The Arome mesoscale project

François Bouttier

Météo-France CNRM/GMAP
42 Av Coriolis 31057 Toulouse, France
francois.bouttier@meteo.fr

ABSTRACT

This paper discusses current issues and plans for mesoscale numerical weather prediction (NWP) modelling. The broader NWP context is reviewed, with examples from the French Arome project, which targets strong convective events (figure 1) using a non-hydrostatic, kilometric-resolution model derived from the IFS/Arpge/Aladin and Mso-NH software. The data assimilation involves the latest techniques used in synoptic scale assimilation, plus specific developments linked to the use of relevant observing systems, structure functions and coupling with a large-scale model. Some additional complexity comes from the need to consider coupled models for specific user needs.

1 Introduction

The Arome (Applications of Research to Operations at MEsoscale) project aims at providing useful predictions of severe fine-scale weather by running data assimilation and forecasts at convection-resolving (i.e. better than 3km) horizontal resolutions, on limited areas and at the one- to two-day range. It is one of several similar projects currently pursued by several NWP organizations:

- the Weather Research and Forecasting Model (WRF) in the USA, (Michalakes et al, 2001, and `http://wrf-model.org/`)
- the mesoscale version of the Unified Model (UM) at the British Met Office (`http://www.metoffice.com/research/`),
- the Consortium for Small-Scale Modelling (COSMO) lead by the Deutscher Wetterdienst (`http://www.cosmo-model.org/`) and participating countries,
- the Arome project driven by Mto-France (`http://www.cnrm.meteo.fr/aladin/`), the French Mso-NH model research community (`http://www.aero.obs-mip.fr/mesonh/`) and countries in the Aladin group,
- several development projects are planned in the Hirlam group: Canada (Laprise et al 1997, Côté et al 1998), Japan, etc.

Specifically, the plan of the Arome project is the development of a new model with non-hydrostatic dynamics and detailed microphysics suitable for explicit simulation of violent convective activity, and a suitable data assimilation system that uses fine-scale observing datasets. This objective raises new issues compared to existing synoptic-scale (and usually global) data assimilation and forecasting systems: fresh approaches are needed for the verification of forecasts, data processing and quality control, the use of forcings from global models, the connection with nowcasting systems, and software organization. The required software is just as complex as in global NWP systems. In order to save manpower and to use the best expertise, the main originality of Arome will be the joint involvement of research and NWP teams in order to produce the best possible system.

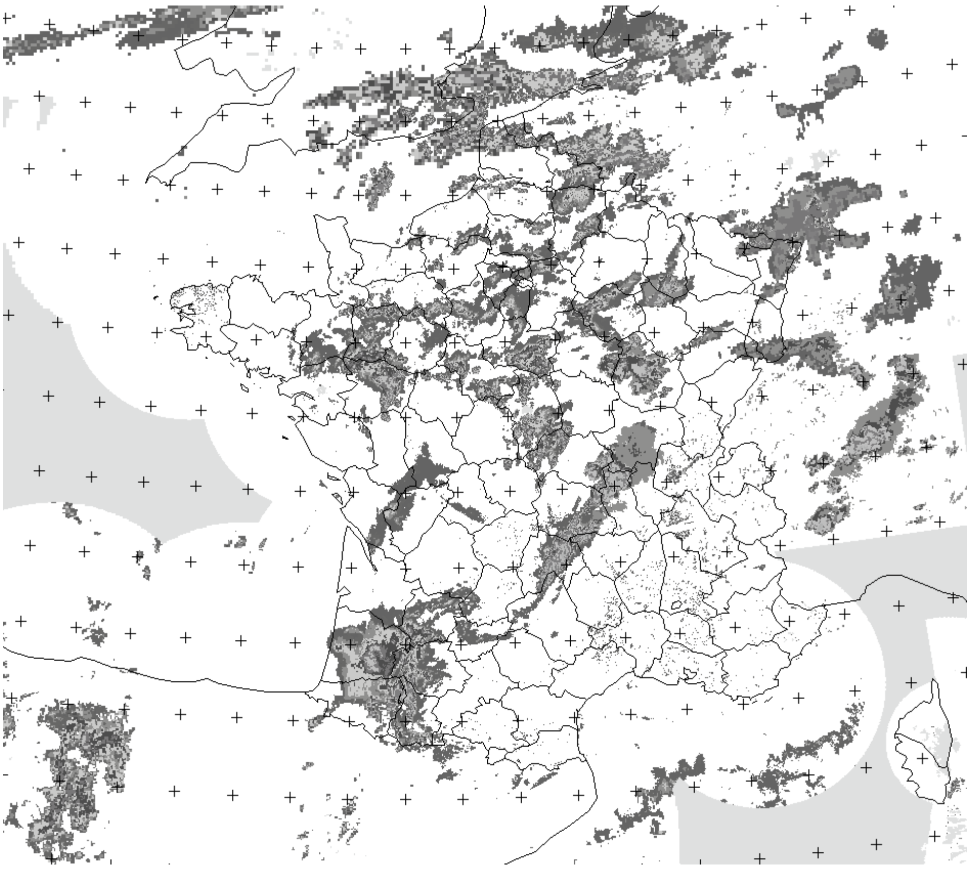

Figure 1: Rain rate over mainland France and its surroundings in an ordinary summer convective situtation (28 August 2003, 15UTC) that can imply locally hazardous weather in the strongest precipitating thunderstorms (here, near the center of France, heading Northeast). It is estimated by processing observed radar reflectivities, which are available at about 2km resolution, and can be assimilated after suitable quality control into kilometric-scale models. The black lines depict administrative borders.

Arome has grown because of several simultaneous causes. The growing awareness of the importance of weather in the public and private sectors has led to demands for better weather forecasts at finer scales and shorter ranges than medium-range NWP. Also, a better integration is needed between NWP centres and hydrological agencies, civil protection, environment monitoring bodies, etc., which is only possible if everyone works on models of compatible resolutions and degrees of physical realism. The continuing rise in computing power and the deployment of high-resolution (from 1 to 30km) operational observing networks will soon make it feasible to issue daily runs on domains the size of a medium country (of the order of 1000x1000km) using models similar to the so-called convection-resolving models (CRMs) already used by the mesoscale research community. The key is that, due to the vertical aspect ratio of convective clouds, convection is much easier to model at horizontal resolutions better than 3km than in the 3-9km range, where convection is partly subgrid, and partly resolved: this makes it difficult to get, say, the right intensity of convective precipitation and the shape of convective cells in this resolution range.

