352

Severe Weather Prediction

T.N. Palmer, J. Barkmeijer, R. Buizza,
C. Jakob, F. Lalaurette, T. Paccagnella[1]
and D. Richardson

Research Department

[1] ARPA-SMR, Bologna, Italy

November 2001

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading, Berks RG2 9AX

library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/pressroom/publications.html

1 Introduction and Objectives

The ECMWF 4-year plan calls for the provision of good forecasts of severe weather at day 4 or 5. This paper discusses some of the conceptual issues that underlie the ability of ECMWF to provide such forecasts (section 2) , and some of the ongoing research projects at ECMWF with a focus on severe weather prediction on these timescales (section 3). These research projects are designed to enhance the capability of the ECMWF Member State National Meteorological Services (NMSs) to issue severe weather predictions for the protection of life and property.

2 Medium-range and short-range forecasting of severe storms

In dealing with a risk of a national or regional emergency threatening life or property, a weather-forecast office will use a range of forecast tools to assess the threat and provide the necessary warnings. Suppose the threat, eg of intense rain for Saturday is first perceived in Monday's forecast. A forecaster will use EPS forecasts (50 members, 80km resolution) and deterministic global forecasts (40km resolution) in the early part of the week, and short-range regional forecasts (10km resolution) and now-casting (radar) in the later part of the week at the stage when the threat is imminent. In this section we discuss the capability of the EPS and deterministic model to make useful medium-range forecasts of the synoptic conditions for severe weather (section 2.1). We then discuss the capability of the EPS and the deterministic model to make quantitative medium-range forecasts of the synoptic conditions for severe weather (section 2.2). Severe and extreme weather are contrasted in section 2.3. Section 2.4 discusses the medium-range predictability of severe weather, while section 2.5 discusses the economic value of probabilistic forecasts of severe weather.

2.1 Can ECMWF's global NWP model and EPS system give useful medium-range forecasts of the synoptic conditions for severe weather?

Severe weather is not necessarily associated with small-scale rapidly- developing vortices. A heat wave may cover a domain many times larger than the truncation scale of the model. There is no fundamental reason why NWP models should not be able to simulate such events. (Such events may also be associated with elevated levels of air pollution, a cause of enhanced mortality. Coupled to suitable tropospheric chemistry models, NWP models should be able to forecast this type of pollution.)

Similarly, coupled to suitable hydrology models, NWP models should be able to simulate river flooding occurring from widespread rain over the river's drainage basin, if the latter spans sufficient grid points of the model. An example of EPS predictions of the rainfall integrated over the Po-river basin during a period of severe flooding is shown in Fig 1 (from Hollingsworth and Viterbo, 2001).

However, there is clearly an important class of severe weather whose space and time scales are relatively small scale. For example, three such events caused great damage in Europe in December 1999 (the Danish storm of 3 and 4 December, and the two French/German storms of 26 and 28 December). These have been studied in some detail by Buizza and Hollingsworth (2001); see also Fig 4 below. It is clear from this and other studies, that, notwithstanding fundamental predictability problems discussed below, the ECMWF model (at TL255 and higher resolution) is capable of simulating the broad synoptic-scale surface pressure fields related to such events in the medium range.

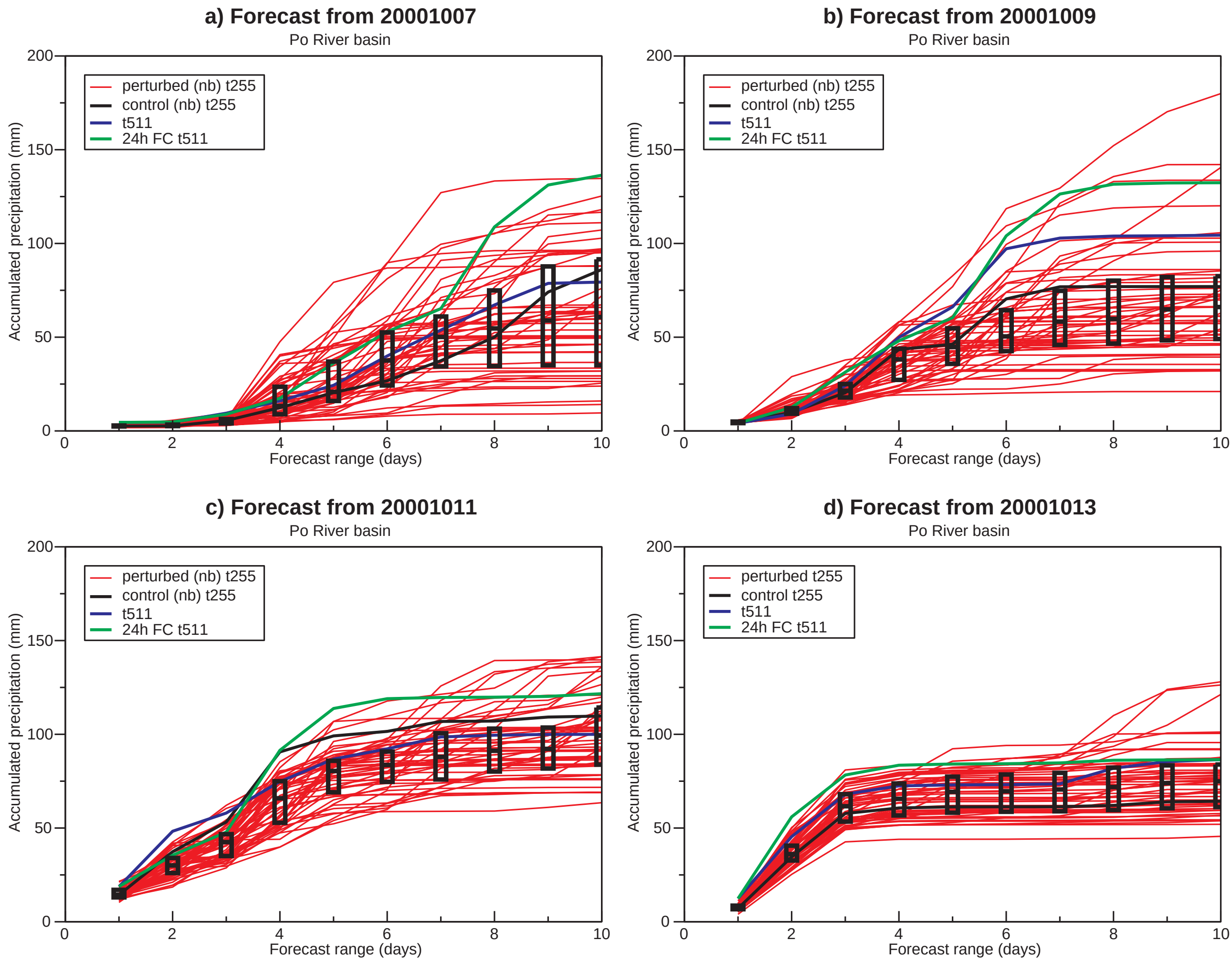


Fig 1 Averaged precipitation over the Po River Basin as a function of forecast time (in days). Blue, black and red curves represent the deterministic forecast at T511, the control forecast at T255 and the 50 perturbed forecasts at T255, respectively; all values curves are accumulated since the beginning of the forecast. Black boxes represent the mean and interquartile values for the ensemble results. The green curve represents the time accumulated succession of 24 hour forecasts for the t511 system and is taken as proxy for truth. Top left: forecast from 20001007; Top right: Forecast from 20001008; Bottom left: Forecast from 20001011; Bottom right: Forecast from 20001013 (Hollingsworth and Viterbo, 2001).

As an example from a more recent period, Fig 2 shows results from a case study of the so-called Halloween storm which passed over the UK on 30 October 2000 (associated with severe flooding). Neither the T319 deterministic forecast, nor the T159 control successfully forecast the low-pressure system. The operational EPS run during this period had erroneously weak initial perturbations. As a result none of the two-day perturbed EPS members satisfactorily forecast the storm (Fig 2a). However, when the EPS was re-run with corrected perturbations (Fig 2b) it indicated there was a risk of a storm, though none of the members was sufficiently intense. Increasing resolution to TL255 (Fig 2c) produced members which were both well positioned and sufficiently intense. Re-running the EPS at TL319 resolution (Fig 2d) did not improve further on the overall performance of the ensemble. This example illustrates the importance of balancing model resolution with ensemble size, an issue discussed in more detail below.

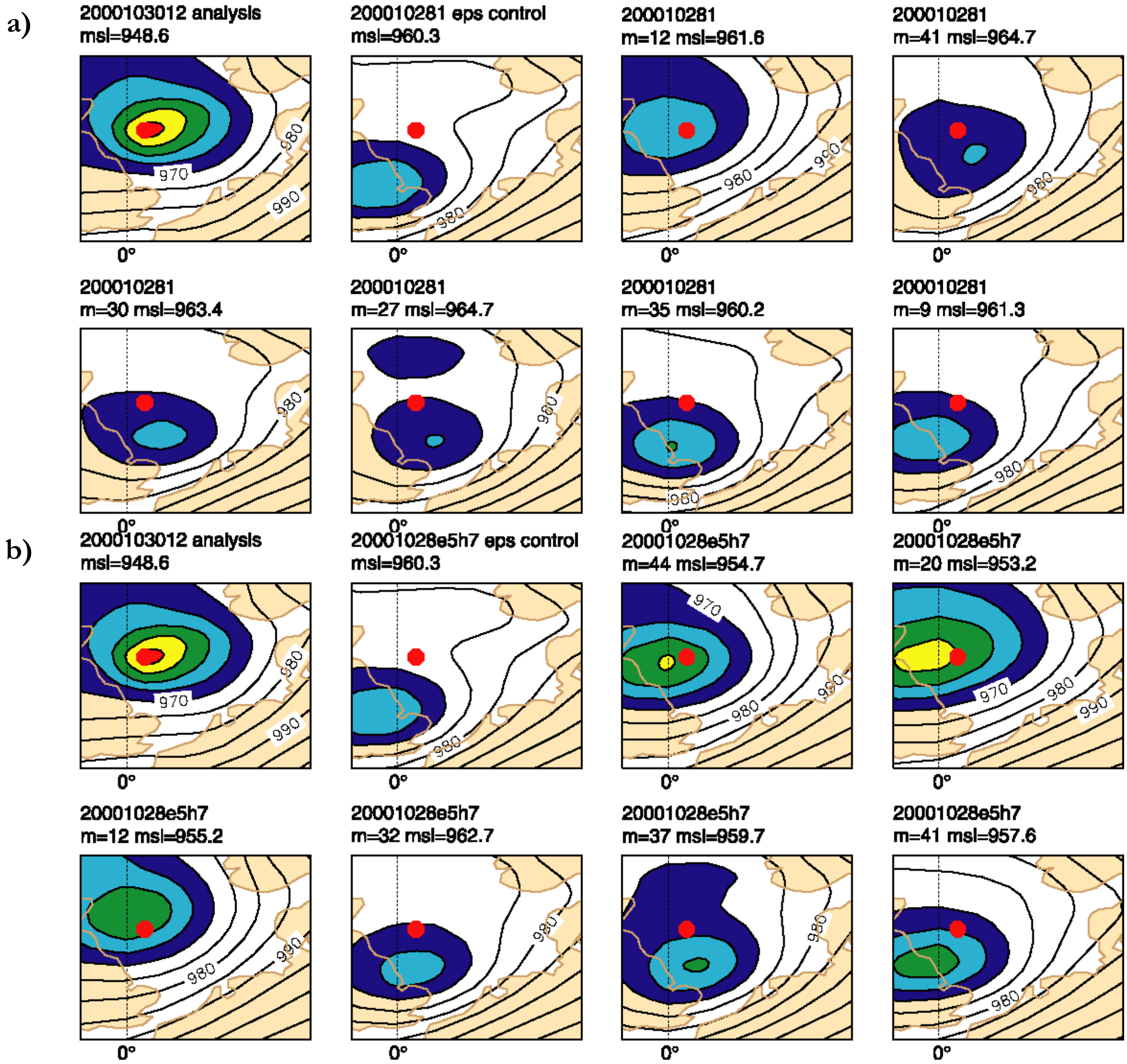


Fig 2a) Mean sea level pressure for the verifying analysis of 30 October 2000, 12UTC (top left panel), the 48 hour operational TL159 EPS control forecast from 28 October 2000 (top middle left panel) and 6 members of the EPS with the storm closest in position compared with the verifying analysis (as depicted with a red dot). The MSL value of the storm and distance to the analysed location are given above each panel. The contour interval is 5hPa. Fig b) as a) but with corrected initial perturbations

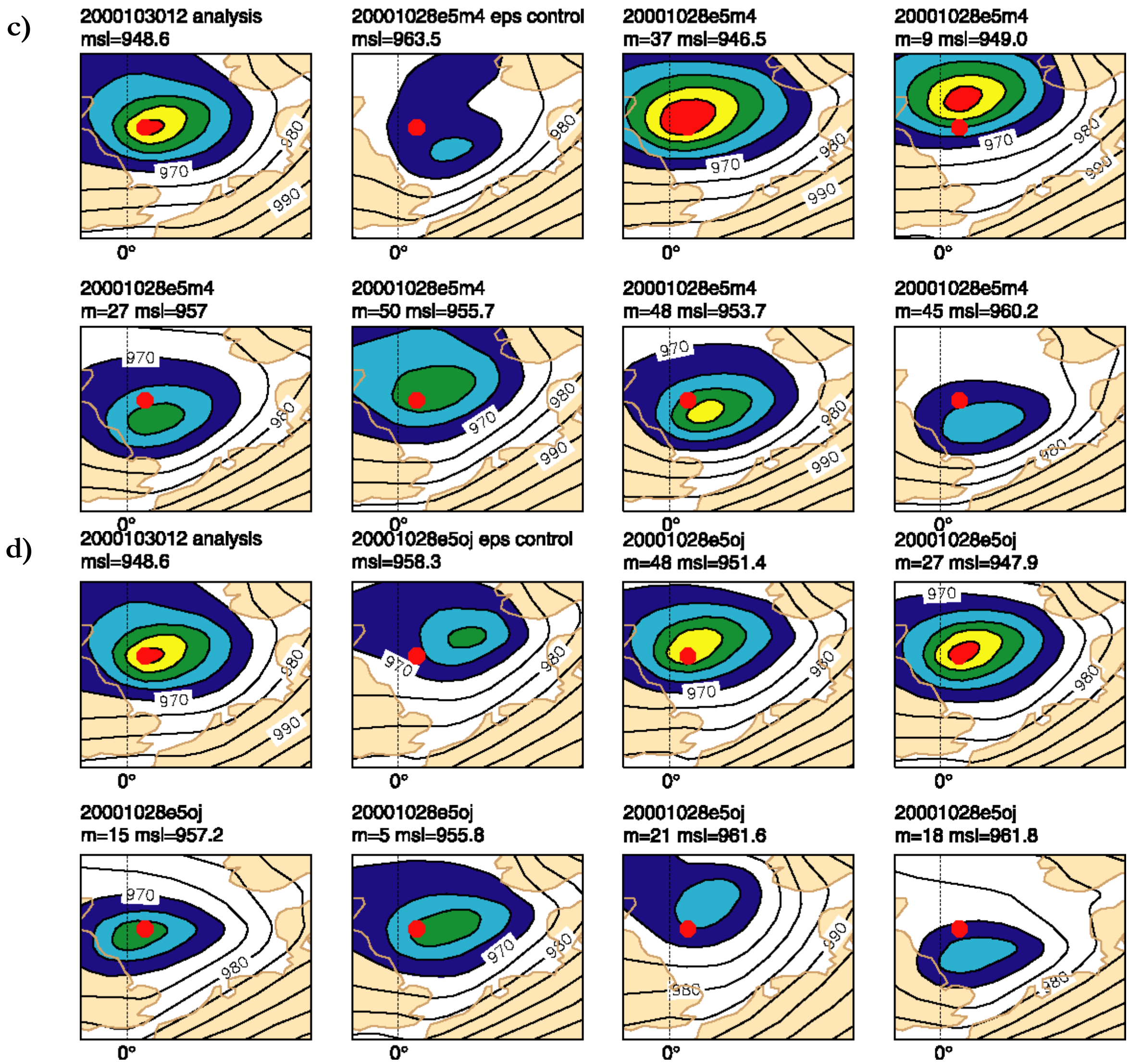


Fig 2c) as 2b) but for TL255. d) as c) but for TL319.

2.2 Can ECMWF's global NWP model and EPS system give quantitative medium-range forecasts of severe weather, such as wind or rain?

The answer to this question depends very much on the scale (in both space and time) of the phenomenon to be forecast. As mentioned above, there are certain large-scale severe events which can certainly be forecast quantitatively by the ECMWF forecast system. On the other hand, one cannot expect an NWP model to be able to simulate well scales close to the truncation scale (Lander and Hoskins, 1996). In this respect one should distinguish between the truncation scale of a model, and the smallest physical feature it can accurately resolve (e.g. Grasso 2000; Durran, 2000). For example, it appears that a mesoscale convective complex with a scale of 10km cannot be well simulated with a grid spacing of more than about a kilometre (e.g. Miller 1978). Davies and Brown (2001) find a broadly similar result when studying flow over orography. They find that two grid points per orographic feature cannot give any appreciable skill in simulating the flow over this feature. With four and six grid points per hill, qualitatively reasonable results are found, but drag is still underestimated (see also Cullen et al.2000 for a similar discussion on resolution and flow over orography).

This is particularly pertinent when studying the ability of NWP models to simulate severe rainfall. Fig 3 shows maximum rainfall as a function of averaging area from a study of a network of 161 rain gauges over Southern Sweden over an 11 year period (Olsson et al 1999). For averaging areas greater than about ~2800km^2, there is little impact of averaging area. However, for averaging areas less than ~2800km^2 there is a large impact of averaging area. This critical area may correspond to the scale which separates large and small mesoscale precipitation processes (the latter associated with many types of severe weather). In view of the remarks above distinguishing truncation scale from resolution, even the TL511 model is unlikely to be able to resolve features with an area of ~2800km^2. Hence it would seem likely that the statistics of extreme rainfall in the TL511 model are not characteristic of extreme rainfall on spatial averages of ~2800km^2 or less. Hence some sort of downscaling scheme will be needed to map ECMWF model fields to severe-weather events.

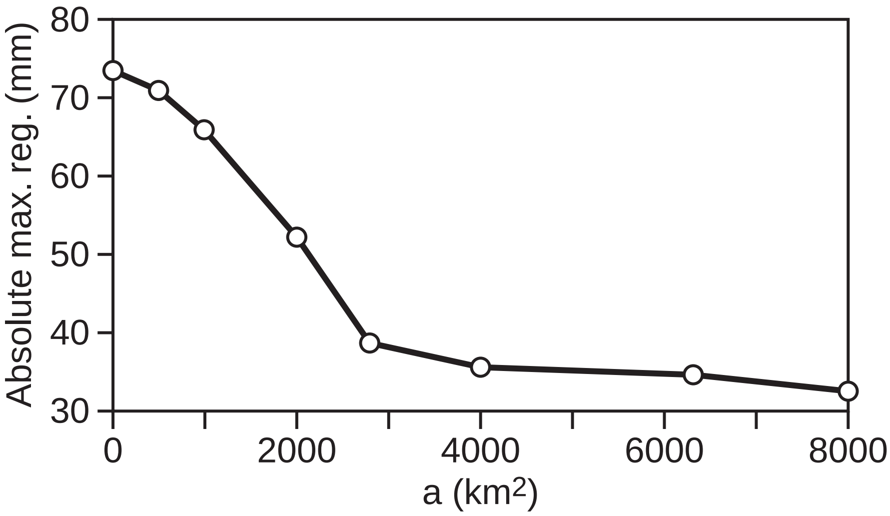


Fig 3 Extreme rainfall as a function of spatial averaging, obtained from a network of 161 raingauges in southern Sweden (Olsson et al, 1999). Mean rainfall did not show the change in qualitative characteristic at scales of ~2800km^2

The bench forecaster currently relies on a short-range high resolution model to provide detailed estimates of the severity of the event. In section 3 we discuss investigations which aim to provide statistical postprocessing of the EPS which will provide useful medium-range estimates of severity.

In discussing the ability of the ECMWF model to simulate wind events on scales of a hundred kilometres or so, it is worth noting that there is now considerable evidence that the atmospheric spectrum of kinetic energy is significantly shallower than a simple two- dimensional enstrophy-cascading spectrum (e.g. Nastrom and Gage, 1985; see also Cho et al, 1999), even in the free atmosphere. By contrast, the ECMWF model spectrum of kinetic energy is never steeper than k^{-3} (M. Hortal, personal communication). This implies that the model will tend to underestimate the intensity of wind events on the small synoptic scales.

