**Climate
of the Past**

Early ship-based upper-air data and comparison with the Twentieth Century Reanalysis

S. Brönnimann[1,2], G. P. Compo[3,4], R. Spadin[1], R. Allan[5], and W. Adam[6]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland
[2]Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Switzerland
[3]University of Colorado, CIRES, Climate Diagnostics Center, Boulder, CO, USA
[4]NOAA Earth System Research Laboratory, Physical Sciences Division, Boulder, CO, USA
[5]ACRE, Met Office Hadley Centre, UK
[6]DWD, Meteorological Observatory Lindenberg, Germany

Abstract. Extension of 3-D atmospheric data products back into the past is desirable for a wide range of applications. Historical upper-air data are important in this endeavour, particularly in the maritime regions of the tropics and the southern hemisphere, where observations are extremely sparse. Here we present newly digitized and re-evaluated early ship-based upper-air data from two cruises: (1) kite and registering balloon profiles from onboard the ship SMS *Planet* on a cruise from Europe around South Africa and across the Indian Ocean to the western Pacific in 1906/1907, and (2) ship-based radiosonde data from onboard the MS *Schwabenland* on a cruise from Europe across the Atlantic to Antarctica and back in 1938/1939. We describe the data and provide estimations of the errors. We compare the data with a recent reanalysis (the Twentieth Century Reanalysis Project, 20CR, Compo et al., 2011) that provides global 3-D data back to the 19th century based on an assimilation of surface pressure data only (plus monthly mean sea-surface temperatures). In cruise (1), the agreement is generally good, but large temperature differences appear during a period with a strong inversion. In cruise (2), after a subset of the data are corrected, close agreement between observations and 20CR is found for geopotential height (GPH) and temperature notwithstanding a likely cold bias of 20CR at the tropopause level. Results are considerably worse for relative humidity, which was reportedly inaccurately measured. Note that comparing 20CR, which has limited skill in the tropical regions, with measurements from ships in remote regions made under sometimes
difficult conditions can be considered a worst case assessment. In view of that fact, the anomaly correlations for temperature of 0.3–0.6 in the lower troposphere in cruise (1) and of 0.5–0.7 for tropospheric temperature and GPH in cruise (2) are considered as promising results. Moreover, they are consistent with the error estimations. The results suggest room for further improvement of data products in remote regions.

1 Introduction

Reanalysis data sets of the 3-dimensional global atmosphere have become the most widely used data sets in geosciences. They serve numerous scientific communities such as impact modeling, risk management, and basic research in atmospheric and climate science. While data for the past 60 years are available from the popular reanalysis products ERA-40 (Uppala et al., 2005) and NCEP/NCAR (Kalnay et al., 1996), longer data sets are desirable for analyses of extreme weather events, for long-term impact studies, or generally for studies of variability on long time scales (reaching back to the preindustrial era) that require high-resolution data.

In order to extend conventional reanalysis projects further back into the past, more historical observations are needed, in particular upper-air observations. Recently, a compilation of historical upper-air data has been published, the Comprehensive Historical Upper-Air Network (CHUAN, Stickler et al., 2010). While CHUAN comprises a large number of historical profiles, the spatio-temporal coverage is very uneven. In particular, CHUAN contains data from only 5 North Atlantic ocean weather ships, and hence coverage is not good

over the oceans. In this paper we present early ship-based upper-air data that can be used in future reanalysis projects.

There is another way to obtain 3-dimensional global atmospheric data besides conventional reanalysis. A long reanalysis data set has recently been produced based on an assimilation of only surface and sea-level pressure (SLP) subdaily observations plus monthly sea-surface temperatures (SSTs) prescribed as boundary conditions using an Ensemble Kalman Filter technique. The feasibility of such a reanalysis was previously demonstrated (e.g., Compo et al., 2006; Whitaker et al., 2004). The "Twentieth Century Reanalysis" or 20CR (Compo et al., 2011) provides global, 6-hourly, 3-dimensional atmospheric data. We are using version 2 in the following, which reaches back to 1871. In a validation against independent historical upper-air data from land-based stations close agreement was found for geopotential height (GPH) in the troposphere over the northern mid-latitudes (Compo et al., 2011). The agreement is worse over the tropics. However, few upper-air data are available for independent validation in the tropics and the southern hemisphere, especially in the maritime regions prior to the 1950s.

Here we present upper-air data from two ship cruises in the tropical and southern ocean in the 1900s and 1930s, respectively, which we have digitized as an extension of the CHUAN data set, for comparison with 20CR, and for use in future reanalysis projects. The data comprise kite and registering balloon profiles from onboard the ship SMS *Planet* on a cruise from Europe around South Africa and across the Indian Ocean to the western Pacific in 1906/1907, and ship-based radiosonde data from onboard the MS *Schwabenland* on a cruise from Europe across the Atlantic to Antarctica and back in 1938/1939.

The paper is organized as follows. In Sect. 2, the data sources, digitizing, and processing steps are described and the comparison strategy is outlined. Results are shown in Sect. 3 and conclusions are drawn in Sect. 4.

2 Data

2.1 Observational data

2.1.1 SMS Planet

The first cruise analysed in this paper is one of the SMS (Seiner Majestät Schiff, i.e., His Majesty's Ship) "Planet" from Kiel (Germany) to Hong Kong in 1906/1907 (Fig. 1a shows the ship positions). The goal of this cruise was the Bismarck Archipelago, Papua New Guinea (at that time a German colony), where geodetic work was performed. On the cruise, oceanographic and atmosheric measurements were made. Leading aerologists were involved in planning, leading, and evaluating this cruise, the mission was well equipped, and the results are well documented (Reichs-Marine-Amt, 1909). Three platforms were used on the ship:
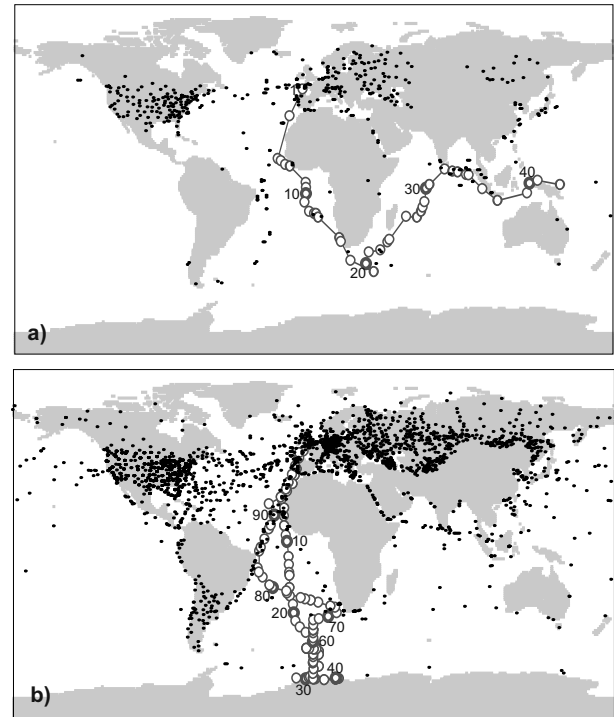

Fig. 1. Ship track and positions of upper-air soundings (open circles) for the cruise of the SMS *Planet* (**a**) and the MS *Schwabenland* (**b**). The dots mark the locations of surface pressure data that were assimilated into the Twentieth Century Reanalysis for 18 May 1906, 12Z (**a**) and 11 February 1939, 12Z (**b**).

pilot balloons, kites, and four registering balloons (i.e., free flying balloons with graphical registering devices, which need to be recovered after the burst of the balloon).

