TAC-0/271
For Author's Correction Only

Theor. Appl. Climatol. 000, 1–20 (2007)
DOI 10.1007/s00704-006-0271-3
Printed in The Netherlands

Theoretical and Applied Climatology

[1] Key Laboratory of Environmental Change and Natural Disaster, College of Resources Science and Technology, Beijing Normal University, Beijing, China
[2] Bjerknes Center for Climate Research/Nansen Environmental and Remote Sensing Center, University of Bergen, Norway
[3] Nansen-Zhu International Research Center, IAP, Beijing, China

Reconstruction of Northern Hemisphere 500 hPa geopotential heights back to the late 19th century

D.-Y. Gong[1,2], **H. Drange**[2,3], and **Y.-Q. Gao**[2,3]

With 13 Figures

Received December 12, 2005; revised March 31, 2006; accepted July 27, 2006
Published online • •, 2006

Summary

In this study the authors have developed a statistical method and have reconstructed Northern Hemisphere 500 hPa heights back to the late 19th century using one temperature and three sea level pressure (SLP) data sets. First, the relationship between ERA40 500 hPa heights and surface temperature and SLP was screened using stepwise multiple regression based on the calibration period of 1958–2002 (1998/2000 according to the availability of SLP data). All selected predictors (temperature and SLP) were significant and their variance contribution was greater than 1%. On average, there were 8.1 variables retained in the final regression equations. Second, the regression equations were applied to compute the 500 hPa height through to the late 19th century for the whole Northern Hemisphere. As the SLP and temperature coverage improved over time, the number of predictors decreased by about 1 in the most recent periods, and the root mean squared error decreased by about 0.8 m. A leave-one-out cross-validation method was used to test the skill and stability of the regression models. The reduction of error during the cross-validation period of 1958–1997 varied from 0.33 to 0.56, depending on the SLP data. Reconstructions were also checked using NCEP/NCAR 500 hPa heights from January 1949 to December 1957, and compared with the historical reconstruction over Europe. Reconstructions show high consistency with these independent data sets. Generally, the reconstruction provides a valuable opportunity to analyze, as well as to validate climate simulations of the variability in free atmosphere circulations over the past one hundred years.

1. Introduction

The analysis of historical atmospheric tropospheric circulation is critically important to global and regional climate change and extremes with regard to its dynamical features (e.g. Luterbacher et al., 2000; Casty et al., 2005a). Long-term observations of sea level pressure (SLP) across the Northern Hemisphere have been available since the late 19th century. Investigations of climate indices derived from SLP (such as the Arctic Oscillation/North Atlantic Oscillation, Southern Oscillation, and so on) has greatly enriched our knowledge of how and why large-scale climate has varied over the last one hundred years or so. In addition to the historical climate analysis, there is increasing interest in the assessment of historical atmospheric circulation variability as simulated by climate models (Casty et al., 2005b; Raible et al., 2005). However, global/hemispheric analyses of the free atmosphere are confined to much shorter periods of about 50 years because routine observations of global upper-level circulation have been available only since the early 1950s. The Short length of upper level data hampers our better understanding of global climate change. It is important and necessary to reconstruct historical geopotential height fields and to extend as far back as possible for both climate study and climate simulation validation. There are several approaches for computing or estimating 500 hPa heights.

(i) The direct method is to add the 500–1000 hPa thickness to the 1000 hPa height. This method is based on physical relationships, since the thickness is proportional to the mean temperature of the air column beneath the 500 hPa level (e.g. Polansky, 2002). This approach requires local SLP and temperature data, which are often missing in the early period and which should be interpolated or reconstructed in advance, which in itself introduces additional uncertainties or biases.

(ii) The second approach is to use climate models to simulate the atmospheric circulation forced by observed boundary conditions such as the sea surface temperature (SST). Similarly, missing upper level data may also be filled by reanalysis using only surface data (Compo et al., 2006). These methods essentially take into account the atmospheric internal dynamics. For example, Bengtsson et al. (2004) assessed the long term trends in reanalysis data sets based on model output under fixed observation distribution and quality. A successful simulation relies heavily on the well-documented surface data. Unfortunately, the quality and/or availability of the historical surface boundary data, including SST, snow cover, sea ice and so on, are generally low in the 19th century and early 20th century (Houghton et al., 2001, Chapter 2). Furthermore, the results of simulation vary with the skill or performance of climate models (Houghton et al., 2001, Chapter 8), and the ratios of signal/noise in geopotential heights are often higher in the tropics than the high latitudes in SST-forced simulations.

(iii) The most frequently used approach is to compute upper level heights using the statistical regression method. Regional to large-scale surface climate is strongly related to the near surface and upper level atmospheric circulation (Klein and Yang, 1986; Houghton et al., 2001). The relationship is consistent irrespective of upscaling (the local scale is related to large scale) or downscaling (the large scale is related to the local scale), the links can be determined and often quantitatively presented as specification relationship in weather and climate analysis. Therefore, geopotential heights can be derived from surface data using the reverse specification method. In early works this idea was adopted to fill in missing observations in some areas or stations (e.g. Namias, 1944). Klein and Dai (1998) developed the idea and computed monthly mean 700 hPa heights with a resolution of $10^\circ \times 10^\circ$ for 1947–1992 over the western half of the Northern Hemisphere from station temperature, precipitation, and SLP. The merit of this method is that it takes into account the relationship between surface climate and upper-air circulation. Therefore, it is often used to compute heights and atmospheric circulation indices. For example, Gong and Wang (2000) examined the possibility of computing Northern Hemisphere 500 hPa heights from 1873 using Jones' SLP (Jones, 1987) and surface temperature data. Luterbacher et al. (2002) reconstructed SLP and 500 hPa heights across Europe back to AD1500 using principal components analysis (PCA) regression. Schmutz et al. (2001) used the same predictors as Klein and Dai (1998), i.e., SLP, temperature, and precipitation to reconstruct the monthly 700, 500, and 300 hPa heights in the European and Eastern North Atlantic regions for the period 1901–1947. They applied canonical correlation analysis based on PCA filtered predictors. Using this kind of statistical approach, Brönnimann and Luterbacher (2004) filled the gaps in upper level circulation data during the period of World War II over the Northern Hemisphere. The importance of other predictors such as radiosonde data, wind direction and cloud observation as well as ship logbook information related to wind for the pre-1850 period has been recently highlighted (García-Herrera et al., 2005). For example, Brönnimann et al. (2004) and Brönnimann and Luterbacher (2004), show that upper level radiosonde data (e.g. Brönnimann, 2003) can be used as an important predictor when computing atmospheric circulation over large areas. However, these upper level observations exist for short time periods only and are sparse in space. Therefore, in our study we employed only surface temperature and SLP records.

The purpose of this paper is to objectively reconstruct 500 hPa heights for the entire Northern Hemisphere back to 1871 by using the reverse specification approach. Data and the method of stepwise regression are summarized in Sect. 2. The properties of the regression equations are presented in Sect. 3. Our reconstructions are compared with other different data sets, and the reconstructive error/skill is estimated in Sect. 4. The usefulness of the reconstruction is discussed in Sect. 5. Finally, Sect. 6 presents the brief conclusion.

2. Data and method

2.1 Data preparation

There are several long-term surface temperature and SLP data sets available since the mid-late 19th century. These are used as predictor data for the reconstruction period. Here we considered three SLP and one temperature data sets, all of them are of hemispheric or global coverage (Table 1) and are summarized below.

(i) HadSLP1, also known as GMSLP3. This is an updated version from GMSLP21f which consists of in-situ marine and land station monthly SLP observations, and is blended with several gridded analysis data sets (from Australia and the USA) to create globally complete fields. The updated version, GMSLP3, is an historical, $5° \times 5°$ gridded monthly data set covering the period 1871–1998 (Basnett and Parker, 1997).

(ii) Jones's SLP for the Northern Hemisphere since 1873 (hereafter referred to as CRUSLP). It is archived on a 5° latitude × 10° longitude mesh, and is available northward of 15° N (Jones, 1987).

(iii) National Center for Atmospheric Research (NCAR) SLP, starts in January 1899 and is gridded on a 5° latitude × 10° longitude mesh (Trenberth and Paolino, 1980).

(iv) Surface temperature data (HadCRUT2, hereafter denoted as T2 for simplicity) have been obtained from the Climatic Research Unit, University of East Anglia. This data set consists of global land and marine surface temperature anomalies since 1850 (Jones et al., 1999, 2001; Jones and Moberg, 2003; Rayner et al., 2003).

There are some additional gridded SLP data sets available. For example, Kaplan et al. (2000) released a reduced space optimal interpolation SLP (OISLP) with a resolution of 4° latitude × 4°

Table 1. Data used for geopotential height reconstruction

Data	Resolution	Coverage	Source
a. SLP data			
HadSLP1	monthly, 5° lat. × 5° lon.	Jan. 1871–Dec. 1998, global	Basnett and Parker (1997)
CRUSLP	monthly, 5° lat. × 10° lon.	Jan. 1873–Dec. 2000, north of 15° N	Jones (1987)
NCARSLP	monthly, 5° lat. × 10° lon.	1899–2003, north of 15° N	Trenberth and Paolino (1980)
b. Surface temperature			
HadCRUT2	monthly, 5° lat. × 5° lon.	Jan. 1870–Dec. 2003, global	Jones and Moberg (2003), Rayner et al. (2003)
c. 500 hPa height			
ERA-40	monthly, 2.5° lat. × 2.5° lon.	Jan. 1958–Dec. 2001, global	http://data.ecmwf.int/data/d/era40_daily

