424

# Monthly forecasting system

## Frédéric Vitart

## Research Department

SAC report October 2003

**Abstract**

A monthly forecasting system has been set up at ECMWF to fill the gap between the EPS and the seasonal forecasting system: it shares characteristics of both systems. It has been run routinely since March 2002, and 30 cases from March 2002 to May 2003 have been verified. Results of this validation suggest that the monthly forecasting system behaves similarly to the EPS during the first ten days of the forecast. Beyond day 10, a comparison between the control of the monthly forecasting system and the control of the EPS suggests that the ocean-atmosphere coupling may impact the forecasts over some regions such as Asia or the North Pacific.

Beyond day 10, the model displays some skill in predicting weekly-averaged 2-metre temperature, precipitation and mean sea-level pressure anomalies relative to the climate from the past 12 years. For the period days 12-18, the ROC score is generally larger than 0.6, and the probabilistic scores are significantly higher than the scores obtained when persisting the anomaly forecast of the previous week. At this time range, the monthly forecasting system usually performs better than climatology and persistence, and therefore the forecasts can be useful. After about 20 days of forecast (days 19-32), the model displays some skill in predicting events with a large threshold. The performance of the system depends strongly on the geographical location with Europe a difficult region. Over Europe, during the verification year, the model forecasts for days 19-32 were generally not very reliable and the probabilistic scores were not significantly better than climatology or persistence. However, the model displayed some useful skill at this time range over North America, Asia and the Southern Extratropics.

In order to investigate the skill of the monthly forecasting system to predict weather regimes, a blocking index was applied to the real-time monthly forecasts from March 2002 to May 2003. Results suggest that the model has moderate skill in predicting blocking events over the Euro-Atlantic region for days 12-18, but no skill at all for days 19-32. However, the model is more successful over the North Pacific. This is also confirmed when verifying the blocking index applied to 12-year model climatology.

The Madden-Julian oscillation (MJO) is an important source of predictability for monthly forecasts. The model has some useful skill in predicting its time-evolution up to 20 days in advance though the variance of the MJO is reduced by a factor two after only 10 days of forecasts. Therefore, the impact of the MJO on the extratratropical circulation is likely to have been underestimated by the monthly forecasting system after day 10.

The model also has skill in predicting the variability of tropical SSTs, particularly the variability linked to ENSO. For instance, the model produces better forecasts of NINO3 SSTs than persistence after day 10. The model also produces useful forecasts of the North Atlantic Oscillation (NAO) during at least the first 20 days of the forecast.

# 1 Introduction

The monthly forecasting system fills the gap between the two currently operational forecasting systems at ECMWF: medium-range weather forecasting and seasonal forecasting. Medium-range weather forecasting produces weather forecasts out to 10 days, whereas seasonal forecasting produces forecasts out to 6 months. The two systems have different physical bases. Medium-range weather forecasting is essentially an atmospheric initial value problem. Since the time scale is too short for variations in the ocean significantly to affect the atmospheric circulation, the ECMWF medium-range weather forecasting system is based on atmospheric-only integrations. SSTs are simply persisted. Seasonal forecasting, on the other hand, is justified by the long predictability of the oceanic circulation (of the order of several months) and by the fact that the variability in tropical SSTs has a significant global impact on the atmospheric circulation. Since the oceanic circulation is a major source of predictability in the seasonal scale, the ECMWF seasonal forecasting system is based on coupled ocean-atmosphere integrations. Seasonal forecasting is also an initial value problem, but with much of the information contained in the initial state of the ocean.

The monthly forecasting system produces forecasts for the time range 10 to 30 days. The time range 10 to 30 days is probably still short enough that the atmosphere retains some memory of its initial state and it may be long enough that the ocean variability has an impact on the atmospheric circulation. Therefore, the monthly forecasting system has been built as a combination of the medium-range EPS and the seasonal forecasting system. It contains features of both systems and, in particular, is based on coupled ocean-atmosphere integrations, as is the seasonal forecasting system.

An important source of predictability in the monthly time range over Europe in the 10-30 day range is believed to originate from the Madden-Julian Oscillation (MJO) (see, for instance, Ferranti et al 1990). Several papers (see, for instance, Flatau et al, 1997) suggest that the ocean-atmosphere coupling has a significant impact upon the speed of propagation of an MJO event in the Indian Ocean and western North Pacific. The use of a coupled system may therefore help to capture some aspects of MJO variability.

The ECMWF monthly forecasting system will be described in Section 2 of this paper. The performance of the system over the first 10 days of the forecasts will be compared to the performance of the EPS system in Section 3. The skill of the system after 10 days will be presented in Section 4. Section 5 will focus on the skill of the model to predict weather regimes and most especially blocking. Section 6 will evaluate the skill of the monthly forecasting system to predict the evolution of the Madden-Julian Oscillation, the importance of which for monthly forecasting has been mentioned above. Finally Section 7 will evaluate the skill of the system to predict variability in SSTs and indices, like the SOI, PNA or NAO, which also can have a significant impact on the forecasts at this time range. The systematic errors in the monthly forecasting system are discussed in the SAC paper entitled 'Systematic model errors' by Thomas Jung.

## 2    Description of the monthly forecasting system

The monthly forecasts are based on an ensemble of 51 coupled ocean-atmosphere integrations (one control and 50 perturbed forecasts). The length of the coupled integration is 32 days, and the frequency of the monthly forecasts is currently every 2 weeks, which is more frequent than seasonal forecasting (once a month), but less frequent than EPS (twice a day). The atmospheric component is IFS with the same cycle as the deterministic forecast. Currently, the atmospheric model is run at $T_L 159$ resolution ($1.125^o \times 1.125^o$) with 40 levels in the vertical. This represents a resolution in between EPS ($T_L 255L40$) and seasonal forecasting ($T_L 95L40$). The oceanic component is the same as for seasonal forecasting system 2. It is HOPE (Wolff et al, 1997), from the Max Plank Institute. The ocean model has lower resolution in the extratropics but higher resolution in the equatorial region, in order to resolve ocean baroclinic waves and processes which are tightly trapped at the equator. The ocean model has 29 levels in the vertical. The atmosphere and ocean communicate with each other through a coupling interface, called OASIS (Terray et al, 1995), developed at CERFACS. The atmospheric fluxes of momentum, heat and fresh water are passed to the ocean every hour and, in exchange, the ocean sea surface temperature (SST) is passed to the atmosphere. The frequency of coupling is higher than in seasonal forecasting (every 24 hours), since high frequency coupling may have some impact on the development of some synoptic scale systems, such as tropical cyclones.

In order to initiate monthly forecasts, initial conditions for both the ocean and atmosphere are required. Atmospheric and land surface initial conditions are obtained from the ECMWF operational atmospheric analysis/reanalysis system. Oceanic initial conditions originate from the oceanic data assimilation system used to produce the initial conditions of the seasonal forecasting system 2. However, this oceanic data assimilation system lags about 12 days behind real-time. The lag is partially due to the fact that the SST, obtained by interpolating in time the weekly OIv2 SSTs produced by NCEP, can be up to 11 days behind real-time. A first option would be to wait for the oceanic initial condition to be created by the data assimilation system to start the
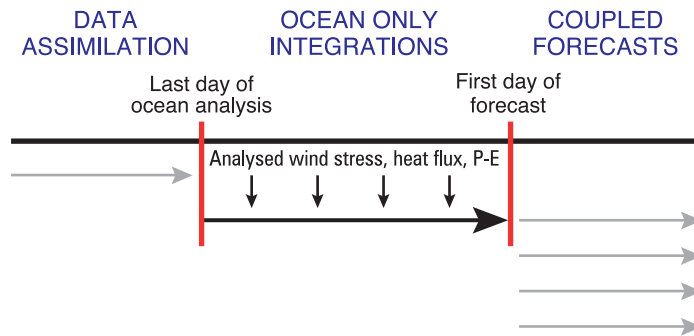
*Figure 1: Diagram representing the setting of the accelerated analysis for the monthly forecasting system. The accelerated analysis starts from the last ocean analysis which usually lags about 12 days behind real-time. The accelerated analysis is performed by integrating the ocean model forced by analyzed wind stress, heat fluxes and P-E from the operational analysis, with a strong relaxation to persisted SSTs. The coupled integration start starts from the ocean state provided by the 12-day run of the accelerated analysis.*

forecast, as in seasonal forecasting. This would mean losing 12 days of forecast and is not therefore suitable for monthly forecasting. A second option would be to persist the SST anomalies of the latest ocean analysis. However, we have some information about the wind stress and heat fluxes during the last 12 days of the ECMWF atmospheric analysis; this information can be used to help determine the present ocean initial state. Therefore, the option that has been chosen for monthly forecasting consists in integrating the ocean model from the last ocean analysis, forced by analyzed wind stress, heat fluxes and P-E (see Fig1). During this 'ocean forecast', the sea surface temperature is relaxed towards persisted SST, with a damping rate of 100 W/$m^2$/K.

The monthly forecasting system is run 51 times from slightly different initial conditions. One forecast, called the control forecast, is run from the operational ocean and atmosphere ECMWF analyses. The 50 additional integrations, the perturbed members, are made from slightly different initial atmospheric and oceanic conditions, which are designed to represent the uncertainties inherent in the operational analyses. The atmospheric component of the coupled GCM is perturbed in the same way as in EPS. The 50 perturbations are produced using the singular vector method (Buizza and Palmer, 1995). These include perturbations in the extratropics and perturbations in some tropical areas by targeting tropical cyclones (Puri et al, 2001). In addition, in order to take into account the effect of uncertainties in the model formulation, the tendencies in the atmospheric physics are randomly perturbed during the model integrations. The current implementation is the same as that used in EPS. For each ensemble member, the stochastic physics (Buizza et al 1999, Palmer 2001) perturbs grid point tendencies of the physics up to 50%. The tendencies are multiplied by a random factor drawn from an uniform distribution between 0.5 and 1.5. The random factor is constant within a $10^o$x$10^o$ domain, for 6 hours. The whole globe is perturbed.

The oceanic initial conditions are perturbed in the same way as in seasonal forecasting. A set of SST perturbations has been constructed by taking the difference between 2 different weekly-mean SST analyses (Reynolds OI and Reynolds 2DVAR) from 1985 to 1999. A second set of SST perturbations has been constructed by taking the difference between Reynolds 2DVAR SSTs and its 1-week persistence. The first set of SST perturbations samples the uncertainties in the SST analysis, whereas the second difference samples the uncertainties due to the fact that the SSTs from NCEP are a weekly-mean product. For each starting date, 25 combinations from these 2 different sets of perturbations are randomly selected and are added to the SSTs produced by the operational ocean analyses with a + and - sign, creating 50 perturbed initial states. In order to have a 3D structure, the SST perturbations are linearly interpolated to an oceanic depth of 40 metres. A set of wind stress perturbations is also calculated by taking the difference between 2 monthly wind stress analyses (ERA15/ECMWF analysis

and a wind analysis from Southampton Oceanic Centre) from 1980 to 1997. Five ocean assimilations (1 control and 4 perturbed) are produced by randomly picking 2 perturbations from the set of wind stress perturbations for each month of data assimilation and adding them with a + or - sign to the analyzed wind stress. The same perturbations cannot be chosen for 2 consecutive months. The wind stress perturbations and SST perturbations are combined to produce the 50 perturbed oceanic initial conditions.

After 10 days of coupled integrations, the model drift begins to be significant. It displays similar patterns to seasonal forecasting after 6 months of integrations, but with less amplitude. The strategy for dealing with model drift is the same as in seasonal forecasting. No 'artificial' terms are introduced to try to reduce the drift of the model and no steps are taken to remove or reduce any imbalances in the coupled model initial state: we simply couple the models together and start to integrate forward. The effect of the drift on the model calculations is estimated from previous integrations of the model in previous years (the back-statistics). The drift is removed from the model solution during the post-processing. In the present system, the climatology (back-statistics) is a 5-member ensemble of 32-day coupled integrations, starting on the same day and month as the real time forecast for each of the past 12 years. For instance, the first starting date of the real-time forecast was 27 March 2002. The corresponding climatology is a 5-member ensemble starting on 27 March 1990, 27 March 1991, ..., 27 March 2001. The 5-member ensemble is thus integrated with 12 different starting dates. This represents a total of 60 integrations and constitutes the 60-member ensemble of the back-statistics. The size of the back-statistics ensemble (60) is of same order as the size of the real-time forecast ensemble (51). The back statistics are created every 2 weeks, alternately with the real-time forecast. During the first week, the back-statistics are created. On the Wednesday of the second week, the corresponding real-time forecast is created. The back-statistics are therefore ready before the real-time forecasting suite starts. For seasonal forecasting, the back statistics are created once and all the real-time forecasts use this climatology for calibration. This is possible because the IFS cycle is the same throughout the lifetime of a seasonal forecasting system. However, for monthly forecasting, the IFS is updated several times per year, which means that the back-statistics needs to be generated for each real-time forecast, and one cannot use the back-statistics of the previous year real-time forecast.

Monthly forecasting products include anomaly, probability and tercile maps produced in a way similar to seasonal forecasting products. However, instead of being based on monthly-means as in seasonal forecasting or instantaneous values as in EPS, they are based on weekly means. It is likely that at that time range, the model has more skill in predicting weekly anomalies than daily values. Fields like surface temperature, 2-metre temperature, precipitation and mean-sea level pressure, have been averaged over 7 days. The 7-day periods correspond to days 5 to 11, days 12 to 18, days 19 to 25 and days 26 to 32. They have been chosen so that they correspond to Sunday to Saturday calendar weeks (the monthly forecasting starting date is on Wednesday). The length of the monthly forecasting system is 32 days, so that it contains 4 of these periods.

Anomaly maps are produced by averaging atmospheric fields over each of the weekly periods and the 51-member ensemble. The plots display the difference between the ensemble mean of the real-time forecast and the ensemble mean of the back-statistics. The product therefore displays the shift of the forecast ensemble mean from the estimated 'climatological' mean (created from ensemble runs over the past 12 years). Probability and tercile maps are also produced by comparing the 51 member ensemble distribution of the real-time forecast to the 60-member ensemble of the model climatology. Figure2 displays an example of a tercile map for forecasts starting on 12 March 2003.

The monthly forecasting system has been run in 'semi-operational' mode since 27 March 2002. This represents a total of 30 real-time cases. The rest of this paper will discuss the verification of these 30 cases. The analysis used to verify the monthly forecasting system is the ECMWF operational analysis for the real-time forecasts and ERA40 reanalysis for the back-statistics. To verify the monthly forecasts of precipitation, the operational or ERA40 forecasts of precipitation between 12 and 36 hours were used as verification data.
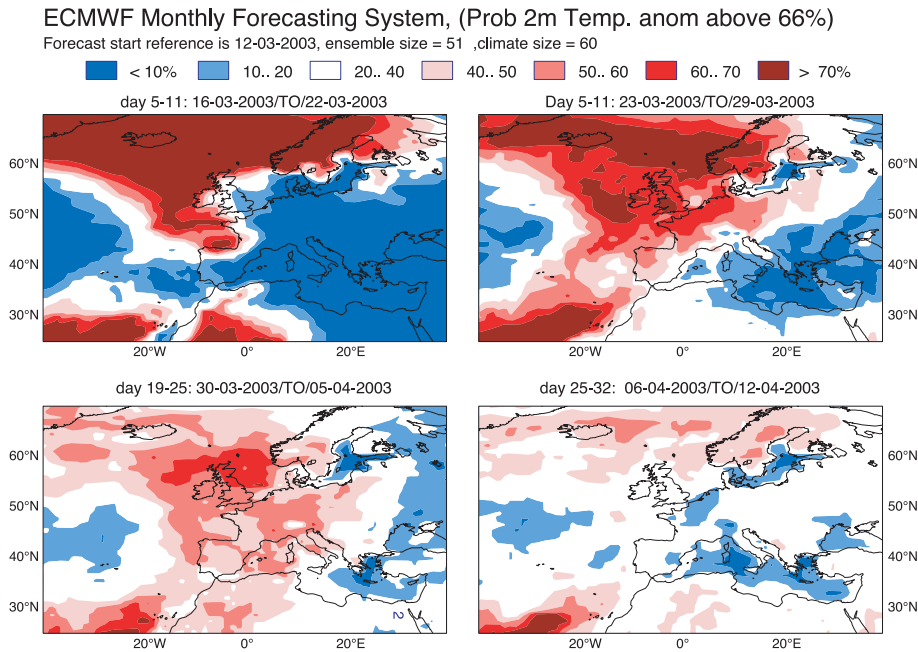
*Figure 2:   Probability of 2-metre temperature anomaly predicted by the monthly forecasting system to be in the upper tercile. Each panel represents one 7-day period. The forecast starting date is 12 March 2003.*

# 3   Evaluation of forecasts up to day 10. A comparison with EPS

In the present section, the skill of the monthly forecasting system during the first 10 days of integrations will be compared to the skill of the EPS. The monthly forecasting system starts every two weeks at 00Z, and since a 51-member ensemble integration of the EPS is realized every day at 00Z (in addition to the 12Z run), it is possible to compare the monthly forecasting system and the EPS over the first 10 days of forecasts. The main differences between the two systems are:

- The monthly forecasting system has a coarser resolution than EPS ($T_L$159 instead of $T_L$255).

- The monthly forecasting system has a longer time step than EPS ( 1 hour instead of 45 minutes).

- The monthly forecasting system is an ocean-atmosphere fully-coupled system, whereas the EPS is based on atmospheric integrations forced by persisted SSTs.

Otherwise, both systems share the same version of IFS and the same atmospheric initial conditions and perturbations (singular vectors + stochastic physics). The first part of this section will compare the anomaly correlation and root mean square error of geopotential height at 500 hPa of the monthly forecasting system and EPS. In the second part, probabilistic scores (ROC and Brier scores) of temperature at 850 hPa and precipitation will be compared.

## 3.1   Deterministic scores

The anomaly correlation and root mean square error of geopotential height at 500 hPa anomalies from the ensemble mean of the monthly forecasting system and the EPS starting at 00Z have been computed every 12 hours. The scores have been averaged over all the 30 cases (Fig3). According to Figure3, EPS seems to perform

slightly better than the monthly forecasting system, probably because of its finer horizontal resolution, but the difference is small. The deterministic scores obtained with the monthly forecasting system and the EPS look similar over the Northern Extratropics. The anomaly correlation of the monthly forecasting system is 0.6 by about day 8, as with EPS.

The scores presented in Figure3 have been calculated over the whole Northern Extratropics. Figure4 displays the anomaly correlations over four different regions in the Northern Extratropics: Atlantic west-Europe, Europe, North Pacific and North America. Although there is some slight variation from one region to another, there is not much difference in the anomaly correlation scores between the monthly forecasting system and EPS. Overall, EPS performs slightly better than the monthly forecasting system, except maybe over Europe and North America after day 7. Over North America, the anomaly correlation reaches 0.6 one day after EPS, but the Kolmogorov-Smirnov test indicates that the difference is not significant within the 90% level of confidence.

## 3.2  Probabilistic scores

Probabilistic scores of temperature at 850 hPa and precipitation obtained with the monthly forecasting system and the EPS have been compared for each of the first 10 days of the forecasts. As for the deterministic scores, there is little difference in the probabilistic scores obtained with the EPS and the monthly forecasting system. EPS consistently performs better than the monthly forecasting system, but, as for the deterministic score, the difference is small (see examples in Fig5 and 6 which show the ROC scores for days 5 and 10 over the Northern Extratropics of the probability that the temperature anomaly at 850 hPa is larger than 2K and the reliability diagram for days 5 and 10 over the Northern Extratropics of the probability that the temperature anomaly at 850 hPa is larger than 8K). This conclusion is valid for all the events we tested and is also true for precipitation (not shown). Therefore, it does not seem to depend strongly on the variable or the event. The same result is obtained when looking at different regions, such as Europe, North America, Tropics or Southern Extratropics or different seasons. Overall, it can be concluded that the monthly forecasting system behaves similarly to EPS during the first 10 days of forecast. However, EPS is likely to perform significantly better than the monthly forecasting system for the prediction of extreme events, such as storms, where the horizontal resolution of the atmospheric model has a significant impact on the forecasts.

## 3.3  Impact of ocean-atmosphere coupling

During the first 10 days of forecasts, the monthly forecasting system behaves much like EPS (see the example in Fig7). Both deterministic and probabilistic scores are somewhat lower than EPS, which is likely due to the difference of horizontal resolution between both systems. At this time range, the ocean-atmosphere coupling does not seem to have a significant impact on the forecasts. However, it may have an impact on the development of some synoptic scale events such as tropical cyclones. To fully address this question, a clean comparison would be needed, where both systems share the same horizontal resolution and time step so that their only difference would be the ocean-atmosphere coupling. Such an experiment has been set up, and will be completed within a few months The results will be published in a forthcoming technical memorandum.

After 10 days of integrations, the ocean state may display enough variance to impact significantly the forecasts. Several papers have pointed out the potentially important role of ocean-atmosphere coupling on the propagation of the Madden-Julian oscillation (see for instance Inness et al, 2003), which is likely to have an impact on the extratropics after 10 days. In order to evaluate the impact of the ocean-atmosphere coupling on the forecasts, the control of the monthly forecasting system has been compared to the control of the EPS. The 50 perturbed members of the EPS are integrated for only 10 days, but the control forecast of the EPS extends to day 20. Anomaly correlations have been applied to weekly-mean anomalies of geopotential height at 500 hPa. The two
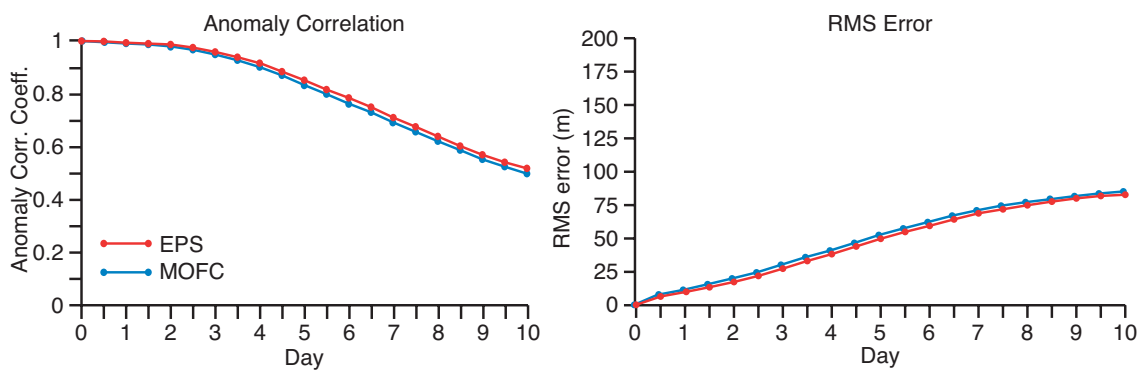
*Figure 3: Anomaly correlation (left panel) and root mean square error (right panel) over the Northern Extratropics as a function of the time-lag (in days). The blue line represents the scores of the ensemble mean of the monthly forecasting system (MOFC), and the red line represents the scores of the ensemble mean of the ensemble prediction system (EPS). The EPS seems to perform slightly better than the monthly forecasting system during the first 10 days of the forecast, but the difference is small.*
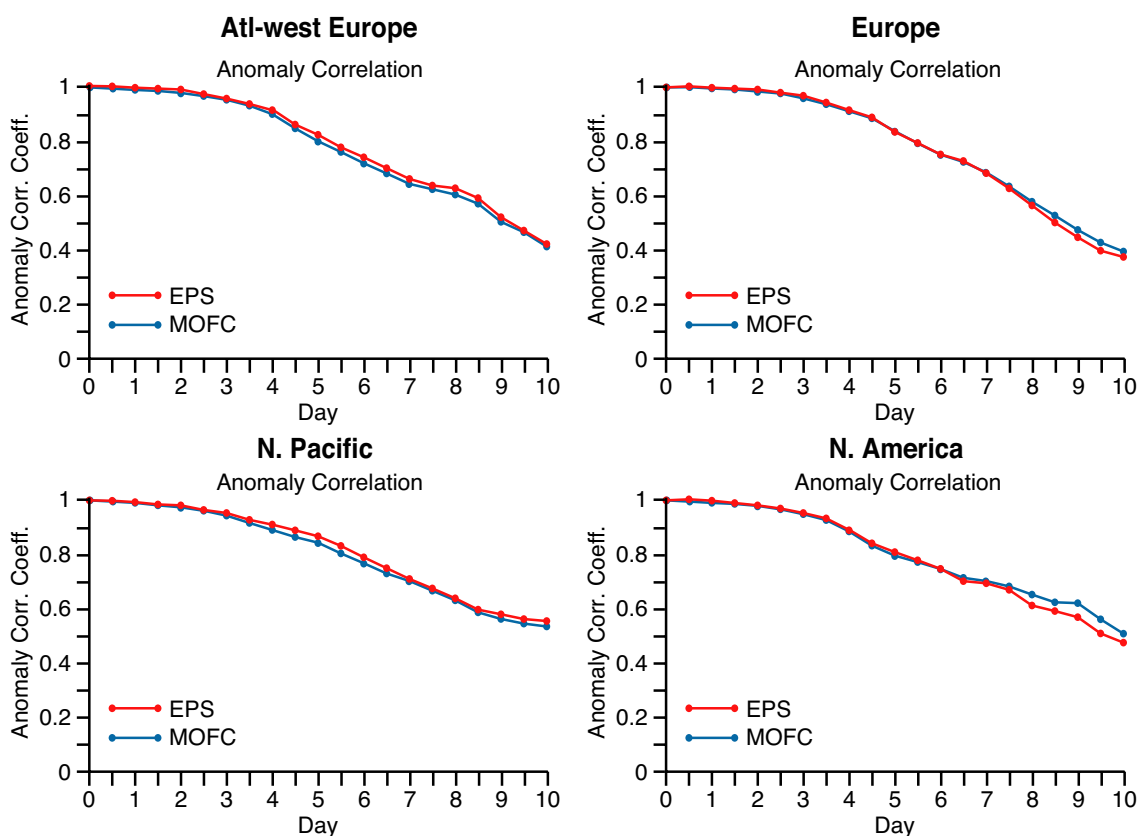


*Figure 4: Anomaly correlation as a function of time-lag over four different regions: Atlantic Western-Europe (top left), Europe (top right), North Pacific (bottom left) and North America (bottom right). The blue lines represent the scores of the monthly forecasting system (MOFC), and the red line represents the scores of the ensemble prediction system (EPS).*
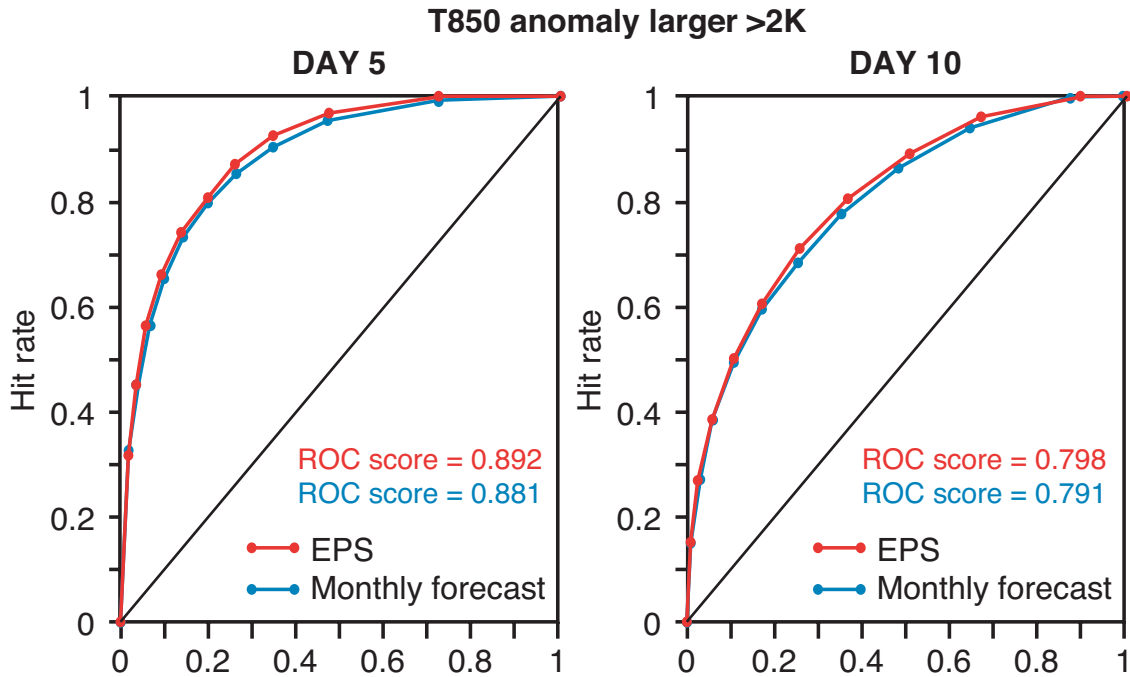
**T850 anomaly larger >2K**



*Figure 5: ROC diagrams of the probability that the 850 hPa temperature anomaly exceeds 2K for day 5 (left panel) and day 10 (right panel). The blue line represents the ROC diagram obtained with the monthly forecasting system. The red line represents the ROC diagram obtained with the EPS.*
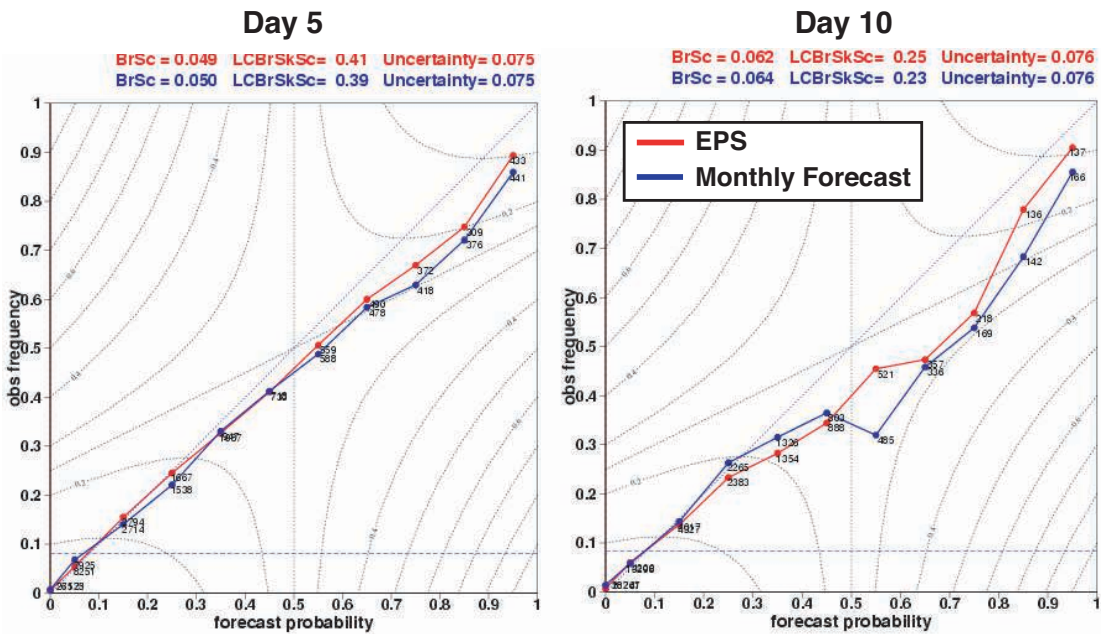


*Figure 6: Reliability diagrams of the probability that the 850 hPa temperature anomaly exceeds 8K for day 5 (left panel) and day 10 (right panel). The blue line represents the reliability diagram obtained with the monthly forecasting system. The red line represents the ROC diagram obtained with the EPS.*
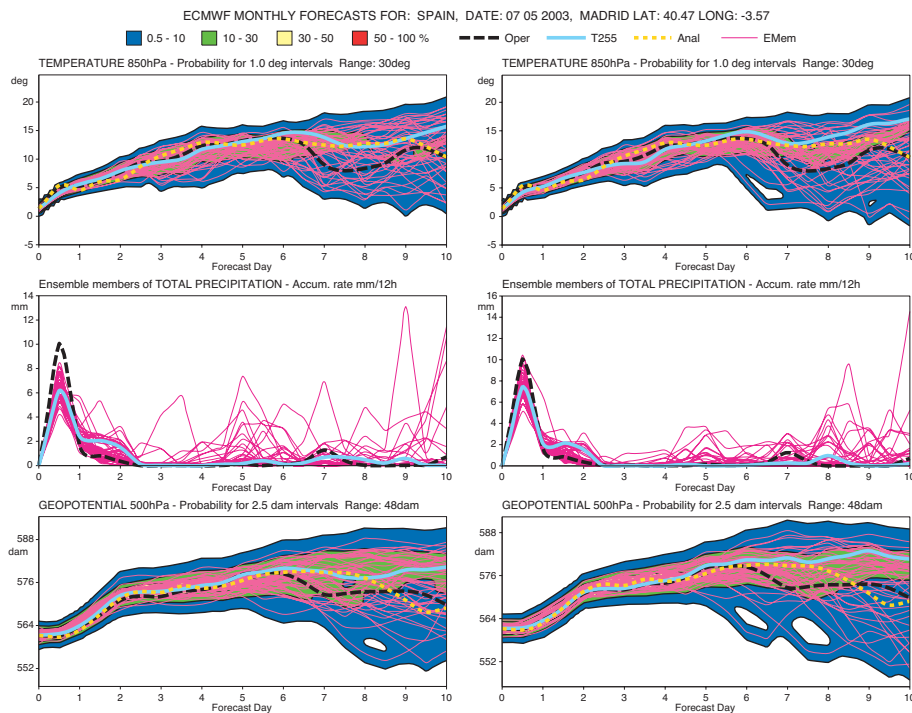
*Figure 7: Plume diagram of 850 hPa temperature (top panels), precipitation (middle panels) and 500 hPa geopotential height (bottom panels) over Madrid for the forecasts starting on 7 May 2003 at 00Z. The left panel represents the EPS forecast, and the right panel represents the monthly forecast. Each magenta line represents one member ensemble forecast. The dark blue shaded area represents the spread of the distribution. The dotted orange line represents the verification from the operational analysis. The continuous cyan line represents the control forecast, and the dashed black line represents the operational deterministic forecast.*

weekly periods considered are days 5 to 11 and days 12 to 18. Figure 8 displays a scatter-plot diagram of the anomaly correlations obtained with the EPS control (y-axis) and with the monthly forecasting system control (x-axis) over the Northern Extratropics. The left panel represents the scores obtained over the period days 5-11. During that period of time, the control EPS performs better than the monthly forecasts in 66% of cases. The difference is significant according to the Kolmogorov-Smirnov test (Knuth, 1981) (KS-test), but it is not very large. This is consistent with the results presented above concerning the 10 first days of the forecast.