One interpretation of the diagnosed $k^{-5/3}$ spectrum (e.g. Lilly, 1983) is that energy associated with forced mesoscale systems (e.g. forced by convective instability, topography etc) is being transferred to (and interacts with) larger scales. As a global NWP model cannot simulate mesoscale dynamical systems, it cannot represent this upward energy transfer. Buizza et al (1999) have suggested that uncertainties in sub-grid variability should be represented stochastically. This stochastic forcing can have a substantial effect on the intensity of isolated vortices (Puri et al, 2001) and even on the model's large scale systematic error (Palmer, 2001).

2.3 Can ECMWF's global NWP model and EPS system give quantitative medium-range forecasts of extreme weather?

Severe weather can be defined as weather which threatens life or property. As mentioned, severe weather can be associated with prolonged synoptic scale events, though more usually is associated with rapidly developing small-scale features in the atmosphere.

On the other hand, extreme weather can be defined from climatological probability distributions. If a climatological distribution of a variable that is well resolved by the model is known, then any forecast of that variable can be assessed as being extreme, irrespective of whether it is severe or not. In many circumstances the linkage between an extreme event and a severe event will be probabilistic (i.e. 1-many). One of the basic forecast products described in section 3 concerns the ability to recognise extreme events produced in the EPS.

2.4 Medium-range predictability of severe weather

Severe weather is often the result of rapid atmospheric development. Notwithstanding the mechanisms of such development, the predictability of severe weather is likely to be lower than normal (though again one needs to note that the development of the `heat-wave' type of severe weather need not be unpredictable in this sense). This sense of relative unpredictability is clearly manifest in the integrated hyetograph of Po basin rainfall shown in Fig 1, where the analysed accumulations of rainfall were captured by only a fraction of EPS members. Similar conclusions can be made for the Halloween storm shown in Fig 2, and the Lothar storm of December 1999 (see Fig 4).

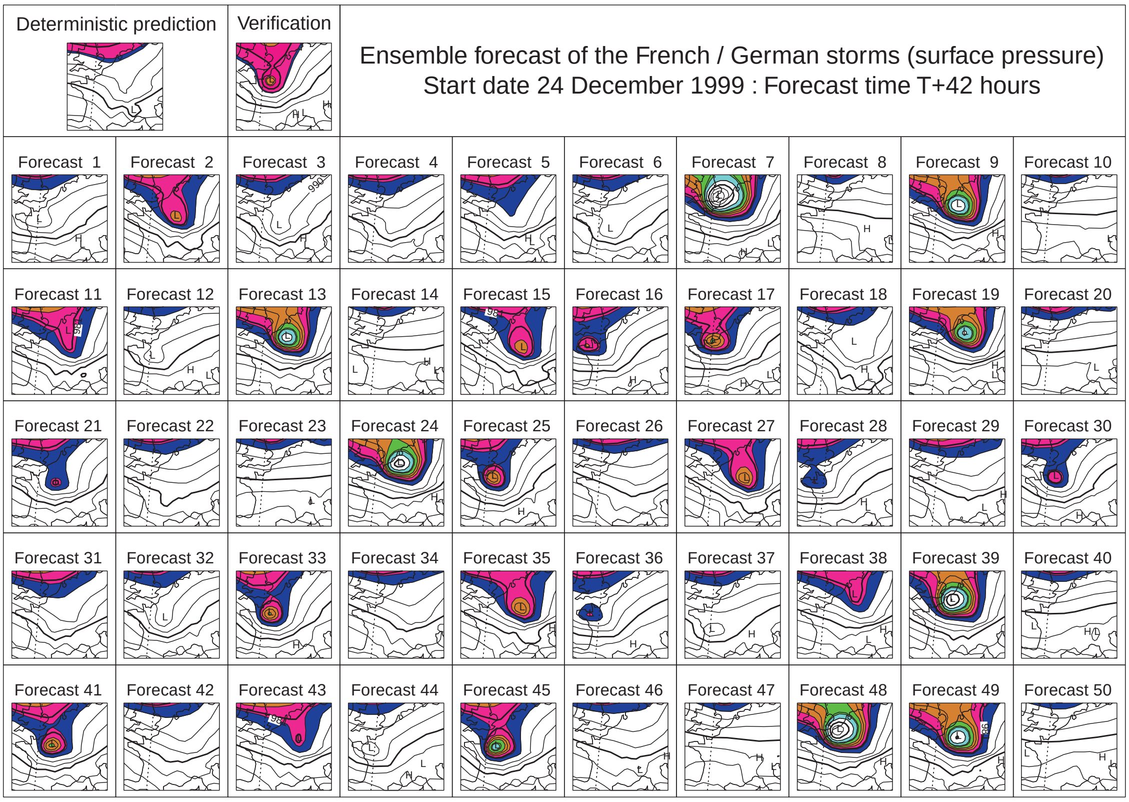


Fig 4. Individual forecasts from a 42 hour ensemble for the storm that devastated parts of mainland Europe on December 26th 1999, indicating its limited predictability. In this case, the high resolution deterministic prediction did not forecast the risk of a severe storm, whilst the EPS indicated that there was a such a risk, vastly exceeding its climatological probability.

The relatively low predictability of such events is not only associated with sensitivity to initial perturbations, but is also due to uncertainties in the model equations. For example, as shown by Puri et al (2001), the intensity of isolated vortices is sensitive to the realisation of the stochastic physics scheme in the ECMWF model (see also Annex 3).

In fact, the predictability of severe weather on scales of 100km or less, may be further lessened by the observed $k^{-5/3}$ spectrum. It is well known (e.g. Lorenz 1969) that the predictability of such spectra are determined by the eddy turn-over time of the scale of interest, and (unlike in k^{-3} turbulence) thus predictability cannot be increased by increasing the accuracy of the initial data on a limited number of scales.

In summary whilst large-scale severe events are predictable, smaller scale rapidly developing systems are likely to have relatively low predictability and can only be predicted using reliable EPS-based systems.

2.5 Quantitative economic use of the EPS

In considering the predictability of severe weather, it is important to be able to assess the utility of probabilistic predictions of severe weather, where the probability of occurrence is significantly less than unity. Consider a reliable probabilistic forecast system for severe weather, and suppose, over a period of years, 10 forecasts of severe weather are made (say for D+5) each with a probability of 20%. If, on each occasion that severe weather occurs, a user can avoid damage valued at 1 million Euros by taking some form of protective action at a cost of 100,000 Euros, then overall the user can expect to save 1 million Euros by taking action on each occasion severe weather was forecast with 20 % probability. On the other hand, in 8 out of 10 situations, the severe weather event would not occur - a significant ‘false alarm’ rate.

When precautionary action is made on a purely economic basis, and when the user is aware of his/her cost/loss ratio (here 0.1), then an 80% false-alarm rate for extreme weather prediction might be considered acceptable. For humanitarian applications (e.g. emergency services), it is clearly much more difficult to quantify this cost/loss ratio and hence to assess the utility of EPS forecasts of extreme weather. However, as this analysis shows, the question of what constitutes ‘useful’ predictability cannot be addressed in a purely meteorological context. An assessment of the characteristics and requirements of end users would be helpful in the development of an extreme-weather forecast capability.

3 Research Developments Relevant To Severe Weather Prediction

For severe weather in the medium range, NMSs have traditionally relied on single high-resolution deterministic forecasts from ECMWF, supplemented with a limited number of forecasts from other operational centres (the so-called `poor man’s’ ensemble). As raised by the discussion in the previous section, this procedure has limitations. Firstly, by its nature, severe weather associated with relatively small-scale rapidly developing disturbances is likely to have limited predictability. Secondly, notwithstanding recent increases in model resolution, prediction of genuinely property-damaging or life-threatening winds or precipitation rates are still not generally obtainable from direct-model output. Based on these considerations, the research projects described in this section are designed to enhance the capability of NMSs to issue severe weather predictions for protection of life and property.

3.1 Product Development

Although, traditionally, forecast products have been developed by the NMSs, since the development of ensemble prediction techniques in particular, ECMWF has taken a more active role in the development of probabilistic forecast products. An example is the Extreme Weather Forecast Index (EFI). One can measure of the ‘extremity’ of a deterministic forecast of some weather variable in terms of a suitably-defined distance from the climatological mean (for a given location and time of year). Similarly one can measure the extremity of an ensemble forecast in terms of a suitably-defined distance between the EPS probability distribution and an appropriate climatological mean distribution. The EFI, based on 2m temperature, 10m wind speed, and precipitation is such a measure, its numerical value lies between –1 and +1. When the EFI is close to +1, then not only is the EPS distribution substantially different from climatology, it differs in the sense of wind speed being anomalously large, precipitation anomalously heavy, or temperature anomalously warm. Maps of EFI can be used by forecasters as an initial alert (‘wake up call’) to the possibility of severe weather. A precise definition of the EFI, with some examples and preliminary assessments, is given in Annex 1.

The climatological distributions needed in the definition of the EFI have so far been defined from archived EPS integrations. However, each time the forecast system is upgraded (e.g. in terms of enhanced model resolution or revised physical parametrisation), these climatological distributions are no longer fully consistent with the operational EPS. This raises the issue of whether a set of hindcast integrations should be made when substantial changes occur to the forecast system. This need not be excessively computationally expensive. For example, 20 years of single deterministic forecasts (e.g. run from reanalysis initial conditions) are equivalent to about two months of 100-member ensembles.

Currently, probability forecast products are based on specific anomaly thresholds for wind, temperature and precipitation. It is planned to supplement these (on the ECMWF web site) with maps showing the wind, temperature and precipitation associated with specific parts of the forecast PDF, e.g. the lower boundary of the upper decile of the EPS. It is also planned to use the EFI climatology to issue probability forecasts forecasting the occurrence e.g. with the upper decile of the climatological distribution. In this way, EPS output will include specific probabilistic products that relate to extreme values.

Action: Further develop specific extreme forecast products from the EPS. Produce a strategy for the production of climatological hindcasts with cost implications.

3.2 Twice-a-day EPS vs 100-member EPS

Currently operational ECMWF forecasts are made once a day. For extreme weather prediction, particularly in the early medium range, then more timely probabilistic forecasts could be made if the operational suite was run twice a day. Real-time forecasts are now being made from 0z and a preliminary evaluation of the impact these additional forecasts have on extreme weather prediction is given in Annex 2.

In order to assess this impact, a set of 100 member ensembles is also being run from 12UTC. The initial conditions for these 100-member ensembles are constructed from a 50-dimensional singular vector space. Hence, the initial perturbations for these 100-member ensembles span directions which are not explicitly sampled by 50-member ensembles. We can ask which validates better: a 100-member ensemble from 12UTC, or a combination of two 50-member ensembles from 12UTC and 00UTC (50+50). The results in Annex 2 go some way to answering this question, though they are based on limited forecast data. Firstly, for the extreme forecast events considered (based on height exceeding two standard deviations below normal), it is definitely advantageous (in terms of Brier Skill Score, ROC area and PEV) to increase ensemble size, either in the 100 configuration or the 50+50 configuration. Whether 100 members is preferable to 50+50 (or vice versa) depends on whether the 100 member ensemble is from the most recent analysis time or not (i.e. whether the 00UTC analysis precedes or succeeds the 12UTC analysis.

In the studies reported in Annex 2, the 100 members of the 50+50 ensemble have been given equal weight in the estimation of forecast probability . On the other hand, forecasts from the older analysis time are likely to be less skilful than forecasts from the more recent analysis time. Statistical estimation of optimal weighting between the 50 member ensembles from earlier and later analysis times will be based on the sample of available forecasts. This optimal estimation can be determined by Bayesian analysis, where the EPS from the earlier analysis time determines the prior probability estimate. The natural extension of such a procedure is that the most recent posterior estimate becomes the prior for the next estimate, analogous to the data assimilation procedure.

Action: Run a more complete set of 100-member and 50+50 EPS forecasts over the coming months, and give a further report to the SAC next year. Discuss with forecasters the impact of 50+50 vs 100 members in the light of NMS forecast office operational scheduling procedures.

3.3 Tropical Cylone EPS

Tropical cyclones provide clear examples of life-threatening weather systems. Prediction of such systems is not only of direct interest to Member States forecast offices, but also because of the interests of European commerce (e.g. in the reinsurance industry). Initial perturbations for the operational EPS are targeted to maximise evolved perturbation energy in the extratropics. Hence ensemble forecasts of tropical cyclone tracks are frequently underdispersive (ie do not include the analysed track). Over the last couple of years, there has been a considerable amount of work on estimation of singular vectors in the tropics (Barkmeijer et al, 2001) and their application to ensemble forecasts of tropical cyclones (Puri et al, 2001). In these studies, the complementary roles of the targeted tropical singular vectors and stochastic physics were noted – the former contributing most to the dispersion in ensemble tracks, the latter contributing most to the dispersion in cyclone intensity.

Results from a set of recent case studies are documented in Annex 3. Although the number of case studies is still limited, it is suggested that there is sufficient evidence that an operational implementation of targeted tropical singular vectors during periods of tropical cyclone activity would be beneficial.

Action: Validate, implement and verify the ensemble tropical cyclone scheme.

3.4 Quantitative validation of severe and extreme events

Whilst it is important to be able to assess the performance of the forecast system on individual cases, a more general methodology for validating forecasts of severe and extreme events is needed. By virtue of the fact that extreme events can be defined from statistical distributions, it is much easier to base validation of NWP forecasts on extreme events.

Potential economic value (PEV; Richardson, 2000) is a good measure of the performance of an NWP model in predicting extreme events for a number of reasons. Firstly, by targeting on a specific dichotomous event, PEV can by construction be used to focus specifically on extreme situations. Secondly, PEV is closely related to two standard probabilistic skill scores: the Brier skill score, and the area under the relative operating characteristic (ROC). Essentially, the integral of PEV with respect to a set of users uniformly distributed in cost/loss ratio gives the Brier skill score; the maximum PEV is related to the ROC area.

Finally, consistent with the discussion above, PEV describes the utility of the ensemble forecast as a tool for risk assessment for a range of potential users.

As an illustration of its utility, Fig 5 shows PEV for D+6 predictions of an extreme event: 500hPa geopotential height less than 2σ below normal. The left hand panel shows the impact of increasing ensemble resolution (from TL159 to TL255), the right hand panel shows the impact of increasing ensemble size (from 50 to 100) members (based on a set of 30 wintertime dates which included the 1999 Danish ‘storm of the century’). In this case, the mean and standard deviation are calculated from the 30-case sample, and all events around the northern hemisphere are included. As discussed below, there is a need to have a reliable indendent estimates of climatological distributions evaluated from gridded analyses and historical forecast data.

It can be seen from these results that there is no unique answer to the question: is it better for prediction of extreme weather to increase model resolution, or to increase ensemble size (i.e. to resolve better the forecast probability distribution)? It can be seen that low cost/loss users benefit more from an increase in ensemble size, whilst a set of users uniformly distributed in cost/loss ratio would, on average, benefit more from an increase in model resolution. Whilst it may not be possible in general to describe user-decision processes in terms of the simple cost/loss model, the analysis clearly suggests that the design of severe or extreme event forecast systems cannot be made without knowledge of the uses and characteristics of users of such systems.

To further emphasise this point, based on the relationship between PEV and Brier skill score Fig 5 indicates that increasing ensemble size from 50 to 100 members has no impact whatsoever on Brier skill score. On the other hand, we could consider a generalisation of Brier skill score, defined as the integral of PEV with respect to a realistic distribution of users. If such a realistic distribution was dominated by cost/loss ratios much less than one, then this generalised Brier skill score would indeed show sensitivity to an increase in ensemble size from 50 to 100 (Richardson, 2001).

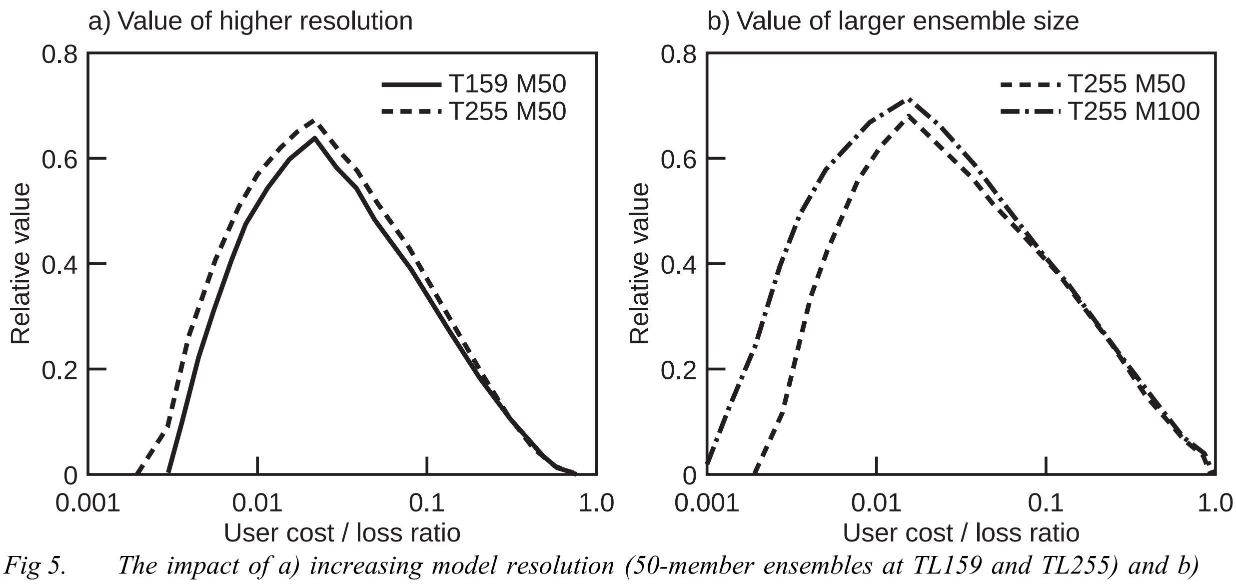


Fig 5. The impact of a) increasing model resolution (50-member ensembles at TL159 and TL255) and b) increasing ensemble size (50-member to 100-member ensembles at TL255 resolution) on the potential economic value of ensemble forecasts for the extreme event: 500hPa geopotential height anomaly less than minus two standard deviations below normal. Based on approximately 30 different initial dates in November/December 1999.

This type of EPS validation can also be applied to rainfall, though, as above, to be quantitatively meaningful, the validation should be done on gridbox-average rainfall, rather than on station data (which, as discussed below, requires some probabilistic downscaling methodology). In a recent study, Mullen and Buizza (2001) have used gridded station precipitation values over the US to validate the EPS. For relatively extreme rainfall events (50mm/day), their analysis also demonstrated the importance of ensemble size.

Action: Generate a European climatological database for rainfall verification from the Member States detailed rain networks. Further develop validation schemes for the EFI and for other forecast products of extreme weather.

3.5 Linking extreme and severe weather

This is essentially the problem of downscaling. For strongly orographically forced problems, the link between extreme weather simulated on the scale of a global NWP model, and severe weather on the mesoscale may be approximately 1-to-1. For such applications, nesting a LAM model inside an NWP model may provide a satisfactory linkage between extreme and severe weather (see section 3.7).

However, as discussed, the linkage is more generally likely to be 1-to-many. This is particularly relevant in the case of precipitation. One possibility is to generalise the model's own parametrisation schemes to include the estimation of a parametrised probability distribution of rain at a point within the gridbox. This is described in Annex 4.

Otherwise, the PDF could be obtained from data analysis external to the model. Suppose the climatological PDF for a chosen variable x at some specific site is $\rho(x)$, calculated e.g. over the ERA-40 period (for a given time of year). Let the PDF of the ERA-40 variables X for the gridbox containing the site in question be $\rho(X)$. The downscaling problem can be phrased in terms of calculating the conditional PDF $\rho(x|X)$ (in theory we might also want to consider adjacent gridboxes in computing these conditional PDFs). An example is illustrated in Fig 6 which gives an empirical cumulative probability distribution of 24-hour accumulated precipitation of SYNOP observations for categories of 24hr short-range accumulated precipitation from the T511 model. The figure shows that even when the deterministic forecast of rainfall is small, there is still a probability of substantial rainfall at a SYNOP station.

