Chapter 1

Ensemble Forecasting

Alfons Callado, Pau Escribà,
José Antonio García-Moya, Jesús Montero,
Carlos Santos, Daniel Santos-Muñoz and
Juan Simarro

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55699

1. Introduction

The atmospheric movements can be described by non-linear differential equations that unfortunately have no analytical solution. The numerical methods to solve the atmospheric non-linear differential equations have been developed in different stages. During the 50s, Charney, Fjørtoft and von Neumann (1950) made a 24-hour forecast of 500 hPa geopotential height using a bidimensional quasi-geostrophic equation. After that, in 1956, Philips showed the close relation between cyclone dynamics and the global circulation using a 2-layer model.

At the beginning of the 70s, the global circulation models emerged (Lynch, 2006). These models are based on a set of non-linear differential equations, which are used to approximate the global atmospheric flow, called primitive equations. During this stage the full primitive equations were implemented without any quasi-geostrophic approximation (Williamson, 2007).

During the 80s, the regional and mesoscale numerical models appeared (Athens &Warner, 1978; Mesinger et al., 1988). The evolution of the models is a direct consequence of the increase of computer resources, and the improvement in observational networks and assimilation methods. This evolution has extended the knowledge on the dynamics and atmospheric microphysical processes.

The last period of the numerical weather prediction was initiated in the 90s. The atmosphere-ocean and atmosphere-ocean-soil coupled models, and the spatio-temporal high resolution models allowed the development of analysis and diagnostic techniques for the weather forecasting (Mechoso & Arakawa, 2003).

Until then, the numerical prediction models' philosophy was based on the deterministic atmospheric behavior. That means, given an atmospheric initial state its evolution can be numerically predicted to give a unique final state. Consequently the efforts of the scientific community were focused on producing the most accurate prediction (Tracton & Kalnay, 1993).

Nevertheless, the formulation of models requires approximations due to unknown variables or known process that cannot be explicitly resolved using the spatio-temporal resolutions a model works with. These processes must be parameterized and this fact generates errors associated with the parameterization used in the model. Although the model could perfectly simulate all the atmospheric processes, it would be impossible to determine a realistic initial state description for all resolutions and in all places using the available observational data (Daley, 1991). Lorenz (1963) showed that small variations on the model initial conditions do not produce a single final solution, but a set of different possible solutions. That is why the predictability of the future atmospheric states is limited in time: the initial condition errors are amplified as the forecast period grows (Lorenz, 1963, 1969).

The traditional deterministic approach gave way to a new paradigm, with richer information than a single solution of the future atmospheric state. The new paradigm includes quantitative information about the uncertainty of the predictive process. The atmospheric non-linear behavior, consequently chaotic, must be treated now in a probabilistic way (Lorenz, 1963).

The improvement of numerical models will permit a better characterization of the atmospheric processes but the models will always have some limitations related to the scales of the simulated processes and the approximations made to solve numerically the equations. Another limitation of the numerical forecasting methods is the lack of observational data with high enough resolution to properly describe the initial state.

Nowadays the observational methods, the assimilation strategies and the own characteristics of the numerical models have inherent limitations that generate uncertainty in the estimation of the possible future atmospheric states. The uncertainty is amplified when the forecast period grows and when the resolution increases. Thus, the probabilistic approach seems an ideal strategy to characterize forecast uncertainty.

The atmospheric state cannot be exactly known. The analysis data always contain an error that only can be estimated. The inaccurate determination of the real atmospheric state drives to the existence of a great number of initial conditions compatible with it. A single model only provides a single solution of the future atmospheric state. The generation of multiple forecasts starting from slightly different but equally-probable initial conditions can characterize the uncertainty of the prediction (Leith, 1974).

The generation of equally probable forecasts starting from multiple realistic initial conditions introduces the probabilistic forecasting concept. A practical approximation to probabilistic forecasting based on meteorological models is the so called ensemble forecasting. The Ensemble Prediction Systems (EPS) are used operationally in several

http://dx.doi.org/10.5772/55699

weather and climate prediction centres worldwide. The European Centre for Medium-Range Weather Forecasts (ECMWF; Molteni et al., 1996) or the Meteorological Service of Canada (Pellerin et al., 2003), among others, produce routinely ensemble predictions. These predictions have been demonstrated to be extremely useful on decision making process.

The EPS is a tool for estimating the time evolution of the probability density function viewed as an ensemble of individual selected atmospheric states. Each of these different states is physically plausible. The spread of the states is representative of the prediction error (Toth & Kalnay, 1997).

Several techniques for constructing the ensemble have been developed and applied. One of the first methods proposed for generating an ensemble of initial states is the random Monte Carlo statistical methodology. It was proposed by Leith (1974), Hollingsworth (1980) and Mullen and Baumhefner (1989), among others.

Perturbative methods that depend on the atmospheric flow are also used. These strategies are based on the generation of perturbations in the subspaces where the initial condition errors grow faster. The breeding vectors (Toth & Kalnay, 1993, 1997) or the singular vectors (Buizza & Palmer, 1995; Buizza, 1997; Hamill et al., 2000) are remarkable examples.

There are other perturbative methods that consider the model sub-grid scale errors by means of varying model physical parameterizations (Stensrud et al., 1998; Houtekamer & Mitchell, 1998; Andersson et al., 1998) or using stochastic physics (Buizza et al., 1999).

The combination of multiple model integrations initialized by multiple initial conditions determined by different analysis cycles is another strategy to generate ensembles. Using different assimilation techniques allows characterizing the uncertainties associated to the initial condition and the uncertainty associated to each model (Hou et al., 2001; Palmer et al., 2004). Finally, taking different global models as different initial conditions has been found to provide better performance than any single model system (Kalnay & Ham, 1989; Wobus & Kalnay, 1995; Krishnamurti et al., 1999; Evans et al., 2000).

The technique based on the use of multiple limited area models (LAM) and multiple initial conditions coming from several global models combined with advanced statistical post-processing techniques (Gneiting & Raftery, 2005a) has been tested in the National Centres for Environmental Prediction (NCEP; Hamill & Colucci, 1997, 1998; Stensrud et al., 1999; Du and Tracton, 2001, Wandishin et al., 2001) during the Storm and Mesoscale Ensemble Experiment (SAMEX; Hou et al., 2001). Such probabilistic predictions have also been generated over the Pacific Northwest coast (Grimit & Mass, 2002) and over the Northeast coast (Jones et al., 2007) of the United States.

The combination of multiple models and multiple analyses is part of the operational suite of NCEP (Du & Tracton, 2001) and the basic idea of the short-range EPS of Washington University (Grimit & Mass, 2002) and the Agencia Estatal de Meteorología (AEMET; García-Moya et al., 2011).

2. Atmosphere as a chaotic system

2.1. Lorenz and non-linearity

Two basic properties can dynamically characterize a chaotic system: the sensitivity to initial conditions and the topologically mixing. Sensitivity to initial conditions implies that infinitesimal changes in the system initial trajectory can lead to big changes in its final trajectory. The Lyapunov exponent (Lyapunov, 1992) gives a measure to this sensitivity to initial conditions as it quantifies the rate of separation of infinitesimally close trajectories. Generally it cannot be calculated analytically and one must use numerical techniques. In Krishnamurthy (1993) it is described how to calculate the Lyapunov exponents of a simple system. The meaning of topological mixing is that the temporal evolution of meteorological quantities in any given region of its phase space will eventually overlap with those of any other given region. This second property is necessary to distinguish between simple unstable systems and chaotic systems.

The classical example provided by Lorenz (1963) is instructive. For this reason we use it in this section, to show briefly some concepts of Chaos Theory. It comes from a simplified model of fluid convection. It consists of a dynamical system with only three degrees of freedom, but it exhibits most of the properties of other more complex chaotic systems. It is forced and dissipative (in contrast to Hamiltonian systems which conserve total energy), non-linear (as its equations contain products of dependent variables) and autonomous (all the coefficients are time independent). The Lorenz (1963) equations are:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= xy - bz \end{aligned} \tag{1}$$

where, in this simplified model, $x(t)$ is proportional to the intensity of convection, $y(t)$ proportional to the maximum temperature difference between up and downward moving fluid portions and $z(t)$ is proportional to the stratification change due to convection. All variables are dimensionless, including time. The solution $\{x(t), y(t), z(t)\}$ is unique provided that the initial conditions $\{x_0, y_0, z_0\}$ are given at time $t = 0$. This means that the system is theoretically deterministic (given a perfect representation of the initial values or the dependent variables and a perfect integration of the non-linear system). The parameters $\{\sigma, r, b\}$ are constant within the time integration and different values provide different solutions thus creating a family of solutions of the dynamical system. Lorenz (1963) chose the values $\sigma = 10$, $r = 28$ and $b = 8/3$ which led to a chaotic solution of the system that is sensitive to small changes in the initial conditions and topological mixing. The dimension of the phase space is equal to the number of dependent variables (three in this case) whereas the dimension of the subspace

http://dx.doi.org/10.5772/55699

reached by a given solution can be smaller as is the case of the Lorenz system. This behaviour can be demonstrated from the divergence of the flow:

$$\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} = -(\sigma + r + b) \tag{2}$$

Which means that an original volume in the phase space V contracts in time to $Ve^{-(\sigma+r+b)t}$. This behaviour is related to the existence of a bounded attracting set of zero volume with dimension smaller than the phase space. An attractor is a set towards which the dynamical system evolves over time. Geometrically, an attractor can be a point, a curve, a surface or even a complicated set with a fractal structure known as a strange attractor. A solution of the Lorenz equations has an initial transient portion and after that it may be settled on a strange attractor. Figure 1 shows exemplarily a numerical solution of the Lorenz system up to $t = 100$ from with initial conditions equal to $x_0 = 0$, $y_0 = 1$ and $z_0 = 0$ using a backward Euler scheme for the time stepping with $dt = 0.01$.

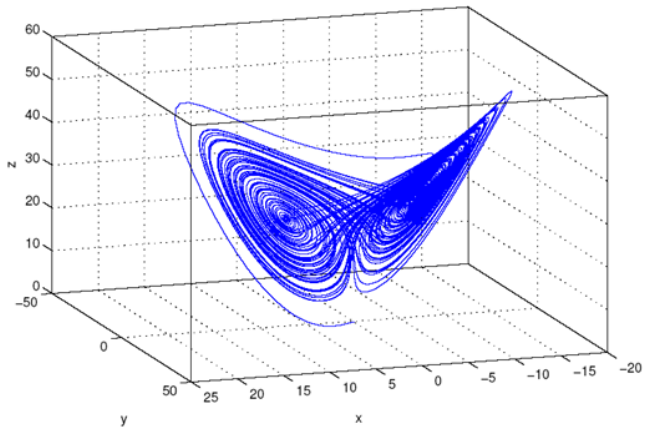

Figure 1. Numerical integration of the Lorenz (1963) system.

As is shown with more detail in next sections the difficulty of weather forecasting is due either to the sensitivity of the atmosphere evolution to small changes in the initial conditions related to the analysis error and to the sensitivity of the atmospheric differential equations to small differences in the numerical schemes used to find a numerical solution or model error. Figure 2 shows the evolution of the Lorenz system for two different but similar initial conditions. The solutions are very similar up to t = 25 approximately in this case and after that the differences become larger. After t = 30 the value of the variables x and y cannot be predicted although z remains more predictable. In general, the time range within which the system remains predictable, depends on the initial condition, and this characteristic is called the flow dependency of the predictability of the system.

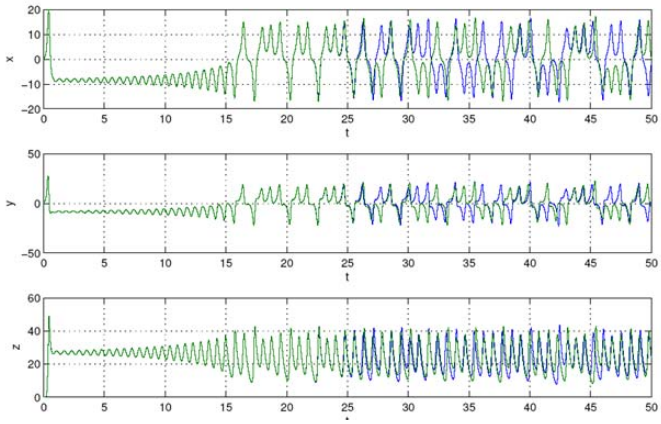

Figure 2. Numerical integration of the Lorenz (1963) system for two different and similar initial conditions. In green $x_0 = 0$, $y_0 = 1$ and $z_0 = 0$, in blue the initial conditions are $x_0 = 0.001$, $y_0 = 1$ and $z_0 = 0$.

The effect of model errors can be shown by changing slightly the constant parameters σ, r, b (Lorenz 1963). In a more complex model, this change would correspond, for example, to a change in the parameterization of the physical processes. Figure 3 shows the temporal evolution of the Lorenz system for two different sets of constant parameters. In this case, the predictability is loose after t = 20 for all the model variables.

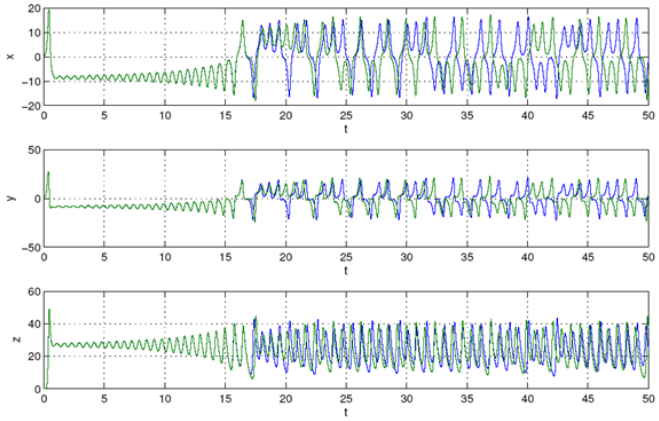

Figure 3. Numerical integration of the Lorenz (1963) system for two different and similar parameters. In green $s = 10$, r = 28 and b = 8/3, in blue a small value 0.001 is added to the parameters.

In light of these results (Lorenz, 1963), the question about the predictability of the atmosphere was raised for the first time, which has involved the efforts of the meteorological community to quantify it over several decades until today.

http://dx.doi.org/10.5772/55699

2.2. The predictability problem

A rigorous analysis of the chaotic properties of such a complex system as the atmosphere can be only achieved in simplified contexts. Atmosphere dynamics has been stated as chaotic and it is well established that there is an effective time barrier beyond which the detailed prediction of the weather may remain impossible (Lorenz, 1969). Predictability, or the degree to which a correct forecast can be made, depends on the spatial and temporal scales (from few hours at the mesoscale to few weeks at the planetary scale) and also on the variable (for instance, surface wind and temperature, precipitation or cloudiness).

Atmospheric chaos, uncertainty, predictability and instability are related concepts. Due to the approximate simulation of atmospheric processes, small errors in the initial conditions and model errors are the two main sources of uncertainties that limit the skill of a single deterministic forecast (Lorenz, 1963). Uncertainty limits the predictability, especially under unstable atmospheric conditions. The atmospheric instabilities related to low predictability conditions are the baroclinic instability at synoptic scales (Buizza & Palmer, 1995) and inertial and potential instabilities (e.g. deep convection) on the mesoscale, among others (Hohenegger & Schär, 2007; Zhang, 2005; Roebber & Reuter, 2002; Emanuel, 1979). This inherent limitation in predictability has led to the concept and development of ensemble prediction systems, which provide probabilistic forecasts to complement the traditional deterministic forecasts (Palmer et al., 1992).

3. Ensemble prediction systems

3.1. Uncertainty sources in numerical weather prediction

As indicated before, due to the chaotic nature of weather, there are several uncertainty or error sources in the Numerical Weather Prediction (NWP) framework that can grow and limit the predictability (Lorenz, 1963, 1969) of atmospheric flow. Forecast errors can arise due to inaccuracies in the initial condition atmospheric state estimates or due to imperfect data assimilation systems (Initial Conditions forecast error source), inadequacies of the NWP models themselves (Model Formulation forecast error source). Processes that take place at spatial scales that are shorter than the truncation scale of NWP models must be parameterized with sometimes inexact approximations thus giving us another source of forecast error (Parameterization forecast error source). One approach of NWP is to use Limited Area Models (LAMs) where the lateral conditions come from a global NWP models. This procedure is another source of forecast error (Lateral Boundary Conditions forecast error source). So far, the main error sources are: Initial Conditions (IC), Model Formulation, Parameterization and Lateral Boundary Conditions (LBC) error sources.

To the extent that these error sources project onto dynamical instabilities of the chaotic atmospheric system, such error will grow with time and evolve into spatial structures favoured by the atmospheric flow of the day. The inherent atmospheric predictability is thus state-dependent.

To estimate these uncertainties or errors, i.e. the predictability, many operational and scientific centres produce ensemble forecasts (e.g. NCEP, ECMWF, etc.). The idea of using ensemble forecasts has been know for many years (Leith, 1974). Since the early 1990s, many centres generate ensemble forecasts. The methodology that is behind is to run multiple (ensemble) forecast integrations from slightly perturbed IC (IC forecast error source), using multiple models or perturbing model formulation (Model Formulation forecast error source). Adding stochastic physics parameterizations (Ehrendorfer, 1997; Palmer, 2001) or using multiple boundary conditions (Lateral Boundary Conditions forecast error source) among others techniques is described below. The discrete distribution of ensemble forecasts can be interpreted as a forecast Probability Density Function (PDF). If an idealized forecast ensemble can be constructed that properly characterizes all sources of forecast errors, then the forecast PDFs would be reliable (see section 5) and skilful (sharper than the climatological PDF). No further information would be needed to make trustworthy forecast-error predictions, since a perfect PDF is a complete statement of the actual forecast uncertainty.

