TECHNICAL MEMORANDUM

738

Sub-seasonal predictions

F. Vitart, G. Balsamo, R. Buizza, L. Ferranti, S. Keeley,
L. Magnusson, F. Molteni and A. Weisheimer

Research Department

October 2014

Special Topic paper on sub-seasonal predictions
presented at the 43rd ECMWF Scientific Advisory Committee, Reading, UK

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

ECMWF has produced 32-day forecasts routinely since March 2002 and operationally since October 2004 to fill the gap between medium-range and seasonal forecasts. The skill of the sub-seasonal forecasts at ECMWF has improved significantly over the past decade. This improvement can be linked to improved skill to predict the Madden Julian Oscillation (MJO), an important source of predictability for the sub-seasonal time range with an average gain of about 1 day of prediction skill per year, and improved tropical-extra-tropical teleconnections associated to the MJO. However, the MJO teleconnections in the Euro-Atlantic sector are still too weak compared to re-analysis. The skill of the ECMWF monthly forecasts to predict the North Atlantic Oscillation (NAO) and sudden stratospheric warmings (SSW), another source of sub-seasonal predictability, has also improved over the past 10 years, although the downward propagations associated to SSWs is much weaker in the model than in ERA Interim. Monthly forecasts display some skill in predicting heat waves, although with a weaker amplitude and mostly when the anti-cyclonic circulation was already present in the initial conditions.

Further improvements in the ECMWF sub-seasonal forecasts are expected in the coming years with, among other changes, the introduction of a dynamic sea-ice model, increased atmospheric and oceanic resolutions and the extension of the re-forecasts. Results from the MINERVA experiment suggest a modest improvement in sub-seasonal skill scores when increasing the horizontal resolution of the atmospheric model from T319 to T639 and T1279. In particular, the NAO forecast skill scores improve when increasing the horizontal resolution from T319 to T639. The extension of the re-forecasts will lead to a more accurate estimation of the forecast anomalies and probabilities. Preliminary results of monthly forecasts with an active sea-ice model are encouraging since these integrations show skill in predicting the evolution of sea ice in the sub-seasonal time-range, particularly in the ice edge. These past and future improvements in the monthly forecasts should make it possible now to extend the monthly forecasts to 45 or 60 days and produce skilful and reliable forecasts beyond day 30.

1 Introduction

The sub-seasonal time range has so far received much less attention than the medium-range or the seasonal time scale as it has long been considered as a “predictability desert”. However, recent research has indicated that important potential sources of predictability for this time range can be exploited through better representation of atmospheric phenomena such as the Madden Julian Oscillation (MJO) and improved coupling with, and initialisation of, the land-ocean-cryosphere and stratosphere. Therefore, there is currently a growing interest in the scientific, operational and applications communities in developing forecasts that fill the gap between medium-range weather forecasts (up to 2 weeks) and long-range or seasonal ones (3–6 months).

ECMWF was one of the first operational centres to produce operational monthly forecasts, about 10 years ago, using at that time a stand-alone forecasting system based on coupled ocean-atmosphere integrations with a T159 atmospheric resolution (Vitart 2004). The coupled model

was run routinely first as an experimental system since March 2002, and became operational in October 2004. Now the vast majority of Global Producing Centres (GPCs) produce operational sub-seasonal forecasts. In some cases (at the Japanese Meteorological Agency, JMA, for instance), a stand-alone forecasting system is used to produce monthly forecasts. Other centres use their seasonal forecasting system to produce sub-seasonal forecast by increasing the frequency of seasonal forecast integrations (e.g. at the Centre for Australian Weather and Climate research, CAWCR) or the number of ensemble members (e.g. at the UK Meteorological Office). ECMWF adopted a different strategy about 10 years ago and decided to make the monthly forecasts an extension of the Ensemble Prediction System since March 2008, thus producing 'seamless' probabilistic forecasts from day 1 to week 4 using the same model.

The goal of the present paper is to review the development and document the skill of the monthly forecasts at ECMWF. Section 2 of this paper will evaluate the current skill of the monthly forecasting system and discuss how it has evolved during the past decade. Section 3 will discuss how various sources of sub-seasonal predictability are represented in IFS. Section 4 will document the skill of the ECMWF monthly forecasts to predict heat waves over Europe. Section 5 will describe planned future changes in the configuration of the monthly forecasts and their expected impact on the forecast skill scores. Section 6 will present an international project, the World Weather Research program (WWRP)-World Climate Research program (WCRP) Sub-seasonal to Seasonal (S2S) prediction project whose main goal is to improve forecast skill and understanding on the sub-seasonal to seasonal timescale, and promote its uptake by operational centres and exploitation by the applications communities.

2 Progress during the past dacade

Monthly forecast (32-day forecasts) have been produced routinely at ECMWF since March 2002, and operationally since October 2004. Since 2002, the configuration of the monthly forecasting system has changed several times (Fig. 1). Following the merging of the monthly and medium-range ensembles in 2008 (Vitart et al. 2008), today the monthly forecasts are generated by extending the 15-day medium-range ensemble integrations to 32 days twice a week (at 00 UTC on Mondays and Thursdays). In other words, seamless forecasts up to 32-days are generated by what is now called the medium-range/monthly ensemble forecast (ENS), one of the key components of ECMWF's Integrated Forecasting System (IFS). ENS includes 51-member run with a horizontal resolution of T639 (about 32 km) up to forecast day 10, and T319 (about 65 km) thereafter. Initial perturbations are generated using a combination of singular vectors and perturbations generated using the ECMWF ensemble of data assimilations (Buizza et al. 2008, Palmer et al. 2007), and model uncertainties are simulated using two stochastic schemes (Palmer et al. 2009). The climatology (re-forecasts) used to calibrate the real-time forecasts is computed using the re-forecast suite that includes only 5 members of 32-day integrations with the same configuration as the real-time forecasts, starting on the same day and month as the real-time forecast over the past 20 years. The re-forecasts are created a couple of weeks before the corresponding real-time forecast. This strategy for re-forecasts (see e.g. Hagedorn et al. 2012 for

a general discussion of the ECMWF re-forecast suite) is different to the one used for seasonal forecasting where the model version is frozen for a few years and the re-forecasts are created only once (Molteni et al. 2011).

The skill of the monthly forecasts is routinely evaluated by scoring the 51-member real time forecasts, mainly against analyses, using a range of measures. For instance, Figure 2 shows skill scores of 2-metre temperature anomalies based on all the real-time forecasts since October 2004. The skill score is the area under the Relative Operating Characteristic (ROC; see Wilks 2011 for a general introduction of probabilistic verification metrics), a measure of the capability of the ensemble to discriminate between occurrence and non-occurrence of events. A ROC score larger than 0.5 indicates that the model is more skilful than climatology. Figure 2 shows a drop of skill with increased time range as expected. For the 12–18 day forecast, the ROC area exceeds 0.7 over large portions of the northern extra-tropics. One week later (i.e. the 19–25 day forecast), the northern extra-tropics still display some skill in predicting 2-metre temperature anomalies, but the highest skills scores are in the tropics. At days 26–32, the skill in the northern extra-tropics is low, although larger than climatology, while skill is largest in the Tropics. A weakness with this type of verification is that it mixes forecasts which have been produced using different versions of the IFS since 2004. This weakness can be addressed by complementing this metric with the approach discussed below.

<table>
<tr><th></th><th>March 2002</th><th>Oct. 2004</th><th>Feb. 2006</th><th>March 2008</th><th>Jan 2010</th><th>Nov. 2011</th><th>June 2012</th><th>Nov. 2013</th></tr>
<tr><td>Frequency</td><td>Every 2 weeks</td><td colspan="5">Once a week</td><td colspan="2">Twice a week (only for real-time)</td></tr>
<tr><td>Horizontal resolution</td><td colspan="3">100km day 0-32</td><td>50km day 0-10
80 km day 10-32</td><td colspan="4">30 km day 0-10
60 km day 10-32</td></tr>
<tr><td>Vertical resolution</td><td colspan="2">40 levels
Top at 10 hPa</td><td colspan="5">62 levels
Top at 5 hPa</td><td>91 levels
Top at 0.01 hPa</td></tr>
<tr><td>Ocean/atmosphere coupling</td><td colspan="3">Every hour from day 0</td><td colspan="4">Every 3 hours from day 10</td><td>Every 3 hours from day 0</td></tr>
<tr><td>Re-forecast period</td><td colspan="3">Past 12 years</td><td colspan="3">Past 18 years</td><td colspan="2">Past 20 years</td></tr>
<tr><td>Re-forecast size</td><td colspan="8">5 members</td></tr>
<tr><td>Initial conditions</td><td colspan="3">ERA 40</td><td colspan="5">ERA Interim</td></tr>
</table>

Figure 1: Evolution of the main changes in the ECMWF monthly forecasts and re-forecasts since 2002.

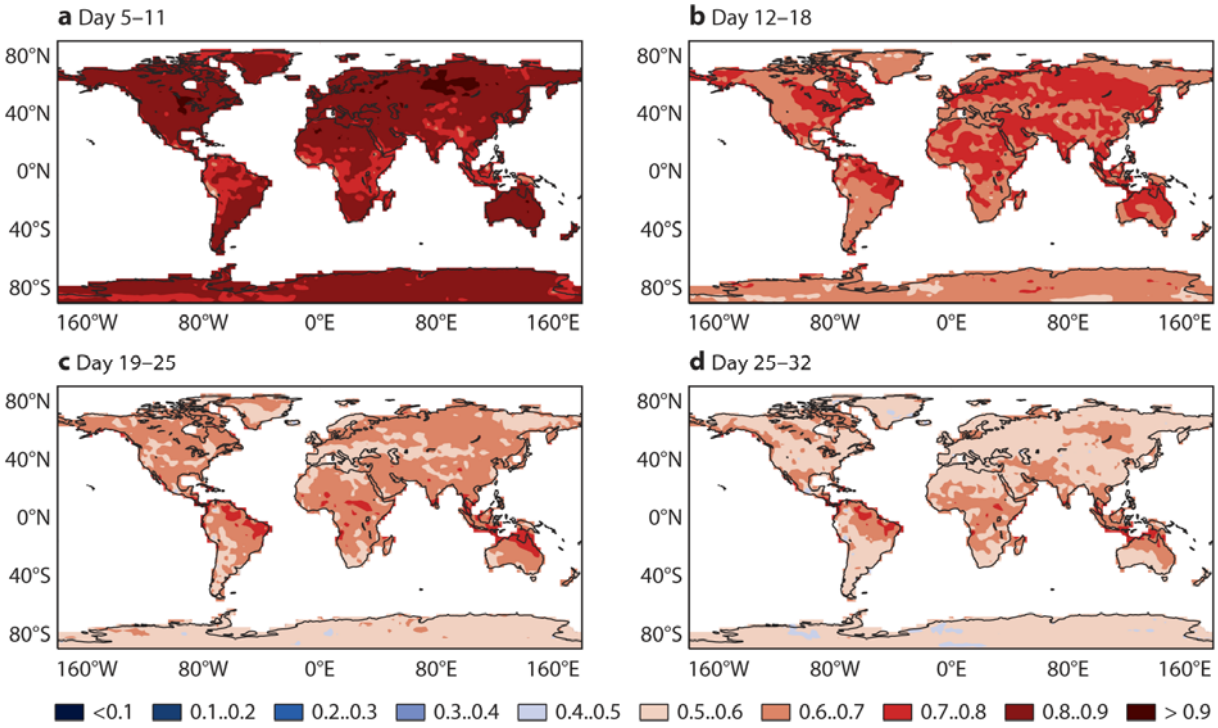

Figure 2: ROC area of the probability of 2-metre temperature anomalies in the upper tercile for the weekly periods day 5-11, day 12-18, day 19-25 and day 26-32. This plot has been produced using all the real-time monthly forecasts since October 2004.

The evolution of the monthly forecast skill scores over the past 10 years could be assessed by comparing the skill scores of the real-time forecasts for each season or each year. However, a major issue with this methodology is that the monthly forecast skill scores are strongly dependant on the large-scale circulation that was predominant during a season. Another option for assessing the evolution of the monthly forecast skill scores that does not suffer from the weakness mentioned above is to use re-forecasts. As shown in Figure 1, the number of re-forecast years has been changing since 2002, but all the re-forecasts since 2002 have the period 1995–2001 in common. The starting days of the re-forecasts may vary by up to 3 days from one year to another, but this does not have a significant impact on the skill scores averaged over a complete year or a season. These scores can be compared for re-forecasts covering the same years and seasons: i.e. all the re-forecasts from 1995 to 2001 that were produced each year between April of a given year until March of the following year. For instance, the scores of 2006 will refer to the scores of all the re-forecasts from 1995 to 2001 that were produced between April 2006 and March 2007 (4 April, 11 April, 18 April.....27 March 1995–2001) using the IFS versions that were operational between April 2006 and March 2007.

Figure 3 displays the evolution of the discrete ranked probability skill score of 2-metre weekly-average winter temperature anomalies since 2002, for three extended-range forecast weekly periods: days 12–18, days 19–25 and days 26–32. The discrete ranked probability skill score is a debiased version of the ranked probability skill score (RPSS) which contains a corrective term

which is a function of the number of categories used to define the probabilities (terciles in the present study) and the ensemble size (5 for the ECMWF re-forecasts) (see for example Müller et al. 2005; Weigel et al. 2007). Forecasts have been compared to analysis over land-points only, and the skill score has been defined by comparing the score of the forecasts and a climatological ensemble with the same membership. Although there is a drop in the probabilistic skill score between days 12–18 and days 19–25, the monthly forecasts still display better skill than climatology (positive RPSS). These results also suggests that there have been improvements in the RPSS scores of 2-metre temperature anomaly re-forecasts over the northern extra-tropics for all three time ranges (days 12–18, days 19–25 and days 26–32) since 2002. The values of the discrete RPSS for days 26–32, although still very low, are now about 50% the values for the previous week (days 19–25) re-forecasts that were produced in 2002. The skill scores of days 19–25 have also improved almost linearly in time and reach about 50% of the skill scores of days 12–18 in the early years of the ECMWF monthly forecasts.

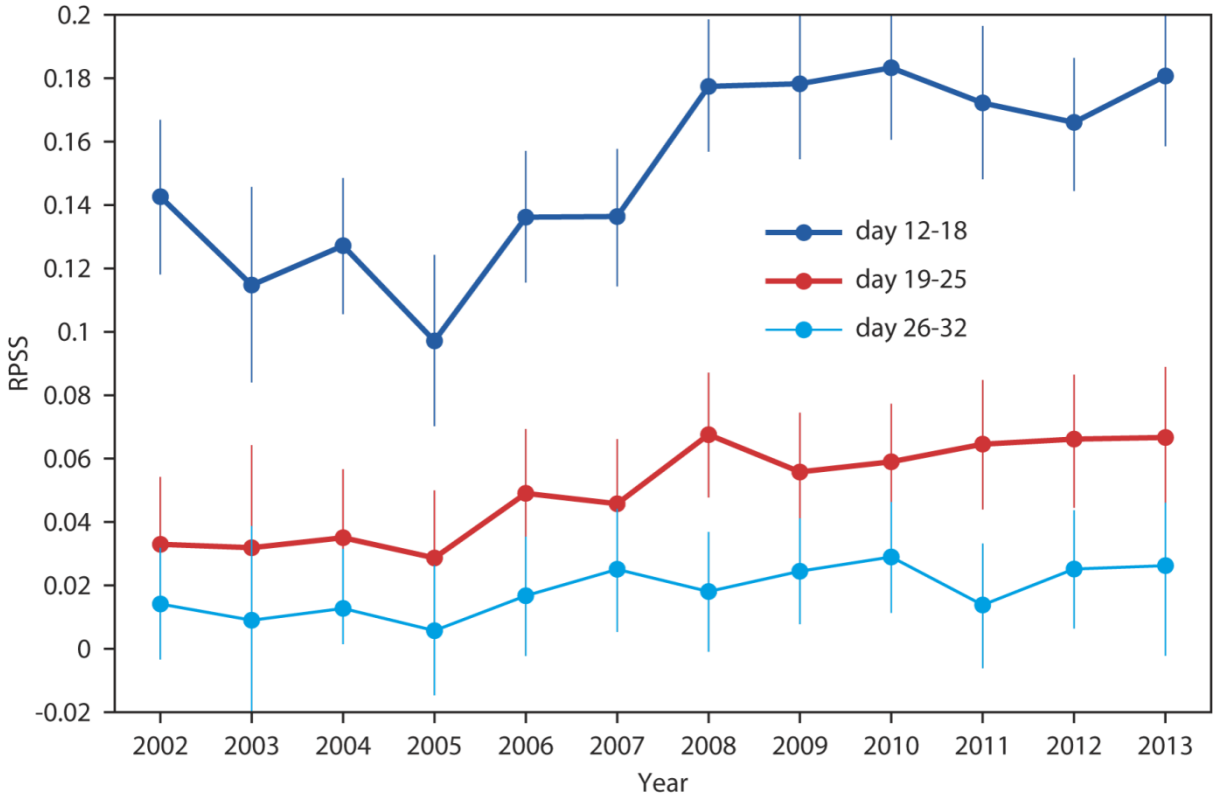

Figure 3: Evolution of the discrete ranked probability skill score (RPSS) of 2-metre temperature weekly mean anomalies over the northern extra-tropics (north of 30°N) since 2002 for days 12–18, days 19–25 and days 26–32. Only land points have been scored. The RPSS has been computed from terciles and for all the ECMWF re-forecasts for the extended boreal winter (October to March).

