P4.12 OPPORTUNITIES FOR HUMAN FORECASTERS TO IMPROVE UPON MODEL FORECASTS NOW AND IN THE FUTURE

Stephen D. Jascourt [1] and William R. Bua [2], UCAR/COMET
Silver Spring, MD [1] and Camp Springs, MD [2]

1. INTRODUCTION

It has been quantitatively shown (for example, in HPC verification of QPF and surface low tracks) that human forecasters can consistently outperform NWP models in the present forecaster/ model mix. How will the current role of the human forecaster in the NWS evolve, now that gridded forecast products are being issued for boxes 5 km or 2.5 km on a side?

The authors have been developing and providing training on NWP models for four years to help forecasters both maximize model usefulness and minimize the impact of model flaws in the forecast process. That experience has reinforced our view that forecasters can intelligently assess and improve upon NWP model forecasts by ·

- Understanding the capabilities and limitations of the NWP model components, such as the physics packages, the analysis method, and the data assimilation system
- Knowing the behavior of NWP models in previous scenarios similar to the current forecast situation, and
- Paying attention to the observations.

This provides a basis for determining which parts of model solution are most likely to be correct, and qualitatively adjusting NWP models for some types of error. This adjusted deterministic forecast addresses biases and specific errors that also will appear in ensemble members, and thus it compliments, rather than duplicates, forecasts of uncertainty provided by ensemble prediction systems.

Shown here are case examples illustrating opportunities to correct for various aspects of NWP models, including their synoptic and mesoscale analyses, specific model artifacts and peculiarities, convective parameterizations, and downscaling issues. While improvements in future models will eliminate some of the specific problems shown here, many of these *types* of problems will remain and will be a difficult challenge for modelers to address. We believe that even as model skill continues to improve, humans will still have opportunities to make major operational forecast corrections for specific events, some of which will have significant societal impact.

Corresponding author address: Stephen D. Jascourt, NWS Training Division – OS 6, SSMC2 Rm. 14230, 1325 East-West Hwy, Silver Spring MD 20910

The key factor for human improvement upon the models is recognizing deficiencies in a particular forecast, such as

- Analysis failures as seen by misfit to weather features cross-corroborated by satellite/radar/raobs/aircraft/profiler/other observations or time series of one type of observation
- Model physical deficiencies
- Regime-dependent model biases (we don't address this further here but it is important)
- Systematic model failures (such as precipitation "bombs" that frequently occur in the GFS) which can be identified and removed from the human forecast (model forecast may usually be good or may be good in other parts of the domain)
- Model inability to accurately predict convective precipitation – model may be useful as guidance for broad-area QPF for a widespread flood event but not for local convection forecast – so use other tools: synoptic and mesoscale presence of instability and focusing mechanism/lift.

As models get better, some deficiencies will remain, some will be removed, and some new ones will be introduced. The role of the human will continue and may be most important in events of significant societal impact, for which the model may do a good job of flagging that some sort of serious weather event is possible but the forecaster will need to be attentive and correct model errors in order to get the forecast just right.

While in this article we show examples that suggest problems with the NCEP Eta and GFS models, it is simply to show the role of the human in the forecast process, not to denigrate the outstanding scientists and modeling effort at NCEP. These models were chosen for examples because they are the models most heavily used by the National Weather Service Weather Forecast Offices. Indeed it is our point that modeling is such a challenge that all models have their warts, and a good forecaster, by knowing what those warts are, can know when the model forecast will be on target and how to adjust it when it will not be.

2. MODEL ARTIFACTS AND PECULIARITIES

Every model now and in the future will have its unique artifacts and peculiarities, the impact of which is usually on the mesoscale.

2.1 Boundary Layer Structure

Mixing depth has been a problem in the Eta model for some time. The shallow convective parameterization warms the lower portion of the cloud-bearing layer, which reduces the lapse rate, shutting down the turbulent kinetic energy and associated mixing. We see an example of this in the sounding in Figure 1. More discussion is in Baldwin, et.al. (2002).

