Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

Sensitivity of CNRM-CM5 seasonal predictability to horizontal resolution

Michel Déqué, Jean Philippe Piedelievre and Eric Maisonnave

SPECS Technical Note 1

January 2014

Series: SPECS Technical Notes

A full list of SPECS technical notes can be found on our website:

http://www.specs-fp7.eu/SPECS/Dissemination.html

Contact: specs@ic3.cat

Index

1. Abstract

The standard version of CNRM-CM5 with a resolution of 150 km for the atmosphere and 1° for the ocean, is compared with a higher resolution version of the model, using 50 km for the atmosphere and 0.25° for the ocean. The study focusses on the seasonal predictability, up to month 7 in the tropical Pacific, and month 4 in the midlatitudes. The improvement in the prediction scores is modest, but certain. The four seasons are examined, and additional experiments (vertical resolution increase, atmosphere-only high resolution) are presented

2. Introduction

Climate modelling as well as weather prediction are scientifically possible, because the atmosphere as well as the ocean, obey, to a certain extent, to the law of fluid mechanics, which were described two century ago. Unfortunately this approach undergoes two limitations: i) the equations have no analytical solution in the general case, ii) they do not take into account the diabatic sources and sinks of water, energy and momentum. The computer revolution of the second half of the 20th century has allowed to solve partly the first point, by replacing the continuous problem by a discrete one which can be solved numerically. Of course, this simplification has complexified the second point, because the mean effect of the unresolved small scales has to be taken into account in the diabatic terms. These terms, often named “physics of the model” as opposed to the “dynamics of the model” rely, except for the radiation transfer, on empirical formulae adjusted from observation campaigns for the one part and by *a posteriori* calibrations for the other part. The diabatic terms are calculated independently for each vertical column at each time steps.

Here comes the question of the horizontal resolution, i.e. the accuracy at which the dynamics is calculated. There is a common belief that the higher the resolution, the better the numerical model. This has been proved for short-range meteorological forecasting. As long as the model resolution is coarser than the observation network, any resolution increase improves the initial condition, and thus the predictability which is related for the major part to the knowledge of the initial state.

For longer-range numerical prediction (beyond the 10-15 day deterministic predictability limit) and for climate modelling, the question is not as simple. The first atmospheric general circulation models were very coarse (Kasahara and Washington, 1967), because of the very limited computer power at that time, and the first resolution increase experiments (Manabe et al., 1970) yielded spectacular results. The improvements are brought by two causes

1) higher resolution makes the solution to the discretized dynamics closer to the actual solution to the continuous equations (the physics has no horizontal resolution because it works on independent columns)

2) the surface boundary conditions (orography, physiography) are distributed on our planet on scales extending from 1 cm to 1000 km

It is possible to evaluate the contribution of the first factor by using an aquaplanet model. WGNE has shown (D. Williamson, 1998, personal communication) that resolutions finer than 100 km do not improve the discrete solution with respect to the continuous solution. But one should keep in mind that the actual solution to the Navier Stokes equations is not the true atmosphere, but a human-built system designed to simplify and mimic the high complexity of nature.

Resolutions higher than 100 km have been proved very successful in the climate regionalization problem (Giorgi 1990). This problem is different from the role of high resolution question in climate modeling. The aim of a regional model is not to reduce the large-scale biases of its driving global model, but to represent with accuracy small-scale features that the global model is unable to mimic. For example the presence of a mountain will modify locally the direction of the wind and the distribution of precipitation in a more realistic way. The global climate modellers have been generally reluctant to devote the whole offer in computation power to the resolution increase: in 40 years, the mean resolution of the global climate models has passed from 1000 km to 100 km, whereas at the same time the mesoscale forecast models went from 100 km to 2 km.

Nevertheless a watch is necessary to be able to attribute the computation power to the relevant needs. The European Hiretycs project (Doblas-Reyes et al., 1998) has shown that increasing atmospheric GCM resolution from 300 km to 100 km had a neutral effect on the large-scale, and might be detrimental when some physical parameterizations are not re-calibrated. The literature about impact of horizontal resolution is very abundant and biased by the positive results. In the European SPECS project, the workpackage WP4.1 addresses the role of high horizontal resolution in seasonal and decadal predictability. We do not claim to bring in this report a definitive answer to this question, which depends on the model used. In addition, a progress when going from 300 km to 150 km does not necessarily imply a further progress from 150 km to 75 km

3. Models and experiments

The base model used here is a modified version of CNRM-CM5 (Voldoire et al., 2013). The ocean part is Nemo version 3.2 with 1° resolution (refined in latitude along the equator) and 42 levels. The atmosphere is Arpege version 5.1 with tl127 spectral truncation (1.4°) and 31 vertical levels. In section 5, we will introduce a vertical resolution with 71 levels by adding 40 levels in the stratosphere. The physical parameterizations (radiation and surface) are different from version 5.2 used in CMIP5. Version 5.1 is described in Déqué (2010).

The high resolution model is based on the same code, but slight differences may have occurred by the fact the the model has been run on a Bull computer (at CEA) whereas the low resolution model has been run on an IBM computer (at ECMWF). In section 6, the high resolution model is run on an IBM as well. The ocean part has a 0.25° horizontal resolution and 75 vertical levels. The atmosphere model uses a tl359 spectral truncation (0.5°). One considers that in a coupled system, the ocean must have a higher resolution than the atmosphere, because the main eddies, which transport heat from the equator to the pole, are of smaller horizontal size in the ocean than in the atmosphere. There are very few differences between the two Arpege versions, besides resolution: the time step has been changed from 30 min to 15 min, and the characteristic length of horizontal diffusion has been changed from 150 km to 50 km. In the case of Nemo, the changes are more substantial, because the two versions have been calibrated at different institutes, the 1° at Met Office then at ECMWF, and the 0.25° at Mercator.

