ERA report series

22   Predictors and grouping for variational bias correction of radiosonde

Marco Milan and Leopold Haimberger

Abstract

Due to various causes, artificial biases can be found in radiosonde temperature observations. The use of biased observations in the analysis leads to systematic errors unless special measures are taken during the data assimilation. This is especially important for the production of reanalysis which are supposed to yield credible trends and low frequency variability.

One such method that avoids the assumption of unbiased input observations is the variational bias correction (VarBC), which is used successfully within the ECMWF (European Centre for Medium-range Weather Forecasts) operational system, mainly to deal with biases in satellite radiance data. In VarBC the bias of the given observation is estimated using a linear predictor model based on a small number of predefined predictors and the corresponding unknown bias parameters. These are estimated together with the model state by including a bias term in the cost function of the variational analysis. Hence VarBC adjusts the bias of groups of observations using the whole state of the atmosphere described by the analysis given from a 4D-Var assimilation system. The resulting corrections are consistent with the other assimilated observations and the model physics.

The radiosonde temperature biases depend mainly on pressure, on solar elevation and on the instrumentation used. The optimal choice of the grouping of radiosonde stations (to get larger samples) and of the bias models is not obvious. While the method should be used in a 4D-VAR setting, its properties can be estimated offline with much less computational effort. In this paper different methods for the grouping and the bias model are tried. Two grouping methods are based on the radiosonde instrumentation metadata, one with and one without clusterization, and a clusterization based only on the radiosonde profile data without metadata. A test without any grouping is also done. At the same time the statistics are compared with the output of two already existing homogeneity adjustment algorithms. The majour outcome of this work is, apart the development of predictors model suitable for VarBC, the detection of the high variability in the bias using the approach based on metadata.

1   Introduction

The existence of biases in the radiosonde temperature data, especially in the upper levels, is extensively described in the literature (Sherwood et al. [2005], Sherwood [2007], Haimberger et al. [2008]). Since they vary in time they also affect the temperature trends in products depending on them (Santer et al. [2005],Dee et al. [2011]). Changes in instrumentation often cause "jumps" of more than $1\,K$ in the radiosonde records. For this reason, a bias correction before or during the assimilation of these data is essential.

The biases in the radiosonde data can be estimated before climate analysis (Haimberger [2007], Haimberger et al. [2012]). The adjusted data should be homogeneous in space and time. This type of adjustments is even more important in the pre-satellite era, because in that period the radiosonde and balloon network is practically the only upper air observing system. Aircraft measurements were quite rare at that time.

The classical homogeneity adjustment approach as it is used in climatology (e.g. Venema et al. [2008]) is not the only possibility to reduce biases. In the context of global data assimilation an automatic procedure is needed for these adjustments. The Variational Bias Correction (VarBC, Derber and Wu [1998] ; Dee [2005], described in Section 2) estimates the systematic errors in the observations using a 4D-Var method with an additional correction term for the bias. In this way, the bias correction depends not only on the raw data itself but on the whole analysis, which means all data assimilated and the model. The method has worked quite well for satellite data and aircraft data but has so far not been applied to radiosondes.

In order to prepare for a VarBC of radiosonde temperatures, predictors and grouping methods for radiosondes are tested offline (outside the assimilation cycle) using existing background departures from existing reanalysis (ERA-Interim Dee et al. [2011] and ERA-40 Uppala et al. [2004]). In ERA-40 lookup tables have been used for radiosonde bias correction (Andrae et al. [2004]). In ERA-Interim adjustments from RAOBCORE_T_13 (Haimberger et al. [2008]) combined with a solar angle dependent bias correction were used.

The bias correction in the previous reanalysis from ECMWF (Andrae et al. [2004]), has defined some criteria which we also consider in the VarBC approach:

- Dependence of the bias on the solar elevation (Sherwood et al. [2005]).

- Dependence of the bias on the instrumentation. I.e., different instrumentation leads to different biases (Gaffen [1994]), which leads also to different corrections for different periods for the same radiosonde station. The variability of the station's biases, which uses the same instrumentation will be checked.

- The bias correction is not applied for the lower levels (in our case below 850hPa), at this height the radiosonde temperature is taken as reference.

- Grouping of more stations with similar characteristics in the instrumentation for better statistics.

The grouping was based on the country where the station was located, using the station ID. The assumption was that in the same country and in the same period similar radiosondes were used, which is sometimes untrue, e.g. over the US from the late 1980s onwards (Haimberger et al. [2008]).

In this paper grouping of the radiosondes is based on the metadata provided by Schroeder [2003], which have been updated until 2013. In this data base more than 3000 kinds of radiosonde instrumentation are documented. To each possible instrumentation setting a number is linked, in order to indicate the sonde_type. Using this library, in most of the cases (more than 90% after 1979), the radiosonde type can be identified by the station ID and the day considered. For each year, the number of used sonde_types might be large (above 80). If we identify a group with a sonde_type, in some cases a single group may have only a small statistic. For this reason, different sonde_types with a similar bias are gathered together using a clustering approach (CL_STYPE), described below. The linear regression which takes into account the sonde_types without clusterization will be called STYPE. A source of difficulties in the clustering approach could be the existence of discontinuities in the bias correction related to the change of the year: a sonde_type can belong to different clusters in two adjacent years. This may lead to different corrections, because the correction is based on the cluster mean. The existence of these situations is tested by looking at the offline monthly bias correction over the whole considered period. The goal of this work is to estimate the mean departure between the temperature in the first guess and the observations (FG). In order to consider the approximation due to the clustering, the computation of the FG is done also without using the clustering; each sonde_type is considered individually.

During this work, some inconsistencies were found in the assumption that radiosondes with the same instrumentation have similar bias. It appears that the metadata information provided in Schroeder's database is not sufficient for describing the bias characteristics. For this reason this hypothesis is tested doing a clusterization not based on the sonde_type but directly on the mean background departures (clustering by station ID, CL_STATID). In this case all statistics are independent from the metadata, which is an advantage in the early years where essential metadata are often missing. In order to check the existence of possible inhomogeneities due to clusterization an approach without any grouping is done as well (STATID, where the data are referenced only to the station ID).

This work is focused on radiosonde data after 1979, when ERA-Interim background departures are available and an offline correction is possible using these data as reference. Before this period, radiosondes are, together with the surface data, the most important data source in the assimilation; consequently the assimilated temperature field is strongly influenced by this type of data. An accurate study of the assimilation chain with and without the satellite data, in order to evaluate the model bias, is desirable; such a study is really important in preparation for the application of the VarBC for radiosonde temperature profiles before the satellite era. This issue will be tested in future work.

In this paper after a brief introduction of the VarBC (Section 2), the data and the method used for the offline study is described in Section 3. Results using CL_STYPE and STYPE are shown in Section 4, as well as the comparison to CL_STATID and STATID. The results will also be compared with background forecasts from ERA-Interim, the current reanalysis of the ECMWF and with the output of a homogenization algorithm based only on neighbouring stations (RICH Haimberger et al. [2008]). Conclusions are drawn in Section 5.

2 VARIATIONAL BIAS CORRECTION

As extensively described in the literature (Lorenc [1986], Zupanski [1997], Daley [1991], Bouttier and Courtier [1999], Kalnay [2003]) the four dimensional variational data assimilation (4D-VAR) is an efficient method of data assimilation, which tries to determine the "true" atmospheric state using a numerical weather prediction (NWP) model and a set of observations. A useful application of 4D-VAR is the generation of a global analysis, based on historical data adding new data information and using more recent models. Well known examples of reanalysis are ERA-40 (Uppala et al. [2005]) and ERA-Interim (Dee et al. [2011]), both from ECMWF.

The VarBC at ECMWF is used in a 4D-Var framework, for simplicity a description using 3D-Var is given. The model state is estimated by minimizing the cost function (Barker et al. [2003]):

$$
\begin{aligned}
J(\mathbf{x}) = \quad & \tfrac{1}{2}(\mathbf{x_b} - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x_b} - \mathbf{x}) \\
& + \tfrac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))
\end{aligned}
\tag{1}
$$

where $\mathbf{x}$ is the unknown model state, $\mathbf{x_b}$ the background estimate, $\mathbf{y}$ the observations, the function $\mathbf{h}(\mathbf{x})$ is the observation operator. $\mathbf{R}$ and $\mathbf{B}$ are the observation and the background error covariance matrix. In the cost function the model and the observations are assumed unbiased.

The assumption of the unbiased observation error ($\mathbf{e}_o$) can be written as:

$$
< \mathbf{e}_o > = < \mathbf{y} - h(\mathbf{x}) > = 0
\tag{2}
$$

Since observations are often biased one needs to avoid this assumption by introducing a method to correct the bias embedded in 4D-VAR. The basic idea in VarBC (Derber and Wu [1998] ; Dee [2005]; Dee [2004]) is the assumption that the systematic observation errors can be described in terms of a limited set of parameters β. In this case the expected value of the observation error can be described by:

$$\begin{array}{rcl} < \mathbf{e}_o > & = & < \mathbf{y} - h(\mathbf{x}) > \neq 0 \\ < \tilde{\mathbf{e}}_o > & = & < \mathbf{y} - \tilde{h}(\mathbf{x}, \beta) > = 0 \end{array} \tag{3}$$

The modified observation operator incorporating the bias uses a linear prediction model based on predefined predictors:

$$\tilde{h}(\mathbf{x}, \beta) = h(\mathbf{x}) + \sum_{i=0}^{N} \beta_i p_i \tag{4}$$

where the p_i are the predictors and the β_i are unknown bias parameters associated with atmospheric variables to update, i.e., in our case the temperature profile.

The control vector $\mathbf{z}$ for the formulation of the cost function is also defined using the parameters β:

$$\mathbf{z}^T = \left[\mathbf{x}^T, \beta^T \right] \tag{5}$$

As usual in variational assimilation, a prior estimation of the analysed field is used. In this sense, the background estimation of the augmented control vector is:

$$\mathbf{z}_b^T = \left[\mathbf{x}_b^T, \beta_b^T \right] \tag{6}$$

where β_b is obtained from a previous analysis of the cycle. The Eq.1 can now be rewritten, integrating the bias terms:

$$\begin{array}{rcl} J(\mathbf{z}) & = & \dfrac{1}{2}(\mathbf{z_b} - \mathbf{z})^T \mathbf{Z}^{-1}(\mathbf{z_b} - \mathbf{z}) \\ & + & \dfrac{1}{2}(\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z}))^T \mathbf{R}^{-1}(\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z})) \end{array} \tag{7}$$

VarBC incorporates some degrees of freedom in the observation operator in order to reduce the bias:

$$\begin{array}{rcl} \tilde{h}(\mathbf{z}) & = & \tilde{h}(\mathbf{x}, \beta) \simeq \tilde{h}(\bar{\mathbf{x}}, \beta) + \mathbf{H}(x - \bar{x}) = h(\bar{x}) + \sum_{i=0}^{N} \beta_i p_i + \mathbf{H}(x - \bar{x}) \\ \mathbf{H} & = & \left(\dfrac{\partial h}{\partial x} \right)^T \end{array} \tag{8}$$

where $\bar{\mathbf{x}}$ is the latest outer loop of $\mathbf{x}$ and the non-linear observation operators are linearised using $\bar{\mathbf{x}}$. The term $\mathbf{H}$ is the Jacobian of the observation operator. In this way the modification to $h(\mathbf{x})$ is additive and linear in the bias parameters. This property is very useful in the computation of the adjoint.

Lacking enough information about the dependence of state and bias parameter background errors VarBC assumes that the parameter (the βs) estimation errors are not correlated with the state estimation errors, therefore $\mathbf{Z}$ is defined as:

$$\mathbf{Z} = \left(\begin{array}{cc} \mathbf{B}_x & 0 \\ 0 & \mathbf{B}_\beta \end{array} \right) \tag{9}$$

where $\mathbf{B}_x$ is the background error covariance matrix (as defined in 4D-VAR), and the additive term $\mathbf{B}_\beta$ the parameter background error covariance. The cost function therefore simplifies to:

$$\begin{aligned} J(\mathbf{x},\beta) = \quad & \tfrac{1}{2}(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}_x^{-1}(\mathbf{x}_b - \mathbf{x}) \\ & + \tfrac{1}{2}(\beta_b - \beta)^T \mathbf{B}_\beta^{-1}(\beta_b - \beta) \\ & + \tfrac{1}{2}(\mathbf{y} - \tilde{h}(\mathbf{x},\beta))^T \mathbf{R}^{-1}(\mathbf{y} - \tilde{h}(\mathbf{x},\beta)) \end{aligned} \tag{10}$$

The $\mathbf{B}_\beta$ is, as said before, uncorrelated with the state estimation errors and it is taken as diagonal. In other words, it depends only on the variance of the observation error, with:

$$\sigma_{\beta_j}^2 = \frac{\sigma_O^2}{N_j} \tag{11}$$

where β_j is the jth bias parameter (predictor) associated with a set of observations in its bias group. N_j is a positive integer which governs how fast the bias adjustment is. In this way, the part in the cost function (Eq. 11) related to the bias parameter background constraint for the jth variable has the same weight as N_j new observations.

3   Data and Model

Before VarBC for radiosonde temperature profiles in the reanalysis is applied, the most suitable predictors and the groups of radiosonde with similar statistics have to be defined. We decided to develop an offline method in order to determine the optimal setting of the parameters. The estimation of the parameter in Eq. 4 does not need to be take place within a 4D-VAR assimilation run but can be done using background departures obtained from a previous reanalysis. The output of this work can be applied in a successive online version and can be stored in an Observation Feedback Archive (OFA, Hersbach et al. [2015])

The input data are temperature and background departures from the ERA-INTERIM reanalysis. In VarBC, the model is assumed unbiased and for this reason the radiosonde bias can be approximated with the mean difference between the radiosonde and the model temperature profile. The goal of this study is to find the best approximation of the radiosonde bias based on a multiple linear regression for different groups.

Each sonde is identified by a station ID and, especially in the recent years, the instrument type is regularly given. If the meta data for the instrumentation is not available the algorithm sets the sonde_type number equal to 0 (Unknown instrument). The number of "Unknown" data can vary between 4% and 10% of the total amount of data in the year.

The radiosonde temperature bias depends on the characteristics of each instrument (Andrae et al. [2004]). The ERA-INTERIM data are analysed from 1979 until 2013. In this period more than 3000 different sonde_types may occur; the complete metadata list is provided by Schroeder [2003].

In order to reduce the number of the possible groups and to enhance the statistics for the application of the VarBC, two different types of clusterization are tested. The first one (CL_STYPE) groups different sonde_types together; the different sonde_types in one group will have similar bias profiles (see Section 3.2) The second one (CL_STATID) groups directly different station IDs without taking into account the metadata. These groupings are performed on a yearly basis in order to reflect the evolution of the radiosondes instruments during the period considered. At the same time the presence of 'jumps' between years for the same station ID may be a source of inconsistencies for VarBC. The existence of such inconsistencies will be investigated in the successive sections, using the computation of monthly biases and their trend, and comparing the bias of a sonde_type in a cluster (CL_STYPE) with the bias of the same sonde_type without using the clustering (STYPE). The same test is done by comparing the bias of a station ID in a cluster (CL_STATID) with the bias of the same station without using the clustering (STATID).

The predictors are chosen among the available quantities the bias depends on, taking into account the physical origins of the bias. The systematic error of the radiosonde temperature profile depends on the pressure level where the temperature is observed, with an enhancement in the stratosphere (Haimberger [2007] ; Randel and Wu [2006]). The radiosonde temperature profile depends also on solar elevation at the time of the sounding, because of the instrumental characteristics of the sonde_type; one example is described in Luers [1997]. Generally, the effect of solar radiation is stronger at the upper troposphere and upwards (Onogi [2000]); for this reason we will apply the solar correction only in the higher levels (pressure lower than 200hPa). Furthermore the bias of the temperature profiles shows a sharp transition between day and night, which has to be considered in the bias model. We do not consider the dependencies of the temperature bias on moisture or on the lapse rate since both predictors have relatively large uncertainties on their own. All predictors are explained in the next section.

3.1   Predictors

For the VarBC, we search for the smallest set of predictors that still fits the predictand (the FG) within an acceptable error margin. These general 'rules' should be kept in mind during the predictor determination.

Four pressure predictors $p_i(P)$ (Fig. 1), are used for four different pressure layers; the predictor itself can have values between 0, when it is not applied, and 1, when it is applied completely. Between the layers, a linear interpolation is done, in order to avoid discontinuities. In this way, for each level, the predictors have the characteristic:

$$\sum_{i=0}^{3} p_i(P) = 1 \qquad (12)$$

The layers are chosen as:

- Lower levels, pressure up to 700hPa (p_3)

- From 150hPa to 850hPa. (p_2)

- From 30hPa to 250hPa. (p_1)

- Upper levels, Pressure below 60hPa. (p_0)

Using independent predictors for different layers has the advantage that the larger bias estimation errors in the stratosphere cannot influence the bias estimates in the lower layers, which would be the case when choosing linear or polynomial regression model. Another reason for this model is the often complicated shape of the bias profiles which cannot be well approximated by low order polynomials. Only in the upper levels (between 200hPa and 10hPa) we tested a linear (in log p) predictor in order to try to reproduce the increase of the solar radiation error with height in these layer (Fig. 2).
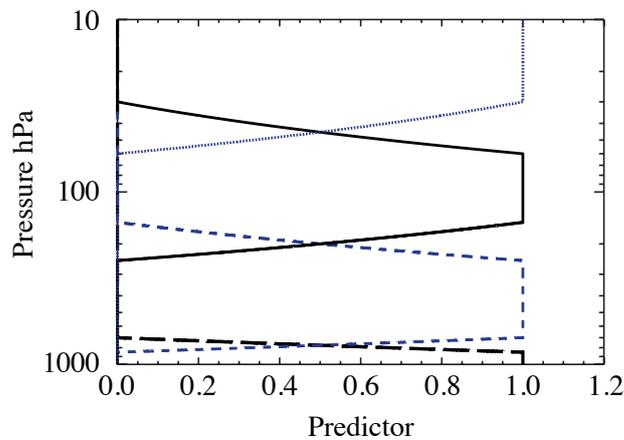
Figure 1: Pressure predictors.
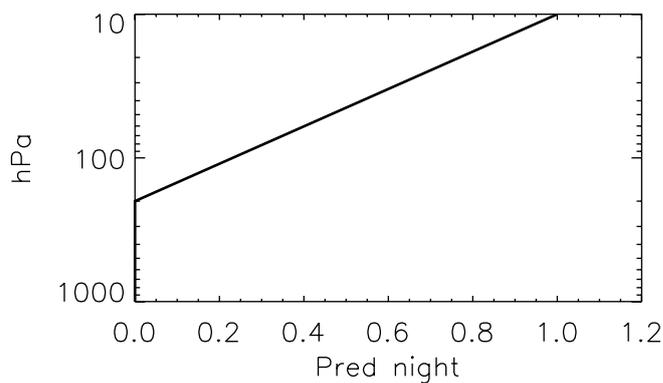
Figure 2: Pressure predictor plog.

The logarithmic predictor (taking into account only the data where the pressure is equal/higher than 10hPa) is set using:

$$plog' = \begin{cases} log_{10}(P/200) & 10\,hPa \leq P \leq 200\,hPa \\ 0 & \text{elsewhere} \end{cases} \tag{13}$$

and normalized using

$$plog = \begin{cases} \frac{|plog'|}{max(|plog'|)} & 10\ hPa \leq P \leq 200\ hPa \\ 0 & \text{elsewhere} \end{cases} \tag{14}$$

The solar elevation is computed using the launching time and the position of the radiosonde (adding one hour in order to give the radiosonde the time to reach the higher levels). For this predictor a polynomial function is chosen. For all given radiosonde types, we computed the polynomial fit (with order from one to five) between FG and the solar elevation, in the higher levels. In the bias models which use the solar elevation, the angles between -7.5^o and 90° are used. The value of -7.5^o for the solar elevation is chosen in order to take into account scattered solar radiation after sunset / before sunrise; the maximal value of solar elevation is 90°, no differences are present between sunrise and sunset; given $\theta_0 = -7.5^o$, the normalized predictor of order n is:

$$\theta'^n = \frac{\theta^n - \theta_0^n}{90^n - \theta_0^n} \tag{15}$$

Negative values of solar elevation smaller than -7.5^o are treated as "night-time", with $\theta = -7.5^o$ ($\theta' = 0$). This is because the bias does not depend on solar angle during the night.

