

S. Brönnimann · J. Luterbacher

Reconstructing Northern Hemisphere upper-level fields during World War II

Received: 6 August 2003 / Accepted: 18 December 2003 / Published online: 2 April 2004
© Springer-Verlag 2004

Abstract Monthly mean fields of temperature and geopotential height (GPH) from 700 to 100 hPa were statistically reconstructed for the extratropical Northern Hemisphere for the World War II period. The reconstruction was based on several hundred predictor variables, comprising temperature series from meteorological stations and gridded sea level pressure data (1939–1947) as well as a large amount of historical upper-air data (1939–1944). Statistical models were fitted in a calibration period (1948–1994) using the NCEP/NCAR Reanalysis data set as predictand. The procedure consists of a weighting scheme, principal component analyses on both the predictor variables and the predictand fields and multiple regression models relating the two sets of principal component time series to each other. According to validation experiments, the reconstruction skill in the 1939–1944 period is excellent for GPH at all levels and good for temperature up to 500 hPa, but somewhat worse for 300 hPa temperature and clearly worse for 100 hPa temperature. Regionally, high predictive skill is found over the midlatitudes of Europe and North America, but a lower quality over Asia, the subtropics, and the Arctic. Moreover, the quality is considerably better in winter than in summer. In the 1945–1947 period, reconstructions are useful up to 300 hPa for GPH and, in winter, up to 500 hPa for temperature. The reconstructed fields are presented for selected months and analysed from a dynamical perspective. It is demonstrated that the reconstructions

provide a useful tool for the analysis of large-scale circulation features as well as stratosphere-troposphere coupling in the late 1930s and early 1940s.

Electronic Supplementary Material Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00382-004-0391-3>

1 Introduction

The analysis of upper-level fields is undoubtedly an important tool in climate research. For instance, the vivid debate on the North Atlantic or Arctic Oscillation (NAO/AO) and the role of the stratosphere in climate variability of the Northern Hemisphere (NH) is based on the analysis of time series of upper-level fields (Thompson and Wallace 1998; Ambaum and Hoskins 2002; Hurrell et al. 2003). A variety of gridded data sets are currently available for climate research. For the most recent period (since about 1978), satellite-based products such as MSU temperatures (Christy et al. 2003; Mears et al. 2003; daily data for three deep atmospheric layers) are available. Further back in time, upper-level data sets are based on radiosonde observations such as the Berlin Stratospheric data (Labitzke et al. 2002; daily/monthly temperature and geopotential height at 100 to 10 hPa for the NH back to 1957) or HadRT (Parker et al. 1997; global gridded monthly temperature at 850 to 30 hPa back to 1958). Radiosonde station series are available back to 1948 from the Comprehensive Aerological Reference Data Set (CARDS) (see also Lanzante et al. 2003). The most widely used gridded meteorological data sets with upper-level variables, however, are re-analysed data sets, i.e. combined surface, radiosonde, and satellite information assimilated using a filtering method and a weather prediction model (ERA-15, ERA-40, NCEP/NCAR Reanalysis, see Kalnay et al. 1996; Kistler et al. 2001). They provide 6-hourly global data up to the middle stratosphere back to 1957 (ERA-40)

S. Brönnimann (✉)
Lunar and Planetary Laboratory, University of Arizona,
PO Box 210092, Tucson, AZ 85721-0092, USA
E-mail: stefanb@lpl.arizona.edu

J. Luterbacher
Institute of Geography, University of Bern, Bern,
Switzerland

J. Luterbacher
NCCR Climate, University of Bern, Bern,
Switzerland

and 1948 (NCEP/NCAR Reanalysis). No upper-level meteorological data set based on observations goes further back in time than 1948.

In order to study past climate variability, upper-level fields for earlier time periods would be beneficial. Several authors have suggested statistical and other approaches for obtaining historical upper-level fields (Kington 1975; Klein and Dai 1998; Schmutz et al. 2001; Luterbacher et al. 2002; Polansky 2002) or reconstructed such fields (Schmutz et al. 2001; Luterbacher et al. 2002; see also Polansky 2002). However, these reconstructions were based on information from the Earth's surface only and were limited to geopotential height (GPH) in the troposphere over certain regions.

Here we present a data set of upper-level fields for the extratropical Northern Hemisphere (NH) for the 1939-1947 period, comprising temperature and GPH and reaching into the lower stratosphere (100 hPa). The chosen time period allows to study the large-scale circulation in the early 1940s and provides a connection to the NCEP/NCAR Reanalysis period. Knowledge of the upper-level circulation is considered especially important in the case of the early 1940s, when exceptionally high total ozone values were observed concurrently with extreme climate anomalies at the ground (see Brönnimann 2003b). A detailed analysis of this period is interesting in the light of current research on stratosphere-troposphere coupling and large-scale climate variability (e.g. Hurrell et al. 2003). Reconstructions based only on information from the Earth's surface are of limited use in this case, and upper-air data prior to 1948 that could be used as predictors in a reconstruction approach were not available at the beginning of the project. However, older upper-air data can be found on paper in various meteorological archives. We digitised, corrected, and validated several tens of thousands of GPH and temperature profiles from radiosonde ascents and weather flights for the 1939-1944 period (Brönnimann 2003a, b) and supplemented them with meteorological data from the Earth's surface. In this work we report on the results of a statistical reconstruction, based on these data, of monthly mean fields of temperature and GPH at 700, 500, 300, and 100 hPa for the extratropical NH (20°-90°N) from 1939 to 1947. Data and methods are briefly presented in Sects. 2 and 3, respectively. Results of sensitivity analyses and validation experiments are discussed in Sect. 4. In addition, reconstructions are presented for selected months in order to demonstrate their usefulness for the analysis of large-scale circulation.

2 Datasets used for the reconstruction

In order to reconstruct and validate upper-level fields for the 1939-1947 period (reconstruction period) by means of statistical models calibrated in a more recent period (calibration period), four data sets are needed: the predictor data in the reconstruction period, the predictor data in the calibration period, the predictor data in the calibration period, and the validation data in the

reconstruction period. In this section, the four data sets are briefly described.

2.1 Predictor data in the reconstruction period

The predictor data can be divided into three groups: surface station temperature data, sea level pressure (SLP) data, and upper-air data. We used monthly mean surface air temperature data from 100 sites, mainly from the NASA-GISS (Hansen et al. 1999, 2001), GHCN V2 (Peterson and Vose 1997) and USHCN (Karl et al. 1990) databases (see electronic supplementary material). Mountain sites or high elevation sites were preferred as they are often above the boundary layer or inversion layers and therefore expected to be better predictors for upper-level temperatures than lowland sites. All data taken from NASA-GISS and GHCN V2 (with one exception, see electronic supplement) stem from the homogenised versions of the corresponding data sets. The USHCN data also underwent a multi-step quality control and homogeneity adjustment procedure. We did not perform additional quality tests except for a comparison with corresponding data from NCEP/NCAR reanalysis in order to fill in missing values in the calibration period (see later).

The SLP data were included in the form of the first 20 latitude weighted principal component (PC) time series north of 20°N from the gridded NCAR data set (Trenberth and Paolino 1980). The PC series were calculated for the 1939-1994 period based on monthly anomalies from the mean annual cycle. For missing values, which mainly concern the northernmost latitudes in February to June 1939, a zero anomaly was assumed.

Historical radiosonde and weather flight data are available from 1939 to 1944 (UA39_44, Version 1.0; Brönnimann 2003a). Here we used monthly mean data, version S (after adjustments and combination of neighbouring series). The target precision of the monthly upper-air data is given as ± 1.6 °C for temperature and ± 30 -80 gpm (increasing with height) for GPH. Quantitative assessments could only be performed for sufficiently long series, but the results suggest that most of them easily meet these criteria (see Brönnimann 2003a for details). The possible bias should be less than ± 0.75 °C for temperature and less than ± 15 -30 gpm for GPH. Several records were not used for the reconstructions in order to have an independent data set for the validation (see Sect. 2.4).