In practice, estimates of all the forecast-error sources mentioned above are inexact, leading to PDFs from real ensemble forecasts with substantial errors in both of the first two moments (mean and variance). These limitations are particularly pronounced for mesoscale prediction of near-surface weather variables, where large underdispersion results from insufficient ensemble size, inadequate parameterization of sub-grid scale processes, and incomplete or inaccurate knowledge of land surface boundary conditions (Eckel & Mass, 2005). Real ensemble forecast distributions, although generated using incomplete representations of weather forecast error sources, often represent a substantial portion of the true forecast uncertainty.

3.1.1. Initial conditions forecast error source

It is clear that the atmospheric state at a given time is not perfectly known; not only the inherent observational errors alone guarantee this, but also the sparse network of observations worldwide that sample the atmosphere only at limited intervals with inexact results. In addition network density and design can yield errors in regional averages (PaiMazumder & Mölders, 2009). Another contribution to IC forecast error source is the Data Assimilation (DA) system used. Every DA system is affected by the characteristic errors of both the observations incorporated in the analysis and of the short-range model forecast, which is typically used as a background or *first guess* field to be adjusted by new observations. IC errors, however small, are exacerbated by the chaotic dynamics of the atmosphere and consequently grow nonlinearly with time.

3.1.2. Model formulation forecast error source

NWP model inadequacy is inevitable given to our inability to represent numerically the governing atmospheric physical laws in full. Contributions to this forecast-error source can be found in the model used which is, of course, a simplified scheme of what really happens in the atmosphere, dynamical formulation, different discretization methods, the numerical

http://dx.doi.org/10.5772/55699

method employed to integrate the model and the different horizontal and vertical discretization resolutions used.

The model formulation forecast error in conjunction with another forecast error sources such as parameterizations has been recognized traditionally by operational forecasters in NWP centres. They usually select *a best model of the day* when producing their operational forecasts. This model selection tries to best handle the evolution of the atmosphere depending on the flow the general situation and the season of the year. The selection is driven by the subjective knowledge than some models behave better than others in some situations due to their formulation.

3.1.3. Parameterization forecast error source

There are several parameterized processes in NWP models: those which are taking place at smaller spatial scales than the truncation scale of the NWP model and are not resolved explicitly by the model as convection. Another one is introduced in a simplified way due to computer time limitations like radiation, and finally processes which are not taking into account in the NWP model dynamic part as microphysics in clouds. All theses processes are called sub-grid processes. It is assumed that sub-grid processes are in equilibrium with grid resolved states and so they can be represented statistically from them. A parameterization is the statistical method used when representing the sub-grid processes. Parameterizations are always imperfect representation of atmospheric processes so they always include inherent errors (Tribbia & Baumhefner, 1988; Palmer, 1997). NWP parameterizations have a time and space scale dependency. At small scales, forecast verification is primarily concerned with the locations and amounts of precipitation and other sensible weather parameters, which are often directly affected by the assumptions used to develop the model parameterization schemes for convection and other processes. Moreover, especially for the higher model resolutions, the implicit equilibrium assumption of sub-grid processes with model state could break down being another source of parameterization uncertainty.

3.1.4. Lateral boundary condition forecast error source

The LBC forecast error is only present in LAMs or regional models, which have as inputs lateral boundary values spatially and temporally interpolated from a coarser resolution grid-point or spectral model. So the coarser model errors are translated into LAMs as LBC error source. For instance, a possible configuration for a LAM EPS could include lateral boundary conditions from an ensemble of global forecasts.

3.2. Techniques used by global models

For many years operational forecasters, particularly medium range forecasters in meteorological services, have had access to some forecast products coming from global NWP centres other than their own. They routinely compare forecasts from different centres to assess the confidence in the forecasts of their own models, and to determine possible alternative forecasts. This set of available forecasts is often called the *Poor Man's* (Ebert,

2001) ensemble because its production is relatively cheap compared to the cost of developing and running a full EPS such as the ECMWF and NCEP ones. It is called *Ad hoc* ensembles by some other authors. Theses ensembles are cheap and easy to create, but they are not generated in a controlled and systematic approach. Not only are they not calibrated but also some ensemble members may be always quite more skilful than others. The hypothesis of equiprobability of the ensemble members is less guaranteed than others EPS which is a major drawback.

Hoffman and Kalnay (1983) proposed a *time-lagged* method or *Lagged Average Forecast* (LAF) method. The time-lagged method uses forecasts from lagged starting times as ensemble members. These members are easy to construct but they lack any scientific motivation. On the contrary, LAF perturbations are realistic short-term forecast errors. However, LAF ensemble forecasting has the disadvantage that most of the times earlier forecasts are considerably less skilful than later forecasts. This drawback can be partly resolved by either using different weights for different members of the ensemble or by scaling back the larger errors to a reasonable size. This procedure is the basis of the Scaled Lagged Average Forecast (SLAF) technique (Ebisuzaki & Kalnay, 1991).

The *multi-model SuperEnsemble* technique (Krishnamurti et al., 1999) is a powerful method to construct EPS. Several different models outputs are put together with appropriate weights to get a combined estimation of weather parameters. Weights are calculated by square minimization in a period that is called *training period*.

A better solution is to sample the different error sources that were indicated before. Depending on the sampling technique we obtain different methods: a Monte Carlo approach as proposed by Leith (1974), Hollingsworth (1980) and Mullen and Baumhefner (1989) among others. In general, the technique consists of sampling all sources of forecast error, by adding or perturbing any input variable (analysis, initial conditions, boundary conditions etc.) and whatsoever meteorological parameter that is not perfectly known. These perturbations can be generated in different ways. The main limitation of the Monte Carlo approach is the need to perform a high number of perturbations in order to have a proper description of the initial uncertainty, which is usually far from the available computational resources. This limitation leads to reduced sampling by just sampling the leading sources of forecast error due to the complexity and high dimensionality of the system. Reduced sampling identifies active components that will dominate forecast error growth.

IC forecast error source have a dominant effect. To sample it several techniques have been available. One of them is the initial perturbations method, which consists of adding small perturbations to the initial analysis, such as NCEP's *breed mode* method (Toth & Kalnay, 1993, 1997; Tracton & Kalnay, 1993). The breed mode method is based on the idea that the analysis created by the data assimilation scheme used will accumulate growing errors. As it can be seen in Figure 4 breeding vectors give a sampling of the growing analysis error: a random perturbation is added to the analysis, evolved in time by integrating the forecast model, rescaled and reintroduced as a perturbation. After several cycles only the fastest growing errors remain.

http://dx.doi.org/10.5772/55699

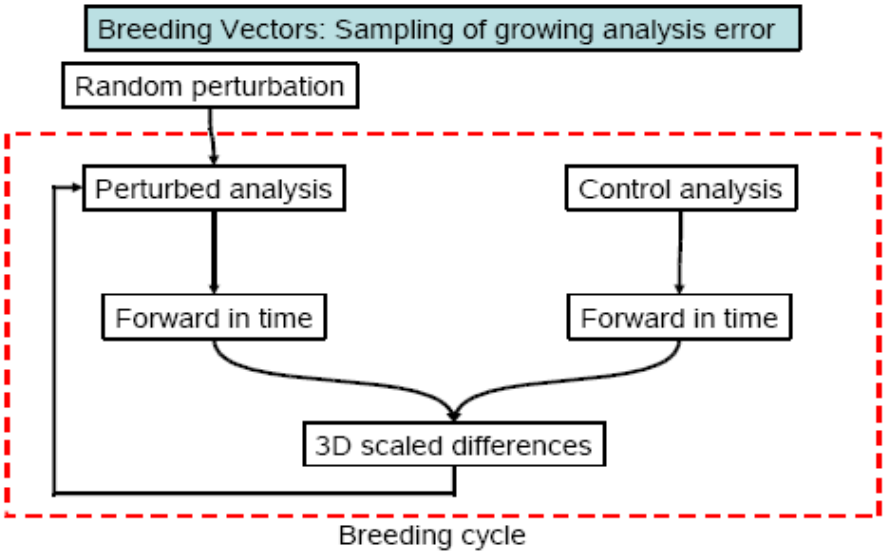

Figure 4. Breeding technique strategy

An alternative to the breed mode is the ECMWF's *singular vector* method (Palmer et al., 1992; Molteni et al 1996) which tries to identify the dynamically most unstable regions of the atmosphere by calculating where small initial uncertainties would affect a 48 hour forecast most rapidly. It needs an adjoint model. Singular vectors give a sampling of the perturbations that produce the fastest linear growth in the future. There are only a relative small number of directions in the phase-space of the atmospheric system along which the most important processes occur. Maximum growth is measured in terms of total energy. The adjoint of the tangent forward propagator with respect to the total norm is defined, and the singular vectors (the fastest growing perturbations) are computed by solving an eigenvalue problem. Singular vector method is schematically described in Figure 5.

In addition to the breeding and singular vector methods there are *Ensemble Transform Kalman Filter* technique (ETKF; Bishop et al., 2001; Wang & Bishop, 2003) and *Ensemble Data Assimilation* (EDA; Houtekamer, 1996; Buizza, 2008). ETKF is similar to the breeding method except that the rescaling factor is replaced by a transformation matrix. It produces an improved ensemble dispersion growth. It is used at the UK Meteorological Office. In EDA, an ensemble of assimilations is created from different analyses which have been generated by randomly perturbing the observations in a manner consistent with observation error statistics.

Model forecast error source, i.e. model formulation and parameterization error sources together, is another component to take into account. To represent model uncertainty several approaches have been used: the m*ulti-model* approach (e.g. DEMETER; ENSEMBLES; TIGGE; Krishnamurti et al, 1999), *multi-parameterizations* or *multi-physics* approach (Houtekamer, 1996), *stochastic parameterizations* (Buizza et al., 1999; Lin & Neelin, 2002), *multi-parameter*

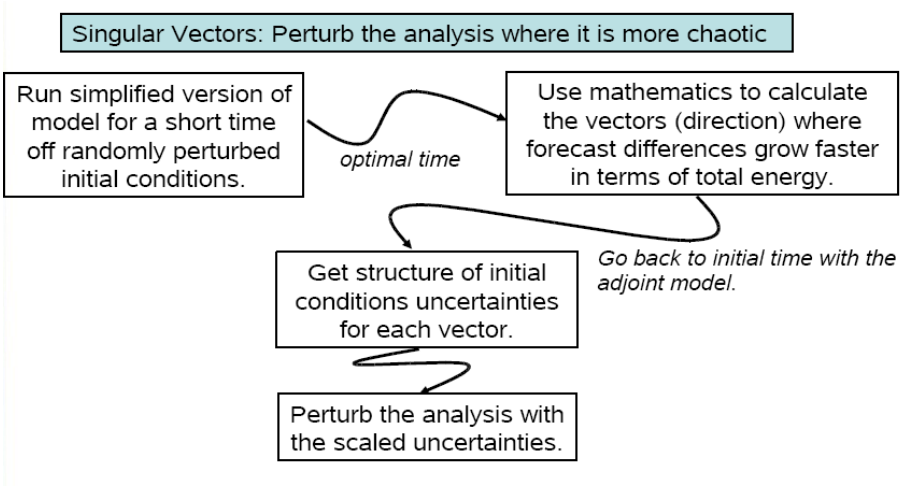

Figure 5. Singular vector technique strategy

approach (Murphy et al., 2004), and *Stochastic-Kinetic Energy-Backscatter* approach (Shutts & Palmer, 2004; Shutts, 2005).

So, in order to sample model forecast error an ensemble forecasts are produced by using different numerical models (multi-model approach). The multi-model approach implies equiprobable members which is not always the case. An alternative method for sampling model forecast errors is using different physical packages (multi-physics approach). Another approach is the *stochastic parameterization* approach applied at ECMWF (Buizza et al., 1999). It is based upon applying stochastically perturbing the total parameterized tendencies with a multiplicative noise. The multi-parameter approach tries to take into account the significant uncertainty in some parameters in NWP models, for instance, by using different values in each ensemble member.

Finally, the *Stochastic-Kinetic Energy-Backscatter* approach addresses a missing physical process, the upscale energy cascade energy from the grid scale to synoptic scales lost due to the excessive dissipation energy in NWP models.

3.3. Techniques used by limited area models

Not only global models can be used in building EPS, but also Limited Area Models (LAM) can be used to create LAM EPS, normally used for the short range. Error sources in LAM EPS are the same as in global EPS, but LAM models require lateral boundary conditions that update the weather situation regularly throughout the integration. These lateral boundary conditions introduce a main source of uncertainty in LAM ensembles. Both LBCs and ICs give their contribution to the spread and skill of the system (Clark et al., 2009). All the techniques discussed so far can be applied to generate LAM EPS. A very popular generating technique is the downscaling of global ensemble forecasts. This technique consists of using the selected

global ensemble members (chosen by clustering) as initial and boundary conditions for a limited area ensemble system. The difficulty is that the perturbations generated from the global EPS are usually effective only on the medium range and large scales. Therefore they are not likely optimal for short range ensemble forecasts. Another technique for sampling lateral boundaries forecast error source is *multi-boundary* technique. In the multi-boundary technique, several different global models supply the lateral boundary conditions needed by the LAM model. One example of the use of the multi-boundary technique is the AEMET Short Range Ensemble Prediction System (AEMET-SREPS; García-Moya et al., 2011). AEMET-SREPS uses the multi-boundary method in addition to the multi-model method. It is built by using a set of LAMs and a set of deterministic global models that supply the initial and boundary conditions. The system is focused on short-range forecast and has been developed to help in the forecast of extreme weather events (gales, heavy precipitation and snow storm) and provides forecasts with good reliability, resolution and discrimination consistently with the analysis in the large-scale flow.

4. Uncertainty representation and weather forecasting products

4.1. Uncertainty representation

In statistics, uncertainty is represented by means of the Probability Distribution Function (PDF). Let us consider a random variable x that we do not know, a priori, anything about its nature. The question is whether we can infer something about it. Let us take n different values of x that belong to the same population. When we construct the histogram of these values, we obtain an approximation of its PDF. As an example, we could think of x as the mean monthly temperature of April at a surface observation station. Then the population would be *the mean monthly temperatures of April at that station*. If we restrict us to only the period 1981-2010, then the n=30 values of x would form the sample space.

The PDF gives us information about the behaviour of the random variable x. For example, let us take the normal or Gaussian PDF of which the analytical formula is:

$$PDF(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

Here σ is the standard deviation and μ is the mean. Figure 6 shows this distribution graphically. Now we can infer something about the nature of the variable x. From Figure 6 we can say that there is a value μ around which all the random variables are distributed symmetrically. Likewise, σ is a measure of the standard deviation of x from its mean. We can think of σ as a mean error (or uncertainty) we would have if we approximated any possible value of x by μ. In resume, the PDF gives us a depiction of all the possible values of x and their associated probabilities of occurrence. This procedure results in an explicit and quantitative way of representing the uncertainty of a random variable.

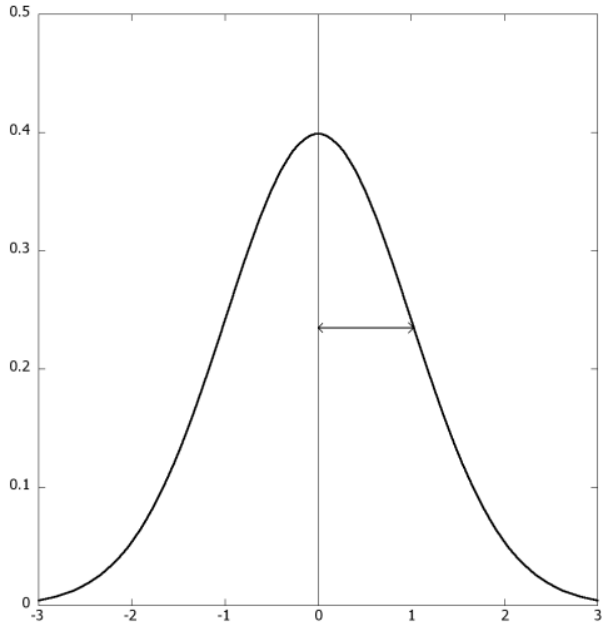

Figure 6. Gaussian PDF with μ = 0 and σ = 1

In the case of a Numerical Weather Prediction (NWP) system there are about 10^8 different random variables x, each one corresponding to each degree of freedom of the model. This fact makes it computationally unfeasible to integrate the Liouville equation (sometimes referred to as Fokker-Plank equation when random processes are included to account for, for example, model error) that describes the time evolution of a PDF. A practical way to resolve this problem is to use an EPS. An ensemble prediction tries to estimate the uncertainty of the forecast by discretizing the forecast PDF for each model parameter at each grid point in N values corresponding to the N ensemble members. As an example, Figure 7 presents the PDF of a 60 hours two metres (2m) temperature forecast of the AEMET-SREPS for the grid point closest to Sevilla, Spain. This ensemble has 25 members, but in this case, there was one that did not integrate properly, so N=24. It is easy to see that the more members the ensemble has, the higher is the resolution of the PDF.

In ensemble prediction, a simplified way of representing the uncertainty of the forecast is the *spread* (Toth & Kalnay, 1997); the standard deviation σ (4) of the PDF quantifies how much the ensemble members deviate, on average, from the mean, and it is often used as a measure of the spread (Wilks, 2006):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - em)^2} \tag{4}$$

http://dx.doi.org/10.5772/55699

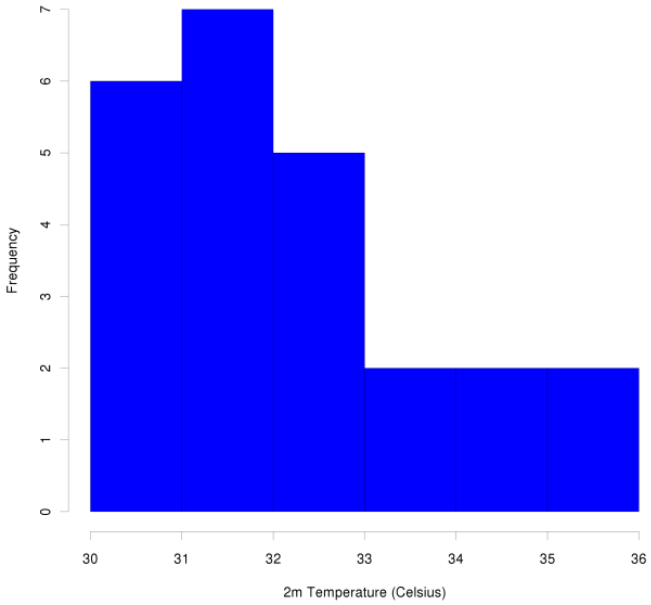

Figure 7. Histogram of 2m Temperature. AEMET-SREPS H+60 predicted values at the grid point closest to Seville. Valid time: 30th June 2011, 12 UTC.

