TECHNICAL MEMORANDUM

514

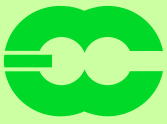
Ensemble Forecasting

M. Leutbecher and T. N. Palmer

Research Department

Submitted to J. Comp. Phys.

February 2007



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

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

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

Abstract

Numerical weather prediction models as well as the atmosphere itself can be viewed as nonlinear dynamical systems in which the evolution depends sensitively on the initial conditions. The fact that estimates of the current state are inaccurate and that numerical models have inadequacies, leads to forecast errors that grow with increasing forecast lead time. The growth of errors depends on the flow itself. Ensemble forecasting aims at quantifying this flow-dependent forecast uncertainty.

The sources of uncertainty in weather forecasting are discussed. Then, an overview is given on evaluating probabilistic forecasts and their usefulness compared with single forecasts. Thereafter, the representation of uncertainties in ensemble forecasts is reviewed with an emphasis on the initial condition perturbations. The review is complemented by a detailed description of the methodology to generate initial condition perturbations of the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). These perturbations are based on the leading part of the singular value decomposition of the operator describing the linearised dynamics over a finite time interval. The perturbations are flow-dependent as the linearisation is performed with respect to a solution of the nonlinear forecast model.

The extent to which the current ECMWF ensemble prediction system is capable of predicting flow-dependent variations in uncertainty is assessed for the large-scale flow in mid-latitudes.

1 Introduction

Numerical weather prediction is, by its very nature, a discipline that has to deal with uncertainties. The initial conditions of a numerical weather prediction model can be estimated only within a certain accuracy. During a forecast some of these initial errors can amplify and result in significant forecast errors. Moreover, the representation of the dynamics and physics of the atmosphere by numerical algorithms introduces further uncertainties associated for instance with truncation errors, with uncertainty of parameters describing sub-grid scale processes such as cumulus convection in a global model. We will refer to these two kinds of errors as initial condition errors and model errors, respectively. For the prediction of the real atmosphere, these two kinds of errors are not really separable because the estimation of the initial conditions involves a forecast model and thus initial condition errors are affected by model errors.

Over the past 15 years, ensemble forecasting became established in numerical weather prediction centres as a response to the limitations imposed by the inherent uncertainties in the prediction process. The ultimate goal of ensemble forecasting is to predict quantitatively the probability density of the state of the atmosphere at a future time. This is a nontrivial task because the actual uncertainty depends on the flow itself and thus varies from day to day. For many users of forecast information, an estimate of the future probability density of a weather-related variable may contain useful information beyond that contained in a single forecast started from the best available estimate of the initial state. There will be useful information in the probability density as long as the distribution differs from the empirical distribution of this variable obtained from past data; the latter distribution is usually referred to as the climatological distribution.

The first theoretical studies on error growth in atmospheric predictions, on the implied limits of prediction and on probabilistic predictions appeared in the late 1950's to early 1970's. This area of meteorological research is referred to as predictability. The pioneering works of Thompson, Lorenz, Epstein and Leith were central to the emergence of predictability research and ensemble forecasting. The roots of ensemble forecasting are analysed in depth in [1]. Readers are referred to this review article for more historical details and references relating to the early research on predictability.

There is growing interest in quantifying uncertainty in many areas of scientific computing [2]. Techniques for uncertainty quantification developed in other disciplines, e.g. oceanography [3], could prove or are already

useful in the context of numerical weather prediction and vice versa. In a nutshell, methods for uncertainty quantification relevant for numerical weather prediction need to cope with the sensitive dependence on initial conditions, the interaction of many spatial and temporal scales, and they should account for the fact that the sources of uncertainty are themselves fairly uncertain.

Numerical weather prediction models use a spatial discretisation to represent the partial differential equations that govern the dynamics of the atmosphere as a system of N ordinary differential equations. Now, we consider the fact that the initial state $\mathbf{x} \in \mathbb{R}^N$ is uncertain but can be described by a probability density function p.d.f. $p(\mathbf{x})$. Then, the evolution of the p.d.f. p under the dynamics of the forecast model is described by Liouville's equation. This is a linear partial differential equation defined on a subset of $\mathbb{R}^N$. Its numerical solution is impractical even for dimensions N as low as 100. Yet, current numerical weather prediction models have phase spaces of dimension $N \sim 10^6 - 10^8$. For such high-dimensional models, even the much simpler problem of merely evolving the covariances of the p.d.f. under the linearised dynamics poses a major challenge. Such a covariance evolution is required in the extended Kalman filter [4], an optimal state estimation algorithm in the limit of linear dynamics. In passing, we note that ensemble forecasting and state estimation, usually referred to as data assimilation in the meteorological context, are closely connected as both require the prediction of uncertainty estimates. The only difference is the time range at which these uncertainty estimates are required. Research in data assimilation is devoted to finding approximations of the extended Kalman filter that avoid the explicit evolution of the full covariance matrix. For instance, the state estimate of weak-constraint four-dimensional variational data assimilation is known to approximate that of a Kalman filter in the limit of a long assimilation time window [5].

The only feasible technique to obtain estimates of the nonlinearly evolved p.d.f. are Monte-Carlo techniques that sample the p.d.f. at initial time and evolve the sampled initial states with the forecast model — or a perturbed version of the forecast model to account for model uncertainty. The sample is usually referred to as *ensemble* and individual elements as *ensemble members*. This approach was envisaged already in the 1960's by Lorenz [1]. In conventional Monte-Carlo approaches, the sampled p.d.f. is known precisely. This is different for the weather prediction problem. The initial p.d.f. itself is uncertain because the estimates of errors determining the initial p.d.f. are uncertain and because the computation of the full p.d.f. for a high-dimensional system can only be achieved using approximations.

This article attempts to give an overview on ensemble forecasting with a particular focus on the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). It is impossible to discuss all aspects of ensemble forecasting in depth in this review article. A broader picture of current research on the predictability of weather and climate can be obtained from the contributions in a recently-published book [6].

This paper is organised as follows. Sec. 2 discusses sources of uncertainty in numerical weather prediction, the prediction of a p.d.f. and the evaluation of probabilistic predictions, and the value of ensemble forecasts in comparison to single forecasts. In Sec. 3, we discuss numerical model configurations for ensemble prediction including that of the ECMWF EPS. In Sec. 4, an overview is given on various methods to represent initial uncertainty, a subsection describes the initial perturbation methodology of the ECMWF EPS in detail. The representation of model uncertainty in ensemble forecasting is discussed in Sec. 5 followed by conclusions in Sec. 6.

2 Uncertainties in weather prediction

2.1 Sources of uncertainty

This section discusses initial condition errors and errors arising from uncertainties associated with the representation of the atmospheric dynamics by numerical algorithms as sources of forecast uncertainty. Although this distinction is useful to structure the story of predictability, we stress once more that initial condition error is inseparable from model error for a real physical system like the atmosphere.

Let us focus on initial condition uncertainties first. The initial conditions for any numerical weather forecast can only be estimated within a certain accuracy. Data assimilation techniques need to account for the uncertainties in the estimation process and can provide estimates of the uncertainty of the initial conditions [7] (this issue). Lorenz [8] studied the growth of forecast errors due to initial condition errors by looking at the rate at which solutions of the ECMWF numerical weather prediction model diverged. He looked at forecasts that are started at subsequent days. If one assumes that initial condition errors have similar statistics as the difference of the 1-day forecast started at day $D-1$ and the initial condition at day D, one can estimate the growth of forecast errors solely due to initial condition errors by looking at the divergence of the two forecasts. Figure 1 illustrates this divergence of two solutions for a particular case. The depicted variable is the geopotential of the 500 hPa pressure surface. In meteorology, geopotential refers to the gravitional potential for unit mass with respect to the mean sea level. If the dependence of gravity on altitude is ignored, the geopotential is proportional to the height above mean sea level. The geopotential has generally high values at low latitudes and low values at high latitudes. This variable is often used to depict the flow on the larger spatial scales in the extra-tropics. To first order the wind is parallel to the isolines of the geopotential and the wind speed is proportional to its gradient (a relationship referred to as geostrophic balance). The integrations are performed with the ECMWF Integrated Forecasting System (IFS) at the resolution of the currently operational EPS (about 50 km in the horizontal and 62 levels in the vertical). The forecast in the left column starts from the estimate of the initial conditions for 15 February 2006, 00 UTC whereas the forecast in the right column starts from the initial condition estimate obtained 12 hours earlier. Panels in the same row show the atmospheric state for the same date and time. There are only minor differences apparent between the 0-hour and 12-hour forecasts (upper panels). However after $7.0/7.5$ days, significant differences have developed in the two forecasts while there are still regions in which the forecasts agree qualitatively (middle panels). Still further into the forecast after $14.0/14.5$ days, the two forecasts barely resemble each other except for the average poleward gradient of geopotential that would be expected from the climatological average(bottom panels).

Statistics accumulated over longer periods confirms that this divergence of solutions from slightly different initial conditions is not atypical [8]. Figure 2 shows the RMS difference of pairs of lagged forecasts valid at the same time for 500 hPa geopotential (north of 20°N). The statistics are based on a sample of 180 cases from Dec 2005 to Feb 2006. The initial doubling rate of the differences is of the order of 2 days. After about 20 days, the growth of the differences has nearly saturated. On average the 20-day forecast and the 20.5-day forecast will be almost as different as two states of the atmosphere that would have been picked randomly from the climatological distribution of atmospheric states.

The above results illustrate the large sensitivity to initial conditions of a particular numerical weather prediction model. But similar results are obtained with any numerical weather prediction model. The high sensitivity of future states to the initial conditions has also been found in more idealised models of the large-scale atmospheric dynamics. In [9] the authors numerically computed the spectrum of Lyapunov exponents for a model of the general circulation of the atmosphere with 1449 variables. They estimated about 100 positive Lyapunov exponents. Lorenz first realised that forced dissipative systems of ordinary differential equations that originate from highly truncated models of fluid flows systematically exhibit sensitive dependence to initial conditions

[10, 11, 12].

An estimate of the actual RMS error of the 500 hPa geopotential forecast is included in Fig. 2. If the RMS difference of lagged forecasts is taken as an estimate of forecast errors due to initial conditions, the difference between the two curves can be attributed to the forecast errors arising from model error [8]. According to this diagnostic, a large fraction of the forecast error appears to be caused by the growth of initial condition errors through the dynamics of the forecast model. While this diagnostic gives a first estimate of the relative contributions of initial condition errors and model errors to the forecast error, it is unknown whether the results depend sensitively on the fact that the "initial errors" are represented by forecast-minus-initial condition differences, which may have different characteristics than the unknown actual initial errors.

Since the first successful numerical weather predictions in the 1950's [13] (this issue), significant advances have been made in the accuracy of numerical weather prediction models. To some extent, these advances have been driven by the steady increase in high-performance computing resources which has permitted the increase in the spatial resolution of numerical weather prediction models over the years. Thereby, more and more scales of motion can be resolved. Further advances can be expected with even higher resolutions as the atmosphere exhibits still significant variability on the spatial scales of todays models' truncation limits (discussed elsewhere in this issue). The accuracy of the numerical weather prediction models has also benefitted significantly from improvements in the way physical processes and motions on the sub-grid scale are represented as well as improvements in the representation of the interaction with the ocean and the land surface. ([14] and other contributions in this issue).

Although the accuracy of numerical weather prediction models has increased and is still increasing, model errors are expected to continue to make an important contribution to forecast errors because of the entanglement of initial condition errors and model errors. In numerical weather prediction, initial conditions at a time t_* are estimated by blending information from observations within a certain time window around t_* with that from the most recent forecast for the time t_* [7]. Thereby, information from all past observations is evolved in time. It is obvious that the accuracy of the initial conditions can be improved by using a more accurate numerical weather prediction model in this estimation process. This dependency of the initial condition error on the model error implies that the relative contributions of initial condition error and model error to the forecast error need not change so much even if significantly more accurate numerical weather prediction models become available.

The atmosphere exhibits variability on spatial scales smaller than and near to the truncation limit. Furthermore, truncation errors due to the spatial discretisations of the partial differential equations tend to be largest on the spatial scales close to the truncation limit. These two factors inevitably lead to significant uncertainty on the near grid-scale. Although this model error contribution is largest on small scales it can affect the errors on scales of mid-latitude high and low pressure systems within a few days. Work on optimal perturbations (singular vectors) shows that the atmospheric dynamics exhibits a preferential growth of perturbation energy to larger spatial scales [15].

The relative contributions of model error and initial condition error to forecast errors are not well known for the weather prediction problem. These can only be known precisely in idealised work in which the system itself is a numerical model. The three major operational ensemble prediction systems exhibit a lack of spread after a forecast-range of about 5 days [16]. A possible explanation for this lack of spread may be that in all three systems the growth of forecast errors due to model errors is not represented adequately.

2.2 Predicting uncertainty

Methods for representing the sources of uncertainty in ensemble prediction systems will be discussed in Secs. 4 and 5. Now, we describe properties of a skilful ensemble prediction system and ways of measuring the "skill"

of an ensemble prediction system.

The ensemble mean is a forecast with generally lower RMS error than that of an unperturbed forecast because the unpredictable scales of motion have been filtered in the ensemble mean and only the signal of the predictable scales remains [17]. With increasing forecast range, the ensemble mean gradually converges to the climatological mean state. A single forecast that has higher spatial resolution than the ensemble forecasts tends to be more skilful than the ensemble mean initially. The forecast range at which the ensemble mean outperforms such a higher resolution single forecast depends, inter alia, on the predictability of the variable that is being considered [18].

