LAMEPS – Limited area ensemble forecasting in Norway, using targeted EPS

Marit H. Jensen, Inger-Lise Frogner* and Ole Vignes, Norwegian Meteorological Institute,
(*held the presentation)

At the Norwegian Meteorological Institute we have run LAMEPS quasi-operationally since mid-February 2005. LAMEPS is run with the Norwegian version of the HIRLAM model and it is driven by members of the ECMWF EPS which are targeted to produce maximum spread amongst ensemble members after 48 hours in northern Europe and adjacent sea areas. This system is abbreviated to targeted EPS, or simply TEPS. We have made a comparison between the 50 member EPS and the 20 member TEPS for an area covering much of the target area.

A multi-model ensemble system (NORLAMEPS) is also used which simply combines LAMEPS and TEPS by using all ensemble members from both systems simultaneously. This combination gives a larger ensemble without extra model runs. Even though the combined system is to some extent an auto-duplication, the ensemble spread is larger because for two reasons: there are un-correlated differences between fields from the different models, and the LAMEPS control forecast with HIRLAM can deviate considerably from the TEPS control with the ECMWF Integrated Forecast System (IFS).

Here we describe the model setup for LAMEPS, TEPS and NORLAMEPS, and show some verification results for the summer and spring of 2005.

Model setup

LAMEPS is an ensemble of runs with the Norwegian version of the limited area model HIRLAM (horizontal resolution of 0.2° with 40 levels). It uses ensemble members from TEPS to perturb both the initial and the lateral boundary conditions.

TEPS uses the same model version and the same set-up as used for the operational EPS. Only 10 singular vectors are calculated as opposed to 25 in the EPS. These singular vectors are targeted to maximize the total energy at final optimization time (48 h) in Northern Europe and adjacent sea areas (see Figure 1). Singular vectors at initial time and 48 hours evolved singular vectors valid at the same time are combined to form initial state perturbations. These are added to and subtracted from the initial state analysis (the "control") with amplitudes based on analysis error estimates. TEPS thus contains 20 ensemble members in addition to the control forecast. The TEPS forecast length is 96 hours and is run at ECMWF once per day at 12 UTC.

Each LAMEPS ensemble member is constructed by running HIRLAM from 20 alternative initial states obtained by adding the 20 TEPS ensemble perturbations (the difference between each TEPS ensemble member and the TEPS control) to the HIRLAM 18 UTC analysis. At the open lateral boundaries the time-developed TEPS ensemble members, corresponding to those used for initial perturbations, are imposed. Thus we obtain 20 different forecasts in addition to the HIRLAM control run. Since HIRLAM starts with an 18 UTC analysis and TEPS with a 12 UTC analysis, the forecasts from LAMEPS are 6 hours shorter than the forecasts from TEPS and EPS. LAMEPS is run at 18 UTC every day and the forecast length is 60 hours.

NORLAMEPS combines the forecasts from TEPS and LAMEPS to provide a single statistic for events, even though they are not entirely independent of each other. Without extra cost, the total number of ensemble members is then 41 in addition to the HIRLAM control forecast. In this way NORLAMEPS is supposed to partly account for uncorrelated forecasts errors caused by model imperfections. The differences between the initial fields in TEPS and LAMEPS are partly caused by these model differences.

Verification methodology

Verification of precipitation forecasts against SYNOP observations is not straightforward because of the very different scales of observations and the forecasts. Furthermore, LAMEPS and TEPS/EPS also have different resolutions. Hence comparison of the different systems using either of the analysed fields from the HIRLAM or the ECMWF IFS model as "the truth" would favour one of the systems. We use the approach proposed by *Ghelli & Lalaurette* (2000) and construct so-called "super-observations" which are representative of precipitation grid-squares. Here all precipitation stations in Norway (several hundreds) inside the verification area (Figure 1) are aggregated to regular grids. Precipitation "super-observations" representative of our 0.2° x 0.2° HIRLAM grid are thus calculated. All the verification of precipitation described here uses such super-observations.