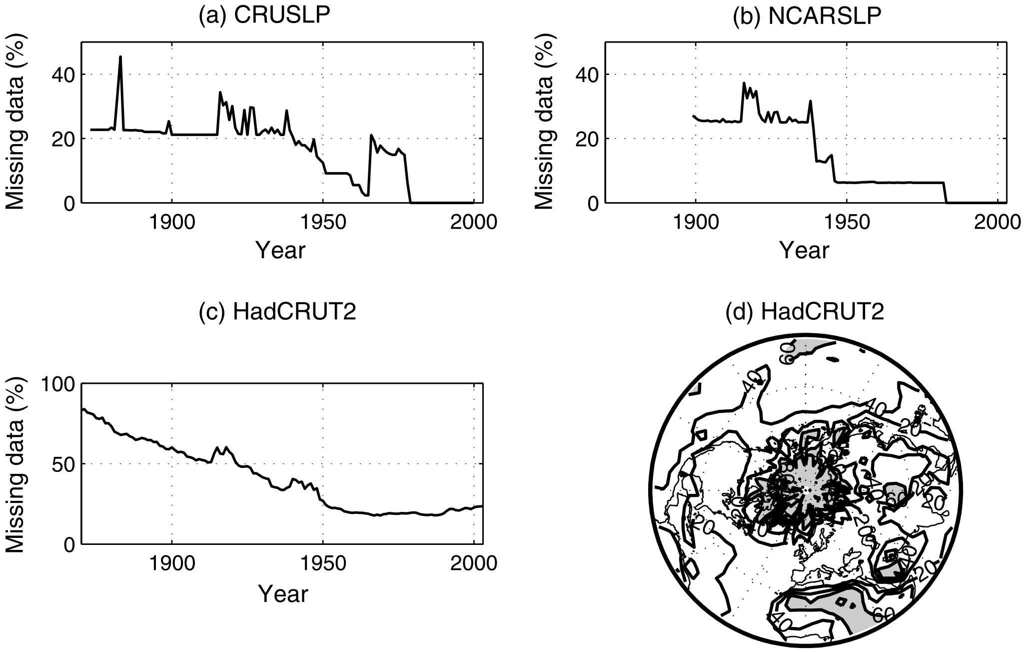

Fig. 1. Missing data (in percentage) for SLP and temperature fields. (**a**) CRUSLP, north of 15° N; (**b**) NCARSLP, north of 15° N; (**c**) HadCRUT2 temperature, north of 0° N; (**d**) the distribution of missing data in HadCRUT2 is for period 1856–2003, areas in excess of 60% are shaded

longitude over global seas for April 1854–December 1992, which is computed using the leading 80 empirical orthogonal functions, based on the Comprehensive Ocean–Atmosphere Data Set (COADS) data sets. Since this data set covers only oceans, it has not been used in the present study. In addition, it should be noted that SLP over oceans in HadSLP1 heavily relies on the quality and availability of COADS SLP records. In the present study HadSLP1 pressure data have been used, therefore the exclusion of OISLP would not reduce the potential SLP information over oceans.

For CRUSLP, NCARSLP, and T2, coverage before the 1950s is low, but improved gradually since this time (Fig. 1). Figure 1d shows the spatial distribution of missing data in T2. In the Arctic Ocean and neighbouring regions, central Pacific, and northern Africa the missing data exceeds 60%. For the CRUSLP and NCARSLP data sets, similar conditions exist. Generally, data availability in northern polar regions is poor during the reconstruction period. Among all data sets only HadSLP1 is complete. The gaps in the original observation are statistically interpolated in advance. However, our reconstructions reveal that the interpolation provides no additional information which would result in a better reconstruction. Similar features were also found in the HadSLP2 data (Allan and Ansell, 2006).

The predictand data used here are ECMWF (European Centre for Medium-Range Weather Forecast) 40-year reanalysis (ERA40) 500 hPa heights available from September 1957–August 2002 which are archived on a regular 2.5° × 2.5° grid. After re-sampling we used a 5° × 5° resolution sub-data set in our reconstruction. It should be pointed out that differences exist between instrumental station data and gridded data, arising from many sources such as the original data uncertainty, quality control procedure, interpolation method, and so on. Reanalysis 500 hPa is different from the observations. These differences might be profound, despite reanalysis relying heavily on observations.

2.2 Methodology

Of the statistic approaches available, some techniques are more widely used than others. One is the climate field reconstruction (CFR) technique (e.g. Smith et al., 1996; Jones and Mann, 2004; Brönnimann and Luterbacher, 2004; Rutherford et al., 2003, 2005; Mann et al., 2005; Casty et al., 2005a; Xoplaki et al., 2005, among others), which does not assume any a priori local relationship between predictors and the climatic field being reconstructed, and depends more heavily on assumptions about the stationarity of relationships between the predictors and large-scale patterns of climate variability than the local calibration technique (Jones and Mann, 2004; Rutherford et al., 2005, and references therein for further information). As a result, the CFR approach often produces a smoother reconstruction. However, the quality of the reconstruction may depend not only on the number/density and quality of predictors but also on the specific locations, since this determines whether or not key large-scale patterns of variance are likely to be captured. In particular, large-scale climate patterns derived from fewer predictors easily drift or can be distorted. Recently, some authors performed the so-called independent climate field reconstruction using a number of different parameters. Interestingly each reconstruction shared no common predictor (Casty et al., 2005a, c). This enables the relation of the fields to one another and helps investigate the possible dynamics and interactions between climate variables.

Alternatively, local multivariate regression is a ‘classic’ technique (e.g. Klein and Dai, 1998; Gong and Wang, 2000, among many others), which allows one to screen all possible local predictors, even those which are remotely located. Possible teleconnections may be identified and included, helping to increase the reconstruction skill. This is particularly true when there are spare observations. Since our aim is to reconstruct the historical 500 hPa heights as skillfully as possible, we did not test the independent reconstruction. To make use of as much predictive information as possible, we prefer to employ the local regression technique, using combined predictors (SLP plus surface temperature). Interestingly, in dense data areas such as Europe, the local regression and PCA-based reconstructions produce highly similar results (see Sect. 4.3).

In the present study we employ a stepwise regression technique to reconstruct the historical 500 hPa heights. The regression equations between surface climate variables (temperature and SLP) and the predictand (ERA40 500 hPa heights) are derived for the calibration period from 1958. The relationship is then applied to the early time period from the late 19th century to 1957, to compute the 500 hPa heights. According to the available predictor data sets, three reconstruction experiments were carried out, namely (a) HadSLP1 + T2 experiment, (b) CRUSLP + T2 experiment, and (c) NCARSLP + T2 experiment. The procedure is summarized below:

(i) All data, including the predictor and predictand data sets, were adjusted and presented as anomalies with respect to 1961–1990. All missing data were simply removed. To maintain a consistent spatial resolution with surface temperature and SLP, the 500 hPa height predictand data were prepared on a $5° \times 5°$ mesh. The reconstructed 500 hPa heights cover 1368 grids (19 in latitude and 72 in longitude) extending from 90° N, 0° E, to 0° N, 355° E.

(ii) After the predictand ERA40 500 hPa height at a specific grid-point and month was selected, the predictor data matrix, which consists of six variables including temperature and SLP for the concurrent month (denoted as month 0), as well as for two adjacent months, i.e. the previous one month (denoted as month -1) and the later one month (denoted as month $+1$), over the whole Northern Hemisphere, for each of the three reconstruction experiments was prepared. The Calibration period for case HadSLP1 + T2 is 1958–1998, for case CRUSLP + T2 is 1958–2000, and for case NCARSLP + T2 T2 is 1958–2001. Surface climate is often related to upper level atmospheric circulation with a time lag. Thus, the use of surface temperature and SLP in month (-1) and month ($+1$) as predictor candidates can provide additional information, particularly in the early period when the observations are sparse. This selection of months has also been used by Klein and Dai (1998), Gong and Wang (2000) and Brönnimann and Luterbacher (2004).

(iii) It should be pointed out that compared to the sample size the number of candidate variables is very large. Prior to the analysis the candidates were screened by simply computing correlations (R) between predictors and ERA40 500 hPa heights, and those variables with moderate-high correlations were screened out. Many of the $|R|$ values were rather small, representing low or no skill in predicting height. These low skill variables were excluded from the stepwise regression process. However, the large data matrix, would require large computing resources, or could even result in computing error and instability during matrix manipulation. A better way to avoid this is to remove these low-skill variables prior to the regression. This low skill threshold was determined at $|R| > 0.25$ (approximately the 90% confidence level). The screening was carried out locally, grid-point by grid-point and month by month, using the entire calibration period data. The result being the inclusion of the most predictive predictors in a much smaller predictor matrix. This screening process reduced the predictor matrix about 40%. Most importantly, this screening process does not reduce the reliability of the finally results. In most of cases the number of variables in the final regression equation remained the same when $|R|$ increased from 0 to 0.1, and changes slightly when $|R|$ increased from 0.1 to 0.2.

(iv) Then stepwise regression was applied to derive the regression equations between the predictand (ERA40 500 hPa heights) and temperatures and SLPs. In an ordinary stepwise regression approach, the selection and removal of variables from the candidate predictors are determined according to their statistical significance. Here the confidence level was set at to 95%, i.e., during the variable-selection process any variables entering the regression equation must be significant at the 95% level, and during the variable-removal process all predictors not significant at the 95% level were eliminated. Even following the application of pre-process screening, as described in step (iii), the predictor matrix was still rather large. Theoretically, adding more variables to the regression equation would increase the explained variance (i.e. larger R^2) – enough variables are present when R^2 approaches 1. In practice, even a random time series would input a little variance into the equation. However, this is meaningless and causes the problem of overfitting (Wilks, 1995). Obviously, a variable contributing only a very small portion of the total variance would bring large uncertainty and bias in the final regression results, even if this variable is statistically significant. In order to produce a stable and robust regression equation, a third criterion was employed: all included variables had to contribute more than 1% of the total variance. This rule applies to each single grid-point. Since the available predictors change with time, the number of final predictors may also change with time and grid-point.

(v) Finally, regression equations were employed to compute the 500 hPa height anomalies in the early periods (January 1871–December 1957 for HadSLP1 + T2, January 1873–December 1957 for CRUSLP + T2, and January 1899–December 1957 for NCARSLP + T2 case).

