Evaluation of Seasonal Forecast Potential for Norwegian Land Temperatures and Precipitation using CCA.

R.E. Benestad

DNMI, September 30, 1999

Contents

1 Background

The main objective of this report is to make an assessment of seasonal forecast prospects for Norwegian temperatures. This report will also serve as a documentation of the pilot study on seasonal prediction carried out at the Norwegian Meteorological Institute (DNMI), and will therefore refer to scripts used in this study. A brief overview of different methodologies will also be presented. A brief description and evaluation of current seasonal forecasting techniques for the Nordic region is also presented by *Førland* et al. (1999).

Although people have made prophecies and forecasts in ancient times, it is only relatively recently that seasonal forecasting has involved scientific methods[1]. The earliest forecasts have been made with statistical models, however, more recently state of the art weather and seasonal predictions have been based on comprehensive and complex physical and dynamical models. Below is an outline of commonly used forecast methodologies, ranked from the simplest techniques to the more advanced method.

[1]Here, we take "scientific methods" to mean methods that fulfill two criteria: i) The experiments/results can be repeated and tested; ii) The results are derived from objective methods.

1.1 Hierarchy of Forecast methodology

i) Climatology, i.e. the January months are cold in Norway while the summers are mild.

ii) Persistence. Both with and without a decay time based on the autocorrelation. Lagged correlation. *Kim & North* (1998) defined an optimal filter which provides a weighting over the predictor combination that yields the minimum least mean-square-errors.

iii) Analogue methods and climate optimals: predictions based on a historical data base (*Barnston*, 1995; *Zorita & von Storch*, 1997).

iv) Classification and cluster analysis (*Vautard* et al., 1999; *Zorita & von Storch*, 1997).

v) Singular Spectrum Analysis (SSA) (*Vautard* et al., 1999; *Ghil & Yiou*, 1996; *Jiang* et al., 1995) and cross-spectral analysis (coherence). SSA finds periodical oscillations and coherence can be used to find oscillations with common frequencies but at different phase. The SSA method bears similarities to extended EOFs.

vi) Principal Oscillation Patterns (POPs): Identify cyclic patterns: $p_r \rightarrow p_i \rightarrow -p_r \rightarrow -p_i \rightarrow p_r$ (*Balmaseda* et al., 1994; *Hasselmann*, 1988).

vii) Complex EOFs: Identify patterns with a fixed phase relationship, such as wave propagation, and can be used to predict the trajectory of a weather system once the initial state is given.

viii) Linear models, $y_{t+l} = \psi x_t$ where the coefficients (model), ψ, are estimated by multivariate regression (MVR), Canonical Correlation Analysis (CCA) or Singular Vector Decomposition (SVD) (*Johansson* et al., 1998; *Zorita & von Storch*, 1997; *Barnston* et al., 1994; *Barnston*, 1995; *Barnston & C.F.Ropelewski*, 1992).

ix) Neural networks: non-linear method for prediction based on past observations (*Zorita & von Storch*, 1997).

x) Coupled hybrid models employing both statistical and dynamical/physical models. Popular in El Niño Southern Oscillation (ENSO) prediction (*Balmaseda* et al., 1994).

xi) Comprehensive coupled dynamical and physical models (general circulation models, or GCMs). For example ECMWF seasonal forecasts (*Stockdale* et al., 1998; *Anderson*, 1995; *Palmer & Anderson*, 1994).

In this report, seasonal hindcasts using linear statistical models, under category *viii*, will be discussed, and we have chosen to examine the CCA technique here. *Benestad* (1999) demonstrated that there was insignificant differences between predictions using a CCA and MVR models at zero lag, and the choice of method is therefore not important. *Vautard* et al. (1999), however, argued that other techniques, such as multichannel singular spectrum analysis (MSSA), may be superior to CCA, implying that the results presented here may not necessarily represent the theoretically best achievable predictions. We will, however, stress that it is important to employ a wide range of different techniques, as the different methods may have different shortcomings and different strengths. A suite of distinct forecast methodologies can therefore give a better estimate of the forecast uncertainties that also to some degree take into account model errors.

Lorenz (1963) demonstrated that chaotic flow, such as atmospheric circulation, although deterministic, may not be predictable over more than a few cycles. For the atmosphere, the limit of predictability is generally assumed to be about two weeks, which is shorter than seasonal time scales. It is therefore important to emphasis that by seasonal forecasting, we do not mean the exact state of the atmosphere at a certain point in time and space, but rather the weather statistics over a given period. In other words, seasonal prediction assumes that certain boundary conditions, such as SST, may bias the phase space trajectory of the climate system, so that the *probabilities* of the occurrence of weather regimes are altered. It is possible that seasonal forecasts cannot predict the *exact* weather statistics over a given time, but only *the most likely* statistical properties. In this case, seasonal prognoses only have merits when given as a probability distribution, rather than just one number. The CCA method doesn't explicitly produce probabilistic forecasts, and we have therefore taken the predictands directly from the model without estimating the probability distribution for the predictions just for evaluation purposes. It is nevertheless possible to quantify an estimate of probability distribution functions by assuming a Gaussian, whose parameters, σ and μ, can be estimated from the past prediction errors.

We will make a distinction between "forecast" and "hindcast", by using the term forecast in conjunction with a prediction of the probability distribution for a climate parameter for a given time for which there is no observation when the forecast is made. Hindcast will henceforth refer to the cross-validation (*Wilks*, 1995, p.194-198) predictions according to a deterministic model, such as $y_{t+l} = \psi x_t$, for which the real value which is predicted is already known and subsequently used in the model validation. Cross-validation analysis essentially means that the predicted value was not used in the model calibration. The term "prediction" will have a more loose mean-

ing referring to the model output, i.e. y_{t+l} in our linear model $y_{t+l} = \psi x_t$. We will henceforth use "prediction" in the meaning forecasts, hindcasts, and model reconstruction of data which also have been included in the calibration analysis.

The term "lag" and "lead time" (l) will be used interchangeably, and will have the same meaning here. The lead time refers to the time lag between the last predictor month and the first predictand target month. For monthly mean forecasts, this implies that a forecast with 1 month lead time makes a prediction for the subsequent month. A seasonal forecast with 1 month lead time makes a prediction of the mean temperature for a 3-month period of which the first month is the month following the predictor month. For instance, a 1-month lead time seasonal prediction starting in January gives a prediction for the February to April (FMA) period. A zero lead time, starting in January, gives a forecast for the January-March (JFM) season, and can be used to estimate the best guess for the subsequent 2 months. In this case, the January predictands may be known, and thus a seasonal prediction with 0-month lead time can be used to predict the mean quantity over the 2 next months. In a sense, the seasonal forecasts implicitly take persistence into account, as past and present observations can be subtracted from the predictands in order to produce a forecast for the remaining time.

The term "predict*or*" will refer to the input (starting) data for the predictive models, x, whereas "predict*and*" are the predicted quantities (model output), y.

1.2 Commonly used Predictors in Seasonal Forecasting

a) Astronomical data (Sun spots). The merits of astronomically based predictions are still questionable. These kinds of forecasts have been around for a long time, for example H.C. Willett (*Godske*, 1956, p.189) and Sir N. Lockyer (late 19th century) have tried to base predictions on sun spots.

b) Past temperatures and past rainfall (persistence forecasts).

c) Sea Surface Temperatures (SSTs): Strong impact on temperature and rainfall patterns in the tropics, but weaker influence in the mid-latitudes (*Colman & Davey*, 1999).

d) Sea Level Pressure (SLP). Captures the state of North Atlantic Oscillation (NAO)

e) Geopotential heights: Captures the state of NAO.

f) Ocean analysis (assimilation) including oceanic subsurface data for coupled model initiation (in ENSO forecasts).

g) Climatic indices such as the Southern Oscillation Index (SOI) and the North Atlantic Oscillation Index (NAOI).

1.3 Other possible Predictors for Seasonal Forecasting

1) Snow extent (*Watanabe & Nitta*, 1999) and ice cover.

2) Solar activity (Length of the solar cycle).

2 Description of observational data sets

2.1 The predictors

The observations discussed in connection with model validation were taken from the NMC (NCEP) ds195.5 data set (monthly mean sea level pressure, SLP, 500hPa geopotential heights, Φ_{500}, and 500hPa temperatures, T_{500}, and the GISST2.2 data sets (sea surface temperatures, SSTs). Although the NCAR ds010.0 and UEA data sets contained longer observational SLP records, the NMC data was preferred to the others as it had better coverage over the Arctic. A comparison between the NCAR, UEA, and NMC SLP suggested that the main features in the different data sets were similar (*Benestad*, 1998b). *Jones* (1987) reported a systematic bias in the Arctic SLP before 1930. Because of lack of good observations, early Arctic SLPs have questionable quality and may introduce errors in the analysis.

The NMC SLP and Φ_{500} records spanned the 1946-1994 period and the "bad" Arctic SLP data were not included in this analysis. The GISST2.2 data set contained SST observations for the 1903-1994 period and sea-ice extent (1949-1994). The data sets and their time spans are listed in table 1. Further description of the data sets is also given in *Benestad* (1998b).

2.2 The predictands

The historical data used as predictands were obtained from the DNMI archives.

Table 1: Data set sizes

Data set	Number of years	Period
NMC ds195.5 SLP	49	1946 - 1994
NMC ds195.5 Φ_{500}	49	1946 - 1994
GISST2.2 SST	92	1903 - 1994
GISST2.2 Sea Ice	46	1949 - 1994

3 Methods

3.1 Statistical predictions

Some GCMs do not capture the westward propagation of low frequency planetary waves between 40 and 60°N (*Doblas-Reyes* et al., 1998), which may have implications for long-range and seasonal forecasts. Statistical methods, on the other hand, may perhaps capture some aspects of these waves (i.e. complex EOFs), and hence utilize this information in the forecasts.

Kushnir & Held (1996) discussed atmospheric model sensitivities to North Atlantic SST anomalies (SSTAs), arguing that the influence of SSTs on the atmosphere is obscure. Most dynamical models suffer from incomplete knowledge of the ocean-atmosphere heat fluxes, and as a results, SST impact studies with different models give conflicting results. Empirical models, on the other hand, are less prone to model misrepresentations despite limitations regarding the assumption that SST-atmospheric relationship is linear and stationary. *Colman & Davey* (1999) reported some predictability of European summer temperatures due to North Atlantic SSTs.

Mullan (1998) proposed that small scale mid-latitudinal SSTAs may be a result of atmospheric forcing rather than being an active forcing agent for subsequent land temperatures. *Johansson* et al. (1998) found little extra skill in CCA hindcasts of North European temperatures when SSTs were included, and their optimal model was constructed with 700hPa geopotential heights and past temperatures. *Vautard* et al. (1999), on the other hand, demonstrated some predictive skills over the US for SST empirical models.

Benestad (1998a) demonstrated that it is crucial to find the optimal predictor (EOF) combination for best predictions. The predictor combinations for the results shown here were chosen through a screening test (*Benestad*, 1998a; *Wilks*, 1995), where the contribution of each EOF was tested in a cross-validation analysis. Only the EOFs which gave a positive contribution towards the correlation score were included in the prediction models. Different EOF combinations were used for different regions and different lead times.

Only a small number of predictions were carried out here, as the number of combinations of different hindcasts with different options is too large to cover all possibilities here. For instance, predictions with only four different initial values were examined, January, April, July and October. The exclusion of the other months as predictors implies that the scores here represent a lower boundary of actual maximal achievable scores. In other words, a one-month lead time hindcast of July starting in June may give a better prediction than a 3-month lead time forecast starting in April.

3.2 Implementation of hindcasting

The hindcast experiments were made using the Matlab codes (version 5.2), *hcst.m*, *optmod.m*, *crossval2.m*, *crossval3.m* and *residuals.m*, which were also used to construct and evaluate statistical downscaling models in *Benestad* (1998a), *Benestad* (1998c) and *Benestad* (1999). These Matlab programs were developed at the Norwegian Meteorological Institute (DNMI), and were designed as universal tools for making experimental hindcasts as well as constructing forecast and downscaling models. The predictors are assumed to be in a gridded field format, read from *netCDF* formatted data files, and the predictands are taken to be individual time series from a number of locations, also assumed to be stored in the *netCDF* format. The codes, *hcst.m* and *optmodX.m*, have a number of control flags which sets a number of options, listed in table 2.

Table 2: Control flags for *optmod.m*

Flag	Function
method	Flag: 1. CCA, 2. SVD, 3. Multivariate Regression (MVR)
lag	Time lag in months.
option	Set option of: 1. ordinary, 2.use odd/even years only, 3. Exclude adjacent points from model calibration, 4. 2+3
prewhit	Flag=1 if pre-whitening before analysis
seasons	Flag=1 if use 3-month seasons.
plotts	Flag=0: only plot the weights and map.
same	flag=1 to use same predictands as last time when number > 25
fcst	Flag=1 for forecast mode: use the residual to determine the probability distribution.

The hindcast tools (*hcst.m*, *optmod.m*, *crossval2.m* and *crossval3.m*) have been used to produce all the figures in this report. The Matlab routines

mapfig.m and *stationmap.m* were used to plot the results and draw the maps of the sites of the predictands.

3.2.1 How to interpret the prediction scores

All the prediction scores quoted here are taken from a cross-validation analysis, which predicts independent data not used in the model calibration. Therefore, negative correlation scores indicate serious misses, for instance that the model predicted warm conditions while cold anomalies occurred. Correlation scores lower than 0.40 may not necessarily indicate real skill; the skill score may vary randomly between zero and 0.40 (due to sampling fluctuations). Correlation scores greater than 0.40 may, however, indicate a connection between the large scale circulation and the local climate, and scores greater than 0.60 will be referred as "useful" for predictions (This threshold is arbitrarily chosen). The definition of useful skill, however, does depend on the use of the forecasts, and these limits must not be regarded as universally valid.

Mullan (1998) and *Katz* (1988) argued that autocorrelation can give non-zero correlations at non-zero lags even if there is only a real correlation for lag zero. The apparent correlation at non-zero lags was explained in terms of serial correlation smearing out the signal. A pre-whitening filter was applied to the data in some sensitivity studies in order to reduce the inflationary effects of autocorrelation on lagged correlation analysis (*Katz*, 1988). The pre-whitened analysis was included in an attempt to deduce whether there was a real connection between the predictands and predictors at a given lag, or whether the skills were due to the autocorrelation smearing out the signal. The scores using pre-whitened data are not representative for the prediction skills, as operational predictions utilizes all information that can be extracted from persistence.

Prediction scores can be assessed in terms of pattern or spatial correlation (*Stephenson*, 1997), however, we will only focus on individual stations in order to evaluate the forecast skills for the respective location. Three regions representing different climate types, southern Norway, western Norway and northwestern Norway, have been selected for detailed validation. These locations do not necessary correspond to the stations with highest skill, but were picked partly because the climate at these places affect a large number of people. Only stations with long complete records that include valid data up to 1994 were chosen here.

