united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

SMR.1403 - 1

Spring Colloquium on the Physics of Weather and Climate: REGIONAL WEATHER PREDICTION MODELLING AND PREDICTABILITY (8 – 19 April 2002)

"Chaos and Ensemble Forecasting"

Andrea MONTANI
Regional Meteorological Service
ARPA Emilia-Romagna
Viale Silvani 6
40122 Bologna
Italy

These are preliminary lecture notes, intended only for distribution to participants

Chaos and weather prediction January 2000

By **Roberto Buizza**[1]

European Centre for Medium-Range Weather

Abstract

The weather is a chaotic system. Small errors in the initial conditions of a forecast grow rapidly, and affect predictability. Furthermore, predictability is limited by model errors due to the approximate simulation of atmospheric processes of the state-of-the-art numerical models.

These two sources of uncertainties limit the skill of single, deterministic forecasts in an unpredictable way, with days of high/poor quality forecasts randomly followed by days of high/poor quality forecasts.

Two of the most recent advances in numerical weather prediction, the operational implementation of ensemble prediction systems and the development of objective procedures to target adaptive observations are discussed.

Ensemble prediction is a feasible method to integrate a single, deterministic forecast with an estimate of the probability distribution function of forecast states. In particular, ensemble can provide forecasters with an objective way to predict the skill of single deterministic forecasts, or, in other words, to forecast the forecast skill. The European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS), based on the notion that initial condition uncertainties are the dominant source of forecast error, is described.

Adaptive observations targeted in sensitive regions can reduce the initial conditions' uncertainties, and thus decrease forecast errors. More generally, singular vectors that identify unstable regions of the atmospheric flow can be used to identify optimal ways to adapt the atmospheric observing system.

Table of contents

1. Corresponding author address: Dr Roberto Buizza, ECMWF, Shinfield Park, Reading RG2 9AX. Tel.:0118-9499653. Email: buizza@ecmwf.int

1. INTRODUCTION

A dynamical system shows a chaotic behavior if most orbits exhibit sensitive dependence (Lorenz 1993). An orbit is characterized by sensitive dependence if most other orbits that pass close to it at some point do not remain close to it as time advances.

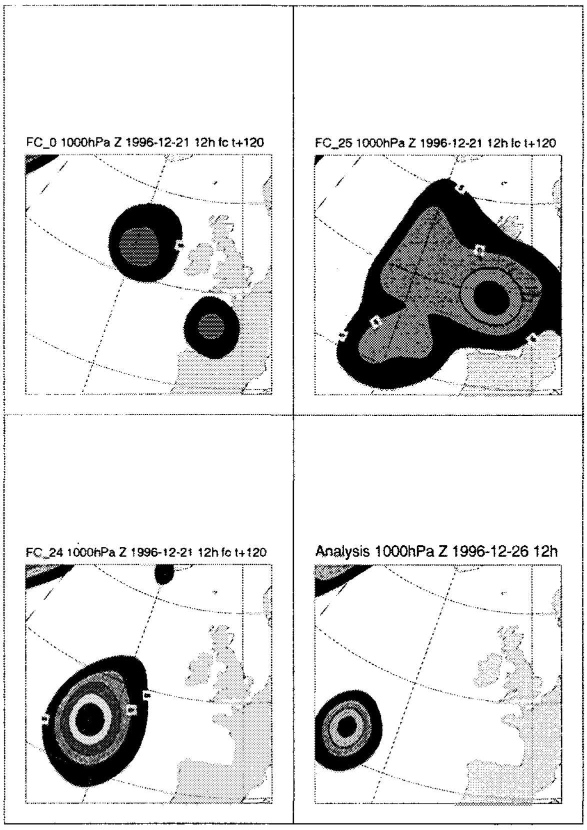

Figure 1. (a)-(c) Forecast for the geopotential height at 1000 hPa (this field illustrates the atmospheric state close to the surface) given by three forecasts started from very similar initial conditions, and (d) verifying analysis. Contour interval is 5 m, with only values smaller than 5 m shown.

The atmosphere exhibits this behavior. Fig. 1 shows three different weather forecasts, all started from very similar initial conditions. The differences among the three initial conditions were smaller than estimated analysis errors, and each of the three initial conditions could be considered as an equally probable estimate of the "true" initial state of the atmosphere. After 5 days of numerical integration, the three forecasts evolved into very different atmospheric situations. In particular, note the different positions of the cyclone forecast in the Eastern Atlantic approaching United Kingdom (Fig. 1 (a)-(c)). The first forecast indicated two areas of weak cyclonic circulation west and south of the British Isles; the second forecast positioned a more intense cyclone southwest of Cornwall, and the third forecast kept the cyclone in the open seas. This latter turned out to be the most accurate when compared to the observed atmospheric state (Fig. 1 (d). This is a typical example of orbits initially close together and then diverging during time evolution.

The atmosphere is an intricate dynamical system with many degrees of freedom. The state of the atmosphere is described by the spatial distribution of wind, temperature, and other weather variables (e.g. specific humidity and surface pressure). The mathematical differential equations describing the system time evolution include Newton's laws of motion used in the form 'acceleration equals force divided by mass', and the laws of thermodynamics which describe the behavior of temperature and the other weather variables. Thus, generally speaking, there is a set of differential equations that describe the weather evolution, at least, in an approximate form.

Richardson (1922) can be considered the first to have shown that the weather could be predicted numerically. In his work, he approximated the differential equations governing the atmospheric motions with a set of algebraic difference equations for the tendencies of various field variables at a finite number of grid points in space. By extrapolating the computed tendencies ahead in time, he could predict the field variables in the future. Unfortunately, his results were very poor, both because of deficient initial data, and because of serious problems in his approach.

After World War II the interest in numerical weather prediction revived, partly because of an expansion of the meteorological observation network, but also because of the development of digital computers. Charney (1947, 1948) developed a model applying an essential filtering approximation of the Richardson's equations, based on the so-called geostrophic and hydrostatic equations. In 1950, an electronic computer (ENIAC) was installed at Princeton University, and Charney, Fjørtoft and Von Neumann & Ritchmeyer (1950) made the first numerical prediction using the equivalent barotropic version of Charney's model. This model provided forecasts of the geopotential height near 500 hPa, and could be used as an aid to provide explicit predictions of other variables as surface pressure and temperature distributions. Charney's results led to the developments of more complex models of the atmospheric circulation, the so-called global circulation models.

With the introduction of powerful computers in meteorology, the meteorological community invested more time and efforts to develop more complex numerical models of the atmosphere. One of the most complex models used routinely for operational weather prediction is the one implemented at the European Centre for Medium-Range Weather Forecasts (ECMWF). At the time of writing (December 1999), its is based on a horizontal spectral triangular truncation T319 with 60 vertical levels formulation (Simmons *et al.* 1989, Courtier *et al.* 1991, Simmons *et al.* 1995). It includes a parameterization of many physical processes such as surface and boundary layer processes (Viterbo & Beljaars 1995) radiation (Morcrette 1990), and moist processes (Tiedtke 1993, Jacob 1994).

The starting point, in mathematical terms the initial conditions, of any numerical integration is given by very complex assimilation procedures that estimate the state of the atmosphere by considering all available observations. The fact that a limited number of observations are available (limited compared to the degrees of freedom of the system) and that part of the globe is characterized by a very poor coverage introduces uncertainties in the initial conditions. The presence of uncertainties in the initial conditions is the first source of forecast errors.

A requirement for skilful predictions is that numerical models are able to accurately simulate the dominant atmospheric phenomena. The fact that the description of some physical processes has only a certain degree of accuracy, and the fact that numerical models simulate only processes with certain spatial and temporal, is the second source of forecast errors. Computer resources contribute to limit the complexity and the resolution of numerical models and assimilation, since, to be useful, numerical predictions must be produced in a reasonable amount of time.

These two sources of forecast errors cause weather forecasts to deteriorate with forecast time.

Initial conditions will always be known approximately, since each item of data is characterized by an error that depends on the instrumental accuracy. In other words, small uncertainties related to the characteristics of the atmospheric observing system will always characterize the initial conditions. As a consequence, even if the system equations were well known, two initial states only slightly differing would depart one from the other very rapidly as time progresses (Lorenz 1965). Observational errors, usually in the smaller scales, amplify and through non-linear interactions spread to longer scales, eventually affecting the skill of these latter ones (Somerville 1979).

The error growth of the 10-day forecast of the ECMWF model from 1 December 1980 to 31 May 1994 was analyzed in great detail by Simmons *et al.* (1995). It was concluded that 15 years of research had improved substantially the accuracy over the first half of the forecast range (say up to forecast day 5), but that there had been little error reduction in the late forecast range. While this applied on average, it was also pointed out that there had been improvements in the skill of the good forecasts. In other words, good forecasts had higher skill in the nineties than before. The problem was that it was difficult to assess a-priori whether a forecast would be skilful or unskillful using only a deterministic approach to weather prediction.

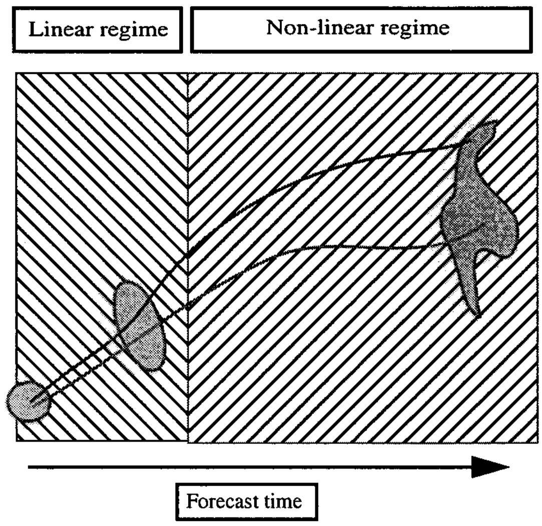

Figure 2. The deterministic approach to numerical weather prediction provides one single forecast (blue line) for the "true" time evolution of the system (red line). The ensemble approach to numerical weather prediction tries to estimate the probability density function of forecast states (magenta shapes). Ideally, the ensemble probability density function estimate includes the true state of the system as a possible solution.

Generally speaking, a complete description of the weather prediction problem can be stated in terms of the time evolution of an appropriate probability density function (PDF) in the atmosphere's phase space (Fig. 2). Although this problem can be formulated exactly through the continuity equation for probability (Liouville equation, see e.g. Ehrendorfer 1994), ensemble prediction based on a finite number of deterministic integrations appears to be the only feasible method to predict the PDF beyond the range of linear error growth. Ensemble prediction provided a way to overcome one of the problems highlighted by Simmons *et al.* (1995), since it can be used to estimate the forecast skill of a deterministic forecast, or, in other words, to forecast the forecast skill.

Since December 1992, both the US National Center for Environmental Predictions (NCEP, previously NMC) and ECMWF have integrated their deterministic high-resolution prediction with medium-range ensemble prediction (Tracton & Kalnay 1993, Palmer *et al.* 1993). These developments followed the theoretical and experimental work of, among others, Epstein (1969), Gleeson (1970), Fleming (1971a-b) and Leith (1974).

Both centres followed the same strategy of providing an ensemble of forecasts computed with the same model, one started with unperturbed initial conditions referred to as the "control" forecast and the others with initial conditions defined adding small perturbations to the control initial condition. Generally speaking, the two ensemble systems differ in the ensemble size, in the fact that at NCEP a combination of lagged forecasts is used, and in the definition of the perturbed initial. The reader is referred to Toth & Kalnay (1993) for the description of the 'breeding' method applied at NMC and to Buizza & Palmer (1995) for a thorough discussion of the singular vector approach followed at ECMWF.

A different methodology was developed few years later at the Atmospheric Environment Service (Canada), where a system simulation approach was followed to generate an ensemble of initial perturbations (Houtekamer *et al.* 1996). A number of parallel data assimilation cycles is run randomly perturbing the observations, and using different parameterisation schemes for some physical processes in each run. The ensemble of initial states generated by the different data assimilation cycles defines the initial conditions of the Canadian ensemble system. Moreover, forecasts started from such an ensemble of initial conditions are used to estimate forecast-error statistics (Evensen 1994, Houtekamer & Mitchell 1998).

Ensemble prediction, which can be considered one of the most recent advances in numerical weather prediction, is the first topic discussed in this work. The development of objective procedures to target adaptive observations is the second topic on which attention will be focused.

The idea of targeting adaptive observations is based on the fact that weather forecasting can be improved by adding extra observations only in sensitive regions. These sensitive regions can be identified using tangent forward and adjoint versions of numerical weather prediction models (Thorpe *et al.* 1998, Buizza & Montani 1999). Once the sensitive regions have been localised, instruments can be sent to those locations to take the required observations using pilot-less aircraft, or energy-intensive satellite instruments can be switched on to sample them with greater accuracy.

After this Introduction, section 2 describes some early results by Lorenz, and illustrates the chaotic behavior of a simple 3-dimension system. In section 3 the main steps of numerical weather prediction are delineated. The impact of initial condition and model uncertainties on numerical integration is discussed in section 4. The ECMWF Ensemble Prediction System is described in section 5. Targeting adaptive observations using singular vectors is discussed in section 6. Some conclusions are reported in section 7. Some mathematical details are reported in two Appendices.

2. THE LORENZ SYSTEM

One of the fathers of chaos theory is Edward Lorenz (1963, 1965). Results from the 3-dimentional Lorenz system

$$\begin{aligned}\dot{X} &= -\sigma X + \sigma Y \\ \dot{Y} &= -XY + rX - Y \\ \dot{Z} &= XY - bZ\end{aligned} \tag{1}$$

illustrate the dispersion of finite time integrations from an ensemble of initial conditions (Fig. 3). The different initial points can be considered as estimates of the "true" state of the system (which can be thought of as any point inside ellipsoid), and the time evolution of each of them as possible forecasts. Subject to the initial "true" state of the system, points close together at initial time diverge in time at different rates. Thus, depending on the point chosen to describe the system time evolution, different forecasts are obtained.

The two wings of the Lorenz attractor can be considered as identifying two different weather regimes, for example one warm and wet and the other cold and sunny. Suppose that the main purpose of the forecast is to predict whether the system is going through a regime transition. When the system is in a predictable initial state (Fig. 3 (a)), the rate of forecast diverge is small, and all the points stay close together till the final time. Whatever the point chosen to represent the initial state of the system, the forecast is characterised by a small error, and a correct indication of a regime transition is given. The ensemble of points can be used to generate probabilistic forecasts of regime transitions. In this case, since all points end in the other wing of the attractor, there is a 100% probability of regime transition.

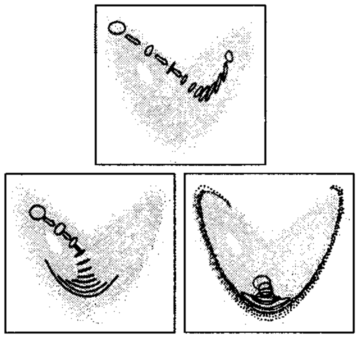

Figure 3. Lorenz attractor with superimposed finite-time ensemble integration.

By contrast, when the system is in a less predictable state (Fig. 3 (b)), the points stay close together only for a short time period, and then start diverging. While it is still possible to predict with a good degree of accuracy the future forecast state of the system for a short time period, it is difficult to predict whether the system will go through a regime transition in the long forecast range. Fig. 3 (c) shows an even worse scenario, with points diverging even after a short time period, and ending in very distant part of the system attractor. In probabilistic terms, one could have only predicted that there is a 50% chance of the system undergoing a regime transition. Moreover, the ensemble of points indicates that there is a greater uncertainty in predicting the region of the system attractor where the system will be at final time in the third case (Fig. 3 (c)).

The comparison of the points' divergence during the three cases indicates how ensemble prediction systems can be used to "forecast the forecast skill". In the case of the Lorenz system, a small divergence is associated to a predictable case, and confidence can be attached to any of the single deterministic forecasts given by the single points. By contrast, a large diverge indicate low predictability.

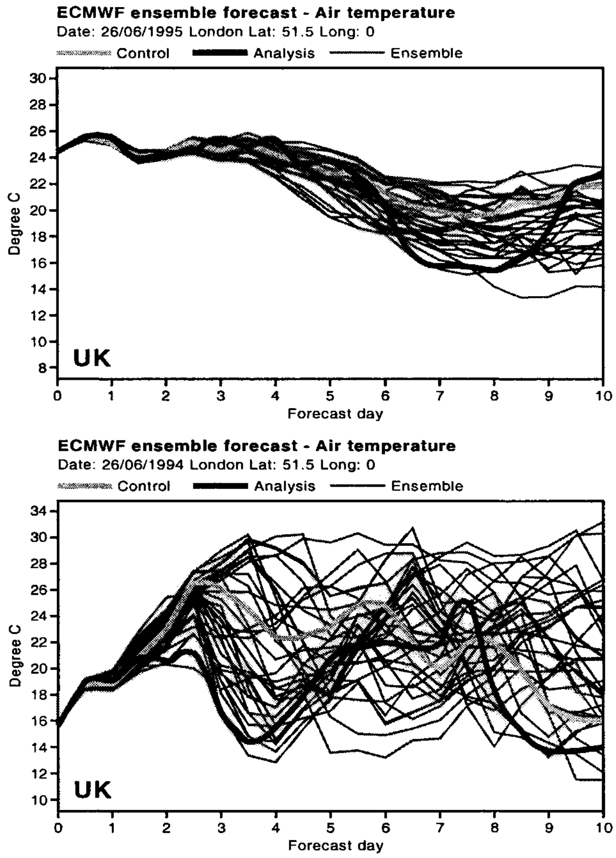

Figure 4. ECMWF forecasts for air temperature in London started from (a) 26 June 1995 and (b) 26 June 1994.

Similar sensitivity to the initial state is shown in weather prediction. Fig. 4 shows the forecasts for air temperature in London given by 33 different forecasts started from very similar initial conditions for two different dates, the 26th of June of 1995 and the 26th of June 1994. There is a clear different degree of divergence during the two cases. All forecasts stay close together up to forecast day 10 for the first case (Fig. 4 (a)), while they all diverge already at forecast day 3 in the second case (Fig. 4 (b)). The level of spread among the different forecasts can be used as a measure of the predictability of the two atmospheric states.

3. Numerical weather prediction

Numerical weather prediction is realised by integrating primitive-equation models. The equations are solved by replacing time-derivatives by finite differences, and spatially either by finite difference schemes or spectral methods. The state of the atmosphere is described at a series of grid-points by a set of state variables such as temperature, velocity, humidity and pressure.