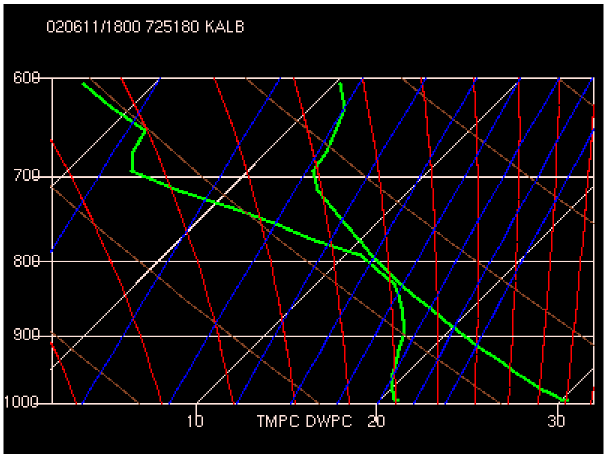


Figure 1. note TKE is only large through the same depth as the moisture is well mixed

Here's another example, from Denver. This one shows it also affects the analysis through assimilation cycling – note how the moisture isn't mixed deep enough even in the analysis.

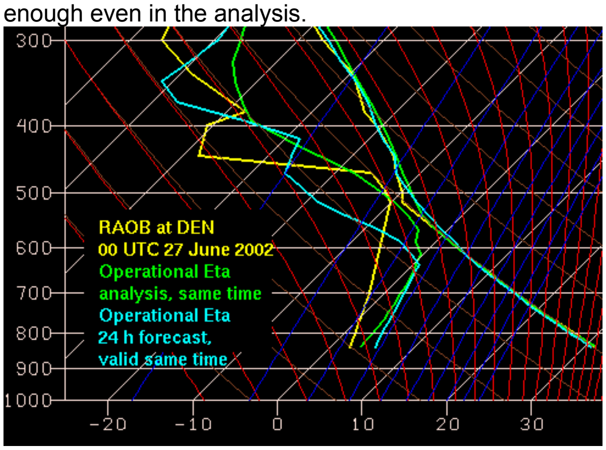


Figure 2. see text for discussion.

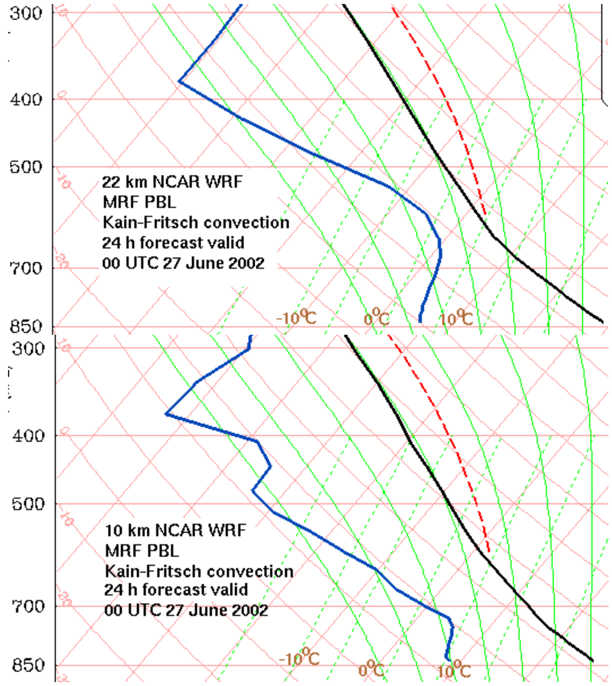


Figure 3. 22-km WRF at top, 10-km WRF at bottom

So this is only a problem in the operational Eta? Here's the same type of problem showing up in NCAR WRF with different PBL scheme, MRF scheme, which is non-local and forces mixing up to the diagnosed to top of the PBL.

Note that the problem is similar in magnitude in the WRF-22 km and Eta-12 km, but is *worse* in the higher resolution 10-km WRF run!

What does this affect?

- CAPE
- 3-d moisture transport
- dispersion of fire smoke or hazardous chemicals
- important for future atmospheric chemistry models

What can the forecaster do to improve upon the model in situations when this type of error is known to be likely?

