Ensemble prediction: a pedagogical perspective

Tim Palmer, Roberto Buizza, Renate Hagedorn, Andy Lawrence, Martin Leutbecher and Lenny Smith

1 Some conceptual questions on ensemble prediction

Despite the fact that the ECMWF Ensemble Prediction System (EPS, *Molteni et al* 1996) brings additional value to ECWMF's dissemination products through its ability to assess flow-dependent weather risk, the EPS is a less straightforward tool to use than the more traditional deterministic forecast. Not surprisingly, therefore, conceptual questions are sometimes asked about the EPS. Here are some examples.

What is the relationship between the spread and skill within the EPS? If the northern hemisphere RMS error of a typical ensemble member is routinely larger than that of the corresponding deterministic forecast, does this imply that this ensemble member is simply a degradation of the deterministic forecast? Should we be striving to reduce the RMS error of ensemble members relative to the deterministic forecast? Does it make a difference if we ask how many ensemble members are better than the deterministic forecast locally, compared with hemispherically? Are the baroclinically-tilted structures often seen in the EPS initial perturbations consistent with our knowledge of analysis error?

Perhaps most important of all is this question: Is it really worth running the EPS when ECMWF has such a high-quality deterministic forecast?

We try to answer these and other related questions in this article.

2 Properties of a perfect Ensemble Prediction System

The scientific basis for ensemble forecasting is encapsulated in Figure 1. In a nonlinear system, here the *Lorenz* (1963) model, but the principle applies to the real atmosphere too, the growth of initial uncertainties during a given forecast period is flow dependent. From some initial states the forecast evolution can be highly predictable, from other initial states it can be highly unpredictable. The ensemble allows us to forecast this flow-dependent predictability.

The illustration in Figure 1 has been formed by integrating the Lorenz equations many times from an ensemble of initial states which form a small circle in the nonlinear system's state space. It is to be imagined that the radius of the circle is some measure of the expected amplitude of initial error.

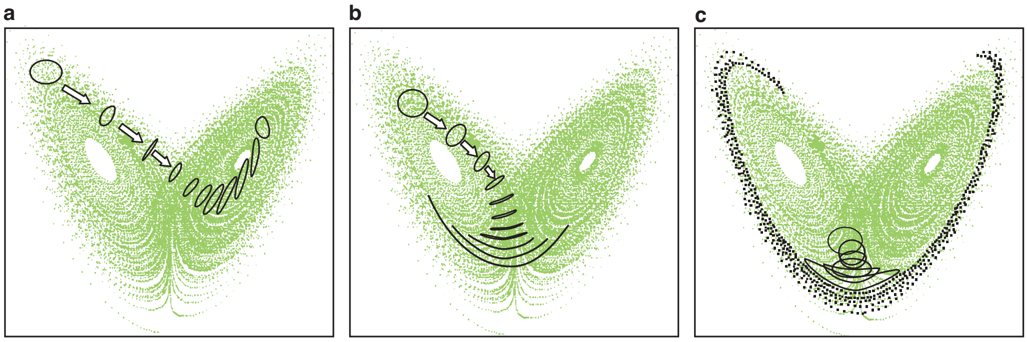

Fig. 1 Scientific basis for ensemble forecasting. In a nonlinear system the growth of initial uncertainty is flow dependent – here illustrated with the Lorenz (1963) model. The set of initial conditions (black circle) is located in different regions of the attractor in (a), (b) and (c).

More generally, the data from which the initial state (at time t_0) is constructed (e.g. atmospheric observations), do not determine this state precisely, but rather determine some probability density function $\rho(X,t_0)$. Essentially $\rho(X,t_0)\, dV$ denotes the probability at time t_0 that the true value X_T of the variable X (for example, 2-metre temperature at London's Heathrow airport) lies in the small volume dV in state space. The objective of an EPS is to estimate the corresponding forecast probability density function $\rho(X,t)$ at forecast time $t>t_0$. In theory, $\rho(X,t)$ can be obtained from $\rho(X,t_0)$ by integrating an equation called the Liouville equation, or its generalisation, the Fokker-Planck equation. In practice, these equations are difficult to solve, even for simple dynamical systems. Instead, and consistent with the methodology applied to obtain Figure 1, $\rho(X,t)$ is estimated by multiple sampling of $\rho(X,t_0)$, integrating each random drawing forwards in time using the given forecast model. Hence, at time t, we can define a perfect EPS as an accurate sampling $\{X_i\}1 \le i \le N$ of the underlying density function $\rho(X,t)$, see Figure 2.

