![Meteorologisk institutt logo] Meteorologisk institutt met.no

Month-to-seasonal forecasting at met.no

Report no. 15/2002

R. E. Benestad

KLIMA
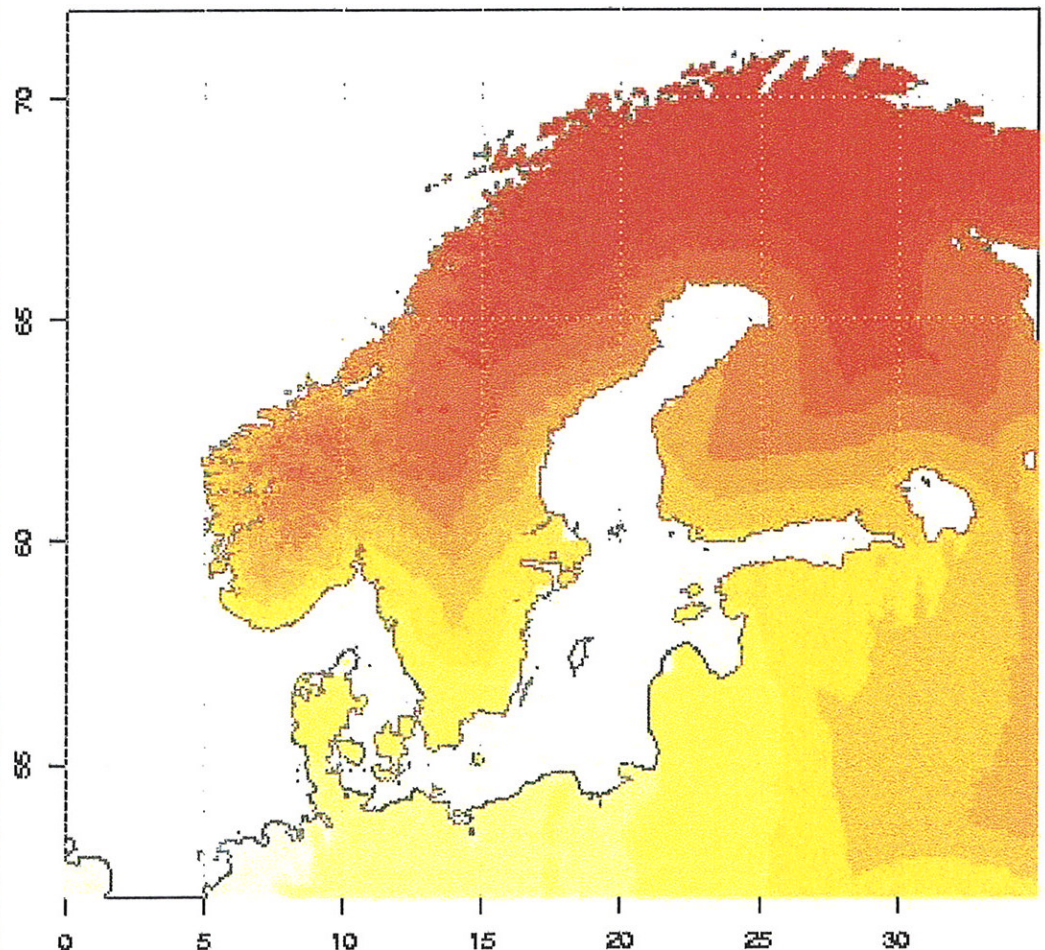
met.no REPORT

NORWEGIAN METEOROLOGICAL INSTITUTE
BOX 43 BLINDERN , N - 0313 OSLO, NORWAY

PHONE +47 22 96 30 00

TITLE:
Month-to-seasonal forecasting at met.no

AUTHORS:

Rasmus E. Benestad

PROJECT CONTRACTORS:

Natsource Tullett Scandinavia

SUMMARY:

Maps of monthly mean sea surface temperature and sea level pressure have been reconstructed for the 1900-2000 period. Analysis of the sea level pressure and sea surface temperature together with monthly mean station observations suggests a real relationship between certain spatial patterns and the subsequent monthly-to-seasonal mean temperature at various locations in Fennoscandia in winter. There is also a possible weak lag relationship in summer, but the results give no indication of any predictive skill in spring. Three prototype month-to-seasonal forecasting models have been tested.
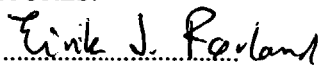
The test results suggest that there may be some predictability associated with monthly mean and seasonal mean temperature. At least part of the predictability of temperature is associated with the high winter-time autocorrelation. The monthly-to-seasonal mean precipitation is generally more difficult to forecast than temperature according to the estimated skill scores.

A new strategy for developing empirical month-to-seasonal forecasting models is proposed. Historical data contain some errors which affect the calibration of the empirical models. One way to deal with errors is to use of weighting functions based on data quality, but in many cases the the quality has not been established. The proposed strategy deals with errors by selecting only a few data points, potentially avoiding bad data. Long historical records of monthly mean observations are sub-sampled iteratively by selecting different random ensembles over a large number of times. The random-ensemble-model (REM) strategy produces a large number of forecasts that can be used to estimate a forecast probability distribution (p.d.f.). The REM model has been tested with climate station records as predictor, but is designed to use any kind of data (e.g. SST fields) as input.
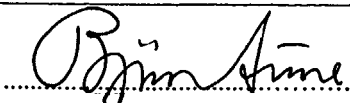
KEYWORDS:

month-to-seasonal forecasting, empirical studies, verification

SIGNATURES:

Eirik J. Førland
Section leader

Bjørn Aune
Head of Climatology Division

Contents

Oslo—Blindern T(2m) climatology

Month
Period: 1890—1999   a

Oslo—Blindern precipitation climatology

Month
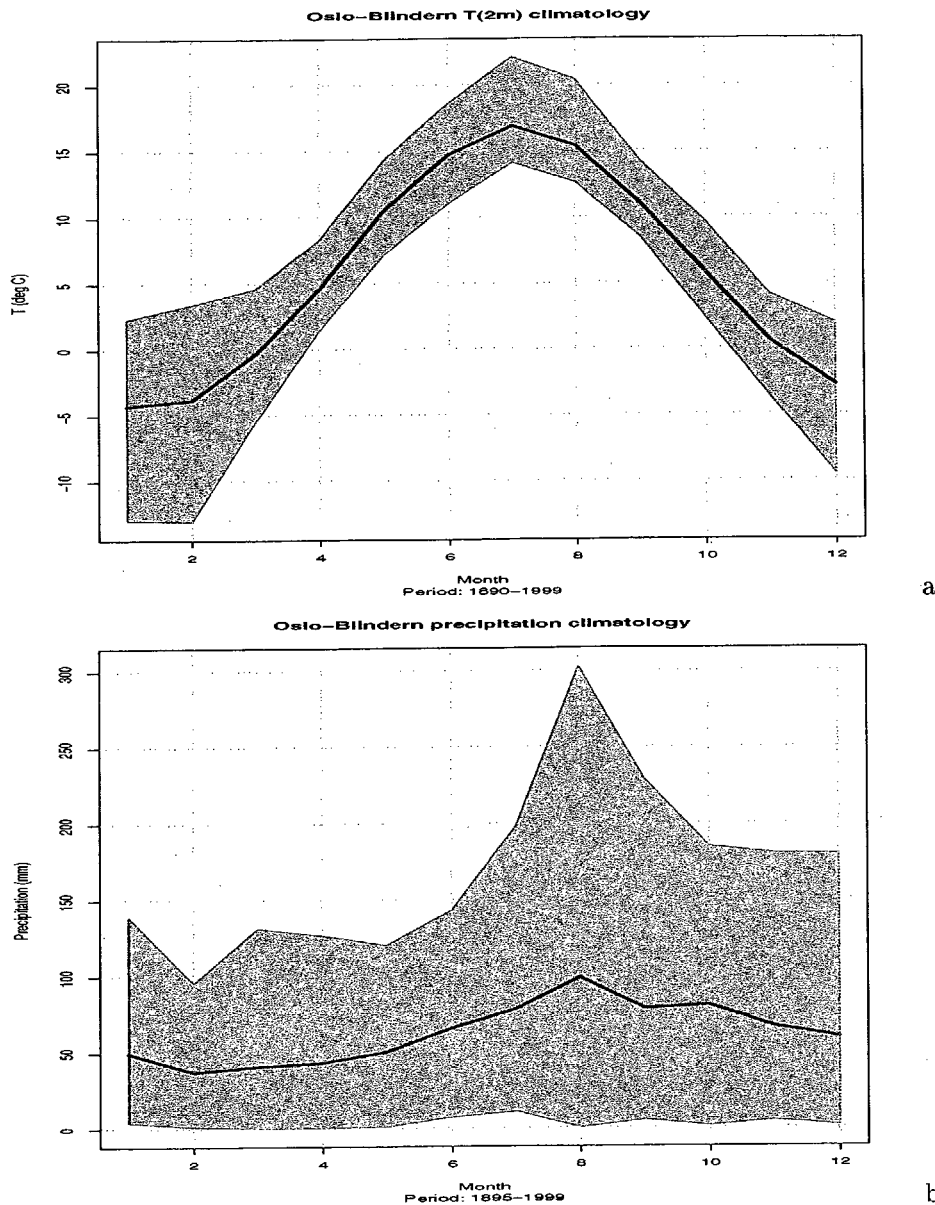**Period: 1895—1999**   b

Figure 0-1. (a) The annual temperature cycle and (b) annual precipitation cycle at Oslo-Blindern, estimated by taking the mean value of all the individual calendar months (January values, then all February values, and so on). The grey area marks the range (maximum and minimum) of values obtained for each month.

0.1 Introduction

It is well-known that the weather is difficult to forecast further ahead than approximately a week. *Lorenz* (1963) established that the atmospheric flow may in principle not be predictable beyond this time horizon. Small errors or perturbation in the initial conditions grow exponentially and may have consequences for the subsequent state of a complex non-linear system. The lack of predictability is a consequence of the so-called "chaos effect". In order to explore various possible outcomes for a weather system, dynamical atmospheric models have used an ensemble approach where many models are run in parallel, but use slightly different starting conditions. The predictability derived from these ensemble type integrations can be inferred from the spread in the solutions. One example of such ensemble forecasts is the seasonal forecasting carried out with a coupled atmosphere-ocean model at the European Centre for Medium-Range Weather Forecasts (ECMWF). A large scatter indicates little predictability. Experience also shows that the use of ensembles improves the weather forecasts.

The "chaos effect" may at first sight seem to indicate that it is not possible to make forecasts for months and seasons ahead. However, weather *statistics* is affected by the boundary conditions (external factors), such as the state of the surface and the energy input. A trivial demonstration of how the boundary conditions affect monthly temperature and precipitation in Oslo is illustrated in Figure 0-1. The temperature in Oslo follows a clear annual cycle, indicating that the temperature statistics are affected by external factors such as the solar inclination (i.e. the incoming solar energy). The annual cycle in the precipitation is less pronounced, which may suggest that the precipitation is less systematically affected by external factors. It is trivial to predict that the July temperature in Oslo will be higher than the January temperature. The question that we want to address is whether we can do better than predicting the climatology (the seasonal cycle).

There are *a priori* reasons to believe that it is in principle possible to do better than just assuming climatological values. For instance, the ocean transports an anomalous amounts of heat into the North Atlantic and the Norwegian Sea, and it is assumed that this heat at some later stage will affect the local climate. The mean influence of oceanic heat transport on the local climate is well-known. Norway, for instance, experiences a substantially milder winter climate than Siberia.

Anomalous snow cover or sea-ice extent may affect the regional albedo, and may therefore influence the regional energy balance both through an albedo effect and because snow melt may require a substantial amount of energy. The physically-based reason for month-to-seasonal predictability seems to be supported by empirical evidence. *Benestad & Tveito* (2002) carried out a survey on possible teleconnections between factors such as sea-ice and local temperature and precipitation variations in the Nordic countries and found indications that suggest there is a connection in spring and autumn between the sea-ice and the subsequent month's temperature (R^2 as high as ~ 0.3). Regression analysis with one-month-lag also indicated a relationship between SSTs in the North Atlantic and temperature anomalies in southern Scandinavia, in agreement with results by *Benestad* (1999a). *Benestad & Tveito* (2002) furthermore reported statistical relationships between the sea-ice and precipitation for the subsequent month over southern Scandinavia during late winter–spring and in the autumn (highest $R^2 \sim 0.3$).

If the surface conditions change slowly, then it is plausible that some of the weather statistics can be predicted over longer time ranges. For instance, the oceans and sea-ice have much higher thermal inertia than the atmosphere and tend to vary more slowly, and the ocean surface may contain precursors for the subsequent atmospheric evolution.

Traditionally, there has been little serious work on forecasting with a time horizon of months to seasons. Only recently, with the advent of a rapid increase in computational power, improved understanding of our climate system, and the improved availability of data, has month-to-seasonal forecasting become feasible. Initially, month-to-seasonal forecasting focussed on the El Niño Southern Oscillation which has a preferred time scale of 2–8 years.

The Norwegian Meteorological Institute (met.no) has recently carried out a number of studies on the potential predictability in Scandinavia (*Benestad*, 2001; *Førland* et al., 1999; *Benestad*, 1999a; *Førland & Nordli*, 1993). A number of different approaches has been utilized for the prediction of subsequent months and seasons. The simplest models consist of statistics such as autocorrelation (*Førland & Nordli*, 1993) or conditional probabilities (Figure 0-2). The observational records suggests long-term trends (warming), such as seen in Figure 0-2(a), and the question is whether these trends are predictable and can be utilised in month-to-seasonal prediction. Figure 0-2(b) suggests that there seems to be a statistical relationship between the January temperature and February temperature in Oslo.

In addition to simple forecasting approaches based on statistics, three more advanced models will be discussed in this report:

a A simple test-model used for hindcast studies (henceforth "test-model").
b A prototype operational model (henceforth "DNMI-model")
c A new model designed for operational use (henceforth the "REM" model)

In addition, reference is made to a "benchmark" model which involves taking the mean of the previous 10 years of anomalies corresponding to predicted value.

A distinction will be made in this report between the terms "prediction", "forecast", and "hindcast". The term "forecast" will refer to predictions made for the future, whereas a "hindcast" is a prediction made for the past but where the values that we are trying to predict have not been used in model calibration (i.e. independent values). The term "prediction" is used in the general sense referring to the

Feb

Time
1890 — 2002   a

Histograms for Feb temperatures
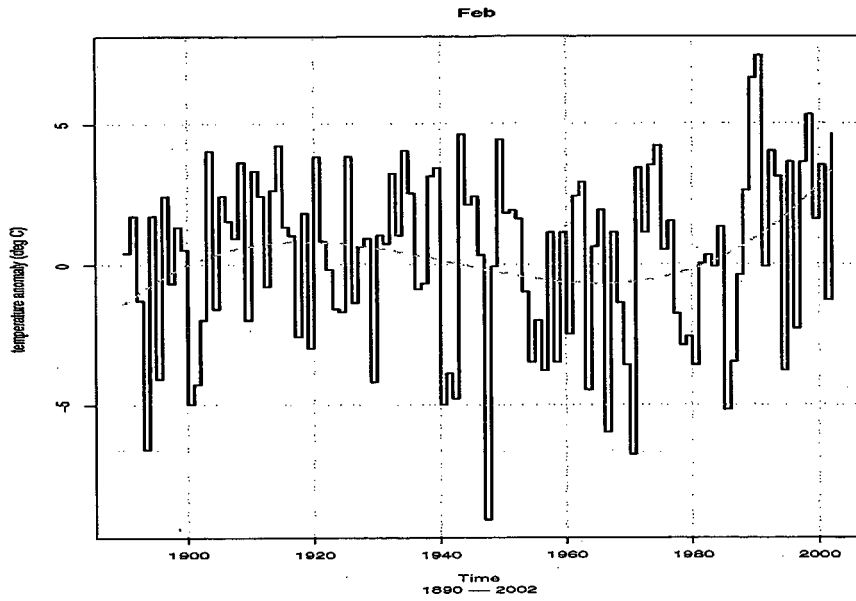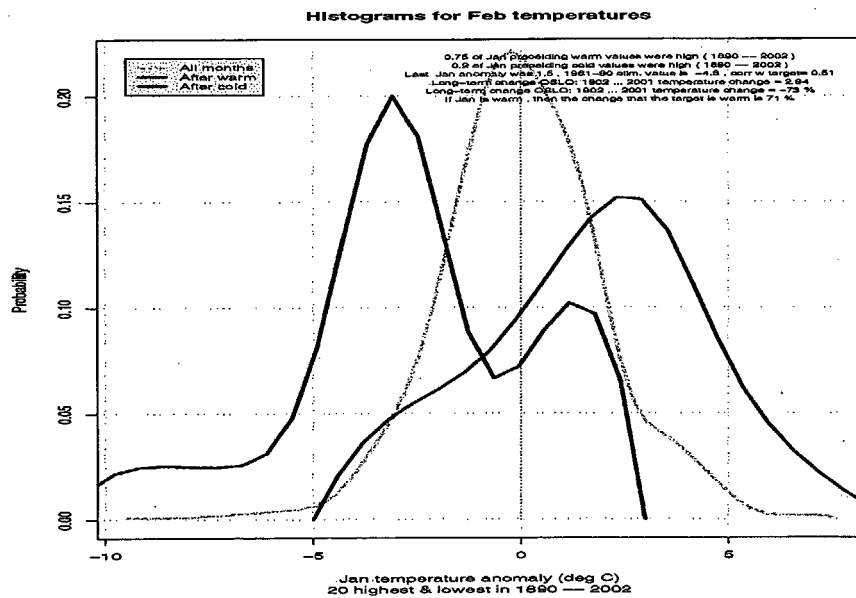
Jan temperature anomaly (deg C)
20 highest & lowest in 1890 — 2002   b

Figure 0-2.   (a) A time series of the February mean temperature at Oslo-Blindern shown together with an estimate of the long-term trend. (b) The distribution of the February temperature for the entire record (grey) and of those February months following the 20 warmest (red) and coldest (blue) January months.

model output. The term "seasonal" in this report normally refers to a 3-month average, which can start in any calendar month (i.e. March–May). The word "climatology" refers to the annual cycle, and an "anomaly" is the deviation from this cycle. A month-to-seasonal forecast may not be just a prediction of a future event, but may also entail a comprehensive analysis of expected errors, probabilities and prediction skill.

This report serves as a documentation of the research into month-to-seasonal forecast at met.no carried out as a pilot programme in partnership with Natsource Tullett (then CBF).

The outline of this report is as follows: a Method & Data chapter is followed by a chapter describing the main results of these studies. The results from the test-model are described as well as the DNMI-model and preliminary results with the REM-model. The reason for describing three different models is that the test-model and the REM-model are designed to use the same input for model calibration. The DNMI-model uses different data for input, and the reason was that it was desirable to set up a model quickly and start making operational forecast for Natsource Tullett Scandinavia on a test-basis.

A more detailed description of these models is given in the following sections. This report concludes with a discussion and conclusions.

TABLE 1. A summary of sources of data used for the reconstruction of the 1900-2000 monthly mean North Atlantic SLP and SST records.

source	period
Benestad (2000) SLP	1873 – 2000
NCEP (*Kalnay* et al., 1996)	1947 – present
COADS SST	1856 – 1992
GISST2.2 SST	1904 – 1994
Reynolds & Smith (1994) SST	1981 – present
Kaplan SST	1856 – 1991
met.no climate archive	~1900 – present
ECMWF surface analysis	January 1985 – present
ECMWF ERA-40 (not yet available)	1958 – 1997

0.2   Methods & Data

0.2.1   Model development

Empirical forecasting requires long and high-quality (homogeneous) data sets for model calibration. The models furthermore need constantly updated predictor fields if they are to be used in an operational forecasting system. The long regional or global data records tend to consist of analyses made for past observations, but a problem is that many of these are not continuously updated (Table 1). The data which are continuously updated, on the other hand, tend not to be sufficiently long for model calibration, possibly with one exception for the NCEP (*Kalnay* et al., 1996) data. Furthermore, comparison of different data sets over a common time interval tends to reveal differences (Fig. 0-3). Even the most recent data based on climate model assimilations, such as the NCEP reanalysis, may suffer from data problems (see appendix A). Thus, the main problem associated with making empirical month-to-seasonal forecasting is the lack of long high quality data records that are continuously updated.

