Monthly Weather Review

Seamless precipitation prediction skill in the tropics and extratropics from a global model

--Manuscript Draft--

Manuscript Number:	
Full Title:	Seamless precipitation prediction skill in the tropics and extratropics from a global model
Article Type:	Article
Corresponding Author:	Matthew C. Wheeler, Ph.D. Bureau of Meteorology Melbourne, Victoria AUSTRALIA
Corresponding Author's Institution:	Bureau of Meteorology
First Author:	Hongyan Zhu
Order of Authors:	Hongyan Zhu
	Matthew C. Wheeler, Ph.D.
	Adam H. Sobel
	Debra Hudson
Abstract:	The skill with which a coupled ocean-atmosphere model is able to predict precipitation over a range of time scales (days to months) is analysed. For a fair comparison across a seamless range of time scales, the verification is performed using data averaged over time windows equal in length to the forecast lead time. At a lead time of one day, skill is greatest in the extratropics around 40-60º latitude, lowest around 20º, and has a secondary local maximum close to the equator. The extratropical skill at this short range is highest in the winter hemisphere presumably due to the higher predictability of winter baroclinic systems. The local equatorial maximum comes mostly from the Pacific, and thus appears to be from the El Niño-Southern Oscillation. As both lead time and averaging window are simultaneously increased, the extratropical skill drops rapidly with lead time, while the equatorial maximum remains approximately constant causing the equatorial skill to exceed the extratropical at leads ≥4 days in austral summer and ≥7 days in boreal summer. At longer lead times, the extratropical skill eventually flattens out or increases, but does not approach the equatorial values. Comparisons with persistence confirm that the model beats persistence for most leads and latitudes, including for the equatorial Pacific where persistence is high. The results are consistent with the view that extratropical predictability is mostly derived from synoptic-scale atmospheric dynamics, while tropical predictability is primarily derived from the response of moist convection to slowly-varying forcing such as from sea surface temperature.
Suggested Reviewers:	Tom Hamill tom.hamill@noaa.gov
	Jagadish Shukla shukla@iges.org
	Duane Waliser duane.e.waliser@jpl.nasa.gov
	Frederic Vitart Frederic.Vitart@ecmwf.int

Cost Estimation and Agreement Worksheet

Click here to download Cost Estimation and Agreement Worksheet: Page_charges_form.pdf

Seamless precipitation prediction skill in the tropics and extratropics from a global model

Hongyan Zhu[a], Matthew C. Wheeler[a], Adam H. Sobel[b], and Debra Hudson[a]

[a] Centre for Australian Weather and Climate Research,

A partnership between CSIRO and the Bureau of Meteorology

Melbourne, Australia

[b] Department of Applied Physics and Applied Mathematics

Department of Earth and Environmental Sciences

Lamont-Doherty Earth Observatory, Columbia University, New York

For submission to *Monthly Weather Review*

Corresponding author address: Dr. Matthew Wheeler, CAWCR/Bureau of Meteorology, GPO Box 1289, Melbourne, VIC 3001, Australia.

E-mail: m.wheeler@bom.gov.au

Abstract

The skill with which a coupled ocean-atmosphere model is able to predict precipitation over a range of time scales (days to months) is analysed. For a fair comparison across a seamless range of time scales, the verification is performed using data averaged over time windows equal in length to the forecast lead time. At a lead time of one day, skill is greatest in the extratropics around 40-60º latitude, lowest around 20º, and has a secondary local maximum close to the equator. The extratropical skill at this short range is highest in the winter hemisphere presumably due to the higher predictability of winter baroclinic systems. The local equatorial maximum comes mostly from the Pacific, and thus appears to be from the El Niño-Southern Oscillation. As both lead time and averaging window are simultaneously increased, the extratropical skill drops rapidly with lead time, while the equatorial maximum remains approximately constant causing the equatorial skill to exceed the extratropical at leads ≥4 days in austral summer and ≥7 days in boreal summer. At longer lead times, the extratropical skill eventually flattens out or increases, but does not approach the equatorial values. Comparisons with persistence confirm that the model beats persistence for most leads and latitudes, including for the equatorial Pacific where persistence is high. The results are consistent with the view that extratropical predictability is mostly derived from synoptic-scale atmospheric dynamics, while tropical predictability is primarily derived from the response of moist convection to slowly-varying forcing such as from sea surface temperature.

1. Introduction

Extratropical and tropical weather have different characteristics. Extratropical weather is dominated by baroclinic disturbances which obtain their energy from the vertical shear in the mean flow and the available potential energy associated with the horizontal temperature gradients that balance that shear (Charney 1947; Lorenz 1955). Precipitation tends to be strongly forced at large scales by isentropic uplift along fronts and dynamical lifting due to the advection of quasi-balanced upper-level potential vorticity anomalies (Bluestein 1993). Tropical weather exists in an environment of much weaker pressure and temperature gradients and (at least in the zones of greater climatological precipitation) higher humidity (Charney 1963, 1969). Tropical precipitation is typically a result of deep convection and closely associated stratiform rain (Schumacher and Houze 2003). The convection is often organized into wave-like disturbances (Wheeler and Kiladis 1999), but it is unclear to what extent these disturbances are dynamically independent entities that organize the convection (as is the case for many extratropical disturbances) as opposed to resulting from spontaneous "self-aggregation" (Mapes 1993; Bretherton et al. 2005) of the convection itself. The lead time at which tropical weather becomes inherently unpredictable is not well known, but is generally thought to be shorter than that for extratropical weather (Shukla 1989; Boer 1995).

At the same time, a substantial literature has established that the tropics are the source of most potential predictability globally on seasonal to interannual time scales (Charney and Shukla 1981; Goddard et al. 2001). There are some extratropical sources of predictability on time scales of weeks to months, such as from stratospheric effects (Baldwin and Dunkerton 1999; Polvani and Kushner 2002), snow cover

(Cohen and Entekhabi 1999), sea ice (Holland et al. 2013) and soil moisture (Koster and Suarez 2003). It appears, however, that tropical sea surface temperature variations, particularly those resulting from the El Niño-Southern Oscillation (ENSO) phenomenon, have a greater impact on the global climate (Hoerling and Kumar 2002). The impacts of ENSO are felt strongly not only in the tropics, but also in many extratropical regions (Kiladis and Diaz 1989), due to atmospheric teleconnections (DeWeaver and Nigam 2004). Of course, at seasonal-to-interannual time scales one is not predicting the daily weather, but only the averages over a month or a season.

The intraseasonal time scale lies between daily weather and seasonal climate. On that intermediate time scale we expect the Madden-Julian oscillation (MJO) to be a source of predictability in the tropics (Waliser et al. 2006), while there may be some additional predictability associated with low-frequency extratropical modes driven by eddy-mean flow interactions (Baldwin et al. 2003).

We expect, then, that the extratropics are more predictable than the tropics at lead times of a day to a week, while the tropics are more predictable at climate scales of months to a year (Shukla 1989; Sobel 2012). Our interest here is in testing whether that expectation is correct, and studying the transition between the two time scales, in both tropical and extratropical latitudes.

We study the relative skill of a particular prediction system in predicting tropical versus extratropical weather and climate across a range of time scales, from daily to monthly. We use a coupled ocean-atmosphere ensemble forecast system that is used operationally for prediction on a range of time scales, from a few days to seasons. We focus on precipitation as it is of interest and has utility in both tropical

and extratropical regions (as opposed to pressure and temperature, which vary much less in the tropics than extratropics and are therefore of less interest there). The model used in the forecast system contains some representation of the main sources of predictability described above, with the exception of the stratospheric sources (Roff et al. 2011) and sea ice variations. Thus, while this is not a true study of potential predictability limits, the prediction skill from the current model should be somewhat comparable to those limits, at least within the realms of our current knowledge. Further, it is of interest to determine the comparative prediction skill that is currently available from an operational system.

The essence of our approach is as follows. We compute the prediction skill at a range of lead times, from one day to one month. As the lead time increases, we also increase the length of the time window over which the data are averaged for verification. This is intended to capture the fact that we are transitioning from weather to climate prediction as the lead time increases, and to allow the transition to occur smoothly. The skill is computed for both total precipitation and anomalies, and comparison is made with the skill achievable by a persistence forecast of the precipitation anomalies. For comparison we also evaluate the forecasts at varying lead time but with a fixed verification window of 1 day.

2. Data and method

a. POAMA-2 ensemble forecast system

We use the Bureau of Meteorology's dynamical Predictive Ocean Atmosphere Model for Australia (POAMA; Alves et al. 2003) version 2 configured for multi-week predictions ("POAMA-2 multi-week" abbreviated to P2-M; Hudson et al. 2013).

Earlier versions of POAMA were designed for seasonal forecasting, however, improvements to the generation of initial conditions to use perturbed atmosphere and ocean initial conditions and a burst ensemble (i.e. an ensemble starting from a single initial time as opposed to a lagged ensemble), and the use of three different model configurations to form a multi-model ensemble, have made P2-M applicable for shorter range forecasts as well, especially at the intraseasonal time scale (Hudson et al. 2013).