The skill of the ensemble mean forecast is a measure that quantifies an aspect of the skill of an ensemble prediction system. However, it does not assess the uncertainty information contained in ensembles directly. Various other metrics are in use which assess the probabilistic information contained in ensemble forecasts. For this paper, we select a few of these metrics, see e.g. [19] for a more comprehensive overview and further references. A variety of metrics is used because they emphasise different aspects of probabilistic forecasts. Conceptually, one can distinguish two aspects that determine the skill and usefulness of probabilistic forecasts: The statistical consistency of the predicted probabilities and the width of the predicted distribution. These two aspects are referred to as reliability and resolution [20].

Let us consider a particular weather event, say whether it will rain more than 1 mm within the next 5 days in London. A probabilistic prediction of this event is said to be reliable if the relative frequency of the event tends to p for a large number of dates for which it was predicted to occur with probability p for any $p \in [0,1]$. More generally, one can formulate a stricter requirement for the reliability of a probabilistic forecast of a scalar variable x: for cases in which the cumulative distribution $P(x)$ is predicted, the sample distribution of the true state of the atmosphere should tend to $P(x)$ in the limit of many cases. We will call such an ensemble *perfectly reliable*. This definition only makes sense for an ensemble with an infinite number of members. Real ensembles with a finite number of members cannot be perfectly reliable due to sampling uncertainty [21]. Statistical consistency can be expressed in the following form for finite size ensembles: All ensemble members and the true state of x are independent draws from the same distribution $P(x)$. This will be referred to as a perfect ensemble. A sample of n realisations from the climate distribution will constitute a perfect ensemble. In the limit of an infinite number of ensemble members the perfect ensemble will be perfectly reliable. The reliability of an ensemble can be improved through calibration based on statistics of past ensemble forecasts for the same season and region [22, 23].

Reliable probability forecasts with a narrow distribution will be more useful than reliable probability predictions with a broad distribution. This aspect is referred to as resolution because a narrow predicted distribution will better resolve whether or not an event is likely to occur than a broad distribution. However, even a broad distribution which deviates somehow from the climatological distribution contains predictive information that can be useful to some forecast users.

An example of a metric to assess the quality of probabilistic forecasts for a scalar variable that measures both reliability and resolution is the continuous ranked probability score [24]. It is defined as the mean squared error of the predicted cumulative distribution

$$\mathrm{CRPS} = \frac{1}{M}\sum_{j=1}^{M}\int_{-\infty}^{+\infty}\Big[P_j(x) - H(x - x_{oj})\Big]^2 \,\mathrm{d}x, \tag{1}$$

where P_j, H, x_{oj} denote the predicted cumulative distribution for case j, the Heaviside step function, and the observed value, respectively. For a deterministic forecast that predicts the value x_{fj}, i.e. $P_j(x) = H(x - x_{fj})$, the CRPS is equivalent to the mean absolute error. The CRPS is related to the Brier score which is a measure used

to evaluate dichotomous events (occurred/not occurred) defined as

$$B = \frac{1}{M}\sum_{j=1}^{M}(p_j - o_j)^2, \tag{2}$$

where p is the predicted probability for the event to occur in the j-th case and o_j is 1 if the event occurred and 0 otherwise. The CRPS can be expressed as the integral $\int \mathrm{d}\xi$ of the Brier score for the event not to exceed the value ξ.

Statistical consistency implies a relationship of the expected variance of the ensemble mean error and the dispersion of the ensemble. Let us consider a perfect ensemble and a sample of M independent cases, e.g. different dates. In the j-th case, the true state $x_{T,j}$ and the N ensemble members $x_{1,j},\ldots,x_{N,j}$ are independent draws from the p.d.f. p_j, which has mean μ_j and standard deviation σ_j. Further, let $\langle\ .\ \rangle_N \equiv \frac{1}{N}\sum_{k=1}^{N}$ denote the expectation computed from the sample of ensemble members, e.g. we write $\langle x_{.,j}\rangle_N \equiv \frac{1}{N}\sum_{k=1}^{N} x_{k,j}$ for the ensemble mean. The dispersion of the ensemble is quantified by

$$s_j^2 = \left\langle \left(x_{.,j} - \langle x_{.,j}\rangle_N \right)^2 \right\rangle_N \tag{3}$$

and the ensemble mean error by

$$\varepsilon_j^2 = \left(x_{T,j} - \langle x_{.,j}\rangle_N \right)^2. \tag{4}$$

We refer to s and ε as RMS spread and RMS error, respectively. Following the standard derivation for an unbiased variance estimate from a sample, we obtain

$$\frac{1}{M}\sum_{j=1}^{M}\left(s_j^2 - \frac{N-1}{N}\sigma_j^2 \right) \quad \longrightarrow \quad 0, \quad \text{for } M \to \infty \tag{5}$$

where convergence is in the probability sense; the term converges to zero almost surely, i.e. with probability one. The ensemble mean RMS error satisfies

$$\frac{1}{M}\sum_{j=1}^{M}\left(\varepsilon_j^2 - \frac{N+1}{N}\sigma_j^2 \right) \quad \longrightarrow \quad 0, \quad \text{for } M \to \infty. \tag{6}$$

The factors $\frac{N-1}{N}$ and $\frac{N+1}{N}$ that account for the finite ensemble size appear in (5) and (6) because spread and error refer to the sample mean $\langle x_{.,j}\rangle_N$ rather than the distribution mean μ_j. Combining (5) and (6) yields

$$\frac{1}{M}\sum_{j=1}^{M}\left(\varepsilon_j^2 - \frac{N+1}{N-1}s_j^2 \right) \quad \longrightarrow \quad 0, \quad \text{for } M \to \infty. \tag{7}$$

This implies that in a perfect ensemble the average ensemble mean RMS error tends with increasing sample size M to the RMS spread times a correction factor close to 1 that accounts for the finite size N of the ensemble. The relationship between spread and error is of particular interest if the spread varies because a perfect ensemble will satisfy (7) not only in an average sense over the entire sample of cases but also for sufficiently large subsamples which are conditioned on the predicted spread. This implies that the spread in a perfect ensemble will predict the standard deviation of the ensemble mean forecast error distribution. Obviously, the variability of the spread has to be larger than that expected by pure sampling uncertainty for this prediction to be useful.

Statistical methods can provide probabilistic forecasts without the need for ensemble forecasts. The climatological distribution has been mentioned previously; another example is a single deterministic forecast turned into a probabilistic forecast by dressing it with a distribution of past forecast errors valid for the same region, season and forecast range. However, such simple statistical methods will always yield the same spread. Some flow-dependence could be introduced in the error statistics by stratifying the sample according to the predicted atmospheric state but sample size is likely to be a major issue with this approach. In contrast, ensembles consisting of multiple integrations of numerical weather prediction models do exhibit temporal and spatial variations of the spread because the conditions of the flow itself modulate the dispersion of nearby atmospheric states. The realism of the spread will depend on how well the sources of uncertainty are represented in the ensemble prediction system. An example of the variability of the ensemble spread is given in (Fig. 3). It shows two consecutive ensemble forecasts of hurricane Katrina's track. Katrina, made landfall near New Orleans on 29 August 2005, 12 UTC. The ensemble forecast issued on 26 August, 00 UTC exhibit a large dispersion of tracks. This dispersion is substantially reduced in the ensemble forecast initialised at 12 UTC on 26 August. The actual track of Katrina was close to the mode of the distribution predicted by the later forecast.

Even for a perfect ensemble, one cannot expect a very high correlation between the ensemble mean error and the spread if the correlation is computed from pairs of error and spread for individual cases because one essentially compares the standard deviation of a distribution with the magnitude of individual realisations. The correlation of spread and ensemble mean error for a perfect ensemble increases with the case-to-case variability of the spread [25]. It is more appropriate to assess the validity of (7) by considering the sample mean RMS errors conditioned on the predicted spread. Such a diagnostic was performed previously for the ECMWF EPS [26]. It is repeated here using data from a more recent version of the operational ECMWF EPS. The data consists of 89 ensemble forecasts (Feb–April 2006) for 500 hPa geopotential height in the Northern Hemisphere mid-latitudes (35°–65°N). As the true state of the atmosphere is unavailable, the ensemble mean error is estimated from analyses, i.e. estimates of the initial state obtained with ECMWF's four-dimensional variational assimilation system. For the considered variable and forecast ranges this is considered to be an acceptable approximation. The data is analysed on a regular $2.5^\circ \times 2.5^\circ$ latitude-longitude grid — 2160 grid points in total. Thus the entire sample consists of 192240 pairs of spread and ensemble mean error. The data has been stratified according to the predicted spread and divided into 20 equally populated bins, where bin boundaries are given by the 5%-,10%,...,95%-percentiles of the spread distribution. Then, the RMS error of the ensemble mean and the RMS spread are computed for each bin. Figure 4 shows the relationship between error and spread for the 2-day, 5-day and 10-day forecast. At all three forecast ranges, the ensemble spread contains useful information about variations of the width of the distribution of the ensemble mean error. The larger the spread of the ensemble the larger is the average ensemble mean RMS error. At the later forecast ranges (5–10 days), the average RMS error is fairly accurately predicted by the ensemble standard deviations. At the shorter forecast ranges (2-days), the spread is less reliable in predicting the variability of the width of the ensemble mean error distribution. For cases with large (small) ensemble spread, the average RMS error is systematically lower (higher) than the spread.

The comparison shown in Fig. 4 does not distinguish between temporal variations of the spread and spatial variations. Some of the skill in predicting variations of the width of the forecast error distribution can arise from correctly predicting the seasonally averaged geographic variations. We consider now the geopotential normalised by the time-average RMS spread as new variable in order to focus on the aspect of the temporal variability only. By definition, the time-averaged spread of the normalised geopotential is geographically uniform. The relationship between spread and average RMS error for the normalised geopotential is shown in Fig. 5. The normalisation reduces the range over which error and spread vary. For instance, the ratio of RMS error in the bin with largest 5% spread to the RMS error in the bin with lowest 5% spread is about 4.5 without normalisation at day 5; this is reduced to about 2.7 with normalisation. With increasing forecast range, the distribution of the predicted spread narrows. Eventually, as the ensemble converges to the climatological

distribution, the spread will be almost constant except for sampling uncertainty and moderate intra-seasonal variations. Figure 5 demonstrates that the EPS is indeed providing information about flow-dependent variations of the width of the ensemble mean error distribution. At the same time, the diagram reveals systematic errors of the flow-dependent error bars predicted by the ensemble. Both calibration as well as improvements of the initial uncertainty representation and model uncertainty representation are expected to lead to significant further improvements of the statistical consistency and thus increase the capability to predict flow-dependent variations of forecast uncertainty.

2.3 The value of ensemble forecasts

Ultimately, the value of a probabilistic ensemble forecast system is as a tool for decision making. A key question is whether an ensemble forecast system has greater value in this respect than a best-guess deterministic forecast. The cost/loss model [27, 28] is relevant in this respect. We imagine a user who is sensitive to a weather event E, in the sense that (s)he will incur a loss L if the event occurs and no protective action is taken. Protective action can be taken at cost C, which we will assume to be less than L. Figure 6 shows the so-called Potential Economic Value (PEV) of the ECMWF Ensemble Prediction System and associated best-guess deterministic forecast (EPS control) for the event E: 24-hour accumulated precipitation exceeding 1mm, over all grid points in Europe, for the period August-October 2005. A PEV of unity indicates a perfect deterministic forecast, a PEV of zero indicates that the value of the forecast system is no better than a decision system based on knowing the climatological frequency $f_{\text{clim}}(E)$ of E - in this latter situation, the optimal decision strategy will be to always take protective action if $f_{\text{clim}}(E) > C/L$ and to never take protective action if $f_{\text{clim}}(E) < C/L$.

It can be seen that the value of the EPS is substantially greater than the value of the control over most of the range of cost/loss values. Indeed, for users with either low C/L or high C/L, the control itself has no value for decision making over and above decisions made with knowledge of the climatological frequency of rain. This can be understood qualitatively as follows. For small C/L it is clearly worth taking protective action even if there is only a small forecast probability of E. By contrast, for large C/L it is only worth taking protective action if it is almost certain that E will occur. More generally [28], for a reliable probabilistic forecast system and for a particular C/L the optimal decision strategy will be to take protective action if the forecast probability of E exceeds C/L. A best-guess deterministic forecast system either predicts E or does not predict E. Its value is therefore sub-optimal for such decision making. Although, the value of the control forecast is lower than that of the ensemble for all C/L, a superior probabilistic forecast can be obtained by optimally combining the EPS with a single high resolution deterministic forecast [18].

Not all decisions are binary decisions. An example of how an ensemble forecast system has greater economic value than a best-guess deterministic forecast system is provided by Weather Roulette [29]. Here the methodology is based on an analysis made by spreading financial stakes across a range of forecast outcomes, in proportion to the probability of that outcome. A more general assessment of the economic value of ensemble forecast systems is discussed in [30].

