ECMWF Feature article

from Newsletter Number 134 – Winter 2012/13

METEOROLOGY

20 years of ensemble prediction at ECMWF

www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/l6nteidw

This article appeared in the Meteorology section of ECMWF Newsletter No. 134 – Winter 2012/13, pp. 16–32.

20 years of ensemble prediction at ECMWF

Jan Barkmeijer, Roberto Buizza, Erland Källén, Franco Molteni, Robert Mureau, Tim Palmer, Stefano Tibaldi, Joe Tribbia

Twenty years ago, on 24 November 1992, the first ensemble forecasts were produced at ECMWF. At that time, ensemble forecasts were issued three times a week, on Friday, Saturday, Sunday, with 33 members at a T63L19 resolution for up to 10 days. Today, the ensemble runs twice a day with 51 members at resolution T639L62 to day 10, and T319L62 from day 10 to 15, and is extended to 32 days twice weekly. It is coupled to a wave model from day 0, and to a dynamical ocean model from day 10 (work is in progress to move this coupling also to day 0). Worldwide, it is recognised as providing the best global, medium-range and monthly probability forecasts (Figure 1). 'Spaghetti' maps are now firmly on the table of most forecasters! This article discusses some of the main points raised during an afternoon of presentations on 3 December 2012, organized to celebrate 20 years of operational ensemble production at ECMWF. The presentations were given by some of the people who contributed to its early design and implementation.

- ***Erland Källén*** (ECMWF) presents some historical background and an overall introduction to the articles
- ***Stefano Tibaldi*** (ARPA Emilia Romagna) brings us back to the 1980s, when discussions started at ECMWF about how users could be provided with an estimate of forecast uncertainty.
- ***Joe Tribbia*** (NCAR) highlights the role that the singular vector strategy played in the development of the ECMWF ensemble and their continuing importance.
- ***Tim Palmer*** (University of Oxford & ECMWF) described how 'flaps of a butterfly wing' led to his involvement in ensemble prediction.
- ***Robert Mureau*** (MeteoGroup) considered the difficulties faced by professionals when trying to present probability forecasts.
- ***Jan Barkmeijer*** (KNMI) focuses on the role of linear models in data assimilation and predictability.
- ***Franco Molteni*** (ECMWF) deals with the monthly and seasonal forecast time range, and discusses ensemble methods applied to this forecast range.
- ***Roberto Buizza*** (ECMWF) presents on-going research work aimed at further improving the ECMWF ensemble system and providing more valuable products to ECMWF users.

Some biographical information about the contributors is given in Box A.

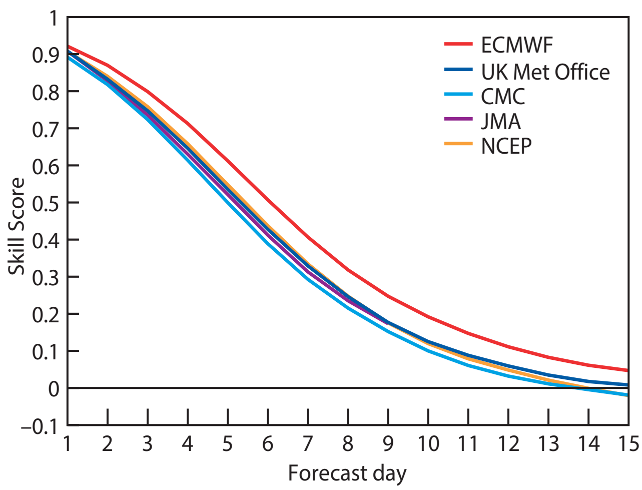

Figure 1 Three-month average continuous rank probability skill score for the probability forecast of the 500 hPa geopotential height over the northern hemisphere extra-tropics for August to October 2011 (dotted lines) and 2012 (solid lines) for the five leading ensembles available in the TIGGE archive: Canadian Meteorological Centre (CMC), Japan Meteorological Agency (JMA), National Centers for Environmental Prediction (NCEP) USA, UK Met Office and ECMWF. Each ensemble has been verified against its own analysis.

 doi:10.21957/l6nteidw

Brief biographical information of the authors

A

- **Erland Källén** worked at ECMWF between 1979 and1982 in the Research Department. After a period at Stockholm University, he returned to ECMWF in July 2009 as Director of Research, his current position.
  - Erland.Kallen@ecmwf.int

- **Stefano Tibaldi** worked in the Research Department of ECMWF between 1977 and 1987. He left as Head of Diagnostics and Predictability Research Section and was replaced in the position by Tim Palmer. Member of ECMWF SAC from 1990 to 1998, he is now the Director General of the Regional Protection Agency of Emilia-Romagna, Italy.
  - stibaldi@arpa.emr.it
  - http://www.arpa.emr.it/

- **Joe Tribbia** visited ECMWF for 6 months in 1991–92, and since then has been working closely with many ECMWF scientists. He is now the Head of the Climate Dynamics and Predictability (CDP) section in the Division of Climate and Global Dynamics (CGD) at the National Center for Atmospheric Research.
  - tribbia@ucar.edu
  - http://ncar.ucar.edu/

- **Tim Palmer** worked at ECMWF between 1986 and 2011 (but continues as a part-time consultant). He is now a Royal Society Research Professor in Climate Physics at the Department of Physics of the University of Oxford.
  - t.n.palmer@atm.ox.ac.uk
  - http://www2.physics.ox.ac.uk/

- **Robert Mureau** worked at ECMWF between 1987 and 1992, in the Research Department. He is now head of research and development at MeteoGroup.
  - r.mureau@weer.nl
  - http://www.meteogroup.com

- **Jan Barkmeijer** worked at ECMWF between 1995 and 2002 in the Research Department. He is now group leader of the Model Section in the Weather Research Department at the Royal Netherlands Meteorological Institute
  - jan.barkmeijer@knmi.nl
  - http://www.knmi.nl

  From January 2013 he was appointed Chair of the ECMWF Scientific Advisory Committee.

- **Franco Molteni** worked at ECMWF in different spells from 1984 to 1996. After a period in Italy, he returned to ECMWF in 2005. He is now the Head of the Probabilistic Forecasting Section.
  - Franco.Molteni@ecmwf.int

- **Roberto Buizza** has been working at ECMWF since 1991. He is now the Head of the Predictability Division of the Research Department.
  - Roberto.Buizza@ecmwf.int

Erland Källén

20 years of ensemble prediction: an introduction

No weather forecast is complete without an estimate of its uncertainty – this has been a long standing truism in meteorology ever since the start of numerical weather prediction in the 1950s. Already *Thompson* (1957) argued that estimating prediction errors is an essential part of dynamic meteorology. He estimated the growth of prediction errors due to uncertainties in the initial state and came to the conclusion that a weather forecast can be no better than a random guess of the weather beyond about a week. *Lorenz* (1969) made a more elaborate analysis of the mechanisms that limit atmospheric predictability and arrived at a predictability estimate of about two weeks.

When ECMWF started with operational weather prediction in 1979 the forecasts did not include uncertainty estimates. It was however recognised that this would be very desirable and a workshop was organised in 1979 to discuss "*Stochastic Dynamic Forecasting*" (*ECMWF*, 1979). An early attempt to compute forecast uncertainties was made by *Hollingsworth* (1979), but the perturbation technique was not sufficiently developed to give useful uncertainty information. Following methods suggested by *Lorenz* (1965) a much improved technique to specify initial condition uncertainty was developed at ECMWF and Tim Palmer headed the team that would implement the first ECMWF ensemble prediction system, today called ENS. We are celebrating the twentieth anniversary of the first operationally-produced ensemble forecasts and the contributors to the commemoration event held at ECMWF gave both a historical account of the developments at ECMWF as well as a look into the future.

All contributors have played an important role in the development of the ECMWF ENS and we are very grateful for their efforts in helping to develop and maintain a world-leading ensemble prediction system as documented in *Hagedorn et al.* (2012). Stefano Tibaldi, Tim Palmer, Joe Tribbia, Robert Mureau and Franco Molteni were all members of the initial ensemble team at ECMWF, with Roberto Buizza and Jan Barkmeijer joining later. All were instrumental in developing the first ensemble system based on the idea that singular vectors (*Farrel*, 1982) can be used to efficiently represent the initial condition uncertainty. Later developments include model error representation through the use of stochastic physics tendencies and various resolution and model upgrades.

In recent years the initial state uncertainty representation has been enhanced through the addition of perturbations derived from Ensembles of Data Assimilations. Personal memories of the difficulties encountered and the successes accomplished can be found in the contributions included in this edition of the *ECMWF Newsletter*. Also included are discussions of how probabilistic forecasts have been received by the users, it is evident that still more work needs to be done in order to widen the acceptance of ensemble forecast information as an integral part of a weather forecast.

Today, 20 years after its operational implementation, the ECMWF ensemble forecasts are much more reliable and skilful than they were two decades ago. The skill has been improving at a rate which is more rapid than the improvement in high resolution, or deterministic, forecast skill. It is also clear that the ensemble skill improvement is dependent on a continued improvement of the accuracy of the high resolution model as well as the accuracy of the initial state. We are confident that continued work and collaboration within the scientific community will lead to further skill advances and an enhanced use of probabilistic forecasts. We are all grateful to the pioneers at ECMWF, please read and enjoy their stories.

doi:10.21957/l6nteidw

Stefano Tibaldi
Why ensembles?

Why, twenty years ago, in 1992, did ECMWF decide to extend the operational production of the so-called 'deterministic' forecasts to 'ensemble' predictions? Or, more appropriately, why did ECMWF, almost fifteen years earlier in the very early 1980s, start a research project to explore the feasibility of producing probability forecasts and/or forecasts explicitly addressing the problem of forecast uncertainty which, in turn, led fifteen years later to a viable operation system, today a pillar of ECMWF's operational forecasting system?