Although upper-air soundings were still in their infancy, this mission seems to have been on the forefront of research and new concepts were developed onboard (e.g., innovative attempts for controlled separation of registering devices from the balloon). The data comprise mostly temperature and wind but the latter is partly qualitative. Relative humidity is reported quantitatively at the surface and only qualitatively at higher levels. Pressure, which was used for calculating the height, is not given in the upper-air section of the report. Approximately one successful ascent per week was performed.

Ascents and descents were analysed by the ship scientists, but descents were only reported graphically as calibration curves for aneroid barometers refer to decreasing pressure, so that altitudes are less accurate for the descent. The ascent data are given as significant levels, indicating key features of the atmospheric profile.

The published report (Reichs-Marine-Amt, 1909) describes in detail the instruments and measurement procedure and also states the uncertainties. For pressure and temperature, the error is given as <2 mm and <2 °C at all heights (1 mm Hg = 1.33 hPa). The random error of the instrument reading is specified as <1 mm and <0.2 °C,

respectively. Concerning systematic errors, there is no mentioning of how radiation and lag errors were treated. All reduction procedures for balloons used on the ship assumed ascent (and descent) velocities of $5 \, \mathrm{m \, s^{-1}}$.

In this study we use only the data from Kiel to Bird Island, which cover 42 soundings performed over a period of 257 days. Four further profiles would be available, but were omitted in this study because they are separated from the other profiles by a gap of four months due to geodetic work. We only used the data from kites plus the four ascents with registering balloons.

After digitizing and quality controlling the data, we interpolated the profiles to fixed geometric altitude levels of 200, 500, 1000, 1500, 2000, ..., 6000, 7000, ..., 14 000 m a.s.l. (above mean sea level). In some cases, the top level was extrapolated to the next level, mostly by less than 100 m (maximum was 284 m). Interpolated and extrapolated values were flagged. No corrections were done for the kite data. Since they were often not continuous and flights sometimes were lengthy, it would have been difficult to assume an ascent velocity. Additionally, since (relative) wind is required for kites to fly, we assume that the ventilation was sufficient so that no correction is needed. Note, however, that both radiation and lag errors would tend to lead to a warm bias.

In the four registering balloon profiles, we corrected lag and radiation errors in the same way as described in Brönnimann (2003), assuming a lag coefficient of 15 s and an ascent velocity of $5 \, \mathrm{m \, s^{-1}}$. Note that we have no source to verify this assumption; it is based solely on our previous experience, which showed these values to be reasonable when no other information is available.

An important part of the comparison concerns the estimation of the errors. A comparison of the kite instrument reading at the start (10 m) and the temperature measured on board the ship (maximum time offset of 1 h) reveals a standard deviation of the difference of 0.86 °C and a warm bias of 0.17 °C. On the one hand, the surface instrument also has an error which is contained in this number, and the maximum time offset adds to the difference. Therefore the error of the kite instrument reading itself might be smaller. On the other hand, flight conditions might be more adverse (including reducing the graphical registration) and hence the error might be larger aloft than at the ground. In fact the difficulty of reading the graphical registration is frequently mentioned.

Based on these considerations we assume that the error of the temperature reading (σ_{instr}) can be quantified by a normal distribution with standard deviation of 0.9 °C.

In order to estimate the temperature error for a given altitude (significant level) one needs to consider also the pressure error (since pressure is used to calculate height). If we assume, based on the error specified above, a standard deviation of the random pressure error of 1 mm Hg (1.33 hPa), this translates into an additional error σ_{alt} (standard deviation) for temperature of 0.1–0.15 °C under most conditions (but it can be larger during strong inversion). To this adds the

systematic error of the pressure reading. The error considered so far concerns the readings from the meteographs such as, for instance, the significant points. An additional error is expected from the interpolation. From the graphical depictions of the ascents given in the report (Reichs-Marine-Amt, 1909) one can try to asses this error, but the drawn curves are smoothed. We assumed an error of 0.1–0.2 °C (σ_{interp}).

Assuming that all three errors are independent, we can estimate the random error for temperature at a specified altitude level as

$$\sigma_{\mathrm{obs}} = \sqrt{(\sigma_{\mathrm{instr}}^2 + \sigma_{\mathrm{alt}}^2 + \sigma_{\mathrm{interp}}^2)}$$

which in this case amounts to around 0.87–0.93 °C. We adopt the higher number in the following.

In addition to this random error, there is a systematic error due to non-corrected radiation and lag errors (note that almost all ascents were performed during the day). They cannot be quantified exactly, but a range can be given. Near the ground, in case of strong inversions the lag error might lead to a cooling and the radiation error is negligible. However, in the free troposphere, both errors tend to lead to a warming. The upper limit of this warming can be estimated from the corrections normally applied to radiosonde temperatures, which would be around 0.8 °C. Note, however, that the kite errors are expected to be considerably smaller (smaller ascent velocity, better ventilation). Finally, there are other possible systematic errors both in the temperature and pressure measurements, e.g., due to calibration issues or processing of instrument readings.

2.1.2 MS *Schwabenland*

The Antarctic expedition of the MS ("Motorschiff", i.e., motor vessel) *Schwabenland*, commissioned by Hermann Göring, served to preserve German interests in the region (Fig. 1b shows the ship positions). Two aircraft that were transported by the ship performed many flights over Antarctica for aerial photogrammetry and cartography. The explored region of Antarctica is called Neuschwabenland (New Swabia) to the present day. On the cruise of the *Schwabenland* from Europe to Antarctica and back, daily to twice daily radiosonde ascents were made.

Results from these soundings have been published in the literature (e.g., Flohn, 1949), but most of the original data material was reportedly destroyed during the war. According to Flohn (1949) and Regula (1958), two radiosonde types were used: the Lang sonde (Reichsamt für Wetterdienst, 1940) and the marine sonde from the Marineobservatorium Wilhelmshaven (Geelhaar, 1942). Heinz Lange was responsible for the radiosonde ascents from the two systems, aided by a technician for each system.

A major fraction of the data were available to us in processed and re-evaluated form. They were part of a compilation of German radiosonde data that was performed in the

Table 1. Estimation of the error of the radiosonde observations at Freiburg i. B. based on the comparison with data from the two mountain sites Säntis and Jungfraujoch. Bold numbers are derived from the other numbers in the table, italics refer to expert estimation, normal printing refers to the residuals calculated from regressing pairwise observations.

