High-resolution Hurricane Forecasts

Chris Davis, Wei Wang, Steven Cavallo, James Done, Jimy Dudhia, Sherrie Fredrick, John Michalakes, Ginger Caldwell, and Tom Engel

National Center for Atmospheric Research[1]

Boulder, Colorado

Ryan Torn

University at Albany, SUNY

Albany, New York

Submitted to

Computing in Science and Engineering

Revised, April, 2010

Corresponding Author:

Christopher A. Davis
P.O. Box 3000
Boulder, CO 80307
cdavis@ucar.edu
Phone: 303-497-8990
Fax: 303-497-8181

[1] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Abstract

The authors discuss the challenges of predicting hurricanes using dynamic numerical models of the atmosphere-ocean system. The performance of particular model is investigated for a large sample of Atlantic tropical cyclones from the 2005, 2007 and 2009 hurricane seasons. The model, derived from the Weather Research and Forecasting (WRF) model, is capable of resolving the deep convective motions within a hurricane and the eye and eye wall of the storm. The use of finer resolution leads to demonstrably improved predictions of storm intensity compared with predictions from coarser resolution models. Possible future real-time applications of this model in a high-performance computing environment are discussed using hurricane Bill (2009) as an example. These applications are well suited to massively parallel architectures.

1. Introduction

The prediction of hurricane intensity remains a daunting challenge even after four decades of research. The intrinsic difficulties lay in the vast range of spatial and temporal scales of atmospheric motions that affect tropical cyclone intensity. The range of spatial scales is literally millimeters to 1000 kilometers or more. Atmospheric dynamical models must account for all these scale simultaneously. Being a nonlinear system, these scales can interact. While approximations must be made to keep computations finite, there is a continued push for finer resolution to capture as many of these scales as possible.

From a computational feasibility standpoint, the minimum grid lengths in a model that captures the full extent of the hurricane circulation (e.g. 1000 km) are a few hundred meters in the horizontal and perhaps 50-100 m in the vertical. Most current weather prediction models[1] are grid-based rather than spectral based (Fourier or some other basis function). Statistical analysis of energy spectra reveal that motions with scales smaller than about 6-7 grid points are not well resolved[2]. Therefore, the minimum physical length scales resolvable are more nearly 1 km in the horizontal and perhaps 300 m in the vertical. But with current computing capability, timely numerical forecasts must be run on much coarser grids.

What does this mean for hurricane forecasts? We believe that it is important to resolve clouds, at least the largest cumulonimbus-producing thunderstorms. These clouds have a horizontal scale of at most a few kilometers and thus can only be resolved with a grid spacing near, or even less than, one kilometer in the horizontal. These clouds span the vertical extent of the troposphere, 12-16 km, and so are relatively easy to resolve in the vertical as long as 30-40 layers are used. Nevertheless, to cover the region affected by a hurricane in a 5-day forecast requires a horizontal domain of perhaps 5,000 km. A volume with a grid increment of 500 m in the horizontal and 250 m in the vertical over a domain of depth 25 km contains roughly 10^{10} grid points.

The equations of motion are first-order in time, hence knowing the model state at one time allows one, in principle, to predict the state a short time later. The length of this "time step" must be limited to ensure numerical stability. This limit prohibits the fluid and waves within the fluid from traveling more than one grid increment in one time step. For grid lengths of a kilometer or so, the time step is roughly 10 seconds; it decreases to 1 second or so on grids of 100 m. Owing to the need to decrease the time step inversely as the resolution increases for computational stability, a doubling of model resolution costs a factor of 16 in computations! More than 400,000 time steps would be needed to integrate a 5-day forecast on a 100-m grid. Couple this with the need to integrate many variables (momentum, water in several forms, temperature, etc.) and you have a problem in which the time tendencies of variables must be estimated about 10^{16}-10^{17} times for a single forecast. Each estimate requires numerous floating point operations, placing the total number of operations in the range of 10^{18}. To produce a forecast in a few hours of wall-clock time requires nearly petascale computing capability.

Because actual hurricane forecasts are integrated on machines that have other demands, and because the capability of these machines is presently measured in teraflops, not petaflops, we must be able to integrate forecasts on coarser grids with other enabling strategies. What strategies and how coarse must these grids be? These questions are at the heart of the high-resolution hurricane (HRH) test that was sponsored by the National Oceanic and Atmospheric Administration (NOAA). The goal was to see whether forecasts on grids of roughly 1-2 km in the horizontal produced better forecasts of hurricane intensity than those on coarser grids of roughly 10 km.

Enabling strategies to allow high-resolution include simplifying the representation of physical processes in the atmosphere, because we cannot account for every turbulent eddy or precipitation particle. Some motions can be resolved, but others must be represented implicitly in terms of resolved scale of motion. In the case of hurricanes there is potentially a significant advantage of representing the deep cumulonimbus clouds explicitly rather than implicitly. The maximum grid spacing at which this explicit representation is possible is roughly 1-4 km; models with grids at 10 km nearly always implicitly represent the effects of deep convection. Clearly the number of computations needed for explicit representation of thunderstorm clouds is much greater than for an implicit representation, but we will demonstrate that the finer resolution improves many aspects of hurricane forecasts.

