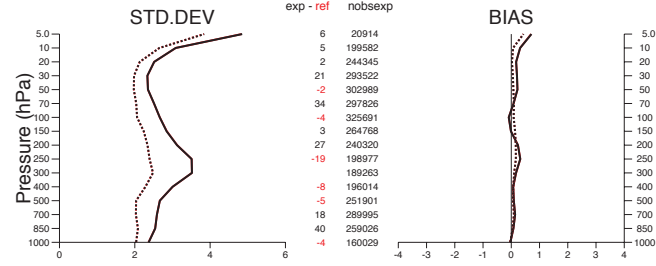
**Figure 4.6:** Radiosonde wind direction departure statistics of the operational run (ref, red) compared to an experiment with radiosonde wind direction VarBC (exp, black). Dotted lines refer to analysis departures, straight lined to background departures. Top panels show used U wind component, bottom panels show used V wind component. Statistics are calculated over a period of 100 days from May to August 2012.

of radiosonde observations. In Figure 4.6 statistics of the used U and V wind components are plotted. The column *nobsexp* shows the number of observations in each pressure level. The column *exp-ref* shows the difference in observations between the experiment and the control run. Positive numbers indicate that more observations were used in the experiment compared to the control run. In this case in most of the levels there are more observations used when the wind VarBC is turned on. This is expected since the quality control should improve the observations. However compared to the overall numbers of observations the numbers are very small which most likely is due to the good quality of the radiosonde network in the Northern Hemisphere (shown by the small bias values on the right side of Figure 4.6).

Statistics for the Southern Hemisphere and the Tropics look alike (not shown) with less observations but still similar amounts used in both runs. The bias is not as small as in the Northern Hemisphere and in the Southern Hemisphere more observations were used after bias correction in higher levels of altitude (between 100 and 20 hPa).

The results for the control and the experiment with the artificial bias added at selected stations show a similar picture. This is not unlikely as it is expected that a change at only 4 stations spread over the globe should not change the overall statistics of the used observations. As an example Figure 4.7 shows the statistics for used U wind component in the Northern Hemisphere (where 3 out of the 4 artificially changed radiosonde stations are located).

The global statistic show very little impact of the wind direction VarBC in a current operational NWP model. This is expected due to the quality of a state of the art radiosonde network like the one currently in operation in most parts of the world. However results might look different when the wind bias correction would be applied to data from the first half of the 20th century. Radiosondes observed less frequent then and reached lower heights in the atmosphere (Ramella Pralungo et al., 2014; Ramella Pralungo and Haimberger, 2014). Therefore these pa-



**Figure 4.7:** Radiosonde statistics of the operational run with added artificial bias (ref, red) compared to an experiment also with artificial bias and with radiosonde wind direction VarBC (exp, black). Dotted lines refer to analysis departure statistics, straight lined to background departure statistics. The panels show statistics for U wind component. Statistics are calculated over a period of 100 days from May to August 2012.

rameters need to be adjusted depending on the usage of the model and the knowledge of the radiosonde network used.

## 5 Conclusions

This work showed that VarBC can also be applied to conventional data when the root of the bias is reasonable well known (This has also been proven by work done on aircraft temperature and surface pressure bias correction done at ECMWF (Isaksen et al., 2012)). The grouping of the data needs to be considered carefully and can change depending on the bias and the information of the observation (transmitted meta data e.g. radiosonde type) available. First tests for radiosonde wind direction VarBC show promising results. This bias correction will be usable in reanalysis efforts as well as NWP models. If used in reanalysis efforts further sensitivity tests will be necessary to adjust the system to cases with sparse observation data. Using this bias correction system a discussion about the linearity of the bias correction also has to be considered. Due to the transformation to wind direction with a rotational matrix it is a valuable argument that the resulting bias correction is not linear but a non-linear bias correction for wind components which are the model variables. This can influence the outcome for wind direction computed from the bias corrected wind components. Further investigation on this topic will give a clearer vision of the influence of this non-linearity.

Due to changes in the ECMWF model version and a change in computing facilities only a limited number of cases and longterm studies were possible. These studies unfortunately did not lead to a case where an artificial bias was completely adjusted over the evaluated period once and forever which would have demonstrated the convergence of the approach. A possible explanation for this could also be the non-linearity of this bias correction within the model variables as discussed above. Also preconditioning efforts could help a faster convergence. Nevertheless the effort taken can help to build a similar system for different future projects such as reanalysis efforts and NWP models.

## Acknowledgements

The author would like to thank ECMWF and especially the reanalysis group for their support and encouragement of this work. Interesting discussions with the group as well as with the staff at the University of Vienna lead to new ideas. This work has been funded by project P21772-N22

of the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Visits at ECMWF were supported through the GRAS-SAF project (CDOP-1 VS10). A background knowledge was achieved while taking part in the graduated Trainee program at ECMWF.

## References

- Andersson, E. and Järvinen, H. (1999). Variational quality control. *Q. J. Roy. Meteor. Soc.*, 125:697–722.
- Dee, D. and Uppala, S. (2009). Variational bias correction of satellite radiance data in the ERA-interim reanalysis. *Q. J. Roy. Meteor. Soc.*, 135:1830–1841.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Mange-Sanz, B., Morcrette, J., Park, B., Paubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.*, 137:553–597. doi:10.1002/qj.828.
- Gruber, C. and Haimberger, L. (2008). On the homogeneity of radiosonde wind time series. *Meteorol. Z.*, 17:631–643.
- Haimberger, L. (2007). Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, 20:1377–1403.
- Hollingsworth, A., Shaw, D. B., Lönnberg, P., Illari, L., Arpe, K., and Simmons, A. J. (1986). Monitoring of observation and analysis quality by a data assimilation system. *Monthly Weather Review*, 114:861–879.
- Isaksen, I., Vasiljevic, D., D., D., and Healy, S. (2012). Bias correction of aircraft data implemented in November 2011. *ECMWF Newsletter*, 131:6.
- Poli, P., Hersbach, H., Tan, D., Thépaut, J., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Trémolet, Y., Holm, E., Bonavita, M., Isaksen, I., and Fisher, M. (2013). The data assimilation system and initial performance evaluation of the ECMWF pilot analysis of the 20th-century assimilating surface observations only (ERA-20C). ERA-report 14, ECMWF.
- Ramella Pralungo, L. and Haimberger, L. (2014). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 - Part II: Homogeneity adjustments for pilot balloon and radiosonde wind data. *Earth Syst. Sci. Data*, submitted.
- Ramella Pralungo, L., Haimberger, L., Stickler, A., and Brönnimann, S. (2014). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 - Part I: Merging and interpolation to 0:00 and 12:00 GMT. *Earth Syst. Sci. Data*, 6:185–200.



