

Reconstruction of monthly 700, 500 and 300 hPa geopotential height fields in the European and Eastern North Atlantic region for the period 1901–1947

Christoph Schmutz*, Dimitrios Gyalistras, Jürg Luterbacher, Heinz Wanner

Institute of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland

ABSTRACT: Gridded geopotential height (gph) fields at 700, 500 and 300 hPa were reconstructed back to 1901 based on sea level pressure, temperature and precipitation data over the European and Eastern North Atlantic region using canonical correlation analysis. Sensitivity analysis showed that the cumulative shared variance of the prefiltered (by means of principal component analysis) gph fields should be on the order of 95 to 99 %, and for the predictors lower than 90 %, in order to achieve reliable reconstructions. The obtained relations between the surface data and the gph fields were robust and the promising cross-validation results were revealed to be reproducible in a second independent period. The main mode of wintertime 700, 500 and 300 hPa gph field variability in the 20th century is a dipole pattern that clearly shows the signature of a regional representation of the Arctic Oscillation (AO). The time series of this mode of variability implied a strengthening of the gph gradients in the troposphere over Europe in the last decades of the 20th century; this is without precedent in the 20th century.

KEY WORDS: Reconstruction · Geopotential height · Europe · North Atlantic · Arctic Oscillation · CCA

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

To understand the low-frequency variability of the tropospheric circulation and to validate global climate models in the European and Eastern North Atlantic region, the reconstruction of geopotential height (gph) fields is of great importance. Moreover, regional impact studies (e.g. for the European Alps) and ocean-atmosphere interaction studies can be conducted on a low-frequency timescale.

The reanalysis project by NCEP/NCAR (National Center for Environment Prediction/National Center of Atmospheric Research) (Kalnay et al. 1996) provides tropospheric data back to 1948. This period is limited by the availability of tropospheric data provided by radio soundings, which became more and more frequent after World War II. The period is only partly representative of the natural low-frequency climate variability because of the increasing evidence of

human-induced climate change (Santer et al. 1996). Moreover, a shift of the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO) indices towards positive values was observed for the reanalyzed periods (Hurrell 1995, Thompson & Wallace 1998). However, since the month-to-month variability exceeds the low-frequency variability by far, the reanalysis period should be valid as a calibration period for the reconstructions. The available long-term NAO and AO indices and sea level pressure (SLP) fields provide a rather limited insight into the long-term and low-frequency variability of the tropospheric circulation. In order to better represent the tropospheric circulation, the reconstruction of gph fields for the 'pre-radio sounding' period is therefore needed. The new data set, reconstructed and described in this article, will therefore allow the tropospheric gph variability before 1948 to be assessed on different levels. Moreover, the gph variability of the second half of the 20th century with an accelerated influence of greenhouse gases and aerosols is placed into a perspective of a period where these influences were drastically smaller. Experiences

*E-mail: schmutz@giub.unibe.ch

with the present reconstruction attempt might be useful for future gph reconstructions further back in time.

Kington (1975) proposed the reconstruction of the daily 500 hPa fields for the European and Eastern North Atlantic sector with historical weather maps from 1781 by inferring the 500–1000 hPa thickness from surface temperature and SLP. However, the reconstructions of the daily charts from the year 1781 were then never conducted using his method (Kington pers. comm. 2000). A recent reconstruction study (Klein & Dai 1998) presented very good results for monthly 700 hPa gph (hereafter denoted as Z_{700}) by a point-wise linear regression technique for the western hemisphere and the first half of the 20th century. They used SLP, gridded surface air temperature, station precipitation data and station temperature data. Preliminary studies by Schmutz et al. (1999) and the good results of Klein & Dai (1998) showed potential for also reconstructing gph fields in the European and Eastern North Atlantic region for the first half of the 20th century.

To our knowledge no attempt has been made to systematically reconstruct monthly gph fields in the European and Eastern North Atlantic region for the period 1901–1947. In order to provide an adequate tropospheric data set for an analysis of the low-frequency variability of precipitation and total ozone in the European Alps (Brönnimann et al. 2000b, Schmutz in press), we adopted a reconstruction procedure which uses no station data within a few hundred kilometres of the Alps. Circumferences in the synoptic analysis will therefore be avoided. Since, for the same reason, no sea surface temperatures (SST) were taken into consideration, ocean-atmosphere studies in the Eastern North Atlantic will be possible with the finally reconstructed data set.

This study is based on the same assumption as every reconstruction study: stationarity of the statistical relationship between the predictors and predictands; thus patterns of the calibration period should also be applicable to the reconstruction period. This assumption appears to be reasonable because the month-to-month variability of the gph is much greater than the low-frequency variability on an interannual to decadal scale.

Here we address the following main questions: (1) How well and robust can we reconstruct Z_{700} , Z_{500} and Z_{300} in the European and Eastern North Atlantic region with temperature, precipitation and SLP data? (2) How sensitive are Z_{700} , Z_{500} and Z_{300} reconstructions to PC (principal component) truncation prior to the fitting procedure? (3) What is the main structure of the low-frequency variability of the reconstructed winter fields (1901–1998)?

A brief overview of the predictor and predictand data sets used is given in Section 2: the reconstruction

method used is discussed. Section 3 summarizes the results of the cross-validated PC truncation experiments. Further, the performance of the weighting scheme used is compared with the results of an experiment where no weighting scheme was applied. Robustness experiments (1948–1960) were used to assess the results of the cross-validation experiments (1961–1990). Finally, the low-frequency variability of the reconstructed upper air fields in winter is analyzed for the period 1901–1998. Conclusions are drawn in Section 4.

2. DATA AND METHODS

Klein & Dai (1998) used temperature (T), precipitation (P) and SLP data to reconstruct Z_{700} fields in the western hemisphere. As their results were good, the same predictors were chosen here. These predictors are also reasonable from a physical point of view, because they are connected to the tropospheric circulation. The distribution of the ridges and troughs in the troposphere determines advection and lifting. Therefore, the considered surface data provide direct information of the tropospheric circulation.

In order to be able to extend the reconstructions back to the 19th century in future attempts, station T and P time series were preferred to gridded data sets. Luterbacher et al. (1999, 2000) used station T and P from the GHCN.2 database (Vose et al. 1992, Peterson & Vose 1997) among other sources for their attempts to reconstruct of SLP fields and NAO indices. The quality of their predictor database was shown to be high. Here, 25 T and 25 P time series regularly distributed over Europe were taken from the GHCN.2 database. Fig. 1 shows the location of the stations used (T and P). The SLP data were taken from NCAR (Trenberth & Paolino 1980). They were on a $5^\circ \times 5^\circ$ grid from 35° N, 40° W to 75° N, 45° E (see Fig. 1). The predictands Z_{700} , Z_{500} and Z_{300} stem from NCEP Reanalysis data (Kalnay et al. 1996) on a $2.5^\circ \times 2.5^\circ$ grid from 35° N, 25° W to 65° N, 30° E (see Fig. 7 for the spatial location of the reconstructions). Six hourly data sets were averaged to monthly means for the period 1961–1990, and daily mean data sets were averaged to monthly means for the period 1948–1960. The NCEP Reanalysis data were provided by the NOAA-CIRES (The National Oceanic and Atmospheric Administration-Cooperative Institute for Research in Environmental Studies) Climate Diagnostics Center, Boulder, CO, from their Web site (<http://www.cdc.noaa.gov>). Prior to the reconstruction procedure, the monthly anomalies were calculated for the predictors and predictands. They are denoted as x' . The base climatology was from 1961 to 1990.

