TECHNICAL MEMORANDUM

694

Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years

Frédéric Vitart

Research Department

February 2013

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

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

Abstract

Sub-seasonal forecasts have been routinely produced at ECMWF since 2002 with re-forecasts produced "on the fly" to calibrate the real-time sub-seasonal forecasts. In this study, the skill of the re-forecasts from April 2002 to March 2012 and covering a common set of years (1995 to 2001) has been evaluated. Results indicate that the skill of the ECMWF re-forecasts to predict the Madden Julian Oscillation has improved significantly since 2002, with an average gain of about 1 day of predictability per year. The amplitude of the MJO has also become more realistic, although the model still tends to produce MJOs which are weaker than in the ECMWF re-analysis. As a consequence, the ability of the ECMWF model to simulate realistic MJO teleconnections over the northern and southern Extratropics has improved dramatically over the past 10 years. Forecast skill scores have also improved in the Extratropics. For instance, weekly mean forecasts of the North Atlantic Oscillation Index are significantly more skillful in recent years than ten years ago. A large part of this improvement seems to be linked to the improvements in the representation of the Madden Julian Oscillation. Skill to predict 2-metre temperature anomalies over the northern Extratropics has also improved almost continuously since 2002, with a gain of almost a week of predictability in the last weeks of the sub-seasonal forecasts. Changes in the horizontal and vertical resolutions of the atmospheric model had only a small impact on the skill scores, suggesting that most of the improvements in the ECMWF sub-seasonal forecasts were due to changes in model physics which were primarily designed to improve the model climate and medium-range forecasts.

1 Introduction

Sub-seasonal forecasting fills the gap between medium-range weather forecasting and seasonal forecasting. It is often considered a difficult time range, since the time scale is sufficiently long so that much of the memory of the atmosphere initial conditions is lost, and it is probably too short so that the variability of the ocean is not large enough, which makes it difficult to beat persistence. However, an important source of predictability is the Madden-Julian Oscillation (see e.g. Ferranti et al. 1990). Other potential sources of predictability include the stratospheric initial conditions that can project in the troposphere over a time scale of a month (Baldwin et al. 2003), the land surface initial conditions (Koster et al., 2011), snow initial conditions (Jeong et al., 2012).

The interest in monthly forecasting was triggered by Miyakoda et al. (1983), who showed how the pronounced blocking event of 1977 was successfully reproduced in 1-month forecasts produced by some general circulation models. Miyakoda et al. (1986) found some marginal skill in eight January 1-month integrations using a 10-day mean filter applied to the prognoses. The report of successful forecasts beyond day 10 triggered a lot of interest at that time. Many of the world's operational prediction centres started to produce a number of large experiments on extended range forecasting (Tracton et al. 1989; Owen and Palmer 1987; Molteni et al. 1986; Deque and Royer 1992). For instance, the European Centre for Medium-Range Weather Forecasts (ECMWF) used operational forecasts to produce a pair of 31-day forecasts starting at two consecutive days for every month from April 1985 to January 1989 (Palmer et al. 1990). These experiments showed some moderate skill after 10 days (Miyakoda et al. 1986; Deque and Royer 1992; Brankovic et al. 1988), particularly when comparing the forecast to climatology. However, most of these experiments failed to display significant skill beyond persisting the medium-range forecasts and therefore did not lead to operational implementation of sub-seasonal forecasts at that time. At ECMWF, the monthly forecasting predictability was revisited in the early 2000th. An experimental monthly forecasting system was set up in 2002 and ran routinely every 2 weeks from March 2002 to October 2004. This experimental forecasting system displayed higher skill than climatology and also higher skill beyond day 10 than persisting medium-range forecasts (Vitart 2004). As a consequence this system became operational in October 2004 and was merged to the ECMWF

Ensemble Prediction System (EPS) in 2008. Since the length of these sub-seasonal forecast is 32 days this system will be thereafter be referred to as the ECMWF monthly forecasting system and the ECMWF sub-seasonal forecasts will be referred to as monthly forecasts.

10 years ago, only a couple of operational centres were producing operational monthly forecasts. Now, the vast majority of the World Meteorological Organisation (WMO) Global Producing Centres are already or on the verge of producing operational sub-seasonal forecasts. A reason for this renewed interest for sub-seasonal forecasts is the perception that numerical models have improved and are producing more skilful monthly forecasts than in the past (e.g. Weaver et al 2011). The goal of the present paper is to evaluate if there has been indeed a positive trend in the ECMWF monthly forecast skill scores. Simmons and Hollingworth (2002) showed that the ECMWF medium-range forecast skill scores display an almost linear positive trend over the northern Extratropics, with a gain of about 1 day of predictability per decade. Do these improvements affect the sub-seasonal forecasts? The skill of the ECMWF monthly forecasting systems has been documented in Vitart (2004) and Weigel et al (2008). However, as it will be described in the following session, the numerical model used to produce these forecasts evolves constantly and therefore these studies offer just a ”snapshot” of the skill of the system at a given time: 2002 and 2003 in Vitart (2004) and 2006 in Weigel et al (2008). Therefore, the present paper will investigate if there is a trend in the evolution of the monthly forecast skill scores, and if the ECMWF monthly forecasting system is now more skilful than 10 years ago.

The evolution of the ECMWF monthly forecast system since 2002 will be described in section 2, along with a discussion of the methodology used in this paper to assess the monthly forecast skill scores. The following section will discuss the evolution of the skill scores of important sources of predictability at the sub-seasonal time range: the Madden Julian Oscillation (section 3), the North Atlantic Oscillation (Section 4) and sudden stratospheric warmings (Section 5). Section 6 will discuss the evolution of 2-metre temperature skill scores and section 7 will discuss the impact of changing horizontal and vertical resolution on the monthly forecast skill scores. Finally Section 8 will conclude this study, along with a discussion.

2 The ECMWF monthly forecasting system

As mentioned in the introduction, an ensemble of monthly forecasts (32-day integrations) has been run routinely at ECMWF since March 2002 and operationally since October 2004 (Vitart, 2004). The ECMWF monthly forecasting system includes a real-time component with a frequency of twice a week since October 2011 (the frequency was every 2 weeks in 2002 and once a week between October 2004 and October 2011). After a few days of model integrations, the model mean climate begins to be different from the analysis. No ”artificial” terms are introduced to try to remove or reduce the drift in the model, and no steps are taken to remove or reduce any imbalances in the coupled model initial state; the models are coupled together and integrated forward. The effect of the drift on the model calculations is estimated from integrations of the model in previous years (the re-forecast). The model drift is removed from the model solution during the post-processing. In the present system, the climatology (re-forecasts) is based on a 5-member ensemble 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 a number of past years. For instance, the first starting date of a real-time monthly forecast was 27 March 2002. The corresponding climatology was a 5-member ensemble starting on 27 March 1990, 27 March 1991,... 27 March 2001. The 5-member ensemble was thus integrated with 12 different starting dates, which represented a 60-member ensemble of re-forecasts. Therefore, the re-forecasts are created on the fly a couple of weeks before the corresponding real-time forecast. This strategy for re-forecasts 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 once for all.

The ECMWF monthly forecasting system has evolved since March 2002. The atmospheric component is the same version of the ECMWF atmospheric model Integrated Forecast System (IFS) as the one used to produce the ECMWF high-resolution operational forecasts and therefore has changed a few times each year. The model physics has changed numerous times during the period 2002-2012. From 2002 until February 2008, the ECMWF monthly forecasts were run separately from the ECMWF medium-range ensemble prediction system (EPS) and from the ECMWF seasonal forecasting system: a 51-member ensemble of 32-day integrations were produced at a T159 (about 1 degree resolution) and with 40 vertical levels coupled to an ocean model (the Hamburg Ocean Primitive Equation model, Wolff et al. 1997). In February 2008, the ECMWF monthly forecasting system has been merged with the EPS using a variable resolution: the first 10 days of integrations at high resolution are now uncoupled (the atmospheric component is forced by persisted SST anomalies), and the resolution is lowered after day 10 with the atmospheric component coupled to an ocean model (see Vitart et al., 2008 for more details). The vertical resolution has increased from 40 to 62 vertical levels in 2006 and the horizontal resolution of the monthly forecasts has increased in 2008 and in 2010.

The main goal of this paper is to evaluate how these changes in the ECMWF monthly forecast system (changes in model physics, horizontal and vertical resolution ...) have affected the monthly forecast skill scores. This could be done by scoring the monthly forecasts which have produced in real-time since 2002. However, a major issue with this methodology is that the monthly forecast skill scores are also strongly dependant on the large-scale circulation that was predominant during a season. For instance,the winter 2009-2010 was exceptionally predictable (e.g. Jung et al 2011). Low frequency variability associated to ENSO can also impact extended range forecast skill scores in the Tropics and Extratropics. This makes it difficult to identify trends in forecast skill scores from a time series of real-time forecast skill scores. Therefore, the present paper will use a different methodology: the re-forecasts covering the same years have been scored and compared. As shown in Figure 1, the number of re-forecast years has changed over the past years, but all the re-forecasts since 2002 have the period 1995-2001 in common. The starting days of the re-forecasts may vary from one year to another, but this should not have a significant impact on the skill scores averaged over a complete year or a season. The most important aspect of this methodology is that the scores are compared over 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 referred 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. The fact that the years go from April to March of the following year ensures a consistency in the model versions used for a complete winter and summer. Weigel et al (2008) scored the ECMWF re-forecast that were produced in 2006. The present paper will extend this study to all the re-forecasts that were produced from April 2002 until March 2012.