Over the period days 12-18, there is no significant difference in the scores obtained with the control EPS and the control monthly forecast over the Northern Extratropics. The monthly forecasting system wins in 16 cases, but loses in 14 cases. However, over some regions like the North Pacific, or East Asia (Fig9), the monthly forecasting system seems to perform significantly better than the EPS over the period days 12-18. Over the North Pacific, the monthly forecasting system wins in 20 cases and loses in 10 cases, and a KS-test indicates a significance of 92%. Over East Asia, the difference is even stronger, with the monthly forecasting system winning 24 cases and losing only 6 cases (significance of 95% according to the KS-test). Over these regions, EPS performs better than the monthly forecasting system during the previous week (period days 5-11). This result suggests that after 10 days of forecast, the ocean-atmosphere coupling may have a significant impact on the forecasts. The fact that this impact is particularly visible over East Asia and the North Pacific could be because these regions are closer to the tropical Pacific and therefore more sensitive to variations in tropical Pacific SSTs and Section 7 will show that the monthly forecasting system has skill in predicting the variations of tropical Pacific SSTs. This result suggests that using a coupled ocean-atmosphere system is a good strategy

**DAY 5-11**

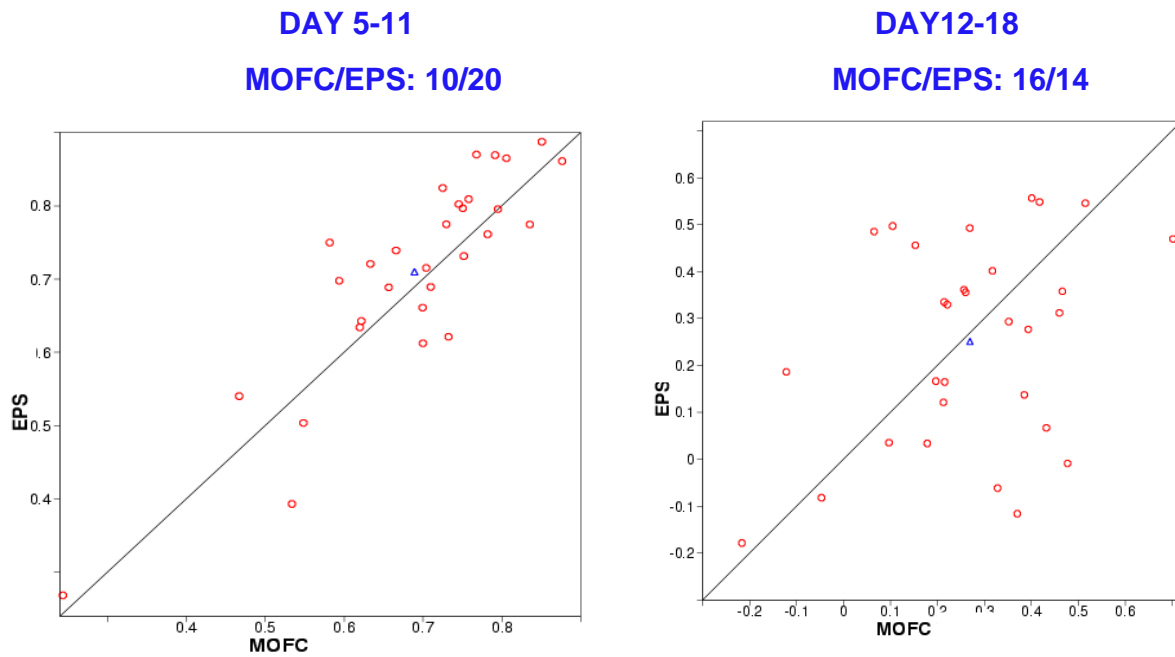**MOFC/EPS: 10/20**

**DAY12-18**

**MOFC/EPS: 16/14**

*Figure 8:  Scatter plot diagram of anomaly correlation of 500 hPa geopotential height over the Northern Extratropics averaged from day 5 to day 11 (left panel) and from day 12 to 18 (right panel). The x-axis represents the scores obtained with the monthly forecasting system and the y-axis represents the scores obtained with EPS over the Northern Extratropics.*
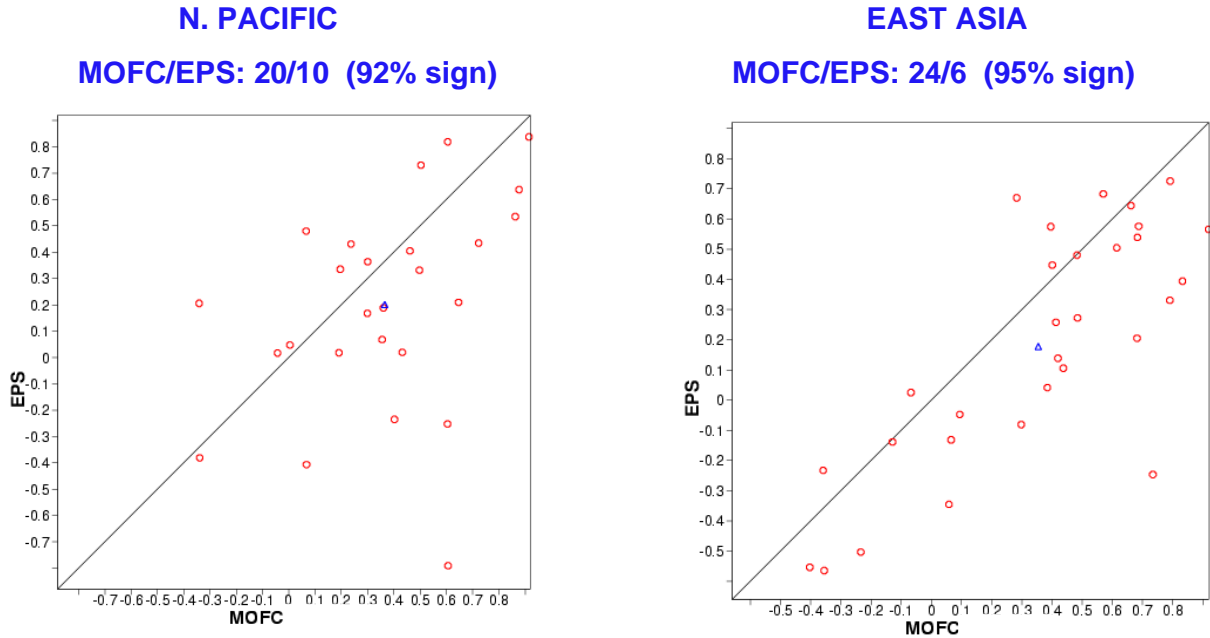
**N. PACIFIC**

**MOFC/EPS: 20/10  (92% sign)**

**EAST ASIA**

**MOFC/EPS: 24/6  (95% sign)**

*Figure 9:  Same as in Figure8, but the left panel represents the scatter plot diagram for the Northern Pacific and the right panel the scatter plot diagram for Asia.  Both are scatter plot diagrams of 500 hPa geopotential height anomaly correlation averaged from day 12 to 18.*

for producing forecasts after day 10.

# 4 Evaluation of the forecasts beyond day 10

The monthly forecasting system has been designed to evaluate the predictability after 10 days of forecast. The present section will evaluate the skill of the monthly forecasting system after day 10.

## 4.1 Deterministic scores

The deterministic scores used operationally for EPS have been extended to the 32-day integrations of the monthly forecasting system. Figure 10 displays examples of anomaly correlation and RMS scores applied to the monthly forecast starting on $1^{st}$ January 2003. In this particular case, the anomaly correlation for most members of the ensemble and also for the ensemble mean reaches 0.6 by day 8. The spread of the ensemble is increasing quickly after day 6 and seems to saturate after about 12 days of forecasts. The top panel represents the 'perfect model' case, where the control forecast is assumed to be the 'truth'. The anomaly correlations and RMS errors are quite close to those obtained in the bottom panels. In the 'perfect model' case, the 0.6 anomaly correlation is reached between day 8 and day 14. If we consider 0.6 as the limit of deterministic usefulness of the system, this result suggests that, in this system, the limit of potential deterministic predictability probably does not exceed 15 days. Not surprisingly, the ensemble mean performs better than individual members.

For all 30 cases, the anomaly correlation and RMS scores of the ensemble mean have been calculated. Figure 11 displays the scores of the ensemble mean for all the 30 cases over the Northern Extratropics. Each line represents one individual case, and the dark line represents the mean over the 30 cases. The 0.6 anomaly correlation is reached between day 7 and day 13, and on average by day 8. The linear correlation diminishes quite sharply after day 8 and reaches 0.3 around day 13. The RMS error reaches climatology around day 14, suggesting that there is little deterministic skill in the monthly forecasting system after about 15 days of forecast.

There is some variability depending on the geographical area (Fig. 12). For instance, over Europe, the anomaly correlations are decreasing much more quickly than over the Northern Extratropics as a whole. The anomaly correlation reaches 0.3 shortly after day 10, and reaches 0 after about day 20. The RMS error reaches climatology after just 11 days of forecasts. Over the North Pacific, on the other hand, the anomaly correlation drops much more slowly with time, reaching 0.3 around day 15 and 0 around day 28. The RMS error over the North Pacific reaches climatology around day 18, which is almost one week later than over Europe. There also seems to be some variability depending on the season. The monthly forecasting system has been operational for just a bit more than one year, so it is difficult to give definitive conclusions on the seasonality of the scores. Nevertheless, Figure 13 suggests that the model performed significantly worse during the Summer 2002 till day 10, which is consistent with the fact that summer is generally a more difficult season for medium-range forecasting over the Northern Extratropics. However, deterministic scores were higher in Summer 2002 after about 15 days of forecasts than in other seasons. After 15 days of forecasts, the performance was slightly worse in Spring 2002 than in other seasons.

All the results presented above are based on daily values. However, after 10 days of forecasts, the model has probably more skill in predicting changes of regimes and the variability of weather over a period of several days. In the present study, we consider weekly-mean system seem to be better than those obtained by persisting the anomaly of the same as the time periods used for the products (see Section 2). Since the starting date is a Wednesday, these weekly periods correspond to a Sunday to Saturday week. The first period corresponds roughly to the last week of the EPS period, whereas the second, third and fourth periods are beyond day 10.
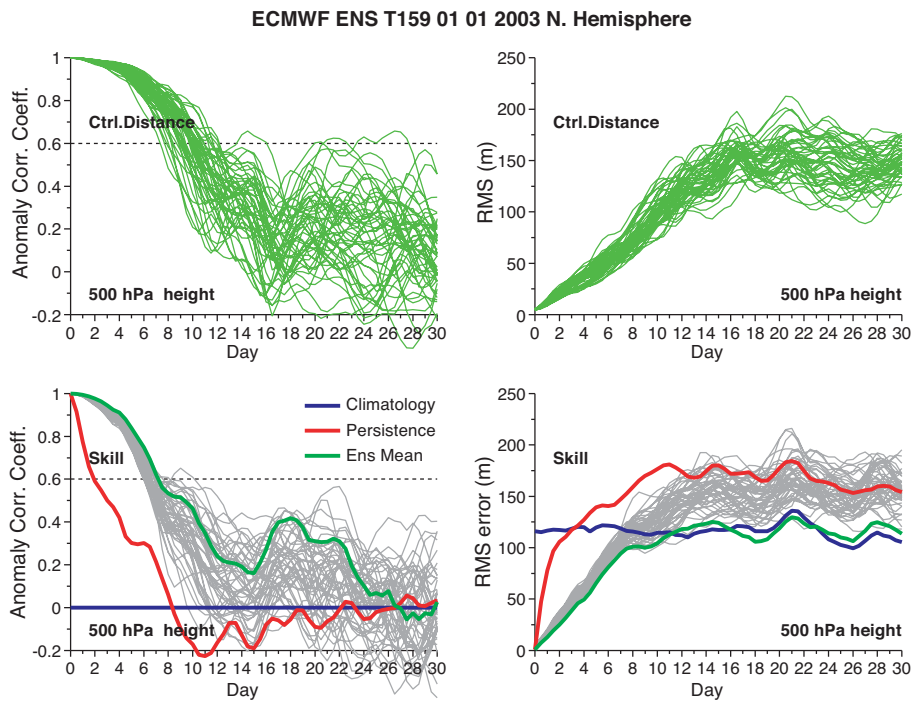
**ECMWF ENS T159 01 01 2003 N. Hemisphere**



*Figure 10: Anomaly correlation (left panel) and RMS error for the monthly forecasting system starting on 1$^{st}$ January 2003. The top panels represent the deterministic scores obtained by considering the control forecast as being 'the truth'. In this panel, each green line represents one member of the ensemble. The bottom panels display the scores relative to the operational analyses. In these panels, each grey line represents one member of the ensemble. The green line represents the scores obtained with the ensemble mean. The red line represents the scores obtained by persisting the initial condition.*

**ECMWF ensemble forecasts – ensemble mean**



*Figure 11: Anomaly correlation (left panel) and root mean square error (right panel) of 500 hPa geopotential height calculated over the ensemble mean. Each red line represents the time evolution of the deterministic scores for one individual case. The heavy red line represents the mean over all the cases. The blue line represents the score obtained with climatology.*
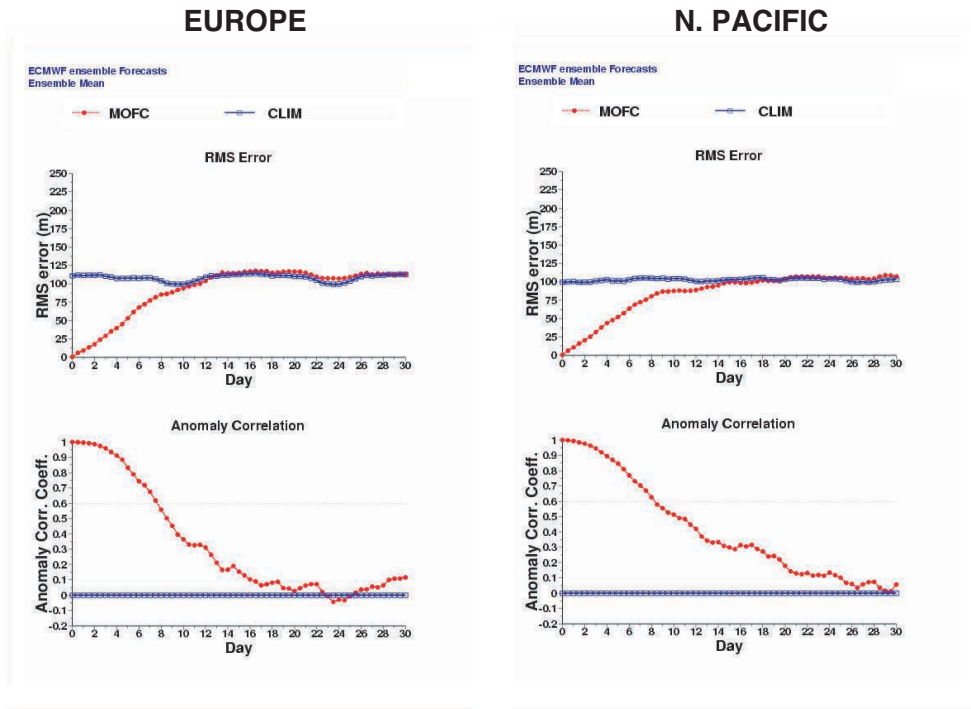
*Figure 12: Anomaly correlation (bottom panels) and root mean square error (top panels) of 500 hPa geopotential height calculated over the ensemble mean and averaged over the 30 cases (red line). The blue line represents the scores obtained with climatology. The left panels represents the deterministic scores over Europe, and the right panels the scores over the North Pacific.*
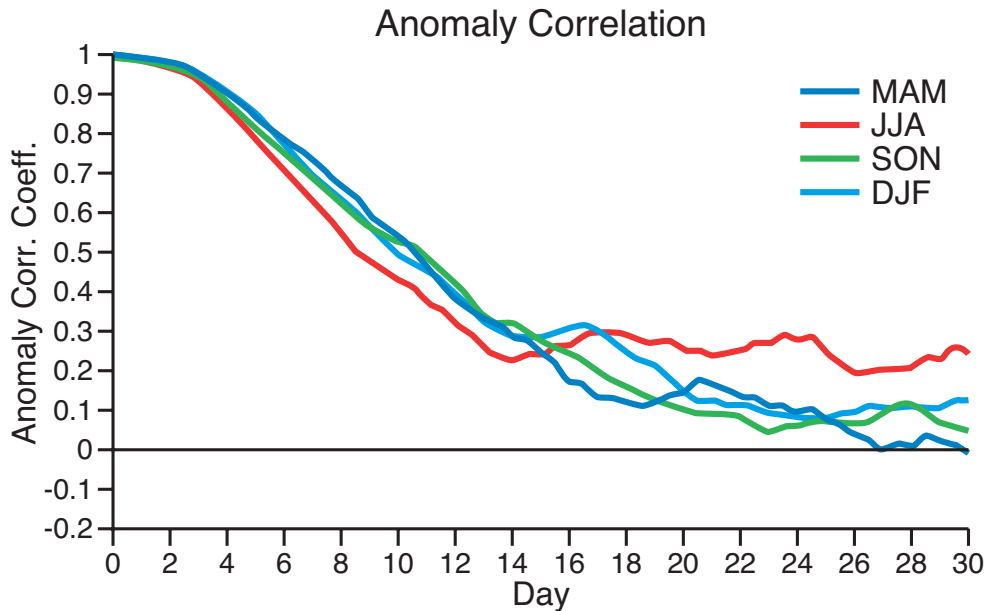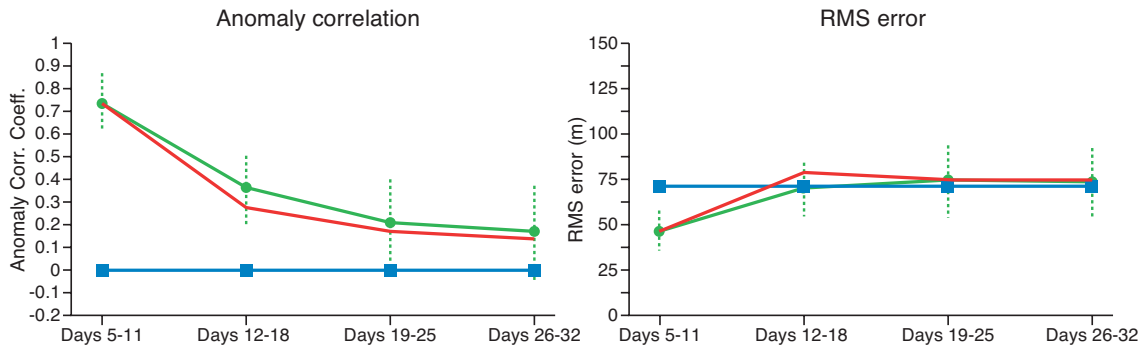


*Figure 13: Time evolution of the anomaly correlation of 500 hPa geopotential height over the Northern Extratropics calculated over the ensemble mean and averaged over the four seasons. Dark blue represents spring, red represents summer, green represents autumn and cyan represents winter.*

*Figure 14:  Anomaly correlation (left panel) and root mean square error (right panel) of weekly-averaged 500 hPa geopotential height calculated over the ensemble mean and averaged over the 30 cases. The green line represents the score of the monthly forecasting system. The dashed green line represents 2 standard deviations. The red line represents the scores obtained by persisting the 500 hPa geopotential height anomaly of the previous week and the blue line represents the score obtained with climatology.*

The anomaly correlation and RMS error have been calculated on this weekly-mean basis (Fig.14). The anomaly correlation is of the order of 0.75 for days 5-11, about 0.35 for the days 12-18, about 0.2 for the days 19-25 and 0.15 for the days 19-32. The RMS error plot suggests that the RMS error of the monthly forecasting system reaches climatology by day 12-18. This confirms that for forecasts beyond day 10, the anomaly correlation is usually much lower than 0.6 and therefore, deterministic forecasts may not be very useful in this time range. However, the deterministic scores of the monthly forecasting system seem to be better than those obtained by persisting the anomaly of the previous week (red curve in Fig.14). This suggests that although the deterministic skill of the monthly forecasting system is moderate after 10 days of forecasts, it is still higher than persistence, implying that the monthly forecasting system may be useful at this time range.

Another way of filtering the data consists in using a spatial filter instead of time averaging. The geopotential at 500 hPa has been filtered using different spectral truncations: T40, T20, T10, T8, T6 and T5. Fig.15 displays an example of geopotential at 500 hPa filtered at these different spectral resolutions. Till T8, the field still resembles the original field but at lower truncations (T5 or T6) most of the synoptic features have disappeared. Anomaly correlations have been computed at different horizontal truncations (Fig.16). The higher the filtering, the more skillful is the forecast, which is consistent with what would be expected: the model has more skill in predicting larger features in the synoptic scale circulation than details. However, the gain is not that large. The gain in predictability between T40 and T8 is about 2 days. With T5, an additional 2 days in predictability can be gained, but at this resolution the forecast may not be very useful ( Fig.15). After 15 days of forecast, there is little daily deterministic skill at any horizontal resolution. Even at T5, the anomaly correlation is below 0.4 after 15 days of integration.

### 4.1.1   Time series

As mentioned in Section 2 , the back statistics of the monthly forecasting system consist of 5-member ensemble integrations over the past 12 years. Although the size of the ensemble is quite small, the back-statistics give some idea on how the monthly forecasting model can predict the interannual variability of different fields. For instance, Figure17 shows the interannual variability of surface temperature anomalies (relative to the model climate computed from the 11 other years of the model climate) averaged over each of the four weekly periods over Northern Europe, and for the forecasts starting on 9 April 1991 to 2002. The red dotted line represents analysis. In that particular example, the model shows some strong skill in the medium-range (days 5-11). The
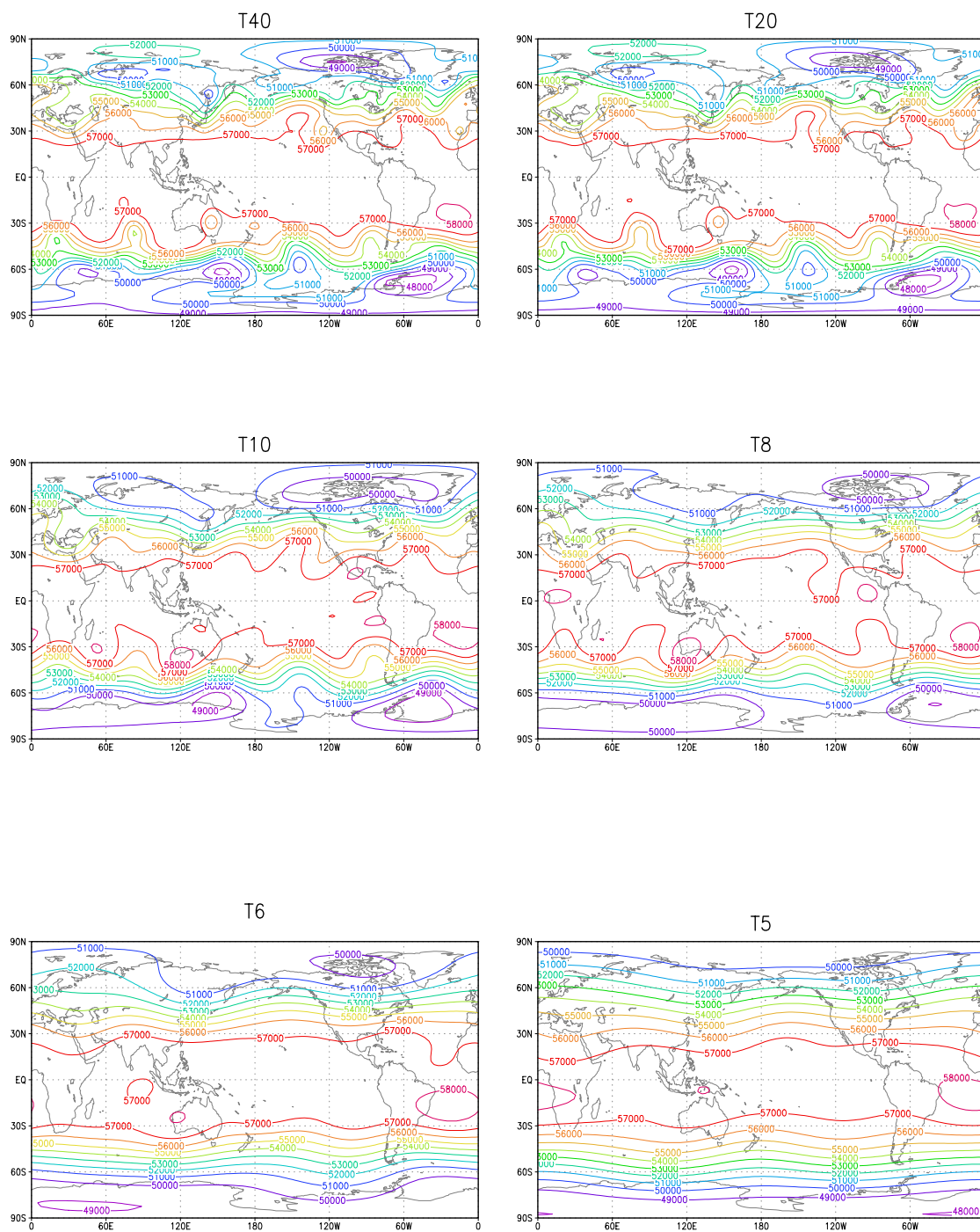
*Figure 15: An example of geopotential at 500 hPa with different spectral resolutions (T40, T20,T10,T8,T6 and T5).*
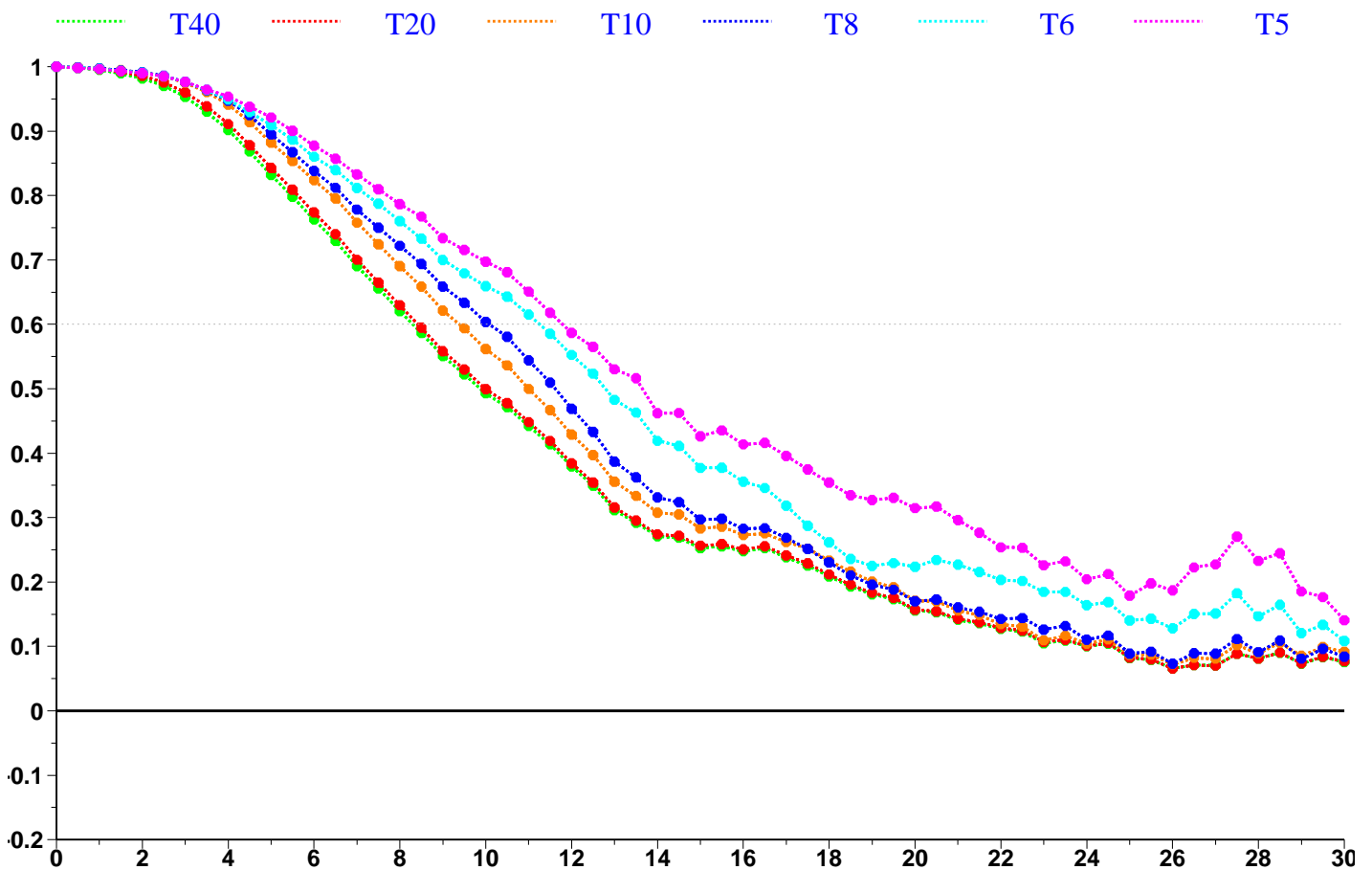
*Figure 16: Time evolution of the anomaly correlation of the ensemble mean of geopotential at 500 hPa over the Northern Hemisphere (North of 20N). The different curves correspond to the anomaly correlations obtained with the model truncated at different spectral resolutions (T40, T20,T10,T8,T6 and T5). T1 has been removed.*

linear correlation is 0.39 in the second period (days 12-18), which is significantly higher than the interannual variability obtained by persisting the surface temperature anomalies of the previous week (correlation of 0.03). Over some regions, like the north-eastern part of US (Fig.18), the monthly forecasting system displays some strong skill in predicting the interannual variability of surface temperature anomalies in the period days 12-18 (correlation of 0.7) and also to a lesser extent in the period days 19-25 with a correlation of 0.38. In the case of the north-eastern part of US, the monthly forecasting system again beats persistence of the anomalies in the previous week for days 12-18 and days 19-25. This result is of course dependent on the geographical location and on time of the year, but it suggests that at times, the model can display strong deterministic skill in the second weekly period.

