The future of ensemble prediction

Ensemble prediction is a technique by which a forecast model is run several times from initial conditions which differ by amounts consistent with uncertainties in the initial state. Since its inception in 1992 (*Palmer et al. 1993)*, the ECMWF ensemble prediction system (EPS) has become an established part of operational forecasting at ECMWF. The original basis for the development of the EPS was the notion that single deterministic forecasts are not consistent with the "scientific method", in the sense that the result from any scientific prediction is not complete without an estimate of the likely error associated with measurement and other experimental inaccuracy (*Tennekes 1991*). It would be easy to estimate the likely error associated with a numerical weather forecast using some average error based on a large number of past cases. However, the atmosphere is chaotic, and that implies not only that forecast accuracy can be sensitive to small uncertainties in starting conditions, but also that the amplification of the initial uncertainty itself depends on the initial state (*Palmer 2000*).

The current configuration of the EPS is briefly described below, and examples are shown of how the EPS can quantify the risk of severe weather in the medium range, in circumstances where single deterministic forecasts fail. This ability to quantify the risk of severe weather implies that the potential economic value of the EPS is much higher than can be obtained from single deterministic forecasts.

The EPS has now entered a more mature phase of development and, as discussed below, there is now an important practical objective to guide its development - as a tool for quantitative risk management in weather-sensitive commercial/humanitarian applications. This maturity is beginning to lead to the consideration of more direct linkage between the EPS and specific user application models. This will make the value of proposed developments to the EPS easier to quantify.

A brief description of the EPS

At the time of writing, the ECMWF EPS comprises 50+1 integrations of the operational ECMWF model at $T_L159L40$ (120 km) resolution (*Buizza et al. 1998*). The initial perturbations are based on the dominant singular vectors (finite-time instabilities) of the forward tangent propagator between day 0 and day 2 (*Buizza and Palmer 1995*). The singular vectors

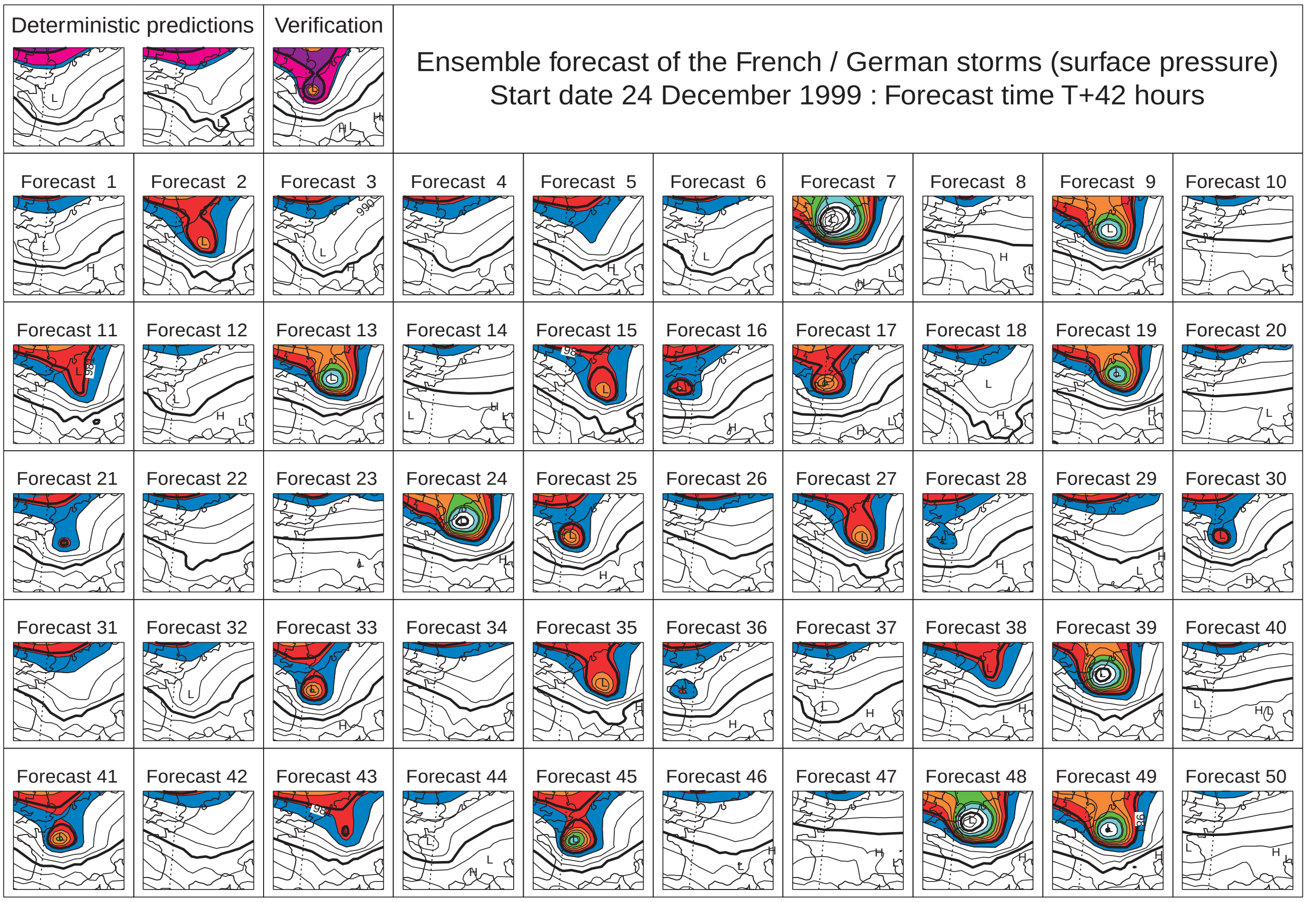
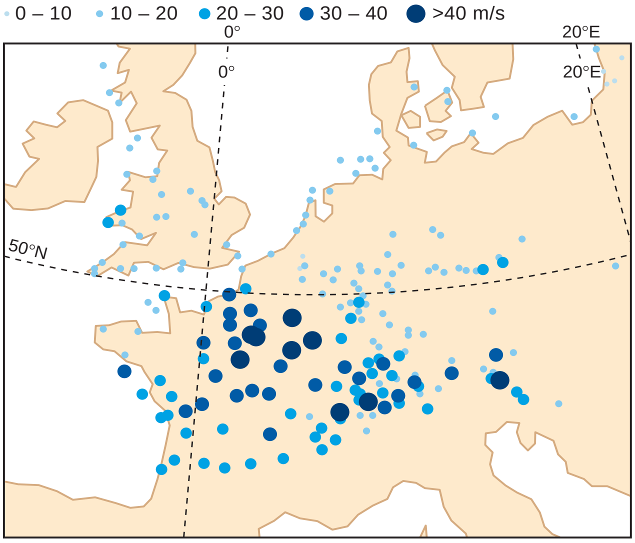