As mentioned above, an advantage of this methodology is that it ensures that all the re-forecasts cover the same seasons and years. There are however a few disadvantages: The ensemble size of the re-forecasts is just 5 members instead of 51 members for the real-time forecasts. This can be an issue when computing probabilistic skill scores which are often affected by ensemble size. Weigel et al (2006) faced the same issue when they scored the ECMWF re-forecasts produced in 2006 and used a correction of the probabilistic skill score which takes into account the ensemble size. It is worth noting that this is less an issue for the present study than for Weigel et al (2006) since the goal of the present study is not to give an assessment of the skill of the monthly forecast system, but to measure the evolution

March 2002	**T159L40** **Coupled to HOPE (1 degree)** **Every 2 weeks** **12 past years** **Use of ERA 40 + operational analysis for initialization**
October 2004	Every week instead of every 2 weeks
February 2006	Increased vertical resolution: T159L62
March 2008	Increased horizontal resolution : - T399L62 for day 0-10 uncoupled - T255L60 after day 10 coupled to HOPE (1 degree) 18 instead of 12 past years Use of ERA Interim instead of ERA40+operational analysis for initialization
January 2010	Increased horizontal resolution: - T639L62 for day 0-10 (uncoupled) - T319L62 after day 10 (coupled)
November 2011	Use of the NEMO ocean model instead of HOPE

Figure 1: Evolution of the main changes in the ECMWF monthly reforecasts from 2002 to 2011.

of the monthly forecast skill scores during the period 2002-2011. The model cycle may have changed more than once the period that has been used to compute the skill scores.This makes the attribution of the changes to a specific change in the model physics more difficult. An alternative would be to run a large set of re-forecasts covering the same period with the various versions of the IFS model, but this would be too expensive to do systematically and impossible to do for old versions of IFS which are no longer supported in the current ECMWF operating systems. Apart from the change from ERA 40 to ERA Interim in March 2008, all the re-forecasts have been initialised from the same dataset. Therefore this verification will not take into account possible improvements due to changes in the ECMWF data assimilation from 2002 to 2011. The following section will discuss the evolution of the skill scores for various aspects of the ECMWF monthly forecasts.

3 Madden Julian Oscillation

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). It has a significant impact on the Indian (Murakami 1976; Yasunari 1979) and Australian monsoon (Hendon and Liebmann 1990). It plays an active role in the onset and development of an El-Niño event (e.g. Kessler and McPhaden 1995) and has an impact on tropical cyclogenesis (e.g. Maloney and Hartmann 2000a; Mo 2000). It also impacts the extratropical weather (Ferranti et al 1990; Cassou, 2008). Therefore it is important for a monthly forecasting system to have skill not only in predicting the evolution of the MJO, but also in simulating the MJO teleconnections.

The goal of the present section is to evaluate the evolution of the skill of the ECMWF re-forecasts from 2002 to 2011 to simulate MJO events and its teleconnections in the Tropics and in the Northern Extratropics.

3.1 Forecast skill scores

The methodology for assessing the skill to predict the MJO follows Gottschalck et al (2009). The Wheeler and Hendon index (WHI, see Wheeler and Hendon 2004) has been applied to all the model hindcasts and to ERA interim (Simmons et al, 2007) over the period 1989-2008 to evaluate the skill of the monthly forecasting system to predict MJO events and to produce composites for different phases of the MJO. The WHI is calculated by projecting the forecasts or analysis on the two dominant combined EOFs of outgoing longwave radiation (OLR), zonal wind at 200 and 850 hPa averaged between 15N and 15S. The WHI has been applied to daily anomalies relative to the 1989-2008 climate instead of the absolute value of the field, in order to remove the impact of seasonal cycle. In addition, a 120-day running mean has been subtracted to remove the variability associated to ENSO. The positive (negative) phase of EOF2 describes suppressed (enhanced) convection over the Indian ocean and enhanced (suppressed) convection over the West Pacific. The positive (negative) phase of EOF1 describes enhanced (suppressed) convection over the Maritime Continent region. Analysis and forecasts are projected onto those two EOFs to describe the phase of the MJO in terms of two time series, PC1 and PC2. The two time series can be plotted as a succession of points in the PC1-PC2 phase space, in such a way that the MJO is described by a clockwise propagation in the phase space. The PC1-PC2 phase space can be divided into 8 sections representing a specific phase of the MJO (see for instance Figure 2 in Gottschalk et al, 2009). Phases 2 and 3 (negative EOF2) correspond to enhanced convection over the Indian ocean, phases 4 and 5 (positive EOF1) correspond to the MJO over the Maritime continent, phases 6 and 7 (positive EOF2) correspond to the MJO over the western Pacific and phases 8 and 1 (negative EOF1) correspond to the active phase of the MJO in the western Hemisphere.

To evaluate the skill of the monthly forecasting system to predict the MJO, a linear bivariate correlation (Lin et al 2008; Rashid et al. 2009) is performed between the PC1 and PC2 time series from the forecast ensemble-mean time series for different lead times and the PC1 and PC2 time series computed from ERA-Interim.

Figure 2 shows the evolution of the MJO bivariate correlation skill score from 2002 until 2011 between the ensemble mean re-forecasts and ERA Interim. In this figure, the three different lines show the forecast day in which the bivariate correlation reached 0.5, 0.6 and 0.8. If we consider MJO bivariate correlation of 0.6 as a limit of MJO predictability, the ECMWF monthly forecasting system displayed skill to predict the MJO up to about 15 days in 2002. In 2011, this predictability limit reaches 25 days, suggesting an averaged gain of about 1 day of predictability per year. The bivariate correlation of 0.5 is now reached beyond day 30 instead of day 22 ten years ago. For the bivariate correlation of 0.8, the gain has been of about 5 days of predictability over the 10-year period. Previous publications (e.g. Bechtold et al, 2008) documented significant improvements in the representation of the MJO in the ECMWF forecasting system with the introduction of a specific version of IFS referred to as cycle 32R3. Figure 2 shows that there was indeed some improvement in the MJO skill scores in 2008, which is the first year following the introduction of this version of IFS. However, Figure 2 also shows that the improvement of the MJO in the ECMWF model is not due to a single model change, but seems to be rather a continuous process, although the improvement is not completely linear. The evolution of the MJO skill scores is similar for winter and summer cases (not shown), but the improvement is more significant for cases when there is already an active MJO in the initial conditions in phases 2 or 3 (not shown).

The evolution of the amplitude error of the MJO, calculated from each individual ensemble member and then averaged does not display an improvement as regular as for the forecast skill scores. According to Figure 3, IFS produced a too weak MJO in the early years of the monthly forecasting system, by more than 30 % after 20 days of forecasts. There has been a clear improvement between 2006 and 2008, which correspond in particular to the changes in model physics introduced with cycle 32R3 (see Bechtold et al, 2008). In 2008, when Cy32R3 was used operationally, the MJO was even slightly too strong as discussed in Vitart and Molteni (2010). Since 2008, the amplitude of the MJO displays a trend towards weaker MJOs, with an amplitude in the recent years about 10 % weaker than in ERA Interim.

Another MJO diagnostic is the error in phase angle in the Wheeler and Hendon phase space diagram between the re-forecast ensemble mean and the reanalysis. Figure 4 shows the evolution of the phase error with time. A negative (positive) value of the angle error indicates an MJO propagation that is too slow (too fast). According to Figure 4, the MJO in the ECMWF monthly forecasts has almost always been too slow. However, there have been some significant improvements over the past 10 years, particularly in the longer forecast time range. In 2002, the model displayed on average an error of more than 10 degrees after 20 days of integrations. In recent years, the phase error has been reduced to less than 5 degrees, indicating that the MJO is still too slow as mentioned in Vitart and Molteni (2010), but faster than it used to be. This improvement is however not visible at the day 10 forecast range.

In summary, there have been very significant improvements in the representation of the MJO in the ECMWF monthly forecasts over the past 10 years. The MJO is stronger, faster and displays much higher forecast skill than 10 years ago. Since the MJO has a global influence at the sub-seasonal time scale (Waliser et al, 2011), the improvement in the representation of the MJO is likely to translate to a better representation of the MJO teleconnections and have a positive impact on the extratropical skill scores.

