Tellus (2001), 53A, 35–55
Printed in UK. All rights reserved

TELLUS
ISSN 0280–6495

Targeted ensemble prediction for northern Europe and parts of the north Atlantic Ocean

By INGER-LISE FROGNER* and TROND IVERSEN, *Department of Geophysics, University of Oslo, PO Box 1022 Blindern, N-0315, Oslo, Norway*

(Manuscript received 27 August 1999; in final form 10 July 2000)

ABSTRACT

The targeting procedure developed at ECMWF is used to make ensembles specially designed for northern Europe and parts of the north Atlantic Ocean. A total of 35 ensembles are integrated, consisting of 20 winter cases and 15 summer cases in 1997, each consisting of 20 members plus one control forecast. The ensembles are run up to day 10, and the ensemble spread inside the target area continues to increase all through the 10 days. Two distinct regimes of increase can be found, the first increase is consistent with the perturbations moving in and through the target area, it is hypothesised that the latter increase in ensemble spread around forecast day 5–7 is connected with increasing non-linearity. The performance of the experimental ensembles is compared to the operational ensemble prediction system (EPS) at ECMWF, both with all 50 members and with only 20 members. The spread increases when the number of members in the ensemble prediction system is increased, and the spread increases inside the target area when targeting is applied. We find that the increase in spread when going from EPS with 50 members to the targeted ensembles is larger than when going from 20 to 50 ensemble members of the operational sets. Clearly targeting must be an option when predicting for a sub-domain of the hemisphere. Looking at other measures, such as the Brier skill score (BSS), relative operating characteristic (ROC) curves and cost/loss analyses, the impact of the targeting is modest for the winter cases, but the impact for the summer cases is evident. For the winter cases a large part of the operational perturbations were located in the same area as the targeted perturbations, and the differences for the two sets are small. For the summer cases the operational perturbations were mostly split between two locations and hence the targeting will give results differing more from the operational.

1. Introduction

Weather forecasting is mathematically a combined initial-boundary value problem for which the forecast quality is critically dependent on the initial conditions (Lorenz, 1963). Although there are clear signatures of periodic nature, such as diurnal and seasonal variability, the transient behaviour is generally dominating. To forecast the weather with skill better than can be obtained by using climatological statistics, one has to know the initial conditions accurately.

The quality of a given numerical prognosis depends on two basic circumstances; how well the numerical model approximates the real atmospheric processes, and how accurately an analysis can be made. The relative importance of these is determined by the growth rates of typical errors in the analyses, and how well the model represents the atmospheric instabilities responsible for these error growth rates. Contemporary numerical weather prediction (NWP) models are believed to describe macro-scale instabilities sufficiently well to produce the correct growth rates for those

* Corresponding author.
e-mail: frogner@geofysikk.uio.no

scales. (The macro-scale motion systems have horizontal radii larger than or of the order of 1000 km) The errors in the macro-scale analyses are statistically much smaller than any climatological data would yield. The useful time range for macro-scale weather forecasting can be defined as the time required for errors present in the analyses to grow to the level defined by climatological fields. This is defined as the saturation level for errors.

Ideally, NWP models should predict probability distributions for each meteorological quantity in each grid-point based on estimates of the initial probability functions for the analysis. Furthermore, model errors should be accounted for by developing stochastic differential equations where probability distributions for uncertain coefficients and terms are used. Due to the very many degrees of freedom this is not possible to achieve in the foreseeable future. Instead, based on the idea first proposed by Lorenz (1965) and by Leith (1974), Monte-Carlo forecasts with a limited number of carefully selected ensemble members are being made. Recently, Buizza et al. (1999a) also perturbed uncertain terms in the model physics. Since there only is sufficient computer power available for a limited set of ensemble members, the ensemble members should be based on initial perturbations of the analysis with non-zero projections on leading Lyapunov vectors (LLVs) (Legras and Vautard, 1996). Perturbations should also have amplitudes comparable to analysis errors. If these two requirements are fulfilled, the ensemble mean should statistically give a better forecast than the control forecast (Leith, 1974) and the spread between the ensemble members should give a qualified estimate of the predictability of the day, and thus the confidence in the weather forecast. It is important to realise, though, that large spread does not necessarily imply large error in the control forecast (Buizza and Palmer, 1998). Whitaker and Loughe (1998) found that spread as a predictor of skill is likely to be most useful when the spread is either very large or very small.

Different macroscale ensemble prediction methods are presently in use with global models at the European Centre for Medium-Range Weather Forecasts (ECMWF) and National Centre for Environmental Prediction (NCEP). The method in use at NCEP (Toth and Kalnay, 1993, 1997) is called breeding of growing modes (BGM). It uses the full governing equations, but it is uncertain if the method selects the most unstable modes.

The method in use at ECMWF (Buizza et al., 1993; Molteni et al., 1996) was in principle proposed by Lorenz (1965). The linear tangent model and its adjoint are used to compute singular vectors (SVs). Initial perturbations are linear combinations of SVs that grow fastest over a specified period in time. Only limited physics is included due to the difficulty in constructing adjoint operators to physical processes.

Both methods have shown to be useful and have been used at the two centres for several years. Legras and Vautard (1996) discuss the difference and similarities of SVs and BGMs. A comparison of Lyapunov vectors and singular vectors is given by Szunyogh et al. (1997) and by Buizza and Palmer (1995). Palmer (1996) gives a general overview of the concept of predictability.

The ECMWF method is applied in this paper. An advantage with the SVs is that they can be calculated in such a way that the perturbations will be confined to a specific geographical area for a specific forecast period by using a local projection operator, this is called targeting. The way targeting is used in the ECMWF model is described in Buizza (1994). In this paper the local projection operator is used on an area covering northern Europe and parts of the north Atlantic Ocean.

The aim of this paper is to see if constraining the perturbations to a specific area will increase the ensemble forecast quality compared to using perturbations constrained to the northern hemisphere. We have set up a targeted ensemble prediction system (TEPS) for northern Europe and parts of the north Atlantic Ocean. It is compared to the operational ensemble prediction system (EPS) at ECMWF. In this paper, we are only perturbing initial conditions. The perturbations computed for the large, operational EPS area are probably split between several locations. The perturbations computed for TEPS are constrained to the much smaller target area. Therefore, fewer perturbations could give as much information in this small target area as more perturbations in the larger area. We try out using only 20 ensemble members as opposed to the 50 members for the operational EPS. TEPS then consists of 21 forecasts, one "control" forecast (started from the operational, unperturbed analysis), and 20 forecasts with perturbed analyses. The methodology used to define these perturbations is described in more detail in Molteni et al. (1996). The quality of the two ensemble systems is in this paper measured by the ordinary root-mean-square-error, Brier skill scores (BSS), relative operating characteristic (ROC) curves and cost/loss analyses. A priori we would expect TEPS to be more capable of predicting extreme weather events, since a larger number of perturbations will be located in the area of interest. This should give a wider range of developments inside the target area than the ordinary EPS. A similar study to ours, but with a longer optimisation time interval and larger target area, was performed by Hersbach et al. (2000).

In Section 2 a short review of singular vectors as they are used at ECMWF is presented. The set-up for the experiments is presented in Section 3. In Section 4, the comparison between the operational EPS and the experimental TEPS can be found, and in Section 5, conclusions and plans for further work are presented.

2. Theory

The mathematical concept of targeted singular vectors is here shortly reviewed. For a more detailed description see Buizza et al. (1993), Buizza (1994) or Palmer (1996).

If in the phase space of a dynamical system, the state vector is denoted by $\boldsymbol{X}$, then the non-linear system of evolution equations can be written as:

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{M}(\boldsymbol{X}), \tag{2.1}$$

where $\boldsymbol{M}$ is an operator which may depend non-linearly on $\boldsymbol{X}$. The time evolution of a small perturbation $\boldsymbol{x}$ of the state vector at time t_0 is assumed to be described by the linear approximation:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{M}_l \boldsymbol{x}, \tag{2.2}$$

where $\boldsymbol{M}_l$ is the Jacobian matrix of $\boldsymbol{M}$ with respect to $\boldsymbol{x}$. As long as some defined norm of $\boldsymbol{x}$ is sufficiently small, (2.2) describes the tangent linear model and $\boldsymbol{M}_l$ is the tangent operator relative to the trajectory $\boldsymbol{X}(t)$.

A forward tangent propagator $\boldsymbol{L}$ that maps small perturbations from the initial time t_0 to a future time t is introduced: $\boldsymbol{v}_i(t) = \boldsymbol{L}\boldsymbol{v}_i(t_0)$. $\boldsymbol{v}_i(t_0)$ are the eigenvectors of $\boldsymbol{L}^*\boldsymbol{L}$ with eigenvalues σ_i^2. $\boldsymbol{L}^*$ is the adjoint of $\boldsymbol{L}$ with respect to an inner product based on total energy.

The σ_i, ranked in terms of magnitude, are called the singular values of $\boldsymbol{L}$, and the vectors $\boldsymbol{v}_i$ its singular vectors. The vector $\boldsymbol{v}_1(t)$ defines the major axis of the forecast probability density function with respect to the chosen norm, $\boldsymbol{v}_2(t)$ the second major axis, and so on. At initial time these axes are given by $\boldsymbol{v}_1(t_0)$, $\boldsymbol{v}_2(t_0)$ respectively. The amplification from t_0 to t associated with these directions are given by the σ_i as long as linearity is approximately valid.

To select perturbations that yield the fastest growing structures in a specific area, a local projection operator is introduced in the equations. In the operational ensembles at ECMWF the area is the northern hemisphere from 30°N to the North Pole.

The method of singular vectors is mathematically appealing, since orthogonal vectors associated with the different linear growth-rates are explicitly found for the forecast optimisation interval. However, there are practical implementation problems. Firstly, it is difficult to construct adjoint operators to all physical processes. Secondly, the matrix $\boldsymbol{L}^*\boldsymbol{L}$ is enormous for regular NWP models. Therefore a highly truncated model (T42) with simplified physical processes is used for the generation of initial perturbations.

