

# Introduction to the special issue on “25 years of ensemble forecasting”

Roberto Buizza 

European Centre for Medium-Range Weather  
Forecasts (Reading UK)/Scuola Superiore  
Sant'Anna (Pisa, Italy)

## Correspondence

Roberto Buizza, ECMWF, Shinfield Park,  
Reading, RG2 9AX, UK.  
Email: roberto.buizza@ecmwf.int

Twenty-five years ago the first operational, ensemble forecasts were issued by the European Centre for Medium-Range Weather Forecasts and the National Centers for Environmental Prediction. These centres were followed in 1996 by the Meteorological Service of Canada, and in the subsequent years by many others. Operational ensemble-based, probabilistic forecasts signed a paradigm shift in weather prediction: for the first time, forecasters and users could have reliable and accurate estimates of the range of possible future scenarios, and not just a single realization of the future. Today, ensembles are used not only to provide reliable and accurate forecasts for the short and medium range, the monthly and seasonal time-scale, but also to provide estimates of the initial state of the atmosphere, and to generate future climate projections. This article provides an overview on how we developed the early ensembles, illustrates the key characteristics of the seven operational, global, medium-range ensembles, and discusses ongoing trends to further improve ensemble performance.

## KEYWORDS

ensemble prediction, numerical weather prediction, predictability, uncertainty estimation

## 1 | THE ECMWF 2017 ANNUAL SEMINAR ON “ENSEMBLE PREDICTION: PAST, PRESENT AND FUTURE”

At the end of 1992, for the first time the European Centre for Medium-Range Weather Forecasts (ECMWF: Buizza and Palmer, 1995; Molteni *et al.*, 1996) and the National Centers for Environmental Prediction (NCEP: Toth and Kalnay, 1993; Tracton and Kalnay, 1993) started issuing operational probabilistic forecasts generated using ensembles. This signed a paradigm shift in weather prediction: for the first time, forecasters and users were given access to not just a single realization of the future but also to a range of possible future scenarios.

ECMWF and NCEP were followed by the Meteorological Service of Canada (MSC) in 1995, and by many others

afterwards. Today, ensembles are used operationally to generate forecasts valid for the short (say up to 2–3 days), medium (say up to 2 weeks) and extended (sub-seasonal and seasonal) forecast ranges. They are also used to estimate analysis uncertainties. Thanks to advances in models, data assimilation schemes and the methods used to simulate initial and model uncertainties, today ensembles are widely used in weather prediction (see, e.g., Molteni *et al.*, 1996; Lalaurette, 2003; Dutra *et al.*, 2013). Ensembles are also used outside the weather sector, in hydrology (Pappenberger *et al.*, 2008; Thielen *et al.*, 2009), in the energy sector (Taylor and Buizza, 2003) and in aviation (Verlinden and Bright, 2017; Osinski and Bouttier, 2018).

Despite the complexity of developing and producing operational ensemble forecasts, it is now widely accepted that issuing a single forecast is not enough. Forecasts must include uncertainty estimations, for example in the form of confidence indicators that allow forecasters to estimate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

how “predictable” the future situations are. These estimates can be expressed in different ways, as a range of possible scenarios, as probabilities or cumulative distribution functions, as clusters highlighting very few possible alternatives or as box-and-whisker plots. Short and medium-range forecasts, monthly and seasonal forecasts, and even decadal forecasts and climate projections are today based on ensembles. Ensembles are used also to estimate the current state of the atmosphere, to provide users with an estimate of the initial-state uncertainty (the analysis error), and to provide a range of conditions that can be used to initialize ensemble forecasts.

This Editorial introduces this Special Issue on “25 years of ensemble forecasting”, which includes contributions that discuss different aspects of ensemble prediction: the design of ensembles, the initialization of ensemble forecasts and model error schemes, high-resolution and convective-permitting ensembles, error growth and predictability, verification metrics and application of ensemble forecasts.

In this article, in section 2 we discuss how ensembles have been gradually introduced in operational weather prediction suites. In section 3, we discuss the configuration of the seven global, medium-range ensembles operational at the time of writing (June 2018). In section 4, we briefly discuss how ensembles are now used also to estimate analysis uncertainties, and for the sub-seasonal and seasonal forecast range. Finally, some key areas of ongoing research aiming to improve further ensemble forecasts are discussed in section 5.

## 2 | THE MOVE TOWARDS ENSEMBLES SIGNALLED A PARADIGM SHIFT IN WEATHER PREDICTION

Since forecasters started using numerical models to predict the weather, they realized that also for the synoptic scales (and not just for local weather), there were cases when forecast errors would remain small for long forecast ranges, while in other cases even a 1-day forecast would be wrong. This operational experience was supported by scientific work that pointed out that due to the chaotic nature of the atmosphere, even very small initial errors could grow very rapidly and affect forecast quality in a very short time (Lorenz, 1969a; 1969b; 1982).

In the 1970s and 1980s, scientists and operational forecasters started investigating whether it would be possible to know in advance, at the time when a forecast is issued, whether the future situation would be easy (or, say, easier than average) to predict. In other words, they were looking for an objective method that could provide a level of forecast confidence. This confidence could be expressed in probabilistic terms, for example in terms of the probability that a specific event (e.g. rainfall more than 50 mm over 6 h) would occur. Otherwise, it could be expressed in terms of a range of weather scenarios, each with an assigned probability of occurrence. In this

way, forecasters could provide their users with the probability of occurrence of each scenario, or of any “tailored” weather condition.

An example of the thoughts of the 1980s comes from Palmer and Tibaldi (1988), who wrote that “*It is apparent, therefore, that a scheme to predict forecast skill will have substantial benefit in the medium range, and is an essential requirement for dynamical extended range forecasting.*” To achieve this, different approaches were tested, mostly based on ensembles, i.e. on mixing and combining many forecasts either started from different conditions, or generated using different models, or using a combination of the two (for some of the earlier works, see, e.g., Hollingsworth, 1980; Hoffman and Kalnay, 1983; Molteni *et al.*, 1996).

### 2.1 | Why are probabilistic forecasts generated using “ensembles” and not by other means?

Ensembles proved to be the only practical way to predict the time-evolution of the probability density function (PDF) of forecast states. Indeed, Ehrendorfer (1994a; 1994b) argued that even integrating a Liouville equation for the forecast PDF for systems with more than very few degrees of freedom might require the use of ensemble methods. This means that ensembles are the only feasible way in numerical weather prediction (NWP), given that the number of degrees of freedom of the phase space spanned by the NWP models is of order  $10^6$  to  $10^{10}$  for existing operational systems.