Predictand	Predictor	temperature (°C)					pressure (hPa except where noted)				
		$\sqrt{(\sigma_{obs}^2+\sigma_{stat}^2+\sigma_{rep}^2)}$	$\sqrt{(\sigma_{obs}^2+\sigma_{rep}^2)}$	σ_{obs}	σ_{stat}	σ_{rep}	$\sqrt{(\sigma_{obs}^2+\sigma_{stat}^2+\sigma_{rep}^2)}$	$\sqrt{(\sigma_{obs}^2+\sigma_{rep}^2)}$	σ_{obs}	σ_{stat}	σ_{rep}
800 hPa	Säntis	2.04	**2.02**	**0.48**	*0.3*	**1.96**	22.20 gpm/2.26 hPa	**2.24**	**1.40**	*0.3*	**1.75**
700 hPa	Säntis	2.42	**2.40**	**1.39**	*0.3*	**1.96**	18.83 gpm/1.72 hPa	**1.69**	–	*0.3*	**1.75**
700 hPa	Jungfraujoch	2.25	**2.23**	**1.06**	*0.3*	**1.96**	24.24 gpm/2.22 hPa	**2.20**	**1.33**	*0.3*	**1.75**
600 hPa	Jungfraujoch	2.28	**2.26**	**1.13**	*0.3*	**1.96**	28.24 gpm/2.25 hPa	**2.23**	**1.38**	*0.3*	**1.75**
Jungfraujoch	Säntis	2.00		*0.3*	*0.3*	**1.95**	1.84 hPa		*0.3*	*0.3*	**1.82**
Säntis	Jungfraujoch	2.02		*0.3*	*0.3*	**1.97**	1.75 hPa		*0.3*	*0.3*	**1.72**
Adopted averages			**2.3**	**1.2**		**1.96**		**2.2**	**1.35**		**1.75**

late 1940s at the observatory of Lindenberg on behalf of the Soviet occupation forces (Beelitz and Robitzsch, 1949; see Brönnimann, 2003). The data were reworked by two leading aerologists of that time, Paul Beelitz, head of the Lindenberg Observatory until 1949, and his successor Max Robitzsch. There is a note describing the data, signed by the two persons:

"Tabelle der Ergebnisse der Radiosondagen.

Anschrift der Station: M.S. Schwabenland (Das Forschungsschiff kreuzte im Atlantik)

Bemerkungen: Das Material stammt vom Forschungs-Motor-Schiff Schwabenland, das auf der Fahrt von der Biscaya bis zur Antarktis erstellt wurde. Es liegen teils die Adiabatenblätter im Original, teils die Blätter der markanten Punkte vor. Die Aufstiege mussten nachrekonstruiert werden. Die Ausgangshöhe (Bordhöhe) wurde mit 15 m in Ansatz gebracht. Das Material ist sehr gut."

Translation: "Table of the results from the radio soundings

Name of the station: M.S. Schwabenland (The research vessel cruised in the Atlantic)

Comments: The material stems from the research motor vessel Schwabenland and was retrieved on the cruise from the Bay of Biscay to Antarctica. It was available partly in the form of adiabatic papers, partly in the form of significant points. The ascents had to be reconstructed. The departure height (board height) was assumed as 15 m. The material is very good."

The source does not specify the sonde type used, but a comparison with the data tabulated in Regula (1958) indicates that it was the material from the Lang sonde. The data from the marine sonde are also tabulated in Regula (1958), but only in abbreviated form (GPH on five pressure levels, humidity and temperature on altitude levels). Only the data from the Lang sonde from Beelitz and Robitzsch (1949) were used in the following.

The data were reworked in the same way as described in Brönnimann (2003), i.e., radiation and lag errors were corrected based on the formulation of Väisälä (1941, 1949) and

Raunio (1950) and the published radiation error of the German sonde (most likely the Lang sonde) by Scherhag (1948). Flohn (1949) analysed the radiation error of that data and concluded that the corrections following the approaches of Väisälä and Scherhag are sufficient. Note that it is not known whether a lag correction was originally applied. The manual for the sonde foresees a lag correction, but was published two years later (Reichsamt für Wettterdienst, 1940). In contrast to our previous work (relating to other data from the same source) we also analysed relative humidity, which in the Lang sonde was measured with a hair hygrometer. However, no corrections were made to humidity. Flohn (1949) states that the error in relative humidity in the Schwabenland ascents was very large for low temperatures and low humidities. For more information on the German radiosondes see DFVLR (1982).

In order to estimate the error of the soundings, we used German radiosonde data from approximately the same time period but from a location that is close to the Alps, namely Freiburg i. B., 1940–1942. Note that in earlier work (Brönnimann, 2003) we assumed that the instrument type used in Freiburg i. B. was a Graw H-38; however, we now think that this also was a Lang sonde (Reichsamt für Wetterdienst, 1940) and hence likely the same sonde type as on the MS *Schwabenland*.

We compared temperature and GPH above the planetary boundary layer at 800, 700, and 600 hPa with station temperature and pressure from mountain sites, namely Säntis (2500 m a.s.l., 140 km away) and Jungfraujoch (3555 m a.s.l., 161 km), respectively. We selected only pairs of observations that were performed within 3 h of each other. Then we performed a linear regression to estimate radiosonde data (predictand) from the corresponding station data (predictor) using one predictor at the time. Finally, we analysed the variance of the residuals. Results are summarized in Table 1. Note that we did not subtract an annual cycle, nor did we include the annual cycle in the regression model.

The agreement between the regression models and the radiosonde data was generally very good, with explained

variances of 83% to 96%. This not only points to a high quality of the data, but also is important to justify the applicability of the general approach. Concerning the analysis of the residuals, we assume again that all errors are independent and hence the standard deviation of the residuals $\sigma_{\text{residuals}}$ is the square root of the sum of the variances of the individual error contributions:

$$\sigma_{\text{residuals}} = \sqrt{(\sigma_{\text{obs}}^2 + \sigma_{\text{stat}}^2 + \sigma_{\text{rep}}^2)}$$

where σ_{obs} is again the error of the observation ($\sigma_{\text{obs}} = \sqrt{(\sigma_{\text{instr}}^2 + \sigma_{\text{alt}}^2 + \sigma_{\text{interp}}^2)}$), σ_{stat} is that of the station reading and σ_{rep} is that of the representativeness (i.e., the error attributed to the distance in time, space, and due to the comparison of free atmospheric data with surface data).

For the error of the station observation, an expert estimation of the standard deviation of the error is $\sigma_{\text{stat}} = 0.3\,°\text{C}$ for temperature and $\sigma_{\text{stat}} = 0.3\,\text{hPa}$ for pressure.