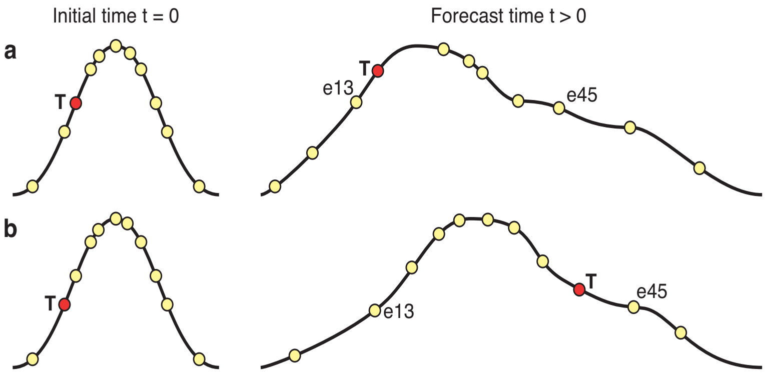

Fig. 2 A schematic showing the evolution of initial probability distributions of truth, to a forecast probability distributions of truth, together with an sample of ensemble members () from a perfect EPS. Truth is shown by the letter "T". In the examples shown (a) could represent a forecast of 2-metre Heathrow temperature today, with (b) representing a forecast of 2-metre Heathrow temperature from last week, or (a) could represent a forecast of 2-metre Heathrow temperature today, whilst (b) could represent a forecast of 2-metre Washington temperature today.

Suppose this procedure is repeated every day, and over a season or so. What mean properties would we expect such an EPS to have? One basic quantity of interest is the second moment of the ensemble - the spread. When the EPS spread is large, then a deterministic forecast from the most likely estimate of initial state (the 4DVAR initial state) will be an unreliable estimate of truth. Conversely, when the EPS spread is small, the corresponding deterministic forecast should be reliable. But what relationship between the spread of the ensemble and the skill, say, of the ensemble-mean deterministic forecast is desirable? This can be answered by considering a perfect EPS, which constitutes a perfect sampling of the underlying probability distribution of the true state of the atmosphere, "truth".

3 The relationship between spread and ensemble mean RMS error

In a perfect ensemble, i.e. a perfect sampling of the underlying probability distribution of truth, then, over a large number of ensemble forecasts, the statistical properties of the true value X_T of X are identical to the statistical properties of a member of the ensemble, X_e (when that member is removed from the ensemble). For the following analysis of spread and skill, we assume that the ensemble size N is sufficiently large that removing one member from the ensemble does not materially affect the results. Hence, for example, the mean squared distance of the J-th member $X_e(J)$ from the ensemble mean $\langle X_e \rangle$ is identical to the mean squared error of the ensemble mean

$$\overline{\left\| X_e(J) - \langle X_e \rangle \right\|^2} = \overline{\left\| X_T - \langle X_e \rangle \right\|^2} \qquad (1)$$

where $\langle \ldots \rangle$ denotes the expectation value with respect to a particular ensemble forecast, and $\overline{\ldots}$ denotes an average over many such ensemble forecasts. Equation (1) holds for any J and it can be applied to a scalar quantity X or to a vector $\boldsymbol{X}$. In the latter case, $\|\ldots\|$ should be understood as the Root Mean Square (RMS) or the Euclidean norm. Taking the expectation $\langle \ldots \rangle$ of Equation (1) yields

$$\overline{\left\langle \left\| X_e - \langle X_e \rangle \right\|^2 \right\rangle} = \overline{\left\| X_T - \langle X_e \rangle \right\|^2} . \qquad (2)$$

Equation (2) implies that the time-mean ensemble spread about the ensemble-mean forecast, should equal the time-mean RMS error of the ensemble-mean forecast.

Figure 3 shows that the relation between spread and ensemble-mean RMS error of the ECMWF EPS is good - though the ensemble is slightly under-dispersive in terms of 850 hPa temperature, and, in the early range, slightly over-dispersive in terms of 500 hPa height. Recently a comparison has been made between operational and quasi-operational EPS systems at different weather centres around the world (*Buizza et al.*, 2005). It was found that the ECMWF system has the best balance between spread and skill, relative to the ensemble mean forecast, throughout the forecast range.

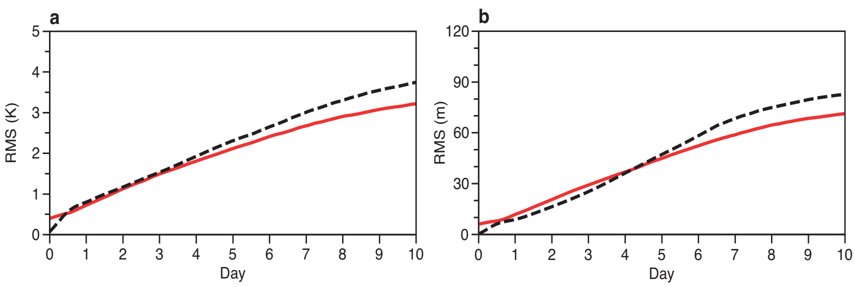

Fig. 3 Spread about ensemble (red solid) mean versus ensemble mean RMS error (black dashed) from the EPS for the northern hemisphere extra-tropics. (a) 850 hPa temperature and (b) 500 hPa height. Average over 39 cases from July 2004 to June 2005 using model cycle Cy29r2 (operational since 28 June 2005).