3 Numerical model configurations for ensemble prediction

3.1 Overview

Ensemble forecasts consist of multiple integrations of numerical weather prediction models. Now, we will discuss suitable configurations of NWP models for ensemble prediction — methods for representing the uncertainties will be discussed in subsequent sections. The amount of available computing resources puts a constraint

on how many ensemble members can be produced, on how fine the spatial and temporal discretisation of the partial differential equations can be and on how far into the future the model can be integrated. Various ensemble prediction systems have emerged within the past 15 years that range from climatic predictions that span several decades to short-range weather prediction up to a few days. Accordingly, the horizontal resolution of the numerical models ranges from a few hundred kilometres for the former to the order of 10 km for the latter. The number of different forecasts in an ensemble tends to be of the order of 10–100 for operational weather prediction systems. Small spatial scales tend to be predictable for shorter time scales than the larger spatial scales [31]. This may be used as argument for using coarser spatial resolutions for the longer forecast ranges. Atmospheric predictions for time scales from the order of weeks onwards are affected by the uncertainty of the evolution of the state of the ocean. Therefore, coupled ocean-atmosphere models are employed for longer-range predictions [32]. Ensembles of regional models that have integration domains for a limited area require that the uncertainty on the lateral boundaries is represented. Ensembles of coarser resolution global model integrations are used to provide dynamically consistent perturbed lateral boundary conditions for regional model ensembles, e.g. [33].

In the following, we focus on ECMWF's medium-range EPS (up to 10 days at the time of writing and 15 days when this article appears). ECMWF also produces forecasts with two lower resolution ensembles for monthly and seasonal predictions. The latter two ensembles are coupled to an ocean model and account also for the uncertainty in the ocean.

3.2 The ECMWF medium-range EPS

The numerical weather prediction model used for all forecasts is a two-time-level semi-Lagrangian global spectral model [34]. The semi-Lagrangian scheme uses a Stable Extrapolation Two-Time-Level Scheme (SETTLS) and a finite element scheme for the vertical discretisation [35, 36]. The atmospheric model is coupled to an ocean wave model [14] and to a land-surface model and comprises a comprehensive set of parameterisations for physical processes such as radiative transfer and moist processes. The initial state of the atmosphere is estimated with a 4-dimensional variational assimilation system. The current horizontal resolution of the assimilation system and deterministic forecasting system corresponds to 25 km. In the vertical, the atmosphere is discretised in 91 layers from the surface to 1 Pa. Some of the recent changes of the assimilation and forecasting system which have lead to improvements in forecast skill over the past years are described in [37].

The EPS model is triangularly truncated at a spherical harmonics total wavenumber 399 which is equivalent to a horizontal resolution of about 50 km. In the vertical, the atmosphere is discretised into 62 layers between the surface and the 5 hPa level (about 35 km altitude). The integration time-step is 1800 s. The ensemble consists of 50 perturbed forecasts and one unperturbed forecast. Ensemble forecasts are generated twice daily for 00 UTC and 12 UTC.

3.3 Discussion

The model used for the first operational ensemble prediction system at ECMWF had a resolution a spectral truncation at total wavenumber 63. Since the resolution has been increased in several steps up to the present truncation at total wavenumber 399. Each of these resolution increases brought significant improvements in the skill of the probabilistic forecasts [38, 39]. In the future, the different ensemble systems at ECMWF will be joined into a single seamless system. This is referred to as variable resolution EPS (VAREPS, [40]).

4 Representation of initial uncertainty

4.1 Overview

Several different techniques have become established for representing initial uncertainty in ensemble weather forecasting. This diversity arises from our limited quantitative knowledge about the relevant sources of uncertainty and the difficulty of conducting fair comparisons of the different techniques using the same numerical model and real observational data. Although we discuss the representation of initial uncertainty and model uncertainty separately; they have to be evaluated jointly because model uncertainty contributes to the initial condition uncertainty. The best technique for representing initial uncertainty is not independent of the nature of the model error and the way it is represented in the ensemble prediction system. Conceptually, one can distinguish between techniques that aim at obtaining a sample from the p.d.f. of initial states and techniques that selectively sample initial uncertainty only in those directions of the state space that are dynamically the most important for determining the ensemble dispersion.

Techniques in the first group employ ensemble based data assimilation techniques to get a sample of initial states. The ensemble of the Canadian Meteorological Service is based on initial conditions from an ensemble Kalman filter in which assimilated observations are perturbed by pseudo-random numbers. The added noise represents observational error [41, 42]. This approach can be viewed as a feasible approximation of the extended Kalman filter [43]. Square root filters are another ensemble based technique which aims at approximating the extended Kalman filter [44]. The latter filters generate perturbations about the ensemble mean through a deterministic algorithm. An ensemble of analysis perturbations is generated from an ensemble of forecast perturbations through a linear transformation such that ensemble based covariances satisfy the Kalman filter covariance update equation. The ensemble transform Kalman filter [45] is a particular square root filter. It is computationally very efficient but does not permit localisation of the ensemble based covariances. Localisation of covariance estimates is used to filter out spurious long-range correlations which arise from sampling uncertainty. Therefore, it is only suitable for generating the perturbations and not an initial state estimate. The ensemble of the UK Meteorological Office uses the ensemble transform Kalman filter to generate the initial perturbations.

Techniques in the second group aim at sampling the dynamically most relevant aspects of the initial uncertainties. The initial perturbations are dynamically constrained. The technique that is used at the US National Center for Environmental Prediction (NCEP) and several other centres is based on so-called bred vectors [46, 47]. A bred vector is obtained through the following procedure. A random perturbation is added to an initial state; then, the perturbed and unperturbed states are evolved with the nonlinear forecast model for a period of time (of the order of 6-24 h); the finite difference of the evolved states is rescaled so that its amplitude matches that of typical analyses error. Next, the new perturbation is added to a new state estimate and the procedure is repeated. The bred vector is the perturbation that emerges after several growth and rescaling steps. The rescaling of the finite difference represents the effect of data assimilation on initial uncertainty in an approximate manner. It adjusts the amplitude of the perturbation, but it does not represent the scale dependent filtering properties of data assimilation algorithms which tend to have a whitening effect on the distribution of errors, i.e. large-spatial-scale errors are better constrained than small-scale errors [48]. Initial perturbations for several ensemble members are obtained by using several independent bred vector computations that start from different random perturbations in the first cycle. The ensemble transform Kalman filter technique can be viewed as a refined breeding method — the re-scaling is replaced by a linear transformation. The ensemble transform Kalman filter technique has been compared with the bred vector method at NCEP for a 10-member ensemble [49]. The study suggests that an ensemble size of 10 members is too small for the ensemble transform Kalman filter to adequately represent inhomogeneities in the global observing network. Recently, NCEP has increased the ensemble size to 14 members (4 times daily) and replaced the breeding by an Ensemble Transform Technique

with rescaling [50]. This method is related to the ensemble transform Kalman filter technique.

The technique that is used to perturb the initial conditions in the ECMWF ensemble prediction system is based on the leading singular vectors of the operator that describes linear perturbation dynamics over a finite time interval, say 2 days. The singular vectors maximise perturbation growth over this time interval. The technique is described in detail in Sec. 4.2. In a sense that will be described formally later, the leading singular vectors identify those directions of initial uncertainty that are responsible for the largest forecast uncertainty at the end of the specified time interval. Due to this property, they provide a convenient way of generating an ensemble with sufficient dispersion in the most uncertain directions.

In an ideal world, where all sources of uncertainty are precisely known, ensemble based assimilation techniques can provide a means of obtaining at least approximately a sample of the actual p.d.f. of initial uncertainty. Idealised studies have compared techniques that aim to sample the actual p.d.f. of initial uncertainty with the bred vector and singular vector approaches in the absence of model error, e.g. [51, 52]. In such studies, observations are generated from a numerical model integration and the same model is used as forecast model. In [51], the authors using a model of the large-scale, quasi-geostrophic dynamics conclude that a perturbed observation ensemble, an ensemble using singular vectors and an ensemble using bred vectors yield equally skilful ensemble mean forecasts. The authors in [52] also used a model of the large-scale, quasi-geostrophic dynamics in the extra-tropics. They conclude that techniques sampling the p.d.f. of initial uncertainty improves the statistical consistency of the ensemble forecast compared to the dynamically constrained techniques. The differences are largest in the early part of the forecast and become fairly marginal around day 5. A study with the 3-variable dynamical system introduced by Lorenz (1963) also concludes that dynamically constrained ensembles are not justified if the p.d.f. of initial uncertainty is available [53].

In the real world, all sources of uncertainty will never be quantified precisely due to the large recurrence time of the atmosphere and the limited spatio-temporal coverage of observational data. The contribution of a source of uncertainty to forecast uncertainty depends not only on the typical magnitude of the errors but also on their spatial and temporal correlations. For instance, some observation types which are used for estimating the initial atmospheric state are known to have spatially correlated errors. For satellite radiances, such correlations are likely to be significant but quantitative estimates are not yet available. Quantitative estimates of observation error correlations have been obtained for indirect wind observations obtained from feature tracking in satellite imagery (atmospheric motion vectors, [54]). This study diagnoses statistically significant correlations at distances of up to about 800 km in the extra-tropics; correlations reach the value 0.2 at distances of about 200–300 km. By and large, spatial and temporal correlations of observation errors are not yet represented in current data assimilation algorithms and this has a potentially degrading effect on the initial state estimate [55]. Observation error correlations will also lead to incorrect estimates of the initial state uncertainty if the correlations are not accounted for in the data assimilation schemes.

The representation of model error within ensemble based assimilation techniques constitutes another major challenge. It is necessary to represent model error in ensemble based data assimilation to obtain reliable uncertainty estimates [42]. Idealised experiments with models of different spatial resolutions suggest that the accuracy of ensemble data assimilation techniques is sensitive to the the specification of the spatial correlations of the model error term [56]. Studies with low-dimensional dynamical systems and with numerical weather prediction models also emphasise the sensitivity of ensemble dispersion to the specification of the temporal correlations of the model error term [57, 58]. Present knowledge about the spatio-temporal correlations of the model error of numerical weather prediction models is rather limited.

The accurate estimation of the p.d.f. of initial uncertainties will require the accurate specification of the important sources of uncertainty. This is conceived as a major challenge for both data assimilation and ensemble prediction. In [16], the authors compare the performance of three global ensemble prediction systems (Canada,

NCEP, ECMWF) but are unable to conclude whether sampling of the initial p.d.f. is superior to the dynamically constrained perturbations because the three ensemble prediction systems use different numerical models and initial conditions of different quality. The singular vector approach that will be described next is an attempt to circumvent the difficulties of obtaining samples of the actual p.d.f. of initial uncertainty.

4.2 Singular vector based initial perturbations

4.2.1 The linear regime and the tangent-linear approximation

The evolution of small amplitude perturbations of a particular solution of a numerical weather prediction model can be approximately described by solving the set of linear equations that is obtained by a first order Taylor expansion of the nonlinear model about the particular solution. In the following, we assume that the partial differential equations that govern the dynamics of the atmosphere have been spatially discretised and can be written as a system of N ordinary differential equations.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = F(\mathbf{x}) \tag{8}$$

Here, $\mathbf{x} \in \Omega \subset \mathbb{R}^N$ and $F : \Omega \to \mathbb{R}^N$. The function F is nonlinear and might even be not continuous. The lack of regularity arises from parameterisations of subgrid-scale physical processes that depend in a discontinuous way on the state $\mathbf{x}$. For what follows, we assume that a reasonable approximation to F exists which is continously differentiable. A significant amount of work has gone into developing differentiable approximations of F, see e.g. [59, 60, 61].

Now, let $\mathbf{x}_r(t)$ denote a particular solution of (8); it will be referred to as the reference trajectory. The dynamics of infinitesimally small perturbations $\mathbf{x} \in \mathbb{R}^N$ about the reference trajectory $\mathbf{x}_r$ is described to first order in $\mathbf{x}$ by the tangent-linear system

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \mathbf{A}(\mathbf{x}_r(t))\,\mathbf{x}, \tag{9}$$

where $A_{jk} = \partial F_j/\partial x_k(\mathbf{x})$ denotes the Jacobi matrix of F. For any solution $\mathbf{x}$ of (9), $\mathbf{x}_r + \varepsilon\mathbf{x}$ approximates a solution of (8) starting at $\mathbf{x}_r(t_0) + \varepsilon\mathbf{x}(t_0)$ to first order in ε. Let us denote by $\mathbf{M}(t_0,t)$ the solution operator of (9), i.e. for any $\mathbf{x}_0 \in \mathbb{R}^N, t \mapsto \mathbf{M}(t_0,t)\mathbf{x}_0$ is a solution of (9) and $\mathbf{M}(t_0,t_0)\mathbf{x} = \mathbf{x}$. The operator $\mathbf{M}$ is referred to as tangent-linear propagator or simply propagator in the following.

Let us now consider how a Gaussian distribution of initial states evolves under the dynamics of (8). The Gaussian distribution will remain close to a Gaussian up to some future time $t_0 + \tau$ if the initial distribution is sufficiently narrow. Its evolution can be approximated by the tangent-linear system (9) assuming that the mean of the Gaussian at t_0 is identical or close to the initial value of the reference trajectory $\mathbf{x}_r(t_0)$. Within this linear approximation, the evolution of the covariances $\mathbf{C}(t)$ of the distribution is governed by

$$\mathbf{C}(t) = \mathbf{M}(t_0,t)\,\mathbf{C}(t_0)\,\mathbf{M}(t_0,t)^{\mathrm{T}}, \tag{10}$$

where superscript $^{\mathrm{T}}$ denotes matrix transpose. This regime is referred to as the *linear regime*. For numerical weather prediction on large horizontal scales (hundreds of km), the linear regime lasts about 1–2 d for initial perturbations with a magnitude corresponding to typical initial condition errors [62, 63, 64]. The dimension of the state vector of numerical weather prediction models is so large ($\sim 10^5$ to 10^8) that the full covariance matrix is too large to be stored, let alone be evolved using (10). However, it is possible to compute those directions at t_0 that evolve into the leading eigenvectors of $\mathbf{C}(t_0 + \tau)$. We will see that these are singular vectors of the propagator $\mathbf{M}(t_0,t_0+\tau)$ if suitable normalisations are used.