Where f_i is the forecasted variable by member *i*, and *em* is the ensemble mean, e.g., the mean of the *N* forecasts. This parameter can be calculated on each grid point. In the case of the 2m temperature forecast (Figure 7) *em* and the spread are 32.3 and 1.5 ºC, respectively. The latter can be interpreted as an estimate of the error of the deterministic forecast so that the higher the spread, the more uncertain is the forecast. Other measures of spread, more robust or resistant, can be alternatively used, e.g. the interquartile range (Wilks, 2006).

4.2. Raw products

Ensembles are composed of members that are deterministic predictions, and allow providing individual deterministic information (García-Moya et al, 2011). This information can either help the traditional staff to understand ensembles, and can provide support in the probabilistic interpretation. Far beyond this, ensembles provide their most useful information when we look at them as intrinsically probabilistic prediction systems. In this context, most of the ensemble outputs reflect this probabilistic nature. Probability is a rich and reasonable model to describe and understand many aspects of the physical world, but the interpretation of ensemble outputs must be learned and used carefully beyond the straightforward interpretation, because (especially for *deterministic forecasters*) some interpretations can be in conflict with *common sense*. Given a grid point with N forecast values x'_i (for an ensemble with N members), we call raw products when only these N values are used straightforward. Three basic examples of raw products exist.

4.2.1. Stamps

The deterministic outputs for all the ensemble members can be plotted as usual meteorological charts (see for instance Figure 8 with MSLP and T850 for the ECMWF EPS). These usual *postage stamp charts* comprise the control forecast (if there is any, top left), the perturbed members (below) and the corresponding high resolution deterministic forecast (if exists, beside the control). The difficulty is that the forecaster would have to deal with an amount of information: 51 scenarios in addition to the high resolution deterministic forecast.

4.2.2. Plumes

In a given location, we can provide N forecast values for that place (either by bi-linear interpolation or some other fine process which could account for height variability). Moreover, we can plot the evolution with forecast step for all the N members, i.e. we would plot N curves. Like on the stamp charts, the control forecast can be highlighted and the high-resolution deterministic forecast can also be plotted (e.g., Figure 9). The difficulty is similar to that one of the stamps namely the necessity to deal with such an amount of information. For a specific location and parameter, plumes can help the forecast guidance and, in fact, often provide information about the uncertainty and general trends.

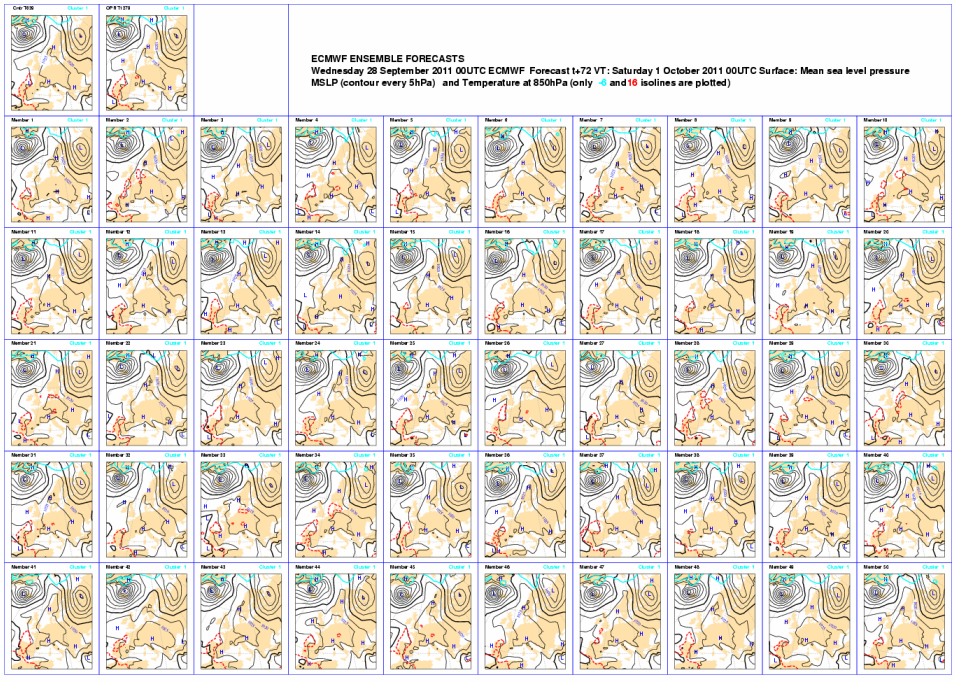

Figure 8. Postage stamps charts of MSLP and 850 hPa temperature T+108 forecasts (see text). Courtesy of ECMWF (2011).

http://dx.doi.org/10.5772/55699

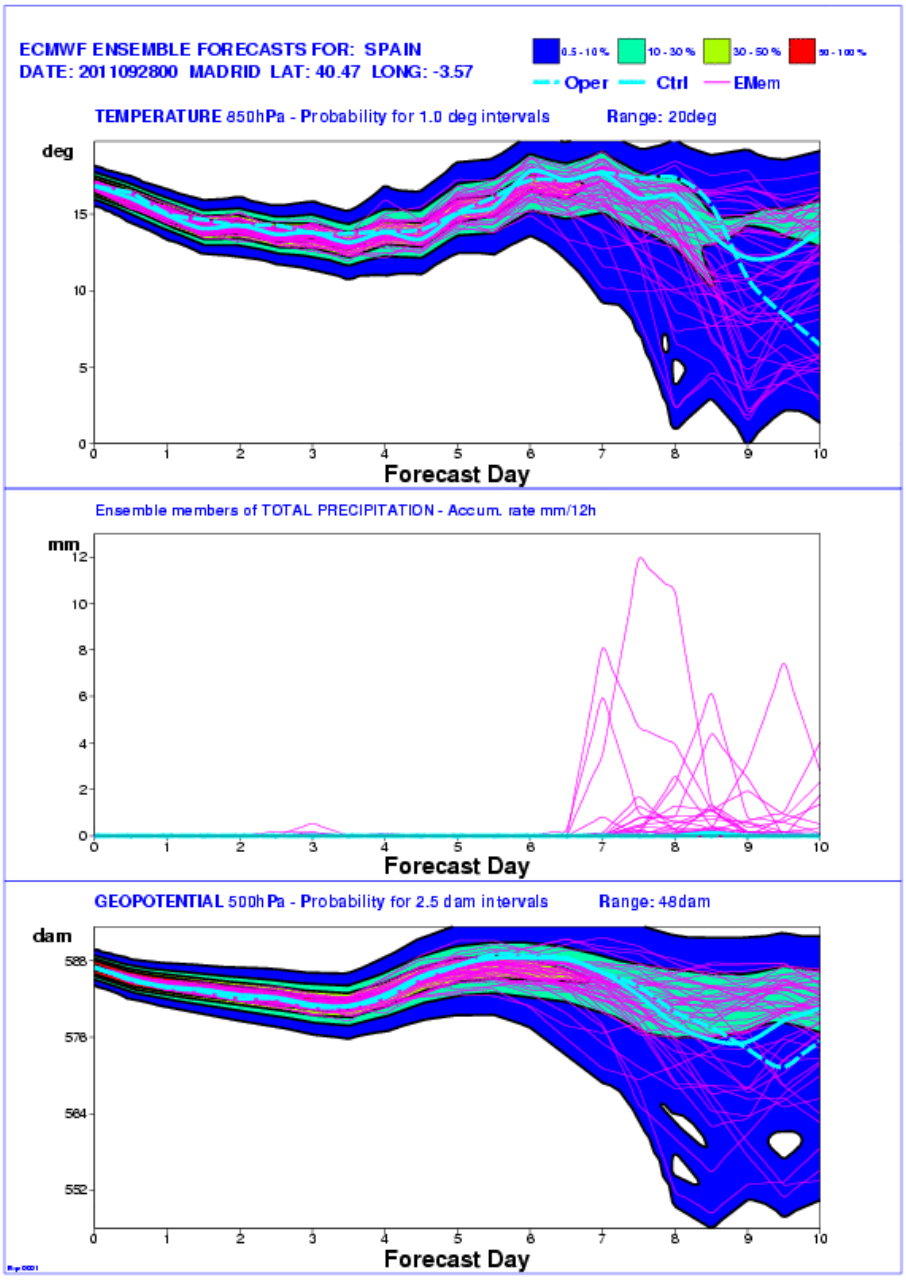

Figure 9. days forecast plumes of 850 hPa temperature at Madrid (see text). Courtesy of ECMWF (2011).

4.2.3. Spaghetti

As a third example, we show the spaghetti charts. For a given (often dynamical) field, it is rather impossible to overlay N member charts. But picking a selected isoline of interest, we can plot one line per member and, thus, the whole plot would contain N lines. Typically, the control is highlighted, and a higher resolution deterministic forecast can be included (e.g. Figure 10). As plumes, this kind of plot can help the forecaster to provide information about the uncertainty.

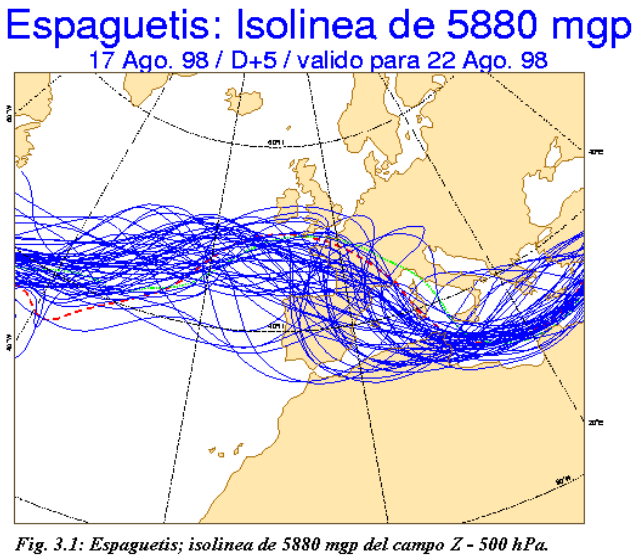

Figure 10. Spaghetti chart of geopotential height (5880-gpm- isoline) for the T+120 forecast at 500 hPa (see text). Courtesy of ECMWF (2011).

All of these raw outputs share the same shortcoming: the inherent difficulty in the forecast guidance for handling the huge amount of information they provide. This issue is addressed using derived probabilistic outputs that compact this information naturally.

4.3. Derived probabilistic products

Probabilistic outputs are derived computationally from the PDF representation, which is assumed to be provided by the ensemble members. These products reflect the probabilistic nature of the ensemble, visually and conceptually. They provide explicit, quantitative and detailed information about uncertainty, and this fact is a real breaking point with respect to deterministic model products. They address the issue of providing compact information in this natural way. Ensembles nowadays can provide ideal complementary information to a higher resolution deterministic model.

http://dx.doi.org/10.5772/55699

For a given grid point, there are N forecast values x'_i for an ensemble with N members. Without further information about the skill of the members, we assume Laplace principle of equal probability, dealing in this case with a discrete PDF. An estimate of the forecast probability p of exceeding a threshold t is given by the well-know formula where the indicator $I(x'_i)$ is usually defined as $I(x'_i)=1$ if $x'_i > t$, $I(x'_i)=0$ otherwise (Ferro 2007b):

$$p = \frac{1}{N}\sum_{i=1}^{N} I(x_i') = \left\{\frac{0}{N}, \ldots, \frac{N}{N}\right\} \tag{5}$$

The corresponding inverse is the computation of percentiles, i.e., for a given probability p, what is the actual forecast value x' for which $p = P(x')$. By adding further information, we can improve the PDF (e.g. by calibration) and the computation would be different. By taking this simple discrete model, we can compute the probabilities of exceeding thresholds, the percentiles for given probabilities, the summary statistics (e.g. mean and standard deviation), etc.

4.3.1. Ensemble mean and spread charts

The ensemble mean (the arithmetic mean of all the ensemble members) is not always a feasible meteorological situation because it is obtained as a result of a statistical operation, not from a numerical model (Buizza and Palmer, 1997). So, it is strongly discouraging in forecast guidance to use the ensemble mean without special care, if at all (García-Moya et al, 2011). However, the ensemble mean is often plotted in charts together with the standard deviation (the latter as a measure of spread), to help with the understanding of the atmospheric flow (e.g. Figure 11). The standard deviation is sometimes normalized.

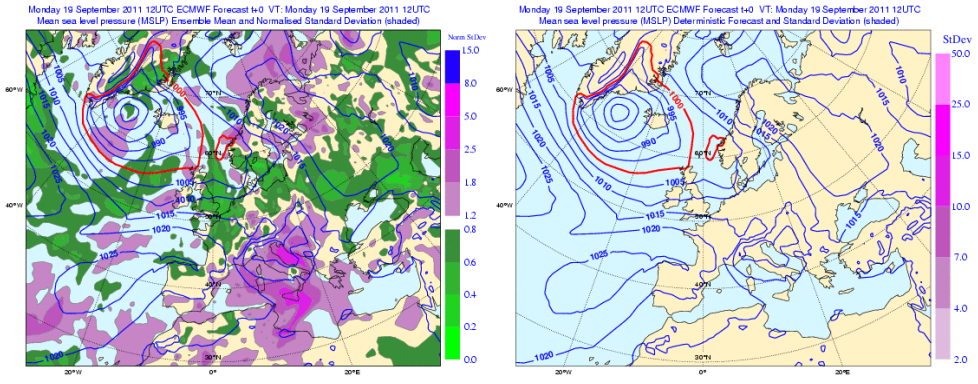

Figure 11. MSLP T+00 ensemble mean (contour) and normalized standard deviation (colours) (see text). Courtesy of ECMWF (2011).

4.3.2. Probability maps and percentile maps

Given a binary event (e.g. precipitation forecast > 5 mm/6h) we can plot the spatial distribution of the forecast probabilities that the EPS provides (see Figure 12). Similar plots can be made for the percentiles. These maps provide the forecasters with useful guidance by showing them where it is more probable for an event of interest to occur (e.g. representative precipitation that exceeds 5 mm/6h).

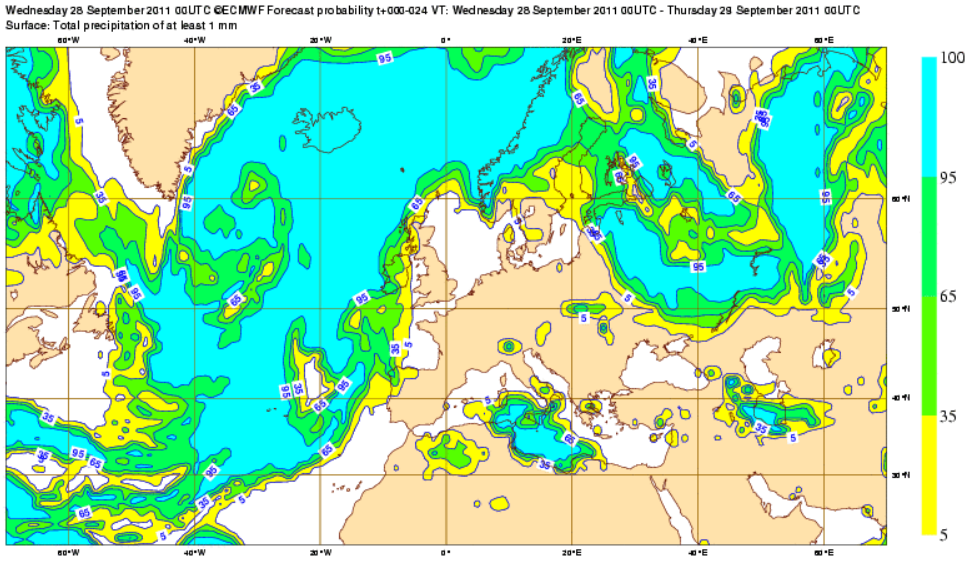

Figure 12. Probability of having accumulated precipitation greater than 1 mm in the interval T+0 to T+24 (see text). Courtesy of ECMWF (2011).

