

Horizontal resolution impact on
short- and long-range
forecast error

Roberto Buizza

Research Department

April 2010

To be published in Q. J. Roy. Meteorol. Soc.

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.



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

© Copyright 2010

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

ABSTRACT

The impact of horizontal resolution increases from spectral truncation T95 to T799 on the error growth of ECMWF forecasts is analyzed. Attention is focused on instantaneous, synoptic-scale features represented by the 500 and 1000 hPa geopotential height and the 850 hPa temperature. Error growth is investigated by applying a three-parameter model, and improvements in forecast skill are assessed by computing the time limits when fractions of the forecast error asymptotic value are reached. Forecasts are assessed both in a realistic framework against T799 analyses, and in a perfect-model framework against T799 forecasts.

A strong sensitivity to model resolution of the skill of instantaneous forecasts has been found in the short forecast range (say up to about forecast day 3). But sensitivity has shown to become weaker in the medium-range (say around forecast day 7) and undetectable in the long forecast range. Considering the predictability of ECMWF operational, high-resolution T799 forecasts of the 500 hPa geopotential height verified in the realistic framework over NH, the long-range time limit $\tau(95\%)$ is 15.2 days, a value that is one-day shorter than the limit computed in the perfect-model framework. Considering the 850 hPa temperature verified in the realistic framework, the time limit $\tau(95\%)$ is 16.6 days for forecasts verified in the realistic framework over NH (cold season), 14.1 days over SH (warm season) and 20.6 days over the tropics.

Although past resolution increases have been providing continuously better forecasts especially in the short forecast range, this investigation suggests that in the future, although further increases in resolution are expected to improve the forecast skill in the short and medium forecast range, simple resolution increases without model improvements would bring only very limited improvements in the long forecast range.

Key words: numerical weather prediction, model resolution, error growth, predictability.

1 Forecast error growth and predictability limits of a numerical weather prediction system

The first estimate of forecast error growth was given by *Charney et al* (1966), who, using a simple model, concluded that a reasonable estimate of the forecast error doubling time was 5 days. Few years later, *Smagorinsky* (1969), using a more refined, 9-level primitive equation model that contained also moist processes, gave a lower estimate of the forecast error doubling time of 3 days. *Lorenz* (1982), analysing 500 hPa geopotential height forecasts from the operational model used at ECMWF at that time, a 15-level primitive equation model with moist processes and orography, further reduced the estimate to 1.85 days for the Northern Hemisphere (NH) winter. This estimate was obtained assuming that forecast error grows following a quadratic equation that was introduced earlier by *Lorenz* (1969), which assumes that the short-term errors grow linearly, while the long-term growth is dominated by a (negative) quadratic term that brings the long range, forecast error to an asymptotic value. Subsequently, *Leith* (1982) included in *Lorenz* (1969)'s forecast-error growth model a constant term designed to simulate the effect of analysis errors on the short range forecast error. *Leith* (1982)'s forecast error growth model was further modified by *Dalcher and Kalnay* (1987), who investigated both the chaotic and systematic component of model error, and modified the model to simulate the effects of non-linear error saturation. *Dalcher and Kalnay* (1987) applied the newly modified model to the data analyzed by *Lorenz* (1982), and introduced two measures of predictability, the **time limits** $\tau(95\%)$ and $\tau(50\%)$, that are the times when the forecast error reaches 95% or 50% of the forecast error asymptotic value. They suggested that the forecast error doubling time, that was used up to that time to assess predictability, was not a very good measure of error growth, but that better measures

were the forecast error growth rate, and time limits such as $\tau(95\%)$ and $\tau(50\%)$. Using the same data set of *Lorenz* (1982), they estimated that the forecast error growth rate was about 0.43 d^{-1} (which corresponds to an error doubling time of about 1.7 days for the NH winter (a shorter value than the 1.85 days found by *Lorenz* in 1982), longer for the long-waves and shorter for the short waves. They concluded that, for the NH winter, the time limit $\tau(95\%)$ was 12 days, and the time limit $\tau(50\%)$ was 5.5 days.

Savijarvi (1995) applied the *Dalcher and Kalnay* (1987)'s model to assess the error growth of forecasts issued by the National Meteorological Centre (NMC) of Washington between 1988 and 1993. *Savijarvi* (1995) introduced a third time limit, $\tau(71\%)$, the time when the forecast error exceeds $1/\sqrt{2}$ of the saturation value [see *Savijarvi* (1995) for more details on the definition of this threshold, which corresponds to the level of climatic variability], and named it the deterministic predictability time limit. He estimated that for the 500 hPa geopotential height during the 2-year period 1991-1993, the NMC forecasts reached this level at about day 9 for the large scales (i.e. waves with total wave number up to 3), and about day 3 for the small scales (i.e. waves with total wave number between 3 and 20). He indicated that if the model error and the initial condition error could be halved, this limit would be reached at day 10 for the large scales and at day 6.6 for the short scales. *Savijarvi* (1995) also estimated that the time limit $\tau(95\%)$ for the NMC system in 1991-1993 was longer than what was given by *Dalcher and Kalnay* (1987), about 20 days for the large scales and 7 days for the short scales, and suggested that this limit will remain at these levels even if the model error and the initial condition error could be halved. *Reynolds et al* (1994) also used the *Dalcher and Kalnay* (1987)'s model to forecasts to investigate the three-dimensional structure of random error growth in the NMC system for the same period, and concluded that both the model and analysis uncertainties contributed to forecast error growth.

More recently, *Simmons and Hollingsworth* (2002) used the *Dalcher and Kalnay* (1987)'s model to study forecast skill improvements of the ECMWF single, high-resolution forecasts. They showed that between winter 1996/97 and winter 2000/01, the forecast error doubling times during the NH winter decreased by $\sim 30\%$ in the short forecast range (say at forecast days 1-4). At days 1 and 2, for example, they quoted doubling times in winter 2000/2001 of 1.14 and 1.42 days, values that are $\sim 50\%$ shorter than the ones quoted by *Lorenz* (1982). The shortening of the doubling times at forecast day 1 and 2 can be related to the more realistic model activity at all scales, due to the increased resolution of the forecast model and to the inclusion of more realistic schemes that simulate physical processes. From Fig. 11 of *Simmons and Hollingsworth* (2002), it is possible to estimate that for the ECMWF system in winter 2000/01, the time limits $\tau(71\%)$ and $\tau(95\%)$ were about 9 and 13 days, respectively.

This work presents an updated assessment of the forecast error growth of synoptic scale features represented by instantaneous (as opposed to averaged over a period of time, e.g. 1 day) ECMWF forecasts of the 500 and 1000 hPa geopotential height and the 850 hPa temperatures. As in *Reynolds et al* (1994) and *Simmons and Hollingsworth* (2002), the *Dalcher and Kalnay* (1987)'s model of forecast-error growth has been used. Through the years, increases in resolution, especially in horizontal resolution, has brought improvements in forecast accuracy. This work provides a clean (clean since forecasts have been produced using different horizontal resolutions but the same model version for the same period) assessment of the impact of horizontal resolution on the short and the long range forecast skill. In particular, the predictability times when the forecast root-mean-square error reaches different levels of the asymptotic limit (25%, 50%, 71% and 95%) is computed. To

better understand the role of model uncertainty on forecast skill, error growth is studied not only in a realistic framework, with forecasts verified against analyses, but also in a perfect-model framework, using high resolution forecasts as verification. The key questions addressed in this work are: how much do the time limits change when resolution is increased from T95 to T799? And how do the past estimates of the so-called predictability time limits compare with these more recent estimates based on winter 2007/08 forecasts?

After this introduction, Section 2 describes the methodology used in this study, Section 3 discusses the impact of resolution on forecast-error growth, and finally some conclusions are drawn in Section 4.

2 Methodology

In this section, the experimental design and the different configurations used in this study are described, then the data-sets used and the verification methods followed are discussed, and finally the forecast-error growth model used to estimate the predictability limits is illustrated.

2.1 Experimental configuration

Although the focus of this investigation is the error growth of single forecasts, the results are based on ensemble experiments that include the control forecast, that starts from the (unperturbed) analysis interpolated at the ensemble resolution, and 4 perturbed members that start from initial conditions perturbed using singular vectors (the same methodology used in the ECMWF operational EPS to construct the initial perturbations has been used; see *Palmer et al 2007* for a recent review). The main reasons why 5-member ensembles instead of single control forecasts are used are the following. First, in the perfect-model framework, the control forecast would have very low initial-condition errors since it starts from the interpolated analysis, and this would make the comparison of results obtained in the perfect-model and the realistic framework problematic. By contrast, the initial uncertainty of the perturbed members is the same in both frameworks. Secondly, results based on the average forecast error growth of the 5 perturbed members are statistically more significant than results based on single control forecasts. Ideally more members would have been included in the ensembles, but unfortunately running more members would have made it impossible to cover a 3-month period, and it was decided that it was more important to sample a whole season rather than increasing the ensemble membership. Moreover, the 5 members have been used to define an ensemble-mean forecast, and investigate the impact of resolution changes on the ensemble-mean forecasts.

Forecasts have been run using the ECMWF model with 62 vertical levels and with six different horizontal resolutions with spectral truncations: T95, T159, T255, T319, T399 and T799 (more precisely, these ensembles have been run with these spectral truncations and their corresponding linear grids in physical space). Vertical resolution has not been varied because a clean comparison of the impact of vertical resolution could not be easily made, since physical parameterization schemes are very sensitive to the number of vertical levels, and each model cycle is usually designed to work with only a limited number of vertical coordinate systems. Considering, for example, the model cycle used in this experimentations, the only two vertical resolutions fully tested and supported were with 60, 62 and 91 vertical levels, with the 91 vertical level version having a higher top of the model (0.01 hPa instead of 0.1 hPa). Thus, a straightforward assessment of the impact of increasing vertical resolution by a factor of about 8, comparable to an increase of horizontal resolution from T95 to

T799, was not possible. Although the ECMWF model has been designed to be able to run with different horizontal resolutions, it is worth recalling that there are some components of the model that depend on horizontal resolution, such as the model orography and the sub-grid orography parameterization scheme (*Brown 2004*); climate files (albedo, land-sea mask, vegetation, surface drag); the adjustment time of the convection scheme, with shorter times used for higher resolution (*Bechtold et al 2008*) and horizontal diffusion.

The 5 ensemble members include the control forecast, which starts from the T799L91 analysis interpolated at the configuration resolution, and four perturbed members, which start from symmetrically perturbed initial conditions. For all resolutions, the perturbed initial conditions have been defined as in the ECMWF operational EPS (see *Palmer et al 2007* for a recent review of the system), by adding and subtracting to the control analysis a perturbation defined by a linear combination of the 50 fastest singular vectors (SVs, *Buizza and Palmer 1995*). The ECMWF operational SVs are computed with spectral triangular truncation T42, and that the initial perturbations are scaled by comparing the SVs with a T42 version of the analysis error estimate provided by the ECMWF 4D-Var system. By construction, the same initial perturbations are used for all ensemble resolutions. It is also worth stressing the fact that since the same analysis is used in all experiments (albeit interpolated from the operational T799 resolution to the experiment resolution), this study does not take into account any potential benefit that increases in the horizontal resolution used in data-assimilation could bring. In this respect, the forecast error sensitivity to horizontal resolution could have been stronger if the lower resolution ensembles were started from a lower resolution analysis.

