Diagnosis of Mesoscale Forecasting Systems

By **Thor Erik Nordeng**

Norwegian Meteorological Institute, Oslo, Norway

1. INTRODUCTION

According to some definitions the mesoscale covers all scales ranging from 2000 km (meso-alfa), through meso-beta (200 to 20 km) to meso-gamma (down to 2 km). Due to major increase in computer power over the last years, even global models running daily in operational mode at major forecasting are capable of resolving the dynamics of the meso-alfa scales. As diagnosis of those models are covered elsewhere, we restrict our self in this paper to limited area models with horizontal grid resolution at 50 km and below.

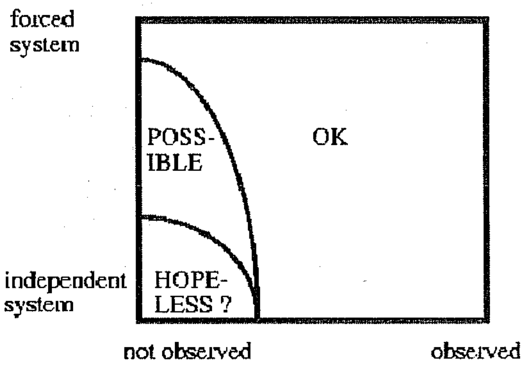
We have chosen a rather wide definition of 'diagnosis', i.e. model performance in terms of evaluation and verification as well as case studies. A more strict definition of diagnosis is to observe the performance of the model system in order to understand which parts of the model system that has to be improved.

Operational mesoscale modelling is performed to satisfy a wide range of objectives. An obvious scientific objective is to be able to forecast weather elements more detailed than can be done with global models. To tailor-make output for use in various application models and to be able to run according to own schedule are other important objectives. Finally, one should not forget that keeping some knowledge and expertise in the field of operational modelling at local centres may be important from educational reasons as well as a safety measure.

In this paper we give a flavour to the topic of mesoscale model diagnostics by a brief description of various mesoscale forecasting systems in use in European countries followed by a short discussion and an example of objective evaluation of one such system. We then turn to case studies demonstrating (but not proving) that MFS's are capable of describing small scale phenomena as they are constructed for. The next paragraph focuses on how to understand the model behaviour by use of diagnostics more or less as is standard routine for global models before we finally round off the paper by a short discussion. But first we give a brief review of the predictability issue for the mesoscales and its relation to verification.

2. PREDICTABILITY AND VERIFICATION

Due to the chaotic nature of the atmosphere deterministic predictions are limited in time. Small errors at the initial time of an integration grow exponentially and will finally dominate the forecast (*Lorenz*, 1969). The predictability range depends on the scale of the system under consideration. The smallest scales have the shortest predictability range. One may therefore argue that running mesoscale models with very high resolution is a waste of computer time since the smallest resolvable scales quickly will loose predictability. However, if the smallest scales are strongly forced with either the larger scale flow or fixed geographical features like e.g. coast lines, topography and ice edge predictability may be much longer than the time-scales given by Lorenz. *Nordeng* (1994) gave an example of a hurricane force mesoscale cyclone in the Bay of Biscaine which could be forecasted with more than 2 days lead time. All structures of the mesoscale low developed during those 2 days and the low developed because its large scale precursors were forecasted correctly. *Lilly* (1986) focused on helicity to explain why some flow apparently have longer predictability than would be expected from the Lorenz theory. Flow with high helicity (three dimensional wind vector parallel to the three dimensional vorticity vector) has less tendency to develop an inertial sub-range and therefore develops a three dimensional turbulence regime more slowly with the result of being more predictable. Mesoconvective systems tend to have high helicity, which may explain why they are so long lived. We would therefore expect that predictability is a function of initial state, resolution and forcing. A schematic plot of predictability for the mesoscales is shown in Fig. 1. Strongly forced flow is predictable. The same is true, but to a lesser degree, for signals which exist initially, but which is not strongly forced. This has important for our attempts to model the mesoscales. We need sophisticated analyses schemes to model the initial stage as accurately as possible as well as a forcing model (or underlying surface) which correctly describes the larger scale flow. In principle we have three regimes (see Fig. 1). Least predictability, if any at all, is found for unforced motion which is not present at the start of the integration.



(after A. M. Bratseth)

Figure 1: A schematic plot of predictability. The horizontal axis describes how well a phenomenon is observed initially, while the vertical axis describes to what degree the phenomenon is forced.