0.2.2   The empirical data

In order to develop and thoroughly test the empirical downscaling methods, 101-year long data records of sea level pressure (SLP) and sea surface temperature (SST) were constructed for the period (1900-2000) by combining various data sets. Since there are differences between the data from the different sources (Figure 0-3), the combination of the different sets is not trivial. The data furthermore come on different grids with different resolution, different time span, and with gaps of missing data. But, as long as the reconstructed record contains a real signal that produces useful predictions, these caveats do not necessarily have to be a problem. The reconstruction of the data was done by projecting past data onto their respective eigen-patterns (U)(Figure 0-4) estimated from the *Reynolds & Smith* (1994) SST and NCEP SLP. The *Benestad* (2000) SLP reconstruction utilized a technique known as optimal interpolation (*Reynolds & Smith*, 1994).

Sea surface temperature reconstruction

Composites of data from sources with valid data have been synthesized at met.no, and these have then been used in a regression analysis in order to project the eigen-patterns from the recent and complete (no "missing data" gaps) *Reynolds & Smith* (1994) SST. The *Reynolds & Smith* (1994) SSTs are routinely updated and are used as predictor in the DNMI-model monthly-to-seasonal forecasting system (the same data are also available though ECMWF, and will be used in the REM-model as well). It is therefore important to use data which are consistent with this set. Figure 0-4 shows a comparison between the composite and reconstruction by projection, and a reasonably good match is evident. However, subsequent attempts to merge the SST reconstructions with the continuously updated Reynolds SSTs from the Climate Diagnostic Center (CDC)* in the USA suggest that there are differences in the two data sets that may cause problems and affect the updated forecasts (Figure 0-5, difference between solid and dashed lines). The data are not without errors, partly because of errors in the observations themselves,

*URL: http://www.cdc.noaa.gov/
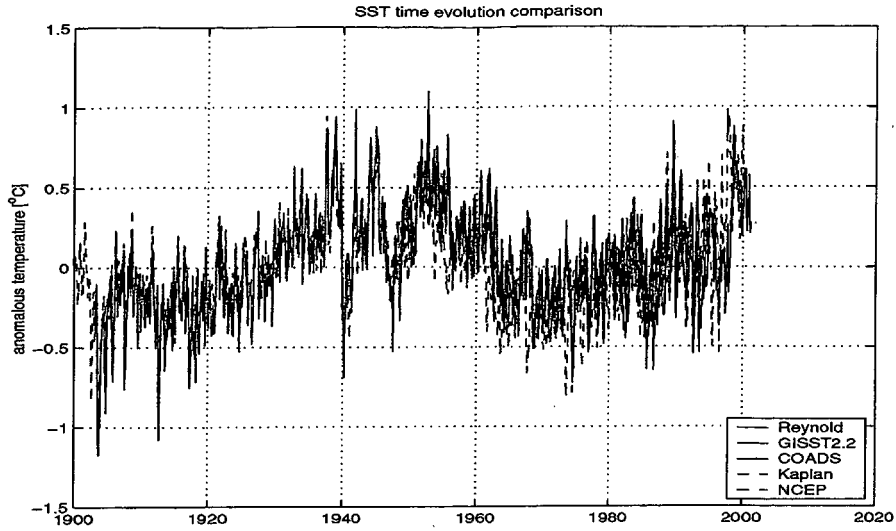
Figure 0-3. Comparison between 5 different SST values estimated for a common area.
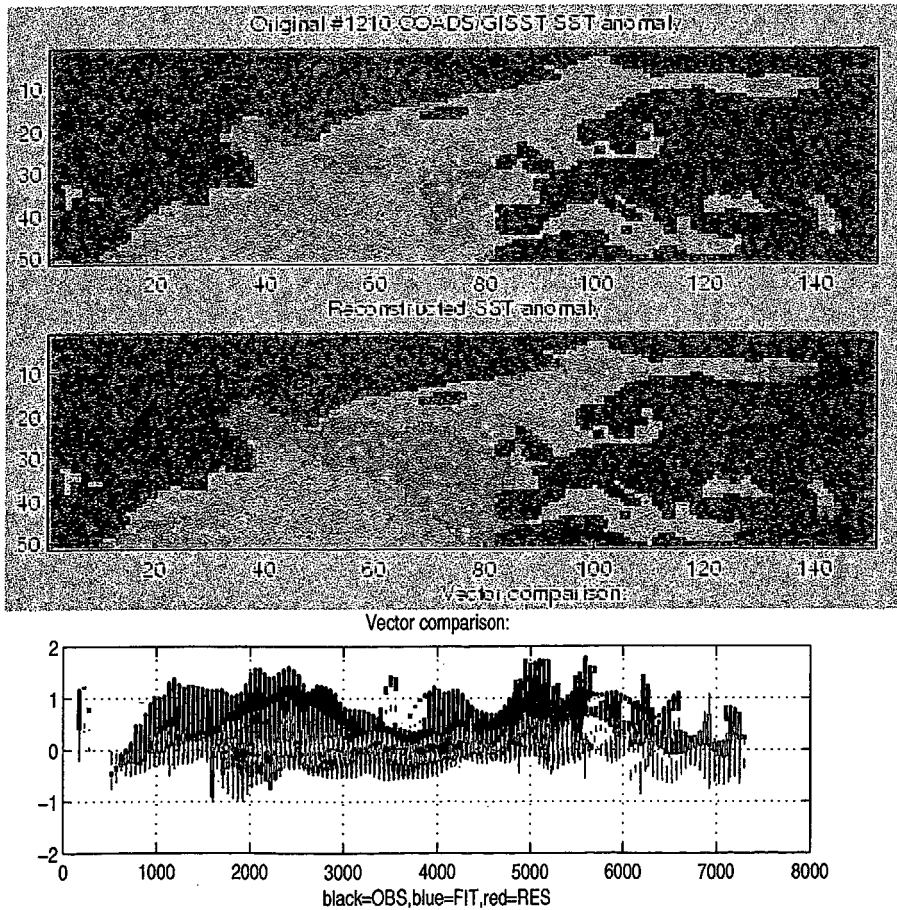
Figure 0-4. Comparison between the 5-source SST composite (upper) and reconstruction by eigen-pattern decomposition (middle). The lower panel shows the same results as a vector where the difference is given in red.

PCs and reconstructed+updated PCs
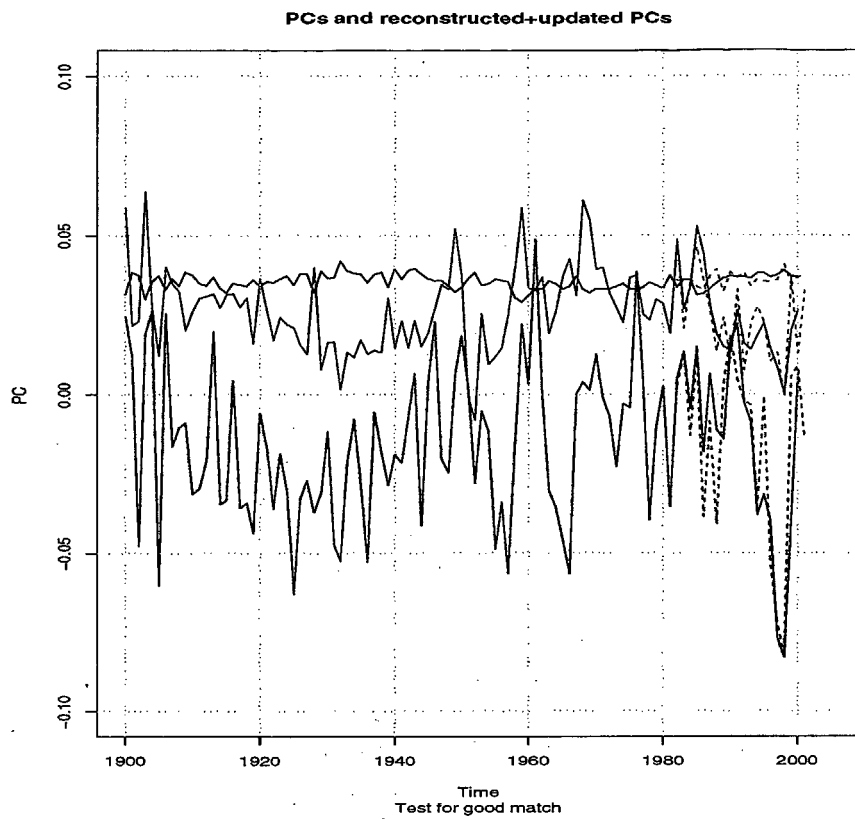
Figure 0-5.   An attempt to merge the historical SST reconstruction (solid lines) with the modern *Reynolds & Smith* (1994) SSTs (dashed lines) suggests that there are differences that may cause problems. The black line shows the leading PC, the red line marks the second and the third PC is shown in blue.
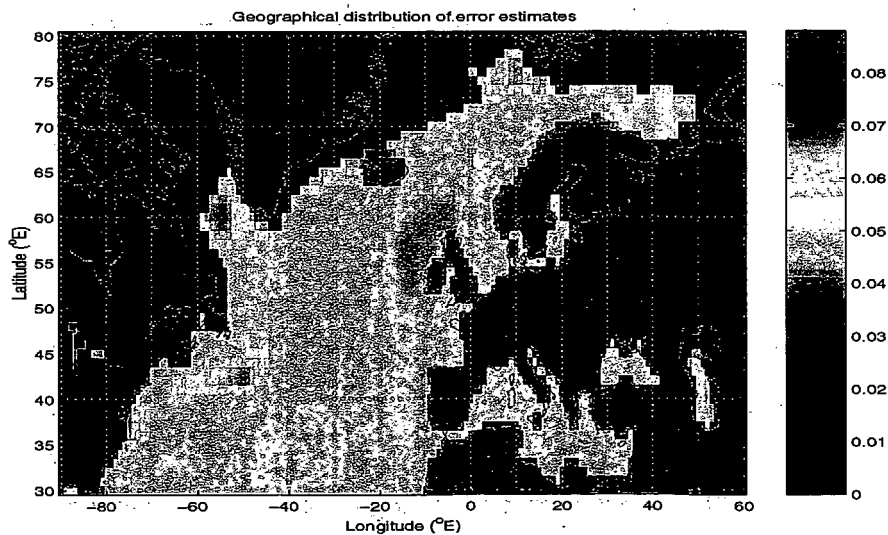
Figure 0-6.   Geographical distribution of rms error between 5-source SST composite and eigen-pattern reconstruction for the period 1981-1995.
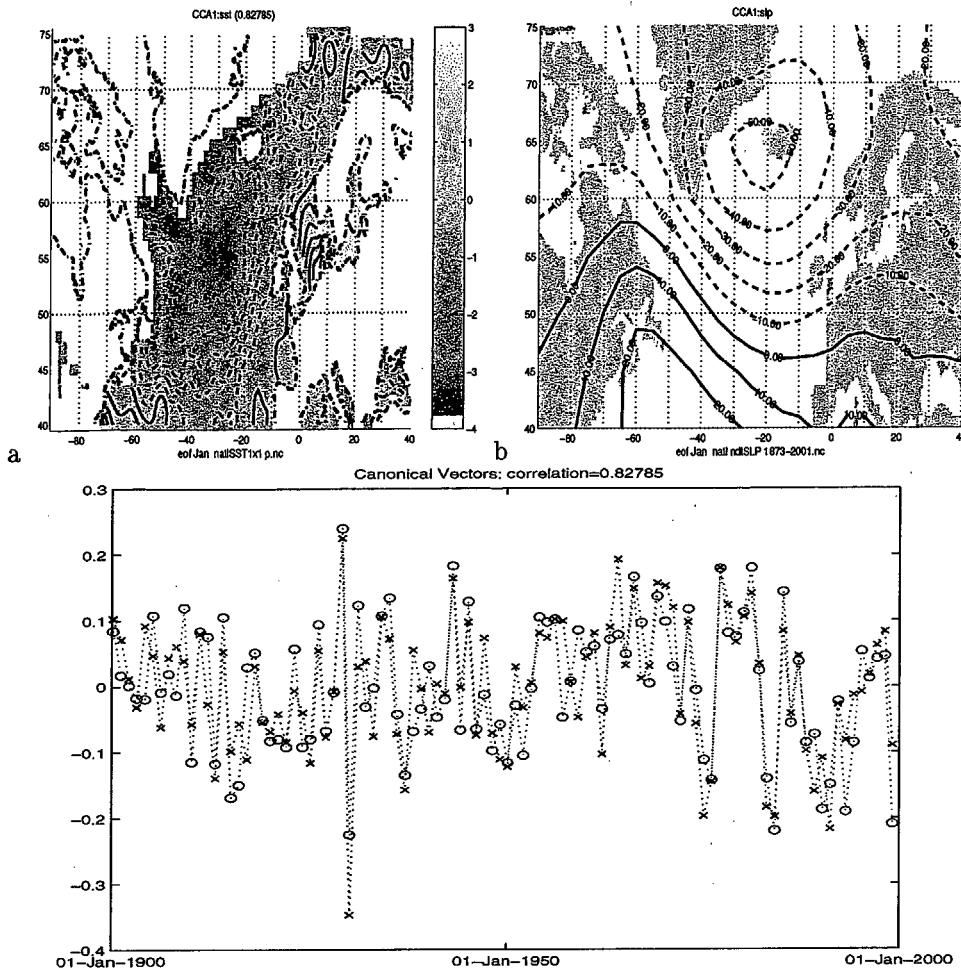
Figure 0-7. The SST (a) and SLP (b) anomalies which are most correlated and their time evolution (c). These results are the leading patterns that emerge from a CCA.

but also due to spatial interpolation and problems associated with the merging of data from different sources.

Figure 0-6 shows the geographical distribution of *root-mean-square error* (rmse) between the composite and the reconstruction. The largest errors are found in the Labrador Sea and east of Greenland. There have been some studies which suggest that the North European weather is sensitive to the surface conditions in the Labrador Sea (*Skeie & Kvamstø*, 2000). Relatively large errors can also be seen along the ice edge.

Sea level pressure reconstruction

The sea level pressure (SLP) data and the 2-meter temperature [T(2m)] fields are described in *Benestad* (2000). These products are based on a combination of the NCEP reanalysis (*Kalnay* et al., 1996) and historical data analyses (*Jones*, 1992; *Jones* et al., 1998) from University of East Anglia Climate Research Unit (UEA CRU) and the Comprehensive Ocean Atmosphere Data Set (COADS) (*Slutz* et al., 1985).

Evaluation of the data quality

One way to test the reconstructed data for a signal that may be used as a precursor for month-to-seasonal forecasting is to apply a canonical correlation analysis (CCA) to the SST and the SLP fields in order to extract spatial patterns which are associated with one another. Results from such an analysis are shown in Figure 0-7, where the SST and the SLP patterns are shown together with their time evolution
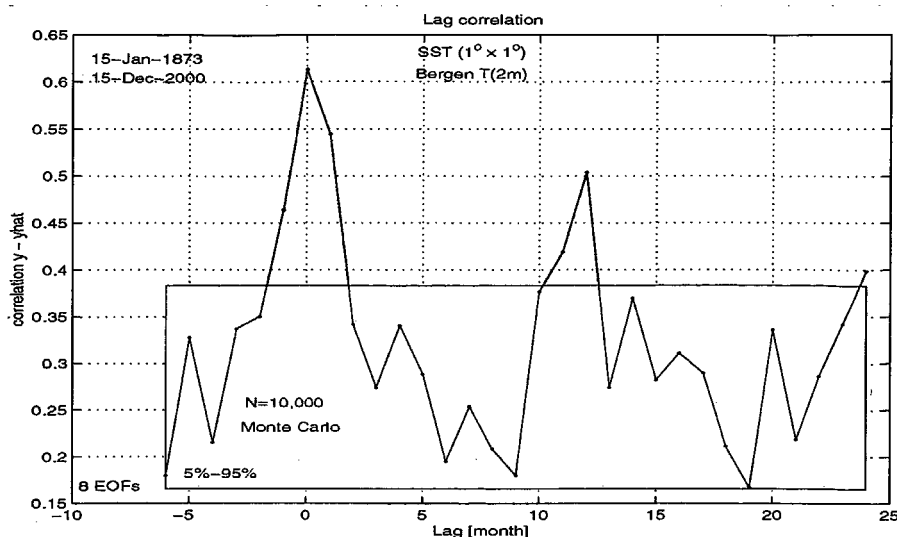
Figure 0-8. Lag correlation plot between Bergen 2-meter temperature and least-squares regression based on January SST anomalies (constructed). Positive lag is equivalent to SST leading Bergen temperature. The filled area show 5%-95% confidence interval based on a white noise 1000 member Monte Carlo test.
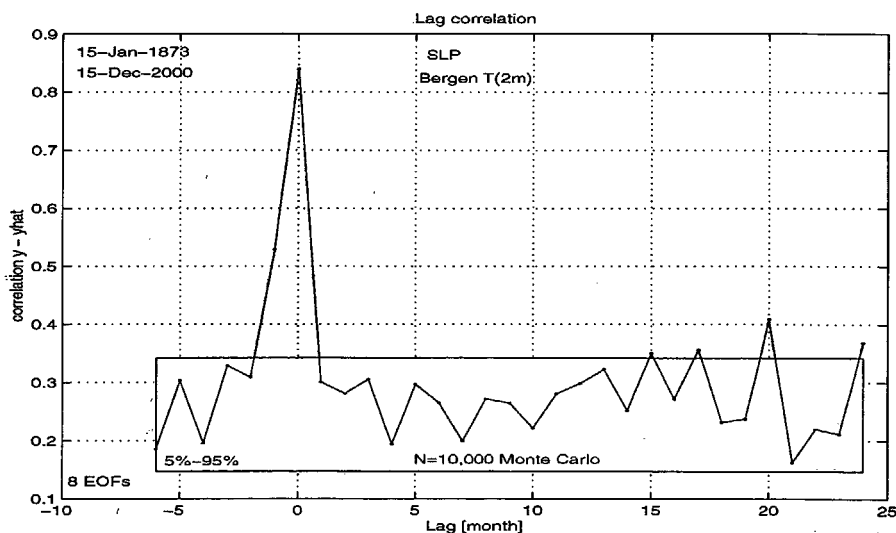
Figure 0-9. Similar to Figure 0-8, but lag correlation plot between Bergen 2-meter temperature and least-squares regression based on January SLP anomalies.

and their correlation. The patterns bear similarities to the North Atlantic Oscillation (NAO), which is a well-known feature in the North Atlantic. The canonical correlation is 0.83, which is indicative of a coupled signal, but still not very high.

Although these data are far from error-free, one may expect from the agreement among the different sources and the CCA results that the reconstructed 101-year SST record (1900-2000) contains the main climatic signals useful for month-to-seasonal forecasting.

The evaluation of the empirical models' capability of predicting future monthly and seasonal mean temperature and precipitation will be based on models developed using the SST reconstruction mentioned above and updated SLP from *Benestad* (2000).

Least-squares fit between the 8 leading EOFs from the SST and SLP data and 2-meter temperature from Bergen (Figures 0-8 and 0-9) have highest correlation with the observations at zero lag. This test suggests that the time stamp of the reconstructed data is correct. These results furthermore confirm the presence of a signal in the SST and the SLP fields that can be associated with temperature variations in Norway. For SST, there is a significant signal at zero lag as well as at plus and minus one month,

but also at 12-month lag. High correlation for the subsequent winter can be explained in terms of the strong winter-time persistence in the oceanic heat anomalies* (*Bhatt* et al., 1998; *Sutton & Allen*, 1997). Furthermore, high correlation at negative lags may be due to December-to-January persistence in the Bergen temperature (or SST). For SLP, the one-month-lag correlation is low, whereas the minus one lag (December temperature) is well outside than the Monte Carlo 10%-confidence interval[†]. The SLP-based lag correlation suggests that there is some persistence in the December-to-January temperature (or possibly SLP).