The perturbed members have been run with the *Buizza et al (1999)* stochastic scheme that simulates the effect of random model errors due to the physical parameterization schemes: at each time step, the scheme adds a stochastic perturbation to the model tendencies. The time-scale of the stochastic perturbations changes with the resolution, with shorter correlation times used when the model resolution increased. This time dependence simulates the belief that as resolution increases, models are supposed to become more accurate, and thus the contribution of the unresolved scales/processes simulated by the stochastic scheme should decrease. The reader is referred to *Buizza et al (1999)* for a description of the stochastic scheme, and for a more detailed discussion on the assumptions behind the time-correlation values used in the scheme. Although stochastic schemes can alter the model climate (see, e.g., the discussion in *Palmer et al 2009*), there have been no indications that the *Buizza et al (1999)* stochastic scheme used in our simulations alters the model climate. Recently, for example, *Jung et al (2005)* studied the impact of two different stochastic schemes on North Pacific weather regimes simulated by T95 model integrations performed with the ECMWF model cycle 26r3 (this cycle was used in operations in 2004). They concluded that the *Buizza et al (1999)* scheme had little influence on the frequency of occurrence of North Pacific weather regimes, while a different stochastic scheme, the backscatter scheme developed by *Shutts (2005)*, led to substantial improvements. Thus, we do not expect that the stochastic scheme could have affected in a resolution-dependent way the model climate of the different ensembles.

Table 1 lists the key characteristics of the different ensemble experiments (resolution, time-step, stochastic physics time-scale), and Figure 1 is a schematic of the ensemble configuration used to run all these experiments.

Configuration	95	159	255	319	399	799
Spectral truncation N	95	159	255	319	399	799
Grid spacing (km at the Equator)	210	125	80	65	50	25
Time step Δt (s)	3600	3600	1800	1800	1200	720
Time-scale stochastic physics (h)	6	6	3	3	2	1.2
CPU time / CPU time (399)	0.02	0.05	0.27	0.43	1	6.68

Table 1. Experiments main characteristics, including the CPU time required to complete 1-day forecast, expressed in units of CPU time required to complete 1-day forecast at T399. The grid spacing has been defined as the size of half a wavelength of the shortest zonal wave resolved by the model at the equator, $L = \pi a/N$, where a is the earth radius (Laprise 1992). The CPU time has been computed assuming that it is proportional to the square of the truncation and is inversely proportional to the time step $CPU(N, \Delta t) = \left(\frac{N}{399}\right)^2 \frac{1200}{\Delta t}$.

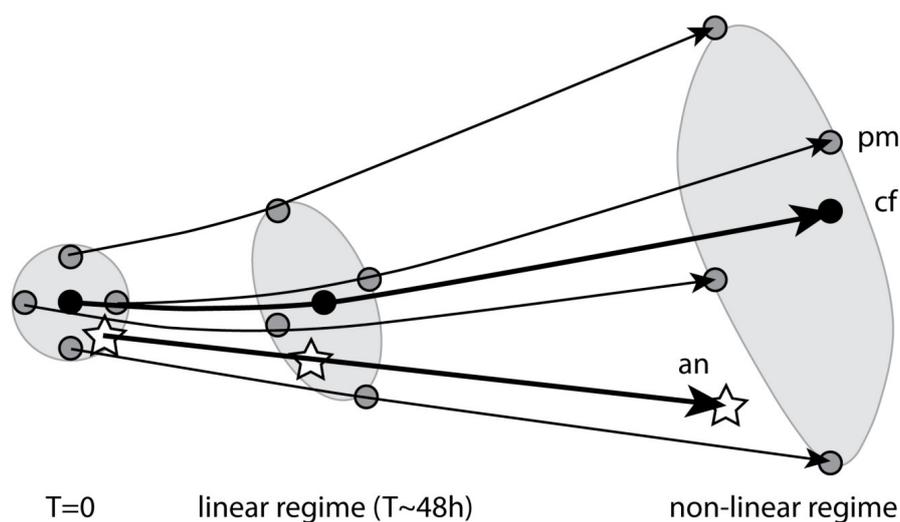


Figure 1. Schematic of the experimental design. The black circle denotes the control (cf), the grey circles denote the perturbed members (pm), and the star denotes the analysis (an).

2.2 Data sets and verification approaches

Forecasts have been run for 90 days from 1 December 2007 to the 28 February 2008, with initial time 12 UTC and a 15 day forecast length. The ensemble membership and the forecast length had to be limited to 5 members and 15 days to be able to complete the experimentation using a reasonable amount of computing resources (the computing time used to complete the experimentation is equivalent to the forecast time required to produce about 16 years of single 10-day T399L62 forecasts, or 2.5 years of single 10-day T799L62 forecasts). Forecast errors have been analyzed for three variables that are normally used to describe the synoptic scales: the 500 and the 1000 hPa geopotential height (Z500, Z1000), and the 850 hPa temperature (T850). Forecasts and analyses have been compared on a regular latitude-longitude grid with 2.5° degree spacing over three areas: the Northern Hemisphere extra-tropics (NH, which includes points with latitude north than 20°N), the Southern Hemisphere extra-tropics (SH, which includes points with latitude south of 20°S) and the

tropics (TR, which includes points with latitude between 20°S and 20°N). Fields have been interpolated from the reduced Gaussian grid to the regular latitude-longitude grid using a bi-linear interpolation method. For reason of space, most of the results will be shown in terms of Z500 over NH, and T850 over NH, SH and the tropics.

As mentioned in the introduction, root-mean-square errors have been assessed in two different frameworks: a realistic framework, in which forecasts have been verified against ECMWF analyses, and a perfect-model framework, in which they have been verified against T799 control forecasts. The T799 analysis that has been used as verification is the ECMWF operational analysis that was generated in D07-JF08 by the ECMWF 12-hour 4-dimensional variational assimilation system. Although using an analysis instead of observations to verify a forecast has the disadvantage that the short-range forecast error might be underestimated, it provides a better coverage of the whole globe for all variables. For the ECMWF system, for example, the average root-mean-square error (rmse) over NH of the T799 analysis of Z500 is estimated to be about 5m, which is about 50% the average rmse of the 1-day T799 forecast and about 25% of the 2-day forecast, which, for the period that has been investigated, are respectively 10.6m and 20.1m). Thus, using analyses instead of observations might lead to substantial (say up to 25%) error underestimation only up to forecast day 2.

In the perfect-model framework, it has been assumed that there are only two sources of forecast error: errors due to the use of a resolution lower than T799, and, in the case of the perturbed members, errors due to initial uncertainties and to model errors represented by the *Buizza et al* (1999) stochastic scheme that simulates model uncertainties due to physical parameterizations. The comparison of the perfect-model and realistic results will highlight the contribution to forecast error of the model error component that has not been simulated (i.e. the use of a lower resolution and the simulation of stochastic model error due to parameterized physical tendencies). As pointed out in section 2.1, the components of the ECMWF model that are resolution dependent are set to make the model climate insensitive to resolution.

2.3 Forecast-error growth model and predictability limits

Forecast error growth has been studied applying the *Simmons et al* (1995)'s version of the error growth model of *Dalcher and Kalnay* (1987), who included in the *Lorenz* (1982) model both the systematic and the random error components, and non-linear error saturation.

Accordingly to this model, the time evolution of the forecast error E is given by the following equation:

$$\frac{dE}{dt} = \alpha E - \beta E^2 + \gamma \quad (1)$$

where the forecast error E of Eq. (1) is the average root-mean-square forecast error (see Appendix A for an explicit definition of the way the average error has been computed). As in *Simmons et al* (1995), Eq. (1) has been written in a discretized form, and the three parameters (α, β, γ) have been estimated by a least-squares fit of the root-mean-square differences between consecutive forecast errors ΔE_j for $j=1,17$, with $t_1=0$ and $t_{17}=360$ hours. Once (α, β, γ) have been determined, the analytical solution of Eq. (1) can be computed (see Appendix A for a more detailed description of the model, and its analytical solution). As in *Bengtsson et al* (2008), forecast error curves have been extrapolated beyond 360-hours.

Following *Dalcher and Kalnay* (1987), the predictability time limit $\tau(95\%)$ has been defined as the time when the forecast error equals 95% of the asymptotic value E_∞ , i.e. when $\frac{E(t)}{E_\infty} = 0.95$, where

$E(t)$ is the analytical solution of Eq. (1). Also, as in *Savijarvi* (1995), the time limit $\tau(71\%)$, i.e. the time when the forecast error equals $1/\sqrt{2}$ of the asymptotic value E_∞ has been computed. Furthermore, the time limits $\tau(25\%)$ and $\tau(50\%)$ have been computed to assess the sensitivity of the short and medium range forecast skill to resolution.

The forecast-error growth model has been applied to the whole (i.e. including both the systematic and random components) forecast error as in *Lorenz* (1982)'s original work and in *Savijarvi* (1995), and not just to the random component of the forecast error, as was done by *Dalcher and Kalnay* (1987). As *Savijarvi* (1995) pointed out, there are two reasons why this approach should be followed: firstly, the error-growth model fit to the whole forecast error data is very good, and secondly, the systematic error is an essential part of the model error that should be taken into considerations when studying forecast predictability. A good fit of the model to the data has been found also in this work (see Section 3), with the forecast error growth curves showing very similar asymptotic values (for the perturbed forecasts, the asymptotic values of resolutions T95, T159, T255, T319, T399 and T799 in the realistic framework are, respectively, 150, 147, 146, 145, 144 and 145 m). The forecast error growth model has been used to extrapolate the error curves beyond the 15 day forecast range for which forecast data are available. Furthermore, it makes it possible to normalize the forecast error curves computed for different model resolution using the appropriate forecast error asymptotic value, thus providing a more correct assessment of the forecast error sensitivity to horizontal resolution.

3 Impact of resolution on forecast error growth

In the first part of this section, the impact of a resolution increase from T95 to T399 on forecast error during the first 15 days is discussed, both in the perfect-model and in the realistic frameworks. This analysis is based on 90-day average rmse computed directly from the experimental data. Then, in the second part of this section, the forecast-error growth model is fit to the experimental data, and the coefficients (α, β, γ) are computed for all resolutions and in both frameworks. Finally, the forecast-error growth curves are then used to assess the impact of resolution on the medium and long range predictability.

3.1 Forecast error computed from the experimental data

Figure 2 shows the average rmse for the control, the ensemble-mean (defined as the average of the control and the 4 perturbed members) and the perturbed members, with errors computed in the perfect-model and realistic frameworks for Z500 over NH. In the perfect-model framework, there is a clear indication of the benefit of increasing the resolution for the whole 15-day forecast range, with stronger benefits detected in the case of the control and the ensemble-mean. The difference between the rmse of the perturbed members run at different resolutions is smaller than the difference between the rmse of the corresponding control forecasts because they all have similar initial condition errors (forecast run at higher resolution start closer to the T799 analysis, see Table 2), while the control forecasts run with higher resolution has much smaller initial errors. For each resolution, the rmse of the ensemble-mean is the lowest due to the filtering effect of the averaging operator. The results

obtained in the realistic framework indicates a clear benefit of increasing resolution from T95 to T255, but a negligible impact of any further increase. As was the case for the results obtained in the perfect-model framework, the impact is more evident for the control and ensemble-mean forecasts.

Configuration	95	159	255	319	399
Initial error control	1.15	0.54	0.27	0.20	0.17
Initial error perturbed member	3.18	3.01	2.98	2.97	2.96

Table 2: Average initial error of the control and perturbed members for Z500 over NH.