One disadvantage of this approach is that the hour of the sounding must be available, this is often not the case especially for the early data. In these cases the solar elevation must be guessed or the dependence of the bias on solar elevation must be disregarded.

The solar elevation predictors are multiplied by an another predictor p_{day}, which uses a linear interpolation between 0 and 1 in the pressure levels between 200hPa and 100hPa and is equal to 1 below 100hPa. An example of the approximation for a Russian radiosonde type (number 1963, name RK5, in the Schroder's table, used from 1979 until 1994), using a polynomial fit of order three, is shown in Fig. 3. In this example the ability of the third order polynomial to follow the FG data is visible. Only the peak around 20 degree is not well represented. For this sonde_type the root mean square difference (RMSD) between data and fit is equal to $1.4 \cdot 10^{-2}$ K. The error bars in Fig. 3 indicate the standard error of the mean of the sample of n independent observations [Wilks, 2011, eq. 5.4].
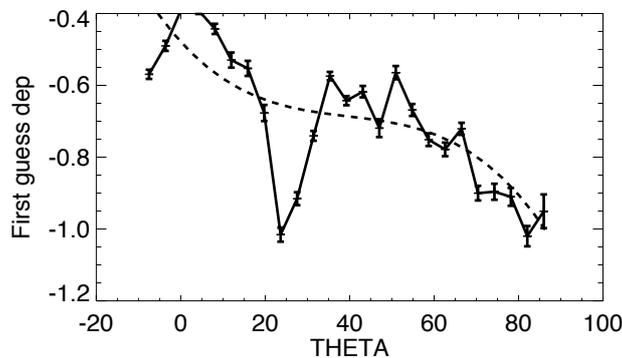
Figure 3: Example polynomial fit for a radiosonde type RK5 (sonde_type 1963), data in the stratosphere, pressure below 200hPa.

Table 1 shows the mean of the root mean square differences (RMSD) between the fits and the data for different polynomial order. The means of the estimated coefficients with their standard deviation are also

	pol. degree 1	pol. degree 2	pol. degree 3
Const	-0.5 ± 1.1	0.95 ± 58.00	$(-2.7 \pm 39.0) \cdot 10^2$
a_1	$(-0.53 \pm 3.30) \cdot 10^{-2}$	-0.12 ± 3.10	21 ± 330
a_2		$(2.0 \pm 0.43) \cdot 10^{-3}$	-0.59 ± 9.2
a_3			$(0.54 \pm 8.6) \cdot 10^{-2}$
RMSD	0.23	0.20	0.18
	pol. degree 4	pol. degree 5	
Const	$(-1.2 \pm 41.0) \cdot 10^3$	$(-1.7 \pm 250.0) \cdot 10^3$	
a_1	$(1.4 \pm 43.0) \cdot 10^2$	$(3.7 \pm 41.0) \cdot 10^2$	
a_2	-6.3 ± 170.0	$(-3 \pm 39.) \cdot 10^1$	
a_3	0.12 ± 3.00	1.2 ± 16.0	
a_4	$(-0.086 \pm 2.000) \cdot 10^{-2}$	$(-2.1 \pm 29.0) \cdot 10^{-2}$	
a_5		$(-0.17 \pm 2.50) \cdot 10^4$	
RMSD	0.16	0.15	

Table 1: Mean of RMSD between fit and data for various polynomial degree of solar elevation. For every term of differrent polynomial the coefficients relative to the degree are given, $a_1 \rightarrow$ first, $a_2 \rightarrow$ second, ...

listed. The best fit with the lowest number of solar angle related predictors is obtained with a polynomial of order three, which means three new predictors for the stratosphere: θ', θ'^2, θ'^3; where θ' is the predictor related to solar elevation.

We tested different combination of predictors using different bias models:

1. linear

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 \tag{16}$$

2. log

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_4 plog \tag{17}$$

3. angle

   During the day

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_4 (\theta' p_{day}) + \beta_5 (\theta'^2 p_{day}) + \beta_6 (\theta'^3 p_{day}) \tag{18}$$

   During the night:

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_7 (plog) \tag{19}$$

4. anglesllog

   During the day

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_4 (\theta' plog) + \beta_5 (\theta'^2 plog) + \beta_6 (\theta'^3 plog) \tag{20}$$

   During the night:

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_7 (plog) \tag{21}$$

5. anglelog

   During the day

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_4 (\theta' p_{day}) + \beta_5 (\theta'^2 p_{day}) + \beta_6 (\theta'^3 p_{day}) + \beta_7 plog \tag{22}$$

During the night:

$$\mathbf{B} = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_3 + \beta_7(plog) \tag{23}$$

The background departures for STATID are used in this test, in this way any secondary effect due to grouping is avoided. We want to test only the effect of the different bias models.

For the 5 different bias models the linear regression is applied to the radiosonde data from 1980 to 2010, as whole set of data also for models with different predictors for day and time, i.e. model *angle* and model *anglesllog*. For each year the RMS of the residual bias is computed ($FG - approximation$), for different thickness of pressure (Fig. 4) as well as the effect of the approximation for different solar elevation classes:

- $\theta \leq -7.5^o$, night.

- $\theta \in]-7.5^o, 7.5^o]$, sunrise/sunset

- $\theta \in]7.5^o, 22.5^o]$

- $\theta > 22.5^o$

with θ the solar elevation angle.
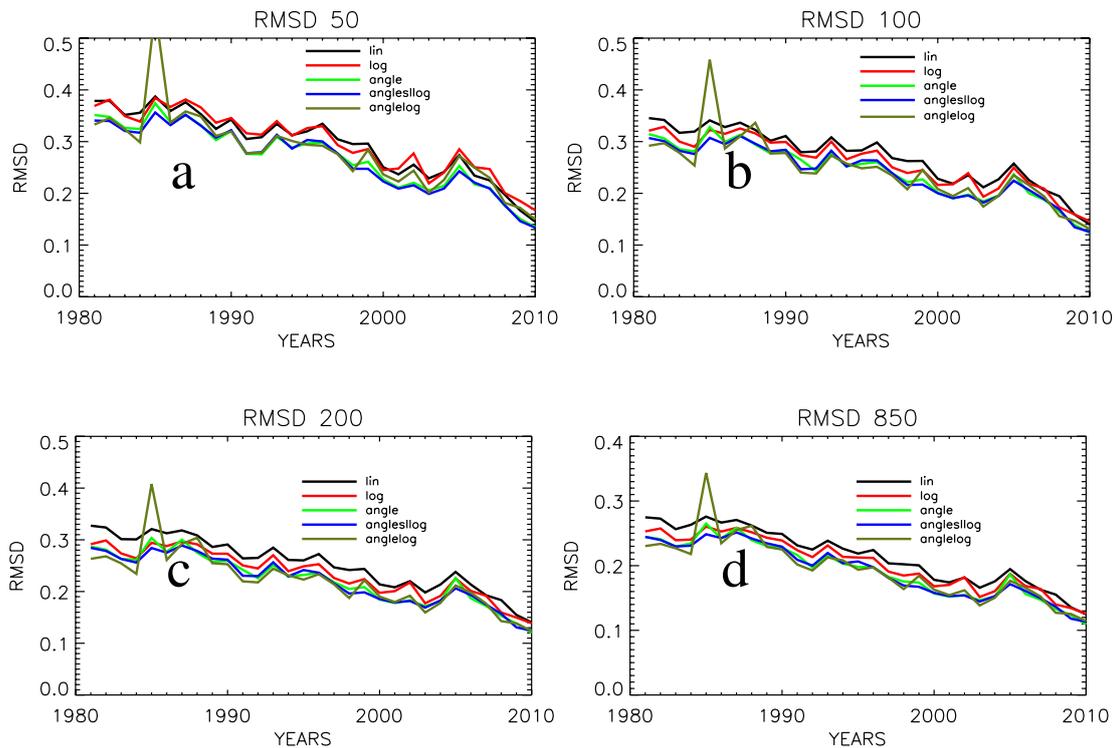
Figure 4:   RMSD of the different approximation algorithms described in Section 3.1. For different thickness of pressure: **a)** [50 : 10]hPa; **b)** [100 : 10]hPa; **c)** [200 : 10]hPa; **d)** [850 : 10]hPa.

For the data, where the solar elevation is in one of the classes, the residual bias is extracted in order to show the error in the bias correction for the chosen solar elevation class (example for the higher levels in Fig. 5). These classes were defined during the ERA40 reanalysis (Andrae et al. [2004]).
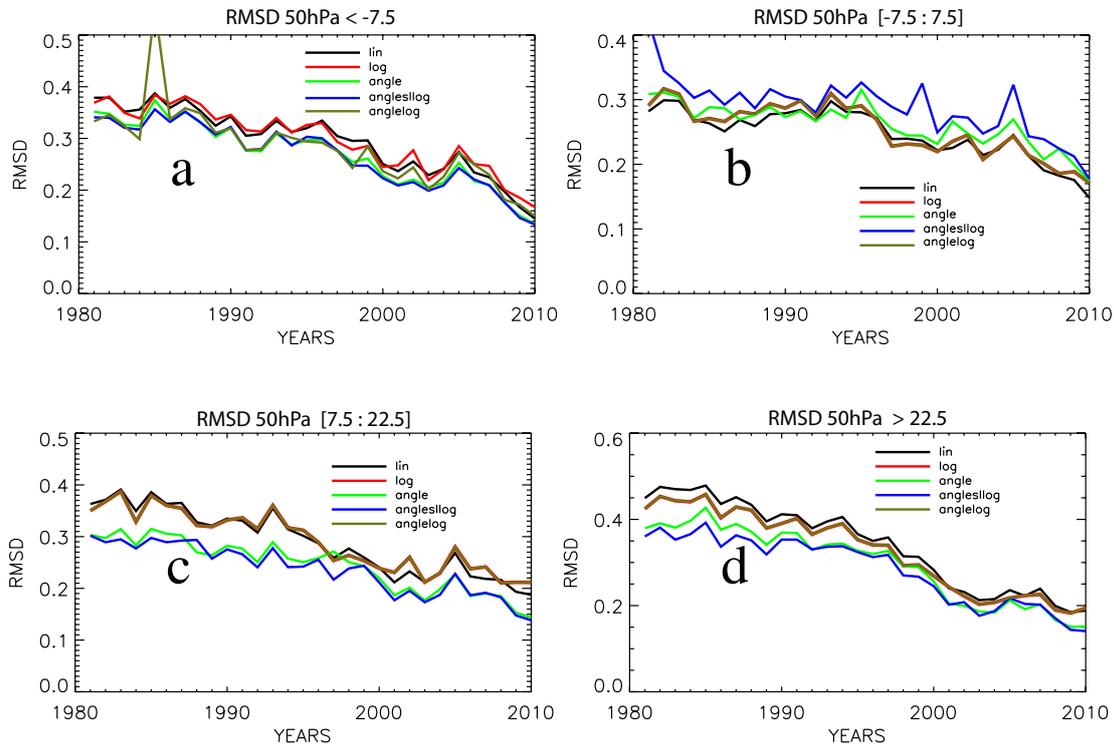
Figure 5:   RMSD of the different approximation algorithms described in Section 3.1. For pressure between 50 and 10hPa, for different solar elevation ranges: **a**) night; **b**) [−7.5 : 7.5]; **c**) [7.5 : 22.5]; **d**) angle greater than 22.5 degree.

Generally, for all pressure layers, better performance is achieved when using the models *anglesllog* and *angle*. Looking to the different solar elevation classes, *anglesllog* performs slightly better than *angle* for the higher solar elevation angles (Fig. 5d), but during the transition between night and day (Fig. 5b) anglesllog shows an inconsistency and a higher RMS. From these results we assume that the more suitable and robust model for the predictors determination is *angle*.

3.2   Radiosonde clustering

Cluster analysis is applied to form groups with similar characteristics (Wilks [2006]) from a pool of objects (in our case the temperature profiles for different sonde_types). It can use hierarchical and non-hierarchical methods (for a distinction between the two approaches as well as the definition of the linkages see Wilks [2006]); here, the first method is used. The applied clustering defines the degree of similarity, where the difference between the temperature profile is defined using a distance (metric, Section 3.2.1). Furthermore, it is used to define the number of clusters and to assign the members to the correct cluster. An overview of cluster analysis is presented in Romesburg [1989]. We first choose a fixed distance *dmax* (using the metric). The members, which are less distant than *dmax*, will be assigned to the same cluster. *dmax* is chosen such that the number of clusters lies within reasonable bounds. We used this technique in order to have continuity in the cluster definition over different years. In this sense, for the case of CL_STYPE, a sonde_type will be grouped in a cluster with similar FG, and the difference between different years should be not larger than the monthly variability of FG. The same is valid for CL_STATID.

3.2.1   Metric

The quantity analysed is the mean FG, i.e., the mean difference, for each pressure level, between the radiosonde temperature and the model temperature.

The distance between the objects in the pool is of primary interest in clustering, because of the necessity to group similar members. For this reason the choice of the metric can lead to different results. The metric chosen has to be linked to the similarities which we want to represent.

In the case of radiosonde temperature profiles, the upper levels are more interesting than the troposphere, given the larger biases present in the stratosphere. The data below 850hPa are not corrected and thus not included in the computation of the metric. This is consistent with our desire not to correct the radiosonde's bias in the boundary layer. Due to the different biases of the radiosonde temperature for different pressure ranges, a weighted Euclidean distance for different levels was chosen.

In the first step, the FG for each sonde_type/station ID in the considered year is divided into three layers:

- From 850hPa to 150hPa, troposphere.

- From 150hPa to 50hPa, first stratospheric layer

- From 50hPa to 10hPa, second stratospheric layer; no or very sparse data, are present above 10hPa

For each layer, taking one sonde_type/station ID, the differences between its mean FG and the mean of the FG for each other sonde_type/station ID are computed. At this point the weighted Euclidean distance is calculated. Taken the distance between two temperature profiles, x and y, a weight w_j is assigned to the j_{th} variable, following:

$$dt_{xy} = \sqrt{\sum_{m=1}^{M} w_m \left(x_m - y_m\right)^2} \qquad (24)$$

In our case, the M variables represent the three layers and the weights are computed taking into account the relative thickness of the layer (in meter) and the importance of the layer in the clusterization, using empirical values :

- troposphere: $w'_1 = ln(\frac{850}{850}) - ln(\frac{150}{850})$

- first stratospheric layer: $w'_2 = \left(ln(\frac{150}{850}) - ln(\frac{50}{850})\right) \cdot 5$

- second stratospheric layer: $w'_3 = \left(ln(\frac{50}{850}) - ln(\frac{10}{850})\right) \cdot 7$

The weights are then normalized, using:

$$w_i = \frac{w'_i}{\sum_{m=1}^{3} w'_m} \qquad (25)$$

The FG depends also on the solar elevation. For this reason we have included also the solar elevation in the metric computation. The same metric as for the whole FG is computed four times for four different solar elevation classes (described in Section 3.1). Unfortunately, the sounding time, and thus the solar

elevation, is not always available; for this reason the data for the metric related to the solar elevation (d_θ) is used only when the hour of sounding is available.

For each solar elavation range, first the same metric as in eq. 24 for every solar elevation is used:

$$d\theta_0 i_{xy} = \sqrt{\sum_{m=1}^{M} w_{mi} \left(x_{mi} - y_{mi}\right)^2} \tag{26}$$

With i the index of the solar elevation range taken into account. In different solar ranges also different numbers of data are present; it is weighted using:

$$d\theta_{xy} = \sum_{i=1}^{4} d\theta_0 i_{xy} \frac{N_i}{N_T} \tag{27}$$

With N_i the number of data present in one solar elevation range, and N_T the total number of data used. In the case d_θ exists ($N_T \neq 0$), the metric used (d) is a mean of dt and $d\theta$.

3.2.2 Clustering technique

Once the metric is defined, in the clustering technique a matrix of the distances can be defined. This symmetric matrix has the dimension $n \times n$, where n is the number of objects in the pool. In each line (same object) all distances to the other objects in the pool are written, the diagonal elements of the matrix are obviously zeros. The distance between two different clusters is not uniquely defined. We used the centroid, i.e., the average distance across all the objects in the cluster. The clusterization is now performed in the following steps:

1. Identification of the two objects, having the shortest distance.

2. Merging of the objects in point 1 into one cluster.

3. Computation of the centroid: the new cluster is linked to the distance to other clusters/objects defined by the centroid.

4. Elimination of the lines and columns related to the two merged objects, from the matrix of the distances.

5. Insertion of a new line and a new column with the distances between the new cluster and the remaining objects, in the matrix of the distances. The symmetric matrix has now the dimension $(n-1) \times (n-1)$

6. Step 1 to 5 are repeated until all objects are in one cluster.

The above method can be plotted in a dendrogram, a general example is shown in Fig. 6. In this example three clusters occur if a distance (d_{max}) of 1 is chosen. In this work, this fixed distance approximately corresponds to a difference of around $1K$ averaged over all the considered pressure layers.

The amount of data with unknown sonde_type is, for each year between 4% and 10%. These would strongly influence the statistic of the cluster in CL_STYPE. For this reason, CL_STYPE contains a group for the radiosonde with unknown sonde_type which is always treated as single cluster. This correction is not needed in the CL_STATID approach.
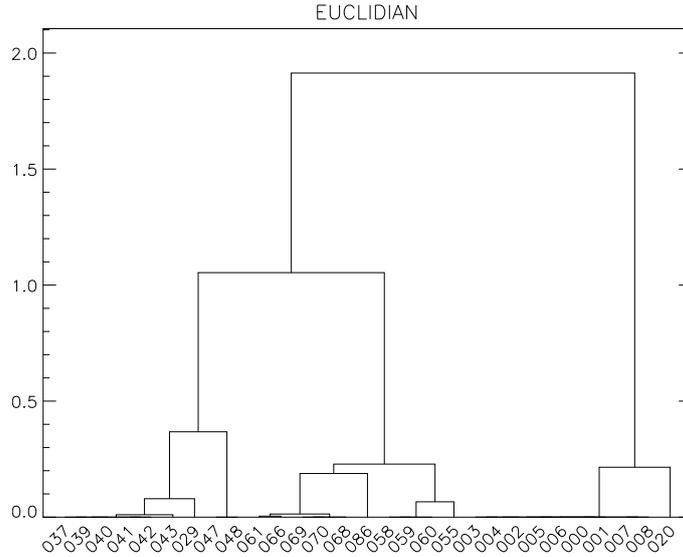
Figure 6: Example of dendrogram for a general objects pool.

4 Results

We applied CL_STYPE inferred from the S. Schroeders metadata file. In the case that the instrumentation is unknown, the sonde_type 0 is given. This is also the case where the instrumentation of the station is obviously false (e.g. radiosondes which are described as 'Wind only').

In CL_STYPE and in CL_STATID the clusterization was done for each year from the start year of ERA-Interim (1979) until 2013. The idea is a yearly update of the clusters. In this way, the changes in radiosonde types or other changes in the station ID are taken into account using a yearly statistic.

