Analysis of gridded sea level pressure and 2-meter temperature for 1873-1998 based on UEA and NCEP re-analysis II.

R.E. Benestad

DNMI, June 12, 2001

Reg Clim

Klima: 03/00

Contents

1 Introduction

Long time series are essential for empirical climate studies, however, many (incomplete) time series suffer from gaps of missing data and regions with no data coverage. Gridded sea level pressure (SLP) and 2-meter temperature (T2M) for the region 90°W-60°E and 30°N-80°N have been synthesized from different data sets in order to obtain more complete data sets. The synthesis was based on projection (P) (regression analysis) and optimal interpolation (OI) between large-scale spatial patterns derived from the National Center for Environmental Prediction (NCEP) reanalysis II product and gridded observations. The end-products are quality controlled data sets with no missing-data gaps.

2 Data and Method

2.1 The data sets

The data sources for the synthesized SLP data set are gridded SLP and T2M obtained from University of East Anglia's (UEA) internet site[1], COADS, NCAR (SLP only), and the NCEP reanalysis II.

The UEA SLP data is described in *Jones* (1992), but the data archive originates from the UKMO. This data set covers the period: 1873-1995. The UEA data[2] are based on several original sources of observations. The period from January 1873 to December 1898, with the exception of the period Sep-1882 to Aug-1883, is based on the *Deutsche Wetterdienst* morning charts 1 (German tapes) and is described in a Branch Memorandum No. 31 (Acquired 1969). The data in the 1882-1883 interval have been obtained from UKMO Synoptic Climatology Staff. The observations for the period January 1899-December 1939 are taken from the US Weather Bureau Extended Forecast Division (US tapes, 1200 hour charts, UK Branch Memorandum No. 25 and No. 34, Acquired 1966-1967 and 1970). The observations for the period January 1940 to December 1948 have been derived from German Offenbach 0001 hour charts (German tapes, acquired 1973), and data from the US Weather Bureau Extended Forecast Division (US tapes, 1200 hour charts, UK Branch Memorandum No. 25) are again used between January 1949 and December 1965. The observations for the period January 1966 to August 20th 1975 are taken from CFO 0001 hour charts by the UKMO Synoptic Climatology Staff, but since August 21st 1975 the data are extracted from

---

[1]http://www.cru.uea.ac.uk/cru/data/pressure.htm

[2]Which are similar to the "WMO" data at DNMI obtained from the UK Met. office

CFO model (10 levels between 1975 and 1982, and 15 levels after 1982) charts of 0001 hour data by the UKMO Synoptic Climatology programs.

The UEA temperature set spans the period 1856-1998, and is described by *Jones* et al. (1998).

The Comprehensive Ocean-Atmosphere Data Set (COADS[3]) is described in *Slutz* et al. (1985), but a brief description is also given by *Benestad* (1998b). The data set consists of historical records of surface marine data for the period Jan 1854 to Dec 1992 compiled by the National Climatic Data Center (NCDC), Environmental Research Laboratories, the cooperative Institute for Research in Environmental Sciences, and the National Center for Atmospheric Research (NCAR). The COADS data include fully quality controlled reports, and are primarily obtained from maritime observations by ships-of-opportunity, and gridded at a spatial resolution of $2° \times 2°$. The COADS data may be regarded as independent to the UEA data. The COADS data are not corrected for inhomogenouities caused by changes in the instrumentation, observational practices and ship type according to *Deser & Blackmon* (1993).

The NCAR ds010.0 data[4] contain daily observations spanning the period 1899-1998, and is described by *Benestad* (1998b). The monthly mean values are estimated from the readings made made at 13:00hZ between 1899 and 1939, and at 12:00hZ after 1939. The entire data record is based on different periods with various sources merged together (*Benestad*, 1998b). *Benestad* (1998b) reported some problems with the NCAR data in the sparsely sampled regions of the Arctic (Bjørnøya at 74.3°N - 19°E) before 1939. These bad data are associated with historical maps from the U.S. National Climate Center.

The NCEP reanalysis II are obtained from the NCEP CD-ROM release in July 1998, and the data are described by *Benestad* (1999a).

Station data from the NACD project (*Frich* et al., 1996) and The Norwegian Meteorological Institute's (DNMI) archives have been used as predictands and independent data used for verification purposes.

2.2   Step 1: Combining incomplete data data sets

The UEA, NCAR, and COADS SLP data sets were interpolated to a common $5° \times 5°$ (lon=[90°W,95°, ... 60°E], lat=[30°N, 35°N,80°N) by using the nearest neighbour point. The nearest neighbour-point scheme was chosen because this was least affected by missing data gaps, which may bias the valid data in

---

[3]URL: http://rainbow.ldgo.columbia.edu/data/NASAentries/nasa3276.html

[4]http://www.scd.ucar.edu/dss/catalogs/gridded.html

a cubic-spline scheme or result in large missing data gaps. The interpolation scheme for the NCEP reanalysis, which had no missing data, was a cubic-splines method.

In the synthesis of a combined data field based on all available observations from several data sources, the departures (anomalies) from the annual mean or a seasonally varying climatology was used. The departures from the mean value was estimated at each grid point by subtracting the temporal average of all the respective valid data values.

The synthesized SLP field was obtained by several steps of processing. The first step involved collecting available valid data from the various sources for each grid box. When more than one source made a contribution, then the average of these was used. In the case when only one data set contained a valid data for the grid box, then only this value was included. SLP data were available from the COADS and UEA for the period 1873-1979, and NCAR from 1899-1998. The period 1979-1995 consisted only of the UEA and NCAR SLP data. Missing data were flagged and excluded from the analysis. All SLP values outside the range of 950-1050hPa were flagged as missing (NaN). Furthermore, similar products were also constructed based on UEA SLP only. Comparisons between the various products suggested that the products based on a single source (UEA) were superior to the "multi-source" products, so the remaining discussion will mainly focus on the "single-source" products (data derived from UEA and NCEP data).

The T2M data were processed in a similar way as the SLP data, but were only based on the UEA and COADS data. Trials with equal weights given to the UEA and COADS data, when both had valid data for a given grid box, again suggested inferior quality of the final product compared to similar analysis based on UEA alone. Part of this problem may be related to the observation made by *Deser & Blackmon* (1993) that the COADS temperatures probably are not homogeneous. For this reason, the observations were taken from only one source (UEA). The UEA temperatures were used up to 1998 and consisted of seasonally dependent anomalies (the original UEA data contained anomalies from the 1961-1990 climatology), as opposed to departures from an multi-annual mean value in the case of SLP.

2.3 Step 2: Computing spatial modes from NCEP reanalysis.

Orthogonal modes were estimated from a singular value decomposition (SVD, see *Press* et al. (1989) and *Strang* (1995)) of the NCEP SLP and T2M anomaly data, after these had been interpolated onto a $5° \times 5°$ grid. The
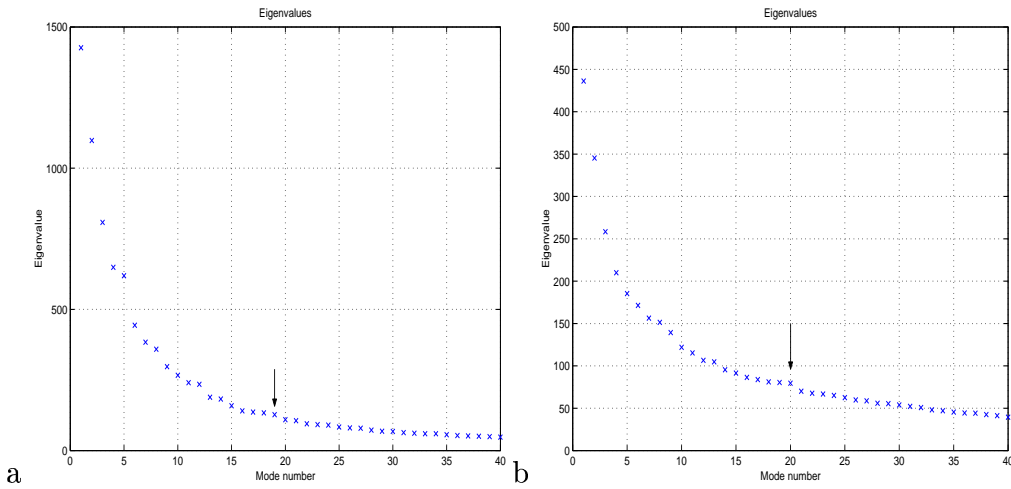
Figure 1: *The eigenvalues associated with the 40 leading modes. Modes with order 19 and higher were assumed to be due to noise.*

SVD essentially returns the eigenvectors of the covariance matrix which are similar to the EOFs, but without a geographical weighting. The SVD was applied to the entire data set and for all seasons, and only a subset of the modes was retained for further analysis.

2.3.1 The NCEP SLP spatial modes

The number of modes chosen for the regression to the evolving SLP patterns was 19. This choice was made from the inspection of Figure 1a, which suggests that the higher order modes have small variance and are associated with a low signal-to-noise ratio. Thus this cut-off may filter out much of the unwanted noise.

Figure 2 presents the first 3 and the last spatial SLP mode. As expected, the leading modes are associated with large-scale (basin-wide) SLP structures, and the higher order modes are characterized by more geographical detail and noise. Mode 19 in Figure 2d shows the smallest spatial scales that are described by this set of spatial modes, and since these are the basis for the reconstruction of the SLP, mode 19 represents the smallest spatial scale in the P-product (best-fit maps).

The leading mode contains the north-south dipole signature of the NAO and the third mode describes the east-west pressure dipole ("Greenland-Fennoscandinavia Oscillation", GFO) responsible for northerly and southerly geostrophic winds over parts of Scandinavia.
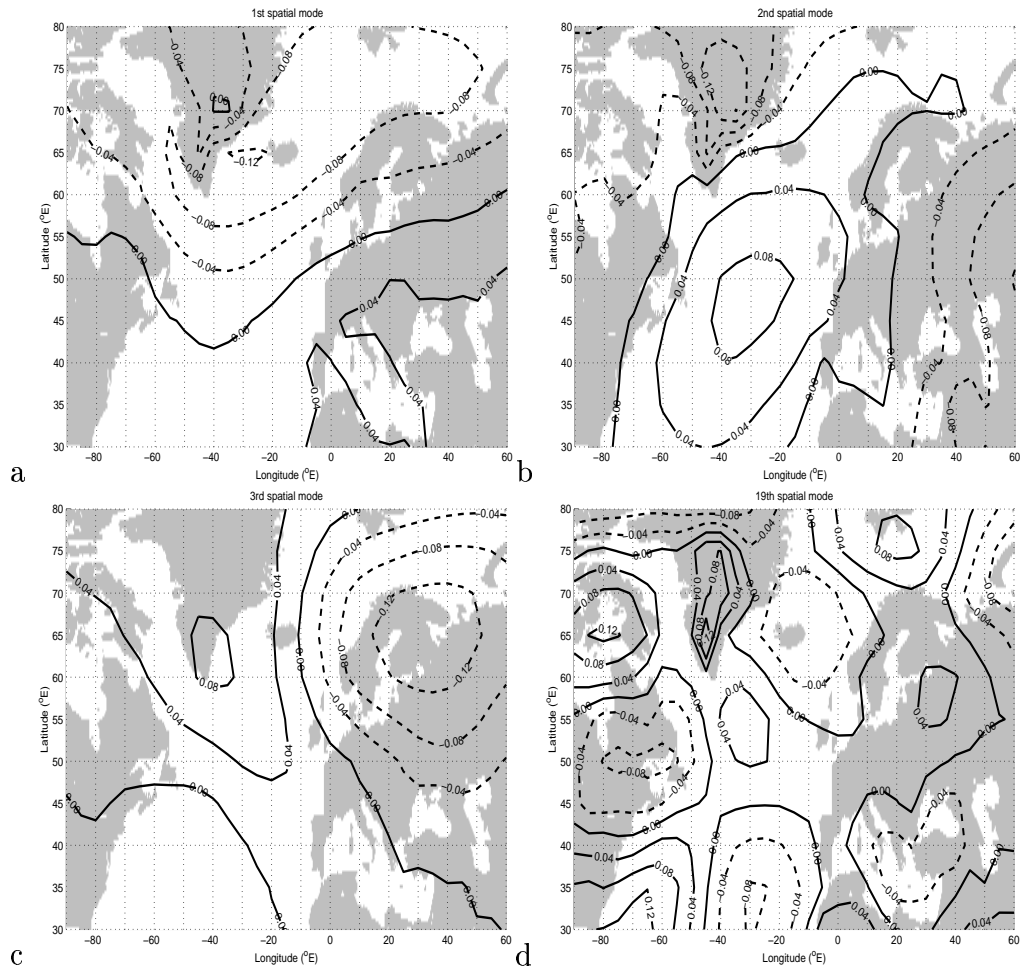
Figure 2: *The spatial SLP patterns associated with three leading modes and mode 19.*
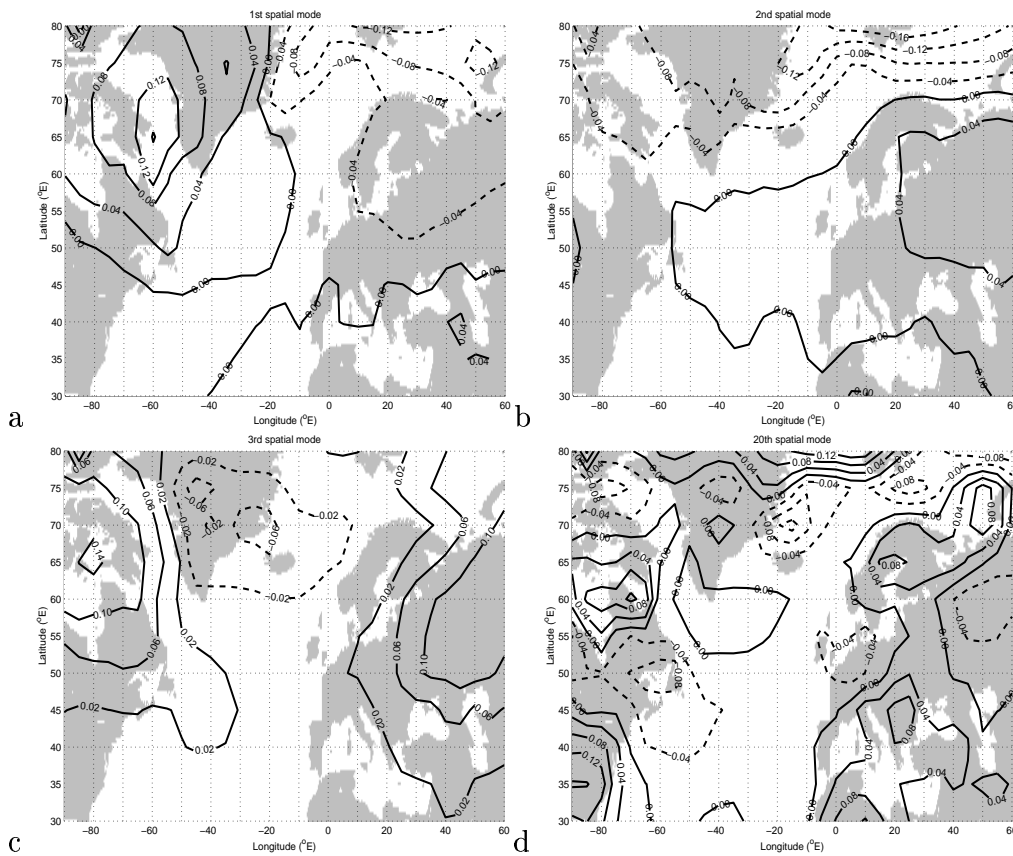
Figure 3: *The spatial T2M patterns associated with three leading and the highest spatial mode used in the regression analysis.*

2.3.2 The NCEP T2M spatial modes

The reconstruction of the best-fit maps T2M was based on the same method as the SLP. The eigenvalues shown in Figure 1b were used to estimate the number of spatial modes that were used in the regression analysis. The T2M reconstruction was based on 20 modes, as opposed to 19 for the SLP. Figure 1b indicates lower noise-levels for the 20 leading modes than for the higher modes, and it is expected that the T2M temperature fields may contain finer geographical details than SLP.