At the time of writing (December 1999) the ECMWF high-resolution deterministic model has a spectral triangular truncation TL319, which is equivalent to a grid-point spacing of about 60 km at mid-latitudes (Fig. 5) and 60 vertical levels (Fig. 6).

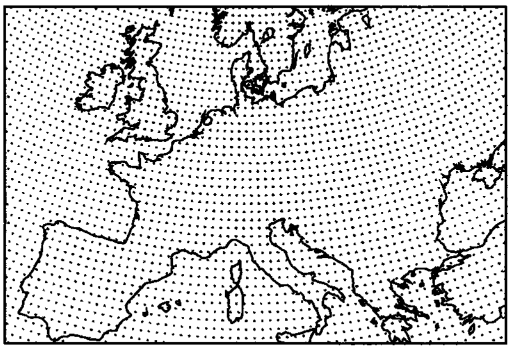

Figure 5. Grid-points over Europe of the ECMWF model.

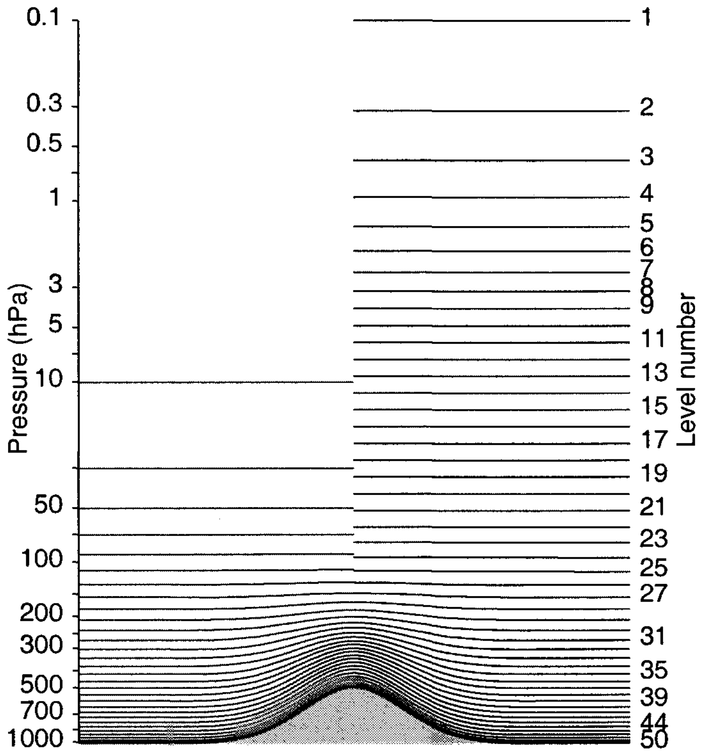

Figure 6. Vertical levels of the ECMWF model.

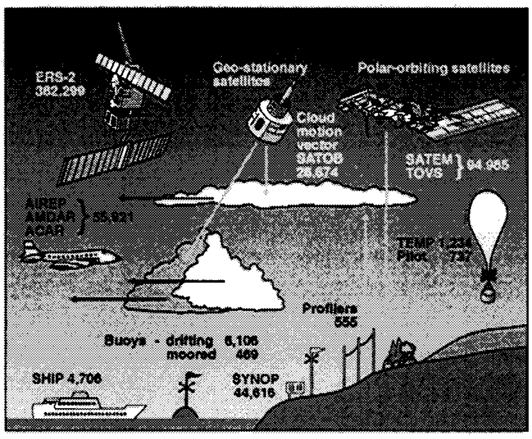

Figure 7. Type and number of observations used to estimate the atmosphere initial conditions in a typical day.

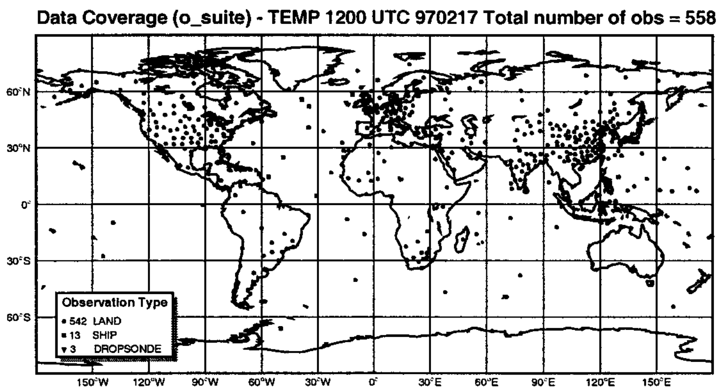

Figure 8. Map of radiosonde locations.

Meteorological observations made all over the world (Fig. 7) are used to compute the best estimate of the system initial conditions. Some of these observations, such as the ones from weather balloons or radiosondes, are taken at specific times at fixed locations (Fig. 8). Other data, such as the ones from aircrafts, ships or satellites, are not fixed in space. Generally speaking, there is a great variability in the density of the observation network. Data over oceanic regions, in particular, are characterised by very coarse resolution.

Observations cannot be used directly to start model integration, but must be modified in a dynamically consistent way to obtain a suitable data set. This process is usually referred to as data assimilation. At the time of writing (December 1999), ECMWF uses a 4-dimensional data assimilation scheme to estimate the actual state of the atmosphere (Courtier *et al.* 1994).

In the ECMWF model, dynamical quantities as pressure and velocity gradients are evaluated in spectral space, while computations involving processes such as radiation, moisture conversion, turbulence, are calculated in grid-point space. This combination preserves the local nature of physical processes, and retains the superior accuracy of the spectral method for dynamical computation.

The physical processes associated with radiative transfer, turbulent mixing, moist processes, are active at scales smaller than the horizontal grid size. The approximation of unresolved processes in terms of model-resolved variables is referred to as parameterisation (Fig. 9). The parameterisation of physical processes is probably one of the most difficult and controversial area of weather modelling (Holton 1992).

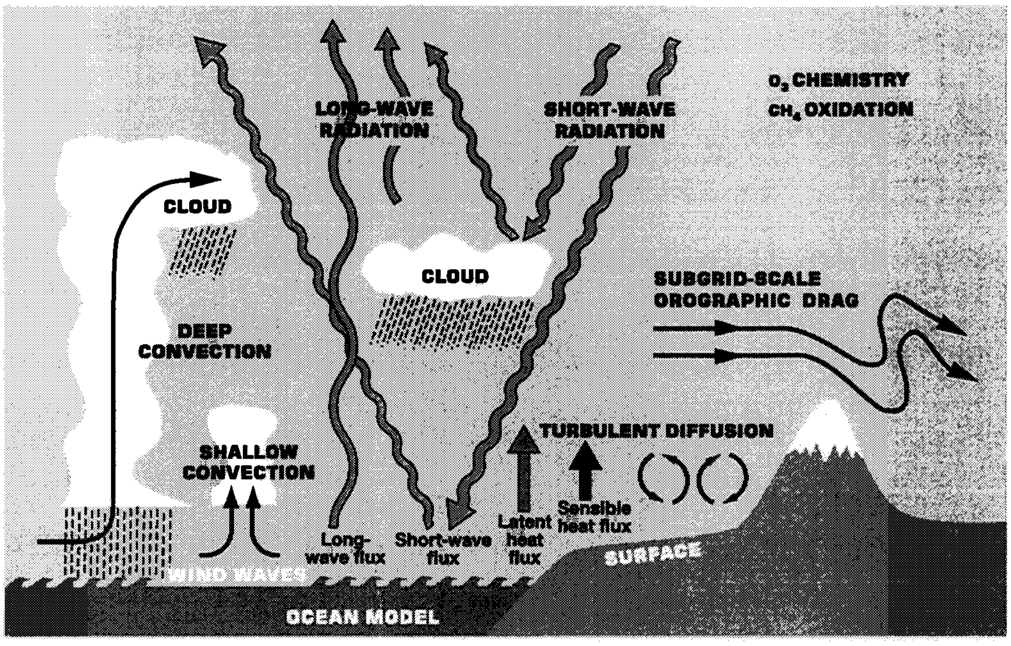

Figure 9. Schematic diagram of the different physical processes represented in the ECMWF model.

4. Sources of forecast error

It has been already mentioned that uncertainties in the initial conditions and in the model are both sources of forecast error. Some indications of the relative importance of the two sources can be deduced from the works of Downton & Bell (1988) and Richardson (1998), who compared forecasts given by the UKMO (United Kingdom Meteorological Office) and the ECMWF forecasting systems. In these studies, substantial forecast differences between the ECMWF and the UKMO operational forecasts could mostly be traced to differences between the two operational analyses, rather than between the two forecast models. On the other hand, recent results from Harrison *et al.* (1999) indicate that the impact of model uncertainties on forecast error cannot be ignored.

These results suggest that an ensemble system should certainly simulate the presence of uncertainties in the initial conditions, since this is the dominant effect, but it should also simulate model uncertainties.

The first version of the ECMWF Ensemble Prediction System (hereafter EPS, Palmer *et al.* 1993, Molteni *et al.* 1996) implemented operationally in December 1992 included only a simulation of initial uncertainties. A similar "perfect model" strategy was followed at the US National Centers for Environmental Prediction (NCEP, Tracton & Kalnay 1993).

Houtekamer *et al.* (1996) first included model uncertainties in the ensemble prediction system developed at the Atmospheric Environment Service in Canada. Following a system simulation approach to ensemble prediction, they developed a procedure where each ensemble member differs both in the initial conditions, and in sub-grid scale parameters. In this approach, each ensemble member is integrated using different parameterizations of horizontal diffusion, convection, radiation, gravity wave drag, and with different orography.

There are certainly good grounds for believing that there is a significant source of random error associated with the parameterized physical processes. For example, consider a grid point over the tropical warm pool area during a period of organized deep convection. By definition, the actual contributions to the tendencies due to parameterized physical processes are often associated with organized mesoscale convective systems whose spatial extent may be comparable with the model resolution. In such a case, the notion of a quasi-equilibrium ensemble of sub-grid-scale processes, upon which all current parameterizations schemes are based, cannot be a fully-appropriate concept for representing the actual parameterized heating (Palmer 1997). For example, even if the parameterized heating fields agree on average (i.e. over many time steps) at the chosen grid point, there must inevitably be some standard deviation in the time-step by time-step difference between observed and modeled heating.

Since October 1998, a simple stochastic scheme for simulating random model errors due to parameterized physical processes has been used in the ECMWF EPS (Buizza *et al.* 1999). The scheme is based on the notion that the sort of random error in parameterized forcing are coherent between the different parameterization modules, and have certain coherence on the space and time scales associated, for example, with organized convection schemes. Moreover, the scheme assumes that the larger the parameterized tendencies, the larger the random error component will be. The notion of coherence between modules allows the stochastic perturbation to be based on the total tendency from all parameterized processes, rather than on the parameterized tendencies from each of the individual modules. In this respect the ECMWF scheme differs conceptually from that of Houtekamer *et al.* (1996). More details about the scheme are reported in the following section.

5. THE ECMWF ENSEMBLE PREDICTION SYSTEM

Routine real-time execution of the ECMWF EPS started in December 1992 with a 31-member T63L19 configuration (spectral triangular truncation T63 and 19 vertical levels, Palmer *et al.* 1993, Molteni *et al.* 1996). A major upgrade to a 51-member TL159L31 system (spectral triangular truncation T159 with linear grid) took place in 1996 (Buizza *et al.* 1998). A scheme to simulate model uncertainties due to random model error in the parameterized physical processes was introduced in 1998.

5.1 The original EPS configuration

Schematically, each ensemble member e_j was defined by the time integration

$$e_j(t) = \int_{t=0}^{t} [A(e_j, t) + P(e_j, t)]\, \mathrm{d}t \tag{2}$$

of the model equations

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P(e_j, t) \tag{3}$$

starting from perturbed initial conditions

$$e_j(t=0) = e_0(t=0) + \delta e_j(t=0) \tag{4}$$

where A and P identify the contribution to the full equation tendency of the non-parameterized and parameterized physical processes, and where $e_0(t=0)$ is the operational analysis at $t=0$.

The initial perturbations $\delta e_j(t=0)$ were generated using the singular vectors of the linear version of the ECMWF, computed to maximize the total energy norm over a 48-hour time interval (Buizza & Palmer 1995), and scaled to have an amplitude comparable to analysis error estimates.

The singular vectors of the tangent forward propagator sample the phase space directions of maximum growth during a 48-hour time interval. Small errors in the initial conditions along these directions would amplify most rapidly, and affect the forecast accuracy. The reader is referred to Appendix A for a more complete mathematical definition of the singular vectors.

Fig. 10 illustrates the typical structure of the leading singular vector used to generate the ensemble of initial perturbations for 17 January 1997. Total energy singular vectors are usually located in the lower troposphere at initial time, with total energy peaking at between around 700 hPa (i.e. around 3000 m), in regions of strong barotropic and baroclinic energy conversion (Buizza & Palmer 1995). During their growth, they show an upscale energy transfer and upward energy propagation. Results have indicated a very good agreement between the regions where singular vectors are located and other measures of baroclinic instability such as the Eady index introduced by Hoskins & Valdes (1990). This is shown in Fig. 11 for the case of 17 January 1997. The reader is referred to Hoskins *et al.* (1999) for recent investigations on singular vector growth mechanisms.

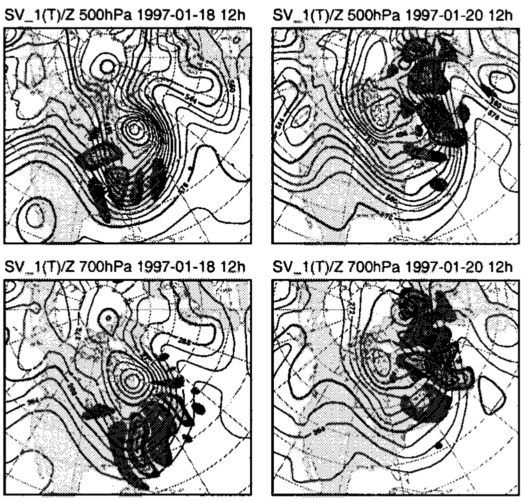

Figure 10. Most unstable singular vector growing between 17 and 19 January 1997 at initial (left panels) and final (right panels) times. The top panels show the singular vector temperature component (shaded blue/green for negative and shaded yellow/red for positive values) and the atmospheric state (geopotential height) at 500 hPa (i.e. approximately at 5000 m). Bottom panels are as top panels but for 700 hPa (i.e. approximately at 3000 m). Contour interval is 8 dam for geopotential height, and 0.2 degrees for temperature at initial time and 1.0 degree at final time.

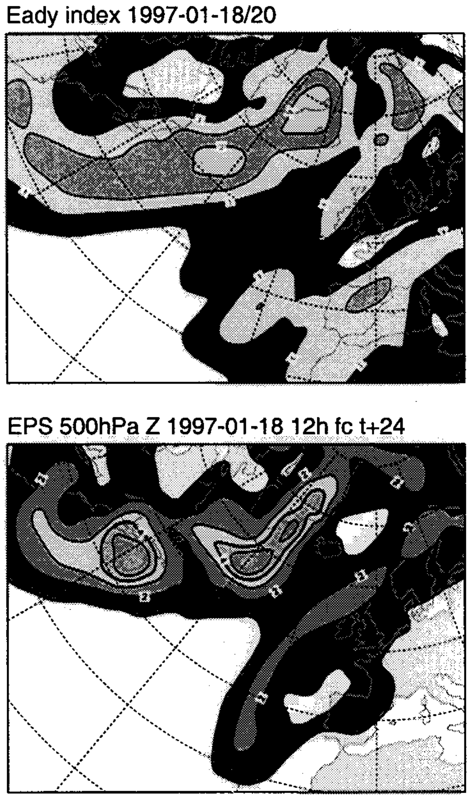

Figure 11. Eady index (top panel) and location of the leading singular vectors identified by the root-mean-square amplitude of the EPS perturbations at forecast day 1 (bottom panel) for the case of 17 January 1997. Contour interval is 1 day-1 for the Eady index, and 1 m for the EPS perturbation amplitude.

Until 26 March 1998, the EPS initial perturbations were computed to sample instabilities growing in the forecast range, and no account was taken of perturbations that had grown during the data assimilation cycle leading up to the generation of the initial conditions (Molteni *et al.* 1996).

A way to overcome this problem is to use singular vectors growing in the past, and evolved to the current time.

The 50 perturbed initial conditions were generated by adding and subtracting 25 perturbations defined using 25 singular vectors selected from computed singular vectors so that they do not overlap in space. The selection criteria were that the leading 4 singular vectors are always selected, and that subsequent singular vectors are selected only if less than 50% of their total energy cover a geographical region where already 4 singular vectors are located.

Once the 25 singular vectors were selected, an orthogonal rotation in phase-space and a final re-scaling were performed to construct the ensemble perturbations. The purpose of the phase-space rotation is to generate perturbations with the same globally averaged energy as the singular vectors, but smaller local maxima and a more uniform spatial distribution. Moreover, unlike the singular vectors, the rotated singular vectors are characterized by similar amplification rates (at least up to 48 hours). Thus, the rotated singular vectors diverge, on average, equally from the control forecast. The rotation is defined to minimize the local ratio between the perturbation amplitude and the amplitude of the analysis error estimate given by the ECMWF data assimilation procedure. The re-scaling allowed perturbations to have local maxima up to $\alpha = \sqrt{0.6}$ larger than the local maxima of the analysis error estimate.

A way to take into account perturbations growing during the data assimilation period was to generate the EPS initial perturbation using two sets of singular vectors. In mathematical terms, since 26 March 1998 (Barkmeijer *et al.* 1999a) the day d initial perturbations have been generated using both the singular vectors growing in the forecast range between day d and day $d+2$ at initial time, and the singular vectors that had grown in the past between day $d-2$ and day d at final time

$$\delta e_j(t=0) = \sum_{i=1}^{25} [\alpha_{i,j} v_i^{d,d+2}(t=0) + \beta_{i,j} v_i^{d-2,d}(t=48\,\mathrm{h})] \tag{5}$$

where $v_i^{d,d+2}(t=0)$ is the i-th singular vector growing between day d and d+2 at time t=0. The coefficients ai,j and bi,j set the initial amplitude of the ensemble perturbations, and are defined by comparing the singular vectors with estimates of analysis errors (Molteni *et al.* 1996). Since 26 March 1998, the selection criteria has been kept as before, but a new scaling factor $\alpha = \sqrt{0.5}$ has been used.