In order to address σ_{rep}, we compared the differences between the mountain stations, again with a regression model, and assumed that the variance of the residuals, after subtracting σ_{stat} of both stations, is a good estimation of σ_{rep}. This is not exactly true as comparing two ground stations does not account for the error of comparing a free atmospheric measurement versus a mountain station. Also, the altitude difference between the stations is larger than between a station and the closest pressure level. Comparing Säntis and Jungfraujoch, for instance, gives $\sigma_{\text{rep}} \approx 1.96\,°\text{C}$ and $\sigma_{\text{rep}} \approx 1.75\,\text{hPa}$ (depending on the choice of x and y) between the stations (see Table 1)

Based on these values we estimate σ_{obs} as $1.2\,°\text{C}$ and 1.35 hPa, the latter of which was transformed back to geopotential meters using climatological temperature profiles (see Sect. 2.3). Note that the error of representativeness is only a rough estimation that might be too large (because in most cases the distance is smaller than between Säntis and Jungfraujoch) or too small (because the ships' locations may be inaccurate due to digitizing errors), but in any case has a large effect on the derived error of the observations. For relative humidity, we do not know the error and hence do not try to quantify any humidity errors in this paper.

A more direct approach is to analyse temperature at 4 km (as given in Regula, 1958) in pairs of ascents from the Lang sonde and the Marinesonde (44 pairs with a time difference <12 h). Various approaches of extrapolating the standard deviation of their difference to zero time shift suggest an observation error near $1\,°\text{C}$ for each sonde, although with a large uncertainty, which is consistent with the above estimates.

2.2 Reanalyses

The observation-based data were compared with 20CR (Compo et al., 2011). We used 6-hourly ensemble mean analysis fields of temperature, GPH and relative humidity as well as the ensemble spread (expressed in the form of the ensemble standard deviation, termed σ_{20CR}) of these fields. We

interpolated the fields to the locations of the ships and chose the standard time closest to the observing time for comparison. In the case of SMS *Planet*, the temperature data were interpolated onto fixed (geometric) altitude levels in order to match the observations.

The error of the interpolation is not exactly known. Note, however, that the spatio-temporal distance of the interpolation of reanalysis data to radiosonde observations ($2° \times 2°$ grid, 6 hourly analyses), is of the same order as that used in the comparison for the case of Freiburg i. B. in Sect. 2.1 (Δx 140–160 km, $\Delta t < 3$ h). If we assume that the interpolation error is constant in time and space and does not have a seasonal cycle, we can use σ_{rep} also as a conservative approximation for the interpolation error in the reanalysis.

When analyzing the early years of the reanalysis, it is important to consider the locations of the pressure measurements that go into 20CR (note that no other information was assimilated). The locations are shown in Fig. 1 for one sample day in each cruise (18 May 1906 and 11 February 1939, respectively). Note that the surface pressure readings from the SMS *Planet* are included in the historical reanalysis, whereas this is not the case for the observation from the MS *Schwabenland*. However, in the case of the SMS *Planet*, there are almost no other pressure data within a few thousand kilometers. In the case of the MS *Schwabenland*, pressure data were more abundant, but not south of 60° S. Hence, both cruises sampled very remote regions of the globe with respect to information assimilated into 20CR.

For some of the comparisons, it is advisable to plot the data in the form of anomalies from a common reference. We used NCEP/NCAR reanalysis (NNR) data for this purpose, namely a climatology of daily mean values as a function of the day of year that is given on the website of NOAA/PSD and refers to the period 1968–1996. These data also were interpolated to the ship's location.

2.3 Comparison

The data were analysed as absolute values and as deviations from the NNR climatology. In addition to the standard measures (bias, correlation), we also analysed whether the difference between observations and 20CR is compatible with both the ensemble spread and the observation error assuming that the two errors are uncorrelated. Specifically, we calculated the fraction of cases for which the difference between observation and reanalysis is outside $\pm 2\,\sigma_{\text{diff}}$ where

$$\sigma_{\text{diff}} = \sqrt{(\sigma_{\text{20CR}}^2 + \sigma_{\text{obs}}^2 + \sigma_{\text{rep}}^2)}.$$

Note that the sum of the errors ($\sigma_{\text{obs}}^2 + \sigma_{\text{rep}}^2$) is even more readily comparable to the cases given in Table 1 than the individual error contributions. Note also that σ_{diff} does not cover all sources of errors and uncertainties, as will be discussed below. For instance, biases (both in 20CR and in the observations) are not included.

Table 2. Results of the comparison of temperature between observations onboard the SMS *Planet* and 20CR. ΔT is the mean difference between 20CR and the observations, T_{anom} refers to the anomalies from NNR climatology, r is the correlation coefficient, P stands for probability, and $P(|\Delta T| > 2\sigma_{\text{diff}})$ is the fraction of differences outside the interval $\pm 2\sigma_{\text{diff}}$. Bold numbers indicate statistical significance (p-values < 0.05 for two-sided tests, paired t-test in the case of the bias).

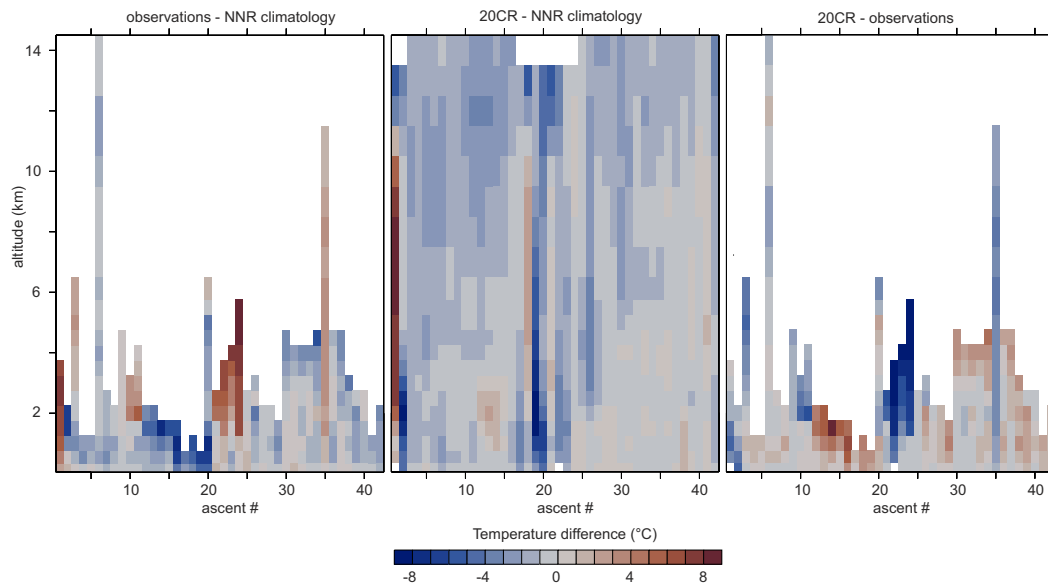| Level (m a.s.l.)	200	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500
N	40	42	40	36	32	25	20	17	15	10	5	5
ΔT [°C]	0.22	**0.73**	**0.92**	**0.48**	−0.90	**−2.01**	−1.58	−1.23	−0.22	−0.21	−3.28	−3.96
$r(T)$	**0.970**	**0.937**	**0.877**	**0.729**	**0.754**	**0.720**	**0.662**	**0.592**	0.458	0.594	0.805	0.770
$r(T_{\text{anom}})$	**0.574**	**0.525**	**0.374**	0.270	**0.406**	**0.409**	0.360	0.221	−0.336	−0.122	0.648	0.748
$P(|\Delta T| > 2\sigma_{\text{diff}})$	0.025	0.024	0.025	0.083	0.156	0.125	0.150	0.177	0.133	0.200	0.200	0.200

Fig. 2. Temperature profiles from observations and 20CR for the cruise of the SMS *Planet*. (left) Anomalies of observations with respect to NNR climatology, (middle) anomalies of 20CR with respect to NNR climatology, (right) difference 20CR minus observations.