For a realistic numerical weather prediction model, it is computationally not feasible to obtain the propagator $\mathbf{M}(t_0,t_1)$ in matrix form. Instead, algorithmic differentiation is used. From the numerical algorithm that describes the nonlinear model F, the numerical algorithm is derived that describes the perturbation dynamics to first order in perturbation amplitude. The new algorithm is referred to as the tangent-linear model. It has two input variables: the starting value for the nonlinear forecast $\mathbf{x}_r(t_0)$ and the perturbation at initial time $\mathbf{x}(t_0)$. An integration of the tangent-linear model to time t_1 provides then $\mathbf{x}(t_1) = \mathbf{M}(t_0,t_1)\mathbf{x}(t_0)$. In a further step, the algorithm representing $\hat{\mathbf{x}}(t_1) \rightarrow \hat{\mathbf{x}}(t_0) = \mathbf{M}(t_0,t_1)^{\mathrm{T}}\hat{\mathbf{x}}(t_1)$ can be derived from the tangent-linear model. This algorithm is referred to as the adjoint model, which is short for adjoint of the tangent-linear model [62, 65]. The input variables for the adjoint model are the initial value of the trajectory $\mathbf{x}_r(t_0)$ and the adjoint perturbation at final time $\hat{\mathbf{x}}(t_1)$.

The development of variational data assimilation techniques (cf. [7]) provided a significant incentive to invest in the development of tangent-linear and adjoint models of numerical weather prediction models. Once tangent-linear and adjoint models are available, the leading singular vectors can be computed via an iterative solution of an eigenvalue problem (see Sec. 4.2.3).

4.2.2 Singular value decomposition of the propagator

The definition of singular vectors involves weighted L^2-norms to measure perturbation amplitude at initial time t_0 and final time t_1. Let us denote these norms by $\| \, \|_i$ and $\| \, \|_f$, respectively. The initial metric carries the information about the distribution of initial errors [66]. Let $\mathbf{C}_0$ denote an *estimate* of the initial error covariance. It can be assumed that $\mathbf{C}_0$ is a positive definite matrix. Otherwise, one can restrict the problem to the subspace in which $\mathbf{C}_0$ is positive definite. Then, we consider the initial norm given by the Mahanalobis distance

$$\|\mathbf{x}\|_i^2 = \mathbf{x}^{\mathrm{T}}\mathbf{C}_0^{-1}\mathbf{x}. \tag{11}$$

For a Gaussian distribution with zero mean and covariance $\mathbf{C}_0$, perturbations with equal initial norm (11) are equally likely to occur.

The final time norm is a convenient measure of forecast error. It provides a normalisation so that errors of different variables like temperature and wind can be compared. The final time norm can be either a norm or a semi-norm. One can introduce projections in state-space in order to focus on certain regions of the atmosphere or certain variables. We denote the final time (semi-)norm by

$$\|\mathbf{x}\|_f^2 = \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x}. \tag{12}$$

Singular vectors refer to a singular value decomposition involving the propagator for a fixed time interval $\mathbf{M} \equiv \mathbf{M}(t_0,t_1)$. The length of the time interval $t_{\mathrm{opt}} = t_1 - t_0$ is referred to as the optimisation time. In order to account for the norms, we consider linear variable transformations $\mathbf{x} \rightarrow \mathbf{C}_0^{-1/2}\mathbf{x}$ and $\mathbf{x} \rightarrow \mathbf{D}^{1/2}\mathbf{x}$. These transform into spaces in which initial time norm and final time norm are identical to the Euclidean norm. Now, we consider the singular value decomposition of the propagator that maps between the transformed spaces

$$\mathbf{D}^{1/2}\mathbf{M}\mathbf{C}_0^{1/2} = \widetilde{\mathbf{U}}\mathbf{S}\widetilde{\mathbf{V}}^{\mathrm{T}} \tag{13}$$

Here, $\mathbf{S}$ is the diagonal matrix containing the decreasing singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$. The orthogonal matrices $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ contain the non-dimensional left and right singular vectors as column vectors.

Let us denote by $\mathbf{V} = \mathbf{C}_0^{1/2}\widetilde{\mathbf{V}}$ and $\mathbf{U} = \mathbf{D}^{-1/2}\widetilde{\mathbf{U}}$ the dimensional singular vectors which represent the actual physical variables like wind and temperature. The transformed right singular vectors $\mathbf{V}$ are usually referred to as initial singular vectors or optimal perturbations (see below). The transformed left singular vectors $\mathbf{U}$

are referred to as normalised evolved singular vectors. The dimensional singular vectors satisfy the following orthonormality conditions

$$\mathbf{V}^{\mathrm{T}}\mathbf{C}_0^{-1}\mathbf{V} = \mathbf{I} \quad \text{and} \tag{14}$$

$$\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U} = \mathbf{I}, \tag{15}$$

where $\mathbf{I}$ denotes the identity matrix. In dimensional form, the singular value decomposition reads

$$\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}\mathbf{C}_0^{-1} \tag{16}$$

and the evolved j-th singular vector is given by

$$\mathbf{M}\mathbf{v}_j = \sigma_j \mathbf{u}_j. \tag{17}$$

Equation (17) implies that the singular value gives the ratio of the final time norm of the j-th singular vector to its initial time norm

$$\frac{\|\mathbf{M}\mathbf{v}_j\|_f}{\|\mathbf{v}_j\|_i} = \sigma_j \tag{18}$$

Singular vectors are also referred to as optimal perturbations because they are optimal in the sense that they maximise the ratio of final time norm to initial time norm. The leading singular vector $\mathbf{v}_1$ identifies the 1-dim subspace which maximises the ratio of norms. The j-th singular vector identifies the 1-dim subspace which maximises the ratio of norms in the subspace $\mathbf{C}_0^{-1}$ orthogonal to the space spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

Now, we consider the decomposition of the forecast error covariance matrix implied by the singular value decomposition [67]. Let $\mathbf{C}_1$ denote the estimate of the forecast error covariance at t_1 which is given by (10). The forecast error covariance matrix in the transformed space in which the final time norm is the Euclidean norm is given by $\widetilde{\mathbf{C}}_1 \equiv \mathbf{D}^{1/2}\mathbf{C}_1\mathbf{D}^{1/2}$. Using the singular value decomposition (13), we get

$$\widetilde{\mathbf{C}}_1 = \widetilde{\mathbf{U}}\mathbf{S}^2\widetilde{\mathbf{U}}^{\mathrm{T}} \tag{19}$$

This equation shows that the left singular vectors yield the empirical orthogonal function (EOF) decomposition of the normalised forecast error covariance estimate $\widetilde{\mathbf{C}}_1$. Thus, the leading initial singular vectors evolve into the leading EOFs of the forecast error covariance estimate.

4.2.3 Singular vectors in NWP models

The leading singular vectors can be identified approximately by iteratively solving eigenproblems. The non-dimensional right singular vectors are solutions of the eigenvalue problem

$$\mathbf{C}_0^{1/2}\mathbf{M}^{\mathrm{T}}\mathbf{D}\mathbf{M}\mathbf{C}_0^{1/2}\tilde{\mathbf{x}} = \sigma^2\tilde{\mathbf{x}} \tag{20}$$

and the initial singular vectors are solutions of the generalised eigenvalue problem

$$\mathbf{M}^{\mathrm{T}}\mathbf{D}\mathbf{M}\mathbf{x} = \sigma^2\mathbf{C}_0^{-1}\mathbf{x}. \tag{21}$$

For simple metrics, i.e. those for which the transformation $\mathbf{C}_0^{1/2}$ is available, the Lanczos algorithm [68] can be used to obtain the leading eigenvectors of the symmetric eigenvalue problem (20). Each iteration of the Lanczos algorithm involves an integration of the tangent-linear model and an integration of the adjoint model over the optimisation interval. For some metrics like the Hessian metric discussed later, the transformation $\mathbf{C}_0^{1/2}$ is not available. Then, a generalised Davidson method can be used to iteratively find the leading eigenvectors of (21) [69, 70].

4.2.4 Construction of initial perturbations from singular vectors

In order to construct perturbations of the initial conditions from singular vectors, an assumption has to be made concerning the distribution of initial errors. Consistent with the choice of the initial time metric in the singular vector computation, we assume that the initial errors are distributed according to a multi-variate Gaussian distribution $N(0,\mathbf{C}_0)$ with zero mean and covariance $\mathbf{C}_0$. This assumption appears appropriate in the absence of more detailed knowledge about the initial uncertainty. If we now consider the error as a random vector $\mathbf{x}$ having the distribution $N(0,\mathbf{C}_0)$, we can uniquely decompose $\mathbf{x}$ into a component in the space L spanned by the leading singular vectors and in a component in the space $L_\perp$ which denotes the $\mathbf{C}_0^{-1}$-orthogonal complement of L. Formally, this can be written as

$$\mathbf{x} = \underbrace{\mathbf{P}_L\mathbf{x}}_{\mathbf{x}_L} + \underbrace{\mathbf{P}_{L_\perp}\mathbf{x}}_{\mathbf{x}_{L_\perp}}, \tag{22}$$

where $\mathbf{P}_L$ and $\mathbf{P}_{L_\perp}$ denote the $\mathbf{C}_0^{-1}$-orthogonal projections on the subspaces L and $L_\perp$, respectively. The components $\mathbf{x}_L$ and $\mathbf{x}_{L_\perp}$ are independent random vectors that are both normally distributed with mean 0 and covariance $\mathbf{P}_L\mathbf{C}_0\mathbf{P}_L^{\mathrm{T}}$ and $\mathbf{P}_{L_\perp}\mathbf{C}_0\mathbf{P}_{L_\perp}^{\mathrm{T}}$, respectively. The singular vector based initial perturbation strategy assumes that only the component of the initial errors in the space of the fastest growing perturbations matters. To remain consistent with the underlying distribution of errors initial perturbations are constructed by sampling from the distribution $N(0,\mathbf{P}_L\mathbf{C}_0\mathbf{P}_L^{\mathrm{T}})$.

A sample of this distribution is obtained by linearly combining singular vectors using random numbers α_j as coefficients. Let $\mathbf{v}_j, j=1,\ldots,M$ denote a set of leading singular vectors computed with the initial time metric $\mathbf{C}_0^{-1}$. Then, the random vector

$$\sum_{j=1}^{M}\alpha_j\mathbf{v}_j, \tag{23}$$

has the distribution $N(0,\mathbf{P}_L\mathbf{C}_0\mathbf{P}_L^{\mathrm{T}})$ if and only if the coefficients α_j are independent normally distributed random numbers with zero mean and unit variance. This equivalences hinges on the fact that the singular vectors form a basis that is orthonormal with respect to $\mathbf{C}_0^{-1}$, Eq. (14). Special considerations concerning the orthonormality of the basis vectors are required if one uses more than one set of singular vectors [71].

A Gaussian distribution does not have compact support and occasionally the coefficients α can become so large that the resulting perturbations are unphysical. Initial perturbations with very large amplitude can trigger numerical instabilities in the integration. Therefore, it is appropriate to replace the Gaussian by a similar distribution that has compact support.

When the initial perturbations are based on (23), the sample mean perturbation is almost always not zero for a finite sample. This makes the ensemble mean in the early forecast ranges worse than an unperturbed forecast starting from the best estimate of the initial state. The sample mean perturbation can be forced to be zero by modifying the sampling approach. A particularly attractive way to achieve this is to enforce a plus-minus symmetry in the sample of initial perturbations. In this approach, pairs of perturbations $(\mathbf{x}_+,\mathbf{x}_-)$ are generated by sampling the coefficients α for $\mathbf{x}_+$ and setting $\mathbf{x}_- = -\mathbf{x}_+$. This method is motivated by the fact that the sign of a singular vector is arbitrary. Using the plus-minus symmetric sampling, the univariate sample distributions in the direction of any singular vector are invariant under the operation of changing the sign of the singular vector. Furthermore, the sample estimates of any odd moment vanishes. For the plain Gaussian sampling (23) this is only achieved in the limit of infinite ensemble size. Note, that there could be a potential problem with plus-minus symmetric perturbations if the ensemble is used to estimate a covariance matrix without using covariance localisation. In the linear regime of perturbation dynamics, using plus-minus symmetric perturbations effectively halves the rank of an already rank-deficient covariance matrix estimate. For

the Ensemble Transform Kalman Filter, a spherical simplex method was shown to be superior to plus-minus symmetric perturbations [72].

In practice, only a crude estimate of $\mathbf{C}_0$ is used in the singular vector computation. Therefore the standard deviation of the sampled Gaussian distribution is a parameter that needs to be determined empirically in order to obtain an adequate dispersion of the ensemble.

