TECHNICAL MEMORANDUM

739

The TIGGE global, medium-range ensembles

Roberto Buizza

Research Department

November 2014

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

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

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

Abstract

TIGGE, the THORPEX Interactive Grand Global Ensemble, is a World Meteorological Organization research programme, launched in March 2005, with the goal to accelerate improvements in the accuracy of 1-day to 2-week high-impact weather forecasts.

The TIGGE archive has been a unique resource for research and development, giving access to hundreds of global, medium-range forecasts issued every day, with different resolution and forecast length. TIGGE has helped scientists to investigate predictability issues and diagnose ensembles' performance by comparing how the different ensembles behave in different cases and seasons. It is also thanks to TIGGE that we have been witnessing a paradigm shift in numerical weather prediction from a single-based, deterministic to an ensemble-based, probabilistic approach, whereby forecasts are more valuable if they are accompanied by uncertainty estimations, expressed e.g. as ranges of possible scenarios or as probabilities of specific events.

This paper introduces a unified framework that can be used to characterize the TIGGE ensembles, compare their characteristics and highlight the rationale behind their design. It presents the configuration of each TIGGE ensemble at implementation time and the time of writing (summer 2014), and discusses the performance of the TIGGE ensembles in predicting synoptic-scales in recent seasons.

There are four main conclusions that can be drawn from this work. Firstly, there is not a unique recipe to generate reliable and skilful ensembles, and indeed the nine TIGGE ensembles show that different methods can be used to simulate the effects of the sources of forecast errors. Secondly, one of the TIGGE ensembles, from ECMWF, still outperforms all the others, with differences in predictive skill for synoptic-scale variables in the medium-range of about 1 day, which corresponds to about 7-8 years of development. Thirdly, international projects such as TIGGE and follow-on projects, will continue to play the essential role of providing the resources (e.g. easy access to standardized multi-system data) to help scientists to understand predictability and further advance ensemble techniques. Because of this, they should continue to be supported. Fourthly, ***ensembles are here to stay, and the future will see ensembles used even more to estimate the probability distribution function of initial and forecast states of ocean, land and atmospheric variables***.

1 TIGGE operational, global, medium-range ensembles

TIGGE is the THORPEX Interactive Grand Global Ensemble (see *Bougeault et al* 2010, and TIGGE and THORPEX in the list of references), where THORPEX is a World Meteorological Organization (WMO) World Weather Research Programme (WWRP) aiming to accelerate improvements in the accuracy of 1-day to 2-week high-impact weather forecasts. TIGGE was launched in March 2005 at a workshop hosted by the European Center for Medium-Range Weather Forecasts (ECMWF; *Richardson et al*, 2004), with the objectives to develop a deeper understanding of the contribution of observation, initial and model uncertainties to forecast error, and to investigate new methods of combining ensembles from different sources and of correcting systematic errors. Since its inception, three centres have been acting as data collectors: ECMWF, CMA (China Meteorological Administration, Beijing, China) and NCAR (National Center for Atmospheric Research, Boulder, Colorado, US). At ECMWF, users can access data generated by the TIGGE ensembles from October 2006 for scientific research, by registering as a *tiggeuser* at the ECMWF TIGGE portal (tigge.ecmwf.int).

The TIGGE archive contains medium-range forecasts (i.e. forecasts valid at least for up to forecast day 10) from nine operational, global ensembles, produced by:

1. BMRC, the Australian Bureau of Meteorology;
2. CMA, the China Meteorological Administration;
3. CPTEC, the Brazilian Center for Weather Prediction and Climate Studies (Centro de Previsao de Tempo e Estudos Climatico);
4. ECMWF, the European Centre for Medium-Range Weather Forecasts;
5. JMA, the Japanese Meteorological Administration;
6. KMA, the Korean Meteorological Administration;
7. MSC, the Meteorological Service of Canada;
8. NCEP, the National Centers for Environmental Prediction;
9. UKMO, the UK Met Office.

The archive includes also forecasts from Meteo-France (MF), which are global (albeit they use a stretched grid with higher resolution over France and lower at its antipodes) but they extend only up to 84 hours. Because of their limited forecast range, the MF ensemble is not discussed in this work.

The TIGGE archive has been, from its inception, a unique and extremely valuable resource. Before TIGGE, these ensembles could have been accessed only from each production centre, with each one archived with a different format and data structure. For some fields, even units were not the same. TIGGE defined a standard, and its archive contains many fields (30 surface fields, 5 upper level fields on 8 pressure levels, every 6 hours for the whole forecast length) from all these ensembles, all archived in the same format and with the same units. Experts and scientists have used the TIGGE data to compare ensembles, to identify their strengths and weaknesses, and to assess the potential value of combining different ensembles to generate multi-model/multi-analysis grand-global ensemble products. As evident from TIGGE-related published literature (see the ECMWF web page for a list of about 70 papers, http://tigge.ecmwf.int/references.html), the archive has been very valuable also for research and development in the fields covered by the two other THORPEX working groups, the Data Assimilation and Observing System (DAOS) and the Predictability and Dynamical Processes (PDP), and other WWRP working groups, including the Working Group on Numerical Experimentation (WGNE), Socioeconomic Research and Applications (SERA) and the Joint Working Group on Forecast Verification Research (JWGFVR).

It is worth reminding that the TIGGE archive has been established as a resource for researchers and not as the provider of operational, multi-system ensemble forecasts. TIGGE data are provided freely for research and education purposes: registration e.g. via the ECMWF TIGGE portal (http://apps.ecmwf.int/datasets/data/tigge/licence/) provides access to the TIGGE data with a delay of 48 hours after initial time of the forecasts. The reader interested in operational multi-model ensembles, could look at two operational ventures established in the U.S.: the North American Ensemble Forecasting System (NAEFS, *Candille* 2009) and the NCEP-FNMOC (Fleet Numerical Meteorology and Oceanography Center) global multi-centre ensemble system dedicated to probabilistic forecasts of wind-wave heights (*Alves et al* 2013).

Before discussing the details of the TIGGE operational, global, medium-range ensembles (hereafter OG-ENS), let's clarify the meaning of these three words, 'operational', 'global' and 'medium-range':

- *operational* means that the ensemble forecasts are issued on a daily basis;
- *global* means that their forecasts cover the whole globe;
- *medium-range* means that they provide forecasts for at least up to 10 days (some of them extend to about 2 weeks, and one to 32 days).

Operational, global medium-range ensemble prediction began in November/December 1992, when ECMWF (*Palmer et al* 1993, *Buizza & Palmer* 1995, *Molteni et al* 1996, *Buizza et al* 2007) and NCEP (*Toth & Kalnay* 1993 and 1997) started issuing global ensemble predictions as part of their operational products. ECMWF and NCEP were followed by MSC (*Houtekamer et al* 1996), who implemented its operational global ensemble prediction system in 1995. Soon afterwards, six other centres started producing global ensemble forecasts operationally: BMRC (*Bourke et al* 1995), CMA (*Su et al* 2014), CPTEC (*Coutinho* 1999), JMA (*Yamaguchi and Majumdar* 2010), KMA (*Goo et al* 2003) and UKMO (*Bowler et al* 2007). This brings the total number of the TIGGE OG-ENS to nine.

Although the discussion of shorter-range ensembles is beyond the scope of this work, readers should be aware that many centres (e.g. Meteo France, as already mentioned; centres involved in the HIRLAM programme; the COSMO Consortium established by the German, Greece, Italian, Polish and Swiss National Meteorological Services; the Spanish Instituto National de Meteorologia) have been running operationally, or are testing, short-range regional ensemble prediction systems. These limited-area and/or short-range systems are not going to be discussed in this paper.

All nine TIGGE OG-ENS have been designed to represent the effects on forecast quality of observation uncertainties, imperfect surface and boundary conditions, data assimilation assumptions (linked to the data-assimilation methods and underlying statistics) and model uncertainties (e.g. due to unresolved processes and the simplified parameterization of physical processes). Each center followed a different approach, and one of the advantages of TIGGE is that it allows comparing them, and assessing how system design affects performance.

In this paper we will review the key characteristics of the nine TIGGE OG-ENS, describe their operational configuration at the time of writing (summer 2014), and discuss their performance. After this introduction, in section 2, we will review how the OG-ENS have been designed, linking their key characteristics to more fundamental predictability work and ideas on how to simulate the different sources of forecast uncertainty. In section 3 we will compare how the OG-ENS perform, looking both at one specific case and at some statistical results. In section 4 we will discuss in general terms the future of ensemble techniques, with a particular focus on the on-going and planned research at ECMWF. Finally, some remarks on the future of ensembles will be made in section 5.

2 The nine TIGGE operational, global, medium-range ensembles

To facilitate the comparison of the TIGGE ensembles we can group the sources of forecast errors into two broad classes, namely initial and model uncertainties. Although the two classes are not mutually exclusive, they are often used to characterize the configuration of each ensemble system. The classes are not mutually exclusive since it is impossible to separate initial and model uncertainties, given that initial conditions are constructed using model-based, data-assimilation procedures. Thus, we should take into account the fact that in some cases, 'initial condition errors' may actually be mainly due to model errors (that affected the initial conditions in the data assimilation process), and schemes designed to simulate the impact on forecasts of initial (model) uncertainties could actually also simulate the effect of model (initial) uncertainties.

Table A lists the main characteristics of the TIGGE OG-ENS, which together provide, at the time of writing (summer 2014), 502 forecasts every day (BMRC stopped production in July 2010, and UKMO stopped production in July 2014), spanning a forecast range between 10 and 16 days (32 days twice a week), with horizontal grid spacing between about 32 and 200 km.

Centre	Initial unc. method (area)	Model unc.	Truncation (degrees, km)	# Vert Lev (TOA, hPa)	Fcst length (d)	# pert mem	#runs per day (UTC)	# mem per day	In TIGGE since (/to)
BMRC (AU)	SV(NH,SH)	NO	TL119 (1.5°; 210km)	19 (10.0)	0-10	32	2 (00/12)	66	Sep-07 (/Jul-10)
CMA (CHI)	BV(globe)	NO	T213 (0.56°; 70km)	31 (10.0)	0-10	14	2 (00/12)	30	May-07
CPTEC (BR)	EOF(40S:30N)	NO	T126 (0.94°, 120km)	28 (0.1)	0-15	14	2 (00/12)	30	Feb-08
ECMWF (EU)	SV(NH, SH, TC) + EDA(globe)	YES	TL639 (0.28°; 35km) TL319 (0.56°; 70km)	91 (0.1)	0-10 10-15/32	50	2 (00/12)	102	Oct-06
JMA (JAP)	SV(NH, TR, SH)	YES	TL479 (0.38°; 50km)	60 (0.1)	0-11	25	2 (00/12)	52	Aug-11
KMA(KOR)	ETKF(globe)	YES	N320 (0.35°; 40km)	70 (0.1)	0-10	23	4 (00/06/12/18)	96	Dec-07
MSC (CAN)	EnKF(globe)	YES	600x300 (0.6°, 75km)	40 (2.0)	10-16/32	20	2 (00/12)	42	Oct-07
NCEP (USA)	ETR(globe)	YES	T254 (0.70°; 90km) T190 (0.95°; 120km)	28 (2.7)	0-7 7-16	20	4 (00/06/ 12/18)	84	Mar-07
UKMO (UK)	ETKF(globe)	YES	N216 (0.45°; 60km)	70 (0.1)	0-15	23	2 (00/12)	48	Oct-06 (/Jul-14)

Table A. Main characteristics (at the time of writing, summer 2014) of the nine TIGGE operational, global medium-range ensembles (OG-ENS), listed in alphabetic order (column 1): initial uncertainty method (column 2), model uncertainty simulation (Y/N, column 3), truncation and approximate horizontal resolution (column 4), number of vertical levels and top of the atmosphere in hPa (column 5), forecast length in days (column 6), number of perturbed members for each run (column 7), total number of members (including the control forecast) per day (column 8) and date since when data have been stored in TIGGE (column 9). Note that all but two ensembles continue to be archived in the TIGGE database: the BMRC OG-ENS data are available only up to 20 July 2010, and the UKMO data only up to 15 July 2014.

All ensembles simulate initial uncertainties with none following exactly the same methodology. Broadly speaking (see sections 2.1-2.9 for details), three OG-ENS use singular vectors (SVs; BMRC, ECMWF and JMA), one uses bred-vectors (BVs; CMA), one uses an ensemble Kalman filter (MSC), one uses perturbations computed using empirical orthogonal functions (EOFs; CPTEC), and three use initial perturbations computed by rotating and/or re-scaling bred-vectors (KMA, NCEP and UKMO). Model uncertainties are simulated in six of the nine OG-ENS using different methods (see sections 2.1-2.9 for details). In terms of spatial resolution, the OG-ENS differ substantially: horizontally their grid spacing ranges from about 200 km (CPTEC) to about 32 km (ECMWF, up to forecast day 10), and vertically the number of levels ranges from 19 (BMRC, with the top of the atmosphere at 10 hPa, about 30 km), to 91 levels (ECMWF, with the top of the atmosphere at 0.1 hPa, about 80 km). Forecast length varies from 10 to 16 days (32 days only few times a week). In terms of membership, each individual ensemble includes between 14 and 50 perturbed members. In terms of daily production frequency, two centres update their ensembles four times a day, six run them twice and one only runs it once a day.

Figure 1 visualises the characteristics of the nine OG-ENS along four key computational cost-drivers: horizontal resolution squared (HR^2, with HR in km; this takes into account that a doubling of the resolution increases the number of grid points by a factor of 4), number of vertical levels (LEV), number of members per day (#M) and forecast length (FCD, in days). These cost drivers, although not exhaustive (e.g. we have not included the time step), can be used to get an estimate of the relative cost of each ensemble. Each ensemble cost-driver has been normalized with respect to the most costly number, i.e. 32 for HR, 91 for LEV, 102 for #M and 16 for FCD. The relative amount of resources required to produce each OG-ENS forecasts per day, r_n, has been defined as the product of the relative cost drivers:

Eq. (1) $$r_n = (HR_n/32)^2 \cdot (LEV_n/91) \cdot (\#M_n/102) \cdot (FCD_n/16)$$

At the time of writing (summer 2014), the ECMWF OG-ENS requires the highest amount of resources to be produced, due its finest resolution and largest membership, with the KMA being the second one (Fig. 1). Compared to the ECMWF OG-ENS, the KMA OG-ENS requires about 42% (of the ECMWF resources), the JMA and the UKMO OG-ENSs about 14%, the MSC OG-ENS about 5%, and the others less than 5%. As will be discussed in section 3.3, there is a rather strong link between relative cost and performance.

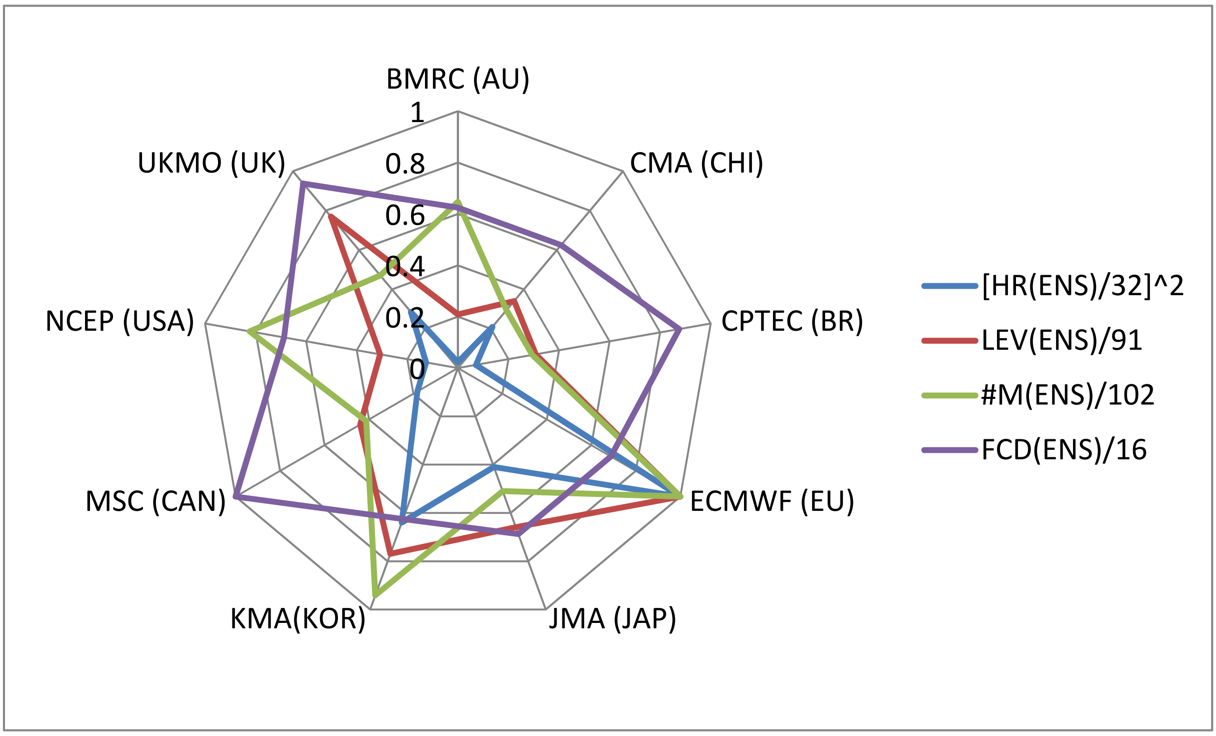

Figure 1. Comparison of the OG-ENS four key computational cost drivers: squared horizontal resolution (HR^2), number of vertical levels (LEV), number of perturbed members per day (#M) and forecast length (FCD, in days). Each ensemble cost-driver has been normalized with respect to the most costly number (32 km for HR, 91 for LEV, 102 for #M and 16 for FCD).

Each TIGGE forecast is defined by the numerical integration of the model equations adopted by the production center to describe the earth-system model (these equations are not exactly the same: the reader is referred to the references listed in sections 2.1-2.9, and references therein, for more information). In other words, each single T-hour forecast starting at day d is given by the time integration of a set of model equations from initial time 0 to time T:

Eq. (2) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_j(t) + P_j(t) + dP_j(t)]dt$$

where A_j is the tendency due to the adiabic processes (say advection, Coriolis force, pressure gradient force), P_j is the tendency due to the parameterized physical processes (say convection, radiation, turbulence, ..) and dP_j represents the tendency due to stochastic model errors and unresolved processes.