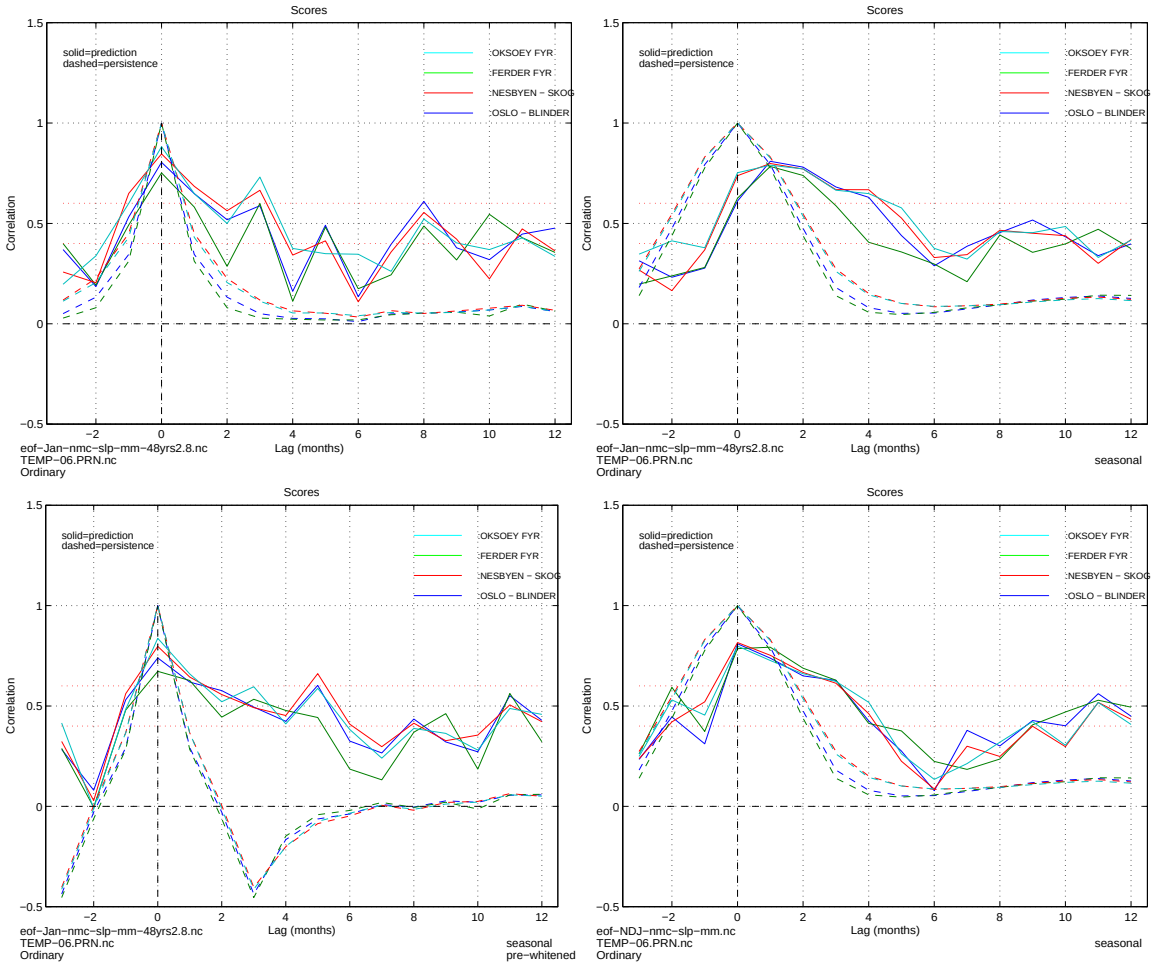

Figure 1: *Skill scores for temperature predictions for southern Norway (Oslo, Nesbyen, Ferder and Oksøy) starting in January. The predictions were based on NMC SLPs. The upper left panel shows scores for prediction of monthly mean temperatures and the upper right panel shows the scores for 3-monthly (seasonal) mean temperatures. The lower left panel shows the prediction of pre-whitened seasonal mean temperatures, whereas the lower right panel shows the scores of seasonal mean temperatures predictions based on seasonal (NDJ) PCA products.*

3.3 Exploring different methods

The best scores were obtained using monthly mean predictor fields to predict seasonally mean (3-monthly) temperatures (figure 1). The persistence in the Norwegian temperatures also indicates greater autocorrelation for the seasonal mean temperatures than monthly mean temperatures, implying greater predictability for the seasonal mean values (upper panels).

Pre-whitened analysis suggests that there also is a correlation between SLP and temperatures at non-zero lags, i.e. that the predictability is not just due to serial correlation in one of the data sets (lower left).

Studies of predictions made with 3-month mean predictor fields suggested that these did not yield much different scores to predictions based on monthly mean predictor fields (lower right). We will therefore focus on the latter type.

4 Temperature predictions based on CCA

Figures 2 to 4 show the locations of the climate stations included in the analysis for southern Norway (Oslo, Nesbyen, Ferder lighthouse, and Oksøy lighthouse), western Norway (Bergen and Ona), and northwestern Norway (Tromsø and Bodø) respectively.

4.1 Sea level pressure hindcasts

4.1.1 Southern Norway

The prediction score based on SLP vary with the lead time and the seasons when the prediction starts (figure 5). The seasonal prediction prospects are good in January, but questionable during the other seasons. There seems to be some marginal skill (correlation higher than 0.60) at seasons other than the winter, but the value of these hindcasts will depend on their use. The peak prediction scores appear during late winter and late summer, at the same times when the autocorrelation in the southern Norwegian temperatures are high (*Førland & Nordli*, 1993). *Vautard* et al. (1999), *Barnston* (1994), and *Johansson* et al. (1998) also reported seasonal variations in the prediction skill scores. A recovery of skill is noted for for 2-month lead hindcasts starting in July.

The inspection of the prediction errors (residuals) after the predictions of the south Norwegian temperatures (figure 6) indicates a less than perfect normal distribution (middle right and lower right). A normally distributed signal is expected to lie on a straight line in a normal distribution test (lower right). Although the serial correlation was low (middle left), the residual of the four sites showed a high correlation. It is assumed that the residuals at each site are random (i.e. Gaussian white noise) and independent of each other if the model captures all the predictable signals in the temperature series. The residual analysis therefore suggests that there may have been a common large scale signal not captured by the model.

The large scale circulation (upper panel) is therefore probably not responsible for most of the temperature variability in southern Norway. However, a NAO like structure was associated with all of the predictability of Norwegian temperatures. The lower left panel of figure 6 shows the predicted and observed temperatures in Oslo. A comparison between the actual temperatures and the residual suggests that the residuals were of approximately same magnitude as the predictions.

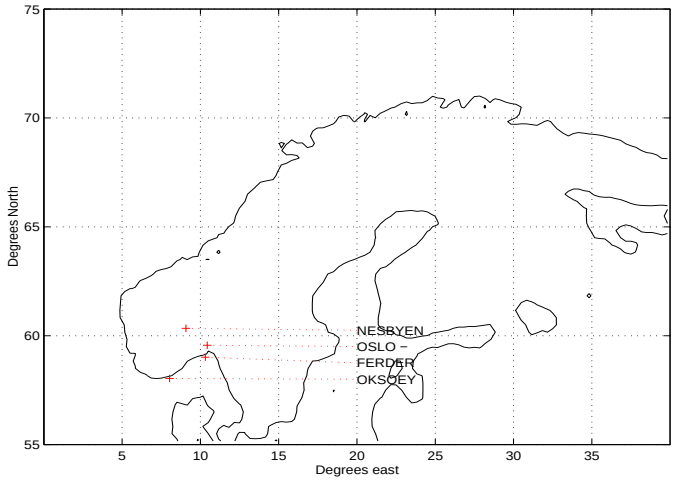

Figure 2: *Map of the land temperature stations in southern Norway.*

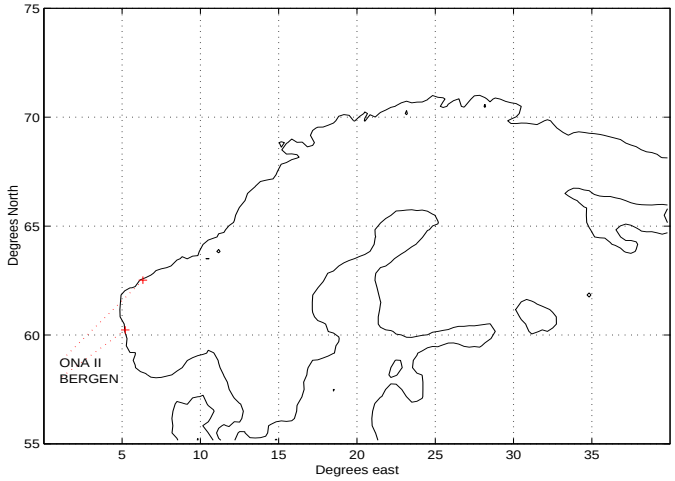

Figure 3: *Map of the land temperature stations in western Norway.*

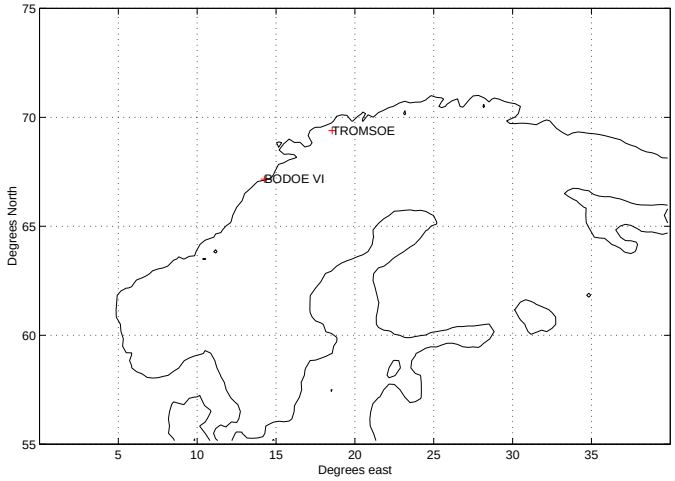

Figure 4: *Map of the land temperature stations in northern Norway.*

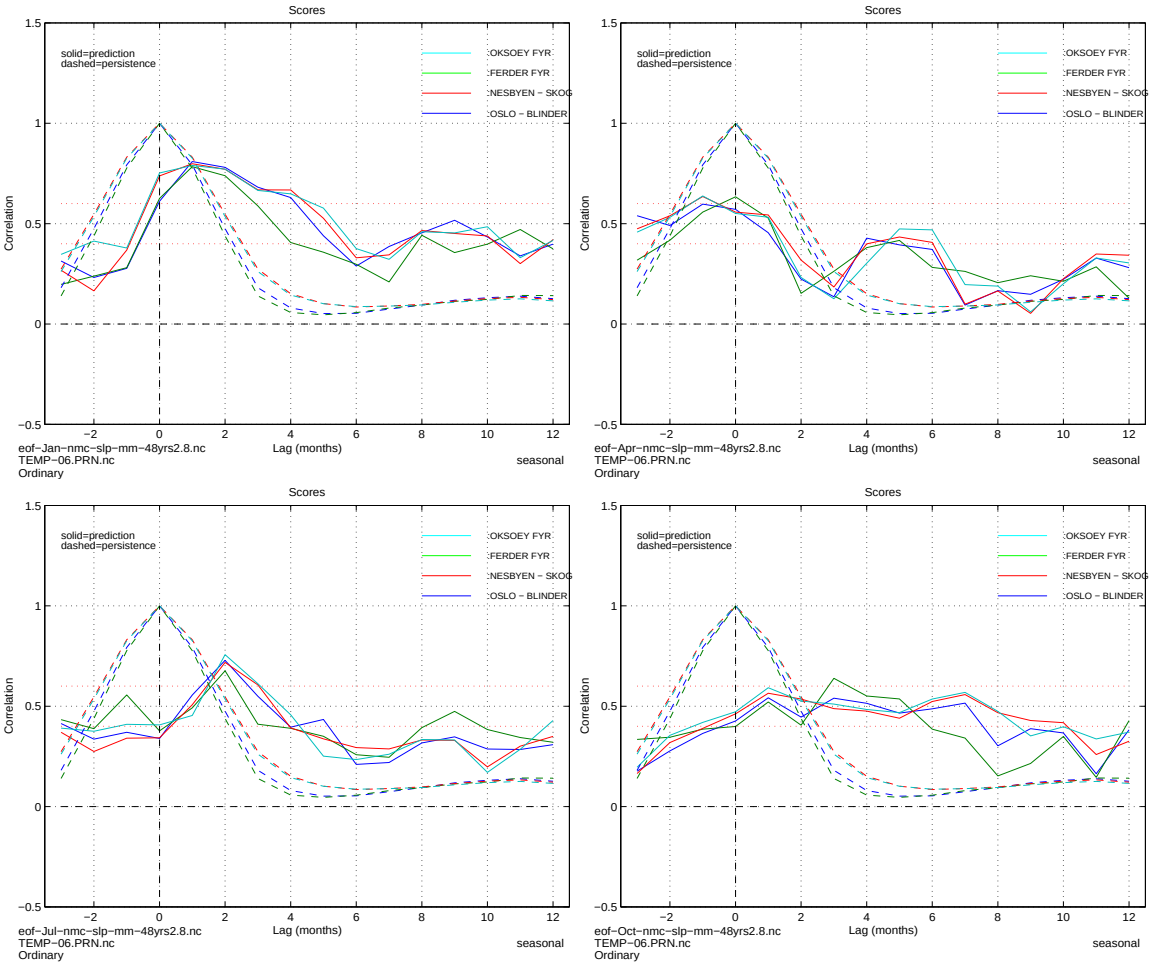

Figure 5: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC SLPs.*

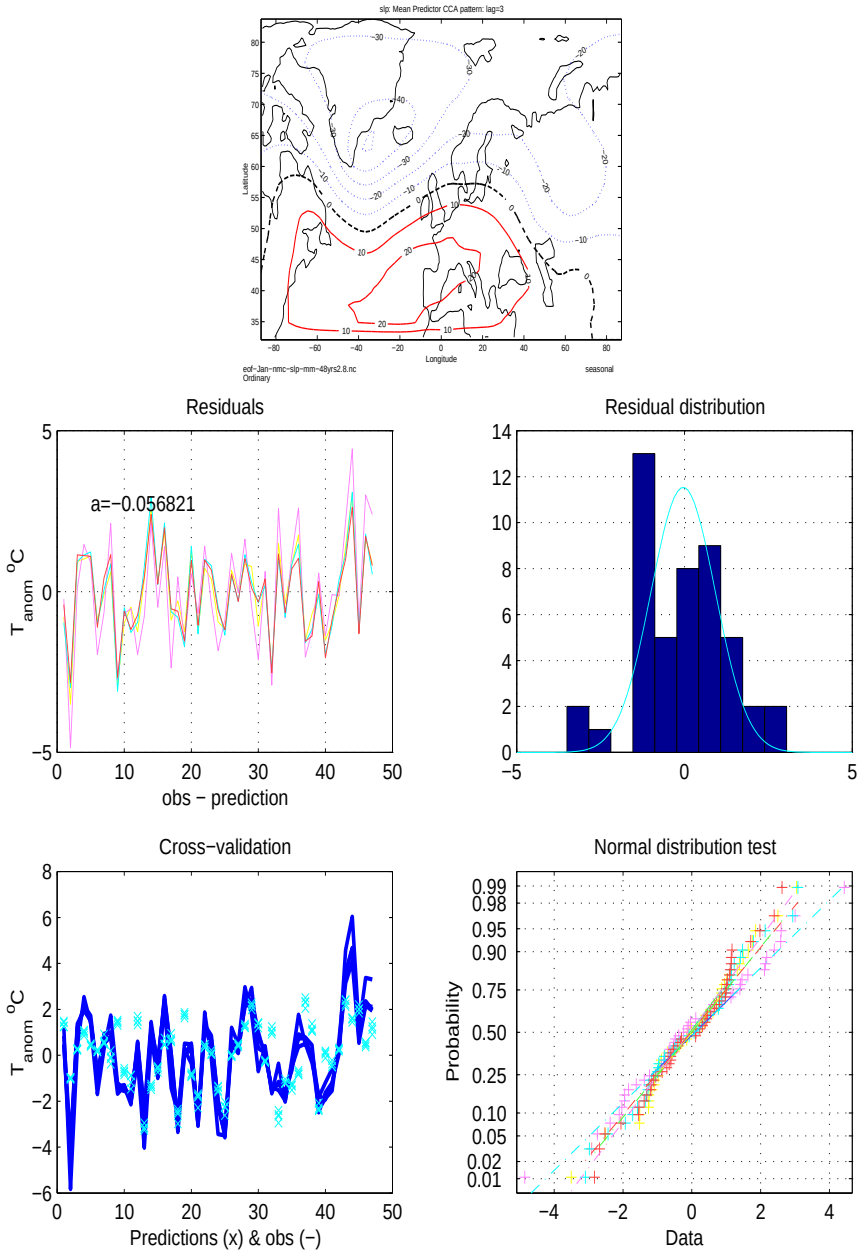

Figure 6: *Example of the leading SLP CCA pattern (left) and the residuals for predictions of the 3-month lead January south Norwegian temperatures. Top: the leading CCA predictor pattern; middle left: the residual time series; middle right: distribution; lower left: predictions and the observations; lower right: normal distribution test.*