4.3 Initial perturbations in the operational ECMWF EPS

Now, we summarise the configuration used for the initial perturbations in the ECMWF EPS. Singular vectors are computed with an initial time norm and final time norm which is referred to as total energy metric. The squared norm is given by the spatially discretised version of the following integral

$$\frac{1}{2}\int_{p_0}^{p_1}\int_S \left(u^2+v^2+\frac{c_p}{T_r}T^2\right)\,\mathrm{d}p\,\mathrm{d}s+\frac{1}{2}R_d T_r p_r \int_S (\ln p_{\mathrm{sfc}})^2\,\mathrm{d}s \tag{24}$$

Here, $u, v, T, \ln p_{\mathrm{sfc}}$ refer to the perturbations of the zonal and meridional wind component, the temperature and logarithm of surface pressure. Furthermore, R_d, c_p, T_r, p_r denote the gas constant and the specific heat at constant pressure of dry air and a reference temperature and pressure, respectively. The integration extends over the entire volume of the model atmosphere with $\int_S \mathrm{d}s$ and $\int_{p_0}^{p_1}\mathrm{d}p$ denoting horizontal integration over the surface of the sphere and vertical integration using pressure p as coordinate. This metric has the advantage that it corresponds to a diagonal matrix in the model state space and thus permits the reduction to an ordinary eigenproblem (20). In [66], the authors compare singular vectors computed with several different simple initial norms. The study shows that total energy is the best of the limited number of considered initial norms in terms of the consistency of the spectral distribution of variance of singular vectors and analysis error estimates.

The singular vectors are optimised for maximum growth over a 48-hour period. This choice is motivated by two aspects: First, the optimisation time has to be sufficiently small for the linearisation to provide a good approximation of the dynamics of perturbations of the nonlinear forecast model that have an initial amplitude of typical analysis errors. Second, the optimisation time should be sufficiently long in order to obtain perturbations that grow rapidly enough in the later forecast ranges.

The expansion into spherical harmonics of the tangent-linear model is truncated at a total wavenumber of 42; this corresponds to a horizontal resolution of about 300 km. The tangent-linear model is linearised about a trajectory which is initialised from a short-range forecast (6-hour). The resulting singular vectors are very similar to ones obtained with a trajectory started from the latest analysis. Hence, the choice of the trajectory in the singular vector computation does not affect the ensemble skill [73]. The benefit of using a slightly less accurate trajectory for the linearisation is that more computing resources can be devoted to the singular vector computation and to the nonlinear forecasts in an operational weather prediction context.

Separate sets of singular vectors are computed for each hemisphere. Singular vectors for a specific region are obtained by replacing the propagator **M** in (20) with **PM**, where **P** denotes a projection of the perturbation to the region of interest. In the hemispheric computations, the evolved perturbations are localised to the extra-tropics of the respective hemisphere (30°–90°latitude). The amplification rates of the leading singular vectors for a particular hemisphere exhibit a seasonal cycle with the largest growth occurring during the dynamically more active winter [15]. Without localisation, the majority of singular vectors would be located in the winter hemisphere. By computing separate sets of singular vectors, equally many singular vectors can be obtained for each hemisphere.

The initial perturbations for the extra-tropics are based on initial singular vectors and on 48-hour linearly evolved singular vectors computed 48-hour prior to the ensemble start time. The component of initial pertur-

bations based on the evolved singular vectors has been introduced to represent the slower growing large-scale component of initial condition errors [74]. The evolved singular vectors are normalised using the total energy norm prior to being used in the Gaussian sampling. The normalisation makes the perturbations consistent with the representation of initial uncertainty implied by the total energy metric. Without normalisation the initial spread of the ensemble would exhibit a few large maxima that correspond to the fastest growing evolved singular vectors. For each hemisphere, extra-tropical initial perturbations are constructed from the leading 50 initial singular vectors and the leading 50 evolved singular vectors. Experimentation using fewer and using more extra-tropical singular vectors suggests that the skill of probabilistic predictions increases with increasing number of extra-tropical singular vectors up to about 50. We expect that the number of singular vectors beyond which the probabilistic skill does not improve significantly depends on the tangent-linear dynamics and the particular singular vector configuration, i.e. optimisation time and optimisation region. In the extra-tropical singular vector computation, the tangent-linear model used is simplified and does not contain parameterisations of physical processes apart from a simplified scheme for vertical mixing and surface friction [75].

The dynamics of perturbations in the tropics is sensitive to the representation of diabatic physics. Simplified versions of the nonlinear physical parameterisations of diabatic processes and their tangent-linear and adjoint models have been developed in [59]. Singular vectors optimised for the entire tropics with this simplified representation of diabatic processes in the tangent-linear model tend to exhibit spurious growth, i.e. the tangent-linear model grossly overestimated the growth of these structures in the nonlinear model [76]. Therefore, perturbations for the tropics have been limited to perturbations in the vicinity of tropical cyclones [77]. In order to obtain structures to which the evolution of the tropical cyclone is sensitive, it is necessary to localise the final time norm to the vicinity of the tropical cyclone. The leading five singular vectors are computed for each optimisation region targeted on a tropical cyclone. It has been shown that these perturbations are effective in generating spread among the tropical cyclone tracks in the ensemble (Fig. 3, [77]). A computation of these initial perturbations is triggered whenever a report of the presence of a tropical cyclone is received. The optimisation region for the tropical cyclone takes into account its likely position at optimisation time based on tracks from the previous ensemble forecast if available. During a transition of a tropical cyclone into the extra-tropics, the optimisation region will move into the optimisation region used for the extra-tropical singular vector computation. In order to avoid that the computation of perturbations for the tropical cyclone duplicates structures already represented in the extra-tropical set of singular vectors, the computation is restricted to the subspace orthogonal to the 50 leading extra-tropical singular vectors [71]. This approach permits statistically consistent initial perturbations even for overlapping optimisation regions.

The initial perturbations are obtained from the singular vectors via the pairwise symmetric Gaussian sampling method described in the previous section. Originally, singular vectors were combined via a selection and rotation algorithm [78]. The Gaussian distribution is truncated at ± 3 standard deviations to avoid excessively large amplitude perturbations that could trigger numerical instabilities. The scaling of the width of the Gaussian distribution is performed independently for each set of singular vectors. The singular vectors are compared with an estimate of the analysis error standard deviation provided by the four-dimensional variational assimilation system using an analysis error variance norm defined as L^2-norm of the perturbation scaled by the estimate of the analysis error standard deviation. The summation in the norm extends over the entire model grid and involves the horizontal wind components, temperature and the logarithm of surface pressure, i.e. the variables defining the a dry model state in grid point space. The standard deviation of the Gaussian for a set of singular vectors is set to

$$\beta = \gamma / \overline{\kappa}, \tag{25}$$

where $\overline{\kappa}$ denotes the average analysis error variance norm for the set of singular vectors. The parameter γ is determined empirically to yield adequate ensemble dispersion. Presently, the EPS uses $\gamma = 0.020$ for the extra-tropical singular vectors and $\gamma = 0.030$ for the singular vectors targeted on tropical cyclones. In the current system, the value of the average norm $\overline{\kappa}$ does not vary much from day to day and the standard deviation β

is basically set by the empirically determined scaling factor. However, in principle β could react to changes in the analysis uncertainty, say if several sources of observational data became unavailable suddenly and the analysis error estimate reflected this change. Figure 7 shows four realisations of initial condition perturbations in the mid-latitudes. One can distinguish perturbations with the typical westward tilt that is familiar of initial singular vectors and the broader and deeper structures of the evolved singular vectors. The standard deviation of the temperature perturbations is typically of the order of about 0.5 K with peak values of about 2 K in the dynamically most active regions. The 4D-Var estimate of the temperature analysis error is spatially more homogeneous. Values of the standard deviation are in a similar range of about 0.5–2 K in the mid-latitudes.

4.4 Discussion

The initial perturbation methodology of the ECMWF has evolved over the years and is likely to evolve further as our knowledge about the sources of uncertainty grows and as new opportunities emerge through improved numerical algorithms and increases in computing power. What are then promising directions of further improving the EPS initial perturbations?

The tangent-linear model used in the computation of extra-tropical singular vectors of the ECMWF EPS is only representing perturbation dynamics of the adiabatic part of the model apart from vertical mixing and surface drag. The impact of representing diabatic processes in the tangent-linear model on the structure and growth rate of extra-tropical singular vectors has been studied in [79] using the parameterisations developed in [59]. They identified that mainly due to the representation of the large-scale, i.e. resolved, condensation in the tangent-linear model the singular vectors become smaller scale and tend to grow faster. In order to resolve the smaller scales, it appears necessary to increase the horizontal resolution in the singular vector computation. At the same time, [79] recommend to shorten the optimisation time from 48 h to 24 h in order to remain in a regime in which the tangent-linear approximation is sufficiently accurate given increased growth rates and smaller spatial scales. Such a singular vector configuration has been tested in the EPS showing positive results for the prediction of some intense extra-tropical cyclones at forecast ranges of about 2 days [80]. However, overall, the shortening of the optimisation time has a slightly detrimental impact on the skill of probabilistic forecasts in the present EPS configuration. Recently, more accurate tangent-linear schemes have been developed for the large-scale condensation and the parameterised convection [60, 61]. Their use for computing singular vectors for the EPS is being explored.

The singular vectors in the ECMWF EPS are computed with a total energy metric at initial time which is a rather crude representation of the initial condition uncertainty. For instance, it does not account for spatial and temporal variations in initial uncertainty. It is possible to use more sophisticated estimates of the initial uncertainty in the initial norm of the singular vector computation. Estimates of the inverse initial error covariance matrix can be obtained from the Hessian of the cost function of variational assimilation algorithms [7]. It is possible to compute singular vectors with the Hessian as initial time metric by solving the generalised eigenproblem (21) [70]. These singular vectors are referred to as Hessian singular vectors. Initial uncertainty estimates based on the Hessian are consistent with the statistical assumptions made in that variational assimilation algorithm. Earlier work comparing singular vectors computed with the Hessian of a 3D-Var cost function as initial time metric concluded that the Hessian singular vectors had generally larger scale structures than the singular vectors computed with the total energy metric [74]. Work with a more recent version of the ECMWF data assimilation system (4D-Var and use of revised background error covariances with spatially more confined correlations) results in Hessian singular vectors that are more similar compared to total energy singular vectors in terms of their spatial scales and the baroclinic tilt against the vertical shear of the flow [81]. Figure 8 shows an example of the leading singular vector computed with three different initial time metrics: total energy, Hessian metric, and a partial Hessian metric that does not include the observation term of the cost function. While the former

Hessian metric provides an analysis error covariance metric, the latter provides a background error covariance metric. The singular vectors computed with the background error covariance metric are deeper and of larger horizontal scale than the singular vectors computed with analysis error covariance metric or the singular vectors computed with the total energy metric. Both the earlier 3D-Var Hessian singular vectors and the more recent 4D-Var Hessian singular vectors were tried as initial perturbation in the EPS. They did not improve the skill of the probabilistic forecasts compared to an EPS using singular vectors computed with the total energy metric [74, 81]. The degree of similarity between the more recent 4D-Var Hessian singular vectors and the total energy singular vectors suggests that there is little scope for improvement. The Hessian singular vector computations use a static specification of the background error covariance. This may severely limit the ability to describe flow dependent variations of initial uncertainty. In a recent publication [82], singular vectors computed with a flow dependent analysis error covariance metric are presented. The covariance was estimated from the analyses of an ensemble Kalman filter. Given improved estimates of the initial error distribution, it will be necessary to reassess the benefit of dynamically constrained sampling of this distribution compared to an unconstrained sampling.

5 Representation of model error

The laws of evolution which govern weather and climate, at least their physical aspects, are well known, and are accurately represented by sets of partial differential equations. These equations nonlinearly couple circulations with different scales and are thus difficult to solve analytically. To solve the governing equations numerically, we project them onto some chosen orthonormal basis, thus determining a set of (up to 10^8) coupled ordinary differential equations. Within these equations, the nonlinear effect of unresolved scales of motion are traditionally represented by simplified deterministic formulae, known as parametrisations. These parametrisations represent the bulk effect of subgrid processes within a grid box, and are justified in much the same way diffusive formulae are justified in statistical mechanics. Hence, for example, parametrisation of deep convection presumes the existence of an ensemble of deep convective plumes, in quasi-equilibrium with the larger scale environment. The associated parametrised convective tendency represents the bulk effect of these plumes in redistributing heat, momentum and water in the vertical column of a given grid box. Similarly, parametrisation of orographic gravity-wave drag presumes the existence of an ensemble of incoherent gravity waves which collectively are associated with a flux of momentum from the surface to some level of presumed wave breaking.

Parametrisations are by their nature approximations. Hence the parametrised convective or orographic tendencies, which represent the mean effect of these processes over many realisations, are usually different from the tendencies associated with the actual convective or orographic subgrid flow. Since the latter is not known, the parametrisation process is necessarily a source of uncertainty in numerical forecasts, and must therefore be represented explicitly in any ensemble forecast system. Without such a source of uncertainty either the ensemble will be underdispersive, or other sources of error, e.g. associated with observational uncertainty, will have to be inflated to prevent underdispersion. In this context, we note again that since the forecast model is used to assimilate observations in creating the initial conditions for a forecast, initial error includes a component due to model error. That is to say, when one speaks of forecast uncertainty as comprising initial error and model error, these two classes of error are not disjoint.

There are currently three general methods for representing model error: the multi-model ensemble, the perturbed parameter ensemble, and stochastic-dynamic parametrization. These different techniques have been discussed and compared in a recent comprehensive review [83] - here we give a brief overview.

The efficacy of the multi-model ensemble relies on the fact that models from different institutes around the world have been developed quasi-independently. Hence, within a multi-model ensemble there will typically

be a set of quite different convection schemes, orographic drag schemes, and numerical approaches. Multi-model seasonal-timescale ensemble forecasts have been shown to be more reliable than single model ensembles, particularly in the tropics [32].

