Increasing NOAA's computational capacity to improve global forecast modeling

A NOAA White Paper

Thomas M. Hamill and Jeffrey S. Whitaker
NOAA Earth System Research Lab, Physical Sciences Division

Michael Fiorino and Steven E. Koch
NOAA Earth System Research Lab, Global Systems Division

Stephen J. Lord
Director, NWS NCEP Environmental Modeling Center

19 July 2010

Corresponding author address

Dr. Thomas M. Hamill
NOAA/ESRL
R/PSD1, 325 Broadway
Boulder, CO 80305-3328
(303) 497-3060
tom.hamill@noaa.gov

Executive Summary

The accuracy of many of the NWS's forecast products, even its regional forecast products, are constrained by the limitations of NOAA's global forecast model. Unfortunately, our global forecasts are less accurate than those from competing numerical weather prediction centers such as the European Center for Medium-Range Weather Forecasts (ECMWF). For most sensible-weather metrics, we lag 1 to 1.5 days (i.e., they make a 3.5-day forecasts as skillful as our 2-day forecast).

Dramatically increased computational resources devoted to global modeling are necessary but not sufficient for NOAA to improve its global forecast modeling capacity. The increased accuracy from a threefold resolution increase can be expected to be roughly comparable to the impact of using an advanced data assimilation technique, e.g., 4D-Var instead of 3D-Var. Depending on the method of evaluation, a threefold increase resolution and the use of advanced data assimilation technique may provide an additional several hours to several days of additional forecast lead time, i.e., with both we may be able to make a 3-day forecast as well as we previously made a 2-day forecast. These improvements are much larger than the improvement that can be expected by dedicating the same amount of resources to augmenting the observing system.

In addition to augmented HPC to support higher-resolution models and advanced data assimilation, there are several other advances that may be able to contribute significantly to improved global forecasts, and these can be facilitated by advanced HPC. These include the development of improved model physical parameterizations, the coupling of forecast state components, the use of ensemble prediction methods, and the development and use of reforecast data sets to correct systematic forecast model deficiencies. Higher-resolution systems may also be able to be deployed with the same resources devoted to HPC through code optimization and processor changes (graphical processing units, or GPUs) instead of, or in addition to central processing units (CPUs).

The improvement in forecasts from adopting a next-generation global modeling system can be demonstrated in 2011 on Oak Ridge National Lab (ORNL) computers provided through stimulus funds. Assuming the demonstration of large performance increases at ORNL, we propose as a target for NOAA to procure the computational resources to allow it to deploy a system comparable to ECMWF's by 2014, a global ensemble with ~15-km grid spacing. Without adoption of GPUs, we estimate approximately 150,000 CPU cores will be needed (two thirds dedicated to operations and parallel testing, one-third to research). With the use of GPUs, the computational capacity may be able to be enlarged by an order of magnitude or more, permitting an additional doubling of model resolution.

Such an investment will represent a tiny fraction of NOAA's current investment in weather satellites. Without a modern data assimilation system and high-quality global forecast model, NOAA's investment in its satellites cannot be fully realized.

1. **Introduction**

Global atmospheric forecast models are the backbone of NOAA's weather predictions. While regional models may provide detail over the US or over hurricanes, the fidelity of this detail is highly constrained by the accuracy of the global forecast model data, which feeds the regional models their lateral boundary conditions. A developing cyclone over Japan can change the weather patterns over North America within two days (Fig. 1), so regional models sometimes produce very poor forecasts unless they are provided with high-quality global forecast data on their boundaries (Fig. 2). *An improved global modeling system is the tide that can lift all of NOAA's weather forecast boats.*

In many important measures, NOAA's global forecast model guidance and its ensemble guidance are much less skillful than those from the European Center for Medium-Range Weather Forecasts (ECMWF; Fig. 3, Hamill et al. 2010), and in other measures NCEP's global forecast guidance is only competitive with guidance from other countries with much smaller populations and world responsibilities (e.g., Fig. 4; Hagedorn et al. 2010). While there are some subtle reasons why we are behind, there are several obvious reasons, too. Our global deterministic model in early 2010 was *less than one-third* the horizontal resolution of ECMWF's (their T1279[1], or ~13.5 km grid spacing at 30° N vs. our T382, or ~45 km). NOAA has not always lagged ECMWF. In 1991, NOAA briefly ran a T126 deterministic forecast model while ECMWF ran a T106 model. Since then, ECMWF has systematically improved their global system (Fig. 5). One major reason why ECMWF's forecasts have improved more quickly than ours has been their more aggressive pace of increasing the model resolution. They increased their deterministic model resolution in 1991 (from T106 to T213); in 1998 (to T319); in 2003 (to T511); in 2006 (to T799) and in 2010 (to T1279). In contrast, from 1991 to 2010, NCEP's Global Forecast System (GFS) model has increased resolution from T126 to T382.

Another area where other forecast centers have made more rapid progress is in the data assimilation techniques used to initialize the forecasts. Every other major numerical weather prediction center, including ECMWF, the UK Met Office, the Canadian Meteorological Centre, Météo France, and the Japanese Meteorological Agency run a four-dimensional variational ("4D-Var") data assimilation system,

---

[1] The T### denotes the global spectral wavenumber at which the forecasts are truncated. For example, T300 indicates that the model is able to resolve a wave as small as 1/300th of the circumference of the earth. Hence, the higher the T-number, the finer the model resolution. The spacing of the grids used for processes such as radiation and parameterizations differs between NCEP and ECMWF, even at the same wavenumber truncation. NCEP uses a higher-resolution "quadratic" Gaussian grid that is necessary to prevent aliasing of quadratic terms with their Eulerian time-stepping scheme. ECMWF, which uses semi-Lagrangian time-stepping scheme that is not subject to aliasing, is able to use a coarser grid. See http://tinyurl.com/242cng7 and www.rclace.eu/File/EUMETNET/raport.pdf.

while NOAA runs a three-dimensional system ("3D-Var"). ECMWF has had an operational 4D-Var since 1997. 4D-Var and the comparable ensemble Kalman filter (EnKF), also in development within NOAA, require 10 to 100 times the computational resources of 3D-Var.

There are other reasons for the large differences in skill between ECMWF and NCEP shown in Figs. 3 and 4. For example, ECMWF assimilates a somewhat wider variety and greater number of observations, in part due to waiting for more data before beginning their assimilation process. They have more staff dedicated to improving the physical parameterizations in their forecast model, the code that describes how the land interacts with the atmosphere, or how an unresolvable, small-scale thunderstorm's effects on the larger-scale weather patterns may be estimated; and they compute in real time a ~20 year, 5-member ensemble of "reforecasts" with their operational medium-range ensemble prediction system to aid in the detection and correction of systematic errors in their forecasts. Their forecast model is also more computationally efficient; by using a "semi-Lagrangian" advection scheme, they can use much longer time steps in their model and hence produce a forecast at the same resolution as ours at a reduced computational expense.

Let's assume that NOAA procures significantly augmented HPC. What will this provide to NOAA and its customers? HPC alone will not permit us to suddenly eliminate the gap with ECMWF, but it can have a substantial impact. In this white paper we attempt to quantify the improvements that can be expected from increasing the resolution, from assimilating more observations, from changing the data assimilation, and so on. We will use standard meteorological metrics, such as improved skill or decreased RMS error; we make no attempt here to quantify the number of lives saved or economic disruptions avoided. Implicitly we are assuming that you, the reader, will be able to see that if the sum of the improvements results in the ability to make a 3-day forecast as well as we used to make a 2-day forecast, this should have major positive consequences (think starting the evacuation of New Orleans 3 days prior to Katrina instead of 2, or avoiding the unnecessary evacuation of Houston for hurricane in advance of Rita).

Below, section 2 attempts to quantify these improvements. We will draw numerical weather prediction experiments from across the US and the globe in order to show that the conclusions about the impact of data assimilation and resolution are not model-specific. Section 3 briefly discusses some other factors that may affect the future performance of forecast models, and section 4 proposes a path forward, a plan for testing an advanced global modeling system and then what NOAA should procure and what R&D NOAA should do to deploy a system competitive with ECMWF in 2014.

2. The impacts of improved modeling techniques provided by increased HPC.

There are many possible ways to improve numerical forecast guidance. These include assimilating new observations; improving the data assimilation technique; refining the forecast model resolution; improving the "physics" in forecast models; adopting coupled model approaches; using ensemble prediction techniques; and conducting "reforecasts" to facilitate statistical post-processing. To varying extents, each of these requires augmented high-performance computing.