Figure 1 Stamp maps of surface pressure (coloured below 980 hPa) for a 42 hour forecast of the first Christmas storm at 06 UTC 26 December 1999. The T_L159 control forecast, the T_L319 operational forecast and the individual ensemble members are shown. Even though the single deterministic forecasts do not forecast the storm, the EPS shows that there is a significant risk of such a storm (indeed a very large risk by comparison with its climatological frequency).

are linearly combined taking into account the expected amplitude of analysis error (*Molteni et al.* 1996). The principal reason for using singular vectors is to "side-step" two problems. Firstly there are many unquantified assumptions in data assimilation and the true analysis-error probability density function (PDF) is, in practice, not well known; this means that a true random sampling of the initial PDF is impossible. Secondly, the number of samples possible is very much smaller than the number of possible choices of initial perturbation. Inadequate sampling will generally lead to an overly confident estimate of forecast reliability.

In addition to initial perturbations, the model equations in the EPS are perturbed using the stochastic physics scheme (*Buizza et al. 1999*). In this scheme, the physical tendencies are perturbed stochastically at each time step and grid point, representing random uncertainty in the formulation of the model equations.

It is proposed to increase the horizontal resolution of the operational EPS to T_L255 (80 km) in the second half of 2000. In addition to operational EPS results, some of the results below are shown for this higher resolution.



Probability of gusts exceeding 40 m/s (T255 H+42 fc)

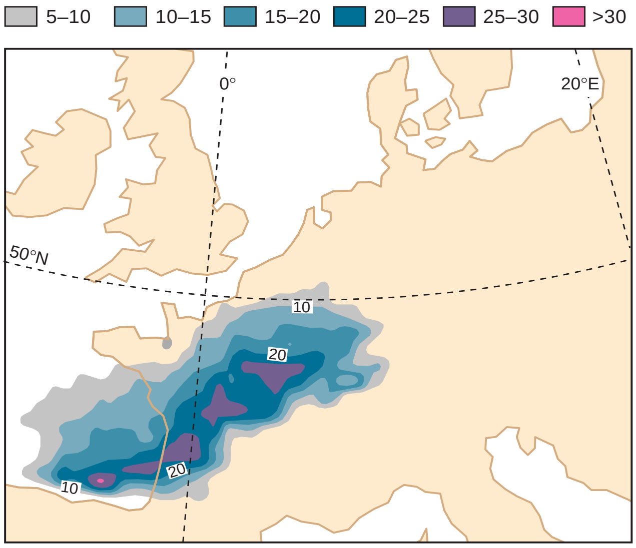


Figure 2 Observed gusts (top panel) and the 42-hour EPS forecast probability of gusts exceeding 40 m/s (computed from the 850 hPa wind speed, bottom panel).

Ensemble predictions of severe weather.

