Review-article

Meteorol. Zeitschrift, N.F. 6, 147–183 (August 1997)

Predicting the uncertainty of numerical weather forecasts: a review

MARTIN EHRENDORFER, University of Vienna

Summary. Weather forecasts produced with numerical weather prediction (NWP) models of the atmosphere possess intrinsic uncertainty. This uncertainty is caused through both errors in the specification of the initial state of the model, as well as errors in the model formulation itself. In the process of NWP, the consideration of both error sources is important, because the nature of atmospheric dynamics is such that it acts to increase errors originating from either error source. In addition to this overall error-growth effect, forecast error possesses considerable day-to-day variability depending, among other things, on the flow regime.

The quantitative and reliable assessment (i. e., prediction) of the uncertainty of weather forecasts is important, both for scientific and economic reasons. Scientifically, quantification of atmospheric predictability asks for the rate at which two initially close trajectories diverge (on average) for given atmospheric dynamics. Such estimates place upper bounds on time horizons over which useful forecasts may be expected. Economically, a reliable estimate of the uncertainty of a particular forecast will lead to increased credibility and utility of weather forecasts.

The description and discussion of strategies and methods to predict the uncertainty of weather forecasts produced with NWP models are the subject of this article. The limited predictability of atmospheric flows, considered here on time scales of days, as it results essentially from the intrinsic error growth in the atmosphere is briefly discussed. The Liouville equation as the theoretical concept for dealing with the prediction problem of forecast uncertainty is described, as it governs the time-evolution of the probability density function (pdf) of the NWP model state. Related concepts more readily applicable in operational contexts are reviewed. Among these concepts are stochastic-dynamic prediction, the lagged-average forecasting technique, and the Monte Carlo approach.

Particular attention is given to the description of methodology and results from currently operational efforts at major forecasting centers directed towards the prediction of forecast uncertainty through multiple (i. e., an ensemble of) NWP model integrations started from different initial states. Such time-evolved ensembles provide partial information about the time-evolved pdf. These *ensemble prediction systems* at the European Centre for Medium-Range Weather Forecasts, as well as the National Centers for Environmental Prediction are discussed in some detail (e. g., with respect to the selection of the initial states of individual ensemble members). Results presently obtained with ensemble prediction systems are highly promising, although various questions related to, for example, the modeling error source, the validation of products from ensemble prediction systems, as well as to the methodology for the selection of initial states remain to be answered. This review is concluded by mentioning briefly similar efforts at other operational NWP centers, as well as applications of the methodology used in ensemble prediction in related contexts such as stability analysis and data assimilation.

Vorhersage der Unsicherheit numerischer Wetterprognosen: eine Übersicht

Zusammenfassung. Wettervorhersagen, die mit numerischen Wetterprognose (NWP) Modellen der Atmosphäre erstellt werden, besitzen inhärente Unsicherheit. Diese Unsicherheit ist bedingt durch Fehler sowohl in der Spezifikation des Anfangszustandes des Modells, als auch in der Modellformulierung selbst. In der NWP ist die Berücksichtigung beider Fehlerquellen von Bedeutung, da aufgrund der Natur der atmosphärischen Dynamik Fehler aus beiden Quellen tendenziell anwachsen. Zusätzlich zu diesem allgemeinen Effekt des Fehlerwachstums besitzt der Vorhersagefehler eine beachtliche Variabilität von Tag zu Tag, die unter anderem von der Strömungssituation abhängt.

Die quantitative und zuverlässige Feststellung (d. h. Vorhersage) der Unsicherheit von Wettervorhersagen ist sowohl aus wissenschaftlichen, wie auch aus ökonomischen Gründen von Bedeutung. Aus wissenschaftlicher Sicht besteht die Quantifizierung der atmosphärischen Vorhersagbarkeit in erster Linie in der Frage nach der Geschwindigkeit, mit welcher zwei anfänglich dicht benachbarte Trajektorien (im Durchschnitt) für gegebene atmosphärische Dynamik divergieren. Ein derartiger Schätzwert führt zur Angabe von Obergrenzen für den Zeithorizont, bis zu dem brauchbare Vorhersagen erwartet werden dürfen. Aus ökonomischer Sicht sollte eine zuverlässige Abschätzung der Unsicherheit einer speziellen Vorhersage zunehmende Glaubwürdigkeit und Verwendbarkeit von Wettervorhersagen zur Folge haben.

Die Beschreibung und Diskussion von Strategien und Methoden zur Vorhersage der Unsicherheit von Wettervorhersagen, die mit NWP Modellen erzeugt werden, sind der Gegenstand dieses Artikels. Die begrenzte Vorhersagbarkeit atmosphärischer Strömungen, betrachtet hier auf einer Zeitskala von einigen Tagen, wird als Resultat des der Atmosphäre inhärenten Fehlerwachstums kurz diskutiert. Die Liouville Gleichung wird als theoretisches Konzept zur Behandlung des Vorhersageproblems der Vorhersageunsicherheit beschrieben, da sie die zeitliche Entwicklung der Wahrscheinlichkeitsdichtefunktion (WDF) des Zustandsvektors des NWP Modells beschreibt. Verwandte Konzepte, die in operationellem Kontext eher anwendbar sind, werden überblicksmäßig diskutiert. Zu diesen Konzepten zählen die stochastisch-dynamische Vorhersage, die Vorhersagetechnik der zeitlich verschobenen Mittelung, sowie der Monte Carlo Zugang.

0941-2948/97/0006-0147 $ 16.65

Spezielle Aufmerksamkeit wird der Beschreibung von Methodik und entsprechenden Ergebnissen gewidmet, welche derzeit im Rahmen von operationellen Programmen an den großen Vorhersagezentren verwendet wird, mit dem Ziel der Prognose der Vorhersageunsicherheit durch mehrfache (d. h. ein Ensemble von) NWP Modellintegrationen, die von unterschiedlichen Anfangszuständen gestartet werden. Derartige zeitlich entwickelte Ensembles liefern teilweise Information über die zeitlich entwikkelte WDF. Diese *Ensemble-Vorhersage-Systeme* am Europäischen Zentrum für Mittelfristige Wettervorhersage, sowie an den National Centers for Environmental Prediction werden im Detail beschrieben (z. B. hinsichtlich der Auswahl der Anfangszustände der einzelnen Ensemblemitglieder). Ergebnisse, die derzeit mit Ensemble-Vorhersage-Systemen erzielt werden, sind vielversprechend, obwohl verschiedene Fragen noch unbeantwortet sind, wie, unter anderem, Fragen im Zusammenhang mit dem Modellfehler, der Validierung von Produkten aus Ensemble-Vorhersage-Systemen, sowie der Methodik zur Auswahl der Anfangszustände. Diese Übersicht schließt mit einer kurzen Erwähnung von ähnlichen Anstrengungen an anderen NWP Zentren, sowie der Anwendung der Methodik aus der Ensemble-Vorhersage in anderem Zusammenhang, wie beispielsweise in Stabilitätsuntersuchungen und der Datenassimilation.

1. Introduction and motivation

The past fifteen years have seen rapid and impressive progress in the area of the operational production of highly accurate and reliable weather forecasts in the short (up to five days) and medium (between six and fifteen days) range by means of numerical weather prediction (NWP) models. This progress is most clearly demonstrated by the continuous improvement of aspects of forecast quality over the last one or two decades. As an example, Fig. 1 shows the increase over the last fifteen years of the length of the time period (in days) over which forecasts of the geopotential height of the 500 hPa pressure surface, made with the global spectral model of the European Centre for Medium-Range Weather Forecasts (ECMWF), are considered to be useful in the sense that a verification score — in this case the anomaly correlation coefficient (see below) — does not fall below a given threshold. As can be seen, this time period was around 5.5 days in the early 1980s, and has extended to somewhere around seven days in the mid 1990s. It needs to be mentioned that, as a one-dimensional verification score, the anomaly correlation coefficient is cleary limited to measuring only one aspect of forecast quality (the accuracy of forecasts of geopotential height in the present context); consequently, other aspects of forecast quality evaluated differently may not necessarily exhibit the same behavior as shown for geopotential height in Fig. 1.

The core of NWP models presently used for producing such forecasts of atmospheric conditions over several days is a suitably discretized version of the laws (approximately) governing the dynamics of atmospheric flow (conservation of momentum, mass, and energy, and an equation of state), augmented by so-called parameterizations of the processes unresolvable on the computational grid of the model (examples of such processes are radiation and convection). The importance of an adequate degree of resolution of the discretization of these equations is reflected through the increase in accuracy in mid 1994 in Fig. 1, at which time the model's resolution was considerably refined. In addition, in order for such models to be useful for prediction purposes, an accurate specification of the initial model state (analysis of the atmospheric state) is required. It is of interest to note in this context that almost all of the improvement in forecast accuracy shown in Fig. 1 is a result of both the improved quality of the forecast model, as well as of the data assimilation procedures used to produce the atmospheric analysis that enters the forecasting process as the initial state, and not so much a result of improvements in the global atmospheric observing system. This finding has been estab-

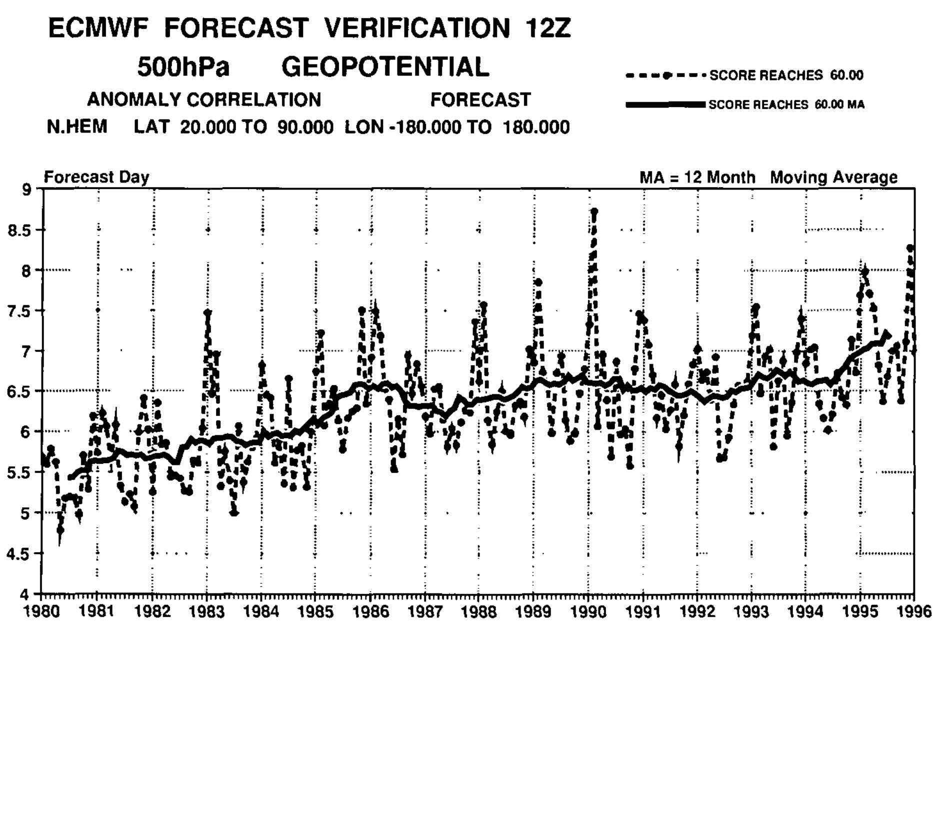


Fig. 1. Lead time (in days) at which the anomaly correlation coefficient for the 500 hPa geopotential ECMWF forecast for the extratropical northern hemisphere falls under the value of 0.6 over the last fifteen years. The dashed curve shows instantaneous values, the solid curve is a 12-month moving average (figure provided by T. PETROLIAGIS, ECMWF).

Abb. 1. Vorhersagezeitpunkt (in Tagen), zu dem der Anomaliekorrelationskoeffizient für die ECMWF Vorhersage des 500 hPa Geopotentials im Gebiet der extratropischen Nordhemisphäre den Wert 0,6 unterschreitet, während der vergangenen 15 Jahre. Die strichlierte Kurve zeigt momentane Werte, die durchgezogene Kurve zeigt einen 12-monatigen gleitenden Mittelwert (Abb. zur Verfügung gestellt durch T. PETROLIAGIS, ECMWF).

lished in course of the ECMWF re-analysis project (e. g., GIBSON et al. 1996). It was found there that the accuracy of forecasts produced with an unchanging forecast model, as well as the same data assimilation procedure remained highly uniform over the time period shown in Fig. 1.

The theoretical foundation for the above-mentioned mathematically and physically based approach to weather prediction was laid by V. BJERKNES (1904) who formulated the problem of weather forecasting in mathematical terms as a general boundary- and initial-value problem. V. BJERKNES (1862–1951) also very succinctly stated what he called "the necessary and sufficient conditions for a rational solution of the forecast problem in meteorology: 1. One has to know with sufficient accuracy the state of the atmosphere at a given time. 2. One has to know with sufficient accuracy the laws according to which one atmospheric state is developing from another" (translated from BJERKNES 1904). Subsequently, L. F. RICHARDSON (1881–1953) proposed to solve the governing hydrodynamical equations in a finite-difference formulation, an idea he actually carried out by hand to produce a numerical forecast (RICHARDSON 1922) that exhibited, however, serious deficiencies (see, e. g., ASHFORD 1985, LYNCH 1992). RICHARDSON's ideas could be carried out in a much more economic and practical way for the first time in the mid 1950s through the advent and use of high-speed computing machines that were first envisaged for meteorological prediction purposes by John VON NEUMANN (1903–1957). These developments led to the first successful numerical weather forecasts (e. g., CHARNEY et al. 1950) and provided demonstration that weather prediction was indeed possible along the lines suggested by V. BJERKNES.

Since then, continuous development of highly sophisticated NWP models has taken place leading to a number of models run at various operational centers. These model developments may all — at least in a broad sense — be summarized as attempts to come closer to the two requirements stated by BJERKNES in 1904. Efforts are concentrating on both more realistic inclusion of governing dynamics and physics (requirement 2), as well as on the most accurate specification of the initial model state possible (requirement 1). Much more complete and detailed accounts of the historic developments leading to the current state-of-the-art methodology in NWP may be found in THOMPSON (1961, 1987a, b) and TRIBBIA and ANTHES (1987) (see also LEWIS 1996). Also, reference is made at this point to the description by RITCHIE et al. (1995) of the NWP model developed and used at ECMWF as a representative for the many today's state-of-the-art NWP models.

In conjunction with the impressive progress made in NWP, as described above, one intriguing feature of the NWP models (and, in turn, to some extent, of the atmosphere) has become clear. This feature is the — sometimes substantial — day-to-day variability of the quality of forecasts. Specifically, it appears that forecasts made with a given NWP model may be highly useful on one occasion (i. e., the a posteriori verification may indicate that the forecast was highly skillful), but rather poor on other occasions (i. e., hardly skillful), for a given lead time. Such variability of skill is considerably less for forecasts in the short range than in the medium or extended range. Clearly, it appears to be highly desirable to know in advance whether a given forecast should be expected to be skillful or not.

An indication of this variability of forecast skill is given in Fig. 2, showing daily verifications during the months of November 1995 (panel a) and December 1995 (panel b) of operational forecasts of the geopotential height of the 500 hPa pressure surface made with the ECMWF model for a lead time of five days. The term "skill" is used here in a generic form to denote a (usually one-dimensional, or scalar) summary measure of the quality of the NWP model forecast; that is, skill is a function of both forecast and observed (or analysed) fields. In the case of Fig. 2 the summary measure used to assess skill is the anomaly correlation coefficient (e. g., MURPHY and EPSTEIN 1989, MURPHY 1997). A forecast with an anomaly correlation of 100 % is perfect, forecasts with an anomaly correlation above 60 % are considered to be useful for weather forecasters in the process of interpreting forecast maps (this threshold of 60 % has been used in Fig. 1 to determine the day plotted there; the usefulness of forecasts with a score better than 60 % has been determined by a subjective evaluation study). It can be seen from Fig. 2 that on most occasions these five-day forecasts are quite skillful, but not so skillful on some occasions (e. g., 17 November 1995). In addition, the variability in skill becomes evident. As an aside, it should be noted that even at this lead time of five days these forecasts present substantial improvements above competing reference forecasts such as climatological or persistence forecasts. Finding ways and strategies that allow one *to specify the skill of a given NWP model forecast a priori* (i. e., at the same time as the forecast is produced), has been one of the major motivations for the numerous studies carried out extensively over the (approximately) last decade that present the major theme of this review. Clearly, this motivation is reinforced through the fact that a priori knowledge of the reliability of a given forecast is accompanied by potentially increased value of the forecast for its end-users, such as persons/institutions involved in decision-making processes with payoff structures dependent on weather conditions (for a recent collection of studies concerned with the economic value of forecasts, see KATZ and MURPHY 1997, MURPHY 1994; see also WILKS and HAMILL 1995).

Viewed slightly differently, the verification scores shown in Fig. 2 indicate that the forecasts are not perfect, but are in some cases very close to perfect, whereas in other cases they are further away from a perfect forecast. In other words, these forecasts are uncertain. One may argue quite generally that it is an intrinsic feature of forecasts (in any field) that they are uncertain. The quantification (or, prediction) of this uncertainty of NWP model forecasts, as well as the discussion of results from studies dealing with this quantification are the subject of this review. More specifically, this review will concentrate on methodology, developments, and studies aimed at the quantification of the

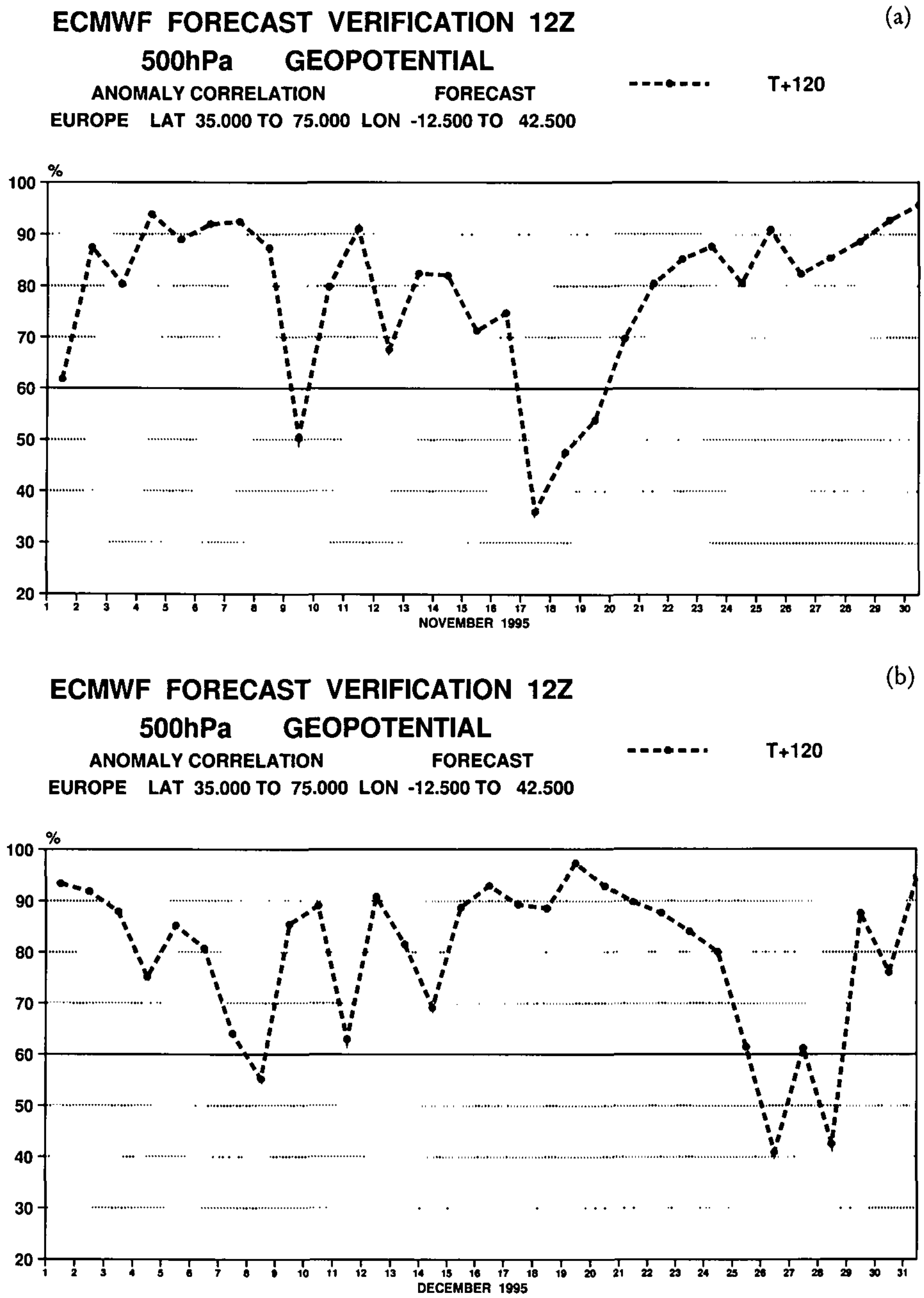


Fig. 2. Anomaly correlation coefficient for the 500 hPa geopotential ECMWF forecast for Europe at a lead time of 5 days for (a) November 1995, (b) December 1995 (figure provided by T. PETROLIAGIS, ECMWF).

Abb. 2. Anomaliekorrelationskoeffizient für die ECMWF Vorhersage des 500 hPa Geopotentials für Europa zu einem Vorhersagezeitpunkt von 5 Tagen für (a) November 1995, (b) Dezember 1995 (Abb. zur Verfügung gestellt durch T. PETROLIAGIS, ECMWF).

uncertainty of forecasts made through dynamical NWP models. This importance of such quantification has been emphasized repeatedly during the last decade (e. g., TENNEKES et al. 1987, KALNAY and DALCHER 1987, PALMER and TIBALDI 1988, TRIBBIA 1991), even though it has been recognized and emphasized already at the beginning of this century (see below). This review will cover parts of the historic developments, and also try to present the theoretical basis for some of the methodology in a proper framework. While this scope may seem severely limited at this point (in view of the areas not discussed; see below), it will become clear that the number of such studies is numerous and has increased considerably over the last decade or so. This increase in number is a reflection of the fact that the area is growing rapidly with new methodologies being presented and various aspects of different questions being posed and studied. It is also a reflection of the recognition that in conjunction with improved deterministic forecasting, to some extent inadequate attention has been given to some aspects of the statistical nature of weather forecasting, such as, for example, the uncertainty quantification of NWP forecasts. Finally, the increasing number of studies in this field also reflects the more recent substantial increases in computer power that allow one to pursue ensemble integration strategies aimed at the assessment of forecast uncertainty. Under this perspective of a rapidly developing, not entirely coherent, field, some points may inadvertently receive inadequate attention here.

The main restriction of this review is that is does not deal with the vast amount of literature describing the quantification of forecast uncertainty (i) before the advent of NWP models, and/or (ii) by purely statistical techniques that utilize NWP output (e. g., the model output statistics technique), and (iii) by subjective techniques (based partly on guidance produced by statistical techniques and NWP models). This restriction was found necessary in order to limit the length of this review, and should not be interpreted to mean that these techniques are of less importance. In fact, the recognition that meteorological predictions possess inherent uncertainty that must/should be quantified was recognized very early, and even long before numerical forecasts were available (e. g., COOKE 1906, VON MYRBACH 1913). For an overview of such studies not discussed here, reference is made to, for example, MURPHY and WINKLER (1984) (see also EHRENDORFER 1989).

The structure of this paper is as follows. A brief review of atmospheric predictability will be given in the next section. The theoretical basis and concepts for the prediction of the uncertainty of NWP model forecasts are discussed in section 3, together with a discussion of various problems that are encountered in the application of these concepts. Various proposals and attempts to exploit these concepts in quasi-operational and operational environments are presented in sections 4 and 5; section 4 is devoted to a discussion of selected developments in the pre-operational phase (mid to late 1980s); in section 5 procedures currently operational at various NWP centers are described. In this latter section, as well as in an accompanying appendix, the rationales behind certain methodologies, are formulated in some detail. The aims that seem to be desirable to achieve are discussed, and views are offered on whether certain aims are achievable through certain procedures. In section 6 various offsprings of the methodology used in uncertainty prediction are briefly discussed, such as the general use of adjoint models in data assimilation, predictability studies, and in a generalization of atmospheric stability analysis. The paper is concluded with a brief discussion of main results that were emphasized in the review, as well as of various open questions.

2. Limits on atmospheric predictability

To explain in detail the day-to-day variations in the degree of the usefulness (skill) of weather forecasts, as shown by example in Fig. 2, appears to be difficult. However, the two general causes that are at the root of this behavior leading to non-perfect forecasts are easily stated. In particular, the limitations on the accuracy of weather forecasts made with dynamical models of the atmosphere can be related to the influence of two distinct error sources (e. g., LORENZ 1990). Namely, the quality of forecasts is limited by the influence of (i) errors in the specification of the initial state of the model (initial state or data error source), as well as (ii) errors in the formulation of the model itself (modeling error source). The specification of boundary conditions necessary within NWP models (e. g., lateral boundary conditions or conditions at the lower boundary of the model's atmosphere) plays a somewhat intermediate role here. For today's state-of-the-art NWP models it is currently believed that both error sources contribute approximately equally to the degradation of forecast quality on time scales of the order of two to five days (see, e. g., TRIBBIA and BAUMHEFNER 1988a), with the initial state error source being of primary importance at shorter lead times, whereas the modeling error source is more dominant at longer lead times. However, due to the interaction of the effects of both error sources, direct assertion of this belief is quite difficult to obtain and requires both very accurate observations and an extensive number of model integrations. It seems clear, however, that large geographic regions without high-quality atmospheric observations (e. g., oceans) necessarily lead to a more inaccurate specification of the initial model state and, consequently, to increases of forecast errors. In this situation, the improvement of forecast quality through improved initial conditions is potentially quite significant (e. g., RABIER et al. 1996; see also last section). Similarly, model formulations near steep and complex terrain or in cases of moist convection are likely inadequate, at least to some extent, leading also to increases in forecast errors.