## APPENDIX

### A New introduced module

The new module built to include VarBC for radiosonde wind direction data

```

MODULE varbc_rsw
!-----
!   VARBC_RSW – Module for variational bias correction
!   (radiosonde wind direciton)
!   Purpose.
!-----
!   Data structures and controls for variational bias correction
!   (RSW data)
!   Author.
!-----
!   Dick Dee (ECMWF)
!   Modifications.
!   Christina Tavalato
!-----
!   Original    2008/10/01
!-----

USE parkind1 , ONLY : JPIM, JPRB
USE yomhook   , ONLY : LHOOK, DR.HOOK
USE yomlun    , ONLY : NULOUT, NULNAM
USE yomcoctp  , ONLY : NSATEM, NSSMI
USE yomcosjo  , ONLY : JOT
USE varbc_pred , ONLY : NPREDCS.RSW
USE yomdb
USE yomct0    , ONLY : NPROC
USE mpl.module, ONLY : MPL.ALLREDUCE, MPL.BARRIER

IMPLICIT NONE
SAVE
PRIVATE
PUBLIC lbc_rsw
PUBLIC varbc_rsw_config , varbc_rsw_groups
PUBLIC varbc_rsw_varbcix , varbc_rsw_pred
PUBLIC varbc_rsw_min , varbc_rsw_groupdescr

INTERFACE varbc_rsw_config
  MODULE PROCEDURE config
END INTERFACE varbc_rsw_config

INTERFACE varbc_rsw_groups
  MODULE PROCEDURE find_groups
END INTERFACE varbc_rsw_groups

INTERFACE varbc_rsw_groupdescr
  MODULE PROCEDURE groupdescr
END INTERFACE varbc_rsw_groupdescr

```

```
INTERFACE varbc_rsw_pred
  MODULE PROCEDURE get_predictors
END INTERFACE varbc_rsw_pred

INTERFACE varbc_rsw_varbcix
  MODULE PROCEDURE set_varbcix
END INTERFACE varbc_rsw_varbcix

INTERFACE varbc_rsw_min
  MODULE PROCEDURE get_min
END INTERFACE varbc_rsw_min

CHARACTER(LEN=*), PARAMETER :: modname = 'VarBC_RSW'
! Table for available station id/sonde type combinations
!


---


INTEGER(KIND=JPIM), PARAMETER :: len_statid = 8
! length of station id string in ODB
TYPE type_stations
  CHARACTER(LEN=len_statid) :: nstatid
  INTEGER(KIND=JPIM) :: nsondes
END TYPE type_stations
INTEGER(KIND=JPIM), PARAMETER :: jpmxns = 2881
! max number of sation ids/sonde type combinations
TYPE(type_stations) :: YSTATIONS(1:jpmxns)
INTEGER(KIND=JPIM), PARAMETER :: jpmxnp = SIZE(NPREDCS.RSW)
! max number of parameters per group
! Table for settings by sondetype
!


---


TYPE type_config
  CHARACTER(LEN=20) :: name ! sonde type
  INTEGER(KIND=JPIM) :: nsondes ! sonde type id
  INTEGER(KIND=JPIM) :: nparam ! number of parameters
  INTEGER(KIND=JPIM) :: npredcs(jpmxnp) ! list of predictors
  LOGICAL :: llconst(jpmxnp) ! flags for constant parameters
  REAL(KIND=JPRB) :: zparams(jpmxnp) ! a priori parameter values
  REAL(KIND=JPRB) :: dfgdep ! histogram range
  INTEGER(KIND=JPIM) :: ncstart ! cold start option
  INTEGER(KIND=JPIM) :: nbgsdvv ! number for background constr.
  LOGICAL :: lreset ! flag to reset parameters
  LOGICAL :: llmode ! flag for mode correction
END TYPE type_config
!


---


! Namelist parameters
!


---


TYPE(type_config) :: yconfig(2)
LOGICAL :: lbc_rsw
LOGICAL :: lbc_vaisala
INTEGER(KIND=JPIM) :: nbgs_vaisala
LOGICAL :: lbc_other
```

---

```

INTEGER(KIND=JPIM) :: nbg_other
LOGICAL                :: lconst           ! Keep all VarBC param. fixed
LOGICAL                :: lmode           ! Correct pdf mode (not mean)
NAMelist /NAMVARBCRSW/ yconfig, &
                        & lbc_rsw, &
                        & lbc_vaisala, &
                        & nbg_vaisala, &
                        & lbc_other, &
                        & nbg_other, &
                        & lconst, lmode

!
INTERFACE
#include "getdb.h"
#include "putdb.h"
END INTERFACE
#include "abor1.intfb.h"
#include "posnam.intfb.h"
! =====
CONTAINS
! =====
SUBROUTINE config
!
!   config - Configure VarBC for RSW data
!   Interface.
!
!   CALL config
!   Author.
!
!   Dick Dee (ECMWF)
!   Modifications.
!   Christina Tavorato
!
!   Original      2008/03/01
!
!
INTEGER(KIND=JPIM) :: is
REAL(KIND=JPRB)    :: zhook_handle
CHARACTER(LEN=*), PARAMETER :: myname = modname//'_'(config)'
IF (LHOOK) CALL DR_HOOK(myname,0,zhook_handle)
WRITE(NULOUT, '(/,a)') myname//':_Configuring _RSW_VarBC'
DO is = 1, SIZE(yconfig,1)
  yconfig(is)%nparam      = 1
  yconfig(is)%npredcs(:)  = NPREDCS_RSW ! p0
  yconfig(is)%zparams(:)  = 0.0_JPRB    ! a priori parameter estimates
  yconfig(is)%llconst(:)  = .false.     ! keep this parameter fixed
  yconfig(is)%ncstart      = 0           ! cold start from 0
  yconfig(is)%nbgstdv      = 0           ! background constraint
  yconfig(is)%llmode       = .false.     ! correct mode of pdf
ENDDO
yconfig(1)%name = 'OTHER'
yconfig(1)%nsondes = -1

```

```
yconfig(1)%dfgdep = 20.0_JPRB
yconfig(1)%nparam = 1
yconfig(1)%npredcs(1:1) = (/0/)
yconfig(1)%ncstart = 0
yconfig(2)%name = 'VAISALA_RS92'
yconfig(2)%nsondes = 81
yconfig(2)%dfgdep = 20.0_JPRB
yconfig(2)%nparam = 1
yconfig(2)%npredcs(1:1) = (/0/)
yconfig(2)%ncstart = 0
! Switches by sonde type:
!
lbc_vaisala = .true.
lbc_other = .true.
! Background error parameters by sonde type:
!
nbg_vaisala = 100
nbg_other = 100
! Panic button:
!
lbc_rsw = .true. ! If this is false then no RSW bias correction
! All parameters constant:
!
lconst = .false.
! Correct mode rather than mean:
!
lmode = .false.
WRITE(NULOUT, '(a)') myname//': &
.....&_Reading_namelist_NAMVARBCRSW, _first_iteration'
CALL posnam(NULNAM, 'NAMVARBCRSW')
READ(NULNAM, NAMVARBCRSW)
IF (.NOT. (lbc_rsw .AND. lbc_vaisala)) yconfig( 2)%nparam = 0
IF (.NOT. (lbc_rsw .AND. lbc_other )) yconfig( 1)%nparam = 0
yconfig(2)%nbgstdv = nbg_vaisala
yconfig(1)%nbgstdv = nbg_other
DO is = 1, SIZE(yconfig,1)
  yconfig(is)%lconst(:) = lconst
  yconfig(is)%lmode = lmode
ENDDO
! Override settings for individual layers/predictors:
!
WRITE(NULOUT, '(a)') myname//': &
.....&_Reading_namelist_NAMVARBCRSW, _second_iteration'
CALL posnam(NULNAM, 'NAMVARBCRSW')
READ(NULNAM, NAMVARBCRSW)
IF (LHOOK) CALL DR.HOOK(myname,1,zhook_handle)
END SUBROUTINE config
!
SUBROUTINE encode_groupkey (cdkey, kstatid)
!
```