There is no doubt that National Meteorological Services are judged, by the public at large, by their ability to provide timely warnings of severe weather. By their nature, severe weather events are often associated with intense atmospheric developments on rather small scales. Such developments usually arise because of strong instability of the flow, suggesting that these forecasts, more than others, will be sensitive to small uncertainties in starting conditions.

In such situations, the guidance from single deterministic forecasts can be unreliable, and this is often manifest in day-to-day inconsistency of such forecasts. In this section, three case studies are shown demonstrating the value of risk assessment derived from corresponding ensemble predictions.

Figure 1 shows individual members of a 42-hour ensemble forecast from a T_L255 ensemble integration for the 26 December 1999 storm over France (in terms of surface pressure). Both the T_L319 operational forecast and the T_L159 control forecast failed to predict the storm at this range (the deterministic forecasts were more successful at both shorter and longer ranges, illustrating the point made above concerning the unreliable nature of single deterministic forecasts of severe weather). It can be seen that many members of the ensemble successfully simulated the storm. As such, the risk of such a storm was forecast with a probability vastly exceeding its climatological probability.

As an example of a possible forecast product, based on such an ensemble, Figure 2 shows the probability of gusts exceeding 40 m/s, based on a simple gust parametrization (that of mixing down air from 850 hPa; Ernst Klinker 2000, personal communication). This parametrization, applied to the ensemble predictions, forecasts a probability up to 30% of such damaging gusts over a swath of northern France. Six hours later (not shown) the region of maximum probability of such gustiness had moved on to Germany. This ensemble forecast, continued to 96 hours, also gave a significant probability of a second storm over France (not shown). The operational EPS (at T_L159 resolution, also not shown) also gave a warning of the storm. However, the gustiness product was somewhat weaker and less well positioned at this lower resolution.

A second example is shown in Figures 3 and 4. The severe snow storm over the US East Coast on 25 and 26 January 2000 was missed by the main operational models. The failure to issue a warning of the risk of severe weather was criticised in the media. The storm caused severe damage with loss of lives, mainly because of the intense snowfall associated with its passage.

Figure 3 shows some of the individual members of the EPS (in terms of 1000 hPa geopotential height and precipitation). It can be seen that the deterministic forecasts have the precipitation bands out to sea, whilst individual EPS members correctly develop an intense low-pressure system over land, with associated precipitation. It can be noted that the ensemble-mean forecast for this event is not