In total 843 variables were used as predictors. The location of all series (surface and upper-air) is shown in Fig. 1a, details can be found in the electronic supplementary material. The surface series cover the extratropical NH well except for the Pacific Ocean. In contrast, the upper-air series are more concentrated on Europe and North America. Note that many of the upper-air series cover only a fraction of the period or have long data gaps. Hence, the data availability for any given month is much more limited (see Fig. 4). Figure 1b shows the number of non-missing variables for each month from 1939 to 1947. Most of the surface series contain relatively few missing values, whereas the availability of upper-air data fluctuates more strongly. At most 400 out of 723 upper-air variables are available at any one time, and the availability has spatial and temporal patterns. North America and Europe are relatively well covered compared to Asia, where upper-air information is sparse and no data at all are currently available after June 1941. The subtropical regions are also underrepresented and the Pacific Ocean is represented only by one surface and one upper-air station, the latter ending in June 1942. As to the vertical distribution, the lower and middle troposphere are well represented in most upper-air series. Upper tropospheric data (300 and 400 hPa) are less frequent over the former Soviet Union than over Western Europe and North America and stratospheric data (200 and 100 hPa) are mostly confined to Europe and North Africa (see Brönnimann 2003b).

Despite the inevitably unequal distribution of predictor series, we expect that they provide the principal information on the large-scale circulation in the northern extratropics in the 1939-1944 period. From 1945 to 1947 only surface series are currently available and less reliable reconstructions are expected.

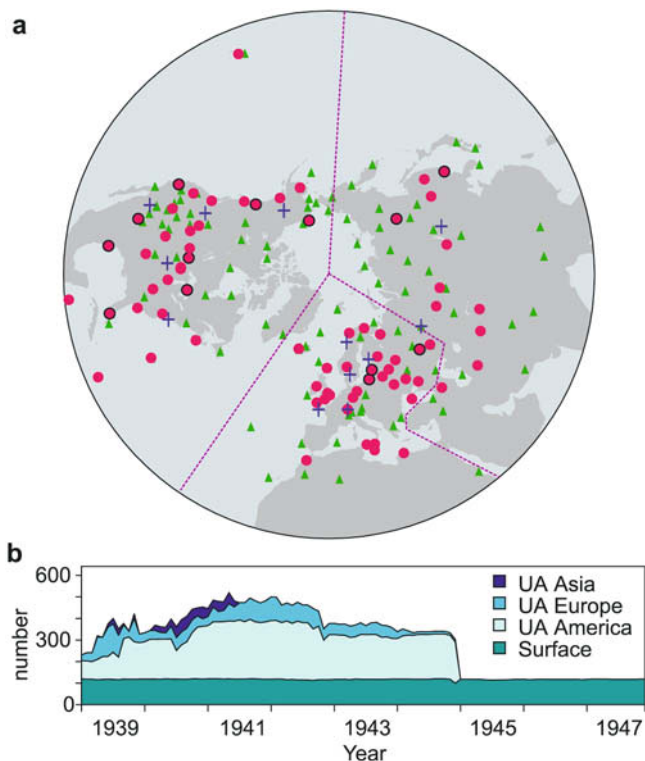


Fig. 1 **a** Map of the predictor variables used for the reconstructions. *Red solid circles* and *blue crosses* denote upper-air (UA) series used for the reconstruction and validation, respectively, *green triangles* denote meteorological stations at the Earth's surface. *Black circles* denote upper-air series used in validation experiment 2. The *dashed lines* group the series by region. **b** Number of non-missing predictor variables, 1939–1947. Upper-air data are separated by geographic region

2.2 Predictor data in the calibration period

The reconstruction approach is based on the assumption that the model is calibrated using different portions of the same predictor data as used for the reconstruction. The SLP-PCs as well as most of the surface station temperature series continue into the present. Several temperature series from the former Soviet Union and China, however, end in the early 1990s. In view of the sparseness of historical upper-air data in this region, we considered these series especially valuable and accordingly defined the calibration period as 1948 to 1994. Missing predictor data in the calibration period should be avoided. Therefore, only data series with at most 10% missing values within this period were chosen. The missing values were replaced after subtracting the long term mean annual cycle and standardising by using corresponding data (i.e. standardised anomalies) from the NCEP/NCAR reanalysis data set (Kalnay et al. 1996; Kistler et al. 2001) at 925 hPa, interpolated to the station location. This is justified by the high correlation (median 0.85) found between the standardised anomalies from both sources.

The historical upper-air data cover only the 1939–1944 period and need to be supplemented with other data in the calibration period. Using CARDS radiosonde data would be possible, but poses several problems. CARDS data might not be consistent with historical data from the same location because of changes in platforms (aircraft, radiosonde), instruments, observation times and frequencies, and correction procedures. In addition, using the same criteria as for surface stations (max. 10% missing values), CARDS data are available for less than half of the locations. We preferred to use corresponding data from the NCEP/NCAR Reanalysis, interpolated to the station location. The historical upper-air data are known to be consistent with NCEP/NCAR Reanalysis data

(Brönnimann 2003a), and there are no missing values. We assume that the reanalysis data are close to the observed values at locations where observations were made and that they are a reasonable estimation elsewhere.

2.3 Predictand data

As predictand data we chose the NCEP/NCAR Reanalysis data set for the 1948–1994 period. No other data set provides temperature and geopotential height data in the troposphere and lower stratosphere during this period. We used monthly mean fields of temperature and GPH at the 700, 500, 300, and 100 hPa levels (hereafter termed T_{700} , Z_{500} etc.) from 20° to 90°N. The data were subsampled to a 5° × 5° grid and the number of longitudes was further reduced at high latitudes (36 at 70°N and 75°N, 18, 9, and 1 at 80°N, 85°N, and 90°N, respectively).

There are several known shortcomings of the NCEP/NCAR Reanalysis data set (see e.g. Santer et al. 1999; Randel et al. 2000 and references therein). The inclusion of satellite data after 1978 and a related processing error led to a step change in temperatures, mostly affecting the lower stratosphere over the tropics. Also, during the earliest period (1948–1957) observation times were different and the amount of upper-air information was smaller and presumably of lower quality. Fewer upper-air observations means that the data product is closer to a model result rather than to an analysis. In addition, there are various smaller problems affecting different parts of the data (see <http://www.cdc.noaa.gov/cdc/reanalysis/problems.shtml>). Hence, the data must be considered less reliable in the early years (Kistler et al. 2001) and inhomogeneities must be expected around 1978 and possibly 1958. In addition to artificial shifts, the real trend in the data due to climate change and stratospheric ozone depletion is also of concern. Despite these problems, which inevitably will have an adverse effect on the reconstruction quality, we think that there are good reasons for using the NCEP/NCAR Reanalysis data as predictand. First, the errors are expected to be smaller in the extratropical NH than in the tropics or in the Southern Hemisphere. Second, we do not use the data for a trend analysis. Our reconstruction approach uses the spatial information and the month-to-month variability, which are expected to be less affected by artificial inhomogeneities. Similarly, the real trend in the data matters only insofar as it leads to a change in the predictor–predictand relation. This might well be the case, but the effect is thought to be small relative to the month-to-month variability. Third, sensitivity experiments can be conducted in order to assess the effect of these inhomogeneities on the reconstructions. Fourth, the final reconstructions can be validated using independent upper-air data.

As we will show in Sect. 4, an acceptable product can be obtained despite these uncertainties in the NCEP/NCAR Reanalysis data. To use only a fraction of the NCEP/NCAR Reanalysis data set also does not provide a solution as the length of the calibration period is important for obtaining good reconstructions.

2.4 Validation data

For the validation of the reconstructed fields we used several historical upper-air series that were not used in the reconstruction (see Sect. 2.1, Fig. 1a, and electronic supplementary material). The series were selected according to five criteria. First, obviously, they must include the reconstructed variables (some series are given on different levels or are outside the reconstructed area). Second, in the context of the quality assessment (described in Brönnimann 2003a), several series were corrected based on information from the Earth's surface. In order to maintain independence, we did not use these series for validation. Third, the quality is not known equally well for all series, and those with a well-known quality were preferred (Brönnimann 2003a). Fourth, the series must have sufficient data at high levels and must have at least five available values. Fifth, all major regions and time periods must be represented. About 13% of

all data were selected in this way and were retained for validation of the final reconstructions. A second validation experiment was performed in which an entirely different sample, comprising 22.5% of the data, was retained (see Sect. 4.4).

