High-Resolution Hurricane Forecasts

Widely varying scales of atmospheric motion make it extremely difficult to predict hurricane intensity, even after decades of research. A new model capable of resolving a hurricane's deep convection motions was tested on a large sample of Atlantic tropical cyclones. Results show that using finer resolution can improve storm intensity predictions.

Predicting a hurricane's intensity remains a daunting challenge even after four decades of research. The intrinsic difficulties lie 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 1,000 kilometers or more.

Atmospheric dynamical models must account for all these scales simultaneously. Being a nonlinear system, these scales can interact. Although forecasters must make approximations to keep computations finite, there's a continued push for finer resolution to capture as many of these scales as possible.

Problem Overview

From the present standpoint of computational feasibility, a model's minimum grid lengths that still allow modeling of the storm and its near environment are a few hundred horizontal meters and approximately 50 to 100 vertical meters. Most current weather prediction uses grid-based rather than spectral-based models (such as Fourier or some other basis function).[1] Statistical analysis of energy spectra reveal that motions with scales smaller than approximately six to seven grid points aren't well resolved.[2] Therefore, the minimum resolvable physical length scales are nearly 1 km horizontally and perhaps 300 m vertically. Given current computing capability, however, timely numerical forecasts must be run on much coarser grids.

What does this mean for hurricane forecasts? We believe that it's 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 be resolved only with a 1 km or less horizontal grid spacing. These clouds span the troposphere's vertical extent—12 to 16 km—and so are relatively easy to resolve in the vertical if 30 to 40 layers are used. Nevertheless, to cover the region affected by a hurricane in a five-day forecast requires a horizontal domain of perhaps 5,000 km. A volume with a grid increment of 500 m horizontally and 250 m vertically over a domain of depth 25 km contains roughly 10^{10} grid points.

Because the equations of motion are first-order in time, knowing the model state at one time lets us, in principle, 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.

1521-9615/11/$26.00
Copublished by the IEEE CS and the AIP

Christopher A. Davis, Wei Wang, Steven Cavallo, James Michael Done, Jimy Dudhia, Sherrie Fredrick, John Michalakes, Ginger Caldwell, and Tom Engel
US National Center for Atmospheric Research
Ryan Torn
State University of New York, Albany

For grid lengths of approximately 1 km, the time step is roughly 10 seconds; it decreases to approximately 1 second on grids of 100 m. Given 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! Hence, we'd need more than 400,000 time steps to integrate a five-day forecast on a 100-m grid. Couple this with the need to integrate many variables (momentum, water in several forms, temperature, and so on) and you have a problem in which the time tendencies of variables must be estimated about 10^{16} to 10^{17} times for a single forecast. Furthermore, 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 we currently measure the capability of these machines 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 NOAA's high-resolution hurricane (HRH) test. The goal was to see whether forecasts on grids of roughly 1 to 2 km horizontally 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 the atmosphere's physical processes because we can't account for every turbulent eddy or precipitation particle. We can resolve some motions, but others must be represented implicitly in terms of resolved scale of motion. In the case of hurricanes, there can be a significant advantage in representing the deep cumulonimbus clouds explicitly rather than implicitly. The maximum grid spacing at which this explicit representation is possible is roughly 1 to 4 km; models with grids at 10 km nearly always implicitly represent the effects of deep convection. Clearly, the number of computations needed to explicitly represent thunderstorm clouds is much greater than for an implicit representation, but as we'll demonstrate, the finer resolution improves many aspects of hurricane forecasts.

We can mitigate part of the added cost by recognizing that it's probably more important to explicitly treat clouds near and within the hurricane's eyewall than clouds far from the storm. Because most of the deep convection occurs within 100 to 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 examine this question later in this article.

In selecting a grid spacing near or just coarser than 1 km, we choose 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 3D turbulence is resolved explicitly.

Such a result has a practical significance. It's questionable whether there's much to be gained by decreasing the horizontal grid spacing below 1 to 2 km unless we decrease it to approximately 100 m or less. This is perhaps a factor of 1,000 beyond current real-time computational capability. This fact supports our focus here on horizontal grid spacing of just over 1 km, and our comparisons of such grid spacing—which represents the forefront of hurricane forecasting—with results obtained at coarser resolution.

Enabling strategies to allow high resolution include simplifying the representation of the atmosphere's physical processes because we can't account for every turbulent eddy or precipitation particle.