A D-day ensemble forecast produces a finite estimate of the conditional PDF, $\rho(X|X_0)$ where X_0 denotes the analysed value of X, D days earlier. By combining $\rho(x|X)$ with $\rho(X|X_0)$ (one based on empirical relationships, the other on numerical integrations of the equations of motion) we arrive at the PDF $\rho(x|X_0)$

Action: Further develop the parametrisation of subgrid PDFs. Develop probabilistic downscaling schemes in collaboration with Member State NMSs.

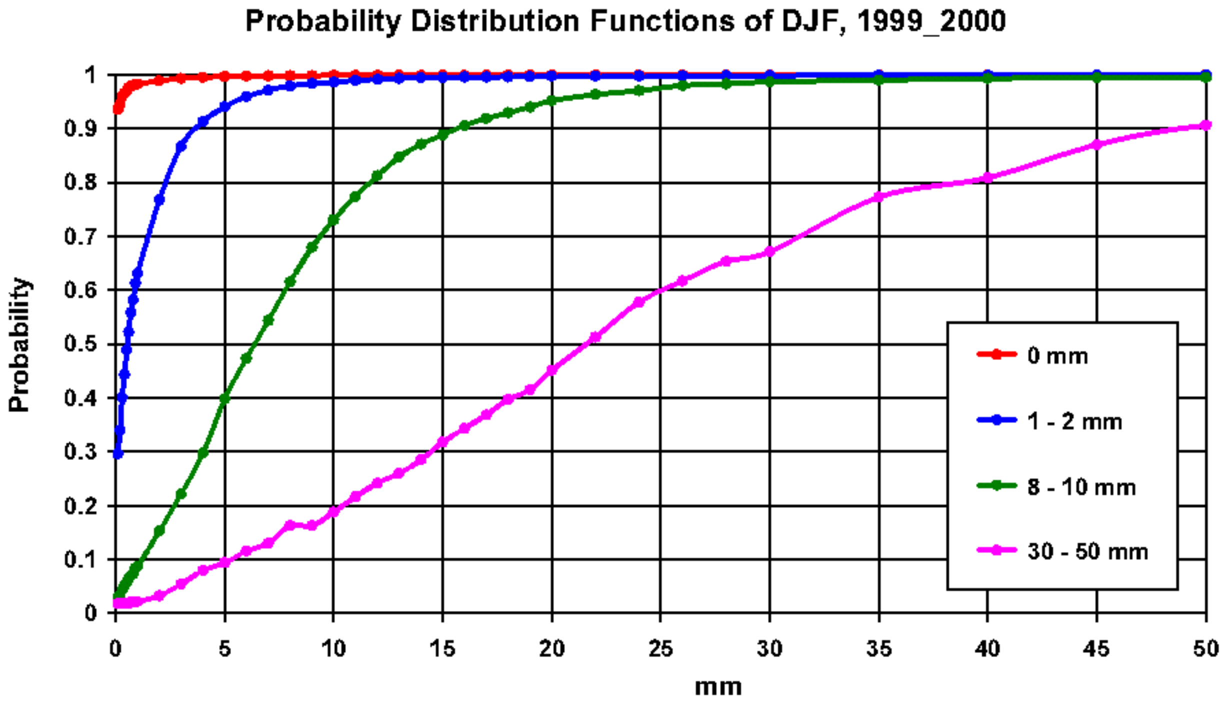


Fig 6 Empirical cumulative distribution functions based on SYNOP observations for categories of model precipitation (mm/day) (T511 forecast 18-42hr range, DJF 2000-2001).

3.6 Other EPS Developments

A number of developments of the EPS system relate to the problem of extreme weather prediction. For example, as illustrated in Fig 2, the specification of the initial EPS perturbations is crucial for reliable probabilistic forecasts of extreme weather. (In Fig 2a, it was shown that with incorrectly specified initial perturbations, no member of the ensemble for the Halloween storm correctly predicted the correct development. With corrected perturbations, Fig 2b, the position of the storm was correctly captured - though not, at TL159 resolution, the intensity of the storm). Research into the specification of singular-vector perturbations continues (including increases in singular vector resolution, moist processes, moisture as an component of the state vector, and using the static or RRKF analysis error covariance metric). In addition, EPS integrations are now supplemented with integrations from analyses made at other operational centres. Finally, tests have been made using, as ensemble initial conditions, ensembles of analyses, based on perturbed observations. To date, the resulting forecast ensembles have been significantly underdispersive. However, further comparison tests are planned, and such perturbations remain an option for future EPS implementation.

In planning the development of the EPS, we return to the question of ensemble size versus model resolution. We have shown examples of the positive impact both of increasing ensemble size and increasing model resolution. Of course, a doubling of ensemble size is much less costly than a doubling of horizontal resolution. Conversely, in order to double resolution, ensemble size would have to be reduced by about a factor of 4. However, studies of PEV suggest that a substantial reduction of ensemble size would be unacceptabe for extreme-events prediction.

However, since small scales lose predictability faster than large scales, the impact of high resolution (e.g. on medium-range forecast skill) is likely to be most important in the first part of the forecast. In order to test this, a variable resolution forecast system has been developed, where the model resolution is changed within the forecast period. Results from deterministic forecast integrations described in Annex 5, appear to confirm that most of the benefit (in terms of conventional 500hPa height scores) of the TL511 model on medium range forecast scores, arises from its use in the first 2-3 days of integration. On the basis of these results, it is planned to asses the feasibility and viability of a variable resolution EPS system to be developed. In any case, such a technique would be useful in developing a seamless system for medium-range, monthly and seasonal forecasting.

As mentioned, the intensity of isolated vortices has been found to be sensitive to the realisation of physical tendencies in the stochastic physics scheme. At present, the stochastic physics scheme is ad hoc, and detailed budget calculations may be needed in order to develop a more soundly-based scheme. Such a scheme may help alleviate the fact that the deterministic ECMWF model is unable to simulate the observed $k^{-5/3}$ spectrum.

Action: Continue research as detailed in 4-yr plan.

3.7 External Collaboration

The research programme into prediction of extreme weather is enhanced through external collaboration. An example is given in Annex 6 (in collaboration with ARP-SMR, Bologna, Italy) where a LAM ensemble is run from a selected set of EPS integrations. An example of prediction of severe weather over Italy is shown. Similar studies (of Scandinavian severe weather) have been made at the University of Oslo and continue through the EU framework-V European Flood Forecasting System Project. Research, in collaboration with KNMI, studying the prediction of severe weather using targeted ensemble perturbations (TEPS) did not show any strong advantage over the operational EPS. Other collaborative projects which involve a component of severe weather prediction include: predictability of blocking, analysis of the effects of moist singular vectors in EPS forecasts (University of Reading), developing links between the EPS and specific application models (London School of Economics), and developing aspects of the stochastic physics scheme (Met.Office). There has been extensive collaboration with the University of Arizona to validate the EPS using gridded precipitation analyses over the USA (Mullen and Buizza, 2001). These studies have confirmed the importance of both large ensemble (100-member) size, and high (TL255) resolution, especially for prediction of relatively high rainfall events. This collaboration is continuing, focussing on specific extreme winter cyclogenesis events. In addition ECMWF has many links with the European NMSs in the validation of the operational forecasts of severe weather (cf the ECMWF Expert User Meetings held each year and a specific secondment from DWD). Finally, we have discussed above the need to develop downscaling schemes in collaboration with Member State NMSs.

Action: Maintain and develop collaborative research as appropriate.

4 Conclusions

The ECMWF strategy calls for the development of a reliable forecast system for prediction of severe weather in the early medium range. As discussed, such a system must necessarily be probabilistic, partly because such severe weather events are often associated with limited synoptic-scale predictability. The extent to which such forecasts are useful is dependent on the decision characteristics of the user. Users who would suffer great loss from such severe weather, but at relatively little cost could take precautionary action to mitigate against such weather, would benefit substantially from a probabilistic severe weather forecast system. Such users would benefit from a system which could reliably discriminate low probability events. The optimal design of a severe weather prediction system requires some knowledge of the characteristics of the users of such forecasts.

Due to spatial and temporal resolution, global numerical prediction models are not intrinsically capable of directly predicting life- or property- threatening weather associated with mesoscale weather events. A number of options were presented to overcome this difficulty: statistical downscaling, parametrising probability distributions of sub-grid variability, embedding a limited area model within selected ensemble members, and predicting extreme weather, as determined by PDFs of model output. The last option requires hindcasts of EPS integrations. Such integrations should be run off-line each time substantially new forecast-system upgrades are implemented. A fraction of computer time should be allocated to the production of such hindcasts (which, in addition, could be included in the ECMWF catalogue in order that users can have information on the expected performance of the forecast system).

The current ECMWF research strategy for the development of a system for predicting severe weather was reviewed. Development of the forecast system included possible twice- a-day prediction, the extension of the EPS into the tropics, with specific targeting of tropical cyclones, the development of a variable resolution forecast model, and the development of parametrisations which estimate probability distributions of point-wise wind and precipitation. Development of products and validation techniques were described.

The SAC may wish to comment on the proposed action items in Section 3.

References

Barkmeijer, J., Buizza, R., Palmer, T.N., Puri, K. and J.-F.Mahfouf, 2001: Tropical singular vectors computed with linearized diabatic physics. *Q.J.R.Meteorol.Soc.*, **127**, 685- 708.

Beljaars, A.C.M., 1987: The influence of sampling and filtering on measured wind gusts. *J.Atmos.Ocean. Technol.*, **4**, 613-626.

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

Buizza, R. and A. Hollingsworth, 2001: Storm Prediction over Europe using the ECMWF Ensemble Prediction System. *Meteorological Apps*, to appear.

Buizza, R., D.S. Richardson and T.N. Palmer, 2001: The new 80-km high resolution ECMWF EPS. *ECMWF Newsletter*, No.90, 2-9.

Cho, J.Y.N., Zhu Y., Newell, R.E., Anderson, B.E., Barrick, J.D., Gregory, G.L., Sachse, G.W., Carroll, M.A. and Albercook G.M., 1999: Horizontal wavenumber spectra of winds, temperature and trace gases during the Pacific Exploratory Missions:1. Climatology. *J. Geophysics. Res.*, **104**, 5697-5716.

Cullen, M., D. Salmond and P. Smolarkiewicz, 2000: Key numerical issues for the future development of the ECMWF numerical model. *ECMWF workshop proceedings on 'Developments for very high resolution global models*', 183-194.

Davies, L. A. and A.R. Brown, 2001: Assessment of which scales of orography can be credibly resolved in a numerical model. *Q.J.R.Meteorol.Soc*., **127**, 1225-1237.

Grasso, L.D., 2000: The differentiation between grid spacing and resolution and their application to numerical modeling. *Bull. Amer. Meteor. Soc*., **81**, 579-580.

Hollingsworth, A. and P. Viterbo, 2001: ECMWF forecasts for the October 2000 flood event in the southern Alps and Po basin. In Preparation.

Lander, J. and B.J.Hoskins, 1966: Believable scales and parametrizations in a spectral model. *Mon.Wea.Rev.,* **125**, 292-303.

Lilly, D.K., 1983: Stratified turbulence and the mesoscale variability of the atmosphere. *J.Atmos.Sci*., **40**, 749-761.

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

Miller M. J., 1978: The Hampstead storm: a numerical simulation of a quasi-stationary cumulonimbus system. *Q.J.R.Meteorol.Soc*., **104**, 413-427.

Mullen and Buizza, 2001: The impact of horizontal resolution and ensemble prediction size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Mon.Wea.Rev*., submitted.

Nastrom, G.D. and Gage, K.S., 1985: A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J.Atmos.Sci*., **42**, 950-960.

Olsson, J., Singh, V.P., and K. Jinno, 1999: Effect of spatial averaging on temporal statistical and scaling properties of rainfall. *J.Geophys.Res*., **104**, 19,117-19,126.

Palmer, T.N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrisation in weather and climate prediction models. *Q. J.R.Meteorol.Soc*., **127**, 279-304.

Puri, K., Barkmeijer, J. and T.N.Palmer, 1999: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Q.J.R.Meteorol.Soc*., **127**, 709-731.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF Ensemble Prediction System. *Q.R. Meteorol. Soc*., **126**, 649-668.

Richardson, D. S., 2001: Measures of skill and value of Ensemble Prediction Systems, their interrelationship and the effect of ensemble size. *Q. J. R.Meteorol. Soc*., to appear.

Annex 1: Extreme Forecast Index

1. Introduction

Both the improvement of numerical forecasts of weather parameters (precipitation, wind, temperatures, cloud cover) and the development of Ensemble Prediction Systems (EPS) have raised the expectations in terms of providing the forecast users with early warnings of extreme events. Direct comparison between observed, local weather parameters and direct model output is however quickly showing the limitations of a direct approach: event the most skilful of global NWP models do not generate wind gusts or rain rates that are a threat for the public lives or properties. Although the usual approach to overcome systematic model errors is to post-process the direct model output using past knowledge of error characteristics (Model Output Statistics, Perfect Prognostic methods, Kalman filters, etc...), such methods are usually only globally optimum: they have a tendency to perform less well precisely in extreme cases that have only a small weight in the global optimisation procedure. The proposal is made here instead to qualify extreme events not with respect to the observed, local climate, but rather to the model climate. In this way we will keep control of the relative frequency of occurrence of the warning generated by the model. The methodology that has been used to develop a representation of model climate is described in Annex 1 section 2.

Common sense seems to dictate the use of very high resolution models in order to cope with extreme weather events. It is true indeed that the higher the resolution, the wider is the range of weather phenomena that can be described explicitly, including potentially damaging ones such as squall lines, tornadoes or tropical cyclones. It is now widely recognised however that not all phenomena that can be represented with realism in high-resolution models are predictable. Therefore it is argued here that a complementary approach can be followed that could provide early warnings of some value up to three or four days in advance. Instead of basing the warnings on a single realisation of a highly chaotic process, it is argued here that a more reliable procedure in the early medium range is to make the warning decision on the basis of the full forecast probability distribution. The Extreme Forecast Index (EFI) has been proposed as a measure of the distance of between the EPS distribution of the day and the model climate distribution relevant for the same location and period of the year. It is described in Annex 1 section 3

2. An EPS pseudo-climate

WMO defines climate as a collection of statistics over three decades (30-years period). Such a period is not currently available at ECMWF: operations only started in 1979, and the 40-years reanalysis project (ERA-40, Simmons et al., 2000) has not been completed yet. For the purpose that we pursue here, it is also important that we ensure the maximum degree of consistency between the climate information and the latest forecast: this cannot be achieved by a consistent reanalysis which only occurs every 10 years or so. The EPS system has been running at TL159 resolution from 16 December 1996 up to 21 November 2000. This gives a period of three full years, which is far too short to cover the climate variability over Europe. We will try however to turn the atmospheric unpredictability into our favour for once, by using the wide range of inaccurate but realistic patterns found in the ensemble of medium range forecasts run since 1996 as information on the climate variability. The underlying assumption is that most of the climate variability over Europe is of unforced, chaotic type - which obviously excludes decadal modes of variability for example. The transient patterns however lose their predictability in less than a week, and therefore their variability is expected to be explored in our super-ensemble of 10 days forecast/ 51 members much quicker than in the real world. It is difficult however to put numbers on how much quicker the climate variability is explored by such a super ensemble. An upper bound is 510 times quicker: this could be achieved only if all scales lose all correlation from one day to the next and from one ensemble member to the other. Some scales indeed decorrelate in less than a day, whereas others stay correlated for several weeks. Rather arbitrarily, it has been decided to keep from each ensemble forecast only the initial condition, the 5-days and the 10-days forecast. They have all been kept as independent realisations, which once again is not a valid assumption - among other implications, this means that we are making the assumption that model systematic errors are developing much more slowly than the differences between ensemble members, so that initial conditions, day 5 and day 10 forecasts can be considered as part of the same probability distribution. Although this would obviously not be true on seasonal scales or in the tropics, we expect this to be a reasonable gross hypothesis in the mid latitudes for 10-days forecasts.

The *pseudo-climate* information gathered for this study was collected independently for each calendar month and each geographical location in the following way:

i) selected parameters are 2m-temperature (over land only), 10-m wind speed, accumulation of precipitation over the last 24, 120 and 240 hours;

ii) analysis and EPS forecasts (lead time 5 and 10 days, control and 50 members) valid for 12UTC any day of the month during years 1997-1999 were extracted over Europe on a 0.5°x0.5° latitude/longitude grid using the ECMWF Meteorological Archiving and Retrieval System (MARS); this is referred later as our super-ensemble: its size is about 10000 members (it varies with the number of days in the calendar month), with the exception of the 10-days accumulation of precipitation whose size is only half (5000) this size;

iii) for every geographical grid point, forecasts were ranked and percentiles were archived; the extreme per-millenium values (0.1% and 99.9%) were also archived, the result being used as empirical distribution functions;

Empirical distribution functions for the parameters retrieved over two locations with very contrasted climates (Iceland and Greece) are shown in Fig A1.1and Fig A1.2. Further evaluation is currently conducted with various reference datasets. First results indicate that, although far from a realistic view of a long-term model climate, this EPS pseudo-climate is a significant step forward compared to what could be gathered from a collection of model analyses over the same period of three years.

3. The EFI formulation

The Extreme Forecast Index is a measure of the distance between the full EPS probability distribution and the EPS climate distribution. It has been designed as an extension of the Continuous Ranked Probability Score (CRPS). In mathematical form:

$$CRPS = \int_{-\infty}^{+\infty} [p_{EPS}(x < \alpha) - p_{c\lim}(x < \alpha)]^2 \, d\alpha \qquad \text{(A1.1)}$$

where α is the forecast variable, p_{EPS} and p_{clim} the probability functions of the EPS forecast and the climate respectively.

As defined in Equation A1.1, the CRPS scales like the variable α. It can easily be rescaled as a non-dimensional index by changing the variable into the climate probability function):

$$CRPS_2 = \int_0^1 [p_{EPS}(x < \alpha(p)) - p]^2 \, dp \qquad \text{(A1.2)}$$

Where $\alpha(p)$ is the inverse probability function returning for a given probability p the value of the meteorological variable that exceeds a proportion p of the climate (percentile).

It is useful to consider the limiting case of a purely deterministic forecast for which forecast probabilities only take values 0 and 1. If p_f is the proportion of climate event exceeding the forecast value α_f, then:

$$CRPS_2 = \frac{1}{3} - p_f(1 - p_f) \qquad \text{(A1.3)}$$

which takes maximum value 1/3 for p=0 or 1 (record breaking forecasts) and minimum value 1/12 for a deterministic forecast of the median of the climate distribution ($p_f = 0.5$). Indeed the zero value can only be achieved by a probability forecast that is always equal to the climate distribution.

The interest is usually for fields exceeding rather than lying below the climate values (strong winds, large amounts of precipitation). Therefore the Extreme Forecast Index is finally defined as:

$$EFI = 3\int_0^1 [p - p_{EPS}(x < \alpha(p))]^2 \, dp \, \mathrm{sign}\left\{1, \int_0^1 [p - p_{EPS}(x < \alpha(p))] dp\right\} \qquad \text{(A1.4)}$$

which is an index lying between -1 and +1: -1 is when all forecast values are extremely small, 0 when all forecast follow exactly the climate distribution, +1 when all forecast values are extremely large. Deterministic forecasts will have EFI values ranging from -1 to -0.25, and from 0.25 to 1, the discontinuity occurring for the median value of the climate distribution. Only for T_{2m} are both extremely warm and cold events highlighted using the usual colour code (blue for cold, red for warm). For other parameters (wind and rain) only forecasts for unusually large amounts (positive EFI) are reported.