Total precipitation from LAMEPS, TEPS, EPS and NORLAMEPS are compared to the super-observations using Rank Histograms, Reliability Diagrams, Brier Skill Scores, ROC curves, and cost/loss analysis. The diurnally accumulated precipitation observations are taken at 06 UTC. Since the forecasts starts at 12 UTC and 18 UTC and are 66/60 hours long, this leaves only two possible time-intervals in the forecast range for verification:

(+12/18 h to +36/42 h) and (+36/42 h to +60/66 h). Note that since LAMEPS is started 6 hours later than TEPS and EPS, the forecast from LAMEPS is 6 hours shorter than the other two forecasts. The distribution of precipitation in Norway is dominated by sharp gradients, caused by predominant westerly winds, a long coastline, and a complex topography. The gradient across the divide between the western and eastern watersheds in Southern Norway is particularly large. The western watersheds receive large amounts of precipitation, while the eastern ones are frequently sheltered by the mountains. Typically the annual difference amounts to a factor of 2 to 3, but in several cases the differences are even much larger. It was noted that agglomerations of samples spanning locations and times with different climatological frequencies can lead to spurious skill measures. To circumvent this problem we verify separately sub-regions with grossly different precipitation climatologies (divided Norway into three - east, west and north) . The precipitation frequencies also vary over the year, and we split the verification results into spring (February-April) and summer (May-July). Averages are calculated using weights reflecting the area of the sub-regions and by the number of days in the two periods.

The ensemble spread

We have chosen to define the spread as the rms-difference between the ensemble members and the ensemble mean as:

$$s = \frac{1}{I} \sum_{i=1}^{I} \sqrt{\frac{1}{N \cdot D} \sum_{n=1}^{N} \sum_{d=1}^{D} \left(e_{ind} - m_{id}\right)^2}$$

where I is the number of grid-points inside the verification area, N is the number of ensemble members, D is the number of cases, e_{ind} is the ensemble member value for member n in the case d and in the specific point i, and m_{id} is the ensemble mean for the same case and in the same point.

	LAMEPS	TEPS	EPS	NORLAMEPS
+36/+42	2.15	2.08	1.56	2.19
+60/+66	2.47	2.38	2.07	2.47

Table 1 Spread around ensemble mean for total precipitation for the four ensemble systems.

Table 1 shows the spread for the systems for total precipitation forecasts. The spread increases with forecast time, which is in line with expected behaviour of unstable systems starting from small perturbations. The EPS has the smallest spread for both forecast times. The main reason for targeting is to constrain the perturbations to a predefined area of particular interest over a certain forecast range. It is expected that a targeted system will have a larger spread between ensemble members in this target area than a system that is not targeted, given the same number of singular vector based perturbations. Thus a 20-member TEPS ensemble has a considerably larger spread between the members than EPS with 50 members. Hence the TEPS ensemble includes a wider selection of fast-growing disturbances over the time-range and area of interest than the EPS ensemble. As a consequence forecasts made from TEPS over the selected time-range can quantify risks of more extreme cases better than those based on EPS. The realism of the increased risk needs to be investigated.

The EPS ensemble with 50 members has considerably smaller spread than both ensembles from TEPS and LAMEPS each of which has only 20 members. The LAMEPS ensemble has a slightly larger spread than the original TEPS ensemble, even though the initial and boundary perturbations are entirely based on TEPS.

NORLAMEPS, which is a simple combination of the TEPS and LAMEPS, has the largest ensemble spread of all the tested systems for forecast range 36/42 h. Hence, LAMEPS triggers slightly different unstable structures in HIRLAM than the global model used for generating TEPS. This difference can partly originate from the higher resolution of LAMEPS and partly from the fact that two different models are used to compute the ensemble members in NORLAMEPS. The spread in the NORLAMEPS ensemble and the associated risks of potentially extreme weather developments should therefore be taken as due to a combination of chaos and model uncertainty. It is impossible to tell to what extent the additional spread from model uncertainty stems from uncorrelated model errors or if it is due to equally realistic but different time developments.

Probabilistic scores

All four ensemble systems are evaluated for total precipitation accumulated over the verification times. The observations of precipitation are mainly at 06 UTC and therefore we use this time for the verification. The verification time for LAMEPS differ from that for TEPS and EPS because of the forecast starts 6 hour for LAMEPS. The verification times are therefore 12-36 h and 36-60 h for LAMEPS, 18-42 h and 42-66 h for TEPS/EPS and a combination of these times for NORLAMEPS. The scores are calculated separately for each of the three sub-domains shown in Figure 2 and for the two seasons (spring and summer). The area-weighted average over the domains and the two seasons are computed.

We have used different standard probabilistic scores for the verification of the four systems. In Figure 2 the Brier Skill Scores for the 12/18-36/42 h and 36/42-60/66 h forecast as a function over precipitation threshold, is shown. Figure 3 shows the area under the ROC curves and Figure 4 the ROC curves (top row) and the cost/loss analysis (bottom row) for the weighted mean over the tree areas and two seasons and two time periods (12/18-36/42 h to the left and 36/42-60/66h to the right). The event threshold is 5 mm/day. Figure 5 shows the ROC curves and cost/loss analysis for event threshold 20 mm/day.