AHW Model

For the HRH test, our group used the advanced hurricane-research weather research and forecasting model (AHW),[4] which was derived from the advanced research weather research and forecasting model (ARW). The AHW is a dynamic atmosphere model that predicts the atmospheric state's time evolution. We integrate predictive equations for the three Cartesian velocity components—entropy, mass, and numerous water phases (vapor, cloud droplets, rain, snow, and ice crystals)—using a discrete time-stepping technique.

The AHW uses grid nesting to locally enhance resolution, which can be achieved in many ways, each of which has pros and cons. Nesting requires blending the model state's update 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 principle, the AHW lets us use an arbitrary number of nests to achieve any desired horizontal resolution. The AHW currently uses only horizontal nesting.

During AHW hurricane forecasts, the nest is repositioned every 15 minutes to recenter 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 (which defines the center of the hurricane). In practice, a storm will progress only a few kilometers between nest movements, so the movements constitute a tiny fraction of the domain size.

It's possible to 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 isn't a simple process.

We represent precipitation, turbulence, radiation, and atmosphere–ocean coupling in AHW using a variety of parameterizations that are fairly simple in their conceptual design. Here, we focus on representing clouds and precipitation because the release of latent heat as water vapor, drawn from the ocean under the storm, condenses within clouds and is the essential fuel for the storm. When a model's grid spacing is much larger than a cloud ("cloud" here refers to a typical cumulonimbus), the only practical strategy is to represent the clouds' net effects in terms of scales of motion that the model can explicitly predict. In a sense, this amounts to prescribing the occurrence of a subgrid-scale thunderstorm when certain conditions are met. The net heating of the air from water condensation is thus prescribed, and the 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.

It's possible to 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 isn't a simple process. The problem is that condensation heating is highly inefficient for producing winds unless you already have a full-fledged hurricane; heating can create buoyancy oscillations, which essentially carry away the storm center's heat.[5] The explicit treatment of thunderstorms produces the cloud and predicts its growth, time step by time step, with nothing predetermined. The motivation in doing so is to achieve a more realistic partitioning of the heat that goes into enhancing the storm's circulation versus heat that is propagated away by buoyancy waves.

A Systematic Test of High-Resolution

The HRH test involved six research groups doing essentially the same task but with different models run in different configurations. The goal was to test the resolution sensitivity of forecast accuracy for hurricane intensity and the amount of computational power needed.

The test involved conducting two sets of 69 simulations covering 10 Atlantic tropical cyclones,[1] each using different horizontal resolution. We then evaluated whether, and by how much, higher resolution improved forecasts. It might seem that higher resolution should always result in superior forecasts, but that isn't actually the case. As you resolve 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. Although other metrics are being considered, the US National Hurricane Center's standard is the root-mean-squared error, which is what we use here.

How We Use Our Model

We initialized the model 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 atmosphere state estimates to decide how to spread information contained in observations.

The *filter is run in a cycle*—that is, a six-hour forecast initialized from the previous analysis time—provides the background upon which observations are analyzed. Typically, the filter cycles for two or more days before we start our first forecast for a given storm. To do real-time forecasts, the filter runs continuously from the hurricane season's start so that an up-to-date analysis is always available that incorporates all past and present observations. For the HRH test, we initiated the filter two days before our first forecast. Assimilated observations included surface pressure, atmospheric soundings (including dropsonde observations from NOAA's Gulf-Stream IV aircraft), data from commercial aircraft, cloud motion vectors from satellite observations, and tropical cyclone position and intensity estimates from the National Hurricane Center.

We initialize the high-resolution forecast from the ensemble member that was closest to the observed storm 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 hours or until the time the observed storm dissipated. The domains' dimensions were 469 × 424 (12 km), 201 × 201 (a 4-km nest), and 241 × 241 (a 1.33-km nest). Because the time step decreased in proportion to the grid spacing, nine time steps of the innermost nest are required for each time step of the coarsest domain. Adding the two moving nests increased 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 hours, while 35 forecasts were integrated the full 126 hours. Initial conditions were identical for each member of a pair.

Forecasts were integrated on the US National Center for Atmospheric Research's bluefire supercomputer, an IBM Power 575 cluster commissioned into service on 30 June 2008 and consisting of 128 Power 575 nodes, each containing 32 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 77 trillion floating-point operations per second (Tflops). The AHW model used five Power 575 nodes, or 160 Power6 processors, and all runs for the HRH test used approximately 115,000 processor hours.