3. Properties of the final regression equations

3.1 Predictor variables

Figure 2 shows the number of predictors in the final regression equations. Although the spatial coverage and availability of data is different among these SLP data sets, the number of predictors in the final regression equations was very similar. On average, there were 8.2 predictors for HadSLP1 + T2 and CRUSLP + T2, and 8.1 for NCARSLP + T2, almost the same. The number of predictors varied greatly from grid-point to grid-point and from month to month. The minimum number of predictors varied between 1 and 3, notice that these minimum numbers appear as the extreme conditions for a couple of grid-points at a specific month and specific year, not an average for the whole Northern Hemisphere over the entire reconstruction period. Among the three cases, the maximum number varied between 15 and 20. On average, temperature and SLP accounted for 59.3 and 40.7% of the total number of predictors. For a target month, temperature predictors in the current month account for 26.5%, the temperature of month (−1) and month (+1) account for 16.1 and 16.7%, respectively. The SLP in the current month accounts for about 24.5%, while the SLP in the adjacent months accounts for much lower percentages in all three reconstruction experiments (see Fig. 2). Obviously, to predict 500 hPa for the target month, simultaneous SLP is very important, while temperature

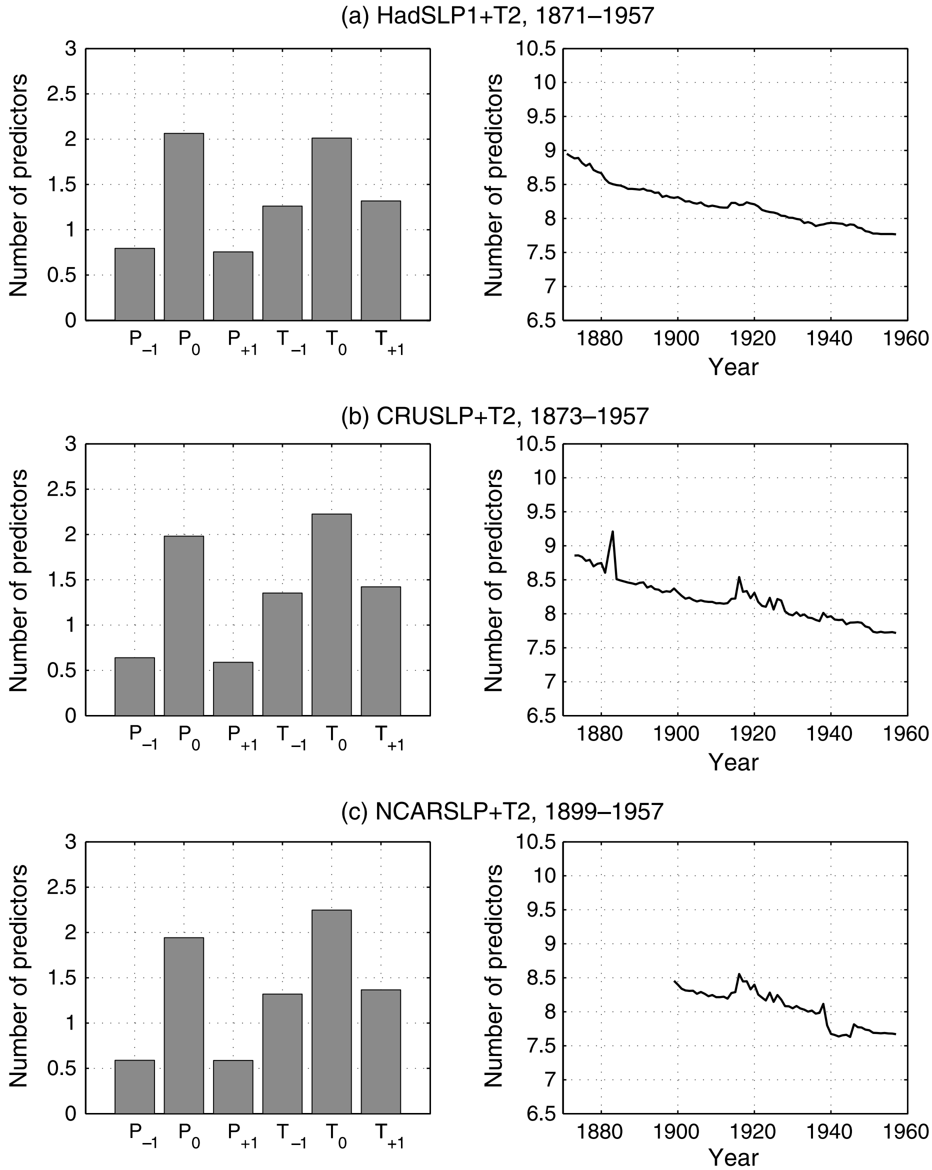

Fig. 2. Number of predictors in the final regression equations for the three reconstruction experiments. Shown in the right panel are the changes of the number of variables for each experiment. The mean total number of predictors for whole reconstruction period is shown in the left panels, where 'P' and 'T' denote SLP and temperature, respectively. The subscripts of '0', '−1', '+1' denote the concurrent, previous one, and next one month, correspondingly

predictors in the simultaneous month, as well as in the preceding and following months are also very important. This may be due to the fact that atmospheric circulation has a short memory in relation to surface climate anomalies. Temperature often has greater persistence. Thus, temperature data in the adjacent months can provide a very useful indication of the upper level atmospheric circulation, either as a preceding surface boundary forcing for the circulation or as a delayed response to the circulation anomalies in the physical sense. Both can establish strong associations between 500 hPa heights and surface temperature. Therefore, inclusion of SLP and temperature in the adjacent months increases the number of efficient candidates.

Figure 3 shows the annual cycle of the number of predictors. All three reconstructions show similar features. The number of variables increased

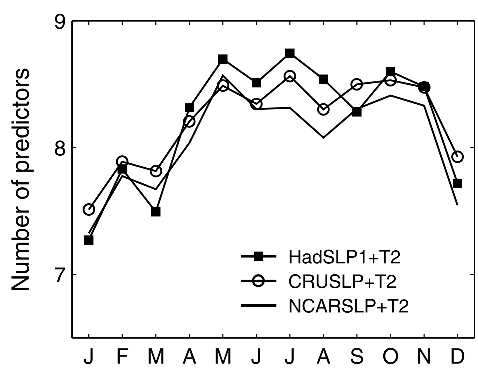

Fig. 3. Number of predictors for different months

from about 7.4 in January to about 8.5 in July, and then dropped to about 7.8 in December. The seasonal difference is clearly evident. On average the number of predictors in winter (December–February) is 7.6. For June–August, the variable number is 8.4. This feature is different from Klein and Dai (1998) results. In their reconstruction of 700 hPa height they used fewer variables in May–August than in December–April (see their Fig. 3b). Our results suggest that to successfully calculate the 500 hPa heights in warm seasons more predictors are needed, while the predictor number is somewhat lower in cold months. This might be due to the fact that in wintertime there are active extra-tropical atmospheric modes, as well as strong associations between the low-frequency variability in surface climate and tropospheric circulation. In winter, a few independent predictors can capture more information on climate dynamics.

Very often the regression algorithm tends to overfit the data by including many variables in the equation. In our stepwise regression and screening, we employed strict criteria to ensure that only the most skillful predictors were selected for the final equation. Highly inter-correlated variables were optimized where replication of their significant co-variance has been eliminated. The number of predictors in the final equation was about 8. It is interesting to note that this number is close to the results of Klein and Dai (1998), in their reconstruction of 700 hPa heights for the Western Hemisphere. They found that a multiple regression equation with 2–6 independent variables usually satisfies the criteria for statistical stability and synoptic reasoning. The methods used here should be helpful in improving the problem of overfitting, but perhaps not the elimination of the issue.

3.2 Goodness of fit

Two criteria were used to examine the goodness of fit and the performance of the multiple regression equations. One is the adjusted square of the multiple correlation coefficient (i.e. the adjusted R square, R_a^2), which is a measure of the actual fit of the best-fitting straight line and is equivalent to the percent reduction of the unexplained variance or reduction of variance. The other is the root mean squared error (RMSE), a measure of absolute error. At any grid-point in any month the RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\phi - \hat{\phi})^2}$$

where ϕ is the predictand, $\hat{\phi}$ is the predicted height, and n is the number of training data. For each year for a specific month, all grid-points were averaged over the Northern Hemisphere to generate an annual value. The average, maximum, and minimum were determined for the entire reconstruction period. The statistics of R_a^2 are plotted in Fig. 4 to show the possible annual cycle. The features of the three reconstructions differ noticeably from each other. The annual mean R_a^2 drops in March for the HadSLP1 + T2 reconstruction, while the conditions for CRUSLP + T2 and NCARSLP + T2 are more stable, with their annual mean R_a^2 showing only very slight change with season. The changes in maximum and minimum R_a^2 are generally in parallel with annual mean values. For the NCARSLP + T2 case, however, the minimum values in September and October show much a larger departure from the annual mean R_a^2 curve. This is probably caused by the poor data quality in some years (e.g. 1941). Overall, using the criterion of R_a^2, the reconstruction of CRUSLP + T2 showed the best performance and stability in all three reconstruction cases.

Table 2 presents the annual mean and seasonal mean RMSE. In all three reconstructions the RMSE in winter is somewhat larger than in summer. In winter the RMSE is about 10–11 m. While in summer it is between 6 and 7 m, about 30–40% smaller. However, this does not necessarily mean that the variance explained by the regression equations in winter is much smaller than in summer because the winter variance of 500 hPa heights is greater than summer, a conclusion also confirmed by RMSE which reflects the absolute error rather than the relative error. The annual cycle of RMSE is very similar for all three reconstruction cases, in terms of their seasonal fluctuations as well as their magnitudes. On average, the RMSEs for HadSLP + T2, CRUSLP + T2, and NCARSLP + T2 are 8.57, 8.75, and 8.73 m, respectively. The geographical distribution of the RMSE was then checked (Fig. 5). Large errors are located in two centres: one in the Arctic Ocean, and the other in the