## 4.2 Probabilistic scores

The deterministic score suggests that after 10 days, the monthly forecasts are generally not useful for issuing deterministic forecasts. The anomaly correlations after day 10 are below 0.6. However, after day 10, the forecasts are essentially probabilistic. In the rest of this section, the probabilistic scores of the monthly forecasting system will be evaluated through the scores obtained with surface temperature, 2-metre temperature, precipi-
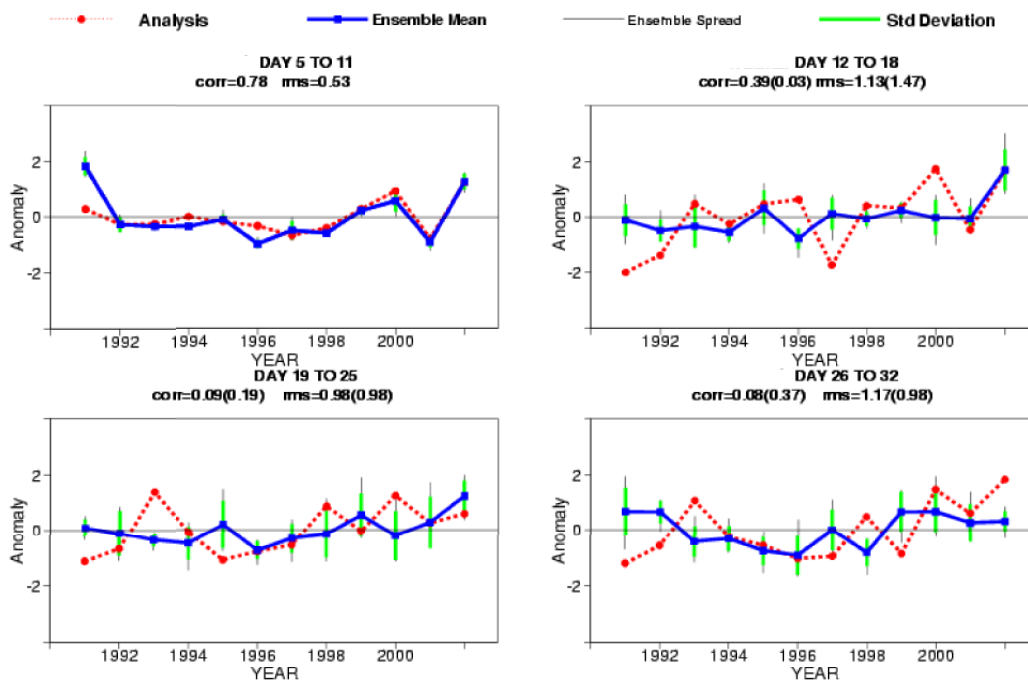
*Figure 17: Interannual variability of surface temperature anomalies over Northern Europe from 1991 to 2002. Forecasts start on 9 April. The blue line represents the mean of the 5 member monthly forecast ensemble, the red line represents the analyzed interannual variability obtained from ERA40. The green vertical line represents one standard deviation, and the thin vertical black line the full spread of the ensemble. The surface temperature anomalies have been averaged over weekly periods. In the top left panel, the surface temperature has been averaged from day 5 to day 11; in the top right panel from day 12 to day 18; in the bottom left panel from day 19 to day 25 and in the bottom right panel from day 26 to day 32. Linear correlations and RMS error between the interannual variability obtained from ERA40 (red curve) and the ensemble mean (blue curve) are indicated in top of the figures. The numbers in parentheses indicate the scores obtained using the persistence of the ensemble mean anomaly of the previous week.*
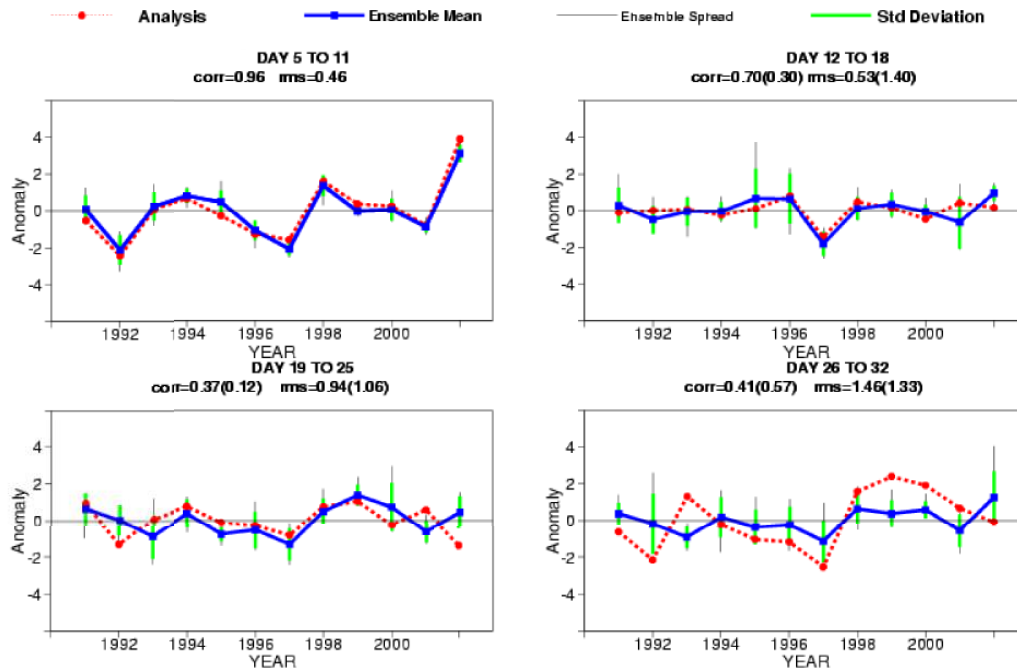
*Figure 18:  Same as Figure17, but for the north-eastern part of the US*

.

tation and mean sea level pressure.  Figure19 displays the ROC diagrams obtained for different periods: days 5-11, days 12-18, days 19-25, days 26-32 and days 19-32 (the two last weeks of the forecast). The event is the probability that the surface temperature is in the upper tercile, over each grid point of the Northern Extratropics.  Only grid points over land are considered.  For the monthly forecast, the upper tercile has been computed relative to the model climatology.  In that respect, the systematic bias of the model has been taken into account. Figure19 displays how the probabilistic scores are decreasing week by week.  Over the first period (days 5-11), the ROC area is of order of 0.8, and drops to 0.7 in the next week.  It drops again to about 0.6 in the following week.  The ROC scores for days 19-25 and days 26-32 are close.  The ROC score for day 19-32 (the two last weeks of the forecast) is close to the score of days 19-25.

The rate at which the ROC score degrades from one week to another depends a lot on the variable considered (Fig.20).  For 2-metre temperature, the ROC area diminishes strongly between days 5-11 and days 12-18.  The scores for the three other periods are close to each other.  For mean-sea level pressure, the drop between days 5-11 to days 12-18 is larger.  There is also a significant drop between days 12-18 and days 19-25.  For days 26-32, the ROC area is almost 0.5, which suggests that there is no skill at this time range.  For precipitation, the ROC scores are lower than those obtained with the other variables, for all the periods and the ROC area for days 12-18 is about 0.6 only.

The ROC diagrams seem to display a significant seasonality (Fig.21), though, once again, the period of verification is too small to allow definitive conclusions.  However, for days 5-11, the monthly forecasting system performed significantly worse in summer 2002 than in the other seasons.  This is consistent with the deterministic scores discussed above, and also with the fact that summer is the most difficult season in the medium-range. In the following week (days 12-18), summer 2002 was also a difficult season, but spring appears to be significantly worse than winter and Autumn 2002.  During the last two weeks of the forecast (days 19-32), there was no significant difference in the scores for summer, winter and autumn 2002.  However the ROC scores in
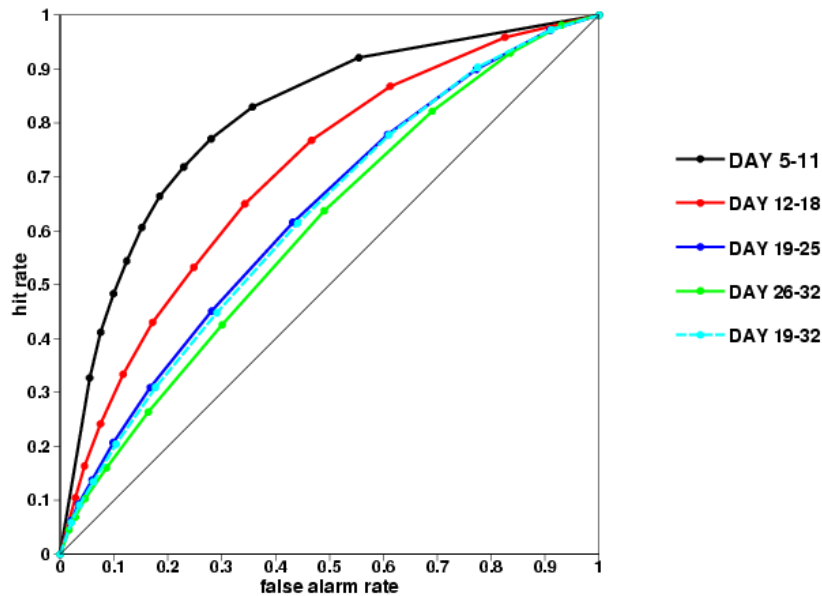
*Figure 19: ROC diagrams of the probability that the surface temperature predicted by the monthly forecasting system is in the upper tercile. The diagrams have been calculated over all the land grid points over the northern Extratropics and over the 30 cases. The black line represents the diagram for the surface temperature averaged from day 5 to day 11, the red line represents the period from day 12 to day 18, the dark blue line represents the period from day 19 to day 25, the green line represents the period from day 26 to 32 and finally the cyan line represents the period from day 19 to day 32.*

spring 2002 were significantly worse than in the other seasons, which is consistent with seasonal forecasting. Although a much larger period of verification would be needed to confirm this result, this suggests that after 20 days of forecast, the monthly forecasting system behaves more like seasonal forecasting. The period from day 12 to day 18 looks more like a transitional period between medium-range and the seasonal time range.

In the rest of this section we will concentrate on the probabilistic scores over the periods days 12-18, which corresponds almost to the first week following the end of the EPS period and period days 19-32 which includes the last two weeks of the forecast.

## 4.3 Days 12-18

This period corresponds almost to the first week after the end of the EPS. At this time range the monthly forecasting system still produces a robust signal. Figure22 shows a typical forecast for this time range. It represents the 2-metre temperature anomaly (relative to the 12-year model climate) averaged from day 12 to day 18 predicted by the monthly forecasting system starting on 6 April 2003. The WMW-test has been applied in order to determine if the 51-member ensemble distribution of the real-time forecast is significantly different from the 60-member ensemble distribution of the model climatology. A Wilcoxon-Mann-Whitney test (WMW-test, see for instance Wonacott and Wonacott 1977) has been applied to estimate whether the ensemble distribution of the real-time forecast is significantly different from the ensemble distribution of the back-statistics. Regions where the WMW-test displays a significance less than 90% are blank. Regions where the WMW-test displays a significance exceeding 95% are delimited by a solid contour (blue or red depending on whether the anomaly is positive or negative respectively). For 2-metre temperature, most of the areas over land are shaded, suggesting that after 10 days of forecast, the model is still able to predict a significant signal (Fig.22). Figure23 represents a typical forecast of precipitation anomaly. For precipitation, which is

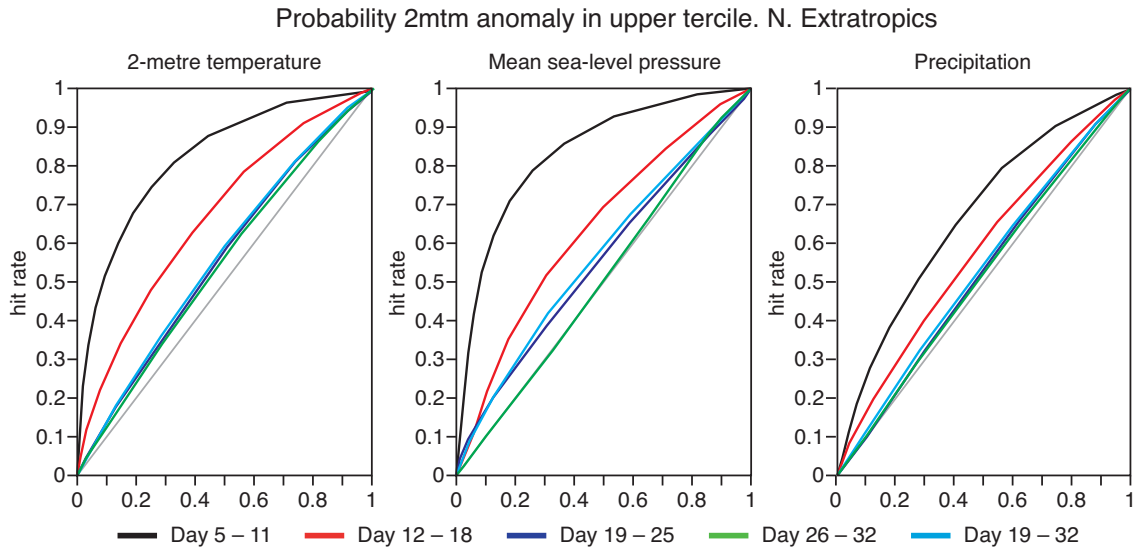Probability 2mtm anomaly in upper tercile. N. Extratropics



*Figure 20: Same as Figure19 for the probability that 2-metre temperature (left panel), mean sea level pressure (middle panel) and precipitation (right panel) predicted by the monthly forecasting system are in the upper tercile. The ROC diagrams have been calculated over all the land grid points over the northern Extratropics.*
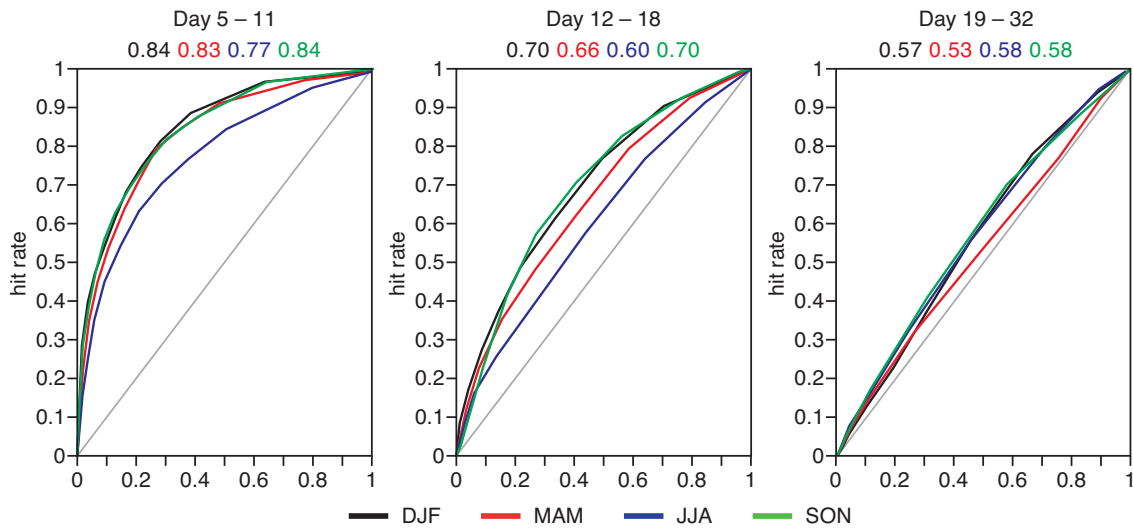


*Figure 21: ROC diagram of the probability that the surface temperature is in the upper tercile for different seasons. The red line represents the diagram for spring 2002, the blue line is for summer 2002, the green line is for autumn 2002 and the black line for winter 2002-2003. The left panel shows the scores for the surface temperature averaged from day 5 to day 11. The middle panel shows the scores for the surface temperature averaged from day 12 to day 18. The right panel shows the scores for the surface temperature averaged from days 19 to 32. The ROC diagrams have been calculated over all the land grid points over the northern Extratropics.*

a much noisier field than 2-metre temperature, blanked areas are much larger than with 2-metre temperature, particularly over some areas, like Europe. It is quite common that the monthly forecasting system does not predict a strong shift in the probability distribution of precipitation over that area. Over North America, South America, and East Asia, the model usually predicts a significant and large signal at this time range.

Figure24 displays the ROC area computed over each land grid point over the period March 2002-May 2003 (30 cases) of the probability that the 2-metre temperature is in the upper tercile. The upper tercile is computed relative to the model climatology, which means that the model systematic bias has been taken into account. According to Figure24, the model showed some skill during the verification period (ROC area larger than 0.5, which corresponds to the score of climatology) over most regions. Over Europe, the ROC area was around 0.6. The score is higher (larger than 0.7) over large portions of North-west America, South Africa and Australia. This figure suggests that overall, the model displays some skill over the period day 12-18. This is confirmed by the ROC diagram computed over the 30 cases and over all the land points in the whole Northern Extratropics (red curve in Figure25), at least for this particular event and period of verification. The score is of course much smaller than the scores obtained in the medium range, when ROC areas are usually larger than 0.8, but it is still significantly higher than 0.5, suggesting that the monthly forecasting system has at least some moderate skill at this time range. The forecast displays also some strong reliability: in Figure25, right panel, higher probabilities in the forecast correspond to a larger observed frequency, and the Brier skill score is positive, suggesting that the monthly forecasting performs better then climatology.

Although the monthly forecasting system displays some skill at days 12-18, if it is not more skillful than the persistence of the previous week, then it would be useless. Therefore, an important test for monthly forecasting consists in comparing its score with the persistence of the previous week of forecast anomaly (day 5-11). The blue curves in Figure25 correspond to the scores obtained when persisting the 2-metre temperature anomalies of the previous week. This is indeed close to the score obtained by persisting the last week of the EPS (see Section 2). Figure25 suggests that the monthly forecasting system performs better than persistence of the previous week. The ROC area is higher (0.67 instead of 0.62) and the forecast is significantly more reliable. The reliability diagram of persistence is much more horizontal than with the monthly forecasting system, which indicates less reliability and the Brier skill score of the persistence of day 5-11 is negative, suggesting that the forecast is less reliable than climatology. The potential economic value (see Richardson 2000 for example) also confirms that the monthly forecasting system has some value for a large range of cost-loss ratios and that it has more value than persisting the previous weekly anomalies (Fig.26).

In order to assess how significant the improvement of the monthly forecasting system is compared to persistence, scatter-plot diagrams of Brier and ROC scores have been plotted in Figure 27. In these plots, the x-axis represents the scores obtained with the monthly forecasting system over the period 12-18, and the y-axis represents the scores obtained with persistence of the previous week. Each circle represents one individual case (there are 30 cases). The Brier scores of the probability that the 2-metre temperature is in the upper tercile is always lower, and therefore more reliable, for the monthly forecasting system than for persistence, with one exception. This demonstrates that the difference of reliability between monthly forecasting and persistence seen in Figure25 is significant (99% significance according to the KS-test). The difference in ROC scores is also significant (92% significance according to the KS-test). However, there have been 7 cases (over a total of 30 cases) where persistence produced a larger ROC score over the Northern Extratropics than the monthly forecasting system. So the difference in ROC scores between the monthly forecasting system and persistence is not as strong as the difference in Brier scores.

The fact that the monthly forecasting system produces better forecasts than persistence for the period day 12-18 seems to be a robust result. Previous figures in this section have shown scores for one single event: the probability of 2-metre temperature in the upper tercile. However, the result does not depend strongly on the threshold. For instance, scoring the probability of 2-metre temperature weekly anomaly larger than
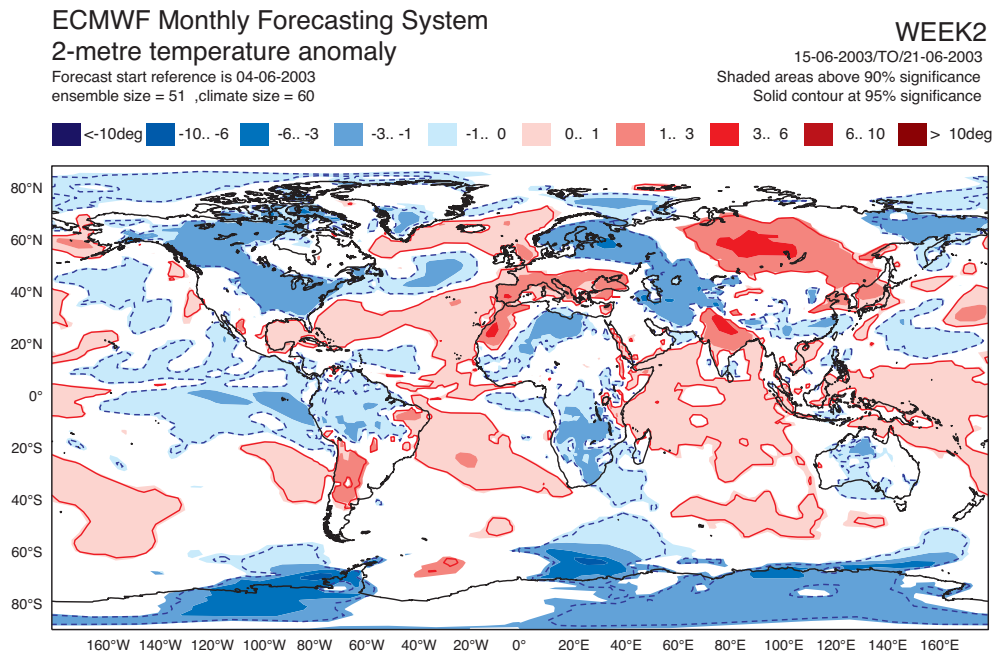
*Figure 22: Monthly forecast of 2-metre temperature averaged from day 12 to day 18. The starting date is the 4 June 2003. Blanked areas represent areas where the ensemble distribution of the real time forecast is not significantly different from the model climatological distribution with a level of confidence higher than 90%.*
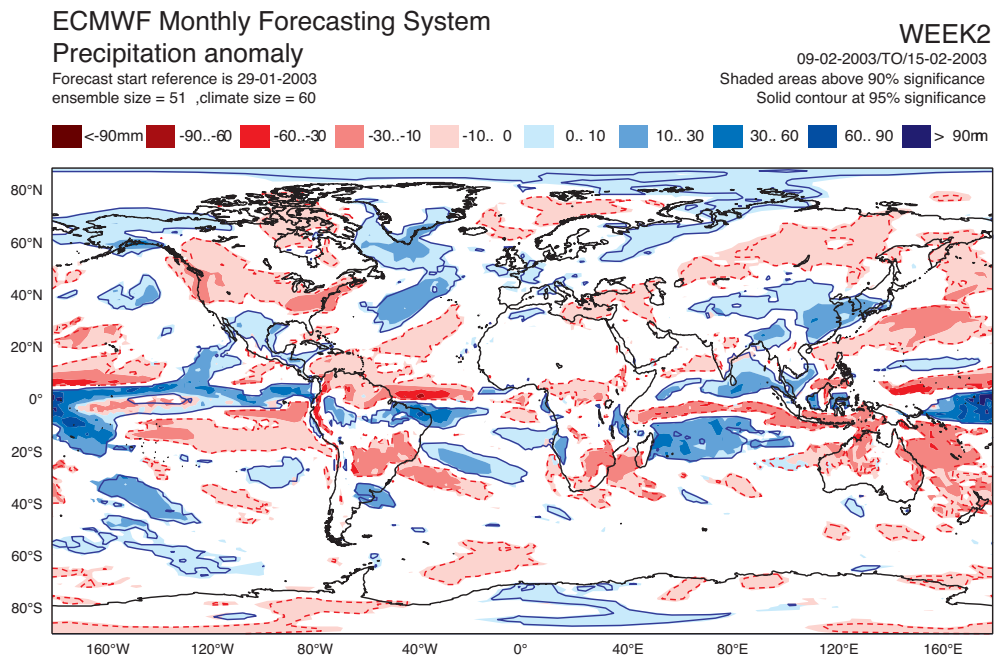


*Figure 23: Same as Figure 22, but for precipitation*

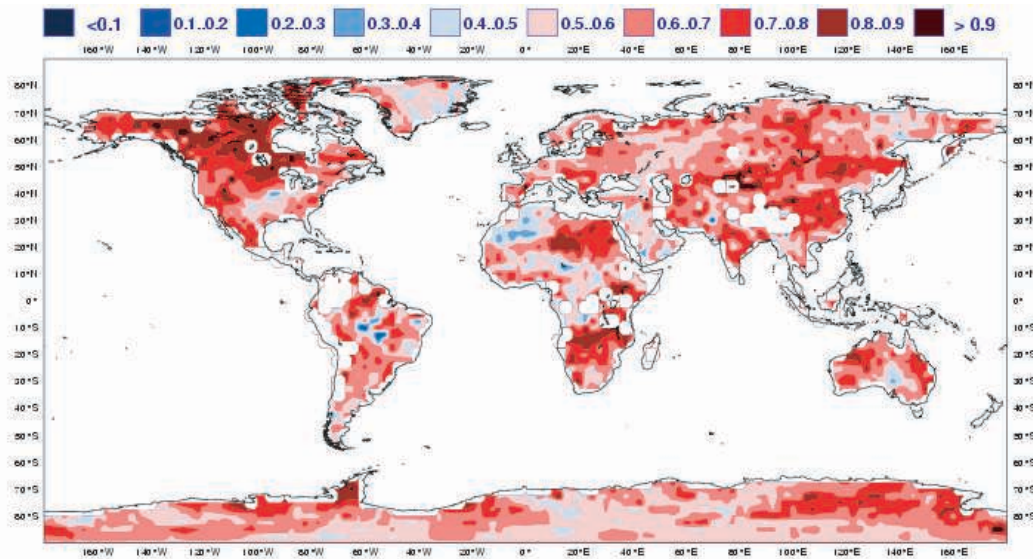**2m-temperature in upper tercile**
**Monthly Forecast Day 12-18**

*Figure 24:   Map of ROC areas of probability that the 2-metre temperature anomaly is in the upper tercile. The verification period is March 2002-May 2003. The red color-scale corresponds to ROC scores higher than 0.5 (better than climatology). The blue color-scale corresponds to ROC scores lower than 0.5 (worse than climatology).*

0K, 1K or 2K (Fig.28) suggests the same conclusion: the monthly forecasting system has some value for the period day 12-18, and it beats persistence of day 5-11. The results are also the same with other variables. Figure29 suggests that for precipitation, the ROC scores are higher for the monthly forecasting system than for persistence, although the ROC area (0.58) is much lower for precipitation than for 2-metre temperature. The monthly forecasts are also more reliable than persistence, with a positive Brier skill score, suggesting that the model performs better than climatology.

Finally, Figure30 suggests that the difference in scores between monthly forecast and persistence is even larger with mean sea-level pressure. The difference of ROC scores is larger than with 2-metre temperature, and the monthly forecasts of mean sea-level pressure display strong reliability, with, once again, a positive Brier skill score, whereas the persistence of the previous week shows a poor reliability, with a reliability diagram close to the horizontal line and a negative Brier skill score of -0.29. The difference between monthly forecast and persistence is significant for precipitation and mean sea-level pressure (Fig.31). Persistence outperforms the monthly forecasting system only on a few occasions. All these results suggest that overall, the monthly forecasting system outperforms persistence and climatology over land points in the Northern Extratropics, suggesting that the monthly forecasting system produces useful forecasts for days 12-18.