The fundamental importance of the role of these two error sources (see also the comment by V. BJERKNES quoted in the previous section) becomes clear when they are considered in conjunction with the now commonly recognized property of NWP models (as well as of the atmosphere) that predicted (future) states are extremely sensitive to small changes in present model (or, atmospheric) states. In other words, experience with NWP models has indicated that two initially slightly different states — each evolving according to the same physical laws — may, and, in general, do, over time develop into states no more similar than two randomly chosen observed states of the atmosphere (e. g., THOMPSON 1957; LORENZ 1982, 1993), with error growth that is exponential at least in the early stages before saturation is reached (NICOLIS et al. 1995). It must be made clear that this inherent error growth is not an artifact of NWP models, but is a consequence of nonlinearity and instability of atmospheric dynamics (PALMER and TIBALDI 1988). It is this internal error-growth mechanism and the associated physical processes responsible for error growth, as well as the average rate at which initially small errors are amplifying that have been the principal subject of predictability studies in recent years (LORENZ 1982, LEITH 1978).

In studying the predictability of nonlinear dynamical systems, LORENZ (1963) provided clear evidence in support of the hypothesis that the atmosphere might indeed possess the property of diverging solutions, and, as such, be of limited predictability. In this seminal work, he investigated a system of three ordinary *nonlinear* differential equations that resembled the atmosphere only remotely, but exhibited *chaotic* behavior; that is, these equations possess (for certain choices of the free parameters) aperiodic solutions with sensitive dependence on their initial conditions (see also, TRIBBIA and ANTHES 1987). In essence then, LORENZ (1963) demonstrated, for the first time, that some simple *deterministic* systems are only predictable for a *finite* time that is dependent on the accuracy with which the initial condition is specified. Several years before then, THOMPSON (1957) demonstrated, in a similar effort, that substantially different predictions resulted from minor changes in initial model states that in turn were related to missing observations. For more detailed recollections concerning these early stages of atmospheric predictability research, reference is made to THOMPSON (1987a), and LORENZ (1993, 1996). The highly significant importance of this early work has been acknowledged repeatedly and widely. For example, TRIBBIA and ANTHES (1987, p. 497) state: "The ramifications of LORENZ's work have had tremendous impact on the fields of applied and pure mathematics, theoretical physics, turbulence theory, mathematical biology, and the philosophy of determinism". Or, after quoting from the abstract of LORENZ (1963), the mathematician STEWART (1989, p. 133) describes the impact of this work on mathematics with the following words: "When I read these words I get a prickling at the back of my neck and my hair stands on end. *He knew! Twenty-four years ago, he knew!* And when I look more closely, I'm even more impressed. In a mere twelve pages Lorenz anticipated several major ideas of nonlinear dynamics, before it became fashionable, before anyone else had realized that new and baffling phenomena such as chaos existed" (italics from STEWART).

The sensitive dependence of the equations governing NWP models on the initial conditions leads yet to another implication. Obviously, the modeling error source has the effect that during the time integration of an NWP model the model state is contaminated with errors, in a systematic and/or random way (for a recent discussion of model errors in NWP models reference is made to TRIBBIA and BAUMHEFNER 1988a, SAHA 1992, BOER 1993, DEE 1995, KANAMITSU und SAHA 1996; see also section 3). These model errors are amplified in much the same way as are the errors in the initial condition (see above). Among the most prominent candidates for contributing to the modeling error source is the combined effect of inaccuracies in the description (parameterization) of processes occurring on unresolved scales on the resolved larger scales of the model. Such effects are only possible in the presence of nonlinear governing equations (NICOLIS 1995, TRIBBIA 1997), since *nonlinearity* is the fundamental reason for scale interactions, very much in the same way as it is responsible for the closure problem in stochastic-dynamic prediction (see below in section 3). This presence of inaccuracies in the parameterizations of unresolved processes in conjunction with nonlinear governing dynamics also necessarily leads, in turn, to limited atmospheric predictability even in the case of error-free initial conditions, because errors in the model description of small, unresolved, parameterized scales inevitably produce similar effects as errors in the initial data, with potential for subsequent amplification. Through the so-called inverse error cascade (e. g., LEITH 1971) errors in the smallest scales induce errors in successively larger scales, until eventually all scales are contaminated (see also, LORENZ 1969a, FORTAK 1973). This contamination of large-scale patterns with errors originating in smaller scales through nonlinear scale interaction leads, in turn, to a time period of limited length over which predictions of detailed weather patterns are possible.

Efforts aimed at the duplication of the result of LORENZ (1963) with sophisticated atmospheric models demonstrated that in that situation, too, two solutions differing only slightly initially diverged with time and eventually became statistically uncorrelated. Consequently, much of the research in atmospheric predictability has focused on investigating at what (average) rate model solutions starting from two such slightly different initial states diverge for a given atmospheric NWP model (see, e. g., CHARNEY et al. 1966, SMAGORINSKY 1969). The Global Atmospheric Research Program (GARP) provided further incentive for studying atmospheric predictability with NWP models (e. g., MIYAKODA et al. 1969, 1972).

Early estimates of such (scale-dependent) error-doubling times were found to be on the order of four to five days; later, LORENZ (1982) obtained estimates of approximately 2.5 days using forecasts produced with the operational NWP model of ECMWF. One of his central results is reproduced in Fig. 3, showing the errors (in a root-mean-square sense) of 500 hPa height forecasts when compared to analyses, as well as when compared to forecasts of the same quantity but for a different lead time, averaged over the period from December 1980 to March 1981. In this figure, the bold line (the upper-most curve) shows the forecast error itself (i. e., the difference between forecast and analysis) as a function of the lead time of the forecast. The lowest curve (labeled 0–1, 1–2, ..., 9–10) shows the difference between two forecasts verifying at the same time, but started from initial times that are 24 hours (1 day) apart; consequently, this curve starts at the value of the forecast error that is obtained for one-day forecasts, and then shows the difference (in a root-mean-square sense) between one- and two-day forecasts, two- and three-day forecasts, and so on up to the difference between the nine- and ten-day forecasts (this measure is sometimes referred to as the *forecast inconsistency* between forecasts one day apart). Similarly, the second curve from below (labeled 0–2, 1–3, ..., 8–10) shows the difference between two forecasts again verifying at the same time, but having been started from initial states that are two days apart.

The interpretation of this figure is manifold. First, and foremost, the lowest curve may be used to estimate the rate of amplification of errors (comparable in size to the one-day forecast error) in the atmosphere, under the assumption that the model used here to produce these forecasts, though evidently not perfect, is a sufficiently accurate description of the atmosphere to represent realistically the divergence of two atmospheric states differing by relatively small initial differences. An extrapolation to smaller errors then leads to the above-mentioned estimate of the error-doubling time of 2.5 days. That value was obtained by LORENZ (1982) under the assumption that the nonlinear terms in the equation governing error growth are quadratic (see also below).

The second implication from this figure is as follows: suppose that the error in the initial state of a model

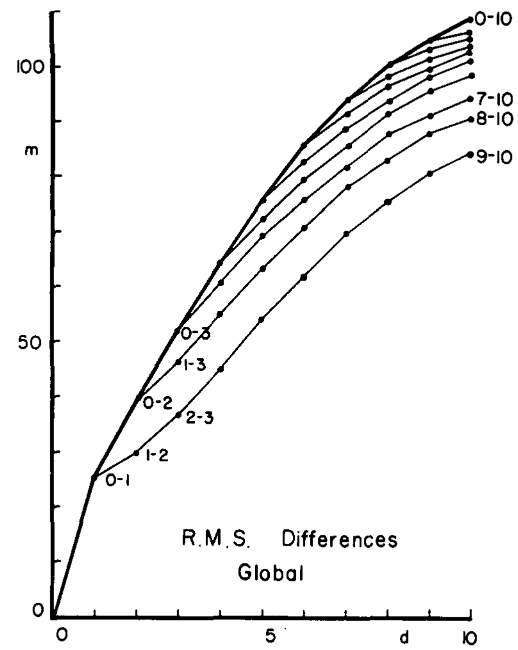


Fig. 3. Global root-mean-square errors between 500 hPa height j-day and k-day ECMWF forecasts (in geopotential meters) valid for the same day, for $j < k$, as a function of k. Values of (j, k) are shown beside some of the points. The heavy curve connects errors between zero-day (i. e., analyses) and k-day forecasts. Thin lines connect errors for constant $k-j$. The forecasts used to produce this figure were operational forecasts for the period 1 December 1980 to 10 March 1981 (figure taken from LORENZ 1982).

Abb. 3. Wurzel aus globalem mittlerem quadratischem Fehler zwischen j-Tages und k-Tages 500 hPa Geopotential ECMWF Vorhersagen (in geopotentiellen Metern) gültig für denselben Tag, für $j < k$, als Funktion von k. Werte für (j, k) sind neben manchen Punkten angegeben. Die dicke Kurve verbindet Fehler zwischen Null-Tages (d. h., Analysen) und k-Tages Vorhersagen. Die dünnen Linien verbinden Fehler für konstantes $k-j$. Die Vorhersagen, die zur Erzeugung dieser Abb. verwendet wurden, waren operationelle Vorhersagen für die Periode von 1. Dezember 1980 bis 10. März 1981 (Abb. aus LORENZ 1982).

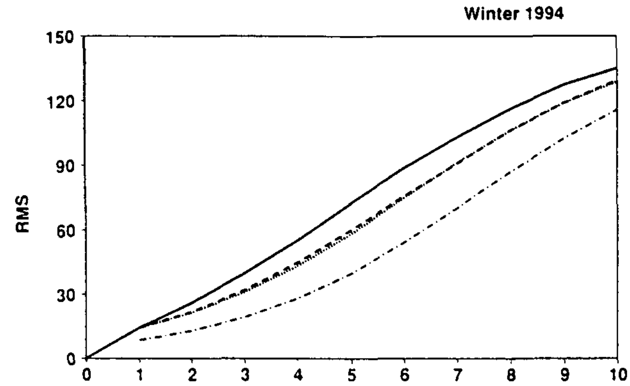


Fig. 4. Root-mean-square errors of 500 hPa height forecasts for winter 1994 over the extratropical northern hemisphere as a function of lead time (solid curve). The dashed curve denotes root-mean-square differences between (k–1)-day and k-day forecasts valid for the same day; the almost incident dotted curve is the result of fitting a Lorenz error-growth model to these differences, while the dashed-dotted curve shows the result of this Lorenz error-growth model for postulated smaller one-day errors (figure taken from SIMMONS et al. 1995).

Abb. 4. Wurzel aus mittlerem quadratischem Fehler von 500 hPa Geopotential ECWMF Vorhersagen für Winter 1994 im Gebiet der extratropischen Nordhemisphäre als Funktion der Vorhersagezeit (durchgezogene Kurve). Die strichlierte Kurve zeigt die Wurzel aus mittleren quadratischen Differenzen zwischen (k–1)-Tages und k-Tages Vorhersagen, gültig für denselben Tag; die fast darüber liegende punktierte Kurve ist das Ergebnis einer Anpassung eines Lorenz Fehlerwachstumsmodells an diese Differenzen, während die strichliert-punktierte Kurve das Ergebnis dieses Lorenz Fehlerwachstumsmodells für postulierte kleinere 1-Tages Fehler zeigt (Abb. aus SIMMONS et al. 1995).

integration is of the size of the current one-day forecast error (approximately 25 gpm in Fig. 3), but that a perfect model could be used instead of the imperfect model. In that situation, the forecast error curve (now the upper-most bold curve) that would be obtained would just be the lowest curve in the diagram, since forecast error would only result from the intrinsic error amplification in the perfect atmospheric model. In other words, the distance between the upper-most and lowest curve in Fig. 3 represents the room for model improvement, assuming that the lowest curve is a good approximation for intrinsic error growth in the atmosphere. Viewed less optimistically, however, one might expect that more realistic (or, improved) models with smaller forecast errors will also lead to more rapid error amplification, thereby reducing the room for improvement that is apparently substantial in Fig. 3, as the lowest curve will move up in that case.

This latter, more pessimistic view is indeed supported by the results from a recent study by SIMMONS et al. (1995) who report error-doubling times on the order of 1.5 days for the currently operational ECMWF forecast model that has horizontal resolution of T213 (i. e., triangular truncation at zonal wavenumber 213) and 31 levels in the vertical (shorthand notation: T213L31). Fig. 4 (taken from SIMMONS et al. 1995) shows the same information as Fig. 3, except for the time period of winter 1994 where the T213L31 operational model has been used to produce the forecasts that enter the calculations to produce the forecast difference curves. Two points become evident from the comparison of Figs. 3 and 4. First, the one-day error has decreased from about 25 to 15 gpm over the period from the early 1980s until 1994. Second, the separation between the forecast error curve and the "perfect-model" error growth curve (lowest curve in Fig. 3, dashed curve in Fig. 4) has approximately halved during this same period. Consequently, the room for improvement in forecast accuracy due to model improvement seems now to be considerably smaller (primarily in view of the faster error-doubling times in this model). To some extent, at least, this more rapid error growth might be the result of the increased resolution of the model used by SIMMONS et al. (1995).

A widely adopted approach for the assessment of the error-doubling times that were mentioned above consists of fitting the free parameters of a conceptual model that hypothetically describes forecast error growth. Such models of varying complexity including the effects of data and modeling error source, as well as nonlinear saturation have been proposed by LEITH (1978), LORENZ (1982), DALCHER and KALNAY (1987), TRIBBIA and BAUMHEFNER (1988a), and STROE and ROYER (1993) (see also SIMMONS et al. 1995, NICOLIS et al. 1995, NICOLIS 1992). It is worth pointing out that the doubling times obtained in these studies should be considered to be representative only for the scales resolved by the model. Resolving increasingly smaller scales is expected to lead to increasingly shorter error-doubling times (e. g., LORENZ 1969a). Such short error-doubling times on very small spatial scales are, in turn, severely limiting the lead times over which useful predictions are possible on these scales; or, stated differently, very high initial state accuracy is required at small scales to allow for accurate prediction at these scales. Along the lines of the results obtained by SIMMONS et al. (1995), it is to be expected that the room for improvements in medium-range prediction should become increasingly smaller if future, more realistic, versions of forecast models continue to lead to decreased error-doubling times, as is presently the case (LORENZ 1982, 1990). However, presumably for each spatial scale a lower limit on its associated error-doubling time should exist.

In view of the evidence provided by the above studies the current estimate for the time range over which useful predictions of detailed synoptic-scale weather patterns should theoretically be possible is approximately *eight days* (BAUMHEFNER 1985, LORENZ 1996, TRIBBIA 1997). This predictability limit for detailed weather patterns must be distinguished from the considerably longer predictability limit for low-frequency, planetary-scale flow patterns such as, for example, the El Niño-Southern Oscillation anomaly (see, e. g., CANE 1992, PENLAND and SARDESHMUKH 1995b, TRIBBIA et al. 1996), or quasi-periodic oscillations like the Madden-Julian oscillation (e. g., MADDEN and JULIAN 1994). Reviews of atmospheric predictability on seasonal time scales may be found, for example, in HOUGHTON (1991), PALMER and ANDERSON (1994), and PALMER (1996). More specific aspects of the predictability problem on these time scales are investigated in, for example, NICOLIS (1987), PALMER (1988, 1993), TIBALDI et al. (1990), TRIBBIA and BAUMHEFNER (1988b, 1993), KUMAR et al. (1996), BAUMHEFNER (1996), and TOTH and KALNAY (1996b). Specific reference is made at this point to the thorough review by ROYER (1993) of research in numerical extended-range prediction during the five years prior to 1993. A selection of reviews of atmospheric predictability studies on a short- to medium-range time scale are LORENZ (1982, 1996), LEITH (1983), SHUKLA (1985), TOMPSON (1985a), ANTHES (1986), GHIL (1987), BOUTTIER (1993b), GÖTZ (1995), PALMER (1996), and HUBER-POCK (1996). Recent predictability studies addressing various issues, such as, for example, the rate of error growth in highly simplified descriptions of the atmosphere, or, the scale dependence of error growth include FRAEDRICH and ZIEHMANN-SCHLUMBOHM (1994), FRAEDRICH and ZIEHMANN (1995), ZIEHMANN-SCHLUMBOHM et al. (1995), FRAEDRICH (1996), TREVISAN and LEGNANI (1995), NICOLIS (1992), STEWART (1994), ERRICO and BAUMHEFNER (1987), and BOER (1994).

The majority of the above studies have investigated atmospheric predictability through numerical experimentation with NWP models. Consequently, the question to what extent the so-found error-doubling times are contaminated by the modeling error source has not been addressed extensively. In an attempt to assess more directly the predictability of the atmosphere itself, LORENZ (1969b) searched a five-year climate record for close analogs to serve as an initial pair, the observed divergence of which would be an indication for true atmospheric error growth. Unfortunately, the closest analogs found were not very close (see also, TREVISAN 1995). Nevertheless, through extrapolation LORENZ (1969b) obtained an estimated error-doubling time of 2.5 days for small errors. Recently, VAN DEN DOOL (1994) has estimated that the probability of finding two analogs (similar to each other within current observational error over the area of the Northern Hemisphere) in a record of ten to hundred years of data is exceedingly small; specifically, he estimated that it would take an observational record of the atmosphere over a time period of 10^{30} years in order to find two atmospheric states that would match to within current observational error. As a consequence, predictability studies have relied heavily on NWP models.

The practical (average) predictability limit of approximately eight days for instantaneous weather patterns as mentioned above must not, however, be understood to imply that it may not on certain individual forecasting situations be possible to produce useful forecasts beyond that time; or, stated differently, it may on occasion be possible to produce forecasts that are more skillful than average for a given lead time (see also Fig. 2). Being well aware of the average predictability limit for the prediction of instantaneous weather patterns, it is the search for ways to separate in an a priori fashion these predictable situations from the unpredictable ones that has stimulated much of the research presenting the theme of this review.

3. Predicting uncertainty: theoretical basis and concepts

a. Theoretical approach

It is important at this stage to describe and discuss in some detail the specific goal(s) to be achieved in the process of predicting the uncertainty of NWP forecasts. First, and foremost, as motivated in the previous two sections, stands the a priori assessment (i. e., prediction) of the skillfulness (or, degree of confidence) of any given deterministic forecast (e. g., KALNAY and DALCHER 1987, PALMER and TIBALDI 1988). Put differently, the goal is to specify the skill of a given forecast in an a priori fashion. The term skill is used at this point in the same generic sense as in the previous sections for measuring in some way the degree of correspondence between a forecast and the relevant verifying analysis (i. e., it may be identified with, e. g., the anomaly correlation coefficient). Nevertheless, the term will require closer consideration further below.

Obviously, the skill (in the above sense) of a forecast is degraded through both the initial state and modeling error source (see section 2). In order to proceed conceptually, uncertainty in the forecast arising through model imperfections is neglected for the moment (perfect-model assumption). It is then quite easily possible to rephrase the above-mentioned goal in a probabilistic framework in the following way. Recognition of the fact that the initial model state can only be known to within some uncertainty (due to imperfections in the observational system and data assimilation algorithms), necessarily leads to the specification of the initial model state in terms of the probability density function (pdf) of the (N-dimensional) model state vector $\mathbf{X}$ at initial time, denoted as $\rho_0(\mathbf{X})$ (the subscript on ρ denotes the time for which the pdf is valid, and not partial differentiation). This pdf ρ_0 describes in a probabilistic format all information available about the initial model state in the phase space of the model. Such a probabilistic format is sufficient to describe uncertainty in a comprehensive way (LINDLEY 1987). Clearly, the actual analysed state of the atmosphere (to be used as initial state in the forecasting process), denoted $\mathbf{X}_0^c$, as well as the true state of the atmosphere (in its appropriate phase space representation), denoted $\mathbf{X}_0^t$, that is in principle unknown, are both realizations of a random variable with pdf ρ_0. This initial pdf ρ_0 evolves under the dynamics of the NWP model into the pdf $\rho_t(\mathbf{X})$ at some later time t. At a very fundamental level, then, any attempt to predict forecast uncertainty must address the question: given initial state uncertainty as summarized through $\rho_0(\mathbf{X})$, as well as the dynamics of a forecast model (see eq. (3.3) below), *predict the time evolution of this initial state uncertainty under the action of the model dynamics.* Beyond that question, utilizing the concept of $\rho_t(\mathbf{X})$, the above-mentioned goal of predicting skill can now be rephrased as: *find the expected distance (skill) between two realizations from* $\rho_t(\mathbf{X})$. This interpretation relies on the fact that — under the perfect-model assumption — both the actual forecast, as well as the verifying analysis (at time t), can be considered to be realizations from $\rho_t(\mathbf{X})$.

Before elaborating in some detail on the ideas put forward in the previous paragraph, it seems worthwhile mentioning that the time evolution of the pdf $\rho_t(\mathbf{X})$ is governed by a linear partial differential equation in the single unknown ρ_t, known as the *Liouville equation (LE),* that may be written in the following format:

$$\frac{\partial\rho_t(\mathbf{X})}{\partial t}+\sum_{k=1}^{N}\Phi_k(\mathbf{X})\frac{\partial\rho_t(\mathbf{X})}{\partial X_k}=-\psi(\mathbf{X})\rho_t(\mathbf{X}) \qquad (3.1)$$

where:

$$\psi(\mathbf{X})\equiv\sum_{k=1}^{N}\frac{\partial\Phi_k(\mathbf{X})}{\partial \mathbf{X}_k} \qquad (3.2)$$

denotes the divergence of the flow in phase space, N denotes the dimension of the model phase space, and subscript k denotes the kth component of the relevant vector. The vector-valued function $\mathbf{\Phi}$ describes the (autonomous), generally nonlinear, dynamics governing the time evolution of the state vector of the NWP model:

$$\frac{d}{dt}\mathbf{X}=\Phi(\mathbf{X}). \qquad (3.3)$$

Imposing the initial condition:

$$\rho_{t=0}(\mathbf{X})=\rho_0(\mathbf{X}) \qquad (3.4)$$

on (3.1), the solution to (3.1) and (3.4) may be written in the following form:

$$\rho_t(\mathbf{X})=\rho_0(\Xi)\exp\left[-\int_0^t\psi[\mathbf{X}(\Xi,t')]dt'\right], \qquad (3.5)$$

where Ξ is the (unique) point in phase space that, under the dynamics (3.3), is mapped into the point $\mathbf{X}$ as time evolves from $t=0$ to t. Thus, in view of (3.5), it suffices, in principle, to evaluate (3.5) and to compute from this solution the expected distance of two realizations (or, any other desired statistics) of ρ_t. Needless to say, there are various complications prohibiting this direct approach.

This probabilistic description of the problem of predicting uncertainty in terms of the LE is based on the continuity argument that total probability (i. e., the integral of the pdf valid at any time over the entire phase space) must be conserved, or, equivalently, that realizations of the dynamical system (3.3) evolving in time cannot spontaneously appear or vanish. The mathematical formulation of this argument is the continuity equation for probability written here in the form (3.1), known as the LE in statistical mechanics (e. g., BALESCU 1975). The derivation of the LE may be carried out from this integral conservation statement (see, e. g., LIN and SEGEL 1988) and is completely analogous to the derivation of the continuity equation in fluid mechanics. The LE bears the name of Joseph LIOUVILLE (1809–1882), presumably because its derivation is facilitated if a general result is used that relates the derivative of the Jacobian of a transformation to the divergence in phase space. This result was first described by LIOUVILLE (1838) while studying problems in the theory of ordinary differential equations (see also LÜTZEN 1990).

The significance of the LE and its potential for applications in the atmospheric sciences were first discussed in a meteorological context by GLEESON (1966, 1970), EPSTEIN (1969), THOMPSON (1972), and FORTAK (1973). As becomes clear from (3.1) together with (3.2), the LE states that the local rate of change of the pdf at any point in phase space must be exactly balanced by the net probability flux across the surface of a small, but finite, volume element surrounding that point in phase space. As such, the LE is entirely analogous to the well-known continuity equation of fluid mechanics.

It is noted here that the LE takes on the form of a Fokker-Planck equation (FPE) (e. g., THOMPSON 1972, 1983; HASSELMANN 1976; MOSS and MCCLINTOCK 1989; SOIZE 1994; RISKEN 1989; GARDINER 1990), in the situation that the model dynamics (3.3) are augmented to include stochastic forcing components of a certain form (i. e., such that the stochastic transport of points in phase space can be described as a diffusive process; THOMPSON 1972). Such forcing components may be added in an attempt to relax the perfect-model assumption made so far or with the goal to study the equilibrium statistical mechanics of a system subject to noise. While extending the governing deterministic dynamics (3.3) to a set of *stochastic* differential equations is conceptually straightforward, the greater difficulty involves the determination of an appropriate and realistic description of the modeling error source in terms of stochastic components to be added to the original NWP model. Some discussion of the issue of model error in the context of uncertainty prediction may be found in DEE (1995) and HOUTEKAMER et al. (1996b) (see also, section 2).

As a brief illustration of the extension of the LE to an FPE, consider the following one-dimensional stochastic differential equation:

$$\frac{\partial X}{\partial t} = \phi(X) + \gamma(t), \tag{3.6}$$

where $\phi(X)$ is an arbitrary (possibly nonlinear) function of X, and $\gamma(t)$ is a stochastic forcing term (taken here in its simplest form as a prototypical term to mimic random model error). Under the assumption that $\gamma(t)$ is a normal random variable, with mean zero and variance Γ, uncorrelated in time (white noise), the time evolution of the pdf $\rho_t(X)$ under the dynamics (3.6) is described by the following FPE:

$$\frac{\partial \rho_t(X)}{\partial t} + \frac{\partial}{\partial X}\left[\phi(X)\rho_t(X) - \frac{\Gamma}{2}\frac{\partial \rho_t(X)}{\partial X}\right] = 0, \tag{3.7}$$

where the first term in brackets results from the deterministic component in (3.6), while the second (diffusive) term results from the stochastic component in (3.6) (see DEGROOT and MAZUR 1984). For more complex random forcings (especially state-dependent forcings) than the one introduced in (3.6) the consideration of the correct interpretation of stochastic differential equations and their probabilistic description is necessary (see, e. g., HASSELMANN 1976, VAN KAMPEN 1992, LASOTA and MACKEY 1994, STRATONOVICH 1992, ARNOLD 1992). Nevertheless, the above example may serve as an illustration of the general idea.