4.3.3. EPS-grams

Box-plots (Wilks, 2006) and similar graphs give a quick, visual and simple representation of a distribution of numbers, a discrete PDF. Extending this idea by including the evolution with forecast time of the main weather parameters at a given location, we obtain plots that are *meteogram based* and often called *EPS-grams*. The building brick is the box-plot: it displays the median, minimum, maximum, percentiles 25 and 75 and sometimes also percentiles 10 and 90. Box-plots are displayed for a series of forecast steps. This procedure is typically applied for the more sensitive parameters to forecast e.g. cloud cover, precipitation, ten metres (10m) wind speed and 2m temperature (e.g. Figure 13). Special care must be taken with EPS-grams interpretation (Persson & Grazzini, 2005) by comparing the location point and the nearest grid-points: distance, land/sea contrast and height must be checked in order to properly use the information that EPS-grams provide. Anyway, the EPS-grams are the most popular and probably useful plots to forecast the weather in a location by taking into account the uncertainties.

http://dx.doi.org/10.5772/55699

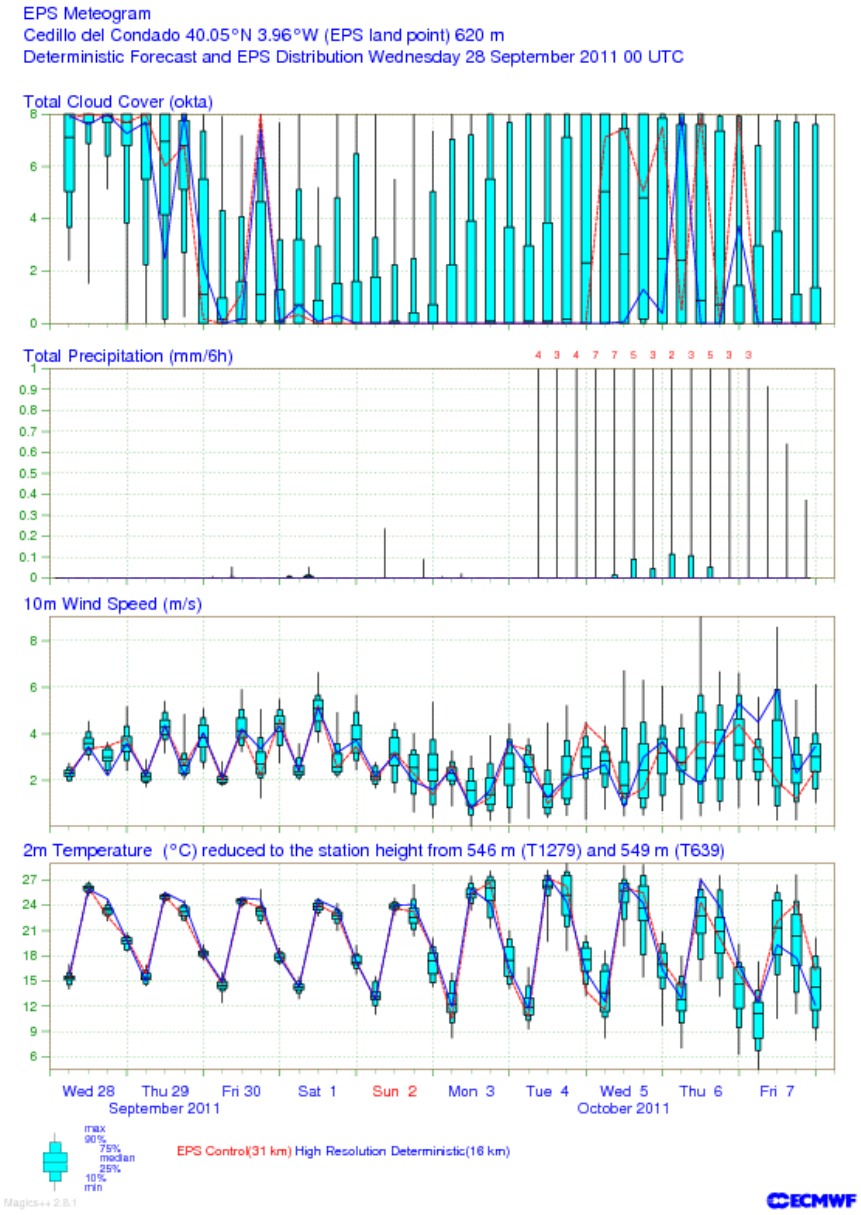

Figure 13. days EPS-gram at Cedillo del Condado (Spain) (see text). Courtesy of ECMWF (2011).

4.3.4. Clustering

A natural way of examining the number of atmospheric scenarios that EPS provide (e.g. stamps) is by similarity: one can gather scenarios in groups that could lead to similar weather conditions (Ferranti, 2010). This process can be done by *eye-ball* or can be carried out computationally by using a clustering algorithm (Ferranti, 2010) that fits the corresponding meteorological requirements. This procedure is often expensive and requires an extra task of defining the similarity criterion, but forecast guidance can improve substantially with the use of clusters. Examples of algorithms used are the Ward algorithm (Ferranti, 2010) and the tubing; both have been used operationally at AEMET (Figure 14). Here clusters have proved to be a very useful guidance in medium range forecasts by summarizing the more important and distinct scenarios (Ferranti, 2010).

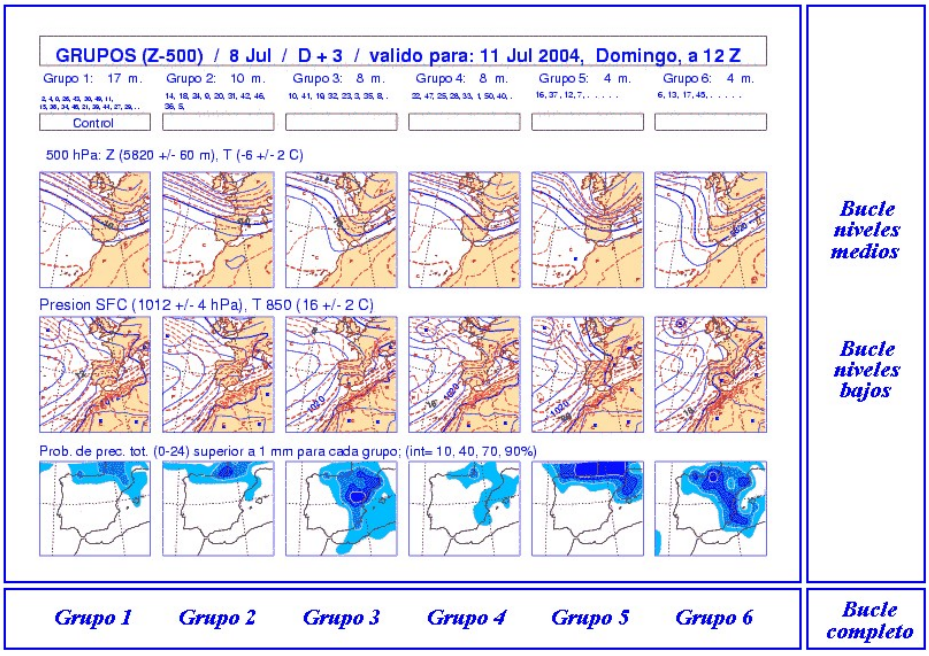

Figure 14. day forecast clusters of 500 hPa height and surface pressure (see text). Courtesy of AEMET (2011)

4.3.5. Extreme forecast index

Extreme events are not always severe, but severe events are often extremes. An index of *extreme forecast* can be computed using the model climatology as a reference, rather than the climatology of observations (Lalaurette, 2003). When observations are used, the forecast is not really prone to fall in the tail of the climatological distribution, and this fact is addressed by using the model climatology. The Extreme Forecast Index (EFI; Lalaurette, 2003) is a quantitative

http://dx.doi.org/10.5772/55699

measure of how extreme is the EPS forecast when compared with the model climatology. The EFI can be plotted in a chart (Figure 15), and this chart is especially useful for weather parameters. Thus the EFI is used by forecasters as an early warning tool to highlight where severe events could happen.

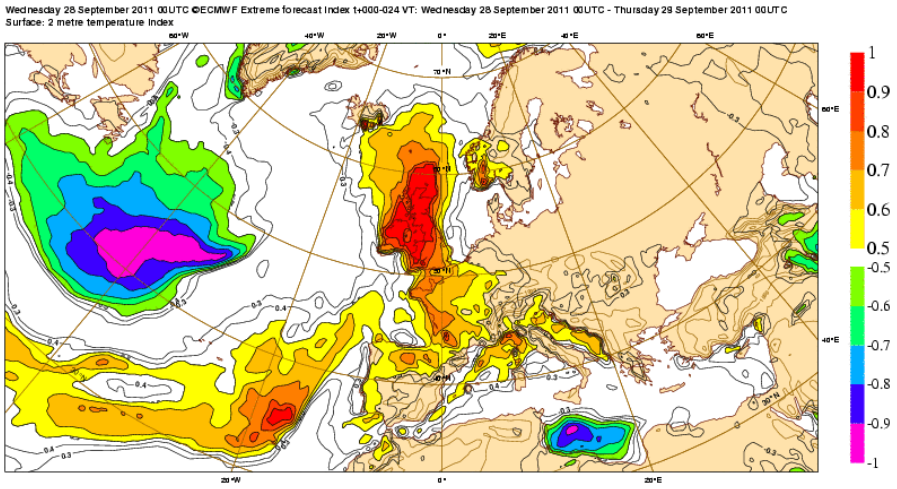

Figure 15. Extreme Forecast Index for range T+0 to T+24 of 2m temperature forecast (see text). Courtesy of ECMWF (2011).

4.4. Interpretation for weather forecasting

As an example for the application of probabilistic forecasting we present a real case of extreme winds that is fully described in Escribà et al. (2010). This section is not intended to be a detailed manual of the utilization of probabilistic products in operational forecasting. More extend and concise information can be found for example at www.ecmwf.int.

At 00 UTC on 24 January 2009 an explosive cyclogenesis in the Gulf of Vizcaya reached its maximum intensity with an observed surface pressures less than 970 hPa on its center. During the cyclone's path through the south of France strong westerly and north-westerly winds occurred over the Iberian Peninsula (> 150 km/h). These winds led to eight casualties in Catalunya, the north-east region of Spain.

In Figure 16 are represented three probabilistic forecasting products, the ensemble mean, the ensemble spread and the probability of having wind speeds greater than 15 m/s (54 km/h). The weather parameter analyzed is 10m wind speed. These fields correspond to the H+60 prediction of the AEMET-SREPS initialized at 00 UTC on 22 January 2009. The ECMWF reanalysis is also shown as verification.

Even though wind speed values plotted in the maps are not extreme, they do not exceed 20 m/s or 72 km/h, this fact has to be taken carefully because we are representing mean values of

wind at a forecast time, i.e. a mean over a time interval equal to the last forecast time step of the forecasting model (which in this case is around 5 minutes). As a first approximation we can estimate the wind gust (maximum wind) as twice the value of the mean wind (this factor can be roughly obtained comparing temporal series of mean wind and wind gusts from observation ground stations). In this case, such an approximation would give us extreme winds of about 150 km/h, similar to those observed.

The ensemble mean (Figure 16) can be thought as a skilful deterministic forecast that comes from the ensemble. When we compare it with the verification we can highlight there is a good agreement in the overall patterns. Looking in more detail we can select three zones where there is more discrepancy: a.) south of Catalunya (yellow ellipse), b.) Aragon and Valencia (white ellipse) and c.) south-east of France (green ellipse). It is especially interesting to analyze zone a.), where the casualties occurred. The question is whether the ensemble can estimate in some way the error in the prediction; the spread is expected to give information on this.

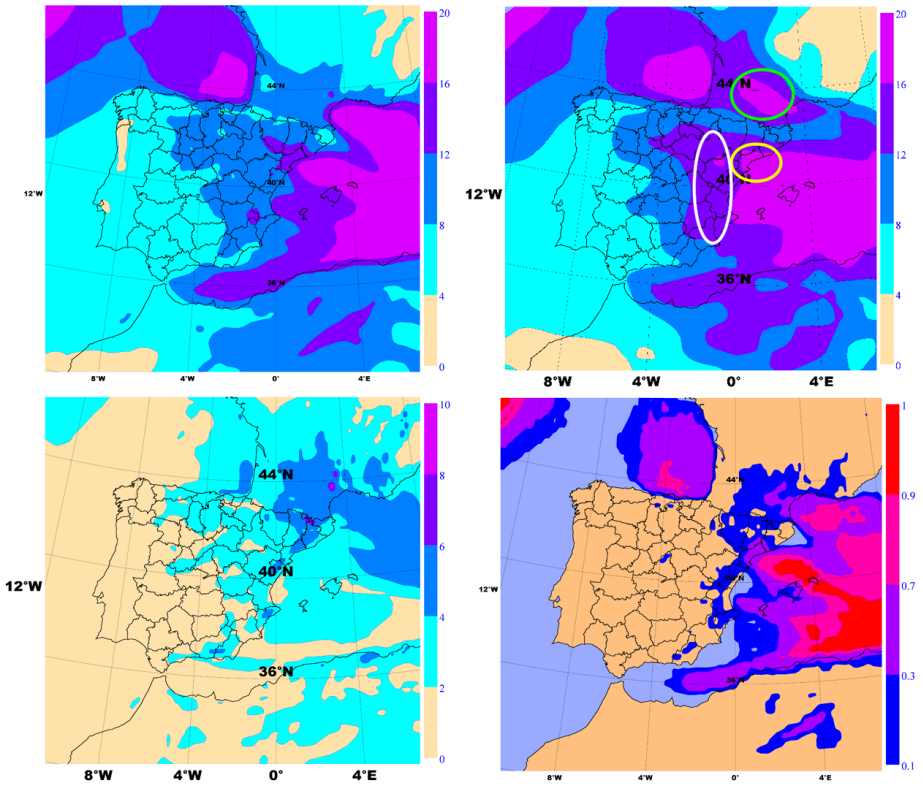

Figure 16. Ensemble mean (top left), ECMWF reanalysis verification (top right), ensemble spread (bottom left) and probability forecast of S10m > 15 m/s (bottom right). The probability field is the only one that is not in m/s. Lead time prediction is H+60 and verification time is 12 UTC of 24 January 2009.

http://dx.doi.org/10.5772/55699

The ensemble spread gives us the areas of more uncertainty in the prediction and is a measure of it. For zones a.) and c.) it has values around 5 m/s, which are considerable. In the case of a.), the spread estimates properly the prediction error, giving valuable information. When making the forecast, we would say that it is possible to have wind speeds greater than 16 m/s. This is not the case in zone c.), where the spread is not enough to explain the discrepancy between the ensemble mean prediction and the verification. In this case, the probabilistic forecast is badly displaced to the east. In zone b.), a lower spread (around 3 m/s) has also the ability to describe the prediction error.

Finally, the probability forecast enforces the general forecast by determining the areas of maximum confidence of occurrence and quantifying this confidence in a number. In this sense, we can say that the probability of having mean wind speed greater than 15 m/s (54 km/h) or wind gusts of more than 100 km/h in zone a.) is between 30% and 70%, which is more than the term *possible*.

5. Ensemble forecast verification

NWP models must be compared with a good representation of the observed atmosphere. This process is often called *forecast verification*, and raises a number of concepts and issues. With verification, we assess the quality and value of forecasts (Murphy, 1993), by using metrics or measures often called scores. By providing detailed information about forecast performance, verification can help in model improvement (developers) and forecast guidance (forecasters). Comprehensive descriptions of standard verification methods can be found in Wilks, 2006 and in Jolliffe & Stephenson, 2003), whereas in Candille & Talagrand, 2005 and in Stensrud and Yussouf, 2007 there is a thorough study of probabilistic forecasts, including ensemble forecasts.

Different frameworks are available for verification. Observations (ob) and forecasts (fc) can be compared, either *whole set to whole set* or *fc to ob* by using their statistical summary properties (measures-oriented approach), as distributions (distributions-oriented), as features (features-oriented), etc. In any case, to compare observations and models is like comparing apples and oranges: they are often different kinds of atmospheric representations, and we need to transform one or both of them into *comparable* representations. This step involves non-trivial issues like interpolation, representativeness, correlation, noise, etc.

An ordinary example of the difficult issue in comparing apples and oranges is the performance assessment of quantitative precipitation forecasts (QPF) from a deterministic model. European meteorological offices provide to the ECMWF 24-hour accumulated precipitation reports from their high-density rain gauges networks. Forecast values are regularly spaced, while observations are not. One comparison method is to interpolate forecast values to observation points (Rodwell, 2010). Special care should be taken with the impact of spatial density of observations and the potential lack of statistical consistency due to spatial dependence between close ones. To address these issues, one can compute an observed quantitative precipitation estimate (QPE) using a simple up-scaling technique (Ghelli & Lalaurette, 2000; Cherubini et al., 2002) whereby stations are assigned to model grid-boxes and then *averaged* to produce one value to

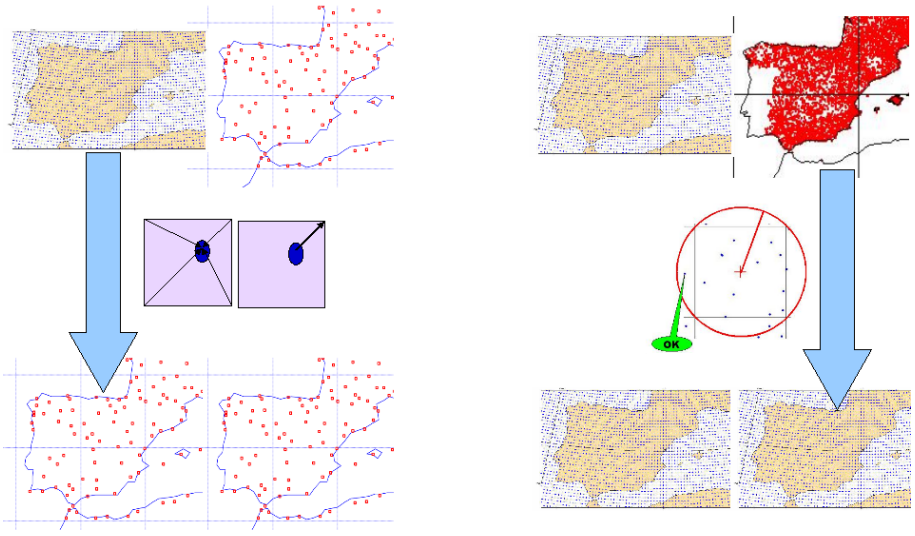

Figure 17. Interpolation to point (left) versus up-scaling (right)

be assigned to the corresponding grid-point. Such a grid value can be for instance a weighted average, and can be compared to the model precipitation forecast, which is also relative to the grid-box areal average, as they both refer to the same spatial scales. Whether to choose interpolation or up-scaling methods it depends on the case (see Figure 17).

Moreover, in the comparison of the performance of QPF from two different forecasting models, further issues arise. The grid spacing of the two models might be different. If observations are gridded to the finer resolution, then the coarser model might be penalized. On the other hand, if observations are gridded to the coarser resolution, the comparison can be fair but the higher resolution model is not given a chance. How to compare the way in which both models represent structures at their own scale is a non-trivial issue. PaiMazumder & Mölders (2009) assessed the impact of network density and design on regional averages using real sites and model simulations over Russia. They find that generally, the real networks underestimate regional averages of sea level pressure, wind speed, and precipitation while overestimate 2m temperature, downward shortwave radiation and soil temperature.