Tests were conducted using zero lag (contemporary observations), and the models reproduced a good part of the temperature and precipitation in regression analyses against station data.

The input data for the test-model

The test-model used SST, SLP, sunspot numbers and the station data itself (e.g. for the preceding month) as input for the step-wise screening.

The input data for the DNMI-model

The DNMI-model used different predictor data than the test-model due to problems merging different data sets. The motivation behind the DNMI-model was to set up a month-to-seasonal forecasting system within a short period of time. In order to achieve this goal, the input data must be in a format that is easy to incorporate into the model system, and one common data format is Unidata's *netCDF*[*] which can be easily read into Matlab using the tool-box *mexcdf*[†]. The NCEP reanalysis data are available in the *netCDF* from the CDC. The predictor data from CDC included skin temperature, sea level pressure, soil wetness. In addition, the monthly mean sunspot number (FTP ftp.ngdc.noaa.gov) was included in the batch of selectors in the step-wise screening process and climate station data from met.no.

The predictor data used by the DNMI-model consist of short records, and one main disadvantage with this model is the greater uncertainty in the calibration due to data size. On the other hand, the DNMI-model does not require the merging of different data sets.

The input data for the REM-model

Temperature and precipitation station records are from the Norwegian Meteorological Institute's (met.no) climate data base. These records are used both as predictors (independent variables) as well as predictands (dependent variables) where a time lag has been introduced to the record. A combination of the reconstructions described above and the analysis from ECMWF will be used as input for the REM-model. The REM-model will also use a new approach using mixed data fields analogous "CPCA" to in *Bretherton* et al. (1992). The use of mixed data fields puts the focus onto so-called "coupled modes" involving more than one physical quantity. The DNMI-model utilised ordinary multiple regression involving several types of predictors, with a step-wise screening based on cross-validation to avoid over-fit.

0.2.3   Model construction: The mathematical framework

Time series representing spatial fields of geophysical data can be decomposed into left and right inverses using a method known as *singular vector decomposition* (SVD)(*Press* et al., 1989; *Strang*, 1995). The results of an SVD also include a diagonal matrix describing the power of the series, and the total decomposition looks like:

$$X = UW^{\frac{1}{2}}V^T \qquad (1)$$

The symbol X in equation (1) denotes the timeseries of data anomalies. We will use the vector notation to represent a spatial pattern of anomalies at a given time t ($\vec{x}_t$).

---

*The annual cycle does not affect these results since the analysis was made between SSTs and monthly sub-sampled temperature by selecting only one calendar month at the time (i.e. all January months).

[†]Estimated using random permutations of the a subsample version of of the data with zero serial correlation. The trends are small and have not been removed.

[*]URL: http://www.unidata.ucar.edu/packages/netcdf/

[†]URL: http://woodshole.er.usgs.gov/staffpages/cdenham/public_html/MexCDF/nc4ml5.html

The advantage of using the SVD products (principal components) is that it can be used to limit the numbers of predictors, and when used in regression the predictors are also orthonormal. Thus, the use of this method greatly reduces the computational demands. The matrices in equation (1) have the following properties: $U^T U = I$, $V^T V = I$, and $W^T = W$. The SVD products can be used in empirical models using spatial fields as predictors. A best-fit ($\hat{x}_t$) to the observed climatic anomalies ($\vec{x}_t$) can be expressed in terms of a combination of "eigen-patterns" (U):

$$\hat{x} = U\vec{\beta} \tag{2}$$

The coefficients $\vec{\beta}$ can be solved using an ordinary *least-squares* fit or solving the equation $\vec{\beta} = U^T \vec{x}$. The principal components (V) can be expressed in terms of the eigen-patterns and the observations: $V^T = W^{-\frac{1}{2}} U^T \vec{x}$. Using the best-fit combination of eigen-patterns ($\hat{x}$) instead of the actual observations gives:

$$\hat{V} = W^{-\frac{1}{2}} U^T U \vec{\beta} = W^{-\frac{1}{2}} \vec{\beta} \tag{3}$$

The eigen-patterns can be found using an SVD analysis of long historical records of data such as sea surface temperature or sea level pressure. Updated observations can then be expressed in terms of these patterns according to equation (2), where the weight of each eigen-pattern is contained in $\vec{\beta}$.

The principal components (V_t) representing the predictors at time t are used in the lagged regression (lag l) employed for the empirical model development.

$$\hat{y}_{t+l} = \vec{a}_1 + \vec{a}_2 V_t^T \tag{4}$$

The quantities $\vec{a}_1$ and $\vec{a}_2$ are model coefficients derived using a best-fit estimation such as least-squares fit.

The development of the month-to-seasonal forecasting included a thorough testing in order to ensure that the codes work the way they were intended. These tests ensure that the matching procedure for the spatial patterns work, the time stamp of the data are correct and match each other, and that the empirical models describe the predictors. All the models used a least-squares optimisation and a step-wise screening to reduce the risk of over-fit.

The test-model

The test-model was based on the mathematical framework outlined in section (0.2.3). The empirical model used a lagged least-squares regression to solve for $\vec{\beta}$ and then to make a prediction according to the statistical model based on the regression coefficients in $\vec{\beta}$: $\hat{Y} = X\vec{\beta}$. The predictors are selected through a step-wise screening process where the results from a cross-validation is used to determine whether a particular predictor should be included according to whether the correlation with independent data was improved.

The DNMI-model

SVD was used to combine data from various sources before the use in the DNMI-model. Each grid-box series was standardised before applying the SVD, and a weighting factor was applied to the station observations in order to increase its importance. The leading modes responsible for 75% of the variance (or a minimum of 5 EOFs) were included from the SLP and the SST data sets and used in the SVD analysis. Similarly, only the principal components that describe 75% of the variance (or a minimum of 15 EOFs) in the SVD analysis that combine the SLP, SST, and monthly mean station observations were used in the regression analysis. A cross-validation analysis was used to determine which predictors should be included in the model, and only those which increased the correlation score by more than 0.03 were included. The 1900-1959 period was used for model calibration whereas the independent 1960-2000 time interval was used for evaluation of the predictions.

TABLE 2.   Number of possible combinations of subsets with r elements from a population of N.

N	r	$^N C_r$
81	30	1.4×10^{22}
81	40	2.1×10^{23}
81	50	2.3×10^{22}
81	60	1.4×10^{19}
91	30	1.0×10^{24}
91	40	1.0×10^{26}
91	50	1.3×10^{26}
91	60	2.0×10^{24}
100	30	2.9×10^{25}
100	40	1.4×10^{28}
100	50	1.0×10^{29}
100	60	1.4×10^{28}

The REM-model

The REM-model adopts a similar mathematical framework as described in section (0.2.3). Instead of a cross-validation procedure for selecting the predictors to be included in the model, the step-wise regression uses the Akaike information criterion (AIC) (*Wilks*, 1995, 300–302). The evaluation of the REM model predictive skill is, on the other hand, based on a cross-validation analysis using the data not included in the model calibration. Recent progress in the month-to-seasonal forecasting system involves modification to the forecasting strategy:

i Add the long-term trend information in order to improve the skill.

ii Carry out a large number of predictions based on random subsets (ensembles) of the historical data → Random Ensemble Model (REM).

iii Improve the verification analysis: Hit-ratio table based on 5 categories [very low, low, around normal, high, very high].

Two new features (*i* and *ii*) have been introduced to the analysis in order to improve the forecasts whereas a third addition (*iii*) provides a new tool for assessing the forecasts.

As the climate is subject to a global warming (*IPCC*, 2001) there is a higher probability for seeing warmer than colder seasons in the future (see Figure 0-2(a)). This information may be utilized in the month-to-seasonal forecasting by simply extrapolating the recent observed linear trend to the near future.

Empirical-based prediction schemes have traditionally been deterministic in nature. One of the disadvantages associated with single deterministic forecasts is the information on the predictability, although some information is embedded in the goodness-of-fit (e.g. variance). Furthermore, ensemble forecasts may provide a crude basis for probability-based forecasts.

Even though empirical models do not explore the chaotic aspects of the systems, they nevertheless may be integrated in an ensemble type forecasts. One way to do this is to compile a forecast of many different types of models, each with different strengths and weaknesses. It is also possible to derive a set of different models by using different subsets (ensembles) of data. The number of different models, m, depends on the number of possible data combinations in terms of subset size, r, and total number of data points, N:

$$m = {}^N C_r = \frac{N!}{r!(N-r)!} \tag{5}$$

If one calendar month is extracted from the 1900-1990 period for calibration ($N = 91$) and subsets of 50 points are used for the calibration, then maximum number of combinations is 1.3×10^{22} (Table 2). Thus, by selecting 1000 different combinations at random, the chance of selecting two or more similar combinations is $P(x > 0) \approx 8 \times 10^{-41}$ assuming a binomial distribution with $p = 1/m = 8 \times 10^{-23}$. Hence, a random selection of subsets can be justified.

There are some questions which need to be resolved regarding the REM strategy. These include: How big sub-samples? What is the optimal area for the predictor field? Which are the best predictor variables? Furthermore, *Feldstein* (2000) has argued that monthly time-scales may not be optimal for detecting teleconnections, and empirical models may be improved by using for instance 10-day or 45-day

means as opposed to monthly mean. The answer to these questions may depend on the season as well as the location. A systematic exploration of the various settings is necessary for answering these questions.

One advantage of the ensemble approach is that it may provide a remedy for certain types of non-stationary and nonlinear relationships (e.g. brings in an aspect of so-called local regression). Bi-modal distributions and clustering of the data (e.g. weather regimes) will become apparent in the REM histograms.

A commonly used measure of skill for probability type forecasts is the *Brier score* (BS) (*Wilks*, 1995, p. 259) which is a mean-squared type measure of probability forecasts where o_k is 1 if an event occurs and $o_k = 0$ if it doesn't occur. We let $\hat{Y}_k$ denote the probability (for instance taken from an e.d.f. /c.d.f. distribution based on past error statistics) forecast. A perfect forecast gives the score BS= 0, whereas for a worthless model BS= 1.

$$\text{BS} = \frac{1}{n} \sum_{k=1}^{n} (\hat{Y}_k - o_k)^2. \tag{6}$$

The scores are estimated from the ensemble mean of independent (cross-validation) forecasts over the 1900-2000 period, except for the Brier-score. The Brier-score is based on all the data, including the dependent ones, in order to derive the predicted probability distributions for each event.

Choice of predictor fields

The present REM month-to-seasonal forecast system is designed to take any relevant predictor, including (i) North Atlantic sea surface temperature (SST) anomalies, (ii) mixed fields of SST, SLP, and 2-meter temperature [T(2m)], and (iii) climate station records. It is important to assess the various predictor choices and different mixed-fields combinations, as well as predictor area, in order to obtain an optimal forecasting model.

0.2.4   The forecasting system configuration

Updating the data

As already stated, an operational forecasting system requires a continual updating of the predictor data. The most flexible approach is to match new data with the old, and the system thereby does not rely on one specific data source. The update may be done by projecting the latest data onto the principal components* (*Benestad*, 1999b) of the DNMI reconstructed data initially established for the predictability studies.

The predictor data (spatial fields used as input) can be updated before the end of the month by using 10-day or 5-day forecasts. The fields updated with *predictions* are more uncertain than if the fields merely consist of *analysis* data. A chrontab job may launch a new job periodically, and an R routine may be used to check whether it's the right time for starting a new forecast (e.g. the beginning of a month). A launch-script calls for an update of the data and starts the forecast at pre-determined times (Figure 0-10)

The month-to-seasonal forecasts will initially use monthly mean re-analysis products from NCEP as predictors. At a later stage, analysis data from ECMWF will also be utilized, but at present these data do not go sufficiently far back in time. The evaluation of the month-to-seasonal forecast products involves hindcasts for the period 1980-present. This time interval is substantially shorter than the evaluation period 1960-2000, implying stronger sampling fluctuations in the score estimates.

The DNMI-model is run manually, and does not produce forecasts at fixed dates. One reason for this is that the DNMI-model takes some of its input from the NOAA Climate Diagnostics Center (CDC) in USA[†]. Although this data is freely available, the date at which they are posted on the Internet may vary. The users of month-to-seasonal forecasts need reliable forecasts that are generated at agreed dates, and there is a need for an automatic fully operational system that does not rely on the presence of key people. The forecasting system should use data that always is available on the first day of each

---

*The principal components describe the temporal evolution of the weights for the spatial modes (EOFs).
[†]URL: http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml

Updated data fields of SST, SLP, T(2m)
from ECMWF.

Analysis & forecast (10–day & 5–day)
Analysis only (1st day of each month).

Start new
forecast? ▷

The REM forecasting
system.
  Merging of old & new data
  Calibration & prediction
  Analysis of errors

◀── Updated observations from
     met.no climate archive.

▶ Automatically generated web
  pages
▶ Data in ASCII format (tables).
▶ Graphics
▶ Analysis of previous forecasts.
  Prediction of forecast skill.

Figure 0-10.   A schematic illustrating how the operational system is designed. The system consists of several modules which can be easily swapped or replaced. This allows easy testing and upgrading or modification of the system is in principle simple. The forecast products may include graphics, tables, statistics or web pages.

month. These points motivated the development of a completely new system which aims to achieve higher predictive skills, be more flexible, and robust. The DNMI model was coded for Matlab, which is an expensive software package that requires a license for the platform running the system. The new system (REM-model), on the other hand, was developed for the free data processing language called R (a "GNU version of S-plus") running on a Linux platform, and does not require special licenses. The new system will henceforth be referred to as the met.no-system, whereas the older model will be referred to as the DNMI model.
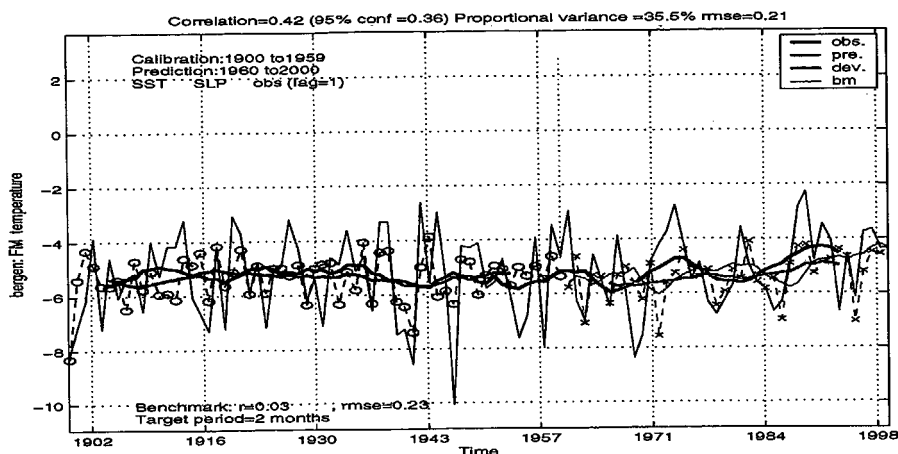
Figure 0-11. Hindcast for the February-March interval made during January (lag=1). The test-model. The black lines represent the observations (thick line shows 5-year low-passed curves), the blue lines show the predictions made for the calibration (dependent) period, and the red lines show the predictions made for an independent evaluation period. The green line shows the results from a benchmark model taking the mean of the previous 10 years anomalies.
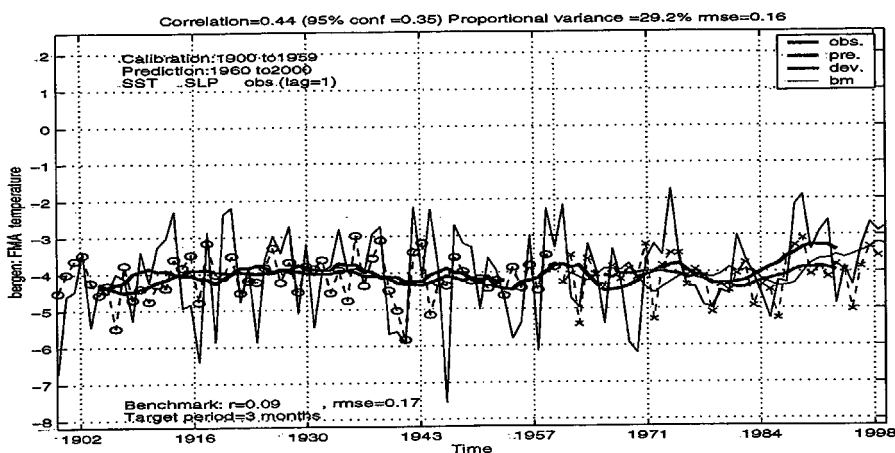
Figure 0-12. Same as Figure 0-11, but showing hindcast for the February-April season made during January (lag=1).

0.3 The results

0.3.1 Hindcasting based on the test-model

Temperature

A test period defined as 1960-2000 was used for the inter-comparison with a benchmark method based on the average of the previous 10 corresponding calendar months (i.e. the previous 10-year climatology). The evaluation used *anomalous correlation* (Pearson):

$$r(\hat{y}, y) = \frac{\sum_{t=1}^{N}(y_t - \overline{y})(\hat{y}_t - \overline{\hat{y}})}{\sqrt{\sum_{t=1}^{N}(y_t - \overline{y})^2 \times \sum_{t=1}^{N}(\hat{y}_t - \overline{\hat{y}})^2}} \qquad (7)$$

and *rmse*:

$$\text{rmse}(\hat{y}, y) = \frac{\sqrt{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}}{N} \qquad (8)$$

scores as a measure of skill. The former is not sensitive to the mean value (offset) of the predictions whereas the latter score is. Since the predictions give the variations about a mean value, a mean value must be added to these results in order to get a useful prediction. In these tests, the mean value for the 30 years preceding the time when the hindcast was made was added to the predictions.

The scores for various models are given in tables 10–19 (in the Appendix) for lags (lead time) of 0–2 months. The tests based on the reconstructed SST, SLP, and the monthly mean station temperature hindcast suggest modest skill during during winter and summer. These results are nevertheless not entirely representative of the maximum achievable scores, partly due to the above-mentioned shortcomings of the SST and SLP reconstruction, but also because these tests only involved SST and SLP and there may be precursor signals in deep ocean conditions and heat transport, snow and ice cover, and soil wetness. Hence, there may be predictors not included that may increase the skill of these models.

Table 10 gives the scores of monthly mean temperature hindcasts for Bergen as well as corresponding scores from the benchmark model (the average of the previous 10 years anomalies). Both anomalous correlation scores and rmse scores are presented. The entries shown in **bold face** have a better score than the benchmark method. It is nevertheless important to keep in mind that the score estimates are associated with some statistical fluctuations. Furthermore, even though the regression model yields better scores than the benchmark model, they may both be useless if both are low (e.g. correlation score less than 0.40). The zero-lag (contemporary) scores are shown as a check reference for the monthly mean values, but these scores also reflect the models' predictive skill for 2-monthly and seasonal (3-monthly) mean values. Hence, a zero-lag prediction for seasonal mean temperature involves the prediction for two months ahead when the present month's temperature is known. It is for instance necessary to subtract $0.33\times$ January value from this seasonal mean and then compute the correlation with the 2-monthly mean February-March (FM) in order to assess the real predictability using zero-lag for more than one month.

The analysis of the monthly mean hindcasts gives encouraging results for the empirical monthly prediction method. At one month's lead time (lag=1), the model scores higher than the benchmark model for 9 of the calendar months in terms of the anomalous correlation score and for 5 months in terms of the rmse score. The fewer high rmse scores obtained by the test-model reflect the difference between the two methods: whereas the benchmark method is a conservative model favouring near-normal values, the regression model predicts larger (more realistic) amplitudes.

The correlation score of 0.50 for 1-month lead time monthly mean temperature hindcast for Bergen starting in January is high and statistically significant at the 5% level, reflecting real predictability. These findings are also in line with earlier analysis based on different data (*Benestad*, 1999a), and seem to be a robust feature.

In Table 11 the high correlation scores ($r = 0.90$ in January-February) associated with the zero-lag prediction of January-February (JF) issued in January *may* suggest that the February mean temperature is highly predictive once the January value is known. The correlation score of 0.42 for one month's lead time is also substantially higher than the corresponding benchmark value and may furthermore be statistically significant at the 5% level (the 5% confidence limit ≈ 0.40). These results suggest a real predictive skill for the two subsequent months.