The initial states of the atmosphere came from Era-interim (produced by ECMWF) available at tl255 (0.7°), then interpolated using a sophisticated extension of the Arpege-IFS code. For the ocean, we used Glorys at 0.25° and PSI2G2 at 1° (produced by Mercator). PSI2G2 is an ocean simulation with Nemo 1° nudged toward Glorys. Because of the period of the Glorys reanalysis, we are limited to the 1993-2009 period (17 years) which includes the successful predictions of the positive ENSO of winter 1997-98 and of the negative NAO of winter 2009-2010. Each year, four starting date have been selected: 1 Feb., 1 May, 1. Aug. and 1 Nov. . The corresponding forecasts will be referred to as Spring, Summer, Autumn, and Winter respectively. According to SPECS specifications, the forecast period covers 7 months. Ten members are generated by adding a tiny perturbation in the initial atmospheric situation. Fifty additional forecasts have been produced up to month 4, so that we have for the average month2-month4 (the traditional target in seasonal prediction) 60 equiprobable members (except in section 6 where we have only 30 members).

The high resolution simulations have been run in the framework of an EC-funded PRACE project named SPRUCE. Additional details on this project can be found in Maisonnave et al. (2013).

4. Results

4.1 Nino 3.4 scores

The greatest success of numerical weather forecasting is the prediction, up to 6 months in advance, of the warm and cold phases of the equatorial Pacific sea surface temperature (sst). The mostly used index is the mean temperature anomaly of the so-called Nino 3.4 area, corresponding to the rectangle 5°S-5°N by 170°W-120°W. Figure 1 shows that the Spring forecast (starting at 1 Feb), which has the least predictability, is degraded by the higher resolution whereas the Autumn forecast (starting at 1 Aug), which has the best predictability, is still improved. In the other two seasons, there is nothing to notice.

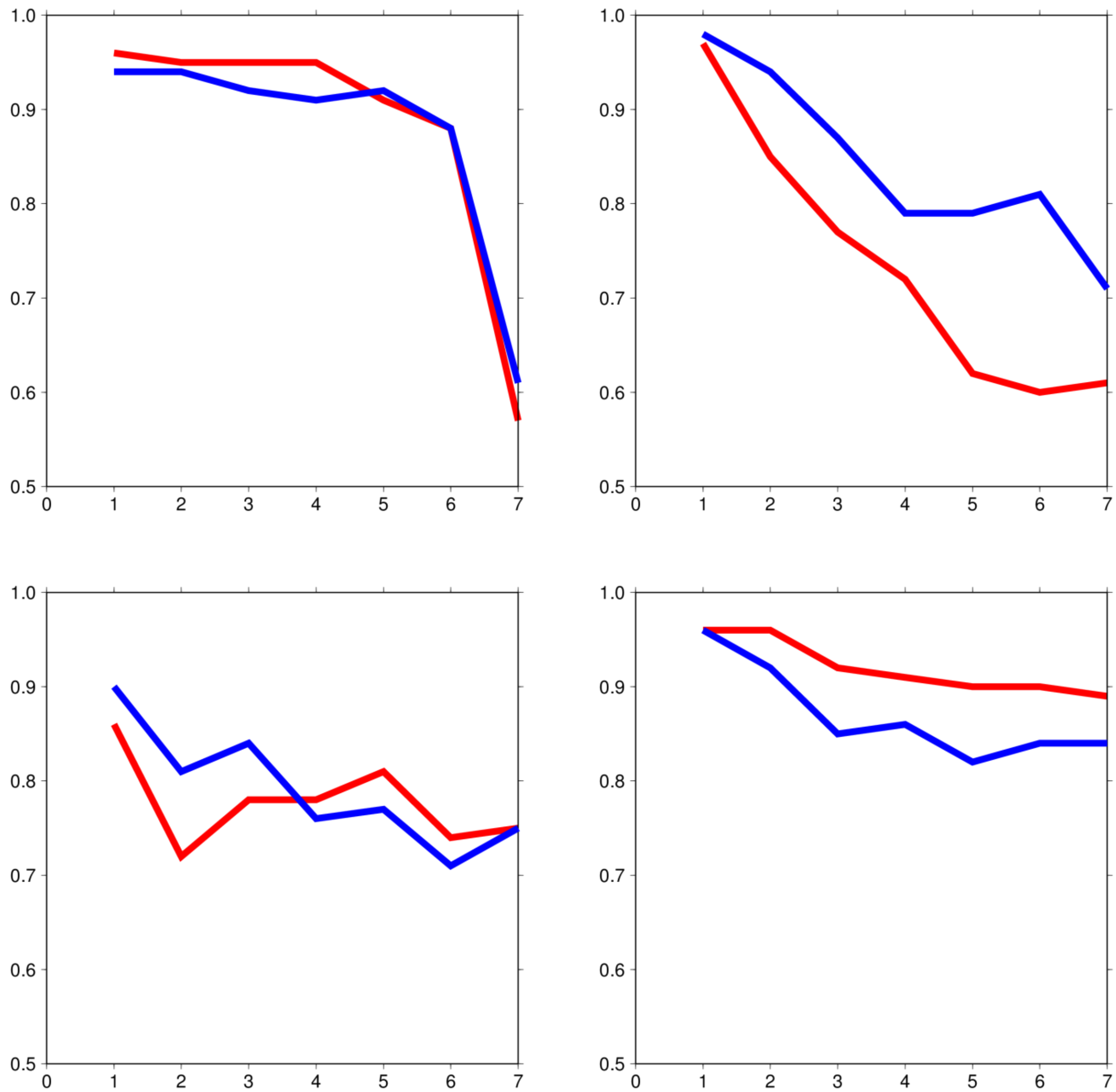

Figure 1: Time correlation between the predicted and observed Nino 3.4 monthly sst as a function of lead time (month); red curve high resolution and blue curve low resolution; top left Winter, top right Spring, bottom left Summer and bottom right Autumn

4.2 NAO index

The ENSO phenomenon has a strong impact on seasonal statistics of many parts of the world. This is not the case in Europe, where the ENSO response still exists, but is tiny compared to the year-to-year variability. Here takes place the North Atlantic Oscillation (NAO). Contrary to ENSO, NAO is not a physical phenomenon, because the alternation of positive and negative phases can be observed from on week to another, as well as from a decade to another. To take a medical comparison, NAO is a thermometer, not a disease. It corresponds to the enhancement or the weakening of the North-South pressure gradient in the northern Atlantic. Here we calculated it with an EOF analysis of the daily 500 hPa height over the Atlantic-Europe area for DJF, from 1979-2012 Era-interim data. We also tried low-pass filtered (beyond 6 days; see Doblas-Reyes and Déqué, 1998) data, without significant changes in the results, except a slight increase in percentage of variance of the first EOFs. The daily annual cycle has been subtracted. The first EOF explains 20% of the total variance. Its pattern is shown in Fig. 2. The second EOF explains 14%, so is clearly separated from the first one. The third EOF explaining 12%, the artificial mixing of EOF2 and EOF3 makes the analysis of the pattern of EOF2 irrelevant. A cluster analysis (Michelangeli et al., 1995), is better suited in this case. When applied to the other seasons, the EOF analysis (unfiltered as well as filtered) provides the same leading mode, but with a lesser percentage of variance and a northward shift of the positive belt. The scores in the other seasons of the first EOF are not significant, so we concentrate here on DJF predictions.