We consider first the potential impact of new observations. Figure 6 provides some evidence suggesting that assimilating additional observations may not be the most cost-effective way to improve forecast guidance. ECMWF has run two reanalyses[2] during the last decade, "ERA-40" using 3D-Var and a T159 forecast model, and "ERA-Interim" with 4D-Var and a T255 forecast model. Each assimilated, effectively, the operational observed data stream available over the reanalysis periods. ERA-Interim's modest increase in resolution, its improved forecast model, and especially the use of 4D-Var led to a modest increase in forecast skill of subsequent forecasts, approximately one-half day extra lead in forecasts relative to ERA-40. That is, a day-5 ERA-Interim forecast was about as skillful in anomaly correlation as a day-4.5 forecast from ERA-40. Note, however, that the skill of the ERA-Interim forecasts did not change appreciably over the nearly two decades, suggesting that the additional remotely sensed observations that were added over those decades had a secondary effect. This suggests that *an incremental dollar spent on improving numerical weather prediction and assimilation methods is likely to provide a much greater beneficial effect than a dollar spent on new observations.* There may be some exceptions to this, such as observations that provide detail on hurricanes.

How much improvement can be attributed to the advanced data assimilation, how much to the forecast model, and how much to other effects such as improving model physics? Figure 7 shows results of experiments conducted at the Canadian Meteorological Centre using a metric that emphasizes the performance of predicting the mid-latitude jet stream. Increasing the resolution of the forecast model threefold had about the equivalent effect of changing the data assimilation from 3D- to 4D-Var. This is similar to the additional computational expense incurred for each. The increased resolution and 4D-Var had a synergistic effect, improving the forecast further when both were combined. In these experiments, the only change to the forecast model was the threefold increase in resolution. In practice, NWP centers find that the effects of increased resolution can be magnified if the forecast model is "tuned" to the new resolution, and improvement is commonly greater when examining other model aspects such as surface temperature or precipitation due in

[2] A reanalysis is a multi-decadal, retrospective analysis of the atmospheric state using a fixed forecast model and data assimilation system; estimates of the atmospheric state are provided twice or four times daily each day over the decades.

part to improved representation of terrain features. The improvements in Fig. 7 thus represent a lower bound.

NOAA has conducted many of its own experiments concerning the effect of model resolution and data assimilation. During 2009, NOAA examined the effects on hurricane tracks from using higher-resolution ensemble prediction system (T382 vs. the operational T126) and an advanced data assimilation method, the EnKF.[3] Figure 8 shows a more substantial positive impact from the increase in resolution and the data assimilation upgrade. The new EnKF-initialized T382 ensemble forecast provided greatly improved forecasts relative to the NCEP operational system, and the experimental forecasts were competitive with those from the ECMWF operational system. However, in metrics similar to those used in Fig. 7, the new data assimilation system and increased resolution made up only about one-half the difference between the NCEP and ECMWF operational systems (Table 1).

NCEP/EMC has made many improvements to its GFS forecast over the past decade, and the relationship of these changes to changes in forecast skill can also provide some evidence about the impact of observations, resolution, assimilation, and other effects. Table 2 lists the changes to the NCEP GFS from 1999 -2009. Figure 9 shows the percent of annual mean 500 hPa anomaly correlation (AC) scores in the Northern and Southern Hemispheres (NH, SH) that are judged "poor," defined as those having an AC score below 0.7 for each year. After 1998, there is a steadily decreasing fraction of poor forecasts in both hemispheres, strong evidence that the overall GFS forecasts are improving. When horizontal resolution increases were implemented in 2000, 2002 and 2005 (1a, 1b, 1c), the fractions of poor forecasts in the NH decreased noticeably thereafter. These improvements were not as noticeable in the SH. Changes in model physics (2a) improved SH and tropical scores (not shown), but had little apparent impact in the NH. Adding AMSU-A in 1999 improved forecasts in both hemispheres but other changes to the data assimilation and observations (3, 4) improved scores when combined with other changes, primarily resolution. This is particularly true over 2007-2009, where the introduction of the GSI 3D-Var data assimilation scheme (Kleist et al. 2009) and COSMIC radio-occultation data (Anthes et al. 2008) appear to have made noticeable improvements.

Physically, what are the mechanisms for the improvement of forecasts with extra resolution? One major aspect is the ability to define the geography more precisely. Figure 10 shows the approximate terrain height for the operational ECMWF and NCEP global ensemble forecasts over North America. Features like California's Coast Range largely disappear with the coarser NCEP terrain. Another problem with coarse resolution models is that they cannot model the detail of the small, weather-producing features and their interaction with the large-scale flow. A hurricane eye wall may have a 100-km diameter, so a model with grid points

[3] NOAA did not have the in-house computational capacity to perform these experiments; the National Science Foundation supplied the HPC.

separated by 90 km (roughly the T190 resolution of the operational GFS ensemble at 30° N) can resolve none of the crucial detail, not the eye, the eye wall, nor the rain bands (Fig. 11; also Gentry and Lackmann 2010). Hence these resolution models cannot be expected to realistically predict the high-impact phenomena such as eye wall replacement cycles and rapid intensification. The practical effect of this can be seen in experiments in 2009 with global forecast models at different resolutions. The bias in the forecast of maximum wind speeds in tropical cyclones was substantially reduced as the resolution was increased (Fig. 12). Figure 13 provides an illustration of the increased fidelity to observations that becomes possible as resolution is dramatically increased.[4]

The adoption of ensemble prediction techniques also has also led to a major positive impact on prediction skill. All major operational NWP centers now run ensemble prediction systems, multiple forecasts from slightly different initial conditions, and possibly using different forecast models. To reduce their computational expense somewhat, at ECMWF and NCEP, the ensembles are conducted with models at half the resolution of the respective deterministic forecasts. Ensemble techniques are necessary because of the chaotic nature of the atmosphere; two model forecasts started from slightly different states will grow to become radically different as forecast lead increases. Ensemble prediction techniques permit a better estimate of the mean state (e.g., Fig. 14, from Toth and Kalnay 1997). The averaging of forecasts filters out the less-predictable aspects while retaining those that are consistent from one member to another. Ensembles also provide quantitative estimates of forecast uncertainty, estimates that can be especially useful for rare, high-impact events (Fig. 15, Palmer 2006, and National Research Council, 2006). Ensemble techniques are also now being used in advanced data assimilation techniques such as the EnKF and EnKF-variational hybrids (Whitaker et al. 2008, Buehner et al. 2010ab).

There are other changes to forecasts that can contribute to increased forecast skill. ECMWF's has dedicated more human and computational resources to developing improved "physical parameterizations." The computational expense of these parameterizations now greatly exceeds the cost of the basic forecast model dynamics, and as forecast parameterizations are improved, they will take up an even larger fraction of the overall computational expense. It is difficult to quantify the specific contribution of improved physical parameterizations to forecast skill, but this effect is significant. Further, we do know that there are many phenomena in the atmosphere whose improper prediction can be traced to deficiencies in parameterizations, such as the Madden-Julian Oscillation (Lin et al. 2006).

---

[4] Grid spacings of O(1 km) as in Fig. 12 will not be possible in the foreseeable future with global models but will be possible by embedding regional prediction models within global models. In section 4, a potential computer upgrade is proposed for 2014, sized to permit ~15-km global ensembles. Nested regional hurricane ensembles with ~1-km grid spacing would be possible using the remaining cycles.

In the future, more realistic forecasts will also be possible if we can couple the ocean, atmosphere, land, cryosphere, and chemistry together in the modeling system rather than modeling them as systems that act independently. These coupled systems will also be more computationally expensive to run.

A final way that has been shown to improve forecast guidance is to post-process it using past forecasts and observations. Systematic model errors can be especially pronounced for these very sensible weather elements that are of greatest interest, such as surface temperature and precipitation. These systematic errors can be mostly corrected before dissemination to the customer if many past forecasts and observations are available. For many phenomena such as long-lead forecasts and forecasts of heavy precipitation, a long set of such "reforecasts" are necessary. The potential impact of reforecasts is shown in Figs. 16 and 17 (Hamill et al. 2006). Unfortunately, reforecasts are computationally expensive; not only must one compute a real-time forecast, but compute forecasts for dates in the past as well. Despite the computational expense, ECMWF runs a 20-year, 5-member ensemble reforecast operationally every seventh day (Hagedorn 2008).

3. **Other factors affecting the future performance of global forecast models**.

Two other technological improvements may allow us to run higher-resolution models for the same expected cost. The first is to deploy models and data assimilation systems that are more computationally efficient and that "parallelize" well. A system that parallelizes well will run approximately in $1/n$ the amount of wall time on n processors compared to its wall time running on a single processor. This may affect the type of models and data assimilation upgrades that are pursued; some approaches parallelize better than others.

Another promising new approach is to adapt forecast models and assimilation systems to run on graphical processing units, or GPUs. ESRL scientists recently demonstrated a 25-fold increase in performance running a simplified global model on GPUs relative to CPUs. Also, the speed of GPUs is currently increasing rapidly and is expected to do so for the foreseeable future, while per-unit CPU performance increases are now slowing. ESRL scientists will be testing and optimizing global models on GPUs in the coming years.