In the MSC OG-ENS, each numerical integration starts from initial conditions defined by an independent data assimilation procedure:

Eq. (3) $$e_j(d,0) = F[e_j(d - T_A; T_A), o_j(d - T_A, d)]$$

where *F[..,..]* denotes the data assimilation process of merging the model first guess $e_j(d - T_A, T_A)$

and the observations spanning the time period T_A (the window covered by the data assimilation process), from *(d-T_A)* to *d*. The first guess $e_j(d - T_A; T_A)$ is given by the T_A-hour integration of the model equations from *(d-T_A)* to *d*. The data assimilation process uses the observations $o_j(d - T_A, d)$, taken between *(d-T_A)* and *d*, and available at the time when the data assimilation procedure starts. More precisely, as it will be discussed in section 2.7, each member initial conditions are selected among the 192 members of their Ensemble Kalman Filter.

In all the other eight OG-ENS, each numerical integration starts from initial conditions defined by adding a perturbation to an unperturbed initial state:

Eq. (4) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

Eq. (5) $$e_0(d,0) = F[e_0(d - T_A; T_A), o_0(d - T_A, d)]$$

where the unperturbed initial conditions are defined by the data assimilation process spanning the time period T_A. The initial perturbations $de_j(d,0)$ are defined in different ways in each OG-ENS.

Equations (2-5) provide a simple, unified framework to describe how the *j-th* member of the 502 TIGGE forecasts is produced every day. In the forthcoming sub-sections, we will discuss how the initial and model uncertainties are simulated in the TIGGE OG-ENS (in alphabetic order). For each system, the discussion will start from the original setting, implemented when the ensemble started operational production, mentioning relevant features and quoting key references, and will finish with a description of the configuration at the time of writing (summer 2014).

2.1 The BMRC (Bureau of Meteorology Research Center, Australia) ensemble

The BMRC OG-ENS started production in July 2001, and stopped in July 2010, when BMRC decided to adopt the UK data-assimilation and forecasting system. BMRC data are available in TIGGE from July 2007 to July 2010. BMRC is planning to re-start their medium-range ensemble production in 2015.

In 2010, when BMRC stopped production, the BMRC OG-ENS had the same configuration as when it started production in 2001. It comprised 33 members, one unperturbed and 32 perturbed ones (*Bourke et al* 1994, 2004), and forecasts were run twice a day up to 10 days. The system had a spectral triangular truncation T_L119L19 (about 1.5 degrees, 160 km in physical space and 19 vertical levels), with the top of the atmosphere at 10 hPa. The forecast model included only a description of land and atmospheric processes (no wave or ocean model component was used).

If we now reconsider Equations (2-5), this is how they can be written for the BMRC OG-ENS:

Eq. (6) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t)]dt$$

where A_0 and P_0 represented the 'unperturbed' model dynamical and physical tendencies (i.e. there was only one dynamical core and one set of parameterisations called with the same parameters, and no model error scheme), and the forecast length T was 10 days. The ensemble did not simulate model uncertainties.

The initial conditions were defined by adding perturbations to the unperturbed initial conditions (ICs):

Eq. (7) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

The unperturbed ICs were produced by interpolating to the ensemble resolution the initial conditions defined by the BMRC 3-dimensional multivariate statistical interpolation scheme. The initial perturbations were defined by T42L19 singular vectors (SVs), computed as in the first version of the ECMWF OG-ENS (see section 2.4 for more details).

The individual perturbations were defined by linear combinations of the singular vectors through an orthogonal phase space rotation followed by an amplitude scaling factor, as was done in the original ECMWF OG-ENS (*Molteni et al 1996)*:

Eq. (8) $$de_j(d,0) = \sum_{k=1}^{32} \alpha_{j,k} SV_k$$

The SVs were computed only over the Southern Hemisphere (SH, for all grid points with latitude λ<20°S), and were scaled to have amplitudes, locally, comparable to analysis error estimates.

2.2 The CMA (China Meteorological Administration) ensemble

The CMA OG-ENS started production in 2001, with a T213L31 resolution (about 0.56 degrees, 70 km in physical space, and 31 vertical levels). CMA OG-ENS data have been available in the TIGGE archive since May 2007.

At the time of writing (summer 2014), the CMA OG-ENS comprises 15 members, one unperturbed and 14 perturbed members (*Su et al* 2014). Forecasts are run twice a day, at 00 and 12 UTC, up to 10 days. The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used). The ensemble does not simulate model uncertainties.

If we now reconsider Equations (2-5), this is how they can be written for the CMA OG-ENS:

Eq. (9) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t)]dt$$

where A_0 and P_0 represents the 'unperturbed' model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters, and no model error scheme), and the forecast length T is 10 days.

The initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (10) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

Eq. (11) $$de_j(d,0) = BV_j(d,0)$$

The unperturbed ICs are produced by interpolating to the T213L31 analysis to the ensemble resolution.

The initial perturbations are defined by bred-vectors, computed as in the original NCEP ensemble (*Toth & Kalnay* 1997; see section 2.8).

2.3 The CPTEC (Centro de Previsao de Tempo e Estudos Climatico, Brazil) ensemble

The CPTEC OG-ENS has been producing ensemble forecasts since 2001. Its data have been available in the TIGGE archive since March 2008.

At the time of writing (summer 2014), the CPTEC OG-ENS has a T126L28 resolution, with a corresponding 0.94 degree grid (about 120 km). In the vertical, it has 28 vertical levels up to 0.1 hPa. It uses a spectral model, has 15 members (one unperturbed and 14 perturbed), runs twice a day (at 00 and 12 UTC) and produces forecasts up to 15 days. The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used). The ensemble does not simulate model uncertainties.

If we now reconsider the Equations (2-5), this is how they can be written for the CPTEC OG-ENS:

Eq. (12) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t)]dt$$

where A_0 and P_0 represents the 'unperturbed' model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters, and no model error scheme), and the forecast length T is 15 days.

The initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (13) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

Eq. (14) $$de_j(d,0) = EOF_j(d,0)$$

The unperturbed ICs are defined by the NCEP operational analysis (see section 2.8), interpolated to the CPTEC resolution.

The initial perturbations are defined using empirical orthogonal functions (*Coutinho* 1999, *Zhang & Krishnamurti* 1999). The method consist of (a) computing 36-hour bred-vectors (by adding random perturbations to unperturbed initial conditions and running pairs of 36-hour forecasts), (b) constructing a time series of these bred vectors, and (c) performing an empirical orthogonal function (EOF) analysis of this time series to obtain the fastest growing perturbations. These EOF-based perturbations are computed for the region 45°S-30°N.

2.4 The ECMWF (European Centre for Medium-Range Weather Forecasts, Europe) ensemble

The ECMWF OG-ENS has been producing ensemble forecasts from 19 November 1992. Its data have been available in the TIGGE archive since October 2006.

At the time of writing (summer 2014), the ECMWF OG-ENS comprises 51 members, one unperturbed and 50 perturbed ones. Forecasts are run with a variable resolution (*Buizza et al* 2007): T_L639L91 (spectral triangular truncation T639 with a linear grid, which corresponds to about 32 km spacing in physical space, and 91 vertical levels) during the first 10 days, and T_L319L91 (i.e. about 65 km spacing) thereafter. Forecasts are run twice a day, with initial times at 00 and 12 UTC, up to 15 days; at 00 UTC on Mondays and Thursdays the forecasts are extended to 32 days (*Vitart et al* 2008). The forecasts are coupled to an ocean wave and a dynamical ocean circulation model.

The ocean wave model is WAM (*Janssen et al* 2005, 2013): it has a 55 km resolution, and 24 directions and 30 frequencies up to day 10, and 12 directions and 25 frequencies afterwards, and it is coupled to the atmosphere every time step,.

The ocean model is NEMO (the Nucleus for European Ocean Modelling), with the ORCA100z42 grid, which has a 1-degree horizontal resolution and 42 vertical layers. NEMO is a state-of-the-art modelling framework for oceanographic research, operational oceanography seasonal forecast and climate studies, developed by the NEMO Consortium (http://www.nemo-ocean.eu/). The reader is referred to *Mogensen et al* (2012a, b), and references therein, for a description of how NEMO and NEMOVAR have been coupled and implemented at ECMWF.

If we now reconsider the Equations (2-5), this is how they can be written for the ECMWF OG-ENS:

Eq. (15) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t) + dP_j(t)]dt$$

where A_0 and P_0 represents the ‘unperturbed’ model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters), dP_j represents the model uncertainty simulated using two model error schemes, the SPPT (*Buizza et al* 1999; *Palmer* *et al* 2009) and SKEB (*Berner et al* 2008, *Palmer et al* 2009) schemes, and the forecast length *T* is 15 or 32 days.

For the atmosphere, the initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (16) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

The unperturbed ICs are given by the ECMWF high-resolution 4-dimensional variational assimilations (4DVAR), run at T_L1279L137 resolution and with a 12-hour assimilation window, interpolated from the T_L1279L137 resolution to the T_L639L91 ensemble resolution.

The perturbations are defined by a linear combination of singular vectors (SVs, *Buizza and Palmer* 1995) and perturbations defined by the ECMWF Ensemble of Data Assimilations (EDA, *Buizza et al* 2008, *Isaksen et al* 2010):

Eq. (17) $$de_j(d,0) = \sum_{a=1}^{8} \sum_{k_a=1}^{50} \alpha_{j,k_a} SV_{k_a} + [\, f_{m(j)}(d-6,6) - < f_{m=1,25}(d-6,6) >]$$

The SVs are computed over up to 8 areas (NH: all grid points with points with latitude λ>30°N; SH: all grid points with latitude λ<30°S; Tropics: up to 6 areas where tropical depressions have been reported). The SVs, the fastest growing perturbations over a 48-hour time interval, are computed at T42L91 resolution. SVs optimized to have maximum total-energy growth over the different areas, are linearly combined and scaled to have amplitudes, locally, comparable to analysis error estimates provided by the ECMWF high-resolution 4DVAR.

Each EDA member is generated by an independent 4DVAR with a lower resolution (T_L399) than the high-resolution 4DVAR (T_L1279). Each EDA member uses perturbed observations, with the observations' perturbations sampled from a Gaussian distribution with zero mean and the observation error standard deviation. Each EDA member non-linear trajectory is generated using also the Stochastically Perturbed Parameterized Tendencies (SPPT) scheme to simulate model uncertainties (see below for a description of the SPPT scheme). Since November 2013, the EDA has been including 25 independent 4DVAR run at T_L399L137 resolution and with a 12-hour assimilation window (the EDA only had 11 members before November 2013).

SV- and EDA-based perturbations are combined as in Eq. (17). The EDA-based perturbations are defined by differences between 6-hour forecasts started from the most recent available EDA analyses (these analyses are valid for 6-hours earlier than the ENS initial time). Differences are computed between each of the 25 perturbed forecasts and their ensemble-mean, and the 25 perturbations are added and subtracted to the unperturbed analysis. SV- and EDA-based perturbations are defined such that full symmetry is maintained in the ENS initial perturbations [i.e. even member (2n) has minus the total perturbation of odd member (2n-1), for n=1, 25].

The definition of the initial conditions for the ocean is different: these are defined by the 5-member ensemble of ocean analyses, produced by NEMOVAR, the NEMO 3-dimensional variational assimilation system (*Mogensen et al* 2013). Each ocean analysis is generated using all available in situ temperature and salinity data, an estimate of the surface forcing from ECMWF short range atmospheric forecasts, sea surface temperature analysis and satellite altimetry measurements. One member is generated using unperturbed wind forcing provided by the high-resolution 4DVAR, while the other 4 members are generated using perturbed versions of the unperturbed wind forcing.

Model uncertainties are simulated only in the free atmosphere (i.e. not in the land surface, nor in the ocean), using two stochastic schemes: the Stochastically Perturbed Parameterized Tendency (SPPT, *Buizza et al* 1999, *Palmer et al* 2010) and the backscatter (SKEB, *Shutts* 2005, *Berner et al* 2008)) schemes. SPPT is designed to simulate random model errors due to parameterized physical processes; the current version uses 3 spatial and time level perturbations. SKEB is designed to simulate the upscale energy transfer induced by the unresolved scales on the resolved scales.

Since March 2008, when the ECMWF medium-range and monthly ensembles were joined, a key component of the ECMWF OG-ENS used to generate some bias-corrected and/or calibrated products, has been the re-forecast suite (*Vitart et al* 2008; *Leutbecher & Palmer* 2008). This suite includes a 5-member ensemble run once a week with the operational configuration (resolution, model cycle, ..) for the past 20 years. These re-forecasts are used to estimate the model climate required to generate some ensemble products (e.g. the Extreme Forecast Index, or weekly-average anomaly maps) and to calibrate the ENS forecasts.

This is how the re-forecasts are used to generate some of the operational ENS products. For each date (e.g. 14 December 2012), 500 forecasts are generated by running 5-member ensembles for 5 weeks for the past 20 years. More precisely, the ICs of these re-forecast ensembles are defined by the same day of the week of the 5 weeks centred on the current date (in this case, 1, 7, 14, 21 and 28 December). These re-forecasts start from ECMWF ERA-Interim re-analyses instead of the operational analyses, since for old dates ERA-Interim provides more accurate ICs. In the re-forecast suite, the simulation of the initial uncertainties is slightly different from the one in the operational suite: the re-forecast ensembles use singular vectors of the day (which are re-computed, as in the operational suite), but use EDA-based perturbations of the current year instead of the 'correct year', since the EDA is not available for the past 20 years (the EDA started running in 2010, see *Buizza et al* 2008 for more details). *Buizza et al* (2008) describes the rationale behind this choice, and shows that despite this choice the re-forecast and the forecast ensembles perform very similarly. Once the 51-member forecasts and the 500-member re-forecasts are completed, products such as probabilities of anomalies (computed with respect to the model climate), or the Extreme Forecast Index (EFI, which needs the model climatological distribution function to be computed, see *Lalaurette* 2003) are generated.

The most recent change of the ECMWF OG-ENS configuration was implemented in November 2013, when the top of the model was raised from 5 to 0.1 hPa, the number of vertical levels was increased from 62 to 91, and the atmosphere and ocean models were coupled from initial time (rather than from day 10 as was done up to that time). Furthermore, surface initial perturbations were added to the ENS initial perturbations (see section 4.3 in *Balsamo et al* 2014), and the coupling to the ocean model NEMO was moved forward in time, from forecast day 10 to the initial time. These changes, together with major upgrades in the model parameterisations, led to further improvements of the ensemble skill, especially in the longer forecast range. The reader is referred to *Vitart et al* (2014) for a discussion of the evolution of the skill of the ECMWF ensemble forecasts over the sub-seasonal forecast range.

2.5 The JMA (Japan Meteorological Agency) ensemble

The JMA OG-ENS has been producing ensemble forecasts since March 2001. Its data have been available in the TIGGE archive since August 2011.

At the time of writing (summer 2014), the JMA OG-ENS (*Yamaguchi and Majumdar* 2010) includes 25 forecasts with a resolution T_L474L60 (spectral triangular truncation with linear grid, 0.375 degrees spacing which corresponds to about 50 km in physical space), with the top of the atmosphere at 0.1 hPa. The most recent change was introduced in March 2014, when the resolution was increased from T_L319 to T_L479, and the daily configuration was changed from producing 51 forecasts once a day (at 12 UTC), to producing 26 twice a day (at 00 and 12 UTC).

Initial uncertainties are simulated using SVs, computed at T63L40 resolution, over the NH and the SH extra-tropics (north and south of 30°) with a 48-hour optimisation time interval, and over the tropics (30°S-30°N) with a 24-hour optimisation time interval. Model uncertainties are simulated using a stochastic scheme similar to the original ECMWF SPPT (*Buizza et al* 1999). The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used).

If we now reconsider the Equations (2-5), this is how they can be written for the JMA OG-ENS:

Eq. (18) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t) + dP_j(t)]dt$$

where A_0 and P_0 represents the 'unperturbed' model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters), dP_j represents the model uncertainty simulated using the JMA stochastic scheme, and the forecast length T is 11 days. The initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (19) $$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

The unperturbed ICs are given by the JMA high-resolution 4DVAR, which has a resolution T_L959L100 (about 20km in physical space), interpolated at the ensemble resolution.

The perturbations are defined by a linear combination of SVs computed over three regions, NH, SH and tropics:

Eq. (20) $$de_j(d,0) = \sum_{a=1}^{3} \sum_{k_a=1}^{25} \alpha_{j,k_a} SV_{k_a}$$

It is worth mentioning that JMA also produces monthly forecasts with a 50-member lagged ensemble, which has a T_L319L60 resolution (about 70 km in grid point space). This monthly ensemble uses bred-vectors (*Toth & Kalnay* 1993, 1997) instead of SVs, to define the initial perturbations. Monthly products are issued once a week, on Fridays, and are based on the 25 forecasts started the previous Wednesday and Thursday at 12 UTC. The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used), and does not simulate model uncertainties. The

JMA monthly forecasts are expected to become available in the Sub-seasonal-to-Seasonal (S2S) archive (see section 5).

2.6 The KMA (Korea Meteorological Agency) ensemble

The KMA OG-ENS has been producing ensemble forecasts since March 2000. The data has been available in the TIGGE archive since October 2007.

The original KMA OG-ENS used initial perturbations defined by a breeding method (*Goo et al* 2003). It included 16 perturbed members, run with a T106L21 resolution, with the 16 initial perturbations defined by rotated bred vectors. The forecasts were run once a day (at 12 UTC) up to 10 days. It did not simulate model uncertainties, and included only a description of land and atmospheric processes (no wave or ocean model component is used).

Since 2011, KMA has been generating its operational forecasts using the Unified Model (UM) and related pre/post processing system imported from the UK Met Office (*Kay et al* 2013). Thus, since then, the KMA OG-ENS has been practically the same as the UKMO OG-ENS (see section 2.9).

