Meteorologische Zeitschrift, Vol. **24**, No. 5, 525–544 (published online August 31, 2015)

Article

Upper-air observations from the German Atlantic Expedition (1925–27) and comparison with the Twentieth Century and ERA-20C reanalyses

Alexander Stickler[1,2*], Samuel Storz[2], Richard Wartenburger[1,2,6], Hans Hersbach[3], Gilbert P. Compo[4,5], Paul Poli[3], Dick Dee[3] and Stefan Brönnimann[1,2]

[1]Oeschger Centre for Climate Change Research, Bern, Switzerland
[2]Institute of Geography, Univ. Bern, Bern, Switzerland
[3]European Centre for Medium-Range Weather Forecasts, Reading, UK
[4]CIRES, Univ. of Colorado, Boulder, USA
[5]PSD-ESRL, NOAA, Boulder, USA
[6]Now at Institute of Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

(Manuscript received March 23, 2015; in revised form July 3, 2015; accepted July 7, 2015)

Abstract

Between April 1925 and June 1927, the research vessel *Meteor* cruised the tropical and southern Atlantic Ocean in the framework of the *German Atlantic Expedition*. One purpose was to systematically explore the vertical structure of the atmosphere. To this end, the ocean was crossed in 14 profiles across parallels of latitude. 801 pilot balloons and 217 kites were launched. The resulting data have been digitised in the framework of the European project ERA-CLIM. Here, they are compared to the Twentieth Century (20CR) and ERA-20C reanalyses, independent datasets based on the assimilation of synoptic pressure and hurricane pressure records, and marine surface winds for the latter, using monthly sea surface temperature and sea ice as boundary conditions. Both reanalyses display similar patterns of systematic differences relative to the observations for temperature, specific humidity, and wind. Furthermore, 20CR and ERA-20C show generally comparable anomaly correlations for all parameters, with the highest values found for pressure and temperature. In the southern extratropics, high (> 0.75) anomaly correlations are found for pressure in both 20CR and ERA-20C, and for temperature in 20CR. Medium (> 0.5) anomaly correlations are found for specific humidity in 20CR. Moderate anomaly correlations (> 0.44) are found for meridional wind in both 20CR and ERA-20C, and for temperature in ERA-20C. In contrast, low anomaly correlations (< 0.44) are found for zonal wind both in 20CR and ERA-20C, and for specific humidity in ERA-20C. In the Tropics, low anomaly correlations are found for all parameters except for pressure, which shows medium anomaly correlations for both 20CR and ERA-20C, and for meridional wind, which shows moderate anomaly correlations for 20CR. In all regions, both reanalyses strongly underestimate the observed range of zonal and particularly meridional wind variability. Even though remaining errors in the observational data cannot be excluded, we estimate that the inherent observational uncertainties do not alter our conclusions. Vice versa, two pieces of evidence support the credibility of the early upper-air data: the robust regressions of both 20CR and ERA-20C against observed pressure and temperature over a large spatial and temporal range, and the similarity between the uncertainties predicted by 20CR and the actual uncertainties determined from the root mean square difference of reanalysis and observation values.

Keywords: Twentieth Century Reanalysis, ERA-20C, upper-air data, ERA-CLIM, CHUAN, kite, pilot balloon, *German Atlantic Expedition*, *Meteor*, Atlantic, temperature, pressure, wind, circulation, humidity, Tropics, extratropics

1 Introduction

There is a considerable lack of knowledge about the atmospheric circulation above the tropical and southern hemispheric ocean basins back in time. This is especially true for the period before 1957, The International Geophysical Year, when the global radiosonde network as present today became fairly well established. However, detailed information on the state of the circulation in the Tropics back in time is very important, e.g. for understanding climatic variability, including changes in the global energy budget.

It may be surprising that many observational data from the early 20th century are in fact available, albeit in much lower absolute quantity compared to nowadays. One problem is that these data are often not yet in digital form, making them useless for modern quantitative study and for derived data products such as reanalyses.

An example of such a rich, historical upper-air data source in a generally data-sparse region is the German research vessel *Meteor*, which extensively collected

*Corresponding author: Alexander Stickler, Oeschger Centre for Climate Change Research and Institute of Geography, Univ. Bern, Hallerstrasse 12, 3012 Bern, Switzerland, e-mail: alexander.stickler@giub.unibe.ch

DOI 10.1127/metz/2015/0683

Gebrüder Borntraeger Science Publishers, Stuttgart, www.borntraeger-cramer.com

aerological data above the tropical Atlantic Ocean during the *German Atlantic Expedition* 1925–27 (Kuhlbrodt and Reger, 1933). Besides a detailed topographical survey of the tropical and South Atlantic (ca. 67,000 echo soundings), and geological, biological and chemical analyses, conceived to advance the knowledge on the global oceanic circulation (Hoheisel-Huxmann, 2007), a further goal of the expedition was to systematically explore the vertical structure (e.g., temperature, pressure, humidity and circulation) of the atmosphere above the Ocean, from the northern subtropics to the southern high midlatitudes. For this purpose, more than 1,000 aerological soundings were launched on board the ship, mainly between 20 ° N and 64 ° S. In this latitudinal range, the Atlantic Ocean was crossed in 14 horizontal transects with latitudinal distances of approximately 7°, stretching more or less along latitude parallels. A distance of 67,535 nautical miles (≈ 125, 075 km) was covered during the whole journey, more than three times the Earth's circumference. This probably represents the spatially most comprehensive upper-air probing above the Atlantic to the present day.

The *Meteor* data are part of a larger, historical upper-air dataset that contains significant amounts of pre-1957 upper-air data (> 1.3 million station days catalogued, > 200, 000 images taken, > 700, 000 station days digitised), spanning large parts of the globe and focussing on so far poorly represented regions such as the Tropics, the polar regions and the Oceans, and on very early 20th century upper-air data from Europe and the US. These data have been made available in digital form by the EU FP7 project ERA-CLIM (European ReAnalysis of global CLIMate observations; http://www.era-clim.eu, Stickler et al., 2014ab). They can help to assess new, totally independent datasets such as the Twentieth Century Reanalysis (20CR; Compo et al., 2011) and ERA-20C (Poli et al., 2013), reanalyses that assimilate exclusively surface pressure observations and hurricane pressure reports (and additionally marine surface winds in the case of ERA-20C), and prescribe monthly sea surface temperature and sea ice as boundary conditions. Eventually, the data will be assimilated into full reanalyses that make use of historical upper-air data.

Earlier studies (Compo et al., 2011; Brönnimann et al., 2011; Zhang et al., 2013) have found surprisingly high correlations of 20CR with observations of temperature and geopotential height in the southern subtropics and midlatitudes for the early 20th century, in spite of the very low input observational density. On the other hand, some studies have also shown that the correlations are generally lower in the Tropics (e.g. Compo et al., 2011) and that large biases exist in this region, e.g., in the wind field above the African monsoon region in the 1940s and 1950s (Stickler and Brönnimann, 2011). A possible conclusion of these results is that a purely pressure-driven reanalysis might be of lower quality over the Tropics compared to midlatitudes, and that the additional assimilation of wind (such as in ERA-20C, albeit only for the surface level) might help to improve the results. Here, we would like to test this hypothesis and assess the quality of the datasets by comparing 20CR and ERA-20C with the observational data from the *Meteor* and with each other.

Section 2 describes the observational and reanalysis datasets, the methods used to interpolate the reanalysis data to the locations and times of the observations, and the methods used to calculate anomalies for the correlation analysis. Section 3 discusses the results of the analysis, separated according to the observational variables (temperature, pressure, specific humidity, zonal and meridional wind). It includes a close examination of their significance. Finally, Section 4 draws conclusions and gives an outlook.

2 Data and methods

In this study, we analyse and compare data from three sources: a primary, observational dataset for the free atmosphere, and the only two previously published atmospheric reanalyses going back before 1948.

2.1 Observational data

We make use of the early, historical upper-air dataset that has been made available by the EU FP7 project ERA-CLIM (Stickler et al., 2014a,b; http://www.era-clim.eu; data and metadata available from the PANGAEA data repository at http://doi.pangaea.de/10.1594/PANGAEA.821222 and from http://www.oeschger-data.unibe.ch/metads). The data have been subject to a raw quality control consisting mainly of range checks. Furthermore, suspicious values were flagged during digitisation and manually re-checked and corrected afterwards, if possible, with the help of digital images of the sources. Finally, the temperature data were additionally screened for large outliers (absolute departures > 30 K) relative to the new, surface only reanalysis ERA-20C (Poli et al., 2013), and those values were again manually re-checked in the sources. The value of 30 K was chosen globally, based on scatter plots of reanalysis departure against ECMWF collection identifiers, as a cut-off value representative of strong outliers. This process led to the correction and/or flagging of 2,325 temperature values for the whole ERA-CLIM upper-air data collection, corresponding to a relatively small share of 2 ‰ of all temperature values collocated with reanalysis values. Therefore, the observational data can still be regarded as virtually independent of ERA-20C. Further details on the data treatment can be found in Stickler et al. (2014b). All the new ERA-CLIM upper-air data will be incorporated in future versions of the Comprehensive Historical Upper-Air Network (CHUAN; Stickler et al., 2010).

For this study, we consider the aerological data that were obtained during the *German Atlantic Expedition* 1925–27 (Kuhlbrodt and Reger, 1933) on

board the research vessel *Meteor*. The *Meteor*, originally conceived as a gunboat for the German Imperial Marine, was never finished as such after the start of World War I. Instead, following a modification in 1923/24, the ship was transferred as a survey vessel to the Imperial Dockyards located in Wilhelmshaven (see e.g. http://de.wikipedia.org/wiki/Meteor_(Schiff,_1915), last accessed 17 January 2014).

Its aerological dataset is split into three separate records (ERA-CLIM record IDs 71–73 in the moving upper-air inventory), which correspond to the three observational platforms that were used during the expedition: pilot balloons to determine wind direction and speed; kites to measure air pressure, air temperature, humidity, wind direction and speed; and finally, registering balloons that observed air pressure, air temperature, and humidity. All kite and registering balloon observations are used from both standard geometrical altitude levels (6, 200, 500, 1,000, 1,500, ... m asl plus highest level reached) and additional, arbitrary geometrical altitude levels.

More than 1,000 ascents with a wide geographical distribution are analysed. The pilot balloon record contains 801 ascents going to altitudes up to 20,500 m asl (median height reached 4,500 m asl). The kite record contains 217 soundings reaching maximum heights of 4,870 m asl (median 2,165 m asl). The few, freely drifting registering balloons complement the kite observations, extending the vertical range of temperature, humidity, and pressure observations to a maximum of 14,700 m asl (median 6,645 m asl). However, the registering balloon record will not be analysed here due to the very low number of observations. The majority of the soundings are located between 20 ° N and 65 ° S. Within this latitudinal range, the Atlantic Ocean was crossed in 14 sections oriented largely along parallels of latitude. A map showing the complete journey is shown in Fig. 1. Fig. 2 shows the locations of all pilot balloon and kite soundings. Fig. 3 illustrates the launch of the pilot balloons and kites on board the *Meteor* from original photos taken during the expedition (Kuhlbrodt and Reger, 1933). More comprehensive information about the records can be found in the ERA-CLIM metadatabase (http://www.oeschger-data.unibe.ch/metads) and on http://doi.pangaea.de/10.1594/PANGAEA.821222.

Figure 1: Map showing the complete trajectory of the journey of the research vessel *Meteor* during the *German Atlantic Expedition* from 16 April 1925 to 2 June 1927 (from Spiess, 1928).

2.2 Reanalyses

The Twentieth Century Reanalysis version 2 (20CR; Compo et al., 2011) output has a horizontal resolution of 2° and 24 levels in the vertical from 1,000 to 10 hPa (the atmospheric model is integrated on a total of 28 hybrid sigma-pressure levels). It covers the period from 1871 onward. It uses a 2008 experimental version of the NCEP Global Forecast System (GFS) coupled atmosphere-land model integrated at a spectral resolution of T62 (grid size approximately 205 km) and at a time step of 20 minutes. A Simplified Arakawa-Schubert convection scheme with momentum mixing (Moorthi et al., 2001) is used. Further parameterisations are described in Compo et al. (2011). An Ensemble Kalman Filter technique (Whitaker and Hamill 2002, Whitaker et al. 2004, Compo et al. 2006, Compo et al. 2011) is used for the assimilation. Here, we use the ensemble mean, which is calculated from 56 equally-likely members.