4 Examples

The EFI map for the 96h forecast valid for 26 December 1999 12UTC shows to which extent the EPS did single out the situation as an extreme departure from the climate distribution of model values for this time of the year (Fig A1.5). Although the ensemble forecast was well below the observed values (25m/s and beyond), a large proportion of ensemble members exceeded the usual range of values found in the model climate distribution (Fig A1.6)

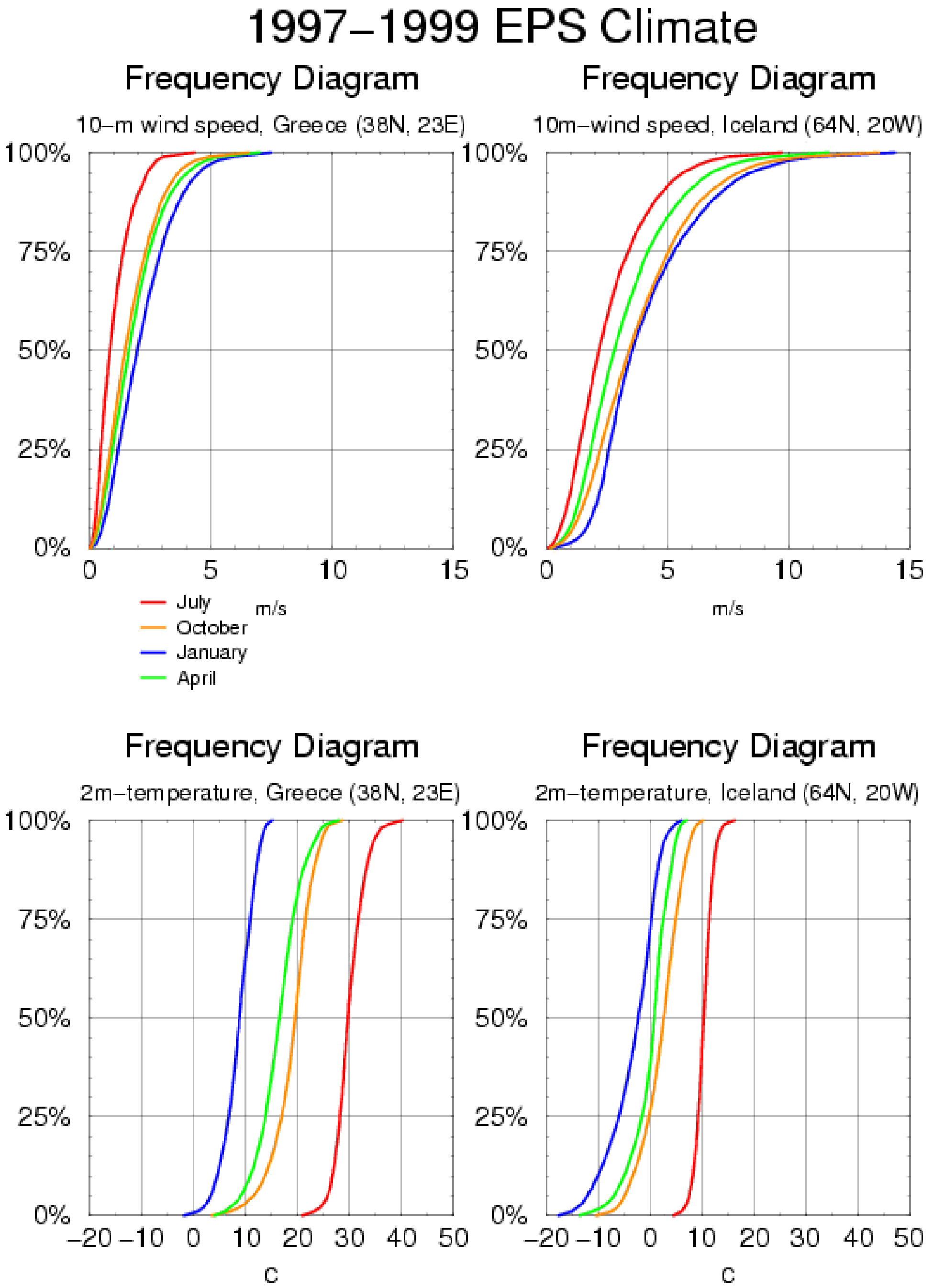


Fig A1.1: Frequency distributions for 10-m wind speed and 2m-temperature in the 1996-1999 EPS climate

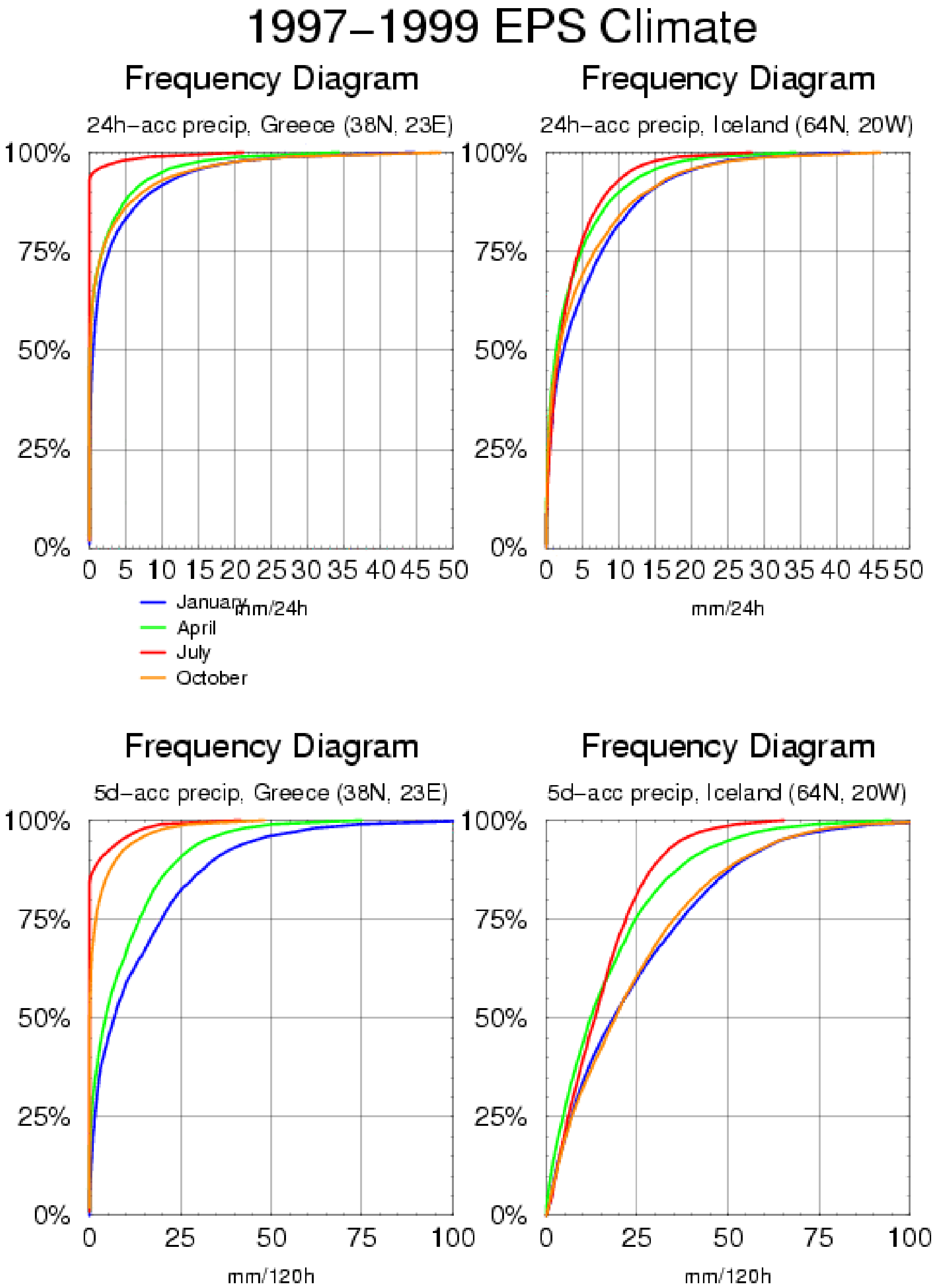


Fig A1.2: Frequency distributions for precipitation (one-day and 5-days accumulation) in the 1996-1999 EPS climate

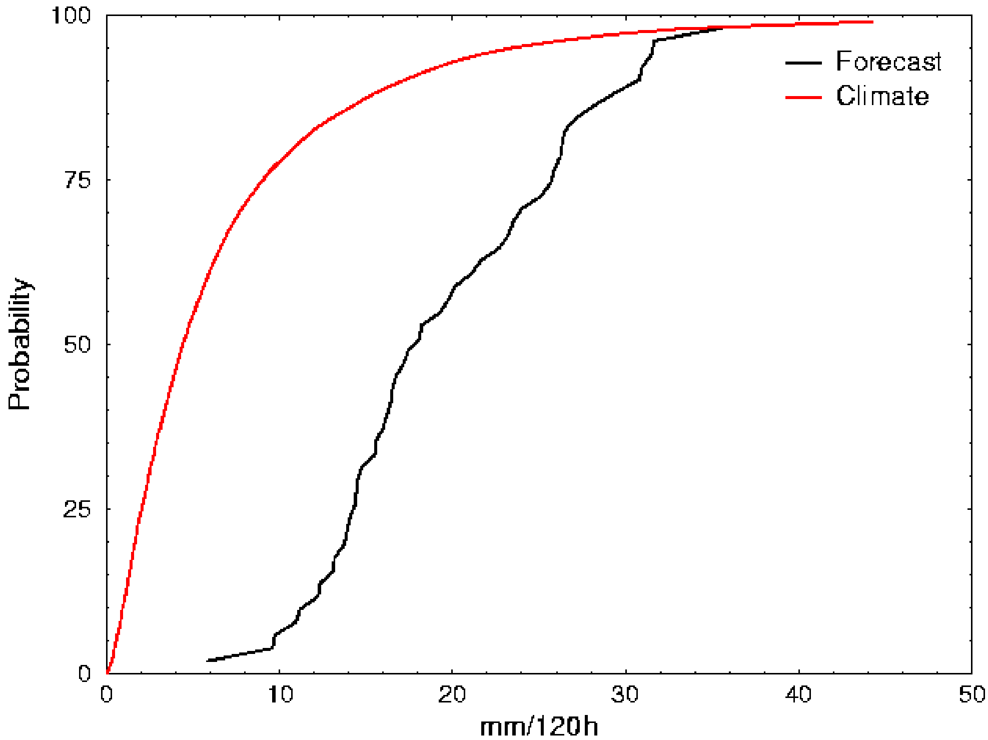


Fig A1.3: Forecast and Climate probability distributions (Base 15/2/2000, precipitation accumulated between 48 and 168h steps)

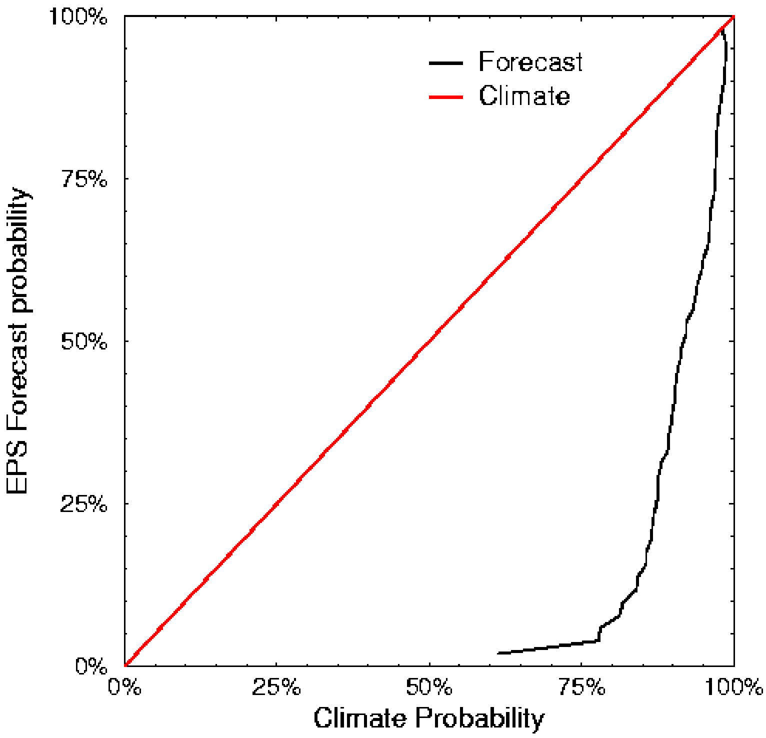


Fig A1.4: Forecast vs Climate Probability diagram. Corresponding EFI is 0.719 (CRPB is +0.402)

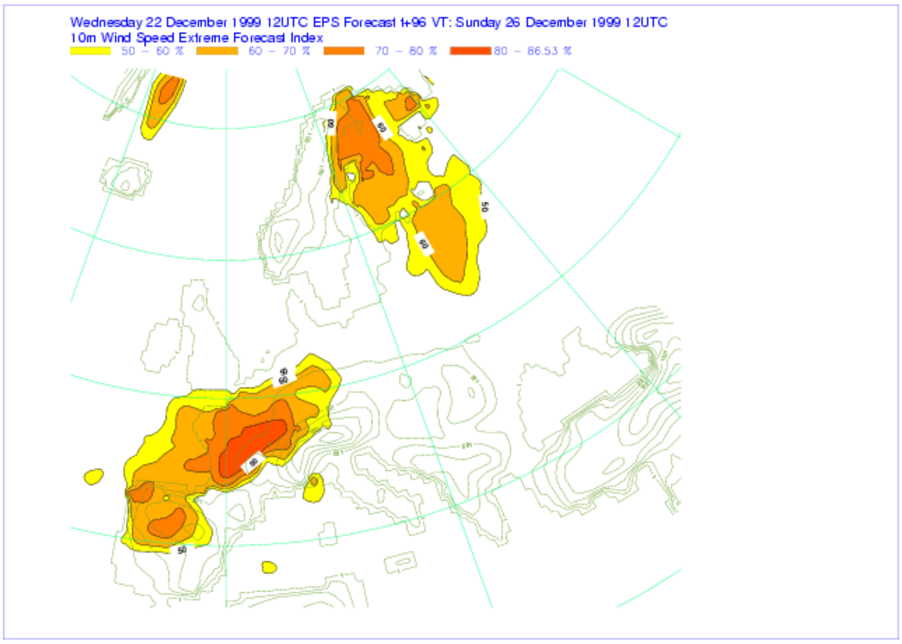


Fig A1.5: Extreme Forecast Index Map for the 26 December 1999 (96h forecast)

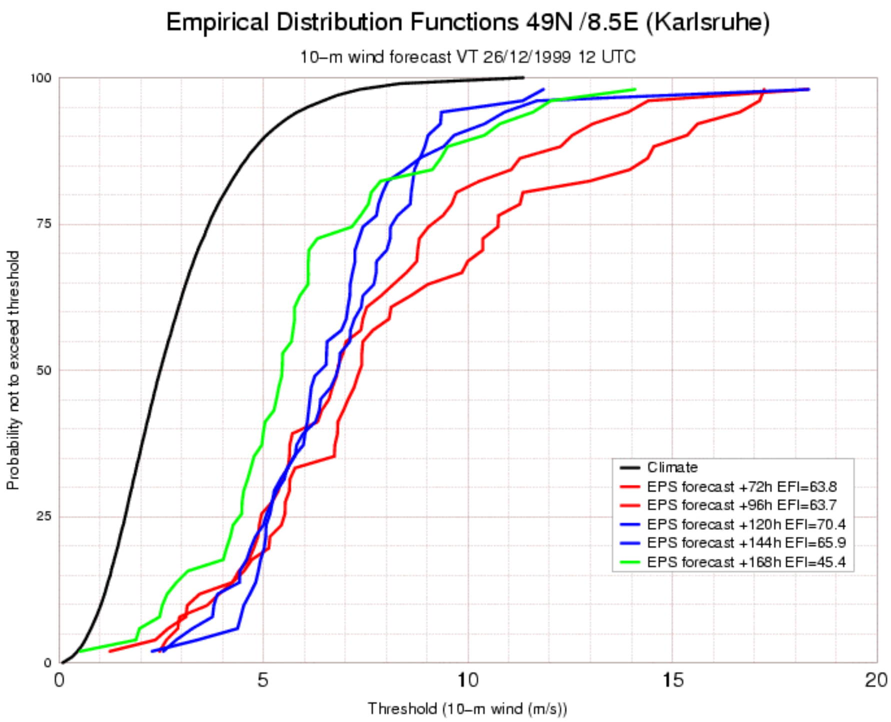


Fig A1.6: Empirical Distribution Functions for Karlsruhe valid on 26 December 1999 12UTC (reported 10m wind speed at this time at the weather station was 25m/s)

5. Validations issues

A major problem that will be faced when designing products targeted for severe weather conditions is that the validation of such products will be limited to a very small number of events. It will certainly be useful to demonstrate that these products do deliver the expected information for past occurrences of severe weather conditions - this is what has just been done in the previous section by providing an EFI map for the "Lothar" storm of December 1999. However a forecast system can only be qualified by the combined measure of its ability to detect the events and to avoid false alarms - the latter being measured by the number of forecasts that produce an event that will not verify. At the scale of a European country, only a few events that can be defined as severe hopefully happen every year, and these can be very different in nature (floods, thunderstorms, wind gusts, snow). In order to gather significant verification statistics for probability forecasts of extreme weather, the only way forward will be to aggregate the verification statistics over areas large enough to ‘catch’ a good number of events. In order to be consistent, it will then be necessary to normalise the events definition to the local climate - ranked verification with respect to both the model and the local climate would be an example.

A different approach has been adopted in a study conducted in collaboration with Météo-France central forecasting office in Toulouse. In order to explore the usefulness of the EFI in operational context, they have computed False Alarm Rates and Non Detection Rates on the basis of the evaluation of the warnings that are issued by the regional forecast offices (they are based on local observations and reports analysed by the local forecasters). Results are shown in Fig A1.7. In average, provided the EFI threshold used to detect the extreme events is set to about 50%, the Non Detection Rate (ratio of observed but not forecast events over the total number of observed events) is rather low (around 20% for precipitation at 72h, but around 50% for the wind). The False Alarm Rate (as defined in this study, it is the ratio of the number of events forecast but not observed to the total number of events forecasted) is rather large, around 60% for rain, around 80% for the wind.

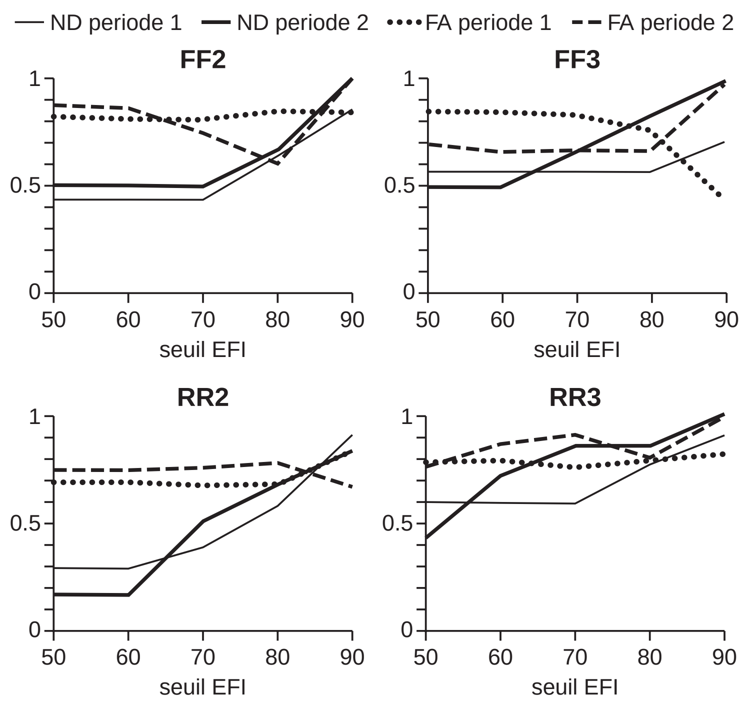


Fig A1.7: Verification of EFI from Météo-France Central Forecasting Office in Toulouse

This triggers a very difficult question to what a reasonable amount of false alarms are. Although it sounds large, a 60% False Alarm Rate means that you will get 6 wrong early warnings issued when 4 prove to be right. A strategy based on warnings issued at random would generate a False Alarm Rate equal to the frequency of non occurrence of the event, which for an extreme event is close to 100%. In order to avoid such shortcomings in the interpretation, the computation of equitable threat scores or similar will need to be the basis of the evaluation of extreme weather forecasts.

Annex 2: The impact of ensemble size on severe weather prediction

1. Introduction

Previous studies have indicated that the impact of ensemble size on forecast performance will depend on the weather parameter of interest and on the decision processes of the individual user. Using a simple cost-loss model of economic decision making, it can be shown that the benefit of increasing ensemble size may be particularly important for rare events and for users with small cost-loss ratios (Richardson 2000, Richardson 2001). Both of these conditions may be expected to apply in severe weather situations.

This paper presents the results of a preliminary study into the effect of increasing the size of the EPS, focusing on the potential benefits for probability forecasts extreme weather events.

The current operational EPS configuration of 51-members at T255 resolution will be referred to as EPS50 and will provide the reference system. This is compared with two alternative larger ensembles. The first comprises 101 members run from the same initial time; the initial perturbations for the 100 member EPS are created using 50 singular vectors (SVs) rather than the 25SVs which are used for EPS50. This 100-member system will be referred to as EPS100. The second additional configuration combines the standard EPS50 with an equivalent EPS50 initialised 12 hours earlier to produce an EPS with a total of 102 members, EPS5050.