The initial perturbations are specified in terms of the spectral coefficients of the 3-dimensional vorticity, divergence and temperature fields (no perturbations are defined for the specific humidity since the singular vector computation is performed with a dry linear forward/adjoint model), and of the 2-dimensional surface pressure field. They are added and subtracted to the control initial conditions to define perturbed initial conditions. Then, 50 + 1 (control) 10-day $T_L159L31$ non-linear integrations are performed.

With the current ECMWF computer facilities, each day the whole EPS (10-day integration, 51 members at $T_L159L40$ with T42L40 singular vectors computed for the Northern and the Southern Hemispheres) takes approximately 150 hours of total computing time (about 10% of this time is used to compute the initial perturbations). By contrast, the high resolution $T_L319L60$ deterministic forecast (10-day integration) takes about 40 hours, and the data assimilation procedure used to generate the unperturbed initial conditions takes about 120 hours (4 cycles per day, once every 6 hours).

5.2 The new EPS configuration

In October 1998, a scheme to simulate random model errors due to parameterized physical processed was introduced (Buizza *et al.* 1999). This scheme can be considered as a simple first attempt to simulated random model errors due to parametrized physical processes. It is based on the notion that random errors due to parametrized physical processes are coherent between the different parametrization modules and have a certain coherence on the space and time scales represented by the model. The scheme assumes that the larger the parametrized tendencies, the larger the random error component.

In the new EPS, each ensemble member e_j can be seen as the time integration

$$e_j(t) = \int_{t=0}^{t} [A(e_j, t) + P_j'(e_j, t)]\, dt \tag{6}$$

of the perturbed model equations

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P_j'(e_j, t) \tag{7}$$

starting from the perturbed initial conditions defined in Eq. (1), where A and P' identify the contribution to the full equation tendency of the non-parameterized and parameterized physical processes. For each grid point $x = (\lambda, \phi, \sigma)$ (identified by its latitude, longitude and vertical hybrid coordinate), the perturbed parameterized tendency (of each state vector component) is defined as

$$P_j'(e_j, t) = [1 + \langle r_j(\lambda, \phi, \sigma) \rangle_{D,T}] P(e_j, t) \tag{8}$$

where P is the unperturbed diabatic tendency, and $\langle \ldots \rangle_{D,T}$ indicates that the same random number r_j has been used for all grid points inside a $D \times D$ degree box and over T time steps.

The notion of space-time coherence assumes that organized systems have some intrinsic space and time-scales that may span more than one model time step and more than one model grid point. Making the stochastic uncertainty proportional to the tendency is based on the concept that organization (away from the notion of a quasi-equilibrium ensemble of sub-grid processes) is likely to be stronger, the stronger is the parameterized contribution. A certain space-time correlation is introduced in order to have tendency perturbations with the same spatial and time scales as observed organization.

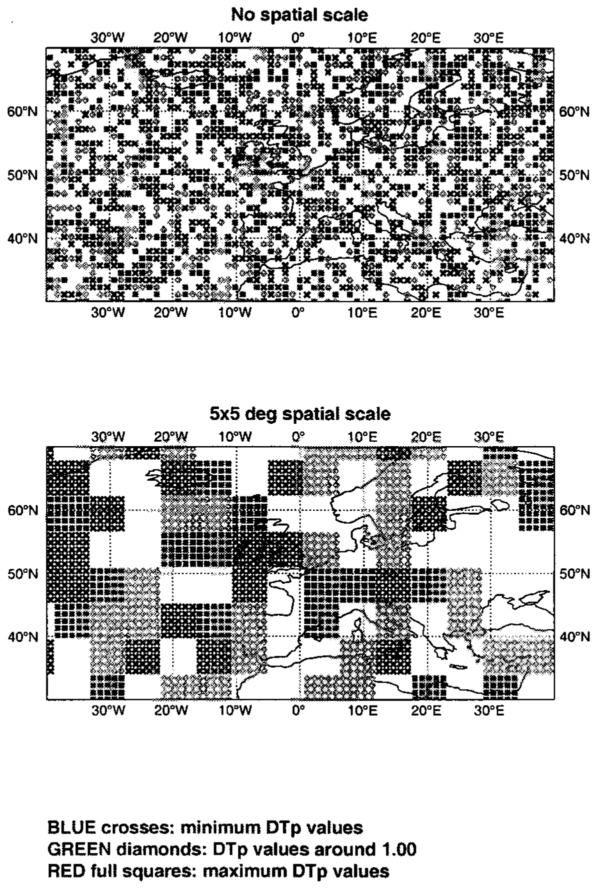

Figure 12. Random numbers used to perturb the tendencies due to parameterised physical processes. The top panel shows the case of no spatial scale, in other words when different random numbers are used at each grid-point. The bottom panel shows the case when the same random number was used for grid-points inside 5-degree boxes. Blue crosses identify grid-points with random numbers $-0.5 \le \eta \le -0.3$, green diamonds points with $-0.1 \le \eta \le 0.1$, and red squares points with $0.3 \le \eta \le 0.5$.

Fig. 12 shows a map of the random numbers used in a configuration tested when developing the so-called stochastic physics scheme. Fig. 12 (a) shows the matrix of random numbers r_j when each grid-point was assigned an independent value, while Fig. 12 (b) shows the matrix when the same random number was used inside 5 degree boxes. Results indicated that even perturbations without any spatial structure (i.e. with random numbers as in Fig. 12 (a)) had a major impact on 10-day model integrations (Buizza *et al.* 1999).

After an extensive experimentation, it was decided to implement in the operational EPS the stochastic physics scheme with random numbers sampled uniformly in the interval $[-0.5, 0.5]$, a 10 degrees box size ($D = 10$), and a 6 hours time interval ($T = 6$).

6. TARGETED OBSERVATIONS

Consider a meteorological system evolving between time t_0 and t, localised at final time t inside a geographical area Σ_t (hereafter verification area). Suppose that extra observations could be taken inside a geographical area Σ_0 at initial time t_0 (hereafter target area), with the purpose of improving the time t forecast inside Σ.

Singular vectors with maximum energy at final time inside a verification area can be used to identify the target area where extra observation should be taken, at initial time, to reduce the forecast error inside the verification area itself (Buizza & Montani 1999). The reader is referred to Appendices A and B for a more complete description of the mathematical formulation.

Other strategies can be used to target adaptive observations. Langland & Rohaly (1996), following the work of Rabier *et al.* (1996) on sensitivity vectors, proposed to use the lower tropospheric vorticity of the forecast state as cost function, and to target the region where the sensitivity field is maximum. A similar technique, but based on the use of a quasi-inverse linear model, was proposed by Pu *et al.* (1997, 1998). Bishop & Toth (1998) introduced the Ensemble Transform technique, in which linear combinations of ensemble perturbations are used to estimate the prediction error variance associated with different possible deployments of observational resources. Finally, following Hoskins *et al.* (1985) and Appenzeller *et al.* (1996), a more subjective strategy based on the use of potential vorticity to analyse atmospheric was also developed.

All these techniques were applied to target observations for the first time during FASTEX, the Fronts and Atlantic Storm Track Experiment (Joly *et al.* 1996, Thorpe & Shapiro 1995, Snyder 1996).

The focus of the FASTEX campaign was the extra-tropical cylonic storms that form over the western and mid Atlantic Ocean, and take about 2 days to develop and move towards Europe. Forecast failures are often associated with these very active atmospheric phenomena.

Fig. 13 shows the tracks of one of the storms observed during FASTEX, IOP 17 (IOP stands for Intensive Observation Period), and the location of various aircrafts that made additional observations between 17 and 19 February 1997 (Montani *et al.* 1999). Singular vectors, computed to have maximum total energy inside a verification region centred on the British Isles, were used to identify the most sensitive regions where observations were made. The comparison of the central pressure of two forecasts, one started from initial conditions computed with and one without the extra observations, with the observed value (Fig. 14) indicates that additional, targeted observations can improve the forecast accuracy. Fig. 15 shows the average impact of the targeted observations on the forecast error. Results indicated up to 20% forecast error reduction.

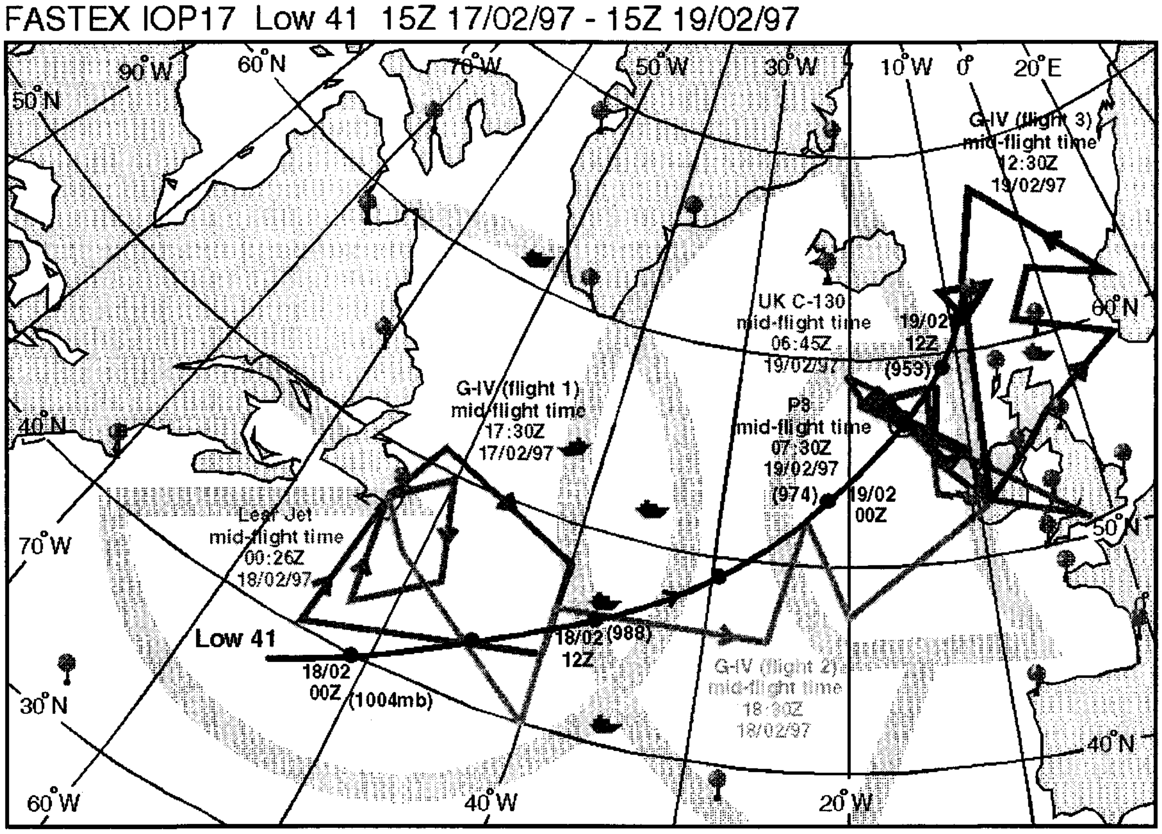

Figure 13. Map summarising the extra observations taken during the FASTEX experiment for IOP 17. The black track identifies the location of the cyclone minimum pressure, the coloured tracks the aircraft missions, and the red symbols additional radio-soundings (from Montani *et al.* 1999).

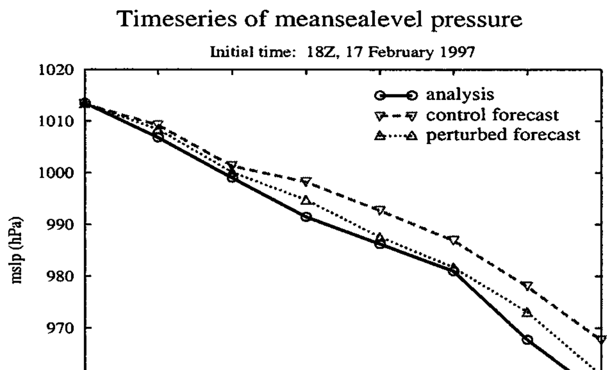

Figure 14. 6-hourly time series of the cyclone central pressure forecast without (blue dash line) and with (red dotted line) extra observations, and observed (black solid line), for IOP 17 (from Montani *et al.* 1999).

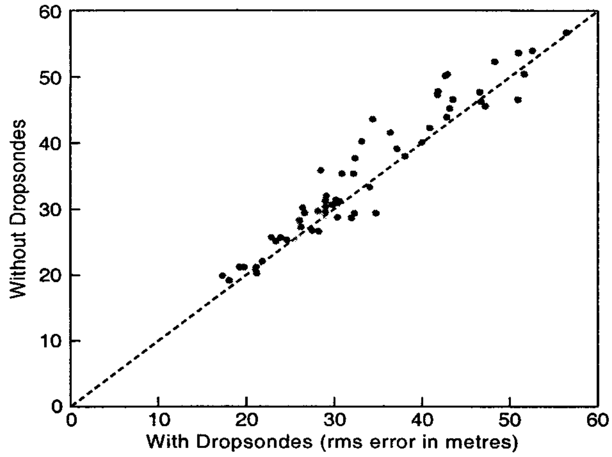

Figure 15. Scatter plot of the mean forecast errors with (x-axis) and without (y-axis) the extra observations, for the 500 and 1000 hPa geopotential height fields over Europe and North Atlantic at forecast day 2, 2.5 and 3 (from Montani *et al.* 1999).

Following the FASTEX campaign, other experiments were performed (e.g. NORPEX, the North Pacific Experiment, CALJET, the California Land-falling Jets experiments). All results confirmed that taking extra observations in sensitive regions could reduce forecast errors.

Robotic aerosondes capable of long-range monitoring could be used operationally in a very near future to fill chronic gaps in the global upper-air sounding network (Holland *et al.* 1992), and take extra observations in objectively identified regions. This follows years of intensive research at the Bureau of Meteorology of Melbourne, Australia, that culminated with the first-ever unmanned aircraft crossing of the Atlantic Ocean in August 1998 (http://www.aerosonde.com/opshist.htm).

7. SUMMARY AND FUTURE DEVELOPMENTS

Two of the most important advances in numerical weather prediction of the last 10 years, the operational implementation of ensemble prediction systems and the development of objective techniques to target adaptive observations, have been discussed.

Ensemble systems provide a possible way to estimate the probability distribution function of forecast states. They have been developed following the notion that uncertainties in the initial conditions and in the model formulation are the main sources of forecast errors. Results have demonstrated that a probabilistic approach to weather prediction can provide more information than a deterministic approach based on a single, deterministic forecast.

Ensemble prediction systems are particularly useful, if not necessary, to provide early warnings of extreme weather events. For example, ensemble systems can be used to predict probabilities of intense precipitation events (Fig. 16). Global ensemble systems can be used to provide boundary and initial conditions for higher-resolution, limited area ensemble prediction systems (Molteni *et al.* 1999, Marsigli *et al.* 1999).

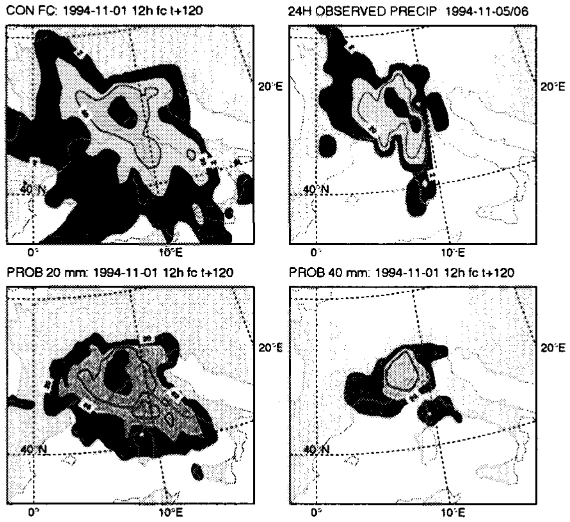

Figure 16. High-resolution TL319L31 forecasts for the flood in Piemonte, Italy, 5-6 November 1994. (a) Precipitation forecast, accumulated between t+96 h and t+120 h, predicted by the control forecast. (b) Observed precipitation field. (c) Ensemble probability forecast of more than 20 mm/d of precipitation. (d) Ensemble probability forecast of more than 40 mm/d of precipitation. Contour isolines 2, 20, 40 and 100 mm/d for precipitation, and every 20% for probabilities.

At ECMWF, work is in progress in many different areas to further improve the current ensemble prediction system.

Linearized versions of the most important physical processes have been developed (Mahfouf 1999), and investigation into the behaviour of the linear models in the computation of tropical singular vectors has started. The tropical target area has been chosen because the current EPS lacks perturbations of the initial condition in this area, where moist processes are of key importance. Results (Barkmeijer *et al.* 1999b, Puri *et al.* 1999) indicate that the inclusion of tropical singular vectors is essential in cases of hurricane prediction. Results indicate that tropical singular vectors are needed to generate a realistic spread among the ensemble of hurricane tracks.

The operational initial perturbations of the ECMWF EPS are constructed using singular vectors with maximum total energy growth. Total energy singular vectors have no knowledge of analysis error statistics. Generally speaking, it would be desirable to use information about analysis error characteristics in the singular vector computation. One way of improving upon this is to use in the singular vector computation statistics generated by the data assimilation system. Work is in progress to use the Hessian of the cost function of the 3-dimensional (or 4-dimensional) variational assimilation system (3D/4D-Var) to define singular vectors (Barkmeijer *et al.* 1998). These so-called Hessian singular vectors are constrained at initial time by analysis error statistics but still produce fast perturbation growth during the first few days of the forecast.

Work is in progress to investigate whether a so-called consensus analysis, defined as the average of analyses produced by different weather centres, is a better estimate of the atmospheric initial state than the ECMWF analysis (Richardson, 1999, personal communication). The operational EPS configuration has been run from the consensus analysis, average of the ECMWF, UKMO (UK Meteorological Office), Météo-France, NCEP (National Centers for Environmental Prediction, Washington) and DWD (Deutscher WetterDienst, Offenbach) analyses. The same perturbations as used in the operational EPS have been added to the consensus analysis to create the 50 perturbed initial conditions. Preliminary results show that the skill of the control forecast is improved if the consensus analysis is used instead of the ECMWF analysis as the unperturbed initial condition. Results also indicate that the difference between the spread in the two systems is rather small, while the ensemble-mean forecast of the system started from the consensus analysis is more skilful.

Experimentation has started to verify whether a resolution increase from Tl159 to Tl255, and from 31 to 60 vertical levels would improve the EPS performance. Preliminary studies of the impact of an increase in horizontal resolution indicate that precipitation prediction improves substantially as resolution increases.