Figure 3 presents the first 3 modes and the highest spatial T2M mode used in the regression analysis. The leading mode is characterised by strong weights over the Labrador Sea, and may be related to the North Atlantic oscillation (NAO) (NB: these modes are *not* the same as EOFs, as these are not weighted according to grid box area, and hence variability at high latitude carries more weight than at lower latitude). The second mode describes a

north-south dipole pattern, and the third mode described a dipole structure with centres of action over Labrador and Iceland. The geographical details in mode 20 gives an indication of the smallest spatial scales that will be present in the T2M reconstruction.

2.4 Step 3: Filling in regions with no data.

The number of modes included in the regressional analysis was determined on the basis of signal-to-noise considerations. Both using low spatial resolution and the discarding of the higher order modes remove some of the small-scale climate signal. However, in the reconstruction going back to 1873, much of the detailed SLP and T2M information over sparse regions is lacking. Furthermore, small-scale geographical features may be difficult to distinguish from noise. Thus, in order to obtain a best possible homogeneous data fields dating back to 1873, one must sacrifice the fine geographical details for more reliable large-scale climatic patterns. In empirical downscaling of future global climate scenarios, it is the large-scale patterns which are important, as the fine geographical details described by the climate models tend to be unreliable (*Grotch & MacCracken*, 1991).

2.4.1 Regression of spatial modes onto available observations

The 2-D maps of SLP and T2M from the UEA data were re-arranged as 1-D vectors, $\vec{b}$, so that the missing data points were removed. Examples of maps of the incomplete data are given in the upper panels of Figure 4. One vector was produced for each time-step, and examples of such vectors are illustrated by the thick black curves in the bottom panel of the same figure. Data from the spatial NCEP modes at corresponding locations were collected in a matrix, X, where each mode is given in the columns and each row represents a particular location. The rows corresponding to locations with no valid data were removed from the matrix, and different matrices were used for the different times of observation. An ordinary least-squares regression was applied to the observations and the modes, $\hat{\vec{y}} = X\vec{b}$, in order to estimate coefficients that describe the best-fit superposition of the spatial modes for each time step. A new data set, henceforth referred to as the "best-fit maps" of the "P (projection) products" (*Kaplan* et al., 1998), was synthesized from the spatial modes weighted by the regression coefficients. Examples of SLP and T2M best-fit maps are shown in the middle panels of Figure 4 as well as by the blue curves in the lower panel. The red curve in the bottom panel is the residual, indicating errors associated with small-scale spatial structures.
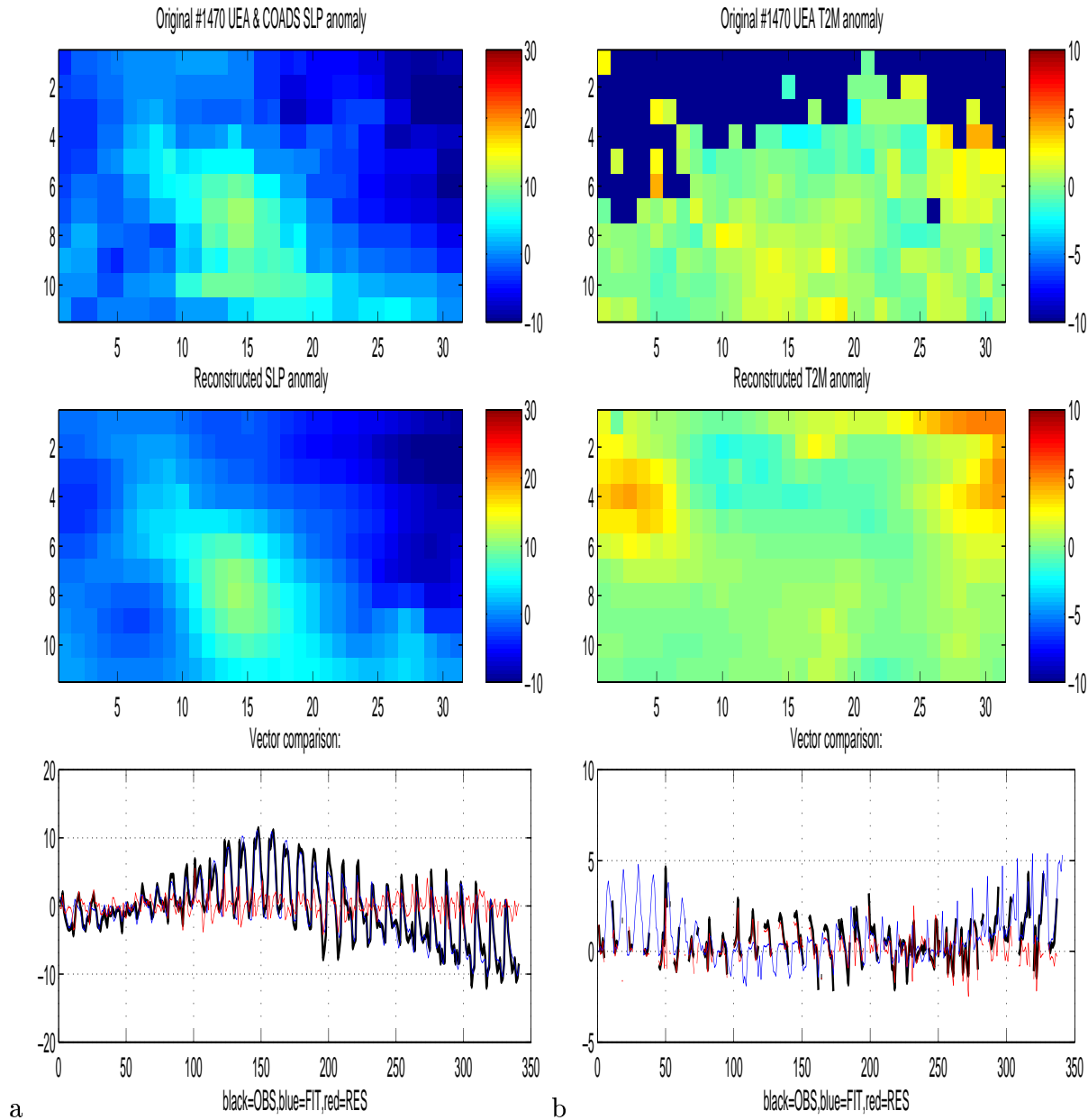
Figure 4: *Example of pattern regression for a) SLP and b) T2M. The upper panel shows the available observations and the middle panel shows the best-fit maps (P-products). The bottom panel displays the same information as the top ones, but in a different format: the black curve represents the available observations, the blue line shows the best-fit in terms of the spatial modes, and the red line shows the residual of the regression.*

Step 3 implicitly assumes that spatial modes for the period 1948-1998 are representative for the whole time period 1873-1998, and that the SLP spatial modes are stationary. There is, however, no guarantee that this is the case, and the result must therefore be checked against the observations. However, this assumption is made in empirical downscaling anyway, for which the use these fields are intended, so this assumption represents no additional limitation to empirical downscaling studies.

2.5   Step 4: Adjustment of the synthesized data.

It is assumed here that the SLP data must be stationary in the sense that the total atmospheric mass doesn't change appreciably over time, and because SLP is taken as $p_{sl} = \int_0^\infty \rho(z)gdz$. Furthermore, there is no reason to believe that the SLP temporal or spatial variance has changed significantly in a systematic fashion over the past century. The UEA and NCEP SLP-products were combined in a fashion that minimized the discontinuity between the two data sets, and the P-product was first adjusted (by the "adjust" field in the netCDF files) with a mean offset added to each grid point before they were scaled ("scaling" field). A similar offset was subtracted from the NCEP-product, and the net result ensured that the two data sets had the same mean values for the overlap-interval.

The best-fit maps of SLP and T2M were produced from the incomplete UEA data set between 1873 and 1947, but the NCEP reanalysis II were used after 1948 for the SLP and 1958 for T2M. The switch between the two sources may introduce discontinuities and inhomogenouities. Adjustments were therefore made to the synthesized SLP to avoid these problems. In order to ensure a smooth transition between the reconstructed and NCEP data at 1948, the SLP was forced to have the same mean value at each grid point for the overlapping period 1948-1995. Moreover, the SLP data were adjusted so that the mean values of the interval period for the P-products were taken as the average of the two mean fields for this period. The T2M data were not adjusted, as the most tests (discussed below) indicated no need for such correction. Furthermore, it is not clear whether such adjustments are warranted. The adjustment field was nevertheless calculated, and stored in the data file ("adjust": see table 1).

Data sets were also constructed based solely on the UEA observations (SLP_U_p.nc holding SLP(1873-1995) and T2M_U_p.nc holding T2M(1873-1998)). There is probably no way avoiding such stitching together of different data sets, as the UEA SLP observations already are obtained from the combination of various sources. One question is therefore whether a combination of the UEA and the NCEP re-analysis is any "worse" than using a SLP

Table 1: Example of Matlab code correcting and rescaling the fields

```
i41=find(tim == tim4(1)) - 1; % = 1020
N=find(tim4 == tim2(end)); % = 111
% Estimate scaling factor that ensures similar Std in regression data
% and NCEP.
%
% Adjust the regression data by a constant offset and scaling.
%
for i=1:i41;
T2M(i,:,:)=T2M(i,:,:) - 0.5*mt2ma + 0.5*mt2ma4;
end
for i=1:nt4;
T2M(i+i41,:,:)=T2M(i+i41,:,:) + 0.5*mt2ma - 0.5*mt2ma4;
end
% % Estimate scaling factor that ensures similar Std in regression data
% and NCEP.
%
scaling=std(t2m4)./std(T2M((i41+1):(i41+N),:,:));
scaling(isnan(scaling))=1;
scaling(find(scaling==0))=1;
%
% Adjust the regression data by a constant offset and scaling.
%
for i=1:i41;
T2M(i,:,:)=T2M(i,:,:).*scaling;
end
```

data set based on German, American and UK archives. The UEA-only data (SLP_U_p.nc) will be compared with the standard SLP products below, but will also be used in future downscaling studies. *Benestad* et al. (1999) found that the UEA T2M data are smoother than the observations, and *Benestad* (1999c) indicated that a too smooth field produces unrealistically large variance.

The SLP time series at each grid point in the best-fit maps were scaled to ensure that they account for same values of standard deviation as the in NCEP data. The T2M data, on the other hand, were not re-scaled.

2.6 Step 5: Producing the final anomalies by 'optimal interpolation'.

According to *Kaplan* et al. (1998), the projection method described in step 3 rank among the worst of the most common statistical methods for producing gridded data. The regression of basis functions, such as EOFs or spatial modes, tend to produce substantial errors in large data voids, and the errors become increasingly problematic with increasingly sparse data coverage. Of four different methods, the ranking from best-to-worst was: *optimal smoothing* (OS), *Kalman filtering* (KF), *optimal interpolation* (OI) and *projection* (P). At most times, the OS and the KF products were similar to the OI results, but the P method differed significantly from the other 3. Both the OS and KF methods are computationally demanding, but the IO may be implemented with less resources.

The IO method of *Reynolds & Smith* (1994) was adopted here for further refinement of the analysis, where the adjusted P-products (best-fit) from step 4 were used as the first-guess for the OI-analysis and the merged data from step 1 as the observations. The spatial correlation weighting function ($<\pi_i\pi_j>$ in *Reynolds & Smith* (1994)) was estimated from correlation analysis of the NCEP data. The data errors were assumed to be uncorrelated ($<\beta_i\beta_j> = \delta_{ij}$).

The Matlab routine *optint.m* was used to do the IO-analysis. The optimal interpolation is the most time consuming part of the analysis[5]. The solution of the equations involved the LDU factorisation (bicg), with convergence errors of 1.1e-15.

2.7 Step 6: Combining the anomalies and climatology.

The SLP data used in the regressional and modal analysis were departures from their respective annual mean values over the entire time period which they represent. The T2M data, were anomalies with respect to the seasonally varying climatology, and did not represent seasonally varying variability. The end products, consequently, consisted of the sum between the SLP and the mean NCEP SLP field[6] and between T2M anomalies and the NCEP T2M climatology[7] respectively.

---

[5]The analysis for 1873-1998, with 1512 months, took 4-5 days on an 180MHz SGI O_2 with 256Mb RAM, and required around 5×10^{11} floating points operations.

[6]For the period 1948-1998.

[7]For the period 1948-1998.

2.8 Description of the code for implementing the analysis

The Matlab scripts used for the entire data synthesis are *constr_slp.m* and *constr_t2m.m* respectively. The Matlab routine *manomaly.m* was used to calculate the NCEP T2M anomalies and *addclim.m* was used for the recombination of the two fields.

The OI results were saved in the files SLP_oi.nc and T2M_oi.nc, whereas the P-products were stored as SLP_p.nc and T2M_p.nc. Additional information about the reconstruction was also saved for quality control and future reference. The data sets included the geographical distribution of the root-mean-squared differences between the NCEP and the best-fit (P) data (represented as 'rmse' in the data files), the adjustments done to the SLP field in step 4 ('adjust' showing the mean differences in the overlapping period and 'scale' describing the scaling factor ensuring similar standard deviations in the best-fit and the NCEP data), and the spatial modes ('modes') used in the regressional analysis.

The quantity called 'R2' contains the information of the R-squared statistics (*Wilks*, 1995), giving an indication of the goodness-of-fit between the modes and the spatial patterns in the P-products of any given time, and is defined mathematically as: $R^2 = \frac{\Sigma_{i=1}^n [\hat{y}(x_i) - \overline{y}]^2}{\Sigma_{i=1}^n y_i^2 - n\overline{y}^2}$ (*Wilks*, 1995). For a perfect regression, then $R^2 = 1$ and 0 for a poor fit[8]. However, this exercise involves an extrapolation as well as interpolation, and these statistics do not represent an objective assessment of the real goodness of fit for the missing regions.

F-statistic or the F-ratio (*Wilks*, 1995, p. 168) is another commonly used measure of how good the regression is, and defined as $F = \frac{\Sigma_{i=1}^n [\hat{y}(x_i) - \overline{y}]^2}{d\Sigma_{i=1}^n [y_i - \hat{y}(x_i)]^2}$, where d is the number of degrees of freedom. If we assume that the F-distribution of random data has a normal distribution, then the F-statistics may be used as a measure of confidence limits. The time series 'F_stats' holds the F-statistics for the regressional analysis indicating how strong the regression results are: large F-ratios imply high confidence that the good fit is not due to just chance.

The quantity 'Prob' indicates the likelihood of finding a similar or better fit between two random patterns, and the variable called 'Npts' describes how many valid data points were used in the construction of the best-fit fields. These statistics have little meaning for the OI-products which only used the best-fit maps as first guess.