---

```

! Encodes group description
! cdkey: encoded string defining a group of RSW data (just station ID)
!-----
CHARACTER(LEN=*) , INTENT(OUT) :: cdkey
CHARACTER(LEN=*) , INTENT(IN)  :: kstatid
REAL(KIND=JPRB) :: zhook_handle
CHARACTER(LEN=*) , PARAMETER :: myname = modname//'_'(encode_groupkey)'
IF (LHOOK) CALL DR_HOOK(myname,0,zhook_handle)
WRITE(cdkey,'(a8)') kstatid
IF (LHOOK) CALL DR_HOOK(myname,1,zhook_handle)
END SUBROUTINE encode_groupkey
!-----

SUBROUTINE decode_groupkey (cdkey, kstatid)
!-----
! Decodes group description
! cdkey: encoded string defining a group of RSW data
!-----
CHARACTER(LEN=*) , INTENT(IN)  :: cdkey
CHARACTER(LEN=*) , INTENT(OUT) :: kstatid
REAL(KIND=JPRB) :: zhook_handle
CHARACTER(LEN=*) , PARAMETER :: myname = modname//'_'(decode_groupkey)'
IF (LHOOK) CALL DR_HOOK(myname,0,zhook_handle)
READ(cdkey,'(a8)') kstatid
IF (LHOOK) CALL DR_HOOK(myname,1,zhook_handle)
END SUBROUTINE decode_groupkey
!-----

SUBROUTINE groupdescr (cdkey, cdescr)
!-----
! Returns group description
! cdkey: encoded string defining a group of RSW data
! cdescr: string suitable for printing
!-----
CHARACTER(LEN=*) , INTENT(IN)  :: cdkey
CHARACTER(LEN=*) , INTENT(OUT) :: cdescr
CHARACTER(LEN=len_statid)      :: istatid
CHARACTER(LEN=80)              :: cldesc
INTEGER(KIND=JPIM)              :: j, ilen
REAL(KIND=JPRB)                :: zhook_handle
CHARACTER(LEN=*) , PARAMETER :: myname = modname//'_'(groupdescr)'
IF (LHOOK) CALL DR_HOOK(myname,0,zhook_handle)
CALL decode_groupkey (cdkey, istatid)
cldesc(:) = '_'
WRITE(cldesc,'(a,4x,2a)') 'RSW', &
& '_statid=', istatid
ilen = LEN(cldesc)
IF (LEN(cdescr)<ilen) CALL ABOR1(myname//':_Increase_string_length')
cdscr(1:ilen) = cldest
IF (LHOOK) CALL DR_HOOK(myname,1,zhook_handle)
END SUBROUTINE groupdescr
!-----

```

---

```
SUBROUTINE find_groups (cdkeys, kng)
!
! find_groups - Find data groups present in ODB requiring VarBC
! Interface.
!
! CALL find_groups (cdkeys, kng)
! cdkeys(1:kng) group descriptions
! Author.
!
! Dick Dee (ECMWF)
! Modifications.
! Christina Tavalato
!
! Original 2008/03/01
!
CHARACTER(LEN=*) , INTENT(OUT) :: cdkeys(:)
INTEGER(KIND=JPIM), INTENT(OUT) :: kng
CHARACTER(LEN=len_statid) :: istatid
INTEGER(KIND=JPIM) :: isondes
INTEGER(KIND=JPIM) :: is, js, ins
REAL(KIND=JPRB) :: zhook_handle
CHARACTER(LEN=*), PARAMETER :: myname = modname//'_'(find_groups)
IF (LHOOK) CALL DRHOOK(myname,0,zhook_handle)
WRITE(NULOUT,'(a)') myname//':_Looking_for_RSW_data_groups..'
IF (nproc>1) CALL MPLBARRIER(CDSTRING=myname)
! Find all station ids for this codetype
!
CALL find_stations(ins)
! Return group descriptions for VarBC
!
kng = 0
IF (lbc_rsw .AND. ins>0) THEN
  DO is = 1, ins ! Look at each sonde
    istatid = YSTATIONS(is)%nstatid
    kng = kng + 1
    IF (kng>size(cdkeys)) CALL ABORT(myname//':_Insufficient_allocation.')
    CALL encode_groupkey (cdkeys(kng), istatid)
  ENDDO
  WRITE(NULOUT,'(a,i3,a)') myname//':_', kng, '_RSW_data_groups_for_VarBC'
ENDIF
IF (LHOOK) CALL DRHOOK(myname,1,zhook_handle)
END SUBROUTINE find_groups
!
SUBROUTINE find_stations (kns)
INTEGER(KIND=JPIM), INTENT(OUT) :: kns
CHARACTER(LEN=len_statid) :: istatid
INTEGER(KIND=JPIM) :: is, isondes
INTEGER(KIND=JPIM) :: iret, info(1), ic, j
REAL(KIND=JPRB) :: zinfo(0)
REAL(KIND=JPRB) :: zhook_handle
```

```

CHARACTER(LEN=*), PARAMETER :: myname = modname//'_'(find_stations)'
#include "openmp_obs.h"
IF (LHOOK) CALL DRHOOK(myname,0,zhook_handle)
WRITE(NULOUT,'(a)') &
  & myname//':_Checking_ODB_for_sonde_station_ids'
IF (nproc>1) CALL MPLBARRIER(CDSTRING=myname)
CALL getdb('VARBCRSW_STATIONS', 0, iret, info, 0, zinfo, 0, &
  & -1, -1, -1, -1, -1, -1)
IF (nproc>1) CALL MPLBARRIER(CDSTRING=myname)
ic = 0
istatid = '_____'
DO j = 1, NROWS.ROBSU ! .. find all distinct stationid's
! IF (ROBHDR(j,MDBSID) ==istatid) CYCLE ! duplicate
  istatid = TRANSFER(ROBSU(j,MDBSID), istatid)
  isondes = ROBSU(j,MDB.SONDE_TYPE.AT.CONV)
  ic = ic + 1
  IF (ic>jpmxns) &
    & CALL ABOR1(myname//':_Insufficient_allocation_-_increase_jpmxns')
  YSTATIONS(ic)%NSTATID = istatid
  YSTATIONS(ic)%NSONDES = isondes
  WRITE(NULOUT,'(3a,i0)') myname//':_statid=',istatid,'_sonde=',isondes
ENDDO
kns = ic
CALL putdb('VARBCRSW_STATIONS', 0, iret, info, 0, zinfo, 0)
IF (LHOOK) CALL DRHOOK(myname,1,zhook_handle)
END SUBROUTINE find_stations
!_____
SUBROUTINE get_predictors (cdkey, knparam, &
  & kpredcs, pparams, ldreset, kstart, pdfgdep)
! Return list of predictors, a-priori parameter values, ..
CHARACTER(LEN=*) , INTENT(IN) :: cdkey
INTEGER(KIND=JPIM), INTENT(OUT) :: knparam, kstart
LOGICAL , INTENT(OUT) :: ldreset
INTEGER(KIND=JPIM), INTENT(OUT) :: kpredcs(jpmxnp)
REAL(KIND=JPRB) , INTENT(OUT) :: pparams(jpmxnp)
REAL(KIND=JPRB) , INTENT(OUT) :: pdfgdep
CHARACTER(LEN=*), PARAMETER :: myname = modname//'_'(get_predictors)'
CHARACTER(LEN=len_statid) :: istatid
LOGICAL :: llmatch
INTEGER(KIND=JPIM) :: isondes, js, is
REAL(KIND=JPRB) :: zhook_handle
IF (LHOOK) CALL DRHOOK(myname,0,zhook_handle)
CALL decode_groupkey (cdkey, istatid)
DO is = 1,jpmxns
  IF (YSTATIONS(is)%nstatid==istatid) THEN
    isondes = YSTATIONS(is)%nsondes
    llmatch = .false.
    IF (yconfig(2)%nsondes/=isondes) THEN
      js = 1
      knparam = yconfig(js)%nparam