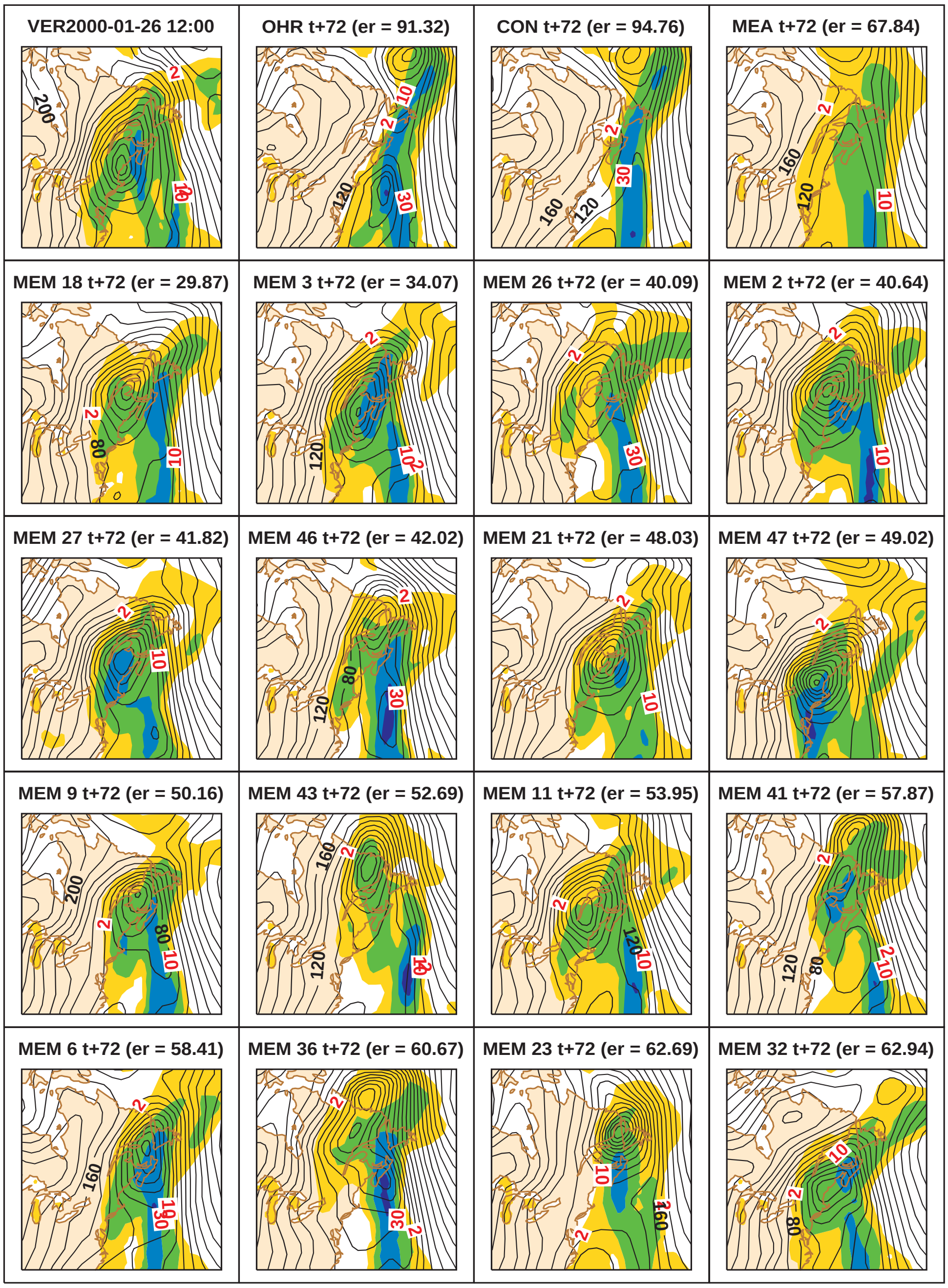


Figure 3 Postage stamp maps for 1000 hPa geopotential height and precipitation for the verifying analysis (first top panel), and for 72-hour forecasts started at 12 GMT on 23 January 2000 of the high-resolution model (second top panel), the EPS control (third top panel), the ensemble-mean (fourth top panel), and the first 16 EPS perturbed-members (subsequent panels) with the smallest root-mean-square-error for geopotential height inside a region centred around the observed cyclone. Contour interval is 20 m for geopotential, and contour isolines are 2, 10, 30 and 60 mm for precipitation.

itself successful. This should be taken as a warning that trying to use the ensemble to provide a modified deterministic (i.e. non-probabilistic) prediction is not itself a reliable procedure.

Figure 4 shows the 0-24 ECMWF operational forecast from 12 UTC on 25 January, which can be considered as a good approximation to the observed precipitation field (1 mm of water is equivalent to 1 cm of snowfall). The 72-hour forecast from the ECMWF operational model failed to predict intense snowfall over the land. By contrast, the 72-hour EPS forecast a probability of 10-to-60% (depending on the geographical location along the US East Coast) of more than 10 mm/d of precipitation, and 2-to-30% probability of more than 20 mm/d. The EPS also indicated a significant probability of enhanced 10 m wind speeds compared with the deterministic forecasts (not shown).

Figure 5 shows an example of an ensemble prediction of the position of tropical cyclone Eline which devastated Mozambique in February 2000 up to 5 days ahead from 18 February 2000. In this example, initial perturbations were made using tropical diabatic singular vectors (*Barkmeijer et al. 2000, Puri et al. 2000*). Such perturbations are not yet computed operationally, but it is hoped to introduce them into the EPS in 2001. The ensemble shows that there is a high probability that the cyclone will strike Mozambique, although at a range of 5 days, the precise position of landfall is uncertain (interestingly the ensemble shows a somewhat bimodal probability distribution of landfall, with the verifying analysis and the high resolution operational forecasts taking different modes).

The potential economic value of probability forecasts

It has been argued that single deterministic forecasts of extreme weather are, by their nature, likely to be unreliable. Ensemble forecasts, on the other hand, should be capable of estimating the risk of extreme weather more reliably. How can this be quantified? Consider a simple decision model used to estimate the potential economic value of weather forecasts (*Murphy 1977, Richardson 1998, 2000*). A user can suffer a loss L if a meteorological event E occurs and no precautionary action is taken. The loss is avoided if precautionary action at cost C is taken. The weather forecasts are used to decide when to take precautionary action.

Figure 6 shows the value of the EPS as a function of the user ratio C/L for the event E: precipitation greater than 10 mm/day (of water equivalent). Also shown for comparison are the values of two deterministic forecasts – the T_L159 control and the ensemble mean. Zero value means that the information provided by the forecasts is of no more use than information associated with a knowledge of the climatological frequency of E. A value of unity would imply a perfect deterministic prediction system.

Both of the deterministic forecasts provide a simple but unreliable criterion for deciding when to take precautionary action: take action when E is forecast; do not take action when E is not forecast. For the EPS, a more sophisticated decision strategy, making use of the forecast probability distribution can be adopted. A user with small C/L should decide always to take precautionary action, except when the probability of E is sufficiently small. A user with C/L close to unity should only take precautionary action when the forecast probability of E is sufficiently high. In general the user should take precautionary action when the probability of E is greater than C/L.

The ensemble-mean forecast gives very poor value, worse than the control, whilst the full ensemble value remains high. The reason for this is straightforward; the ensemble mean is a smooth forecast, and will consistently underpredict the more extreme types of events. On the other hand, probability forecasts made using the ensemble probability distribution prove much more valuable in forecasting the risk of this precipitation event.