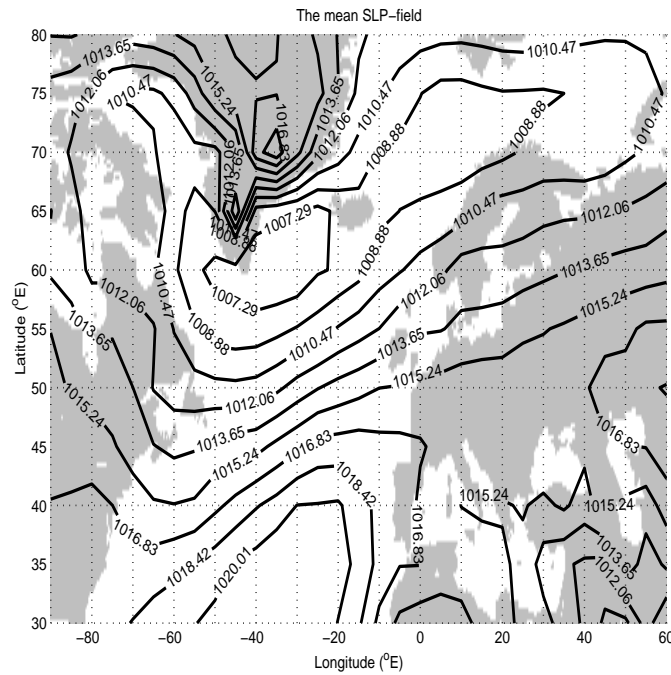[8]In some of the product, the R^2 was estimated as $R^2 - 1$ because of a glitch

Figure 5: *The mean OI SLP pattern*

3 Results

3.1 SLP

3.1.1 Quality control

One crude quality test is to examine the average statistics and compare with expectations. The mean SLP field is dominated by the mean NCEP SLP field because the reconstructed data were based on departures from a mean field, and the end result is the sum of these departures and the mean NCEP SLP. The mean SLP field has the well-known Icelandic low-pressure system and the Azores high. There are strong SLP gradients over Greenland, which are caused by the topography.

A first-line control was the visual inspection of the original incomplete geographical SLP patterns and a comparison with the best-fit maps, as in the upper and middle panels of Figure 4. The lower panel of Figure 4 shows the same information, but in a form that allows easier comparison (black curve= original incomplete data, blue curve=best-fit, and red-line indicates the residuals). Similar checks were made for every 10 time-steps.

Another type of quality control was made my examining various statistics about the regression, known as the "R-squared" numbers and "F-statistics".
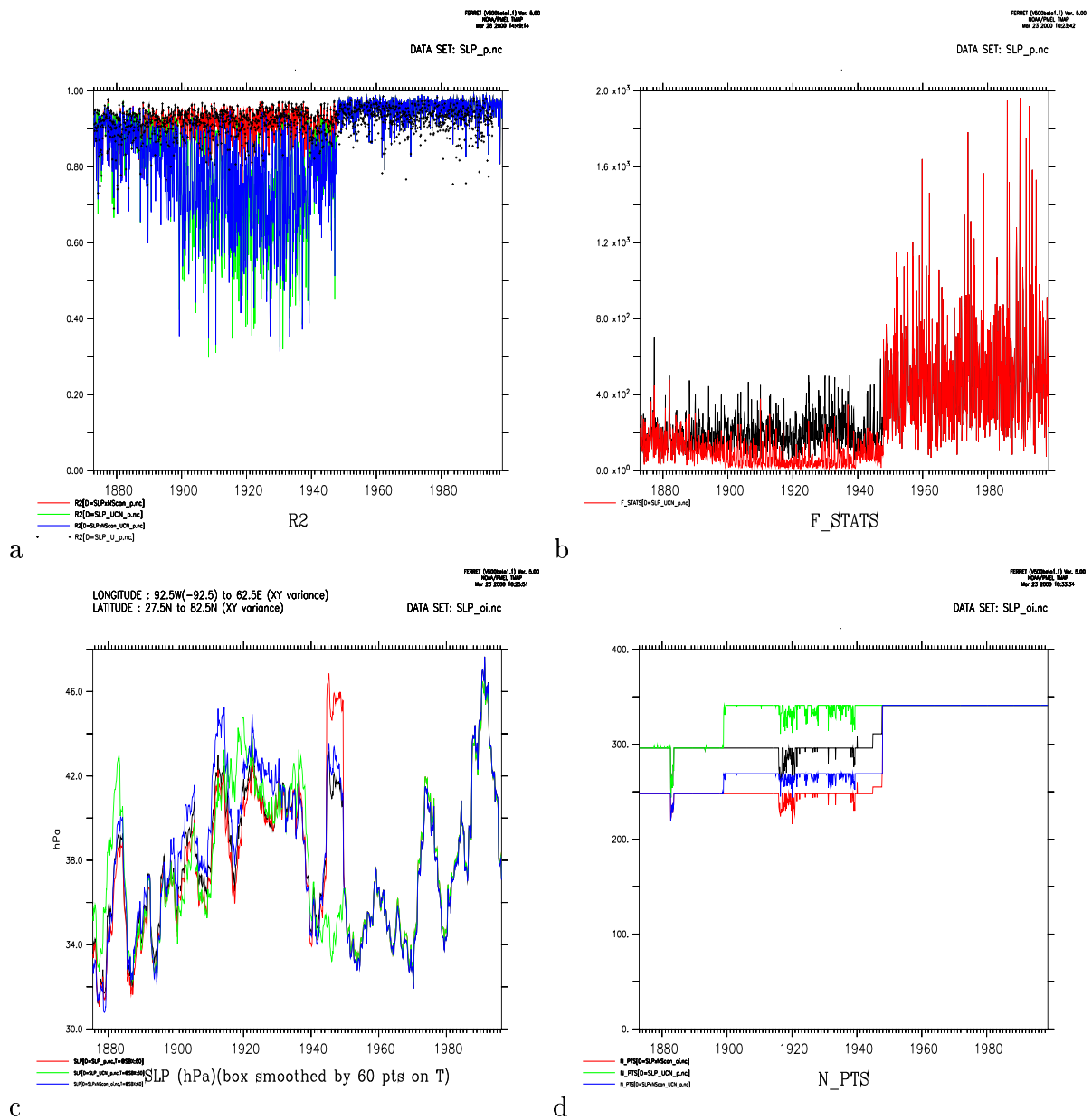
Figure 6: *The a) R^2 and b) F-statistics from the regression of NCEP modes onto the data. Panel c) shows the spatial variance and d) indicates the number of observations used for the regression. The black curves in panels a) and b) show the results for the standard products (using UEA and NCEP SLP only), whereas the red curves show the same statistics for the multi-source products (based on UEA, NCAR, COADS, and NCEP data). Colour codes for panel c) black = standard OI, red = standard P, green = multi-source P, and blue = test OI. For panel d), the number of observations for the standard product are shown in black, test product in red, multi-source in green, and test-multi-source in blue.*
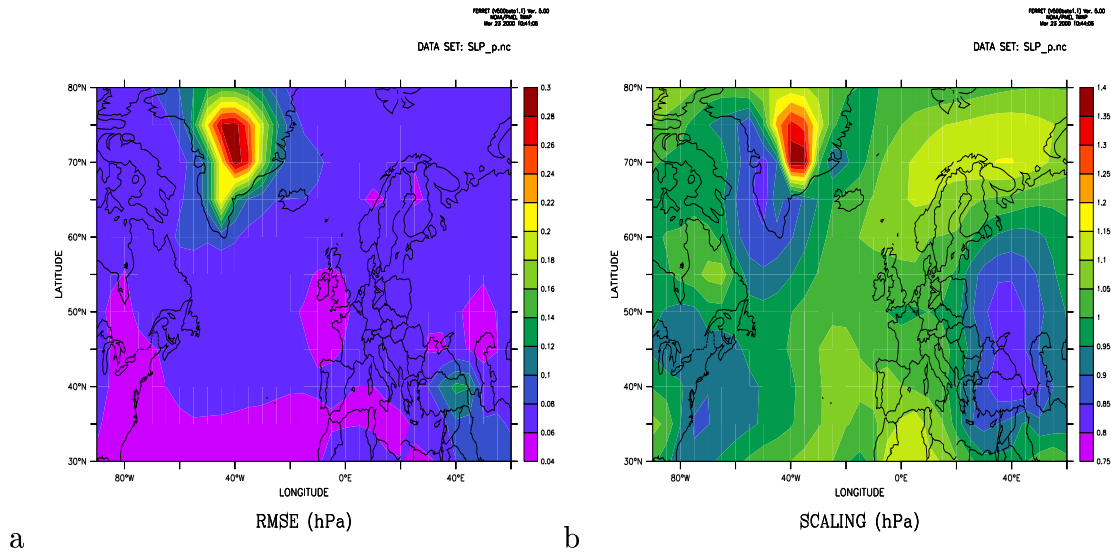
a

b

Figure 7: *a) Estimates of the root-mean-squared errors between the best-fit SLP patterns and the NCEP data for the overlap-interval 1948-1995 and b) the variance ratio of the NCEP to the original (unscaled) best-fit analysis.*
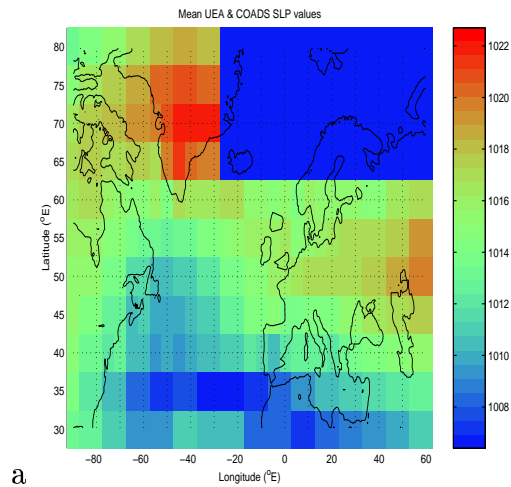