```

```
      kpredcs(1:knparam) = yconfig(js)%npredcs(1:knparam)
      pparams(1:knparam) = yconfig(js)%zparams(1:knparam)
      kstart          = yconfig(js)%ncstart
      lreset          = yconfig(js)%lreset
      pdfgdep         = yconfig(js)%dfgdep
      llmatch = .true.
      EXIT
    ENDIF
    llmatch = .false.
    IF (yconfig(2)%nsondes==isondes) THEN
      js = 2
      knparam          = yconfig(js)%nparam
      kpredcs(1:knparam) = yconfig(js)%npredcs(1:knparam)
      pparams(1:knparam) = yconfig(js)%zparams(1:knparam)
      kstart          = yconfig(js)%ncstart
      lreset          = yconfig(js)%lreset
      pdfgdep         = yconfig(js)%dfgdep
      llmatch = .true.
      EXIT
    ENDIF
    IF (.NOT. llmatch) CALL ABOR1(myname//':_No_match.')
  ENDIF
ENDDO
IF (LHOOK) CALL DRHOOK(myname,1,zhook_handle)
END SUBROUTINE get_predictors
!-----
SUBROUTINE get_min ( cdkey, knparam, kbgstdv, pdfgdep )
!-----
!   get_min - Return parameters used in the minimisation for a particular
!             group of RSW data
!   Author.
!-----
!   Dick Dee (ECMWF)
!   Modifications.
!   Christina Tavalato
!-----
!   Original      2008/03/01
!-----
CHARACTER(LEN=*) , INTENT(IN) :: cdkey
INTEGER(KIND=JPIM) , INTENT(IN) :: knparam
INTEGER(KIND=JPIM) , INTENT(OUT) :: kbgstdv
REAL(KIND=JPRB) , INTENT(OUT) :: pdfgdep
CHARACTER(LEN=len_statid) :: istatid
INTEGER(KIND=JPIM) :: js, isondes, is, ic, jpmxnp
REAL(KIND=JPRB) :: zhook_handle
LOGICAL :: llmatch
CHARACTER(LEN=*) , PARAMETER :: myname = modname//'_'(get_min)'
IF (LHOOK) CALL DRHOOK(myname,0,zhook_handle)
CALL decode_groupkey (cdkey, istatid)
DO is = 1,jpmxns
```



```

IF (ystations(is)%nstatid==istatid) THEN
  isondes = ystations(is)%nsondes
  llmatch = .false.
  IF (yconfig(2)%nsondes/=isondes) THEN
    js=1
    kbgstdv = yconfig(js)%nbgstdv
    pdfgdep = yconfig(js)%dfgdep
    llmatch = .true.
    EXIT
  ENDIF
  llmatch = .false.
  IF (yconfig(2)%nsondes==isondes) THEN
    js=2
    kbgstdv = yconfig(js)%nbgstdv
    pdfgdep = yconfig(js)%dfgdep
    llmatch = .true.
    EXIT
  ENDIF
  IF (.NOT. llmatch) CALL ABOR1(myname//':_No_match.')
ENDIF
ENDDO
IF (LHOOK) CALL DRHOOK(myname,1,zhook_handle)
END SUBROUTINE get_min
!
SUBROUTINE set_varbcix (cdkeys, kix, kng)
!
!   set_varbcix - Set VarBC group indices in ODB
!   Author.
!
!   Dick Dee (ECMWF)
!   Modifications.
!   Christina Tavolato
!
!   Original      2008/03/01
!
INTEGER(KIND=JPIM), INTENT(IN) :: kix(kng), kng
CHARACTER(LEN=*) , INTENT(IN) :: cdkeys(kng)
INTEGER(KIND=JPIM) :: iret, ilen, jg, info(1), jobs, ibody, jlev
INTEGER(KIND=JPIM) :: isondes
REAL(KIND=JPRB) :: zinfo(0), zhook_handle
CHARACTER (LEN=len_statid) :: istatid, istid
CHARACTER(LEN=*), PARAMETER :: myname = modname//'_'(set_varbcix)'
#include "openmp_obs.h"
IF (LHOOK) CALL DRHOOK(myname,0,zhook_handle)
WRITE (NULOUT,'(/,a)') &
  & myname//':_Setting_VarBC_group_indices_(varbc_ix@body)_in_ODB'
CALL GETIDB('VARBC.RSW',0,iret,info,0,zinfo,0,-1,-1,-1,-1,-1,-1)
ilen = iret
DO jg = 1, kng
  CALL decode_groupkey (cdkeys(jg), istatid)

```

```

DO jobs = 1, ilen
  istid = TRANSFER(ROBHDR(jobs,MDBSID), istid)
  IF (istid == istatid) THEN
    DO jlev = MLNKH2B(jobs), MLNKH2B(jobs+1) - 1
      ROBODY(jlev,MDB.VARBC_IX_AT_BODY) = kix(jg)
      WRITE(NULOUT, '(a,i0)') 'Varbc_Index',jlev,kix(jg)
    ENDDO
  ENDIF
ENDDO
CALL PUTDB('VARBC.RSW',0,iret,info,0,zinfo,0)
IF (LHOOK) CALL DR_HOOK(myname,1,zhook_handle)
END SUBROUTINE set_varbcix
!
END MODULE VARBC_RSW

```

---

## B List of modified routines

Following routines were modified for the inclusion of VarBC for radiosonde wind direction data. The routines are coded at ECMWF in the IFS branch da6\_CY38R2\_varbcRSWtestm\_exp:

- /ifs/module/varbc\_setup.F90

This routine includes all the different VarBC modules, calls the setup routines and the bias correction one by one. The module presented in appendix A was introduced in this routine and the setup runs alongside the other VarBC modules.