4.1 First guess departure profiles

For each cluster, the linear regression of FG with the bias model described above is computed using all data from the sonde_types or station ID in the cluster. For each year since 1979 we plotted the FG of the two types of clusters, the linear regression and we overplot also the mean FG of the members in the cluster in order to visualize the spread within the cluster.

In the case of CL_STYPE and STYPE, the error bars, at each pressure level in the plots, describe the variance of the stations within the same sonde_type. If $FG_i(j)$ is the mean FG for station j and the sonde_type i, the error of the sonde_type (i) over all stations J with the same sonde_type (i) is computed using:

$$\varepsilon_i = \sqrt{\sum_{j=1}^{J} \frac{\left(FG_i(j) - \overline{FG_i}\right)^2}{J}} \qquad (28)$$

ε_i is an estimation of the variability of the station biases within the same sonde_type at a certain pressure level. In this way, it is possible to recognize if the radiosonde is separate from different sonde_types

within a cluster. In order to estimate the approximation due to the clustering, the linear regression is done also without the use of clustering. An advantage of the latter approach is that the approximation will be better, but some sonde_types may have not enough data to calculate a robust bias estimate.

The error bars in CL_STATID and in STATID are the variance of the data for each level, when the number of the data in the level for the station ID is more than 20, otherwise the smallest and the largest values are taken.

The output of the offline regression is also evaluated for four different solar elevation classes, the same ranges as used in the metric definition 3.2.1. For the data where the solar elevation is in one of the classes, the FG is extracted in order to show the error in the bias correction for the chosen solar elevation class. Unfortunately, the sounding time, and thus the solar elevation, is not always available; for this reason the fit is calculated where the FG has the correspondent solar elevation.

Generally, the accuracy of the radiosonde temperature profiles depends on the sonde_type/station ID. Some sonde_types/station IDs show profiles which abruptly change level by level. Nevertheless, the algorithm yields a reasonable approximation for all sonde_types/station IDs. The FG for the cluster 32 (for CL_STYPE) in the year 1991 and its approximation is shown as example in Figs. 7 and 8 (for different ranges of solar elevation), the FG of the elements of the cluster are also overplotted (MEMBER FG). The clusterization efforts to group temperature profiles with similar characteristics (in the mean), shows differences within the ranges used in the cluster definition. The bias in this cluster depends strongly on the solar elevation. The linear regression approximates the mean of the FG departure well. In this case the difference in the FG between the data during the night (Fig. 8a) and for higher solar elevation (Fig. 8d) is visible especially in the higher levels. The FG during the night changed from -1.5K to 0K, while in the case of higher solar elevation the FG can reach values of less than -2K between 100hPa and 20hPa. In both cases the bias model described in Section 3.1 yields an approximation which can follow the FG profile.
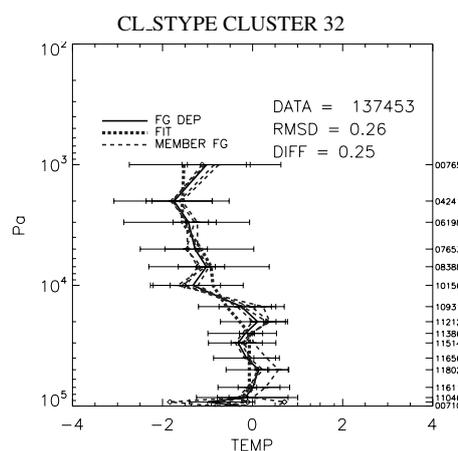
Figure 7: Year 1991, FG for CL_STYPE cluster number 32 (all members carry different types of radiosonde, all made by VIZ manufacturing). Solid line FG of the cluster. Dashed line FG of the cluster members. Dotted line the FG fit using the algorithm.

The FG of the cluster 19 in year 1998 (for CL_STYPE) and its approximation is shown as another example in Fig. 9. As visible in the cases presented here (Figs. 7, 9 and 13), common for many sonde_types, the variability within a cluster is lower than the station variability within a sonde_type. We computed the variability using eq. 28, For some sonde_types, the bias of the station can differ by more
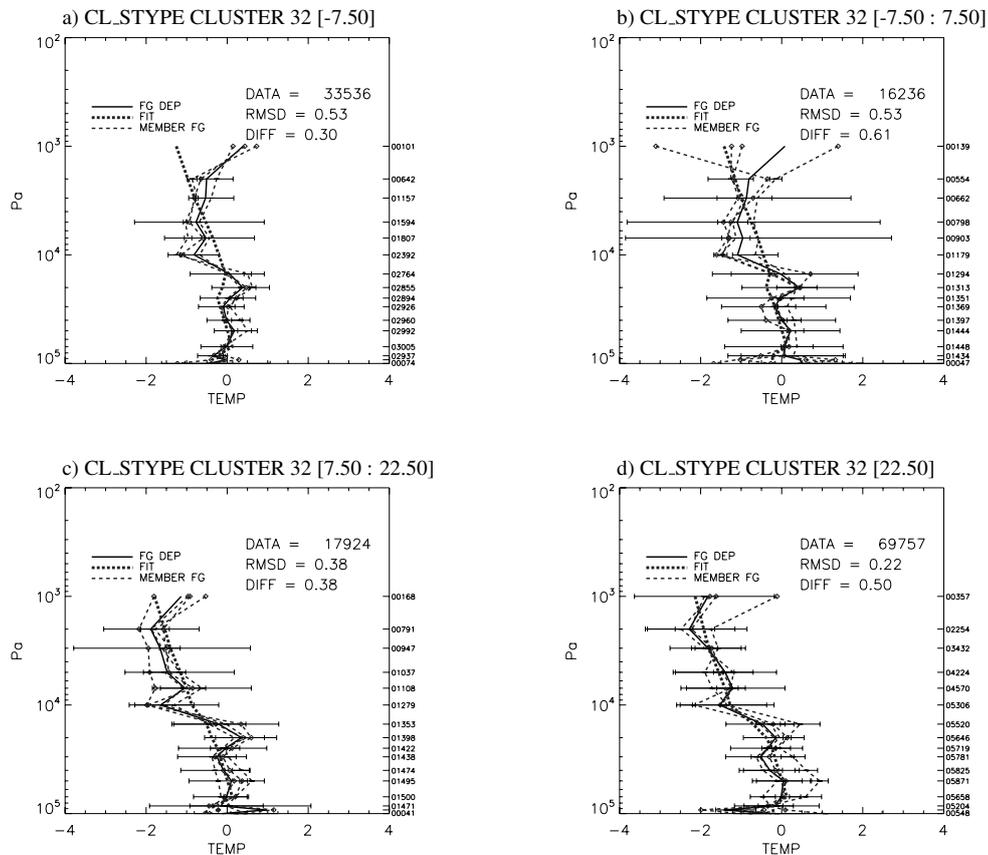
Figure 8:   Same as Fig. 7, FG for different angle ranges. **a:** $\theta \leq -7.50^o$; **b:** $\theta \in\,]-7.50^o, 7.50^o]$; **c:** $\theta \in\,]7.50^o, 22.50^o]$; **d:** $\theta > 22.50^o$. Lines as in Fig. 7.

than 4 K. This bias can clearly not be solely described by the sonde_type.

The cluster 19 in the year 1998 consists of two Indian sonde_types:

- I32 India 1680 MHz Mark 3 radiosonde 1992 WMO, number 431

- I31 India 401 MHz Mark 3 radiosonde 1992 WMO, number 430

The approximations using two different clusters from CL_STATID are shown in Figs. 10a and 10b. The profiles belong to station IDs which have the same sonde_type (WMO code I31, number 430), thus are part of the statistics in Fig. 9a. It is important to note that for some levels the FG of the clusters Fig. 10a shows differences larger than in Fig. 10b: at 70hPa the FG of cluster 94 is less than 2 K, while the cluster 182 is more than 3.5 K.

The behaviour of FG in different ranges of solar elevation is shown, for CL_STYPE in Fig. 11 and in Fig. 12 for CL_STATID. In this case the temperature profiles are weakly dependent on the solar elevation and approximations are very similar. The linear regression does not fit the FG well at higher levels. This feature (common also in other groups) is due to the sparse data in the 10hPa level and to strong differences between the FG at 20hPa and at 10hPa.

In order to check the sensitivity of the linear regression to the clusterization based on sonde_type, the
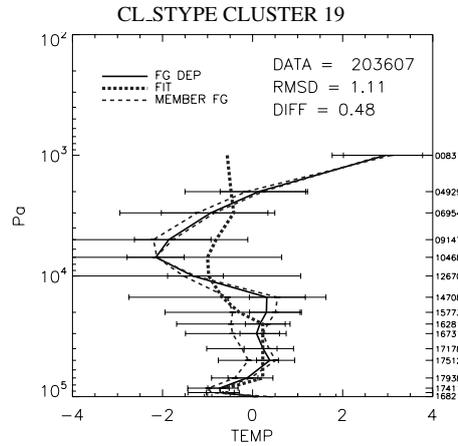
Figure 9: Year 1998, FG for CL_STYPE number 19, I31 (430) and I32 (431). First Guess departures (FG).
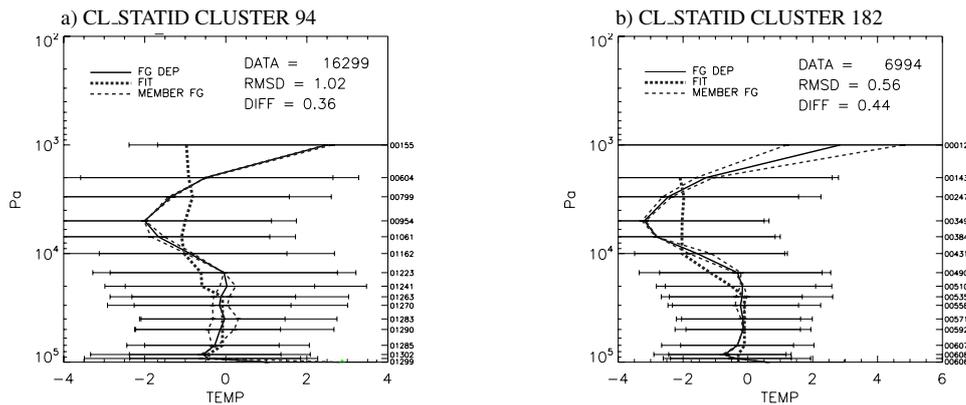
Figure 10: Year 1998, CL_STATID. Cluster number 94, stations: 42182, 42369. Number 182, stations: 42874, 42647 . All stations in the clusters with sonde_type I31(430). Lines as in Fig. 7

comparison between CL_STYPE and STYPE is done. One example of the BIAS is shown for the group 32 in the year 1991 (Fig. 13). This cluster consists of four VIZ sonde_types:

- ZF1c VIZ model 1192 NWS 1680 MHz, number 2876

- ZF1 VIZ model 1192 NWS 1680 MHz, number 2875

- VIZ B in USA, introduced in 1988, number 3135

- VIZ specifically [VIZ A was "pressure commutated"], number 3128
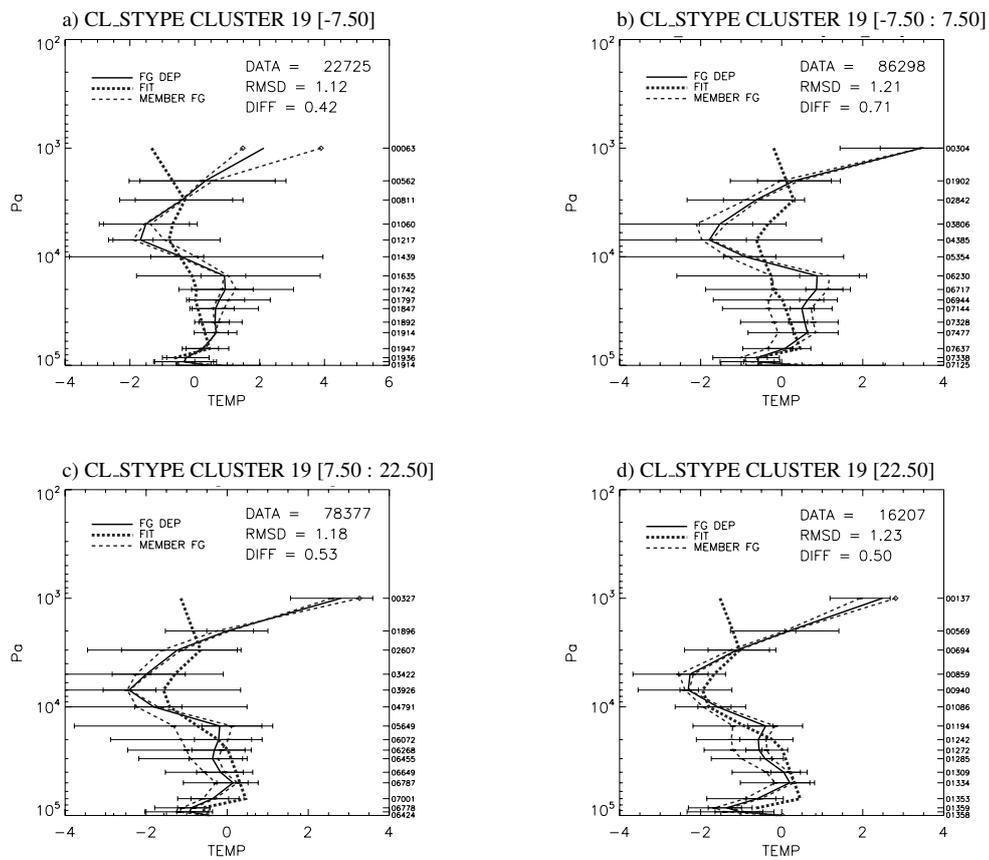
Figure 11:   Same as Fig. 9 FG for different angle ranges. **a:** $\theta \leq -7.50^o$; **b:** $\theta \in]-7.50^o, 7.50^o]$; **c:** $\theta \in]7.50^o, 22.50^o]$; **d:** $\theta > 22.50^o$. Lines as in Fig. 7.
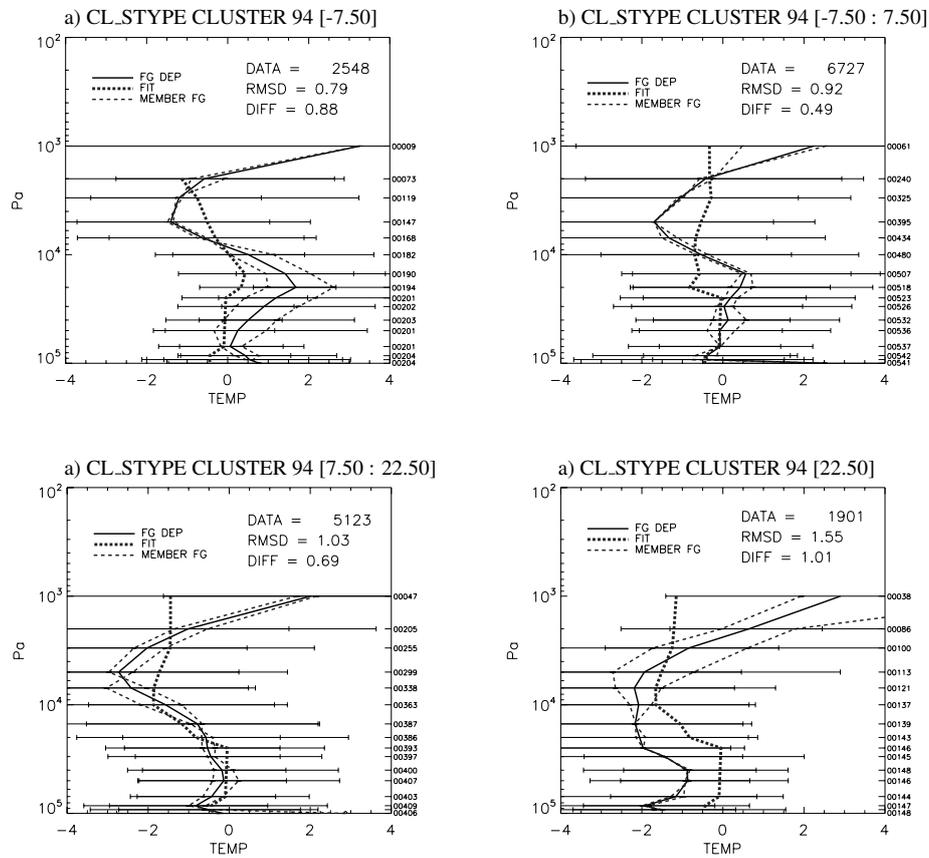
Figure 12:   Same as Fig. 10 for cluster 110 , FG for different angle ranges. **a**: $\theta \leq -7.50^o$; **b**: $\theta \in]-7.50^o, 7.50^o]$; **c**: $\theta \in]7.50^o, 22.50^o]$; **d**: $\theta > 22.50^o$. Lines as in Fig. 7.
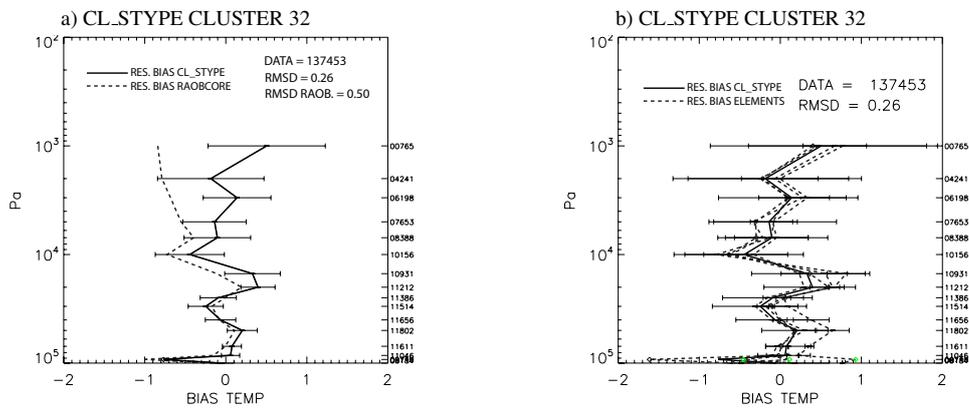
Figure 13: Year 1991, cluster based on sonde_type number 32, sonde_types in the cluster: ZF1c (2876), ZF1 (2875), CE (3135) and CI (3138). Comparison for residual BIAS **a**: RAOBCORE - CL_STYPE **b**: elements of CL_STYPE.
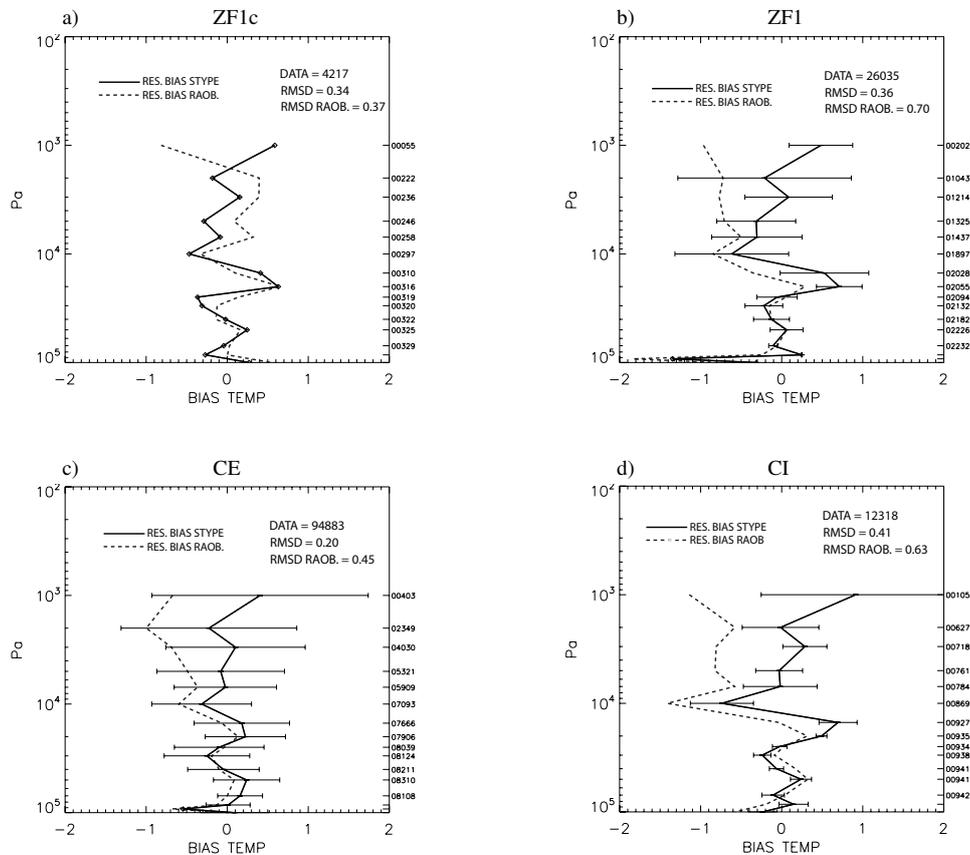
Figure 14: Year 1991, Elements in the cluster 32 of CL_STYPE. Residual BIAS for STYPE **a**: ZF1c (2876), **b**: ZF1 (2875), **c**: CE (3128), **d**: CI (3135).