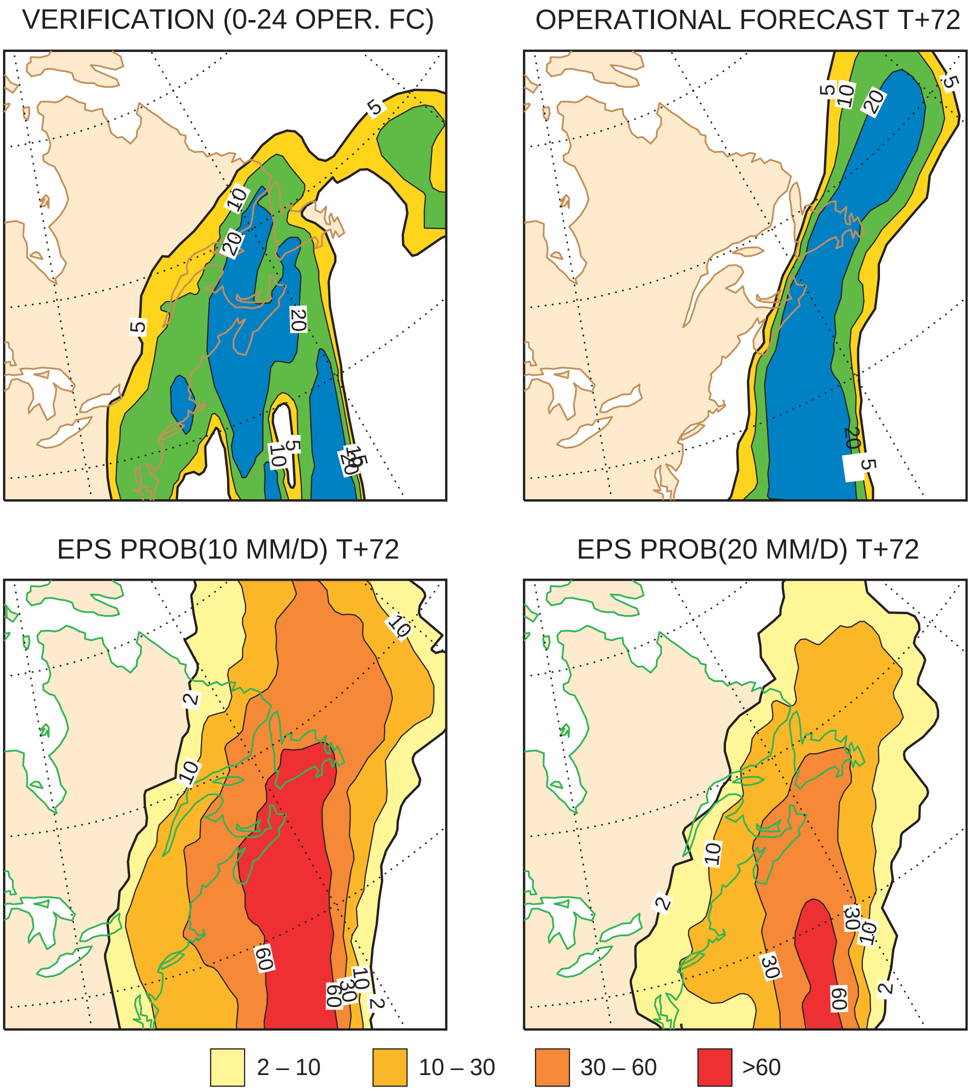


Figure 4 Verification field for precipitation (defined as the 0-24 hour high-resolution forecast started at 12 UTC on 25 January 2000, top left panel), 72-hour forecast from the high-resolution model (right top panel), and 72-hour EPS forecasts of the probability of more than 10 mm/d (bottom left panel) and 20 mm/d (bottom right panel) of precipitation. Contour isolines are 5, 10, and 20 for precipitation and 2%, 10%, 30% and 60% for probabilities.