The newly available ERA-20C extended climate reanalysis (Poli et al., 2013) has been produced at the ECMWF in the framework of ERA-CLIM and covers the period 1900–2010. The atmosphere, ocean wave, and land surface model configurations are those of the ECMWF Integrated Forecast System (IFS) operational model as of December 2012, described in the documentation of IFS version cycle 38R1 (ECMWF, 2013). The model uses a convection parameterisation following Tiedtke (1989). The atmospheric part of the model was integrated at a spectral resolution of T159 (grid size approximately 125 km) on 91 model levels (in addition output is also available on 37 pressure levels from 1,000 to 1 hPa), and with an integration time step of 30 min. The data assimilation system relies on a ten-member Ensemble of Data Assimilations (EDA) (Isaksen et al., 2010). Each member employs a 4D-Var analy-

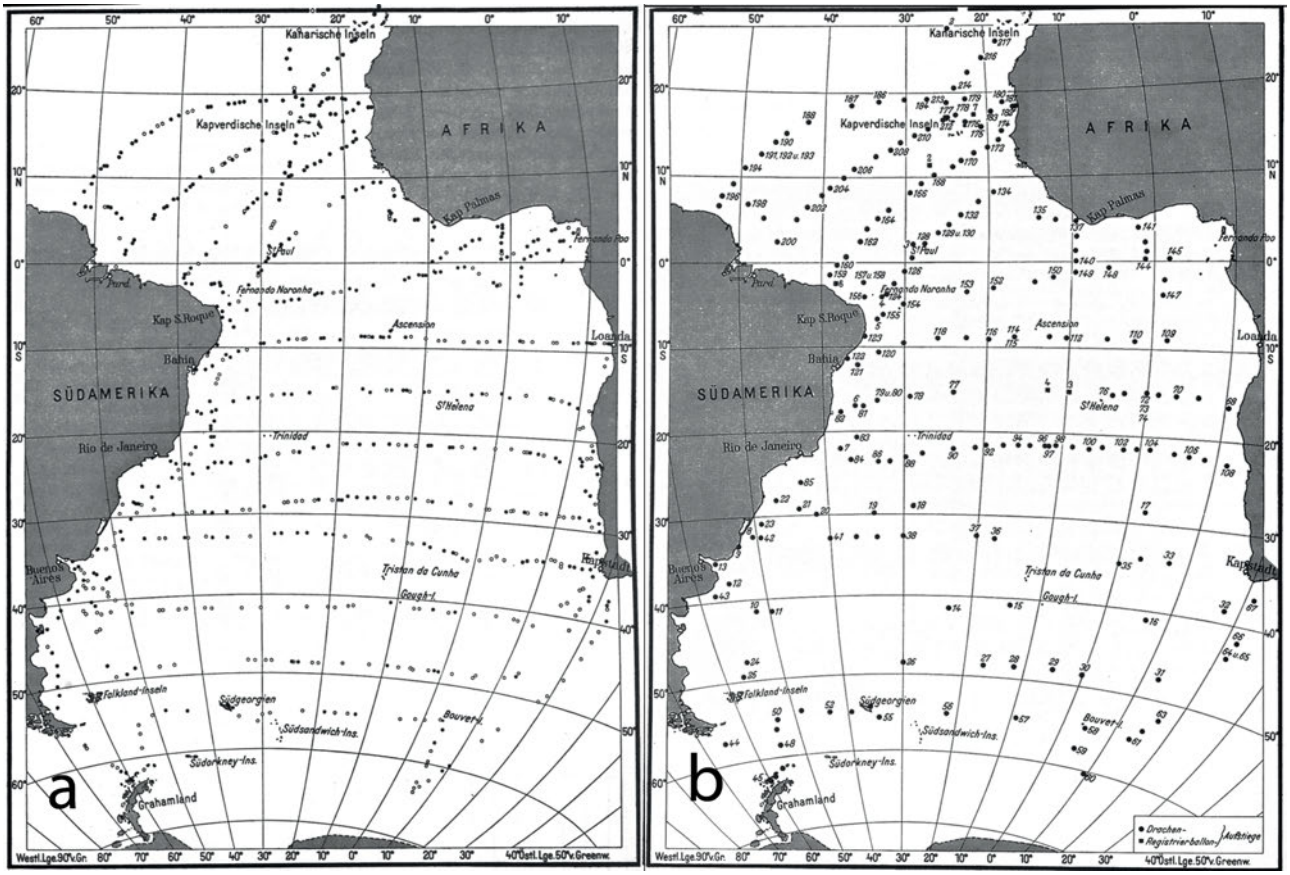

Figure 2: Maps showing the locations of (a) all pilot balloon ascents (b) all kite ascents made from the research vessel *Meteor* during the German Atlantic Expedition from 16 April 1925 to 2 June 1927 (from Kuhlbrodt and Reger, 1933)

sis scheme with an analysis window of 24 h instead of 12 h as in ERA-Interim (Dee et al., 2011) and in the ECMWF operational forecasts. A variational bias correction (VarBC; Dee, 2004) is employed within the IFS 4D-Var to correct for biases in surface pressure observations. The analysed fields we study here are from the deterministic run.

Both 20CR and ERA-20C assimilate synoptic pressure from the International Surface Pressure Data Bank (ISPD V2 for 20CR, V3.2.6 for ERA-20C; http://rda.ucar.edu/datasets/ds132.0/) and from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, version 2.5; http://icoads.noaa.gov/; Worley et al., 2005; Woodruff et al., 2010). Only slight differences exist around the tropical and south Atlantic between the used versions of ISPD: In particular, data from the island of St. Helena in the central South Atlantic are available in V 3.2.6 that were not yet present in V2. Furthermore, both reanalyses assimilate hurricane "best track" surface pressure values from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010). Very recent investigations by one of the co-authors at ECMWF have shown that many hurricane track pressure observations were erroneously discarded in ERA-20C, especially in the first half of the 20th century. Fig. S1 shows that, for the years 1900 to 1949, 20CR fitted the tropical cyclone central pressure estimates from IBTrACS well, whereas ERA-20C did not (note that the number of IBTrACS hurricane track pressure observations has increased a lot between the time of the creation of 20CR and the time of creation of ERA-20C, explaining the different numbers of collocations in the figure). Further investigations into the source of this problem identified the quality control step in ERA-20C that checks for constant time series in order to reject the data from stuck pressure sensors as cause. This procedure is inadequate for observations of tropical cyclone central pressure, where the intensity may be constant for several days, while only the cyclone position changes. However, we emphasise that this issue should not affect our analysis, since the *Meteor* observations north of 10 °N were done exclusively from April to May 1925 and from February to May 1927, outside the North Atlantic hurricane season.

In addition to surface pressure and unlike 20CR, ERA-20C assimilates surface marine winds. Further differences between the input datasets to the reanalyses exist in the applied boundary conditions: Whereas 20CR uses sea surface temperature and sea ice from HadISST1.1 (Rayner et al., 2003), ERA-20C uses one of the 10 equally plausible sea surface temperature/sea ice evolutions from the more recent

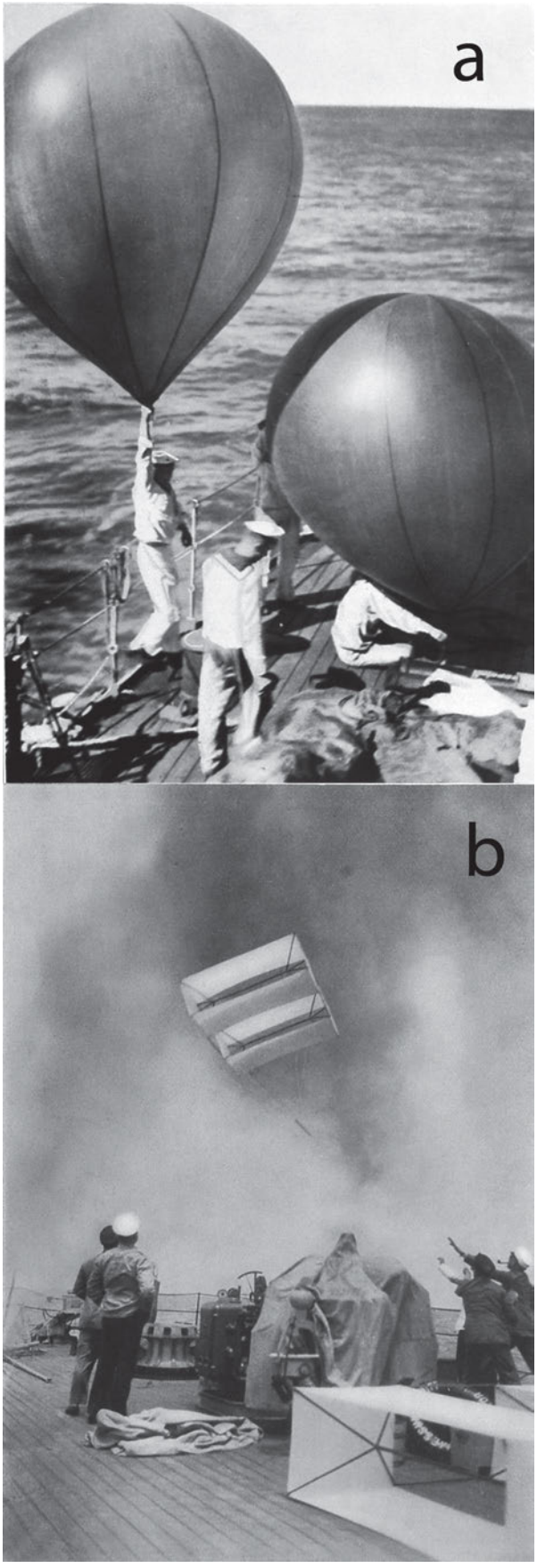

Figure 3: Photos from the original data publication by Kuhlbrodt and Reger (1933) showing (a) the filling of a registering balloon and (b) the landing of a kite during the *German Atlantic Expedition* 1925–27 on board the *Meteor*.

dataset HadISST2.1.0.0 (http://www.metoffice.gov.uk/hadobs/hadisst2/; Rayner et al., manuscript in preparation). And whereas during the period of this study, 20CR uses monthly volcanic aerosols, bi-annual averages of the time-varying global mean CO_2 concentration, and annual values of the incoming solar radiation as described in Saha et al. (2010), ERA-20C applies a transient atmospheric forcing and land surface evolution as prescribed in the CMIP5 experiments (WCRP, 2011). Hersbach et al. (2015) give a complete description of the model setup.

Fig. 4 shows the global distribution of synoptic pressure stations from the International Surface Pressure Data Bank (ISPD; http://rda.ucar.edu/datasets/ds132.0/) from which data have been assimilated into both reanalyses during the *German Atlantic Expedition* (exemplarily showing the year 1926 from V2; 1925 and 1927 look very similar). The number of stations is limited to five around the tropical and South Atlantic (two at the southern tip of South America, one in northern South America, one on the Canary Islands, and one in central East Africa). As mentioned above, the newer ISPD version used in ERA-20C additionally has data from St. Helena.

To illustrate the density of assimilated ship surface pressure observations contained in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, version 2.5; http://icoads.noaa.gov/; Worley et al., 2005; Woodruff et al., 2010), Fig. 5 shows the cumulative distribution of sea-level pressure observations in ICOADS for the year 1926. In the Atlantic, ship observations are relatively abundant along the main shipping routes from Europe to North and Central America, and from Europe to Brazil and Argentina, with 150–200 observations (or one observation every 2–4 days) for each $1° \times 1°$ grid cell, and even higher values along the European coast (up to 500–600 observations or ca. 1–2 observations per day on average). The density of observations is significantly lower in other regions of the North Atlantic and along the eastern shipping route of the South Atlantic towards the Gulf of Guinea and the Cape of Good Hope (30–100 observations or one observation only every 4–12 days on average). In a smaller part of the central Atlantic between Africa and Brazil and in large areas of the South Atlantic, the density is very sparse with less than 3 observations in a year.

Because they are not using any upper-air data, both 20CR and ERA-20C are completely independent from the ERA-CLIM observational upper-air dataset. Their intercomparison presents an interesting assessment of the reliability of the reanalyses as well as the quality of the observational data, particularly in the case of the sparsely observed regions (even in terms of a surface quantity like sea-level pressure) traversed by the *Meteor*. To date, no published reanalysis has made use of the considerable amount of historical upper-air data before 1948, even though reanalyses are expected to significantly profit from assimilating further historical upper-air data.

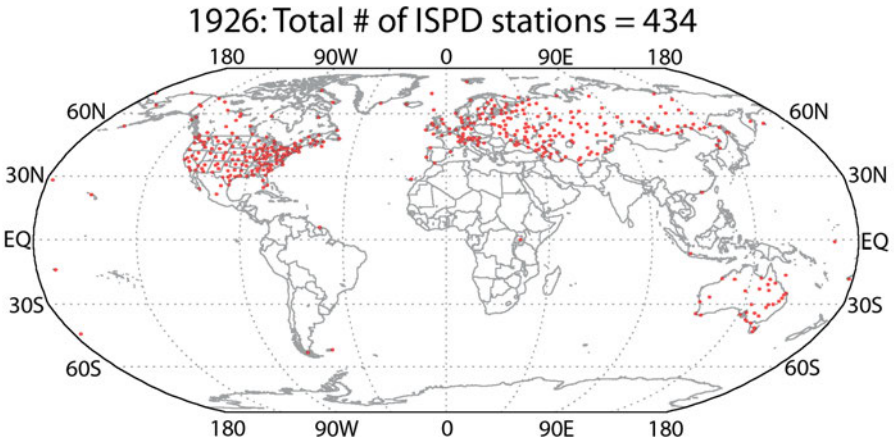

Figure 4: Global distribution of the International Surface Pressure Data Bank (ISPD Version 2) synoptic pressure stations for the year 1926 (adopted from: http://www.esrl.noaa.gov/psd/data/ISPD/v2.0/img/Map_ispd-1926.png, last visited 17 January 2014). Some additional stations, particularly the island of St. Helena, are available for ISPD V3.2.6, used in ERA-20C (Figure slightly adopted, courtesy of Xungang Yin).