The scores for seasonal (3-monthly) mean predictions are presented in Table 12, indicating potential for empirical seasonal prediction of the Bergen temperature during winter. The winter-time scores are both substantially higher than the benchmark method and just statistically significant at the 5% level.

Hindcasts based on a number of different model settings were evaluated. Tables 13 shows hindcasts scores for monthly mean, 2-monthly mean, seasonal mean temperature in Bergen when sunspot data (monthly mean sunspot number and 6-month low-passed sunspot value) were included as a predictor (there is a long history for using sunspots for climate prediction (?)). In some cases the scores are improved and in other cases deteriorated. Parts, if not all, of the variations in the prediction scores can be accounted for by sampling fluctuations. Thus the sunspot data does apparently not make a systematic improvement to the forecast capability. Much of the precursor signal can be seen in the station data itself.

Table 14 shows the hindcast scores from models only utilizing the Bergen monthly mean temperature. The autocorrelation can account for most of the predictive skill during winter, and the 1-month lead for February using just observations is associated with better skill than for the models also including SST and SLP. The reason why the "standard" (using SST, SLP, and contemporary observation
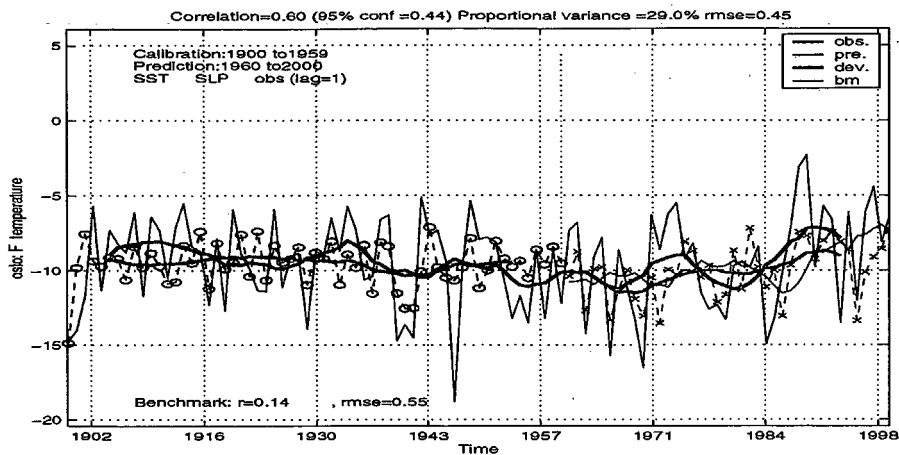
Figure 0-13. Evaluation of the prediction of February mean temperature in Oslo.

as input/predictor) regression model yields lower scores than the model taking only the contemporary monthly mean station observation is that some of the precursor signal in the station data is swamped by "noise" in the SST and SLP (i.e. the method for extracting the information is not perfect as some information tends to "leak" out). The models only trained with SLP and SST (Table 15) can also improve on the results from the benchmark model, although it is difficult to resolve the predictability due persistence from what is "caused" by the SST and SLP anomalies (high scores at zero-lag suggest that there is a relationship between the contemporary large-scale SST and SLP anomalies and the station observations).

Other locations

Analyses of hindcasted temperature for Oslo, Stockholm, Copenhagen, and Helsinki give qualitatively similar results as the above analysis for Bergen. The one-month-lead hindcast results for Oslo February temperature, based on SST, SLP, and past station values, are show in Figure 0-13. The correlation score for the independent evaluation period (1960–2000) is 0.60. The correlation score for the February mean temperature in Oslo is lower than for February–April. Similar analysis for two-month targets suggest that the models skills are not limited to monthly or seasonal mean values (Figures 0-14 and 0-15). In other words, the skill-estimates seem to be robust. Figure 0-15 suggests that two-month spring-time temperatures in Helsinki also may have some predictability.

Table 16 lists the hindcast scores for the seasonal mean temperature in Oslo, suggesting good potential for seasonal prediction during winter (lag-1 anomalous correlation of 0.74). The scores for seasonal forecasts at other locations are as high as 0.6-0.7 for one month's lead time in winter. The models fail to give skillful predictions for the summer temperature (e.g. Figure 0-16).

At 2 months' lead time, the scores are generally bad, and such forecasts are probably not useful. Nevertheless, a month-to-seasonal forecast at 1 month's lead time represents a prediction for the next 2-4 months.
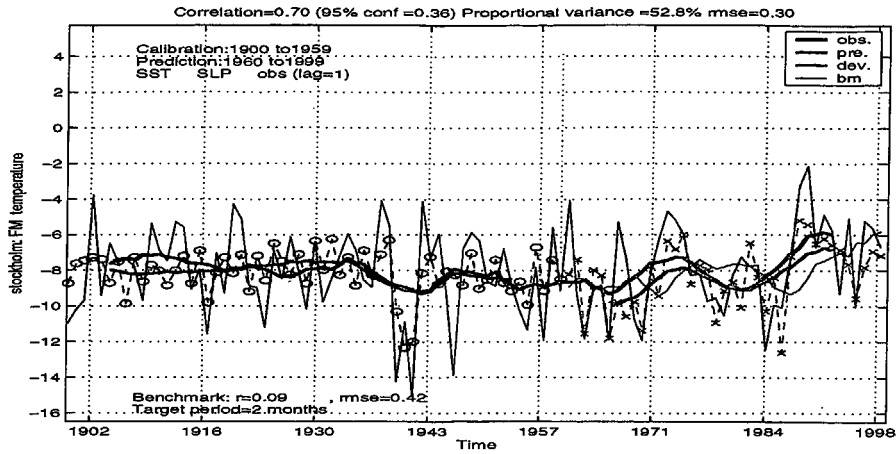
Figure 0-14.   Evaluation of the prediction of February-March mean temperature in Stockholm.
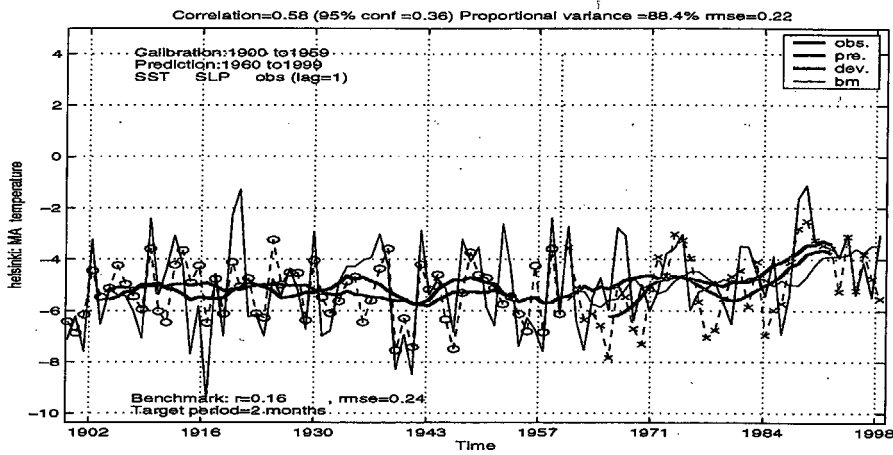
Figure 0-15.   Evaluation of the prediction of March-April mean temperature in Helsinki.
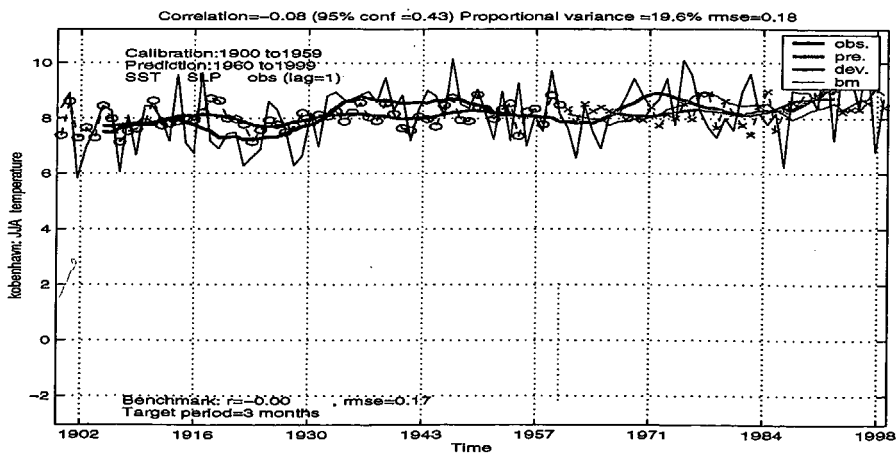
Figure 0-16.   Evaluation of the prediction of May-July mean temperature in Copenhagen.

R.E. Benestad

Correlation=0.46 (95% conf =0.36) Proportional variance =7.3% rmse=17.16
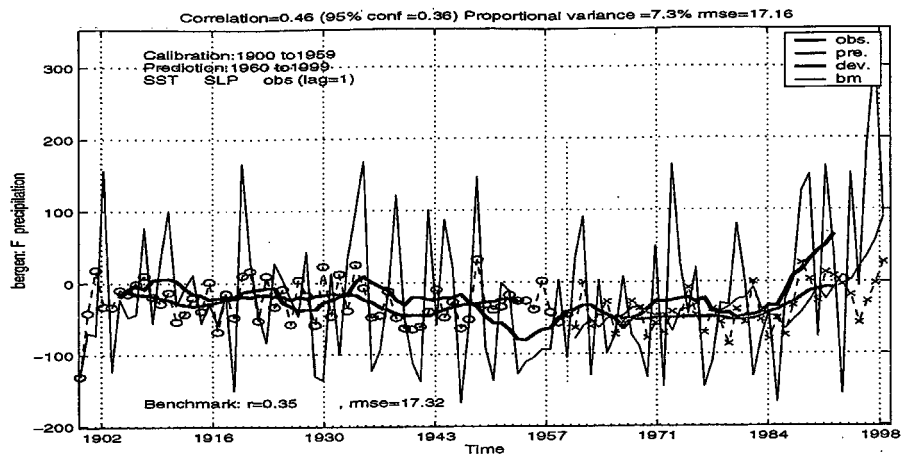
Figure 0-17. Hindcast for the Bergen February mean precipitation made during January (lag=1).

Correlation=0.52 (95% conf =0.36) Proportional variance =10.0% rmse=10.36
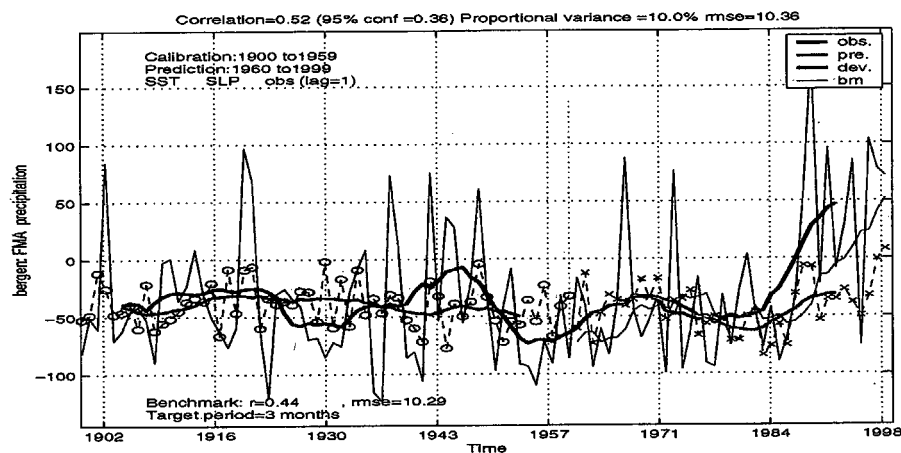
Figure 0-18. Precipitation hindcast for the February-April interval made during January (lag=1).

Precipitation

The monthly mean hindcasts for the precipitation in Bergen suggest that the test-model in some cases has skill better than the benchmark model and higher than the 5% confidence level. The model does, however, not predict the recent upward trend in the winter time precipitation (Figure 0-17). The prediction of seasonal mean values gives higher correlation score than the monthly mean amount(Figure 0-18), but also the for seasonal mean values do the models miss the recent trend. Table 20 (in the appendix) gives an evaluation of the hindcast for seasonal mean precipitation in Bergen (Figure 0-17), indicating best prospects for predictability in January. One month's lead time correlation score of 0.52 is significant at the 5% significance level, but it is interesting to note that also the benchmark method yields high scores in January. Much of this precipitation is related to the NAO (*Benestad & Tveito*, 2002). The results suggest virtually no skill for the late spring, summer and autumn. As the precipitation is much more localized than temperature anomalies, it is expected to see greater differences in the scores from location to location. Table 21 presents the corresponding analysis for Oslo. The prospects for lag-1 seasonal prediction of precipitation for Oslo appear to be greatest during summer and autumn (Table 21, Figure 0-19). Although the regression method achieves better scores than the benchmark method, both tend to be low and the usefulness of these predictions may be questioned. Also in Stockholm (Table 22) do the spring- and summer-time rainfall predictions obtain better scores than the benchmark model, albeit much of these differences may be due to sampling fluctuations. The predictability for seasonal mean rainfall over Copenhagen looks bleak according to the results in Table 23, and the situation is not much better for Helsinki (Table 24). All this means that North Atlantic SST and SLP are not suitable as
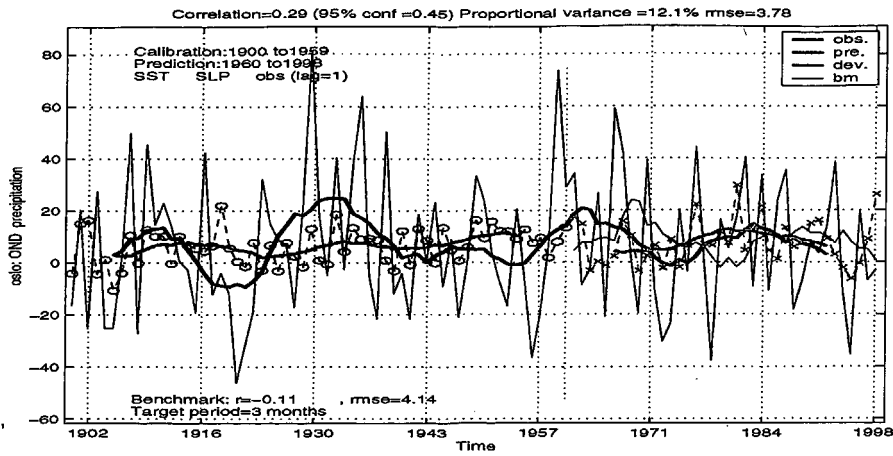
Figure 0-19. Hindcast for the Oslo October-December precipitation made during September (lag=1).

predictors for seasonal mean precipitation for these locations, nor is there much autocorrelation in the precipitation series apart from in Bergen during winter. The apparent moderate scores for Oslo (Table 21) may be real, but there is no guarantee that these are not coincidental (see section (0.5)).
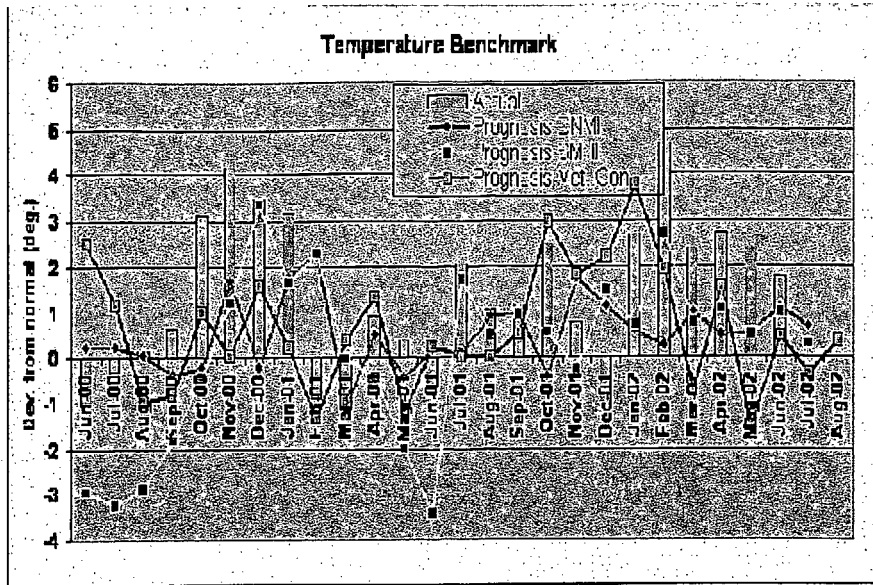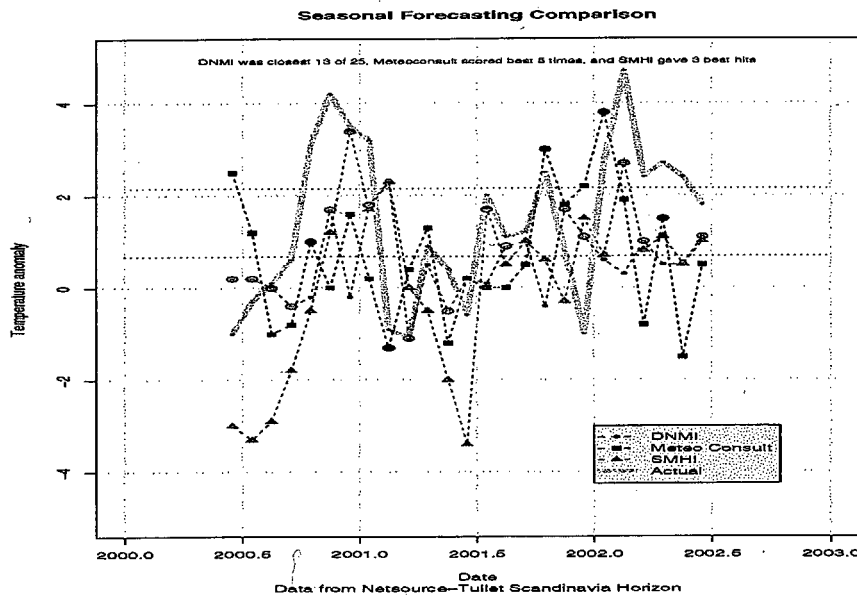
Figure 0-20.   (a) The model comparison obtained from NTS Horizon web site for weighted temperature over Norway, Sweden and Finland (The weighting is done by NTS). (b) A reproduction of the forecast also showing the best forecast (black circle).

0.4   The Forecasts

0.4.1   The DNMI-model

Evaluation of skill

The prototype month-to-seasonal forecasting system was developed in a joint research project together with Natsource-Tullett Scandinavia* (NTS, former CBF) as a part of their Horizon project. The forecasts from this model have been evaluated by NTS and are shown in Figure 0-20(a). Of 25 cases (June 2000–July 2002), the DNMI[†] model was closest to the true value for 13 of the 25 cases, whereas MeteoConsult

---

*URL: http://www.cbf.no/.
[†]"Det Norske Meteorologiske Institutt": The Norwegian Meteorologival Institute.
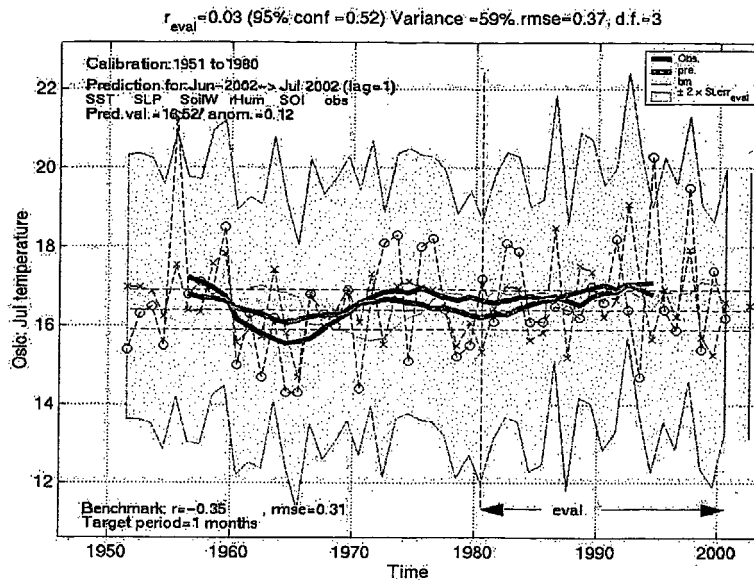
Figure 0-21. Example of a forecast made with the DNMI model. The forecast was based on June 2002 observations and analysis and the target month was July 2002 (lead time of 1 month). Monthly mean temperature.