3. Set-up of experiments

3.1. Targeted singular vectors (TSVs)

To compute the targeted singular vectors, the local projection operator selects a much smaller target area than the northern hemisphere. The target area was chosen to focus northern Europe and parts of the north Atlantic Ocean (Fig. 1). The targeted singular vectors are calculated with the same models as the operational ECMWF SVs, with a horizontal resolution T42, 31 levels in the vertical, and an optimisation interval of 48 h. This enables a straightforward comparison between the two sets. Occasionally stratospheric modes occur, and to avoid these modes the upper 8 model levels

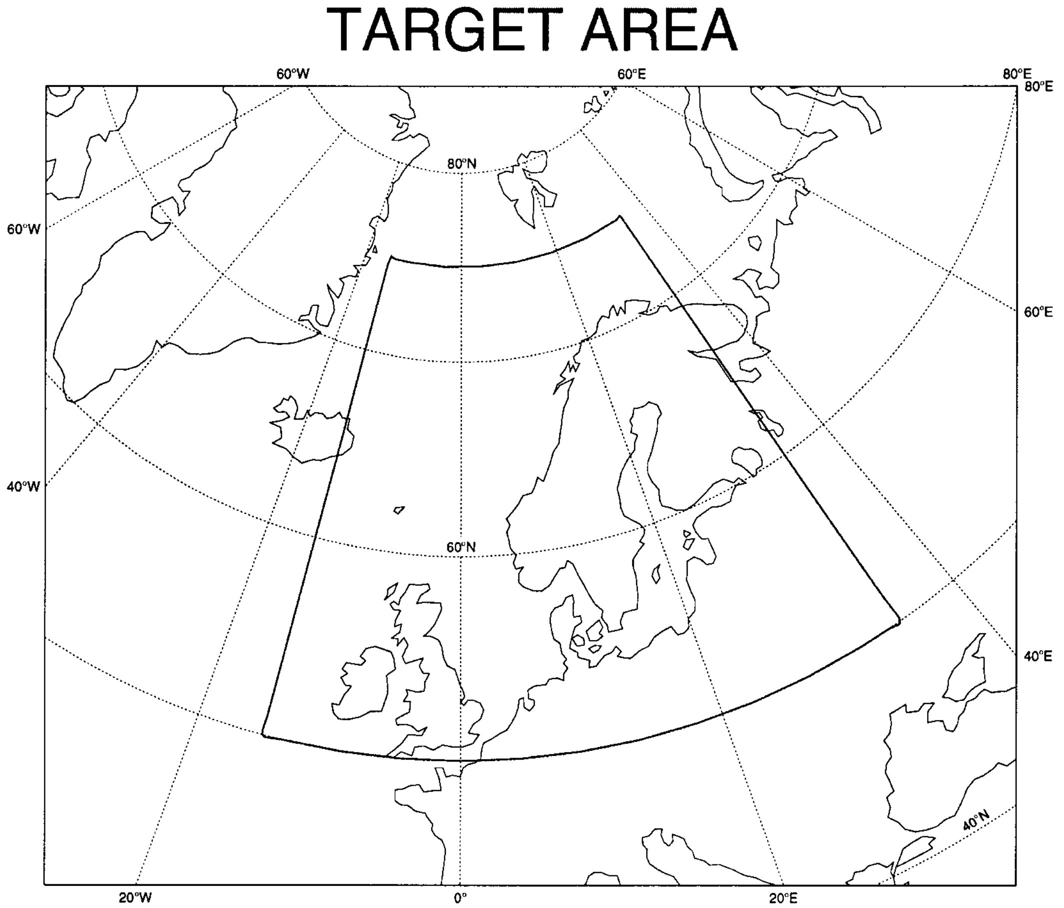

Fig. 1. Target area for experimental singular vectors (TSVs). The area is from 50.0 to 75.0°N, and from 15°W to 35.0°E.

were excluded in the target domain (Hartmann et al., 1996).

3.2. *Targeted ensembles (TEPS)*

The TSVs are used to make perturbations which are integrated to yield the ensembles. The model used was the $T_L159L32$, that is the same model as used for the operational ensembles. The version of the model that was operational on a particular day is used for the experimental runs for this same day. Therefore there are no model differences between the two sets. The operational perturbations were integrated up to day 10, and so are the TEPS. With such a small target area, the significant portions of the original perturbations will propagate out of the target area after a few days. For operational weather forecasting there is little reason to make targeted ensemble runs much longer than the optimisation period. After this period one would expect that some of the ensemble perturbations will decay and the ordinary EPS should perform better than TEPS. The reason for doing longer integrations is the possibility of non-linear amplification of the quasi-stationary perturbations that are left in the target area through interactions with geographically localised forcing.

3.3. *Experiment periods*

35 targeted ensembles were created, 20 winter cases and 15 summer cases. The winter period

chosen is 1–20 February 1997 and the summer period 1–15 June 1997.

The winter period includes a transition between weather regimes, and there was strong cyclonic activity from the west over Scandinavia. This was the FASTEX period (Joly et al., 1997). The period started with a ridge that moved eastwards and out of the target area on the 3rd day. From this date until 10 February the 500 hPa flow over northern Europe was almost zonal. At day 10 a cyclone entered from the north-west, and the target area was dominated by this low for the following 5 days, before a ridge re-emerged. Fields that describe the average over 5 days for the different weather regimes are presented in Fig. 2. The transition from a low-index flow pattern (Fig. 2a) to a high-index flow pattern (Fig. 2b) is clearly seen. In Fig. 2c we see a trough in the eastern part of the target area. The transition to a low-index pattern in the target area is evident in Fig. 2d.

With much activity in and around the target area, it turned out that a major part of the northern hemispheric singular vectors (NHSVs) were concentrated in this area already. The improvement of TEPS over EPS may then not be very large, therefore another period is also chosen. This period was chosen from the criterion that there should be comparable activities in the two major storm tracks in the northern hemisphere.

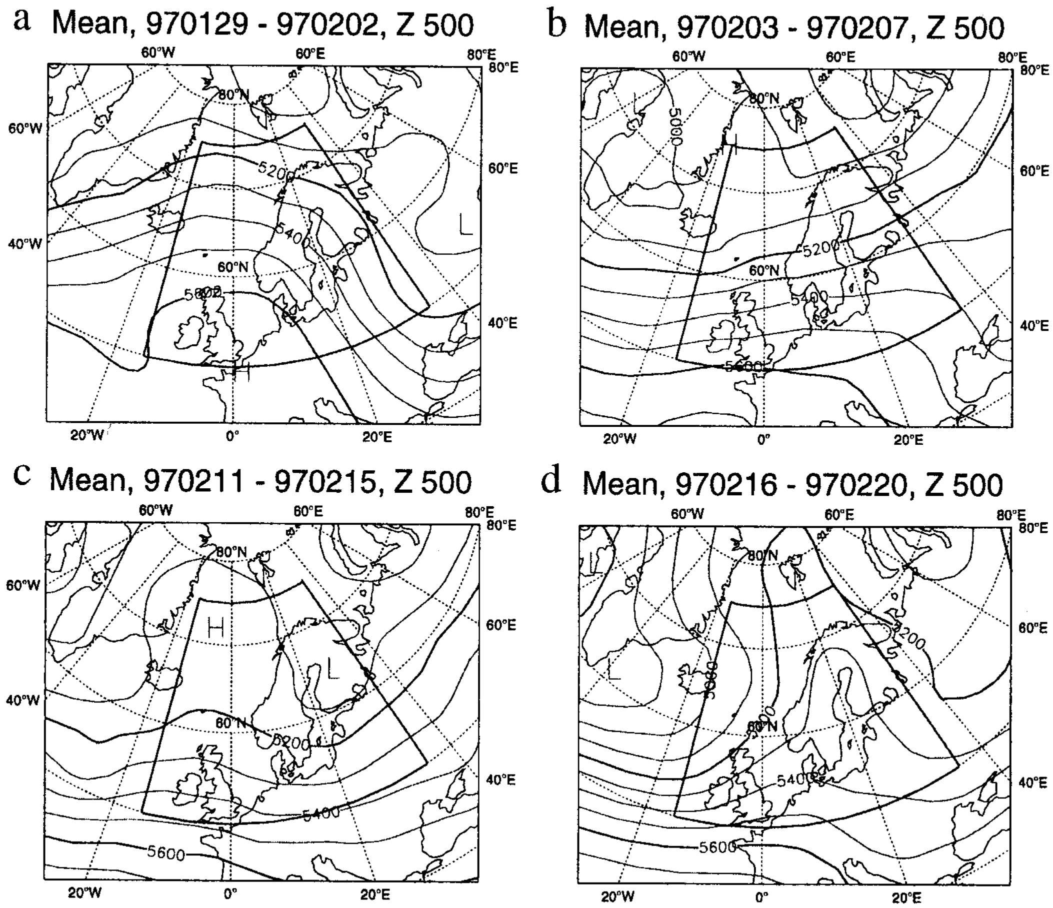

Fig. 2. Mean over 5 days in 500 hPa geopotential height, (a) from 29 January to 2 February, (b) from 3 February to 7 February, (c) from 11 February to 15 February and (d) from 16 February to 20 February. The contour interval is 100 m.

For the summer period chosen, the number of NHSVs in the two tracks is almost equal. This led to more TSVs than NHSVs in the target area (not shown), and should lead to an expected improvement of TEPS over EPS. The summer period started with a strong anticyclone in the Atlantic which slowly disappeared, and for the remaining period the area was dominated by a cyclone in the West moving slowly northwards. The average over 5 days for these two weather regimes are presented in Fig. 3.

3.4. *Verification methods*

The parameters used for verification and comparison are the 500 hPa geopotential height (Z 500), the two-meter temperature (2T), and the large scale precipitation (LSP). For 2T the ensembles were compared to the operational ECMWF analysis. For LSP such an analysis is not available. Instead, the difference between +24 h and +12 h from the operational T213 forecast started at 00 or 12 UTC is used. For the summer period the 00 UTC forecast was unfortunately not available, and instead the difference between +36 h and +24 h started from 12 UTC accumulated between noon and midnight is used.