a

Figure 8: *Map over the mean SLP from the available observations and the blanked out region (blue area)*
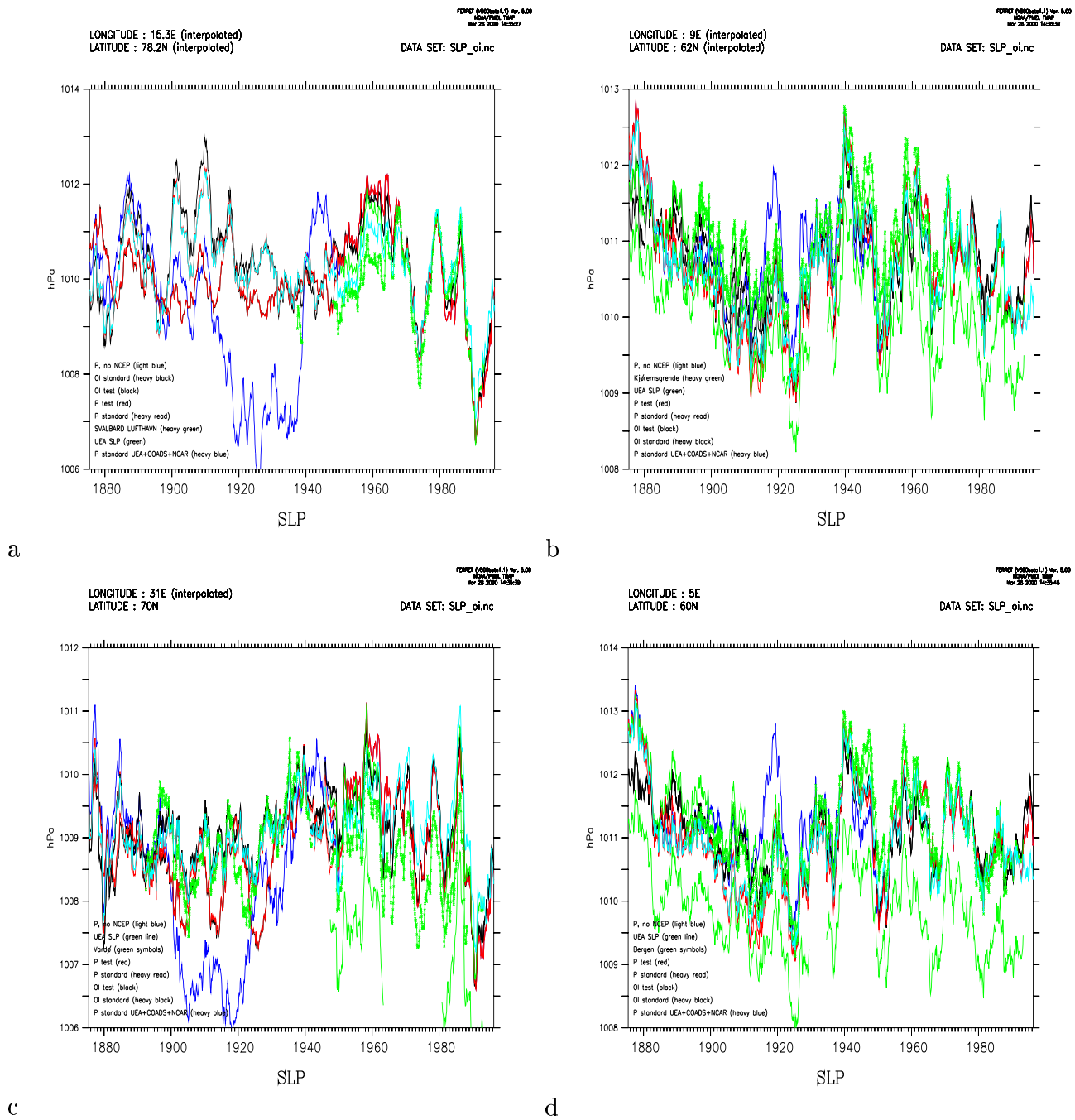
Figure 9: *The comparison between interpolated SLP time-series for a) Svalbard (78.2° N, 15.3° E), b) Kjøremsgrende (62° N, 9° E) and c) Vardø (70° N, 31° E), and (d) and Bergen (60° N, 5° E). The black curve is the OI reconstructed SLP field based on the maximum available observations (thick) and with observations in the region 60N-80N and 10W-60E masked (thin) as missing (see Figure 8), whereas the red line represents a similar P reconstructions, and the UEA SLP data as thin green lines. Two green curves are shown for the UEA data, one unadjusted and one which is adjusted to have same mean value as the reconstructions/NCEP. The actual station observations are shown as green crosses. The light blue line shows the P products based on NCEP modes regressed onto UEA from 1873-1995. The SLP has been 5-year low-pass filtered (boxcar).*
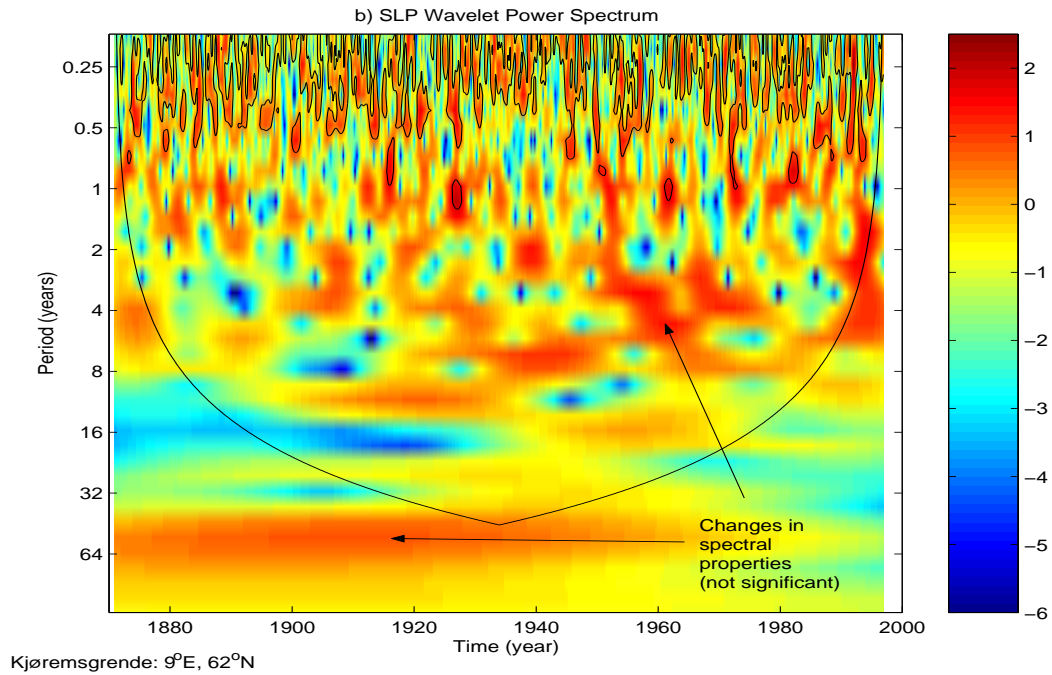
Figure 10: *Wavelet analysis of the SLP from Kjøremsgrende indicates variations in the low-frequency spectral properties, but these are not statistically significant above the 95% level. The x-axis denotes time whereas the y-axis gives the time scales of the SLP variability. The colours indicate the strength of SLP fluctuations at given date (given by the position on the x-axis) and associated with a particular (range of) time scale(s) (y-axis).*
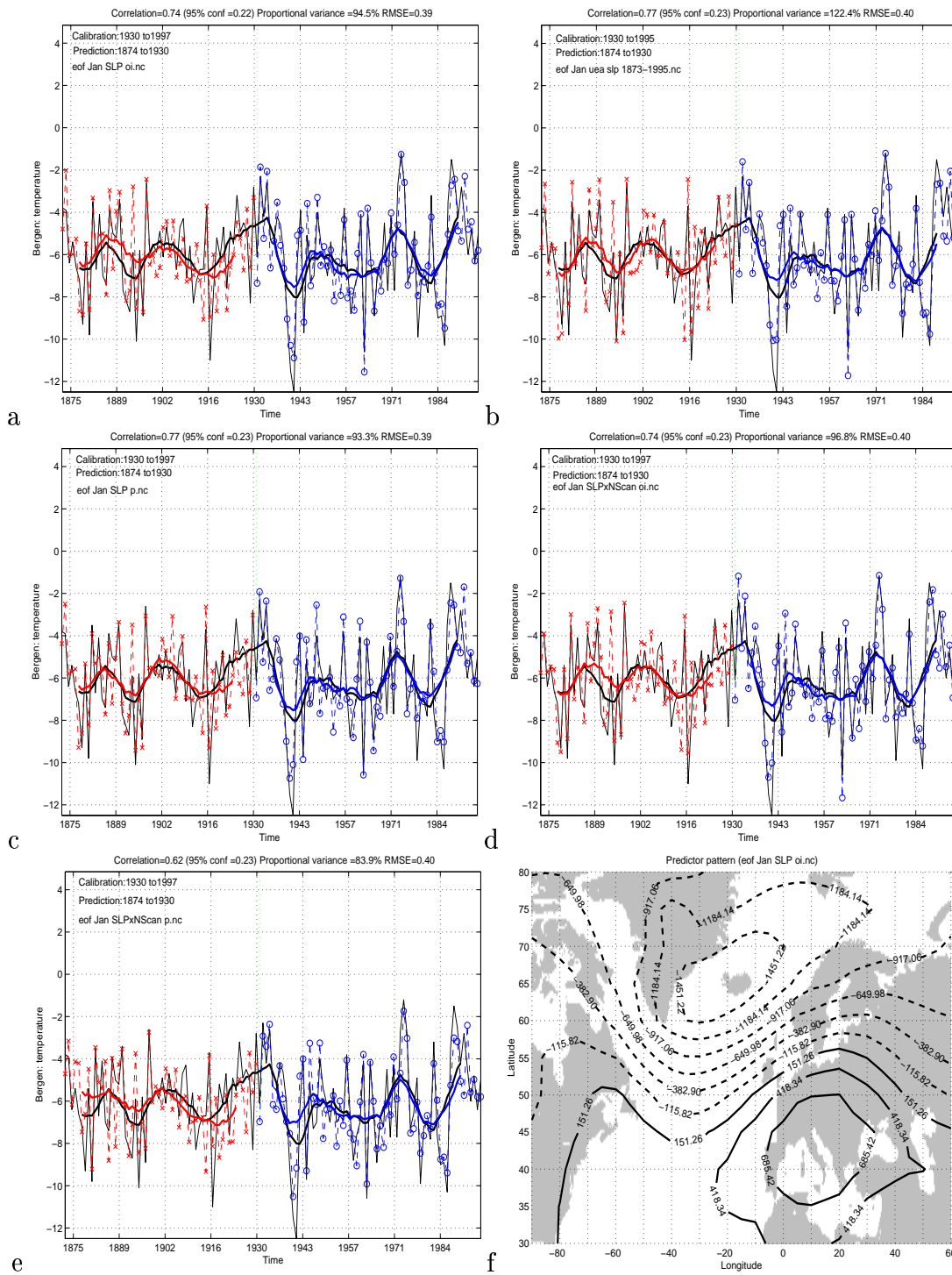
Figure 11: *Panels a)-e) show the results from tests where a simple empirical SLP model was used to predict temperatures in the period 1873-1947, not used in the calibration of the model, and f) shows the SLP predictor pattern (standard OI) having the strongest relationship with the temperature in Bergen. The predictions in a) are based on the standard OI-product, b) on UEA, c) on the best-fit projection (P) products, d) on OI SLP data with 60N-80N, 0-60E masked out, and e) the P-SLP with 60N-80N, 0-60E masked out.*

Figure 6a gives the R^2 scores for the regression of the modes onto the observations, and the results suggest generally good fits, and of course, with better scores for the NCEP data. It is evident from Figure 6a that the multi-source products are associated with lower values for R^2, and worse fit between the combined data and the NCEP modes. The F-statistics shown in Figure 6b indicates, in agreement with the R^2 scores, that the fit between the modes and the NCEP data was associated with high confidence whereas the match with the UEA record had lower confidence (10-400).

A test of homogeneouity was made by examining the geographical variance (the variance in x and y directions) of the SLP at each time interval (Figure 6c). The results suggested that SLP the may not be entirely stationary, as the geographical variance was often higher towards the end of the data record when the NCEP data were used as a basis for regression. There was even a slight tendency of long-term increase in the geographical variance within the NCEP interval, although the question whether the increase was significant is not been addressed here. There is also a suggestion of a systematic increase in the geographical variance from 1873 to 1920, in addition to low-frequency variations.

Figure 6d shows the number of observations used to find the best-fit SLP patterns, and the number of observations was approximately constant for the 1873-1947 period and the 1948-1998 period, but there was a sharp increase in observations from the UEA data set to the NCEP reanalysis II.

A comparison was made between differences in mean values in the overlap-interval (1948-1995) between the SLP reconstructions based on the regression onto the UEA data and the NCEP data ('adjust'). Figure 7a shows only small differences in these mean fields, with largest errors over Greenland. Furthermore, the ratio of the (temporal) variance of NCEP to the best-fit maps was examined to check whether there may be discontinuities in the change between the two products (Figure 7b). Largest standard deviation differences were found in the variance near eastern Greenland (UEA describes lower variance than NCEP) and over Labrador Sea and Asia Minor (UEA describes higher variance than NCEP), indicating possible problems with the data in these regions.

More sophisticated tests were applied to the data to check whether the SLP reconstruction is reliable. These tests involved comparisons with similar products as the "standard versions" (SLP_oi.nc, SLP_p.nc, T2M_oi.nc and T2M_p.nc), but based on a subset of the available observations. In this case, the observations in the region 60°N-80°N and 10°W-60°E (Northern Scandinavia) were excluded from step 1 in the analysis ("Blanked-out" or masked), and these products will henceforth be referred to as the "test-products" (SLPxNScan_oi.nc and SLPxNScan_p.nc). In Figure 8 the masked

region is shown in dark blue. A comparison of interpolated SLP time-series from the various products and station observations (NACD) in the blanked-out region gives a stringent test as to whether the final product is reliable. This kind of test also gives an indication of the sensitivity to the number of observations.

Figure 9a shows such a comparison for the SLP over 4 different locations. The station Kjøremsgrende (Figure 9b) at Dombås and Bergen (Figure 9d) are located at the edge of the masked out region (data void), whereas Svalbard and Vardø (Figure 9a and c) are well within the this area. It is evident from Figure 9 that there are some differences between the different SLP reconstructions before 1960 as well as between the gridded data and the station readings (green curves). Some discrepancies were particularly evident during the 1910-1930s in the multi-source product, and the differences are greatest over Svalbard and Vardø with lowest data density. The results from the best-fit analysis based solely on the UEA data (no regression to the NCEP re-analysis in the period 1948-1998, but only to UEA SLP 1873-1995) are represented by the light blue curve. There are some differences between the UEA- and NCEP-based results for the period 1948-1995, although these are generally small.

There is a systematic decrease in SLP over Kjøremsgrende up to 1920 followed by low-frequency variations. The results from a wavelet analysis (*Torrence & Compo*, 1998) of this series is shown in Figure 10. Although variations with time scales ranging from 2-8 years have become more prominent, the analysis suggests that these changes in the spectral properties are not statistically significant. Likewise, the analysis hints at a reduction variations with time scales longer than 32 years, but these results are outside the wavelet confidence window. The results of the wavelet analysis therefore give no evidence of the SLP over Dombås being inhomogeneous.

Figure 6d shows the total number of observations used in the best-fit analysis, indicating that the blanked-out region only represented a small proportion of the total amount of data. This result suggests that neither the OI nor the P analyses are particularly sensitive to the number of observation in the data void, and the finding that the OI results are less sensitive than the best-fit maps is in agreement with *Kaplan* et al. (1998). The data in the regions sparsely covered with observations are nevertheless expected to be subject to highest uncertainties before the 1930s. The SLP constructions were similar where there were observations, i.e at Kjøremsgrende (Figure 9b).

It is evident that the OI analysis has a tendency to follow the UEA data whenever available. In general, the gridded analyses does not follow the station observation closely, suggesting that these products are not suitable for interpolation to small scales. As mentioned earlier, the fine geographical

details have been sacrificed for more robust estimation of large-scale features. It therefore not expected that the time series from the reconstructed fields are identical to the historical observations when examining *small-scale* variability such as time series interpolated for a given climate station.

A quality test of the *large-scale* field structure was made by using the synthesized data to reconstruct temperature series for various location (downscaling). The crux of this test is to develop an empirical model that describes the local temperature from the large-scale SLP patterns for a given period (1930-1998) and then use this model and SLP from a different period (1873-1930) to predict the local temperature. If the prediction is good, then we may have confidence in the SLP-product. In this case, the test also gives an indication that the empirical relationship between the large scale SLP and the local temperature has been stationary over the past 126 years. The empirical model used here was based on ordinary regression applied to the calibration period. A step-wise regression analysis was used to identify the modes that contribute towards bergen temperatures, and a cross-validation correlation analysis was used to test whether the individual mode improved the prediction skill.

Figures 11a-e show the results from such a test for Bergen January temperature for a number of data products. The correlation score shown was calculated for the independent temperatures of the evaluation period only (red curve). Predictions based on the OI-products are compared with predictions using the incomplete UEA SLP data set, the P-products, and test data produced with a subset of the available observations. Different OI analyses using different data sources were tried (not shown), and the best predictions are obtained with the UEA data.

The results of all these prediction tests are indeed similar to the station observations, suggesting that the large-scale patterns in the SLP field are representative of the past variations and that the relationship between the SLP and the Bergen January temperature is close to stationary. The predictor pattern with strongest relation to the Bergen temperature has a strong NAO character, but with the southern centre of action shifted eastward (Figure 11d).

The highest scores were obtained with models derived from the UEA and standard P data. Both the standard OI and the test OI products were associated with similar skill (0.74), but marginally lower than the highest scores (0.77). The correlation score for the test P product, on the other hand was clearly lower (0.62) than when all available data were used. The interpretation of these results is that the projection (P) method is more sensitive to large regions with missing data, which is in accordance with *Kaplan* et al. (1998).

All the predictions produces spurious cold events in January 1914 and 1918, but the OI and P products gave less cooling than the UEA data. The data also produced a fictitious warm January month at the end of the 1800s. Most of the "early century warming" (1915-1935) was captured by the models, although the low-passed curves show a slight underestimation or delay of the warming.

3.1.2 Pressure gradients

Figure 12 shows pressure gradients calculated for a north-south transect between *a)* Bjørnøya and Uppsala, *b)* Iceland and the Azores (NAOI), and east-west transects between *c)* Vardø-Jakobshavn, and *d)* Uppsala-Torshavn. The black lines are the OI SLP, red represents P SLP, green are the corresponding SLP differences from the UEA data, and blue shows the SLP differences between the station observations. Assuming that the station data (blue) represent the "truth", it is evident that there is a good match between the various SLP differences in the north-south sections in Figure 12a (SLP(Bjørnøya) - SLP(Uppsala)). The estimates of the NAOI from the various data sets in Figure 12b correspond reasonably well with the official NAOI index (blue) (*Jones* et al., 1997), although all data sets indicate too small variations. The analyses appear to be out of phase with both the NAOI and the corresponding estimate from UEA around 1920 and before 1900.

Estimates of the east-west SLP differences are shown in 12c and d. Both OI and P SLP data indicate too high values for the Vardø-Jakobshavn SLP difference, but since this transect is in the Arctic data sparse region, the quality of the data in this region is questionable. Furthermore, Jakobshavn is in Greenland which is associated with large and sharp SLP gradients. The east-west transect for Uppsala-Torshavn (Figure 12d) where the data quality is higher suggests that the UEA SLP differences are biased towards too high values, whereas the analyses are more similar to the observations (blue curve). These results seem to suggest that the zonal SLP differences (meridional geostrophic winds) are difficult to reproduce. In either case, the SLP measurements contain random or systematic errors, and when subtracting two large ($\approx$1000hPa) noisy (a few hPa) measurements with similar value, one is prone to relatively large errors.