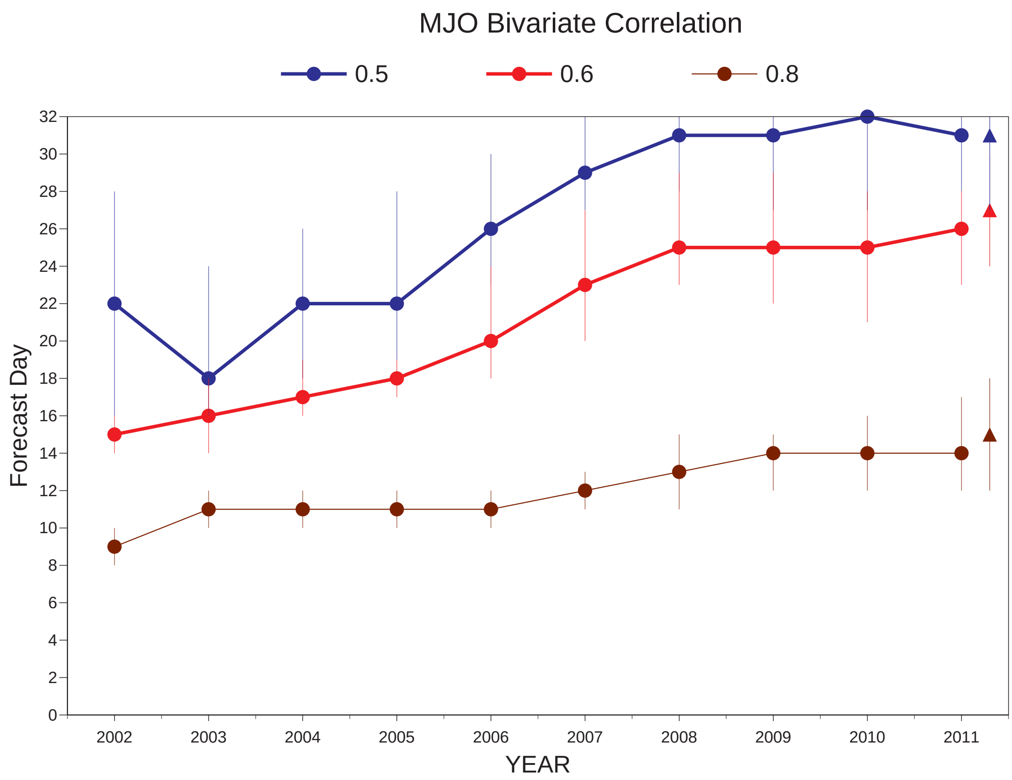

Figure 2: Evolution of the MJO skill scores (bivariate correlations applied to WHI) since 2002. The MJO skill scores have been computed on the ensemble mean of the ECMWF re-forecasts produced during a complete year. The blue, red and brown lines indicate respectively the day when the MJO bivariate correlation reaches 0.5, 0.6 and 0.8. The triangles show the skill scores obtained when rerunning the 2011 re-forecasts with the version of the IFS which was implemented operationally in June 2012 (Cycle 38r1) to give an indication of the MJO forecast skill scores expected in the coming year.

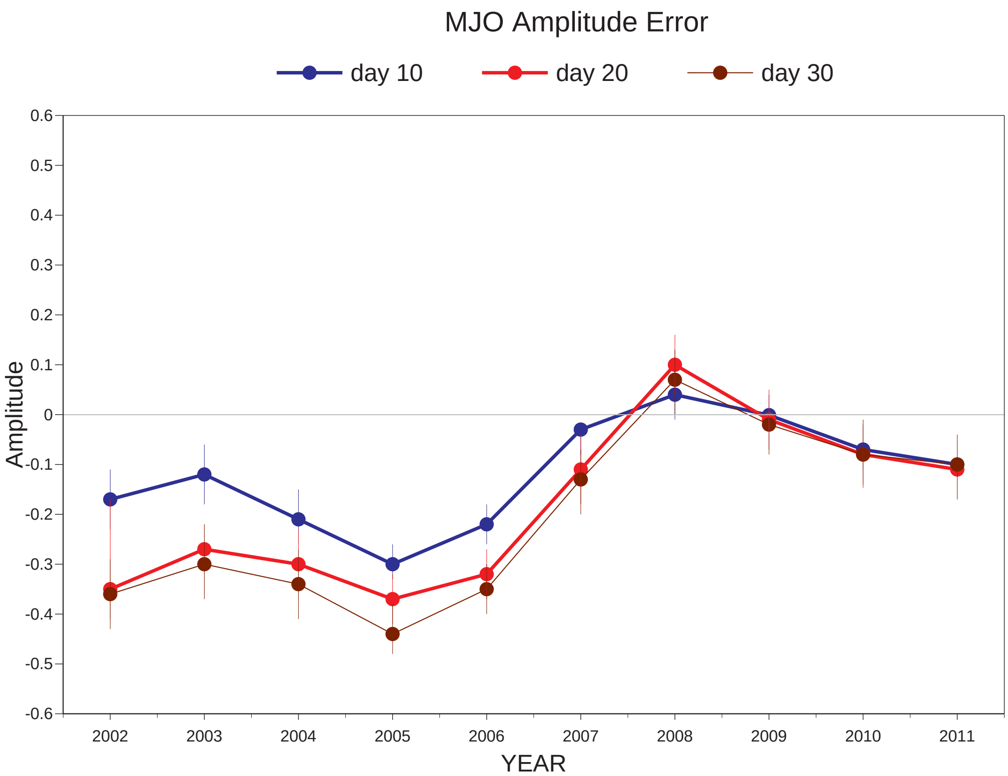

Figure 3: Same as Figure 2 but for the amplitude error of the ensemble mean of the re-forecasts relative to the mean MJO amplitude obtained from ERA Interim. Negative (positive) numbers indicate that the MJO simulated by IFS is weaker (stronger) than in the ECMWF reanalysis.

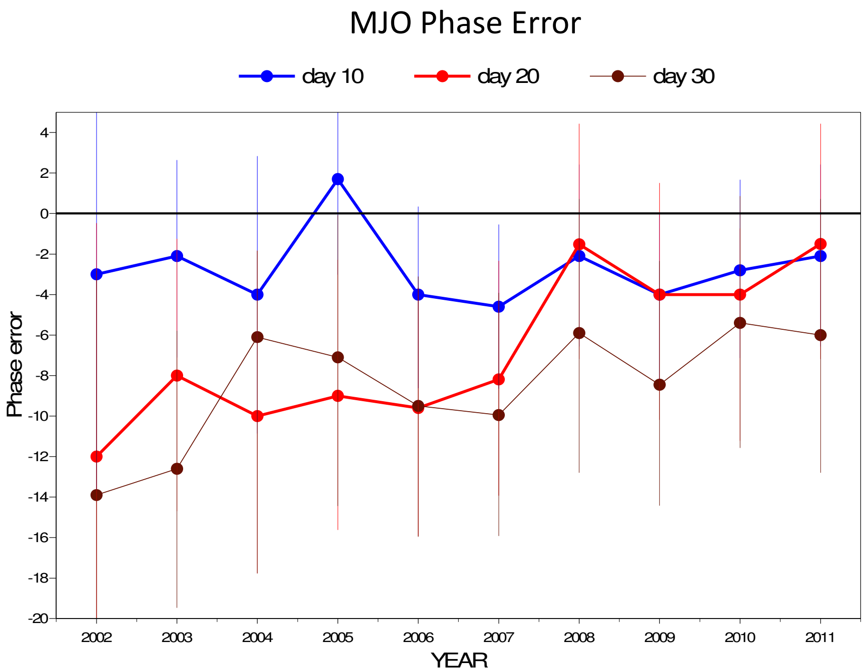

Figure 4: Same as Figure 2 but for the phase angle error of the ensemble mean of the re-forecasts relative to the mean MJO phase angle obtained from ERA Interim. Negative (positive) numbers indicate that the MJO simulated by IFS is slower (faster) than in the ECMWF reanalysis.

3.2 Extratropical teleconnections

Stronger MJOs in the model simulations are likely to lead to stronger teleconnections. Using reanalysis data covering the period 1974-2007, Cassou (2008) and Lin et al (2008) showed that the impact of the MJO on European weather is the strongest about 10 days after the MJO is in Phase 3 or Phase 6 (e.g. Fig. 3 of Cassou, 2008). The probability of a positive phase of the North Atlantic Oscillation (NAO) 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). The impact of the MJO on two other Euro-Atlantic weather regimes, the Atlantic Ridge and Scandinavian blocking, is much weaker. Vitart and Molteni (2010) showed a set of ECMWF re-forecasts using the version of IFS known as cycle 32R3 displayed a realistic MJO teleconnections over the northern Extratropics, consistent with the impact Cassou (2008) and Lin (2008) found in re-analysis data. This section will focus on the 500 hPa geopotential height composites 10 days after an MJO in Phase 3 with an amplitude larger than a standard deviation to evaluate if the MJO teleconnections on the northern and southern Extratropics have improved over the past 10 years by comparing the re-forecasts produced each year from 2002 until 2012 with ERA Interim. Only the re-forecasts covering the extended winter season are considered (from October to March for the Northern Extratropics and April to September for the Southern Extratropics).

According to Figure 5, the MJO teleconnections (10 days after an MJO in Phase 3) are significantly more realistic over the northern Extratropics in 2011 than in 2002. The amplitude of the teleconnections is much larger in 2011, most especially in the Euro-Atlantic region, where the re-forecasts produced in 2012 simulate a stronger positive NAO anomaly than in 2002. However, as already mentioned in Vitart and Molteni (2010), the impact of the MJO on the NAO is still underestimated in the 2011 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. Figure 5 suggests that a similar improvement took place over the southern Extratropics. As for the northern Extratropics, the MJO displays stronger teleconnections over the southern Extratropics in the re-forecasts produced in 2011 than in the re-forecasts produced in 2002, although the most recent version of the ECMWF forecasting system still underestimates the amplitude of the teleconnections compared to ERA Interim. The teleconnection patterns look also more realistic in 2011 than in 2002, particularly in the South Pacific basin. The same conclusions are valid for the composites of 500 hPa geopotential height 10 days after an MJO in Phase 6 (not shown), except that the composites 10 days after an MJO in phase 6 do not display a stronger 500 hPa geopotential anomaly over the northern Pacific compared to ERA Interim.