4 The relationship between the RMS error of the ensemble mean and the RMS error of an ensemble member

What are the implications of a good balance between spread and skill? Figure 2 shows two schematic probability distribution functions (PDFs) for a perfect EPS. These could represent forecasts of Heathrow temperature today and Heathrow temperature last week, or alternatively forecasts for Heathrow and Washington temperatures today. In Figure 2(a), truth is (by chance) close to ensemble member 13 and far from ensemble member 45, whilst in Figure 2(b), the converse is (by chance) true. How far is truth from an ensemble member on average over many cases?

In a perfect ensemble, we have

$$\overline{\left\langle \left\| X_e - X_T \right\|^2 \right\rangle} = \overline{\left\langle \left\| \left(X_e - \left\langle X_e \right\rangle \right) + \left(\left\langle X_e \right\rangle - X_T \right) \right\|^2 \right\rangle} \\ = 2\,\overline{\left\langle \left\| X_T - \left\langle X_e \right\rangle \right\|^2 \right\rangle} \qquad (3)$$

where the last equality exploits Equation (2) from Box A and the fact that the term involving the (inner) product vanishes because $\left\langle X_e - \left\langle X_e \right\rangle \right\rangle = 0$. In the short range, where the ensemble mean approximates well the unperturbed control forecast we have

$$\overline{\left\langle \left\| X_e - X_T \right\|^2 \right\rangle} = 2\,\overline{\left\langle \left\| X_T - X_c \right\|^2 \right\rangle} \qquad (4)$$

where X_c denotes the forecast value associated with the unperturbed deterministic control forecast. Equations (3) and (4) imply that the RMS distance between a perturbed ensemble member and truth will be, on average, $\sqrt{2} - 1 \approx 41\%$ larger than the distance between the ensemble mean, and in the short range the control, and truth.

Diagnostics of the ECMWF EPS are qualitatively consistent with this property of a perfect EPS. However, this fact has led to some conceptual difficulties amongst users of the EPS. *Does it mean that a perturbed EPS member is no better than a forecast which has simply been degraded everywhere relative to the control?* No! For example, the circulation over the northern hemisphere as a whole comprises a number of quasi-independent synoptic systems – i.e. a forecast PDF is multi-dimensional, with the number of dimensions corresponding to the number of effective degrees of freedom in the northern hemisphere flow. In a randomly-chosen member from a perfect EPS we can expect some of these synoptic systems to be more accurately predicted than the control, but others will not. By contrast, a uniformly-degraded deterministic forecast will, by construction, be everywhere worse than the control. The difference is critical.

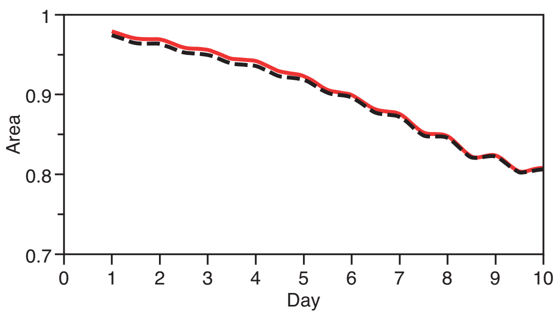

Fig. 4 Area under the Relative Operating Characteristic for positive anomalies of 850 hPa temperature for the operational EPS (red solid) and for an experiment in which the amplitude of the EPS perturbations has been reduced (black dashed; initial singular vectors by 30%, evolved singular vectors by 50%). Average for the northern hemisphere extra-tropics over 29 cases in April/May 2005. Both experiments used model cycle Cy29r2.

Conversely, if we tried to make each perturbed member more skilful relative to the control, will this make a better EPS? No! A simple way to make perturbed members more skilful, is to reduce the amplitude of the initial perturbations. Figure 4 shows results from a set of experiments where just this has been done. The skill of the resulting EPS has degraded, even though individual perturbed members are, on average, more skilful. Confusing? The problem with reducing spread is that the resulting EPS suffers from being tied too closely to the "apron strings" of the control. When the control forecast is evolving through an unstable and therefore unpredictable part of state space, the resulting EPS will not give a realistic indication of the magnitude of this unpredictability and the resulting probabilistic forecast will be over-confident.

Consider a related question: how many times should we expect, in a perfect EPS, a perturbed ensemble member to be "better than" the control forecast, in the early range of the forecast when the control is essentially the same as the ensemble-mean forecast? The answer to this question depends on how large an area we base our assessment of "better than". For a variable like Heathrow 2-metre temperature (with one-dimensional Gaussian PDF), it can be shown (see the Appendix) that a perturbed ensemble member has a 35% chance of being closer to truth than the control - hence in a perfect EPS, 35% of ensemble members should be "better" than the control. However, recall that if an ensemble member is close to truth at Heathrow, it need not be close to truth at Washington, and vice versa. The larger the area over which the (deterministic) skill of the ensemble member is validated, i.e. the larger the dimension of the forecast PDF, the smaller is the probability that a randomly-chosen member will be more skilful than the control. This effect can be quantified by considering a multi-dimensional Gaussian (corresponding to multiple degrees of freedom in the flow), and again asking how many times a perturbed ensemble member is "better than" the control forecast using RMS error as measure (see Appendix for mathematical details). For a 2-dimensional Gaussian 28% of members are better, for a 10-dimensional Gaussian 7% of members are better, and for a 100-dimensional Gaussian only 10^{-4} % of members are better!