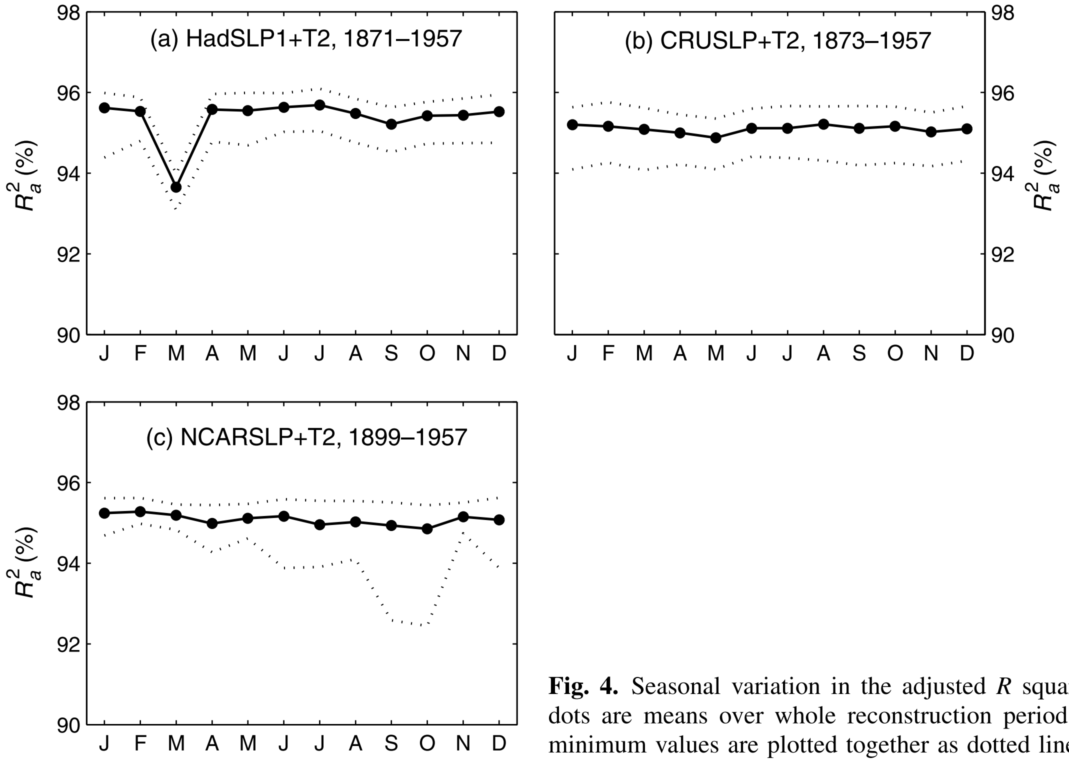

Fig. 4. Seasonal variation in the adjusted R square. Lines with filled dots are means over whole reconstruction period, the maximum and minimum values are plotted together as dotted lines

Table 2. Statistics for the root mean squared error of the regression equations. Unit m

	HadSLP1 + T2	CRUSLP + T2	NCARSLP + T2
Dec., Jan., Feb.	10.70	11.0	10.93
Mar., Apr., May	9.16	9.05	8.91
Jun., Jul., Aug.	6.41	6.73	6.84
Sep., Oct., Nov.	8.02	8.20	8.27
Yearly average	8.57	8.75	8.73

north Pacific around the Aleutian Islands. The RMSE gradually decreases from about 15 m in the high latitudes to less than 6 m in the middle and low latitudes. All three cases show similar spatial features even though there are differences in the availability and coverage of SLP, suggesting that the regression equations are generally stable among the three experiments with respect to the fitting errors.

3.3 *The influence of missing data*

As shown in Sect. 2, the number of missing data varies with time. For instance, SLP for December 1944 is entirely missing in NCARSLP. There are many missing values, especially in the first part of the 20th century. In the early period there are more than 20–50% missing data in all predictor data (except HadSLP1), after which time data availability improves gradually (Fig. 1). How does the missing data influence the performance of the regression equations? Temporal changes were examined based on (1) the number of variables, (2) RMSE, and (3) the adjusted R square. Together with the mean number of predictors in Fig. 2 the changes of the number of predictors as a function of time are shown. Obviously, the number of predictors varies with the improvement of data coverage and availability. There is a long-term decreasing trend, while during the two of World Wars the number of predictors clearly increased. Through the reconstruction period, the number of predictors dropped by about 1. It is interesting to note that the HadSLP1 + T2 displays very similar features to those of CRUSLP + T2 and NCARSLP + T2, although HadSLP1 does not

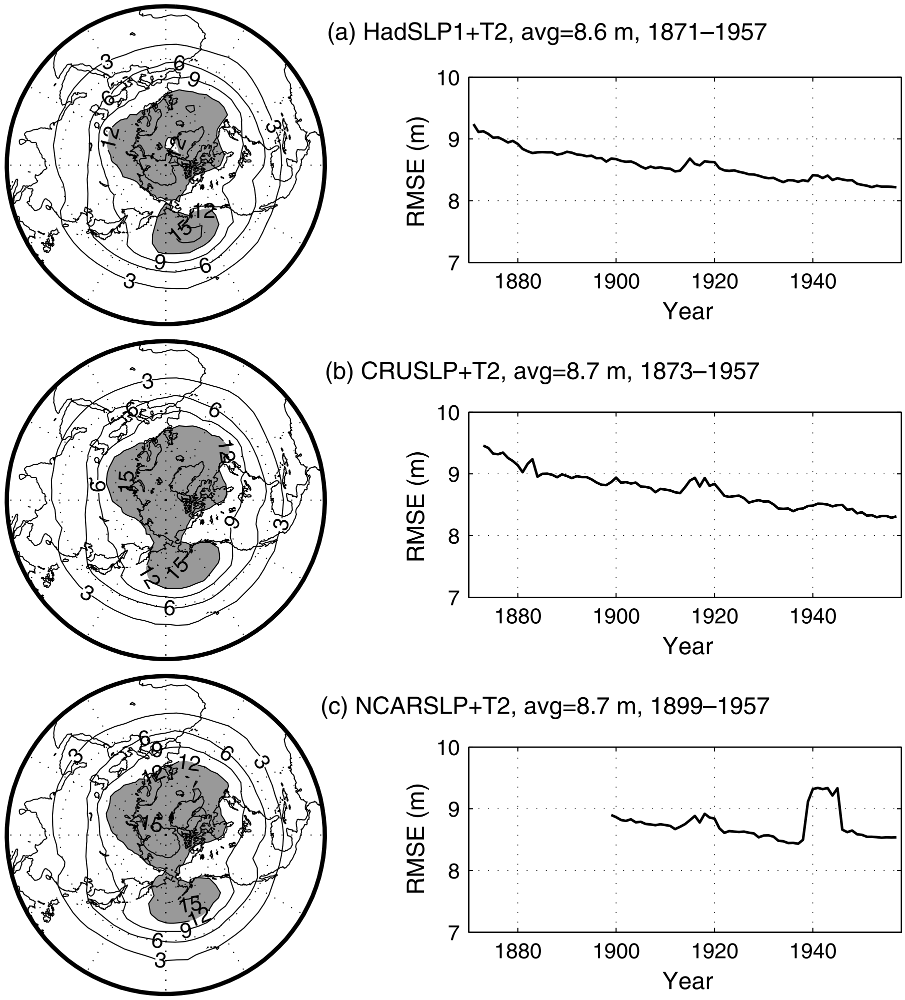

Fig. 5. Distribution of the root mean squared errors (RMSE). Regions where values greater than 12 m are shaded in order to highlight the error centers. Their changes with time are shown in the right panel

suffer from missing data. This should be due to the limitation of availability of the useful information, even though the real missing data in the original observations were interpolated by statistical methods. The meteorological signals contained in the original observations are essential. The interpolation of SLP or temperature cannot input additionally useful information for reconstruction of the 500 hPa heights.

As data availability improved, RMSE decreased simultaneously (Fig. 5). In the HadSLP1 + T2 and CRUSLP + T2 reconstructions, the errors dropped from 9 m in the 1870s to about 8.2 m in the 1950s. During the two World Wars there are small rises. With the exception of NCARSLP + T2, which shows a moderate rise during World War II. Overall there should be slightly higher uncertainty during the early period of reconstructions of HadSLP + T2 and CRUSLP + T2. For NCARSLP + T2 reconstruction there is somewhat larger uncertainty during World War II.

Next, temporal changes in the adjusted R square were examined. The three cases show similar features: an upward trend through the reconstruction period and a notable drop during World Wars I and II (figure not shown). From the 1880s to the 1950s R_a^2 rose slightly by 1% in both HadSLP1 + T2 and CRULSLP + T2 reconstructions. In the NCARSLP + T2 experiment the R_a^2 increased only 0.2% from the 1900s to the 1950s, but during World War II it clearly dropped. These results indicate that missing data do influence the performance of the regression equations. The gradual improvement of observation records led to a decrease in the number of predictors and RMSE, and also resulted in a weak increase in R_a^2, even though long-term changes were small in magnitude. With respect to R_a^2 the influence of missing data on the stability and performance of the regression equations was somewhat smaller for CRUSLP + T2 and HadSLP1 + T2 reconstructions.

4. Verification

4.1 Cross-validation

Validation can be used simply to estimate the generalization error of a given equation, or it can be used for equation selection by choosing one of several equations that have the smallest estimated generalization error. In this case, however, there is a short calibration period. If more years were retained for validation purposes, the calibration samples would be smaller and the result would very likely produce a regression with higher instability and larger error. Instead cross-validation was performed by applying a leave-one-out method. To facilitate comparison the validation period for all three experiments was 1958–1997. For each case, each month at each grid-point, calibration and estimation were repeated 40 times. Based on the cross-validation results, the reduction of error statistic (RE) (Cook et al., 1994) was used as a diagnostic of reconstructive error/skill, which is defined as:

$$RE = 1 - \frac{\sum(\phi - \hat{\phi})^2}{\sum(\phi - \phi_c)^2}$$

where ϕ is the ERA40 500 hPa heights, $\hat{\phi}$ is the estimation, and ϕ_c is the mean of the calibration period. The computation of RE was carried out locally (the sums extend over 40 years, for each month) and/or at the hemispheric level (the sums extend over both years and grid-points). $RE = 0$ represents the threshold for no skill in the reconstruction. $RE > 0.2$ provides an indication of useful reconstruction.

RE scores for all three reconstruction experiments are summarized in Table 3. CRUSLP + T2 and NCARSLP + T2 have similar and high skills, REs are +0.51 and +0.56, respectively. The HadSLP1 + T2 score is somewhat smaller at +0.33. The RE scores change with season, maximum scores occurring in winter, while lower skills showing up in summer. This is consistent with previous studies (for example Brönnimann and Luterbacher, 2004). Interestingly, the seasonal difference in CRUSLP + T2 and NCARSLP + T2 cases is notably smaller than in the HadSLP1 + T2 experiment. The HadSLP1 + T2 RE in summer is 0.18, the smallest in all cases. Moreover, the three experiments show differing stability in the RE scores from 1958 to 1997. CRUSLP + T2 and NCARSLP + T2 have similar, higher stabilities. RE scores in the HadSLP1 + T2 models have greater year-to-year fluctuations after the late 1970s (see Fig. 6). These should provide a higher level of confidence for the CRUSLP + T2 and NCARSLP + T2 reconstructions.

