

# ECMWF Feature article

from Newsletter Number 153 – Autumn 2017

METEOROLOGY

25 years of ensemble  
forecasting at ECMWF



[www.ecmwf.int/en/about/news-centre/media-resources](http://www.ecmwf.int/en/about/news-centre/media-resources)

doi:10.21957/bv418o

*This article appeared in the Meteorology section of ECMWF Newsletter No. 153 – Autumn 2017, pp. 20-31.*

## 25 years of ensemble forecasting at ECMWF

Roberto Buizza, David Richardson

Twenty-five years ago ECMWF was one of the first forecasting centres which started to issue operational ensemble forecasts. The availability of such forecasts marked a paradigm shift in weather prediction: for the first time, forecasters and users were provided with reliable and accurate estimates of the range of possible future scenarios, and not just with a single realisation of the future. Today ensembles are used not only in forecast mode, to provide forecasts for the short- and medium-range, the monthly and the seasonal timescale, but also in analysis mode, to provide estimates of the initial state of the Earth system. These ensemble-based forecasts and analyses provide more complete information than single, deterministic forecasts, for example through indices of the risk of severe events; probabilities of the occurrence of weather events; the range of possible values at specific locations; alternative weather scenarios; and weekly-mean anomalies.

ECMWF ensembles have been developed, implemented and maintained thanks to the work of very many people at ECMWF and in its Member States, and of visitors who, over the years, have spent time working with us to understand their performance and to improve them further. This started well before 1992, with trials that helped us to identify the strategy to be followed, and it is continuing today, thanks to the interactions with scientists within global projects such as the World Meteorological Organization's TIGGE and S2S (sub-seasonal to seasonal prediction) projects. As we explain in this article, ensemble forecast performance is strongly linked to the quality of the model and the assimilation system used to generate the initial conditions; the assimilation of an increasing number of observations; the strategy that we have followed to simulate initial and model uncertainties; and the ensemble forecast configuration. Neither the performance of ensemble forecasts nor the range of ensemble-based products that we currently offer would be what they are today without those many contributions.

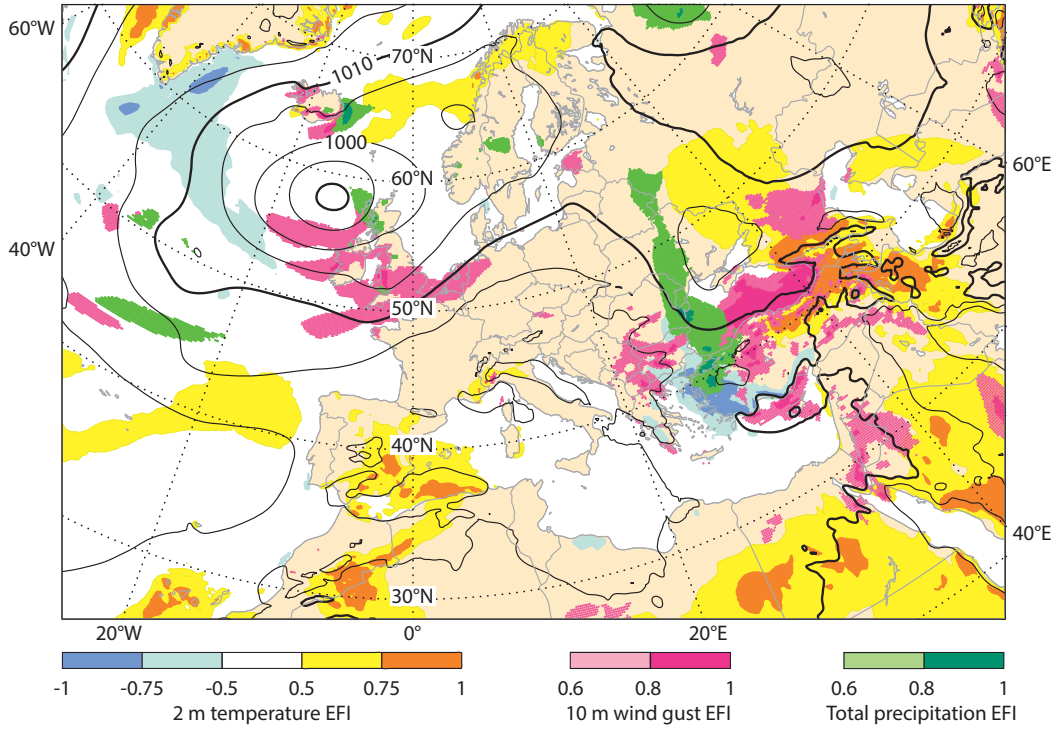
This article presents some examples of ensemble-based forecasts and explains their added value; it briefly reviews how we got where we are today, starting from ECMWF's first ensemble forecasts in 1992; it discusses the key characteristics of an ensemble system and the design of the ensembles operational today at ECMWF; it charts the evolution of ensemble forecast quality; and it describes the development of ensemble-based products to meet different user requirements. Finally, it looks to the future to highlight areas where further improvements can be made.

### Example forecasts

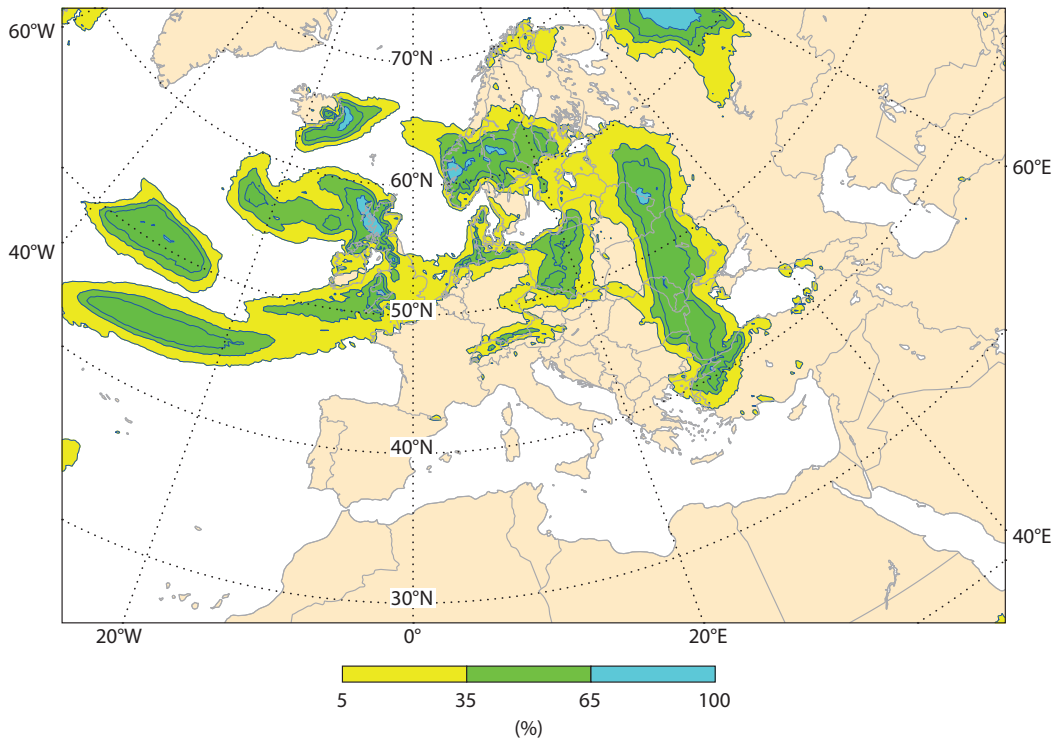
Figures 1 to 5 give an impression of the breadth of information that ensemble-based products can provide.