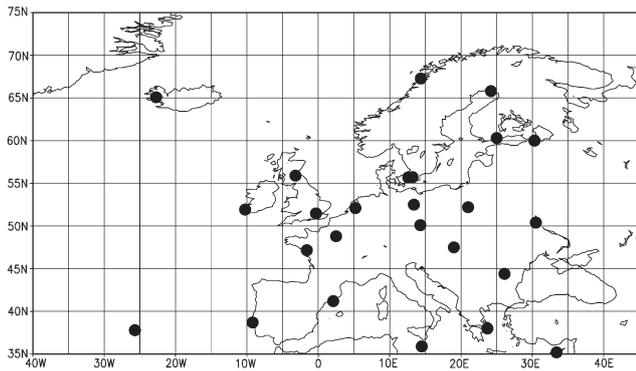


Fig. 1. Geographical locations of the predictor data. (●) Positions of the 25 T and P stations. The SLP is gridded

Most reconstruction approaches use multivariate regression techniques (e.g. Jones et al. 1987, Klein & Dai 1998, Luterbacher et al. 1999). Here, the number of variables was reduced with a principal component analysis (PCA) in order to avoid overfitting and to remove the noisy part of the predictor information. Only valuable information is kept with the retained PCs. After the PCA a canonical correlation analysis (CCA) was applied to the predictor and predictand fields (Barnett & Preisendorfer 1987, Wilks 1995, von Storch & Zwiers 1999). The CCA selects 2 patterns of 2 different fields such that the correlation between the time expansion coefficients are maximised. Then a second pair of patterns is calculated in the same manner, but their time expansion coefficients are not correlated to the first pair, and so on. The optimal relations obtained are then used to reconstruct the gph.

Results from Livezey & Smith (1999a) and the companion papers from Livezey & Smith (1999b) and Smith & Livezey (1999) showed that the applied weighting schemes (see also Eqs A1, A2 & A3 in Appendix 1) and the number of retained PCs are important for the subsequent CCA results. As a result of cross-validated sensitivity tests, they found that the PC truncation is of

greater importance than the data weighting prior to the CCA. They suggested distinct ratios for the retained number of PCs for each season and type of field specification. Since the variance concentration on the first PCs of a geophysical field can vary considerably with the location, season and variable, the suggestions of Livezey & Smith (1999a) had to be re-evaluated for the present reconstruction attempt.

Here, the optimal predictor-predictand relationship was calculated based on the CCA method of Barnett & Preisendorfer (1987). As well as the monthly predictors and predictands used, Table 1 gives an overview of the reconstruction experiments conducted.

Unless noted, the predictors and predictands were weighted prior to the PCA. In order to take into account the over-representation of the higher latitudes in the regular SLP and gph grids, these fields were latitude weighted (see Appendix 1). The T and P fields were standardized.

Since no comparable reconstruction exists for the considered region, the final reconstructions were compared to the results of a Z_{700} reconstruction of another region (western hemisphere, Klein & Dai 1998) using the root-mean-square error. This comparison provides rough but valuable information on the order of magnitude of the errors.

The reduction of error (RE) (Murphy & Epstein 1989, Wilks 1995, von Storch & Zwiers 1999) was used as a skill score in the different experiments. For each grid point j the departures to the monthly climatology was considered. The skill score for the whole fields were the areal weighted mean values RE_{areal} of RE_j . Values of the RE between 0 and 1 (0 to 100%) can be considered as a skillful reconstruction, with 1 as a perfect specification. Values less than or equal 0 indicate a trivial reconstruction in the sense that using climatology would allow at least an equally good estimation.

For the years 1961–1990 cross-validation experiments were conducted separately for each predictand month and for each combination of retained PCs considered. There was always a set of 3 fitting months, with the central month as the predicted one. For example, for a DJF (December, January, February) experiment only January was validated (30 times in the period 1961–1990). For each reconstructed month all computational steps were repeated, including the calculation of the PCA. For the next combination of retained PCs, the whole procedure was repeated. After having determined the reconstruction skill of the considered combination of retained PCs, the maximum RE_{areal} for the 3-dimensional function was determined (see also Fig. 2). This

Table 1. Conducted experiments and monthly predictors and predictands used. CCA: canonical correlation analysis T : 25 stations with temperature data all over Europe. P : same stations as for T but with precipitation data. SLP: sea level pressure. See also Fig. 1

Experiment	Period	Months	Predictors	Predictands	Weighting
CCA cross-validation	1961–1990	All	T, P, SLP	Z_{700} Z_{500} Z_{300}	Yes
CCA cross-validation	1961–1990	Jan	T, P, SLP	Z_{700} Z_{500} Z_{300}	No
CCA cross-validation	1948–1960	All	T, P, SLP	Z_{700} Z_{500} Z_{300}	Yes

was done for each month and level individually. Finally, for each maximum of RE_{areal} the proportion of variance shared by the XPCs (predictor PCs) and the YPCs (predictand PCs) was calculated.

In order to ensure that no biased results (i.e. because of autocorrelation) with a spurious RE were generated by the cross-validation experiments, robustness tests were conducted using the optimal PC truncation for the verification period 1948–1960. The same optimal relations were then used to reconstruct Z_{700} , Z_{500} and Z_{300} for the period 1901–1947.

Z_{700} , Z_{500} and Z_{300} were simultaneously analyzed with a combined PCA (Bretherton et al. 1992, von Storch & Zwiers 1999) for the reconstructions (1901–1947), observations (1948–1998) and for the combination of both periods (1901–1998). The analysis was done for the winter months (DJF) using areally weighted input data.

The spectral behaviour of the time series of the PCs was analyzed with a wavelet analysis (Torrence & Compo 1998) using a Morlet wavelet base. The statistical significance of a given local wavelet power spectrum was tested with a 10% significance level against the null hypothesis that the time series is the result of a red-noise process with a respective empirical lag-1 autocorrelation coefficient.

3. RESULTS AND DISCUSSION

3.1. PC truncation experiments

The sensitivity to PC truncation of the reconstruction method used was analyzed for each level using cross-validation, following the guidelines of Livezey & Smith (1999b). The results for all 3 levels are listed in Table 2. For each month and level the optimal PC truncation is listed with the number of retained XPCs and YPCs and their shared variances. The last column denotes RE_{areal} , which is maximized for these optimal models.

The highest RE_{areal} (76%) was calculated for Z_{700} for October—the month with the best reconstruction skill. With increasing height the skill decreased. The lowest values were around 46% for Z_{300} for July and August—the months with the lowest reconstruction skills.

Table 2 shows that most of the models with an optimal PC truncation have a very high proportion of shared variance for the YPCs of 90 to 99%. In contrast to this, the shared variance of the XPCs was generally much lower, with values between 65 and 90%. The optimal PC truncations of Table 2 were used for the final reconstructions of the period 1901–1947 and also for the robustness test (1948–1960).

No such coherent picture was found when considering the number of retained PCs. For most months and

Table 2. The number of PCs (principal components) retained for the subsequent CCA, their shared variance and the RE_{areal} for the models with the optimal PC truncations for all 3 levels and each month separately. Cross-validation experiments from 1961–1990. XPC: number and cumulative shared variance of the retained PCs of the predictors (T , P and SLP); YPC: same as XPC, but for the predictands (Z_{700} , Z_{500} and Z_{300})

	XPC		YPC		RE_{areal} (%)
	No. of PCs	Shared variance (%)	No. of PCs	Shared variance (%)	
Z_{700}					
Jan	12	83.3	10	99.5	73.4
Feb	18	89.5	4	92.2	67.7
Mar	8	72.8	10	99.2	67.0
Apr	10	74.1	10	98.9	58.5
May	16	80.9	10	98.3	72.8
Jun	22	88.0	10	98.1	58.9
Jul	12	73.8	14	99.2	54.1
Aug	8	66.3	12	98.9	55.3
Sep	16	83.8	8	97.5	68.2
Oct	16	84.9	8	97.9	76.0
Nov	6	68.0	6	96.5	67.2
Dec	14	85.1	6	97.2	72.6
Z_{500}					
Jan	14	85.9	10	99.3	67.0
Feb	22	92.6	4	90.2	67.1
Mar	10	76.9	6	94.8	63.3
Apr	22	90.3	10	98.4	61.0
May	16	80.9	8	95.8	67.7
Jun	24	89.7	10	97.7	57.9
Jul	22	87.4	14	99.2	52.9
Aug	10	70.8	10	98.0	50.8
Sep	18	86.3	8	97.2	65.6
Oct	16	84.9	8	97.5	71.7
Nov	6	68.3	6	95.6	62.5
Dec	10	79.1	6	96.5	69.3
Z_{300}					
Jan	10	80.3	4	89.6	57.8
Feb	6	70.8	4	88.5	60.2
Mar	12	80.4	6	94.0	57.4
Apr	14	81.2	8	96.1	58.6
May	12	74.3	8	95.2	60.5
Jun	8	65.6	20	99.6	55.9
Jul	10	69.9	20	99.6	46.9
Aug	16	81.2	6	91.9	46.3
Sep	18	86.3	14	99.3	59.4
Oct	22	90.5	8	96.9	68.1
Nov	6	68.0	6	94.7	56.6
Dec	6	69.6	6	95.6	65.7

levels the number of retained XPCs was higher than the number of retained YPCs. This is due to the fact that the variance concentration of the first XPCs was lower than that of the first YPCs.