- Use other bases, such as inflections in the model temperature sounding, to estimate PBL depth
- Mix moisture and other tracers up to that depth
- Estimate CAPE and convective potential based on this adjusted mixing ratio
- Note that this same problem will appear in ensemble means if the underlying physics of the various ensemble members have this problem

2.2 *Surface Condition: Snow/Ice Cover*

The snowcover is updated once daily in the 6 UTC cycle using a snowcover analysis that collects from 18 UTC the day before to 18 UTC the last day, then gets modified by model-predicted melting and model-predicted snowfall in the assimilation cycle. The bottom line is that major changes in the snow field generally lag by around a day in the Eta and GFS model cycles. Figure 4 shows a case in which the model snowcover is both too extensive and, where snowcover was actually observed, the model snowpack appears to be too dense/thick, slowing retreat of the snowline from melting during the forecast. The 6-hour forecast 2-meter temperatures are shown along with a map of observed temperatures at that time. Forecasts are around 10 deg F too low where the actual snowcover had melted. Forecasters can expect to see more forecast errors of this sort where the snow line has retreated from melting after 18z the previous day. The warm sector temperatures appear to be well forecast where the model has bare ground, north up to Cincinnati and east to Charlestown, WV. Temperatures further north over Michigan and Wisconsin are also close at this time, as they ought to be in a 6-hour forecast.

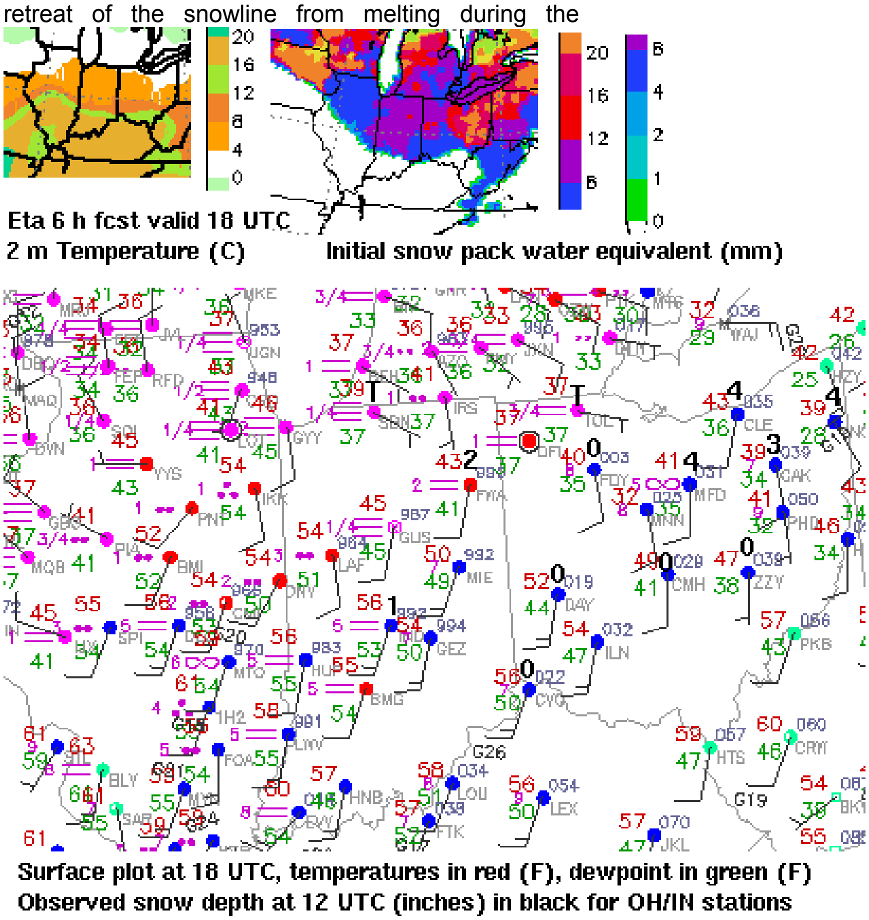


Figure 4.
See text for discussion.

Future of problem: real-time remote sensing observations and better assimilation cycling of snow will improve snowcover analyses. However, hourly METAR snowfall amount observations are of such poor quality that they ought not be used in the assimilation, and cloud cover will impede use of satellite data. This will continue to be a problem as well for ice coverage on the Great Lakes, where substantial changes can occur underneath a week-long overcast cloud deck.

Forecast impacts:

- Warm front movement/location
- Surface temperature
- Clouds
- Boundary layer depth
- Lake-effect precipitation

What can the forecaster do to improve upon the model?