At the time of writing (summer 2014), the KMA OG-ENS is based on 24 members, one control and 23 perturbed ones, with initial perturbations generated using an Ensemble Transformed Kalman Filter (ETKF) method with localization (*Bowler et al* 2008; *Kai & Kim* 2014). It has a horizontal resolution of approximately 40 km and 70 vertical levels (N320L70).

Model uncertainties are simulated using stochastic-physics schemes that consist of 'random parameters' and 'stochastic convective vorticity' schemes (*Bowler et al.* 2008).

If we now reconsider the Equations (2-5), this is how they can be written for the KMA OG-ENS:

Eq. (21) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_0(t) + dP_j(t)]dt$$

where A_0 and P_0 represents the 'unperturbed' model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters), dP_j represents the model uncertainty simulated using the KMA stochastic scheme, and the forecast length T is 10 days.

The initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (22) $$e_j(d,0) = e_0(d,0) + ETKF_j(d,0)$$

The unperturbed ICs are given by the KMA version of the UKMO 4DVAR system.

The perturbations are defined by ETKF perturbations (see section 2.9).

2.7 The MSC (Meteorological Centre of Canada) ensemble

The MSC OG-ENS has been producing ensemble forecasts since February 1998. Its data have been available in the TIGGE archive since October 2007. Since August 2011, the resolution of MSC OG-ENS has been N600xN300 (600 grid points in longitude and 300 in latitude), which corresponds to about 0.6 degrees, 75 km.

The MSC perturbed-observation approach attempts to represent all sources of uncertainty by adding random perturbations to as many system components as possible. The Canadian Ensemble Kalman Filter (EnKF) has been used to supply initial conditions (*Houtekamer et al* 2009, 2014), with sources of uncertainty simulated by different sets of random perturbation schemes (*Houtekamer et al.* 1996; *Houtekamer & Lefaivre* 1997, *Anderson* 1997).

Compared to the 1995 configuration, the number of the MSC ensemble members has increased from the original 8 to 16, and stands now at 20. The number of the members of the MSC EnKF used to define the initial conditions has also increased from 48 to 96, and now stands at 192.

The most recent change of the MSC OG-ENS was implemented in December 2013, when the treatment of the surface temperature of the sea (SST) was changed. In the new ensemble, the SST evolves while persisting the anomaly (deviation from the climatology). Since December 2013, monthly forecasts have been produced on Thursdays by extending the forecast length to 32 days. Furthermore, re-forecasts based on 4-member ensembles have been run once a week for the past 18 years, so that the model climate can be estimated, and calibrated products can be generated.

At the time of writing (summer 2014), the MSC OG-ENS includes 21 members, one unperturbed and 20 perturbed, and is run twice a day (at 00 and 12 UTC) up to forecast day 16. Once a week, at 00 UTC of each Thursday, the ensemble is extended to 32 days. The initial conditions are obtained directly from the Canadian EnKF. Model uncertainties are sampled using four schemes: isotropic perturbations at the initial time, different physical physical parameterisations, stochastic physical tendency perturbations and stochastic kinetic energy backscatter (*Houtekamer et al* 2009). The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used).

If we now reconsider the Equations (2-5), this is how they can be written for the MSC OG-ENS:

Eq. (23) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_j(t) + dP_j(t)]dt$$

where A_0 represents the 'unperturbed' model dynamical core, Pj represents the physical tendencies, which varies in each member since different parameterisation schemes and/or different parameters are used, and dP_j represents the model uncertainty simulated using different stochastic schemes.

Each member ICs are defined by one of the EnKF members:

Eq. (24) $$e_j(d,0) = ENKF_{k(j)}(d,0)$$

It is worth mentioning that the MSC and the NCEP OG-ENS forecasts are exchanged in real time to generate the multi-model products of the North American Ensemble Forecast System (NAEFS, *Candille* 2009; see also the Canadian web site: http://weather.gc.ca/ensemble/naefs/index_e.html). NAEFS is a joint project involving MSC, the United States National Weather Service (NWS) and the National Meteorological Service of Mexico (NMSM). NAEFS, launched in November 2004, provides users with operational products generated by blending the MSC and the NCEP OG-ENSs. The research, development and operational costs of the NAEFS system are shared among the three partners.

2.8 The NCEP (National Centers of Environmental Prediction, United States of America) ensemble

The NCEP OG-ENS has been producing ensemble forecasts since December 1992. Its data have been available in the TIGGE archive since March 2007.

The original version of the NCEP OG-ENS simulated only initial uncertainties using bred-vectors (BVs, *Toth & Kalnay* 1993, 1997). The breeding method involves the maintenance and cycling of perturbation fields that develop between two numerical model integrations. These fields, once re-scaled, define the initial perturbations. In its original form with a single global rescaling factor, the BVs represented a non-linear extension of the Lyapunov vectors (*Boffetta et al.* 1998). In the operational NCEP OG-ENS, multiple breeding cycles are used, each initialized at the time of implementation with independent perturbation fields ("seeds"). The original system was based on 10 perturbed ensemble members, run both at 00 and 12 UTC every day out to 16 days lead-time. For both times, the generation of the initial perturbations was done in 5 independent breeding cycles, originally started with different perturbations, using a regional rescaling algorithm. Since then, the method used to define the initial perturbations has been upgraded several times.

At the time of writing (summer 2014), the NCEP OG-ENS consists of four runs a day (at 00, 06, 12 and 18 UTC), with a T382L64 resolution up to forecast day 7, and a T126L64 resolution from day 7 to 16. Each run includes one unperturbed and 20 perturbed forecasts. The forecast model includes only a description of land and atmospheric processes (no wave or ocean model component is used).

The initial perturbations are now generated using the 'Ensemble Transform with Rescaling' (ETR; *Wei et al* 2006, 2008) technique. The ETR method is an extension of the original breeding approach (in an ensemble with only two members, both methods should produce the same perturbations). To improve the simulation of initial uncertainties in cases of tropical storms, the perturbed initial conditions are generated using also a tropical storm relocation method (*Liu et al* 2006, *Snyder et al* 2010).

In the current NCEP OG-ENS, model uncertainties are represented using the Stochastic Total Perturbation Scheme (STTP, *Hou et al* 2008), designed to represent model uncertainty by adding a stochastic forcing term to the total tendency. For each 6-hour forecast period, this term is defined by a linear combination of the past ensemble tendencies. In the linear combination, the total tendencies are rescaled so that, on average, the ensemble standard deviation matches the error of the ensemble-mean.

If we now reconsider the Equations (2-5), this is how they can be written for the NCEP OG-ENS:

Eq. (25) $$e_j(d;T) = e_j(d;0) + \sum_{k=1}^{N} \Delta T_{j,k} + \Delta S_{j,k}$$

Eq. (26) $$\Delta T_{j,k} = \int_{T_k}^{T_k+6} [A_{0,j}(t) + P_{0,j}(t)]dt$$

Eq. (27) $$\Delta S_{j,k} = \sum_{m=1}^{N} w_{m,k} \Delta T_{m,k}$$

where $A_{0,j}$ and $P_{0,j}$ represents the 'unperturbed' model dynamical and physical tendencies (i.e. there is only one dynamical core and one set of parameterisations called with the same parameters). The *j-th* subscript indicates that each finite-time tendency is different for each ensemble member, since it is computed starting from a different initial state.

Every 6 hours, for each member, the 6-hour tendency $\Delta T_{j,k}$ is computed by integrating ahead in time the model equations for 6 hours. Once the 6-hour tendency $\Delta T_{j,k}$ has been computed, the stochastic perturbation $\Delta S_{j,k}$ is defined by a linear combination of all 6-hour tendencies. It should be evident now why the scheme is called STTP, where 'T' stands for 'total': since the original tendencies include also the dynamical tendencies (by contrast, the ECMWF SPPT method does not perturb the tendencies due to the dynamics). Once the STTP term has been computed, the initial states are then advanced by 6-hour by adding $(\Delta T_{j,k} + \Delta S_{j,k})$. Then the process is repeated.

The initial conditions are defined by adding perturbations to the unperturbed ICs:

Eq. (28) $$e_j(d,0) = e_0(d,0) + ETR_j(d,0)$$

The unperturbed ICs are given by the NCEP T382L64 4-dimensional variational assimilation system.

As discussed in section 2.7, the NCEP and the MSC OG-ENS forecasts are exchanged in real time to generate the multi-model products of the North American Ensemble Forecast System (NAEFS, http://weather.gc.ca/ensemble/naefs/index_e.html).

2.9 The UKMO (United Kingdom Meteorological Office) ensemble

The UKMO OG-ENS started production in August 2005, and stopped in July 2014. Its data have been available in the TIGGE archive since October 2006. UKMO is investigating the possibility to re-start generating medium-range ensembles using a lagged approach.

The UKMO OG-ENS used an Ensemble Transform Kalman Filter approach (ETKF, *Wei et al* 2006, *Bishop et al* 2001, *Bowler et al* 2007, 2008), to generate the initial perturbations. The ETKF is a simplified version of the Ensemble Kalman Filter, a data assimilation scheme which updates the mean state of the atmosphere and the error covariance in that estimate using background information obtained from an ensemble. The ETKF can be viewed as a transformation of the NCEP error breeding scheme.

The initial perturbations were defined by a linear combination of the forecast perturbations from the previous cycle of the ensemble. The weights were calculated by considering the spread of the ensemble in the space of the observations (*Wang et al* 2004), ensuring that the perturbations are centred on the control analysis, and that they are orthogonal. The perturbations were inflated to ensure that the ensemble has the correct spread for the next analysis time (corresponding to t+12 hours), with the inflation factor calculated on-the-fly, so that the system automatically re-tuned itself to model changes.

Two stochastic physics schemes were used to represent the effects of structural and sub-grid-scale model uncertainties: the Random Parameters (RP) scheme and the Stochastic Convective Vorticity (SCV) scheme. The RP scheme treated a selected group of parameters as stochastic variables (*Lin & Neelin* 2000; *Bright & Mullen* 2002), and involved perturbing a total of 8 parameters from 4 different physical parameterisations (large-scale precipitation, convection, boundary Layer and gravity wave drag). The main aim of the SCV scheme (*Gray & Shutts* 2002) was to represent potential vorticity anomaly dipoles similar to the one typically associated with a meso-scale convective systems.

In July 2014, when production was stopped, the UKMO OG-ENS included one unperturbed and 24 perturbed members, run twice a day with a 60 km resolution and 70 vertical levels, and up to a forecast range of 15 days. The forecast model included only a description of land and atmospheric processes (no wave or ocean model component is used).

If we now reconsider the Equations (2-5), this is how they can be written for the UKMO OG-ENS:

Eq. (29) $$e_j(d;T) = e_j(d;0) + \int_0^T [A_0(t) + P_j(t) + dP_j(t)]dt$$

where A_0 represented the ‘unperturbed’ model dynamical tendencies (i.e. there was only one dynamical core), P_j represented the fact that the physical parameterisations were integrated by perturbing some key parameters as defined by the RP scheme, dP_j represented the model uncertainty simulated using the SCV scheme, and the forecast length T was 15 days.

The initial conditions were defined by adding perturbations to the unperturbed ICs:

Eq. (30) $$e_j(d,0) = e_0(d,0) + ETKF_j(d,0)$$

The unperturbed ICs were given by the UKMO high-resolution 4-dimensional variational assimilations, and the perturbations were defined by ETKF perturbations.

3 Performance of the TIGGE ensembles

It is evident from section 2 that there is not a *unique recipe* to generate reliable and skilful ensembles, and indeed the nine TIGGE OG-ENS use different methods to simulate the effects of the sources of forecast errors. Hereafter we will assess their performance, and discuss its link with their design and configuration. First, we will look at similarities and differences in the ensemble-mean and the spread (measured by the ensemble standard deviation, i.e. the spread measured around the ensemble-mean) for the case of 10 January 2013, and then we will look at seasonal average performance. We will focus on synoptic scales, represented by the 500 hPa geopotential height and the 850 hPa temperature, and on the 24-hour accumulated precipitation. The discussion will be limited to the eight TIGGE ensembles available for the period discussed (i.e. all but the BMRC one, which has stopped production in 2010).

Similar comparisons were performed in the past. In one of the first comparisons, *Park et al* (2008) concluded that there was a large difference between the performances of the single ensembles. For the 500 hPa geopotential height over the Northern Hemisphere (NH) in the medium-range (say around forecast day 5), the difference in predictability between the worst and the best control or ensemble-mean forecasts was about 2 days, while the difference between the worst and the best probabilistic predictions was larger, about 3-4 days.

More recently, *Hagedorn et al* (2012) not only discussed the TIGGE ensembles' performance, but also investigated the possibility of combining the TIGGE ensembles into a multi-model one. The skill of this multi-model ensemble was compared with the skill of an ensemble defined by calibrated ECMWF OG-ENS, with the calibration based on the ECMWF ensemble re-forecast suite (see section 2.4). This ECMWF calibrated ensemble was generated using re-forecasts for the past 18 years (see section 2.4). Considering the statistical performance of global probabilistic forecasts of 850 hPa and 2-meter temperatures, they concluded that a multi-model ensemble containing the nine TIGGE ensembles did not improve on the performance of the best single-model, the ECMWF OG-ENS. However, a reduced multi-model system, consisting of only the four best ensembles (ECMWF, MSC, NCEP, UKMO) showed an improved performance. They also concluded that the ECMWF OG-ENS was the main contributor for the improved performance of the multi-model ensemble; that is, if the multi-model system did not include the ECMWF contribution, it was not able to improve on the performance of the ECMWF OG-ENS alone. These results were shown to be only marginally sensitive to the choice of verification dataset.

Also in 2012, *Yamaguchi et al* (2012) looked at the tropics, and compared tropical cyclone (TC) track predictions from each TIGGE ensemble, with tracks generated by what they called a multi-center grand ensemble (MCGE) that included all of them. Their work considered 58 TCs in the western North Pacific from 2008 to 2010. In the verification of TC strike probabilities, the Brier skill score of the MCGE was larger than that of the best single ensemble, which was the ECMWF OG-ENS in the medium-range. By contrast, the reliability was improved by the MCGE, especially in the high-probability range.

3.1 Ensemble-mean and spread of the TIGGE ensembles for one specific case: comparison of forecasts started at 12UTC of 10 January 2013

First, let's consider OG-ENS forecasts for the 10th of January 2013 (12UTC), to illustrate how similar or different the ensemble-mean and the spread of the OG-ENS are. This case was randomly selected within the most recent seasons for which at least eight TIGGE ensembles were available, and for which routine scores were available in the ECMWF data-base.

Figures 2-9 show the eight TIGGE ensemble mean and spread, measured by the ensemble standard deviation. The fields are shown over a Euro-Atlantic sector at initial time (12 UTC of 10 January 2013) and at t+48h. This geographical region has been selected instead of the whole NH so that we can appreciate and see, in the maps, synoptic-scale differences.

If we start considering the ensemble-means, at initial time they are very similar, while at t+48h some small differences start to appear. This indicates that the eight centres' analyses are very similar in this region. Generally speaking, this is valid for the extra-tropical regions, but it is not true over the tropical region (say between 20°S-20°N), where the analyses differ more, as shown, e.g., in *Park et al* (2008)

By contrast, initial time differences are much larger in terms of ensemble spread. For example, the spread of the CPTEC OG-ENS (Fig. 3) is zero, since by construction perturbations are defined only up to 30°N. The JMA OG-ENS (Fig. 5) has spread geographically confined, since it is using only singular vectors that are localized in space (their spread is very similar to the one of the original ECMWF OG-ENS, which used only singular vectors). The comparison between the JMA OG-ENS and the ECMWF OG-ENS (Figs. 4-5) illustrates the impact of using EDA-based perturbations (the ECMWF OG-ENS) instead of the evolved SVs (the JMA OG-ENS): this replacement made the ECMWF initial perturbations less localized geographically and vertically (*Buizza et al* 2008).

The KMA, MSC and UKMO OG-ENSs have the largest initial spread: this is to compensate for their initial perturbations' slower growth rate. As shown, e.g., in *Buizza et al* 2005, their initial perturbations, generated using the EnKF or EKTF methods, grow slower than forecast error, and thus to get the right level of spread in the medium-range the initial perturbations have to be set with rather large amplitudes (see also Fig. 10 and related discussion in section 3.2).

At forecast time t+48h, the spread of all ensembles are more similar than at initial time, in terms of both coverage and local maxima. Broadly speaking, they are also rather similar in terms of structures, with maxima localized in the same areas, where the jet-stream is stronger or where cyclonic developments occur. This is not surprising since all these ensembles have been designed for the medium-range, say the 3-10 forecast range, and to achieve this, they have all been configured to have, on average, the right level of spread from about forecast day 2-3.