Only a few of the experiments presented above (Table 2) led to a comparable number of retained PCs as in Livezey & Smith (1999b). Our experiments implied a lower number of XPCs and a higher number of YPCs than suggested in Livezey & Smith (1999b). To some extent these discrepancies can be explained by differing variables, geographical locations and resolutions of the analyzed fields. For example, the shared variance for the first PCs changes with season and

height. The variance concentration for the first PCs also depends on the predictors and predictands used. The spatial autocorrelation is the key factor for this. If a spatially very homogeneous data set is considered (e.g. Z_{700} , Z_{500} or Z_{300}), only a small number of PCs are needed to account for 90 to 95% of the variance. If a spatially less homogeneous data set (T , P or SLP) is considered, more than twice as many PCs are needed. Therefore, we prefer to use the cumulative shared variance of the retained PCs rather than the number of retained PCs.

The number of retained PCs can also be seen under the aspect of the spatial degrees of freedom (DOF; Fraedrich et al. 1995, Wang & Sheng 1999). For the European and Eastern North Atlantic sector, Fraedrich et al. (1995) found robust monthly DOF for different gph and different temporal and spatial sample sizes. Their DOF varied for low-pass-filtered daily data between 5 in winter and 10 in summer. Wang & Sheng (1999) showed with different methods that the DOF by Fraedrich et al. (1995) might be underestimated. Nevertheless, they also found a clear seasonality with a marked increase in the DOF in summertime for the northern hemisphere. Here, the empirically defined number of retained PCs also showed a marked seasonal dependency (Table 2) for both the XPCs and YPCs. The absolute DOF of the low-pass-filtered daily gph data by Fraedrich et al. (1995) are of the same order of magnitude as the retained YPCs (gph).

Considering the sensitivity of the skill to PC truncation, it was possible to distinguish 2 groups. Fig. 2 illustrates this fact for 2 typical examples (June and December) for each level. For August to March RE_{areal} —depending on the shared variance of XPC and YPC—showed a flat sensitivity ‘topography’ (right-hand side of Fig. 2) and therefore a quite robust picture. A more convex form was found for April to July (left-hand side of Fig. 2). For both groups RE_{areal} was rather sensitive, with clearly lower RE_{areal} for combinations of high values of shared variance for XPC (>90%) and YPC (>98%).

The same analysis as in Table 2 and Fig. 2 was also conducted for the distribution of RE_j over the grid points for each month and level (not shown). Generally the results for the lower and upper quartiles, the minimum, the median and the maximum were consistent with the results for RE_{areal} . Because of the skewed distribution of RE_j , the median was generally much higher than that for RE_{areal} .

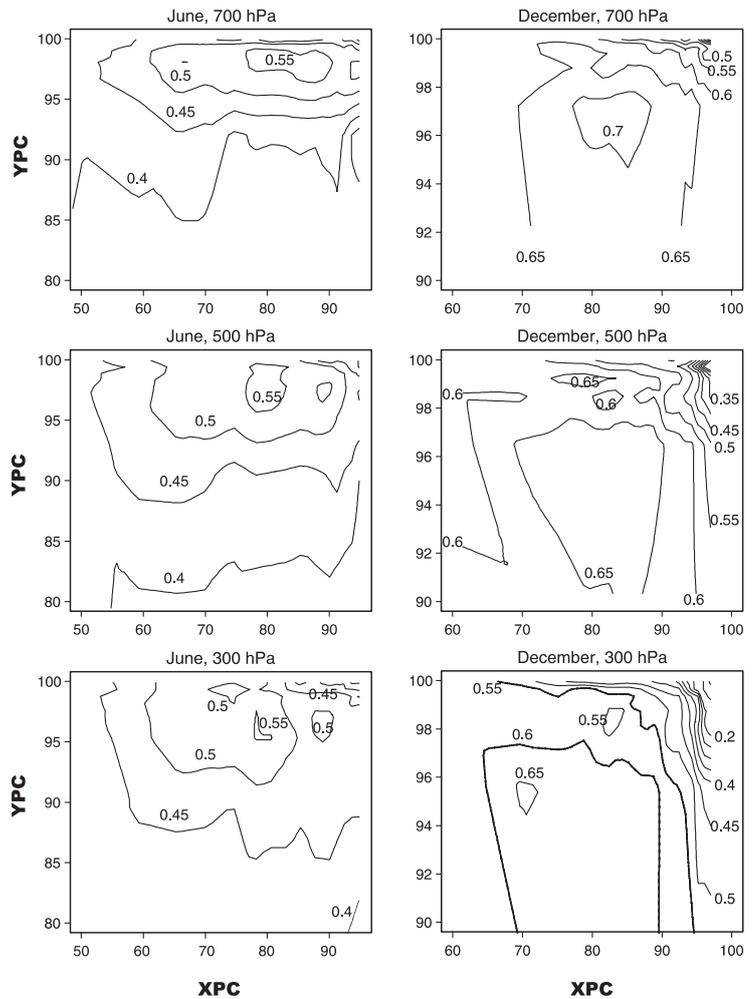


Fig. 2. RE_{areal} depending on the PC truncations for each level and for the cross-validation experiments. The x and y axes denote the proportion of shared variance in % by the retained XPCs and YPCs. (Note the slightly different scales on the x and y axes between June and December)

3.2. Weighting schemes

In order to test the sensitivity of the skill to the data weighting prior to the CCA, a cross-validation experiment was conducted for January with the same method as shown in Table 2 and Fig. 2. For this experiment the grid points and stations of the predictor and predictand data sets were neither weighted nor standardized prior to the calculations of the covariance matrix. Therefore, in Eq. (A3) the last option was applied.

The results of this experiment are presented in Table 3. The number of optimally retained XPCs and YPCs and related shared variances differed somewhat to the results in Table 2, but the skill (RE_{areal}) was much lower in the experiment without weighting than for the experiments with a weighting scheme. The difference was around 15%, which is much more than the loss

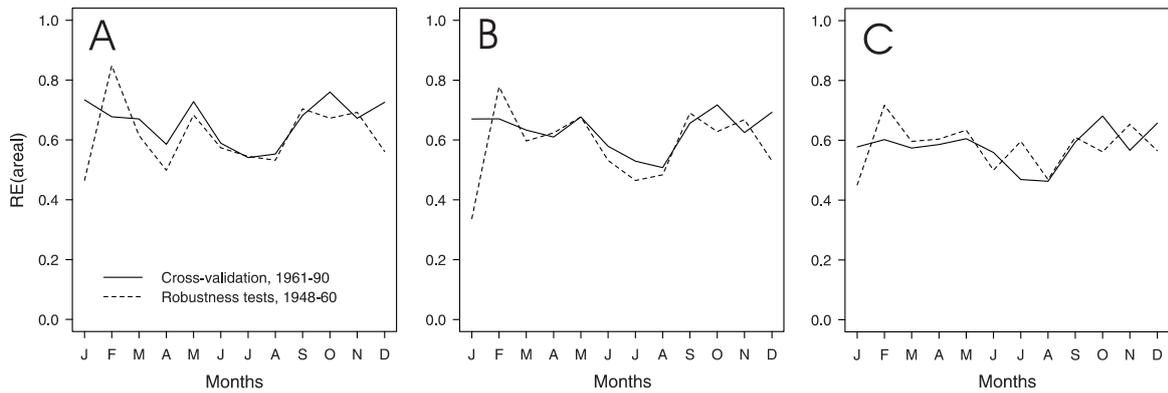


Fig. 3. Comparison of RE_{areal} for each month between the cross-validation experiments and the robustness tests. (A) Z_{700} , (B) for Z_{500} and (C) for Z_{300}

of skill if a suboptimal combination of XPCs and YPCs was considered in the experiments with weighted data.