31 cases were run in each configuration for the period 27 November to 27 December 1999. EPS100 was initialised at 00UTC on each date and the standard EPS50 was run from both 00UTC and 12UTC. Note that because of the different numbers of SVs used to generate the initial perturbations, EPS50 at 00UTC is not equivalent to a sample of 50 members from EPS100.

With only a small sample of cases it is not possible to derive robust statistics for specific severe weather events. Instead we evaluate probability forecasts of the event “500hPa more than 2 standard deviations below normal” to give an indication of the potential effects of ensemble size. Scores are calculated only for the extra-tropical Northern Hemisphere, again to increase the sample size. The ensembles are assessed using potential economic value V, ROC area and Brier skill score (BSS). In most cases absolute differences in skill are small. To indicate the improvement of one configuration over another we use a measure of relative improvement (RI) defined as a gain in predictability over the reference system. For a performance measure S, we can define the skill of configuration A at forecast step t hours, SA(t), relative to that of system B at step t, SB(t) as

$$RI = \frac{S_A(t) - S_B(t)}{S_B(t-12) - S_B(t)} \times 100\% \qquad \text{(A2.1)}$$

So a relative improvement of 100% indicates that forecasts from system A are as good as forecasts from system B started 12 hours later. The choice of 12 hours as the reference interval is made to match the difference in initialisation time in EPS5050 configuration.

2. Results

We compare both 100-member configurations, EPS100 and EPS5050, with the reference system EPS50 initialised at 00UTC. Both larger configurations show a substantial increase in ROC area (Fig A2.1). Beyond day 2, RI is generally greater than 100%, bringing a gain in predictability of 12 hours or more. EPS100 has a consistent advantage over EPS5050. Fig A2.2 shows potential economic value, V, for the day 6 forecasts. The benefit of the 100-members is most apparent for small cost-loss ratios (C/L). Users with high values of C/L will see little difference between the 50 and 100-member configurations. Again, the best performance comes from the EPS100 system. Differences in Brier skill score are smaller than for ROC area (Fig A2.3), but there is a consistent positive signal for EPS100. EPS5050 is generally slightly less skilful than the reference EPS50 for this measure.

Based on these results, generating a 100-member EPS by generating more SVs at a single data time is more beneficial than the alternative approach of increasing ensemble size including members from earlier start times.

However, there is another potential benefit of the time-lagged EPS5050 approach. This system can be updated twice daily rather than once per day for the equivalent cost EPS100. The potential gains to be achieved by this more frequent updating are shown in Figs A2.4 to A2.6. In these figures we use EPS100 (from 00UTC) as the reference system. We compare it to two time-lagged configurations. The first, EPS5050 as defined previously, combines 50 members from 00UTC with 50 members initialised 12 hours earlier. The second one combines 50 members from 00UTC with 50 members initialised 12 hours later; this will be referred to as EPS5050+.

The added benefit of including forecasts from a later analysis time are seen in both ROC area and BSS. The relative improvement is rather noisy for ROCA (a sign of the limited sample size) and therefore is not shown. For BSS, the indication is that the RI is consistently above 40% but generally less than 100%. This suggests that EPS5050+ offers improvement over EPS100, but the benefit is not as great as would be seen by running EPS100 every 12 hours. Differences in value between EPS100 and EPS5050+ are small for all users, but there is a small advantage for EPS5050+ (Fig A2.5).

3. Conclusions and future plans

This initial study has identified the potential benefits for forecasting extreme events to be gained from increasing ensemble size. Two methods of generating larger ensembles were considered. Both of these demonstrated benefits over the current 50-member configuration. This was particularly so for users with small cost-loss ratios; differences between the systems would be less apparent for users with larger C/L.

Overall, the greatest benefit comes from including as many members as possible from the most recent analysis time. Generating more perturbations using more SVs is more beneficial than simply increasing ensemble size by including members from earlier start times. However this advantage must be weighed against the potential additional benefit of twice daily updating which would be possible for 50+50 configuration. Which approach is most beneficial in practice may depend on the operational scheduling of the forecasts and the time at which they become available and are used by the forecasters. If the EPS will only be used once per day, the best approach is likely to be to use 100 members from the most recent possible start time. However if EPS forecasts are updated or used at various times throughout the day, the benefit of the 50+50 system will become more apparent.

Given the limited sample size in this preliminary study, specific severe weather events and synoptic aspects have not been addressed. There is some indication that in particular instances of severe weather the EPS100 may give a better indication of possible scenarios than the 50+50 system (A Persson, personal communication). A larger sample is needed to investigate these aspects further as well as to generate more robust statistics.

Since both EPS100 and EPS5050 offer potential benefits, the relative advantages of each configuration need to be assessed over a larger sample. As part of the investigation of severe weather prediction the standard EPS configuration is already being run routinely at 00UTC as well as the operational 12UTC run. To compare this with EPS100, 100 members are also being run, every four days, from 12UTC. This will then build up a large sample of forecasts with which to compare the systems. A more detailed evaluation will be undertaken than was possible with the limited data of the preliminary study. This will include, in addition to the standard upper air verification, weather parameters (precipitation, surface winds) and synoptic events.

References

Richardson, D. S, 2000. Skill and economic value of the ECMWF ensemble prediction system, *Q.J.R.Meteorol. Soc*., **126**, 649-668.

Richardson, D. S, 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q.J.R.Meteorol. Soc*.To appear.

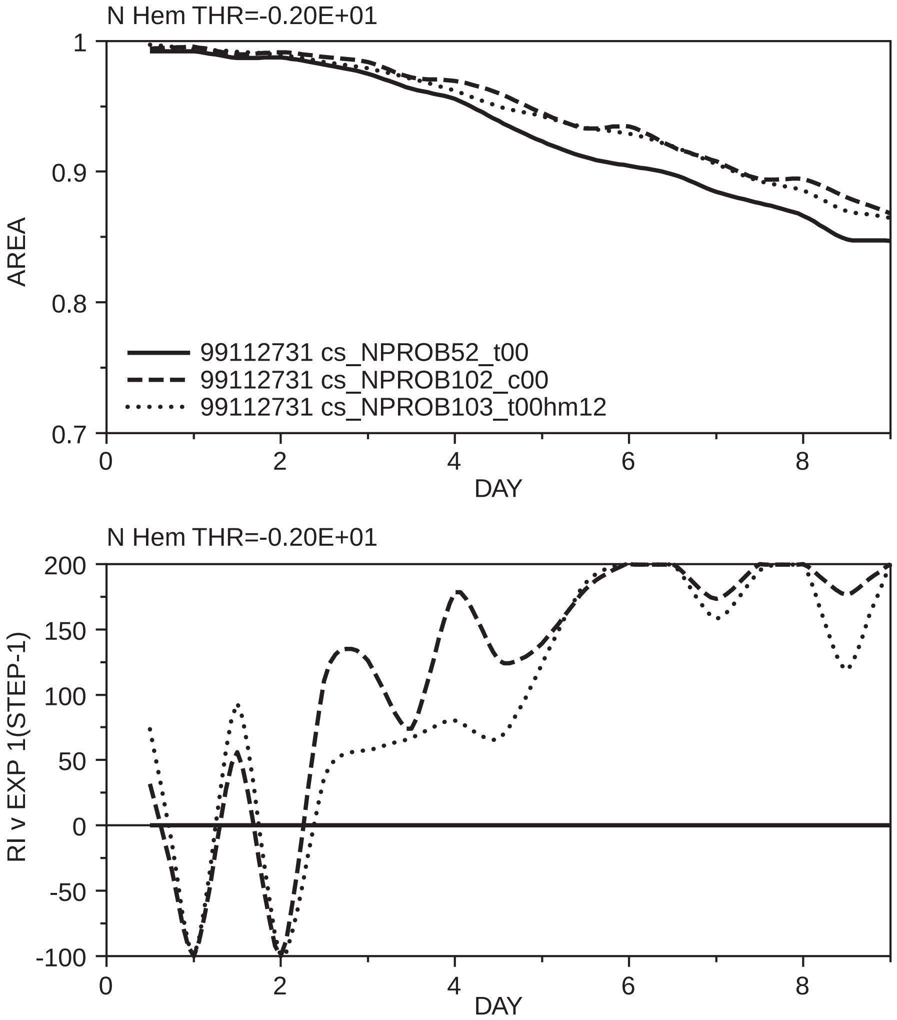


Fig A2.1: ROC area for EPS probability forecasts of the event '500 hPa anomaly more than 2 standard deviations below normal' over the extra-tropical Northern Hemisphere for 31 cases (Nov/Dec 1999). Upper panel: ROC area; lower panel: Relative improvement RI. Curves show EPS50 (solid line); EPS100 (dashed line); and EPS5050 (dotted line). See text for details.

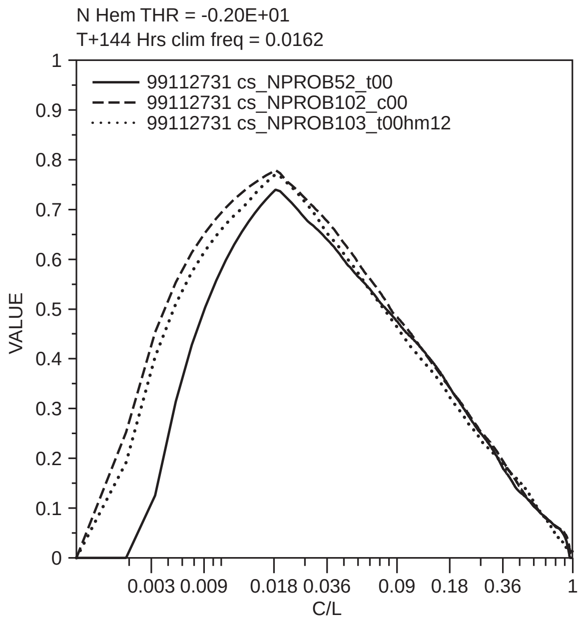


Fig A2.2: Potential economic value V for day 6 EPS probability forecasts of the event '500 hPa anomaly more than 2 standard deviations below normal' over the extra-tropical Northern Hemisphere for 31 cases (Nov/Dec 1999). Curves show EPS50 (solid line); EPS100 (dashed line); and EPS5050 (dotted line).

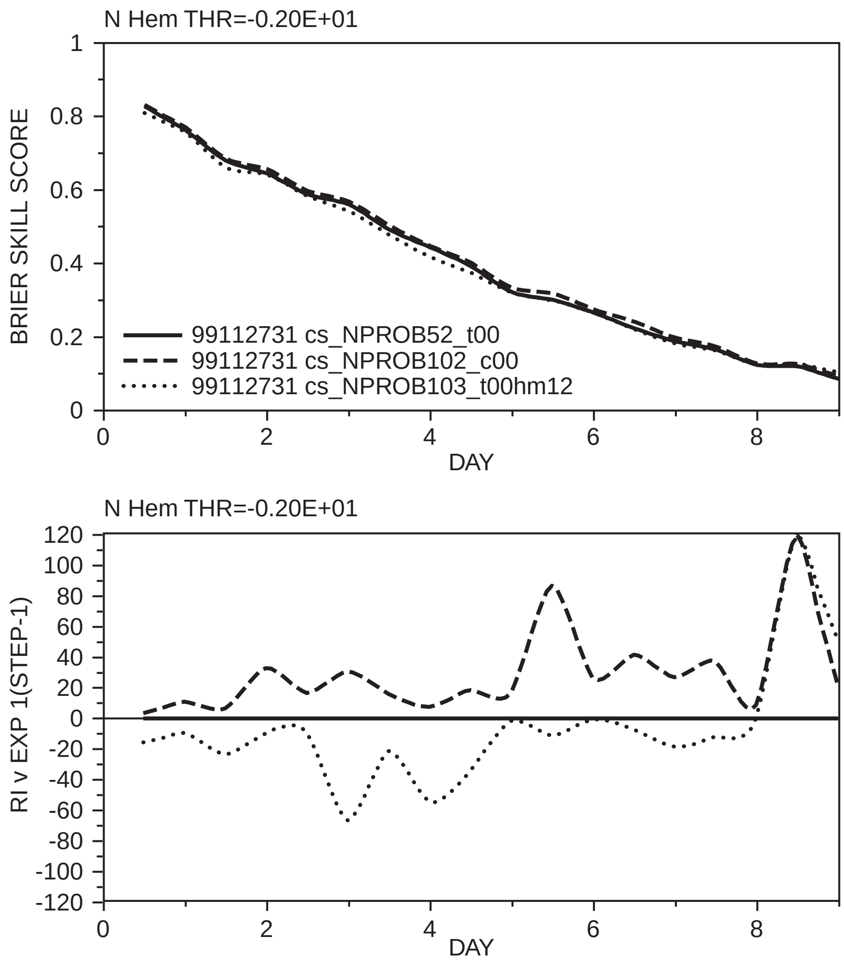


Fig A2.3: As Fig A2.1, but for Brier skill score.

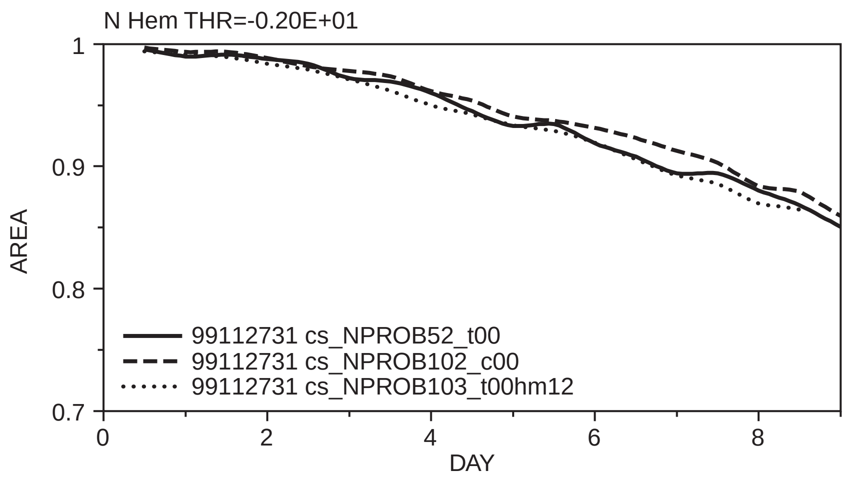


Fig A2.4: As Fig A2.1 but for EPS100 (solid line); EPS5050+ (dashed line); and EPS5050 (dotted line). RI not shown.

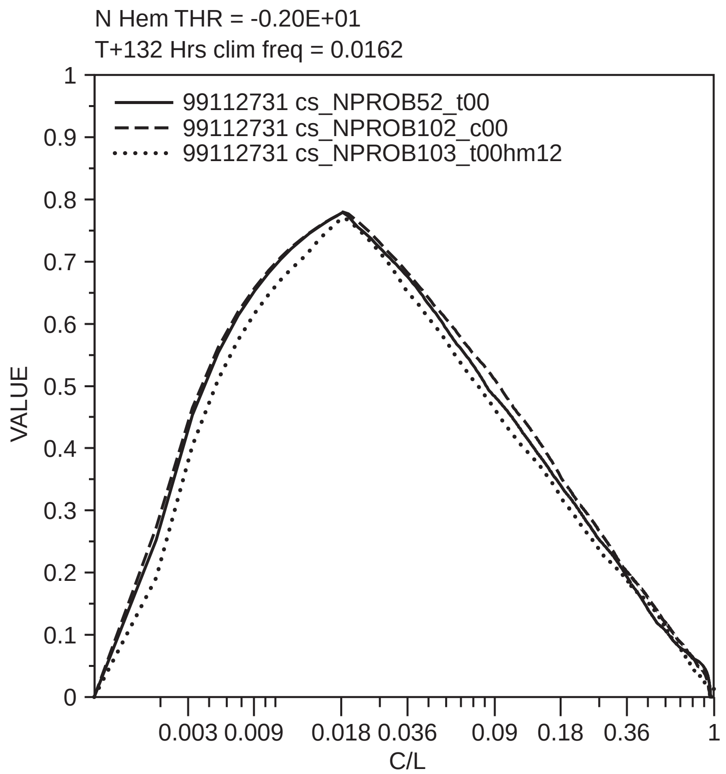


Fig A2. 5: As Fig A2.2 but for EPS100 (solid line); EPS5050+ (dashed line); and EPS5050 (dotted line).

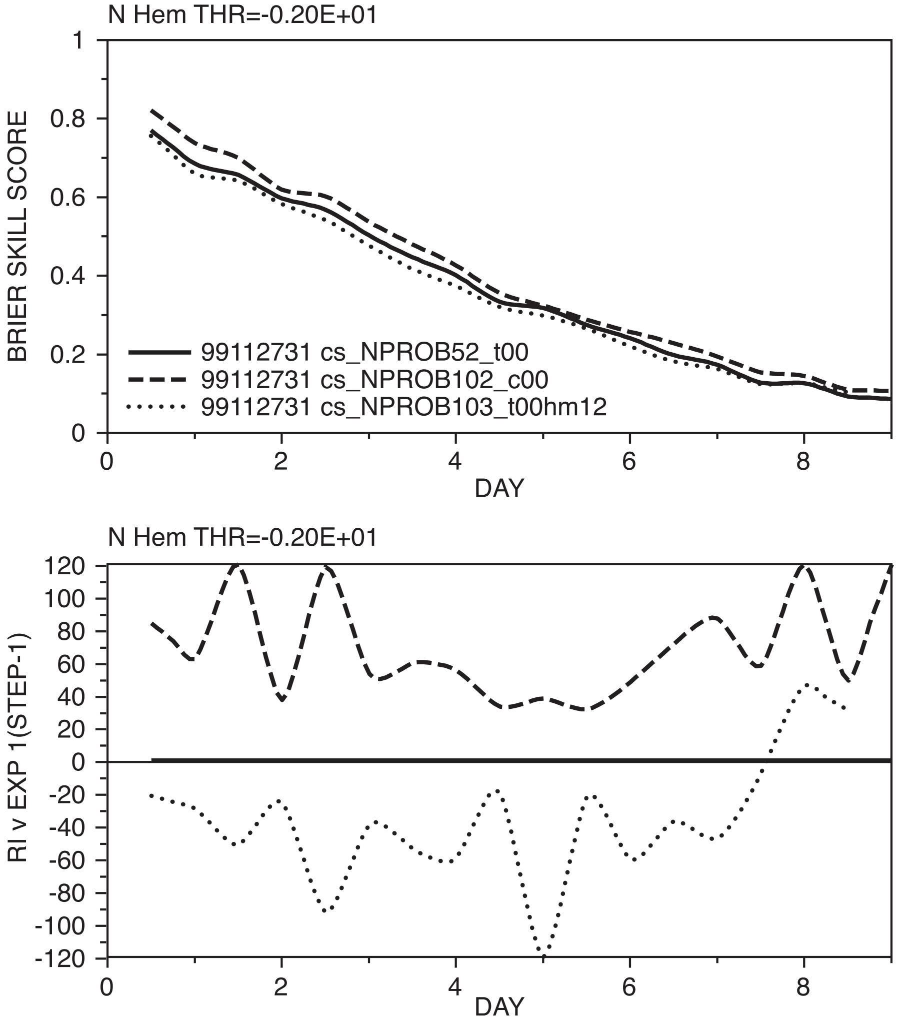


Fig A2.6: As Fig A2.3 but for EPS100 (solid line); EPS5050+ (dashed line); and EPS5050 (dotted line).

Annex 3: Tropical cyclone ensemble forecasting

1. Introduction

With the development of linearized versions of important components of the ECMWF forecast model, such as, vertical diffusion, subgrid-scale orographic effects, large-scale condensation, long-wave radiation and deep cumulus convection, it is possible to determine singular vectors (SVs) for situations where physical processes may contribute significantly to perturbation growth. Integrations with finite-size perturbations have shown that the fit between the linear and nonlinear model improves when physics is included in the linear model (Mahfouf, 1999). It therefore seems natural to extent the area of initial EPS perturbations to the tropical region by adding perturbations based on diabatic SVs to the already existing extra-tropical perturbations determined by SVs using simplified physics.