4.1.2 Western Norway

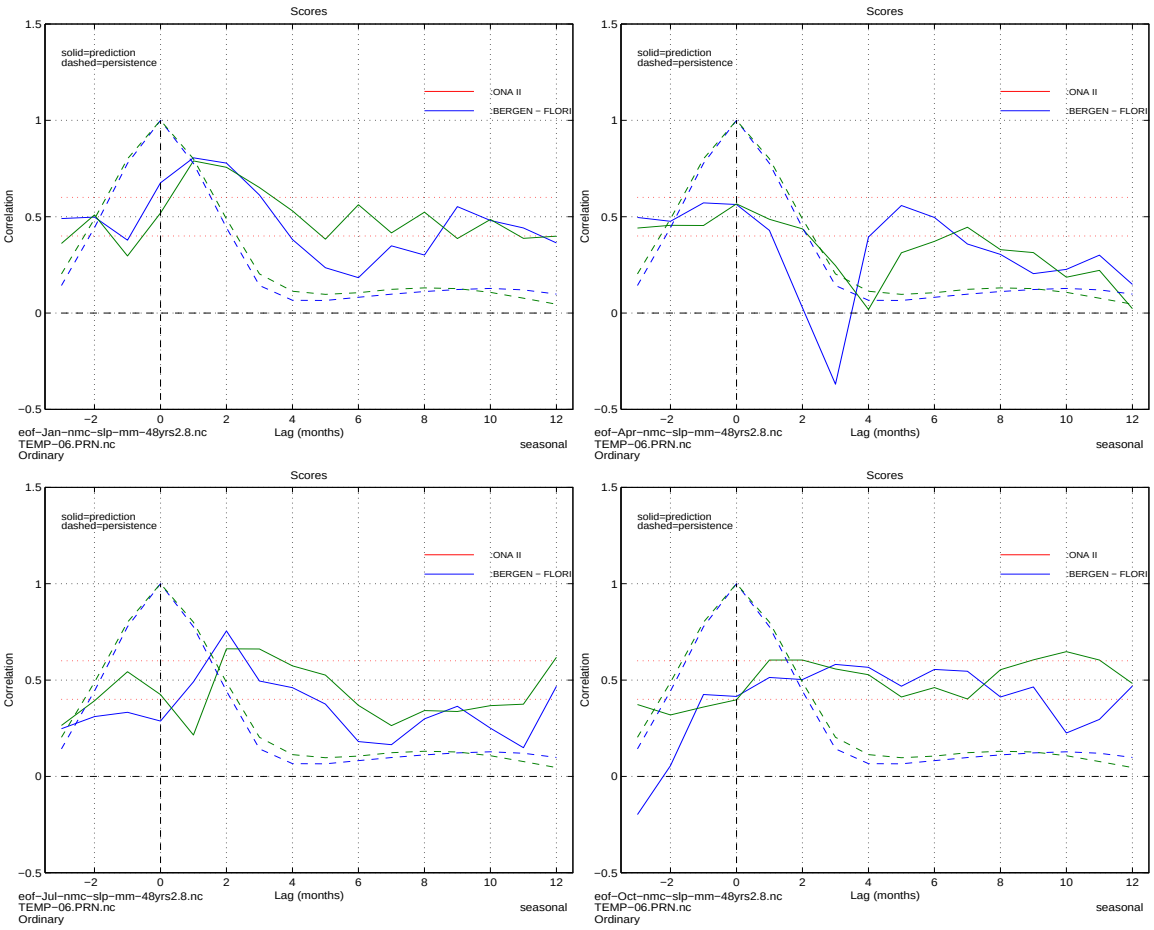

Figure 7: *Skill scores for temperature predictions for Bergen and Ona starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC SLPs.*

The prediction scores for western Norway (figure 7) were roughly similar to those of southern Norway (figure 5), apart from lower spring scores. The best predictions were those starting in January, and the scores for the other seasons were questionable. The negative cross-validation scores seen for the April predictions are indicative of model over-fit, and seasonal temperature predictions with up to 4 months lead time were worse than predictions based on persistence. There were also indications of "recovery of skill" for 2-month lead hindcasts starting in July (target September). The autumn scores were relatively insensitive to the lead time beyond 2 months, although were considered as marginally significant.

4.1.3 Northwestern Norway

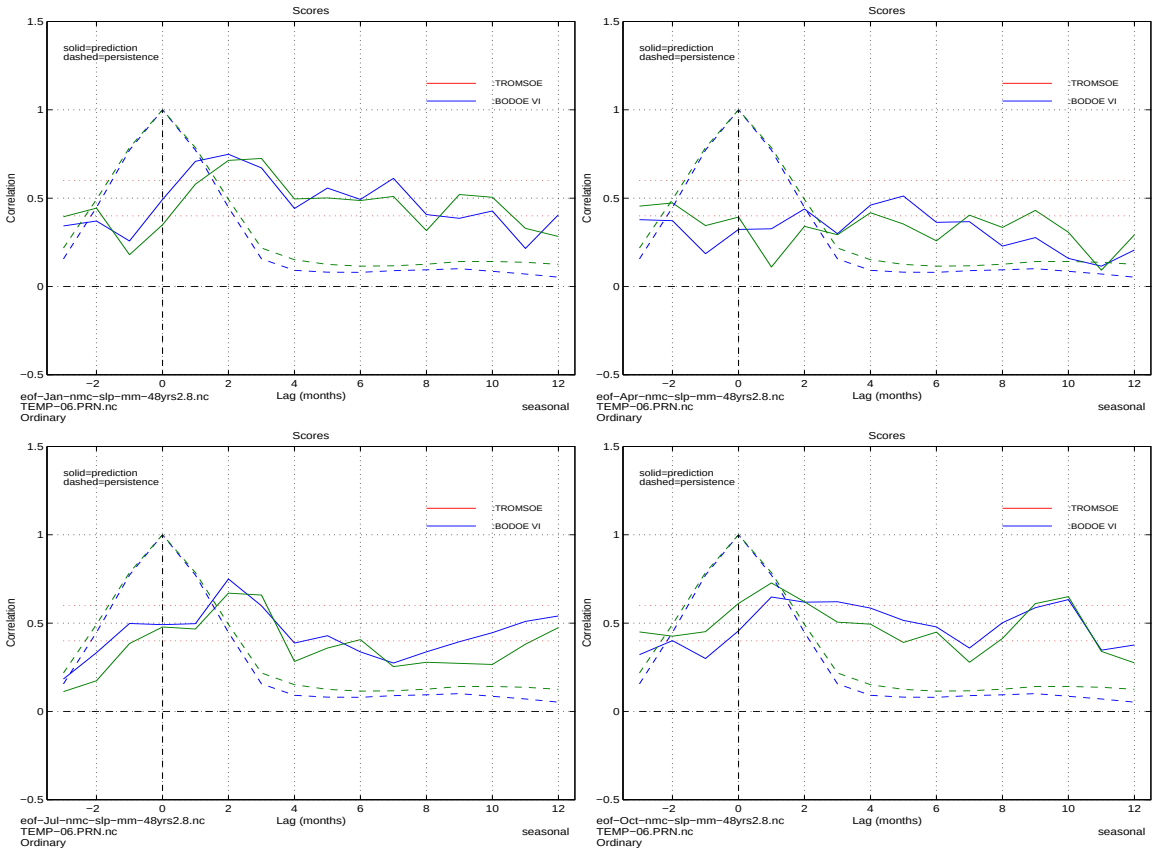

Figure 8: *Skill scores for temperature predictions for Tromsø and Bodø starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC SLPs.*

The prediction scores for northwestern Norway peak at 2-month lead times for hindcasts starting in January and July (figure 8). The spring time predictions (upper right) were associated with poor scores whereas hindcasts starting in October (lower right) were associated with relatively high scores for a wide range of lead times. A slight skill recovery was seen in the October scores at 9-10 month lead time, however, this recovery may be due to statistical fluctuation.

The common feature of the skillful hindcasts in all 3 climate regions was a predictor pattern which resembled the NAO feature. This observation is in line with results from other studies such as *Watanabe & Nitta* (1999) and *Benestad* (1998a). The NAO therefore appears to have a great impact on the Norwegian climate.

4.2 Sea surface temperature hindcasts

4.2.1 Southern Norway

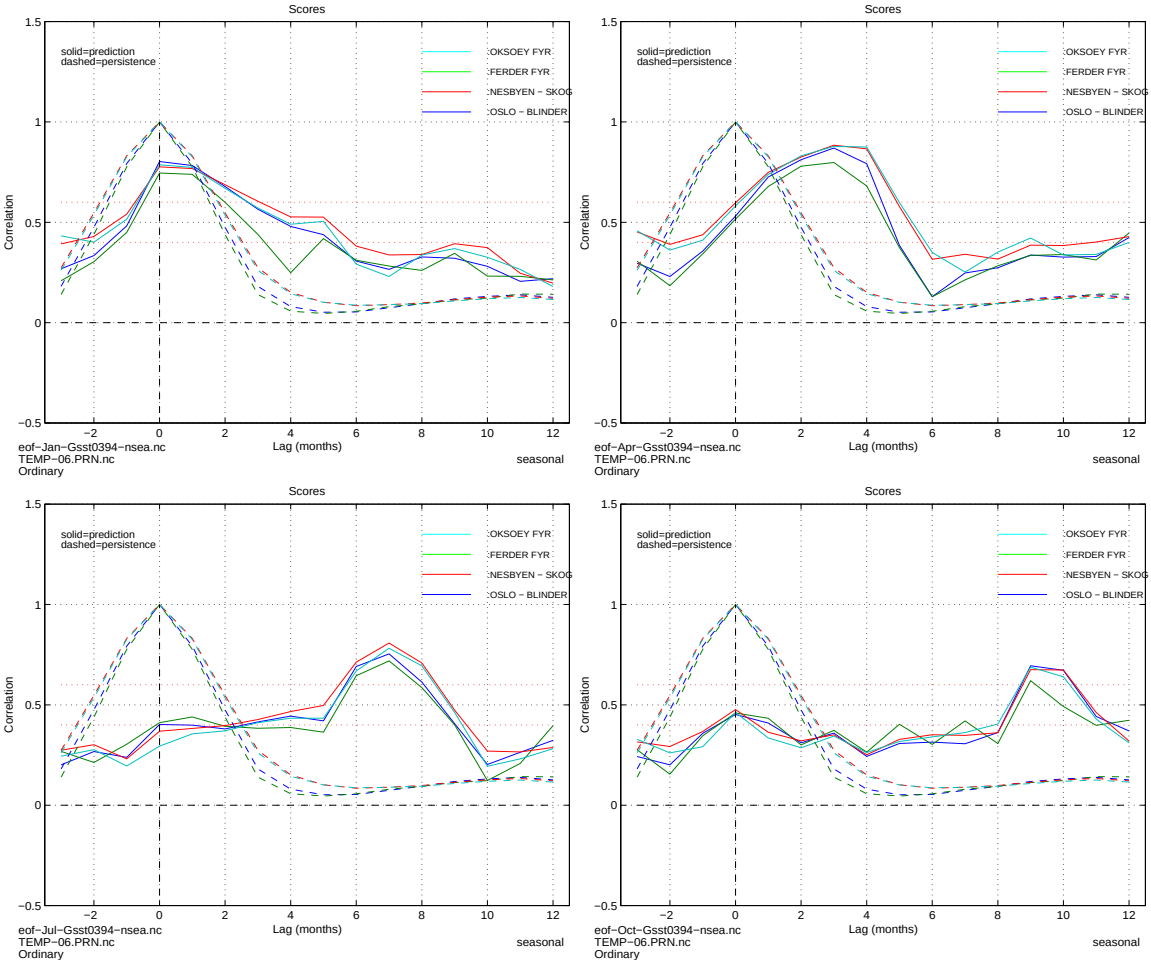

Figure 9: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 SSTs.*

The seasonal predictions based on SSTs from the Nordic Seas (the Norwegian Sea, the North Sea, the Baltic Sea, and the Barents Sea) were in general good for all seasons, except the autumn (figure 9). The predictions starting in January achieved similar skills scores to the corresponding SLP predictions, however, the spring SST predictions gave substantially higher scores for lead times longer than 2 months. The late summer (July and August) had especially high predictability. Hindcasts starting in July gave best predictions

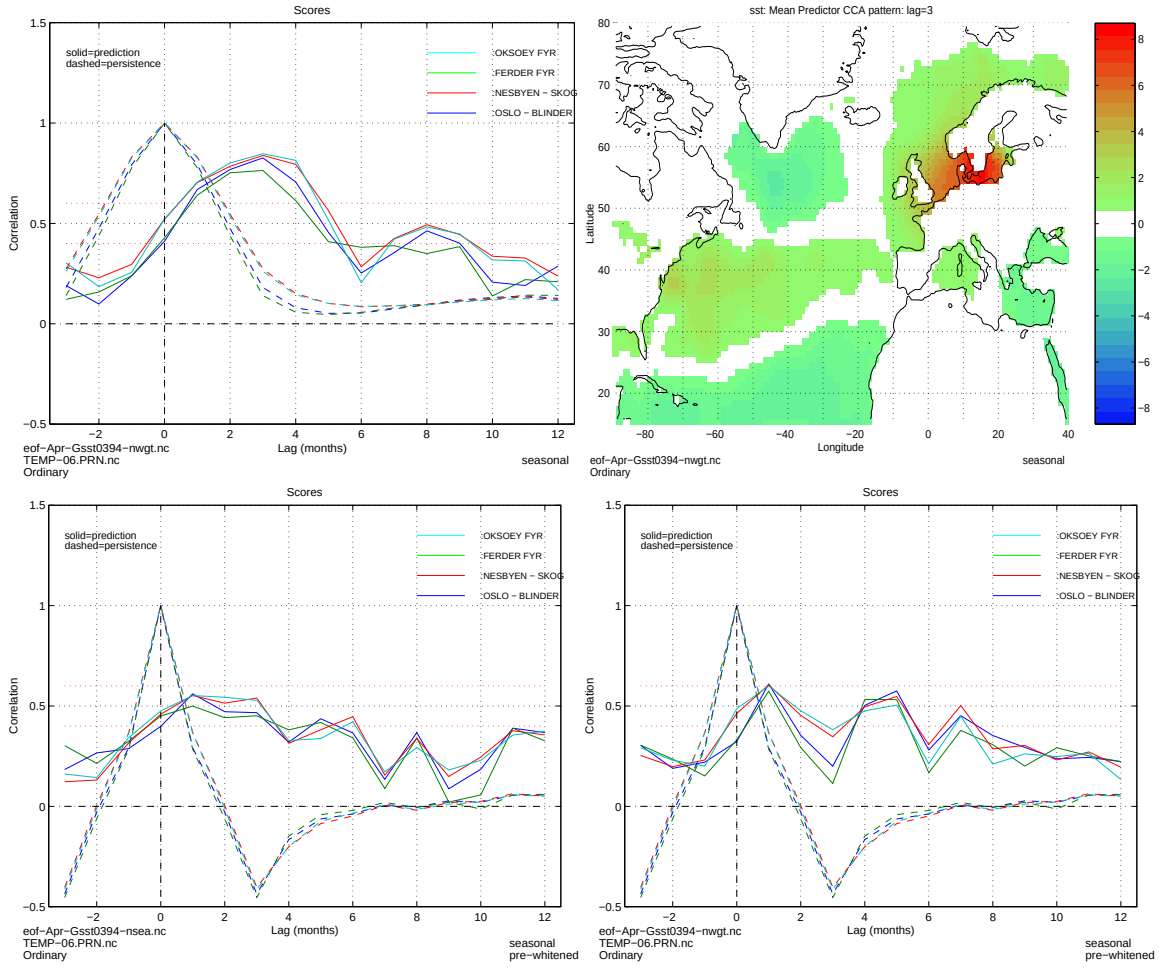

Figure 10: *Upper left: Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in April and based on GISST2.2 SSTs from the entire North Atlantic (upper left). Upper right: the leading CCA SST pattern associated with the maximum scores in upper left panel. Lower: skill scores of the south Norwegian temperatures after pre-whitening based on the Nordic Seas SSTs (left) and North Atlantic SSTs (right).*

for the subsequent February-April period, whereas the autumn predictions hinted at marginal prediction scores for the following summer. Dramatic skill recovery was seen for all hindcasts starting in all seasons but January. Skill recovery has also been reported by *Colman & Davey* (1999), who explained this recovery in terms of the eastward advection of SSTAs across the North Atlantic basin.

The prediction shown here only used SSTs from the adjacent seas, however, *Benestad* (1998a) demonstrated that the leading CCA pattern may be

part of a basin scale North Atlantic SST anomaly. Figure 10 shows various test scores from predictions based on SSTs from the entire North Atlantic, and the results were similar to those of the Nordic Seas SST model. The SSTA predictor pattern with the strongest influence on the Norwegian temperatures was associated with strong amplitudes in the North sea (similar to the leading CCA SST pattern shown in figure 11, upper panel). This regional SST pattern was also consistent with those basin scale patterns identified by *Colman & Davey* (1999), and is shown in the upper right panel of figure 10. The North Atlantic SST pattern is furthermore similar to the pattern which has been associated with the NAO (*Watanabe & Nitta*, 1999; *Deser & Blackmon*, 1993).

The two lower panels in figure 10 also show a comparison between the scores obtained from the Nordic and North Atlantic models. Here the data had been pre-whitened prior to the calibration of the models in order to identify real predictive signals other than pure persistence. The two SST models produced similar score functions, indicating that the local SSTs contained most of the predictive signal. The pre-whitened analysis gave no dramatic skill recovery at about 3-month lag, which may indicate that this recovery of skill may be due to to late spring and early summer SST persistence.

The residuals from the CCA predictions were approximately normally distributed and were associated with insignificant trend (figure 11). The similar time evolution in the residuals between the various sites suggest that there were predictable signals not captured by the SST models. The residuals also had similar magnitude as the predictions, further indicating that the SSTs can only account for some of the land temperature variability.