The Brier skill score (BSS) (Brier, 1950, Stanski et al., 1989) and relative operating characteristic (ROC) curves (Mason, 1982) are measures used on probabilistic forecasts. Brier skill score expresses the fractional improvement over climatology. It has a range of $-\infty$ to 1 where 1 indicates a perfect forecast. ROC curves are curves where the *false alarm rate* (the percentage of forecasts of the event given that the event did not occur) is along the abscissa and the *hit rate* (the probability of detection of the event) is along the ordinate. The perfect forecast is represented by a curve that lies along the left-hand side of the graph up to the upper left corner, then along the upper side. The area under the ROC curve is a good measure of how close to these two lines the curve is. An area of one represents a perfect forecasting system and an area of 0.5 represents useless forecasts. For more details on the BSS and ROC measures, see Section 7.

Cost/loss analyses (Katz and Murphy, 1997; Richardson, 1998) are associated with a specific decision-making problem. It takes into account the actual benefit a user could obtain by using forecasts instead of, e.g., climatology when deciding whether to take precautions against a hazardous weather event. The economic value of the forecast system is computed as a function of the cost/loss ratio. A more detailed description of the cost/loss analyses can be found in the Section 7.

The climatological frequency of an event is needed in the calculation of BSS and in the cost/loss analyses. The climatological frequency used is determined separately for the winter and summer sets, and for 12 h and 00 h verification time. The analyses for all time steps within the sets were used. For the 20 winter cases this resulted

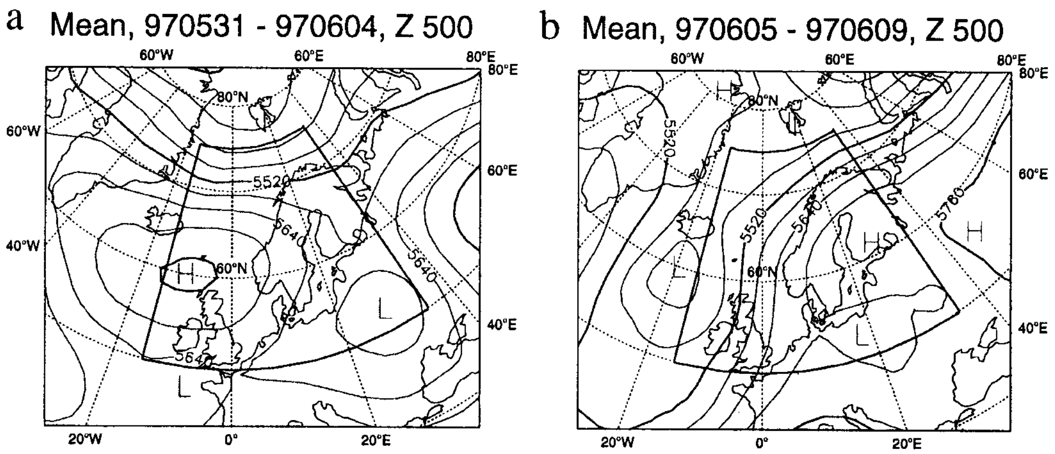

Fig. 3. Mean over 5 days in 500 hPa geopotential height, (a) from 31 May to 4 June and (b) from 5 June to 9 June. The contour interval is 60 m.

in climatology based on 29 days, when excluding overlapping dates. For the 15 summer cases, the climatology is based on 24 days.

4. Results and discussion

4.1. Ensemble spread

The ensemble spread of a parameter is defined as the average spread for all ensemble members over all points inside the target area:

$$s = \frac{1}{I} \sum_{i=1}^{I} \sqrt{\frac{1}{N} \sum_{j=1}^{N} (e_{ij} - c_i)^2},$$

where N is the number of ensemble members, I is the number of points inside the target area, e_{ij} is the parameter value for ensemble member number j in point number i, and c_i is its value for the control run in point number i. The constraint on the singular vectors to the target area should result in more unstable structures in this particular area than accounted for by the operational NHSVs. This leads to an expected higher spread for TEPS than for EPS, and an expected advantage of higher ability to detect extreme weather events. Extreme weather events mean rare, and not necessarily hazardous, events. In Fig. 4, the spread around the control inside the target area for the fifteen summer cases is shown for TEPS along with the operational EPS for only the first 20 members and for all 50 ensemble members (EPS 20 and EPS 50). The result for the twenty winter cases is similar (not shown). The spread for TEPS is larger than the spread for both 20 and 50 members of the operational EPS up to day 4. For the whole northern hemisphere, however, the spread for TEPS is, as expected, smaller than the spread for the operational EPS for all forecast times (not shown).

The largest difference in spread is obtained around forecast time +48 h, which is the optimisation time for the targeted singular vectors. After +48 h, the spread continues to increase, but at a slightly lower growth rate. Later in the forecast period the spread again starts to increase at a larger growth rate. This will be discussed in the next section. It is worth noting that the increase in spread for TEPS over EPS is much larger than the increase in spread for operational EPS when increasing from 20 to 50 members.

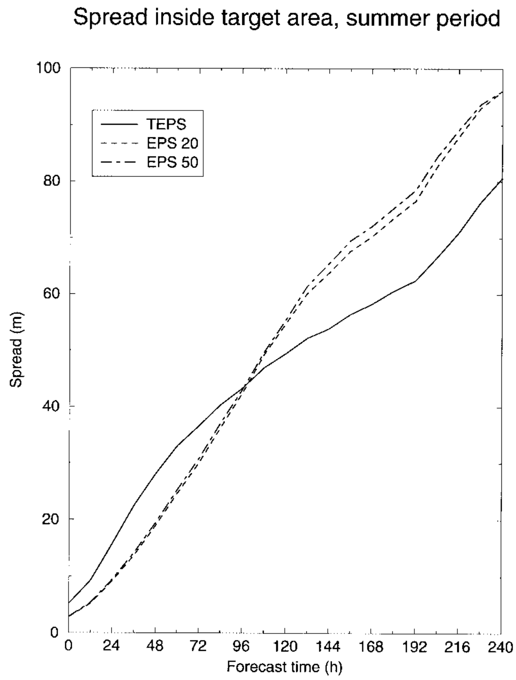

Fig. 4. 500 hPa geopotential height ensemble spread inside the target area for the summer period, as a function of forecast time.

The spread around the ensemble mean has also been calculated. This gives the same qualitative results (not shown) but with slightly smaller values.

Fig. 5 shows a typical spread diagram for one particular day, in this case 15 June 1997. As in Fig. 4 both TEPS, EPS 20 and EPS 50 are shown. Again; TEPS has a higher spread up to about forecast day 4. From a map view of the spread around the control for 15 June 1997 (not shown), it is seen that the spread for a particular forecast time is very similar for the three forecast systems, although the amplitudes are different. This is in accordance with Hartmann et al. (1995) who showed that a singular vector that grows in a specific area will be represented with the same structure whether the optimisation area is the northern hemisphere or just the part where the singular vector exists. Hersbach et al. (2000) also found that targeting acts as a selection and reordering of the NHSVs. The remainder of the TSVs are not contained in the NHSVs, but consists of slower growing error patterns. Also the vertical

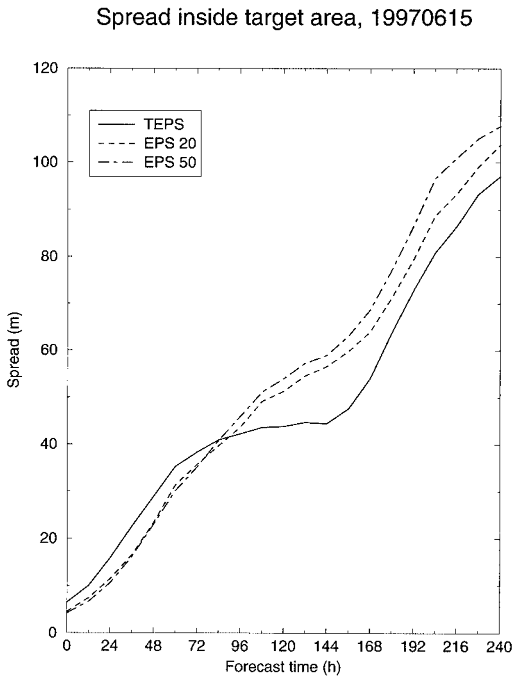

Fig. 5. As Fig. 4, but for only one summer case, 15 June 1997.

and horizontal distributions of total energy are similar for TSVs and NHSVs.

Fig. 6 displays the map view of the spread computed for all 20 ensemble members for TEPS for all 20 winter cases:

$$S = \sqrt{\frac{1}{ND} \sum_{d=1}^{D} \sum_{j=1}^{N} (e_{d_j} - c_d)^2},$$

where $D = 20$ is the number of days in the winter set, e_{d_j} is the parameter value for day d and ensemble member j and c_d is the control value for day d. All other variables are as described earlier. In the first panel (day 1) the spread clearly has its maximum in the western part of the map. Around the optimisation time (day 2) the maximum spread is localised inside the target area, with the lines of constant spread almost describing circles around it. For forecast day 5, the maximum is clearly shifted towards the east. Later in the forecast period the spread is more uniformly distributed over the whole displayed area (forecast day 9). The results for the summer period are similar (not shown).

It seems reasonable to conclude that the targeting satisfies the purpose of increasing the spread inside the target area at the optimisation time (48 h). Outside the target area, however, it is seen that the spread for TEPS generally has smaller values than the two operational ensembles. Clearly, the perturbations are more confined to the target area than the northern hemispheric perturbations. Fig. 5 showed that the integrated spread inside the target area for TEPS increased up to about forecast day 4 when it became almost constant for some days, while the spread for EPS continued to increase.

4.2. Increase in spread between forecast day 5 and 10

In Fig. 4 we saw that the spread does not decrease after the optimisation time of 48 h. The perturbations do not leave the target area completely. Instead the spread increases throughout the 10-day forecasts. Partly this somewhat surprising result may be due to non-linear effects, but it was also demonstrated by Simmons and Hoskins (1979) that upstream development occurs in baroclinic flows. After the first growing perturbation has passed, new developments occur continuously in the trail.