The early days

The seed for probability forecasts was planted very early at ECMWF by Tony Hollingsworth in a (failed, alas!) experimental attempt, back in 1979 (*Hollingsworth*, 1980). He followed *Leith* (1974) in applying straight Monte Carlo techniques to meteorological modelling by perturbing model variables with random perturbations with amplitudes of the order of the analysis error as estimated by the optimum interpolation (OI) technique. The gallant attempt turned out to be doomed because such random perturbations were not projecting enough on meteorological modes and were therefore rapidly wiped out by model dissipative mechanisms (mostly horizontal diffusion).

The idea remained somewhat dormant until Henk Tennekes raised the matter during discussions at ECMWF's Scientific Advisory Committee (SAC) at around 1986, discussions which are usually referred to by quoting his statement that "no forecast is complete without a forecast of forecast skill" (*Tennekes et al.*, 1987). My personal recollection is that his opinion, forcefully and convincingly stated, sufficiently influenced the rest of the SAC and the then Director (Lennart Bengtsson) and Head of Research (David Burridge) to set the ball rolling, at least at ECMWF (no need to convince Tony Hollingsworth, he was already convinced).

If you ask yourself where the scientific interest in probability forecasting was coming from, you have to recollect that the long-standing attempt of producing probability forecasts by solving the evolution equations for the pdfs (probability distribution functions) of atmospheric variables (the so-called classical stochastic-dynamic forecasting techniques, e.g. *Epstein*, 1969) had failed. In practice this was due, among other things, to the rapidly increasing complexity of the problem when moving away from the prediction of the mean mean toward that of the higher order moments.

But there were other reasons why the times were ripe. For example, the attempt of using the 12 or even 6 hour lagged-average forecasting technique (*Hoffmann & Kalnay*, 1983) to estimate forecast spread, hoping that it would provide an estimate of forecast skill, had also given somewhat disappointing results, see *Palmer & Tibaldi* (1988).

Additionally, no time can be truly ripe if the technical instruments are not ready (remember how long we had to wait for Richardson's NWP ideas to come to fruition). At that time (the second half of the 1980s) it had become technically possible to perform large-scale numerical experiments: the enormous growth of available computing power had made it possible to construct and operate at full steam that 'numerical laboratory' that Axel Wiin-Nielsen, Syukuro Manabe and others had started implementing 10–15 years earlier. In fact it was exactly in those years that a Numerical Experimentation Section was created in the ECMWF Research Department.

Why probability forecasting?

But again, why probability forecasting? There is nothing wrong with progressively improving single, 'deterministic' forecasts by increasing model accuracy and decreasing initial condition errors, is there? But the meteorological atmosphere is a chaotic system on time scales of a few days, maybe weeks, depending on the spatial scales of interest (and the climatic system is also chaotic, but on much longer timescales). Also the behaviour of our numerical simulations of the atmosphere would continue to be affected by the problems typical of model simulations of chaotic dynamical systems even if (a) we could have perfect initial conditions and (b) we could write perfectly accurate evolution equations and (c) solve them with perfect numerical schemes, only just because of the limited number of significant digits used by any computer (*Lorenz*, 1963).

Looking at the problem from a slightly more fundamental point of view, a forecast explicitly cast in probability terms is better not only because it provides the user with an estimate of the error 'of the day', but because it is more 'truthful'. So a probability forecast conveys a message which explicitly reminds the user of the fact that associated to the forecast there is always a forecast uncertainty which should be considered, computed and taken into account when making any practical use of the forecast (see Figure 2). In fact even 'deterministic' forecasts are in reality probability forecasts in disguise, since an error bar (even if only an average error bar) can and should always be associated with it. That error bar implies a probability distribution of predicted future states around a central value.

Increasing the use of probability forecasts?

But, having said all that, do problems with spreading the use of ensemble forecasts or probability forecasts remain even today, after twenty years of operational production and dissemination? This would indeed appear to be the case, at least in some situations, as often the following problems are still outstanding.

- There are very poor statistics concerning the verification of rare events (e.g. extreme events) which are often the main target of ensemble forecasts.
- Some forecast users do not always interpret the concept of probability associated with ensemble forecasts in the correct way. It is sometimes difficult to explain that it is possible for two different forecasting systems produce different probabilities of occurrence of the same meteorological phenomenon because they have errors of different nature, size and structure and that this is perfectly legitimate and does not necessarily imply that one is wrong and the other one is right (they may in fact be both right, or wrong, for that matter!).
- The entire system of civil protection alerts, for example, is currently based (at least in some countries) on a conceptually deterministic use of meteorological and hydrological forecasts: will the river overflow the bank? The answers "yes" or "no" are allowed but not "maybe"!

Recall for a moment the experience of Charlie, Tim Palmer's golfing friend, the builder who laid concrete immediately before a frost on the basis of a (wrong) 'deterministic' forecast (*Palmer*, 2006). Had he used a probability forecast, he could have been much better off, but only after applying a quantitative cost/loss analysis to his concrete laying job. Was he ready at the time? Did he know all the necessary action/no action cost figures? And, even more importantly, would he be culturally ready even today?

Helping decision makers

A civil protection plan to evacuate thousands of citizens in view of a probable flood requires a bit more effort than costing the laying of three thousand square yards of concrete. So to modify a civil protection plan to make it consistent with weather and hydrological forecasts that are cast in probability terms requires a mix of cultural, scientific, technical and communication skills. Depending on the cost/loss ratio of every action to be taken as a consequence of the weather alert, different decisions might have to be taken, and the costing of some losses/damages (human life, population health, psychological consequences of loss of homes and property, etc.) might pose severe difficulties. Special development efforts might be needed before decision making and intervention procedures can be thought and formulated in probability terms.

The technical, economic and productive systems of some Central and Northern European countries might be already using probability ensemble predictions to their full value. Can we say the same of other public bodies and of operational meteorology, hydrology and civil protection of Mediterranean Europe, where objective and quantitative fact-based decision making has not yet completely penetrated minds and society?

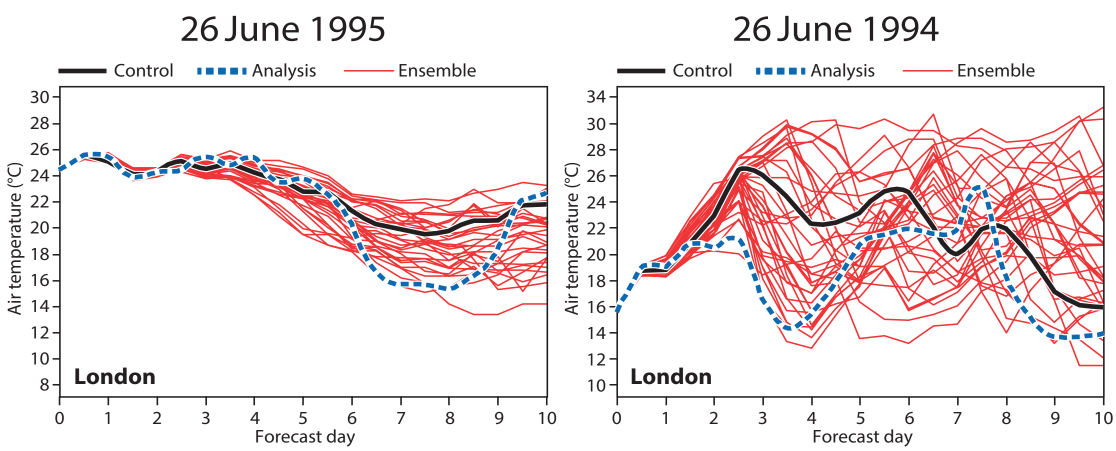

Figure 2 Two ensemble forecasts of air temperature on the same day of two different years at London illustrating the flow dependence of forecast errors (the errors 'of the day'). If the forecasts are coherent (small spread) the atmosphere is in a more predictable state than if the forecasts diverge (large spread).

doi:10.21957/l6nteidw

Joe Tribbia

Singular vectors: are they still valuable?

Twenty-one years ago I had the good fortune to be an invited visitor to ECMWF and to work in the Predictability Section; this was in pre-launch stage of the ensemble forecasting system. It had been agreed in discussions with Tim Palmer that my tasks would be associated with the use of Singular Vectors (SVs) as a way of sampling the initial state uncertainty and thereby initialize the ensemble for the prediction of forecast reliability. Tim was not only the Section Head but also the intellectual leader who pushed hard for the development of ensemble forecasts based on SVs.

Leading the way with SVs

As is often the case, the Centre was leading the operational community into a new and challenging area of predictive science and the work I was doing during my visit would play a small part in the development of useful ensemble forecasts. Prior to my arrival almost all of the groundwork had been laid by two members of the Predictability Section, Franco Molteni and Robert Mureau. They had done some preliminary experiments using a quasi-geostrophic (QG) model to generate SVs that were interpolated onto the operational model's grid as a precursor to the self-consistent SVs generated using the numerical approximations of the forecast model (see Figure 3). These QG SVs appeared to be producing perturbed ensemble members that were dynamically active and held the promise of finally achieving rational spread–skill relationships. This was something many groups were seeking within the context of tackling dynamical extended-range forecasting, which was defined as numerical weather prediction beyond the deterministic range (in those days, week two to one month lead).

In addition to the foundation already in place due to the testing of QG ensemble perturbations, the Centre's Head of Research, Tony Hollingsworth, and Phillipe Courtier had seen to it that the ECMWF forecast model would have tangent forward versions of each subroutine along with their complementary adjoints, ensuring that my main tasks would be primarily to connect the components and insert an eigen-solver at the proper point to enable SV perturbations to be computed. At the end of my visit, I handed over this capability to a bright, young researcher the Centre had hired into the Predictability Section, Roberto Buizza, and the success of SV-based ensemble forecasts was assured.

The future for SVs

Twenty years later, the ECMWF ensemble forecasts are the best in the world due to the combination of having the most accurate model and data assimilation components, the use of stochastic physics, calibration of the forecast skill and the SV method of generating initial perturbations for the ensemble (c.f. *Hagedorn et al.*, 2012). However, the future use of the SV technique is in question as ECMWF moves forward. The most ominous threat, ironically, comes from the ensemble forecast itself which has been shown to be an effective and economical tool for the prediction of forecast reliability.