Apart from the computing cost aspect, one could say that a 2-km scale model is easier to design than a 5-km scale one because of this convection resolution problem. The actual modelling issue is a bit more complex, since (1) there still are some interesting phenomena that can be modelled without problems in the 3-9km resolution range, like dry adaptation of the atmosphere to orography, and (2) there is some hope that current research on this resolution range will eventually lead to convection schemes that solve existing problems. There are, however, some overwhelming advantages to focusing on higher resolutions:

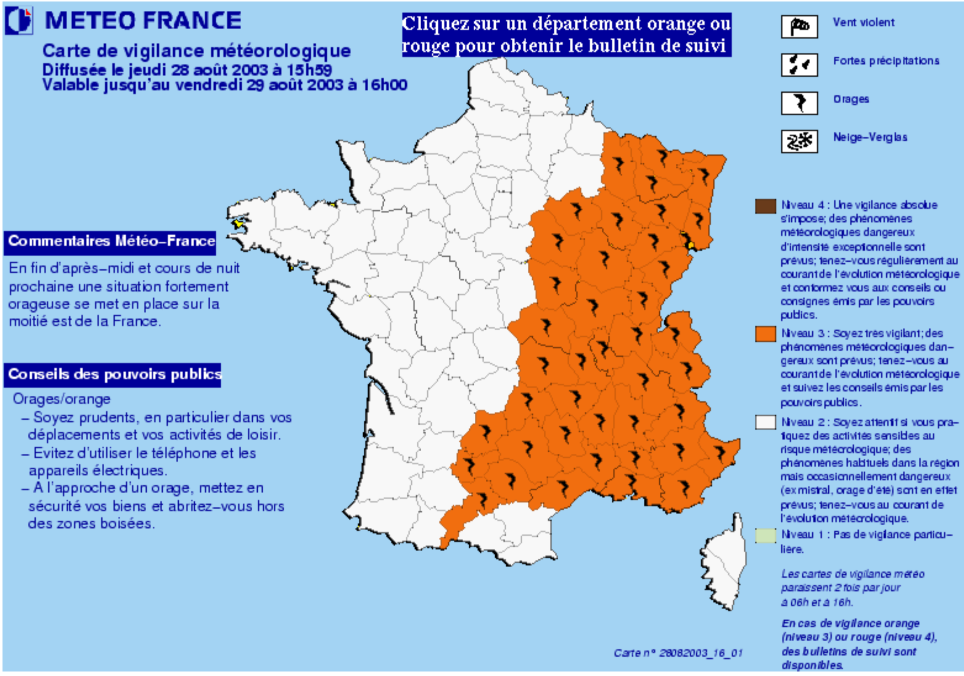

Figure 2: Meteorological awareness product publicly released by the Mto-France forecasters on `www.meteo.fr` *for the next 24 hours on 28 August 2003, 14UTC. The grey levels on the territory indicate levels of danger on a scale of 1 to 4 (highest danger). On this situation, the West and South are on level 2, the rest is on level 3 due to the risk of heavy thunderstorms. Better short-range, fine-scale NWP systems are necessary to help delineate the risk areas, timing in a more useful way.*

- using the same tools as research teams makes scientific interaction with them more efficient,
- explicit representation of convective clouds facilitates their analysis and initialization,
- a 2 to 3-km model resolution is consistent with the planned resolution of observing systems such as geostationary satellites and radars,
- these resolutions are well suited for coupling with ancillary models for hydrology, nowcasting, urban air quality, vegetation, etc.
- computer evolution means that regional NWP at resolutions worse than 3km will have lost much of its appeal (except for the geographically larger countries) by 2010 in comparison with the evolution of planned global models e.g. at ECMWF. Higher resolutions are more attractive for the long term as soon as we have the observations to initialize them.

There is a special problem with forecast ranges and cutoff practice (i.e. forecast production delays). Most existing NWP models provide little user value at very short ranges (less than 6 hours). This may be due to spin-up problems leading to unrealistic cloud and precipitation forecasts, to the use of very long cutoffs, or to the lack of use of observations of quickly-evolving phenomena such as satellite or radar images. Briefly put, very short-range NWP is not yet competitive with nowcasting based on image visualisation. There have been some attempt to bridge the gap between nowcasting and NWP e.g. in the Nimrod post-processing system (Golding 1998). An ambition of the Arome project is to try and apply short-range NWP techniques directly to nowcasting systems through direct assimilation of radar and satellite image data at nearly full resolution and real time.

The key motivation for Arome may thus be summarized as follows:

- to provide useful input to the preparation of meteorological awareness products,
- to significantly improve (with respect to available large-scale models) analyses and forecasts of strong convective events and associated phenomena: precipitation, gusts, hail, in terms of location and timing,
- to improve model output of actual weather parameters, including visibility, turbulence, temperature, local effects,
- to maximize working efficiency by seeking synergies between relevant institutes (NWP offices, research labs, European bodies), starting with the software tools (IFS, Arpge/Aladin, Mso-NH for the atmospheric part)

The fundamental ingredients which are detailed in the following sections are:

- limited-area model domain at least 1000km wide in order to provide some local predictability over at least the 24-hour forecast range,
- horizontal resolution better than 3km, vertical better than 400m, dynamics timestep of the order of 1 minute,
- prognostic representation of 3-D wind (i.e. non-hydrostatic dynamics), temperature, mass, water vapour, relevant cloud and precipitation water species, turbulent kinetic energy, plus as many chemicals as required by the users,
- assimilation of observed data at temporal resolutions from 10 to 60 minutes,
- variational assimilation including non-linear radiance and radar reflectivity data,
- optimized structure functions for the analysis of humidity, planetary boundary layer, convective cell structures and orographic effects,
- careful treatment of fine-scale verification and limits of predictability.