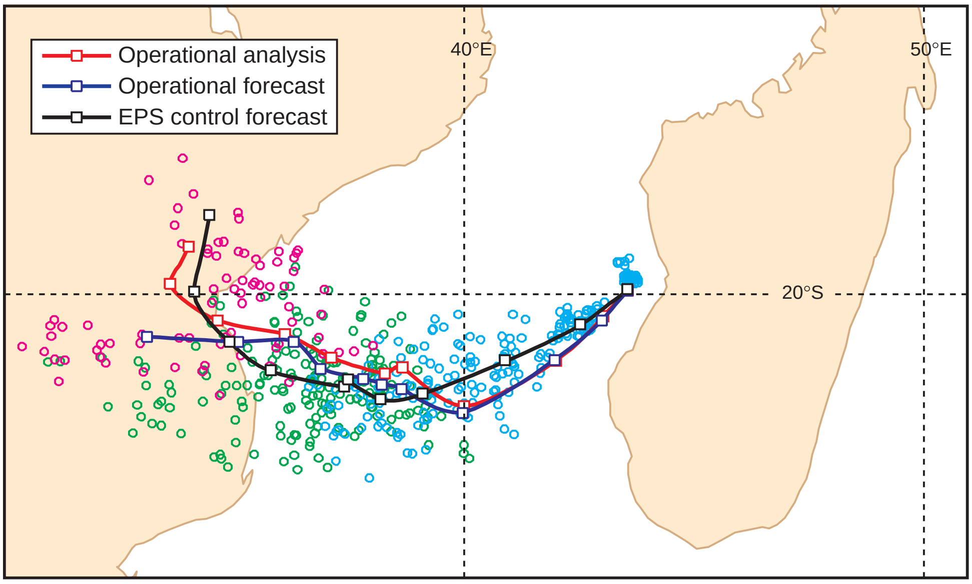


Figure 5 An experimental EPS showing the ensemble of tropical cyclone positions up to 5 days ahead for cyclone Eline which devastated Mozambique in February 2000. Symbols denote the cyclone position in the EPS 0-to-24 hour (blue), 48-to-72 hour (green) and 96-to-120 hour (purple) forecasts. The red line shows the observed path (square symbols identify the position every 12 hours), the blue-line the path predicted by the high-resolution model and the black line the path predicted by the EPS control.

EPS as a quantitative tool for risk management

The previous section has illustrated the value of the EPS as a quantitative tool for risk management; by contrast it has been shown that single deterministic forecasts are less valuable. In order to develop this notion of value, EPS output needs to be linked directly with specific user application models. Many examples of such user application models can be envisaged: prediction of damage from flood, storm, drought, prediction of electricity demand, prediction of the pay-out on some weather-related financial contract. A schematic of this notion is illustrated in Figure 7. Output from each member of the EPS is first passed through an empirical model to correct for model bias, and to apply down-scaling to give values appropriate to specific geographical points. For each member of the ensemble the corrected forecast data is fed into the user application model. The end product will be a probability distribution of damage/demand/pay-out. If the probability of abnormally high damage/demand/pay-out is sufficiently high, then the user can take appropriate action. The trigger for such action depends on the user, for very high potential loss, the threshold probability for such an abnormality might be rather low. In this way, the need for direct probabilistic forecasts of weather parameters is obviated, and questions such as "How do customers make use of probability forecasts?" are circumvented.

As an example of such a procedure, *Hoffschildt et al. (2000)* considered the problem of ship routing, computing over a season of forecasts, an ensemble of optimal ship routes from Brest to New York, based on the individual members of the EPS (see also *Janssen, 2000)*. In situations where the ensemble of ship routes is consistent, the master can plot his route with confidence. In cases where the ensemble of ship routes is broad there is uncertainty in the optimum route, and if he has this option, the master might be advised to stay in port, until a more clear-cut option is available. If this is not an option then, as shown by *Hoffschild et al. (2000)*, the most likely optimal route based on the ensemble can be chosen. Because of non-linearity in the prediction process, this "most likely" optimum route may differ from the optimum route provided by either of the deterministic forecasts. In general *Hoffschildt et al. (2000)* found that significant fuel saving could be had using the most likely optimal ship route defined by the EPS, compared with the single estimate of optimal ship route defined by the high-resolution T_L319 single deterministic forecast.

This type of "end-to-end" analysis of the EPS is still in its infancy. The true value of the EPS will not be realised until it becomes an established procedure. This may require a radical change in the way in which EPS data are used in Member State National Meteorological Services.

Computational Demands for the EPS

The EPS is computationally demanding; although, on the other hand, it is a perfect application for multi-processor supercomputers. There are four principal components that contribute to the cost of the EPS:

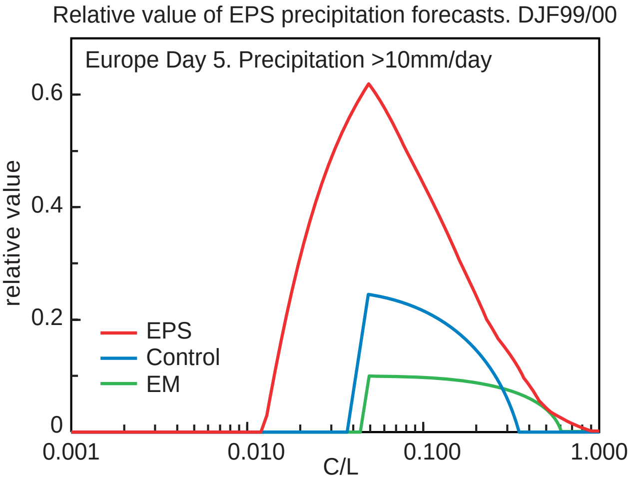


Figure 6 Potential economic value (see Richardson, 2000 for details) for forecasts of 10 mm/day, for a range of users with variable cost/loss ratios. Solid line: EPS, dashed line: control forecast, chain dashed line: ensemble mean.