The main improvement in the representation of the MJO teleconnections happened in 2008, when IFS cycle 38r3 was introduced. This coincides with the time when the MJO became more intense in the re-forecasts (see Fig. 3). However, the MJO teleconnections have also become slightly more realistic since 2008, despite a slight decline of the amplitude of the MJO, particularly over the Euro-Atlantic sector, where the 2011 re-forecasts display a stronger NAO signal than in 2010 (not shown).

3.3 Modulation of tropical cyclones by the MJO

The impact of the MJO on tropical cyclone activity has been documented in numerous observational studies for the western North Pacific (Nakazawa 1988, Liebmann et al 1994), the eastern North Pacific (Molinari et al, 1997; Maloney and Hartmann 2000a), the Gulf of Mexico (Maloney and Hartmann 2000b, Mo, 2002), the South Indian Ocean (Bessafi and Wheeler, 2006; Ho et al 2006) and the Australian

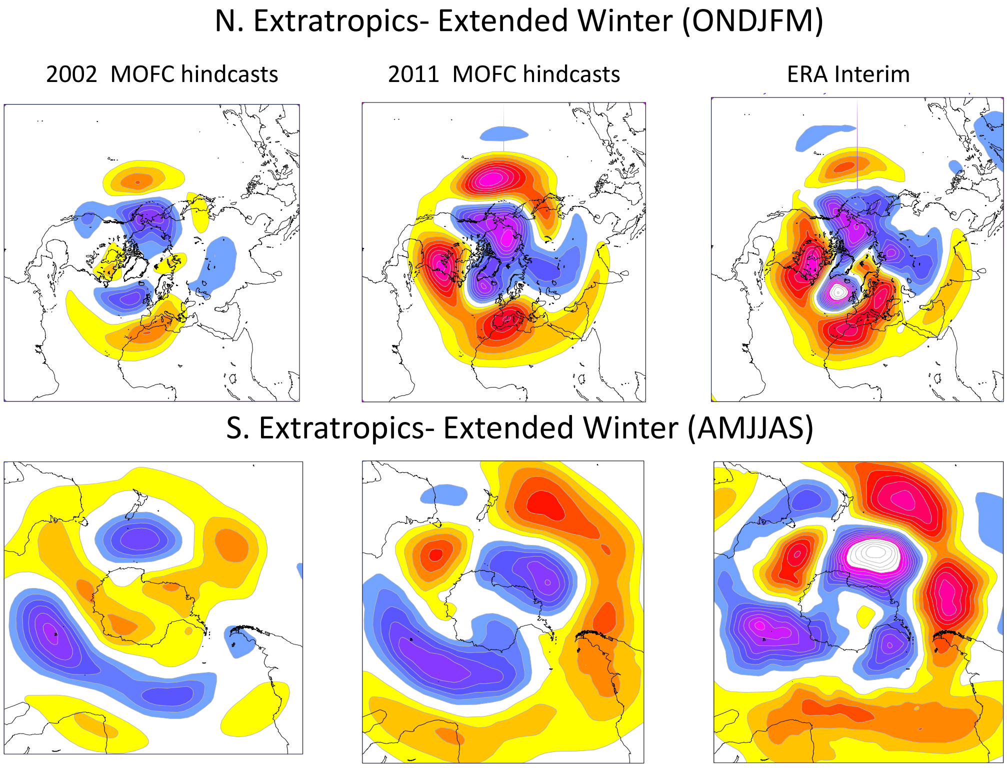

Figure 5: MJO Phase 3 10-day lagged composites of 500 hPa geopotential height anomaly over the northern Extratropics (top panels) and southern Extratropics (bottom panels) for all the October to April re-forecasts that were produced in 2002 (left panel), in 2011 (middle panel) and ERA Interim (right panel). 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.

region (Hall et all 2001).Vitart (2009) showed that a large set of ECMWF re-forecasts using the IFS cycle 32 r3 (operational in 2008) was able to simulate the modulation of tropical cyclone activity by the Madden Julian Oscillation over all the ocean basins. The impact of the MJO on observed tropical storms comes mainly from mid-level relative humidity and low-level absolute vorticity (Camargo et al 2009).

The model tropical cyclones are tracked using the methodology described in Vitart et al (1997), revised in Vitart et al (2003). 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. The climatology of tropical storms has improved over the past 10 years in the ECMWF re-forecasts. Figure 6 shows, for instance, than in 2002 the ECMWF forecasting system produced about as many storms in the South Pacific than in the South Indian Ocean, and too few near the north western coast of Australia. The geographical distribution of tropical storms over the southern Hemisphere is more realistic in the re-forecasts of 2011. This is also the case for the northern Hemisphere (not shown), where the re-forecasts of 2002 tended to produce tropical storm genesis over the eastern part of the North Atlantic basin and South of 25 North, whereas the re-forecasts of 2011 display a realistic tropical storm climatology over the North Atlantic. When there is an MJO in phase 2 or 3 in the model, the re-forecasts of 2002 and 2011 display more tropical cyclone activity over the Indian Ocean and less over the South Pacific and near the Maritime Continent (right panels in Fig. 6), which is consistent with observational studies and with Vitart (2010). Interestingly, the re-forecasts of 2002, despite simulating too weak MJOs, display a modulation of tropical cyclone activity by the MJO similar to the modulation of the MJO in the 2011 re-forecasts. The differences between the right and left panels in Figures are of similar amplitude. Therefore this suggests that it is the presence of an MJO more than its amplitude that affects the tropical cyclone activity in the model, whereas the previous section suggested that the northern Hemisphere teleconnections were affected by the amplitude of the MJO.

4 North Atlantic Oscillation

The prediction of the North Atlantic Oscillation (NAO) is of particular importance for the prediction of European weather. The positive and negative NAO are amongst the most frequent weather regimes in the Euro-Atlantic region. An NAO index has been constructed by projecting the daily 500 hPa height anomalies over the Northern Hemisphere onto a pre-defined NAO pattern. The NAO pattern was defined as the first leading mode of Empirical Orthogonal Function (EOF) applied to the NCEP re-analysis of monthly mean 500mb height during the 1950-2000 period. NAO skill scores have been produced for each year from 2002 until 2011 by applying the NAO index to the re-forecasts and to ERA interim and computing the linear correlation between the ensemble means of the re-forecasts and ERA Interim. In this section only extended winter cases (from October to March) are considered. Figure 7 shows the time series of the NAO skill scores since 2002, and is equivalent to Figure 2, but for the NAO instead of the MJO. Figure 7 shows that there has been some improvements in the prediction of the daily values of the NAO with a gain is of about 4 days of predictability 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 improvement in the NAO skill scores seems to be almost linear.

Monthly forecasts products at ECMWF are mainly expressed in terms of weekly means (see Vitart 2004 for more details) since at the extended range the model has generally more skill in predicting weekly anomalies than daily values. The 7-day periods correspond to forecast days 5-11, day 12-18, day 19-25 and day 26-32. They have been chosen that way so that they correspond to Monday to Sunday calendar weeks for the monthly forecast starting on Thursday 00 UTC. Since November 2011, monthly forecasts

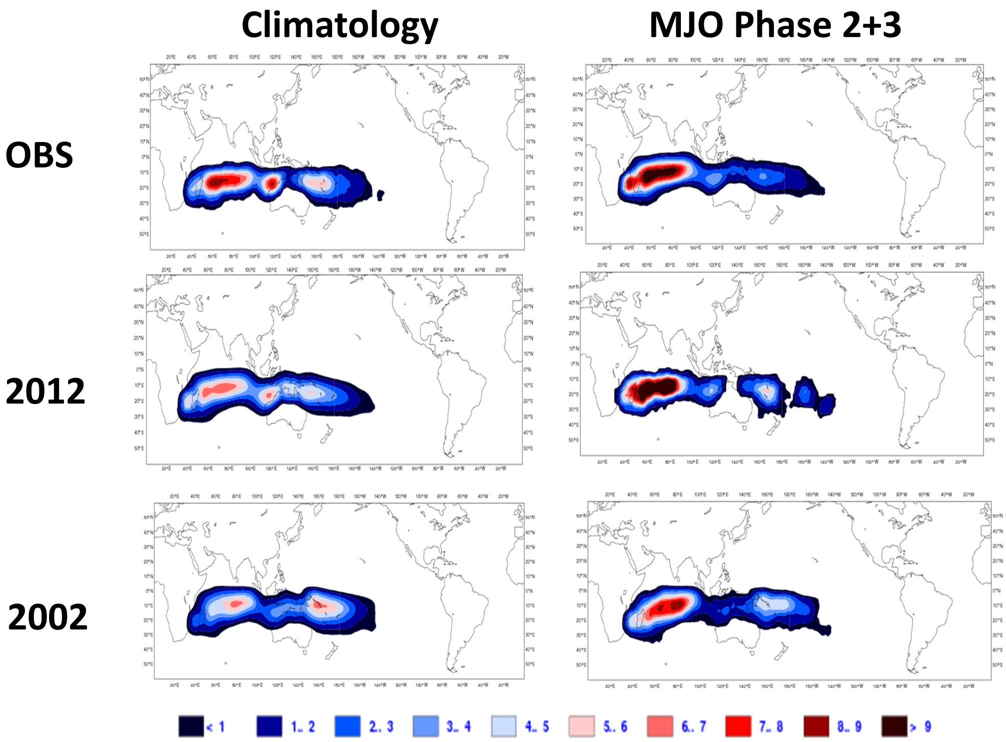

Figure 6: Tropical storm density during the period October to March 1995-2001 (left panels) and when there is an MJO in Phase 2 or 3 (right panels). The top panels show observations (from Joint Typhoon Warning Center), the middle panel shows the tropical storm densities from the re-forecasts produced in 2011, and the bottom panel shows the tropical storm densities from the re-forecasts produced in 2002. The tropical storm density is calculated by computing the number of tropical storms passing within 500 kilometres and then normalise that number by the total number of tropical storms over the whole basin.