The dimensionality of the problem is not the only reason why approaches other than ensemble-based are not used to estimate the time evolution of the PDF of forecast states. Talking about weather prediction, Ehrendorfer (1994b) states that “... *the applicability of the Liouville equation in context more realistic than considered in this work may be subject to considerable problems,*” and that “... *an approach based on the Liouville equation is generally considered to be impracticable in the context of forecasting forecast skill.*”

### 2.2 | Is there any evidence that ensemble forecasts are more valuable than single ones?

There are at least two reasons why ensemble-based, probabilistic forecasts are more valuable than single forecasts. The first reason, as mentioned above, is that they make it possible not only to predict the most likely scenario but also to estimate the probability that an alternative event, or in general any event of interest, can occur. In other words, ensembles provide users with more complete and valuable information, and one way to measure the difference in value between single and ensemble forecasts is to use the Potential Economic Value metric (PEV: Richardson, 2000). PEV is based on a simple cost–loss model, whereby a user can decide to pay an amount (cost)  $C$  to protect against a loss  $L$ , linked to a specific weather event. Forecasts’ value can then be assessed by

considering users with different C/L ratios, and by constructing a curve that shows the savings that users can make if they used the forecasts. Clearly, PEV is a function of the reliability and accuracy of the forecasts: a poor (either unreliable or inaccurate) ensemble will not be able to outperform a good, single forecast. Richardson (2000) and Buizza (2001) showed that the ECMWF ensemble provides more valuable forecasts than the single, ECMWF high-resolution model. Since its first use, routine PEV computations at ECMWF have continuously confirmed this result.

The second reason why ensemble-based, probabilistic forecasts are more valuable than single forecasts is that an ensemble provides forecasters with more consistent (i.e. less changeable) successive forecasts. This can be assessed if one compares consecutive ensemble forecasts, issued 24 h apart and valid for the same verification time. Results indicate that the most recent ensemble jumps less, i.e. is more consistent with respect to the one issued 24 h earlier, than the corresponding single forecasts (Zsoter *et al.*, 2009).

### 3 | WHAT CHARACTERIZES AN ENSEMBLE CONFIGURATION?

Clearly, each institution that issues ensemble-based forecasts has developed its ensemble using its model and its data-assimilation system. If we consider, for example, ECMWF, the model used by its ensembles is the same model used to generate its high-resolution forecast, and the data-assimilation system used to generate the ensemble initial conditions is the same as the one used to generate the high-resolution initial conditions. In the past 25 years, both the model and the data assimilation have been evolving, their resolutions have been increasing substantially, and these changes have clearly affected the evolution of the ensemble performance.

Hereafter, we are not going to review model and data assimilation aspects, but we are going to focus on some of the key characteristics of an ensemble configuration:

- a Initial perturbation strategy;
- b Model uncertainty simulation strategy;
- c Resolution, forecast length and number of members.

With the term “initial perturbation strategy” we mean the way uncertainties linked to observation quality and coverage, and to data-assimilation assumptions, are simulated in the ensemble. With model uncertainty simulation strategy, we mean the approach followed to simulate model errors, linked e.g. to physical parametrizations, and the fact that models have a finite resolution. These two aspects, together with ensemble resolution (horizontal and vertical), the forecast length and the number of members, define the key characteristics of an ensemble system and affect its accuracy and reliability.

#### 3.1 | Initial perturbation strategy

Initial-condition uncertainties arise, for example, because observations are affected by observation errors, and do not cover the whole globe with the same quality and frequency. Furthermore, the process of estimating the initial state of the system, from which a forecast is computed, is based on some statistical assumptions and approximations.

There is not a unique perturbation strategy to simulate initial uncertainties. In the first version of the ECMWF ensemble (Molteni *et al.*, 1996), initial uncertainties were simulated using singular vectors (SVs: Buizza *et al.*, 1993; Buizza and Palmer, 1995): perturbations with the fastest growth over a finite time interval. Compared to random initial perturbations, SVs are characterized by a faster growth rate, which is very similar to the forecast error growth rate. SVs remained the only type of initial perturbation used in the ECMWF ensemble until 2008, when the ensemble of data assimilations (EDA) started being used together with SVs (Buizza *et al.*, 2008; Isaksen *et al.*, 2011). SVs are still an essential component of the ECMWF ensemble, and they keep providing dynamically relevant information about initial uncertainties that could have a strong, negative impact on forecast errors.

In contrast to the ECMWF ensemble, the first version of the NCEP ensemble used bred-vectors (BVs) to simulate initial uncertainties. The BV cycle (Toth and Kalnay, 1997) aims to emulate the data-assimilation cycle, and it is based on the notion that analyses generated by data assimilation will accumulate growing errors by the virtue of perturbation dynamics. This is because neutral or decaying errors detected by an assimilation scheme in the early part of the assimilation window will be reduced, and what remains of them will decay due to the dynamics of such perturbations by the end of the assimilation window. In contrast, even if growing errors are reduced by the assimilation system, what remains of them will amplify by the end of the assimilation window.

The ECMWF and the NCEP ensembles were followed, in 1995, by the Canadian ensemble, which was designed to simulate a wider range of error sources, linked to initial uncertainties due to observation errors and data assimilation assumptions, and was also linked to model uncertainties (Houtekamer *et al.*, 1996). Because of this, the Canadian initial perturbation strategy could take into account uncertainties linked to observations’ quality and coverage, and initial condition errors linked to model uncertainties.

In rather general terms, the initial perturbation strategies used by all the other ensembles operational today (see Table 1 and the references reported in section 3.3 for more details regarding the global medium-range ensembles) have been inspired by these three approaches, although their detailed implementation differs from them, and includes upgrades and changes. The reader is referred to the references included in section 3.3 for more details.