TABLE 3.   The actual forecasted temperature ($\hat{T}$) and precipitation ($\hat{rr}$) anomalies for Oslo-Blindern based on the DNMI-model. The Pearson correlation between the observed values and the predictions is 0.01 for temperature and -0.17 for rainfall (the p-values are 0.98 and 0.57 respectively, indicating that these are not statistically significant from zero). It is important to keep in mind that the number of forecasts is small and this preliminary evaluation therefore is very uncertain.

Year	Month	Obs T	$\hat{T}$	Obs rr	$\hat{rr}$	Date of fcst
2001	6	-0.6	1.02	45.0	20.67	22-Jun-2001
2001	7	1.3	0.39	82.3	13.83	18-Jul-2001
2001	8	0.6	1.84	92.8	-4.95	07-Aug-2001
2001	9	0.7	-0.62	66.9	-40.63	11-Sep-2001
2001	10	2.6	-0.90	149.3	-8.83	17-Oct-2001
2001	11	1.4	1.70	37.3	35.51	08-Nov-2001
2001	12	-0.3	-0.78	31.2	14.29	12-Dec-2001
2002	1	1.5	0.55	75.6	31.00	08-Jan-2002
2002	2	4.6	1.10	57.9	3.40	08-Feb-2002
2002	3	2.1	NA	32.4	NA	NA
2002	4	2.4	-0.57	40.4	-1.83	12-Apr-2002
2002	5	2.1	0.32	99.2	5.01	06-May-2002
2002	6	1.4	0.65	60.5	-19.10	17-Jun-2002
2002	7	1.0	0.12	133.0	15.37	10-Jul-2002

was best in 5 of the cases and SMHI in 3. Four of the cases were "drawn" between the DNMI model and one of the other models (Figure 0-20(b)). Of the 25 DNMI forecast, 14 had the right sign whereas 12 Meteo Consult and 16 of the SMHI forecast had the correct polarity. The Pearson correlation between the DNMI forecast is 0.11 (p-value = 0.60), compared to 0.16 for Meteo Consult (p-value = 0.45) and 0.58 for SMHI (p-value = 0.00). The root-mean-squared error (rmse) is 1.81 for DNMI, 2.03 for Meteo Consult and 1.66 for SMHI. Another way of assessing the forecast is to classify the value in a given category, say low, normal and high (each with an equal chance of being populated, the values shown as dotted horizontal lines in Figure 0-21(a)).

The DNMI model did not predict any of the large anomalies. Thus, although the DNMI model compares well with the two other models, there may be room for improving the forecast skill.

Table 3 shows an evaluation of the actual forecasts made for Oslo-Blindern in the period June 2001 – July 2002. A correlation analysis between the predicted and observed anomalies suggests that the DNMI-model appears to be only marginally better than using the climatology as a prediction*.

---

*The evaluation interval is too short to make a strong statement about this.

Additional information derived from the analysis

There are many ways in which month-to-seasonal forecasts can be presented. A common graphical format is a time series plot showing the past and present predictions together with the "truth", thus facilitating a visual evaluation of the past predictions. Figure 0-21 shows an example of such a presentation, where the yellow area denotes a measure of uncertainty ($\pm 2 \times$ standard error of the evaluation period), black line with circles denotes the actual monthly mean temperature, and the red line with crosses represents the predictions. Low-passed values are also shown as thick curves, showing the tendency for the model of over- or underestimating temperature. The thin green line shows the benchmark predictions. The 1961–1990 climatological mean temperature is marked by the middle dashed horizontal line as are the levels ($\pm 0.43\sigma$) dividing the observed distribution into 3 equal areas (upper and lower horizontal dashed line) so that the theoretical probability[†] of the observed value being in either of the three categories high, normal or low is 0.33. The most recent forecast is shown as a dark red "x" together with error bars.

A different format is the probability distribution plots shown in Figure 0-22. Fig 0-22(a) shows a histogram of the past prediction errors from an independent evaluation period centered around the most recent prediction together with a best-fit Gaussian curve. Because of the short predictor data records, the population of error estimates is small, and the histogram gives only a sketchy impression of a Gaussian shape at best. The errors also tend to be much greater than the size of the categories "cold", "normal", and "warm". The x-axis can therefore be categorized in terms of these dashed vertical lines as "low", "normal", and "high" values. Panel (b) shows the corresponding e.d.f. and c.d.f., from which the probability of getting low, normal, and high values can be read directly. Moreover, due to the small error sample, the distributions only give a crude and uncertain measure of probabilities. The Brier score for both the regression and the benchmark method are shown in Fig.0-22(b).

One way of gauging the real predictive skill of a forecast model is to look at the statistics telling how often a severe event took place and how often the model predicted such an anomaly. Figure 0-23 show hit-ratio tables which shows how often high temperature occurred in Bergen (upper row) and how often high temperatures were predicted (left column). The blue numbers along the diagonal (lower-left-to-upper-right) give the number of correct forecasts, whereas those boxes off this diagonal represent misses of various degrees (red numbers show a total miss). It is apparent that the benchmark method is a very "conservative" model giving more "normal value" predictions, and is not a very good model despite obtaining high anomalous correlation and rmse scores.

The probability of making a correct forecast can be estimated according to:

$$P_{\text{hit}} = \frac{N(y_{\text{low}}, \hat{y}_{\text{low}}) + N(y_{\text{norm}}, \hat{y}_{\text{norm}}) + N(y_{\text{hi}}, \hat{y}_{\text{hi}})}{N_{\text{total}}}. \tag{9}$$

The hit-ratio can be defined as the sum of correct predictions over the total number of forecasts (The sum of blue numbers in Figure 0-23). The hit-ratios computed for the DNMI model (for the whole of Scandinavia) is 0.48, Meteo Consult 0.40, and SMHI 0.44 (expectation value of a random guess is 0.33). It is possible to estimate confidence limits for this probability by carrying out Monte Carlo integrations using a large number of same length data series with random Gaussian numbers. Note that the probability scores will vary with the levels used to divide the observations into the categories [low, normal, high]: too high levels (i.e. the "normal" bin gets too wide) will lead to inflated score (and not very useful forecast) whereas too low levels will make the scores more sensitive to misses of near-normal conditions. The probability scores and corresponding 5% confidence levels are shown for the hindcasts in Fig. 0-23 both for the regression model and the benchmark.

Analysis of the range of model coefficients derived from the cross-validation analysis can also give some indication of the robustness and uncertainty associated with the predictions. Figure 0-24 shows the model coefficients together with error bars indicating the spread in the various estimates. When there is a clear and strong stationary relationship between x_t and y_{t+l}, then there will be small error bars. A high level of noise and weak and uncertain relationships will tend to give greater scatter in the coefficient estimates and hence larger error bars. In the example shown in Figure 0-24, the coefficients are well defined.

One interesting observation that can be made from Figure 0-20 is that the anomalies are biased

---

[†]Assuming a normal distribution.
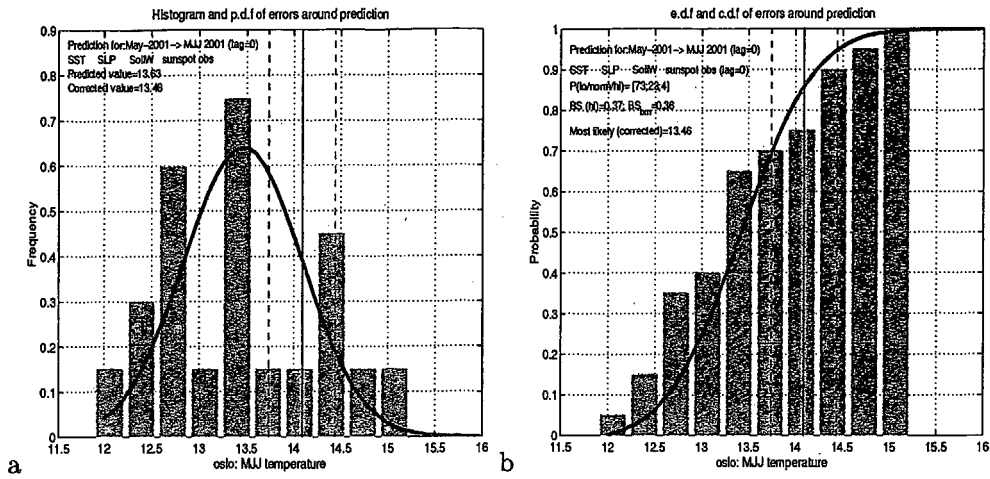
Figure 0-22. A monthly temperature forecast shown as probability distributions based on past errors: shown as histogram / p.d.f. and e.d.f. / c.f.d. The blue vertical gives the 1961–1990 climatological value and the dashed lines mark divide the three categories "cold", "normal", and "warm".
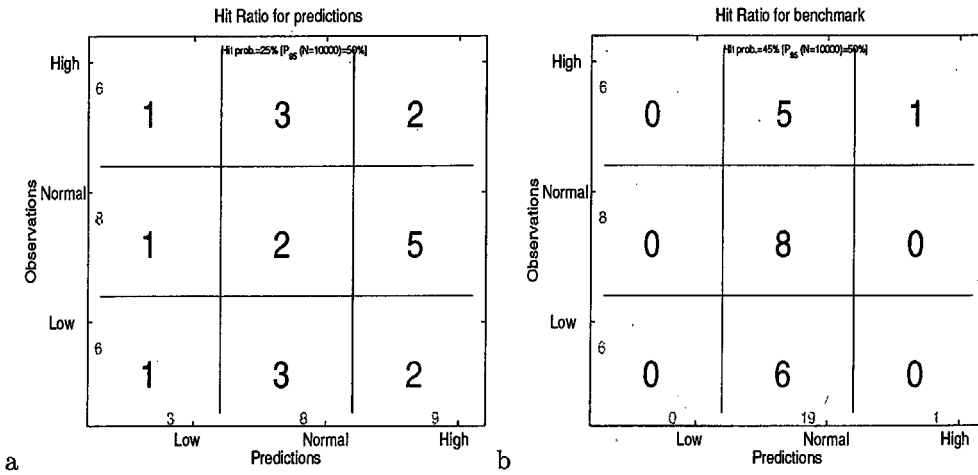
Figure 0-23. Hit ratio statistics for Oslo May–July mean temperature forecasts at one-month's lead.
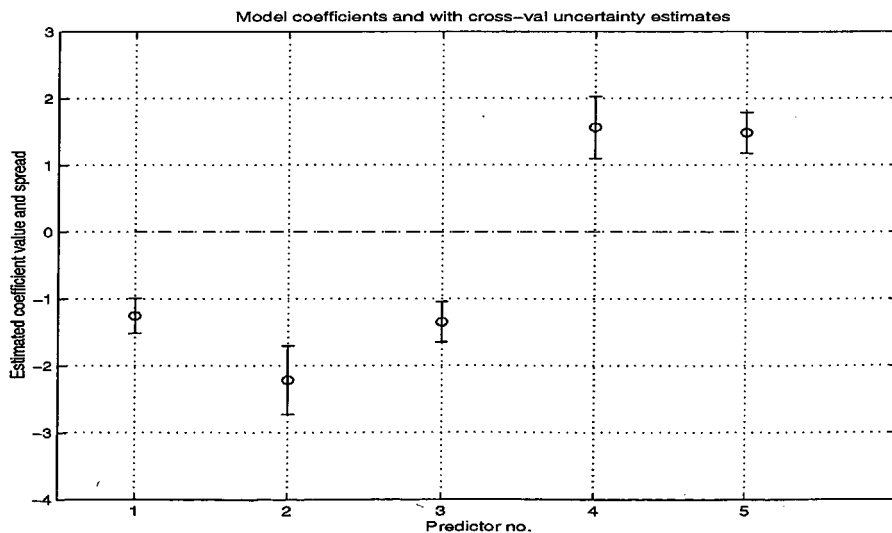
Figure 0-24. Best-estimate model coefficients from cross-validation analysis shown with a uncertainty measure based on the cross-validation standard deviation.

TABLE 4.   The climate station predictors covering 1900–1980 used for calibration and cross-validation in the examples
shown.

Bjørnholt	T(2m)
Dunderlandsdalen	T(2m)
Geilo	T(2m)
Halden	T(2m)
Namdalseid	T(2m)
Oksøy	T(2m)
Oksøy	precipitation
Oksøy	snow-depth
Skjåk	T(2m)
Sulitjelma	T(2m)

TABLE 5.   Locations where station records are used as predictors. (N), (DK), (FIN), (F), (S) refer to Norway, Denmark,
Finland, Faeroes, and Sweden respectively.

Kjøremsgrendi (N)	Oslo-Blindern (N)	Nesbyen (N)
Bergen-Florida (N)	Værnes/Trondheim (N)	Tromsø (N)
Karasjok (N)	Vardø (N)	København (DK)
Helsinki (FIN)	Turku (FIN)	Tampere (FIN)
Jyväskylä(FIN)	Kuopio (FIN)	Sodankylä(FIN)
Torshavn (F)	Göteborg (S)	Karlstad (S)
Uppsala (S)	Stockholm (S)	Stensele (S)
Abisko (S)		

towards positive values (min=-1.100, median=1.200, mean=1.416, and max=4.700). The reason for this shift is the long-term changes in the temperature associated with decadal variations and long-term warming trends.

0.4.2   The REM-model

The REM model is a kind of "brute force"-solution to the forecasting problem for which there are errors in the predictor data. This forecasting strategy makes good use of the resources that modern computers can offer, and a complete forecast on a double 800 MHz CPU machine running Linux takes several hours, depending on the ensemble size, number of locations, and number of lags. The ensemble size, size of the subset used for model calibration, lags, type of predictor, and locations are set at the beginning of each forecast. The predictor data is pre-processed before being used in the REM model, and any type of data can be used as inputs provided they are of correct format. The forecasting system is set up so that it may start overnight and use computing resources when the computer otherwise is idle. The model can be described as "generic" and can be applied to any time series. This means that the model can easily be set up for the prediction of precipitation or other parameters as well as temperature, and it may used for any geographical region given sensible input data.

Figure 0-25 shows some preliminary results obtained with the REM-model using climate station values as predictor. The advantage of using climate station data (Table 4) is that only long records may be included and there is no need for merging different data sets. The climate station data is also available soon after the end of each month. The disadvantage with using station data only is that pre-cursory signals may be weaker (for instance more local "noise") or even absent in the local station records. Hence, these forecasts may not give the correct impression about the maximum achievable skill of the REM-model.

Figure 0-25 shows a cross-validation analysis of 100 one-month-lag predictions for Oslo. There is a tendency for the range of predicted values for each month to vary with time. The correlation between the cross-validation analysis and the observed temperature record is 0.30 for the period 1900–1999 (however, there are gaps of missing data). For a model producing random numbers, the range would be expected to be constant and the correlation is expected to be zero. The REM-model is not perfect and predicts a number of high temperatures in the 1910s and during the late 1990s when the observations suggest low values. Figure 0-26(a) shows the distribution of the ensemble of forecasts by the REM model (blue) and the errors associated with the ensemble mean values. The hit-ration (Figure 0-26(b)) for Oslo of 0.26 is higher than what is expected for a model producing random values.

Cumulated independent hindcasts
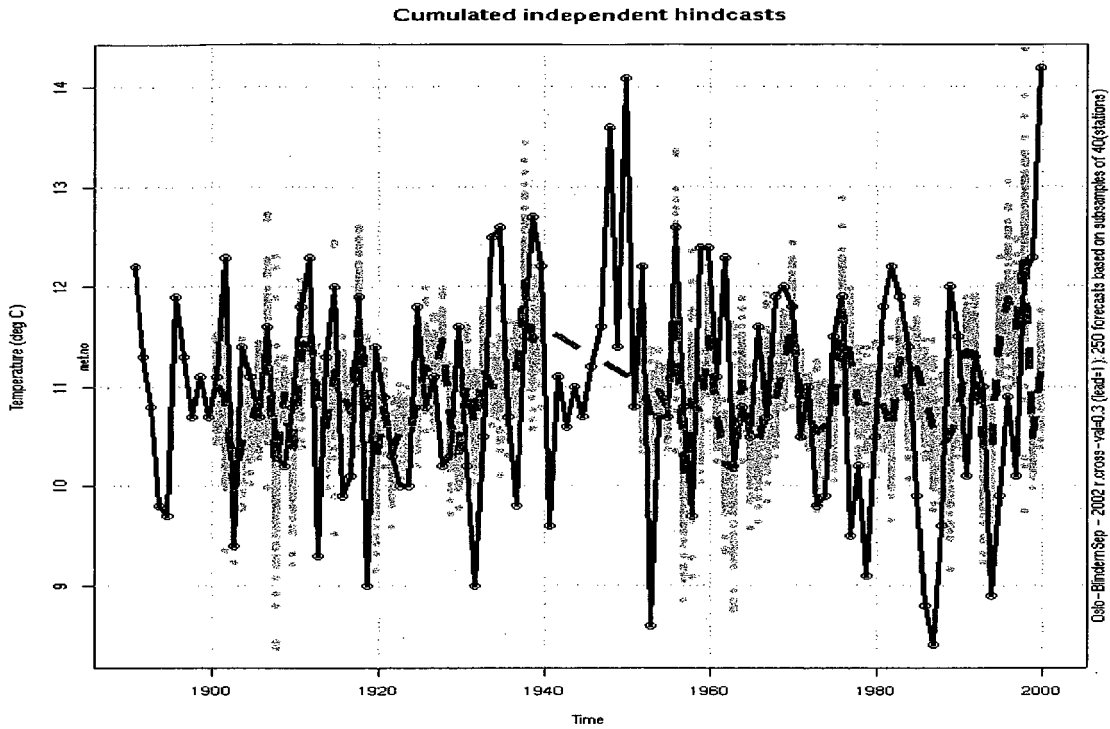
Figure 0-25. Example of output from the REM-model using climate station data as input. The cross-validation results from an ensemble of 250 forecasts.

TABLE 6. Analysis of variance (ANOVA) of the results from a regressional analysis of forecasted temperature anomaly against geographical parameters. The columns give the estimated coefficients, error, t-value and probability associated with the regression against distance to the coast, altitude, latitude and longitude. The residual standard error is 0.2717 on 17 degrees of freedom. The estimated multiple R^2 is 0.6386, and an adjusted R^2 0.5536. The f-statistic is 7.511 on 4 and 17 DF, and the probability that the null-hypothesis of no geographical dependence (p-value) is 0.001123 (Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).

	Estimate	Std. Error	t-value	$\Pr(> |t|)$
(Intercept)	5.2997211	1.0606950	4.996	0.000110***
dist	-0.0842264	0.0745746	-1.129	0.274404
alt	-0.0002072	0.0002687	-0.771	0.451374
lat	-0.0680486	0.0191566	-3.552	0.002450**
lon	-0.0020978	0.0123226	-0.170	0.866834

Evaluation of skill

The box-plots in Figure 0-27 show how the correlation scores vary with lag. There is a rapid drop from zero to one-month lag, as expected. The one-month lag generally has higher correlation scores than the 2-month lag, and this suggests that there is indeed some predictive skill at lags of one month.

0.4.3 Spatial analysis of the forecasts

It is possible to use the forecasts for a number of stations in Scandinavia to infer change for other locations in the same vicinity if the forecasted values follow geographical patterns. Table 6 gives a summary of a regression of predicted temperature anomaly at various locations (dependent variable) against geographical parameters. The forecasts exhibit a clear latitudinal dependence with lower values at higher latitudes.