Model performance has to be assessed. Verifying mesoscale model systems are however not a straight forward task. Simply to adopt methods from synoptic scale models should be avoided simply because a detailed forecast is heavily penalized in terms of standard quality measures such as RMS error and correlation. As shown by Simmons (this volume) two sine waves having same amplitude and wavelength but being slightly more than 60 degree out of phase has larger RMS difference than we would obtain by comparing a sine wave with a flat (zero) distribution. In practical terms this means that if we try to forecast a high resolution phenomenon and succeed in terms of strength and size, but fails slightly in terms of position, we would be better off by forecasting nothing! Our standard scores are strongly biased towards smooth fields and it is difficult to show objectively that high resolution models can add valuable information. Some years ago at the Norwegian Meteorological Institute we run an experimental version of our limited area model in a quasi operational suite for the southern part of Norway. The horizontal resolution was 5 km. Subjective evaluation by duty forecasters gave the model a high score, but objective verification could not show improvements over our operational 50 km model. This also leads to the question about representativeness in terms of observations and model resolution. It may be necessary to develop methods to "filter" observations or develop verification methods which can recognize significant meteorological events and give them high scores if they are forecasted in "roughly" correct place at approximately correct time. An example from the 10 km resolution operational model at the Norwegian Meteorological Institute is shown in Fig.2 for an area just outside Oslo. Contour lines of topographic height used in the model together with true position of lakes and rivers are shown. There is a mismatch between the position of the relatively large lake in the middle of the map and the lowest topographic height. The river running out of its SW corner is even running uphill! Features like this will obviously give problems when trying to verify model forecasts against observations which to a strong degree are determined by local topography and land-use.

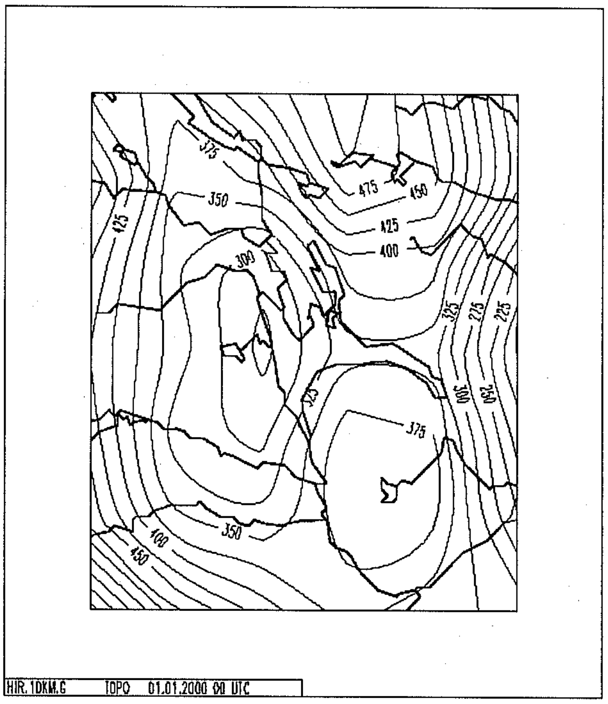


Figure 2: A section of the integration area for the HIRLAM10 model at the Norwegian Meteorological Institute for an area just outside Oslo showing topographic height, rivers and lakes. The plot covers an area of 5 by 6 grid lengths in the mode (50 by 60 km).

3. MODEL TYPES

A number of model types are used in forecasting the mesoscales, and it has to be so due to its wide definition. It is normally recognized that one needs a nonhydrostatic formulation of the forecast model when the grid-length is less than a few kilometres. It may e.g. be shown that for an unstable atmosphere the growth rate of convection increases beyond all limits for decreasing horizontal scales when using the hydrostatic assumption while it approaches the Brunt-Vaisala frequency for a nonhydrostatic assumption. In a stable environment with forced circulations scale analysis indicates that the hydrostatic assumption may be used even with as high horizontal resolution as 1 km provided that the horizontal wind is not too strong (less than 5m/s), Smith (1980).

A number of European countries are now running operational limited area mesoscale models with horizontal grid-resolution below 20 km at their national meteorological service, but only a few of those systems have their own data assimilation system. Some countries import initial conditions and lateral boundary conditions from another countries by bilateral agreement, while others run their own global or limited area model system with a coarser mesh to provide initial conditions and lateral boundary values. Those who attempt to perform a full data assimilation cycle employ nudging (e.g. Germany), optimal interpolation (e.g. Denmark) or variational methods (e.g. UK). In UK a preprocessing system for moisture is used to construct proxy data based on observations of cloud base and cloud top from the standard synoptic network as well as satellite and radar observations. At the time of writing only Germany (resolution 7 km) is running with the non-hydrostatic assumption.