The atmospheric component of P2-M is run in spectral space with a triangular truncation at wavenumber 47 (approximately a 250km grid) and 17 vertical levels. It includes a land component which is a simple bucket for soil moisture and three soil levels for temperature. The ocean model has a zonal resolution of 2°, a meridional resolution of 0.5° within 8° of the equator increasing to 1.5° near the poles, and 25 levels. While the atmospheric model has a relatively coarse resolution compared to modern numerical weather prediction models, it is comparable to what has commonly been used for seasonal prediction over the last decade and is considered adequate to resolve the key sources of predictability discussed in the introduction. Further details of these model components are provided in Hudson et al. (2013) and references therein.

Also important are the methods employed for producing initial conditions, including perturbations to the initial conditions which are needed to make a forecast ensemble. The unperturbed initial conditions are provided by separate data assimilation schemes for the ocean versus the atmosphere and land. The atmosphere and land initial conditions are created by nudging zonal wind, meridional wind, atmospheric temperature, and humidity in the atmosphere-land component of the

model (when run prior to hindcasts or forecasts being made, and forced with observed sea surface temperatures) towards an observationally based analysis (Hudson et al. 2011). The analysis used is the European Centre for Medium Range Weather Forecast's 40 year Re-analysis (ERA-40; Uppala et al. 2005) for the period 1980 to August 2002, and the Bureau of Meteorology's operational global numerical weather prediction (NWP) analysis thereafter. Ocean initial conditions are derived using a pseudo-ensemble Kalman filter data assimilation system (Yin et al. 2011). In-situ ocean temperature and salinity observations are assimilated and corrections to currents are generated based on the ensemble cross-covariances with temperature and salinity.

Perturbations to the initial conditions of the central member are generated using a coupled breeding scheme which produces perturbations to all components of the coupled system in a consistent fashion. Ten different perturbed states are produced which, together with the unperturbed central member, provides for 11 different initial states from which to start a burst ensemble (Hudson et al. 2013).

In addition to the perturbed initial states, which allow for an estimate of forecast uncertainty due to sensitivity to initial condition errors, a multi-model ensemble comprising three different model configurations is used to provide a sample of model uncertainty. The three configurations are differentiated by their use of: (a) standard physics with no flux correction; (b) bias correction of fluxes at the air-sea interface; and (c) as in (a), except with modified atmospheric physics in the form of an alternative shallow convection parameterization. Each model configuration is run with the 11 different initial conditions to provide a 33-member ensemble. The coupled breeding of initial states uses version (a).

142 The above description of the P2-M system applies to both the hindcasts (i.e.
143 forecast runs that are started using initial states from previous times) as well as
144 forecasts run in real time. In this work we analyse the skill of the hindcasts only.
145 However, given the same configuration of the hindcast and real-time systems, we
146 expect that the skill of the hindcasts should be comparable to a suitably large sample
147 of real-time forecasts, assuming a relatively stable climate. The hindcasts we analyse
148 have start times on the 1st, 11th and 21st days of each month of the year. To match
149 the period of available global daily precipitation observations (see section 2b), we
150 analyse the period between 1996 and 2009 only.

151 *b. Observations*

152 The observational data set for verification in this paper is the Global
153 Precipitation Climatology Project (GPCP) daily precipitation with 1° resolution
154 (Huffman et al. 2001). The GPCP data is a blended product derived from both station
155 observations and satellite measurements. The satellite data is sourced from both
156 geostationary and polar-orbiting platforms. When this work commenced, the available
157 daily GPCP data (version 1.1) extended from October 1996 to August 2009, which is
158 the period we have chosen to evaluate the model hindcasts. We map the GPCP data to
159 the model grid by first interpolating the GPCP data to a 0.5º grid, and then averaging
160 in the zonal and meridional directions to match the POAMA grid spacing. Our
161 analysis therefore concentrates on precipitation that is area-averaged over a scale of
162 about 250 km, providing a reasonable representation of most synoptic-scale weather.
163 Known problems exist in the GPCP data at high latitudes (Bolvin et al. 2009),
164 however our results and survey of the literature give us enough confidence to show
165 the skill calculations to a latitude of 80º.

166 *c. Measures of prediction skill*

167 We assess skill by comparing the P2-M forecasts with the verifying GPCP
168 observations. We computed a number of different verification measures, each having
169 different strengths and weaknesses (not shown), and verified that the conclusions are
170 not sensitive to which ones we use. We therefore choose to show the simplest
171 measures for this paper: the correlation of the ensemble mean total precipitation with
172 the observed verification data (hereafter CORt) and the correlation of the ensemble
173 mean precipitation anomalies with the observed anomalies (hereafter CORa). These
174 correlations are computed over time (i.e. using data from many different verification
175 windows), separately for each grid point and each lead time. In the case of CORt, this
176 measure is affected by both the model's ability to accurately represent the
177 climatological seasonal cycle in its forecasts, and the variability. In the case of CORa,
178 the seasonal cycles are removed from both the observations and forecasts by
179 removing their respective climatologies. For the model this is the hindcast
180 climatology, which is a function of both lead time and start day and month. CORa
181 therefore is affected only by the model's ability to forecast the variability about that
182 climatology.

183 Computing and showing CORt is a more usual practice for the weather
184 prediction community (e.g. Ebert 2001), whereas concentrating on anomalies (i.e.
185 CORa) is more usual for the seasonal prediction community (e.g. Cottrill et al. 2013).
186 This is partly because users of weather information are more interested in total
187 precipitation, whereas users of climate information are more interested in whether
188 future conditions may be wetter or drier than normal (i.e. anomalies). Another reason
189 is that the numerical weather prediction community tend not to produce large hindcast

datasets (which are necessary for computing a model climatology) whereas seasonal prediction systems require hindcasts to take account of the climate drift that becomes noticeable at longer lead times (Stockdale 1997). Perhaps the main disadvantage of these two verification measures is that they ignore the probabilistic nature of the ensemble. The other disadvantage is that the correlation is insensitive to mean bias. However, given that in this work we are more interested in the relative skill between regions and lead times, we feel their simplicity outweighs these disadvantages.

For both CORt and CORa the correlation is calculated at each grid point as:

$$COR = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$

where x is the ensemble mean forecast precipitation (using totals for CORt and anomalies for CORa), y is the observed precipitation value (totals or anomalies), n is the number of verification times, and each sum is calculated over n values.

To study the differences in skill between different seasons, we show computations of CORt and CORa for the contrasting seasons of December-January-February (DJF) and June-July-August (JJA) for which n is 117 and 108 respectively (13 or 12 years × 3 months per season × 3 forecast starts per month). When computing the correlation for a particular season, one may at first think that CORt will be equal to CORa since the correlation automatically removes the respective time mean values from the two fields that are being correlated. However, the seasonal cycle is not constant across a 3 month season, so in practice CORt and CORa are not the same.

211 Further detail on the calculation of the climatological season cycles is as
212 follows: For both the observed and model forecast precipitation, the exact same years
213 are used to compute the climatology (i.e. October 1996 to August 2009). We are also
214 careful to use the exact same days of the year from the verifying observations as from
215 the model. For example, consider the forecast of the 2nd week from the initial
216 condition of 11 December 2001. The dates of the 2nd week are 19-25 December 2001.
217 The observed climatology for this forecast is computed by averaging the precipitation
218 data for 19-25 December for all 13 years (i.e. 7×13 days of observed data). The model
219 climatology for this forecast is computed by averaging the model precipitation for 19-
220 25 December from all 33 ensemble members of all 13 years of forecasts that were
221 initialized from 11 December (i.e. 7×33×13 days of model data). Note that unlike
222 Hudson et al. (2013), we do not need to compute a different hindcast climatology for
223 each of the 3 different model configurations because of the linearity of the ensemble
224 mean calculation (this is not the case for probabilistic forecasts).

225 *d. Forecast time window definition*

226 As stated in the introduction, we take the approach of widening the time
227 averaging window of the forecasts and verifying observations when looking at
228 progressively longer lead times. For example, for a forecast lead time of 1 day we use
229 an averaging window of 1 day, and for a lead time of 1 week we use a window of 1
230 week. A schematic of this approach and the terminology we use to label it is provided
231 in Fig. 1. Our intention is to provide a seamless transition from weather to climate
232 prediction in this analysis of skill. Note that "1d1d" is what is usually called "day 2"
233 in other papers, and "1w1w" is what is usually called "week 2". The longest window
234 and lead time combination we consider is 4 weeks (i.e. 4w4w). 4w4w is roughly

equivalent to “month 2” in other papers, noting that a month is roughly 4 weeks long. We also study the intermediate window/lead times of 2d2d, 4d4d, and 2w2w, providing a total of 6 different time scales. Later in the paper we also evaluate forecasts using the more traditional approach of varying the lead time but with a fixed verification window of 1 day. Using the terminology discussed above, this latter analysis focusses on forecasts for 1d0d to 1d2w, where 1d0d is equivalent to the first 24 hours of the forecast (see Fig. 9).

e. Seasonal definition

We show our computations of CORt and CORa for the seasons of DJF and JJA only (as described above). Note that we use the starting date (i.e. initial condition) of the model forecasts to determine the season rather than the verifying time. This means that for the 4w4w calculations the verification times extend up to ~7 weeks after the end of each season (noting that the latest hindcast each season is initialized on the 21st of the month). For example, the 4w4w calculations for JJA will include verifying data from 30 June to 16 October.