- /ifs/module/varbc\_pred.F90

In this routine the predictors for each VarBC module are set. A constant predictor was included for the VarBC radiosonde wind direction module.

- /ifs/op\_obs/hop.F90

- /ifs/op\_obs/hopad.F90

- /ifs/op\_obs/hoptl.F90

In these routines, describing the observation operator its adjoint and tangent linear, wind components were transformed from the u and v components to wind direction for the bias correction. The transformation and introduction of a background term for wind direction in hop.F90 (line 1169 and following) is shown below. Note that as mentioned in section 3.3.1 the rotation rather than its linearization is used.

```

IF (LBC_RSW .AND. LVARBC .AND. LL_TEMP .AND. CLV(JVNM) == 'U') THEN
  ALLOCATE(ZXPPB(KDLEN,IMXCOUNT,1),ZXPRED(0:JPMXNPRED,ILEN), &
    & ZOMF(KDLEN,IMXCOUNT),IXVARBC(KDLEN,IMXCOUNT), &
    & ZUVO(KDLEN,IMXCOUNT,2),ZUVB(KDLEN,IMXCOUNT,2), &
    & ZUVN(KDLEN,IMXCOUNT,2),ZXPPN(KDLEN,IMXCOUNT,2))

  ZOMF(:,:) = 0.0_JPRB
  ZUVO(:,:,:) = RMDI
  ZUVB(:,:,:) = RMDI
  ZUVN(:,:,:) = RMDI
  ZDIR = 0.0_JPRB

  DO JNLV = 1,IMXCOUNT
    DO JOBS = 1,ILEN
      IF (JNLV > ICOUNT(JOBS)) CYCLE
      IBODY = MLNKH2B(JOBS) + (IPNLV(JOBS,JNLV)-1)
      IXVARBC(JOBS,JNLV) = ROBODY(IBODY,MDB.VARBC_IX_AT_BODY)
      IF (ROBODY(IBODY,MDBVAR) /= RMDI .AND. &
        & ROBODY(IBODY+1,MDBVAR) /= RMDI .AND. &
        & ROBODY(IBODY,MDBOMF) /= RMDI .AND. &
        & ROBODY(IBODY+1,MDBOMF) /= RMDI) THEN
        ZUVO(JOBS,JNLV,1) = ROBODY(IBODY,MDBVAR)
        ZUVO(JOBS,JNLV,2) = ROBODY(IBODY+1,MDBVAR)
        ZUVB(JOBS,JNLV,1) = ZXPP(JOBS,JNLV,1)
        ZUVB(JOBS,JNLV,2) = ZXPP(JOBS,JNLV,2)
      IF (SQRT(SUM(ZUVO(JOBS,JNLV,:)*ZUVO(JOBS,JNLV,:))) &
        & *SQRT(SUM(ZUVB(JOBS,JNLV,:)*ZUVB(JOBS,JNLV,:))) &

```

```
& /= 0.0_JPRB) THEN
ZOMF(JOBS,JNLV) = &
& ACOS(SUM(ZUVO(JOBS,JNLV,:) * ZUVB(JOBS,JNLV,:)) &
& / (SQRT(SUM(ZUVO(JOBS,JNLV,:) * ZUVO(JOBS,JNLV,:)) &
& * SQRT(SUM(ZUVB(JOBS,JNLV,:) * ZUVB(JOBS,JNLV,:)))))
ZDIR=ZUVO(JOBS,JNLV,1)*ZUVB(JOBS,JNLV,2) - &
& ZUVO(JOBS,JNLV,2)*ZUVB(JOBS,JNLV,1)
IF (ZDIR < 0.0_JPRB .AND. ZOMF(JOBS,JNLV) /= RMDI) &
& ZOMF(JOBS,JNLV) = - ZOMF(JOBS,JNLV)
ENDIF
ENDIF
ENDDO
ENDDO

CALL VARBC_PRED_RSW (ILEN, IMXCOUNT, ICOUNT(1:ILEN), &
& IXVARBC(1:ILEN,:), JPMXNPRED, ZXPRED, LSTATS=LIFSTRAJ)

CALL VARBC_BIAS (IXVARBC, IMXCOUNT, ICOUNT, KDLEN, ILEN, &
& JPMXNPRED, ZXPRED, ZXPPB(:, :, 1))

DO JNLV = 1,IMXCOUNT
DO JOBS = 1,ILEN
IF (JNLV > ICOUNT(JOBS)) CYCLE
IBODY = MLNKH2B(JOBS) + (IPNLV(JOBS,JNLV)-1)
IF (ZXPPB(JOBS,JNLV,1) /= RMDI .AND. &
& ROBODY(IBODY,MDBVAR) /= RMDI .AND. &
& ROBODY(IBODY+1,MDBVAR) /= RMDI) THEN
ZUVN(JOBS,JNLV,1) = &
& ROBODY(IBODY,MDBVAR)*(COS(ZXPPB(JOBS,JNLV,1))-1) &
& - ROBODY(IBODY+1,MDBVAR)*SIN(ZXPPB(JOBS,JNLV,1))
ZUVN(JOBS,JNLV,2) = &
& ROBODY(IBODY,MDBVAR)*SIN(ZXPPB(JOBS,JNLV,1)) &
& + ROBODY(IBODY+1,MDBVAR)*(COS(ZXPPB(JOBS,JNLV,1))-1)
ZXPPN(JOBS,JNLV,1) = &
& ZUVN(JOBS,JNLV,1) - ROBODY(IBODY,MDBVAR)
ZXPPN(JOBS,JNLV,2) = &
& ZUVN(JOBS,JNLV,2) - ROBODY(IBODY+1,MDBVAR)
ENDIF
ROBODY(IBODY,MDBTORB) = ZXPPN(JOBS,JNLV,1)
ROBODY(IBODY+1,MDBTORB) = ZXPPN(JOBS,JNLV,2)
ENDDO
ENDDO
DEALLOCATE(ZXPPB, ZXPRED, ZOMF, IXVARBC, ZUVO, ZUVB, ZUVN, ZXPPN)
ENDIF
```

At the end of the routine the observation type radiosonde is added to an IF statement to introduce bias correction when dealing with radiosonde data (line and 2596 following).

```
IF (IOBSTYPE == NAIREP .OR. IOBSTYPE == NISAC .OR. &
& IOBSTYPE == NIEMP) THEN
CALL HDEPART(KDLEN, ILEN, KDBDY, IMXBDY, IOBSTYPE, ICMBDY, IVNMRQ, ZHOFX)
```

**ELSE**

...

Here the rotation matrix rather than its linearization is used. Addressing the fact that the bias correction is non-linear in the model variables (it is linear for wind direction which is not a model variable) there is an argument for using the linearization of the rotation matrix in the minimisation that needs to be discussed when applying this bias correction.