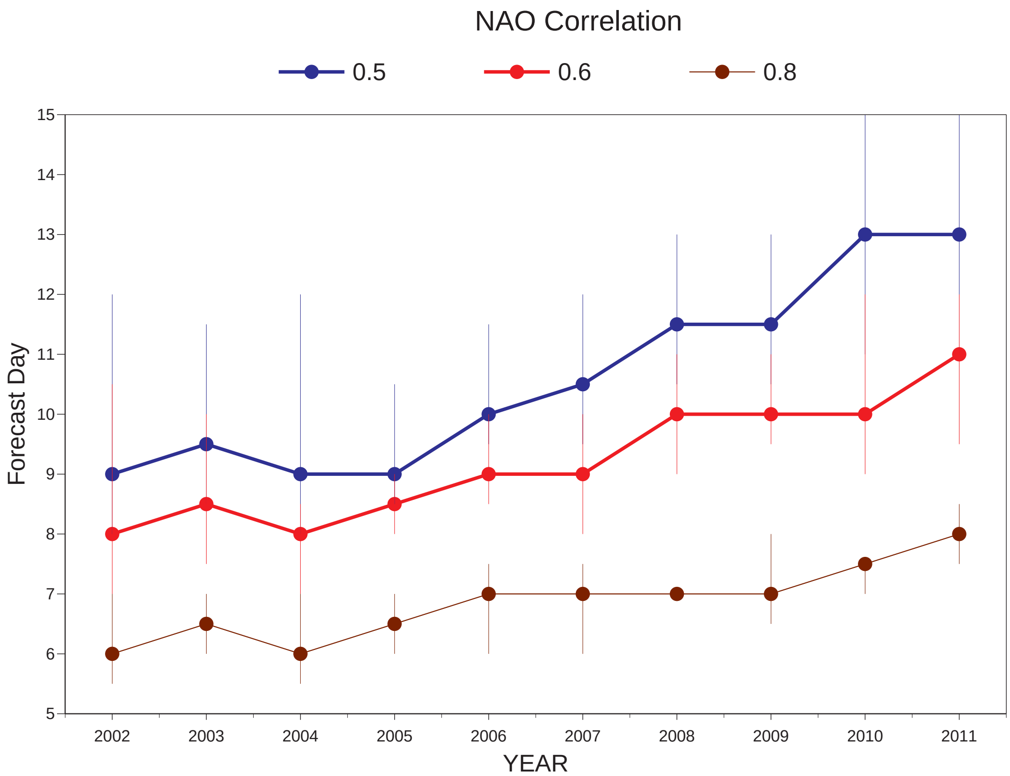

Figure 7: Evolution of daily NAO skill scores since 2002. The daily NAO skill scores (correlations applied to the NAO index) have been computed on the ensemble mean of the ECMWF re-forecasts produced from October to March 1995-2001 and ERA Interim. The blue, red and brown lines indicates the day when the NAO index correlation reaches respectively 0.5, 0.6 and 0.8.

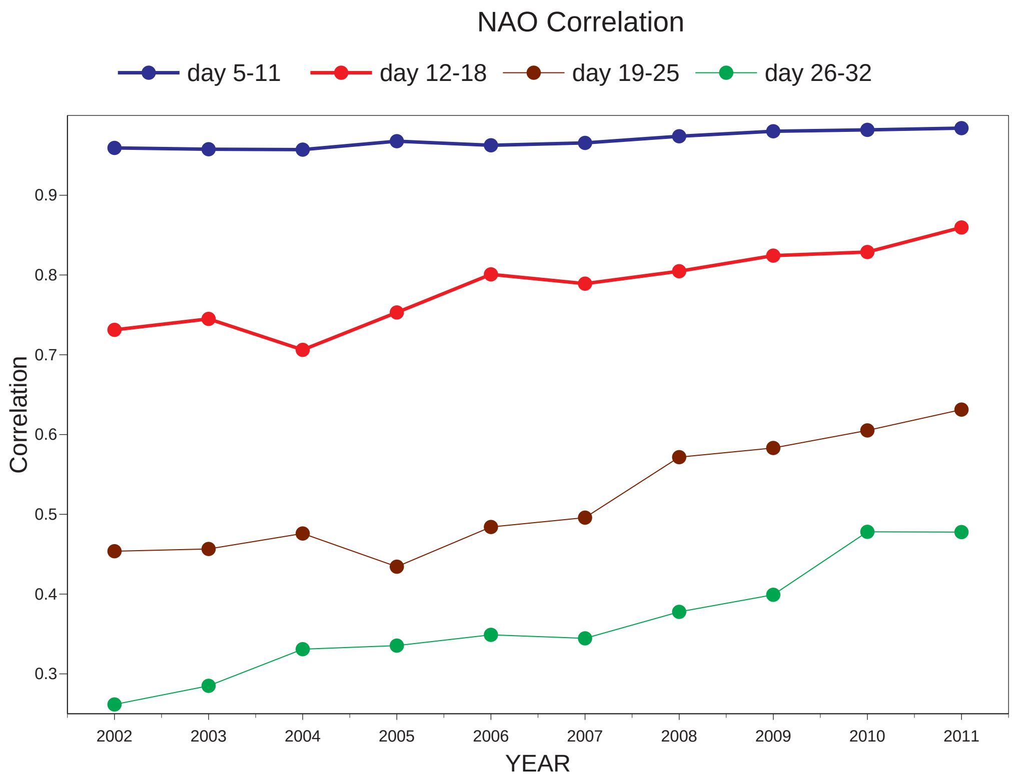

Figure 8: Evolution of the NAO skill score applied to weekly means (day 5-11, 12-18, 19-25 and 26-32) since 2002. The NAO skill score is the correlation between the NOA index computed from the ensemble weekly means and the NAO index computed from ERA Interim. The blue, red, brown and green lines represent the NAO scores of respectively day 5-11, day 12-18, day 19-25 and day 26-32.

are also produced on Mondays with weekly mean periods corresponding to day 1-7, day 8-14, day 15-21 day 22-28, but these weekly periods will not be discussed in the present paper. Figure 8 shows the evolution of the NAO skill scores applied to the 4 weekly periods mentioned above. This figure confirms the improvement in the prediction of the NAO over the past 10 years for all the weekly periods. For instance, the NAO skill score for the weekly period day 19-25 which was lower than 0.5 in 2002, exceeds 0.6 in 2011. Interestingly the NAO skill score of day 26-32 (last week of the ECMWF monthly forecasts) have reached the same level of skill as the forecasts of day 19-25 10 years ago, suggesting a gain of about a week of predictability for the NAO prediction.

As mentioned in the previous section, the MJO teleconnections over the northern Extratropics project into a positive or negative NAO depending on the phase of the MJO (e.g. Cassou 2008, Lin et al. 2008, Vitart and Molteni 2010). Therefore part of the improvements in the NAO prediction may originate from improvements in the prediction of the MJO. To determine if this is the case, the NAO skill scores have been computed by considering only the cases when there is an MJO in the initial conditions (solid line in Figure 9) and the cases when there is no MJO in the initial conditions (dashed line in Figure 9). Figure 9 shows the NAO skill scores for the forecast range day 19-25. According to Figure 9, the prediction of the NAO has improved even when there was no MJO in the initial conditions, suggesting that part of the improvement in the prediction of the NAO were not related to the MJO. However, the improvements in

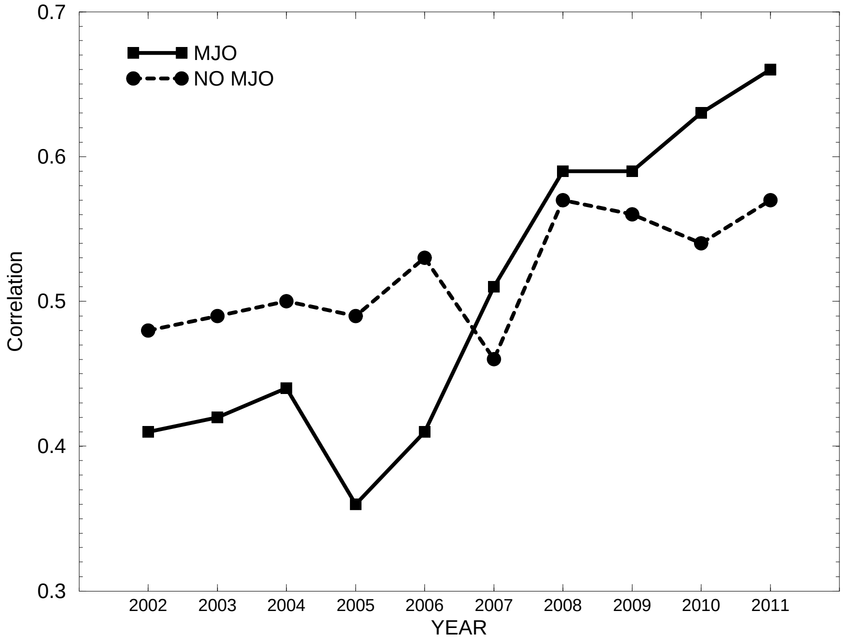

Figure 9: Evolution of the NAO skill scores (correlation between the ensemble mean NAO index and the NAO index computed from ERA Interim) for the period day 19-25 for all the cases when there is an MJO in the initial conditions (amplitude of the MJO index larger than 1 independently of the phase) and when there is no MJO in the initial conditions (amplitude of the MJO index less than 1).

