1. INTRODUCTION

With current horizontal resolutions, the models used for numerical weather prediction (NWP) are approaching limits of validity of the hydrostatic approximation. Considerable experience with nonhydrostatic models has been accumulated in simulating convective clouds and storms, but this experience may not be directly or entirely applicable to NWP. Namely, NWP deals with a much wider range of temporal and spatial scales.

Concerning the criteria that a successful nonhydrostatic NWP model should satisfy, there are several obvious choices. Apparently, the accuracy of the nonhydrostatic model must not be inferior to that of mature hydrostatic models running at the same resolution. Moreover, having in mind the uncertainties concerning the benefits that can be expected from nonhydrostatic dynamics at transitional resolutions, the nonhydrostatic model must be computationally efficient. Finally, the model dynamics should be capable of reproducing strongly nonhydrostatic flows at very high resolutions. Although such resolutions are far beyond the resolutions that will be used in NWP in the foreseeable future, this condition must be satisfied in order to demonstrate that the model is nonhydrostatic.

Having in mind these considerations, a novel approach (Janjic et al., 2001; Janjic, 2003) has been applied in the NCEP Nonhydrostatic Mesoscale Model (NMM) that has been developed within the Weather Research and Forecasting (WRF) initiative. Namely, instead of extending a cloud model to larger spatial and temporal scales, the hydrostatic approximation is relaxed in a hydrostatic model formulation based on modeling principles proven in practice. These principles have been set up by Janjic (1977, 1979, 1984) and have been applied and thoroughly tested in NWP and regional climate applications of the NCEP Eta model. By relaxing the hydrostatic approximation, the applicability of the model formulation is extended to nonhydrostatic motions, and at the same time, the favorable features of the hydrostatic formulation are preserved. In other words, following an evolutionary approach, the nonhydrostatic NWP model is built on NWP experience.

With this approach, the nonhydrostatic equations are split into two parts: (a) the part that corresponds to the hydrostatic system, except for corrections due to the vertical acceleration, and (b) the part that allows computation of the corrections appearing in the first system. No linearization or additional approximation is required. The nonhydrostatic effects are introduced in the form of an add on nonhydrostatic module that can be turned on or off depending on model resolution so that the nonhydrostatic model can be run in the hydrostatic mode at lower resolutions with reduced computational cost.

2. HIGHLIGHTS OF THE MODEL EQUATIONS

Following the usual practice in NWP, the model equations are formulated using a vertical coordinate based on mass (or hydrostatic pressure) (Janjic et al., 2001; Janjic, 2003). With this choice, the mass as well as a number of other first order and quadratic quantities can be conserved in the discretized system without major difficulties. Moreover, in the mass coordinate the nondivergent flow remains on coordinate surfaces.
A similar argument applies to adiabatic flows in isentropic coordinates. However, important flow regimes on the meso scales are characterized by weak stability and strong diabatic forcing, the features that render the isentropic coordinate less appealing.

The full system of nonhydrostatic equations is presented in Janjic et al. (2001). Further details and updates on the numerical methods used are given in Janjic (2003) and later on in this paper. At this point, only some very general implications for the discretization will be reviewed.

Even though the hybrid pressure-sigma vertical coordinate is actually employed in the model, for simplicity, the sigma coordinate (Phillips, 1957) will be used here as a representative of the mass based coordinates. Then, the nonhydrostatic continuity equation can be written as

\[ w = \frac{1}{g} \left( \frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \nabla \Phi + \sigma \frac{\partial \Phi}{\partial \sigma} \right) = \frac{dz}{dt} \]  

(2.1)

Here, \( w \) is vertical velocity, \( g \) is gravity, \( \Phi \) is geopotential, \( \mathbf{v} \) is the horizontal wind vector, \( \sigma \) is the vertical velocity in the sigma coordinate, and \( z \) is the height of coordinate surfaces. Then, the ratio of the vertical acceleration to gravity has the form

\[ \varepsilon \equiv \frac{1}{g} \frac{dw}{dt} \]  

(2.2)

Since the geopotential is computed from hydrostatic pressure, temperature and nonhydrostatic pressure, \( \Phi, w, \) and \( \varepsilon \) are not independent variables. An important consequence of this for the discretization is that an independent prognostic equation should not be used e.g. for vertical velocity. In the continuous case the presence of such an independent additional prognostic equation would be of no consequence since the system would be still internally consistent. However, in the discrete system, at different places, the same variable would be computed in different, generally inconsistent ways. A usual manifestation of such inconsistencies is enhanced noisiness, a problem that many nonhydrostatic models have been plagued with.

Another important implication for the discretization follows from the fact that the parameter (2.2) enters the nonhydrostatic equations for the basic dynamical variables through a factor of the form \( 1 + \varepsilon \) (Janjic et al., 2001; Janjic 2003). If \( \varepsilon \) is zero, the system becomes hydrostatic. On the synoptic scales \( \varepsilon \) is small and approaches computer round-off error.

In atmospheric flows, even in cases of vigorous vertical accelerations, \( \varepsilon \) remains several orders of magnitude smaller than unity. Thus, high accuracy of computation of \( \varepsilon \) is not of paramount importance since the computational errors are of an even higher order than \( \varepsilon \) itself.

3. CLASSICAL NONHYDROSTATIC SOLUTIONS

In order to test the validity of the approach in the limit of highly nonhydrostatic flows, a two-dimensional model in the vertical plane was developed and run in a series of classical nonhydrostatic tests (Janjic et al., 2001). However, as pointed out in Janjic (2003), and later on in this paper, the computational algorithms used in the model have evolved significantly since then. In order to verify that the new computational methods have not affected the favorable features of the model, results are shown here of the reruns of the cold and warm bubble tests using the latest computational techniques. For more details about these and other nonhydrostatic tests the reader is referred to Janjic et al. (2001).