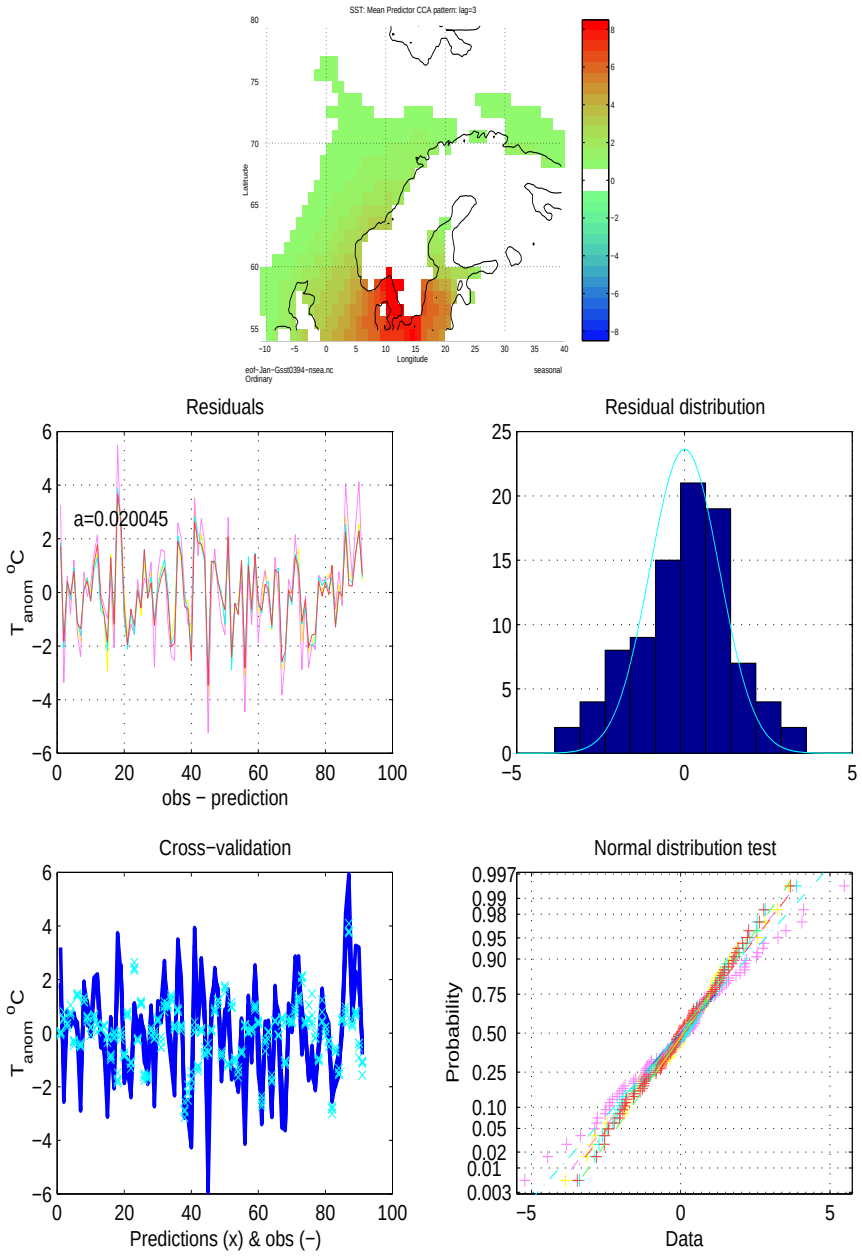

Figure 11: *Same as figure 6, but for the January SST CCA models.*

4.2.2 Western Norway

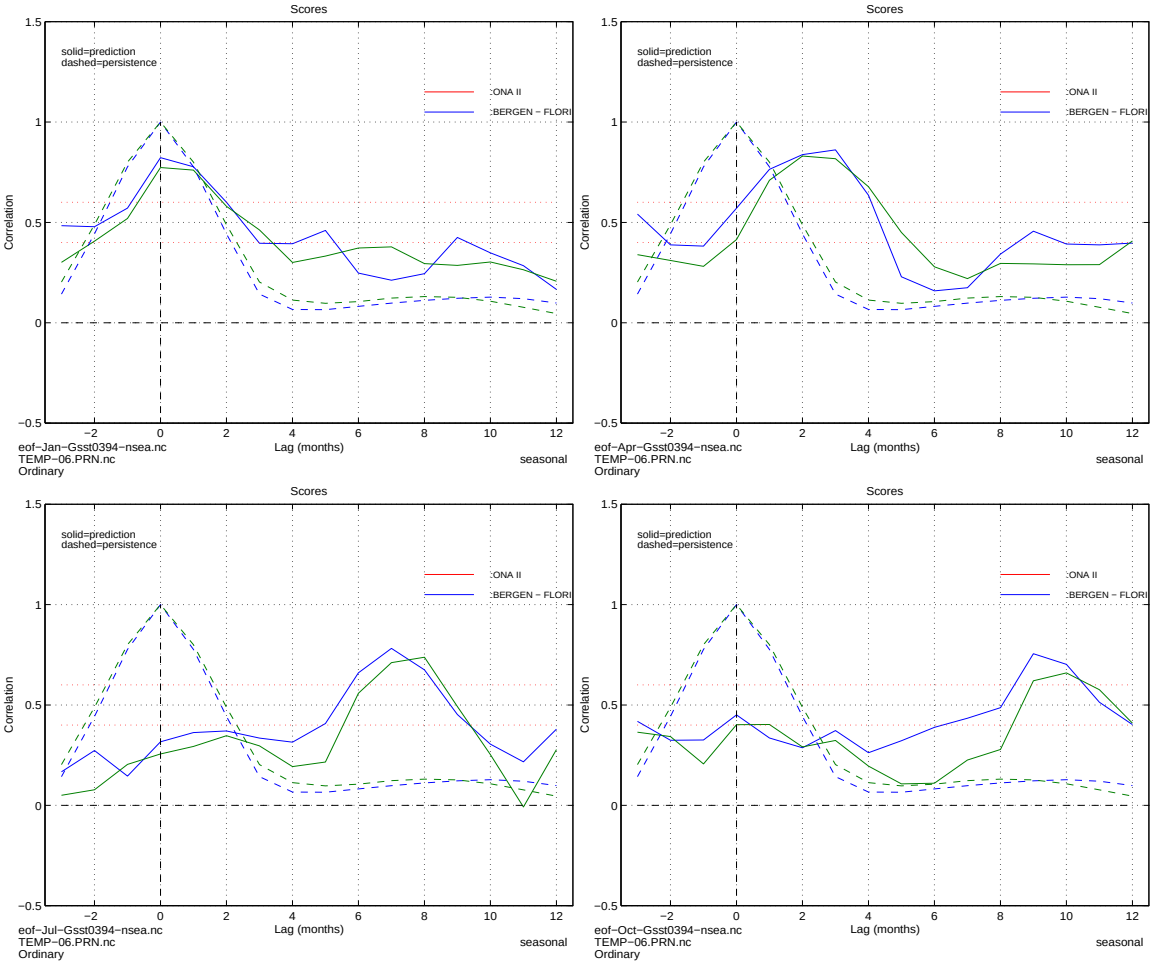

Figure 12: *Skill scores for temperature predictions for Bergen and Ona starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 SSTs.*

4.2.3 Northwestern Norway

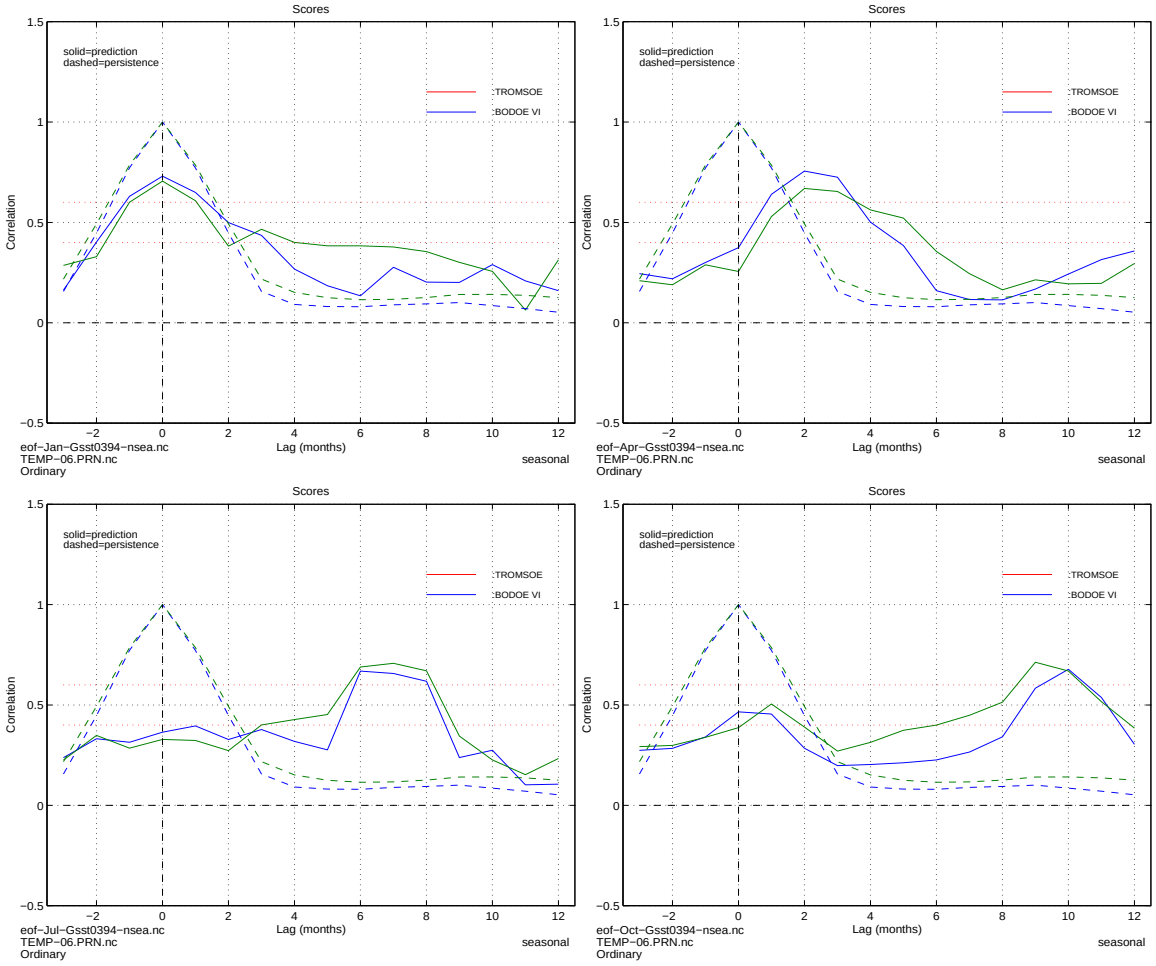

Figure 13: *Skill scores for temperature predictions for Tromsø and Bodø starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 SSTs.*

The prediction scores for the west coast (figure 12) and northwestern temperatures (figure 13) were similar to those of the southern part of Norway (figure 9), suggesting that the large scale North Atlantic SST anomalies have far reaching influence on the local climate. The fact that the temperatures in all different climate regions indicated a skill recovery at 9-month lead time for predictions starting in October may suggest that this peak may be real.

4.3 500hPa diagnostics hindcasts

4.3.1 Southern Norway

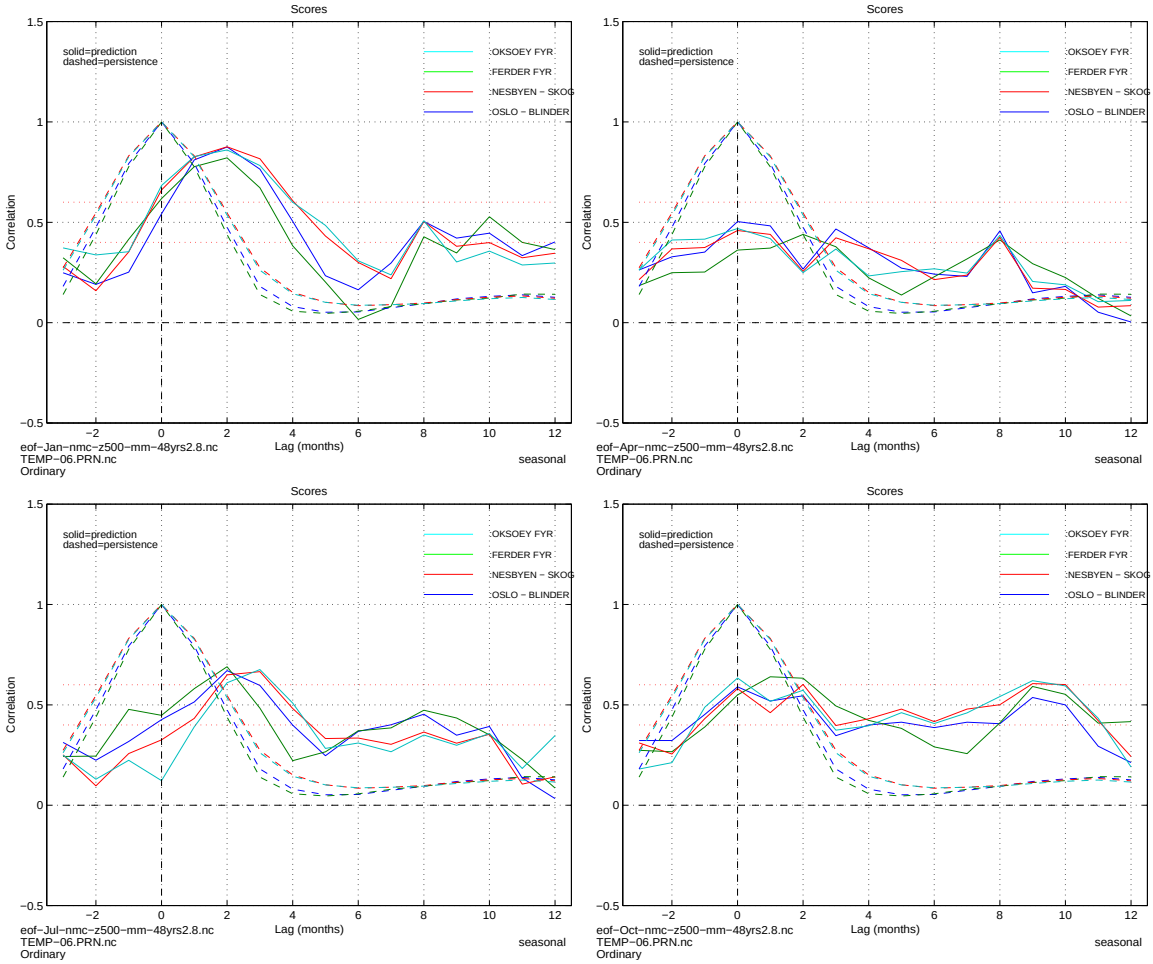

Figure 14: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC 500hPa geopotential heights.*

Seasonal hindcasts based on 500hPa geopotential heights and that start in January have good hindcast prospects for lead times between 0-4 months for the southern part of Norway (figure 14). The skill scores were low for predictions starting in April, but both July and October predictions suggested good prospects for some lead times.

4.3.2 Western Norway

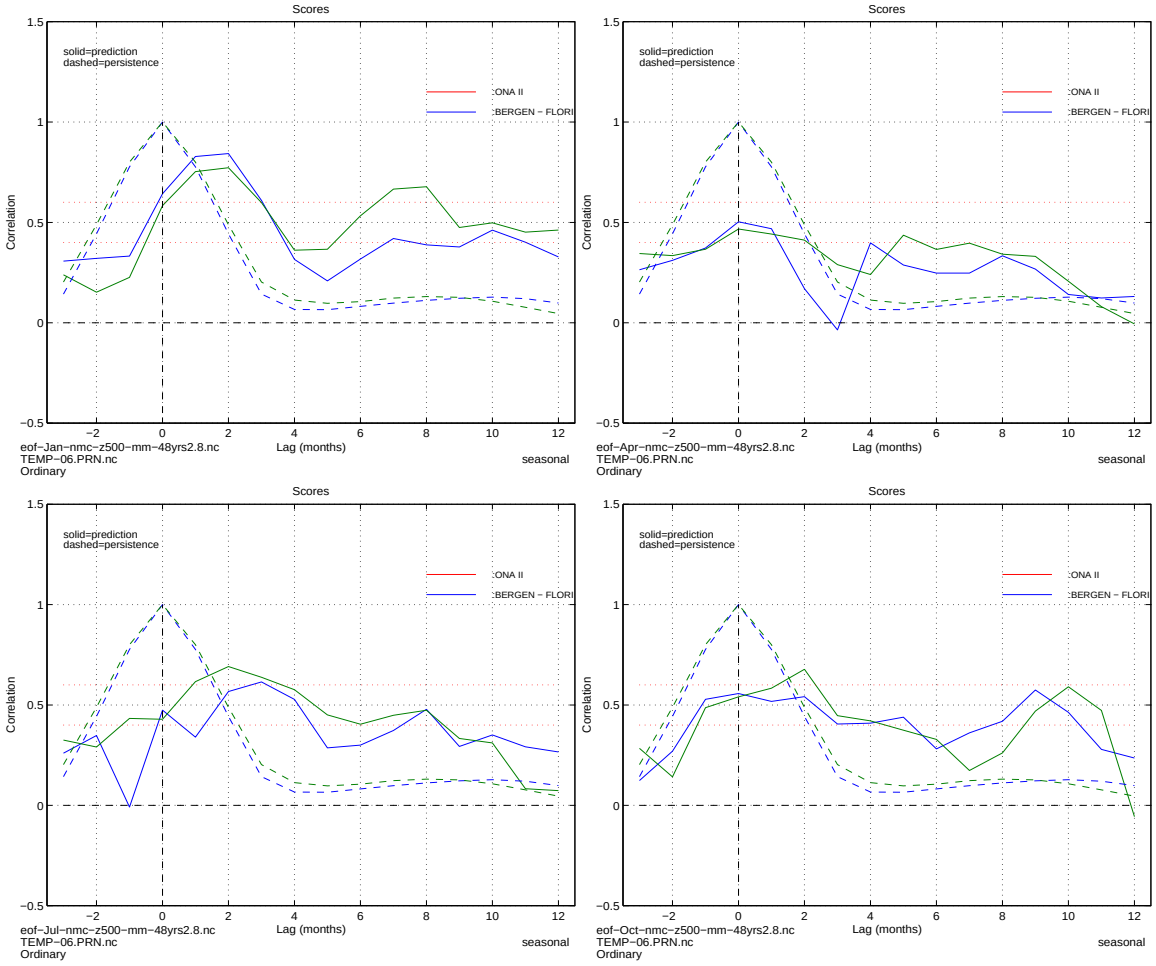

Figure 15: *Skill scores for temperature predictions for Bergen and Ona starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC 500hPa geopotential heights.*

The seasonal prediction prospects for the west coast were similar to those for southern Norway. There was also a more prominent recovery of skill at 8-month lead time for the January prediction.