For any (in particular, nonlinear) governing dynamics of the form (3.3), the LE is a linear, first-order (second-order in the case of the FPE) partial differential equation, with the pdf as a function of time and phase space coordinates being the single dependent variable, and, as such, accessible to the solution given in (3.5). The LE (or, its associated FPE) has, among other applications, been the starting point for investigations concerning equilibrium statistics in turbulent flows (e. g., THOMPSON 1972, 1983, 1985b, 1986a; SALMON et al. 1976) and climate problems (e. g., THOMPSON 1985c). Quite generally, the balance between deterministic forcing and diffusion caused through noise, as displayed in the stationary form of the FPE and generally known as the *fluctuation-dissipation relation*, provides the starting point for investigations of equilibrium statistics attained by such flows (see, e. g., FARRELL and IOANNOU 1993a, b; DELSOLE and FARRELL 1995, 1996; PENLAND 1989; PENLAND and MATROSOVA 1994; ALEXANDER and PENLAND 1996; NEWMAN et al. 1997). One of the prime applications of the LE in the atmospheric sciences is, however, in the problem of predicting forecast uncertainty, as this equation provides the natural and general approach to this problem as described above (e. g., EPSTEIN 1969; FORTAK 1973; THOMPSON 1986a; PAEGLE and ROBL 1977; EHRENDORFER 1994a, b; GÖTZ 1996).

b. Concepts

With reference to the LE as the general framework for the prediction of forecast uncertainty, EPSTEIN (1969) postulated that the description of the atmospheric state at any time can sensibly only be given in probabilistic terms (see above). Recognizing the difficulties involved with and the computational efforts necessary for producing the solution of the LE over the entire phase space, EPSTEIN (1969) proposed to consider instead solving the equations governing the time-evolution of first, second, and possibly higher moments of the pdf of the model state vector. These equations were subsequently referred to as the *stochastic-dynamic equations* (see also FORTAK 1973). Clearly, this stochastic-dynamic prediction approach requires the derivation and numerical implementation of these equations that increase rapidly in number as moments beyond the first are considered. In addition, this approach calls for a solution to the closure problem to a sufficient degree of accuracy, that arises because, in the presence of nonlinear governing dynamics, the stochastic-dynamic equations for moments up to a given order will contain terms involving moments of order greater than this given order.

Stochastic-dynamic prediction has been studied only in a limited number of investigations. Theoretical progress, however, in this area was made by THOMPSON (1985d) who showed in the context of the spectral barotropic vorticity equation that (under certain assumptions) a closed set of stochastic-dynamic equations for the variances could be obtained; roughly speaking, these assumptions involved neglect of covariances between modes in different triads, and parameterization of covariances between modes in the same triad. Although this simplification leads to an enormous reduction of the computational burden encountered, only a limited number of further applications of stochastic-dynamic prediction exist (see also THOMPSON 1986b, and section 4).

LEITH (1974) proposed a strategy for dealing with the prediction of forecast uncertainty that represents a computationally more easily tractable alternative to both the LE and stochastic-dynamic prediction. This strategy, referred to as *Monte Carlo (MC) method*, consists of generating a (large) ensemble of initial model states through (random) sampling from ρ_0, and evolving each of the resulting initial states by the NWP model. For sufficiently large sample sizes, reliable estimates of (moments of) the time-evolved pdf ρ_t may be computed from the time-evolved ensemble members (e. g., HAMMERSLEY and HANDSCOMB 1964, FISHMAN 1996). This approach accounts for uncertainty in the initial state insofar as each of the initial states selected is a likely candidate for the initial model state, consistent with the accuracy of the available analysis of the atmospheric state (as described through ρ_0). *The MC approach is (in modified forms) at the basis of all currently operational efforts that are aimed at gaining information about the time-evolved pdf* ρ_t, and, in turn, at assessing the skill of forecasts in an a priori fashion, since these efforts are largely based on multiple model integrations (see section 5).

Given such multiple NWP model integrations from an application of the MC method, goals beyond the ones stated at the beginning of this section may seem feasible: for example, the mean forecast of such a set of model integrations might be computed in an attempt to achieve improved accuracy over individual integrations (i. e., improvement of the control forecast might be attempted through filtering). Or, probabilities for the occurrence of certain events, as well as higher sample moments might be computed from such a set of multiple model integrations (see, e. g., FRAEDRICH and ZIEHMANN 1995; HOUTEKAMER and DEROME 1994, 1995; see also section 5).

c. Operational restrictions and fundamental problems

The MC method is computationally less demanding than solving the relevant stochastic-dynamic equations, or even the relevant LE. Nevertheless, its computational cost is still prohibitively high in the context of state-of-the-art operational NWP models due to the large dimension of their associated phase space. Continued research efforts at operational centers (e. g., MOLTENI et al. 1996; TOTH and KALNAY 1996a, 1997; HOUTEKAMER et al. 1996a; see also section 5) are therefore concentrating on the design of sampling strategies that allow one to gain information about the evolution of ρ_0 efficiently even in the case when the sample size is restricted to be (extremely) small (when compared to the dimension N of the model phase space). Such sampling strategies are collectively referred to here as *ensemble prediction methods.* Some of these are based on the concept of directions of most rapid error growth (see section 5a), as described by appropriately constructed initial perturbations. The design of other sampling strategies is, at least partly, such that they allow collection of information about ρ_0 in the situation where specific properties of this initial pdf are not known explicitly or even unknown (see section 5b).

Inspite of this *operational restriction* that the number of model integrations that can be carried out with an operational NWP model in order to study the temporal evolution of $\rho_0(\mathbf{X})$ is severely limited (and much smaller than the dimension of the model phase space), the MC method is a perfectly valid and powerful procedure for evaluating the solution (3.5) to the LE through evolving an ensemble *randomly* drawn from ρ_0 with the dynamics (3.3). In fact, the MC method very closely resembles computing the solution to the LE (3.5) directly for various points $\mathbf{X}$ in phase space. However, under that restriction the MC method may encounter serious sampling problems when moments of $\rho_t(\mathbf{X})$ are estimated from multiple model integrations. For sampling that should be more efficient than random sampling because it fills phase space in a *quasi-random* manner, reference is made at this point to MOROKOFF and CAFLISCH (1995), and SENDOV and ANDREEV (1994) (see also SCHUBERT et al. 1992).

Beyond that operational restriction, two additional *fundamental problems* leading to complications within this approach must be mentioned. First, and foremost, in real-world applications the NWP dynamics (3.3) is different from the true atmospheric dynamics that supposedly links the true atmospheric states used for subsequent validation. In other words, due to the clearly existing modeling deficiencies in the NWP model, the pdf evolution as assessed through the MC approach will (to some extent) be different from the pdf evolution relevant for the true atmospheric dynamics. As a consequence, even though both analysed state $\mathbf{X}_0^c$ and true state $\mathbf{X}_0^t$ should be regarded as realizations from the initial pdf ρ_0, it will no longer be clear that the true state at time t can necessarily be regarded as member of ρ_t, that is obtained by employing the NWP model dynamics (3.3). As mentioned before, it may, in an attempt to compensate for detrimental effects arising from making the perfect-model assumption, be possible to augment the model dynamics (3.3) by stochastic representations of model error in order to obtain more realistic dynamics. However, at a very fundamental level modeling deficiencies will remain unavoidable, and, in turn, when employed within the MC approach, will lead to a time-evolved pdf that is different from the pdf that one would obtain when using the true atmospheric dynamics. In other words, even without sampling restrictions, the fact that model dynamics (used in the MC approach or the LE) possess deficiencies necessarily leads at best to an approximation of the time-evolved pdf. In order for such an approximation to be useful for estimating forecast uncertainty, it is necessary that the modeling error source is smaller than the initial state error source (see section 2). However, since this difficulty is also encountered in NWP in general, where it must be assumed that the model is a faithful representation of the atmosphere, it will not be discussed in any more detail from this point on.

The second difficulty arises since among the goals of primary interest in the design of ensemble prediction methods is, as mentioned previously, not so much the prediction of the time evolution of ρ_0 itself, but rather a prediction of the (expected) skill of a single (control) forecast (see, e. g., MOLTENI et al. 1996; see sections 3a and 5a). For concreteness, skill may, for example, be assessed through the anomaly correlation coefficient (see section 1). This goal is more difficult to achieve than the prediction of the evolution of ρ_0 (that is governed by a linear partial differential equation, i. e., the LE; see above) for two reasons. The first reason is related to model errors leading to inaccurate prediction of ρ_t. The second difficulty in achieving this goal arises since presumably there does not necessarily exist a simple relationship between properties of the time-evolving pdf ρ_t and the skill of the control forecast. Even without the presence of model errors — a situation in which both control and verifying trajectory can be thought of as being realizations from the time-evolving pdf ρ_t — the expected prediction error will, in general, not be simply related to more easily derivable properties of ρ_t, such as an overall variance measure, even for moderately complex (nonlinear) NWP models. Nevertheless, it would seem, at least in principle, possible to derive on the basis of ρ_t an expected prediction error (or, skill measure) measuring in some form the expected distance of two realizations taken from ρ_t (like the control forecast and the verifying trajectory). In turn, obviously the presence of model errors leads to complications. Currently, empirical approaches are investigated at operational centers that allow one to move closer to this goal to be achieved by ensemble prediction methods (see also, section 5a).

4. Pre-operational uncertainty prediction in NWP: selected activities

The discussion in the previous section indicates that the theoretical foundations for the prediction of the uncertainty inherent in NWP forecasts, specifically the LE, as well as the stochastic-dynamic prediction and the MC prediction approach are firmly established, at least in principle. Nevertheless, certain theoretical aspects of these concepts still deserve further attention (e. g., the closure problem). In addition, these concepts have been put to use in operational contexts only comparatively recently. Three, not entirely unconnected reasons seem to be responsible for the fact that these concepts have been exploited operationally only rather recently (approximately at the beginning of the 1990s). First, computing capacities before then were generally not powerful enough to allow for multiple integrations of the NWP models used for operational weather prediction at that time. Second, during the 1980s there did not exist a well-established or widely-accepted rationale or method for generating in an efficient and meaningful way realizations from ρ_0 to be subsequently used for multiple model integrations. Third, the model error present in operational models during that time was generally considered too large (compared to the initial state error; see section 3c) for their successful use in uncertainty prediction. However, various pre-operational attempts were carried out during these years to attack the problem of uncertainty prediction; some of them are discussed here.

In the area of stochastic-dynamic prediction, FLEMING (1971a, b) and PITCHER (1977) studied various questions, among them the problems of closure as well as of model errors. Also, THOMPSON (1985d, 1986b, 1988) studied various aspects of the closure problem, mostly within barotropic atmospheric dynamics. As mentioned before, it appeared possible to derive a closed system for variances under certain assumptions that was sufficiently accurate for short enough time intervals. Nevertheless, the need to derive equations for higher-order moments that exists here has prevented the stochastic-dynamic prediction approach from being pursued in operational contexts. It is noted at this point that the prediction step in the Kalman filter bears a strong relationship to the stochastic-dynamic prediction approach insofar as the Kalman filter equations deal explicitly with the covariances of the components of the model state vector (see, e. g., JAZWINSKI 1970, BOUTTIER 1996; see also section 6); however, as pointed out by COHN (1993) the closure chosen for standard forms of the extended Kalman filter (in nonlinear problems) appears to exhibit an inconsistency when compared to the stochastic-dynamic equations (this inconsistency relates to the coupling between first- and second-order moment equations).

An innovative approach, called lagged-average forecasting, that avoids the expensive production of multiple forecasts was proposed by HOFFMAN and KALNAY (1983). Their basic idea consists of studying a collection of forecasts, verifying at the same time, but initialized at various preceeding times (i. e., lagged in time), as a surrogate for realizations from the time-evolved pdf ρ_t. This collection of forecasts is routinely available from the operational forecast runs; thus, this technique does not require additional computing resources. Results reported from the application of the lagged-average forecasting technique emphasize the usefulness of this approach (e. g., KALNAY and DALCHER 1987; MURPHY 1988, 1990; DALCHER et al. 1988; BRANKOVIĆ et al. 1990; see also section 5b).

In this pre-operational period, experiments based on purely statistical methods (e. g., multiple linear regression) were performed in an attempt to identify directly relationships between a variety of predictors (e. g., the error in the estimate of the initial state, or specific features characterizing the initial state) and the actual skill of the forecast (measured, for example, through the anomaly correlation coefficient). While ideally suited for operational use due to their negligible cost (when compared to multiple integrations of the NWP model) results from relevant studies clearly show the limitations of these empirical methods (e. g., MOLTENI and PALMER 1991, PALMER and TIBALDI 1988, COLUCCI and BAUMHEFNER 1992). The issue of predicting directly the skill of a forecast (see the more precise discussion in the previous section) will be brought up briefly again in section 5a in the context of the consideration of the relationship between the skill (of the control forecast) and the spread within the time-evolved ensemble. In addition, during that same period, the MC approach was widely used, although not operationally, in a number of studies testing its feasibility (e. g., MURPHY 1988, 1990), and using it as standard to check the accuracy of closure assumptions (e. g., EPSTEIN 1969, FLEMING 1971a, THOMPSON 1985d).

5. Operational uncertainty prediction in NWP: Ensemble prediction systems

Significantly increased computing capabilities that became available at major operational NWP centers at the beginning of the 1990s moved the actual execution of (parts of) the theoretical program outlined in sections 3a, b into the realm of a real possibility. This increase in computing capabilities was at the same time accompanied by the accumulating evidence that state-of-the-art models had in a sense reached a certain level of saturation regarding their degree of complexity (particularly in terms of resolution). Further, model errors were apparently reduced considerably (e. g., SIMMONS et al. 1995). As a consequence, it seemed that these additional computing capabilities might lead, when used to attack the problem of predicting the uncertainty of NWP forecasts, to highly increased usefulness of the forecast products. On the other hand, at that stage it was anticipated that improvements that would result from using those enhanced computing capabilities for further increasing the complexity of the NWP model itself would be comparatively minor.

Nevertheless, in view of the tremendous degree of complexity of the NWP models used for operational forecasting at that (and present) time (the number of the degrees of freedom N of such models is presently on the order of several million), a procedure based on a modified MC approach had to be adopted at these NWP centers. This modification concerns primarily the fact that initial states are not selected *randomly*, but rather in a very systematic (or *biased*) way described in some detail below that explicitly utilizes dynamical information about the atmospheric flow. Sampling approaches utilizing dynamical information offered the hope that they would lead to a most efficient use of available computing resources, given that the limits on these resources seemed to preclude the exploitation of the full potential of the LE in the context of such models, as well as the derivation and subsequent integration of stochastic-dynamic equations. Consequently, at these centers research efforts were directed towards the question of how to select these initial states in a manner that would be most efficient for learning about the time-evolved pdf ρ_t. The crucial importance of that question was obvious in light of the very small (when compared to N) number of parallel model integrations allowable.

The discussion of the resulting implementations of ensemble prediction in this section will concentrate on the following four major efforts: namely, developments at the ECMWF, the National Centers for Environmental Prediction (NCEP, formerly known as the National Meteorological Center) in the United States, the United Kingdom Meteorological Office (UKMO), and Recherche en prévision numérique (RPN) at the Atmospheric Environment Service in Canada. One of the major differences of these systems is the way how initial perturbations are generated. Here, attention will be focused on the methodology utilized at ECMWF; the methodologies used at the three other centers will be briefly discussed. A few brief comments will be made concerning the validation of products from ensemble prediction systems.

a. European Centre for Medium-Range Weather Forecasts

Introductory remarks

At ECMWF the generation of members in the initial ensemble (i. e., the sampling from the initial pdf ρ_0 relies heavily on the concept of finite-time most unstable structures, introduced in meteorology in the context of a predictability study by LORENZ (1965). The precise definition of these structures that are also referred to as singular vectors (SVs) is given in the appendix. Initial ensemble members are constructed by adding linear combinations of these optimal perturbations to a reference (or, control) initial state (denoted as $\mathbf{X}_0^c$ in section 3a). As discussed in the appendix, these optimal perturbations are obtained as the solutions of an optimization problem aimed at maximization of an error measure (also referred to as cost function). That problem may — under the assumption of tangent-linear error growth (see below) — be reduced to a symmetric eigenproblem. In this situation, the optimal perturbations are linearly related to the so-called right SVs (see, e. g., GOLUB and VAN LOAN 1989, NOBLE and DANIEL 1977) of the resolvent of the tangent-linear model equations. The basic rationale for this approach is that the SVs — by representing those directions in phase space along which errors will amplify most rapidly — also capture, in a sense to be made precise, the most important properties of the time-evolving pdf ρ_t; a pedagogical illustration for this idea is given below. Adopting such a rationale that potentially promises an efficient sampling strategy, seems to be of particular importance in view of the severe restrictions imposed by computational resources on the number of ensemble members allowable in the situation of operational NWP models.

Before describing in more detail the methodology used at ECMWF and a selection of relevant results, an illustration for the basic idea behind this approach is illustrated in Fig. 5 through a pedagogical example. This figure shows the exact solution of the LE, as evaluated from (3.5), for the three-dimensional dynamical system proposed by LORENZ (1984) as a prototype for the atmospheric general circulation. The equations of this system are repeated here for reference (as a special case of eq. (3.3)):

$$\frac{d}{dt}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -Y^2 - Z^2 - aX + aF \\ XY - bXZ - Y + G \\ bXY + XZ - Z \end{pmatrix}, \qquad (5.1)$$

where the parameters a, b, F, and G are chosen here to take on the values 0.25, 4.0, 8.0, and 1.25, respectively. For this choice of parameters, the system is chaotic, in the sense that the largest Lyapunov exponent (e. g., PARKER and CHUA 1989) is positive. Eq. (5.1) is nondimensional, with the unit of time being five days. The properties of this system with regard to atmospheric predictability have been studied extensively (see, e. g., NICOLIS 1992, NICOLIS et al. 1995).

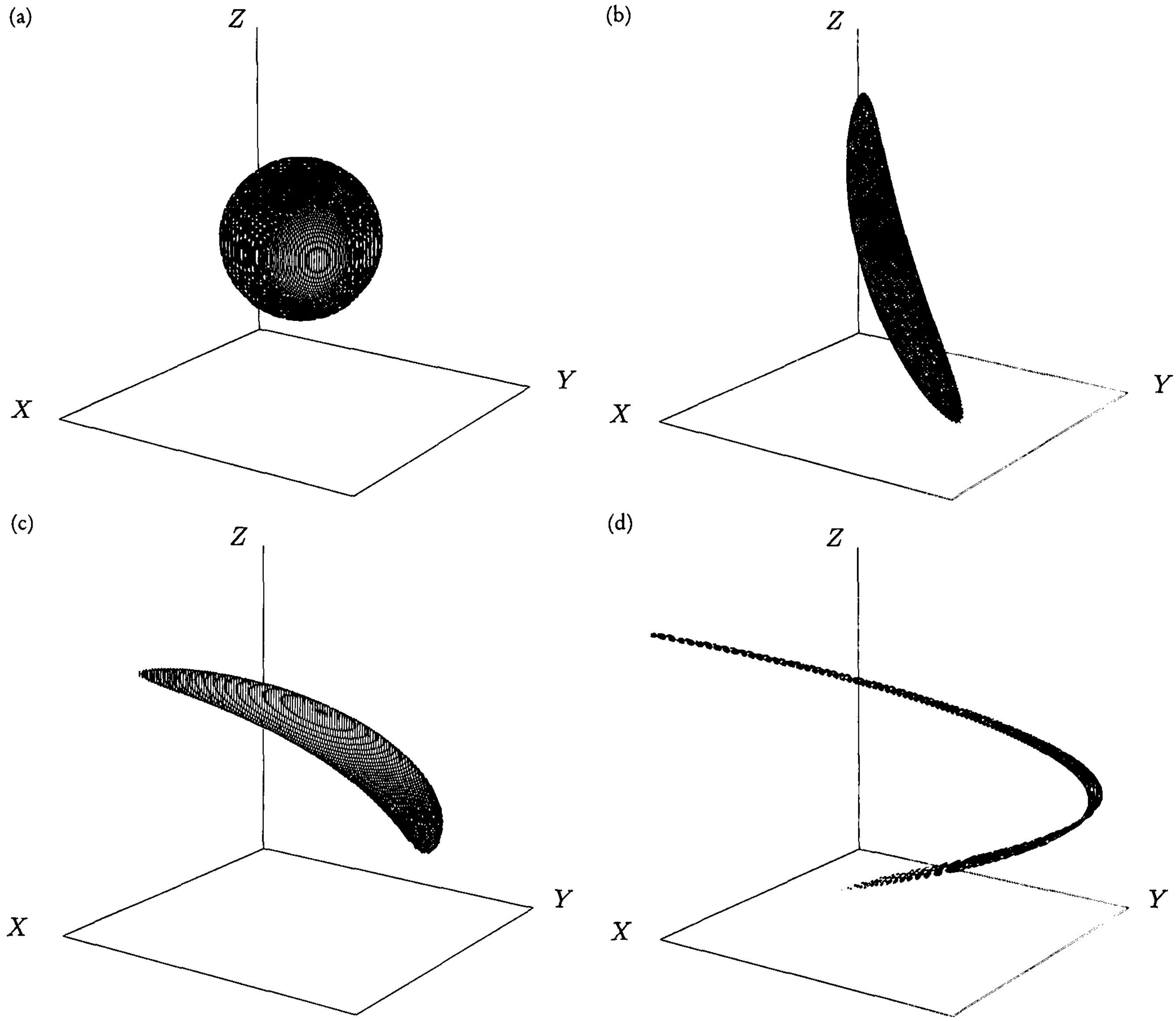


Fig. 5. Time evolution in phase space of an initially Gaussian probability density function under the dynamics of the LORENZ (1984) system (parameter values are a=0.25, b=4.0, F=8.0, G=1.25). Panel a) shows the initial pdf, panels b), c), and d) show the pdf after one, two, and three nondimensional time steps, respectively. Plotting of the pdfs is schematically, in the sense that surfaces shown are the surfaces within which the pdf-values are larger than one. The initial pdf has mean of $(X, Y, Z)=(2, -1, 0)$, with diagonal covariance matrix (all three variances equal to 0.01^2).

Abb. 5. Zeitliche Entwicklung im Phasenraum einer Wahrscheinlichkeitsdichtefunktion (WDF), die anfangs Gauß ist, durch die Dynamik des LORENZ (1984) Systems (Parameterwerte sind a=0,25, b=4,0, F=8,0, G=1,25). Bild a) zeigt die anfängliche WDF, Bilder b), c) und d) zeigen die WDF nach ein, zwei und drei dimensionslosen Zeitschritten. Die WDFs sind schematisch dargestellt, in dem Sinn, daß die gezeigten Flächen jene Flächen sind, innerhalb derer die WDF-Werte größer als eins sind. Die anfängliche WDF hat Mittelwert $(X, Y, Z)=(2, -1, 0)$, mit diagonaler Kovarianzmatrix (alle drei Varianzen sind gleich $0{,}01^2$).

In producing the results shown in Fig. 5, the initial pdf $\rho_0(\mathbf{X})$ (panel a) has been taken as multivariate normal, with diagonal covariance structure. The pdfs are presented in this figure through the surfaces that separate values of ρ larger than one (inside the surface) from those that are smaller than one (outside the surface); clearly, in this graphical display, ρ_0 appears as a sphere. It may be seen that the time-evolving pdf (shown in panels b, c, and d after one, two, and three nondimensional time steps, respectively) is stretching during the first time step very rapidly along a preferred direction, while it is shrinking along other directions. Loosely speaking, errors (or, more precisely, the variance) grow along certain directions, while they decay along other directions. This growth along a (number of) preferred direction(s) is presumably well described through linearized dynamics (see below), whereas nonlinear effects become clearly important for the evolution of the pdf at later times as shown in panels c) and d). The manifestation of nonlinearity appears through the deformation of the initial sphere into shapes different from ellipsoids.

It is noted here for completeness that for the system under consideration any volume in phase space will, as time proceeds, shrink to zero, since the sum of the Lyapunov exponents is negative for the parameter values chosen for the computations carried out to produce this figure; that is, the present system is dissipative (see, e. g., NICOLIS 1995). However, along the lines of the main argument that leads to the LE, the *integral over the pdf* enclosed by any material surface does remain constant in time (even for a dissipative system); this, in turn, is a consequence of the fact that all points initially on that material surface are mapped by the flow into points on a deformed material surface at a later time, so that no realizations can spontaneously appear or vanish (see, e. g., EGGER 1996, ARNOLD 1989). Note, however, that the surfaces shown in Fig. 5 do not represent material surfaces in this sense, since the surfaces in panels b)–d) are not the maps by the flow of the surface in panel a); they are rather simply the surfaces that separate values of ρ larger than one from those smaller than one (see above).

Studying the deformation of a material surface under the dynamics (5.1) yields further insight into the gradual appearance of nonlinear effects, as well as the importance of preferred directions during the initial validity of the tangent-linear approximation. In this context, panels a)–d) of Fig. 6 show the time evolution of such a *material surface* under (5.1). At the initial time, this material surface is taken to be the surface of a sphere that is centered at the mean of ρ_0, with a radius of 0.02 units (this choice corresponds to two standard deviations of the pdf shown in Fig. 5a). The material surfaces are shown here schematically through a set of seven differently colored "latitude circles" lying on the surfaces of the three-dimensional objects. (Note that the perspective of viewing leads to a distortion of the three-dimensional objects; the degree of that distortion becomes evident from panel a) displaying the initial *sphere*). It can be seen from Fig. 6 that during an initial period (i. e., panels b and c for nondimensional times 0.5 and 0.9, respectively) the material surface is deforming linearly through rotating and stretching. This stretching of the surface is again occurring along the direction of the variable Z. At the later time $t = 3.5$ (panel d), the surface has been deformed in a highly nonlinear way. In Figs. 6a)–d) the time-evolved material surface is shown in a frame moving with the time-dependent reference trajectory started at the mean of ρ_0; that is, the surfaces are shown in terms of deviations (denoted by primes in the labeling of the axes) from the time-evolving reference trajectory. This reference trajectory is shown in Fig. 6e), with the red dots marking the times for which the surfaces are displayed in the previous panels. Note that this procedure ensures that the nonlinear deformation of the material surface is displayed.

The behavior of system (5.1) shows — within the tangent-linear regime — rapid spreading of probability in a certain direction, while probability is contracting in other directions. Material surfaces are distorted in a qualitatively similar fashion (Figs. 5 and 6). If this behavior is also typical for atmospheric dynamics, it is reasonable to study error growth along the directions of most rapid amplification for an operational NWP model (these directions are — in a tangent-linear context — represented by the SVs, in a sense made precise below), since these directions will capture most efficiently the relevant error growth. In addition, given these directions, a sampling process within an ensemble prediction system may be designed to explicitly prefer these directions.