The Extreme-Forecast-Index (EFI) forecast issued on 27 July 2017 (Figure 1) identified southeastern Europe as an area that could be affected by anomalous precipitation and wind anomalies. In terms of precipitation, products such as the probabilistic forecast of rainfall in excess of 5 mm/day confirm this (Figure 2). The forecast identifies northwest Turkey as a region that could be affected by rainfall events. A forecaster interested in more local weather, say for Istanbul, could then click on the EFI map and generate a 15-day ENS meteogram for this city (Figure 3). This product shows the whole range of possible values that surface variables such as cloud cover, precipitation, wind and temperature can reach, and it contrasts them with average, climatological values. Figure 3 shows that, indeed, Istanbul is expected to experience anomalous precipitation on 28 and 29 July. It also indicates that the two-metre maximum temperature for 28 July will be very low, close to the climatological minimum for this time of the year.

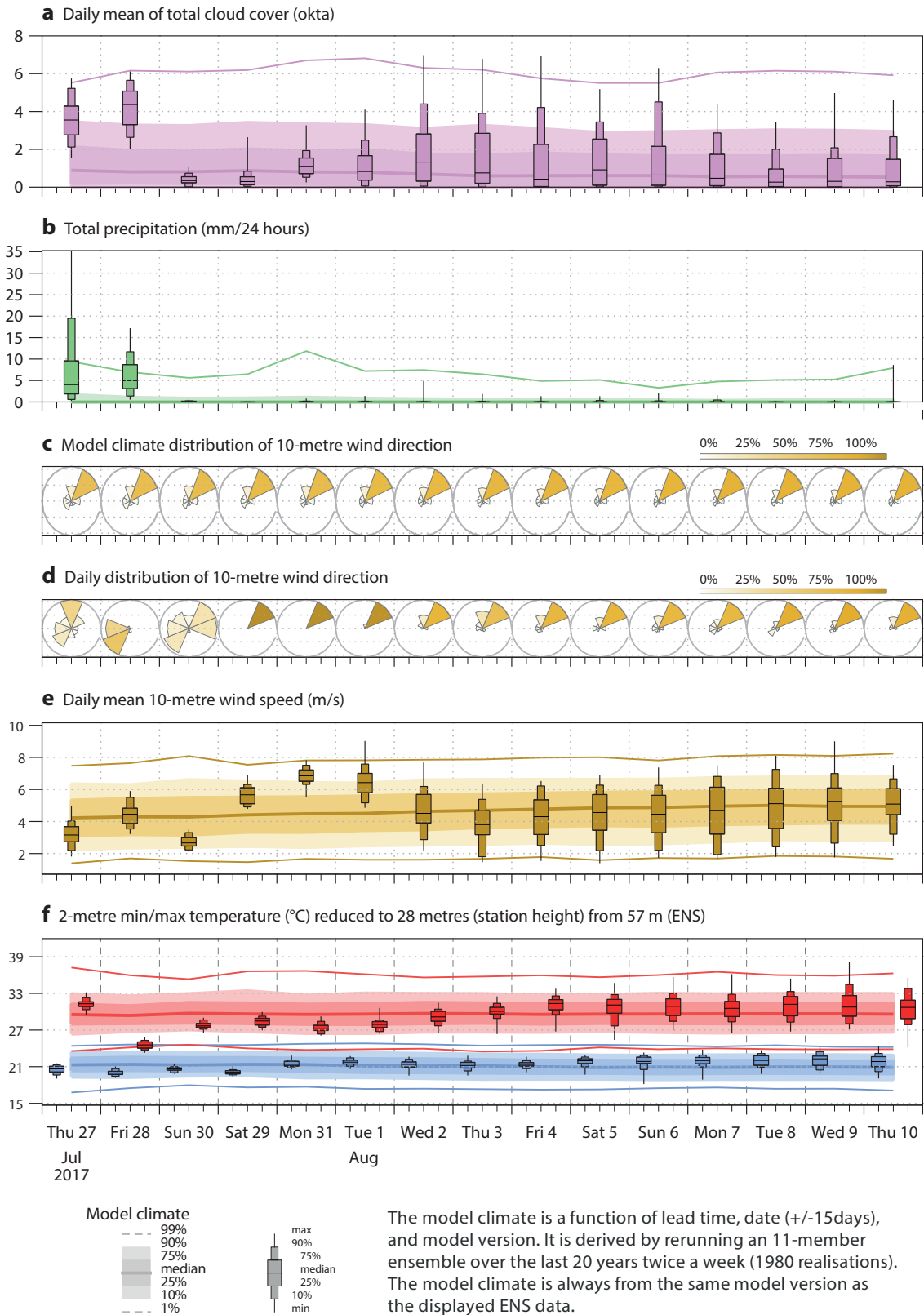
The meteogram for Istanbul also indicates that, after day four, the city will experience anomalous winds. To understand the synoptic-scale pattern associated with this event, we can look at the clusters for the 500 hPa geopotential height over Europe (Figure 4). They indicate that there is very little uncertainty over southeastern Europe (the three clusters have a very similar circulation in that region), with a low-pressure anomaly. Indeed, the small difference between the weather scenarios over Turkey for 31 July (Figure 4, right-hand column) is also reflected in the small spread in the wind forecast for Istanbul for that day (Figure 3, second diagram from the bottom).