3. Results

a. CORt – correlation with the ensemble mean totals

Maps of CORt for the contrasting seasons of DJF and JJA and for the window/lead time combinations of 1d1d and 4w4w are displayed in Fig. 2. Positive values indicate positive skill in the sense that there is an in-phase relationship between the forecast and observed values. For 1d1d, a positive skill is achieved everywhere except over the subtropical dry zones over Africa, the eastern Atlantic, and eastern Pacific. In DJF the highest large-scale 1d1d CORt (> 0.5) is achieved over the north

Pacific and north Atlantic, while in JJA the region of highest large-scale CORt is over the mid-latitudes of the southern hemisphere. This is consistent with previous work (Ebert et al. 2003) that shows that extratropical precipitation is generally easier to predict for short lead times in winter when it is associated mainly with synoptic-scale systems such as fronts, whereas in summer it is more often associated with convective systems such as thunderstorms that are harder to predict. This short-range seasonality in the extratropics will become more apparent in the zonally-averaged skill plots in Figs. 3, 6, and 10.

Interestingly, the 1d1d CORt maps (Fig. 2) also indicate some patches of very high skill in the equatorial zone especially over the Indian and Pacific Ocean sectors in DJF. This was not initially anticipated given our review of published papers as discussed in the introduction. We did not expect such high skill in the tropics at short lead times.

At the longer window/lead time scale of 4w4w, the CORt maps of Fig. 2 indicate greatest skill (CORt > 0.7) over the tropical Pacific, especially in DJF. This is obviously a result of the predictability provided by ENSO. Greatest precipitation skill (CORt $\geq$ 0.9) is achieved over the central-eastern equatorial Pacific because this is where precipitation is most strongly related to the SST variations of ENSO (Weare 1987). Indeed, these maps look much like the maps of SST skill for POAMA provided in Cottrill et al. (2013). Further, DJF is when ENSO events typically reach their peak SST anomaly, so greater precipitation prediction skill from forecasts initialized in DJF is somewhat expected.

A further interesting feature from Fig. 2 is the band of CORt > 0.3 extending around the globe at the latitudes of 50-65ºS for 4w4w in DJF. Our initial thought was

that this may be related to the southern annular mode and its relationship with ENSO (L'Heureux and Thompson 2006). This relationship is known to be strongest in DJF. However, as we will show later, this signal mostly disappears when the skill associated with the climatological seasonal cycle is removed (i.e. in CORa), indicating that it stems from a pronounced seasonal cycle that is well represented by the model during DJF for those latitudes.

Information from the intermediate window/lead times is presented in Fig. 3 which shows the zonally-averaged CORt for the model forecasts for five different lead times and averaging windows, extending from 1d1d to 4w4w. In the extratropics at short lead times, greater skill in winter than summer, as discussed above, is readily apparent. In both the northern and southern hemisphere the zonally-averaged CORt is greater than 0.5 in winter and less than 0.5 in summer. As the window/lead time increases, the CORt skill in the extratropics generally decreases until 1w1w at which point it appears to approximately level-off such that the 4w4w CORt is on average somewhat higher. Interestingly, at the 4w4w time scale, the CORt in DJF is on average higher than that in JJA in both hemispheres (and in the tropics).

Turning now to the deep tropics (i.e. within about 10º of the equator), the variation of skill with increasing window/lead times is much different to that described above for the extratropics. Indeed, Fig. 3 nicely shows how the skill remains remarkably constant with increasing window/lead time in the tropics. In fact, the skill increases somewhat with window/lead time during DJF.

Another way to look at the variation of skill in the tropics versus extratropics with increasing window/lead time is presented in Fig. 4. In this figure we add the 2w2w time scale, and space the time scales along the x-axis according to their

306 logarithm. We can now see more clearly that for the extratropics in both hemispheres,
307 there tends to be a minimum in CORt skill for the 2w2w time scale in all latitude
308 bands in both seasons, except for the 70ºS-50ºS band in DJF (which has its minimum
309 at 1w1w). This indicates that the 2nd half of the first month (or equivalently weeks 3
310 and 4 together) are the most unpredictable when evaluated this way. In contrast, the
311 tropical latitudes show very little variation of CORt with time scale.

312 *b. CORa – correlation with the ensemble mean anomalies*

313 As we described in Section 2c, CORt may be influenced by the ability of the
314 model forecasts to represent the observed seasonal cycle. If there is a strong seasonal
315 cycle that is accurately represented by the model then CORt will be higher, but if the
316 model gets the seasonal cycle reversed, CORt will be lower. CORa on the other hand,
317 removes the effects of the climatological seasonal cycle, and it is the more usual way
318 of showing the correlation skill in seasonal prediction studies.

319 Comparing the CORa maps in Fig. 5 with the CORt maps in Fig. 2, the most
320 obvious difference is generally lower values for CORa for the longer window/lead
321 time window, but with very little change for 1d1d. The reason for this difference
322 (between 1d1d and 4w4w) is because a longer averaging window gets a greater
323 contribution to its total variance from the seasonal cycle. Removing the contribution
324 from the seasonal cycle makes the model performance look worse at the 4w4w time
325 scale, especially in regions away from the ENSO-dominated tropical Pacific. The
326 most obvious location for this apparently lower skill (when looking at CORa
327 compared to CORt) is over the Southern Ocean around 55ºS in DJF. As discussed in
328 the previous section, we initially thought high CORt in this region may be associated
329 with the southern annular mode (see also Lim et al. 2013). However, given the

absence of this signal in CORa, it appears to instead be associated with an accurate representation of the seasonal cycle. This reduction in apparent skill when measured with CORa is similar to the effect described by Hamill and Juras (2006).

Looking at the maps of Fig. 5 in more detail, there are a few regions of relatively high 4w4w skill that stand out. In the tropics, the ENSO-dominated signal in the equatorial Pacific extends westward into the islands of Indonesia and Papua New Guinea in JJA, and more towards the Philippines to the north in DJF, consistent with the empirical findings of McBride et al. (2003). In the northern hemisphere there are patches of relatively high skill in the North Pacific and western United States in DJF, consistent with our expectation from knowledge of the Pacific-North American (PNA) pattern (Kumar and Hoerling 1998). In the southern hemisphere there is skill in the South Indian Ocean and Western Australia in DJF, and eastern Australia in JJA. The latter is expected given the known influence of ENSO in Australia (McBride and Nicholls 1983). Importantly, these correlation skill values are greater than what are achievable from persistence alone (Simmonds and Hope 1997; next sub-section). Other interesting patches of high 4w4w CORa are in north Africa and the southeast Pacific in JJA, and the western equatorial Indian Ocean in DJF.

When viewing the zonally-averaged CORa values (as a function of latitude and window/lead time) in Fig. 6, a very similar conclusion is reached as we obtained when looking at CORt. That is, that prediction skill decreases with window/lead time in the extratropics (outside of about 10º of the equator), but stays much the same in the tropics. Similarly, when looking at the alternative display of Fig. 7, which includes the additional time scale of 2w2w, we can see this variation with window/lead time clearly. We can also see at what point the skill in the tropics (when taken as a whole)

354 begins to exceed that in the extratropics: for DJF it occurs at the 4d4d time scale, and
355 for JJA it occurs at the 1w1w scale.

356 *c. Comparison with persistence*

357 An important component of predictability is the prediction skill that can come
358 from persistence, so it is of interest to see how these results compare. Fig. 8 presents
359 the correlation skill for persistence forecasts for 4 different time-scales (labelled as
360 P1d1d, P4d4d, P2w2w, and P4w4w), and also shows the correlations for the 1d1d and
361 4w4w model forecasts for comparison. These persistence calculations used
362 precipitation anomalies (i.e. CORa), and like for the model forecasts an averaging
363 window equal in length to the lead time was used. For example, the P1d1d calculation
364 uses the observed precipitation anomaly on the day before the model initial condition
365 as the forecast, whereas the P4w4w calculation used the precipitation anomaly
366 observed for the 4 weeks leading up to the initial condition.

367 In general, it can be seen in Fig. 8 (and with comparison to Fig. 6) that the
368 zonally-averaged CORa from the model tends to be higher than that for persistence,
369 especially for the shorter time scales. Viewing maps of the persistence skill (not
370 shown) confirms that this is generally the case for individual locations as well. Even
371 at the longer 4w4w time scale, the model CORa exceeds or approximately equals the
372 persistence skill (i.e. for P4w4w) for most latitudes equatorward of 50º. This is an
373 encouraging result for the model because persistence has historically been difficult to
374 beat at this range, as discussed by Vitart (2004).

375 Poleward of 50º, however, there are some notable peaks in P4w4w that are not
376 replicated in the model forecasts, located around 70ºS in DJF, and 65ºS and 75ºN in

JJA. The maps of persistence skill (not shown) indicate that these peaks correspond to regions where large and persistent anomalies in sea ice cover occur (Parkinson and Cavalieri 2008; Wheeler 2008), and an influence of sea ice on precipitation appears quite possible (Weathery 2004). We note that POAMA-2 uses prescribed sea ice from a multi-year climatology, so is not able to reproduce this persistence skill, but it is something that may be improved by the incorporation of varying sea ice and sea ice anomalies in the initial condition in future versions of the model.

d. Fixed time-averaging window of 1 day

Instead of increasing the time window at the same rate as the lead time, in Figs. 10 and 11 we present the prediction skill as a function of lead time and latitude for a fixed time window of 1 day (see schematic of the window and lead definitions used in Fig. 9). In this analysis we show CORt only. As expected, the skill drops off much more rapidly (and monotonically) with lead time with a fixed window than it does when the window is increased. Importantly, however, the rate at which the CORt skill drops is much less in the tropics than the extratropics providing the same general conclusion as before, that is, that there is a general transfer of skill from the extratropics to tropics as lead time is increased. The lead time at which the skill in the tropics tends to surpass the skill in the extratropics is shown to be at about 4 days in DJF about 14 days in JJA. These values are respectively similar to and a little longer than the values found when the window length was allowed to vary as well (Figs. 4 and 7). Having a slightly longer estimate from this 1 day window calculation makes sense, because it uses effectively earlier information from the forecasts. For example, 1d2w is equivalent to day 15 whereas 2w2w is equivalent to days 15 to 28.