Following Straka et al. (1993), in a neutrally stratified atmosphere an initial cold disturbance was introduced. The integration domain extended 40 km in the \( x \) direction, and the free surface was located at 442 hPa, or at about 6400 m. The center of the initial disturbance was in the middle of the domain in the \( x \) direction, 20 km away from either of the lateral boundaries. The boundary between the pressure and the sigma coordinate systems was set at about 900 hPa, in the region where the major disturbances were developing. As in the main test in the Straka et al. (1993) study, the horizontal resolution was 100 m, and the vertical resolution was 100 m on the average. The potential temperatures after 300 s (top), 600 s (middle) and 900 s (bottom) are displayed in Fig. 1. The area shown extends from the center of the domain to 19200 m to the right, and from the surface to 4600 m. The contour interval is 1K. Comparison of the results obtained in this test with those presented in Janjic et al. (2001), and Straka et al. (1993), reveals that the new numerical algorithms perform as good as or better than those used in the Janjic et al. (2001) tests. In particular, the pressure and sigma ranges of the hybrid vertical coordinate evidently fit to each other seamlessly.
Fig. 1. The cold bubble test. Potential temperatures after 300 s, 600 s and 900 s in the right hand part of the integration domain extending from the center to 19200 m, and from the surface to 4600 m. The grid size is $\Delta z \approx \Delta x = 100$ m and $\Delta t = 0.3$ s. The contour interval is $1^\circ$K.

In the warm bubble test (Janjic et al., 2001) an initial warm disturbance of the potential temperature was introduced in a neutral atmosphere. The integration domain extended 20 km in the $x$ direction. The free surface was located at 135 hPa, or at about 13500 m. The boundary between the pressure and sigma ranges of the hybrid vertical coordinate was set at about 560 hPa, again at a pressure level swept over by the ascending bubble. The center of the initial disturbance was in the middle of the domain, 10 km away from either of the lateral boundaries. The horizontal resolution was 100 m, and the vertical resolution was 100 m on the average. The potential temperature deviations after 360 s (upper left panel), 540 s (upper right panel), 720 s (lower left panel) and 900 s (lower right panel) are presented in Fig. 2. The area shown extends 16 km along the $x$ axis, and from 1000 m to 13200 m along the $z$ axis. The contour interval is $1^\circ$K. The rate of ascent and the intensity of the disturbance again agree very well with those reported earlier in Janjic et al. (2001).

4. HORIZONTAL GRID AND HORIZONTAL COORDINATES

Winninghoff (1968) and Arakawa and Lamb (1977) examined frequencies of gravity-inertia waves on various types of rectangular horizontal grids obtained using second-order centered differences. Generally, better agreements with the exact frequencies were achieved on the staggered grid $C$ and on the semi-staggered grid $B$ (or $E$) than on the other considered grids. The grids $B$, $C$ and $E$ are shown in Fig. 3. In the figure, $h$ denotes the mass point variables, while the horizontal velocity vector and the velocity components are denoted, respectively, by $v$, $u$ and $v$. However, the staggered grid and the semi-staggered grids also have problems (e.g. Janjic, 2003). The problems on the staggered grid arise due to the averaging of the velocity components in the Coriolis force terms. On the other hand, in order to illustrate the problems on the semi-staggered grids, consider the linearized shallow water equations on an infinite plane

$$
\frac{\partial u}{\partial t} = -g \frac{\partial h}{\partial x} + fu, \quad \frac{\partial v}{\partial t} = -g \frac{\partial h}{\partial y} - fv, \\
\frac{\partial h}{\partial t} = -H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right).
$$

(4.1)

Here, $u$ and $v$ are the velocity components, $h$ is the height of the free surface, $g$ is gravity, $f$ is the Coriolis parameter assumed to be constant,
and \( H \) is the mean depth of the fluid. The system (4.1) discretized in the most straightforward way on the \( B \) grid has the form
\[
\frac{\partial u}{\partial t} = -g\delta x h_y^v + fv, \quad \frac{\partial v}{\partial t} = -g\delta y h_x^v - fu, \\
\frac{\partial h}{\partial t} = -H(\delta x x \chi + \delta y y \chi),
\]
(4.2)

In (4.2), the symbol \( \delta \) and the overbar, respectively, represent the simplest two-point centered differencing and averaging operators applied in the direction indicated by the accompanying subscript or superscript. Following Janjic (1984), the velocity components on the \( B \) grid may be written in terms of the velocity potential \( \chi \) and the stream function \( \psi \) as
\[
u = \delta x \tilde{\chi}_y - \delta y \tilde{\psi}_x, \quad v = \delta y \tilde{\chi}_x + \delta x \tilde{\psi}_y.
\]
(4.3)

After substituting (4.3) into (4.2), and rearrangement, one obtains
\[
\frac{\partial \tilde{\chi}}{\partial t} = -gh + f\psi, \quad \frac{\partial \tilde{\psi}}{\partial t} = -f\chi, \\
\frac{\partial h}{\partial t} = -H(\delta x x \tilde{\chi}_y + \delta y y \tilde{\chi}_x).
\]
(4.4)

Here repeated subscripts and superscripts indicate repeated applications of the operators they are accompanying. As can be seen from (4.4), the transformed equations of motion are solved exactly, and the reason for the \( B \) grid problems is the insufficiently accurate computation of the Laplacian due to the averaging of the derivatives in the continuity equation (Janjic 2003).

Note that the finite difference equations (4.4) are defined on a nonstaggered grid carrying all three variables \( \chi, \psi \) and \( h \) at each grid point (Janjic, 1984). This grid is also shown in Fig. 3. It was named \( Z \) grid by Randall (1994) (Gavrilov, 2004). However, on the \( Z \) grid, the continuity equation can be written also as
\[
\frac{\partial h}{\partial t} = -H(\delta x x \chi + \delta y y \chi),
\]
(4.5)
i.e., without the averaging in the divergence term that was responsible for the \( B \) grid problems. Unfortunately, application of (4.5) in case of more complex equations requires costly conversions between the velocity components and the velocity potential and the stream function.

