# Ensemble forecasting

**François Bouttier[1], Roberto Buizza[2]**

1. Centre national de recherches météorologiques, Météo-France, Toulouse, France (francois.bouttier@meteo.fr)

2. Professor, Scuola Universitaria Superiore Sant'Anna Pisa, Pisa, Italy (roberto.buizza@santannapisa.it)

**ABSTRACT:** Twenty-five years ago the first operational ensemble forecasts were issued. This triggered a paradigm shift in weather prediction: for the first time, forecasters and users were able to 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 forecasts for the short and medium-range, the monthly and seasonal time scales, but also to estimate the initial state of the atmosphere. In this article, we briefly review how we got here, starting from the establishment of the global ensembles in the 1990s, to the use of very high-resolution, limited-area ensembles for the short-range. We discuss what are the key characteristics of an ensemble system, and why they provide more valuable information than single forecasts. Finally, we look to the future, arguing that it is time that we all **think ensemble**!

## A paradigm shift in weather prediction: the move towards ensembles

Since numerical models have been used to predict the weather, forecasters have realized that there are cases when forecast errors would remain small even 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 small initial errors could grow very rapidly and affect forecast quality in a very short time.

Scientists and operational forecasters started investigating whether it would be possible to know in advance, when a forecast is issued, whether the situation was easy (or, say, easier than average) to predict. In other words, they were looking at an objective method that could be used to provide a level of forecast confidence. This confidence could be expressed in probabilistic terms, for example by giving the probability that a specific event (e.g. rainfall in excess of 50 mm over 6 hours) would occur. It could also be expressed in terms of weather scenarios, each with an assigned probability. In this way forecasters could assess the range of possible weather that could occur in the future, and provide their users with the probability of occurrence of each scenario.

In the 1980s, different approaches were tested, all based on ensembles, i.e. on mixing and combining a number of forecasts either started from different conditions, or generated using different models, or by combining the two.

The main idea behind an ensemble approach is very simple (Figure 1). Consider a variable such as temperature, at a specific location. If a forecas-

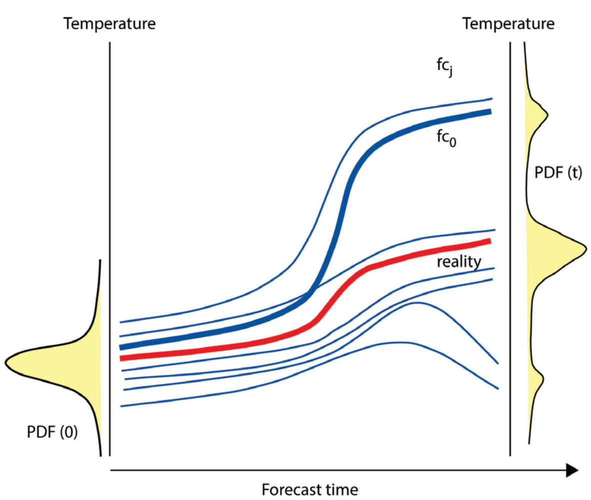


*Figure 1. Schematic of an ensemble prediction system, based on an ensemble of N forecasts (the blue lines) to predict reality (the red line). Both at initial time and at forecast time t, the N forecasts can be used to estimate the Probability Density Function (PDF) of atmospheric states. The PDF can then be used to estimate, e.g., the mean value, the range of possible outcomes, the probability that certain thresholds are exceeded. The "spread", i.e. the level of divergence between the forecasts, can be used as an estimate of the confidence of the forecast: in a reliable ensemble, a small spread, i.e. forecasts close to each other, indicates a high confidence, a more predictable case. By contrast, a large spread indicates a less predictable situation.*

ter has access to an ensemble of N forecasts instead of only one single forecast, s/he can estimate the range of possible outcomes, and/or the most probable value, and/or the probability that temperature would be higher, or lower, than a certain value. A forecaster may also like to visualize the outcomes as N maps of possible meteorological scenarios.

In the 1980s, different techniques were tried to develop reliable and accurate ensembles. An ensemble is reliable if there is, on average, a correspondence between a forecast probability and the probability of occurrence. In a reliable ensemble, if an event is predicted with an 80% probability, then this event occurs 80% of the time (on average over a large sample of cases). An ensemble is accurate when the average error of a probabilistic forecast is small. In the USA, ensembles based on lagged forecasts (i.e. forecasts started at different times and days, e.g. the 9 forecasts issued every 6 hours over the past 2 days), were tried. This method delivered forecasts with a reasonable quality for the medium forecast range (say after one week), but not for the shorter forecast range, since the 'oldest' forecasts were too old to be accurate. At ECMWF a different method was tried: ensembles of forecasts were initialized at the same time, but with initial conditions perturbed in a random way. This method did not deliver good results, since the random perturbations did not lead to very different forecasts, and the ensemble suffered from under-dispersion and very poor reliability, since the forecasts remained too similar to provide valuable information about possible future scenarios.