The results from the REM model can be used in a multiple regression against geographical parameters such as the distance from the coast, altitude, latitude, and longitude. More complex models may also include the gradient of the topography, but if the number of locations is small, then the inclusion of too many independent variables is likely to result in an over-fit. Figure 0-28 shows a map constructed from a

Random Ensemble Model forecast distribution

Error distrib.

Probability

0.8

0.6

0.4

0.2

0.0

10   11   12   13   14   15   16   17

Temperature anomaly (C)
Oslo-Blindern Sep - 2002 .Normal= 11.29

Oslo-Blindern Sep - 2002 r.cross-val=0.3 (lead=1) 250 forecasts based on subsamples of 40 stations)

a

Hit-Miss table: Temperature

	Hot	Warm	Normal	Cold	Freezing
Freezing	0	4	9	8	1
Cold	0	5	5	4	1
Normal	0	4	12	0	1
Warm	0	6	8	3	0
Hot	0	7	9	2	0

Predicted met.no

Oslo-Blindern Sep - 2002 (lead=1)x 250 forecasts based on subsamples of 40 (stations)

Observed
[p-hit= 0.26 , expected p-val= 0.2]

b

Figure 0-26. Example of output from the REM-model using climate station data as input. (a) histogram of all forecasts in Figure 0-25 (blue) providing a basis for a probability estimates and the error distribution (grey), and (c) the hit-ratio-analysis based on the 250-forecast-ensemble mean.
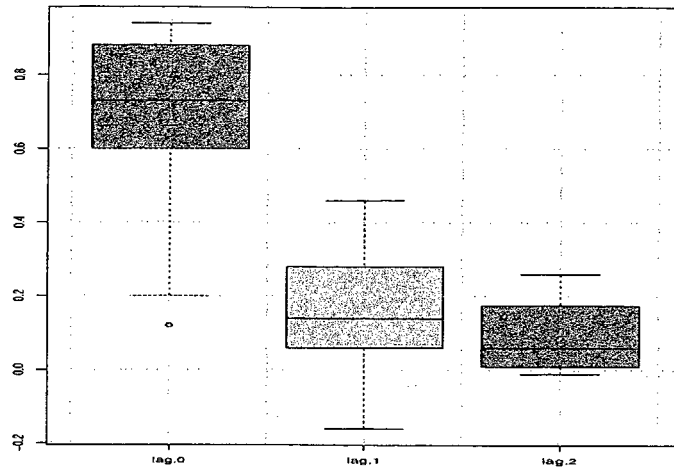
Figure 0-27.   Preliminary correlation scores from the REM-model using station values as both input and output (Tables 4 and 5). The labels on the x-axis "lag.0", "lag.1", etc refer to one-month and two-months lagged predictions.

multiple regression with only 22 different stations (Table 5) using 4 geographical parameters (Table 6). The analysis shown here may include too few locations and the prediction from the regression analysis may be over-fit.

The residual from this regression can be spatially interpolated through the means of a kriging analysis if the residuals contain a spatial structure, and spatial maps can be constructed by adding the multiple regression results and the interpolations from the kriging analysis. The map shown in Figure 0-28 did not include kriging since the residuals contained no spatial structure.

The development of the REM model has only involved temperature forecasts, but future work will also include precipitation.
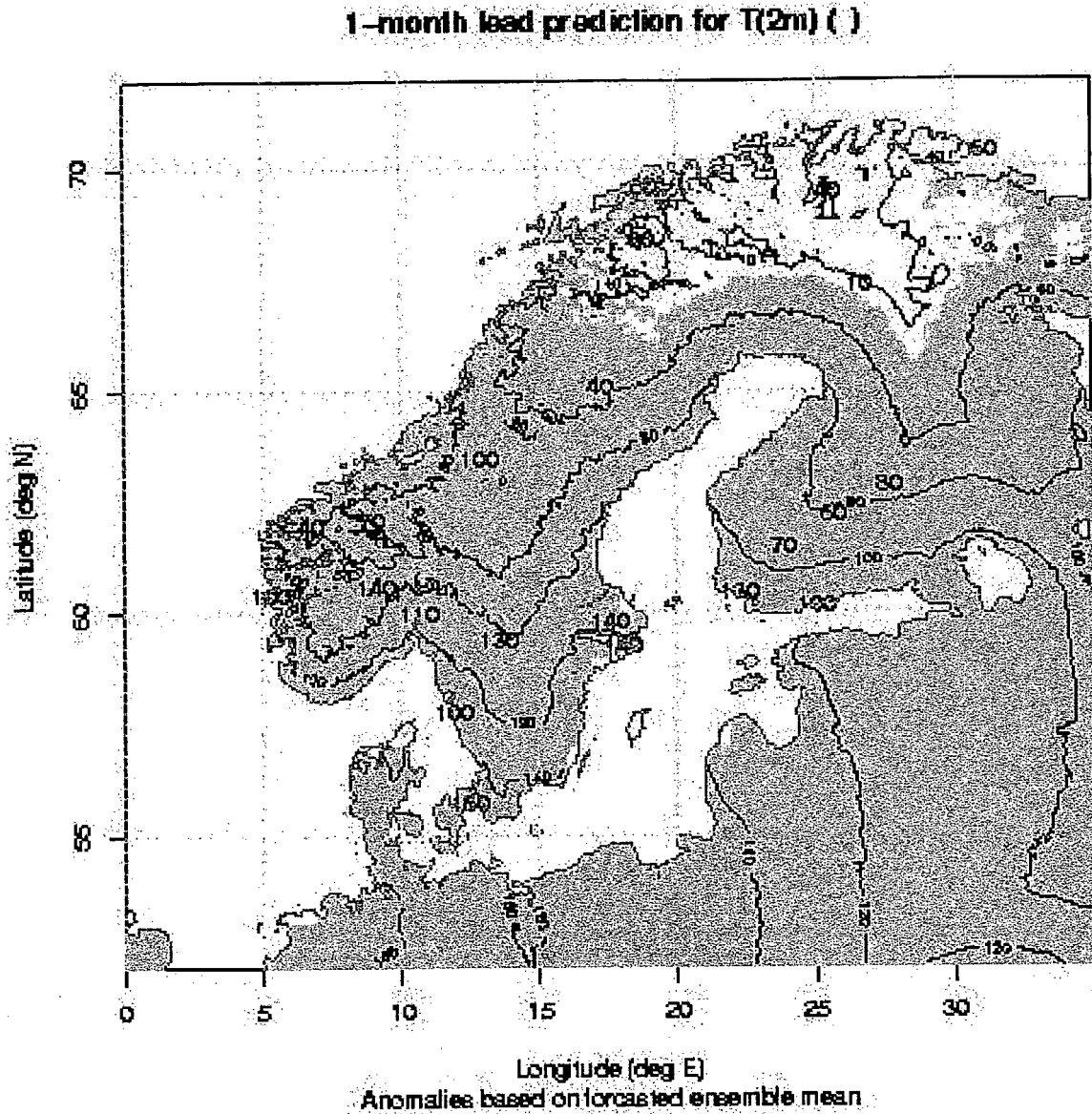
Figure 0-28. An example of how geographical information can be included in the analysis to make spatial maps of forecasts. This figure shows the one-month lead forecast for September 2002, and is derived from a multiple regression with 22 locations (Table 5) and 4 independent variables. The temperature anomalies have bee n multiplied by 100 in this figure.

TABLE 7.   Probabilities of getting n cases with higher score than the benchmark if the probability of obtaining a better score for each case is 50% (both methods are equally good) according to a binomial distribution.

n=1	n=2	n=3	n=4	n=5	n=6
0.3	1.6	5.4	12.1	19.3	22.6

n=7	n=8	n=9	n=10	n=11	n=12
19.3	12.1	5.4	1.6	0.3	0.0

TABLE 8.   Counts (N) of superior (in terms of the anomalous correlation score) seasonal mean temperature forecasts based on SST, SLP, and past observations, and estimated significance (p in %) [N / p].

location	lag=0	lag=1	lag=2
Bergen	11 / 0.3	2 / 1.6	0 / 0
Oslo	12 / 0.0	2 / 1.6	0 / 0
Stockholm	12 / 0.0	3 / 5.4	2 / 1.6
Copenhagen	12 / 0.0	2 / 1.6	2 / 1.6
Helsinki	11 / 0.3	3 / 5.4	2 / 1.6

0.5   Discussion and conclusion

One of the greatest obstacles to month-to-seasonal forecasting is the lack of long high-quality data records for the calibration of empirical models and for validation of empirical and dynamical models. Temperatures and SLP have traditionally been used as predictors since there are long historical record of these parametres. It is reasonable to believe that data on oceanic heat transport, heat content, snow and sea-ice extent may contain information that can be used to infer near-future weather statistics, but records of these types of data tend to be too short for empirical modelling. One interesting question is whether this obstacle can be bypassed by the use of e.g. an ocean model forced with best available wind and temperature data to reconstruct the oceanic heat transport and heat content. Empirical models could be tested using vertical cross-sections of the ocean as predictor data. Ocean models may be driven with up-to-date atmospheric forcings, or past reconstructions may be merged with remote sensing data if there is a good agreement between the simulations and the observations. Altimeter data, sea-ice, and snow-cover from remote sensing platforms have, if at all, rarely been used in empirically based seasonal forecasting, but it may be possible to utilise this data if they can be merged with assimilated ocean model data in order to construct sufficiently long data records. In other words, a combination of using dynamical and empirical models may potentially lead to improved month-to-seasonal forecasting.

The interpretation of the forecasting scores must be done with caution if many different locations are considered. The chance of obtaining n higher scores out of N can be estimated for the null-hypothesis that the forecast models are equally skillful as a benchmark model. The likelihood that one of equally good models scores higher for an individual case is $p = 0.5$, but if the comparison is made for more than one case, the likelihood must be estimated assuming a binomial distribution. The scores shown in the Tables 10–24 are computed for the 12 different calendar months, so that the value for N can be taken as 12. It is assumed for simplicity that there is no seasonal dependency in model skill.

TABLE 9.   Counts (N) of superior (in terms of the anomalous correlation score) seasonal mean precipitation forecasts based on SST, SLP, and past observations, and estimated significance (p in %) [N / p].

location	lag=0	lag=1	lag=2
Bergen	11 / 0.3	5 / 19.3	5 / 19.3
Oslo	12 / 0.0	10 / 1.6	8 / 12.1
Stockholm	8 / 12.1	7 / 19.3	5 / 19.3
Copenhagen	11 / 0.3	7 / 19.3	8 / 12.1
Helsinki	11 / 0.3	8 / 12.1	7 / 19.3

$$P_r(X = n) = \begin{pmatrix} N \\ \\ n \end{pmatrix} p^n(1-p)^{N-n} \tag{10}$$

The probability of obtaining n cases with higher score is listed in Table 7 and Tables 8 and 9 present statistics showing how many cases (out of 12) there are where the forecast model obtained higher anomalous correlation score than the benchmark model as well as the probability of seeing the same number of cases if the two models were equally good. For the monthly mean temperature the results suggest that the model outperforms the benchmark model for zero lag (temperature averaged over present month and the 2 following months, Table 8). The model is also generally better than the benchmark model for 1-month's lead time. For lags greater than one month, the benchmark model tends to be superior to the forecast model, suggesting either that the forecast model predictions are worse than a random guess at these lead times or that there is some low-frequency signal which makes the benchmark model better than a pure guess.

The high anomalous correlation scores in January and February at 1-months lead time are significant at the 5% level.

When it comes to precipitation (Table 9), the forecast model is significantly better than the benchmark model at all locations except for Stockholm at lag 0. For 1-month lead time, the forecast model is significantly better in Oslo than the benchmark model at the 2.7% level, but apart from this case, the two prediction methods appear to be equally good. Precipitation is much more localized, which can explain the regional differences in the correlation scores.

The prediction scores for precipitation at lead times greater than one month tend to be lower than 0.40 which is approximate the 5% confidence level for the correlation (estimated using a Monte Carlo technique). Because there are many independent score estimates, we must expect to see some cases exceeding this value (e.g. Bergen, Jan, Table 20) if the scores are randomly distributed. The benchmark model based on the previous 10-year climatology achieves high scores for precipitation in Bergen during winter. There is a clear difference between the benchmark scores of Bergen and locations such as Oslo, Stockholm, Copenhagen, and Helsinki.

These results merely reflect the skill of simple regression type models and are not necessarily representative of the highest practical predictability that is achievable. A combination of various models based on different methodology is likely to improve the skill further. Hence these results are an encouraging starting point for empirical forecasting for the temperature in northern Europe. For precipitation, on the other hand, forecasting according to these results appears to be more tricky. The prospect of forecasting rainfall amounts will rely on finding other predictors that have a real association with the precipitation.

0.6   Documented problems in the NCEP data

0.6.1   NCEP Reanalysis PSFC problem 1948-1967

Evaluation - Reruns for Jan, Jul 1953

By pure serendipitity, a problem has been uncovered with the encoding of surface and mean sea level pressure* that affect the period 1948-1967. The problem originated in the erroneous conversion of a portion of the TD13 dataset into the bufr ADPSFC type messages. Surface pressures in the past have suffured from a "Y2K" like problem in that values were stored with 3 digits ,pp.p, and a flag that indicated which hundreds units should be added (700,800,900,1000,etc). The discovery arose from a case study of the Reanalysis handling of the North Sea Gale of Jan 31:Feb 2, 1953 In addition to the forecast and verication grids, Suru made observation plots where the data rejected by the Reanl were plotted in red. The PSFC 03Z Feb 1, 1953 revealed the problem. Most of the obs within the 1000mb contour were rejected for being unphsically high pressure values. The increment plots revealed they were ~100 mb too large. If one mentally replaces the leading "10" with a "9", the values create a reasonable picture of the cyclone.

Once we saw that pattern, we told Jack Woollen what we had found, so that he could research the origin of these reports and the decoding procedures. Jack called back the next day and said he had found a tell-tail error in translating dataset TD13. Furthermore, it apeared that 4-6% of the surface obs in the set were affected.

In order to assess the geographical distribution over time, I have made two sequences maps. I had created grib files of the obs with surface pressures greater than 1050mb. I then used these in conjunction with the files Wesley has made available for plot and for download, and computed the percentages of obs <1050mb. This approach gives a sense of the location and volume of obs that were miscoded. I sampled all the January files for each year 1948-1968. Dataset TD13 ended in 1967. To see that the problem ended there see the file for Jan 1968. PLOTS of % January ADPSFC obs with psfc < 1050mb '40s Jan 1948 Jan 1949 '50s Jan 1950 Jan 1951 Jan 1952 Jan 1953 Jan 1954 Jan 1955 Jan 1956 Jan 1957 Jan 1958 Jan 1959 '60s Jan 1960 Jan 1961 Jan 1962 Jan 1963 Jan 1964 Jan 1965 Jan 1966 Jan 1967 Jan 1968 PLOTS of % monthly 1953 ADPSFC obs with psfc < 1050mb Jan 1953 Feb 1953 Mar 1953 Apr 1953 May 1953 Jun 1953 Jul 1953 Aug 1953 Sep 1953 Oct 1953 Nov 1953 Dec 1953

This error has good news and bad news. The good news is that the "bad data" were omitted. The analyses are not contaminated. The error is one of omission, not comission.

The bad news is that the omission of data is correlated with Northern Hemisphere extra-tropical cyclones, thereby requiring the dynamics of the data assimilation process to completely determine the details of the surface pressure without the benefit of direct measurement when the pressure lowered below 1000 mb. In addition (and the source of the serendipity) European obs seems to bear the brunt of the problem, unlike US obs which were unaffected.

Evaluation - Reruns for Jan, Jul 1953

A methodology to identify the miscoded sfc press and mslp obs proved more involved that I first thought.

1. Read all sfc obs for all synoptic times
2. Store sfc obs sorted by observed hour and location
3. Do a time series analysis of locations that reported either pressure value over 1049.9 mb.
4. Look for tendencies of adjacent times (restricted to no more than 24 hrs) of 75mb. This kept cold Siberian highs from being viewed as mis-coded.
5. Flag the "excessive" sequence.
6. When flagged mslp/ unflagged sfc press found, check the mslp d-value minus sfcp d-value difference.

   i. Small (<25mb): flag psfc for correction

*URL http://lnx21.wwb.noaa.gov/images/psfc/psfc.html

   ii. Large (>75mb, usually ~100mb) psfc is ok

   iii. Indeterminate (>25 and <75): flag for rejection.

7. Write a synoptically time sorted file for the assimilation bufr editor to read and apply

In order assess the impact of encoding error, a winter and summer month were rerun with corrected values of surface pressure. We chose Jan 1953 and July 1953 arbitrarily. An assessment of the impact on the monthly mean surface pressure fields is given here. It is my opinion that the error does not warrant a rerun.

0.6.2 NCEP Surface temperature (air.sfc), 2 meter temperature (air.2m), and TMAX temperature (tmax.2m). URL: Surface temperature (air.sfc), 2 meter temperature (air.2m), and TMAX temperature (tmax.2m).

The skin temperature is determined diagnostically. It is the temperature required to balance the fluxes at the surface.

In air, sensible heat flux, long wave radiation (up and down) depend on the skin temperature. In the ground, the heat flux into the soil depends on the skin temperature.

The problem occurs when the winds are weak (less than 0.75 m/s), and the stability conditions are just right. Under these conditions, the parameterization for the thermal exchange coefficient can break down. This caused the model to blow-up (divide by zero). (After running this code for years, the bug surfaced.) A bug fix was introduced (instead of dividing by zero, dividing by a small number). However, the fundamental problem remained. Under light winds, the thermal exchange coefficient can be close to zero. For argument's sake, let's assume that the exchange coefficient is zero. In this case, the sensible heating is zero, and the solar heating has to be balanced by latent heat flux, the net long-wave flux and a ground heat flux. Since the latent heat flux is not dependent on the skin temperature, the reduction of the sensible heat flux must be balanced by an increased upward long wave flux and a downward ground heat flux. In order to make this balance work, the skin temperature will become unreasonably hot. On the next time step, the winds may pickup and skin temperature will return to normal (skin temperature has no inertia). The affect on the lowest sigma level is minor as the net fluxes remain reasonable. This phenomenon will not happen over water as the skin temperature (SST) is fixed.

TMAX and possibly TMIN are affected by this problem. TMAX is the worst possible case - it will capture spikes that only last 1 time step

The surface temperature will be affected.

The 2m temperature is interpolated from the skin temperature and the bottom sigma-layer and thus will be affected.

Sensible heat flux and the upward long wave radiation flux are also affected.

A suggested work around is to reject instantaneous data if the following two conditions are true:

- The data is over land.
- The 10 meter wind is less than .75 m/s.

This test rejects   2% of the total number of grid points. It will not work for averaged quantities and Tmax, Tmin. In addition, this test rejects an awful amount of good data. In the monthly means produced at NCEP this screening test will not be used because ejecting so much good data would be worse than including a few bad points (which get smoothed out in the monthly averages).

It may also be possible to reject data greater than N standard deviations from the mean. No tests have been conducted with this screening method.

0.6.3 Problems with aircraft data January 1976 to December 1978

The corrected data have been on-line at CDC since 14 March 1998.

Prior to March 14, 1998 the data for January 1976 to December 1978 were corrupted due to a miscommunication among the US Air Force, NCAR and NCEP regarding a data set of aircraft reports for this period only. Along the way the sense of E-W longitude was flipped, causing a sizable portion of these aircraft data for these three years to be mis-positioned.

0.6.4   Problems with the June 1997 (CDAS) data

The Climate Data Assimilation System (CDAS) data for the NCEP/NCAR Reanalysis project for June 1997 was recalculated after it was originally released at the beginning of July. This recalculation was necessitated by three factors:

1. The strong El Nio.
2. Problems with the radiances coming form NOAA-12 (started   1 June).
3. The grib identification for the ingest ice anl's were changed during the month - the original run ceased ingesting these data as of the date of the change.

The corrected version of the June 1997 CDAS data has been on-line at CDC since 20 Oct 1997.