4. MODEL PERFORMANCE IN TERMS OF OBJECTIVE VERIFICATION

As already discussed it is difficult to beat coarse resolution models in terms of standard objective verification methods such as RMS errors and correlation. Table 1 shows scores for the HIRLAM models running operationally in Norway (10 and 50 km resolutions). The errors are of the same order, but HIRLAM50 tends to be slightly better than HIRLAM10 for temperature while the opposite is true for winds. When land-sea contrasts, land use or major topographic features are important for describing the flow, the fine mesh models apparently gives important details. In fact, the numbrs show that HIRLAM10 is better than HIRLAM50 for winds in all seasons and for all three parameters. Fig. 3 shows wind roses from a light house at the coast of Norway. The high resolution model is able to add information as compared to the coarser mesh models for this site. For other sites where very small scale local topographic features are important a 10 km resolution may not score better than a coarser resolution model. Similar behaviour is reported from other countries running mesoscale model systems (not shown). High resolution models add information which is judged as important for duty forecasters, but the detailed information does not necessarily improve an objective score such as RMS errors. The Danish Meteorological Institute reports however that their high resolution model improves hit rates and false alarm rates for strong winds (Leif Laursen, personal communication).

SEASON	TEMPERATURE (2m)						WIND FORCE (10m)					
	HIRLAM10			HIRLAM50			HIRLAM10			HIRLAM50		
	ME	SDE	RMSE	ME	SDE	RMSE	ME	SDE	RMSE	ME	SDE	RMSE
Autumn	-0.5	2.3	2.8	-0.4	2.2	2.6	0.5	2.3	2.6	0.7	2.4	2.8
Winter	-0.5	3	3.8	-0.1	2.8	3.3	0.7	2.7	3.2	0.9	2.8	3.4
Spring	-1.7	2.3	3.2	-2.2	2.2	3.3	0.2	2.5	2.7	0.4	2.6	2.9
Summer	-1.6	2.3	3	-2.2	2.2	3.3	0.1	2.3	2.5	0.1	2.4	2.6

Average values for each station

SEASON	TEMPERATURE (2m)						WIND FORCE (10m)					
	HIRLAM10			HIRLAM50			HIRLAM10			HIRLAM50		
	ME	SDE	RMSE	ME	SDE	RMSE	ME	SDE	RMSE	ME	SDE	RMSE
Autumn	51	65	59	96	82	88	94	99	101	53	48	46
Winter	39	49	35	107	97	111	88	92	95	58	54	51
Spring	81	71	80	63	73	64	84	95	96	60	49	48
Summer	1-7	57	86	39	89	60	84	105	93	62	41	53

Number of stations where each model scores best

Table 1 . Performance of HIRLAM10 and HIRLAM50 as compared to Norwegian synop stations. Accumulated statistics started in autumn 1998. (ME = mean error, SDE = standard deviation, RMSE = root mean square error).

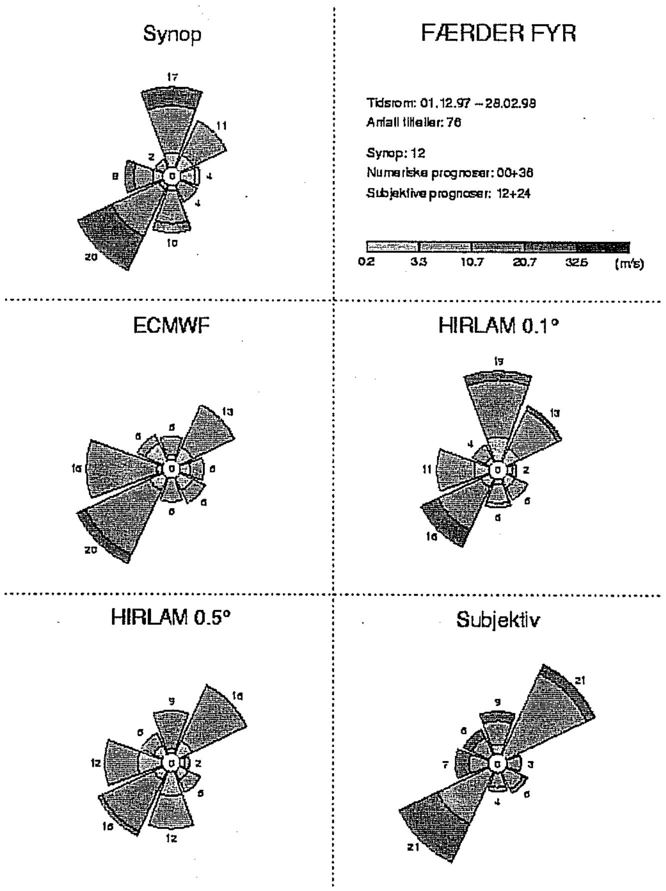
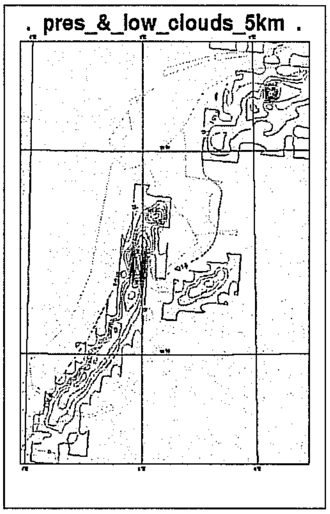