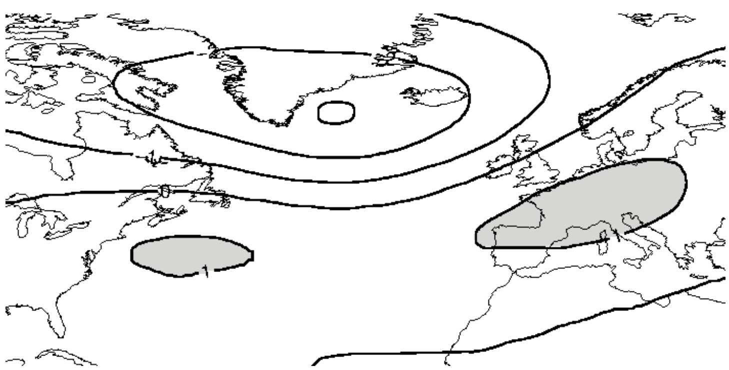

Figure 2: Spatial pattern of EOF1 of 500 hPa height in DJF (dimensionless units, contour interval 1, shading above 1); this pattern is used to calculate the NAO index

As shown in Table 1 (see explanation below), the time correlation is 0.46 for low resolution and 0.51 for high resolution. The question is: is this difference significant, given the small size of the sample (17 values). There are two possible ways of building pdfs or confidence intervals:

a) the Monte-Carlo method:

The NAO index for a given season is an average over space and time of geopotential height. It can be reasonably considered as a gaussian variable. Let us make the further assumption that (*F*,*V*) is a gaussian vector, where *F* is a forecast and *V* the corresponding verification. Then we can write:

$$F=\alpha+\beta V+Z \; . \qquad (1)$$

where α and β are the linear regression coefficients, and *Z* a gaussian variable independent of V, of mean 0 and variance $\sigma^2=\sigma_1^2+\sigma_2^2$. The first part of this variance is the inter-annual variance, the second part is the intra-ensemble variance. They are estimated with the 17 years and the 50 members respectively. Note that with this simple model, the intra-ensemble variance does not depend on the year (this is approximately the case with our data). If we consider an ensemble mean forecast, instead of an individual forecast, Eq. (1) is still valid, since it is linear, but the variance of the white noise *Z* is:

$$\sigma^2=\sigma_1^2+\frac{\sigma_2^2}{n} \qquad (2)$$

where *n* is the size of the ensemble. It is easy, thanks to a gaussian generator, to produce 5000 series of 17 pseudo forecasts of NAO index, and therefore 5000 random correlation coefficients.

If we replace in Eq. (1) the linear regression by the actual ensemble mean forecast for a given year, and set $\sigma_1=0$ in Eq. (2) we get a narrower distribution, because we get rid on the uncertainty about the time period 1993-2009 and take into account only the ensemble members sampling. This approach is more robust, but introduces a bias in the correlations: the random NAO forecasts are actual NAO forecasts plus a random noise, so their prediction skill is reduced. However, we will use it in the following for a better comparison with method b)

Table 1 shows, for *n*=10 (the SPECS recommended values) and *n*=60 (the actual value) the mean of this sample. Figure 3 shows the corresponding pdfs for the lower and higher resolutions.

b) the sub-sampling method:

A non-parametric way consists, for a given Winter, to select *n* members amongst the 60 individual members. With *n*=10, there are about 75 billion possibilities, and each year, another choice can be made, which rises the number of possible correlations to the power 17. If we produce 5000 drawings as in a) the estimates are not independent, because the same member can be used in two estimates. With *n*=10, there are only 6 possible fully independent sub-samples, which is not enough to estimate a pdf. This is not a problem for the mean, but as *n* increases, the pdf becomes artificially narrow. At the limit case *n*=60, there is a single possible sub-sample, which is the full sample: the pdf is therefore a Dirac function. Table 1 and Figure 3 show the results with this method. The conclusions of a) remain valid, but the pdfs are slightly shifted toward higher correlation values, and have a little less spread. Part of this spread reduction comes from the “consanguinity” of the random drawings.

	Monte-Carlo method		Sub-sampling method	
	Low-Res	High-Res	Low-Res	High-Res
10 members	0.33	0.39	0.34	0.40
60 members	0.42	0.48	0.46	0.51

Table 1: Mean time correlation of the NAO indices with different ensemble sizes, based on the pdf with two methods of generation. In the case of the sub-sampling method with 60 members (last two columns and last row), this mean is just the time correlation of the actual series, because a single sub-sample is available

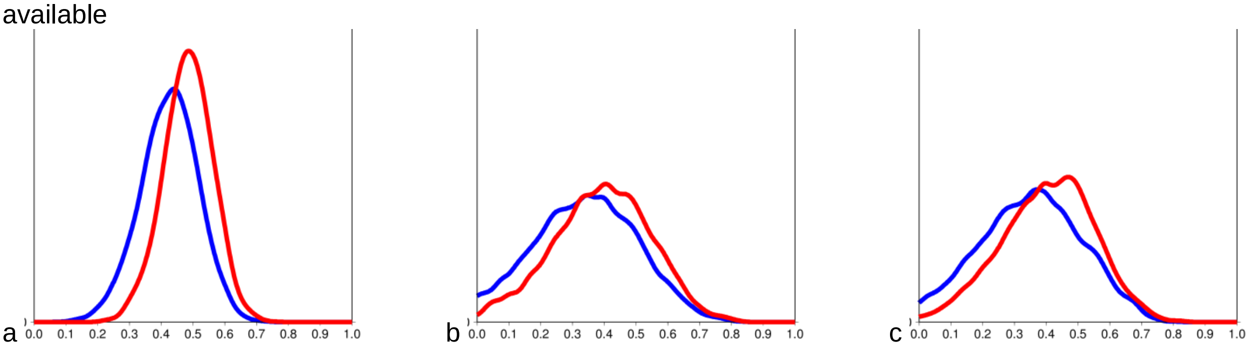

Figure 3: Pdf of the time correlation between the predicted and observed NAO indices for the high resolution (red curve) and the low resolution (blue curve): 60-member ensembles with Monte-Carlo method (a), 10-member ensembles with Monte-Carlo method (b), and 10-member ensembles with sub-sampling method (c)