Part of the added cost can be mitigated by recognizing that it is probably more important to treat explicitly clouds near and within the eye wall of the hurricane than clouds far from the storm. Because most of the deep convection occurs within 100-200 km of the hurricane center, this is where the higher model resolution must be. Local resolution refinement is clearly an enabling strategy for producing timely forecasts with fine resolution. But how far out from the center must high-resolution extend? We address this question in section 4.

In selecting a grid spacing near or just coarser than 1 km we make a choice to resolve the deep convective eddies that span the troposphere, but not to resolve the turbulent eddies that represent mechanically generated turbulence. Turbulence acts as a break on storm intensity[3]. Furthermore, simulated storm intensity tends to increase with decreasing grid spacing until three-dimensional turbulence is resolved explicitly.

Such a result has a practical significance. It is questionable whether there is much to be gained by decreasing the horizontal grid spacing below 1-2 km unless one decreases it to around 100 m or less. This is perhaps a factor of 1000 beyond current real-time computational capability. This fact supports the focus of the present article on horizontal grid-spacing of just over 1 km, and comparing such grid spacing, which represents the forefront of hurricane forecasting, with results obtained at coarser resolution.

The hurricane model our group used for the HRH test was the Advanced Hurricane-research WRF (AHW)[4], derived from the Advanced Research WRF (ARW), where WRF is the Weather Research and Forecasting Model. The AHW is a dynamic model of the atmosphere meaning that it predicts the time evolution of the atmospheric state. Predictive equations for the three Cartesian velocity components, entropy, mass and numerous phases of water (vapor, cloud droplets, rain, snow and ice crystals) are integrated using a discrete time-stepping technique.

The AHW uses grid nesting to locally enhance resolution. Local resolution enhancement may be achieved in many ways, each of which has pros and cons. Nesting requires blending the update of the model state within and outside the nest at each time step. Nesting is interactive, meaning that the outer domain affects the inner as information sweeps across the nest boundary and into the nest from the coarse domain. In addition, the nest produces structures that can flow out into the coarse domain. In the AHW an arbitrary number of nests may be used, in principle, to achieve any desired horizontal resolution. The AHW currently employs nesting only in the horizontal directions, not in the vertical.

During AHW hurricane forecasts, the nest is repositioned every 15 minutes to re-center the vortex. The cost for moving the nest is equivalent to several nested time steps. Nest repositioning is automated based on the location of the pressure minimum. In practice a storm will progress only a few kilometers between nest movements, so the movements constitute a tiny fraction of the domain size.

The representation of the precipitation, turbulence, radiation and atmosphere–ocean coupling in AHW is accomplished using a variety of parameterizations that are fairly simple in their conceptual design. For the present article, the primary issue is the representation of clouds and precipitation, since the release of latent heat as water vapor, drawn from the ocean under the storm, condenses within clouds is the essential fuel for the storm. When the grid spacing of a model is much larger than a cloud (where "cloud" here refers to a typical cumulonimbus), the only practical strategy is to represent the net effects of clouds in terms of scales of motion the model can explicitly predict. In a sense, this amounts to prescribing the occurrence of a sub-grid-scale thunderstorm when certain conditions are met. The net heating of the air from the condensation of water is thus prescribed and this heating is transferred to the resolved motions. If the process is realistically modeled, the heating realized drives the inflow near the surface that carries with it the higher angular momentum air from the surroundings resulting in stronger winds. One can produce hurricanes in a relatively coarse-resolution model to the extent that the parameterization of clouds realistically redistributes the latent energy derived from the ocean. However, such redistribution is not a simple process.

The problem is that condensation heating is very inefficient for producing winds unless one already has a full-fledged hurricane; heating can create buoyancy oscillations which essentially carry away the heat from the storm center[5]. The explicit treatment of thunderstorms produces the cloud and predicts its growth, time-step by time-step, with nothing pre-determined. The motivation in doing so is to achieve a more realistic partitioning of the heat that goes into enhancing the circulation of the storm versus heat that is propagated away by buoyancy waves.

The remainder of the paper demonstrates the benefits of high-resolution models for hurricane prediction and discusses the need for high-performance computing to achieve this. Section 2 summarizes results from a recently completed test of fine versus relatively coarse horizontal resolution used to simulate hurricane intensity and structure. Section 3 shows results from real-time forecasts of Atlantic storms during the 2009 hurricane season. Section 4 discusses potential future applications of high resolution hurricane models using forecasts of hurricane Bill (2009) as an example. Our overall findings are summarized in Sec. 5.

2. A systematic test of high-resolution

The HRH test involved six research groups doing essentially the same thing, but with different models run in different configurations. The goal was to test the sensitivity of forecast accuracy of hurricane intensity to model resolution, and hence to the amount of computational power needed. The test involved conducting two sets of 69 simulations covering 10 Atlantic tropical cyclones[2], each using different horizontal resolution, and then evaluating whether, and by how much, higher resolution improved forecasts. It may seem that higher resolution should always result in superior forecasts, but that is not actually the case. As one resolves more scales of motion and more variability, the increased variance can lead to larger errors in individual cases. The standard metrics for quantifying accuracy use squares of differences between forecast and observed intensity, where intensity is defined as the maximum one-minute-sustained wind at 10 meters elevation. Hence, outliers are amplified. While other metrics are being considered,

[2] The storms from 2005, and number of forecasts for each, were: Emily (10), Katrina (6), Ophelia (11), Philippe (6), Rita (7), and Wilma (11). The rest were from 2007: Felix (8), Humberto (2), Ingrid (4) and Karen (4).

root-mean-squared error is the standard used by the National Hurricane Center and will be adopted here.