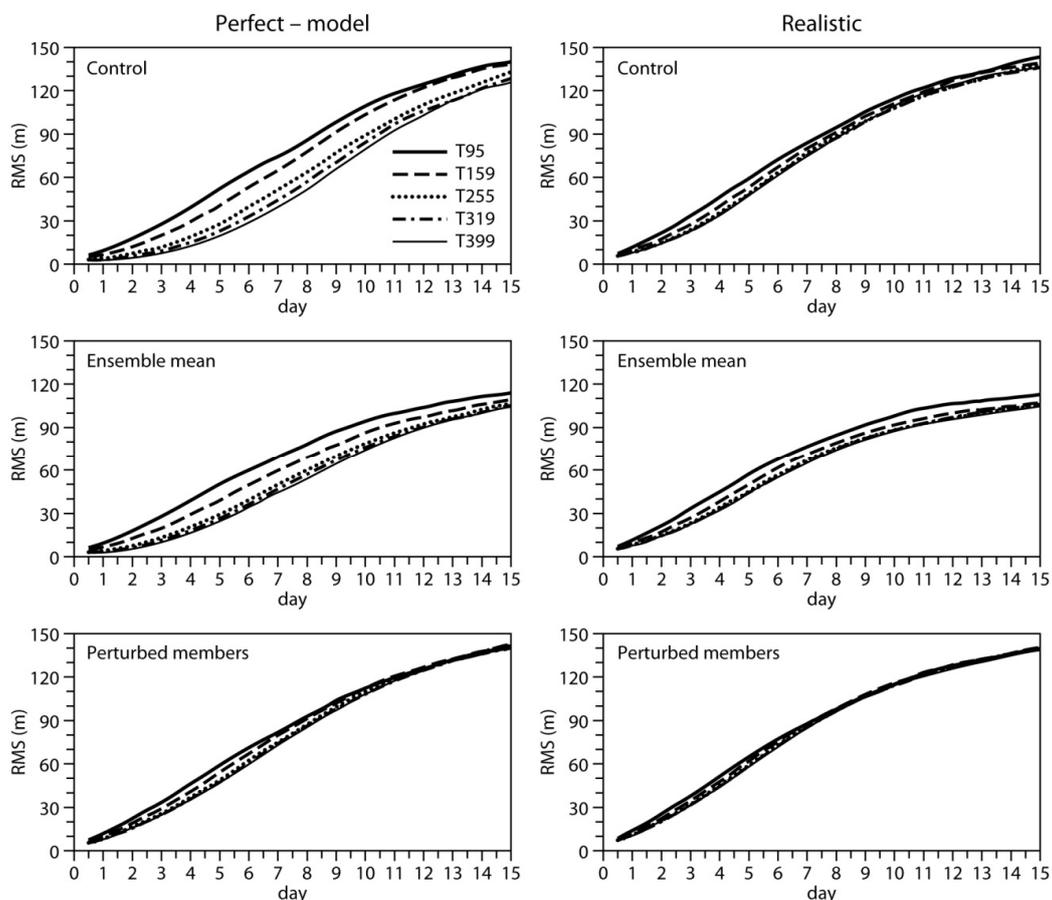


Figure 2: 90-day average root-mean-square-error (RMS, in meters) computed for Z500 over NH. Top panels: rmse of the control forecast in the perfect-model (left) and the realistic (right) framework run with resolutions T95 (solid), T159 (dashed), T255 (dotted), T319 (chain-dashed) and T399 (thin solid). Middle panels: as top panels but for the ensemble-mean. Bottom panels: as top panels but for the perturbed members.

For any two ensemble configuration with resolution T_x and T_y , the rank-sum Mann-Whitney-Wilcoxon non-parametric test $RMW(T_x, T_y)$ has been used to assess the statistical significance of the difference between the distributions of the 90 forecast errors for each forecast step. The test (see, e.g., Wilks 1995 for a definition) gives the probability, herein expressed as a percentage, that any two error

distributions, of which Fig. 2 shows the 90-day average, are samples of the same underlying overall distribution. For any two distributions, the test has been computed using a boot-strapping technique randomly re-sampling the two distributions 5000 times. Being non-parametric, the test has the advantage of not relying on any assumption on the distribution of the scores. Note that in the computation of the test it has been assumed that there is no correlation between the grid-point average root-mean-square errors of consecutive days.

Generally speaking, the test confirms the average results shown in Fig. 2, i.e. that the greater the difference in resolution is, the more statistically significant the difference between the two average forecast error curves is, and that short-range differences are more statistically significant than long-range differences. As an example, Figure 3 shows the test value computed $RMW(T799, Ty)$ computed between the forecast error distribution function of the T799 perturbed forecasts and the distributions of the perturbed forecasts (for reason of space, the test is shown only for the perturbed forecasts since most of the results discussed in this paper refer to the perturbed forecast errors) of experiments T399, T319, T255, T159 and T95, in the realistic and perfect-model frameworks. In the perfect-model framework, the test $RMW(T799, Ty)$ is less than 10% for up to forecast day 12 for all Ty , while in the realistic framework the test is less than 10% only for some resolutions Ty in the short- and medium-range. This difference between the perfect-model and the realistic test values reflects the difference between the corresponding average forecast error curves shown in Fig. 2.

Table 3 lists the forecast error doubling times at forecast days 1-3 for two resolutions, T95 and T399. Before discussing the results, recall that the forecast error has been computed not against observations but against the analysis, which has been generated using the ECMWF 4D-Var assimilation system and thus, most likely, projects more strongly onto the model manifold than the observations. It is interesting to note that for both resolutions the doubling times computed in the perfect-model and the realistic frameworks are very similar, indicating that the model error component not simulated in the perfect-model framework has a negligible impact on the forecast error growth in the short forecast range. Table 3 also shows that both in the perfect-model and the realistic frameworks, the doubling times are shorter for the higher resolution, more active T399 model, which is more penalized a the lower resolution ones. The fact that the T399 model is more active can be detected by comparing the spread of the two systems: at forecast days 2 and 4, the T95 spread is already about 10% and 16% lower than the spread of the T399 ensemble (not shown). *Toth et al* (2002) found a similar effect of increasing the resolution of single forecasts as discussed further in Section 4.

Figure 4 shows the average rmse for the control, the ensemble-mean and the perturbed members for T850 computed over NH in the perfect-model and the realistic frameworks, and Figs. 5 and 6 show the corresponding results for the Southern Hemisphere (SH) and the tropics. Generally speaking, results for T850 confirm the conclusions drawn from the Z500 results: the impact of resolution is very clear in the perfect-model framework, while there is only a small positive impact when increasing resolution from T95 to T255 in the realistic framework. The T850 forecast error asymptotic level depends on the region. The comparison of the NH and SH curves show that the asymptotic level is higher during the cold season over NH than during the warm season over SH, and that the asymptotic level is reached earlier during the warm season over SH.

Doubling time control forecast	T95		T399	
	Perfect-model	Realistic	Perfect-model	Realistic
Day 1	0.73	0.72	0.67	0.62
Day 2	1.23	1.27	0.97	1.06
Day 3	1.64	1.72	1.15	1.38

Table 3a. Doubling times of the control forecast error for Z500 over NH in the perfect-model and the realistic framework computed from the forecast-error growth model solution (see Appendix A).

Doubling time perturbed members	T95		T399	
	Perfect-model	Realistic	Perfect-model	Realistic
Day 1	0.86	0.84	0.91	0.84
Day 2	1.38	1.38	1.31	1.28
Day 3	1.82	1.83	1.61	1.61

Table 3b. Doubling times of the perturbed-members for Z500 over NH in the perfect-model and the realistic framework computed from the forecast-error growth model solution $\tau(t)$ (see Appendix A).

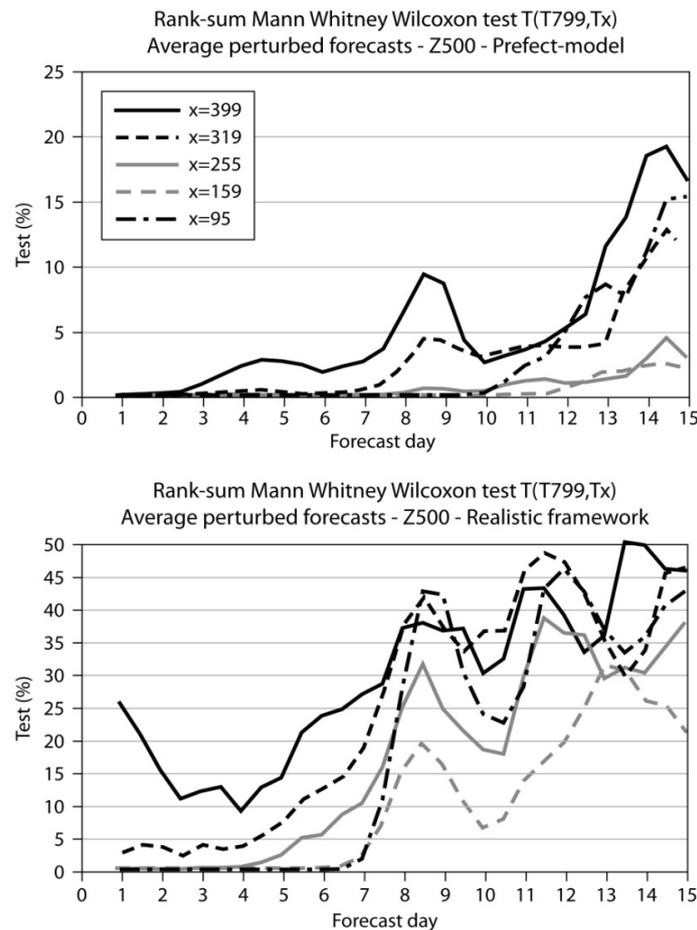


Figure 3. Rank-sum Mann-Whitney-Wilcoxon test $T(T799, Tx)$ computed between the distributions of the perturbed forecast errors of the T799 ensemble and the ensembles run at resolution T399, T319, T255, T159 and T95. The test has been computed for the error distributions of Z500 forecasts over NH in the perfect-model (top) and the realistic (bottom) frameworks.

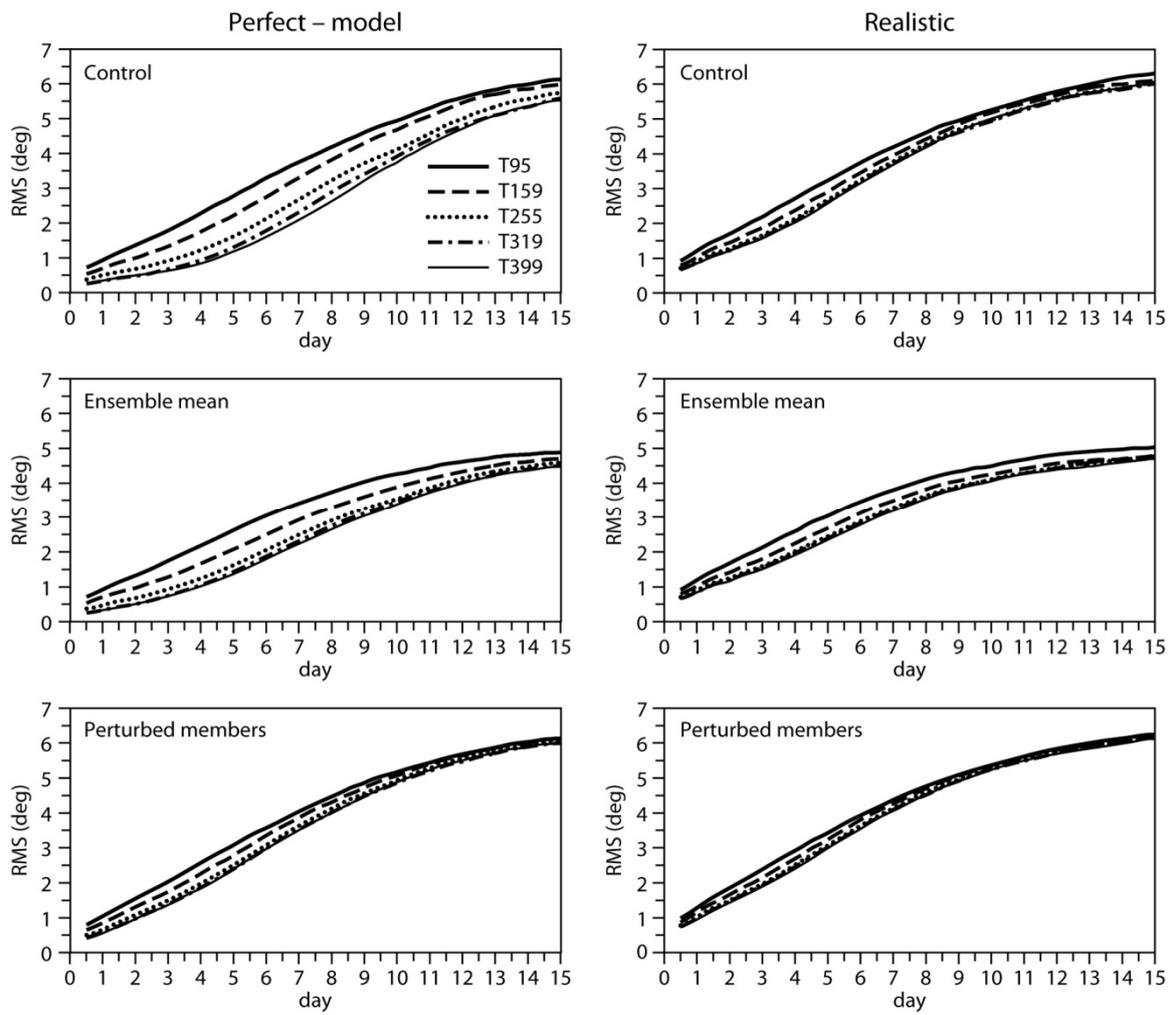


Figure 4: As Fig. 2 but for T850 over NH.