Fig. 5 shows a typical feature for all the 35 cases; between day 3 and day 6 there is a plateau for TEPS, and for some individual cases a weak local maximum. The timing of this plateau varies slightly from case to case, therefore the mean for the fifteen summer cases in Fig. 4 does not show this plateau clearly. For this single case, TEPS has a larger spread up to about forecast day 4. Between forecast day 3 and 6 the spread is almost constant, whilst shortly after there is a sudden increase. The spatial distribution of the spread is shown in Fig. 7. The spread starts to increase northwest of Scotland at day 6 and spreads eastwards into the centre of the target area during amplification. The two maxima west and east of northern Scandinavia also amplify and contribute to the increase in spread. During the remaining forecast period, the increase in spread is mainly due to amplification of a quasi-stationary pattern inside the target area. One could hypothesise that the cause of the increasing spread in the latter part of the forecast period is due to perturbations that have travelled around the globe, but this is not

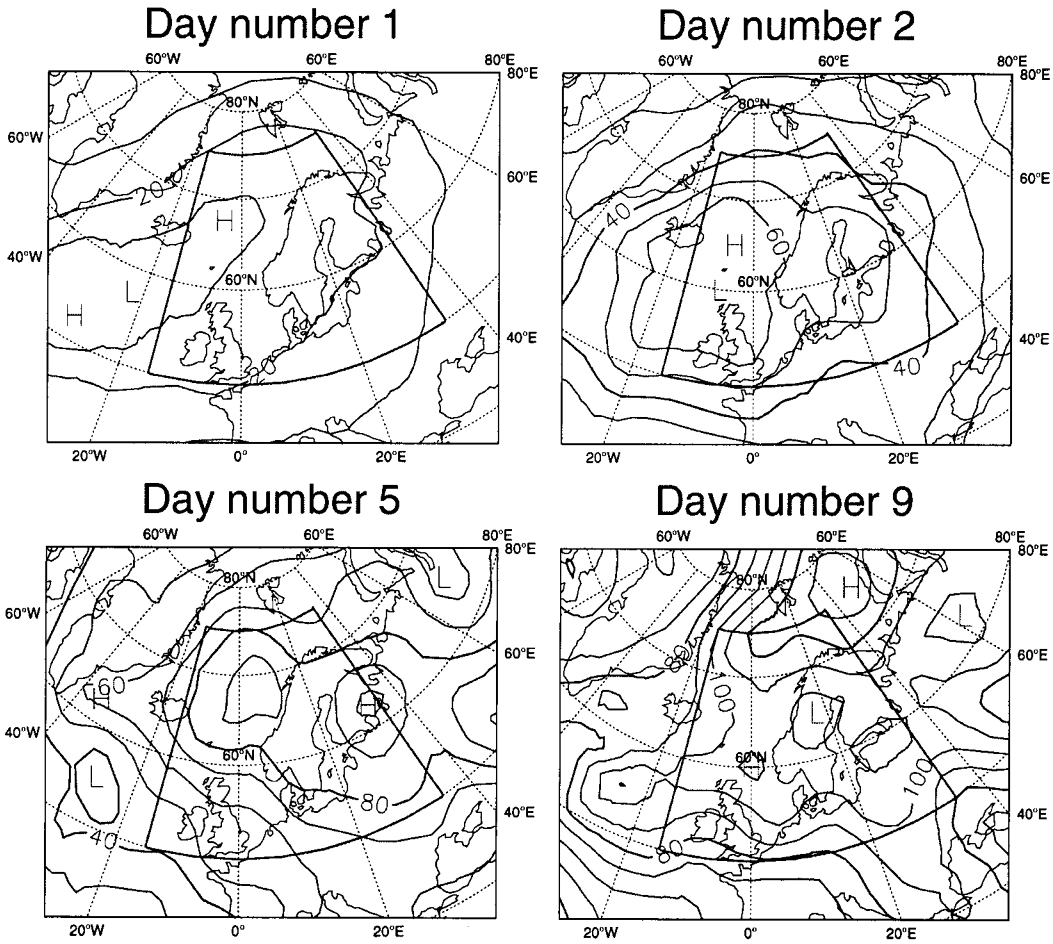

Fig. 6. Ensemble spread for TEPS for the winter period, in 500 hPa geopotential height. Displayed are forecast days 1, 2, 5 and 9. The contour interval is 10 m.

the case. The original perturbations can be followed, and evidently the time is much too short for the perturbations to complete one cycle.

Fig. 8 shows the spread, s, around the control for a few selected TEPS members for 15 June 1997. For a single member ($N = 1$), the expression for s reduces to the spatial mean of the absolute value of the difference between the ensemble member and the control. Shown are results for ensemble member number 3, 4, 19 and 20, i.e., two pairs of perturbations. These are amongst the members that experience the largest increase in spread over the last forecast days, but all except two members (member 2 and 15) show similar characteristics. All four members show a weak maximum before they experience a rapid growth in spread.

Since member number 3 and 4 project with exactly opposite signs onto the same combination of singular vectors at $t = 0$, they will add to zero as long as the development is linear. Fig. 9 shows the map view of a non-linearity index for ensemble members 3 and 4 for 15 June 1997, calculated in the following way:

$$I_{\mathrm{NL}}(i) = |(e_{i1} - c_i) + (e_{i2} - c_i)|,$$

where, in this case, e_{i1} is member number 3 and e_{i2} is member number 4. The index i indicates grid-point no. i inside the target area, and c_i is the control forecast in point i. I_{NL} is zero at forecast

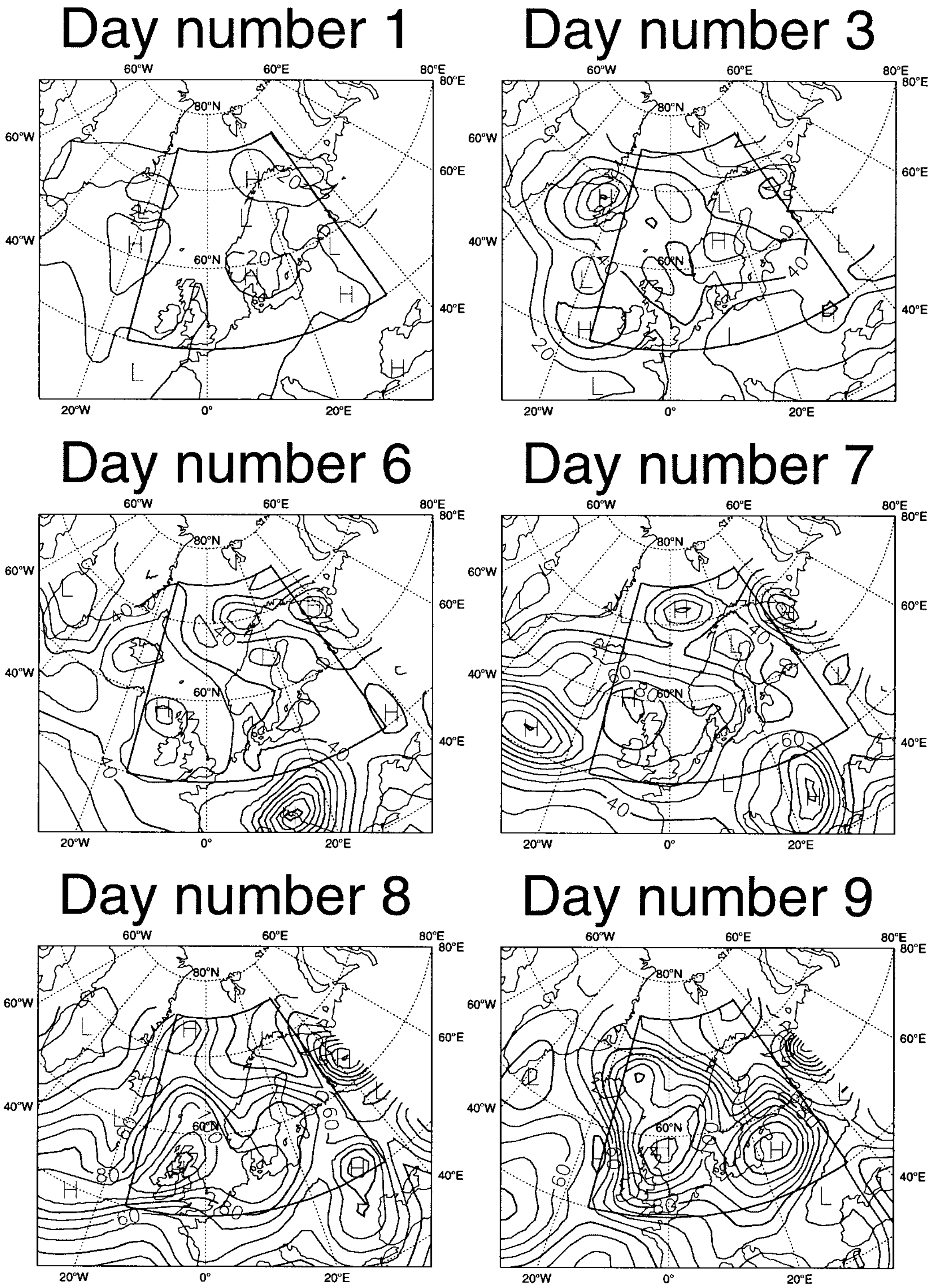

Fig. 7. Ensemble spread for TEPS for 15 June 1997, in 500 hPa geopotential height. Displayed are forecast day no. 1, 3, 6, 7, 8 and 9. The contour interval is 10 m.