The performance of the optimal PC truncations with no weighting in Table 3 was ranked among the results

with a weighting prior to the CCA. On Z_{700} this led to the 58th best model out of 195 tested combinations, the 45th best model for Z_{500} and the 33rd best model for Z_{300} . This means, that the best CCA models without

weighting of the data were not competitive with the results of the weighted data. Regularly gridded data fields should be areally weighted in order to account for the increasing network density in higher latitudes. This helps to level out the over-representation of the northern regions in the CCA models. The standardization of T and P helps to filter out the dominance of stations with high variability.

The comparison of the sensitivity 'topographies' (as in Fig. 2) of the experiments with and without a weighting scheme showed that an adequate and reasonable weighting scheme together with the optimal PC truncation is the key to achieving a high reconstruction skill.

Livezey & Smith (1999b) found that the PC truncation is more important than the weighting scheme after recalculating the specifications of Barnston (1994) and Barnston & Smith (1996). Here, a somewhat different picture was found. However, we agree that—with an adequate weighting scheme—the PC truncation is very important.

3.3. Robustness tests and spatial performance of the reconstruction

The results of cross-validation experiments could be affected by low-frequency variability (von Storch & Zwiers 1999). Data independency cannot be fully guaranteed

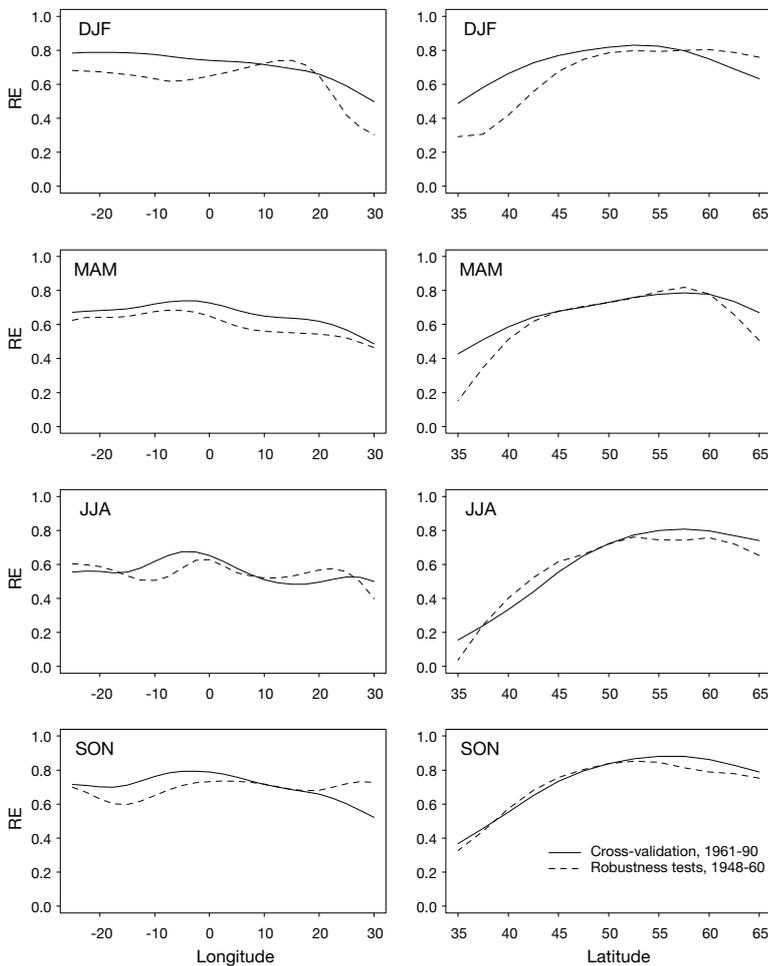


Fig. 4. Comparison of the meridional mean RE (left-hand side) and zonal mean RE (right-hand side) for each season between the cross-validation experiments and the robustness tests for Z_{700}

Table 3. The number of PCs retained for the subsequent CCA, their shared variance and RE_{areal} for the models with optimal PC truncations for all 3 levels in January without weighting prior to the PC truncation experiments (RE_{areal}^*) and with weighting (RE_{areal}^{**})

Level	XPC		YPC		RE_{areal}^* (%)	RE_{areal}^{**} (%)
	No. of PCs	Shared variance (%)	No. of PCs	Shared variance (%)		
Z_{700}	8	81.4	10	99.5	56.7	73.4
Z_{500}	8	81.4	4	92.0	56.0	67.0
Z_{300}	8	81.4	6	96.1	41.1	57.8

and the skill of the reconstructions might be overestimated. Therefore, a careful clarification should be undertaken. A second series of experiments was thus conducted to estimate the robustness of the previously discussed results. The models were fitted in the period 1961–1990 (see Table 1) and verified on independent data sets for all levels and months in the period 1948–1960. Fig. 3 shows RE_{areal} of the robustness tests for all 3 levels together with the results of the cross-validation experiments.

It can clearly be seen that the RE_{areal} is similar for the 2 series of experiments for all levels and most of the months. For some months, all levels showed an improvement of skill in the period 1948–1960. Considerably lower RE_{areal} were found in December and January. It is not clear what caused this loss of skill. A thorough examination of this phenomenon showed that a substantial loss of information was experienced in January mainly over the Iberian Peninsula and in December over the southeasterly part of the reconstructed domain (over Greece and Turkey). One possible explanation might be that in the period 1948–1960 (robustness tests) the gph of the Azores High was generally lower and its centre more to the west than in the period 1961–1990 (cross-validation experiments), leading to a poorer reconstruction of this feature between 1948 and 1960.

Figs 4, 5 & 6 reveal the meridional and zonal mean RE for each season and level for both series of experiments (cross-validation and robustness experiments). The meridional and the zonal mean RE were generally in good agreement. For Z_{700} and Z_{500} a slight decline in skill towards the east (continental Europe) and a substantial loss of skill towards the south (Mediterranean region) was observed for every sea-

son. For Z_{300} , a poor skill was achieved for all seasons for latitudes south of 45° N and north of 60° N, while a pronounced meridional drop towards the east was found mainly for the winter months (DJF).

The general decrease of skill towards the lower latitudes, which can be observed in any of the experiments conducted (Figs 4, 5 & 6) at all levels, was also reported by Klein & Dai (1998) for the western hemisphere. In these regions the variability in the reconstruction periods is lower than in the observation periods.