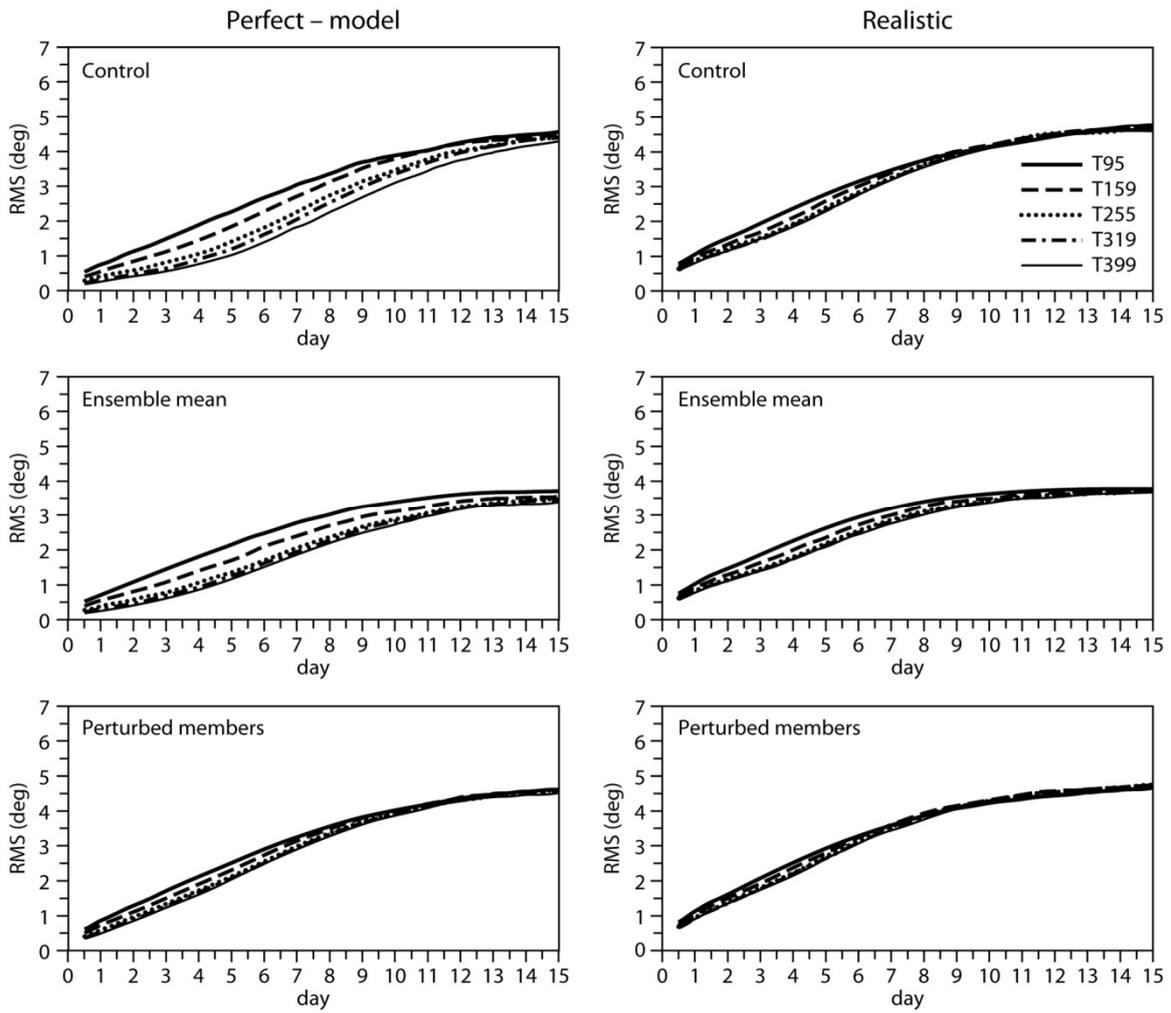


Figure 5: As Fig. 2 but for T850 over the Southern Hemisphere (SH).

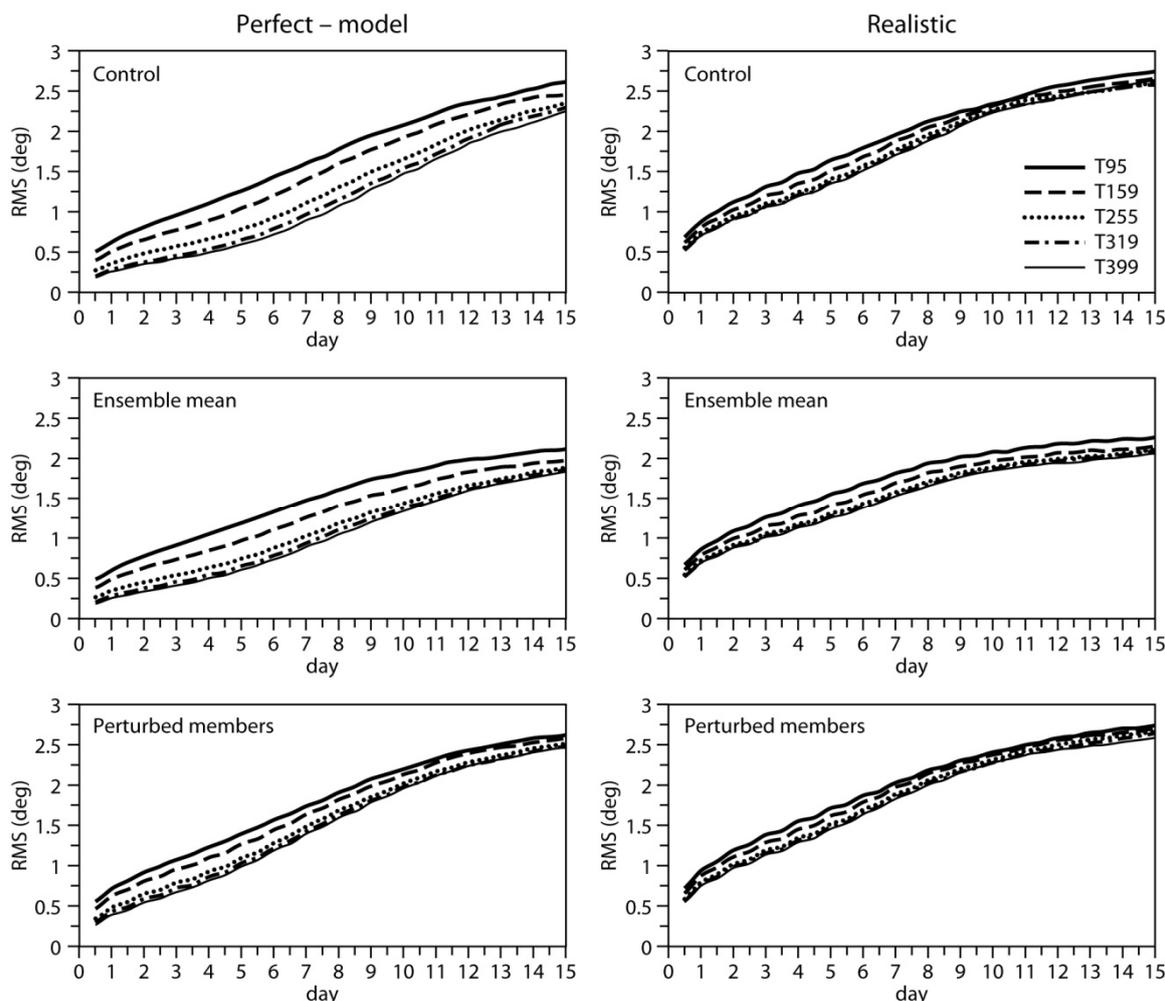


Figure 6: As Fig. 2 but for T850 over the tropics.

3.2 Estimation of the forecast-error growth model from the experimental data

The coefficients (α, β, γ) of the forecast error growth model have been computed for all forecasts in both frameworks using the discretized form of the forecast-error growth model (see Appendix A for more details). The fit of the forecast-error curve to the experimental data is, in general, rather good as indicated by two diagnostics. Figure 7 shows that the scatter plots of the change in forecast error as a function of the average forecast error computed from the data and from the model are close, although not perfect. The poor fit at very short forecast ranges (say up to forecast day 1) is partly due to super-exponential error growth, as was pointed out by *Bengtsson et al* (2008). The differences between the data and the model scatter plots indicate that although the forecast-error growth model captures the key features of forecast error growth, it is not capable to properly describe it at all forecast ranges. As a further measure of model fit, the correlation coefficient between the forecast-error data and the model values have been computed. Results indicate that in both the perfect-model and the realistic frameworks, the correlation coefficients are above 90% for the control forecasts and above 95% for the perturbed forecasts for all resolutions (not shown).