4. A plan for upgrading computational resources.

To provide more quantitative evidence for the impact of upgrades to the global model, we propose a parallel test of a high-resolution global ensemble forecast and data assimilation system using "stimulus" computers soon to be installed at Oak Ridge National Labs (ORNL). We anticipate several months after their installation before these computers are filled with climate applications, and data assimilation and ensemble forecast software can be rapidly ported and parallel tested during this period. We anticipate running a 20-member ensemble daily using NOAA global forecast models to 10 days lead, with the models at approximately 15-

km grid spacing. Forecasts would be compared against NCEP operational forecast using standard metrics and disseminated widely for forecaster feedback. This is the approximate resolution that we expect ECMWF forecasts to operate at in 2014.

Presuming a successful demonstration, NOAA would then have quantitative evidence to justify large HPC upgrades in 2014 and beyond. Again, as a benchmark, let us assume we will deploy an equivalent resolution data assimilation and forecast system to ECMWF's expected 2014 system. Currently ECMWF is running their global ensemble at T639 (roughly 30 km resolution) – this is currently more than three times the resolution of NCEP's operational ensemble. The ECMWF expects to be running ensembles at twice that resolution (15 km) by 2014. A 15-km global ensemble run within a hybrid 4D-Var/EnKF data assimilation system with on the order of 50 ensemble members will require on the order of 50,000 CPU cores, if 10-day forecasts are to be completed in 1 hour's wall time (Hamrud 2010). The cost of the data assimilation would be dominated by the cost of cycling the ensemble forecasts forward 6 hours for the EnKF, so this should be a negligible fraction of the overall medium-range ensemble computational expense. Forecasts beyond 10 days would be computed at a lower resolution, say T319, making the cost of the subsequent monthly forecast again a fraction of the cost of the 10-day forecast. Other resource intensive applications, such as short-range, limited-area ensembles and reforecasts could be computed using the available cycles during different times of the day.

Assuming a backup system is available for operations and parallel testing (another 50, 000 CPUs) and a similar system is available for research (another 50,000 CPUs) this brings the total number of needed CPUs to 150,000. While such systems are expensive now, the per-CPU cost will decrease significantly the next four years.

A major investment in model/data assimilation and software development will be required to effectively make use of new computer resources. An intensive effort will need to improve the computational efficiency and the performance of NOAA operational global forecast model. This will require additional R&D into improved physical parameterizations and representations of model uncertainty. A concurrent effort to accelerate the development of an advanced hybrid variational/ensemble data assimilation system will be needed. Success will require unprecedented cooperation between NOAA research and operational labs, with a focus on developing a ***single*** (but flexible and extensible) ensemble prediction and data assimilation system that is second to none in forecast skill.

References

Anthes, R. A., and others, 2008: The COSMIC-FORMOSAT-3 Mission. *Bull. Amer. Meteor. Soc.*, **89**, 313-333.

Bougeault, P., M. Fuentes, B. Raoult, Z. Toth, C. Bishop, B. Brown, D. Schuster, S. Worley, D. Burridge, D. Parsons, D. H. Chen, E. Ebert, T. M. Hamill, K. R. Mylne, R. Swinbank, W. Tennant, J. Nicolau, T. Paccagnella, Y.-Y. Park, P. de Silva-Dias, Y. Takeuchi, and L. Wilson, 2010: The THORPEX interactive grand global ensemble (TIGGE). *Bull Amer. Meteor. Soc*., in press.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description and single-observation experiments. *Mon. Wea. Rev*., **138**, 1550-1566.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: one-month experiments with real observations. *Mon. Wea. Rev*., **138**, 1567- 1586.

Gentry, M., and G. M. Lackmann, 2010: Sensitivity of tropical cyclone structure and intensity to horizontal resolution. *Mon. Wea. Rev*., **138**, 688-704.

Hagedorn, R., 2008: Using the ECMWF reforecast dataset to calibrate EPS forecasts. ECMWF Newsletter, Autumn 2008, 8-13.

Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2010: Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Mon. Wea. Rev*., conditionally accepted. Available from renate.hagedorn@ecmwf.int.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33-46.

Hamill, T. M., Whitaker, J. S., M. Fiorino, and S. J. Benjamin, 2010: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. Mon. Wea. Rev., conditionally accepted. Available at http://www.esrl.noaa.gov/psd/people/tom.hamill/HFIP-ensfcst-MWR-hamilletal_rev1.pdf .

Hamrud, M., 2010: *Report from IFS Scalability Project*. ECMWF Technical Memorandum 617, 19 pp. Available at http://www.ecmwf.int .

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691-1705.

Lin, J.-L., K. M. Weickmann, G. N. Kiladis, B. E. Mapes, K. R. Sperber, W. Lin, M. C. Wheeler, S. D. Schubert, A. Del Genio, L. J. Donner, S. Emori, J.-F. Gueremy, F. Hourdin, P. J. Rasch, E. Roeckner, and J. F. Scinocca, 2006: Tropical Intraseasonal Variability in 14 IPCC AR4 Climate Models. Part I: Convective Signals. *J. Climate*, **19**, 2665-2690.

Mesinger, F., and others, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

National Research Council, 2006: Completing the forecast: characterizing and communicating uncertainty for better decisions using weather and climate forecasts. National Research Council, 178 pp.

Palmer, T. N., 2006: Predictability of weather and climate: from theory to practice. Chapter 1 of *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds. Cambridge Press, 702 pp.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463-482.

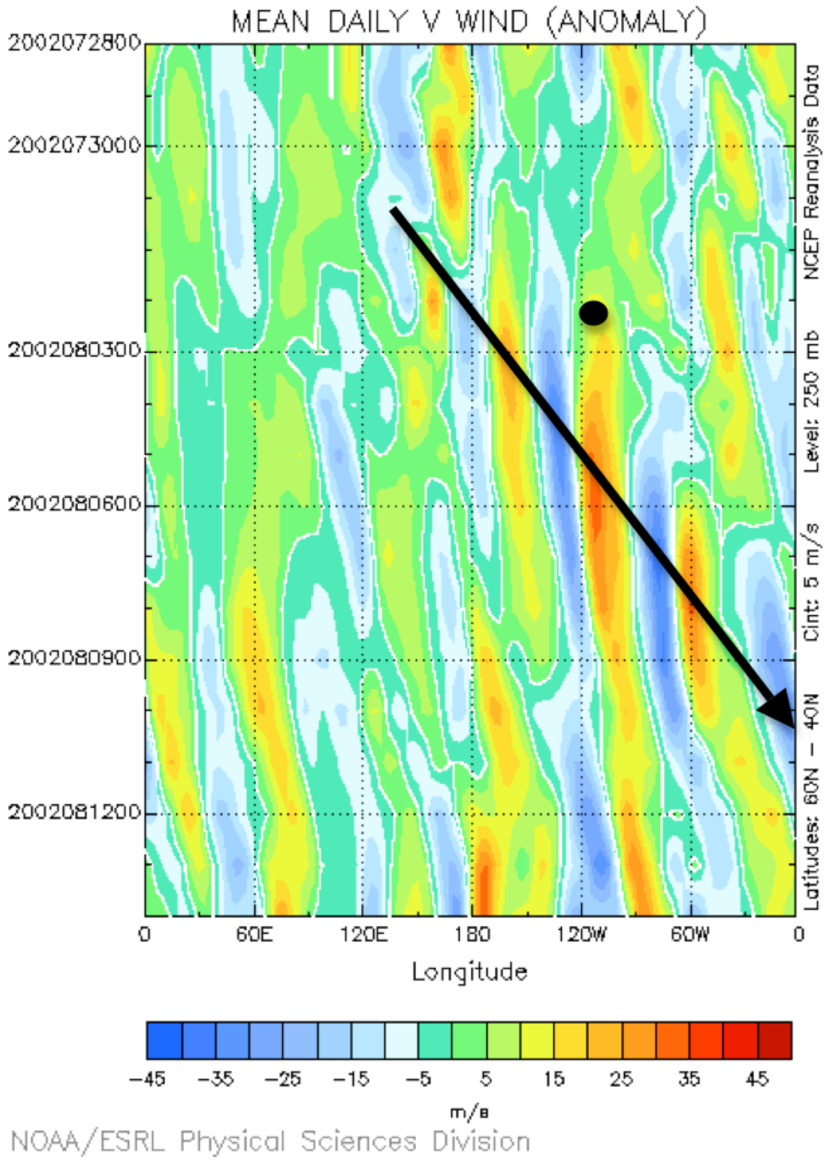


Figure 1: Hovmöller (time-longitude) diagram of the 250-mb meridional wind component (ms^{-1}) for the period 28 July - 14 August 2002 and the latitudinal belt 40-60º N. On 1 August 2002, a growing cyclone near Japan (tail end of arrow) caused downstream Rossby-wave development (arrow) that began to affect western North America two days later (dot). Extreme flooding in central Europe occurred at the end of this period, on 11 August 2002. A skillful global forecast of the cyclogenesis east of Japan and the subsequent Rossby wave dispersion was thus necessary for a skillful short-range forecast over the US and for a medium-range forecast over Europe. Image provided by the NOAA/ESRL Physical Sciences Division, Boulder Colorado from their Web site at www.esrl.noaa.gov/psd/.