0.7   The test-model: Scores

0.7.1   Temperature scores

TABLE 10.   Prediction scores (anomalous correlation / RMS error) for *monthly* mean Bergen temperature using SLP, SST, and the Bergen monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. Those scores which beat the benchmark are shown in bold. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.99 / 0.10**	**0.50 / 0.30**	-0.31 / 0.24
Feb	**0.99 / 0.08**	**0.39** / 0.20	**0.17** / 0.17
Mar	**1.00 / 0.02**	-0.26 / 0.19	-0.38 / 0.18
Apr	**1.00 / 0.02**	-0.21 / 0.20	**0.23 / 0.22**
May	**1.00 / 0.01**	**0.11 / 0.23**	-0.19 / 0.23
Jun	**0.99 / 0.03**	**-0.16** / 0.25	**0.15 / 0.21**
Jul	**0.97 / 0.05**	**0.20 / 0.21**	-0.12 / 0.24
Aug	**0.99 / 0.03**	-0.07 / 0.23	-0.20 / 0.21
Sep	**0.98 / 0.05**	**-0.11** / 0.25	**-0.11** / 0.28
Oct	**1.00 / 0.02**	**-0.02** / 0.28	-0.34 / 0.31
Nov	**0.99 / 0.05**	**0.23 / 0.29**	**0.14 / 0.39**
Dec	**0.99 / 0.04**	**0.09 / 0.39**	**0.17 / 0.34**
benchmark			
Jan	0.35 / 0.36	0.07 / 0.34	-0.21 / 0.22
Feb	0.34 / 0.31	-0.21 / 0.22	0.05 / 0.17
Mar	0.26 / 0.19	0.05 / 0.17	0.00 / 0.18
Apr	0.43 / 0.15	0.00 / 0.18	-0.22 / 0.24
May	0.34 / 0.17	-0.22 / 0.24	-0.18 / 0.20
Jun	0.24 / 0.22	-0.18 / 0.20	-0.18 / 0.22
Jul	0.23 / 0.18	-0.18 / 0.22	0.24 / 0.21
Aug	0.31 / 0.20	0.24 / 0.21	-0.18 / 0.21
Sep	0.46 / 0.19	-0.18 / 0.21	-0.13 / 0.28
Oct	0.22 / 0.19	-0.13 / 0.28	-0.27 / 0.32
Nov	0.22 / 0.25	-0.27 / 0.32	0.06 / 0.40
Dec	0.13 / 0.29	0.06 / 0.40	0.07 / 0.34

TABLE 11.   Prediction scores (anomalous correlation / RMS error) for *2-monthly* mean Bergen temperature using SLP, SST, and the Bergen monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part.

start	lag 0	lag 1	lag 2
Jan	**0.90 / 0.16**	**0.42** / 0.21	-0.32 / 0.16
Feb	**0.90 / 0.11**	**0.37** / 0.14	-0.30 / 0.13
Mar	**0.79 / 0.08**	**0.02** / 0.14	-0.23 / 0.16
Apr	**0.67 / 0.09**	-0.21 / 0.16	**-0.07** / 0.18
May	**0.69 / 0.12**	**0.07** / 0.15	**-0.21** / 0.19
Jun	**0.68 / 0.12**	**0.01** / 0.19	0.02 / 0.17
Jul	**0.73 / 0.11**	0.12 / 0.16	-0.17 / 0.18
Aug	**0.68 / 0.12**	-0.07 / 0.15	-0.35 / 0.18
Sep	**0.71 / 0.11**	**0.05** / 0.20	**-0.21** / 0.25
Oct	**0.64 / 0.13**	-0.28 / 0.25	**-0.01 / 0.29**
Nov	**0.68 / 0.16**	-0.06 / 0.30	**0.16 / 0.33**
Dec	**0.79 / 0.19**	**0.23 / 0.32**	**0.14 / 0.23**
benchmark			
Jan	0.37 / 0.30	0.03 / 0.23	-0.02 / 0.14
Feb	0.33 / 0.21	-0.02 / 0.14	-0.12 / 0.13
Mar	0.36 / 0.13	-0.12 / 0.13	-0.08 / 0.17
Apr	0.41 / 0.12	-0.08 / 0.17	-0.19 / 0.15
May	0.30 / 0.15	-0.19 / 0.15	-0.24 / 0.17
Jun	0.26 / 0.14	-0.24 / 0.17	0.16 / 0.15
Jul	0.24 / 0.15	0.16 / 0.15	0.12 / 0.14
Aug	0.43 / 0.14	0.12 / 0.14	-0.08 / 0.18
Sep	0.38 / 0.13	-0.08 / 0.18	-0.23 / 0.24
Oct	0.24 / 0.17	-0.23 / 0.24	-0.05 / 0.31
Nov	0.13 / 0.22	-0.05 / 0.31	0.12 / 0.33
Dec	0.23 / 0.28	0.12 / 0.33	0.03 / 0.23

TABLE 12. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Bergen temperature using SLP, SST, and the Bergen monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.87 / 0.14**	**0.44 / 0.16**	-0.15 / 0.12
Feb	**0.88 / 0.09**	**0.42 / 0.10**	-0.26 / 0.11
Mar	**0.63 / 0.08**	**-0.16 / 0.12**	**0.00 / 0.13**
Apr	0.29 / 0.11	-0.20 / 0.15	**-0.06 / 0.18**
May	**0.57 / 0.10**	**-0.05 / 0.15**	**0.02 / 0.15**
Jun	**0.58 / 0.12**	**0.03 / 0.16**	-0.08 / 0.13
Jul	**0.59 / 0.11**	-0.12 / 0.15	-0.22 / 0.14
Aug	**0.58 / 0.10**	-0.12 / 0.15	-0.35 / 0.16
Sep	**0.47 / 0.13**	**-0.11 / 0.19**	**-0.08 / 0.24**
Oct	**0.26 / 0.16**	-0.37 / 0.24	0.03 / 0.28
Nov	**0.47 / 0.20**	0.03 / 0.28	**0.21 / 0.25**
Dec	**0.68 / 0.21**	**0.24 / 0.25**	0.08 / 0.17
benchmark			
Jan	0.35 / 0.23	0.09 / 0.17	-0.08 / 0.11
Feb	0.37 / 0.16	-0.08 / 0.11	-0.17 / 0.11
Mar	0.34 / 0.10	-0.17 / 0.11	-0.05 / 0.13
Apr	0.30 / 0.11	-0.05 / 0.13	-0.30 / 0.15
May	0.30 / 0.12	-0.30 / 0.15	0.01 / 0.13
Jun	0.24 / 0.14	0.01 / 0.13	0.16 / 0.12
Jul	0.36 / 0.12	0.16 / 0.12	0.07 / 0.14
Aug	0.45 / 0.11	0.07 / 0.14	-0.18 / 0.17
Sep	0.34 / 0.12	-0.18 / 0.17	-0.10 / 0.24
Oct	0.15 / 0.15	-0.10 / 0.24	0.04 / 0.28
Nov	0.18 / 0.22	0.04 / 0.28	0.10 / 0.25
Dec	0.29 / 0.26	0.10 / 0.25	0.09 / 0.17

TABLE 13. Prediction anomalous correlation scores for monthly mean, 2-month mean, seasonal mean temperature in Bergen. Calibration period = 1900-1959, evaluation interval=1960-2000 using SLP, SST, sunspot data, and previous month's temperature. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2	lag 3
monthly				
Jan	**0.98 / 0.11**	**0.49 / 0.30**	-0.30 / 0.24	
Feb	**0.99 / 0.09**	**0.39 / 0.20**	**0.19 / 0.17**	
Mar	**0.98 / 0.04**	-0.11 / 0.20	**0.03 / 0.19**	
Apr	**1.00 / 0.02**	**0.03 / 0.18**	**0.22 / 0.22**	
May	**1.00 / 0.01**	**0.05 / 0.23**	-0.19 / 0.23	
Jun	**0.98 / 0.04**	-0.18 / 0.25	**0.06 / 0.22**	
Jul	**0.96 / 0.06**	**0.14 / 0.22**	0.02 / 0.23	
Aug	**0.99 / 0.03**	0.02 / 0.23	**-0.16 / 0.21**	
Sep	**0.97 / 0.05**	**-0.12 / 0.24**	**-0.03 / 0.27**	
Oct	**1.00 / 0.02**	**0.03 / 0.27**	-0.31 / 0.31	
Nov	**0.99 / 0.05**	**0.24 / 0.29**	**0.11 / 0.40**	
Dec	**0.99 / 0.04**	-0.13 / 0.42	**0.18 / 0.34**	
2-month				
Jan	**0.89 / 0.17**	**0.38 / 0.21**	-0.33 / 0.16	
Feb	**0.90 / 0.11**	**0.36 / 0.14**	**-0.05 / 0.13**	
Mar	**0.76 / 0.09**	**-0.01 / 0.13**	-0.26 / 0.16	
Apr	**0.69 / 0.09**	-0.20 / 0.19	**-0.09 / 0.18**	
May	**0.68 / 0.12**	**-0.02 / 0.16**	**-0.19 / 0.19**	
Jun	**0.65 / 0.12**	**0.01 / 0.19**	0.09 / 0.17	
Jul	**0.71 / 0.11**	0.20 / 0.17	-0.15 / 0.17	
Aug	**0.71 / 0.11**	-0.01 / 0.15	**-0.03 / 0.18**	
Sep	**0.71 / 0.11**	**-0.07 / 0.20**	-0.20 / 0.25	
Oct	**0.65 / 0.13**	-0.28 / 0.25	**-0.01 / 0.29**	
Nov	**0.68 / 0.16**	-0.08 / 0.30	0.16 / 0.33	
Dec	**0.81 / 0.18**	**0.26 / 0.32**	**0.21 / 0.22**	
seasonal				
Jan	**0.86 / 0.14**	**0.42 / 0.16**	**0.31 / 0.10**	
Feb	**0.88 / 0.09**	**0.37 / 0.11**	**-0.13 / 0.12**	
Mar	**0.59 / 0.09**	**0.04 / 0.11**	-0.13 / 0.14	
Apr	**0.33 / 0.11**	-0.24 / 0.16	**-0.06 / 0.18**	
May	**0.57 / 0.10**	**-0.01 / 0.15**	-0.05 / 0.14	
Jun	**0.58 / 0.12**	**0.08 / 0.15**	-0.04 / 0.14	
Jul	**0.60 / 0.11**	-0.08 / 0.15	0.06 / 0.14	
Aug	**0.60 / 0.10**	0.05 / 0.14	**0.05 / 0.16**	
Sep	**0.47 / 0.13**	**-0.12 / 0.19**	-0.10 / 0.24	
Oct	**0.26 / 0.16**	-0.38 / 0.25	-0.06 / 0.29	
Nov	**0.38 / 0.21**	0.02 / 0.27	**0.18 / 0.25**	
Dec	**0.70 / 0.21**	**0.27 / 0.24**	**0.18 / 0.17**	

TABLE 14. Prediction anomalous correlation scores for monthly mean, 2-month mean, and seasonal mean temperature in Bergen. Calibration period = 1900-1959, evaluation interval=1960-2000 using only the monthly temperature from the month when the forecast is issued. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
monthly			
Jan	0.99 / 0.10	0.57 / 0.29	0.25 / 0.20
Feb	0.99 / 0.08	0.34 / 0.20	0.17 / 0.17
Mar	1.00 / 0.02	-0.06 / 0.17	-0.36 / 0.18
Apr	1.00 / 0.02	-0.12 / 0.19	-0.05 / 0.23
May	1.00 / 0.01	0.11 / 0.23	-0.28 / 0.21
Jun	0.99 / 0.03	-0.18 / 0.22	0.14 / 0.21
Jul	0.97 / 0.05	0.20 / 0.21	-0.11 / 0.23
Aug	0.99 / 0.03	-0.07 / 0.23	-0.38 / 0.20
Sep	0.98 / 0.05	-0.10 / 0.22	-0.12 / 0.28
Oct	1.00 / 0.02	0.01 / 0.27	-0.28 / 0.31
Nov	0.99 / 0.05	0.12 / 0.29	0.07 / 0.39
Dec	0.99 / 0.04	0.28 / 0.38	0.08 / 0.34
2-month			
Jan	0.90 / 0.16	0.57 / 0.19	0.34 / 0.13
Feb	0.90 / 0.11	0.40 / 0.13	0.13 / 0.12
Mar	0.79 / 0.08	-0.13 / 0.13	-0.24 / 0.16
Apr	0.67 / 0.09	-0.28 / 0.17	-0.41 / 0.15
May	0.69 / 0.12	0.05 / 0.15	-0.31 / 0.18
Jun	0.75 / 0.10	0.020 / 0.17	0.02 / 0.17
Jul	0.73 / 0.11	0.12 / 0.16	-0.27 / 0.16
Aug	0.68 / 0.12	-0.07 / 0.15	-0.17 / 0.18
Sep	0.71 / 0.11	-0.070 / 0.20	-0.30 / 0.25
Oct	0.64 / 0.13	-0.25 / 0.24	-0.06 / 0.29
Nov	0.68 / 0.16	0.06 / 0.29	0.07 / 0.33
Dec	0.79 / 0.19	0.27 / 0.32	-0.06 / 0.23
seasonal			
Jan	0.87 / 0.14	0.59 / 0.14	0.23 / 0.11
Feb	0.88 / 0.09	0.43 / 0.10	-0.15 / 0.11
Mar	0.63 / 0.08	-0.13 / 0.11	-0.35 / 0.13
Apr	0.29 / 0.11	-0.40 / 0.13	-0.37 / 0.15
May	0.57 / 0.10	-0.10 / 0.15	-0.29 / 0.15
Jun	0.64 / 0.11	-0.02 / 0.15	-0.10 / 0.13
Jul	0.59 / 0.11	-0.03 / 0.13	-0.22 / 0.14
Aug	0.58 / 0.10	-0.12 / 0.15	-0.25 / 0.16
Sep	0.47 / 0.13	-0.26 / 0.18	-0.07 / 0.23
Oct	0.26 / 0.16	-0.23 / 0.24	-0.00 / 0.27
Nov	0.47 / 0.20	0.14 / 0.27	0.12 / 0.25
Dec	0.69 / 0.21	0.16 / 0.25	0.10 / 0.17

TABLE 15. Prediction anomalous correlation scores for monthly mean, 2-month mean, and seasonal mean temperature in Bergen. Calibration period = 1900-1959, evaluation interval=1960-2000 using only SLP, and SST. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
monthly			
Jan	**0.66 / 0.31**	**0.38 / 0.34**	-0.47 / 0.23
Feb	**0.70 / 0.26**	-0.04 / 0.27	-0.09 / 0.18
Mar	**0.60 / 0.18**	-0.09 / 0.18	-0.42 / 0.18
Apr	**0.50 / 0.17**	-0.09 / 0.20	**0.27 / 0.22**
May	**0.63 / 0.15**	**-0.13 / 0.24**	**-0.14 / 0.22**
Jun	**0.56 / 0.19**	**-0.06 / 0.23**	0.05 / 0.22
Jul	**0.66 / 0.14**	**-0.10 / 0.24**	-0.33 / 0.27
Aug	**0.40 / 0.20**	-0.36 / 0.28	-0.41 / 0.20
Sep	**0.56 / 0.18**	-0.09 / 0.24	-0.25 / 0.28
Oct	**0.66 / 0.15**	**0.14 / 0.26**	-0.22 / 0.30
Nov	**0.41 / 0.24**	**0.12 / 0.30**	**0.08 / 0.40**
Dec	**0.71 / 0.21**	**0.13 / 0.39**	**0.26 / 0.33**
2-month			
Jan	**0.60 / 0.28**	**0.33 / 0.22**	**0.24 / 0.14**
Feb	**0.62 / 0.19**	**0.46 / 0.13**	-0.24 / 0.13
Mar	**0.50 / 0.13**	**0.10 / 0.12**	-0.24 / 0.16
Apr	0.37 / 0.12	**-0.07 / 0.17**	**-0.04 / 0.17**
May	**0.41 / 0.15**	**0.02 / 0.16**	**-0.00 / 0.18**
Jun	**0.27 / 0.16**	**-0.04 / 0.19**	-0.01 / 0.17
Jul	**0.33 / 0.17**	-0.09 / 0.20	-0.16 / 0.16
Aug	-0.09 / 0.21	-0.27 / 0.16	-0.21 / 0.18
Sep	**0.44 / 0.15**	-0.16 / 0.19	-0.41 / 0.26
Oct	**0.44 / 0.16**	**-0.01 / 0.23**	0.01 / 0.29
Nov	0.29 / 0.22	**-0.03 / 0.30**	0.09 / 0.34
Dec	**0.35 / 0.27**	**0.35 / 0.31**	**0.12 / 0.22**
seasonal			
Jan	**0.65 / 0.20**	**0.35 / 0.17**	**0.28 / 0.10**
Feb	**0.74 / 0.13**	**0.19 / 0.11**	-0.25 / 0.13
Mar	**0.45 / 0.10**	-0.35 / 0.11	**0.05 / 0.13**
Apr	0.11 / 0.11	-0.14 / 0.14	**-0.09 / 0.18**
May	0.12 / 0.13	**0.01 / 0.16**	-0.17 / 0.17
Jun	0.24 / 0.15	-0.00 / 0.16	-0.13 / 0.14
Jul	0.25 / 0.16	0.03 / 0.14	-0.20 / 0.14
Aug	0.02 / 0.15	-0.42 / 0.15	-0.27 / 0.16
Sep	0.22 / 0.14	-0.23 / 0.19	-0.30 / 0.26
Oct	**0.30 / 0.15**	-0.13 / 0.23	**0.05 / 0.27**
Nov	**0.20 / 0.22**	-0.00 / 0.27	**0.11 / 0.26**
Dec	**0.44 / 0.24**	**0.35 / 0.24**	**0.34 / 0.16**

TABLE 16. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Oslo temperature using SLP, SST, and the local monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	0.88 / 0.23	0.74 / 0.23	-0.31 / 0.19
Feb	0.93 / 0.13	0.56 / 0.14	-0.07 / 0.13
Mar	0.81 / 0.10	0.20 / 0.11	-0.06 / 0.15
Apr	0.30 / 0.13	-0.32 / 0.16	0.00 / 0.19
May	0.52 / 0.11	-0.09 / 0.18	-0.09 / 0.17
Jun	0.53 / 0.14	-0.11 / 0.19	-0.15 / 0.16
Jul	0.73 / 0.11	0.25 / 0.14	-0.10 / 0.16
Aug	0.76 / 0.10	0.15 / 0.15	0.13 / 0.21
Sep	0.41 / 0.15	0.19 / 0.22	-0.14 / 0.31
Oct	0.40 / 0.20	0.19 / 0.29	0.14 / 0.40
Nov	0.48 / 0.26	0.23 / 0.39	0.25 / 0.38
Dec	0.66 / 0.31	0.22 / 0.38	0.12 / 0.30
benchmark			
Jan	0.41 / 0.36	0.22 / 0.29	0.22 / 0.17
Feb	0.46 / 0.26	0.22 / 0.17	-0.06 / 0.12
Mar	0.45 / 0.15	-0.06 / 0.12	-0.08 / 0.14
Apr	0.24 / 0.11	-0.08 / 0.14	-0.16 / 0.18
May	0.26 / 0.12	-0.16 / 0.18	-0.15 / 0.17
Jun	0.27 / 0.16	-0.15 / 0.17	-0.11 / 0.15
Jul	0.35 / 0.15	-0.11 / 0.15	-0.13 / 0.16
Aug	0.38 / 0.14	-0.13 / 0.16	-0.24 / 0.23
Sep	0.38 / 0.15	-0.24 / 0.23	0.04 / 0.30
Oct	0.30 / 0.20	0.04 / 0.30	0.08 / 0.41
Nov	0.30 / 0.28	0.08 / 0.41	0.17 / 0.39
Dec	0.32 / 0.37	0.17 / 0.39	0.22 / 0.29