As visible in the Figs. 13 and 14, the bias is well corrected if the algorithm is applied to the cluster as well as to the single sonde_type. A strong variability is present within the same sonde_type, which means that using this method for different stations with the same instrumentation (at least according to the metadata) the same bias correction is applied, but they have different bias behaviour.

The reason of this source of error can be originated from erroneous metadata, or from an incorrect assumption that stations with the same instrumentation have similar FG biases. The last assumption may be inappropriate if the same sonde_type is used at different latitudes. We assume the model bias equal to 0. In reality its bias is much lower than the one of the radiosonde but it is not negligible; different model biases at different latitudes can lead to the above mentioned variability within the same sonde_type. One example is the French sonde_type Mesural MH73A, which was used in the eighties in the Antarctic and also in Polynesia. For this reason the clusterization using only station ID is desirable.

4.2 Temperature trend spatial homogeneity

The existence of temperature biases in the radiosonde data can lead to inaccurate approximation in the temperature trend. For this reason an accurate study of the temperature trend is a necessary condition to extract information about the existences of biases in the radiosonde profiles.

In this work, it is assumed that nearby stations have similar trends, i.e. the trend would be spatially homogeneous. The temperature trend is computed for different levels for the years 1981-2010, when at least 20 of 30 years of data have been available. For a comparison with the data adjusted with RAOB-CORE and RICH, the format of the data has to be converted; which means every sounding given the local time has to be related to sounding at 00 UTC and at 12 UTC. Soundings between 9 UTC and 15 UTC are associated to 12 UTC, while soundings between 21 and 3 UTC are associated to 00 UTC; the mean of the two sounding is used as daily mean for the trend. As mentioned in Section 3.1, one problem in the available radiosonde data (especially in the early data) is that the hour is not always available. For this reason we can not use these data in our multi linear regression, which takes into account the solar elevation and thus time. For the study of the trend homogeneity, an objective method is needed. The metric (C) used is the one described in Haimberger [2007], which takes into account the different trends of the stations and their distances using an inverse exponential approach. The lower the value C is, the higher is the spatial homogeneity of the trends.

The trends from fields of ERA-Interim background forecasts are quite smooth and their value of the metric C can be used as benchmark. In the Figs. 15, 16, 17, 18 and 19 the temperature trends for raw data, CL_STYPE, CL_STATID, RAOBCORE, RICH and for the background (defined as "Model") are shown, at 00 UTC and at 12 UTC. As example the pressure level at 100hPa is taken. As expected in both cases the background field is highly homogeneous (value of C around 100 or less), the raw data show strong negative trends in the Chinese region and some negative trends in USA and Europe. These trends are partially corrected using the offline methods CL_STYPE and STYPE (Figs. 15b, 16b, 17a, 17b). Locally, CL_STYPE at 00 UTC shows larger trends in South-East Asia and Oceania, STYPE at 12 UTC shows larger trends in Alaska. The fields for RAOBCORE (Figs. 15c and 16c) and for RICH (Fig.s 15d and 16d) are very similar and show in both cases a lower C (higher homogeneity) than the multiple linear regression with grouping based on sonde_types. One of the reasons of this discrepancy is the existence of large spatial inhomogeneities in South-East Asia and India, which are not properly corrected using these approaches. One reason is, as already reported in Section 4.1, the presence of different biases for different stations within the same sonde_type. Second only 4 predictors are used for pressure layers, which means the correction is not done level by level (as in RAOBCORE and RICH) but nearby levels are grouped together. Using the approaches based on station ID the temperature trends show a clear improvement in the homogeneity and in the metric C compared to the corrections based on sonde_types (Figs. 18 a, 18b, 19a and 19b). An improvement in the values of the temperature trends (taking the background as benchmark) is observable, the C metric is still bigger than the ones of RICH and RAOBCORE, probably due to the not well corrected stations in India.

Theoretically, the clusterization can introduce artificial biases. Therefore a comparison between CL_STYPE and STYPE as well as between CL_STATID and STATID is done. Generally, the bias is corrected using the statistic of a cluster in a single year; in a different year the clusters change and also the bias may change. In order to analyse the sensitivity of the trend homogeneity to the clusterization, the maps of the temperature trends using STYPE are shown in Figs. 17a and b and for STATID in Figs. 19a and b. The differences between the results of CL_STYPE and STYPE are not evident, for the trends at 00 UTC (Figs. 15b and 17a), the mean (the C metric is comparable) and also the fields show similar values. At 12 UTC (Figs. 16b and 17b) the value of C is smaller for STYPE, but looking at the trend fields the higher inhomogeneities are still present and the differences are not well pronounced, a larger trend is present in Alaska / North West Canada for STYPE. The same test for CL_STATID and STATID show really homogeneous fields (especially at 12 UTC), furthermore the C metric and the fields have similar values.
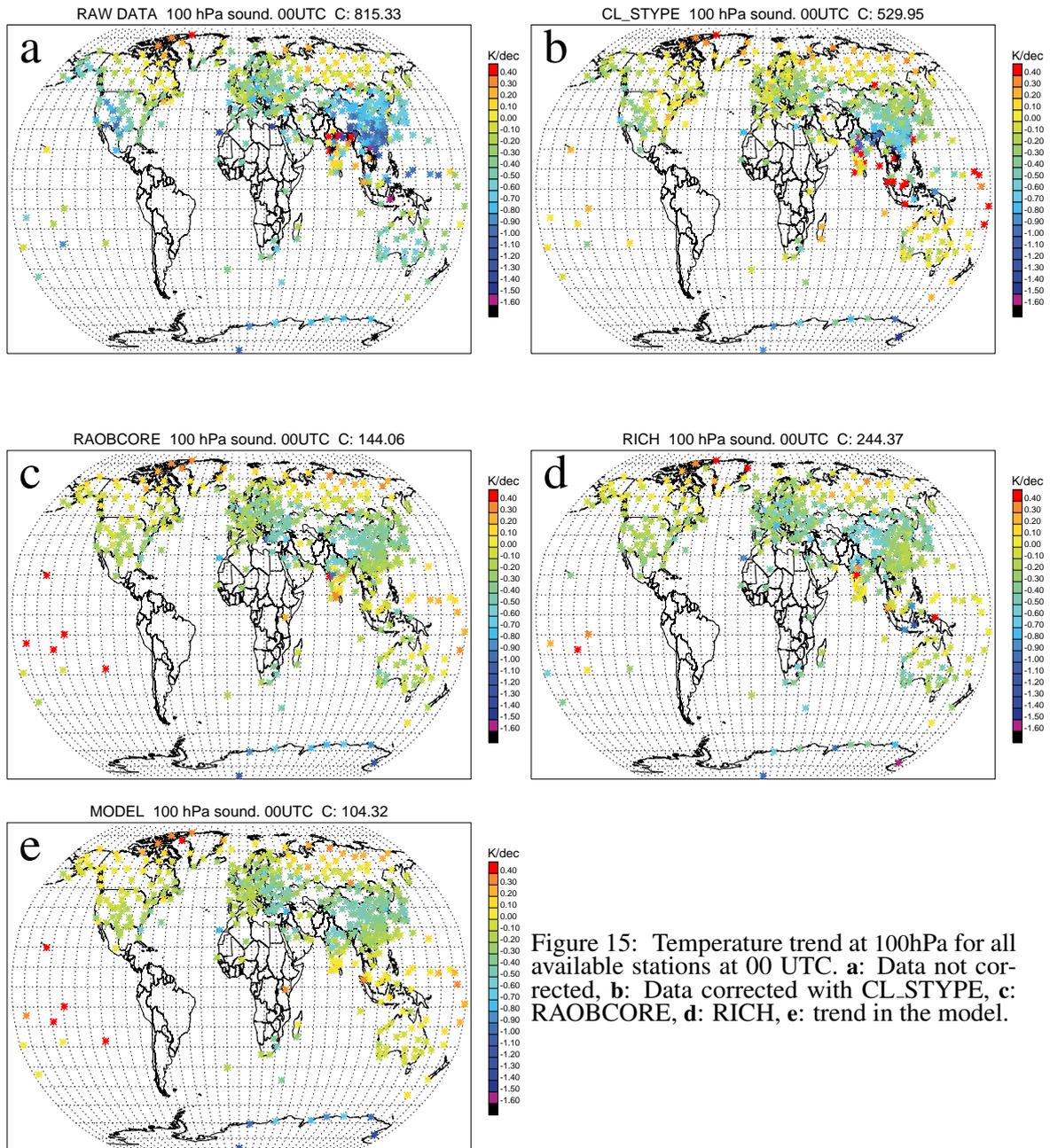
Figure 15: Temperature trend at 100hPa for all available stations at 00 UTC. **a**: Data not corrected, **b**: Data corrected with CL_STYPE, **c**: RAOBCORE, **d**: RICH, **e**: trend in the model.
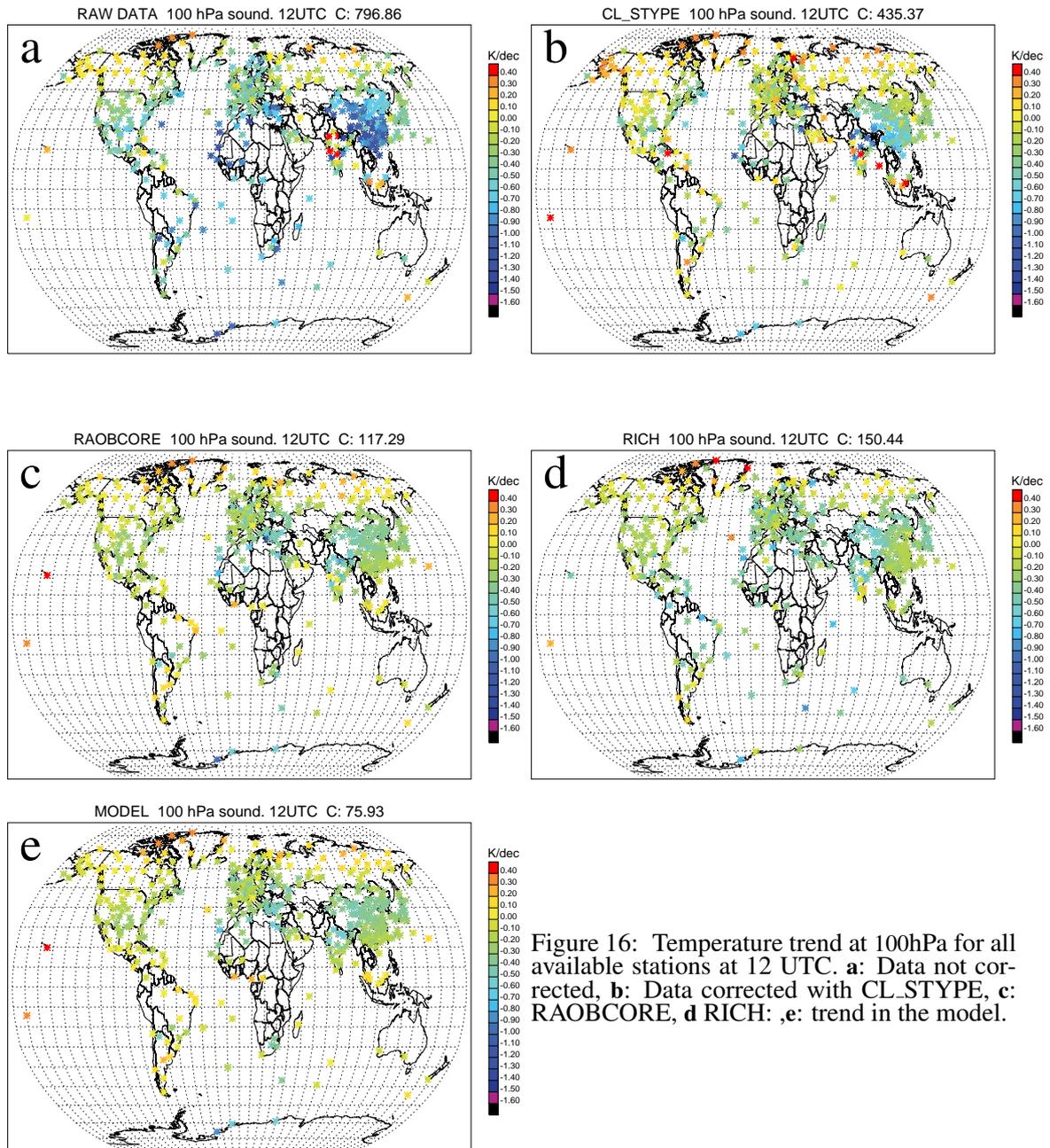
Figure 16: Temperature trend at 100hPa for all available stations at 12 UTC. **a**: Data not corrected, **b**: Data corrected with CL_STYPE, **c**: RAOBCORE, **d** RICH: ,**e**: trend in the model.
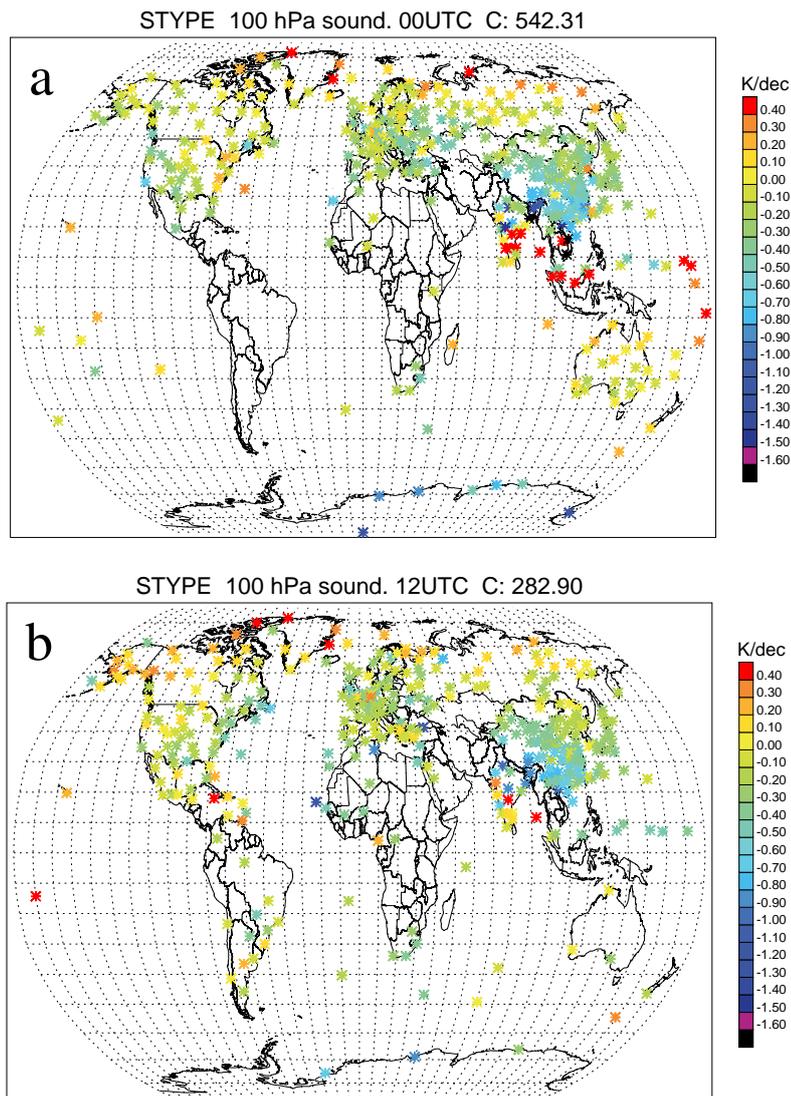
Figure 17: Temperature trend at 100hPa for data corrected with STYPE **a**: 00 UTC, **b**: 12 UTC.
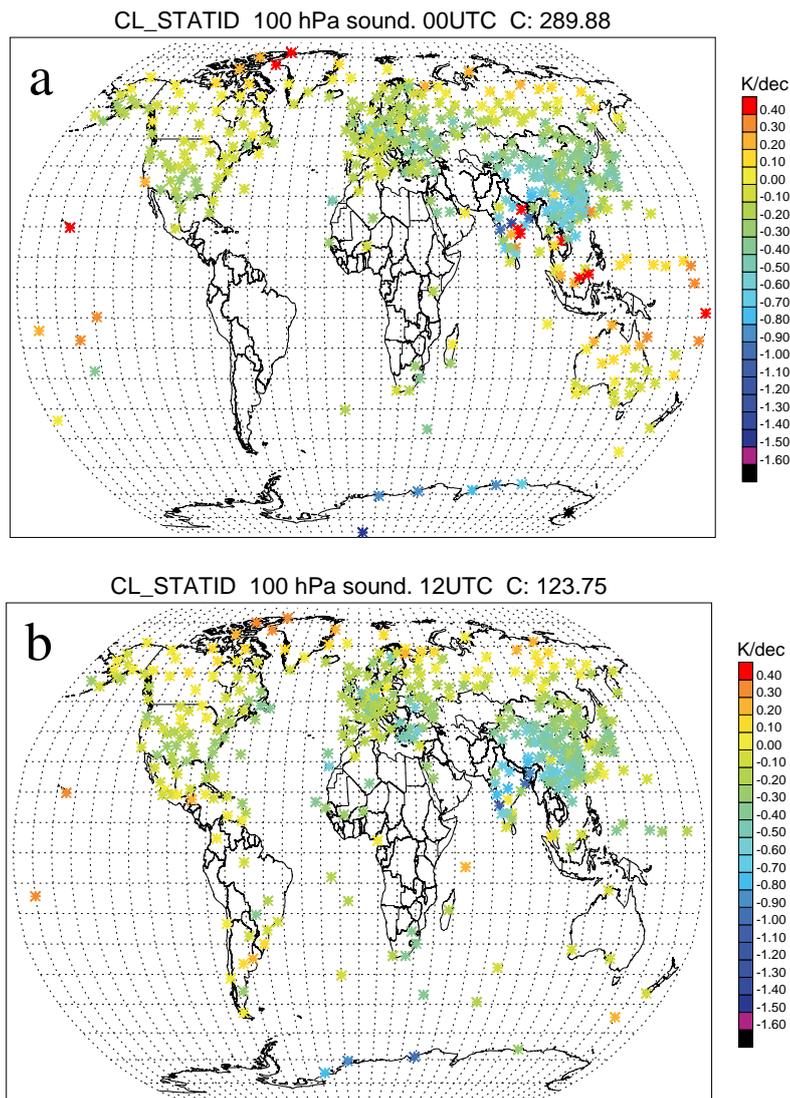
Figure 18: Temperature trend at 100hPa for data corrected with CL_STATID, **a**: 00 UTC, **b**: 12 UTC.