The model was initialized using an ensemble Kalman filter[6,7]] consisting of 96 members at 36-km grid spacing. The ensemble Kalman filter uses statistics from an ensemble of estimates of the state of the atmosphere to decide how to spread information contained in observations. The filter cycles, meaning that a 6 h forecast initialized from the previous analysis time provides the background upon which observations are analyzed. Typically the filter cycles for 2 or more days before we start our first forecast for a given storm. To do real-time forecasts, the filter runs continuously from the start of the hurricane season so that an up-to-date analysis is always available that incorporates all past and present observations. For the HRH test, the filter was initiated two days before our first forecast. Assimilated observations included surface pressure, rawinsonde (including dropsonde observations from the Gulf-Stream IV aircraft of the National Oceanic and Atmospheric Administration), data from commercial aircraft, cloud motion vectors from satellite observations, and tropical cyclone position and intensity estimates from the National Hurricane Center.

A single ensemble member was chosen for the high-resolution forecast. This member was chosen as the one closest to the observed intensity at initialization time. Pairs of forecasts, one with a single 12-km grid, the other with storm-centered, moving nests of 4-km and 1.33-km grid spacing, were integrated to 126 h or until the time the observed storm dissipated. The dimensions of the domains were: 469x424 (12-km); 201x201 (4-km nest); and 241x241 (1.33-km nest). Because the time step decreased in proportion to the grid spacing, 9 time steps of the innermost nest are required for each time step of the coarsest domain. The addition of the two moving nests increases the overall computations by roughly a factor of four compared to the 12-

km domain alone. Of the 69 pairs of forecasts initialized, 57 forecasts extended for at least 72 h, while 35 forecasts were integrated the full 126 h. Initial conditions were identical for each member of a pair.

Forecasts were integrated on NCAR's *bluefire* supercomputer, an IBM Power 575 cluster commissioned into service on June 30, 2008 and consisting of one hundred twenty-eight Power 575 nodes, each containing thirty-two POWER6 processors. The POWER6 processor is clocked at 4.7 GHz and capable of four floating point operations per clock; thus NCAR's bluefire supercomputer has a peak computation rate of seventy-seven trillion floating point operations per second (Teraflops). The AHW model, as configured for the runs described in this article, used five Power 575 nodes, or 160 POWER6 processors, and all runs for the HRH test utilized approximately one hundred fifteen thousand processor hours.

While these runs did not necessarily tax the computational capability of the *bluefire* system, they are significant in that they were performed during normal batch production computing. Just five years ago, a single comparable run would have used over half of NCAR's most powerful computer system and all runs would have required dedicating that system to AHW runs for two weeks – something we would not have considered at the time. It is thus readily conceivable, that five years hence, we will be able to integrate the entire ensemble forward at high-resolution rather than selecting, as described above, a single member for a deterministic forecast.

Hurricane intensity is defined as the maximum one-minute-sustained wind at 10 meters elevation. The overarching finding from the resolution test is that increasing resolution improves forecasts of hurricane intensity and some of the structural aspects of hurricanes, and the results have been demonstrated to be statistically significant[8]. The improvement is about 8% for the root-mean-square intensity error. Recall that the inclusion of the nests increases the computations by a factor of four. An 8% improvement may seem small given the computational cost. However, operational hurricane intensity prediction has improved by only a few percent in the past two decades or more. In comparison, the present results are encouraging.

The bias of intensity was relatively small for the high-resolution forecasts. At most lead times out to 72 hours the high-resolution forecasts had a bias of only a few knots (1 knot = 0.514 m s^{-1}). The coarse-resolution forecasts were biased low by 5-10 knots. This low bias of intensity got progressively worse as horizontal resolution decreased. The primary reason is that the inner core of the hurricane, where the strong winds are, becomes poorly resolved as the grid spacing becomes as large as the core. Recall that 6-7 grid points are necessary to truly resolve the kinetic energy at a given spatial scale. It would indeed be a large hurricane that is well resolved on a grid as coarse as, say, 36 km.

The shortcomings of coarser resolution are seen clearly as a function of observed storm intensity (Table 1). For tropical storms (TS) and hurricanes of category 1 and 2 intensity on the Saffir Simpson scale, the high-resolution forecasts exhibited slightly larger errors. However, for category 4 and 5 storms, the errors of the low-resolution forecasts became much larger than the errors in the high-resolution forecasts. Thus, high-resolution is most beneficial for predicting the most destructive hurricanes in this particular sample.