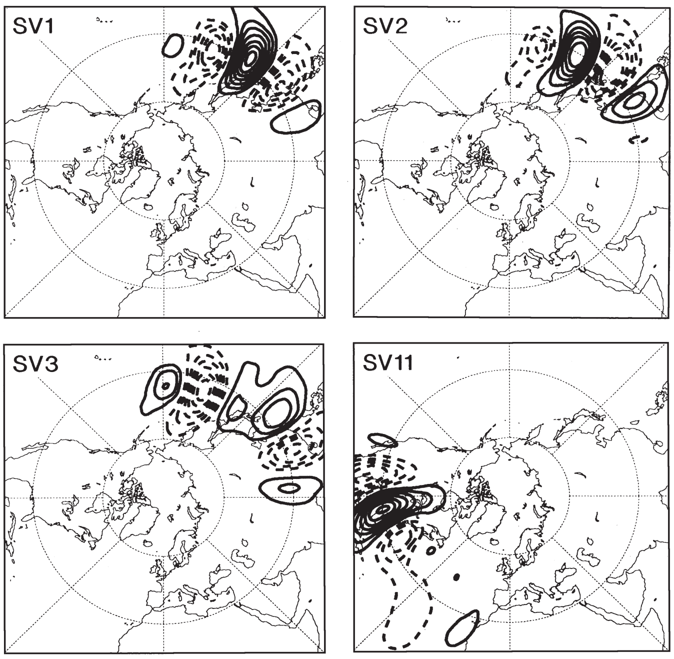

Figure 3 Until 1992, ECMWF did not have a tangent forward and adjoint version of its primitive equation model. Thus early experimentation on the use of dynamically conditioned, optimal perturbations was done by extrapolating to the T63L19 resolution of the ECMWF model singular vectors computed with a T21L3 quasi geostrophic model. This figure shows one of the first examples of the T21L3 QG-SVs growing over a 12-hour optimisation time period from 12 UTC on 2 December 1988. The quasi-geostrophic SV perturbations numbers 1, 2, 3 and 11 are shown in terms of 500 hPa geopotential height (from *Mureau et al.*, 1991).

With the confidence gained from twenty years of ensemble prediction, ECMWF is experimenting with ensemble methods of data assimilation, like the Ensemble Kalman Filter (EnKF). If the current 4D-Var assimilation method is replaced with an EnKF-based assimilation scheme, the next logical step would be to use the ensemble members from the EnKF as the initial states for the ensemble members. (Preliminary experiments shown by Roberto Buizza indicate that not much, if any, skill would be lost in the ensemble forecasts in using such a strategy.) This would permit a self-consistent, statistically-optimal ensemble method of predicting both expected value of the forecast and its expected error covariance.

Even if SV initialization is eventually superseded by an alternative method of ensemble member generation at ECMWF, the Centre would be wise to continue to maintain the capability to generate SVs in their forecast system. The reason is simple: SVs are not only practically useful in generating ensemble members, they are also dynamically useful as an efficient basis for performing diagnostic analysis. This is clear from the dual nature of SVs which can be derived either as the most rapidly growing perturbations or the EOFs (Empirical Orthogonal Functions) at a future time of a distribution that is initially Gaussian. In the latter interpretation of SVs, these vectors are the most efficient that can be used to characterize the distribution. Also, while for the typical weather event ensemble filtering methods of generating initial states for ensemble members may be indistinguishable in a probability forecast from the SV method of initialization, in the atypical extreme event SV initialization may be a necessary ingredient for an early warning of a significant storm. This would be consistent with the fastest growing disturbance derivation of SVs and give a further rationale for their use.

Whatever the future brings the history of SVs and their use at ECMWF has demonstrated that the efforts of the Predictability Section twenty years ago have benefited both the Atmospheric Sciences and Society.

Tim Palmer

Is the butterfly effect real? A practical example

Although I have spent much of my life developing and promoting the benefits of ensemble forecasting, this was not something I had planned to do, nor had I predicted that this is something I would do when I started my meteorological research. My entry into this field was an example of the ‘butterfly effect’ at work – see Figure 4.

At the Met Office

I joined the Met Office having completed a doctorate in theoretical physics in 1977. My initial research at the Met Office was in stratospheric dynamics. In the early 1980s I was lucky enough to spend a year at the University of Washington working with Jim Holton, and returned to the Office as an expert in stratospheric dynamics. This led me to being promoted to the rank of Principal Scientific Officer. The only problem was that I now had to head a group, and there was already a well-established stratospheric group leader who was a world leader in the field. And he wasn’t going anywhere! The only group-leader vacancies were at the Office’s Training College (then next door to ECMWF) and in the long-range forecasting branch.

The long-range forecasting branch, sometimes called the ‘Synoptic Climatology’ branch, was known for making long-range forecasts (typically a month ahead but sometimes longer) using statistical empirical models. Monthly forecasts using such techniques were made for a number of fee-paying customers; typically the utilities companies in the UK. My job, should I choose to accept it, was to introduce dynamical methods into long-range forecasting (Andrew Gilchrist was the driving force behind this project). For some time, I was not really sure what the better choice would be – Training College or Synoptic Climatology, and in any case I was rather baffled by the system that prevented me from being able to continue with the work which had established my name in the middle atmosphere field. In the end a few butterflies flapped their wings (or maybe did not) and I ended up choosing Synoptic Climatology.

The Met Office was already using a global climate model to study climate change. I had to adapt and test the model for use in monthly and seasonal forecasting. It was fairly obvious right from the start that it would be necessary to study these forecasts using ensembles of integrations, rather than single ‘deterministic’ forecasts, and there was already some literature on this, notably, by Kiku Miyakoda from GFDL. Doug Mansfield had already done some of the early work in this area in the Synoptic Climatology branch using a hemispheric model. Doug and I were then joined by James Murphy, who subsequently made a name for himself introducing ensemble techniques into the Hadley Centre’s climate change forecast system.

After a year or two of research in this area, James and I finally introduced these ensemble techniques into the operational monthly forecast system, the one that produced forecasts issued to the utilities companies and others, producing dynamically based probabilities. Here we blended the dynamical and empirical probability forecasts. The paper *Murphy & Palmer* (1986) documents what I believe was the first ever

doi:10.21957/l6nteidw

operational ensemble weather forecast. Unfortunately, these probabilities were not all that reliable, not least because the model had severe biases in its mid-latitude flow. This led me to think about parametrization of orographic gravity wave effects, but that is another story.

At ECMWF

It turned out that ECMWF was acquiring quite a reputation for innovative and exciting research, and, after a couple of years or so in the Synoptic Climatology branch, I applied for a job at ECMWF, working with Stefano Tibaldi. When I joined in 1986 there was interest at ECMWF in developing techniques to 'forecast the forecast skill' as it was then called (a phrase, coined, I believe, by Henk Tennekes). However, it seemed to me that what ECMWF really needed was an operational ensemble forecast system for producing completely probabilistic forecasts for the medium range, a bit like the one that had developed for the monthly forecast system at the Met Office. In fact Tony Hollingsworth had already started to think about Monte Carlo forecasting at ECMWF. However, he had rapidly found a problem that beset all attempts to use 'random' perturbations to create ensembles of initial conditions. In Tony's experiments the random perturbations were actually decaying as they were integrated away from the initial conditions, the complete opposite of what a chaotic system should do!

Before working actively on developing a medium-range ensemble forecasting system, Stefano Tibaldi and I looked empirically at what types of forecast flows were more predictable and what ones less so. Using a barotropic model developed by Adrian Simmons, I was able to show that the results of our empirical analysis could be explained in terms of the growth of small perturbations on barotropic basic states associated with the predictable or unpredictable composite forecast flows. To my initial surprise, however, I found that normal mode theory was utterly incapable of explaining how these small perturbations grew in the barotropic model; the basic states where initial perturbation growth was largest were more stable in the normal mode sense than the basic states where perturbations grew rapidly!

At this time, I recalled a talk by Brian Farrell from Harvard which I had heard whilst still at the Met Office. Brian said forcefully that normal mode growth was irrelevant for explaining baroclinic instability (Farrell, 1982). At the time I did not really understand why he was saying this, but I finally understood it when trying to understand results from the barotropic model. Essentially what Brian was saying is that the growth of small perturbations is governed by processes which relate to the fact that the linear evolution operators are not self adjoint. Together with Brian Hoskins and a PhD student, Zuojun Zhang, we were finally able to understand the results from the barotropic model.

This research had a strong impact on me as I began to think about ways to construct a medium-range ensemble forecast system which could overcome the problems found by Tony Hollingsworth. With Franco Molteni, it led us to formulate a strategy for perturbing the initial conditions for an ensemble forecast, based on singular vector perturbations. We still use this strategy today, although the Ensemble of Data Assimilations method (EDA) plays an increasingly important role for creating ensemble perturbations. However, since the IFS dynamics is rather dissipative at high wavenumbers (*Augier & Lindborg*, 2012), it is not completely clear to me how well EDA can adequately account for observation and initial model uncertainty at high wavenumbers (without artificially inflating the perturbations) and one should probably not drop the singular-vector perturbations too readily.

The flap of those butterflies wings in the early 1980s has had a profound effect on my career; one, I think, that was for the better!