Methods a) and b) agree on the fact that the time correlation increases by 0.06 when increasing resolution. But they lead to different pdfs. Method b) has the advantage to be non-parametric and to be applicable to the case of anomaly correlations between fields (see section 4.3). But as the size *n* increases, the spread becomes underestimated; the mean remains unbiased but poorly estimated. Method a) does not suffer from this drawback, but either we combine the uncertainty due to year-to-year variability and uncertainty due to member-to-member variability and we get a larger spread for the pdf, or we add a white noise to actual forecasts and we get a smaller mean for the pdf.

4.3 Large areas ACC

Seasonal forecasting is not just targeted for ENSO and NAO. In this section we will examine, for surface temperature, precipitation and 500 hPa height the sensitivity of robust scores with respect to horizontal resolution.

First, let us have a look at the mean prediction error. This quantity is easy to predict and is subtracted a posteriori from the prediction. Bias and predictability are two different concepts in numerical modelling, like the value of a function and the value of its derivative, but if the bias is reduced without any ad-hoc tuning (like we do here by increasing resolution) we can expect that the model is on the right way of converging toward reality. In addition, bias estimates, contrary to correlations, are statistically stable with 17 years of data.

Figure 4 shows that the winter bias of 500 hPa height is significantly reduced by the higher resolution, except over the North-East Pacific. Figure 5 shows that for precipitation, the bias is untouched, or even increased by a higher resolution. Figure 6 shows that for surface temperature, the bias is similar, except in the winter polar regions where it increases as resolution increases. Over most of the oceans, the predictions are 1°C over the observations.

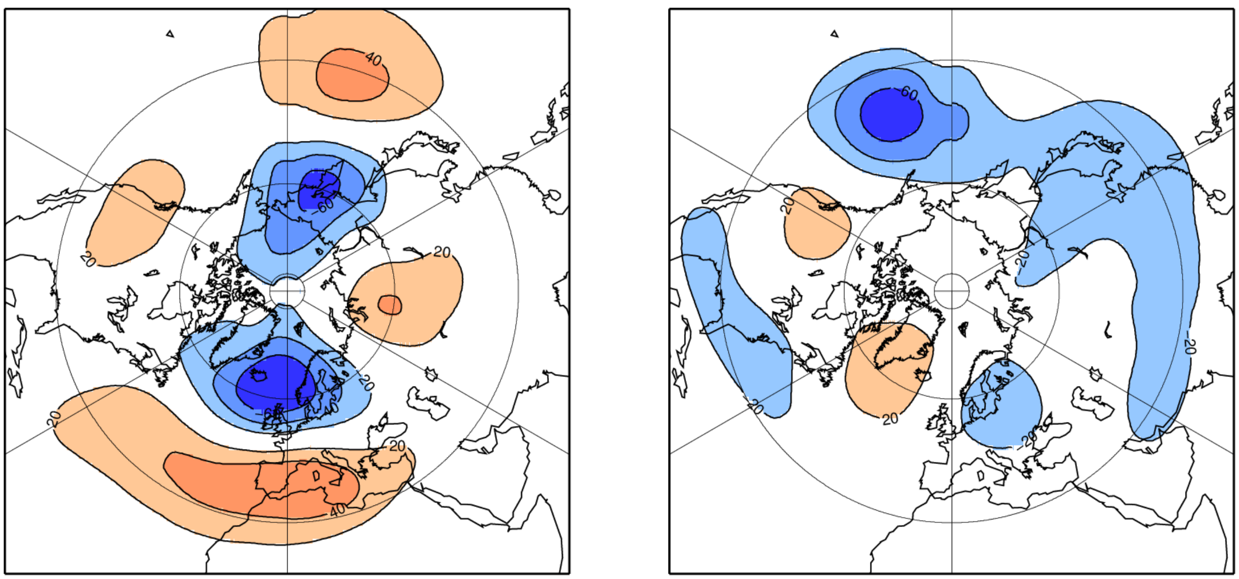

Figure 4: Systematic error of 500 hPa height in DJF (month 2-4) for low resolution (left) and high resolution (right); contour interval 20 m.