2 The Arome model

The correct definition of a data assimilation system starts by careful consideration of the model. Several features of a mesoscale model have an impact on the analysis algorithm. An obvious topological difference with a large-scale, global model is the presence of a lateral boundary condition. In theory there should be no discontinuity in resolution at the interface between global and mesoscale model. The technique of variable resolution in a global or limited-area model (LAM) has been studied by several teams but so far it has failed to lead to competitive NWP systems. Also, there are practical advantages to the flexibility of a modular global/LAM approach, where several concurrent LAMs can be configured independently from the global model, as opposed to a completely integrated approach whereby a single variable-resolution system would be needed for each region of interest. Although it is recognised that meteorological forecast quality is harmed by numerical problems in existing global/LAM coupling techniques, this flexibility translates in terms of availability of operational resources in such a way that the LAM approach makes very high resolution affordable on small areas. Nevertheless, the numerical formulation of lateral boundary condition coupling remains a difficult and concerning problem that deserves serious study. It contains at least four subproblems: discontinuities in the dynamics, in the orography (which can be solved by making it match across the boundaries), in the resolution-dependent behaviour of the physics, and the update frequency of boundary conditions (limited by efficiency constraints on the global model or by telecommunication bandwidth; kilometric-scale meteorology probably requires at least hourly updates of boundary conditions). One needs to consider both problems of ingoing (how to input large-scale waves into the LAM) and outgoing (how to prevent spurious wave reflexion) information at the LAM lateral boundaries.

The numerical formulation of modern LAMs is dominated by terrain-following vertical coordinates which are efficient in the modelling of weather that does follow the orography. This means in practice that the maximum slope allowed at the surface is of the order of 30 degrees. The use of ever higher horizontal resolutions is likely to lead to a growing interest in altitude-based model coordinates (physically more relevant for highly stratified weather such as fog in valleys) and in the representation of steep or vertical terrain, such as cliffs, building, factories or vehicles. This seems to be required for accurate modelling of urban weather and industrial impact studies, probably in ten years or so.

The representation of stratosphere is probably not needed at high horizontal resolutions: LAMs do not need to forecast the evolution of the atmosphere higher than the mid-stratosphere or so. A high-quality stratospheric background is however important for data assimilation in the radiative transfer models used as observation operators, and it is anticipated that this information will be provided as a forcing by large-scale global model using sophisticated stratospheric dynamics and chemistry.

The horizontal discretisation may either be gridpoint or spectral (with field biperiodization). The distinction is rather academic now that most spectral models use gridpoint semi-Lagrangian advection, and all models apply some sort of field smoothing in order to prevent the growth of non-physical discretisation waves in the dynamics. Spectral models with appropriately designed collocation grids do not exhibit any Gibbs effect (spectral non-physical artifacts) due to spectral transforms, and many so-called gridpoint models use spectral numerical solvers in their timestepping.

There is more competition in the field of the dynamical formulation itself and in the way the timestepping is performed. Dynamics at resolutions of the order of a kilometer or less have to be non-hydrostatic: this is essential for the correct representation of orographically induced waves in many situations, and for the modelling of deep convection where vertical velocities can reach tens of metres per seconds. Non-hydrostatic dynamics can be formulated in a number of ways e.g. compressible or anelastic, in height-based or mass-based vertical coordinate, with a timestepping based on high-order explicit or partially implicit schemes.

The physics/dynamics interface needs to be carefully designed because the definition of a model air parcel ceases to be obvious when the model grid approaches the scale of turbulent eddies, and there are complex microphysical and precipitation processes involved. For instance, assuming local thermodynamical equilibrium or horizontal homogeneity is unlikely to be acceptable in strongly convective clouds, and raises issues in the very definition of fluxes in the model. Precipitation in particular exchanges heat and momentum with its environment, takes much longer than a model timestep to fall, and exits the model when it reaches the surface, creating fluxes that may have a strong impact on the modelling of tropical cyclones, for instance. The interaction between these processes and the non-hydrostatic dynamics needs to be carefully designed in order to ensure that each significant aspect is represented exactly once in the model timestep, either in the dynamics or in the effect of the physical parametrisations.

The relevant physical parametrisations themselves are different from the ones used in large-scale models: some disappear because the corresponding processes are explicitly represented on the grid of the mesoscale model (e.g. gravity wave drag on orography, deep convection), others are needed in order to handle the next level of approximation implied by the higher degree of detail in these models:

- turbulent mixing (usually handled as vertical diffusion in large-scale models) needs to take into account horizontal mixing effects, the shape of the surface, and the timescale of evolution of subgrid turbulence, which usually requires a prognostic handling of turbulent kinetic energy. Turbulence is the main relevant subgrid-scale issue in kilometric-scale models, which is reminiscent of the problem of parametrizing subgrid convection in large-scale models.

- microphysical processes need to be quite detailed for an accurate representation of cloud structures and precipitation. Again, this involves specific prognostic fields due to the relevant timescales and the importance of tridimensional advection. Water may be present in a variety of states: cloud liquid water and ice, graupel, precipitating rain, snow, hail... Microphysical parametrisations need to handle both fast pro-