5.1. A first requirement: Deterministic performance of ensemble members

The assessment of the deterministic quality of the ensemble members is a first requirement in the development of an EPS. When the quality of the ensemble members is similar, then any member can be weighted equally in the computation of a probabilistic forecasts, i.e. they are assumed to be equiprobable. Once provided this individual quality, then some other properties can be considered (see below). In addition, the ensemble mean should show a better deterministic performance than any individual member in terms of Root Mean Square Error (*RMSE*) (Leith, 1974; Murphy, 1988; Whitaker & Loughe, 1998; Ziehmann, 2000).

http://dx.doi.org/10.5772/55699

As a visual representation, either time-series or evolutions with forecast step for bias and *RMSE* are usually depicted for synoptic parameters (e.g. Z500) for each member and also for the ensemble mean. As an example, Figure 18 shows *BIAS* and *RMSE* evolution with forecast length for the different ensemble members and the ensemble mean.

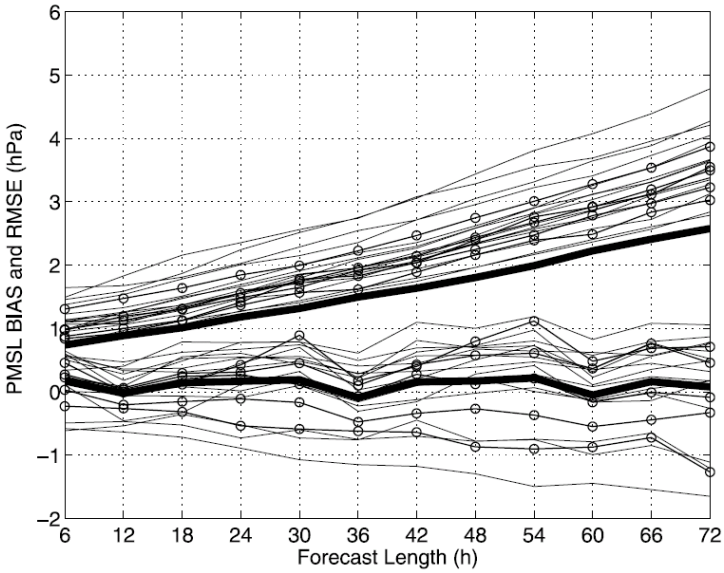

Figure 18. Evolutions with forecast length of mean sea level pressure (*MSLP*) *BIAS* (bottom group) and *RMSE* (upper group) computed for each member (thin lines) and for the ensemble mean (thick line). 'Normal operating' members are highlighted (circles) (García-Moya et al., 2011).

5.2. Large scale flow: Statistical consistency with the observations/analysis

As a probabilistic forecast, an EPS must be statistically consistent with the observations in the large scale flow given the EPS domain is large enough. At this scale, the model analyses of upper-air dynamical fields (e.g. 500 hPa geopotential height, Z500) can be used for comparison, by providing a larger sample and covering the whole integration domain, and so by giving no priority to land areas where the density of observations is higher. Verification against SYNOP/TEMP observations is expected to give worse but qualitatively similar results. This statistical consistency can be assessed in several ways; two methods are shown here: the rank histogram and the spread-error diagram.

On each grid-point, either the analysis or each of the ensemble member values are assumed to be independent realizations of the same atmospheric process and hence equally probable. Here, the rank of the analysis is an integer number according to the position of the analysis value in the sorted list of forecast values. The rank histogram (Anderson, 1996; Hamill & Colucci, 1997, 1998; Hamill, 2000; Candille & Talagrand, 2005) can be used to check if the

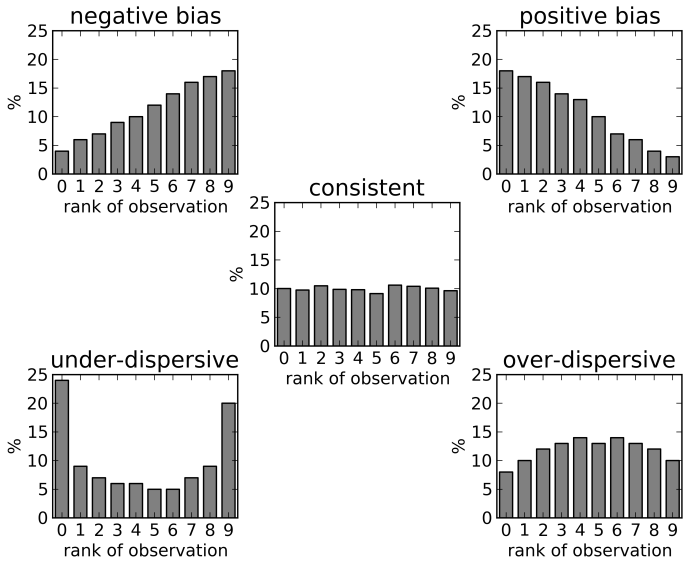

Figure 19. Examples of rank histograms (see text).

verifying observation is statistically indistinguishable from the set of forecast values (*reliable*). Such a system must show an approximately flat-shaped rank histogram (Figure 19 middle). Some outliers ("U" shape, Figure 19 bottom left) would indicate sub-dispersion that are typical in current EPS operational systems, while overdispersion would correspond to the opposite (inverted "U" shape, Figure 19 bottom right). The bias would produce rank histograms with positive (negative) or negative (positive) slope (bias) (see Figure 19 top left and top right, respectively).

Furthermore, the ensemble spread (often measured by the standard deviation with respect to the ensemble mean or the control forecast if possible) and the error of the ensemble (measured by the *RMSE* with respect to the analysis for either the control forecast or the ensemble mean) should show a linear relationship and a similar growth rate with respect to forecast step (Buizza & Palmer, 1997; Whitaker & Loughe, 1998). An EPS is expected to sample the uncertainties of NWP models (ensemble *spread*), as well as to give explicit and quantitative information about the predictability of the atmosphere (represented by the ensemble error). According to this, an ensemble can be statistically consistent (*calibrated*) or, on the other hand, can be underdispersive (quite common in operational ensembles) or overdispersive (e.g. Figure 20).

5.3. Weather parameters: Binary events

For the performance assessment of weather parameters (e.g. precipitation, 2m temperature, 10m wind), with larger variability in space and time, the use of observations is encouraged, as they are not as smooth as upper-air field analyses. In the distributions-oriented framework (for a detailed description see Joliffe & Stephenson, 2003; Wilks, 2006), the performance of an

http://dx.doi.org/10.5772/55699

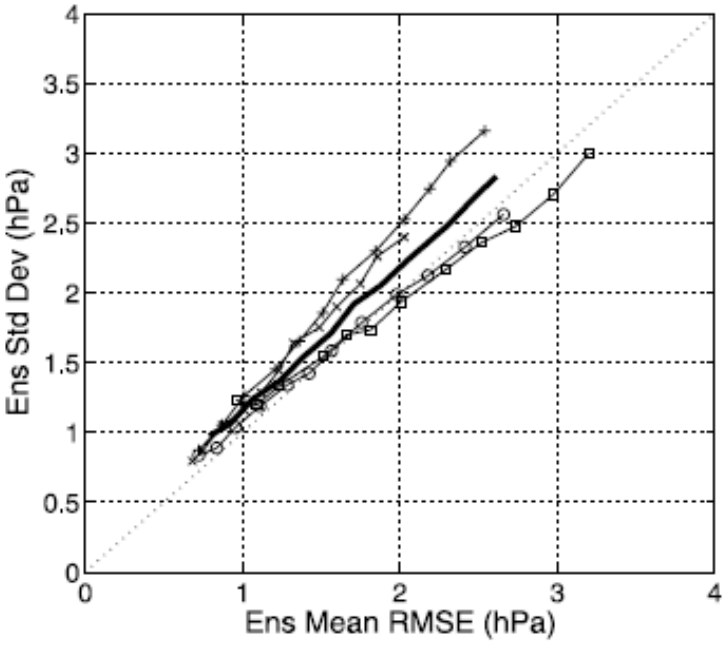

Figure 20. Spread-error diagrams showing five EPS: one of them consistent (solid black), two of them overdispersive (x crosses, + crosses) and two of them underdispersive (circles, squares).

EPS can be done measuring its response to a set of binary events (occurrence / non-occurrence, e.g. to exceed a threshold). The EPS behavior in this context is described by different properties: reliability, resolution, sharpness and discrimination. Brier (Skill) Score together with *ROC* curves provide a framework to give measures to theses properties. Moreover, the benefits of using a forecasting system can be shown with the so-called relative value, a quantity that depends on the forecast user's cost/loss ratio. In this framework, the joint distribution of forecasts and observations gives complete support for the computation of scores.

5.3.1. Formal framework

By considering a binary event X of the parameter x exceeding a threshold t ($\{X: x>t\}$ e.g. rain over 5 mm), we compare a forecast probability p (number of member values exceeding t, p = $\{0/N, 1/N,... N/N\}$) and the corresponding a posteriori observation probability p_0, which is usually taken as p_0 = {0,1}, depending on whether the event took place or not. However, if observational uncertainty was considered, then p_0 could take any value in the interval [0,1]. In this probability space, a natural extension of the *RMSE* is the Brier Score, defined as BS = E $[(p - p_0)^2]$, where E[] is the expectation value over all forecast-observation pairs. BS is negatively oriented and BS=0 if and only if p=p_0 for any pair, while BS=1 indicates the worst possible forecast.

The *joint distribution* (Murphy, 1988) of forecasts and observations can be represented by two distributions that describe completely the system performance: $g(p)$ and $p'(p)$, where $g(p)$ is the forecast probability distribution (relative frequency of forecasts with probability p) and $p'(p)$ gives the conditional observation distribution (relative frequency of forecasts with probability p and for which the event did happen). The expectation values can be computed through a partition in probability space according to the possible forecast probability values, i.e., the number of members (Santos & Ghelli, 2011):

$$E\left[y\right] = E\left[E_p\left[y\right]\right] = \int_0^1 y\, g(p)\, dp \approx \sum_{p=0/N}^{p=N/N} y\, g(p) \tag{6}$$

The base rate p_c=E[$p'(p)$]=E[$E_p[p_0]$] is the frequency of occurrence of the event. Using these two distributions, a decomposition of the *BS* can be done (Jolliffe & Stephenson, 2003; Candille & Talagrand, 2005):

$$BS = E\left[p - p_0^2\right] = E\left[(p - p'(p))^2\right] - E\left[(p'(p) - p_c)^2\right] + p_c(1 - p_c) = B_{rel} - B_{res} + B_{unc} \tag{7}$$

the components are meaningful: *Reliability* (B_{rel}) measures the straight correspondence between probabilistic forecasts p and conditional observation frequencies $p'(p)$, and can be improved by re-labeling of probability intervals (a process that should be called *re-labeling* calibration to avoid confusion). *Resolution* (B_{res}) gives a measure of variability of conditional observations $p'(p)$ around the base rate, and cannot be improved by re-labeling, thus it gives an upper bound for inherent skill. For a perfectly reliable system the reliability component vanishes (p=$p'(p)$ for all cases), and the resolution is equal to the *sharpness*, a measure of variability of the forecast probability distributions, or how often different forecast probabilities occur without taking into account the observations. The uncertainty component (B_{unc}) is the variance of the probabilistic observations p_0 and corresponds to the value of the *BS* using the sample climatology as forecast (i.e. issuing always a forecast probability p=p_c a system is perfectly reliable B_{rel}=0, and shows no resolution B_{res}=0); it depends only on the observations and is usually taken as a reference for the Brier Skill Score (*BSS*), if special care is taken with the interpretation (Mason, 2004). The same decomposition can be applied to the *BSS* (Candille & Talagrand, 2005):

$$\left.\begin{aligned} BSS_{rel} &= \frac{E\left[(p - p'(p))^2\right]}{p_c(1 - p_c)} \\ BSS_{res} &= 1 - \frac{E\left[(p'(p) - p_c)^2\right]}{p_c(1 - p_c)} \end{aligned}\right\} \quad BSS = 1 - BSS_{rel} - BSSs \tag{8}$$

http://dx.doi.org/10.5772/55699

To give a summary of performance measures comprising the response to several thresholds, the Ranked Probability Score (and the corresponding skill score) can be used, either the discrete or the continuous version (Hersbach, 2000).

A complementary measure of ensemble performance is the *discrimination* or ability of a system to distinguish the ocurrence or non-ocurrence of a binary event X given the observations according to Signal Detection Theory (Kharin & Zwiers, 2003). The discrimination is related to the hit rate (H) and the false alarm rate (F) for a given base rate p_c (Candille & Talagrand, 2008):

$$\begin{cases} H(s) = \dfrac{1}{p_c}\int_s^\infty g(p)p'(p)dp \\ F(s) = \dfrac{1}{1-p_c}\int_s^\infty g(p)(1-p'(p))dp \end{cases} \tag{9}$$

As a measure of discrimination, the area A under the Relative Operating Characteristics (*ROC*) curve (H versus F) is often used, with A=0.5 for the sample climatology (no skill) and A=1 for a perfect forecast (Santos & Ghelli, 2011). *ROC* Skill Area (*RSA*) can be used instead: if A is the area under the *ROC* curve, RSA=2A-1 gives values in the interval [-1,1], 1 for a perfect forecast, 0 for no skill and -1 for a potentially perfect forecast after calibration. Discrimination is related to resolution, but they do not measure exactly the same property and, especially if observational uncertainty is present, they can show different and indicative behaviour. While *BSS* is potentially insensitive to extreme events, *RSA* is not (Gutiérrez et al., 2004), whereas *RSA* can be insensitive to some kinds of forecast biases (Kharin & Zwiers, 2003).

Another interesting complementary property, beyond performance, is the Economic *Relative Value* (*RV;* Richardson, 2000). By crossing a *contingency table* (forecast yes/no by occurrence yes/no of the event) with an *expenses matrix* (preventive action yes/no by occurrence yes/no, that includes the cost C of the action and the loss L in case of occurrence), it can be computed the relative economic reduction (*RV*) of using the forecast comparing with the sample climatology. *RV* depends on the base rate p_c and also on C and L, i.e. it depends also on the user.

5.3.2. Visual presentation

The properties described above can be visualized in several ways. *Sharpness histogram*: *g(p)* distribution is put in a histogram along probability intervals. A predicting system with good sharpness would issue forecast probabilities close to 0 and 1. Sharpness is often plotted as an inset on the attributes diagram. *Attributes diagram*: *p'(p)* distribution is plotted on the Y axis against probability intervals p on the X axis. Several straight lines are plotted as reference: the diagonal (representing perfect reliability), the no-resolution line (corresponding to the sample climatology as forecast), and the no-skill line (forecasts with no skill w.r.t. the climatology, i.e. B=B_{unc}). Figure 21 (left) illustrates this. Some examples of forecasting systems are idealized in Figure 21 (right).

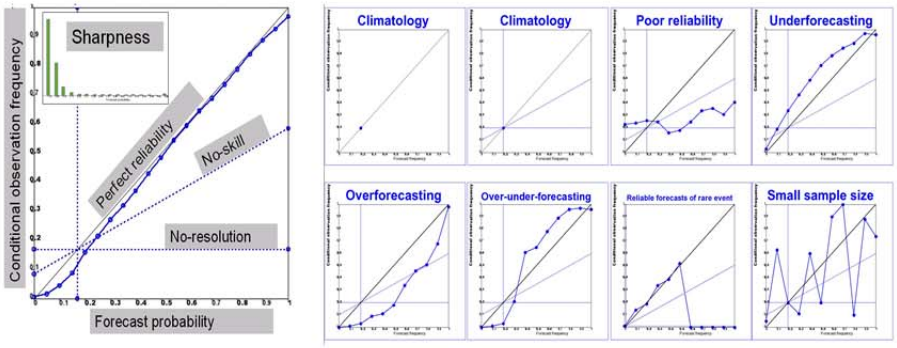

Figure 21. left) Attributes diagram for an almost perfectly reliable forecasting system, showing the sharpness histogram; (right) Attributes diagrams for idealized examples of forecasting systems.

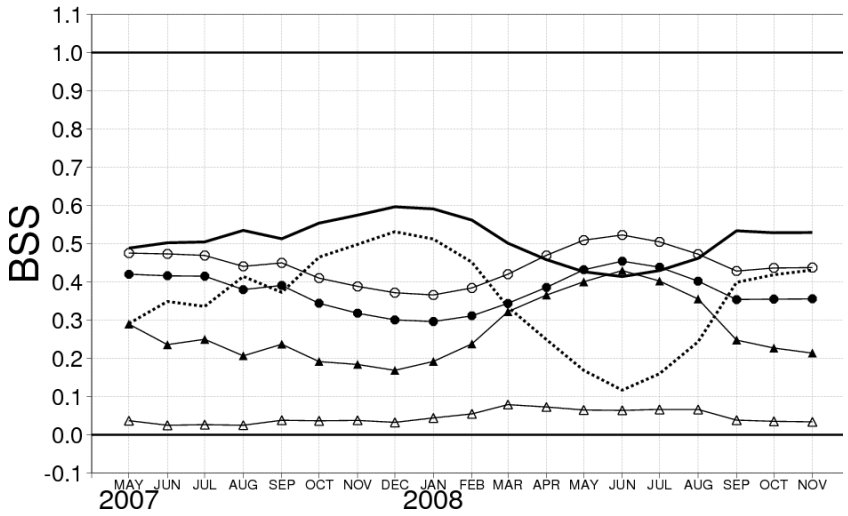

Figure 22. Time series for two different systems 24 h accumulated precipitation forecast (T+30 to T+54), of *BSS* (solid, dotted) and its components BSS_{rel} (triangles) and BSS_{res} (circles). Santos and Ghelli (2011)

BSS decomposition time series: *BSS* (and its components BSS_{rel} and BSS_{res}) time series are plotted in curves. As *BSS* is positively oriented (the larger the BSS, the better is the performance) and its components are not, special care must be taken (see Figure 22) (Santos and Ghelli, 2011). *ROC curves*: the hit rate (*Y* axis) is plotted against the false alarm rate (*X* axis) (see Figure 23 left). Here a deterministic forecast is compared to an EPS. *RV envelopes*: the *RV* can be plotted on the *Y* axis, the cost-loss ratio C/L on the *X* axis and provide one curve for a deterministic model. For an *N* members ensemble, *N* curves can be plotted (we can plot *RV* for any probability interval in the partition described above), or eventually, the envelope that covers all the

http://dx.doi.org/10.5772/55699

N curves. The user can decide on the *C*/*L* intervals for optimal use of the forecasting system. In this sense, an EPS can be also compared with a deterministic model (see Figure 23, right).