Again, this result causes confusion. In a specific ensemble forecast, if we plot the northern hemisphere RMS error of the perturbed ensemble members of an ensemble on top of the control or high resolution deterministic forecast, then because there are so many degrees of freedom over the whole northern hemisphere, it is likely, from the argument above, that none of the members will be more skilful than the control. On the other hand, as discussed above, any one perturbed member may well be more skilful than the control over a specific region, such, for example, as Southern England. This is illustrated by Figure 5 which shows the percentage of perturbed forecast with smaller RMS error than the control forecast for regions of various sizes.

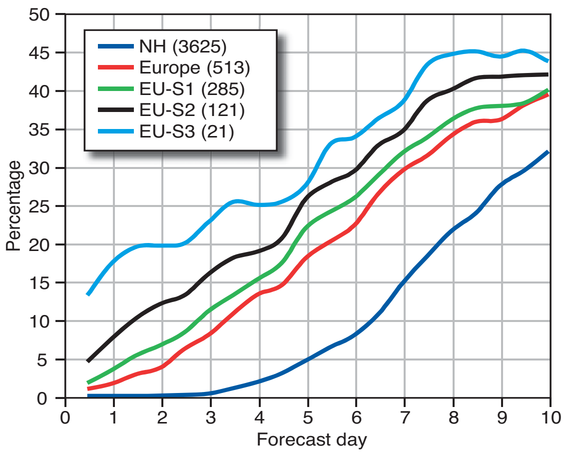

Fig. 5 Percentage of perturbed forecasts with lower RMS error than the control forecast for regions of various sizes; the number of verification grid points is given in parentheses (NH refers to the northern hemisphere extra-tropics and EU-S1-3 are European sub-domains of increasingly smaller size corresponding to large, medium-size and small member states).

The dependence of the number of ensemble members more skilful than the control on the number of degrees of freedom in the flow, is a reason why this type of diagnostic is not calculated routinely, and is certainly not one of the standard measures of skill against which the EPS is assessed. So, this raises the following question: What types of diagnostic are useful for assessing the performance of the EPS against the control or the high-resolution forecast? Indeed, can we assess quantitatively whether it really is worth running the EPS when ECMWF has such a high-quality deterministic forecast? The following sections address this question.

5 The EPS ensemble mean versus high-resolution deterministic forecast

The simplest product from the EPS is the ensemble-mean forecast. How does this product compare with the high-resolution deterministic forecast? *Rodwell* (2006) has shown that in terms of 500 hPa height, the ECMWF high-resolution deterministic forecast outperforms the EPS ensemble mean in the first few days of the forecast. This is not surprising; in terms of 500 hPa height, the EPS ensemble mean is essentially equal to the control forecast in the first couple of days of the forecast, and the control forecast is run at a lower resolution than the high-resolution deterministic forecast. However, the results are more interesting for variables like precipitation or potential vorticity (an intrinsic model variable – i.e. based on wind, pressure and temperature – with a spectrum of variability which is more comparable with "sensible weather" than 500 hPa height, at sub-cyclone scales). For these variables, the EPS ensemble mean is, on average, virtually as skilful as the high-resolution deterministic forecast in the early ranges of the forecast (and more skilful thereafter), despite the EPS being run at lower resolution. The reason for this is that fields like precipitation and potential vorticity (unlike 500 hPa height) have significant partially-unpredictable scales, even in the short range. The nonlinear filtering effect of the ensemble mean is effective in removing such unpredictable scales.

6 EPS versus deterministic forecast for binary decision making

The real value of the EPS over deterministic forecasting lies in decision making particularly for users who can quantify their "value at risk", i.e. the value of assets at risk to specific types of adverse weather event, and can take mitigating action at known cost.

Here is an example, which appears facetious, but illustrates the principle well. A colleague once phoned on a Monday morning, wanting to know whether or not it was going to rain the coming Saturday evening. He said he was having a garden party, and wanted to know whether or not to hire a marquee. He had to decide whether or not to hire the marquee in the next couple of hours. It was explained that predicting rainfall with certainty, so far ahead and for such a small area (his back garden), was virtually impossible; at best it would only be possible to give a probabilistic assessment of whether or not it would rain. What use is that, he asked? It was enquired whether the Queen was coming to the party. If the Queen was coming, then the marquee should be hired if the probability of rain exceeds 1% (i.e. if any member of the EPS predicts rain). On the other hand, if the queen was not coming but the town mayor was, then perhaps the marquee should be hired if there is more than a 10% chance of rain. However, if the party was just for friends from the pub, then perhaps it was only necessary to hire the marquee if the chance of rain exceeds 70%.