- Monitor snow and ice coverage in the model analysis, model forecast, and all available observations
- Monitor snowpack water equivalent in the model and observations to determine if there may be important errors in the amount of water to melt
- Adjust the forecast based on where discrepancies are expected in locations of bare ground or open lake water during the forecast
- Note that the same adjustments will be needed for all ensemble members using the same snow pack and ice cover

2.3 Cloud-Radiation Feedback

The Eta model has too much solar insolation reaching the surface, both in clear sky conditions and through cloud cover. In the example shown in Figure 5, the shortwave radiation received at the ground resembles that on a clear day, and correspondingly the model has warmed the boundary layer enough to erode the low clouds early in the day, resulting in large positive temperature errors during the daytime, drying at the surface and even throughout the boundary layer, as the inversion is weakened sufficiently to increase entrainment/exchange with the free atmosphere.

This problem could be related to insufficient cloud thickness at the initial time, to slight errors in wind direction affecting the amount of upslope, vertical diffusion or physical mixing processes that eroded the clouds prematurely, poor partitioning of surface fluxes – perhaps too much into sensible heating, perhaps drizzle or light snow was actually occurring that reduced sensible heating, or other factors. Cloud-radiation feedback loops can also involve the land surface model, boundary layer mixing and vertical diffusion, and can be rather complex to try to untangle.

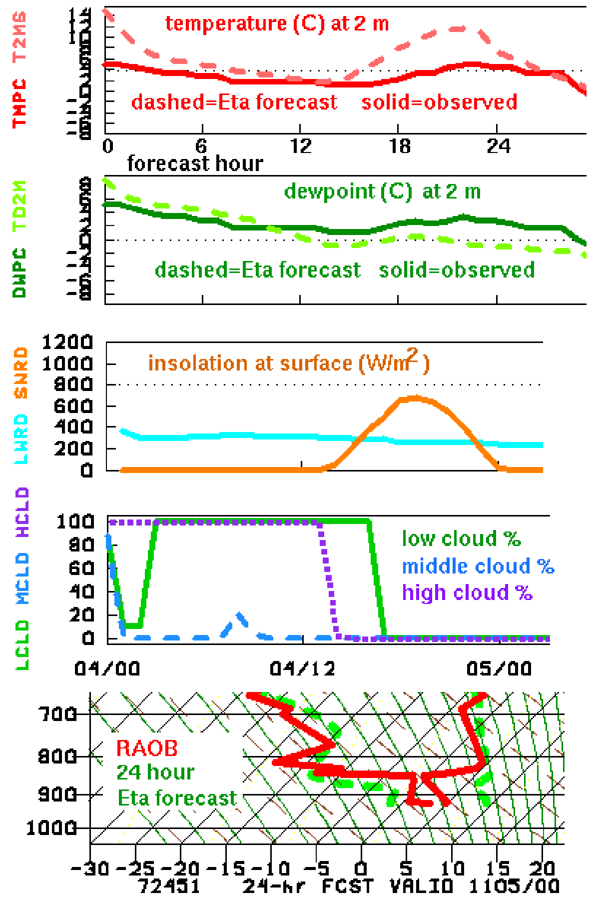


Figure 5. see text for discussion

In the future, when problems such as that shown here are reduced, models will reproduce the natural sensitivity to cloud cover. In a high-resolution model, this can lead to convective cloud shading, which is likely to be misplaced and mistimed given the level of skill at predicting the exact location, coverage, and timing of convection. This could spawn a cascade of mesoscale forecast errors in a properly sensitive model. The more detail a model can create, the more it will be able to create realistic-appearing artifacts.

Forecast impacts:

- Surface temperature and dewpoint
- Cloud cover
- Possibly PBL depth

What can the forecaster do to improve upon models in the future when the radiation-cloud coupling is improved but problems of sensitivity exist?

- Ensembles may provide the best guidance
- Humans will need to identify systematic biases such as that shown here for the Eta model in order to improve upon ensemble forecasts

3. LARGE-SCALE ANALYSIS PROBLEMS

There seems to be a commonly held impression that the problem of large-scale analysis has been solved with modern data assimilation systems ingesting an enormous volume of global remote-sensing data. Most of the variance in the atmosphere occurs at large scales (red noise spectrum), the large scale regime determines possible mesoscale states, cyclone tracks, whether moisture sources such as the Gulf of Mexico can be tapped or are shut down, and so on. The success of the large-scale analysis is manifested in long lead-time predictions in very recent years for major weather events – even if the details could not be predicted in advance, to be able to predict many days in advance, for instance, that a major east-coast cyclogenesis event would occur is a remarkable achievement.

