DATA ASSIMILATION FOR MESOSCALE MODELS

Nils Gustafsson
Swedish Meteorological and Hydrological Institute
Norrköping, Sweden

1 INTRODUCTION

Mesoscale atmospheric models are utilized for many purposes. For short range operational numerical weather prediction, such models are applied several times a day at grid resolutions ranging from a few kilometers to hundreds of kilometers. Mesoscale models are also applied for research purposes. In order to understand many mesoscale atmospheric phenomena it is necessary to apply models in order to simulate these phenomena. Regional climate simulation is another application area. For most mesoscale model applications, it is necessary to combine observed data with the abilities of the models to simulate the mesoscale atmospheric phenomena, i.e. it is necessary to apply mesoscale data assimilation. There are many similarities between the general data assimilation problem for e.g. global atmospheric models but several new problems will show up in mesoscale data assimilation.

The purpose of this paper is to review the status of data assimilation for mesoscale atmospheric models. Particular emphasis is paid on data assimilation for mesoscale limited area models, since such models are frequently applied by weather services for short range numerical weather prediction. We will start with a review of problems associated with mesoscale data assimilation in section 2 and operational and future observing system for mesoscale models will be reviewed and discussed in section 3. The status of development of mesoscale data assimilation techniques will be reviewed in section 4 and a particular review of activities within the international HIRLAM project will, finally, be given in section 5.

2 PROBLEMS IN MESOSCALE DATA ASSIMILATION

The task of data assimilation for mesoscale limited area models involves several particular problems in comparison with the general data assimilation problem for synoptic scale global forecast models. Examples of such particular problems are:

- Mesoscale data assimilation includes a wide range spatial scales. In addition to the synoptic scales of motion, we will have to treat also the mesoscale part of the spectrum. In general we will not have observations to recover the whole spectrum, so we will have to rely on model simulated features for the smaller scales of motion.

- In addition to circulation systems forced by flow instabilities, many interesting mesoscale phenomena are strongly forced by inhomogeneities in the lower boundary conditions, e.g. land-sea differences. This will require improved initial data on these lower boundary conditions and for some phenomena coupled oceanographic models including data assimilation will be required.

- Geostrophic adjustment theory tells us that the wind field becomes more important for the smaller horizontal scales. For the assimilation of wind information to be efficient, adjustment theory also tells us that the wind must be given for deep layer enough. Thus, access to high quality wind profile information becomes more important for mesoscale data assimilation. For shallow mesoscale boundary layer structures, mass field information is likely to be important, however.

- Mesoscale flow is in general less geostrophic and more divergent than synoptic scale flow. This makes initialization with traditional initialization techniques, like the non-linear normal mode initialization, less useful. In particular for mesoscale non-hydrostatic models, appropriate normal modes for such initialization are not available.

- Since mesoscale models generally are applied over regional areas only, lateral boundary conditions impose severe problems also with regard to data assimilation.

- Moist processes become increasingly important with increased spatial resolutions. This makes the assimilation of moisture parameters more important, in particular since the short forecast range of mesoscale model applications makes improved spin-up of moist processes crucial.

2.1 Forced mesoscale systems versus instabilities

Considering the basic mechanisms for their origin, one may distinguish between at least two different types of mesoscale circulation systems: (1) Circulation systems caused by internal flow instabilities, e.g. mesoscale low pressure developments and, (2) Circulation systems forced by inhomogeneities of the lower boundary conditions, e.g. sea and land breezes.

For the forecasting and simulation of these two different types of mesoscale circulation systems, we need different data assimilation strategies. The first type of mesoscale circulation systems requires a good description of the initial internal flow, i.e. the upper air wind, temperature and pressure fields, and in this respect the assimilation problem is not much different from the more general atmospheric data assimilation problem for synoptic scales of motion. The mesoscale instabilities may add requirements on refined spatial and temporal resolutions of the flow characteristics responsible for the mesoscale instabilities.

The second class of mesoscale circulation systems may put more particular demands on the mesoscale data assimilation, since we need to describe the lower boundary conditions carefully. Consider, for example, the mesoscale convective snowbands that occur in winter-time cold air outbreaks over the open water surfaces of the Baltic Sea. One example of such convective snowbands is shown in the satellite image given in Figure 1. It has been shown (Andersson and Gustafsson, 1994), that these convective snow-bands can be predicted by mesoscale models, provided the sea ice conditions and the sea surface temperatures are accurately described. For the particular case in Figure 1, the detailed geometry of the ice conditions in the western part of the Bay of Finland and in the archipelagos of Finland and Sweden turned out to crucial for the proper simulation of the major snowband along the Swedish east coast. The establishment of initial sea ice conditions, with fine details as those seen in Figure 1, is likely to require coupled oceanographic models including components for the sea ice. Such coupled models require also oceanographic data assimilation techniques to be applied.

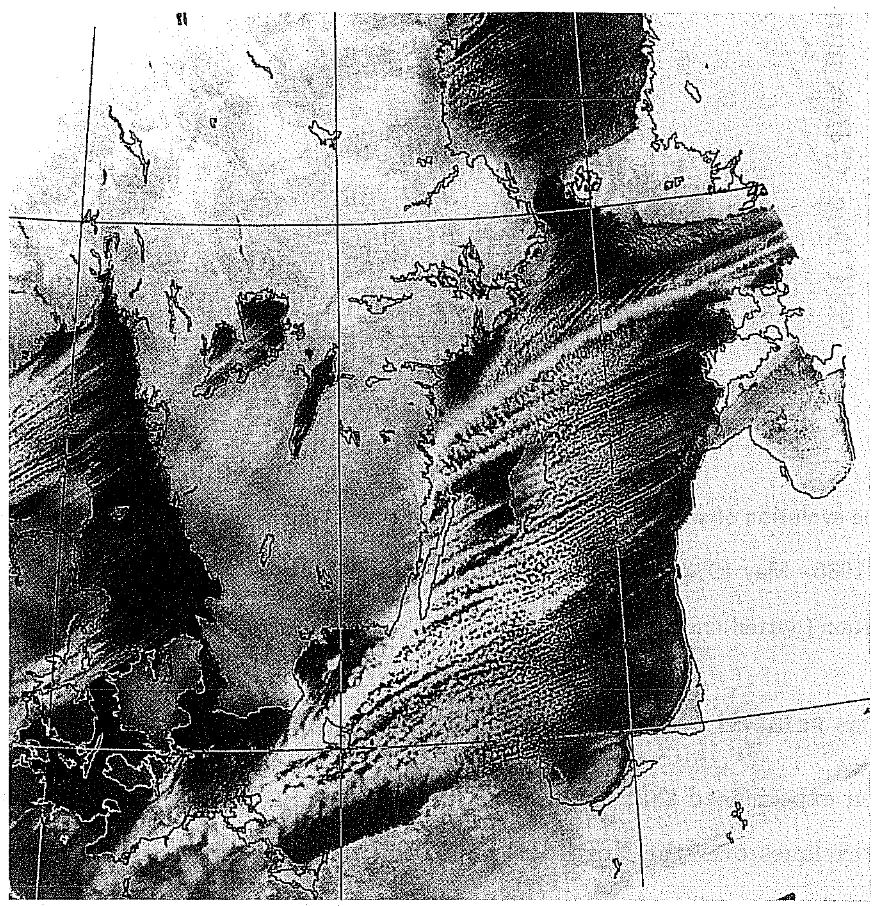

Figure 1: Infrared NOAA-9 satellite image 11 January 1987 12.35 UTC.

The Swedish Meteorological and Hydrological Institute (SMHI) is carrying out an atmospheric re-analysis exercise for the cold winter 1986-87 within the framework of the BALTEX experiment. A simple coupled oceanographic model was added to the atmospheric model for this re-analysis exercise. In the first re-analysis trial, no data assimilation was applied in the coupled oceanographic model and, as it may be expected, a slow drift away from the observed conditions in the Baltic Sea ice and water temperature conditions was experienced. Adding a simple assimilation of sea surface temperatures improved the situation significantly. Once the sea surface temperatures were right, also the forming/melting and advection of sea ice were simulated accurately. The time evolution of the sea surface temperature according to observations and according to the sea and ice model with and without data assimilation for one particular subarea of the oceanographic model is given in Figure 2. Notice the differences in the ice-covered period with and without data assimilation.