This idea is exploited in the ensemble prediction system at ECMWF. It may be formalized in the following way, assuming tangent-linear perturbation growth. This tangent-linear hypothesis expresses the assumption that the time-evolution of perturbations $\mathbf{X}' \equiv \mathbf{X} - \overline{\mathbf{X}}$ from a reference model trajectory $\overline{\mathbf{X}}$ of small, but finite size, can be described with sufficient accuracy over a certain period of time through the *tangent-linear model* (5.2) that is obtained through linearization of the nonlinear model equations, such as (3.3), along the reference model trajectory:

$$\frac{d}{dt}\mathbf{X}' = \mathbf{L}\mathbf{X}'. \tag{5.2}$$

Here the tangent-linear dynamics $\mathbf{L}$ is obtained as the first derivative of the nonlinear dynamics evaluated along the reference model trajectory $\overline{\mathbf{X}}$:

$$\mathbf{L}_{ij} \equiv \frac{\partial \Phi_i(\mathbf{X})}{\partial \mathbf{X}_j}\Big|_{\overline{\mathbf{X}}}. \tag{5.3}$$

For example, the dynamics tangent-linear to the system (5.1) are given — as a special case of (5.2) — by:

$$\frac{d}{dt}\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} -a & -2\overline{Y} & -2\overline{Z} \\ \overline{Y} - b\overline{Z} & \overline{X} - 1 & b\overline{X} \\ b\overline{Y} + \overline{Z} & b\overline{X} & \overline{X} - 1 \end{pmatrix} \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix}, \tag{5.4}$$

where quantities with an overbar satisfy the nonlinear system (5.1). As such, the tangent-linear model (5.2) together with (5.3) represents the first-order approximation (in the perturbation $\mathbf{X}'$) to the fully nonlinear time evolution of perturbations that is governed by:

$$\frac{d}{dt}(\mathbf{X} - \overline{\mathbf{X}}) = \Phi(\mathbf{X}) - \Phi(\overline{\mathbf{X}}). \tag{5.5}$$

The approximation (5.2) to (5.5) leads to both conceptual simplifications and computational savings (e. g., nonlinear optimization problems are reduced to eigenproblems); however, when making this approximation, its validity must be tested thoroughly. Studies carried out so far indicate that this assumption is reasonably valid for perturbations comparable in size to current analysis errors over time periods of about two days (e. g., LACARRA and TALAGRAND 1988, ERRICO et al. 1993b, BUIZZA 1995a, TANGUAY et al. 1995). For example, BUIZZA (1995a) investigated the relative magnitudes of first- and higher-order Taylor expansion coefficients in the context of a primitive-equation T63L19 model (without moist physics). One of his results is reproduced in Fig. 7. It can be seen from this figure that the first-order term (solid line) is the largest term up to five days, and that at two and a half days the second-order term (dashed line) becomes half as big as the first-order term. Clearly, these results are

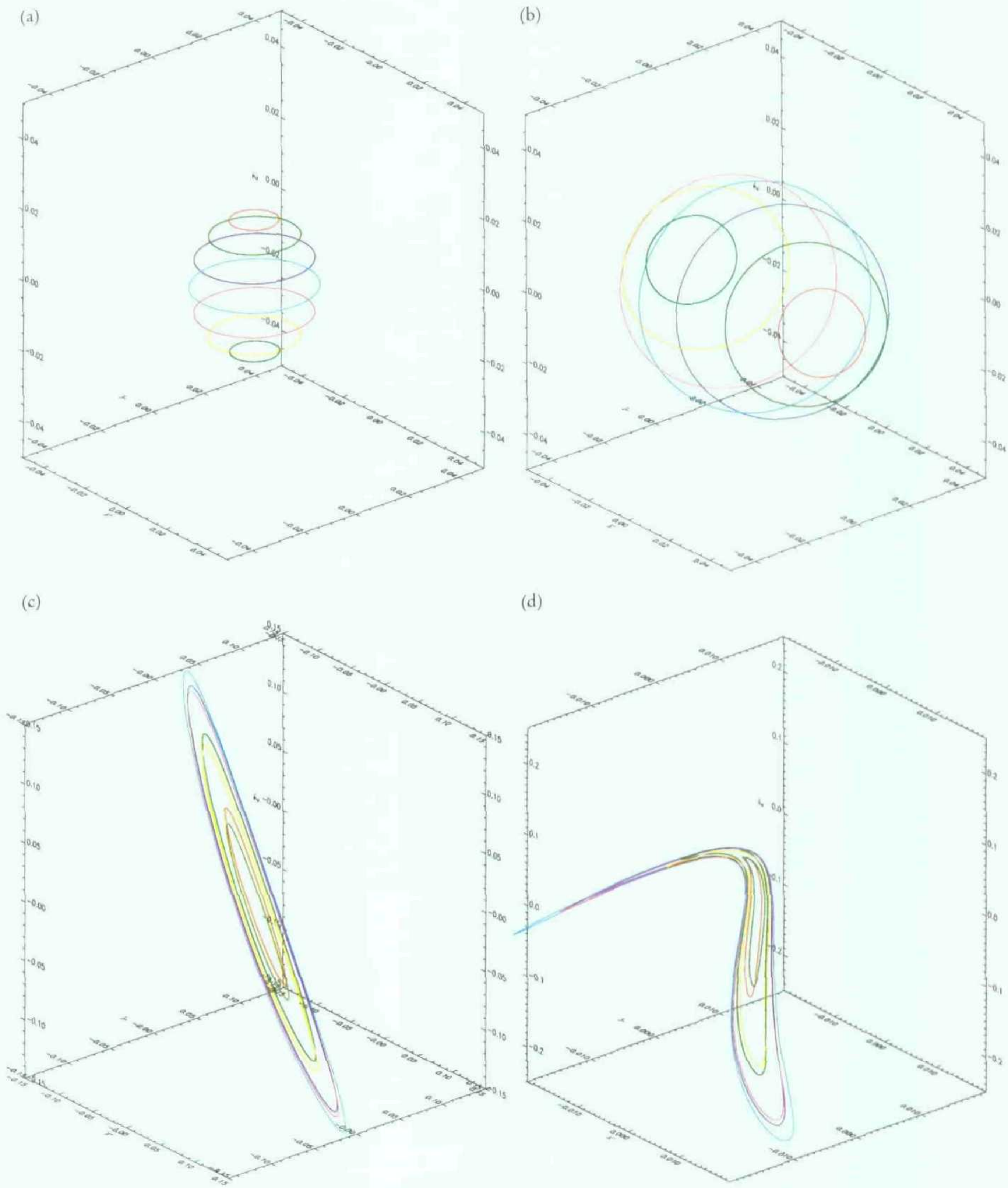


Fig. 6. Panels a)–d) show the (nonlinear) time evolution in phase space of a material surface (taken as a sphere at initial time) under the dynamics of the LORENZ (1984) system (same parameter values as in Fig. 5), at nondimensional times 0, 0.5, 0.9, and 3.5, respectively. Material surfaces are shown schematically through a set of seven differently colored "latitude circles" lying on the surfaces of the three-dimensional objects. Note that the perspective of viewing leads to a distortion of these objects (the degree of that distortion becomes clear from panel a) displaying the initial sphere). The initial sphere is centered at $(X, Y, Z)=(2, -1, 0)$ and has a radius of 0.02 units. Note the change in scale by a factor of three (in all three directions) going from panels a), b) to c), and the change by a factor of 0.1 (directions X' and Y') and 0.6^{-1} (direction Z') going from panel c) to d). Panel e) shows the time evolution of the center of the initial sphere that is subtracted from the surfaces shown in a)–d) before plotting. Red dots indicate the times relevant for panels a)–d).

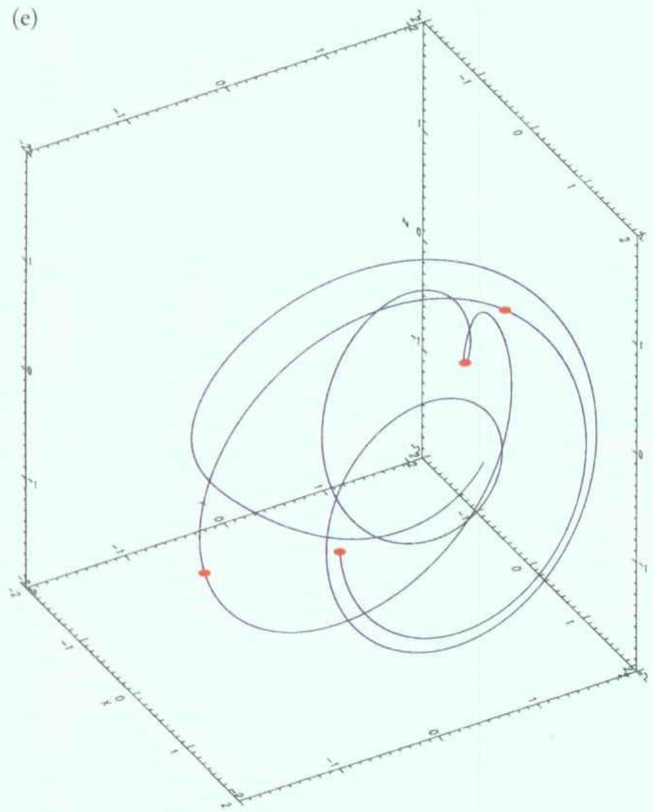


Abb. 6. Bilder a)–d) zeigen die (nichtlineare) zeitliche Entwicklung im Phasenraum einer materiellen Fläche (spezifiziert als Kugel zur Anfangszeit) unter der Dynamik des LORENZ (1984) Systems (mit denselben Parameterwerten wie in Abb. 5), zu den dimensionslosen Zeiten 0, 0,5, 0,9 und 3,5. Die materiellen Flächen sind schematisch gezeigt durch einen Satz von sieben verschiedenfarbigen „Breitenkreisen", welche auf der Oberfläche der dreidimensionalen Objekte liegen. Man beachte die Verzerrung der Objekte durch die Perspektive der Zeichnung (der Grad dieser Verzerrung wird deutlich durch die Darstellung der Kugel in Bild a)). Die Kugel zur Anfangszeit hat den Mittelpunkt (X, Y, Z)=(2, −1, 0) und einen Radius von 0,02 Einheiten. Man beachte die Skalenänderung um einen Faktor drei (in allen drei Richtungen) zwischen Bildern a), b) und c) und die Skalenänderung um einen Faktor 0,1 (Richtungen X' und Y') und $0{,}6^{-1}$ (Richtung Z') zwischen Bildern c) und d). Bild e) zeigt die zeitliche Entwicklung des Mittelpunkts der Kugel zum Anfangszeitpunkt, welcher von den Flächen, die in Bildern a)–d) gezeigt sind, vor dem Zeichnen subtrahiert wurde. Rote Punkte bezeichnen die Zeitpunkte, die für Bilder a)–d) relevant sind.

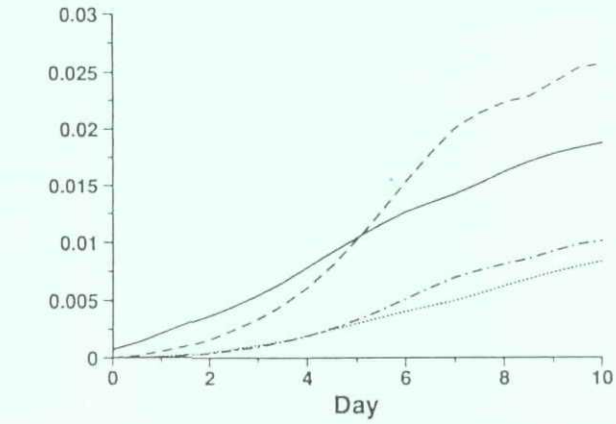


Fig. 7. Accuracy of the tangent-linear approximation as a function of lead time. Shown are the magnitudes of mean Taylor expansion coefficients for a forecast started from initial condition obtained by perturbing the analysis from 28 December 1992. The magnitude of these coefficients is computed in a root-mean-square sense in the physical domain. Solid line: first-order term, dashed: second-order term, dotted: third-order term, chain-dashed: fourth-order term. The time evolution is studied with a primitive-equation T63L19 model; initial perturbation structures are obtained using fast growing perturbations (through singular vectors) (figure taken from Buizza 1995a).

Abb. 7. Genauigkeit der tangentenlinearen Approximation als Funktion des Vorhersagezeitraums. Gezeigt sind die Werte gemittelter Taylor Entwicklungskoeffizienten für eine Vorhersage, die von einer Anfangsbedingung gestartet wurde, die sich durch Störung der Analyse vom 28. Dezember 1992 ergab. Die Werte dieser Koeffizienten sind im Gitterpunktsbereich als Wurzel aus mittleren quadratischen Größen berechnet. Durchgezogene Linie: Term erster Ordnung, strichliert: Term zweiter Ordnung, punktiert: Term dritter Ordnung, strichliert-punktiert: Term vierter Ordnung. Die zeitliche Entwicklung wurde mit einem T63L19 Modell basierend auf den primitiven Gleichungen untersucht; die Anfangsstörungen ergeben sich aus schnell wachsenden Störungen (aus den singulären Vektoren) (Abb. aus BUIZZA 1995a).

to some extent dependent on the norm used to measure the magnitude of perturbations. In that context it is worth noting that TANGUAY et al. (1995) have reported a strong scale dependence of the validity of the tangent-linear approximation. Very little evidence exists todate as to the validity of the tangent-linear approximation when moist convective processes are considered (see, e. g., ERRICO and EHRENDORFER 1995, EHRENDORFER et al. 1996).

In the formalization presented in the appendix, attention is given to the optimal prediction of the forecast error covariance matrix $\mathbf{S}_t$, summarizing second-moment information about the time-evolved pdf ρ_t. It is outlined there that $\mathbf{S}_t$ can, assuming the validity of tangent-linear error growth, be approximated in an optimal way through time-evolving appropriately constructed SVs. Optimality refers in this context to the reconstruction of a maximum fraction of variance, given a prespecified number of allowable (tangent-linear) model integrations. The appropriate construction of these SVs must take into account the initial error covariance structure (or, analysis error covariance matrix) (e. g., FISHER and COURTIER 1995, HOUTEKAMER 1995, PALMER 1996, EHRENDORFER and TRIBBIA 1997). It is shown that under these conditions the time-evolved SVs are the empirical orthogonal functions of the tangent-linearly approximated form of the forecast error covariance matrix at time t, denoted $\hat{\mathbf{S}}_t$. As such, when a given number of time-evolved SVs is used to reconstruct $\hat{\mathbf{S}}_t$ one can be assured that a maximum possible fraction of the total variance in $\hat{\mathbf{S}}_t$ is recovered.

The effectiveness of this procedure to predict the forecast error covariance matrix through SVs (in the following referred to as SV method) is illustrated for the variance of Z of the Lorenz system (5.1) in Fig. 8 (from EHRENDORFER

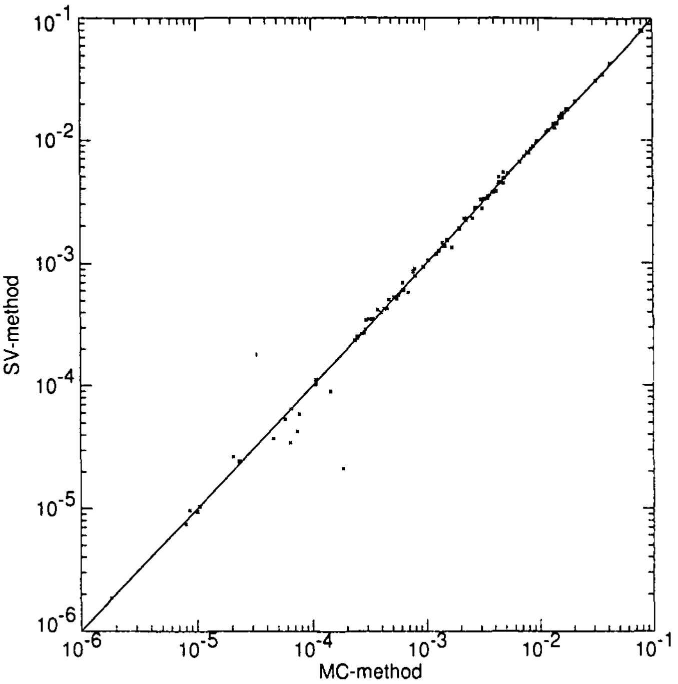


Fig. 8. Illustration of the SV method for the LORENZ (1984) system (same parameter values as in Fig. 5). Shown is the agreement between the variance (at nondimensional time of two units) of the third variable in the Lorenz system (i. e., third diagonal entry of the forecast error covariance matrix) as obtained through the MC approach (abscissa, sample size 10^5) and the SV method (ordinate, two time-evolved SVs used for reconstruction). Each point corresponds to one of the one hundred cases investigated, differing by the choice of the mean of the initial probability density function. The double-logarithmic nature of both axes is chosen to emphasize the wide range of variances encountered (initial variances are 0.01^2 in each of the one hundred cases) (figure taken from EHRENDORFER and TRIBBIA 1997).

Abb. 8. Illustration der SV Methode für das LORENZ (1984) System (mit denselben Parameterwerten wie in Abb. 5). Gezeigt ist die Übereinstimmung zwischen der Varianz (zum dimensionslosen Zeitpunkt zwei) der dritten Variablen im Lorenz System (d. h., die dritte Größe auf der Diagonale der Vorhersagefehler-Konvarianzmatrix) berechnet durch die MC Methode (Abszisse, Stichprobenumfang 10^5) und durch die SV Methode (Ordinate, zwei zeitlich entwickelte SVs zur Rekonstruktion verwendet). Jeder Punkt entspicht einem aus einhundert untersuchten Fällen, welche sich durch die Wahl des Mittelwerts der anfänglichen Wahrscheinlichkeitsdichtefunktion unterscheiden. Die doppellogarithmische Einteilung beider Achsen verdeutlicht die weite Spanne von entstehenden Varianzen (Anfangsvarianzen sind $0{,}01^2$ in jedem der einhundert Fälle) (Abb. aus EHRENDORFER und TRIBBIA 1997).

and TRIBBIA 1997). For this three-dimensional system, the first two SVs were used to construct $\hat{S}_t$, for t equal to two nondimensional time units. The performance of this method was tested for one hundred different initial pdfs (differing by the choice of the mean of the pdf), and verified against corresponding MC estimates of the variance of Z with very high sample size. The good agreement of the two estimates emphasizes the usefulness of the approach based on SVs. This good agreement is also a reflection of the fact that the tangent-linear approximation is quite accurate over this time period of two time units for the size of perturbations to be expected on the basis of the initial pdfs (see also Figs. 5 and 6).

The operational system

The above-outlined result that appropriately constructed SVs evolve into the EOFs of the forecast error covariance matrix at time t (given that the tangent-linear approximation is valid), provides a rather strong motivation for basing a sampling procedure in an ensemble prediction system on the leading SVs. This argument provides one of the rationales for the strategy pursued at ECMWF (see also BETTI and NAVARRA 1995). It is important, however, to keep in mind the limitations imposed on this strategy by making the tangent-linear approximation. Reference is made here to the discussion in section 6 for the wide range of potential applicability of SVs in other situations, such as, for example, in a generalized approach to stability analysis. At this point, it may suffice to note that error growth described through SVs may — over finite times — be much faster and quite different from conventional normal-mode exponential growth. In addition, the time-dependent nature of the operator **L** allows for implicit consideration of the specific synoptic situation in the examination of error growth.

In the currently operational version of the ensemble prediction system at ECMWF (at the time of writing, February 1997), a set of 50 (nonlinear) model integrations (32 up to December 1996), starting from different initial states, is carried out up to a forecast duration of ten days. These integrations are performed with a truncated form of the operational model at triangular resolution of T159 with 31 vertical levels (T159L31; T63L19 up to December 1996; BUIZZA 1996, and personal communication). In addition to these 50 perturbed integrations, one T159L31 integration is carried out from the unperturbed initial state (the same initial state that is used at the finer resolution for the operational T213L31 forecast). The initial conditions for the 50 perturbed forcasts are obtained by adding and subtracting to the operational analysis 25 orthogonal perturbations defined as linear combinations of leading SVs. For various details on the operational configuration of the ensemble prediction system at ECMWF, as well as most recent changes, reference is made to MOLTENI et al. (1996) and BUIZZA (1996, 1997); refer also to the information given in Table 1.

The linear combinations of these leading SVs are created such that they are spatially more uniform than the individual SVs themselves; in addition, these perturbations are scaled such that the perturbation amplitude is comparable to the size of analysis errors (some experimentation has been carried out with increasing and decreasing the initial perturbation amplitudes by modest amounts). This form of the operational version of the ensemble prediction system has been operational since December 1992 with various modifications; for example, initially ensemble forecasts were prepared only three times a week, but since May 1994

Table 1. Some characteristics of the presently operational ensemble prediction systems at ECMWF and NCEP.
Tabelle 1. Einige Merkmale der derzeit operationellen Ensemble-Vorhersage-Systeme an ECMWF und NCEP.

Ensemble prediction system characteristics	ECMWF (European Centre for Medium-Range Weather Forecasts)	NCEP (National Centers for Environmental Prediction)
operational since	19 December 1992	7 December 1992
generation of initial perturbations	finite-time optimization over 48 hours (36 hours until August 1994)	breeding cycle with rescaling every 24 hours (6 hours before 1994)
number of ensemble members (at time of writing — February 1997)	1 control (T213L31) 1 control (T159L31) 50 perturbed integrations (T159L31)	2 controls (T126 and T62) at 00 UTC 10 perturbed integrations (T62) at 00 UTC 1 control (T126) at 12 UTC 4 perturbed integrations (T62) at 12 UTC
major changes	— 1 May 1994: multiple integrations daily (weekends only before then) — 14 March 1995: SVs computed at T42L19 (T21L19 before then) — December 1996: SVs computed at T42L31 — December 1996: 50 perturbed integrations (32 integrations at T63L19 before then)	— seven independent breeding cycles since March 1994 (one cycle before then)
selected references	MOLTENI et al. (1996) BUIZZA (1994a, 1996, 1997)	TOTH and KALNAY (1993, 1996a, 1997) TRACTON and KALNAY (1993) TOTH et al. (1997)

ensemble forecasts have been run daily. Also, the resolution of the models used for SV computations and subsequent time integration has increased substantially since then (see Table 1, and BUIZZA et al. 1997a, b; HARTMANN et al. 1995). The present form of the operational ensemble prediction system is the result of careful experimentation with a three-level quasi-geostrophic and a primitive-equation model. Results from these studies, reported by MUREAU et al. (1993), and BUIZZA et al. (1993) indicated that the construction of perturbations on the basis of SVs for the purpose of ensemble prediction was more successful than other techniques considered (see also BUIZZA 1994b, 1995b).

At the time of writing, the computation of the SVs is carried out with a T42L31 (T42L19 up to December 1996) tangent-linear version of the operational forecast model (with no moist physics considered) using the so-called total-energy norm (see, e. g., BUIZZA et al. 1993) at both initial and final times for an optimization time interval of 48 hours. This choice for the norm specifies the operators **M** as well as **V** (see appendix), where **M** (at the final time) includes a projection operator that is used to restrict perturbation growth to the northern hemisphere extratropics (i. e., latitude > 30° N) (see BUIZZA 1994a). This tangent-linear model version — that is evaluated along a predicted trajectory — includes only a very crude representation of the physical processes present in the full model version (only surface drag and vertical diffusion are included). Also, using the total-energy norm at the initial time is clearly quite different from using the inverse of the analysis error covariance matrix to constrain the SVs initially (see eq. (A.3) in the appendix). Even though there are some indications that at least some of the properties of **V** are included in the total-energy norm (MOLTENI et al. 1996), important properties, such as location and flow dependence, or differences between rotational and gravitational flow components are neglected. Research is currently underway to understand technical and conceptual issues in dealing with the problem of including (an approximation to) the analysis error covariance matrix in the SV computations (see, e. g., FISHER and COURTIER 1995, BARKMEIJER 1996, and personal communication).

The computation of the SVs is carried out by partly solving the symmetric eigenproblem (A.4) (see appendix). In that context there are two technical points worth mentioning that relate to the fact that the dimension of this eigenproblem is so large (even for a T42L31 model) that explicit matrix representation of the operators is prohibitive. First, in-core eigensystem solvers cannot be used to deal with this eigenproblem. Instead, it is necessary to employ semi-direct methods, such as the Lanczos procedure (see, e. g., LANCZOS 1950, STRANG 1986, WATKINS 1993). Robust computer implementations of the Lanczos procedure for (extremely) large symmetric eigenproblems are available (e. g., SIMON 1984, GRIMES et al. 1994), and are also at use in the ECMWF ensemble prediction system. The Lanczos algorithm proceeds iteratively, and will provide the leading portion of the eigenspectrum of a symmetric matrix **H**, given that the action of **H** on an arbitrary vector can be computed for an arbitrary vector. Second, the large dimension of the eigenproblem requires the representation of the operators composing **H** in coded form. This applies in particular for the resolvent $\mathbf{R}_t$, as well as for its *adjoint* $\mathbf{R}_t^{\mathrm{T}}$. It is therefore necessary to have available in coded form the adjoint tangent-linear resolvent of the model under consideration. Constructing such code for given $\mathbf{R}_t$ is conceptually straightforward, but time-consuming and error-prone. Consequently, major efforts are currently directed at the automatic generation of adjoint codes (e. g., THACKER 1990, BERZ et al. 1996).

Products and validation

The basic product available on a daily basis from the ensemble prediction system at ECMWF is clearly the set of 50 perturbed model integrations, together with the unperturbed integration, as well as the routinely available T213L31 forecast. These perturbed integrations carry information about the time-evolved pdf (MOLTENI et al. 1996; see also PETROLIAGIS et al. 1997, EMMRICH 1996). Displaying (or, summarizing) and validating this information is, by itself, a major undertaking. Among the displays that are routinely disseminated to the meteorological services of the ECMWF member states are stamp maps, clustering results, probability plumes, and probability maps. The information in these displays is generated from the time-evolved ensemble. On the *stamp maps* the 50 (or 52) forecasts for 500 hPa geopotential height can be shown on one or two sheets of paper; thus, they give a quick overview of all the time-evolved ensemble members.

Clusters formed on the basis of the 500 hPa height forecasts provide condensed information of different flow types present in the time-evolved ensemble. An example of such cluster information is given in Fig. 9 valid at a lead time of 6 days for 24 December 1996 (12 GMT). The flow situations in the three clusters are not dramatically different, but marked differences exist. Also, it is of interest to note that the populations in the three clusters are not widely different.