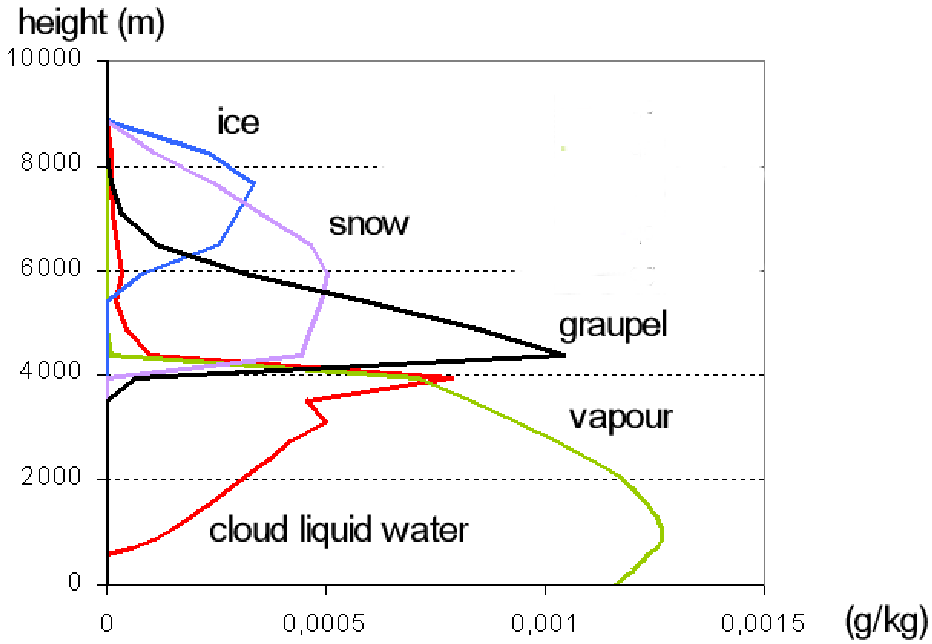

Figure 3: Vertical distribution of the microphysical species in a 1-D simulation of a deep convective cloud using the Arome/Mso-NH microphysics scheme.

cesses (like phase changes) and slow processes (e.g. falling of precipitation) with respect to the timestep of the dynamics (figure 3).

- radiation needs to take three-dimensional effects into account, such as shading in mountainous areas, and the precise interaction with clouds.
- the representation of surfaces will naturally be as detailed as allowed by the high horizontal resolution, e.g. with a representation of cities and heterogeneities in land surface cover, offering interesting opportunities for chemical modelling.
- additional parametrisations may be required for specific needs, such as shallow convection if it is not satisfactorily handled by the turbulence and microphysics, or parametrisations of ice or the ocean mixing layer e.g. for the prediction of coastal fog.

It is recognized that the importance of physics in mesoscale processes warrants the use of parametrisations that account for most of the numerically cost of a model timestep, i.e. more than in large-scale models. The large variation in relevant timescales of the represented processes may lead to calling each parametrisation scheme at frequencies that can be higher or lower than the dynamics timestep, in order to find a good compromise between cost and physical accuracy.

The choice is for the Arome model to use a physical package from the Mso-NH model (Lafore et al 1998, Caniaux et al 1995, Cuxart et al 2000, Masson 2000, Pinty 1998) and to plug it into a new version of the Aladin community model (Bubnová et al 1995) with new, efficient non-hydrostatic implicit semi-lagrangian numerics (Bnard, in press).

3 Issues in convective-scale data assimilation

As in any data assimilation system, the emphasis is on analyzing the variables and features that are going to drive the prediction of the targeted phenomenon i.e. mesoscale convective systems at ranges from a few hours to one or two days. This optimal control problem can take several routes:

- direct observation of the relevant fields. This is the ideal situation, rarely achieved due to limitations in data coverage and instrumental technology. There is a lack of observations of the inside of convective clouds, or of near-surface variables at kilometric resolutions.
- indirect analysis through multivariate couplings in the observation operators, the background error model, and the linearized model (in 4D-Var only). Most of the coupling is likely to be provided by balance properties built into the model of background error statistics.
- forcing of some variables by others in the integration of the model. This works well for variables that adjust quickly to others (e.g. vertical wind, cloud water), even if they are prognostic variables. The constraint is that the adjustment times must be shorter than the required forecast ranges (less than 3 hours at mesoscale, at least 12 hours in global models) and the process should not create spurious features in the model output, such as oscillatory behaviour.

Since a LAM is driven by a large-scale data assimilation and modelling system, the emphasis is on the analysis of small-scale features. This assumes that the LAM data assimilation has a means to import large-scale update information from its coupling system (more on this later). The mesoscale research community has experience with convection forecast sensitivity to the initial conditions (Ducrocq et al 2000), which indicates that it is essential to analyze clouds, humidity and low-level fields at fine scales. Correct initialization of potential vorticity anomalies is also known to play a key role in forecasting precursors of cyclogenesis and frontogenesis. LAM data analysis at fine scales implies that care must be taken in the use of observations near the lateral boundaries and the orography, which probably deserves more attention than in large-scale assimilation systems. Ultimately, if the LAM data assimilation is meant to be mainly useful is some types of severe weather, the whole system needs to be tuned with respect to this kind of weather (whereas large-scale systems have to be permanently able to deal with all kinds of weather, using appropriate algorithmic compromises).

Timeliness is essential for LAM forecasts to be useful in real-time NWP. Most of the added value of using a LAM (with respect to large-scale NWP systems) is expected to be in the short ranges: at ranges beyond one or two days, one hardly expects small atmospheric scales to be predictable, except for those implied by trivial orographic adjustment of large-scale features . Conversely, there is increasing pressure to design NWP systems that are usable at very short ranges (less than 6 hours) in order to fill the current gap between NWP and nowcasting techniques (primarily based on image extrapolation, and rarely used beyond the two hour range). It is desirable that the NWP products do not disagree with the latest available satellite and radar imagery data since they attempt to represent the same scales and phenomena. Thus, a kilometric-scale LAM NWP system such as Arome should target forecast ranges from one or two hours to 24-36 hours. This means that the forecasts need to be delivered to the forecasters and customers in no more than an hour after real time, and updated every one or two hours. It means in turn that observations must be acquired in 15 to 30 minutes, and that fast data analysis algorithms must be used. This means that the requirements of mesoscale NWP in terms of observations and algorithms are quite different from the global NWP systems that typically allow several hours for data acquisition and analysis.