**TABLE 1** Key characteristics of the seven global, medium-range ensembles operational at the time of writing (June 2018), 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 members for each run (column 7), and number of runs per day (column 8). The ECMWF ensemble is also run up to 6 days at 0600 and 1800 UTC, see Table 2

Centre	Initial unc. method (area)	Model unc.	Truncation (degrees, km)	No. of Vert Lev (TOA, hPa)	Fcst length (d)	No. of Members (pert + unpert)	No. of Runs per day (UTC)
CMA	BV(globe)	No	T213 (0.56°; 70 km)	31 (10.0)	10	14 + 1	2 (0000/1200)
CPTEC	EOF(40°S:30°N)	No	T126 (0.94°, 120 km)	28 (0.1)	15	14 + 1	2 (0000/1200)
ECMWF	SV(NH, SH, TC) + EDA(globe)	YES	Tco639 (0.14°; 18 km)	91 (0.01)	0–15	50 + 1	2 (0000/1200) up to 15 days (2 per week up to 46 days)
			Tco319 (0.28°; 36 km)		15/46		
JMA	LETKF and SV(NH, TR, SH)	YES	TL479 (0.35°; 40 km)	100 (0.01)	11	25 + 1	2 (0000/1200)
KMA	ETKF(globe)	YES	N400 (0.28°; 32 km)	70 (0.1)	12	24 + 1	2 (0000/1200)
MSC	EnKF(globe)	YES	800 × 400 (0.45°, 50 km)	40 (2.0)	16/32	20 + 1	2 (0000/1200)
NCEP	ETR(globe)	YES	T574 (0.30°; 34 km)	64 (2.7)	0–8	20 + 1	4 (0000/0600/1200/1800)
			T372 (0.60°; 55 km)		8–16		

### 3.2 | Model uncertainty strategy

Model uncertainties arise because the models that we use to generate weather forecasts are imperfect, simulate only certain physical processes on a finite mesh, and do not resolve all the scales and phenomena that occur in the real world.

The Canadian ensemble implemented in 1995 was the first one to include also a simulation of model uncertainties. Following the Canadian example, the simulation of model uncertainties was introduced in the ECMWF ensemble in 1999, using a stochastic approach to simulate the effect of model errors linked to the physical parametrization schemes (Buizza *et al.*, 1999). This was the first time that a stochastic term was introduced in numerical weather prediction.

At present, four main approaches are followed in ensemble prediction to represent model uncertainties (see, e.g., Palmer *et al.* (2009) and Buizza (2014) for a review):

- A multi-model approach, where different models are used in each of the ensemble members; models can differ entirely or only in some components (e.g. in the convection scheme);
- A perturbed parameter approach, where all ensemble integrations are made with the same model but with different parameters defining the settings of the model components; one example is the Canadian ensemble (Houtekamer *et al.*, 1996);
- A perturbed-tendency approach, where stochastic schemes designed to simulate the random model error component are used to simulate the fact that tendencies are known only approximately; one example is the ECMWF Stochastically Perturbed Parametrization Tendency scheme (SPPT: Buizza *et al.*, 1999);
- A stochastic backscatter approach, where a Stochastic Kinetic Energy Backscatter scheme (SKEB: Shutts, 2005) is used to simulate processes that the model cannot resolve, e.g. the upscale energy transfer from the scales below the model resolution to the resolved scales; an example

is the SKEB scheme that was used at ECMWF in the medium-range/monthly ensemble (ENS) from November 2010 to June 2018 (the scheme was switched off in the most recent upgrade of the ECMWF ensemble on 5 June 2018 since it was not delivering any significant improvements and was increasing the computation cost by about 2.5%); SKEB is still used in the ECMWF operational seasonal ensemble SEAS5, which was implemented in November 2017 (and was not upgraded in June 2018).

The design of model uncertainty schemes to be used in ensembles is a key area of active research in many institutions (see, e.g., Raynaud *et al.*, 2012; Piccolo and Cullen, 2016; Leutbecher *et al.*, 2017; Lock, 2018).

### 3.3 | Resolution, forecast length and number of members

Resolution, both horizontal and vertical, forecast length and the number of members are three further key characteristics of an ensemble configuration. They are also key cost drivers. Theoretical work done in the 1970s and 1980s suggested that one needs at least about 10 members to be able to have a good ensemble-mean forecast (Leith, 1974), i.e. to have enough members to filter out the unpredictable scales. Follow-on works indicated that further increasing the ensemble size led to improved performance (Buizza and Palmer, 1998). Today, most of the operational ensemble forecast systems have between 20 and 50 members (Leutbecher, 2018), with ensembles used to estimate analysis uncertainties having a size going up to more than 250 members (Houtekamer, 2018).

Considering that we need to generate forecasts in a reasonable amount of time (say about 1 h) and that we have a finite amount of computing resources, compromises must be taken when an ensemble configuration is defined. Ideally, we would like to have high resolution in physical and probabilistic space, i.e. a very high spatial resolution and as many members as possible. Furthermore, we would also like

to extend the forecast length as long as skilful forecasts for at least certain spatio-temporal scales can be issued, i.e. up to the time for which forecasts outperform a climatological forecast (for a discussion of the forecast skill horizon, the reader is referred to Buizza and Leutbecher (2015)). At ECMWF, for example, to be able to serve as many users' demands as possible (in terms of forecast quality and resolution), we decided to compromise on the design of the ensembles, and use three different resolutions to generate forecasts valid for the medium-range, the monthly and the seasonal time-scales. By using a variable resolution approach (Buizza *et al.*, 2007; Vitart *et al.*, 2008) and decreasing the resolution past the medium-range, we had been able to afford to generate operationally ensemble forecasts with a longer forecast length, with a minimum impact on the ensemble performance. By further reducing the resolution, we had been able also to generate operationally seasonal forecasts (Molteni *et al.*, 2011). Results so far have suggested that this strategy, to "remove" scales in the model integration as they lose predictability (i.e. as forecast skill on these scales is lost) and as their impact on the slightly larger scales is reduced, is a very effective way to use computing resources to generate operational forecasts.

Table 1 shows the key characteristics of the seven ensemble prediction systems operational at the time of writing (June 2018) that provide global forecasts valid for the medium- and the extended-range (say for between 7 and 15 days). They are generated at the following institutions:

- 1 CMA, the China Meteorological Administration;
- 2 CPTEC, the Brazilian Center for Weather Prediction and Climate Studies (Centro de Previsão de Tempo e Estudos Climáticos);
- 3 ECMWF, the European Centre for Medium-Range Weather Forecasts;
- 4 JMA, the Japanese Meteorological Administration;
- 5 KMA, the Korean Meteorological Administration;
- 6 MSC, the Meteorological Service of Canada;
- 7 NCEP, the US National Centers for Environmental Prediction.