The value of the EPS against the deterministic control for such binary decision making is assessed routinely at ECMWF in the form of Potential Economic Value (Figure 6). Here the x-axis of Figure 6 denotes the user cost-loss ratio. We can relate the cost-loss ratio to the probabilistic threshold above. For example, suppose the colleague valued his potential knighthood at £50,000; this would be the value at risk if the Queen got wet. If hiring the marquee costs £500 (mitigating cost), then it would be appropriate to decide to hire the marquee if the probability of rain exceeded C/L=1%, i.e. if just one EPS member forecasts rain. On the other hand, the value of local business at risk if the town mayor got wet might only be worth £5,000, in which case the relevant cost/loss ratio would only be 10%.

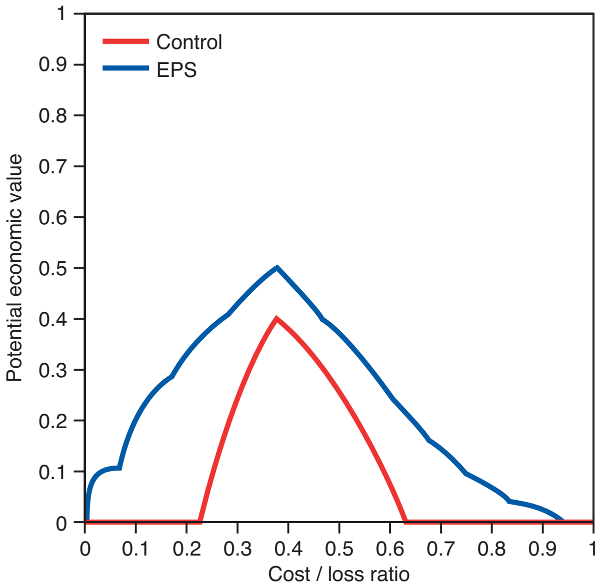

Fig. 6 Potential economic value for the ECMWF EPS control and the EPS itself based on forecasts of whether or not it will rain (24-hour precipitation exceeding 1 mm, August-October 2005, Europe). The potential economic value ranges between 0 (climatological forecast) and 1 (perfect forecast). For users with either low cost/loss ratio, or high cost/loss ratio, the control itself has no value for decision making (over and above decisions made with knowledge of the climatological frequency of rain).

The colleague responded that neither the queen nor the mayor was coming, but the mother in law was! On this basis, he decided he would hire the marquee if the probability of rain exceeded 25%. The EPS for Saturday showed the probability of rain was 10%. He didn't hire the marquee. (It didn't rain, and the colleague was a convert to probability forecasting!)

Realising the true economic value of the EPS requires knowledge of the forecast customers' specific circumstances. Perhaps this will be a key role for the forecaster in the future – a detailed interaction with the customer to determine the most appropriate probabilistic thresholds tailored to his or her specific needs.

7 EPS versus deterministic forecasts for weather trading

Not all decisions are simple binary decisions. Consider a simple gambling game – perhaps not so different to that played by energy traders – where you are betting on the Heathrow temperature seven days from now. Should you just bet on one temperature, or spread your bets across a range of temperatures, e.g. in proportion to the EPS-based probability of occurrence? Assume the "casino" you are betting against has determined the payout for a correctly-forecast temperature, based on a Gaussian distribution whose mean is the ECMWF high-resolution forecast of

Heathrow temperature, and whose standard deviation is taken from past error statistics of the high-resolution forecast. This is the so-called Weather Roulette problem first posed by Leonard Smith (London School of Economics) and Mark Roulston (Pennsylvania State University). The gamble starts on the first of January with an initial stake of £1. All the winnings are reinvested. Based on day 7 forecasts, Figure 7(a) shows that, after a year, the gambler using the EPS will have made more than £10^{30} against the casino! It turns out that the EPS gambler will win against the casino at all forecast ranges, though the payout is largest at about day 6-7.