A more complete comparison of the properties of the remaining two possibilities, the staggered grid \( C \) and the semi-staggered grids \( B \) and \( E \) can be found, e.g., in Janjic and Mesinger (1984, 1989). These considerations do not give decisive advantage to either of the two choices. The problems on the semi-staggered grids \( B \) and \( E \) are restricted mainly to the shortest waves, and there is an effective technique for filtering the low frequency, short-wave noise resulting from the inaccurate computation of the divergence term (Janjic, 1979). More sophisticated, nondissipative methods (“deaveraging” and “isotropisation”) for dealing with the problem also were proposed (Janjic et al., 1998), leading to dramatic improvements of the finite-difference frequencies of the short gravity-inertia waves on the semi-staggered grids. On the other hand, in the case of slow internal modes, and/or weak static stability, the \( C \) grid develops problems in the entire range of the admissible wave numbers (Arakawa and Lamb, 1977). Note that phenomena occurring in a weakly stable atmosphere represent a very important class of mesoscale motions.

The linearized anelastic nonhydrostatic system is a better starting point for addressing the choice of the horizontal grid in the mesoscale models than the linearized shallow water equations (4.1) (communicated by Klemp, 1997; Janjic, 2003). As before, the problems with the anelastic system on the \( C \) grid are mainly due to the averaging of the Coriolis force. Depending on the horizontal and vertical grid sizes, in cases of weak static stability these problems may develop throughout the admissible wave-number range, including the...
longest waves (Janjic, 2003). On the $B$ grid the problems are again mainly due to the averaging within the divergence term, but remain restricted to the shortest resolvable scales, and hardly can be detected in real data simulations with currently used horizontal resolutions (Janjic, 2003). Since the problems on the semi-staggered grid $B$ are restricted to the shortest waves, and can be effectively controlled (Janjic, 1979; Janjic et al., 1998), the preference was given to the semi-staggered grids. For historical reasons, the $E$ grid is used in the initial version of the NCEP Nonhydrostatic Meso Model, although a $B$ grid version (NMM-B) with analogous properties also exists (Janjic, 2003).

The longitude-latitude coordinates are rotated in the model in such a way that the coordinate origin is located in the middle of the integration domain. In this way, the reduction of the longitudinal grid-size is minimized as the southern and the northern boundaries of the integration domain are approached, and, therefore, longer time steps can be used.

5. VERTICAL COORDINATE AND VERTICAL STAGGERING

The most widely used approach for representing topography are terrain-following coordinates such as the sigma coordinate (Phillips, 1957) and its extensions such as the hybrid sigma-pressure coordinate of Arakawa and Lamb (1977), or the hybrid eta coordinate of Simmons and Burridge (1981). A rare exception has been the step-mountain blocking used in the NCEP Eta model. Originally proposed by Bryan (1969), and subsequently widely used in oceanography, this technique was implemented in the sigma coordinate by Mesinger et al. (1988). Yet another approach coming from oceanography is the shaved cell method (e.g. Adcroft et al. 1997). Steppeler et al. (2002) implemented this method in the dynamical core of the Lokal Modell of the German Weather Service.

The advantage of the step-like mountain blocking is that the coordinate surfaces are quasi-horizontal. This, however, is not without consequences. For example, internal discontinuities are introduced at the vertical sides of the steps that replace the mountain slopes, and lateral boundary conditions are required at these discontinuities. Note that the accuracy of finite-differencing at the points next to the internal boundaries is reduced. If the no slip boundary conditions (Bryan, 1969; Mesinger et al., 1988) are used in order to preserve the major favorable features of the finite-differencing schemes (Janjic, 1977, 1979, 1984), a nonphysical sink of momentum is introduced. Yet another problem is the representation of the physical processes in the planetary boundary layer (PBL). If one wants to represent these processes in a reasonably uniform way throughout the integration domain, including both low-lying and elevated terrain, an approximately equidistant spacing of the vertical levels is required in the lowest few kilometers of the atmosphere. However, the vertical resolution needed to achieve this goal is still too high. In addition, several recent studies (Adcroft et al., 1997; Gallus, 2000; Gallus and Klemp, 2000; Janjic and DiMego, 2001; Gavrilov, 2002) indicate that more problems should be expected at higher resolutions.

The shaved cell approach has even more problems with complex lower and internal boundary conditions. Also, as with the step-mountains, the vertical resolution is reduced over elevated terrain that poses problems for physical parameterizations. Addressing this problem, the group at the German Weather Service is working on an approach where the contributions of the model’s physical parameterizations are evaluated on a separate, terrain following grid.

Thus, the terrain-following hybrid pressure-sigma vertical coordinate (Arakawa and Lamb, 1977) has been chosen as a compromise (Janjic, 2003). With the hybrid coordinate, the coordinate surfaces are flat above and away from the mountains. Over the mountains the hybrid coordinate has increased vertical resolution, and the equations are continuous, without the computational internal boundary conditions. Since the hydrostatic pressure is currently used as the vertical coordinate above 400 hPa, the possible inaccuracies due to the sloping coordinate surfaces are restricted only to about the lower half of the mass of the atmosphere. Note that, generally, the largest errors in the sigma coordinate occur in the stratosphere. Thus, the most serious problems associated with the sloping sigma surfaces are eliminated.

The example in Fig. 4 is shown in order to illustrate the advantages of the hybrid coordinate. In this example the two-dimensional model used in the nonhydrostatic tests was integrated for 12 hours using 10 km resolution in the horizontal and 60 equidistant layers in the vertical. The integration domain with cyclic lateral boundary
conditions was 200 grid distances wide and extended up to 10 hPa. The experiments started from an atmosphere at rest and in hydrostatic equilibrium. In such an atmosphere the pressure gradient force is zero. Following Phillips (Phillips, 1974; Janjic, 1977), the vertical profile of geopotential had the form

\[
\Phi = 1054.5 + 80397.3 \, z - 7659.0 \, z^2 + 1110.0 \, z^3
\]

\[
z = -\ln p + \ln 100000.
\]

In the middle of the domain, a mountain about 1900 m high was defined specifying the surface pressure as

\[
p_s = 100000. - \frac{20000}{1 + \left(\frac{x-x_c}{x_a}\right)^2}.
\]

Here \( x_c \) is the center of the domain, and \( x_a \) is the 10 grid distances wide mountain halfwidth. The spurious \( u \) wind component developing due to the pressure gradient force errors in the sigma coordinate is shown in the upper panel of Fig. 4. The corresponding errors in the hybrid coordinate with the boundary between the pressure and sigma domains at about 400 hPa are shown in the lower panel of the figure. The plots cover the entire integration domain in the horizontal, and extend up to 30 km in the vertical. The ticks on the vertical axes are drawn every 1000 m. Dashed lines represent negative values. As can be seen from the figure, the errors in the sigma coordinate are for an order of magnitude larger and reach about 4.5 m s\(^{-1}\). The error in the hybrid coordinate remains close to zero everywhere, and its maxima reach only about 0.5 m s\(^{-1}\). As pointed out, the sigma coordinate errors are concentrated mainly at higher altitudes, in the area of the tropopause and in the stratosphere.