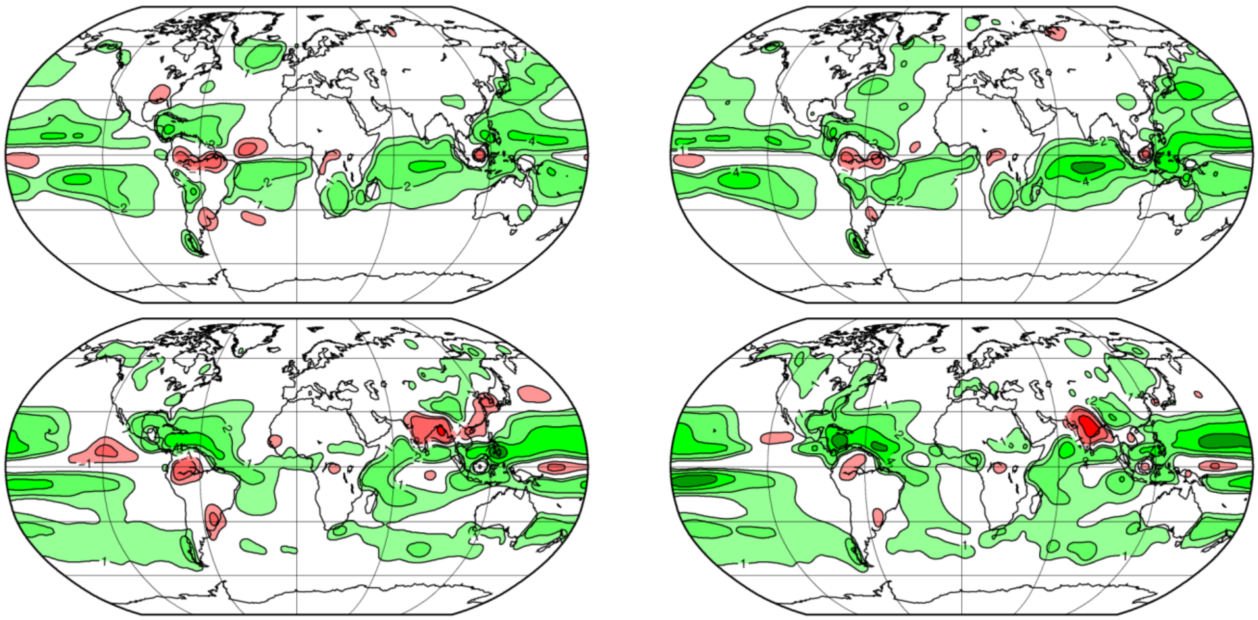

Figure 5: Systematic error of precipitation at month 2-4 for low resolution (left) and high resolution in DJF (top) and JJA (bottom); contours ±1 ±2 ±4 and ±6 mm/day

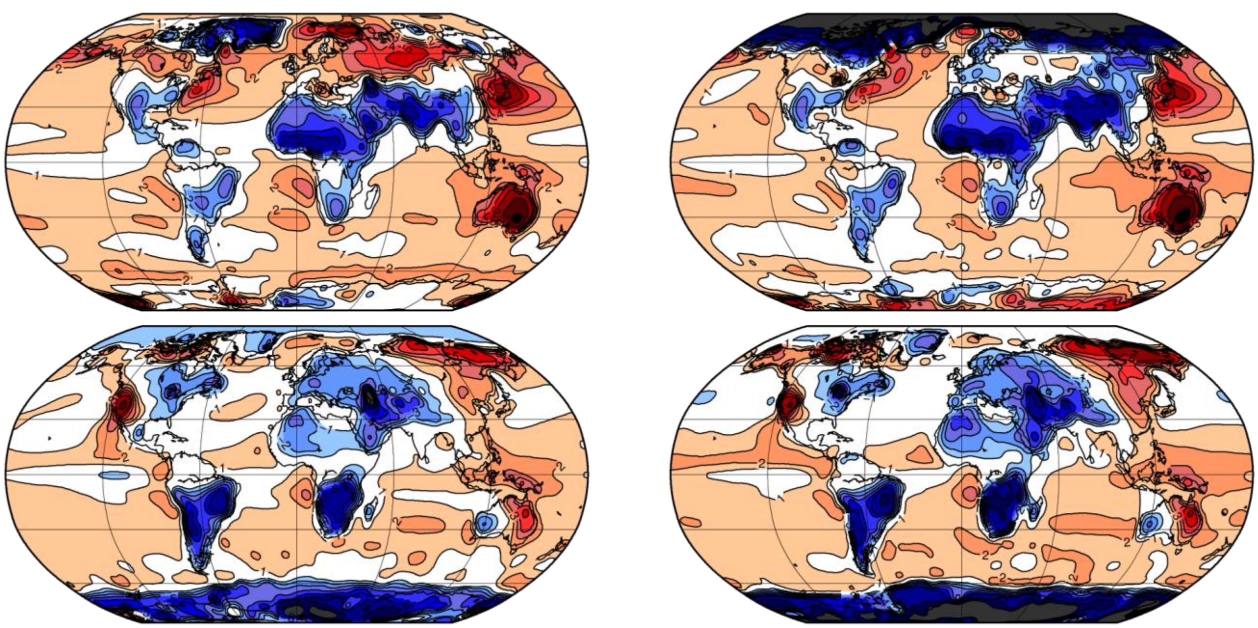

Figure 6: As Figure 5 for surface temperature; contour interval interval 1°C, zero contour omitted

Given the fact that local correlation estimates based on 17 cases are highly sample-dependent, except over the tropical Pacific where the actual correlation is high, maps similar to Figures 4-6 do not convey useful information. To get a relevant signal, and decide whether higher resolution has improved predictability, a simple and efficient solution consists of calculating averages over wide areas. A good partitioning of the globe is to separate the tropics (30°S-30°N) and the mid-latitudes (30°N-90°N and 30°S-90°S). In fact the so-called mid-latitudes include the polar regions, but weighted by the cosine of the latitude, their contribution is negligible. To get a better stability of the estimates, a "cooking recipe" consists of averaging the *Z*-transform of the correlation *r*:

$$Z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) \tag{3}$$

Z is close to *r* when *r* is small, but is huge when *r* is close to 1 or -1. Indeed an empirical correlation of 0.90 is much more significant than a correlation of 0.10, so this method of averaging favours the high positive or high negative correlations in the averaging process. See Déqué (2010) for a comprehensive review of scoring methods in the case of two time series.

This simple method has two limitations:

-the predictability depends on the year: a winter like 1997-98 has seen a big event and is particularly well predicted by most models

-the predictability is not a local property, but concerns also teleconnection patterns, e.g. NAO

So another metrics has been early proposed in long-range forecasting evaluation, named anomaly correlation coefficient (ACC):

$$ACC = \frac{[(F-\bar{F})(V-\bar{V})]}{\sqrt{[(F-\bar{F})^2][(V-\bar{V})^2]}} \tag{4}$$

where F is forecast, V is verification, overbar is time (i.e. multiyear) average, and square bracket is spatial average. So one can calculate an ACC for each year over a given area. When synthesizing this metrics over 17 years, simple time averaging is not a good idea. One can use the above Z-transform. One can alternatively use the method described in Déqué and Royer (1992), which has good properties of stability and relevance. It consists simply in averaging with respect to time the three moments (terms between square brackets) in Eq (4), and applying the square root and the ratio at the last step of the process. If the same principle is applied when spatially averaging the correlation *r*, then time average of ACC and spatial average of *r* yield the same value.

Table 2 provides, for the 4 seasons (month 2-4 averages), for precipitation, 500 hPa height and surface temperature and for the 3 parts of the globe, the mean ACC obtained with the two resolutions.