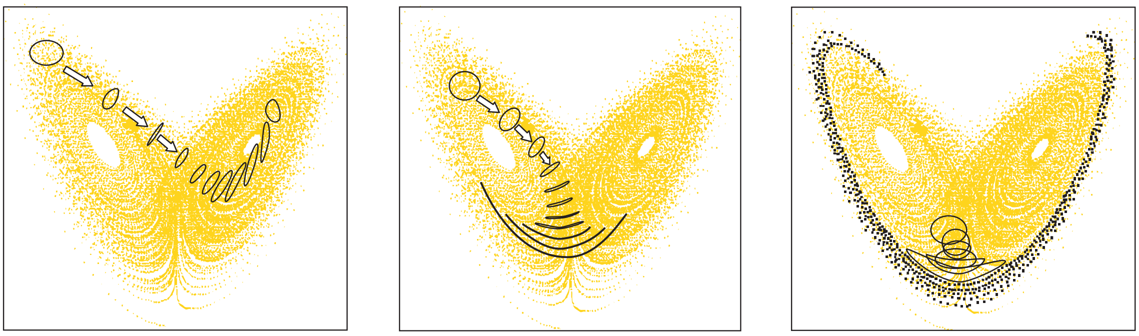

Figure 4 The rationale for ensemble forecasting can be demonstrated using the iconic Lorenz '63 model. In a nonlinear system predictability is a function of initial state. Ensemble predictions make it possible to forecast such flow-dependent predictability.

Robert Mureau
Do people really want probability forecasts?

I worked at ECMWF between 1987 and 1992, and am proud to have been, from the very beginning, involved in ensemble forecasting. I have been and still am a firm believer in the use of ensembles for making probability forecasts. I fully agree with those who say that the probability distribution for a forecast parameter gives the most comprehensive information about the forecast. And, yet, when you ask me to answer the question in the title, I will have to take a deep breath and have to answer the question with a firm “no”: people do not want probability forecasts. They should want it, but they do not. That is a frustrating answer, but it is reality. I think it is time we recognize and accept the hesitations of the user. Perhaps we should think of different strategies to persuade people to use probability forecasts.

Barbecue summers and hamburgers

Probability forecasting in meteorology and the attempts to promote probability forecasting go back to, at least, the 1960s when Allan Murphy and Ed Epstein started their pioneering work. That was tedious stuff, very theoretical and never very appealing to the simple user, despite the excellent examples that were provided (*Katz & Murphy*, 1997). When more computer power became available and we began to understand how to tackle the issue of generating perturbations that would survive the initialization (Hollingsworth, 1980) probability forecasting became much more transparent. We could now follow the effect of small errors and could literally see the forecasts diverge, just as Edward Lorenz had experienced in his very first modelling experiments.

At ECMWF, Horst Böttger, Bernard Strauss and Anders Persson realized from the very beginning that training forecasters in the use of the ensembles would be important and that it had to be part of the training courses run by the Met Ops Department at ECMWF. Those courses certainly did their job. An experienced (Dutch) forecaster once told me, slightly nervously, after attending the Training Course: “I go home now, but I don’t dare to make a forecast anymore”. He did go home, he did make many more forecasts, but as ‘deterministic’ as he ever had done before. That surprised me, but he explained that the daily operational routine expected you to fill in tables and present the forecast in numbers. And also, as he was trained in the traditional way, he saw it as his duty to show people the way in the dark world of uncertainty, and give into the wish of the users to make the decision for them.

There is a constant battle between on the one hand the professional (scientist, forecaster) and on the other hand the media and the decision makers. In the old days scientists were revered and well-respected members of society – people listened. Nowadays, they appear in popular talk shows and are constantly challenged to present their message in a way that everybody understands. That is good (out of the ivory tower), but accidents happen.

As mentioned above, people want straight answers. The well-prepared professional who goes to the studio with cautious arguments lined up and probabilities assessed, has to ‘compete’ with the talk show host and may get, after the extensive and thorough explanation has been given, still the question: “but is it going to rain?”. Sometimes the professional caves in, sometimes the media simply misquote. In the late eighties you were given the impression that if, in Britain, you had eaten one hamburger in your life, you were doomed. Mexican flu, seasonal forecasting (barbecue summers), climate change… there are many examples. The public remembers all the above examples as false alarms. Of course these examples all support the case for probabilistic statements. But very often you do not get the chance: your interview would probably not be broadcast unless you make strong statements. A probability statement is seen as a weak statement, a ‘cop out’. This is the world the professional lives in.

Still the same question after twenty years

The fact that the organizers of the event to mark twenty years of ensemble prediction at ECMWF asked me, explicitly, to give a talk with the above title illustrates the point. (I was particularly puzzled by the choice for the word “really” in the question). But it was justified to ask the question: most of us are still giving very similar presentations to twenty years ago, and, worse, are getting the same negative, sometimes outright cynical, responses. We must have done something wrong. We should try to understand better why people do not (want to) hear the message.

The most commonly quoted explanation is that people find statistics difficult to interpret. That is only true up to a point. If statistics can be linked to experience the problem does not seem to be too big. People understand that a horse with 12–1 odds to win is more likely to win than a horse with 30–1 odds. It is easy to relate odds to how many races a horse has won in the past. For the same reason it is somehow easy to grasp that a hurricane can make landfall in a relative wide coastal zone. People can go back into their

doi:10.21957/l6nteidw

memory for that. However, in general, if you tell someone that an event is to happen with a probability of 80%, that person will find it very difficult to accept that it is possible that the event might not happen. Intuitively most users will translate everything higher than 66% into a “yes”. And vice versa, everything lower than 33% into a “no”. Those of you who ever managed to ‘sell’ a probabilistic warning system, and asked for which threshold the user want to receive a warning mail, will know that the response will almost invariably be a request for the 66% threshold. Intuitively that is close to a decision conversion. You need many cost-loss discussions (and learning experiences) with the user to make him choose lower thresholds.

Also society has become more demanding. We want to manage the world as if it is our back garden. We have come to the point that natural disasters have become almost unacceptable. We want everything to be safe and guaranteed. We have insurances against almost anything. We build houses in low-lying areas (polders, flood plains), and expect the state to protect us 100% of the time. When it snows, we hardly accept warnings: we still want to go out and demand the civil authorities to keep the streets clear. We do not accept misses, and we hate false alarms. They are too disruptive in our precious time schedules. So, as a professional, you can hardly win. In Italy, recently, six seismologists who, in 2009, assured the public that there would be no earthquake in the region of L’Aquila (they made the mistake of converting a very low probability into a “no”) were convicted in an official court of law for manslaughter.

The future

Am I pessimistic? No. At MeteoGroup we encounter many different types of customers and we always tell them about the many possibilities of the ensemble system. Sometimes with success, particularly with professional customers who have clear requirements and are able to formulate critical thresholds. The energy trading world is a good example of a community who understand the concept of risks. A probability forecast of wind power or solar power under or over certain thresholds has proven to be a useful product.

A good example of effectively using probability forecasts is the warning system which we have set up for Dutch Rail which gets warnings based on the 10% probability of snowfall of more than a certain amount (Figure 5). When the probability threshold is exceeded they will seriously consider adjusting the train schedule such that there will be fewer trains which will be easier to manage if there are snow problems A set of hindcasts has shown that such a system generates a number of false alarms. Dutch Rail appreciates the statistics and is prepared to explain to the public that it is a choice between waiting an extra half an hour for the train or getting stuck at a station for a large part of the day.

We have to continue further developing the ensemble system and improving the probability forecasts, particularly for the short range (convective systems, fog…). And we have to continue persuading the user to appreciate the value of probabilities. But we have to be more appreciative of the reluctance of people to use such a system. Maybe then we will be more successful.

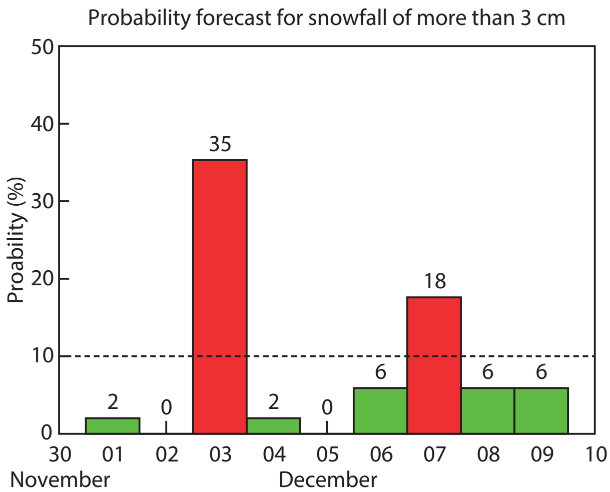

Figure 5 Example of a probability warning plot for Dutch Rail for snowfall in The Netherlands from the forecast starting at 00 UTC on 1 December 2012. Ensemble members are counted for snow for anywhere in the country, with the restriction that the temperature is below 1°C. The warning system indicated a risk of snowfall from 1 December onwards. On the 3rd light snow occurred and on the 7th there was serious snow in The Netherlands. Dutch Rail decides to contact MeteoGroup if there is a risk of snow in the forecast for 2 days ahead. On 3 December the train schedules were adjusted. Thanks to Etienne Weijers from Pro Rail for allowing this data to be presented.

Jan Barkmeijer
Is there a bright future for linear models?

Do linear models have a bright future? Without hesitation I would answer this question with an affirmative “yes”. Between 1995 and 2002 I worked in the Predictability Section, which I consider as a highlight in my working career. During that time I got intensively involved in the application of linear models and was, and still am, impressed by what they can teach us. For sure there are areas where the role of linear models has changed or even diminished in recent years. In the context of probability forecasting at ECMWF this is illustrated by the increasing use of the Ensemble of Data Assimilations (EDA) to derive initial condition perturbations at the expense of singular vectors (SVs).

Despite the introduction of the EDA, it must be said however that SVs still play a role in tuning the spread–skill relation to a satisfactory level. At the same time it seems quite plausible that, with the increasing spatial resolution, linearized versions of forecast models will become too difficult to develop. Standard operational limited area models have now reached a grid resolution of a few kilometres and are equipped with highly complex and nonlinear physics packages, which, for example, describe the dynamics of various hydrometeors such as rain, snow, cloud ice and cloud water. To what extent SVs can probe dominant perturbations growth mechanism for such high-resolution models is still a matter of debate. Therefore it is intriguing that even for these mesoscale models the 4D-Var algorithm is still producing sensible results (e.g. in case of assimilating radar data). Another area where the use of linear models seems to be limited at first sight is climate predictions. After all the term ‘linearity assumption’ alone is already rather out of place here.