One of the primary shortcomings of intensity forecasts is the prediction of rapid intensity changes. Here we adopt the definition of an increase of maximum sustained wind of 25 knots in 24 h. To define skill, we compute the equitable threat score (ETS). If *a* denotes the number of correct forecasts of rapid intensity change (hits), *b* denotes the number of forecasts of rapid intensity change that do not occur (false alarms), *c* denotes the number of times rapid intensity change occurred but was not predicted (misses), and *d* denotes the number of correct null forecasts (correct forecasts of no rapid intensification), the ETS is defined:

$$ETS = \frac{a - \varepsilon}{a + b + c - \varepsilon}; \quad \varepsilon = \frac{(a + b)(a + c)}{a + b + c + d}$$

where ε is the number of hits expected due to random guessing. The ETS for the high-resolution AHW forecasts was 0.16, for the coarse-resolution AHW forecasts was 0.11 and for the human-generated forecasts was 0.04. It should be noted that the human-generated forecasts tend to be conservative about intensity changes. This is a conscious decision but it also reflects the fact that it is very difficult to tell from observations when a storm will rapidly intensify.

The skill of predicting rapid intensification is clearly evident in the high resolution forecasts. We suspect the reason is that while the conditions for rapid intensification are set by the environment of a storm in many cases, the vortex response to this environment is more realistic at higher resolution. This is probably due to better resolution of the inner core, but also the more realistic treatment of thunderstorms as referred to earlier. One might think of the high resolution model as more “agile”.

The prediction of storm location was not changed in a statistically significant way by the addition of high resolution. This result is not surprising and echoes the findings of other researchers. Hurricane track can be well predicted even in global weather prediction models that do not resolve the eye and associated inner core of hurricanes. The reason is usually explained that a large component of track prediction is effectively “steering” of the vortex by the flow averaged over the depth of vortex. This averaged flow typically represents the position and intensity of weather systems with length scales of 1000 km or more. Such scales are well resolved in a global model of even modest resolution.

3. Real-time Hurricane Forecasts

The AHW forecast model was also applied to real-time hurricane prediction in 2009. Some minor changes were made to the forecasting system. The primary change was the use of a nested grid-spacing of 12 km in the ensemble forecasts to better capture stronger storms. We compared 50 forecasts from AHW with operational forecasts from the Hurricane-WRF (HWRF) model from the National Centers for Environmental Prediction. The numerical methods and physical parameterizations are formulated differently in HWRF than those in AHW, and these forecasts were integrated on a coarser grid spacing of roughly 9 km, compared to the inner nest grid spacing of 1.33 km in AHW; therefore, this was not a controlled comparison. The point was to demonstrate the possibility of improved capabilities for high-resolution forecast models. As may be inferred from section 2, an important difference here is the treatment of convection, which is mainly implicit in HWRF and explicit in AHW. As may be inferred from the differences in resolution, the computing requirements for HWRF are several times smaller than for AHW.

The 2009 hurricane season in the Atlantic was unusual. No storms made landfall as hurricanes and storms were generally weak. Notable exceptions were hurricanes Bill and Fred, but these posed little threat to populated areas. Figure 1a shows the root-mean-squared intensity errors for six storms: Bill, Danny, Erika, Fred, Henri and Ida. There were some large differences in model performance that were storm dependent. The AHW forecasts were better for Bill, Danny, Erika, and Henri while HWRF forecasts of hurricane Fred were superior. This sample of forecasts is homogeneous, meaning that we only compared HWRF and AHW forecasts with the same valid time and lead time. However, no verification is performed after dissipation of the observed storm even if a viable storm still exists in the forecast.

Danny, Erika and Henri were all weak storms embedded in hostile environmental conditions. The primary factor inhibiting their development was the increase of the horizontal wind with height, referred to as vertical wind shear, which tended to tilt the storms. Looked at from the side, the displacement from the surface circulation center to the center at 5 kilometers above the ground can easily be 50-100 km in a tilted storm. Tilted storms tend to produce their convection displaced from the surface circulation center and this is an inefficient configuration for intensification. Hurricanes usually intensify when a ring of convection envelopes the center and begins to contract radially inward. This tends not to happen in storms experiencing strong vertical shear. Here “strong” is a relative term, but usually refers to a systematic increase of the wind by at least 2-3 m s^{-1} for every kilometer in the vertical. From the boundary layer to the middle troposphere (5-6 km AGL) the corresponding wind increase would be 10-15 m s^{-1}. The AHW portrayed a more realistic tilted structure of the storms, and the explicit treatment of convection more realistically kept the convection displaced from the storm center by more than 150 km.

The results for hurricane Fred were the reverse. The HWRF was able to intensify and weaken Fred at nearly the right time whereas the AHW produced realistic intensification and weakening of Fred, but it was consistently late by about one day. This created large errors for the AHW forecast. The problem in this case may have been the proximity of the storm in AHW to the lateral boundary and an associated error on the forecast vertical wind shear. Lateral boundary conditions for state variables are supplied from a global forecast model that has its own errors that will be inherited by AHW.

Position errors (Fig. 1b) indicate that overall the HWRF forecasts produced smaller position errors, but primarily for weak storms. The largest relative improvements of HWRF over AHW were for Danny and Ida. The AHW forecast position for Ida was biased westward relative to the observed position. This occurred because of the erroneous formation of a large-scale cyclone over the western Gulf of Mexico that captured Ida and steered it on a westward path for about one day whereas the real track was almost due northward.

Overall, the benefits of additional resolution in the AHW, which generally improved the intensity forecasts, were countered by worse track performance in some cases. These track errors likely stemmed from errors on spatial scales larger than the hurricane. These results remind us of the need for accurate forecasts on scales ranging from much larger than the storm (1000 km or more) down to the scale of the eye wall (10 km or less) to address both the position and intensity aspects of hurricane forecasts.