The late 1980s and the early 1990s saw the development and testing of more promising methods both at ECMWF and at NCEP. 1992 saw the implementation of the first two operational ensemble systems at the European Centre for Medium-Range Weather Forecasts (ECMWF) in Europe, and at the National Centers for Environmental Prediction (NCEP) in the United States of America. They were followed by the Météorological Service of Canada (MSC) in 1995, and by others a few years later, both for the global scale and for specific regions.

These implementations generated 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.

## Are ensemble-based, probabilistic forecasts more valuable than single ones?

Today, it is widely accepted that forecasts have to include uncertainty estimations, confidence indicators that allow forecasters to estimate how 'predictable' the future situations are. 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 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 more accurately estimate the analysis error.

There are at least two reasons why ensemble-based, probabilistic forecasts are more valuable than single forecasts. The first reason 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 more generally, any event of interest can occur. In other words, ensembles provide users with more complete information, with extra pieces of information about the future weather scenario. One way to measure such a difference is to evaluate the Potential Economic Value (PEV; Richardson 2000) of a forecasting system.

The PEV is based on a simple user model, called "cost-loss": Given a predefined binary weather event (e.g. subzero temperatures, gale-force winds, etc), a user is assumed to have the choice between either paying an insurance premium C (the "cost") to protect his/herself against the consequences of this event, or taking the risk to incur damage worth L (the "loss") in case the event occurs. We assume that the user always takes the best option, and it can be shown that C and L only matter through their "cost-loss" ratio, C/L. For any given forecasting system, the PEV measures the average expenses caused by forecast errors: thus, it can be used to quantify the savings brought by replacing a forecasting system by a better one.

As an example, Figure 2 shows the average PEV for the ECMWF single high-resolution forecast and the medium-range/monthly ensemble (ENS) probabilistic forecast of four events:

• 2-meter temperature cold anomaly (with respect to climatology) lower than 4 degrees;

• 2-meter temperature warm anomaly (with respect to climatology) greater than 4 degrees;

• 10-meter wind-speed stronger than 10 m/s;

• Total precipitation larger than 1 mm.

Figure 2 shows that the ENS-based probabilistic forecasts have a higher PEV for all ranges of users. Consider, for example, a user who can protect against a loss of L=1,000,000 euros by spending C=100,000 euros, if s/he knew the temperature and rainfall amount in 6 days. This user has a cost/loss ratio of 0.1. Suppose that this user has access to ECMWF ensemble and high-resolution forecasts. For this user, Fig. 2 shows that ensemble-based, probabilistic forecasts are more valuable than single, high-resolution forecasts. In particular, for precipitation (Fig. 2, bottom-right panel), the 6-day single high-resolution forecasts have no value. This analysis can be generalized to a non-monetary framework, for instance one can measure the value of ensemble predictions for limiting human casualties during severe weather events (e.g. Hewson and Tsonevsky, 2016).

The second reason why ensemble-based, probabilistic forecasts are more valuable is that an