Although these runs didn't necessarily tax the bluefire system's computational capability, they're 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 wouldn't have considered at the time. It's thus readily conceivable that five years hence, we'll be able to integrate the entire ensemble forward at high resolution rather than selecting a single member for a deterministic forecast.

Our Results

The resolution test's overarching finding is that increasing resolution improves forecasts of both hurricane intensity and some of hurricanes' structural aspects; we've demonstrated the results to be statistically significant.[8] The improvement is about 8 percent for the root-mean-square intensity error. As we noted earlier, the inclusion of the nests increases the computations by a factor of four. An 8 percent improvement might 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 intensity bias 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 ms^{-1}). The coarse-resolution forecasts were biased low by 5 to 10 knots. This low bias of intensity got progressively worse as horizontal resolution decreased. The primary reason is that the hurricane's inner core—where the strong winds are—becomes poorly resolved as the grid spacing becomes as large as the core. Recall that six to seven grid points are necessary to truly resolve the kinetic energy at a given spatial scale. It would indeed be a large hurricane that's well resolved on a grid as coarse as, say, 36 km.

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

An intensity forecast's primary shortcoming is in predicting rapid intensity changes. Here, we

Table 1. Mean difference of absolute error.*		
Saffir-Simpson intensity category (v_{obs})	Mean difference of absolute error (knots)	Number in sample
Tropical depression (< 34 knots)	–1.0	61
Tropical storm ($34 \leq v_{obs} < 64$)	3.0	161
1 ($64 \leq v_{obs} < 83$)	3.0	69
2 ($83 \leq v_{obs} < 96$)	3.3	35
3 ($96 \leq v_{obs} < 113$)	0.4	43
4 ($113 \leq v_{obs} < 135$)	–14.2	54
5 ($v_{obs} \geq 135$)	–15.6	48

Absolute error is $\overline{|v_{HR} - v_{obs}| - |v_{LR} - v_{obs}|}$, where v is the maximum wind (knots); HR and LR are high-resolution and low-resolution, respectively; obs refers to the best track data; and the over bar indicates a sample average. Negative values denote larger errors in the coarse-resolution forecasts.

adopt the definition of "rapid" as an increase of the maximum sustained wind by 25 knots or more in 24 hours. To define skill, we compute the equitable threat score (ETS) as

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

where

- a is the number of correct forecasts of rapid intensity changes (hits),
- b is the number of forecasts of rapid intensity changes that didn't occur (false alarms),
- c is the number of times rapid intensity changes that occurred but weren't predicted (misses), and
- d is the number of correct null forecasts (correct forecasts of no rapid intensification), and
- ε is the number of hits expected due to random guessing.

The ETS was 0.16 for the high-resolution AHW forecasts, 0.11 for the coarse-resolution AHW forecasts, and 0.04 for the human-generated forecasts. Human-generated forecasts tend to be conservative about intensity changes; this is a conscious decision that also reflects the fact that it's 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 for this is that although a storm's environment often sets the conditions for rapid intensification, the vortex response to this environment is more realistic at higher resolution. This is probably because there's better inner-core resolution, as well as more realistic treatment of thunderstorms. We can thus think of the high-resolution model as more agile.

Adding high resolution didn't change storm location prediction in any statistically significant way. This result isn't surprising, and it echoes other researchers' findings. A hurricane's track can be well predicted even in global weather prediction models that don't resolve the hurricane's eye and associated inner core. Typically, this is because a large component of track prediction is effectively the "steering" of the vortex by the wind averaged over the vortex's depth. This averaged flow typically represents the location and strength of weather systems with length scales of 1,000 km or more. Such scales are well resolved in a global model of even modest resolution.