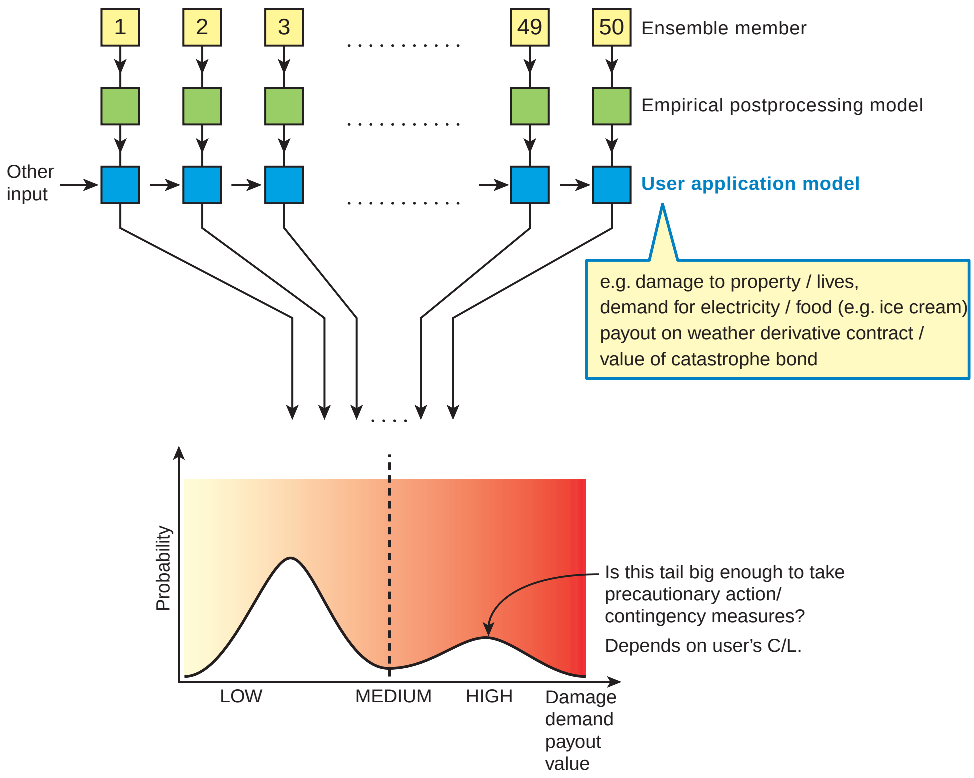


Figure 7 A schematic diagram showing that for quantitative risk assessment, the output of the EPS should be directly coupled to user application models. The output of such a system will be a probability forecast of user specific variables (damage, demand, pay out etc). From this, the user can directly assess if the risk of an abnormal situation warrants precautionary action.

- the construction of the initial perturbations;
- the resolution of the integrating model;
- the number of ensemble members;
- the length of the integration.

The construction of initial perturbations involves singular vectors, and this takes about 10% of the total cost of the EPS. The computation of diabatic singular vectors using the Hessian of the analysis cost function as initial metric increases the cost of this computation of singular vectors by about a factor of 4. Adding tropical singular vector computations would represent a further increase in cost. However, in view of planned increases in model resolution and ensemble size, there is no reason to suppose the computational cost for the initial perturbations will change significantly as a function of total cost, and this is therefore not the dominant concern when considering computer costs.

Balancing the need to have the best possible resolution for the nonlinear forecast model against providing a sufficiently large ensemble is a difficult judgement to make. The EPS is currently run at a resolution of about half of the high-resolution single deterministic forecast. However, as a tool for quantitative risk assessment of severe weather, it is clearly desirable to be able to run ensembles with a model whose resolution is such as to be able to simulate severe weather, otherwise the EPS will systematically underpredict the probability of severe weather. Insofar as T_L319 and T_L511 show a sensitivity of resolution to the simulation of severe weather, then these resolutions should be considered as eventual targets for the EPS. An intermediate increase in EPS resolution (to $T_L255L60$) is planned in late 2000, and documention of the improvement of the T_L255 ensemble over the T_L159 ensemble is currently in progress.

On the other hand, it is also necessary to ensure adequate ensemble size. For example, given the true forecast PDF, suppose that there is a 10% risk of some severe weather event (whose climatological probability may be orders of magnitude smaller than 0.1) – of the Christmas storm discussed above. Suppose that the EPS were to sample randomly from the true forecast PDF (which itself is problematic, see above), then with a 50-member ensemble one would expect 5 members to predict the event. Consider a particular 50-member ensemble in which no member predicts the event. Using a simple chi-squared test with a 1% confidence value it would be impossible to reject the hypothesis that such an ensemble could not have been drawn randomly from the true forecast PDF. This is an indication of inadequate ensemble size. By contrast, if the event was not predicted by any member of a 100-member ensemble, the chi-square test would indicate that this ensemble was not a random drawing of the true PDF.