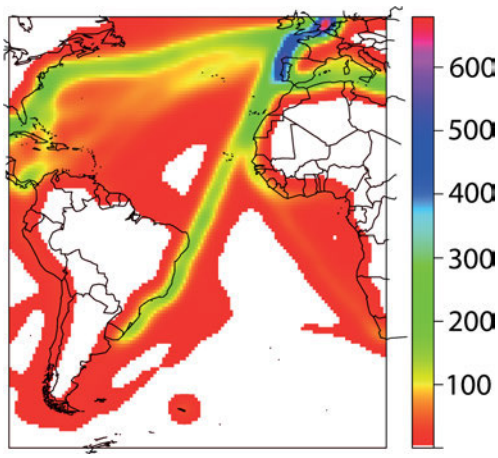

Figure 5: Regional cumulative distribution of International Comprehensive Ocean-Atmosphere Data Set (ICOADS, version 2.5) sea-level pressure observations for the year 1926. The plot was created with the R statistics software (http://www.r-project.org/), using the spatstat package and the density function with a Gaussian smoothing kernel.

2.3 Methods

Here, we use the $2° \times 2°$ temperature, pressure, wind, and specific humidity data on pressure levels extracted from the ensemble mean analysis of 20CR and from the ERA-20C deterministic run for the launch times and locations of the observations in the observational dataset. For 20CR, ERA-20C, and an NCEP/NCAR 50-Year Reanalysis (NNR; Kistler et al., 2001) four-times daily 30-year climatology (1981–2010, available on http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.pressure.html), the extraction in time was performed by taking the reanalysis data at the closest available reanalysis time to the launch time of the ascents. For kite ascents, which sometimes took several hours for ascent and descent, observation levels with a difference > 3 h between the launch time of the ascent and the observation were discarded to ensure comparability. In all cases, the horizontal, spatial extraction was done at the closest grid points of the reanalyses. As an alternative to selecting the closest grid point, an interpolation to the exact location might be done. However, the differences are arguably small compared to other sources of uncertainty, including the displacement of the balloon with height. The vertical interpolation from the pressure levels of the reanalysis to the geometrical altitude levels of the observational data was performed using the reanalyses' geopotential height data, assuming linear vertical profiles with respect to geometrical height. Before beginning the analysis, two positive, unrealistic outlier temperature values (> 50 °C) were deleted from the kite temperature data for the subsequent study. Additionally, the observations and reanalysis values corresponding to three unrealistic altitude and two unrealistic pressure values were identified by using scatter plots of observed pressure vs. altitude. The values connected to wrong altitudes were discarded from the study, the wrong pressure values in the observations were corrected by going back to images of the original source.

Both the observational and the extracted reanalysis data were first pooled into latitude-height bins (size $10° \times 500$ m, and $10° \times 1,000$ m for levels above 4,000 m asl in the case of pilot balloon data, centred on whole tens of degrees and multiples of 250 and 500 m, respectively). Then, anomalies of the observations and of reanalysis fields were calculated relative to the NNR climatology, interpolated (as in the case of

Table 1: Mean/Median Spearman rank correlation coefficients between the 20CR/ERA-20C reanalysis and the *Meteor* observed anomalies of different variables during the *German Atlantic Expedition* 1925–27. The values given are calculated as average over all latitude-altitude bins with $N > 15$. *Values for kite observations up to 4,500 m asl. #Values for pilot balloon observations up to 21.5 km asl.

	Tropics (25 ° S–25 ° N)		Southern Extratropics (65 ° S–25 ° S)	
	20CR	ERA-20C	20CR	ERA-20C
Temperature*	0.32/0.34	0.25/0.24	0.80/0.84	0.47/0.55
Pressure*	0.68/0.75	0.66/0.75	0.82/0.91	0.78/0.84
Specific Humidity*	0.22/0.21	0.31/0.35	0.62/0.65	0.42/0.43
Zonal Wind*	0.43/0.50	0.52/0.52	0.44/0.58	0.40/0.37
Meridional Wind*	0.27/0.26	0.25/0.33	0.52/0.63	0.75/0.73
Zonal Wind#	0.34/0.41	0.35/0.32	0.41/0.43	0.28/0.34
500–1,000 m asl	0.55/0.53	0.67/0.61	0.46/0.43	0.57/0.55
lowest 3 km asl	0.52/0.54	0.56/0.58	0.43/0.50	0.41/0.50
Meridional Wind#	0.47/0.50	0.24/0.26	0.49/0.58	0.48/0.63
500–1,000 m asl	0.55/0.50	0.63/0.62	0.56/0.53	0.71/0.71
lowest 3 km asl	0.35/0.38	0.38/0.37	0.57/0.59	0.63/0.66

the 20CR/ERA-20C data) to the times and locations of the observations. The horizontal archived resolution of NNR is 2.5 ° on 17 levels in the vertical. 20CR, ERA-20C and the observations are then compared by 1) calculating the time mean difference (bias) b_X of the reanalyses relative to the observations and by 2) calculating Spearman rank correlation coefficients $r_{S,X}$ between the time series of the anomalies of parameter X in the reanalyses and the observations. Spearman rank correlation coefficients have the advantage of being more robust against outliers than regular (Pearson's) correlation coefficients, in which the values or anomalies themselves are directly correlated with each other instead of with their ranks.

The 20CR anomaly correlations do not need to be corrected (as for an ensemble mean calculated from ensemble members of a simulation), and are directly comparable to the ERA-20C anomaly correlations: firstly, the ensemble mean in the Ensemble Kalman Filter used in 20CR is the same solution as the single 4-D Var solution under certain, usually valid assumptions; and secondly, the variance of both the 20CR ensemble mean and of ERA-20C are generally very similar.

We used the statistical software R (http://www.r-project.org) for all analyses described here. The minimum number of observation-reanalysis pairs used for the calculation of the Spearman rank anomaly correlation values shown in the figures is 15. All anomaly correlations with $N > 15$ were found to be significant at the 90 % level according to the significance test described below, and mean and median values of the anomaly correlations in Table 1 are given for all bins (non-significant and significant), again with a pair number $N > 15$.

We decided to use a minimum threshold for N for the anomaly correlation plots because a statistical significance of a correlation at the 90 % significance level for a relatively small number of values is not very meaningful with respect to the magnitude of the correlation, i.e., the confidence interval for the correlation coefficient becomes very large. E.g., with $Si = 90\,\%$ and $N = 10$, a correlation coefficient of $r = 0.5$ would have a confidence interval (at the same significance level) of $(-0.07, 0.82)$, i.e., not even the sign of the correlation would be certain. For $r = 0.6$, the confidence interval would begin just above zero $(0.07, 0.86)$. And for large correlation coefficients, e.g., $r = 0.9$, the confidence interval would still be relatively wide $(0.69, 0.97)$. However, $r = 0.5$ and $r = 0.6$ are typical values of correlation coefficients found in our study. Therefore, we opted for a higher threshold. In case of $N = 16$, the same confidence bands become $(0.09, 0.77)$, $(0.23, 0.82)$ and $(0.77, 0.96)$. The choice to use a threshold value of $N > 15$ for the correlation plots is arbitrary to a certain degree, but can be understood as a compromise between even smaller confidence bands at larger N and the number of available observations, particularly at higher altitude levels. Furthermore, it does not alter the conclusions of the present study.

The statistical significance of the systematic differences is calculated using a one-sided, heteroskedastic t test at the 99 % level (see e.g. Eq. 8–9 of Schönwiese, 1992). The significance of the Spearman rank correlation values is estimated using a two-sided z test on Fisher's transform at the 90 % level (see e.g. Eq. 10–28 and 10–31 of Schönwiese, 1992).

3 Results and discussion

3.1 Temperature

The quantitative comparison of temperature from the reanalyses and observations shows very good agreement for some parts of the study domain, particularly in the extratropics, while other areas show more moderate consistency. Figs. 6a and 7a show the time and zonal mean difference of the temperature in 20CR/ERA-20C relative to the kite observations from the *Meteor*. Both exhibit a similar bipolar pattern with positive differences at lower levels up to 1,500–2,500 m asl between 45 ° S and 5 ° S (up to +3.5 K in ERA-20C), and negative differences at higher levels above 2,000 m asl between 35 ° S and 5 ° N (up to −7 to −8 K in both reanalyses). In the

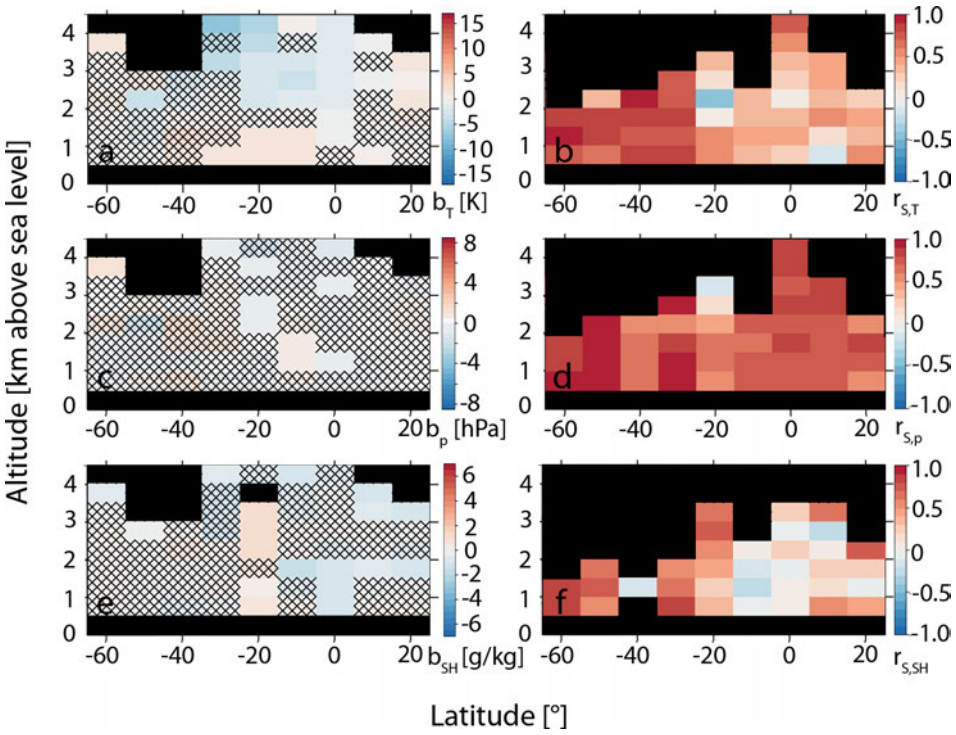

Figure 6: Zonal mean (a, c, e) differences *b* and (b, d, f) Spearman rank anomaly correlations r_S between 20CR and the *Meteor* kite observations of (a, b) temperature, (c, d) pressure, (e, f) specific humidity. Bins without observation-reanalysis pairs (difference) or with less than 15 observation-reanalysis pairs (correlation) are shown in black, bins with differences that are not significant at the 99 % level (a, c, e, one-sided t test) are hatched. All correlations are significant at the 90 % level (b, d, f, two-sided z test). The lowest level (0–500 m asl) has been left out because the vertically interpolated 20CR data have gaps there

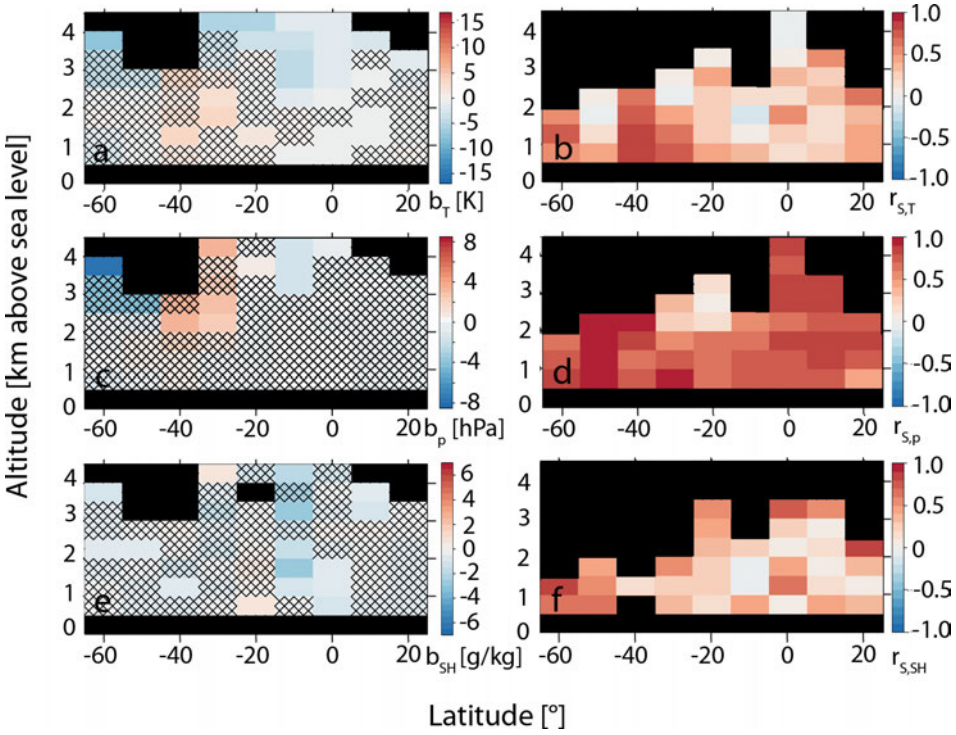

Figure 7: As Fig. 6, but for ERA-20C instead of 20CR.

mean over all available bins with a number of reanalysis-observation pairs $N > 15$, 20CR is slightly colder than the observations ($\overline{b_T} = -0.30$ K), whereas ERA-20C exhibits almost the same temperature as the observations ($\overline{b_T} = -0.01$ K). The standard deviation of the time mean reanalysis-observation difference calculated over all bins with $N > 15$ is 1.3 K for both reanalyses, very close to the estimated uncertainty of the 20CR ensemble mean (time mean of the 20CR ensemble spread of 1.5 K for the same bins).