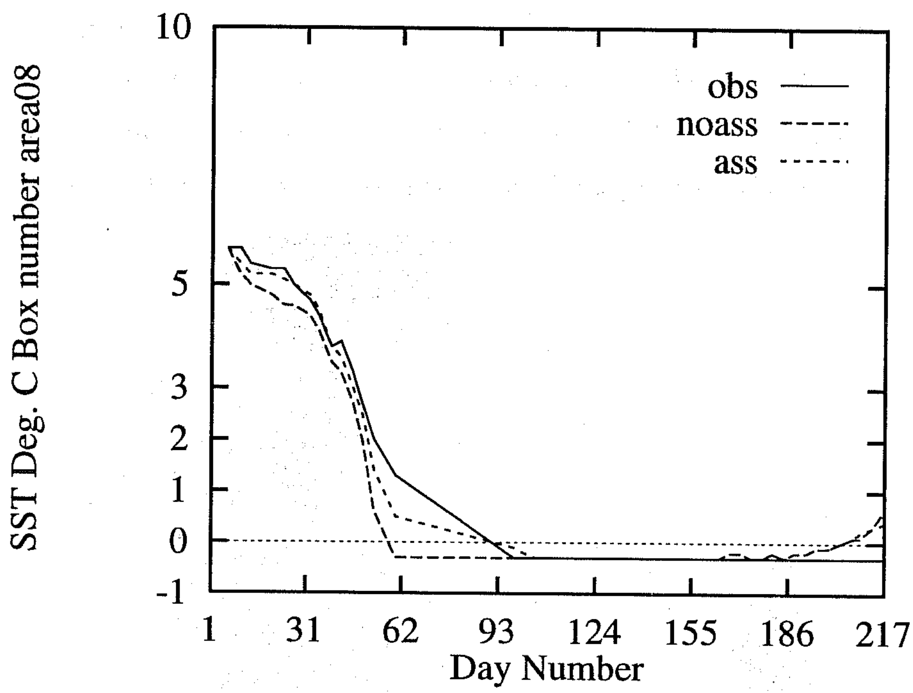

Figure 2: Time evolution of sea-surface temperatures in the subbasin "Archepelago of Finland" in the Baltic Sea during November 1986 - May 1987; observations (full line), model without data assimilation (dashed line) and model with data assimilation (dotted line).

2.2 Problems related to lateral boundary conditions

It has often been experienced that the limited area models in Europe have difficulties in forecasting fast developing cyclones over the North Atlantic. Some of these forecasts have been investigated in detail, and it has been possible to trace the origin of forecast failures, for example, to poor lateral boundary conditions. Several such forecast failures were discussed by Gustafsson, Lönnberg and

Pailleux (1996). Theoretical investigations on the propagation of initial data errors in numerical forecast model integrations were first discussed by Charney (1949) and in the recent monograph by Phillips (1990) the work done since then is summarized. Using maximum values of Rossby wave group velocities, Phillips found that

".... if a one-day forecast is to be made for a point near the ground in middle latitudes, the state of the atmosphere at the beginning of the forecast must be known within the following volume: 13 kilometers upward; 3 400 kilometers to the north and south; 3 400 kilometers to the east; 1 100 kilometers to the west. (The west and east boundaries of this region should be shifted westward by an amount equal to the average wind speed times one day to account for the advection by the mean zonal winds.) ...."

It has been demonstrated (Gustafsson, 1990), that the use of inadequate lateral boundaries during the data assimilation may have a more negative impact on the forecast quality than the use of inadequate lateral boundary condition fields directly during the forecast phase. The main reason for this is that the advection of observed information from data-dense areas (e.g. North America) to data-sparse areas (e.g. the North Atlantic) is replaced by the advection of lower quality ("old") forecast information during the data assimilation.

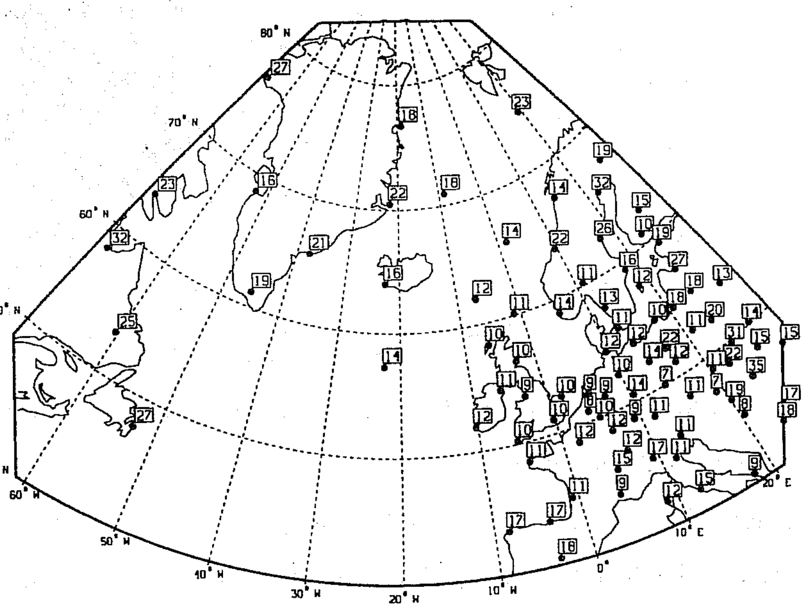

Figure 3: R.M.S. differences between 500 hPa radiosonde height observations and FMI 6 hour HIRLAM forecasts for January 1992 and 00 UTC initial data.

The first-guess for a limited area analysis is normally a six hour forecast. At the boundaries, the forecast values are a blend of the predicted values from the inner domain and of external boundary forecasts. The first-guess values near the boundaries might then be based on older forecasts than the first-guess in the inner area. Figure 3 and 4, provided by Peter Lönnberg FMI, show the rms-values of observed-minus-forecast values for radiosondes in the FMI HIRLAM area. The rms-values at the boundaries are significantly higher than those in the inner area. Even the error values in the North Atlantic area are clearly smaller than the North American values. The contribution to the high errors at the North American radiosonde stations must come from a poor background field as the observations themselves are of high quality. Figure 3 shows the situation in January 1992 with the first-guess compared to 00 UTC radiosonde data. The 00 UTC first-guess at the boundary was in January 1992 a 36 hour forecast from ECMWF. Corresponding statistics, but for 12 UTC data, are shown in Figure 4. The boundary first-guess values are a 24 hour ECMWF forecast in Figure 4. As expected the rms-values are much lower at the boundaries at 12 UTC than at 00 UTC, while the values in the inner area are comparable. The rms-values presented in Figures 3-4 support the hypothesis that large deviations between background field and observations might excite noise at the boundaries. Noise from the western boundary propagates quickly to the European continent and contributes to the fast error growth in the HIRLAM forecasts after 24-36 hours.

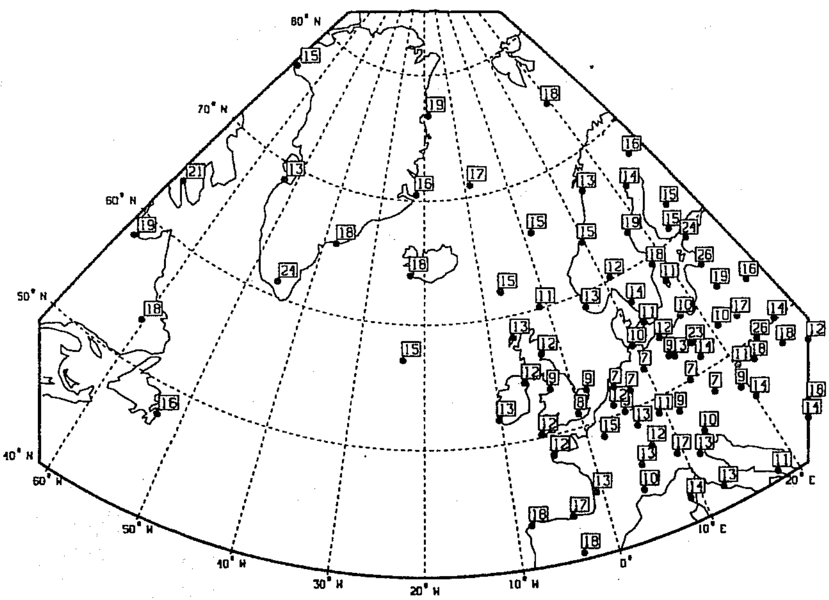

Figure 4: R.M.S. differences between 500 hPa radiosonde height observations and FMI 6 hour HIRLAM forecasts for January 1992 and 12 UTC initial data.

Operational experience suggests that the handling of the lateral boundaries and the use of small integration areas are significant error sources. Old boundaries generate significant errors in the data assimilation cycle. Global, or at least hemispheric, data assimilation may be necessary to achieve higher accuracy of the initial state. Most smaller weather services will have no choice, however, the only possible way to continue independent numerical weather prediction is by application of limited area models. Every effort must then be taken to obtain best possible lateral boundaries. One way out of this problem for the smaller weather services would be cooperation among several smaller weather services for running data assimilation systems over larger integration areas combined with national applications of very high resolution models, confined to smaller integration areas.