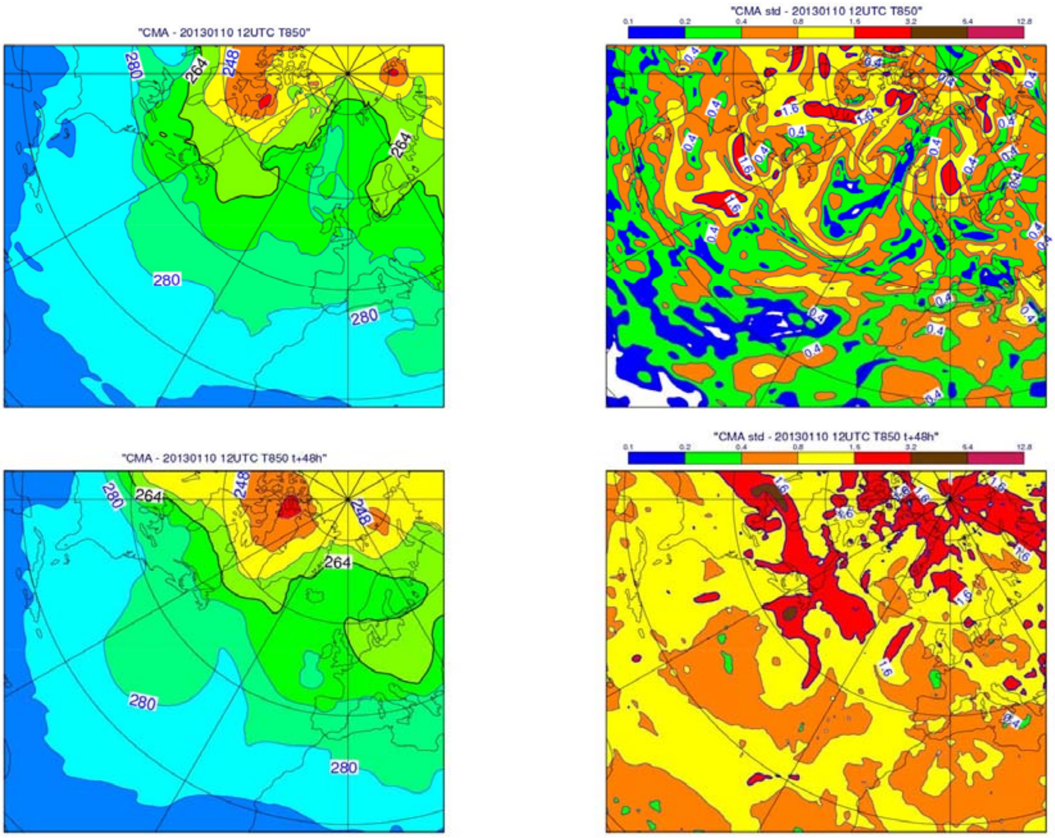

Figure 2. CMA OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

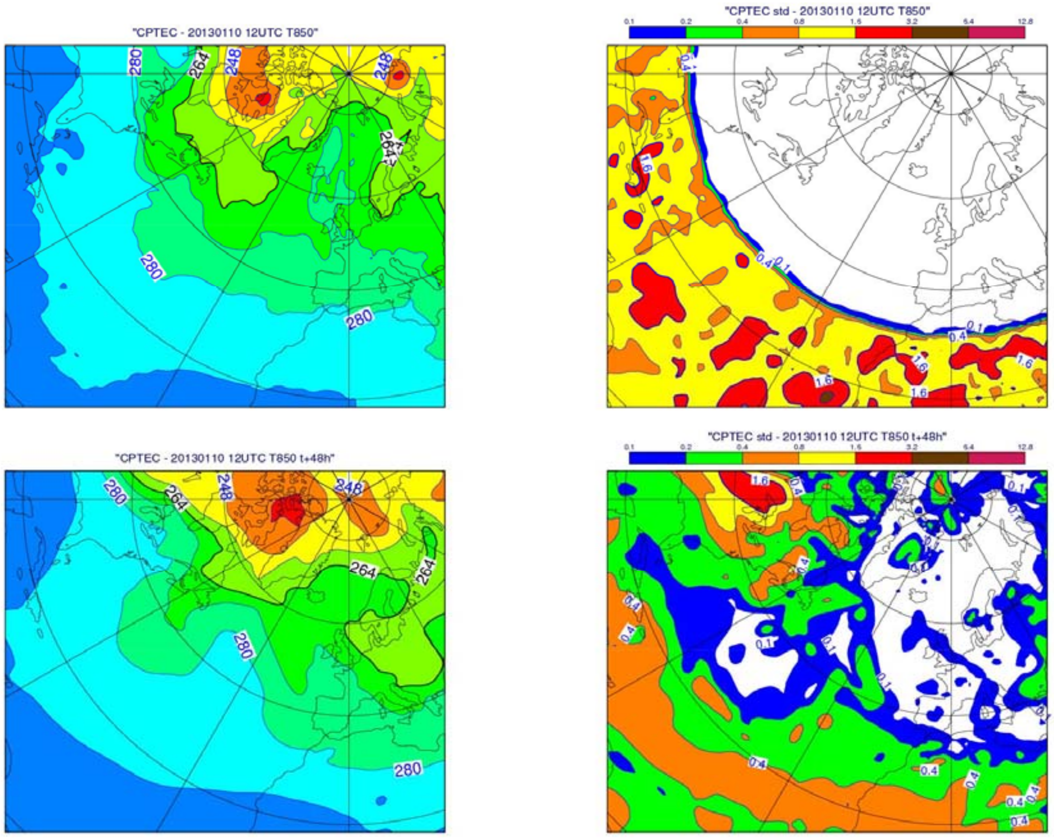

Figure 3. CPTEC OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

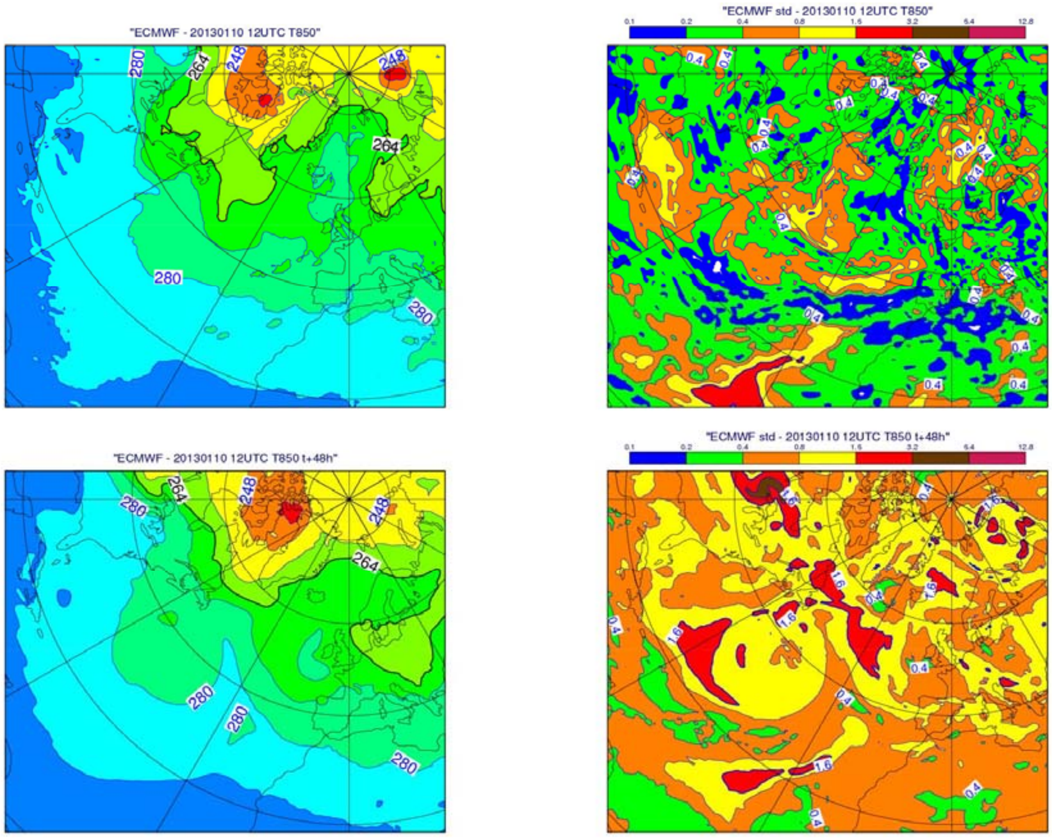

Figure 4. ECMWF OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

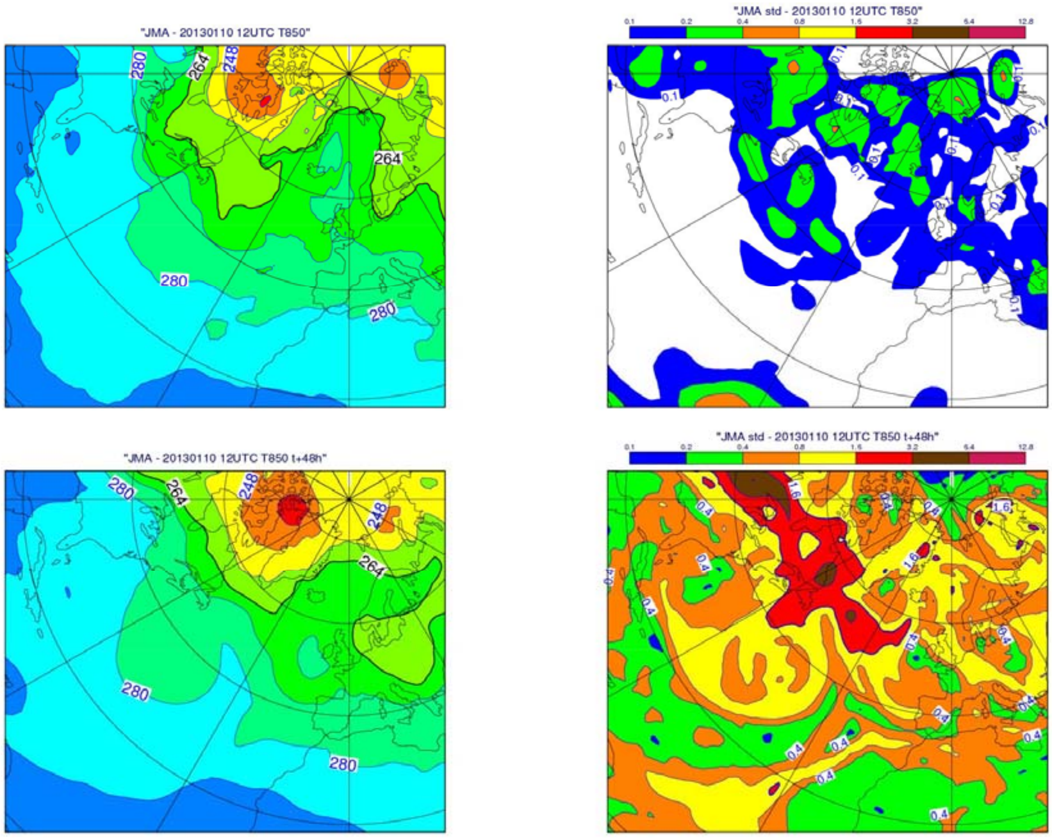

Figure 5. JMA OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

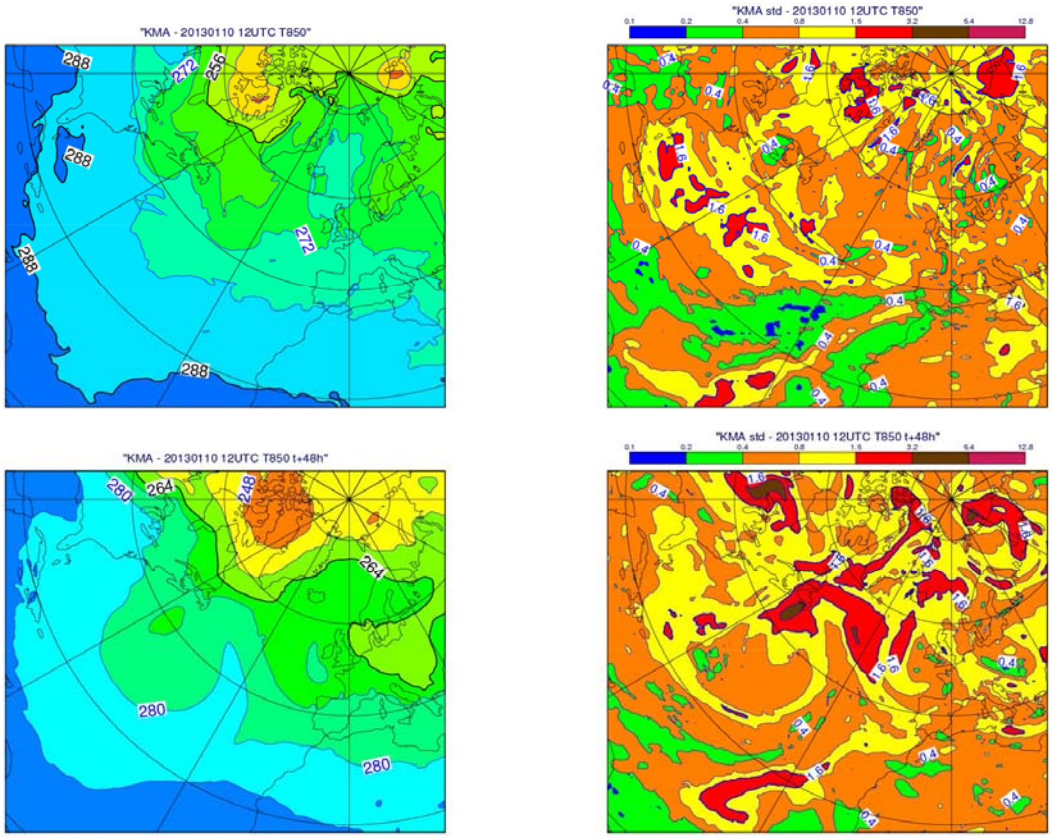

Figure 6. KMA OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

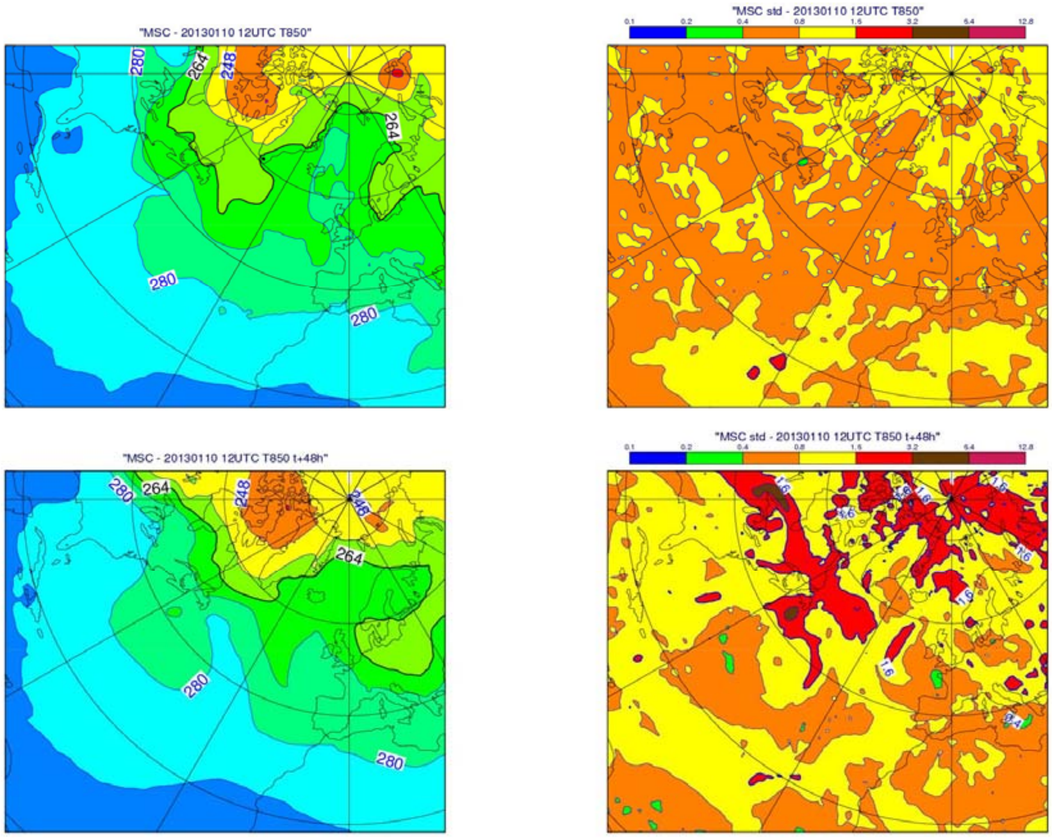

Figure 7. MSC OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

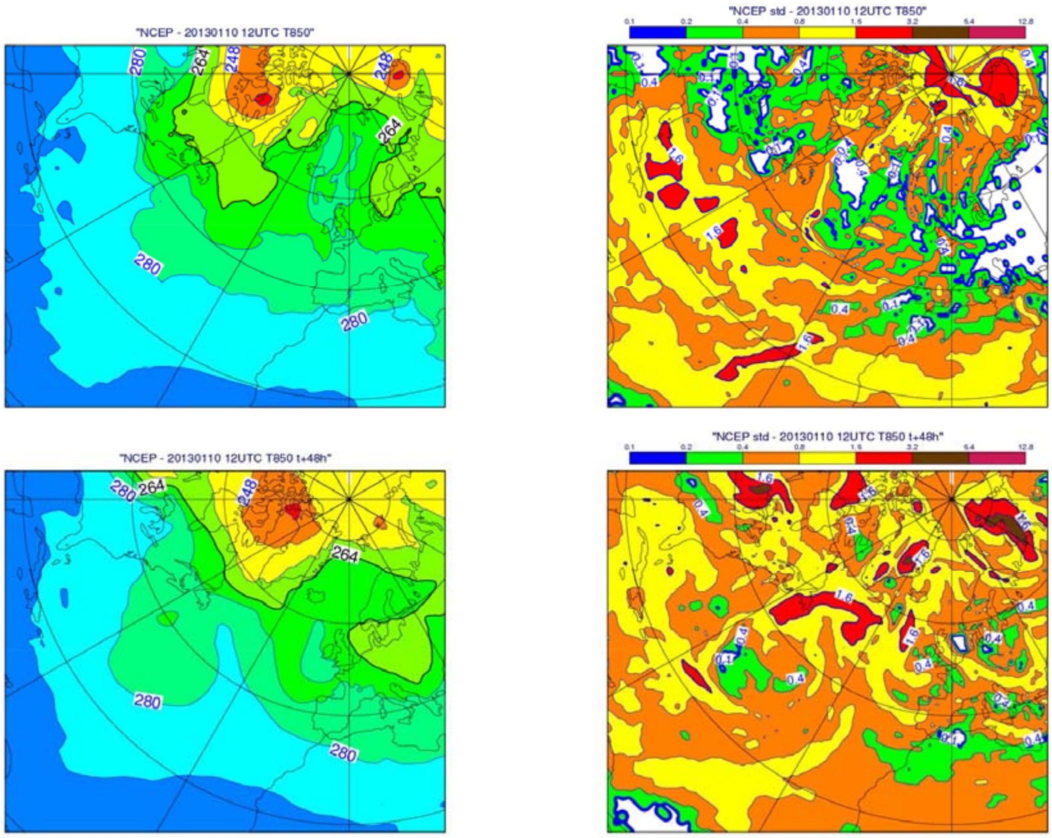

Figure 8. NCEP OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

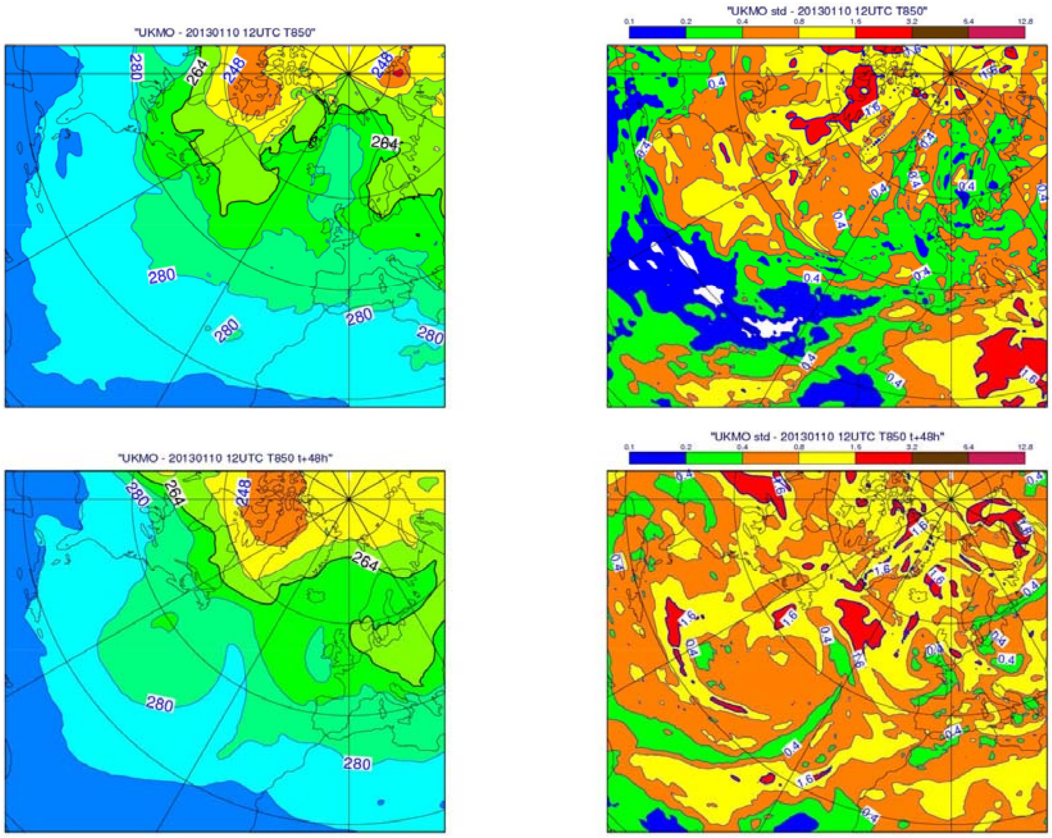

Figure 9. UKMO OG-ENS ensemble-mean (left panels) and spread (right panels) at initial time (top panels) and after 48 hours (bottom panels). The field shown is temperature at 850 hPa. The contour interval for the ensemble-mean is 8 degrees, and the spread shading is at 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 degrees.

Although single cases are interesting and can provide some indications on the characteristics of the different OG-ENS, a proper comparison of the performance of the different ensembles has to be based on a large number of cases. To this aim, in section 3.2 we will compare the average ensemble spread for a whole season, winter 2012-13 (i.e. December 2012, and January-February 2013). Then, in section 3.3 we will compare the accuracy of the ensembles' probabilistic forecasts.

3.2 Average ensemble spread of the TIGGE ensembles for winter 2012-13

Figure 10 shows the winter 2012-13 spread, measured for temperature at 850 hPa and average spatially over the Northern Hemisphere extra-tropics (NH, 20°N-90°N) and the tropics (TR, 30°S-30°N), from the initial time to forecast day 16.

The NH average results confirm the indications given by the case study, with the JMA OG-ENS showing the smallest initial spread and the CMA, KMA, MSC and UKMO OG-ENS the largest. If we compare the spread at t+6h (first point) with the values at 48-72 hours, it is evident, for example, the faster growth rate of SV-based perturbations (JMA OG-ENS) compared to the EnKF/ETKF ones (CMA, KMA, MSC and UKMO OG-ENSs).

The TR average results show a different picture, with the MSC OG-ENS showing the largest values. This is partly due to the fact that the EnKF initial perturbations are larger, but also because the computation of SVs in the tropics, where moist processes play a dominant role, is still sub-optimal, because they are computed at a very coarse resolution and still with a dry total energy norm (*Coutinho et al* 2004). Furthermore, tropical SVs are computed only for limited regions, and thus they do not cover the whole tropical band.

3.3 Average skill of the TIGGE ensembles probabilistic forecasts in summer and winter, over the tropics and the Northern Hemisphere extra-tropics

As in the previous section, first we are going to look at the winter 2012-13, 850 hPa temperatures over the NH and the TR regions. Then, we will consider more recent seasons for which routine performance measures have been computed at ECMWF for the available TIGGE ensembles. In all these verifications, each ensemble has been verified against its own analysis.

Performance has been assessed considering three metrics: reliability, measured by the match between the ensemble spread and the error of the ensemble-mean, the root-mean-square-error (RMSE) of the ensemble-mean, and the accuracy of probabilistic forecast measured by the continuous ranked probability score. Reliability is a very important attribute of ensembles. In a reliable ensemble, an event that is predicted to have a p% probability to happen, should occur p% of the times. Thus reliability can be measured by looking at average scatter diagrams of forecast probabilities versus occurrence frequencies. A reliable ensemble is also an ensemble that includes the verification within its forecast distribution. For this to happen, the average spread of the ensemble measured by the standard deviation has to match, on average, the error of the ensemble-mean (*Palmer et al* 2006).

The similarity between the ensemble spread and the ensemble-mean error is the first metric that we are going to analyse. The second metric is the root-mean-square-error (RMSE) of the ensemble-mean, the first-order moment of each ensemble probability distribution function. The third metric is the continuous ranked probability score (CRPS), the equivalent of the RMSE for probabilistic forecasts. It measures the average distance between the forecast probability density function and the observed density function, which is a delta function if observation errors are not taken into account (our case), or a very narrow distribution if they are taken into account. The CRPS is zero for a perfect forecast.