4. Toward Petascale Hurricane Forecasts

Results presented so far have been obtained with limited computational capability. Next-generation high-performance computing (HPC) systems comprising tens of thousands of CPUs may allow real-time forecast domains that cover an entire storm at resolutions of 1km or finer, and even allow for ensembles of such configurations to be integrated. In order to test forecasting as well as computational issues in advance of this capability, we have run the Hurricane Bill case on 16-thousand processors (4-thousand nodes) of the IBM Blue Gene/Q system at the Argonne Leadership Computing Facility (ALCF). Previously, WRF sustained 50 teraflops on 150-thousand MPI tasks of a Cray XT5 for a 2-billion cell idealized "Nature Run" simulation[9],[10].

Here the simulation used a somewhat larger coarse 12-km grid, but much larger nests. The innermost nest, 1.33-km grid spacing, was expanded laterally to 750x750 points and moved with the storm. This nest covered a domain that was 1000km by 1000km square, enough to cover all but the outermost tips of the storm's rain bands (Fig. 2). Compared to the nature run, this is much smaller (24 million cells), but it is roughly ten times larger than the forecasts previously described. Each simulation-minute required approximately 3.1 trillion floating point operations. To perform a forecast useful in real time, the simulation must run at least 20 times faster than wall-clock time. At this pace, a 5-day forecast requires 3-6 hours to run which implies a sustained performance of at least 1 teraflop. The performance of the AHW configuration with the large innermost domain appears in Fig. 3.

Blue Gene/P supports threads for shared memory parallelism between the CPUs on each node (unlike the earlier Blue Gene/L). We took advantage of this and used OpenMP within each node and MPI message passing between nodes to reduce the number of messages between nodes and to reduce per-node memory use. The Blue Gene CPUs also have SIMD units for boosting on-processor performance. Unfortunately, on the Blue Gene/P these work only for double precision (64-bit) floating point data. WRF can be run in double precision, but needs only single (32-bit) floating point precision. Additional message and memory traffic outweighed the benefits of the faster CPU performance; therefore, the SIMD units were unused for this Blue Gene hurricane simulation. Based on communication with ALCF and IBM, we expect that the next system in the series, Blue Gene/Q, will address this problem by supporting single precision SIMD processing. Model output used Parallel NetCDF[11].

The results of three simulations of hurricane Bill are shown in Fig. 4. One forecast was actually done in real-time using the same domain configuration as for the HRH test. The simulation with the large nests, but same outer domain produced about the same result as the real-time forecasts. However, the simulation with a larger outer domain produced a better intensity and position forecast (position not shown) than either of the first two. Further, we note that none of these forecasts produced the correct weakening of Bill on the 21st of August. While just a single case, it suggests that provided the core of the storm and intense precipitation are contained within a high-resolution nest, the size of that nest may not matter a great deal. However, the size of the outer domain could be important. Our relatively large errors for hurricane Fred also provide evidence of the importance of the size of the outer domain.

5. Summary

There are many challenges regarding the prediction of hurricanes. Forecasters rely heavily on guidance from dynamical prediction models at time ranges beyond a day or so. We have concentrated on forecasts of storm intensity because this is historically a challenging endeavor and one that computing power may help address. We have shown that there are advantages to reducing the grid spacing of prediction models to a few kilometers in the horizontal so that the representation of moist convection can be explicit in models rather than completely parameterized. In carefully controlled simulations wherein the only aspect varied was the horizontal grid spacing (and physical parameterizations consistent with this change in grid spacing), to full real-time comparisons with the operational model, increased resolution reduces the intensity error by about 8%. Because the refinement in resolution is localized to the inner core of the storm, these higher-resolution forecasts can be accomplished with relatively modest computational resources. Enlarging the area covered by high resolution may not be cost effective for improving forecasts compared to enlarging the outer domain.

We must always remain aware of the need to improve the accuracy of the prediction of storm position as well as intensity. We saw that simply increasing resolution does not appear to effect the accuracy of storm track. Furthermore, the comparison of HWRF and AHW indicates that resolving well the inner core is not essential for a model to produce a superior track forecast. The optimal forecast system is one that combines the locally fine resolution for improving the prediction of storm intensity with the skillful prediction on the large scales of motion for the benefit of track forecasts.

The ultimate goal is to push prediction skill toward what is allowed by the intrinsic limits of predictability of the atmosphere. However, limits of predictability are highly case dependent. That is why the best way forward is probably to integrate large ensembles of high-resolution forecasts so that the uncertainty can be directly estimated for any situation. This represents a push toward highly scalable large-computing applications that are already possible and should be pursued further.

Acknowledgements

The authors acknowledge funding from the Hurricane Forecast Improvement Project (HFIP) that supported portions of this work. The authors also acknowledge Sid Ghosh of NCAR’s Computing and Information Systems Laboratory for his assistance with the real-time forecasts.

References

[1] Haltiner, J. T., and R. T. Williams, 1980: *Numerical weather prediction and dynamic meteorology*. 2nd ed. John Wiley and Sons, 477 pp.

[2] Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019-3032.