The *probability plumes* serve to give an impression of the dispersion among the individual ensemble members in a probabilistic way throughout the forecast range; as an example, for the temperature at 850 hPa at a given location, these plumes, give the (normalized) probabilities for this

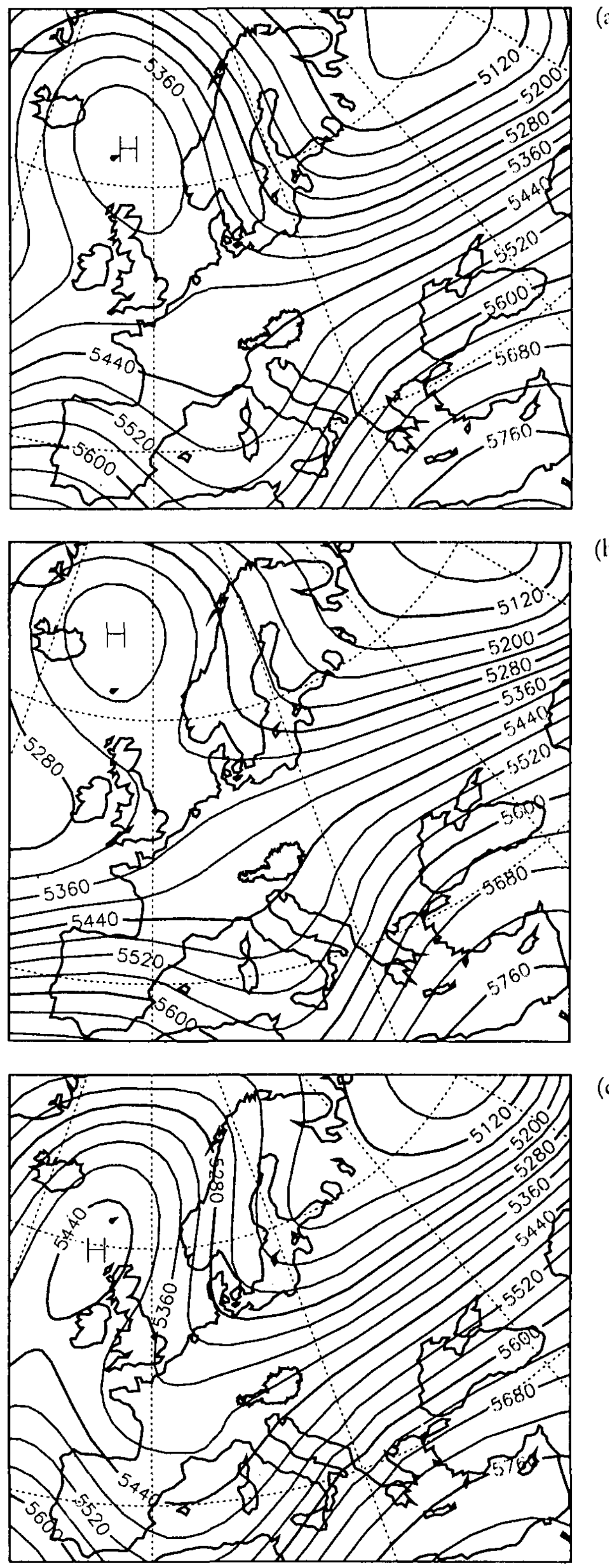


Fig. 9. Representation of the 51 members (as operational at the time of writing) of the ECMWF ensemble prediction system in terms of three clusters of 500 hPa geopotential height fields valid for a lead time of 6 days (verifying at 24 December 1996, 12 GMT). Individual forecasts have been carried out with a T159L31 version of the operational ECMWF model with initial perturbations constructed from T42L31 SVs. Cluster populations are 19 (cluster 1, panel a), 18 (cluster 2, panel b), 14 (cluster 3, panel c) (figure provided by Austrian Weather Service Zentralanstalt für Meteorologie und Geodynamik).

Abb. 9. Darstellung der 51 Mitglieder (wie operationell zum Zeitpunkt des Schreibens) des ECMWF Ensemble-Vorhersage-Systems anhand von drei Clustern des 500 hPa Geopotential Feldes gültig für einen Vorhersagezeitraum von 6 Tagen (Verifikationszeitpunkt ist der 24. Dezember 1996, 12 GMT). Die einzelnen Vorhersagen wurden mit der T159L31 Version des operationellen ECMWF Modells ausgeführt, wobei die anfänglichen Störungen auf der Grundlage von T42L31 SVs erzeugt wurden. Besetzung der Cluster ist 19 (Cluster 1, Bild a), 18 (Cluster 2, Bild b), 14 (Cluster 3, Bild c) (Abb. zur Verfügung gestellt durch Zentralanstalt für Meteorologie und Geodynamik).

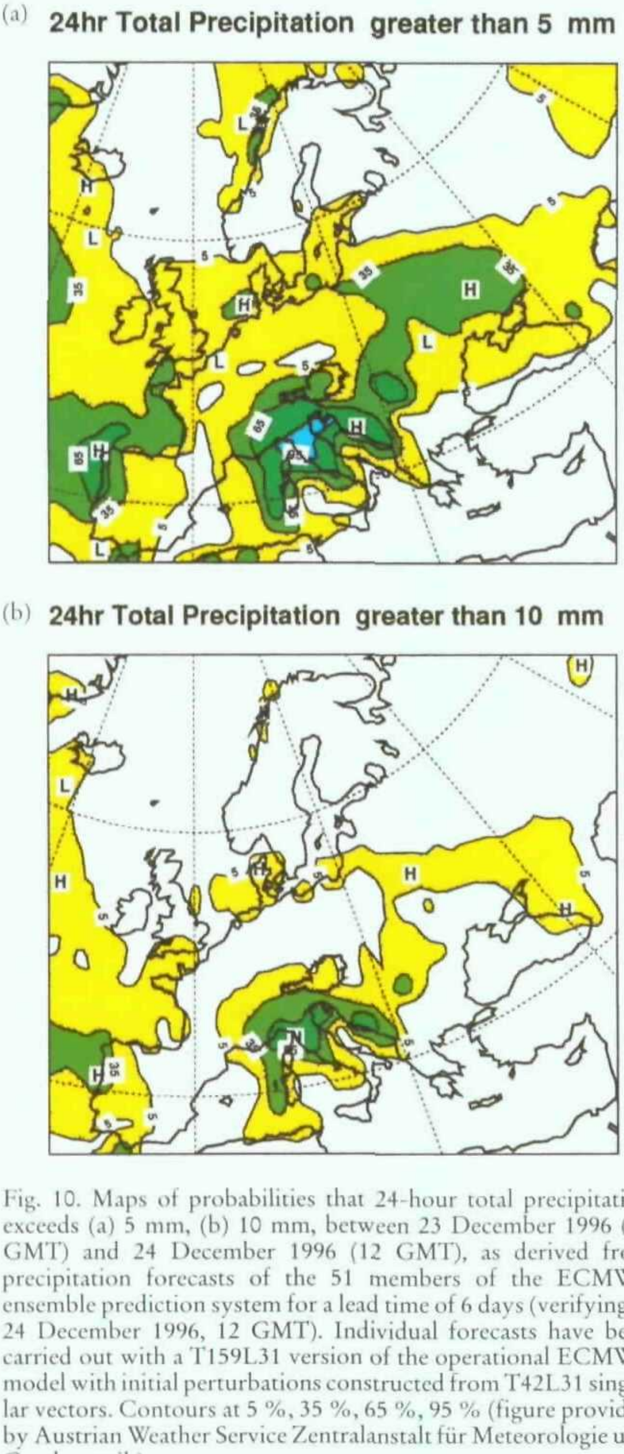


Fig. 10. Maps of probabilities that 24-hour total precipitation exceeds (a) 5 mm, (b) 10 mm, between 23 December 1996 (12 GMT) and 24 December 1996 (12 GMT), as derived from precipitation forecasts of the 51 members of the ECMWF ensemble prediction system for a lead time of 6 days (verifying at 24 December 1996, 12 GMT). Individual forecasts have been carried out with a T159L31 version of the operational ECMWF model with initial perturbations constructed from T42L31 singular vectors. Contours at 5 %, 35 %, 65 %, 95 % (figure provided by Austrian Weather Service Zentralanstalt für Meteorologie und Geodynamik).

Abb. 10. Karten der Wahrscheinlichkeit für das Ereignis, daß der 24-stündige Niederschlag (a) 5 mm, (b) 10 mm überschreitet, und zwar im Zeitraum von 23. Dezember 1996 (12 GMT) bis 24. Dezember 1996 (12 GMT); diese Wahrscheinlichkeiten wurden aus den Niederschlagsvorhersagen der 51 Mitglieder des ECMWF Ensemble-Vorhersage-Systems gültig für einen Vorhersagezeitraum von 6 Tagen abgeleitet (Verifikationszeitpunkt ist der 24. Dezember 1996, 12 GMT). Die einzelnen Vorhersagen wurden mit der T159L31 Version des operationellen ECMWF Modells ausgeführt, wobei die anfänglichen Störungen auf der Grundlage von T42L31 SVs erzeugt wurden. Isolinien bei 5 %, 35 %, 65 %, 95 % (Abb. zur Verfügung gestellt durch Zentralanstalt für Meteorologie und Geodynamik).

temperature to fall into 1 K intervals. Finally, the *probability maps* are obtained by counting in a frequentist approach the number of occurrences of a given event within the time-evolved ensemble (e.g., the occurrence of precipitation greater than a certain amount, temperature anomaly greater or smaller than a certain threshold). An example of such an operationally produced probability map for 24 December 1996 (12 GMT) is shown in Fig. 10 for two different precipitation thresholds, again at a lead time of 6 days.

One characteristic property that may be derived from an ensemble of forecasts is the spread within the ensemble (the term spread is described more precisely below). One would like, ideally, to be able to assess the skill of the control forecast given the ensemble members, or, specifically, given the spread of the ensemble (see section 3). Therefore, since one might expect that small spread of the ensemble is related to high skill of the control forecast, investigating the relationship between some measure of spread within the ensemble and the a posteriori skill of the control forecast represents an important aspect in the design and study of an ensemble prediction system. However, for the reasons discussed in section 3c, such relationships must be expected

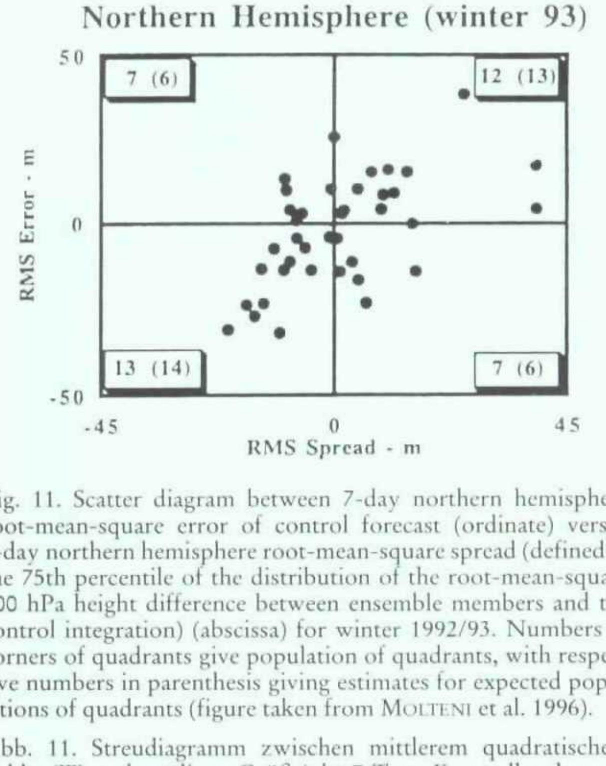


Fig. 11. Scatter diagram between 7-day northern hemisphere root-mean-square error of control forecast (ordinate) versus 7-day northern hemisphere root-mean-square spread (defined as the 75th percentile of the distribution of the root-mean-square 500 hPa height difference between ensemble members and the control integration) (abscissa) for winter 1992/93. Numbers in corners of quadrants give population of quadrants, with respective numbers in parenthesis giving estimates for expected populations of quadrants (figure taken from MOLTENI et al. 1996).

Abb. 11. Streudiagramm zwischen mittlerem quadratischem Fehler (Wurzel aus dieser Größe) der 7-Tages Kontrollvorhersage auf der Nordhemisphäre (Ordinate) und mittlerer quadratischer Streuung auf der Nordhemisphäre (Wurzel aus dieser Größe) (definiert als 75stes Perzentil der Verteilung der mittleren quadratischen Differenzen (Wurzel aus dieser Größe) der Höhe der 500 hPa Fläche zwischen Ensemble Mitgliedern und der Kontrollintegration) (Abszisse) für Winter 1992/93. Die Zahlen in den Ecken der Quadranten geben die Besetzung der Quadranten an, mit den Zahlen in Klammern als Schätzwerte für zu erwartende Besetzungen der Quadranten (Abb. aus MOLTENI et al. 1996).

to be complex and far from trivial. As an example, Fig. 11, taken from MOLTENI et al. (1996), shows the relationship between spread and skill at a seven-day lead time for the northern hemisphere during winter 1993. Here spread is defined as the 75th percentile of the distribution of the root-mean-square 500 hPa height difference between the perturbed ensemble members and the control integration. It becomes obvious that the degree of correspondence between ensemble spread and skill of the control forecast is not high. However, in view of sampling issues, as well as the problem of model errors, it is unclear how the functional form of these relationships should even be expected to look. Further discussion of this issue and of the difficulties involved here can be found, for example, in TRACTON et al. (1989), BRANKOVIĆ et al. (1990), BARKER (1991), HOUTEKAMER (1992, 1993), BARKMEIJER (1993), BARKMEIJER et al. (1993), ROYER (1993), WOBUS and KALNAY (1995), ANDERSON and STERN (1996), and ANDERSON (1996c).

The validation of ensemble forecasts (e. g., STRAUSS and LANZINGER 1996a, b; LANZINGER and STRAUSS 1996; ZHU et al. 1996) is still an area in which it is to some extent unclear how to proceed (e. g., MOLTENI et al. 1996; see also section 6). Indeed, the evaluation of different types of forecasts itself is an area of ongoing active research (e. g., MURPHY 1997). In the validation of ensemble forecasts many of the verification methods applicable for probabilistic forecasts are also applicable. For example, when ensemble forecasts are converted into probabilistic forecasts, such as probability of precipitation forecasts (see, e. g., ANDERSON 1996b), they may be verified using, for example, the Brier score (BRIER 1950), or standard reliability diagrams. An example of such a reliability diagram for probabilistic temperature forecasts (for the event of a cold temperature anomaly of more than 4 K) at a lead time of six days derived from operational ensemble forecasts during spring 1995 is given in Fig. 12. In this diagram the observed conditional frequency of the event in question — conditioned on those cases when a specified probability of the event was used — is plotted against the relevant forecast probabilities. Ideally, for perfect forecasts this conditional observed frequency is equal to the forecast probability used, since then, on average, the forecast probability and the observed frequency are the same. It can be seen from Fig. 12 that these probabilistic temperature forecasts exhibit a substantial amount of overforecasting for high forecast probabilities used, in the sense that forecast probabilities used are too high compared to observed frequencies. However, the accuracy of these forecasts is such that they present a marked improvement over climatological forecasts (the Brier skill score is given here as 32.5 %). For the extremely high degree of reliability that is presently achieved for subjective precipitation probability forecasts (based on numerical guidance) in the United

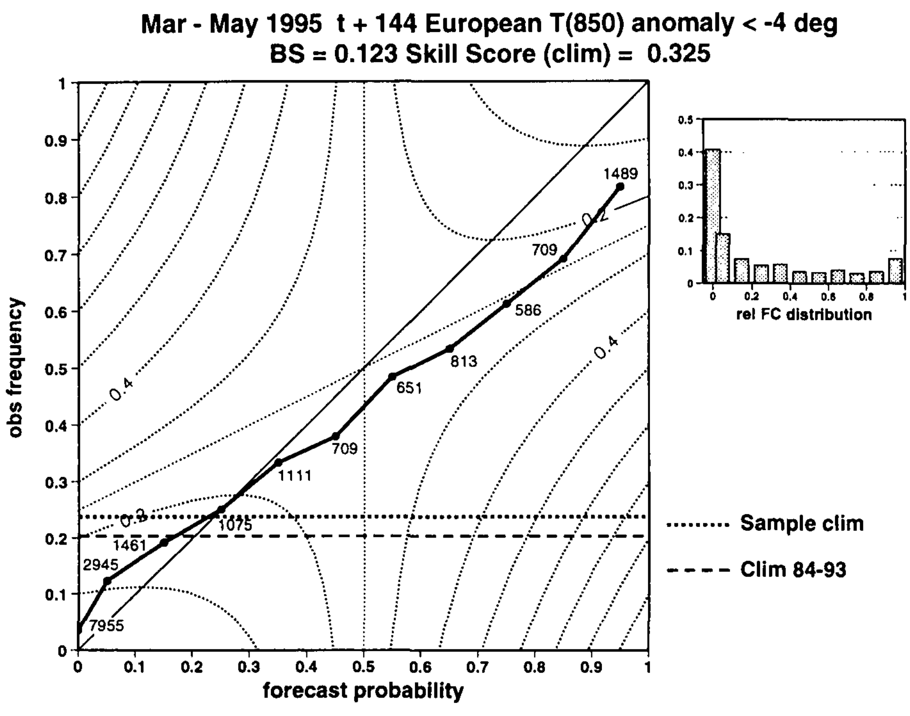


Fig. 12. Reliability diagram for 6-day probability forecasts from the ECMWF ensemble prediction system for 850 hPa temperature cold anomalies of more than 4 K for spring 1995. Verification is performed against the analysis over the European area. Numbers next to points on reliability curve indicate in absolute terms how many times that specific probability was forecast (these numbers are shown in the small histogram in relative terms). The two horizontal lines show sample (dotted) and long-term (dashed) observed frequency of the event. Isolines in the diagram show the Brier score. The overall Brier score of these probability forecasts is 0.123, resulting in a skill score of 32.5 % (above climatology). Note the degree of overforecasting for higher probabilities (figure taken from STRAUSS and LANZINGER 1996a).

Abb. 12. Verläßlichkeitsdiagramm für 6-Tages Wahrscheinlichkeitsvorhersagen aus dem ECMWF Ensemble Vorhersage System für 850 hPa Temperaturanomalien (zu kalt) größer als 4 K für Frühjahr 1995. Die Verifikation wurde anhand der Analyse über Europa durchgeführt. Die Zahlen neben den Punkten auf der Verläßlichkeitskurve zeigen in absoluter Weise an, wie oft eine gewisse Wahrscheinlichkeit vorhergesagt wurde (diese Zahlen sind in dem kleinen Histogramm relativ dargestellt). Die beiden horizontalen Linien zeigen Stichproben (punktiert) und Langzeit (strichliert) beobachtete Häufigkeit des Ereignisses. Isolinien im Diagramm zeigen das Brierscore. Das gesamte Brierscore dieser Wahrscheinlichkeitsvorhersagen ist 0,123, woraus sich ein Geschick von 32,5 % ergibt (gegenüber der Klimatologie). Man beachte den Grad der Übervorhersage für größere Wahrscheinlichkeiten (Abb. aus STRAUSS und LANZINGER 1996a).

States, reference is made at this point to MURPHY and WINKLER (1984) (their Fig. 2).

A suggestion for validation based on the principle of indistinguishability of the ensemble members and relevant observations has been proposed by O. TALAGRAND (1996, personal communication). This suggestion is based on the idea that an observed value should be indistinguishable from the corresponding values produced by the ensemble. In other words, one may ask the question whether there is a systematic way in which the observed values differ from the values of the ensemble. This question may be investigated by checking whether the observed value tends to fall into a preferred place within the ensemble values. If that is the case, then the observational distribution is different from the distribution of the ensemble values. A plot of the relative frequencies with which given places in the ensemble distribution are occupied by the observed values, as a function of the place (class) defined by the ensemble distribution, constitutes the so-called *Talagrand diagram.* More details on these diagrams may be found in STRAUSS and LANZINGER (1996a, b), LANZINGER and STRAUSS (1996), HARRISON et al. (1995), ANDERSON (1996b), and BUIZZA (1997). An example of such a diagram is given in Fig. 13. It becomes evident from this figure that, at present, the verifying analysis is clearly distinct from the set of ensemble members, since the end-intervals of the histogram are occupied more frequently than the inner intervals. One of the reasons that might be responsible for that behavior appears to be that the spread in the ensemble is too small.

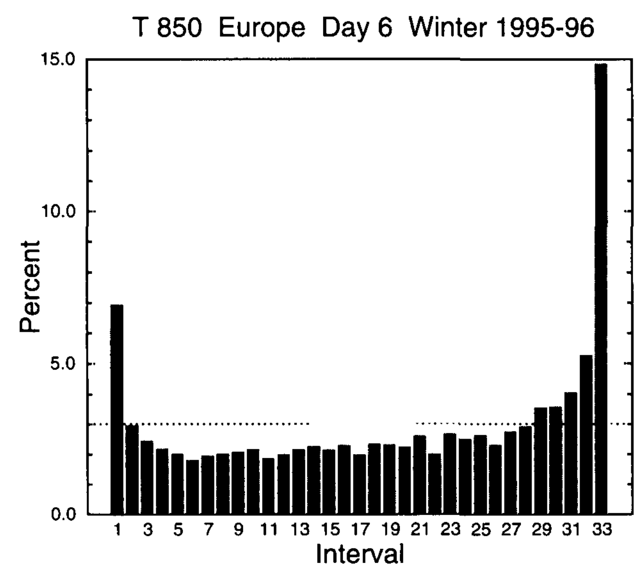


Fig. 13. Talagrand diagram for 6-days 850 hPa temperature forecasts from the ECMWF ensemble prediction system, for winter 1995/96, over Europe. Shown are the relative frequencies with which given places in the ensemble distribution are occupied by the observed values (figure taken from STRAUSS and LANZINGER 1996b).

Abb. 13. Talagrand Diagramm für 6-Tages 850 hPa Temperaturvorhersagen aus dem ECMWF Ensemble-Vorhersage-System, für Winter 1995/96, über Europa. Gezeigt sind die relativen Häufigkeiten, mit denen eine gegebene Stelle in der Ensemble Verteilung durch einen beobachteten Wert eingenommen wird (Abb. aus STRAUSS and LANZINGER 1996b).

A general result from validating ensemble forecasts is that the mean of the ensemble — as just one piece of useful information derivable from the ensemble — performs in general better than the control forecast (at least beyond a certain lead time; see, e. g., MOLTENI et al. 1996). Improving the skill of deterministic forecasts through ensemble averaging appears therefore possible, and may be of importance when one is interested at prediction beyond the deterministic range (see BAUMHEFNER 1996, ROYER 1993). However, such improvement — that effectively results from statistical filtering of unpredictable components in the forecast — provides only limited information about the time-evolution of moments higher than the first of the pdf.

Quite generally, the question of how to validate ensemble forecasts is related to the objective that one tries to achieve with the ensemble prediction system. In that sense, for example, covariances predicted on the basis of SVs along the lines discussed above (i. e., through the SV-method) can be validated against covariances predicted through the MC approach. Or, one might envisage the detailed comparison of a sparsely sampled pdf (as represented through the 50 ensemble members) with the pdf that is obtained from a MC experiment with very high ensemble size, which in effect would be an approximation to the pdf as time-evolved through the LE. A number of further comments regarding the validation of ensemble forecasts are made in section 6.

b. National Centers for Environmental Predictions

The strategy pursued at NCEP to sample in an economic and efficient way from the initial pdf is quite different from the approach based on SVs practiced at ECMWF. This NCEP technique, referred to as the *breeding method* (breeding of growing vectors) is described in detail by TOTH and KALNAY (1993, 1996a, 1997), and KALNAY and TOTH (1996) and will be discussed here in connection with selected results briefly. This method consists of the following steps: (i) an arbitrary perturbation is added to an operational analysis at a specific time; (ii) two integrations with the NWP model are carried out for one day; one started from the operational analysis, the other started from the perturbed initial state (this time period of one day is presently used operationally); (iii) the difference between the two integrations after one day is rescaled in amplitude to the size of the perturbation used in step (i) and this rescaled difference is then used to repeat the above steps. Through this approach perturbations are being bred that grow along the forecast trajectory. The breeding cycle, as described above, has a strong relationship with an operational assimilation cycle since in both cases a nonlinear NWP model is used to produce a short-range forecast. It is noted that the breeding cycle operates with operational analyses (i. e., observations enter the breeding cycle).

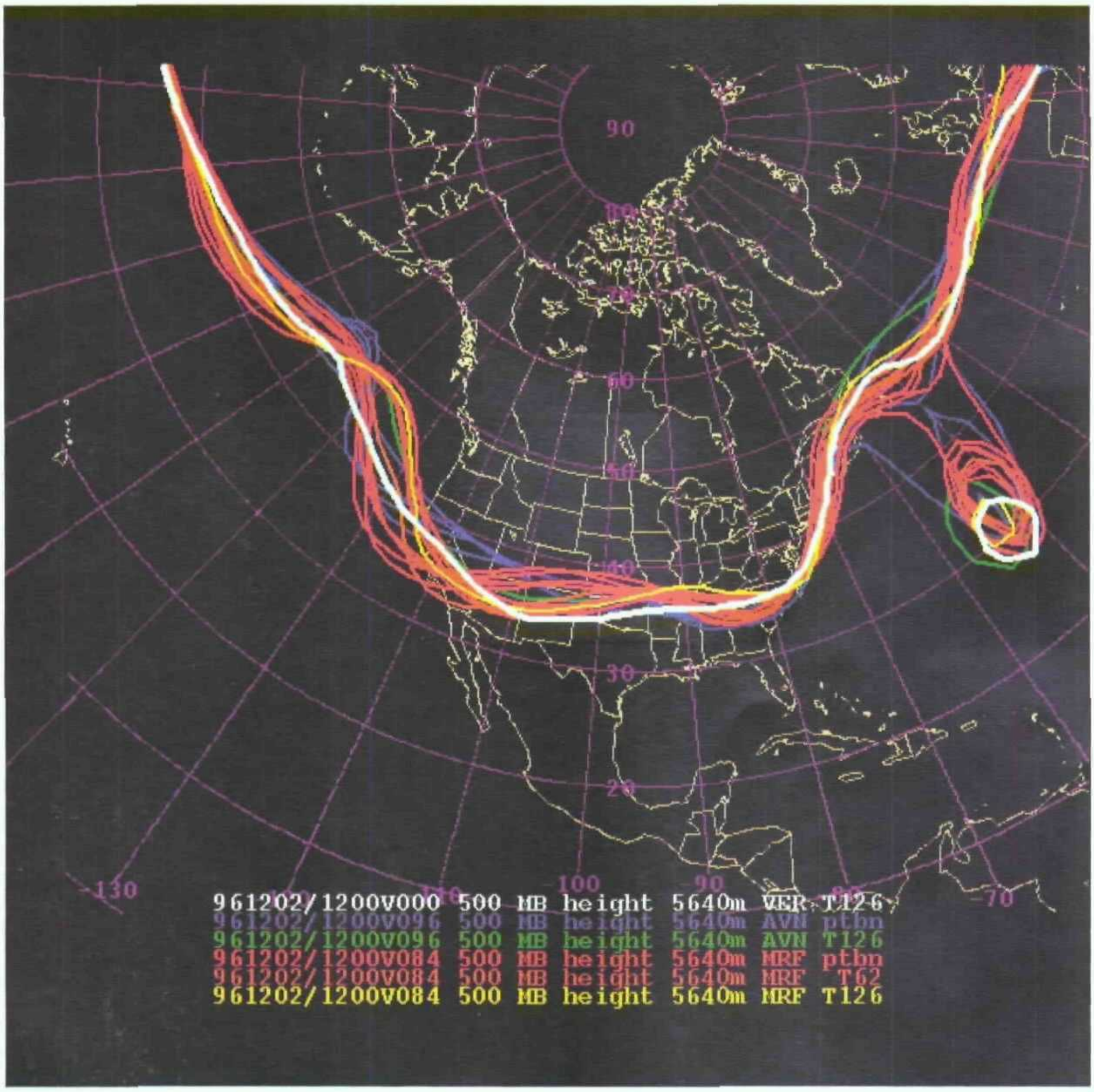
The main idea behind the breeding of growing perturbations presenting the core of the ensemble prediction system at NCEP is based on the rationale that perturbations applied to the control initial state must be chosen along the directions in which error growth will (or, is likely to) occur. With that goal in mind, as TOTH and KALNAY (1997) emphasize, the breeding technique is an alternative (to the SV based approach) methodology for estimating and representing the subspace of growing perturbations, without using adjoint equations (as required for the computation of SVs; see section 5a). In fact, the NCEP breeding method mimics the computation of the leading Lyapunov exponent and vector for a dynamical system, that characterize the divergence properties of a dynamical system (see below). Nevertheless, important differences exist between breeding vectors, SVs, and Lyapunov vectors, as LEGRAS and VAUTARD (1996) make clear.