Real-Time Hurricane Forecasts

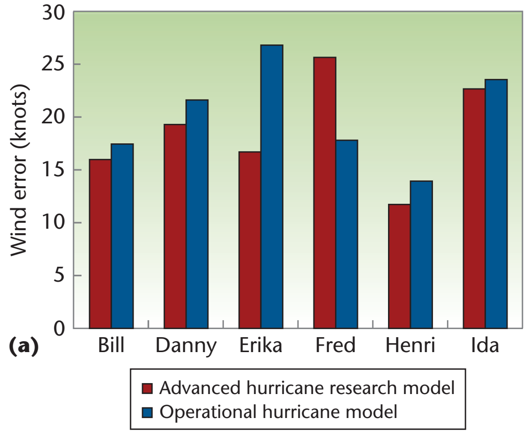
We also applied the AHW forecast model to real-time hurricane prediction in 2009, with minor changes to the forecasting system. The primary change was to use a nested grid spacing of 12 km in the ensemble forecasts to better capture stronger storms. We compared 50 forecasts from AHW with operational forecasts from the US National Centers for Environmental Prediction's Hurricane Weather Research and Forecasting model (HWRF).

HWRF's numerical methods and physical parameterizations are formulated differently than those in AHW, and HWRF integrated these forecasts on a coarser grid spacing (approximately 9 km) compared to AHW's inner nest grid spacing of 1.33 km. Given this, it wasn't a controlled comparison. The point was to demonstrate the possibility of improved capabilities for high-resolution forecast models. Also, the way the two systems treat convection is a key difference: it's mainly implicit in HWRF and explicit in AHW. Finally, as the resolution difference imply, HWRF's computing requirements are several times smaller than those 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. Some of the major differences in model performance were storm dependent: AHW forecasts were better for Bill, Danny, Erika, and Henri, while HWRF offered superior forecasts of Hurricane Fred. Our forecast sample is homogeneous; we compared only HWRF and AHW forecasts with the same valid time and lead time. However, we didn't perform verification after the observed storm's dissipation, even when a viable storm still existed in the forecast.

Danny, Erika, and Henri were weak storms embedded in hostile environmental conditions. The primary factor inhibiting their development was the increase of the horizontal wind with height—or *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 to 100 km in a tilted storm. Tilted storms tend to produce their convection displaced from the surface circulation center; this is an inefficient configuration for intensification. Hurricanes usually intensify when a ring of convection envelops 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 it usually refers to a systematic wind increase of at least 2 to 3 ms^{-1} for every vertical kilometer. From the boundary layer to the middle troposphere (5 to 6 km above ground level, or AGL) the corresponding wind increase would be 10 to 15 ms^{-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. HWRF intensified and weakened Fred at nearly the right time, whereas AHW produced realistic intensification and weakening of Fred, but was consistently late by approximately one day. This created large errors for the AHW forecast. The problem might have been the storm's proximity 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, which AHW will inherit.



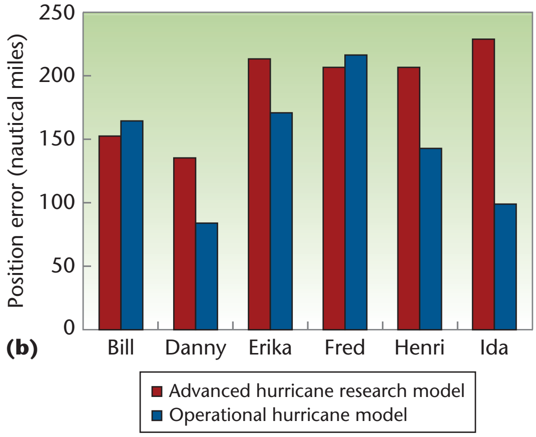


Figure 1. A comparison of two weather research and forecasting (WRF) models: our advanced hurricane-research WRF model (AHW) and the US National Centers for Environmental Prediction's hurricane WRF (HWRF) model. (a) Root-mean-squared errors for maximum wind for each of several storms during the 2009 Atlantic hurricane season. (b) The same comparison, but with root-mean-squared position errors (units are nautical miles, 60 nmi = 111.1 kilometers = 1 degree of latitude). The number of six hourly forecast valid times for each storm is as follows: Bill, 255; Danny, 35; Erika, 78; Fred, 200; Henri, 61; and Ida, 120.

Position errors (Figure 1b) indicate that, overall, the HWRF forecasts produced smaller position errors, but this was primarily true 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 approximately one day, whereas the real track was almost due northward.