In a multi-model ensemble forecast system, the size of the ensemble is limited by the number of available models, typically on the order of 10. In order to create large ensembles, the perturbed-parameter approach [84, 85] has been developed. In this methodology, individual uncertain parameters within the set of parametrisations of a single specific model are perturbed according to some estimates of parameter uncertainty. Hence, for example, elements of the ensemble could be run with the standard value of the convective entrainment parameter tripled. So far, most examples of perturbed parameter ensembles have been made in the context of climate change. In [86], the authors discuss the use of very short-range forecast tendencies to constrain perturbed parameter ensembles.

The multi-model ensemble assumes that on any given occasion, at least one of the multi-model parametrisations is capable of giving the correct grid-box mean tendency. In the perturbed parameter ensemble, it is assumed that the correct tendency can be obtained by a suitable parameter perturbation of a given parametrisation. The stochastic-dynamic parametrisation approach [87] does not assume that the correct tendency can be given by a deterministic bulk formula. Indeed, in the stochastic-dynamic approach it is assumed that the assumption of an ensemble of subgrid processes at any time-step and for any gridbox, is a flawed assumption. For example, in particularly active regions of convection a given grid box may be dominated by a single mesoscale convective complex, the very antithesis of the ensemble plume model. In the stochastic-dynamic approach, a specific realisation of the stochastic process is associated with a single specific realisation of the subgrid circulations.

A simple stochastic parametrisation has been discussed in [58], where it is shown that probabilistic forecast scores are improved by including such a scheme in the ensemble forecast system. This scheme is presently used to represent model uncertainty in the operational ECMWF EPS. A more sophisticated scheme based on the idea of backscattering of kinetic energy dissipated by the model [88] has been shown to be capable of ameliorating some of the more robust systematic errors in climate models associated with mid-latitude blocking [89]. The spatio-temporal forcing pattern in one version of this scheme is based on stochastic-dynamic cellular automata; in another version it is based on a spectral Markov process.

At present no comparative studies of these three techniques for representing model uncertainty have been performed. However, under the European Union project ENSEMBLES, a detailed comparison will be undertaken in seasonal and decadal forecast mode.

The representation of model uncertainty is still at a rudimentary stage of development. One technique which may prove useful in developing stochastic parametrisations is the method of coarse-grained budgets from cloud-resolving models. Early results in this direction [90] lend support to the simple stochastic scheme proposed in [58].

6 Conclusions

As emphasised many years ago by Tennekes et al [91], a prediction of the likely error of a weather forecast is as important as a prediction of any particular meteorological variable; predicting predictability is as important as predicting rainfall. However, because the atmosphere is a nonlinear dynamical system, the likely error of a weather forecast varies from initial state to initial state. That is to say, the growth of inevitable uncertainties and errors in making a weather forecast is flow dependent. Ensemble forecast systems have been developed to estimate such flow-dependent estimates of forecast uncertainty. In this paper we have focused on the development of the ECMWF medium-range ensemble prediction system, now operational for 15 years.

The inevitable uncertainties and errors that degrade the potential skill of a weather forecast include uncertainties in the observations, used to initialise the predictions, and in the forecast models themselves. These sources of error are not disjoint; for example, errors interpolating from point or pixel-scale observations to grid-point fields arise because of the finite truncation scale of the forecast model. Indeed, there remains considerable uncertainty in how to specify the statistical properties of the errors in observations and in the forecast model. For the former, there are issues associated with the spatio-temporal correlation of observation error from space-based platforms. For the models, the nature and upscale propagation of parametrisation error onto the resolved flow is still not well understood theoretically.

For these reasons we have discussed the different types of philosophy underlying perturbation strategies in ensemble prediction: those which develop perturbations within a framework where the statistical properties of uncertainty are assumed known, and those where such statistical properties are themselves uncertain and where conventional estimates are too conservative (leading to under-dispersive ensembles). The singular-vector strategy, used at ECMWF and discussed in this paper, lies within the latter type of approach. However, in the future, as more sophisticated methods to represent observation and model uncertainty are developed it may be possible to develop perturbation strategies, eg based on ensemble data assimilation techniques, from a given statistical estimate of uncertainty.

Ensemble prediction systems generate probability forecasts and a variety of methods are available to assess the skill of such probability forecasts, some of which are discussed in this paper. However, the real value of ensemble prediction systems are not the probability forecasts per se, but their ability to influence decisions across a range of applications sectors. A general methodology of how a reliable ensemble prediction system can be more valuable for decision making than a corresponding best-guess deterministic forecast was discussed. However, as operational ensemble prediction continues to develop, so specific examples of the value of ensemble prediction for decision making will increase. Such specific examples will most likely arise when a specific application model is coupled to each individual member of the ensemble prediction system. In this way, weather becomes an intermediate variable in the prediction system and a probabilistic forecast of the relevant user variable becomes the primary output from the ensemble. In order for such end-to-end systems to become viable in an operational context, it will be important for sufficient calibration runs to be made before any new operational system can be implemented.

Ensemble forecasting has now become an established technique in medium-range prediction, having been operational for 15 years. However, ensemble prediction systems have now been developed on almost all timescales, from nowcasting to climate change. Both meteorologists, and users of meteorological forecasts, now recognise that, no matter what the timescale of interest, there is little merit in producing forecasts without also producing corresponding forecasts of likely accuracy.

Acknowledgements

Andy Lawrence kindly provided Fig. 8. Comments by a reviewer and the editors helped improving the manuscript.

References

[1] J. M. Lewis, Roots of ensemble forecasting, Mon. Wea. Rev. 133 (2005) 1865–1885.

[2] G. E. Karniadakis, J. Glimm, Preface to special issue on uncertainty quantification in simulation science, J. Comp. Phys. 217 (2006) 1–4.

[3] P. F. J. Lermusiaux, Uncertainty estimation and prediction for interdisciplinary ocean dynamics, J. Comp. Phys. 217 (2006) 176–199.

[4] S. E. Cohn, An introduction to estimation theory, J. Meteor. Soc. Japan 75 (1997) 257–288.

[5] M. Fisher, M. Leutbecher, G. A. Kelly, On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, Quart. J. Roy. Meteor. Soc. 131 (2005) 3235–3246.

[6] T. Palmer, R. Hagedorn (Eds.), Predictability of Weather and Climate, Cambridge Univ. Press, 2006, 702 pp.

[7] I. M. Navon, Data assimilation for numerical weather prediction: A review, J. Comp. Phys.(this issue).

[8] E. N. Lorenz, Atmospheric predictability experiments with a large numerical model, Tellus 34 (1982) 505–513.

[9] S. Vannitsem, C. Nicolis, Lyapunov vectors and error growth patterns in a T21L3 quasigeostrophic model, J. Atmos. Sci. 54 (1997) 347–361.

[10] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130–141.

[11] E. N. Lorenz, A study of the predictability of a 28-variable atmospheric model, Tellus 17 (1965) 321–333.

[12] E. N. Lorenz, The Essence of Chaos, Univ. Washington Press, 1993.

[13] P. Lynch, The origins of computer weather prediction and climate modelling, J. Comp. Phys.(this issue).

[14] P. Janssen, Progress in ocean wave forecasting, J. Comp. Phys.(this issue).

[15] R. Buizza, T. N. Palmer, The singular-vector structure of the atmospheric global circulation, J. Atmos. Sci. 52 (1995) 1434–1456.

[16] R. Buizza, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, Y. Zhu, A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, Mon. Wea. Rev. 133 (2005) 1076–1097.

[17] C. E. Leith, Theoretical skill of Monte Carlo forecasts, Mon. Wea. Rev. 102 (1974) 409–418.

[18] M. Rodwell, Comparing and combining deterministic and ensemble forecasts: How to predict rainfall occurrence better, ECMWF newsletter (106) (2006) 17–23, (available at http://www.ecmwf.int/publications/newsletters/pdf/106.pdf).

[19] G. Candille, O. Talagrand, Evaluation of probabilistic prediction systems for a scalar variable, Quart. J. Roy. Meteor. Soc. 131 (2005) 2131–2150.

[20] A. H. Murphy, A new vector partition of the probability score, J. Appl. Met. 12 (1973) 595–600.

[21] D. S. Richardson, Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size, Quart. J. Roy. Meteor. Soc. 127 (2001) 2473–2489.

[22] F. Atger, Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration, Mon. Wea. Rev. 131 (2003) 1509–1523.

[23] T. Gneiting, A. E. Raftery, A. H. Westveld III, T. Goldman, Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, Mon. Wea. Rev. 133 (2005) 1098–1118.

[24] H. Hersbach, Decomposition of the continous ranked probability score for ensemble prediction systems, Weather and Forecasting 15 (2000) 559–570.

[25] J. S. Whitaker, A. F. Loughe, The relationship between ensemble spread and ensemble mean skill, Mon. Wea. Rev. 126 (1998) 3292–3302.

[26] O. Talagrand, R. Vautard, B. Strauss, Evaluation of probabilistic prediction systems, in: Proc. of Workshop on Predictability, ECMWF, Reading, UK, 1997, pp. 1–25.

[27] A. H. Murphy, The value of climatological, categorical and probabilistic forecasts in the cost-loss situation, Mon. Wea. Rev. 105 (1977) 803–816.

[28] D. S. Richardson, Skill and relative economic value of the ECMWF ensemble prediction system, Quart. J. Roy. Meteor. Soc. 126 (2000) 649–667.

[29] T. Palmer, R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher, L. Smith, Ensemble prediction: A pedagogical perspective, ECMWF newsletter 106 (2006) 10–17.

[30] T. N. Palmer, The economic value of ensemble forecasts as a tool for risk assessment: from days to decades, Quart. J. Roy. Meteor. Soc. 128 (2002) 747–774.

[31] J. J. Tribbia, D. P. Baumhefner, Scale interactions and atmospheric predictability: An updated perspective, Mon. Wea. Rev. 132 (2004) 703–713.

[32] T. N. Palmer, A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, M. Deque, E. Diez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J. Gueremy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J. Terres, M. C. Thomson, Development of a European multimodel ensemble system for seasonal-to-interannual prediction (Demeter), Bull. Am. Meteor. Soc. 85 (2004) 853–872.

[33] I.-L. Frogner, T. Iversen, Targeted ensemble prediction for northern Europe and parts of the north Atlantic ocean, Tellus 53 A (2001) 35–55.

[34] C. Temperton, M. Hortal, A. Simmons, A two-time-level semi-Lagrangian global spectral model, Quart. J. Roy. Meteor. Soc. 127 (2001) 111–127.

[35] M. Hortal, The development and testing of a new two-time-level semi-Lagrangian scheme (SETTLS) in the ECMWF forecast model, Quart. J. Roy. Meteor. Soc. 128 (2002) 1671–1687.

[36] A. Untch, M. Hortal, A finite-element scheme for the vertical discretization of the semi-Lagrangian version of the ECMWF forecast model, Quart. J. Roy. Meteor. Soc. 130 (2004) 1505–1530.

[37] A. J. Simmons, A. Hollingsworth, Some aspects of the improvement in skill of numerical weather prediction, Quart. J. Roy. Meteor. Soc. 128 (2002) 647–677.

[38] R. Buizza, D. S. Richardson, T. N. Palmer, Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles, Quart. J. Roy. Meteor. Soc. 129 (2003) 1269–1288.

[39] M. Untch, A., M. Miller, R. Buizza, P. Janssen, Towards a global meso-scale model: The high resolution system TL799L91 and TL399L62 EPS, ECMWF newsletter (108) (2006) 6–13, (available at http://www.ecmwf.int/publications/newsletters/pdf/108.pdf).

[40] R. Buizza, J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, T. N. Palmer, F. Vitart, The new ECMWF VAREPS (variable resolution ensemble prediction system), Quart. J. Roy. Meteor. Soc.(submitted, manuscript available at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/401-500/tm499_rev.pdf).

[41] P. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, H. L. Mitchell, A system simulation approach to ensemble prediction, Mon. Wea. Rev. 124 (1996) 1225–1242.

[42] P. L. Houtekamer, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, B. Hansen, Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, Mon. Wea. Rev. 133 (2005) 604–620.

[43] G. Burgers, P. J. van Leeuwen, G. Evensen, Analysis scheme in the ensemble Kalman filter, Mon. Wea. Rev. 126 (1998) 1719–1724.

[44] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, Ensemble square root filters, Mon. Wea. Rev. 131 (2003) 1485–1490.

[45] C. H. Bishop, B. J. Etherton, S. J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, Mon. Wea. Rev. 129 (2001) 420–436.

[46] Z. Toth, E. Kalnay, Ensemble forecasting at NMC: The generation of perturbations, Bull. Am. Meteor. Soc. 74 (1993) 2317–2330.

[47] Z. Toth, E. Kalnay, Ensemble forecasting at NCEP and the breeding method, Mon. Wea. Rev. 125 (1997) 3297–3319.

[48] R. Daley, Atmospheric Data Analysis, Cambridge Univ. Press, 1991, 457 pp.

[49] M. Wei, Z. Toth, R. Wobus, Y. Zhu, C. H. Bishop, X. Wang, Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP, Tellus 58A (2006) 28–44.

[50] M. Wei, Z. Toth, R. Wobus, Y. Zhu, Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system, Tellus*(submitted)*.

[51] P. L. Houtekamer, J. Derome, Methods for ensemble prediction, Mon. Wea. Rev. 123 (1995) 2181–2196.

[52] T. M. Hamill, C. Snyder, R. E. Morss, A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles, Mon. Wea. Rev. 128 (2000) 1835–1851.