However, the performance of the monthly forecasting system varies from one region to another. Over North America or the Southern Extratropics, the model displayed some strong potential economic value over the 30 cases, and performed significantly better than persistence (Fig.32). Over the Tropics, the model displayed less skill, but still performed better than persistence. Over Europe, the potential economic value of the model was much less than over the other regions, and not significantly better than persistence, suggesting that Europe is a more difficult region at this time range than North America or the Southern Extratropics. Figures33 and 34 display respectively the difference of ROC and Brier skill scores between the monthly forecasting system

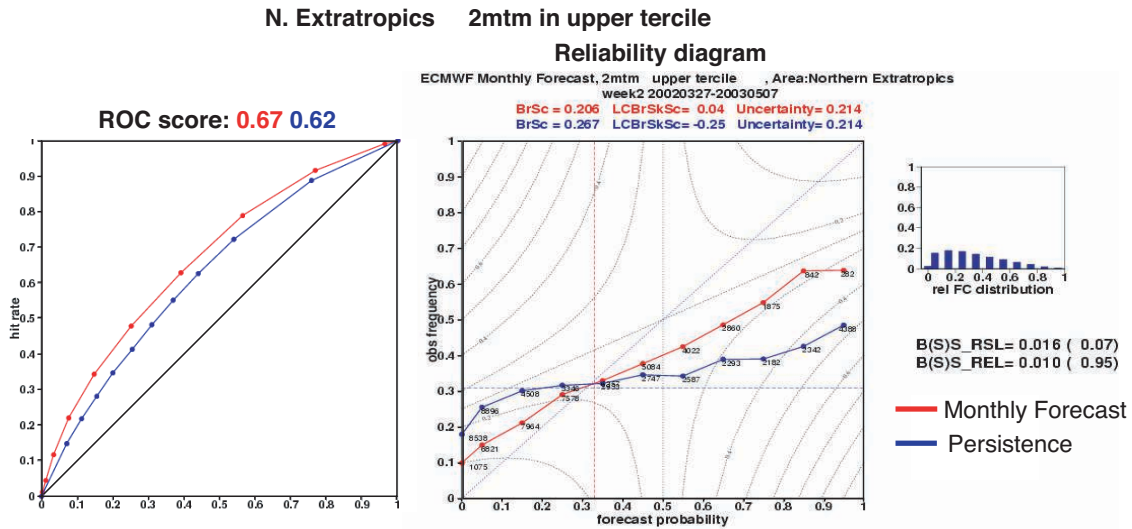**N. Extratropics   2mtm in upper tercile**

*Figure 25:  ROC (left panel) and reliability (right panel) diagrams of the probability that the 2-metre temperature is in the upper tercile. Only land points in the northern Extratropics have been considered. The red curves represent the diagrams obtained with the monthly forecasting system. The blue curves correspond to the diagrams obtained by persisting the anomalies from the previous week (days 5-11).*

and persistence.  Overall, the monthly forecasting system performs better than persistence.  However, over Scandinavia and the eastern edge of Europe, persisting the anomalies of the past week produced better forecasts than the monthly forecasting system during the period of verification.  This is not the case over the rest of Europe, however, where the monthly forecasting system performs better than persistence, particularly when using the Brier skill score (Fig.34).

### 4.3.1    Some examples

The fact that the monthly forecasting system performs better than persistence over the period days 12-18 suggests that the model has some skill in predicting changes of weather from one week to another, as the following examples over Europe will demonstrate. In the following examples, only deterministic forecasts will be shown, for comparison with observations. However, probabilistic forecasts are also available, like tercile or probability maps.

The first example (Fig.35) is the monthly forecast starting on 12 March 2003. In this case, the model predicted cold anomalies over most of Europe for days 5-11, except Scandinavia.  However, it predicted a different situation over western Europe for days 12-18, with warm anomalies over western Europe, which were not present in the initial conditions nor in the forecast of the previous week. This forecast verified well (Fig.35, bottom panels). The intensity of the anomalies is less in the forecast than in observations, mostly because the figure displays the ensemble mean, and as the spread of the ensemble is getting large, the mean of the ensemble rarely displays a large anomaly for days 12-18. The model displayed some skill in predicting the evolution fo the anomaly of geopotential height at 500 hPa from days 5-11 to days 12-18 over western Europe Fig.36).

The second example (Fig.37) is the monthly forecast starting two weeks after the previous case, on 26 March 2003. In this case, the model predicted cold anomalies over all Europe for days 12-18, which were not present in the forecast of days 5-11.  This forecast is in good agreement with analysis (bottom panel of Fig.37). The
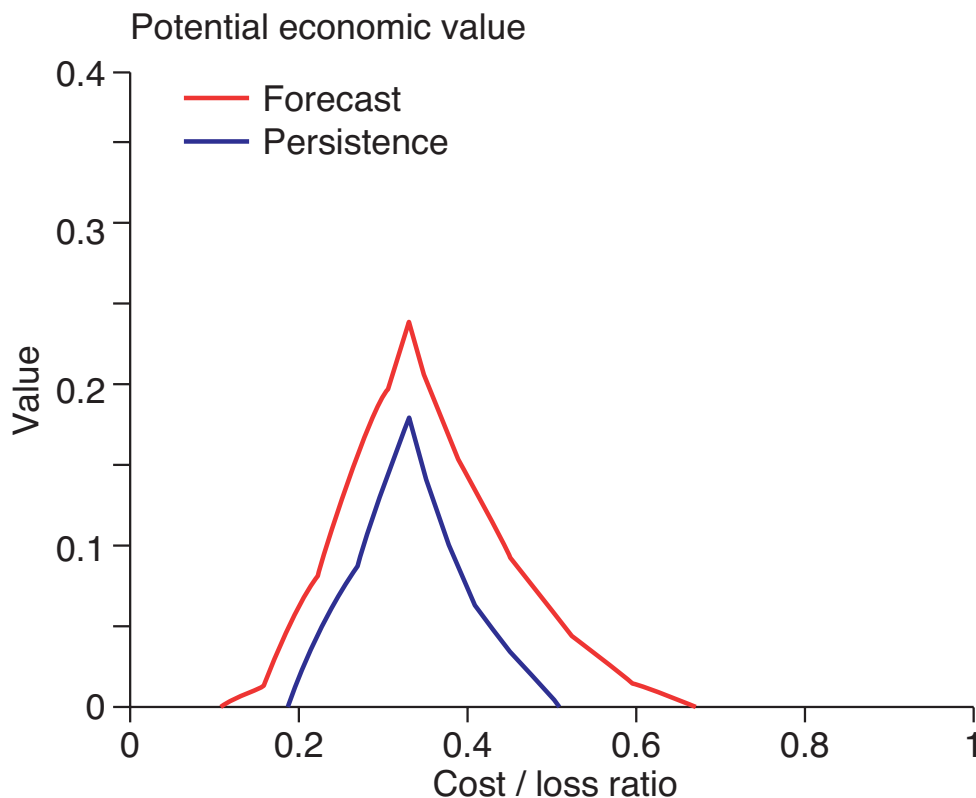
## Potential economic value



*Figure 26: Potential economic value of the probability that the 2-metre temperature is in the upper tercile. The red curve represents the monthly forecasting system. The blue curve corresponds to the potential economic value obtained by persisting the anomalies from the previous week (days 5-11).*
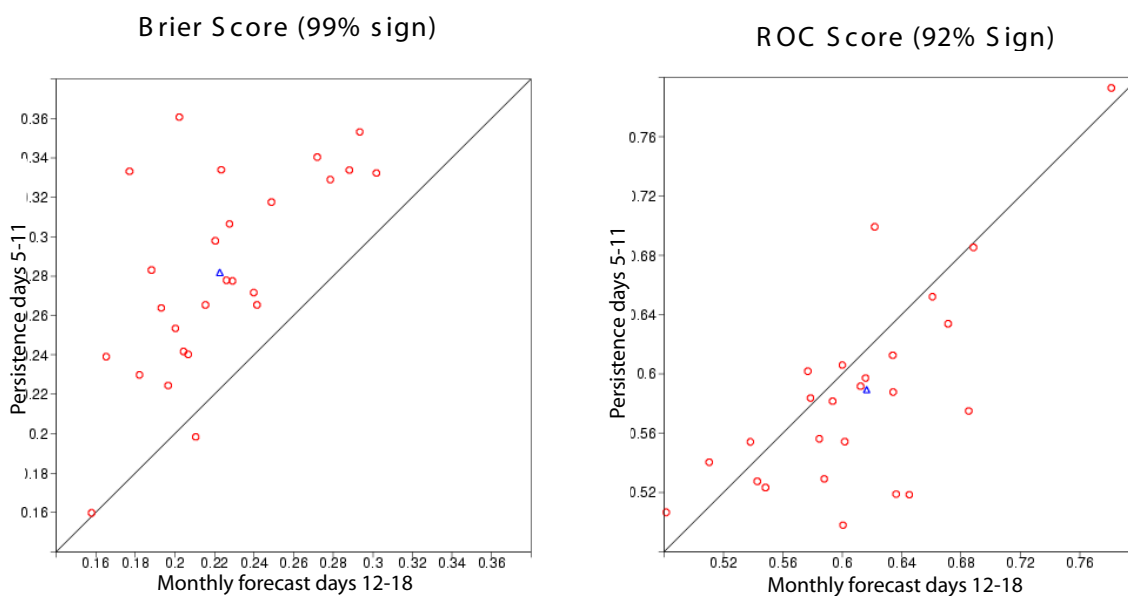


*Figure 27: Scatter-plot diagram of Brier scores (left panel) and ROC scores (right panel) of the probability that the 2-metre temperature is in the upper tercile. Each circle represents one case (of a total of 30).*

## Potential Economic Value 2m-temp N. Extra.



*Figure 28: Potential economic value of the probability that the 2-metre temperature anomaly is greater than 0K (blue curves), 1K (red curves) and 2K (green curves). The solid line represents the potential economic value obtained with the monthly forecasting system. The dotted line represents the potential economic value obtained with persistence.*



*Figure 29: Same as Figure 25, but for precipitation.*

**N. Extratropics    Mean sea level pressure in upper tercile**



*Figure 30:   Same as Figure 25, but for mean sea-level pressure.*



*Figure 31:  Scatter-plot diagram of Brier scores of the probability that the precipitation (left panel) or mean sea-level pressure (right panel) is in the upper tercile*
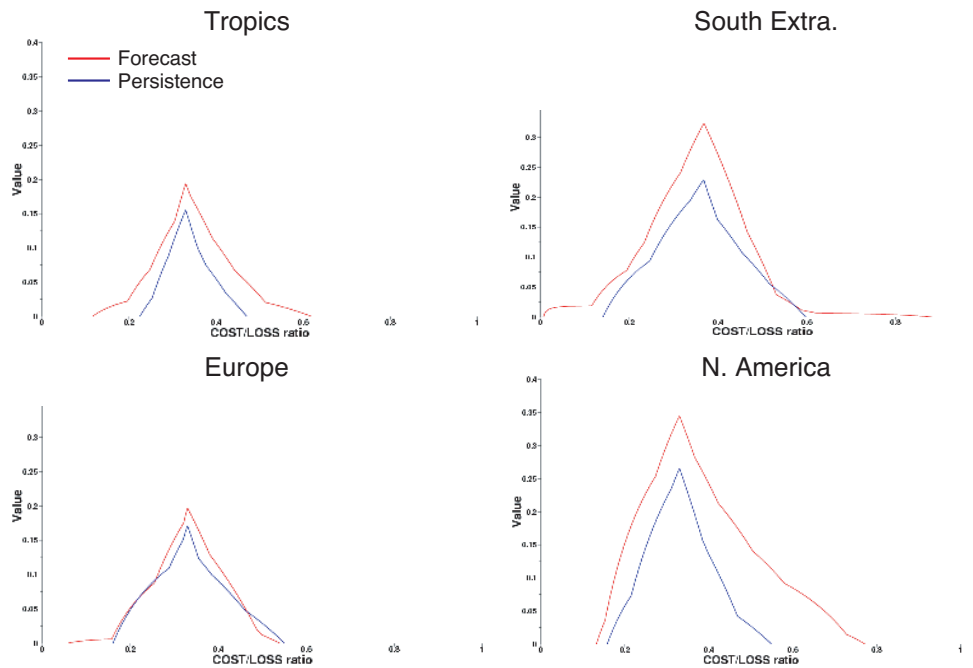
*Figure 32:   Potential economic values of the probability that 2-metre temperature is in the upper tercile. Red curves represent the potential economic value of the monthly forecasting system, whereas the blue curves represent the potential economic value of the persistence of the previous week. The scores have been computed over different regions: Tropics (top left panel), Southern Extratropics (top right panel), Europe (bottom left panel) and North America (bottom right panel).*

forecast of geopotential height at 500 hPa indicates that the model predicted a clear change of regime between days 5-11 and days 12-18 over Europe, with some good agreement with observations (Fig.38). In particular, the model correctly predicted a strong negative anomaly of geopotential height at 500 hPa over central Europe, which was responsible for the cold weather at that time. This example suggests that the model may have some skill in predicting changes of regime from one week to another. This aspect will be explored using blocking indices in Section 5.

The model forecasts are not always correct, as can be seen in Figure39. In this example, the forecast starting on 23 April 2003 for the period days 5-11 was in good agreement with observations: warm temperature anomalies over South-western Europe and cold anomalies over north-eastern Europe. However, the monthly forecasting system predicted that the cold anomaly would intensify, and the warm anomaly over western Europe would diminish. The opposite happened (bottom of Fig.39). The model persisted a negative anomaly of geopotential height at 500 hPa from days 5-11 to days 12-18 over Scandinavia (Fig.40), whereas the analysis shows that this anomaly disappeared during the second week.

## 4.4   Days 19-32

After about 20 days, the ensemble forecast is getting closer to the climatological distribution. Figure41 displays a typical forecast of 2-metre temperature for days 19-25. Blanked areas represent areas were the WMW-test does not detect any significant difference (within the 90% level of confidence) between the ensemble forecast and the climatological distributions. Over land, blanked areas largely dominate, suggesting that at this time range, the model does not display many strong signals. However, there are still some areas where the model
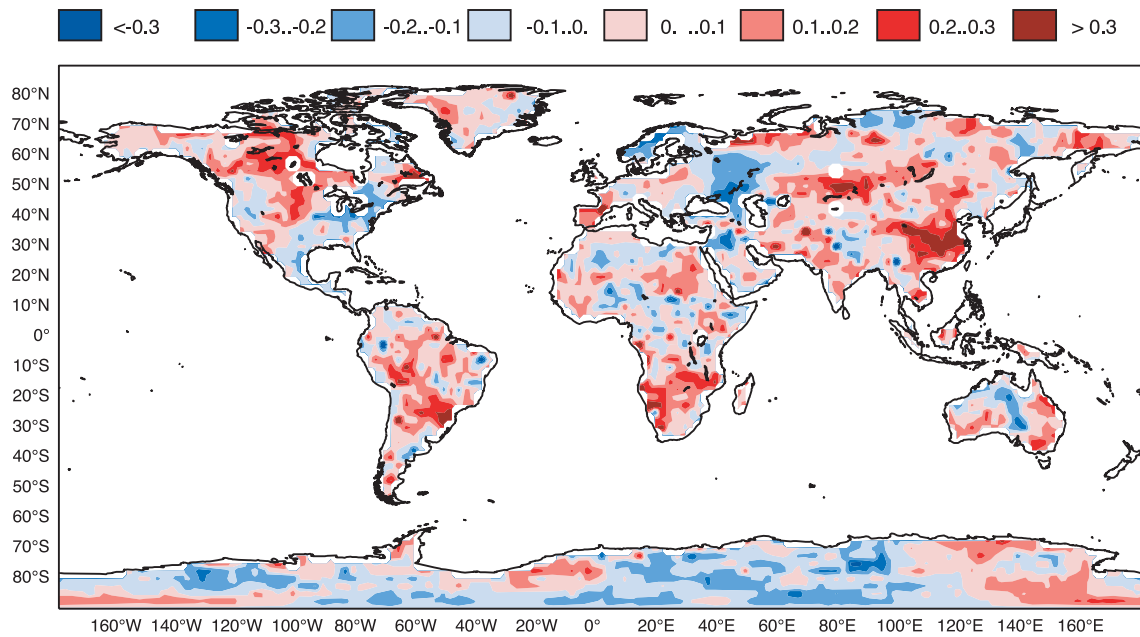
*Figure 33:   Difference between the ROC scores of the monthly forecasting system and the ROC scores obtained using persistence of the previous week. The red colors indicate positive values (the monthly forecasting system performs better than persistence), and the blue colors negative values (persistence performs better than the monthly forecasting system).*
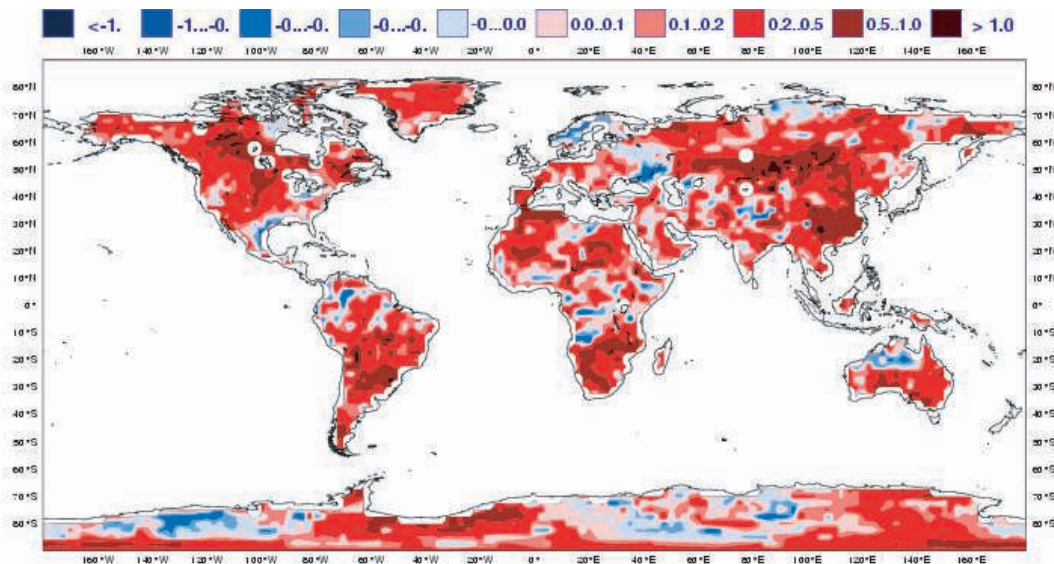


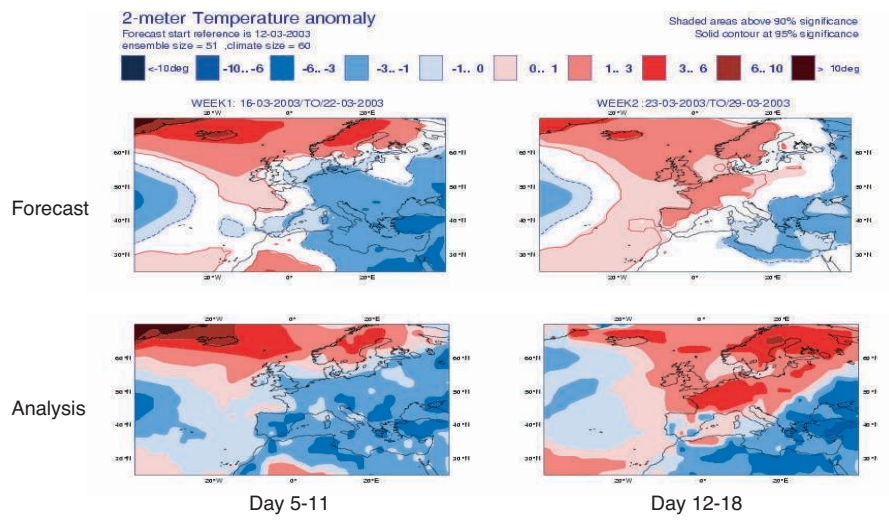*Figure 34:   Same as Figure 33, but for the Brier skill score.*

*Figure 35: Monthly forecast of 2-metre temperature anomalies (top panels) and verification from operational analysis and ERA40 (bottom panels). The starting date of the monthly forecast is the 12 March 2003. The bottom panel represents the forecast of days 5-11, and the right panels the forecasts of days 12-18.*
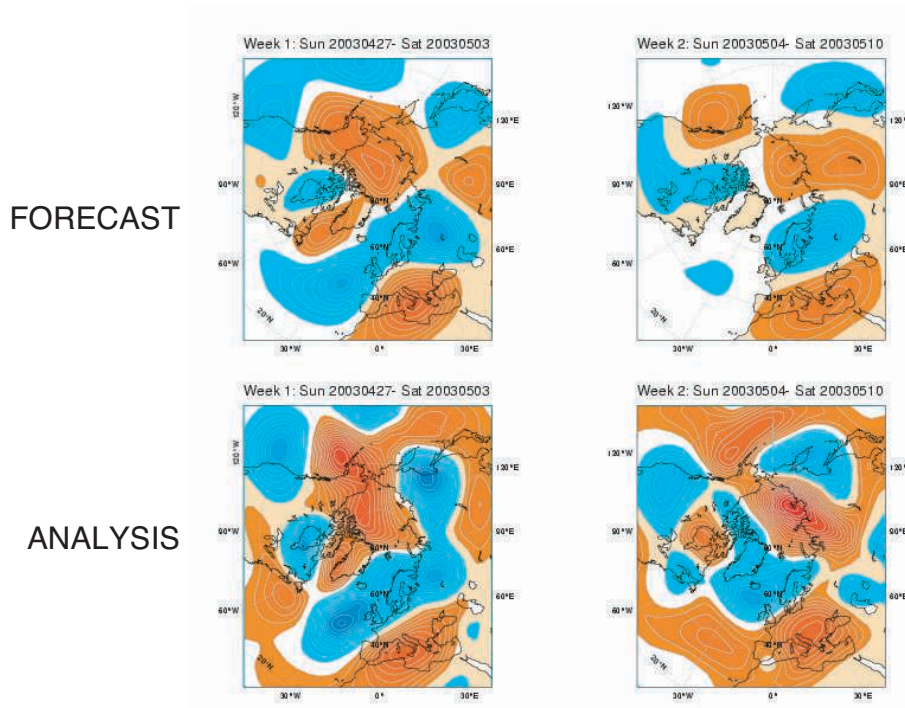


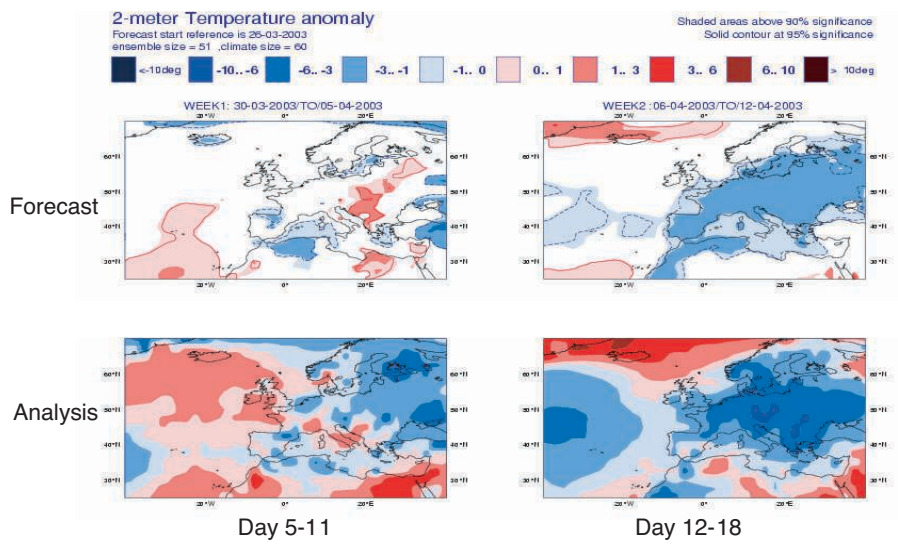*Figure 36: Same as Figure 35, but for geopotential height anomalies at 500 hPa.*

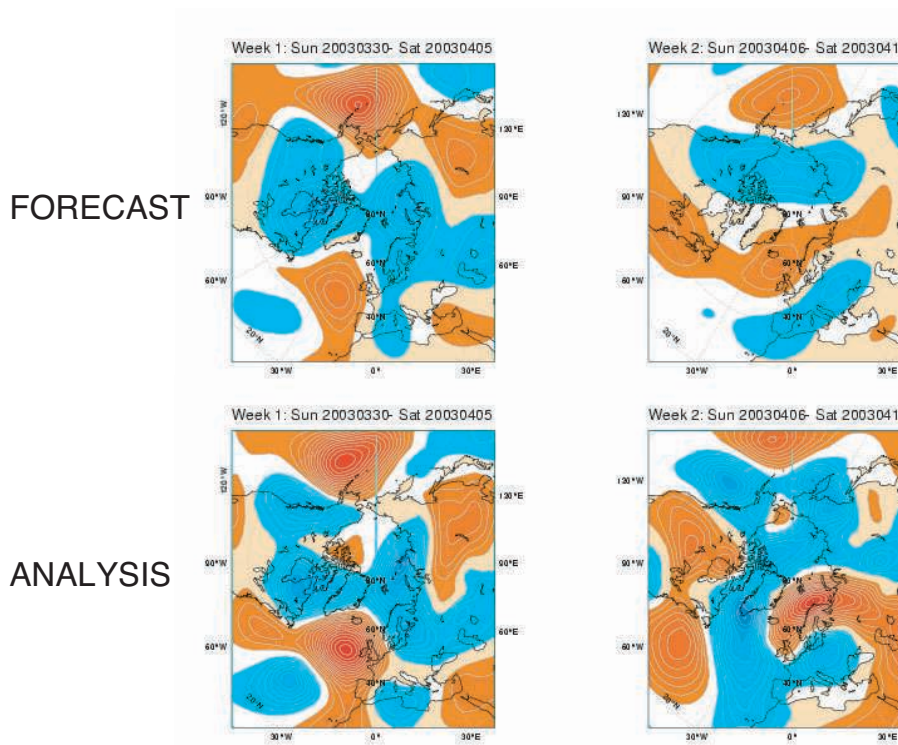*Figure 37: Same as Figure35, but for the forecast starting on 26 March 2003.*



*Figure 38: Same as Figure36, but for the forecast starting on 26 March 2003.*
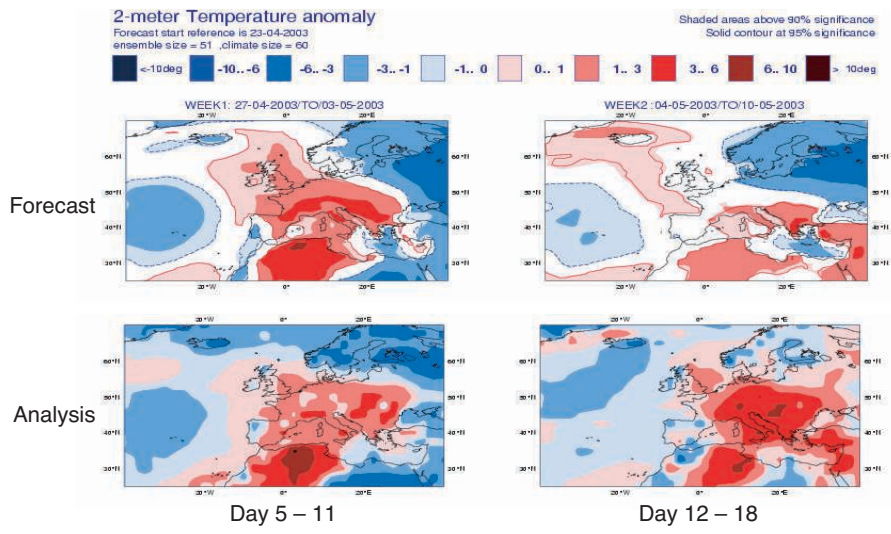
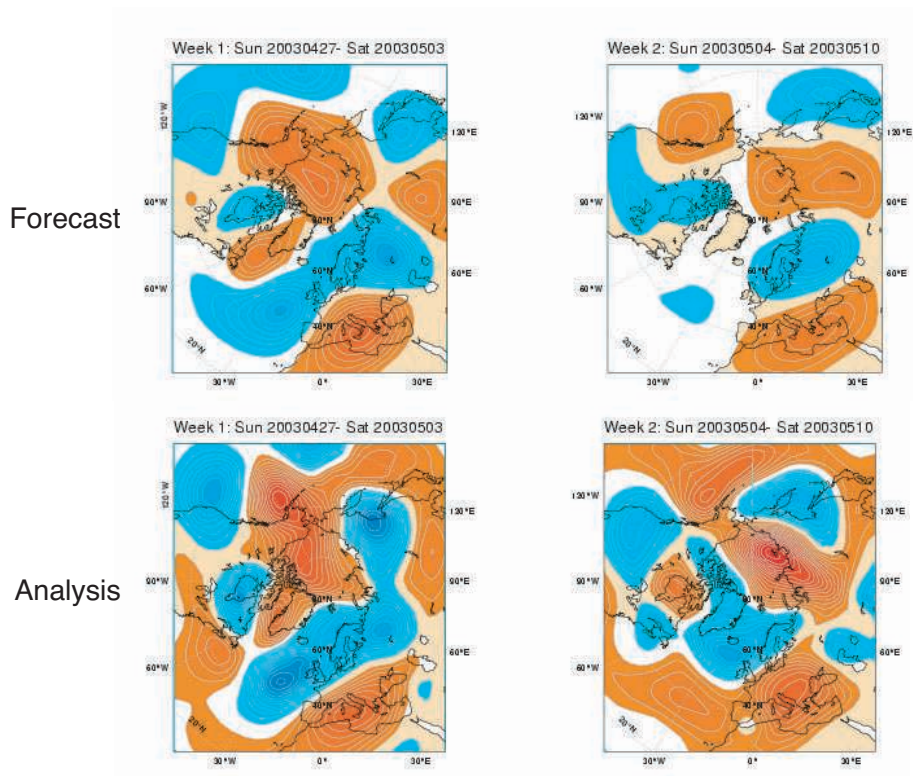*Figure 39: Same as Figure35, but for the forecast starting on 23 April 2003.*



*Figure 40: Same as Figure36, but for the forecast starting on 23 April 2003.*

forecast is significantly different from climatology, for example, over southern Europe in Figure41. Of course, the geographical location of those areas varies from one starting date to another.
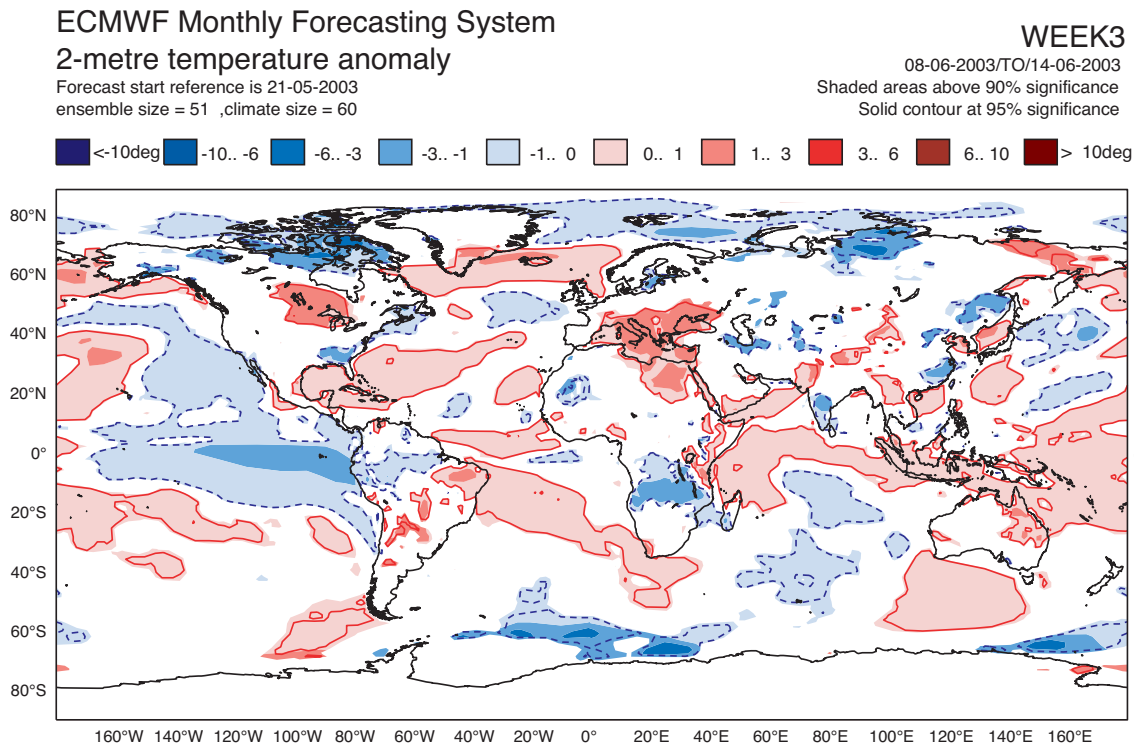


*Figure 41:   Monthly forecast of 2-metre temperature averaged from day 19 to day 25. The starting date is the 21 May 2003. Blanked areas represent areas where the ensemble distribution of the real time forecast is not significantly different from the model climatological distribution with a level of confidence higher than 90%.*

The different fields (2-metre temperature, precipitation, mean-sea-level pressure ..)  have been averaged over the two week period days 19-32 (the two last weeks of the monthly forecast). Brier skill scores, ROC areas, potential economic values have been computed over land grid points and over the 30 cases. The scores have also been compared to the scores obtained by persisting the anomalies of the 2 previous weeks of the forecast (days 5-18). This will be referred to as persistence. Results suggest that after about 20 days of forecast, the value of the forecast is very dependent on the threshold of the event (Figures 42 and 43). For small thresholds, such as the probability that the 2-metre temperature anomaly is larger than 0, the potential economic value is quite low, and the model does not perform significantly better than persistence (Fig.42). However, for higher thresholds, for example the probability that the 2-metre temperature anomaly is larger than 2K, the model displays some value for lower cost-loss ratios, which exceeds the value of persistence. This suggests that even at this time range, the monthly forecasting system could still be useful.