4. Conclusions

We have analysed the skill with which an operational forecast system is able to predict precipitation over a range of time scales from a day to months. We focus on the contrasting results obtained for different latitude bands and at different lead times. To emphasize the seamless transition between weather and climate prediction, we have verified the model predictions after averaging both the forecasts and observed verification data over a time window equal to the forecast lead time. We performed skill calculations both on the total fields, and on anomalies computed by removing the appropriate climatological seasonal cycles from both the forecasts and the verification data. The skill measures we present are based on correlations between the forecasts and observed computed over time for each grid point. Calculations are made for the contrasting DJF and JJA seasons with ~13 years of model hindcasts.

At a lead time of one day, prediction skill is greatest in the extratropics around 40-60º latitude, lowest around 20º latitude and poleward of 70º, and has a secondary local maximum close to the equator. The extratropical skill at this short range is highest in the winter hemisphere presumably due to the high day-to-day predictability of winter baroclinic weather systems and associated fronts. In the summer hemisphere extratropics it is less, evidently due to the greater difficulty in predicting summer thunderstorms and the weaker summer baroclinic systems, but it still exceeds the 1 day prediction skill near the equator. The local equatorial maximum in the zonal mean is derived from the central and eastern Pacific, and thus appears (even at one-day lead time) to be related to ENSO.

As both lead time and averaging window are simultaneously increased, the extratropical skill drops rapidly at early lead times, while the equatorial maximum decreases much more slowly or stays approximately constant. The near-equatorial

425 skill becomes equal to or greater than that at any other latitude band at around 4 days
426 lead time in DJF, and 7 days in JJA. At longer lead times, the extratropical
427 correlations eventually flatten out or increase with lead time, but do not approach the
428 near-equatorial values.

429 Importantly, the model prediction skill exceeds the skill of a persistence (of
430 anomalies) forecast in most locations, especially at shorter lead times. For predictions
431 of a 4 week average at a lead of 4 weeks (i.e. 4w4w) the model skill remains better
432 than persistence equatorward of about 50º, but is dramatically worse than persistence
433 in a few locations near the sea ice edges in the Arctic and Antarctic.

434 To compare with our method of using an increasing window size with
435 increasing forecast lead, we also computed the skill for varying lead times but a fixed
436 averaging window of one day, a calculation more similar to the typical practice in
437 weather forecast verification. The correlations at longer leads are smaller than those
438 computed at the same leads but with averaging windows equal to the lead time, as
439 expected. However, perhaps more surprisingly, the slower decay of equatorial skill
440 found with variable averaging windows is also found with the fixed one-day
441 averaging window, so that at sufficiently long leads (~4-7 days) the equatorial skill
442 still exceeds that in the other latitude bands.

443 The broad picture we are left with is that on time scales of a few days or less,
444 extratropical precipitation is more predictable than tropical, while at time scales of a
445 week or longer, tropical precipitation, within about 10º of the equator, is more
446 predictable than extratropical. This broad picture is remarkably robust to the details
447 of how one does the calculations. While the absolute values of the skill depends on
448 season, and on whether the averaging window is fixed or increasing with lead time, in

all cases the near-equatorial zone becomes more predictable than the extratropics at lead times of 4-7 days.

This picture appears consistent with the view that extratropical predictability is mostly derived from the model's ability to simulate synoptic-scale atmospheric dynamics with rapid growth of initial state error (Lorenz 1969), while predictability in the deep tropics is mostly derived from the response of moist convection to slowly-varying forcing such as from sea surface temperature (Charney and Shukla 1981). If there is any surprise here, it is that tropical influences can provide greater predictability than extratropical atmospheric dynamics at time scales as short as 4-7 days.

Finally, we advocate the usefulness of computing and displaying forecast skill globally across a large range of time scales as we have done here. Using precipitation as the verifying variable provides what we think is a fair comparison between the tropics and extratropics and the technique of increasing the averaging window size at the same rate as increasing the lead time provides the fairest comparison between different time scales. Recently, the need for seamless verification approaches has been promoted by Ebert et al. (2013), and while other approaches do exist (DelSole and Tippett 2009), we feel the simplicity of our approach is an important advantage. Future work is planned to analyse other forecast systems (especially those employing a model with higher resolution) and to further investigate the skill as measured by verification measures that take into account the probabilistic nature of the ensemble.

Acknowledgements

471 We thank Beth Ebert and Huqiang Zhang for advice on verification methods, Harry
472 Hendon for advice on predictability and POAMA, Frederic Vitart for discussions of
473 tropical versus extratropical prediction skill, Eunpa Lim for suggestions on the title
474 and discussions about the southern annular mode, and Phil Reid for advice on sea ice.
475 Beth Ebert, John McBride, and Mike Tippett kindly read earlier versions of this paper.
476 Hongyan Zhu receives funding support from the Australian Climate Change Science
477 Program.

478

479

480 **References**

481 Alves, O., and Coauthors, 2003: POAMA: Bureau of Meteorology operational
482 coupled model seasonal forecast system. *Science for Drought—Proc. National*
483 *Drought Forum*, Brisbane, Australia, Queensland government, 49–56.

484 Baldwin, M. P., and T. J. Dunkerton, 1999: Propagation of the Arctic Oscillation from
485 the stratosphere to the troposphere. *J. Geophys. Res.*, **104**, D24, 30937–30946.

486 Baldwin, M.P., D.B. Stephenson, D.W.J. Thompson, T J. Dunkerton, A.J. Charlton,
487 and A. O'Neill, 2003: Stratospheric memory and extended-range weather
488 forecasts. *Science*, **301**, 636–640.

489 Bluestein, H.B., 1993: *Synoptic-dynamic meteorology in midlatitudes*. Volume II.
490 Observations and theory of weather systems. Oxford University Press, New
491 York. 606 pp.

492 Boer, 1995: Analyzed and forecast large-scale tropical divergent flow. *Mon. Wea.*
493 *Rev.*, **123**, 3539-3553.

494 Bolvin, D. T., R. F. Adler, G. J. Huffman, E. J. Nelkin, and J. P. Poutiainen, 2009:
495 Comparison of GPCP monthly and daily precipitation estimates with high-
496 latitude gauge observations. *J. Appl. Meteor. Climatol.*, **48,** 1843–1857.

497 Bretherton, C. S., P. N. Blossey and M. Khairoutdinov, 2005: An energy-balance
498 analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci.*,
499 **62**, 4273-4292.

500 Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J.*
501 *Meteor.*, **4**, 136–162.

Charney, J. G., 1963: A note on large-scale motions in the tropics. *J. Atmos. Sci.*, **20**, 607–609.

Charney, J. G., 1969: A further note on large-scale motions in the tropics. *J. Atmos. Sci.*, **26**, 182–185.

Charney, J.G. and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics*, Editors: Sir James Lighthill and R. P. Pearce, Cambridge University Press, 99-109.

Cohen, J., and D. Entekhabi, 1999: Eurasian snow cover variability and northern hemisphere climate variability. *Geophys. Res. Lett.*, **26**, 345-248.

Cottrill, A., H. Hendon, E. Lim, S. Langford, K. Shelton, A. Charles, D. McClymont, D. Jones, and Y. Kuleshov, 2013: Seasonal forecasting in the Pacific using the coupled model POAMA-2. *Wea. Forecasting*, **28**, 668–680.

DeWeaver, E., and S. Nigam, 2004: On the forcing of ENSO teleconnections by anomalous heating and cooling. *J. Climate*, **17**, 3225–3235.

DelSole, T., M.K. Tippett, 2009: Average predictability time. Part II: seamless diagnoses of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1188–1204.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461-2480.

Ebert, E.E., U. Damrath, W. Ergen, M.E. Baldwin, 2003: The WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc.*, **84**, 481-492.

524 Ebert, E., L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Gober, S.
525 Joslyn, B. Brown, T. Fowler, and A. Watkins, 2013: Progress and challenges in
526 forecast verification. *Meteorol. Appl.*, **20**, 130–139.

527 Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher, and M.A. Cane,
528 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J.*
529 *Climatol*., **21**, 1111–1152.

530 Hamill, T.M. and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the
531 varying climatology? *Q. J. R. Meteorol. Soc.*, **132**, 2905–2923.

532 Hoerling, M.P., and A. Kumar, 2002: Atmospheric response patterns associated with
533 tropical forcing. *J. Climate*, **15**, 2184–2203.

534 Holland, M.M., E. Blanchard-Wrigglesworth, J. Kay, and S. Vavrus, 2013: Initial-
535 value predictability of Antarctic sea ice in the Community Climate System
536 Model 3. *Geophys. Res. Lett*., **40**, 1-4, doi: 10.1002/grl.50410.