the NAO skill scores have been much stronger when there was an MJO in the initial conditions than when there was no MJO in the initial conditions, suggesting that the major part of the improvements in NAO skill scores displayed in Figures 7 and 8 come from improvements in the prediction of the MJO. Figure 9 also suggests that in the first years of the ECMWF monthly forecasts, the presence of an MJO had a negative impact on the NAO skill scores, with lower NAO skill scores when there was an MJO in the initial conditions. However, since 2008, the presence of an MJO in the initial conditions has a positive impact on the NAO skill scores. This can be explained by the fact that the MJO amplitude was too weak before 2007, making it difficult for the ECMWF model to reproduce correctly its teleconnections in the Extratropics as shown in Figure 5. Since 2008, the impact of the MJO on the Extratropics is much more realistic and the ECMWF model can now reproduce a realistic impact of the MJO on the NAO.

5 Sudden Stratospheric Warmings

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 a 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).

In this study, the difference between the temperature at 50 hPa over the North Pole and the temperature at 50 hPa averaged around 60 North is used as an index for SSWs. Fields at 50 HPa are archived only since 2004. Therefore this section will consider just the re-forecasts that were produced after 2004. Figure 10 shows that the ECMWF forecasting system displays skill to predict the SSWs for up to about 18 days (correlation of 0.6). The skill of the ECMWF monthly forecasts to predict SSWs has shown little improvement since 2002 (Fig. 10). The only noticeable improvement in forecasting skill 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. Since 2006 the vertical resolution of the ECMWF monthly forecasts has remained the same. Re-forecast experiments which have been performed with a vertical resolution of 91 vertical levels with a top level at 0.01 hPa instead of 62 vertical levels display higher skill to predict the SSW index (Fig. 11). This confirms that vertical resolution, particularly in the stratosphere (the 91 and 62 vertical resolutions were identical in the troposphere) has 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 12 shows the lag correlation between the SSW and NAO indices. This figure indicates that the lag correlation becomes more negative 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 12 shows that the amplitude of the lag correlation diminishes 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 a very spell 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.

6 2-metre temperature anomalies over the Northern Extratropics

Weigel et al (2008) provided a fully probabilistic evaluation of the ECMWF monthly re-forecasts that were produced in 2006 for weekly averaged forecasts of 2-metre temperature. In this publication, the verification was based on a modified version of the ranked probability skill score (RPSS; Epstein 1969; Murphy 1969,1971). The classical RPSS is a squared measure comparing the cumulative probabilities of categorical forecast and observation vectors relative to a climatological forecast strategy. A detailed description of the RPSS is provided in Wilks (2006). An advantage of the RPSS is that it is sensitive to distance in that a forecast is increasingly penalised the more its cumulative probability differ from the actual outcome. However, a big caveat of the RPSS is its strong negative bias for small ensemble size (e.g. Buizza and Palmer 1998). Muller et al (2005) and Weigel et al (2007) have derived a debiased version of the RPSS, the so-called discrete ranked probability skill score which contains a corrective term function of the number of categories used to define the probabilities (terciles in the present studies) and the ensemble size (5 for the ECMWF re-forecasts). Applying the discrete RPSS to the ECMWF

Figure 10: Evolution of the SSW skill scores since 2002. The daily SSW skill scores (correlations applied to the SSW index) 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.

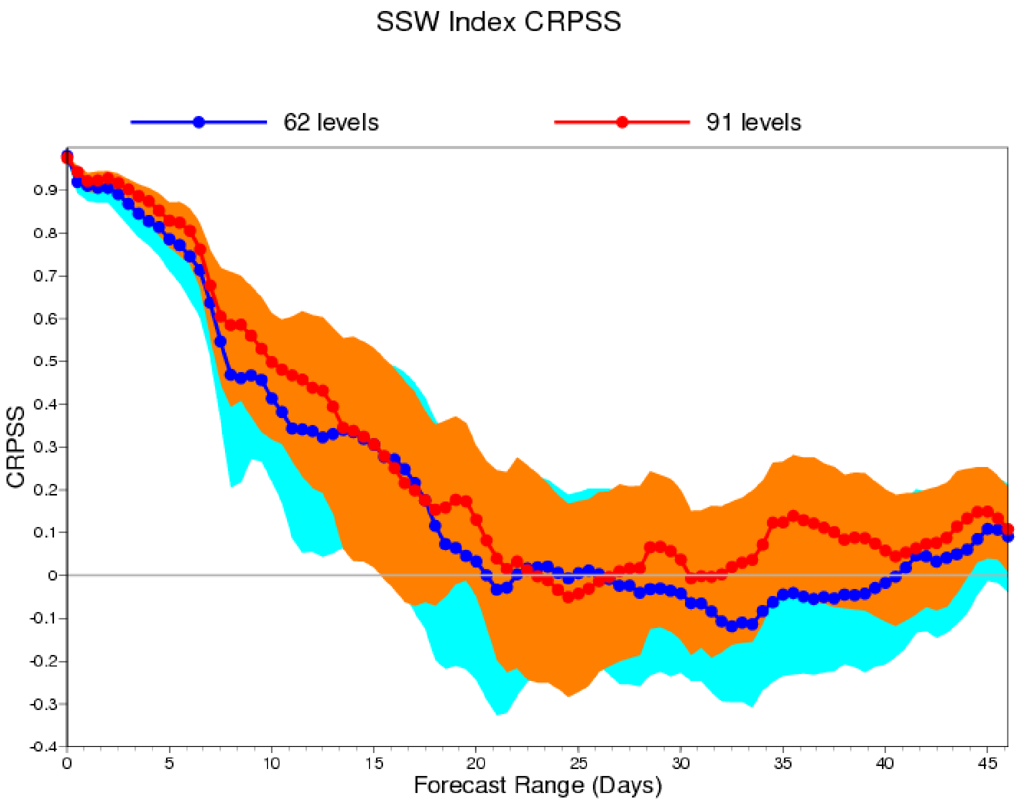

Figure 11: Ranked probability skill scores of the SSW index computed from 80 15-member 45-day reforecasts starting the 1st February, May, August and November 1989 to 2008 using IFS cycle 36r4. The red curve shows the RPSS scores obtained when using 91 vertical levels and the blue curve shows the RPSS skill scores obtained with the 62 vertical levels control experiment.

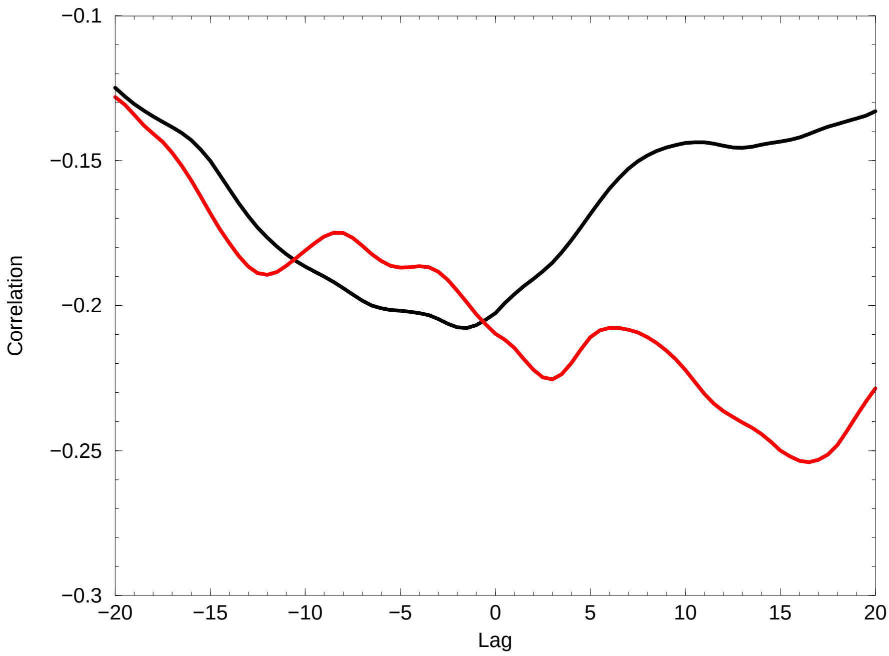

Figure 12: Lag correlation between the NAO and SSW index as a function of days preceeding (negative x-axis) or following (positive x-axis) a SSW. The black line shows the lag correlation obtained from all the re-forecasts produced between October 2011 and March 2012 and covering the years 1995 to 2001 and for each ensemble member separately. The red line shows the corresponding verification using ERA Interim.

re-forecasts produced in 2006, Weigel et al (2008) found that the ECMWF monthly forecasts of 2-metre temperature anomalies generally outperform persistence and climatology, but that it displayed very little skill over the northern Extratropics after day 18.