Except for low thresholds, the monthly forecasting system generally produces better forecasts than persistence for days 19-32. Figure44 displays a scatter-plot diagram of Brier skill score over the Northern Hemisphere. The events scored are the probabilities that the mean sea-level pressure (left panel) and precipitation (right panel) are in the upper tercile. In the vast majority of cases, the Brier skill score is larger with the monthly forecasting system than with persistence. The skill of the monthly forecasting system at this time range is, however, strongly dependent on the geographical location (Fig.45). Over Europe, the model displays a ROC score close to 0.5, suggesting that the model has little skill over this area. This is confirmed when computing the ROC score over the whole area (Fig.46). The reliability of 2-metre temperature anomaly larger than 2K is
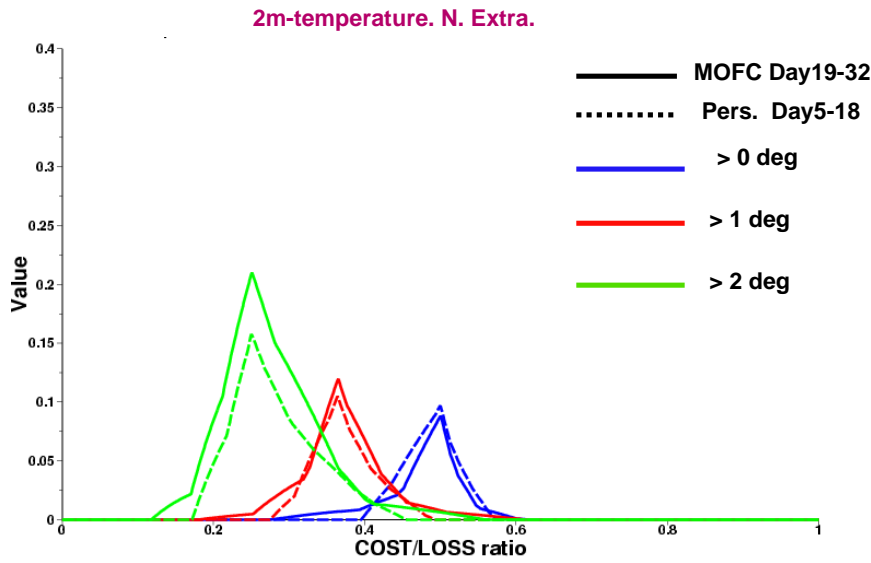
*Figure 42: Potential economic value of the probability that the 2-metre temperature anomaly averaged from day 19 to day 32 is greater than 0K (blue curves), 1K (red curves) and 2K (green curves). The solid line represents the potential economic value obtained with the monthly forecasting system. The dotted line represents the potential economic value obtained with persistence.*
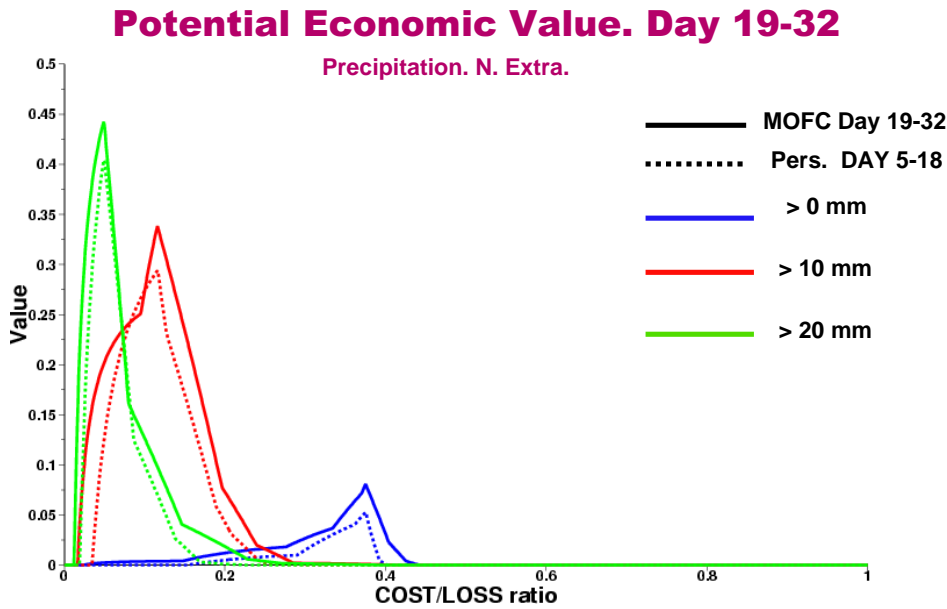


*Figure 43: Potential economic value of the probability that precipitation anomaly accumulated from day 19 to day 32 is greater than 0 (blue curves), 10mm (red curves) and 20mm (green curves). The solid line represents the potential economic value obtained with the monthly forecasting system. The dotted line represents the potential economic value obtained with persistence.*

also quite small over Europe (Fig.47, left panel). For precipitation (Fig.48, left panel), the model displays some reliability over Europe, but the Brier skill score, although positive, is very low. This suggests, that the monthly forecasting system does not display very useful skill over Europe after 20 days of forecasts, at least over the 30 cases that have been verified.

However, over other regions, like North America, the Tropics or the Southern Extratropics, the model displays ROC scores exceeding 0.6 on average (Fig.45). Over those regions, the model also displays some value at the 19-32 day time range, and also usually beats persistence, particularly over the Southern Extratropics (Fig.46). The 2-metre temperature forecasts over North America and the Southern extratropics are quite reliable, with positive Brier Skill scores (Fig.47). This is also true for precipitation (Fig.48)

### 4.4.1   An example: heat wave over France in August 2003

This section will show an example of monthly forecasting up to 30 days. The example chosen is the heat wave that occurred over Europe during the Summer 2003. This heat wave was particularly intense during the first 2 weeks of August, with catastrophic consequences. The present study focuses on the week from August 3 to August 9, where 2-metre temperature anomalies relative to the past 12-year climate were close to 10 degrees over some parts of France (Figure49, top left panel). The top right panel in Figure49 displays the monthly forecasts starting on 30 July 2003. It represents the ensemble mean of 2-metre temperature anomaly (relative to the model climatology) for the period days 5-11. At that time range, the model predicted a strong positive anomaly of 2-metre temperature, although the intensity of the ensemble mean is less than in the analysis (top left panel). Almost all the ensemble members predicted a significant heat wave, but only a few members of the ensemble predicted an intensity as strong as in the analysis.

The monthly forecasting system runs every two weeks, but for this test-case, a hindcast starting on 25 July 2003 has been realized to evaluate the weekly evolution of the forecast. Figure50 displays stamp maps of the 2-metre temperature forecasts averaged over the period days 12-18. A subjective test suggests that fifteen members of the ensemble predict a weekly 2-metre temperature anomaly larger than 3K over most of Europe ( members indicated with a blue line in Figure50). Four ensemble members display an anomaly with the same amplitude as in the analysis. In the 60 member-ensemble model climatology corresponding to this real-time forecast, only four members display a weekly 2-metre temperature anomaly larger than 3K, and none of them has an amplitude comparable to the analysis. This suggests that although most ensemble members do not predict a significant heat wave over Europe for the period days 12-18, the model predicted a probability for this extreme event higher than in the model climatology. The ensemble mean (Figure49, bottom left panel) displays a significant and positive anomaly in 2-metre temperature averaged over days 12-18 over most of Europe. However since the anomaly is averaged over all the ensemble members, its magnitude is less than in the analysis.

Stamp maps for the period days 19-25 of the monthly forecast starting on the 16 July 2003 (Figure51) indicate that 12 ensemble members predict a weekly 2-metre temperature anomaly larger than 3K, with one forecast displaying a positive 2-metre temperature anomaly as strong as in the analysis. In the model climatology corresponding to this forecast, only 4 ensemble-members display a weekly 2-metre temperature anomaly larger than 3K, and none of them has an amplitude comparable to the analysis. This suggest that 20 days in advance, the model predicted a higher probability than usual of a significant heat wave over western Europe. The ensemble mean (Figure49, bottom right panel) displays a strong ( although weaker than in the analysis) and significant positive anomaly over a large portion of Europe. Such a strong signal in the ensemble mean is quite unusual for this time range.

ms lp in upper tercile

MOFC/PERS: 21/5 (sign:99%)

Precip in upper tercile

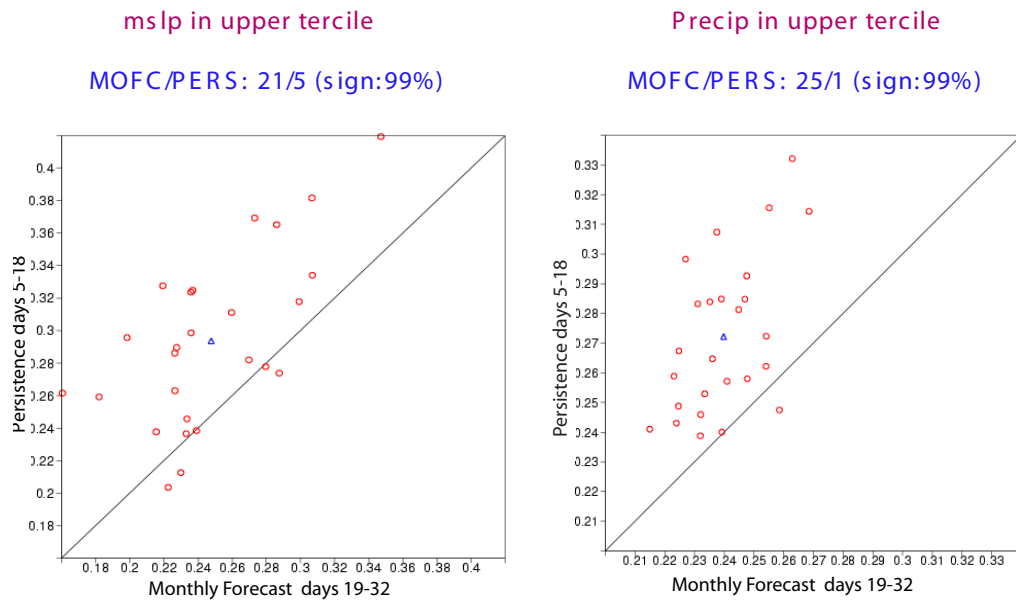MOFC/PERS: 25/1 (sign:99%)



*Figure 44:    Scatter-plot diagram of Brier scores of the probability that the mean sea-level pressure (left panel) and precipitation (right panel) are in the upper tercile. Each circle represents an individual case.*
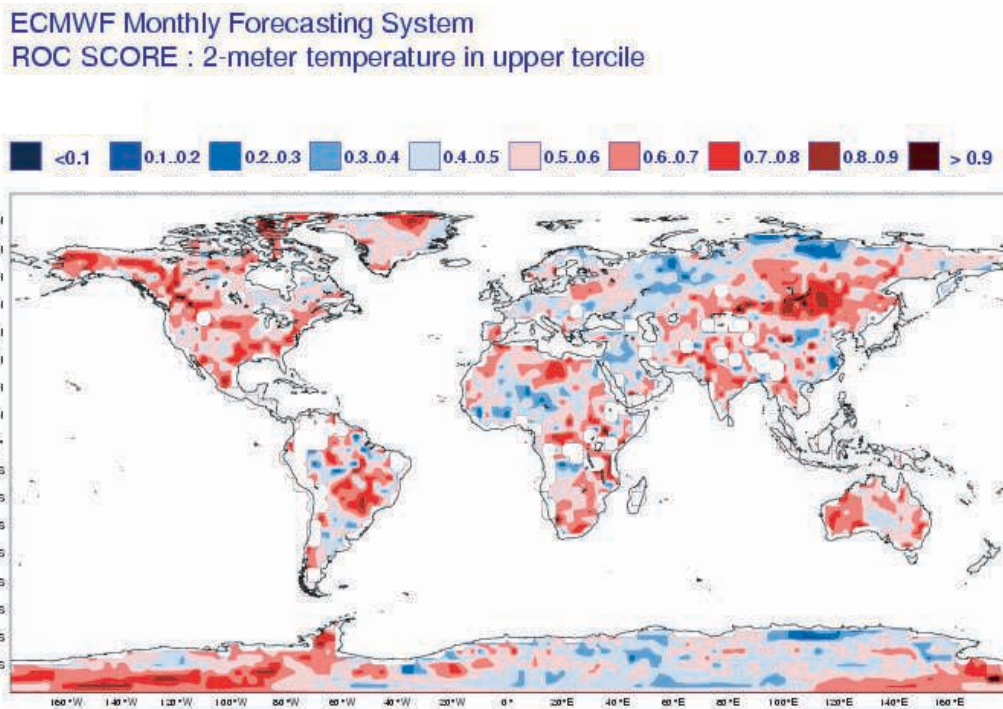


*Figure 45:   Map of ROC areas of probability that the 2-metre temperature anomaly averaged from day 19 to day 32 is in the upper tercile. The verification period is March 2002-May 2003. The red color-scale corresponds to ROC scores higher than 0.5 (better than climatology). The blue color-scale corresponds to ROC scores lower than 0.5 (worse than climatology).*
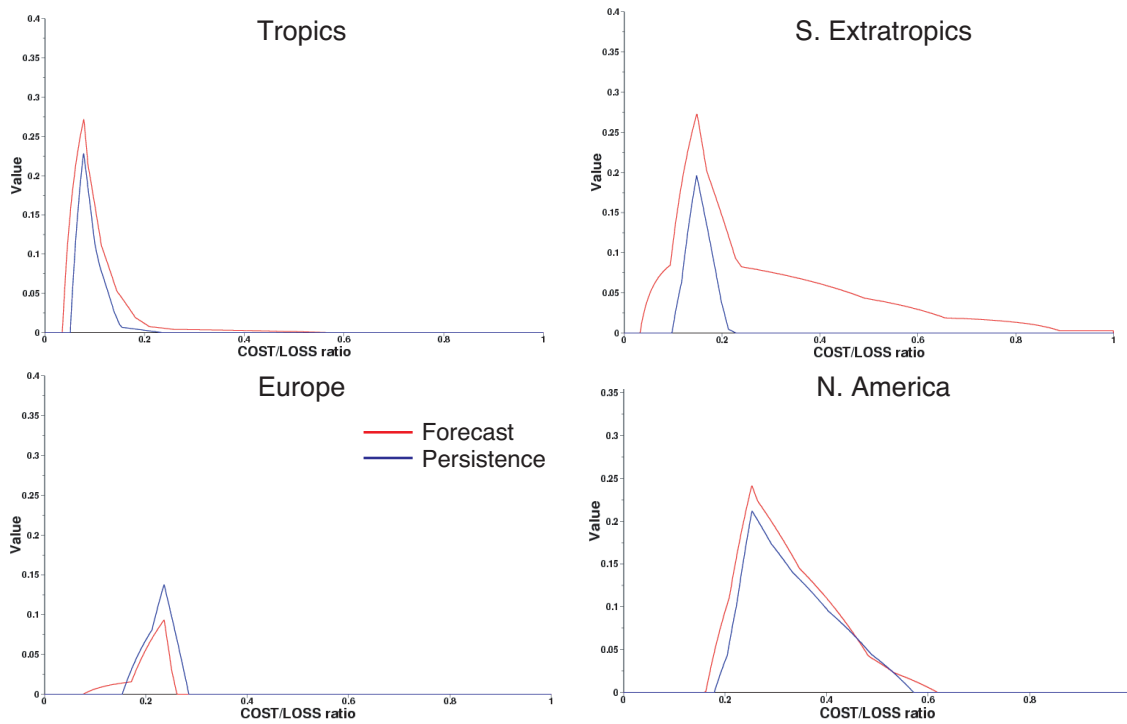
*Figure 46:   Potential economic values of the probability that 2-metre temperature averaged from day 19 to day 32 is in the upper tercile. Red curves represent the potential economic value of the monthly forecasting system, whereas the blue curves represent the potential economic value of the persistence of the previous week. The scores have been computed over different regions: Tropics (top left panel), Southern Extratropics (top right panel), Europe (bottom left panel) and North America (bottom right panel).*
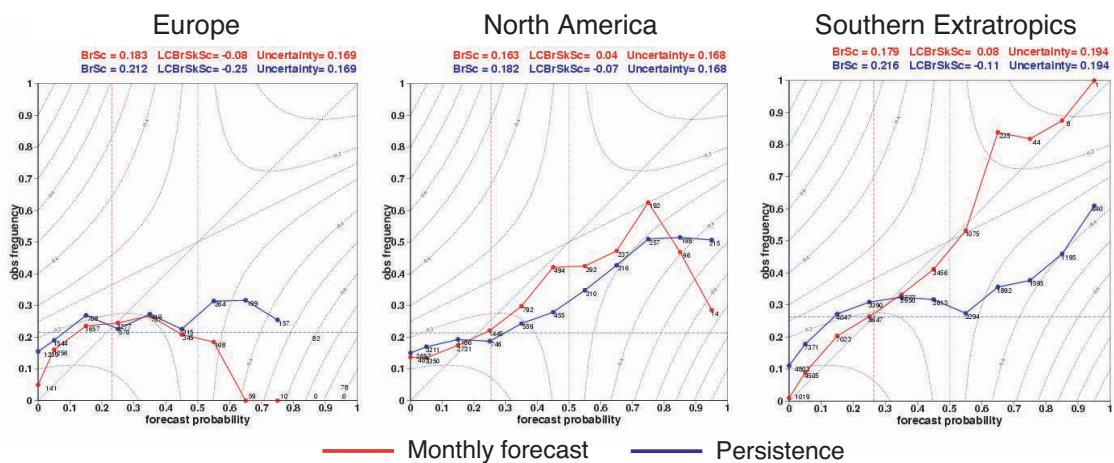


*Figure 47:   Reliability diagram of the probability that the 2-metre temperature anomaly averaged from day 19 to day 32 is larger than 2K over a) Europe, b) North America and c) Southern Extratropics.*
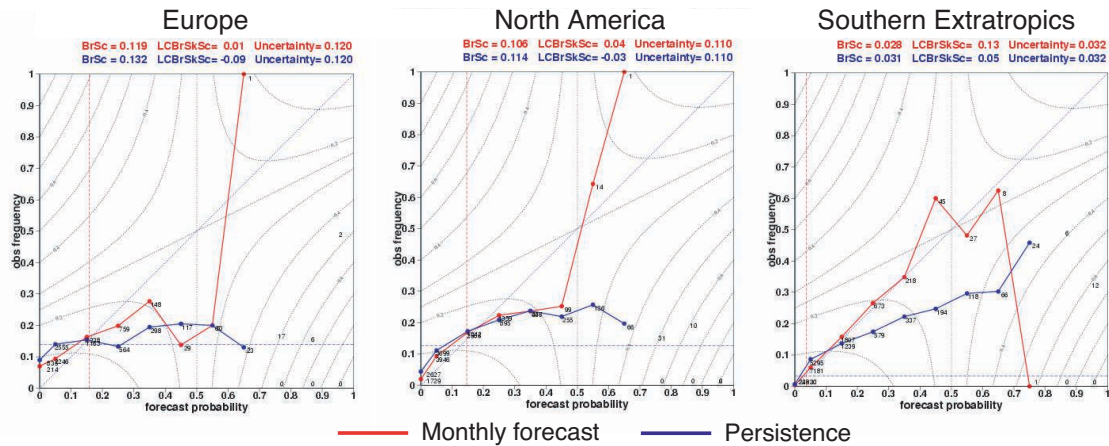
*Figure 48: Reliability diagram of the probability that the accumulated precipitation anomaly averaged from day 19 to day 32 is larger than 10 mm over a) Europe, b) North America and c) Southern Extratropics.*

# 5 Weather regimes

The previous section has explored the skill of the monthly forecasting system in predicting weekly-mean anomalies for each grid point. However, most of the skill of the monthly forecasting system is likely to originate from its skill in predicting changes in the large-scale circulation, and more specifically changes in weather regimes. The goal of this section is to evaluate if the monthly forecasting system is able to predict changes in weather regimes.

## 5.1 Regime clusters

Six weather regimes based on 500 hPa geopotential have been selected using ERA15 data (Chessa, 2000). In this approach, the 500 hPa geopotential for each member of the ensemble is compared to each of the six predefined weather regimes, and is associated to the closest. The number of ensemble members corresponding to each cluster is then calculated (see Fig.52 for example), and this number is translated into a probability for each cluster. This diagnostic is used in EPS with instantaneous values of 500 hPa geopotential. For the monthly forecasting system, it has also been applied to the 500 hPa geopotential averaged over weekly periods.

In order to evaluate the skill of the model to predict the probability of the 6 different clusters, the Brier score has been computed (Fig.53). According to Figure53, the model performs better than climatology for the first 3 periods, but not for days 26-32. The model also performs better than persistence during the first 3 periods, suggesting that the monthly forecasting system can be useful in predicting the population of the different clusters up to day 25.

## 5.2 Blocking index

A different approach from regime clustering consists in defining blocking indices. In this context, the weather pattern over a given region, like Atlantic-Europe, is classified as either blocked or zonal. In the present paper, we consider one of these indices, which will be referred to as the Pelly-Hoskins index (Pelly and Hoskins, 2003). This index diagnoses blocking events from the meridional temperature gradient of potential temperature
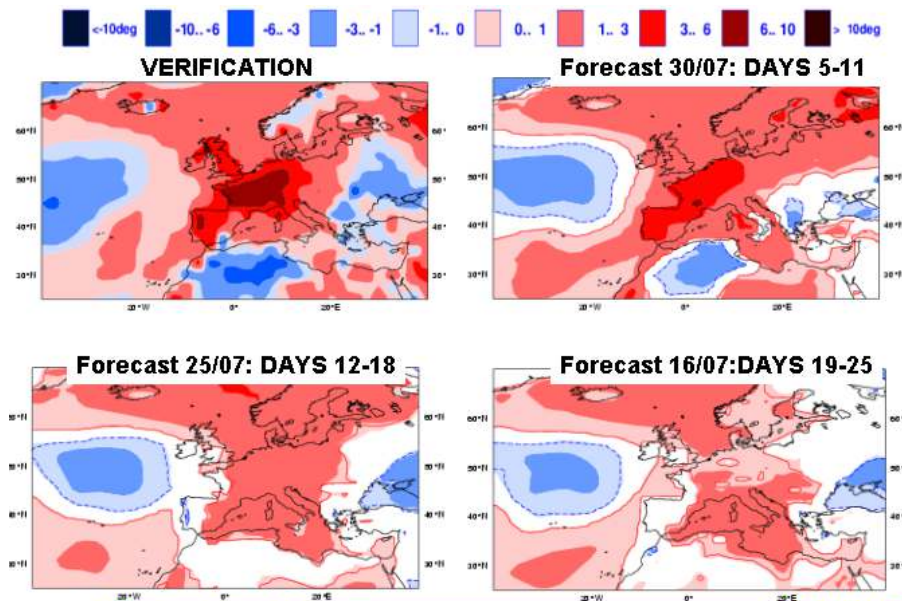
*Figure 49: 2-metre temperature anomalies averaged over the period 3-9 August. The top left panel represents the 2-metre temperature anomaly from the ECMWF operational analysis. The anomalies are relative to the past 12-year climate computed using ERA40 and operational analysis. The top right panel represents the ensemble mean of 2-metre temperature anomaly from the monthly forecast starting on 30 July 2003. The 2-metre temperature anomaly has been averaged over the period days 5-11 and the anomaly is relative to the model climatology. Areas where there is no significant difference between the ensemble distribution of the the real-time forecast and the model climatology according to the WMW-test have been blanked. The bottom left panel represents the 2-metre temperature anomaly from the hindcast starting on 25 July 2003, and averaged over all the ensemble members and the period days 12-18. The bottom right panel represents the 2-metre temperature anomaly from the monthly forecast starting on 16 July 2003, and averaged over all the ensemble members and the period days 19-25.*

on a potential vorticity surface.

### 5.2.1   Real-time forecast

The index has been computed for each member and for each realization of the real time forecast ensemble. For each day of the forecast, the predicted probability of blocking has been computed and compared with analysis. A daily Brier score has then been calculated and averaged over the 30 cases. It has then been normalized by the Brier score obtained from climatology. A value of the score lower than 1 indicates skill in the monthly forecasting system. A score close to 1, on the other hand, indicates that the model does not perform better than climatology. According to Figure54, the monthly forecasting system has some skill in predicting the occurrence of blocking on a daily basis during the first 6 days of the forecast. However, after about 10 days, the model does not display significant skill in predicting if one specific day will be blocked or not. In the rest of this subsection we will investigate the skill of the model to predict blocking episodes over a weekly averaged period, instead of over one single day.