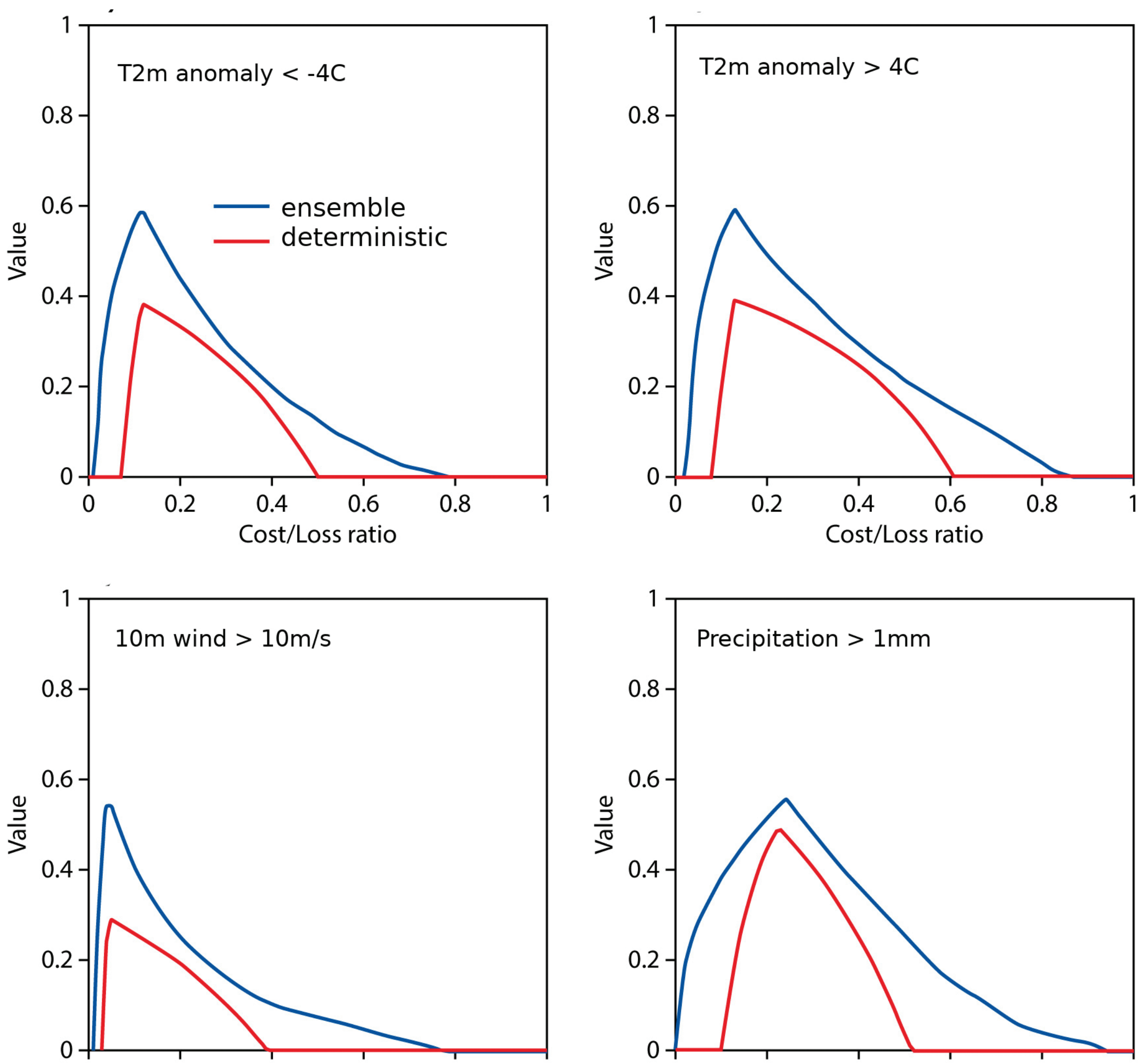


*Figure 2. Potential Economic Value (PEV) of ECMWF single high-resolution forecasts (red lines) and ENS-based probabilistic forecasts (blue lines), for cost loss ratios C/L ranging from 0 to 1, for four different forecasts: 2-meter temperature cold anomaly lower than 4 degrees (top-left panel), 2-meter temperature warm anomaly greater than 4 degrees (top-right panel), 10-meter wind-speed stronger than 10 m/s (bottom-left panel) and total precipitation larger than 1mm (bottom-right panel). PEV average values have been computed considering the ECMWF operational forecasts for October-November-December 2016, verified against SYNOP observations.*

ensemble system provides forecasters with more consistent (i.e. less changeable) successive forecasts. This can be easily verified if one considers consecutive ensemble-mean forecasts, issued 24-hours apart and valid for the same verification time. Results indicate that they jump less, i.e. are more consistent, than the corresponding single forecasts. In other words, ensemble-based, dynamical averaging makes successive forecasts more consistent.

Generally speaking, the main reason why ensemble-based, probabilistic forecasts are more reliable, accurate and valuable, is that by using the whole ensemble we can filter out the unpredictable scales (e.g. by constructing averages among all, or a selection of the available forecasts), or we can assess in a reliable way whether there is any chance that certain events (e.g. heat-waves leading to droughts, or extreme rainfall leading to flooding) might occur.

## Who uses ensemble prediction?

Ensembles are mostly useful for extreme weather prediction (high precipitation, winds and violent thunderstorms). They certainly have great potential for other meteorological applications (such as road condition management or wind power prediction), but these benefits have only begun to materialize: in practice, the use of ensembles has been hampered by technical cost and conceptual complexity. Ensembles produce large amounts of data that are still imperfect; using them requires significant training efforts, because translating ensemble information into decisions can be tricky.

Some forecasters claim that it is harder to apply human expertise to ensembles than to deterministic forecasts. Ironically, most of them routinely compare forecasts from various models and meteorological centres, which is a makeshift ensemble prediction. Ensemble forecasts add provable value to the forecasting process, but bringing this value to real users will still require much training and modernization of forecasting habits: today, we are facing dozens of years of experience in the use of deterministic models, compared to much less effort dedicated to using ensembles.