Together, these seven institutions produce every day more than 500 medium-range forecasts, with a horizontal resolution that ranges from about 16 to about 120 km. All simulate initial uncertainties, but only five also simulate model uncertainties (Buizza (2014) provides an overview of their key characteristics and performance). They all use different techniques to generate the initial perturbations:

- The CMA ensemble uses bred-vectors (BV), perturbations designed to emulate the analysis cycle (Toth and Kalnay, 1993; 1997);
- The CPTEC ensemble uses Empirical Orthogonal Functions (EOF; Coutinho, 1999; Zhang and Krishnamurti, 1999);

- The ECMWF ensemble uses a combination of SVs and EDA-based perturbations (Buizza and Palmer, 1995; Buizza *et al.*, 2008);
- The JMA ensemble uses a combination of SVs and perturbations generated using a Local Ensemble Transform Kalman Filter (LETKF; Miyoshi and Sato, 2007);
- The KMA ensemble uses an Ensemble Transformed Kalman Filter method with localization (Bowler *et al.*, 2008; Kai and Kim, 2014);
- The MSC ensemble uses an Ensemble Kalman Filter (EnKF; Houtekamer *et al.*, 2009; 2014);
- The NCEP ensemble uses perturbations generated with an Ensemble Transform with Rescaling (ETR; Wei *et al.*, 2006; 2008) technique.

The fact that they also use different approaches to simulate model uncertainties indicates that there is not a unique way to generate reliable and accurate probabilistic forecasts. This is not to say that ensemble performance is independent of the ensemble's design. This was, for example, highlighted by Buizza *et al.* (2005), who compared the performance of the three global ensembles operational in 2004 at ECMWF, MSC and NCEP. They concluded that performance depends not only on the quality of the model and the data-assimilation system, but also on the methodology followed to simulate the sources of forecast error. Park *et al.* (2008) also documented the large difference between the performance of the single ensembles: for the day-5 prediction of synoptic scales over the Northern Hemisphere (identified by the 500 hPa geopotential height), the difference between the best and the worst ensemble skill was about 3 days.

Forecasts from the ensembles listed above are exchanged routinely, and can be accessed in delayed mode (48 h after they have been generated) thanks to the WMO TIGGE project. TIGGE is the World Meteorological Organization THORPEX Interactive Grand Global Ensemble, which started in 2004; TIGGE data can be accessed from the ECMWF web site: <https://www.ecmwf.int/en/research/projects/tigge>. TIGGE data can be used, for example, to understand the impact of using different ensemble methodologies on the ensembles' performance (see, e.g., Park *et al.*, 2008; Hagedorn *et al.*, 2012).

#### 4 | ENSEMBLES ARE USED ALSO FOR THE SUB-SEASONAL AND SEASONAL TIME-SCALES, AND TO ESTIMATE ANALYSIS UNCERTAINTIES

In the 1990s–2000s, operational very high-resolution forecasts started including also ensemble components, run with limited-area models nested in global ensembles, and ensembles started being used also to provide extended-range (monthly and seasonal) forecasts. Following the Canadian

example, ensembles were also developed to estimate analysis uncertainties at ECMWF (Isaksen *et al.*, 2011) and Météo-France (Berre *et al.*, 2007). In this article, we are not going to discuss limited-area ensembles, because the techniques that they use to simulate initial and model uncertainties are similar to the techniques used in the medium-range ensembles, reviewed in section 3. Hereafter, we continue to discuss global ensembles for the extended range and to estimate analysis uncertainties.

#### 4.1 | Extended-range ensembles, and reforecast ensembles

Since the beginning of the 2000s, ensembles have been used to generate operational sub-seasonal (monthly: Vitart, 2004) and seasonal forecasts. Extended-range ensembles have, generally speaking, a coarser resolution than the medium-range ensembles, to limit the production costs (see, e.g., Table 2 for ECMWF). Compared to the medium-range ensembles, most of them also include a dynamical ocean model, to be able to simulate better the propagation of coupled ocean–atmosphere phenomena, like the organized convection associated with the Madden–Julian Oscillation.

Using ensembles is even more essential for the extended time range, to be able to extract reliable and accurate signals. It is worth mentioning that there is increasing evidence to suggest that for this time range the number of ensemble members should be higher than for the medium-range ensembles, say about 100–200, compared to about 25–50 for the medium-range (T. Stockdale, personal communication, 2018).

A key ingredient of extended-range ensembles is that they rely on reforecast ensembles (Hamill *et al.*, 2006; Hagedorn *et al.*, 2008), i.e. on ensemble forecasts generated with the operational ensemble configuration but for a large number of past cases. Extracting predictable signals for the extended range has benefitted from the use of ensembles of reforecasts, since they allow us to estimate in a better way the ensemble error characteristics (e.g. model biases, or whether they tend to under- or over-estimate certain phenomena: Hagedorn *et al.*, 2012). At ECMWF, for example, the reforecast suite of the medium-range/monthly time-scale covers the past 20 years (an 11-member ensemble is run twice a week, for each week of the past 20 years), and the reforecast suite of the seasonal ensemble covers the past 30 years (a 15-member ensemble is run once a month, for the past 30 years). Reforecast ensembles are essential also to be able to have a statistically significant estimate of the skill of monthly and seasonal forecasts, to understand the predictability of different phenomena and how to extract the predictable signals from the forecasts. The need of having an ensemble of reforecasts makes the cost (in terms of computer power, data handling and data storage) of the monthly and seasonal ensembles very high, and this is one of the key reasons why they are characterized by a coarser resolution (Table 2).

#### 4.2 | Ensembles of analyses

Since its inception in 1995 (Houtekamer *et al.*, 1996), the Canadian ensemble included an ensemble of analyses, generated using an ensemble Kalman filter (EnKF). The initial conditions of the ensemble forecasts were defined by one of the members of the EnKF. The EnKF has been providing MSC Canada with information about uncertainties in the analysis. ECMWF (Buizza *et al.*, 2008; Isaksen *et al.*, 2011) and Météo-France (Berre *et al.*, 2007) started producing an Ensemble of Data Assimilations in 2008.

The ECMWF EDA is based on an ensemble of  $N$  separate data assimilation procedures (where  $N$  is 25 at the time of writing, and is planned to increase to 50 in 2019), each using perturbed observations and a model uncertainty scheme. Observations are perturbed to simulate observation errors, linked to the instruments' characteristics and to their representativeness. As for the case of ensemble forecasts, model uncertainties are simulated in the ensembles of data assimilations to take into account the fact that the models used to define the analysis (i.e. the forecast initial conditions) are not perfect.

Table 3 lists the key characteristics of the EDA used at ECMWF. As mentioned above, since 2008 the ECMWF EDA is used in combination with the SVs to define the initial conditions of the medium-range/monthly ensemble. The addition of EDA-based perturbations has had a major impact on the ensemble reliability and accuracy in the short forecast range over the extratropics, and for the whole forecast range over the Tropics (Buizza *et al.*, 2008).

### 5 | A LOOK INTO THE FUTURE

Looking into the future, three main trends can be detected in the way ensembles are planned to be upgraded:

- A move towards an Earth-system approach to modelling and assimilation;
- A move towards a seamless approach in the design of the analysis, medium-range, sub-seasonal and seasonal ensembles;
- A move towards higher resolution.