[53] J. L. Anderson, The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results, Mon. Wea. Rev. 125 (1997) 2969–2983.

[54] N. Bormann, S. Saarinen, G. Kelly, J.-N. Thépaut, The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data, Mon. Wea. Rev. 131 (2003) 706–718.

[55] Z. Liu, F. Rabier, The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study, Quart. J. Roy. Meteor. Soc. 128 (2002) 1367–1386.

[56] T. M. Hamill, J. S. Whitaker, Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches, Mon. Wea. Rev. 133 (2005) 3132–3147.

[57] D. S. Wilks, Effects of stochastic parametrizations in the Lorenz '96 system, Quart. J. Roy. Meteor. Soc. 131 (2005) 389–407.

[58] R. Buizza, M. Miller, T. N. Palmer, Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Quart. J. Roy. Meteor. Soc. 125 (1999) 2887–2908.

[59] J.-F. Mahfouf, Influence of physical processes on the tangent-linear approximation, Tellus 51 A (1999) 147–166.

[60] A. M. Tompkins, M. Janisková, A cloud scheme for data assimilation: Description and initial tests, Quart. J. Roy. Meteor. Soc. 130 (2004) 2495–2517.

[61] P. Lopez, E. Moreau, A convection scheme for data assimilation: Description and initial tests, Quart. J. Roy. Meteor. Soc. 131 (2005) 409–436.

[62] J.-F. Lacarra, O. Talagrand, Short-range evolution of small perturbations in a barotropic model, Tellus 17 (1988) 321–333.

[63] I. Gilmour, L. A. Smith, R. Buizza, On the duration of the linear regime: Is 24 hours a long time in weather forecasting?, J. Atmos. Sci. 58 (2001) 3525–3539.

[64] C. A. Reynolds, T. E. Rosmond, Nonlinear growth of singular-vector-based perturbations, Quart. J. Roy. Meteor. Soc. 129 (2003) 3059–3078.

[65] R. Errico, What is an adjoint model?, Bull. Am. Meteor. Soc. 78 (1997) 2577–2591.

[66] T. N. Palmer, R. Gelaro, J. Barkmeijer, R. Buizza, Singular vectors, metrics, and adaptive observations, J. Atmos. Sci. 55 (1998) 633–653.

[67] M. Ehrendorfer, J. J. Tribbia, Optimal prediction of forecast error covariances through singular vectors, J. Atmos. Sci. 54 (1997) 286–313.

[68] G. H. Golub, C. F. Van Loan, Matrix Computations, 3rd Edition, Johns Hopkins University Press, 1996, 694 pp.

[69] E. R. Davidson, The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices, J. Comp. Phys. 17 (1975) 87–94.

[70] J. Barkmeijer, M. Van Gijzen, F. Bouttier, Singular vectors and estimates of the analysis-error covariance metric, Quart. J. Roy. Meteor. Soc. 124 (1998) 1695–1713.

[71] M. Leutbecher, Representing initial condition uncertainties with multiple sets of singular vectors optimised for different criteria, Quart. J. Roy. Meteor. Soc.*(submitted)*.

[72] X. Wang, C. H. Bishop, S. J. Julier, Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?, Mon. Wea. Rev. 132 (2004) 1590–1605.

[73] M. Leutbecher, On ensemble prediction using singular vectors started from forecasts, Mon. Wea. Rev. 133 (2005) 3038–3046.

[74] J. Barkmeijer, R. Buizza, T. N. Palmer, 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System, Quart. J. Roy. Meteor. Soc. 125 (1999) 2333–2351.

[75] R. Buizza, Sensitivity of optimal unstable structures, Quart. J. Roy. Meteor. Soc. 120 (1994) 429–451.

[76] J. Barkmeijer, R. Buizza, T. N. Palmer, K. Puri, J.-F. Mahfouf, Tropical singular vectors computed with linearized diabatic physics, Quart. J. Roy. Meteor. Soc. 127 (2001) 685–708.

[77] K. Puri, J. Barkmeijer, T. N. Palmer, Ensemble prediction of tropical cyclones using targeted diabatic singular vectors, Quart. J. Roy. Meteor. Soc. 127 (2001) 709–731.

[78] F. Molteni, R. Buizza, T. N. Palmer, T. Petroliagis, The ECMWF ensemble prediction system: Methodology and validation, Quart. J. Roy. Meteor. Soc. 122 (1996) 73–119.

[79] M. M. Coutinho, B. J. Hoskins, R. Buizza, The influence of physical processes on extratropical singular vectors, J. Atmos. Sci. 61 (2004) 195–209.

[80] A. Walser, M. Arpagaus, C. Appenzeller, M. Leutbecher, The impact of moist singular vectors and horizontal resolution on short-range limited-area ensemble forecasts for two European winter storms, Mon. Wea. Rev. 134 (2006) 2877–2887.

[81] A. R. Lawrence, M. Leutbecher, T. N. Palmer, Comparison of total energy and Hessian singular vectors: Implications for observation targeting, Quart. J. Roy. Meteor. Soc.(manuscript in preparation).

[82] M. Buehner, A. Zadra, Impact of flow-dependent analysis-error covariance norms on extratropical singular vectors, Quart. J. Roy. Meteor. Soc. 132 (2006) 625–646.

[83] T. N. Palmer, G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, M. Leutbecher, Representing model uncertainty in weather and climate prediction, Annu. Rev. Earth Planet. Sci. 33 (2005) 163–193.

[84] J. M. Murphy, D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, D. A. Stainforth, Quantifying uncertainties in climate change using a large ensemble of global climate model predictions, Nature 430 (2004) 768–772.

[85] D. A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, M. R. Allen, Uncertainty in predictions of the climate response to rising levels of greenhouse gases, Nature 433 (2005) 403–406.

[86] M. Rodwell, T. N. Palmer, Assessing model physics with initial forecast tendencies: application to climate change uncertainty, Quart. J. Roy. Meteor. Soc. XXX, *(submitted)*.

[87] T. N. Palmer, A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models, Quart. J. Roy. Meteor. Soc. 127 (2001) 279–304.

[88] G. Shutts, A kinetic energy backscatter algorithm for use in ensemble prediction systems, Quart. J. Roy. Meteor. Soc. 131 (2005) 3079–3102.

[89] T. Jung, T. N. Palmer, G. J. Shutts, Influence of stochastic parametrization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model, Geophys. Res. Letters 32 (2005) L23811.

[90] G. J. Shutts, T. N. Palmer, Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem, J. Clim. 20 (2007) 187–202.

[91] H. Tennekes, A. P. M. Baede, J. D. Opsteegh, Forecasting forecast skill, in: Proc. of Workshop on Predictability in the Medium and Extended Range, ECMWF, Reading, UK, 1987, pp. 277–302.

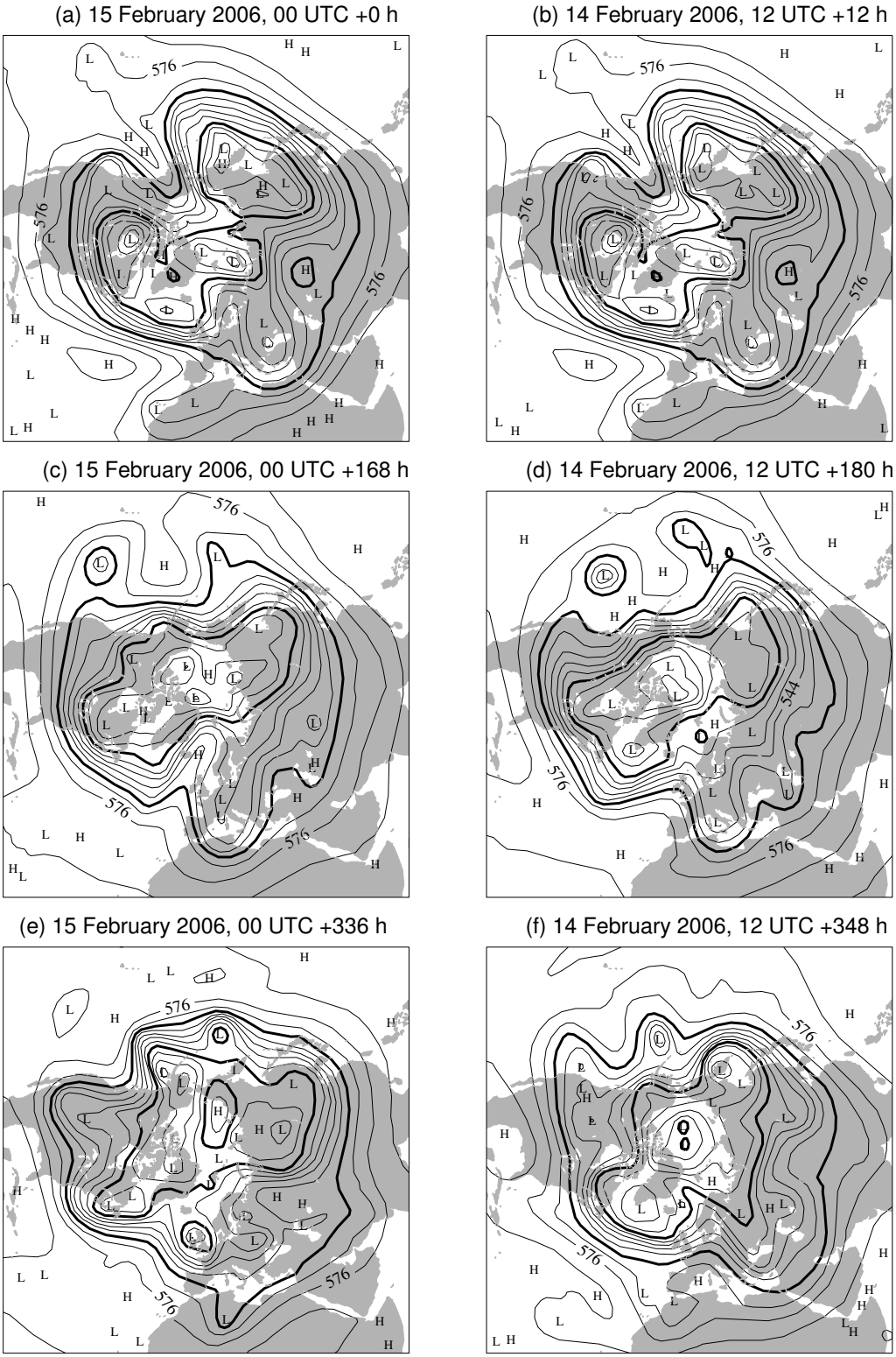


Figure 1: Geopotential of the 500 hPa pressure surface (contours every $800\,\mathrm{m^2\,s^{-2}}$*) for two forecasts started 12 hours apart: 15 February 2006, 00 UTC for (a,c,e) and 14 February 12 UTC for (b,d,f). The forecast steps are 0.0 and 0.5 d in (a) and (b); 7.0 and 7.5 d in (c) and (d); 14.0 and 14.5 d in (e) and (f). Polar stereographic map of the Northern Hemisphere.*

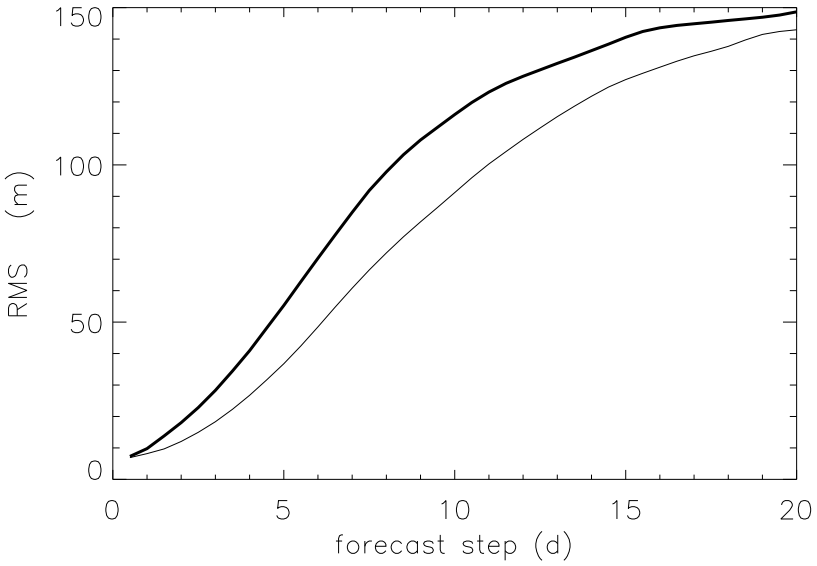


Figure 2: 500 hPa geopotential height RMS error (heavy line) of the unperturbed control forecast of the ECMWF EPS and RMS difference of two subsequent (12-hour lag) control forecasts valid at the same time (thin line). RMS values have been computed for the Northern Hemisphere (north of 20°N) for the period Dec 2005–Feb 2006 using data from the new high-resolution EPS (truncation at wavenumber 399, 62 levels in the vertical) which became operational in Feb '06.