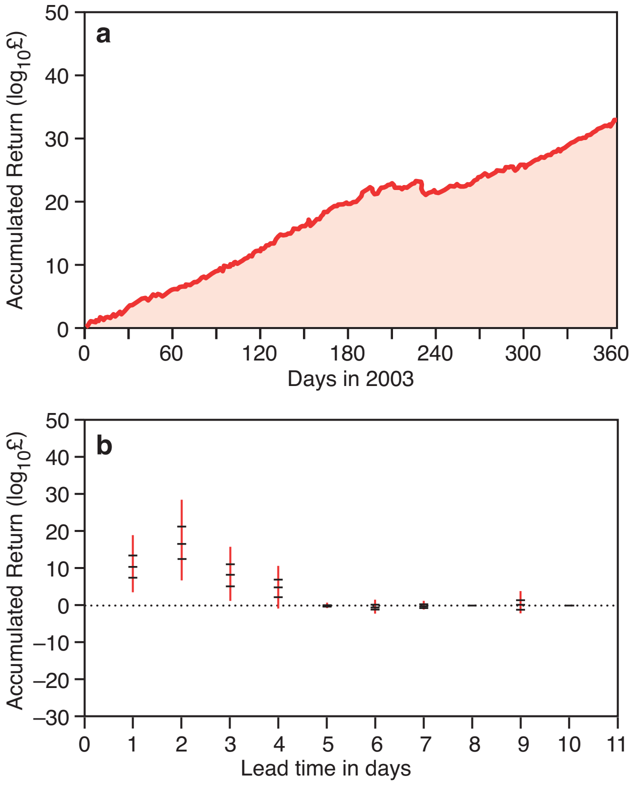

Fig. 7 Weather Roulette. (a) Return accumulated over 1 year when betting against the casino which sets the odds for 2-metre temperature at Heathrow seven days ahead according to the dressed high-resolution deterministic forecast, whereas the gambler distributes the available money in proportion to dressed EPS probabilities. (b) Accumulated return after 1 year of betting on 2-metre temperatures at Heathrow for lead-times from 1 to 10 days. Here the odds are set by the dressed EPS probabilities and the gambler distributes the available money in proportion to the optimal blend of high-resolution and EPS probabilities. The vertical bars represent the range of returns determined by bootstrapping the data, with horizontal lines marking the 5%/95% intervals and the median.

Suppose the gambler had access to the high-resolution deterministic. Could he improve his strategy by combining the high-resolution deterministic forecast with the EPS. Figure 7(b) shows that for lead-times up to 4 days, a betting system based on an optimal blend of high-resolution and EPS probabilities leads to a positive return when played against odds based solely on the EPS. However, after about day 4 there appears little extra value in adding the high-resolution deterministic forecast to the EPS. (*Rodwell*, 2005, discusses the potential impact of adding the high-resolution deterministic forecast to the EPS in terms of precipitation.)

Weather Roulette is an example of a validation technique which compares the EPS and deterministic forecast in a form where both have been optimally dressed in the form of probability forecasts. It clearly demonstrates the value of the EPS throughout the forecast range. These results confirm other studies that has shown that the ECMWF EPS provides valuable forecasts for business applications in energy demand prediction (*Taylor & Buizza* 2003) and in pricing financial instruments (*Taylor & Buizza* 2006).

8 EPS perturbation methodology

Sometimes it is asked whether the EPS is superior to a simple lagged ensemble comprising some of the most recent high-resolution forecasts. It is superior for a number of reasons. Firstly, it is impossible to create a probability forecast with any substantial resolution using just five or so members (with more than about five members, the lagged members become too unskilful at the effective initial time, to represent analysis error). Matters are actually worse than this, since the individual forecasts in a lagged ensemble are partially correlated with respect to one another. That is, on average, the error covariances in a lagged ensemble are significantly larger than the error covariances between members of the EPS. Figure 8 shows a comparison of the EPS with a lagged ensemble comprising the five most recent high-resolution deterministic forecasts. Figure 8(a) shows that the percentage of ensemble members better than the control is similar in both cases, whilst Figure 8(b) shows that the skill of the EPS is substantially greater than that of the lagged ensemble. A Weather Roulette analysis supports this conclusion: the EPS outperforms the simple lagged high-resolution forecasts.

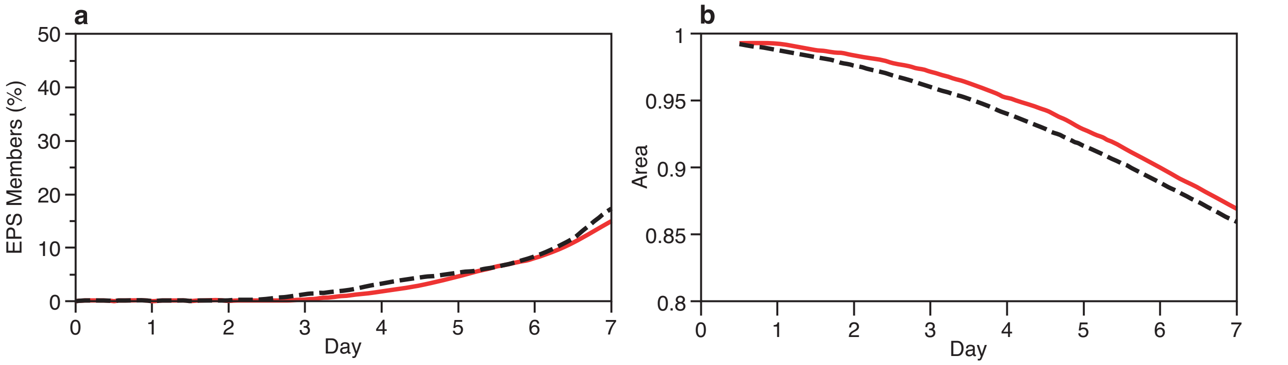

Fig. 8 Comparison of the EPS (red solid) against a lagged ensemble (black dashed) comprising the most recent five high-resolution deterministic forecasts of the 500 hPa height for the northern hemisphere extra-tropics. (a) Percentage of members with lower RMS error than control forecast. (b) Area under the Relative Operating Characteristic for positive anomalies of 500 hPa height. December-February, 2004/05, 90 cases.

The initial perturbation strategy for the ECMWF EPS is to draw randomly from an initial Gaussian PDF based on (a) the leading initial-time singular vectors of the first 48 hours of the forecast flow, and (b) the evolved singular vectors from the previous 48 hours (e.g. *Molteni & Palmer*, 1993, *Barkmeijer et al.*, 1999). The former are rapidly-growing, small-scale perturbations, the latter are weakly-growing, large-scale perturbations.