The geopotential and the nonhydrostatic pressure are defined at the interfaces of the layers, while all three velocity components and temperature are carried in the middle of the model layers (Janjic, 1977). The vertical velocity is defined at the E grid mass points.

6. SPATIAL DISCRETIZATION

The discretization principles applied in the NMM, and thoroughly tested in practice in its hydrostatic predecessors, the Eta and the HIBU models, have been (Janjic, 1977, 1984):

- Conservation of selected integral properties, and in particular, following Arakawa, the control over the nonlinear energy cascade by the conservation of energy and enstrophy in case of nondivergent flow;
- Cancellation of the contributions of the pressure gradient force and the \( \omega \alpha \) term of the thermodynamic equation to the total energy generation, and consequently consistent
transformation between the kinetic and potential energy; and
- Minimization of the errors due to sloping sigma surfaces.

Although designed following the same general principles, the specific numerical schemes employed evolved significantly over time and over about two orders of magnitude in horizontal resolution. For example, the problem of the sloping sigma surfaces was first addressed by minimizing the sigma coordinate pressure gradient force errors (Janjic, 1977), then by the step-mountain blocking (Mesinger et al., 1988) in the Eta model, and finally by the already discussed hybrid pressure-sigma coordinate in the NMM. The treatment of the $\omega\alpha$ term has certainly played an important role in the treatment of orography as well (Janjic, 1977). Yet, perhaps the most significant upgrade was the introduction of the new schemes for calculating the contribution of the nonlinear advection terms and the horizontal divergence operators (Janjic, 1984; Gavrilov and Janjic, 1989). In the current model formulation, all divergence operators are computed using the fluxes between each point and its eight nearest neighbors (Janjic, 1984).

In the case of rotational flow and cyclic boundary conditions, the Janjic (1984) scheme for horizontal advection of momentum on the $E$ grid conserves the following properties:

- Enstrophy as defined on the staggered grid $C$
  \[
  \sum_{i,j} \left( \delta_{x',x'} \psi + \delta_{y',y'} \psi \right)^2 \Delta A ,
  \]
  and also, separately,
  \[
  \sum_{i,j} \left( \delta_{x',x'} \psi \right)^2 \Delta A , \quad \sum_{i,j} \left( \delta_{y',y'} \psi \right)^2 \Delta A \\
  \sum_{i,j} \left( \delta_{x',x'} \psi \delta_{y',y'} \psi \right) \Delta A ;
  \]
- Rotational kinetic energy as defined on the staggered grid $C$
  \[
  \sum_{i,j} \frac{1}{2} \left( \delta_{y',\psi} \right)^2 \Delta A \quad \text{and} \quad \sum_{i,j} \frac{1}{2} \left( \delta_{x',\psi} \right)^2 \Delta A ;
  \]
- Rotational kinetic energy as defined on the semi-staggered grid $E$
  \[
  \sum_{i,j} \frac{1}{2} \left( \delta_{y,\psi} \right)^2 \Delta A \quad \text{and} \quad \sum_{i,j} \frac{1}{2} \left( \delta_{x,\psi} \right)^2 \Delta A ;
  \]
- Rotational momentum as defined on the staggered grid $C$
- Rotational momentum as defined on the semi-staggered grid $E$.

The $Z$ grid equivalent of the $E$ grid (Janjic, 1984) used to define the quantities (6.1)-(6.3) is shown in Fig. 5 together with the orientation of the coordinate axes $x, y$ and $x', y'$ appearing in (6.1)-(6.3). As before, $\chi$ and $\psi$ are the velocity potential and the stream function, respectively, and $h$ stands for mass point variables. The symbol $\Delta A$ denotes the area of the grid boxes, and the summation sign with the subscripts $i,j$ represents the summation in the horizontal.

In case of general flow, the scheme conserves:

- Kinetic energy as defined on the semi-staggered grid $E$

Fig. 5. The Z grid equivalent of the E grid. Orientations of the coordinate axes $x, y$ and $x', y'$ are indicated.
\[
\sum \frac{1}{2} \left[ (\delta_x \chi - \delta_y \psi)^2 + (\delta_y \chi + \delta_x \psi)^2 \right] dV, \quad (6.4)
\]
and also,
\[
\sum \frac{1}{2} \left[ (\delta_x \chi)^2 \right] dV
\]
and
\[
\sum \frac{1}{2} \left[ (\delta_y \chi)^2 \right] dV;
\]

- Momentum as defined on the semi-staggered grid \( \mathbb{E} \).

In (6.4), the summation is performed over all grid points, and the symbol \( dV \) denotes the grid box volume in hydrostatic vertical coordinates. The scheme for horizontal advection of temperature conserves the first and the second moments of temperature.

Both of the two discussed options for dealing with the problem of the semi-staggered grids with the frequencies of the shortest gravity-inertia waves (Janjic, 1979; Janjic et al. 1998) were tested in the model. The deaveraging option is computationally efficient and requires only a few simple iterations on the hydrostatic pressure tendency. Although it is non-dissipative, and therefore physically better founded, the dramatic improvement of the frequencies of the shortest gravity-inertia waves achieved requires that the time step be significantly reduced which leads to reduced computational efficiency of the model. Thus, it is not obvious which of the two options should be given preference in practice (Janjic, 2003).

Concerning the basic dynamical variables, quadratic conservative vertical advection is used. In addition to the material surface boundary conditions requiring that the vertical velocity be equal to zero at the top and at the bottom of the model’s atmosphere, vertical boundary conditions are needed for the nonhydrostatic deviation of pressure. It is assumed that the nonhydrostatic pressure deviation vanishes at the top of the model’s atmosphere, while its vertical derivative vanishes at the bottom (Janjic et al., 2001).