2.3 Problems related to moisture assimilation and spin-up

The time integration of each numerical forecast model from the initial analysis fields is associated with a "spin-up" of the dynamical and physical processes included in the model. The state of "balance" that is a result of this spin-up is quite a subtle one, and the time-scale for the spin-up can be considerable, depending on the nature of the physical and dynamical processes involved. Every time new observations are influencing the model state through the data assimilation cycle, this subtle balance is disturbed and the model integration has to go through a new period of spin-up. With the application of high resolution limited area models, we are interested in the quality of forecasts ranging from 6 hours to a few days. The short end of this forecast time range is well within the time period needed for spin-up of certain processes. Taking this conflict between spin-up and forecast range into account, it is of course necessary to take every possible measure to reduce the spin-up time and to avoid data assimilation procedures which destroy the subtle balance created by the model spin-up. An illustration to the spin-up of the operational SMHI HIRLAM mesoscale forecast model, including cloud water as an explicit forecast model variable, is given in Tables 1 and 2. The tables present results from verification of the predicted total cloudiness against satellite (AVHRR) derived total cloudiness (Karlsson, 1996). Table 1 gives the verification results for August 1994, when the cloud water field was initialized with zero values, and Table 2 gives the verification results for March 1995, when a 6 h forecast had replaced the zero values as an initialization for cloud water. The 6 h cloud water forecast obviously helps to reduce the spin-up of the cloudiness forecast, although the results should be interpreted with some caution, since the cloudiness characteristics in August generally are different from those in April, with increased dominance of clouds of convective origin.

	AVHRR	+06h	+12h	+24h	+36h	+48h
Cloudiness(%)	57.8	39.3	43.3	46.3	47.3	47.9
Bias(%)	-	-18.5	-14.5	-11.5	-10.5	-9.8
RMS(%)	-	19.8	15.7	13.1	12.6	12.0

Table 1: Comparison of monthly mean SMHI HIRLAM cloudiness forecast fields with monthly mean AVHRR derived cloudiness fields for August 1994.

	AVHRR	+06h	+12h	+24h	+36h	+48h
Cloudiness(%)	68.2	58.5	63.5	67.4	69.5	70.6
Bias(%)	-	-9.6	-4.6	-0.8	1.3	2.4
RMS(%)	-	11.0	6.8	5.2	5.7	5.6

Table 2: Comparison of monthly mean SMHI HIRLAM cloudiness forecast fields with monthly mean AVHRR derived cloudiness fields for March 1995.

3 OBSERVING SYSTEMS FOR MESOSCALE ASSIMILATION

We may generally state that conventional observing systems, like the radiosonde network, have poor spatial resolutions in comparison with the phenomena we want to describe in the initial data for mesoscale models. Therefore we need to try to utilize remote sensing data for mesoscale data assimilation. This adds significant difficulties to the data assimilation task, since the parameters measured by remote sensing techniques are most commonly not direct model parameters, like temperature, wind and humidity profiles measured by the radiosonde network. In addition, remote sensing observations are often associated with complicated observational error structures. Provided the appropriate assimilation techniques can be developed, however, remote sensing data have great potential for use in mesoscale data assimilation. We will review a few of these potential possibilities below. Other important sources of data for mesoscale models are:

- Aircraft data, in particular with regard to the rapid development of automatic data collection and processing systems such as ACARS. ACARS reports are collected within the framework of air traffic control systems. Wind and temperature profiles from take-off and landing of the aircrafts are available in addition to the single level data from the flight tracks. ACARS have been introduced operationally over the North American continent and a similar system is being built up over Europe.
- In the long term the use of satellite image information is an important data assimilation challenge in the context of e.g. 4-dimensional variational assimilation. In the short term, some empirical techniques may possibly be tried in order to improve the current humidity analysis or to improve the initial cloud representation in the model.
- Wind profilers
- High resolution TOVS data
- Scatterometer wind data
- Surface data from e.g. automatic weather stations with high resolution in space and time

3.1 Weather Radar Data

Weather services around the world have invested heavily in networks of weather radars, mainly for manual nowcasting and very short-range forecasting purposes. The output signal from weather radars is reflectivity caused by precipitation and some of these weather radars also produce doppler mode information. The reflectivity signal may be converted to precipitation intensity and the doppler mode signals may be used to produce radial wind vectors. Examples of weather radar networks are the NEXRAD system in the USA and the NORDRAD system covering large parts of the Nordic countries.

The potential for using weather radar for mesoscale data assimilation is obvious. Radar precipitation with a horizontal resolution of a few kilometers should be useful for e.g. diabatic initialization purposes and the radial wind vector information from doppler radars should be useful for recovering the wind field. The success in utilization of weather radar data for mesoscale data assimilation has been rather limited. This is so because of the very complicated error structures in radar information. To convert radar information to precipitation, for example, one needs to convert radar reflectivity to precipitation intensity, and this is not easy due to the dependence on size of the rain droplets which is unknown. One need also to eliminate enhanced reflectivity from melting ice and snow particles as well as reflectivity from the surface due to anomalous radar beam propagation.

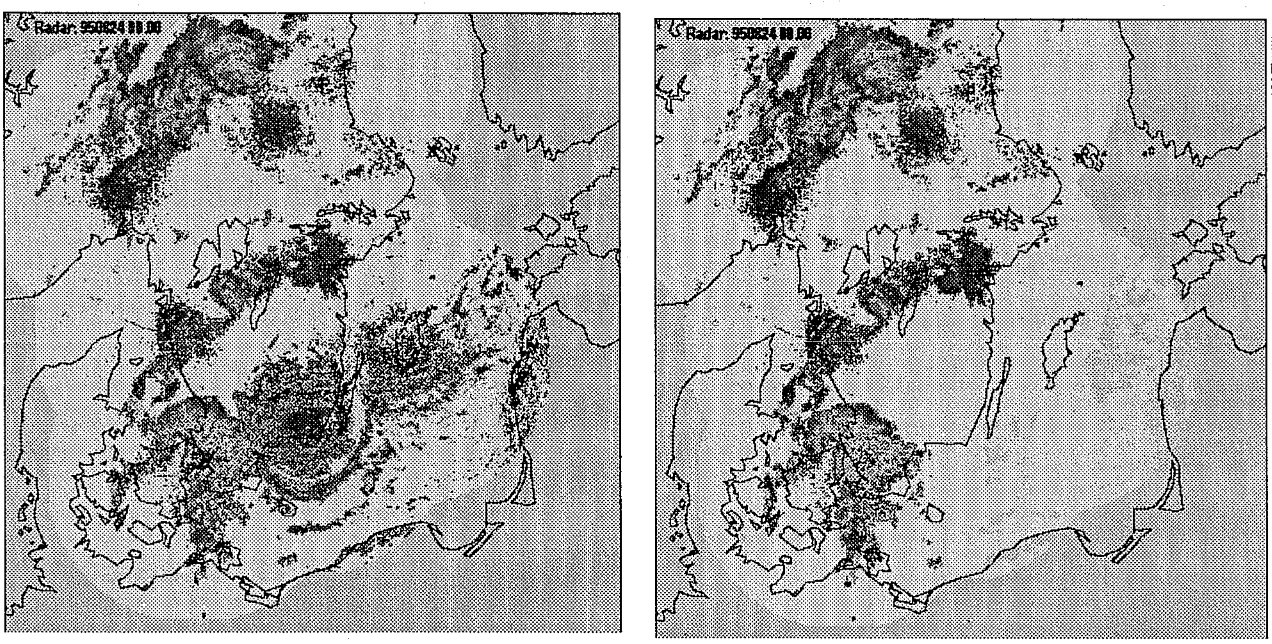

Figure 5: Precipitation data from the NORDRAD network, 24 August 1995 00.00 UTC. Without (left) and with (right) filtering of ground echoes due to anomalous radar beam propagation.

The SMHI has embarked on a mesoscale analysis project. The objective of the project is primarily to produce standalone objective analysis fields to be used for nowcasting and severe weather monitoring purposes. Precipitation is one of the most important parameters for this mesoscale analysis project and radar precipitation data is one of the most important input data sources, in particular over the data sparse water surfaces of the Baltic Sea. The use of radar data has caused serious problems with regard to quality control. Figure 5 illustrates some of these problems. In the original radar data, there is an intensive "real" precipitation area in the southwestern part of Sweden caught by the Copenhagen radar. In addition to this, there are false equally intensive radar precipitation areas indicated by the radars in the south-eastern corner of Sweden and on the island of Gotland. Notice also the false intense precipitation area along the coast of the Baltic states. The false precipitation patterns are, in fact, reflections from the ground, caused by anomalous radar beam propagation. In this particular case, it was possible to eliminate most of the false precipitation patterns by utilizing the doppler radar signals, indicating that the radar beams were reflected by the non-moving land and sea surfaces. In other cases, such an elimination algorithm is not sufficient, however, a particular problem is reflections from sea surface waves with non-zero doppler velocities.

The Doppler weather radars also produce radial wind vectors. Within the framework of variational assimilation, the radial wind vectors can be used directly together with dynamical and filtering constraints. Another possibility is to derive a single wind profile from each Doppler radar (VAD, Velocity Azimuth Display), by assuming horizontally homogeneous wind conditions at each vertical level in the vicinity of the radar. Such VAD-profiles can be obtained, e.g. every 15 minute, from each doppler radar site. Figure 6 (Tage Andersson, personal communication) illustrates the good agreement between VAD wind profiles and radiosonde wind profiles from observation sites close to Gothenburg in Sweden.