537 Hudson, D., O. Alves, H.H. Hendon and G. Wang, 2011: The impact of atmospheric
538 initialisation on seasonal prediction of tropical Pacific SST. *Clim. Dyn.*, **36**,
539 1155-1171.

540 Hudson, D., A.G. Marshall, Y. Yin, O. Alves, H.H. Hendon, 2013: Improving
541 intraseasonal prediction with a new ensemble generation strategy. Accepted to
542 *Mon. Wea. Rev.*

543 Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B
544 McGavock, J. Susskind, 2001: Global precipitation at one-degree daily
545 resolution from multi-satellite observations. *J. Hydrometeor*., **2**, 36-50.

546 Kiladis, G.N., and H.F. Diaz, 1989: Global climatic anomalies associated with
547 extremes in the Southern Oscillation. *J. Climate*, **2**, 1069–1090.

548 Koster, R.D., and M.J. Suarez, 2003: Impact of land surface initialization on seasonal
549 precipitation and temperature prediction. *J. Hydrometeor*., **4**, 408–423.

550 Kumar, A., M.P. Hoerling, 1998: Annual Cycle of Pacific–North American seasonal
551 predictability associated with different phases of ENSO. *J. Climate*, **11**, 3295–
552 3308.

553 L'Heureux, M.L., D.W.J. Thompson, 2006: Observed relationships between the El
554 Niño–Southern Oscillation and the extratropical zonal-mean circulation. *J.*
555 *Climate*, **19**, 276–287.

556 Lim, E.-P., H.H. Hendon, and H. Rashid, 2013: Seasonal predictability of the southern
557 annular mode due to its association with ENSO. Accepted to *J. Climate*.

558 Lorenz, E.N., 1955: Available potential energy and the maintenance of the general
559 circulation. *Tellus*, **2**, 157-167.

560 Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer.*
561 *Meteor. Soc.*, **50,** 345-349.

562 Mapes, B.E., 1993: Gregarious tropical convection. *J. Atmos. Sci.*, **50**, 2026–2037.

563 McBride, J.L., M.R. Haylock, and N. Nicholls, 2003: Relationships between the
564 Maritime Continent heat source and the El Niño-Southern Oscillation
565 phenomenon. *J. Climate*, **16**, 2905-2914.

566 McBride J.L., and N. Nicholls, 1983: Seasonal relationship between Australian
567 rainfall and the Southern Oscillation. *Mon. Wea. Rev.*, **111**, 1998–2004.

568 Parkinson, C.L., and D.J. Cavalieri, 2008: Arctic sea ice variability and trends, 1979-
569 2006. *J. Geophys. Res.*, **113**, doi:10.1029/2007JC004558.

570 Polvani, L. M., and P. J. Kushner, 2002: Tropospheric response to stratospheric
571 perturbations in a relatively simple general circulation model. *Geophys. Res.*
572 *Lett.*, **29**, doi: 10.1029/2001GL014284.

573 Roff, G., D.W.J. Thompson, and H.H. Hendon, 2011: Does increasing model
574 stratospheric resolution improve extended-range forecast skill? *Geophys. Res.*
575 *Lett.*, **38**, doi: 10.1029/2010GL046515

576 Schumacher, C., and R.A. Houze Jr., 2003: Stratiform rain in the tropics as seen by
577 the TRMM Precipitation Radar. *J. Climate*, **16**, 1739–1756.

578 Shukla, J., 1989: Tropical forecasting: predictability perspective. *Aust. Met. Mag.*, **37**,
579 141-153.

580 Simmonds, I., and P. Hope, 1997: Persistence of Australian Rainfall anomalies. *Int. J.*
581 *Climatol.*, **17**, 597-613.

582 Sobel, A. H., 2012: Tropical weather. *Nature Education Knowledge*, **3** (10):2.

583 Stockdale, T.N., 1997: Coupled ocean-atmosphere forecasts in the presence of climate
584 drift. *Mon. Wea. Rev.*, **125**, 809-818.

585 Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor.*
586 *Soc.*, **131,** 2961–3012.

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

588 Waliser, D., K. Weickmann, R. Dole, S. Schubert, O. Alves, C. Jones, M. Newman,
589 H.-L. Pan, A. Roubicek, S. Saha, C. Smith, H. van den Dool, F. Vitart, M.
590 Wheeler, and J. Whitaker, 2006: The experimental MJO prediction project. *Bull.*
591 *Amer. Meteor. Soc.*, **87**, 425-431.

592 Weare, B.C., 1987: Relationships between monthly precipitation and SST variations
593 in the tropical Pacific region. *Mon. Wea. Rev.*, **115**, 2687–2698.

594 Weatherly, J.W., 2004: Sensitivity of Antarctic precipitation to sea ice concentrations
595 in a general circulation model. *J. Climate*, **17**, 3214-3223.

596 Wheeler, M.C., 2008: Seasonal climate summary southern hemisphere (summer 2007-
597 08): mature La Nina, an active MJO, strongly positive SAM, and highly
598 anomalous sea-ice. *Aust. Meteor. Mag*., **57**, 379-393.

599 Wheeler, M., and G.N. Kiladis, 1999: Convectively coupled equatorial waves:
600 Analysis of clouds and temperature in the wavenumber-frequency domain. *J.*
601 *Atmos. Sci.*, **56**, 374-399.

602 Yin, Y., O. Alves and P.R. Oke, 2011: An ensemble ocean data assimilation system
603 for seasonal prediction. *Mon. Wea. Rev*., **139**, 786-808.

604

605

606 **Figure captions**

607 Fig. 1. Schematic of the time window and lead time definitions used in this analysis.
608 The horizontal axis represents forecast time from the initial condition. “1d1d” refers
609 to an averaging window of 1 day at a lead time of 1 day. Similarly, “2d2d” represents
610 an averaging window of 2 days at a lead time of 2 days, and so on. Note that “1d1d” is
611 what is usually called “day 2” in other papers, and “1w1w” is what is usually called
612 “week 2”.

613 Fig. 2. Maps of CORt for model forecasts at 1d1d (upper panels) and 4w4w (lower
614 panels). The left panels are for the DJF season and right panels are for JJA.

615 Fig. 3. Zonally-averaged CORt for model forecasts at different time window/lead
616 combinations for DJF (top panel) and JJA (bottom panel).

617 Fig. 4. Zonally-averaged CORt over specified latitude ranges versus forecast
618 window/lead times from 1d1d to 4w4w for DJF (top panel) and JJA (bottom panel).
619 Note that the spacing of the time scale is based on the logarithm of the lead time.

620 Fig. 5. As in Fig. 2, but for CORa.

621 Fig. 6. As in Fig. 3, except for CORa.

622 Fig.7 As in Fig. 4, except for CORa.

623 Fig. 8. As in Fig. 6, except showing zonally-averaged CORa for persistence forecasts
624 (labelled as P1d1d, P4d4d, P2w2w, P4w4w) and the model forecasts of 1d1d and
625 4w4w for comparison (same as in Fig. 6).

626 Fig. 9. As in Fig.1, except showing the window and lead time definitions used for the
627 calculations in Figs. 10 and 11 (i.e. with a fixed 1-day averaging window).

628 Fig. 10. As in Fig. 3, except using a constant 1 day averaging window as defined in
629 Fig. 8.

630 Fig. 11. As in Fig. 4 except using a constant 1 day averaging window as defined in
631 Fig. 8.

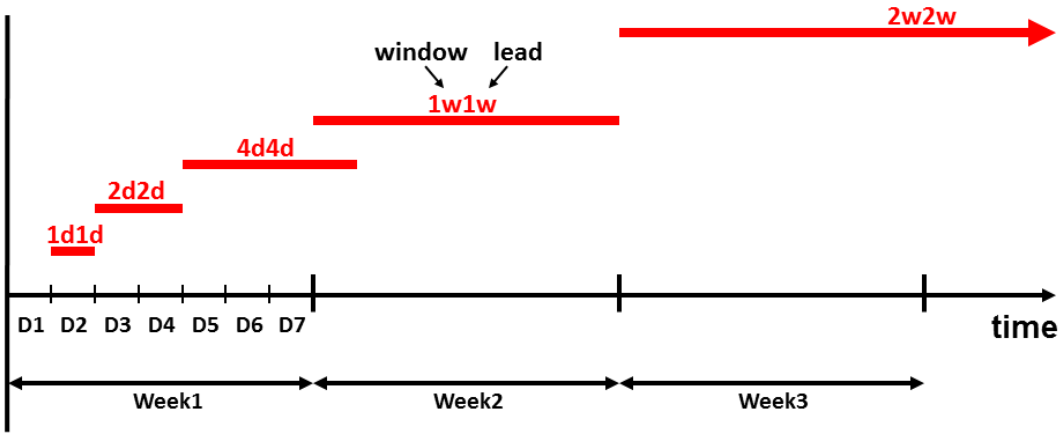

Fig. 1. Schematic of the time window and lead time definitions used in this analysis. The horizontal axis represents forecast time from the initial condition. "1d1d" refers to an averaging window of 1 day at a lead time of 1 day. Similarly, "2d2d" represents an averaging window of 2 days at a lead time of 2 days, and so on. Note that "1d1d" is what is usually called "day 2" in other papers, and "1w1w" is what is usually called "week 2".