With the linearization of the rotation matrix as presented in section 3.3.1 in mind the code changes in the adjoint and tangent linear routine of the observation operator are computed likewise.

- /odb/ddl/varbc\_rsw\_rbody.sql
- /odb/ddl/varbc\_rsw\_rohdr.sql
- /odb/ddl/varbc\_rsw\_stations.sql

These are the sql requests called to identify the radiosonde wind data. They also include a call for the radiosonde type if this is reported. The call to the observation data base (ODB) is performed when the VarBC groups are built.

For the experiments with an artificial bias also the following routine was changed:

- /odb/bufr2odb/bufr2odb\_temp.F90

In this routine the observed data are converted to wind components and stored in the ODB. Before the conversion the artificial wind direction bias was introduced to certain stations (identified by their WMO-ids).

After a check for the date (artificial bias are only introduced 4 days into the experiment) the following lines are used to introduce the artificial bias. These lines add the 15 degrees to the wind direction value and check if this leads to a value greater than 360 degrees. If this happens the new wind direction is corrected to be between 0 and 360 degrees (introduced to the code after line 835 - only one station (Marion Island) is shown in this example, the code for other stations only changes in the call for the station-id).

```

IF(idate > 20120502 .AND. idate < 20121130) THEN
  WRITE(0,*) 'VarBC_add_bias_on:',idate
  WRITE(0,*) 'VarBC_RSW_add_no_wind_direction_bias:station:', &
    & adjustl(csgn(i))
  WRITE(0,*) 'Wind:', vals(i_011001)
  IF(adjustl(csgn(i)) == '68994') THEN
    WRITE(0,*) 'VarBC_Marion_Island '
    body(jobs,j_obsvalue)=vals(i_011001)+15.
    IF(body(jobs,j_obsvalue) > 360.) THEN
      body(jobs,j_obsvalue)=body(jobs,j_obsvalue)-360.
    ENDIF
  ENDIF
  ...
ENDIF

```

To execute VarBC for radiosonde wind direction bias the logical of the namelist needs to be set in the following scripts:

- /scrips/gen/ifstraj

- /scripts/gen/ifsmin

This is done by setting the following logical to *.TRUE.* in both scripts when including the different name lists (ifstraj: line 1567 and following; ifsmin: line 1039 and following):

```
&NAMVARBCRSW  
    LBC_RSW=.TRUE. ,  
/
```

The setting should be done in this scripts to include VarBC for radiosonde wind direction however this setting can be overwritten in the VarBC module introduced in appendix A.

## 6 Final Remarks and Conclusions

### 6.1 Final Remarks

This thesis discussed several aspects concerning the quality of meteorological observations. Using observations of the past and present comes with some challenges ascertaining the quality of the observations used and the quality of the result. Over the last decades, with the increase of computational power and the increase in complexity of the numerical weather prediction (NWP) models, the amount of observations used in each system cycle has also increased. Therefore new ways of quality control and bias correction had to be introduced.

#### 6.1.1 Variational Quality Control

During my stay as a Graduated Trainee at ECMWF the opportunity arose to develop a VarQC system that uses a Huber norm to describe the data innovations rather than the previously used Gaussian and flat curve. This distribution was first considered by Andersson and Järvinen (1999) but at that time the decision was made to implement a Gaussian plus flat distribution. The idea to implement a Huber norm instead has been prepared in talks at ECMWF and I was appointed to look into possible observation types that would benefit from such a quality control. Statistics were calculated for all data sets and the decision was made to apply the new Huber norm VarQC on all conventional data except humidity observations as their distributions would need to be normalised first.

Results proved the system to be successful with less rejections in cases where the background forecast differed substantially from the observation due to a fast developing small scale weather event. The system also provided better analyses in case of tropical storms. High quality dropsondes used to observe tropical storms were assigned high weights, therefore benefiting the analysis.

During the implementation of the system other quality control steps were revisited and due to the robustness of the Huber norm the first-guess limits could be broadened. In the process, observation error statistics were also looked at and revisited and changes were made to represent the radiosonde observation errors better. Also adjustments to the surface pressure bias correction were made to fit the settings of the new system.

With everything in place the new developed VarQC system was implemented in the ECMWF operational system in September 2009 and has been in use since then.

Next to the quality control of observations within a data assimilation system for an operational model, biases in long term observation records need to be adjusted as well before the results of the data set can be interpreted in a climatological way.

### **6.1.2 Bias Correction using Homogenisation Methods**

One of the most important climatological data sets is the radiosonde data set. This conventional data set provides not only surface data but also data from the upper atmosphere. The first radiosondes were operated in the 1930s (even earlier data from registering balloons is available) however a dense radiosonde coverage can only be found from 1958 (the International Geophysical Year) onwards. This date is often used as the starting point for climatological data sets and reanalysis data sets. Therefore homogenised upper air records from 1958 onwards are useful as a climate reference as well as input data for reanalysis. Different to other homogenised radiosonde data sets, the presented data sets RAOBCORE and RICH are retrieved by automated systems that use all available radiosonde observations on a daily time scale (Most of the other homogenised radiosonde data sets only provide monthly data). Reanalysis data sets from ECMWF are used as a reference to determine the artificial break points in the observation time series and either those data sets or neighbouring radiosonde stations are used to adjust the break in the time series.

The results were compared to various other upper air data sets including other homogenised radiosonde data sets, satellite data and GPS radio occultation data. (A)MSU satellite data available from 1979 onwards provides a data set that covers the whole globe with the setback that the data only gives information on a broad atmospheric layer. GPS radio-occultation data has a higher vertical resolution but is only available from the early 21st century onwards (Hajj et al., 2002). Records so far do not span over a 30 year climatological period. For intercomparison all data sets were calculated to match MSU-equivalent layers. For a better understanding of the uncertainties in the homogenised data sets RAOBCORE and RICH, an ensemble of resolutions with different parameter settings was evaluated. Those results show consistency with the satellite and GPS radio occultation data. Interestingly also the warming of the tropical mid-troposphere predicted by climate models can also be found in the adjusted data.

Homogenised data sets seem to be a good source for climate monitoring and are also very valuable as input data sets for future reanalysis.

### **6.1.3 Comparison of Upper-Air Temperature Data**

A homogenised long-term upper air data set from radiosonde records allows the comparison to different other systems that observe the atmosphere. Recently the introduction of GPS radio occultation data allowed a radiosonde-to-satellite comparison at a high vertical resolution. Therefore data does not have to be converted into the previously used 4 broad atmospheric layers anymore. Although temperature time series from GPS radio occultation data are only available from the early 2000s onwards they will be an extremely useful upper-air dataset in the future alongside radiosonde data since they do not need to be calibrated and the observation



operator is linear with very high accuracy (Healy and Thépaut, 2006).