Over most of the study region, the 20CR temperature anomalies agree better with the observed ones than those in ERA-20C. Figs. 6b and 7b show the zonally averaged Spearman rank correlation coefficients between the 20CR/ERA-20C and *Meteor* temperature anomalies. 42/39 out of 44 altitude-latitude bins with $N > 15$ and significant results ($Si = 90\,\%$, z test) or 95 %/89 % display positive correlations, numbers that can be expected for larger N at this moderate significance level if the "true" (as opposed to estimated) correlation would be positive everywhere (the same holds true for similar statements in the following subsections on pressure and zonal/meridional wind, but not for specific humidity in the case of 20CR, see Section 3.3). The mean correlation coefficient is higher for 20CR than for ERA-20C (0.49 vs. 0.33). Both reanalyses show higher correlations in the southern extratropics (65 ° S–25 ° S) than in the Tropics (25 ° S–25 ° N). For 20CR, correlation values > 0.8 only occur south of 25 ° S (mean 0.80), while the correlations between 25 ° S and 25 ° N are generally moderate to low (< 0.6 and mostly < 0.5; mean 0.32). For ERA-20C, the mean correlation coefficient is 0.47 (median 0.55) in the extratropics and 0.25 in the Tropics.

In the direct comparison between the reanalyses during those times when *Meteor* observations are available, ERA-20C tends to be warmer than 20CR above 1,500 m asl between 45 ° S and 15 ° S and cooler between 15 ° S and 25 ° N and towards the mid troposphere south of 45 ° S (see Fig. S3a in the Supplement). Similar to the correlations between the reanalyses and the observations, the anomaly correlations between the reanalyses are higher on average in the southern extratropics compared to the Tropics (Fig. S3b).

Scatter plots of the absolute temperature values of the reanalyses (as opposed to the anomalies with respect to the NNR climatology) against the observations (Fig. 8, upper row) demonstrate that the (Pearson's) correlations of both reanalyses with observations are high and very close to each other (although slightly higher for 20CR) when analysing the complete dynamic range of the data including their large spatial scale (65 ° S–25 ° N, 0–4,500 m asl) and the annual cycle ($r = 0.95$ vs. $r = 0.93$). Also, the regression coefficient of 20CR is closer to 1 (0.97 vs. 0.93) and the intercept is smaller (+0.32 ° C vs. +0.61 ° C) than for ERA-20C. For ERA-20C, the scatter is generally larger over the whole range of values, and temperatures tend to be underestimated towards the lowest observed values (< −10 °C).

Taylor diagrams of the binned 20CR and ERA-20C versus observed temperature data (anomalies and total values, Fig. 9ab) confirm that the (Pearson's) correlations with 20CR tend to be higher than those with ERA-20C, particularly in the southern extratropics. For both 20CR and ERA-20C, the ratio of the reanalysed to the observed standard deviation covers a relatively large range. For the anomalies, the smallest values of ca. 0.25–0.3 are detected for some bins in the Tropics in ERA-20C, and particularly in 20CR. The largest values of 1.25–1.75 are detected for some extratropical bins, particularly in ERA-20C, but only relatively few bins show values > 1.1. For the total values, the detected spread of the ratio of the reanalysed to the observed standard deviation is somewhat smaller (ERA-20C: range ca. 0.5–1.6 vs. 0.25–1.8, 20CR: range ca. 0.3–1.35 vs. 0.23–1.3), but the correlations of the absolute temperature values are only slightly higher at most than those of the anomaly correlations. This, together with the very high correlation of absolute values found in the scatter plots, points to a good spatial (latitudinal and vertical) representation of absolute temperature, but an only slightly better representation of the seasonal compared to the higher frequency temporal variability in both reanalyses.

3.2 Pressure

While some biases are present, the observed variations of pressure as a function of time and altitude agree extremely well with both reanalyses for almost all regions and altitudes. Figs. 6c and 7c show the time and zonal mean difference of air pressure as a function of altitude in 20CR and ERA-20C, respectively, relative to the kite observations from the *Meteor*. Most differences are not highly significant. However, the significant differences are much larger (up to +3.8 hPa and −8.8 hPa) in ERA-20C than in 20CR. Positive differences in ERA-20C occur mainly at 25 ° S–45 ° S, negative ones at 15 ° S–5 ° S and south of 45 ° S. These differences might imply a poleward shift of the Atlantic sector southern hemisphere subtropical high pressure belt in ERA-20C compared to the observations and 20CR. They could also be caused by individual synoptic systems that are not as well represented in ERA-20C, dominating the bias statistics. In the mean over all available bins with $N > 15$, both 20CR and ERA-20C show practically the same pressure values compared to the observations ($\overline{b_p} = +0.05$ hPa vs. $\overline{b_p} = +0.21$ hPa). The standard deviation of the time mean reanalysis-observation differences calculated over the bins with $N > 15$ is small in both cases with 0.7 hPa and 1.0 hPa. As with temperature, the values are again slightly smaller than the value expected from the estimated uncertainty of the 20CR ensemble mean (the time average of the ensemble spread is 1.1 hPa over the same bins). Interestingly, the *Meteor* observed pressure anomalies $\overline{p'_{\mathrm{obs}}}$ are consistently negative on average (relative to the 1981–2010 climatology) in all latitudinal belts from 25 ° N to 35 ° S, where less

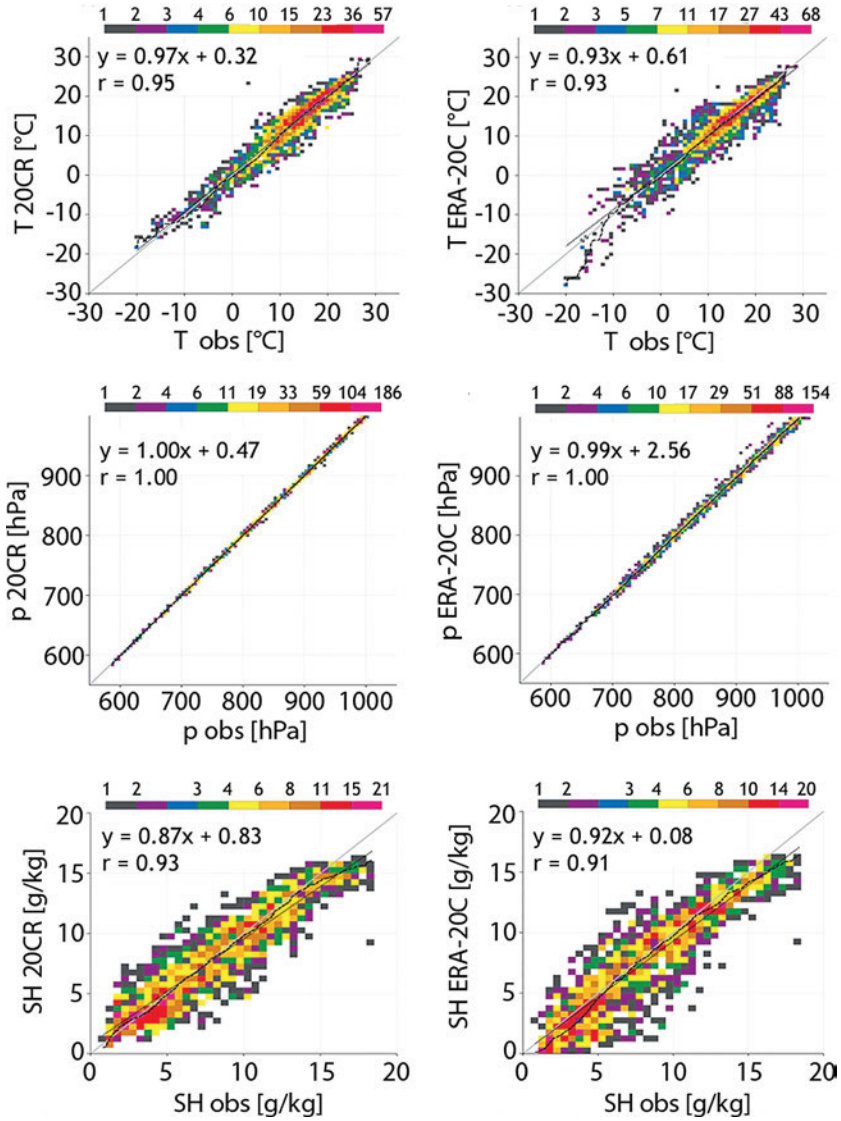

Figure 8: Binned scatter plots of 20CR (left panels) and ERA-20C (right panels) versus *Meteor* observed temperature (upper row), pressure (middle row) and specific humidity (lower row) counts. The bins widths are ca. 0.8 K, 5 hPa and 0.5 g/kg in each direction. Also shown are the 1:1 lines (grey), the linear regression lines (dotted) and cumulative distribution functions (dashed) of the reanalysis data versus the observed values and the Pearson's correlation coefficients.

"noise" from weather systems would be expected on the sub-monthly time scale compared to the midlatitudes, i.e. where we might be able to see 20th century long-term trends (see supplementary Fig. S2). These negative differences (and possibly positive trends from the mid-1920s to the late 20th century) are at least broadly consistent with the rather positive NNR sea-level pressure (SLP) long-term trends of up to 3 hPa in the latitudinal band from 45 ° S to 45 ° N in the Atlantic basin during 1948–1998 reported in Gillett et al. (2003). The strongly positive differences south of 55 ° S are also consistent with the strongly negative SLP trend of up to −12 hPa in NNR between 55 ° S and 65 ° S found in the same study. However, due to the high synoptic variability in the southern midlatitudes, it is not clear whether the relatively few *Meteor* observations can be considered representative of the mid-1920s climatology of air pressure in the Atlantic south of 55 ° S.

Figs. 6d and 7d show the zonally averaged Spearman rank correlation coefficients between the 20CR/ERA-20C and *Meteor* pressure anomalies. 43/44 out of 44 altitude-latitude bins with $N > 15$ and significant results ($Si = 90\,\%$, z test) display positive correlations. The mean correlation coefficients are similar and higher than those for temperature anomalies with 0.73/0.71. The correlation values are higher on average south of 25 ° S (20CR: mean 0.82, median 0.91, ERA-20C: mean 0.78, median 0.84), but the difference between the datasets in the region between 25 ° S and 25 ° N is not as large as in the case of temperature (20CR: mean 0.68, median 0.75, ERA-20C: mean 0.66, median 0.75).