Let us consider an ensemble-based fog forecast, for instance. In order to make an informed forecast, one needs to apply statistical and/or physical corrections to numerical model output, to mix information from various ensembles and deterministic models, to take into account recent observations and nowcasts. This information then needs to be summarized as a preferred scenario, and the probabilistic forecast information has to be concisely communicated to various users. The message will be different for users who need to know if there is even the slightest chance of fog, and for users who only want a warning if fog is quasi certain. An automated forecasting process that would not include all these steps would be much less effective than a human forecaster who uses more traditional methods... or even a clever user of readily accessible meteorological websites. In a nutshell, the value of ensemble forecasts will not materialize until forecast post-processing systems are substantially improved to bridge the current gap between numerical model output and end users (including forecasters).

## Global ensembles for medium-range: how are they designed?

Ensembles are designed to simulate the sources of forecast errors linked to initial conditions and model uncertainties. Model uncertainties arise because the models that we use to generate weather forecasts are imperfect, they only simulate certain physical processes on a finite mesh, and they do not resolve all the scales and phenomena that occur in the real world. Initial condition uncertainties arise because observations are affected by observation errors, and do not cover the whole globe with the same frequency. Furthermore, the process of estimating the initial state of the system from which a forecast is computed, is based on statistical assumptions and approximations, including the imperfections of the model used to assimilate the observations

In the first version of the ECMWF global ensemble (Molteni et al. 1996), initial uncertainties were simulated using singular vectors (SVs), which are the perturbations with the fastest growth over a finite time interval, to simulate initial uncertainties (Buizza and Palmer 1995). SVs provided a very good basis to define the initial perturbations of the ECMWF ensemble: compared to random initial perturbations, they were characterized by a much quicker amplification, similar to the forecast error growth rate. SVs remained the only type of initial perturbations used in the ECMWF ensemble until 2008, when the ensemble of data assimilations (EDA) started to be used in combination with singular vectors (Buizza et al 2008). SVs are an essential component of the ECMWF ensemble, and they keep providing dynamically-relevant information about initial uncertainties that could have a strong contribution to forecast errors.

There are several ways to simulate initial and model uncertainties. Indeed, in the first version of NCEP global ensemble, bred-vectors (BVs) were used to simulate initial uncertainties instead of SVs. The BV cycle aims to emulate the data-assimilation cycle: it is based on the notion that analyses generated by data assimilation will accumulate growing errors by the virtue of perturbation dynamics (Toth and Kalnay 1997). This is due to the fact that 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 their dynamics during the assimilation window. In contrast, growing errors, even if they are reduced by the assimilation system, 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 developed following a different approach. In Canada, they decided to adopt a Monte Carlo approach, designed to simulate both initial uncertainties due to observation errors and data assimilation assumptions, and model uncertainties (Houtekamer et al 1996). The Canadian ensemble was the first one that included a simulation of model uncertainties, and it tried to include as many error sources as possible.

Following the Canadian example, the simulation of model uncertainties was introduced in the ECMWF ensemble in 1999, using, for the first time in numerical weather prediction, a stochastic approach to simulate the effect of model errors linked to the physical parameterisation schemes (Buizza *et al.* 1999). Since then, many other operational ensembles have also included schemes to simulate model uncertainties (see e.g. Buizza 2014, for a review of the main characteristics of the operational global ensembles).

At present, four main approaches are followed in ensemble prediction to represent model uncertainties (see Palmer et al 2009 for a review):

- A multi-model approach, where different models are used in each ensemble members; models can differ entirely or only in some components (e.g. in the convection scheme); (Descamps *et al.*, 2015)
- 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 only approximately known: one example is the ECMWF Stochastically Perturbed Parametrization Tendency scheme (SPPT, Buizza et al 1999);
- A stochastic back-scatter approach, where a Stochastic Kinetic Energy Backscatter scheme (SKEB) is used to simulate processes that the model cannot resolve, such as the upscale energy transfer from scales below the model resolution to the resolved scales: an example is the current ECMWF SKEB scheme (there is a plan to switch it off in the future since it does not appear to provide any significant benefit).

## What characterizes an ensemble configuration?

We have previously discussed two key aspects that define the characteristics of an ensemble: the methodology used to simulate initial uncertainties, and the simulation of model approximations. Other important characteristics of an ensemble are its horizontal and vertical grid resolution, and the number of ensemble members. 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, i.e. to have enough members to filter out the unpredictable scales. Today, most of the operational ensembles have between 20 and 50 members. While this may be enough for some applications, the prediction of rare events typically requires at least 50 to 100 members.