[3] Rotunno, R., Y. Chen, W. Wang, C. Davis, J. Dudhia, and G. J. Holland, 2009: Large-eddy simulation of an idealized tropical cyclone. *Bull. Amer. Meteor. Soc*., **90**, 1783-1788.

[4] Davis, C., W. Wang, S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, H. Reeves, R. Rotunno, and Q. Xiao, 2008: Prediction of landfalling hurricanes with the advanced hurricane WRF model. *Mon Wea. Rev*., **136**, 1990-2005.

[5] Schubert, W. H., and J. J. Hack, 1982: Inertial stability and tropical cyclone development. *J. Atmos. Sci.*, **39**, 1687-1697.

[6] Anderson, J., T Hoar, K. Raeder, H. Liu, Hui, N. Collins, R. Torn, A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. Bull. Amer. Meteor. Soc., 90, 1283-1296.

[7] Torn, R. D., 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Mon Wea. Rev*., Submitted.

[8] Davis, C. A., W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. and Forecasting*. Submitted.

[9] Michalakes, J., J. Hacker, R. Loft, M. O. McCracken, A. Snavely, N.Wright, T. Spelce, B. Gorda, and B. Walkup. WRF Nature Run. *Supercomputing, 2007 ACM/IEEE conference.*

[10] Peter Johnsen, Cray Inc. Personal communication.

[11] Jianwei, L., L. Wei-keng, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale. Parallel netCDF: A High-Performance Scientific I/O Interface. *Supercomputing, 2003 ACM/IEEE Conference*.

Category of Intensity (v_{obs})	Mean Difference of Absolute Error (knots)	Number in Sample
TD (< 34 knots)	-1.0	61
TS ($34 \le v_{obs} < 64$)	3.0	161
1 ($64 \le v_{obs} < 83$)	3.0	69
2 ($83 \le v_{obs} < 96$)	3.3	35
3 ($96 \le v_{obs} < 113$)	0.4	43
4 ($113 \le v_{obs} < 135$)	-14.2	54
5 ($v_{obs} \ge 135$)	-15.6	48

Table 1. Mean difference of absolute error $\overline{|v_{HR} - v_{obs}| - |v_{LR} - v_{obs}|}$, where v is the maximum wind (knots), subscripts *HR* and *LR* refer to high-resolution and low-resolution, subscript *obs* refers to the best track data, and the overbar indicates a sample average. Negative values denote larger errors in the coarse-resolution forecasts. Categories 1-5 refer to the Saffir-Simpson scale (with wind ranges indicated).