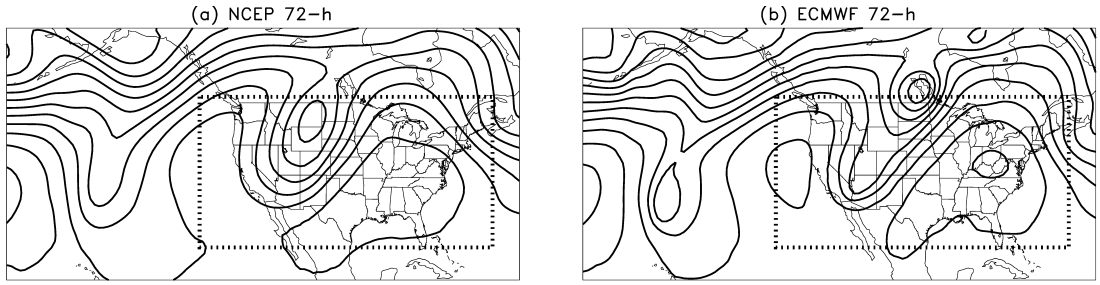


Figure 2: 72-h deterministic forecasts of 500-hPa geopotential height from (a) NCEP and (b) ECMWF operational forecasts, initialized at 1200 UTC 10 October 2008 (which was much closer to the analyzed state 3 days later). Box denotes a hypothetical domain for a high-resolution regional model. Given the large differences in the large-scale flow, the regional model would be unlikely to produce the detail inside the box consistent with ECMWF's forecast given NCEP's lateral boundary conditions; the boundary conditions would strongly affect the solution inside the box. Data downloaded from TIGGE database (Bougeault et al. 2010).

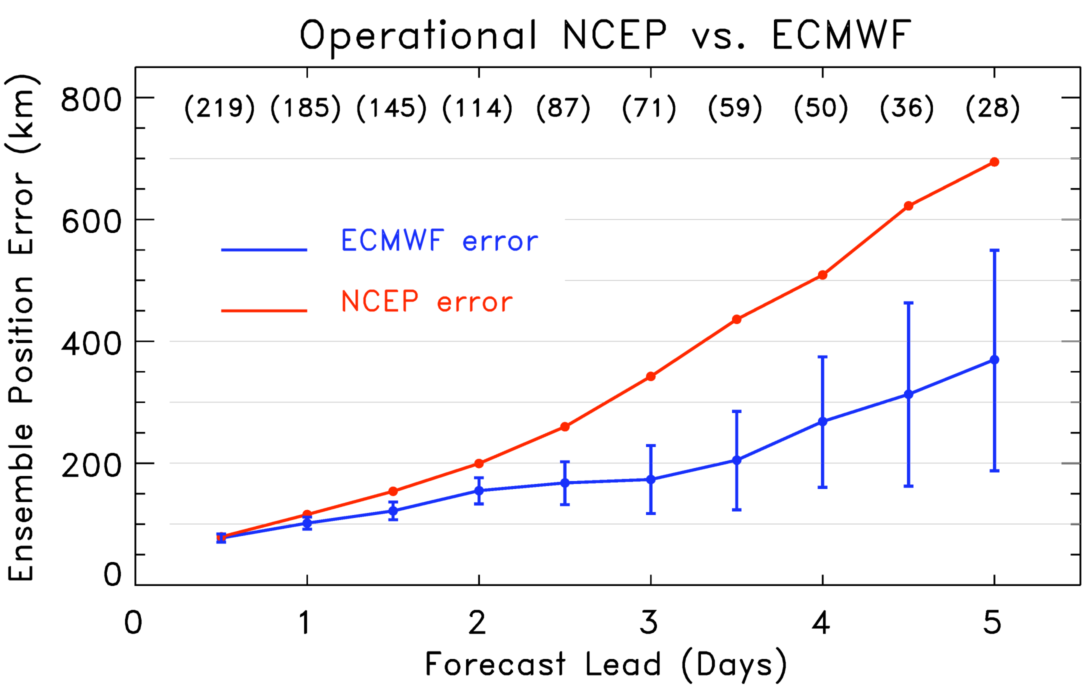


Figure 3: Comparison of NCEP and ECMWF tropical cyclone track forecast mean absolute position error for tropical cyclones between 1 August and 4 October 2009. Error bars denote the confidence intervals; differences outside the intervals are statistically significant at the 5% level. Counts at the top are the number of tropical cyclones that were successfully tracked by both models. Data from Hamill et al. (2010).

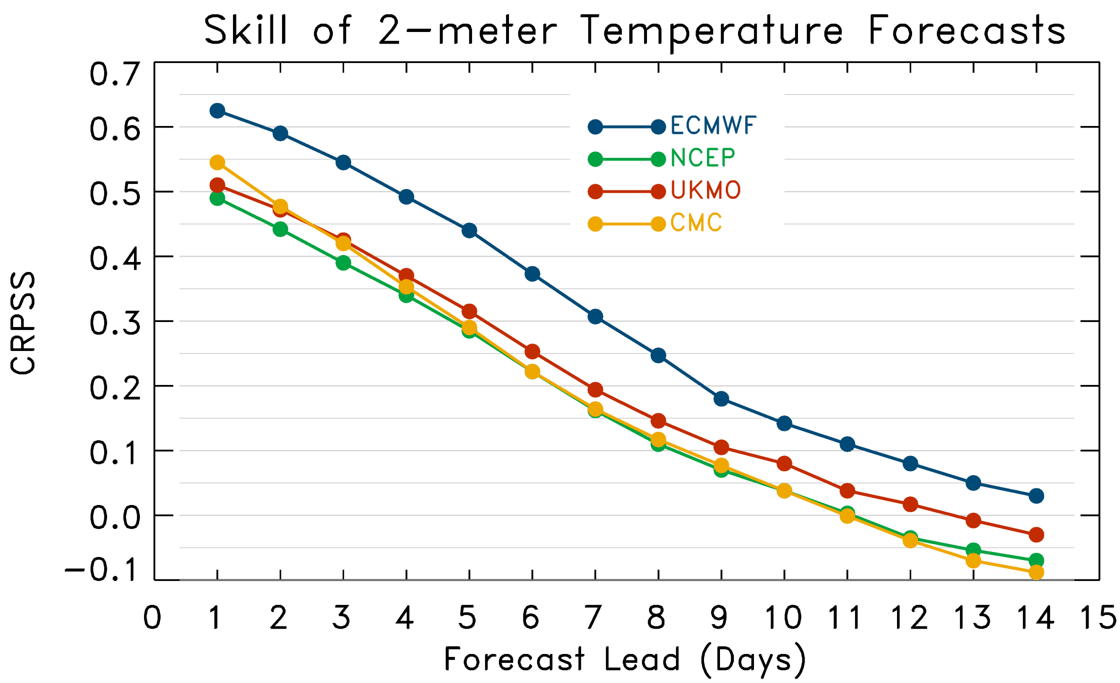


Figure 4: Continuous ranked probability skill score (1.0=perfect, 0.0=skill of climatology) for global 2-m temperature forecasts from four major operational global models. Forecasts are verified against the ERA-interim reanalysis, and each forecast model's output has been bias-corrected using the previous 30-days' differences between the mean forecast and the ERA-interim analysis. "UKMO" denotes UK Met Office forecasts, "CMC" denotes Canadian Meteorological Centre forecasts. From Hagedorn et al. (2010).

(a)

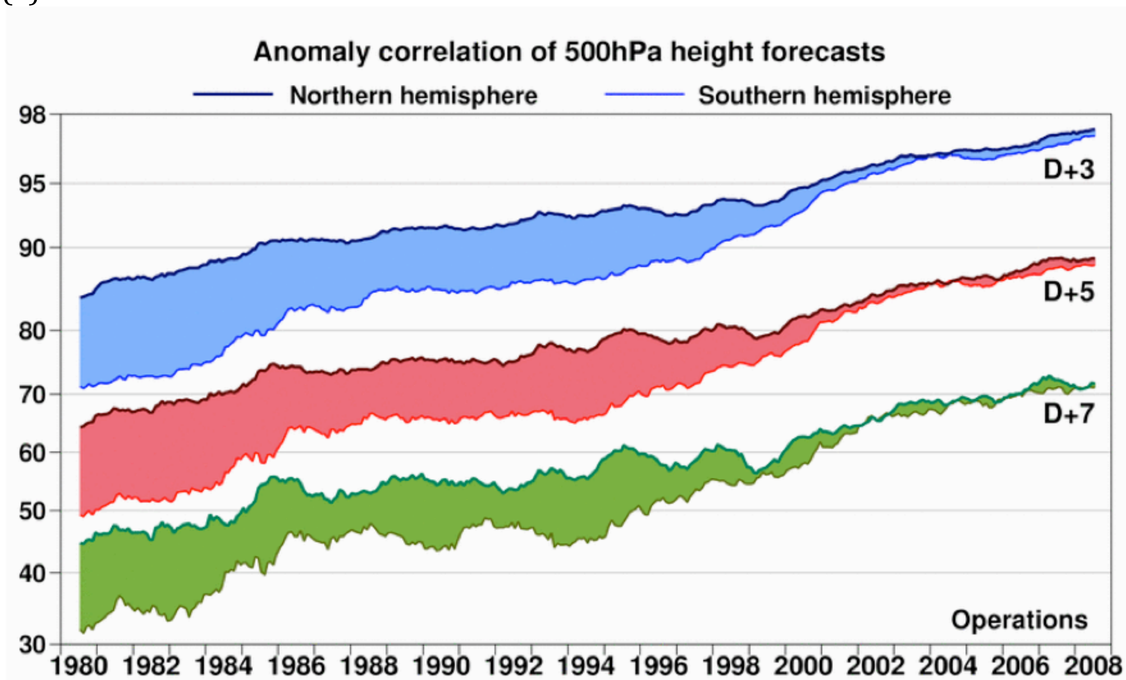


Figure 5: Improvement in the 500-hPa anomaly correlation of ECMWF's operational deterministic forecasts for 3-, 5-, and 7-day forecasts. Courtesy of Adrian Simmons, ECMWF.