Taking into account users' demands, and given that we need to generate forecasts in a reasonable amount of time (say about 1 hour) with finite computing resources, compromises have to be taken when an ensemble configuration is defined. Ideally, we would like to use as many members as possible (say in the order of 100), the highest resolution possible (to be able to simulate the finest scales), and to extend the forecast length for as long as possible, to provide a wider audience with ensemble-based, probabilistic forecasts. Unfortunately, it is not possible, and this is why, for example, ECMWF uses three different resolutions to generate ensembles for the medium-range, the monthly and the seasonal time scale. Yet, none of these resolutions are as high as the ones used by (e.g.) Météo-France to generate their limited-area short-range forecasts. Such resolutions would be prohibitively expensive with the forecast ranges used at ECMWF. Table 1 summarizes the characteristics of ECMWF and Météo-France ensembles.

sembles, say about 100-200, compared to about 25-50 for the medium-range.

Extracting predictable signals for the extended range has benefitted from the use of ensembles of re-forecasts, i.e. forecasts generated using the current schemes but over past years. At ECMWF, the re-forecast 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 re-forecast suite of the seasonal ensemble covers the past 30 years (a 15-member ensemble is run once a month, for the past 30 years). Ensemble re-forecasts are essential for obtaining a statistically significant estimate of the skill of monthly and seasonal forecasts, and to extract predictable signals from them.

| | Maximum forecast horizon | Resolution (horizontal/vertical) | Number of members (production frequency) | reforecast (number members, archive length) |
|---|---|---|---|---|
| ECMWF (lateral boundary project) | 6.5 days | 18 km / 91 levels | 51 (x2 / day) | No |
| ECMWF (medium range) | 15 days | 18 km / 91 levels | 51 (x2 / day) | (22/week, 20 years) |
| ECMWF (monthly forecast) | 46 days | 36 km / 91 levels | 51 (x2 / week) | (22/week, 20 years) |
| ECMWF (seasonal forecast) | 7 months | 80 km / 91 levels | 51 (x1 / week) | (15/month, 30 years) |
| Météo-France regional prediction (Arome EPS) | 45 hours | 2.5 km / 91 levels | 16 (x4 / day) | (12/day, 1 year) |
| Météo-France global prediction (PEARP) | 4.6 days | 10 km / 90 levels over europe | 35 (x2 / day) | (80/month, 30 years) |
| Météo-France seasonal forecast | 7 months | 80 km / 91 levels | 51 (x1 / month) | (15/month, 20 years) |

*Table 1. Key characteristics of the ECMWF and Météo-France ensemble predictions, in terms of maximum forecast length, grid resolution (horizontal and vertical), number of members (with forecast update frequency), and reforecasting system (forecast frequency and length of history). Note the PEARP system resolution is location dependent and also issues boundary conditions for Arome-EPS. (Spring 2019 data)*

## Ensembles are also used for the sub-seasonal and seasonal time scales

Since the beginning of the 2000s, global ensembles have been used to generate monthly and seasonal forecasts. These extended-range ensembles have a coarser resolution than the medium-range ensembles, in order to limit the production costs (See Table 1). Compared to the medium-range ensembles, most of them also include a dynamical ocean model, to be able to better simulate the propagation of coupled ocean-atmosphere phenomena, like the organized convection associated with the Madden-Julian Oscillation.

Using ensembles is essential at extended time ranges (monthly to seasonal), where no reliable nor accurate signal could be extracted from single forecasts. Increasing evidence suggests that for this time range the number of ensemble members should be higher than in medium-range ensembles, say about 100-200, compared to about 25-50 for the medium-range.

These two facts (the need for a large membership and the need for re-forecasts) make the monthly and seasonal ensembles very expensive in terms of computing power. This is why they are characterized by a relatively coarse resolution (Table 1).

## Ensembles in atmospheric analyses and reanalyses

The initial states of numerical forecasts are prepared from observations using an algorithm called "data assimilation" that produces numerical snapshots of the atmosphere called "analyses'. Since its inception in 1995, the Canadian ensemble has 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.

Since 2008, ECMWF and Météo France have used a slightly different approach called Ensemble Data Assimilation (EDA), whereby one runs an ensemble of N separate data assimilation procedures, each using perturbed observations and model uncertainty schemes. Observations are perturbed to simulate the fact that observations are not perfect due to observation errors, and to take into account observation representativeness errors, consistently with the way they are used in the assimilation procedures. As in ensemble forecasts, ensembles of data assimilations contain simulations of model uncertainties to represent the fact that the models used to define the analyses (i.e. the forecast initial conditions) are not perfect. Table 2 lists the key characteristics of the Ensemble of Data Assimilations used at ECMWF and Météo France.