3 Sources of sub-seasonal predictability and their representation in the model

Several sources of predictability for the sub-seasonal time range have been identified over the past decades. The present section will discuss the links of the extra-tropical circulation and sub-seasonal predictability with (a) the Madden Julian Oscillation and the Tropics, (b) sudden stratospheric warmings and the stratosphere and (c) land conditions, how these links are simulated in the ECMWF ensemble and their impact on the predictive skill of ECMWF monthly forecasts.

3.1 MJO and its teleconnections

The Madden-Julian Oscillation (MJO) is a main source of predictability in the tropics on time scales exceeding one week but less than a season (Madden and Julian, 1971). The Wheeler and Hendon index (WHI, see Wheeler and Hendon, 2004) has been applied to all the model re-forecasts and to ERA-Interim (the ECMWF most recent and most accurate version of the re-analysis of the past decades, Dee et al 2011) over the period 1995–2001 to evaluate the skill of the monthly forecasting system in predicting MJO events and to produce composites for the eight phases of the MJO.

Figure 4 shows the evolution of an MJO skill score from 2002 until 2013 between the ensemble mean re-forecasts and ERA-Interim. In this figure, the line shows the forecast day in which the bivariate correlation reached 0.6. If we consider the MJO bivariate correlation of 0.6 as a limit of MJO prediction skill, the ECMWF monthly forecasting system displayed skill to predict the MJO up to about 15 days in 2002. In 2013, the limit of 0.6 was reached around day 27, suggesting an averaged gain of about 1 day of lead-time per year. The difference of MJO skill scores between 2002 and 2013 is statistically significant within the 5% level of confidence. This skill should continue to improve with the forthcoming implementation of model cycle 40R3, planned for the end of 2014 (Fig. 5), thanks to the implementation of the revised organised convective detrainment term and the revised convective momentum transport. The evolution of the amplitude error of the MJO, calculated from each individual ensemble member and each individual forecast and then averaged, does not display an improvement as regular as for the forecast skill scores (see Fig4a in Vitart 2014). According to Vitart (2014), the ensemble members of the ECMWF monthly forecasts produced a too weak MJO in the early years of the monthly forecasting system, with the amplitude about 30% too low beyond forecast day 20. There has been a clear improvement between 2006 and 2008 linked to important changes to the model physics:

- The introduction of a parameterisation of ice supersaturation (Tompkins et al. 2005, Vitart et al. 2007) in 2006

- The introduction of a new radiation parameterisation (McRad) (Morcrette et al, 2007) in 2007
- The introduction of a new convective scheme in IFS cycle 32R3 in 2008 (Bechtold et al. 2008).

In 2008, when Cy32r3 was used operationally, the MJO was even slightly too strong. Since 2008, the amplitude of the MJO displays a trend towards weaker MJOs, with amplitudes in the recent years only about 10% weaker than in the ERA-Interim analyses.

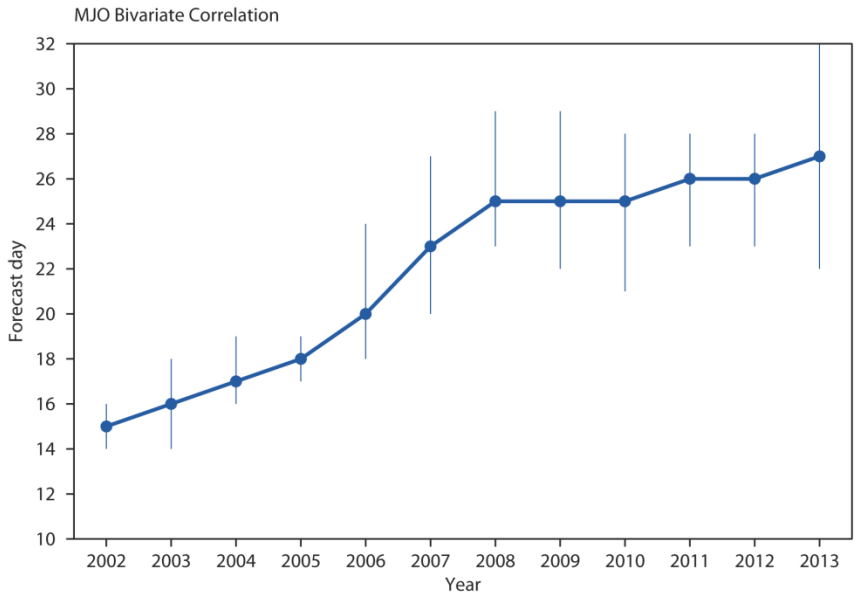

Figure 4: Evolution of the MJO skill scores (bivariate correlations applied to WHI) since 2002 as indicated by the days when the MJO bivariate correlation reaches 0.6. The MJO skill scores have been computed on the ensemble mean of the ECMWF re-forecasts produced during a complete year. The vertical bars represent the 95% confidence interval computed using a 10,000 bootstrap re-sampling procedure.

3.1.1 MJO impact in the Extratropics

Using reanalysis data, Cassou (2008) showed that there is a link between the MJO and North Atlantic Oscillation (NAO) during DJF. The probability of a positive phase of the NAO (i.e. difference of atmospheric pressure at sea level between the Icelandic low and the Azores high) is significantly increased about 10 days after the MJO is in Phase 3 (Phase 3 + 10 days), and significantly decreased about 10 days after the MJO is in Phase 6 (Phase 6 + 10 days). The probability of a negative phase of the NAO is decreased (increased) about 10 days after the MJO is in Phase 3 (Phase 6). According to Figure 6, the MJO teleconnections (10 days after an MJO in Phase 3) are more realistic over the northern extra-tropics in 2013 (middle panel) than in 2002 (left panel) compared to ERA-Interim (right panel). The re-forecasts produced in 2013 simulate a stronger positive NAO anomaly than in 2002. However, the impact of the MJO on the NAO is still

underestimated in the 2013 re-forecasts compared to ERA-Interim. On the other hand, the ECMWF forecasting system overestimates the positive 500 hPa geopotential anomaly over the northern Pacific. The same conclusions are valid for the composites of 500 hPa geopotential height 10 days after an MJO in Phase 6 (not shown). These results indicate that any future improved MJO teleconnections are likely to impact the monthly forecast skill scores in the northern Extratropics, and in particular the skill of the model to predict the NAO.

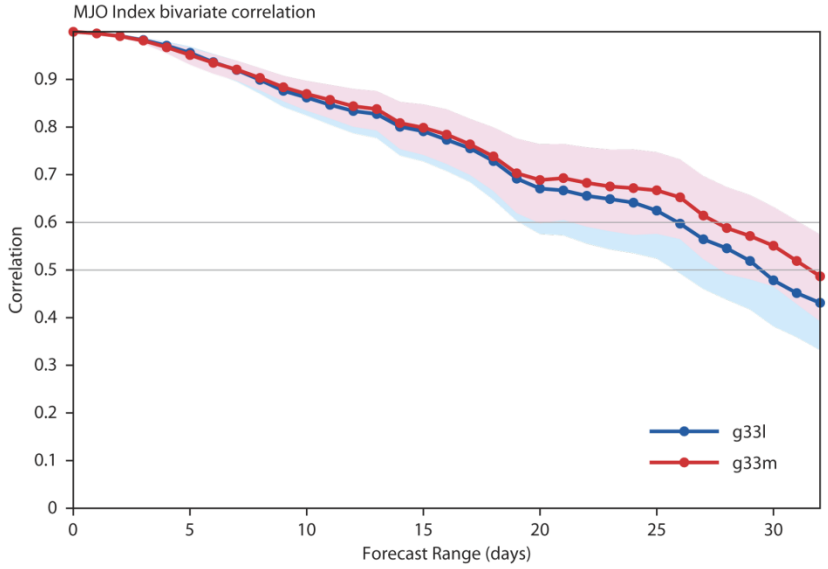

Figure 5: MJO bivariate correlation as a function of lead time computed from a set of re-forecasts using cycle 40r1 (blue curve) and cycle 40r3 (red curve). The shaded areas represent the 95% level of confidence using a 10,000 bootstrap re-sampling procedure. The re-forecasts start on 1st February/May/August and November 1989-2008.

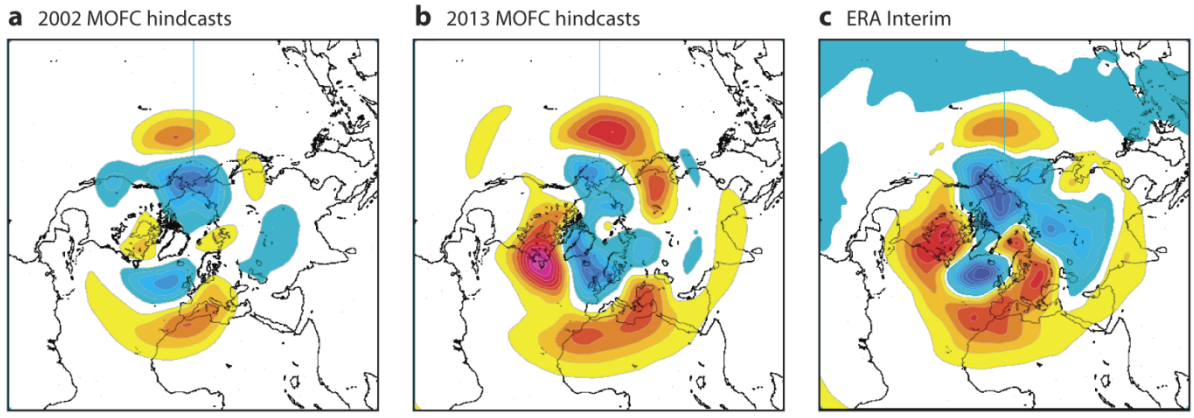

Figure 6: MJO Phase 3 10-day lagged composites of 500 hPa geopotential height anomaly over the northern extra-tropics for all the October to April re-forecasts that were produced in (a) 2002, (b) 2013 and (c) ERA-Interim. Red and orange colours indicate positive anomalies. Blue colours indicate negative anomalies. The lowest contour is at 10 metres and the contour interval is 5 metres.

An NAO index has been constructed by projecting the daily 500 hPa height anomalies over the northern hemisphere onto a pre-defined NAO pattern based on an EOF (Empirical Orthogonal Function) analysis. The NAO pattern was defined as the first leading mode of EOF applied to the reanalysis of monthly mean 500 hPa height during the 1950–2000 period produced by NCEP (National Centers for Environmental Prediction). NAO skill scores have been produced for each year from 2002 until 2013 by applying the NAO index to the re-forecasts and to ERA-Interim, and by computing the linear correlation between the ensemble-mean re-forecasts and ERA-Interim.

Figure 7 shows that there has been improvement in the prediction of the daily values of the NAO with a gain of about 4 days of lead time for a correlation of 0.5, 3 days for a correlation of 0.6 and 2 days for a correlation of 0.8. As for the MJO, the improvement in the prediction of the NAO cannot be attributed to a single change of the ECMWF forecasting system. The difference of NAO skill scores between 2002 and 2013 are statistically significant with the 95% level of confidence. Vitart (2014) showed that a large portion of the improvements in the NAO skill scores can be attributed to the improvements in the prediction of the MJO.

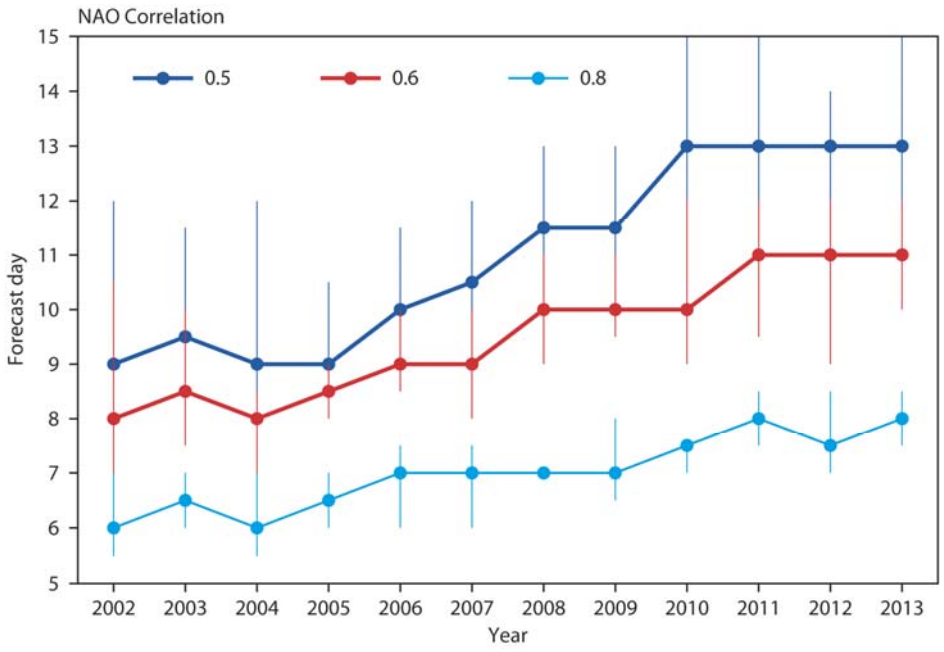

Figure 7: Evolution of daily NAO skill scores since 2002 as indicated by the days when the NAO index correlation reaches 0.5, 0.6 and 0.8. The daily NAO skill scores (correlations applied to the NAO index) have been computed on the ensemble mean of the ECMWF re-forecasts produced from October to March 1995–2001 and ERA-Interim. The vertical bars represent the 95% confidence interval computed using a 10,000 bootstrap re-sampling procedure.

3.1.2 An example of MJO impact on the Extratropics circulation

March 2013 was exceptionally cold over most of Europe and was the second coldest March in the UK since 1910. The cold temperature anomalies also extended over large portions of North

America and Asia. This cold event lasted for about a month, and was associated with a negative phase of the NAO. Several drivers could have contributed to this cold event: weather in the Tropics, the stratosphere, conditions in the North Atlantic and the state of the Arctic. These drivers are not necessarily independent, and the cold March 2013 could be due to a combination of interlinked causal factors. The cold event March 2013 coincided with the propagation of a MJO event into the western tropical Pacific. Studies (e.g. Cassou 2008, Lin et al, 2008) have shown that an active phase of the MJO over the western Pacific is conducive to negative NAO. Was this MJO event responsible for the cold event over Europe?

To establish if there was a link between this MJO event and the cold March 2013, a composite of 2-metre temperature anomalies of the 10 ensemble member forecasts which predicted the strongest MJO event was compared to a composite of 2-metre temperature anomalies of the 10 ensemble member forecasts which predicted the weakest MJO for each individual forecast system (as for 2-metre temperature, MJO forecasts varied greatly between ensemble members). Results from the ECMWF model (Fig. 8) and the NCEP model (not shown) suggest that there was a strong link between the MJO event and the cold anomaly over Europe, with the ensemble members with a strong MJO predicting 2-metre temperature anomaly patterns more consistent with observations than the ensemble members with a weak MJO. The ensemble forecasts from Environment Canada showed the same link between good MJO forecasts and good 2-metre temperature forecasts (not shown).

This link between MJO and 2-metre temperature anomaly forecasts could be due to an impact of the MJO on extratropical weather or to an impact of extratropical weather on the MJO. To assess the causality, further sensitivity experiments were conducted with the ECMWF model where the weather over the tropical band (20S-20N) was relaxed towards analysis. In this experiment, the MJO evolution in tropical latitudes, is therefore, almost perfect. Results (Fig.9) show that the relaxation experiment produces 2-metre temperature anomalies over the Northern Extratropics, most especially over Europe, which are much closer to analysis than the model integrations without relaxation. The difference in 2-metre temperature anomalies between the relaxation and the control experiments is consistent with a canonical MJO response in the Extratropics. Therefore, these experiments confirmed the role of tropical convection on the cold March 2013 over Europe and part of North America.