Yet, it is precisely these two topics, high-resolution modelling and climate predictions, which provide examples that support the view that linear models will remain useful for the coming years. The first example describes how linear models can be employed in climate predictions so that forecasts with increased amplitudes into a prescribed anomaly direction can be realized. The second example focuses on the definition of the linearization trajectory, which is required to execute linear models. Usually the linearization trajectory is produced by a nonlinear model run. If the perturbations become too large or nonlinearities too dominant, the ‘linearity assumption’ breaks down and this nonlinear trajectory becomes sub-optimal. Variational data-assimilation is an area where this approach may be of use.

Example A: Adding flavour to climate forecasts

Determining regional climate change is not straightforward as global change, variations in the atmospheric circulation and local feedback all contribute in an intricate interplay. For example, the increase of the global mean temperature cannot fully explain why more westerly winds than usual during winter will result in warmer winters in Western Europe. Since climate forecasts tend to disagree on changes in the large-scale atmospheric circulation this also implies that regional climate change comprises a less predictable component. Still, in order to produce time series of local climate variables for the future climate, one often relies on statistical techniques to obtain local climate information. A drawback of these approaches is that often the physical consistency between different meteorological variables is violated.

By using a minimization technique it is possible to adjust a climate run to produce, for example, an atmospheric circulation over the North Atlantic sector, which is characterized by a more persistent westerly circulation in winter. The method has already been applied in other studies (*Widmann et al.*, 2010, and references therein) for models of medium complexity. It has the advantage that the average atmospheric circulation is modified while at the same time synoptic-scale variability is able to adjust to these large-scale circulation adjustments.

Figure 6 presents the first results of this minimization method as obtained with the ECMWF IFS (Integrated Forecasting System). The method consists of two steps that are repeatedly performed during the model integration.

An optimal tendency perturbation is determined such that its linear response after a certain lead time, here 5 days, has a maximal contribution in the direction of a prescribed target pattern.

The optimal tendency perturbation is subsequently applied in the climate run during the next 5 days.

Figure 6a shows the NAO (North Atlantic Oscillation) target pattern. The impact of using tendency perturbations in 5-day chunks during a 3-month run is given in Figure 6b in terms of the 3-monthly mean difference between perturbed and unperturbed run and again. Clearly, the tendency perturbations have resulted in a stronger projection onto the NAO target pattern. At the same time this model run has produced dynamically and physically consistent data which may help to identify and quantify feedback processes between atmospheric dynamics and boundary conditions.

doi:10.21957/l6nteidw

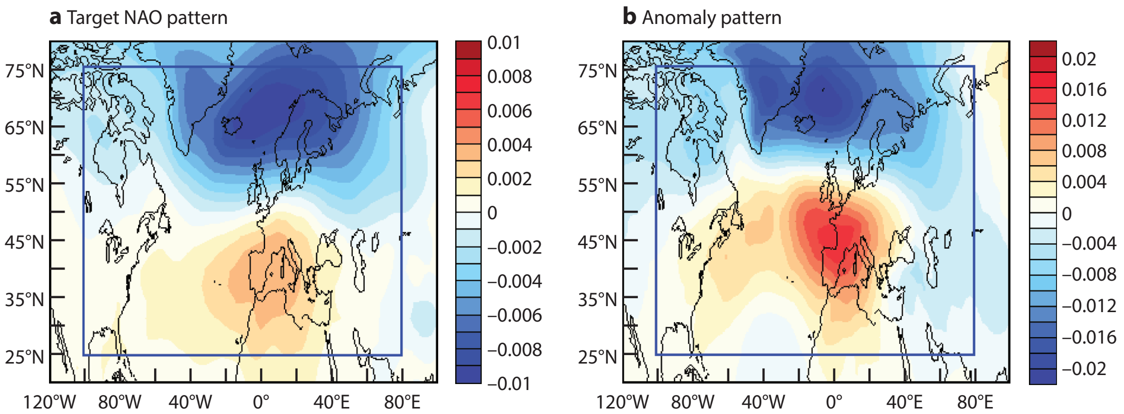

Figure 6 (a) Target NAO pattern used to determine 5-day forcings and (b) anomaly pattern given by a 3-monthly mean difference between the forced and unforced model run. The results are given in terms of the logarithm of surface pressure. Thanks to Lucinda Rasmijn for allowing this figure to be used.

Example B: Gaussian quadrature linearization trajectories

The use of linear models is limited for those time ranges for which the linear assumption is valid. By this we mean that the difference between two nonlinear model runs, and with an initial difference magnitude comparable to analysis increments, can be described by the associated linear version of the nonlinear model. Often to achieve this great effort is required to develop linearized models that capture as many features as possible of the full nonlinear model. Despite these efforts, the time window during which the linear assumption is valid ranges from around a day for models at the synoptic scale to only several hours for cloud resolving models at the kilometre scale. In order to be able to run a linear model a linearization trajectory is required. Usually such a trajectory is provided by a nonlinear model. By choosing, however, an optimal linearization trajectory it is possible to considerably extend the lead times for which the linear assumption holds. For example, in a quasi-geostrophic model an increase from a few days to over 200 days was realized.

In Figure 7 results are displayed for 10 cases. Around day 175 the initial numerical noise has reached such a level that the angle between the linearly and nonlinearly evolved perturbations starts to become non-zero. Around day 200 the differences have become too large to be neglected. While quasi-geostrophic models have only quadratic nonlinearities, higher order nonlinearities can easily be accounted for by introducing an ensemble of linearization trajectories with weights given by Gaussian quadrature points. For more details see *Stappers & Barkmeijer* (2012).

A natural application of this approach lies in 4D-Var. It is known that combining linear and nonlinear models in the 4D-Var algorithm can lead to instabilities caused by differences in phase speed of gravity waves. Using optimal linearization trajectories the distinction between inner and outer loops in 4D-Var is no longer necessary as the trajectory is already updated in the inner loop and with negligible additional computational costs. As the innovation vector is not modified in this approach it is also straightforward to derive exact equations for the adjoint-based observations impact.

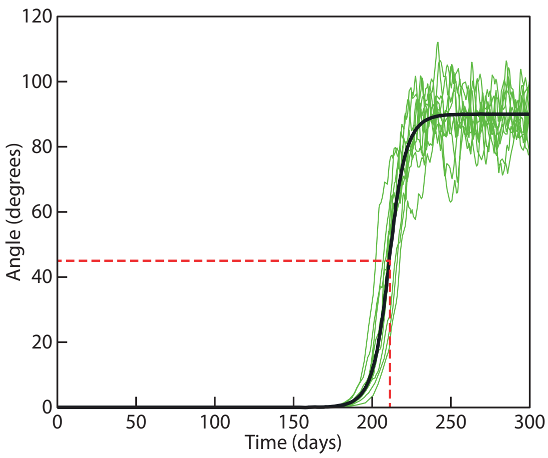

Figure 7 Angle between the nonlinearly and linearly evolved perturbation (10 cases) as a function of lead time and using the optimal linearization trajectory in the tangent linear model. Thanks to Roel Stappers for allowing this figure to be used.

Franco Molteni

What progress did we achieve with the ensemble approach to long-range predictions?

Ensemble prediction at ECMWF was born in the early 1980s with the goal of extending numerical prediction from the medium to the long range, and I was fortunate enough to get involved in the very early stages of such a research project. After my graduation, while working as a meteorological consultant for the Italian Electricity Board, I met Stefano Tibaldi (who at the time was the Head of the Numerical Experimentation Section at ECMWF), and asked him if ECMWF had any plan to use numerical models for long-range predictions. His answer, in short, was "we intend to do that, would you be interested in joining us?", and this is how my career was shaped!

I joined the Numerical Experimentation Section in January 1984 as a visiting scientist, and found that some 'deterministic' monthly range forecasts were already being run by Ulrich Cubasch. However, Stefano was keen to test the lagged-average ensemble approach advocated by *Hoffman & Kalnay* (1983); so, with Ulrich and Stefano, we ran four case studies of monthly forecasts with 9-member ensembles started from operational analyses lagged by 6 hours, using the ECMWF spectral model at T21 (yes, T21!) resolution. We soon realized that the model was drifting fast towards its own climatology, so we also ran the model from 10 earlier initial dates in order to get an estimate of the systematic error and subtract the model bias from the predicted anomalies. Thanks to a bit of 'beginner's luck', the results were sufficiently encouraging to convince the ECMWF management to invite me back in 1985 to repeat the experiments at T42 resolution. A summary of those experiments can be found in *Molteni et al.* (1987).

Research on extended-range predictability continued at ECMWF for more than a decade, using prescribed sea-surface temperature as a boundary condition for ensemble integrations of the atmospheric model. After 1987, under the leadership of Tim Palmer, the focus shifted from the monthly to the seasonal-scale, with a specific interest in the atmospheric response to the ENSO (El Niño–Southern Oscillation) phenomenon.

Seasonal predictions

In the mid 1990s, the time was ripe for starting experimental seasonal predictions using a coupled ocean-atmosphere model, and a new group was established under the leadership of David Anderson. This group managed to produce a remarkable prediction of the major El Niño event of 1997–98 (e.g. *Stockdale et al.*, 1998), and such a success was instrumental in moving seasonal prediction from a research project to an operational (albeit experimental) activity. The coupled system introduced in 1997 has been upgraded three times since then, with the latest configuration (referred to as System 4) having been implemented in November 2011 (*Molteni et al.*, 2012). The ECMWF seasonal predictions have maintained a world-class standard in the last 15 years, especially with regard to the forecasts of ENSO events, as highlighted by the recent review by *Barnston et al.* (2012).