The centered conservative schemes used for advection of the basic dynamic variables develop problems in case of advection of positive definite scalars with large spatial variation, such as specific humidity, cloud water, or turbulence kinetic energy. For this reason, an upgraded version of the scheme used for advection of passive substances in the NCEP Eta model (Janjic, 1997) is applied. The scheme consists of three steps. In the first step an upstream biased scheme is used to advect the passive substance. In the second step, antifiltering is applied, optimized in such a way as to minimize computational dispersion in sheared flows. In the third step, forced conservation is imposed.

7. TIME DIFFERENCING

Following the same proven principles for time discretization as in its hydrostatic predecessors, the hydrostatic core of the equations used in the NMM is split into two energy conserving subsystems (Janjic, 1979, 2003; Janjic et al., 2001). These two subsystems are solved using different time stepping methods. The splitting is not done by automatically separating all the advection terms since the advection of pressure in the omega-alpha term of the thermodynamic equation should not be separated from the contribution of the pressure gradient force. An economical forward–backward scheme (Ames, 1969; Gadd, 1974; Janjic and Wiin-Nielsen, 1977; Janjic, 1979) is used for the gravity-inertia waves. The Adams–Bashforth scheme is applied for the contributions of the horizontal advection terms. The advection time step is the same as that for the gravity-inertia waves. Note that the ratio between the advection time step and the gravity-inertia time step is restricted to only about 2 by the CFL criterion on the semi-staggered grids, where longer gravity-inertia steps can be used than on the staggered grid \( \mathbb{C} \). The trapezoidal scheme for the Coriolis force terms (Janjic and Wiin-Nielsen, 1977) has been replaced by the Adams-Bashforth scheme in order to avoid overestimating the amplitude of the divergent part of the flow (Janjic and Wiin-Nielsen, 1977).

For the vertical advection, the Matsuno scheme has been replaced in later versions of the NMM by the neutral Crank-Nicholson scheme. Namely, as a legacy from the Eta model, which required very high vertical resolution in order to resolve topography and the PBL over elevated terrain, the NMM is often run with higher vertical resolution than necessary with the hybrid coordinate, so that the vertical advection sometimes dangerously approaches or exceeds the limit imposed by the CFL criterion.

As pointed out in Janjic (2003), a novelty in the treatment of the nonhydrostatic terms is that the iterative method for solving the vertical implicit
pressure equation discussed in Janjic et al. (2001) has been replaced by a direct solver. This modification has brought a further improvement of the computational efficiency of the model.

8. OPERATIONAL APPLICATIONS

Since July 2002, the NMM has been run operationally in NCEP High Resolution windows in six nested domains (West, Central, East, Alaska, Hawaii, Puerto Rico) shown in Fig. 6. The horizontal resolution is 8 km for all domains except for Alaska where the horizontal resolution is 10 km. On 21 June 2005 the gridpoint spacing of the nests was decreased to about 5.1 km. The model has 60 unequally spaced levels in the vertical. In addition, the model is used for fire weather forecasting and other purposes on call.

Fig. 6. The six High Resolution Windows: Western, Central, Eastern, Alaska, Hawaii and Puerto Rico domains embedded into the Eta domain (solid black line).

The topography is defined as grid-box means of the USGS 30'' global Digital Elevation Model data. Except in ten rows along the lateral boundaries, no smoothing or filtering is applied. Although unfiltered topography introduces forcing at the smallest resolvable scales that can adversely affect particularly precipitation forecasts, this choice was made hoping that realistic topography would result in better forecasting of low-level flow over complex terrain, and in particular, in more realistic simulation of channeling effects of topography.

The model as run in the nested domains does not have its own dedicated data assimilation system. The initial and boundary conditions are defined by interpolation of the operational Eta model data. The Eta model is run with 12 km resolution and 60 levels in the vertical.

In the two small domains the model is run twice a day (Hawaii cycles 00Z and 12Z, Puerto Rico cycles 06Z and 18Z). In the remaining four domains, the model is run once a day starting from 00Z (Alaska), 06Z (West), 12Z (Central) and 18Z (East). The forecasts are computed up to 48 hours.

The computational efficiency of the model has been very high, and substantially higher than the computational efficiency of most nonhydrostatic models. Moreover, further significant improvement of the computational efficiency of the model is possible. The model has been highly reliable.

Results of a recent comprehensive diagnostic study demonstrate the potentials of the WRF-NMM (http://www-ad.fsl.noaa.gov/fvb/rtvs/wrf/retro_runs/). In this study the forecasts have been evaluated using the standard NCEP verification package. This package computes a number of traditional scores versus observations, which in the free atmosphere basically measure how well large-scale features are predicted. The scores computed for near surface variables are more representative for smaller-scale features.

The WRF-NMM scores obtained in this study have been very good in relative terms despite numerous factors that have had adverse effect on the quality of the forecasts. Among these, perhaps the most important are inconsistently defined initial and boundary conditions, small integration domains, an imperfect physical package, unfiltered mountains, as well as seemingly insignificant computational inaccuracies and inconsistencies in the preparation of the initial and boundary conditions. As an example of such seemingly insignificant computational inaccuracies, after the completion of the study, an inconsistency in conversion of geopotential data into temperature in the initial and boundary conditions was found that was causing significant positive bias in temperature at upper model levels. The impact of such a bias is clearly visible in the study in the scores for temperature and geopotential at the highest model levels.

For more information on the model performance as measured by the NCEP standard verification system, the reader is referred to the quoted study and other NCEP web sites. In subsequent sections various other aspects of model performance will be discussed.
9. THE NONLINEAR DYNAMICS OF THE NMM

The problem of accumulation of energy at the small scales due to false nonlinear energy cascade and other computational problems has been known for a long time (c.f. Phillips, 1956, 1959; Arakawa, 1966, 1972; Sadourny, 1975a, 1975b). If unchallenged, this process may lead to numerical instability. Historically, the simulated spectrum has been kept under control either by removing the spurious energy at the small scales by numerical filtering, or by preventing excessive nonlinear energy transport to the small scales by energy and enstrophy conservation as proposed by Arakawa (1966). The Arakawa approach is physically better founded, but requires application of carefully designed, complex finite-difference schemes.