3   Results

3.1   Comparison for SMS Planet

Figure 2 shows vertical temperature differences between observations and NNR climatology (left), between 20CR and NNR climatology (middle), and between 20CR and observations (right). The corresponding statistics are given in Table 2.

The weather during this cruise was rather normal (see Fig. 2 left). The main anomaly features are the very first two profiles, then a series of profiles with rather cool conditions in the lowest 1000 m (#12–20, extratropical South Atlantic and Southern Ocean), and a sequence of profiles with a very strong, high inversion (#21–24, Southern Indian Ocean). The weather log for ascents #12–20 mentions frequently calm conditions aloft, often cloudy, and generally difficult conditions for kite flights. The inversions in ascents #21–24 were

reported in the log and concurred with reportedly very dry conditions at these altitudes and sometimes weak winds. At most other times, temperatures were close to climatology.

The 20CR depicts the anomalies during the first two profiles. It also shows the cool, unstable conditions in ascents #17–20 (the southernmost ascents of the cruise). However, the reanalysis does not capture the strong inversions. This affects the mean differences between the two data sets, which would be within ±1 °C without profiles #21–24, but with these profiles reach 2 °C and are statistically significant at most levels.

The correlations (see Table 2) are very high for the observed values of temperature in the lower troposphere but moderate (0.3–0.6 up to 3 km, somewhat less for the 1500 m a.s.l. level which is near the boundary layer top) for the anomalies. Again, this is largely due to the lacking inversion. Without profiles #21–24 anomaly correlations would be on the order of 0.6–0.8 (somewhat lower at 1500 m a.s.l.),
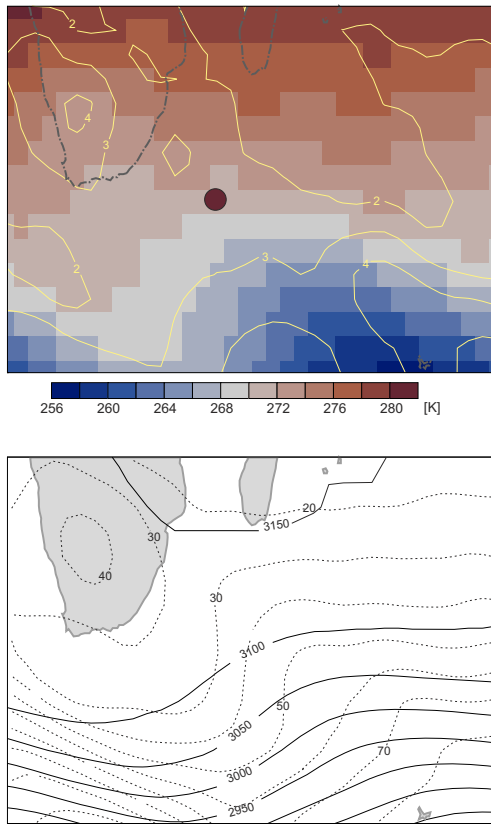
Fig. 3. Data from 20CR at 700 hPa on 18 May 1906, 12:00 UTC. Top: ensemble mean temperature (colours) and ensemble spread (yellow lines). The large circle indicates the location and value of the observed temperature onboard SMS *Planet* (ascent number 22). Bottom: ensemble mean GPH (solid lines) and ensemble spread (dashed). The area displayed is (20° S–50° S, 10° E–80° E).

which would be considered very high even when compared to results by Compo et al. (2011) that were derived from later periods. However, the presence or not of inversions might be an important regional feature. The inversion occurred southeast of Port Elizabeth, South Africa, over the South Indian Ocean. This region is further south than the usual extent of the trade winds and trade wind inversion but might be affected by inversions occasionally.

Figure 3 shows 700 hPa temperature and GPH on 18 May 1906, 12Z, from 20CR. Shown are the ensemble mean (colours and contours) as well as the spread (yellow and dashed lines). Also shown is the observed temperature (interpolated to the geometric altitude of the ensemble mean 700 hPa surface). The difference between observations and 20CR exceeds 10 °C and is clearly outside $\pm 2\sigma_{\text{diff}}$. The high pressure system is much further north, near Madagascar, and does not influence the region. The ensemble spread is relatively small. Hence in this case we suggest that 20CR does not capture the feature.

The last row in Table 2 shows the fraction of differences between 20CR and the observations that is outside $\pm 2\sigma_{\text{diff}}$. If all errors were included in the estimation of σ_{diff}, a fraction of 5% would indicate that reanalysis and observations are in agreement given the specified errors. However, the inversion issue affects 10% of the profiles. In fact, near the ground the fraction is below 5%, but then the fraction increases to over 10% and hence slightly higher than expected. This might be due to biases. The likely warm bias in the observations is not accounted for in σ_{diff}, and 20CR also might have biases (biases are not covered by the ensemble spread).

3.2 MS *Schwabenland*

Now, we consider the cruise of MS Schwabenland from December 1938 to April 1939. In contrast to the SMS Planet, the data from the MS *Schwabenland* pose particular problems. Figure 4 shows, for two variables (500 hPa GPH and 700 hPa temperature), sonde observations and co-located analysis values from 20CR as well as the anomalies with respect to NNR climatology. Also shown are the corresponding uncertainties $2\sigma_{\text{obs}}$ and $2\sigma_{\text{20CR}}$, respectively. It is apparent that the observations and 20CR have substantial differences around ascents #50–61. Temperatures in the observations appear to be 10 °C higher, and GPH also is higher. Moreover, this difference is far larger than the errors of the two series.

One could simply omit that data as possibly erroneous. However, the description of the data asserts a high quality (see quote), and there is an obviously high correlation (0.86 for 700 hPa temperature) between anomalies in observations and reanalysis within that series of 12 profiles. The same deviation is also found in the data tabulated in Regula (1958) and it is also found in the data from the Marinesonde.

The nature of the error is difficult to specify. We therefore analysed the vertical profile of observations minus reanalysis (Fig. 5) in order to use the profile shape to attribute the error to an underlying (usually simple) problem (see Grant et al., 2009 for more details). Uncorrected radiation and lag errors can be ruled out as they are far too small to explain a 10 °C difference (although the shape of the lag error seems correct). Other potential problems such as unit errors, a constant offset in temperature or a constant offset in pressure would produce difference profiles with distinctly different shapes.