This situation is reflected in RE_{areal} , thus resulting in lower values and, hence, in reduced reconstruction skill. Fig. 1 shows that the southern part of the considered region has a sparse station network. Rigorous tests on independent data conducted by Schmutz et al. (2000) and Luterbacher et al. (1999) for NAO index

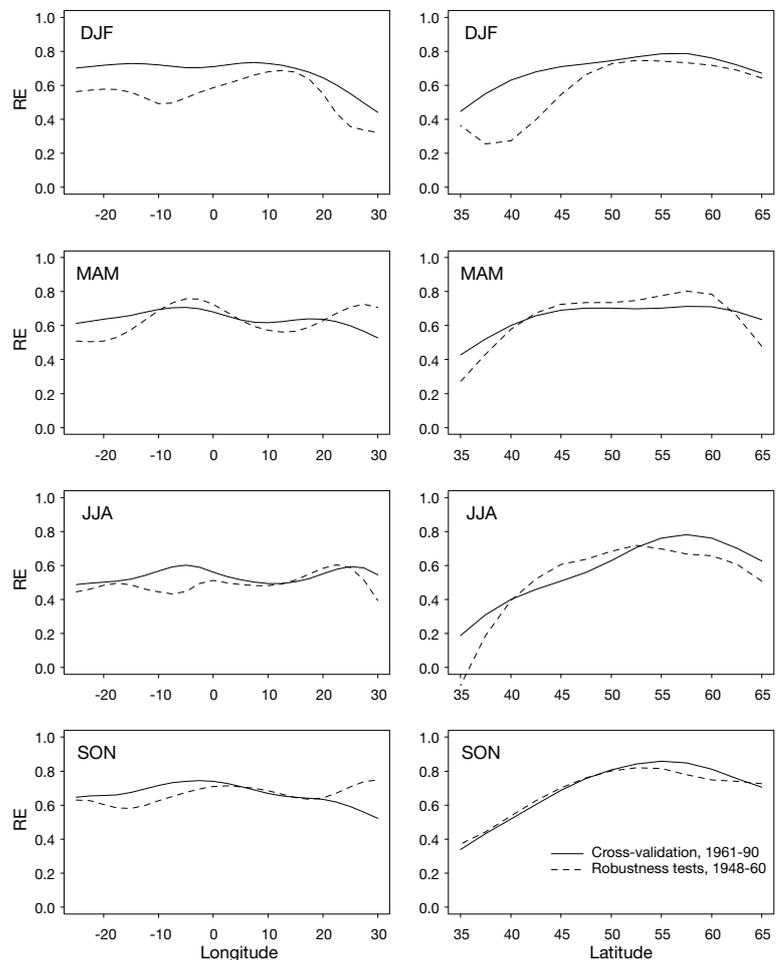


Fig. 5. Comparison of the meridional mean RE (left-hand side) and zonal mean RE (right-hand side) for each season between the cross-validation experiments and the robustness tests for Z_{500}

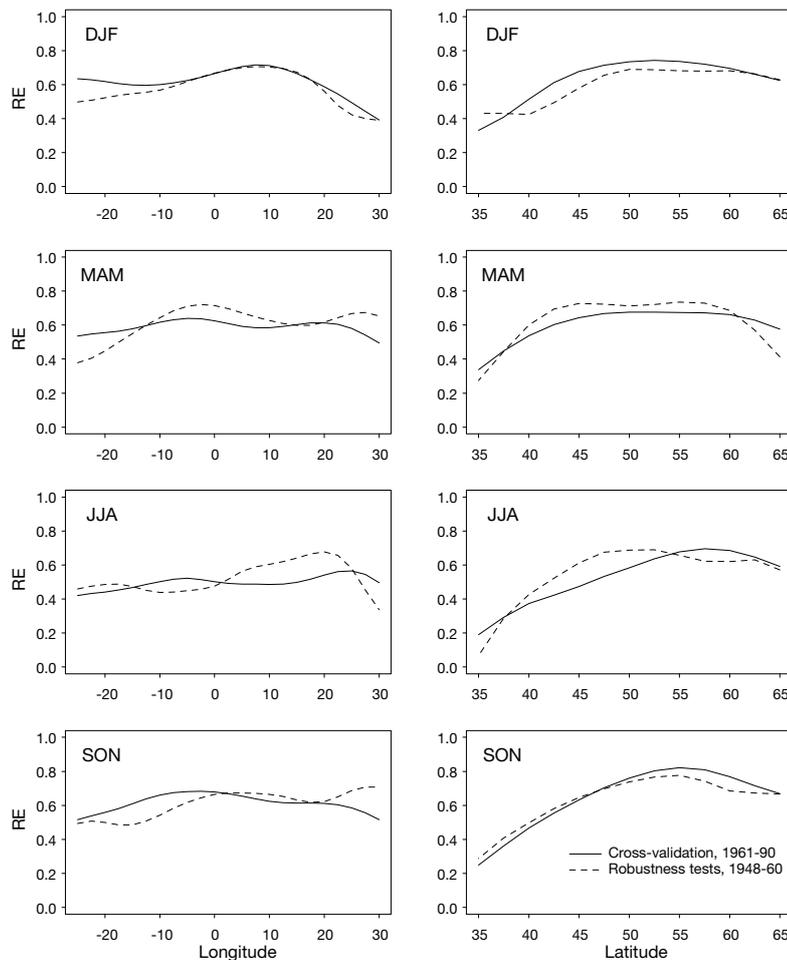


Fig. 6. Comparison of the meridional mean RE (left-hand side) and zonal mean RE (right-hand side) for each season between the cross-validation experiments and the robustness tests for Z_{300}

reconstructions showed that a high-density predictor network with complementary climate information is an important factor for better specification results. Consequently, we see potential for a better specification of the gph with a denser data network in the Mediterranean area. Since mainly the subtropical area and eastern Europe are affected by poorer reconstructions, other explanations are also possible. The signature of the mid-latitude planetary waves, which is captured by the predictors used is less clear at lower latitudes. Therefore, the predictors used (T , P and SLP) contain only parts of all the information. In autumn, winter and spring a certain loss of information due to decoupling effects between tropospheric variables and surface data has to be taken into account. Additional data that cope with these problems are required to obtain a specific improvement; data for wind direction and speed as well as temperature data from mountain stations would potentially improve the skill in the discussed regions.

The geographical distribution of the RE_j for each level is presented in Figs 7 & 8 for the months with the best (Fig. 7) and worst (Fig. 8) reconstruction skills according to the cross-validation experiments (left-hand side). The RE_j of the corresponding months for each level for the robustness tests can be found on the right-hand side of the figures. For both, independent of high or low RE_j values, the region with the highest reconstruction skill was generally the northwestern part of Europe, including the British Isles and the Eastern North Atlantic. This was confirmed for all levels and months (not shown). The spatial patterns of the reconstruction skill for a given month were consistent for each level (see Fig. 7 for October; the same was also observed for the other months, not shown). The highest skill was observed for Z_{700} followed by Z_{500} and Z_{300} .

Even though the Z_{700} specifications by Klein & Dai (1998) are not directly comparable to our results, since their reconstruction was for a different region, a rough comparison by means of the root-mean-square error can be conducted. Table 4 shows that the results from Klein & Dai (1998) were in fact of the same order of magnitude as our Z_{700} specifications. Somewhat better results were achieved with the presented study, but both studies showed a marked increase in the root-mean-square error in wintertime.

3.4. Principal mode of low-frequency variability

For the period 1901–1947, the gph fields were reconstructed according to the optimal results of Table 2. For the winter months a combined PCA was conducted for the reconstructions (1901–1947) together with the measured data (1948–1998) for all levels (Z_{700} , Z_{500} and Z_{300}). The results of the first PC from a combined PCA over the whole period (1901–1998) are given in Figs 9 & 10. Fig. 9 reveals the loading patterns of the first eigenvector, and Fig. 10 shows the corresponding seasonally averaged time coefficients. The first PC revealed a consistent dipole structure on each level.

The same analysis was also conducted on the split data sets for the reconstruction period (1901–1947) and observation period (1948–1998). The same spatial patterns were found for all periods (not shown).