4 Relevant observations

The most informative data from satellite in NWP data assimilation is from polar-orbiting platforms such as EPS, NPOESS, EOS etc. (see J.-N. Thpaut's presentation in this book), with a frequency of about 4 hours

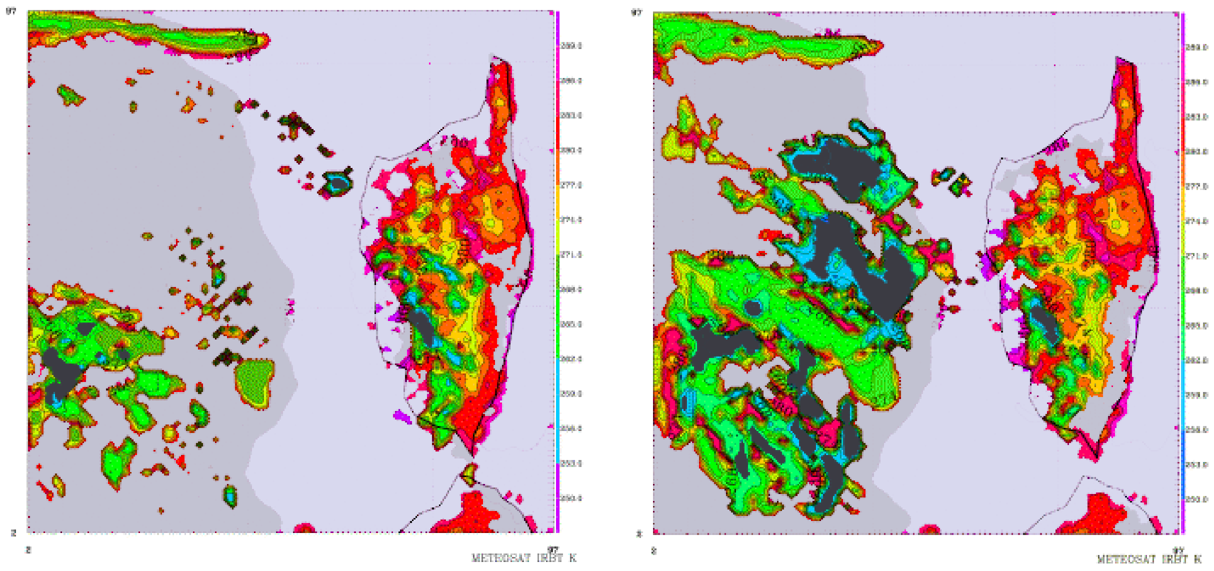

Figure 4: Impact of assimilating full-resolution clear-sky infrared radiances (water vapour channels) from the MSG satellite on the 6-hour forecast of 2.5km-resolution mesoscale Mso-NH model, using the Arome 3D-Var algorithm to compute the starting analysis. The panels show the forecast of infrared outgoing radiances (showing deep convective cloud top structures in dark, white areas are clear skies) without (left panel) and with (right panel) MSG data assimilation. The domain is the Mediterranean sea, west of the island of Corsica, encompassing 200x200km on the plot. The right-hand image is in better agreement with the observed infrared satellite image (not shown), indicating genuine predictability of clouds and rain down to about 20km-scale in the current version of Arome (to be further improved).

around 2008 on any given region. This frequency is low, but the information content is potentially valuable, with high spectral and spatial resolution if a clever combination of sounders and imagers is used. Conversely, geostationary data is less informative in each pixel, with less spectral channels and spectral resolution, but the coverage is much better: every 30 minutes or so (MSG, GOES, etc. depending on the geographical area), with a resolution of a few kilometres (figure 4). Clearly, both observing systems have complementary roles to play.

The timeliness requirements mean that fast data transmission and preprocessing must be available, usually with local data acquisition (Eumetsat has set up a convenient acquisition system over Europe, EARS, and transportable local reception hardware and software (e.g. AAPP from the Eumetsat NWP Satellite Application Facility) is available for several polar-orobiting satellites). Consequently, many satellite instruments that are commonly assimilated in global NWP systems are useless for mesoscale applications. Late data that provides essential information on the atmosphere may however be to some extent using appropriate real time catchup assimilation strategies that combine long- and short-cutoff analyses, but this can be expensive, as it amounts to running the data assimilation system several times to at each analysis date.

The satellite-derived information that will be useful for mesoscale NWP comprises:

- infrared sounders with high spectral (thus, vertical) resolution such as AIRS, IASI and CrIS. They provide useful data on mid- and upper-tropospheric humidity and temperature, the major problem being that they do not work below cloud tops (most interesting convective weather involves extensive cloud cover e.g. cirrus or cumulonimbus anvils).
- infrared sensors also provide data on cloud top height and cloud shape. Suitable data assimilation algorithms need to be developed for these highly nonlinear, but important, features.
- microwave sounders such as AMSU and the SSM/I family (if it can be acquired) provide data on temperature, humidity and cloud liquid water and ice, even in cloudy scenes. They are potentially very

useful over sea despite their relatively low spatial resolution. The major problem is the interpretation of microwave radiances over land surfaces, which we hope will improve in future years.

- microwave sounders and scatterometers also provide information on sea surface winds, which have potential value at mesoscales (simulations of strong orographically-induced convective precipitation in Mediterranean countries have shown high sensitivity to low-level wind on nearby sea surfaces)
- GPS data can provide humidity information with either low vertical or horizontal resolution, depending on the technology used, but reliable humidity data is always desirable for improved prediction of clouds and precipitation.
- Doppler wind lidar (DWL) is still a technology of the future, with current projects only seeking to provide very sparse coverage. It is also affected by the presence of clouds. Nevertheless, it has huge potential for mesoscale applications, since it could observe tropospheric wind over any area, including land and low levels, with no serious competition from any other satellite-based technology.
- Atmospheric-motion winds (SATOBs) can be produced at fine scales. Although current experience with SATOB data shows limited impact in large-scale NWP, some larger benefits may be found at finer scales.

Data from in-situ instruments and ground-based wind profilers mostly suffers from a lack of coverage, even in the better-equipped countries. The main exception is near busy airports, where abundant aicraft report data provide fine high-resolution information which is going to be useful in NWP for air traffic. Special consideration shall be given to surface stations, which, thanks to automatisation (land stations and buoys) can provide precious boundary layer data for which there is no real competition from remote-sensing so far: pressure, wind, humidity, quantitative precipitation rate, time-integrated precipitation, cloud type and base, visibility. Automated surface networks can reach an effective resolution of 30 kilometres over large areas.