4.3.3 Northwestern Norway

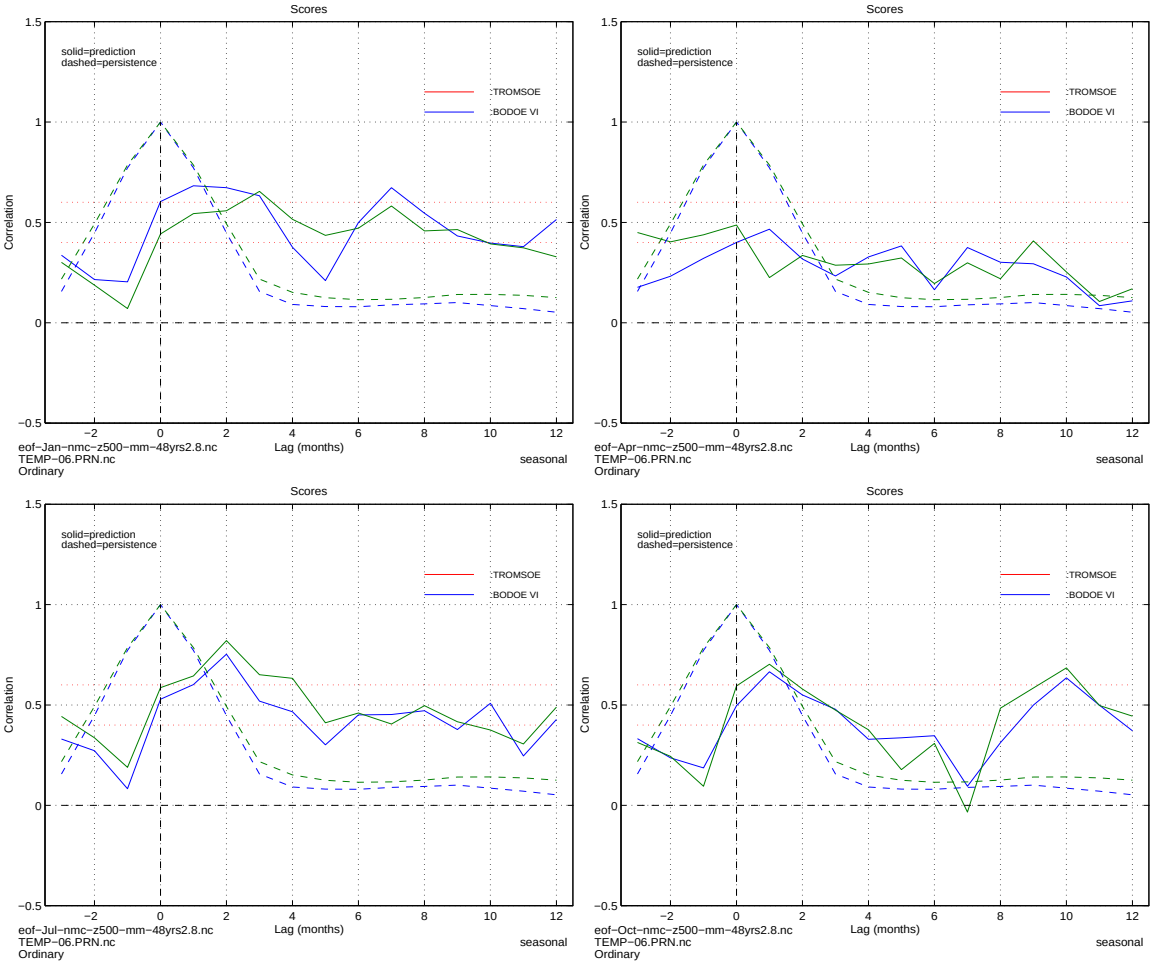

Figure 16: *Skill scores for temperature predictions for Bergen and Ona starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on NMC 500hPa geopotential heights.*

The hindcast skill scores in northwestern Norway suggested greatest predictability for 500hPa geopotential height based hindcasts starting in July and October (figure 16). A pronounced skill recovery was found at 10-month lead time for the October prediction (target month: following July-September).

4.4 Sea-ice hindcasts

4.4.1 Southern Norway

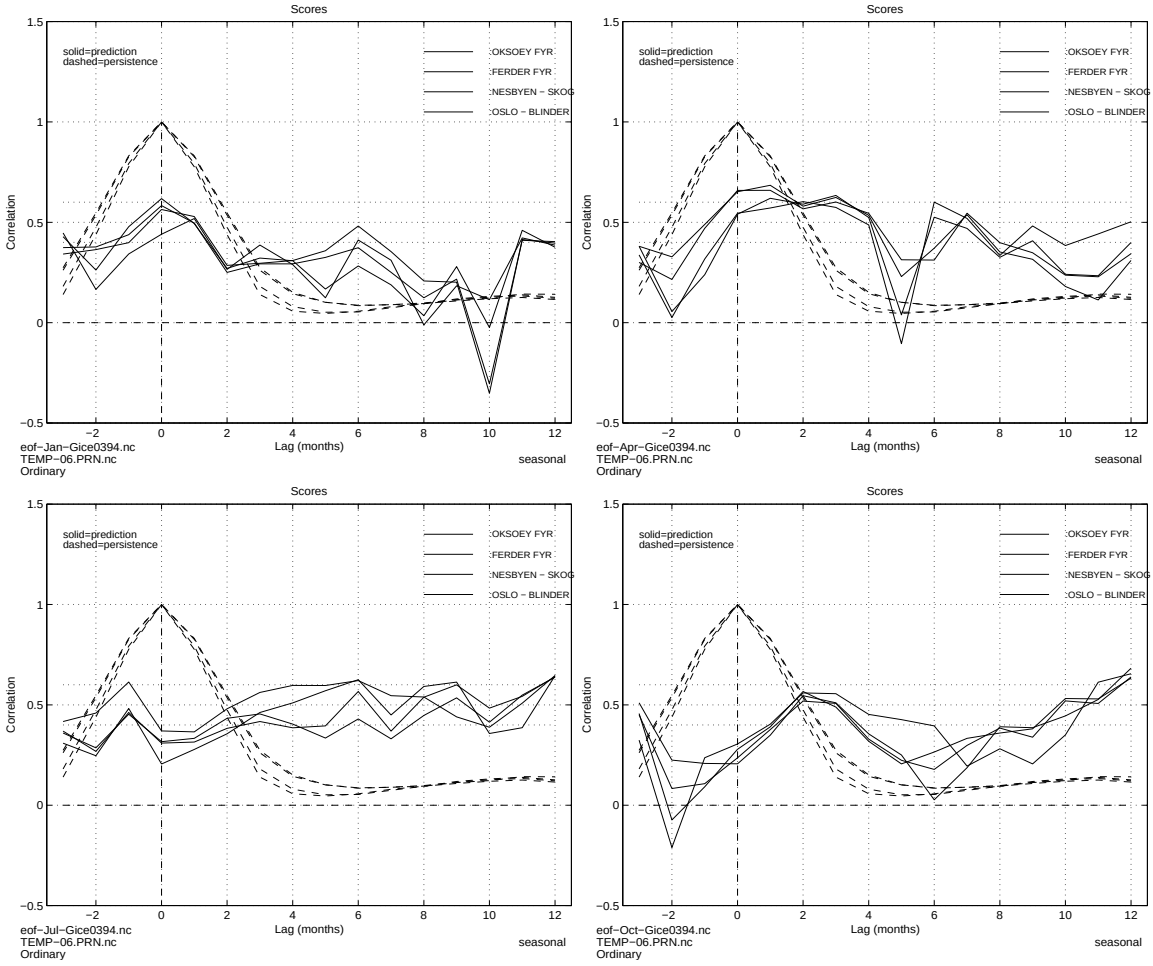

Figure 17: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 sea-ice.*

The prediction schemes based on sea-ice alone achieved only marginal skill scores (figure 17), with greatest predictability for the April hindcasts. The January hindcast only yielded high scores for 0-month lead time. The July predictions, on the other hand, indicated some skill for long lead times, and the October hindcasts scores peaked around December-January and the subsequent autumn showed some recovery of skill. How much of this skill is true remains to be found out in future experimental forecasts projects. It is

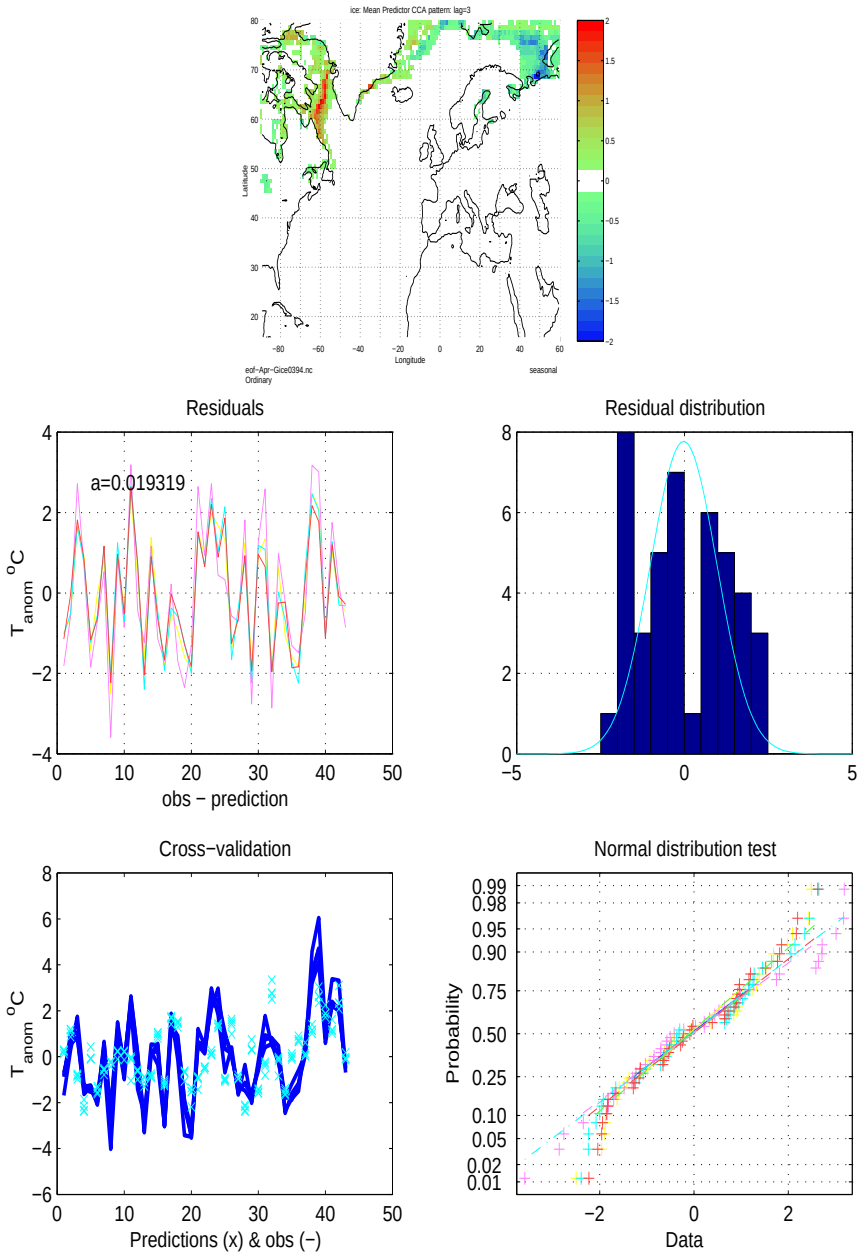

Figure 18: *Same as figure6, but for the April sea-ice CCA models.*

possible that the forecast information in the sea-ice observations may give additional skill for empirical models based on more than one predictor quantity.

The ice predictor pattern indicates an anti-phase relationship between the ice extent in the Labrador Sea and the Greenland-Iceland-Norwegian (GIN)

Sea (figure 18, top). This pattern may be linked with the NAO and *Deser & Blackmon* (1993) have noted a strong lagged correlation between sea ice in the Labrador sea and North Atlantic SSTs 1-2 years later. The residuals are of similar magnitude to the predicted values and highly correlated with each other, suggesting that the ice models did not reproduce all of the predictable signal. Middle right and lower right panels indicate that the residuals were not normally distributed.