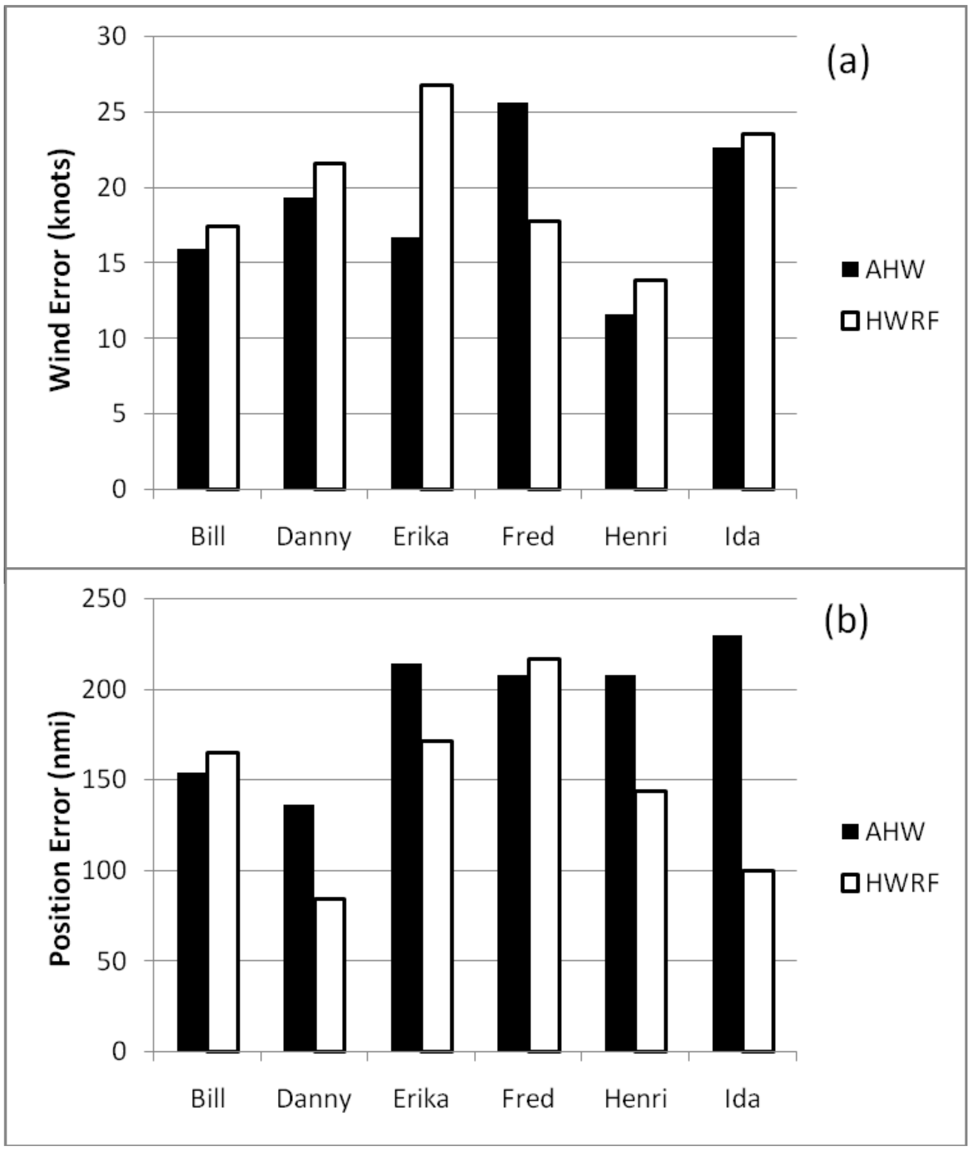

Figure 1. (a) Root-mean-squared errors for maximum wind for each of several storms during the 2009 Atlantic hurricane season (units are knots); (b) as in (a) but for root-mean-squared position errors (units are nautical miles, 60 nmi = 111.1 km = 1 degree of latitude)The number of six-hourly forecast valid times for each storm is; Bill: 255; Danny: 35; Erika: 78; Fred: 200; Henri: 61; and Ida: 120.

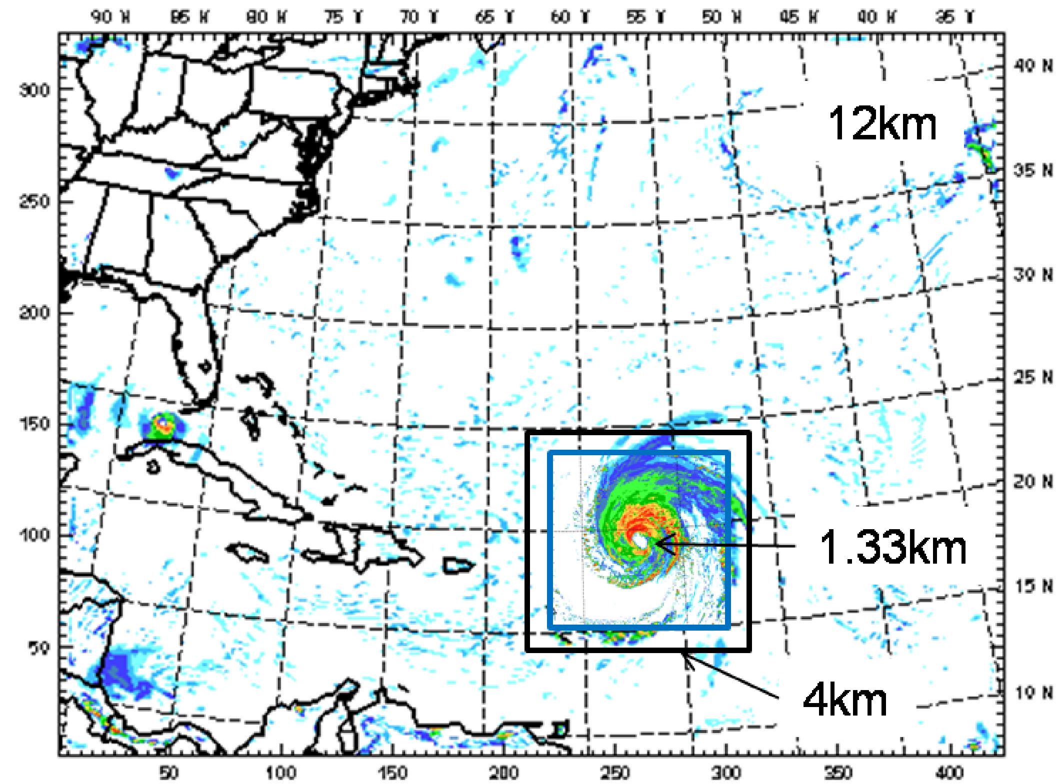

Figure 2. 3-Domain configuration using the large inner nests. The field is a pseudo-radar-reflectivity derived from the falling precipitation particles in the AHW, valid at an altitude about 40 m above the surface. Nest boundary of the domain with 4-km grid spacing is shown in black; nest boundary of the innermost nest is in blue.