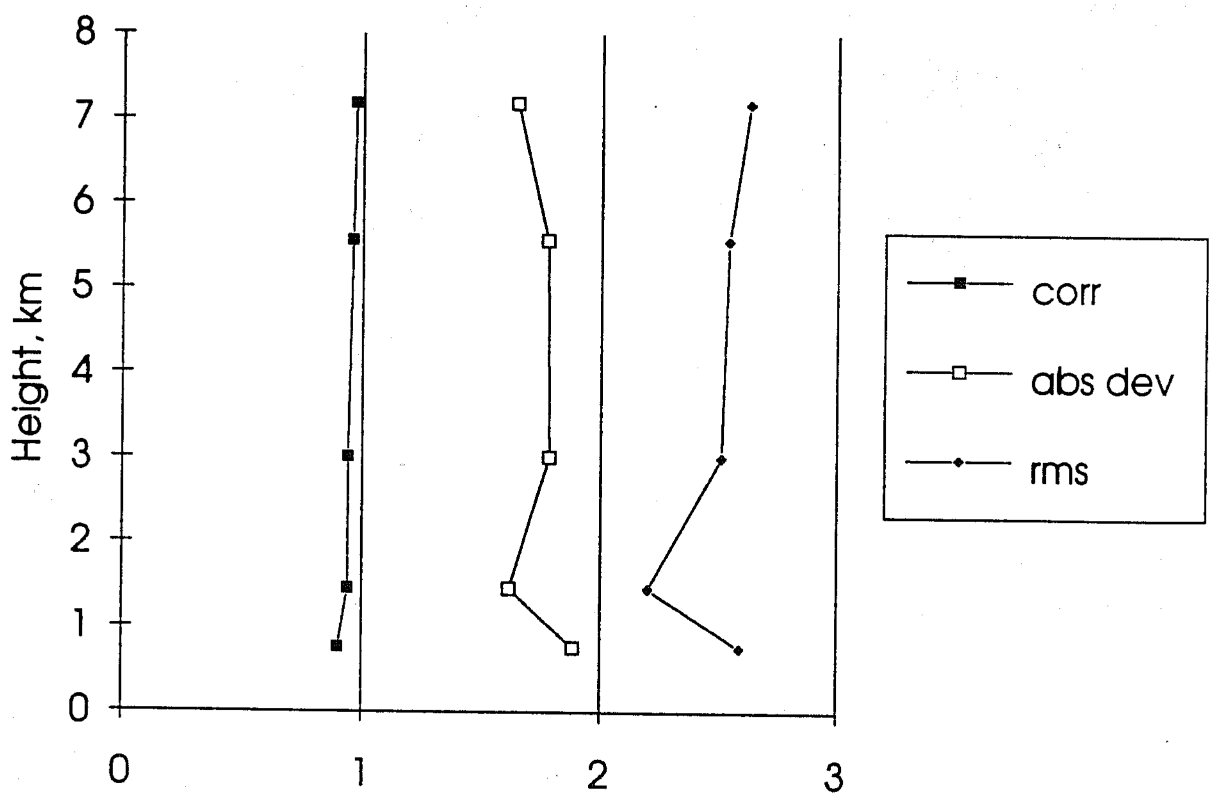

Figure 6: Comparison between VAD wind profiles from the Jonsered doppler radar and radiosonde winds from Landvetter. Correlation coefficients, mean absolute deviation and r.m.s. deviation are given for the scalar wind velocity. The data periods include 6 December 1994 - 14 February 1995 and 28 June - 30 November 1995.

3.2 GPS measurements of water vapor

Radio occultation of the terrestial atmosphere is now possible through the use of signals transmitted by the satellites of the Global Positioning System (GPS). By receiving the signals on satellites in low orbit, it is possible to derive accurate, high-resolution profiles of refractivity. Vertical profiles of temperature and water vapor can then be retrieved from these refractivity measurements. The radio wave propagation delay of the GPS signals received at GPS stations on the earth may also be used to determine very accurate vertically integrated water vapor measurements with high time resolution. This idea is being tested within the BALTEX experiment. Approximately 5 Finnish and 20 Swedish GPS stations were operated for this purpose during August - November 1995. Very accurate vertically integrated water vapor amounts, according to comparison with radiosonde and water vapor radiometer measurements, were retrieved. Figure 7 (Gunnar Elgered, personal communication) shows the time evolution of the integrated water vapor given by the GPS retrieval and by the microwave radiometer at the Onsala Observatory in western Sweden.

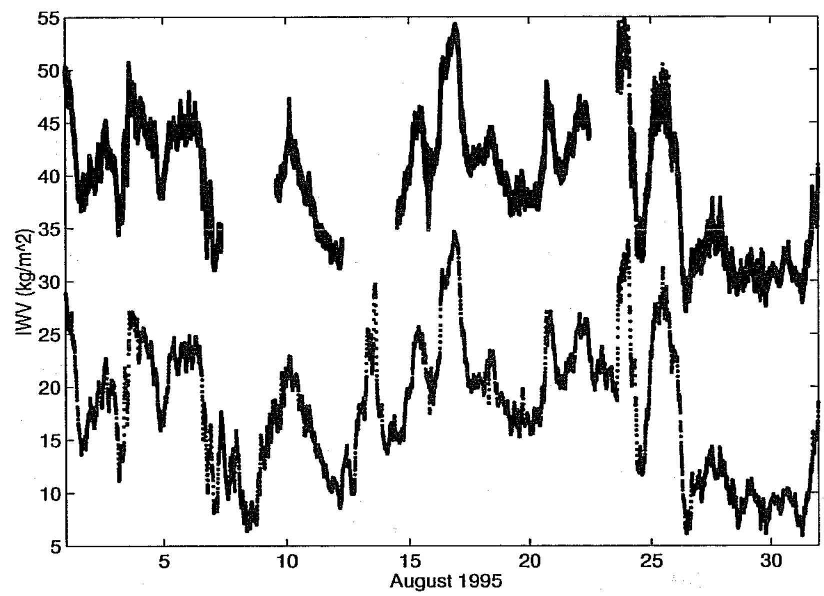

Figure 7: Integrated water vapor (IWV) given by a microwave radiometer (upper curve;offset $20kgm^{-2}$) and GPS retrieval (lower curve) at the Onsala observatory during August 1995.

Kuo et al. (1996) have investigated the possibilities of assimilating vertically integrated water vapor measurements into a mesoscale non- hydrostatic model by means of 4-dimensional variational data assimilation (4DVAR). The experiment was carried out as an observing system experiment. The special soundings collected during in SESAME (Severe Environmental Storms and Mesoscale Experiment) 1979 were used to construct a high quality upper-air analysis data set, from which "simulated" integrated water vapor measurements were generated. In addition, the vertical water vapor profiles were used for verification purposes. Although the experiment was carried out without any background error constraint in the assimilation, that indeed should help to distribute the integrated water vapor amounts in the vertical, and although the assimilation window in the 4DVAR experiment was as short as 1 hour, it was possible to demonstrate that a significant portion of the vertical variation of the water vapor profiles was retrieved by the 4DVAR procedure. In addition, it was shown that ground measurements of water vapor could also contribute to improve the retrieval of the water vapor profiles in the lower part of the troposphere.

4 MESOSCALE DATA ASSIMILATION METHODS

An overview of data assimilation techniques utilized for operational mesoscale numerical weather prediction is given in Table 3 below. It may be noticed that forward intermittent data assimilation based on optimum interpolation and non-linear normal mode initialization is most commonly applied operationally at the present time. The application of continuous data assimilation with observation nudging based on analysis corrections at the UK Met Office is the main exception.

Model group/ Weather Service	Objective analysis method	Initialization method	Assimilation scheme	Development effort
HIRLAM	3D Optimum Interpolation	Non-linear normal mode	Forward Intermittent	3DVAR, 4DVAR
France	3D Optimum Interpolation	Digital filter	Forward Intermittent	3DVAR, 4DVAR
Germany	3D Optimum Interpolation	Non-linear normal mode	Forward intermittent	Nudging
United Kingdom	Analysis corrections	Divergence damping	Nudging, Forward continuous	3DVAR, 4DVAR
USA, NCEP	3D Optimum Interpolation	Digital filter	Forward intermittent	3DVAR, 4DVAR
Canada	3D Optimum interpolation	Digital filter	Forward intermittent	3DVAR, 4DVAR
Japan	3D Optimum interpolation	Non-linear normal mode	Forward intermittent	
Australia	3D Optimum Interpolation	Digital filter	Forward intermittent	

Table 3: Overview of operational mesoscale data assimilation techniques applied for short range numerical weather prediction.

With regard to future developments, there is dominance for variational techniques. The German

Weather Service is an exception with nudging for their future mesoscale data assimilation. The mesoscale forecasting may be considered as a dynamical adaptation of a global forecast over a rather small area. For this purpose nudging, and in particular nudging of moisture parameters, may be an appropriate approach (Verner Wergen, personal communication).