Nastrom and Gage (1985) examined measurements made by commercial aircraft and found that one-dimensional kinetic energy spectra along their flight-paths in the lower stratosphere and in the upper troposphere follow the \(-3\) slope on the larger scales, and the \(-5/3\) slope on the scales from several hundred kilometers to several kilometers. Several possible explanations for the \(-5/3\) spectral shape at the small scales have been proposed (e.g. Gage, 1979; Lilly, 1983; Gage and Nastrom, 1986; Tung and Orlando, 2003). They include the downscale nonlinear energy cascade and an inverse cascade from smaller to larger scales. Using two-parameter quasi-geostrophic dynamics, Tung and Orlando (2003) demonstrated that given enough time, the \(-5/3\) spectral range can be generated through downscale cascade of energy. Similar statistical properties of the spectra were obtained in extended simulations using the GFDL SKYHI model with modest, but still higher than usual resolution for a climate model (e.g., Hamilton et al., 1999; Kosyik and Hamilton, 2001).

It is interesting what shape of energy spectrum can be realistically expected in short-range integrations of mesoscale models. The statistical properties of atmospheric spectra typically are investigated in extended integrations (tens or hundreds of days), and the spectra are averaged over long periods (tens or hundreds of days) in order to ensure that statistical equilibrium is reached. The need for extended integrations and long averaging periods arises due to the time scale of the nonlinear cascade. In addition, the mesoscale integration domains are typically smaller than the large-scale atmospheric disturbances that feed the downscale nonlinear cascade, so that the large-scale disturbances cannot transfer the energy to the small scales in an adequate manner across the domain scale. Thus, if the energy in the small-scale part of the spectrum is missing in the initial data, where can it come from in the integration? Considering the temporal and spatial limitations, it appears that physical or spurious sources of energy other than the downscale nonlinear cascade from the large-scale motions are needed in order to develop and maintain the \(-5/3\) spectra in mesoscale atmospheric models.

In order to clarify this issue, several tests have been performed using two versions of the NMM, the parallel version on the \(E\) grid and the PC version on the \(B\) grid (NMM-B). Both versions were designed applying exactly the same principles in the discretization of the model dynamics, and share the same physical package (Janjic, 2003). They are both well qualified for investigating atmospheric spectra. They conserve energy and enstrophy, which generally improves the accuracy of nonlinear dynamics and controls the downscale nonlinear energy cascade restricting spurious energy transfer toward smaller scales. In addition, the NMM and the NMM-B use hybrid pressure-sigma vertical coordinate, so that, except for the errors propagating from below, in the upper troposphere and in the stratosphere where the Nastrom and Gage (1985) spectrum was observed, they are free of the sigma coordinate errors that are largest at higher altitudes. Thus, two major sources of small-scale noise, the spurious energy cascade and the sigma coordinate errors, are well controlled. The energy conservation eliminates the need for excessive dissipation (either explicit or built into the finite-difference schemes), and explicit formulation of major dissipative processes allows precise “dosage” of dissipation.

A test with the operational setup of the NMM over the Central Domain (8km, 60 levels), but in the sigma coordinate, generated spectra on constant pressure surfaces in the upper troposphere and in the stratosphere that were very similar to the observed ones. An example of the 300 hPa spectrum (white diamonds) averaged over forecast times from 24 to 36 hours is shown in Fig. 7. The forecast started from the September 08, 2003, 12Z GFS data. In agreement with the observations, the model develops the spectrum (white diamonds) following the \(-3\) (gray squares) and \(-5/3\) (black triangles) slopes. However, the
interpretation of this result is not straightforward. Namely, when the operational setup of the NMM is run in the hybrid coordinate, the spectrum at 300 hPa remained steeper than the $-5/3$ slope at the small scales, and approached the $-3$ slope (not shown). Furthermore, as can be seen from Fig. 8, the spectrum of the square of unfiltered topography in the Central domain (white diamonds) closely follows the $-5/3$ slope in the mesoscale range. The topography spectrum, together with the lack of the $-5/3$ range in the hybrid coordinate run, indicate that the model spectrum may have been simply saturated at the small scales by the noise due to the sigma coordinate errors. If so, the computational noise would be mistaken for the atmospheric spectrum.

As another example, the time average over forecast hours 36-48 of the spectra at 300 hPa (white diamonds) obtained in the NMM run for the case of Hurricane Isabel in the East domain with the resolution of 8 km and 60 levels is shown in Fig. 9 using the same arrangement as before. The model was run from the September 17, 2003, 18Z Eta data. In order to facilitate and accelerate development of the small-scale part of the spectrum, the lateral diffusion was turned off, and only weak divergence damping was retained. The East domain is much less mountainous than the Central domain, and in contrast to the previous run in the Central domain, this run was performed using the hybrid coordinate. Thus, the impact of the sigma coordinate errors was substantially reduced in the upper troposphere and in the stratosphere. As can be seen from the figure, the spectrum spun up by the model agrees remarkably well with the observed spectrum. Apparently, the tropical storm provided a sufficiently strong physical energy source at the small scales, so that the model spectrum indeed could be generated by at least some of the mechanisms responsible for generating the spectrum in the real atmosphere.

In order to further investigate possible mechanisms responsible for the development of the spectrum observed by Nastrom and Gage (1985), the NMM-B was run using the resolution of 15 km in the horizontal and 32 levels in the vertical in a domain of the same size as the Central domain, but over the Atlantic Ocean. Hence, the possibility of mountains influencing the energy
spectrum was eliminated. As before, the lateral diffusion was turned off in order to facilitate and speed-up accumulation of energy at the small scales, and only weak divergence damping was retained. The model was run from the September 7, 2003, 12Z GFS data. During most of the 48 hour integration a vigorous extra-tropical system was present near the northern boundary of the domain. Hence, again, there was a significant physical source of energy at the small scales through the forecast period.