In these figures one can see that LAMEPS has a considerably lower score for the low precipitation thresholds. For mid to high thresholds NORLAMEPS has very good scores, showing that LAMEPS gives extra and valuable information to the TEPS.

We have also looked at the rank histograms and the Reliability diagrams for the four systems (not shown here). The rank histograms indicate a bias in all four systems where they all underestimate the variability. The underestimation is small in TEPS and especially large in EPS. The Reliability diagrams show good reliability up to about 70% for TEPS and NORLAMEPS, after which the two systems over-forecast the higher probabilities. LAMEPS and EPS over-forecast the probabilities from about 30-40%.

Discussion and conclusion

From the results we have seen that LAMEPS is able to produce more spread than EPS for precipitation over Norway. For events with small precipitation amounts the probabilistic scores for TEPS and EPS are better than for LAMEPS, but for larger precipitation amounts LAMEPS scores better.

The combined system NORLAMEPS gets the largest spread and also best probabilistic scores from mid to high precipitation amounts. The combination of LAMEPS and TEPS adds value to the two individual systems. The improvement by NORLAMEPS can partly be due to the increase in resolution for LAMEPS, and partly from the fact that it combines results from two different model systems.

The comparison of the 20 members TEPS and 50 members EPS is interesting for the Norwegian verification area. The increase in ensemble spread of TEPS compared to EPS demonstrates the advantage of targeting. Also for many of the probabilistic scores TEPS is better or comparable to EPS, even though it has fewer members.

The verification has shown that TEPS gives better results for the short range verification period. This may indicate that the method of perturbing TEPS gives very good results early in the forecast range, but that the weather system then moves out of our target area. In this case the ordinary EPS has an advantage. One way of dispensing with this is to combine TEPS and EPS as input to LAMEPS.

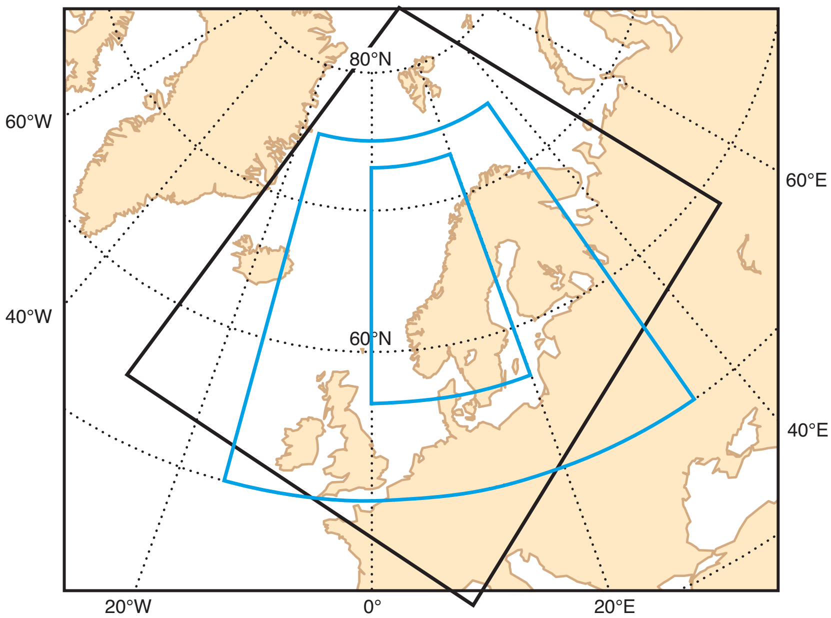

Fig. 1 Areas used in the experiments. The large black area is the HIRLAM integration area, the large blue area is the target area for optimization of singular vectors, and the small blue area is the verification domain