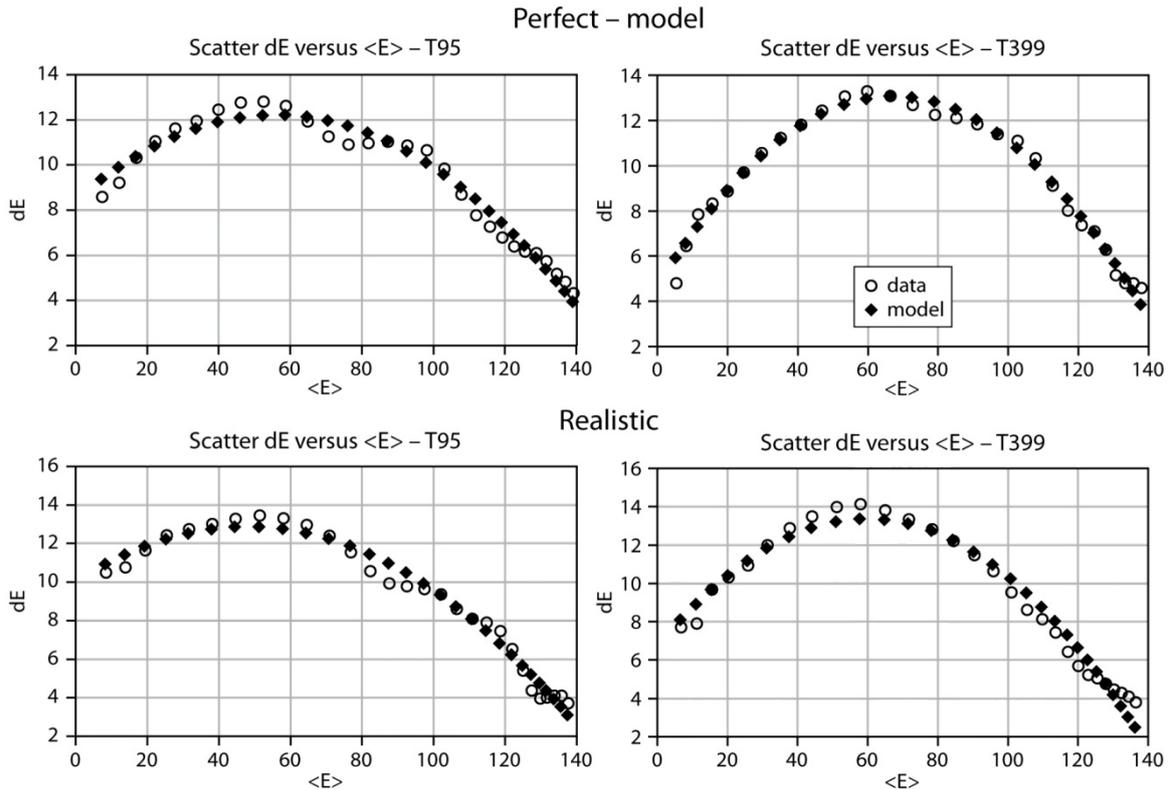


Figure 7: Scatter diagram of the error growth versus the average error computed from the data (open circles) and the model (full diamonds) for Z500 over NH. Top panels: scatter plot for the perturbed members in the perfect-model framework for resolution T95 (left) and T399 (right). Bottom panels: as top panels but for the realistic framework.

3.3 Resolution impact on medium and long range predictability

Figures 8 and 9 show the forecast-error growth model curves for the control and perturbed members, with the model coefficients estimated from the data and the curves normalized by the forecast error limit E_∞ ,

$$\eta(t) \equiv \frac{E(t)}{E_\infty} = 1 - \frac{1 + \frac{S}{aE_\infty}}{1 + C_1 e^{C_2 t}} \quad (2)$$

[See the Appendix for a deduction of the error growth solution and for the relationship between the parameters in Eq. (2) and the coefficients (α, β, γ) of the forecast error growth model Eq (1).]

The normalized curves shown in Figs. 8-9 reflect the rmse data shown in Fig. 7: errors are smaller in the perfect-model than in the realistic framework, the control errors are smaller than the perturbed member errors, and the sensitivity to resolution is larger for the control forecasts, since the high-resolution curves start closer to the analysis, as was pointed out earlier (see also Table 2).

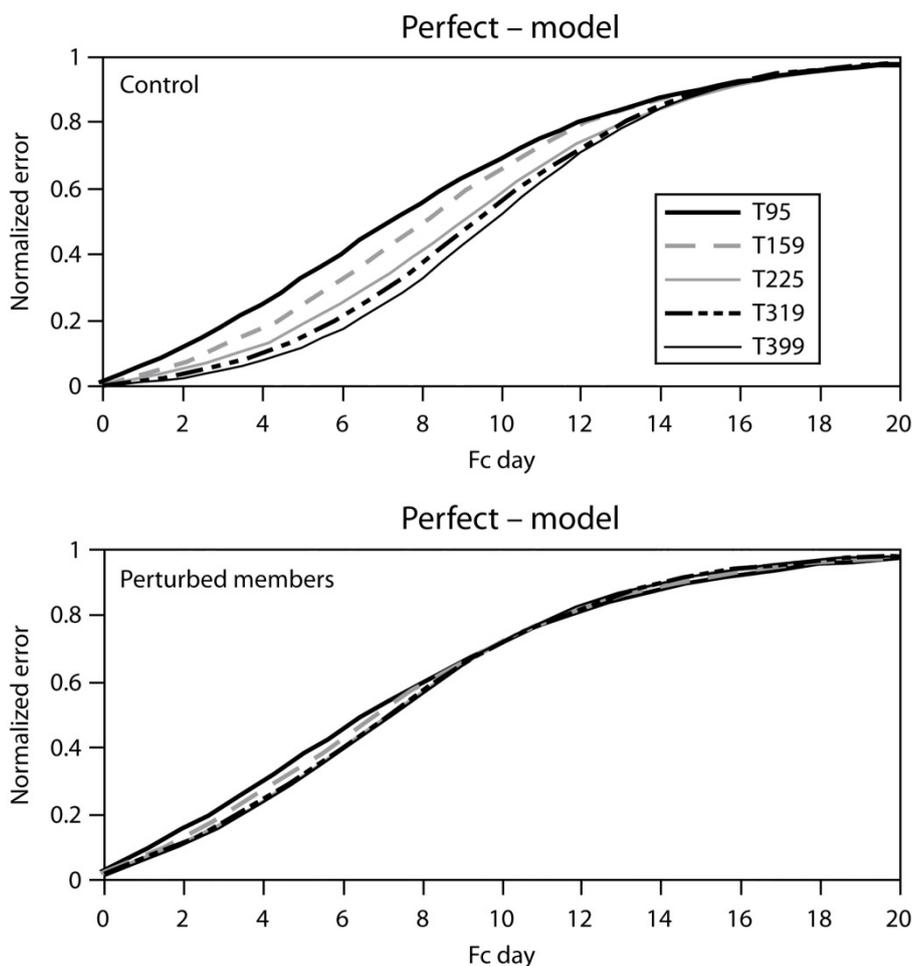


Figure 8: Model error curves computed in the perfect-model framework for Z500 over NH. Top panel: model error curves for the control forecast for the first 10 forecast day for resolutions T95 (solid), T159 (dashed), T225 (dotted), T319 (chain-dashed) and T399 (thin solid). Bottom panel: as top panel but for the perturbed members.

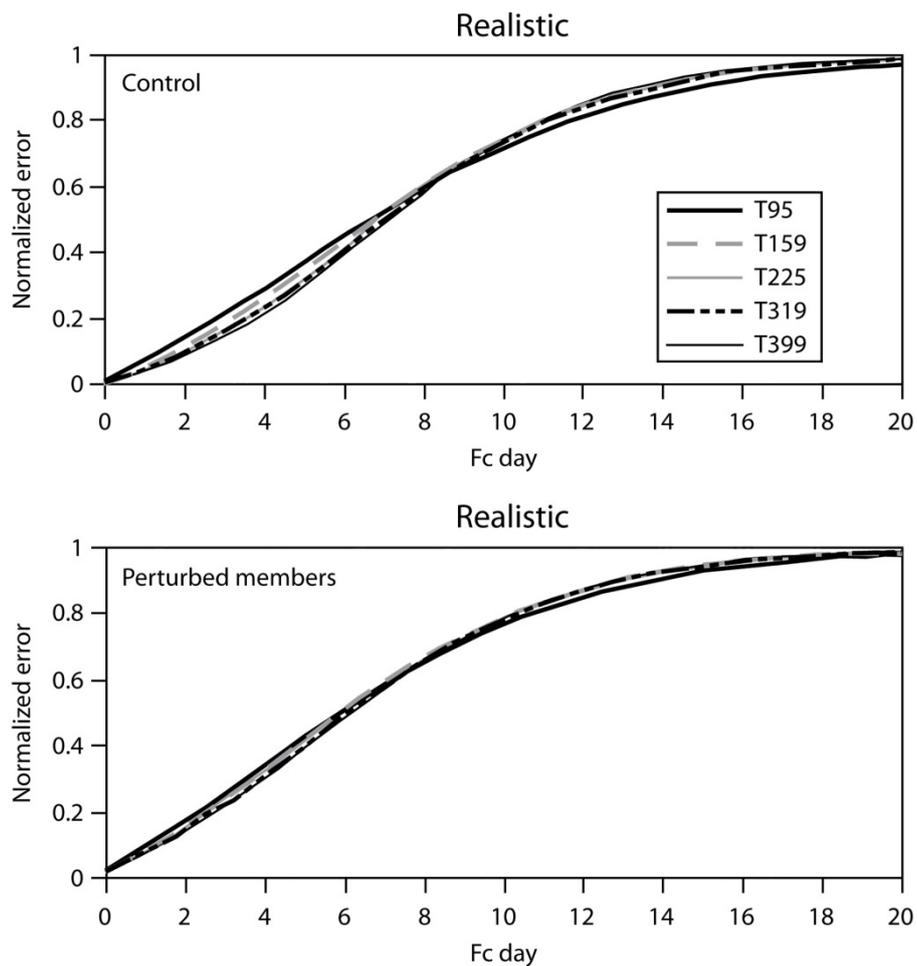


Figure 9: Model error curves computed in the realistic framework for Z500 over NH. Top panel: model error curves for the control forecast for the first 10 forecast day for resolutions T95 (solid), T159 (dashed), T225 (dotted), T319 (chain-dashed) and T399 (thin solid). Bottom panel: as top panel but for the perturbed members.

To highlight the impact of resolution on forecast error growth at different forecast ranges, the forecast times $\tau(25\%)$, $\tau(71\%)$ and $\tau(95\%)$ have been computed from the normalized curves (recall that e.g. $\tau(25\%)$ is the time at which $\eta(t) = 0.25$). The top panels of Fig. 10 show that when resolution is increased by about a factor of 4 from T95 to T399, the control forecast $\tau(25\%)$ increases by 3.0 days (from 4.1 to 7.1d) in the perfect-model framework (by construction, $\tau(100\%)$ is infinite for T799 since in the perfect-model case the T799 control forecast is used as verification). By contrast, it increases only by 0.9d (from 3.5 to 4.4d) in the realistic framework. For the perturbed members, $\tau(25\%)$ increases by ~ 0.8 d in the perfect-model case, and by 0.5d in the realistic framework. It is not surprising that the increase in predictability of the control in the perfect-model framework is the largest, almost 3-times longer than the one in the realistic framework, since not only the model but also the initial condition accuracy improves (see Table 2) as the resolution increased. By contrast, the increase in predictability of the perturbed members in the perfect-model and the realistic case differ by a smaller amount: this is because the initial condition accuracies are more similar (Table 2). The middle and bottom panels of Fig. 10 shows the forecast times when $\tau(50\%)$ and $\tau(71\%)$: these panels

show that as the forecast length increases, the positive impact of resolution decreases and eventually disappears. Results show that $\tau(50\%)$ for the control forecast increases by 2.5d (0.4d) in the perfect-model (realistic) frameworks, and $\tau(50\%)$ for the perturbed forecasts increases by 0.5d (0.2d) in the perfect-model (realistic) frameworks. The impact of resolution is much smaller, and in some cases negative, on $\tau(71\%)$. Table 4 lists the impact of the resolution increase of about a factor of 4, from T95 to T399, on $\tau(25\%)$, $\tau(50\%)$, $\tau(71\%)$ and $\tau(95\%)$ for the perturbed members.