Figure 9(a) shows a typical initial singular vector as used in the EPS. It has sometimes been questioned whether such baroclinically-tilting structures really are a feature of analysis errors. Similarly, the uniqueness of these singular vector structures has also been questioned, since they depend on the choice of an initial metric. So, can the use of singular vectors for initial EPS perturbations be justified from sound physical principles?

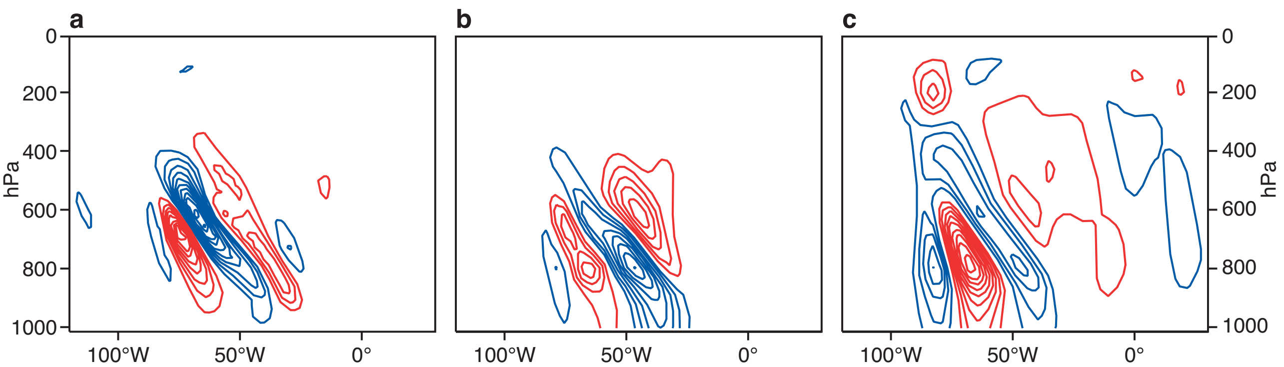

Fig. 9 Example of a singular vector (temperature cross-section at 50°N): (a) based on a total energy inner product, (b) based an analysis error covariance matrix inner product, and (c) based on a background error covariance matrix inner product (positive values in red; negative values in blue).

The natural inner product to use for singular-vector calculations is the analysis error covariance metric, since, evolved to forecast time, these singular vectors map directly onto the eigenvectors of the forecast error covariance matrix (*Ehrendorfer & Tribbia*, 1997). Figure 9(a) shows a leading singular vector calculated with respect to a total energy inner product. *Palmer et al*. (1998) argued that such an inner product should approximate well the inner product formed from the analysis error covariance metric.

It is now possible to calculate singular vectors using the 4DVAR analysis error covariance matrix. Figure 9(b) shows the same singular vector as Figure 9(a), but using the analysis error covariance metric, rather than the total energy metric. Figure 9(b) is very similar to Figure 9(a), suggesting that the EPS perturbations are indeed consistent with the statistics of analysis error. It is interesting to note that the structure of these singular vectors is strongly influenced by the presence of the observation error covariance matrix in the total analysis error covariance matrix. The role of observations is to constrain large-scale well-observed parts of the analysis. Without this constraint, i.e. by not including the observation error covariance matrix in the analysis error covariance matrix, the leading singular vectors are broader and deeper than they would otherwise be – more similar to breeding vectors (Figure 9(c)).

9 The ECMWF EPS: a valuable tool for decision making

The EPS is a valuable tool for decision making in applications sensitive to weather. Certain properties of the EPS have been studied, and some conceptual misunderstandings have been addressed. Above all, there can be little doubt that the resources devoted by ECMWF to the EPS are well justified.

On the other hand, there is certainly scope for improving the EPS, and ECMWF will be working with partners from the Member States on many aspects of the EPS. Such improvements will include:

- Increase in EPS resolution from T_L255L40 to T_L399L62;
- Unified ensembles for medium-range and monthly timescales (*Buizza et al* 2006);
- Development of back statistics from latest model cycles to calibrate probabilities;
- Incorporation of moist processes in the extra-tropical singular vector computations;
- Development of stochastic parametrisations to represent model error;
- Use of ensemble data assimilation in place of evolved singular vectors;
- Development of statistical schemes which will allow incorporation of high-resolution deterministic forecast into the EPS probability products;
- Comparison of ECMWF EPS against THORPEX grand multi-model ensemble.

In particular, the development of stochastic parameterisation and ensemble data assimilation will lead to a more realistic representation of model and initial uncertainties in the tropics, where the current EPS is, overall, underdispersive.

More information about ensemble methods for forecasting predictability can be found in "*Predictability of Weather and Climate*", edited by Tim Palmer and Renate Hagedorn, which is due to be published by Cambridge University Press in 2006. The book addresses predictability from the theoretical to the practical points of view, on timescales from days to decades.