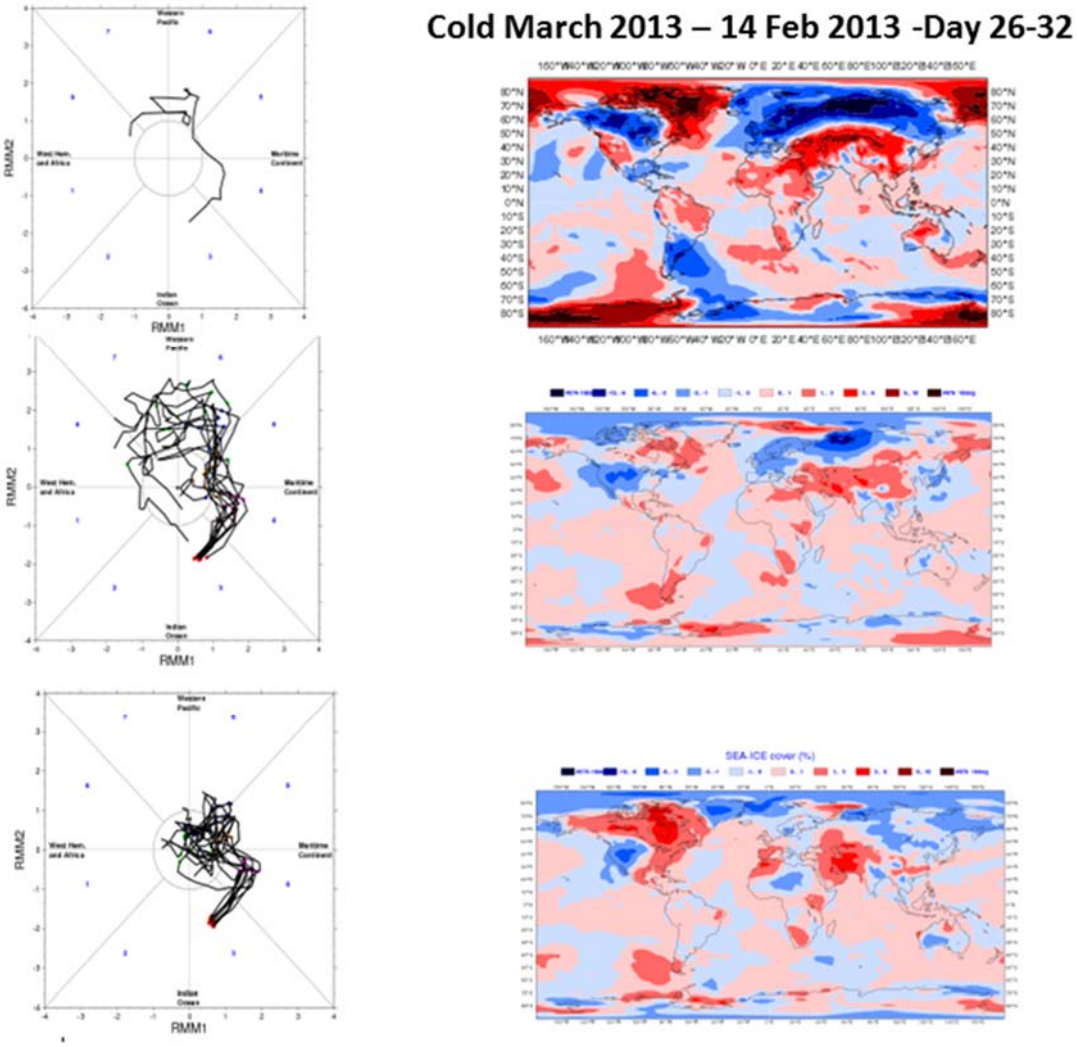

Figure 8: The left panels show the MJO propagation in the Wheeler-Hendon during the 32-day integrations starting on 14 February 2013 for the 10 ensemble members which predicted a strong MJO event propagating in the western pacific (middle panel) and for the 10 ensemble members which predicted the weakest MJO propagation(bottom panel). The top left panel shows the verification using ERA Interim. The right panel show the associated 2-metre temperature anomalies diagrams averaged over the period 11-17 March 2013, from the analysis (top panel), the 10 ensemble members which predicted the strongest MJO (middle panel) and the 10 ensemble members predicting the weakest MJO event (bottom panel) from the ECMWF forecasts starting on 14 February 2013 (time range day 26-32).

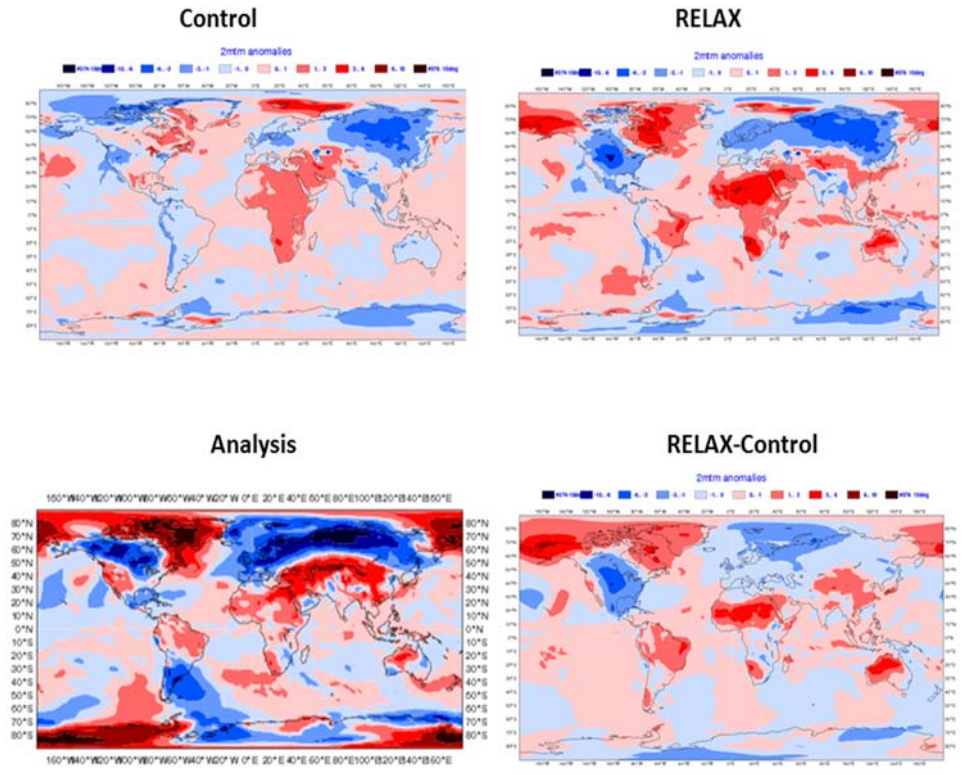

Figure 9: 2-metre temperature anomalies averaged over the period 11-17 March 2013 from the control experiment starting on 14 February 2013 (top left), the experiment with the tropics relaxed towards analysis (top right), analysis (bottom left). The bottom right panel shows the difference of 2-metre temperature anomalies between the relaxation and the control experiments.

3.1.3 Impact of the MJO on tropical cyclone activity

The impact of the MJO on tropical cyclone activity has been documented in numerous observational studies (e.g. Nakazawa 1988, Maloney and Hartmann 2000). The model tropical cyclones are tracked using the methodology described in Vitart et al. (1997). As in observations, model tropical storms display a maximum 10-metre wind velocity exceeding 17 m/s. This tracker has been applied to all the model re-forecasts. According to Vitart (2014), IFS reproduces the observed modulation and eastward propagation of tropical cyclone activity when the MJO propagates from the Indian Ocean (phase 2 or 3) to the western Hemisphere (Phases 8 or 1). The model also reproduces accuratetly the modulation of tropical cyclone activity over the Northern Hemisphere during the period June to November (Vitart 2009). Based on the ability of IFS to simulate the impact of the MJO on tropical cyclone activity and the skill of the model to predict MJO events, sub-seasonal forecasts of tropical cyclone activity over weekly periods are issued routinely from the monthly forecasts. Vitart et al. (2010) showed that these sub-seasonal forecasts of tropical cyclone activities are reliable up to week 4 over some basins, and outperform statistical models.

3.2 Sudden stratospheric warmings and their impact in the lower troposphere

Sudden stratospheric warmings (SSWs), where the polar vortex of westerly winds in the winter hemisphere abruptly (i.e. over the course of a few days) slows down or even reverses direction, accompanied by a rise of stratospheric temperature by several tens of Kelvins, are considered another potential source of predictability at the sub-seasonal time scale. Baldwin and Dunkerton (2001) showed strong apparent downward propagation of easterly and westerly anomalies from the stratosphere to the troposphere on monthly timescales. Importantly, this tends to be followed by easterly (negative NAO/AO) conditions in the troposphere. Perturbation experiments also reproduce negative NAO in response to weakened stratospheric winds on both sub-seasonal and longer timescales (for example Boville 1984, Norton 2003, Scaife et al. 2005).

The difference of 50hPa temperature between 90N and 30N averaged over all the longitudes is used as an index for SSWs (similar results were obtained when using a different SSW index based on the zonal wind at 10 hPa and 60N). Fields at 50 hPa are archived only since October 2004. Therefore this section will consider only the re-forecasts that were produced after 2004. Figure 10 shows that the correlation of 0.6 is reached around day 24 in 2013. A noticeable improvements in forecasting skill since 2004 occurred in 2006 when the vertical resolution of the ECMWF monthly forecasting system increased from 40 to 62 vertical levels, with a top level at about 5 hPa instead of 10 hPa before 2006. The SSW skill scores have significantly further improved in 2013 (period from October 2013 to March 2014) when the vertical resolution of the monthly forecasts was increased again from 62 to 91 vertical levels with a top level at 0.01 hPa. This confirms that vertical resolution, particularly in the stratosphere (the 91 and 62 vertical resolutions were identical in the troposphere) had a positive impact on the skill of IFS to predict SSWs.

For successful monthly forecasts, it is not only important for the forecasting system to display skill in predicting SSWs, it is also important to simulate the impact of SSWs on the tropospheric weather, most especially its impact on the NAO. Figure 11 shows the lag correlation between the SSW and NAO indices. This figure indicates that the lag correlation increases in absolute terms in the days following a SSW in ERA Interim. Although the absolute value of the lag correlation is not very high (0.25), this indicates that the probability of a negative NAO increases in the days following a SSW, consistent with previous studies (e.g Boville 1984). However, Figure 11 shows that the amplitude of the lag correlation diminishes in absolute terms instead of increasing in the days following a SSW in the ECMWF monthly forecasts, suggesting that the ECMWF forecasting system under-represents this impact of the stratosphere on the troposphere. Case studies, like the stratospheric warming of February 2012 which may have led to cold weather over Europe, also suggest that the impact of the stratosphere on the troposphere is too weak in the current version of IFS (not shown). All the re-forecasts produced since 2002 display a similar behaviour. Figure 12 shows an example of the zonal wind anomalies at 30N averaged over Europe during the January 2013 SSW event. According to this figure the SSW event which started at the beginning of

January 2013 was well predicted 15 days in advance, but zonal wind anomalies below 50 hPa were strongly underestimated in the model compared to reanalysis. This figure confirms that IFS has some issues in producing the downward propagation associated to a SSW. This is likely to affect the monthly forecast skill scores over Europe when a SSW takes place (about once a year during the winter season).

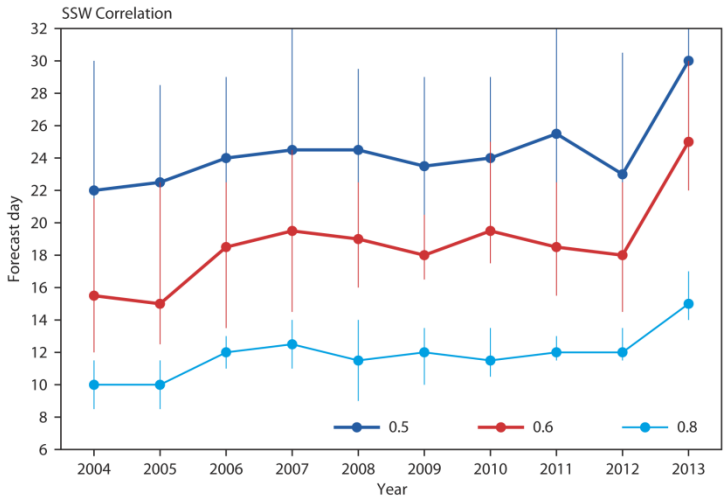

Figure 10: Evolution of the SSW skill scores since 2004. The daily SSW skill scores (correlations applied to a SSW index which is based on the difference of temperature at 50 hPa between the North pole and 30N average over all the longitudes) have been computed on the ensemble mean of the ECMWF re-forecasts from October to March and ERA Interim. The blue, red and brown lines indicate the day when the SSW index correlation reaches respectively 0.5, 0,6 and 0.8. The vertical bars represent the 95% confidence interval computed using a 10,000 bootstrap re-sampling procedure.

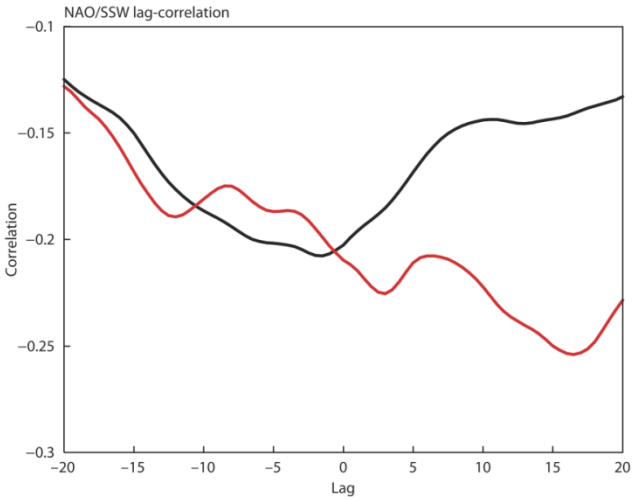

Figure 11: Lag correlation between the NAO and SSW index as a function of days preceding (negative x-axis) or following (positive x-axis) a SSW. The black line shows the lag correlation obtained from all the re-forecasts produced between October 2011 and March 2012 and covering the years 1995 to 2001 and for each ensemble member separately. The red line shows the corresponding verification using ERA Interim.

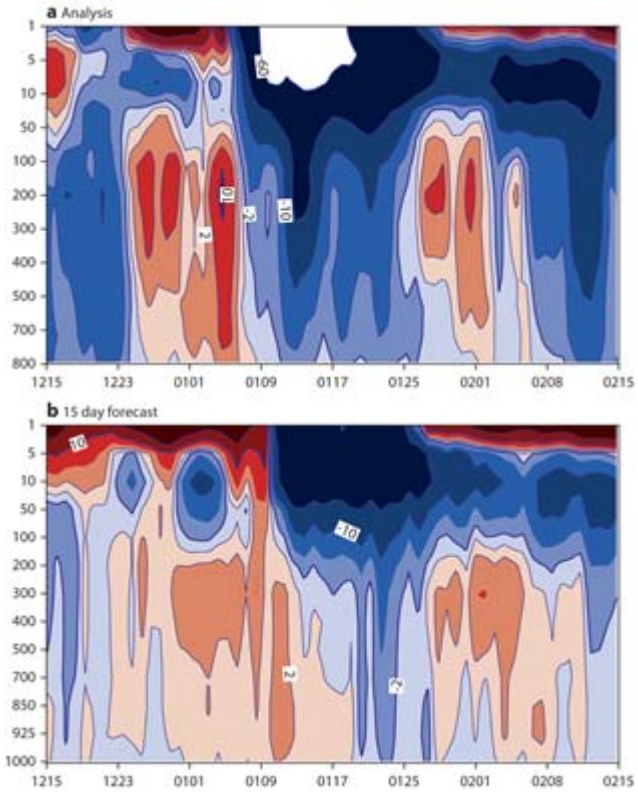

Figure 12: Evolution of zonal wind anomalies at 60N averaged over Europe from 15 December 2012 to 15 February 2013 (x axis) as a function of the pressure level (vertical axis). The reforecasts were produced once a day with a T639 resolution and 91 vertical levels, with 15 members and IFS cycle 38R2.

3.3 Land initial conditions

This section analyses the impact of soil initial conditions on sub-seasonal forecasts. Section 3.3.1 will discuss the impact of soil moisture, whereas Section 3.3.2 will evaluate the sub-seasonal predictability associated with snow cover.

3.3.1 Predictability associated to soil moisture: GLACE-2 project

The GLACE (Global Land Atmosphere Coupling Experiment) and GLACE2 projects have engaged several research centres to conduct experiments with different Earth System Models sharing a common protocol and equal boundary conditions. The configuration of the IFS used in GLACE experiments is documented in Vitart et al. (2008) and it forms the base of the ECMWF monthly forecasting system connected to the ensemble prediction system (Molteni et al. 2011). The first GLACE initiative, to which ECMWF did not participate, led to important scientific discussions on

the role of land surface in the long-term forecast. The main results of this project involved studies of the predictability of the temperature at 2 meters (T2m) and precipitation (P) expressed as anomalies relative to the forecast average monthly climate (Koster et al. 2004). A limitation of that study was the lack of quantification of the degree of predictability and the fact that the period was too short to allow statistically significant conclusions to be drawn.

The GLACE2 project was designed to quantify the gain in predictability for anomalies of 2-meter temperature and precipitation, when accurate initial conditions are available to describe the land. The GLACE2 experimental protocol consists of running a first series of model integrations where the land has been properly initialized (Series 1) and a second series of integrations (Series 2) where the land initial conditions have been randomized.

Three different studies have focused respectively on the United States, Europe and the global domain as documented in Koster et al. (2009), van den Hurk et al. (2012) and Koster et al. (2010), respectively. The three studies show the added value of a realistic initialization of the soil moisture, measured in terms of predictability gain of 2m temperature and precipitation over the US and Europe (see van den Hurk et al. 2012) with forecasts lead-time ranging from 16 to 60 days ahead. Continental US generally has a higher potential predictability and this is also mirrored in the actual prediction results. It is remarkable that for temperature at 2m over Europe the predictability gain obtained from better soil moisture initial conditions, extends significantly beyond the ranges typically considered in NWP (and up to 16-30 days ahead). Results by Weisheimer et al. (2011) also indicate that forecasts are capable of detecting extreme events such as the summer of 2003.