The spatial distribution of RE scores are presented in Fig. 7. Most of the useful reconstructions ($RE > 0.2$) are located in the extra tropics (north of about 30° N). The maximum RE scores are located in northern America, northern Pacific, northern Atlantic, northwestern Asia, and Europe. Compared to other continents, the middle to low latitudes of Asia shows relatively small values in all three reconstructions. Overall, the RE scores are lower in low latitudes in all three experiments, with minimum centres appearing in north Africa and the neighbouring Atlantic, and east Pacific about 20–25° N. Additional reconstructions were generated using the recently updated HadSLP2 data set (Allan and Ansell, 2006). The RE scores of which are very similar to HadSLP1 in terms of temporal features and the spatial distribution.

To some degree, these features are similar to the spatial distribution of the root mean squared errors (Fig. 5). RMSE centres are most obvious in the Arctic Ocean, where lower RE scores are also found (particularly in the HadSLP1 + T2 experiment). The error centre in the northern Pacific, however, disappears in the RE maps. Instead, the northern Pacific is characterised by high reconstructive skills. Brönnimann and Luterbacher

Table 3. The RE scores for cross-validation period 1958–1997

	HadSLP1 + T2	CRUSLP + T2	NCARSLP + T2
Dec., Jan., Feb.	0.55	0.60	0.64
Mar., Apr., May	0.31	0.50	0.57
Jun., Jul., Aug.	0.18	0.47	0.58
Sep., Oct., Nov.	0.30	0.47	0.47
Annual	0.33	0.51	0.56

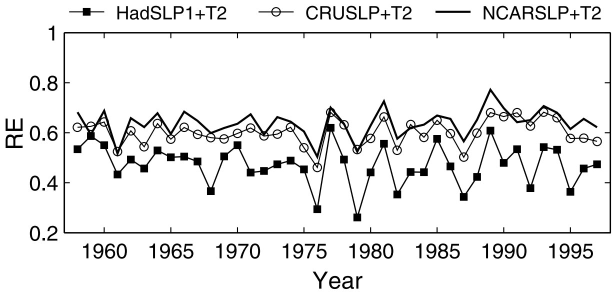

Fig. 6. Time series of the RE scores for the cross-validation period of 1958–1997

(2004) RE scores for 700 hPa and 500 hPa reconstructions for the 1939–1944 period also have very similar spatial features (c.f. their Fig. 2). Low and high reconstructive skills seem to be confined to geographical locations, whether these relate to the predictability of atmospheric circulation or to data availability and quality are questions which need further discussion.

4.2 Comparison with NCEP/NCAR reanalysis 500 hPa heights

This subsection further compares the reconstructions with National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data (Kalnay et al., 1996), which are of the same spatial resolution of $2.5^\circ \times 2.5^\circ$ as ERA-40 data but covers a longer time period. Thus, 500 hPa heights for a pre-calibration period of 1949–1957 were used as independent observation data for verification. The regression stability and the data quality of the computed 500 hPa heights were checked in the conventional way. The consistency between the reconstruction and NCEP/NCAR reanalysis can easily be judged by the spatial correlations of the height anomalies. Circulation patterns often experience high spatial autocorrelation, reducing the efficient number of degrees of freedom. The significance level of the spatial correlation can be estimated using the Monte Carlo method by randomly re-arranging the variable order. Since this study focussed on generating the reconstructions strict significant tests were not conducted. Here the correlations were simply computed from all 1368 grid-points. Data in high and low latitudes are dealt equally. Correlations were calculated for each month for each of the three reconstruction cases. Figure 8 shows the results for all months during 1949–1957 (108 months in total). The CRUSLP + T2 and NCARSLP + T2 reconstructions display smaller month-to-month differences, suggesting more stable consistency with NCEP/NCAR reanalysis 500 hPa heights. Mean

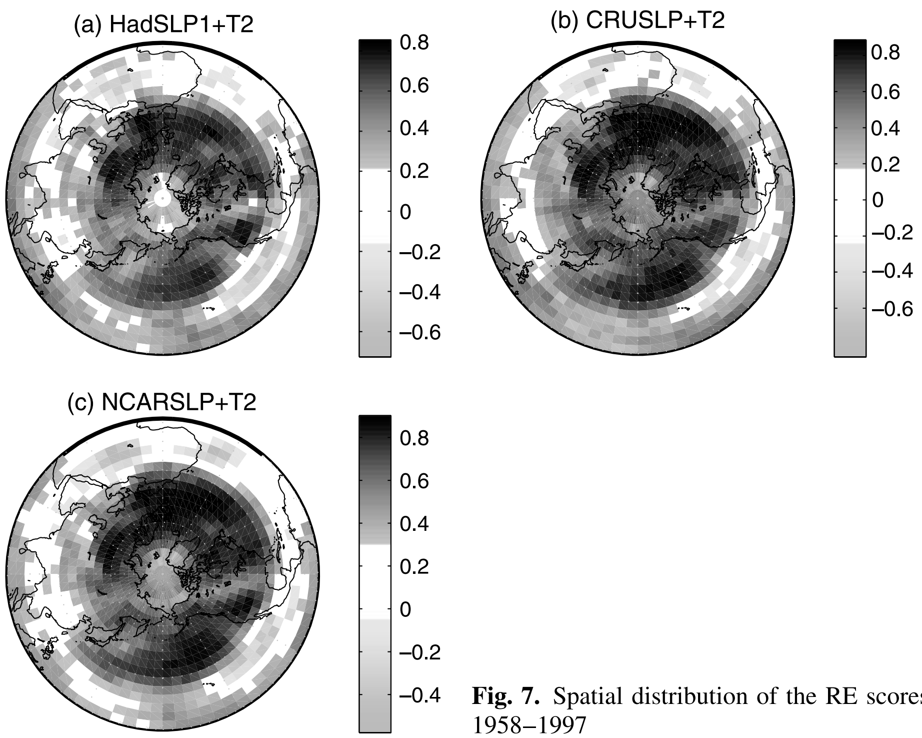

Fig. 7. Spatial distribution of the RE scores for the cross-validation period of 1958–1997

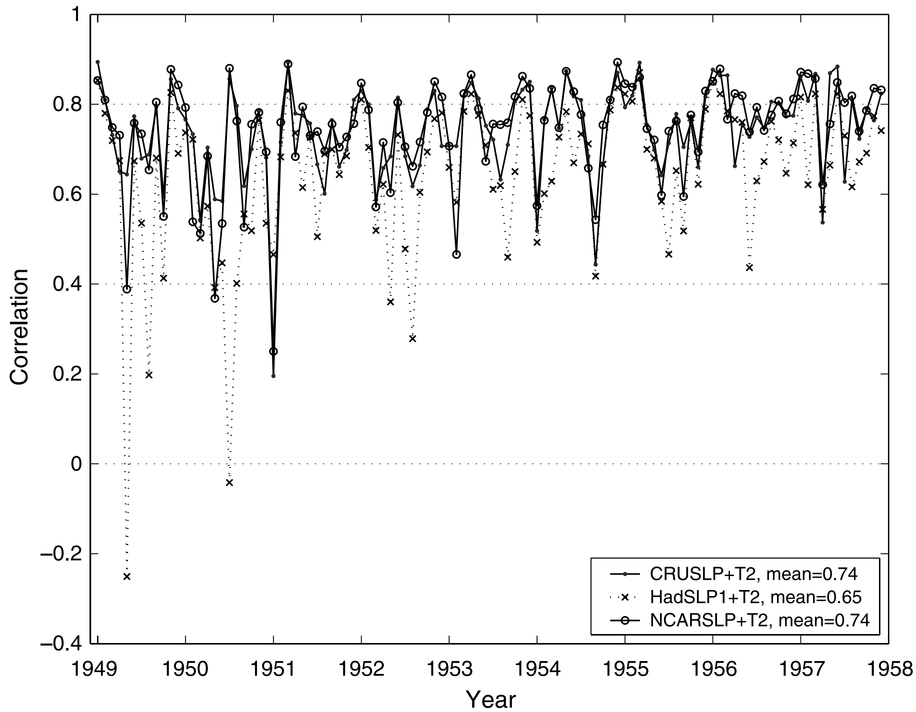

Fig. 8. Spatial correlation coefficients between the computed 500 hPa height anomalies and NCEP/NCAR reanalysis height anomalies over northern hemisphere. Three cases are plotted together for comparison

spatial correlations are also the same, 0.74. The HadSLP1 + T2 reconstruction shows a much larger month-to-month difference in the correlation, especially in the early 1950s. Two months with low correlation are outstanding. One is July 1950, with correlation near zero, and the other is May 1949, with a negative correlation of −0.25. The mean spatial correlation in 1949–1957 is 0.65, smaller than the other two reconstructions. The reason(s) why HadSLP1 + T2 reconstruction is different from other two cases are however, as yet, unclear.

The seasonal cycle of the spatial correlations is evident in HadSLP1 + T2, changing from 0.72 in winter to only 0.57 in summer. While for the CRUSLP + T2 and NCARSLP + T2 reconstructions the seasonal difference is almost indistinguishable, varying slightly between 0.76 and 0.73 (see Table 4).

To demonstrate how consistent the spatial patterns are between the reconstructions and NCEP/NCAR reanalysis 500 hPa height anomalies, the similarity in their features has been examined. Figure 9 displays an example for January 1957. In the NCEP/NCAR reanalysis map, the positive anomaly centres are located in the northern Pacific, East Asia, southern U.S., and Europe. A negative centre appears in eastern Canada and Greenland. These features are fairly well reproduced in the CRUSLP + T2 reconstruction, their spatial correlation is 0.86. The other two reconstructions contain similar features. For example, the NCARSLP + T2 is correlated with NCEP/NCAR reanalysis at 0.87 (Fig. 9).