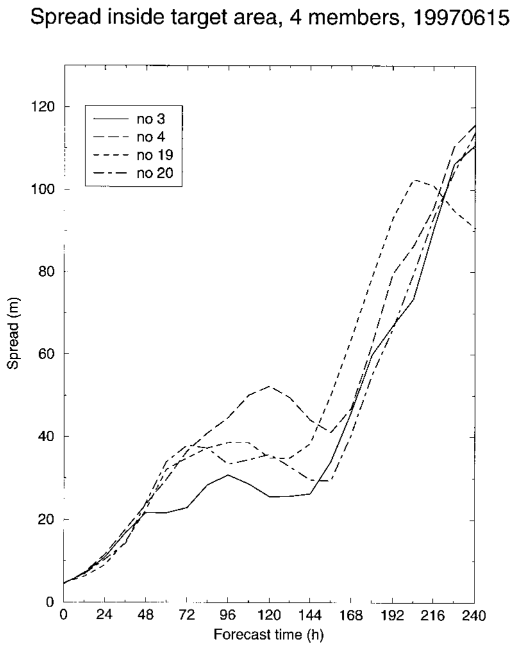

Fig. 8. 500 hPa geopotential height spread for four single ensemble members inside the target area from 15 June 1997, as a function of forecast time.

time 0 and also later if the evolution is linear, and it increases with the degree of non-linearity. I_{NL} shows a steady increase the first days. Around day 7–8 I_{NL} shows a marked increase inside the target area. This is due to a shift in the location between members 3 and 4 (not shown). This was also around the time for the sudden increase in spread for the two members and could indicate that the increase in spread after forecast day 5 is connected with non-linearity. The solid line in Fig. 10 shows I_{NL} from Fig. 9 averaged over the target area. This curve clearly resembles the curve for the spread inside the target area with an increase up to day 5, then a plateau and finally a sudden, steep increase. Fig. 11 shows the same as Fig. 9 for member number 19 and 20, for which it is even more clear that the sudden increase in spread around day 6 is non-linear. Integrated over the target area (dashed line in Fig. 10) I_{NL} indeed shows a marked increase at day 6. Note that the non-linearity in Fig. 11 is found in the same location as in Fig. 9. This was also the case for several other members (not shown), and could indicate that non-linearity is connected to the background flow. At the initial time for this particular forecast, there was a frontal zone over Norway and the region continued to be active for the next days.

I_{NL} was also calculated for all pairs of ensemble members for the whole summer period. Integrated over the target area (not shown) the non-linearity index shows a smooth increase from 0 m to approximately 100 m. The plateaus seen in Fig. 10 is not evident, this is due to the different timing of the plateaus for the different cases.

4.3. *The quality of ensemble prediction measured by BSS and area under ROC curves*

4.3.1. Large scale precipitation. In Fig. 12, the area under the ROC curves and the BSS for TEPS, EPS 20 and EPS 50 for the 20 winter cases inside the target area are shown. The spatial resolution of the grid used for verification is 1.5°/1.5° (the area contains 578 points). The event threshold is 7.5 mm/12 h compared to climatology, where climatology is computed from 5 years of monthly mean fields. An area less than 0.7 is often taken as a limit for when the forecast is useless (Buizza et al., 1999b). For the first 5 days, all 3 forecasting systems are over this limit. (It should be noted that verification before day 1.5 for the winter cases and before day 2.0 for the summer cases is not possible, because up to this time the verifying analysis is based on a forecast started at the same time or earlier than the ensemble.) As measured by the area under ROC curves, TEPS gives similar results as EPS 20 up to about 3.5 days. For the Brier skill score, TEPS performs comparable to the operational EPS 50 up to about day 3.5. Similar results are obtained for different thresholds. Fig. 13 is the same as Fig. 12 for the 15 summer cases. There is a small improvement for TEPS for day 2 in terms of area under the ROC curve. TEPS becomes worse than EPS 50 already after about 2.5 days, but stays better than EPS 20 up to day 4. According to BSS, TEPS is better than the operational EPS between day 2 and 3.5. Different thresholds give similar results, and generally forecast day 2 gives, not surprisingly, the largest improvement. The differences are, however, relatively small.

4.3.2. Two-meter temperature. The improvement for TEPS over EPS is larger for two-meter temperature than for LSP, especially for large thresholds (positive and negative). In Fig. 14, the area under the ROC curves and the BSS for two-meter

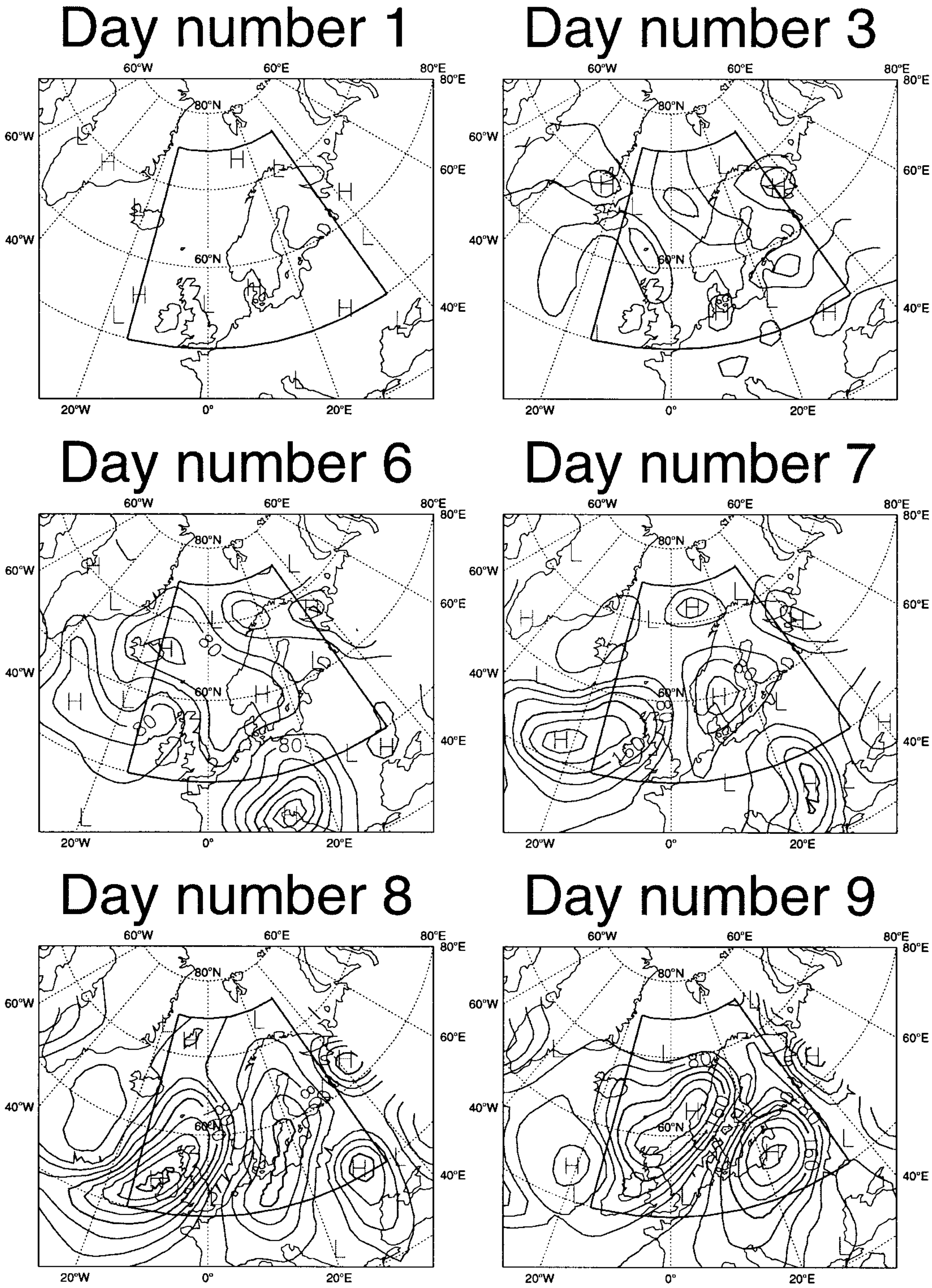

Fig. 9. Non-linearity index for member number 3 and 4 from 15 June 1997. Displayed are forecast day no. 1, 3, 6, 7, 8 and 9. The contour interval is 40 m.

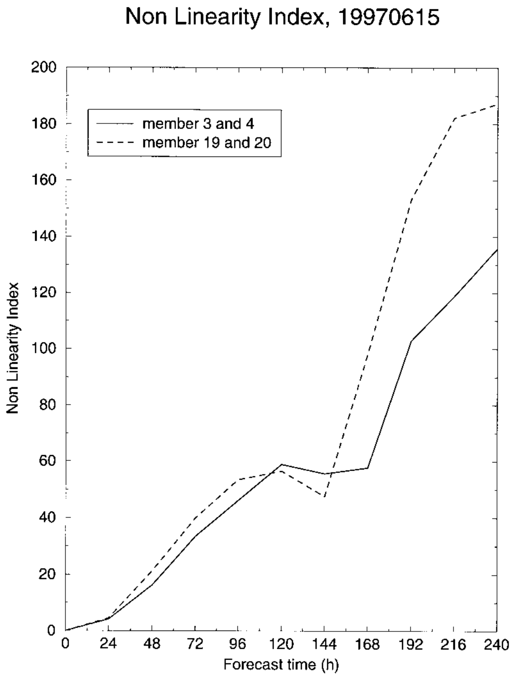

Fig. 10. Non-linearity index integrated over the target area for 5 June 1997, solid line member number 3 and 4 and dashed line member number 19 and 20.

temperature for the 20 winter cases for a threshold of warmer than 8°C compared to climatology, is shown. The climatology is computed from 5 years of monthly mean fields, and there is a discrimination between noon and midnight. For the area under the ROC curves, TEPS is better or equal to the operational EPS with 50 members up to day 3 and the operational EPS with 20 members up to day 4.5. For BSS, TEPS is better or similar to EPS 50 and EPS 20 up to about day 4.5. The same conclusions hold also for other thresholds, with the tendency for the improvement of TEPS over EPS for forecast day 2 to increase as the absolute value of the threshold increases. This means that for forecast day 2, TEPS is better for the more extreme events. For the summer cases the results are similar (not shown).

4.4. *The quality of ensemble predictions measured by ROC curves and cost/loss analyses*

The ROC curves for forecast day 2 for which the area was discussed in the previous section, are in the upper panels of Figs. 15, 16 and 17, respectively. In all curves we see a gap between the upper right corner (the event is always predicted) and the top most point (one or more ensemble members predicted the event). This is due to the events being rare (p_cli = 0.028, p_cli = 0.014 and p_cli = 0.091, respectively, where p_cli is the climatological frequency). Rare events give a low false alarm rate because the number of non-occurrences is the denominator of the false alarm rate. In all three cases, the curves are almost identical up to the last point (i.e., one or more ensemble members predicted the event). For Figs. 16 and 17 the top curve is the TEPS curve, for 15 it is the EPS 50 curve. The improvement in area discussed in the previous section is mainly due to this uppermost point. For Figs. 16 and 17, the results for TEPS are an increase in the hit rate relative to the two EPS sets, but the false alarm rates are mainly unchanged. Since the denominators in the calculation of the hit rate and the false alarm rate are the same for the three sets, the increase in the hit rate value is due solely to an increase in the number of correct forecasts of the event. Since the false alarm rate is mainly unchanged, the number of incorrect forecasts of the event does not increase substantially. TEPS is therefore better at detecting rare events compared to the two other EPS sets.