In two companion papers (Barkmeijer et al., 2001; Puri et al., 2001) several properties of tropical SVs are described and their impact on the EPS performance is studied, focusing especially on ensemble forecasting of tropical cyclones (TCs). One of the results reported there was that using the entire tropical region, e.g. 20°S - 20°N, as target area in the SV computation was not very useful in tropical cyclone (TC) ensemble forecasting. Although in such a set-up, SVs are computed which produce the largest perturbation growth for the tropical region in terms of total energy, usually only a few, if any, SVs were associated with the TC dynamics. The ensembles with perturbations based on such SVs showed spread in the TC intensity, but this was mainly due to the so-called stochastic physics. This scheme tries to simulate uncertainties in the model description (Buizza et al., 2000). The spread in the TC tracks was negligible.

To benefit from tropical diabatic SVs in TC ensemble forecasting, it was necessary to define target areas in the vicinity of TC locations. As a consequence, multiple SV computations are required when several TCs coexist. In order to reduce the overall numerical costs for determining perturbations in the tropics, a multi-Gaussian sampling technique was adopted as described in Ehrendorfer (1999). For each TC, 50 initial EPS perturbations are determined by sampling from the leading five SVs and added to the already existing perturbations.

2. Tropical singular vectors

The tropical SV computation does not differ from the extra-tropical computation with respect to the defining norms. Both at initial and final time the total energy norm is used. The main difference is the use of diabatic linear models. Experiments have shown that the simplified physics used in the extra-tropics does not yield SVs which are associated with TCs. An example of this is given in Fig A3.1. Both computations, with optimization time 48 hours and with simplified or diabatic linear models, use the same 30° x 40° (latitudinal and longitudinal direction) target area around the TC location, denoted by a large dot. The leading SVs using simplified physics seem not associated with the TC and do not show the spiraling structures around the TC location present in the diabatic SVs. Ensemble experiments confirmed the superior properties of diabatic SVs. Diabatic SV based perturbations were able to produce useful spread in the TC tracks as opposed to perturbations based on SVs with simplified physics.

However, the use of diabatic linear models is not sufficient to ensure useful SVs for TC ensemble forecasting. Fig A3.2 shows the locations of the leading 25 SVs when a tropical strip 30°S - 30°N or 20°S - 20°N is used as target area instead of a 30° x 30° target area around the only TC at (10.0°N ; 135.0°E). Even for the reduced strip 20°S - 20°N, the majority of SVs are not located in the vicinity of the TC. It is for this reason that in the subsequent ensemble integrations targeting has been applied in the SV computation for each TC.

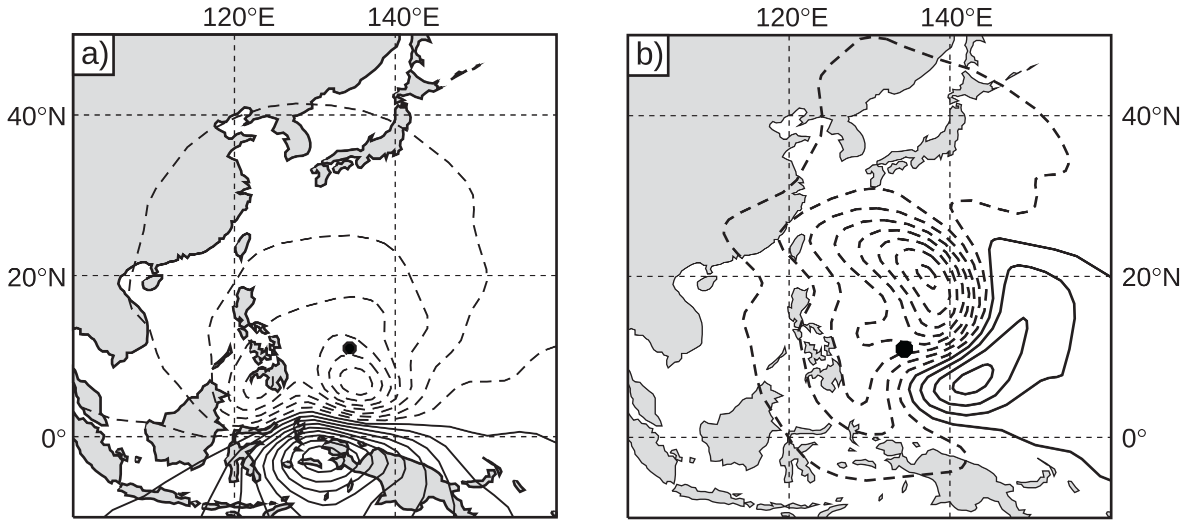
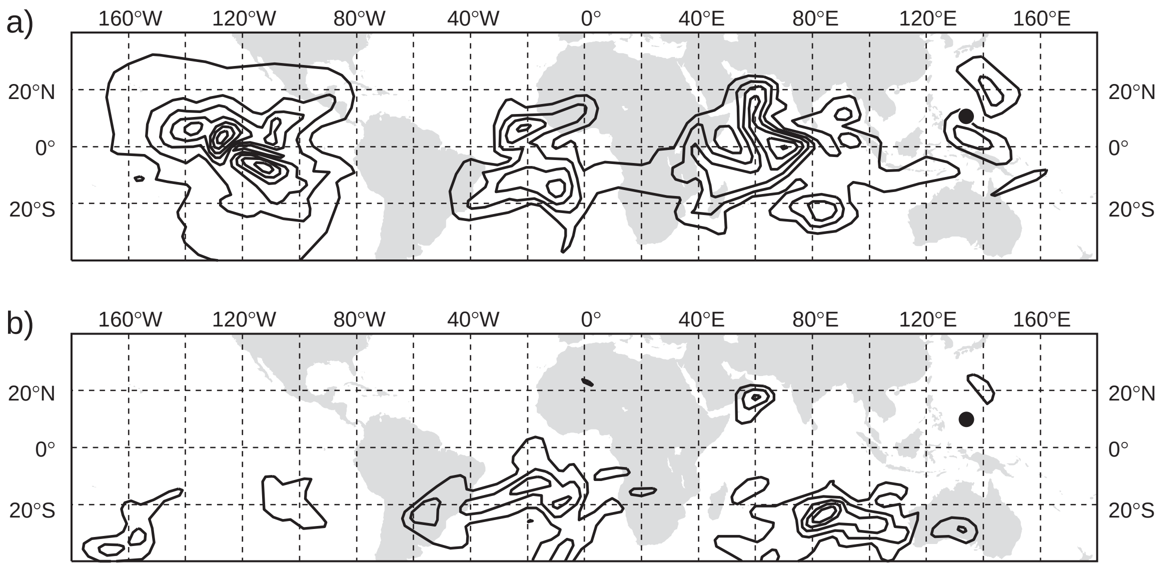


Fig A3.1 Streamfunction fields at 500 hPa of the leading SV, targeted for Zeb and starting from 10 October 1998 at 12UTC, using linear models with (a) simplified or (b) diabatic physics. The location of Zeb is denoted with a large dot.



FigA3. 2: Streamfunction RMS fields at 500 hPa of the leading 25 SVs for tropical cyclone Zeb 11 October 1998 12UTC and targeted for (a) 20°S - 20°N and (b) 30°S - 30°N. The position of Zeb is denoted with a large dot.

2.1 Targeting

The number of SV computations in the tropical region may vary from day to day depending how many target areas are defined. However, to limit the numerical costs, the maximum number of target areas is set to four. The procedure of defining tropical target areas is as follows:

1. Always select the Carribean area: 0°N - 20°N and 90°W - 50°W. The motivation being that weather systems originating from the Carribean area may influence medium-range European forecasts.

2. Identify in the tropical strip 25°S - 25°N all systems of type **Hurricane, Typhoon, Cyclone,Severe tropical storm** or **Tropical storm** (i.e., tropical systems with WMO classification index larger than 1) and define for each a target area given by a 30° x 40° box (latitudinal and longitudinal direction) around the location intersected by the strip 25°S - 25°N.

3. Merge the closest target areas in case more than four target areas are defined. Suppose there are N target areas with Western longitude coordinate a_j and $0° \leq a_1 < a_2 <\ldots< a_{N+1} < 360°$. Concatenate areas j and $j + 1$ with the smallest distance $a_{j+1} - a_j$, where area N+1 is identical to area 1. Repeat this process until 4 areas remain.

3. Ensemble results

To study the impact of targeted tropical SVs on TC tracks, but also on the extra-tropical probability scores, 12 cases have been run as listed in Table A3.1. All experimentation was performed with a T_L255L40 ECMWF model resolution and with model cycle 23R3, including the stochastic bugfix (introduced in operations 19 December 2000) and the corrected perturbation size (RNORM=2, introduced in operations 16 Februari 2001). Since only a limited number of SVs (around 10) are available per target area, it was not possible to employ the operational rotation technique as used in the extra-tropics (Molteni et al., 1996). Instead, multi-Gaussian sampling (Ehrendorfer, 1999) has been applied in the tropics to produce 50 perturbations from the 5 leading SVs per target area. Schematically the EPS perturbations can be written as

$$p_i = \sum_{\text{target areas}} \sum_{j=1}^{5} a_{ij} SV_j \qquad\qquad i = 1,\ldots,50,$$

where p_i , i = 1,…,50 are the 50 extra-tropical (+/-) perturbations and a_{ij} are random numbers drawn from a normal distribution with mean 0 and a standard deviation to ensure that the same initial ensemble spread results as with the operational perturbation technique, for more details see Ehrendorfer (1999).

Date	Area	Name	Type
19980819	Atlantic	Bonnie	cyclone
19981011	Pacific	Zeb	typhoon
20000107	Indian Ocean	Babiola	cyclone
20000820	Pacific	Bilis	typhoon
20000924	Atlantic	Isaac	hurricane
20000927	Atlantic	Isaac	hurricane
20001001	Atlantic	Keith	hurricane
20010104	Indian Ocean	Ando	cyclone
20010119	Indian Ocean	Charly	tropical storm
20010212	Indian Ocean	Vincent	cyclone
20010227	Indian Ocean	Paula	cyclone
20010228	Indian Ocean	Paula	cyclone

Table A3.1. List of tropical cyclone cases.

3.1 Cyclone tracks

The tropical perturbations have a clear impact on the EPS cyclone tracks. Fig A3.3 shows the tracks for cyclone ANDO of the 50 ensemble members for a forecast period of 4 days and starting from 4 January 2001. In the operational EPS, as shown in Fig A3.3a, all tracks follow the same route up to day 3 and there is a high probability that the cyclone will make landfall over La Reunion between day 2-3. The experimental EPS, however, makes this scenario less likely by indicating other possible cyclone tracks.

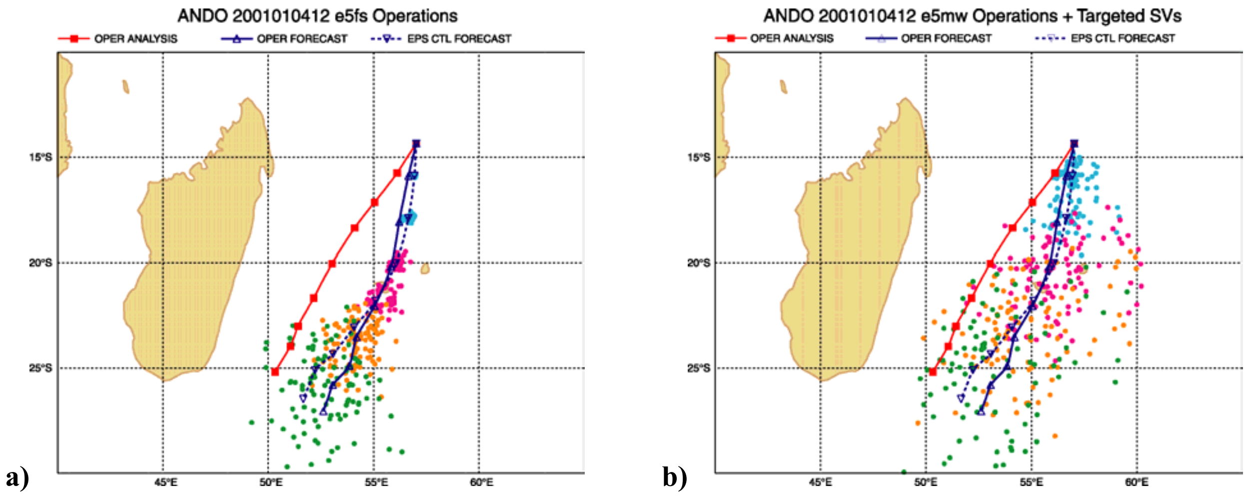


Fig A3.3: Tropical cyclone positions up to 4 days ahead for cyclone ANDO in the (a) operational and (b) experimental EPS using targeted tropical perturbations, started form 4 January 2001. Symbols denote the cyclone position in the EPS for 0-to-24 hour (blue), 24-to-48 hour (purple), 48-to-72 hour (orange) and 72-to-96 hour (green) forecasts. The red line denotes the analysed track (symbols denote the position every 12 hours), the solid blue line the track predicted by the operational forecast and the dotted blue line the track predicted by the EPS control forecast.

To investigate whether tropical perturbations are beneficial for the skill of the TC tracks in the ensemble, distances between the ensemble tracks and the analysis for the TC location have been computed for certain forecast periods and averaged over the 12 cases. Fig A3.4 gives the number of members, averaged over the 12 cases, closer to the analysed TC location for selected distances and forecast periods. For a lead time of 1 day, the EPS unperturbed control forecast is quite capable to position the TC accurately. As seen from Fig A3.4, initial perturbations may slightly deteriorate the skill of the ensembles members for this short forecast range. However, for longer lead times the number of skilful members is larger in the experimental ensemble.

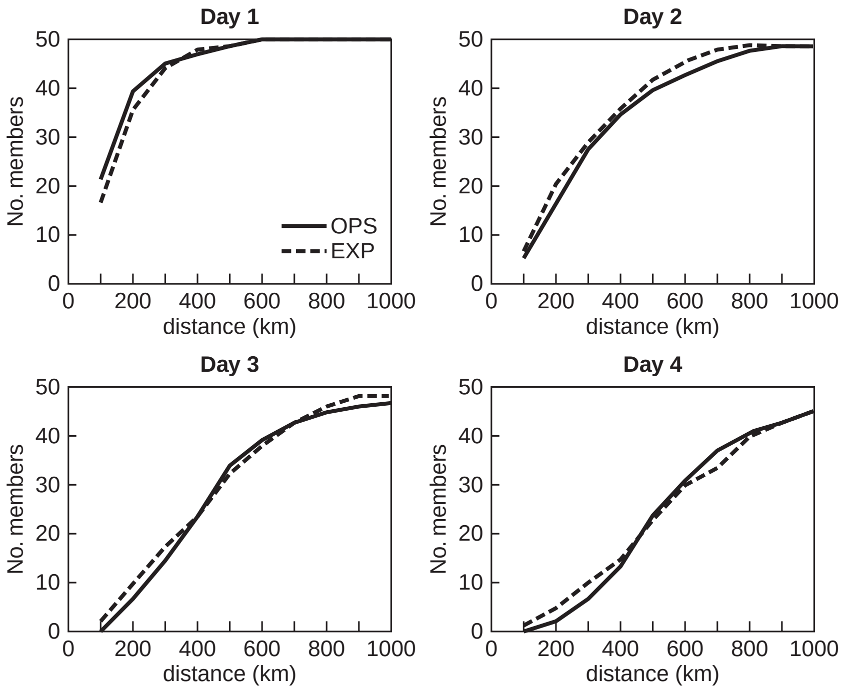


Fig A3.4: Number of ensemble members (averaged over 12 cases) closer to the analysed position of the tropical cyclone as a function of distance, for the operational EPS (solid line) and the EPS with tropical perturbations (dashed lines). The forecast period is given above each panel.

3.2 Probability scores

Additional ensembles have been run for the period September-October 2000, starting from 1 September 2000 and every 5 days. Note that even when tropical cyclones are absent, always the Carribean area is targeted. Supplemented with cases from Table A3.1 for the same period, this gave a sample comprising 15 cases for which probability scores were computed for geopotential height at 500 hPa and total precipitation accumulated over 12 hours.

The standard probability scores were used, such as, the area under the Relative Operating Characteristic (ROC) curve as a measure of the capability of the EPS to discriminate between hit and false alarm rates, Brier score and Brier skill score computed with respect to a climatological forecast.

Generally speaking, the impact of tropical perturbations on the probability scores is neutral for 500 hPa geopotential height and slightly positive for total precipitation for the tropics and NH. An example of the latter parameter is shown in Fig A3.5 for the event ‘accumulated rain over 12 hours larger than 5 mm’ for the NH.

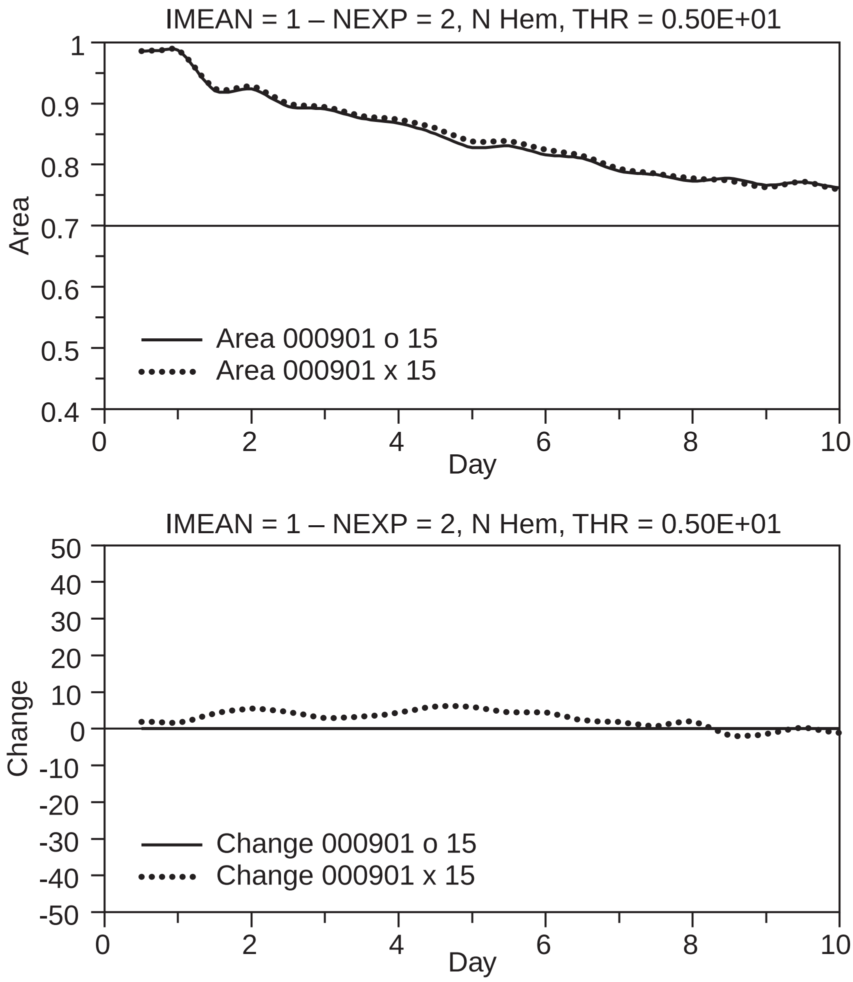


Fig A3.5: Area under the ROC-curve for the event 'accumulated rain over 12 hours larger than 5 mm', and evaluated for the NH (averaged over 15 cases).

4. Conclusion

The current EPS lacks initial perturbations in the tropical region and will for this reason not inform optimally about probable future states of the atmosphere for this region. Sampling from the leading SVs which were targeted for tropical cyclones resulted, on average, in more skilful cyclone tracks for lead timesof 2-4 days. In some cases there was a clearly improved guidance on possible TC tracks (see Fig A3.3). The impact on Z500 probability scores was neutral; for precipitation accumulated over 12 hours probability forecasts were slightly better for the tropics and NH.

In order to have SV targeting procedure without any manual intervention, an automatic procedure to read GTS bulletins containing warning on tropical systems is being tested at the moment.

References

Barkmeijer, J., Buizza, R., Palmer, T.N., Puri, K., and J.-F. Mahfouf, 2001: Tropical singular vectors computed with linearized diabatic physics. *Quart. J. Roy. Meteor. Soc.*, **127**, 685-708.

Buizza, R., Miller, M. and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Ehrendorfer, M., 1999: Prediction of the uncertainty of numerical weather forecasts: Problems and approaches. In *Proceedings of the ECMWF Workshop on Predictability*, 20 - 22 October 1997, pp. 27-99.

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

Molteni, F., Buizza, R., Palmer, T.N. and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269-298.

Puri, K., Barkmeijer, J. and T.N. Palmer, 1999: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709-731.