3.3.2 Predictability associated with snow: SNOWGLACE project

As demonstrated for soil moisture, one should expect that strong anomalies in other slowly evolving surface reservoirs play a role in enhancing predictability. This is due to the "memory" effect that continental surfaces can play in weather forecasting for medium and long range. For example, the snow covers a fairly large fraction of the northern hemisphere land in winter. The representation of snow is very important to be able to describe the degree of decoupling between the surface and the atmosphere. In winter, the freezing land delays the thermodynamic process of cooling at the surface (by latent heating). When snow deposits and accumulates on the ground, its high insulating power inhibits the soil freezing process (the insulated soil typically remains at 0°C underneath the snow pack, (see Beljaars et al. 2007) and allows the air to cool down faster. In fact, the nocturnal radiative cooling in the presence of snow is much more efficient because it applies almost exclusively to cool the snow layer. In the presence of a large accumulation of snow, its density is a variable with a slow evolution, which depends on the history of the snowpack, and plays a role in the seasonal prediction (via memory effects, see Dutra et al. 2010).

It is very plausible to assume that snow plays a key role in global weather and climate at all forecasting range. Several studies (Cohen and Entekhabi, 1999; Orsolini and Kvamstø, 2009; Peings et al. 2011) have shown such sensitivity, but quantifying the predictability associated with snow remains a challenge for the scientific community. The success of the GLACE and GLACE2

projects led to the SNOWGLACE exploratory project conducted in collaboration with the NILU (Norwegian Institute for Air Research) laboratory and inspired by the GLACE2 project (as explained in Orsolini et al. 2013). Pairs of 2-month ensemble ECMWF forecasts were started every 15 days from the 15th of October through the 1st of December in the years 2004–2009, with either realistic initialization of snow variables based on re-analyses (series 1), or else with ‘‘scrambled’’ snow initial conditions from an alternate autumn date and year (series 2).

Initially, in the first 15 days, the presence of a thicker snowpack cools surface temperature over the continental land of Eurasia and North America. At a longer lead of 30-day, it causes a warming over the Arctic and the high latitudes of Eurasia due to an intensification and westward expansion of the Siberian High. It also causes a cooling over the mid-latitudes of Eurasia, and lowers sea level pressures over the Arctic. The impact of realistic snow initialization upon the forecast skill in snow depth and near-surface temperature is estimated for various lead times. Following a modest skill improvement in the first 15 days over snow-covered land, a forecast skill improvement up to the 30-day lead time over parts of the Arctic and the Northern Pacific was also found, which can be attributed to the realistic snow initialization over the land masses (Fig. 13).

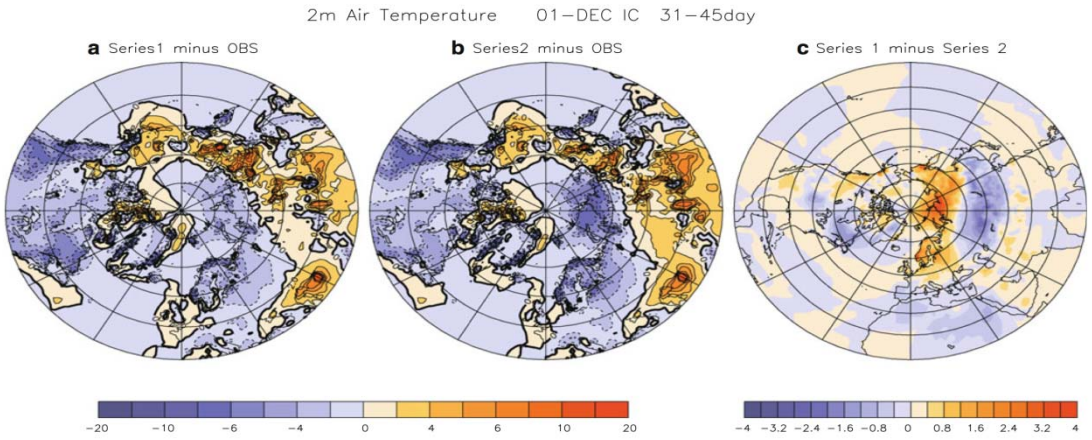

Figure 13: Near-surface temperature biases and differences at 30-day lead. T2m differences between the ensemble-mean of Series 1 and ERA-Interim (a), between the ensemble-mean of Series 2 and ERA-Interim (b), and between ensemble-means of Series 1 and Series 2 (c). They are shown for the start date of DEC 1 and valid for a 15-day sub-period.

3.4 Impact of stochastic physics on model biases and reliability

As mentioned in Section 2, the operational ensemble forecasts use two stochastic physics schemes (the Stochastically Perturbed Physical Tendency scheme and the Stochastic Kinetic Energy Backscatter scheme) to simulate model uncertainties (Palmer et al 2009). A previous study on the impact of these schemes on the coupled ECMWF seasonal forecasting system reports a significant reduction of tropical model biases (Weisheimer et al., 2014). To test the impact of stochastic perturbations on systematic errors and forecast performance on the monthly time scale, monthly hindcasts with CY40R1 have been integrated with a T319(T255)L91 resolution for the period

1989-2008 using 1st of Feb, May, August and November start dates and 15 ensemble members with and without representations of model uncertainties.

Figure 14 demonstrates the impact of stochastic perturbations on the mean precipitation state for the August start dates. The positive rainfall bias near the equator reaching from the Eastern Pacific through the tropical Atlantic and Africa is reduced and the too dry conditions in the Western Pacific is reduced when stochastic physics is activated.

The statistics of tropical storms is significantly improved due to stochastic parameterisations (not shown). The number of tropical storms increases, most likely thanks to improved large-scale variability, over most ocean basins leading to a reduced underestimation of the number compared to observations and ERA-I. As an example of the impact of stochastic physics on the forecast performance we analysed the reliability of tercile rainfall events over the tropics. Stochastic processes lead to a significant improvement in the reliability of these events during the first two weeks of the forecasts (d4-d11 and d11-d18), see Fig 15. The impact on ensemble spread is also being investigated. Results from the bivariate MJO index analysis suggest a substantial increase in ensemble spread during the entire forecast range due to stochastic parameterisations in the atmosphere.

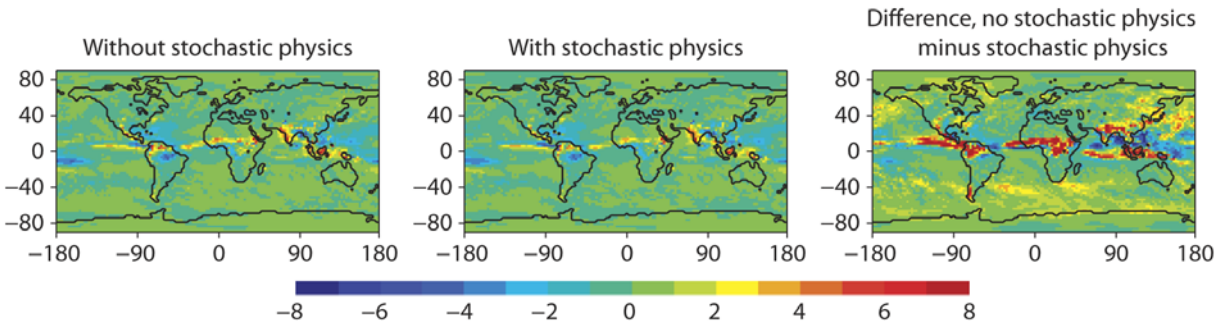

Figure 14: Precipitation biases in the monthly forecasts in August averaged over the month without stochastic physics (left), with stochastic physics (middle) and the difference no stochastic physics minus stochastic physics.

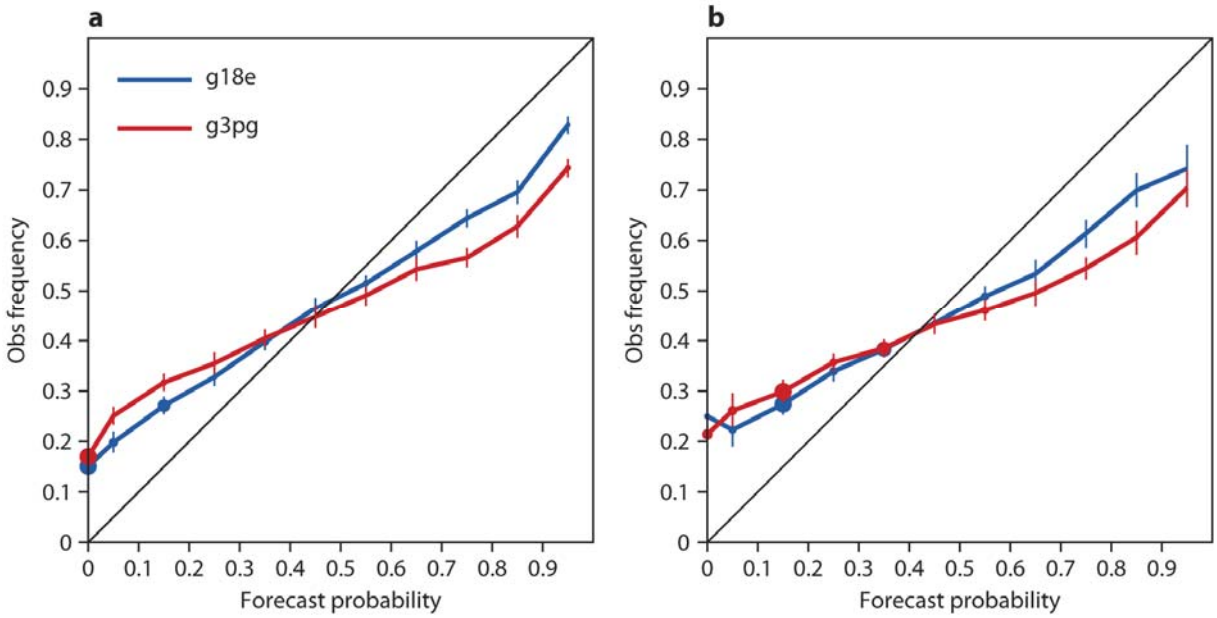

Figure 15: Reliability diagrams for forecasts of upper tercile precipitation events in the Tropics during week1 (left) and week2 (right) without stochastic physics (red) and with stochastic physics (blue). The size of the solid circles is proportional to the number of cases populating each individual bin.

4 Heat Wave prediction

An important application of the sub-seasonal forecasts is the prediction of extreme events. This section will evaluate the prediction of an example of extreme events: the Heat Waves (HW) which have high societal impact. The prediction of such extreme events a few weeks in advance would be particularly useful for society. Since heat waves can last more than a week they could be the type of weather events that the sub-seasonal prediction systems can forecast. Several studies have shown that HWs over Europe during the warm season are related to persistent and large-scale high pressure systems (Della Marta et. al 2007). Those systems are sometime associated with global teleconnections linked to tropical organized convection (MJO) (Cassou 2005). Because of their low-frequency nature and their teleconnections they can exhibit predictability on the sub-seasonal time scale. A further source of predictability arises from the effect of soil moisture conditions in the amplification of the temperature anomalies. The soil conditions can play a role only in combination with persistent anti-cyclonic summer weather regimes (Quesada et al. 2012). Therefore accurate skill in predicting persistent large-scale high pressure systems is instrumental to forecast HW.

Four types of HW events have been identified during the period May to September 2005-2012, when real-time ECMWF extended range forecasts were available: 3 HW events over Eastern Europe (EE), 2 over Western-Central Europe (WE), 3 over the Northern Sea (NS) and 3 over Russia (RU). Two-metre temperature composites of ERA-Interim weekly means anomalies for the

4 types of HW events are shown in Figure 16. For the HW detection an objective criterium that identifies coherent and persistent regions with temperatures exceeding the upper 90th centile of the local climatological distribution has been used. The weekly mean anomalies are computed with respect to a weekly mean climate of the most recent 18 years consistent with the weekly mean anomalies of the forecast. The EE events exhibit maximum temperature over Poland and Romania (Fig.16a). The composite for the RU events (Fig.16b) includes the 2010 case with max temperature over Northern Russia. The NS composite (Fig.16c) show very warm condition over Scandinavia and UK. The WE composite anomaly (Fig.16d) is reminiscent of the 2003 HW event although this is not included in the analysis. The corresponding composites for the geopotential height at 500 hPa present an anticyclonic anomaly in phase with the maximum temperature (not shown).

The ideal method to evaluate the skill of the ECMWF extended range ensemble in predicting HW is to use a selection of objective verification measures for probabilistic forecasts. In reality verification requires a far larger sample than what is available. This is typically the case for any investigation that involves extreme events. Hence the cases were investigated individually and we have used the composites of ensemble mean anomalies to summarize the results. The 2m temperature composites, based on weekly mean anomalies of ensembles forecasts at 12-18 days, are shown in Figure 17. Generally the forecasts identify with a certain degree of accurracy the location of warm anomalies although the amplitude is underestimated.

Two out of the three EE events were well forecasted. In the case of a poor EE forecast, the high pressure over Scandinavia present during the week leading to the HW onset was wrongly persisted and consequently the warm anomalies were not predicted over the right location. Both WE events were reproduced althought the onset of one event was forecasted with one week of delay. Both events persisted for about 3 weeks and during this time the verifying anomalies were included in the range of predicted temperatures. Forecasts gave accurate predictions for the 2 out of the 3 HW events over the Northern sea which were all associated with a Scandinavian blocking. The poor forecast, by persisting the anomalous circulation of the previous week, missed the high pressure anomalies over Scandinavia. The most persistent HW events were detected in Russia. Two out of 3 Russian heat waves were forecast at 12-18 days although the onset of the event was captured with about one week of delay in both cases.

Figure 18 shows timeseries of the ensemble forecast distribution at day 12-18 (blue box and wiskers) for the weekly mean anomalies that verified the week starting from 21 June to 22 Septembre 2010. The values are averaged over an area between 30-50E and 70-50N. The verifying anomalies are represented by red dots and the ERA-Interim climate distribution is represented by the yellow box and wiskers. Figure 18 shows the extent of the 2010 HW event over Russia and the fact that most of the time the analysis is included in the predicted range. It is also clear that the start of the warm conditions is reproduced with a delay by the forecast at 12-18 days.

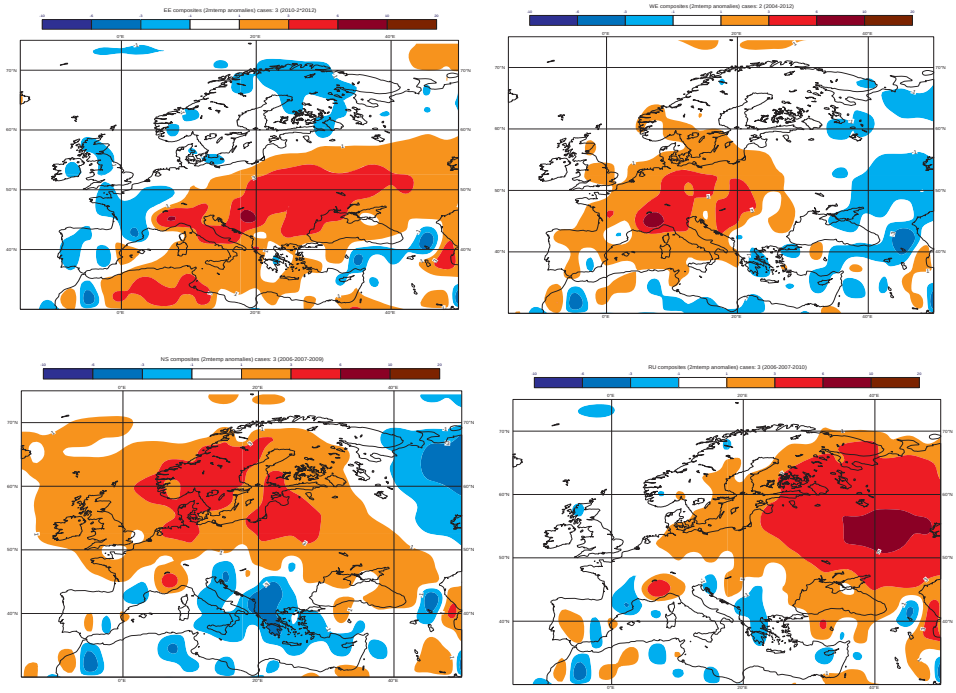

Figure 16: 2m temperature composites from era-interim weekly means anomalies: top left) Eastern Europe, top right) Western Europe, bottom left) North Sea and bottom right) Russia HW event.