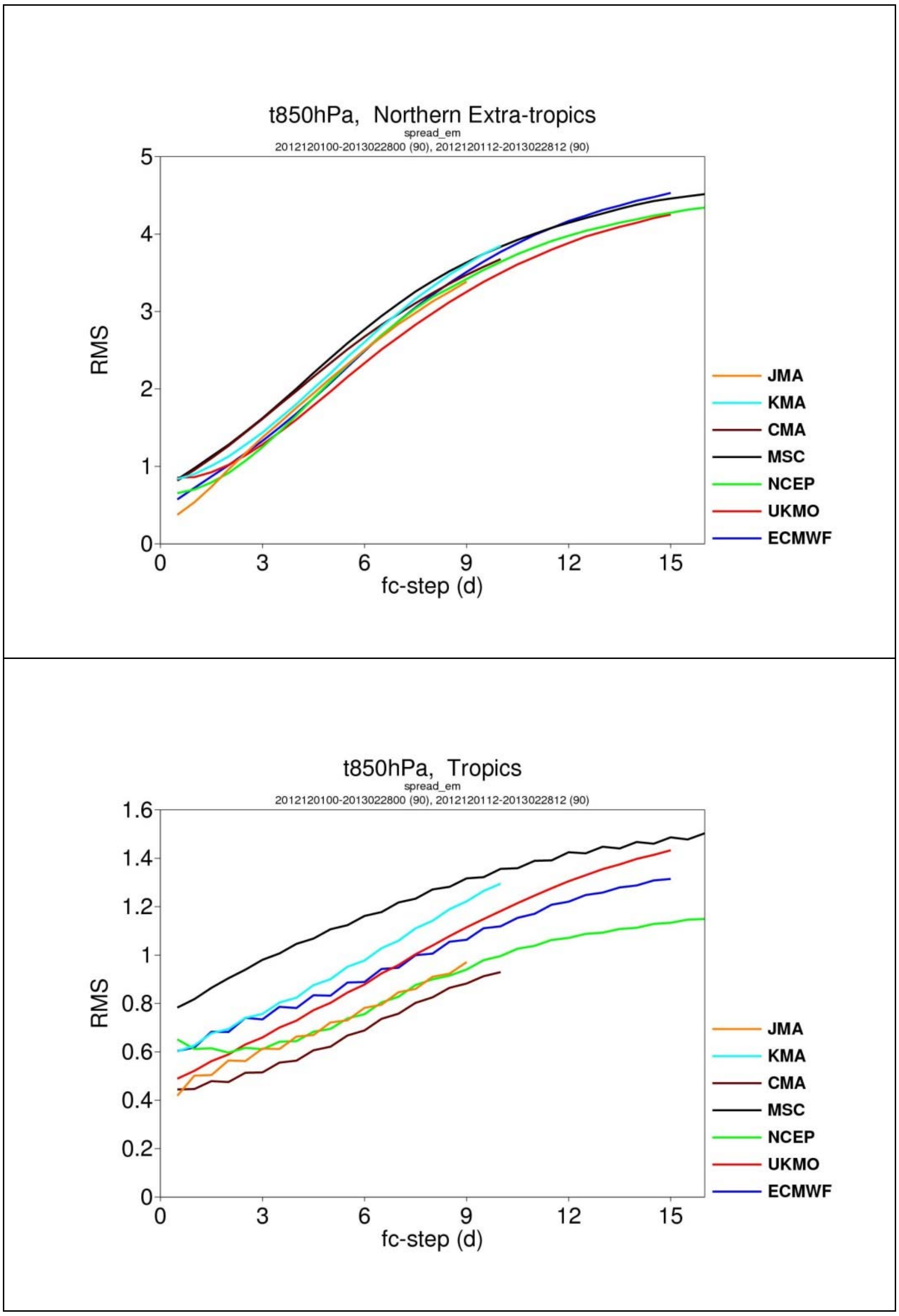

Figure 10. Winter 2012-13 average ensemble spread over the Northern Hemisphere (20°N-90°N, top panel) and the tropics (30°S-30°N, bottom panel) of 8 TIGGE ensembles (the BMRC OG-ENS stopped production in July 2010), for temperature at 850 hPa. The ensemble spread has been defined by the ensemble standard deviation.

Figure 11 shows the RMSE of the eight TIGGE ensemble-mean forecasts (dashed lines) and the ensemble spread (solid lines, also shown in Fig. 10). Over the NH, the plot shows firstly that the ECMWF OG-ENS ensemble-mean forecast (blue dashed line) has the lowest RMSE and the CMA OG-ENS has the largest. Secondly, it shows that for the ECMWF OG-ENS the average spread and ensemble mean error curves (solid and dashed blue curves) are the closest. They are followed by the MSC (black lines), NCEP (green lines) and JMA (orange) OG-ENSs. Over the TR, the NCEP OG-ENS has the lowest RMSE. In terms of reliability, the MSC OG-ENS shows the closest match between the spread and ensemble-mean error curves, followed by the ECMWF, NCEP and UKMO OG-ENS.

Figure 12 shows the CRPS of the eight ensembles for winter 2012-13. Results show that over the NH, the ECMWF OG-ENS performs best, with the JMA and the NCEP OG-ENSs second best. Over the TR, results are similar between forecast day 2 and 6, while afterwards the NCEP OG-ENS has the lowest CRPS.

Figures 13-14 show some more recent comparison of five of the TIGGE OG-ENS which are routinely monitored and compared at ECMWF (ECMWF, JMA, MSC, NCEP and UKMO). Results refer to the most recent summers and winters for which data were available (JJA, Fig. 13, and DJF, Fig. 14), for 500 hPa geopotential height and 850 hPa temperature forecasts over the NH. Accuracy is measured by the CRPSS, which is the CRPS skill score, computed using the observed climatology as reference. Results confirm that the ECMWF OG-ENS performs best, with differences in skill in the medium-range (say at about forecast day 7) of about 1 day (in other words, the ECMWF OG-ENS t+192h forecast has the same CRPSS as the second best OG-ENS t+168h forecast).

Finally, Figs. 15-16 show the most recent available results for four of the TIGGE ensembles (ECMWF, JMA, NCEP and UKMO) for the prediction of 24-hour accumulated precipitation, verified against observations at synoptic stations, for summer 2013 and winter 2013-14. These results indicate that, in general, the skill for the prediction of precipitation at synoptic stations is lower than the skill for the prediction of the large-scale flow, as represented e.g. by the 500 hPa geopotential height and the 850 hPa temperature. While these latter forecasts are skilful (i.e. better than a probabilistic forecast based on climatology) for the whole 15-day forecast range, for precipitation they are skilful only for up to a maximum of 10 days over the Northern Hemisphere extra-tropics. Over the tropics, only in winter 2013-14 one ensemble, the ECMWF OG-ENS, provided skilful forecasts for the first 5 days.

The results shown in Figs. 11-16 confirm earlier works (see, e.g., *Hagedorn et al* 2012 and references therein, and the TIGGE publications listed on the ECMWF tigge website, tigge.ecmwf.int), that, on average, the ECMWF OG-ENS is providing the most skilful global forecasts, followed by MSC, NCEP and JMA. It is interesting to look back at the relative cost of the eight TIGGE ensembles, and compare how performance and cost rankings are linked. The ECMWF OG-ENS, the best performer, is also the most expensive one, and that there is a 70% correlation between the ranks of the eight OG-ENS relative costs and performance. This indicates that higher resolution, better accuracy and model complexity, which are essential to achieve superior performance, require computing power to deliver forecasts in a timely manner (at ECMWF, e.g., we aim to be able to complete the 51 15-day forecasts in a maximum of 2 hours using only about 30-50% of the available computing power, to have some redundancy in case of need).

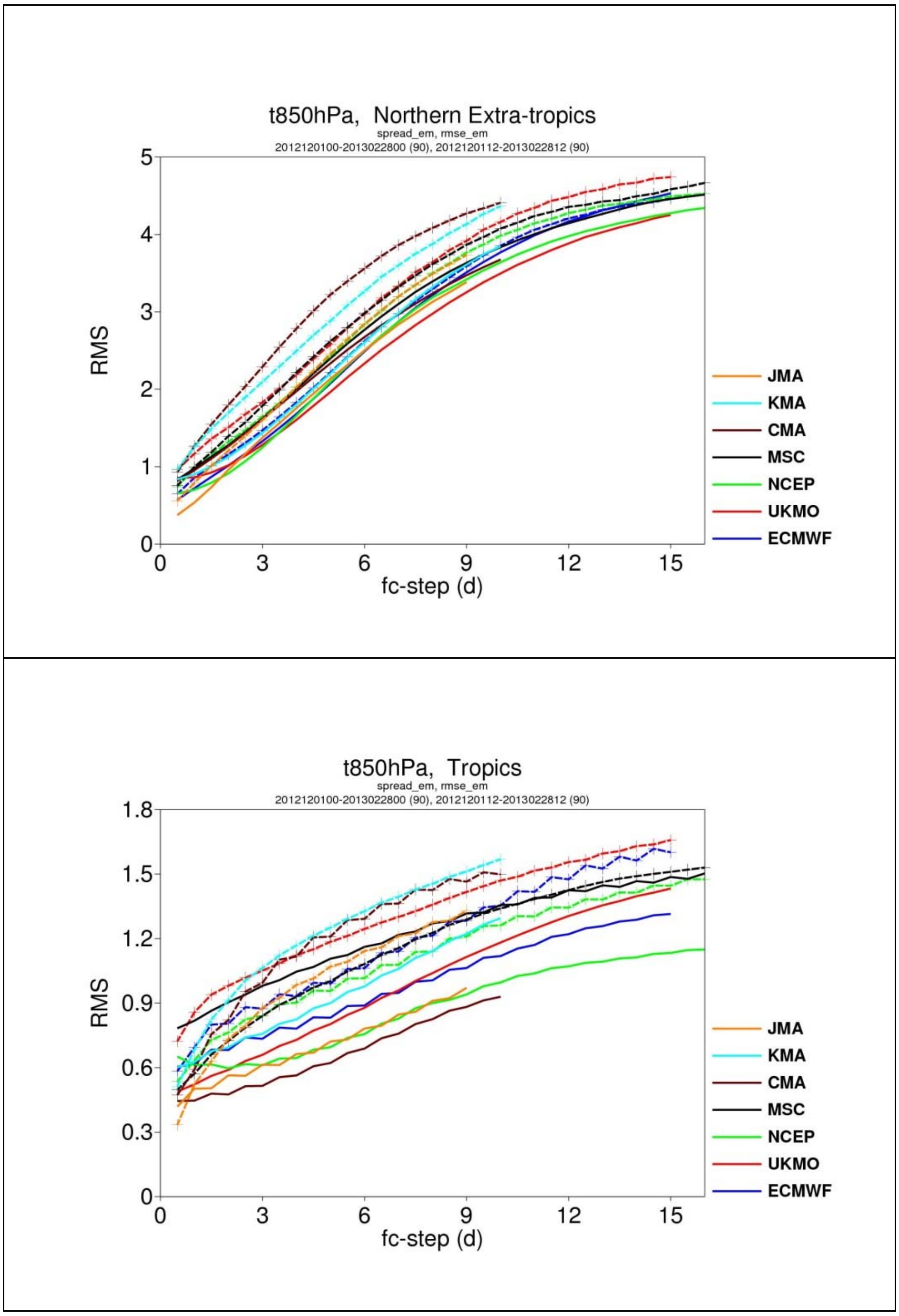

Figure 11. Winter 2012-13 average ensemble spread (solid lines) and root-mean-square-error of the ensemble-mean (dashed lines) over the Northern Hemisphere (20°N-90°N, top panel) and the tropics (30°S-30°N, bottom panel) of 8 TIGGE ensembles (the BMRC OG-ENS stopped production in July 2010), for temperature at 850 hPa. The spread lines are the same as in Fig. 10. Each ensemble has been verified against its own analysis.