In this section, the discrete RPSS has been applied to all the re-forecasts of 2-metre temperature anomalies that were produced since 2002. Figure 13 displays the evolution of the discrete RPSS of 2-metre temperature anomalies since 2002 for the weekly periods day 12-18, day 19-25 and day 26-32. Figure 13 shows that there is a significant drop in the probabilistic score between day 12-18 and day 19-25, but the monthly forecasts still display better skill than climatology (positive RPSS) as mentioned in Vitart (2004) and Weigel et al.(2008). This figure also suggests that there have been improvements in the RPSS scores of 2-metre temperature anomaly re-forecasts over the northern Extratropics for three time ranges (day 12-18, day 19-25 and day 26-32) since 2002. The values of the discrete RPSS for day 16-32, although still very low, are now close to the values for the previous week (day 19-25) re-forecasts that were produced 10 years ago. The skill scores of day 19-25 have also improved in time almost linearly and get close to the skill scores of day 12-18 in the early years of the ECMWF monthly forecasts. This result confirms that the ECMWF monthly forecasts have improved over the past 10 years. The results presented in Figure 13 are for the whole year, but a similar figure for the extended winter season indicates a similar result (not shown).

7 Impact of model resolution

The previous sections documented that the skill of the ECMWF monthly forecasts have improved significantly over the past 10 years. Except for the prediction of stratospheric sudden warmings, the improvement has been regular and cannot be fully attributed to a single change in the forecasting system. A large fraction of the improvements is likely to come from changes to the model physics, like the introduction of Cycle 32r3 end of 2007 (Bechtold, 2010). Since a new version of IFS involve generally more than changes in the physics it is often difficult to determine which exact change in the physics parameterisation had an impact. For instance, Hirons et al (2012) showed the results of sensitivity experiments that have been run to determine which specific change in the physics of IFS that was introduced in CY32R3 was responsible for the improved MJO prediction. It is beyond the scope of this paper to determine which specific changes in model physics are responsible for the improvement of the ECMWF monthly forecasts. However, a possible source of improvement for the ECMWF monthly forecasts could be the changes in model configuration, in particular the increases of vertical or horizontal model resolutions. To determine if the changes in model configuration had an impact on the monthly forecast skill scores, an experiment has been set up which reproduced all the re-forecasts that were produced to calibrate the operational real-time forecasts from October 2011 until April 2012 and covering the past 18 years with the same version of IFS as the one used operationally to produced the monthly forecasts from October 2011 until April 2012, but with the model configuration that was used in 2002: resolution of TL159L40 (about 1 degree resolution) coupled from day 0 (instead of T639L62 uncoupled up to day 10 and T319L62 coupled to an ocean model afterwards). This experiment will be thereafter referred to as OLDMOFC, the operational re-forecast of 2011 will be referred to as OPER.

The skill to predict the MJO propagation is very close, but slightly higher in OLDMOFC than in the operational re-forecasts of 2011 (not shown). This slight improvement is most likely due to the fact that IFS is coupled to an ocean model from day 0 in OLDMOFC, instead of day 10 in OPER (Vitart et al, 2008). However the amplitude of the MJO is reduced by about 5% in OLDMOFC compared to OPER after 10 days of model integrations. This difference in the amplitude of the MJO is statistically significant. A sensitivity experiment similar to OLDMOFC has been performed, but with resolution of

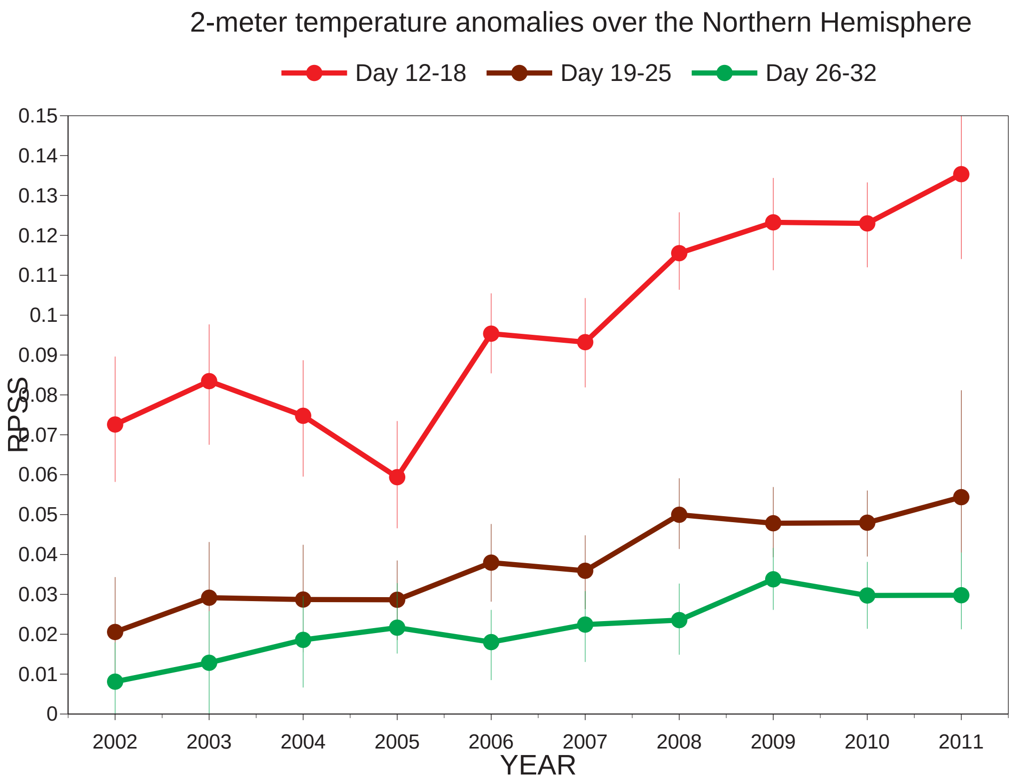

Figure 13: Evolution of the discrete ranked probability skill score (RPSS) of 2-metre temperature weekly mean anomalies over the northern Extratropics (North of 30N) since 2002. Only land points have been scored. The RPSS has been computed from terciles and for all the ECMWF re-forecasts covering all seasons. The red line shows the RPSS of day 12-18, the brown line represents the RPSS of day 19-25 and the green line the RPSS of day 26-32.

T159L62. This new experiment does not show a difference in MJO amplitude compared to OPER, which suggests that it is the change of vertical resolution that is responsible for the change in MJO amplitude. This result is consistent with Figure 3 which shows an increase in the amplitude of the MJO in 2006 when the 62 vertical level resolution was introduced. These results also suggest that the increase of horizontal resolution over the past 10 years had little impact on the MJO prediction.

Table 1 shows the discrete ranked probability skill scores of weekly mean 2-metre temperature anomalies over the northern Extratropics in OPER and OLDMOFC. The probabilistic skill scores are slightly higher in OPER than in OLDMOFC, particularly in the first 2 weeks, where the difference is statistically significant. However, the difference of skill scores between OPER and OLDMOFC is small at all time ranges compared to the difference of skill scores between 2002 and 2011, suggesting that the changes in model resolutions since 2002 explain only a small fraction of the improvements since 2002. Therefore, the numerous changes in model physics since 2002 are likely to be the main reason behind the improvements in monthly forecast skill scores at ECMWF.

	Day 5-11	Day 12-18	Day 19-25	Day 26-32
OPER	0.48 (0.01)	0.19 (0.02)	0.08 (0.01)	0.04 (0.01)
OLMOFC	0.44 (0.01)	0.16 (0.02)	0.07 (0.01)	0.038 (0.01)

Table 1: Discrete ranked probability skill score of 2-metre temperature anomalies over the northern Extratropics (North of 30N) computed over the weekly periods: day 5-11, day 12-18, day 19-25 and day 26-32 for OPER and OLDMOFC. The 5-member re-forecasts cover the period October to April 1993-2010. The numbers in parenthesis indicate 1 standard deviation.

8 Conclusion

Monthly forecasts are produced operationally at ECMWF since 2002. This study has shown that the skill of the ECMWF monthly forecasts has improved over the past 10 years. The improvements in the skill scores are particularly high for the prediction of the MJO which is an important source of predictability at the sub-seasonal time scale. Over the northern Extratropics, the prediction of 2-metre temperature anomalies has also increased over the past 10 years, particularly for day 12-18 where the CRPSS scores have almost doubled over the past 10 years. The skill for day 26-32 has now reached the forecast skill at day 19-25 10 years ago. The skill for day 19-25 is getting closer to the skill for day 12-18 10 years ago. Similar improvements are visible in the upper-air fields, such as the NAO. A large part of the improvement in the NAO skill scores seem to originate from improvements in the prediction of the MJO. However, the prediction of the sudden stratospheric warmings does not display any significant improvement since 2002, except in 2006 when the vertical resolution of the forecasting system has increased. The improvements in the monthly re-forecast skill scores reported in this study are likely to be an underestimation 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.