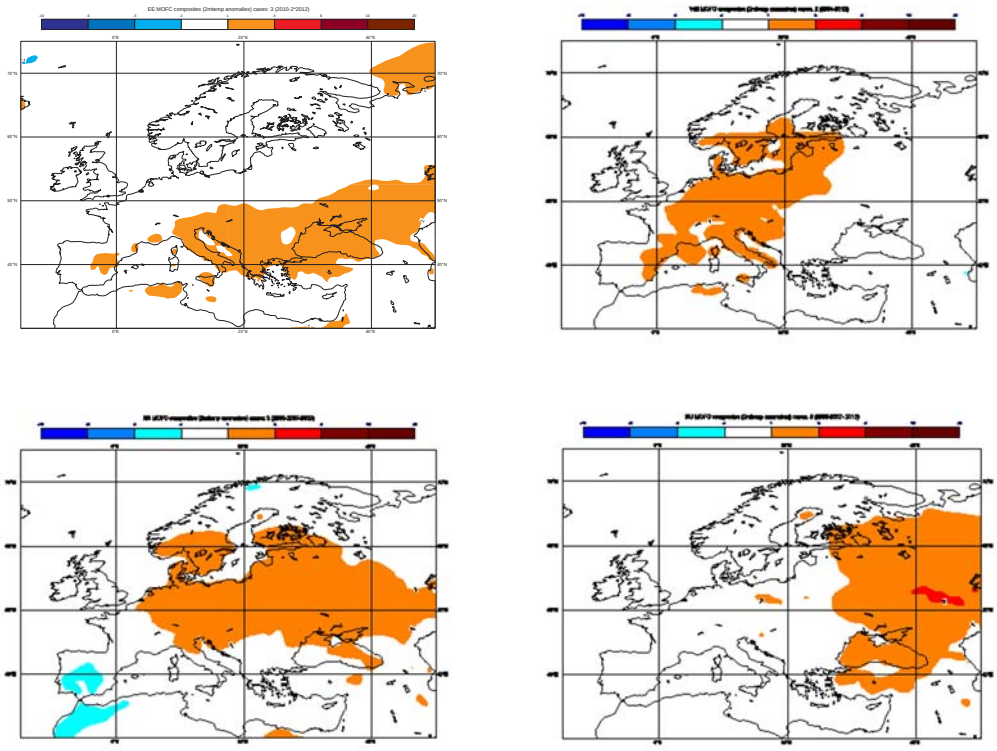

Figure 17: 2 m temperature composites from the ensembles forecast at 12-18 days verifying on the same dates as in Figure 1.

It is difficult to draw any firm conclusions from the limited sample available. Generally the successful predictions persisted the anti-cyclonic circulation which was already present in the initial conditions. In contrast most of the non-successful predictions did not have an anti-cyclonic circulation in the initial conditions. The skill in predicting heat waves at the extended range may therefore be limited by the ability of the forecast model to represent transitions to anti-cyclonic circulation regimes, which is consistent with the cause of medium-range forecast busts identified in Rodwell et al. (2012). Experiments are also planned to evaluate the role of tropical forcing and initial soil moisture conditions in each event.

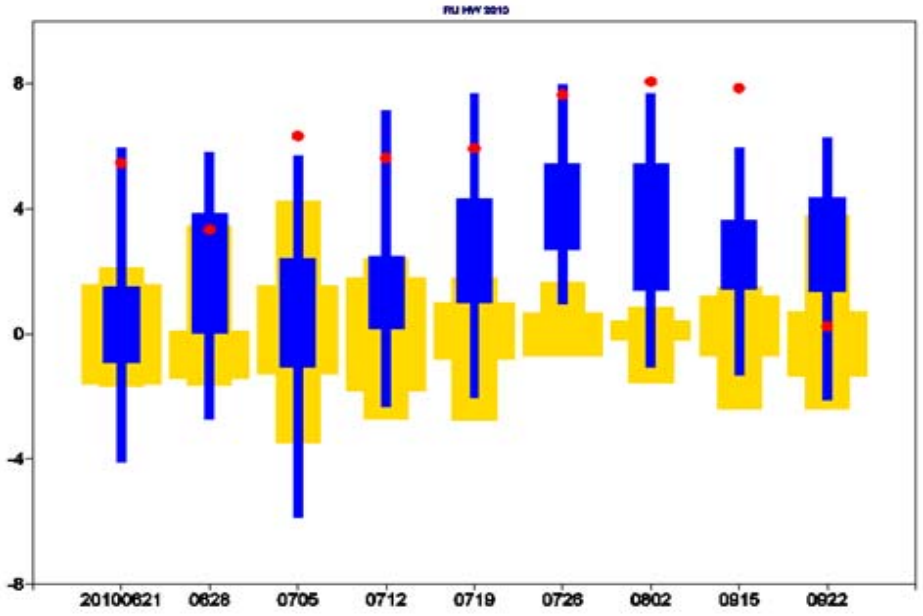

Figure 18: Evolution of 2m temperature weekly mean anomalies from June to September 2010. The values are averaged over an area between 30-50E and 70-50N. The verifying anomalies are represented by red dots and the era-interim climate distribution is represented by the yellow box and wiskers.

5 Future model changes

In this section we discuss three major changes in the monthly configuration, planned to be implemented in the next couple of years to further improve the skill of monthly forecasts and provide users with valuable forecasts in the sub-seasonal forecast range: (a) the extension of the re-forecast suite membership, (b) the increase in the horizontal resolution and (c) the inclusion of a sea-ice model and d) the extension of the forecast length to 45-60 days.

5.1 Extension of the re-forecast membership list

In the current configuration, the re-forecast dataset consists of 5 ensemble members integrated on the same day and same month as the Thursday monthly forecasts over the past 20 years (20 start dates) (see Fig. 1). This represents a total of 100 ensemble members. The Thursday monthly forecasts are calibrated using only the re-forecasts starting the same day and month (100 members), whereas the Monday real-time forecasts are calibrated using the re-forecasts associated to the previous and next Thursday (a total of 200 members).

The calibration of the real-time forecast is performed by a simple bias correction: the anomalies are computed by removing the ensemble mean climate from the ensemble mean of the real-time forecasts. More sophisticated techniques, like Bayesian methods would require many more re-forecast years than are currently produced. The re-forecasts are also used to calibrate probabilistic forecasts, such as the probability to be in the upper or lower tercile, by defining the probability boundaries. The larger the size of the ensemble re-forecasts, the more accurate the probability boundaries are calculated. The re-forecasts are also used to determine if the forecast anomalies are statistically significant or not, by applying a WMW test (Wonacott and Wonacott1977) to the ensemble distribution of the real-time forecasts and the re-forecasts. The monthly forecast plots show only the areas where the two ensemble distributions are different within the 90% level of confidence.

A major issue with the current configuration of the re-forecast dataset is that it is produced only once a week, creating a difference between the calibrations of Monday and Thursday real-time forecasts. Its ensemble size is also too small to allow a proper assessment of the monthly forecast skill scores, since probabilistic forecasts can be very sensitive to the ensemble size. Furthermore, each re-forecast ensemble includes only 5 members while the daily forecast ensembles include 51 members. Figure 19 shows an example of reliability diagram of the probability that 2-metre temperature is in the upper tercile over Europe obtained with a 5-member, 11-member and 51-member ensemble. The reliability diagram obtained with the 5-member ensemble suggests that the 2-metre temperature forecasts are unreliable over Europe, with a reliability curve almost flat, whereas the 11 and 51-member ensemble members display a reliability curve close to the diagonal. This is a reason why the re-forecasts are rarely used for model skill assessment. The monthly forecast skill assessment is provided by a research experiment with 80 ensemble integrations starting on 1st February, May, August and November 1989 to 2008 with an ensemble size of 15 members.

To address both these issues, the Ensemble re-forecast dataset will be extended when cycle 40r3 becomes operational. In this new configuration, the re-forecasts will be produced twice a week, with an ensemble size of 11 members.

When this extension of the re-forecasts becomes operational, the monthly forecasts will be calibrated over a complete 1 week window centered on the start date of the forecast. For instance, the Thursday monthly forecasts will be calibrated with the re-forecasts associated with the

Thursday start date and those associated with the previous and following Monday. All the re-forecasts will have the same weight. This will have the advantage of increasing considerably the number of ensemble members used to assess the model climatology from 100 (5*20) in the current configuration to 660 (11*20*3). This should allow a more accurate assessment of extreme event probabilities in the model climatology (deciles for instance).

Since the current re-forecast ensemble climatology is already quite large (100-member ensemble), the impact of the extension on the calibrated monthly forecast product is generally modest. The anomalies produced with the current and new climatology are quite similar to a first order. However, there are, at times, clearly detectable local differences (not shown).

Extending the re-forecast window from 1 week to 5 weeks, as is done to generate the widely used Extreme Forecast Index (EFI) product, would have the advantage of increasing the ensemble size of the climatology used to calibrate the real-time forecast, but will have the disadvantage of introducing a spurious signal due to the seasonal cycle. Figure 20 shows the difference of 2-meter temperature computed from a 1-week window and a 5-week window climatology. In this example the windows are centred to the day which is at the peak of the seasonal cycle of 2-metre temperature. The differences can reach up to 1 degree and are of the same order of magnitude as the 2-metre temperature anomalies in week 4. Therefore, a 1-week window will be used to calibrate the monthly forecasts when the extended re-forecasts will become operational.

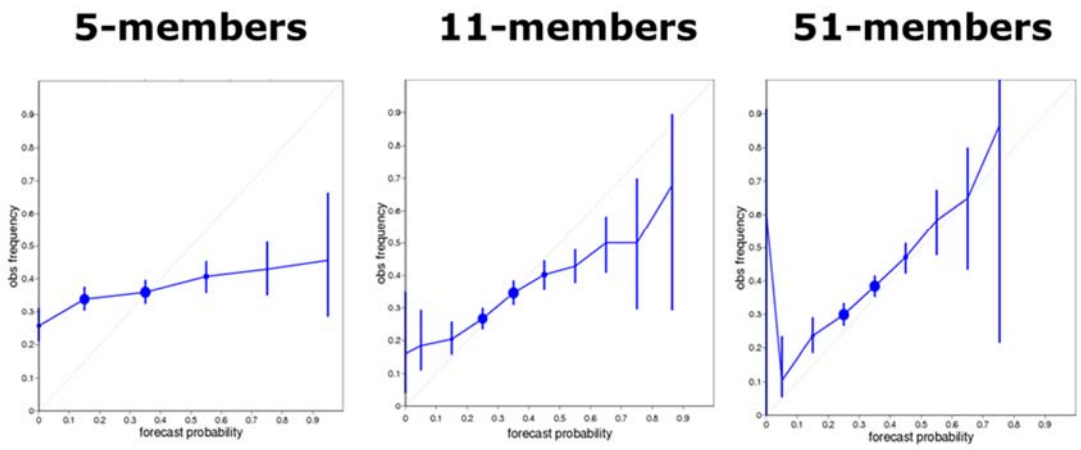

Figure 19: Reliability diagram of the probability that 2-meter temperature anomalies are in the upper tercile for the time range day 26-32. This diagram has been produced from a set of 5-member (left panel), 11-member (middle panel) and 51-member (right panel) re-forecasts starting on 1st November 1980-2010 at a T639 resolution (MINERVA experiment). The solid circles are proportional to the number of cases populating a specific bin.

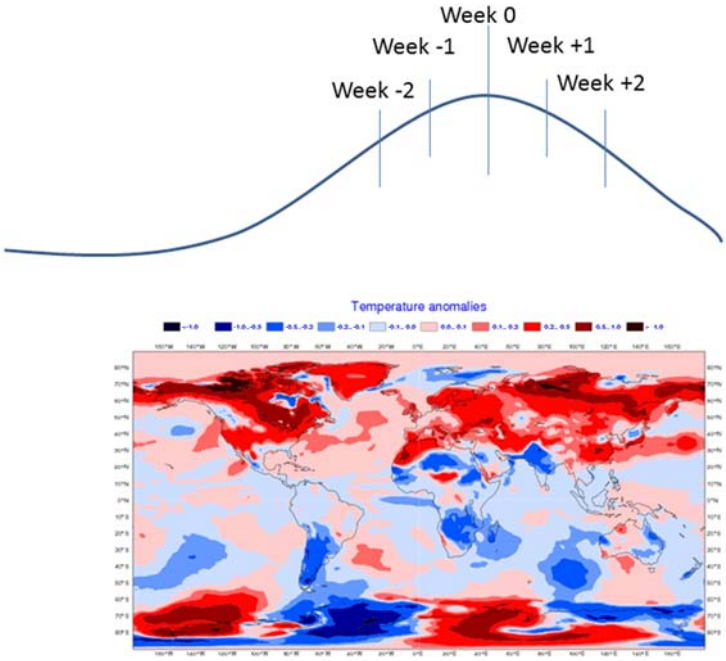

Figure 20: Difference between the 2-metre temperature climatology produced using a 1-week window and a 5-week window. The central date is the 6th of June and the time range is day 26-32.

5.2 Impact of atmospheric horizontal resolution

The impact of increasing the horizontal resolution in the land and atmospheric component has been assessed from experiments produced for the MINERVA project. The MINERVA project (http://www.cpc.ncep.noaa.gov/products/outreach/workshops/CDPW38/3_Wednesday/1_Morning/Kinter.pdf) is a COLA (Center for Ocean, land and Atmospheric research), ECMWF and NCAR (National Center for Atmospheric Research) collaboration aiming to explore the impact of increased atmospheric resolution on model fidelity and prediction skill in a coupled, seamless framework. The project, part of the NCAR Advanced Scientific Discovery Program, consisted of running 7-month integrations with IFS cycle 38r1 (which was operational at ECMWF between June 2012 and June 2013) at three horizontal resolutions (T319, T639 and T1279) on the NCAR Yellowstone supercomputer (72 K-core IBM iDataPlex). The integrations were performed with 91 vertical levels. About 28 million core-hours have been used for the model integrations. The model integrations started on 1 May and 1st November 1980 to 2011 with an ensemble size of 51 members, except for the T1279 integrations which cover a shorter period (1990-2011) with a smaller ensemble size (15 members). The integrations have now been completed, and a subset of the output data has been archived at ECMWF in MARS. In the MINERVA experiments, IFS was coupled to NEMO with an oceanic horizontal resolution of 1 degree (same resolution as currently in operation). This section will discuss the impact of atmospheric resolution over the first 2 months of forecasts in the MINERVA experiments.

5.2.1 Model Climatology

The impact of increasing the horizontal atmospheric resolution on the model biases has been evaluated for several variables. The atmospheric horizontal resolution has overall only a small impact on the sea surface temperature biases (not shown). For month 1, the model biases look very similar at the three different resolutions. At month 2, the increased atmospheric horizontal resolution reduces the cold biases in the sub-tropical North Atlantic and Central Pacific, but increases the warm biases near Africa in the tropical Atlantic. The impact of resolution on the biases of precipitation or 500 hPa geopotential height is also very modest (not shown) with similar patterns and intensity in all three resolutions. The increased resolution also has a negligible impact on the climatology of blocking events, whereas Jung et al (2012) mentioned an improved representation of blocking events when increasing the IFS resolution from T159 to T511 in simulations where IFS was forced by observed SSTs. It is possible that this result was model dependent or that blocking is more sensitive to resolution changes between T159 and T319 (lowest resolution of MINERVA) than beyond. The increased resolution does, however, have a very significant impact on the model variability, with for instance increased and therefore improved skewness and kurtosis for variables like vorticity at 850 hPa (Fig. 21).

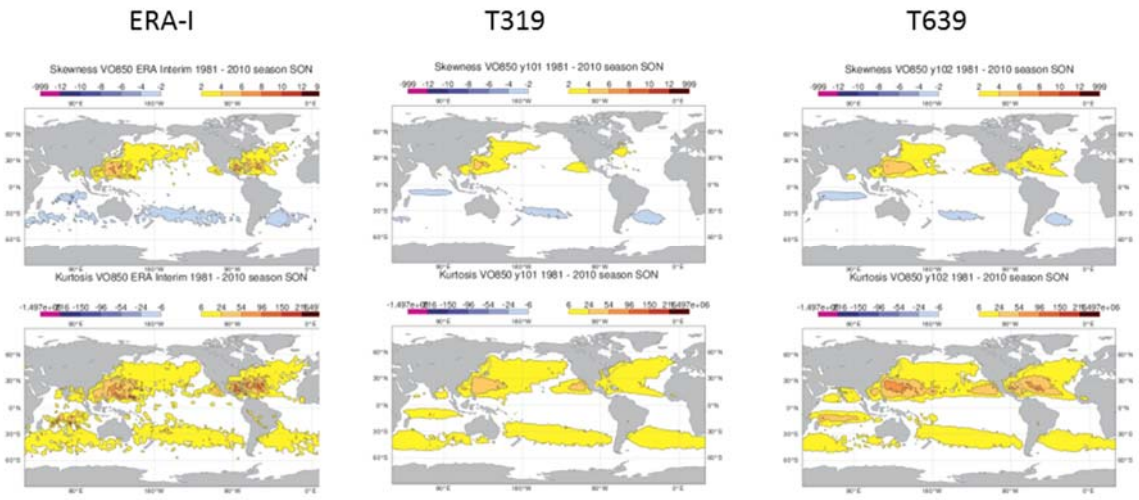

Figure 21: Skewness (top panels) and kurtosis (bottom panels) of vorticity at 850 hPa over the period September to November (most active period of the Atlantic tropical storm season) in ERA Interim (left panels), Minerva T319 (middle panels) and Minerva T639 (right panels) integrations starting on 1st May 1981-2010.