Figure 3: Wind roses from the Færder light house in the Oslo fjord for the period December 98 to February 99 showing observed wind (Synop) and forecasted wind from the ECMWF, HIRLAM50 and HIRLAM10 models. Subjective forecasts from duty meteorologists are also shown

5. MODEL PERFORMANCE IN TERMS OF CASE STUDIES

This kind of evaluation is more favourable for high resolution model systems as it focuses on synoptic (which involves subjective evaluation) rather than pure objective criteria. An obvious pitfall is to demonstrate good performance (and forget all the poor ones) rather than evaluate the model in terms of good and bad performance. The example chosen here is probably no exception!

Our example is taken from The Dutch meteorological Institute (KNMI) and shows forecasted low clouds over the coastal ares of the Netherlands (Fig. 4) with a 5 km resolution model and a 20 km resolution model. Fig. 5 shows a verifying satellite picture. The higher resolution model is capable of describing the coastal convergence from both sides of the peninsula between the North Sea and the Markermeer giving rise to the cloud band.



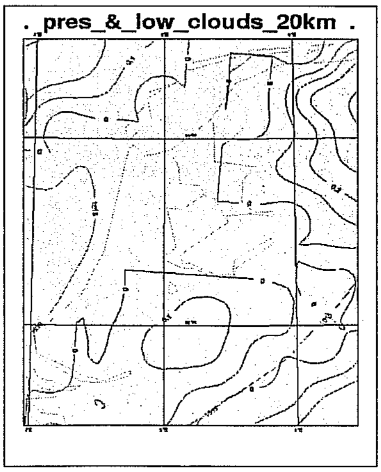


Figure 4: Forecasted low clouds over the Netherlands with a 5 km resolution model (top panel) and a 20 km resolution model (lower panel)

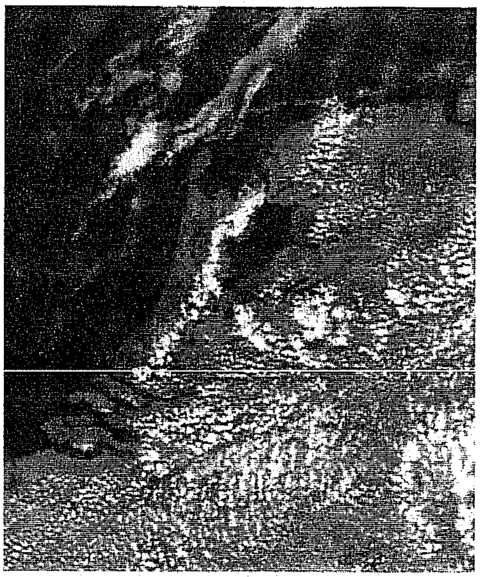


Figure 5: Verifying satellite picture for figure 4. Main feature to look at is the cloud band between the North Sea and the Markermeer.

It is easy to find good as well as poor examples of model performance. High resolution models tend to describe a lot of mesoscale features (as they are supposed to do), but our impression by subjectively evaluating these models over the years is that the described phenomena indeed are found in nature as well, not necessarily at the exact correct place at the correct time (see discussion in section 2), but close enough to give good guidance for the forecasters.