The value curves for forecast day 2 corresponding to the described ROC curves are in the lower panels of Figs. 15, 16 and 17. For the two cases where the area under the ROC curve was largest for TEPS, we see a marked improvement of TEPS over both EPS 20 and EPS 50 for a wide range of cost/loss ratios. Also the maximum value is generally higher. The improvement over the EPS sets is mainly to the left of the peak in the two figures, i.e., for relatively small cost/loss ratios. Therefore, users with cost/loss ratios in this range will benefit from using TEPS rather than the operational EPS. In Figs. 16 and 17, we see that the improvement is larger when going from EPS 50 to TEPS than when going from EPS 20 to EPS 50. For LSP in the winter (Fig. 15), there is no improvement for TEPS (except for a small range of cost/loss ratios in the lower right corner). One might expect that users with high potential loss will be very interested in such an improvement seen in Figs. 16, 17. Small cost/loss ratios mean that a user has either low cost and/or a high potential loss. The results for 2T in the summer

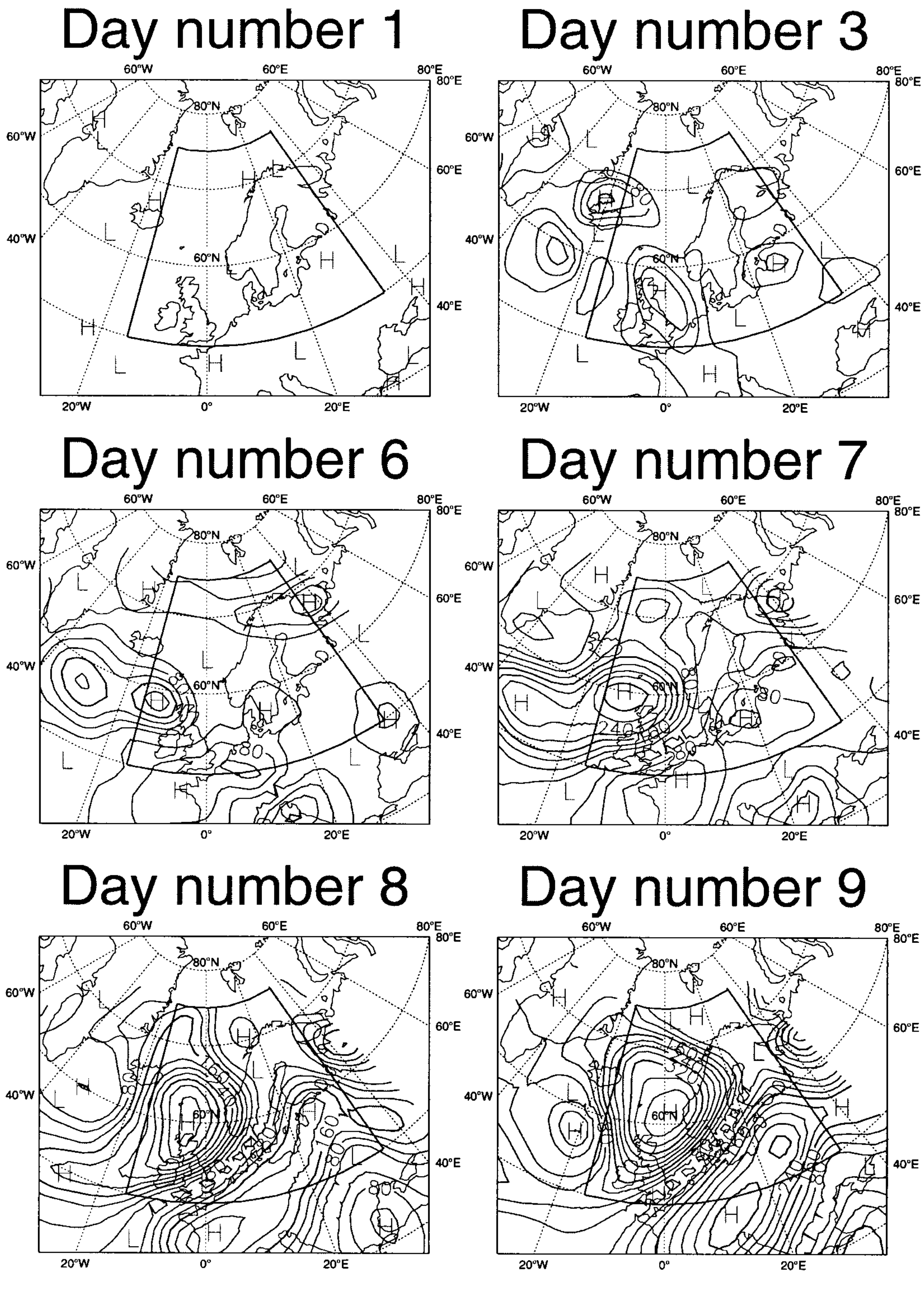

Fig. 11. Same as Fig. 9, but for member 19 and 20 from 15 June 1997. The contour interval is 40 m.

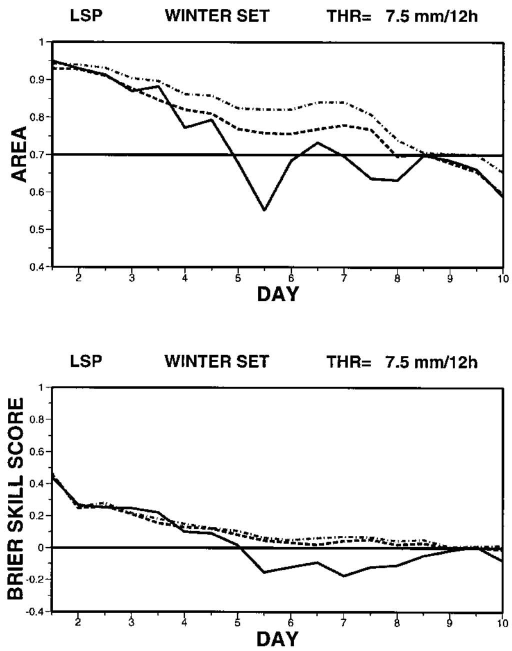

Fig. 12. Area under ROC curves (top panel) and Brier skill score (bottom panel) for TEPS (solid line), EPS 20 (dashed line) and EPS 50 (chain-dashed line) for more than 7.5 mm/12 h LSP compared to climatology, computed for the target area and all 20 winter cases.

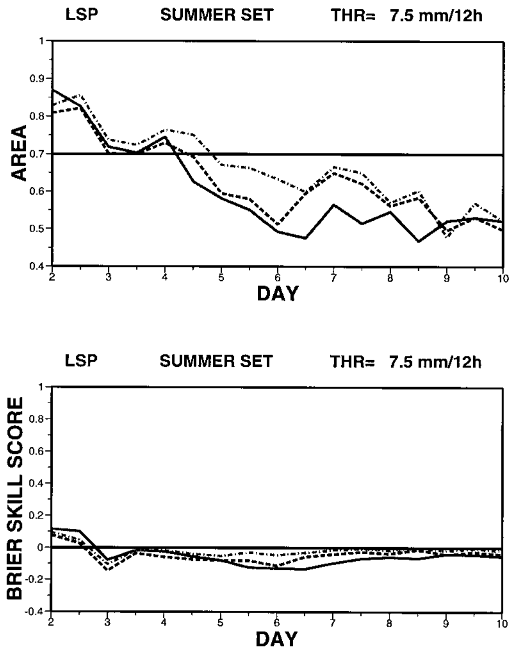

Fig. 13. As Fig. 12, but for the 15 summer cases.

are similar to the winter results, except the improvement of TEPS is in that case also evident for forecast day 3 (not shown).

5. Summary and conclusions

The aspects of targeted ensemble prediction for northern Europe and parts of the Atlantic are investigated. A total of 35 ensembles were computed, 20 winter sets and 15 summer sets. The set-up resembles that of the operational ECMWF EPS-system, except for the target area and that only 20 ensemble members are integrated. The aim is to see if constraining the perturbations to a specific area will increase the quality of ensemble prediction in that area, compared to the operational, even with fewer members in the ensemble.

A main issue is to increase the spread inside the target area in order to describe a larger portion of the probability distribution for the initial state. For all 35 cases the spread increases relative to the operational EPS up to about forecast day 4. The increase in spread for TEPS over EPS is also larger than the increase when going from 20 to 50 members for the operational EPS. If larger spread for a sub-domain of the northern hemisphere is desired, it appears from this result that targeting is a better approach than increasing the ensemble size.

TEPS was run for 10 days. We found that the ensemble spread increased throughout the 10-day forecasts for all 35 cases, but with a smaller increase in the middle of the forecasting period. The first increase in spread is quasi-linear and is consistent with the view that generated perturbations move through the target area. The later increase is non-linear with quasi-stationary spatial patterns.

The experimental TEPS and the operational EPS with all 50 and only 20 members were compared by using Brier skill score (BSS), relative operating characteristic (ROC) curves, the area under these curves and cost/loss analysis. The parameters used in the comparison were large scale precipitation (LSP) and two-meter temperature (2T).