4.4.2 Western Norway

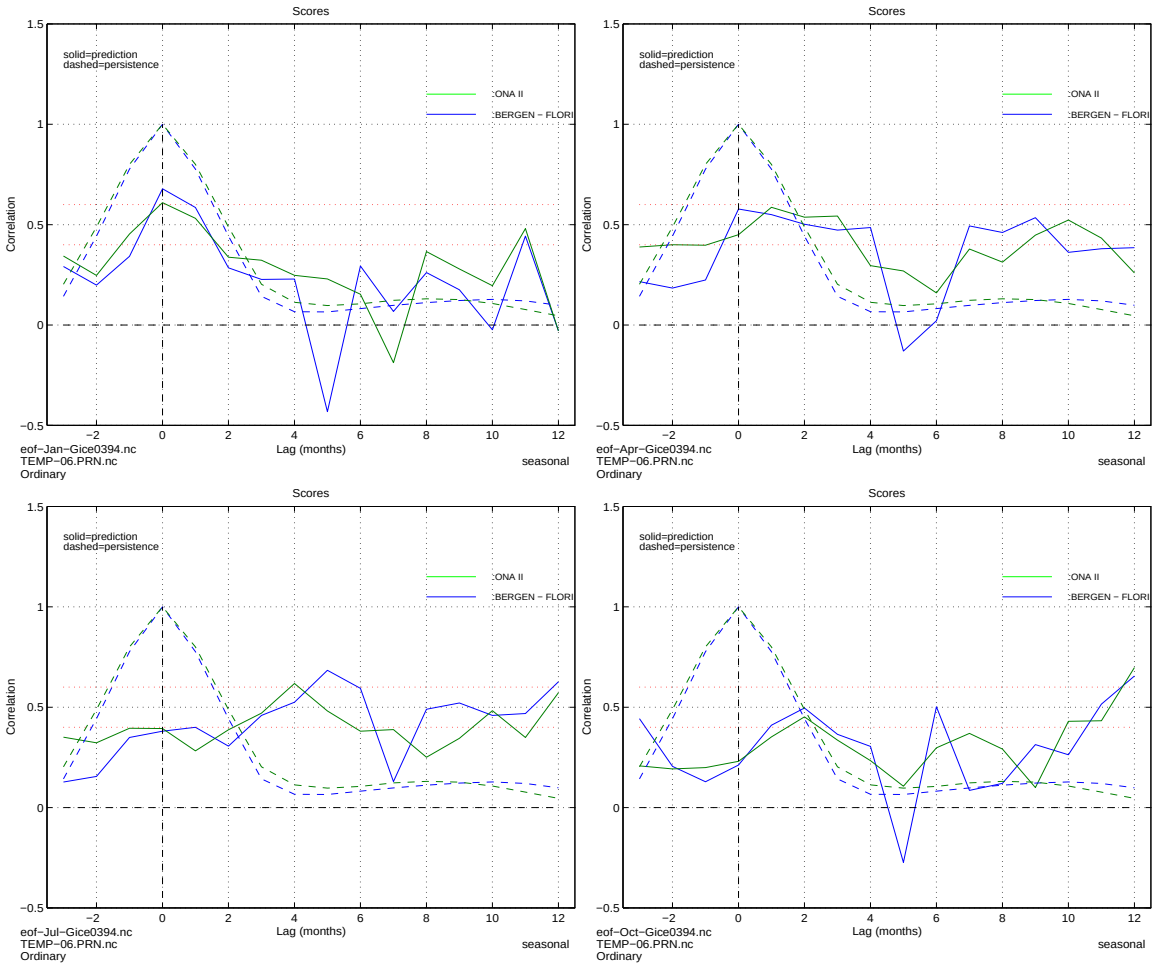

Figure 19: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 sea-ice.*

4.4.3 Northwestern Norway

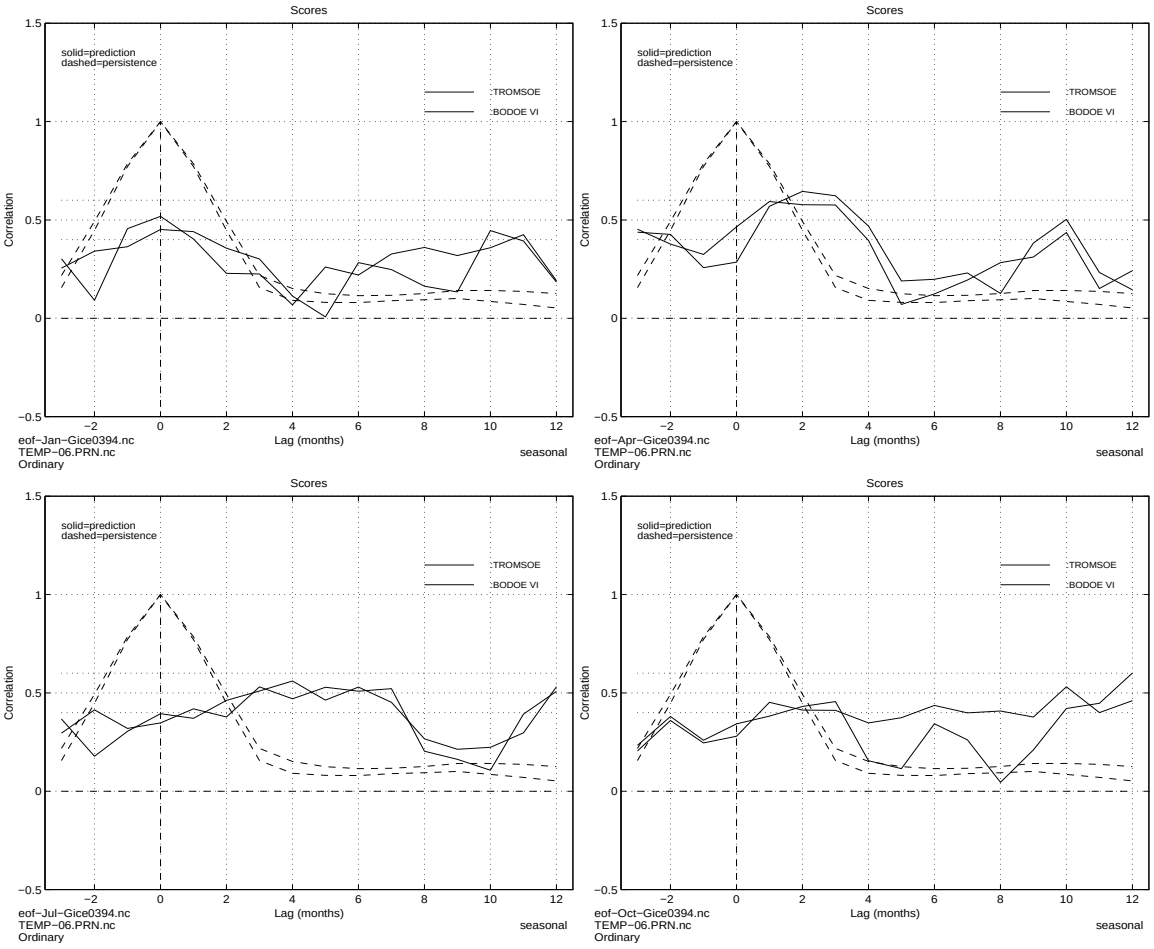

Figure 20: *Skill scores for temperature predictions for Oslo, Nesbyen, Ferder and Oksøy starting in January (upper left), April (upper right), July (lower left) and October (lower right). The predictions were based on GISST2.2 sea-ice.*

It was surprising to see similar sea-ice model prediction score for southern, western and northern Norway.

5 Precipitation predictions based on CCA

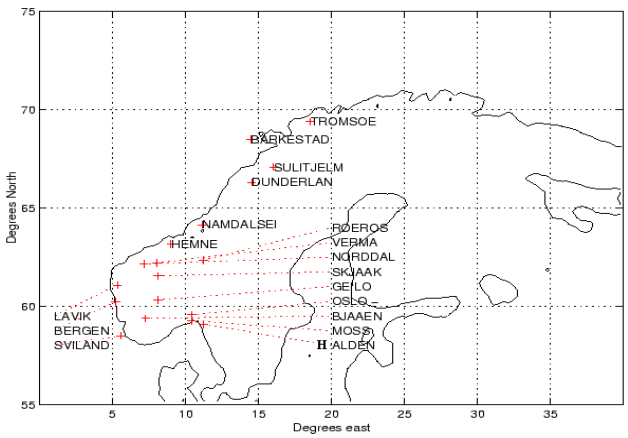

Figure 21: Map of stations measuring precipitation.

Figure 21 shows the locations of the 18 precipitation stations used as predictands in the following discussion. The results shown in this section were obtained from empirical models calibrated with all the sites shown in figure 21 ("nationwide model"), as opposed to the predictions in the previous section where different temperature models were developed for different regions ("regional models"). The need for regional models is just as great for precipitation predictions, however, the objective of this section is merely to indicate the potential for predictability of rainfall for the sites with highest skill (the west coast of Norway). All the results shown in the figures are in terms of precipitation anomalies ($\times 10^{-4} m$).

5.1 SLP predictions

Figure 22 shows January seasonal 1-month lead time (target interval: Feb-Apr) precipitation hindcasts for Bergen (top left) and Tromsø (top right), together with the leading CCA pattern (bottom left) and predictand weights and scores (bottom right). Four of the stations in the southwestern part of Norway had predictions with a high cross-correlation score, whereas the predictions for northern and eastern Norway had poor skill. The predictor pattern indicates that the precipitation in southwestern Norway was associated with westerly geostrophic wind, bringing in moist and mild maritime air from the North Atlantic.

The prediction scores (left) and leading CCA pattern (right) for 1-month lead time hindcasts starting at the spring, summer and autumn seasons are

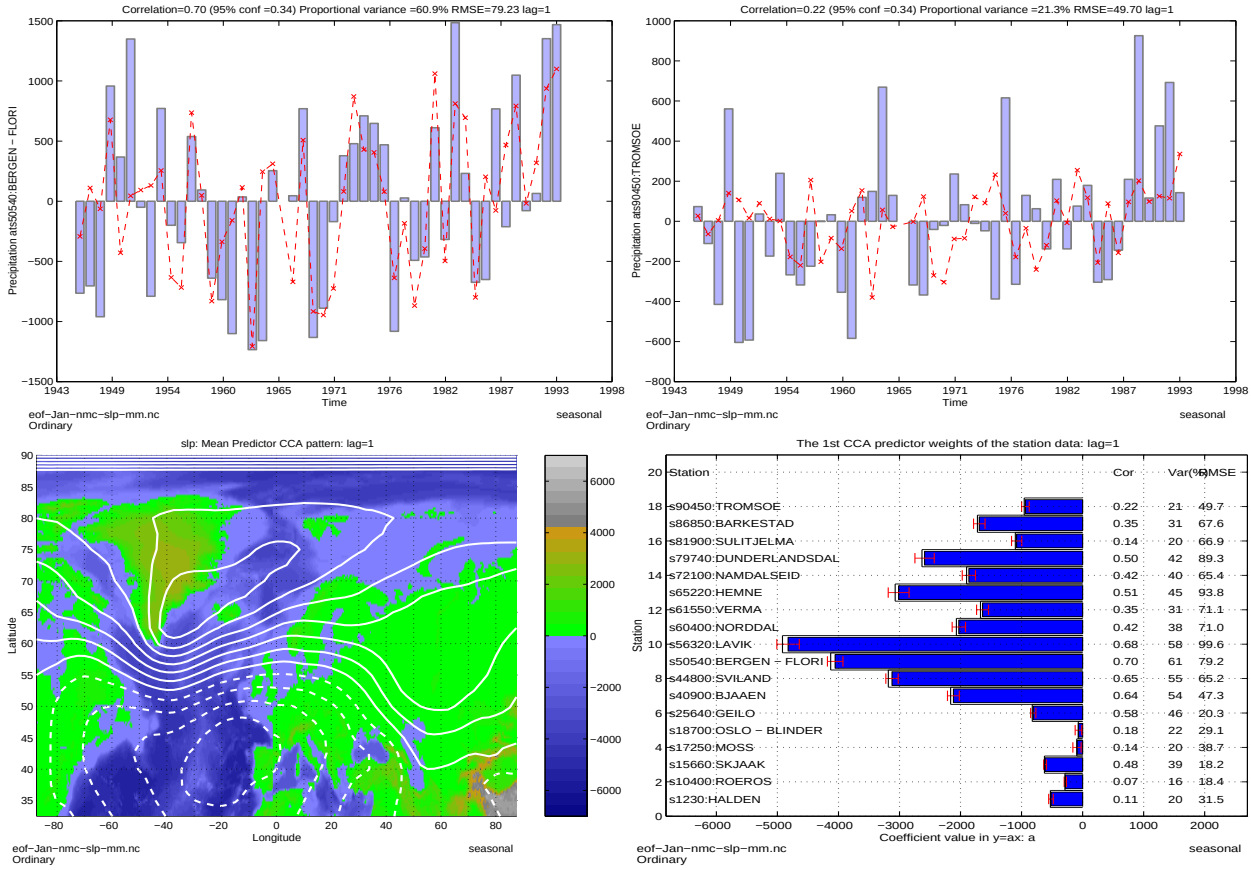

Figure 22: The seasonal hindcast predictions for Feb-Apr precipitation with lead times of 1 month at Bergen (upper left) and Tromsø (upper right). The corresponding leading CCA SLP patterns is shown in the bottom left panel and the predictand weights and prediction scores at the bottom right. The units are in the upper panels are 0.1mm and the quantities shown are anomalies.

shown in figure 23. The April predictions rely to a strong degree on NAO like SLP structure, whereas the prediction patterns for the predictions starting in the summer or the autumn resemble a distorted and displaced NAO structure. The important geostrophic wind for all seasons has a strong zonal component, and the locations with the highest prediction scores are found on the west coast of Norway for predictions made during spring time, in the south Norwegian mountains (Geilo) during summer time and around Trondheim during the autumn season.

In April (figure 24) useful prediction skills for 2-month lead time were found in Bergen and Lavik, both on the western coast of Norway. Table

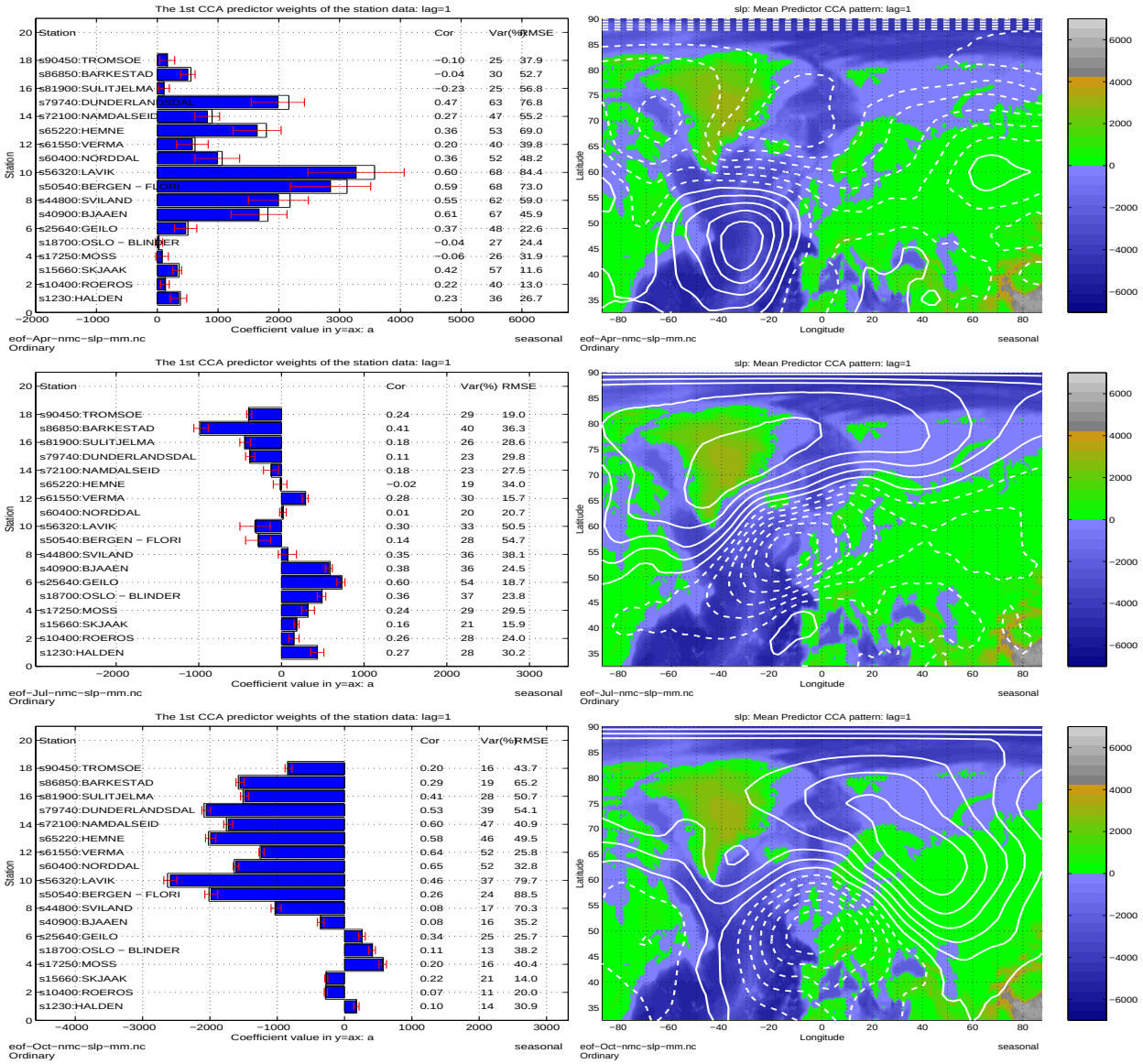

Figure 23: Predictand weights and scores (left) and leading SLP CCA patterns for 1-month lead time seasonal precipitation predictions starting in April (upper), July (middle), and October (bottom) using SLP as predictor.