#### **6.1.4 Variational Bias Correction**

Methods to bias correct observations within the data assimilation are applied to more and more data types (Dee and Uppala, 2009; Isaksen et al., 2012). It is an option that allows a bias correction within the system while simultaneously assimilating the data. Therefore no bias correction step (e.g. adjusting known biases of instrument types reported in the metadata) before assimilation has to be considered. Due to the possibility of different bias parameters and dependences of the bias a good bias description is possible. The interesting question in the future with online bias correction such as variational bias correction will be the anchoring of the system. This used to be done by conventional observations - especially radiosondes. Including radiosonde data into a VarBC system might need a different observation system for the model anchoring. This might be done by GPS radio occultation data which is an upper-air data set with high vertical resolution and good spacial coverage (Melbourne et al., 1994; Kursinski et al., 1997; Hajj et al., 2002).

The introduced radiosonde wind direction VarBC module is a first attempt to include radiosondes in variational bias correction. There is, however, still a need to test the system to verify its correctness. Also, the impact of a linear bias model to a variable that is not a model variable which results in a non-linear bias in the model variables needs to be investigated. More tests will show if it is sensible to apply this correction in a NWP model (this might not be necessary due to the very good quality of radiosonde data and the benefit of having radiosondes as an unadjusted anchoring system) or if the benefits of this VarQC are more suitable for reanalysis efforts where historic radiosonde records are assimilated. These historic records need a homogenisation method applied and this could be an option to do the bias correction online.

## **6.2 Conclusions**

Monitoring, quality control, homogenisation and bias correction are important tools to improve the quality of meteorological observations for assimilation. This is the case for NWP forecast models as well as climate reanalysis. Improving the input data will always lead to a better output of the model.

Advances in this field in the past years have led to the fact that nowadays as much time and effort is put into the data assimilation as in the actual forecast model itself. Constant improvement on observation systems, recordings of meta data and introductions of new data sets allow a better understanding of the observations and therefore a better understanding of their possible dependences, faults and biases. Addressing bias correction as a part of modern meteorological forecast models will lead to a better model output and therefore better forecasts. Not only is

this beneficial for current NWP models but observation quality is a very important part of any climate reanalysis effort. Understanding possible biases of historic observations and introducing automated systems to check and adjust them will make it possible to use more historic data than ever before and to recreate the atmospheric state of the last century for climate studies.

Even with modern observation systems and the technical advances of the recent years it will never be certain that meteorological observations are unbiased. Therefore quality control and bias correction will remain an important issue in the future of NWP models and reanalysis.

## References

- Andersson, E. and Järvinen, H. (1999). Variational quality control. *Q. J. Roy. Meteor. Soc.*, 125:697–722.
- Blunden, J. and Arndt, D. S. (2012). State of the Climate in 2011. *Bulletin of the American Meteorol. Soc.*, 93(7):S1–S264.
- Blunden, J., Arndt, D. S., and Baringer, M. O. (2011). State of the Climate in 2010. *Bulletin of the American Meteorol. Soc.*, 92(6):S1–S266.
- Christy, J., Spencer, R., and Lobl, E. (1998). Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate*, 11:2016–2041.
- Christy, J., Spencer, R., Norris, W., Braswell, W., and Parker, D. (2003). Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *J. Atmos. Ocean. Tech.*, 20:613–629.
- Courtier, P., Thépaut, J., and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. Roy. Meteor. Soc.*, 120:1367–1387.
- Dee, D., Rukhovets, L., Todling, R., da Silva, A., and Larson, J. (2001). An adaptive buddy check for observational quality control. *Q. J. Roy. Meteor. Soc.*, 127:2451–2471.
- Dee, D. and Uppala, S. (2009). Variational bias correction of satellite radiance data in the ERA-interim reanalysis. *Q. J. Roy. Meteor. Soc.*, 135:1830–1841.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Mange-Sanz, B., Morcrette, J., Park, B., Paubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.*, 137:553–597. doi:10.1002/qj.828.
- Friederichs, P. and Thorarinsdottir, T. (2012). Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics*, 23:579–594.
- Gruber, C. and Haimberger, L. (2008). On the homogeneity of radiosonde wind time series. *Meteorol. Z.*, 17:631–643.
- Haimberger, L. (2007). Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, 20:1377–1403.
- Haimberger, L., Tavolato, C., and Sperka, S. (2008). Towards elimination of the warm bias in historic radiosonde temperature records - some results from a comprehensive intercomparison of upper-air data. *J. Climate*, 21:4587–4606.
- Haimberger, L., Tavolato, C., and Sperka, S. (2012). Homogenization of the global radiosonde data set through combined comparison with reanalysis background series and neighbouring stations. *J. Climate*, 25:8108–8131.

- Hajj, G., Kurinski, E., Romans, L., Bertiger, W., and Leroy, S. (2002). A technical description of atmospheric sounding by GPS occultation. *J. Atmos. Sol-Terr. Phys.*, 64:451–469.
- Healy, S. and Thépaut, J. (2006). Assimilation experiments with CHAMP GPS radio occultation measurements. *Q. J. Roy. Meteor. Soc.*, 132:605–623.
- Hollingsworth, A., Shaw, D. B., Lönnberg, P., Illari, L., Arpe, K., and Simmons, A. J. (1986). Monitoring of observation and analysis quality by a data assimilation system. *Monthly Weather Review*, 114:861–879.
- Huber, P. J. (1964). Robust estimates of a location parameter. *Ann. Math. Statist.*, 35:73–101.
- Ingelby, N. B. and Lorenc, A. C. (1993). Bayesian quality control using multivariate normal distributions. *Q. J. Roy. Meteor. Soc.*, 119:1195–1225.
- Isaksen, L., Vasiljevic, D., Dee, D., and Healy, S. (2012). Bias correction of aircraft data implemented in November 2011. *ECMWF Newsletter*, 131:6.
- Kursinski, E., Hajj, G., Schonfield, J., Linfield, R., and Hardy, K. (1997). Observing earth’s atmosphere with radio occultation measurements using global positioning system. *J. Geophys. Res.*, 102:23429–23465.
- Ladstädter, F., Steiner, A. K., Foelsche, U., Haimberger, L., C., T., and Kirchgast, G. (2011). An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes and GPS radio occultation. *Atmos. Meas. Tech.*, 4:1965–1977. doi:10.5194/amt-4-1965-2011.
- Lorenc, A. C. and Hammon, O. (1988). Objective quality control of observations using bayesian methods - theory and practical implementation. *Q. J. Roy. Meteor. Soc.*, 112:515–543.
- Mears, C. and Wentz, F. (2009). Construction of the RSS V3.2 lower-tropospheric temperature dataset from the MSU and AMSU microwave sounders. *J. Atmos. Ocean. Tech.*, 26:1493–1509.
- Mears, C., Wentz, F., and Thorne, P. (2012). Assessing the value of microwave sounding unit-radiosonde comparisons in ascertaining errors in climate data records of tropospheric temperatures. *J. Geophys. Res.*, 117(D19103).
- Melbourne, W., Davis, E., Duncan, C., Hajj, G., Hardy, K., Kurinski, E., Meehan, T., Young, L., and Yunck, T. (1994). The application of spaceborne GPS to atmospheric limb sounding and global change monitoring. *JPL Publication*, 147:94–18.
- Ramella Pralungo, L. and Haimberger, L. (2014a). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 - Part II: Homogeneity adjustments for pilot balloon and radiosonde wind data. *Earth Syst. Sci. Data*, 7:335–383.
- Ramella Pralungo, L. and Haimberger, L. (2014b). New estimates of tropical mean temperature trend profiles from zonal mean historical radiosonde and pilot balloon wind shear observations. *J. Climate*, submitted.