Table 4. Statistics for the spatial correlations between reconstruction and NCEP/NCAR reanalysis 500 hPa height. Based on 9-year period of January 1949–December 1957 (108 months)

	HadSLP1 + T2	CRUSLP + T2	NCARSLP + T2
Dec., Jan., Feb.	0.72	0.76	0.76
Mar., Apr., May	0.65	0.74	0.73
Jun., Jul., Aug.	0.57	0.73	0.74
Sep., Oct., Nov.	0.66	0.73	0.75
Maximum	0.87	0.89	0.89
Minimum	−0.25	0.20	0.25
Average	0.65	0.74	0.74

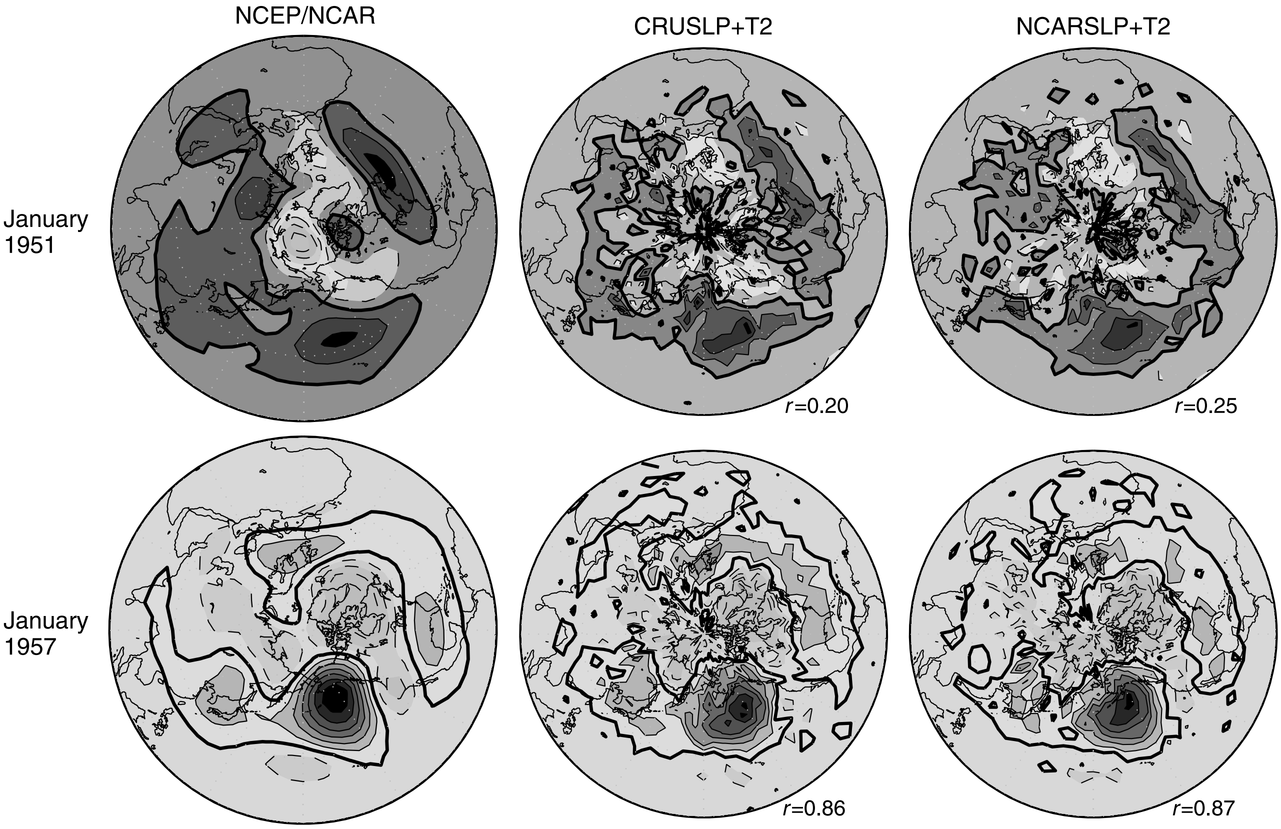

Fig. 9. Reconstructed 500 hPa height anomalies from CRUSLP + T2 and NCARSLP + T2 experiments for January 1951 and January 1957. Compared with NCEP/NCAR reanalysis 500 hPa height anomalies. Spatial correlation with NCEP/NCAR are shown too. Zero contours are shown in bold, positive contours are shown in solid lines, and negative contours are in dashed lines. Contour intervals: 40 m

Figure 8 shows there are some low correlations for several months. As an example, Fig. 9 displays the CRUSLP + T2 reconstruction for January 1951. The correlation with NCEP/NCAR 500 hPa heights is only +0.20, the lowest in all 108 months of CRUSLP + T2 reconstructions. It is very interesting to note that even though the overall spatial correlation is low, there is good consistency in the mid-low latitudes. In particular the pattern of the anomalous centres in the northern Pacific, northern Atlantic and Ural Mountains is very similar. The positions and intensities of mid-latitude anomalous centres are also very similar. Unfortunately, a great difference shows up over the Arctic regions. The same situation appears in the HadSLP1 + T2 and NCARSLP + T2 reconstructions. Figure 9 also shows the NCARSLP + T2 reconstruction for January 1951. Although its correlation of 0.25 is the lowest in all NCARSLP + T2 reconstructions, the spatial pattern in middle and low latitudes is quite similar to the CRUSLP + T2 reconstruction as well as to the NCEP/NCAR reanalysis. Similar conditions exist in other low correlation months.

The comparison with NCEP/NCAR reanalysis data clearly illustrates that the reconstructions in the middle latitudes are of higher consistency among the three reconstruction experiments and observations. The discrepancy mainly arises over the Arctic regions. This is likely due to poor data availability and probably also due to larger RMSE in northern polar regions. In addition, the dense grid in the high latitude contributes a larger portion to the correlation which can distort the correlation incorrectly biasing it toward the high latitudes. This can exaggerate the difference between the reconstruction and NCEP/NCAR reanalysis 500 hPa data and underestimate the spatial correlation.

The above investigation suggests that the reconstruction reliability in North Pole regions is lower than in middle latitudes. In all three reconstructions the CRUSLP + T2 and NCARSLP + T2 are more consistent with observations during the analysis period of 1949–1957.

4.3 Comparison with European reconstruction

Luterbacher et al. (2002) reconstructed gridded sea level pressure over the eastern North Atlantic–European region (30° W–40° E, 70° N–30° N) back to AD1500, using principal component regression analysis based on a combination of early station temperature, pressure, precipitation, and documentary proxy data. Using the same procedure, they later reconstructed 500 hPa heights on a 2.5° × 2.5° grid for the same region using NCEP/NCAR reanalysis 500 hPa heights as the predictand based on the calibration period 1948–1990. The method and data sets used in their study are different from this current study, therefore their results are different. In this subsection a comparison of the current reconstructions with the Luterbacher et al. (2002) results are discussed.

A grid point for analysis of 0° E, 60° N was randomly chosen (Fig. 10). This is an ocean gridpoint where reconstruction is not usually as good

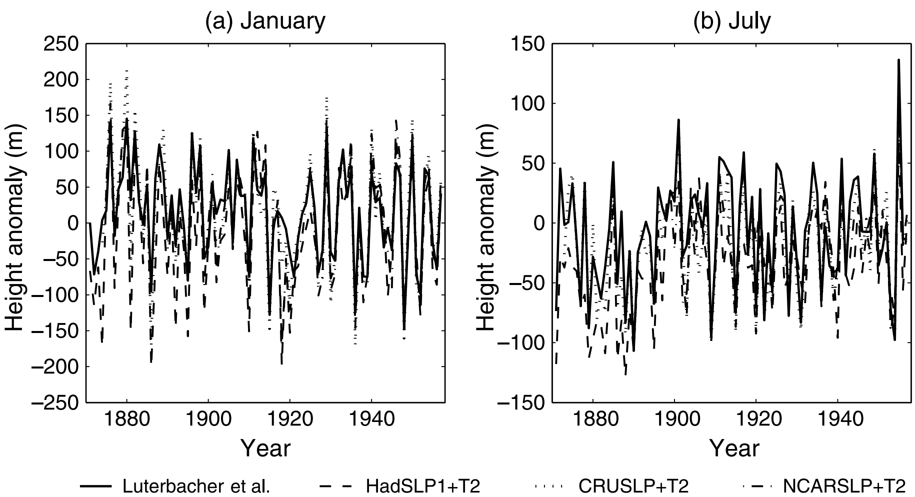

Fig. 10. Time series of the reconstructed 500 hPa height anomalies at a grid (0° E, 60° N) for (**a**) January and (**b**) July

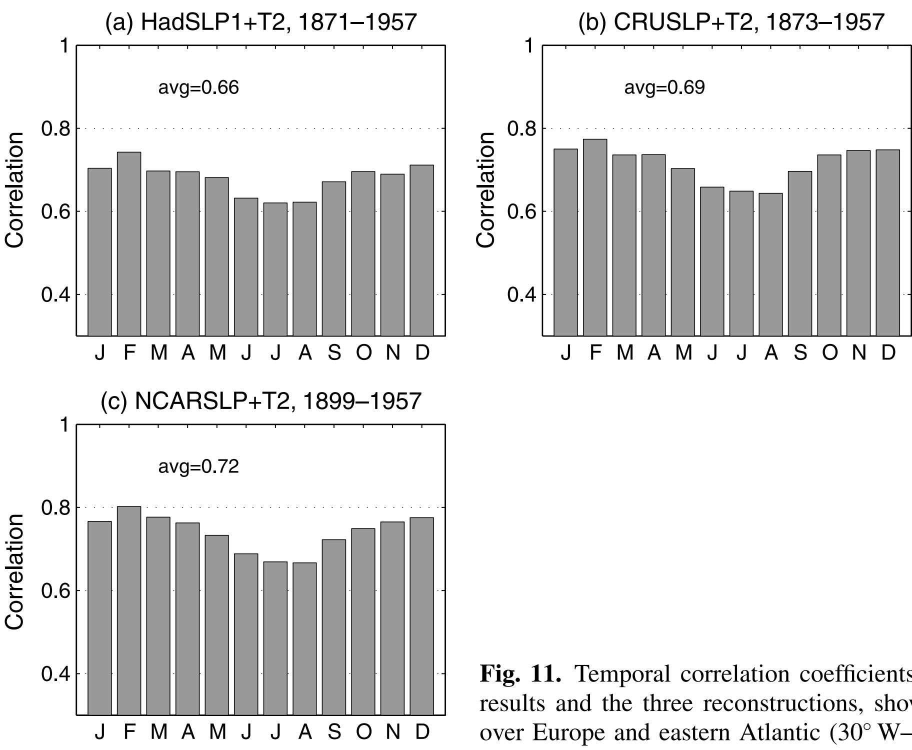

Fig. 11. Temporal correlation coefficients between Luterbacher et al. (2002) results and the three reconstructions, shown as the regional means averaged over Europe and eastern Atlantic (30° W–40° E, 70° N–30° N)