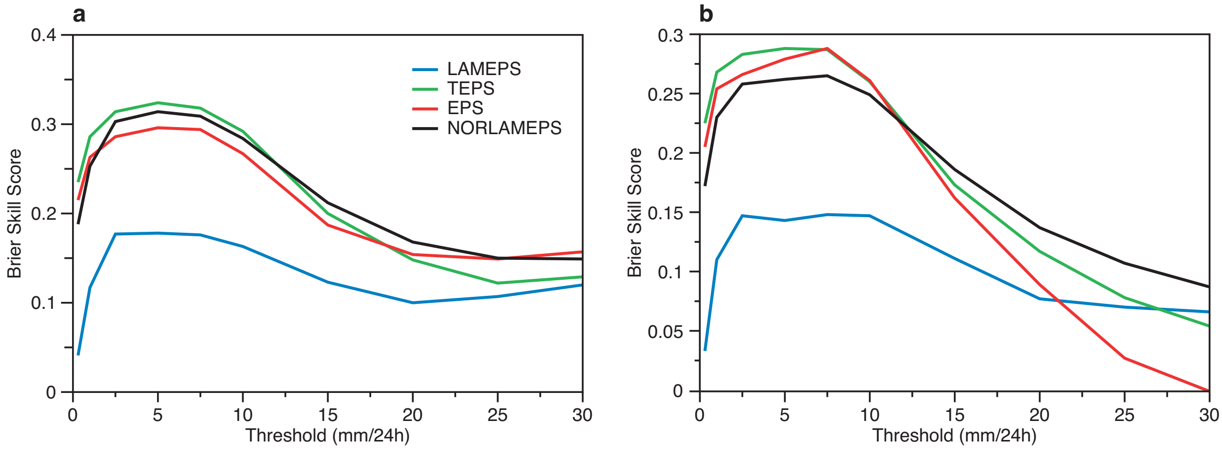

Fig. 2 Brier Skill Scores for precipitation as a function of threshold for (a) 12/18-36/42 h and (b) 36/42-60/66 h forecasts. Mean over all verification areas and both seasons

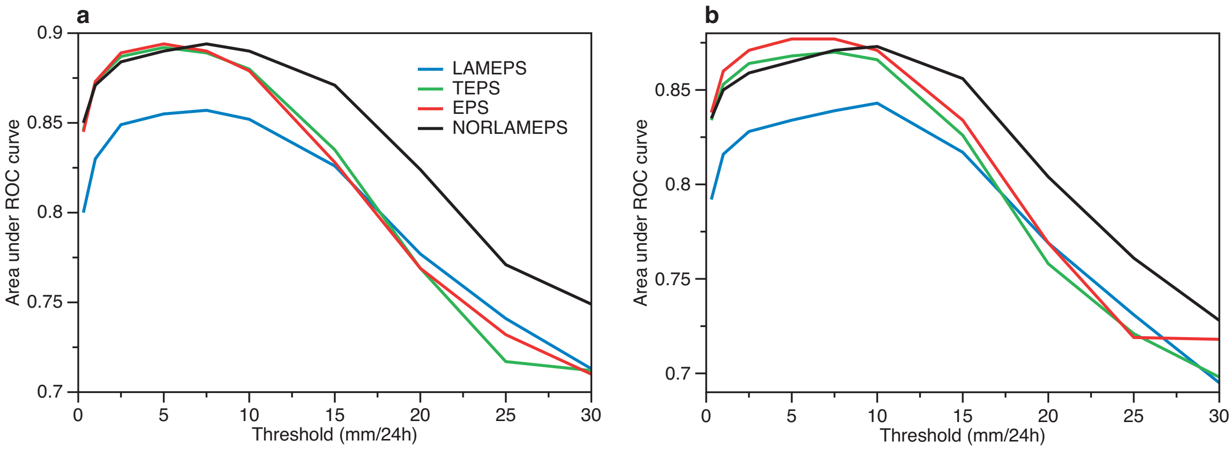

Fig. 3 Area under the ROC curve for precipitation as a function of threshold for (a) 12/18-36/42 h and (b) 36/42-60/66 h forecasts. Mean over all verification areas and both seasons