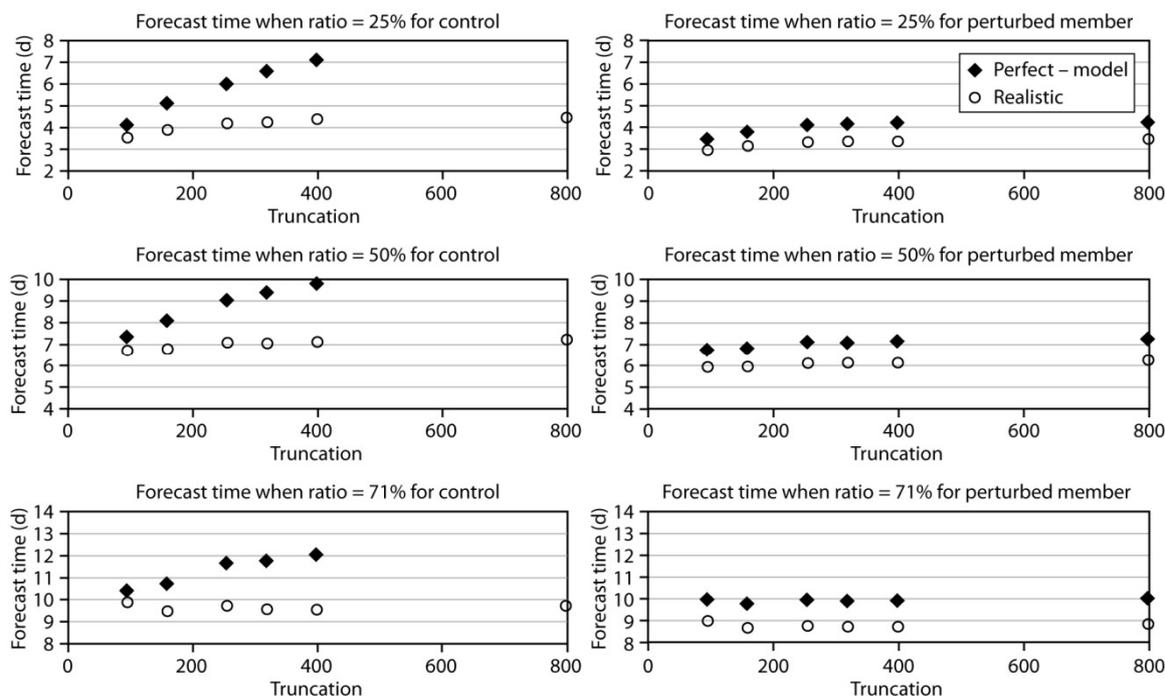


Figure 10: Predictability time limits $\tau(\eta)$ when the average forecast error reaches 25% (top), 50% (middle) and 71% (bottom) of the asymptotic limit for the control (left panels) and the perturbed members (right panels) in the perfect-model (black symbols) and the realistic (grey symbols) framework for Z500 over NH. (Note that for the T799 control in the perfect-model framework τ is infinite for since, by construction, the T799 control forecast is used as verification.)

	Perfect-model			Realistic		
	T95	T399	Difference	T95	T399	difference
$\tau(25\%)$	3.4	4.2	0.8	2.9	3.4	0.5
$\tau(50\%)$	6.7	7.2	0.5	5.9	6.1	0.2
$\tau(71\%)$	9.9	9.9	0	9.0	8.7	-0.3
$\tau(95\%)$	18.1	16.4	-1.7	16.8	15.1	-1.7

Table 4: Impact of resolution increase from T95 to T399 on the predictability times $\tau(25\%)$, $\tau(50\%)$, $\tau(71\%)$ and $\tau(95\%)$ computed for the perturbed members in the perfect-model and realistic framework, for Z500 over NH.

The fact that the impact of resolution decreases with forecast time is more evident in Figure 11, which compares the difference between the forecast-error of the T95, T159, T255, T319 and T399 forecasts and the T799 forecast for all forecast ranges, computed in both frameworks. Figure 11 shows that, in both verification frameworks, a resolution increase has the largest positive impact in the short forecast range, with a peak at around forecast day 1, with the benefit becoming smaller after forecast about day 9 in the perfect-model framework and day 7 in the realistic framework. The fact that there is a 2-day difference between the perfect-model and the realistic results indicate that a reduction of model error (i.e. the elimination of the model error component that affects the realistic and not the perfect-model experiments), could increase the benefit of current, and future resolution increases. The fact that differences are negative for T95 after forecast day 8 is an indication of the limitations of the 3-parameter forecast error growth model to fully describe forecast error growth.

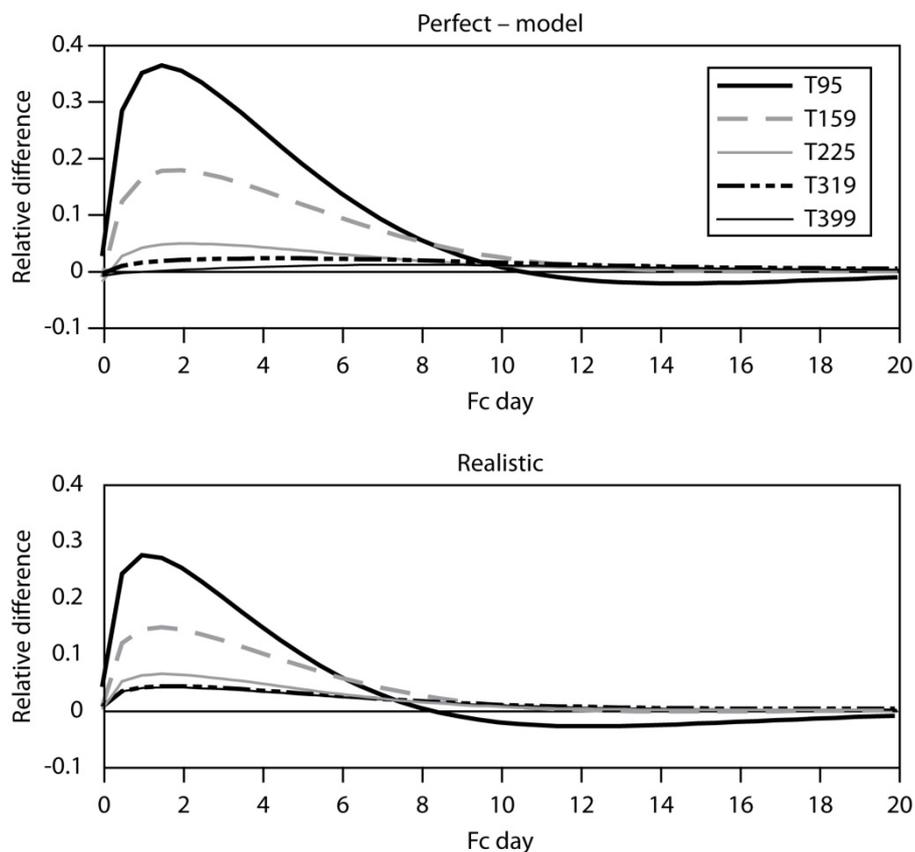


Figure 11: Relative difference between the $rmse(pf)$ of resolution T_x and of configuration T_{799} in the perfect-model (top panel) and the realistic (bottom panel) framework, for resolutions ($T_x=$) T95 (solid), T159 (dashed), T255 (dotted), T319 (chain-dashed) and T399 (thin solid), for Z500 over NH.

Figures 12-16 summarize the main results of this work. Figures 12, 13 and 14 show the time limits $\tau(25\%)$ and $\tau(71\%)$ computed using the normalized, error growth curves for the perturbed members for Z500, Z1000 and T850 over NH [the time limit $\tau(95\%)$ has not been shown because, as pointed out above, already $\tau(71\%)$ shows a very weak sensitivity to resolution]. The time limits computed using Z1000 (Fig. 13) and T850 (Fig. 14) confirm the two key conclusions drawn above, precisely that a resolution increase has a larger impact on the shorter forecast range and that the impact of resolution decreases as resolution increases.

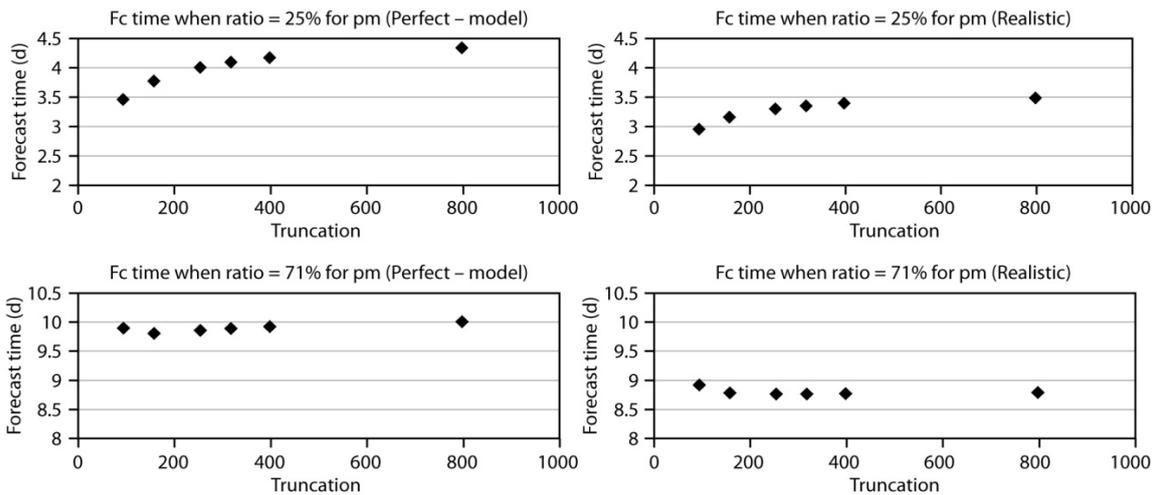


Figure 12: Forecast times $\tau(25\%)$ (top panels) and $\tau(71\%)$ (bottom panels) for Z500 over NH for the perfect-model (left panels) and the realistic (right panels) framework.

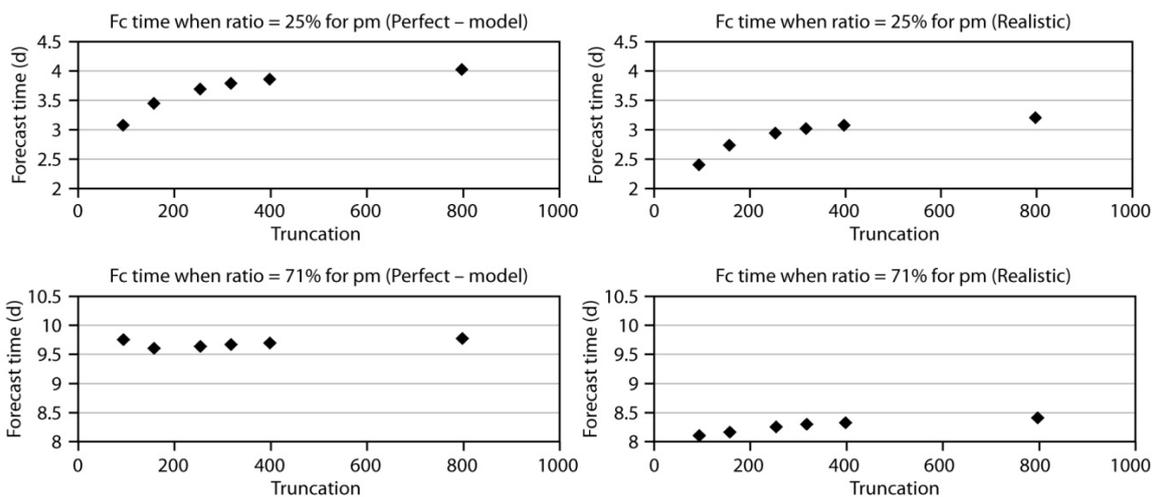


Figure 13: As Fig. 12 but for Z1000 over NH.