An experiment with the original set-up of the ECMWF monthly forecasting system and the current version of IFS suggests that most of the improvements of monthly forecast skill scores are due to changes in the model physic parameterisations. The various changes in model physics over the past 10 years were not designed to improve specifically sub-seasonal predictions, but were introduced to reduce model systematic errors and improve the physics parameterisations. It is only recently that monthly forecast skill scores are used in the process of validating a new operational version of IFS. Therefore, it seems that improvements in the ECMWF medium-range forecasts have led to better sub-seasonal forecasts. The results also suggest that the increased vertical resolution introduced in 2006 had a positive impact on the sudden stratospheric warming skill scores.

Further improvements to the ECMWF ensemble prediction systems are planned for the coming years and should help maintain the increase in sub-seasonal forecast skill. These improvements include coupling the atmospheric model to the ocean model from day 0 instead of from day 10, include a sea-ice model in the forecasting system instead of persisting sea-ice, perturb land-surface initial conditions, use of a high-resolution ocean model in addition to further changes in model parameterisation and horizontal and vertical resolution. Improvements in the stratosphere-troposphere interactions are also likely to lead to improved monthly forecasts.

It would be interesting to compare the ECMWF subseasonal prediction skill scores with other operational centres and determine if similar improvements have been observed in other operational centres, but sub-seasonal forecasts are currently not exchanged between operational centres. A new joint World Weather Research Program (WWRP) and World Climate Research Program project called the sub-seasonal to seasonal prediction project (S2S) has been set up for a period of 5 years starting in 2013. One of its goals will be to create a database of sub-seasonal to seasonal forecasts from operational centres. This project will allow to monitor the skill of various forecasting systems, and also investigate the usefulness of multi-model sub-seasonal forecasts.

REFERENCES

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

Brankovic, C., F. Molteni, T. N. Palmer, and U. Cubasch, 1988: Extended range ensemble forecasting at ECMWF. *Proc. ECMWF Workshop on Predictability in the Medium and Extended Range*, Reading, United Kingdom, ECMWF, 4587.

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.

Boville B. A., 1984. The influence of the polar night jet on the tropospheric circulation in a GCM. *J. Atmos. Sci.*, **41**. 11321142.

Buizza, R. and T.N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2508-2518.

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

Déqué, M., and J. F. Royer, 1992: The skill of extended-range extratropical winter dynamical forecasts. *J. Climate*, **5**, 13461356.

Epstein, E.S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985-987.

Ferranti, L., T. N. Palmer, F. Molteni and E. Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation and its impact on medium and extended range prediction. *J. Atm. Sci.*, **125**, 2177-2199.

Gottschalk J., M. Wheeler, K. Weickmann, F. Vitart, N. Savage, H. Hendon, H. Lin, M. Flatau, D. Waliser, K. Sperber, W. Higgins and A. Vintzileos, 2009: Establishing and Assessing Operational Model MJO Forecasts: A Project of the CLIVAR Madden-Julian Oscillation Working Group. *Bull. Am. Meteor. Soc.*, **91**, 1247-1258.

Hall, J. D., A. J. Matthews, and D. J. Karoli, 2001: The modulation of tropical cyclone activity in the Australian region by the Madden-Julian Oscillation. *Mon. Wea. Rev.*, **129**, 2970-2982.

Ho, C.-H., J.-H. Kim, J.-H. Jeong, and H.-S. Kim., 2006: Variation of tropical cyclone activity in the South Indian Ocean: El Niño-Southern Oscillation and Madden-Julian Oscillation effects. *J. Geophys. Res.*, **111**, D22191.

Hendon, H. H. and B. Liebmann, 1990: A composite study of onset of Australian summer monsoon. J. Atmos. Sci., **48**, 2909-2923.

Jeong J.-H., H. W. Linderholm, S.-H. Woo, C. Folland, B.-M. Kim, S.-J. Kim, D. Chen, 2012: Impacts of Snow Initialization on Subseasonal Forecasts of Surface Air Temperature for the Cold Season. *J. climate*, in press. doi: http://dx.doi.org/10.1175/JCLI-D-12-00159.1

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

Klotzbach, P. J., and W. M. Gray, 2003: Forecasting September Atlantic basin tropical cyclone activity. *Wea. and Forecasting*, **18**, 1109-1128.

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, D.M. Lawrence, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W.J. Merryfield, S. Seneviratne, T. Stanelle, B.J.J.M. van den Hurk, F. Vitart and E.F. Wood, 2010: The contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophys. Res. Lett., 37, L02402,doi:10.1029/2009GL041677.

Lin. H., G. Brunet and J. Derome, 2004: An Observed Connection between the North Atlantic Oscillation and the MaddenJulian Oscillation *J. Climate*, **22**, 364-380.

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, 2000a: Modulation of Eastern North Pacific Hurricanes by the Madden-Julian Oscillation. *J. Climate*, **13**, 1451-1460.

Maloney, E. D. and D. L. Hartmann, 2000b: Modulation of hurricane activity in the Gulf of Mexico by the Madden-Julian Oscillation, *Science*, **287**, 2002-2004.

Miyakoda, K., T. Gordon, R. Caverly, W. Stern, J. Sirutis, and W. Bourke, 1983: Simulation of a blocking event in January 1977. *Mon. Wea. Rev.*, **111**, 846869.

Miyakoda, K., J. Sirutis, and J. Ploshay, 1986: One month forecast experiments without anomaly boundary forcings. *Mon. Wea. Rev.*, **114**, 23632401.

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

Molinari, J., S. Knight, M. Dickinson, D. Vollaro, and S. Skubis, 1997: Potential vorticity, easterly waves and eastern Pacific tropical cyclogenesis. *Mon. Wea. Rev.*, **125**, 2699-2708.

Molteni, F., U. Cubasch, and S. Tibaldi, 1986: 30- and 60-day forecast experiments with the ECMWF spectral models. *Proc. ECMWF Workshop on Predictability in the Medium and Extended Range*, Reading, United Kingdom, ECMWF, 51107.

Muller, 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.

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

Murphy, A.H., 1969: On the ranked probability skill score. *J. Appl. Meteor.*, **8**, 988-989.

Murphy, A.H., 1971: A note on the ranked probability skill score. *J. Appl. Meteor*, **10**, 155-156.

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.

Owen, J. A., and T. N. Palmer, 1987: The impact of El-Niño on an ensemble of extended-range forecasts. *Mon. Wea. Rev.*, **115**, 21032117.

Palmer, T. N., C. Brankovic, F. Molteni, and S. Tibaldi, 1990: Extended range predictions with ECMWF models. I: Interannual variability in operational model integrations. *Quart. J. Roy. Meteor. Soc.*, **116**, 799834.

Rashid, H., H. H. Hendon, M. C. Wheeler, and O. Alves: Prediction of the Madden-Julian Oscillation with the POAMA dynamical prediction system. *Climate Dynamics*, *36 (3)*, 649-661.

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.

Simmons, A. J. and A. Hollingworth, 2002: Some aspects of the improvement of skill of weather numerical prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647-677.

Simmons, A., S. Uppala, D. P. Dee and S. Kobayashi, 2007: ERA-Interim: New ECMWF reanalysis products from 1989 onwards, *ECMWF Newsletter*, **110**, 25-35.

Tracton, M. S., K. Mo, W. Chen, E. Kalnay, R. Kistler, and G. White, 1989: Dynamical Extended Range Forecast (DERF) at the National Meteorological Center. *Mon. Wea. Rev.*, 117, 16041635.

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

Vitart, F., R. Buizza, M. A. Balmaseda, G. Balsamo, J.-R. Bidlot, A. Bonet, M. Fuentes, A. Hofstadler, F. Molteni and T. Palmer, 2008: The new VAREPS-monthly forecasting system: a first step towards seamless prediction. *Quart. J. Roy. Meteor. 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. and F. Molteni, 2010: Simulation of the MJO and its teleconnections in the ECMWF forecast system. *Quart. J. Roy. Meteor. Soc*, **136**, 842-855.

Waliser, D. L. and Coauthors, 2011: The Year of Tropical Convection (May 2008 to April 2010): Climate variability and weather highlights. Bull. Amer. Meteor. Soc.,
doi: http://dx.doi.org/10.1175/2011BAMS3095.1

Weaver S.-J., W. Wang, M. Chen, A. Kumar, 2011: Representation of MJO Variability in the NCEP Climate Forecast System. *J. Climate*, **24**, 4676-4694.

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

Weigel, A., D. Baggenstos, M.A. Liniger, F. Vitart and C. Appenzeller, 2008: Probabilistic verification of monthly temperature forecasts. *Mon. Wea. Rev.*, **136**, 5162-5182.

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.S., 2006: *Statistical Methods in Atmospheric Sciences*. International Geophysics Series, Vol. 91, Academic Press, 627 pp.

Wolff, J.O., E. Maier-Raimer, and S. Legutke, 1997: The Hamburg ocean primitive equation model. Deutsches Klimarechenzentrum Tech. Rep. 13, Hamburg, Germany, 98 pp.

Woolnough, S.J., F. Vitart and M. A. Balmaseda, 2007: The role of the ocean in the Madden-Julian Oscillation: Implications for MJO prediction. *Q. J. R. Meteorol. Soc.*, **133**, 117-128.

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