Two experiments were run with the described setup. In one of them the physical package was turned off, while in the other the model was run with the full physics. The evolution of the NMM-B spectrum at 300 hPa over the Atlantic Ocean (white lines) at 300 hPa in the East domain. The –3 (gray squares) and –5/3 (black triangles) slopes are shown for comparison. Run starting from 18Z, 09/17/2003 (Hurricane Isabel), Eta data, 8 km, 60 levels resolution. No lateral diffusion, weak mass divergence damping. Fig. 9. Time average over 36-48 hours of the NMM spectra (white diamonds) at 300 hPa in the East domain. The –3 (gray squares) and –5/3 (black triangles) slopes are shown for comparison. Run starting from 18Z, 09/17/2003 (Hurricane Isabel), Eta data, 8 km, 60 levels resolution. No lateral diffusion, weak mass divergence damping.

10-11, by the end of the forecast, the model (white lines/diamonds) failed to develop the –5/3 (black lines/triangles) spectral range, and the slope of the spectrum remained generally steeper than the –5/3 at all times.

On the other hand, the evolution of the NMM-B spectrum at 300 hPa over the Atlantic Ocean (white lines) in the test with physics is shown in Fig. 12 at 6-hour intervals (0-24 hours, top to bottom in the left column, and 30-48 hours, top to bottom in the right column). The –3 (gray lines) and –5/3 (black lines) slopes are again shown. The 36-48 hours average of the spectra (white diamonds) is shown in Fig. 13. As can be seen from Figs. 12-13, this time the model (white lines/diamonds) developed the –3 (gray lines/squares) and –5/3 (black lines/triangles) spectral ranges that agree very well with observations. However, as can be seen from Fig. 12, it needed up to 24 hours to do so. Note that the sharp drop-off of the spectrum that is usually seen in numerical simulations (e.g. in Fig. 7) at the small-scale end of the spectrum is missing here due to weak dissipation.

As expected from theoretical considerations, and as confirmed by the examples shown, the nonlinear dynamics of the NMM work well, and reproduce the spectrum observed in the atmosphere successfully. Moreover, the simulated spectrum shows no sensitivity to horizontal resolution in the considered resolution range. However, a sufficiently strong physical or spurious energy source on the small scales is needed in order to develop the –5/3 spectral range. In the examples considered, the physical energy sources were the physical parameterizations in case of Hurricane Isabel and the storm over the Atlantic. The spurious energy was supplied by the sigma system errors in the run over the Central domain.

Unfortunately, it is often difficult to establish whether the appropriate conditions for reproducing atmospheric spectra in short-range mesoscale runs are satisfied. As experience has shown, unless special care is taken, computational noise due to model imperfections is likely to contaminate the small-scale part of the spectrum. The dissipation may keep the level of the noise at a tolerable level, but the noise spectrum modified by the dissipation should not be mistaken for the observed atmospheric spectrum, nor can it be taken as a proof of accuracy of nonlinear energy transfers (Sadourny, 1975a). The fundamental problem here is to keep the physical, and eliminate spurious energy sources at the small scales. In
the NMM two major sources of small-scale computational noise, the sigma coordinate errors.

Fig. 10. Time evolution of the NMM-B spectra (white lines) at 300 hPa over Atlantic Ocean in the experiment with no physics. The –3 (gray lines) and –5/3 (black lines) slopes are shown for comparison. Run starting from 12 UTC, 09/07/2003, GFS data, 15 km, 32 levels resolution. No lateral diffusion, weak mass divergence damping. Plots every 6 hours, top to bottom, 0-24 left column, 30-48 right.
It is also interesting how the NMM nonlinear dynamics reproduce the energy spectrum in the case of three-dimensional turbulence. In order to address this problem, the NMM-B was run with horizontal resolution of 1 km, and an average vertical resolution of about 500 m. The horizontal domain had 112 by 112 points. Double periodic boundary conditions were specified along the lateral boundaries. Similarly as in Takemi and Rotunno (2003), the model was initialized with the vertical thermodynamic structure of the Fort Sill storm of May 20, 1977, and the initial wind was set to zero. The spectrum of $w^2$ at the 700 hPa level (c.f. Takemi and Rotunno, 2003) corresponding to decaying turbulence generated by moist convection was obtained by averaging the spectra between forecast hours 3 and 4. The time averaged spectrum (white diamonds) is shown in Fig. 14 together with the $-5/3$ slope (black triangles). The agreement between the computed and the theoretical spectrum is again evident.

10. HIGHLIGHTS OF MODEL PERFORMANCE

In terms of performance, generally, the model has been highly competitive with mature operational high-resolution NWP models, despite the fact that it has been handicapped by inconsistent initial and boundary conditions taken from the Eta model, relatively small integration domains, and suboptimal tuning of the physical package. Although it is difficult to compare directly the performance of the NMM to that of e.g. the Eta model, statistical scores and numerous examples (Black et al., 2002, Janjic et al. 2003) indicate that the NMM adds value to the forecasts of the driving Eta model. This applies particularly to the details of flow over complex terrain.

However, as illustrated in Fig. 15, the most dramatic differences between the Eta and the NMM are seen in vertical structures developing due to effects of topography. The panels in the left column are from the 12 km Eta run, and the panels in the right column are from the NMM East domain run. The middle and bottom panels of the two columns represent 12 hour and 15 hour forecast cross sections, respectively, starting from 18Z, January 7, 2003. The cross sections are taken along the blue lines in the top panels. The topography is indicated in the top panels by color shading with the contours at 100, 175, 250, 375, 500, 750, 1000, 1250 etc. meters, and by the shaded area at the bottom of the cross sections. The blue and brown contour lines indicate the negative (upward) and positive (downward) vertical velocity $\omega = dp/dt$, respectively. The contour interval is 0.2 Pa s$^{-1}$. The potential temperature is represented by the dashed red contour lines with the contour interval of 4 degrees. The background color shading in the cross sections represents isotachs with the contour interval of 10 knots.