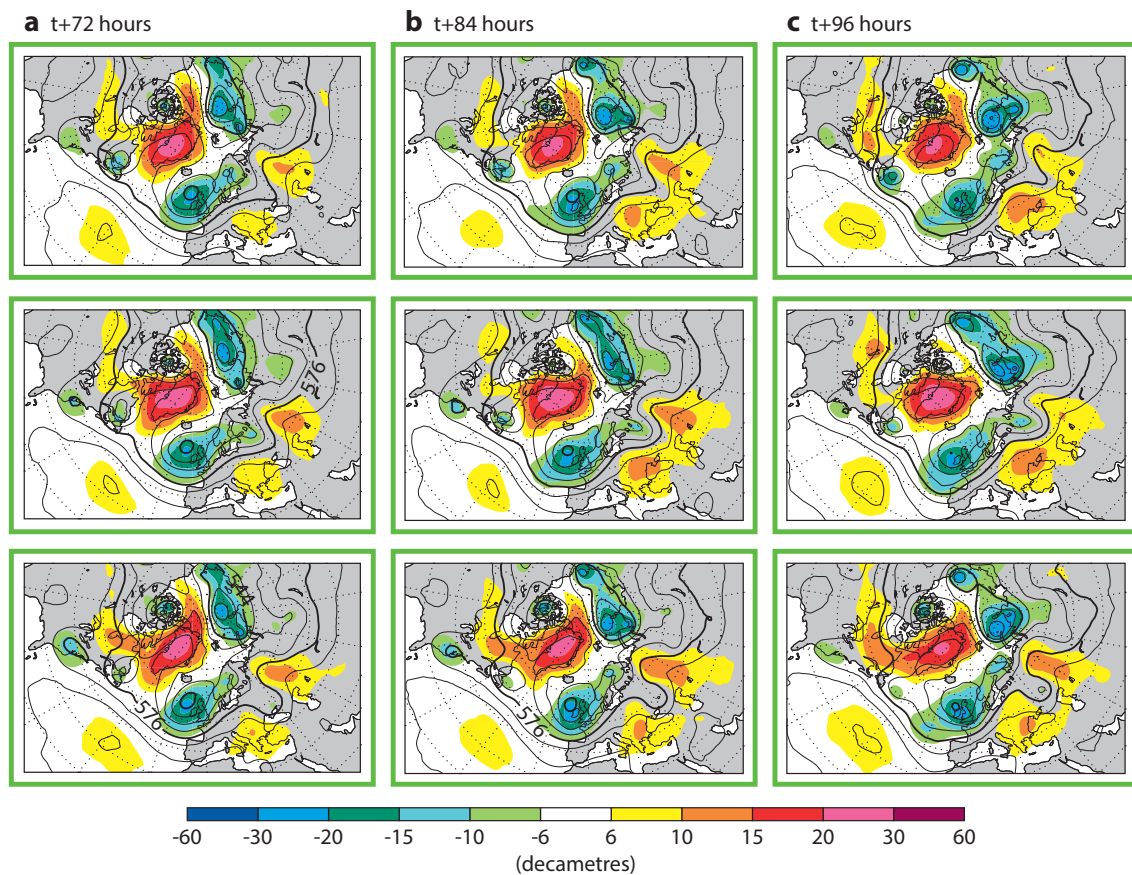
**Figure 1** ENS-based Extreme Forecast Index (EFI) forecast issued on 27 July 2017 at 00 UTC and valid for 29 July. The map shows areas where the ensemble forecast distribution differs substantially from the model climatological distribution for three variables: 2-metre temperature, 10-metre wind gusts and precipitation. The black contours show the ensemble-mean forecast for mean sea level pressure. For example, the EFI map shows that southeastern Europe is predicted to experience extreme wind gusts and precipitation.



**Figure 2** ENS-based probability of 24-hour precipitation in excess of 5 mm, issued on 27 July 2017 at 00 UTC and valid for 29 July.



**Figure 3** 15-day ENS meteogram for Istanbul, issued on 27 July 2017 at 00 UTC, showing (a) daily mean total cloud cover, (b) total daily precipitation, (c) model climate 10-metre wind direction distribution, (d) daily 10-metre wind direction distribution, (e) daily mean 10-metre wind speed, and (f) 2-metre minimum and maximum temperature. For each variable, the plot shows the ensemble distribution (box-and-whiskers in most cases) and the model climate so that users can assess the range of possible future weather scenarios and how this compares with the model climate.



**Figure 4** ENS-based 500 hPa geopotential height forecast clusters starting from 27 July 00 UTC for the European region, valid for (a) t+72 hours, (b) t+84 hours, and (c) t+96 hours. Cluster 1 (top row) includes 31 of the 51 ENS members, cluster 2 (middle row) 11 members and cluster 3 (bottom row) 9 members. Each panel shows the 500 hPa geopotential height (black contours) predicted by a representative member (RM) of the cluster and the anomaly (shading, computed with respect to the model climate). Each day, up to a maximum of five clusters are generated by an algorithm that defines in an objective way how many clusters are needed to represent the ensemble forecast distribution and identifies the RMs. The frame colour of each plot represents different climatological regimes. The green colour shown here stands for a negative North Atlantic Oscillation (NAO).

Ensembles also provide very valuable information for longer time ranges. An example is given in Figure 5, which shows a series of monthly ensemble-mean forecasts of the weekly-average two-metre temperature anomaly over Europe, predicted for the week of 19 to 26 of June 2016. The plots show that the ensemble was able to predict the heat wave that affected Europe up to four weeks ahead.

### 1992: the start of a paradigm shift

From the early days of numerical weather prediction (NWP), it was clear that there are some cases when forecast errors remain small even for long forecast ranges, and others when even a 1-day forecast is wrong. This operational experience was supported by scientific studies that pointed out that, due to the chaotic nature of the atmosphere, even small initial errors can grow very rapidly and affect forecast quality at a very short range.

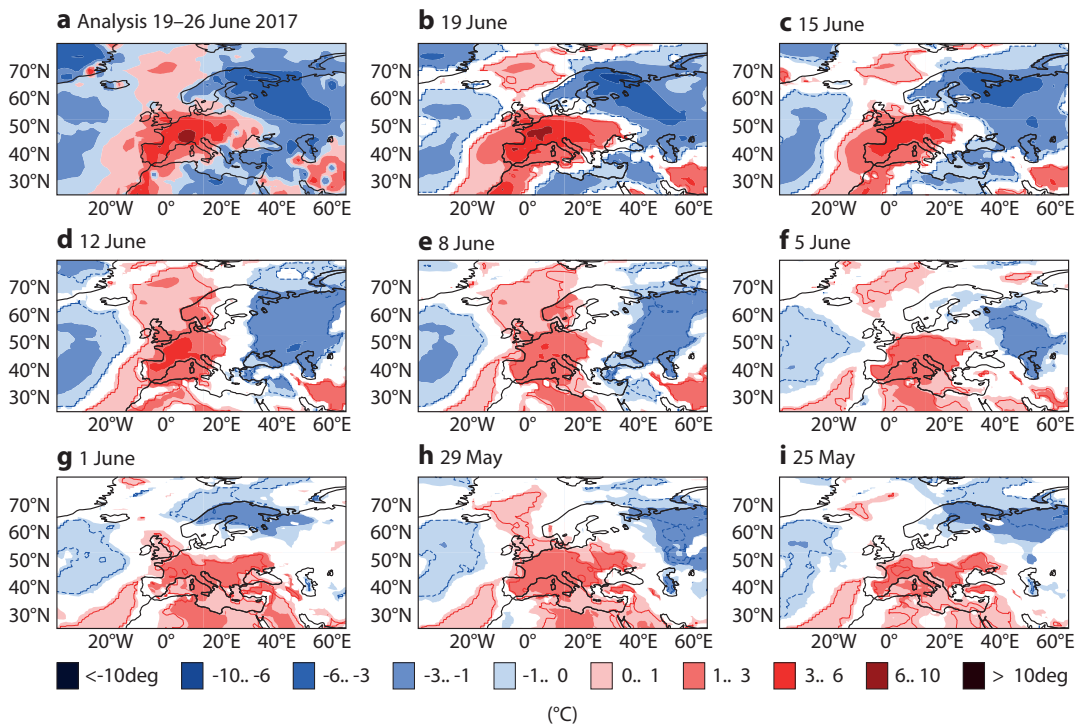
In the seventies and the eighties, we started investigating whether we could determine in advance, say when a forecast is issued, whether the future weather was easier or more difficult to predict than on average. In other words, we were looking for an objective method that could provide us with a level of forecast confidence. At that time different approaches were tested at the major NWP centres. It quickly became clear that the only feasible way to address this problem was to use ensembles. The main idea behind an ensemble approach is very simple: generate  $N$  forecasts, each of them designed to take into account possible uncertainties, and use the  $N$  forecasts to estimate the range of possible outcomes, and/or the most probable set of values, and/or the probability that temperature (or other variables) will be higher or lower than a certain value.