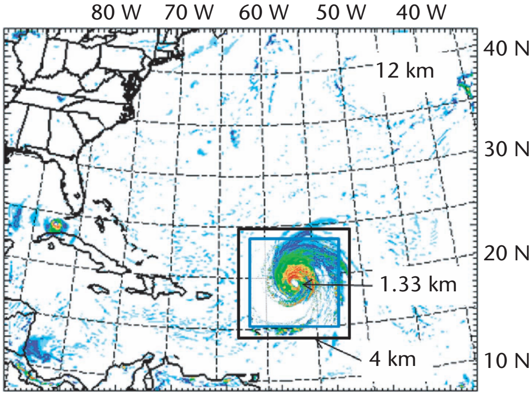


Figure 2. A three-domain configuration using the large inner nests. The field is a pseudo-radar-reflectivity derived from the falling precipitation particles in the Advanced Hurricane-Research Weather Research and Forecasting model (AHW), valid at an altitude approximately 40 meters above the surface. The domain's nest boundary with a 4-km grid spacing is shown in black; the innermost nest's boundary is in blue.

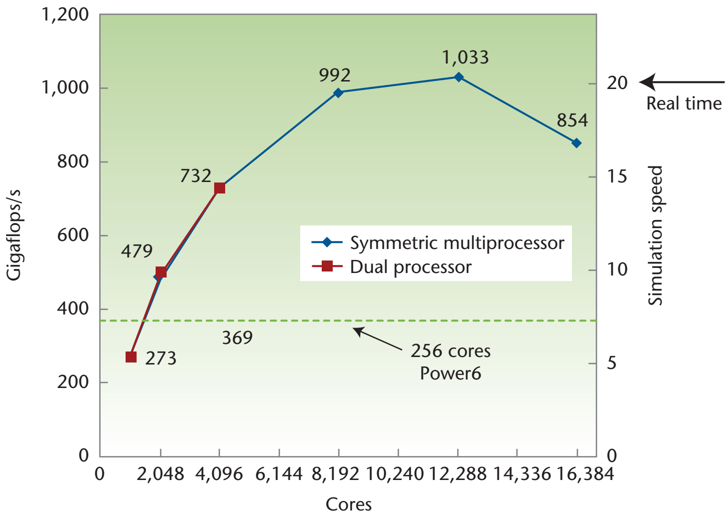


Figure 3. Weather research and forecasting (WRF) performance for Hurricane Bill with increasing numbers of Blue Gene/P CPU cores. The blue line indicates a single MPI task with four OpenMP tasks per node; red is using two two-way threaded tasks per node. The arrow indicates the minimum simulation rate needed for real-time hurricane forecasting. For comparison, we also show the performance on 256 cores of bluefire.ucar.edu, an IBM Power6 system.

Overall, the benefits of AHW's additional resolution, 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 indicate the need for accurate forecasts on scales ranging from much larger than the storm (1,000 km or more) down to the eyewall's scale (10 km or less) to address the hurricane forecasts' position and intensity aspects.

Toward Petascale Hurricane Forecasts

Thus far, the results we've presented have been obtained with limited computational capability. Next-generation high-performance computing (HPC) systems comprising tens of thousands of CPUs could allow real-time forecast domains that cover an entire storm at resolutions of 1 km or finer, and even allow integration of ensembles of such configurations.

To test forecasting as well as computational issues in advance of this capability, we've run the Hurricane Bill case on 16,000 processors (4,000 nodes) of the IBM Blue Gene/P the Argonne Leadership Computing Facility (ALCF). Previously, WRF sustained 50 teraflops on 150,000 Cray XT5 MPI tasks for a two-billion-cell idealized "nature run" simulation.[9]

In our simulation, we used a somewhat larger coarse 12-km grid, but much larger nests. We expanded the innermost nest (1.33-km grid spacing) to 750 × 750 points and moved with the storm. This nest covered a domain that was 1,000-km by 1,000-km square—enough to cover all but the outermost tips of the storm's rain bands (see Figure 2). Compared to the nature run, this is much smaller (24 million cells), but it's roughly 10 times larger than the forecasts we've 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 five-day forecast requires 3 to 6 hours to run, which implies a sustained performance of at least 1 Tflop. Figure 3 shows the AHW configuration's performance with the large innermost domain.

Unlike the earlier Blue Gene/L, Blue Gene/P supports threads for shared-memory parallelism between the CPUs on each node. To take advantage of this, we 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/P also has single instruction, multiple data (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 faster CPU performance's benefits; therefore, we didn't use the SIMD units 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. For model output, we used Parallel NetCDF.[10]

Figure 4 shows the results of three Hurricane Bill simulations. 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 than either of the first two (the position forecast isn't shown here). Furthermore, none of these forecasts produced the correct weakening of Bill on 21 August.