The first trend towards an Earth-system approach is linked to results obtained in the past two decades that showed that by adding relevant processes we can improve the quality of the existing forecasts, and further extend the forecast skill horizon at which dynamical forecasts lose their value. Buizza and Leutbecher (2015), for example, looked at the evolution of the skill of the ECMWF ensemble from 1994 to date, and concluded that “*Forecast skill horizons beyond 2 weeks are now achievable thanks to major advances in numerical weather prediction. More specifically, they are made possible by the synergies of better and more complete models, which include more accurate simulation of relevant*

**TABLE 2** Key characteristics in June 2018 of the ECMWF ensembles of forecasts for the limited-area community (row 1), the medium-range (row 2), the sub-seasonal range (row 3) and the seasonal range (row 4), in terms of resolution (truncation and resolution in km; column 2), number of vertical levels and top of the atmosphere (column 3), forecast length (column 4), number of members (column 5), and frequency (column 6). ENS for the monthly extension starts from the ENS medium-range forecasts. SEAS5 forecasts are extended to 13 months every quarter (on the 1st of February, May, August, November). ENS reforecasts include 11 members, run twice a week, for the most recent 20 years. SEAS5 reforecasts include 25 members, run once per month for 36 years (1980–2016)

ECMWF ensembles of forecasts	Horizontal resolution: truncation (km)	No. of Vertical Levels (TOA, hPa)	Fcst length (days/months)	No. of Members (pert + unpert)	Frequency of operational production
ENS for LAM (BC-project)	Tco639 (18 km)	91 (0.01)	0–6.5 day	50 + 1	2 per day (0600/1800)
ENS for the medium-range	Tco639 (18 km)	91 (0.01)	0–15 day	50 + 1	2 per day (0000/1200)
ENS for the monthly extension	Tco319 (36 km)	91 (0.01)	15–46 day	50 + 1	2 per week (0000, Mon/Thu)
SEAS5	Tco319 (36 km)	91 (0.01)	7/13 months	50 + 1	1 per month (0000 on the 1st)

**TABLE 3** Key characteristics in June 2018 of the ECMWF Ensemble of Data Assimilations, in terms of resolution (horizontal truncation and grid resolution in km; number of vertical levels and top of the atmosphere), assimilation method (4-dimensional variational and assimilation window), number of members (including times when the ensemble of analyses is produced), and whether they include a simulation of observation and model uncertainties

	Resolution		Assimilation method (window)	Members (frequency)	Uncertainties	
	Hor. Truncation (km)	Vert. Lev (TOA)			Obs	Model
ECMWF Ensemble of Data Assimilations (EDA)	Tco639 (18 km)	137 (0.01 hPa)	4D-Var (12 h)	25 (0000, 1200 UTC)	Yes	Yes

*physical processes (e.g. the coupling to a dynamical ocean and ocean waves), improved data-assimilation methods that allowed a more accurate estimation of the initial conditions, and advances in ensemble techniques.”*

The second trend towards a seamless approach comes partly from scientific reasons and partly for technical reasons. From the scientific point of view, for example, there is evidence that processes that were thought to be relevant mainly for the extended range are also relevant for the short range. An example comes from the introduction of a dynamical ocean in the ECMWF ensembles. At ECMWF, for example, we started using a coupled ocean–land–atmosphere model for the seasonal and the monthly time-scales, and we introduced it in the medium-range ensemble only later, when we realized that it could contribute to improving its reliability and accuracy. From the technical point of view, having an integrated approach whereby the same model is used in analysis and prediction mode, from day 0 to year 1, simplifies production, maintenance and the implementation of upgrades. Furthermore, it helps the diagnostics and evaluation of a model version, since tests carried out over different time-scales can help identify undesirable behaviours that could lead to forecast errors.

The third trend towards higher resolution comes from the need to resolve better the smaller scales, and their interaction with the slightly-less-small scales, and so on. All scales are relevant in weather prediction, and errors propagate from the smallest to the larger scales. For example, at ECMWF improvements in the treatment of convection in the Tropics led to improvements in the development and propagation of organized convection (Bechtold *et al.*, 2014), which had positive impact on the skill of monthly forecasts over Europe (Vitart, 2013).

If we consider the current ensembles, we should not forget that even if they use resolutions of 18–120 km (see Table 1), they are able to resolve in a realistic way only scales that are about 5–6 times their resolution. This is because the scales closer to the model grid spacing are not simulated in an accurate way (e.g. to avoid numerical instabilities, strong diffusion operators are used, which makes the energy spectra at the finest scales far from the observed spectra). This means, for example, that frontal dynamics is still poorly resolved, and this can have strong impact on the prediction of synoptic-scale features in the medium-range, and of low-frequency variability (e.g. the North Atlantic Oscillation, or European blocking) in the monthly time-scale.

Thus, at the time of writing (June 2018) the ECMWF global ensemble has an effective resolution of about 100 km up to forecast day 15 (although its grid spacing is 18 km), and the highest-resolution limited-area ensemble run, e.g. at Météo-France, has an effective resolution of about 15 km (although its grid spacing is about 3 km). If we want to be able to predict phenomena such as intense wind gusts or extreme precipitation linked to convective events, it is essential that we aim to further increase the models’ resolution. At ECMWF, the 2016–2025 Strategy talks about aiming to increase the ensemble resolution to about 5 km by 2025, with limited-area models expected to run at least with a 1 km resolution by that time. These plans, as discussed above, should not forget the need to increase also resolution in probability space, if we want to provide users with more reliable forecasts also for very rare events, and this can only be achieved by increasing the ensemble size.

Another area where progress should be expected is in the definition of the ensemble initial conditions. Ensembles of analyses and forecasts should be linked closer together,

to improve the consistency in the simulation of initial uncertainties. In terms of modelling, physical processes that are not yet included in the models but are relevant for weather prediction should be included to make the forecasts more and more realistic. At ECMWF, for example, as part of our strategy to move towards Earth-system models and assimilation systems, we are investigating the potential role of including an interactive aerosol in our ensemble (A. Benedetti and F. Vitart, personal communication, 2018). Model error schemes should be redesigned to simulate better the model uncertainties linked to each process, and to the numerical schemes used in the numerical integrations. Coupled models, which are now used in ensemble forecasts but not yet in ensembles of data assimilations, should be used in both to provide better and more consistent initial conditions for the coupled ensembles. Similarly, coupled initial perturbations (or, even better, ensembles of coupled initial conditions) should be used to initialize the coupled ensembles.