3.2 Comparison between large-scale climate patterns

Figures 13 to 15 shows the results from a Canonical Correlation Analysis (CCA) (*Benestad*, 1998a; *Wilks*, 1995; *Bretherton* et al., 1992) on various combinations of data sets. A comparison between the OI SLP (a,c) and
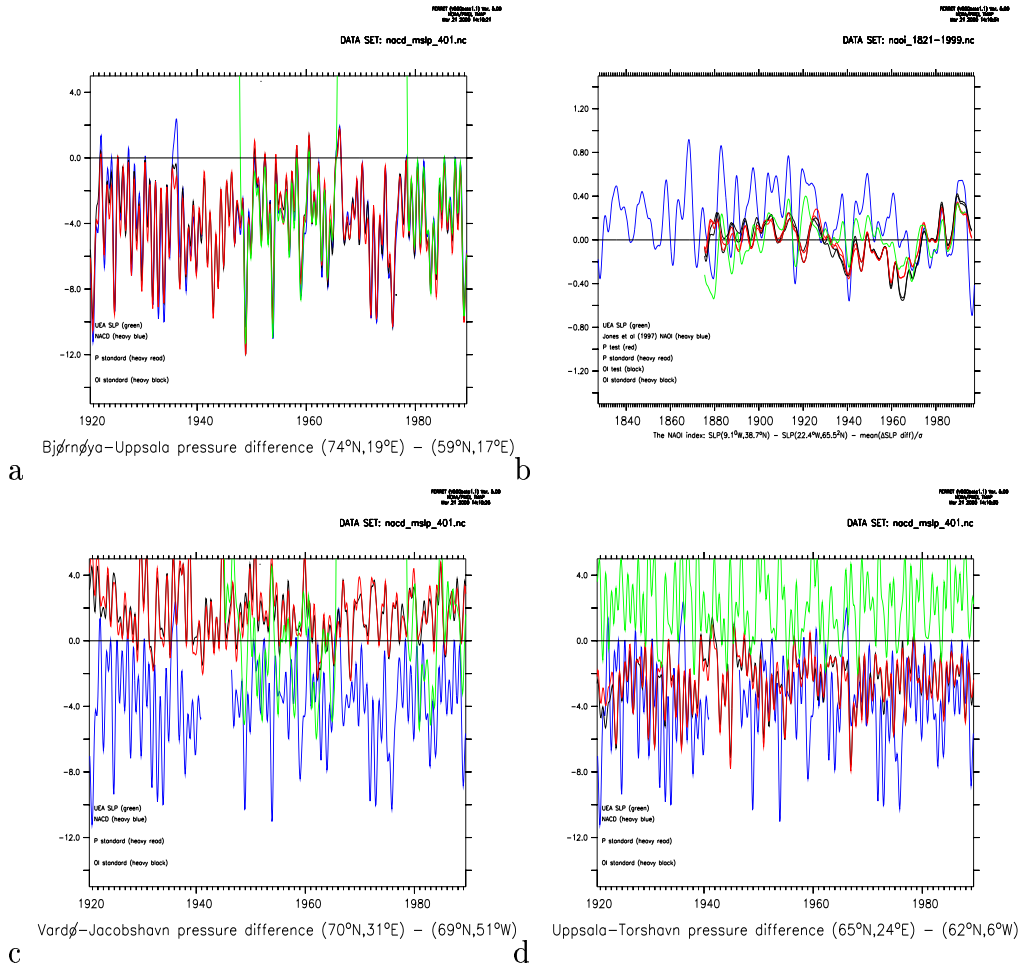
Figure 12: *SLP differences (gradients) estimated from observations and various gridded data sets. Black=OI SLP, red = P SLP, Blue=observations (NACD, and NAOI), and green=UEA.*
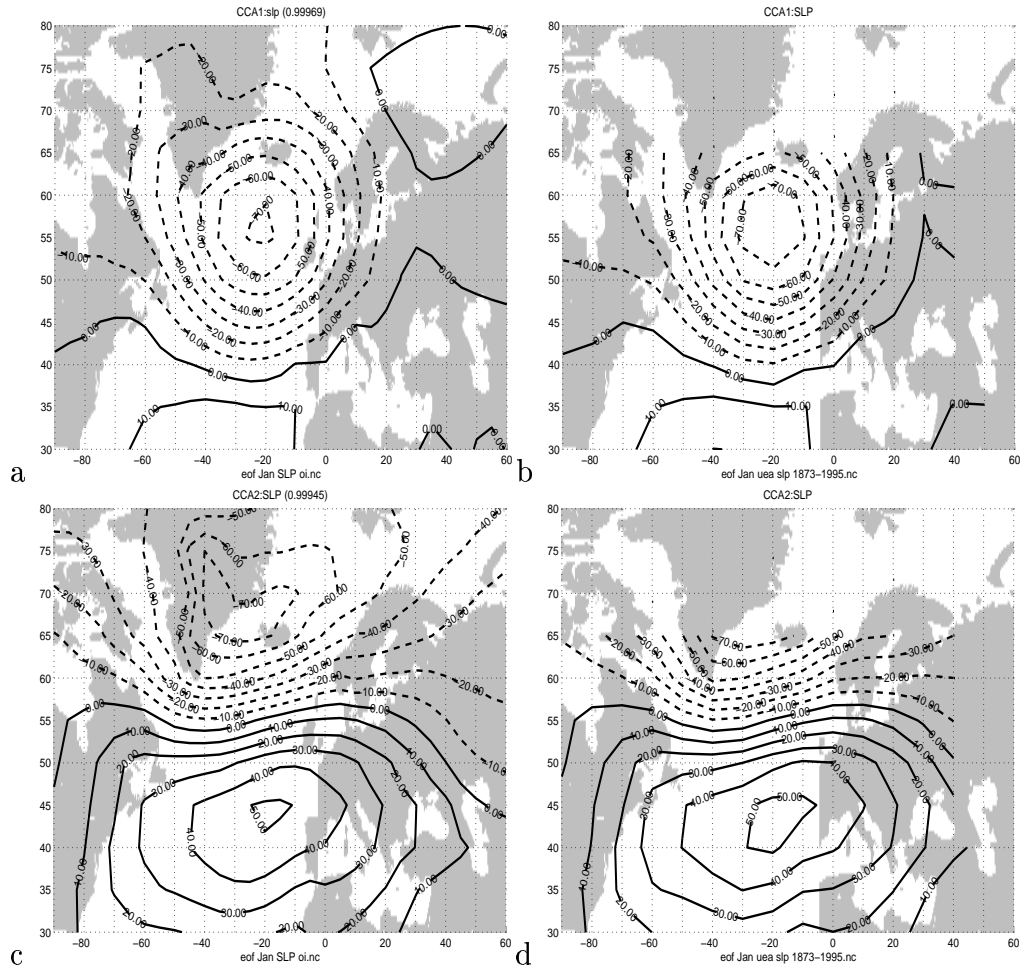
Figure 13: *The two leading CCA patterns for OI SLP and UEA SLP indicating similar spatial structures with similar temporal evolution.*
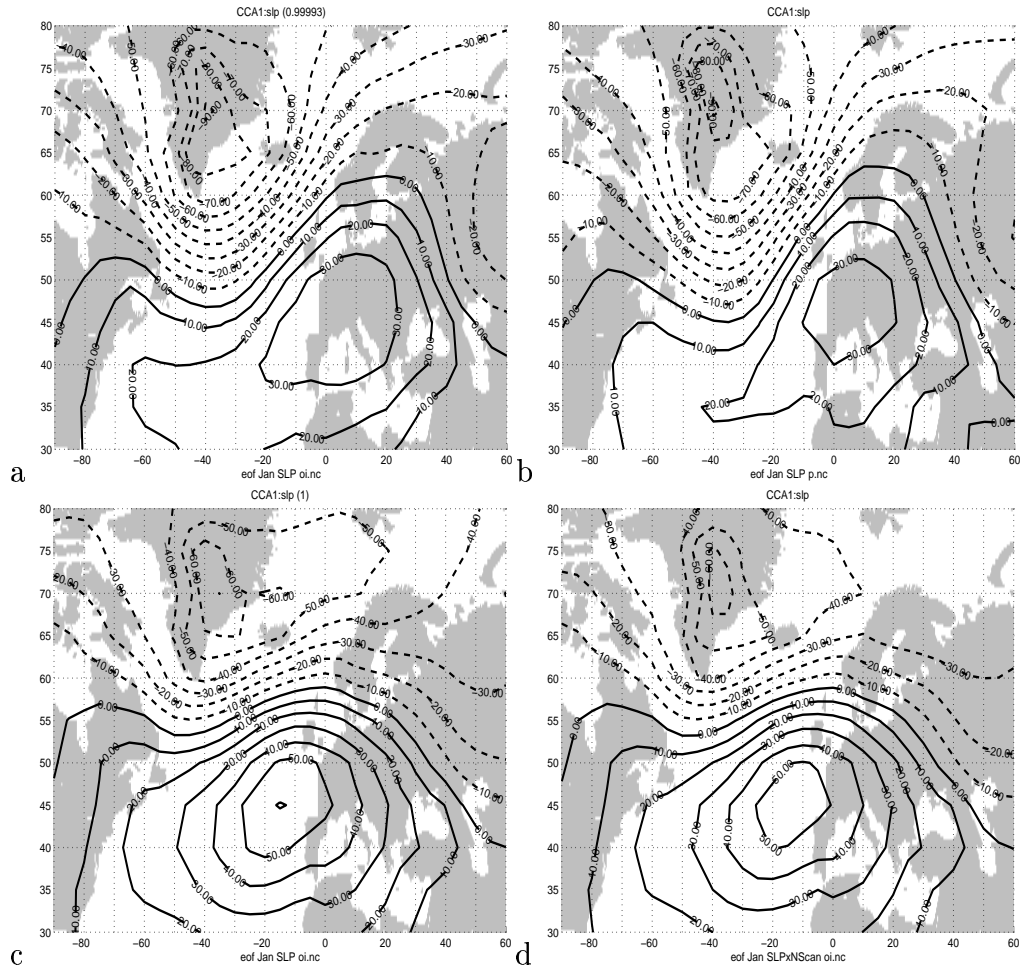
Figure 14: *Upper: the leading CCA patterns for OI and P SLP products. Lower: the leading CCA patterns for standard and test OI SLP.*
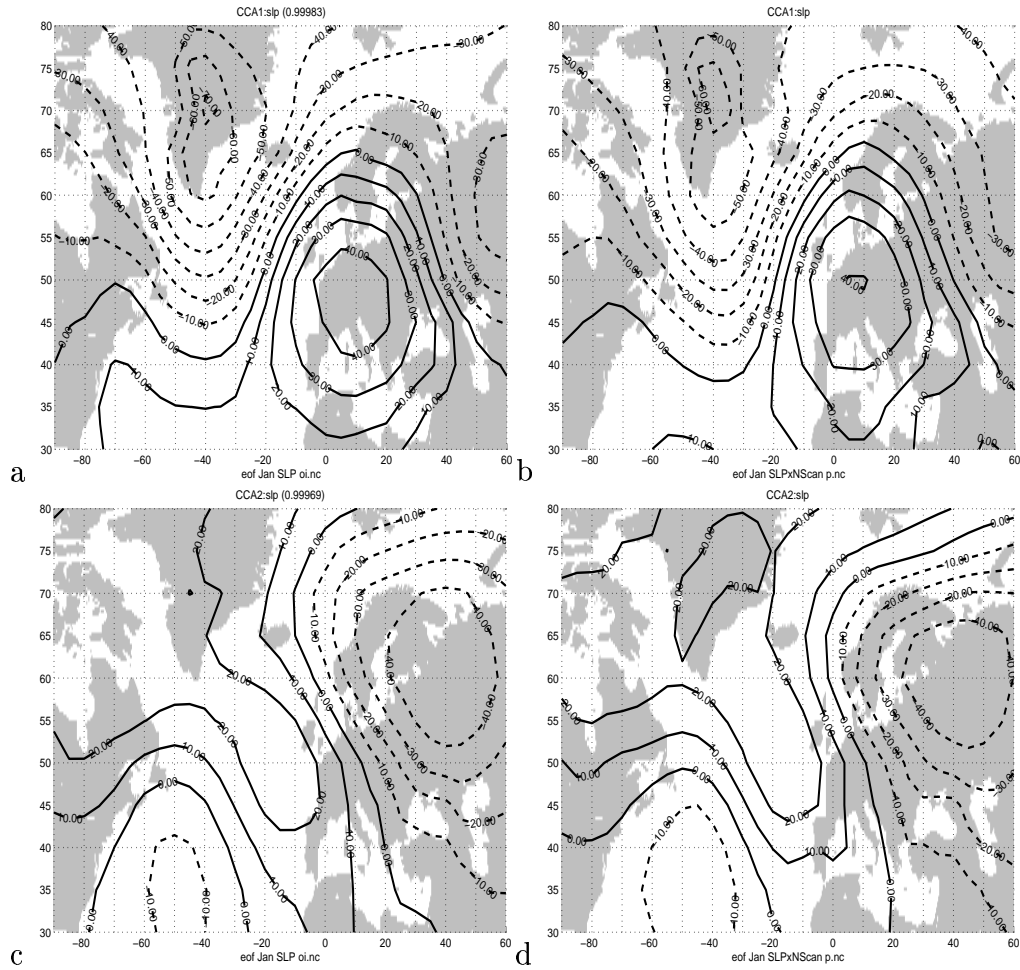
Figure 15: *The two leading CCA patterns for the standard OI and test P SLP products.*

UEA (b,d) data is shown in Figure 13, which confirms that the two data sets describe similar spatial SLP patterns with similar temporal evolution. The leading CCA patterns (a,b) describe a monopole south of Iceland, whereas the second CCA patterns contain strong signatures of the NAO.

The upper panels of Figure 14 show the similarities as well as differences between the OI and P SLP products. The greatest differences are found over Greenland-Iceland and over the Azores. The lower panels of Figure 14 shows a similar analysis for the standard OI SLP and test OI SLP: the greatest differences are found in the data sparse region of Greenland, Iceland, and Svalbard.

The results from a comparison between standard OI SLP and test P SLP are shown in Figures 15. Differences are evident over Greenland, the Barents Sea, and the subtropical Atlantic in the leading CCA pattern and over Labrador Sea, southeastern Greenland, and Hudson Bay in the second CCA pattern.

3.3 Testing the data fields in conjunction with prediction schemes

Finally the OI SLP data products were tested in conjunction with empirical downscaling models intended for studies of future local climate scenarios. The empirical models are described in *Benestad* (2001) and *Benestad* (1999b), and involve the use of common EOFs (*Barnett*, 1999). The tests consisted of using the (CCA) empirical downscaling models calibrated from 1873-1950 to predict the temperature evolution after 1950. The results of these tests are for Oksøy (a), Bergen (b), Kjøremsgrende (c), and Vardø (d) are shown in Figure 16, and the models gave a good description of the recent temperature variations, although not a perfect reproduction. The models did not capture all of the low-frequency swings. The conclusion of these test results is that the OI SLP data are suitable for empirical downscale studies based on the empirical models of *Benestad* (1999b). The relationship between the large-scale SLP patterns and the local temperatures in the four location has been approximately stationary in the last 126 years.