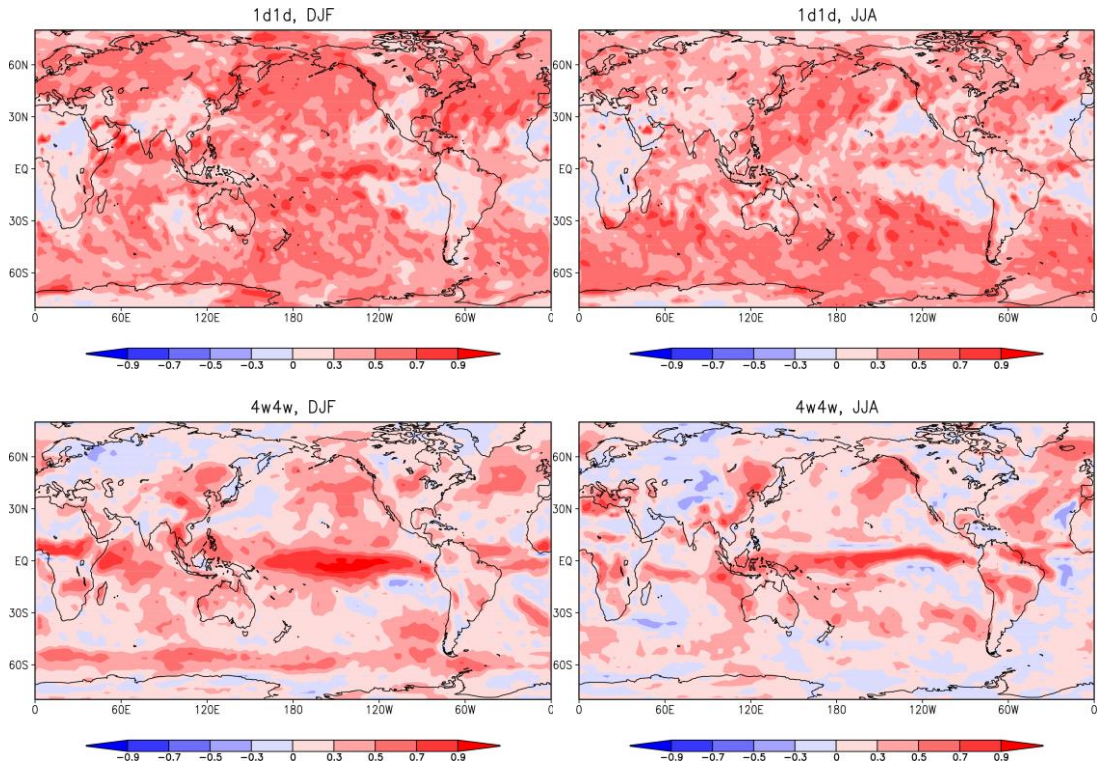

Fig. 2. Maps of CORt for model forecasts at 1d1d (upper panels) and 4w4w (lower panels). The left panels are for the DJF season and right panels are for JJA.

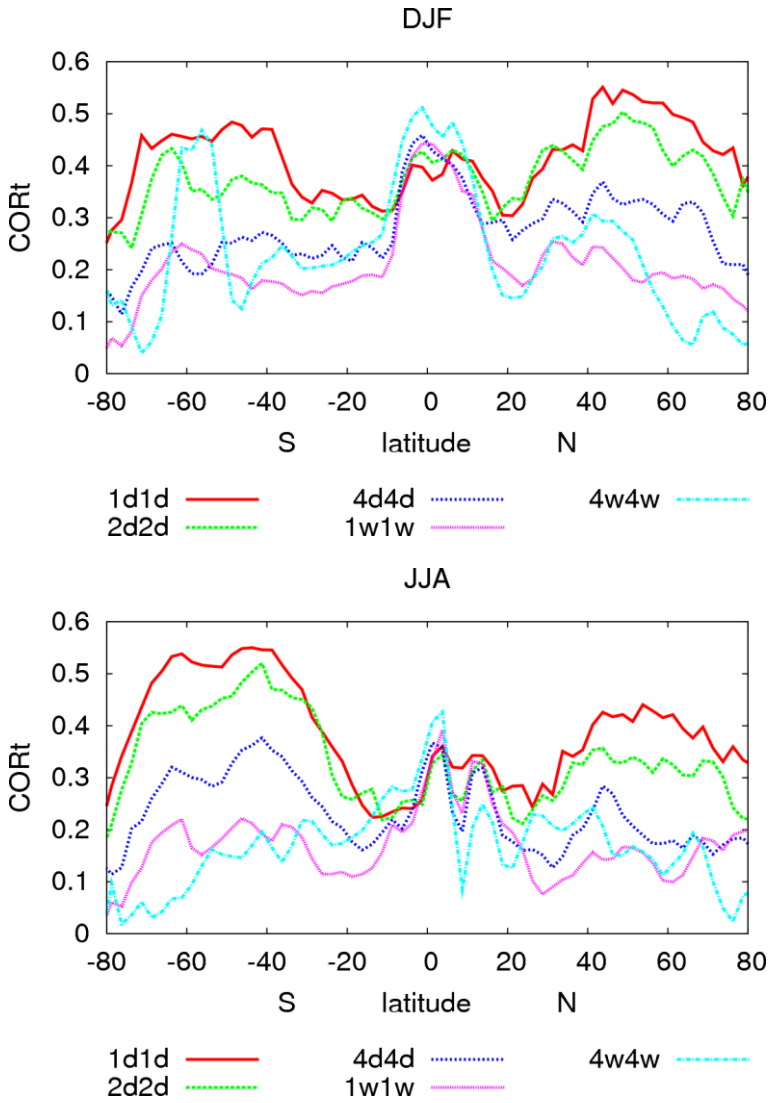

Fig. 3. Zonally-averaged CORt for model forecasts at different time window/lead combinations for DJF (top panel) and JJA (bottom panel).

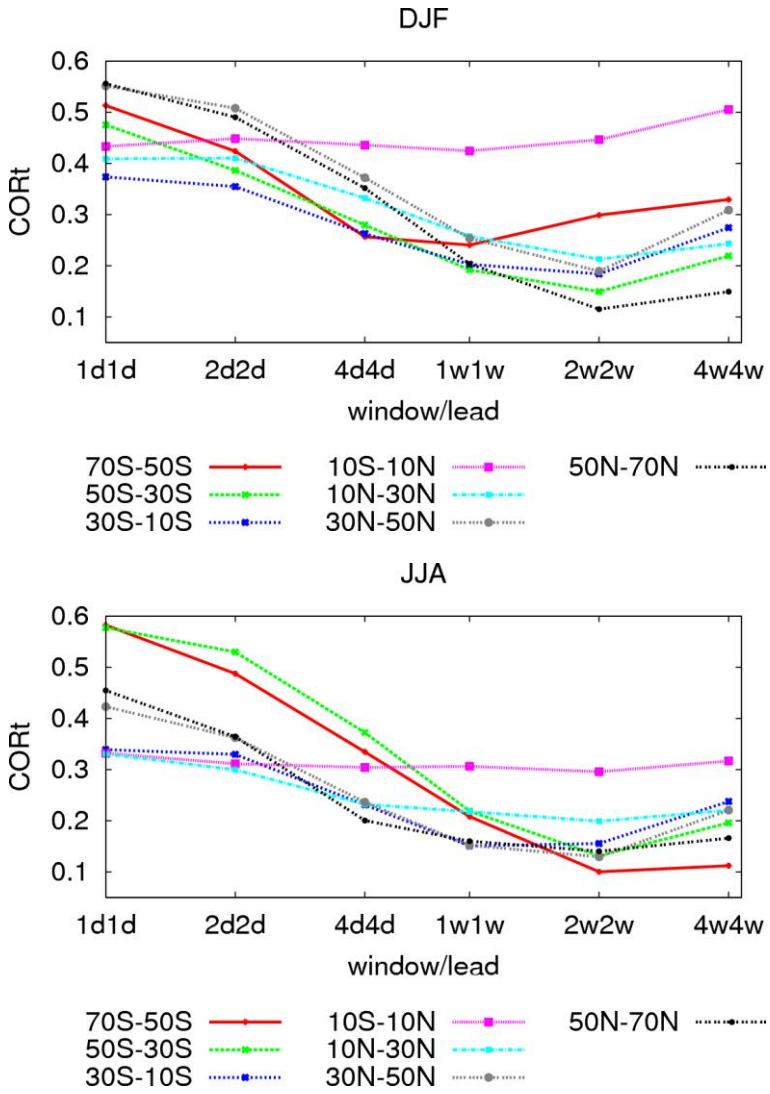

Fig. 4. Zonally-averaged CORt over specified latitude ranges versus forecast window/lead times from 1d1d to 4w4w for DJF (top panel) and JJA (bottom panel). Note that the spacing of the time scale is based on the logarithm of the lead time.