The geographical distribution of the first EOF (Fig. 9) clearly shows the signature of the AO over the Euro-

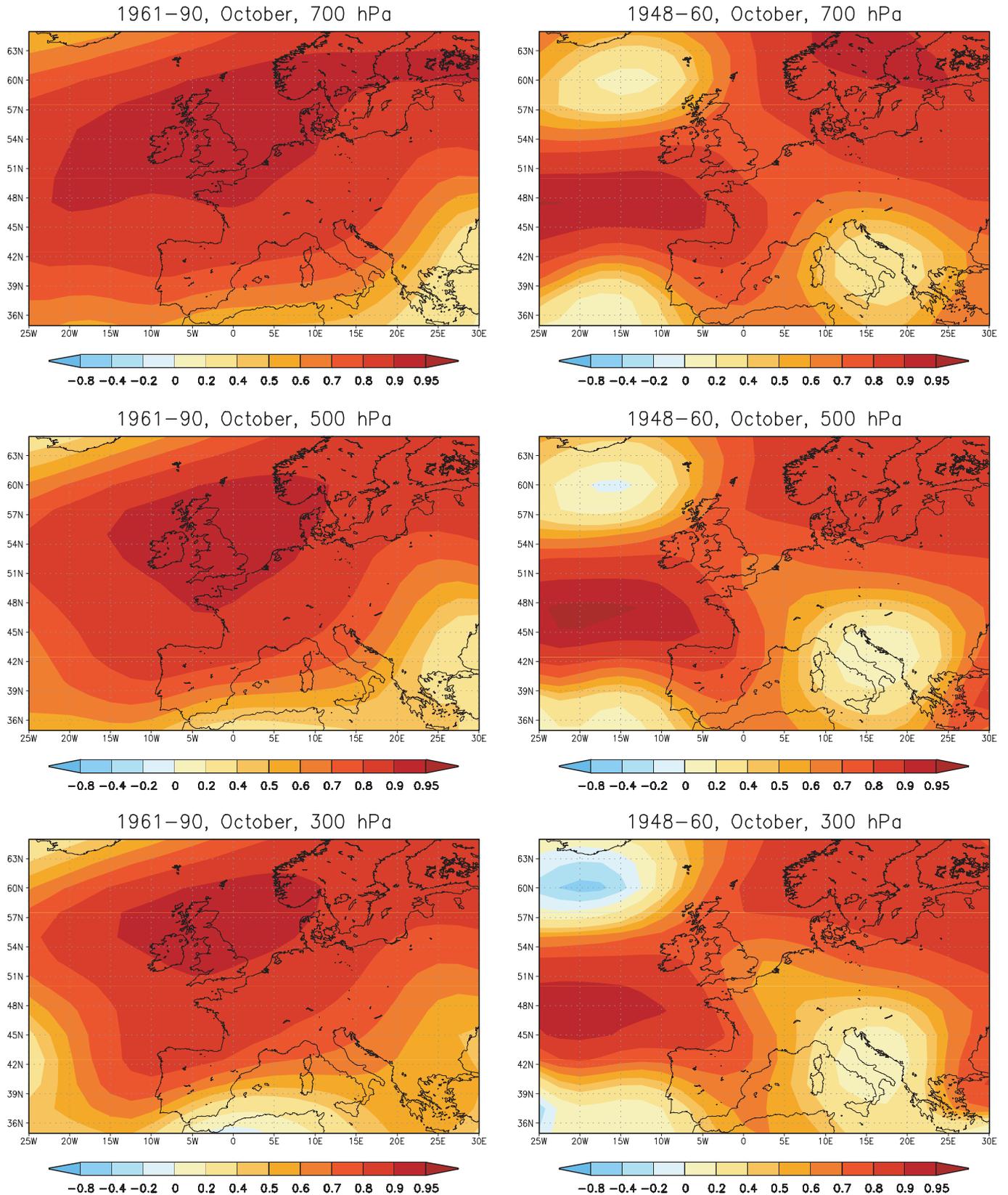


Fig. 7. Geographical distribution of the RE_i for the months with the best reconstruction skill in the cross-validation period, 1961–1990 (left-hand side), for each level. The corresponding results of the robustness tests (1948–1960) are given on the right-hand side

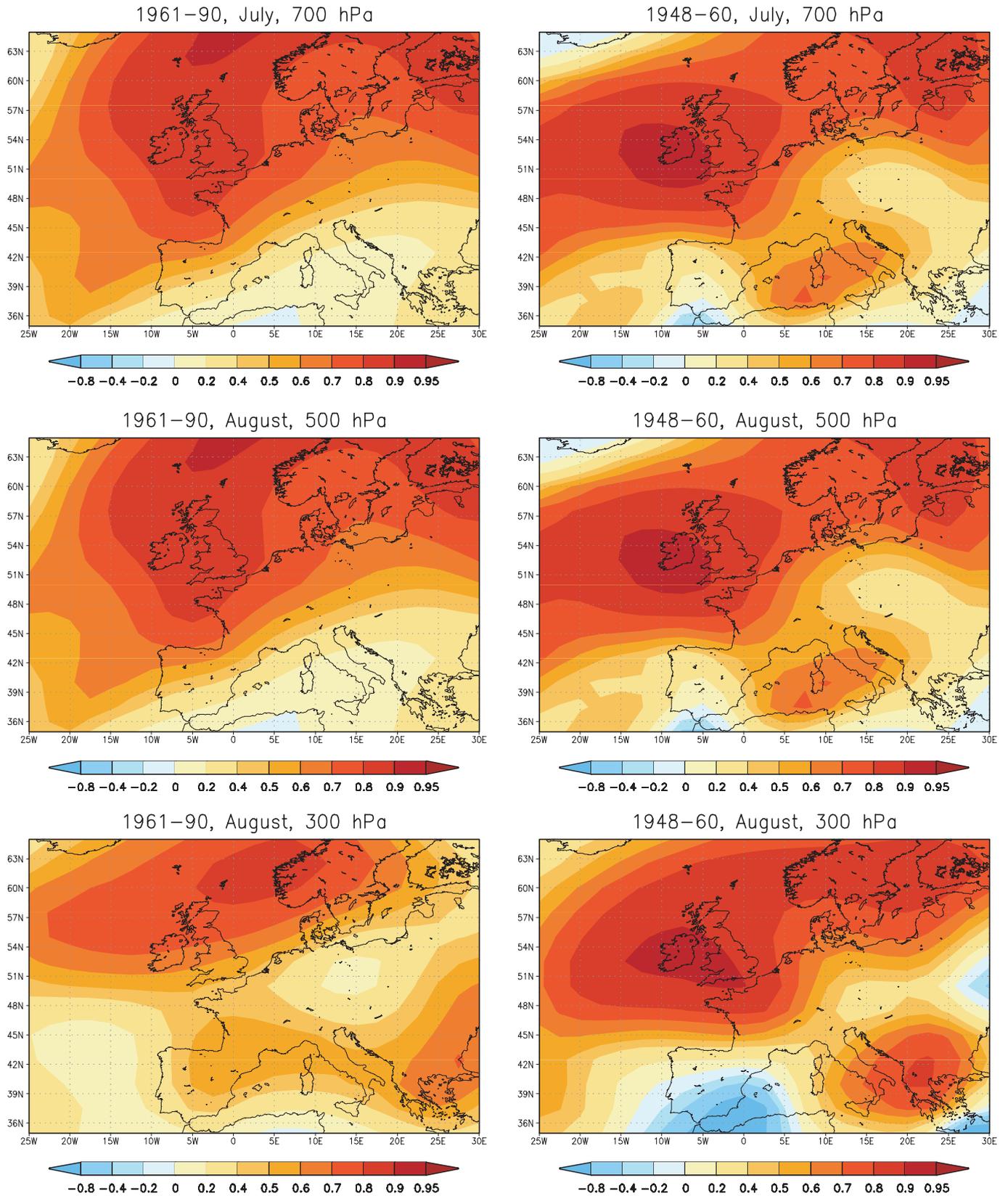


Fig. 8. Geographical distribution of the RE for the months with the worst reconstruction skill in the cross-validation period, 1961–1990 (left-hand side), for each level. The corresponding results of the robustness tests (1948–1960) are given on the right-hand side

Table 4. Comparison of the root-mean-square error (m) for Z_{700} reconstructions. The left column shows the results for the western hemisphere obtained by Klein & Dai (1998). The right column shows the results of our study for the European and Eastern North Atlantic area for the cross-validation period

Months	Klein & Dai (1998)	This study
DJF	30	26.6
MAM	26	19.6
JJA	24	14.3
SON	26	17.2
Year	26.5	19.4

pean and Eastern North Atlantic sector (Thompson & Wallace 1998, 2000, Brönnimann et al. 2000a, Wallace 2000). The first PC can therefore be interpreted as the AO time series.