3 shows the prediction skill for hindcasts starting at different times of the year. The predictions starting in April may have prospects of producing useful *seasonal mean* predictions for up to 2 months in advance. In July, marginal skill can be seen in the 2-month lead time predictions for Bergen (table 3) while the 2-month lead time hindcasts starting in October were

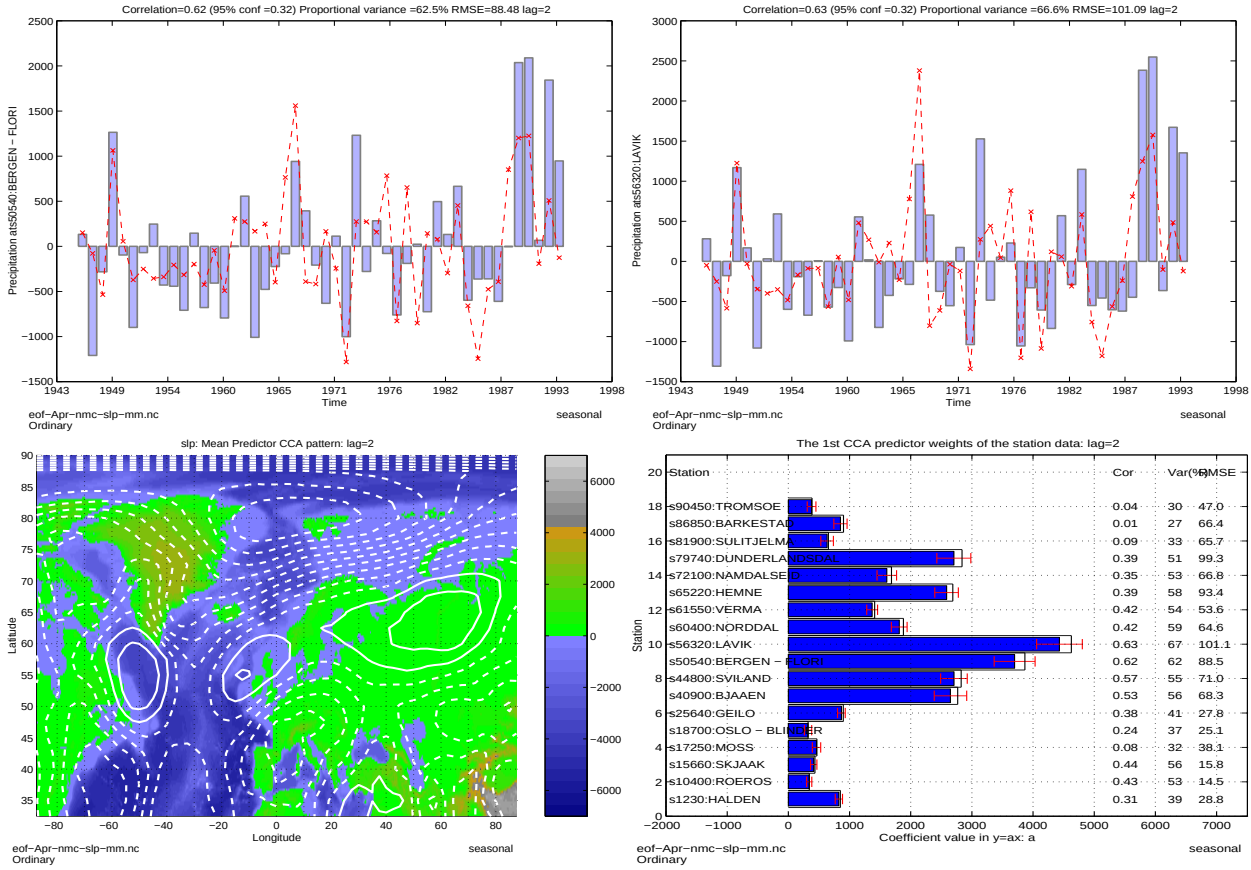

Figure 24: The seasonal hindcast predictions for Jun-Aug precipitation with lead times of 2 months at Bergen (upper left) and Lavik (upper right). The corresponding leading CCA SLP patterns is shown in the bottom left panel and the predictand weights and prediction scores at the bottom right. The units are in the upper panels are 0.1mm and the quantities shown are anomalies.

poor.

Table 3: Hindcast cross-variance correlation scores for SLP predictions for seasonal mean precipitation. Scores greater than 0.59 are shown in bold, and the 95% statistical significance limit was around 0.30.

Lag	Oslo	Bergen	Tromsø
January			
0 (DJF)	-0.07	**0.73**	0.52
1 (JFM)	0.18	**0.70**	0.22
2 (FMA)	0.32	0.23	0.39
3 (MAM)	0.42	0.16	0.19
April			
0 (MAM)	-0.16	**0.67**	-0.24
1 (AMJ)	-0.04	0.59	-0.10
2 (MJJ)	0.24	**0.62**	0.04
3 (JJA)	0.13	0.39	0.45
July			
0 (JJA)	0.13	0.39	0.45
1 (JAS)	0.36	0.14	0.24
2 (ASO)	-0.22	0.44	-0.20
3 (SON)	0.02	0.30	0.08
October			
0 (SON)	0.40	0.31	0.20
1 (OND)	0.11	0.26	0.20
2 (NDJ)	-0.10	-0.36	-0.26
3 (DJF)	-0.18	0.11	0.21

5.2 SST predictions

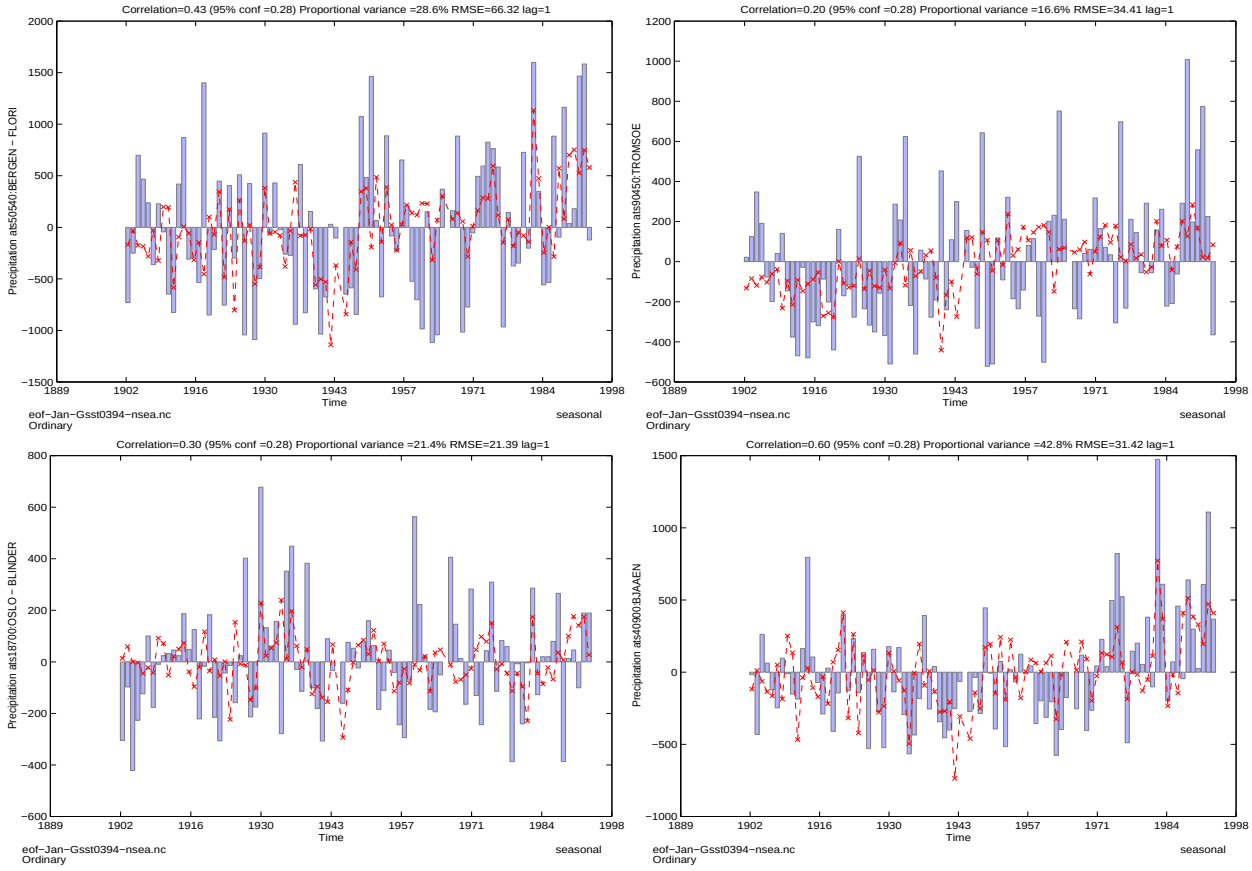

Figure 25: The Feb-Apr seasonal mean hindcast predictions for precipitation with lead times of 1 month for Bergen (upper left), Tromsø (upper right), Oslo (lower left) and Bjåen (lower right). The predictors were SST and the quantities shown are anomalies (units=0.1mm).

Figure 25 shows the 1-month lead time predictions made for the January to March (JFM) season at Bergen (top left), Tromsø (top right), Oslo (bottom left) and Bjåen (bottom right) employing Nordic Seas SST anomalies. Although the 1-month lead SST prediction scores were lower than those based on SLPs, the prediction skill of seasonal hindcast at Bjåen may be considered as useful (lower right). The leading CCA pattern (figure 26, left panel) indicates that most of the predictability came from SST anomalies in the North Sea, Skagerrak, Kattegat, and the Baltic Sea, and the predictand weights (right panel) suggest that enhanced precipitation is associated with warmer sea surface.

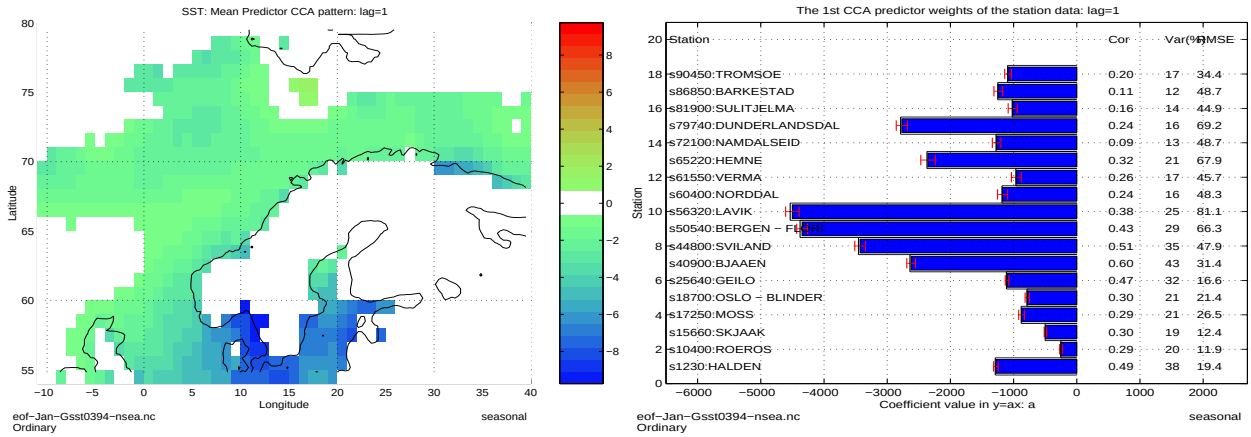

Figure 26: The leading CCA SST pattern (right) and hindcast predictand weights (left) and prediction scores for seasonal mean precipitation with lead times of 1 month.

The correlation scores of predictions with 2-months and 3-months lead time were all below the 95% confidence limit (table 4), contrary to the seasonal temperature predictions, and the poor long lead scores were possibly due to low persistence in the precipitation records and/or weak dependence on the SSTs.

Table 4: Hindcast cross-variance correlation scores for SST predictions for seasonal mean precipitation. Scores greater than 0.59 are shown in bold, and the 95% statistical significance limit was around 0.30.

Lag	Oslo	Bergen	Tromsø
January			
0 (DJF)	0.35	0.38	0.21
1 (JFM)	0.30	0.43	0.20
2 (FMA)	0.35	0.16	0.27
3 (MAM)	0.15	0.32	-0.05
April			
0 (MAM)	0.19	0.53	0.28
1 (AMJ)	0.07	0.36	0.24
2 (MJJ)	0.08	0.17	0.12
3 (JJA)	0.05	0.22	-0.06
July			
0 (JJA)	-0.04	0.42	0.33
1 (JAS)	0.19	0.21	0.07
2 (ASO)	0.21	0.10	0.16
3 (SON)	0.21	0.10	0.22
October			
0 (SON)	0.01	0.19	0.19
1 (OND)	-0.08	0.32	0.27
2 (NDJ)	0.31	0.15	0.14
3 (DJF)	0.13	-0.04	-0.10

5.3 500hPa geopotential height predictions

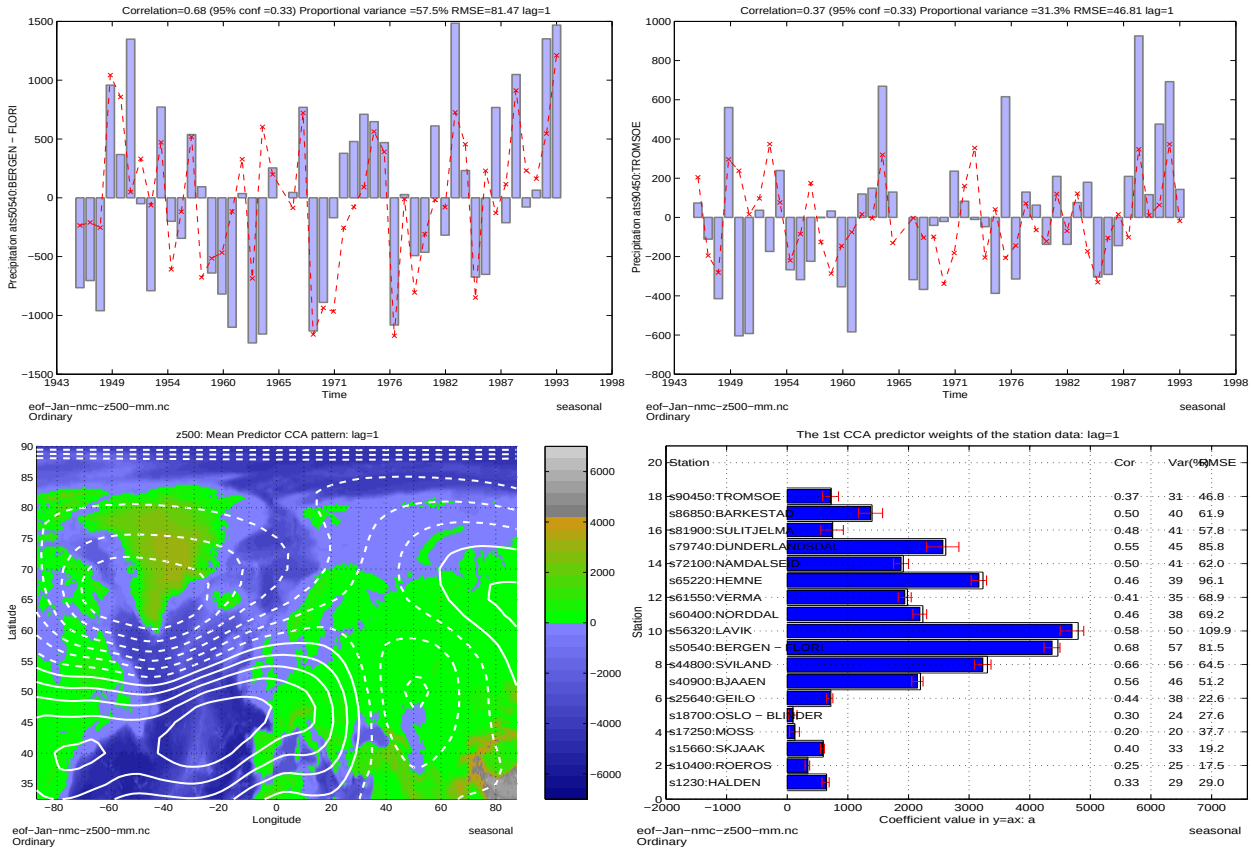

Figure 27: The Feb-Apr seasonal precipitation hindcast predictions with lead times of 1 month at Bergen (upper left) and Tromsø (upper right). The corresponding leading CCA Φ_{500} patterns is shown in the bottom left panel and the predictand weights and prediction scores at the bottom right.

Predictions with 1-month lead time of seasonal precipitation at Bergen (top left) and Tromsø (top right) using 500hPa geopotential heights are shown in figure 27. The predictor pattern (figure 27, bottom left) bears some resemblance to the NAO dipole structure, and the predictions at both Bergen and Sviland had useful skill (figure 27, bottom right).

The skills (left) and predictor patterns (right) of 500hPa geopotential heights predictions of seasonal precipitation with 1-month lead time in figure 28 indicate useful prediction scores for Sviland during autumn, but no useful scores during the spring and summer time. The 2-month lead time skills were lower than those of shorter lead time, indicating no re-occurrence in skill (table 5).