In the 1980s, different techniques were tried to develop reliable and accurate ensembles. These two adjectives, ‘reliable’ and ‘accurate’, are key, since they define whether an ensemble is capable of providing valuable information. An ensemble is reliable when there is, on average over many cases (say a season), a good correspondence between a forecast probability and the probability of occurrence. More precisely, in a reliable ensemble, if an event is predicted with an 80% probability, it occurs 80% of the time when such a prediction is made. An ensemble is accurate when the average error of the ensemble mean is small. In a reliable ensemble, the average spread is equal to the average error of the ensemble mean. An ensemble is sharp when the spread of the ensemble members is small (so event probabilities tend towards 0 or 100%). A good ensemble forecast is as sharp as possible while still being reliable.

In the 1980s in the US initial tests used lagged ensembles, which mixed forecasts started at different times and on different days, e.g. the nine forecasts issued every six hours over the past two days. ECMWF tried to generate ensembles starting all at the same time, but with initial conditions perturbed in a random way. Results indicated that the US method delivered forecasts with a reasonable quality for the medium forecast range, beyond about a week, but not for the shorter forecast range, since the ‘oldest’ forecasts were too old to be accurate. The ECMWF methods did not deliver good results since the random perturbations did not lead to very different forecasts: the forecasts remained too similar to provide valuable information on possible future scenarios.

The beginning of the 1990s saw the development and testing of more promising methods both at ECMWF and at the US National Centers for Environmental Predictions (NCEP). 1992 saw the implementation of the first two operational ensemble systems in those two places. In 1995 the Meteorological Service of Canada (MSC) followed suit and others a few years later, both at the global scale and for specific regions.

These implementations generated a paradigm shift in operational NWP from a deterministic approach, based on a single forecast, to a probabilistic one, in which ensembles are used to estimate the probability density function of initial and forecast states. Products such as the ones shown in Figures 1–5 would not exist if it was not for the development and operational implementation of these ensembles.



**Figure 5** Analysis and ENS-based monthly forecasts for weekly-average anomalies of 2-metre temperature between 19 and 26 June 2016, showing (a) the observed anomaly, defined by the average of ECMWF analyses over that period, and showing the forecasts starting on (b) 19 June, (c) 15 June, (d) 12 June, (e) 8 June, (f) 5 June, (g) 1 June, (h) 29 May and (i) 25 May. Each forecast panel shows the ensemble-mean 2-metre anomaly.

### Added value of ensemble forecasts

Today it is widely accepted that forecasts have to include uncertainty estimations, confidence indicators that allow forecasters to estimate how 'predictable' the future is. Short- and medium-range forecasts, monthly and seasonal forecasts, and even decadal forecasts and climate projections are today based on ensembles, so that the uncertainty associated with the forecast can be estimated. Furthermore, ensembles are widely used to provide an estimate of the initial state uncertainty, to estimate the analysis error more accurately.

Ensemble-based, probabilistic forecasts are more valuable than single forecasts. This is mainly due to the fact that they provide probabilities for different events to occur. In other words, ensembles provide users with more complete information about future weather scenarios. One way to measure such a difference is to apply simple cost-loss models using a measure called the Potential Economic Value (PEV) of a forecasting system (Richardson 2000). Another reason why ensemble-based, probabilistic forecasts are more valuable than single forecasts is that they provide more consistent (i.e. less changeable) successive forecasts. For example, consecutive ensemble-mean forecasts issued 24-hour apart and valid at the same time are generally found to jump less than corresponding single forecasts such as the high-resolution forecast or the ensemble control forecast (the ensemble member that starts from the 'most likely' initial state, defined by the unperturbed analysis). By using the whole ensemble, the unpredictable features are averaged out, and the predictable features (the signals) can be extracted.

### Design of medium-range global ensembles

Ensembles are designed to simulate the sources of forecast errors linked to initial condition and model uncertainties. 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. Initial condition uncertainties arise because observations are affected by observation errors and do not cover the whole globe with a uniform density 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.

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 (Buizza & Palmer, 1995). SVs provided a very good basis to define the initial perturbations of the ECMWF ensemble: compared to the random initial perturbations tried in the 1980s, they led to a very good growth rate in the spread of the ensemble, 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 being used, together with singular vectors (Buizza *et al.*, 2008). EDA-based perturbations were added to improve the simulation of initial errors linked to the characteristics of the observing system (observation errors, coverage, scalability...). Today SVs remain 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.

There are different ways to simulate initial and model uncertainties. For example, in the first version of NCEP's 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 errors that have a tendency to grow by virtue of perturbation dynamics (Toth & Kalnay, 1997). On the one hand, errors that have a tendency to stay constant or to decay will be reduced when detected by an assimilation scheme in the early part of the assimilation window. What remains of them will decay by the end of the assimilation window due to the dynamics of such perturbations. On the other hand, even if errors that have a tendency to grow are reduced by the assimilation system, what remains of them will have amplified by the end of the assimilation window.

In 1995 the ECMWF and the NCEP ensembles were followed by the Canadian ensemble. The Canadians adopted 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 to include a simulation of model uncertainties. It tried to include as many sources of error as possible.

Following the Canadian example, a stochastic scheme designed to simulate model uncertainties was introduced in the ECMWF ensemble in 1999 (Buizza *et al.*, 1999). Since then, many other operational ensembles have also included such schemes to simulate model uncertainties. Buizza (2014) provides a review of the main characteristics of the operational global ensembles available in the TIGGE database.

At present, as detailed by Palmer *et al.* (2009), four main approaches are followed in ensemble prediction to represent model uncertainties:

- A multi-model approach, where different models are used to construct ensembles; models can differ entirely or only in some components (e.g. in the convection scheme);
- A perturbed parameter approach, where all ensemble integrations are performed 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 back-scatter approach, where a Stochastic Kinetic Energy Backscatter scheme (SKEB) 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 currently used in the ECMWF ensemble, which is due to be switched off in 2018 since, in its current formulation, it does not deliver any significant benefits.

### Ensemble configurations

Two key aspects that define the characteristics of an ensemble are the methodology used to simulate initial uncertainties and the approach adopted to simulate model approximations. A third key characteristic of an ensemble is its resolution, both horizontal and vertical. A fourth aspect is the forecast length, and the fifth key aspect of an ensemble configuration is the number of ensemble members.