Annex 4: Prediction of extremes from downscaling parametrisations

1 Precipitation

1.1 The problem

With the introduction of CY21R4 the model predicts not only the grid-mean convective and stratiform precipitation, but also the fraction of the grid-box that is covered by stratiform precipitation, hereafter referred to as precipitation fraction. Additionally, while post-processing convective precipitation as grid-mean value, the model internally always assumed that this precipitation is falling over a small part of the grid-box (5-10%). Using the information on the area-coverage with precipitation it is in principle straightforward to produce a probabilistic forecast of precipitation at any given time.

If, say, the precipitation fraction is 25% and the grid-mean stratiform precipitation is 1mm/h, the probabilistic forecast would be that there is a 25% chance for rain in the grid, with the amount being 4mm/h. This is a much more useful forecast than predicting 1mm/h for the entire grid and verification against individual SYNOP stations is simplified to similar means as used to verify EPS probability forecasts. There are, however, two major caveats: i) if coexistent, convective and stratiform precipitation areas need to be overlapped; and ii) precipitation is neither measured nor archived from models as instantaneous values.

Overcoming i) is not too difficult given the ECMWF model physics. Here a very strong link between convective and stratiform clouds (and hence precipitation) exists, that justifies assuming maximum overlap between the two areas. The second of the two problems is much more difficult to overcome, since it involves modelling the overlap of the (only discreetly known) precipitation events that occur over the accumulation time. This requires knowledge about the space-time correlations of precipitation under various flow conditions. Note that whatever overlap model is employed, it needs to be designed to be independent of the sampling interval (=model time-step) of the individual precipitation events.

1.2 An idealised model

In order to investigate some of the issues involved in `translating' grid mean precipitation and fraction information into a probabilistic product a toy model was developed. The model attempts to overlap in time the individual precipitation events. It uses a discretization approach that divides a model grid-box into 20 smaller sub-boxes and predicts the precipitation in each of the sub-boxes. This approach is similar to that of Jakob and Klein (1999) who used it to overlap many vertical levels of varying cloud cover. Here the vertical levels are replaced by time and the cloud cover by precipitation fraction. In order to ensure independence of the sampling interval (time-step) the model predicts the overlap of individual precipitation events as a function of wind-speed. The underlying assumption is that the precipitation area is advected through a grid-box with the mean wind. Fig A4.1 shows the distribution of precipitation produced by the toy model over 6 hours assuming constant values for grid-mean stratiform precipitation of 5mm/d, grid-mean convective precipitation of 10mm/d, stratiform precipitation fraction of 0.3, and convective precipitation fraction of 0.1. The assumed windspeed is constant at 3m/s. The 4 different lines represent different assumed model time steps from 12 to 3600 s. The horizontal line shows the grid-mean value of total precipitation, which is the information currently used as model forecast of precipitation. The model grid-size is assumed to be 120 km.

It is obvious that the rainfall distribution produced by the toy model is significantly different from a homogeneous mean currently used. In fact, with the parameter settings used, there is a 15% chance for any station in the grid not to experience any precipitation in the 6 hour interval, while there is a higher than 50%

probability that the precipitation measured at anyone location is larger than the grid-mean value. It is also evident, that with the settings used here, time-step independence is relatively well ensured. This is so, because the boundary of the precipitation area (30% after maximum overlapping convective and stratiform areas), remains within the grid box, which is evident from the fact that there are still sub-boxes without any precipitation.

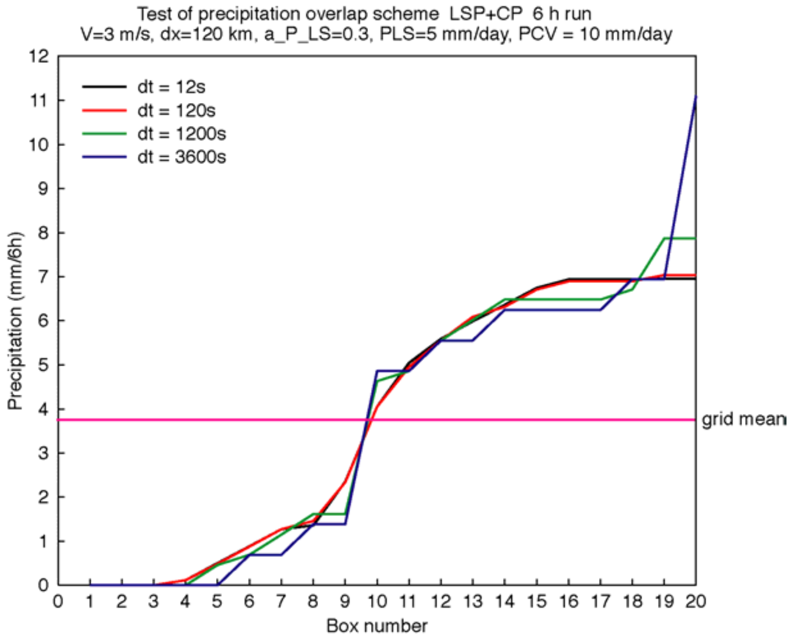


Fig A4.1. Distribution of precipitation within a single model grid box over 6 hours produced by the toy model (see text). The parameter settings assume constant values for grid-mean stratiform precipitation of 5 mm/d, grid-mean convective precipitation of 10 mm/d, stratiform precipitation fraction of 0.3, and convective precipitation fraction of 0.1. The assumed windspeed is constant at 3 m/s. The lines represent results with different model time steps. The horizontal line indicates the grid-mean precipitation.

If one increases the windspeed to 10m/s (Fig A4.2) the latter is not true anymore and additional assumptions about grid-box boundary conditions need to be made. For lack of any better method periodic boundary conditions are assumed in Fig A4.2, i.e., any precipitation moving out of the grid-box at one side is assumed to move in again at the other side. Although the word boundary condition implies spatial features, one needs to keep in mind that ultimately one wants to devise a statistical model of cloud overlap, which needs however to include some length and time scale to assure time-step independence. Fig A4.2 shows various interesting features. First of all, for time-steps less or equal to 1200 s, time-step independent results are achieved. The precipitation distribution shows much less variance, and deviations from the grid-mean are smaller. This is so, because at the higher assumed wind-speed the precipitation area moves through the grid-box much more rapidly and the precipitation gets spread out more. It is also evident that for very long time steps (3600 s) the results are quite different from shorter time-steps. This is partly due to the choice of parameters but might indicate that there are numerical constraints on successfully applying the simple toy model, similar to a CFL criterion for advection.

1.3 Summary

A simple toy model was developed to explore the potential of using already existing model information to produce probabilistic precipitation forecasts using the deterministic model. Despite using extreme simplifications of the full problem, it was shown that, using grid-mean information only (as done now), is potentially extremely misleading with respect to the information used in parametrizing precipitation within the model. This has consequences for both the model verification and the prediction of extreme rainfall maxima, if those occur on the sub-grid scale. Further research and some collaboration with hydrologists is necessary to generate a more realistic model of space-time overlap of precipitation events. The potential benefits for forecasting severe weather events might be substantial.

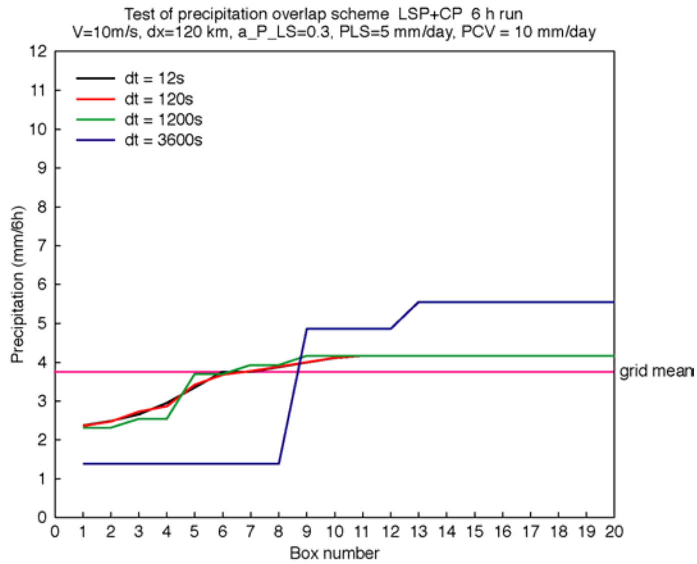


Fig A4.2. As Fig A4.1 but with an assumed constant wind speed of 10 m/s.

2 Wind gusts at 10 m height

Wind gusts are reported in strong wind conditions, because the ‘extremes’ are relevant for storm damage. The word "extreme" needs some specification, as it depends on the filtering characteristics of the anemometer system, how extreme the observed wind will be. A fast response system will measure higher gusts than a very heavy slowly responding anemometer. In order to get uniform observations, WMO defines a wind gust as the maximum of the wind averaged over 3 second intervals. The requirement for persistence over a few seconds is relevant because only an extreme that lasts for a few seconds causes enough drag on structures to result in damage.

Because of the importance of extreme wind forecasts, wind gusts are now parametrized and post-processed in the IFS model and stored in the MARS archive. The principle is that the standard deviation of the horizontal wind is estimated and that the difference between gust and mean wind is made proportional to this standard deviation on the basis of universal turbulence spectra. Wind standard deviations and spectra are based on standard Monin Obukhov similarity, where horizontal wind fluctuations are characterized by surface friction, surface buoyancy flux, and boundary layer depth (see Beljaars, 1987 for a description of the theory and comparison to observations). From the controlling parameters it is clear that the effects of surface friction (through surface roughness) and stability are captured. However, the approach might be less adequate for gusts in baroclinic situations and gusts due to strong convective events.

Fig A4.3 shows an example of a strong wind situation with the mean wind in the upper panel and gusts in the lower panel. The observed mean and gusts from SYNOP stations are shown by printed numbers. It should be noted that the mean wind is an instantaneous field whereas the gust field represents the maximum gust since the last post-processing time (referring to the previous 3 hour interval in this case).

The gust field has the expected characteristics: gusts are less affected by coastlines than the mean wind and the gusts are substantially higher than the mean. The extremes in the South of England and the coast of Brittany and Normandy are reasonably well captured, but more inland in France the model gusts are too low. As can be expected from gust observations, they vary substantially from one location to the other, so a big sample will be needed in order to get a good impression of possible biases.

Operational validation of wind gusts has begun. First results show some negative biases (-2/-3m/s); however, such verifications are biased by the fact that wind gusts are available from SYNOP messages only when gusts are in excess of certain threshold values.

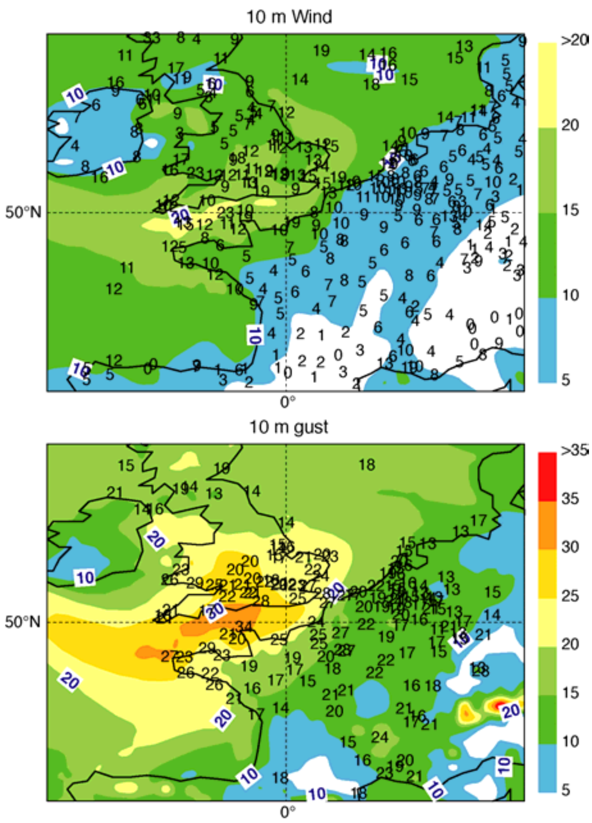


Fig A4.3. Wind at 10m (top panel) and maximum probable gust over the previous 3 hours (bottom panel) for 8 December 2000 6 UTC. The contours are from a 42 hour forecast; the printed numbers are observations in m/s.

In principle, point-wise forecasts of any specific wind gust value can be treated as probabilistic in much the same way that point-wise rainfall can. Indeed the computation of maximum wind gust described in this section is based on an assumed probability distribution of point-wise wind.

References

Beljaars, A.C.M., 1987: The influence of sampling and filtering on measured wind gusts. J.Atmos.Ocean. Technol., 4, 613-626.

Jakob, C., and S. A. Klein, 1999: The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ECMWF model. Q.J.R.Meteorol.Soc., 125, 941-965.

Annex 5: VAREPS (VAriable Resolution EPS)

3 The impact of resolution and membership on the EPS

The ECMWF Ensemble Prediction System (EPS) based on a finite number of integrations of a version of the ECMWF non-linear model can be used to estimate the probability density function (PDF) of forecast states. The cost of the EPS depends mostly on the resolution of the non-linear model and on the ensemble membership (Fig A5.1). For example, an EPS with 11 members with a spectral truncation uses approximately the same CPU time as a 101-member T_L159L40 EPS.

The performance of the EPS depends on both resolution and membership. High resolution is needed to be able to resolve well physical features in the forecast, whilst large membership is required to be able to resolve well the forecast PDF. The need to disseminate forecasts in a timely fashion, limits the total amount of CPU time that can be devoted operationally to the EPS. Once this limit has been set, the definition of the optimal combination of membership and resolution is not trivial and immediate, since different users would consider different combinations as optimal. At ECMWF, the configuration of the operational EPS run daily has been defined to try to satisfy a broad range of users by considering a large and comprehensive set of accuracy measures (Buizza 1997, Talagrand et al 1999, Buizza et al 1999).

The membership of the ECMWF operational EPS was set to 32 members at the start, in December 1992 (Molteni et al 1996) and was then increased to 51 members in December 1996 (Buizza et al 1998). Since the beginning, CPU time availability has forced the EPS to use a resolution lower than the resolution of the analysis and of the high-resolution deterministic forecast (Table A5.1). Note that the EPS has always been using a constant resolution throughout the whole 10-day integration period.

The fact that higher resolution induces more accurate EPS forecasts was documented in Buizza et al (1998 and 2001). This was the reason of the two EPS upgrades of 11 December 1996 (from 33 members at T63L19 to 51 members at T_L159L31) and of 21 November 2000 (from T_L159L40 to T_L255L40). Buizza et al (1998) detected also a positive effect of a larger membership, and this lead to the decision to increase the ensemble size from 33 to 51 in December 1996. Further evidence of the impact of the ensemble size was reported by Buizza and Palmer (1998) who compared ensembles with 4, 8, 16 and 32 member at T63L19 resolution. Mullen & Buizza (2001) concluded that ensemble size is very important when predicting severe and rare flooding events over the US. Mullen & Buizza (2001) concluded that having a large ensemble size is a very important ingredient to be able to predict the probability of rare events.

4 VAREPS: A possible EPS configuration?

One of the weaknesses of the current EPS is that it does not use the ECMWF analysis at its full accuracy. The EPS initial conditions are generated by interpolating the T_L511L60 analysis to the EPS T_L255L40 resolution, thus loosing details that are predictable (and may have a predictable effect on larger scales) during the first few days of time integration.

The proposed VAriable Resolution EPS (VAREPS) would use the same resolution as the analysis (i.e. T_L511L40) during the first few days of time integration until a forecast time T_{TRUNC}, and then the current EPS resolution (T_L255L40). This would move the interpolation from time 0 to time T_{TRUNC} when predictability is lost in the small-scales. In terms of CPU cost, a VAREPS single integration costs less than a T_L511L40 integration and thus more members can be run at the same cost (Table A5.2). For example, a single VAREPS2 forecast, with truncation applied at forecast day 2 (i.e. T_L511L40 up to day 2 and T_L255L40 from day 2 to day 10) needs approximately 2.6 time the CPU time required to run a T_L255L40 10-d forecast.

There are two possible benefits of a VAREPS:

- the analysis is used at its full resolution and truncation is postponed into the forecast to the time when predictability is lost at the smallest scales;
- the cost of running a single forecast is reduced and thus membership can be set to higher values.

Before testing the VAREPS, single deterministic experiments should be performed to assess the impact of applying a truncation at time TTRUNC.

5 Impact of truncation on single deterministic forecasts

The performance of a set of 10-d VAriable Resolution (VAR) single deterministic forecasts run with truncation applied at forecast day 2, 3 and 5 (Table A5.3) has been compared with the performance of 10-d T_L255 and T_L511 forecasts. All these forecasts have been run with 40 vertical levels (L40), starting from the same T_L511L60 analysis interpolated as required, and with the same model cycle for 30 summer (August 1999) and 30 winter (December 1999) cases.

The accuracy of the forecasts have been measured by computing the anomaly correlation and the root-mean-square error over the Northern Hemisphere and Europe for the prediction of the 500 hPa geopotential height field, and the root-mean-square error of the prediction of the 850hPa vector-wind over the Tropics.

Denote by sc(CONF,d) the monthly average score 'sc' of the day-d forecast run in configuration CONF, where CONF is one of the configurations listed in Table A5.3. Define the Relative Improvement index RI(CONF,d) as follows:

$$RI(CONF,d) = \frac{sc(CONF,d) - sc(T_L 225,d)}{sc(T_L 225,d-1) - sc(T_L 225,d)}$$

The Relative Improvement index RI can be considered as a skill score defined using as reference the score of a T_L255 forecast issued 1 day before. For each forecast-day d the index RI(CONF,d) gives the difference in skill of the day-d forecast given by configurations CONF and T_L255, normalized by the difference in skill of the day-d and the day-(d-1) T_L255 forecasts. In other words, RI(CONF,d) gives the gain in predictability compared to a 1-d gain in predictability at truncation T_L255.

For example, results indicate that RI(VAR3,3)=0.34 for sc='anomaly correlation for the 500hPa geopotential height over NH during winter'. This means that the VAR3 forecast has an anomaly correlation score

Fig A5.2 shows the winter-average Relative Improvement index RI for configurations T_L511, VAR5, VAR3 and VAR2 at forecast day 3, 5 and 7 computed for 5 different scores:

- root-mean-square error for the prediction of the 500hPa geopotential height over the NH (first group of bars);

- root-mean-square error for the prediction of the 500hPa geopotential height over the Europe (second group of bars);
- anomaly correlation for the prediction of the 500hPa geopotential height over the NH (third group of bars);
- anomaly correlation for the prediction of the 500hPa geopotential height over the Europe (fourth group of bars);
- root-mean-square error for the prediction of the 850hPa vector wind height over the Tropics (last group of bars).

Results indicate that all forecast configurations have a higher skill than T_L255. Considering, for example, the anomaly correlation of the day-5 forecast for the 500hPa geopotential height (Fig A5.2, middle panel, fourth group of bars) results indicate a 0.30 RI for all configurations. Note that for all forecast steps and all scores, the RI values are similar for all configurations.

Fig A5.3 is the equivalent of Fig A5.2 but for the 30-winter cases. Again, results are always positive for all configurations and depend only weakly on truncation.

The results shown in Fig A5.3 and A5.4 suggest that applying a truncation at forecast day 2 or 3 has a weak effect on the skill of a forecast. In other words, the benefit of running from a high-resolution analysis (T_L511) is maintained even if a forecast started at T_L511 resolution is truncated during the time integration.

6 Future plan

The results discussed above suggest that VAREPS could be a viable configuration. The plan is to compare the skill of the operational 51*$T_L511L40$ EPS with a 21*VAREPS2 ensemble run with resolution $T_L511L40$ up to forecast day 2 and than with resolution $T_L255L40$. The CPU time required to run this latter configuration is about 5% larger than the cost of the operational EPS. It will be studied whether such a system is more capable to provide severe weather warnings than the current operational ensemble, especially for very difficult situations as the storms that crossed Europe during December 1999 (Buizza & Hollingsworth 2001).

References

Buizza, R., 1997: Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF Ensemble Prediction System. Mon. Wea. Rev., 125, 1, 99-119.

Buizza, R., & Palmer, T. N., 1998: Impact of ensemble size on the skill and the potential skill of an ensemble prediction system. Mon. Wea. Rev., 126, 9, 2503-2518.

Buizza, R., & Hollingsworth, A., 2001: Storm prediction over Europe using the ECMWF Ensemble Prediction System. Met. Appl., submitted.

Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., & Wedi, N., 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. Q. J. R. Meteorol. Soc., 24, 550, 1935-1960.