Since 2008, at ECMWF, the ECMWF EDA is used in combination with SVs to define the initial conditions of the medium-range/monthly ensemble (Buizza et al 2008). The addition of EDA-based perturbations has had a major impact on the ensemble reliability and accuracy in the short forecast range over the extra-tropics, and for the whole forecast range over the tropics.

## The original motivation for fine scale ensemble prediction

In the early 2000's, the usefulness of global ensemble prediction was beginning to be recognized for synoptic scale prediction at scales larger than 500km. This was more or less the effective resolution of current global ensemble prediction systems, as dictated by computing resources. Perturbation algorithms were rather mature, so most of the subsequent progress was expected to come from computer upgrades. In the meantime, the potential of limited area, kilometric-scale models was becoming obvious - starting by experiments at NCAR, USA - for the prediction of severe weather events such as violent thunderstorms, strong winds, and intense orographic precipitation. Their finer mesh allowed them to tap into new sources of atmospheric predictability (Marsigli et al 2001). It ended the old scaremongering according to which it was useless to run models at higher resolutions than the global ones, because fine scale events could not be predicted since there was significant uncertainty at the synoptic scales. Indeed, a few pioneering centres began running operational fine scale limited area models with great success, first deterministically, then as ensembles

## Predicting heavy rain and thunderstorms

Intense orographic precipitation is probably the phenomenon that best demonstrates the usefulness of high model resolution, because lower resolution ones tend to severely underestimate their intensities. Numerical experiments have shown that these events can become fairly predictable if a high enough resolution is used. This is because, even if the synoptic forcing is imperfectly known, local orography often is a key ingredient that is easy to specify. The huge human and material losses linked to high Mediterranean flood events have led some modelling teams to heavily invest in high resolution forecasting tools.

In 2001, the Arpa-SMR group in Bologna was the first European one to run a fine scale ensemble prediction in real time (Marsigli et al 2001). It used a homegrown model (Lambo, that covered Nor-

*Table 2: main characteristics of atmospheric ensemble data assimilation systems at ECMWF and Météo-France, in terms of grid resolution (as in Table 1), number of members, data assimilation algorithms (3D-Var and 4D-Var i.e. 3D and 4D variational techniques) (Spring 2019 data).*

| | Resolution (horizon / vertical) | Number of members | assimilation algarithm |
|---|---|---|---|
| CEPMMT (global) | 36 km / 91 levels | 25 | 4D-Var |
| Météo-France (global Arpege) | 50 km / 105 levels | 25 | 4D-Var |
| Météo-France (regional Arome) | 3.8 km / 90 levels | 50 | 3D-Var |

thern Italy with a 20 km mesh) driven by members of the ECMWF ensemble prediction system. The dynamical flow adaptation to local orography was enough to significantly improve the rainfall forecasts. Arpa-SMR have since then upgraded their system. Their example motivated Météo-France to test their own ensemble called "Arome EPS" over the Mediterranean regions during the 2012 HyMeX-SOP1 field experiment (Vié *et al* 2011). The USA, too, have demonstrated great experience in high resolution ensemble prediction.

Another key objective of high resolution ensembles is the prediction of heavy thunderstorms, even if they are not driven by orography. These plainland thunderstorms are a major issue in the American Midwest and in some European regions such as Germany. They are tricky to forecast because, in comparison with orographic precipitation, they are much more sensitive to the model initial condition and their predictability is very limited (typically, to less than 24 hours): numerical prediction systems often fail to produce useful warnings for these thunderstorms. Thus, ensemble prediction systems in these regions tend to be more recent, with more complex ensemble perturbation schemes (typically, a Kalman-filter type representation of initial uncertainty, and perturbations to the physical parametrisations).

## The rise of kilometric ensembles

The German national weather service (DWD) was the first to dedicate large human and computer resources to developing such a system, Cosmo-DE-EPS at 2.8km horizontal resolution, which is well designed for tracking convection because it is frequently refreshed (every 3 hours) with relatively short forecast ranges (up to 21 hours). This system used 5 different physical parametrisation packages and 4 lateral boundary conditions derived from several global prediction systems. Cosmo-DE-EPS has since then been upgraded and it remains a leader in its class.