Theoretical work done in the 1970s and 1980s suggested that one needs at least about 10 members for a good ensemble-mean forecast, i.e. one which has enough members to average out the unpredictable scales or features. But are 10 members enough to get a good probability distribution forecast, and not just a good ensemble-mean? Results obtained in the 1990s and 2000s based on the comparison of ensemble sizes of up to a few hundred indicated that reliability and accuracy are very sensitive to ensemble size. On synoptic scales (scales of a few hundred kilometers), increasing the ensemble size from say 10 to about 50 was found to have a clear and detectable impact on ensemble forecast performance. Further increases beyond 50 have a smaller effect, on average, but can still have a detectable impact if one wants to predict rare events. When ensembles are used to estimate the analysis uncertainties, increasing the ensemble size to a few hundred members was found to bring clear benefits. Today most operational ensemble forecasts have between 20 and 50 members, while ensembles of analyses have up to about 300 members (although this number can be substantially lower depending on the computational cost of the analysis method).

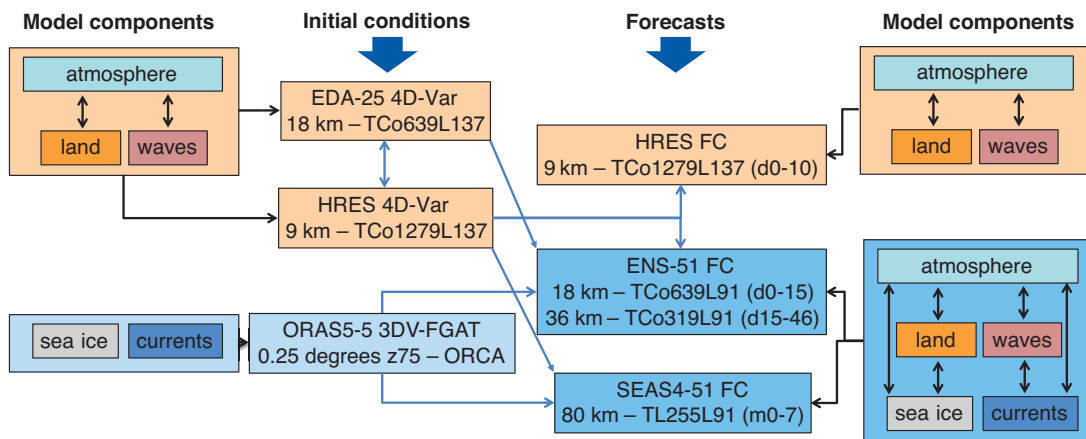
Resolution, forecast length and the number of ensemble members are key cost drivers of ensemble production. Given that we need to generate forecasts in a reasonable amount of time if we want them to be valuable (say about 1 hour), and that we have a finite amount of computing resources, compromises have to be made when an ensemble configuration is defined. Ideally, we would like to use as many members as possible and the highest resolution possible to be able to simulate also the finest scales so that we can provide detailed forecasts, including of severe weather. We would also like to extend the forecast length as much as possible to provide a bigger set of users with ensemble-based, probabilistic forecasts.



The compromise struck in ECMWF’s Integrated Forecasting System (IFS) is to use a number of different configurations for the ocean, the Ensemble of Data Assimilations, boundary condition generation, medium-range and monthly ensemble forecasts (ENS), and seasonal forecasts (see Table 1 for details).

	Forecast length	Resolution		Number of members	Re-forecasts (number of members; number of years)
		Horizontal	Vertical (number of layers)		
ORAS5 (Ocean reanalysis System-5)	--	100 km	75 layers	5 (once a day)	–
EDA (Ensemble of Data Assimilations)	--	18 km	137 (to 0.01 hPa)	25 (00, 12 UTC)	–
ENS for boundary-condition generation	6.5 days	18 km	91 (to 0.01 hPa)	51 (06, 18 UTC)	–
ENS for the medium range	15 days	18 km	91 (to 0.01 hPa)	51 (00, 12 UTC)	Yes (22/week; 20 years)
ENS for the monthly range	46 days	36 km	91 (to 0.01 hPa)	51 (00 UTC on Mon and Thu)	Yes (22/week; 20 years)
SEAS4 (seasonal-range ensemble)	7 and 13 months	80 km	91 (to 0.01 hPa)	51 (1st of each month)	Yes (15/month; 30 years)

**Table 1** Key characteristics of the ECMWF ensembles. All ensembles simulate initial and model uncertainties.



Atmosphere grids: TCo (cubic-octahedral Gaussian reduced grid) or TL (Gaussian linear grid)  
 Ocean grid: ORCA (tri-polar grid)

**Figure 6** Diagram showing the links between the three ECMWF ensembles (EDA, ENS and SEAS4) and Earth system components used to generate the initial conditions and the forecasts.

Figure 6 illustrates how the three ECMWF ensembles and the high-resolution analysis and forecast are linked together:

- The 25-member EDA and the high-resolution analysis (both using 4D-Var) are used to generate the initial conditions of the two coupled forecast ensembles, ENS and SEAS4;
- The 5-member ORAS5 (Ocean Reanalysis and Analysis, version S5) is used to initialise the dynamic ocean and sea-ice components of the two coupled ensembles, ENS and SEAS4;
- The high-resolution analysis is used to generate the initial conditions of the single, high-resolution forecast (HRES).

Figure 6 also illustrates the Earth-system components included in the initial condition and forecast models:

- The EDA, the high-resolution analysis and the HRES use the ECMWF land and atmosphere model and the ECMWF wave model (ECWAM);
- The ocean analysis ORAS5 uses the NEMO (Nucleus of European Modelling of the Ocean) ocean model (see <https://www.nemo-ocean.eu/>) and the LIM2 (Louvain-la-Neuve) sea-ice model (see <http://www.elic.ucl.ac.be/repomodx/lim/>);
- The two coupled ensembles ENS and SEAS4 use all model components: IFS+ECWAM+NEMO+LIM2.

The high-resolution forecast is due to be coupled with the ocean and sea-ice models (NEMO+LIM2) from the beginning of 2018, after the implementation of IFS Cycle 45r1.

### Ensembles for sub-seasonal and seasonal timescales

Since the beginning of the 2000s, ensembles have also been used to generate monthly and seasonal forecasts. These extended-range ensembles are global and have a coarser resolution than the medium-range ensembles to limit production costs (see Table 1). Since extracting predictable signals for the extended range is very difficult, these ensembles have been complemented by re-forecast suites, which are smaller-size ensembles with the same configuration as the operational ensembles (apart from their size) generated for the last few decades. After the medium-range and the monthly ensembles were joined in 2008, with the implementation of the VAREPS approach, we have been able to exploit the re-forecasts to design new and better products for the medium range too.

Today the re-forecasts are used to estimate the ensemble characteristics (reliability and accuracy, and model biases), and they help to generate forecast products across the whole forecast range, for example products such as the EFI (shown in Figure 1), the 15-day meteograms (Figure 3) and some of the extended-range probabilistic forecasts (e.g. the ones shown in Figure 5) by providing a model climatology.