STATID 100 hPa sound. 00UTC C: 356.37

STATID 100 hPa sound. 12UTC C: 199.28
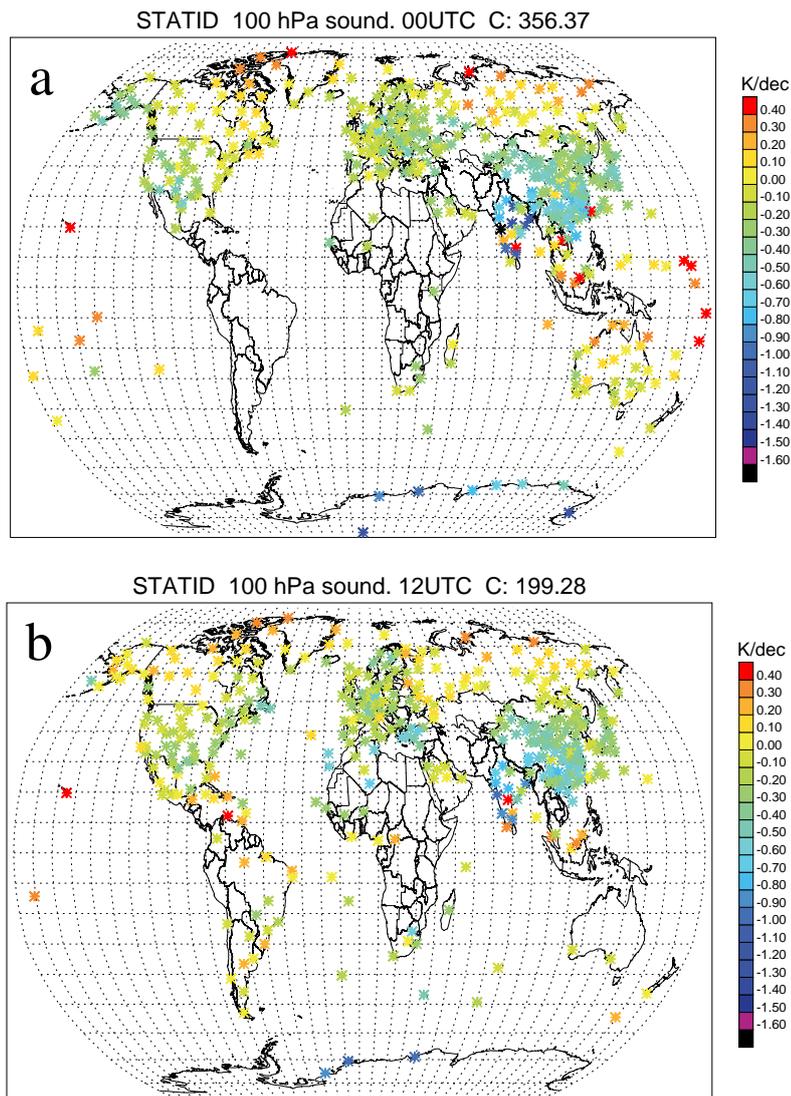
Figure 19: Temperature trend at 100hPa for data corrected with STATID, **a**: 00 UTC, **b**: 12 UTC.
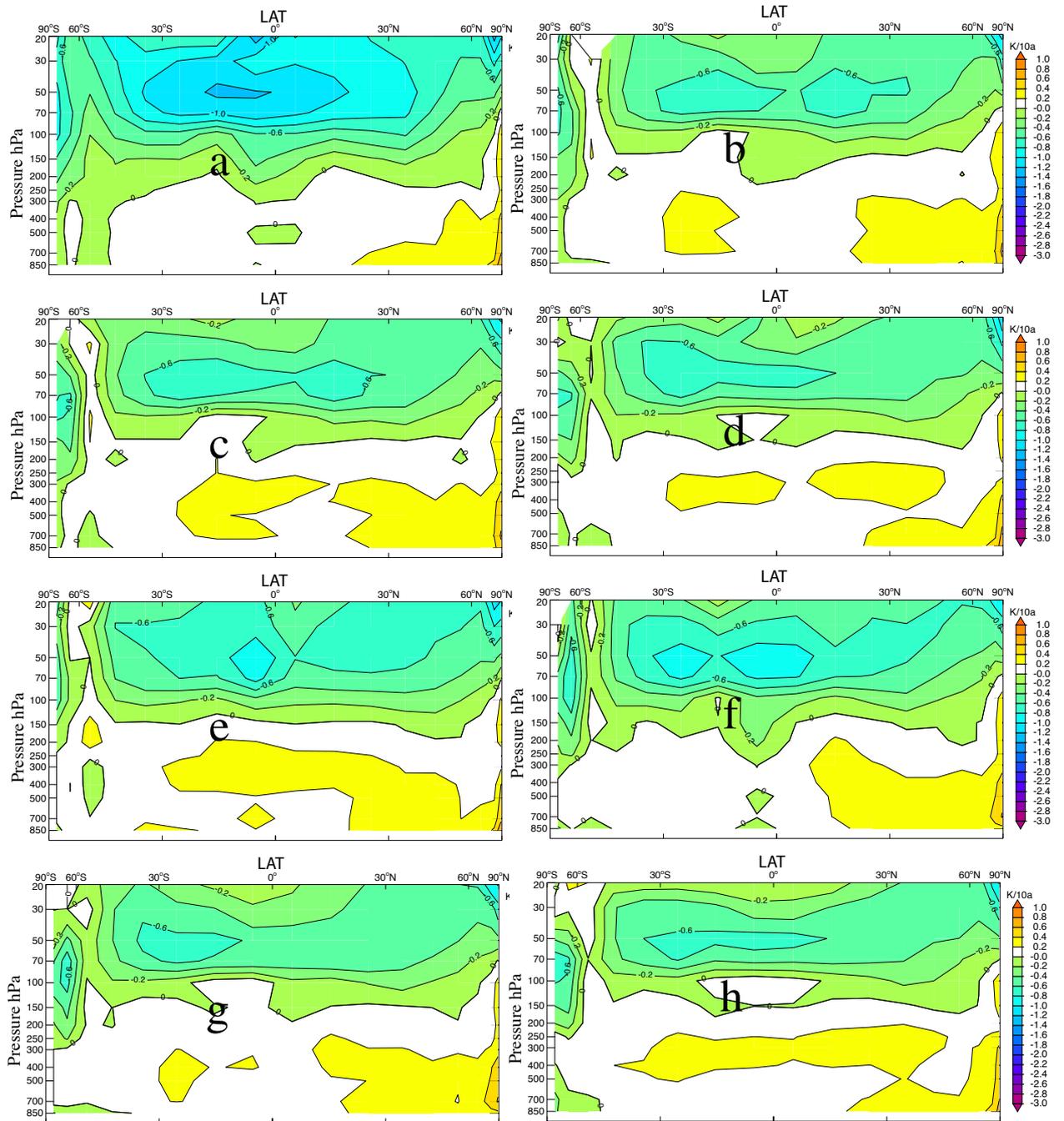
Figure 20:   Daily zonal mean temperature trend, on decadal scale (*K/decade*). **a**: data not corrected, **b**: CL_STYPE, **c**: STYPE **d**: RAOBCORE, **e**: RICH, **f**: CL_STATID, **g**: STATID, **h**: trend in the model.

The effect of the adjustments on the zonal mean of the trends for all pressure levels is shown in Fig. 20. We consider the stations where at least 240 monthly means in a total of 360 are present. As usual the benchmark for the homogeneity is the model trend (Fig. 20g). In the background a negative trend in the upper levels (pressure lower than 100hPa) is visible as well as a positive trend in the lower levels. The radiosondes have the feature to overestimate the negative trend in the upper levels (Fig. 20a), this property is described also in the literature (e.g., Sherwood et al. [2005], Randel and Wu [2006]). All types of bias correction algorithms succeed in the attenuation of the negative trend in the upper levels (Figs. 20b,c,d,e,f,g). The results using CL_STYPE (Fig. 20b) look very much like the ones from RAOBCORE, and both are very similar to the background trend. The corrected radiosonde data using RICH and CL_STATID (Fig. 20e,f) show a quite different field for the levels around 50hPa, the negative trend in this region is more pronounced than using the other corrections. This is not the case for the corrections using STATID, which is really similar to the background.

4.3   Temperature trend time homogeneity

In the offline approach, the correction from VarBC is approximated using the linear regression of the FG based on the predictors. One of the features which the bias correction should provide is the temperature homogeneity for each station in the considered time. The change in instrumentation or other changes, for a radiosonde station can lead to different biases and thus to time inhomogeneities. In this section we study the bias time series of two stations and the corrections computed using the CL_STYPE, STYPE, CL_STATID and STATID at the 100hPa level. A comparison with the results from RAOBCORE and RICH is also made.

The first test station has the ID number 70219 and is situated in Alaska (lat 60.79, lon -161.84). The monthly bias of the original data (Fig. 21a) in the early years shows a mean of around 1 K, followed by a jump from 1990 to 1996 where the bias is between 2K and 3K. The recent years have a very low bias. The data corrected with RAOBCORE (Fig. 21b) are homogeneous, only the variance is higher in the early years. RICH (Fig. 21e) reduces the bias but a limited residual value still exists $(0.12 \pm 0.37K)$, the stronger problems are in the data before 1996. CL_STYPE (Fig. 21c) and STYPE (Fig. 21d) reduce the bias but some months in the period between 1990 and 1996 have a strong residual bias, leading to high variability. CL_STATID (Fig. 21f) shows a reduction of the bias, however some months between 1990 and 1995 have an high variability in the residual bias; STATID shows better results for the residual bias, however with an relatively high variance for the months between 1990 and 1996. All methods have very similar corrections in the early and in the most recent periods (Fig. 22); CL_STYPE, STYPE, CL_STATID and STATID show a seasonal effect in the corrections. Between 1990 and 1996 CL_STYPE (Fig. 22a) and STYPE (Fig. 22b) have corrections similar to RAOBCORE and RICH but with a very strong variability. CL_STATID (Fig. 22c) is more similar to RICH and RAOBCORE, differences still exist in 1992 and between 1993 and 1996, this is not the case for STATID (Fig. 22d).

70219 ORIGINAL 100hPa 00UTC
bias=0.83 ± 0.91

70219 RAOBCORE 100hPa 00UTC
bias=0.00 ± 0.34

70219 CL_STYPE 100hPa 00UTC
bias=0.20 ± 0.40

70219 STYPE 100hPa 00UTC
bias=0.16 ± 0.41

70219 RICH 100hPa 00UTC
bias=0.12 ± 0.37

70219 CL_STATID 100hPa 00UTC
bias=0.30 ± 0.50

70219 STATID 100hPa 00UTC
bias=0.06 ± 0.34
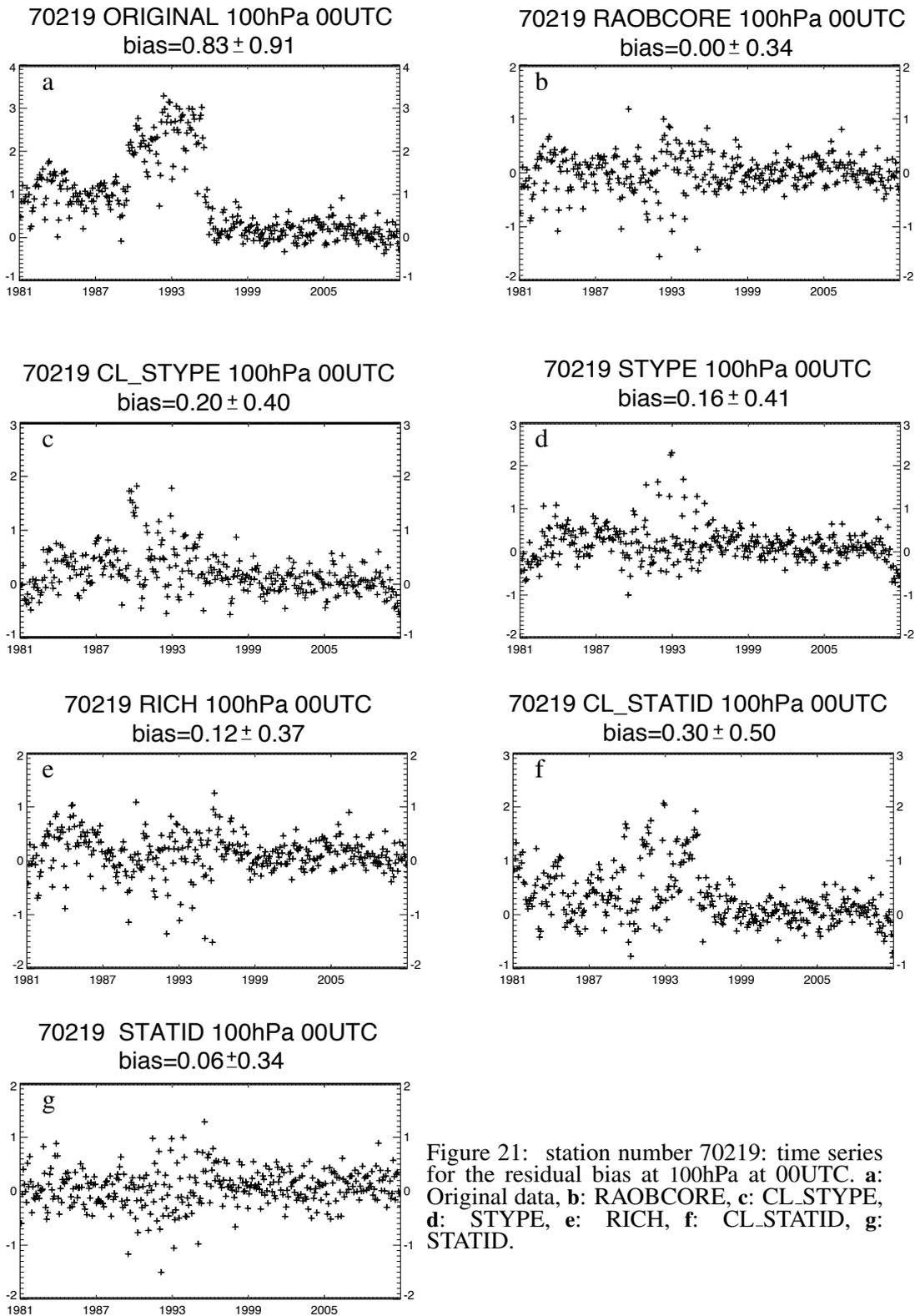
Figure 21: station number 70219: time series for the residual bias at 100hPa at 00UTC. **a**: Original data, **b**: RAOBCORE, **c**: CL_STYPE, **d**: STYPE, **e**: RICH, **f**: CL_STATID, **g**: STATID.
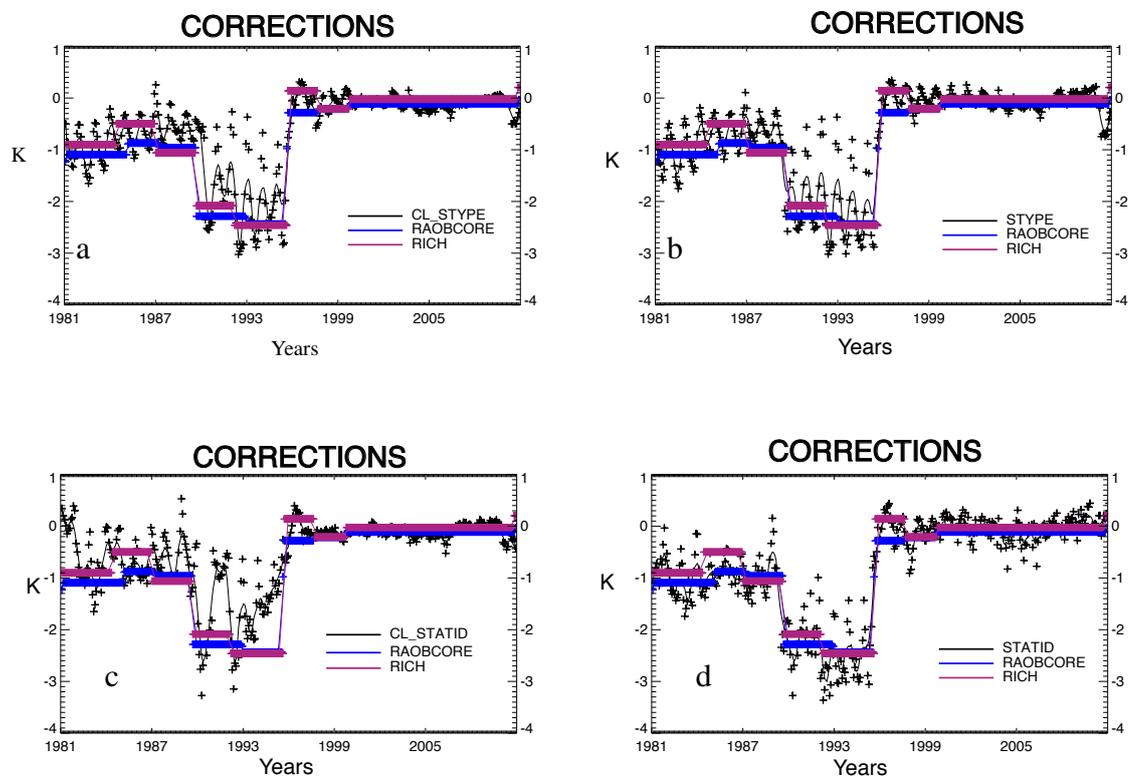
Figure 22: Station 70219, time series for the corrections at 100hPa; plus signs: monthly data, lines: running mean over 6 months. **a**: RAOBCORE, RICH and CL_STYPE, **b**: RAOBCORE, RICH and STYPE, **c**: RAOBCORE, RICH and CL_STATID, **d**: RAOBCORE, RICH and STATID.

The second test station has the ID number 91413 and is situated in the Pacific Ocean (lat 9.4895, lon 138.0716). As well visible from Fig. 23a, in the monthly bias of the original data at 00 UTC (10 local time), a jump is present around the year 1995. At that time the VIZ radiosondes (Schroeder type number 2961) were replaced by a Vaisala radiosonde (type number 2678). In the figure, the bias with its standard deviation is given. It is important to note the strong positive bias in the early years.

For both sonde_types (a VIZ type, 2961, and a Vaisala type, 2678), the behaviour of the residual bias at 100hPa is examined more closely in Fig. 24 and Fig. 25. In Figs. 24a and 25a the fit is computed using CL_STYPE. In the plots the residual bias of only the data with the studied sonde_type (sonde_type "number") is also present. In Figs. 24b and 25b the fit is computed using STYPE. From the previous analysis the jump in the station is present around the year 1995, and has a value of more than 1K (Figs. 23c and d). For both sonde_types such difference can not be attributed to the clusterization because the bias for the sonde_types is very similar (less than 0.4K).

The correction applied in ERA-INTERIM (Fig. 26), based on RAOBCORE, leads to a well homogenized time series (Fig. 23b). It reduces the bias but a residual value still exists (0.80 ± 0.54K). A similar behaviour is visible in the data corrected using RICH (Fig. 23e), in this case the jump around 1995 is still visible but strongly attenuated. The corrections applying CL_STYPE (Figs. 23 c and d) can correct the bias in the right direction but not completely. A large residual bias is still present in the corrected data; the results for CL_STYPE and STYPE are similar, the residual bias is 1.33 ± 0.81K using the clusterization and 1.30 ± 0.79K without. At the same time the correction applying CL_STATID (Fig. 23f), leads to a well homogenized time series with a low residual bias (0.52 ± 0.61K), very similar results are visible for STATID Fig. (23g).