Finally, an ensemble approach to data assimilation is under test. Following Houtekamer *et al.* (1996), but with the ECMWF approach to represent model uncertainties, work has started at ECMWF to generate an ensemble of initial perturbations using the ECMWF 3D/4D-Var data assimilation. The purpose of this work is to investigate whether a better estimate of the "true" state of the atmosphere can be computed using this probabilistic approach.

8. CONCLUSION

The weather is a chaotic system, and numerical weather prediction is a very difficult task.

This work has demonstrated that the application of linear algebra (i.e. the use of singular vectors computed by solving an eigenvalue problem defined by the tangent forward and adjoint versions of the model) to meteorology can help in designing new ways to numerical weather prediction (Buizza 1997).

The same technique can be applied to any dynamical system, in particular to very complex systems with a large dimension. The basic idea is that there are only few, important directions of the phase-space of any system along which the most important processes occur. A successful prediction of the system time evolution should sample these directions, and describe the system evolution along them.

9. ACKNOWLEDGEMENTS

The ECMWF Ensemble Prediction System is the result of the work of many ECMWF staff members and consultants. It is based on the Integrated Forecasting System/Arpege software, developed in collaboration by ECMWF and Meteo-France. The work of many ECMWF and Meteo-France staff and consultants is acknowledged. I am grateful to Robert Hine for all his editorial help.

APPENDIX A SINGULAR VECTOR DEFINITION

Farrell (1982), studying the growth of perturbations in baroclinic flows, showed that, although the long time asymptotic behavior is dominated by discrete exponentially growing normal modes when they exist, physically realistic perturbations could present, for some finite time intervals, amplification rates greater than the most unstable normal mode amplification rate. Subsequently, Farrell (1988, 1989) showed that perturbations with the fastest growth over a finite time interval could be identified solving the eigenvalue problem of the product of the tangent forward and adjoint model propagators. His results supported earlier conclusions by Lorenz (1965) that perturbation growth in realistic models is related to the eigenvalues of the operator product.

Kontarev (1980) and Hall and Cacuci (1983) first used the adjoint of a dynamical model for sensitivity studies. Later on, Le Dimet & Talagrand (1986) proposed an algorithm, based on an appropriate use of an adjoint dynamical equation, for solving constraint minimization problems in the context of analysis and assimilation of meteorological observations. More recently, Lacarra & Talagrand (1988) applied the adjoint technique to determine optimal perturbations using a simple numerical model. Following Urban (1985) they used a Lanczos algorithm (Strang, 1986) in order to solve the related eigenvalue problem. For a bibliography in chronological order of published works in meteorology dealing with adjoints up to the end of 1992, the reader is referred to Courtier *et al.* (1993).

After Farrell and Lorenz, calculations of perturbations growing over finite-time intervals were performed, for example, by Borges & Hartmann (1992) using a barotropic model, and by Molteni & Palmer (1993) using a barotropic and a 3-level quasi-geostrophic model at spectral triangular truncation T21. Buizza (1992) and Buizza *et al.* (1993) first identified singular vectors in a primitive equation model with a large number of degrees of freedom.

Let χ be the state vector of a generic autonomous system, whose evolution equations can be formally written as

$$\frac{\partial\chi}{\partial t} = A(\chi) \tag{9}$$

Denote by $\chi(t)$ an integration of Eq. (9) from t_0 to t, which generates a trajectory from an initial point χ_0 to $\chi_1 = \chi(t)$. The time evolution of a small perturbation x around the time evolving trajectory $\chi(t)$ can be descibed, in a first approximation, by the linearized model equations

$$\frac{\partial x}{\partial t} = A_l x \tag{10}$$

where $A_l = \left.\frac{\partial A(x)}{\partial x}\right|_{\chi(t)}$ is the tangent operator computed at the trajectory point $\chi(t)$.
Let $L(t, t_0)$ be the integral forward propagator of the dynamical equations linearized about a non-linear trajectory $\chi(t)$

$$x(t) = L(t, t_0)x(t_0) \tag{11}$$

that maps a perturbation x at initial time t_0 to the optimization time t. The tangent forward operator L maps the tangent space Π_0, the linear vector space of perturbations at χ_0, to Π_1, the linear vector space at χ_1.

Consider two perturbations x and y, e.g. at χ_0, a positive definite Hermitian matrix E, and define the inner product $(\ldots;\ldots)_E$ as

$$(x;y)_E = \langle x;Ey\rangle \tag{12}$$

on the tangent space Π_0 in this case, where $\langle\ldots;\ldots\rangle$ identifies the canonical Euclidean scalar product,

$$\langle x;y\rangle = \sum_{i=1}^{N} x_i y_i \tag{13}$$

Let $\|\ldots\|_E^2$ be the norm associated with the inner product $(\ldots;\ldots)_E$

$$\|x\|_E^2 = (x;x)_E = \langle x;xE\rangle \tag{14}$$

Let L^{*E} be the adjoint of L with respect to the inner product $(\ldots;\ldots)_E$,

$$(L^{*E}x;y)_E = (x;Ly)_E \tag{15}$$

The adjoint of L with respect to the inner product defined by E can be written in terms of the adjoint L^* defined with respect to the canonical Euclidean scalar product,

$$L^{*E} = E^{-1}L^{*}E \tag{16}$$

From Eqs. (11) and (16) it follows that the squared norm of a perturbation x at time t is given by

$$\|x(t)\|_E^2 = (x(t_0);L^{*E}Lx(t_0))_E \tag{17}$$

Equation (17) shows that the problem of finding the phase space directions x for which $\|x(t)\|_E^2/\|x(t_0)\|_E^2$ is maximum can be reduced to the search of the eigenvectors $v_i(t_0)$

$$L^{*E}Lv_i(t_0) = \sigma_i^2 v_i(t_0) \tag{18}$$

with the largest eigenvalues σ_i^2.

The square roots of the eigenvalues, σ_i, are called the singular values and the eigenvectors $v_i(t_0)$ the (right) singular vectors of L with respect to the inner product E (see, e.g., Noble & Daniel, 1977). The singular vectors with largest singular values identify the directions characterized by maximum growth. The time interval $t - t_0$ is called optimization time interval.

Unlike L itself, the operator $L^{*E}L$ is normal. Hence, its eigenvectors $v_i(t_0)$ can be chosen to form a complete orthonormal basis in the Nth -dimensional tangent space of the perturbations at χ_0. Moreover, the eigenvalues are real, $\sigma_i^2 \geq 0$.

At optimization time t, the singular vectors evolve to

$$v_i(t) = L(t, t_0)v_i(t_0) \tag{19}$$

which in turn satisfy the eigenvector equation

$$LL^{*E}v_i(t) = \sigma_i^2 v_i(t_0) \tag{20}$$

From Eqs. (17) and (20) it follows that

$$\|v_i(t)\|_E^2 = \sigma_i^2 \tag{21}$$

Since any perturbation $x(t) / \|x(t_0)\|_E$ can be written as a linear combination of the singular vectors v_i, it follows that

$$\max_{\|x(t_0)\|_E}\left(\frac{\|x(t)\|_E}{\|x(t_0)\|_E}\right) = \sigma_1 \tag{22}$$

Thus, maximum growth as measured by the norm $\|\ldots\|_E$ is associated with the dominant singular vector v_1 .

Given the tangent forward propagator L, it is evident from Eq. (18) that singular vectors' characteristics depend strongly on the inner product definition and to the specification of the optimization time interval.

The problem can be generalized by selecting a different inner product at initial and optimization time. Consider two inner products defined by the (positive definite Hermitian) matrices E_0 and E, and re-state the problem as finding the phase space directions x for which

$$\frac{\|x(t)\|_E}{\|x(t_0)\|_E} = \frac{\langle Lx(t_0);ELx(t_0)\rangle}{\langle x(t_0);E_0x(t_0)\rangle} \tag{23}$$

is maximum. Applying the transformation $y = E_0^{1/2}$, the right hand side ofEq. (18) can be transformed into

$$\frac{\langle LE_0^{-1/2}y(t_0);ELE_0^{-1/2}y(t_0)\rangle}{\langle y(t_0);y(t_0)\rangle} = \frac{\langle y(t_0);E_0^{-1/2}L^*ELE_0^{-1/2}y(t_0)\rangle}{\langle y(t_0);y(t_0)\rangle} \tag{24}$$

Since

$$E_0^{-1/2}L^*ELE_0^{-1/2} = (E^{-1/2}LE_0^{-1/2})^*(E^{-1/2}LE_0^{-1/2}) \tag{25}$$

the phase space directions which maximize the ratio in Eq. (24) are the singular vectors of the operator $E^{-1/2}LE_0^{-1/2}$ with respect to the canonical Euclidean inner product. With this definition, the dependence of the singular vectors' characteristics on the inner products is made explicit.

At ECMWF, due to the very large dimension of the system, the eigenvalue problem that defines the singular vectors is solved by applying a Lanczos code (Glub & Van Loan 1983).

APPENDIX B PROJECTION OPERATORS

The set of differential equations that defines the system evolution can be solved numerically with different methods. For example, they can be solved with spectral methods, by expanding a state vector onto a suitable basis of functions, or with finite-difference methods in which the derivatives in the differential equation of motions are replaced by finite difference approximations at a discrete set of grid points in space. The ECMWF primitive equation model solves the system evolution equations partly in spectral space, and partly in grid point space.

Denote by x_g the grid point representation of the state vector x, by S the spectral-to-grid point transformation operator, $x_g = Sx$, and by Gx_g the multiplication of the vector x_g, defined in grid-point space, by the function $g(s)$:

$$\begin{aligned} g(s) &= 1 \forall s \in \Sigma \\ g(s) &= 0 \forall s \notin \Sigma \end{aligned} \tag{26}$$

where s defines the coordinate of a grid point, and Σ is a geographical region.

Define the function $w(n)$ in spectral space as

$$\begin{aligned} w(n) &= 1 \forall n \in \Omega \\ w(n) &= 0 \forall n \notin \Omega \end{aligned} \tag{27}$$

where n identifies a wave number and Ω is a sub-space of the spectral space.

Consider a vector x. The application of the local projection operator T defined as

$$T = S^{-1}GS \tag{28}$$

to the vector x sets the vector x to zero for all grid points outside the geographical region Σ. Similarly, the application of the spectral projection operator W to the vector x sets to zero its spectral components with wave number outside Ω.

The projection operators T and W can be used either at initial or at final time, or at both times. As an example, these operators can be used to formulate the following problem: find the perturbations with (i) the fastest growth during the time interval $t - t_0$, (ii) unitary E_0-norm and wave components belonging to Ω_0 at initial time, (iii) maximum E-norm inside the geographical region Σ and wave components belonging to Ω_1 at optimization time. This problem can be solved by the computation of the singular values of the operator

$$K = E^{-1/2} T S_1 L S_0 E_0^{-1/2} \tag{29}$$

REFERENCES

Appenzeller, Ch., Davies, H. C., Popovic, J. M., Nickovic, S., & Gavrilov, M. B., 1996: PV morphology of a frontal-wave development. Met. and Atm. Phys., 58, 21-40.

Barkmeijer, J., van Gijzen, M., & Bouttier, F., 1998: Singular vectors and estimates of the analysis error covariance metric. Q. J. R. Meteor. Soc., 124, 549, 1695-1713.

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

Barkmeijer, J., Buizza, R., Palmer, T. N., & Puri, K., 1999b: Tropical singular vectors computed with linearized diabatic physics. Q. J. R. Meteor. Soc., submitted.

Bishop, C. H., & Toth, Z., 1998: Ensemble Transformation and Adaptive Observations. J. Atmos. Sci., under revision.

Borges, M., & Hartmann, D. L., 1992: Barotropic instability and optimal perturbations of observed non-zonal flows. J. Atmos. Sci., 49, 335-354.

Buizza, R., 1992: Unstable perturbations computed using the adjoint technique. ECMWF Research Department Technical Memorandum No. 189, ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Buizza, R., 1994: Sensitivity of Optimal Unstable Structures. Q. J. R. Meteorol. Soc., 120, 429-451.

Buizza, R., 1997: The singular vector approach to the analysis of perturbation growth in the atmosphere. Ph. D. thesis, University College London, Gower Street, London.

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

Buizza, R., & Montani, A., 1999: Targeting observations using singular vectors. J. Atmos. Sci., 56, 2965-2985.

Buizza, R., Tribbia, J., Molteni, F., & Palmer, T. N., 1993: Computation of optimal unstable structures for a numerical weather prediction model. Tellus, 45A, 388-407.

Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., & Wedi, N., 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. Q. J. R. Meteorol. Soc., 124, 1935-1960.

Buizza, R., Miller, M., & Palmer, T. N., 1999: Stochastic simulation of model uncertainties. Q. J. R. Meteorol. Soc., 125, 2887-2908.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. J. Meteor., 4, 135-162.

Charney, J. G., 1948: On the scale of atmospheric motions. Geofys. Publ., 17, 1-17.

Courtier, P., Freydier, C., Geleyn, J.-F., Rabier, F., & Rochas, M., 1991: The Arpege project at Météo-France. Proceedings of the ECMWF Seminar on Numerical methods in atmospheric models, Vol. II, 193-231. ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Courtier, P., Derber, J., Errico, R., Louis, J.-F., & Vukicevic, T., 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. Tellus, 45A, 342-357.

Courtier, P., Thepaut, J.-N., & Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol. Soc., 120, 1367-1388.

Downton, R. A., & Bell, R. S., 1988: The impact of analysis differences on a medium-range forecast. Meteorol. Mag., 117, 279-285.

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. Mon. Wea. Rev., 122, 703-713.

Ehrendorfer, M., & Tribbia, J. J., 1995: Efficient prediction of covariances using singular vectors. Preprint Volume, 6th International Meeting on Statistical Climatology, Galway, Ireland, 135-138.

Epstein, E. S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.

Evensen, G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geoph. Res., 99, 10,143-10,162.

Farrell, B. F., 1982: The initial growth of disturbances in a baroclinic flow. J. Atmos. Sci., 39, 8, 1663-1686.

Farrell, B. F., 1988: Optimal excitation of neutral Rossby waves. J. Atmos. Sci., 45 (2), 163-172.

Farrell, B. F., 1989: Optimal excitation of baroclinic waves. J. Atmos. Sci., 46 (9), 1193-1206.

Fleming, R. J., 1971a: On stochastic dynamic prediction. I: the energetics of uncertainty and the question of closure. Mon. Wea. Rev., 99, 851-872.

Fleming, R. J., 1971b: On stochastic dynamic prediction. II: predictability and utility. Mon. Wea. Rev., 99, 927-938.

Gleeson, T. A., 1970: Statistical-dynamical predictions. J. Appl. Meteorol., 9, 333-344.

Golub, G. H., & Van Loan, C. F., 1983: Matrix computation. North Oxford Academic Publ. Co. Ltd., pp. 476.

Hall, M. C. G., & Cacuci, D. G., 1983: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. J. Atmos. Sci., 46, 9, 1193-1206.

Holton, J. R., 1992: An Introduction to Dynamic Meteorology. Academic Press Inc., 511 pp.

Holland, G. J., McGeer, T., & Youngren, H., 1992: Autonomous aerosondes for economical atmospheric soundings anywhere on the globe. Bull. Amer. Met. Soc., 73, 1987-1998.

Hoskins, B. J., & Valdes, P. J., 1990: On the exhistence of storm tracks. J. Atmos. Sci., 47, 1854-1864.

Hoskins, B. J., McIntyre, M. E., & Robertson, A. W., 1985: On the use and significance of isentropic potential vorticity maps. Q. J. R. Meteorol. Soc., 111, 877-946.

Hoskins, B. J., Buizza, R., & Badger, J., 1999: The nature of singular vector growth and structure. Q. J. R. Meteorol. Soc., under revision.

Harrison, M. S. J., Palmer, T. N., Richardson, D. S., & Buizza, R., : 1999: Analysis and model dependencies in medium-range ensembles: two transplant case studies. Q. J. R. Meteorol. Soc., 125, 2487-2515.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., & Mitchell, H., 1996: A system simulation approach to ensemble prediction. Mon. Wea. Rev., 124, 1225-1242.

Houtekamer, P. L., & Mitchell, H., 1998: Data assimilation using an ensemble Kalman filter. Mon. Wea. Rev., 126, 796-811.

Jacob, C., 1994: The impact of the new cloud scheme on ECMWF's Integrated Forecasting System (IFS). Proceedings of the ECMWF/GEWEX workshop on Modeling, validation and assimilation of clouds, ECMWF, Shinfield Park, Reading RG2 9AX, 31 October - 4 November 1994, pp. 464.

Joly, A., Jorgensen, D., Shapiro, M. A., Thorpe, A. J., Bessemoulin, P., Browning, K. A., Cammas, J.-P., Chalon, J.-P., Clough, S. A., Emanuel, K. A., Eymard, R., Gall, R., Hildebrand, P. H., Langland, R. H., Lamaître, Y., Lynch, P., Moore, J. A., Persson, P. O. G., Snyder, C., & Wakimoto, R. M., 1996: The Fronts and Atlantic Storm-Track Experiment (FASTEX): Scientific objectives and experimental design. Report Number 6. The FASTEX Project Office, Météo-France, CNRM, 42 Avenue Coriolis, Toulouse, France. (Also submitted to Bull. Am. Met. Soc.).

Kontarev, G., 1980: The adjoint equation technique applied to meteorological problems. Technical Report No. 21, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK.

Lacarra, J.-F., & Talagrand, O., 1988: Short range evolution of small perturbations in a barotropic model. Tellus, 40A, 81-95.

Langland, R H, & Rohaly, G D, 1996: Adjoint-based targeting of observations for FASTEX cyclones. American Meteorological Society pre-prints of the 7th conference on mesoscale processes, September 9-13, 1996, Reading, UK, pp. 618.

Le Dimet, F.-X., & Talagrand, O., 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus, 38A, 97-110.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. Mon. Wea. Rev., 102, 409-418.

Lorenz, E. N., 1963: Deterministic non-periodic flow. J. Atmos. Sci., 20, 130-141.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. Tellus, 17, 321-333.

Lorenz, E. N., 1993: The essence of chaos. UCL Press, pp 227.

Mahfouf, J.-F., 1999: Influence of physical processes on the tangent-linear approximation. Tellus, 51A, 147-166.

Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Molteni, F., & Buizza, R., 1999b: High resolution ensemble prediction. Part II: an application to four flood events. Q. J. R. Meteorol. Soc., submitted

Morcrette, J.-J., 1990: Impact of changes to the radiation transfer parameterisation plus cloud optical properties in the ECMWF model. Mon. Wea. Rev., 118, 847-873.

Molteni, F., & Palmer, T. N., 1993: Predictability and finite-time instability of the northern winter circulation. Q. J. R. Meteorol. Soc., 119, 1088-1097.

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.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., & Paccagnella, T., 1999: High resolution ensemble prediction. Part I: representative members definition and global applications. Q. J. R. Meteorol. Soc., submitted.

Montani, A., Thorpe, A. J., Buizza, R., & Unden, P., 1999: Forecast skill in the ECMWF model using targeted observations during FASTEX. Q. J. R. Meteorol. Soc., in press.

Noble, B., & Daniel, J. W., 1977: Applied linear algebra, Prenctice-Hall, Inc., pp. 477.

Palmer, T. N., 1997: On parametrizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography. Proceedings of the ECMWF workshop on New insights and approaches to convective parametrization, ECMWF, Shinfield Park, Reading RG2-9AX, UK, 328-337.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., & Tribbia, J., 1993: Ensemble prediction. Proceedings of the ECMWF Seminar on Validation of models over Europe: vol. I, ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Pu, Z.-X., Kalnay, E., Sela, J., & Szunyogh., 1997: Sensitivity of forecast error to initial conditions with a quasi-inverse linear method. Mon. Wea. Rev., 125, 2479-2503.

Pu, Z.-X., Kalnay, E., & Toth, Z., 1998: Application of the quasi-inverse linear and adjoint NCEP global models to targeted observations during FASTEX. Pre-prints of the 12th Conference on Numerical Weather Prediction, 11-16 January 1998, Phoenix, Arizona, 8-9.

Puri, K., Barkmeijer, J., & Palmer, T. N., 1999: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. Q. J. R. Meteorol. Soc., submitted.

Rabier, F., Klinker, E., Courtier, P., & Hollingsworth, A, 1996: Sensitivity of forecast errors to initial conditions. Q. J. R. Meteorol. Soc., 122, 121-150.

Richardson, L. F., 1922: Weather Prediction by Numerical Process. Cambridge University Press (reprt. Dover, New York).

Richardson, D. S., 1998: The relative effect of model and analysis differences on ECMWF and UKMO operational forecasts. Proceedings of the ECMWF Workshop on Predictability, ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, C., & Wergen, W., 1989: The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. Meteorol. Atmos. Phys., 40, 28-60.

Simmons, A. J., Mureau, R., & Petroliagis, T., 1995: Error growth and predictability estimates for the ECMWF forecasting system. Q. J. R. Meteorol. Soc., 121, 1739-1771.

Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. Bull. Am. Meteor. Soc., 77, 953-961.

Somerville, R. C. J., 1979: Predictability and prediction of ultra-long planetary waves. Pre-prints of the Amer. Meteor. Soc. Fourth Conference on Numerical Weather Prediction, Silver Spring, MD, 182-185.

Strang, G., 1986: Introduction to applied mathematics. Wellesley-Cambridge Press, pp 758.

Thorpe, A. J., & Shapiro, M. A., 1995: FASTEX: Fronts and Atlantic Storm Track Experiment. The Science Plan. Available form the FASTEX Project Office, July 1995, pp. 25.

Thorpe, A. J., Buizza, R., Montani, A., & Palmer, T. N., 1998: Chaotic control for weather prediction. Royal Society New frontiers in Science Exhibition, June 1998, The Royal Society, 6 Carlton House Terrace, London SW1Y-5AG, UK.

Tiedtke, M., 1993: Representation of clouds in large-scale models. Mon. Wea. Rev., 121, 3040-3060.

Toth, Z., & Kalnay, E., 1993: Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Met. Soc., 74, 2317-2330.

Tracton, M. S., & Kalnay, E., 1993: Operational ensemble prediction at the National Meteorological Center: practical aspects. Weather and Forecasting, 8, 379-398.

Urban, B., 1985: Error maximum growth in simple meteorological models (in French). Meteorologie Nationale Internal Report.

Viterbo, P., & Beljaars, C. M., 1995: An improved land surface parametrisation scheme in the ECMWF model and its validation. J. Clim., 8, 2716-2748.

Von Neumann, J., & Richtmeyer, R. D., 1950: A method for the numerical calculation of hydrodynamical shocks. J. Appl. Phys., 21, 232.

The Prediction of Uncertainty in Numerical Weather Forecasting

T.N.Palmer
ECMWF
Shinfield Park
Reading RG2 9AX
UK

Abstract

The predictability of weather and climate forecasts is determined by the projection of uncertainties in both initial conditions and model formulation, onto flow-dependent instabilities of the chaotic climate attractor. Since it is essential to be able to estimate the impact of such uncertainties on forecast accuracy, no weather or climate prediction can be considered complete without a forecast of the associated flow-dependent predictability. The problem of predicting uncertainty can be posed in terms of the Liouville equation for the growth of initial uncertainty, or a form of Fokker-Planck equation if model uncertainties are also taken into account. However, in practice, the problem is approached using ensembles of integrations of comprehensive weather and climate prediction models, with explicit perturbations to both initial conditions and model formulation; the resulting ensemble of forecasts can be interpreted as a probabilistic prediction.

Many of the difficulties in forecasting predictability arise from the large dimensionality of the climate system, and special techniques to generate ensemble perturbations have been developed. Special emphasis is placed on the use of singular-vector methods to determine the linearly unstable component of the initial probability density function. Methods to sample uncertainties in model formulation are also described. Practical ensemble prediction systems are described, and examples of resulting probabilistic weather forecast products shown. Methods to evaluate the skill of these probabilistic forecasts are outlined. By using ensemble forecasts as input to a simple decision-model analysis, it is shown that that probability forecasts of weather have greater potential economic value than corresponding single deterministic forecasts with uncertain accuracy.

1 Introduction

Most meteorologists would rate the development of the global weather and climate prediction model as amongst the most important scientific developments in our field over the last 50 years. Using such models we can make forecast of daily weather out to 10 days, forecasts of El Niño and its global impact on timescales of seasons, and make projections of anthropogenic climate change decades ahead.

Historically, such predictions have tended to be made in a deterministic mode ('a blocking anticyclone will develop five days from now'); however, although the limited the predictability of weather forecasts had been clearly stated over 40 years ago (Thompson, 1957), forecast error bars were rarely stated. This point was made forcibly some years later by Tennekes et al (1991), and from around that time, a serious attempt has been made to express numerical weather forecasts in probabilistic terms using ensemble prediction techniques (eg Palmer et al , 1993; Toth and Kalnay, 1993). In the decade of the 1990s, ensemble prediction has become an established tool in numerical weather prediction; probability forecasting is here to stay.

This paper deals with probabilistic weather prediction from its theoretical basis, through an outline of practical methodologies, to an analysis of validation techniques including estimates of potential economic value. In section 2, we consider how to forecast uncertainty, assuming a perfect deterministic forecast model. The evolution equation for the probability density function (pdf) of the climate state vector is the Liouville equation; an example of its solution is given for illustration. However, application to the real climate system is severely hampered by two fundamental problems. The first is directly associated with the dimensionality of the climate equations; current numerical weather prediction models comprise $O(10^7)$ individual scalar variables. The second problem (not unrelated to the first) is that, in practice, the initial pdf is not itself well known.

To amplify on this last remark, a description of contemporary data assimilation schemes is described in section 3. Such schemes are based on minimising a cost function which combines these observations with a background estimate of the initial state provided by a short-range model forecast from an earlier set of initial conditions. In principle, given Gaussian error statistics, the Hessian or second derivative of the cost function determines the initial pdf. However, in practice, there are significant shortcomings in our ability to estimate this pdf.

The number of degrees of freedom in comprehensive climate and weather prediction models is not determined by any scientific constraint (there is no obvious 'gap' in the energy spectrum of atmospheric motions), but rather by the degree of complexity than can be accommodated using current computer technology. As such, there are inevitably processes occurring in the atmosphere and oceans which are only partially resolved and cannot be accurately described by a parametrised closure approximation. Section 4 describes two recent attempts to represent the pdf associated with this uncertainty in the computational representation of the equations of motion of climate: the multi-model ensemble, and stochastic parametrisation.

Section 5 describes the basis behind attempts to make probability forecasts from ensembles of model integrations. If the ensemble of forecast phase-space trajectories evolve though a relatively stable part of the climate attractor, then resulting probability forecasts will be relatively sharp. Conversely, if the ensemble passes through a particularly unstable part of the attractor, then the corresponding forecast probability may be little different from a long-term climatological frequency. Examples of practical probability forecast products are given.

The question of how to validate probability forecasts is discussed in section 6. One particular techniques are described based on a root mean square distance between the probability forecast of a dichotomous event and the corresponding verification. This measure allows one to formulate the notion of reliability of probability forecasts.

A fundamental question when assessing probability forecasts is whether a useful level of skill has been attained. To discuss this quantitatively, a simple cost/loss decision model is applied in section 7. It is shown, that the (potential) economic value of probability weather forecasts for a variety of users, is higher than the corresponding value from single deterministic forecasts.

Concluding remarks are made in section 8.

2 A theoretical approach to probability forecasting

The evolution equations in a climate or weather prediction model are conventionally treated as deterministic. These (N dimensional) equations, based on spatially-truncated momentum, energy, mass and composition conservation equations will be written schematically as

$$\dot{X} = F[X] \tag{1}$$

where X describes an instantaneous state of the climate sytem in N-dimensional phase space. Equation 1 is fundamentally nonlinear and deterministic in the sense that, given an initial state X_a, the equation determines a unique forecast state X_f.

The meteorological and oceanic observing network is sparse over many parts of the world, and the observations themselves are obviously subject to measurement error. The resulting uncertainty in the initial state can be represented by the pdf $\rho(X, t_a)$; given a volume V of phase space, then $\int_V \rho(X, t_a) dV$ is the probability that the true initial state X_{true} at time t_a lies in V. If V is bounded by an isopleth of ρ (ie co-moving in phase space), then, from the determinism of equation 1, the probability that X_{true} lies in V is time invariant. Hence, (similar to the mass continuity equation in physical space), the evolution of ρ is given by the Liouville conservation equation (introduced in a meteorological context by Gleeson, 1966, and Epstein, 1969)

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial X}(\dot{X}\rho) = 0 \tag{2}$$

where $\dot{X}$ is given by equation 1. In the second term of this equation, there is an implied summation over all the components of X .

Fig 1 illustrates schematically the evolution of an isopleth of $\rho(X, t)$. For simplicity we assume the initial pdf is isotropic (eg by applying a suitable coordinate transformation). In the early part of the forecast, the isopleth evolves in a way consistent with linearised dynamics; the N-ball at initial time has evolved to an N-ellipsoid at forecast time t_1. For weather scales of $0(10^6)$ km, this linear phase lasts for about 1-2 days into the forecast. Beyond this time, the isopleth starts to deform nonlinearly. The third schematic shows the isopleth at a forecast range in which errors are growing nonlinearly. Predictability is finally lost when the forecast pdf $\rho(X, t)$ has evolved irreversibly to the invariant distribution ρ_{inv} of the attractor. This is shown schematically in Fig 1 using the Lorenz (1963) attractor.

The growth of the pdf through the forecast range is a function of the initial state. This can be seen by considering a small perturbation δx to the initial state X_a. From equation 1, the evolution equation for δx is given by

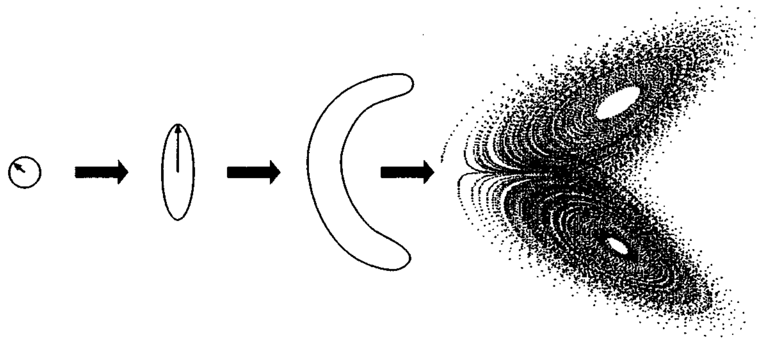

Figure 1: Schematic evolution of an isopleth of the probability density function (pdf) of initial and forecast error in N-dimensional phase space. (a) At initial time, (b) during the linearised stage of evolution. A (singular) vector pointing along the major axis of the pdf ellipsoid is shown in (b), and its pre-image at initial time is shown in (a). (c) The evolution of the isopleth during the nonlinear phase is shown in (c); there is still predictability, though the pdf is no longer Gaussian. (d) Total loss of predictability, occurring when the forecast pdf is indistinguishable from the attractor's invariant pdf.

$$\delta\dot{x} = J\delta x \tag{3}$$

where the Jacobian is defined as

$$J = dF/dX \tag{4}$$

Since $F[X]$ is at least quadratic in X, then J is at least linearly dependent on X. This dependency is illustrated in Fig 2 showing the growth of an initial isopleth of an idealised pdf at three different positions on the Lorenz (1963) attractor. In the first position, there is little growth, and hence large local predictability. In the second position there is some growth as the pdf evolves towards the lower middle half of the attractor. In the third position, initial growth is large, and the resulting predictability is correspondingly small.

The Liouville equation can be formally solved to give the value of ρ at a given point X in phase space at forecast time t (Ehrendorfer, 1994a, Palmer, 2000). Specifically

$$\rho(X,t) = \rho(X', t_a)/\det M(t, t_a) \tag{5}$$

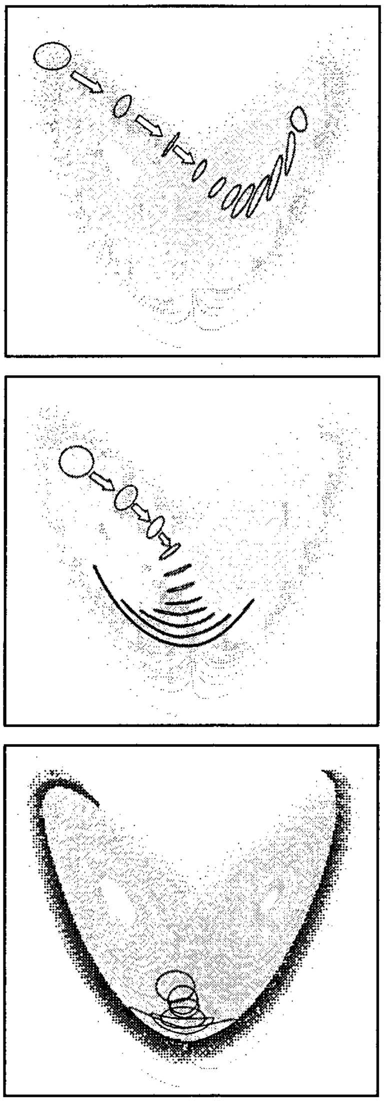

Figure 2: Phase-space evolution of an ensemble of initial points on the Lorenz (1963) attractor, for three different sets of intial conditions. Predictability is a function of inital state.

where

$$M(t, t_a) = \exp \int_{t_a}^{t} J(t')dt' \tag{6}$$

is the so-called forward tangent propagator, mapping a perturbation $\delta x(t_a)$, along the nonlinear trajectory from X' to X', to

$$\delta x(t) = M(t, t_a)\delta(t_a) \tag{7}$$

and X' in this equation corresponds to that initial point, which, under the action of equation 1 evolves to the given point X at time t.

A simple example which illustrates this solution to the Liouville equation is given in Fig 3, for a 1 dimensional Riccati equation (Ehrendorfer, 1994a)

$$\dot{X} = aX^2 + bX + c \tag{8}$$

where $b^2 > 4ac$, based on an initial Gaussian pdf. The pdf evolves away from the unstable equilibrium point at $X = -1$ and therefore reflects the dynamical properties of equation 8. Within the integration period, this pdf has evolved to the nonlinear phase.

The forward tangent propagator plays an important role in meteorological data assimilation systems; see section 3 below. However, even though the forward tangent propagator may exist as a piece of computer code, this does not mean that the Liouville equation can be readily solved for the weather prediction problem. For example, the determinant of the forward tangent propagator is determined by the product of all its singular values. For a comprehensive weather prediction model, a determination of the full set of $O(10^7)$ singular values is currently impossible. Secondly, the inversion of 1 to find an initial state X', given a forecast state X, is itself problematic. Thirdly, a particular type of weather at a particular location is not related 1-1 with a state X of the climate system. For example, to estimate the probability of it raining in London two days from now, we would have to apply equation 5 and the inversion to find X', to each state X_{rain} on the climate attractor, for which it is raining in London.

An alternative to using the solution form 5 is to integrate the partial differential equation 2 by randomly sampling the initial pdf, and integrating each sampled point using 1; the Monte-Carlo solution. However, the problem of dimensionality continues to be a significant issue. If phase-space is N dimensional, then, even in the linear phase, $O(N^2)$ integrations will be needed to determine the forecast error covariance matrix. In the nonlinear phase,

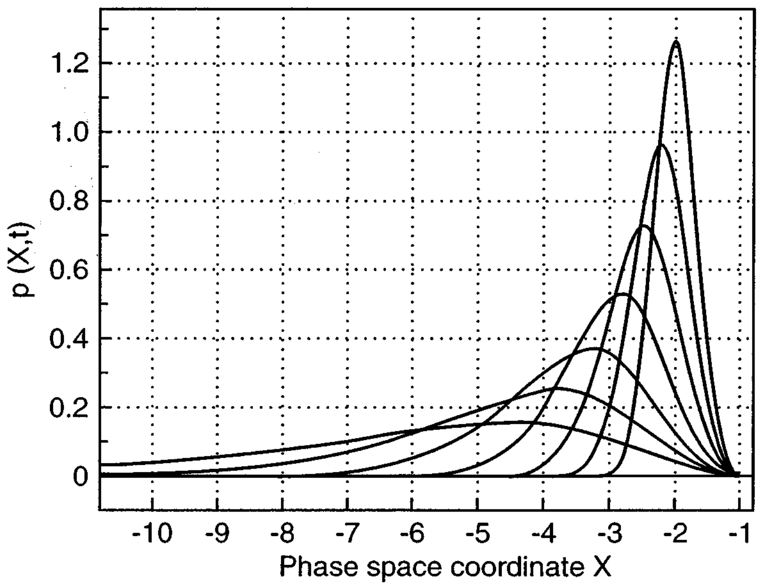

Figure 3: An analytical solution to the Liouville equation for an initial Gaussian pdf (shown peaked on the right-hand side of the figure) evolved using the Riccati equation (see text). From Ehrendorfer (1994a).

many more integrations are needed to determine the pdf, as it begins to wrap itself around the attractor. Ehrendorfer (1994b) has shown that even for a 3-dimensional dynamical system, a Monte-Carlo sampling of $O(10^2)$ points can be insufficient to determine the pdf within the nonlinear range.