ECMWF is upgrading the seasonal ensemble to SEAS5, which will have the same resolution as the ENS monthly extension. SEAS5 is due to become operational in November 2017.

### Ensembles of analyses and reanalyses

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 each of the ensemble members are defined by one of the members of the EnKF. The EnKF has been providing the Meteorological Service of Canada with information about uncertainties in the analysis.

At ECMWF and Météo-France, we started producing an Ensemble of Data Assimilations in 2008. We run an ensemble of  $N$  separate data assimilation procedures, each using perturbed observations and a model uncertainty scheme. Observations are perturbed to simulate the fact that observations are not perfect due to observation errors, and to take into account observation representativeness errors. Model uncertainties are simulated to take into account the fact that the models used to define the analysis are not perfect.

Since 2008, the ECMWF EDA has been 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 ensemble reliability and accuracy in the short forecast range over the extratropics, and for the whole forecast range over the tropics.

Since 2002, ECMWF has been producing a five-member ensemble of ocean analyses and reanalyses to initialise the ocean component of coupled ensembles. The medium-range/monthly ENS started using the ocean ensemble in 2008 from day 10, when it was merged with the monthly ensemble (see Box A). Since 2013, ENS has been using the ocean ensemble ORAS5 from initial time.

## ENS configuration changes

A

This list includes the main changes made to the medium-range/monthly ensemble (ENS) since its first day of real-time production and dissemination on 19 December 1992:

- **Dec 1992:** the ensemble starts running three days a week (Fri-Sat-Sun, at 00 UTC); initial uncertainties are simulated using only the initial-time SVs with a T21L19 resolution, computed with a 36-hour optimisation time interval, over the whole globe; only the initial-time SVs are used, and the initial perturbations are symmetric; the forecast resolution is T63L19 (~320 km); forecasts are run up to 10 days; ENS includes 33 members; there is no simulation of model uncertainties, no coupling to an ocean/sea-ice model, and no re-forecast suite;
- **Feb 1993:** to address the fact that SVs were concentrating mainly in the southern hemisphere (SH), the Local Projector Operator (LPO) was introduced, to allow SVs to be located in the northern hemisphere (NH); this had a major impact on ensemble reliability over the NH;
- **Aug 1994:** to improve the ensemble spread, the optimisation time interval (OTI) of the SV computation was increased from 36 to 48 hours; this improved the perturbation growth also beyond the OTI; from 1 May 1994, ENS forecasts were generated every day, once a day, at 00 UTC;
- **Mar 1995:** the horizontal resolution of the SVs was increased to T42; this improved perturbation growth and thus ensemble reliability;
- **Mar 1996:** a second set of SVs was introduced, targeted to grow over the SH; this had a major impact on ensemble reliability over the SH;
- **Dec 1996:** the resolution of ENS was increased to TL159L31 (~120 km), and the number of members was increased from 33 to 51;
- **Mar 1998:** a second set of SVs, called evolved SVs, that grow during the two days before the initial date, were added to the initial-time SVs; the evolved SVs simulated the effect of errors growing during the data-assimilation period; their addition improved ensemble reliability (spread) especially in the short range;
- **Oct 1998:** the stochastic model error scheme (SPPT) was introduced to simulate the effect of model uncertainties linked to physical parameterization; this had a large impact on ensemble reliability (spread) over the whole forecast range, and especially over the tropical region;
- **Oct 1999:** vertical resolution was increased from 31 to 40 levels;
- **Nov 2000:** ENS horizontal resolution was increased from TL159 to TL255 (~80 km);
- **Jan 2002:** SVs targeted to grow over the tropical region, in areas where tropical depressions were identified, were added; they led to improved spread over the tropical region, and especially in cases of tropical storms;
- **Sep 2004:** the sampling strategy applied during the generation of ENS initial perturbations was changed to Gaussian sampling;
- **Jun 2005:** the Gaussian sampling method was revised;
- **Feb 2006:** ENS horizontal resolution was increased from TL255L40 to TL399L62 (~60 km);
- **Sep 2006:** ENS was extended to 15 days, with the use of variable resolution (VAREPS), whereby the forecast resolution was truncated at day 10 from TL399 to TL255;
- **Mar 2008:** the medium-range ensemble (ENS) and the monthly ensemble were merged, using the VAREPS technique; ENS was run to 32 days once a week (Mon at 00 UTC); ENS was then coupled to the dynamical ocean model HOPE from forecast day 10; the ENS re-forecast suite with a 5-member ensemble run once a week for the past 18 years was introduced;
- **Sep 2009:** the stochastic model error scheme was revised;
- **Jan 2010:** the horizontal resolution was increased from TL319 to TL639 (~35 km) in the first 10 days, and from TL255 to TL319 (~70 km) from day 10 to day 32;
- **Jun 2010:** a new set of initial perturbations, generated using the 10-member Ensemble of Data Assimilations (EDA), was introduced in ENS; the EDA-based perturbations improved the simulation of the perturbations linked to the data-assimilation cycle, and replaced the evolved SVs; this led to improvements in ensemble reliability (spread), especially in the short forecast range and over the tropical region;
- **Nov 2010:** a second scheme, the stochastic backscatter (SB) scheme, was introduced to simulate model error;
- **Nov 2011:** a new ocean model was introduced: NEMO with a 1-degree resolution (~100 km) replaced HOPE; the ENS extension to 32 days started being run twice a week (Mon and Thu at 00 UTC);
- **Jun 2012:** the EDA-based perturbations were revised to include perturbations in the surface fields, and the re-forecast suite was enlarged to cover the past 20 years;
- **Nov 2013:** the vertical resolution was increased from 62 to 91 vertical levels, and the coupling to the ocean model was moved from day 10 to day 0; this led to major improvements in the prediction of phenomena over the tropics, such as the Madden-Julian Oscillation (MJO);
- **May 2015:** forecast length was extended from 32 to 46 days, and the re-forecast suite was enlarged to include two 11-member ensembles run every week (on Mon and Thu), covering the past 20 years;
- **Mar 2016:** the horizontal resolution was increased to TCo639L91 (cubic-octahedral grid; ~18 km) up to day 15, and to TCo319L91 (~36 km) from day 15 to day 46;
- **Nov 2016:** the ocean model resolution was increased from 1 degree to 0.25 degrees (~25 km), and the number of vertical layers from 42 to 75; the interactive sea-ice model LIM2 was introduced.

### Changes in configuration over 25 years

Since its inception in 1992, the medium-range ensemble has changed configuration several times. A chronology of the main configuration changes is provided in Box A. It is interesting to compare the ENS configuration implemented in operations in December 1992 with that of 2016.