TABLE 17. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Stockholm temperature using SLP, SST, and the local monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	0.86 / 0.22	0.69 / 0.23	0.67 / 0.14
Feb	0.92 / 0.15	0.64 / 0.15	0.25 / 0.13
Mar	0.83 / 0.11	0.34 / 0.12	0.07 / 0.18
Apr	0.37 / 0.14	-0.24 / 0.19	0.12 / 0.21
May	0.56 / 0.13	-0.18 / 0.22	-0.13 / 0.23
Jun	0.46 / 0.17	0.05 / 0.21	-0.01 / 0.16
Jul	0.80 / 0.12	0.29 / 0.14	0.01 / 0.15
Aug	0.70 / 0.11	0.13 / 0.15	-0.19 / 0.19
Sep	0.49 / 0.14	-0.25 / 0.23	-0.37 / 0.32
Oct	0.29 / 0.18	-0.51 / 0.31	-0.28 / 0.42
Nov	0.32 / 0.25	0.18 / 0.39	0.28 / 0.40
Dec	0.66 / 0.31	0.14 / 0.41	0.17 / 0.31
benchmark			
Jan	0.35 / 0.38	0.15 / 0.31	0.23 / 0.18
Feb	0.42 / 0.28	0.23 / 0.18	-0.06 / 0.13
Mar	0.45 / 0.16	-0.06 / 0.13	-0.09 / 0.16
Apr	0.23 / 0.12	-0.09 / 0.16	-0.03 / 0.19
May	0.31 / 0.15	-0.03 / 0.19	0.01 / 0.20
Jun	0.34 / 0.17	0.01 / 0.20	-0.05 / 0.15
Jul	0.49 / 0.17	-0.05 / 0.15	-0.09 / 0.15
Aug	0.43 / 0.14	-0.09 / 0.15	-0.27 / 0.18
Sep	0.28 / 0.14	-0.27 / 0.18	-0.17 / 0.29
Oct	0.18 / 0.17	-0.17 / 0.29	-0.03 / 0.42
Nov	0.15 / 0.26	-0.03 / 0.42	0.07 / 0.42
Dec	0.26 / 0.38	0.07 / 0.42	0.15 / 0.31

TABLE 18.   Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Copenhagen temperature using SLP, SST, and the local monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.85 / 0.19**	**0.69 / 0.20**	**0.64 / 0.14**
Feb	**0.92 / 0.13**	**0.70 / 0.14**	**0.47 / 0.11**
Mar	**0.88 / 0.10**	**0.26 / 0.12**	-0.09 / 0.15
Apr	**0.44 / 0.12**	-0.07 / 0.15	-0.15 / 0.20
May	**0.67 / 0.10**	-0.08 / 0.18	-0.27 / 0.20
Jun	**0.44** / 0.15	-0.01 / 0.18	-0.23 / 0.13
Jul	**0.86 / 0.09**	**0.50 / 0.11**	-0.01 / 0.13
Aug	**0.73 / 0.10**	**0.19** / 0.13	0.05 / 0.14
Sep	**0.51** / 0.12	**-0.13** / 0.18	**0.12 / 0.21**
Oct	**0.31** / 0.13	-0.25 / 0.22	-0.16 / 0.32
Nov	**0.25 / 0.20**	**0.22 / 0.30**	**0.27 / 0.32**
Dec	**0.71 / 0.23**	**0.27 / 0.32**	**0.20** / 0.27
benchmark			
Jan	0.37 / 0.30	0.18 / 0.26	0.23 / 0.17
Feb	0.43 / 0.24	0.23 / 0.17	0.02 / 0.12
Mar	0.44 / 0.15	0.02 / 0.12	-0.01 / 0.15
Apr	0.26 / 0.11	-0.01 / 0.15	-0.00 / 0.17
May	0.29 / 0.13	-0.00 / 0.17	0.03 / 0.18
Jun	0.36 / 0.15	0.03 / 0.18	-0.07 / 0.13
Jul	0.44 / 0.16	-0.07 / 0.13	0.02 / 0.11
Aug	0.45 / 0.12	0.02 / 0.11	-0.25 / 0.13
Sep	0.38 / 0.10	-0.25 / 0.13	-0.20 / 0.23
Oct	0.24 / 0.12	-0.20 / 0.23	-0.03 / 0.32
Nov	0.13 / 0.21	-0.03 / 0.32	0.09 / 0.34
Dec	0.27 / 0.29	0.09 / 0.34	0.18 / 0.26

TABLE 19.   Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Helsinki temperature using SLP, SST, and the local monthly mean temperature at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.82 / 0.27**	**0.61 / 0.26**	**0.61 / 0.15**
Feb	**0.92 / 0.15**	**0.59 / 0.16**	**0.37 / 0.13**
Mar	**0.82 / 0.13**	**0.25 / 0.15**	**0.13** / 0.18
Apr	**0.47 / 0.14**	-0.04 / 0.18	**0.24** / 0.18
May	**0.46** / 0.15	-0.21 / 0.19	-0.15 / 0.17
Jun	**0.60 / 0.14**	0.07 / 0.18	-0.26 / 0.17
Jul	**0.68 / 0.12**	-0.06 / 0.17	-0.05 / 0.20
Aug	**0.41** / 0.14	**0.10 / 0.19**	**-0.00** / 0.22
Sep	**0.51 / 0.17**	**0.19** / 0.21	0.03 / 0.34
Oct	**0.30** / 0.22	-0.45 / 0.40	-0.17 / 0.50
Nov	0.13 / 0.35	**0.05** / 0.47	**0.25 / 0.44**
Dec	**0.63 / 0.36**	**0.32 / 0.43**	**0.22 / 0.32**
benchmark			
Jan	0.34 / 0.42	0.07 / 0.34	0.19 / 0.18
Feb	0.38 / 0.30	0.19 / 0.18	-0.07 / 0.14
Mar	0.42 / 0.16	-0.07 / 0.14	-0.27 / 0.17
Apr	0.22 / 0.13	-0.27 / 0.17	-0.22 / 0.17
May	0.23 / 0.15	-0.22 / 0.17	-0.12 / 0.16
Jun	0.31 / 0.15	-0.12 / 0.16	0.01 / 0.16
Jul	0.44 / 0.14	0.01 / 0.16	-0.07 / 0.20
Aug	0.29 / 0.14	-0.07 / 0.20	-0.06 / 0.20
Sep	0.28 / 0.18	-0.06 / 0.20	-0.13 / 0.34
Oct	0.27 / 0.19	-0.13 / 0.34	-0.02 / 0.47
Nov	0.19 / 0.31	-0.02 / 0.47	0.04 / 0.47
Dec	0.28 / 0.43	0.04 / 0.47	0.07 / 0.34

0.7.2   Precipitation scores

TABLE 20.   Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Bergen precipitation using SLP, SST, and the local monthly mean precipitation at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	0.83 / 9.29	0.52 / 10.36	0.38 / 8.20
Feb	0.76 / 8.14	0.28 / 8.57	0.02 / 6.07
Mar	0.90 / 4.32	0.36 / 5.35	0.15 / 6.45
Apr	0.63 / 4.51	-0.30 / 6.48	-0.15 / 7.16
May	0.49 / 5.31	-0.06 / 7.03	0.03 / 9.15
Jun	0.37 / 6.08	-0.24 / 9.46	0.04 / 9.37
Jul	0.18 / 8.86	-0.34 / 10.70	-0.12 / 12.19
Aug	0.34 / 9.30	-0.11 / 12.34	-0.07 / 13.40
Sep	0.56 / 9.49	-0.00 / 11.83	0.10 / 12.91
Oct	0.56 / 10.02	-0.14 / 13.84	-0.02 / 14.12
Nov	0.63 / 10.12	0.15 / 13.23	0.24 / 14.07
Dec	0.72 / 10.67	0.13 / 14.19	0.20 / 11.43
benchmark			
Jan	0.56 / 11.24	0.44 / 10.29	0.20 / 8.62
Feb	0.62 / 9.11	0.20 / 8.62	-0.11 / 6.03
Mar	0.42 / 7.83	-0.11 / 6.03	0.03 / 6.24
Apr	0.33 / 5.43	0.03 / 6.24	-0.13 / 6.84
May	0.34 / 5.68	-0.13 / 6.84	-0.10 / 9.54
Jun	0.22 / 6.38	-0.10 / 9.54	0.02 / 9.68
Jul	0.26 / 8.65	0.02 / 9.68	0.03 / 11.92
Aug	0.32 / 8.58	0.03 / 11.92	-0.02 / 12.15
Sep	0.35 / 10.47	-0.02 / 12.15	0.14 / 12.89
Oct	0.26 / 11.00	0.14 / 12.89	0.32 / 12.44
Nov	0.37 / 11.71	0.32 / 12.44	0.39 / 12.57
Dec	0.53 / 11.15	0.39 / 12.57	0.44 / 10.29

TABLE 21.   Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Oslo precipitation using SLP, SST, and the local monthly mean precipitation at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	0.74 / 2.21	0.00 / 2.79	-0.14 / 2.97
Feb	0.42 / 2.49	-0.03 / 2.85	0.02 / 2.46
Mar	0.50 / 2.31	0.03 / 2.37	-0.20 / 3.08
Apr	0.25 / 2.33	-0.09 / 3.05	-0.04 / 4.12
May	0.35 / 2.56	0.16 / 3.90	0.11 / 4.29
Jun	0.36 / 3.68	0.09 / 4.44	0.18 / 4.11
Jul	0.56 / 3.54	0.33 / 3.71	0.01 / 4.17
Aug	0.45 / 3.65	0.20 / 3.90	-0.05 / 4.07
Sep	0.44 / 3.59	0.29 / 3.78	-0.15 / 3.46
Oct	0.75 / 2.61	0.27 / 3.00	0.16 / 2.78
Nov	0.65 / 2.44	-0.07 / 3.15	-0.15 / 3.54
Dec	0.51 / 2.40	0.21 / 3.12	0.20 / 2.62
benchmark			
Jan	0.31 / 2.94	0.10 / 2.65	-0.10 / 2.66
Feb	0.31 / 2.47	-0.10 / 2.66	-0.05 / 2.49
Mar	0.14 / 2.49	-0.05 / 2.49	-0.04 / 2.74
Apr	0.21 / 2.28	-0.04 / 2.74	-0.01 / 4.05
May	0.25 / 2.50	-0.01 / 4.05	-0.08 / 4.47
Jun	0.27 / 3.68	-0.08 / 4.47	-0.05 / 4.23
Jul	0.26 / 4.08	-0.05 / 4.23	-0.26 / 4.25
Aug	0.22 / 3.82	-0.26 / 4.25	-0.11 / 4.14
Sep	0.24 / 3.79	-0.11 / 4.14	-0.19 / 3.24
Oct	0.31 / 3.76	-0.19 / 3.24	-0.17 / 2.96
Nov	0.41 / 2.92	-0.17 / 2.96	0.04 / 3.20
Dec	0.30 / 2.67	0.04 / 3.20	0.10 / 2.65

TABLE 22. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Stockholm precipitation using SLP, SST, and the local monthly mean precipitation at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.55 / 1.17**	-0.12 / 1.63	-0.08 / 1.71
Feb	0.40 / 1.41	-0.09 / 1.67	-0.25 / 1.94
Mar	0.51 / 1.42	**0.14 / 1.79**	**-0.09 / 3.03**
Apr	0.38 / 1.69	0.07 / 2.91	-0.18 / 3.75
May	**0.28 / 2.80**	**-0.10 / 3.52**	-0.12 / 3.16
Jun	**0.61 / 2.59**	**0.36 / 2.87**	**-0.02 / 3.12**
Jul	**0.55 / 2.61**	**-0.18 / 3.00**	**-0.13 / 2.54**
Aug	**0.59 / 2.38**	**-0.02 / 2.43**	-0.16 / 2.86
Sep	**0.40 / 2.34**	**-0.02 / 2.80**	-0.17 / 2.19
Oct	**0.77 / 1.74**	-0.15 / 2.18	-0.16 / 2.03
Nov	0.22 / 2.11	-0.33 / 2.36	-0.12 / 1.67
Dec	**0.74 / 1.26**	-0.29 / 1.62	**0.12 / 1.54**
benchmark			
Jan	0.39 / 1.29	0.05 / 1.55	0.12 / 1.62
Feb	0.46 / 1.40	0.12 / 1.62	0.00 / 1.86
Mar	0.53 / 1.47	0.00 / 1.86	-0.21 / 3.15
Apr	0.42 / 1.64	-0.21 / 3.15	-0.20 / 3.59
May	0.20 / 2.82	-0.20 / 3.59	-0.10 / 3.29
Jun	0.23 / 3.16	-0.10 / 3.29	-0.21 / 3.13
Jul	0.29 / 2.92	-0.21 / 3.13	-0.30 / 2.51
Aug	0.17 / 2.89	-0.30 / 2.51	-0.04 / 2.80
Sep	0.10 / 2.34	-0.04 / 2.80	0.06 / 1.97
Oct	0.35 / 2.52	0.06 / 1.97	-0.05 / 1.93
Nov	0.40 / 1.77	-0.05 / 1.93	-0.06 / 1.45
Dec	0.29 / 1.77	-0.06 / 1.45	0.05 / 1.55

TABLE 23. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Copenhagen precipitation using SLP, SST, and the local monthly mean precipitation at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	**0.75 / 1.62**	-0.10 / 2.53	-0.12 / 2.70
Feb	**0.46 / 2.09**	**-0.24 / 2.67**	**-0.23 / 2.86**
Mar	**0.52 / 2.19**	-0.42 / 3.06	**-0.05 / 2.74**
Apr	**0.43 / 2.32**	**0.20 / 2.31**	-0.26 / 3.54
May	**0.49 / 2.06**	**0.18 / 3.18**	**0.28 / 3.12**
Jun	**0.45 / 2.89**	-0.08 / 3.60	-0.16 / 3.31
Jul	**0.59 / 2.61**	**-0.34 / 2.98**	**0.05 / 2.84**
Aug	**0.37 / 2.77**	**-0.11 / 2.93**	**0.00 / 2.58**
Sep	**0.68 / 2.22**	**0.04 / 2.51**	**-0.07 / 3.31**
Oct	**0.30 / 2.62**	-0.25 / 3.51	-0.38 / 3.37
Nov	0.29 / 3.00	-0.30 / 3.27	-0.25 / 2.70
Dec	**0.79 / 1.86**	**0.21 / 2.31**	**0.10 / 2.38**
benchmark			
Jan	0.29 / 2.24	-0.47 / 2.50	-0.41 / 2.84
Feb	0.37 / 2.22	-0.41 / 2.84	-0.41 / 2.92
Mar	0.10 / 2.56	-0.41 / 2.92	-0.13 / 2.60
Apr	0.01 / 2.66	-0.13 / 2.60	0.16 / 3.36
May	0.22 / 2.32	0.16 / 3.36	0.15 / 3.38
Jun	0.41 / 2.95	0.15 / 3.38	-0.03 / 3.05
Jul	0.42 / 2.93	-0.03 / 3.05	-0.21 / 2.89
Aug	0.29 / 2.74	-0.21 / 2.89	-0.32 / 2.66
Sep	0.21 / 2.66	-0.32 / 2.66	-0.20 / 3.18
Oct	0.16 / 2.41	-0.20 / 3.18	-0.14 / 3.12
Nov	0.32 / 2.83	-0.14 / 3.12	-0.09 / 2.48
Dec	0.27 / 2.80	-0.09 / 2.48	-0.47 / 2.50

TABLE 24. Prediction scores (anomalous correlation / RMS error) for *seasonal* mean Helsinki precipitation using SLP, SST, and the local monthly mean precipitation at the time when the forecast was made as predictors. Calibration period = 1900-1959, evaluation interval=1960-2000. The benchmark scores are given in the lower part. (Monte Carlo 5% confidence estimates for correlation is around 0.40)

start	lag 0	lag 1	lag 2
Jan	0.48 / 1.81	0.01 / 2.28	0.08 / 2.04
Feb	0.56 / 1.73	-0.06 / 2.01	-0.10 / 2.21
Mar	0.52 / 1.62	-0.10 / 2.20	-0.06 / 2.74
Apr	0.35 / 1.88	-0.23 / 2.72	-0.17 / 3.44
May	0.29 / 2.56	-0.05 / 3.49	-0.07 / 2.97
Jun	0.57 / 2.79	0.14 / 2.99	-0.15 / 3.97
Jul	0.55 / 2.55	-0.01 / 3.54	-0.02 / 3.55
Aug	0.47 / 3.17	-0.06 / 3.47	-0.39 / 3.53
Sep	0.60 / 2.72	0.06 / 3.58	-0.02 / 2.70
Oct	0.76 / 2.29	0.13 / 2.58	0.04 / 2.68
Nov	0.49 / 2.43	-0.21 / 3.07	-0.33 / 2.56
Dec	0.56 / 2.23	-0.25 / 2.38	-0.12 / 2.28
benchmark			
Jan	0.11 / 2.02	-0.27 / 2.24	-0.11 / 1.95
Feb	0.09 / 2.04	-0.11 / 1.95	-0.18 / 2.12
Mar	0.28 / 1.79	-0.18 / 2.12	-0.07 / 2.85
Apr	0.20 / 1.93	-0.07 / 2.85	-0.02 / 3.60
May	0.29 / 2.52	-0.02 / 3.60	-0.05 / 3.03
Jun	0.33 / 3.21	-0.05 / 3.03	-0.21 / 3.59
Jul	0.43 / 2.70	-0.21 / 3.59	-0.11 / 3.57
Aug	0.20 / 3.31	-0.11 / 3.57	-0.14 / 3.64
Sep	0.21 / 3.33	-0.14 / 3.64	-0.04 / 2.69
Oct	0.26 / 3.33	-0.04 / 2.69	-0.04 / 2.62
Nov	0.26 / 2.42	-0.04 / 2.62	-0.19 / 2.23
Dec	0.24 / 2.35	-0.19 / 2.23	-0.27 / 2.24

Bibliography

Benestad, R.E., 1999a. *Evaluation of Seasonal Forecast Potential for Norwegian Land Temperatures and Precipitation using CCA*. Klima 23/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E., 1999b. *S-mode and T-mode EOFs from a GCM modeller's perspective: Notes on the linear algebra*. Klima 24/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E., 2000. *Analysis of gridded sea level pressure and 2-meter temperature for 1873-1998 based on UEA and NCEP re-analysis II*. KLIMA 03/00. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E., 2001 (May 30). *On seasonal forecasting for Fennoscandia: Evaluation of Empirical Model Hindcasts*.

Benestad, R.E., 2002. *Solar Activity and Earth's Climate*. Berlin and Heidelberg: Praxis-Springer.

Benestad, R.E., & Tveito, O.E., 2002. *A survey of possible teleconnections affecting Fennoscandia*. KLIMA 11/02. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Bhatt, U.S., Alexander, M.A., Battisti, D.S., Houghton, D.D., & Keller, L.M., 1998. Atmosphere-Ocean Interaction in the North Atlantic: Near-Surface Climate Variability. *Journal of Climate*, 11, 1615–1631.

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.

Ellner, Stephen P., 2001. Review of R, Version 1.1.1. *Bulletin of the Ecological Society of America*, 82(2), 127–128.

Feldstein, S.B., 2000. The Timescale, Power Spectra, and Climate Noise Properties of Teleconnection Patterns. *Journal of Climate*, 13, 4430–4440.

Førland, E., & Nordli, P.Ø., 1993. *Autokorrelasjon i nedbør og temperatur*. Klima 11/93. DNMI.

Førland, E.J., Benestad, R.E., Hanssen-Bauer, I., Mamen, J., & Smits, J., 1999. *Seasonal Forecasting for Norway*. KLIMA 29/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway. (In Norwegian).

Gentleman, R., & Ihaka, R., 2000. Lexical Scope and Statistical Computing. *Journal of Computational and Graphical Statistics*, 9, 491–508.

IPCC., 2001. *IPCC WGI THIRD ASSESSMENT REPORT*. Summary for Policymakers. WMO.

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.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Wollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., & Joseph, D., 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, 77(3), 437–471.

Lorenz, E., 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.

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.

Skeie, P., & Kvamstø, N.G., 2000. Atmospheric response to variations in the Labrador Sea ice-cover as indicated by ensemble simulations with ARPEGE. *Pages 135-138 of:* Iversen, T., & Høiskar, B.A.K. (eds), *RegClim*. General Technical report, no. 4. http://www.nilu.no/regclim/: NILU.

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.

Sutton, R.T., & Allen, M.R., 1997. Decadal predictability of North Atlantic sea surface temperature and climate. *Nature,* **388,** 563–567.

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