As pointed out by BUIZZA and PALMER (1995) in an enlightening discussion on similarities and differences between SVs and breeding vectors, the perturbations obtained through the breeding technique can be interpreted as the result of the action of the tangent-linear model resolvent $\mathbf{R}_t$ (see section 5a) on a given initial perturbation. This interpretation is appropriate, since the nonlinearly evolved perturbations are rescaled to small amplitudes every 24 hours so that perturbation growth is effectively linearized. Also, if the breeding is carried out over a sufficiently long time, with the effect that a long sequence of resolvents operates on a random initial perturbation (as it does in fact happen operationally due to the repeated execution of the above steps (i) through (iii); see below), the result will converge to a variant of the so-called dominant Lyapunov vector — a locally defined property of the dynamical system (i. e., the NWP model) under consideration. It is important to mention that the incorporation of observations through the analysis cycle (see above) leads to a computation of the bred vectors along a discontinuous trajectory, while Lyapunov vectors are defined in the literature along a continuous trajectory. This dominant (or leading) Lyapunov vector may be thought of as a generalization of the dominant normal mode, when the situation of a time-independent basic state (e. g., the model state $\overline{\mathbf{X}}$ defined in the context of eq. (5.2)) is generalized to that of a time-dependent basic state (for a precise definition, see, PARKER and CHUA 1989; see also NICOLIS et al. 1995, TREVISAN and LEGNANI 1995, and VANNITSEM and NICOLIS 1997). BUIZZA and PALMER (1995) also point out that over any given 24-hour time interval the perturbation growth effected through the breeding method must be expected to be smaller than the optimized growth (see eqs. (A.2) and (A.3) in the appendix) of the leading SV, which can be attributed to the fact that the breeding vectors are computed without the use of adjoint equations. For an in-depth discussion of the relationship between the SV based approach as practiced at ECMWF and the NCEP breeding method from a dynamical systems point of view, reference is made to LEGRAS and VAUTARD (1996). Also, SZUNYOGH et al. (1997) describe a detailed comparison of both methodologies in the context of a low-resolution version of the global operational NCEP NWP model (see also ANDERSON 1996a, c; TOTH et al. 1997).

A basic assumption to be met in order that the perturbations created during the breeding process are an efficient sample from the initial pdf is that the errors in the initial NWP model state at a particular time are dominated by those instabilities of the flow that have developed over a series of previous assimilation cycles. To develop these instabilities, perturbations are carried through the breeding cycle as outlined above. To support the validity of this assumption, TOTH and KALNAY (1996a) present evidence that the breeding vectors are in fact representing analysis errors realistically which would imply that sampling from ρ_0 can be achieved through this method without explicit knowledge of the properties of ρ_0 (see also, IYENGAR et al. 1996). From that assumption, it also becomes clear that an apparent difference between the breeding method and the method practiced at ECMWF is given through the fact that the perturbations at NCEP are generated based on *past* information about the flow, whereas the SVs computed at ECMWF contain *future* information about potential instabilities of the flow. Clearly, future information is only available through the necessarily imperfect forecast model; nevertheless, incorporating future information about instabilities is likely to be advantageous.

Another difference between the methods used at both centers is that the ECMWF approach based on SVs relies on the tangent-linear approximation to solve an optimization problem for maximum error growth. No tangent-linear (and corresponding adjoint) model are needed in the breeding method. However, it may be argued that in view of the high degree of accuracy of the tangent-linear approximation for perturbation sizes and optimization times considered in the breeding process (see also, section 5a), both tangent-linear and nonlinear time evolution of errors should give rather similar results; however, refraining from the use of a tangent-linear model eliminates the possibility of explicitly maximizing error growth. Nevertheless, such similarity between tangent-linear and nonlinear error evolution has to this point been supported primarily through investigations with dry models (i. e., without moist convection or precipitation; see section 5a). To account for moist processes present in nonlinear NWP models within tangent-linear models is complicated (e. g., BUIZZA et al. 1996, EHRENDORFER et al. 1996); such a tangent-linear description is not necessary for the breeding method that evolves perturbations nonlinearly with diabatic processes included.

In general, the merits and/or weaknesses of either method must be judged on whether sampling from the initial pdf is done in an efficient way, since this sampling is the basic prerequisite for a successful representation of the time-evolved pdf. More specifically, this requirement asks, among other issues, for the explicit or implicit consideration of the analysis error covariance structure $\mathbf{V}$ in the construction of the initial perturbations (e. g., FISHER and COURTIER 1995, HOUTEKAMER 1995, EHRENDORFER and TRIBBIA 1997). At present, no conclusive answer can be given as to



NCEP Global Ensemble Forecast
start time 96112900, forecast hour 084, z 500mb = 5640m

Fig. 14. Position of the 5640 m geopotential height contour line on the 500 hPa pressure surface as predicted from the 17 ensemble members in the NCEP ensemble prediction system, at a lead time of 84 hours, verifying at 2 December 1996 (12 GMT). The yellow, orange, and green lines mark the high resolution control forecasts, red and blue lines denote the perturbed forecasts, and the white line is the verifying analysis (figure reproduced from NCEP world-wide-web ensemble products home page).

Abb. 14. Position der 5640 m geopotentiellen Höhe Isolinie auf der 500 hPa Druckfläche wie vorhergesagt durch die 17 Ensemble Mitglieder im NCEP Ensemble-Vorhersage-System, zu einem Vorhersagezeitpunkt von 84 Stunden, mit Verifikationszeitpunkt 2. Dezember 1996 (12 GMT). Linien in gelb, orange und grün markieren die hochauflösenden Kontrollvorhersagen, rot und blau zeigen gestörte Vorhersagen an und die weiße Linie ist die verifizierende Analyse (Abb. reproduziert aus dem Internet).

which of the present implementations is more effective in that context (see also, section 6).

The NCEP ensemble prediction system has been operational since December 1992. The presently operational configuration of the breeding cycle at NCEP is such that a total of seventeen ensemble integrations is available in real time over a lead time of 16 days. This number of ensemble members is the result of using a total of seven pairs of perturbations (added with opposite signs to the operational analysis), where five pairs are added at 00 UTC, and the remaining two pairs at 12 UTC. In addition, there are three control integrations, as detailed in Table 1 (see also, Fig. 1 in TOTH et al. 1996a). All integrations are carried out with the NCEP Medium Range Forecast (MRF) model at the resolutions described in Table 1. The presence of ensemble members originating at initial times 12 hours apart makes clear that the bred ensembles also exploit the properties of the lagged-average forecasting technique (see section 4). Also, the operational system is set up such that the bred perturbations are determined as the difference between the two 24-hour forecasts resulting from a positively and negatively perturbated initial state; consequently, the breeding is part of the ensemble forecasting process and requires no additional computing resources beyond those necessary to carry out the ensemble forecasts themselves.

Various products are derived from the time-evolved ensembles. A selection of these is described by TRACTON and KALNAY (1993). Some of the products are similar in format to the products available from the ensemble prediction system at ECMWF. For example, NCEP prepares cluster information and probability maps. A particularly informative display of the differences in the flow that might be expected as a result of initial state uncertainty, as predicted through the individual ensemble members, are maps showing in a composite chart the position of a given height contour (usually the 5640 m height contour) on the 500 hPa pressure surface. An example of such a *spaghetti-diagram* is shown in Fig. 14 for the ensemble forecast at a lead time of 84 hours valid for 2 December 1996 (12 GMT). It can be seen that on this occasion the contours are not substantially different with the implication that similar flow types might be anticipated on the basis of this ensemble.

As a second example of the products prepared at NCEP from the ensemble of forecasts, a probability map for precipitation exceeding two different thresholds is shown in Fig. 15. These probabilities are derived in the same frequentist fashion from the ensemble as at ECMWF. The large amount of information present in the ensemble of forecasts prohibits discussing derived products here in further detail. As a way of automatically disseminating (a selection of) the operationally prepared products to everyone interested, NCEP has set up a page accessible through the internet (address: http://sgi62.wwb.noaa.gov:8080/ens/enshome.html; see also, http://www.cdc.noaa.gov) at which charts of all (or most) operational products can be perused and/or downloaded.

Major efforts are also underway at NCEP to validate objectively the quality of the products generated through the ensemble prediction system (see also the comments in section 5a, and section 6). Some of the verification measures are similar to the ones used at ECMWF (e.g., reliability diagrams). One result confirmed at NCEP is that ensemble-mean forecasts are more skillful than the single control forecast. In concluding this subsection, reference is made to TOTH et al. (1996a, 1996b) and ZHU et al. (1996) for recent results of validating the NCEP ensemble forecasts.

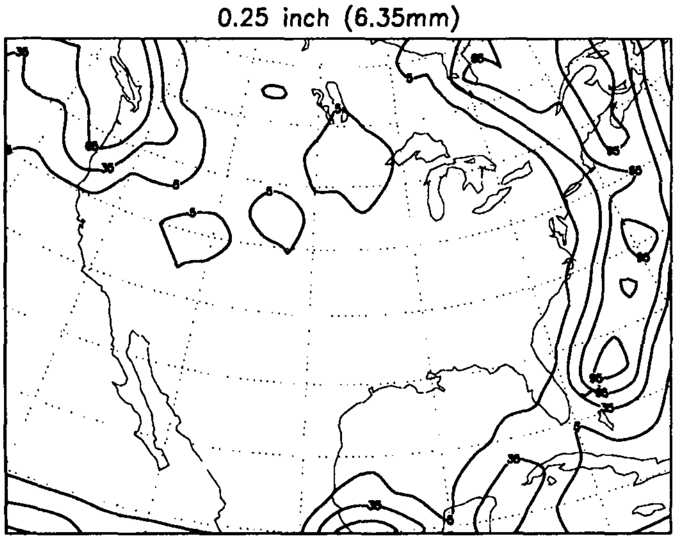
c. United Kingdom Meteorological Office

The distinguishing feature of the ensemble prediction system at the UKMO is the production of joint ensembles by combining the ECMWF ensembles with ensembles run using the UKMO Unified model (UM). One of the basic motivations behind the idea of employing different models for multiple model integrations is the fact that operational forecasts from different centres do not infrequently diverge in the medium (and also in the short) range. Further, there is at present no clear indication whether this divergence should be attributed primarily to the different operational analyses used at different centres or rather to differences in model formulations. In that context, it has been demonstrated that minor analysis differences can cause major forecast differences (e.g., RABIER et al. 1996, see also section 2); however, the role of model differences is less well documented. Through the UKMO approach to ensemble prediction, it might be expected that some of these questions could be, at least partially, resolved.

The UKMO ensemble prediction system is not run operationally, and as such is not available to forecasters in real time. However, joint ensembles were produced on a regular basis on Saturdays and Sundays in the period from 29 October 1994 to 5 March 1995, giving 38 joint ensembles in total. For each of these ensembles the ECMWF component of the experiment consisted of the 33 members of the ECMWF T63L19 ensemble with T21L19 optimal perturbations used to define the perturbed initial states (this was the standard ECMWF ensemble prediction system during that period; see Table 1). The UKMO component of the experiment consisted of 32 integrations carried out from initial states defined as the sum of the operational UKMO analysis and the perturbations used at ECMWF. Detailed documentation of this experiment, as well as a full discussion of the results may be found in HARRISON et al. (1995, 1996), and RICHARDSON and HARRISON (1996) (see also, HARRISON 1994; later experiments performed with joint ensembles are described by RICHARDSON et al. 1996). Only a few of the main results from the above experiment will be discussed below.

First, the experiment demonstrated that model integrations using perturbations computed with the ECMWF system did diverge from each other when using the UM. It became also evident that the skill of the mean of the combined ensemble was on average greater than that of the single model ensembles (on average, the anomaly correlation of the combined ensemble mean fell below 0.6 approximately half a day after individual ensemble means reached that score). Further, the combined ensemble showed an improvement over single ensembles in terms of increased reliability of probabilistic forecasts (e.g., for positive 500 hPa height anomalies). It has remained unclear, however, whether it is model or analysis differences that explain variability in forecast accuracy between different operational centers, as both seemed important, depending — among other issues — on forecast range, synoptic situation, and geographical region. Also, researching, in the context

(a)



(b)

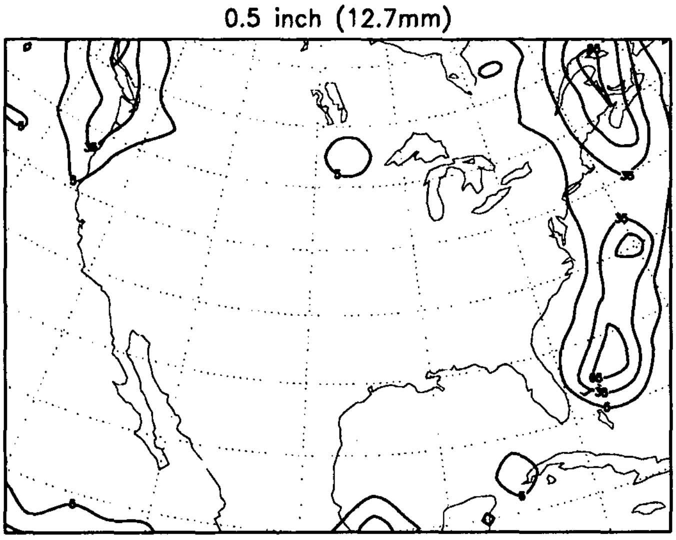


Fig. 15. Maps of probabilities for 24-hour precipitation for the period 2 December 1996 (12 GMT) to 3 December 1996 (12 GMT) to exceed (a) 6.4 mm (0.25 inch), (b) 12.7 mm (0.5 inch) derived from predictions with the 17 ensemble members in the NCEP ensemble prediction system. The lead time is 84 hours, initial time is 29 November 1996 (00 GMT) (figure provided by Zoltan TOTH and Yuejian ZHU).

Abb. 15. Karten der Wahrscheinlichkeit für das Ereignis, daß der 24-stündige Niederschlag (a) 6,4 mm (0,25 Inch), (b) 12,7 mm (0,5 Inch) überschreitet, und zwar im Zeitraum von 2. Dezember 1996 (12 GMT) bis 3. Dezember 1996 (12 GMT); diese Wahrscheinlichkeiten wurden aus den Vorhersagen der 17 Mitglieder des NCEP Ensemble-Vorhersage-Systems abgeleitet. Der Vorhersagezeitraum ist 84 Stunden, Anfangszeitpunkt ist der 29. November 1996 (00 GMT) (Abb. zur Verfügung gestellt durch Zoltan TOTH and Yuejian ZHU).

of the experiment, the relationship between ensemble spread and forecast skill (see section 5a, and also the relevant discussion in section 3), has been inconclusive and shown that the operational prediction of forecast skill is not yet fully feasible. Finally, it became clear that the SV perturbations created at ECMWF do not capture differences between the UKMO and the ECMWF analyses; this behavior may be a reflection of the fact that these analysis did not project strongly on the growing model subspaces in the cases investigated (HARRISON et al. 1996).

UKMO has set up a *Working Group for the Development and Use of Ensemble Forecasting* to consider all aspects of ensemble prediction from its design to the communication of its results to users (RICHARDSON and HARRISON 1996). One of the recommendations of this working group has recently been to make ensemble forecasting operational at UKMO (see also HARRISON et al. 1996).

d. Recherche en Prévision Numérique

At RPN of the Canadian Atmospheric Environment Service (Dorval, Québec, Canada) a system simulation approach is used to define the ensemble prediction system. As in the systems discussed in the previous sections, the primary goal is to provide initial model states from which to perform a set of medium-range forecasts. However, at RPN, in generating these initial model states many of the components of the processing system that leads from observations to analyses (i. e., the data assimilation system) are varied, in addition to varying the input observations. Thus, in this approach the full data assimilation system is not regarded as a fixed entity, but operates in modified forms (consistent with known weaknesses of the system) on modified sets of observations. This combination of perturbing observations, as well as components of the assimilation system (in particular, the assimilating model) then results in a set of perturbed analyses which are then integrated forward in time with perturbed prediction models. Clearly, the way in which observations, assimilation system, as well as assimilating and prediction models are varied does present a crucial part in the implementation of this method. It is currently implemented such that initial ensembles obtained are largely unbiased (at least less biased than the initial ensembles obtained from techniques based on optimal perturbations or breeding). Nevertheless, the RPN system simulation approach has some resemblance to the breeding method carried out at NCEP. A description of the present state of the ensemble prediction system at RPN that is not operational at the time of writing is given by HOUTEKAMER et al. (1996a, b). Preliminary results indicate that the method may be promising for ensemble prediction, but the small number of ensemble forecasts carried out so far does not allow to substantiate any definite claims at this point.

In addition, at RPN, various careful investigations of other methods proposed for ensemble prediction have been carried out. For example, the breeding method (see section 5b) has been tested with regard to its potential to lead to an improved ensemble-mean forecast (HOUTEKAMER and DEROME 1994). HOUTEKAMER and DEROME (1995) compared the breeding method, the optimal perturbation approach, and the system simulation approach with regard to their potential to generate initial perturbations in ensemble prediction systems; all methods gave ensemble means that showed an improvement over the control integration in the context of a quasi-geostrophic model. Also, HOUTEKAMER (1995) described ensemble prediction experiments in which covariance information was used for the construction of optimal perturbations (see also section 5a).

6. Concluding remarks and related aspects

Errors in forecasts produced with NWP models result from both errors in the specification of the initial model state, as well as errors in the model formulation. While this general fact was recognized already at the beginning of this century, it has remained difficult to study separately the impact of both error sources, as well as to quantify their joint impact on the quality of NWP forecasts in specific situations. During the early period of NWP, forecast errors due to simplified model formulations were dominating the total forecast error. However, as models have improved over time due to (among other things) increased resolution and more detailed consideration of physical processes, the potential for small initial errors to contribute substantially to total forecast error due to the amplification of instabilities in the (model) atmosphere has increased. Consequently, this property of error amplification of the (model) atmosphere entails a limit on the average predictability of the atmosphere for any nonzero errors in the initial state, even in the hypothetical situation of a perfect model. Such a limit may be defined either through an atmospheric error-doubling time, or, equivalently, through the time that it takes on average for two close atmospheric states to become as different as two randomly selected observed atmospheric states (this squared average distance between two randomly selected atmospheric states is twice the climatological variance of the variables considered).

Given this potential of small initial errors to influence NWP forecast errors significantly, the question has arisen to what extent that may happen for specific flow situations. This question may, in turn, be addressed with the state-of-the-art NWP models through approaches based on the LE, such as, for example, modified MC approaches discussed in this review. In such an approach, NWP models are assumed to describe in a reasonably realistic way the time evolution of the atmosphere (and, consequently, atmospheric error growth). Viewed slightly differently, the rather successful (on many occasions) medium-range prediction of instantaneous circulation patterns with NWP models has been followed by the need to specify in advance the goodness of such forecasts. This need becomes particularly clear in light of the large degree of day-to-day variability of the a posteriori verification of such forecasts.

In this paper the main emphasis has been on providing a comprehensive review of the methodology applicable and presently applied for the prediction of the uncertainty of forecasts produced with NWP models, as well as of results in operational and quasi-operational environments obtained with such methodology. Starting with a brief review of classical predictability studies in connection with more recent assessments of atmospheric error-doubling times (currently estimated to be between 1.5 and 2 days for scales presently explicitly resolved in NWP models), the LE and its solution have been discussed as the starting point for the description of the time-evolution of initial state uncertainty given a specific atmospheric model (deterministic or, if desired, with added random components for simulating certain types of model error). Given the time-evolved pdf, the questions raised above can be answered, in principle. However, within that approach, the ignorance about specifying model error enters as a fundamental problem. The estimation of expected skill may also prove to be fundamentally difficult due to the presence of model error (in addition to problems related to sampling from the time-evolved pdf). Also, the specification of the initial pdf in realistic NWP contexts poses a problem that is far from trivial. Beyond these problems, the major restriction arising in operational contexts has been pointed out; namely, due to limits on computational resources, the size of samples from the initial pdf to be evolved over time must be taken to be (extremely) small compared to the number of degrees of freedom of the model. Nevertheless, the ratio of sample size to the degrees of freedom in the NWP model becomes more favorable, when dynamical constraints relevant for error growth are taken into account (e. g., locality of error growth, localization of baroclinic waves in storm tracks, relevant time scales). Time-evolving such samples can be viewed as an effort to mimic the time evolution of the pdf (or its most important properties), in the absence of sufficient computational resources to evaluate the solution of the LE.

Being aware of the difficulties involved with addressing the fundamental problems noted above, research at operational centers is presently focusing on the problem imposed through computational restrictions: namely, study strategies for efficiently creating initial states consistent with the ignorance about the operationally analysed state. The collection of these states is then interpreted as a sample (with certain properties, possibly non-random) from the initial pdf ρ_0 to be integrated in time with the NWP model. Such a strategy together with the computational environment to carry out multiple model integrations started from these states is referred to as an ensemble prediction system.

In this review developments for the four ensemble prediction systems at ECMWF, NCEP, UKMO, and RPN have been discussed in more datail. In that discussion several issues related to the validation of the products from such ensemble prediction systems were also addressed (see also below). In the system at ECMWF initial perturbations are generated using the general notion of finite-time most rapid error growth, as described through the SVs of the tangent-linear model resolvent. By employing the tangent-linear model resolvent, error growth is restricted to tangent-linear error growth; again, this restriction is made for reasons of computational savings and conceptual simplifications, and could be relaxed in principle.

For completeness, it has to be mentioned at this point that beyond their use within ensemble prediction systems, SVs, or more general, the notion of finite-time unstable structures, have a wide potential of further applicability. SVs were first considered in a meteorological context by LORENZ (1965) in a predictability study with a low-dimensional barotropic model. As discussed in section 5a and the appendix, they possess the property of amplifying most rapidly over a finite-time interval in a tangent-linear framework for a given measure (norm). As such, SVs are a highly useful tool in studying various questions, including questions of atmospheric predictability and growth arising from instabilities (e. g., MUKOUGAWA et al. 1991, FARRELL and MOORE 1992, MOORE and FARRELL 1993, YODEN and NOMURA 1993, PALMER 1993, MOLTENI and PALMER 1993, URBAN 1993, VUKIĆEVIĆ 1993, EHRENDORFER and ERRICO 1995, NICOLIS et al. 1995; see also NEWMAN and SARDESHMUKH 1995, PENLAND and SARDESHMUKH 1995a, SARDESHMUKH et al. 1997). SVs are optimized over finite times and obtained from a tangent-linear form of the relevant evolution equations and may thus possess very different properties than exponentially growing shape-preserving normal-mode solutions to linear (autonomous) perturbation equations, as pointed out by FARRELL (1988, 1989, 1990) in the context of models of increasing complexity. More recently, FARRELL and IOANNOU (1996a, b) have reviewed and demonstrated the applicability of SVs for such generalized stability analysis based on SVs that is not necessarily restricted to time-independent basic states used for linearizations (as required for the conventional computation of normal modes). Such a generalization of stability analysis in terms of SVs beyond shape-preserving normal-mode solutions is of particular importance since, in general, linearized operators (around time-varying basic states) are nonnormal for given inner products (i. e., they do not commute with their adjoints). This nonnormality allows for the possibility of finite-time growth faster than exponential, even if no unstable normal modes are present (see, e. g., FARRELL and IOANNOU 1993a, b for a more detailed discussion; see also, EHRENDORFER et al. 1996). The role of non-modal finite-time growth (as represented through SVs) as compared to normal-mode growth of perturbations in atmospheric dynamics (e. g., CHARNEY 1947, EADY 1949, DRAZIN and REID 1981) is presently under investigation (e. g., BORGES and HARTMANN 1992; FARRELL and IOANNOU 1993a, b; CHANG and MAK 1995; BORGES and SARDESHMUKH 1995, 1997; PENLAND and SARDESHMUKH 1995b; HARTMANN et al. 1996).

The computation of SVs for realistic NWP models requires in coded form the adjoint of the tangent-linear model resolvent (in addition to a semi-direct eigensolver; see section 5a). At this point it is appropriate to refer briefly to the large class of problems that can be addressed when adjoints of linearized models are available (see also, ERRICO 1997). Examples are sensitivity studies (e. g., ERRICO and VUKIĆEVIĆ 1992; RABIER et al. 1992; LANGLAND et al. 1995, 1996; BISHOP and TOTH 1996; EMANUEL et al. 1996; FREDERIKSEN 1997), predictability studies (e. g., OORTWIJN and BARKMEIJER 1995, EHRENDORFER and ERRICO 1995, SARDESHMUKH et al. 1997), and the efficient minimization of cost functions in variational data assimilation (e. g., TALAGRAND 1989; THÉPAUT et al. 1993, 1996; see also THOMPSON 1969). A recent prominent example giving insightful results as to the sensitivity of the forecast error to the initial model state on the basis of an adjoint model was given by RABIER et al. (1996). Their results showed that the gradient of the forecast error consisted of rapidly growing components present in the analysis error (see also BUIZZA et al. 1997a, GELARO et al. 1997). In the context of the wide applicability of adjoint models, a good overview of present research in that area may be found in the workshop reports by ERRICO et al. (1993a), PRAGÉR et al. (1995), SNYDER (1996), and BROOKS et al. (1995).