For the winter sets, the impact of targeting was minimal for LSP, although TEPS performs com-

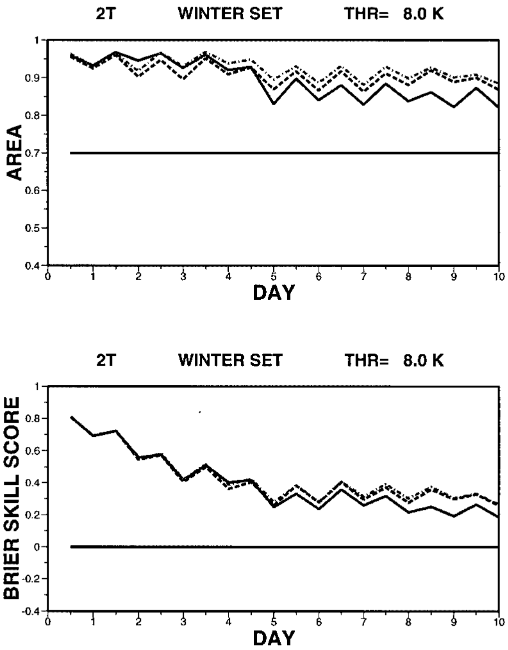

Fig. 14. Area under ROC curves (top panel) and Brier skill score (bottom panel) for TEPS (solid line), EPS 20 (dashed line) and EPS 50 (chain-dashed line) for warmer than 8°C compared to climatology, computed for the target area and all 20 winter cases.

parable to EPS 50 for most thresholds for the first days, except for large thresholds when TEPS performed a little bit worse than EPS 20 and EPS 50. It was established that most of the northern hemispheric singular vectors (NHSVs) already were concentrated around the target area in the winter. The summer period was characterised by comparable activities in the two major low pressure belts in the northern hemisphere, and hence the number of NHSVs was comparable in the two tracks. For this period TEPS for LSP for day 2 is better than both EPS 20 and EPS 50 for all thresholds. For day 3 the score was usually between EPS 20 and EPS 50.

For 2T, the impact of targeting was evident also for the winter period with scores better or equal to EPS 50 at day 2, and scores between EPS 20 and EPS 50 for day 3. For the summer period TEPS scores better than both EPS 20 and EPS 50 for all thresholds at least up to forecast day 3.

From these results it seems reasonable to tentatively conclude that targeting is a good approach

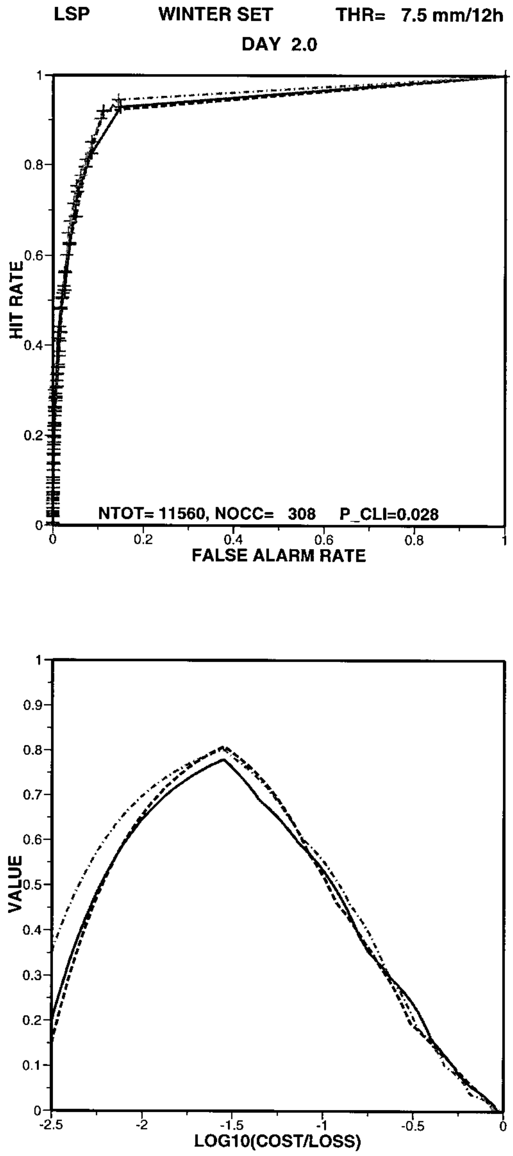

Fig. 15. ROC curve (top panel) and cost/loss analysis (bottom panel) for TEPS (solid line), EPS 20 (dashed line) and EPS 50 (chain-dashed line) for more than 7.5 mm/12 h compared to climatology, computed for the target area and all 20 winter cases at forecast day 2. NTOT is the number of points the event is tested for, NOCC is the number of times the event occurred in the sample and P_CLI is the climatological frequency.

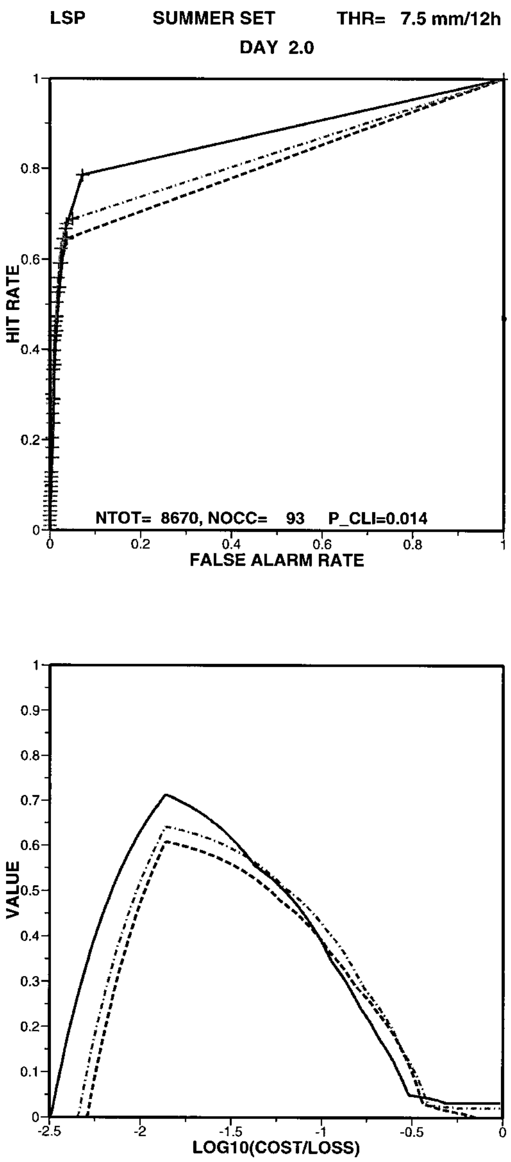

Fig. 16. As Fig, 15, but for the 15 summer cases.

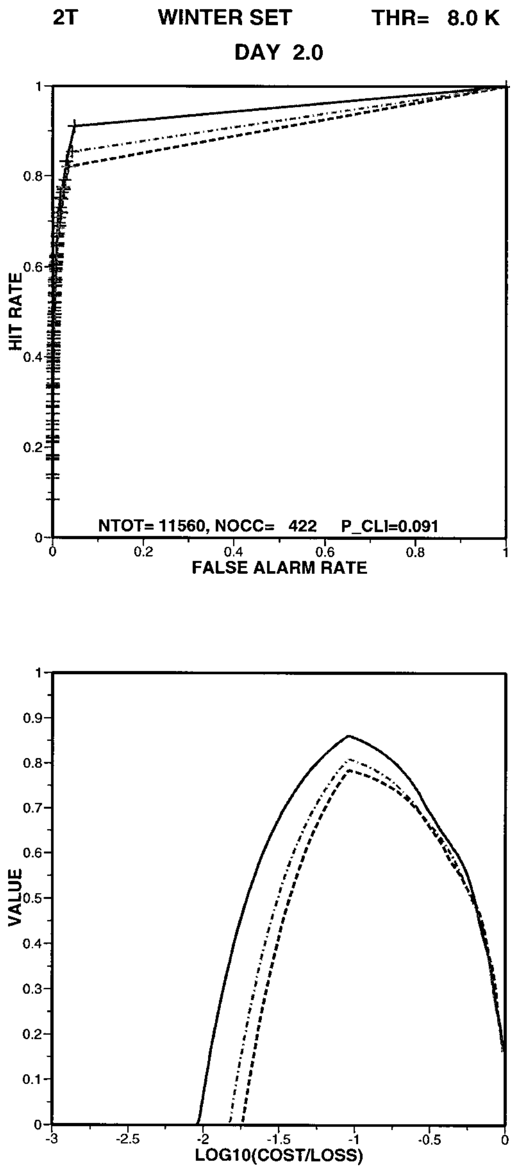

Fig. 17. ROC curve (top panel) and cost/loss analysis (bottom panel) for TEPS (solid line), EPS 20 (dashed line) and EPS 50 (chain-dashed line) for warmer than 8°C compared to climatology, computed for the target area and all 20 winter cases at forecast day 2. NTOT is the number of points the event is tested for, NOCC is the number of times the event occurred in the sample and P_CLI is the climatological frequency.

for improving the performance of ensemble forecasting for a limited area for the first 2–3 forecast days when using ensemble members based on 2-days SVs. For most cases 20 targeted ensemble members are able to give as much or more information as 50 NHSVs. This should be confirmed by a larger number of cases.

A different approach for enhancing the performance inside a sub-domain of the northern hemisphere is to run ensembles with a limited area model (LAM). For making ensembles with LAM it is desirable to perturb the lateral boundary conditions as well as the initial conditions. Gustafsson et al. (1998) have performed a sensitivity analysis for LAM in which the adjoint was used and found out that both the initial and boundary fields have influence. We have started work where the targeted ensembles created here will be used as initial and boundary conditions in the HIRLAM model. The aim will be to see if ensemble forecasting will improve LAM-prediction.

6. Acknowledgements

We are grateful to the staff at the European Centre for Medium-Range Weather Forecasts (ECMWF) for letting Inger-Lise Frogner stay for 6 months working on this project. We are also grateful to ECMWF for letting us use their computer resources. We acknowledge T. Palmer, J. Barkmeijer, D. Richardson, R. Buizza and H. Hersbach for providing parts of the verification tools and for stimulating discussions. We are also grateful to the two anonymous reviewers for valuable comments on the manuscript.

7. Appendix

We have used validation statistics that are not common in regular deterministic forecasting. This appendix summarises briefly their formulas and properties.