Partly independent data sets for validation can also be obtained by splitting the calibration period into two subperiods, one of which is used for calibration and one for reconstruction (see later).

3 Reconstruction method

3.1 Standardising and weighting scheme

All variables in both the predictand and predictor data set were expressed as standardised anomalies with respect to the 1948–1994 mean annual cycle. Thereafter, a weight field was calculated. In the predictand data set, equal weight was attributed to each field (combination of variable and level, e.g. Z_{300}) and each grid cell was weighted by its area. Similarly, in the predictor data, we assigned each series to a level (surface, 850–700 hPa, 600–400 hPa, 300–100 hPa), grid cell (using the predictand grid), and variable (temperature, GPH) and attributed equal weight to each field (combination of variable and level). Within the fields we weighted each series according to the area of the grid cell and the number of variables in the cell or, in the case of the SLP-PC time series, according to the variance. A different weight file was produced for each combination of non-missing predictor variables in the reconstruction period (practically each month). Because in the PC analysis (see later) the weights pertain to the covariances, we multiplied the predictors and predictands by the square root of the corresponding weight. A sensitivity analysis was also performed with unweighted predictors.

In the 1939–1944 period, the data from the Earth's surface constitute around 30% of the non-missing predictor variables (Fig. 1b). By construction, they contribute 25% of the total weight. Accordingly, the upper-air data comprise 70% of the variables and contribute 75% of the weight. The largest individual weight (mostly > 10) was always attributed to the first SLP-PC, whereas the weight of the most important upper-level series was between 2.5 and 5 (i.e. about the same weight as the fifth SLP-PC).

3.2 Forming statistical models

For each combination of non-missing predictor variables in the reconstruction period, a separate statistical model was formed. Because the relationship between the predictor variables and the predictands is likely to change with season (Schmutz et al. 2001; Luterbacher et al. 2002), we selected the corresponding calendar month as well as the two neighbouring months for calibrating the models. For instance, for reconstructing the fields for January 1939, we used only data from the months of December, January, and February in the calibration period. Hence, 141 months were available for calibrating each model.

The reconstructions were performed as described previously (Luterbacher et al. 2002). In the calibration period, PC analyses were performed separately on the predictor data as well as on the predictand data. Only the first few PCs were kept, retaining a certain fraction of the variance (sensitivity analyses were performed using different fractions). Each predictand PC time series was then expressed as a function of all predictor PC time series using multiple regression models. The predictor PC time series in the reconstruction period were then calculated as a function of the predictor variables in the reconstruction period and the predictor PC scores obtained in the calibration period. The predictand PC time series in the reconstruction period were calculated from the predictor PC time series using the regression models estimated in the calibration period. The anomaly fields in the reconstruction period were then obtained as a linear combination of the predictor PC time series and the predictand PC scores obtained in the

calibration period. Weighting and standardising were reversed and the long-term mean annual cycle added. Finally, the fields were regridded to a $5^\circ \times 5^\circ$ grid using linear interpolation.

3.3 Sensitivity analyses and validation

Sensitivity experiments were performed addressing the fraction of explained variance in the predictor data (80, 85, 90, or 95%) and in the predictand data (90 or 95%), the weighting scheme for the predictors (weighting/no weighting) as well as the choice whether all eight fields (four temperature and four GPH fields) should be reconstructed at once, individually, or in subgroups. Table 1 gives an overview of the most important experiments. The results of each experiment were evaluated using two split-sample validations (SSVs, see later). This led to the choice of an appropriate configuration for the final reconstructions of the 1939–1947 period. These final reconstructions were then validated using the retained upper-air data from the 1939–1944 period. A second validation experiment of the final configuration was performed using a different sample of historical upper-air data for validation (see Fig. 1, Table 1, and electronic supplementary material).

The SSV is a special case of cross-validation. It is a repetition of the entire reconstruction procedure, subdividing the calibration period into a calibration and a validation period. For the first of our SSVs we used the first 31 years (1948–1978) for calibration and the last 16 years (1979–1994) for validation, for the second SSV we used the last 31 years (1964–1994) for calibration and the first 16 years (1948–1963) for validation. Note that the two 16-year periods correspond well with the supposed inhomogeneities in the NCEP/NCAR data set, especially the supposed break around 1978. Hence, these inhomogeneities are expected to have a maximum negative effect on the outcome of the validation, which provides a restrictive test of the quality of the reconstructions. However, the SSV approach does not account for errors in the upper-air predictor data, as they stem from the same data set as the predictand data (see Sect. 4.4).

Several error statistics were used to assess the model performance, the most important being the reduction of error (RE , Lorenz 1956; Cook et al. 1994), defined as

$$RE = 1 - \frac{\sum_t (x_{rec} - x_{obs})^2}{\sum_t (x_{null} - x_{obs})^2} \quad (1)$$

where t is time, x_{rec} is the reconstructed value, x_{obs} is the observed value and x_{null} is a null hypothesis or no-knowledge prediction (e.g. constant, climatology, random, persistence, etc.). In our case, since we reconstruct anomalies, the chosen null hypothesis is a zero anomaly (i.e. the mean annual cycle in the calibration period). Values of RE can be between $-\infty$ and 1 (perfect reconstructions). An RE of 0 indicates that the reconstruction is not better than the null hypothesis. A random number with correct variance would yield an RE of -1 . $RE > 0$ determined in an independent period is normally considered an indication that the model has predictive skill. For values greater than around 0.2, RE is close to the coefficient of determination R^2 . In the following, we consider a reconstruction useful if $RE > 0.2$, which roughly corresponds to $R^2 = 0.2$ to 0.25 or $r = 0.45$ to 0.5.

As each statistical model pertains to a specific calendar month, there are 16 opportunities to validate each model in a 16-year period. Hence, Eq. (1) sums over 16 time steps. For each month in the reconstruction period (i.e. each model), a spatial field of RE is obtained from each SSV. For the comparison of different sensitivity experiments, it is more practical to condense the information into one number. Because the distribution of RE over time at one grid point is typically skewed, we used both the median and the mean value at each grid point, denoted RE_{median} and RE_{mean} , to characterise the reconstruction quality. RE_{median} is usually higher than RE_{mean} . A single number for each experiment and field (combination of variable and level) can be obtained by an area-weighted average of RE_{median} or RE_{mean} . We used the average from 25° to 90°N as RE at 20°N sometimes showed a different behaviour