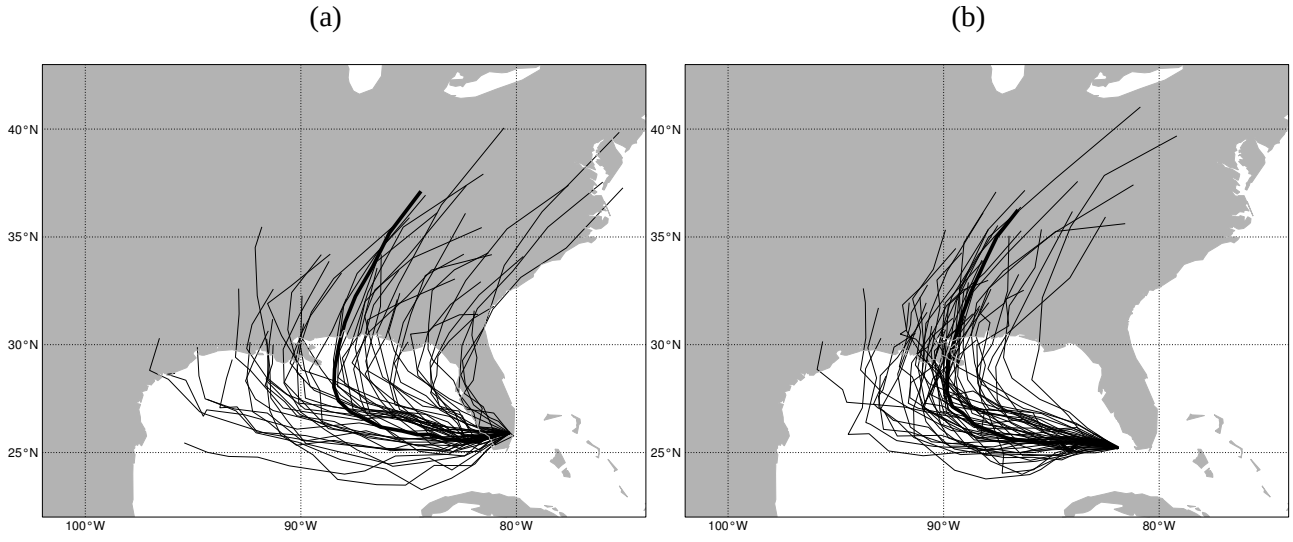


Figure 3: Five-day track forecasts for hurricane Katrina made with the ECMWF deterministic forecast system (heavy line) and the ECMWF ensemble prediction system (thin lines). Forecasts initialised on 26 August 2005, 00 UTC (a) and 12 UTC (b).

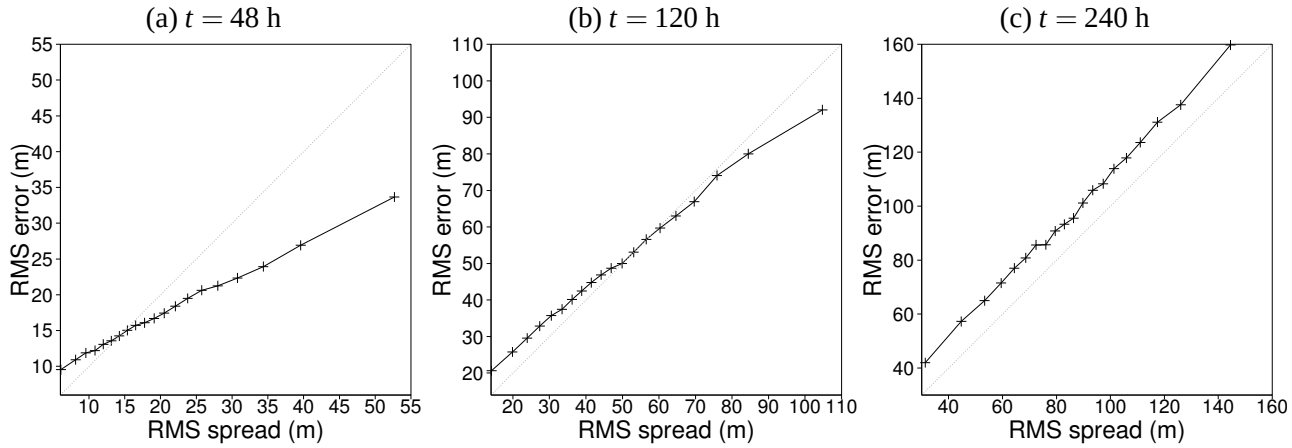


Figure 4: RMS error of the ensemble mean averaged over subsamples of increasing predicted RMS spread for 500 hPa geopotential height in the mid-latitudes 35°–65°N. Forecast ranges 2 d (a), 5 d (b) and 10 d (c). 89 cases in Feb–Apr 2006. The 20 bins are equally populated.

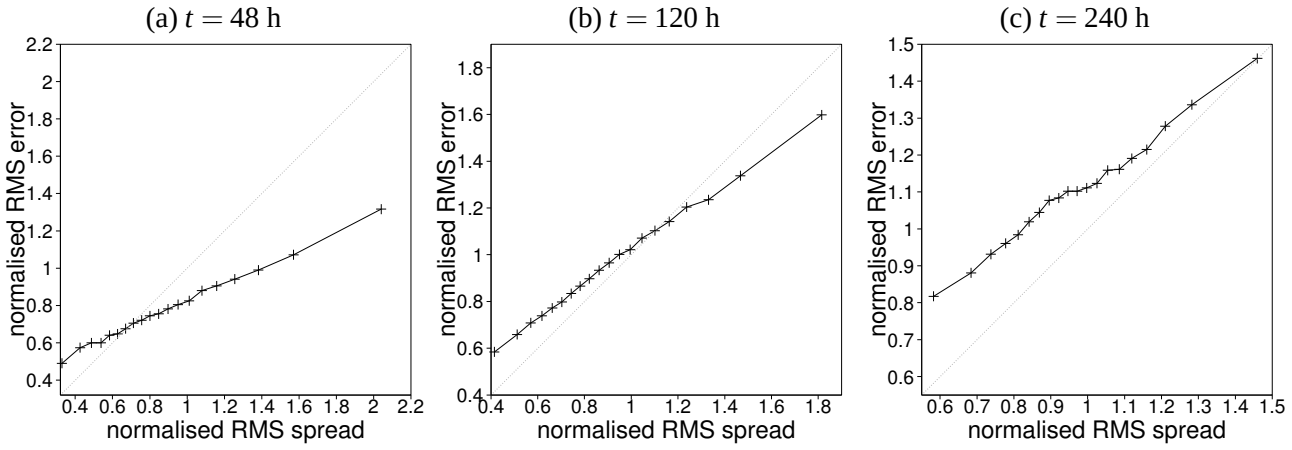


Figure 5: As previous Fig. but for 500 hPa geopotential normalised by the seasonally averaged spread in order to geographically homogenise the data.

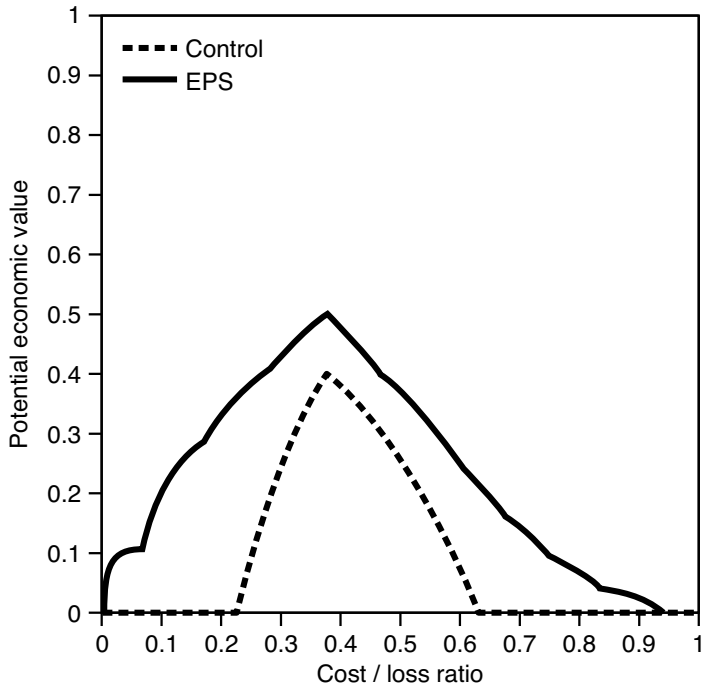


Figure 6: Potential economic value for the ECMWF EPS (solid) and control forecast (dashed) based on six-day forecasts of whether or not it will rain (24-hour precipitation exceeding 1mm, August-October 2005, Europe).

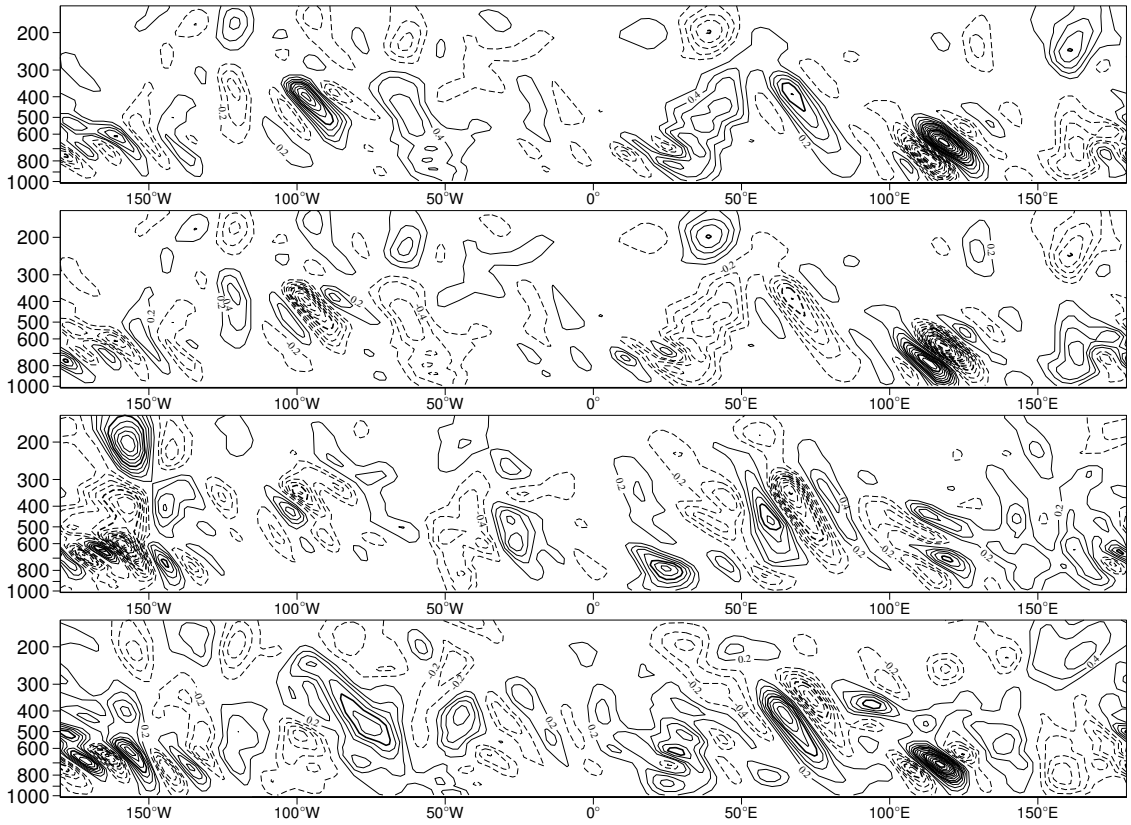


Figure 7: Temperature (every 0.2 K, negative values dashed, positive values solid, heavy contour every 1 K) in a vertical section at 50°N for four realisations of initial condition perturbations corresponding to perturbed forecasts 1, 2, 5 and 50 (from top to bottom) of the ECMWF EPS initialised at 21 March 2006, 00 UTC. Vertical axis: pressure (hPa), horizontal axis: degrees longitude.

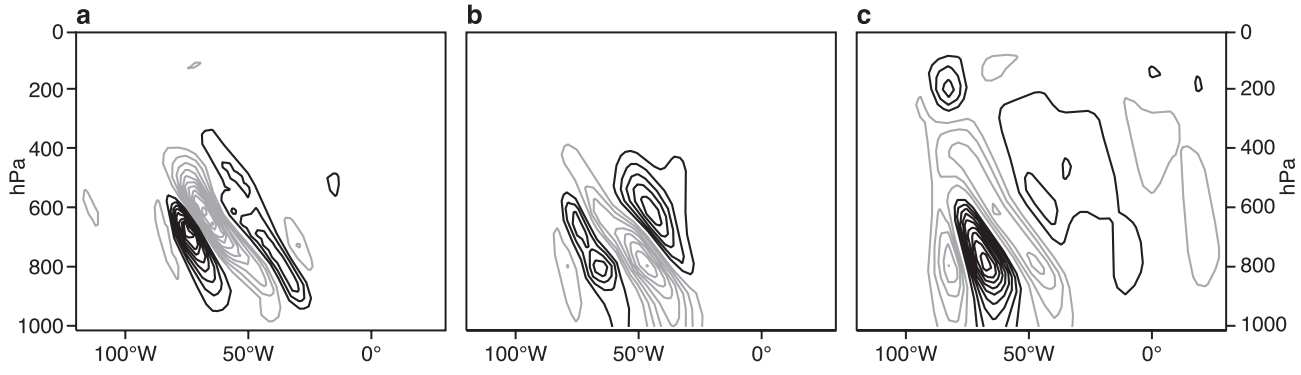


Figure 8: Temperature perturbation (every 0.01 K, positive: black, negative: grey, zero contour omitted) in vertical section at 50°N of leading singular vector computed with different initial metrics: (a) total energy, (b) Hessian of 4D-Var cost function, (c) Hessian of background part of 4D-Var cost function.