Radar data is essential for short-range precipitation forecasts, and depiction of strong convective systems, since hardly any other observing system sees inside clouds. Modern radars can measure Doppler wind components, reflectivity that is linked to water and ice species, polarization information which helps in the interpretation of reflectivity and the characterization of water and ice species, and vertical distribution of reflectivity (thanks to appropriate scanning strategies) which help in quality control and physical interpretation e.g. for translating upper-level reflectivities into surface precipitation rates. Even richer information is expected in the future from the developing radar technology. The main limitation is the coverage of the radar network, which tends to be restricted to industrial areas with many gaps e.g. due to surface masks and orography, and the difficulty in preprocessing the radar signal (transmission, quality control, bias correction).

In summary, the most crucial data for mesoscale data assimilation (on top of the data used by the large-scale assimilation that provides the boundary conditions) is going to be (assuming the LAM covers a well-equipped region):

1. radar data using modern instruments,
2. surface automated networks (land and buoys), aircraft near major airports,
3. geostationary satellite radiances and cloud products,
4. polar-orbiting infrared satellite radiances
5. polar-orbiting satellite microwave and scatterometer data (over sea)

The use of satellite data can largely be copied from what large-scale NWP systems already do. On the other hand, much work is needed on radar data, which is specific to mesoscale applications (but may eventually be useful in large-scale assimilation, too). In terms of observing network planning, mesoscale observing systems experiments will be needed as soon as efficient data assimilation systems are available. The main underlying question is whether one should insist on high-quality data with limited coverage (e.g. as sophisticated and expensive profilers or radars), or on high coverage in space and time from simpler instruments (e.g. as improved geostationary satellites, or more frequent polar-orbiting satellites).

5 Data assimilation algorithms

Although 4D-Var has proved to be highly effective in global data assimilation, its applicability to time-critical mesoscale NWP is questionable for two reasons:

- 4D-Var is inherently slow because it processes data in batches. In order to benefit from flow-dependent structure functions, one needs long enough time windows, which makes the time-critical part of the data assimilation expensive. The trick of using an incremental approach (computing 4D-Var at a low resolution) may not work as well in mesoscale as in synoptic scales, because the paradigm of 'large-scale waves that force small-scale features' does not apply to many important mesoscale phenomena: the triggering of convection by low-level convergence lines (e.g. coastal breeze) or orography is a mainly small-scale process. Cloud microphysics are completely local processes that do not react well to reinterpolations. It is likely that a satisfactory mesoscale analysis system will act to be at the same resolution as the model.
- 4D-Var heavily relies on linearization hypotheses that are questionable at fine scales. Cloud and boundary layer fields contain small-scale, stiff patterns that are going to defeat the linearization of local field perturbations with respect to changes in wind. Many fields to analyse (starting with humidity and cloud liquid water) have non-gaussian probability density functions with complex, non-linear balance relationships.

There are indeed some successful experiences with mesoscale 4D-Var. The works of Sun and Crook (1998, among other studies) pertain to cloud-scale models, with very high resolutions, short time windows, and abundant data, using mainly 3D Doppler winds that constrained the analysis problem well at the scale of a single convective system. One may wonder whether this would apply to NWP applications where data coverage is very partial. Also, the Japanese JMA team uses an operational 4D-Var at 20-km incremental resolution. Although their system has a lower resolution that what is discussed here, experience from their system is going to be very instructive.

In the Arome project, the chosen strategy focusses on the development of a comprehensive, full-resolution 3D-Var system at 2km horizontal resolution before moving to 4D-Var when computer power allows it. 3D-Var already provides many of the benefits of 4D-Var when used in FGAT mode (first-guess at the appropriate time for the comparison to observations, i.e. $y - H(x_b)$ is 4D but matrix K is computed in 3D), including the use of weakly nonlinear observation operators, and is a good development ground for observation operators, initialization techniques, background error model (Berre 2000), and large-scale coupling strategies. The two areas where experimentation with 4D-Var is highly desirable are the exploitation of Doppler wind information, and the interpretation of surface rainfall/cloud top information in terms of cloud internal structure, because they imply a complex inversion procedure. Future will tell whether there problems can be treated without going to full 4D-Var e.g. using sophisticated balance constraints in the background error model.

Structure functions need to be improved on top of the formulations that already exist in large-scale data assimilation. Large-scale spectral non-separability may be less important at small-scales, but weather-dependency deserves more attention: on a small region, the typical shape and amplitude of forecast errors is likely to vary a lot from one analysis to the next. Even on a single case, errors inside a convective system are likely to be very different ahead or behind this system, because of the complex organisation of airmasses and microphysical fields at these scales. The main area for concern is the vertical planetary boundary layer structure, in order to use low-level observations efficiently and to depict inversion layers when appropriate. The humidity analysis is a closely related topic, and it plays a major role in the evolution of most cloud systems. Appropriate structure functions can be designed using physical common sense and appropriate background error calibration procedures. The next step will be the development of more automatic structure function specification techniques, using simplified forms the Kalman filter. Rather sophisticated and expensive algorithms can be considered, as long as they do not imply an unreasonable computer burden on the time-critical part of the NWP production system. A special problem will be the specification of horizontal structure functions close to the lateral boundary conditions, since the analysis increments need to be compatible with the coupling constraints imposed on the LAM forecast model.

Usually, large-scale error autocorrelation models are tensor products of horizontal and vertical operators, following a suitable vertical coordinate. Mesoscale analysis in mountainous terrain will require specific studies, as we are dealing with steeper and steeper slopes, and considering that some phenomena (such as fog) do not follow the slopes. Fields like low-level winds are likely to exhibit very complex error structures over and around orography.

Statistical studies of forecast errors show less geostrophic balance at small scales than at synoptic scales in the atmosphere. This is even more pronounced when the weather is dominated by mesoscale features such as deep convection, orographically induced structures, sea breeze or rain bands, usually with lifespans that are shorter than the geostrophic adjustment time. Also, the initialisation of mesoscale humidity often has an impact on the forecast, unlike at synoptic scales where variations in humidity are mostly driven by the dry model dynamics. There is a view in synoptic-scale meteorology that prominent dynamical systems (e.g. fronts, cyclogenesis, barotropic or baroclinic waves) can be completely initialized using the knowledge of only part of the atmospheric state (e.g. the distribution of mass or of potential vorticity): the non-observed parameters can be reconstructed by the model dynamics, dynamical initialization or balance constraints in the data assimilation. This view is only weakly useful at scales smaller than a few kilometres, where all fields (pressure, temperature, wind, humidity) have to be observed and analyzed. Current experience suggests that one can still rely on the hydrostatic balance to avoid explicit analysis of vertical wind, and on properties of the physics to let the model produce an adjustment of turbulent kinetic energy and microphysical variables (with adjustment times of one hour or less) with respect to the other fields. Direct analysis of these fields may still be useful if relevant observations are available.