5.2.2 Forecast skill scores

The MJO displays little sensitivity to an increase in IFS horizontal resolution. The skill scores and the amplitude of the MJO (not shown) are almost identical at T319, T639 and T1279. The MJO teleconnections over the northern Extratropics 10 days after an MJO in phase 6 are displayed in

Figures 22. Although the representation of the MJO does not seem to be affected by the increased horizontal resolution, the MJO teleconnections display some improvements with horizontal resolution, although the number of forecasts may be too small to be sure that the differences are not due sampling. Increasing the resolution from T319 to T639 has little impact on the MJO teleconnections after an MJO in Phase 3 (not shown), but it increases the amplitude of the teleconnections after an MJO in Phase 6 (Fig. 22) over North Atlantic positive anomalies, projecting into a stronger negative NAO. In the T1279 integrations, the position of the negative anomalies over the North Atlantic is more consistent with re-analysis than in the T319 and T639 integrations for the teleconnections associated to an MJO in Phase 3 (not shown). However, increasing the resolution from T639 to T1279 seems to have little impact on the teleconnections following an MJO in phase 6 (Fig. 22). Even at T1279, the amplitude of the MJO teleconnections is much weaker than in ERA Interim.

The NAO index based on an EOF projection has been calculated over the common period 1990-2010 for the November start dates for each model and over weekly periods (week 1 to 4). The skill scores (correlation with analysis) are displayed in Figure 23. According to Figure 23, the NAO skill score improves when increasing the resolution from T319 to T639 in week 2, but not in the other weekly periods. The difference in week 2 is statistically significant at the 90% level of confidence. The T1279 integrations display the highest NAO skill scores over the 4 weekly periods. In week 4, the difference between the T1279 and the T639 integrations is statistically significant according to a 10,000 bootstrap re-sampling procedure. The NAO skill scores have also been evaluated for longer time ranges: month 2 and December-January-February (DJF) for the MINERVA integrations starting on 1st November 1980-2010. Results show that the NAO skill scores for month 2 are higher at T639 with a correlation of 0.5 than at T319 with a correlation of 0.37. At the seasonal time scale, T639 integrations also display higher NAO scores (correlation of 0.51 for DJF) than the T319 integrations (correlation of 0.26 for DJF).

Increasing the resolution from T319 to T639 leads to improved RPSS of weekly and monthly mean geopotential height anomalies (not shown) in the northern Extratropics for the forecasts starting on 1st May 1980-2010. The forecasts starting on 1st November 1980-2010 are also more skilful at T639 than at T319, but the difference is smaller than for the forecasts starting on 1st May.

The precipitation skill scores are generally lower than temperature or geopotential height skill scores. Over the Tropics, the model displays positive RPSS only in the first 2 weeks of the forecasts starting on 1st May. In the Extratropics, precipitation skill scores are higher, with positive RPSS up to day 26-32. As for 500 hPa geopotential height, the RPSS of weekly mean precipitation are higher at T639 than at T319 over the Northern Extratropics and in the Tropics (Fig. 24), although the difference in RPSS skill scores is not statistically significant within the 95% level of confidence. Increasing the resolution from T639 to T1279 also has a positive impact on the RPSS of 500 hPa geopotential height and temperature at 850 hPa (Fig. 25) in the extended range over the northern Extratropics. The RPSS skill scores are statistically significantly higher at T1279 than at T639 for the time range day 31-60 at the 95% level of confidence for 500 hPa geopotential height (Fig. 25a) and 850 hPa temperature anomalies (Fig. 25b). However, the skill scores of precipitation (not

shown) are not significantly different. In the Tropics, the skill scores display little sensitivity to the increase of resolution from T639 to T1279 (not shown).

In summary, the MINERVA experiment showed that increasing the horizontal atmospheric resolution from T319 to T639 and T1279 had little impact on the model mean biases but improved the model variability. Forecast skill scores are generally improved when increasing the resolution although more cases would be needed to demonstrate that this impact is statistically significant. However, increasing the resolution from T319 to T639 improves significantly the NAO skill scores in week 2 as well as in seasonal forecasts. Additional experiments are planned with a ¼ degree ocean (instead of the 1 degree ocean resolution in MINERVA) to assess the impact of increased oceanic resolution.

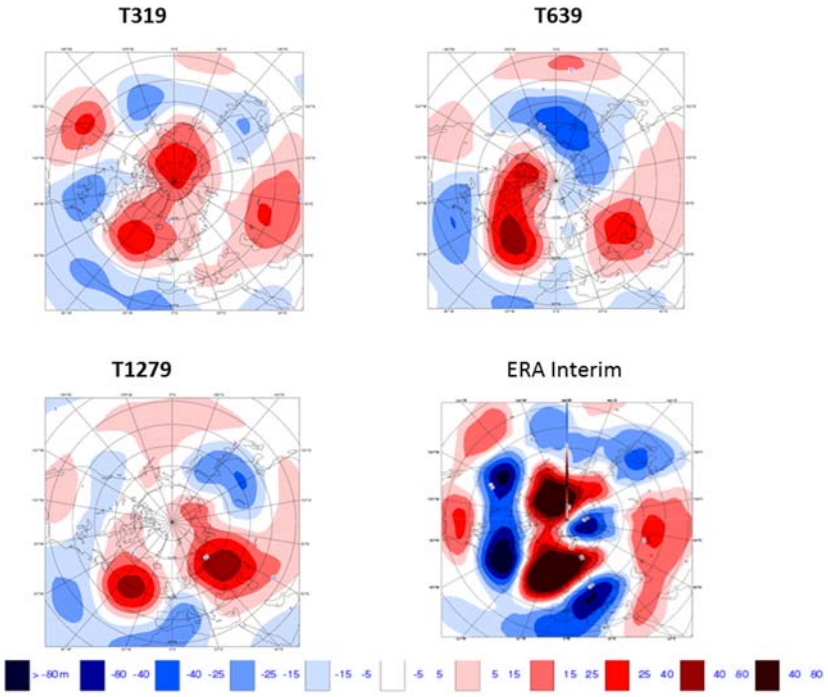

Figure 22: Composites of 500 hPa geopotential height anomalies 10 days after an MJO in Phase 6 for the MINERVA integrations starting on 1st November 1990-200 at t319 (top left panels), t639 (top right panels) and t1279 (bottom left panels). The verification from ERA Interim is displayed in the bottom right panel.

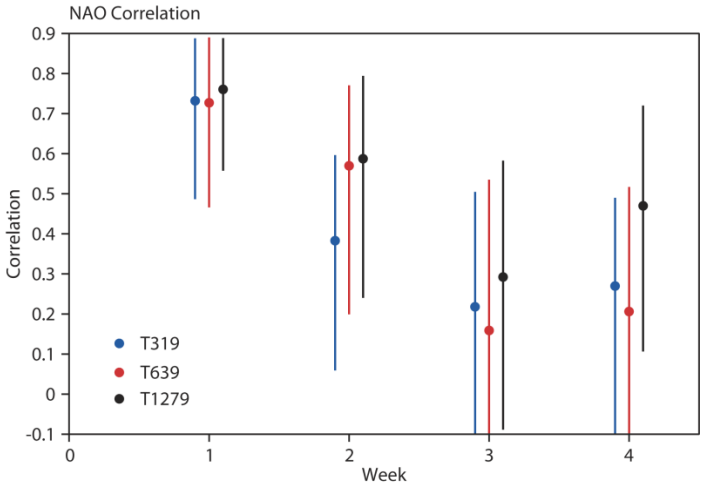

Figure 23: NAO weekly-mean skill scores from MINERVA integrations starting on 1st November 1990-200 at t319 (blue), t639 (red) and t1279 (black). The circles correspond to the correlation between the model integrations and ERA Interim. The vertical lines represent the 95% level of confidence using a 10,000 bootstrap re-sampling procedure.

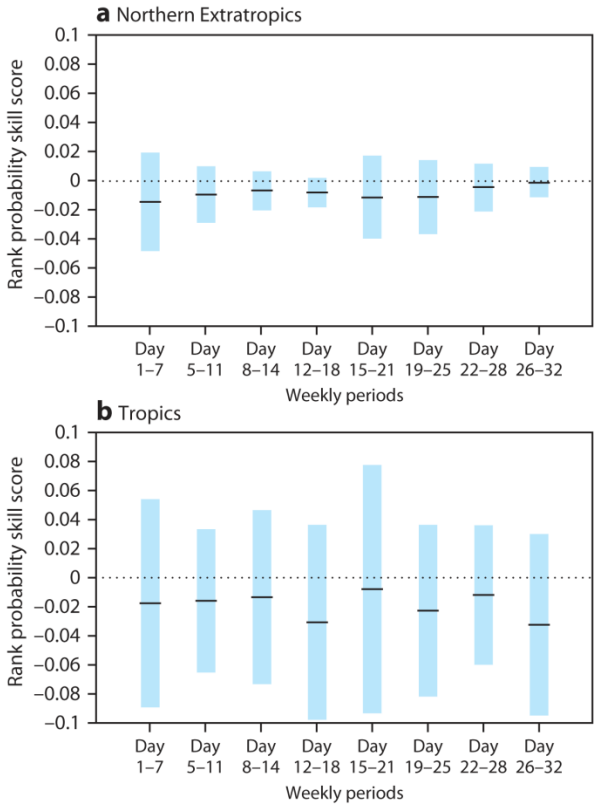

Figure 24: Difference of RPSS skill scores of total precipitations between the T319 and T639 integrations starting on 1st May 1980-2011 for the northern Extratropics (Top panel) and in the tropics (bottom panel). Negative (positive) values indicate that the T639 integrations are more (less) skilful than the T319 integrations. The vertical bars represent the 95% level of confidence using a 10,000 bootstrap resampling procedure.

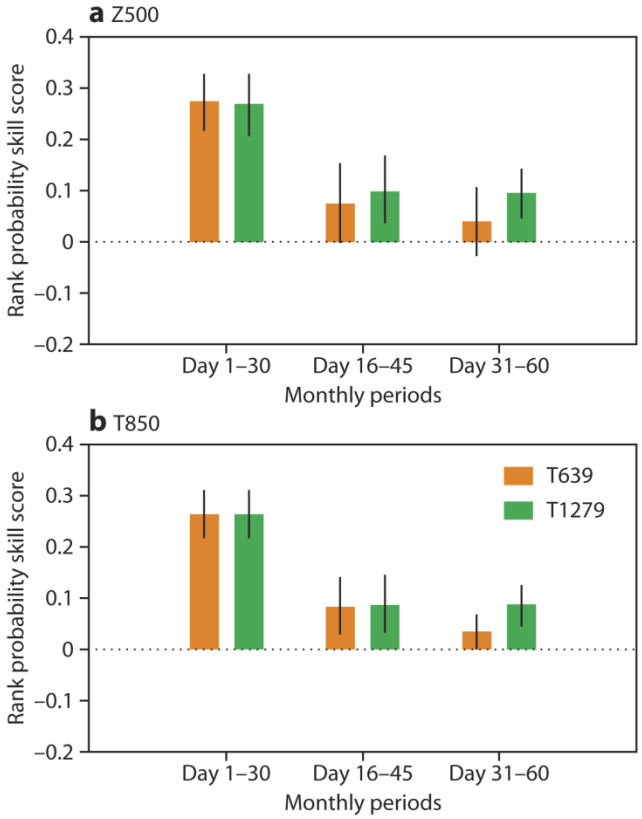

Figure 25: Ranked probability skill scores of 500 hPa geopotential height (top panel) and temperature at 850 hPa (bottom panel) anomalies from the 15-member MINERVA integrations starting on 1st November 1990-2010 at T639 (orange bars) and T1279 (green bars) for three monthly mean periods: day 1-30, day 15-45 and day 30-60. The vertical lines represent the 95% level of confidence using a 10,000 bootstrap re-sampling procedure.

5.3 Impact of sea ice model on the ENS monthly system

To assess the impact of a dynamical sea ice model on sub-seasonal forecasts, the 1 degree NEMO ocean model has been coupled with the dynamic-thermodynamic sea ice model LIM2. The ice and ocean model are coupled every time step and are coupled to the IFS every 3 hours. The monthly forecast system is fully coupled to the ocean-ice model from day 0 (the tendency coupling has been deactivated so that small ocean features which are present in the high-resolution ocean initial conditions will be lost due to the lower resolution ocean model that is used).

The LIM2 dynamic-thermodynamic ice model is described in detail in Fichefet and Morales Maqueda (1997). The thermodynamic calculations are done with a two layer ice model, which may have a layer of snow on top. The ice thickness can vary and the thickness of the two layers within the thermodynamic model change in time to reflect this variation. The presence of brine pockets within the ice and the non-uniformity of the ice thickness are parameterised within the thermodynamic calculations. Ice temperatures from the top layer of the model (ice, or snow if present) are passed to IFS. Sea ice dynamics are modelled using a viscous-plastic scheme (Hibler, 1979).

For the experiments with the sea ice model activated, the ocean and sea ice initial conditions are provided by a NEMOVAR 1 degree analysis which uses the same version of NEMOVAR as MyOcean2 0.25 degree analysis (ORAP5.0, see description here: http://www.myocean.eu/web/69-myocean-interactive-catalogue.php?option=com_csw&view=details&product_id=GLOBAL_REANALYSIS_PHYS_001_017). The reanalysis uses observed sea ice concentration within the 3DVar framework to produce ice concentration and thickness initial conditions for LIM2. There is only one ocean and sea ice ensemble member for each start date instead of the 5 ocean ensemble members for the control run initial conditions.

A 20-year re-forecast set of 15 ensemble members has been integrated, using 1 February, May, August and November start dates. The differences between the monthly forecast system with the sea ice model activated (SEAICE) and the current standard set up of the monthly forecast system (CONTROL) have been calculated. The results are presented in terms of skill scores which are compared to ERA-Interim and then case studies where the performance of the sea ice has been compared to the OSTIA reanalysis (1989-2007). OSTIA reanalysis was used as the sea ice field is better represented at high latitudes than in ERA-Interim; the ERA-Interim has a fixed field above (84N), it is important for recent sea ice decline to capture changes closer to the pole.

First, the impact of changing to an active sea ice model on the large scale circulation of the atmosphere has been evaluated. The Rank probability skill score for the standard monthly forecast system (CONTROL) and the new set up with the sea ice model activated (SEAICE) has been assessed based on the analysis of ERA-Interim. Results (not shown) suggests that adding the sea ice model is largely neutral. There is a slight degradation in skill at weeks one and two but this is not statistically significant. At weeks 3 and 4, where bigger changes in the sea ice field are expected when considering the fully coupled system , some improvements in the forecast skill are visible but again this is not statistically significant. The results are similar for the southern hemisphere as well (not shown).

The skill of forecasting the evolution of the sea ice itself which is a useful forecast variable in terms of shipping and for interaction with the ocean waves has been assessed. The mean sea ice concentration bias in week 4 of prediction for the whole reforecast period has been compared to ERA-Interim for autumn and spring when the sea ice concentration changes the most rapidly in the seasonal cycle (not shown). The most striking difference between the two experiments is that the biases are of opposite sign. LIM2 has a positive bias in the North Atlantic in spring and autumn, probably due to sea surface temperature biases in the NEMO model. In the Antarctic LIM2 displays a positive bias in spring and a negative bias in autumn. The autumn positive bias may be due to errors in the initial conditions, but may also indicative of missing processes in the coupled system that lead sea ice break up such as the wave action which acts to break up the ice and hinder the refreezing of the ice. In the CONTROL setup, the biases are due to the year to year variability compared to recent climatology. The biggest spread in ice conditions in the southern hemisphere is in the freezing season, which is where we see the biggest biases in CONTROL.

The dynamic modelling of sea ice will have the biggest impacts when the sea ice evolution is not well captured by the climatology. Since the turn of the century ice conditions in the Arctic have shown a rapid change with increased melting in the summer season. At the end of the 1990's sea ice still tended to fill the whole of the Arctic basin in September when the sea ice minimum is reached. In 2007 increased melting combined with atmospheric circulation which was favourable to flush ice from the Arctic basin through the Fram Strait led to sea ice minima that had not been seen in recent history. Year 2007 was used as a case study for assessing how the monthly system is able to predict the sea ice evolution as it requires the combination of thermodynamic and dynamic processes to capture the extreme event.

Figure 26 shows the difference between the ensemble mean predicted sea ice concentration for 2nd September and that "observed" (as provided by the OSTIA reanalysis). The results show the errors in the sea ice field are much smaller when a coupled sea ice model is used and the model is able to capture the very anomalous decline in sea ice cover.

Finally the skill of the sea ice model has been assessed for different seasons. Figure 27 displays the correlations for the last week of the forecast period between the predicted sea ice concentration and that of OSTIA. The correlations are large in the ice edge areas and larger than in the Control integration (not shown). In the central area, the correlations are negative in both Control and SEAICE. This is likely due to the fact that pack ice tends to dominate in this region with very small variations in ice concentration. Therefore, any small deviation from observations can create negative correlation. The fact that the correlations exceed 0.6 (and often 0.8) in around the sea ice edge suggests that predictions of the sea ice concentration are very promising.