Table 1 Overview and results of selected sensitivity experiments

Run	R^2_x	R^2_y	W_x	Mode	1939-1944					1945-1947							
					Z_{700}	Z_{500}	Z_{300}	Z_{100}	T_{700}	T_{500}	T_{300}	T_{100}	Z_{700}	Z_{500}	Z_{300}	T_{700}	T_{500}
Area-weighted average of RE_{mean} from SSVs																	
A	85	95	Yes	1	0.44	0.43	0.40	0.31	0.30	0.34	0.18	0.13	0.47	0.44	0.40	0.32	0.29
B	90	95	Yes	1	0.49	0.48	0.44	0.35	0.33	0.32	0.21	0.14	0.49	0.46	0.40	0.33	0.28
C	95	95	Yes	1	0.54	0.52	0.47	0.37	0.34	0.23	0.12	0.50	0.47	0.40	0.31	0.26	
D	95	95	No	1	0.50	0.44	0.46	0.35	0.36	0.34	0.21	0.13	0.43	0.41	0.36	0.30	0.23
E	85	95	Yes	2	0.44	0.42	0.40	0.31	0.30	0.29	0.18	0.13	0.47	0.44	0.40	0.32	0.29
F	90	95	Yes	2	0.50	0.48	0.44	0.35	0.31	0.32	0.22	0.14	0.49	0.46	0.41	0.33	0.28
G	95	95	Yes	2	0.54	0.52	0.47	0.36	0.34	0.33	0.23	0.12	0.50	0.47	0.40	0.31	0.26
H	95	95	Yes	3	0.54	0.52	0.47	0.37	0.34	0.33	0.23	0.12	0.50	0.47	0.40	0.31	0.26
Area-weighted average of RE_{median} from SSVs																	
A	85	95	Yes	1	0.48	0.47	0.45	0.37	0.35	0.34	0.24	0.18	0.52	0.49	0.45	0.37	0.33
B	90	95	Yes	1	0.54	0.53	0.51	0.41	0.39	0.38	0.27	0.20	0.54	0.51	0.47	0.39	0.34
C	95	95	Yes	1	0.59	0.58	0.54	0.43	0.42	0.40	0.30	0.20	0.56	0.53	0.47	0.38	0.33
D	95	95	No	1	0.55	0.55	0.52	0.41	0.41	0.39	0.28	0.20	0.51	0.50	0.45	0.38	0.32
E	85	95	Yes	2	0.48	0.48	0.46	0.36	0.35	0.34	0.24	0.18	0.52	0.49	0.45	0.37	0.33
F	90	95	Yes	2	0.54	0.53	0.51	0.41	0.37	0.38	0.28	0.20	0.54	0.51	0.47	0.38	0.34
G	95	95	Yes	2	0.60	0.58	0.54	0.42	0.42	0.40	0.30	0.20	0.56	0.53	0.47	0.38	0.33
H	95	95	Yes	3	0.59	0.58	0.54	0.43	0.41	0.40	0.29	0.20	0.56	0.53	0.47	0.38	0.33
RE using historical upper-air data																	
H + F					0.78	0.77	0.72	0.25	0.69	0.63	0.52	0.48					
RE using historical upper-air data (experiment 2)																	
H + F					0.63	0.65	0.63	0.57	0.66	0.61	0.42	0.38					

R^2_x and R^2_y are the variances retained on the predictor and predictand side, respectively. W_x indicates whether or not the predictor data were weighted. Modes 1, 2, and 3 indicate respectively whether all fields were reconstructed together, individually, or all together except 100 hPa temperature. Columns Z_{700} through T_{100} (for 1939-1944) and Z_{700} through T_{500} (1945-1947) give area-weighted averages (25° - 90° N) of RE_{mean} and RE_{median} , averaged from both SSVs. The bottom two rows give RE for the final reconstructions in the 1939-1944 period calculated from retained historical upper-air data and results from the second validation experiment. Numbers in bold indicate the combination of models chosen for the final reconstructions

than elsewhere. In most cases only the averaged results from both SSVs are shown. Because of the different predictors used, the 1939-1944 and 1945-1947 periods were studied separately. Only Z_{700} , Z_{500} , Z_{300} , T_{700} , and T_{500} were analysed for the latter period.

Other error measures used for the model evaluation include the bias and the root mean squared error (RMSE). Approximate 95% confidence intervals can be defined as 2 RMSE, averaged from both SSVs. Note that all of these measures can also be calculated for derived properties, such as the anomaly of a grid point value from the entire field at each time step or the anomaly from the reconstruction-period mean at each grid point, or any other variable derived from the data. This allows assessment of how well the spatial patterns or the month-to-month variability are reconstructed, regardless of the bias.

4 Results and discussion

4.1 Sensitivity experiments

Table 1 gives the area-weighted average of RE_{mean} and RE_{median} averaged from both SSVs for the most important sensitivity experiments. Results are listed separately for the 1939-1944 and 1945-1947 periods. They show that generally good reconstructions can be expected for GPH at all levels and for T_{700} and T_{500} . Results are somewhat worse for T_{300} and clearly worse for T_{100} . The detailed validation results for the experiment chosen for the final reconstructions are discussed in Sect. 4.3. In this section, only the differences between the sensitivity experiments are discussed.

These differences are surprisingly small. To start with, the weighting of the predictors has positive effects on the results for GPH, especially at lower levels, whereas the effect is very small for temperature and for upper levels. The effect on RE_{mean} is similar to that on RE_{median} . Since the weighting is followed by a variable reduction, the weighting problem is related to the problem of the choice of the fraction of explained variance to be retained in the PC analysis on the predictor side. Preliminary sensitivity experiments (using no weighting; not shown) clearly imply that the predictand variance is best set to 95%, in agreement with other studies (Schmutz et al. 2001). Experiments A, B, and C (reconstructing all fields together) and E, F, G (reconstructing fields separately) show the effect of using 85, 90, or 95% of the variance on the predictor side (experiments with 80% were clearly worse and are not shown). For most fields, best results are found when 95% is chosen, for some fields 90% and 95% show equal results and in few cases 90% is better. It is worth noting that T_{100} shows a different behaviour than the other fields, which could be due to the fact that it has not much of its variability in common with the other fields. It might therefore be preferable to reconstruct this field separately. For the other fields, the difference between reconstructing all fields together or separately is very small. Additional experiments using suitable subgroups of the fields also gave the same results.

In order to obtain the highest possible quality, we chose the reconstructions of T_{100} from experiment F,

which gives the best results for both RE_{median} and RE_{mean} (Table 1). All other fields were reconstructed together (experiment H, using 95% of the predictor variance). This procedure is thought to give the best overall results in the 1939-1944 period. Note that the differences between most of the experiments are small and other choices would lead to very similar results.

4.2 Effect of a bias or inhomogeneity in the NCEP/NCAR reanalysis data

RE accounts for both the bias and the variance of the reconstructions. The contribution of the bias is normally small, but it can nevertheless have some effect at upper levels. Biases are found in the SSVs at the 100 hPa level, especially for temperature. They can be caused by real trends (due to climate change) or inhomogeneities in the data, as discussed previously. For the purpose of this study, it does not matter what causes the bias. It is interesting to note, however, that for T_{100} results are worse in the first SSV than in the second, which could be due to an inhomogeneity around 1978. For most other fields, for which the effect of the latter inhomogeneity is supposedly small, the opposite is the case, which could be due to a generally lower data quality in the 1948-1963 period, possibly due to sparse radiosonde data. These results are consistent with a negative effect of the quality of the predictand data on the reconstructions. Note, however, that these effects are accounted for in the SSVs and all derived skill measures. Moreover, independent data can be used to assess the quality of the actual reconstructions for the 1939-1944 period (see later).

Because of the bias in the SSVs, the spatial variability and the month-to-month variability are reproduced more reliably than RE in Table 1 implies. For instance, the area-averaged RE_{median} for T_{100} rises from 0.20 to 0.23 if the mean value of the field is subtracted at each time step (area-averaged RE_{mean} increases from 0.14 to 0.17). Hence, the usefulness of the reconstructions depends on the application. Spatial interpretations are more robust than arguments based on absolute values.

4.3 Validation results for the chosen reconstruction approach

In this section we discuss the results from the SSVs for the approach chosen for the final reconstruction. As can be seen in Table 1 (numbers in bold), all fields have an area-averaged $RE_{median} \geq 0.2$ (averaged from both SSVs). RE_{mean} is lower, between 0.14 and 0.54. While these aggregated numbers are useful for the comparison of sensitivity experiments, the spatial and temporal details in RE deserve more attention. Figure 2 shows fields of RE_{median} for the 1939-1944 period (averaged from both SSVs). As expected, the skill is best over areas with a high density of predictor variables, i.e. North America and Europe, and worse at polar and