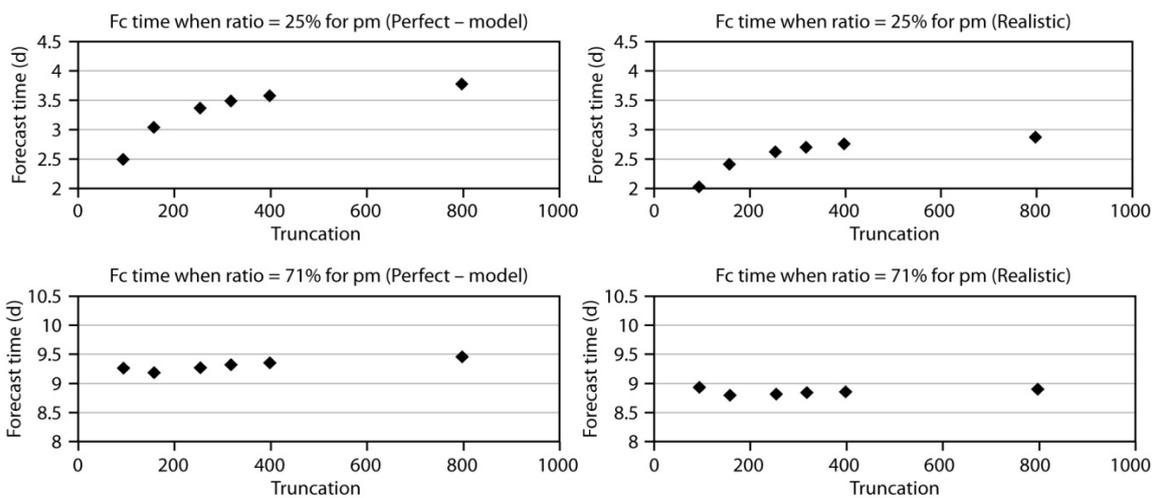


Figure 14: As Fig. 12 but for T850 over NH.

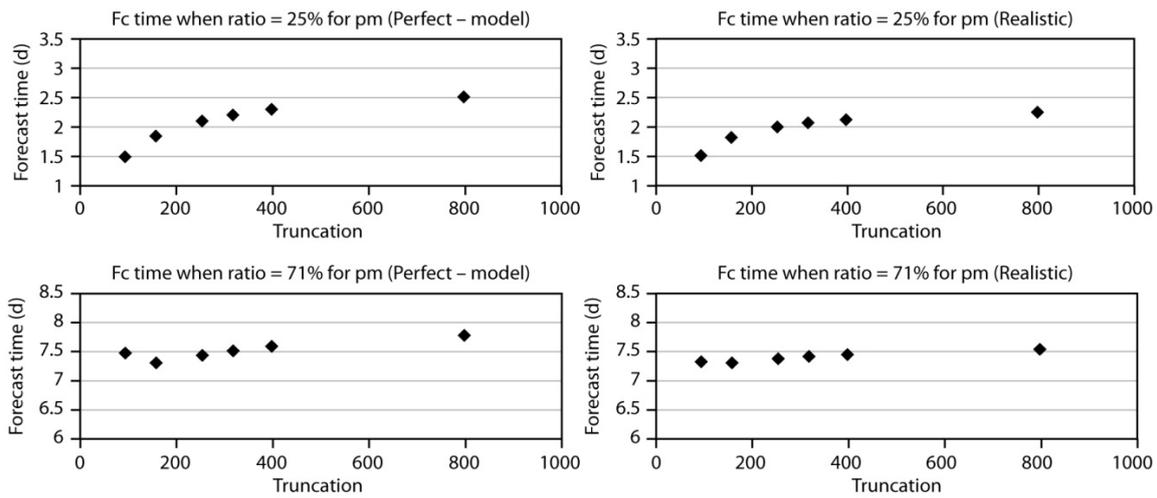


Figure 15: As Fig. 12 but T850 over SH.

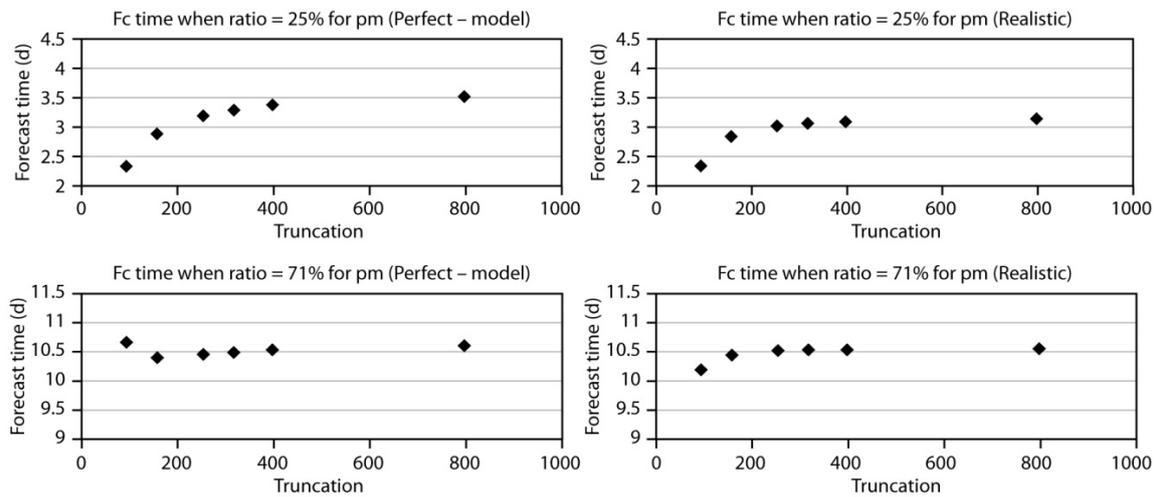


Figure 16: As Fig. 12 but for T850 over the Tropics.

Figures 15 and 16 show the time limits $\tau(25\%)$ and $\tau(71\%)$ computed using the normalized, error growth curves for the perturbed members for T850 over SH and the tropics. These figures support the two conclusions drawn above, although the time limits are quantitatively different. The comparison of the T850 results for NH (Fig. 14) and SH (Fig. 15) show that the predictability limits are shorter during the warm season over the SH than during the cold season over the NH: for example, $\tau(71\%)$ for T799 in the realistic case is 9 days over NH and 7.5 days over SH. For the tropics, the short-range limit $\tau(25\%)$ is similar to the NH one, but the long-range limit $\tau(71\%)$ is longer, 10.5 days for the T799 forecasts verified over the tropics, compared to 9 days for forecasts verified over NH and 7.5 days for forecasts verified over SH.

4 Discussion and conclusions

The error growth of instantaneous forecasts has been analyzed, and the sensitivity of predictability to horizontal resolution has been assessed using ECMWF forecasts. More precisely, the impact of horizontal resolution on short, medium and long range predictability time limits, defined as the times when the forecast root-mean-square error reaches different levels of the forecast error asymptotic limit (25%, 50%, 71% and 95%) has been discussed. Attention has been focused on the 500 and 1000 hPa geopotential height and the 850 hPa temperature forecasts during in winter 2007/08 (December 2007, January and February 2008, 90 cases) verified over the Northern Hemisphere, the Southern Hemisphere and the tropics. The *Dalcher and Kalnay* (1987)'s model of forecast-error growth has been used to extrapolate the forecast-error growth beyond the 15 forecast days for which forecast data have been generated, and to normalize the forecast error curves of forecasts run with different resolutions using the appropriate asymptotic limit given by the error growth model. Forecast error growth has been studied both in a realistic framework, with forecasts verified against analyses, and in a perfect-model framework, using high resolution forecasts as verification. Results obtained in the perfect-model framework gave an upper bound on the skill gains times that could be expected in reality.

Before summarizing the key results, it is important to mention two important caveats. Firstly, analyses have been used as verification in the realistic framework instead of observations. This has the benefit of providing a better coverage of the whole globe for all variables, but has the disadvantage that the short-range forecast error might be underestimated. Although this leads to an under-estimation of the short-range forecast error (say by up to 25% for forecast lengths shorter than 2 days), it is difficult to gauge whether this could lead to an over or and under-estimation of the impact of horizontal resolution on predictability. Secondly, since the same analysis is used in all experiments (albeit interpolated from the operational T799 resolution to the experiment resolution), this study does not take into account any potential benefit that increases in the horizontal resolution used in data-assimilation could bring. This might have reduced the sensitivity to horizontal resolution of the short-range predictability limits, but it is doubtful whether it could have any impact on the medium- and long-range predictability limits.

The key results of this work on six issues are summarized hereafter.

► Predictability time limits estimated in the perfect-model and the realistic frameworks

Results obtained in the two frameworks lead to the same overall conclusions, although quantitatively the predictability times obtained in the perfect-model framework are longer than the corresponding times obtained in the realistic framework. This is particularly true in the medium range for the control forecasts. The perfect-model values give an upper bound on the predictability that could be achieved by current, state-of-the-art numerical weather prediction models, such as the one used at ECMWF. Considering the impact of a 4-fold increase of resolution from T95 to T399 on the error of perturbed forecasts over NH, perfect-model results listed in Table 4 (left columns) indicate gains in the short range skill that are about 60% longer than what the realistic results indicate [e.g., 0.8 instead of 0.5 days for $\tau(25\%)$]. Differences become larger in the medium range, with skill gains up to 100% longer than the realistic results indicate [e.g., 0.5 instead of 0.2 days for $\tau(50\%)$]. These differences indicate that simple resolution increases without model developments would bring only small improvements in the medium and long forecast ranges.

► Usefulness of simple, parametric forecast error growth models

The simple forecast error growth model first proposed by *Lorenz* (1969, 1982), modified by *Dalcher and Kalnay* (1987) and used, among others, by *Savijarvi* (1995), *Simmons and Hollingsworth* (2002) and *Bengtsson et al* (2008) has proven again to be a useful tool to investigate forecast error growth beyond the forecast range spanned by available forecast data. In case of the comparison of the error growth of forecasts characterized by different asymptotic limits, it provides an easy way to normalize the different forecast error curves. Despite this, our investigation also confirmed *Bengtsson et al* (2008)'s indication that this forecast error growth model has difficulties in describing the error growth at very short forecast ranges.

► Impact of resolution on predictability time limits over NH winter

Results obtained in both the perfect-model and the realistic frameworks indicate that increasing resolution leads to longer time limits in the short forecast range, but has a small impact in the long forecast range (see Table 4 for a summary of the NH results). These results are in line with the impact that resolution increases reported in the following published works:

Simmons and Hollingsworth (2002), who reported improvement in the short range forecast skill by about 1 day per decade (see their Fig. 4), due to a combination of a resolution increase of about a factor of 2 (from T213 to T511 between 1992 and 2000), model changes and improvements in data assimilation. Our results indicate an increase of $\tau(25\%)$ for the control forecast in the realistic framework of a resolution increase from T159 to T399 of 0.5 days (from 3.9 to 4.4 days). The gain computed in this work in the realistic framework is lower, possibly because forecasts did not benefit from model or data-assimilation improvements (forecasts have been performed using the same model cycle and starting from the same initial conditions).

Tracton and Kalnay (1993) and *Toth and Kalnay* (1993), who discussed the implementation of the NCEP ensemble prediction in 1992 with a variable resolution, higher up to day 5 and then lower, because they found that beyond 5 days horizontal resolution did not improve the skill of their system. *Toth et al* (2002) also reported that for single forecasts the use of higher resolution can reduce the short range forecast error, but could have a small or even detrimental effect in the long range.

Buizza et al (2003), who reported gains in the predictability of single control forecasts of about 0.5 days at around forecast day 5 when the ensemble resolution was increased from T159 to T255, and the analysis resolution was increased from T319 to T511 (see their Fig. 4). Our results indicate an increase of $\tau(50\%)$ for the control forecast in the realistic framework of a resolution increase from T159 to T255 of 0.3 days (from 6.8 to 7.1 days): our gain is lower because *Buizza et al* (2003) T255 forecasts also benefited from the increase of the analysis resolution.