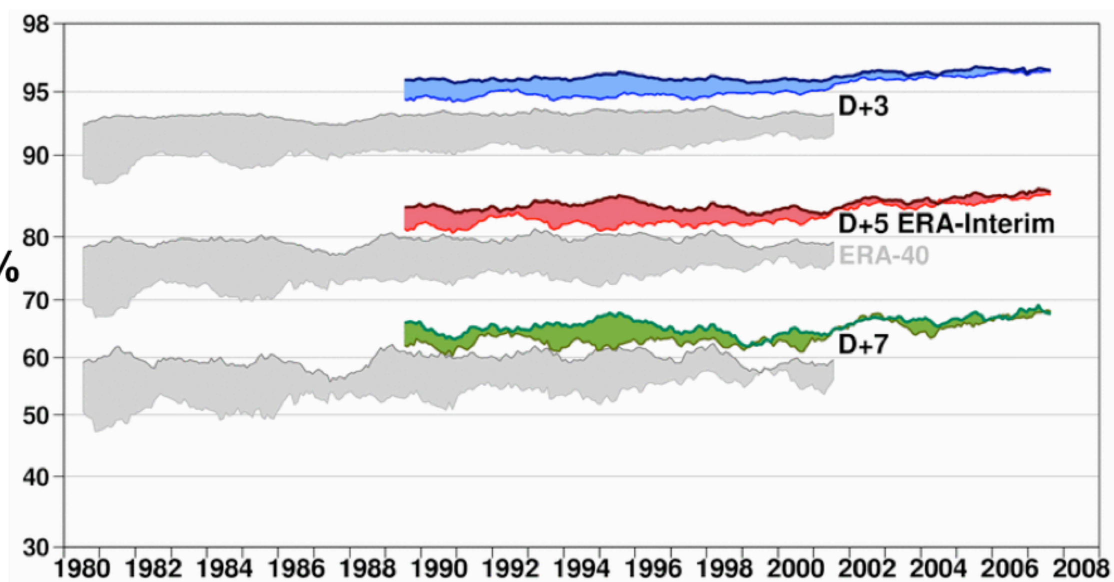


Figure 6: Anomaly correlation of 3-, 5-, and 7-day deterministic forecasts initialized from ERA-40 reanalysis (grey) and ERA-Interim (colored). All ERA-40 forecasts used the same T159 forecast model, and all ERA-Interim forecasts used the same fixed T255 forecast. Courtesy of Adrian Simmons, ECMWF.

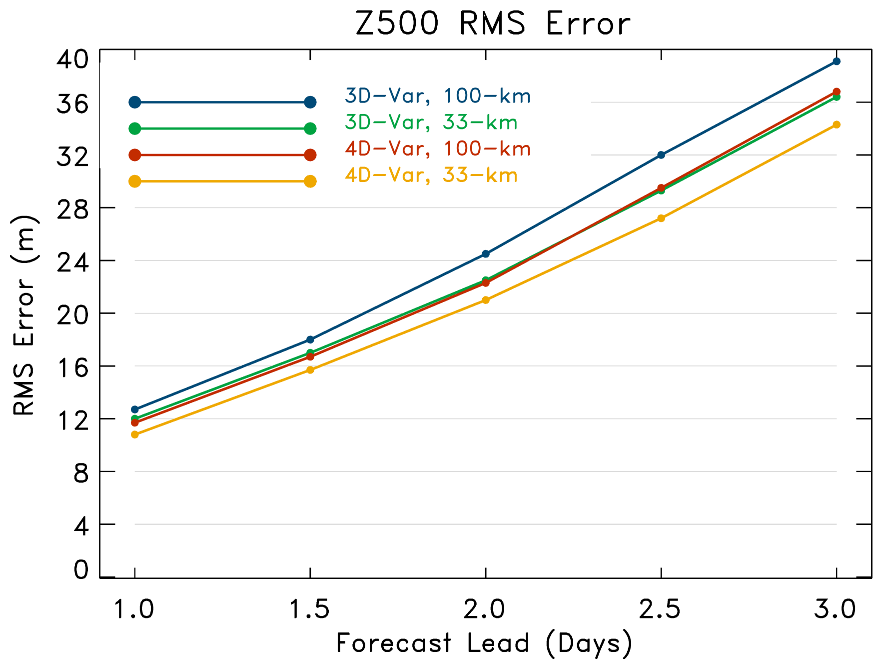


Figure 7. Error of 500-hPa geopotential height forecasts from the Canadian Meteorological Center global forecast modeling system with 3D- and 4D-Var data assimilation and grid spacings of 100 and 33 km. c/o Stephane LaRoche and Gilbert Brunet, Environment Canada.

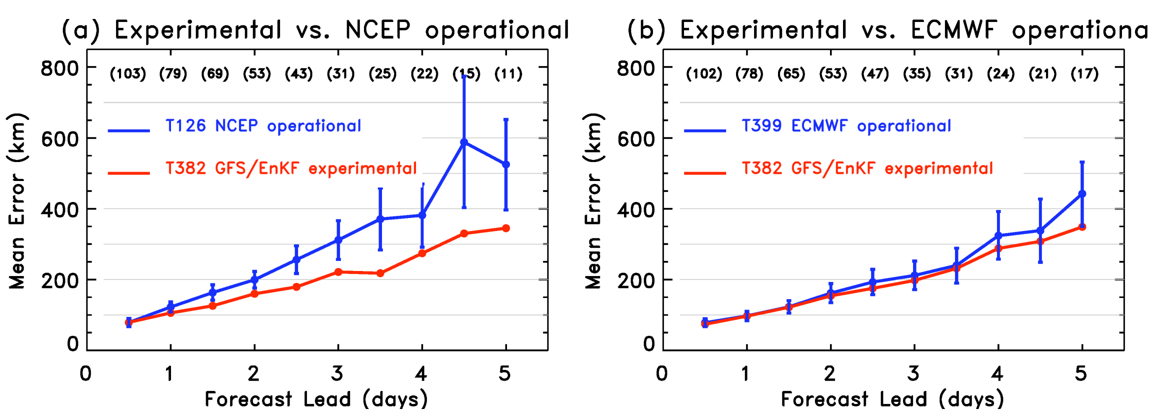


Figure 8: (a) 2009 tropical cyclone track errors from experimental global forecasts with a threefold increase in resolution and an ensemble Kalman filter relative to NCEP operational forecasts. Error bars denote confidence intervals; forecasts outside of the range of the confidence intervals are statistically significant at the 5 percent level. Numbers in parentheses denote the number of cyclones that were tracked simultaneously by both models. (b) as in (a), but against ECMWF's operational forecasts. From Hamill et al. (2010).

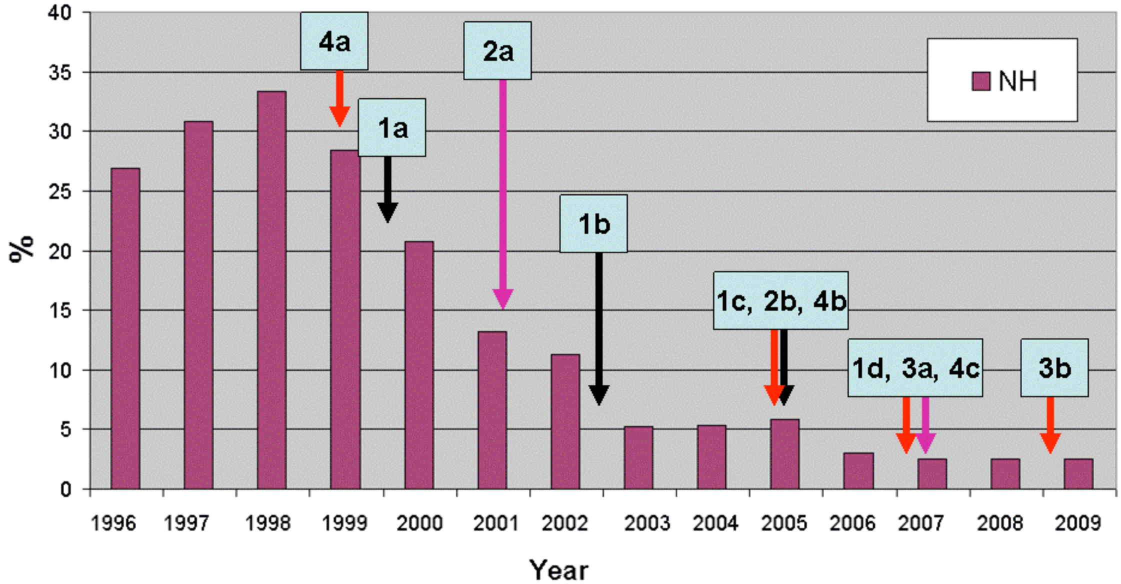


Figure 9: The percentage of 500-hPa Northern-hemisphere "poor" forecasts, defined as those with an anomaly correlation of 0.7 or below, for each year from 1996-2009. Arrows indicate the changes, listed in Table 2, made to the GFS.