A strong connection exists between data assimilation methodology (e. g. GHIL et al. 1997, IDE et al. 1997) and ensemble prediction methods based on SVs, through their joint use of adjoint models, as well as through the fact that the results from the data assimilation system (specifically, the uncertainty of the operational analysis) provide the starting point for the ensemble prediction system (e. g., FISHER and COURTIER 1995). Further, it appears that SVs might be a highly useful tool in efficient implementations of the Kalman filter (BOUTTIER 1996). Some of these connections are discussed in recent work on data assimilation represented here through a brief selection of references (e. g., COURTIER et al. 1993, 1994; THÉPAUT et al. 1993, 1996; JÄRVINEN et al. 1996; COHN 1993; COHN et al. 1994; TODLING and COHN 1994; COHN and TODLING 1996; GHIL and TODLING 1996; BOUTTIER 1993a, 1994; PIRES et al. 1996; TANGUAY et al. 1995; MÉNARD and DALEY 1996).

The sampling strategy at NCEP, referred to as the breeding method, avoids the construction of optimally growing SVs and thus the necessity to employ a tangent-linear model. Some indications exist that the breeding cycle may allow for implicit incorporation of important properties of the initial pdf ρ_0. Nevertheless, finite-time error growth rates obtained with the breeding method are smaller than those obtained from the explicit maximization resulting in the set of SVs. Presently, the ensemble size at NCEP is about a third of the ensemble size utilized at ECMWF.

From the comparative discussion of the ensemble prediction systems at ECMWF and NCEP it has become clear that the process of validating the products from these systems is an important problem (e. g., LANZINGER and STRAUSS 1996; WILSON 1996). For example, it has recently been stated that "the validation methodology needs to be developed further. At the moment, there is no well established way of comparing ensembles from two different systems. This has serious practical implications, in particular for the testing of changes to the system configuration: change to model formulation, to the initial perturbations, increased ensemble size." (quoted from Report of the Working Group on the Use of Medium-Range Forecast Guidance, ECMWF 1996). Conventional verification procedures, such as computation of skill scores (like anomaly correlation) can be applied, but important aspects of the ensemble can be missed by using only such procedures.

Selected results from validation studies of ensemble prediction systems discussed here indicate that probabilistic forecasts derived from ensembles exhibit only a modest amount of reliability (see, e. g., the reliability diagram in Fig. 12) and that observations are, in general, clearly distinguishable from the set of time-evolved ensemble members (see, e. g., the Talagrand diagram in Fig. 13). Detailed investigations of the comparative behavior and properties of ensemble prediction systems (primarily of those at ECMWF and NCEP) are currently in progress (e. g., ZHU et al. 1996, BAUMHEFNER 1997, personal communication). Such comparisons also consider methodology appropriate for quite inexpensive generation of initial perturbations (e. g., ERRICO and BAUMHEFNER 1987).

The ensemble prediction system at the UKMO is presently not operational and provides the time evolution of the same initial perturbations with two different models. A brief overview of some of the results has been given in section 5c. It is at this point important to mention that the generation of initial states at these three centers (i. e., ECMWF, NCEP, and UKMO) results in (potentially) highly biased (in favor of the directions of most rapid error growth) samples and thus clearly does not form a random sample. Biasedness of a specific form (e. g., in favor of SVs) is quite advantageous for the purpose of efficient covariance prediction, in the sense outlined in section 5a and the appendix, but for certain other applications it may turn out to be quite undesirable. In fact, since the initial bias will supposedly prevail to some extent as the ensemble is integrated in time, care must be exercised when averaging is subsequently performed over the ensemble members. Nevertheless, by using the same initial perturbation with positive and negative sign in the initial ensemble, the biasedness should not, at least in the tangent-linear situation, lead to a negative impact in computing the mean of the time-evolved ensemble. Biased ensembles may, however, have implications for the estimation of various other statistics; this issue has recently been addressed by ANDERSON (1996c). Some of the ensemble prediction efforts at RPN are clearly directed towards moving away from biased samples (HOUTEKAMER 1996, personal communication). Also, the proposed ideas for simulating certain steps in the Kalman filter through evolving ensembles require unbiased samples (e. g., EVENSEN 1994, 1997; EVENSEN and VAN LEEUWEN 1996). Further, it is presently unclear what the role of biased initial samples is as the tangent-linear regime is gradually left and nonlinearities become important in the medium-range (e. g., BARKMEIJER 1996).

Due to space restrictions, no separate section was devoted in this review to the description of research in the areas of predictability and ensemble prediction at the Royal Netherlands Meteorological Institute (KNMI). Various conceptual ideas have been investigated by this group and are documented in, for example, HOUTEKAMER (1991, 1992, 1993), BARKMEIJER et al. (1993), BARKMEIJER (1992, 1993, 1995), OORTWIJN and BARKMEIJER (1995), and BARKMEIJER et al. (1996). With regard to a brief description of plans for implementing ensemble prediction systems at other operational centers (e. g., Japan, South Africa, India) reference is made here to TOTH and KALNAY (1997).

The recent developments in the area of ensemble prediction are aimed primarily at the prediction of the uncertainty of NWP model forecasts. The theoretical basis for all these developments may well be traced back to the LE describing uncertainty propagation governed by a dynamical system. The application of this concept to today's high-dimensional NWP models leads to several problems that are currently under investigation; approaches presently proposed and studied to deal with these problems and relevant results have been discussed here. With regard to assessing the uncertainty of NWP forecasts, success from applications of these approaches has been demonstrated, to some extent. Nevertheless, it remains to address further questions in this context in more detail, some of which seem to be: (i) the relationship between the specific objectives of ensemble prediction and the validation of the resulting products; (ii) the assessment of the degree to which the uncertainty of NWP forecasts is in fact predictable from a theoretical point of view; (iii) the tradeoff between the resolution of the NWP model used for the ensemble integrations and the number of ensemble integrations; (iv) the detrimental effect of model error on the quality of the ensemble; (v) the role of nonlinear effects in the time evolution of the pdf, as approximated through a small sample. It must be expected that future experimentation with ensemble prediction systems and extension of these systems will continue to introduce new and interesting perspectives into the process of weather forecasting and will in due time reveal large benefits for the users of forecasts that will justify the investments made during this period of developing ensemble prediction systems.

Acknowledgments

It is a pleasure to thank Jan BARKMEIJER (ECMWF), Roberto BUIZZA (ECMWF), Peter HOUTEKAMER (RPN), and Zoltan TOTH (NCEP), for providing detailed information about the implementation of the ensemble prediction systems at their centers. I am also grateful to Roberto BUIZZA, Professor Edward N. LORENZ (Massachusetts Institute of Technology, USA), Franco MOLTENI (Centro di Calcolo Interuniversitario dell'Italia Nord-Orientale, Italy), Thomas PETROLIAGIS (ECMWF), Bernard STRAUSS (Météo-France, France), Zoltan TOTH, Yuejian ZHU (NCEP), and to the Zentralanstalt für Meteorologie und Geodynamik (Wien, Austria) for providing figures and/or permitting the reproduction of published figures in this article. Detailed comments by Dave BAUMHEFNER (National Center for Atmospheric Research (NCAR), USA), Roberto BUIZZA, Ron ERRICO (NCAR), Professor Michael HANTEL (University of Vienna), and Zoltan TOTH on earlier versions of this manuscript are highly appreciated. I thank Professor Allan H. MURPHY (Prediction and Evaluation Systems, Oregon, USA) and Joe TRIBBIA (NCAR) for discussions related to the contents of this article. In particular, I would like to thank Professor Helmut PICHLER (University of Innsbruck) for the invitation to write this article. This work was carried out as part of the project P11234-GEO (Forecast Skill Prediction for Weather Prediction Models) funded by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung.

Appendix: Predicting forecast error covariances

Attention is given here to the optimal prediction of the forecast error covariance matrix $\mathbf{S}_t$. It is outlined that $\mathbf{S}_t$ can, under the tangent-linear hypothesis, be approximated in an optimal way through time-evolving appropriately constructed SVs (the presentation here follows closely the work of EHRENDORFER and TRIBBIA 1997). Optimality here refers to reconstruction of a maximum fraction of variance given a prespecified number of allowable model integrations. The appropriate construction of these SVs must take into account the initial error covariance structure (or, analysis error covariance matrix) denoted here as $\mathbf{V}$ (e. g., FISHER and COURTIER 1995, HOUTEKAMER 1995). It is first noted, that within a tangent-linear context, the time evolution of $\mathbf{V}$, followed by a projection (expressed through $\mathbf{M}$), is given by:

$$\hat{\mathbf{S}}_t = \mathbf{M}\mathbf{R}_t\mathbf{V}\mathbf{R}_t^{\mathrm{T}}\mathbf{M}^{\mathrm{T}} = \underbrace{\left(\mathbf{M}\mathbf{R}_t\mathbf{V}^{1/2}\right)}_{\equiv\mathbf{G}_t}\left(\mathbf{M}\mathbf{R}_t\mathbf{V}^{1/2}\right)^{\mathrm{T}}, \quad (\mathrm{A}.1)$$

where the superscript $^{\mathrm{T}}$ denotes the transpose, and $\hat{\mathbf{S}}_t$ denotes the tangent-linearly approximated form of $\mathbf{S}_t$. Consider next the question of finding the specific structure $\mathbf{y}$ that maximizes the following scalar (quadratic) cost function:

$$J(\mathbf{y}) = (\mathbf{M}\mathbf{R}_t\mathbf{y})^{\mathrm{T}}(\mathbf{M}\mathbf{R}_t\mathbf{y}), \quad (\mathrm{A}.2)$$

subject to the constraint:

$$\mathbf{y}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y} = 1, \quad (\mathrm{A}.3)$$

where $\mathbf{R}_t$ denotes the resolvent of the system that is tangent-linear to system (3.3), that is, system (5.2) (the resolvent $\mathbf{R}_t$ describes the solution of the tangent-linear system in the form of a mapping between initial perturbation and perturbation at time t). $\mathbf{M}$ denotes an arbitrary projection operator, that at the same time can be used to incorporate a norm, and $\mathbf{V}$ denotes the analysis error covariance matrix (i. e., the forecast error covariance matrix at initial time). The set of structures resulting from this maximization problem, is named the set of *singular vectors.* Since in the above maximization problem linearized error evolution is considered, these singular vectors are obtained by solving the following symmetric eigenproblem:

$$\left(\mathbf{M}\mathbf{R}_t\mathbf{V}^{1/2}\right)^{\mathrm{T}}\left(\mathbf{M}\mathbf{R}_t\mathbf{V}^{1/2}\right) = \hat{\mathbf{Y}}\Lambda\hat{\mathbf{Y}}^{\mathrm{T}}, \quad (\mathrm{A}.4)$$

where the columns of the matrix $\hat{\mathbf{Y}}$ are the orthonormal (i. e., $\hat{\mathbf{Y}}^{\mathrm{T}}\hat{\mathbf{Y}} = \mathbf{I}$) eigenvectors of $\mathbf{G}_t^{\mathrm{T}}\mathbf{G}_t$. The set of SVs is subsequently obtained through:

$$\mathbf{Y} = \mathbf{V}^{1/2}\hat{\mathbf{Y}}. \quad (\mathrm{A}.5)$$

The set of structures $\mathbf{Y}$, the first of which solves the maximization problem (A.2, 3), is denoted as the set of SVs, primarily because they are linearly related to the right singular vectors $\hat{\mathbf{Y}}$ of $\mathbf{G}_t$ that appear in the singular value decomposition (SVD) of $\mathbf{G}_t$ (see, e. g., STRANG 1986, GOLUB and VAN LOAN 1989). This SVD may be written as:

$$\mathbf{G}_t = \Pi\Lambda^{1/2}\hat{\mathbf{Y}}^{\mathrm{T}}. \quad (\mathrm{A}.6)$$

Here, Π and $\hat{\mathbf{Y}}$ are referred to as the left and right singular vectors of $\mathbf{G}_t$, respectively, and the diagonal elements of the diagonal matrix $\Lambda^{1/2}$ (positive square-root of Λ) are the corresponding singular values. Clearly, since the SVs $\hat{\mathbf{Y}}$ diagonalize the matrix $\mathbf{G}_t^{\mathrm{T}}\mathbf{G}_t$, they allow derivation of the eigenstructure of $\hat{\mathbf{S}}_t$ in the following form:

$$\hat{\mathbf{S}}_t = \mathbf{G}_t\mathbf{G}_t^{\mathrm{T}} = \Pi\Lambda\Pi^{\mathrm{T}}, \quad (\mathrm{A}.7)$$

where:

$$\Pi = \mathbf{G}_t\hat{\mathbf{Y}}\Lambda^{-1/2}, \quad (\mathrm{A}.8)$$

obtained through time-evolving the set of SVs $\hat{\mathbf{Y}}$, represents the orthonormal eigenstructure of $\hat{\mathbf{S}}_t$, and, as such, can be viewed as the empirical orthogonal functions (EOFs) needed to construct $\hat{\mathbf{S}}_t$. The result (A.7) indicates that the time-evolved SVs, when constructed using information about the initial covariance matrix $\mathbf{V}$, are the EOFs of the forecast error covariance matrix at time t. As such, when a given number of time-evolved SVs is used to reconstruct $\hat{\mathbf{S}}_t$ one can be assured that a maximum possible fraction of the total variance in $\hat{\mathbf{S}}_t$ is recovered. Or, stated differently, when k tangent-linear integrations are allowable, it is optimal to propagate the first k SVs with $\mathbf{G}_t$, given that optimality is measured by how much of the total variance can be reconstructed through these k integrations.

References

Alexander, M. A., C. Penland, 1996: Variability in a mixed layer ocean model driven by stochastic atmospheric forcing. — J. Climate **9**, 2424–2442.

Anderson, J. L., 1996a: Selection of initial conditions for ensemble forecasts in a simple perfect model framework. — J. Atmos. Sci. **53**, 22–36.

— 1996b: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. — J. Climate **9**, 1518–1530.

— 1996c: Impacts of dynamically constrained initial conditions on ensemble forecasts. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virginia, 56–57.

Anderson J. L., W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts in a perfect model setting. — J. Climate, **9**, 260–269.

Anthes, R. A., 1986: The general question of predictability. — In: Ray, P. S. (Ed.), Mesoscale meteorology and forecasting, Amer. Meteorol. Soc., 636–656.

Arnold, L. G., 1992: Stochastic differential equations: theory and applications. — Krieger, 228 pp.

Arnold, V. I., 1989: Mathematical methods of classical mechanics. — Springer, 508 pp.

Ashford, O. M., 1985: Prophet — or professor? The life and work of Lewis Fry Richardson. — Hilger, 304 pp.

Balescu, R., 1975: Equilibrium and nonequilibrium statistical mechanics. — Wiley, 742 pp.

Barker, T. W., 1991: The relationship between spread and forecast error in extended-range forecasts. — J. Climate **4**, 733–742.

Barkmeijer, J., 1992: Local error growth in a baroptropic model. — Tellus **44A**, 314–323.

— 1993: Lokal skill prediction for the ECMWF model using adjoint techniques. — Month. Weath. Rev. **121**, 1262–1268.

— 1995: Approximating dominant eigenvalues and eigenvectors of the local forecast error covariance matrix. — Tellus **47A**, 495–501.

— 1996: Constructing fast-growing perturbations for the nonlinear regime. — J. Atmos. Sci. **53**, 2838–2851.

Barkmeijer, J., P. Houtekamer, X. Wang, 1993: Validation of a skill prediction method. — Tellus **45A**, 424–434.

Barkmeijer, J., R. Mureau, J. Oortwijn, T. Opsteegh, 1996: Research on regional predictability at KNMI. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **I**, 247–274.

Baumhefner, D. P., 1985: The relationship between present large-scale forecast skill and new estimates of predictability error growth. In: Holloway G., B. J. West (Eds.), Predictability of fluid motions, American Institute of Physics, 169–180.

— 1996: Numerical extended-range prediction: forecast skill using a low resolution climate model. — Month. Weath. Rev. **124**, 1965–1980.

Berz, M., C. Bischof, G. Corliss, A. Griewank (Eds.), 1996: Computational differentiation — techniques, applications, and tools. — Society for Industrial and Applied Mathematics, 421 pp.

Betti, A., A. Navarra, 1995: Prediction of the evolution of the variance in a barotropic model. — J. Atmos. Sci. **52**, 367–379.

Bishop, C. H., Z. Toth, 1996: Using ensembles to identify observations likely to improve forecasts. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, 72–74.

Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. — Meteorol. Z. **21**, 1–7.

Boer, G. J., 1993: Systematic and random error in an extended-range forecasting experiment. — Month. Weath. Rev. **121**, 173–188.

— 1994: Predictability regimes in atmospheric flow. — Month. Weath. Rev. **122**, 2285–2295.

Borges, M. D., D. L. Hartmann, 1992: Barotropic instability and optimal perturbations of observed nonzonal flows. — J. Atmos. Sci. **49**, 335–354.

Borges, M. D., P. D. Sardeshmukh, 1995: Barotropic Rossby wave dynamics of zonally varying upper-level flows during northern winter. — J. Atmos. Sci. **52**, 3779–3796.

— — 1997: Application of perturbation theory to the stability analysis of realistic atmospheric flows. — Tellus, **49A**, 321–336.

Bouttier, F., 1993a: The dynamics of error covariances in a barotropic model. — Tellus **45A**, 408–423.

— 1993b: La prévisibilité. — La Météorologie **8**, 7–17.

— 1994: A dynamical estimation of forecast error covariances in an assimilation system. — Month. Weath. Rev. **122**, 2376–2390.

— 1996: The Kalman filter. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. I, 221–245.

Branković, C., T. N. Palmer, F. Molteni, S. Tibaldi, U. Cubasch, 1990: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. — Quart. J. Roy. Meteorol. Soc. **116**, 867–912.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. — Month. Weath. Rev. **78**, 1–3.

Brooks, H. E., M. S. Tracton, D. J. Stensrund, G. DiMego, Z. Toth, 1995: Short-range ensemble forecasting (SREF): report from a workshop. — Bull. Amer. Meteorol. Soc. **76**, 1617–1624.

Buizza, R., 1994a: Localization of optimal perturbations using a projection operator. — Quart. J. Roy. Meteorol. Soc. **120**, 1647–1682.

— 1994b: Sensitivity of optimal unstable structures. — Quart. J. Roy. Meteorol. Soc. **120**, 429–451.

— 1995a: Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. — Quart. J. Roy. Meteorol. Soc. **121**, 1705–1738.

— 1995b: The impact of orographic forcing on barotropic unstable singular vectors. — J. Atmos. Sci. **52**, 1457–1472.

— 1996: Ensemble prediction at the European Centre for Medium-Range Weather Forecasts. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 20–49.

— 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. — Month. Weath. Rev. **125**, 99–119.

Buizza, R., R. Gelaro, F. Molteni, T. N. Palmer, 1997a: The impact of increased resolution on predictability studies with singular vectors. — Quart. J. Roy. Meteorol. Soc., **123**, 1007–1033.

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

Buizza, R., T. N. Palmer, J. Barkmeijer, R. Gelaro, J. F. Mahfouf, 1996: Singular vectors, norms and large-scale condensation. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, 50–52.

Buizza, R., T. Petroliagis, T. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, N. Wedi, 1997b: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. — Quart. J. Roy. Meteorol. Soc., under revision.

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

Cane, M. A., 1992: Tropical Pacific ENSO models: ENSO as a mode of the coupled system. — In: Trenberth, K. (Ed.), Climate System Modeling, Cambridge University Press, 583–614.

Chang, J.-C. J., M. Mak, 1995: Nonmodal barotropic dynamics of intraseasonal disturbances. — J. Atmos. Sci. **52**, 896–914.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. — J. Meteorol. **4**, 135–162.

Charney, J. G., R. Fjörtoft, J. v. Neumann, 1950: Numerical integration of the barotropic vorticity equation. — Tellus **2**, 237–254.

Charney, J. G., R. G. Fleagle, H. Riehl, V. E. Lally, D. Q. Wark, 1966: The feasibility of a global observation and analysis experiment. — Bull. Amer. Meteorol. Soc. **47**, 200–220.

Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. — Month. Weath. Rev. **121**, 3123–3149.

Cohn, S. E., N. S. Sivakumaran, R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. — Month. Weath. Rev. **122**, 2838–2867.

Cohn, S. E., R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. — J. Meteorol. Soc. Japan **74**, 63–75.

Colucci, S. J., D. P. Baumhefner, 1992: Initial weather regimes as predictors of numerical 30-day mean forecast accuracy. — J. Atmos. Sci. **49**, 1652–1671.

Cooke, W. E., 1906: Forecasts and verifications in western Australia. — Month. Weath. Rev. **34**, 23–24.

Courtier, P., et al., 1993: Variational assimilation at ECMWF. — ECMWF Research Dept. Technical Memorandum No. 194, 84 pp. [available from: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK].

Courtier, P., J.-N. Thépaut, A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. — Quart. J. Roy. Meteorol. Soc. **120**, 1367–1387.

Dalcher, A., E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. — Tellus **39A**, 474–491.

Dalcher, A., E. Kalnay, R. N. Hoffman, 1988: Medium range lagged average forecasts. — Month. Weath. Rev. **116**, 402–416.
Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. — Month. Weath. Rev. **123**, 1128–1145.
DeGroot, S. R., P. Mazur, 1984: Non-equilibrium thermodynamics. — Dover, 510 pp.
DelSole, T., B. F. Farrell, 1995: A stochastically excited linear system as a model for quasigeostrophic turbulence: analytic results for one- and two-layer fluids. — J. Atmos. Sci. **52**, 2531–2547.
— — 1996: The quasi-linear equilibration of a thermally maintained, stochastically excited jet in a quasigeostrophic model. — J. Atmos. Sci. **53**, 1781–1797.
Drazin, P. G., W. H. Reid, 1981: Hydrodynamic stability. — Cambridge University Press, 525 pp.
Eady, E. T., 1949: Long waves and cyclone waves. — Tellus **1**, 33–52.
Egger, J., 1996: Volume conservation in phase space: a fresh look at numerical integration schemes. — Month. Weath. Rev. **124**, 1955–1964.
Ehrendorfer, M., 1989: Meteorologische Wahrscheinlichkeitsvorhersagen: Einige Aspekte. — Wetter u. Leben **41**, 123–139.
— 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. — Month. Weath. Rev. **122**, 703–713.
— 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. — Month. Weath. Rev. **122**, 714–728.
Ehrendorfer, M., R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. — J. Atmos. Sci. **52**, 3475–3500.
Ehrendorfer, M., J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. — J. Atmos. Sci. **54**, 286–313.
Ehrendorfer, M., J. J. Tribbia, R. M. Errico, 1996: Mesoscale predictability: an assessment through adjoint methods. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. I, 157–183.
Emanuel, K. A., E. N. Lorenz, R. E. Morss, 1996: Adaptive observations. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, 67–69.
Emmrich, P., 1996: Mittelfristige Wettervorhersage. — Promet Meteorol. Fortbildung **25**, 12–25.
Epstein, E. S., 1969: Stochastic dynamic prediction. — Tellus **21**, 739–759.
Errico, R. M., 1997: What is an adjoint model? — Bull. Amer. Meteorol. Soc., in press.
Errico, R. M., D. P. Baumhefner, 1987: Predictability experiments using a high-resolution limited-area model. — Month. Weath. Rev. **115**, 488–504.
Errico, R. M., M. Ehrendorfer, 1995: Moist singular vectors in a primitive-equation regional model. — Preprint Volume, Tenth Conference on Atmospheric and Oceanic Waves and Stability, Big Sky, Montana, 235–238.
Errico, R. M., T. Vukićević, 1992: Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. — Month. Weath. Rev. **120**, 1644–1660.
Errico, R. M., T. Vukićević, P. Courtier, J. Derber, J.-F. Louis, 1993a: Workshop on adjoint applications in dynamic meteorology. — Bull. Amer. Meteorol. Soc. **74**, 845–847.
Errico, R. M., T. Vukićević, K. Raeder, 1993b: Examination of the accuracy of a tangent linear model. — Tellus **45A**, 462–477.
European Centre for Medium-Range Weather Forecasts, 1996: Fifth Workshop on Meteorological Operational Systems. — Workshop Proceedings, 321 pp.
Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. — J. Geophys. Res. **99(C5)**, 10143–10162.
— 1997: Advanced data assimilation for dtrongly nonlinear dynamics. — Month. Weath. Rev. **125**, 1342–1354.
Evensen, G. P. J., van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. — Month. Weath. Rev. **124**, 85–96.
Farrell, B. F., 1988: Optimal excitation of neutral Rossby waves. — J. Atmos. Sci. **45**, 163–172.
— 1989: Optimal excitation of baroclinic waves. — J. Atmos. Sci. **46**, 1193–1206.
— 1990: Small error dynamics and the predictability of atmospheric flows. — J. Atmos. Sci. **47**, 2409–2416.
Farrell, B. F., P. J. Ioannou, 1993a: Stochastic forcing of the linearized Navier-Stokes equations. — Physics of Fluids A **5**, 2600–2609.
— — 1993b: Stochastic dynamics of baroclinic waves. — J. Atmos. Sci. **50**, 4044–4057.
— — 1996a: Generalized stability theory. Part I: autonomous operators. — J. Atmos. Sci. **53**, 2025–2040.
— — 1996b: Generalized stability theory. Part II: nonautonomous operators. — J. Atmos. Sci. **53**, 2041–2053.
Farrell, B. F., A. M. Moore, 1992: An adjoint method for obtaining the most rapidly growing perturbation to oceanic flows. — J. Physical Oceanogr. **22**, 338–349.
Fisher, M., P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. — ECMWF Research Dept. Technical Memorandum No. 220, 26 pp. [available from: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK].
Fishman, G. S., 1996: Monte Carlo — concepts, algorithms, applications. — Springer, 698 pp.
Fleming, R. J., 1971a: On stochastic dynamic prediction. I. The energetics of uncertainty and the question of closure. — Month. Weath. Rev. **99**, 851–872.
— 1971b: On stochastic dynamic prediction. II. Predictability and utility. — Month. Weath. Rev. **99**, 927–938.
Fortak, H., 1973: Prinzipielle Grenzen der deterministischen Vorhersagbarkeit atmosphärischer Prozesse. — Ann. Meteorol. (N. F.) **6**, 111–120.
Fraedrich, K., 1996: Das Lorenz-Modell: Ein Paradigma für Wetter und Vorhersagbarkeit. — Promet Meteorol. Fortbildung **25**, 62–79.
Fraedrich, K., C. Ziehmann-Schlumbohm, 1994: Predictability experiments with persistence forecasts in a red-noise atmosphere. — Quart. J. Roy. Meteorol. Soc. **120**, 387–428.
Fraedrich, K., C. Ziehmann, 1995: Praktische Vorhersagbarkeit: Persistenz in rotem Rauschen. — Meteorol. Z., N. F. **4**, 139–149.
Frederiksen, J. S., 1997: Adjoint sensitivity and finite-time normal mode disturbances during blocking. — J. Atmos. Sci. **54**, 1144–1165.
Gardiner, C. W., 1990: Handbook of stochastic methods. — Springer, 442 pp.
Gelaro, R., R. Buizza, T. N. Palmer, E. Klinker, 1997: Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. — J. Atmos. Sci., under revision.
Ghil, M., 1987: Dynamics, statistics and predictability of planetary flow regimes. — In: Nicolis, C., G. Nocolis (Eds.), Irreversible phenomena and dynamical systems analysis in geosciences, Reidel, 241–283.
Ghil, M., K. Ide, A. Bennett, P. Courtier, M. Kimoto, M. Nagata, M. Saiki, N. Sato (Eds.), 1997: Data Assimilation in Meteorology and Oceanography: Theory and Practice. Meteorol. Soc. Japan, Universal Academy Press, 386 pp.