An example of an ENSO prediction is given in Figure 8, which shows the first 14-month prediction of SST anomaly in the NINO-3 region (150°W–100°W, 5°N–5°S) performed by System 4, superimposed to the actual evolution of the anomaly. Although predicting the impact of ENSO in some regions of the world remains problematic, ENSO prediction can be considered a success story for long-range ensemble forecasting at ECMWF and for the international scientific community in general.

Monthly predictions

But what about the monthly time scale? The early experiments of the 1980s showed some encouraging results, but also highlighted the limitations of the ensemble techniques and numerical models used at that time. ECMWF decided to go back to monthly forecasting with a coupled system in 2002, with products being released operationally in 2004. Since March 2008, monthly forecasts are run as a reduced-resolution extension of the medium-range ensemble (*Vitart et al.*, 2008). This allows monthly forecast to exploit the progress in ensemble perturbation strategy and model resolution which have been introduced in the medium-range ensemble configuration. Monthly-scale predictions are particularly challenging, because in the sub-seasonal time range the effects of internal atmospheric variability are often dominant over the 'forcing' from the surface conditions.

An important sub-seasonal phenomenon, the Madden-Julian Oscillation (MJO), has proved to be a particularly difficult and elusive target for numerical modellers. However, major development in convective parametrization occurred at ECMWF and other leading centres in the last few years have significantly improved the quality of MJO predictions. Since the MJO generates teleconnections in both the northern and the southern extratropics, with important effects on Euro-Atlantic flow regimes (e.g. *Cassou*, 2008; *Vitart & Molteni*, 2010), one should expect improvements in MJO predictions to be reflected in the skill of monthly forecasts over the Euro-Atlantic region.

doi:10.21957/l6nteidw

Improvements in MJO predictions are clearly shown in Figure 9, taken from a recent study by *Vitart* (2013). Here, the correlation between the ensemble-mean and the observed North Atlantic Oscillation (NAO) index is shown using re-forecasts run with the operational monthly forecast systems used from 2002 to 2012. Two curves are displayed: one for the skill in periods of no significant MJO activity, and one for periods when MJO amplitude exceed one standard deviation. Both curves show an increase of skill with time. However, in the earlier years, the inability of the ECMWF model to represent properly the MJO and its teleconnections produced a better NAO skill when the MJO was inactive. Viceversa, in the last 5 years, improvements in modelling the MJO-NAO connections resulted in a better NAO skill during active MJO periods, leading to a substantial increase in NAO predictability in the second half of the monthly forecast range. Again, this can be regarded as a major modelling success, but one that needs a well calibrated ensemble system to be fully exploited.

The remaining challenges

Looking ahead at the next generation of ECMWF long-range forecasting systems, we face a number of challenges. Some of them regard the formulation of the coupled model: for example, we are working to implement an ocean model with increased resolution (¼° grid, 75 vertical levels) and a dynamical sea-ice module, and we need to improve our representation of stratospheric and land-surface processes. Progress is also needed in ensemble perturbation strategies: we are moving towards a coupled ensemble data assimilation system, providing consistent initial perturbations in the atmosphere and the ocean, and a representation of model uncertainties closely connected with specific components of the physical parametrization package.

Finally, we should not forget that our understanding of the interactions between atmospheric flow regimes and ocean and land-surface conditions is still unsatisfactory. Dynamical understanding is no less important than model resolution and computer power in driving advances in long-range predictions.

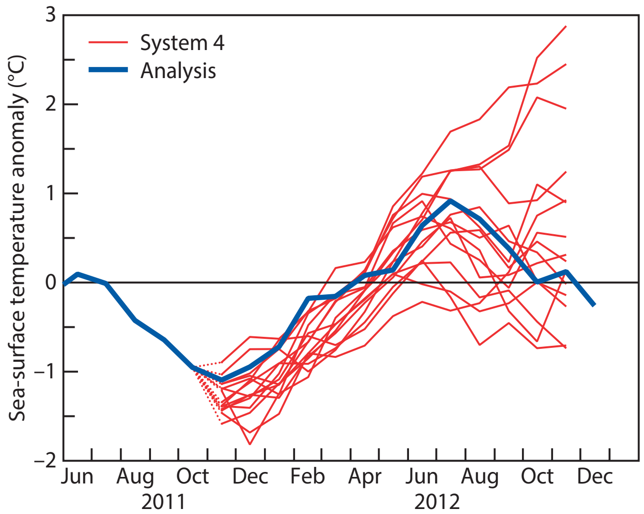

Figure 8 Predicted evolution of the monthly mean sea-surface temperature anomaly for NINO-3 SST anomaly from the 14-month forecast from System 4 started on 1 November 2011. Shown are the ensemble members and verifying analysis.

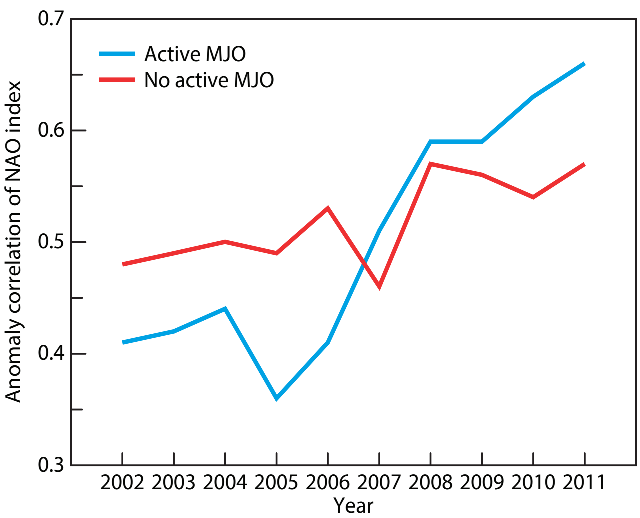

Figure 9 Anomaly correlation of ensemble mean forecasts of the NAO index for cases with no active Madden-Julian Oscillation (MJO) and those with an active MJO (amplitude > 1 standard deviation) in winters 1995 to 2001, from the re-forecasts performed to calibrate operational monthly forecasts in years 2002 to 2001.

Roberto Buizza
What next?

I joined ECMWF in October 1991, and started working on the development of the first version of the ECMWF ensemble prediction system under the supervision of Tim Palmer, with Franco Molteni, Robert Mureau and Joe Tribbia, who was visiting ECMWF for a six-month period. My first piece of work was computing singular vectors using the ECMWF tangent forward and adjoint dynamical model. I then developed the first linear and adjoint physical parametrization scheme, a vertical diffusion and surface drag scheme, which was required to compute meteorologically-sound singular vectors, and in the 3D-Var and 4D-Var assimilation schemes. This was followed by research on various predictability aspects, from adaptive targeting using singular vectors and model error simulation using stochastic methods, to applications of ensemble weather forecasts in the energy sector, in finance and flood prediction. I grew (older!!) as the ensemble evolved to become the recognized best medium-range probabilistic system.

In this contribution, I will discuss very briefly ECMWF's on-going research to improve the quality and reliability of its ensemble forecasts. I have grouped the work in four main areas.

- Model simulation of physical processes, taking model uncertainty into consideration.
- Data assimilation, including analysis uncertainty estimation.
- Ensemble forecast configuration.
- Ensemble products and calibration.

Box B summarizes the key characteristics of the current operational configuration, many of which will be referred to in the following discussion.

Model simulation of physical processes, taking uncertainty into consideration

Work to improve the simulation of physical processes also taking uncertainty into consideration will continue and is expected to lead to more realistic simulations of model uncertainties.

The fundamental role of past model improvements on ensemble prediction has been confirmed recently by a study of the time evolution of the performance of single and probability forecasts. Results provided by Martin Leutbecher have indicated that having an accurate simulation of the physical processes is essential to provide skilful medium- and long-range predictions.

A recent example of the role of model advances in improving the performance of long-range prediction is the positive impact that changes in the convection scheme had on the simulation of the Madden-Julian Oscillation (MJO), a major mode of intra-seasonal variability, which interacts with weather and climate systems on a near-global scale. Improvements introduced in January 2010 in the entrainment and detrainment, and the convection closure formulations has led to advances in the representation of atmospheric variability and in the propagation of the MJO signal through the entire integration period. Another recent example is given by the role of improvements in the land surface hydrology, convection and radiative parametrization in the prediction of the heat wave that affected the 2003 European summer. Finally, it is worth mentioning the on-going work to increase the number of vertical levels and revision of the model parametrizations in the stratosphere. Preliminary results indicate that the changes under testing are leading to a better simulation of the quasi-biennial oscillation. Tests have started to assess whether this improvement has any effect on the monthly time scale.

Work to improve the physical parametrizations will increasingly include the development of better approaches to improving the simulation of model uncertainties. The current operational ensemble uses a combination of two stochastic schemes, designed to simulate random model errors due to the parametrized physical processes and to the upscale energy transfer due to unresolved scales. The plan is to revisit the current formulations, and assess whether different approaches better linked to the physical schemes could lead to similar positive impacts. There are several promising lines of research.

One area under investigation in collaboration with the Finnish Meteorological Institute is based on an automatic estimation of the distribution (rather than one single value) of model parameters to be used to sample the parameters in each member of the ensemble.

Another area of research follows from the preliminary results obtained in collaboration with the UK and the Spanish Meteorological Institutes that suggested it is worth exploring whether tendency perturbations would be better computed separately for each single parametrization schemes than for the total tendency.

 doi:10.21957/l6nteidw

Configuration of the ECMWF operational ensemble forecasts

B

The ECMWF global medium-range forecast comprises a high-resolution forecast (HRES) and an ensemble of lower-resolution forecasts (hereafter named ENS rather than EPS, following a recent revision of the terminology used at ECMWF, see *ECMWF Newsletter No. 133*, 1–13). The follwing provides some details about the configuration of the operational ensemble forecasts.