4.1 Experiences from the HIRLAM data assimilation based on OI

The core of the present operational HIRLAM data assimilation system is the limited area version of the ECMWF Optimum Interpolation (OI) scheme for 3-dimensional multivariate analysis of the wind- and mass-fields and for univariate analysis of the humidity field (Lorenc, 1981, Lönnberg and Shaw, 1987). The adaptation of the global ECMWF OI analysis scheme for HIRLAM purposes was essentially of a technical nature - the analysis calculations needed to be modified from those of global latitude/longitude geometry to limited areas with rotated latitude/longitude geometry. Some initial sensitivity experiments indicated that the quality control algorithms, the analysis structure functions and the data selection algorithms of the global scheme performed reasonably well also for the HIRLAM purposes, at least for applications with horizontal grid resolutions of the order of 50 km. Without any further tuning of the analysis structure functions, the HIRLAM data assimilation system has also been applied operationally at grid resolutions of the order of 20 km and 5 km. Several attempts have been carried out to improve the present HIRLAM data assimilation scheme:

- Lönnberg and Eerola (1996) have collected information on the quality of radiosonde data in the European area by means of observation minus first guess statistics. It turns out that a significant number of European radiosonde stations are associated with un-acceptable systematic as well as random observational errors. A "gray" list of radiosonde stations has been constructed. These stations are subject to a more stringent background error quality control.

- The HIRLAM OI implementation assumes stationary statistics apart from a simple dependence of the forecast error variance on the analysis error variance of the previous cycle. This scheme is flow independent as it depends only on the data density of earlier analyses. In areas of moderately dense network of independent observations it would be possible to estimate some parameters of the background error covariances (assuming that the observation error statistics are stationary). The updated statistics would then influence the analysis as well as quality control decisions. Dee and Cats (1995) have investigated the use of a one-parameter covariance model in which the magnitudes of the observation and forecast error variances were adjusted by the same factor. The adjustment value was calculated separately for each analysis box. In this technique the ratio between observation and forecast error is kept constant; only the variances are adjusted by the same amount. Thus, the adjustment does not affect the weights given to observations, only the quality control decisions. This parameter estimation was tested for a few cases only, with a rather modest resulting impact.

- Within the framework of a new surface parameterization scheme for HIRLAM, a soil parameter data assimilation scheme is being developed based on ideas of Mahfouf (1991). This soil assimilation scheme requires input of 2 meter temperature and humidity analyses. An analysis scheme for these parameters, utilizing an-isotropic analysis structure functions taking land-sea and orography differences into account, has been developed (Beatriz Navascues, personal communication). One example of such analysis structure functions, with reference to a horizontal position situated in the River Ebro valley, is given in Figure 8.

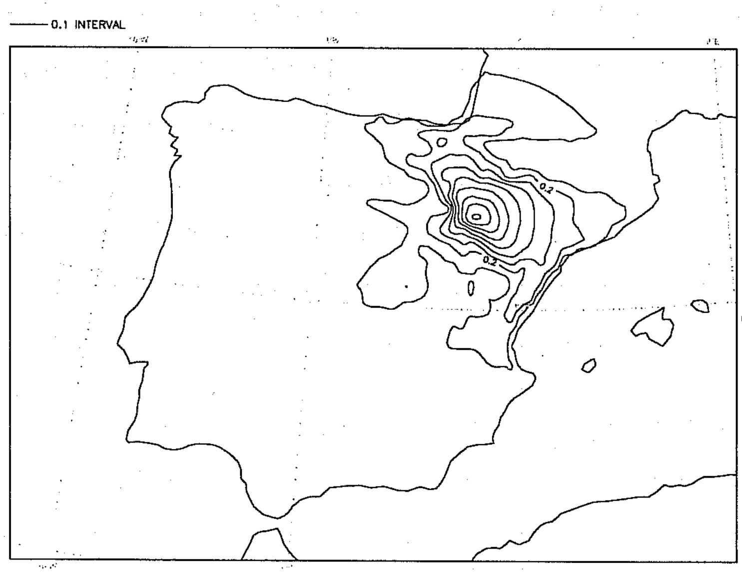

Figure 8: Anisotropic analysis structure functions for analysis of 2 meter temperature. The reference station for the correlation functions in the figure is situated in the River Ebro valley.

- An attempt to introduce flow-dependent analysis structure functions by carrying out the analysis in a geostrophic coordinate system has been tried at the Norwegian Meteorological Institute (T.-E. Nordeng, personal communication). The analysis first guess field was interpolated to a regular grid in geostrophic coordinates and the coordinates of the observations were transformed accordingly. The analysis was then carried out using the normal horizontally isotropic and vertically/horizontally separable analysis structure functions. After the analysis, the analysis fields were interpolated back to the normal coordinate system. Through the geostrophic coordinate transformation, a baroclinic vertical tilt was introduced in addition to a horizontal stretching, resulting in horizontally anisotropic structure functions along fronts etc. The first trial to utilize this analysis technique resulted in slightly improved baroclinic developments (T.-E. Nordeng, personal communication).

4.2 Mesoscale data assimilation in isentropic coordinates

Benjamin (1989) and his colleagues at the NOAA Forecast System Laboratory in Boulder have developed a Mesoscale Analysis and Prediction System (MAPS), specifically designed to handle 3 hourly surface and aircraft observations. The model as well as the analysis scheme of MAPS are formulated in a hybrid isentropic/sigma vertical coordinate. Isentropic vertical coordinates are of particular interest for analysis purposes, since the high vertical resolution of radiosonde observations may be used for a proper analysis of e.g. frontal structures and inversions also in between the positions of the radiosonde stations. A similar effect may be achieved by formulating the analysis structure functions in isentropic coordinates also within the framework of an analysis with the model variables in a pressure-based vertical coordinate.

4.3 Mesoscale assimilation based on nudging

In its simplest version the nudging technique only needs a forecast model and a mechanism to insert observed information at the nearest model point in space and time. Crude forcing of the information, i.e. replacement of the model value with the observed value, leads to rejection of the inserted information. The most important aspect of the nudging technique is the control of the projection into the model space. Most assimilation schemes based on nudging include a spreading of the observation in space by means of simplified objective analysis techniques. An assimilation scheme based on the nudging technique must also have an observation processing package and quality control modules. The level of sophistication of these modules needs to be similar to those used by any other technique. Any nudging which is applied to the model equations should be small and applied several times during a forecast for a time span of several hours. Specification of the weighting to be assigned at each time step requires major tuning. The local analysis in the vicinity of an observation should be balanced geostrophically and should filter the noise in the observations.

The U.K. Meteorological Office utilizes a data assimilation system based on nudging with the analysis correction method (Lorenc et al., 1991). The analysis correction method is a successive correction analysis scheme, designed to converge towards statistical interpolation results after several iterations. The iterations are carried out over several timesteps during the nudging process. The U.K. Meteorological Office has also developed a Moisture Observation Pre-processing System (MOPS) with the aim to improve moisture parameter initial conditions for the unified mesoscale model (Macpherson et al., 1995). Radiosonde relative humidity reports and SYNOP relative humidity reports are assimilated directly by the analysis correction method, the SYNOP relative humidity observations have proved to be beneficial in visibility forecasting. Other data sources, specifically cloud imagery, radar precipitation, cloud and precipitation observations from SYNOPs, are pre-processed to provide a 3-dimensional cloud field that is converted to a relative humidity field using the same algorithms as used by the model physics. The MOPS cloud data has been shown to provide better results in the earlier part of the forecast (<12 hour) of precipitation, cloud cover, cloud base and fog.

4.4 Mesoscale data assimilation based on variational techniques

Variational techniques for meteorological data assimilation were introduced by Sasaki (1958). In the early applications, variational techniques were mainly used for filtering of fields, that had already been put on a grid by objective analysis techniques. The filtering generally included various dynamical constraints, e.g. geostrophy or mass-conservation. With regard to mesoscale applications, variational techniques have been of particular importance for filtering and interpretation of radar wind information and other remote sensing data.

Since meteorological forecast models are able to accurately simulate many meteorological phenomena, it is a natural step to try to utilize the forecast model equations as constraints in data assimilation. This was made possible with the introduction of the concept of adjoint models in variational data assimilation by Le Dimet and Talagrand (1986) and by Lewis and Derber (1985). They showed that the data assimilation problem in NWP can be solved through the adjoint technique in a way which is much cleaner than current operational techniques. Then, since the mid 80s, an important effort has been put on variational techniques in research on data assimilation. A global 3D variational algorithm has been run operationally in Washington since 1991 (Parrish et al. 1992) and ECMWF runs a global 3D version of the variational assimilation of the IFS/ARPEGE (Courtier et al.,1993) operationally.