One possible method for effectively increasing ensemble size is to run the ensemble more frequently, e.g. twice per day, instead of once. An optimal probability forecast could then be made by combining the two individual ensembles.

Finally, there is the question of integration length. In principle, providing the ensemble is able to produce reliable probabilities, then the EPS can be extended beyond 10 days (however, in practice, implementation of such a strategy should await a better simulation of the dominant intraseasonal modes of variability in the model - there is evidence to suggest that the representation of such modes may well improve with an interactive ocean). However, increasing the forecast range to 20 days will approximately double the cost of the EPS (it would more than double the cost if the model is run with an interactive ocean).

The EPS is undoubtedly computationally expensive, and the factors discussed above all increase its computational cost significantly. A crucial question is whether such an increase can be justified. This depends on what really is the value of the EPS to the ECMWF Member States. It has been argued that if the links between EPS output and application model input were more strongly developed, then the true value of the EPS as a quantitative tool for risk management would be realised. The development of this methodology will be a key to the development of the EPS itself.

References

Barkmeijer, J., **R. Buizza**, **T. N. Palmer** & **K. Puri**, 2000: Tropical singular vectors with linearised diabatic physics. *Q. J. R. Meteorol. Soc.*, to appear.

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

Buizza, R., **T. Petroliagis**, **T.N. Palmer**, **J. Barkmeijer**, **M.Hamrud**, **A. Hollingsworth**, **A. Simmons** and **N. Wedi**, 1998: The Impact of Model Resolution and Ensemble Size on the Performance of an Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **124**, 1935-1960.

Buizza, R., **M.J. Miller** & **T.N. Palmer**, 1999: Stochastic simulation of model uncertainties in the ECMWF Ensemble Prediction System. *Q.J.R. Meteorol. Soc.*, **125**, 2887-2908.

Hoffschildt, M., **J.-R. Bidlot**, **B. Hansen** & **P.A.E.M. Jannssen**, 2000: Potential benefit of ensemble forecasts for ship routing. *ECMWF Technical Memorandum No. 287.*

Janssen, P.A.E.M., 2000: Potential benefits of ensemble prediction of waves. *ECMWF Newsletter 86*

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

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

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

Puri, K., **J. Barkmeijer** & **T.N. Palmer**, 2000: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Q. J. R. Meteorol. Soc.*, to appear.

Palmer, T.N., 2000: The prediction of uncertainty in weather and climate forecasting. *Rep. Prog. Phys.*, **63**, 71-116.

Richardson, D., 1998: Obtaining economic value from the EPS. *ECMWF Newsletter No. 80.*

Richardson, D., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **126**, 649-667.

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

T.N. Palmer, J. Barkmeijer, R. Buizza, E. Klinker and D. Richardson

---

A revised land-surface analysis scheme in the Integrated Forecasting System

Two significant changes on the land-surface analysis were implemented in the operational ECMWF 4D-Var system during 1999. In March (Cy19r2), an analysis of two-metre temperature and relative humidity based on univariate optimum interpolation using SYNOP observations was put into the operational system. In July (Cy21r2), the previous operational soil moisture analysis based on a nudging scheme was replaced by an analysis of soil moisture and soil temperature based on optimum interpolation. These modifications have had a positive impact on the quality of ECMWF near-surface products and are also expected to improve the quality of products generated by the 40-year reanalysis project (ERA-40). The purpose of this paper is to describe and illustrate this revised land surface analysis scheme.

In August 1993, the relaxation of both the temperature and moisture content at the bottom of the soil layer to their climatological values was replaced by a zero flux condition for heat transfers and a free-drainage condition for water transfers. By summer 1994, the land surface had drifted to an excessively dry state over most of the Northern Hemisphere continents, with a detrimental impact on forecast scores. With no relaxation to climatology and no initialisation of the soil prognostic variables, nothing prevented the land-surface scheme from drifting to an unrealistic state. Such drifts are the consequence of a positive feedback with the atmosphere when the land-surface scheme experiences systematic errors in the atmospheric forcing, or when some physical processes are not properly described. Two conclusions were drawn: soil prognostic variables need to be initialised and short-range forecast errors of near-surface parameters contain information on the state of the soil.

A simple soil analysis scheme (a nudging scheme) was implemented in December 1994. Soil-moisture increments $\Delta\Theta$ at each analysis cycle were assumed to be proportional to increments of specific humidity Δq at the lowest model level produced by the atmospheric analysis:

$$\Delta\Theta = C_{veg} D \Delta q \ ,$$

where D is a global constant coefficient and C_{veg} is the vegetation fraction (introduced in order to reduce correction over deserts). This nudging scheme has proved to be robust enough to control soil-moisture drifts in the root zone during summer. However, recent comparisons of the