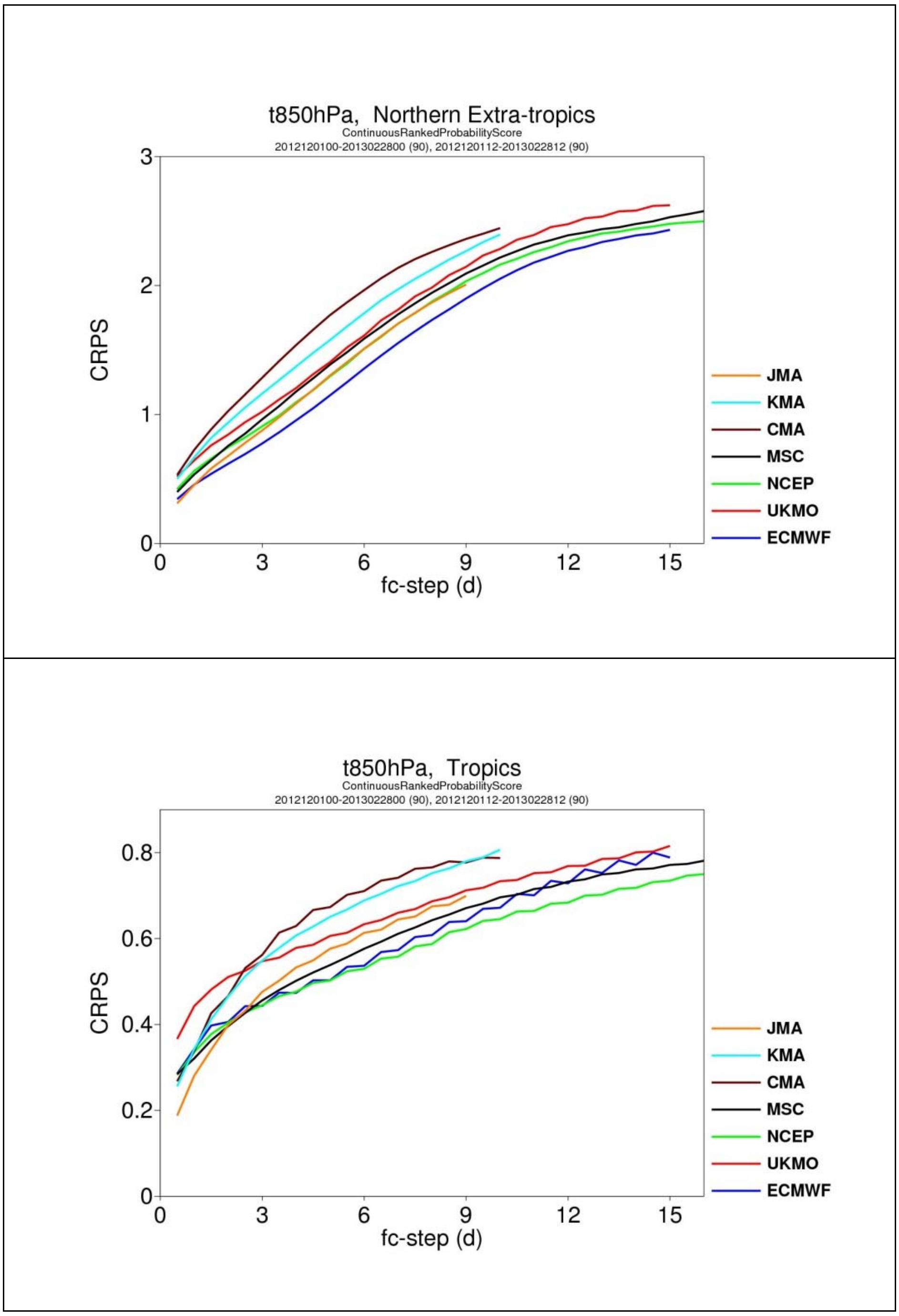

Figure 12. Winter 2012-13 average continuous ranked probability score (CRPS) over the Northern Hemisphere (20°N-90°N, top panel) and the tropics (30°S-30°N, bottom panel) of 8 TIGGE ensembles, for temperature at 850 hPa. Each ensemble has been verified against its own analysis.

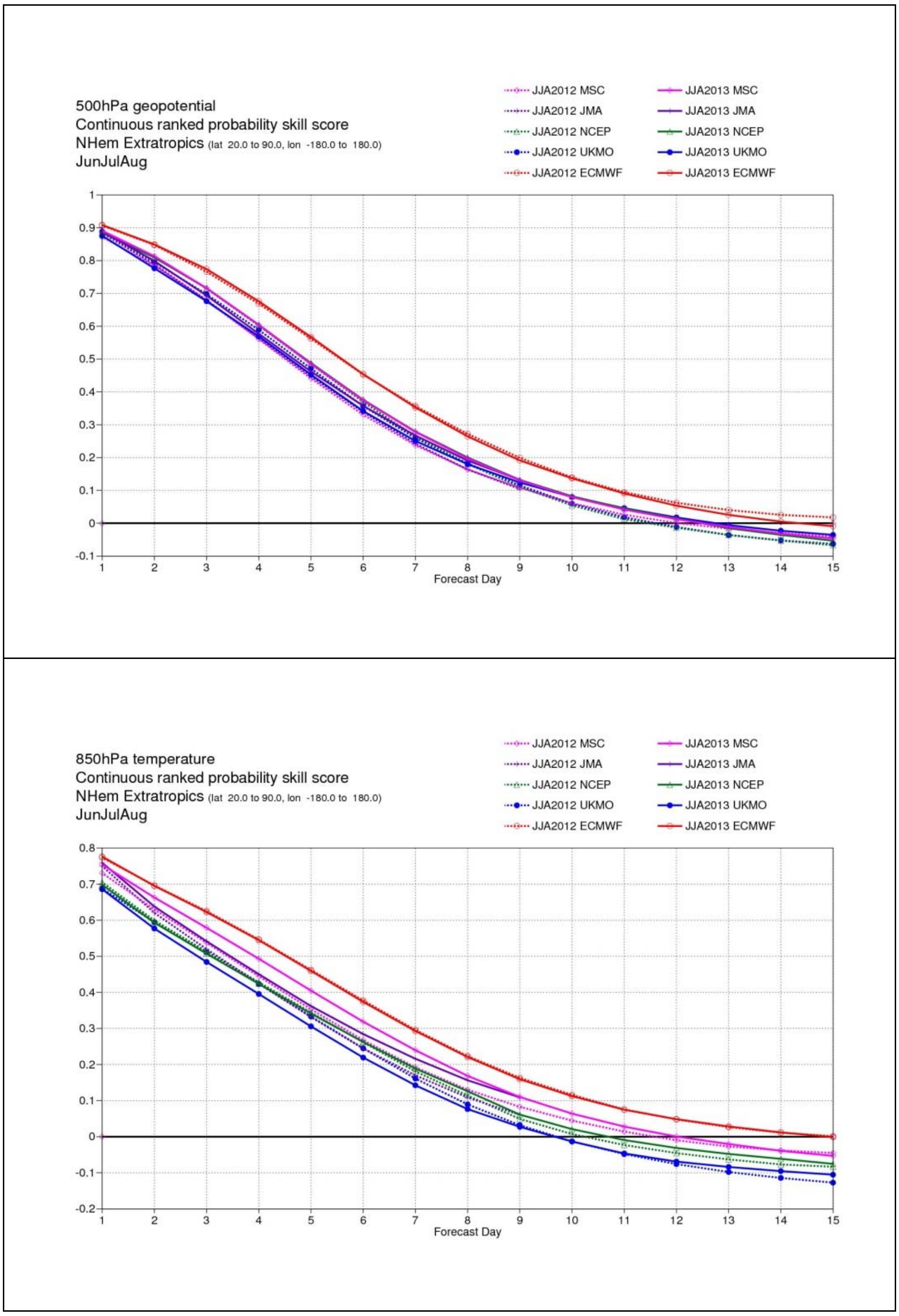

Figure 13. Summer 2013 (JJA, solid lines) and summer 2012 (JJA, dashed lines) average continuous ranked probability skill score (CRPSS) over the Northern Hemisphere (20°N-90°N) for the geopotential height at 500 hPa (top panel) and the temperature at 850 hPa (bottom panel) of five TIGGE ensembles from ECMWF, MSC, JMA, NCEP and UKMO. Each ensemble has been verified against its own analysis.

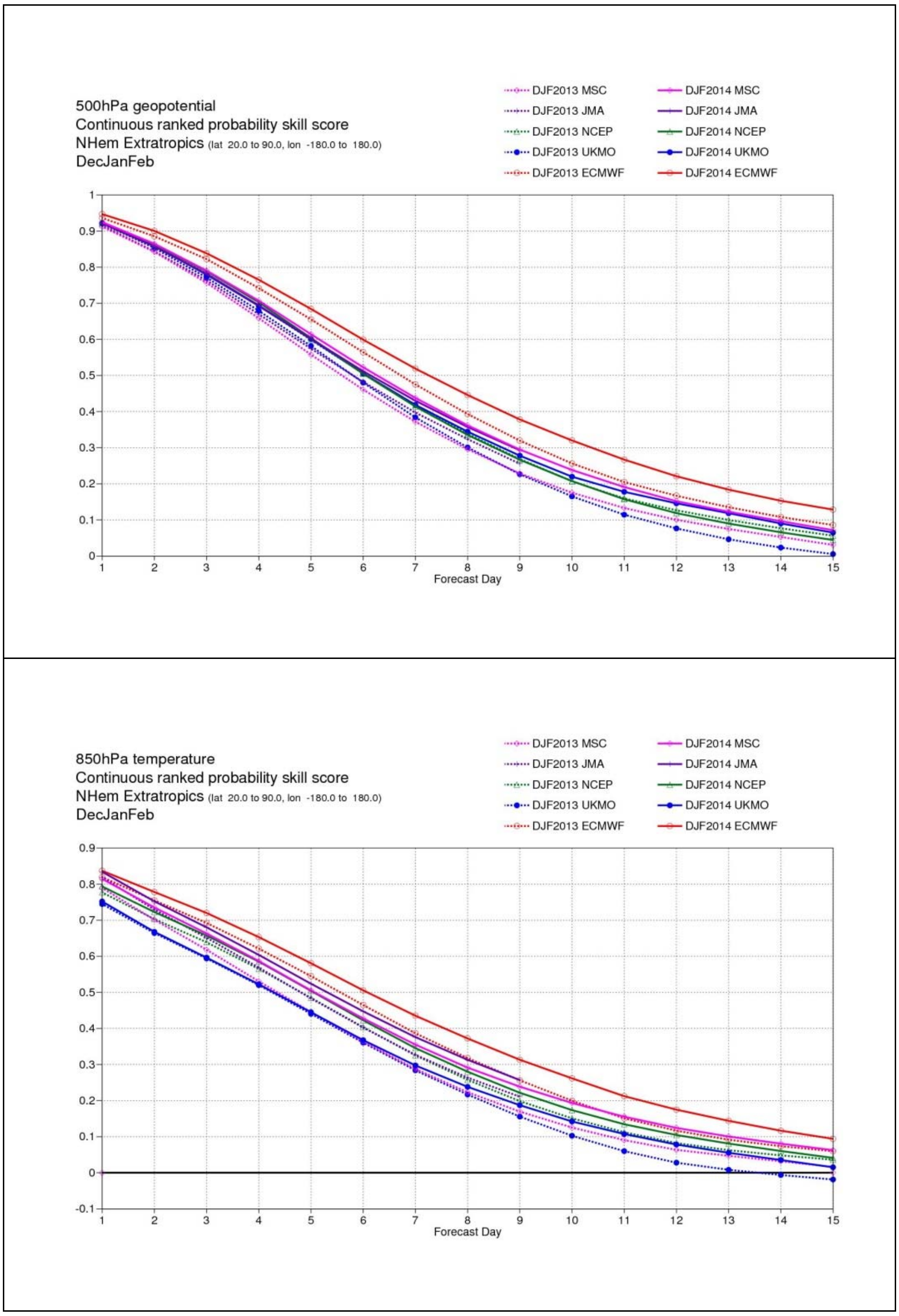

Figure 14. Winter 2013/14 (D13JF14, solid lines) and winter 2012/13 (D12JF13, dashed lines) average continuous ranked probability skill score (CRPSS) over the Northern Hemisphere (20°N-90°N) for the geopotential height at 500 hPa (top panel) and the temperature at 850 hPa (bottom panel) of five TIGGE ensembles from ECMWF, MSC, JMA, NCEP and UKMO. Each ensemble has been verified against its own analysis.

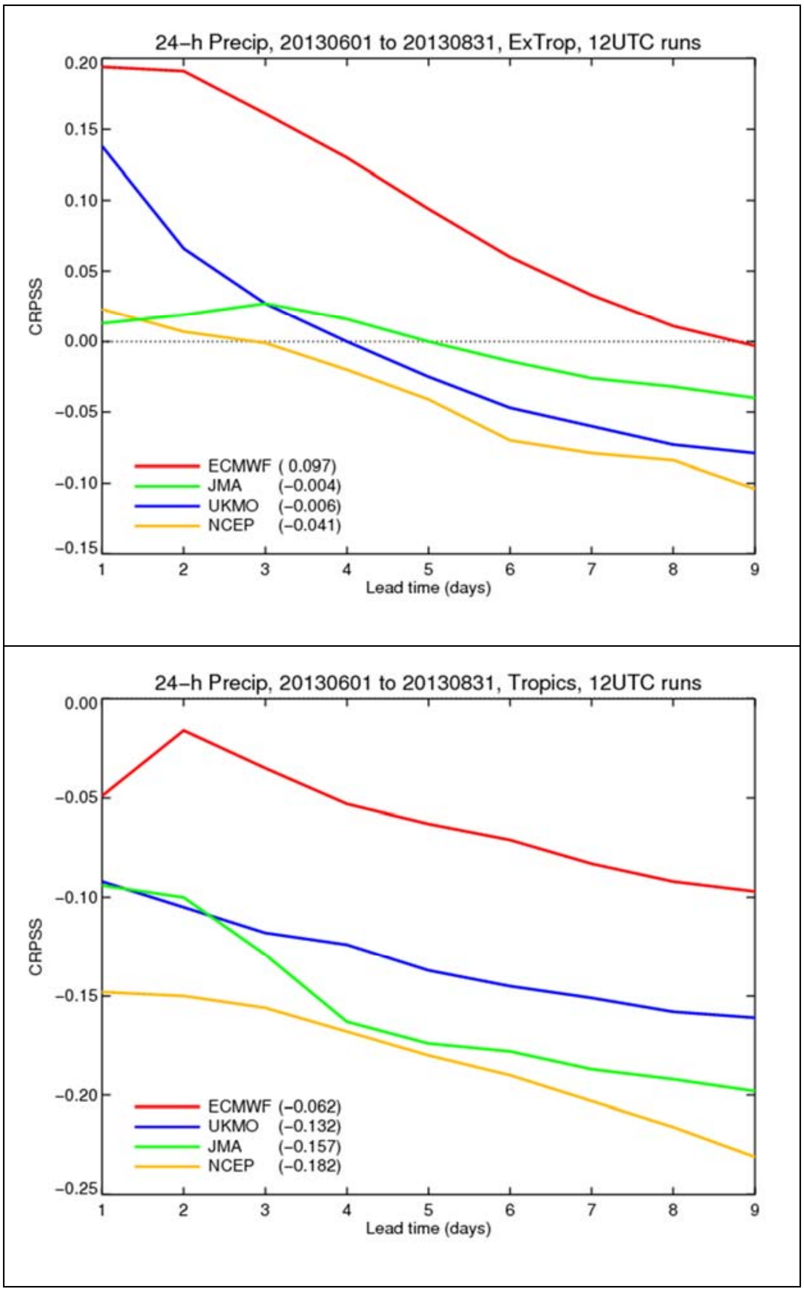

Figure 15. Summer 2013 (JJA) average continuous ranked probability skill score (CRPSS) for 24-hour accumulated precipitation over the Northern Hemisphere extra-tropics (20°N-90°N, top panel) and the tropics (30°S-30°N, bottom panel), for the four TIGGE ensembles from ECMWF, JMA, NCEP and UKMO. Verification is against observations at synoptic stations.

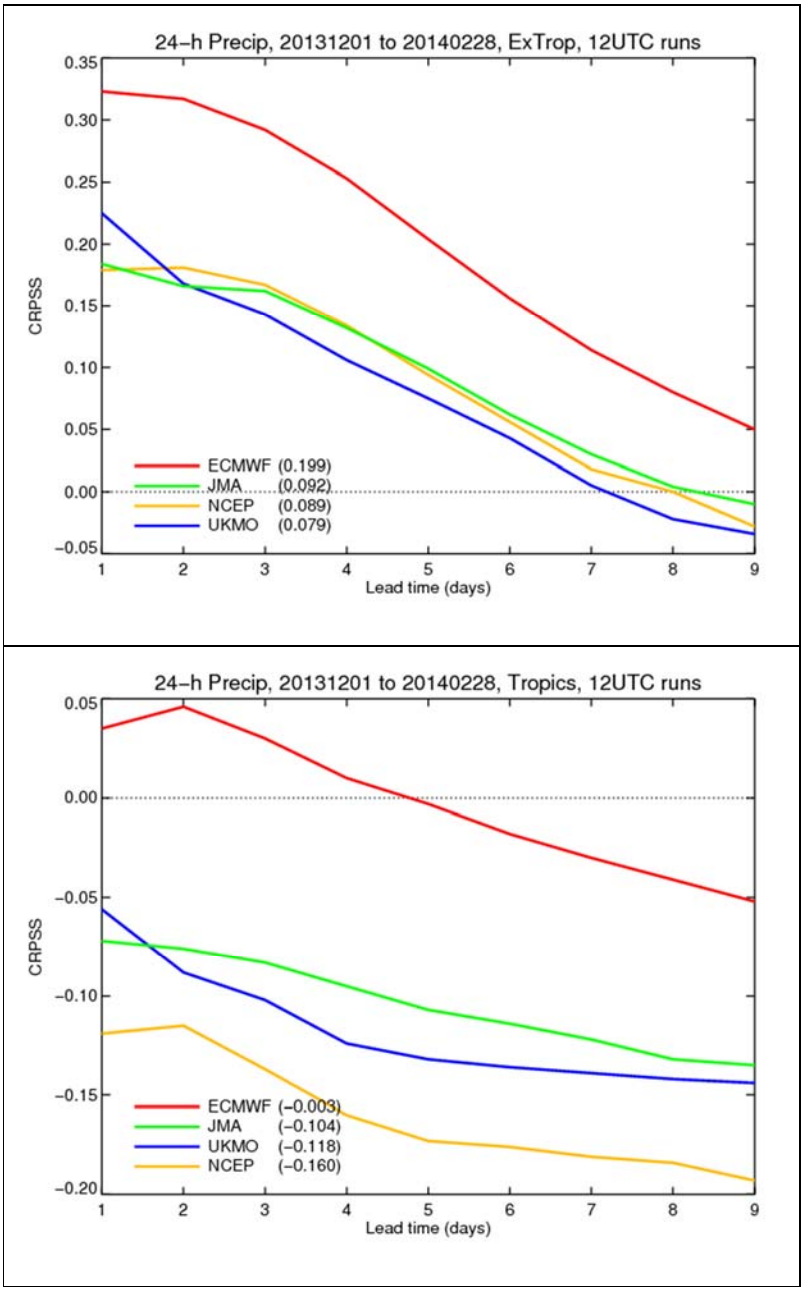

Figure 16. Winter 2013-14 (D13JF14) average continuous ranked probability skill score (CRPSS) for 24-hour accumulated precipitation over the Northern Hemisphere extra-tropics (20°N-90°N, top panel) and the tropic s (30°S-30°N, bottom panel), for the four TIGGE ensembles from ECMWF, JMA, NCEP and UKMO. Verification is against observations at synoptic stations.