The ECMWF 2016–2025 Strategy, adopted by the Council in December 2015 (ECMWF, 2015) indeed talks about building an Earth-system model that includes all relevant processes (“... develop an integrated global model of the Earth system to produce forecasts with increasing fidelity on time ranges up to one year ahead”). It talks also about adopting an ensemble approach for all time-scales (“Operational ensemble-based analyses and predictions that describe the range of possible scenarios and their likelihood of occurrence and that raise the international bar for quality and operational reliability”).

This last point raises the question of whether, as part of the ECMWF operational forecasts, ECMWF should continue to generate one high-resolution analysis and one high-resolution forecast. So far, experimental tests have indicated that re-centring the ECMWF ensemble on the ECMWF high-resolution analysis leads to better ensemble forecasts. Results have also indicated that adding the high-resolution forecast to the 51 ensemble members into a mixed-resolution, 52-member ensemble leads to better short-range forecasts. As part of the work that we are performing to implement the ECMWF 2016–2025 Strategy, we have started to investigate further the performance of a mixed-resolution ensemble, which could include  $N$  members at a lower resolution and  $M$  (with  $M > 1$ , but with  $M < N$  or even  $M < < N$ ) members at higher resolution (M. Leutbecher and F. Vitart, personal communication, 2018). Advances in post-processing and calibration techniques (Vannitsem *et al.*, 2018) should make it possible to merge ensembles with different resolutions.

## 6 | CONCLUSIONS

In this article, we have briefly reviewed how ensembles have become an essential component of the operation suites of most of the meteorological centres. The main characteristics of the seven medium-range global ensembles operational at the time of writing (June 2018) have been illustrated. The

establishment of ensemble methods as key tools not only for estimating forecast uncertainty at all forecast ranges but also for estimating analysis uncertainties has been discussed. Finally, some key trends that are being followed by many operational centres to improve further their performance have been highlighted.

In the articles included in this Special Issue, the reader can find more detailed discussions of some of the aspects that we discussed above: ensemble size, initial condition and model error representation, error growth and scale-dependent predictability, diagnostics of ensemble-based probabilistic forecasts and applications of ensemble forecasts.

On the issue of ensemble size, it is not clear whether the current size of about 20-to-50 is large enough to provide reliable probabilistic forecasts, especially of the tails of the distributions, which includes the events that users with low cost/loss ratios would be more interested in (Leutbecher, 2018). Results from Canada, for example, indicate that having a large enough ensemble size is very important if one wants to provide a reliable analysis error estimate. Indeed, the Canadian Ensemble Kalman Filter has 256 members, while their ensemble forecast system has only 21 (Houtekamer, 2018).

Together with ensemble size, resolution is a key cost-driver in ensemble prediction. Although some users would want to trade resolution for ensemble size, care should be taken in doing so. Resolution is key if one wants to predict phenomena characterized or influenced by small scales, such as extreme wind storms or convection events. Frogner (2018) and Weyn (2018) discuss some of the issues and challenges linked with the design of very high-resolution and convective-permitting ensembles. Designing ensemble systems capable of providing skilful information for these scales is very challenging, partly because initializing the small scales is very difficult.

Investigating the predictability of different phenomena such as the North Atlantic Oscillation (Weisheimer, 2018) and understanding forecast error growth (Buizza *et al.*, 2018), and more generally the scale dependency of the forecast skill horizon (Hoskins, 2013; Buizza and Leutbecher, 2015) are also key to help us to advance the design of future ensembles. Moving towards more complete Earth-systems can help to extract predictable signals; the ocean is an example of an Earth-system component that has been added to operational global ensembles in the past years, and interactive aerosol and the inclusion of some atmospheric gases (e.g. ozone, CO<sub>2</sub>, other greenhouse gases) might be the next components that could help further extend the forecast skill horizon. In this Special Issue, Zanna (2018) discusses the state of the art in ocean modelling and recent advances in the design of ocean ensembles, and Xian (2018) raises issues linked to the prediction of aerosol. Palmer (2018) also provides some very interesting views on how ensemble prediction will evolve in the next 25 years.

Two other key topics discussed in this Special Issue are ensemble verification, and how best to communicate uncertainty to the users. Siegert (2018) illustrates how



complex the verification of ensemble performance is. He discusses the sensitivity of one score that is used to assess the performance of ensemble systems, the ignorance score, shows that it depends on the ensemble size, and proposes to apply an ensemble-adjustment. On the second topic, Fundel (2018) discusses the difficulty to communicate probabilities to lay audiences and the user's prevalent resistance to probability forecasts, and reports the experience of the German weather service (DWD) in introducing probabilistic forecast products to three exemplary user groups.

Let me conclude by saying that, in the past 25 years, ensembles have demonstrated to be extremely valuable and essential sources of information. In weather prediction, ensemble performance has been improving by about 2 days per decade in the medium-range (Buizza and Leutbecher, 2015), and even by up to 1 week per decade for the monthly time-scale for large-scale phenomena such as the Madden–Julian Oscillation (Vitart, 2013; Vitart *et al.*, 2014). Their performance will continue to improve, provided that we can advance in the areas discussed above (modelling, including model error simulation, data assimilation and ensemble initialisation, ensemble design and membership), and their use will continue to increase, provided that we can help users to take decisions using probabilistic information.

This latter one is, in my view, another very hard challenge that we are facing. One possible way forward in this area is to interact more with experts working in other fields where ensembles are used to estimate a range of possible outcomes, such as the insurance sector and economics (see, e.g., Ravazzolo and Vahey, 2013; Sun and Jin, 2016). By working together with them and social scientists that have been studying decision theory under severe uncertainty (see, e.g., Busemeyer *et al.*, 1993; Comes *et al.*, 2011), we could identify the best way to promote the use of ensemble-based, probabilistic forecasts in our field, and thus promote the use of probabilistic-based information. More work along these lines should also be promoted.

## ACKNOWLEDGEMENTS

I would like to thank Lesley Gray and Douglas Parker for their thorough review of this article, and for their very valuable comments and suggestions.