Looking at the corrections, those from CL_STYPE and STYPE are smaller than the others (Fig. 26a, b). The values of the corrections from RAOBCORE are very similar to the ones using CL_STATID (Fig. 26c) and STATID (26d). For STATID some months can be considered as outliers, the discrepancy is due to the lack of data when using the STATID method. RICH shows similar corrections as RAOBCORE at the beginning and at the end of the period; in the middle of the time series RAOBCORE adjusts the BIAS in a smoother way, whereas RICH shows a strong jump in the adjustments.

Since the jump at station 91413 occurs in 1995, we analyse the temperature bias profiles in the years 1994 (Figs. 27 and 28) and 1996 (Fig. 29 and 30). The figures show the residual bias profiles using the CL_STYPE (Figs. 27a and 27c , 29a and 29c) and STYPE (Figs. 27b, 29b). The residual biases using CL_STATID and STATID are depicted in Figs. 28a and 28b for year 1994, and Figs. 30a and 30b for year 1996. Looking at the sonde_type approach, the profiles of the residual bias computed with and without clusterization are very similar; a variability of around 1K within the station of the same sonde_type is observable (Figs. 27b and 29b). In both years, the elements of the cluster in the residual bias applying CL_STATID are very similar one to each other, and the residual bias itself has always an absolute value lower than 1 K; STATID has similar results but one station shows different residual bias between 70 and 200hPa. Looking at the residual bias in the same cluster for RAOCORE (Figs. 29e and 27e) and RICH (Figs. 29f and 27f); in both years (1994 and 1996) the residual biases of RAOBCORE and RICH are comparable with the ones of CL_STATID: the root mean square of the cluster as well as the mean of the root mean square of the elements in the cluster are similar.

For each considered station at least 240 monthly approximations of FG are done (from a maximum of 360), which lead to the same number of evaluated biases. For each station the standard deviation of the bias can be computed, as well as the average of the absolute values of the residuals. These quantities measure the variability of the bias evaluation, the most important difference is that the standard deviation is more sensitive to the extreme values. In the presence of a jump, this variability is higher because of

91413 ORIGINAL 100hPa 00UTC
bias = 1.76±1.13

91413 RAOBCORE 100hPa 00UTC
bias = 0.80± 0.54

91413 CL_STYPE 100hPa 00UTC
bias=1.33±0.81

91413 STYPE 100hPa 00UTC
bias=1.30±0.79

91413 RICH 100hPa 00UTC
bias=0.50±0.61

91413 CL_STATID 100hPa 00UTC
bias=0.54 ± 0.64

91413 STATID 100hPa 00UTC
bias=0.51±0.63

Figure 23: Station number 91413: time series for the residual bias at 100hPa at 00UTC. **a**: Original data, **b**: RAOBCORE, **c**: CL_STYPE, **d**: STYPE, **e**: RICH, **f**: CL_STATID, **g**: STATID.

Figure 24: Time series for the residual bias at 100hPa, for sonde_type 2961 (a VIZ type), **a**: CL_STYPE and the single sonde_type, **b**: STYPE and RAOBCORE.
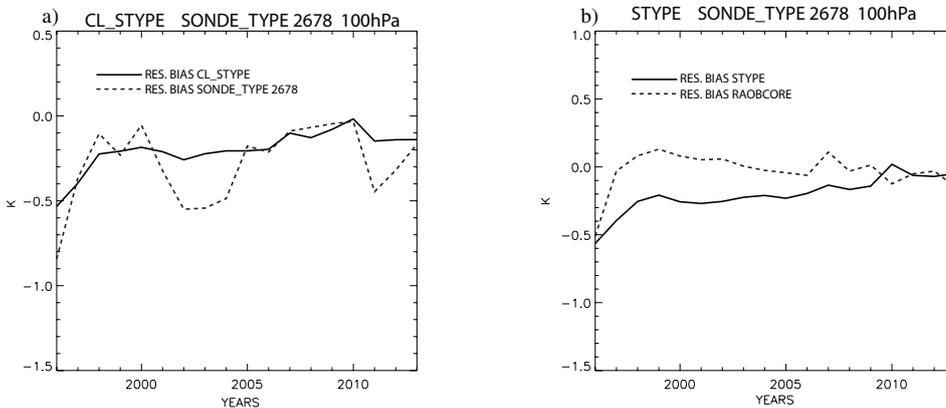
Figure 25: Time series for the bias at 100hPa, for sonde_type 2678 (a Vaisala type), **a**: CL_STYPE and the single sonde_type, **b**: STYPE and RAOBCORE.

the different biases before and after the jump. A general measure for the homogeneity can be the mean of the standard deviation or the mean of the average absolute values of the residues (FG-estimated bias).

The values for standard deviation and for the absolute value of the residues are given for different levels in Table 2 and Table 3. All values for the standard deviation are very similar, with some relative improvement in the average for the levels lower than 200hPa. RICH and RAOBCORE perform slightly better in the levels with pressure higher/equal than 200hPa, nevertheless the variability is very high. Looking at the absolute values of the residuals, the improvement compared to the raw data is more evident. One possible explanation is that the linear regression algorithms do not succeed in the correction (of radiosondes temperature profiles) where the difference in the biases is high. For this reason a thorough examination of the bias distribution is desirable (Section 4.4).
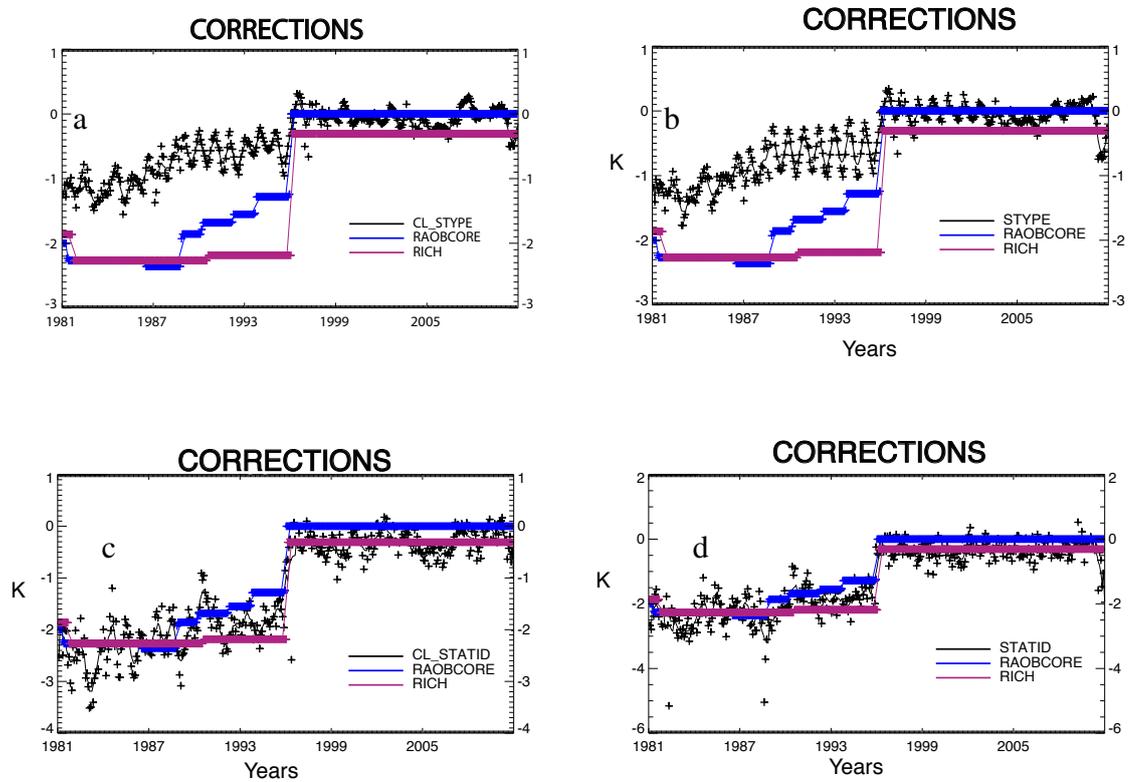
Figure 26:   Station 91413: time series for the corrections at 100hPa; plus signs: monthly data, lines: running mean over 6 months. **a**: RAOBCORE, RICH and CL_STYPE, **b**: RAOBCORE, RICH and STYPE, **c**: RAOBCORE, RICH and CL_STATID, **d**: RAOBCORE, RICH and STATID. .
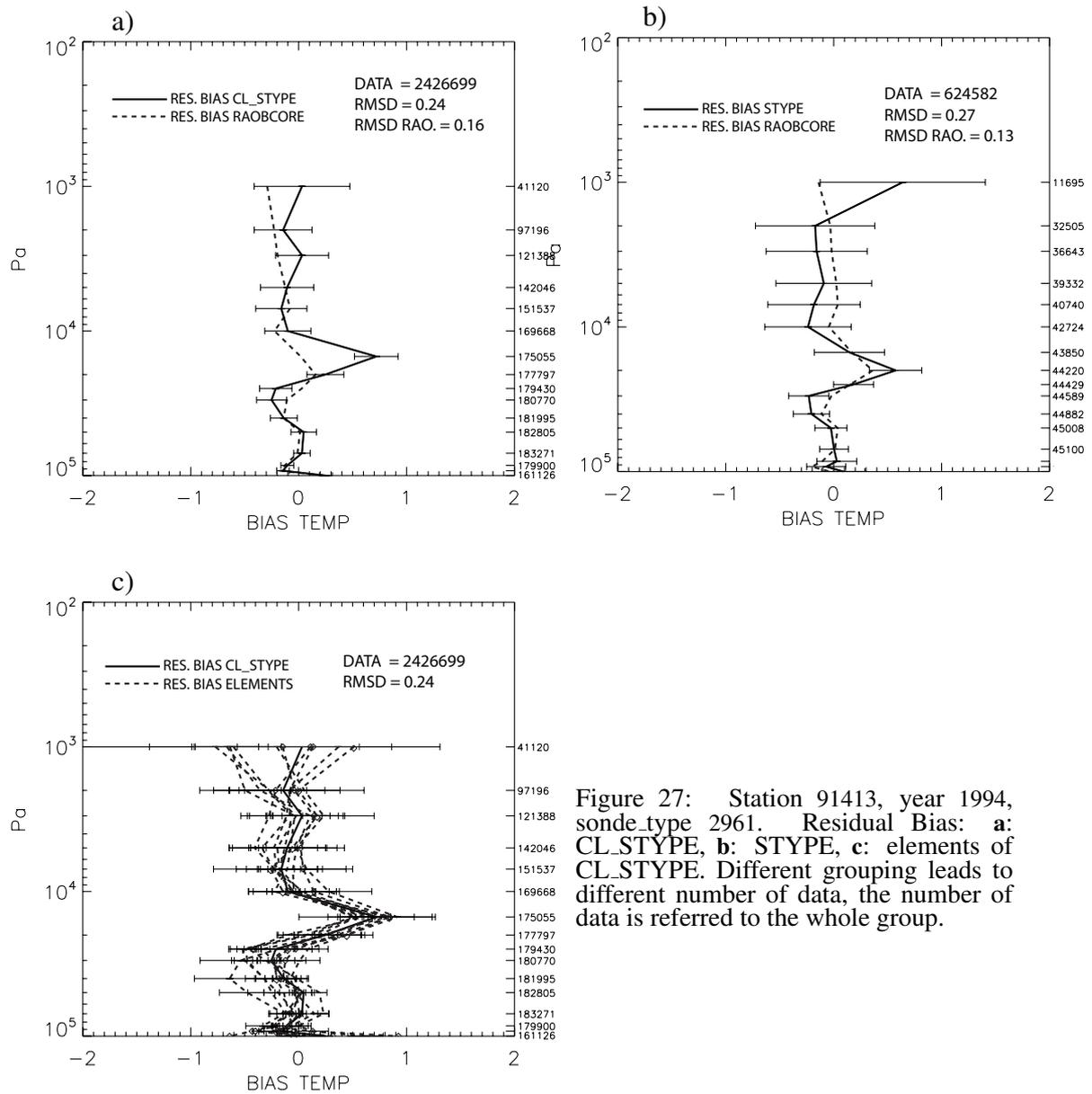
Figure 27: Station 91413, year 1994, sonde_type 2961. Residual Bias: **a**: CL_STYPE, **b**: STYPE, **c**: elements of CL_STYPE. Different grouping leads to different number of data, the number of data is referred to the whole group.
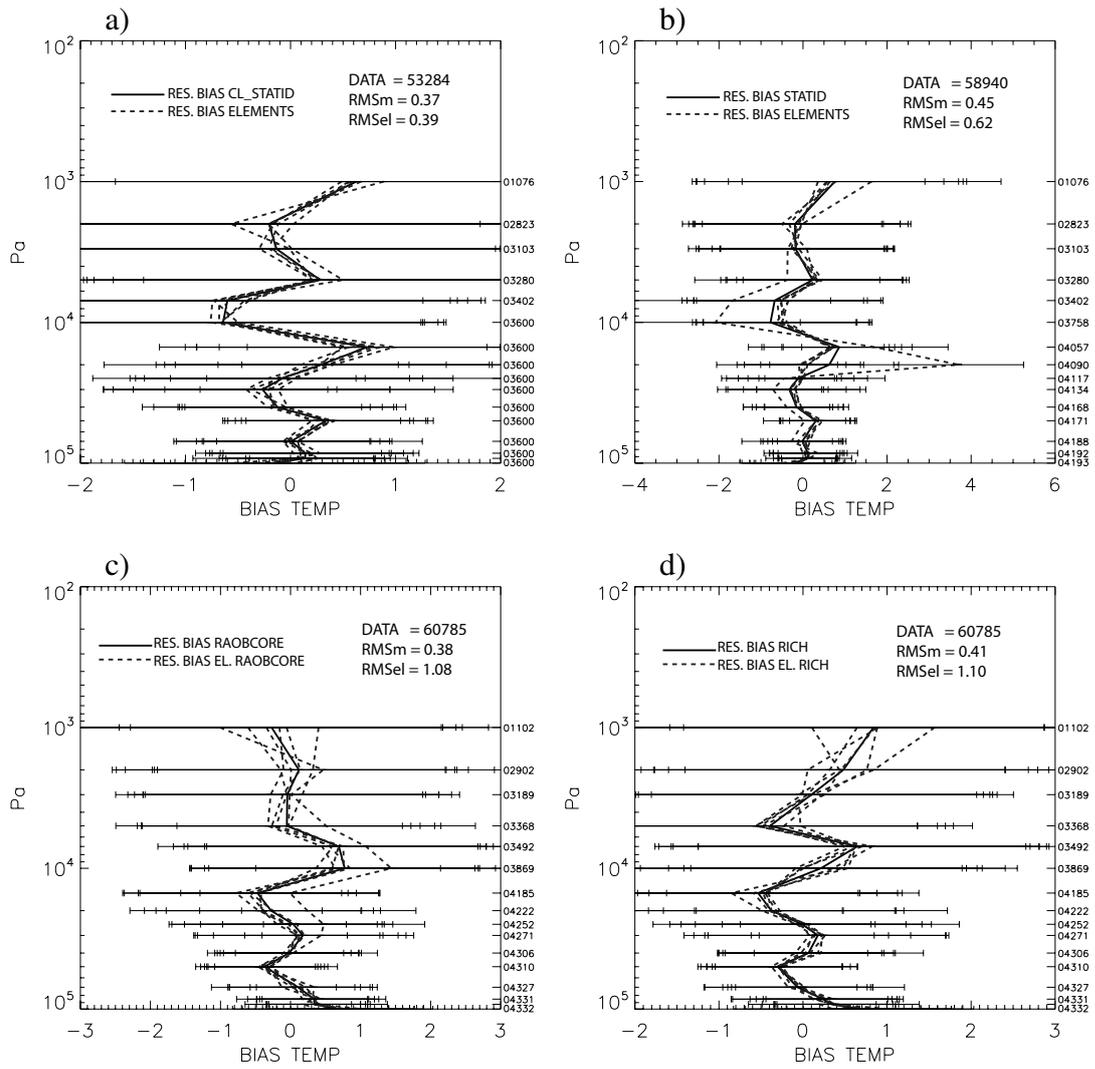
Figure 28: Station 91413, year 1994, sonde_type 2961. Residual Bias: **a**: elements of CL_STATID; **b**: STATID; **c**: RAOBCORE, **d**: RICH.
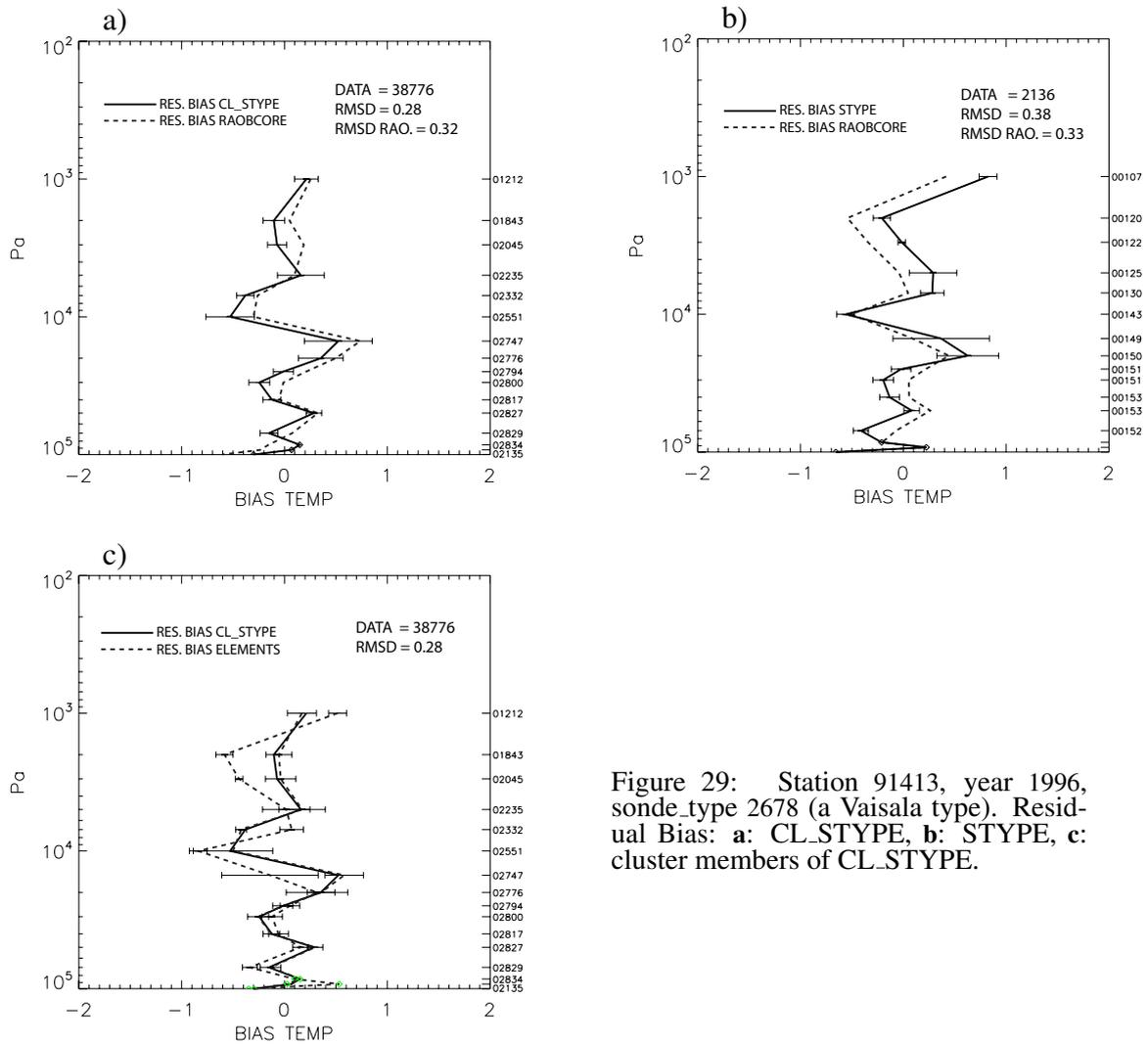
Figure 29: Station 91413, year 1996, sonde_type 2678 (a Vaisala type). Residual Bias: **a**: CL_STYPE, **b**: STYPE, **c**: cluster members of CL_STYPE.
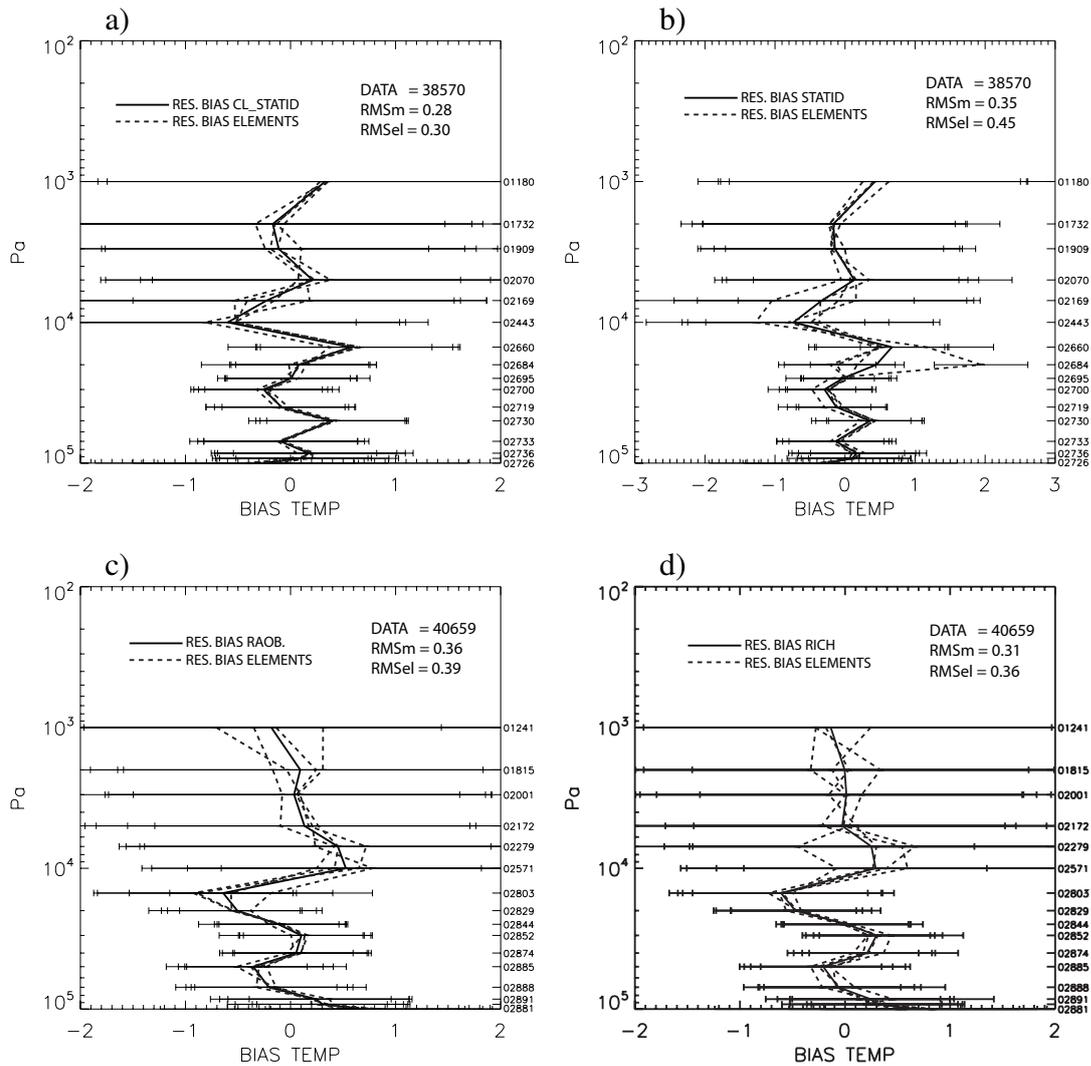
Figure 30: Station 91413, year 1996, sonde_type 2678. Residual Bias: **a** CL_STATID; **b**: STATID; **c**: RAOBCORE, **d**: RICH, and of the stations in the cluster.