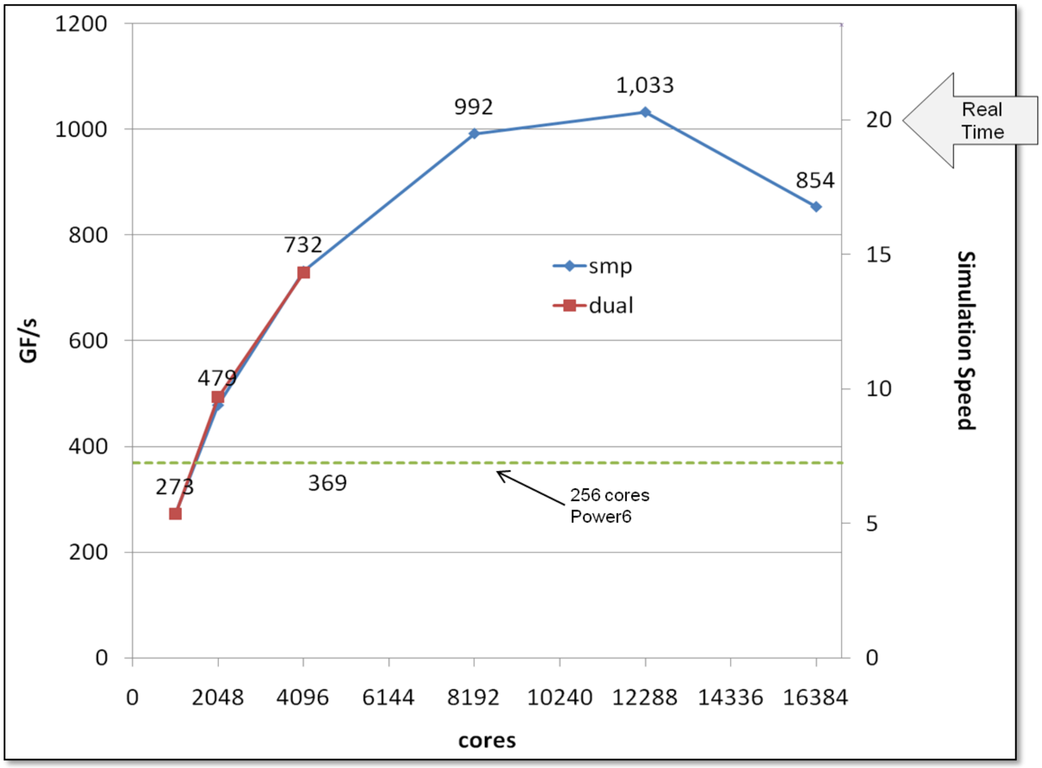

Figure 3. WRF performance for Hurricane Bill with increasing numbers of BG/P cpus (cores). The blue line is using a single MPI task with 4 OpenMP tasks per node; red is using two 2-way threaded tasks per node. The arrow indicates the minimum simulation rate needed for real-time hurricane forecasting. Performance on 256 cores of bluefire.ucar.edu, an IBM Power6 system, is shown for comparison.

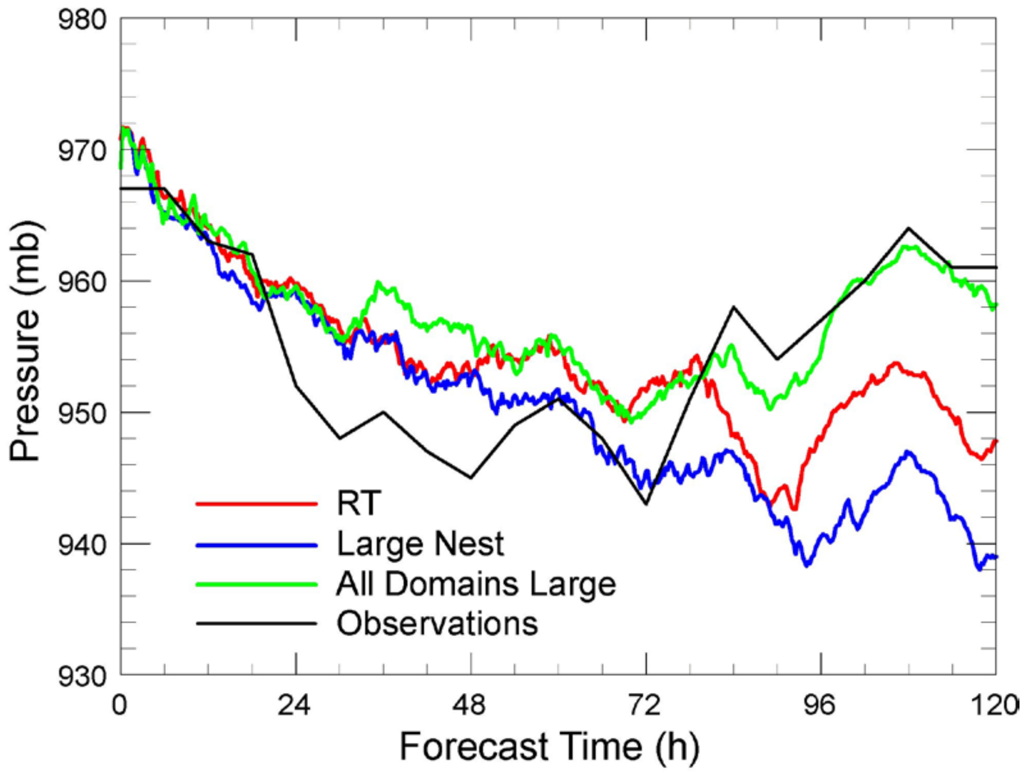

Figure 4. Time series of sea-level pressure (millibars, mb, 1 mb = 0.0295 inches of mercury) at the center of Bill. All forecasts were begun at 00 UTC 18 September, 2009. Red curve is for the real-time forecast ; blue line is the large-nest, small-coarse-domain simulation and green is the large-nest, large-domain simulation. Black is the observations from the National Hurricane Center derived mainly from satellite and hurricane reconnaissance aircraft data.