The time series of the first PC indicated a broad range of variations with episodes of strongly negative values (tendency for decreased southwesterlies over western Europe) mainly in the middle of the century. In the last 3 decades a strong tendency towards positive values was observed. Obviously, these decades showed the behaviour of the main wintertime tropospheric mode over Europe, which is unique in the whole 20th century.

The first PC was analyzed for its spectral properties using a wavelet analysis. The wavelet power spectrum showed distinct interannual to decadal variability, but no persistent periodicity was found over the whole period. Only a few peaks reached a significance level of 10%. For the first PC, periodicities were found with weak spectral power at 2.5, 6–8, 10–16 and 30–40 yr (Fig. 11). The 10% significance level of the spectral power showed a patchy picture. Appenzeller et al. (1998) and Luterbacher et al. (1999) described the intermittent character of the NAO. This also holds for the first wintertime PC of the gph fields during the whole 20th century.

4. CONCLUSIONS AND OUTLOOK

Considering the assumption of stationary relations between the surface variables and the tropospheric circulation, it was possible to accurately reconstruct tropospheric gph fields in the European and Eastern North Atlantic region in the 20th century based on surface data (T , P and SLP). The reconstructions were robust, and the promising cross-validation results were confirmed in experiments for an independent period. Limitations in the reconstructions were found in the subtropics and in the southeastern part of the domain. A loss of skill was also observed with increasing height

of the reconstructed level. An adequate weighting (e.g. areal weighting for gridded data) of the predictors and predictands should be applied prior to the PCA.

Prior to the fitting procedure a prefiltering of the predictors and predictands has to be conducted. The cumulative shared variances of the optimally retained YPCs should be very high (on the order of 95 to 99%), while the values for the optimally retained XPCs should be generally lower than 90%. An empirical rule of thumb that not more than 90% of the shared variance of the predictors together with 98% of the predictands should be retained for fitting purposes in the same reconstruction could be applied.

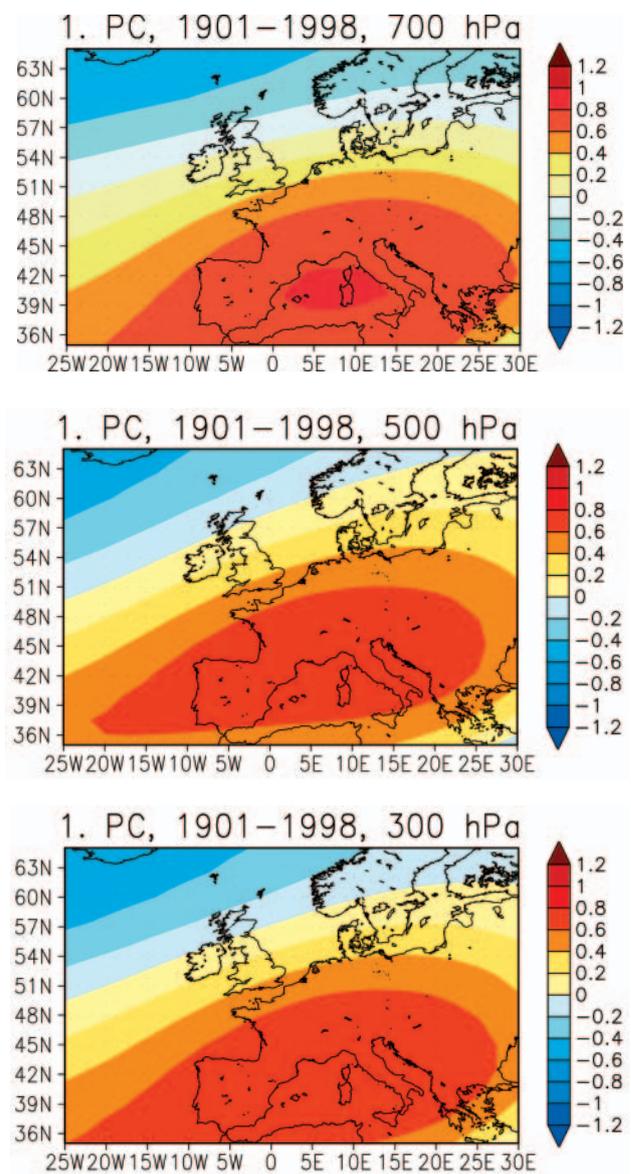


Fig. 9. Eigenvector of the first PC from a combined PCA between 1901 and 1998 (DJF). For the period 1901–1947 the reconstructed values were used. Units are arbitrary

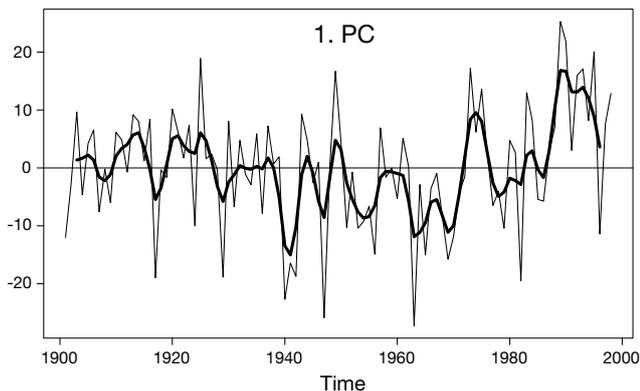


Fig. 10. Seasonally averaged (DJF) time series from the eigenvector of Fig. 9. The thick line denotes the low-pass-filtered time series (triangular 1-3-5-3-1). Units are arbitrary

The main mode of tropospheric variability for the winter months (DJF) of the whole 20th century strongly resembles the AO. A generally weaker meridional gradient—with episodes with very weak gradients—was observed in the first half and in the middle of the century. Moreover, the strengthening of the polar vortex in winter in the last low decades was also observed in the troposphere and is without precedent in the first half of the 20th century for the European and Eastern North Atlantic sector.

Signatures of low-frequency variability can be found in the combination of the reconstructed and observed data sets, but no preferred significant periodicity characterizes the main modes of variability.

Future attempts for reconstructions should seek to also include mountain station data, mainly in the medi-

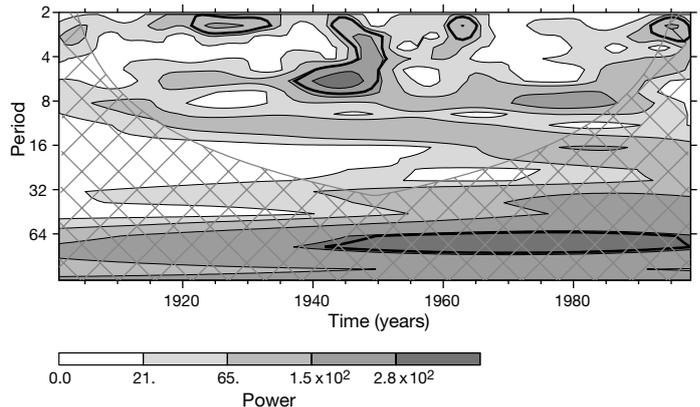


Fig. 11. Wavelet power spectrum with a Morlet wavelet base of the first PC from the combined PCA (Z_{700} , Z_{500} and Z_{300}) for the period 1901–1998 (winter averages, same as Figs 9 & 10). The thick contour line denotes the 10% significance level based on red noise processes with an empirical lag-1 autocorrelation of 0.3. Cross-hatched parts of the power spectrum are contaminated by 0-padding of the time series (reduced variance)

terranean area. The fairly good database for at least the second half of the 19th century and the good results found with this study leads us to believe that there is potential for a reconstruction of the gph fields back to the mid-19th century.