## ORCID

Roberto Buizza  <http://orcid.org/0000-0002-3754-2271>

## REFERENCES

- Bechtold, P., Sandu, I., Klocke, D., Semane, N., Ahlgrimm, M., Beljaars, A., Forbes, R. and Rodwell, M. (2014) *The role of shallow convection in ECMWF's Integrated Forecasting System*. ECMWF Research Department Technical Memorandum No. 725. Reading: ECMWF, pp. 29.
- Berre, L., Pannekoucke, O., Desroziers, G., Stefanescu, S.E., Chapnik, B. and Raynaud, L. (2007) A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. *Proceedings of the ECMWF Workshop on Flow-Dependent Aspects of Data Assimilation*, Reading: ECMWF, Shinfield Park, Reading RG2 9AX, pp. 151–168.
- Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. and Shutts, G.J. (2008) The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134, 703–722.
- Buizza, R. (2001) Accuracy and economic value of categorical and probabilistic forecasts of discrete events. *Monthly Weather Review*, 129, 2329–2345.
- Buizza, R. (2014) *The TIGGE ensembles*. ECMWF Research Department Technical Memorandum No. 739. Reading: ECMWF, pp. 52. Available at: <http://www.ecmwf.int> [Accessed 1 November 2014].
- Buizza, R. and Leutbecher, M. (2015) The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, 141, 3366–3382.
- Buizza, R. and Palmer, T.N. (1995) The singular vector structure of the atmospheric general circulation. *Journal of the Atmospheric Sciences*, 52, 1434–1456.
- Buizza, R. and Palmer, T.N. (1998) Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126(9), 2503–2518.
- Buizza, R., Tribbia, J., Molteni, F. and Palmer, T.N. (1993) Computation of optimal unstable structures for a numerical weather prediction model. *Tellus*, 45A, 388–407.
- Buizza, R., Miller, M. and Palmer, T.N. (1999) Stochastic representation of model uncertainties in the ECMWF EPS. *Quarterly Journal of the Royal Meteorological Society*, 125, 2887–2908.
- Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M. and Zhu, Y. (2005) A comparison of the ECMWF, MSC and NCEP Global Ensemble Prediction Systems. *Monthly Weather Review*, 133(5), 1076–1097.
- Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G. and Vitart, F. (2007) The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quarterly Journal of the Royal Meteorological Society*, 133, 681–695.
- Buizza, R., Leutbecher, M. and Isaksen, L. (2008) Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 134, 2051–2066.
- Buizza, R., Stockdale, T. and Janousek, M. (2018) Error growth in the ECMWF medium-range ensemble. (under revision).
- Busemeyer, J., Townsend, R. and James, T. (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459.
- Comes, T., Hiete, M., Wijngaards, N. and Schultmann, F. (2011) Decision maps: a framework for multi-criteria decision support under severe uncertainty. *Decision Support Systems*, 52(1), 108–118. <https://doi.org/10.1016/j.dss.2011.05.008>.
- Coutinho, M.M. (1999) *Ensemble prediction using principal-component-based perturbations*. Thesis in Meteorology, National Institute for Space Research (INPE), pp. 136 (in Portuguese).
- Dutra, E., Diamantakis, M., Tsonevsky, I., Zsoter, E., Wetterhall, F., Stockdale, T., Richardson, D. and Pappenberger, F. (2013) The extreme forecast index at the seasonal scale. *Atmospheric Science Letters*, 14(4), 256–262. <https://doi.org/10.1002/asl2.448>.
- ECMWF. (2015) *ECMWF Strategy 2016–2025*. Reading: ECMWF. Available at: [https://www.ecmwf.int/sites/default/files/ECMWF\\_Strategy\\_2016-2025.pdf](https://www.ecmwf.int/sites/default/files/ECMWF_Strategy_2016-2025.pdf) [Accessed 1 June 2015].
- Ehrendorfer, M. (1994a) The Liouville equation and its potential usefulness for the prediction of forecast skills. Part I: Theory. *Monthly Weather Review*, 122, 703–713.
- Ehrendorfer, M. (1994b) The Liouville equation and its potential usefulness for the prediction of forecast skills. Part II: Applications. *Monthly Weather Review*, 122, 714–728.
- Frogner, I.L. (2018) Convection-permitting ensembles: challenges related to their design and use. (under revision).
- Fundel, V. (2018) Advancing the dialogue between research and applications of probabilistic weather information. (under revision).
- Hagedorn, R., Hamill, T.M. and Whitaker, J.S. (2008) Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: Two-meter temperatures. *Monthly Weather Review*, 136, 2608–2619. <https://doi.org/10.1175/2007MWR2410.1>.
- Hagedorn, R., Buizza, R., Hamill, M.T., Leutbecher, M. and Palmer, T.N. (2012) Comparing TIGGE multi-model forecasts with re-forecast calibrated ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138, 1814–1827.