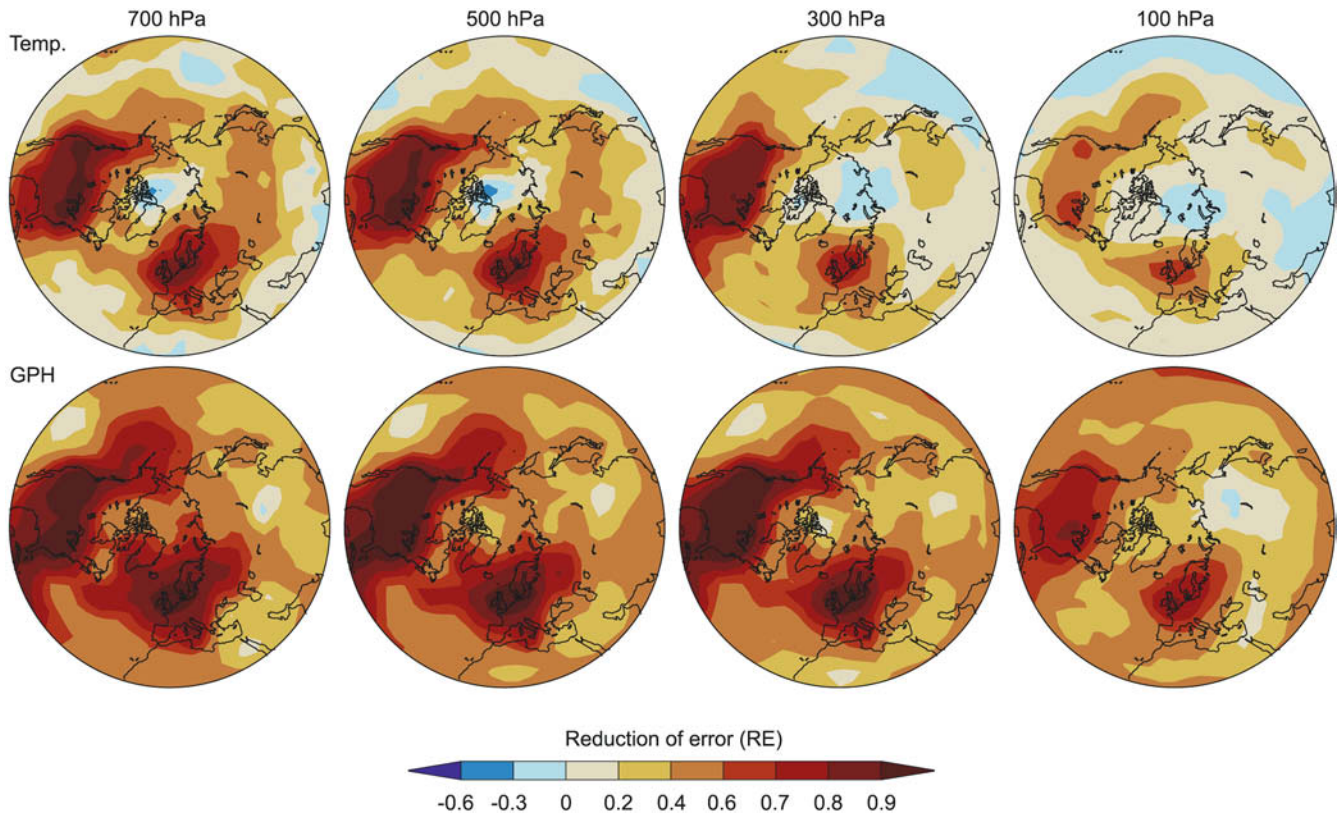


Fig. 2 Maps of RE_{median} for (top) temperature and (bottom) GPH at different levels for the 1939-1944 period

subtropical latitudes and over Asia. It is interesting to note that RE_{median} is reasonably high also over large parts of the North Pacific despite the fact that there are almost no predictors from this region. The reconstruction skill is even acceptable over the North Pole for GPH, but not for temperature.

As noted in Sect. 4.1 there are large differences between the fields. RE_{median} values are much higher for GPH than for temperature and higher for levels close to the surface (where the availability of predictor data is better) than for upper-levels. Considering a threshold of 0.2, acceptable reconstructions can be expected most of the time for GPH at all levels and for T_{700} and T_{500} . Results are worse for T_{300} and clearly worse for T_{100} . In the latter case, reasonable reconstructions can generally be expected only for the latitude band of around 35° to 65°N , with the exception of the Asian sector. Note, again, that spatial patterns are better reproduced than absolute values and hence some of the reconstructions might be useful for certain applications despite a relatively poor overall quality. Note also that the variability between the statistical models (i.e. the month-to-month variability of RE) is considerable (see Sect. 4.5. for examples).

Figure 3 shows time series of area-weighted averages of RE (from both SSVs) from 1939 to 1947. There is a clear annual cycle in RE for all fields. The minimum RE values for Z_{100} are reached in spring, for all other fields in summer. Hence, reconstructions are expected to be better in winter than in summer. This is due to the

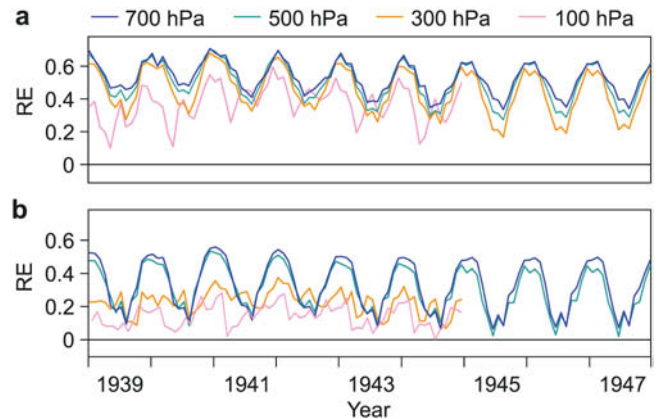


Fig. 3 Time series of area-weighted averages (25° - 90°N) of RE for **a** GPH and **b** temperature at different levels from 1939 to 1947

dominance of large-scale circulation patterns in winter such as the NAO/AO or the Pacific North American pattern (e.g. Barnston and Livezey 1987), which are more easily captured even with a limited number of predictors. Within the 1939-1944 period, the skill for Z_{100} and T_{100} is best in 1941 and 1942, when the number of available predictor variables is highest (see Fig. 1b). In the 1945-1947 period the reconstruction skill is lower than in the 1939-1944 period especially during the summer months. This is due to the vertical coupling of atmospheric circulation in summer, which is weaker and of a smaller spatial scale than in winter (e.g. Brönni-

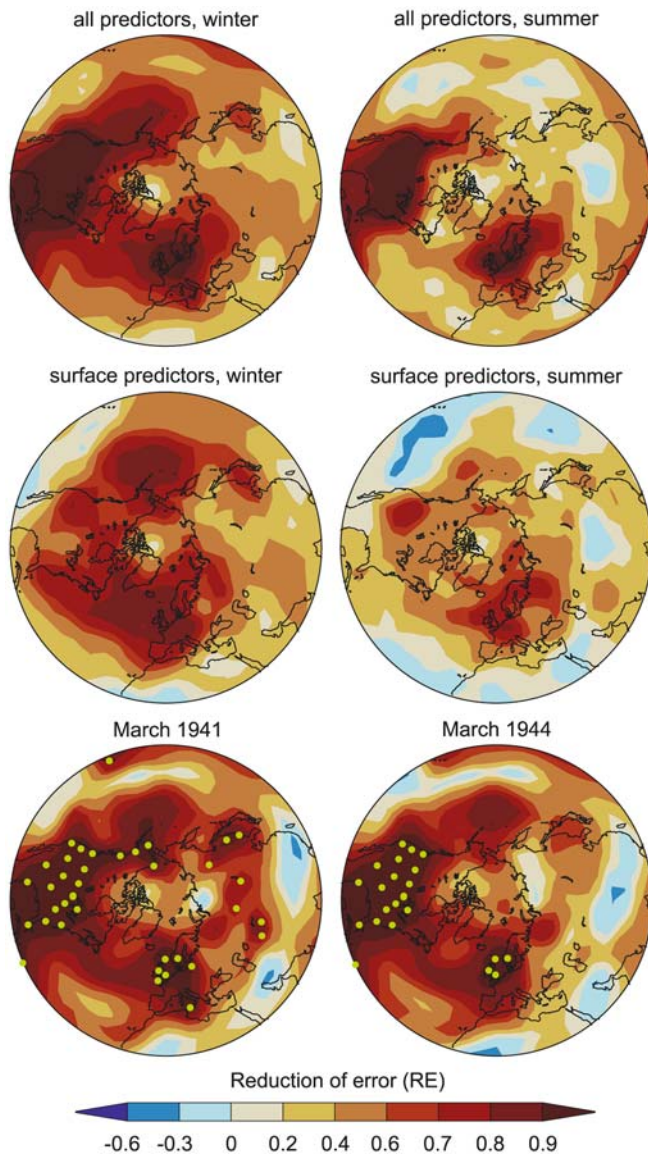


Fig. 4 Maps of RE_{median} for 300 hPa GPH for different sets of predictor variables for winter (November to April) and summer (May to October). Green dots in the bottom row give the location of upper-air predictors used

mann et al. 2000). Nevertheless, except for temperature in summer, acceptable reconstructions were obtained also in this period, using only predictors from the Earth's surface.