3.4 T2M

3.4.1 Quality control

Figure 4b is again an example of a first-stage quality control of T2M, based on a visual comparison between the best-fit and original spatial T2M patterns. The test was the same as for the SLP. The lower panel provides an easy
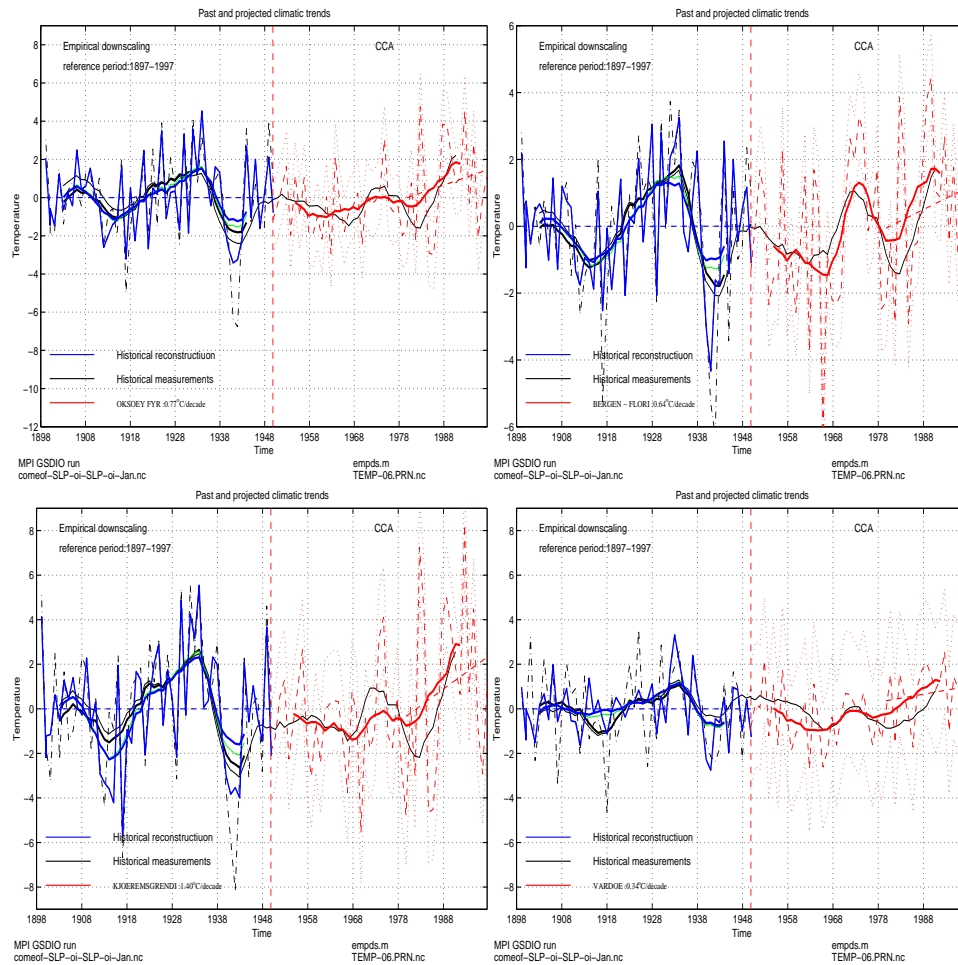
Figure 16: *Using separate periods for the calibration and evaluation of the CCA downscaling models to predict January temperatures for a) Oksøy, b) Bergen, c) Kjøremsgrende, and d) Vardø. The results were derived using common EOFs of the OI SLP, and the same routine employed in development of climate scenarios based on AOGCM results.*
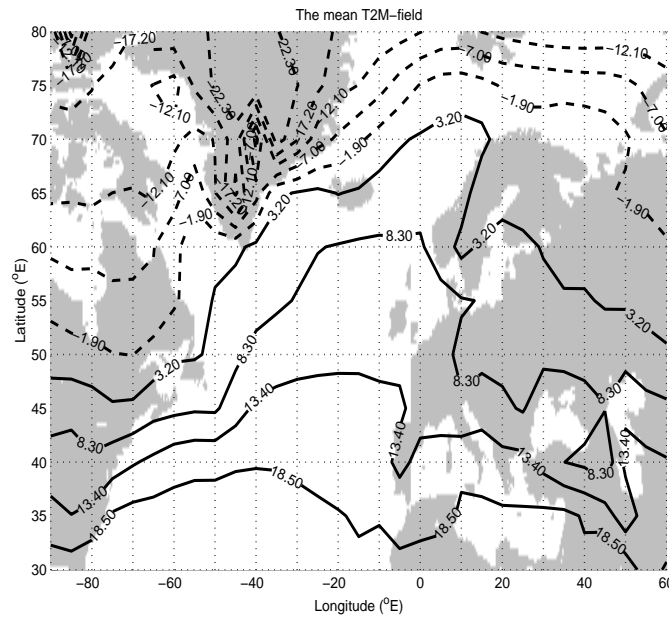
Figure 17: *The mean T2M pattern*

comparison between the original data and the best-fit. This control gave assurances of reasonable reconstruction of the spatial pattern.

A secondary quality control involves the mean temperature fields. Figure 17 presents a map of the reconstructed mean T2M for the period 1873-1998. As expected, the polar regions are cooler than the sub-tropics. The warm ocean currents are associated with high temperatures and the ocean is on average warmer than the continents.

The regression statistics, such as the R-squared and F-ratio (Figure 18a and b), indicated that the best-fits were associated with high level of confidence and closely matched the original data. To investigate whether the temperature was subject to systematic changes in terms of amplitudes, the geographical variance was plotted in Figure 18c. Although, there are some variations, the curve resembles a white noise process with no long-term systematic changes after the 1920s. From this point of view, the T2M reconstruction may be regarded as stationary for this latter period. The larger variances in the early part of the record may be indicative of questionable data quality, for instance due to a reduction in the number of observations during WW1 (see Figure 18d). The times of reduced quality (where there are no original observations) do not show up in the R^2 and the F statistics, as these merely assess the goodness of fit between the valid data and the modes. But, Figure 18d shows the number of observations used for the best-
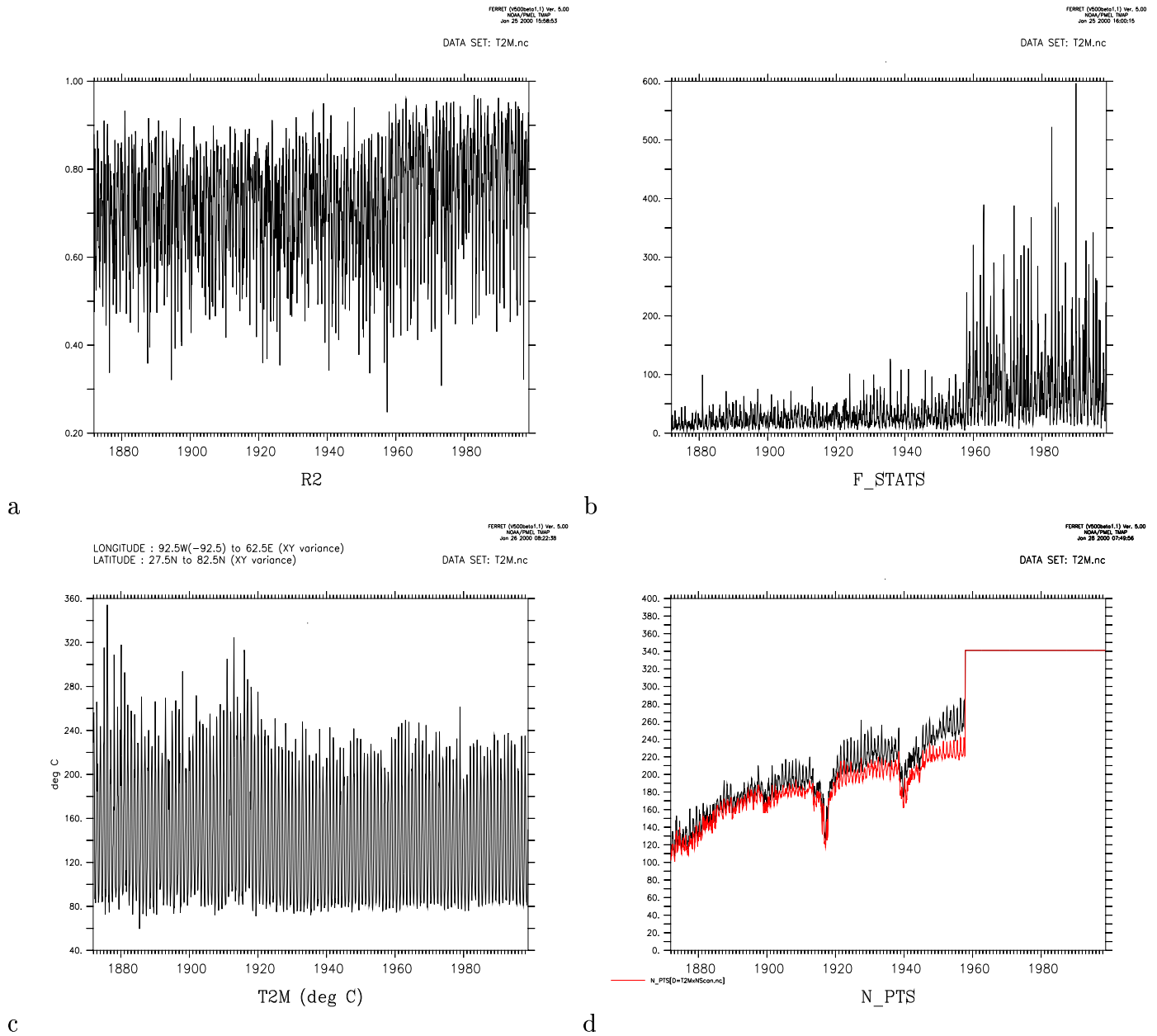
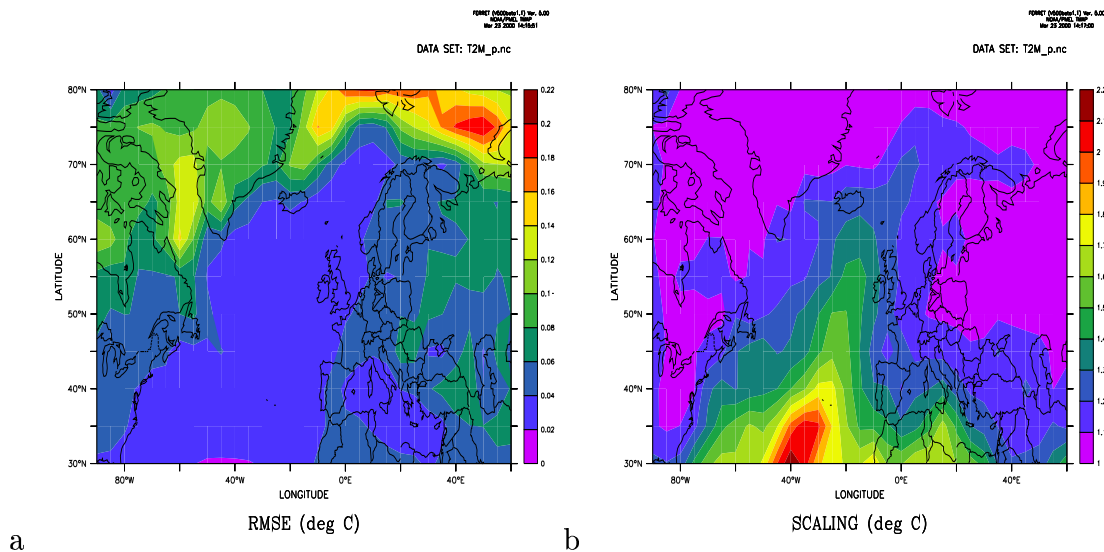Figure 18: *Same as Figure 18 but for the T2M data.*

Figure 19: *a) Estimates of the root-mean-squared errors between the best-fit T2M patterns and the NCEP data for the overlap-interval 1958-1995 and b) the variance ratio of the NCEP to the original (unscaled) best-fit analysis.*

fit estimation increased systematically until 1958, after which the amount of data was constant. The dips in the number of observations are associated with the two world wars.

The root-mean-square errors and scaling ratios for the overlap period (1958-1998) are shown in figure 19. The greatest errors were found in the Arctic and over the oceans (Labrador Sea and Barents Sea). The differences in the standard deviation were found over the Atlantic ocean, and curiously in the vicinity the ocean currents (the North Atlantic Drift and the Norwegian current) and over the North Atlantic interior (the T2M from *Jones* et al. (1998) have smaller amplitudes than NCEP re-analysis).

A comparison (Figure 20) between the 2-meter temperature at 5°E-60°N and the Bergen-Florida temperature revealed a cold gridded bias of approximately 2°C. This bias does not necessarily imply that the gridded temperature is wrong, as the 5° × 5° grid describes a large-scale quantity whereas the Bergen temperature measured in a fjord is much influenced by local conditions. Likewise, a comparison between the gridded temperature and measurements made at the top of nearby mountains, would probably suggest a warm bias. Most of this bias stems from the NCEP data, but the discrepancy may also be explained by the fact the reconstructions do not include small-scale details.

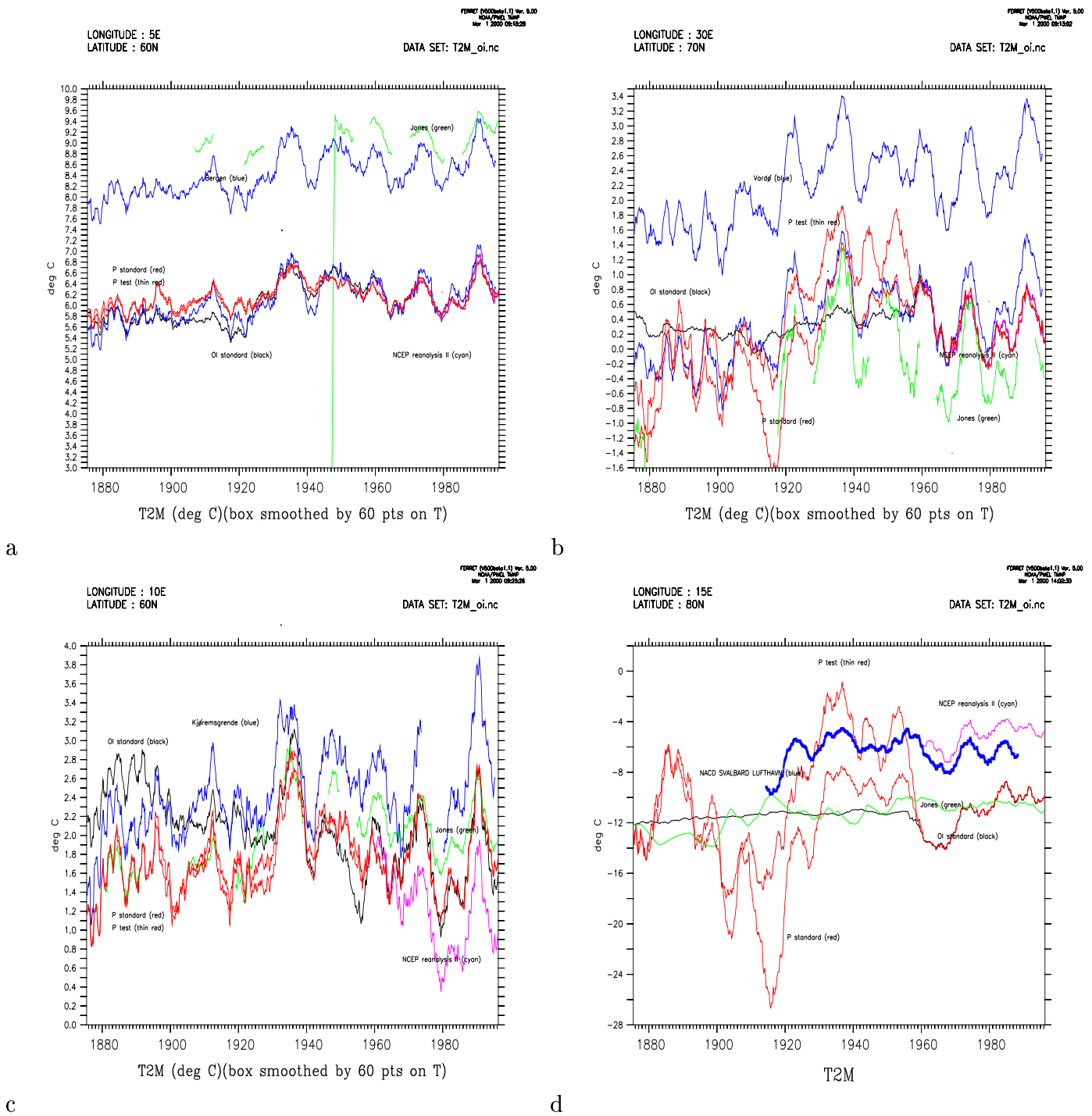The comparison between reconstructed temperature and observations