The very characteristic belly shape of the temperature difference profile (Fig. 5 right, red curve) can be very easily explained, however, by a shift of the temperature profile by a constant altitude. Since some of the profiles were taken from adiabatic diagrams, which usually have altitude as a vertical axis, such an error might occur when copying the data, for instance. We have no direct evidence for such an error. However, the fact that correcting for an assumed constant offset of 1500 m brings the difference profile (Fig. 5 right, purple curve) at all levels to very close agreement with that for the rest of the record (Fig. 5 right, blue curve) may be considered indicative of this type of error. Note that we do
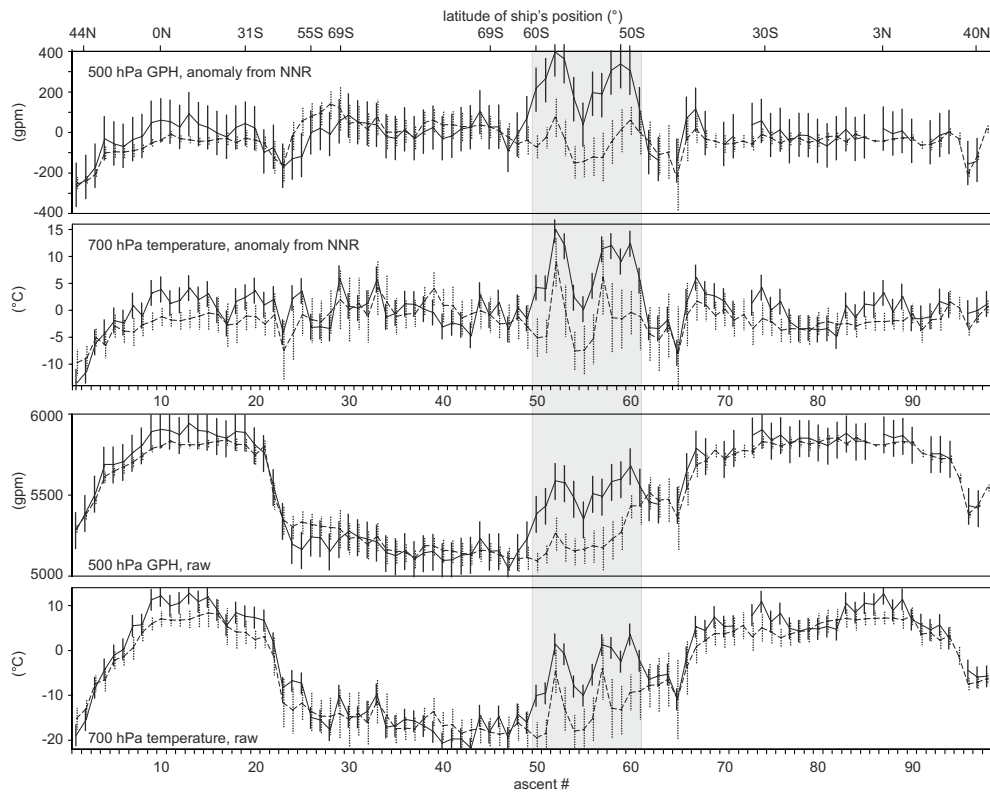
Fig. 4. 500 hPa GPH and 700 hPa temperature from the MS *Schwabenland* (solid lines) and 20CR (dashed lines) as a function of the ascent number. The top two panels display the data in the form of anomalies from the NNR climatology, the bottom two panels show the raw data. The vertical bars give $2\sigma_{obs}$ (solid) and $2\sigma_{CR}$ (grey dashed) respectively. The shaded region of ascents 50 to 61 highlights the suspected erroneous observations.
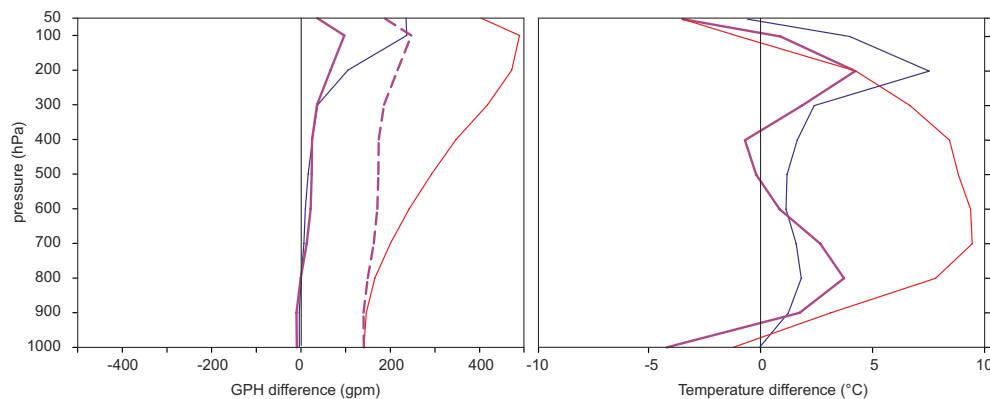

Fig. 5. Vertical profile of the difference between observations from the MS *Schwabenland* and 20CR for GPH (left) and temperature (right). Red = uncorrected profiles for the ascents #50–61, blue = all other ascents, thick purple = corrected profiles of ascents #50–61 (dashed: no altitude offset at 1000 hPa). Note that the 1000 hPa level was below sea level during ascents #50–61 and hence these values were extrapolated on the original data sheets.

not expect the difference to vanish at all levels. There may be remaining errors in the observations or in the reanalysis. For instance, a cold bias at the tropopause in 20CR might affect the uppermost three levels in Fig. 5. Also, note that the 1000 hPa level was below sea level during all ascents #50–61. The values for this level were hence extrapolated in our sources (but nevertheless should be considered in Fig. 5, in contrast to the rest of the paper, as they are likely a product of the same data processing).

Table 3. Results of the comparison of temperature (T), GPH (Z), and relative humidity (rH) between observations onboard the MS *Schwabenland* and 20CR. Subscript "anom" refers to the anomalies from NNR climatology, r is the correlation coefficient, P stands for probability, and $P(|\Delta T| > 2\sigma_{\mathrm{diff}})$ is the fraction of differences outside the interval $\pm 2\sigma_{\mathrm{diff}}$.