A research effort is needed to try and derive new balance constraints that can be applied in background error modelling at mesoscales. It is intuitively obvious that there is some physical balance mechanism behind the organisation of convective clouds, mesoscale storms, stably stratified layers, or breezes. The problem is to formulate it in a numerically suitable way for the variational analysis, and (perhaps most importantly) to be able to activate various types of balance constraints at the right places, and at the right times. The expected benefit is the enforcement of multivariate structure functions which would allow the assimilation of organized mesoscale systems on the basis of incomplete observations such as the ones provided by remote-sensing instruments. This would be a mathematical equivalent of the process by which the human brain imagines three-dimensional meteorological objects, simply by looking at radar and satellite images.

6 Large/small scale coupling

The coupling of LAM data assimilation with large-scale systems is not yet a solved problem. A quite advanced development has been applied in the regional LAM 4D-Var system at the Japan Meteorological Agency (JMA), through the use of a large-scale constrint term in the variational cost-function. Still, there is a lack of fundamental studies of the issue. Basically, one wants a LAM forecast that uses large-scale information as well as possible. Except during the first minutes of the forecast, or in cases of very stationary large-scale weather, a LAM forecast (and data assimilation) is significantly influenced by both its own initial state and the large-scale forcing. This is because large-scale atmospheric waves tend to travel very quickly (group velocity well over 100km/h): if the large-scale forcing is very poor, there is little hope that the LAM forecast will be useful. On the other hand, there is experimental evidence that small perturbations to the initial state of a LAM forecast can last for a long time (e.g. three days in a 3000km wide LAM over Europe). It gives confidence that, with a reasonably good large-scale forcing (such that large-scale errors will not corrupt any small-scale information in the model), a LAM data assimilation will indeed provide LAM forecasts with added value, compared to LAM forecasts in pure local adaptation mode (i.e. started from an interpolation of a large-scale analysis). This is expected despite imperfections in the large-scale forcing and weaknesses in the numerical formulation of the large/small-scale coupling in the LAM forecast.

One could run a LAM data assimilation system in algorithmic isolation from the large-scale assimilation system, the large-scale information coming only through the model's boundary conditions. Such a system will inevitably develop discrepancies between the large- and small-scale analyses for a variety of reasons:

- difference in the physical behaviour of the two models,
- differences in the use of observations and cutoff times,
- chaotic growth of perturbations in both models.

This is undesirable because it implies unphysical discrepancies in the history of lateral boundary conditions for the LAM forecast, inconsistencies between the large- and small-scale forecast products, and imperfect communication of the information about atmospheric large scales to the LAM system. Considering that global assimilation now reach 30km effective resolution, and operational LAM assimilations are expected to treat domains thousands of kilometres wide, there is a range of important spatial frequencies (of the order of 100-1000km) that will be much better assimilated in global than in LAM assimilations (since global systems have access to more observations and avoid geometrical aliasing problems at the LAM lateral boundaries), but that the LAM system needs to know accurately.

Clearly, one needs a procedure to force the LAM analysis to be consistent with the corresponding large-scale state (an analysis, or a recent forecast if the global analysis is not available in time for the LAM system), near the LAM lateral boundaries, and in the larger scales represented by the LAM. A pragmatic procedure called 'blending' has been tested in the Aladin European model community, by which the largest scales of the LAM are reset to the corresponding ones in the forcing global model, and a digital filter initialization is applied in order to avoid problems due to e.g. orographic discrepancies. Another approach is to go back to the mathematical formulation of the LAM data assimilation problem, and to consider the large-scale forcing as just another prior set of information, like the LAM background, the observation and the balance constraints. A rigorous probabilistic approach would require the knowledge and modelling of error statistics in the large-scale forcing, and error correlations with other terms (there may be strong error correlations with the LAM background fields near the lateral boundaries). One can be pragmatic about it and simply add a conveniently designed 'large-scale penalty' term to the 3D/4D-Var LAM cost function. This amounts to the procedure used at JMA, and it is being developed and theoretically studied in the framework of the Aladin/Arome models at fine scales. The underlying idea is that the job of the LAM analysis is restricted to analysing the smallest scales.

A clean mathematical approach to the coupling of large/small-scale data assimilation systems is difficult. Specific research is needed to understand which hypotheses are acceptable or not. In a real-time framework the implementation of the coupling is facing three interacting problems:

- numerical properties of the formulation of lateral boundary conditions in the LAM forecast model,
- analysis cutoff practices in the global and regional data assimilation system, which will lead to compromises regarding the freshness of large-scale analyses and lateral boundary conditions available to the LAM system,
- numerical properties of formulation of large/small-scale coupling in the LAM analysis.

In particular one needs to balance the qualitative advantage of waiting for a fresh large-scale analysis, with the implied loss in timeliness of the LAM analysis and forecast. Complicated real-time catchup strategies may be needed, with analyses done using several data cutoffs in both large- and small-scale assimilation systems.

There are many potential sources of statistical suboptimality in such a coupled system. Treating the LAM background and the large-scale forcing as data with statistically independent errors: intuitively, errors in the large-scale system will produce related errors in the past history of LAM model boundary conditions, and the boundary conditions available to the analysis. If the large-scale forcing is the result of an analysis, then its errors are likely to be correlated with observation errors in the LAM analysis, since some observations would be used in both. A workaround would be to enforce a strict scale separation between the two systems i.e. the LAM analysis should only deal with spatial scales that are not represented in the large-scale forcing. This may work well in the conceptual framework of inorganized convection, where there large-scale conditions conducive to convection are provided by the large-scale assimilation, and the LAM analysis is only there to determine the precise location and structure of each small convective cell. In reality, however, there often is some organization at all scales, and one can imagine that a large scale front or lowpressure system cannot be modified independently from the smaller-scale precipitating structures it is driving. This scale separation problem is fundamental to LAM mesoscale modelling and will require much study in the future. Perhaps one needs to develop specific multiscale background error correlation models, or the problem is too difficult to be solved directly and one should focus on designing mesoscale predictability estimation algorithms that will measure whether the LAM forecast is expected to beat or not the large-scale forecast products, depending on the situation.