Errico and Vukicevic (1992) have developed the adjoint of the PSU-NCAR mesoscale model and they used this adjoint model successfully for studying the sensitivity of forecast errors to initial conditions. Zupanski (1993) developed the adjoint of the NMC eta level regional forecast model and used this adjoint for experiments with regional four-dimensional data assimilation. Zupanski added a term penalizing high frequency divergence oscillations in the cost function, and the control variable of the minimization included a model bias term in addition to the initial conditions. Zupanski and Zupanski (1995) have shown in some recent experiments that precipitation data may be used to improve short range precipitation forecasts. Within the framework of variational data assimilation, there is also a possibility to control the lateral boundary conditions. It is not clear, however, whether this is an advantage for operational short-range numerical weather prediction, since during the forecast step it is necessary anyhow to rely on lateral boundary conditions provided by e.g. a global forecast model.

4.5 Mesoscale initialization techniques

The non-linear normal mode initialization (Machenhauer, 1977) has been established as a robust and efficient initialization technique for global synoptic scale numerical weather prediction. This technique can also be applied with the diabatic processes included. Diabatic non-linear normal mode initialization has been applied successfully to mesoscale limited area models (Huang et al., 1994). In its diabatic form, non-linear normal mode initialization may also be forced by "observed" diabatic heating rates, derived from e.g. satellite image data and it has been shown that such forced diabatic normal mode initialization may improve e.g. precipitation forecasts during the first 12 hours of the forecast range. It is not clear, however, whether such diabatic normal mode initialization will improve precipitation forecasts beyond the first few hours of the forecast model integration. The normal mode initialization modifies only the gravity wave part of the initial data, and the Rossby wave part of the initial data may dominate the forecast after an initial adjustment period.

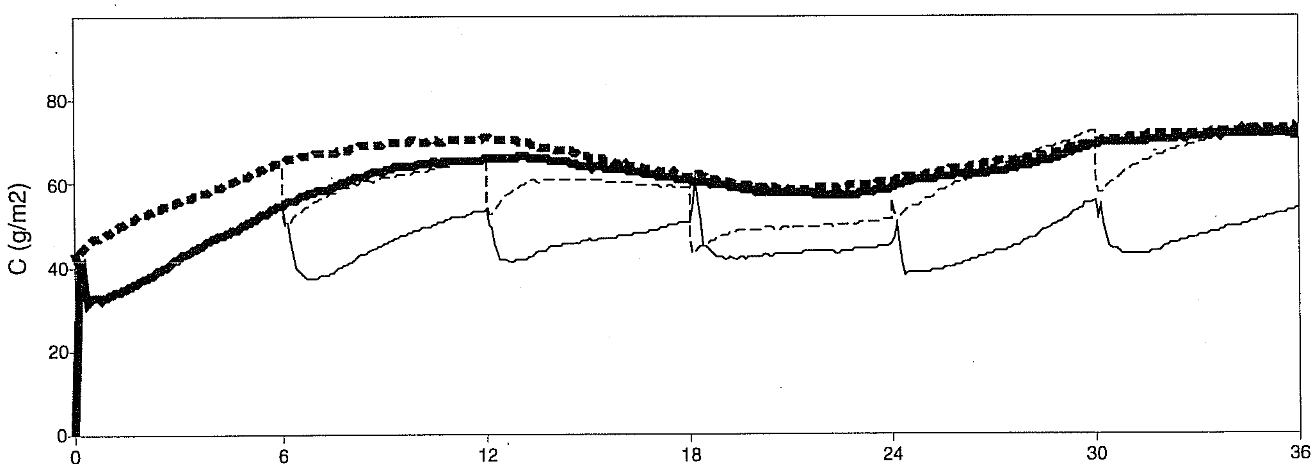

Figure 9: Vertically integrated cloud water averaged over the model domain C (gm^{-2}) as a function of time. The results for the adiabatic normal mode initialization are shown by full lines and the results for the diabatic digital filter initialization by dashed lines. The first 36 h forecast is given together with 6 h data assimilation results.

Digital filters (Lynch and Huang, 1992) offer another, more flexible, solution to the initialization of mesoscale limited area models. Since the normal modes of the forecast model need not to be known for digital filter initialization, digital filters are ideal for e.g. initialization of non-hydrostatic mesoscale models. Digital filters can be applied for diabatic initialization (Huang et al. 1994). Diabatic digital filter initialization has also been applied successfully to a model including cloud water as a explicit forecast model variable. Figure 9 is taken from the study by Huang (1996) and it shows that the spatially averaged cloud water amount during the 6 hour data assimilation cycles reaches a level similar to those during longer forecast model integrations, when the diabatic digital filter initialization is utilized. This is not the case when diabatic digital filter initialization is not applied.

5 VARIATIONAL DATA ASSIMILATION FOR HIRLAM

The design a new data assimilation system for HIRLAM was recently studied by Gustafsson et al. (1996). It was suggested that a new data assimilation for HIRLAM should be based on variational data assimilation techniques. Main arguments in favor of the variational techniques are (1) the possibilities to improve on initial baroclinic structures and (2) the possibilities for a more rational use of observations that are non-linearly coupled to the forecast model variables. The research and development work for a HIRLAM variational data assimilation started in February 1995. The target is 4DVAR, 4-dimensional variational data assimilation. The natural step towards 4DVAR is 3DVAR, 3-dimensional variational data assimilation. The following research and development tasks have been identified as necessary for the HIRLAM variational data assimilation:

- The tangent-linear and the adjoint of the adiabatic spectral HIRLAM model, including non-linear normal mode initialization, have been coded and tested successfully.
- The tangent-linear and adjoint of a few physical parameterization schemes have been coded. A general strategy for formulation of "adjoint" physics needs to be established.
- A software system for pre-processing of observational data and retrieval of observational data within the variational data assimilation has been be developed and will be maintained in collaboration between ECMWF, Meteo-France and HIRLAM.
- A first version of a background error constraint has been coded and tested.
- A basic software framework for HIRLAM 3DVAR and 4DVAR, including minimization with the INRIA package M1QN3, has been coded and tested.
- A weak digital filter constraint has been coded and preliminary tested.
- Observation operators are presently being tested. These include spatial interpolations, post-processing of various observed quantities, variationally based quality control etc.
- Extended tests of 3DVAR are expected to be carried out during 1997.
- A formulation of incremental 4DVAR needs to be developed.
- Extended tests of 4DVAR are expected to be carried out during 1998.

5.1 Tangent-linear and adjoint of the spectral HIRLAM

For the development of the first version of the adjoint HIRLAM, it was decided to use the spectral formulation of the model, see Gustafsson (1991). There are two reasons for starting with the spectral version: (1) The spectral version of HIRLAM is a more "modern" code based on Fortran 90 and utilization of automatic arrays, (2) In general, it is easier to develop adjoints of spectral models since Fourier transforms are self-adjoint and since no efforts are needed to develop adjoints of complicated finite difference operators. A manual coding technique was used to develop the first version of the adjoint of the adiabatic part of the spectral HIRLAM including horizontal diffusion. The tangent-linear and the adjoint of the non-linear normal mode initialization were developed as well. For each subroutine containing any non-linear expressions, the corresponding tangent-linear subroutine was first coded. Then the adjoints of each tangent-linear (and originally linear) subroutine were coded in a statement-by-statement fashion. By considering each statement of the tangent-linear and linear subroutines as a complex matrix operator, the corresponding adjoint statement(s) were derived by taking the complex conjugate and transpose of this matrix operator. An important and very time-consuming phase in the development of the adjoint code by this manual technique was of course the testing and verification of the correctness of each subroutine.

5.2 The background error constraint

For the background error constraint in the first version of HIRLAM 3DVAR, the control variable vector is transformed in such a way that the background error covariance matrix of the transformed control vector becomes diagonal. The transforms are similar to the transforms applied in the ECMWF 3DVAR (Courtier et al., 1993). Fourier transforms as applied in the spectral HIRLAM are used in the horizontal. Eigen-vectors of vertical forecast error correlation matrices are used for the transformations in the vertical. In order to obtain initial data for HIRLAM that are well balanced, a constraint on the balance between the wind field and the mass field analysis increments is needed. For the ECMWF 3DVAR, this is done by projection on the normal modes of the forecast model, followed by stronger penalty on the fast gravity modes than on the Rossby modes. This technique is not easily applicable to the HIRLAM model, since the HIRLAM Rossby modes used for the normal mode initialization are based a constant Coriolis parameter and this would result in a poor balance constraint. One alternative would have been to apply the NMC approach (Parrish et al., 1992) based on increments of vorticity, divergence and the un-balanced part of the mass field as control variables. Again, this is not so easily applicable for the HIRLAM problem, since the transformations between wind field components and vorticity/divergence are not unique for a limited area. Therefore, as a first trial for the HIRLAM 3DVAR, we are using ageostrophic wind component increments and the full mass field increments in the control vector. This approach has the further advantage of avoiding the non-unique transformations between temperature/surface pressure and linearized geopotential.