Although just a single case, this suggests that if the storm's core and intense precipitation are contained within a high-resolution nest, the size of that nest might not matter much. However, the outer domain's size could be important. Our relatively large errors for Hurricane Fred further support the importance of the outer domain's size.

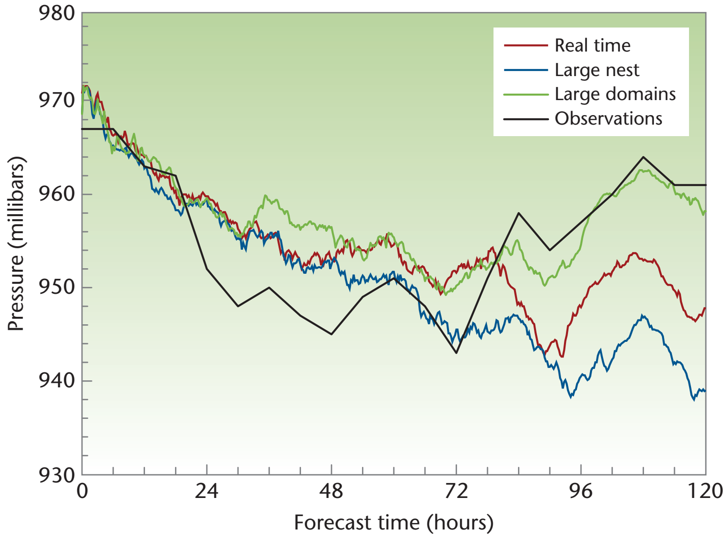


Figure 4. Time series of sea-level pressure at Hurricane Bill's center (1 millibar = 0.0295 inches of mercury). All forecasts started at 00 UTC 18 September 2009. The red curve shows the real-time forecast; the blue line shows the large-nest, small-coarse-domain simulation; the green line shows large-nest, large-domain simulation; and the black line shows observations from the US National Hurricane Center, derived mainly from satellite and hurricane reconnaissance aircraft data.

Hurricane prediction has many challenges. Forecasters rely heavily on guidance from dynamical prediction models at time ranges beyond a day or more. We concentrated on storm intensity forecasting because it's been a challenging endeavor historically, and it's one that computing power might help us address.

As our results show, there are advantages to reducing prediction model grid spacing to a few horizontal kilometers so that the moist convection representation can be explicit in models rather than completely parameterized. In carefully controlled simulations where only the horizontal grid spacing (and physical parameterizations consistent with this change in grid spacing) varied, the increased resolution reduced intensity error by approximately 8 percent. Because we localized resolution refinement to the storm's inner core, we can accomplish these higher-resolution forecasts with relatively modest computational resources. Enlarging the area covered by high resolution might not be cost effective for improving forecasts compared to enlarging the outer domain.

We must remain continuously aware of the need to improve predictions of storm position accuracy as well as intensity. As our results show, simply increasing resolution doesn't appear to effect storm track accuracy. Furthermore, the comparison of HWRF and AHW indicates that resolving the inner core well isn't essential for a model to produce a superior track forecast. The optimal forecast system is one that combines the locally fine resolution to improve intensity prediction with skillful prediction on the large scales of motion to benefit track forecasts.

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

Acknowledgments

Our work was partially funded by the Hurricane Forecast Improvement Project. We thank Sid Ghosh of the US National Center for Atmospheric Research's Computing and Information Systems Laboratory for his assistance with the real-time forecasts. The National Center for Atmospheric Research is sponsored by the US National Science Foundation.

References

1. J.T. Haltiner and R.T. Williams, *Numerical Weather Prediction and Dynamic Meteorology*, 2nd ed., John Wiley and Sons, 1980.
2. W.C. Skamarock, "Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra," *Monthly Weather Rev.*, vol. 132, no. 12, 2004, pp. 3019–3032.