7 Conclusions

There is a steady push for both research and operational NWP models to deal with higher and higher resolutions on limited areas. NWP with a few metres of resolution may become a reality within a few years for very specific applications, such as urban air quality monitoring or atmospheric modelling on special sites. It eventually means dealing with steep, and even vertical, slopes, which will lead to new challenges in the numerical formulation of models and assimilation algorithms. Continuing progress will be needed in the development of parametrisations of microphysical processes, turbulent mixing and radiation, as well as their regularized counterparts for use in future high-resolution 4D-Var, and the determination of the balance properties they imply, either for improved background error modelling, or for a posteriori physical initialization of the forecasts.

Observing networks need to be adapted to the future needs of mesoscale data assimilation. This should become easier when we have developed such operational assimilation systems to a state where we can demonstrate that we are indeed able to make good use of existing observing systems, and where we can measure the value of new observing networks in order to express the needs of mesoscale needs precisely. There already are some good networks that have potential value for mesoscale NWP: geostationary satellites, radars, mesoscale surface networks. Future improvements are expected from more sophisticated radars with improved network coverage, Doppler wind lidars, denser automated surface networks, all-weather satellite data, lightning detection, and observations targeted for mesoscale NWP. Image processing techniques need to be developed for optimal use of remote-sensing data, and current experience gathered in the nowcasting community will be helpful.

The major deliverable that is expected from new mesoscale NWP system is quantitative precipitation forecasting (QPF). This raises interesting challenges in modelling and verification as the emphasis can be put on differing interpetations of model precipitation output: extreme precipitation values (e.g. for flash flood prediction), cumulated amounts over several hours (for hydrology), precipitation types (snow, rain, hail), occurence of precipitation. Users may be concerned about the precise location and timing of precipitating events such as fronts, showers, or of other events like thunderstorms, coastal surges, or fog formation and dissipation. The relevant meteorological fields may be non-smooth and highly inhomogeneous, with very small-scale features that may or may not be predictable. The associated parameter distributions may be non-Gaussian and situation-dependent. It means that work is needed on the probabilistic interpretation of mesoscale model output, starting with the transposition of predictability and diagnostic tools already developed for large-scale NWP systems.

On top of this growing demand for improving weather information, there are requirements for a multidisciplinary approach to environmental problems: pollution, floods, forest fires, avalanches and landslides... It means that mesoscale models and their data assimilations will have to be coupled with models designed by other communities for non-NWP needs:

- land surface models depicting land use, soil, vegetation, cities and snow;
- hydrological models to handle fast and slow floods, droughts and agricultural water availability;
- models for superficial ocean layers and the cryosphere;
- chemical models for gases and aerosols;
- other models for fire, visibility, aircraft turbulence...

Most of these models do not need to be interactively coupled with the NWP data assimilation. Still, they will pose many problems of NWP system postprocessing, grid definition, software maintenance, validation of integrated systems, and the workforce spent on interaction between an ever-increasing number of NWP centres, laboratories, international institutes and coordinating bodies. The broader key issues will be the optimal fusion between as many data fusion as possible, model intercomparison in order to preserve some healthy competition between future operational mesoscale NWP centres, and comprehensive validation, including probabilistic risk management aspects, of all aspects of the quality of the model output.

8 References

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Wea. Rev.,* **128,** 644–667.

Bubnová, R., G. Hello, P. Bénard and J.-F. Geleyn, 1995: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/ALADIN NWP system. *Mon. Wea. Rev.,* **123,** 515–535.

Caniaux, G., J.-L. Redelsperger and J.-P. Lafore, 1994: A numerical study of the stratiform region of a fast-moving squall line. Part I. General description, and water and heat budgets. *Journal of Atmospheric Sciences*,**51,** 2046-2074.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch and A. Staniforth, 1998: The operational CMC/MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395.

Cuxart, J., Bougeault, Ph. and Redelsperger, J.L., 2000: A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q. J. R. Meteorol. Soc.,* **126,** 1-30.

Ducrocq, V., J.-P. Lafore, J.-L. Redelsperger and F. Orain, 2000: Initialisation of a fine scale model for convective system prediction: a case study. *Quart. J. Roy. Meteorol. Soc.,* **126,** 3041–3066.

Golding, B., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteorol. Appl.,* **5,** 1–16.

Lafore, J.-L., J. Stein, N. Asencio, P. Bougeault, V. Ducrocq, C. Fischer, P. Héreil, J.-L. Redelsperger, E. Richard and J. Vilà-Guerau de Arellano, 1998: The Meso-NH atmospheric simulation system. Part I: adiabatic formulation and control simulations. *Ann. Geophys.,* **16,** 90-109.

Laprise, R., D. Caya, G. Bergeron and M. Giguère, 1997: the formulation of the André Robert MC2 model. *Atmos-Ocean*, special vol.35, no.1, 466–475.

Masson V., 2000: A physically-based scheme for the urban energy budget in atmospheric models. *Bound. Layer Meteor,* 1994, 357-397.

Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock (2001): "Development of a Next Generation Regional Weather Research and Forecast Model" in *Developments in Teracomputing*: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology. Eds. Walter Zwieflhofer and Norbert Kreitz. World Scientific, Singapore. pp. 269-276.

Pinty, J.-P. and P. Jabouille, 1998: A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. Proc. Conf. of Cloud Physics, Everett, WA, USA, Amer. Meteor. soc., Aug. 1999, 217 - 220.

Stockwell R.W., Kirchner F., Kuhn M., Seefeld S., A new mechanism for regional atmospheric chemistry modeling, *J. Geophys. Res.*, **102,** 25847-25879, 1997.

Sun J., and N. A. Crook, 1998: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint: Part II. Retrieval experiments of an observed Florida convective storm, *J. Atmos. Sci.*, **55,** 835-852.