6. GLOBAL DIAGNOSTICS

It is standard routine to evaluate global models by running them in a climate mode, i.e. for several months, in order to check that the energy balance in the model is correct. The same exercise is usually not done for limited area models constructed for short range weather prediction though in principle it can be done. The outcome of such an exercise is however not conclusive. In their prognostic mode these models are forced by forecasted fields from a larger scale model (e.g. a global model) on the lateral boundaries. Results will strongly be biased towards the large scale solution which is imposed on the lateral boundaries and model errors may be concealed. Fig. 6 shows however an example of such a model evaluation taken from the Rossby Climate Centre in Sweden (SWECLIM) where the method revealed serious problems with the soil water scheme. The HIRLAM model is used as a regional climate model forced with lateral boundary values from a global climate centre (here: The Hadley centre, UK). The top panel shows a 10 year average from present day climate (control run) with the original soil water scheme while the bottom panel shows the same 10 year run but with a new scheme. The figures show net water balance (i.e. precipitation minus evaporation and runoff).

It's clear that the old scheme was seriously unbalanced and gave values over most of Europe in the range 50 to 200 mm per year. The new scheme however gives reasonable values in the range -20 to 50 mm per year. Note

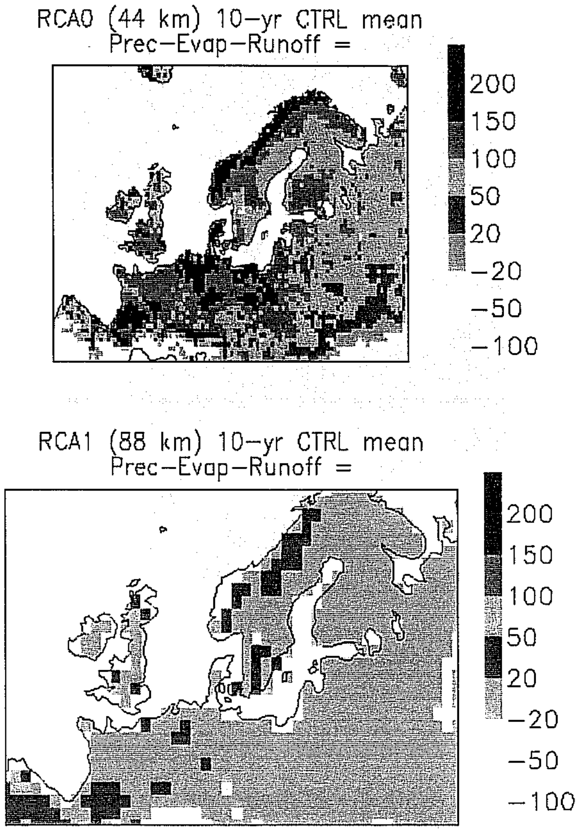


Figure 6: Net water balance in two versions of the Rossby centre regional climate model. Unit mm per year.

that different horizontal resolution was used but this did not have any important effect on the results, (Markku Rummukainen, personal communication). Such errors are difficult to detect in normal case study experiments or when running in an operational mode.

7. CONCLUSION

Standard procedure for evaluating mesoscale model systems is forecast verification and case studies. While case studies often demonstrates that high resolution models indeed improve mesoscale structures this is not as evident from objective verification. Standard verification methods penalize detailed forecasts. In addition we know that predictability is shorter for small scale dynamical systems. All in all this indicates that we should not expect to see major improvements in terms of standard objective scores when evaluating high resolution model systems. However, it have been demonstrated that for some parameters (e.g. wind) high resolution models do verify better than coarser resolution models.

With a few exceptions mesoscale model systems running operationally may be regarded as sophisticated dynamical and physical adaptation models. Detailed structures are developed during the forecasts from external large scale forcing, topography and land use. We should therefore not expect these models to perform significantly better than coarser resolution models. Full benefit from high resolution models can only be expected when high resolution observations can successfully be assimilated into the models. This involves constructing sophisticated assimilation schemes (probably 4D-VAR) with the capacity of assimilating unconventional data like clouds, total column water, rain rate, radar winds etc. Several research teams have started this work and results should be expected within a few years.

8. REFERENCES

Lilly, D.K., 1986: The Structure, Energetics and Propagation of Rotating Convective Storms. Part II: Helicity and Storm Stabilization. *J. Atmos. Sci.,* ***43****, 126-140.*

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

Nordeng, T.E., 1994: Numerical simulations of mesoscale substructures and physical processes within extratropical cyclones. In: The Life Cycles of Extratropical Cyclones, Vol. 1. Invited Papers. International symposium in Bergen, Norway, 27 June - 1 July, 1994.

Smith, R.B., 1980: Linear theory of stratified hydrostatic flow past an isolated mountain. *Tellus,* ***32****, 348-364.*