The pressure biases of ERA-20C relative to 20CR (Fig. S3c) look similar than those of ERA-20C relative to observations in the latitude band 35 ° S–5 ° S, implying that 20CR generally has a smaller pressure

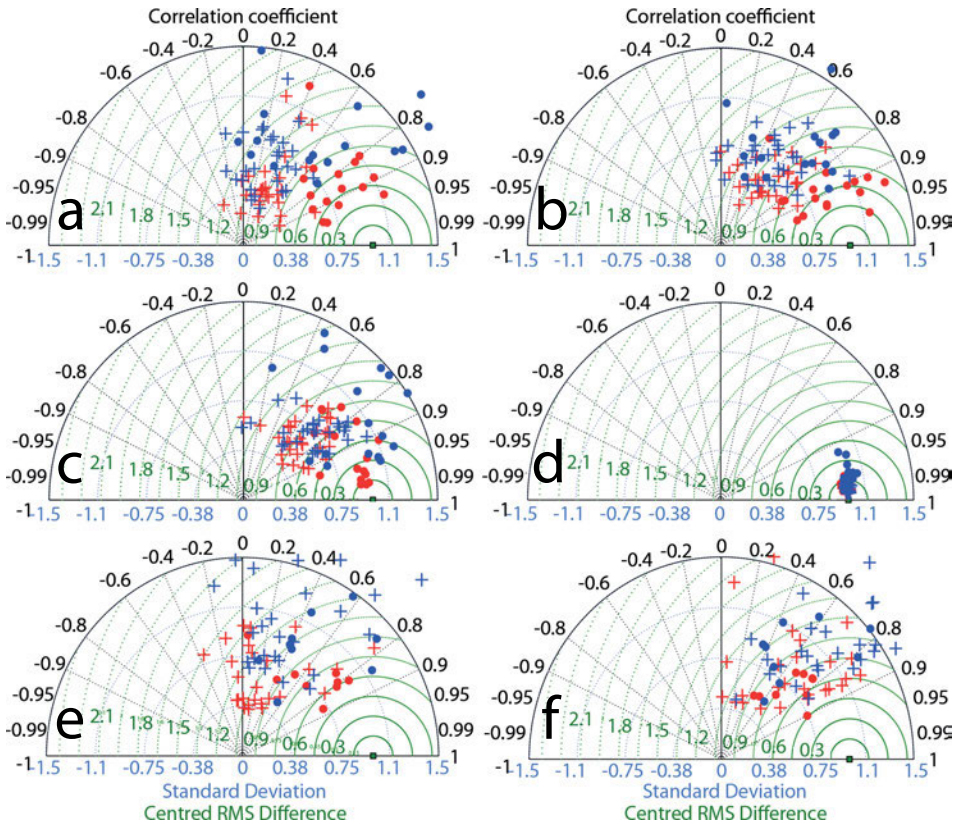

Figure 9: Taylor diagrams for anomalies (left column) and absolute values (right column) of 20CR (red) and ERA-20C (blue) versus *Meteor* observed temperature (upper row), pressure (middle row) and specific humidity (lower row) for each latitude-altitude bin with $N > 15$ observations-reanalysis pairs. Values for extratropical bins are marked by dots, values for tropical bins by crosses. The standard deviations have been standardised by the observed value. Note that red dots and crosses, and blue crosses are present, but not visible in (d), since they are overlain by blue dots.

bias than ERA-20C. Anomaly correlations between ERA-20C and 20CR are generally medium to high (Fig. S3d).

Scatter plots of raw pressure values from the reanalyses against those from the observations show results very close to the 1:1 line for both datasets (Fig. 8 middle row), as expected from the very strong relationship between altitude and atmospheric pressure. The regression of 20CR exhibits a slightly smaller scatter ($r = 1.000$) than that of ERA-20C ($r = 0.999$). Also the intercept is larger for ERA-20C (+2.6 hPa) compared to 20CR (+0.5 hPa). For both reanalyses, the largest differences relative to the observations tend to occur at the highest pressure values, i.e. closer to the surface.

For pressure, the Taylor diagrams (Fig. 9cd) confirm the very high (Pearson's) correlations of the total values for both ERA-20C and particularly 20CR in the extratropics and in the Tropics. Also the ratio of the reanalysed to the observed standard deviation is close to 1 for them. For the anomalies, on the other hand, the ratio covers a larger range from ca. 0.4 to 1.05 in 20CR, and from ca. 0.55 to 1.5 in ERA-20C (values > 1.1 appear only in the extratropics in ERA-20C). In agreement with the results found in Figs. 6 and 7, the (Pearson's) correlation coefficients are consistently high for the extratropics in 20CR (and with the exception of three outliers also for ERA-20C), and generally medium to high in the tropics. The results for the absolute values point to a good to very good representation of the pressure variability on the seasonal time scale in both reanalyses.

3.3 Specific humidity

Comparison of humidity between the reanalyses and observations shows some agreement and perhaps more than might have been expected for such a difficult quantity to measure and simulate. Figs. 6e and 7e show the time and zonal mean difference of specific humidity in 20CR/ERA-20C relative to the kite observations from the *Meteor*. The largest significant positive difference occurs in 20CR (+1.3 g/kg), whereas the largest significant negative difference appears in ERA-20C (−3.1 g/kg). The overall pattern of differences is similar in the Tropics in 20CR and ERA-20C: the reanalyses tend to be wetter than the observations at the southern edge of the Tropics (around 20 °S), and drier further towards the north (although not all differences are highly significant). In the mean over all available bins with $N > 15$, both 20CR and ERA-20C show slightly lower specific humidity values than the observations ($\overline{b_{SH}} = -0.2$ g/kg and $\overline{b_{SH}} = -0.6$ g/kg). Comparing the variability (not shown), it may be surprising

that the standard deviations of the time mean reanalysis-observation differences (rms) calculated over these bins are equal (0.7 g/kg) for both datasets, and are slightly smaller than the estimated 20CR ensemble mean uncertainty (time mean ensemble spread over the same bins is 1.1 g/kg).

Over most of the study region, the ERA-20C humidity anomalies agree better with the observed than 20CR. Figs. 6f and 7f show the zonally averaged Spearman rank correlation coefficients between the 20CR/ERA-20C and *Meteor* specific humidity anomalies. Whereas only 29 out of 37 altitude-latitude bins with $N > 15$ and significant results ($Si = 90\,\%$, z test) display non-negative correlations in the case of 20CR, this number is higher (35) for ERA-20C. In the case of 20CR, this number is smaller than what would be expected by chance at $Si = 90\,\%$ and for "true" positive correlations everywhere. This suggests that "true" (and not only estimated) negative anomaly correlations do indeed exist. For 20CR, the mean correlation coefficient of 0.32 is much lower than in the case of temperature and particularly pressure anomalies. In contrast, for ERA-20C, the mean correlation coefficient of 0.35 is clearly lower than the ERA-20C value found for pressure anomalies, but higher than the one calculated for temperature anomalies. There are no clear spatial patterns in the specific humidity anomaly correlations. Very high and very low correlations appear in both the extratropics as well as in the Tropics. Nevertheless, the general statement that correlations are higher on average south of 25 ° S (20CR: mean 0.62, ERA-20C: mean 0.42) compared to north of that latitude (0.22 vs. 0.31) holds also true for this variable, although the difference is not as marked in the case of ERA-20C.

The systematic differences between ERA-20C and 20CR particularly reveal lower specific humidities in ERA-20C relative to 20CR (and also to observations) in the latitude band 15 ° S–5 ° S (Fig. S3e). The anomaly correlations between the reanalyses are relatively high between 5 ° S and 25 ° N (where they are also higher than between each reanalysis and the observations) and south of 55 ° S (Fig. S3f), but low between 45 ° S and 5 ° S.

Scatter plots of 20CR and ERA-20C against observed specific humidity reveal a slightly larger scatter of the ERA-20C values than the 20CR values around the regression line ($r = 0.91$ vs. $r = 0.93$). Both ERA-20C and 20CR seem to underestimate specific humidity at high observed values (cumulative distribution functions that deviate from the 1:1 line at specific humidities > ca. 13–14 g/kg). In general, the scatter plots show a similar pattern with a strong scatter particularly in the lower medium range of humidities (5–10 g/kg, larger for ERA-20C), and a "saturation" between 15 and 17 g/kg in the reanalyses, likely pointing to an underestimation of the highest humidity values in the tropical lower troposphere.

The Taylor diagram of the binned 20CR and ERA-20C versus observed specific humidity anomalies (Fig. 9e) confirms the clearly higher (Pearson's) anomaly correlations in the extratropics compared to the Tropics (cf. Table 1) for both reanalyses. 20CR shows a consistently lower variance than the observations in both the extratropics and the Tropics on the intraseasonal time scale (standard deviation ratios between 0.4 and 1), whereas ERA-20C exhibits a large spread of the ratio between ca. 0.45 and 1.65. Total values for the Tropics (crosses in Fig. 9f) show much larger correlations compared to the anomalies, suggesting that the annual cycle is well represented in both reanalyses at most altitudes but the fluctuations are not represented as well (with the notable exception of some bins at the highest available levels at 15 ° S–25 ° S and 5 ° S to 25 ° N, possibly above the boundary or trade wind inversion layer). Only three bins have correlations < 0.4 in this case. Also, the possible underestimation of the variance in 20CR relative to the observations, as already suggested by the regression coefficient of 0.87 in Fig. 8, is much less pronounced for the total values, at least in the Tropics.

3.4 Zonal and meridional wind

While there are regions of excellent agreement, overall the reanalysed and observed wind from both the kites and pilot balloons agree moderately well, as with the temperature and specific humidity from the kites. Figs. 10a and 11a show the time and zonal mean difference of the zonal wind in 20CR/ERA-20C relative to the pilot balloon observations from the *Meteor*. As already mentioned in Section 2, the vertical range of the pilot balloon observations extends much higher than that of the kite observations, up to about 20,500 m asl. Despite this, we restrict the discussion for the Tropics to altitudes below 16 km asl, since we do not expect these "surface input" reanalyses to correctly represent the QBO and other stratospheric variations. The largest significant positive difference of the reanalyses relative to the pilot balloon observations occurs between 35 ° S and 45 ° S at 15–16 km asl in 20CR (+19.6 m/s), the largest negative difference between 55 ° S and 65 ° S at 7–8 km asl in ERA-20C (−15.6 m/s). Absolute differences > 10 m/s in the pilot balloon observations are only found at higher altitudes (above 6 km asl). Much smaller values are found in the lower troposphere, in agreement with the kite observations (cf. supplementary Figs. S4 and S5). There, the sign of the significant differences generally agrees for the kite and the pilot balloon observations. Both reanalyses exhibit a similar large-scale, bipolar difference pattern relative to the observations with positive differences at latitudes from 25 ° S to 55 ° S (implying stronger than observed westerlies), and negative differences at higher levels above a slanted plane rising from 10,000 m asl around the Equator towards 13,000–16,000 m asl between 25 ° S and 15 ° S (implying weaker than observed westerlies or even easterlies instead of observed westerlies). Additionally, 20CR shows negative differences at lower levels from 5 ° N to 25 ° N that represent weaker than observed westerlies, whereas ERA-20C shows westerly differences in the upper troposphere in the same latitudinal belt. In the mean

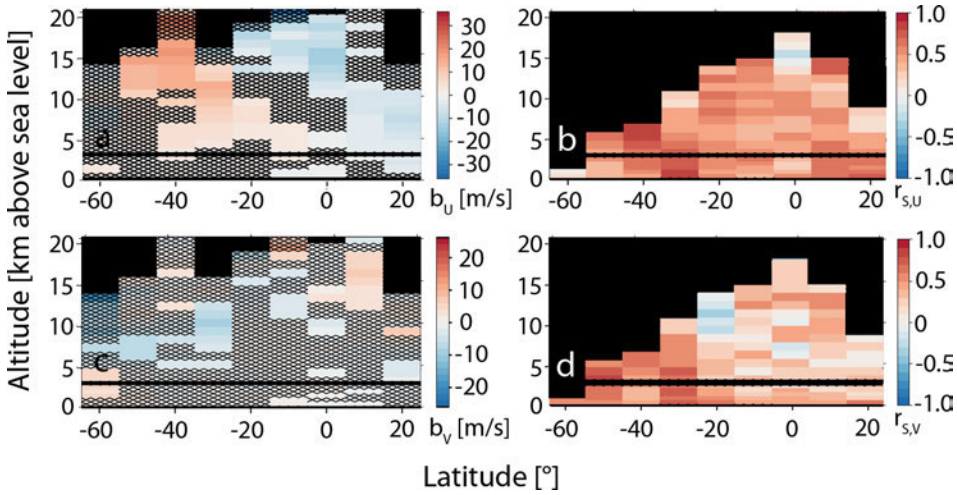

Figure 10: As Fig. 6, but for the zonal mean (a, c) differences b and (b, d) Spearman rank anomaly correlations r_S between 20CR and the *Meteor* pilot balloon observations of (a, b) zonal wind, (c, d) meridional wind. Note the different vertical range compared to Fig. 6. The level between 3,000–3,500 m asl has no observational data (black).