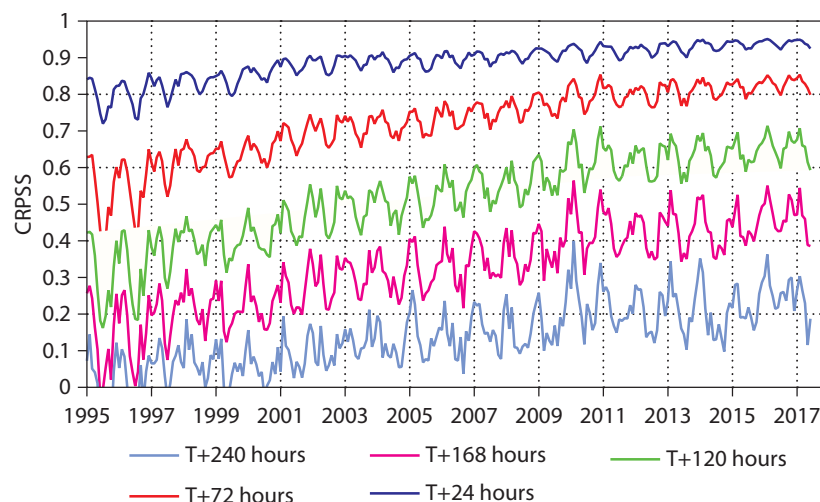
In terms of the key cost drivers of ensemble production, the main changes are:

- the horizontal resolution has increased by a factor of 20, from about 320 km to about 16 km;
- the vertical resolution has increased by a factor of almost 5, from 19 to 91 vertical levels;
- the forecast length has been extended from 10 to 46 days;
- the number of ensemble members has increased from 33 to 51;
- the frequency of ENS forecast production has increased: in 1992 we produced 99 ensemble forecasts each week (3x33), while today we produce 1428 ensemble forecasts each week up to 6.5 days (4x51x7); of these 1428 forecasts, 714 are extended up to 15 days (2x51x7); of these 714 forecasts, 102 are extended up to 46 days (2x51);
- today we also produce ensemble re-forecasts: every week, we generate 440 ensemble forecasts up to 46 days (2x11x20).

### Evolution of ensemble forecast quality

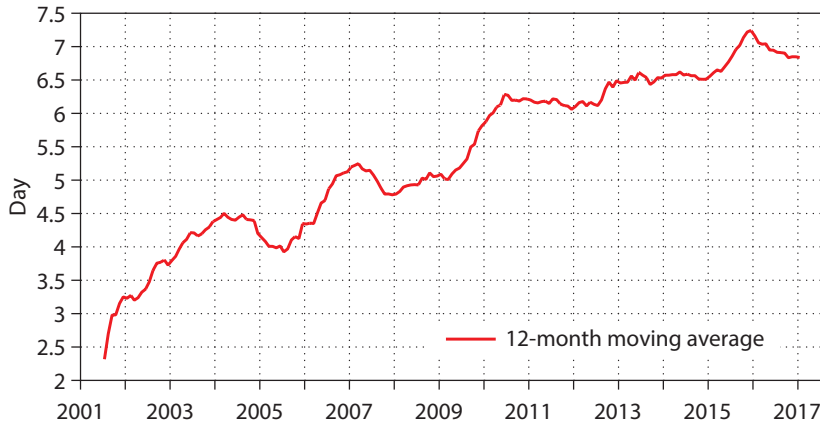
Thanks to model upgrades, improvements in the data assimilation system, the use of more observations, and the ENS configuration changes discussed above, the ENS performance has increased substantially during the past 25 years.

Figure 7 shows the time evolution of the skill of ENS forecasts for 500 hPa geopotential height over the northern hemisphere from 1 January 1995 to today. Skill is measured by the Continuous Ranked Probability Skill Score (CRPSS), which compares the Continuous Ranked Probability Score (CRPS) of ensemble forecasts with that of a reference forecast, such as climatology. CRPS measures how close ensemble distributions are to observed values. The CRPS for a deterministic forecast is equal to the mean absolute error. CRPSS has a value of 1 for a perfect forecast, and is zero for a forecast that has the same skill as a statistical forecast based on climatology. Figure 7 shows that for 500 hPa geopotential height, a variable that describes the large scales in the free atmosphere, ENS forecasts have improved by about 1.5 days per decade. For example, today's 5-day forecasts (green line) are as skilful as 3-day forecasts (red line) were in 2001. This represents a predictability gain of about 2 days over a 16-year period.



**Figure 7** Time evolution, from 1 January 1995 to today, of the Continuous Ranked Probability Skill Score (CRPSS) of ENS forecasts of 500 hPa geopotential height over the northern hemisphere, for lead times of 24 hours, 72 hours, 120 hours, 168 hours and 240 hours. Forecasts are verified against operational analyses. The more or less regular pattern of peaks and troughs in each line stems from differences in predictability related to the seasons: winter weather tends to be more predictable than summer weather.

If we look closer to the surface, results are even more striking. For precipitation, for example, Figure 8 shows that between 2002 and 2017 the lead time when the CRPSS dropped below 0.1 increased from about forecast day 3 to about forecast day 7, equating to a predictability gain of about 4 days over a 15-year period. Similar improvements are found for other variables and other regions (e.g. Europe, not shown).



**Figure 8** Time evolution of the forecast lead time when the CRPSS for the prediction of 24-hour accumulated precipitation drops below 0.1. Forecasts are verified against observations at SYNOP network weather stations in the extratropics.

Looking at longer forecast lead times, it is worth remembering that in 2006 the medium-range ensemble was extended to 15 days, and that in 2008 ENS was joined to the monthly ensemble and extended to 32 days. In 2015 it was further extended to 46 days. These extensions were justified by the fact that forecasts for these extended lead times had been improving as well. Clearly, for these forecast ranges only spatially large scales and time-averaged fields can be predicted with a certain level of skill. Results documented in *Buizza & Leutbecher (2015)* and *Vitart et al. (2014)* show that for these large-scale, low-frequency phenomena the forecast skill horizon has been extended to several weeks. The reader is also referred to *Vitart et al. (2014)* for a comprehensive overview of how the skill of ECMWF monthly forecasts evolved during the preceding 15 years.

### Product development

Key to the successful use of ensemble forecasts is the ability to extract and communicate the information that is relevant to each user's decision-making process.

Alongside the development of the ENS perturbation methodologies, there has been substantial work and progress in the development of ensemble-based forecast products to address a range of different user requirements and enable forecasters to extract the appropriate information from the ENS.

When the ensemble was first introduced, the number of ENS-based products was limited. 'Stamp' maps showed each ENS member at forecast day 7, allowing the user to quickly assess by eye the range of possible weather states. This was complemented by cluster products that objectively grouped the set of ENS members into a small number of different scenarios that showed the different forecast evolution for 5 to 7 days ahead over Europe. For a small set of pre-defined locations, 'plume diagrams' showed the evolution of a small number of surface parameters through the forecast range. These products were issued to users by fax.

Nowadays users have access to a wide range of ENS data and products that process and present the ensemble information in different ways according to the needs of the user. The focus at ECMWF is to provide generic products that will be useful to assist operational weather forecasters. Many users complement the ECMWF products by doing their own post-processing to generate specific products tailored to their individual needs.