However, the reality is that despite these successes, there are still notable shortcomings in the large-scale analysis that the forecaster can identify and use to adjust the model forecast. Diagnostics which Steve Silberberg used to post on the AWC model verification web page, comparing forecast-analysis differences at different initial times with the same valid time, show that general sense of error patterns with a given synoptic system tend to be repeated from one model run to the next – this is clearly the influence of the first guess in the analysis and suggests that there is an underlying synoptic-scale structure to analysis errors of a given system. Often, IR or WV satellite imagery reveals parcel trajectories arcing anticyclonically into, through, and out of a subtropical jet core which is located hundreds of kilometers from where the model analysis has the jet core. Sometimes, a southwesterly subtropical plume appears in the water vapor imagery to be more impressive than suggested by the model output, and the precipitation shield ends up tracking further to the north than predicted by the model. In the case of the Presidents' Day 2003 east-coast snowstorm, over several days the models progressively predicted the storm to occur later and track further north, as systematic errors in the analysis had the upstream upper low not deep enough, the jet core with the southern stream not far west enough. Also, the preceding northern stream system was not handled quite right, so that the jet entrance region as it moved off the coast created more impressive anticyclogenesis than previously predicted. The deeper southern stream western system allowed embedded waves to take longer to round the trough and the system was more cut off than in the models, all resulting in it moving slower, which allowed the northern stream wave train enough time to advance eastward to avoid confluence from blocking northward progress of the east coast cyclone.

A useful tool for examining some of these analysis errors quantitatively is the analysis-radiosonde discrepancy field. NCEP has a web page showing observation increments and analysis increments interpolated to the station location, so that the difference is the portion of the observation not taken into account by the analysis. One should expect to these to be small except when there is conflicting information that suggests a possibly large observation error, and one should expect these to have no structure at wavelengths longer than the correlation lengths used in the analysis covariance structure. However, sometimes, large and systematic differences do occur, and these are signals that the forecaster can use to identify large-scale adjustments needed to the model forecasts. Unfortunately, no such tool presently exists in AWIPS.

How applicable is this problem to the future? Improvements in resolution and in mesoscale data assimilation are not going to have much affect on this problem. Satellite-based Doppler winds through the depth of the troposphere, should those someday become available, could significantly improve this situation.

4. MESOSCALE ANALYSIS PROBLEMS

Effective utilization of mesoscale data in mesoscale models is at the cutting edge of research today, and operational models leave much to be desired in this area, and even with improvements, there will still be room for humans to play an important role in the forecast process.

4.1 Large Errors Render Ensembles Useless

When the error in the analysis is well beyond the size of ensemble perturbations, the correct forecast will lie outside the ensemble envelope. Furthermore, by identifying the analysis error, a human forecaster can know what direction to change a forecast rather than simply have ensemble-based information regarding uncertainty. An example of such a case is shown in Figure 6.