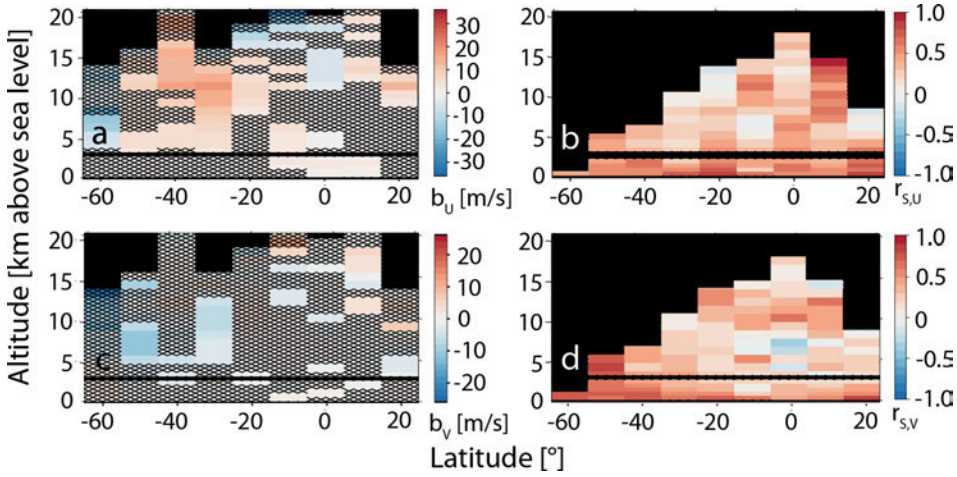

Figure 11: As Fig. 10, but for ERA-20C instead of 20CR.

over all available bins with $N > 15$, 20CR shows almost no difference to observed zonal wind values ($\overline{b_U}$ = +0.2 m/s), while ERA-20C shows more westerly winds ($\overline{b_U}$ = +1.6 m/s). Panels a and c of Fig. S6 show histograms of the zonal wind differences relative to the observations for both 20CR and ERA/20C. The more westerly winds on average than in the observations in ERA-20C are also reflected in the histogram, which shows a much higher frequency of b_U between 0 and 5 m/s than between 0 and −5 m/s, and a much higher frequency between 5 and 10 m/s than between −5 and −10 m/s. The distributions for both 20CR and ERA-20C have higher frequencies on the positive side between 10 and 20 m/s than between −10 and −20 m/s, but at the same time longer tails on the negative side. The standard deviations of the time mean reanalysis-observation differences calculated over the same bins are quite large with 7.0 m/s (20CR) and 6.4 m/s (ERA-20C). These numbers are a bit larger than the expected uncertainty of 20CR when using the time mean of the ensemble spread as an estimate (5.7 m/s), suggesting that the reanalysis uncertainty of the zonal wind might be slightly underestimated here.

Figs. 10b and 11b show the zonally averaged Spearman rank correlation coefficients between the 20CR/ERA-20C and *Meteor* pilot balloon zonal wind anomalies. 111/108 out of 114 altitude-latitude bins with $N > 15$ and significant results ($Si = 90\,\%$, z test) display non-negative correlations. As for specific humidity and temperature, the mean and median correlation coefficients of 0.44/0.34 and 0.51/0.34 are clearly lower than in the case of pressure anomalies, and the correlations for ERA-20C are clearly lower than those for 20CR. The correlations do not show a clear pattern in latitude nor altitude. Whereas for 20CR values in the extratropics tend to be higher than in the Tropics (20CR: mean 0.44, median 0.58, vs. mean 0.43, median 0.50), there is no such tendency in the case of ERA-20C (mean 0.28, median 0.34 vs. mean 0.35, median 0.32). Interestingly, the average ERA-20C correlations for the second lowest level (500–1,000 m asl) and the lowest 3 km asl are higher

than the 20CR correlations (0.63 vs. 0.51 and 0.50 vs. 0.48), as might be expected from the additional assimilation of surface marine winds in that reanalysis (see also Table 1). However, the positive effect does not seem to extend much into the mid to upper troposphere/lower stratosphere.

Fig. 10c and 11c show the time and zonal mean difference of the meridional wind in 20CR/ERA-20C relative to the pilot balloon observations from the *Meteor*. In both reanalyses, the largest significant positive and negative differences are about +8 m/s and −10 m/s and occur above the lower troposphere. Significant differences occur scattered over all latitudes and over different altitudes. Although the differences are not always highly significant at the bin level, the spatial structure is somewhat organised and similar with negative differences in the mid to upper troposphere from 65 ° S to 5 ° N and positive differences at 5 ° N–25 ° N above 9–12 km asl. In the mean over all available bins with $N > 15$, both 20CR and ERA-20C show slightly more northerly meridional winds ($\overline{b_V} = -0.4$ m/s and $\overline{b_V} = -0.8$ m/s) relative to the pilot balloon observations. Panels b and d of Fig. S6 show histograms of the meridional wind differences b_V relative to the observations. In both cases, the differences are virtually normally distributed, with a tendency for slightly more negative values particularly in the tails of the distributions. The standard deviations of the time mean reanalysis-observation differences calculated over the same bins are again relatively large and similar with 3.0 m/s and 3.4 m/s, and are smaller than the estimated 20CR uncertainty for the same bins (time average of the ensemble spread of 5.1 m/s).

Finally, Figs. 10d and 11d show the zonally averaged Spearman rank correlation coefficients between the 20CR/ERA-20C and *Meteor* pilot balloon meridional wind anomalies. 106/105 out of 114 altitude-latitude bins with $N > 15$ and significant results ($Si = 90\,\%$, z test) display positive correlations. As for specific humidity, temperature, and zonal wind, the mean correlation coefficients of 0.37/0.31 are clearly lower than in the case of pressure anomalies, and the correlation is slightly lower in the case of ERA-20C. The correlations show a strong dependence on the latitudinal range with higher values on average south of 25 ° S (20CR: mean 0.52, median 0.63 vs. mean 0.27, median 0.26; ERA-20C: mean 0.48, median 0.63 vs. mean 0.24, median 0.26). Again, the average ERA-20C correlations for the second lowest level (500–1,000 m asl) and the lowest 3 km asl are higher than the 20CR correlations (0.67 vs. 0.55 and 0.49 vs. 0.45), suggesting a positive impact of the additionally assimilated observational data for this variable (see also Table 1).

As with the bias statistics, for the lower troposphere ($< 4,500$ m asl), similar correlation results are generally found for both reanalyses when analysing zonal or meridional wind data from the kite observations instead of the (more abundant) pilot balloon observations (see supplementary Figs. S4 and S5).

The spatial structure of the zonal wind differences of ERA-20C relative to 20CR reveals mainly more westerly winds in ERA-20C compared to 20CR in the upper troposphere from 35 ° S–5 ° N in the mid to upper troposphere from 5 ° N–25 ° N, and generally in the tropical lower troposphere (see Fig. S7a in the Supplement). The zonal wind anomaly correlations of ERA-20C and 20CR tend to be medium to high everywhere except in the mid to upper troposphere between 25 ° S and 5 ° S, a region for which the anomaly correlation of ERA-20C with the observations was also found to be low (cf. Figs. S5b and 11b).

Scatter plots of 20CR and ERA-20C against observed raw zonal wind show a higher correlation for 20CR than for ERA-20C ($r = 0.75$ vs. $r = 0.68$, Fig. 12 upper row). Both the 20CR ensemble mean and ERA-20C tend to underestimate the observed range of zonal wind speeds (regression slopes of 0.73 and 0.77). ERA-20C shows larger differences relative to the observations than 20CR (> 25 m/s), particularly on the positive side of the regression line and for observed westerly velocities of 10 to 35 m/s.

The scatter plots for meridional wind show a higher, albeit moderate correlation for 20CR ($r = 0.44$), and a very weak correlation ($r = 0.32$) for ERA-20C (Fig. 12 lower row). Both reanalyses, particularly 20CR (as can be seen from the large deviation of the cumulative distribution function from the 1:1 line), strongly underestimate the observed variability of the meridional wind (slopes of only 0.26 and 0.28 in the regression).

The Taylor diagrams for the anomalies and for the total values of the zonal wind show a tendency for higher correlations for 20CR compared to ERA-20C (Fig. 13ab), consistent with the Spearman rank anomaly correlations in Figs. 10 and 11. Similar to the scatter plots for the entire range of reanalysed vs. observed values (Fig. 12), the Taylor diagrams suggest a reduced variance of zonal wind also on the latitude-altitude bin level in both reanalyses, and particularly in 20CR, compared to the observations. In the case of 20CR, almost all bins show a relative standard deviation between 0.4 and 1, in both the total value and the anomalies. Again in agreement with Figs. 10 and 11, no clear distinction is visible between the behaviour in the extratropics and in the Tropics, except that tropical bins tend to show lower relative variances than extratropical bins. For zonal wind, there is a slight increase of the correlation when using absolute values instead of anomalies for some tropical bins. Nevertheless, the results point to generally similar deficiencies in the representation of the annual cycle (absolute values) compared to the higher frequency variability.

The Taylor diagrams for the meridional wind reveal a tendency for higher correlation coefficients in the extratropics compared to the Tropics for both reanalyses (Fig. 13cd), as already suggested by Table 1 and Figs. 10 and 11. The lower relative variances found for the entire range of observed values (Fig. 12) are again additionally detected on the latitude-altitude bin level for total as

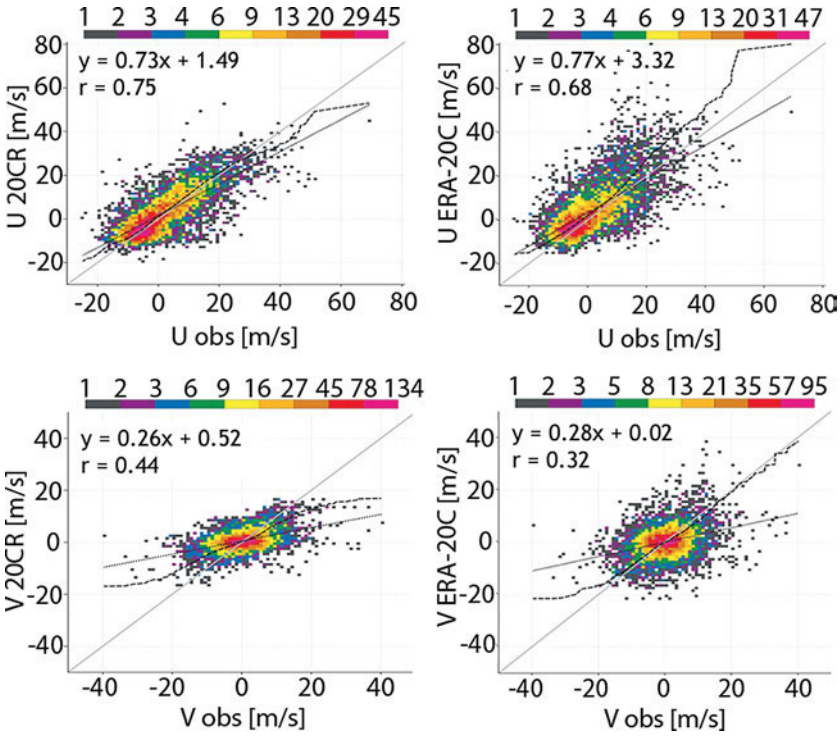

Figure 12: As Fig. 8, but for 20CR (left panels) and ERA-20C (right panels) versus *Meteor* observed zonal wind (upper row) and meridional wind (lower row) from pilot balloons. The bins widths are ca. 1 m/s in each direction. Also shown are the 1:1 lines (grey), the linear regression lines (dotted) and cumulative distribution functions (dashed) of the reanalysis data versus the observed values and the Pearson's correlation coefficients.

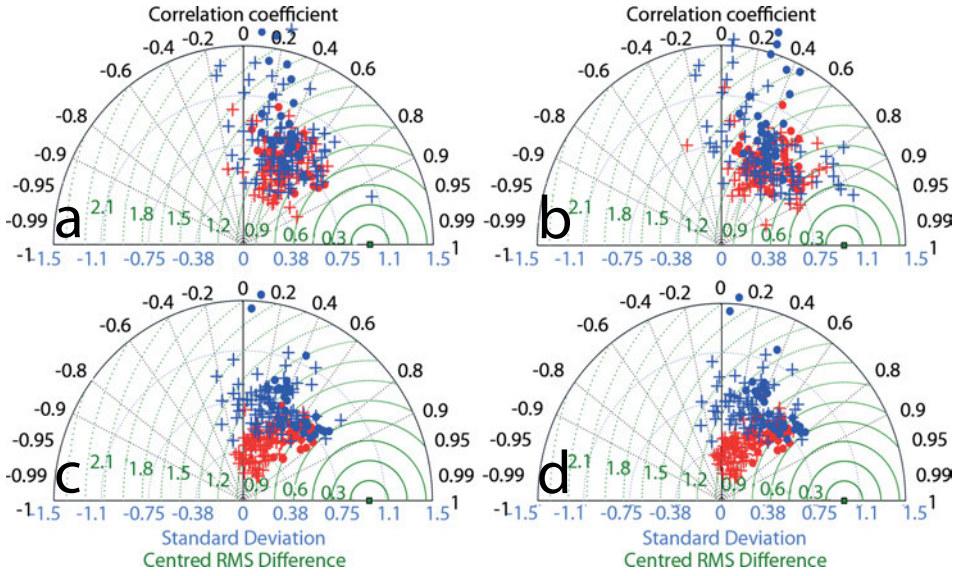

Figure 13: As Fig. 9, but for zonal wind (upper row) and meridional wind (lower row).

well as for the anomaly values, and for both reanalyses. Also similar to the result found in Fig. 12, 20CR shows even lower relative standard deviations (range ca. 0.2 to 0.8 for anomalies) compared to ERA-20C (range ca. 0.4 to 1.55). As for zonal wind, there is a clear tendency for lower relative variances in the Tropics compared to the extratropics. Again, the results point to similar deficiencies in the representation of the annual cycle (absolute values) compared to the higher frequency variability.