		30°N-90°N		30°N-30°S		30°S-90°S	
		LR	HR	LR	HR	LR	HR
MAM	Prec.	.18	.18	.47	.39	.17	.16
	Z500	.21	.26	.68	.64	.23	.18
	Tsur	.31	.32	.46	.42	.27	.23
JJA	Prec.	.05	.07	.40	.37	.17	.15
	Z500	.07	.17	.55	.55	.25	.27
	Tsur	.22	.27	.48	.46	.24	.19
SON	Prec.	.08	.10	.51	.52	.24	.22
	Z500	.08	.21	.61	.62	.46	.39
	Tsur	.23	.22	.52	.52	.25	.22
DJF	Prec.	.26	.24	.61	.61	.16	.21
	Z500	.37	.35	.74	.75	.31	.34
	Tsur	.35	.28	.52	.52	.32	.29

Table 2: Mean ACC of low resolution (LR) and high resolution (HR) over the 3 parts of the globe for precipitation, 500 hPa height and surface temperature; statistically significant superiority is shaded.

The scores are most of the time close to each other in Table 2. In order to determine whether the improvement or degradation is significant, a sub-sampling technique has been applied as in section

4.2: 95% confidence intervals and medians of 30-sized ensembles have been computed, and a difference in scores is considered as significant when each median is outside the other confidence interval. The result is shown by shaded numbers in Table 2. Out of the 72 cells of the table, 10 exhibit a superiority of the higher resolution, and 4 exhibit a superiority of the lower resolution. So one can synthesize by a small improvement of the scores when increasing in resolution. This result is in agreement with section 4.1 (ENSO) and 4.2 (NAO).

5. Vertical resolution

When increasing horizontal resolution in the ocean, we increased the vertical resolution as well, because we have little control on NEMO versions. In the case of ARPEGE, we have decided to keep constant the maximum of things when increasing horizontal resolution. But SPECS investigates the role of the stratosphere in WP4.3, and we can, thanks to the SPRUCE project, combine a high horizontal resolution, and a stratospheric version. A similar numerical experiment has been conducted for November starts only, with a version of ARPEGE with 71 vertical levels instead of 31. The 40 additional levels concern the stratosphere, and the tropospheric resolution is the same as in section 4.

In the case of ENSO prediction till month 7 (10 members only), Figure 7 shows that a higher stratospheric resolution has no impact on the scores. Changing the vertical resolution in the troposphere has not been considered here, since it would have implied a new calibration of the physics, which is a time and resource demanding exercise at such a resolution. The physical parametrizations concerning individual columns, one can, in a first approximation, keep them unchanged when increasing horizontal resolution. This is not the case in the vertical, at least in the troposphere, and in particular in the planetary boundary layer (first 1500 m above the surface). In unpublished studies, we have seen ENSO score improvements, at low horizontal resolution, when we increase the vertical resolution in the troposphere and the stratosphere jointly. It is speculated that increasing stratospheric vertical resolution needs a similar increase in vertical tropospheric resolution to be beneficial in term of ENSO scores, but this question is out of the scope of this study.

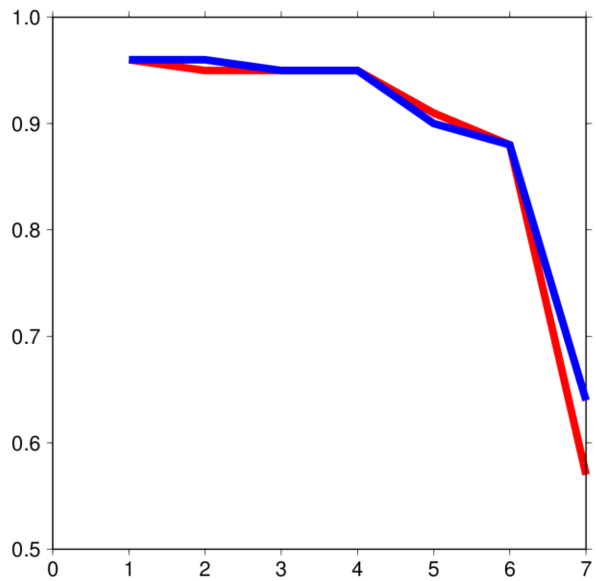

Figure 7: Time correlation between the predicted and observed Nino 3.4 monthly sst as a function of lead time (month, November to May); red curve high horizontal resolution and blue curve horizontal and vertical high resolution

As far as NAO index is concerned, the time correlation is 0.51 for the low vertical resolution (as shown in Table 1). With high vertical resolution, the score increases to 0.55. Figure 8a shows that there is a higher probability to get an NAO time correlation between 0.60 and 0.70 with 71 vertical levels than with 31 levels; but the jump in mean correlation is not very significant, even less than it was in Figure 3a. Figure 8b, when compared with Figure 4b, shows a reduction of the 500 hPa height bias in the northern hemisphere. The addition of a better stratosphere in a high horizontal resolution model is thus a further progress in the northern hemisphere winter circulation.

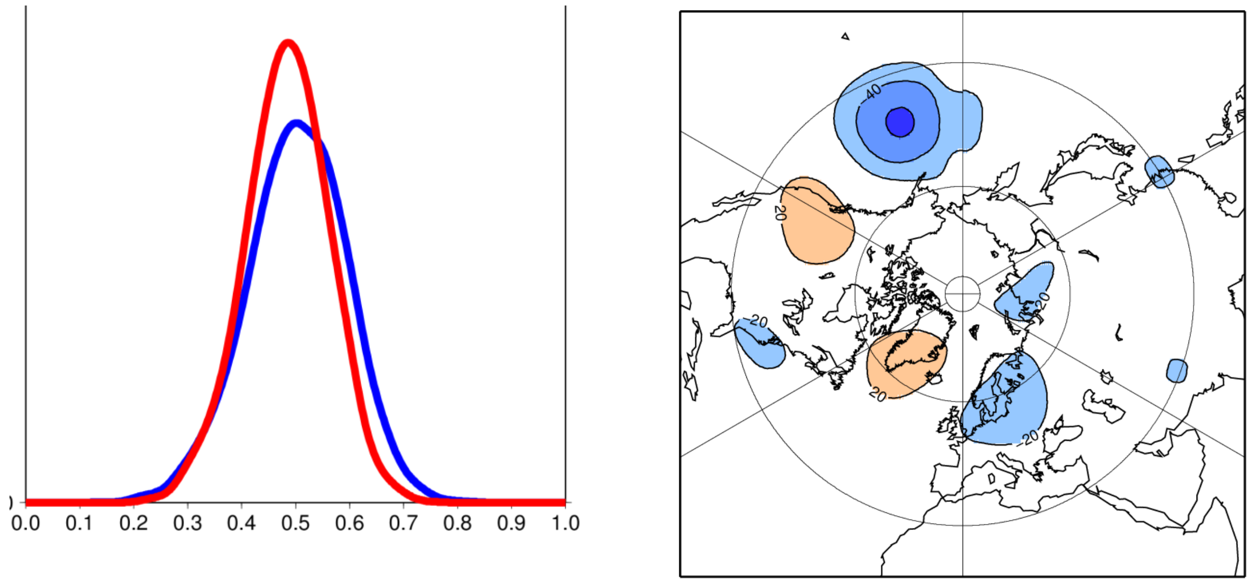

Figure 8: left panel : as Fig.3a for high vertical (blue curve) vs low vertical resolution (red curve); right panel: as Fig.4 for high vertical resolution