Appendix. Perfect ensembles sampled from Gaussian distributions

This appendix investigates how often a member of an ensemble is better than the control forecast for a particular *perfect ensemble* scenario. This idealized situation is adopted as it permits a semi-analytical solution. We assume the following:

- The system and forecasts of it are n-dimensional vectors.
- The control forecast (most likely state) is an unbiased estimate of the true state.
- The error of the control forecast (control-minus-truth) is distributed according to an isotropic Gaussian distribution.
- The "ensemble" is given by the same Gaussian distribution. The results that will be discussed are independent of ensemble size. The ensemble could consist of any number of members drawn from the Gaussian distribution or alternatively one can consider the Gaussian distribution itself as the probabilistic forecast.
- Thus, the control-minus-truth differences and the control-minus-ensemble member differences are distributed according to the same isotropic Gaussian distribution. Without loss of generality, we can assume that the control forecast is zero (otherwise we can discuss everything in terms of differences with respect to the control forecast).
- The Euclidean norm will be used to measure the error of a forecast.

Let us start with the one-dimensional case. The Gaussian with standard deviation σ is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

The probability of an ensemble member x to be closer to the true state y than the control forecast 0 is given by the double integral:

$$\rho_1 = \int_{\mathbb{R}} p(y) \int_{|x-y|<|y|} p(x)\,dx\,dy$$

where $\mathbb{R}$ denotes the set of real numbers. In this equation, the integral over x yields the probability that an ensemble state x is closer to a given true state y than the control. These probabilities are then weighted with the probability that y occurs in the outer integral. Numerical evaluation yields ρ_1=0.35.

Now, we turn to the n-dimensional case. We consider an isotropic Gaussian distribution. The probability that an ensemble member x is closer to truth y in the Euclidean norm than the control forecast 0 can be expressed by the $2n$-dimensional integral:

$$\rho_n = \int_{\mathbb{R}^n} p(\mathbf{y}) \int_{\|\mathbf{x}-\mathbf{y}\|<\|\mathbf{y}\|} p(\mathbf{x}) d^n x \, d^n y$$

where

$$p(x_1,..,x_n) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{1}{2} \sum_{j=1}^{n} \frac{x_j^2}{\sigma^2} \right)$$

and where $\|\mathbf{z}\| = \left(\sum_{j=1}^{n} z_j^2 \right)^{1/2}$ denotes the Euclidean norm.

As in the one-dimensional case, ρ_n is independent of the standard deviation σ. Exploiting the spherical symmetry of the Gaussian PDF, the $2n$-dimensional integral can be reduced to a three-dimensional integral for any n. The latter integral can be evaluated numerically (Table 1). The results show that as the dimension increases a perturbed forecast is less likely to be better than the control forecast. For dimensions larger than 100, the probability drops to values below 10^{-6}.

	Dimension n								
	1	2	3	4	5	10	20	50	100
ρ_n	0.35	0.28	0.22	0.18	0.16	0.07	0.02	4×10^{-4}	1×10^{-6}

Table 1 Probability that the RMSE of a perturbed member is smaller than the RMSE of the control forecast for an isotropic Gaussian in n dimensions.

References

Barkmeijer, J., Buizza, R., & **Palmer, T. N.**, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2333-2351.

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

Buizza, R., 2005: EPS skill improvements between 1994 and 2005. *ECMWF Newsletter No. 104*, ECMWF, Shinfield Park, Reading RG2 9AX, UK, 10-14.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei & **Y. Zhu**, 2005: A comparison of the ECMWF, MSC and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076-1097.

Buizza, R., Bidlot, J., Fuentes, M., Holt, G., Hamrud, M., Palmer, T. N., Vitart, F., & **Wedi, N.**, 2006: The new ECMWF Variable Resolution Ensemble Prediction System (VAREPS). In this Proceedings of the 10th workshop on Meteorological Operational Systems, ECMWF, November 2005. Available from ECMWF, Shinfield Park, Reading, RG2-9AX, UK.

Ehrendorfer, M. & **J. Tribbia**, 1997: Optimal prediction of forecast error covariances through singular vectors. *J.Atmos.Sci.*, **54**, 286-313.

Lorenz, E.N., 1963. Deterministic non-periodic flow. *J.Atmos.Sci.*, **42**, 433-471.

Molteni, F. & **T.N.Palmer**, 1993: Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.*, **119**, 269-298.

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

Palmer, T.N., R.Gelaro, J. Barkmeijer, & **R. Buizza**, 1998: Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.*, **55**, 633-653.

Rodwell, M., 2006: Comparing and combining deterministic and ensemble forecasts: How to predict rainfall better. *ECMWF Newsletter No. 106*, available from ECMWF, Shinfield Park, Reading, RG2-9AX, U.K., pg 17-23.

Taylor, J., & Buizza, R., 2003: Using Weather Ensemble Prediction in Energy Demand Forecasting. *Intern. J. Forecast.*, **19**, 57-70.

Taylor, J., & **Buizza, R.**, 2006: Density forecasting for weather derivative pricing. *Intern. J. Forecast.*, in press.