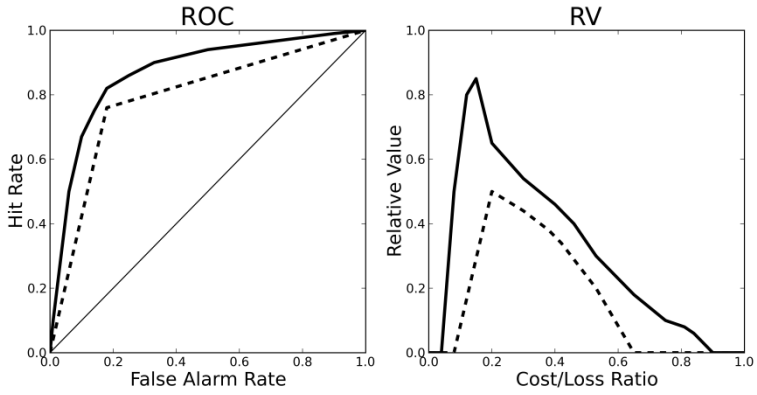

Figure 23. Comparison of deterministic (dash) and EPS (solid) precipitation forecast performance, using empirical *ROC* curves (left) and relative value envelopes (right).

5.4. Verification issues and prospective

Computational resources, object-oriented programming languages and data-base improvements give a boost to forecast verification. In the last decades, the forecast verification community has started to address some important issues that have an impact in the interpretation of verification scores and introduction of new conceptions. Either EPS specific or not, some of these issues are of large interest and hence are introduced here.

5.4.1. Pooling versus stratification

To compute statistically significant scores, samples must be large (many fc-ob pairs) and the corresponding significance tests should be applied (e.g. t-Student). On the other hand, mixing non-homogeneous sub-samples (e.g. different seasons) can lead to misleading performance information. In this context, the dimensionality problem (Murphy, 1988) can be a critical stumble in practice (Candille & Talagrand, 2008): computing completely consistent (from the strictly mathematical point of view) scores can turn out to be an infeasible task, and this issue often leads to a practical compromise: large samples are created without mixing different ones, according to possibilities. E.g.: the splitting of seasonal behavior that could be hidden in the overall average.

5.4.2. Flow-dependent verification

Flow-dependent sample stratification can improve other traditional ways of stratifying (e.g. seasonal), and nowadays can be tackled with clustering techniques (Ferranti & Corti, 2010).

5.4.3. Sampling uncertainty

Given the dimensionality problem and pooling-stratification compromise, the available fc-ob pairs are, in practice, relatively small samples from the all possible realizations of the model and observing systems. Thus, the scores computed are only sample measures of the population quantities, and there is a sampling uncertainty related to this process (PaiMazumder & Mölders, 2009). Any verification report should include this uncertainty, with error bars, confidence intervals, etc. (see e.g. Efron & Tibshirani, 1997; Bradley et al., 2008).

5.4.4. Spatial scales of forecasts and observations

Spatial scales are a key point (see examples above). For example, the double penalty (Ebert & Gallus, 2009) is a well-know related issue. The relatively recent development of new methods that account for spatial patterns, e.g. CRA (Ebert & Gallus, 2009), MODE (Davis et al., 2009), SAL (Wernli et al., 2008) are still under research, but show promising results and can lead to a framework of diagnostic verification (closer to subjective verification in the sense that provides information that can better help model developers and weather forecasters). For a comprehensive overview, see (Gilleland et al., 2009).

5.4.5. Extreme and severe weather

Extreme and severe weather are often introduced together, but they are not the same. Extreme events are rare events, with low base rates and belong to the tail of the corresponding climatological distributions. Severe events are those that have an impact (human and material) on society. Severe weather verification must include extra information from outside the meteorological context, whereas verification of extreme events is still in an early stage (Ferro, 2007a; Casati et al., 2008), and some alternative scores are under research, e.g. the Extreme Dependency Score (EDS; Stephenson et al., 2008; Ghelli & Primo, 2009).

5.4.6. Observational uncertainty

In forecast verification it has been traditionally assumed that the observation error is negligible when compared with the forecast error. This assumption can be consistent for longer forecast ranges, when the forecast error is much larger than the observation uncertainty. Several studies have extended the verification problem to a more general framework, in which observations are described together with their uncertainty. They show sometimes a surprising result: traditional scores generally underestimate EPS performance (e.g., Saetra et al., 2004; Candille & Talagrand, 2008; Santos & Ghelli, 2011).

5.4.7. Ensemble size

Differences in ensemble size can have an impact on performance assessment (e.g. compare a 16 members EPS with a 51 members EPS). The difference in size would, in principle, give better performance to the larger EPS a fact that should be at least taken into account. This issue is addressed by various authors (e.g., Buizza & Palmer, 1998; Ferro, 2007b; Ferro et al., 2008).

http://dx.doi.org/10.5772/55699

6. Statistical post-processing

EPS evidence systematic errors like do the deterministic NWP models. Calibration is the process of correction of the ensemble PDF to adjust it to the actual (and unknown) forecast uncertainty. The main point of calibration techniques is to use the information of the former prediction's skill to correct the current probabilistic forecast.

Different methodologies have been proposed recently to build calibrated probabilistic forecasts from ensembles, including Bayesian Model Averaging (Raftery et al., 2005), Logistic Regression (Hamill et al., 2008) and Extended Logistic Regression (Wilks, 2009), Non-homogeneous Gaussian Regression (Gneiting et al., 2005b) and Ensemble Dressing (Roulston & Smith, 2003; Wang & Bishop, 2005).

6.1. Bayesian Model Averaging (BMA)

Bayesian Model Averaging (BMA) is a statistical post-processing method that generates calibrated and sharp predictive PDFs from EPS (Raftery et al., 2005). The BMA predictive PDF of a weather variable is a weighted average of PDFs centred on the individual bias-corrected forecasts. The weights reproduce the predictive skill of the ensemble member over a training period.

If the forecast errors are approximately Gaussian distributed such as surface temperature or sea level pressure, BMA can be applied (e.g. Raftery et al., 2005; Wilson et al., 2007). For non-Gaussian error distributions using a mixture of skewed PDFs allows to extend the BMA methodology to this kind of weather parameters; A combination of point mass at zero and a power-transformed gamma distribution, for instance, can be applied to quantitative precipitation (Sloughter et al., 2007) and a mixture of gamma distributions with different shapes and scale parameters can be used to improve wind speed probabilistic forecasts (Sloughter et al., 2010).

The BMA predictive PDF is a summation of weighted PDFs of each individual ensemble member (Leamer, 1978; Kass & Raftery, 1995; Hoeting et al., 1999):

$$PDF\left(y \mid f_1,\ldots,f_m;\theta_1,\ldots,\theta_m\right)=\sum_{i=1}^{m} w_i PDF_i\left(y_i \mid f_i,\theta_i\right) \tag{10}$$

Where f_i is the ensemble member deterministic forecast, y represents the forecasted variable, and $_i$ are the characteristic parameters of the ith individual PDF from the ith ensemble member. Each of these individual PDFs associated to each ensemble member is weighted based on the ensemble member's relative performance during the training period. The weights w_i are probabilities, i.e. non-negative and add up to 1. The BMA weights w_i and the parameters $_I$ are estimated by maximum likelihood (Wilks, 2006) using the training data. This estimate cannot be done analytically so an expectation maximization (EM) iterative algorithm is used (Dempster et al., 1977; McLachlan & Krishnan, 1997).

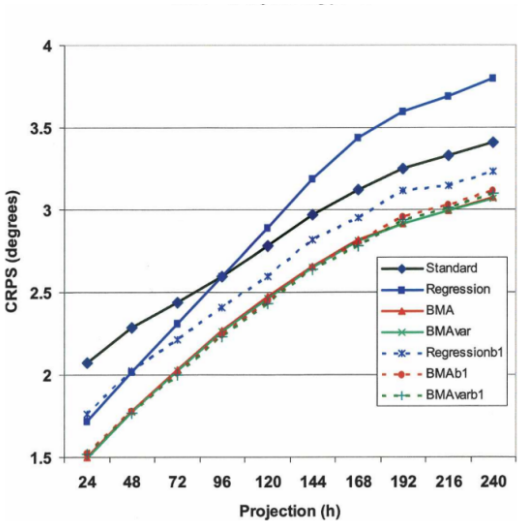

Figure 24. CRPS of surface temperature ensemble predictions at 21 ground stations over Canada. Curves correspond to raw ensemble and the six other ensembles corrected by: Linear regression, BMA, a variation of BMA, and the same three methods with the bias previously corrected. 40 days are used as training period., From Wilson et al. (2007). See their article for details. © American Meteorological Society. Reprinted with permission.

Figure 24 illustrates the potentiality of BMA as a method for ensemble calibration. Continuous ranked probability score (CRPS) performance score (Hersbach, 2000) is represented for various ensemble predictions of surface temperature. CRPS index is understood to be for a probabilistic forecast the equivalent of the mean absolute error for a deterministic forecast. The different curves correspond to the raw ensemble and to the six ensembles calibrated using different statistical techniques. It is straightforward to see that the ensembles corrected with BMA perform better than any of the others.

6.2. Logistic Regression (LR) and Extended LR

The logistic regression and extended logistic regression techniques are described in detail in Wilks (2009). Logistic Regression (LR) approximates the Cumulative Distribution Function (CDF) of the predicted parameter y by the following equation (Wilks, 2009):

$$CDF(q) = PDF(y \le q) = \frac{e^{f(x)}}{1 + e^{f(x)}} \tag{11}$$

Where q is a selected prediction threshold and:

$$f(x) = b_0 + b_1 x_1 + \cdots + b_n x_n \tag{12}$$

http://dx.doi.org/10.5772/55699

Being $\{x_{1,\ldots,} x_n\}$ the regression predictors and $\theta = \{b_0, b_{1,\ldots,} b_n\}$ the unknowns to be estimated during the training process. Equation (11) has a characteristic *S* shape with values bounded on the $0 < CDF(q) < 1$ interval. The name logistic comes to the fact that the regression equation is linear on the logistic scale:

$$\ln\left[\frac{CDF(q)}{1-CDF(q)}\right] = f(x) \tag{13}$$

Typical predictors for LR when calibrating ensemble predictions are ensemble mean, ensemble spread or a function of them (Hamill et al., 2008). As thresholds q, using the representative climatological quantiles of the meteorological parameter y ensures a statistical uniformity in the process of regression.

According to Wilks (2009) θ unknowns are generally estimated using maximum likelihood (Wilks, 2006), but other estimation techniques could give better performance, for example the minimization of the continuous ranked probability score (Hersbach, 2000).

By construction Equation (13) is fitted separately for every threshold q and this fact involves several problems. We consider for example the parameter precipitation and two thresholds, q_1 = 2mm and q_2 = 10mm. After the training we have two different regression equations for each threshold, $f_1(x)$ and $f_2(x)$, which in general are not parallel. The non-parallelism of the functions implies that for some values of the predictors $\{x_{1,\ldots,} x_n\}$ these curves will cross leading to the unrealistic result of $CDF(q_1) > CDF(q_2)$. Another problem arises when we want to estimate the *CDF* of a threshold for which regressions have not been fitted. This process requires some kind of interpolation of *CDFs* which is not statistically coherent. Finally, the more equations are to be fitted the more unknowns have to be estimated.

To overcome these problems, Wilks (2009) proposed a new approach to Equation (11) that consists of including a function $g(q)$ in the exponent which increases with threshold q:

$$CDF(q) = PDF(y \le q) = \frac{e^{f(x)+g(q)}}{1+e^{f(x)+g(q)}} \tag{14}$$

Thus, a unique regression estimation for any value of q is needed, which implies the parallelism of the functions $f(x)$ for the different thresholds (the unknowns $\{b_0, b_{1,\ldots,} b_n\}$ are always the same). This approximation is known as Extended Logistic Regression (ELR).

It is important to point out as an advantage that LR (and ELR) has no statistical restriction to be used with non-Gaussian parameters such as precipitation or wind. At the same time this technique can be applied to ensembles whose members are non-distinguishable.

As an example of ELR, Schmeits and Kok (2010) calibrated ECMWF ensemble predictions of precipitation over an area that covers Netherlands (see the article for details). After studying the performances of different shapes for $g(q)$ and the predictors, they select :

$$g(q) = b_2\sqrt{q}$$
$$x_1 = \overline{\left(\sqrt{x}\right)}_{ens} \tag{15}$$

Being x_1 the ensemble mean of the square root of the predicted precipitation. Then the equation to be regressed is:

$$f(x) + g(q) = b_0 + b_1\overline{\left(\sqrt{x}\right)}_{ens} + b_2\sqrt{q} \tag{16}$$

Figure 25 represents a reliability diagram which compares the forecast probabilities of having precipitation lower or equal than 5 mm with the observed frequencies of this event. For a perfect reliable forecast all points would be in the diagonal, so in this case ELR calibration clearly improves the performance of the raw forecast.

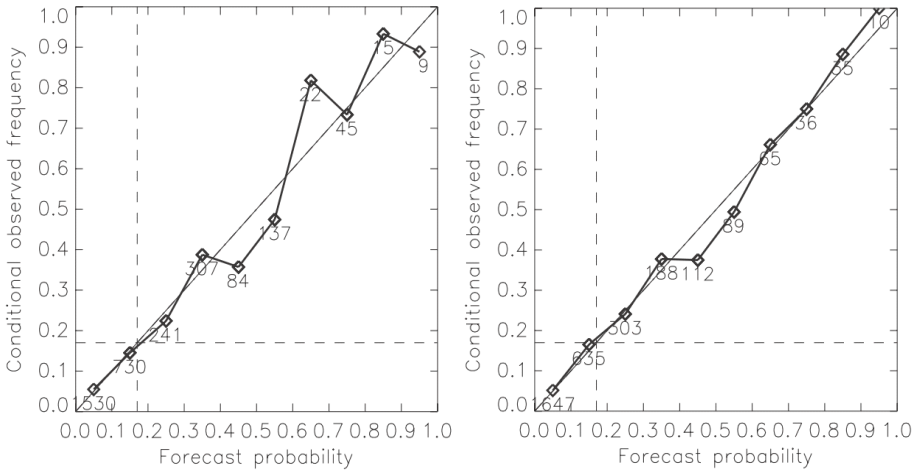

Figure 25. Reliability diagrams of H+126 area mean precipitation forecasts for the raw ensemble (left) and the ELR calibrated ensemble (right). In this case threshold q = 5 mm. From: Schmeits & Kok (2010). © American Meteorological Society. Reprinted with permission.

6.3. Non-homogeneous gaussian regression

The non-homogeneous Gaussian Regression (NGR) technique was proposed by Gneiting et al. (2005b). In its general form, the predictive PDF estimated by NGR is assumed to be a perfect Gaussian with the mean value being a bias-corrected weighted average of the ensemble members forecasts and the variance a linear function of the ensemble variance. That is (Gneiting et al., 2005b):

http://dx.doi.org/10.5772/55699

$$PDF(y) = \frac{1}{\sqrt{2\pi(c+ds^2)}} e^{-\frac{\left[y-(a+b_1x_1+\cdots+b_kx_k)\right]^2}{2(c+ds^2)}} \quad (17)$$

Here y is the weather quantity to be predicted, s is the ensemble spread and $\{x_1, \ldots, x_k\}$ are the k ensemble predictions of parameter y. $\theta = (a, b_1, \ldots, b_k, c, d)$ are the unknowns of this expression that have to be estimated by regression using the training data, which consist in a series of former forecast-observation pairs. The term non-homogeneous refers to the fact that the variances of the regression errors are not the same for all the values of $\{x_1, \ldots, x_k\}$ (they depend on s) as it is assumed in linear regression.

This way of representing the predictive PDF allows a natural understanding of the regression coefficients. Coefficient a is a bias-correction of the ensemble weighted mean. The weights $\{b_1, \ldots, b_k\}$ can be negative but for an easier interpretation Gneiting et al. (2005b) recommended constraining them to be non-negative which is done during the training process. On one hand side, they represent the performance of the ensemble members over the training period, with respect to the other members. On the other hand, they also reflect the correlations between ensemble members. Gneiting et al. (2005b) showed how a five members ensemble with three members using the same global data as initial and boundary conditions (so highly correlated) is automatically reduced to a three members ensemble after NGR calibration leaving only non-zero weights for the members that use different data as initial conditions. Variance coefficients c and d are constrained to be non-negative and they are a measure of the spread-skill relationship. For large values of d NGR variance is correlated to ensemble variance (s^2) so a significant correlation of the spread with the skill of the ensemble weighted mean is obtained. If spread and skill are independent of each other, d values will be negligible and it is c what represents the variance of the NGR calibrated mean.

Compared to other techniques (e.g. BMA; Raftery, 2005), NGR has the advantage that can be applied to ensembles whose members are non-distinguishable, such as the ECMWF ensemble prediction system (Molteni et al., 1996). In this case, NGR is simplified by constraining the $b_1 = \ldots = b_k$ coefficients to be equal, which at the same time agrees with the assumption of equiprobability of members. Now the analytical PDF (17) is reduced to:

$$PDF(y) = \frac{1}{\sqrt{2\pi(c+ds^2)}} e^{-\frac{\left[y-(a+bx_m)\right]^2}{2(c+ds^2)}} \quad (18)$$

Where x_m is the ensemble mean.