► Short-, medium and long-range predictability limits over NH, SH and the tropics

The short-range predictability time limit $\tau(25\%)$ has shown a strong sensitivity to resolution: for T850 in the realistic case, for example, it has increased by about 50% when horizontal resolution has increased from T95 to T799: from 2 to 3 days over NH, from 1.5 to 2.2 days over SH and from 2.3 to 3.2 days over the tropics. Results indicate that also a horizontal resolution increase from T399 to T799 leads to small but detectable improvements. By contrast, the medium-range limit $\tau(71\%)$ has not shown any sensitivity to resolution. Similar conclusions can be drawn by considering the perfect-

model results. The long range time limit $\tau(95\%)$ computed in this work for T255 Z500 forecasts verified over NH is 16.6 days in the perfect-model framework and 15.3 days in the realistic framework. These values are about 2 day longer than the 13 days estimated by *Simmons and Hollingsworth* (2002) for the T255 system that was operational at ECMWF at the time of their investigation. The corresponding predictability limits for the T799 system are practically the same as the T255 ones (16.6 and 15.2 days, respectively, in the perfect-model and the realistic frameworks). Considering T799 forecasts of the 850 hPa temperature (T850) verified in the realistic framework, the time limit $\tau(95\%)$ is 16.6 days for NH (cold season), 14.1 days for SH (warm season) and 20.6 days over the tropics.

► Forecast error doubling times over NH

Consistently with the positive impact of a resolution increase, our results have indicated that the forecast error doubling time decreases when resolution increases, especially in the short range (day 1-4), in agreement with the findings of *Simmons and Hollingsworth* (2002). For example, they reported a $\sim 10\%$ decrease in the doubling time at forecast day 2, from 1.59 days in 1996/97, when the ECMWF model had a T213 resolution, to 1.42 day in 2000/01, when the resolution was T511 (see their Table 1). Our results for the control forecast in the realistic framework indicate a $\sim 12\%$ reduction of the doubling-time at forecast day 2 (from 1.21 days to 1.06 days) when resolution was increased from T255 to T399. For the perturbed forecasts in the realistic framework, our results indicate a $\sim 5\%$ reduction (from 1.34 days to 1.28 days). As mentioned above, the fact that the decrease is larger for the control forecast is because higher-resolution control forecasts start closer to the analysis.

► Predictability limit sensitivity to horizontal resolution and model improvements

To conclude this discussion, let us consider the question that was posed in the introduction: what is the sensitivity of the short and long range forecast skill to the model resolution? Our results have indicated a strong sensitivity of forecast skill to model resolution in the short range, but no sensitivity in the long range. If, as in *Dalcher and Kalnay* (1987), the predictability limit is defined as $\tau(95\%)$, i.e. the forecast time when the forecast error reaches 95% of the saturation level, our estimates indicate that for the prediction of instantaneous, synoptic scale features as represented by Z500, our estimates are that the limit is 16.6 and 15.2 days, respectively, for T799 perturbed forecasts verified in the perfect-model and the realistic frameworks. These values are 3-to-4 days longer than the 12 days quoted by *Dalcher and Kalnay* (1987), which was obtained applying the forecast error growth model to the ECMWF Z500 forecasts used by *Lorenz* (1982). Although a direct comparison of these estimates is not possible, since they used forecasts issued by weather models of different complexity for different periods, and used different versions of the forecast error growth model, the comparison suggests that we have not yet converged to a number that measures the real predictability limit of the atmosphere, and future works might further lengthen this estimate. Our results suggest that rather than resolution, it is model improvements that might lead better predictions and longer predictability limits.

Acknowledgements

Dr A Beljaars, Prof. E Källén and Dr Tim Palmer are thanked for providing valuable comments to an earlier version of this paper. Dr Dan Cornford, Prof. Eugenia Kalnay and two anonymous reviewers

are thanked for their revision works that lead to an improved manuscript. Anabel Owen is thanked for her very meticulous editorial work that led to higher quality figures.

Appendix A: the forecast-error growth model

The forecast-error growth model used to estimate the predictability limits was first proposed by *Lorenz* (1982), then used and slightly modified by *Dalcher and Kalnay* (1987) and by *Simmons et al* (1995). Accordingly to this model, the time evolution of the forecast error E is given by the following equation:

$$(A.1) \quad \frac{dE}{dt} = \alpha E - \beta E^2 + \gamma .$$

In this study, the forecast error $E(t)$ of Eq. (A.1) is the average root-mean-square error of a forecast, verified either against a T799 analysis in the realistic framework or against a T799 forecast in the perfect-model framework. The average forecast error has been computed by using grid-point weights $w_k = \cos(\text{lat}_k)$ to take into account the spherical geometry of the earth:

$$(A.2) \quad E(t) = \sqrt{\frac{1}{N_{day}} \frac{1}{N_G} \sum_{j=1}^{N_{day}} \sum_{k=1}^G w_k (f_{j,k}(j,t) - a_{j,k}(j+t))^2}$$

Where in Eq. (A.2) $f_{j,k}(j,t)$ is the forecast at time t started on day j at grid-point k , and $a_{j,k}(j+t)$ is the corresponding verifying analysis.

Equation (A.1) can also be written as:

$$(A.3) \quad \frac{dE}{dt} = (aE + S)\left(1 - \frac{E}{E_\infty}\right) ,$$

with:

$$(A.4) \quad \begin{aligned} S &\equiv \gamma \\ E_\infty &\equiv \frac{\alpha}{2\beta} + \sqrt{\frac{\alpha^2}{4\beta^2} + \frac{\gamma}{\beta}} . \\ a &\equiv \beta E_\infty \end{aligned}$$

In Eq. (A.3), a is the rate of growth of the forecast error, S simulates the effect of model error deficiencies on the error growth, and E_∞ is the asymptotic value (see *Dalcher and Kalnay*, 1987, for more details). Equation (A.3) has solution:

$$(A.5) \quad \eta(t) \equiv \frac{E(t)}{E_\infty} = 1 - \frac{1 + \frac{S}{aE_\infty}}{1 + C_1 e^{C_2 t}} ,$$

with

$$(A.6) \quad \begin{aligned} C_1 &\equiv \frac{E(0) + S/a}{E_\infty - E(0)} , \\ C_2 &\equiv a + S/E_\infty \end{aligned}$$

where $E(0)$ is the error at initial time. Note that Eq. (A.5) can be used to compute the forecast error doubling time $\tau(t) = \frac{\ln(2)}{\alpha + \gamma/E(t)}$ at each forecast step t .

Equation (A.1) can be written in a discretized form for the forecast step j :

$$(A.7) \quad \frac{\Delta E_j}{\Delta t} = \alpha \bar{E}_j - \beta \bar{E}_j^2 + \gamma ,$$

where for each forecast step $j=1, \dots, N$, Δt is equal to 12 hours, $\Delta E_j = (E_{j+1} - E_j)$ is the forecast error increase between steps j and $(j+1)$, and $\bar{E}_j = 0.5(E_j + E_{j+1})$ is the average forecast error between the two steps. As in *Lorenz (1982)* and *Simmons et al (1995)*, the three parameters (α, β, γ) of Eq. (A.1) can be estimated by a least-square fit of the root-mean-square differences between consecutive forecast errors ΔE_j for $j=1, 17$, with $t_1=0$ and $t_{17}=360$ hours. Once (α, β, γ) have been determined, the coefficients (a, S, E_∞) and (C_1, C_2) can be computed using the relationships in Eqs. (A.4) and (A.6).

As in *Dalcher and Kalnay (1987)*, **predictability time limits $\tau(\eta)$** can be defined as the forecast times when the forecast error equals a fraction η of the asymptotic value E_∞ . By using Eq. (A.5), it is straightforward to compute the predictability limit as a function of the ratio $\eta(t)$,

$$(A.8) \quad T_{\text{lim}} = \frac{1}{C_2} \ln \left(\frac{\frac{S}{aE_\infty} + \eta(t)}{\frac{1}{C_1} \frac{1}{1 - \eta(t)}} \right) .$$

In this work, four time limits will be computed, as the time when $\eta(t)$ equals 0.25, 0.5, $1/\sqrt{2}$ (0.71) and 0.95.

References

- Bechtold, P., M., Koehler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., and Balsamo, G., 2008: Advances in simulating atmospheric variability with the ECMWF model: from synoptic to decadal time-scales. *ECMWF Technical Memorandum no. 556*, pp. 22. Available from <http://www.ecmwf.int/publications/>
- Bengtsson, L. K., Magnusson, L., and Källén, E., 2008: Independent estimations of the asymptotic variability in an ensemble forecast system. *Mon. Wea. Rev.*, 136, 4105-4112.
- Brown, A.R., 2004: Resolution dependence of orographic torques. *ECMWF Technical Memorandum no. 433*, pp. 18. Available from <http://www.ecmwf.int/publications/>
- Buizza, R., and T.N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **125**, 2887-2908.
- Buizza, R., Richardson, D. S., and Palmer, T. N., 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Q. J. R. Meteorol. Soc.* **129**, 1269-1288.
- Charney, J. G., Fleagle, R. G., Fiehl, H., Lally, V. E., and Wark, D. Q., 1966: The feasibility of a global observation and analysis experiment. *Bull. Americ. Meteorol. Soc.*, **47**, 200-220.
- Dalcher, A., and Kalnay, E., 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39**, 474-491.
- Jung, T., Palmer, T. N., and Shutts, G. J., 2005: Influence of stochastic physics on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *ECMWF Technical Memorandum no. 473*, pp. 7. Available from <http://www.ecmwf.int/publications/>
- Leith, C. E., 1982: Statistical methods for the verification of long and short range forecasts. ECMWF Seminar on 'Problems and Prospects in Long and Medium Range Weather Forecasts', 14-18 September 1981, 313-333. Available from <http://www.ecmwf.int/publications/>
- Lorenz, E., 1969: Atmospheric predictability as revealed by natural occurring analogues. *J. Atmos. Sci.*, **26**, 636-646.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505-513.
- Palmer, T N, Buizza, R., Leutbecher, M., Hagedorn, R., Jung, T., Rodwell, M, Vitart, F., Berner, J., Hagel, E., Lawrence, A., Pappenberger, F., Park, Y.-Y., van Bremen, L., Gilmour, I., and Smith, L., 2007: The ECMWF Ensemble Prediction System: recent and on-going developments. A paper presented at the 36th Session of the ECMWF Scientific Advisory Committee. *ECMWF Technical Memorandum no. 540*, pp 55. Available from <http://www.ecmwf.int/publications/>
- Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer M., and Weisheimer, A., 2009: Stochastic parametrization and model uncertainty. ECMWF Technical Memorandum no.598, pp. 42. Available from <http://www.ecmwf.int/publications/>

- Reynolds, C. A., Webster, P. J., and Kalnay, E., 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281-1305.
- Savijarvi, H., 1995: Error growth in a large numerical forecast system. *Mon. Wea. Rev.*, **123**, 212-221.
- Shutts, G. J., 2005: A kinetic back-scatter algorithm for use in ensemble prediction systems. *Q. J. Roy. Meteorol. Soc.*, **131**, 3079-3102.
- Simmons, A. J., and Hollingsworth, A., 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **128**, 647-677.
- Simmons, A. J., Mureau, R., and Petroliaigis, T., 1995: Error growth and predictability estimates for the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.*, **121**, 1739-1771.
- Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. *Bull. Amer. Meteorol. Soc.*, **50**, 286-311.
- Toth, Z., and Kalnay, E., 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317-30.
- Toth, Z., Zhu, Y., Szunyogh, I., Iredell, M., and Wobus, R., 2002: Does increased resolution enhance predictability?. Paper J1.8 in *Proceedings of the AMS Symposium on 'Observations, data assimilation and probabilistic prediction'*, 13-17 January 2002, Orlando, Florida. Available from American Meteorological Society, Boston, USA.
- Tracton, M. S., and Kalnay, E., 1993: Operational ensemble forecasting at the National Meteorological Center: practical aspects. *Weather Forecast.*, **8**, 379-398.
- Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press, Inc., San Diego, pp. 467 (ISBN 0-12-751965-3).