	20hPa		50hPa		100hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	1.09 ± 0.59	0.99 ± 0.40	0.89 ± 0.73	0.83 ± 0.57	0.76 ± 0.59	0.73 ± 0.51
CL_STYPE	0.94 ± 0.46	0.90 ± 0.37	0.75 ± 0.62	0.76 ± 0.55	0.67 ± 0.56	0.64 ± 0.51
STYPE	0.95 ± 0.50	1.01 ± 0.50	0.82 ± 0.59	0.73 ± 0.49	0.82 ± 0.60	0.68 ± 0.54
CL_STATID	0.94 ± 0.50	0.90 ± 0.41	0.72 ± 0.50	0.71 ± 0.49	0.65 ± 0.51	0.60 ± 0.42
STATID	0.93 ± 0.48	0.92 ± 0.43	0.72 ± 0.52	0.72 ± 0.49	0.65 ± 0.50	0.61 ± 0.41
RAOBCORE	0.92 ± 0.50	0.81 ± 0.32	0.71 ± 0.58	0.65 ± 0.46	0.59 ± 0.48	0.56 ± 0.40
RICH	0.91 ± 0.38	0.85 ± 0.31	0.71 ± 0.55	0.67 ± 0.46	0.60 ± 0.46	0.55 ± 0.35
	200hPa		500hPa		700hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	0.69 ± 0.49	0.68 ± 0.46	0.44 ± 0.28	0.44 ± 0.28	0.43 ± 0.24	0.45 ± 0.25
CL_STYPE	0.61 ± 0.40	0.67 ± 0.42	0.42 ± 0.25	0.44 ± 0.27	0.44 ± 0.26	0.48 ± 0.30
STYPE	0.79 ± 0.48	0.60 ± 0.41	0.45 ± 0.33	0.44 ± 0.28	0.47 ± 0.33	0.47 ± 0.27
CL_STATID	0.61 ± 0.40	0.58 ± 0.35	0.44 ± 0.32	0.44 ± 0.32	0.49 ± 0.32	0.49 ± 0.31
STATID	0.64 ± 0.44	0.59 ± 0.33	0.50 ± 0.44	0.54 ± 0.45	0.56 ± 0.44	0.60 ± 0.44
RAOBCORE	0.53 ± 0.34	0.52 ± 0.34	0.36 ± 0.21	0.37 ± 0.22	0.39 ± 0.21	0.41 ± 0.23
RICH	0.57 ± 0.33	0.54 ± 0.32	0.40 ± 0.21	0.41 ± 0.21	0.42 ± 0.21	0.46 ± 0.22

Table 2: Mean of bias standard deviation with its standard deviation, at different pressure levels.

	20hPa		50hPa		100hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	1.01 ± 0.58	1.04 ± 0.50	0.99 ± 1.03	0.93 ± 0.88	0.93 ± 0.66	0.94 ± 0.70
CL_STYPE	0.76 ± 0.36	0.78 ± 0.34	0.66 ± 0.58	0.67 ± 0.49	0.58 ± 0.45	0.57 ± 0.42
STYPE	0.75 ± 0.38	0.78 ± 0.34	0.66 ± 0.56	0.65 ± 0.47	0.59 ± 0.44	0.56 ± 0.42
CL_STATID	0.67 ± 0.30	0.65 ± 0.24	0.56 ± 0.40	0.55 ± 0.38	0.52 ± 0.38	0.49 ± 0.31
STATID	0.67 ± 0.30	0.66 ± 0.24	0.57 ± 0.42	0.55 ± 0.38	0.53 ± 0.38	0.49 ± 0.31
RAOBCORE	0.72 ± 0.43	0.70 ± 0.34	0.59 ± 0.59	0.61 ± 0.81	0.52 ± 0.39	0.54 ± 0.55
RICH	0.82 ± 0.32	0.75 ± 0.29	0.63 ± 0.42	0.60 ± 0.36	0.60 ± 0.38	0.53 ± 0.31
	200hPa		500hPa		700hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	0.66 ± 0.44	0.67 ± 0.43	0.40 ± 0.23	0.37 ± 0.23	0.38 ± 0.26	0.38 ± 0.22
CL_STYPE	0.53 ± 0.35	0.53 ± 0.35	0.36 ± 0.20	0.39 ± 0.23	0.41 ± 0.29	0.42 ± 0.25
STYPE	0.54 ± 0.34	0.51 ± 0.34	0.37 ± 0.22	0.38 ± 0.23	0.41 ± 0.30	0.42 ± 0.24
CL_STATID	0.46 ± 0.29	0.45 ± 0.23	0.33 ± 0.18	0.35 ± 0.16	0.41 ± 0.26	0.41 ± 0.20
STATID	0.47 ± 0.29	0.45 ± 0.23	0.34 ± 0.19	0.35 ± 0.16	0.42 ± 0.26	0.42 ± 0.20
RAOBCORE	0.47 ± 0.27	0.45 ± 0.33	0.32 ± 0.19	0.32 ± 0.26	0.35 ± 0.22	0.35 ± 0.21
RICH	0.49 ± 0.25	0.48 ± 0.25	0.36 ± 0.19	0.38 ± 0.18	0.37 ± 0.17	0.39 ± 0.17

Table 3: Mean of absolute value of the residue with its standard deviation, at different pressure levels.

	20hPa		50hPa		100hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	1.42 ± 0.92	1.71 ± 1.02	2.17 ± 0.93	1.80 ± 0.91	1.39 ± 0.62	1.51 ± 0.75
CL_STYPE	0.38 ± 0.16	0.39 ± 0.15	0.52 ± 0.28	0.40 ± 0.16	0.45 ± 0.21	0.34 ± 0.18
STYPE	0.57 ± 0.19	0.77 ± 0.33	0.66 ± 0.18	0.59 ± 0.17	0.47 ± 0.21	0.44 ± 0.16
CL_STATID	0.39 ± 0.15	0.40 ± 0.14	0.53 ± 0.29	0.39 ± 0.15	0.44 ± 0.18	0.32 ± 0.15
STATID	0.28 ± 0.15	0.25 ± 0.09	0.42 ± 0.21	0.32 ± 0.15	0.42 ± 0.15	0.31 ± 0.14
RAOBCORE	0.40 ± 0.16	0.52 ± 0.26	0.36 ± 0.13	0.71 ± 0.34	0.28 ± 0.08	0.44 ± 0.17
RICH	0.70 ± 0.46	0.70 ± 0.47	0.46 ± 0.23	0.39 ± 0.20	0.41 ± 0.14	0.39 ± 0.19
	200hPa		500hPa		700hPa	
TYPE	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**	**00 UTC**	**12 UTC**
Raw data	0.63 ± 0.20	0.63 ± 0.32	0.21 ± 0.06	0.19 ± 0.08	0.17 ± 0.03	0.15 ± 0.06
CL_STYPE	0.25 ± 0.09	0.22 ± 0.07	0.13 ± 0.04	0.14 ± 0.05	0.24 ± 0.06	0.20 ± 0.08
STYPE	0.38 ± 0.14	0.32 ± 0.08	0.19 ± 0.07	0.20 ± 0.08	0.24 ± 0.05	0.19 ± 0.08
CL_STATID	0.25 ± 0.08	0.23 ± 0.07	0.13 ± 0.04	0.15 ± 0.05	0.23 ± 0.06	0.20 ± 0.07
STATID	0.20 ± 0.09	0.16 ± 0.06	0.14 ± 0.04	0.15 ± 0.04	0.23 ± 0.05	0.21 ± 0.07
RAOBCORE	0.18 ± 0.05	0.18 ± 0.07	0.10 ± 0.03	0.11 ± 0.04	0.12 ± 0.02	0.11 ± 0.03
RICH	0.21 ± 0.07	0.20 ± 0.09	0.13 ± 0.05	0.13 ± 0.06	0.11 ± 0.05	0.12 ± 0.07

Table 4: Mean of MS with its standard deviation, at different pressure levels.

4.4   Distribution of the biases

In the case of a perfect correction of the radiosonde data the residual bias should be normally distributed around zero with variance tending to zero. The bias distributions of the raw and the corrected radiosonde data can be estimated from their frequencies in given bins (total number of bins *N*):

$$MS = \frac{1}{N} \sum_{i=1}^{N} x_i^2 \tag{29}$$

where x_i is the value of the residual bias in each bin.

This metric can be computed for all years and all levels. We again take 100hPa as example, in Figs. 31 and 32 the distributions are plotted for the raw data (a) and for the data corrected using CL_STYPE (Figs.31b and 32 b), STYPE (Figs. 31c and 32 c), CL_STATID (Figs. 31f and 32f) and STATID (Figs. 31g and 32g). The data adjusted with RAOBCORE (Fig.s 31d and 32d) or RICH (Figs. 31e and 32e) are also shown. The raw data distribution has mean of 0.8K with considerable variance and a positive tail. The corrected data using CL_STYPE and STYPE are comparable (the MS and the distribution). The positive tail is reduced and the distribution has a mean of less than 0.2K. CL_STATID and STATID improve even more the results for the mean and for MS. The distribution of residual bias of RAOBCORE is narrower than the sonde_type approach but it has a positive mean of around 0.3K, which leads to the larger MS at 12UTC. In RICH the mean is closer to 0, compared with the other corrections, but at the same time the MS is higher, due to the broader distribution (higher variance).

The MS of the residual Bias for all methods computed over each years (and then averaged), for 6 different pressure levels, is shown in Table 4; the MS is computed each month, we considered only the stations where at least 240 of 360 months are present. Generally, in the troposphere (levels 500 and 700hPa), all methods are comparable, even the raw data has still a low MS. From the higher levels up to 200hPa, the better MS is given by CL_STATID and STATID.
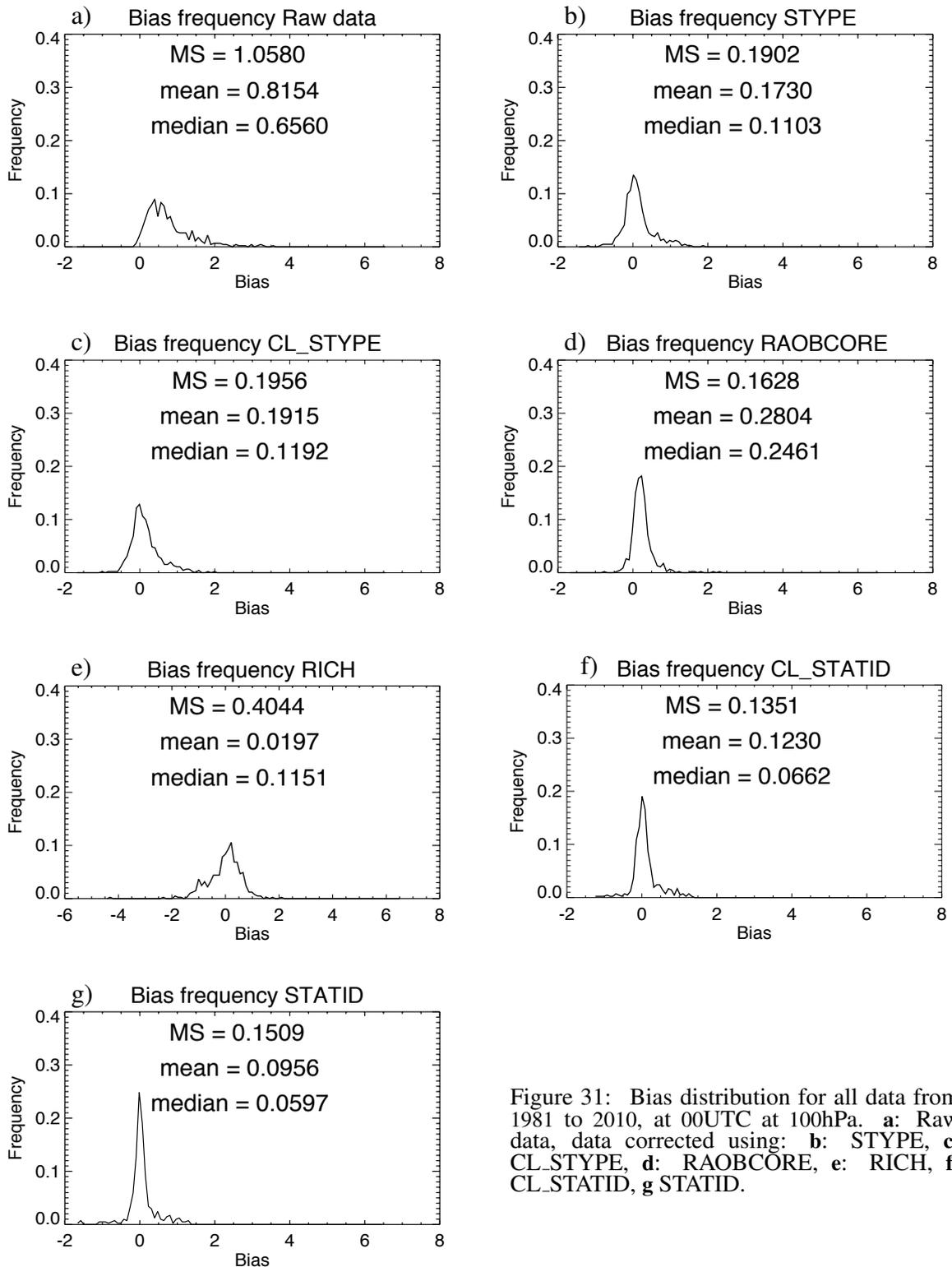
Figure 31: Bias distribution for all data from 1981 to 2010, at 00UTC at 100hPa. **a**: Raw data, data corrected using: **b**: STYPE, **c**: CL_STYPE, **d**: RAOBCORE, **e**: RICH, **f**: CL_STATID, **g** STATID.
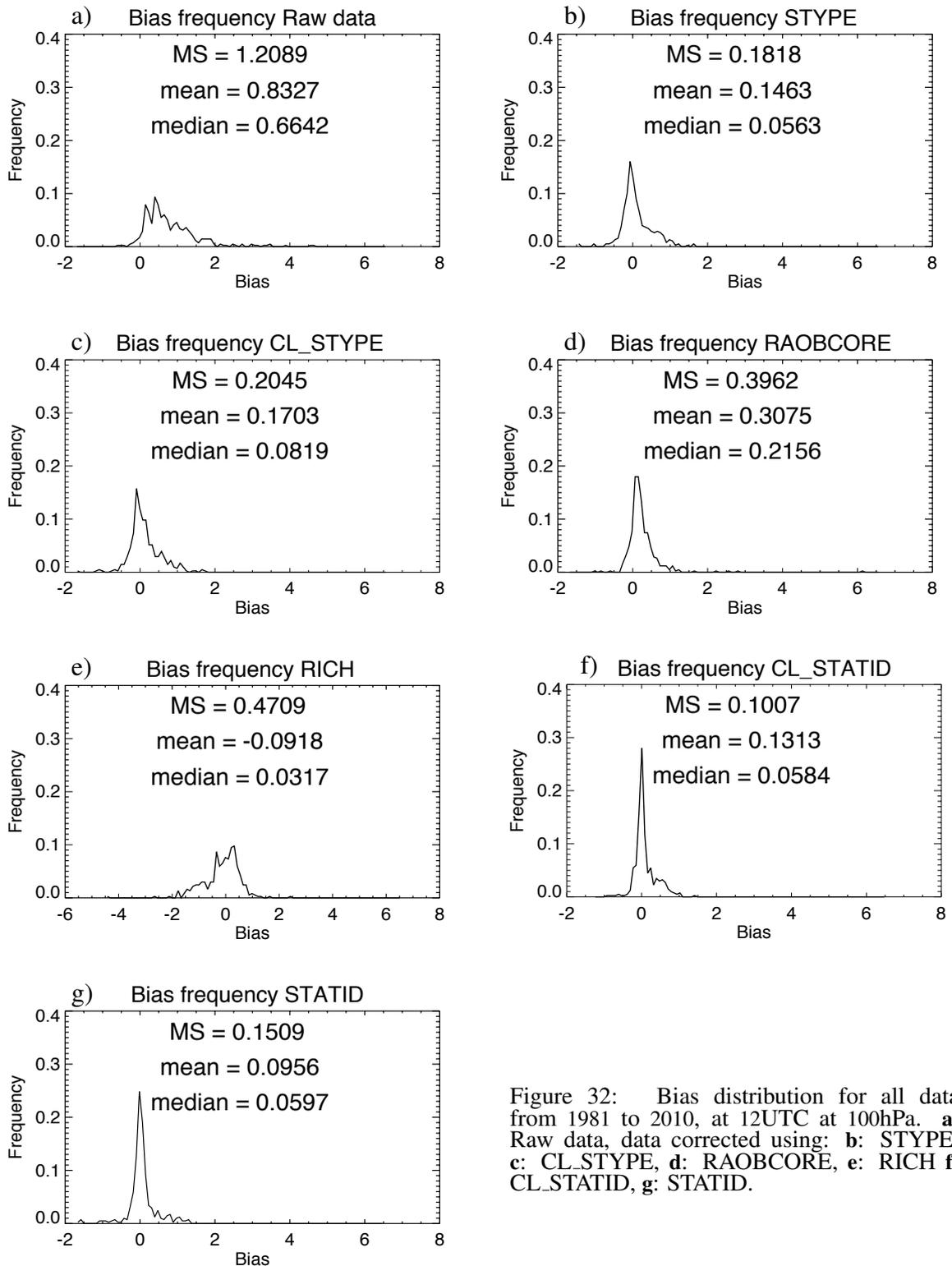
Figure 32:   Bias distribution for all data from 1981 to 2010, at 12UTC at 100hPa. **a**: Raw data, data corrected using: **b**: STYPE, **c**: CL_STYPE, **d**: RAOBCORE, **e**: RICH **f**: CL_STATID, **g**: STATID.
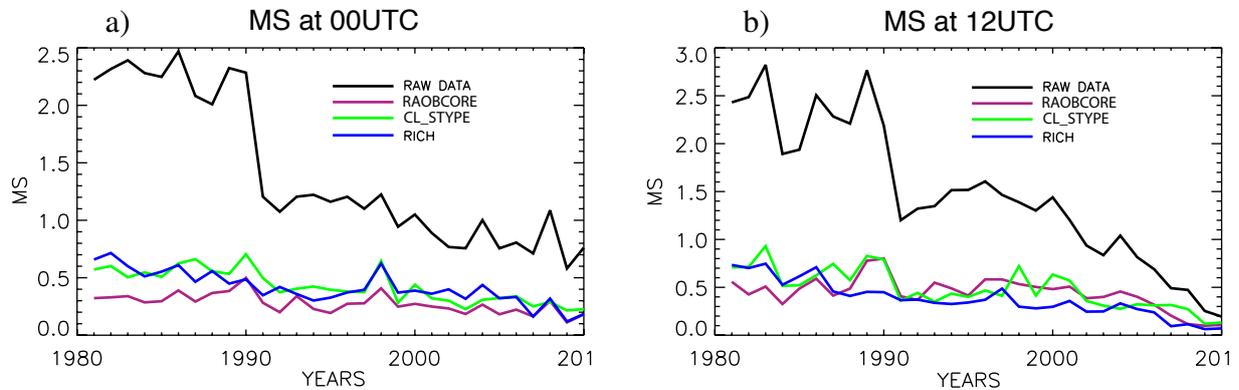
Figure 33: MS of the bias distribution for all data at 100hPa, CL_STYPE. **a**: 00UTC, **b**: 12UTC.