4 Is there a future for ensembles? How can we further improve them?

In the last 25 years, following the pioneering work of people like *Thompson* (1957), *Epstein* (1969), *Lorenz* (1969 a, b), *Leith* (1974), we have witnessed a paradigm shift in operational numerical weather prediction from a deterministic approach, based on a single forecast, to a probabilistic one, whereby multiple ensembles are used to estimate the probability density function of initial and forecast states. 1992 saw the implementation of the first two operational ensemble systems at ECMWF in Europe, and at NCEP in the United States of America. They were followed by MSC in Canada in 1995 and by others a few years later. Building on the positive results of the global ensembles, limited-area, short-range ensembles started being developed (see e.g. the activities under the TIGGE-LAM project, http://www.smr.arpa.emr.it/tiggelam/).

Today, it is widely accepted that forecasts have to include uncertainty estimations, confidence indicators that allow forecasters to estimate how 'predictable' the situation is. These estimates can be expressed in different ways, as a range of possible scenarios, or as probabilities that events of interest can occur. Today, short and medium-range forecasts, monthly and seasonal forecasts, and even decadal forecasts and climate projections are also based on ensembles, so that not only the most likely scenario but also its uncertainty can be estimated. Furthermore, ensembles are also widely used to provide an estimate of the initial state uncertainty, to estimate more accurately the analysis error.

Thus, the answer to the first question posed in the title of this section is: ***yes, there is a future for ensembles***, at least until we do not find an alternative method to provide better estimates of the probability density function of initial and forecast states.

To address the second question, let's look at the past 20 years, in particular to the ECMWF OG-ENS, to highlight which changes and upgrades led to the continuous improvements.

Figure 17 shows how the CRPSS for the 500 hPa geopotential heights and the 850 hPa temperatures have been evolving from 1 May 1994 to date (1 May 1994 is the date when the ECMWF OG-ENS started being run every day, at 12 UTC; between 19 November 1992 and 30 April 1994 the ensemble was only run on Fridays, Saturdays and Sundays, due to lack of computing resources to produce them daily). By comparing the three lines, we can see that the CRPS of the t+7d forecast in 2014 has reached values similar to the ones of the t+3d forecast in 1994. In other words, the forecast skill has been improving by about 1.5-2.5 days per decade.

Improvements have been due to a combination of model (better physical parameterisations and numerical schemes) and data assimilation upgrades (e.g. from the optimum interpolation scheme operational in 1992 to a 3-dimensional and then 4-dimensation variational system, which now uses a 12-hour assimilation window). These upgrades, combined with the increased number and quality of satellite observations, have led to better initial conditions. Improvements have also been due to increases in horizontal and vertical resolution, made possible by the continuous increases of ECMWF computing power, and from substantial upgrades in its configuration that have made the ensemble more reliable thanks to a advances in the simulation of initial and model uncertainties. Thanks to the combination of smaller initial errors, more accurate models, and improved methods to simulate initial and model uncertainties, the skill improvements documented in Fig. 17 have been achieved.

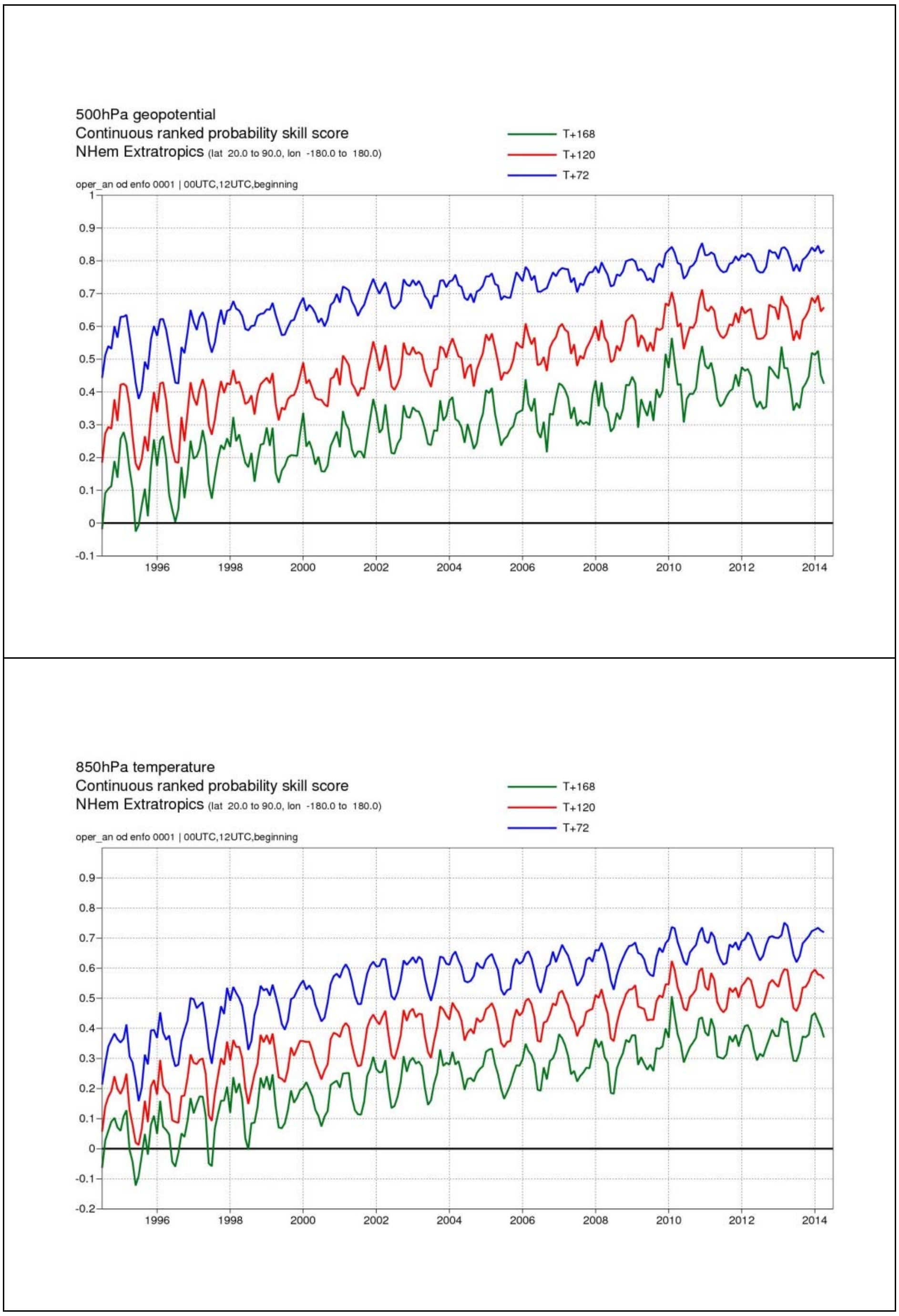

Figure 17. Time series of monthly average CRPSS for the prediction of the 500 hPa geopotential height (top panel) and the 850 hPa temperature (bottom panel) of the ECMWF OG-ENS, over the Northern Hemisphere extra-tropics (20°N-90°N), for day 3 (blue line), day 5 (red line) and day 7 (green line). Verification is against ECMWF operational analyses.

Table B lists some of the major upgrades of the ECMWF OG-ENS configuration, in resolution, membership, methodology used to simulate initial and model uncertainties, and coupling to an ocean model. Care should be taken when linking Table B and the evolution of the ensemble performance as shown in Fig. 17, since Table B does not list the whole range of very important changes and upgrades of the model physics and the data assimilation technique used at ECMWF.

Figure 17 shows that the improvements were larger in the early days, between 1994 and 2004. This is because the first versions of the ECMWF OG-ENS were rather under dispersive, and it took some years of research and development to diagnose the problems, understand how to address them and thus improve the ensemble's reliability.

Table B, e.g., lists 1998 as a year of major improvements in the simulation of initial and model uncertainties: evolved singular vectors were introduced to simulate errors that have been growing during the data assimilation time period (*Barkmeijer et al* 1999), and the first version of the ECMWF stochastic physics scheme SPPT was implemented (*Buizza et al* 1999). 1999 and 2000 saw two major resolution increases, and 2002 saw the introduction of SV-based initial perturbations in the tropics (*Barkmeijer et al* 2001). After 2004, the rate of improvement has been slightly slower, but it is still positive at about 1-1.5 days per decade (the exact value depends on the variable, the region and the forecast range considered).

Contrasting the ECMWF OG-ENS in 1992 and today gives a measure of the fundamental configuration changes implemented in the ensemble:

- In November 1992, when operational production started, the ensemble had 33 members and was run at T63L19 resolution up to 10 days, only three times a week (at 00 UTC on Fridays, Saturdays and Sundays). It had initial uncertainties simulated using initial-time singular vectors computed at T21L19 resolution over the whole globe and with a 36-hour optimisation time interval. It did not simulate model uncertainties.
- Today, the ensemble has 51 members, and is run twice a day (at 00 and 12 UTC) at T_L639L91 resolution up to 10 days and a T_L319L91 resolution from day 10 to 15. Twice a week it is extended to 32 days. Initial uncertainties are simulated using a combination of initial-time singular vectors computed at T42L91 resolution over difference regions and with a 48-hour optimisation time interval, and perturbations defined by the ECMWF ensemble of data assimilations (EDA, with 25 T_L399L137 members). The ensemble simulates model uncertainties using two schemes, the revised 3-time level stochastically perturbed parameterized tendencies (revSTP) and the backscatter (BS) schemes. The forecasts are also run coupled to an ocean model (NEMO) from initial time. Initial uncertainties in the ocean are simulated by using initial conditions from the 5-member ocean real-time analysis scheme (ORAS4, *Mogensen et al* 2012).

Some of the ECMWF OG-ENS characteristics have only slightly changed since its inception, e.g. the use of the singular vectors to simulate initial uncertainties, and the fact that they are still computed with a rather coarse resolution (T42) when compared to the resolution of the forecast model (T_L639). T42 is still the best cost/effective resolution for the SV computation: experiments performed in the past decade when the forecast resolution was increased, had suggested that increasing the SV resolution would lead to substantially higher computational costs without delivering any improvement. Other aspects have changed substantially: for example, a major change in the simulation of initial uncertainties was the replacement of the evolved singular vectors with EDA-based perturbations to simulate initial uncertainties that have been growing during the past 12 hours, during the data-assimilation period.

Another major change was been the introduction of the simulation of model uncertainties, in October 1998, when the first version of the SPPT scheme (*Buizza et al* 1999) was implemented. This was the first time that numerical weather prediction recognized and accepted the idea that stochastic (i.e. non 'deterministic') terms, when added to the equations, can improve the probabilistic forecasts. Another major advance was the introduction of the coupling to the ocean model in 1998, when the medium-range and the monthly ensembles were merged (*Vitart et al* 2008). This recognized the role of slowly-evolving, ocean-driven phenomena on medium-range and monthly atmospheric forecasts.

Between 1992 and 2014 the number of members has increased only slightly, from 33 to 51. By contrast, horizontal resolution has increased by about a factor of 10, from T63 to T_L639, with another major upgrade being planned for 2015, when the forecasts are expected to move from a $T_L639L91$ (about 32 km) to a T_L1023 (about 20 km) horizontal resolution.

	Description	Initial uncertainty estimation (SVs and EDA)							Model uncertainty estimation and forecast configuration						
		HRES	VRES	OTI	Area	past	future	sampl	HRES	VRES	Tend	#	Mod Unc	Coupling	refc suite
Nov 1992	Oper Impl	T21	L19	36h	globe	NO	SVINI	simm	T63	L19	10d	33	NO	NO	NO
Feb 1993	SV LPO	"	"	"	NHx	"	"	"	"	"	"	"	"	"	"
Aug 1994	SV OTI	"	"	48h	"	"	"	"	"	"	"	"	"	"	"
Mar 1995	SV hor resol	T42	"	"	"	"	"	"	"	"	"	"	"	"	"
Mar 1996	NH+SH SV	"	"	"	(NH+SH)x	"	"	"	"	"	"	"	"	"	"
Dec 1996	resol/mem	"	L31	"	"	"	"	"	TL159	L31	"	51	"	"	"
Mar 1998	EVO SV	"	"	"	"	SVEVO	"	"	"	"	"	"	"	"	"
Oct 1998	Stoch sch SPPT	"	"	"	"	"	"	"	"	"	"	"	STP	"	"
Oct 1999	vert resol	"	L40	"	"	"	"	"	"	L40	"	"	"	"	"
Nov 2000	FC hor resol	"	"	"	"	"	"	"	TL255	"	"	"	"	"	"
Jan 2002	TC SVs	"	"	"	(NH+SH)x+TC	"	"	"	"	"	"	"	"	"	"
Sep 2004	sampling	"	"	"	"	"	"	Gauss	"	"	"	"	"	"	"
Jun 2005	rev sampl	"	"	"	"	"	"	"	"	"	"	"	"	"	"
Feb 2006	resolution	"	L62	"	"	"	"	"	TL399	L62	"	"	"	"	"
Sep 2006	VAREPS	"	"	"	"	"	"	"	TL399(0-10) / TL255(10-15)	"	15d	"	"	"	"
Mar 2008	VAREPS-mon	"	"	"	"	"	"	"	"	"	15d/32d	"	"	HOPE from d10	5*18y
Sep 2009	Rev SPPT	"	"	"	"	"	"	"	"	"	"	"	"	"	"
Jan 2010	hor resol	"	"	"	"	"	"	"	TL639(0-10) / TL319(10-15)	"	"	"	revSTP	"	"
Jun 2010	EDA EPS	"	"	"	"	EDA	"	"	"	"	"	"	"	"	"
Nov 2010	Rev Stoch scheme	"	"	"	"	"	"	"	TL639(0-10) / TL319(10-15)	"	15d/32d	"	revSTP+BS	"	"
Nov 2011	New ocean model	"	"	"	"	"	"	"	"	"	"	"	"	NEMO from d10	"
Jun 2012	Rev EDA-pert & refc suite	"	"	"	"	"	"	"	"	"	"	"	"	"	5*20y
Nov 2013	vert resol & coupling from d0	T42	L91	48h	(NH+SH)x+TC	EDA	SVINI	Gauss	TL639(0-10) / TL319(10-15)	L91	15d/32d	51	revSTP+BS	NEMO from d0	5*20y

Table B. Time evolution of the ECMWF OG-ENS configuration between inception (Nov 1992) and today (summer 2014). Columns 3-9 refer to the key characteristics of the initial perturbations added to simulate initial uncertainties: horizontal resolution (HRES) and number of vertical levels (VRES), optimisation time interval used to compute the singular vectors (OTI), optimisation area (Area), method used to simulate uncertainties that have been growing during the data-assimilation period (past) and the future (future), and sampling technique used to define the initial perturbations. Columns 10-16 refer to the key characteristics of the forecasts: horizontal resolution (HRES) and number of vertical levels (VRES), forecast length (Tend, in days), number of members (#), simulation *of model uncertainties (Mod Unc), coupling to an ocean model (Coupling) and size of the re-forecast suite in terms of members and years (refc suite).*

Since March 2008, when the medium-range and the monthly ensembles have been merged, focus has been extended from the 3-to-10 day to the 3-to-30 day range, into the sub-seasonal range. The second half of the sub-seasonal range has seen major improvements in the past decade, thanks, as already mentioned, to a combination of model improvements, the introduction of coupled ocean-land-atmosphere models, advances in data assimilation that led to better initial conditions for the atmosphere and the ocean, and the development of better schemes to simulate model uncertainties. *Vitart* (2013) indicated that the skill of the ECMWF re-forecasts to predict the Madden Julian Oscillation has improved significantly since 2002, with an average gain of about 1 day of predictability per year. *Vitart et al* (2014) showed that the ability of the ECMWF model to simulate realistic MJO teleconnections has also improved dramatically over the past 10 years, with, for example, gains in predictability of about 1 week in the prediction of weekly average 2-metre temperature anomalies over the Northern Hemisphere extra-tropics.

Let's now go back to the questions of whether there is a future for ensembles and how we can keep improving them. This brief analysis of the evolution of the ECMWF OG-ENS configuration, and the results discussed in this paper and published in the literature (see e.g. the references listed in this paper), indicates the following:

a) ensembles provide more complete estimates of the state of the earth-system and of its future states (forecasts), and their use will continue to increase, until a different, more effective method to estimate the initial-time and forecast probabilities is developed;
b) skilful ensembles are based on an accurate model that simulates all relevant processes, and start from accurate initial conditions: by further improving models and data assimilation algorithms we can increase the skill of probabilistic forecasts;
c) ensembles need to account properly for all sources of forecast uncertainties to provide skilful and reliable forecasts: thus, to further increase their skill it is necessary to continue to improve the simulation of initial and model uncertainties;
d) ensemble re-forecasts (i.e. forecasts of past cases spanning many years generated using the operational system) are increasingly used to estimate the ensemble's model climatological distribution and assess the ensemble's predictability (*Gneiting et al* 2005; *Hamill & Whitaker* 2007); they will become an essential and integral part of future configurations.