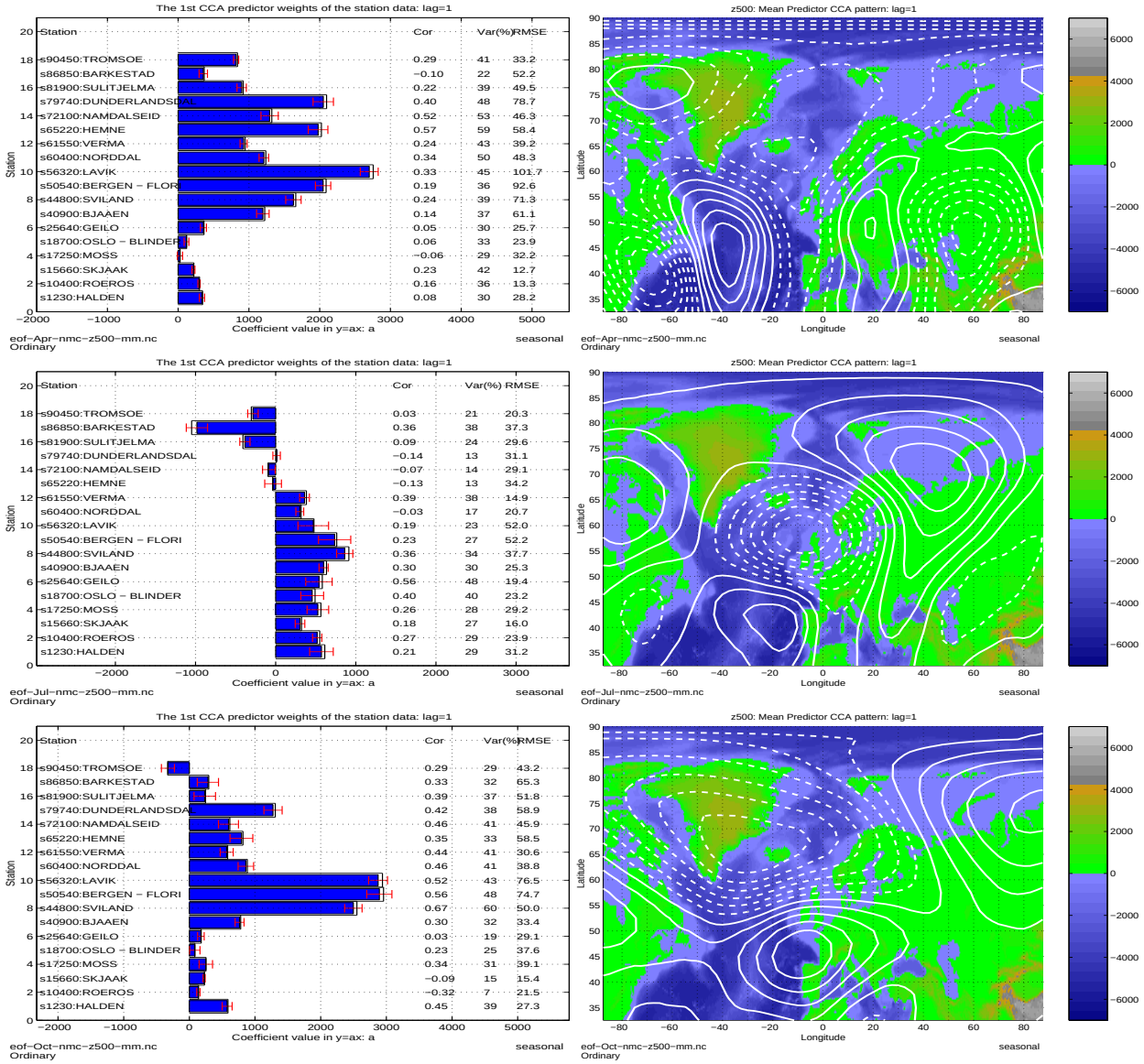

Figure 28: Predictand weights and scores (left) and leading Z500 CCA patterns for 1-month lead time seasonal precipitation predictions stating in January (upper) April (second from the top), July (third from the top) and October (bottom) using Z500 as predictors.

Table 5: Hindcast cross-variance correlation scores for z500 predictions for seasonal mean precipitation shown. Scores greater than 0.59 are shown in bold, and the 95% statistical significance limit was around 0.30.

Lag	Oslo	Bergen	Tromsø
January			
0 (DJF)	0.02	**0.82**	**0.67**
1 (JFM)	0.34	**0.68**	0.37
2 (FMA)	0.14	0.32	0.12
3 (MAM)	0.13	-0.05	-0.05
April			
0 (MAM)	0.04	0.56	0.00
1 (AMJ)	0.06	0.19	0.29
2 (MJJ)	0.05	0.18	0.30
3 (JJA)	-0.01	0.05	0.10
July			
0 (JJA)	0.58	0.22	0.50
1 (JAS)	0.40	0.23	0.03
2 (ASO)	-0.22	0.08	-0.06
3 (SON)	0.08	0.04	-0.05
October			
0 (SON)	0.37	0.44	0.38
1 (OND)	0.23	0.56	0.29
2 (NDJ)	-0.01	0.10	0.10
3 (DJF)	-0.33	0.04	0.02

5.4 Sea-ice predictions

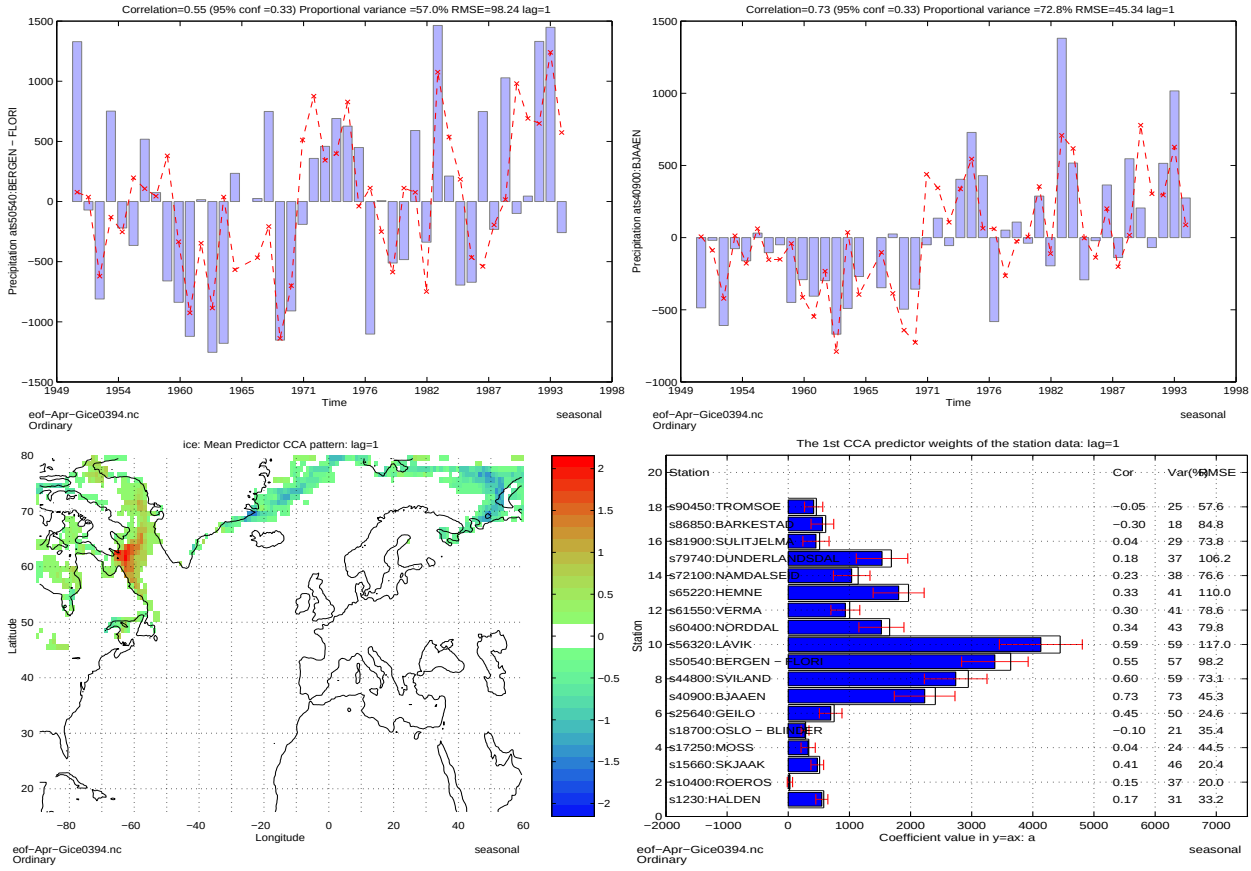

Figure 29: The May-Jul seasonal precipitation hindcast predictions with lead times of 1 month at Bergen (upper left) and Bjåen (upper right). The corresponding leading CCA sea-ice patterns is shown in the bottom left panel and the predictand weights and prediction scores at the bottom right.

Figure 29 shows May-Jul seasonal precipitation hindcasts for Bergen (top left) and Bjåen (top right) employing April sea-ice cover, and it appears that precipitation at Sviland and Bjåen are correlated with the expansion of the Labrador ice sheet and the retraction of the ice edge along the east coast of Greenland and the Polar Sea. Table 6 gives an overview of the prediction scores for the sea-ice models, indicating that apart from the two locations with high prediction skills for the 1-month lead time April-June (AMJ) hindcasts, the sea-ice models were poor models in terms of precipitation predictions.

It is difficult to explain how the sea-ice extension may affect the Norwe-

Table 6: Hindcast cross-variance correlation scores for sea-ice predictions for seasonal mean precipitation. Scores greater than 0.59 are shown in bold, and the 95% statistical significance limit was around 0.30.

Lag	Oslo	Bergen	Tromsø
January			
0 (DJF)	-0.18	**0.69**	0.55
1 (JFM)	-0.21	0.42	-0.20
2 (FMA)	-0.21	0.25	-0.14
3 (MAM)	-0.14	0.44	-0.06
April			
0 (MAM)	0.00	**0.63**	0.22
1 (AMJ)	-0.10	0.55	-0.05
2 (MJJ)	-0.34	0.55	-0.25
3 (JJA)	0.07	0.47	0.25
July			
0 (JJA)	0.36	0.28	**0.67**
1 (JAS)	-0.34	0.39	-0.23
2 (ASO)	-0.04	-0.32	-0.20
3 (SON)	-0.03	-0.17	-0.05
October			
0 (SON)	-0.09	-0.02	-0.08
1 (OND)	-0.05	0.26	0.33
2 (NDJ)	-0.16	-0.04	0.38
3 (DJF)	0.10	0.12	-0.32

gian rainfall without more in-depth studies using numerical models, however, it is possible that large scale signals, such as NAO, is associated with both the ice formation/melting as well as influencing the Norwegian precipitation.

Table 7: Best Scores and Predictors. The number on the right hand indicated the forecast lead time. Scores less than 0.30 are left blank.

Month	Oslo	Bergen	Tromsø
Precipitation			
DJF			
JFM	0.34 z500 1	0.70 SLP 1	0.37 z500 1
FMA	0.35 SST 2	0.32 z500 2	0.39 SLP 2
MAM	0.42 SLP 3	0.44 ICE 3	
AMJ		0.56 SLP 1	
MJJ		0.62 SLP 2	
JJA		0.47 ICE 3	0.45 SLP 3
JAS	0.40 z500 1	0.39 ICE 1	
ASO		0.44 SLP 2	
SON		0.30 SLP 3	
OND		0.56 z500 1	0.33 ICE 1
NDJ	0.31 SST 2		0.38 ICE 2

6 Discussion and conclusion

The seasonal prediction analysis gave evidence for some historical predictive skill associated with Norwegian surface temperatures. Some of the skill is thought to be related to the persistence of climatic anomalies, but there are also indication of a lagged relationship between the large scale features and the local temperatures. The potential seasonal predictability can only be assessed further through an experimental operational forecast experiments.

Table 7 presents the best skill scores for seasonal precipitation prediction achieved in this study. The seasonal prediction scores for precipitation did not yield as promising results as for temperatures, however, the precipitation on the west coast of Norway was nevertheless associated with at least some predictability. The southeastern and northern Norway did not show any significant predictability with regard to precipitation. SLP appeared to give the highest skill scores, however, other quantities such as SST, Φ_{500}, and sea-ice gave sometimes better predictions than the SLP. The analysis of the predictability of precipitation was far from exhaustive, and further work is needed to get a more complete picture of seasonal predictability of precipitation. The short data records examined here will furthermore imply large sampling fluctuations and insufficiently long time series for model calibration.

It is possible that the empirical models can be improved by including more than one parameter in the predictor data set. However, the NAO appears to produce signals in all data sets that are associated with predictive skills, and it is possible that most of the spatial patterns in the various fields examined here contain the same signal. The different residual statistics for the different predictor quantities suggest that the various predictor quantities did not account for the same signals. There may also be room for further improvement in model skill by employing better methods, and multichannel singular spectral analysis (*Vautard* et al., 1999) has been suggested as a promising candidate. Refinement of the empirical models may possibly involve the removal of linear trends in the time domain prior to the model calibration, and the linear trend may be used as an additional predictor. However, the analysis of residuals indicated that the residual trends were small, suggesting that the separation of linear trends may only contribute to marginal hindcast improvements for the cases examined here (western Norway).

One important result from this study was that one empirical hindcast model could not produce optimal predictions for several different climatic regions, but each region required one empirical model which was especially optimised for the local climate.

7 Acknowledgment

This work was supported by the Norwegian Electricity Federation ("*Energiforsyningens Fellesorganisasjon*", EnFO), contractno. 551017.

References

Anderson, D.L.T. 1995. The legacy of TOGA: seasonal climate prediction. *In:* ECMWF (ed), *Predictability.* Seminar proceedings, vol. 2. ECMWF.

Balmaseda, M.A., Anderson, D.L.T, & Davey, M.K. 1994. ENSO prediction using a dynamical ocean model coupled to statistical atmosphere. *Tellus*, **46A**, 497–511.

Barnston, A. 1995. Brief Summary of NMC's Canonical Correlation Analysis (CCA), Optimal Climate Normals (OCN), and NMC Coupled Model Forecasts for U.S. Surface Climate. *NOAA: Experimental Long-Lead Forecast Bulletin*, **4**, 45–46.

Barnston, A. G., & C.F.Ropelewski. 1992. Prediction of ENSO Using Canonical Correlation Analysis. *Journal of Climate*, **5**, 1317–1345.

Barnston, A.G. 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. *Journal of Climate*, **7**, 1513–1564.

Barnston, A.G., van den Dool, H.M., Zebiak, S.E., Barnett, T.P., Ji, M., Rodenhuis, D.R., Cane, M.A., Leetmaa, A., Graham, N.E., Ropelewski, C.R., Kousky, V.E., O'Lenic, E.A., & Livezey, R.E. 1994. Long-Lead Seasonal Forecasts - Where do we stand? *Bullit. Am. Met. Soc*, **75**, 2097–2114.

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

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

Benestad, R.E. 1998c. *SVD applied to Statistical Downscaling for Prediction of Monthly Mean Land Surface Temperatures: Model Documentation.* Klima 30/98. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1999. *MVR applied to Statistical Downscaling for Prediction of Monthly Mean Land Surface Temperatures: Model Documentation.* Klima 2/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Colman, A., & Davey, M. 1999. Prediction of summer temperature, rainfall and pressure in Europe from preceding winter North Atlantic Ocean temperatures. *International Journal of Climatology*, **19**, 513–536.

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

Doblas-Reyes, F.J., Déqué, M., Valero, F., & Stephenson, D.B. 1998. North Atlantic wintertime intraseasonal variability and its senaitivity to GCM horizontal resolution. *Tellus*, **50A**, 573–595.

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.

Ghil, M., & Yiou, P. 1996. Spectral Methods: What they Can and Cannot Do for Climate Times Series. *In:* Anderson, D.L.T., & Willebrand, J. (eds), *Decadal Variability.* NATO ASI series, vol. 44. Springer.

Godske, C.L. 1956. *Hvordan blir været?* 1 edn. Oslo: J.W. Cappelen.

Hasselmann, K. 1988. PIPs and POPs: The reduction of complex dynamical systems using Principal Interaction and Oscillation Pattern. *Journal of Geophysical Research*, **93**, 11015–11021.

Jiang, N., Ghil, M., & Neelin, D. 1995. Forecasts of Equatorial Pacific SST Anomalies Using an Autoregressive Process Using Singular Spectrum Analysis. *NOAA: Experimental Long-Lead Forecast Bulletin*, **4**, 24–27.

Johansson, Å., Barnston, A., Saha, S., & van den Dool, H. 1998. On the Level and Origin of Seasonal Forecast Skill in Northern Europe. *Journal of the Atmospheric Sciences*, **55**, 103–127.

Jones, P.D. 1987. The early twentieth century Arctic high - fact or fiction? *Climate Dynamics*, **1**, 63–75.

Katz, R.W. 1988. Use of Cross Correlations in the Search for Teleconnections. *Journal of Climatology*, **8**, 241–253.

Kim, K-Y., & North, G.R. 1998. EOF-based Linear Prediction Algorithm: Theory. *Journal of Climate*, **11**, 3046–3056.

Kushnir, Y., & Held, I.M. 1996. Equilibrium Atmospheric Response to North Atlantic SST Anomalies. *Journal of Climate*, **9**, 1208–1220.

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

Mullan, A.B. 1998. Southern Hemisphere Sea-Surface Temperatures and their Contemporary and Lag Association with New Zealand Temperature and Precipitation. *International Journal of Climatology*, **18**, 817–840.

Palmer, T.N., & Anderson, D.L.T. 1994. The prospect for Seasonal Forecasting - A Review Paper. *Q.J.R.M.S.*, **120**, 755–793.

Stephenson, D.B. 1997. Correlation of Spatial Climate/Weather Maps and the Advantage of using Mahalanobis Metric in predictions. *Tellus*, **49A**, 513–527.

Stockdale, T.N., Anderson, D.L.T., Alves, J.O.S., & Balmaseda, M.A. 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370–373.

Vautard, R., Plaut, G., Wang, R., & Brunet, G. 1999. Seasonal Prediction of North American Surface Temperatures Using Space-Time Principal Components. *Journal of Climate*, **12**, 380–394.

Watanabe, M., & Nitta, T. 1999. Decadal changes in the atmospheric circulation and associated surface climate variations in the northern hamisphere winter. *Journal of Climate*, **12**(2), 494–510.

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

Zorita, E., & von Storch, H. 1997. *A survey of statistical downscaling results.* Tech. rept. 97/E/20. GKSS.