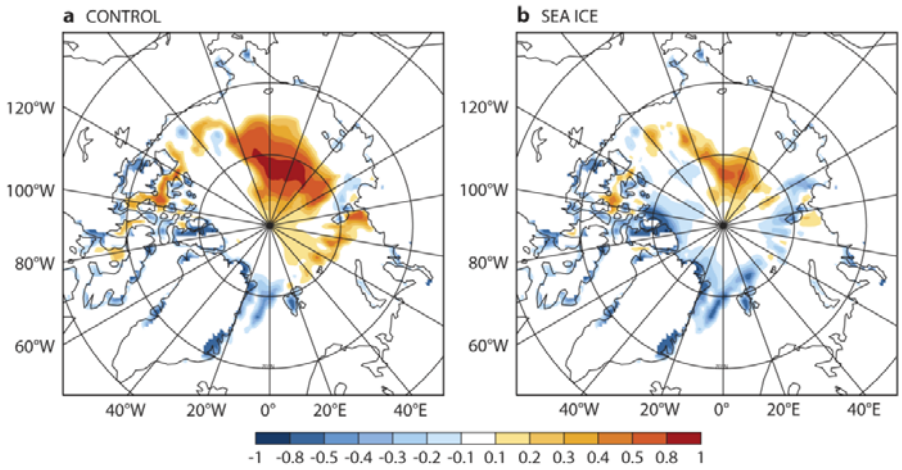

Figure 26: Difference between ensemble mean forecast sea ice concentration and OSTIA at day 32 (2nd September) for 2007. The bottom panel shows the ensemble mean for the active sea ice model and the top panel shows the sea ice field for all ensemble members of the standard set up which are relaxed towards the 2002-2006 climatology.

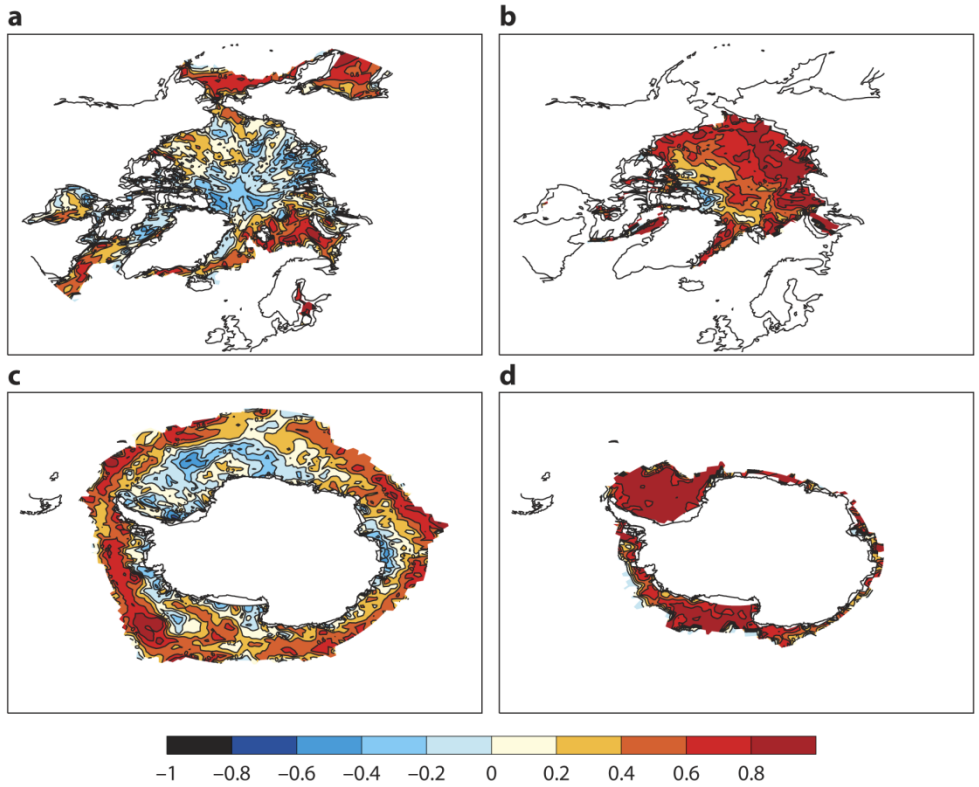

Figure 27: Correlation between week four ensemble mean sea ice concentration forecast and OSTIA for the years 1989-2007. Ensemble mean correlation is shown for winter (left panels) and summer (right panels) for the North Pole (top panels) and the South Pole (bottom panels). Colours in red hues show positive correlation and blue hues show negative. The contour interval is 0.2 with the first contour starting at + or - 0.2. The dark red contours correspond to a positive correlation of 0.8.

5.4 Extension to 45 or 60 days

The previous sections have shown that the skill of the ECMWF monthly forecasts have significantly improved over the past decade. In particular, the skill at week 4 (day 26-32) is now significantly higher than it used to be 10 years ago. In addition, some important predictors for sub-seasonal prediction such as the MJO or Sudden Stratospheric Warmings (SSW) display predictive skill beyond day 32 (Vitart 2014). These results indicate that skilful forecasts could be issued beyond the current 32 days, and that it could be time now to extend the length of the monthly forecasts to 45 days or 60 days. The fact that a full cycle of the MJO is about 60 days is a further argument for extending the ECMWF monthly forecasts. However, an increase in the length of the monthly forecasts needs to be justified by demonstrated skill beyond day 32, most especially over Europe which is our main focus.

In order to evaluate if the current coupled IFS-NEMO ensemble has skill beyond day 32, a series of 45-day integrations have been produced once a month over the period 1989-2008 (a total of 20*12=240 starting dates). The ensemble size is 15 members and the version of IFS used in this experiment is cycle 38R1 (the version that was operational in 2012). The model resolution is the

same as in the ensemble currently in operation: T639 for the first 10 days and T319 after day 10, with 91 vertical levels and the top of the atmosphere at 0.01 hPa. Skill scores have been computed over 15-day periods (day 1-15, 16-31, 31-45) so that the last 15-day period corresponds to a period beyond the time frame of the current monthly forecasts. According to Figure 28, geopotential height at 500 hPa and temperature at 850 hPa display positive RPSS during all three 15-day periods, although, as expected, there is a significant drop in the forecast skill scores from one period to the next. For the time range day 31-45, the RPSS is small, but positive over the northern Extratropics. According to a 10,000 bootstrap re-sampling procedure, the RPSS is significantly positive within the 95% level of confidence.

Over Europe, the RPSS for day 31-45 is even larger than over the Northern Extratropics, and also positive within the 95% level of confidence. These results suggest that the model displays positive skill up to day 45, and that a forecast range extension from 32 to 45 days could deliver valuable forecasts.

As a further test that this is indeed the case, the RPSS of the 31-45 day forecasts have been compared to the RPSS of persisting the day 16-31 forecasts. Results indicate that these later display a negative RPSS, close to -0.02 for both Z500 and T850, and the difference of skill scores between persistence and day 31-45 forecasts is statistically significant. This result confirms that extending the monthly forecasts to 45 days would provide useful and skilful forecasts, better than both climatology and persisting 16-31 day forecasts.

Extending the monthly forecasts up to 60 days would also be useful, and would allow in particular more frequent updates of month 2 of the seasonal forecasting system. Figure 25, from the MINERVA integrations, shows that there is a significant positive skill for day 31-60 and that this skill is likely to increase with increased resolution (see MINERVA results in Section 5.2.2) .

Another way of assessing the skill of the sub-seasonal forecasts in the Extratropics consists in associating each forecast to a weather regime over the Euro-Atlantic sector and computing the associating probabilistic skill score. In this study, the four weather regimes described in Cassou (2008) have been used to classify 500 hpa geopotential height anomalies: positive NAO, negative NAO, Atlantic ridge and Scandinavian blocking. The weather regime skill scores have been assessed using the continuous ranked probability score (CRPS, e.g. Wilks 2011). According to Figure 29, the forecasts display some skill to predict the probability to be in a specific weather regime up to about 45 forecast days, when the CRPS becomes negative. Over the Pacific-North America sector, the skill limit extends to about 55 days. The amplitude of the CRPS increases when the resolution is increased to T639, particularly in the extended range, although the day when CRPS becomes negative remains almost the same as at T319. The CRPS calculations in Figure 29 include the skill of all 4 Euro-Atlantic regimes. However, the CRPS varies a lot from one regime to another. For the Euro-Atlantic sector, the model displays much higher CRPS for positive NAO and Atlantic ridge, but low skill beyond day 10 to predict Scandinavian blocking and negative NAO (not shown). In addition, increasing the horizontal resolution from T319 to T639 seems to benefit more the prediction of positive NAO than the prediction of negative NAO (not shown).

In summary, extending the monthly forecast up to day 45 would provide skilful and useful forecasts beyond day 30. This would increase the cost of the operational ENS by less than 4 % (this is because the most expensive part of ENS is producing the first 10 forecast days, since they are run at higher resolution than the forecasts beyond 10 days).

In addition, it is worth considering that most operational centres are now producing sub-seasonal forecasts up to 45 or 60 days. In the S2S database of sub-seasonal to seasonal forecasts (see Section 4), ECMWF sub-seasonal forecasts are now the shortest, when other centres like NCEP provide sub-seasonal forecasts up to day 45.

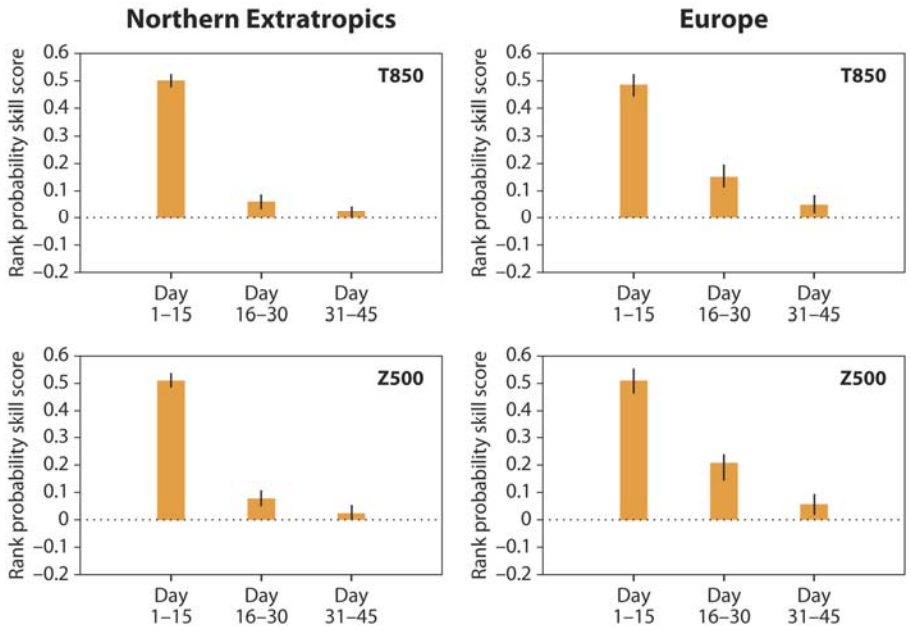

Figure 28: Ranked probability skill scores of 500 hPa geopotential height (top panel) and temperature at 850 hPa temperature (bottom panel) over the northern Extratropics (left panels) and Europe (right panel) for 3 15-day averaged periods: day 1-15, day 16-30 and day 31-45. The RPSS has been computed from a series of 45-day 15-member re-forecasts starting on the first of each month over the period 1989-2008.

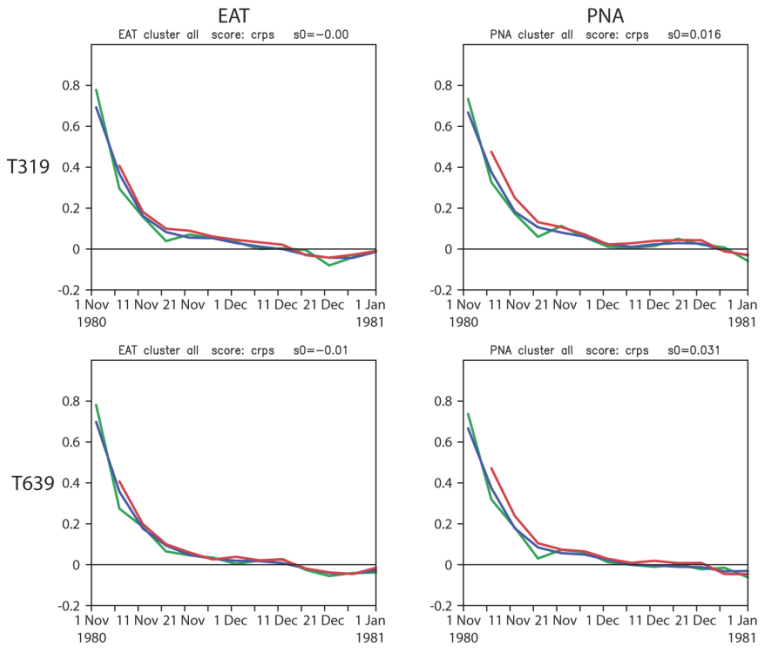

Figure 29: Continuous Ranked Probability Score (CRPS) of the probability to be in one of the 4 Euro-Atlantic weather regimes (left panels) and Pacific-North America weather regimes (right panels) as a function of lead time for all the MINERVA integrations starting on 1st November at T319 (top panels) and T639 (bottom panel). The green line shows the CRPS for 5-day means, the red line shows the CRPS for 15-day mean. The blue line shows the 5-day mean CRPS after smoothing.

6 The sub-seasonal to seasonal (S2S) prediction project

Recent publications (e.g. Brunet et al. 2010; Hurrell et al. 2009; Shapiro et al. 2010; Shukla et al. 2010) have stressed the importance of and need for collaboration between the weather and climate communities to better tackle shared critical issues, and most especially to advance sub-seasonal to seasonal prediction. Such an initiative would help bridge the gap between the numerical weather and short-term climate communities and be an important step towards a seamless weather/climate prediction system. Weather, climate, and Earth-system prediction services would greatly benefit from this joint effort. Based on this proposal and on the potential for improved forecast skill at the sub-seasonal to seasonal time range, a sub-seasonal prediction (S2S) research project has been established. Frédéric Vitart (ECMWF) and Andrew Robertson (International Research Institute) are the co-chairs of the S2S project (http://www.wmo.int/pages/prog/arep/wwrp/new/S2S_project_main_page.html). Its main goal is to improve our understanding of the predictability on the sub-seasonal to seasonal timescale, improve forecast skill over this time range and promote its uptake by operational centres and exploitation by the applications community (Vitart et al, 2012).

To achieve its main goal, an extensive database is being established, containing sub-seasonal (up to 60 days) forecasts and reforecasts (sometimes known as hindcasts), modelled in part on the

THORPEX Interactive Grand Global Ensemble (Richardson et al, 2005) database for medium range forecasts (up to 15 days) and the Climate-System Historical Forecast project (CHFP) for seasonal forecasts. Models from 11 operational centres will contribute to the S2S database which will be hosted at ECMWF. Figure 30 shows the details of each operational model which will contribute to the S2S database.

The design of the S2S database is challenging since unlike for TIGGE, the model configuration for sub-seasonal prediction varies greatly from one operational centre to another. For instance, some models are integrated every day (e.g. UKMO, NCEP), others are integrated several times a week (e.g. ECMWF, EC, CAWCR), and others have a monthly frequency (e.g. Météo-France, CMA). The ensemble size and model resolution also vary greatly between the 11 ensembles. Some re-forecasts are produced once (e.g. NCEP), others are produced "on the fly" to take into account model changes as at ECMWF.

The S2S database will provide access to near real-time forecasts with a 3-week delay from 2015. This database will be very useful since it will allow each centre to compare the performance of their ensembles with the others, thus helping to identify strengths and weaknesses of systems' design. This database will also help answer some important scientific questions relative to the sub-seasonal to seasonal prediction, like the benefit of multi-model sub-seasonal forecasting.

The research topics of the WWRP/WCRP Sub-seasonal to Seasonal Prediction project (S2S) are being organized around a set of five sub-projects (Madden-Julian Oscillation, Monsoons, Africa, Extremes and Verification), each intersected by the cross-cutting research and modelling issues, and applications and user needs discussed above. These research topics are also very relevant to ECMWF research activities. For instance, a main topic of the MJO sub-project will be to investigate the MJO Maritime Continent problem (as described in Section 2). S2S is endorsing a field campaign over the Maritime Continent in 2017/2018. The simulation of tropical-extratopical MJO teleconnections will also be an important topic for S2S. It is hoped that these sub-projects will provide a vehicle for broad community research engagement in sub-seasonal to seasonal prediction.