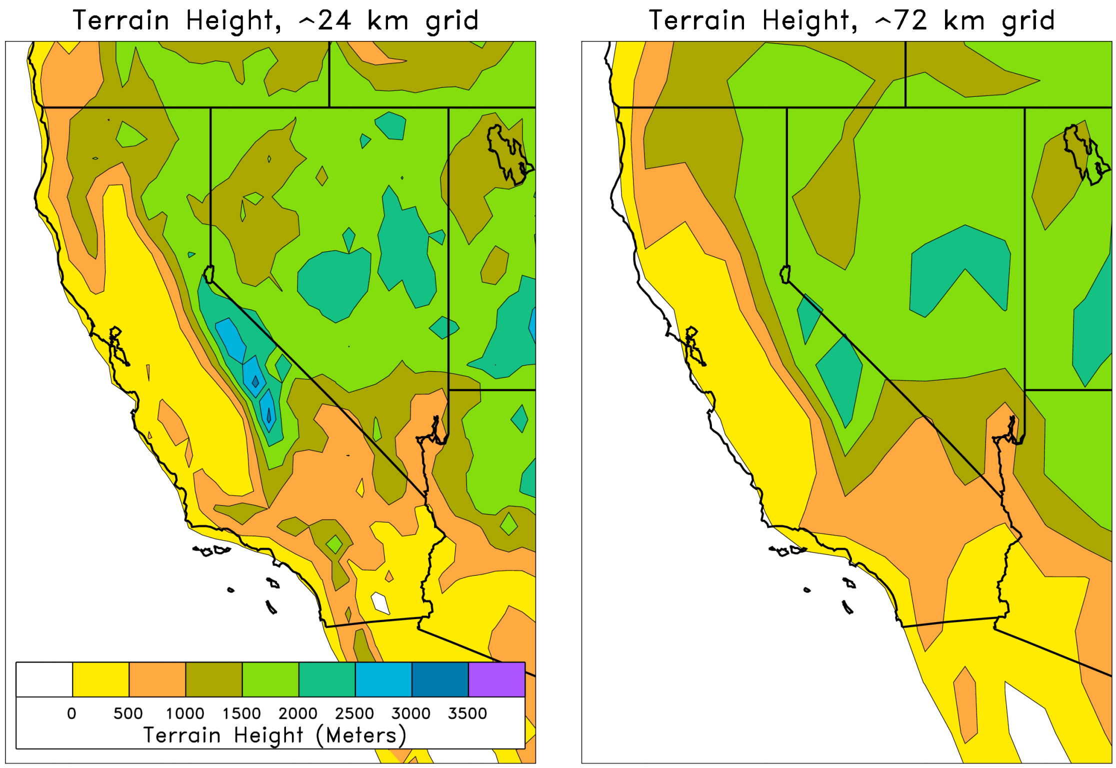


Figure 10: Approximation of the terrain used in ECMWF's operational T639 (~24-km at 45° N) forecast model and NCEP's T190 (~72-km at 45° N) forecast model, here shown for the southwestern US. Actual terrain fields used in the spectral models have additional unrealistic "artifacts" due to the numerical phenomenon of "spectral ringing." Internally generated by T. Hamill using US Navy high-resolution terrain data set.

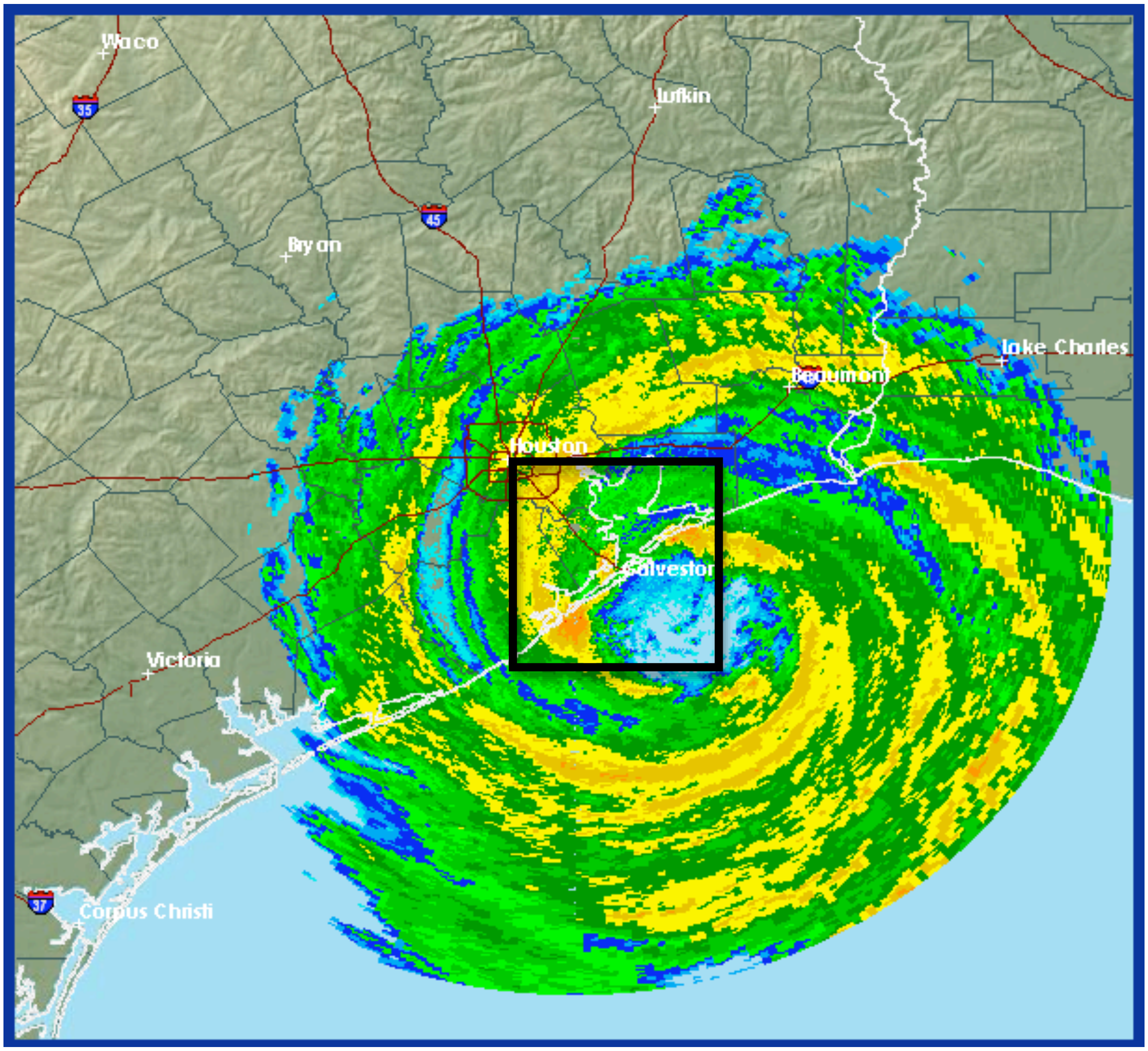
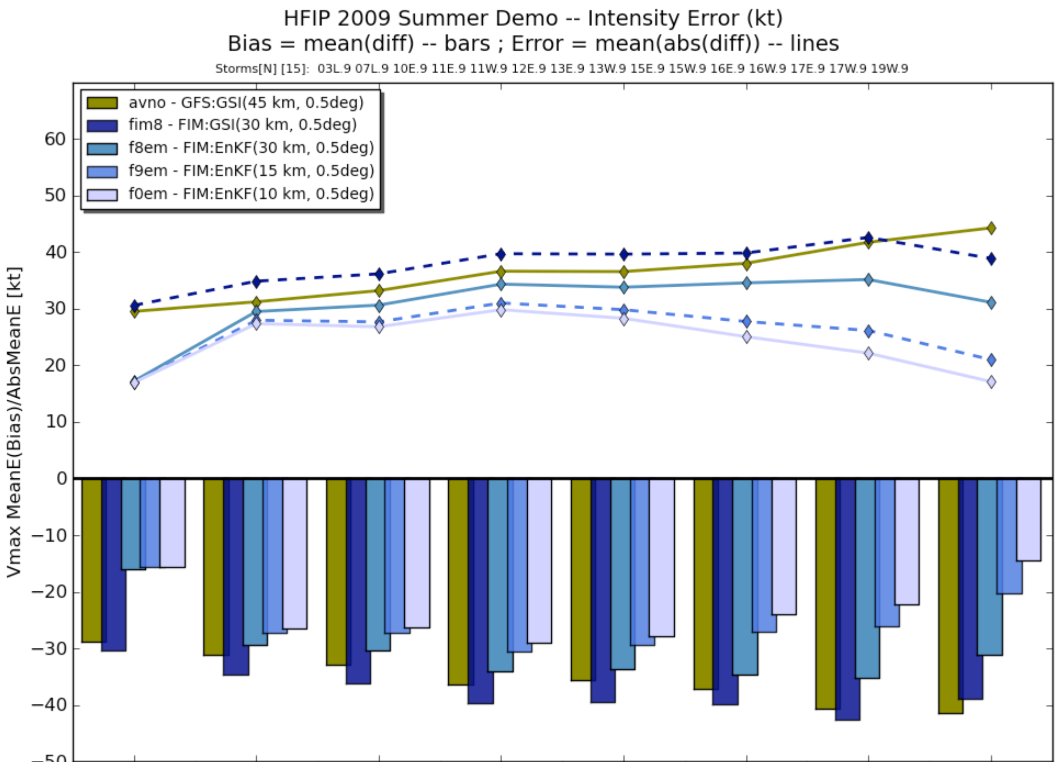