Ghil, M., R. Todling, 1996: Tracking atmospheric instabilities with the Kalman filter. Part II: two-layer results. — Month. Weath. Rev. **124**, 2340–2352.

Gibson, J. K., A. Hernandez, P. Kallberg, A. Nomura, E. Serrano, S. Uppala, 1996: Current status of the ECMWF re-analysis (ERA) project. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 283–290.

Gleeson, T. A., 1966: A causal relation for probabilities in synoptic meteorology. — J. Appl. Meteorol. **5**, 365–368.

— 1970: Statistical dynamical predictions. — J. Appl. Meteorol. **9**, 333-344.

Golub, G. H., C. F. Van Loan, 1989: Matrix computations. — The Johns Hopkins University Press, 642 pp.

Götz, G., 1995: Predictability of nonlinear dynamical systems. — Idojaras **99**, 1–32.

— 1996: On the possibilities of forecasting forecast skill. — Idojaras **100**, 229–246.

Grimes, R. G., J. G. Lewis, H. D. Simon, 1994: A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. — SIAM J. Matrix Analysis and Applications **15**, 228–272.

Hammersley, J. M., D. C. Handscomb, 1964: Monte Carlo methods. — Methuen, 178 pp.

Harrison, M. S. J., 1994: Ensembles, higher-resolution models and future computing power — a personal view. — Weather **49**, 398–406.

— 1996: Ensemble forecasting at the UK Meteorological Office. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, J21–J23.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, R. Buizza, T. Petroliagis, 1996: Joint ensembles from the UKMO and ECMWF models. — Proceedings 1995 ECMWF Seminar on Predictability, Vol **II**, 61–120.

Harrison, M. S. J., D. S. Richardson, K. Robertson, A. Woodcock, 1995: Medium-range ensembles using both the ECMWF T63 and unified models — an initial report. — UKMO Techn. Rep. No. 153, 25 pp.

Hartmann, D. L., R. Buizza, T. N. Palmer, 1995: Singular vectors: the effect of spatial scale on linear growth of disturbances. — J. Atmos. Sci. **52**, 3885–3894.

Hartmann, D. L., T. N. Palmer, R. Buizza, 1996: Finite-time instabilities of lower-stratospheric flow. — J. Atmos. Sci. **53**, 2129–2143.

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

Hoffmann, R. N., E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. — Tellus **35A**, 100–118.

Houghton, J., 1991: The Bakerian Lecture 1991: The predictability of weather and climate. — Phil. Trans. R. Soc. Lond. A **337**, 521–572.

Houtekamer, P. L., 1991: Variation of the predictability in a low order spectral model of the atmospheric circulation. — Tellus **43A**, 177–190.

— 1992: The quality of skill forecasts for a low-order spectral model. — Month. Weath. Rev. 120, 2993–3002.

— 1993: Global and local skill forecasts. — Month. Weath. Rev. **121**, 1834–1846.

— 1995: The construction of optimal perturbations. — Month. Weath. Rev. **123**, 2888–2898.

Houtekamer, P. L., J. Derome, 1994: Prediction experiments with two-member ensembles. — Month. Weath. Rev. **122**, 2179–2191.

— — 1995: Methods for ensemble prediction. — Month. Weath. Rev. **123**, 2181–2196.

Houtekamer, P. L., L. Lefaivre, J. Derome, 1996a: The RPN ensemble prediction system. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **II**, 121–146.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, H. L. Mitchell, 1996b: A system simulation approach to ensemble prediction. — Month. Weath. Rev. **124**, 1225–1242.

Huber-Pock, F., 1996: Vorhersagbarkeit und Ensemble-Vorhersage. — ÖGM Bull. **96/2**, 2–32.

Ide, K., P. Courtier, M. Ghil, A. C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential and variational. — J. Meteorol. Soc. Japan, **75**, 181–189.

Iyengar, G., Z. Toth, E. Kalnay, J. S. Woollen, 1996: Are the bred vectors representative of analysis errors? — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, J64–J66.

Järvinen, H., J.-N. Thépaut, P. Courtier, 1996: Quasi-continuous variational data assimilation — Quart. J. Roy. Meteorol. Soc. **122**, 515–534.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. — Academic Press, 376 pp.

Kalnay, E., A. Dalcher, 1987: Forecasting forecast skill. — Month. Weath. Rev. **115**, 349–356.

Kalnay, E., Z. Toth, 1996: The breeding method. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **I**, 69–82.

Kanamitsu, M., S. Saha, 1996: Systematic tendency error in budget calculations. — Month. Weath. Rev. **124**, 1145–1160.

Katz, R. W., A. H. Murphy (Eds.), 1997: Economic value of weather and climate forecasts. — Cambridge University Press, 222 pp.

Kumar, A., M. Hoerling, M. Ji, A. Leetmaa, P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. — J. Climate **9**, 115–129.

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

Lanczos, C., 1950: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. — J. Res. Nat. Bureau Standards **45**, 255–282.

Langland, R. H., R. L. Elsberry, R. M. Errico, 1995: Evaluation of physical processes in an idealized extratropical cyclone using adjoint sensitivity. — Quart. J. Roy. Meteorol. Soc. **121**, 1349–1386.

— — — 1996: Adjoint sensitivity of an idealized extratropical cyclone with moist physical processes. — Quart. J. Roy. Meteorol. Soc. **122**, 1891–1920.

Lanzinger, A., B. Strauss, 1996: EPS evaluation at ECMWF. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 87–101.

Lasota, A., M. C. Mackey, 1994: Chaos, fractals, and noise. Stochastic aspects of dynamics. — Springer, 472 pp.

Legras, B., R. Vautard, 1996: A guide to Liapunov vectors. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **I**, 143–156.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. — J. Atmos. Sci. **28**, 145–161.

— 1974: Theoretical skill of Monte Carlo forecasts. — Month. Weath. Rev. **102**, 409–418.

— 1978: Objective methods for weather prediction. — Ann. Rev. Fluid Mechanics 10, 107–128.

— 1983: Predictability in theory and practice. — In: Hoskins, B. J., R. P. Pearce (Eds.), Large-scale dynamical processes in the atmosphere. Academic Press, 365–383.

Lewis, J. M., 1996: Philip Thompson: pages from a scientist's life. — Bull. Amer. Meteorol. Soc. **77**, 107–113.

Lin, C. C., L. A. Segel, 1988: Mathematics applied to deterministic problems in the natural sciences. — Soc. for Industrial and Applied Mathematics, 609 pp.

Lindley, D. V., 1987: The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. — Statistical Sci. **2**, 17–24.

Liouville, J., 1838: Note sur la théorie de la variation des constantes arbitraires. — J. Mathématiques pures et appliquées **3**, 342–349.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. — J. Atmos. Sci. **20**, 130–141.

— 1965: A study of the predictability of a 28-variable atmospheric model. — Tellus **17**, 321–333.

— 1969a: The predictability of a flow which possesses many scales of motion. – Tellus **21**, 289–307.

— 1969b: Atmospheric predictability as revealed by naturally occurring analogues. — J. Atmos. Sci. **26**, 636–646.

— 1982: Atmospheric predictability experiments with a large numerical model. – Tellus **34**, 505–513.

— 1984: Irregularity: a fundamental property of the atmosphere. — Tellus **21**, 739–759.

— 1990: Effects of analysis and model errors on routine weather forecasts. — Proceedings 1989 ECMWF Seminar on Ten years of medium-range weather forecasting, 115–128.

— 1993: The essence of chaos. — University of Washington Press, 227 pp.

— 1996: Predictability — a problem partly solved. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **I**, 1–18.

Lützen, J., 1990: Joseph Liouville 1809–1882: Master of pure and applied mathematics. — Springer, 884 pp.

Lynch, P., 1992: Richardson's barotropic forecast: a reappraisal. — Bull. Amer. Meteorol. Soc. **73**, 35–47.

Madden, R. A., P. R. Julian, 1994: Observations of the 40–50-day tropical oscillation — a review. Month. Weath. Rev. **122**, 814–837.

Ménard, R., R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. — Tellus **48A**, 221–237.

Miyakoda, K., G. D. Hembree, R. F. Strickler, I. Shulman, 1972: Cumulative results of extended forecast experiments. Part I: Model performance for winter cases. — Month. Weath. Rev. **100**, 836–855.

Miyakoda, K., J. Smagorinsky, R. F. Strickler, G. D. Hembree, 1969: Experimental extended predictions with a nine-level hemispheric model. — Month. Weath. Rev. **97**, 1–76.

Molteni, F., R. Buizza, T. N. Palmer, T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. — Quart. J. Roy. Meteorol. Soc. **122**, 73–119.

Molteni, F., T. N. Palmer, 1991: A real-time scheme for the prediction of forecast skill. — Month. Weath. Rev. **119**, 1088–1097.

— — 1993: Predictability and finite-time instability of the northern winter circulation. — Quart. J. Roy. Meteorol. Soc. **119**, 269–298.

Moore, A. M., B. F. Farrell, 1993: Rapid perturbation growth on spatially and temporally varying oceanic flows determined using an adjoint method: application to the Gulf stream. — J. Physical Oceanogr. **23**, 1682–1702.

Morokoff, W. J., R. E. Caflisch, 1995: Quasi-Monte Carlo integration. — J. Computational Physics **122**, 218–230.

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

Mukougawa, H., M. Kimoto, S. Yoden, 1991: A relationship between local error growth and quasistationary states: case study in the Lorenz system. — J. Atmos. Sci. **48**, 1231–1237.

Mureau, R., F. Molteni, T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. — Quart. J. Roy. Meteorol. Soc. **119**, 299–323.

Murphy, A. H., 1994: Assessing the economic value of weather forecasts: an overview of methods, results and issues. — Meteorol. Applications **1**, 69–73.

— 1997: Forecast verification. — In: Katz, R. W., A. H. Murphy (Eds.), Economic value of weather and climate forecasts, Cambridge University Press, 19–74.

Murphy, A. H., E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. — Month. Weath. Rev. **117**, 572–581.

Murphy, A. H., R. L. Winkler, 1984: Probability forecasting in meteorology. — J. Amer. Statistical Assoc. **79**, 489–500.

Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. — Quart. J. Roy. Meteorol. Soc. **114**, 463–494.

— 1990: Assessment of the practical utility of extended range ensemble forecasts. — Quart. J. Roy. Meteorol. Soc. **116**, 89–125.

Newman, M., P. D. Sardeshmukh, 1995: A caveat concerning singular value decomposition. — J. Climate **8**, 352–360.

Newman, M., P. D. Sardeshmukh, C. Penland, 1997: Stochastic forcing of the wintertime extratropical flow. — J. Atmos. Sci. **54**, 435–455.

Nicolis, C., 1987: Climatic predictability and dynamical systems. — In: Nicolis, C, G. Nicolis (Eds.), Irreversible phenomena and dynamical systems analysis in geosciences, Reidel, 321–354.

— 1992: Probabilistic aspects of error growth in atmospheric dynamics. — Quart. J. Roy. Meteorol. Soc. **118**, 553–568.

Nicolis, G., 1995: Introduction to nonlinear science. — Cambridge University Press, 254 pp.

Nicolis, C., S. Vannitsem, J.-F. Royer, 1995: Short-range predictability of the atmosphere: mechanisms for superexponential error growth. — Quart. J. Roy. Meteorol. Soc. **121**, 705–722.

Noble, B., J. W. Daniel, 1977: Applied linear algebra. — Prentice-Hall, 477 pp.

Oortwijn, J., J. Barkmeijer, 1995: Perturbations that optimally trigger weather regimes. — J. Atmos. Sci. **52**, 3932–3944.

Paegle, J., E. Robl, 1977: The time behavior of the probability density function of some simplified atmospheric flows. — J. Atmos. Sci. **34**, 979–990.

Palmer, T. N., 1988: Medium and extended range predictability and stability of the Pacific/North American mode. — Quart. J. Roy. Meteorol. Soc. **114**, 691–713.

— 1993: Extended-range atmospheric prediction and the Lorenz model. — Bull. Amer. Meteorol. Soc. **74**, 49–65.

— 1996: Predictability of the atmosphere and oceans: from days to decades. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **I**, 83–141.

Palmer, T. N., D. L. T. Anderson, 1994: The prospects for seasonal forecasting — a review paper. — Quart. J. Roy. Meteorol. Soc. **120**, 755–793.

Palmer, T. N., S. Tibaldi, 1988: On the prediction of forecast skill. — Month. Weath. Rev. **116**, 2453–2480.

Parker, T. S., L. O. Chua, 1989: Practical numerical algorithms for chaotic systems. – Springer, 348 pp.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. — Month. Weath. Rev. **117**, 2165–2185.

Penland, C., L. Matrosova, 1994: A balance condition for stochastic numerical models with application to the El Nino-Southern Oscillation. — J. Climate **7**, 1352–1372.

Penland, C., P. Sardeshmukh, 1995a: Error and sensitivity analysis of geophysical eigensystems. — J. Climate **8**, 1988–1998.

— — 1995b: The optimal growth of tropical sea surface temperature anomalies. — J. Climate **8**, 1999–2024.

Petroliagis, T., R. Buizza, A. Lanzinger, T. N. Palmer, 1997: Potential of the ECMWF ensemble prediction system in cases of extreme weather events. — Meteorol. Applications., **4**, 69–84.

Pires, C., R. Vautard, O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. — Tellus **48A**, 96–121.

Pitcher, E. J., 1977: Application of stochastic dynamic prediction to real data. — J. Atmos. Sci. **34**, 3–21.

Prager, T., T. Vukićević, J.-N. Thépaut, J.-F. Louis, P. Gauthier, R. M. Errico, J. Derber, P. Courtier, 1995: Second workshop on adjoint applications in dynamic meteorology. — Bull. Amer. Meteorol. Soc. **76**, 375–379.

Rabier, F., P. Courtier, O. Talagrand, 1992: An application of adjoint models to sensitivity analysis. — Beitr. zur Physik d. Atmosphäre **65**, 177–192.

Rabier, F., E. Klinker, P. Courtier, A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. — Quart. J. Roy. Meteorol. Soc. **122**, 121–150.

Richardson, D. S., M. S. J. Harrison, 1996: The use of ensemble forecasting techniques at the UK Meteorological Office, now and in the future. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 74–80.

Richardson, D. S., M. S. J. Harrison, K. B. Robertson, A. P. Woodcock, 1996: Joint medium-range ensembles using UKMO, ECMWF and NCEP ensemble systems. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, J26–J27.

Richardson, L. F., 1922: Weather prediction by numerical process. — Cambridge University Press.

Risken, H., 1989: The Fokker-Planck equation. — Springer, 472 pp.

Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, M. Hamrud, 1995: Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. — Month. Weath. Rev. **123**, 489–514.

Royer, J. F., 1993: Review of recent advances in dynamical extended range forecasting for the extratropics. — In: Shukla, J. (Ed.), Prediction of interannual climate variations, Springer, 49–69.

Saha, S., 1992: Response of the NMC MRF model to systematic-error correction within integration. — Month. Weath. Rev. **120**, 345–360.

Salmon, R., G. Holloway, M. C. Hendershott, 1976: The equilibrium statistical mechanics of simple quasi-geostrophic models. — J. Fluid Mechanics **75**, 691–703.

Sardeshmukh, P. D., M. Newman, M. D. Borges, 1997: Free barotropic Rossby wave dynamics of the wintertime low-frequency flow. — J. Atmos. Sci. **54**, 5–23.

Schubert, S., M. Suarez, J.-K. Schemm, E. Epstein, 1992: Dynamically stratified Monte Carlo forecasting. — Month. Weath. Rev. **120**, 1077–1088.

Sendov, B., A. Andreev, 1994: Approximation and interpolation theory. — In: Ciarlet, P. G., J. L. Lions (Eds.), Handbook of Numerical Analysis, Vol. **III**, Elsevier, 406–421.

Shukla, J., 1985: Predictability. — In: Saltzman, B., S. Manabe (Eds.), Issues in atmospheric and oceanic modeling, Part B, Advances in Geophysics, **28**, Academic Press, 87–122.

Simmons, A. J., R. Mureau, T. Petroliagis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. — Quart. J. Roy. Meteorol. Soc. **121**, 1739–1771.

Simon, H., 1984: The Lanczos algorithm with partial reorthogonalization. — Math. Comp. **42**, 115–136.

Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. — Bull. Amer. Meteorol. Soc. **50**, 286–311.

Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. – Bull. Amer. Meteorol. Soc. **77**, 953–961.

Soize, C., 1994: The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions. — World Scientific, 321 pp.

Stewart, D. A., 1994: Predictability of low-frequency planetary waves in a simple low-resolution model. — Month. Weath. Rev. **122**, 405–423.

Stewart, I., 1989: Does God play dice? The mathematics of chaos. — Basil Blackwell, 317 pp.

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

Stratonovich, R. L., 1992: Nonlinear nonequilibrium thermodynamics I. — Springer, 361 pp.

Strauss, B., A. Lanzinger, 1996a: Validation of the ECMWF ensemble prediction system. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **II**, 157–166.

— — 1996b: Verification of the ensemble prediction system (EPS). — ECMWF Newsletter **72**, 9–15.

Stroe, R., J. F. Royer, 1993: Comparison of different error growth formulas and predictability estimation in numerical extended-range forecasts. — Ann. Geophysicae **11**, 296–316.

Szunyogh, I., E. Kalnay, Z. Toth, 1997: A comparison of Lyapunov and optimal vectors in a low resolution GCM. — Tellus, **49A**, 200–227.

Talagrand, O., 1989: Four-dimensional variational assimilation. — Proceedings 1988 ECMWF Seminar on Data assimilation and the use of satellite data, Vol. **II**, 1–30.

Tanguay, M., P. Bartello, P. Gauthier, 1995: Four-dimensional data assimilation with a wide range of scales. — Tellus **47A**, 974–997.

Tennekes, H., A. P. M. Baede, J. D. Opsteegh, 1987: Forecasting forecast skill. — Proceedings ECMWF Workshop on Predictability in the Medium and Extended Range, 277–302.

Thacker, W. C., 1990: Large least-square problems and the need for automating the generation of adjoint codes. — Lectures in Applied Mathematics **26**, 645–677.

Thépaut, J.-N., P. Courtier, G. Belaud, G. Lemaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: a case study. — Quart. J. Roy. Meteorol. Soc. **122**, 535–561.

Thépaut, J.-N., R. N. Hoffman, P. Courtier, 1993: Interactions of dynamics and observations in four-dimensional variational assimilation. — Month. Weath. Rev. **121**, 3393–3414.

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

— 1961: Numerical weather analysis and prediction. — Macmillan, 170 pp.

— 1969: Reduction of analysis error through constraints of dynamical consistency. — J. Appl. Meteorol., **8**, 738–742.

— 1972: Some exact statistics of two-dimensional viscous flow with random forcing. — J. Fluid Mechanics **55**, 711–717.

— 1983: Equilibrium statistics of two-dimensional viscous flows with arbitrary random forcing. — Physics of Fluids **26**, 3461–3470.

— 1985a: A review of the predictability problem. — In: Holloway, G., B. J. West (Eds.), Predictability of fluid motions, American Institute of Physics, 1–10.

— 1985b: Moment closures for equilibrium statistics of randomly-forced triads. — J. Atmos. Sci. **42**, 631–638.

— 1985c: A statistical-hydrodynamical approach to problems of climate and its evolution. — Tellus **37A**, 1–13.

— 1985d: Prediction of probable errors of predictions. — Month. Weath. Rev. **113**, 248–259.

— 1986a: Probability distribution of modal amplitudes in interacting triads with arbitrary random forcing. — J. Atmos. Sci. **43**, 1517–1524.

— 1986b: A simple approximate method of stochastic-dynamic prediction for small initial errors and short range. — Month. Weath. Rev. **114**, 1709–1715.

— 1987a: Oral history. — Interview by J. Tribbia, A. Kasahara, 52 pp. [Available from the National Center for Atmospheric Research, P. O. Box 3000, Boulder, CO 80307, U.S.A.]

— 1987b: The maturing of the science. — Bull. Amer. Meteorol. Soc. **68**, 631–637.

— 1988: Stochastic-dynamic prediction of three-dimensional quasi-geostrophic flow. — J. Atmos. Sci. **45**, 2669–2679.

Tibaldi, S., T. N. Palmer, C. Branković, U. Cubasch, 1990: Extended-range predictions with ECMWF models: influence of horizontal resolution on systematic error and forecast skill. — Quart. J. Roy. Meteorol. Soc. **116**, 835–866.

Todling, R., S. E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. — Month. Weath. Rev. **122**, 2530–2557.

Toth, Z., E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. — Bull. Amer. Meteorol. Soc. **74**, 2317–2330.

— — 1996a: Ensemble forecasting at NCEP. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **II**, 39–60.

— — 1996b: Climate ensemble forecasts: how to create them? Idojaras **100**, 247–256.

— — 1997: Ensemble forecasting at NCEP and the breeding method. — Month. Weath. Rev., **125**, in press.

Toth, Z., E. Kalnay, S. Tracton, R. Wobus, J. Irwin, 1996a: A synoptic evaluation of the NCEP ensemble. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 50–73.

— — — — — 1997: A synoptic evaluation of the NCEP ensemble. — Weather and Forecasting, **12**, 140–153.

Toth, Z., I. Szunyogh, E. Kalnay, 1996b: Singular, Lyapunov and bred vectors in ensemble forecasting. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, 53–55.

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

Tracton, M. S., K. Mo, W. Chen, E. Kalnay, R. Kistler, G. White, 1989: Dynamic extended range forecasting (DERF) at the National Meteorological Center: practical aspects. — Month. Weath. Rev. **117**, 1604–1635.

Trevisan, A., 1995: Statistical properties of predictability from atmospheric analogs and the existence of multiple flow regimes. — J. Atmos. Sci. **52**, 3577–3592.

Trevisan, A., R. Legnani, 1995: Transient error growth and local predictability: a study in the Lorenz system. — Tellus **47A**, 103–117.

Tribbia, J. J., 1991: The problem of predicting forecast skill: current status and prospects for the future. — Extended Abstracts Submitted to the ICTP/WMO International Technical Conference on Long-Range Weather Forecasting Research, Trieste, Italy, World Meteorological Organization, 283–286.

— 1997: Weather prediction. — In: Katz, R. W., A. H. Murphy (Eds.), Economic value of weather and climate forecasts, Cambridge University Press, 1–18.

Tribbia, J. J., R. A. Anthes, 1987: Scientific basis of modern weather prediction. — Science **237**, 493–499.

Tribbia, J. J., D. P. Baumhefner, 1988a: The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. — Month. Weath. Rev. **116**, 2276–2288.

— — 1988b: Estimates of the predictability of low-frequency variability with a spectral general circulation model. — J. Atmos. Sci. **45**, 2306–2317.

— — 1993: On the problem of prediction beyond the deterministic range. — In: Shukla, J. (Ed.), Prediction of interannual climate variations, Springer, 251–265.

Tribbia, J. J., P. Gent, H.-L. Lee, 1996: ENSO predictability and prediction studies at NCAR. — Proceedings 1995 ECMWF Seminar on Predictability, Vol. **II**, 253–259.

Urban, B., 1993: A method to determine the theoretical maximum error growth in atmospheric models. — Tellus **45A**, 270–280.

Van den Dool, H. M., 1994: Searching for analogues, how long must we wait? — Tellus **46A**, 314–324.

Van Kampen, N. G., 1992: Stochastic processes in physics and chemistry. — North Holland, 465 pp.

Vannitsem, S., C. Nicolis, 1997: Lyapunov vectors and error growth patterns in a T21L3 quasi-geostrophic model. — J. Atmos. Sci., **54**, 347–361.

Von Myrbach, O., 1913: Verläßlichkeit von Wetterprognosen (Ein Vorschlag). — Meteorol. Z. **30**, 496–497.

Vukićević, T., 1993: Possibility of skill forecast based on the finite-time dominant linear solutions for a primitive equation regional forecast model. — J. Atmos. Sci. **50**, 1777–1791.

Watkins, D. S., 1993: Some perspectives on the eigenvalue problem. — SIAM Rev. **35**, 430–471.

Wilks, D. S., T. M. Hamill, 1995: Potential economic value of ensemble-based surface weather forecasts. — Month. Weath. Rev. **123**, 3565–3575.

Wilson, L. J., 1996: Verification of weather element forecasts from an ensemble prediction system. — ECMWF Workshop Proceedings, Fifth Workshop on Meteorological Operational Systems, 114–126.

Wobus, R. L., E. Kalnay, 1995: Three years of operational prediction of forecast skill at NMC. Month. Weath. Rev. **123**, 2132–2148.

Yoden, S., M. Nomura, 1993: Finite-time Lyapunov stability analysis and its applications to atmospheric predictability. — J. Atmos. Sci. **50**, 1531–1543.

Zhu, Y., G. Iyengar, Z. Toth, S. Tracton, T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. — Preprint Volume, Eleventh Conference on Numerical Weather Prediction, Norfolk, Virgina, J79–J82.

Ziehmann-Schlumbohm, C., K. Fraedrich, L. A. Smith, 1995: Ein internes Vorhersagbarkeitsexperiment im Lorenz-Modell. — Meteorol. Z., N. F. **4**, 16–21.

Doz. Dr. Martin Ehrendorfer
Institute for Meteorology and Geophysics
University of Vienna
Hohe Warte 38
A-1190 Vienna, Austria
e-mail:
martin.ehrendorfer@univie.ac.at

Received 7 January 1997, in revised form: 3 March 1997