Buizza, R., Barkmeijer, J., Palmer, T. N., & Richardson, D., 1999: Current Status and Future Developments of the ECMWF Ensemble Prediction System. Meteorol. Appl., 6, 1-14.

Buizza, R., Richardson, D., & Palmer, T. N., 2001: The new 80-km high-resolution ECMWF EPS. ECMWF Newsletter n. 90, 2-9, ECMWF, Shinfield Park, Reading RG2-9AX, UK.

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

Mullen, S., & Buizza, R., 2001: The Impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. Mon. Wea. Rev., submitted.

Talagrand, O., Vautard, R., & Strauss, B., 1999: Evaluation of probabilistic prediction systems. Proceedings of the ECMWF Workshop on Predictability, ECMWF, Reading, 20-22 October 1997, 1-26.

Period	Analysis Resol	HRES FC Resol	EPS	
			Resol	Size
19 Dec 1992 – 10 Dec 1996	T213L31	T213L31	T63L19	33
11 Dec 1996 – 30 Mar 1998	T213L31	T213L31	TL159L31	51
1 Apr 1998 – 12 Oct 1999	T_L319L31	T_L319L31	T_L159L31	51
13 Oct 1999 – 20 Nov 2000	T_L319L60	T_L319L60	T_L159L40	51
21 Nov 2000 - now	T_L511L60	T_L511L60	T_L255L40	51

TableA5. 1. List of changes of the resolution on the ECMWF analysis and high-resolution deterministic forecast, and list of changes of the resolution and ensemble size of the EPS.

Configuration	Normalized Cost of a single 10d forecast	Ensemble Size	Total cost of ensemble
TL255	1.00	51	51.00
TL511	9.04	6	54.24
VAREPS5	5.02	11	55.19
VAREPS3	3.41	17	57.98
VAREPS2	2.61	21	54.75

Table A4.2. Column 1: list of configurations. Column 2: CPU time of a single 10-d forecast run with horizontal resolution TL255, TL511, and with resolution TL511 up to forecast day 5 (VAREPS5), 3 (VAREPS3) and 2 (VAREPS2) (a unit cost is the cost of running a 10-d TL255 forecast). Column 3: ensemble size. Column 4: normalized cost of running an ensemble with the size specified in column 3. Values refer to forecasts run with 40 vertical levels.

Config	Forecast day									
	1	2	3	4	5	6	7	8	9	10
T511										
VAR5										
VAR3										
VAR2										
T255										

Table A5.3. Schematic of the VAR configurations used to run single deterministic forecasts.

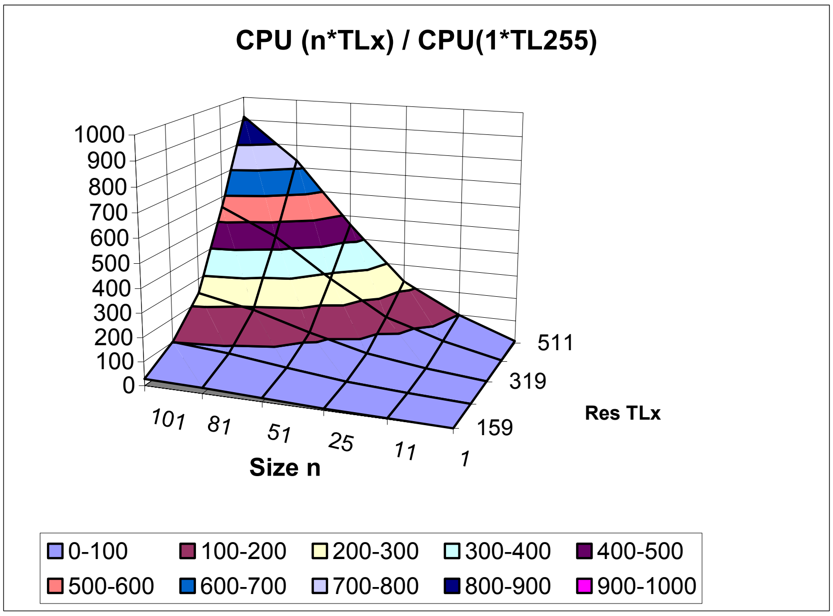


Fig A5.1. Normalized CPU time cost of the ECMWF EPS as a function of ensemble size and the horizontal resolution (a unit cost is the cost of running a single 10-d TL255 forecast). Values refer to forecast run with 40 vertical levels.

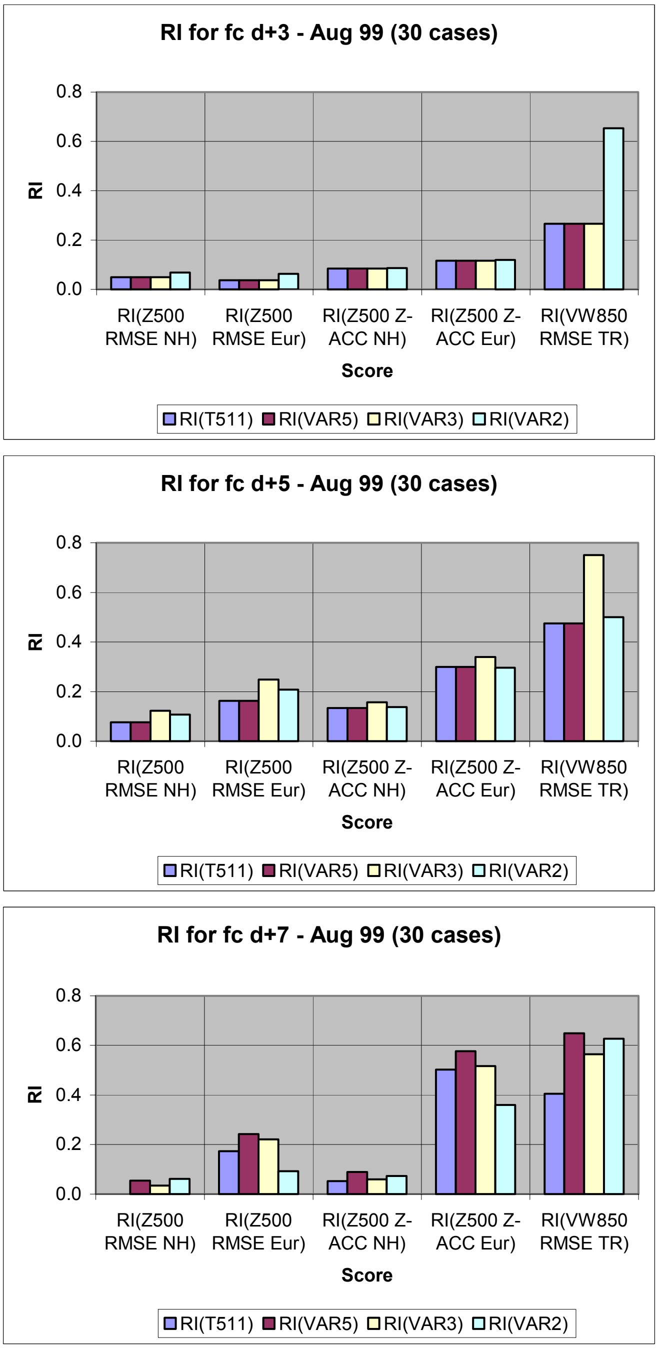


Fig A5.2. Winter-average relative improvement index RI at forecast day 3 (top), 5 (middle) and 7 (bottom) computed for 5 different scores: RMSE for the 500hPa geopotential height forecast over NH (first group of bars) and Europe (second group of bars), ACC of the 500hPa geopotential height forecast over NH (third group of bars) and Europe (fourth group of bars) and for the RMSE of the vector-wind over the Tropics (last group of bars). Color bars refer to forecast in configuration TL511 (first bar, violet), VAR5 (second bar, red), VAR3 (third bar, white) and VAR2 (fourth bar, cyan).

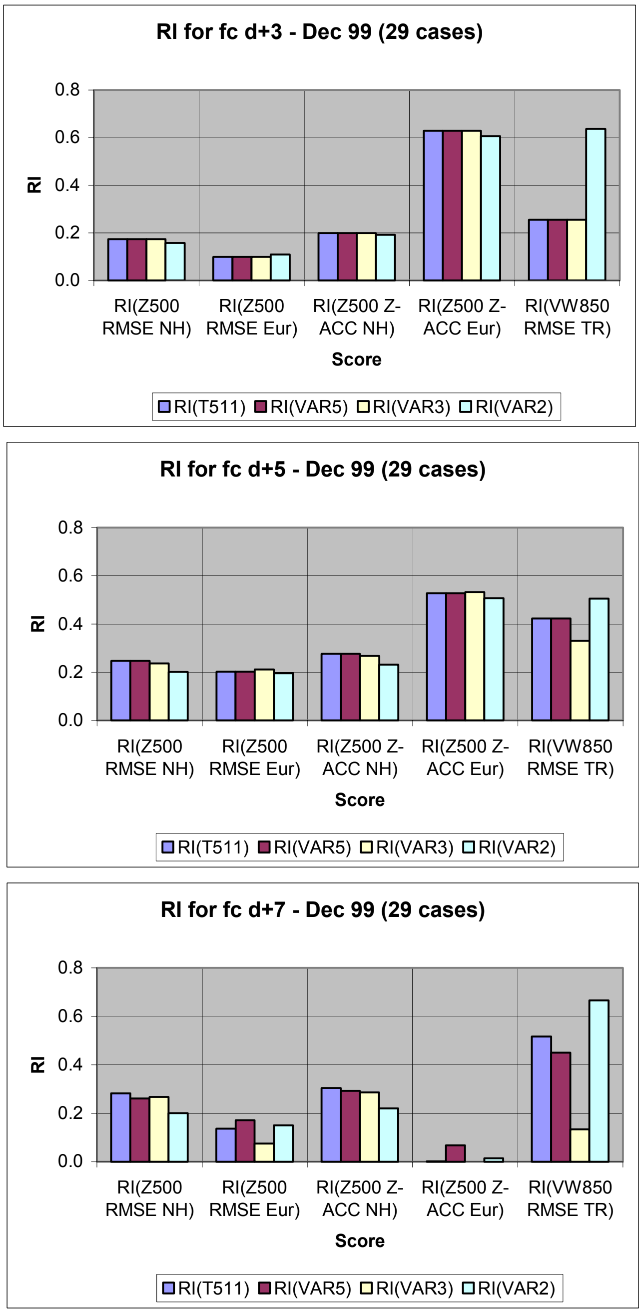


Fig A5.3. As Fig A5.2 but for summer-average results.

Annex 6: LEPS, the Limited area model Ensemble Prediction System at ARPA-SMR

The Ensemble approach allows to associate a probability of occurrence to meteorological events, providing further scope to quantitative precipitation forecasting. Current operational implementations of the Ensemble Prediction technique are currently based upon GCMs essentially covering the global scale. In the limited area model environment, it appears to be difficult to produce perturbations of the initial conditions which can efficiently grow for time ranges longer than 12--24 hours, depending on the size of the integration domain, possibly due to the driving/damping effects of the lateral boundary conditions. The natural approach to regionalize and adapt the global-scale EPS on the local scale should be to nest limited-area models in each EPS member. The obvious drawback of this procedure is connected to the large amount of computer resources required and to the intense data-flow necessary if the LAM integrations are not performed in the same centre where the EPS is produced.

At ARPA-SMR, the dynamical adaptation of the ECMWF EPS on the local scale through Limited Area Model nesting, is founded on an ensemble reduction technique where only few members of the entire global ensemble are selected to drive LAM integrations (Molteni et al., 2001; Marsigli et al., 2001). The reduction procedure is carried on by performing, on a restricted area (53-35N; 5W-25E), an independent cluster analysis of the 51 EPS members by a complete linkage method, where clustering is based on dynamic and thermodynamic fields in the lower-to-middle troposphere. The number of clusters is fixed to 5. A Representative Member (RM) for each of the 5 clusters is then defined by selecting the cluster member with the minimum distance from the other members of the same cluster and the maximum distance from all the remaining EPS members. These 5 RMs provide initial and boundary conditions for 5 LAM integrations up to three-to-five days ahead. The 5 individual LAM integrations are then used to compute a-priory probability of occurrence of meteorological events of interest, e.g. the exceeding of a given accumulated precipitation threshold, by combining them with weights proportional to the population of the cluster they represent.

Another practically important feature of the ARPA-SMR methodology is the use of the concept of ‘super-ensemble’. Rather than using only one ECMWF EPS set, more (up to three) consecutive daily EPS sets, progressively lagged in time, are used, providing initial sets of up to 153 individual members. From preliminary evaluations, the use of the super-ensemble technique increases the reliability of the computed a-priory probability of occurrence of the predicted event (surpassing of a precipitation threshold). The limited area model employed so far is LAMBO (Limited Area Model BOlogna), the hydrostatic limited-area model in operational use at ARPA-SMR. LAMBO is based on the NCEP ETA model and has an operational horizontal resolution of 10 or 20 km with 32 vertical levels.

The LEPS technique is under evaluation on a set of individual case studies of severe precipitation events in Europe. A systematic statistical verification of LEPS performance is also ongoing for the period Sept--Oct--Nov 1999, during which special observational data-set are available in correspondence of the Special Observing Period of the MAP Programme. For this period, LEPS is integrated using the Targeted Ensemble Prediction System, developed at ECMWF mainly by KNMI (Hersbach et al., 2000) with the aim to enhance the ensemble spread over Europe in the short and early medium range.

In the following figures, an example of the results of the LEPS application is shown in the case of the Soverato flood event (September 2000), which occurred in the Calabria region (the ‘tip’ of the ‘Italian boot'). Precipitation exceeding 340 mm/day was observed in the land areas of the Calabria region, this causing landslides, great disruption and losses of life.

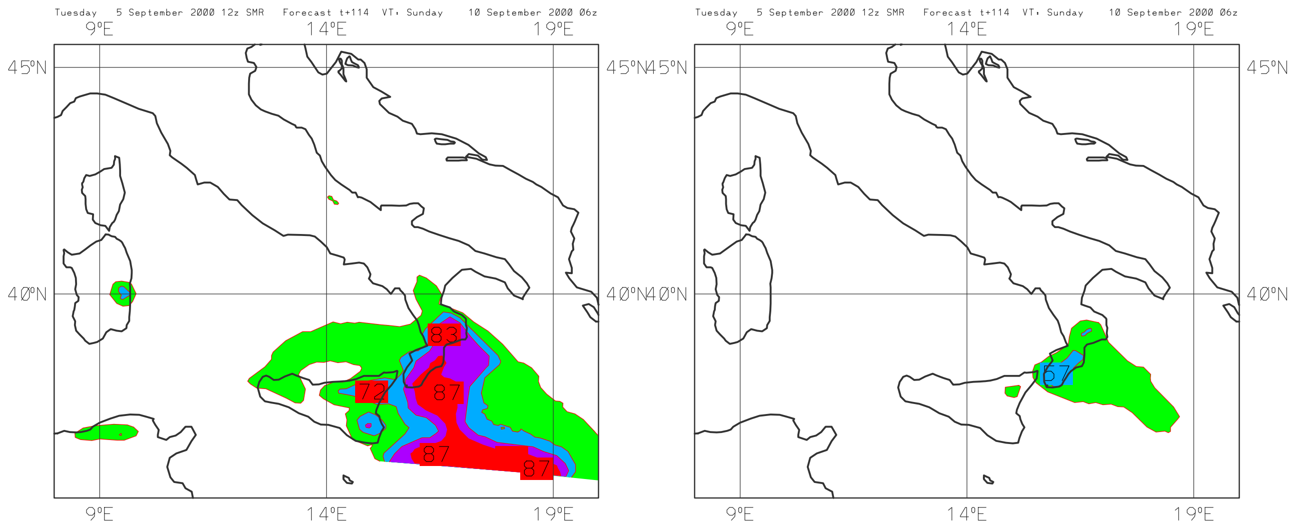


Fig A6.1: Probability maps (valid at 00UTC 10 September 2000; 60 hour forecast) for 24h accumulated rainfall exceeding 20 and 100 mm (left and right panel, respectively) as predicted by the super-ensemble. Contours every 20%.

Fig.1 shows the rainfall probability maps (exceeding the 20 and 100 mm/day thresholds) generated by the super--ensemble methodology 60 hours before the event.Both localisation and intensity of the flood are properly captured, despite the fact that these maps are obtained by combining only 5 (weighted) LAMBO integrations. On the other hand, the two panels of Fig.2 show the results obtained when LAMBO is nested on all EPS members (the so-called 'brute-force approach'). It can be noticed how the land regions highlighted as the most likely to be subjected to heavy rain are essentially the same in both figures. It could be said that both methods provide approximately similar results (Montani et al., 2001). Hence, little information appears to be lost when the super--ensemble LEPS methodology is applied to reduce the ensemble size and extract the crucial information about the weather evolution scenarios highlighted in the ensemble.

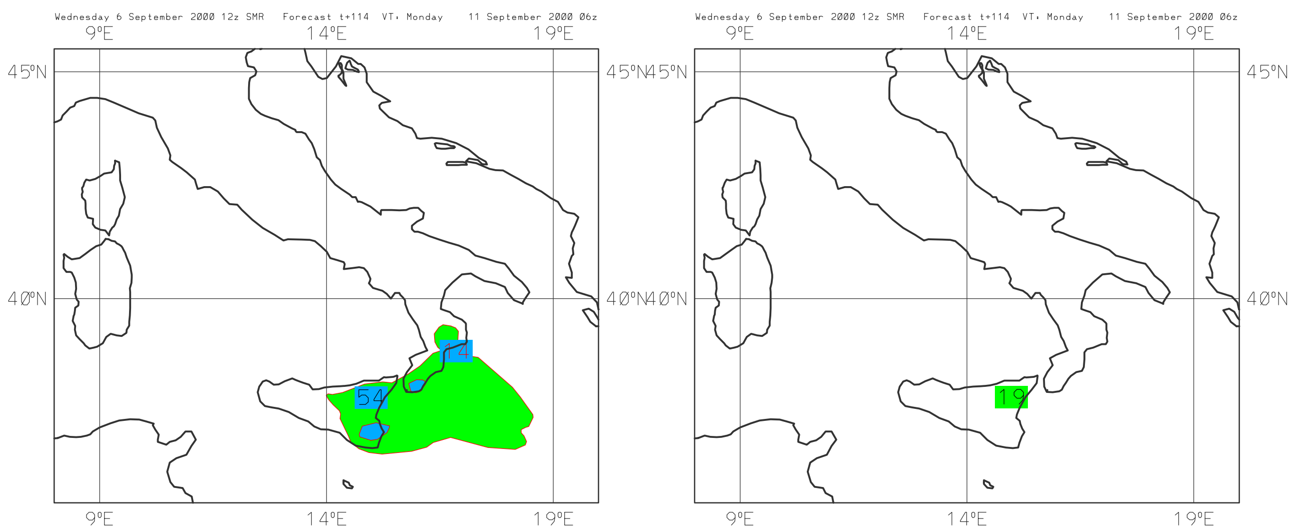


Fig.A6. 2: The same as Fig.1 but for LAMBO nested on all EPS members

The LEPS technique is still currently being tested and developed. The first major change envisaged will be on the type of LAM employed, since the non-hydrostatic Lokal Modell replaces LAMBO. The name of the project will also reflect this change, from LEPS to COSMO-LEPS, to officially frame this project in the COSMO (COnsortium for Small-scale MOdelling) project. During 2001-2002, COSMO-LEPS will be tested on a regular/quasi--operational basis, to allow a comprehensive verification effort, based both on subjective and objective criteria.

Acknowledgments

The fundamental contributions of Chiara Marsigli, Andrea Montani and Fabrizio Nerozzi to the entire project are gratefully acknowledged.

References

Hersbach, H., Mureau, R., Opsteegh, J. D. and Barkmeijer, J., 2000: A short--range to early--medium--range Ensemble Prediction System for the Europen Area. *Mon.Wea.Rev*, **128**, 3501--3519.

Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S., Molteni, F. and Buizza, R., 2001: A strategy for High--Resolution Ensemble Prediction. Part II: Limited-area experiments in four Alpine flood events. *Q.J.R.Meteorol.Soc*., in press.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T., 2001: A strategy for High-Resolution Ensemble Prediction. Part I: Definition of Representative Members and Global Model Experiments. *Q.J.R.Meteorol.Soc*., in press.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Buizza, R. and Tibaldi T., 2001: Performance of global and limited--area ensemble forecasts for a flash flood case. *Q.J.R.Meteorol.Soc*., to be submitted.