The number of blocking days has been calculated over each weekly period (days 5-11, days 12-18, days 19-25, days 26-32), for each member of the real-time forecast ensemble and for all the 30 cases. It has also been calculated for the ensemble climatology. Figure55 displays the distribution (median, 25 %, 75 %, maximum and minimum) of blocking days anomalies for each case (x-axis) over the Euro-Atlantic sector. The same calcula-
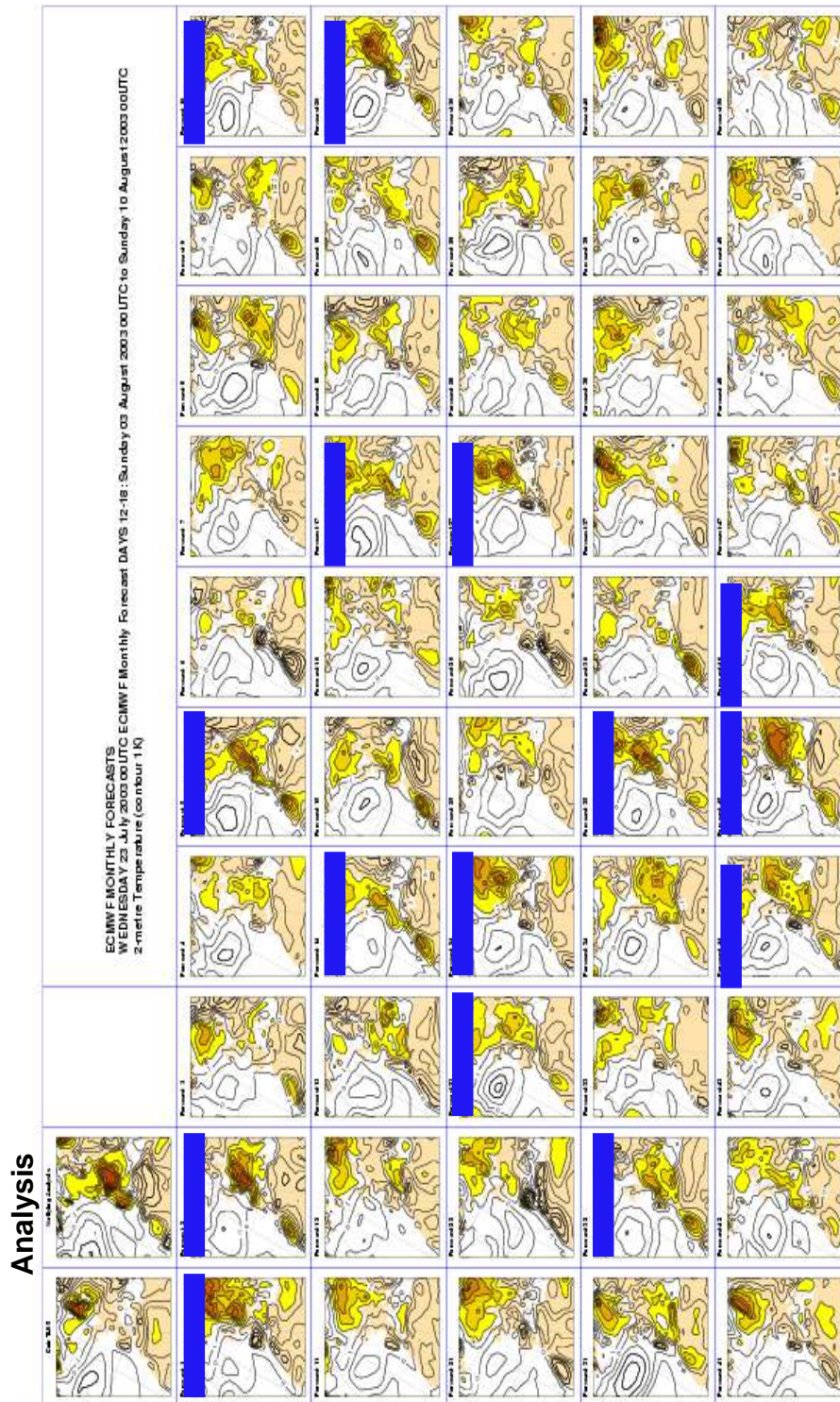
*Figure 50:   Stamp maps of 2-metre temperature anomalies relative to the past 12-year climate for the monthly forecast starting on 25 July 2003 and averaged over the period days 12-18. The contour interval is 1K. Areas where the anomaly exceeds 2K are shaded. The analysis is the second map in the top of the panel. The control of the monthly forecast is the map at the top left of the panel. All the other maps represent an individual member of the ensemble. The blue bar indicates ensemble members which display a 2-metre temperature exceeding 3K.*
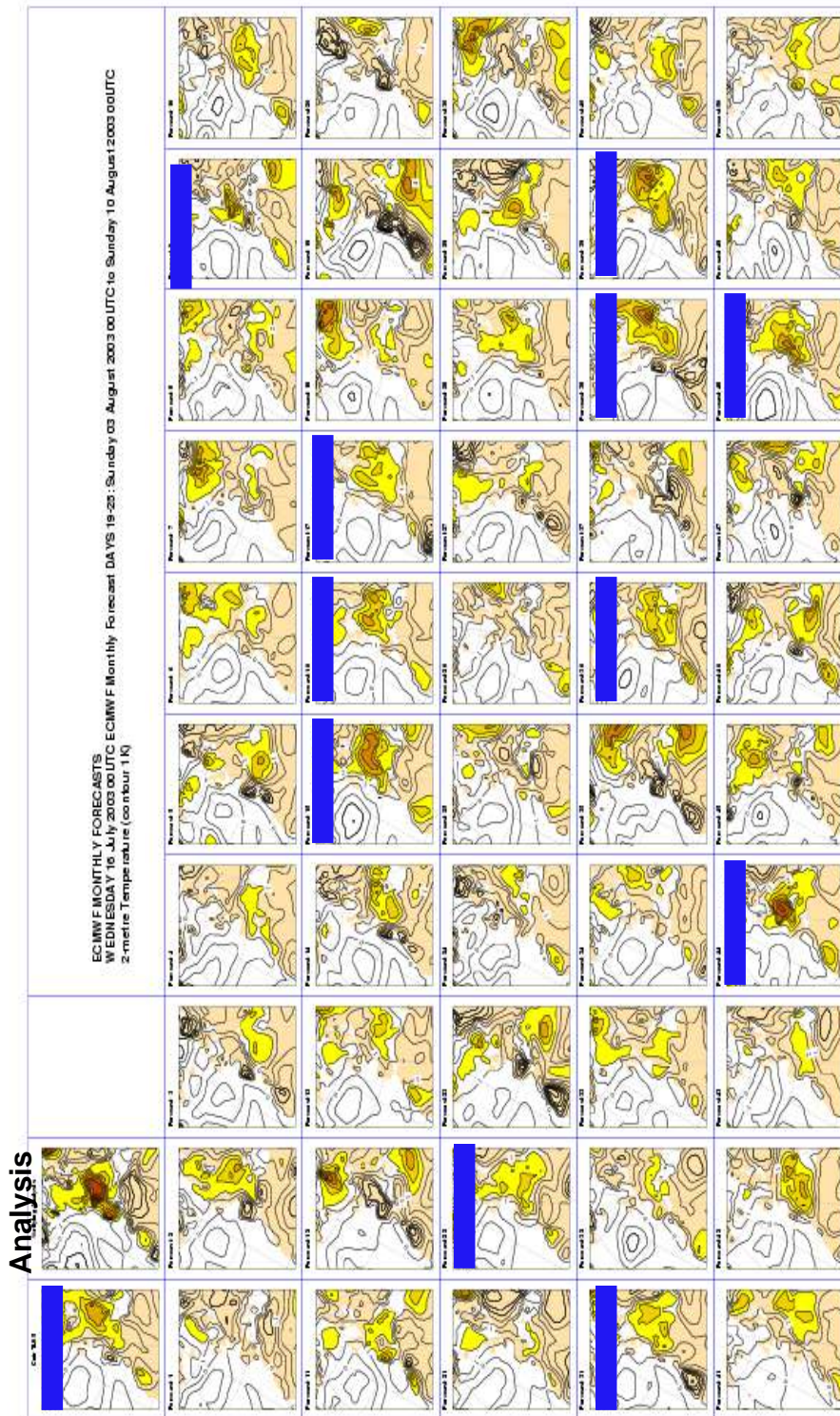
*Figure 51:   Same as Figure50 but for the monthly forecast starting on 16 July 2003 and the 2-metre temperature anomalies have been averaged over the period days 19-25.*
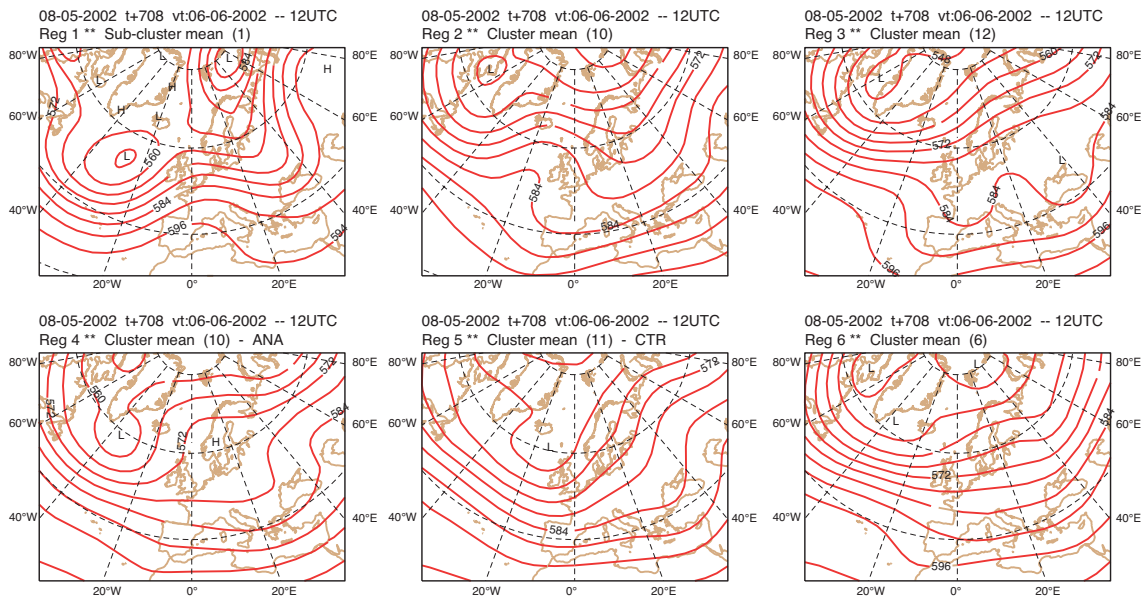
*Figure 52: The six regimes used for the classification. The number in parenthesis represents the number of ensemble members associated to each of the six predefined clusters on 6 June 2002 from the monthly forecast starting on 8 May 2002. ANA indicates the cluster associated to the analysis and CTR the cluster associated to the control of the monthly forecasting system.*



*Figure 53: Evolution of the Brier score associated to the probability for each cluster. The green line represents the score of the monthly forecasting system, calculated over the 30 real-time cases, and for each of the 4 weekly periods (days 5-11, days 12-18, days 29-25 and days 26-32). A score of 0 corresponds to a perfect deterministic forecast. The orange line represents the scores obtained by persisting the probability of the previous week. The red line represents the score obtained by persisting the first week of the forecast, and the blue line represents the score of climatology.*

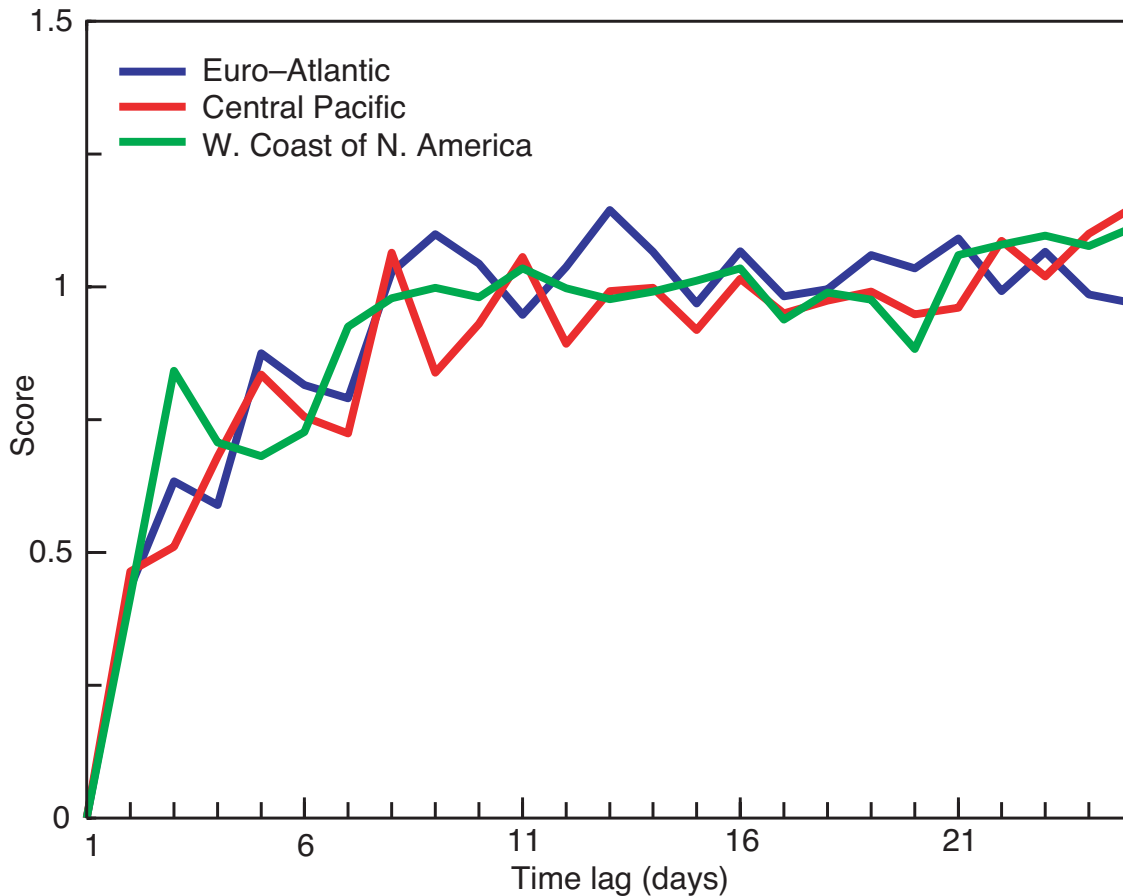*Figure 54: Time evolution of the Brier score obtained with the Pelly-Hoskins index applied to the 51-member real-time monthly forecast divided by the score obtained with climatology. The verification period is March 2002-May 2003. The blue line represents the score over the Euro-Atlantic sector, the red line represents the score over the Central Pacific and the green line represents the score obtained over the western coast of North America.*

tion of blocking events has been applied to the operational analysis and ERA40 (red lines in Fig.55). According to Figure55, the model displayed some skill in predicting the evolution of the blocking index averaged over the period days 5-11 from March 2002 to April 2003. The linear correlation between the time evolution of the median of the ensemble and the analysis is 0.44 over the full verification period. Figure55 highlights some 'bad spells' when the model did not perform well at this time range, like in July-August and also in January. From January to April, the Euro-Atlantic sector was generally blocked, and the model at this time range successfully predicted more blocking days than climatology in February and March.

For the period days 12-18 (Fig.56), the linear correlation between the time evolution of the mean of the ensemble and the analysis is quite low (0.28), although it is higher than the linear correlation obtained if persisting the blocking index of the previous week (-0.10). The model successfully predicted more blocking than climatology for the period February to April. During these three months, the model displayed strong skill in predicting the time evolution of the blocking index anomaly. However, as for the period days 5-11, the model was less successful over the first part of the blocking period, in December and January, maybe because the first part of this long blocking episode was less predictable.

Over the two last weeks of the monthly forecasts (Fig.57), the model displays very low skill in predicting

the variability of the blocking index anomaly with a linear correlation of -0.10. The model predicted less blocking activity than climatology over the period January to April 2003, in contradiction with observations. These results suggest, that over the Euro-Atlantic sector, the monthly forecasting system displays some skill in predicting the anomaly of blocking index averaged over one week up to about 20 days.

The Pelly-Hoskins index has also been calculated over the Central North Pacific sector. Over that region, the monthly forecasting system displays much stronger skill than over the Euro-Pacific sector. For the period days 5-11, the linear correlation between the median of the ensemble and analysis is 0.6. For the period days 12-18, the linear correlation is 0.44, much higher than when persisting the anomaly of the previous week (0.19). At this time range, the model has displayed particularly strong skill since November 2002 (Fig.58), but was less successful during summer 2002. For the period days 19-32, the linear correlation is 0.2, which is low, although higher than over the Euro Atlantic sector and higher than persistence. The above results suggest that overall, the monthly forecasting system has much stronger skill in predicting the blocking index over the North Pacific region than over Europe. This is probably due to the fact that over the North Pacific region, the blockings are more dependent on tropical Pacific variability than over the Euro-Atlantic sector, and the monthly forecasting system has some skill in predicting such variability, as will be discussed in the Section 7.

### 5.2.2    Some examples

Figures59, 60, 61 and 62 display a few examples of forecasts of 500 hPa geopotential taken from the period January to April 2003. As mentioned above, the model failed to predict the blocking anomalies in December and January. The first example (Fig.59) is the forecast of 500 hPa geopotential anomaly averaged over the 2-week period from 5 January 2003 to 18 January 2003. The left panel represents the anomaly of 500 hPa geopotential (relative to the 12-year climatology) computed from the operational analysis and ERA40 for the climatology. There was a strong blocking over Europe, with a negative 500 hPa geopotential anomaly over Europe and a positive anomaly of geopotential over the Atlantic. The forecast starting on 1st January 2003 (verification time is day 5 to day 18) failed to predict this pattern (top right panel of Fig.59). Instead, the model predicted a much more zonal flow. The forecast starting two weeks earlier (verification period is day 19 to day 32) also failed to predict a blocking over Europe.

Figure60 displays the analysis and forecasts for the period from 16 February 2003 to 1st March 2003. During that period, there was a strong blocking over Europe, which was well predicted starting on 12 February 2003. The forecast starting 2 weeks earlier, on 28 January 2003, also quite successfully predicted the main patterns of the 500 hPa geopotential anomalies (Fig.60, bottom right).

Finally, Figure61 displays the analysis and forecasts for the period from 27 April to 10 May 2003. During that period the 500 hPa geopotential anomalies suggest a zonal flow, which followed the long period of blocking. This is an interesting test case, since in the initial condition of the monthly forecasts starting on 9 and 23 April 2003, the weather regime over the Euro-Atlantic zone was blocked. The monthly forecast starting on 23 April 2003 successfully predicted this transition from a blocked to a zonal regime for days 5 to 18 (Fig.61, top right). The monthly forecast starting 2 weeks earlier, on 9 April 2003, displayed just a weak signal (Fig.61, bottom right).

### 5.2.3    Interannual variability

The period January to April 2003 was characterized by increased blocking over the Euro-Atlantic sector relative to climatology. The interannual variability of the Pelly-Hoskins index from 1991 to 2003 and averaged over the period January to April confirms than in 2003 the synoptic situation of the Euro-Atlantic sector displayed more
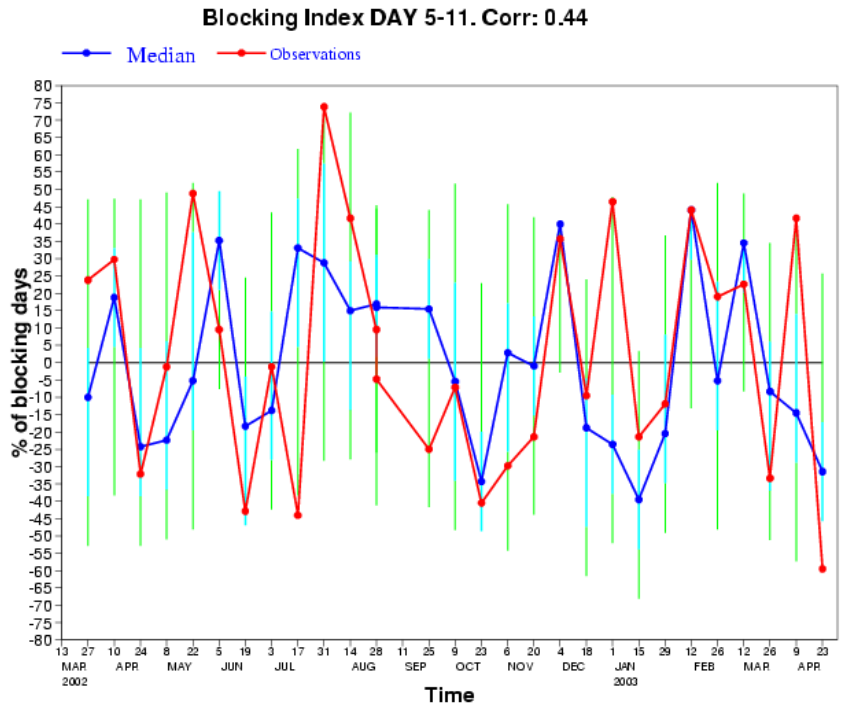
*Figure 55: Time evolution of the Pelly-Hoskins index anomaly (relative to the past 12 years climatology) averaged over the period days 5-11 from 27 March 2002 to 23 April 2003. The red line represents the index obtained with analysis. The blue line represents the time evolution of the median of the 51-member monthly forecast ensemble. The cyan line represents the distance between the 25% and 75% level of the ensemble distribution. The green line represents the distance between the maximum and minimum of the monthly forecast ensemble distribution.*
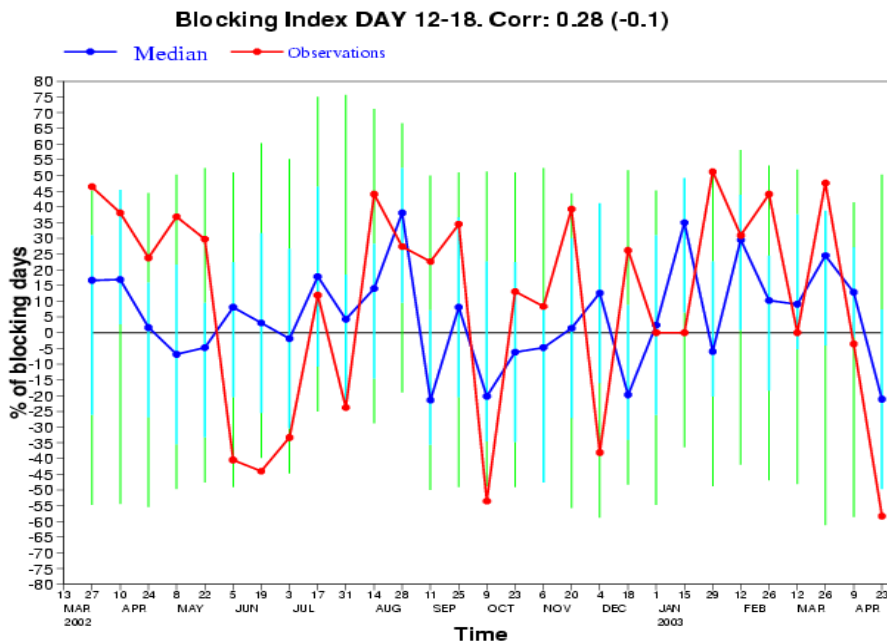


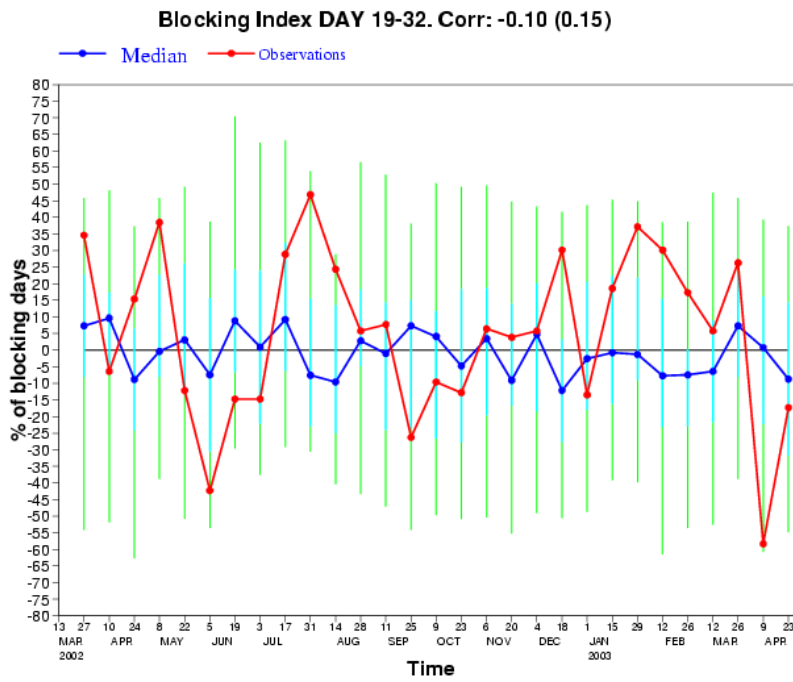*Figure 56: Same as Figure 55 but for the period days 12-18*
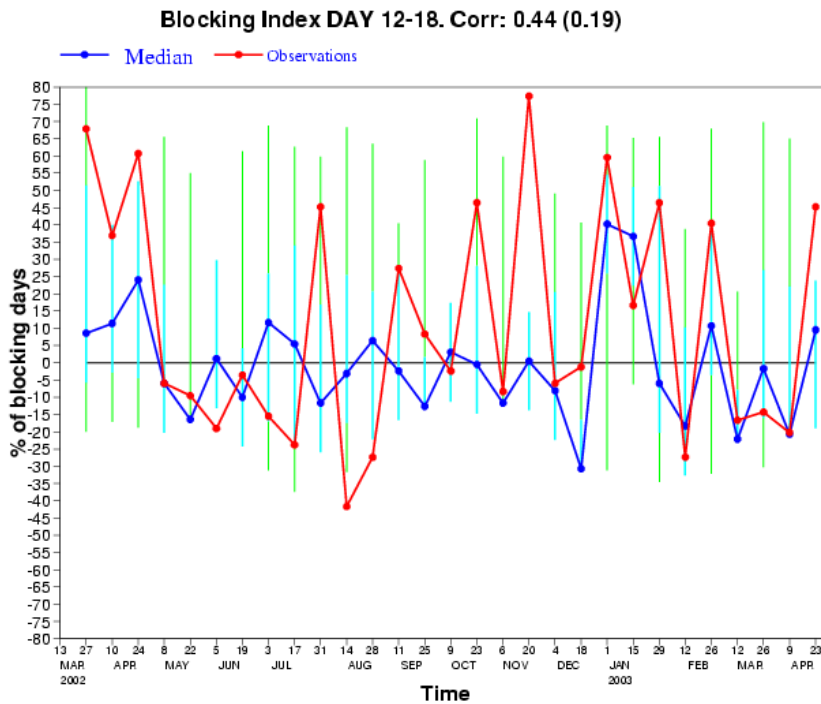
*Figure 57:   Same as Figure55 but for the period days 19-32*



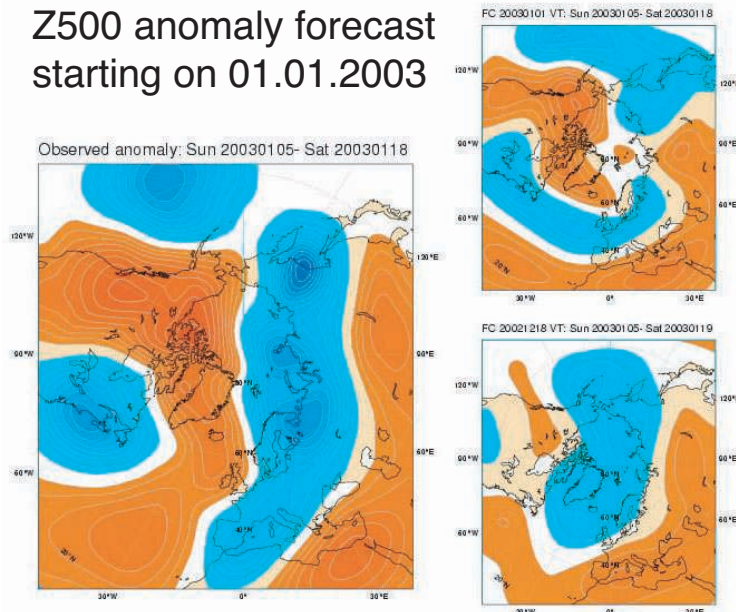*Figure 58:   Same as Figure56 but for the Central Pacific.*

*Figure 59: Anomaly of 500 hPa geopotential height averaged from 5 to 18 January 2003 and relative to the past 12 years climatology. Blue colors are for negative anomalies and red colors are for positive anomalies. The contour interval is 2 dam. The left panel represents the anomaly obtained with analysis and ERA40, and the right panels represent monthly forecasts. The top right panel corresponds to the monthly forecast starting on 1st January 2003 (time range is days 5-18), and the bottom right panel corresponds to the monthly forecast starting on 18 December 2002 (time range is days 19-32).*

blocking than usual. However, it was not exceptional. In 1996, the positive anomaly was much stronger than in 2003 (Fig.62). Overall, the blocking index averaged over the period January to April has a strong interannual variability. In 1994, 1994, 2001 and 2002, the flow was significantly more zonal than usual, whereas it was more blocked in 1991, 1992, 1993, 1996 and 2003. Figure62 shows the interannual variability of the 5-member ensemble mean (blue line), averaged over all the forecasts starting from 1st January to 30 April and averaged over the period from day 12 to day 18. At this time range, the interannual variability of the monthly forecast is still strong, and the linear correlation between the interannual variability of the ensemble mean and the corresponding analysis is 0.54 (94% significance). This suggests that at this time range (days 12-18) the model has some skill in predicting the interannual variability of the blocking index over the Euro-Atlantic region. However, for the period days 19-32, the model displays little skill in predicting the interannual variability of the blocking index (Fig.63). As discussed above the model predicted less blocking than climatology for 2003, in contradiction with observations. The model also failed to predict that the flow in 1996 would be exceptionally blocked. On the other hand, it successfully predicted a more zonal flow in 2001 and 2002. The linear correlation between the model ensemble mean and the analysis is just 0.27.

As discussed above, the model seems to have much stronger skill in predicting the blocking index over the North Pacific than over the Euro-Atlantic region. Figure64 shows the interannual variability over the Central Pacific for days 12-18. At this time range, the model predicted successfully more blocking in 1995 and 2003, and a more zonal flow from 1999 to 2002, although it failed to predict significantly more blocking in 1997. The linear correlation between the ensemble mean and the analysis is 0.68 (99% significance), which is somewhat higher than the correlation obtained over the Euro-Atlantic sector (0.54). For the period days 19-32 (Fig.65), the model also seems to display some skill in predicting the interannual variability of the blocking index. At this time range, the linear correlation is 0.54 (94% significance), which is much higher than the linear correlation

*Figure 60: Same as Figure59 but for the 500 hPa geopotential height averaged from 16 February to 1st March 2003. The top right panel corresponds to the monthly forecast starting on 12 February 2003 (time range is days 5-18), and the bottom right panel corresponds to the monthly forecast starting on 29 January 2003 (time range is days 19-32).*
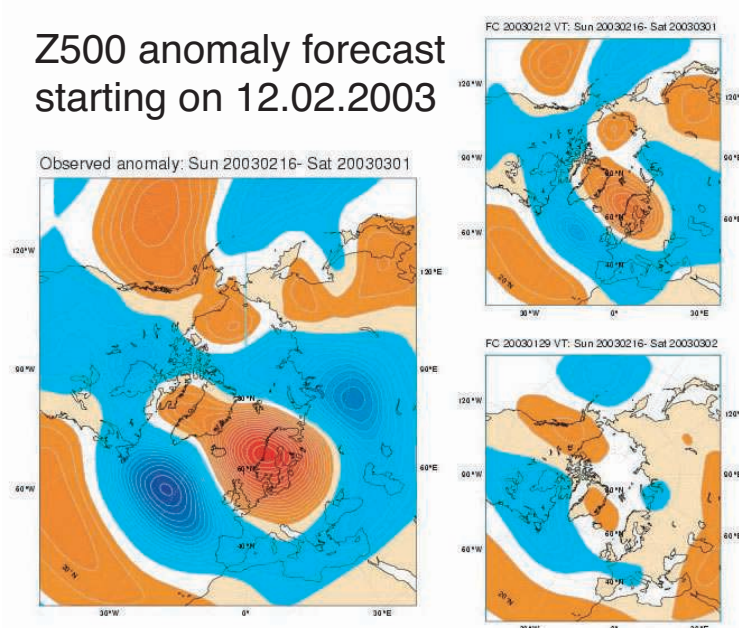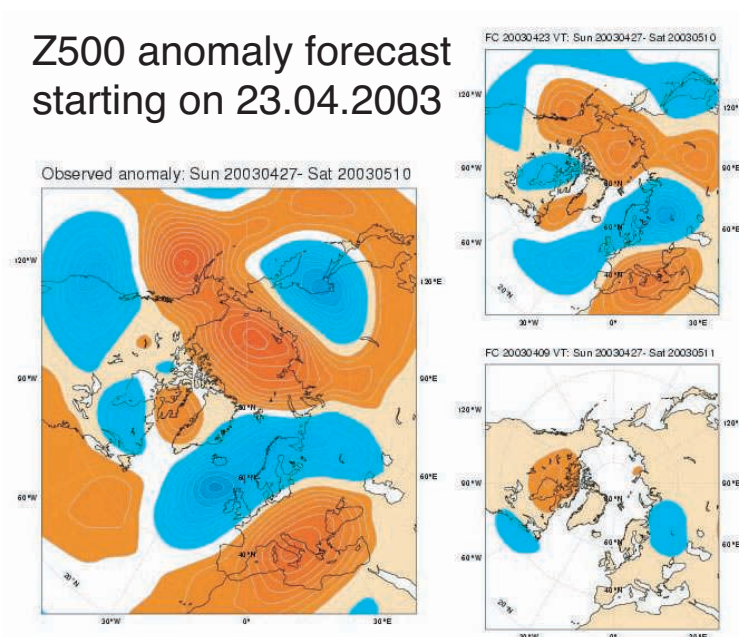


*Figure 61: Same as Figure59 but for the 500 hPa geopotential height averaged from 27 April to 10 May 2003. The top right panel corresponds to the monthly forecast starting on 23 April 2003 (time range is days 5-18), and the bottom right panel corresponds to the monthly forecast starting on 9 April 2003 (time range is days 19-32).*

obtained over the Euro-Atlantic sector (0.27). This result is in agreement with the results in the previous subsection obtained with the real-time monthly forecast over the period March 2002-April 2003. In the Euro-Atlantic sector, the model seems to have skill in predicting the blocking index up to about 20 days. The model has more skill over the North Pacific region, particularly for days 19-32. This may explain why Europe is a difficult region for the monthly forecast after day 20, whereas the monthly forecasting system still displays skill over North America after 20 days (Section 4).