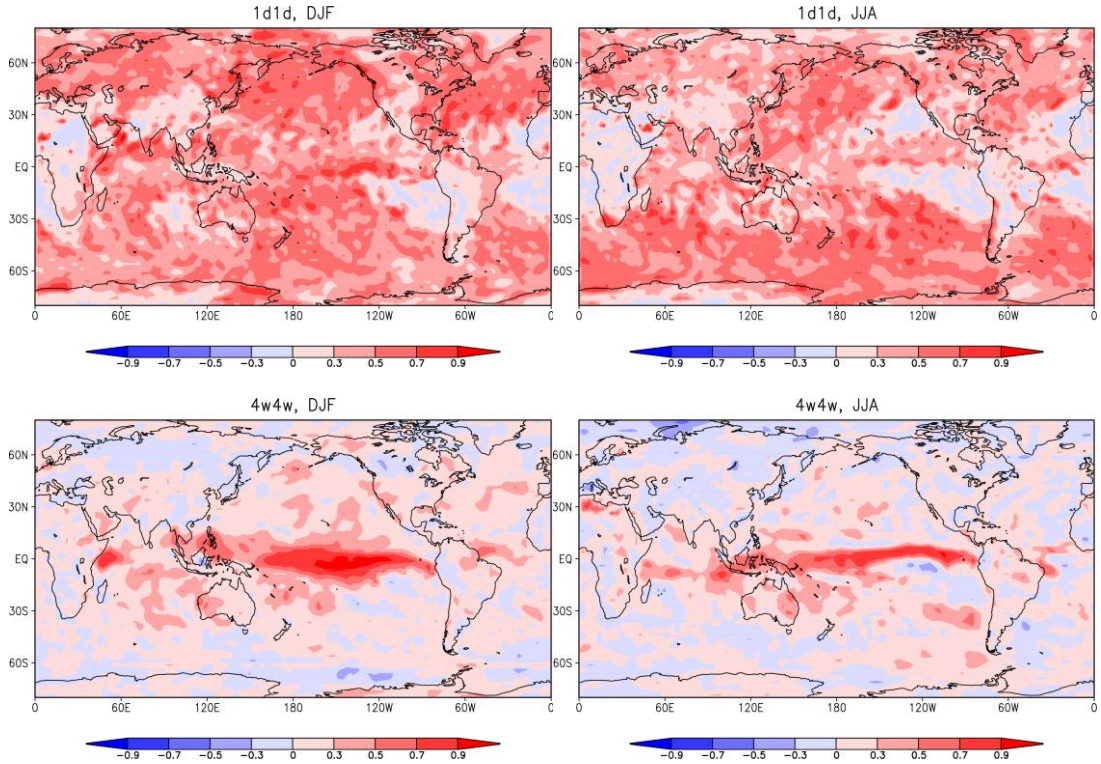

Fig. 5. As in Fig. 2, but for CORa.

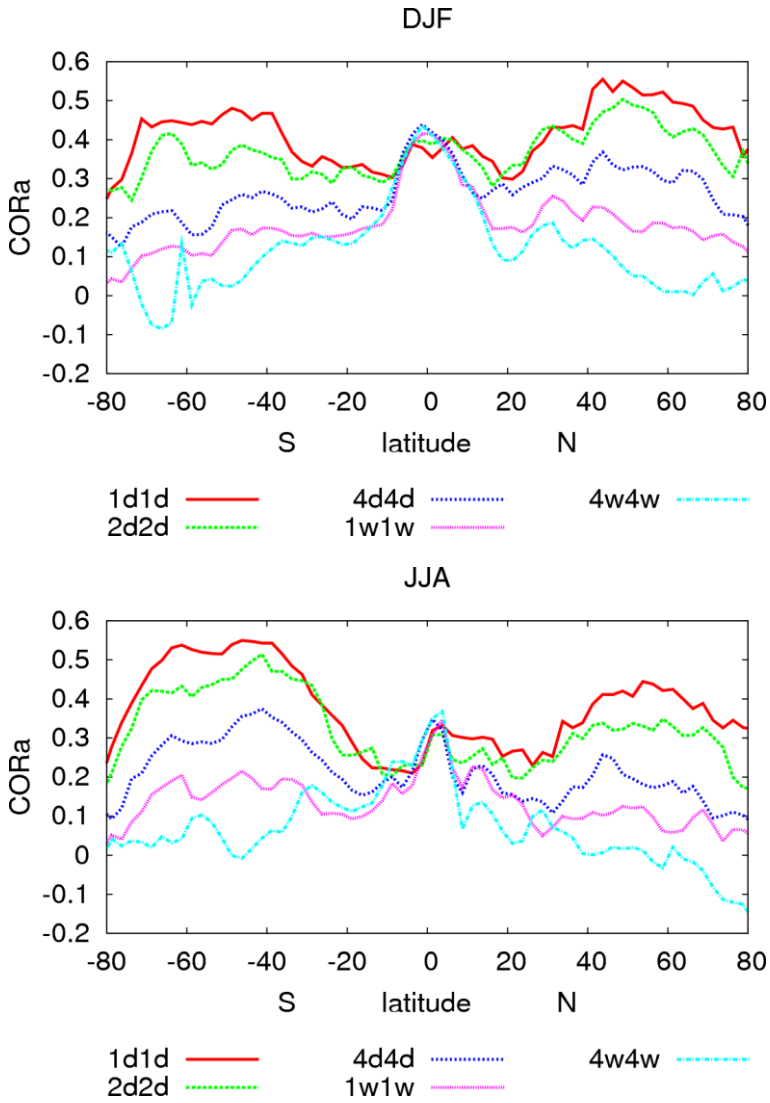

Fig. 6. As in Fig. 3, except for CORa.

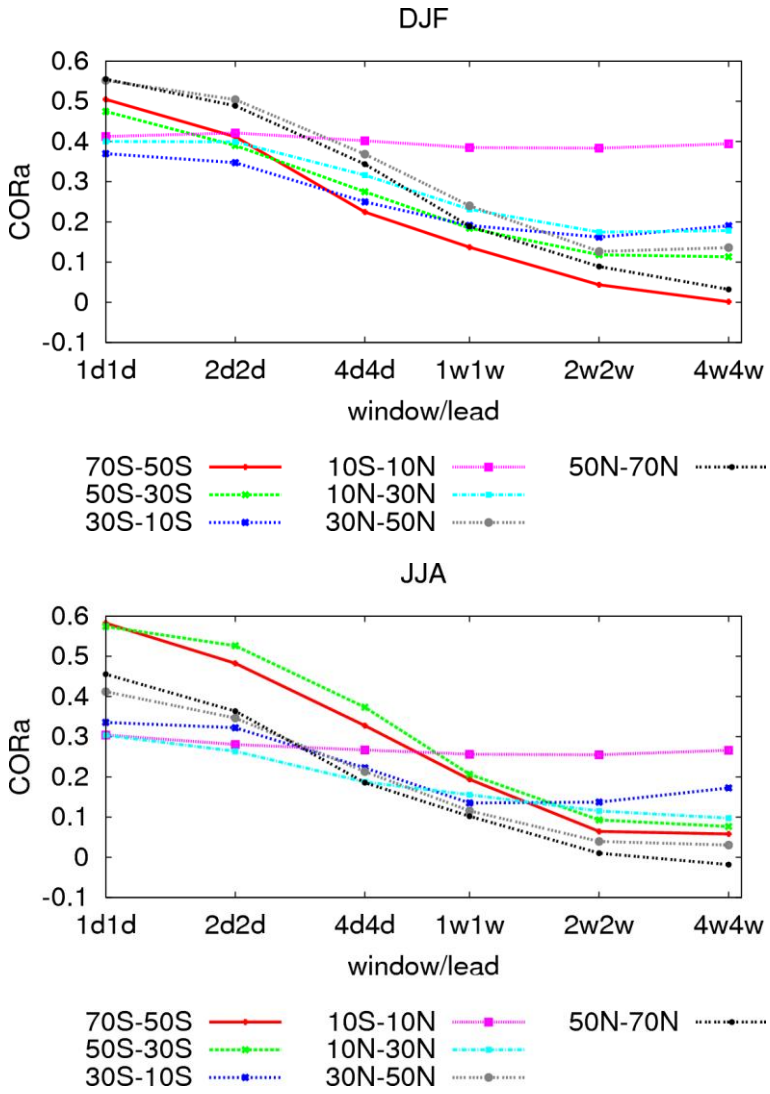

Fig.7 As in Fig. 4, except for CORa.