In future studies, the low frequency variability of the reconstructed gph fields will be analyzed together with long-term total ozone time series in Europe and with long-term precipitation time series in the Alpine region.

The reconstructed data set is freely available from the first author upon request.

Appendix 1. Weighting schemes

Unless specifically noted, the gridded SLP, Z_{700} , Z_{500} and Z_{300} fields were areally weighted (see Eq. A1) and T and P were standardized Eq. (A2), prior to the calculations of the covariance matrix for the PCA (ϕ_j : latitude; s_j : empirical standard deviation of the grid point j). The areal weights of the gridded data are denoted as w'_j , and w''_j stands for the standardization. Eq. (A3) summarizes the weights for each grid point, j . There are $n = 1 \dots m$ predictor fields and 1 predictand field with $j = 1 \dots k$ columns (stations or grid points) The total variance for each field (VAR_n) is the sum of the weighted variances Eq. (A4) The sum of all field variances is denoted as VAR_{tot} in Eq. (A5). In order to equally weight the total variance of each of the predictor and predictand fields, Eq. (A6) was applied for the relative weight w_n^{rel} of each field. Each CCA model was fitted for 3 mo; therefore the weights ($w_n^{rel} w^*_j$) in Eq. (A7) were applied grid point-wise for the differences of the monthly anomalies $x'_{n,j}$ to the mean of each considered monthly field. The covariance matrix for the PCA was then calculated for the matrix \mathbf{x}'' .

$$w'_j = \sqrt{\cos \phi_j} \quad (\text{A1})$$

$$w''_j = \frac{1}{s_j} \quad (\text{A2})$$

$$w^*_j = \begin{cases} w'_j & \text{Latitude weighting} \\ w''_j & \text{Standardization} \\ w'_j w''_j & \text{Latitude weighting and standardization} \\ 1 & \text{No weighting} \end{cases} \quad (\text{A3})$$

$$VAR_n = \sum_{j=1}^k w^{*2}_j s^2_j \quad (\text{A4})$$

$$VAR_{tot} = \sum_{n=1}^m VAR_n \quad (\text{A5})$$

$$w_n^{rel} = \sqrt{\frac{1}{m} \frac{VAR_{tot}}{VAR_n}} \quad (\text{A6})$$

$$\mathbf{x}''_{n,j} = w_n^{rel} w^*_j (\mathbf{x}'_{n,j} - \bar{\mathbf{x}}''_{n,j}) \quad (\text{A7})$$

Acknowledgements. The authors wish to thank Mary Brown for proofreading the English text. This work was made possible by the Swiss National Science Foundation.

LITERATURE CITED

- Appenzeller C, Stocker TF, Anklin M (1998) North Atlantic Oscillation dynamics recorded in Greenland ice cores. *Science* 282:446–449
- Barnett T, Preisendorfer R (1987) Origins and levels of monthly and seasonal forecasts skill for United States surface air temperature determined by canonical correlation analysis. *Mon Weather Rev* 115:1825–1850
- Barnston AG (1994) Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J Clim* 7: 1513–1564
- Barnston AG, Smith TM (1996) Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J Clim* 9:2660–2697
- Bretherton CS, Smith C, Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. *J Clim* 5:541–560
- Brönnimann S, Luterbacher J, Schmutz C, Wanner H, Staehelin J (2000a) Variability of total ozone at Arosa, Switzerland, since 1931 related to atmospheric circulation indices. *Geophys Res Lett* 27:2213–2216
- Brönnimann S, Schmutz C, Luterbacher J, Staehelin J (2000b) Total ozone and climate variability over Europe. In: NASDA (ed) *Atmospheric ozone—Proceedings of the Quadrennial Ozone Symposium, Sapporo. Supplement*, p 765–766
- Fraedrich K, Ziehm C, Sielmann F (1995) Estimates of spatial degrees of freedom. *J Clim* 8:361–369
- Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* 269:676–679
- Jones PD, Wigley TM, Briffa KR (1987) Monthly mean pressure reconstruction for Europe (back to 1780) and North America (back to 1858). Technical Report 37, DOE, US Dept of Energy, Washington, DC
- Kalnay E, Kanamitsu M, Kistler R, Collins W, and 18 others (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471
- Kington JA (1975) The construction of 500-millibar charts for the eastern North Atlantic-European sector from 1781. *Meteorol Mag* 104:336–340
- Klein WH, Dai Y (1998) Reconstruction of monthly mean 700-mb heights from surface data by reverse specification. *J Clim* 11:2136–2146
- Livezey RE, Smith TM (1999a) Considerations for use of the Barnett and Preisendorfer (1987) algorithm for canonical correlation analysis of climate variations. *J Clim* 12: 303–305
- Livezey RE, Smith TM (1999b) Covariability of aspects of North American climate with global sea surface temperatures on interannual to interdecadal timescales. *J Clim* 12:289–302
- Luterbacher J, Schmutz C, Gyalistras D, Xoplaki E, Wanner H (1999) Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys Res Lett* 26:2745–2748
- Luterbacher J, Rickli R, Tinguely C, Xoplaki E, and 29 others (2000) Reconstruction of monthly mean sea level pressure over Europe for the Late Maunder Minimum period (1675–1715) based on canonical correlation analysis. *Int J Climatol* 20:1049–1066
- Murphy AH, Epstein ES (1989) Skill scores and correlation coefficients in model verification. *Mon Weather Rev* 117: 572–581
- Peterson TC, Vose RS (1997) An overview of the global historical climatology network temperature database. *Bull Am Meteorol Soc* 78:2837–2850
- Santer BD, Wigley TML, Barnett TP, Anyamba E (1996) Detection of climate change and attribution of causes. In: IPCC (eds) *Climate change 1995. The science of climate change*. Cambridge University Press, p 407–443
- Schmutz C (in press) Seasonal precipitation trends of the twentieth century in the Alpine region and its vicinity: reanalysis and processes. *Clim Res*
- Schmutz C, Luterbacher J, Gyalistras D, Wanner H (1999) Reconstruction of monthly 700 hPa and 500 hPa geopotential height fields over Europe and the Eastern North Atlantic: method development and validation. *Geophys Res Abstr* 556
- Schmutz C, Luterbacher J, Gyalistras D, Xoplaki E, Wanner H (2000) Can we trust proxy-based NAO index reconstructions? *Geophys Res Lett* 27:1135–1138
- Smith TM, Livezey RE (1999) GCM systematic error correction and specification of the seasonal mean Pacific-North America region atmosphere from global SSTs. *J Clim* 12: 273–288
- Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys Res Lett* 25:1297–1300
- Thompson DWJ, Wallace JM (2000) Annular modes in the extratropical circulation. Part I: month-to-month variability. *J Clim* 13:1000–1016
- Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79:61–78
- Trenberth K, Paolino DA (1980) The Northern Hemisphere sea-level pressure data set: trends, errors and discontinuities. *Mon Weather Rev* 108:866–872
- von Storch H, Zwiers FW (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge
- Vose RS, Schmoyer RL, Steurer PM, Peterson TC, Heim R, Karl TR, Eischeid JK (1992) *The Global Historical Climatology Network (GHCN): long term monthly temperature, precipitation, sea level pressure, and station pressure data*. Technical Report 53, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN
- Wallace JM (2000) North Atlantic Oscillation/annular mode: two paradigms—one phenomenon. *Q J R Meteorol Soc* 126:791–805
- Wang X, Sheng SS (1999) Estimation of spatial degrees of freedom of a climate field. *J Clim* 12:1280–1291
- Wilks DS (1995) *Statistical methods in the atmospheric sciences. An introduction*. Academic Press, New York

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: April 24, 2000; *Accepted:* October 14, 2000
Proofs received from author(s): August 9, 2001