3.5 Significance of differences including measurement uncertainty

The significance test applied above to the mean differences between the observations and the 20CR/ERA-20C did not yet take into account any intrinsic observation errors. By adding such an estimate to the bias analysis, we can assess whether the observational uncertainty can account for the found biases, or whether the reanalyses are

Table 2: Estimates of systematic (random) errors of single historical upper-air observations ($n = 1$) from different studies. *Values derived from total accuracy σ_{total} and random error σ_{random} as $\sigma_{\text{systematic}} = \sqrt{(\sigma^2_{\text{total}} - \sigma^2_{\text{random}})}$, assuming independence of systematic and random errors. #Values chosen for attributing detected, significant differences between observations and reanalyses to deficiencies of the latter.

Reference	Wind speed	Temperature	Pressure	Specific humidity
Reichs-Marine-Amt (1909)		2 K*# (0.2 K)	2.3 hPa*# (1.33 hPa)	
Jasperson (1982)	< 0.5 m/s			
Brönnimann et al. (2011)		(1.2 K)	(1.35 hPa#)	
Sun et al. (2013)		≤ 2 K@$p \geq$ 30 hPa		
Wartenburger et al. (2013)	(3 m/s#)	(1.5 K#)	(1.3 hPa)	
Ladstädter et al. (2015)				< 20 %# (< 50 %#)

indeed not in agreement with the observations including their uncertainties.

A number of studies give estimates for the uncertainty of single, historical upper-air observations (see Table 2). In our context, it appears most reasonable to subdivide the uncertainty into a systematic part, i.e. the part of the error that does not reduce when averaging over a larger number of observations, and a random part that is reduced as $\frac{1}{\sqrt{n}}$ with the number of observations n (see Table 2). Then, significant differences (according to the simple t test) that are larger than two times the total accuracy of the observations can be attributed to the reanalysis with a confidence of 95 %, assuming that the measurement accuracy is normally distributed. The total accuracy in each latitude-altitude bin can be calculated as $\sigma_{\text{total}} = \sqrt{\left(\sigma^2_{\text{systematic}} + \frac{\sigma^2_{\text{random}}}{n}\right)}$ with the number n of observation-reanalysis pairs, assuming statistical independence of the systematic and random part of the total error. As a conservative estimate, we have applied the larger values marked with a hash in Table 2 for the calculation of σ_{total}. For wind speed, the systematic error of the measurements (upper limit of 0.5 m/s given in Jasperson (1982), see Table 2) is negligible compared to the random error for $n < 5$, and generally compared to the magnitude of the detected differences (Figs. 10ac and 11ac). Ladstädter et al. (2015) give estimates of systematic and random differences between specific humidity from standard, modern equipment radiosondes (Vaisälä RS90/92) and from the Global Reference Upper-Air Network (GRUAN) compared to GPSRO observations. They find systematic differences of up to 15–20 % and random differences in the order of 50 % for RS90/92 radiosondes in the lower to mid troposphere relative to GPSRO, whereas the agreement between RS90/92 and GRUAN is much better. Here, we take these rather conservative values as upper limits for the uncertainty assessment.

Under these assumptions and for 20CR, only the negative temperature differences at 30 ° S above 3,000 m asl are outside $2 \times \sigma_{\text{total}}$ (cf. Fig. 6a). Likewise, for pressure, only the positive difference at 50 ° S/2,500–3,000 m asl is outside $2 \times \sigma_{\text{total}}$ (cf. Fig. 6c). And for specific humidity, only the positive, significant difference at 15 ° S–25 ° S / 3,000–3,500 m asl is outside $2 \times \sigma_{\text{total}}$ (cf. Fig. 6e). For zonal and meridional wind in contrast, the significant differences identified in Fig. 10ac are without exception outside the estimated $2 \times \sigma_{\text{total}}$. Observational errors might be larger than at fixed land stations in this case due to the movement (translational, rolling) of the ship that would have certainly complicated the tracking of the balloons/kites.

For ERA-20C, the largest significantly negative temperature differences (< −4 K) above 2,500 m asl south of 55 ° S and at 35 ° S–5 ° S are outside $2 \times \sigma_{\text{total}}$ (cf. Fig. 7a). With respect to pressure, only the strongly negative difference at 3,500–4,000 m asl south of 55 ° S is larger than $2 \times \sigma_{\text{total}}$ (cf. Fig. 7c). For specific humidity, the significant differences are outside the 95 % confidence interval of the observations south of 45 ° S, and at 5 ° S–15 ° S above 3,000 m asl (cf. Fig. 7e). Finally, in the case of zonal and meridional winds all differences significant according to the t test (cf. Fig. 11ac) are larger than $2 \times \sigma_{\text{total}}$.

4 Conclusions and outlook

This paper compares a recently digitised, rich source of historical upper-air data for the tropical and South Atlantic during the mid-1920s (ca. 1,000 profiles), obtained on board the research vessel *Meteor* during the *German Atlantic Expedition*, with two totally independent reanalyses (20CR, ERA-20C). The value of the newly available atmospheric profiles lies in their potential to be used in future reanalysis projects, to better constrain the past state of the atmosphere in an otherwise sparsely observed region of the globe. The reanalyses themselves differ only slightly in their synoptic input data (besides updated datasets, mainly additional marine winds in ERA-20C), while they differ more strongly with respect to the boundary conditions, atmospheric models, and assimilation techniques.

Our study has analysed the systematic differences and anomaly correlations between temperature, pressure, humidity, and wind in the historical observations and in the reanalyses. We find an overall good agreement between the historical observation data and the reanalyses. Very few of the anomaly correlations are negative. Many of the biases are small. Nevertheless, the comparison has identified some significant differences between

Table 3: Mean bias $[\overline{b_x}]$ and standard deviation of the differences $\sigma(\overline{b_x})$ between the 20CR/ERA-20C reanalysis and the *Meteor* observations of different variables X during the *German Atlantic Expedition* 1925–27. The values given are calculated as average over all latitude-altitude bins with $N > 15$. *Values for kite observations up to 4,500 m asl. #Values for pilot balloon observations up to 21.5 km asl.

	[timeavg(b_X)]		σ(timeavg(b_X))	
	20CR	ERA-20C	20CR	ERA-20C
Temperature*	−0.30 K	−0.01 K	1.3 K	1.3 K
Pressure*	+0.05 hPa	+0.21 hPa	0.7 hPa	1.0 hPa
Specific Humidity*	−0.2 g/kg	−0.6 g/kg	0.7 g/kg	0.7 g/kg
Zonal Wind*	+0.3 m/s	+1.8 m/s	1.7 m/s	2.0 m/s
Meridional Wind*	+0.4 m/s	+0.4 m/s	2.1 m/s	1.9 m/s
Zonal Wind#	+0.2 m/s	+1.6 m/s	7.0 m/s	6.4 m/s
Meridional Wind#	−0.4 m/s	−0.7 m/s	3.0 m/s	3.4 m/s

the reanalyses and the observations, and we find clearly different strengths of anomaly correlations depending on the parameter and the considered region (extratropics and Tropics). The main findings of the study can be summarised as follows:

1. Systematic and significant ($Si = 99\,\%$) differences of the reanalyses relative to the observations outside the estimated 95 % observational confidence interval:

   (a) Temperature: In 20CR, temperatures are lower than in the observations above 3,000 m asl between 25 ° S and 35 ° S (in the order of −5 K). In ERA-20C, temperatures are lower than in the observations above 3,500–4,000 m asl south of 55 ° S and between 35 ° S–15 ° S, and at 15 ° S–5 ° S/2,500–3,500 m asl (order −5 K). These large negative differences cannot be explained by radiation errors of the measurements, since the latter are expected to be much smaller in the lower to mid troposphere (typical values of < 0.3 K at pressure values > 300 hPa, see e.g. Dirksen et al., 2014).

   (b) Pressure: Differences are mostly inside the observational 95 % uncertainty range (ERA-20C, 20CR).

   (c) Specific humidity: In 20CR, the southern edge of the Tropics (15 ° S–25 ° S, outside observational uncertainty at 3,000–3,500 m asl) is wetter than in the observations (order +1 to +2 g/kg). In ERA-20C, the latitudinal band 15 ° S–5 ° S (outside observational uncertainty above 3,000 m asl) is much drier than in the observations (order −2 to −4 g/kg), and also tends to be drier south of 45 ° S (order up to 1 g/kg).

   (d) Zonal wind: Both 20CR and ERA-20C show stronger than observed westerlies between 25 ° S and 55 ° S (order +10 to +20 m/s). They show weaker than observed westerlies or even easterlies instead of observed westerlies at higher levels between 10,000 and 16,000 m asl between 15 ° S and 5 ° N (order −5 to −10 m/s, only around the equator in the case of ERA-20C). In the northern hemisphere Tropics, 20CR has weaker than observed mid to upper tropospheric westerlies from 5 ° N to 25 ° N (order −5 to −10 m/s), whereas ERA-20C shows stronger than observed upper tropospheric westerlies in the same latitudinal band (order +5 to +15 m/s).

   (e) Meridional wind: Both 20CR and ERA-20C show much more neutral meridional wind conditions in the mid to upper troposphere around 50 ° S and 30 ° S than the observations, which clearly display southerly winds of the order of 10 m/s on average in that region.

2. The standard deviations of the time mean reanalysis-observation difference calculated over the bins with > 15 observation-reanalysis pairs are generally very similar for ERA-20C and 20CR. The time mean systematic differences, averaged over the same bins, are slightly smaller in 20CR for pressure, specific humidity and zonal wind, and slightly smaller in turn in ERA-20C for temperature (Table 3).

3. The estimated uncertainties of the 20CR ensemble mean are very similar to the standard deviations of the time mean reanalysis-observation differences for temperature, pressure, specific humidity, and meridional wind, suggesting that the respective 20CR uncertainty estimates are conservative, and that the effective observational errors might be smaller than the conservative literature values given in Table 2. For zonal wind, the estimated 20CR ensemble mean uncertainty is only a little smaller than what would be expected from a conservatively estimated error of 3 m/s (see Table 2) for observed wind speed. The really good agreement between the rms statistics and the 20CR uncertainty conversely implies that the 20CR ensemble appears to be reliable compared to these independent observational data taken from many different climate regimes.

4. Spearman rank anomaly correlations ($Si = 90\,\%$) between the reanalyses and observations (Table 1, time and spatial averages over the latitudinal ranges 65 ° S–25 ° S and 25 ° S–25 ° N):

(a) Generally, 20CR and ERA-20C show comparable anomaly correlations for all parameters, with the highest values found for pressure and temperature, when accounting for using an ensemble mean instead of a single model realisation. The ensemble mean of 20CR tends to show slightly higher anomaly correlations than ERA-20C.

(b) For 20CR, moderate to low average anomaly correlations (< 0.5) with all variables except pressure are found in most of the Tropics (25 ° S–25 ° N), whereas the average anomaly correlation coefficients are high (temperature, pressure, > 0.75), medium high (specific humidity, meridional wind below 4.5 km asl from kites, > 0.5), moderate (meridional wind up to 20.5 km asl from pilot balloons), and low (zonal wind up to 20.5 km asl from pilot balloons) in the southern extratropics (25 ° S–65 ° S), depending on the variable considered.

(c) For ERA-20C, low anomaly correlations except for pressure and zonal wind below 4.5 km asl from kites are found on average in the Tropics, whereas anomaly correlations are on average high (pressure, meridional wind below 4.5 km asl from kites), moderate (temperature, meridional wind up to 20.5 km asl from pilot balloons), or low (specific humidity, zonal wind up to 20.5 km asl from pilot balloons) in the southern extratropics.

(d) The only observed variables for which on average more than 50 % of the high frequency variability are represented in the reanalyses are southern extratropical pressure (20CR and ERA-20C) and temperature (20CR).

(e) A positive impact of the additionally assimilated surface marine winds can be detected in ERA-20C compared to 20CR wind fields for the lower troposphere up to about 3 km asl (see Section 3.4).

5. Ratio of standard deviations and seasonal variability in the reanalyses relative to the observations (Taylor diagrams Figs. 9 and 13):

(a) 20CR specific humidity shows a consistently lower variance than the observations in both the extratropics and the Tropics on the intraseasonal time scale (standard deviation ratios between 0.4 and 1 for the anomalies, see Fig. 9e).

(b) Both reanalyses, and particularly 20CR, have a reduced variance of zonal, and particularly meridional wind on the latitude-altitude bin level compared to the observations (relative standard deviations of the anomalies between 0.4 and 1 for zonal and between 0.2 and 0.8 for meridional wind for 20CR), with the lowest values generally found in the Tropics.