- Hamill, T.M., Whitaker, J.S. and Mullen, S.L. (2006) Reforecasts: an important dataset for improving weather predictions. *Bulletin of the American Meteorological Society*, 87(1), 33–46. <https://doi.org/10.1175/BAMS-87-1-33>.
- Hoffman, R.N. and Kalnay, E. (1983) Lagged average forecasting, an alternative in Monte Carlo forecasting. *Tellus*, 35A, 100–118.
- Hollingsworth, A. (1980) An experiment in Monte Carlo forecasting procedure. *Proceedings of the ECMWF Workshop on Stochastic Dynamic Forecasting* (ECMWF, 1980). Reading: ECMWF, Shinfield Park, Reading, RG2 9AX, UK.
- Hoskins, B.J. (2013) Review article: the potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, 139, 573–584.
- Houtekamer, P.L. (2018) Using the hybrid gain algorithm to sample data assimilation uncertainty. (under revision).
- Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H.L. (1996) A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124, 1225–1242.
- Houtekamer, P.L., Mitchell, H.L. and Deng, X. (2009) Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, 137, 2126–2143.
- Houtekamer, P.L., Deng, X., Mitchell, H.L., Baek, S.-J. and Gagnon, N. (2014) Higher resolution in an operational ensemble Kalman filter. *Monthly Weather Review*, 142, 1143–1162.
- Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M. and Raynaud, L. (2011) *Ensemble of data assimilations at ECMWF*. ECMWF Technical Memorandum No. 636. Reading: ECMWF, 48 p. Available at: <https://www.ecmwf.int/en/elibrary/technical-memoranda> [Accessed 1 September 2011].
- Kai, J. and Kim, H. (2014) Characteristics of initial perturbations in the ensemble prediction system of the Korea Meteorological Administration. *Weather and Forecasting*, 29, 563–581.
- Lalauette, F. (2003) Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quarterly Journal of the Royal Meteorological Society*, 129(594), 3037–3057. <https://doi.org/10.1256/qj.02.152>.
- Leith, C. (1974) Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 102, 409–418.
- Leutbecher, M. (2018) Ensemble size: how suboptimal is less than infinity? (under revision). <https://doi.org/10.1002/qj.3387>.
- Leutbecher, M., Lock, S.J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N. and Weisheimer, A. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339.
- Lock, S.J. (2018) Treatment of model uncertainty from radiation by the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme and associated revisions in the ECMWF ensembles. (under revision).
- Lorenz, E.N. (1969a) The predictability of a flow which possess many scales of motion. *Tellus*, 21(3), 289–307.
- Lorenz, E.N. (1969b) Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26, 636–646.
- Lorenz, E.N. (1982) Atmospheric predictability experiments with a large numerical model. *Tellus*, 34, 505–513.
- Miyoshi, T. and Sato, Y. (2007) Assimilating satellite radiances with a Local Ensemble Transform Kalman Filter (LETKF) applied to the JMA Global Model (GSM). *SOLA*, 3, 37–40.
- Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T. (1996) The new ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122, 73–119.
- Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. and Vitart, F. (2011) *The new ECMWF seasonal forecast system (System 4)*. ECMWF Research Department Technical Memorandum No. 656. Reading: ECMWF, p. 49. Available at: <http://old.ecmwf.int/publications/> [Accessed 1 June 2011].
- Osinski, R. and Bouttier, F. (2018) Short-range probabilistic forecasting of convective risks for aviation based on a lagged-average-forecast ensemble approach. *Meteorological Applications*, 25, 105–118.
- Palmer, T.N. (2018) The ECMWF ensemble prediction system: looking back (more than) 25 years and projecting forward 25 years. (under revision). <https://doi.org/10.1002/qj.3383>.
- Palmer, T.N. and Tibaldi, S. (1988) On the prediction of forecast skill. *Proceedings of the ECMWF Workshop on 'Predictability in the Medium- and Extended Range'*. Reading: ECMWF, pp. 263–309.
- Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J., Steinheimer, M. and Weisheimer, A. (2009) *Stochastic parametrization and model uncertainty*. ECMWF Research Department Technical Memorandum No. 598. Reading: ECMWF, Shinfield Park, Reading, RG2 9AX, UK, pp. 42.
- Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H.L., Buizza, R. and de Roo, A. (2008) New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophysical Research Letters*, 35, L10404. <https://doi.org/10.1029/2008GL033837>.
- Park, Y.-Y., Buizza, R. and Leutbecher, M. (2008) TIGGE: preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134, 2029–2050.
- Piccolo, C. and Cullen, M.J.P. (2016) Ensemble data assimilation using a unified representation of model error. *Monthly Weather Review*, 144, 213–224.
- Ravazzolo, F. and Vahey, S.P. (2013) Forecast densities for economic aggregates from disaggregate ensembles. *Studies in Nonlinear Dynamics and Econometrics*, 18(4), 367–381. <https://doi.org/10.1515/snde-2012-0088>.
- Raynaud, L., Berre, L. and Desrozier, G. (2012) Accounting for model error in the Météo-France ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 138, 249–262.
- Richardson, D.S. (2000) Skill and economic value of the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 126, 649–668.
- Shutts, G.J. (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131, 3079–3100.
- Siegert, S. (2018) The ensemble-adjusted Ignorance score for forecasts issued as Normal distributions?. (under revision).
- Sun, H.S. and Jin, Z. (2016) Estimating credit risk parameters using ensemble learning methods: an empirical study on loss given default. *Journal of Credit Risk*, 12(3), 43–69. <https://doi.org/10.21314/JCR.2016.212>.
- Taylor, J. and Buizza, R. (2003) Using weather ensemble prediction in energy demand forecasting. *International Journal of Forecasting*, 19, 57–70.
- Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A. (2009) The European flood alert system. Part 1: Concept and development. *Hydrology and Earth System Sciences*, 13, 125–140.
- Toth, Z. and Kalnay, E. (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological Society*, 74, 2317–2330.
- Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125, 3297–3319.
- Tracton, M.S. and Kalnay, E. (1993) Operational ensemble prediction at the National Meteorological Center: practical aspects. *Weather and Forecasting*, 8, 379–398.
- Vannitsem, S., Wilks, D. and Messner, J. (2018) *Statistical Post-processing of Ensemble Forecasts*. Dordrecht: Elsevier (eBook ISBN: 9780128122488; Paperback ISBN: 9780128123720).
- Verlinden, K.L. and Bright, D.R. (2017) Using the second-generation GEFS reforecasts to predict ceiling, visibility, and aviation flight category. *Weather and Forecasting*, 32, 1765–1780. <https://doi.org/10.1175/WAF-D-16-0211.1>.
- Vitart, F. (2004) Monthly forecasting at ECMWF. *Monthly Weather Review*, 132, 2761–2779.
- Vitart, F. (2013) *Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years*. ECMWF Research Department Technical Memorandum No. 694. Reading: ECMWF, pp. 28.
- Vitart, F., Buizza, R., Alonso Balmaseda, M., Balsamo, G., Bidlot, J.R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F. and Palmer, T.N. (2008) The new VAREPS-monthly forecasting system: a first step towards seamless prediction. *Quarterly Journal of the Royal Meteorological Society*, 134, 1789–1799.
- Vitart, F., Balsamo, G., Buizza, R., Ferranti, L., Keeley, S., Magnusson, L., Molteni, F. and Weisheimer, A. (2014) *Sub-seasonal predictions*. ECMWF Research Department Technical Memorandum No. 734. Reading: ECMWF, pp. 47.

- Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. and 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. and Zhu, Y. (2008) Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A*, 60, 62–79.
- Weisheimer, A. (2018) How confident are predictability estimates of the winter North Atlantic Oscillation? (under revision).
- Weyn, J.A. and Durran, D.R. (2018) The scale dependence of initial-condition sensitivities in simulations of convective systems over the southeastern US. (under revision). <https://doi.org/10.1002/qj.3367>.
- Xian, P. (2018) Current state of the global operational aerosol multi-model ensemble: an update from the International Cooperative for Aerosol Prediction (ICAP). (under revision).
- Zanna, L. (2018) Uncertainty and scale interactions in ocean ensembles: from seasonal forecasts to multi-decadal climate predictions. (under revision).
- Zhang, Z. and Krishnamurti, T.N. (1999) A perturbation method for hurricane ensemble predictions. *Monthly Weather Review*, 127, 447–469.
- Zsoter, E., Buizza, R. and Richardson, D. (2009) “Jumpiness” of the ECMWF and UK Met Office EPS control and ensemble-mean forecasts. *Monthly Weather Review*, 137, 3823–3836.

**How to cite this article:** Buizza R. Introduction to the special issue on “25 years of ensemble forecasting”. *Q J R Meteorol Soc* 2018;1–11. <https://doi.org/10.1002/qj.3370>