- Ramella Pralungo, L., Haimberger, L., Stickler, A., and Brönnimann, S. (2014). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 - Part I: Merging and interpolation to 0:00 and 12:00 GMT. *Earth Syst. Sci. Data*, 6:185–200.
- Seidel, D., Angell, J., Christy, J., Free, M., Klein, S., Lanzante, J., Mears, C., Parker, D., Schabel, M., Spencer, R., Sterin, A., Thorne, P., and Wentz, F. (2004). Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets. *J. Climate*, 17:2225–2240.
- Tavolato, C. and Isaksen, L. (2011). Data usage and quality control for ERA-40, ERA-Interim and the operational ECMWF data assimilation system. *ERA report series*, 7.
- Tavolato, C. and Isaksen, L. (2014). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Q. J. Roy. Meteor. Soc.* published online, in press.
- Tavolato-Wötzl, C. (2015). Variational bias correction for radiosonde wind direction. Technical report, ECMWF.
- Uppala, S. M., Kållberg, P. W. and Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Hasler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E. V., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. (2005). The ERA-40 re-analysis. *Q. J. Roy. Meteor. Soc.*, 131(612):2961–3012. doi:10.1256/qj.04.176.



## **Acknowledgements**

This work would not have been possible without the contributions of many. Some helped with the scientific work and interesting discussions about the topic others supported me with encouragement and words of support. Therefore I would like to thank you all, especially:

Words of thanks have to go to my supervisor Dr. Leopold Haimberger who encouraged me to do this work, supported me while working in his group at the University of Vienna and who always found the time to answer my questions and discuss interesting aspects. Thank you!

I have great gratitude for my local supervisor Lars Isaksen during my stay as a Graduated Trainee at the ECMWF. I learned a lot during our discussions and found a lot of motivation in his encouragement.

Next to them I really appreciated all the conversations I had with the staff at the University of Vienna and ECMWF which lead to interesting new ideas and provided a different view on things.

Throughout my work my family and friends have been supportive, they listened to me, they proofread my work and provided helpful comments - even sometimes with no understanding of the topic I was talking or writing about - and encouraged me to finish this thesis. Without them the last years would not have been as they were! So this is for all of you who believed in me and realised that talking to me about the weather has nothing to do with polite small talk.





## CURRICULUM VITAE

### PERSONAL DETAILS:

Name: Christina Margareta Tavolato-Wötzl  
Citizenship: Austrian

### ACADEMIC EDUCATION:

2001 – 2006: Study of Meteorology at the University of Vienna  
October 2006: Degree in Meteorology *Magistra der Naturwissenschaften*  
2007, 2009 – 2015: Doctoral study in Meteorology at the University of Vienna

### PROFESSIONAL EXPERIENCE:

Since 2012: Scientific worker at MeteoServe GmbH  
2009 – 2012: PhD. position at the project "Global in situ upper air data for climate change research" (PI: Leopold Haimberger, FWF - Project: P21772-N22) at the University of Vienna  
2009 – 2011: Lecturer for undergraduate studies at the University of Vienna (Problems in statistic, theoretical meteorology, satellite meteorology)  
2007 – 2009: Graduate Trainee at the European Centre for Medium-range Weather Forecast (ECMWF) within the data assimilation section  
2005 – 2007: Scientific worker at the project "Homogenization of the global radiosonde temperature dataset" (PI: L. Haimberger, FWF - Project: P18120-N10) at the University of Vienna

### SELECTED PRESENTATIONS:

DACH 2013: Poster Presentation: C. Tavolato, R. Kaltenböck: *Monitoring der Dual-Pol Radare in Österreich* - in german.  
ICR4 2012: M. Milan, C. Tavolato and L. Haimberger: *Variational bias correction for radiosonde data*  
EGU 2011: Poster presentation: C. Tavolato and L. Haimberger: *Variational bias correction for radiosonde wind direction bias*  
DACH 2010: C. Tavolato and L. Haimberger: *Variationelle Fehlerkorrektur für Windrichtungsfehler von Radiosondendaten* - in german.  
RMetS Meeting 2009: L. Isaksen and C. Tavolato: *Accounting for non-Gaussian observation error*  
EGU 2007: C. Tavolato and L. Haimberger: *Global mean stratospheric warm bias of 1K of radiosonde temperatures in the 1980s* - invited as a solicited talk.

#### SELECTED PUBLICATIONS:

- TAVOLATO-WÖTZL, C., 2015: *Variational Bias Correction for Radiosonde Wind Direction* - technical report, ECMWF.
- TAVOLATO, C., and ISAKSEN, L., 2014: *On the use of a Huber norm for observation quality control in the ECMWF 4D-Var* - Q. J. Roy. Meteor. Soc., published online
- HAIMBERGER, L., TAVOLATO, C. and SPERKA, S., 2012: *Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighbouring stations* - J. Climate: 25, 8108-8131
- DEE, D. P., UPPALA, S. M., SIMMONS, A. J., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M. A., BALSAMO, G., BAUER, P., BECHTOLD, P., BELJAARS, A. C. M., VAN DE BERG, L., BIDLOT, J., BORMANN, N., DELSOL, C., DRAGANI, R., FUENTES, M., GEER, A. J., HAIMBERGER, L., HEALY, S. B., HERSBACH, H., HÓLM, E. V., ISAKSEN, L., KÅLLBERG, L., KÖHLER, M., MATRICARDI, M., McNALLY, A. P., MONGE-SANZ, B. M., MORCRETTE, J.-J., PARK, B.-K., PEUBEY, C., DE ROSNAY, P., TAVOLATO, C., THÉPAUT and J.-N., VITART, F., 2011: *The ERA-Interim reanalysis: configuration and performance of the data assimilation system* - Quarterly Journal of Roy.Met.Soc.: 137, 553-597
- LADSTÄDTER, F., STEINER, A. K., FOELSCH, U., HAIMBERGER, L., TAVOLATO, C. and KIRCHENGAST, G., 2011: *An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes, and GPS radio occultation* - Atmos. Meas. Tech. Discuss.: 4, 2127-2159
- TAVOLATO, C. and ISAKSEN, L., 2011: *Data usage and quality control for ERA-40, ERA-Interim and the operational ECMWF data assimilation system* - ERA report series, No.7
- HAIMBERGER, L., TAVOLATO, C. and SPERKA, S., 2008: *Towards elimination of the warm bias in historic radiosonde temperature records - some new results from a comprehensive intercomparison of upper air data* - J. Climate: 21, 4587-4606.
- TAVOLATO, C., 2006: *Vergleich globaler korrigierter Radiosondentemperaturdatensätze mit MSU-Satellitendaten* (Diploma Thesis, Supervisor: Prof. Dr. Michael Hantel) - in german.