Apparently, the NMM develops much more vigorous mountain waves than the Eta, and the wavelength of the mountain waves in the NMM is much shorter than in the Eta. The effect of vertical transport of momentum is visible in the bottom panels, particularly in the case of the NMM. Apparently, flying through the Eta cross sections would be quite a different experience from flying through the NMM profiles.
Fig. 12. Time evolution of the NMM-B spectra (white lines) at 300 hPa over Atlantic Ocean in the experiment with physics. The –3 (gray lines) and –5/3 (black lines) slopes are shown for comparison Run starting from 12 UTC, 09/07/2003, GFS data, 15 km, 32 levels resolution. No lateral diffusion, weak mass divergence damping. Plots every 6 hours,
Fig. 13. Time average over 36-48 hours of the NMM-B spectra (white diamonds) at 300 hPa over Atlantic Ocean in the experiment with physics. The $-3$ (gray squares) and $-5/3$ (black triangles) slopes are shown for comparison. Run starting from 12 UTC, 09/07/2003, GFS data, 15 km, 32 levels.

An experimental 12 hour forecast of the sea level pressure starting from September 17, 2003, 12Z GFS data (Hurricane Isabel) is shown in Fig. 16. The predicted pressure in the center of the storm was 952 hPa, while the observed value at that time was 953 hPa. The 30 hour forecast of the accumulated 3 hour precipitation for this case is shown in Fig. 17. The landfall occurred at about this forecast time, and as can be seen from Fig. 17, it was accurately predicted by the model. This example demonstrates that the NMM has the ability to spin up and maintain realistically deep tropical storms and to predict their tracks accurately. The work on the development of the Hurricane WRF based on the WRF-NMM is well under way (e.g. Surgi et al., 2004).

Particularly interesting results were obtained during the joint 2004 SPC/NSSL Spring Program (Kain et al., 2005). This was a carefully controlled experiment in which the model was run once per day starting from 00Z at near-cloud-resolving resolution of 4.5 km and without parameterized convection. As an example, the 24 hour forecast of accumulated 1 hour precipitation valid at 00Z April 21 is shown in the upper panel of Fig. 19.

Fig. 14. Decaying 3D turbulence. Time average over hours 3-4 of the NMM B spectra of $w^2$ (white diamonds) at 700 hPa in a domain with double periodic boundary conditions. While the verifying radar reflectivity is shown in the lower panel. As can be seen from the figure, the predicted precipitation pattern and timing were very similar to the verification. During the experiment, the model demonstrated the ability to spin up severe convective systems more frequently, and with a stronger signal, than if this were happening only by chance. This was reflected in the verification scores (Kain et al., 2005) that showed that the WRF-NMM with near-cloud-resolving resolution for the first time clearly outperformed the operational NCEP mesoscale forecasts with parameterized convection. This result also suggests that further improvements in deterministic forecasting of severe weather phenomena still may be achieved with increased resolution.

11. CONCLUSIONS

The NCEP nonhydrostatic mesoscale model (NMM) (Janjic et al., 2001; Janjic, 2003) has been formulated building on the experiences of high-resolution hydrostatic numerical weather forecasting. In this way, the favorable features of hydrostatic model formulation are preserved in the range of validity of the hydrostatic approximation.
Fig. 15. The 12 km Eta (left column) and the 8 km NMM East domain (right column) cross sections taken along the blue lines in the top panels. The middle and bottom panels are 12 hour and 15 hour forecasts, respectively, starting from 18Z, January 7, 2003. The topography is indicated in the top panels by color shading with the contours at 100, 175, 250, 375, 500, 750, 1000, 1250 etc. meters, and by the shaded area at the bottom of the cross sections. The blue and brown contour lines are the negative (upward) and positive (downward) values of vertical velocity \( \omega = \frac{dp}{dt} \), respectively. The contour interval is 0.2 Pa s\(^{-1}\). The dashed red contour lines are potential temperature with the contour interval of 4 degrees K. The background color shading in the cross sections represents isotachs with the contour interval of 10 knots.
Fig. 16. An experimental 12 hour forecast of the sea level pressure starting from September 17, 2003, 12Z (Hurricane Isabel). The NCEP GFS data were used for initial and boundary conditions.

Fig. 17. The 30 hour forecast of the accumulated 3 hour precipitation indicating the location of the landfall. The forecast started at September 17, 2003, 12Z (Hurricane Isabel). The NCEP GFS data were used for the initial and boundary conditions.

No linearization or additional approximation is required with this approach.

The nonhydrostatic module can be turned on and off, so that the same model can be run in both hydrostatic and nonhydrostatic modes. This allows easy comparison of hydrostatic and nonhydrostatic solutions of an otherwise identical model. This feature also allows the model to be run in the hydrostatic mode at lower resolutions with reduced computational cost. This is an advantage in the case of unified global and regional forecasting systems designed for a wide range of horizontal resolutions.

At very high resolutions, a two-dimensional version of the model successfully reproduced the classical nonhydrostatic solutions demonstrating the soundness of the formulation. Compared to the hydrostatic version of the model, no additional computational boundary conditions at the top have been needed in real data runs in a wide range of horizontal resolutions.

The nonlinear dynamics of the NMM model demonstrate the ability to reproduce the observed atmospheric spectrum (Nastrom and Gage, 1985). Moreover, at higher resolution, the NMM
successfully reproduces the theoretical spectrum of decaying three-dimensional turbulence.

Despite application of sophisticated numerical methods, the computational efficiency of the model is very high, and substantially higher than the computational efficiency of most nonhydrostatic models. Moreover, further significant improvement of the computational efficiency is possible. This will allow further increase of the resolution and application of more sophisticated physical parameterizations. The model has been highly reliable.

In terms of performance on the synoptic scales, generally, the model has been highly competitive with mature high-resolution NWP models, despite the fact that it has been handicapped by inconsistent initial and boundary conditions, relatively small integration domain, and little retuning of the physical parameterizations. Plans for future work address each of these issues. The model demonstrates ability to add value to the forecasts produced by the driving model (the Eta model).

Significant differences between the NMM and the NCEP hydrostatic high resolution Eta model can be seen on smaller scales. The differences are particularly striking in mesoscale vertical structures developed by the two models over complex topography. In addition to resolution, the representation of mountains and the nonhydrostatic dynamics are believed to have played a role in producing such different results.

With near-cloud-resolving resolution, the model was able to spin up severe convective systems more frequently, and with a stronger signal, than if this were happening only by chance. This result indicates that further improvements in deterministic forecasting of severe weather phenomena still may be possible with increased resolution.

REFERENCES

______, 1979: Forward–backward scheme modified to prevent two–grid–interval noise and