as for land grid-points. In addition, this grid-point does not show the greatest consistency with the Luterbacher et al. (2002) results. The three reconstructions are plotted together with the Luterbacher et al. (2002) values at this grid-point. Clearly, all four timeseries vary in-phase with very high consistency in both January and July. This encourages a comparison for all grid-points. Temporal correlations were calculated at each grid-point, which were found to be generally high in the three reconstructions. The consistency is greater over most of Europe, and relatively small in some southern grid-points near North Africa (figures not shown). To get an idea how the consistency varies with season, all temporal correlations were averaged for each month over the whole region. The results are plotted in Fig. 11. Minimum correlations occurred in summer (June–August) in all three reconstructions, while the maximum correlations appeared in cold seasons. Annual means of the temporal correlations for HadSLP1 + T2, CRUSLP + T2, and NCARSLP + T2 are 0.66, 0.68, and 0.72, respectively. The data sets and methods used in our reconstructions are very different from those of Luterbacher et al. (2002), but produced very similar 500 hPa height patterns and temporal variability in the eastern North Atlantic and European regions. This does not necessarily mean that the same goodness of fit applies equally to other grid-points beyond Europe, the consistency between these different reconstructions adds additional confidence to the current reconstructions with regard to the methodology and approach.

5. Preliminary applications

To demonstrate the usefulness of the reconstructions, this section briefly shows two preliminary applications. One focuses on explaining winter temperature extremes in East China, and the other tests ENSO related atmospheric anomalies in the reconstruction period. In both applications only the CRUSLP + T2 reconstruction is employed, because it shows better reconstructive skill as demonstrated in the previous sections.

5.1 Winter temperature extremes in East Asia

Winter temperature extremes in East Asia are of importance in monitoring the climate response to global warming. Mid-high latitude Asia has experienced rapid wintertime temperature variations during the last one hundred years or so (Houghton et al., 2001). Wang et al. (1999) estimated the occurrence of very cold winters using observation records for Beijing and Shanghai, which are the longest observations in China available since the late 19th century. Empirical orthogonal function analysis of December–February mean temperature anomalies showed that the leading spatial mode has the same sign at almost all stations. This unipole mode accounts for nearly half of the total variance (49%). Therefore, most of the stations experience the same or very similar temporal features in winter temperature, i.e., the cold temperatures tend to occur at many stations simultaneously, mainly due to the fact that the East Asian winter monsoon dominates over most of east China. Therefore, this provides a chance to analyze extremely cold winters in the last one hundred years based on a couple of stations. Beijing and Shanghai mean temperatures provide a reliable alternative estimation of national average temperatures. Based on the long-term observations of temperature anomalies (T') for the period 1880–1957 five extremely cold winters were identified using the criterion of $T' \leq -1.65\sigma$, where σ is the standard deviation in the whole period 1880–1957. This defines a climate extreme at the 95% level for a one-tailed *t*-test. Extremely cold winters include 1884/85 ($T' = -2.2\,°C$), 1892/93 ($-2.7\,°C$), 1935/36 ($-2.8\,°C$), 1944/45 ($-2.8\,°C$), and 1956/57 ($-2.2\,°C$). Similarly, three of the warmest winters were identified by the criterion of $T' \geq 1.65\sigma$, including 1934/35 ($+1.7\,°C$), 1945/46 ($+1.3\,°C$), and 1948/49 ($+1.5\,°C$).

With regard to East Asian temperature variability and extremes many recent studies have paid more attention to near surface circulation systems such as the Siberian High in the SLP field (Gong and Wang, 1999; Gong and Ho, 2002, 2004; Wu and Wang, 2002). To get a better understanding of the mid-tropospheric circulation anomalies in association with the winter temperature extremes, a composite analysis based on the warmest three and coldest five winters identified above was conducted. Figure 12 displays the 500 hPa difference (warmest winters minus coldest winters). A well-defined features show up: strong negative anomalies appear in a broad region

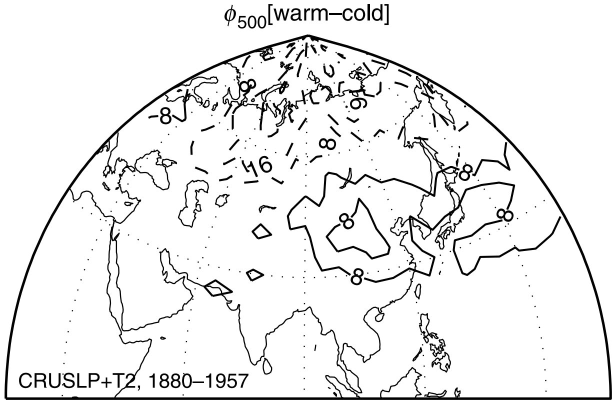

Fig. 12. Composite map of December–February 500 hPa heights using CRUSLP + T2 reconstruction with respect to the winter temperature extremes in East China. Warmest three winters (1934/35, 1945/46, 1948/49) minus coldest five years (1884/85, 1892/93, 1935/36, 1944/45, 1956/57). Unit: m

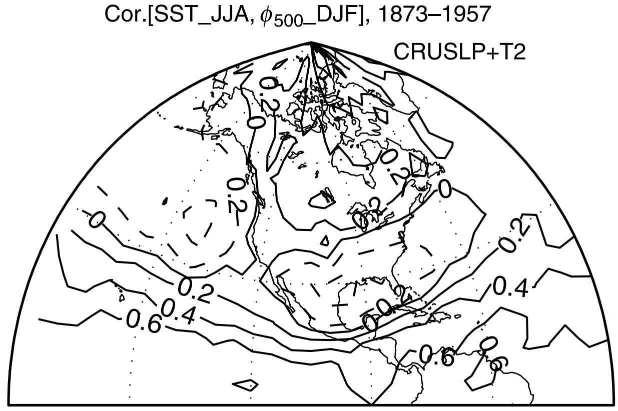

Fig. 13. Correlation of June–August Niño3 SST with December–February 500 hPa height of CRUSLP + T2 reconstruction. Data period is 1873–1957. SST leads height by 6 months

mainly covering about 60–70° N, 60–80° E with a centre located over the Ural Mountains, and positive signals in the mid-latitudes of East Asia, chiefly located over 40–45° N, 100–120° E. This circulation pattern indicates that the enhanced ridge over the Ural Mountains and deepened trough over East Asia plays an important role in the occurrence of extremely cold winters in East China. This situation provides favourable conditions for cold air masses to extend southward over East Asia under the steering of upper circulation along the enhanced airstream in the east flank of the ridge and the rear of the trough (Zhang and Lin, 1992; Gong et al., 2001).

5.2 ENSO-related circulation over Pacific/North America sector

The second application focusses on ENSO-related changes in 500 hPa height in the Pacific-North America sector. ENSO is one of the major signals for monitoring and predicting extra tropical atmospheric circulation and climates. Brief results are presented which show that the time-lag ENSO signals are well documented in reconstructed extra tropic 500 hPa heights.

Figure 13 displays the time-lag correlation between June and August Niño3 SST (Kaplan et al., 1998) and December–February 500 hPa height anomalies. Notice that the results shown here are for the SST leading 500 hPa height by half year. Centres of negative anomalies are evidently located in the northwestern North Pacific and the southeastern U.S., while positive anomalies appear in Canada as well as in a broad region over the tropical Pacific. This well-defined structure clearly manifests the Pacific/North America (PNA) teleconnection pattern. This feature is even more significant in the simultaneous correlation map. The PNA pattern is the dominant mode of seasonal/interannual variability over the Pacific/North America sector and plays a very important role in Northern Hemisphere circulation and climate predictability (e.g. Leathers et al., 1991; Renwick and Wallace, 1995, among many others). Although PNA is an internal mode, its variability is linked to tropical Pacific SST forcing (Horel and Wallace, 1981). It is a fact that in the reconstruction procedure, SSTs are used as predictors, but only SSTs for months -1, 0, and $+1$ are included in the predictor candidate matrix. The strong persistence of the ENSO signals involving the PNA pattern in the historical time period in our reconstructions is impressive, even for 500 hPa height lags behind Niño3 SST as long as 6 months. Strong circulation links beyond 2-month time-lag should not be considered as a simultaneous fitting to the SSTs, instead atmospheric dynamics should play, at least partial, roles. Therefore, our reconstruction provides an opportunity to better understand how the northern extra tropical climate has responded to ENSO during the last one hundred years.

In general, the reconstructions might provide useful information for general climate model validation and the assessment of the global climate simulation of the last century, which is one of the essential issues in global climate study (Houghton et al., 2001).

6. Conclusion

This study has developed an objective method and calculated the monthly mean 500 hPa height anomalies for all Northern Hemisphere grid-points (5° latitude × 5° longitude resolution) from surface temperature and three different SLP data sets. For each target month, predictors were determined from the candidates of target and two adjacent months. A stepwise program was used to derive the multiple regression equations. Averaging over all grid-points and all months, the number of predictors in the final regression equations was about 8.1. Then the 500 hPa heights prior to 1958 were reconstructed using these regression equations.

RE scores for CRUSLP + T2 and NCARSLP + T2 cases were higher than for HadSLP1 + T2 during the cross-validation period of 1958–1997. These two cases also display a more stable RE than the HadSLP1 case. Reconstructions based on CRUSLP and NCARSLP data sets were highly correlated with NCEP/NCAR reanalysis at 0.74. For the case of HadSLP1 the correlation is somewhat smaller, 0.65. Long-term reconstructions by Luterbacher et al. (2002) over Europe and eastern North Atlantic are also highly consistent with our results. On average the temporal correlations to our results vary from 0.66 to 0.72 during the period 1871/1873/1899–1957.