Figure 20: *Comparison between temperature variability in a) Bergen b) Vardø, c) Kjøremsgrende and d) Svalbard airport and corresponding interpolated gridded temperature from various gridded data sets. Black line denotes the OI-products, cyan represents the NCEP reanalysis II, green 2-meter temperatures from Jones (1998), blue curves are the station measurements, and red lines indicate the test P-product. Thin red and black lines show the test P and OI. In some panels, two curves for the station temperatures are shown, one adjusted to have similar mean value as the OI-products.*
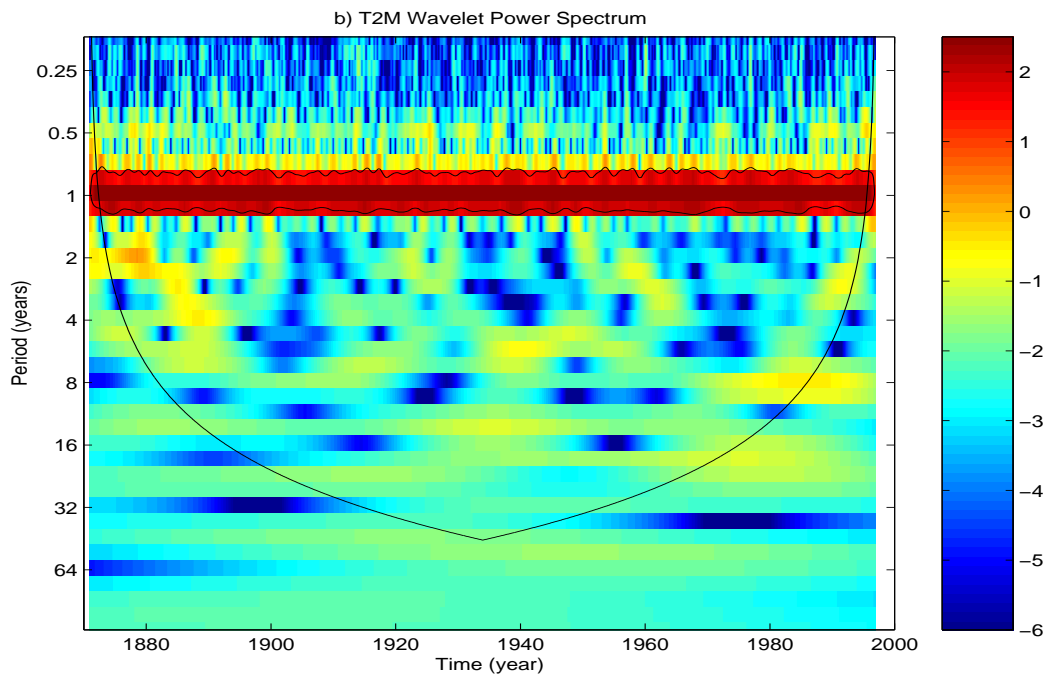
Figure 21: *Wavelet analysis of the T2M from Kjøremsgrende shows no substantial variations in the low-frequency spectral properties which are statistically significant above the 95% level.*
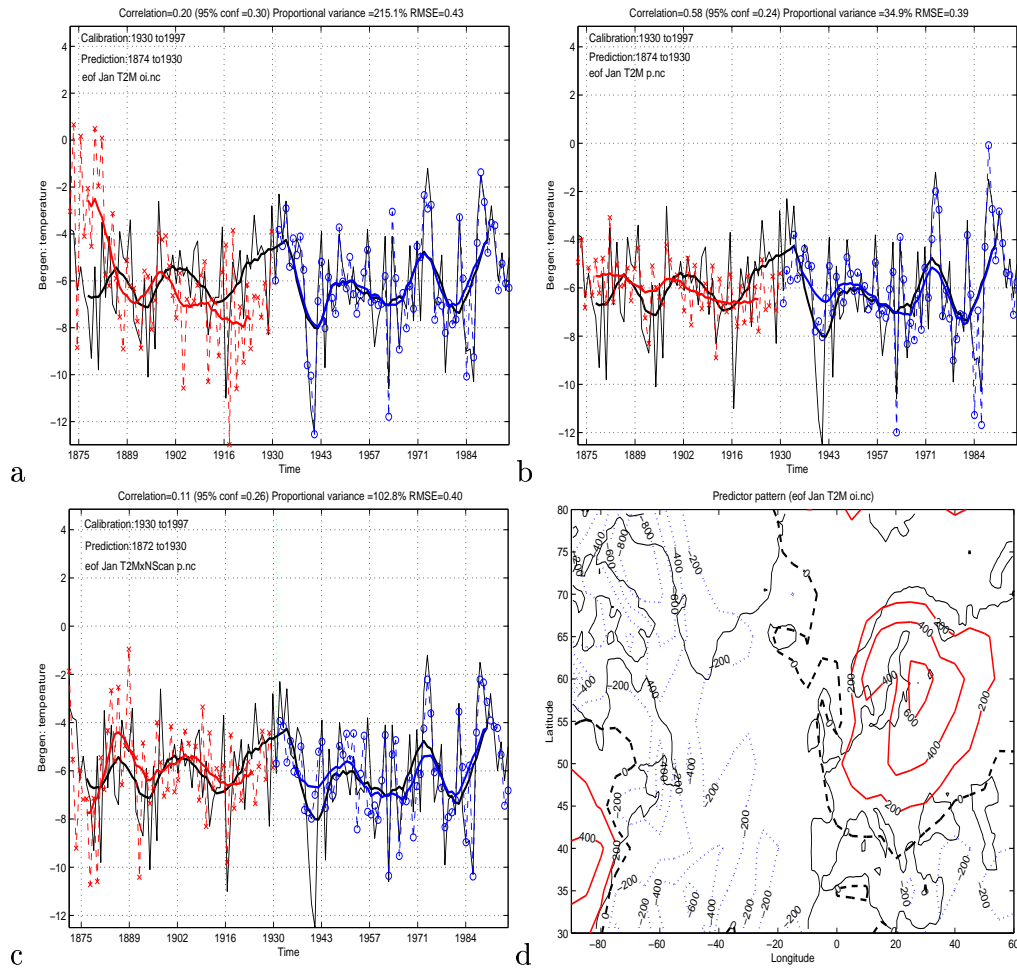
Figure 22: *Panels a)-c) show the results from tests where an empirical T2M downscaling model was used to predict January temperatures in the period 1873-1947, not used in the calibration of the model, and d) shows the T2M predictor pattern having the strongest relationship with the temperature in Bergen.*

from climate stations is not necessarily a stringent test of the real quality of the data, as these products are derived only in terms of large-scale patterns and since some of the station data have been used in the construction of the original gridded (UEA) and reanalysis (NCEP) data. Because of spatial coherence, stations not included in the gridded data sets are not entirely independent of these, as they are correlated with neighbour station records which may have been included. Thus, these comparisons do not provide clear-cut evaluations between predictions and independent data. The reconstructions may be tested more stringently by adopting two methods: exclude data from remote regions where there are few high-quality observations (eg. Svalbard and Greenland), or by evaluating predictions from the large-scale patterns and downscaling models with station values. Once more, the downscaling approach provides a test on the stationarity between the large-scale pattern and the local climate, and can only be used to verify the data set. As mentioned earlier, poor predictions are not necessarily due to errors in the reconstructions, but may also be a result of non-stationary relationship between the predictors and the predictands.

Comparisons were also made between station observations (NACD) and interpolated values from three different reconstructions of T2M, one of which is a test-product which was based on a subset of the total available observations. Figure 8 shows the region where observations in this test-product were excluded from the best-fit construction of the test T2M P-product, and Figure 20 shows differences between the three data sets. The standard OI reconstruction is shown in heavy black, and the red curves represents the corresponding P-products (heavy= standard, thin line=test-product). The NACD station observations given by the blue lines, whereas the green curves represent the *Jones* et al. (1998) temperatures.

Figure 20a shows a comparison between the reconstructed T2M and the Bergen-Florida temperatures. It is evident that the reconstruction reproduces most of the variations in the Bergen temperature, although there are some differences in amplitude and timing. It is, however, not expected that these curves should be identical, as they are not directly comparable. Whereas the observations represent a local climate, the temperature reconstruction describe large-scale temperature fields (see mode 20 in Figure 3d) on a $5° \times 5°$ grid. Similar analyses were made for Vardø (Figure 20b), Kjøremsgrende (Figure 20c) and Svalbard (Figure 20d). Vardø and Svalbard are located within the data void in the test-products, whereas the other locations are on the edge of this masked region. The OI results are different to the P-products in Vardø and Svalbard, but here the latter data are more realistic as the OI data are too smooth and have too low variance before 1958. It is possible that this problem arises because the data coverage

is too sparse (see Figure 4b for typical data coverage) and the correlation distance too small to give a realistic value for T2M (O-E Tveito, personal communications). The OI T2M results for Svalbard are clearly unrealistic before 1958.

The various reconstructions are similar at Kjøremsgrende, but the OI results represent a warmer climate than the best-fit products before 1930. There are small differences between the various best-fit reconstructions over mainland Norway, but there are substantial differences between the different P data sets over Svalbard between 1900 and 1960. Before 1900, the standard P and test P T2M products show similar values over Svalbard. One explanation for this observation is that there were few observations in the masked out region before 1900 (see Figure 18d), so that the standard and the test-products almost become identical for the earliest records (i.e masking out missing data has no effect). This result indicates a reduction of data quality for records earlier than 1960 as the number of good observations diminishes as we go back in time.

A wavelet analysis was applied to a temperature series interpolated from the P T2M at 62°N-9°E (Kjøremsgrende) to test for homogeneouity (Figure 21). This test gave no indication of inhomogeneouity for this location. The annual cycle is prominent for the entire record, in accordance to expectations. Although statistically insignificant (below 95% level), there are hints of variations in the spectral properties (slightly more prominent decadal variations after 1980).

The T2M products were subject to further quality controls where the large-scale spatial structures were tested by using part of the data to calibrate empirical downscaling models and the remaining independent data for making predictions. As for the same test for the SLP data, these results also served as a test of stationarity between the large-scale spatial patterns and local variability. Figures 22a-c show the predictions of Bergen January temperatures, and it is evident that the P-product yields a better prediction than OI. The predictions based on the OI-products have large errors in the early part of the record. Figure 22d shows a spatial map of the regression coefficients, indicating the spatial structure of the large-scale temperature pattern that is related to the January temperature in Bergen has opposite polarities over the Labrador Sea and Scandinavia, and resembles the temperature pattern associated with the NAO.

Thus, whereas the SLP OI-product is marginally superior to the SLP P-products in terms of these test results, the best-fit T2M maps (P-products) appear to be more reliable than the OI.

3.5 Testing the data fields in conjunction with prediction schemes

Figure 23 shows the results from 4 tests using common EOF based empirical downscale models with different T2M products. The two upper panels show tests with the same T2M data (standard P), but for which the models have been calibrated over different periods. In both cases, the models give a good reproduction of the January temperatures for Oksøy, and the best-fit linear trend estimates for 1980-1998 are similar. The test P results gives a poorer prediction, and the best-fit linear trend is even negative because it misses the recent upswing. The OI data, on the other hand, suggests a strong warming since 1980, and captures the recent warming but exaggerates the 1945-1960 mildly warm period.

3.6 Comparison between large-scale climate patterns

A CCA on the P (left) and OI (right) T2M products (Figure 24) suggests similar large scale structures, but with different details. The leading CCA pattern for OI T2M suggests stronger weights over Labrador Sea and Finland, but also has a slightly more "noisy" appearance. For the second CCA patterns, there are differences over eastern Canada and Labrador, as well as over the Greenland-Iceland-Norwegian Sea and Eastern Europe.