To achieve this question, let us look at the mean ACC scores for the 3 parts of the globe. Table 3, which partly resumes the results of Table 2 (LVR is identical to HR) shows that the scores are hardly modified, with a slight decrease in the southern mid-latitudes. It is interesting to note that the increase in NAO correlation and the reduction of Z500 bias are accompanied, as in section 4, by a decrease in NH Z500 MACC (0.37, then 0.35, then 0.34).

		30°N-90°N		30°N-30°S		30°S-90°S	
		LVR	HVR	LVR	HVR	LVR	HVR
DJF	Prec.	.24	.24	.61	.62	.21	.17
	Z500	.35	.34	.75	.73	.34	.30
	Tsur	.28	.28	.52	.52	.29	.27

Table 3: Mean ACC of low vertical resolution (LVR) and high resolution (HVR) over the 3 parts of the globe for precipitation, 500 hPa height and surface temperature

6. High resolution atmosphere-only

At this stage, we have seen that the improvement, if any, is modest. The question is, do we have a small contribution of both the atmosphere and ocean components, or do we have an improvement by one component and a degradation by the other. To test this hypothesis, we have repeated the experiment of section 4, but keeping the low resolution (and the associated oceanic starting conditions) in NEMO. ARPEGE is used at Tl359 with 31 vertical levels. Only 30 members have been generated for the 17 years and 4 seasons. The 60 members of the original high resolution are used, by sub-sampling, to evaluate the statistical significance of the score differences.

As far as ENSO is concerned, Figure 9, based on 10 members per forecast, shows that the Spring degradation we observed in section 4.2 was due to the high resolution in the ocean. On the other hand, the improvement we had in Autumn is shared between the atmosphere and the ocean, because the blue curve in Figure 9 is closer to the red curve than in Figure 1. In these two Figures, the red curve is the same and corresponds to high resolution in the atmosphere and in the ocean. One can also notice that in Summer, a higher resolution ocean degrades the scores, or in other words, that a higher resolution atmosphere increases the scores.

The correlation coefficient of the NAO index averaged over 30 members per forecast is 0.38. If we select 30 members out of the 60 members of the full high resolution exercise, we get a median of 0.48, and a 95% interval of [0.32-0.63]. Having in mind that the spread of this interval is underestimated by the “consanguinity” of the random drawings, we can state that this score degradation is not significant, and that a higher resolution ocean may improve the NAO prediction, or may not. The northern hemisphere bias (not shown) has a similar pattern as in Figure 4b, with positive anomalies over Greenland and California. But a uniform positive bias is superimposed, so that the systematic error is positive everywhere.

Table 4 shows the mean ACC with 30-member forecasts in both oceanic resolution. Over the northern midlatitudes, higher oceanic resolution improves the scores in MAM and JJA, but degrades in SON and DJF. In the tropics and in the southern mid-latitudes, higher oceanic resolution is followed, when significant, by a score increase.

One must remain very careful about the conclusions of this section. With a higher oceanic resolution:

1. the ENSO is degraded, but the tropical scores are improved
2. the NAO is improved, but the DJF northern geopotential height is degraded