The distribution and the MS is computed for each year (the MS is computed each month, we considered only the stations where at least 6 months of data in the year are present), in order to check the time evolution of the MS. We assumed that the early data are more biased than recent data. In Figs. 33 and 34 the time evolution of the MS (at 100hPa) for the raw data, CL_STYPE and STYPE are plotted, at 00UTC and 12UTC. The differences in the MS for the raw data (and also for the corrected data from RAOBCORE) between the two figures are due to the different data sample, we take into account only the soundings adjusted by the linear regression and a larger sample is present using the clusterization. The improvement of the raw data quality with time is well visible. The mean of the MS decreases from 2.5K to around 0.5K, a strong jump is present in the early nineties. The MS at 00UTC for the data corrected using RAOBCORE is for certain periods of time better than the one using CL_STYPE, the differences are lower in the later period. The MS of CL_STYPE and STYPE are in all cases comparable.

A peak in the MS of STYPE adjusted data is visible at the year 1994, this is due to a problem in the Chinese radiosondes. In Fig. 34b the MS without the Chinese data is plotted, and no peak is visible. The MS of the corrected data based on CL_STATID is shown in Fig. 35, for STATID in Fig. 36. At 00UTC, the values for CL_STATID are comparable with RICH and RAOBCORE, some differences are visible between 1990 and 1998. The results for STATID are of the same magnitude, some spikes are present, probably due to the lack of data. At 12UTC the results for CL_STATID are comparable and sometimes better than RICH and RAOBCORE, in STATID still 2 spikes are present.
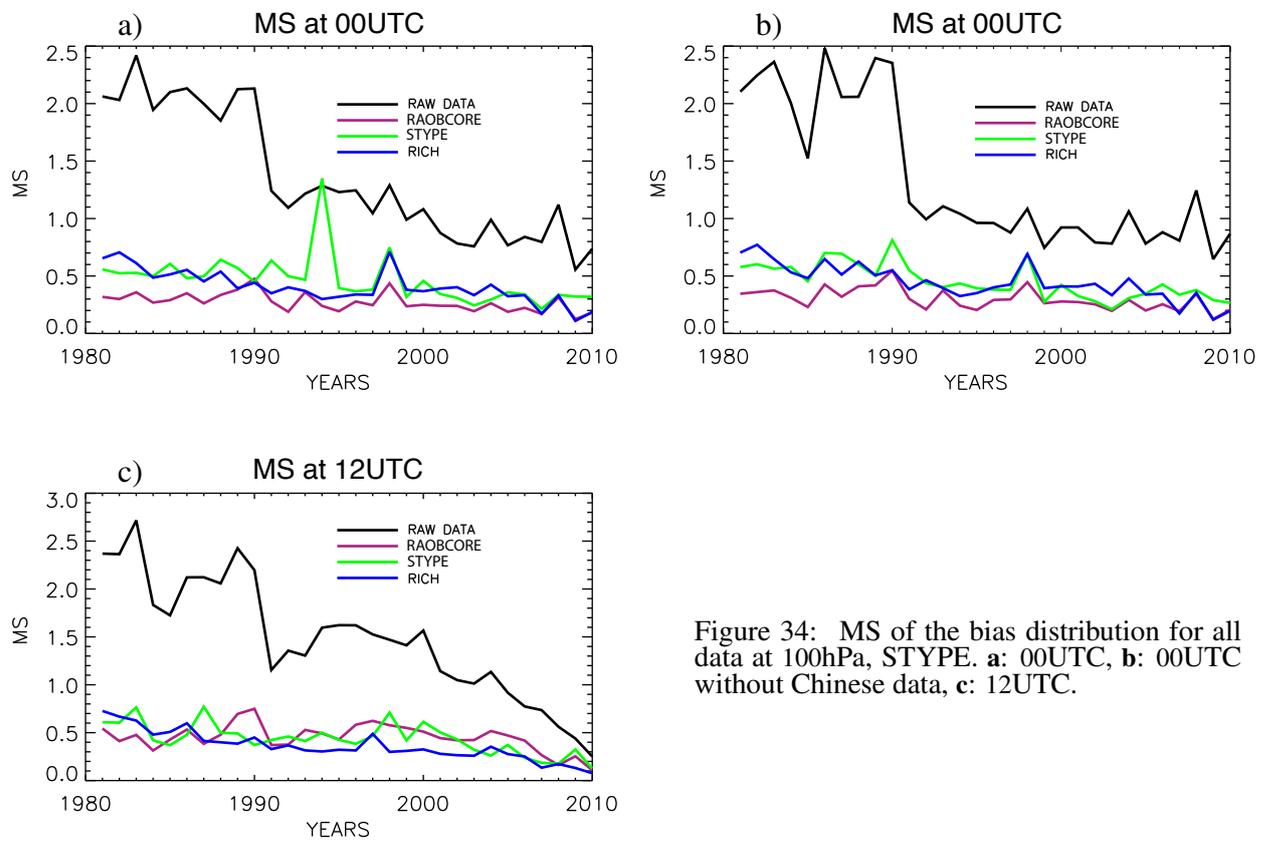
Figure 34: MS of the bias distribution for all data at 100hPa, STYPE. **a**: 00UTC, **b**: 00UTC without Chinese data, **c**: 12UTC.
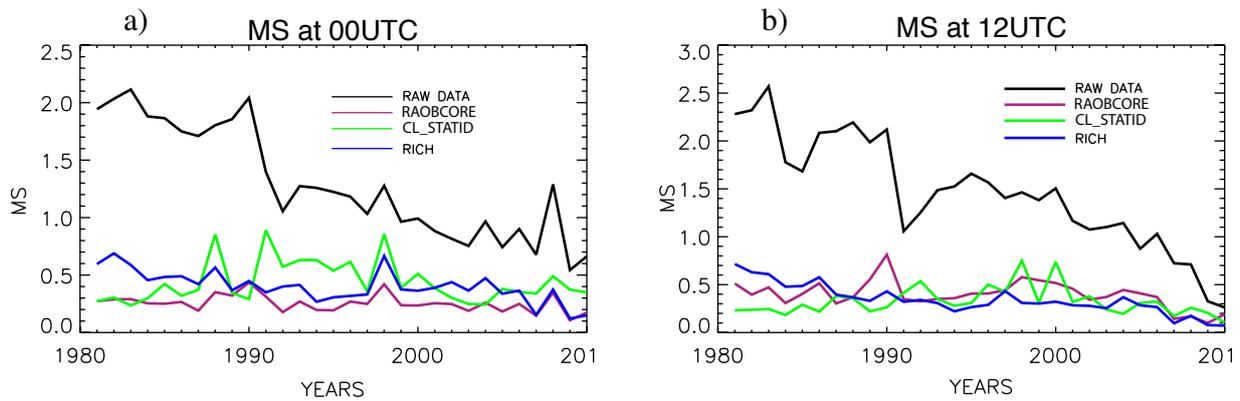
Figure 35: MS of the bias distribution for all data at 100hPa, CL_STATID. **a**: 00UTC, **b**: 12UTC.
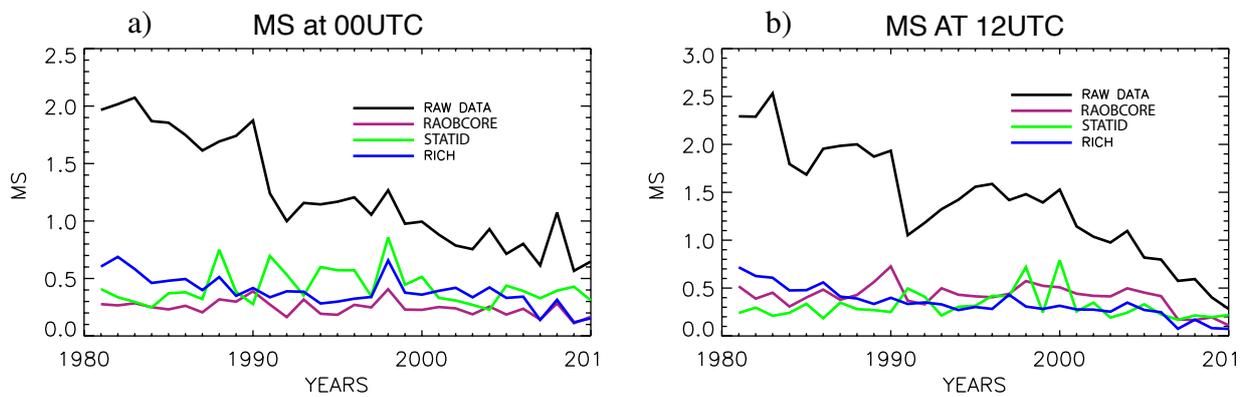
Figure 36: MS of the bias distribution for all data at 100hPa, STATID. **a**: 00UTC, **b**: 12UTC.

5 Summary

This paper has searched for suitable models to describe radiosonde temperature bias profiles with a small set of predictors. The bias model should be suitable for implementation in a large-scale variational bias correction system as it is operational at ECMWF. Background departures (FG) from ERA-Interim between 1979 and 2010 were used as main data source.

The method developed is based on a linear regression, using pressure levels and solar angle as predictors, and a clusterization, based on a weighted Euclidean metric, with different weights for different levels and taking into account the solar elevation. Two different approaches are presented, one based on metadata (derived from Schroeder's database) which group stations with the same sonde_type (CL_STYPE) and one which clusterizes the stations (CL_STATID) based on their FG. In order to evaluate the effect of the clusterization a comparison with the raw data and with a linear regression without any clusterization is done, for the sonde_type approach using STYPE and for the station ID approach using STATID. The results of the corrected radiosonde profiles are also compared with background forecast time series and with profiles corrected with the existing RAOBCORE and RICH bias adjustment systems (Haimberger et al. [2012]).

An important outcome of this work is the strong documented variability of the bias for stations with the same sonde_type, which contrasts our assumption when we started this work. The documented information about radiosonde instrumentation is in too many cases not sufficient to characterize the bias. Generally speaking the global mean residual biases are smaller when CL_STYPE and STYPE are used, but the spatial trend heterogeneity is not much improved. CL_STATID and STATID, show improvements to the raw data comparable or better than RICH and RAOBCORE. In some cases the lack of statistics using the STATID approach can lead to relatively high scatter in the solutions of the linear regression, for this reason a approach using CL_STATID seems preferable. Overall the results suggest that such a suitable linear model based on clusters of stations with similar bias characteristics, variational estimation of biases may work as well as classical approaches based on time series analysis, such as RAOBCORE and RICH.

A work in progress is testing the use of the online variational bias correction of radiosonde temperature profiles, using the output of this study. VarBC need to set time-constants to the observation groups depending on the amount of the data in the group. One advantage of the clusterizaton could be the implementation of different time-constant depending on the amount of the data present in the cluster, given the VarBC the possibility of adjust faster (slower) bigger (smaller) clusters (Andrew Lorenc personal communication).

References

U. Andrae, N. Sokka, and K. Onogi. The radiosonde temperature bias correction in ERA-40. ERA-40, Project Report Series No. 15, ECMWF, Reading, UK, 2004.

D. M. Barker, W. Huang, Y.-R. Guo, and A. Bourgeois. A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note NCAR/TN-453 + STR, NCAR, UCAR Communications, P.O. Box 3000, Boulder, CO 80307., 2003.

F. Bouttier and P. Courtier. Data assimilation concepts and methods. Training course notes, ECMWF, 1999.

R. Daley. *Atmospheric Data Analysis.* Cambridge University Press, 1991.

D. P. Dee. Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, 131:3323–3343, 2005.

D.P. Dee. Variational bias correction of radiance data in the ECMWF system. Technical report, ECMWF, Reading, UK, June 2004.

D.P. Dee, S. M Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hlm, L. Isaksen, P. Kallberg, M. Khler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, CTavolato, J.N. Thpaut, and F. Vitart. The ERA Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, 137.656:553–597, 2011.

J.C. Derber and S. W. Wu. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, 126:2287–2299, 1998.

D.J. Gaffen. Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.*, 99:1984–2012, 1994.

L. Haimberger. Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. *J. Climate*, 20:1377–1403, 2007.

l. Haimberger, C. Tavolato, and S. Sperka. Toward Elimination of the Warm Bias in Historic Radiosonde Temperature Records Some New Results from a Comprehensive Intercomparison of Upper-Air Data. *J. Climate*, 21.18:4587–4606, 2008.

l. Haimberger, C. Tavolato, and S. Sperka. Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *J. Climate*, 25.23:8108–8131, 2012.

H. Hersbach, P. Poli, and D. Dee. The observation feedback archive for ICOADS and ISPD data sets. ERA Report Series 18, ECMWF, Reading, UK, 2015.

E. Kalnay. *Atmospheric modelling, data assimilation and predictability.* Cambridge University Press, 2003.

A.C. Lorenc. Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112:1177–1194, 1986.

J.K. Luers. Temperature Error of the Vaisala RS90 Radiosonde. *J. Atmos. Oceanic Technol.*, 14:1520–1532, 1997.

K Onogi. The long-term performance of the radiosonde observing system to be used in ERA-40. ERA-40, Project Report Series No. 2, ECMWF, Reading, UK, 2000.

W.J. Randel and F. Wu. Biases in Stratospheric and Tropospheric Temperature Trends Derived from Historical Radiosonde Data. *J. Climate*, 19:2094–2104, 2006.

H.C. Romesburg. *Cluster Analysis for Researchers*. Wadsworth/ Lifetime Learning Publications, 1989.

B.D. Santer, T.M.L. Wigley, C. Mears, F.J. Wentz, S. A. Klein, D.J. Seidel, K.E. Taylor, P.W. Thorne, M.F. Wehner, P.J. Gleckler, J.S. Boyle, W.D. Collins, K.W. Dixon, C. Doutriaux, M. Free, Q. Fu, J.E. Hansen, G. S. Jones, R. Ruedy, T. R. Karl, J. R. Lanzante, G. A. Meehl, V. Ramaswamy, G. Russell, and G. A. Schmidt. Amplification of surface temperature trends and variability in the tropical atmoschere. *Science*, 309:1551–1555, 2005.

S. R. Schroeder. Completing instrument metadata and adjusting biases in the radiosonde record to allow determination of global precipitable water trends. Preprints, 12th Symp. On Meteorological Observations and Instrumentation, Long Beach, CA, 2003.

S.C. Sherwood. Simultaneous detection of climate change and observing biases in a network with incomplete sampling. *J. Climate*, 20:4047–4062, 2007.

S.C. Sherwood, J. Lanzante, and C. Meyer. Radiosonde daytime biases and late-20th century warming. *Science*, 309:1556–1559, 2005.

S. Uppala, P. Kållberg, A. Hernandez, S. Saarinen, M. Fiorino, Li Xu, K. Onogi, N. Sokka, U. Andrae, and V. Da Costa Bechtold. ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957-2002. *ECMWF Newsletter*, 101:2–21, 2004.

S.M Uppala et al. The ERA40 reanalysis. *Quart. J. Roy. Meteor. Soc.*, 131.612:2961–3012, 2005.

V. Venema, O. Mestre, and E. Aguilar. Description of the cost-home monthly benchmark dataset with temperature and precipitation data for testing homogenisation algorithms. Technical report, Universität Bonn, 2008. http://www.meteo.uni-bonn.de/mitarbeiter/venema/themes/homogenisation/description_monthly_benchmark_dataset.pdf.

D.S. Wilks. *Statistical methods in the atmospheric sciences*. Elsevier, 2 edition, 2006.

D.S. Wilks. *Statistical methods in the atmospheric sciences*. Elsevier, 3 edition, 2011.

D Zupanski. A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Mon. Wea. Rev.*, 125:2274–2292, 1997.