Yet another method of solution of the Liouville equation is possible, writing equation 2 in terms of an infinite hierarchy of equations for the moments of ρ, and applying some closure to this set of moments (Epstein, 1969). This method is certainly useful for evolving the pdf within the linearised phase, and indeed forms the basis of the so-called Kalman filter approaches to data assimilation (see section 3 below). Ehrendorfer (1994b) has shown that in the nonlinear phase, substantial errors in estimating the first and second moments of ρ can arise from neglecting third and higher order moments.

In conclusion, whilst a formal analytic solution can be found to the problem of predicting the forecast pdf, there are practical problems associated with the dimension of the underlying dynamical system.

3 The probability density function of initial error

In order to discuss how the pdf of initial error can be estimated in weather and climate prediction, it is necessary to outline the method by which observations are used to determine the initial conditions for a deterministic weather or climate forecast.

In meteorology and oceanography, data assimilation is a means of obtaining a forecast initial state which in some well-defined sense optimally combines the available observations for a particular time with an independent background state (Daley, 1991). This background state is usually a short-range forecast (eg 6 hour) from an estimate of the initial state valid at an earlier time, and this carries forward information from observations from earlier times. A very simple example of the basic notion can be illustrated by considering two different independent estimates, s_o and s_b, of a scalar s. Suppose that the errors associated with these two estimates are random, unbiased and normally distributed, with standard deviations σ_o and σ_b respectively. Then the maximum-likelihood estimate of s is the state s_a which minimises the cost function

$$J(s) = \frac{(s - s_b)^2}{2\sigma_b^2} + \frac{(s - s_o)^2}{2\sigma_o^2}. \tag{9}$$

The least-squares solution

$$s_a = s_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}[s_o - s_b] \tag{10}$$

is easily found. The error associated with s_a is normally distributed with variance given by

$$\partial^2 J/\partial s^2 = \sigma_b^{-2} + \sigma_o^{-2} = \sigma_a^{-2} \tag{11}$$

The data assimilation technique used in weather prediction (eg at ECMWF) is a multi-dimensional generalisation of this technique (Courtier et al, 1994, 1998). The analysed state X_a of the atmospheric state vector is found by minimising the (incremental) cost function

$$J(X) = \frac{1}{2}(X - X_b)^T B^{-1}(X - X_b) + \frac{1}{2}(HX - Y)^T O^{-1}(HX - Y) \tag{12}$$

where X_b is the background state, B and O are covariance matrices for the pdfs of background error and observation error respectively, H is the so-called observation operator, and Y denotes the vector of available observations. For example, if Y includes a radiance measurement taken by an infrared radiometer onboard a satellite orbiting the earth then HX includes an estimate of the infrared radiance that would be emitted by a model atmosphere as represented by the state vector X. Similarly, if Y includes a surface pressure measurement taken at some point p on the earth's surface, then HX includes the surface pressure at p given X. Since X is finite dimensional, the operator H inevitably involves an interpolation to p. Similar to equation 11, the Hessian of J is given by (Fisher and Courtier, 1995)

$$\nabla\nabla J = B^{-1} + H^T O^{-1} H \equiv A^{-1} \tag{13}$$

We refer to A as the analysis error covariance matrix.

In the current ECMWF operational data assimilation system, the background error covariance matrix B is estimated using climatological forecast error statistics from past cases (Parrish and Derber, 1992); in particular, B is not dependent on the present state of the atmospheric circulation. This is believed to introduce considerable imprecision in the estimate of the initial pdf as given by 13. This estimate can be improved; within the linearised regime (cf Fig 1), the forecast error covariance matrix F implied by equations 5 and 7 can be written

$$F(t) = M(t, t_a)A(t_a)M^T(t_a, t) \tag{14}$$

where M is the tangent propagator along the trajectory between the initial state X_a and the forecast state X at time t. Since the time between consecutive analyses (typically 6 hours) is broadly within this linearised regime, then a flow-dependent estimate of the background error covariance matrix at time t_a can be obtained by propagating the analysis error covariance matrix from the earlier analysis time t_{a-1}, ie

$$B(t_a) = M(t_a, t_{a-1})A(t_{a-1})M^T(t_{a-1}, t_a) \tag{15}$$

The propagator M, and its transpose M^T are essential components of 4-dimensional data assimilation (Courtier et al, 1994) where observations are assimilated over a time window. Using M, an increment δx can be evaluated at the same time that an observation is taken. Given the dimension of comprehensive weather prediction models, M is not known in matrix form, and is represented in operator form (cf equation 7). Similarly the transpose M^T is also represented in operator form M^* (see equation 18 below) and is known as the adjoint (tangent) propagator.

However, equation 15 is computationally intractable for numerical weather prediction, requiring $O(10^{14})$ individual linearised integrations of M for a complete specification of the propagated matrix MAM^T. Three possible solutions have been proposed. The first is essentially a Monte-Carlo solution, whereby a random sampling of A is evolved using M (Evensen, 1994; Andersson and Fisher, 1999). The second proposal involves solving the propagation equation 15 with an intermediate complexity model (Ehrendorfer, 1999). The final proposal (the so-called reduced-rank Kalman filter; Fisher, 1998) is to propagate A explicitly only in the appropriate unstable subspace defined by the dominant flow-dependent local instabilities of the attractor. Broadly, speaking, the proposal is to have the best possible knowledge of the initial pdf in that part of phase space from which forecast errors are most likely to grow. At present these three different proposals are being evaluated.

Since the notion of local flow-dependent instability features strongly in later sections of this paper, it is worth outlining some more detail on how these instabilities can be estimated. First consider a Euclidean inner product $< .., .. >$ so that for any perturbations δx, δy,

$$< \delta x, \delta y >= \sum \delta x_i \delta y_i \tag{16}$$

In terms of $< .., .. >$ the adjoint tangent propagator M^* is defined by

$$\delta y(t_a) = M^*(t_a, t)\delta y(t) \tag{17}$$

where

$$< \delta y, M\delta x >=< M^*\delta y, \delta x > \tag{18}$$

for an arbitrary pair of perturbations $\delta x(t_a)$, $\delta y(t)$.

The analysis error covariance matrix A defines a secondary inner product

$$(.., ..) =< .., A^{-1}.. > \tag{19}$$

Here A^{-1} is the covariant form of an analysis error covariance metric, $g_{A^{-1}}$ (Palmer et al, 1998). Hence the perturbation $\delta x(t)$ which has maximum Euclidean amplitude at t and unit $g_{A^{-1}}$ norm at initial time t_a is given by

$$\max_{x(t_a)\neq 0} \frac{< \delta x(t), \delta x(t) >}{< \delta x(t_a), A^{-1}\delta x(t_a) >} = \max_{x(t_a)\neq 0} \frac{< \delta x(t_a), M^* M\delta x(t_a) >}{< \delta x(t_a), A^{-1}\delta x(t_a) >} \tag{20}$$

This is equivalent to finding the dominant eigenvector of the generalised eigenvector equation

$$M^* M\delta x(t_a) = \lambda A^{-1}\delta x(t_a) \tag{21}$$

Formally, by taking the square root of A, equation 21 can be transformed to a singular vector equation which can be solved using a Lanczos algorithm (Strang, 1986; Buizza and Palmer, 1995). More generally, equation 21 is solved using a generalised Davidson algorithm (Barkmeijer et al, 1998). We refer to the solutions $\delta x(t_a)$ in equation 21 as $g_{A^{-1}}$-singular vectors of M.

The set of dominant $g_{A^{-1}}$ singular vectors of M (with largest singular values $\sigma_i = \sqrt{\lambda_i}$) defines an unstable subspace in the tangent space at t_a. It comprises the set of most rapidly-growing directions defined locally on phase space, relative to a basic-state trajectory between t_a and t, subject to the constraint that the initial perturbations are normalised with respect to the initial pdf. At forecast time t, these singular vectors have evolved into the major axes of the forecast error ellipsoid, or, equivalently, into the eigenvectors of the forecast error covariance matrix. The first two parts of Fig 1 show schematically a dominant $g_{A^{-1}}$ singular vector at initial and forecast time.

4 The probability density function of model error

So far, we have assumed the 'classical' chaotic paradigm, that loss of predictability occurs only because of inevitable uncertainty in initial conditions. However, there are also inevitable uncertainties in our ability to represent computationally the governing equations of climate. These uncertainties can contribute both to random and systematic error in model performance. In practice, as discussed below, it is not easy to separate the predictability problem into a component associated with initial error and a component associated with model error.

As mentioned, weather and climate models have a resolution of O(100km) in the horizontal. This immediately raises the problem of closure - how to represent the effect of partially resolved or unresolved processes onto the resolved state vector X. The effects of topography and clouds are examples. In weather and climate prediction models, equation 1 is generally expressed as

$$\dot{X} = G[X] + P[X; \alpha(X)] \tag{22}$$

where $P[X; \alpha(X)]$ stands for some parametrised representation of unresolved processes, and $G[X]$ represents terms in the equations of motion associated directly with resolved scales. Conceptually, a parametrisation is usually based on the notion of a statistical ensemble of sub-grid scale processes within a grid box, in some secular equilibrium with the grid-box mean flow. This allows equation 22 to be written

$$\dot{X}_j = G_j[X] + P[X_j; \alpha(X_j)] \tag{23}$$

where X_j and G_j represent the projection of X and G into the subspace associated with a single grid box x_j in physical space. Borrowing ideas from statistical mechanics, a familiar parametrisation might involve the diffusive approximation, where α would be a diffusion coefficient which might depend on the Richardson number of the large-scale flow. However, whilst diffusive closures do in fact play a role for example in representing the effects of the turbulent planetary boundary in the lower kilometer of the atmosphere, they are certainly insufficient, and, in some circumstances, may be fundamentally flawed. To see this, it is enough to concentrate on one relevant process - atmospheric convection.

Often, such instability is released through overturning circulations whose horizontal scales are small compared with the smallest resolved scale of a global weather or climate model. However, at the large-scale end of the spectrum of convectively-driven circulations is the organised mesoscale convective complex (Moncrieff, 1992), with horizontal scales of perhaps 100km. They can be simulated explicitly in regional models with O(1km) resolution, but such resolution is not practicable for global weather and climate models.

The form of parametrisation given in equation 23 is appropriate for describing cumulus and simple cumulonimbus. For example, in a contemporary convective parametrisation (eg Betts and Miller, 1986) if the resolved-scale vertical temperature gradient at x_j is convectively unstable, then over some prescribed timescale (given by α) P will operate to relax X_j back to stability. On the other hand, the existence of organised mesoscale convective complexes poses a problem for parametrisations of this form. In particular, the basic assumption of a quasi-equilibrium of sub-gridscale convectively-forced motions (with the implication that the kinetic energy released by overturning circulations is dissipated on sub-grid scales, rather than injected into the large scale) cannot be fully justified.

One means of addressing the erroneous assumption of deterministic locality in equation 23 would be to add to equation 23 a stochastic energy source term $\mathcal{S}(X_j;\alpha)$. Recognising this, Buizza et al (1999) have proposed the simple stochastic form

$$\dot{X}_j = G_j[X] + \beta P[X_j;\alpha(X_j)] \tag{24}$$

where β is a stochastic variable representing a random variable drawn from a uniform pdf between 0.5 and 1.5. Stochastic representation of sub-grid processes is a technique already utilised in turbulent flow simulations (Mason and Thompson, 1992).

An example of the impact of the random effect of the stochastic parametrization represented by equation 24 is given in Fig 4 which shows sea-level pressure over part of Australia and the west Pacific from four 2-day integrations of

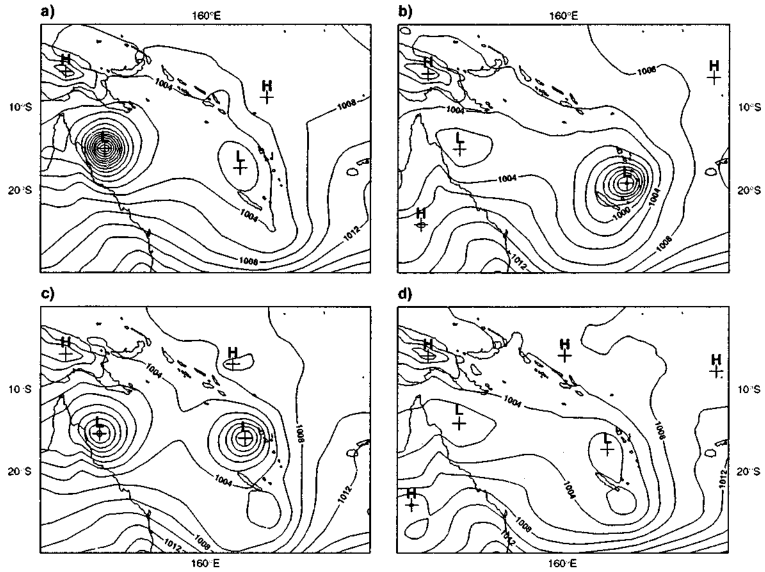

Figure 4: Four 2-day integrations of the ECMWF model from identical starting conditions but different realisations of the stochastic parametrisation scheme represented by equation 24 with parameter settings as given in Buizza et al (1999). The field shown is sea-level pressure over parts of Australia and the west Pacific. The depressions in the pressure field represent potential tropical cyclones.

the ECMWF model. The integrations have identical starting conditions, but different realisations of the pdf represented by β. The figure shows two tropical cyclones. The intensity of the cyclones can be seen to be very sensitive to the realisation of the stochastic parametrisation. In Fig 4a) the western cyclone is intense; in Fig 4b) the eastern cyclone is intense; in Fig 4c) they are both intense; in Fig 4d) neither are intense. This rather extreme example clearly shows the difficulty in predicting tropical cyclone development, and its sensitivity to model parametrisation.

The purpose of the discussion above is to point out that although current parametrisations have been enormously successful in representing subgrid processes, there are inevitable uncertainties in the representation of such processes. In section 2 we represented the evolution of the initial pdf given the deterministic equation 1 in terms of a Liouville equation. In the idealised case where model uncertainties are represented by an additive Gaussian white noise with zero mean and variance Γ , then equation 2 becomes a Fokker-Planck equation (eg Hasselmann, 1976; Moss and McClintock, 1989). However, in practice, a realistic state-dependent stochastic forcing (eg in equation 24) would be too complex for this simple representation to be directly relevant.

For ensemble forecasting (discussed in section 5), there are two other commonly-used techniques for representing model uncertainty. The first technique is the multi-model ensemble (Harrison et al, 1999; Palmer et al, 1999). In the second technique, the values α of the parameters used in the deterministic equation 22 of one particular model, are perturbed in some stochastic manner (cf Houtekamer et al, 1996).

On the other hand, these latter techniques should be seen as conceptually distinct from the type of stochastic physics scheme described schematically in equation 24. In multi-model ensembles, and ensembles with stochastic α, the model perturbations account for the fact that the expected value of the pdf of sub-grid processes is not itself well known. (Hence, for example, there are many different atmospheric convection parametrisation schemes in use around the world; the Betts-Miller scheme desribed above is but one of these. The existence of this ensemble of convection schemes is an indication that the expectation value of the pdf of the effects of sub-grid convection is not known with complete confidence!) By contrast, the stochastic physics scheme described in equation 24 is an attempt to account for the fact that in circumstances of convective organisation, the pdf of sub-grid processes is not especially sharp around the mean. This would argue for the combined use of multi-model ensembles and stochastic physics parametrisation.

5 Probability forecasting by ensemble prediction

In the discussion above, the problem of forecasting uncertainty in weather and climate prediction has been formulated in terms of a Liouville, or, including model uncertainty, a form of Fokker-Planck equation. In practice these equations are solved using ensemble techniques ie from multiple integrations of the governing equations from perturbed initial conditions, using either multiple models, or stochastic parametrizations, to represent model uncertainty. Insofar as the perturbed initial conditions constitute a random sampling of the initial pdf, this could be described as a Monte-Carlo approach. However, the methodologies discussed below for estimating the initial perturbations do not necessarily constitute such a random sampling, for reasons to be discussed. As such, the more general term 'ensemble prediction' is used.

On timescales of 1 day or so, the extratropical weather forecast problem could be described conceptually as the prediction of the development of individual baroclinic disturbances, and their associated weather (fronts, cloud, rainfall etc). On the other hand, the atmospheric circulations exhibit a degree of organisation on timescales of 10 days or so. More precisely there is evidence of 'circulation regimes' characterised by persistence on timescales much longer than an individual weather system, but with transitions between regimes characterised by the faster timescale eg of the dominant baroclinic instability (eg Kimoto and Ghil, 1993).

There is evidence of similar behaviour in tropical atmospheric circulations. Consider, for example, the Asian monsoon. Individual monsoon 'depressions', associated with an instability of the monsoon jet stream, develop on timescales similar to (or faster than) extratropical instabilities, and thus characterise the short-range forecast problem. On the other hand, the so-called active and break phases of the monsoon (which characterise the intraseasonal fluctuations of the monsoon) appear to have distinct regime dynamics with typical residence timescales on the order of 10-20 days (Webster et al, 1998).

Hence on the timescale of 10 days, it is important to be able to estimate uncertainty in predictions of persistence or change in regime type. In the last few years, ensemble forecasting on this 10-day timescale has become an established part of operational global weather prediction (Palmer et al, 1993, Toth and Kalnay, 1993, Houtekamer et al, 1996, Molteni et al, 1996). Different strategies have been proposed for determing the ensemble of starting conditions. Conceptually, the strategies can be delineated according to whether the initial perturbations merely sample observation error, or whether (given the dimension of phase space and the uncertainty in our knowledge of the initial pdf) the initial perturbations are constrained to lie on some dynamically unstable subspace. In the latter case, this dynamical subspace is defined in different ways. Within the meteorological community, there has been a very lively debate on the merits and deficiencies of the different strategies.