# 6   The Madden-Julian Oscillation

The Madden-Julian Oscillation (MJO) is a 30-60 day oscillation and the dominant mode of variability in the Tropics on time scales exceeding one week and less than a season (Madden and Julian, 1971). It is a large-scale, quasi-periodic eastward-moving disturbance in the surface pressure, tropospheric temperature and zonal winds over the equatorial belt. The MJO has a significant impact on the Indian summer monsoon (Yasunari 1979). Clouds associated with the active phase of the monsoon propagate northward through the Indian Ocean and Indian subcontinent at about 1 degree latitude per day (Murakami 1976). Yasunari (1979) associated these northward moving clouds to the MJO. It has a significant impact on the Australian monsoon (Hendon and Liebman 1990) and may play an active role in the onset and development of an El-Niño event (Kessler and McPhaden 1995). It also has an impact on tropical cyclogenesis over the eastern North Pacific (Maloney et al 2000) and over the Gulf of Mexico (Mo, 2000).

It can also affect the extratropics. Ferranti et al (1990) demonstrated that the MJO had a significant impact on Northern Hemisphere weather, including over Europe. They demonstrated, that there was a link between the MJO and PNA and North Atlantic Oscillation (NAO) teleconnection patterns. They provided evidence that an improved representation of the MJO in the ECMWF forecast model (achieved in that case by relaxing the tropical circulation towards analysis) could lead to a considerable increase of skill for the extratropics after 10 days of forecast. This result suggests that the MJO is an important source of predictability for the intraseasonal time range (more than 10 days and less than a season). Therefore it is very important for the monthly forecasting system to have skill in predicting the evolution of the MJO.

In order to evaluate the skill of the monthly forecasting system to predict the MJO, an EOF analysis of velocity potential anomalies (relative to the past 12 -year climate) at 200 hPa has been performed on operational analyses and on each member of the monthly forecast. The velocity potential has been averaged between 5S and 5N along the whole equatorial band. For each member of the ensemble, all the 32-day forecasts of 200 hPa velocity potential anomalies (relative to the model climate) have been concatenated into a single file, before the EOF analysis was performed. The same process has been applied to the analysis, using ERA40 data to compute the climate. Figure66 displays the two dominant EOFs in the analysis (left) and in the control run (as an example) of the monthly forecasting system (right). In both analysis and monthly forecast, the two dominant EOFs each represent about 36% of the total variance of 200 hPa velocity potential, and together, they represent more than 70% of the total variance. The patterns are very similar in the analysis and in the monthly forecast.

The two dominant EOFs (EOF1 and EOF2) have a phase shift of about 90 degrees of longitude, which means that they capture the eastward propagation of velocity potential that is characteristic of an MJO event. For instance Figure67 displays a Hovmoller diagram (times goes from bottom to top) of velocity potential anomaly at 200 hPa, from the operational analysis for the period from 24 April to 26 May 2002 (left panel). This Hovmoller plot displays a clear eastward negative propagation, which is the signature of an MJO. The right panel of Figure67 displays the Hovmoller diagram of the velocity potential reconstructed using EOF1 and EOF2. The signal is much less noisy than when using the raw data, which is not surprising, since this EOF reconstruction method is equivalent to filtering the data. The eastward propagation is well represented in the

Blocking Index over over the Euro-Atlantic Sector
Forecast starting from 1st January to 30 April
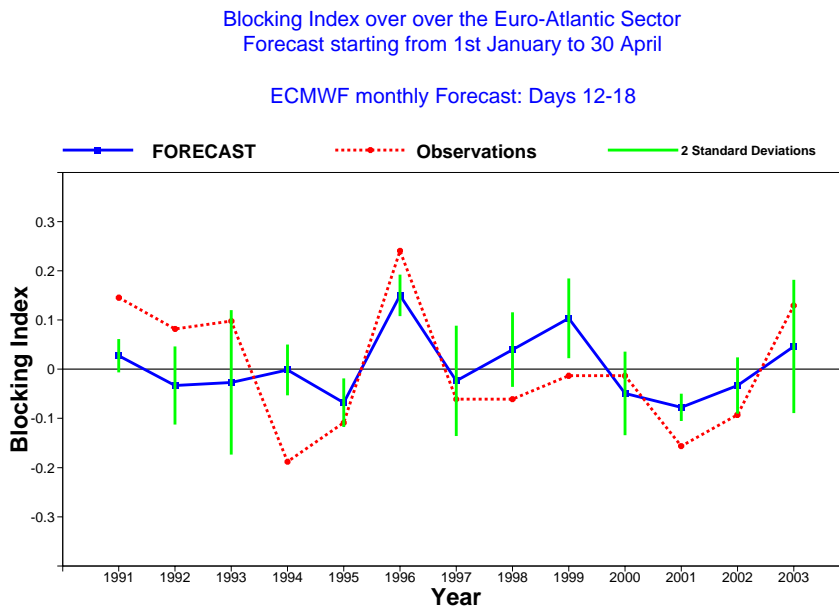
ECMWF monthly Forecast: Days 12-18



*Figure 62: Interannual variability of the Pelly-Hoskins index anomaly over the Euro-Atlantic region from 1991 to 2003. The index has been averaged from day 12 to day 18 and over all the forecasts between the 1st January to the 30 April (blue curve). The green line represents 2 standard deviations and the red line corresponds to the verification from ECMWF operational analysis and ERA40.*

Blocking Index over over the Euro-Atlantic Sector
Forecast starting from 1st January to 30 April
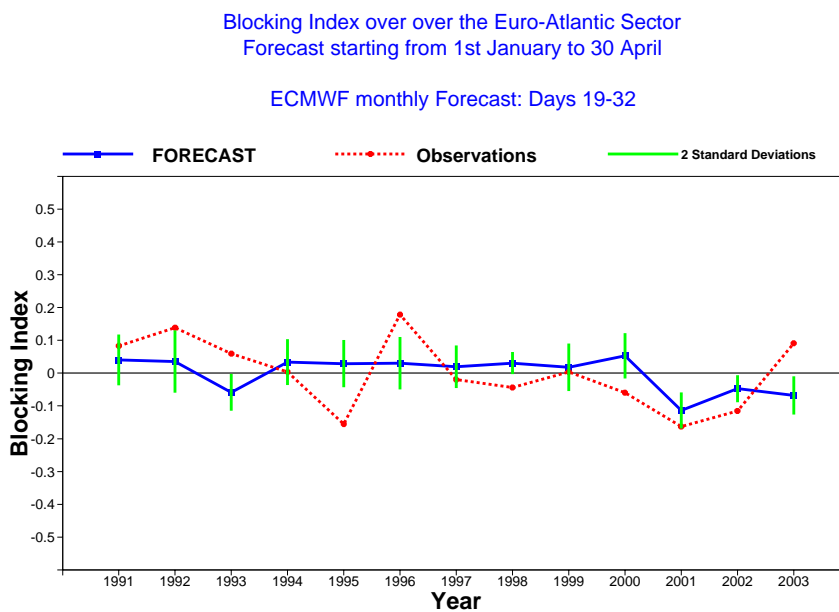
ECMWF monthly Forecast: Days 19-32



*Figure 63: Same as Figure 59, but the index has been averaged over the two-week period from day 19 to day 32.*

Blocking Index over over the Central Pacific Sector
Forecast starting from 1st January to 30 April
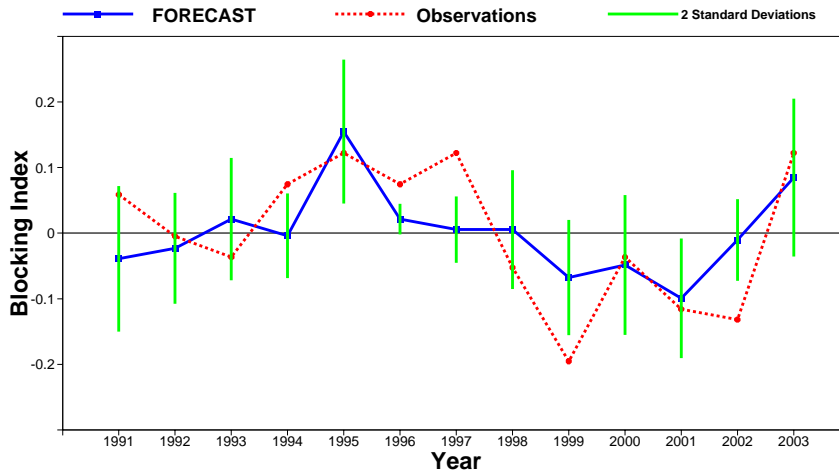
ECMWF monthly Forecast: Days 12-18



*Figure 64: Same as Figure 59 but for Central Pacific.*

Blocking Index over over the Central Pacific Sector
Forecast starting from 1st January to 30 April
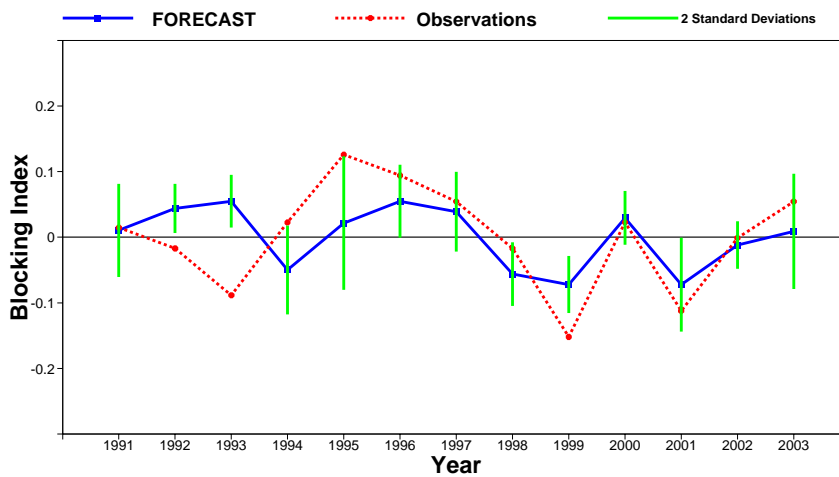
ECMWF monthly Forecast: Days 19-32



*Figure 65: Same as Figure 60 but for Central Pacific.*

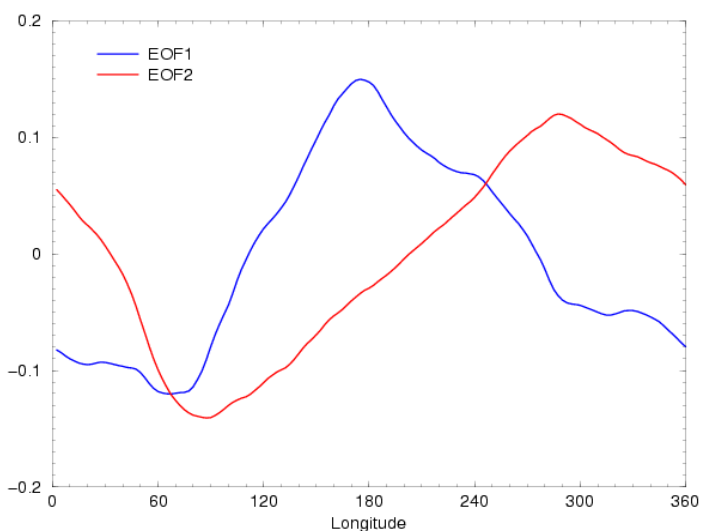reconstructed field, suggesting that the two EOFs capture the MJO propagation.

The same EOF reconstruction method has been applied to each member of the monthly forecasting system. The right panel in Figure68 shows the Hovmoller diagram of the EOF-reconstructed velocity potential of the control forecast starting on 27 March 2002. The y-axis represents the 32 days of the model integration. In this particular case, an MJO event was already present in the initial condition over the central Pacific (negative contours). The model predicts well its eastward propagation, in good agreement with observations (left panel), at least till about day 20. However, the predicted amplitude is only half of the observed amplitude after only a few days of integrations. After about 20 days, the control forecast did not predict any significant eastward propagation, unlike in observations. In Figure68, negative contours correspond to an MJO event.

In order to evaluate the skill of the monthly forecasting system to predict the MJO, for each longitudinal point (averaged from 5N to 5S) and for each time lag, the time series based on the 30 cases of the EOF-reconstructed velocity potential from the ensemble mean has been correlated with the time series computed using operational analysis and ERA40 (Figure69, middle panel). The linear correlation is above 0.6 up to 15 days over most longitudinal points, and remains high till about day 25. The linear correlations obtained are much higher than those obtained when persisting the initial condition (Fig.69, right panel). This suggests that the model has some skill in predicting the evolution of the MJO during at least the first 20 days of the forecast. The linear correlations are also higher than those obtained when using raw data of velocity potential instead of the EOF-reconstructed data. This is not surprising, since as mentioned above, the EOF-reconstructed data represents a filtered version of the velocity potential, and therefore is likely to be easier for the model to predict.

The skill of the monthly forecasting system to predict the Madden-Julian oscillation is confirmed when performing an anomaly correlation of the EOF-reconstructed velocity potential at 200 hPa averaged between 5S and 5N over the whole longitudinal band for each day of the forecast and for each case. The result has been averaged over the 30 cases. Figure70 displays the evolution of the anomaly correlation as a function of the time lag (x-axis). The blue line represents the scores obtained with the ensemble-mean. The anomaly correlation reaches 0.6 by about day 10, and stays around 0.5 till about day 25. After day 25, the anomaly correlation decreases quickly. This suggests that the monthly forecasting system has some robust skill in predicting the evolution of the MJO during at least the first 25 days of the forecast. The scores are higher than with the control forecast (Fig.70), which is not surprising, since ensemble-means generally perform better than individual ensemble members. The score is also higher than when persisting the initial state (red line in Figure70). The RMS error of the ensemble-mean is lower than the RMS error obtained with persistence of the initial conditions (Fig.70, right panel). Interestingly, it is also lower than the RMS error obtained with climatology till around day 20. Therefore, the model displays useful skill in predicting the evolution of the MJO for about 20 days. This is in good agreement with the skill of statistical forecasts, such as the one by Wheeler and Weickmann (2001). This skill is likely to improve the scores of the monthly forecasting system in the extratropics.

Although the monthly forecasting system displays some skill in predicting the evolution of the MJO up to 20 days, it does not maintain the intensity of the MJO for more than a few days (Fig.71). The variance of the principal component (PC1) which is a measure of the intensity of the MJO is reduced by almost a factor 2 after only 10 days into a forecast. The dramatic reduction of variance is also present in PC2 (Fig.72), although to a lesser extent. The time-scale of this decrease in the variance of PC1 and PC2 (about 10 days) is consistent with the time it takes the model to reach its radiative-convective equilibrium. Figure73 displays one member of the monthly forecasting system in the PC1 and PC2 phase-space. The black squares correspond to the forecasts after 24 hours, and the red squares correspond to the monthly forecast after 10 days. It appears clearly from this figure, that the red squares are much closer to the origin (0,0) than the black squares, which means that the amplitude of the MJO is significantly decreased after 10 days of forecast. The green squares correspond to the monthly forecasts after 30 days. The SAC paper 'Systematic model errors' by Thomas Jung discusses the reduction of MJO variance in more details. Interestingly, some experiments described in that SAC paper

## EOF1 and EOF2 from analysis
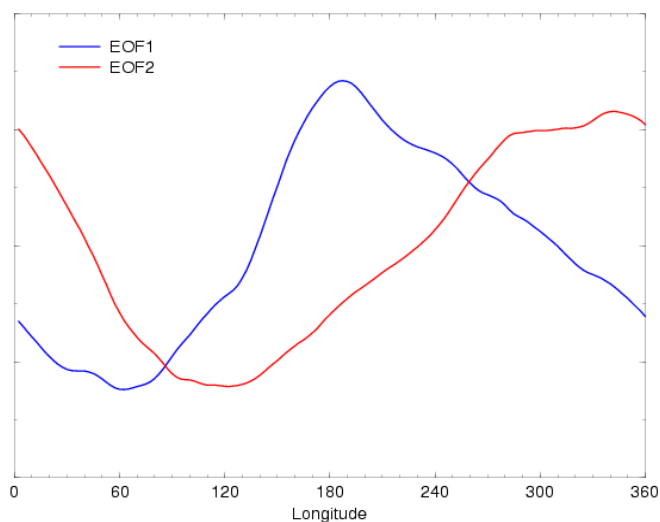
## EOF1 and EOF2 from MOFC



*Figure 66:   First 2 EOFs of velocity potential at 200 hPa computed from the analysis (left panel) and from the monthly forecasting system (right panel).  For the monthly forecasting system, all lead times are included. The blue curve corresponds to the first EOF and the red curve corresponds to the second EOF. The x-axis represents the longitude.*



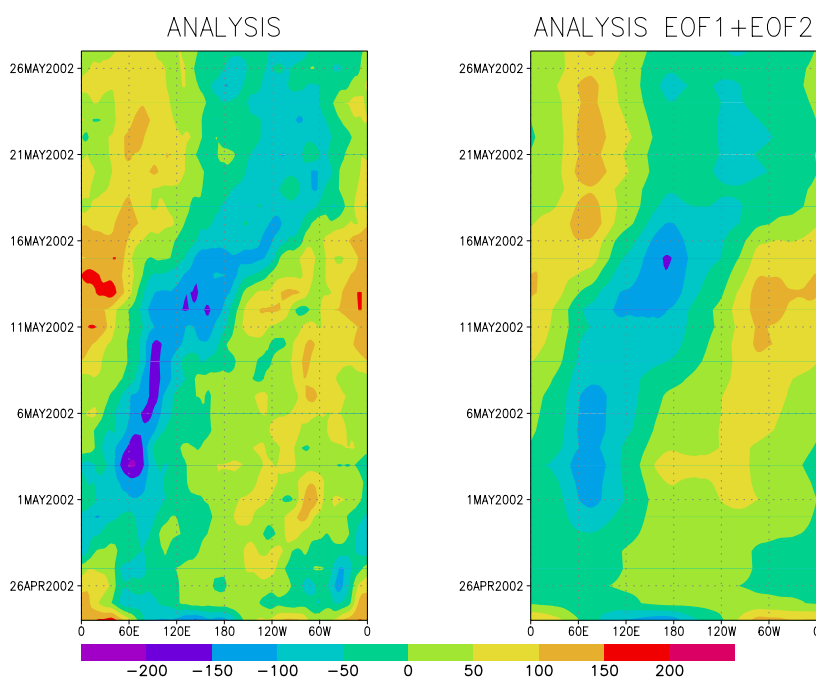*Figure 67:   Hovmoller diagram of 200 hPa velocity potential anomaly from 24 April to 26 May 2002 from ERA40. Blue colors represent negative values and red colors correspond to positive value.  The MJO corresponds to negative values.  The Hovmoller diagram of the raw data is displayed in the left panel, and the Hovmoller diagram after EOF reconstruction (from EOF1 and EOF2) is displayed in the right panel.*
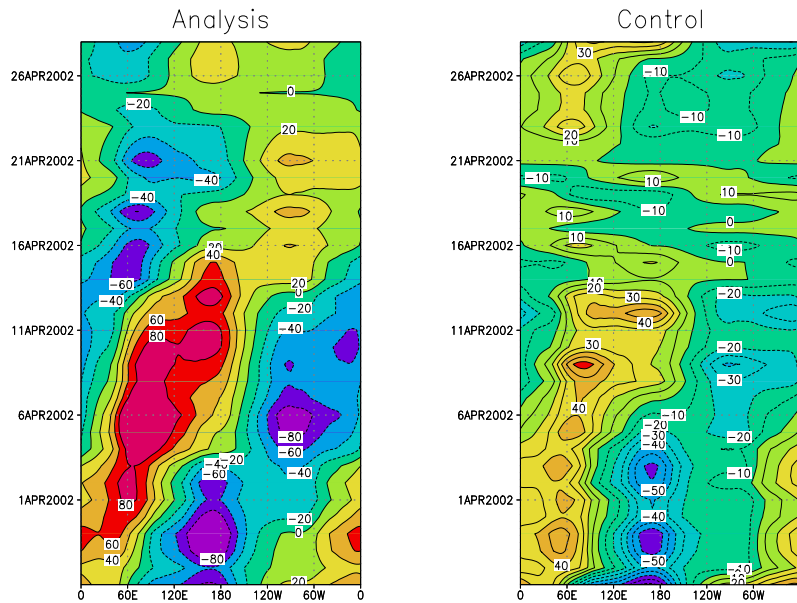
*Figure 68: Hovmoller diagram of 200 hPa velocity potential anomaly after EOF reconstruction from 27 March to 26 April 2002. The left panel shows the analysis and the right panel the control of the monthly forecast starting on 27 March 2002.*



*Figure 69: Hovmoller diagram of linear correlation between the time series of 200 hPa velocity potential anomaly predicted by the monthly forecasting system and the analysis (left panel). The middle panel shows the anomaly correlation obtained with the 200 hPa velocity potential anomaly after EOF reconstruction, instead of the raw data (the forecasts have been projected on the EOFs from the analysis). The right panel shows the anomaly correlation obtained with persistence of the initial condition (day 0). The correlations have been computed from the 30 cases from March 2002 to May 2003.*

*Figure 70: Anomaly correlation (left panel) and root mean square error (right panel) between the 200 hPa velocity potential anomaly after EOF reconstruction and the analysis. The b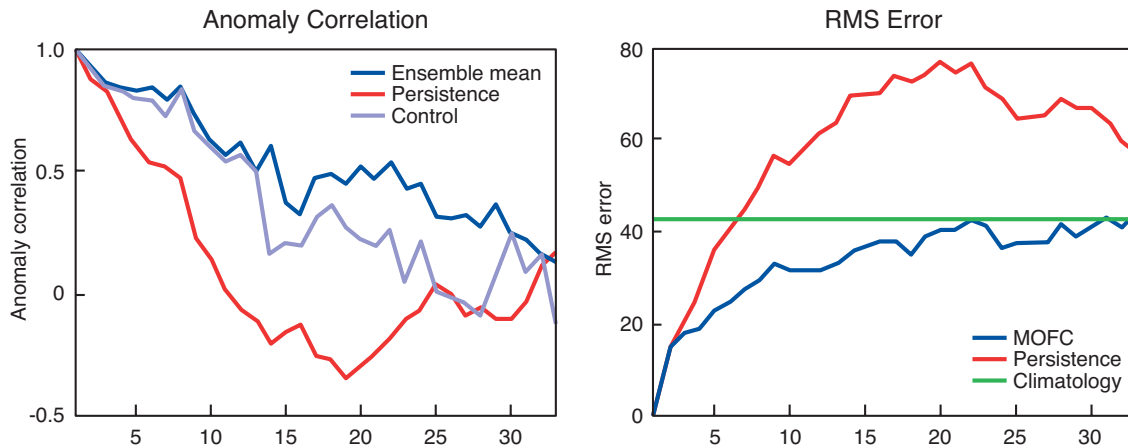lue line represents the scores obtained with the ensemble mean. The magenta line corresponds to the score obtained with the control forecast. The red line represents the scores obtained with the persistence of the initial condition. The green line represents the climatology. The scores have been computed from the 30 cases from March 2002 to May 2003.*

suggest that this problem may be resolution dependent.

All the 51 members of the monthly forecasting system have been projected into the PC1-PC2 phase-space. PC1 and PC2 are the two principal components associated with EOF1 and EOF2 computed from analysis and displayed in the left panel of Figure66). In the PC1-PC2 phase-space (see for instance Figure73), the distance from the origin represents the amplitude of the MJO and the positive x-axis (PC1 positive and PC2 equals to zero) corresponds to a velocity potential with the same shape as the EOF1 (blue curve) in Figure66 and therefore with a minimum at about 60E. Since the MJO is associated with *negative* values of the velocity potential, the positive x-axis corresponds roughly to an MJO at about 60E. In the same way, the positive y-axis corresponds to an MJO at about 90E. The negative x-axis corresponds to an MJO located at the dateline and the negative y-axis to an MJO at 60W. Therefore, the top right quadrant corresponds to the MJO over the Indian Ocean between approximatively 60E and 90E. The top left quadrant corresponds to the MJO over the central Pacific. The bottom left quadrant corresponds to an MJO further east between 180E and 60W. Finally the bottom right quadrant, corresponds to the MJO between 60W and 60E. So an MJO starting over the Indian Ocean and propagating all around the globe will be the arc of a circle starting in the top right quadrant in the PC1-PC2 phase space and moving anti-clockwise. The size of the domain corresponding to each quadrant varies since the two dominant EOFs (left panel in Figure66) are not exactly 90 degrees out of phase.

From the 30 cases that have been verified, it appears that the skill of the model to predict the MJO depends strongly on the initial position in the PC1-PC2 phase-space (Fig.74). In Figure 74, a cluster of bad cases at day 10 is present in the top right quadrant, which represent cases where the initial condition includes convection in the Indian Ocean. On the other hand, when the initial condition includes convection east of 90 East (negative values of PC1) the forecast system performs significantly better. Figure 74 suggests strongly that the monthly forecasting system has difficulties in propagating an MJO from the Indian Ocean to the Pacific. This is also true for verification at day 20, instead of day 10 (not shown). For day 20, most of the best cases have initial conditions clustered in regions of high negative PC1. When the initial condition includes an MJO in the Indian Ocean, the model seems to have difficulties in propagating the MJO across the maritime continent. The following examples (Fig. 75 to Fig. 78) illustrate this point.

**PC1**



*Figure 71: Time evolution of the amplitude of the principal component 1 (PC1). The black line corresponds to the analysis. Each green line represents the variance of one member of the ensemble averaged over the 30 cases. The solid red line represents the variance averaged over the 51 members of the ensemble. The variance has been computed over 30 cases from March 2002 to May 2003.*

**PC2**



*Figure 72: Same as Figure 71, but for the principal component 2 (PC2).*

In Figure 75, there is some MJO propagation over the Indian Ocean during the first days of the forecast. The model predicted well the propagation till day 5. However, when the MJO reached the eastern edge of the Indian Ocea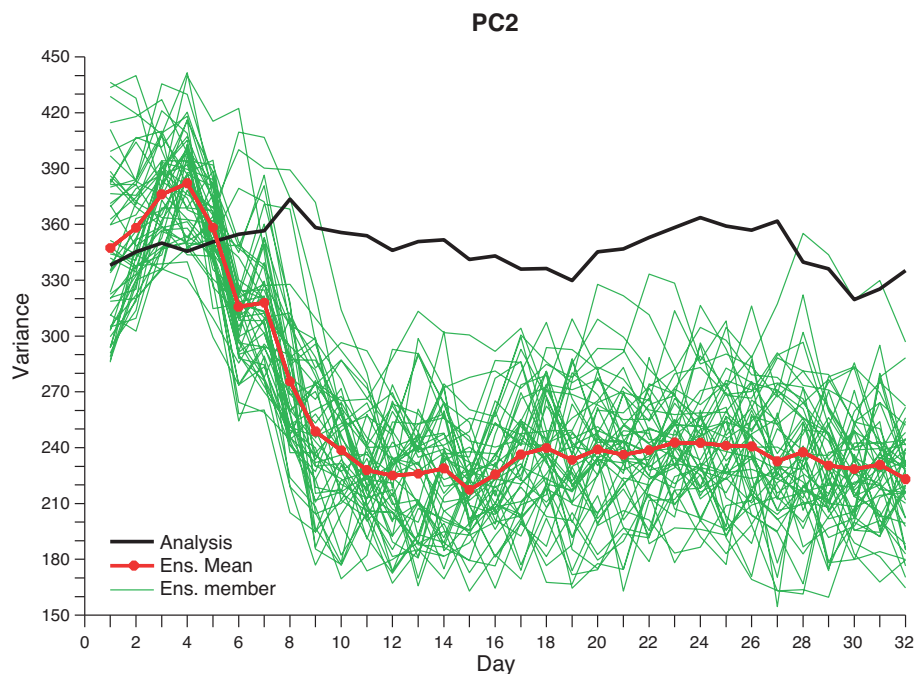n (around the positive y-axis), most members of the ensemble predict a weakening of the MJO for days 10 and 15, in contradiction to observations, where the MJO was in the west Pacific by day 15. Figure 76 shows another example of the monthly forecasting system failing to predict the propagation of the MJO over Indonesia. In Figure 76, the MJO starts over Indonesia. The forecast for day 5 was poor, with the model predicting the signal staying at the same place, but weakening. This tendency is confirmed for forecast days 10 and 15.

On the other hand, when the forecast starts with an MJO in the Pacific (Figures 77 and 78), then the model seems to successfully predict its propagation. In Figure 77, for instance, the initial condition includes an MJO event in the Central Pacific. After 20 days of forecast, 80% of the members of the ensemble predict its occurrence in the eastern Pacific and the observed position is inside the cloud defined by all the ensemble members. In Figure 78, the initial condition includes an MJO event near the dateline. The model successfully predicted its propagation during the first 20 days of the forecast. At day 20, the majority of ensemble members are in the same quadrant as the analysis, which, once again, is inside the cloud defined by all the monthly forecast ensemble members.

The monthly forecasting system does not have problems in producing eastward propagation of convection in the Tropics, but the model seems to have difficulty in propagating the convection from the Indian Ocean to the Pacific Ocean. This may be due to errors in the model mean state. The fact that the model physics did not maintain the intensity of the MJO suggests that part of the skill of the model to predict the propagation of the MJO was probably 'wasted' and did not contribute as much as expected to improve the skill of the monthly forecast system in the extratropics after 10 days.

# 7  Other sources of predictability in the intraseasonal time scale

This section investigates the skill of the monthly forecasting system to predict some modes of variability other than the Madden-Julian oscillation that are important for predictability between 10 and 30 days. First, the skill of the coupled system to predict SSTs will be evaluated as this is likely to impact its skill in predicting the atmospheric circulation. In particular, a rapid warming or cooling of SSTs in the tropical eastern and central Pacific due to ENSO is likely to have an impact on the atmospheric circulation in the intraseasonal time range. The skill of the model to predict the evolution of ENSO indices will be discussed in Section 7.2. Finally, for Europe, an important source of predictability comes from the North-Atlantic oscillation (NAO), which will be discussed in Section 7.3.

## 7.1  SSTs

The skill of the model to predict SSTs has been evaluated in a similar way to surface temperature over land in Section 3. Probabilistic scores have been computed over the 30 real-time cases. The event considered in this section is the probability that the SST anomalies are in the upper or lower tercile. As in Section 2, SSTs have been averaged over different periods: days 5-11, days 12-18 and days 19-32.