Other operational kilometric-resolution ensemble systems were implemented around 2012 at the Met Office (MOGREPS-UK), in 2017 at Météo-France (Arome-France-EPS, Bouttier *et al* 2015) and in the USA, among others. Each has its own peculiarities that reflect differing national priorities (Figure 3). Today, most major meteorological services are implementing operational kilometric scale ensembles. Unlike global systems, it is difficult to mutualize their production between several countries, because these systems only cover small geographical areas, due to their high numerical cost. Still, there is much ongoing data exchange for the computation of numerical products (for instance in the EU SESAR project for European aviation) and for the scientific and technical development of ensemble systems

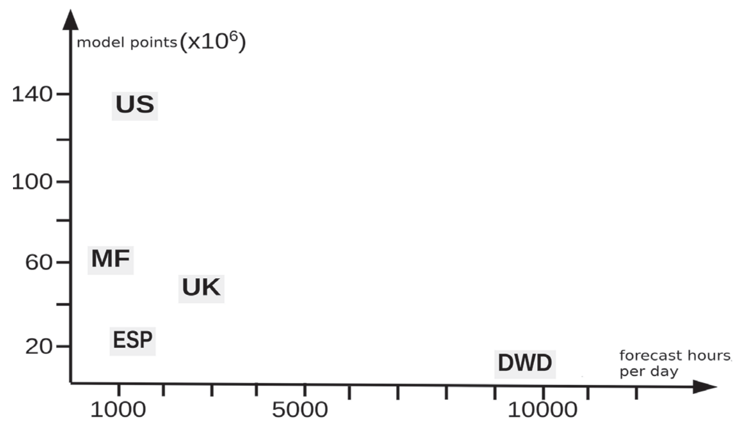


*Figure 3. Comparison of five 2018 kilometric scale operational ensembles: US (NCEP, USA), MF (Météo-France), UK (Met Office, United Kingdom), ESP (AEMET, Spain), DWD (Germany). The lower points mark the ones with the most expensive model, the rightmost ones have the greatest number of members, update frequency and forecast range, measured by the number of hours of model runs per day.*

## Is there a benefit in having both global and limited-area ensembles?

It should be clear now, from the discussions above, that there is a clear benefit for the users in having both a global ensemble at a coarser resolution, and a high-resolution ensemble focused on a region of interest, as it is the case for ECMWF and Météo-France. Thanks to this combination, forecasters and users can look for warnings of the possibility that high-impact weather events can occur a few days ahead, using the ECMWF global ensemble. As the events get closer to real time, one can use the higher-resolution, short-range forecasts based on the Météo-France ensemble to obtain more detail.

The higher-resolution ensemble can be used as a magnifying lens, able to provide information on the spatial and temporal scales that are not resolved in the global ensemble. Their combination makes it possible for forecasters and users to have a long-term view of what could happen a few weeks or months ahead, and at the same time to assess a few hours or days in advance whether there is a chance that extreme, localized weather events could affect some specific locations.

This is illustrated in Figure 4 on a catastrophic flooding case (Cannes, Oct 2015): ECMWF's global ensemble warned about one week in advance of the possibility of heavy rain. The Météo-France global ensemble gave a more geographically precise signal 3 days in advance, and Arome-France-EPS predicted very high precipitation over the right urban area, with some better indication of the intensities to expect.

## A look to the future

Looking to the future, three trends can be detected in the way ensembles are being 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 is linked to results obtained in the past two decades that showed that by adding relevant processes we can further improve the quality of the existing forecasts, and we can 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, that include more accurate simulation of relevant 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 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 only relevant 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. In the beginning, a coupled ocean-land-atmosphere was only used for the seasonal and the monthly time scales, and then it was introduced into the medium-range ensemble once it was realized that it improved its reliability and accuracy. From a 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 the maintenance and upgrading processes. The diagnostic and evaluation of a model version over different time scales can help identify undesirable behaviours that could produce forecast errors.

The third trend comes from the need to better resolve the smaller scales, their interaction with the slightly-less-smaller-scales, and so on. All scales are relevant in weather prediction, and errors propagate from the smallest to the larger scales. If we consider the current ensembles, we should not forget that even if they use grid resolutions of 2-20 km (see Table A), they are capable of realistically resolving only scales that are about 5-6 times their grid resolution. This is because the scales closest to the model grid spacing cannot be simulated in an accurate way, for technical reasons. Thus, looking at Table 1, today's ECMWF global ensemble and Météo-France's regional ensembles have effective resolutions of about 15 km, respectively. If we want to be able to predict phenomena, such as intense wind storms of heavy precipitation events, it is thus essential that