A recent analysis of the problem has been performed by Anderson (1997) using the Lorenz (1963) model. First, points are sampled at regular intervals on the attractor. A pdf $\mathcal{O}$ of 'observation error' is then assumed. A sampled point on the attractor is perturbed with a realisation of $\mathcal{O}$; this is the 'analysed state'. A 2-member ensemble is then generated by perturbing about the analysed state with perturbations $\pm p$, where p is again drawn from $\mathcal{O}$. The integration from the state on the attractor is taken as 'truth'. Based on standard measures of skill (see section 6 below), this ensemble is more skilful than any other 2-member ensemble, eg based on singular-vector perturbations of the analysed state. Similar results have been found by Hamill et al (1999) based on an intermediate quasi-geostrophic model. This study concludes that it is enough in generating an ensemble of initial conditions to randomly perturb the observations Y (see section 3) consistent with the observation error covariance matrix O.

On the other hand, this type of experiment is rather idealised. On the basis of the discussion in section 3, a sampling of the initial pdf obtained from equation 12 by perturbing observations, is likely to be a gross underestimate of (for example) the second moment of the initial pdf. Unfortunately, as discussed, there are so many unquantified uncertainties in the actual details of the data assimilation procedure, it is not at all straightforward to quantify this underestimation.

The strategies that use dynamically-constrained perturbations, in some sense bypass the quantification of these uncertainties, and focus on perturbations that are necessarily growing, and hence are likely to contribute to significant forecast error. Two types of dynamically-constrained perturbation have been proposed: the first based on 'bred vectors' (Toth and Kalnay, 1997) and the singular vectors (Buizza and Palmer, 1995) discussed above.

A study of differences between ensembles initalised using bred vectors and singular vectors has recently by Trevisan et al (1999) using the intermediate nonlinear quasi-geostrophic model of Reinhold and Pierrehumbert (1982) described above. A preliminary study of the predictability properties of the model was made by performing relatively large Monte Carlo simulations to each of 2,500 initial states (on the model attractor). Much smaller ensemble integrations were then constructed using the two types of dynamically constrained perturbation. The study focussed on cases of regime transition. It was found that in general, the bred-vector ensemble provided an average error distribution more similar to the Monte Carlo distribution, whilst the singular-vector ensemble provided a more reliable estimate of an upper bound on error growth. For example, a prediction of a low probability of transition of weather regime was found to be much more reliable using the singular vector perturbations than using the bred-vector perturbations in these dynamically-constrained ensembles.

This is of some relevance for user confidence in probability forecasts. As discussed in more detail in section 7, consider a user who could potentially suffer a catastrophic loss L if an event E occurs (in this case a circulation regime change bringing severe weather of some type). The user can take protective action at cost C, which might itself be substantial, but can be presumed to be less than L. Based on the ensemble forecast, one can estimate a probability p that E will occur. If p is sufficiently low, then the user should be able to assume that protective action is not necessary. On these occasions of very low p, the incurrence of a loss L would be seen as a failure of the forecast system, and user confidence in the system would be seriously compromised. Of course, an a posteriori constant offset could be added to the forecast probability, so that very small probabilities would never be issued. But in this case, the ensemble would be no realistic value to the user as a decision tool, since the issued probabilities would always be sufficiently high that protective action would always be taken.

The implication of this analysis is that it is important, arguably paramount, to be sure that the user can be confident in a forecast where the ensemble spread is small. For this reason, it is important in an operational environment to ensure that the initial perturbations are not conservative in the sense of not spanning phase space directions where the pdf is underestimated (due to unquantified errors in the data analysis procedures), and where forecast error growth could be large because of dynamical instability. For these reasons, the ECMWF ensemble system is based in part on initial perturbations using rapidly growing singular vectors (with approximations to the $g_{A^{-1}}$ metric). Examples of the beneficial impact of singular vector perturbations over other types of perturbation are given in Mureau et al (1993) and Gelaro et al (1998).

At the time of writing, the ECMWF Ensemble Prediction System (EPS) comprises 51 forecasts of the ECMWF forecast model (see Buizza et al, 1998: Buizza et al, 1999). As above, the control forecast is run from the operational ECMWF analysis. The 50 perturbed initial conditions are made by adding and subtracting linear combinations of the dominant 25 singular vector perturbations to the operational analysis. In addition, the 50 perturbed forecasts are also run with the stochastic parametrisation defined in equation 24.The EPS is run every day and basic meteorological products are disseminated to all the national meteorological services of the ECMWF Member States. These products often take the form of probability forecasts for dif-

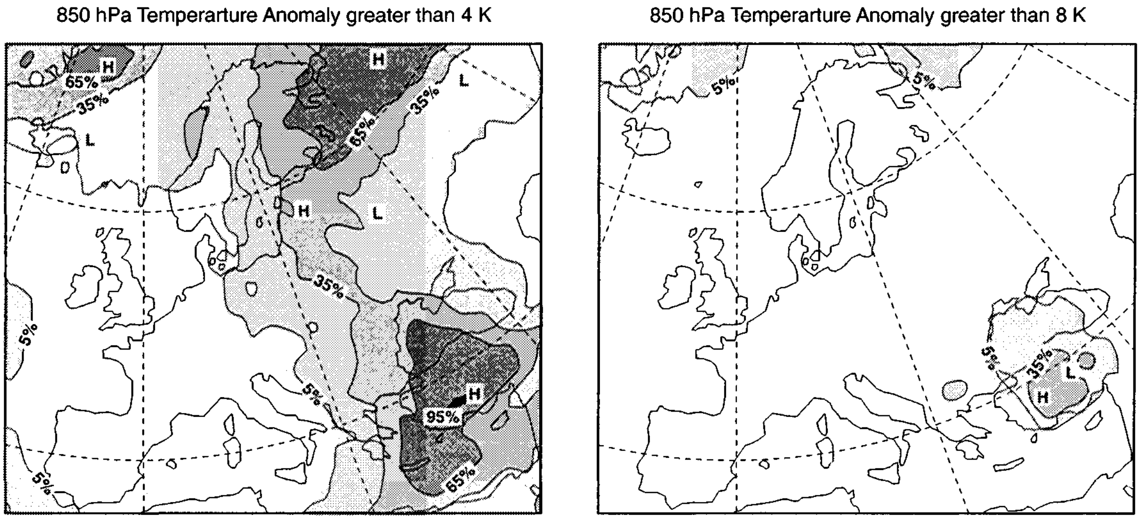

Figure 5: An example of an operational forecast product from the 50-member ECMWF ensemble prediction system. Probability that a) 850hPa temperature is at least 4C above normal, b) 8C above normal, based on a 6-day forecast.

ferent events E, based on the fraction of ensemble members for which E is forecasts. For example, Fig 5 shows day 6 probability forecasts over Europe for the events $E_{>8}, E_{>4}$ defined as: difference of lower tropospheric temperature from a long-term climatology is $> 8C, > 4C$, respectively.

Based on the EPS, many of the European national meteorological services are providing their customers detailed probability forecasts of possible weather parameters for site-specific locations. For example, based on the EPS, the United Kingdom Meteorological Office provides forecast pdfs of surface temperature and rainfall for various specific locations in the UK. Joint probability distributions are also estimated (eg the probability of precipitation with surface temperature near or below freezing).

As discussed above (cf Fig 2), a feature of nonlinear dynamical sytems is the dependence of error growth on initial state. Fig 6 illustrates this, based on two 10-day EPS integrations from starting dates exactly one year apart, the meteorological forecast variable being surface temperature over London. The unperturbed control forecast and verification are also shown. In the first example, the growth of the initial perturbations is relatively modest, and the associated forecast temperature pdf is relatively sharp. In the second example, the initial perturbations grow rapidly and the associated forecast temperature pdf is broad. Notice in the second example that the control integration is already very unskilful 3 days into the forecast. The large ensemble spread provides an a priori warning to the user that decisions made from single deterministic forecasts during this period could be extremely misleading.

Although forecasters have traditionally viewed weather prediction as deterministic, a culture change towards probabilistic forecasting is in progress. On the other hand, it is still necessary to demonstrate to users of weather forecasts that reliable probability forecasts provide greater value than imperfect deterministic forecasts. Such a demonstration will be given in section 7 below.

6 Verifying probability forecasts

As discussed, the output from an ensemble forecast can be used to construct a probabilistic prediction. In this section, we discuss a basic measures of skill for assessing a probability forecast: the Brier Score. This measure is based on the skill of probabilistic forecasts of a binary event E, as discussed in section 5 above. For example E could be: temperatures will fall below 0C in three days time; average rainfall for the next three months will be at least one standard deviation below normal; seasonal-mean rainfall will be below average and temperature above average, and so on.

Consider an event E which, for a particular ensemble forecast, occurs a fraction p of times within the ensemble. If E actually occurred then let $v = 1$, otherwise $v = 0$. Repeat this over a sample of N different ensemble forecasts, so that p_i is the probability of E in the ith ensemble forecast and $v_i = 1$ or $v_i = 0$, depending on whether E occurred or not in the ith verification $(i = 1, 2 \ldots N)$.

The Brier score (Wilks, 1995) is defined by

$$b = \frac{1}{N} \sum_{i=1}^{N} (p_i - v_i)^2, \; 0 \leq p_i \leq 1, \; v_i \epsilon \{0, 1\} \tag{25}$$

From its definition $0 \leq b \leq 1$, equalling zero only in the ideal limit of a perfect deterministic forecast. For a large enough sample, the Brier score

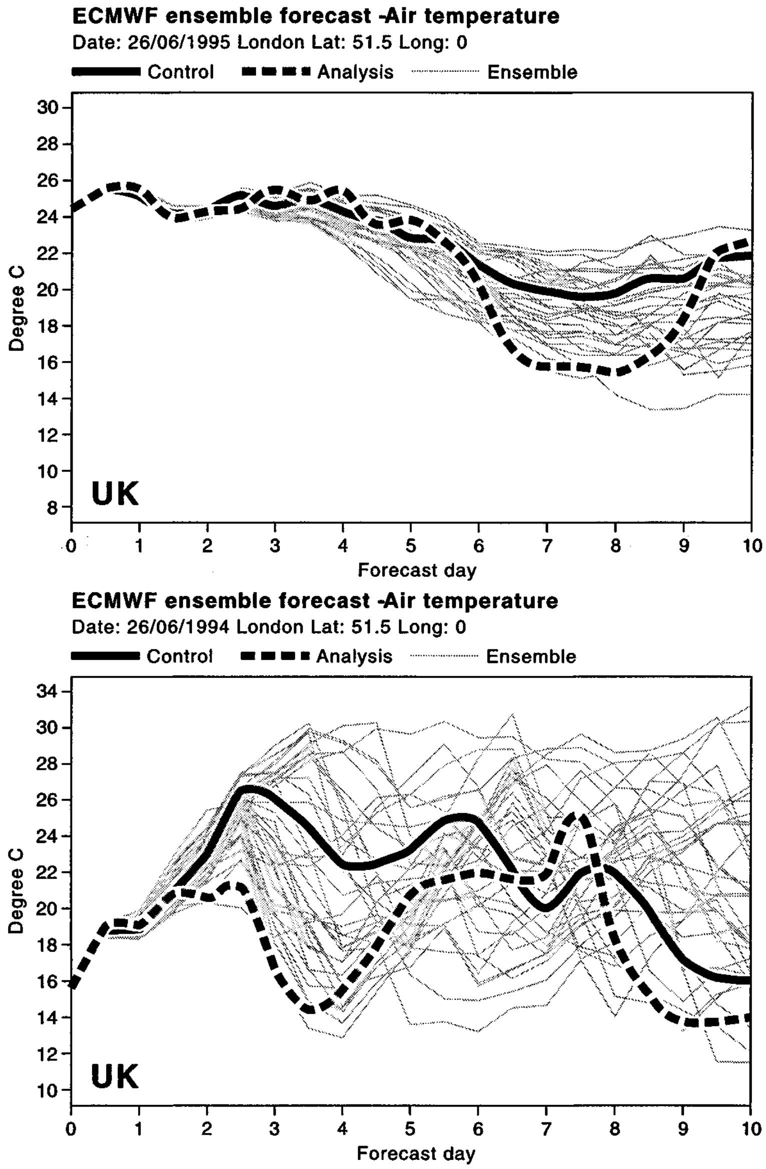

Figure 6: Time series of ensemble forecast integrations for surface temperature over London from starting conditions exactly 1 year apart. The unperturbed control forecast (heavy solid) and the verifying analysis (heavy dashed) are also shown. Top: a relatively predictable period. Bottom: an unpredictable period. (In these examples, the ensembles comprised 32 perturbed forecasts and the model did not include the stochastic representation as described in equation 24.)

can be written as

$$b = \int_0^1 [p-1]^2 o(p)\rho_{ens}(p)dp + \int_0^1 p^2[1-o(p)]\rho_{ens}(p)dp \tag{26}$$

where $\rho_{ens}(p)dp$ is the relative frequency that E was forecast with probability between p and $p+dp$, and $o(p)$ gives the proportion of such cases when E actually occurred. To see the relationship between (25) and (26) note that $\int_0^1 [p-1]^2 o(p)\rho_{ens}(p)dp$ is the Brier score for ensembles where E actually occurred, and $\int_0^1 [p-0]^2(1-o(p))\rho_{ens}(p)dp$ is the Brier score for ensembles where E did not occur.

Simple algebra on (26) gives Murphy's (1973) decomposition

$$b = b_{rel} - b_{res} + b_{unc} \tag{27}$$

of the Brier score, where

$$b_{rel} = \int_0^1 [p - o(p)]^2 \rho_{ens}(p)dp \tag{28}$$

is the reliability component,

$$b_{res} = \int_0^1 [\bar{o} - o(p)]^2 \rho_{ens}(p)dp \tag{29}$$

is the resolution component

$$b_{unc} = \bar{o}[1-\bar{o}] \tag{30}$$

is the uncertainty component, and

$$\bar{o} = \int_0^1 o(p)\rho_{ens}(p)dp \tag{31}$$

is the (sample) climatological frequency of E.

A reliability diagram (Wilks, 1995) is one in which $o(p)$ is plotted against p for some finite binning of width δp. In a perfectly reliable system $o(p) = p$ and the graph is a straight line oriented at 45° to the axes, and $b_{rel} = 0$. Reliability measures the mean square distance of the graph of $o(p)$ to the diagonal line.

Resolution measures the mean square distance of the graph of $o(p)$ to the sample climate horizontal line. A system with relatively high b_{res} is one where the dispersion of $o(p)$ about $\bar{o}$ is as large as possible. Conversely, a forecast system has no resolution when, for all forecast probabilities, the event verifies a fraction $o(p) = \bar{o}$ times.

The term b_{unc} on the right-hand side of (27)ranges from 0 to 0.25. If E was either so common, or so rare, that it either always occurred or never occurred within the sample of years studied, then $b_{unc} = 0$; conversely if E occurred 50% of the time within the sample, then $b_{unc} = .25$. Uncertainty is a function of the climatological frequency of E, and is not dependent on the forecasting system itself. It can be shown that the resolution of a perfect deterministic system is equal to the uncertainty.

When assessing the skill of a forecast system, it is often desirable to compare it with the skill of a forecast where the climatological probability $\bar{o}$ is always predicted (so $\rho_{ens}(p) = \delta(p - \bar{o})$). The Brier score of such a climatological forecast is $b_{cli} = b_{unc}$ (using the sample climate), since, for such a climatological forecast $b_{rel} = b_{res} = 0$. In terms of this, the Brier skill score, B, of a given forecast system is defined by

$$B = 1 - \frac{b}{b_{cli}}. \tag{32}$$

$B \leq 0$ for a forecast no better than climatology, and $B = 1$ for a perfect deterministic forecast.

Skill-score definitions can similarly be given for reliability and resolution, ie

$$\begin{aligned} B_{rel} &= 1 - b_{rel}/b_{cli} && (33) \\ B_{res} &= b_{res}/b_{cli} && (34) \end{aligned}$$

For a perfect deterministic forecast system, $B_{rel} = B_{res} = 1$. Hence, from equations 27 and 32

$$B = B_{res} + B_{rel} - 1 \tag{35}$$

Fig 7 shows two examples of reliability diagrams for the ECMWF EPS taken over all day-6 forecasts from December 1998- February 1999 over Europe (cf Fig 5). The events are $E_{>4}$, $E_{>8}$:- lower tropospheric temperature being at least 4C, 8C greater than normal. The Brier score, Brier skill score, and Murphy decomposition are shown on the figure.

The reliability skill score B_{rel} is extremely high for both events. However, the reliability diagrams indicate some overconfidence in the forecasts.

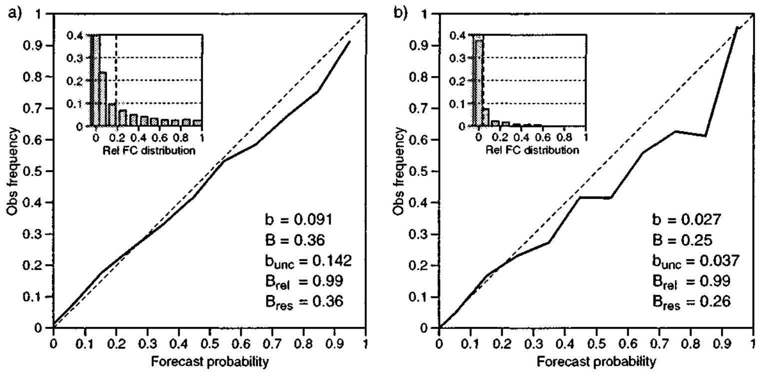

Figure 7: Reliability diagram and related Brier score skill score and Murphy decomposition for the events: a) 850hPa temperature is at least 4K above normal and b) at least 8K above normal, based on 6-day forecasts over Europe from the 50-member ECMWF ensemble prediction system from December 1998-February 1999. Also shown is the pdf $\rho_{ens}(p)$ for the event in question.

For example, on those occasions where $E_{>4}$ was forecast with a probability between 80% and 90% of occasions, the event only verified about 72% of the time. However, it should be remembered that the integrand in equation 28 is weighted by the pdf $\rho_{ens}(p)$, shown next to each in each reliability diagram. In both cases, forecasts where $p > 0.4$ are relatively rare and hence contribute little to B_{rel}.