- ***Membership.*** 51 members, runs twice a day at 00 and 12 UTC with a 15 day forecast range; twice weekly it is extended to 32 days (on Mondays and Thursdays at 00 UTC).
- ***Atmosphere resolution.*** For the atmosphere, variable horizontal resolution, with a spectral triangular truncation T639 (about 32 km) up to day 10 and T319 (about 65 km) afterwards, with a spatial linear grid; in the vertical, ENS uses 62 levels up to 5 hPa.
- ***Ocean wave resolution***. ENS is coupled to the WAM wave model with 55 km resolution, and 24 directions and 30 frequencies up to day 10, and 12 directions and 25 frequencies afterwards.
- ***Ocean currents resolution***. ORCA100z42 grid, with a 1-degree horizontal resolution and 42 vertical layers.
- ***Ocean currents coupling***. ENS runs with persisted sea-surface temperature (SST) anomalies up to day 10, and coupled to the NEMO ocean model afterwards.
- ***Uncertainty simulation*** – ENS has been designed to simulate initial and model uncertainties;
- ***Initial uncertainties.*** Atmosphere initial uncertainties are simulated by adding to the HRES-4DVAR T1279L91 (16 km) analysis two sets of perturbations generated using:
  – 50 forecast singular vectors computed at T42 resolution over different regions of the globe (NH, SH, tropics) with maximum total-energy growth over 48 hours.
  – 6-hour forecasts at T399 resolution started from the 11 perturbed members of the Ensemble of Data Assimilations (EDA).
- ***Ocean currents uncertainties***. Ocean initial uncertainties are simulated by using the NEMOVAR ensemble of ocean analysis.
- ***Model uncertainties in the atmosphere*** – Model uncertainties are simulated only in the atmosphere using two stochastic schemes:
  – The stochastically perturbed parametrized tendency (SPPT) scheme is designed to simulate random model errors due to parametrized physical processes; the current version uses 3 spatial and time level perturbations.
  – The stochastic back-scatter (SKEB) scheme is designed to simulate the upscale energy transfer induced by the unresolved scales on the resolved scales.
- ***Model climate and calibration***. Some of the ensemble products (e.g. the Extreme Forecast Index, or weekly-average anomaly maps) are constructed by comparing the most recent ensemble forecast with the model climate estimated using the re-forecast suite, which includes 500 forecasts. For each date (e.g. 14 December 2012), these 500 forecasts are defined by combining 5-member forecasts run for 5 initial dates centred on the current date (1, 7, 14, 21 and 28 December) of the past 20 years (these ensembles start from ERA-Interim central analysis, use singular vectors of the day but EDA-based perturbations computed for the current year since the EDA has been running only since 2010).

Data assimilation, including the estimation of analysis uncertainty

One of the crucial aspects of the design of ensemble prediction is the definition of the ensemble of initial states. In the current ensemble configuration, initial conditions are defined by adding perturbations generated using singular vectors and those based on the Ensemble of Data Assimilations (EDA) to a central analysis defined by the high-resolution four-dimensional analysis (interpolated to the ensemble resolution). Thus their quality depends on both the central (unperturbed) analysis and the initial perturbations.

With respect to the central analysis, results so far have indicated that centring the ensemble on the high-resolution analysis provides the best results. It is expected that this design will remain in place, at least until the EDA quality and resolution improves and its membership increases to 51 (from the current 11). Improvements in the data-assimilation algorithm (e.g. due to increased resolution, improvements in the statistical assumptions and extension of the assimilation time window) are expected to lead to improvements in the ensemble performance.

With respect to the initial perturbations, the introduction in 2010 of EDA-based perturbations addressed two known weaknesses of the old operational ensemble.

- Singular vectors are only marginally sensitive to observation characteristics (error in general, including coverage and representativeness).
- Singular vectors are too localized in space if compared to analysis error estimates, and poorly sample some areas of the world (e.g. the tropical band).

The plan for the future is to have an even closer coupling between the EDA and the ensemble, so that the forecasts benefit from planned EDA advances (e.g. in the planned increase in the number of members, in the resolution of the outer and/or inner loops, in the specification of the observation errors used to perturbed the observations in the EDA members). A tighter coupling between the EDA and the ensemble will also be achieved by the addition of EDA-based land-surface perturbations. Preliminary results indicate that including soil moisture and soil temperature perturbations will lead to spread increases in regions where the forecast is under dispersive (e.g. in the early time range for variables close to the surface). As shown in Figure 10, after 24 hours, EDA-based land-surface perturbations would induce a larger soil-moisture spread than EDA-based upper-level perturbations over Brazil (top panels). This signal propagates to the free atmosphere and leads in some specific locations to 2-meter temperature spread of a comparable amplitude to the spread induced by the EDA-based upper-level perturbations (bottom panels).

Another area where work has started to improve the simulation of initial uncertainties is to include sea-surface temperature perturbations, mimicking what is currently done in the seasonal prediction System 4 using the ensemble of 5-member ocean analyses. The inclusion of EDA-based land-surface and sea-surface temperature perturbations can be seen as the first steps of a strategy that aims to improve the simulation of initial uncertainties of the coupled ocean-land-atmosphere ensemble.

The planned tighter coupling of the EDA and the ensemble will also imply implementing a consistent approach in the simulation of model uncertainty in both the EDA and the ensemble. Today, the EDA uses only one of the two model uncertainty schemes used in the ensemble, namely the SPPT (the stochastically perturbed parametrized tendency scheme is designed to simulate random model errors due to parametrized physical processes). The other scheme, which is based on SKEB (stochastic back-scatter scheme is designed to simulate the upscale energy transfer induced by the unresolved scales on the resolved scales), induces undesirable features when the EDA is used to compute background error statistics for the high-resolution 4D-Var. The aim is to develop a unified approach to the simulation of model uncertainties that can be used both in assimilation and prediction mode. This will be taken into considerations during the revision and upgrade of the current model error schemes.

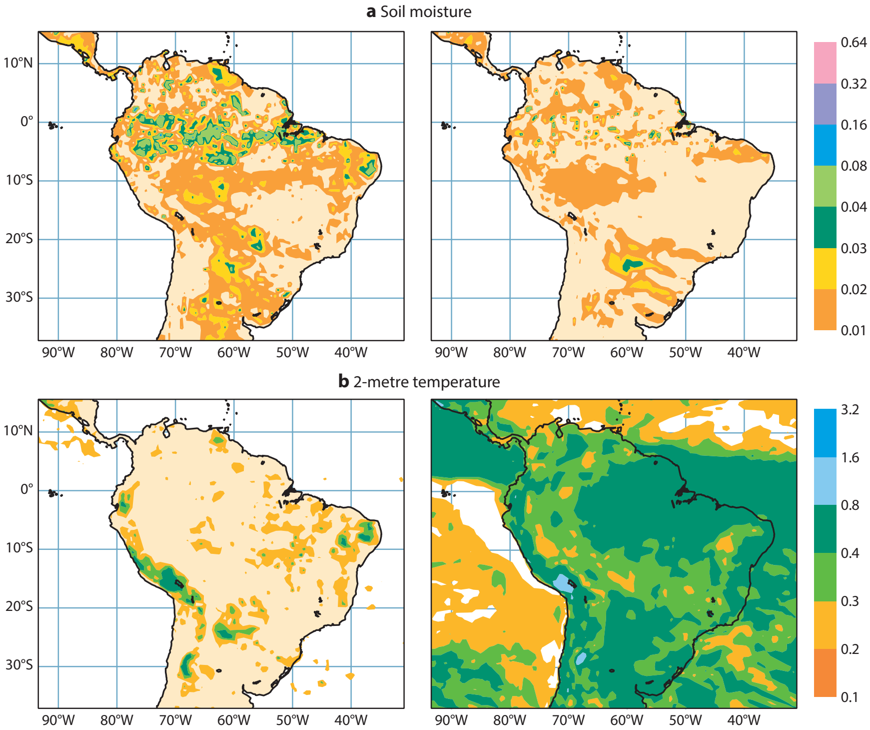

Figure 10 Average ensemble standard deviation at 24 hours from 8 cases for (a) soil-moisture and (b) 2-metre temperature of ensembles run in two different configurations: with only EDA-based land-surface perturbations (left panels) and with only EDA-based upper-level perturbations (right panels).

doi:10.21957/l6nteidw

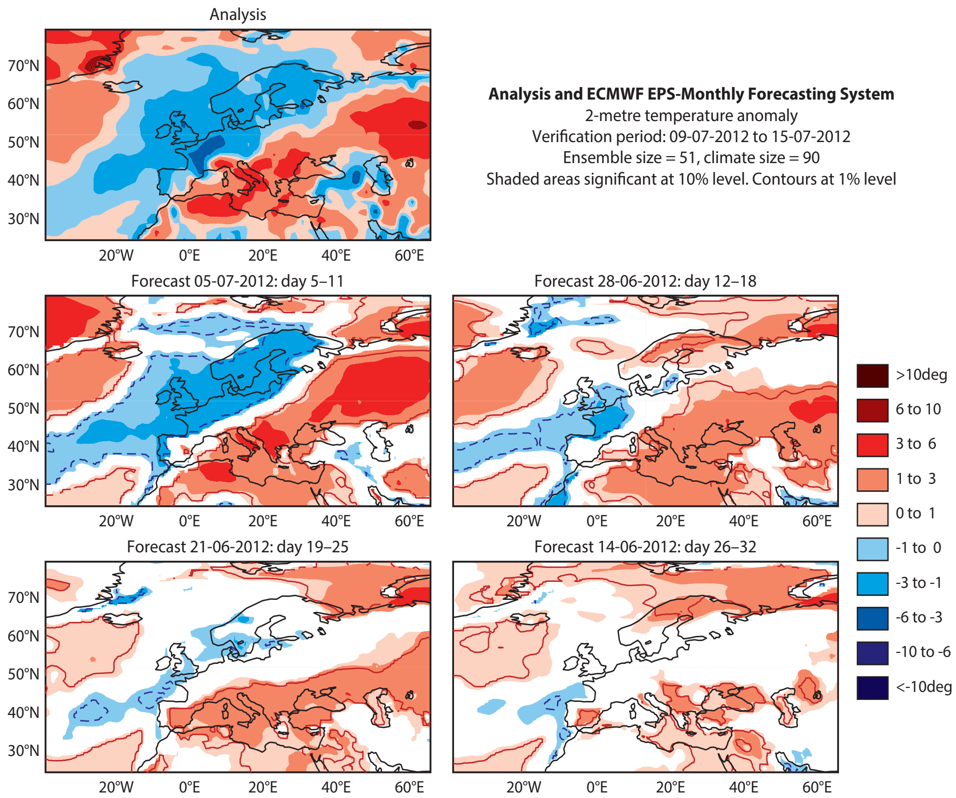

Figure 11 This is an example of a monthly forecast showing weekly-average 2-metre temperatures valid for the week 9–15 July 2012 (top panel), when Northwest Europe was experiencing cold anomalies and Southern Europe hot, summery weather! The other four panels show the weekly average forecasts issued on 5 July (5–11 days), 28 June (12–18 days), 21 June (19–25 days) and 14 June (26–32 days). The temperature anomalies are shaded when the ensemble forecast distribution is statistically significantly different from the climatological distribution. In some areas (e.g. UK and Italy) cold and warm conditions were correctly predicted up to three weeks in advance.