Due to the Gaussian shape of the analytical expression (Eq. 17), this technique is expected to be especially useful for weather parameters that have Gaussian distributions such as temperature or pressure. Figure 26 represents a real experiment by Hagedorn et al. (2008) where GFS (Toth & Kalnay, 1997), ECMWF and a multi-model (combining GFS and ECMWF) ensemble

predictions of surface temperature are calibrated using NGR all else being equal. Continuous Ranked Probability Skill Score (CRPSS) is used as a performance measure for probabilistic forecasts (Jolliffe & Stephenson, 2003). The higher its values (closer to 1), the better the probabilistic forecast will be. When CRPSS reaches 0 it means that the probabilistic forecast has the same skill than the climatology (Hagedorn et al., 2008; see for details). In this case, it is clear the benefit of calibrating the ensembles with NGR.

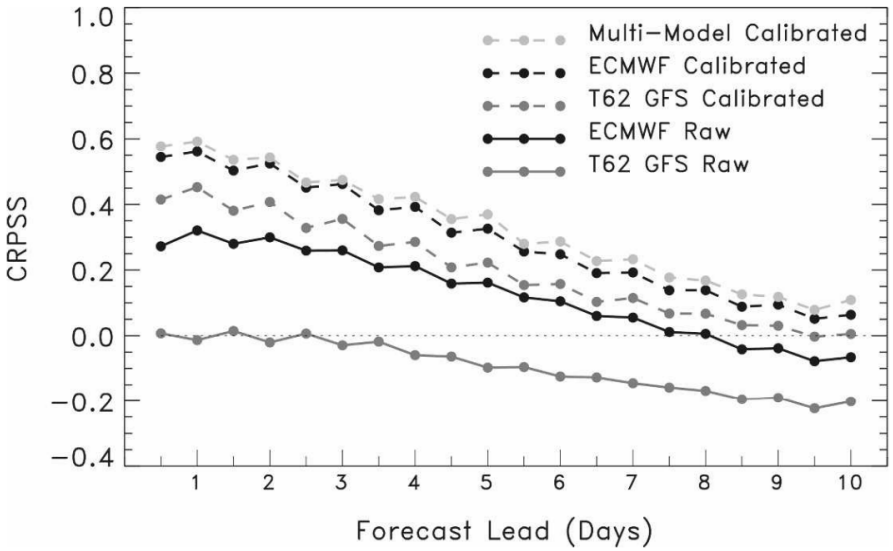

Figure 26. CRPSS of surface temperature forecasts with and without calibration. From: Hargedorn et al. (2008). © American Meteorological Society. Reprinted with permission.

A classical and widely extended technique for estimating the θ unknowns is maximum likelihood (Wilks, 2006). Nevertheless Gneiting et al. (2005b) demonstrated, for NGR probabilistic forecasts of temperature and surface pressure, that estimating θ by minimization of the continuous ranked probability score (Hersbach, 2000) gives clearly a better calibration of the PDF.

The adequate length of the training period for an operational approach is not unique. Raftery et al. (2005) showed that by using the previous 25 days the prediction intervals are the narrowest maintaining the right coverage of the verification (see the article for details). However, Wilks and Hamill (2007) used 45 days. An experimental study of the optimal length for one or more specific locations is desirable.

6.4. Ensemble dressing

The dressing technique is a statistical post-processing technique based on combining each member of a dynamical ensemble with its own statistical error ensemble.

http://dx.doi.org/10.5772/55699

Roulston & Smith (2003) proposed the use of a simple resampling scheme called *best member* method: individual members of an ensemble are *dressed* with an error distribution derived from the error made by the *best* member of the ensemble. The best member is defined as the member that is closer to the verification and the uncertainty of it is the one that is added to the rest of members. Identification of the best member is performed by means of multivariate forecasts although only univariate forecasts are dressed. The number of forecast variables required is estimated by looking at the fraction of the false best members (FBM). These FBM are defined using a distance on the vector space of the verification. If the N ensemble members are described as d-dimensional vectors, x_i *(i=1,…,N)* and y is the verification, the normalized distance is defined as (Roulston and Smith, 2003):

$$R_{i,d}^2 = \sum_{k=1}^{d} \frac{\left(x_{i,k} - y_k\right)}{\sigma_k^2} \tag{19}$$

Where d is the number of forecast variables being considered and $_k$ is the standard deviation of the kth component of the forecast vector. The best member is the one which the minimum $R_{i,d}^2$ although new additional variables are included. FBMs are the ones whose minimum distances are not maintained when new variables are added. The fraction of FBMs computed using the previous historic forecasts allows for a minimum number of variables required to obtain the best member to dress the dynamical ensemble. The best member error is also determined from historical ensemble forecasts by computing the differences between the best member and the corresponding verification.

Wang and Bishop (2005) showed by stochastic simulations that the best-member method can lead both to underdispersive or overdispersive ensembles. In addition to this, Wilks (2006) demonstrated that the dressed ensemble cannot be reliable. In order to alleviate these problems a new multivariate dressing method based on the second moment constraint is proposed. Ensemble bias is removed before building training statistics for the dressing kernel assuming that each ensemble is drawn for stochastic process. To dress the ensemble, statistical perturbations ε are added to each ensemble member. The covariance matrix Q is defined as (Wang & Bishop, 2005):

$$Q = <\varepsilon\varepsilon^T> = E\Omega E^T \tag{20}$$

Where the columns of E contain the eigenvectors of Q and the diagonal matrix contains the corresponding eigenvalues. Positive eigenvalues indicate that the ensemble is underdispersive in the directions of the corresponding eigenvectors and thus dressing is necessary. The Q matrix can be expressed as a function of the ensemble member forecasts and the verification values. The new dressing perturbation generator is defined as (Wang & Bishop, 2005):

$$\varepsilon = c_1 e_1^+ + c_2 e_2^+ + \cdots + c_l e_l^+ \tag{21}$$

Where e_i^+, $i = 1, 2,..., I$, are the eigenvectors corresponding to the positive eigenvalues. The coefficients c_i are univariate random variables generated from a normal distribution with mean equal to zero and variance equal to the ith positive eigenvalue of Q.

The comparison of the original best-member dressing method with the second moment constraint dressing method confirms that the spread of the best-member dressed ensemble is indeed underdispersive, or even becomes overdispersive, depending on factors such as the undressed ensemble size, how underdispersive the undressed ensemble is and the nature of the subspace from which the best member is identified. On the other hand, for underdispersive ensembles, the second moment constraint dressing kernel correction always returns about the right amount of dispersion.

Although underdispersion is a common characteristic of an EPS, some variables have an overdispersive behavior (Feddersen & Andersen, 2005). Fortín et al. (2006) proposed to dress and weight each member differently to improve the reliability of the forecast and to correct the variable under or overdispersed. This method is very similar to BMA (Raftery et al., 2005) and today has only been applied to one-dimensional variables.

7. A brief description of some current state-of-the-art ensemble prediction systems

The Australian Bureau of Meteorology (BoM), Brazilian Centro de Previsao do Tempo e Estudos Climatico (CPTEC), China Meteorological Administration (CMA), the European Centre for Medium-Range Weather Forecasting (ECMWF), Japan Meteorological Agency (JMA), Korea Meteorological Administration (KMA), Meteorological Service of Canada (MSC), Météo-France (MF), UK Met Office (UKMO) and US National Centres for Environmental Prediction (NCEP), among others, run ensemble prediction systems.

The ECMWF-EPS is a global ensemble that is optimized for the medium range. It uses the singular vector technique (Ensemble Data Assimilation is under research and is starting to be used operationally together with SV) for providing the set of initial perturbations, as well as stochastic parameterizations to account for model errors. The ECMWF-EPS comprises 51 members with 62 vertical levels and a spectral horizontal resolution of T639.

In the NCEP ensemble the initial perturbations are obtained by the Ensemble Transform with Rescaling (ETR) technique. It also uses stochastic perturbations to account for model errors. It runs 20 members with 28 levels and a spectral horizontal resolution of T126.

The MetOffice ensemble, called MOGREPS, works with 24 members. It uses the ETKF (Ensemble Transform Kalman Filter) technique with scaling of perturbations using radiosonde and ATOVS observations (Bowler et al., 2008). The horizontal resolution is 0.83 degrees in longitude and 0.56 in latitude, and the number of vertical levels is 70.

The Japan Meteorological Agency (JMA) ensemble has 51 members. It uses singular vectors for the calculation of the initial perturbations and stochastic representation of physical

http://dx.doi.org/10.5772/55699

parameterizations for accounting model error. The number of vertical levels is 60 levels and the spectral horizontal resolution is T319.

A source of information for studying global ensembles is the TIGGE project, which is a key component of the THORPEX Interactive Grand Global Ensemble, a World Weather Research

Programme for improving the accuracy of high-impact weather forecasts. In the TIGGE project, the forecasts of 10 global ensembles are archived, permitting the comparison of methods and results.

Limited area ensembles are developed for higher resolutions (nowadays from 2 km up to 25 km) and shorter time ranges (from 18 to 72 hours) than those of global ensembles. When a model can explicitly resolve convection (due to its characteristics and high resolution configuration) it can represent more realistically typical precipitation patterns in the forecast field. However, convection forecasts (as well as other small scale processes) are very limited by their deterministic predictability (which is small due to its chaotic behaviour). Therefore, even in the short forecast range of only 24 hours, the prediction of details in the convection (such as location and timing of a thunderstorm) are usually very uncertain. Limited area ensembles can add information to deterministic high resolution forecasts and for this reason many operational weather centres are developing limited area ensembles.

Limited area ensembles running in operational centres are based on high resolution non-hydrostatic models, such as ALADIN (developed and maintained by a consortium of 16 National Meteorological Services, led by Météo-France), COSMO (developed by a consortium of seven NMS, led by Deutscher Wetterdienst), WRF (developed in the United States of America by NCAR, NOAA and others), HARMONIE (a model which shares code with ALADIN, developed by a consortium of 10 NMS) and the Unified Model (MetOffice). These ensembles can have an assimilation cycle which uses a wide class of meteorological observations, in some cases including radar data trying to represent as much as possible actual precipitation processes. They need lateral boundary conditions typically coming from global ensembles or coarser limited area ensembles.

Just to mention two examples of limited area ensembles, we resume below the characteristics of the COSMO-DE ensemble (Deutscher Wetterdienst) and the AEMET-SREPS. Other operational limited area ensembles are: the Norwegian targeted EPS LAMEPS (Frogner & Iversen, 2002), the Hungarian LAMEPS based on ALADIN (Hágel & Horányi, 2007), the multi-model GLAMEPS (ALADIN and HIRLAM consortium) and the limited area version of MOGREPS (MetOffice).

The COSMO-DE Ensemble is based on the convection resolving model COSMO-DE. It produces 2.8 km grid forecasts up to 18 hours, runs every 3 hours and assimilates estimated precipitation rates from RADAR.

The AEMET-SREPS uses the multi-model and multi-boundaries techniques for sampling initial, lateral conditions and model errors. It uses five global models for initial and lateral conditions and five limited area models running with every lateral and initial conditions (MM5, UM, HIRLAM, COSMO and HRM) thus producing 25 members. The horizontal

resolution is 25 km and produces forecasts up to 72 hours twice every day. It covers a wide area (includes North Atlantic, Europe and North of Africa).

8. Conclusions and future directions

The most reliable and skilful theoretical forecasts from the current observed state of the atmosphere can be obtained through a Probability Distribution Function (PDF) which describes a comprehensive set of possible future states. The only practically feasible methodology to assess a forecasted PDF is using an Ensemble Prediction System (EPS), that is, a PDF sample of different but equally plausible Numerical Weather Prediction (NWP) forecasts (EPS members). Furthermore from the practical point of view EPS forecasts have been showed to be more reliable and skilful than a forecast from one single NWP model, even when the latter has a higher resolution.

This better performance is due to the fact that the set of non-linear equations which describe the future evolution of the atmosphere have a chaotic behaviour. This means that any uncertainty in the prediction process like two slightly different initial states could grow and lead to quite significantly distinct forecasted states. As a consequence, the predictability associated to any forecasted atmospheric state is always spatially and temporally limited but depending in each forecast on the uncertainty magnitude and the particular atmospheric situation.

The sources of errors and uncertainties which limit the predictability are mainly due to: a.) inaccuracies in the initial atmospheric state, estimated from available observations with their associated observational error and limited representativeness and imperfect assimilation systems, b.) inadequacies of the NWP models, related to dynamical NWP model formulation and physical parameterizations and c.) for Limited Area Model (LAM) EPSs, approximations and errors from Lateral Boundary Conditions. So any reliable and skilful EPS has to take into consideration all of these uncertainty sources by using different methodologies. Some of these methodologies are, for instance (and respectively): a.) singular vectors, bred vectors, Ensemble Transform Kalman Filter (ETKF) and Ensemble Data Assimilation (EDA), b.) multi-model, multi-physics or multi-parameterizations, multi-parameters and stochastic parameterizations, and c.) multi-boundaries.

In addition, as any forecasting system, EPS quality and value, that is the overall performance, has to be evaluated through objective verification, assessing the necessary and complementary set of properties (with the corresponding tools): consistency (rank histogram), reliability (spread-error and attributes diagrams), resolution (resolution component of Brier Score), discrimination (Relative Operating Characteristic curves), sharpness (Sharpness Histogram), skill (Brier Skill Score) and relative value (Relative Value Diagram). The main goal of verification, apart from assessing EPS forecast performance, is that EPS developers can be leaded to what and how to improve the EPS.

A number of EPS products, which take into account the forecast probabilities and the predictability concept, can serve the forecast guidance. The EPS products can be raw ones (e.g. stamps,

http://dx.doi.org/10.5772/55699

plumes or spaghettis) or derived (e.g. ensemble mean and spread charts, probability and percentile maps, EPS-grams, clusters and extreme forecast indexes). Before the products production it would be advisable to calibrate the EPS in order to remove its systematic errors. Some statistical post-processing techniques for ensemble calibration are Bayesian Model Averaging, Logistic and Extended Logistic Regression, Non-homogeneous Gaussian Regression and Ensemble Dressing.

Regarding to future scenarios, the weather forecast process is expected to improve at all temporal scales, from the first hours to the climatic scales, and to be done at finer spatial scales, due to the fact of having more and better observations, better NWP models, increased supercomputer resources, etc. Even though it is expected a reduction of errors and uncertainties, they will be always present and limit the predictability. This means that the majority of ideas and methodologies explained in this chapter are going to be useful for the next generation of weather forecasting systems, although new ones are expected to be developed. Then some of the future directions of work are outlined.

From the point of view of the current experience in EPS development, the multi-model and multi-analysis (from independent Global NWP models) approach, has showed to have better performance than any theoretical methodology based on a single model. This fact means that there is not enough knowledge about the different model uncertainties and that there can even be other unknown sources of error. Anyway, the latter approach is expected to be used intensively in the next EPS generations and even to overcome the former approach.

On the other hand, the better the EPS performance seems to be, the more number of error sources are considered and even the more methodologies are used together. Thus, future EPS developments are going to follow this line, partly because it increases the spread counteracting the common EPS shortcoming, the underdispersion. Anyway particular attention has to be paid not to increase spuriously the EPS spread, that is, without increasing the skill.

As it has been mentioned before, other EPS developments will come from having better assimilation techniques, better generation of initial conditions and better methodologies to tackle model errors and uncertainties. This fact means that, in a foreseeable future, EPSs will become more complex.

Finally in the current and the next decade there will be an important increase in the horizontal and vertical resolutions of the EPSs linked to the NWP model developments for smaller grid spacing. Thus, spatial resolutions of the next Global and LAM EPSs generations are going to be, respectively, close to the non-hydrostatic scale (e.g. 8-16 km), and about the meso-gamma or convection-resolving scale (e.g. 1-4 km). One consequence of this will be that verification will have to evolve to an objected-oriented way (e.g. SAL or MODE techniques). Because of uncertainties grow faster as the resolved scales are smaller, due to a more intrinsic chaotic nature, another consequence will be that the only feasible methodology to forecast the weather at these scales will be ensemble forecasting.

Author details

Alfons Callado, Pau Escribà, José Antonio García-Moya, Jesús Montero, Carlos Santos, Daniel Santos-Muñoz and Juan Simarro

Agencia Estatal de Meteorología (AEMET), Spain

References

[1] Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9, pp. 1518–1530.

[2] Andersson, E. & Coauthors. (1998). The ECMWF implementation of the three-dimensional variational assimilation (3D-Var). III: Experimental results. *Quarterly Journal of the Royal Meteorological Society*, 124, pp. 1831-1860.

[3] Athens, R. & Warner, T. (1978). Development of hydrodynamic models suitable for air pollution and other mesometeorological studies. *Monthly Weather Review*, 106, pp. 1045-1078.

[4] Bishop, C. H., Etherton, B. J. & Majumdar S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129, pp. 420–436

[5] Bowler, N. E. (2008). Accounting for the effect of observation errors on verification of MOGREPS. *Meteorological Applications.,* 15, pp. 199-205.

[6] Bradley A. A., Schwartz S. S. & Hashino T. (2008). Sampling Uncertainty and Confidence Intervals for the Brier Score and Brier Skill Score. *Weather and Forecasting*, 23:5, pp. 992-1006.

[7] Buizza, R. & Palmer, T. (1995). The singular vectors structure of the atmospheric general circulation. *Journal of Atmospheric Sciences*, 52, pp. 1434-1456.

[8] Buizza, R. & Palmer, T. (1997). Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF Ensemble Prediction System. *Monthly Weather Review*, 125, pp. 99-119.

[9] Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A. & Wedi, N. (1998). Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 124, pp. 1935-1960.

[10] Buizza, R., Miller, M. & Palmer, T. (1999). Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 125, pp. 2887-2908.

[11] Buizza, R., Leutbecher, M., & Isaksen, L. (2008). Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 134, pp. 2051-2066.

[12] Candille, G., & Talagrand, O. (2005). Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131, pp. 2131–2150.

[13] Candille, G. & Talagrand, O. (2008). Impact of observational error on the validation of ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 134, pp. 959-971.

[14] Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G. & Mason, S. (2008). Forecast verification: current status and future directions. *Meteorological Applications*, 15, pp. 3–18. doi: 10.1002/met.52.

[15] Charney, J. G., Fjørtoft, R. & von Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. *Tellus*, 2, pp. 237-254.