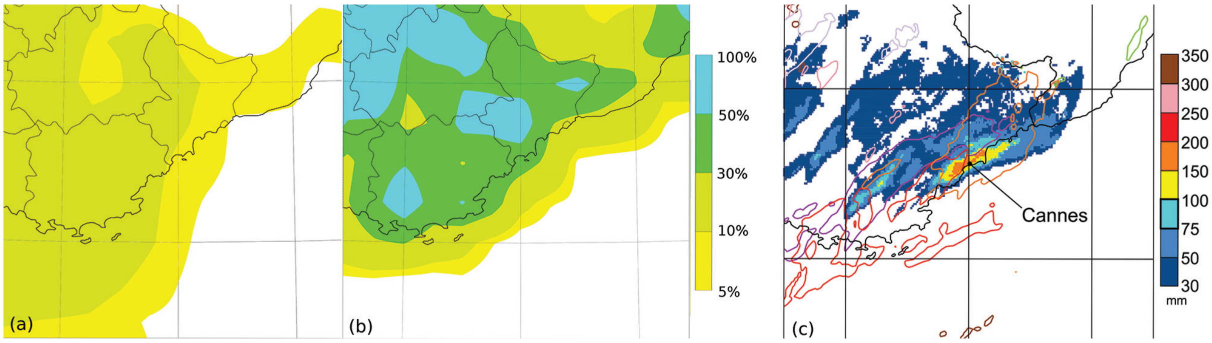


*Figure 4. Ensemble prediction on the Cannes city area for the 3 Oct 2015: a) ECMWF 5-day ensemble prediction - the colours indicate the probability of moderate to high precipitation (more than 20mm in 24 hours). b) Arpege 3-day ensemble forecast (PEARP system), with the same colour scheme; c) Arome-EPS 21-h forecast. The contours delineate the area of high precipitation forecast (model reflectivities>44dBz). The colour shading shows the radar observation of actually observed precipitation (Figure provided by O. Nuissier, CNRM).*

we increase the models' resolution to a few hundred meters for the limited-area models, and to about 1 km for the global models.

## In conclusion ... think ensemble!

The future will see new applications of ensembles. Their reliability and accuracy will further improve thanks to advances in model design, data assimilation methods, and in the schemes used to simulate the initial and model uncertainties. Resolution will be increased, to improve large-scale predictions and to start resolving finer, relevant scales. Ensembles of analyses and forecasts will be more closely linked together, to improve their performance. Physical processes that are not yet included in the models but are relevant for weather prediction will be included, to make the forecasts more and more realistic.

## References

Bouttier F., Raynaud L., Nuissier O., Ménétrier B., 2015. Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Q. J. R. Meteorol. Soc.*, 142, 390-403. doi: 10.1002/qj.2622

Buizza R., 2014. *The TIGGE medium-range, global ensembles*. ECMWF Research Department Technical Memorandum No. 739, 53 p. (disponible sur www.ecmwf.int).

Buizza R., Leutbecher M., 2015. The forecast skill horizon. *Q. J. R. Meteorol. Soc.*, 141, 3366-3382.

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.

Buizza R., Miller M., Palmer T.N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, 125, 2887-2908.

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

Descamps L., Labadie C., Joly A., Bazile E., Arbogast P., Cébron P., 2015. PEARP, the Météo-France short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, 141, 1671-1685. doi: 10.1002/qj.2469

Hewson T., Tsonevsky I. 2017: Flash floods over Greece in early September 2016. ECMWF Newsletter no.150, winter 2017, 2-3. (disponible sur www.ecmwf.int )

Houtekamer P.L., Lefraive L., Derome J., 1996. A system simulation approach to ensemble prediction. *Mon. Weather Rev., 124*, 1225-1242.

Labadie C., Cébron P., Raynaud P. Descamps L., Arbogast P. 2017: Le vendeur de glaces et le chaos. La Méteorologie, 96, 21-25.

Marsigli C., Montani A., Nerozzi F., Paccagnella T., Tibaldi S., Molteni F., Buizza R., 2001. A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. *Q. J. R. Meteorol. Soc.*, 127, 2095-2115.

Molteni F., Buizza R., Palmer T.N., Petroliagis T., 1996. The new ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.*, 122, 73-119.

Palmer T.N., Buizza R., Doblas-Reyes F., Jung T., Leutbecher M., Shutts G.J., Steinheimer M., Weisheimer A., 2009. *Stochastic parametrization and model uncertainty*. ECMWF Research Department Technical Memorandum No. 598, 42 p. (disponible sur www.ecmwf.int).

Raynaud L., Berre L., Desroziers G., 2011. An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system. *Q. J. R. Meteorol. Soc.*, 137, 607-619.

Richardson D.S., 2000. Skill and economic value of the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, 126, 649-668.

Richardson D.S., 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. R. Meteorol. Soc.*, 127, 2473-2489.

Toth Z., Kalnay E., 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, 125, 3297-3319.

Vié B., Nuissier O., Ducrocq V., 2011. Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Weather Rev.*, 139, 403-423.