Figure 11: Hurricane Ike radar reflectivity at 0607 UTC 13 September 2008. Box represent the approximate size of a grid box in the current operational T190 GFS ensemble.



	0h	12h	24h	36h	48h	72h	96h	120h
avno	30; -29[74]	31; -31[57]	33; -33[50]	37; -36[43]	37; -36[37]	38; -37[22]	42; -41[14]	44; -42[8]
fim8	31; -30[74]	35; -35[57]	36; -36[50]	40; -40[43]	40; -40[37]	40; -40[22]	43; -43[14]	39; -39[8]
f8em	17; -16[74]	29; -29[57]	31; -30[50]	34; -34[43]	34; -34[37]	35; -35[22]	35; -35[14]	31; -31[8]
f9em	17; -16[74]	28; -27[57]	28; -27[50]	31; -31[43]	30; -29[37]	28; -27[22]	26; -26[14]	21; -20[8]
f0em	17; -16[74]	27; -27[57]	27; -26[50]	30; -29[43]	28; -28[37]	25; -24[22]	22; -22[14]	17; -14[8]

Figure 12: Maximum wind speed error and bias of deterministic forecasts from operational and experimental NOAA models. Lines in the top part of the plot show the mean absolute error; bars on the bottom part of the plow show the mean wind speed bias. Model data plotted are the operational GFS initialized with the GSI 3D-Var ("AVNO"); the NOAA ESRL FIM model initialized with the GSI at 30-km resolution ("FIM8"); the FIM at 30-km resolution initialized with an EnKF ("F8EM"); and the 15-km resolution FIM ("F9EM") and 10-km FIM ("F0EM"). Internal results c/o Mike Fiorino, NOAA.

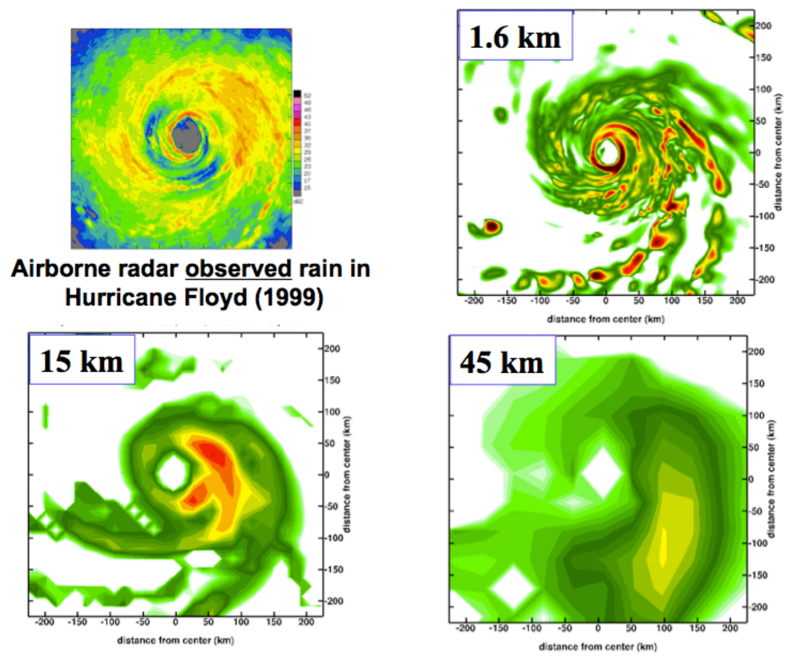


Figure 13: Hurricane Floyd radar reflectivity compared to MM5 model simulations showing the effect of varying the model grid resolution from (b) 1.6 km to (c) 15 km, and (d) 45 km, representative of current research hurricane models, operational regional models, and operational global models, respectively.

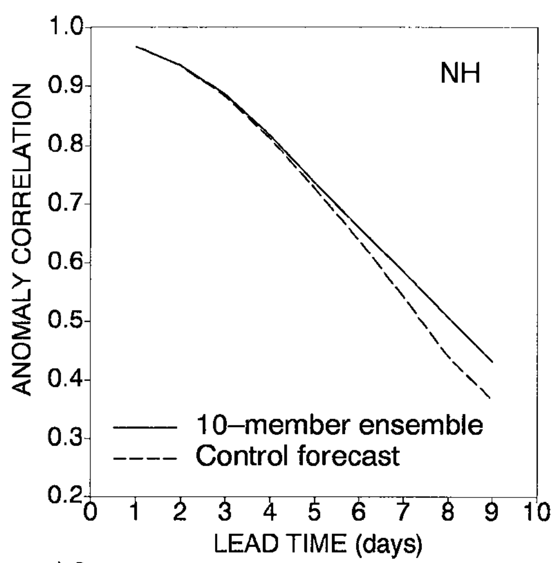


Figure 14: Illustration of the effect of ensemble averaging. Here, the anomaly correlation of 500-hPa geopotential height (larger is better) is compared between a single deterministic forecast and an average of a 10-member ensemble started with perturbed initial conditions. From Toth and Kalnay (1997).

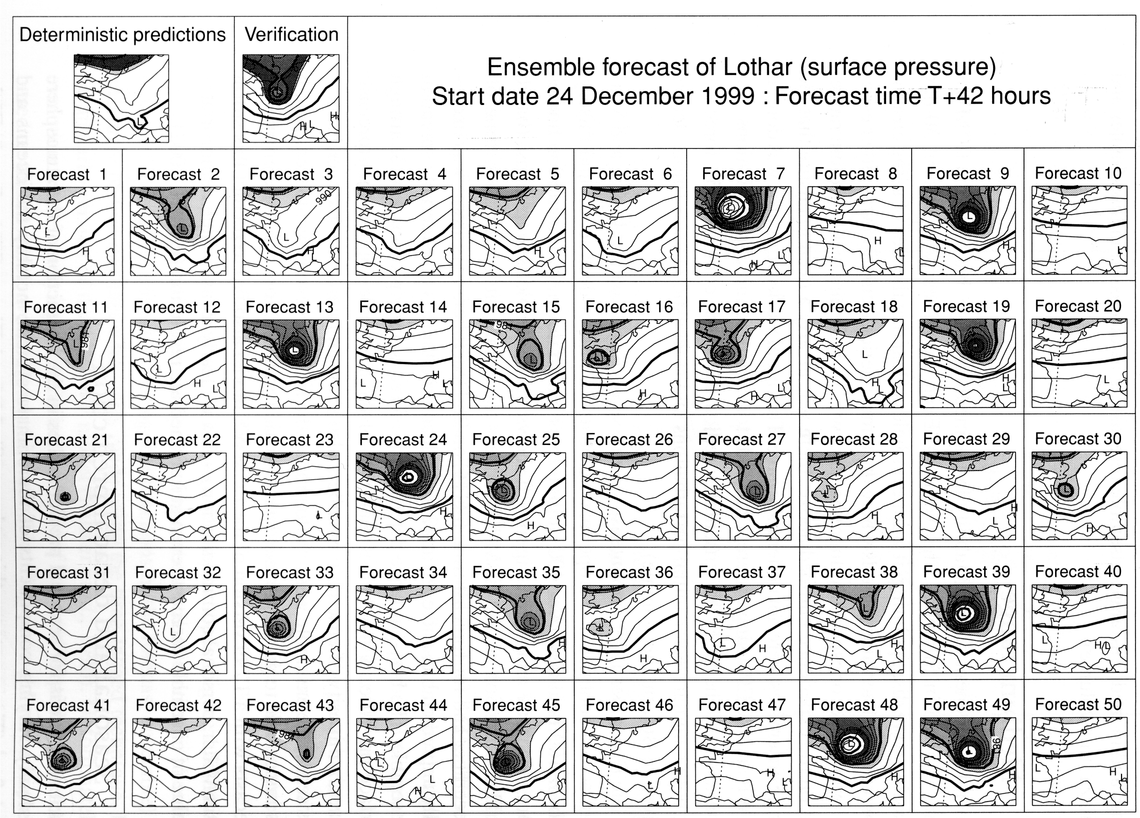


Figure 15: "Stamp map" forecasts of the damaging December 1999 "Lothar" storm in Europe. In all panels, the contours are sea-level pressure 42 h after the initial time of the forecasts, 1200 UTC 24 December 1999. England is in the upper-left corner of each panel, France in the center. The 42-h deterministic forecast (upper-left) failed to produce any storm, while the verification (also upper left) showed that in fact an intense storm (with strong damaging winds) was observed. Approximately one third the 50 forecast members also predict the development of a similar very intense storm. From Palmer (2006).