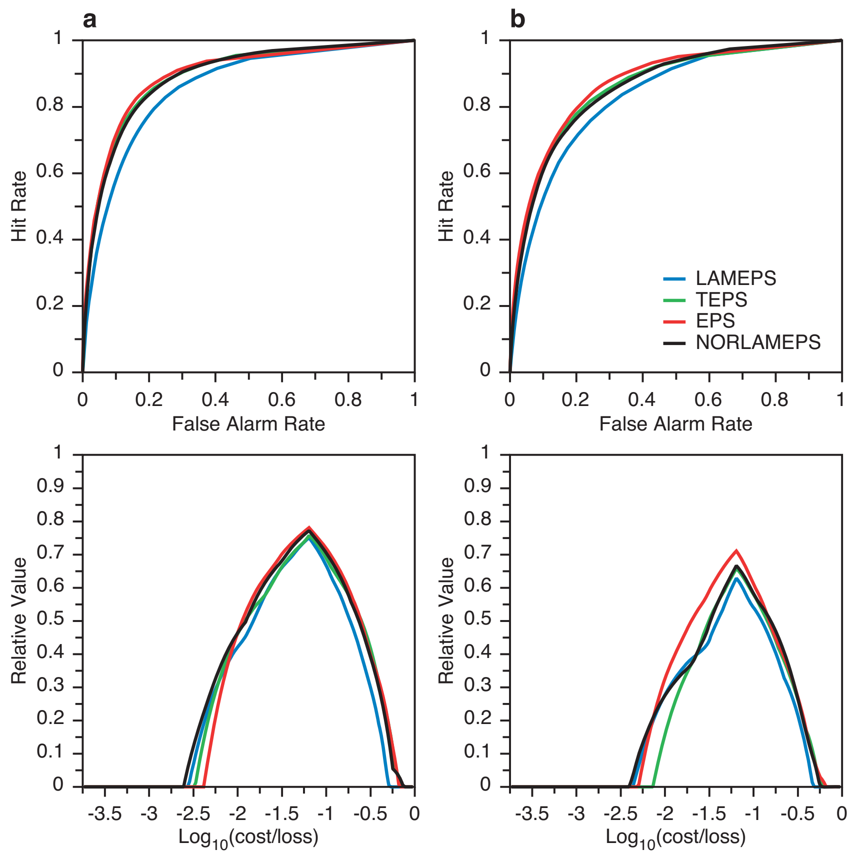

Fig. 4 **a)** ROC curves (top) and cost/loss analysis (bottom) for total 24 hour precipitation for 12/18-36/42 h forecasts. **b)** as **a)** but for 36/42 – 60/66 h forecasts. The threshold is 5 mm/day. Mean over all verification areas and both seasons. Blue – LAMEPS, green – TEPS, red – EPS and black – NORLAMEPS. NTOT is the total possible times the event can occur while NOCC is the number of times the event occurs in the sample. P_CLI is the total precipitation sample climatology.

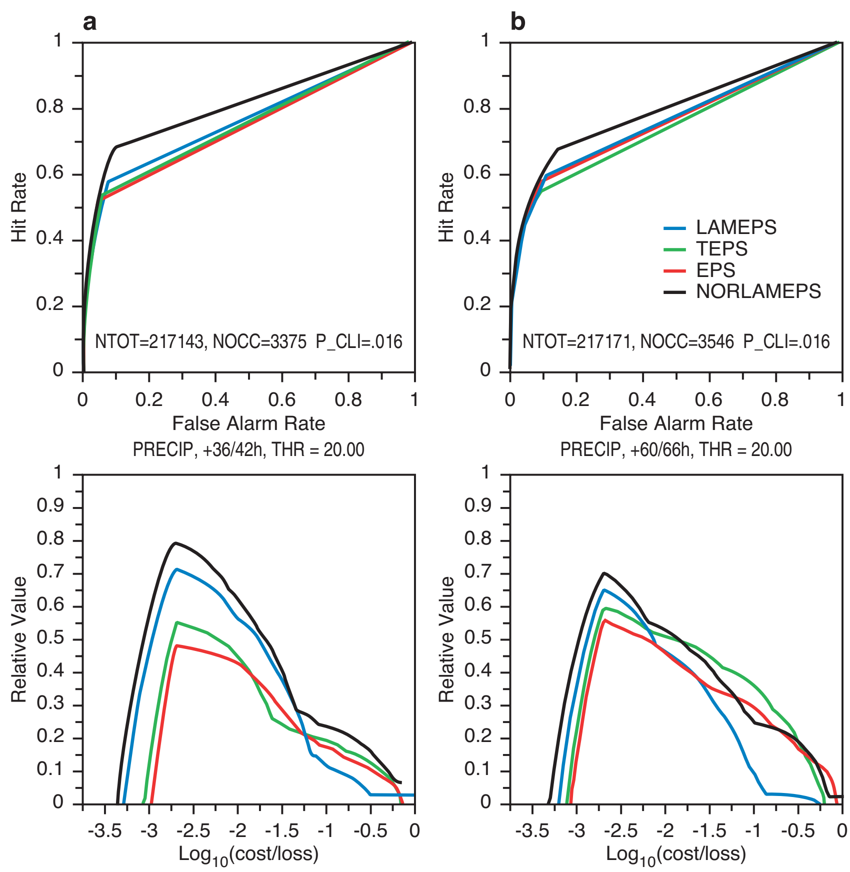

Fig. 5 As Figure 4, but event threshold 20 mm/day