Our reconstructions provide the possibility for analyzing and understanding Northern Hemispheric climate variations in the middle troposphere since the late 19th century as two brief applications demonstrated, and may also be used for climate simulation validation.

However, it should be pointed out that the reconstructions are somewhat noisy (for example, Fig. 9). Possible reasons include that (1) the calibration period might be too short to derive a stable relationship, and (2) the regression method is heavily based on some highly sensitive local predictors. To overcome these shortcomings, future work is planned to perform an accompanying reconstruction using PCA-regression and taking into account newly updated SLP and temperature data sets.

Acknowledgments

Authors thank Drs. David Parker and Tara Ansell at the Hadley Centre, and Steven Worley of UCAR for their help in providing SLP data sets. Thanks are due to P. D. Jones of the University of East Anglia for providing temperature and CRU SLP data, and to Jürg Luterbacher of the Institute of Geography of University of Bern for providing 500 hPa reconstructions over the eastern North Atlantic-European region. The High-Performance Computing Centre at University of Bergen provided the computing resources for this study when the first author visited the Nansen Environmental and Remote Sensing Center. The authors thank the three reviewers for their constructive comments on an early version of the manuscript. Yongqi Gao was supported by the Chinese Academy of Sciences Hundred Talent Project.

References

Allan RJ, Ansell TJ (2006) A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850–2004. J Climate (accepted)

Basnett T, Parker D (1997) Development of the Global Mean Sea Level Pressure Data Set GMSLP2. Climate Research Technical Note, 79, Hadley Centre, Meteorological Office, Bracknell, U.K. 16 pp plus Appendices

Bengtsson L, Hagemann S, Hodges KI (2004) Can climate trends be calculated from reanalysis data? J Geophys Res 109(D11111). doi: 10.1029/2004JD004536

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

Brönnimann S, Luterbacher J (2004) Reconstructing northern hemisphere upper-level fields during World War II. Climate Dynamics 22: 499–510

Brönnimann S, Luterbacher J, Staehelin J, Svendby TM, Hansen G, Svenøe T (2004) Extreme climate of the global troposphere and stratosphere in 1940–42 related to El Niño. Nature 431: 971–974

Casty C, Handorf D, Sempf M (2005a) Combined winter climate regimes over the north Atlantic/European sector 1766–2000. Geophys Res Lett 32(L13801). doi: 10.1029/2005GL022431

Casty C, Handorf D, Raible CC, González-Rouco JF, Weisheimer A, Xoplaki E, Luterbacher J, Dethloff K, Wanner H (2005b) Recurrent climate winter regimes in reconstructed and modeled 500 hPa geopotential height fields over the north Atlantic/European sector 1659–1990. Climate Dynamics 24: 809–822

Casty C, Wanner H, Luterbacher J, Esper J, Böhm R (2005c) Temperature and precipitation variability in the European Alps since 1500. Int J Climatol 25(14): 1855–1880

Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100-year reanalysis using only surface pressure data. Bull Amer Meteor Soc 87: 175–189

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

García-Herrera R, Können GP, Wheeler DA, Prieto MR, Jones PD, Koek FB (2005) CLIWOC: a climatological database for the world's oceans 1750–1854. Climatic Change 73: 1–12

Gong D-Y, Ho C-H (2002) Siberian High and climate change over middle to high latitude Asia. Theor Appl Climatol 72: 1–9

Gong D-Y, Ho C-H (2004) Intra-seasonal variability of wintertime temperature over East Asia. Int J Climatol 24(2): 131–144

Gong D-Y, Wang S-W (1999) Long-term variability of the Siberian High and the possible influence of global warming. Acta Geographica Sinica 54(2): 125–133 (in Chinese)

Gong D-Y, Wang S-W (2000) Experiments on the reconstruction of historical monthly mean northern hemispheric 500 hPa heights from surface data. J Trop Meteor 16(2): 148–154 (in Chinese)

Gong D-Y, Wang S-W, Zhu J-H (2001) East Asian winter monsoon and Arctic Oscillation. Geophys Res Lett 28: 2073–2076

Horel JD, Wallace JM (1981) Planetary scale atmospheric phenomena associated with the Southern Oscillation. Mon Wea Rev 109: 813–829

Houghton JT, Ding Y-H, Griggs DJ, Noguer M, van der Linden PJ, Dai X-S, Maskell K, Johnson CA (2001) Climate Change 2001: the Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovermenal Panel on Climate Change. Cambridge, U.K.: Cambridge University Press, 881 pp

Jones PD (1987) The early twentieth century Arctic high – fact or fiction? Climate Dynamics 1: 63–75

Jones PD, Mann ME (2004) Climate over past millennia. Rev Geophys 42(RG2002). doi: 10.1029/2003RG000143

Jones PD, Moberg A (2003) Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. J Climate 16: 206–223

Jones PD, New M, Parker DE, Martin S, Rigor IG (1999) Surface air temperature and its variations over the last 150 years. Rev Geophys 37: 173–199

Jones PD, Osborn TJ, Briffa KR, Folland CK, Horton B, Alexander LV, Parker DE, Rayner NA (2001) Adjusting for sampling density in grid-box land and ocean surface temperature time series. J Geophys 106D: 3371–3380

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 KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77: 437–431

Kaplan A, Cane MA, Kushnir Y, Clement AC, Blumenthal MB, Rajagopalan B (1998) Analysis of global sea surface temperature 1856–1991. J Geophys Res 103: 18567–18589

Kaplan A, Kushnir Y, Cane MA (2000) Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. J Climate 13(16): 2987–3002

Klein WH, Dai Y (1998) Reconstruction of monthly mean 700-mb heights from surface data by reverse specification. J Climate 11(8): 2136–2146

Klein WH, Yang R (1986) Specification of monthly mean surface temperature anomalies in Europe and Asia from concurrent 700 mb monthly mean height anomalies over the Northern Hemisphere. J Climatol 6: 463–484

Leathers DJ, Yarnal B, Palecki MA (1991) The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. J Climate 4(5): 517–528

Luterbacher J, Rickli R, Tinguely C, Xoplaki E, Schüpbach E, Dietrich D, Hüsler J, Ambühl M, Pfister C, Beeli P, Dietrich U, Dannecker A, Davies TD, Jones P, Slonosky V, Ogilvie AEJ, Maheras P, Kolyva-Machera F, Martin-Vide J, Barriendos M, Alcoforado MJ, Nunes MF, Jónsson T, Glaser R, Jacobeit J, Beck C, Philipp A, Beyer U, Kaas E, Schmith T, Bärring L, Jönsson P, Rácz L, Wanner H (2000) Reconstruction of monthly mean sea level pressure over Europe for the late Maunder minimum period (1675–1715). Int J Climatol 20: 1049–1066

Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Reconstrction of sea level pressure fields over the eastern north Atlantic and Europe back to 1500. Climate Dynamics 18(7): 545–561

Mann ME, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J Climate 18: 4097–4107

Namias J (1944) Construction of 10000-foot pressure charts over ocean areas. Bull Amer Meteor Soc 25: 175–182

Polansky B (2002) Reconstructing 500-hPa height fields over the northern hemisphere. Msc thesis, University of Washington

Raible CC, Stocker TF, Yoshimori M, Renold M, Beyerle U, Casty C, Luterbacher J (2005) Northern hemispheric trends of pressure indices and atmospheric circulation patterns in observations, reconstructions, and coupled GCM simulations. J Climate 18: 3968–3982

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kaplan A, Kent EC (2003) Globally complete analyses of sea surface temperature, sea ice and night marine air temperature, 1871–2000. J Geophys Res 108(D14):4407. doi: 10.1029/2002JD002670

Renwick JA, Wallace JM (1995) Predictable anomaly patterns and the forecast skill of the northern hemisphere wintertime 500-mb height field. Mon Wea Rev 1232: 114–2131

Rutherford S, Mann ME, Delworth TL, Stouffer RJ (2003) Climate field reconstruction under stationary and nonstationary forcing. J Climate 16: 462–479

Rutherford S, Mann ME, Osborn TJ, Bradley RS, Briffa KR, Hughes MK, Jones PD (2005) Proxy-based northern hemisphere surface temperature reconstructions: sensitivity to methodology, predictor network, target season and target domain. J Climate 18: 2308–2329

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. Climate Research 18(3): 181–193

Smith TM, Reynolds RW, Livezey RE, Stokes DC (1996) Reconstruction of historical sea surface temperatures using empirical orthogonal functions. J Climate 9(6): 1403–1420

Trenberth KE, Paolino DA (1980) The northern hemisphere sea-level pressure data set: trends, errors and discontinuities. Mon Wea Rev 104: 1354–1361

Wang S-W, Gong D-Y, Chen Z-H (1999) Serious climatic disasters of China during the past 100 years. Quart J Appl Meteor 10(Suppl): 4353 (in Chinese)

Wilks DS (1995) Statistical methods in the atmospheric sciences: an introduction. U.S.A.: Academic Press, 467 pp

Wu B-Y, Wang J (2002) Winter Arctic Oscillation, Siberian High and East Asian winter monsoon. Geophys Res Lett 29: 1897. doi: 10.1029/2002GL015373

Xoplaki E, Luterbacher J, Paeth H, Dietrich D, Steiner N, Grosjean M, Wanner H (2005) European spring and autumn temperature variability and change of extremes over the last half millennium. Geophys Res Lett 32(L15713). doi: 10.1029/2005GL023424

Zhang J-C, Lin Z-G (1992) Climate of China. John Wiley & Sons and Shanghai Scientific and Technical Press, 376 pp

Authors' addresses: Dao-Yi Gong (e-mail: gdy@bnu.edu.cn), Key Laboratory of Environmental Change and Natural Disaster, College of Resources Science and Technology Beijing Normal University, Beijing 100875, China; Helge Drange, Bjerknes Center for Climate Research/Nansen Environmental and Remote Sensing Center, University of Bergen, Norway; Yong-Qi Gao, Nansen-Zhu International Research Center, IAP, Beijing 100029, China.