Thus, there is a future for ensembles, and improvements will come from the adoption of more complex and fully coupled models, capable to simulate all relevant processes of the ocean, cryosphere, land and atmosphere system. These coupled (eventually earth-system) ensembles will start from fully coupled initial conditions, and will simulate in a consistent and coherent way initial and model uncertainties of all model components (ocean, land, cryosphere, atmosphere, ..). Re-forecast suites will become an integral part of ensemble forecasting, to allow the estimation of the model climatological distribution and the generation of calibrated products.

Grand, multi-system ensembles could become more common, especially for the extended forecast range (say for the seasonal to inter-annual, and possibly longer, range). This is the range where results so far have indicated that blending forecasts generated using different systems could help filtering out the unpredictable signals, and compensate for single-system systematic biases. This has indeed being happening for medium-range forecasting in the US/Canada/Mexico with NAEFS, and in Europe/US for seasonal forecasting with EUROSIP (the EUROpean Seasonal to Interannual Prediction system), which is based on the four seasonal ensembles generated by ECMWF, Meteo France, NCEP and UK Met Office.

Multi-system forecast products could also be one of the main deliverables of the WMO sub-seasonal-to-Seasonal (S2S) project, one of the proposed WWRP/THORPEX and WCRP joint research project which aims to improve forecast skill and understanding on the sub-seasonal to seasonal timescale.

5 Final remarks on operational global medium-range ensembles

This review paper should have given the reader a comprehensive overview of the TIGGE global, medium-range ensembles at the time of writing (December 2014). This overview and the analysis of the TIGGE ensembles' performance also allow us to draw four main conclusions.

The discussion has clearly shown that although all TIGGE ensembles have been designed with the same goal to estimate the probability density function of forecast states, different techniques have been used to simulate initial and model uncertainties. Thus, the *first conclusion* that we can draw is that there is not a *"unique recipe"* to generate reliable and skilful ensembles.

Considering the performance of the TIGGE ensembles, the *second conclusion* that we can draw is that the system design affects performance, and that the ECMWF ensemble still outperforms all the others, with differences in predictive skill for synoptic-scale variables in the medium-range of about 1 day, which corresponds to about 7-8 years of development.

The TIGGE project was established mainly as a resource for research and development in predictability and ensemble forecasting, and not for operational production, thus aspects linked to this latter challenge has not been discussed. The reader interested in looking into them, could look at two operational ventures established in the U.S.: the North American Ensemble Forecasting System (NAEFS) and the NCEP-FNMOC (Fleet Numerical Meteorology and Oceanography Center) global multi-centre ensemble system dedicated to probabilistic forecasts of wind-wave heights (*Alves et al* 2013).

In the past two years, two projects have started that are going to complement TIGGE: TIGGE-LAM, with focus on the short forecast range, and S2S, with focus on the sub-seasonal forecast range.

TIGGE-LAM (http://www.smr.arpa.emr.it/tiggelam/) is an extension of TIGGE for the limited area modelling (LAM) community. The TIGGE-LAM forecasts, produced on grids between 12 and 2 km resolution, provide detailed information for the short range, up to a few days ahead. TIGGE-LAM will enable users to compare models and improve the methodologies for the generation and application of regional ensemble forecasts. It will also provide valuable feedback to global ensemble developments as the resolution of these systems is planned to increase significantly in the coming years.

S2S (http://www.wmo.int/pages/prog/arep/wwrp/new/S2S_project_main_page.html; *Vitart et al* 2012) is a WMO project that aims to improve our understanding of the predictability on the sub-seasonal to seasonal timescale, improve forecast skill over this time range and promote its uptake by operational centres and exploitation by the applications community. To achieve its main goal, following TIGGE's example, it is establishing an extensive database containing sub-seasonal (up to 60 days) forecasts and reforecasts (sometimes known as hindcasts from up to 11 operational centres).

The *third conclusion* that we would like to draw is that TIGGE, TIGGE-LAM, S2S and similar projects, are essential and fundamental resources that can help understanding how best to design these ensembles, so that we can extract predictable signals from our forecasts.

Finally, considering the future of ensembles, the *fourth conclusion* that we would draw is that ensembles are here stay, and in the future they will be used even more both at initial time and at all forecast ranges, to help us take into account initial and model uncertainties in our predictions.

Acknowledgements

TIGGE has been successful thanks to the commitment of many scientific and technical experts who, worldwide, at the operational production and archiving centres, worked hard to build the extremely valuable TIGGE archive, and shared ideas, knowledge and information to advance our understanding of the predictability of the atmospheric system. TIGGE would not have been possible without them.

Considering the ECMWF medium-range/monthly ensemble (ENS), ENS has been the result of the continuous work of ECMWF staff, consultants and visitors who have continuously improved the ECMWF model, analysis, diagnostic and technical systems, and of very successful collaborations with its member states and other international institutions. The work of everyone is acknowledged.

Finally, Jean Bidlot, Erland Källén and Simon Lang are thanked for their valuable comments to an earlier version of this manuscript.

References

Alves, J.-H. G. M., P. Wittmann, M. Sestak, J. Schauer, S. Stripling, N. B. Bernier, J. McLean, Y. Chao, A. Chawla, H. Tolman, G. Nelson, and S. Klotz, 2013: The NCEP–FNMOC Combined Wave Ensemble Product: Expanding Benefits of Interagency Probabilistic Forecasts to the Oceanic Environment. *Bull. Amer. Meteor. Soc.*, **94**, 1893–1905.

Balsamo, G., A. Agustì-Panareda, C. Albergel, A. Beljaars, S. Boussetta, E. Dutra, T. Komori, S. Lang, J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay, I. Sandu, N. Wedi, A. Weisheimer, F. Wetterhall, E. Zsoter, 2014: Representing the Earth surfaces in the Integrated Forecasting System: Recent advances and future challenges. *ECMWF Research Department Technical Memorandum n. 729*, pp. 50 (available from ECMWF, Shinfield Park, Reading RG2-9AX, UK).

Barkmeijer, J., Buizza, R., & Palmer, T. N., 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2333-2351.

Barkmeijer, J., Buizza, R., Palmer, T. N., Puri, K., & Mahfouf, J.-F., 2001: Tropical singular vectors computed with linearized diabatic physics. *Q. J. R. Meteorol. Soc*., **127**, 685-708.

Berner J., G. Shutts, M. Leutbecher, and T.N. Palmer, 2008: "A Spectral Stochastic Kinetic Energy Backscatter Scheme and its Impact on Flow-dependent Predictability in the ECMWF Ensemble Prediction System". *J. Atmos. Sci.*, **66**, 603-626.

Bishop, C. H., Etherton, B. J. & Majumdar, S. J., 2001: Adaptive sampling with the ensemble transform kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.*, **129**, pp. 420-436.

Borges, M., and Hartmann, D. L., 1992. Barotropic instability and optimal perturbations of observed non-zonal flows. *J. Atmos. Sci.*, **49**, 335-354.

Bougeault, P., Z. Toth, C. Bishop, B. Brown, D. Burridge, D. Chen, E. Ebert, M. Fuentes, T. Hamill, K. Mylne, J. Nicolau, T. Paccagnella, Y.-Y. Park, D. Parsons, B. Raoult, D. Schuster, P. Silva Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson and S. Worley, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Met. Soc.*, **91**, 1059–1072.

Bourke, W., T. Hart, P. Steinle, R. Seaman, G. Embery, M. Naughton, & L. Rikus, 1995: Evolution of the Bureau of Meteorology's Global Assimilation and Prediction system. Part 2: resolution enhancements and case studies. *Aust. Met. Mag.*, **44,** 19-40**.**

Bourke, W., Buizza, R., & Naughton, M., 2004: Performance of the ECMWF and the BoM Ensemble Systems in the Southern Hemisphere. *Mon. Wea. Rev.* **132**, 2338-2357.

Bowler, N E, Arribas, A, Mylne, K R, & Robertson, K B, 2007: Numerical weather prediction: the MOGREPS short-range ensemble prediction system. Part I: system description. UK Met. Office NWP Technical Report No. 497, pp. 18.

Bowler, N E, Arribas, A, Mylne, K R, Robertson, K B, & Shutts, G J, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteo. Soc.*, **134**, 703-722.

Bright, D.R. & Mullen, S.L., 2002: Short-range ensemble forecasts of precipitation during the southwest monsoon. *Weather and Forecasting*, **17**, 1080-1100.

Buizza, R., Tribbia, J., Molteni, F., & Palmer, T. N., 1993: Computation of optimal unstable structures for a numerical weather prediction model. Tellus, 45A, 388-407.

Buizza R., & T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Buizza R., & Palmer, T. N., 1998: Impact of ensemble size on the performance of an ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, **124**, 1935-1960.

Buizza R., M. Miller, & T. N. Palmer. 1999: Stochastic representation of model uncertainties in the ECMWF EPS. *Q. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Buizza R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, & Y. Zhu. 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.,* **133**, 1076-1097.

Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., & Vitart, F., 2007: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Q. J. Roy. Meteorol. Soc.*, 133, 681-695.

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

Candille, G., 2009: The multi-ensemble approach: The NAEFS example. *Mon. Wea. Rev.*, *137*, 1655–1665.

Coutinho, M. M., 1999: Ensemble prediction using principal-component-based perturbations. Thesis in Meteorology, National Institute for Space Research (INPE), pp. 136 (in Portuguese).

Coutinho, M. M., Hoskins, B. J., & Buizza, R., 2004: The influence of physical processes on extratropical singular vectors. *J. Atmos. Sci.*, **61**, 195-209.

Ehrendorfer, M., & J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286-313.

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

Farrell, B. F., 1982. The initial growth of disturbances in a baroclinic flow. *J. Atmos. Sci.*, **39**, 8, 1663-1686.

Farrell, B. F., 1988. Optimal excitation of neutral rossby waves. *J. Atmos. Sci.*, **45** (2), 163-172.

Farrell, B. F., 1989. Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46** (9), 1193-1206.

Gneiting, T., Raftery, A. E., Westweld III, A. H., & Goldman, T., 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*., **133**, 1098-1118.

Goo, T.-Y., S.-O. Moon, J.-Y. Cho, H.-B. Cheong, & W.-J. Lee, 2003: Preliminary results of medium-range ensemble prediction at KMA: Implementation and performance evaluation as of 2001. *Korean J. Atmos. Sci*., **6**, 27-36.

Gray, M.E.B. & Shutts, G.J., 2002: A stochastic scheme for representing convectively generated vorticity sources in general circulation models. APR Turbulence and Diffusion Note No. 285, Met Office, UK.

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

Hamill, T. M., & Whitaker, J. S., 2007: Ensemble calibration of 500 hPa geopotential height and 850 hPa and 2-meter temperatures using re-forecasts. *Mon. Wea. Rev.*, **135**, 3273-3280.

Houtekamer, P L & L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev*, **125**, 2416-2426.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie & H. L. Mitchell, 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Houtekamer, P. L., Mitchell, H. L., & Deng, X., 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2126-2143.

Houtekamer, P. L., Deng, X., Mitchell, H. L., Baek, S.-J., & Gagnon, N., 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1143–1162.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher & L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. *ECMWF Research Department Technical Memorandum n. 636*. Available from ECMWF, Shinfield Park, Reading RG2-9AX (see also http://old.ecmwf.int/publications/).

Janssen, P., J.-R. Bidlot, S. Abdalla and H. Hersbach, 2005: Progress in ocean wave forecasting at ECMWF. *ECMWF Research Department Technical Memorandum n. 478.* Available from ECMWF, Shinfield Park, Reading RG2-9AX (see also http://old.ecmwf.int/publications/).

Janssen, P.A.E.M., O. Breivik, K. Mogensen, F. Vitart, M. Balmaseda, J.-R. Bidlot, S. Keeley, M. Leutbecher, L. Magnusson and F. Molteni, 2013: Air-sea interaction and surface waves. *ECMWF Research Department Technical Memorandum n. 712*. Available from ECMWF, Shinfield Park, Reading RG2-9AX (see also http://old.ecmwf.int/publications/).

Kai, J, & kim, H, 2014: Characteristics of initial perturbations in the ensemble prediction system of the Korea Meteorological Administration. *Wea. Forecasting*. DOI: 10.1175/WAF-D-13-00097.1, in press.

Lalaurette, F. 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Q. J. Roy. Meteorol. Soc*., **129**, 3037-3057.

Leith, C. E., 1965: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev*., 102, 409-418.

Leutbecher, M & Palmer T. N., 2008: Ensemble forecasting. *J. Comp. Phys.* **227,** 3515–3539.

Lin, J.W.B. and J.D. Neelin, 2000: Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophysical Research Letters*, **27**, 3691-3694.

Liu, Q., S. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane relocation in global ensemble forecast system. *Pre-prints, 27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., P5.13.

Lorenz, E. N., 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321-333.

Lorenz, E. N., 1969a: The predictability of a flow which possess many scales of motion. *Tellus*, **XXI**, 3, 289-307.

Lorenz., E. N., 1969b: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci*., 26, 636-646.

Mogensen, K., M. Alonso Balmaseda, A. Weaver, 2012a: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. *ECMWF Research Department Technical Memorandum n. 668*, pp 59. Available from ECMWF, Shinfield Park, Reading RG2-9AX (see also http://old.ecmwf.int/publications/).

Mogensen, K., S. Keeley and P. Towers, 2012b: Coupling of the NEMO and IFS models in a single executable. *ECMWF Research Department Technical Memorandum n. 673*, pp. 23. Available from ECMWF, Shinfield Park, Reading RG2-9AX (see also http://old.ecmwf.int/publications/).

Molteni F., R. Buizza, T. N. Palmer, & T. Petroliagis. 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Q. J. Roy. Meteor. Soc.*, **122**, 73-119.

NAEFS, the North American Ensemble Forecasting System, see NAEFS web page hosted by NCEP: http://www.emc.ncep.noaa.gov/gmb/ens/NAEFS.html, and web page hosted by MSC Canada: http://weather.gc.ca/ensemble/naefs/index_e.html.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., & Tribbia, J., 1993: Ensemble prediction. Proceedings of the *ECMWF Seminar on Validation of models over Europe: vol. 1*, ECMWF, Shinfield Park, Reading RG2-9AX, UK, 285 pp (available from ECMWF, Shinfield Park, Reading RG2-9AX, UK).

Palmer, T., Buizza, R., Hagedorn, R., Lawrence, A., Leutbecher, M., & Smith, L., 2006: Ensemble prediction: a pedagogical perspective. *ECMWF Newsletter n. 106*, available from ECMWF, Shinfield Park, Reading RG2-9AX, UK, pg 10-17 (see also the ECMWF web site: http://www.ecmwf.int/en/about/news-centre/media-resources).

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

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Research Department Technical Memorandum No. 598, pp. 42 (available from ECMWF, Shinfield Park, Reading RG2-9AX, UK).

Park, Y.-Y., Buizza, R., & Leutbecher, M., 2008: TIGGE: preliminary results on comparing and combining ensembles. *Q. J. R. Meteorol. Soc.*, **134**, 2029-2050.

Richardson, D., Buizza, R., & Hagedorn, R., 2005: First workshop of the THORPEX Interactive Grand Global Ensemble (TIGGE). WMO/TP report No. 1273, WWRP/THORPEX No. 5, available from WMO, 7bis, Av. de la Paix, Case postale No. 2300, CH-1211 Geneva 2, Switzerland, pp 39 (http://www.wmo.int/pages/prog/arep/wwrp/new/documents/TIGGEFirstWorkshopReport.pdf) .

Shutts, G. 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Q. J. Roy. Meteor. Soc.*, **131**, 3079-3100.

Snyder, A. D. S., Pu, H., and Zhu, Y., 2010: Tracking and Verification of East Atlantic Tropical Cyclone Genesis in the NCEP Global Ensemble: Case Studies during the NASA African Monsoon Multidisciplinary Analyses. *Weather and Forecasting*, **25**, 1397-1411.

Su, X., Yuan, H., Zhu, Y., Luo, Y., and Wang, Y., 2014: Evaluation of TIGGE ensemble predictions of Northern Hemisphere summer precipitation during 2008-2012. *Jou. Geo. Res., Atmos.,* under revision.

S2S, the WMO sub-seasonal prediction (S2S) research project. See the S2S web pages hosted by WMO, http://www.wmo.int/pages/prog/arep/wwrp/new/S2S_project_main_page.html.

THORPEX, THe Observing system Research and Predictability Experiment. See THORPEX web page hosted by WMO: http://www.wmo.ch/thorpex/.

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, 9, 275-295.

TIGGE. See the report of the 1st TIGGE Workshop, held at ECMWF in 2005: WMO/TD-No. 1273 WWRP-THORPEX No. 5 (available from the WMO web site at www.wmo.int/thorpex).

Toth, Z. & E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Toth, Z., & Kalnay, E., 1997: Ensemble Forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Toth, Z., Zhu, Y., & Marchok, T., 2001: The ability of ensembles to distinguish between forecasts with small and large uncertainty. *Weather and Forecasting*, **16**, 463-477.

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

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

Vitart, F., 2013: Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years. *ECMWF Research Department Technical Memorandum* No. 694, pp. 28 (available from ECMWF, Shinfield Park, Reading RG2-9AX, UK).

Zhang, Z., & Krishnamurti, T.N., 1999: A perturbation method for hurricane ensemble predictions. *Mon. Wea. Rev.*, **127**, 447-469.

Wang, X., C.H. Bishop, & S.J. Julier, 2004: Which is better, an ensemble of positive/negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590-1605.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C., & Wang, X., 2006: Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus A*, **58**, 28-44.

Wei, M., Toth, Z., Wobus, R., & Zhu, Y., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system *Tellus A*, **60**, 62-79.

Yamaguchi, M., and Majumdar, S. J., 2010: Using TIGGE Data to Diagnose Initial Perturbations and Their Growth for Tropical Cyclone Ensemble Forecasts. *Mon. Wea. Rev.*, **138**, 3634–3655.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences-An Introduction. International Geophysics Series, Vol.59, Academic Press, 467 pp.