### **Extreme Forecast Index**

The Extreme Forecast Index (EFI) was specifically designed to alert forecasters to occasions of potentially extreme weather (Lalurette, 2003; Zsoter, 2006). The EFI compares the current ensemble forecast to the model climate distribution (generated by running a large set of re-forecasts over the last 20 years). It highlights areas where the current ENS forecasts are showing an enhanced likelihood of unusual weather. A large EFI indicates that the weather is likely to be extreme in the context of what can occur locally.

The EFI is one of the most popular ensemble products with forecasters. It can be especially useful in forecasting around the world, when forecasters may not have detailed knowledge of the regional climate. Since the EFI focuses on anomalies relative to the local climate, it is especially relevant for impact-based forecasting, where local extremes (or return periods) are more relevant than fixed event thresholds.

### **Storm tracks**

Also relevant for severe weather forecasting are specific sets of products for extra-tropical and tropical cyclones. In both cases the cyclones are tracked in each ENS member and a range of products show the evolution of certain features along the forecast track, such as central pressure and maximum wind associated with the system. See the separate article on the hurricanes Harvey and Irma in this Newsletter for examples.

Both sets of products are designed to show information about the tracks and intensities of storms in the forecast and to help the forecasters quickly answer questions such as where and when severe storms will occur; how intense they will be; and where there may be a risk of a severe tropical cyclone in the coming days or weeks.

Many ENS products are available on the ECMWF website and many are now interactive, allowing the user to for example click on a location of high EFI to examine the details of the full ENS distribution at that location. ecCharts is an interactive web application that enables users to explore the ECMWF forecasts in even more detail. It allows them to zoom in on any area of interest; to select and overlay different forecast parameters; to compare and combine HRES and ENS forecasts; to compute probabilities for specific events of their own choosing (for example by selecting a precipitation threshold and time interval); and even to define combined events (such as the probability of both heavy precipitation and extreme wind).

There are several other products not mentioned here which highlight different aspects of the ensemble distribution, for use by forecasters in different situations. Each is designed to extract the most relevant information from the ensemble and to present it to the forecaster as clearly as possible, so the forecaster can focus on their job without having to spend time themselves trying to process the huge amount of information in the ensemble.

### **A look to the future**

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

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

The first trend is justified by results obtained in the past two decades that have shown 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.

The second trend is partly motivated by scientific developments and partly by technical requirements. From a scientific point of view, there is evidence that processes that were thought to be relevant for the extended range are also relevant for the short range. An example is the introduction of a dynamic ocean in ECMWF ensembles. We started using a coupled ocean–land–atmosphere model for the seasonal and monthly timescales. We also introduced it in the medium-range ensemble once we realised that it could help to improve 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 maintenance and the implementation of upgrades. Furthermore, it facilitates the diagnostics and evaluation of a model version, since tests carried out over different timescales can help to identify undesirable behaviour that could lead to forecast errors.

The third trend stems from the need to better resolve the smaller scales and their interaction with slightly less small scales, and so on. All scales from the microphysics within individual clouds to large-scale weather systems covering thousands of kilometres are relevant in weather prediction, and errors propagate from the smallest to the larger scales. If we consider the current operational ensemble, we should not forget that even if it uses a grid spacing of 18 km in the first 15 days, it can actually resolve in a realistic way only scales that are about 5 to 6 times the grid spacing. This is because the scales closest to the model grid spacing are not simulated in an accurate way. Thus, today the ECMWF global ensembles (EDA and ENS) have an effective resolution of about 100 km. Even the highest-resolution limited-area ensembles, such as the ones in operation at Météo-France and the UK Met Office, have a resolution of about 2 km, which corresponds to an effective resolution of about 10 km.

ECMWF's ten-year Strategy adopted in 2016 sets ambitious goals in line with these requirements. These include the introduction of a 5 km global ensemble by 2025. Even this will, however, not be the end of the road. If we want to be able to predict weather events such as intense wind storms or heavy precipitation at the scales at which they occur, it will be essential for model resolution to be increased to a few hundred metres for limited-area models and, in the long term, possibly to even finer resolutions than 5 km globally.

### **In conclusion ... ensembles are the way forward!**

The past 25 years have seen major advances in ensemble prediction, both in the way ensembles are generated and in ensemble products. Forecasts have become more accurate and reliable thanks to improvements in the initial conditions (i.e. in the use of observations and in the data assimilation system used to generate them); in the quality of forecast models; and in ensemble configurations. The introduction of additional relevant Earth system processes, such as the coupling to dynamic ocean and sea-ice models, has also led to improvements, and it has helped us to 'tame' the butterfly effect (*Buizza et al.*, 2015). The use of re-forecasts has made it possible to extract more meaningful signals from the raw forecast data.

We are confident that the future will see the use of ensembles also in areas where they are not yet used. Ensemble reliability and accuracy will continue to improve as a result of further advances in models, data assimilation methods, and the schemes used to simulate the initial and model uncertainties. Resolution will be increased to better simulate small-scale processes that are not currently resolved and to capture their important interactions with larger-scale processes. Ensembles of analyses and forecasts will be linked closer together to improve their performance. Physical processes that are not yet included in the models but that are relevant for weather prediction will be included, to make the forecasts more and more realistic.

The time is right: ensembles are the way forward!

**Further reading**

**Barkmeijer, J., R. Buizza, E. Källén, F. Molteni, R. Mureau, T. Palmer, S. Tibaldi & J. Tribbia**, 2013: 20 years of ensemble prediction at ECMWF. *ECMWF Newsletter No. 134*, 16–32.

**Buizza, R. & T.N. Palmer**, 1995: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 9, 1434–1456.

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

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

**Buizza, R.**, 2014: The TIGGE medium-range, global ensembles. *ECMWF Technical Memorandum No. 739*.

**Buizza, R., & M. Leutbecher**, 2015: The Forecast Skill Horizon. *Q.J.R. Meteorol. Soc.*, **141**, Issue 693, Part B, 3366–3382.

**Buizza, R., M. Leutbecher & A. Thorpe**, 2015: Living with the butterfly effect: a seamless view of predictability. *ECMWF Newsletter No. 145*, 18–23.

**Houtekamer, P.L., L. Lefraive & J. Derome**, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

**Lalurette, F.** 2003. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Q. J. R. Meteorol. Soc.*, **129**, 3037–3057.

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

**Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer & A. Weisheimer**, 2009: Stochastic parametrization and model uncertainty. *ECMWF Technical Memorandum No. 598*.

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

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

**Vitart, F., G. Balsamo, R. Buizza, L. Ferranti, S. Keeley, L. Magnusson, F. Molteni & A. Weisheimer**, 2014: Sub-seasonal predictions. *ECMWF Technical Memorandum No. 738*.

**Zsótér, E.** 2006: Recent developments in extreme weather forecasting. *ECMWF Newsletter No. 107*, 8–17.

© Copyright 2017

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

The content of this Newsletter is available for use under a Creative Commons Attribution-Non-Commercial-No-Derivatives-4.0-Unported Licence. See the terms at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

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