The differences in reconstruction skill between using only surface predictors and using all predictors yields information on the relative importance of the upper-air series, which can be important for guiding future re-evaluation work. Figure 4 shows a more detailed comparison on the example of Z_{300} . As noted, upper-air data are especially important in summer (May to October), whereas surface data alone provide good reconstructions in the cold season (November to April). In both seasons, the differences are largest over the continental areas, i.e. in North America between the Appalachians and the

Rocky Mountains and in Russia. In contrast, upper-air data do not contribute to better reconstructions over the Atlantic and Pacific north of 40°N . The bottom row of Fig. 4 shows a comparison of two statistical models with upper-air data, one of which includes data in Hawaii, Alaska, and Siberia, the other not. Surprisingly, upper-air data from the Pacific coast have only a small influence on the reconstruction skill over the North Pacific. They slightly affect the skill over the Arctic and their greatest effect is over the continents. For future re-evaluation efforts this means that data from Asia, the Arctic, and the subtropics would be most beneficial.

4.4 Validation of the final reconstructions

After having discussed the results from the SSVs, we now turn to the validation of the actual reconstructions. There are important differences between these two “validations”: On the one hand, the SSVs provide a pessimistic estimation of the skill as they use only 31 years of data for calibrating the models, whereas the final reconstruction uses 47 years and hence is considered to provide a more accurate model and better reconstructions. On the other hand, in the two 16-year validation periods of the SSVs, the upper-air predictors were taken from the NCEP/NCAR Reanalysis data set, which was at the same time used for the validation. Hence, in contrast to the 1939–1944 period, the upper-air predictors in the SSV approach have no errors and therefore the SSVs provide an optimistic estimation.

The only way of validating the final reconstructions is to use independent data. Several historical upper-air series were retained for this purpose. From these data, RE can be calculated in the same way as explained above except that for this purpose all observations were pooled. Results are shown in Table 1 and Fig. 5. Generally, higher RE values were found than in the SSVs. This is expected as the series used for the validation are mostly in areas with good data coverage and hence high reconstruction skill. If the spatial representativity is considered, then the RE values obtained from historical upper-air data seem to be lower than the ones obtained from SSVs. This is also not surprising, as all errors in the validation data are reflected as reconstruction errors. In fact, simple experiments for 300 hPa temperature at a typical midlatitude site assuming realistic errors in the validation data (90% of the monthly mean values within $\pm 1^{\circ}\text{C}$, mean offset of 0.5°C) reveal that even if the reconstructions were perfect, RE is expected to be only around 0.77. Hence, the RE values displayed in Table 1 can be considered as excellent.

In order to address the effect of the sampling bias (validation series in well-sampled areas), we performed a second validation experiment for which an entirely different, much larger sample of historical upper-air data was retained (note that these are not the final reconstructions). We selected sites at the edge of or outside the well-sampled regions whenever possible (see Fig. 1).

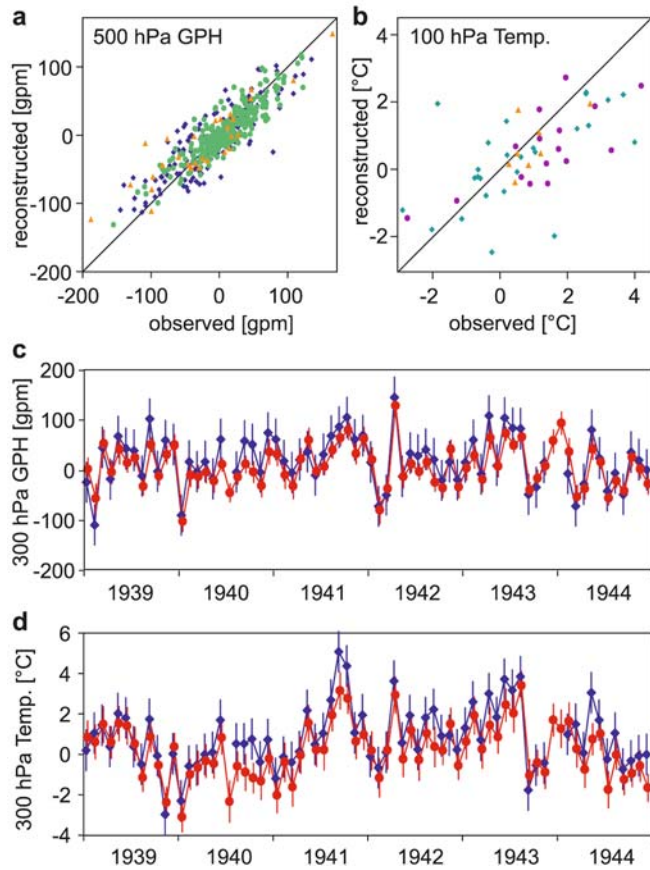


Fig. 5 **a** Observed and reconstructed anomalies of 500 hPa GPH for sites in North America (green dots), Europe and Africa (blue diamonds), and Asia (orange triangles) in 1939-1944. **b** Same as **a** for 100 hPa temperature at Brest (green diamonds), Freiburg (purple circles), and Bodø (orange triangles) in 1939-1944. **c** Time series of observed (blue diamonds) and reconstructed (red circles) anomalies of 300 hPa GPH at Omaha. Error bars give the assumed uncertainty of the observations and the 95% confidence intervals for the reconstructions, respectively. **d** Same as **c** for 300 hPa temperature. All anomalies are with respect to the calibration period (1948-1994)

Without these data, which are known to be important predictors, and with a much lower number of variables, we expect clearly worse reconstructions and at the same time validate these reconstructions in more remote regions. Nevertheless, the results (Table 1) reveal a good skill for all fields.

In the following, more details on the validation of the final reconstructions are given. Figure 5 shows the observed and reconstructed anomalies for Z_{500} and T_{100} with respect to the calibration period (1948-1994). Z_{500} is very well reconstructed. The European and Asian sites show a slightly larger scatter than the American sites, but the agreement is excellent for all groups. For T_{100} only three series from Europe could be used for validation: Brest and Freiburg in Western and Central Europe, respectively, and Bodø in the Norwegian Arctic. As expected the scatter is larger and there are two or three outliers; all of them contained within series from Brest. Comparisons with neighbouring stations suggest

that at least the outlier in the upper left part of the plot is due to erroneous validation data from Brest. Despite these outliers, the agreement is acceptable.

The lower part of Fig. 5 shows observed and reconstructed time series for Z_{300} and T_{300} at Omaha, Nebraska, from where a nearly complete, independent record is available for validation (Brönnimann 2003b). The reconstructions follow the observed values very closely; the explained variance is larger than 80% in both cases. RE values are 0.78 and 0.72 for Z_{300} and T_{300} respectively, which is lower than what is expected from SSVs (see Fig. 2). The reason for this might be errors in the observed data from Omaha, mainly in the form of a bias. The mean differences amount to 14 gpm for GPH and 0.65 °C for temperature, which is large but within the uncertainty of the historical upper-air data. The error bars show an assumed uncertainty in the observations of ± 1 °C and ± 40 gpm as well as the 95% confidence interval of the reconstructions determined from both SSVs. The error bars overlap in almost all cases. Accounting for these uncertainties, the agreement between reconstructions and observations is excellent. The main signal is picked up in both cases.

In all, the validation with independent data suggests good reconstructions over the Northern midlatitudes for all fields and confirms the results obtained from the SSVs (Sect. 4.3). It is planned to digitise more historical upper-air series, which will provide further opportunities to validate the reconstructions.