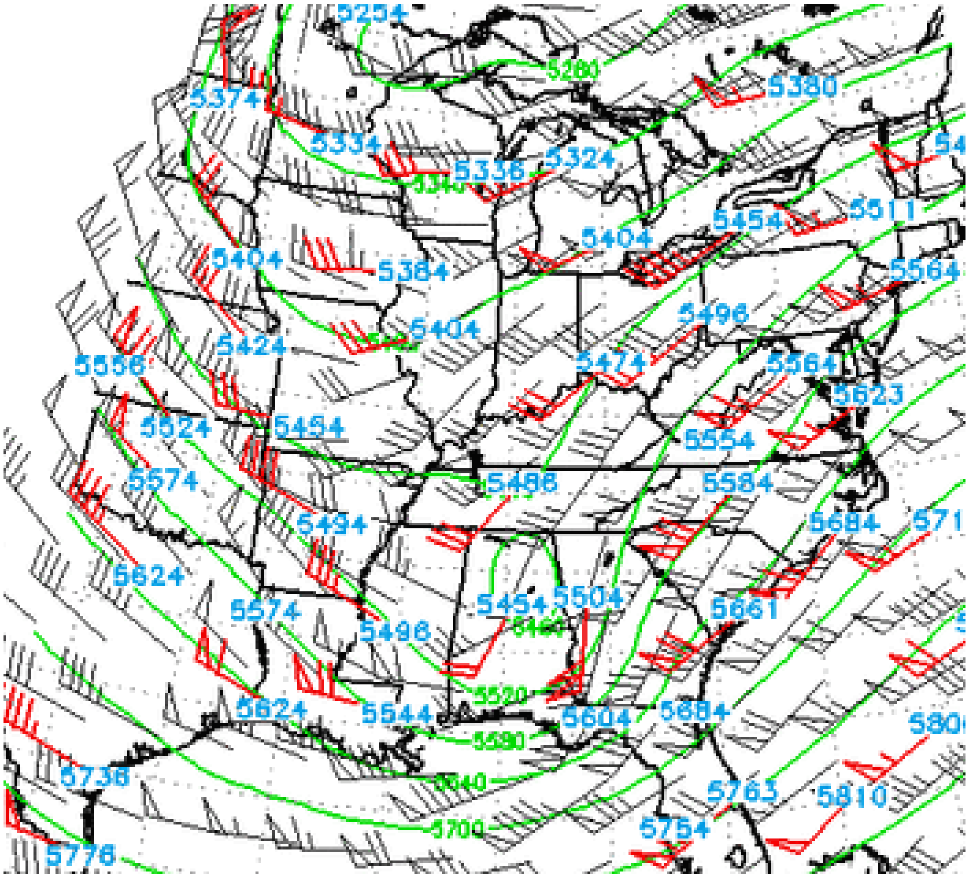


Figure 6. 12 UTC January 6, 2002 GFS analysis heights in green, winds in gray with radiosonde heights in blue and winds in red. Note the 100 knot wind observation at Peachtree City (Atlanta)

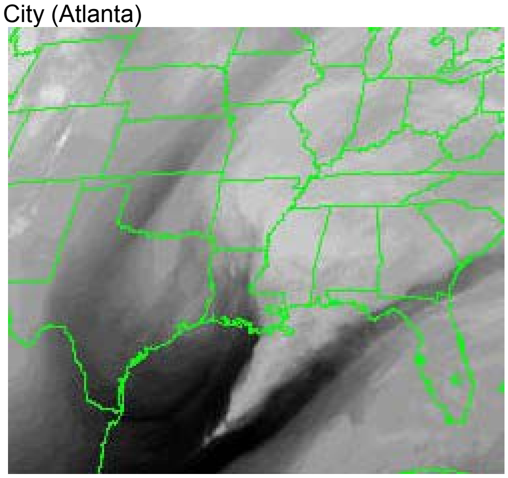


Figure 7. Water vapor imagery from 0015 UTC 6 January 2002

Note the 50-knot error in wind speed over western Georgia, the sharpness of the wind shift in the trough, and the sharpness of the speed gradient around the trough axis. The GFS and Eta analyses and the analyses of every ensemble member were poor, and none of these models or their ensembles predicted the snowstorm which commenced over Pennsylvania and New York shortly after this time. The observations clearly indicate a stronger wave with a sharper trough than predicted, which would push the precipitation shield further north and west into the cold air. A satellite image taken twelve hours earlier (Figure 7) shows the sharp trough with diffluent cirrus from convection over the Gulf of Mexico – a scenario that often leads to trouble with the models.

Poor first guesses are probably a large contributor to analysis errors such as this, and first guesses will improve as models and assimilation systems improve. But for the foreseeable future, forecasters can still improve upon models by being alert for large mesoscale errors such as this one.