4 Coupling between large-scale SLP and T2M patterns

It is interesting to look at the coupling between different quantities which may reveal which physical processes are at work in the real climate system, but such analytical results can also be used in the evaluation of data. Figure 25 shows the two leading patterns from a CCA between the OI SLP and the P T2M analyses. The leading CCA pattern reproduces the well known NAO signatures in the T2M (left), characterised by anomalies of opposite polarities over Labrador Sea and northern/eastern Europe (a), and SLP with the north-south dipole centered over Iceland and the Azores (b). A strong NAO (inverse of the leading CCA patterns shown here) implies cold conditions over Labrador and warmer climate over northern/eastern Europe. Frigid polar air is advected (geostrophic wind) from the polar regions down over Labrador, whereas milder maritime air is transported across the Norwegian Sea to Scandinavia.
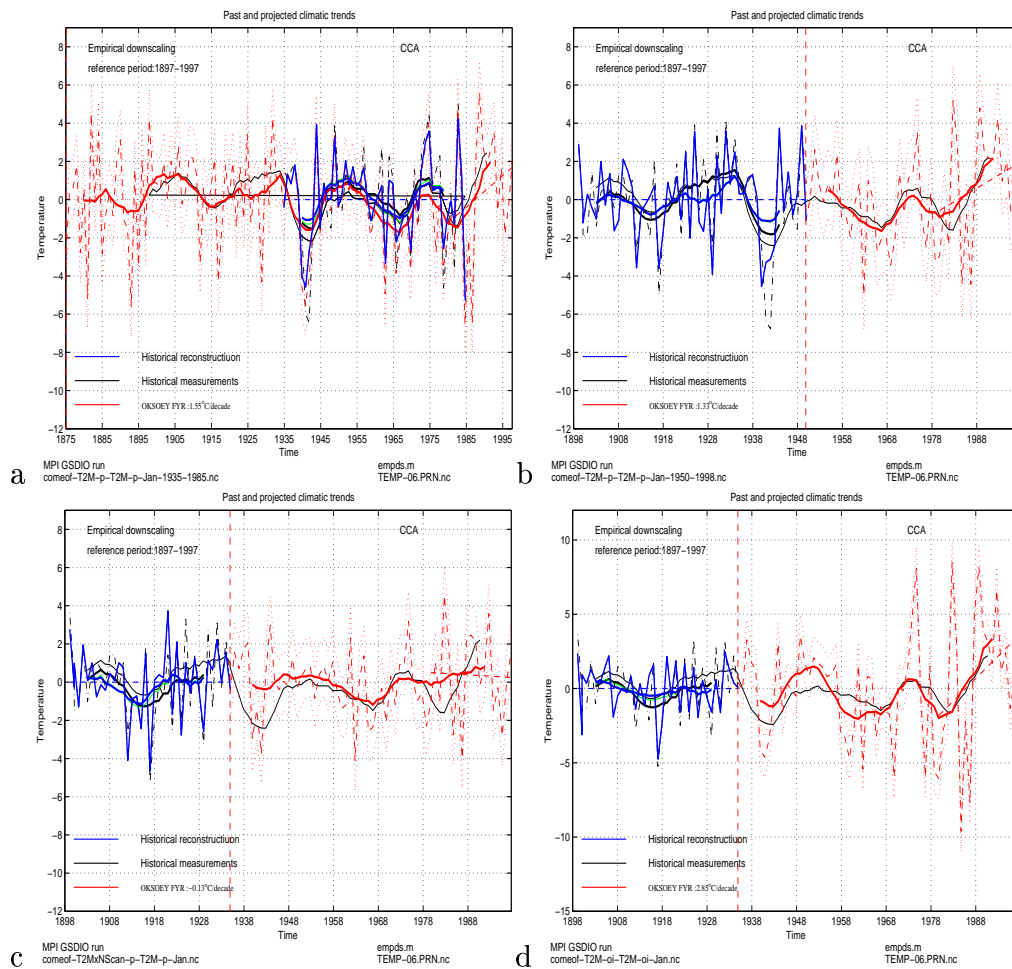
Figure 23: *Using separate periods for the calibration and evaluation of the CCA downscaling models to predict January temperatures for Oksøy. The results were derived using common EOFs for the standard P T2M (a and b) data, test P T2M (c), OI T2M (d), and the same routine employed in development of climate scenarios based on AOGCM results.*
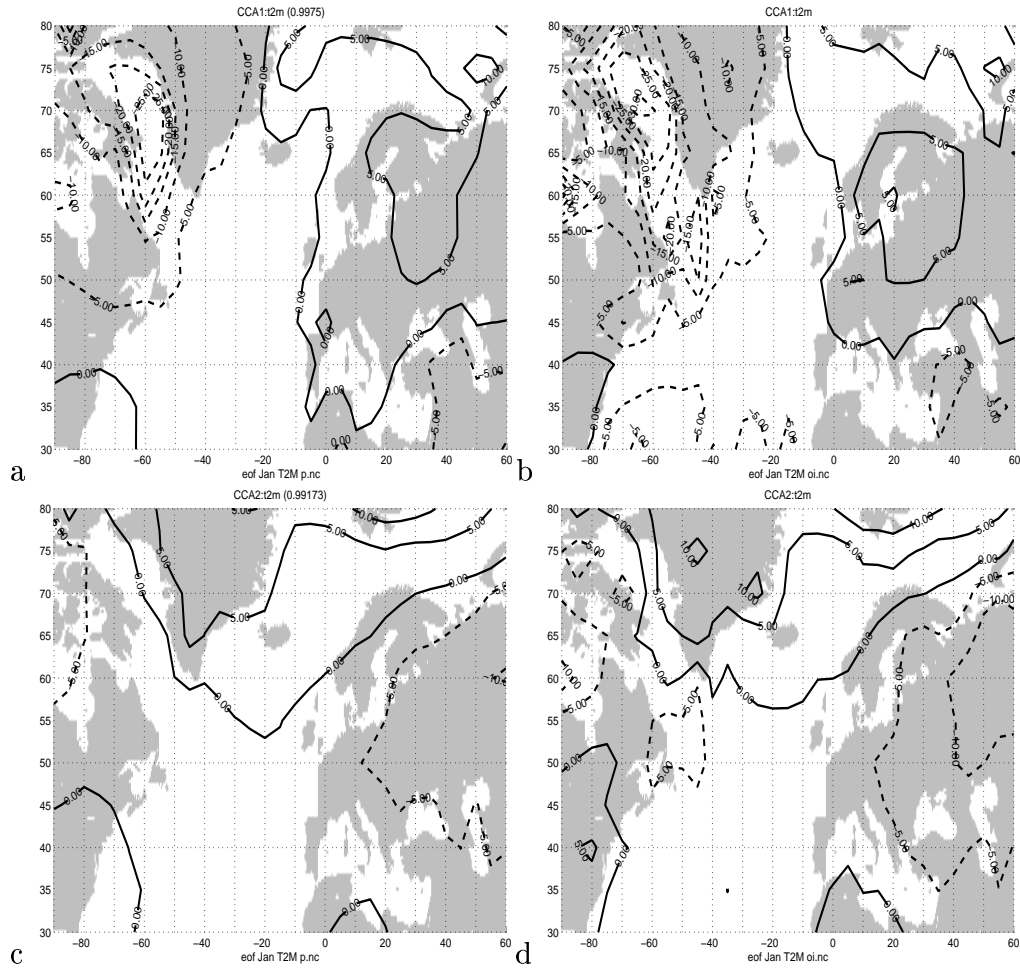
Figure 24: *The two leading CCA patterns for OI and P T2M indicating similar spatial structures with similar temporal evolution.*
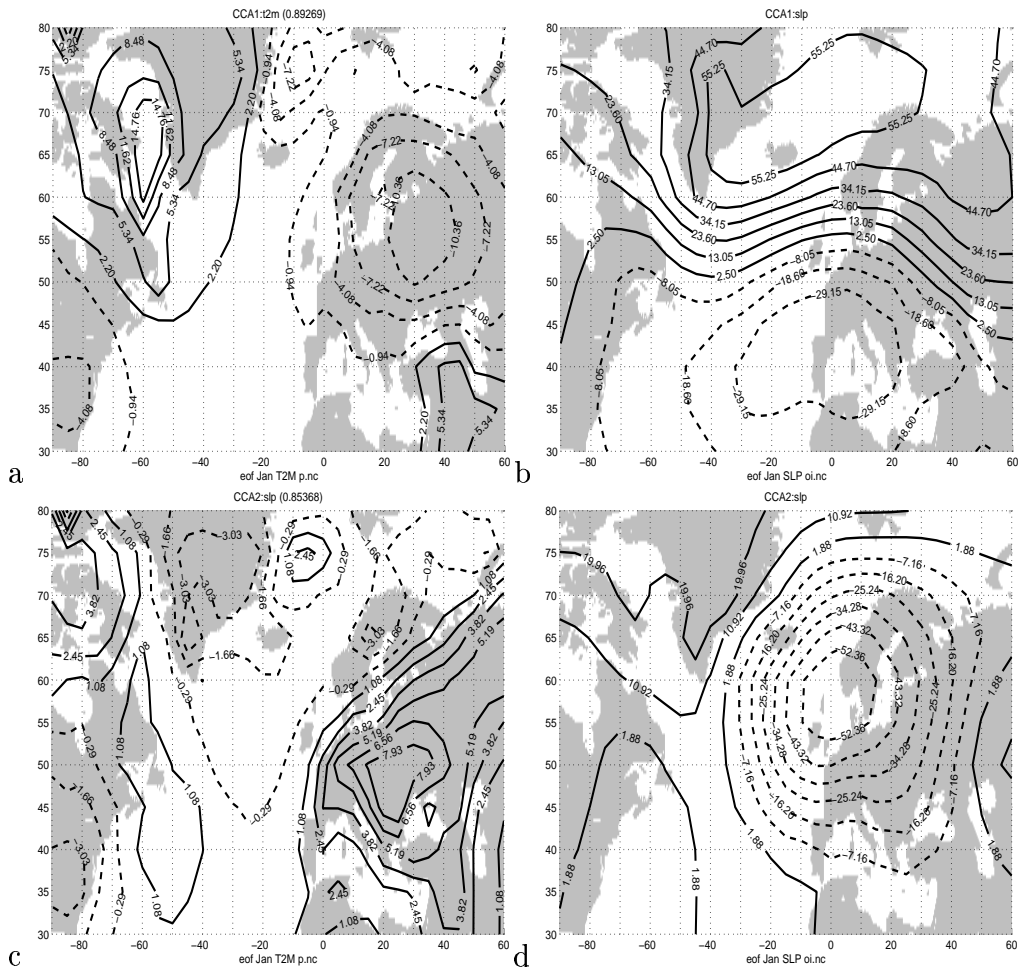
Figure 25: *The two leading CCA patterns for OI SLP and P T2M showing the geographical distribution of the coupling strength between SLP and T2M.*

The second CCA pattern describe warming (cooling) over central Europe (c) which is connected with a low (high) pressure system over the North sea (d). High pressure systems over the North Sea are often associated with cold outbreaks over central Europe as cold air is advected southward from eastern Russia. The temperature and pressure weights other than those associated with the centres of action in the second CCA patterns are weak, and it is uncertain whether there is any significance attached to these weak signals.

5   Discussion

Data reconstruction based on a few observations will always be a risky business because the amount of information really is limited to those observations. However, if there is some *a priori* information about the subject, then it may be possible to fill in some of the unknowns. In climate studies, one can use expectations about how the system behaves as *a priori* information. It is for instance possible to make predictions based on physical principles to give a more complete picture of the climate system, and this has been done in several data (re)analysis projects such as the NCEP reanalysis II. The only draw-back is that we have no complete knowledge of the physical processes that may play a role in the climate system, the climate models are far from perfect, and there is always the danger of producing misleading results. This problem may be illustrated by comparing different analysis products, such as the ECMWF reanalysis (ERA) and the NCEP reanalysis, which sometimes give quite different values (eg *Arpe & Rhodin* (1999)). However, extensive quality control and validation justifies the use of the reconstructed data both based on physical models as well as statistical methods.

It is expected that there also are some errors in the observations, for instance due to the influence of the local environment, instrument biases, mis-reading and mis-typing of the observations, relocation of stations, changes to the instruments and observational practices, or inappropriate analysis. The reconstruction of the historical data using spatial modes from the NCEP reanalysis may have filtered out some noise and may even corrected for some systematic errors, such as fictitious Arctic high pressure (*Jones*, 1992) and underestimation of variance in the T2M (*Benestad* et al., 1999).

The primary purpose for producing these data sets is for use in empirical downscaling studies of future climate scenarios. Such studies require long time series, and these products represent the best attempt to meet these requirements. Since empirical downscaling implicitly assumes a stationary relationship between the large-scale climate patterns and the local climates, the fact that these data sets were produced under the same assumption does

not pose any further limitations to the validity of the downscaling results. Tests for stationarity gave positive results, suggesting that these products are suitable for empirical downscaling studies. However, these limitations may be more important if these data sets are used for other types of studies, and in some cases, these products may even be unsuitable. We have shown that these products are inappropriate for detailed regional climate studies, other than empirical downscaling. The data quality is furthermore questionable over the Arctic region. The data may be used for validation of *large-scale* spatial structures of climate anomalies in climate model studies for non-polar regions.

The data products were validated in conjunction with empirical CCA downscaling models based on the common EOF method described by *Benestad* (1999b), and the conclusion from these validation results is that the empirical models derived from these reconstructed fields are capable of capturing most of the temperature changes associated with large-scale climate anomalies.

An important question is whether the reconstructed data based on UEA in the early period and the NCEP re-analysis when these become available are subject to discontinuities or inhomogenouities. The tests so far do not suggest that there are such problems with the end products. For example, the use of SLP products to predict the January temperature for Bergen produce similar results as models based on the UEA data. It is also important to keep in mind that the UEA may not be entirely homogeneous either, as this data set consists of a merge between various archives. Here the view of these analyses is that these data sets give approximately the right picture of the past 126 years of climate evolution, however, these data are not expected to be exactly right. Only the extensive use of these data sets will therefore reveal how reliable these are. Extensive testing of these products is recommended in conjunction with climatological studies. These data sets will be used in the Norwegian *RegClim* (Regional climate developement under global warming) project.

6   Summary

Two sets of almost stationary and homogeneous 1873-1998 SLP and T2M data have been reconstructed with no missing values. Both projection and OI data sets were produced, and a set of quality tests suggest that the OI-product is superior for the SLP but is inferior to the P-product for T2M. A number of quality tests described herein suggest that these products reproduce most of the large-scale climatic evolution. Evaluation against station

records suggest some discrepancies, but since only the large-scale climate variability was analysed, one cannot expect a perfect reconstruction of local time series. These data sets are particularly suitable for use in empirical downscaling studies and the study of large-scale climate variability. The products, however, are unsuitable for the description of local and regional climates.

7   Acknowledgement

References

Arpe, K., & Rhodin, A. 1999. Differences in the hydrological cycle from different reanalyses - which one shall we believe? *Page 42 of:* ECMWF (ed), *Second International Conference on Reanalysis*ECMWF, for ECMWF/WCRP.

Barnett, T.P. 1999. Comparison of Near-Surface Air Temperature Variability in 11 Coupled Global Climate Models. *Journal of Climate*, **12**, 511–518.

Benestad, R.E. 1998a. *CCA applied to Statistical Downscaling for Prediction of Monthly Mean Land Surface Temperatures: Model Documentation.* Klima 28/98. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1998b. *Description and Evaluation of the Predictor Data sets used for Statistical Downscaling in the RegClim.* Klima 24/98. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1999a. *Conversion of the NCEP re-analysis data to the netCDF format and quality control.* KLIMA 31/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1999b. *Evaluation of the common EOF approach in linear Empirical Downscaling of Future ECHAM4/OPYC3 GSDIO Climate Scenarios.* KLIMA 35/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1999c. *Pilot Studies of Enhanced Greenhouse Gas Scenarios for Norwegian Temperature and Precipitation from Empirical Downscaling.* Klima 16/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 2001. A comparison between two empirical downscaling strategies. *Int. J. Climatology*, **accepted**.

Benestad, R.E., Hanssen-Bauer, I., Førland, E.J., Tveito, O.E., & Iden, K. 1999. *Evaluation of monthly mean data fields from the ECHAM4/OPYC3 control integration.* Klima 14/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Bretherton, C.S, Smith, C., & Wallace, J.M. 1992. An Intercomparison of Methods for finding Coupled Patterns in Climate Data. *Journal of Climate*, **5**, 541–560.

Deser, C., & Blackmon, M.L. 1993. Surface climate variations over the North Atlantic ocean during winter: 1900-1989. *Journal of Climate*, **6**, 1743–1753.

Frich, P., Alexandersson, H., Ashcroft, J., Dahlström, B., G.R., Demarée, Drebs, A., van Engelen A.F.V, E.J., Førland, Hanssen-Bauer, I., Heino, R., Jónsson, T., Jonasson, K., P.Ø., Nordli, Schmidth, T., Steffensen, P., Tuomenvirta, H., & Tveito, O.E. 1996. *North Atlantic Climatological Dataset (NACD Version 1) - Final Report.* Scientific report 1. DMI.

Grotch, S., & MacCracken, M. 1991. The use of general circulation models to predict regional climate change. *Journal of Climate*, **4**, 286–303.

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., & Wigley, T. M. L. 1998. Northern Hemisphere surface air temperature variations, 1851–1984. *J. Clim. Appl. Met.*, **25**, 161–179.

Jones, P.D. 1992. The early twentieth century Arctic High - fact or fiction? *Climate Dynamics*, **1**, 63–75.

Jones, P.D., Jonsson, T., & Wheeler, D. 1997. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *International Journal of Climatology*, **17**, 1433–1450.

Kaplan, A., Cane, M.A., Kushnir, Y., Clement, A.C., Blumenthal, M.B., & Rajagopalan, B. 1998. Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, **103**(C9), 18,567–18,589.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. 1989. *Numerical Recipes in Pascal.* Cambridge University Press.

Reynolds, R.W., & Smith, T.M. 1994. Improved global sea surface temperature analysis using optimum interpolation. *Journal of Climate*, **7**, 929–948.

Slutz, R.J., Lubker, S.J., Hiscox, J.D., Woodruff, S.D., Jenne, R.L., Steurer, P.M., & Elms, J.D. 1985. *Comprehensive Ocean-Atmosphere Data Set; Release 1.* Tech. rept. Climate Research Program, Boulder, Colorado.

Strang, G. 1995. *Linear Algebra and its Application.* San Diego, California, USA: Harcourt Brace & Company.

Torrence, C., & Compo, G.P. 1998. A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61–78.

Wilks, D.S. 1995. *Statistical Methods in the Atmospheric Sciences.* Orlando, Florida, USA: Academic Press.