Level (hPa)	1000	900	800	700	600	500	400	300	200	100	50
$n(Z)$	98	96	96	95	95	91	86	79	70	56	21
$n(T)$	55	96	96	96	95	92	85	80	70	53	21
n (rh)	59	78	78	77	75	55					
ΔZ [gpm]	4.4	4.2	−0.2	−7.8	−11.5	−17.3	−25.8	−37.2	−99.9	−212.3	−169.2
ΔT [°C]	0.0	−1.3	−2.0	−1.7	−1.1	−1.0	−1.4	−2.3	−7.1	−3.6	1.6
Δrh [%]	3.7	0.2	7.6	11.4	12.9	9.2					
$r(Z)$	0.858	0.915	0.947	0.966	0.975	0.981	0.982	0.984	0.981	0.920	0.664
$r(T)$	0.954	0.957	0.964	0.969	0.970	0.970	0.963	0.936	0.057	0.965	0.921
$r(\mathrm{rh})$	0.357	0.196	0.313	0.321	0.322	0.389					
$r(Z_{\mathrm{anom}})$	0.472			0.608	0.655	0.673	0.658	0.640	0.662	0.303	0.233
$r(T_{\mathrm{anom}})$	0.626			0.664	0.582	0.562	0.517	0.432	0.236	0.572	0.220
$r(\mathrm{rh}_{\mathrm{anom}})$	0.312			0.132	0.151	−0.005					
$P(|\Delta Z| > 2\,\sigma_{\mathrm{diff}})$	0.133	0.051	0.041	0.042	0.042	0.033	0.093	0.101	0.229	0.571	0.476
$P(|\Delta T| > 2\sigma_{\mathrm{diff}})$	0.104	0.061	0.112	0.073	0.063	0.076	0.047	0.163	0.529	0.132	0.095

Recalculating GPH from the shifted temperature profile (assuming a zero height error for the surface) also removes the vertical structure in the GPH difference profile entirely, but leaves a constant offset (Fig. 5 left, dashed purple curve) arguably because the height offset at 1000 hPa is not zero. Hence, we further assumed a constant offset of 150 gpm at the surface (Fig. 5 left, purple curve). After this correction, the curve fits very well to that for the rest of the record.

Note that the corrections may seem arbitrary, and indeed they are to some extent. Also, by using 20CR as a reference we lose independence. On the other hand, we have determined only two parameters and (largely) remove different biases for 20 variables simultaneously (10 levels, 2 variables per level). Note also that around 200 observation pairs have been used to constrain the two parameters and that the nature of the error is plausible. For these reasons we utilize these corrections although we note that there may be remaining problems.

The comparison of time-height sections for the corrected data as well as the statistical analysis are given in Fig. 6 and Table 3, respectively. Note that in contrast to Fig. 2 (for SMS *Planet*), we now have pressure as the vertical coordinate, and we also show GPH and relative humidity. Note also that, because of the mismatch between the reported pressure levels and that of the NNR climatology, 800 and 900 hPa cannot be shown in anomaly form (interpolating the climatology to these levels might add additional uncertainties especially since these levels are close to the boundary layer top), but they can be shown in the direct comparison.

There is one visually very prominent feature in this figure, namely very cool temperatures near the tropopause in 20CR. This has been found also in other comparisons and likely points to a cold bias in 20CR (Compo et al., 2011), though

historical upper-air data could also be warm biased in the stratosphere. Apart from that, the agreement between reanalysis and observations is very good for temperature and GPH (Table 3). We find anomaly correlations of 0.5–0.7 throughout the troposphere, which is at least as good as expected from Compo et al. (2011). Also, correlations for temperature anomalies are high in the lower stratosphere (they drop at the tropopause level).

Most GPH and temperature differences are within $\pm 2\sigma_{\mathrm{diff}}$ in the lower to middle troposphere (see also vertical bars in Fig. 4). Around 5–10% are outside the spread, increasing towards the upper troposphere and stratosphere (and also increasing at 1000 hPa). Similar as for the SMS *Planet*, this might be due to biases, which increase with altitude (arguably in both the radiosonde data and in 20CR). Since biases are not accounted for in σ_{diff}, we consider these results to be consistent with the estimated errors.

Relative humidity is shown in the lowermost panel. Measuring relative humidity with weather balloons is difficult even in the present day and hence we do not expect high-quality data from the 1930s (as is also stated in Flohn, 1949). In the lower troposphere, we do find some agreement between the two (anomaly correlations are significant at 1000 hPa, which is consistent with Flohns judgement (1949) of smaller errors with higher temperatures), while huge discrepancies arise in the middle troposphere. As we know nothing about the errors in humidity, we do not estimate σ_{diff}. Evidently, relative humidity remains a challenge in historical data sets.

One of the strongest anomaly features on this cruise occurred in mid- February 1939 (within the corrected interval). Temperature in the lower troposphere increased by 10 °C within a day and then decreased again by the same
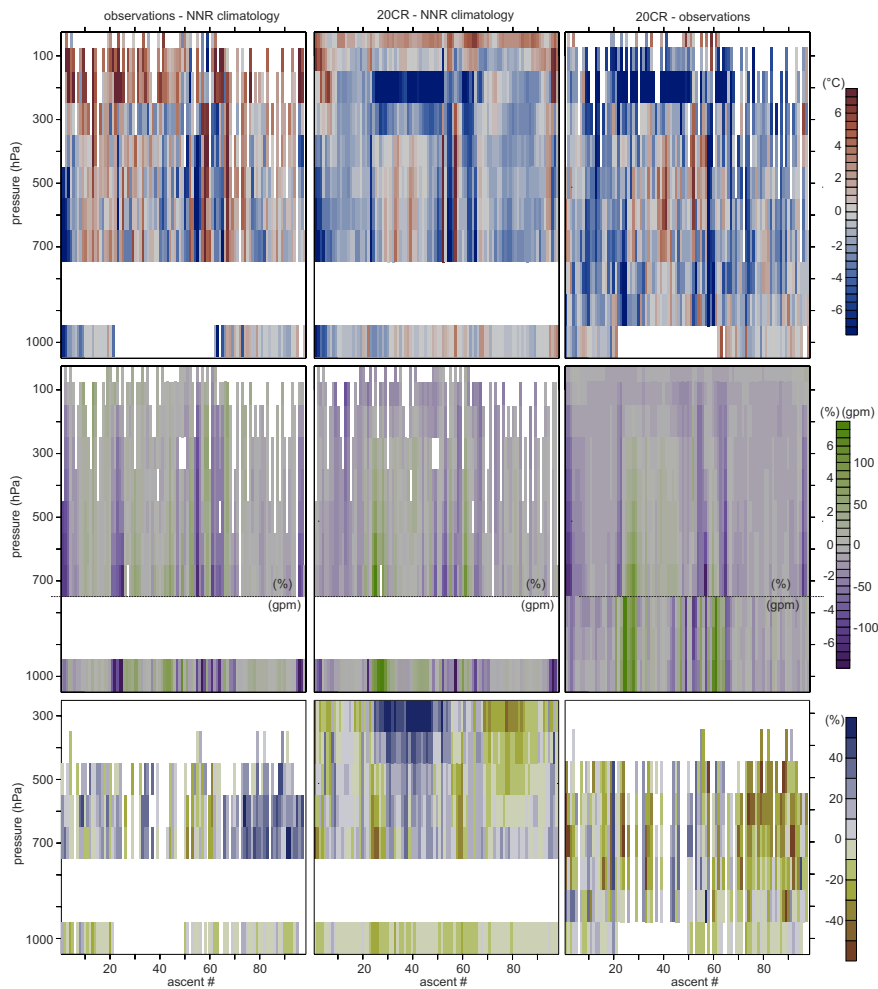
Fig. 6. Vertical profiles from observations and 20CR for the cruise of the MS *Schwabenland* for temperature (top), GPH (middle) and relative humidity (bottom). (left) Anomalies of observations with respect to NNR climatology, (middle) Anomalies of 20CR with respect to NNR climatology, (right) difference 20CR minus observations.

amount. Figure 7 shows fields of 700 hPa GPH and temperature from 20CR, together with the ensemble standard deviation for 700 hPa GPH for the 11, 13, and 16 February.