Ensemble forecast configuration

Twenty years ago, ensembles were produced three times a week, at 12 UTC, and included 33 members run at T63L19 resolution. Since then, many changes in the ensemble configuration were introduced.

- Daily production started on 1 May 1994, still once a day.
- In 1998, each member was run with a coupled ocean wave model (WAM).
- From 2004 forecasts were issued twice a day, at 00 and 12 UTC.
- Between 1992 and today, membership increased from 33 to 51, and resolution from T63L19 to T639L62.

In 2008, the medium-range and monthly ensembles were joined, with the adoption of a variable-resolution strategy, forecasts were extended to 15 days every day and to 32 days once a week, and the coupling to the ocean current model was introduced (an example of a monthly forecast is given in Figure 11). At the same time, re-forecasts started been generated to allow the bias-correction and calibration of some products.

Since 2000, the ECMWF ensemble forecasts have been used to provide initial and boundary conditions to limited-area ensembles run by ECMWF Member States (among them, for example, the ones developed by the COSMO Consortium, Norway and France) and to drive ensemble flood forecasts (e.g. the European Flood Awareness System).

Is the current operational configuration the best for our users, or would they benefit more from a different set-up? For example, several global ensembles (e.g. the Canadian and the American ones) are now issued four times a day to provide users with a more frequent forecast updates. Should ECMWF follow their example?

Some of our Member States have informed us that the 00 UTC ensemble arrives too late on forecasting desks to be used early in the morning. We have also been informed that the 00 UTC ensemble is completed too late to be used to drive limited-area ensemble forecasts that have to generate forecasts by, say, 6 am local time. Should ECMWF at least also run an ensemble at 18 UTC to help the limited-area community?

To accommodate the users' requests while keeping the production cost to affordable levels, a possibility would be to move from symmetric 00 and 12 UTC production schedules, towards a four-times-a-day production, with each schedule not necessarily with the same membership, resolution and forecast length. For example, we could extend the 51 ensemble members to 15/32 days only at 00 UTC, but limit the 6, 12 and 18 UTC ensembles to 5–10 days, possibly with a smaller membership to reduce production costs. Possible changes to the operational configuration along these lines could be explored with the aim of making ECMWF ensemble forecasts of more value to our users.

Ensemble products and calibration

The last topical area in which further advances are expected is in the product generation, especially exploiting the re-forecast dataset to calibrate the ensemble forecasts.

Today, only a few products use the re-forecast dataset to either bias-correct some of the fields or to estimate the model climate. The latter is required to translate the ensemble probabilities into indices that highlight how far an ensemble forecast distribution is from the climatological one. An example of such a product is the extreme forecast index (EFI), which is routinely computed for surface variables such as total precipitation, 2-metre temperature and 10-metre wind speed.

As shown by *Hagedorn et al.* (2012), re-forecast calibrated ECMWF ensembles are more skilful than multi-model ensembles that include four of the best global ensembles. Following these results, a study has been initiated, in collaboration with the University of Heidelberg, to assess how we can further exploit the re-forecast data to generate a wider range of calibrated ensemble products. These could include, for example, meteograms at all grid points, and/or at specific locations, of calibrated ensemble distributions. This study will also look at which methodologies would be more appropriate to use for the different weather variables. It is expected that work in this area in collaboration with Academia and Member States could lead to more accurate, reliable and valuable ensemble forecasts.

1975, 1992, 2012, and beyond

We hope that this article has given an interesting historical overview on how ECMWF evolved its approach to numerical weather prediction. In 1975, in his article in the proceeding of the first '*ECMWF Seminar on the Scientific Foundation of Medium Range Weather Forecasts*', Cecil Leith stated that:

"Numerical weather prediction can never be exact owing to errors in the determination of the initial state and to external error sources arising from discrepancies between the dynamics of numerical models and that of the real atmosphere"

It took us until 1991 to develop and implement as part of ECMWF's forecasting suite the first ensemble prediction system that simulated the effect of 'initial state' error sources, and another decade to take 'external error sources' into account. Today, 20 years after operational implementation, the ECMWF medium-range ensemble forecasts are reliable and skilful up to one month ahead.

We are confident that continuous work and collaboration with the scientific community will lead to further advances and use of probabilistic forecasts.

Further reading

Augier, P. & **E. Lindborg,** 2012: A new formulation of the atmospheric spectral energy budget with application to two high resolution general circulation models. http://arxiv.org/abs/1211.0607

Barnston, A.G., M.K. Tippett, M.L. L'Heureux, S. Li & **D.G. DeWitt,** 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Am. Meteorol. Soc.,* **93**, 631–651

Cassou, C., 2008: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature,* **455**, 523–527.

ECMWF, 1979: *Proceedings from Workshop on Stochastic Dynamic Forecasting,* ECMWF, Reading, UK, 17–19 Oct. 1979

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

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

Hagedorn, R., R. Buizza, M.T. Hamill, M. Leutbecher & **T.N. Palmer,** 2012: Comparing TIGGE multi-model forecasts with re-forecast calibrated ECMWF ensemble forecasts. *Q. J. Roy. Meteorol. Soc.,* **138**, 1814–1827.

Hoffman, R. & **E. Kalnay,** 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **354**, 100–118.

doi:10.21957/l6nteidw

Hollingsworth, A., 1979: An experiment in Monte Carlo forecasting. In Proc. *ECMWF Workshop on Stochastic Dynamic Forecasting,* ECMWF, Reading, UK, 17–19 Oct. 1979, 65–85.

Katz, R.W. & **A.H. Murphy,** 1997: *Economic Value of Weather and Climate Forecasts,* Cambridge University Press.

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

Leith, C.E., 1975: Statistical and statistical-dynamical methods in medium range weather forecasts. In *Proceedings of ECMWF Seminar on the Scientific Foundation of Medium-Range Weather Forecasts,* ECMWF, Reading, UK, 1–12 Sept. 1975, 307–379.

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

Lorenz, E.N., 1969: The predictability of a flow which contains many scales of motion. *Tellus,* **21**, 289–307

Molteni, F., U. Cubasch & **S. Tibaldi,** 1987: 30 and 60-day forecast experiments with the ECMWF spectral models. In *Proceedings of First ECMWF Workshop on Predictability in the Medium and Extended Range,* ECMWF, Reading, UK, 17–19 Mar. 1986, 51–108.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen,T. Palmer & **F. Vitart,** 2011: The new ECMWF seasonal forecast system (System 4). *ECMWF Tech. Memo. No. 656.*

Mureau, R., F. Molteni & **T.N. Palmer,** 1991: Ensemble prediction using dynamically-conditioned perturbations. In *Proc. ECMWF Workshop on New Developments in Predictability,* ECMWF, Reading, UK, 13–15 Nov. 1991, 143–174.

Murphy, J.M. & **T.N. Palmer,** 1986: Experimental monthly forecasts for the United Kingdom. Part II: A real-time long-range forecast by an ensemble of numerical integrations. *Meteorol. Mag.,* **115**, 337–344.

Palmer, T.N., 2006: Predictability of Weather and Climate from theory to practice. In *Predictability of Weather and Climate, Palmer and Hagedorn Eds,* Cambridge Univ. Press, 1–29.

Palmer, T.N. & **S. Tibaldi,** 1988: On the prediction of forecast skill. *Mon. Wea. Rev.,* **116**, 2453–2480

Stappers, R. & **J. Barkmeijer,** 2012: Optimal linearization trajectories for tangent linear models. *Q. J. R. Meteorol. Soc.,* **138**, 170–184.

Stockdale, T.N., D.L.T. Anderson, J.O.S. Alves & **M. Alonso Balmaseda,** 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature,* **392**, 370–373.

Tennekes, H., A.P.M. Baede & **J.D. Opsteegh,** 1987: Forecasting forecast skill, In *Proceedings of ECMWF Workshop on Predictability in the Medium and Extended Range,* 17–19 March 1986, 277–302.

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

Vitart, F., R. Buizza, M. Alonso Balmaseda, G. Balsamo, J.-R. Bidlot, A. Bonet, M. Fuentes, A. Hofstadler, F. Molteni & **T.N. Palmer,** 2008: The new VAREPS-monthly forecasting system: a first step towards seamless prediction. *Q. J. R. Meteorol. Soc.,* **134**, 1789–1799

Vitart, F. & **F. Molteni,** 2010: Simulation of the MJO and its teleconnections in the ECMWF forecast system. *Q. J. R. Meteorol. Soc.,* **136**, 842–855.

Vitart, F., 2013: Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years. Submitted to *Q. J. R. Meteorol. Soc*.

Widmann, M., H. Goosse, G. van der Schrier, R. Schnur & **J. Barkmeijer,** 2010: Using data assimilation to study extratropical Northern Hemisphere climate changes over the last millenium. *Climate of the Past*, **6**, 727–644.