[16] Cherubini, T., Ghelli, A., & Lalaurette, F. (2002). Verification of precipitation forecasts over the Alpine region using a high-density observing network. *Weather and Forecasting*, 17, pp. 238–249.

[17] Clark, A. J., Gallus, W. A., Xue, M. & Kong, F. (2009). A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Weather and Forecasting*, 24, pp. 1121–1140.

[18] Daley, R. (1991). *Atmospheric Data Analysis*. Cambridge University Press, Cambridge.

[19] Davis, C. A., Brown, B. G., Bullock, R. & Halley-Gotway, J. (2009). The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program. *Weather and Forecasting*, 24, pp. 1252–1267, doi: 10.1175/2009WAF2222241.1

[20] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Ser. B, 39, pp. 1–38.

[21] Du, J. & Tracton, M. (2001). Implementation of a real-time short range ensemble forecasting system at NCEP: An update. Preprints, *Ninth Conference on Mesoscale Processes*, pp. 355-360. Ft. Laurderdale, FL: American Meteorological Society.

[22] Ebert, E. E. (2001). Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Monthly Weather Review*, 129, pp. 2461–2480.

[23] Ebert, E. E., Gallus, W. A. (2009). Toward Better Understanding of the Contiguous Rain Area (CRA) Method for Spatial Forecast Verification. *Weather and Forecasting*, 24, pp. 1401–1415.

[24] Ebisuzaki, W., & Kalnay, E. (1991). Ensemble experiments with a new lagged average forecasting scheme. *WMO Research Activities in Atmospheric and Oceanic Modeling* Rep. 15, 308 pp.

[25] Eckel, F. A. & Mass, C. F. (2005). Aspects of effective mesoscale, short-range ensemble forecasting, *Weather and Forecasting*, 20, pp. 328-350.

[26] Efron, B. & Tibshirani, R. (1997). Improvements on Cross-Validation: The 632+ Bootstrap Method. *Journal of the American Statistical Association*, 92, No. 438, pp. 548-560.

[27] Ehrendorfer, M. (1997). Predicting the uncertainty of numerical weather forecasts: a review. *Meteorologische Zeitschrift*, N.F. 6, pp. 147-183.

[28] Emanuel, K. A. (1979). Inertial Instability and Mesoscale Convective Systems. Part I: Linear Theory of Inertial Instability in Rotating Viscous Fluids. *Journal of the Atmospheric Sciences*, 36, pp. 2425-2449.

[29] Escribà, P., Callado, A., Santos, D., Santos, C., García-Moya, J.A. & Simarro, J. (2010). Probabilistic prediction of raw and BMA calibrated AEMET-SREPS: the 24 of January 2009 extreme wind event in Catalunya. *Advances in Geosciences*, 26, pp. 119-124.

[30] Evans, R. E., Harrison, M. & Graham, R. (2000). Joint medium range ensembles from the Met. Office and ECMWF systems. *Monthly Weather Review* , 128, pp. 3104-3127.

[31] Feddersen, H. & Andersen, U. (2005). A method for statistical downscaling of seasonal ensemble predictions. *Tellus*, 57A, pp. 398–408.

[32] Ferranti, L. & Corti, S. (2010). Ensemble prediction skill in relation with large scale circulation patterns. *EMS Annual Meeting Abstracts*, Vol. 7, EMS2010-769, 2010, 10th EMS / 8th ECAC

[33] Ferro, C. A. T. (2007a). A Probability Model for Verifying Deterministic Forecasts of Extreme Events. *Weather and Forecasting*, 22, pp. 1089–1100.

[34] Ferro., C. A. T. (2007b). Comparing Probabilistic Forecasting Systems with the Brier Score. *Weather and Forecasting* 22:5, pp. 1076-1088.

[35] Ferro, C. A. T., Richardson, D. S. & Weigel, A. P. (2008). On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15, pp. 19–24.

[36] Fortín, V., Favre, A. C. & Saïd, M. (2006). Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. *Quarterly Journal of the Royal Meteorological Society*, 132, pp. 1349–1369.

[37] Frogner, I. L. & Iversen, T. (2002). High-resolution limited-area ensemble predictions based on low-resolution targeted singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 128, pp. 1321–1341.

[38] García-Moya, J. A., Callado, A., Escribà, P., Santos, C., Santos-Muñoz, D. & Simarro, J. (2011). Predictability of short-range forecasting: a multimodel approach. *Tellus A*, 63, pp. 550–563.

[39] Ghelli, A. & Lalaurette, F. (2000) Verifying precipitation forecasts using upscaled observations. *ECMWF Newsletter* 87, ECMWF, Reading, United Kingdom, pp. 9–17.

[40] Ghelli, A. & Primo, C. (2009). On the use of the extreme dependency score to investigate the performance of a NWP model for rare events. *Meteorological Applications.*, 16, pp. 537–544.

[41] Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B. & Ebert, E. E. (2009). Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24, pp. 1416–1430.

[42] Gneiting, T. & Raftery, A. E. (2005a). Weather forecasting with ensemble methods. *Science* , 310, pp. 248-249.

[43] Gneiting, T., Raftery A.E., Westveld III, A.H. & Glodman, T. (2005b). Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*. 133, pp. 1098-1118.

[44] Grimit, E. & Mass, C. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* , 17, pp. 192-205.

[45] Gutiérrez, J. M., Cofiño, A. S., Cano, R. & Rodríguez, M. A. (2004). Clustering methods for statistical downscaling in short-range weather forecasts. *Monthly Weather Review,* 132, pp. 2169–2183.

[46] Hagedorn, R., Hamill, T. M., & Whitaker, J. S. (2008). Probabilistic Forecast Calibration using ECMWF and GFS Ensemble Reforecasts. Part I: Two-meter temperatures. *Monthly Weather Review,* 136, pp. 2608-2619.

[47] Hágel & Horányi (2007). The ARPEGE/ALADIN limited area ensemble prediction system: the impact of global targeted singular vectors. *Meteorologische Zeitschrift,* 16, Number 6, December 2007, pp. 653-663.

[48] Hamill, T. & Colucci, S. (1997). Verification of ETA-RSM short-range ensemble forecast. *Monthly Weather Review*, 125, pp. 1312-1327.

[49] Hamill, T. & Colucci, S. (1998). Evaluation of ETA-RSM probabilistic precipitation forecasts. *Monthly Weather Review*, 126, pp. 711-724.

[50] Hamill, T., Snyder, C. & Morss, R. (2000). A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Monthly Weather Review*, 128, pp. 1835-1851.

[51] Hamill, T. M., Hagedorn R. & Whitaker, J. S. (2008). Probabilistic Forecast Calibration using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation. *Monthly Weather Review*. 136, pp. 2620-2632.

[52] Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting,* 15, pp. 559-570.

[53] Hoffman, R. N. & Kalnay, E. (1983). Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, 35, pp. 100–118.

[54] Hollingsworth, A. (1980). An experiment in Montecarlo Forecasting. *Workshop on Stochastic-Dynamic Forecasting*, pp. 65-85. Reading, United Kingdom.

[55] Hohenegger, C. & Schär, C. (2007). Predictability and error growth dynamics in cloud-resolving models. *Journal of Atmospheric Sciences*, 64, pp. 4467-4478.

[56] Hoeting, J. A., Madigan, D. M., Raftery, A. E. & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial (with discussion). *Statistical Science*, 14, pp. 382–401.

[57] Hou, D., Kalnay, E. & Droegemeier, K. (2001). Objetive verification of the SAMEX'98 ensemble forecast. *Monthly Weather Review,* 129, pp. 73-91.

[58] Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. & Michell, H. (1996). A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124, pp. 1225-1242.

[59] Houtekamer, P. L. & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126, pp. 796-811.

[60] Jolliffe, I. T. & Stephenson, D. B. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, New York.

[61] Jones, M., Colle, B. & Tongue, J. (2007). Evaluation of a mesoscale short-range ensemble forecast system over the Northeast United States. *Weather and Forecasting* , 22, pp. 36-55.

[62] Kalnay, E. & Ham, M. (1989). Forecasting forecast skill in the Southern Hemisphere. Extended Abstracts. *Third Int. Conf. on Southern Hemisphere Meteorology and Oceanography*, pp. 24-27. Buenos Aires, Argentina.: American Meteorological Society.

[63] Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association,* 90, pp. 773–795.

[64] Kharin, V. V. & Zwiers, F. W. (2003). On the *ROC* score of probability forecasts. *J. Clim.* 16, pp. 4145–4150.

[65] Krishnamurthy, V. (1993). A Predictability Study of Lorenz's 28-Variable Model as a Dynamical System. *Journal of the Atmospheric Sciences,* Vol. 50, No. 14, pp.2215-2229.

[66] Krishnamurti, T. N., Kishtawal, C., LaRow, T., Bachiochi, D., Zhang, Z., Willford, C., et al. (1999). Improved weather and seasonal climate forecast from multimodel superensemble. *Science* , 285, pp. 1548-1550.

[67] Lalaurette, F. (2003) Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quarterly Journal of the Royal Meteorological Society*, 129, pp. 3037–3057.

[68] Leamer, E. E. (1978). *Specification Searches*. Wiley, New York.

[69] Leith, C. E. (1974). Theoretical skill of Monte Carlo forecast. *Monthly Weather Review*, 102, pp. 409-418.

[70] Lin, J. W. B. & Neelin, J. D. (2002). Considerations for stochastic convective parameteriza-tion. *Journal of Atmospheric Sciences,* Vol. 59, No. 5, pp. 959–975.

[71] Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*. , 20, pp. 130-141.

[72] Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, 21, pp. 1-19.

[73] Lyapunov, A. M. (1992). *The general problem of the stability of motion*, Translated by A. T. Fuller, London: Taylor & Francis, ISBN 978-0748400621

[74] Lynch, P. (2006). "The ENIAC Integrations". The Emergence of Numerical Weather Prediction. *Cambridge University Press,* pp. 206–208, ISBN 9780521857291.

[75] Mason, S. J. (2004). On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Monthly Weather Review,* 132, pp. 1891–1895.

[76] McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, 274 pp.

[77] Mechoso, C. R. & Arakawa, A. (2003). *General ciculation. Models*. In J. R. Holton, J. A. Curry and J. A. Pyle, Encyclopedia of Atmospheric Sciences, pp. 861-869. Oxford: Academic Press.

[78] Mesinger, F., Janjic, Z., Nickovic, S., Gavrilov, D. & Deaven, D. G. (1988). The step mountain coordinate: model description and performance for cases of alpine cyclogenesis and for a case of an Appalachian redevelopment. *Monthly Weather Review*, 116, pp. 1493-1518.

[79] Molteni, F., Buizza, R., Palmer, T. & Petroliagis, T. (1996). The ECMWF Ensemble Prediction System: Methodology and Validation. *Quarterly Journal of the Royal Meteorological Society*, 122, pp. 73-119.

[80] Mullen, S. & Baumhefner, P. (1989). The impact of initial condition uncertainty on numerical simulations of large-scale explosive cyclogenesis. *Monthly Weather Review*, 117, pp. 2800-2821.

[81] Murphy, J. M. (1988). The impact of ensemble forecasts on predictability. *Quarterly Journal of the Royal Meteorological Society*, 114, pp. 463–493.

[82] Murphy, A. H. (1993). What is a good forecast?, An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8, pp. 281–293.

[83] Murphy, J., Sexton, D., Barnett, D., Jones, D., Webb, M., Collins, M. & Stainforth, D. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430, pp. 768–772.

[84] PaiMazumder, D. & Mölders, N. (2009). Theoretical assessment of uncertainty in regional averages due to network density and design. *Journal of Applied Meteorology and. Climatology*, 48, pp. 1643-1666.

[85] Palmer, T. N., Mureau, R., Buizza, R., Chapelet, P. & Tribbia, J. (1992). Ensemble prediction. *ECMWF Research Department Technical Memorandum*, 188, 45 pp.

[86] Palmer, T. N. (1997). On parameterizing scales that are only somewhat smaller than the smallest resollved scales, with application to convection and orography. *Proceedings of the ECMWF Workshop on New Insights and Approaches to Convective Parameterization*, 4-7 November 1996, pp. 328-337.

[87] Palmer, T. N. (2001). A nonlinear dynamical perspective on model error: Aproposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, 127, pp. 279–304.

[88] Palmer, T., Alessandri, A., U.Andersen, Cantelaube, P., Davey, M. et al. (2004). Development of a European multimodel ensemble system for seasonal to inter-annual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85, pp. 853-872.

[89] Pellerin, G., Lefaivre, F., Houtekamer, P. & Girard, C. (2003). Increasing the horizontal resolution of ensemble forecast at CMC. *Nonlinear Processes in Geophysics*, 10, pp. 463-468.

[90] Persson, A., Grazzini, F. (2005). User guide to ECMWF forecast products. Meteorological Bulletin M3.2, ECMWF, Reading, United Kingdom, 115 pp.

[91] Raftery, A. E., Balabdaoui, F., Gneiting, T. & Polakowski, M. (2005). Using Bayesian Model Averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, pp. 1155-1174.

[92] Richardson, D. S. (2000). Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126, pp. 649–667.

[93] Roebber, P. J. & Reuter, G. W. (2002). The sensitivity of precipitation to circulation details. Part II: Mesoscale modeling. *Monthly Weather Review*, 130, pp. 3–23.

[94] Rodwell, M. J., Richardson, D. S., Hewson, T. D. & Haiden, T. (2010). A new equitable score suitable for verifying precipitation in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 136, pp. 1344–1363.

[95] Roulston, M. S. & Smith, L. A. (2003). Combining dynamical and statistical ensembles. *Tellus* A, 55, pp. 16–30.

[96] Saetra, Ø., Hersbach, H., Bidlot, J. R. & Richardson, D. S. (2004). Effects of Observation Errors on the Statistics for Ensemble Spread and Reliability. *Monthly Weather Review*, 132, pp. 1487-1501.

[97] Santos, C. & Ghelli, A. (2011). Observational probability method to assess ensemble precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society*, doi: 10.1002/qj.895.

[98] Schmeits, M. J & Kok, K. J. (2010). A comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression using ECMWF Ensemble Precipitation Reforecasts. *Monthly Weather Review,* 138, pp. 4199-4211.

[99] Shutts, G., & Palmer, T. N. (2004). The use of high resolution numerical simulations of tropical circulation to calibrate stochastic physics scheme. *Proceeding ECMWF/ CLIVAR Workshop on Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO*, Reading, United Kingdom, ECMWF, pp. 83-102.

[100] Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131, pp. 3079–3102.

[101] Sloughter, J. M., Raftery, A. E., Gneiting T. & Fraley C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135, pp. 3209–3220.

[102] Sloughter, J.M., Gneiting, T. & Raftery, A.E. (2010). Probabilistic Wind Speed Forecasting using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association*, 105, pp. 25-35.

[103] Stensrud, D., Bao, J. W. & Warner, T. (1998). Ensemble forecasting of mesoscale convective systems. In A. M. Soc (Ed.), *12th Conference on Numerical Weather Prediction, Phoenix , AZ.* Preprints, pp. 265-268.

[104] Stensrud, D., Brooks, H., Du, J., Tracton, M. & Rogers, E. (1999). Using ensembles for short-range forecasting. *Monthly Weather Review*, 127, pp. 433-446.

[105] Stensrud, D. J. & Yussouf, N. (2007). Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecast system. *Weather and Forecasting*, 22, pp. 3–17.

[106] Stephenson, D. B., Casati, B., Ferro, C. A. T. & Wilson, C. A. (2008). The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorological Applications*, 15, pp. 41–50.

[107] Toth, Z. & Kalnay, E. (1993). Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, 74, pp. 2317-2330.

[108] Toth, Z. & Kalnay, E. (1997). Ensemble forecasting at NCEP: The breeding method. *Monthly Weather Review*, 125, pp. 3297-3318.

[109] Tracton, M. S. & Kalnay, E. (1993). Operational ensemble forecasting prediction at the National Meteorological Centre: Practical aspects. *Weather and Forecasting*, 8, pp. 379-398.

[110] Tribbia, J. J. & Baumhefner, D. P. (1988). The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Monthly Weather Review*, 116, pp. 2276-2288.

[111] Wandishin, M., Mullen, S., Stensrud, D. & Brook, H. (2001). Evaluation of a short-range multimodel ensemble system. *Monthly Weather Review*, 129, pp. 729-747.

[112] Wang, X. & Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of Atmospheric Sciences*, 60, pp. 1140–1158.

[113] Wang, X. & Bishop, C. H. (2005). Improvement of ensemble reliability with a new dressing kernel. *Quarterly Journal of the Royal Meteorological Society,* 131, pp. 965–986

[114] Wernli, H., Paulat, M., Hagen, M., Frei, C. (2008). SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136, pp. 4470–4487.

[115] Whitaker, J. S. & Loughe, A. F. (1998). The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review, 126*, pp. 3292–3302.

[116] Wilks, D. S., (2006). *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.

[117] Wilks, D. S. & Hamill, T.M. (2007). Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Review*. 135, pp. 2379-2390.

[118] Wilks, D. S. (2009). Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications,* 16, pp. 361-368.

[119] Williamson, D. L. (2007). The evolution of dynamical cores for global atmospheric models. *Journal of the Meteorological Society of Japan,* 85B, pp. 241-269.

[120] Wilson, L. J., Beauregard, S., Raftery, A. E. & Verret, R. (2007). Calibrated Surface Temperature Forecasts from the Canadian Ensemble Prediction y stem Using Bayesian Model Averaging (with discussion). *Monthly Weather Review*, 135, pp. 1364-1385. Discussion pages 4226-4236.

[121] Wobus, R. & Kalnay, E. (1995). Three years of operational prediction of forecast skill. *Monthly Weather Review*, 123, pp. 2132-2148.

[122] Ziehmann, C. (2000). Comparison of a single-model EPS with a multimodel ensemble consisting of a few operational models. *Tellus*, 52A, pp. 280–299.

http://dx.doi.org/10.5772/55699

[123] Zhang, F. (2005). Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Monthly Weather Review*, 133, pp. 2876-2893.