In order to obtain the model variable vector X from the control variable vector χ the following series of inverse transforms are applied

$$X = X_b + T^{-1}\chi = X_b + A^{-1} S F^{-1} V^{-1} P^{-1} \chi \tag{1}$$

where the forward transforms are defined by

- Subtraction of the background field X_b to calculate the increment field $\delta X = X - X_b$
- A: Calculation of ageostrophic wind component increments by subtraction of geostrophic wind component increments, as determined from the tangent-linear massfield increment gradients
- S^{-1}: Normalization with forecast error standard-deviations
- F: Fourier transform to spectral space
- V: Vertical transform, i.e. projection on the eigen-vectors of the vertical forecast error correlation matrices
- P: Pre-conditioning by normalization with a horizontal spectral density function and the vertical eigen-values

From the definition of the covariance matrix B for the background error and the transformation T we may derive

$$B = E < (X - X_b)(X - X_b)^* >= T^{-1} E < \chi\chi^* > (T^{-1})^* = T^{-1}(T^{-1})^* \tag{2}$$

where the operator $E <>$ stands for mathematical expectation. Multiplication of the covariance matrix B with a model state step function vector δ, with zeroes in all elements besides in the position of one particular model variable in one particular spatial position, will result in a vector COV containing the covariances between that particular variable and spatial position and all the variables in the model vector. The transformation form $T^{-1}(T^{-1})^*$ of the covariance matrix B provides a practical way to carry out this calculation. The implied covariances are given by

$$COV = T^{-1}(T^{-1})^* \delta \tag{3}$$

In Figure 10 we have plotted a selection of implied background error covariances with reference to the u-component of the wind at level 8 in one horizontal position. From the (auto-)covariances with respect to the u-component in other horizontal positions but at the same level, we can notice the elliptic-shaped covariance structure, typical of non-divergent wind component covariances utilized for traditional OI analysis schemes. In the vertical cross-section for the u-component we can notice the decrease of the covariances in the vertical. Plotted are also the (cross-)covariances between the u-component and the temperature. Here we can notice a consistent geostrophically induced structure in the temperature field, as derived from a single u-component observation. Also the effects of the tropopause can be noticed. These implied structures give us some confidence in the correctness of the utilized transformations.

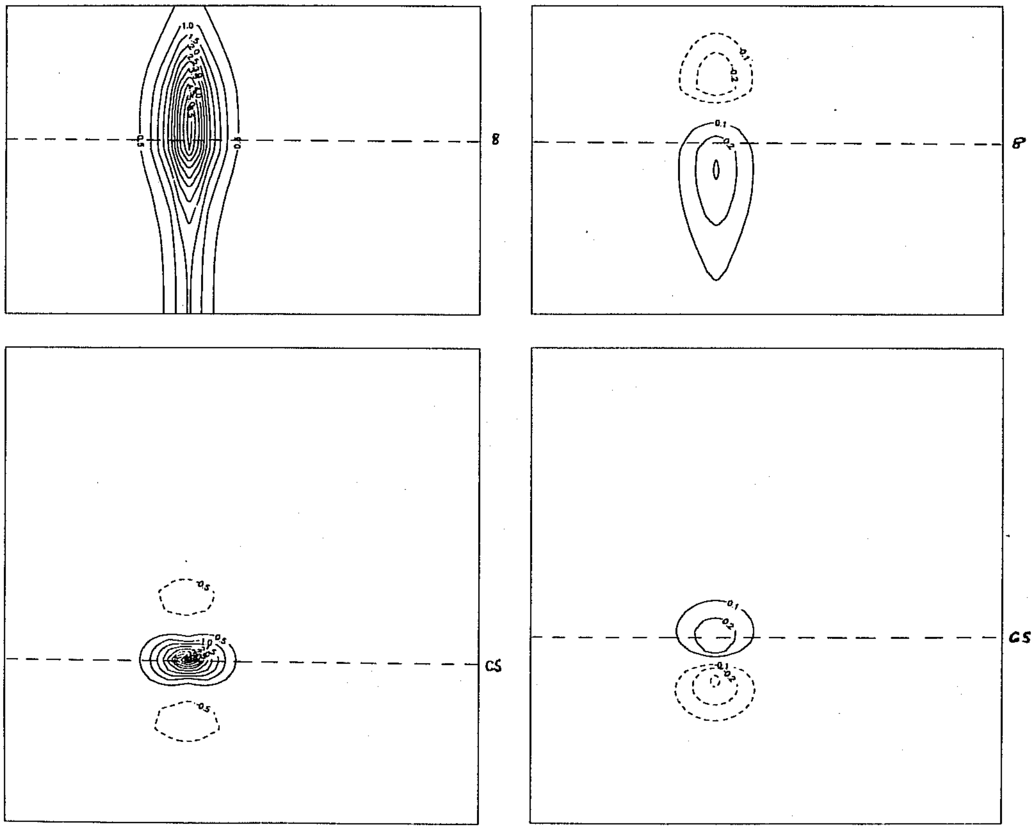

Figure 10: Selection of implied background error covariances with reference to the u-component in gridpoint (20,18) at level 8. The (auto-)covariances for the u-component are given in the figures to the left and the (cross-)covariances for the temperature are given to the right. The lower figures show the horizontal covariances at level 8 and the upper figures vertical cross-sections of covariances through the horizontal lines marked CS in the figures below.

5.3 A weak digital filter constraint

Digital filters have successfully been applied for initialization of numerical weather prediction models (Lynch and Huang, 1992). Digital filters may also be included as weak constraints in 4DVAR (Gustafsson, 1992). A weak digital filter constraint formulation has been coded and tested for the HIRLAM 4DVAR. Consider a digital filter applied over the time step number interval (-n,n):

$$X_0^{DF} = \sum_{t=-n}^{n} g_t X_t \tag{4}$$

where X_t denotes the un-filtered model solution, X_0^{DF} the filtered model state at the mid-point of the time interval and g_t the digital filter weights. For the 4DVAR we would like the solution to be obtained by the minimization process to vary slowly in time. This means that we would like the filter to have a very small effect, if applied on the final solution. This can be formulated as the following constraint J_c to be minimized:

$$J_c = \frac{\gamma}{2}(X_0 - \sum_{t=-n}^{n} g_t X_t)^2 = \frac{\gamma}{2}(\sum_{t=-n}^{n} \tilde{g}_t X_t)^2 \tag{5}$$

where the second expression has been obtained by a slight re-definition of the digital filter weights. The gradient of this quadratic constraint with respect to the model state variables follows easily

$$\frac{\partial J_c}{\partial X_t} = \gamma \tilde{g}_t \sum_{r=-n}^{n} \tilde{g}_r X_r \tag{6}$$

A problem with this weak digital filter constraint is the selection of the arbitrary relative weight γ given to the digital filter constraint. The sensitivity of the filtering effects with respect to the value of this weight needs to be tested.

5.4 Sensitivity experiments and "Poorman's 4DVAR"

One successful application area for adjoint models is sensitivity studies (Rabier et. al. 1995). The adjoints of numerical weather prediction models are used to relate the forecast error, measured by the difference between a forecast and its verifying analysis, to a small perturbation (sensitivity field) in the initial state. Adding the sensitivity field to the initial state, a new forecast (sensitivity forecast) is then run. Using both analyses and observations for verification, Gustafsson and Huang (1996) demonstrated the consistent improvement in the sensitivity forecasts. The improvement made by sensitivity forecast can not be considered as an improvement in a real forecast sense due to the use of the "future" analysis. However, it is evident that the quality of the sensitivity forecast up to the analysis time is improved. In other words, even in a real forecast sense, the sensitivity forecast can provide an intermittent data assimilation system with better first-guess fields, which may lead to improved analyses. Based upon these arguments, the following Poorman's 4DVAR has been tested: (1) Starting from t=-6h, the nonlinear forecast model is integrated forward to produce the preliminary first-guess field at t=0. (2) With the observations collected around t=0 and the preliminary first-guess field, the OI analysis is performed to produce the preliminary analysis. (3) Using the difference between the preliminary first-guess and the preliminary analysis as input, the adjoint model is integrated backward in time to produce the sensitivity field at t=-6h. (4) Adding the sensitivity field to the analysis at t=-6h, the nonlinear forecast model is integrated forward in time (sensitivity forecast) to produce a new first-guess. (5) The final OI analysis is performed based on the new first-guess and observations.

The major motivation for the poorman's 4DVAR is the ability of the backward adjoint and forward non-linear integration cycle to act as a dynamical filter on the preliminary OI analysis increments. The analysis increments of the final OI analysis application are smaller and, consequently, the ensuing forecast would be associated with a less serious spin-up of e.g. baroclinic structures as well as diabatic processes.