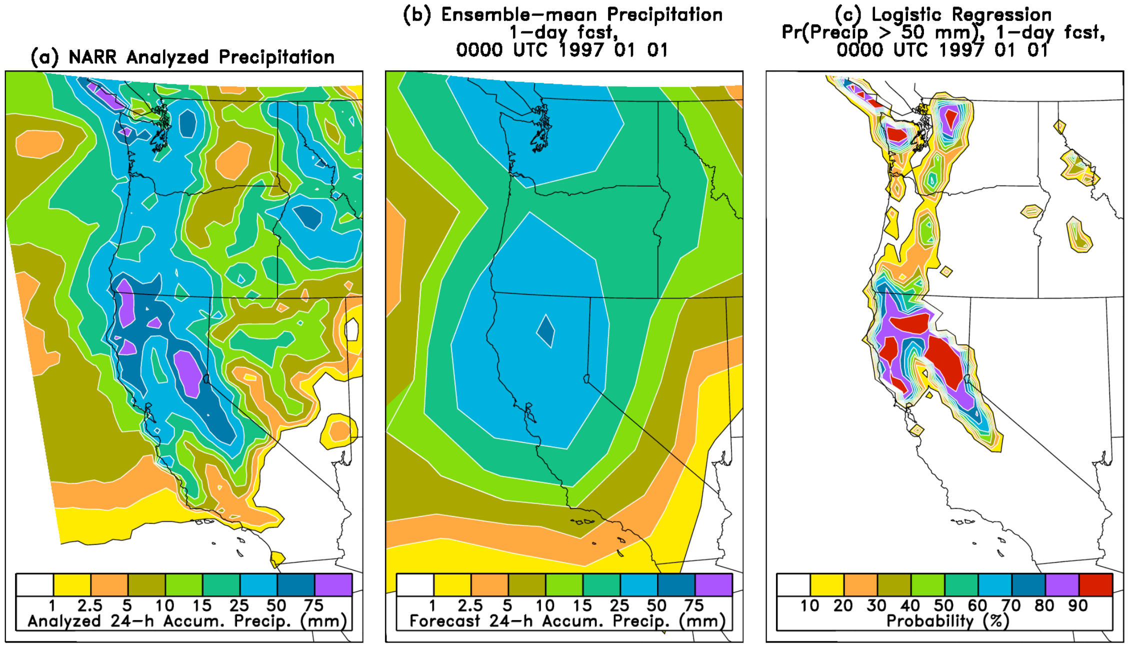


Figure 16: Example of the potential benefit that a reforecast data set can provide through correction of systematic error through its statistical downscaling. Panel (a) shows the analyzed precipitation from the North American Regional Reanalysis (Mesinger et al. 2006). Panel (b) shows the ensemble-mean precipitation forecast from the T62 reforecast data set. Panel (c) shows an estimate of the probability of greater than 50 mm rainfall (~ 2 inches) during this period. The probabilities in panel (c) were developed using a regression model trained on past ensemble-mean forecasts and the past analyzed precipitation. The regression coefficients were then applied to the rainfall pattern in (b). The reforecast is thus able to pick up where the coarse-resolution forecast is raining too much (in the San Joaquin Valley) or too little (along the Coast Range, or the Sierra Nevada Front) on average. For more details, see Hamill et al. (2006).

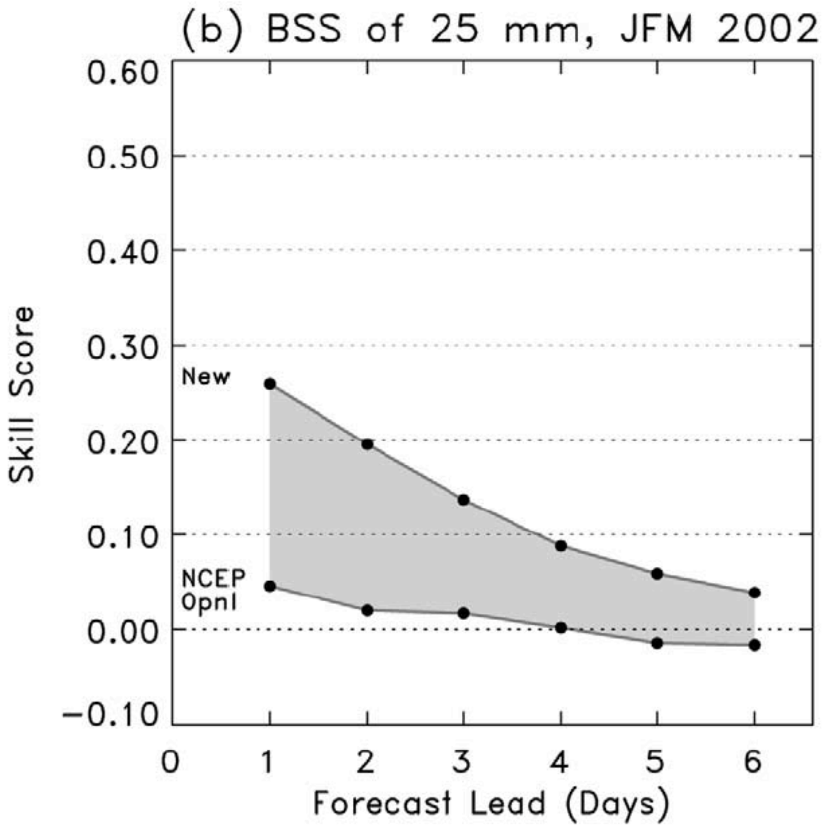


Figure 17: Brier skill score (a measure of probabilistic forecast skill) of precipitation forecasts greater than 25 mm, estimated directly from the NCEP global ensemble prediction system and after post-processing using reforecasts. From Hamill et al. (2006).

	ECMWF T399	GFS/3D-Var T126	GFS/EnKF T382
RMSE 500-mb height, N. Hem.	32.35	39.98	36.22
RMSE 500-mb height, S. Hem.	51.14	63.12	56.72
AC 500-mb height, N. Hem.	0.888	0.832	0.854
AC 500-mb height, S. Hem.	0.891	0.829	0.856

Table 1: Root-mean square errors (RMSE; lower is better) and anomaly correlations (AC; higher is better) of 72-h forecasts from the 2009 operational ECMWF T399 ensemble-mean forecasts, the operational GFS-based ensemble at NCEP (3D-Var initial condition, T126 forecast model), and the experimental T382 GFS ensemble initialized with the EnKF. All errors are measured with respect to the own products's analyses, and all verifications are performed on a 2.5-degree latitude-longitude grid. Results from internal calculations by J. Whitaker, NOAA.

Type of Change	Change	Date
1. Horizontal and/or vertical resolution	1a. Horizontal resolution increase: from 100 km to 70 km and L28 to L42	2/2000
	1b. Horizontal resolution increase: 70 km to 55 km and L42 to L64	11/2002
	1c. Horizontal resolution increase: 55 km to 38 km	5/2005
	1d. Change vertical coordinate from sigma to sigma-pressure	5/2007
2. Model physics	2a. Prognostic cloud water, cumulus momentum transport	5/2001
	2b. Reduce background vertical diffusion	5/2005
3. Data assimilation	3a. Introduce Gridpoint Statistical Interpolation (GSI)	5/2007
	3b. Flow-dependent weighting of background variances and Variational Quality Control	2/2009
4. Adding new observations	4a. AMSU-A and HIRS-3	3/1999
	4b. AIRS central spot, AQUA AMSU-A	5/2005
	4c. COSMIC, full resolution AIRS, METOP HIRS, AMSU-A, MHS	5/2007

Table 2. List of major changes to the NCEP GFS by Type of Change (1-4). Each change in column 2 is referred to in Figure 9.