4.5 An analysis of the reconstructed fields

In order to demonstrate the usefulness of the reconstructions to study the large-scale circulation in the early 1940s, an analysis of selected fields is presented in this section. We show anomaly fields (with respect to the 1961-1990 mean annual cycle) for August 1940, February 1941, and January 1942. These months were chosen because upper-air data, surface temperature, and SLP fields for the same months were already presented in Brönnimann (2003b, Fig. 9), allowing a direct comparison. Figure 6 shows T_{500} , Z_{500} , T_{100} , and Z_{100} for these three months. Areas with $RE < 0.2$ are displayed as light shaded areas. It is clear that the reconstructed fields for August 1940 are less reliable than the ones for the winter months. This holds especially for temperature. Even for T_{500} useful reconstructions can only be expected over the continental midlatitudes. In contrast, for Z_{100} in winter there are only small areas with $RE < 0.2$, pointing to reliable reconstructions even in the lower stratosphere.

Several of the tropospheric features already discussed in Brönnimann (2003b) are also found in the reconstructions such as the intensified Aleutian low and the strong temperature contrasts in the two winter months. It is interesting to address the circulation anomalies at 100 hPa. Both winter months were characterised by positive GPH anomalies over the polar region, pointing

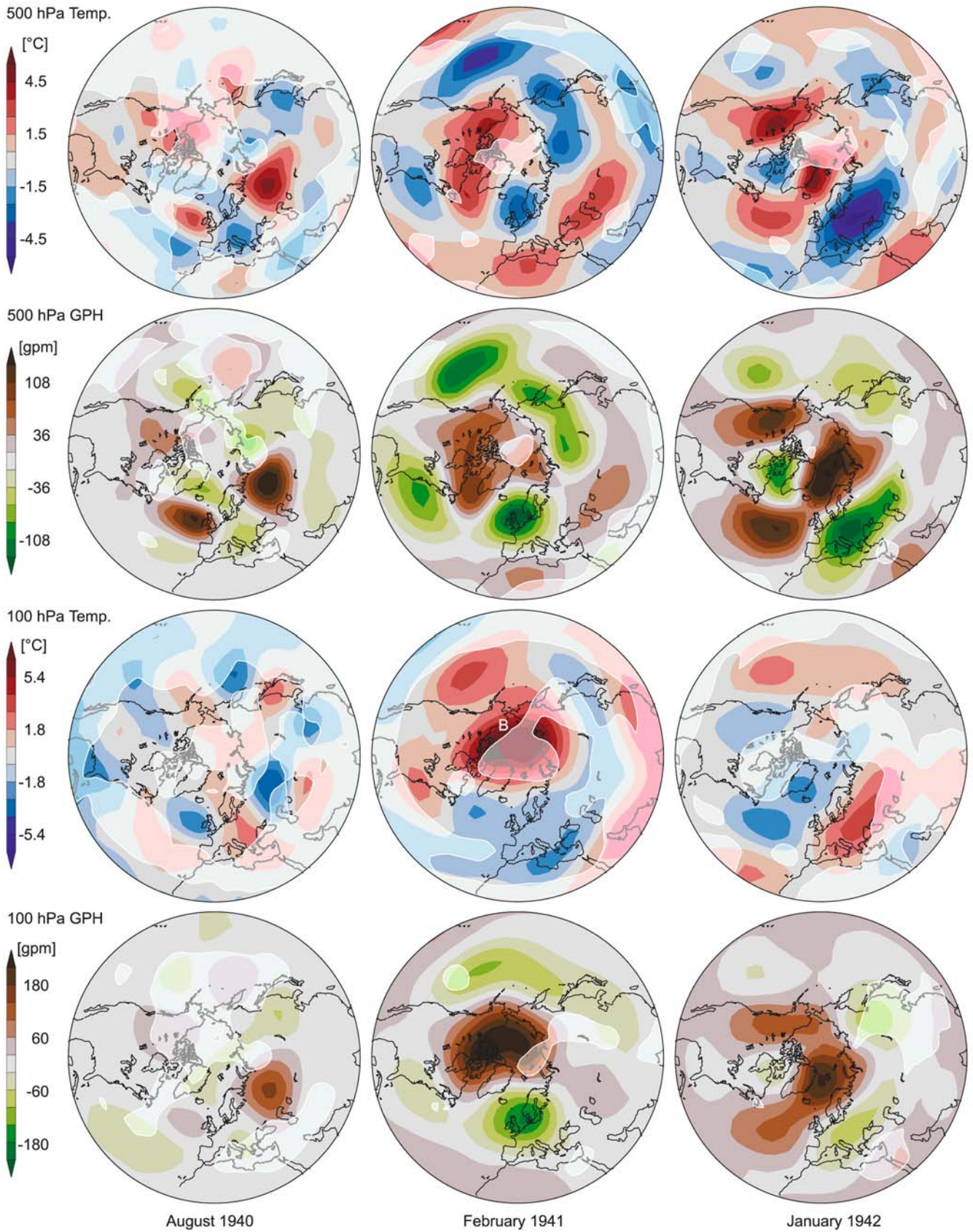




Fig. 6 Reconstructed anomaly fields of 500 and 100 hPa GPH and temperature for August 1940, February 1941, and January 1942. Light shading denotes areas with $RE < 0.2$. Anomalies refer to the 1961-1990 period

to a weak or disturbed polar vortex. The absolute fields show very small pressure gradients between the polar and subpolar areas in February 1941 (not shown), whereas in January 1942 the gradients are somewhat stronger and confine, in the monthly mean field, a narrow, elongated vortex stretching from Baffin Island to Yakutsk. The 100 hPa temperatures were very different in the two months. In January 1942, the reconstructed temperatures show anomalies at midlatitudes, but were close to normal in the polar region. In contrast, in February 1941 positive anomalies of more than 8 °C are estimated over the Arctic. Unfortunately, in both cases the interpretation is hampered by low RE values (< 0.2) over most of the polar region. In order to test whether the high polar temperatures in February 1941 were real or not, we used radio sounding data from Barrow, Alaska (B in Fig. 5), which is inside the polar region. This record was also used for the reconstructions, but only up to 300 hPa. Data from higher levels were digitised, but not re-evaluated in the same way because the data are given on fixed altitude levels and the interpolation to pressure levels introduces a bias (see Brönnimann 2003b). Nevertheless, they can give some additional information. The data for the 14 km altitude level (ca. 130 hPa, the highest level with sufficient data) show an increase in temperature of 10 °C from January to February 1941. This is even more than in the reconstructions at 100 hPa. Similar comparisons for other sites also reveal a good correspondence between reconstructions and observations. Therefore, we consider the reconstructed field to exhibit the main features of the real field. It is possible that this warm monthly mean anomaly over the polar region was related to a major warming in the middle stratosphere, as similar anomalies in T_{100} in the NCEP/NCAR Reanalysis data almost exclusively occur with major warming events at 30 hPa (data since 1952, see Labitzke and Naujokat 2000) in the same or previous month. Further Arctic radiosonde records will be digitised specifically for this period.

The two winter months were also different in the troposphere and at the surface. The reconstructed 500 hPa GPH field in February 1941 shows positive anomalies over the Canadian Arctic, Greenland, and Svalbard, surrounded by negative anomaly centres. The spatial pattern exhibits a strong similarity with the AO (Thompson and Wallace 1998); the AO index value for this month is -1.87 (NAO index from the Climatic Research Unit: -0.42). The vertical structure of the positive GPH anomaly shows the well-known north-westward tilt with height (see Wanner et al. 2001). A possible stratospheric warming in February 1941 would be consistent with a weak polar vortex and a negative AO index (Hartmann et al. 2000) and could have been

caused by increased upward and poleward propagating planetary wave activity (related to a strong zonal wave number 1), decelerating the zonal flow in the middle stratosphere and inducing anomalous downwelling over the North Pole (Polvani and Waugh 2004; Limpasuvan et al. 2004). However, much more upper-air data and daily upper-level fields are clearly necessary in order to test this hypothesis.

The situation in January 1942 was more complex. The Icelandic low was shifted northwestward to southern Greenland, while weather in Europe was dominated by a strong Scandinavian high at the surface. This pattern, causing extremely low temperatures in large parts of Europe, does not project well onto the AO (index: -0.35) or NAO (+0.75). The pronounced 500 hPa GPH anomaly over the Norwegian Sea extends to 100 hPa, but seems comparably weakened and less tilted. The 100 hPa temperature anomalies at midlatitudes can be explained by the planetary wave structure, which was strongly dominated (in the monthly mean fields) by zonal wave number 2.