3. R. Rotunno et al., "Large-Eddy Simulation of an Idealized Tropical Cyclone," *Bulletin Am. Meteorological Soc.*, vol. 90, no. 12, 2009, pp. 1783–1788.
4. C. Davis et al., "Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model," *Monthly Weather Rev.*, vol. 136, no. 6, 2008, pp. 1990–2005.
5. W.H. Schubert and J.J. Hack, "Inertial Stability and Tropical Cyclone Development," *J. Atmospheric Science*, vol. 39, no. 8, 1982, pp. 1687–1697.
6. J. Anderson et al., "The Data Assimilation Research Testbed: A Community Facility," *Bulletin Am. Meteorological Soc.*, vol. 90, no. 9, 2009, pp. 1283–1296.
7. R.D. Torn, "Performance of a Mesoscale Ensemble Kalman Filter (EnKF) During the NOAA High-Resolution Hurricane Test," *Monthly Weather Rev.*, preprint, 23 Aug. 2010; doi:10.1175/2010MWR3361.1.
8. C.A. Davis et al., "Does Increased Horizontal Resolution Improve Hurricane Wind Forecasts?" *Weather and Forecasting*, preprint, 24 Aug. 2010; doi:10.1175/2010WAF2222423.1.
9. J. Michalakes et al., "WRF Nature Run," *J. Physics: Conf. Series*, vol. 125, no. 1, 2008; doi:10.1088/1742-6596/125/1/012022.
10. L. Jianwei et al., "Parallel netCDF: A High-Performance Scientific I/O Interface," *Proc. Int'l Conf. Supercomputing*, IEEE Press, 2003; doi:10.1109/SC.2003.10053.

Christopher A. Davis *is a senior scientist at the US National Center for Atmospheric Research. His research interests include hurricanes, convection, numerical weather prediction, and forecast verification. Davis has a PhD in meteorology from the Massachusetts Institute of Technology. He is a member of the American Meteorological Society and the American Geophysical Union. Contact him at cdavis@ucar.edu.*

Wei Wang *is an associate scientist at the US National Center for Atmospheric Research. Her research interests include hurricanes, convection, extreme weather, and numerical modeling. Wang has a PhD in meteorology from The Pennsylvania State University. Contact her at weiwang@ucar.edu.*

Steven Cavallo *is a postdoctoral fellow at the US National Center for Atmospheric Research. His research interests include arctic dynamics, numerical modeling and development, tropical meteorology, Vortex dynamics, mesoscale numerical weather prediction, and data assimilation. Cavallo has a PhD in meteorology from the University of Washington. He is a member of the American Meteorological Society and the American Geophysical Union. Contact him at cavallo@ucar.edu.*

James Michael Done *is a project scientist at the US National Center for Atmospheric Research. His research interests include tropical cyclogenesis and easterly waves as well as the effects of climate variability and change on tropical cyclones. Done has a PhD in meteorology from the University of Reading, UK. He is a member of the American Geophysical Union and Royal Meteorological Society. Contact him at done@ucar.edu.*

Jimy Dudhia *is a project scientist at the US National Center for Atmospheric Research. His research interests include physical parameterizations, weather prediction, and numerical weather prediction. Dudhia has a PhD in meteorology from Imperial College, London. He is a fellow of the Royal Meteorological Society and a member of the American Meteorological Society. Contact him at dudhia@ucar.edu.*

Sherrie Fredrick *is a software engineer at the US National Center for Atmospheric Research. Fredrick has a BS in physics from University of Northern Colorado and a BS in computer science from Colorado University (Denver). Contact her at sherrie@ucar.edu.*

John Michalakes *is a software engineer at the US National Center for Atmospheric Research. His research interests include high-performance computing. He has an MS in computer science from Kent State University. Contact him at michalak@ucar.edu.*

Ginger Caldwell *works in allocations management at the US National Center for Atmospheric Research. Her research interests include metrics, statistical measures of quality, metrics, computational resource allocation, and measuring computational resource use. She is a member of the ACM. Contact her at cal@ucar.edu.*

Tom Engel *is a high-performance computing specialist the US National Center for Atmospheric Research. His research interests include high-performance computing architectures, operating systems, performance analysis, and tuning. Engel has an MS in astrogeophysics from the University of Colorado. He is a member of American Meteorological Society, IEEE, and the ACM. Contact him at engel@ucar.edu.*

Ryan Torn *is an assistant professor from the State University of New York, Albany. His research interests are in atmospheric sciences, including numerical weather prediction, data assimilation, predictability, and tropical cyclones. Torn has a PhD in meteorology from the University of Washington. He is a member of the American Meteorological Society. Contact him at torn@atmos.albany.edu.*