Sub-seasonal real-time Operational Forecasts

	Time-range	Resol.	Ens. Size	Freq.	Hcsts	Hcst length	Hcst Freq	Hcst Size
ECMWF	D 0-32	T639/319L62	51	2/week	On the fly	Past 18y	weekly	5
UKMO	D 0-60	N96L85	4	daily	On the fly	1989-2003	4/month	3
NCEP	D 0-60	N126L64	16	daily	Fix	1999-2010	daily	4
EC	D 0-35	0.6x0.6L40	21	weekly	On the fly	Past 15y	weekly	4
CAWCR	D 0-120	T47L17	33	weekly	Fix	1989-2010	3/month	33
JMA	D 0-34	T159L60	50	weekly	Fix	1979-2009	3/month	5
KMA	D 0-30	T106L21	20	3/month	Fix	1979-2010	3/month	10
CMA	D 0-45	T63L16	40	6/month	Fix	1982-now	monthly	48
Met.Fr	D 0-60	T63L91	41	monthly	Fix	1981-2005	monthly	11
SAWS	D 0-60	T42L19	6	monthly	Fix	1981-2001	monthly	6
HMCR	D 0-60	1.1x1.4 L28	10	monthly	Fix	1979-2003	monthly	10

Figure 30: This table shows the main characteristics of the 11 models which will contribute to the S2S database

7 Conclusions

Monthly forecasts have been produced operationally at ECMWF since 2002. At the time of writing (July 2014), monthly forecasts are produced twice a week with the ECMWF coupled ensemble (ENS) which includes 51 forecasts, run daily with a T639 (about 32 km) horizontal resolution up to day 10 and a T319 (about 64 km) resolution from day 10 to 32, and with 91 vertical levels up to 0.01 hPa in the atmosphere. ENS is coupled to a 1-degree ocean model (NEMO) from initial time since November 2013. ENS simulates initial uncertainties in the atmosphere using a combination of singular vectors and perturbations defined by the ECMWF ensemble of data assimilations, and in the ocean using the 5-member ocean ensemble of assimilations (ORA-S4, Mogensen et al 2012). ENS simulates model uncertainties in the atmosphere using two stochastic schemes. Monthly forecasts are calibrated using a re-forecast suite that includes 5-member ensembles run once a week for the past 20 years.

This study has shown that the skill of the ECMWF monthly forecasts has improved since 2002. The improvements in the skill scores are particularly high for the prediction of the MJO which is an important source of predictability on the sub-seasonal time scale. Over the northern Extratropics, the prediction skills of the NAO and 2-metre temperature over the northern Extratropics have also increased over the 10-year period, particularly for day 12-18. Vitart (2014) showed that a large portion of the improvements in the NAO skill scores can be attributed to improvements in the prediction of the MJO. For 2-metre temperature, the skill for day 19-25 in 2012 is getting close to the skill for day 12-18 in 2002. Similar improvements are visible in the upper-air fields, especially for large-scale phenomena such as the NAO. The improvements in the monthly re-forecast skill scores reported in this study are likely to give a conservative estimation of the improvements in the real-time forecasts since this study does not take into account improvements in the generation of atmospheric initial conditions, except for the change from ERA 40 to ERA Interim in 2008. Over the past 10 years, the quality of the initial conditions has improved thanks to better data assimilation schemes, model improvements and the use of new observing systems.

The improvement in the monthly forecasting skill scores can be partially explained by the fact that the model is now able to exploit the predictability associated with the MJO. 10 years ago the MJO was too weak after a few days of model integrations for the extratropical weather forecast to benefit from this source of predictive skill. In fact, Vitart (2014) shows that, 10 years ago, the monthly forecast skill scores used to be worse when there was an MJO in the initial condition, whereas, now, the skill scores tend now to be higher when there is an MJO in the initial condition.

However there are still important scientific issues which prevent the model from fully exploiting the predictability associated to the MJO and stratospheric initial conditions. Regarding the MJO for instance, the model often struggles to propagate the MJO across the Maritime Continent. Figure 8 showed that this has the potential of significantly impacting the weather forecast over Europe 3 to 4 weeks in advance. Another important issue with the MJO is the fact that the MJO teleconnections seem much weaker in the model than in ERA Interim over the Euro-Atlantic sector (NAO projection). It is not entirely clear if this is a real issue or it is due to sampling (20 years of reanalysis may not be enough to fully assess the amplitude of the MJO teleconnections). If this is a real issue, then the monthly forecast is not yet fully exploiting all the predictability associated with the MJO.

Another important issue for monthly forecasting is the impact of sudden stratospheric warmings (SSWs) on the European winter weather, and more generally the quality of the simulation of stratosphere-troposphere interactions in the ECMWF model. Section 3.2 showed that the coupled model displays remarkable skill in predicting SSWs up to 3 weeks in advance, but the impact of the SSW on the NAO and the lower troposphere is much weaker than observed. These issues have also been reported in other operational centres (e.g NCEP), and some will be addressed by the WWRP-WCRP sub-seasonal to seasonal prediction project (S2S) through the analysis of a TIGGE-like database of sub-seasonal to seasonal forecasts, and possibly coordinated experiments.

These issues also suggest that the current versions of IFS still not fully exploits all the various sources of sub-seasonal predictability and that there is still room for the monthly forecast skill scores to improve in the future.

Planned forthcoming changes to the ECMWF operational ensemble should help maintain the increase in sub-seasonal forecast skill and provide users with valuable and skilful sub-seasonal forecasts. These changes include extending the re-forecasts, using a dynamical sea-ice model in the forecasting system instead of persisting sea-ice and using a high-resolution ocean model in addition to further changes in model parameterisation and horizontal resolution. Results from the MINERVA experiment have shown some improvement in forecast skill scores, particularly for the prediction of the NAO in week 2. However, increasing the atmospheric resolution from T639 to T1279 leads to only a modest improvement of skill scores over Europe. It is possible that an increase of the ocean resolution from 1 degree to a quarter of a degree will have a stronger impact on the skill scores over Europe, through a better representation of the Gulf Stream in the North Atlantic and a better NAO prediction. Results of coupled integration with a quarter of a degree version of NEMO have not been shown in this report, since experiments with the coupled system with a ¼ degree version of NEMO have just started at the time of writing. Additional improvements are expected from the implementation of a sea-ice model, coupled wave-ocean mixing and a better representation of initial condition (e.g. coupled data assimilation) and model uncertainties in the ocean. The past and future improvements in the monthly forecasts should make it possible now to extend the monthly forecasts to 45 or 60 days and produce skilful and reliable forecasts beyond day 30.

Acknowledgements

The authors are grateful for comments and support to this report provided by Erland Källén and Alan Thorpe. We acknowledge editorial support by Samantha Moreby, Anabel Bowen and Simon Witter.

References

Baldwin, M.P. and T.J. Dunkerton, 2001: Stratospheric Harbingers of Anomalous Weather Regimes. Science, 294, 581-584.

Bechtold., P., M. Koehler, T. Jung, P. Doblas-Reyes, M. Leutbecher, M. Rodwell and F. Vitart, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time scales, Quart. J. R. Meteorol. Soc., 134, 1337-1351.

Beljaars, A., G. Balsamo, A. Betts and P. Viterbo, 2007: Atmosphere/surface interactions in the ECMWF model at high latitudes. Proc. of ECMWF Seminar on Polar meteorology, 4-8 September 2006, ECMWF, 153-168.

Boville B. A., 1984. The influence of the polar night jet on the tropospheric circulation in a GCM. J. Atmos. Sci., 41. 1132–1142.

Brunet, G., M. Shapiro, D. Hoskins, M. Moncrieff, R. Dole, G.N. Kiladis, B. Kirtman, A. Lorenc, B. Mills, R. Morss, S. Polavarapu, D. Rogers, J. Schaake and J. Shukla, 2010: Collaboration of the weather and climate communities to advance subseasonal to seasonal prediction. Bulletin of the American Meteorological Society, 1397-1406.

Buizza, R., Leutbecher, M., & Isaksen, L., 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. Q. J. R. Meteorol. Soc., 134, 2051-2066.

Cassou C, Terray L P., 2005: Tropical Atlantic influence on European heat waves. J. Climate 18 2805-2811

Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. Nature, doi:10.1038/nature07286.

Cohen, J. L., and D. Entekhabi,1999: Eurasian snow cover variability and North- ern Hemisphere climate predictability, Geophys. Res. Lett., 26, 345 – 348, 1999.

Dee, D.P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., doi: 10.1002/qj.828, also available as ERA Report Series, n. 9, pp. 71.

Della Marta P. M. Luterbacher, H.von Weissenfluh, E. Xoplaki M.Brunet H Wanner, 2007: Summer heat waves over western Europe 1880-2003, their relationship to large-scale forcings and predictability Clim Dyn 29 251-275.

Dutra, E., G. Balsamo, P. Viterbo, P.M.A. Miranda, A. Beljaars, C. Schär, and K. Elder, 2010: An improved snow scheme for the ECMWF land surface model: description and offline validation. J. Hydrometeor., 11, 899-916. doi: 10.1175/2010JHM1249.1., also available as ECMWF Tech. Memo. 607.

Fichefet, T., and M.A. Morales Maqueda, 1997: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. Journal of Geophysical Research, 102, 12,609-12,646.

Hagedorn, R., Buizza, R., Hamill, M. T., Leutbecher, M., & Palmer, T. N., 2012: Comparing TIGGE multi-model forecasts with re-forecast calibrated ECMWF ensemble forecasts. Q. J. Roy. Meteorol. Soc., 138, 1814-1827.

Hibler, W. D., 1979: A Dynamic Thermodynamic Sea Ice Model. J. Phys. Oceanogr., 9, 815–846.

van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. Koster, S. Seneviratne and H. Camargo Jr, 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, Clim. Dyn., doi:10.1007/s00382-010-0956-2.

Hurrell, J., G. Meehl, D. Bader, T. Delworth, B. Kirtman, and B. Wielicki 2009: A unified modelling approach to climate prediction. Bull Am Met Soc., 90,1819-1832.

Jung, T., M. J. Miller, T. N. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. A. Cash, J. L. Kinter III, L. Marx, C. Stan, K. I. Hodges, 2012: High-Resolution Global Climate Simulations with the ECMWF Model in the Athena Project: Experimental Design, Model Climate and Seasonal Forecast Skill. J. Climate, 25, 3155-3172.

Koster RD, Dirmeyer PA, Guo Z, Bonan G, Cox P, Gordon C. Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu C, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud Y, Taylor C. Verseghy D, Vasic R, Xue Y and Yamada T., 2004 : Regions of strong coupling between soil moisture and precipitation. Science, 305.

Koster, R. D.. S. Mahanama, T. Yamada, G. Balsamo, M. Boisserie, P. Dirmeyer, F. Doblas-Reyes, T. Gordon, Z. Guo, J.-H. Jeong, D. Lawrence, Z. Li, L. Luo, S. Malyshev, W. Merryfield S. I. Seneviratne, T. Stanelle, B. van den Hurk, F. Vitart, and E. F. Wood, 2009: The Contribution of Land Surface Initialization to Subseasonal Forecast Skill: First Results from the GLACE-2 Project, Geophys. Res. Lett., 2009GL041677R.

Koster, R.D., S. P. P. Mahanama, T. J. Yamada, G. Balsamo, A. A. Berg, M. Boisserie, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J.-H. Jeong, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. van den Hurk, F. Vitart, and E. F. Wood, 2010: The Second Phase of the Global Land-Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill, J. Hydrometeor., doi: 10.1175/2011JHM1365.1.

Lin, H., G. Brunet and J. Derome, 2008: Forecast skill of the Madden-Julian Oscillation in two Canadian atmospheric models. Mon. Wea. Rev., **136**, 4130-4149.

Madden, R. A. and P.R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. J. Atm. Sci., 5, 702-708.

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

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer and F. Vitart, 2011: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. memo 656.

Mogensen, K., M. Alonso Balmaseda, A. Weaver, 2012: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. ECMWF Research Department Technical memorandum No. 668, pp 59 (available from ECMWF).

Morcrette, J.-J, P. Bechtold, A. Beljaars, A. Benedetti, A. Bonet, F. Doblas-Reyes, J. Hague, M. Hamrud, J. Haseler, J.W. Kaiser, M. Leutbecher, G. Mozdzynski, M. Razinger, D. Salmond, S. Serrar, M. Suttie, A. Tompkins, A. Untch and A. Weisheimer, 2007: Recent advances in radiation transfer parametrizations. ECMWF Technical Memorandum, 539, 50 pp.

Müller, W.A., C. Appenzeller, F.J. Doblas-Reyes, and M.A. Liniger, 2005: A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. J. Climate, 18, 1513-1523.

Nakazawa, T., 1988: Tropical super clusters within intraseasonal variations over the western Pacific. J. Meteor. Soc. Japan,64, 17-34.

Norton, W. A., 2003. Sensitivity of northern hemisphere surface climate to simulation of the stratospheric polar vortex. Geophys. Res. Lett., 30.1627.

Orsolini, Y.J. and N. Kvamstø, 2009 : The role of the Eurasian snow cover upon the wintertime circulation: decadal simulations forced with satellite observations. J. Geophys. Res., 114, D19108, doi:10.1029/2009JD012253.

Orsolini, Y. J., R. Senan, G. Balsamo, F.J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, R.E. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts, *Clim. Dyn. 01/2013; DOI:10.1007/s00382-013-1782-0.*

Palmer, T N, Buizza, R., Leutbecher, M., Hagedorn, R., Jung, T., Rodwell, M, Vitart, F., Berner, J., Hagel, E., Lawrence, A., Pappenberger, F., Park, Y.-Y., van Bremen, L., Gilmour, I., & Smith, L., 2007: The ECMWF Ensemble Prediction System: recent and on-going developments. A paper presented at the 36th Session of the ECMWF Scientific Advisory Committee. ECMWF Research Department Technical Memorandum n. 540, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pp. 53.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer M., & Weisheimer, A., 2009: Stochastic parametrization and model uncertainty. ECMWF Research Department Technical Memorandum n. 598, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pp. 42.

Peings, Y., H. Douville, R. Alkama, and B. Decharme, 2011 : Snow contribution to springtime atmospheric predictability over the second half of the twentieth century, Clim Dyn, 37(5-6), 985.

Quesada B. R. Vautard P. Yiou M. Hirschi S. Seneviratne, 2012: Asymmetric European summer heat predictability from wet and dry southern winter and springs. Nature climate change DOI:10.1038/NCLIMATE1536

Richardson, D., R., Buizza and R. Hagedorn, 2005: Final report of the 1st Workshop on the THORPEX Interactive Grand Global Ensemble (TIGGE). WMO TD No. 1273, WWRP-THORPEX No. 5 pdf

Rodwell, M and co-authors, 2012: Characteristics of occasional poor medium-range forecasts for Europe. ECMWF Newsletter, **131**, 11-15.

Scaife A.A., Knight J.R., Vallis G.K., Folland C.K., 2005. A stratospheric influence on the winter NAO and North Atlantic surface climate. Geophys. Res. Lett., 32: L18715, DOI:10.1029/2005GL023226.

Shapiro and others, 2010 An Earth-system Prediction Initiative for the 21st Century. BAMS doi: 10.1175/2010BAMS2944.1

Shukla J. et al 2010: Towards a new generation of world climate research and computing facilities BAMs, 91, 1407-1412.

Tompkins, A.M., K. Gierens and G. Radel, 2005: Ice supersaturation in the ECMWF Integrated Forecast System. ECMWF technical Memorandum, 481, 15 pp.

Vitart, F., J.L. Anderson and W.F. Stern, 1997; Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. J.Climate, 10, 745-760.

Vitart, F, 2004: Monthly forecasting at ECMWF. Mon. Wea. Rev., 132(12), 2761-2779.

Vitart, F., S. Woolnough, M. A. Balmaseda and A. Tompkins, 2007: Monthly forecast of the Madden-Julian Oscillation using a coupled GCM. Mon. Wea. Rev., 135, 2700-2715.

Vitart, F., A. Bonet, M. Balmaseda, G. Balsamo, J.-R. Bidlot, R. Buizza, M. Fuentes, A. Hofstadler, F. Molteni, and T. Palmer, 2008: The new VarEPS- monthly forecasting system: a first step towards seamless prediction. Q. J. R. Meteorol. Soc., 134, 1789-1799.

Vitart, F., 2009: Impact of the Madden Julian Oscillation on tropical storms and risk of landfall in the ECMWF forecast system. Geophys. Res. Lett., L1 5802, doi:10.1029/2009GL039089.

Vitart, F., A. Leroy and M.C. Wheeler, 2010: A comparison of dynamical and statistical predictions of weekly tropical cyclone activity in the Southern Hemisphere. Mon. Wea. Rev, 138, 3671–3682.

Vitart, F., A. Robertson and D. Anderson, 2012: Subseasonal to Seasonal Prediction Project: bridging the gap between weather and climate. WMO Bulletin, 61(2), 23-28.

Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. Quart. J. Roy. Meteor. Soc, 140, 1889-1899.

Weigel, A.P., M.A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. Mon. Wea. Rev., 135, 118-124.

Weisheimer A, Doblas-Reyes P, Jung T and Palmer T. 2011. On the predictability of the extreme summer 2003 over Europe, Geophy. Res. Let., 38, doi:10.1029/2010GL046455.

Weisheimer, A., S. Corti, T.N. Palmer and F. Vitart, 2014: Addressing model error through atmospheric stochastic physical parameterisations: Impact on the coupled ECMWF seasonal forecasting system. Phil. Trans. R. Soc. A , 372,201820130290, doi: 10.1098/rsta.2013.0290.

Wheeler, M.C. and H.H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. Mon. Wea. Rev., 132, 1917-1932.

Wilks, D, 2011: Statistical methods in atmospheric sciences. Academic Press, pp 704, ISBN: 978-0-12-385022-5.

Wonacott, T.H. and R.J. Wonacott, 1977: Introductory statistics. John Wiley, 650 pp.