All three months presented in Fig. 6 are typical for the circulation in the 1940-1942 period, which was characterised at the surface by a strong Aleutian low, low temperatures over Europe, warm temperatures in Alaska and the Arctic, and predominantly negative AO and NAO indices for three winters in a row. A detailed analysis of the circulation during the early 1940s is beyond the scope of this work and will be presented in a future article. This brief analysis shows that, despite some limitations, the reconstructed fields can provide important information about the large-scale circulation up into the lower stratosphere. It also shows that the reconstruction skill strongly varies in time and space.

5 Conclusions

Monthly mean fields of 700, 500, 300, and 100 hPa GPH and temperature were reconstructed for the extratropical NH, 1939-1947. The reconstructions were based on historical upper-level data and data from the Earth's surface. Validation experiments within the calibration period as well as validations of the final reconstructions using independent upper-air data reveal a satisfying overall quality. However, there are very important differences between the fields, spatial differences, and seasonal and temporal differences that need to be considered when working with the reconstructed fields. Generally, GPH reconstructions are mostly reliable except over eastern Siberia. Temperature fields at 700 and 500 hPa are also reliable in the 1939-1944 period except for the western North Pacific, subtropical Asia and the polar region. Temperatures at higher levels (300 and 100 hPa) should be interpreted with care, although we have demonstrated that they can provide important information and may be useful for certain applications focusing on the spatial patterns rather than on the absolute values. It is important that the reconstructions

are only analysed together with corresponding fields of *RE* or *RMSE*, as their quality varies strongly.

An analysis of the reconstructed fields for selected months (including a possible stratospheric warming event in February 1941) shows that the reconstructions are suitable for studying the large-scale circulation of the northern extratropics. They can be used to address stratosphere-troposphere coupling in the late 1930s and early 1940s.

Finally, we would like to point out that historical upper-air observations may contribute important information to studies of past climate variability. Some of the records that can be found on paper in meteorological archives reach back to the 1920s. Although the quality of the data is largely unknown, and is probably the most important limitation, it could be possible in the future to extend reconstructions further back in time by at least several years.

Acknowledgements S.B. was funded by the Swiss National Science Foundation and the Holderbank-Foundation. J.L. was funded by the Swiss National Competence Center for Research in Climate. Support by the Stiftung Marchese Francesco Medici del Vascello is gratefully acknowledged. The reconstructed fields and corresponding error measures as well as the historical upper-air data and detailed documentation can be downloaded from the website http://sinus.unibe.ch/~broenn/UA39_44/. The authors wish to thank D. Dietrich (Univ. Bern) for statistical advice.

References

- Ambaum MHP, Hoskins BJ (2002) The NAO troposphere-stratosphere connection. *J Clim* 15: 1969-1978
- Barnston AG, Livezey RE (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon Weather Rev* 115: 1083-1126
- Brönnimann S (2003a) Description of the 1939-1944 upper-air data set (UA39_44) Version 1.0. University of Arizona, Tucson, USA, pp 40
- Brönnimann S (2003b) A historical upper-air data set for the 1939-1944 period. *Int J Climatol* 23: 769-791 DOI 10.1002/joc.914
- Brönnimann S, Luterbacher J, Schmutz C, Wanner H, Staehelin J (2000) Variability of total ozone at Arosa, Switzerland, since 1931 related to atmospheric circulation indices. *Geophys Res Lett* 27: 2213-2216
- Christy JR, Spencer RW, Norris WB, Braswell WD, Parker DE (2003) Error estimates of Version 5.0 of MSU/AMSU bulk atmospheric temperatures. *J Atmos Oceanic Tech* 20: 613-629
- Cook ER, Briffa KR, Jones PD (1994) Spatial regression methods in dendroclimatology - a review and comparison of two techniques. *Int J Climatol* 14: 379-402
- Hartmann DL, Wallace JM, Limpasuvan V, Thompson DWJ, Holton JR (2000) Can ozone depletion and greenhouse warming interact to produce rapid climate change? *Proc Natl Acad Sci USA* 97: 1412-1417
- Hansen J, Ruedy R, Sato M, Imhoff M, Lawrence W, Easterling D, Peterson T, Karl T (2001) A closer look at United States and global surface temperature change. *J Geophys Res* 106: 23,947-23,963
- Hansen J, Ruedy R, Glascoe J, Sato M (1999) GISS analysis of surface temperature change. *J Geophys Res* 104: 30997-31022
- Hurrell J, Kushnir Y, Ottensen G, Visbeck M (eds) (2003) The North Atlantic Oscillation. Climatic significance and environmental impact. AGU, Geophysical Monograph Series 134, Washington DC, USA
- Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77: 437-471
- Karl TR, Williams CN Jr, Quinlan FT, Boden TA (1990) United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, pp 389
- Kington JA (1975) The construction of 500-millibar charts for the eastern North Atlantic-European sector from 1781. *Meteorol Mag* 104: 336-340
- Kistler R et al (2001) The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull Am Meteorol Soc* 82: 247-267
- Klein WH, Dai Y (1998) Reconstruction of monthly mean 700-mb heights from surface data by reverse specification. *J Clim* 11: 2136-2146
- Labitzke K, Naujokat B (2000) The lower Arctic stratosphere in winter since 1952. *SPARC Newsletter* 15: 11-14
- Labitzke K et al (2002) The Berlin Stratospheric Data Series, CD from Meteorological Institute, Free University Berlin, Germany
- Lanzante JR, Klein SA, Seidel DJ (2003) Temporal homogenization of monthly radiosonde temperature data. Part I: methodology. *J Clim* 16: 224-240
- Limpasuvan V, Thompson DWJ, Hartmann DL (2004) On the life cycle of Northern Hemisphere stratospheric sudden warming. *J Clim* (in press)
- Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. M.I.T. Statistical Forecasting Project Report 1, Contract AF 19: 604-1566
- Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Reconstruction of sea-level pressure fields over the eastern North Atlantic and Europe back to 1500. *Clim Dyn* 18: 545-561 DOI 10.1007/s00382-001-0196-6
- Mears C, Schabel M, Wentz F (2003) A reanalysis of the MSU Channel 2 tropospheric temperature record. *J Clim* 16:3650-3664
- Parker DE, Gordon M, Cullum DPN, Sexton DMH, Folland CK, Rayner N (1997) A new gridded radiosonde temperature data base and recent temperature trends. *Geophys Res Lett* 24: 1499-1502
- Peterson TC, Vose RS (1997) An overview of the Global Historical Climatology Network temperature data base. *Bull Am Meteorol Soc* 78: 2837-2849
- Polansky BC (2002) Reconstructing 500-hPa height fields over the Northern Hemisphere. Msc Thesis, University of Washington, Seattle, USA, pp 38
- Polvani LM, Waugh DW (2004) Upward wave activity flux as precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *J Clim* (in press)
- Randel WJ, Wu F, Gaffen DJ (2000) Interannual variability of the tropical tropopause from radiosonde data and NCEP reanalysis. *J Geophys Res* 105: 15,509-15,523
- Santer BD, Hnilo JJ, Wigley TML, Boyle JS, Doutriaux C, Fiorino M, Parker DE, Taylor KE (1999) Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J Geophys Res* 104: 6305-6333
- Schmutz C, Gyalistras D, Luterbacher J, Wanner H (2001) Reconstruction of monthly 700, 500 and 300 hPa geopotential height fields in the European and Eastern North Atlantic region for the period 1901-1947. *Clim Res* 18: 181-193
- Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys Res Lett* 25: 1297-1300
- Trenberth KE, Paolino DA (1980) The Northern Hemisphere sea level pressure data set: trends, errors, and discontinuities. *Mon Weather Rev* 108: 855-872
- Wanner H, Brönnimann S, Casty C, Gyalistras D, Luterbacher J, Schmutz C, Stephenson D, Xoplaki E (2001) North Atlantic Oscillation - concept and studies. *Surv Geophys* 22: 321-381