(c) The seasonal cycle of temperature in the reanalyses correlates only slightly better with the observations than the higher frequency anomalies. For pressure and specific humidity, we find clearly higher correlations on the seasonal compared to the higher frequency timescale. Finally, for zonal and meridional wind, the correlations on the seasonal timescale are similar (low to moderate) than on the intraseasonal timescale.

Our findings are in line with the results of previous studies that have found a lower quality of 20CR in the Tropics, but relatively high anomaly correlations in the southern extratropics (Compo et al., 2011; Stickler and Brönnimann, 2011; Brönnimann et al., 2011; Brönnimann and Stickler, 2013; Zhang et al., 2013). However, more variables are considered here, particularly wind and humidity. Furthermore, the analysed period is quite early and the data have an exceptionally good coverage for a data sparse region such as the tropical and South Atlantic. Regarding wind, all results obtained are robust with respect to the choice of the observational platform (kite or pilot balloon).

The present study is one of the first to compare ERA-20C with historical observations in the Tropics and in the southern hemisphere. The new ERA-20C reanalysis performs less well than the 20CR 56 member ensemble mean in the representation of the short term variability of temperature, specific humidity, and zonal wind in the southern extratropics, and of meridional wind in the Tropics. On the other hand, it performs better than 20CR in the representation of the synoptic variability of specific humidity in the Tropics, and its zonal and meridional wind variances are closer to those found in the observations. Still, in this comparison at least, the assimilation of marine winds does not seem to have markedly improved the quality of ERA-20C over that of 20CR, with the noticeable exception of lower to mid tropospheric winds. Since the rest of the assimilated data are quite similar, we suggest that the differences in the used models (model formulation or parametrisations) or in the assimilation techniques (Ensemble Kalman Filter vs. 4D-Var) are the most likely candidates to have caused the differences in the result of the reanalyses. However, a closer examination of the exact causes of the found reanalysis-observation differences is outside the scope of this paper.

The differences and generally much lower anomaly correlations in the Tropics (even at low levels) are in agreement with 1) a generally very small number of observations (as in the southern extratropics), 2) a weaker physical constraint of atmospheric circulation by surface pressure variations due to the small horizontal component of the Coriolis force and 3) the much smaller variance of surface pressure (observations and reanalyses) in the Tropics compared to the extratropics. Point 2) is also supported by the facts that a) zonal and particularly meridional wind are the parameters for which the highest and lowest observed values tend to be underestimated

by the reanalyses (see Fig. 12), and that b) Taylor diagrams show a smaller wind variance in the reanalyses compared to the observations, particularly in the Tropics (see Figs. 9 and 13). The detected increase in differences (and partly decreasing correlation) with altitude in the extratropics (particularly for temperature, zonal and meridional wind) are consistent with a decreasing quality of the reanalyses and decreasing number of observation / increasing observational error towards upper levels.

The generally relatively low correlations found for specific humidity (except in the extratropics for the 20CR ensemble mean) might be caused by deficiencies either in the circulation or moisture sources and sinks (evaporation fluxes and precipitation) of the reanalyses, or in errors of the measurements themselves. Humidity measurements from radiosondes are nowadays still considered less reliable than pressure and temperature observations, particularly at low absolute values in the upper troposphere and lower stratosphere. Even though the historical humidity measurements here are only from the moister, lower troposphere, they may be flawed by uncertainty. On the other hand, the root mean square anomaly of the observations, i.e. the standard deviation of the mean observation-climatology difference, is not much larger than the root mean square anomaly in 20CR/ERA-20C (1.0 g/kg vs. 0.6 g/kg / 0.9 g/kg for all bins with significant differences, compared to absolute observed values between 1.3 g/kg and 15.0 g/kg, mean over all significant bins: 6.3 g/kg). These numbers suggest that the observed values realistically follow the dynamic range of the natural specific humidity distribution in the covered atmospheric region (25 ° N–65 ° S, 0–4,500 m asl), and that it might rather be a deficiency of the reanalyses or atmospheric models therein that is responsible for the detected differences. Furthermore, the statistically significant differences are mostly outside the estimated $2 \times \sigma_{\text{total}}$ error of the observations in the significant bins, at least up to relative observation errors of 15–20 %. Finally, differences of both signs do exist which cannot be explained by rather constant measurement biases. However, even though such errors seem large at a first glance, they cannot be totally excluded in the historical data.

All upper-air data digitised in ERA-CLIM will be quality controlled in more detail and included in the Comprehensive Historical Upper-Air Network (CHUAN; Stickler et al., 2010). This observation database will serve future reanalysis projects at ECMWF and other centres. The digitisation of the catalogued, historical data is being continued in the framework of ERA-CLIM2, the follow-up project to ERA-CLIM, which started in January 2014.

Acknowledgments

AS was funded by the EU FP7 projects ERA-CLIM (Grant 265229) and ERA-CLIM2 (Grant 607029). The digitisation of the historical *Meteor* upper-air data was financed through ERA-CLIM. SB and RW received funding from the SNF project EVALUATE (SNF 200021-130407). The 20CR data are from the NOAA ESRL website. The ERA-20C data are available from the ECMWF website. GPC is supported by the US Department of Energy Office of Science (BER) and the NOAA Climate Program Office.

References

Brönnimann, S., A. Stickler, 2013: Aerological observations in the Tropics in the Early Twentieth Century. – Meteorol. Z. **22**, 349–358. DOI:10.1127/0941-2948/2013/0458.

Brönnimann, S., G.P. Compo, R. Spadin, R. Allan, W. Adam, 2011: Early ship-based upper-air data and comparison with the Twentieth Century Reanalysis. – Clim. Past **7**, 265–276.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. – Bull. Amer. Meteor. Soc. **87**, 175–190.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M.C. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, S.J. Worley, 2011: The Twentieth Century Reanalysis Project. – Quart. J. Roy. Meteor. Soc. **137**, 1–28.

Dee, D., 2004: Variational bias correction of radiance data in the ECMWF system. – Proc. of the ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP, 97–112, ECMWF, Reading.

Dee, D.P., S.M. Uppala, A.J. 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. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.B. Monge-Sanz, J.J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.N. Thépaut, F. Vitart, 2011: 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.

Dirksen, R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data. – Atmos. Meas. Tech. **7**, 4463-4490. DOI:10.5194/amt-7-4463-2014. processing for the Vaisala RS92 radiosonde

ECMWF, 2013: IFS documentation CY38R1. – Available at: http://old.ecmwf.int/research/ifsdocs/CY38r1/.

Gillett, N.P., F.W. Zwiers, A.J. Weaver, P.A. Stott, 2003: Detection of human influence on sea-level pressure. – *Nature* **422**, 292–294.

Heiskanen, W.A., H. Moritz, 1967: Physical Geodesy. – Freeman, San Francisco.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, D. Dee, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. – Quart. J. Roy. Meteor. Soc. **140**, published online, DOI:10.1002/qj.2528.

Hoheisel-Huxmann, R., 2007: Die Deutsche Atlantische Expedition 1925–1927: Planung und Verlauf. –Convent, Hamburg, 125 pp.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. – ECMWF Technical Memoranda **636**, available at: http://old.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm636.pdf.

Jasperson, W.H., 1982: The limiting accuracy of wind profiles obtained by tracking rising balloons. – J. Appl. Meteor. **21**, 816–822.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, M. Fiorino, 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. – Bull. Amer. Meteor. Soc. **82**, 247–268.

Knapp, K.R., M.C. Kruk, D.H. Levinson, H.J. Diamond, C.J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. – Bull. Amer. Meteor. Soc. **91**, 363–376. DOI:10.1175/2009BAMS2755.1.

Kuhlbrodt, E., J. Reger, 1933: Die Aerologischen Methoden und das Aerologische Beobachtungsmaterial. – In: Defant, A. (Ed.): Deutsche Atlantische Expedition auf dem Forschungsschiff *„Meteor“*. – Wissenschaftliche Ergebnisse, Vol. XV. de Gruyter, 1–305.

Ladstädter, F., A.K. Steiner, M. Schwärz, G. Kirchengast, 2015: Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002 to 2013. – Atmos. Meas. Tech. **8**, 1819–1834. DOI:10.5194/amt-8-1819-2015.

Moorthi, S., H.-L. Pan, P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. – Technical Procedures Bulletin **484**, NOAA, NWS: Silver Spring, MD, available at: http://www.nws.noaa.gov/om/tpb/484.htm

Poli, P., H. Hersbach, D. Tan, D. Dee, J.-N. Thépaut, A. Simmons, C. Peubey, P. Laloyaux, T. Komori, P. Berrisford, R. Dragani, Y. Trémolet, E. Holm, M. Bonavita, L. Isaksen, M. Fisher, 2013: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). – ERA Report Series **14**, available at: http://old.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_14.pdf

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. – J. Geophys. Res. **108**, DOI:10.1029/2002JD002670.

Reichs-Marine-Amt, 1909: Forschungsreise S.M.S. “Planet” 1906/07. II. Band: Aerologie (contributed by Hergesell, Köppen, Kurtz, Lübbert, Maurer, Schlenska, Schweppe). – Verlag Karl von Siegismund, Berlin.

Saha, S., S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. Van Den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R.W. Reynolds, G. Rutledge, M. Goldberg, 2010: The NCEP climate forecast system reanalysis. – Bull. Amer. Meteorol. Soc. **91**, 1015–1057, DOI:10.1175/2010BAMS3001.1.

Schönwiese, C.-D., 1992: Praktische Statistik für Meteorologen und Geowissenschaftler, 2. Auflage. – Gebrüder Borntraeger, Berlin and Stuttgart, 231 pp.

Spiess, F., 1928: Die Meteor-Fahrt. Forschungen und Erlebnisse der Deutschen Atlantischen Expedition 1925–1927. – Reimer, Berlin.

Stickler, A., S. Brönnimann, 2011: Significant bias of the NCEP/NCAR and twentieth-century reanalyses relative to pilot balloon observations over the West African Monsoon region (1940–57). – Quart. J. Roy. Meteor. Soc. **137**, 1400–1416.

Stickler, A., A.N. Grant, T. Ewen, T.F. Ross, R.S. Vose, J. Comeaux, P. Bessemoulin, K. Jylhä, W.K. Adam, P. Jeannet, A. Nagurny, A.M. Sterin, R. Allan, G.P. Compo, T. Griesser, S. Brönnimann, 2010: The Comprehensive Historical Upper Air Network. – Bull. Am. Meteor. Soc. **91**, 741–751. DOI:10.1175/2009BAMS2852.1.

Stickler, A., S. Brönnimann, M.A. Valente, J. Bethke, A. Sterin, S. Jourdain, E. Roucaute, M.V. Vasquez, D.A. Reyes, R. Allan, D. Dee, 2014a: ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses. – Bull. Amer. Meteor. Soc., published online. DOI:10.1175/BAMS-D-13-00147.1.

Stickler, A., S. Brönnimann, S. Jourdain, E. Roucaute, A. Sterin, D. Nikolaev, M.A. Valente, R. Wartenburger, H. Hersbach, L. Ramella-Pralungo, D. Dee, 2014b: Description of the ERA-CLIM historical upper-air dataset. – Earth Syst. Sci. Data **6**, 29–48. DOI:10.5194/essd-6-29-2014.

Sun, B., A. Reale, S. Schroeder, D.J. Seidel, B. Ballish, 2013: Toward improved corrections for radiation-induced biases in radiosonde temperature observations. – J. Geophys. Res. **118**. DOI:10.1002/jgrd.50369.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. – Mon. Wea. Rev. **117**, 1779–1800.

Wartenburger, R., S. Brönnimann, A. Stickler, 2013: Observation errors in early historical upper-air observations. – J. Geophys. Res. Atmos. **118**, 12,012–12,028, DOI:10.1002/2013JD020156.

WCRP, 2011: Coupled Model Intercomparison Project –Phase 5. – Special Issue of the CLIVAR Exchanges Newsletter No. 56, Vol. 15, No. 2, International CLIVAR Project Office, Southampton.

Whitaker, J.S., T. Hamill, 2002: Ensemble data assimilation without perturbed observations. – Mon. Wea. Rev. **130**, 1913–1924.

Whitaker, J.S., G.P. Compo, X. Wei, T.M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. – Mon. Wea. Rev. **132**, 1190–1200.

Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith, C. Wilkinson, 2010: ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. – Int. J. Climatol. **31**, 951–967. DOI:10.1002/joc.2103.

Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker, N. Lott, 2005: ICOADS release 2.1 data and products. – Int. J. Climatol. **25**, 823–842. DOI:10.1002/joc.1166.

Zhang, Q., H. Körnich, K. Holmgren, 2013: How well do reanalyses represent the southern African precipitation? – Climate Dyn. **40**, 951–962. DOI:10.1007/s00382-012-1423-z.

The pdf version (Adobe Java Script must be enabled) of this paper includes an electronic supplement:
Table of content – Electronic Supplementary Material (ESM)

Figures S1, S2, S3, S4, S5, S6, S7