ROC scores for days 12-18 are in general larger than 0.7, and also generally slightly higher than the scores obtained by persisting SST anomalies from days 5-11. For the period days 19-32 (Fig.79), the ROC score is also higher than 0.7 over most areas. However, there are some tropical regions like the North Indian Ocean or Tropical Western Pacific, where the ROC score was much lower (Fig.79). Figure79 indicates that the SSTs predicted by the coupled system display a ROC score generally higher than the ROC score obtained by persist-
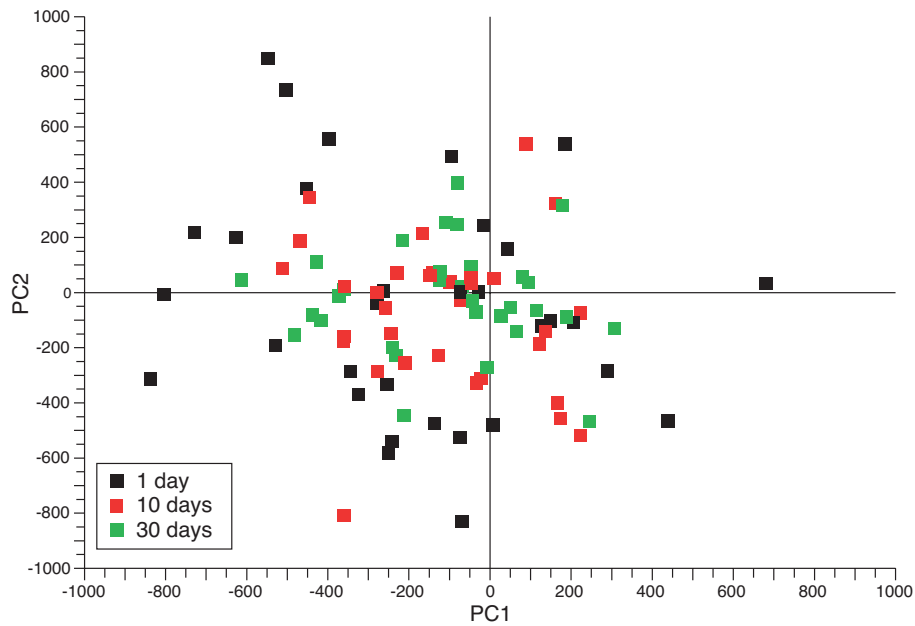
*Figure 73: Control of the monthly forecasting system in the PC1-PC2 phase space. analysis. The dotted line represents the position averaged over all the members of 30 cases). The red squares correspond to the monthly forecast after 10 days. The green squares correspond to the monthly forecast after 30 days.*



*Figure 74: Initial positions of the 30 real-time monthly forecasts from March 2002 to May 2003 in the PC1-PC2 phase-space. The blue circles represent the positions of the 8 best cases (25% of the cases) and the red diamonds represent the positions of the 8 worst cases at day 10. The skill of the forecast has been measured using the RMS error between the position of the analysis in the phase-space at day 10 and each individual member of the ensemble.*
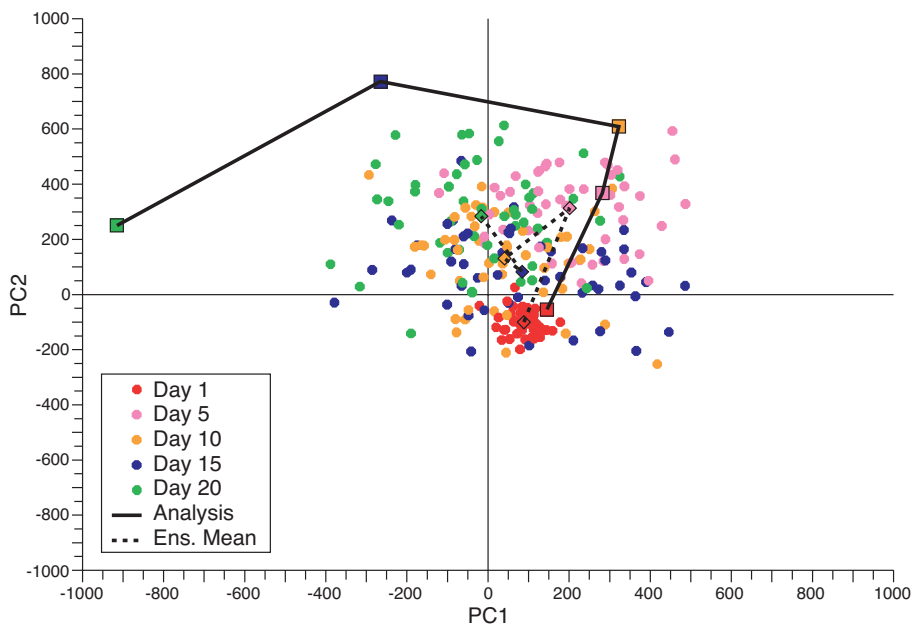
*Figure 75: Monthly forecast starting on 24 April 2002 in the PC1-PC2 phase space. Each circle represents one member of the monthly forecast after 1 day (red), 5 days (magenta), 10 days (orange), 15 days (blue) and 20 days (green). The big squares and the solid line represent the operational analysis. The dotted line represents the position averaged over all the members of the ensemble. The top right quadrant corresponds to an MJO activity between in the Indian Ocean between 60E to 90E. The top left quadrant corresponds to an MJO activity between 90E and 180E. The bottom left quadrant corresponsd to an MJO east of the dateline but west of 60W. Finally, the bottom right quadrant corresponds to an MJO activity between 60W and 60E.*
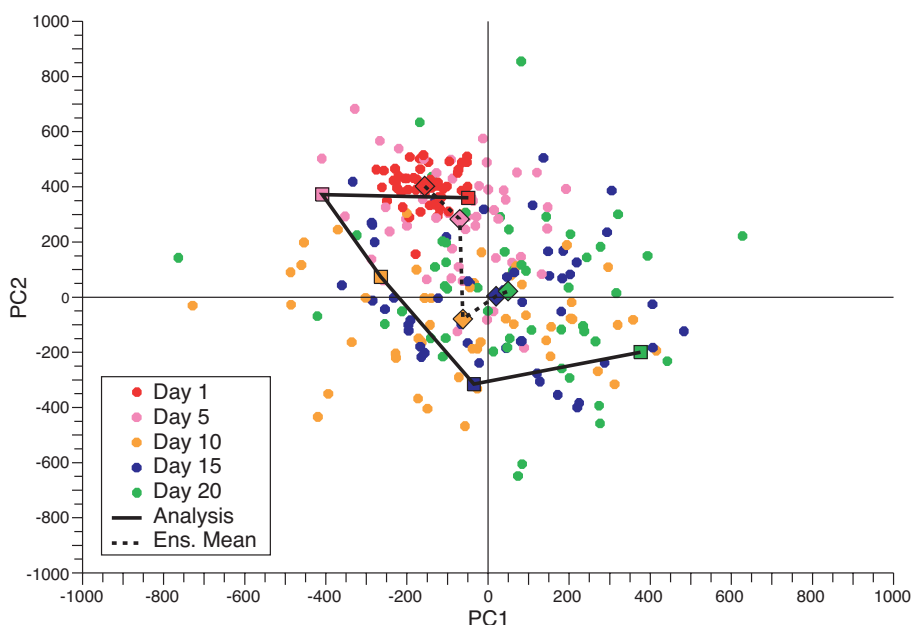


*Figure 76: Same as Figure 75 but for the monthly forecast starting on 26 March 2003.*
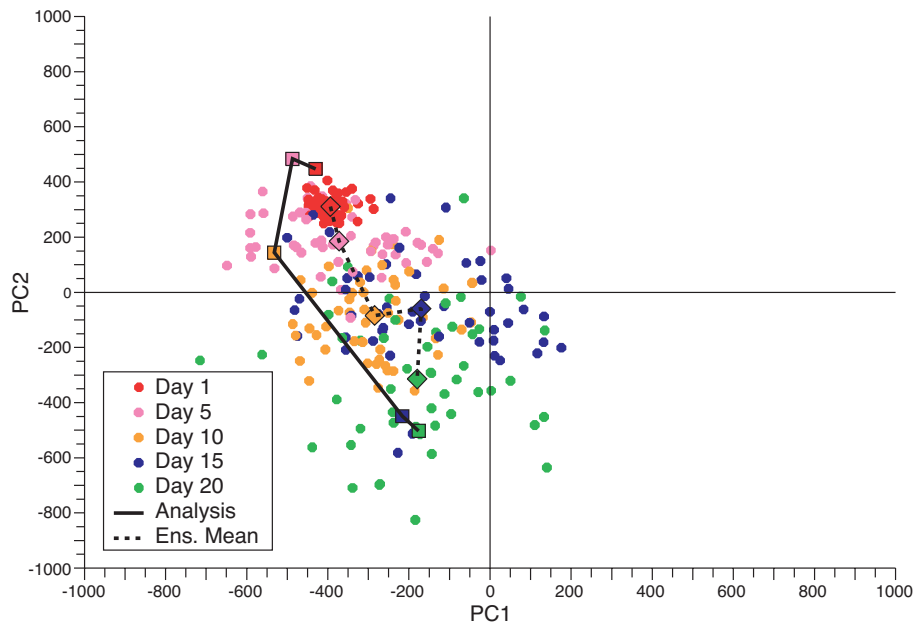
*Figure 77: Same as Figure75 but for the monthly forecast starting on 1 January 2003.*
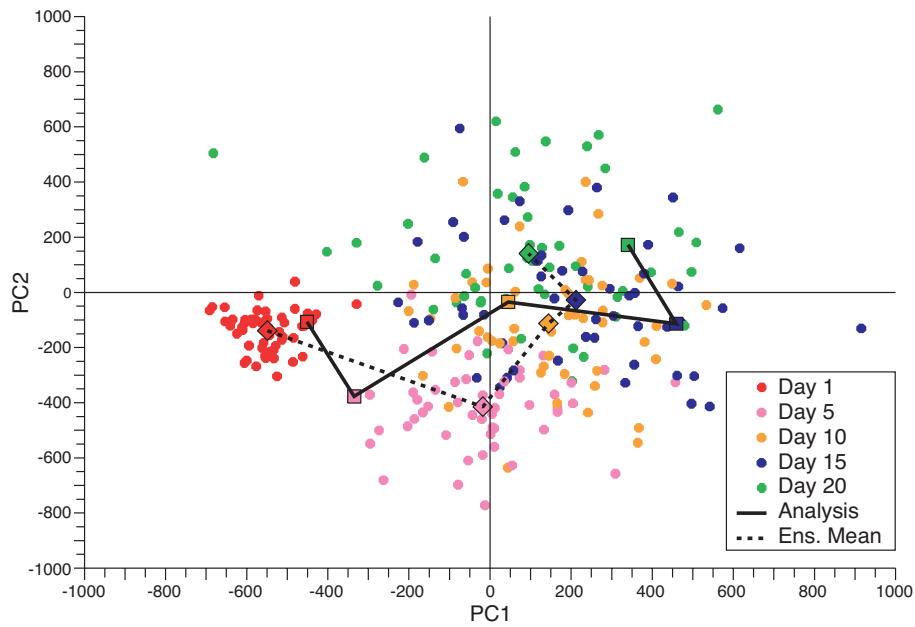


*Figure 78: Same as Figure75 but for the monthly forecast starting on 4 June 2003.*

## a) Monthly Forecast
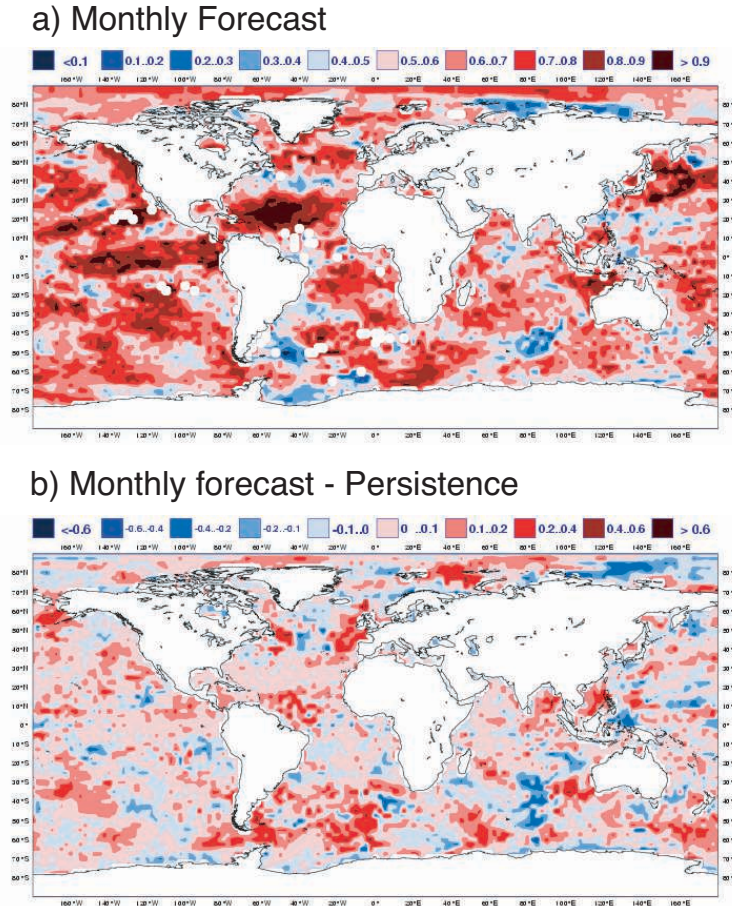


## b) Monthly forecast - Persistence



*Figure 79: ROC score over each ocean point of the probability that the sea surface temperature averaged from days 19 to 32 is in the upper tercile (top panel). Red colors indicate a ROC area larger than 0.5, and blue colors indicate ROC areas less than 0.5 (model is less skillful than persistence). The bottom panel displays the difference of ROC areas between the monthly forecast and the persistence of the SST anomalies from days 5 to 18. Red colours indicate that the monthly forecasting system performs better than persistence, and blue colours indicate that persistence performs better than the monthly forecasting system.*

ing the SST anomalies of the previous two weeks (days 5-18). This result suggests, that at least over the period April 2002-May 2003, the coupled system displayed some skill in predicting the evolution of SSTs. This is particularly clear over the Northern Atlantic, which is likely to impact the weather over Europe. This result confirms that using coupled ocean-atmosphere integrations instead of forcing the atmospheric component by persisted SSTs as in EPS is probably a good strategy for the intraseasonal time range, which is likely to improve the skill of the system in the time range from days 10 to 20.

## 7.2 ENSO

A period of 1 year of validation is not enough to study the El-Niño Southern Oscillation (ENSO). Therefore, the 12-year model hindcasts, used to calibrate the monthly forecasting system, have been used to evaluate the skill of the coupled model to predict the evolution of ENSO, although 12 years is still not enough, since only a few ENSO events occurred during this period. Nevertheless, ENSO indices based on SSTs averaged over

**a) NINO3**

**b) NINO 34**



*Figure 80: RMS errors (top panels) and anomaly correlations (bottom panels) of NINO3 (left panels) and NINO3.4 (right panels) indices as a function of the forecast lead time. The green solid line represents the scores of the ensemble mean of the monthly forecasting system based on 30 cases. The vertical lines represent 2 standard deviations computed over the 30 cases. The red line represents the scores obtained by persisting the initial condition. The blue line corresponds to the score of the climatology.*



*Figure 81: Same as Figure 80 but for weekly-mean values of the SOI index.*

NINO3 (90W-150E, 5N-5S)) and NINO3.4 (170W-120W, 5N-5S) have been computed for all the monthly hindcasts and compared to the SST analyses for each day of the forecast (Reynolds OIv2) (Fig.80). The anomaly correlations of the monthly forecast of the NINO3 index anomalies are higher than those obtained by persisting the initial anomalies of NINO3 index. For the NINO3.4 index, the monthly forecast performs better than persistence after about day 10, indicating that the monthly forecasting system has some skill in predicting the evolution of the NINO indices.

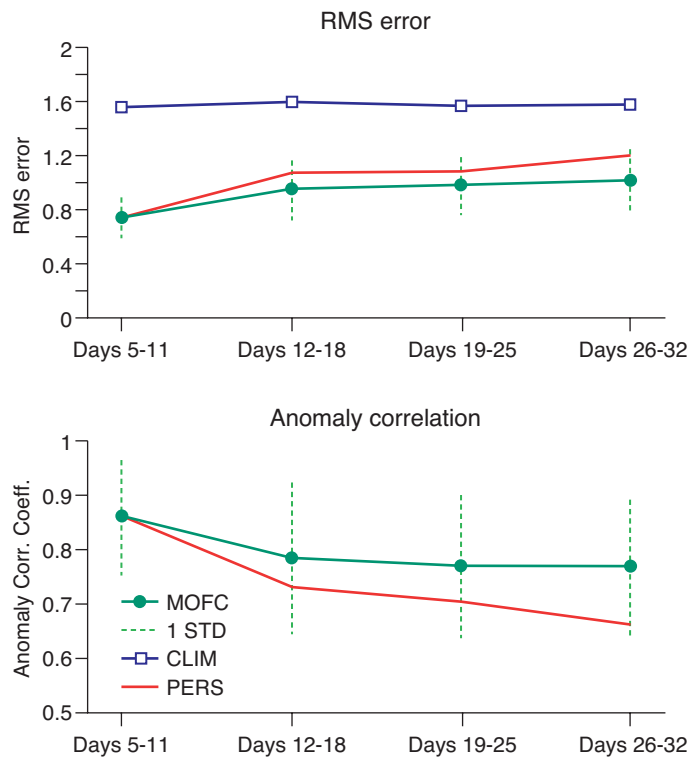Another widely used index for ENSO, is the Southern Oscillation Index (SOI), which is defined as the difference in pressure between Darwin and Tahiti. As for the NINO indices, the index has been applied to the 12-years of monthly hindcasts, and to ERA40 for validation. Anomaly correlations and RMS error have been computed to assess the skill of the monthly forecasting system (Fig.81). The monthly forecasting system has skill in predicting the weekly-mean of the SOI index and this skill exceeds the skill obtained when persisting the forecast of the previous week. This result confirms that the coupled ocean-atmospheric model has some useful skill in predicting the evolution of ENSO during the first 30 days of the forecast.

## 7.3   NAO

The North Atlantic Oscillation (NAO) is known to have a significant impact on the weather over Europe. During a positive phase of the NAO (geopotential height anomaly at 500 hPa is positive over Europe and negative over Greenland), there is more precipitation over Scandinavia, drier conditions over Spain and Portugal, and Central and North Europe tend to be warmer whereas Greenland tends to be cooler. The impact of a negative phase of NAO is the opposite of the impact of the positive phase of the NAO.

To evaluate the skill of the monthly forecasting system to predict the NAO, an NAO index has been applied to all the 5-member monthly hindcasts. The index was computed by projecting the predicted geopotential height at 500 hPa into the first EOF obtained from NCEP Reanalysis (1948 to 2000) over the region (90W-60E, 20N-90N). To focus on the winter season, only those forecasts were used that start in the period 15 November to 15 March of the following years from 1990 to 2002 (108 monthly forecasts). Anomaly correlation coefficients for 5-member ensemble-mean forecasts are about 0.78 for days 5-11, 0.50 for days 12-18, 0.30 for days 19-25 and 0.08 for the last week of the monthly forecasts (days 26-32) (Fig.82). Particularly for days 12-18 and days 19-25, the monthly forecasting system is significantly more skillful than the forecasts based on persistence in which the mean forecast anomalies from the previous seven-day period were persisted.

# 8   Summary and discussion

## 8.1   Summary

A monthly forecasting system has been set up at ECMWF. This system which is based on coupled ocean-atmosphere integrations has been run routinely since March 2002. Real-time forecasts are produced every two weeks. The present paper discusses the verification of 30 cases.

During the first 10 days of the forecast, the skill of the monthly forecasting system is close to that of the EPS. Deterministic (anomaly correlations or RMS scores) and probablistic scores, such as ROC or Brier skill scores, indicate that EPS performs slightly better than the monthly forecasting system during that period of time, most likely because of its finer horizontal resolution ($T_L255$ instead of $T_L159$), but the differences are small. This seems to indicate that the ocean-atmosphere coupling does not seem to significantly affect the scores in the medium-range. However, between day 10 and day 20 the control run of the monthly forecasting system seems
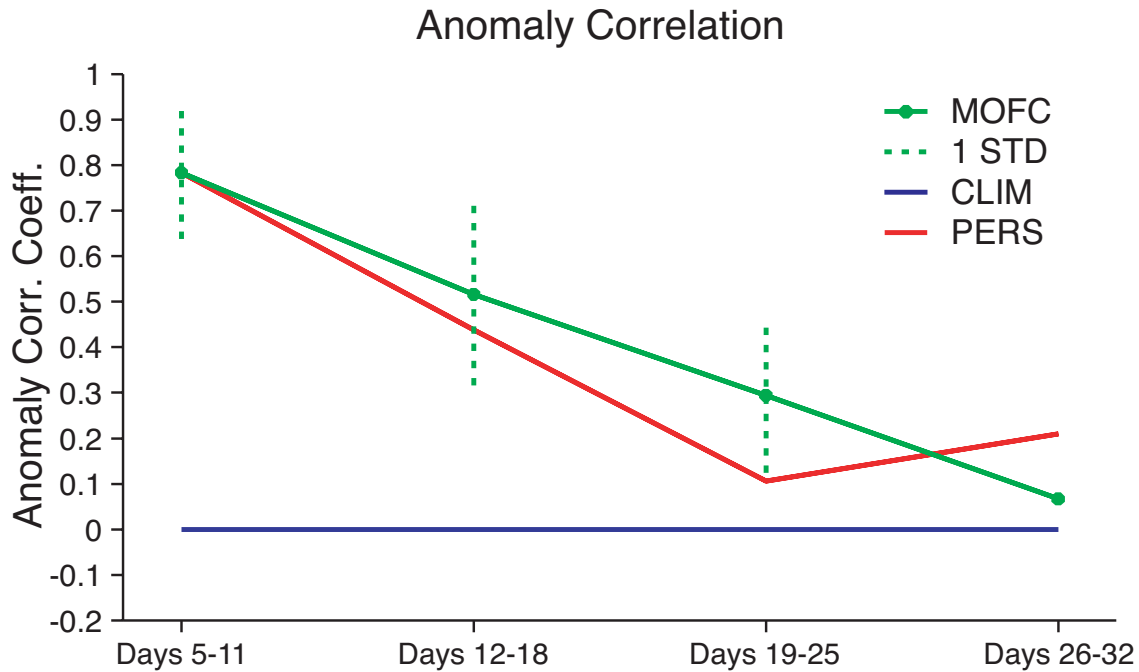
## Anomaly Correlation



*Figure 82: Same as Figure81 but for the anomaly correlation of weekly-mean values of the NAO index from 15 October to 15 February.*

to perform better than the control run of the EPS over the Asian and Pacific region (only the control member of the EPS is integrated for 20 days). This is likely due to the impact of the ocean-atmosphere coupling. The coupled system has some skill in predicting the variability in tropical SSTs. During the first 10 days of forecasts, this is unlikely to make a significant difference, but after 10 days, the variability in tropical SSTs may be strong enough to start to impact significantly the extratropical atmospheric circulation. The North Pacific and Asian regions would be the first to be affected by a change in tropical Pacific SSTs. The fact that the monthly forecasting system seems to perform better than EPS after day 10 over some regions supports the choice of a coupled ocean-atmosphere system for monthly forecasting rather than the use of persisted SSTs as in EPS.

Over the period days 12-18, the monthly forecasting system produces forecasts that are generally better than climatology or persistence. Therefore, the monthly forecasting system is probably useful for forecasts at this time range. Summer seems to be a difficult season as in the medium-range. The monthly forecasting system for days 12-18 beats persistence of days 5-11 in almost all cases when using the Brier skill score. During the period of verification, the model was particularly skillful over North America, Central Asia and the Southern Extratropics. Over Europe, the model was less successful, particularly over Scandinavia and eastern Europe.

During the two following weeks of the monthly forecast from day 19 to day 32, the coupled model performs generally better than persistence. At this time range, the model's skill increases with higher thresholds. The model shows some skill over some areas like North America and the Southern Extratropics. Europe on the other hand seems to be a difficult region after 20 days of forecast.

The coupled model seems to have skill in predicting the propagation of the MJO during at least the first 20 days of the forecast, which is comparable to the skill of some statistical models, although the amplitude is underestimated by the monthly forecasting system. The model seems also to have skill in predicting the ENSO and NAO variability till about day 20. Blocking indices have been applied to the monthly forecasting system. The results suggest that the model has some moderate skill in predicting the variability of the blocking index

over the Euro-Atlantic region till about day 20, but it has no skill at all after day 20. Over the Central Pacific region, the model displays much stronger skill in predicting blocking, which may explain why the model has more skill over Asia and North America than over Europe over the period days 12-18.

## 8.2   Discussion

The main conclusion of the present paper is that the monthly forecasting system performs clearly better than climatology and persistence till about day 20. Therefore, there is no reason for stopping forecasts at day 10. Forecasts between day 10 and 20 can be useful. After 20 days, the forecasts can still be useful over some regions such as North America or Southern Extratropics, but over Europe the scores were not significantly better than climatology. However, the results presented in this paper originate from little more than 1 year of verification. It is likely that the skill can display some interannual variability. In particular, it is possible that the conclusion about European skill may be too pessimistic. This year has been quite exceptional over Europe, with a persistent blocking situation during last winter and last spring. Atmospheric GCMs have difficulties in maintaining a blocking event for more than a few days. The NAO is an important source of long-term predictability over Europe and Section 5 shows that the monthly forecasting system has some skill in predicting the NAO index till about day 20. However, the NAO index was close to normal during most of last winter. Therefore, last winter may have been a particularly tough test for monthly forecasting over Europe, and it is likely that the model should perform better over Europe in some years. On the other hand, the development of an El-Niño event in the tropical Pacific during the period of verification may have increased the skill of the model over the Pacific and North America. It is likely that during years without El Niño or La Niña, the predictability over North America will be reduced.

Model improvements may also improve the skill of the monthly forecasting system over Europe in the coming years. Section 3 presented the example of the European floods where an extratropical downstream propagating wave played an important role in the occurrence of heavy precipitation over Central Europe. The model successfully predicted the occurrence of this wave, but an error in the phase-shift when the wave moved from North America to the Atlantic prevented the model from predicting the trough moving over Central Europe that was responsible for the flooding. This error in the phase shift appears already in the EPS after about day 6. If this is due to model errors and not to a lack of predictability, then improvements in the simulation of such downstream propagating waves is therefore likely to benefit the monthly forecasting system, and increase its skill after day 20. However, it is not clear at present what is the predictability of such downstream propagating waves.

The model shows some skill in predicting the evolution of the MJO till about day 20, though the variance of the velocity potential at 200 hPa is dramatically reduced after a few days of integration. The performance of the model may depend on the IFS cycle, the resolution and the time-step. Further work will investigate the sensitivity of the MJO simulation to those factors. It is likely that the monthly forecasts over the extratropics can be improved.

Further studies will be conducted on the monthly forecasting system. In terms of diagnostics and forecasts, more attention will be given to evaluate the skill of the system to predict the occurrence of extreme events during a given period of time. The present analysis focused essentially on predicting time-averaged anomalies (over one or two weeks). The extension of the extreme forecast index (EFI, see Lalaurette 2002) to the monthly forecasting system may be a way of evaluating the skill of this system to predict extreme events. Further plans also include investigating the impact of higher resolution during the first 10 days of forecasts. A first series of experiments will investigate the possibility of using the EPS forecasts at day 10 as atmospheric initial conditions for monthly forecasts from day 10 to day 32. Oceanic initial conditions would be provided by integrating the ocean model for 10 days, forced by the fluxes predicted by the EPS. This configuration would

save some computer resources, since the coupled system would be integrated for 22 days instead of 32 days. The cost of the ocean-only integrations (used for the 10 first days of the forecast) is small compared to the cost of the fully coupled ocean-atmospheric system. In addition, the 10 first days of the forecasts (the EPS) will have a higher horizontal resolution than the current monthly forecasting system, which could help to improve the forecasts after day 10. A second series of experiments will investigate the possibility of using a coupled ocean-atmospheric model during the 32 days of integrations as in the present monthly forecasting system, but with the same horizontal resolution as EPS ($T_L$255) till day 10, and a resolution of $T_L$159 afterwards. This will allow a clean comparison with EPS in order to assess the impact of the ocean-atmosphere coupling in the medium-range. The only difference with the first set of experiments, is that the 10 first days of integrations will be coupled, whereas the first setting is based on atmosphere only integrations till day 10 (the EPS).

REFERENCES

Buizza, R. and T.N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Buizza, R. M. Miller and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Chessa, P.A., 2000: Classifications and validation of the ECMWF EPS perturbed forecasts using pre-defined weather regimes. *ECMWF technical memorandum*, **317**, 29 pp.

Ferranti, L., T.N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177-2199.

Flatau, M., P. J. Flatau, P. Phoebus, P. Niiler, 1997: The Feedback between Equatorial Convection and Local Radiative and Evaporative Processes: The Implications for Intraseasonal Oscillations. *J. Atmos. Sci.*, **54**, 2373-2386.

Grazzini, F. and G. van den Grijn, 2003: Central European floods during summer 2002. *ECMWF Newsletter*, **96**, 18-28.

Hendon, H.H., and B. Liebmann, 1990: A composite study of onset of Australian summer monsoon. *J. Atmos. Sci.*, **47**, 2909-2923.

Inness, P. M., Slingo, J. M., 2003: Simulation of the MaddenJulian Oscillation in a Coupled General Circulation Model. Part I: Comparison with Observations and an Atmosphere-Only GCM. *J. Climate* **16**, 345-364.

Kessler, K.S. and M.J. McPhaden, 1995: Oceanic equatorial Kelvin waves and the 1991-1993 El-Nino. *J. Climate*, **8**, 1757-1774.

Knuth, D.E., 1981: Seminumerical Algorithms. Vol. 2. The art of computer programming, Addison-Wesley, 688 pp.

Lalaurette, F., 2002: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *ECMWF Tech. Memorandum*. 373.

Madden, R.A., and P.R. Julian, 1971: Description of the 40-50 day oscillation in the Tropics. *J. Atmos. Sci.*, **43**, 3138-3158.

Maloney, E.D. and D.L. Hartmann, 2000: Modulation of eastern North Pacific hurricanes by the Madden-Julian oscillation. *J. Climate*, **13**, 1451-1460.

Mo, K.C., 2000: The association between intraseasonal oscillations and tropical storms in the Atlantic basin, *Mon. Wea. Rev.*, **128**, 4097-4107.

Murakami, T., 1976: Cloudiness fluctuations during the summer monsoon. *J. Meteor. Soc. Japan*, **54**, 175-181.

Palmer, T.N. 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.

Pelly, J.L., and B.J. Hoskins: A New Perspective on Blocking. *J. Atmos. Sci.*, **60**, 743-755.

Puri, K., Barkmeijer, J., Palmer, T.N., 2001: Tropical singular vectors computed with linearized diabatic physics. *Quart. J. Roy. Meteor. Soc.*, **127**, 709-731.

Richardson, D.s., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649-667.

Terray, L., E. Sevault, E. Guilyardi and O. Thual, 1995: The OASIS coupler user guide version 2.0, *CERFACS, technical report*, TR/CMGC/95-46.

Wheeler, M., and K.M. Weickmann, 2001: Real-Time Monitoring and Prediction of Modes of Coherent Synoptic to Intraseasonal Tropical Variability. *Mon. Wea. Rev.*, **129**, 2677-2694.

Wolff, J.O., E. Maier-Raimer and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model. *Deutches klimarechenzentrum, Hamburg, Technical Report*, **13**, 98 pp.

Wonnacott, T.H., and R.J. Wonnacott, 1977: Introductory Statistics, John Wiley, 650 pp.

Yasunari, T., 1979: Cloudiness fluctuations associated with the Northern Hemisphere summer monsoon. *Meteor. Soc. of Japan*, **58**, 225-229.