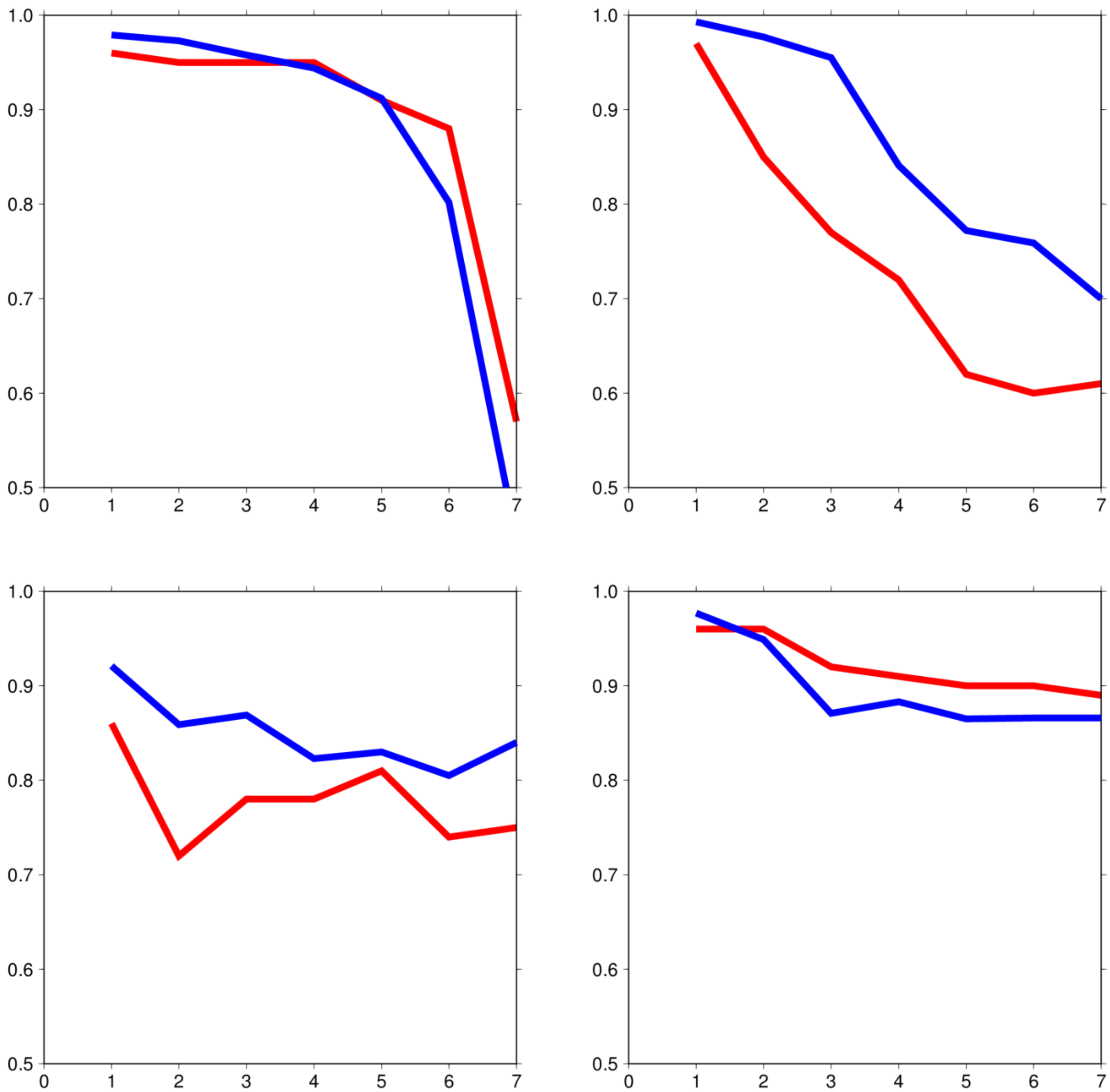

Figure 9: Time correlation between the predicted and observed Nino 3.4 monthly sst as a function of lead time (month); red curve high resolution and blue curve high resolution atmosphere-only; top left Winter, top right Spring, bottom left Summer and bottom right Autumn

		30°N-90°N		30°N-30°S		30°S-90°S	
		AHR	FHR	AHR	FHR	AHR	FHR
MAM	Prec.	.15	.17	.38	.39	.15	.16
	Z500	.25	.25	.65	.64	.30	.18
	Tsur	.29	.32	.42	.42	.23	.23
JJA	Prec.	.08	.06	.35	.37	.15	.14
	Z500	.06	.16	.52	.54	.16	.25
	Tsur	.21	.26	.45	.45	.21	.19
SON	Prec.	.08	.09	.47	.51	.21	.22
	Z500	.16	.18	.57	.61	.40	.38
	Tsur	.26	.22	.50	.51	.18	.22
DJF	Prec.	.26	.23	.52	.61	.16	.19
	Z500	.36	.34	.73	.75	.26	.32
	Tsur	.35	.27	.51	.52	.26	.28

Table 4: As Table 2 for atmosphere-only high resolution (AHR) and full high resolution (FHR). Mean ACC is calculated with 30 members; statistically significant superiority is shaded.

7. Conclusions

The good news is that, with a new forecast system which had never been tested or tuned, with a verification period which is unusual, our forecast scores are not degraded. This proves that our methodology of forecast evaluation is not biased by a practice tending to stress what is good and hide what is bad. One can even say that a higher resolution produces a slight improvement. But given the huge computation cost (in particular for the oceanic component), it is not reasonable to devote all resources to horizontal resolution. For example, going from 10 members at low resolution to 60 members at low resolution produces more improvement than going to 10 members at high resolution, and the cost factor is less (6 versus 20 and a perfect scalability). There are other ways to explore like the physics and the vertical resolution. It is hard to make a summary, because high resolution in the atmosphere and high resolution in the ocean may have counteracting effects. As shown with the NAO index, the 17-year verification period is too short to discriminate the small differences between the scores, although we have worked with big ensembles (60 members).

This weak conclusion is thus period-dependent, and studies based on the 20th century reanalyses (at least for the atmosphere) is a perspective for further studies. This conclusion is also, and for the major part, model-dependent, and SPECS is a good place to exchange the experience on that question

8. References

Déqué, M. and J.F. Royer (1992). The skill of extended-range extratropical winter dynamical forecasts. J. Climate 5, 1346-1356

Déqué, M. (2010). Regional climate simulation with a mosaic of RCMs. Meteorologische Zeitschrift 19, 259-266

Doblas-Reyes F.J., M. Déqué, F. Valero and D.B. Stephenson (1998). North Atlantic wintertime intraseasonal variability and its sensitivity to GCM horizontal resolution. Tellus 50, 573-595

Doblas-Reyes F.J. and M. Déqué (1998). A flexible bandpass filter design procedure applied to midlatitude intraseasonal variability. Mon Wea Rev 126, 3326-3335

Giorgi, F (1990). Simulation of regional climate using a limited area model nested in a general circulation model. J. Climate 3, 941-963

Kasahara, A. and W.M. Washington (1967). NCAR global general circulation model of the atmosphere. Mon Wea Rev 95, 389-402

Maisonnave, E., C. Cassou, L. Coquart, M. Déqué, N. Ferry, J.F. Guérémy, J.-P., Piédelièvre, L. Terray and S. Valcke (2012): PRACE Project Access for Seasonal Prediction with a high ResolUtion Climate modEl (SPRUCE) ,Working Note, WN/CMGC/12/29 SUC au CERFACS, URA CERFACS/CNRS No1875, France. Available at www.cerfacs.fr/globc/publication/technicalreport/2012/SPRUCE_WN.pdf

Manabe S., J. Smagorinsky, J.L. Holloway and M.M. Stone (1970). Simulated climatology of a general circulation model with a hydrological cycle. III: Effect of increased horizontal resolution. Mon Wea Rev 98, 175-212

Michelangeli P.A., R. Vautard and B. Legras (1995). Weather regimes: recurrence and quasi stationarity. J Atmos Sci 52, 1237-1256

Voldoire, A. E. Sanchez-Gomez, D. Salas y Mélia, B. Decharme, C. Cassou, S.Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave , M.-P. Moine, S. Planton, D. Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart and F. Chauvin, (2013). The CNRM-CM5.1 global climate model : description and basic evaluation , Climate Dyn. 40, 2091-2121

9. Acknowledgements

We acknowledge that the results in this paper have been achieved using the PRACE Research Infrastructure resource Curie based in France at TGCC (project ID 2012060982) and the European project SPECS (GA 308378).