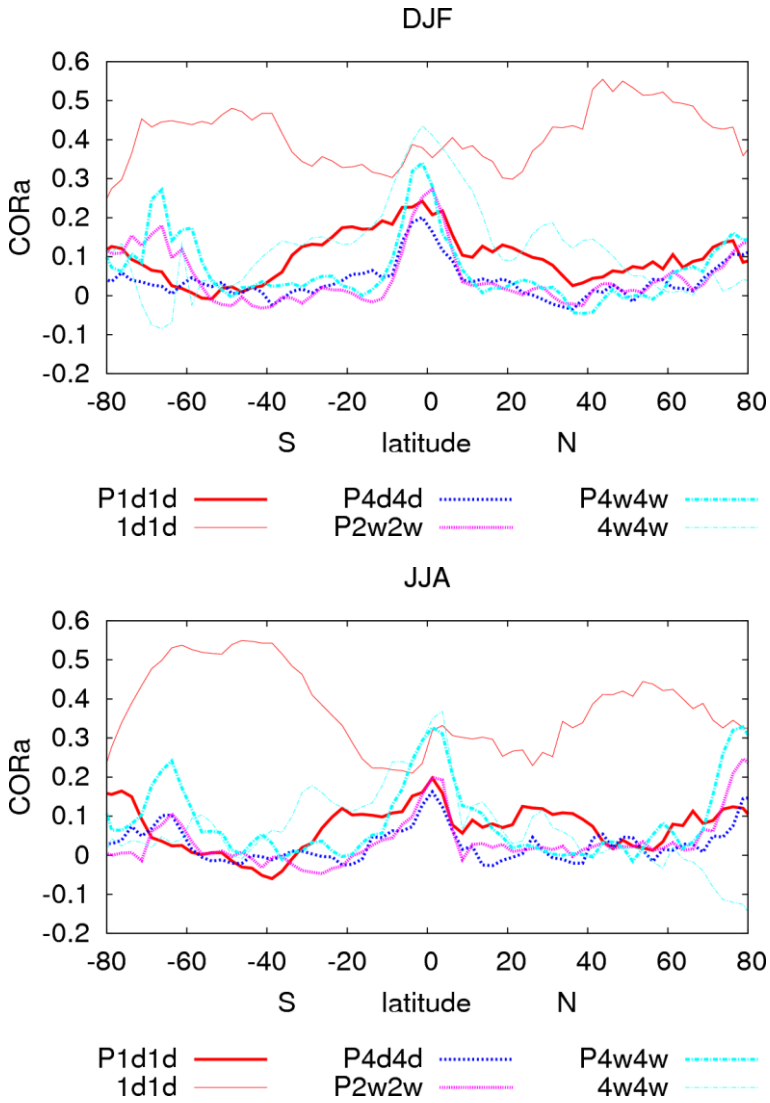

Fig. 8. As in Fig. 6, except showing zonally-averaged CORa for persistence forecasts (labelled as P1d1d, P4d4d, P2w2w, P4w4w) and the model forecasts of 1d1d and 4w4w for comparison (same as in Fig. 6).

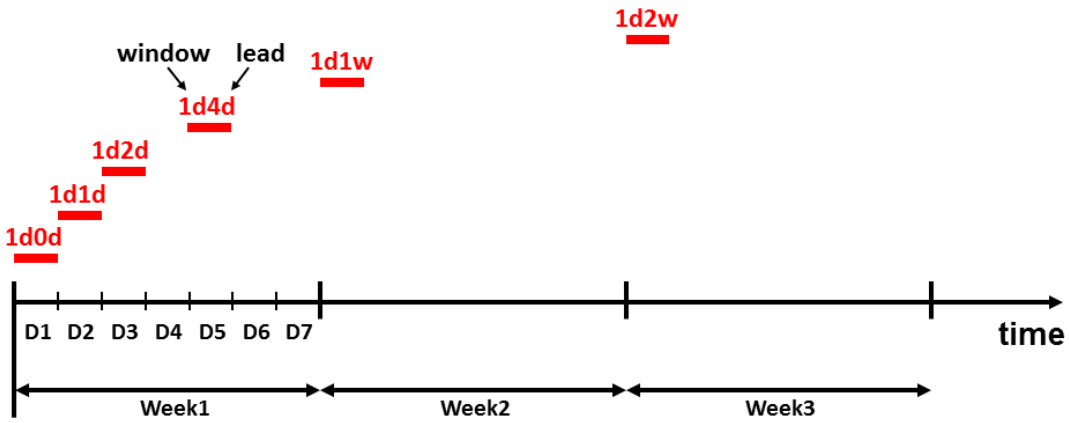

Fig. 9. As in Fig.1, except showing the window and lead time definitions used for the calculations in Figs. 10 and 11 (i.e. with a fixed 1-day averaging window).

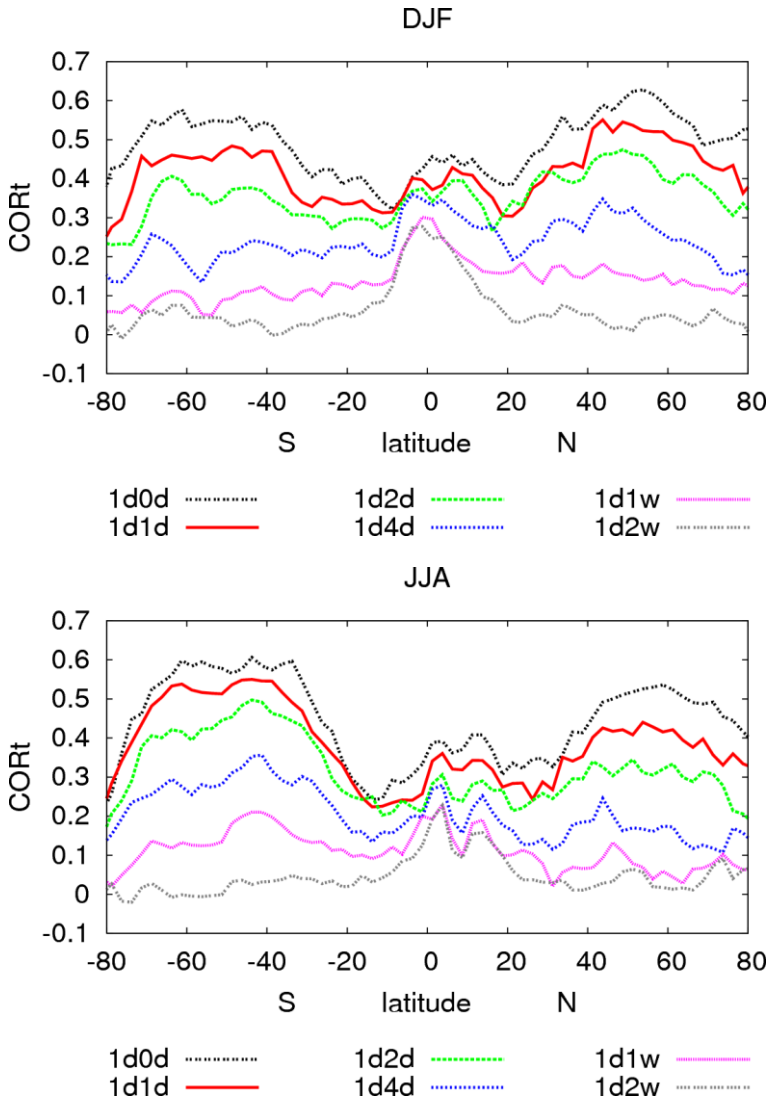

Fig. 10. As in Fig. 3, except using a constant 1 day averaging window as defined in Fig. 8.

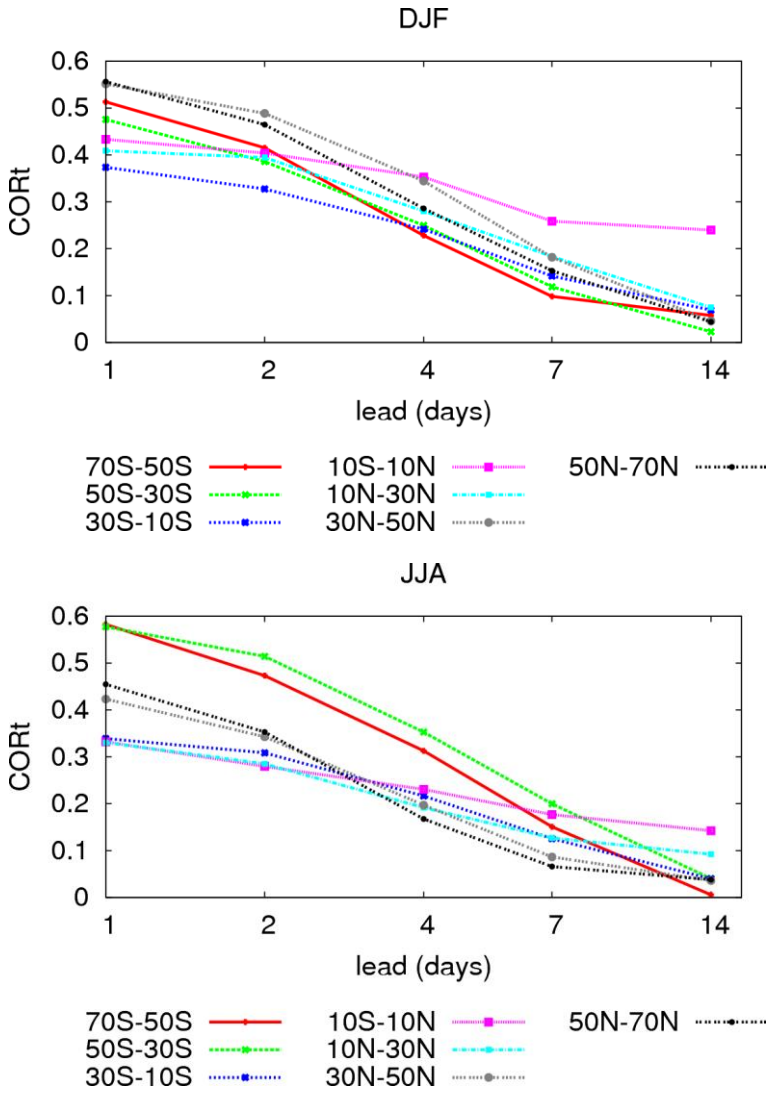

Fig. 11. As in Fig. 4 except using a constant 1 day averaging window as defined in Fig. 8.