(A) *The Brier skill score (BSS)*

BSS (Brier, 1950, Stanski et al., 1989) is used on probabilistic forecasts. The forecast probability f is the fraction of ensemble members that predict the event in question. If O is a Boolean variable that equals 1 if the event occurs and 0 if not, we can define the Brier score (BS):

$$\mathrm{BS} = \frac{1}{N} \sum_{i=1}^{N} (f_i - O_i)^2,$$

where N is the number of cases (i.e., all dates considered and all points inside the target area). The score is negatively oriented. Zero is perfect and 1 is no skill. BS is strongly influenced by the climatological frequency of the event. The rarer the event, the better the Brier score that can be obtained, without showing any skill. Therefore, the Brier skill score (BSS) is used instead:

$$\mathrm{BSS} = \frac{\mathrm{BS_c} - \mathrm{BS_f}}{\mathrm{BS_c}},$$

where $\mathrm{BS_f}$ is the Brier score for the forecast and $\mathrm{BS_c}$ is the Brier score based on a reference forecast (here climatology). $\mathrm{BS_c}$ is calculated by $f = c$, the climatological probability:

$$\mathrm{BS_c} = s(c-1)^2 + (1-s)c^2,$$

where s is the sample frequency.

BSS has a range of $-\infty$ to 1, where BSS equals one for the perfect forecast. If BSS equals zero, $\mathrm{BS_c} = \mathrm{BS_f}$ and no improvement over climatology is obtained, and for skilled forecasts BSS > 0. BSS is unstable if it is applied to forecasts of rare events. The score should not be used for samples with very few occurrences of the event.

(B) *Relative operating characteristic (ROC) curves*

Also ROC curves are used to test certain events (Mason, 1982), by appraising the ability of the forecast to discriminate between two alternate outcomes. Consider a two category contingency table (Table 1).

X are referred to as "hits", Y as "misses", Z as "false alarms" and W as "correct rejections". From these we define the hit rate:

$$\mathrm{HR} = \frac{X}{X+Y} = \frac{X}{p},$$

Table 1. *Contingency table for an event. X, Y, Z, W and TOTAL are number of occurrence of events*

		Event forecasted: yes	Event forecasted: no	
Event observed	yes	X	Y	$X+Y$
	no	Z	W	$Z+W$
		$X+Z$	$Y+W$	total

and the false alarm rate:

$$\mathrm{FAR} = \frac{Z}{Z+W} = \frac{Z}{(1-p)},$$

where p is the climatological probability of the event. These can then be plotted as a single point in a diagram with the false alarm rate along the abscissa and the hit rate along the ordinate. Ideally, the point should lie in the upper left corner of the diagram. This idea is generalised and used for probability forecasts. The data are binned into classes, according to the number of ensemble members that predicts the event. For each category the number of occurrences and non-occurrences is calculated. A threshold for forecasting the event is then chosen, e.g., when n of the ensemble members predict the event. All categories with n or more ensemble members indicating the event is taken as a forecast of the event, and all categories below as event not forecasted. This can be repeated for different thresholds. The result will be a curve, called the *relative operating characteristic* (ROC) curve. A perfect forecast is represented by a curve that lies along the left-hand and upper sides of the graph. The area under the ROC curve is a good measure of how well the system performs. An area of one is the highest achievable, and the area decreases toward 0 as the curve moves away from the left and top sides of the diagram. An area of 0.5 represents a useless forecast, because it means equally as many false alarms as hits.

(C) *Cost/loss analyses*

By using a cost/loss analysis (Katz and Murphy, 1997; Richardson, 1998), a user of a forecasting system can decide whether to take action or not against a possible hazardous weather event. If action should be taken depends on the cost (C) of taking preventing action, the loss (L) if the event occurs and action was not taken, and the climatological probability of the event (p). For a deterministic forecast it is natural to take action if the event is predicted, and otherwise do nothing. For ensemble prediction one has to decide how many members should predict the event before action should be taken. Different users may have different thresholds for taking action, depending on their cost/loss ratio. A decision maker is assumed to minimise his/her expected loss over a large number of cases. The economic value is measured by comparing the expected utilities with and without the forecast. When choosing between two possible actions, the decision maker will either incur a cost C, a loss L or zero cost. If protective action is taken the cost will be C whether the event occurs or not. If protective action is not taken, a loss will only occur when the weather event occurs.

The mean expense if protective action is always taken is C. If protective action is never taken, the mean expense will be pL. If no other information than climatology is available, the optimal strategy would be to always initiate protective action if $C/L < p$ and never otherwise. If one had a perfect forecasting system, the minimum mean expense would be pC, which means only protect when the weather event occurs and never incur any loss. Fig. 18 shows an expense diagram with the cost/loss ratio along the abscissa and mean expense per loss along the ordinate. The line marked "perfect" is the lower limit for the mean expense, which is the expense one would get with a perfect forecasting system. The heavy, solid line is the minimum mean expense given climate information. The region in which a better forecasting system can give a lowering of the expenses, is therefore inside the triangle in the diagram.

The mean expense (ME) per loss is defined as:

$$\mathrm{ME} \equiv \frac{YL + (Z+X)C}{L},$$

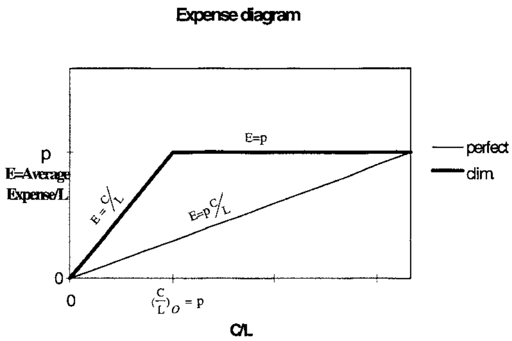

Fig. 18. Mean expense per unit loss as a function of the cost/loss ratio. Heavy solid line represents decisions based on climate information. The line marked "perfect" represents the expense obtained with perfect knowledge of the future. Both lines depend on the frequency of the event, as well as the cost/loss ratio.

or by introducing HR and FAR:

$$\mathrm{ME} = Y + (Z+X)\frac{C}{L}$$
$$= \mathrm{FAR}\,\frac{C}{L}(1-p) - \mathrm{HR}p\left(1-\frac{C}{L}\right) + p.$$

The economic value of the forecast is then defined as follows:

$$V = \frac{\mathrm{ME(climate)} - \mathrm{ME(forecast)}}{\mathrm{ME(climate)} - \mathrm{ME(perfect)}}$$
$$= \frac{\min\left(\frac{C}{L}, p\right) - \mathrm{FAR}\,\frac{C}{L}(1-p) + \mathrm{HR}p\left(1-\frac{C}{L}\right) - p}{\min\left(\frac{C}{L}, p\right) - p\frac{C}{L}},$$

where ME(climate) stands for the mean expense when using climate, ME(forecast) means that the forecasting system is used, and ME(perfect) is the mean expense we would obtain if we had a perfect forecasting system. V is then the value of the forecasting system, as a proportion of what would be achieved by a perfect forecast. A maximum value always occurs for $C/L = p$. A perfect forecasting system would give a value of 1, and for a climate forecast V would equal 0. If $V > 0$ there is a benefit for the user.

The value obtained for probabilistic forecast also depends on the probability threshold chosen for when to take action or not. This choice converts the probability forecast to a deterministic forecast. All forecasts with a higher probability for the event to occur are taken as a forecast that the event will occur. Those with lower probability are taken as forecasts that the event will not occur. For each possible probability threshold a value curve can be obtained. The envelope of these curves then forms the optimal value curve, i.e., the maximum possible value obtained with the forecasting system. Such envelope curves of V as a function of C/L are displayed in Section 4.

REFERENCES

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**, 1–3.

Buizza, R. 1994. Localization of optimal perturbations using a projection operator. *Q. J. R. Meteorol. Soc.* **120**, 1647–1681.

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

Buizza, R. and Palmer, T. N. 1998. Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.* **126**, 2503–2518.

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

Buizza, R., Miller, M. and Palmer, T. N. 1999a. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**, 2887–2908.

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

Gustafsson, N., Källén, E. and Thorsteinsson, S. 1998. Sensitivity of forecast errors to initial and lateral boundary conditions. *Tellus* **50A**, 167–185.

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

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

Hersbach, H., Mureau, R., Opsteegh, J. D. and Barkmeijer, J. 2000. A short to early-medium range ensemble prediction system for the European area. *Mon. Wea. Rev.*, in press.

Joly, A., Jorgensen, D., Shapiro, M. A., Thorpe, A., Bessemoulin, P., Browning, K. A., Cammas, J-P., Chalon, J-P., Clough, S. A., Emanuel, K. A., Eymard, L., Gall, R., Hildebrand, P. H., Langland, R. H., Lemaitre, Y., Lynch, P., Moore, A., Persson, P. O. G., Snyder, C. and Wakimoto, R. M. 1997. The fronts and Atlantic storm-track experiment (FASTEX): scientific objectives and experimental design. *Bull. American Met. Soc.* **78**, 1917–1940.

Katz, R. W. and Murphy, A. H. 1997. *Economic value of weather and climate forecasts.* Cambridge University Press, 222 pp.

Legras, B. and Vautard, R. 1996. A guide to Liapunov vectors. *Proceedings of the ECMWF Seminar on Predictability*, 4–8 September 1995. Reading, England, vol. 1, 143–156.

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

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

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

Mason, I. 1982. A model for assessment of weather forecasts. *Aust. Met. Mag.* **30**, 291–303.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T.

1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

Palmer, T. N. 1996. Predictability of the atmosphere and oceans: from days to decades. *Proceedings of the ECMWF Seminar on Predictability*, 4–8 September 1995. Reading, England, vol. 1, 83–141.

Richardson, D. 1998. Obtaining economic value from the EPS. *ECMWF Newsletter* **80**, 8–12.

Simmons, A. J. and Hoskins, B. J. 1979. The downstream and upstream development of unstable baroclinic waves. *J. Atmos. Sci.* **36,** 1239–1254.

Stanski, H. R., Wilson, L. J. and Burrows, W. R. 1989. Survey of common verification methods in meteorology. Atmospheric Environment Service Research Dep, 89-5, 114 pp (available from: Forecast Research Division, 4905 Dufferin St., Downsview, ON M3H 5T4, Canada).

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

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bul. American Met. Soc.* **74**, 2317–2330.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.

Whitaker, S. and Loughe, A. F. 1998. The relationship between ensemble spread and the mean skill. *Mon. Wea. Rev.* **126**, 3292–3302.