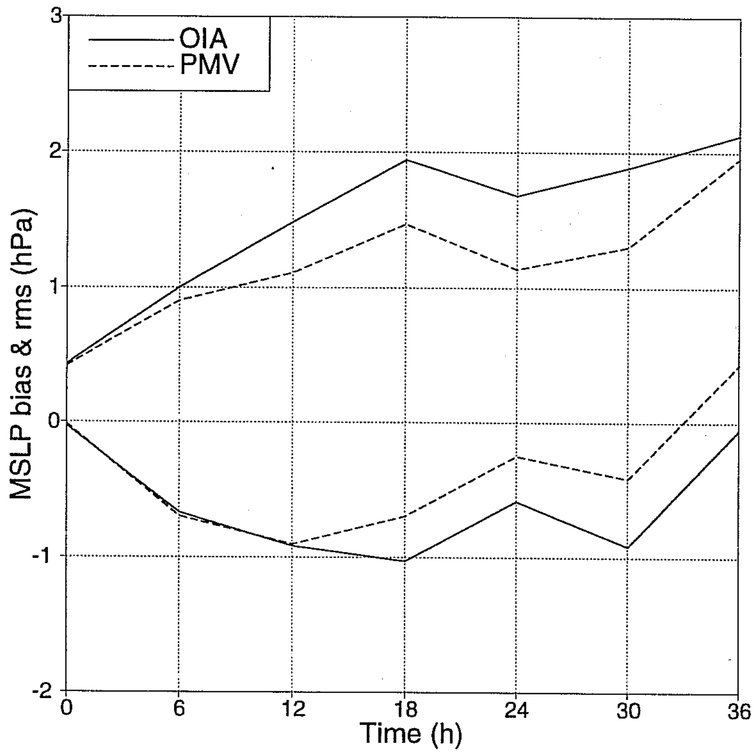

Figure 11: Observation verification of mean sea level pressure as a function of forecast length for standard (OIA) and Poorman's 4DVAR (PVM) initial data for the 36 h forecasts starting from 1200 UTC 14 September 1994. The lower two curves are for the bias. The upper two curves are for the rms.

The idea described above for utilization of the adjoint model was tested for a 5 day period, 13 - 18 September 1994, and the results were compared with those from the standard OI-based intermittent data assimilation (Huang et al. 1996). In general, it was proven that the analysis increments were smaller during the final analysis cycle, after the increments from the backward adjoint and forward non-linear sensitivity forecast runs were added to the analysis first guess fields. It also turned out the forecasts from the analyses based on these improved first guess fields gave slightly better verification scores, the improvement in predictability at the 36 hour forecast range was of the order 3-6 hours. In particular, the forecast of one fast low pressure development over the European area was significantly improved with regard to the position of the low pressure system. A closer inspection of the differences between the two forecast runs, revealed that the backward adjoint and forward non-linear sensitivity forecast runs added an increment to the first guess field with a vertically tilting baroclinic structure. Mean sea level pressure forecast scores for this particular forecast event are included in Figure 11. We may conclude that application of an adjoint model may be used to improve real time operational forecasts also within the framework of OI-based intermittent data assimilation.

6 CONCLUDING REMARKS

Particular problems related to data assimilation for mesoscale limited area models were reviewed, some aspects of observing systems of potential importance for such data assimilation were discussed and an overview of available data assimilation techniques with emphasis on mesoscale assimilation was also given. From this review of problems, observations and assimilation techniques we may conclude:

- High quality lateral boundary conditions are highly needed for mesoscale data assimilation and forecasting.
- We need to get access to improved observing systems for moisture parameters and we need to improve moisture assimilation (and initialization) techniques.
- 3-dimensional and 4-dimensional variational assimilation techniques provide good technical frameworks for utilization of remote sensing data.

The status of development of the HIRLAM variational data assimilation was reviewed. Results with sensitivity experiments with the adjoint of the spectral HIRLAM, as well as experiments with the use of the adjoint HIRLAM model within the framework of an OI-based intermittent data assimilation system, indicate potential abilities of 4-dimensional variational data assimilation to improve forecasts of baroclinic developments also on shorter time scales.

7 References

Andersson, T. and N. Gustafsson, 1994: Coast of Departure and Coast of Arrival: Two Important Concepts for the Formation and Structure of Convective Snowbands over Seas and Lakes. *Mon. Wea. Rev.*, **122**, 1036-1049.

Benjamin, S., 1989: An Isentropic Mesoα-Scale Analysis System and Its Sensitivity to Aircraft and Surface Observations. *Mon. Wea. Rev.*, **117**, 1586-1603.

Charney, J., 1949: On a physical basis for numerical weather prediction of large-scale motions in the atmosphere. *J. Meteor.*, **6**, 371-385.

Courtier, P., J.-N. Thepaut and A. Hollingsworth, 1994: A Strategy for operational Implementation of 4D-VAR, using an incremental approach. *Q.J.R. Meteorol. Soc.*, **120**, 1367-1387.

Courtier, P., Andersson, E., Heckley, W., Kelly, G., Pailleux, J., Rabier, F., Thepaut, J-N., Unden, P., Vasiljevic, D., Cardinali, C., Eyre, J., Haseler, J., Hollingsworth, A., McNally, A. and Stoffelen, A., 1993: Variational Assimilation at ECMWF. ECMWF Technical Memorandum No. 194, September 1993.

Dee, D. and G. Cats, 1995: Tuning OI covariance parameters: Experiments with the HIRLAM system. In proceedings of the WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Tokyo, 13-17 March, pp 225-228.

Errico, R. M. and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU - NCAR mesoscale model. *Mon. Wea. Review*, **120**, 1644-1660.

Gustafsson, N., 1990: Sensitivity of limited area model data assimilation to lateral boundary condition fields. *Tellus*, **42A**, 109-115.

Gustafsson, N., 1991: The HIRLAM Model. Seminar proceedings on Numerical Methods in Atmospheric Models. Volume II, pp. 115-146. Available from the ECMWF.

Gustafsson, N., 1992: Use of a digital filter as weak constraint in variational data assimilation. Workshop proceedings on Variational assimilation, with special emphasis on three-dimensional aspects. pp 327-338. Available from the ECMWF.

Gustafsson, N., Lönnberg, P. and Pailleux, J. 1996: Data assimilation for high resolution limited area models. *J. Meteorol. Soc. Japan*, in press.

Gustafsson, N. and Huang, X.-Y. 1996: Sensitivity experiments with the spectral HIRLAM and its adjoint. *Tellus*, **48A**, 501-517.

Huang, X.-Y., Cederskov, A. and Källen, E., 1994: A comparison between digital filtering initialization and nonlinear normal mode initialization in a data assimilation system. *Mon. Wea. Review*, **122**, 1001-1015.

Huang, X.-Y., 1996: Initialization of cloud water content in a data assimilation system. *Mon. Wea. Review*, **124**, 478-486.

Huang, X.-Y., Gustafsson, N. and Källen, E., 1996: Using an adjoint model to improve an optimum interpolation based data assimilation system. *Tellus*, in press.

Karlsson, K.-G., 1996: Validation of modelled cloudiness using satellite-estimated cloud climatologies. *Tellus*, in press.

Kuo, Y.-H., Zou X. and Y.-R. Guo, 1996: Variational assimilation of precipitable water using a nonhydrostatic mesoscale Adjoint Model. Part I: Moisture Retrieval and Sensitivity Experiments. *Mon. Wea. Review*, **124**, 122-147.

Le Dimet, F.X. and O. Talagrand, 1986: Variational Algorithms for Analysis and Assimilation of Meteorological Observations. Theoretical Aspects. *Tellus*, **37A**, 309-322.

Lewis, J.M. and J.C. Derber, 1985: The Use of Adjoint Equations to Solve a Variational Adjustment Problem with Advective Constraints. *Tellus*, **37A**, 309-322.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A.C., Bell, R.S., and B. Macpherson, 1991: The Meteorological Office Analysis Correction data assimilation scheme. *Quart. J. Roy. Met. Soc.*, **117**, 59-89.

Lynch, P. and Huang, X.-Y., 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019-1034.

Lönnberg, P. and D. Shaw, 1987 (editors). ECMWF Data Assimilation - Scientific Documentation. ECMWF Research Manual 1. 10/87 2nd Revised Edition.

Lönnberg, P. and K. Eerola, 1996: Treatment of non-Gaussian observation errors in the HIRLAM system. Workshop proceedings on Non-linear Aspects of Data Assimilation, ECMWF, Reading, U.K., 9-11 September 1996, to appear.

Mahfouf, J.-F., 1991: Analysis of soil moisture from near-surface parameters: A feasibility study. *J. Appl. Meteor.*, **30**, 1534-1547.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253-271.

Macphersson, B., B.J. Wright, W.H. Hand and A.J Maycock, 1995: The Meterological Office mesoscale data assimilation scheme. Forecasting Research Division, Scientific Paper No. 33, Available from Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, United Kingdom.

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Phillips, N., 1990: Dispersion processes in large-scale weather prediction. WMO - No. 700. World Meteorological Organization, Geneva, Switzerland

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. 1995: Sensitivity of two-day forecast errors over the Northern Hemisphere to initial conditions. *Q.J.R.Meteorol.Soc.*, **122**, 121-150.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. of Met. Soc. Japan*, **36**, 77-88.

Zupanski, M., 1993: Regional Four-Dimensional Variational Data Assimilation in a Quasi-Operational Forecasting Environment. *Mon. Wea. Review*, **121**, 2396-2408.

Zupanski, M. and D. Zupanski, 1995: Recent developments of the NMC's regional four-dimensional variational data assimilation system. In proceedings of the WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Tokyo, Japan, 13-17 March 1995, pp 367-372.