The general synoptic situation is dominated by high synoptic activity. On the 11th of February the ship was located in a trough, with cold air penetrating far equatorwards. A second disturbance is located near the South Orkney Islands and develops into a strong low pressure system that moves eastward. Two days later the ship is located in the warm sector ahead of a cold front. The latter appears also as a peak in the ensemble spread for 700 hPa GPH. Three days later the circulation has switched to a more zonal regime, and the ship is again located in the cold polar air mass.

For this specific period, 20CR and observations fit very well, despite the remote location of the ship (near 60° S). SLP data from several ships in this region were assimilated into 20CR. Although this information was small in number, it was obviously sufficient to reproduce this feature.

4 Conclusions

In order to extend global 3-dimensional data sets back into the past, more historical upper-air data are needed, especially over remote locations such as the world oceans. Here we have presented early, ship-based, upper air data from two cruises from Europe to the tropics and even to Antarctica. We have tried to quantify the errors and found random temperature errors of 0.9–1.2 °C and pressure errors of 1.35 hPa (standard deviations). We have also suggested the presence of biases that would add to these errors.

Comparison with 20CR shows generally close agreement. The fact that we do find agreement on a day-to-day time scale in such remote places suggests that even over sparsely observed regions of the globe in the 1st half of the 20th century, successful reanalysis is possible. The results also show that, importantly, error estimations are consistent. At the same time, we also demonstrate that some features may not be well
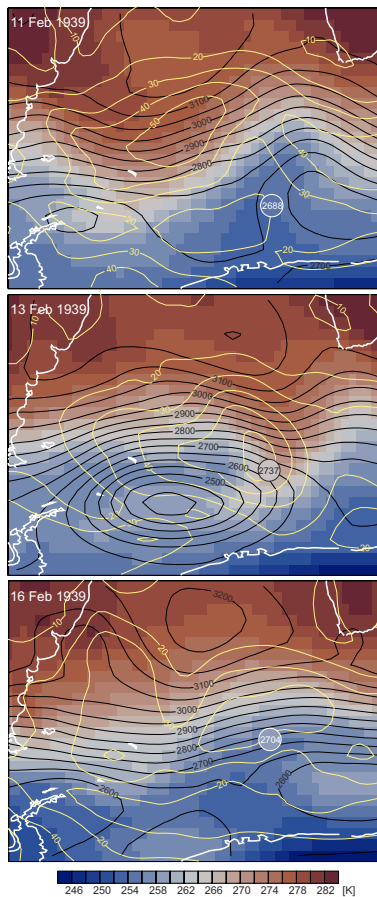
Fig. 7. GPH (solid lines) and temperature (colours) at 700 hPa for three days in February 1939 (12:00 UTC). The yellow lines indicate the ensemble spread of GPH. The large circles and numbers indicate the observed temperature and GPH, respectively. The area displayed is (70° W–30° E, 74° S–20° S).

represented. More efforts are needed to improve both quality and quantity of historical upper-air data. In fact, much more ship-based, upper-air data could be digitised for the first half of the twentieth century (e.g., see listings in Flohn, 1949). These data could then be used in future reanalysis efforts, such as those of the ERA-CLIM project, which will develop the next long European reanalysis. The data presented in this paper will be incorporated into CHUAN (Stickler et al., 2010).

References

Beelitz, P. and Robitzsch, M.: Arbeiten aus dem Aerologischen Archiv des Observatoriums Lindenberg Kr. Beeskow. Zusammengestellt von Direktor Dr. Paul Beelitz und Professor Dr. Max Robitzsch. Zusammenstellung der Radiosonden – Aufstiege von Mitteleuropa 1939 – 1944 (Nach dem Archiv des Observatoriums Lindenberg), I. Teil, Oktober 1949, II. Teil, Dezember 1949.

Brönnimann, S.: A historical upper air data set for the 1939–1944 period, Int. J. Climatol., 23, 769–791, 2003.

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

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

DFVLR (Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt) (Ed.): Die Entwicklung der deutschen Radiosonden von 1930–1955, 1982.

Flohn, H.: Grundzüge der allgemeinen atmosphärischen Zirkulation auf der Südhalbkugel auf Grund der aerologischen Ergebnisse der Deutschen Antarktischen Expedition "Schwabenland" 1938/39, Arch. Met. Geoph. Biokl. A (Meteorology and Atmospheric Physics), 2, 17–64, 1949.

Gehlhaar, W.: Die meteorologischen Messelemente der Marineradiosonde, Arch. dt. Seewarte u. Marineobs., Vol. 62, Nr. 1, Hamburg, 1942.

Grant, A. N., Brönnimann, S., Ewen, T., and Nagurny, A.: A New Look at Radiosonde Data Prior to 1958, J. Climate, 22, 3232–3247, 2009.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77, 437–471, 1996.

Raunio, N.: Amendments to the computation of the radiation error of the Finnish (Väisälä) radiosonde, Geophysica, 4, 14–20, 1950.

Regula, H.: Die Wetterverhältnisse während der Expedition und die Ergebnisse der meteorologischen Messungen, in Deutsche Antarktische Expedition 1938/39, edited by: Kottas, A. and Ritscher, A., Wissenschaftliche Ergebnisse. Vol. 2, No. 1, 16-40. Hamburg, 1958.

Reichsamt für Wetterdienst: Technische Vorschrift für die Radiosonde (Modell Lang) 1940 (Vorläufige Form), Berlin, Germany, 1940.

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

Scherhag, R.: Neue Methoden der Wetteranalyse und Wetterprognose, Springer, Berlin, Germany, 1948.

Stickler, A., Grant, A. N., Ewen, T., Ross, T. F., Vose, R. S., Comeaux, J., Bessemoulin, P., Jylhä, K., Adam, W. K., Jeannet, P., Nagurny, A., Sterin, A. M., Allan, R., Compo, G. P., Griesser, T., and Brönnimann, S.: The Comprehensive Historical Upper-Air Network, Bull. Amer. Meteorol. Soc., 91, 741–751, 2010.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteorol. Soc., 131, 2961–3012, 2005.

Väisälä, V.: Der Strahlungsfehler der finnischen Radiosonde, Mitteilungen des Meteorologischen Instituts der Universität Helsinki No. 47, 1941.

Väisälä, V.: Solar radiation intensity at the ascending radiosonde, Geophysica, 3, 37–55, 1949.

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