To see why probability forecasts of $E_{>4}$ have higher Brier skill scores than probability forecasts of $E_{>8}$, consider equation 35. From Fig 7, whilst B_{rel} is the same for both events, B_{res} is larger for $E_{>4}$ than for $E_{>8}$. This can be seen by comparing the histograms of $\rho_{ens}(p)$ in Fig 7 which are more highly peaked for $E_{>8}$ than for $E_{>4}$; there is less dispersion of the probability forecasts of the more extreme event about the climatological frequency of the event, than the equivalent probability forecasts of the more moderate event. This is hardly surprising; the more extreme event $E_{>8}$ is relatively rare (its climatological frequency is ~ 0.04) and most of the time is forecast with probabilities which almost always lie in the first probability category ($0 \leq \delta p \leq 0.1$). In order to increase the Brier score of this relatively extreme event, one would need to increase the ensemble size so that finer probability categories can be reliably defined. (For example, suppose an extreme event has a climatological probability of occurrence of p_{rare}. Let us suppose that we want to be able to forecast probabilities of this event which can discriminate between probability categories with a band width comparable with this climatological frequency, then the ensemble size S_{ens} should be$\gg 1/p_{rare}$.) With finer probability categories, the resolution component of the Brier score can be expected to increase. Providing reliability is not compromised, this will lead to higher overall skill scores.

However, this raises a fundamental dilemma in ensemble forecasting given current computer resources. It would be meaningless to increase ensemble size by degrading the model (eg in terms of 'physical' resolution) making it cheaper to run, if by doing so it could no longer simulate extreme weather events. Optimising computer resources so that on the one hand, ensemble sizes are sufficiently large to give reliable probability forecasts of extreme but rare events, and on the other hand that the basic model has sufficient complexity to be able to simulate such events, is a very difficult balance to define

7 The economic value of probability forecasts

Although B provides an objective measure of skill for ensemble forecasts, they do not determine measures of usefulness for seasonal forecasts. In an attempt to define this notion objectively, we consider here a simple decision model (Murphy, 1977; Katz and Murphy, 1997) whose inputs are probabilistic forecast information and whose output is potential economic value.

Consider a potential forecast user who can take some specific precautionary action depending on the likelihood that E will occur. Let us take some simple examples relevant to seasonal forecasting. If the coming winter is mild (E:- seasonal-mean temperature above normal), then overwintering crops may be destroyed by aphid growth. A farmer can take precautionary action by spraying. If the growing season is particularly dry (E:- seasonal-mean rainfall at least one standard deviation below normal), then crops may be destroyed by drought. A farmer can take precautionary action by planting drought-resistant crops. In both cases taking precautionary action incurs a cost C irrespective of whether or not E occurs (cost of spraying, or cost associated with reduced yield and possibly with more expensive seed). However, if E occurs and no precautionary action has been taken, then a loss L is incurred (crop failure).

This simple 'cost-loss' analysis is also applicable to much shorter range forecast problems (Richardson, 1999). For example, if the weather event was the occurrence of freezing conditions leading to ice on roads, and the precautionary action was to salt the roads, then C would correspond to the cost of salting, and L would be associated with the increase in road accidents, traffic delays etc.

In general, the expense associated with each combination of action/inaction and occurrence/non-occurrence of E is given in the decision-model contingency matrix

		Occurs	
		No	Yes
Take Action	No	0	L
	Yes	C	C

It is presumed that the decision maker wishes wishes to maximise profits, or at least minimise overall expenses.

If only the climatological frequency $\bar{o}$ of E is known, there are two basic options: either always or never take precautionary action. Always taking action incurs a cost C on each occasion, whilst never taking action incurs a loss L only on the proportion $\bar{o}$ of occasions when E occurs, giving an expense $\bar{o}L$.

If forecast data used in section 6 above were used by a hypothetical decision maker, would his/her expenses would be reduced beyond what could be achieved using $\bar{o}$ alone? Consider first a deterministic forecast system with characteristics described by the forecast-model contingency matrix

		Occurs	
		No	Yes
Forecast	No	α	β
	Yes	γ	δ

The user's expected mean expense M per unit loss is

$$M = \frac{\beta L + (\gamma + \delta) C}{L} \tag{36}$$

This can be written in terms of the hit-rate H and the false-alarm F defined as

$$\begin{aligned} H &= \delta/(\beta+\delta) \\ F &= \gamma/(\alpha+\gamma). \end{aligned} \tag{37}$$

so that

$$M = F\frac{C}{L}(1-\bar{o}) - H\bar{o}(1-\frac{C}{L}) + \bar{o} \tag{38}$$

For a perfect deterministic forecast $H = 1$, $F = 0$, hence

$$M_{per} = \bar{o}\frac{C}{L} \tag{39}$$

To calculate the mean expense per unit loss knowing only $\bar{o}$, suppose first the decision maker always protects, then $M = C/L$. Conversely, if the decision maker never protects then $M = \bar{o}$. Hence if the decision maker knows only $\bar{o}$, M can be minimised by either always or never taking precautionary action, depending on whether $C/L < \bar{o}$, or $C/L > \bar{o}$ respectively. The mean expense per unit loss associated with a knowledge of climatology only is therefore

$$M_{cli} = \min(\frac{C}{L}, \bar{o}). \tag{40}$$

The value V of forecast information is defined as a measure of the reduction in M over M_{cli}, normalised by the maximum possible reduction associated with a perfect deterministic forecast, ie

$$V = \frac{M_{cli} - M}{M_{cli} - M_{per}} \tag{41}$$

For a forecast system which is no better than climate, $V = 0$; for a perfect deterministic forecast system $V = 1$.

An ensemble forecast gives hit and false-alarm rates $H = H(p_t)$, $F = F(p_t)$, as a function of probability thresholds p_t. Hence V is defined for each p_t, ie $V = V(p_t)$. Using (38), (39) and (40)

$$V(p_t) = \frac{\min(\frac{C}{L}, \bar{o}) - F(p_t)\frac{C}{L}(1-\bar{o}) + H(p_t)\bar{o}(1-\frac{C}{L}) - \bar{o}}{\min(\frac{C}{L}, \bar{o}) - \bar{o}\frac{C}{L}}. \tag{42}$$

For given C/L and event E, the optimal value is

$$V_{opt} = \max_{p_t} V(p_t). \tag{43}$$

Fig 8 shows examples of optimal value as a function of user cost/loss ratio for the ECMWF day 6 ensemble weather prediction system and the event $E_{>4}$ (as in Fig 7). The solid curve is the optimal value for the ensemble system, the dashed curve shows value for a single deterministic forecast (the unperturbed 'control' integration in the ensemble). Peak value tends to occur for $C/L \sim \bar{o}$; for such users, it makes little difference to the climatological expense M_{cli} whether they always protect, or never protect. The figure also illustrates the fact that the ensemble forecast has greater 'value' than a single deterministic forecast. For some cost/loss ratios (eg $C/L > 0.6$), the deterministic forecast has no value, whereas the ensemble forecast does have value. The reason for this can be understood in terms of the fact that for a probabilistic forecast, different users (with different C/L) would take precautionary action for different forecast probability thresholds. A user who would suffer a catastrophic loss ($C/L << 1$) if E occurred, would take precautionary action even when a small probability of E was forecast. A user for whom precautionary action was expensive in comparison with any loss ($C/L \sim 1$) would take precautionary action only when a relatively large probability of E was forecast. The result demonstrates the value of a reliable probability forecast.

8 Concluding remarks

Our climate is a complex nonlinear dynamical system, with spatial variability on scales ranging from individual clouds to global circulations in the atmosphere and oceans, and temporal variability ranging from hours to millenia. Weather and climate scientists interact with society through the latter's demands for accurate and detailed environmental forecasts: of weather, of El Niño and its impact on global rainfall patterns, and of man's effect on climate. The complexity of our climate system implies that quantitative predictions can only be made with comprehensive numerical models which encode the relevant laws of dynamics, thermodynamics and chemistry for a multi-constituent multi-phase fluid. Typically such models comprise some millions of scalar equations, describing the interaction of circulations on scales ranging from tens of kilometres to tens of thousands of kilometres; from the ocean depth to the upper stratosphere. These equations can only be solved on the world's largest supercomputers.

However, a fundamental question that needs to be addressed, both by

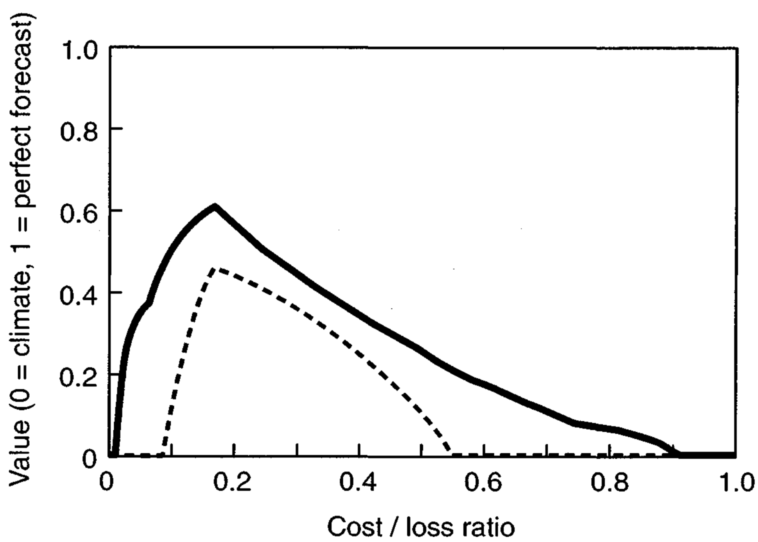

Figure 8: Potential economic value of the ECMWF ensemble prediction system as a function of user cost/loss ratio of day 6 weather forecasts over Europe (for the period December 1998 -February 1999) for the event $E_{<-4}$:-850hPa temperature at least 4C below normal. Solid: value of the ECMWF ensemble prediction system. Dashed: value of a single deterministic forecast (the unperturbed 'control' forecast of the EPS system). From Richardson (1999).

producers and users of such forecasts, is the extent to which weather and climate are predictable; after all, much of chaos theory developed from an attempt to demonstrate the limited predictability of atmospheric variations. In the past, the topic of predictability has been a somewhat theoretical one, somewhat removed from the practicalities of prediction. A famous climatologist remarked some years ago: 'Predictability is to prediction as romance is to sex!'. However, the remark is perhaps not so apposite today; the science of predictability of weather and climate in now an integral part of the practical prediction problem - the two cannot be separated. The predictability problem can be formulated eg through a Liouville equation; however, in practice, estimates of predictability are made from multiple (ensemble) forecasts of comprehensive weather and climate prediction models. The individual members of the ensemble differ by small perturbations to quantities that are not well known. The predictability of weather is largely determined by uncertainty in a forecast's starting conditions, though the effects of uncertainty in representing computationally the equations that govern climate (for example, how to represent the effects of convective instabilities in a model that cannot resolve individual clouds) are not negligible.

Chaos theory implies that all such environmental forecasts must be expressed probabilistically; the laws of physics dictate that society cannot expect arbitrarily accurate weather and climate forecasts. These probability forecasts quantify uncertainty in weather and climate prediction. The duty of the meteorologist is to strive to estimate reliable probabilities; not to disseminate predictions to society with a precision that cannot be justified scientifically. Examples were shown that, in practice, the economic value of a reliable probability forecast (produced from an ensemble prediction system) exceeds the value of a single deterministic forecast with uncertain accuracy.

However, ensemble forecasting poses a fundamental dilemma given current computing resources. To be able to simulate extreme events requires models with considerable complexity and resolution. On the other hand, estimating reliably changes to the probability distributions of extreme and hence relatively rare events, requires large ensembles. One thing is certain; the more the need to provide reliable forecasts of uncertainty in our predictions of weather and climate, the more the demand for computer power exceeds availability, notwithstanding the unrelenting advance in computer technology. Indeed the need for quantitative predictions of uncertainty in weather and climate science is a relevant consideration in the design of future generations of supercomputers; ensemble prediction is a perfect application for parallel computing! There can be little doubt that the benefits to society of reliable regional probability forecasts of extreme weather events, seasonal variability and anthropogenic climate change justify such technological developments.

Acknowledgements

I would like to thank colleagues at ECMWF, Drs J. Barkmeijer, R. Buizza, F. Lalaurette, K.Puri and D. Richardson for innumerable discussions and for help in producing many of the figures in this paper.

References

Anderson, J.L,, 1997: Impact of dynamical constraints on the selection of initial conditions for ensemble predictions: low-order perfect model results. Mon.Wea.Rev., 125, 2969-2983.

Andersson, E and M. Fisher, 1999: Background errors for observed quantities and their propagation in time. ECMWF workshop on Diagnosing Data Assimilation Systems. ECMWF Shinfield Park, Reading, RG2 9AX, UK.

Barkmeijer, J., Gijzen, Van M., and F. Bouttier, 1998: Singular vectors and estimates of the analysis error covariance metric. Q.J.R.Meteorol.Soc., 124, 1695-1713.

Betts, A.K., and M.J.Miller, 1986: A new convective adjustment scheme. Part II:single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. Q.J.Meteorol.Soc., 112, 693-709.

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

Buizza, R., M.J.Miller and T.N.Palmer, 1998: Stochastic simulation of model uncertainties in the ECMWF ensemble prediction system. Q.J.Meteorol.Soc., to appear.

Buizza, R., T. Petroliagis, T.N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons and N.Wedi, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. Q.J.R. Met.Soc., 124, 1935-1960.

Buizza, R., A.Hollingsworth, A.Lalaurette and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. Weather and Forecasting, 14, 168-189.

Courtier, P., Thèpaut, J-N., and A.Hollingsworth, 1994: A strategy for operational implementationa of 4D-Var using an incremental approach. Q.J.R. Meteorol.Soc., 120, 1367- 1387.

Courtier, P., Andersson, E., Heckley, W., Pailleux, P., Vasilevic, D., Hamrud, M., Hollingsworth, A. Rabier, F. and Fisher, M., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. Q.J.R.Meteorol.Soc., 124, 1783-1808.

Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press. Cambridge, UK. pp457.

Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: theory. Mon.Wea.Rev., 122, 703-713.

Ehrendorfer, M., 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II:Applications. Mon.Wea.Rev., 122, 714-728.

Ehrendorfer, M., 1999: Kalman filtering and atmospheric predictability. In: Proceedings of the ECMWF workshop on diagnosis of data assimilation systems. November 1988. ECMWF, Shinfield Park, Reading RG2 9AX.

Epstein, E.S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic ocean model. J. Geophys. Res., 99 (C5), 10 143- 10162.

Fisher, M., 1998: Development of a simplified Kalman filter. ECMWF Technical Memorandum No 260. ECMWF Shinfield Park, Reading, RG2 9AX, UK.

Fisher, M. and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Technical Memorandum No 220. ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Gelaro, R., R.Buizza, T.N.Palmer and E.Klinker, 1998: Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. J.Atmos.Sci., 55, 1012-1037.

Gleeson, T.A., 1966: A causal relation for probabilities in synoptic meteorology. J. App. Met., 5, 365-368.

Hamill, T.M., C. Snyder and R.E.Morse, 1999: A comparison of probabilistic forecasts from bred, singular vector and perturbed observation ensembles. Mon. Wea. Rev., submitted.

Harrison, M., T.N.Palmer, D.S.Richardson and R. Buizza, 1998: Analysis and model dependencies in medium-range ensembles: two transplant cast studies. Q.J.R.Meteor.Soc., 125, 2487-2515.

Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. Tellus, 28, 473- 485.

Houtekamer, P.L., L.Lefaivre, J. Derome, H.Richie, and H.L.Mitchell, 1996. A system simulation approach to ensemble prediction. Mon.Wea.Rev., 124, 1225-1242.

Katz, R. W., and Murphy, A. H., 1997. Forecast value: prototype decision-making models. In Economic value of weather and climate forecasts, Katz, R. W., and Murphy, A. H., Eds.. Cambridge University Press, 222 pp

Kimoto, M., and M. Ghil, 1993: Multiple flow regimes in the northern hemisphere winter. Part I: Methodology and hemispheric regimes. J.Atmos.Sci., 50, 2625-2643.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. J.Atmos.Sci., 42, 433-471.

Mason, P.J. and D.J.Thompson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. J.Fluid. Mech., 242, 51-78.

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

Moncrieff, M.W., 1992: Organised convective systems: archetypal dynamical models, mass and momentum flux theory and parametrization. Q.J.R.Meteorol.Soc., 118, 819-850.

Moss, F. and P.V.E. McClintock, Eds., 1989: Noise in Nonlinear Dynamical Systems. Vol 1, Theory of Continuous Fokker-Planck Systems. Cambridge University Press, 353pp.

Mureau, R., F.Molteni and T.N.Palmer, 1993: Ensemble prediction using dynamically-conditioned perturbations. Q.J.R.Meteorol.Soc., 119, 299-323.

Murphy, A.H., 1973: A new vector partition of the probability score. J.Appl. Meteor., 12, 595-600.

Murphy, A., H., 1977. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. Mon. Wea. Rev., 105, 803-816.

Palmer, T.N., F. Molteni, R. Mureau, R. Buizza, P.Chapelet and J. Tribbia, 1993: Ensemble Prediction. 1992 ECMWF Seminar Proceedings.

Palmer, T.N., 1998: Singular vectors, metrics and adaptive observations. J.Atmos.Sci., 55, 633-653.

Palmer, T.N., C. Brankovic and D.S. Richardson, 1999: A probability and decision model analysis of PROVOST seasonal multi-model ensemble integrations. Q.J.R.Met.Soc., to appear

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. Mon. Wea. Rev., 120, 1747-1763.

Reinhold, B.B. and R.T.Pierrehumbert, 1982: Dynamics of weather regimes: quasi- stationary waves and blocking. Mon.Wea.Rev., 110, 1105-1145.

Richardson, D.S., 1998: Skill and economic value of the ECMWF ensemble prediction system. Q.J.R.Met.Soc., to appear.

Strang, G., 1986: Introduction to applied mathematics. Wellesley-Cambridge Press. pp758.

Tennekes, H., 1991: Karl Popper and the accountability of numerical forecasting. In New developments in Predictability. ECMWF Workshop Proceedings, ECMWF, Shinfield Park, Reading UK, 1991.

Thompson, P.D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. Tellus, 9, 275-295.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. Bull.Am.Met.Soc., 74, 2317-2330.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. Mon.Wea.Rev., 12, 3297-3319.

Trevisan, A. , Pancotti, F. and F. Molteni, 1999: Ensemble prediction in a model with flow regimes. Q.J.R.Met.Soc., submitted.

Webster, P.J., V.O. Magana, T.N.Palmer, J.Shukla, R.A.Thomas, M.Yanai and T.Yasunari, 1998: Monsoons: Processes, predictability and the prospects for prediction. J.Geophys.Res., 103, 14451-14510.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press. 467pp