4.2 Errors Caused by Isotropic Errror Covariances

It is commonly recognized that mesoscale data assimilation will require appropriate mesoscale structure in the background error covariances, and NCEP already is working on developing this capability. However, it may be a while until models become good at this. Here is a dramatic example of the type of problem that more often occurs less dramatically:

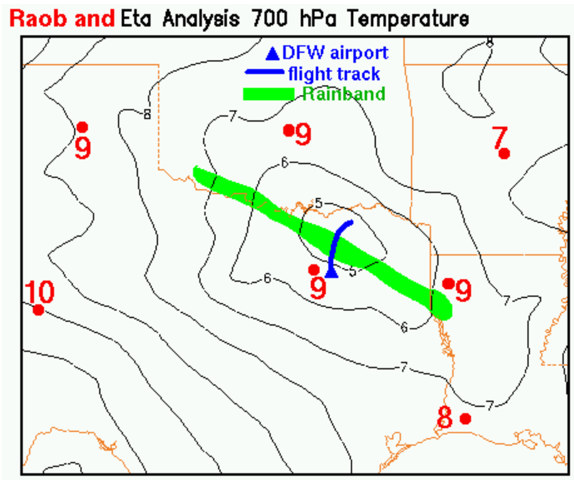


Figure 8. see text for discussion

Figure 8 shows the operational Eta model temperature analysis (°C)at 700 hPa in black contours along with radiosonde temperatures plotted in red. A rain band producing evaporative cooling was positioned across north Texas and aircraft flying into and out of Dallas-Fort Worth (DFW) airport passed through the rain band at around 700 hPa. The data from this flight track were extrapolated by the analysis into the pattern of cool temperatures shown in Figure 8 a long way outside the rain band.

Forecasters can improve upon the model by being alert to improper structure in the analysis and need to build in the proper structure (such as cool temperatures in immediate proximity to the rain band) when considering their nowcast and forecast.

4.3 Small-scale Features

For a week, beginning July 3, a series of derecho events tore across the central US. A satellite loop from the morning of July 4 revealed an intense vortex, showing intense rotory motion in the water vapor imagery. One image, from 12 UTC, overlayed with 300 hPa winds from the radiosondes in yellow and from the Eta operational analysis in blue, is shown in Figure 9.

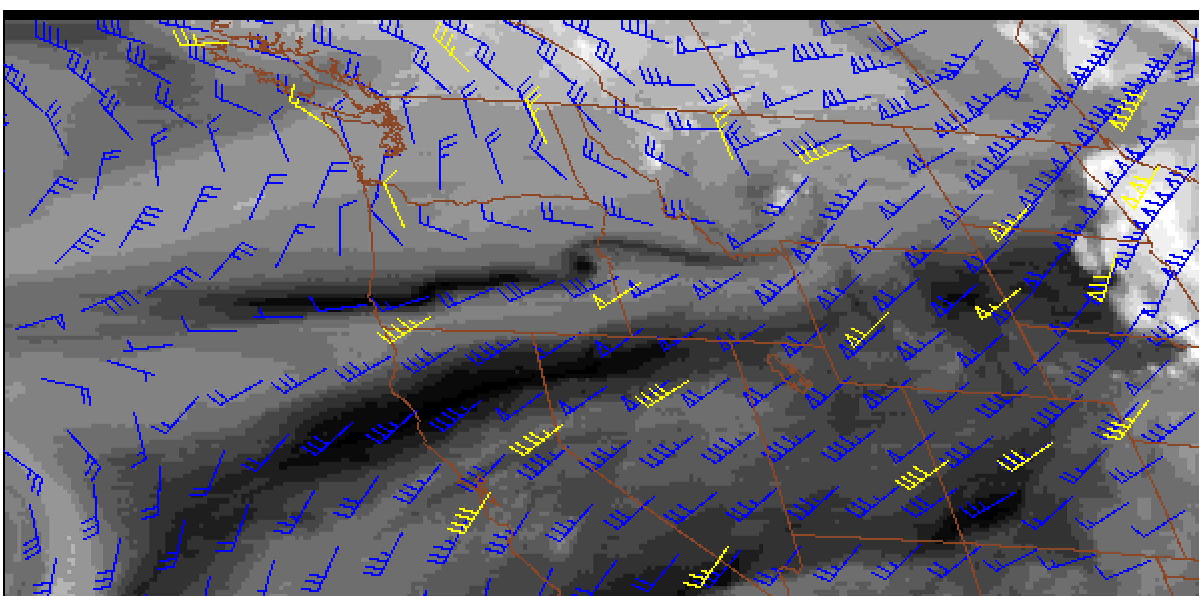


Figure 9. Water vapor image of vortex over southern Montana and string of smaller vorticies from off Oregon coast to the Montana vortex. Eta analysis 300 hPa winds are in blue and radiosonde winds in yellow.

The rotor was large enough to be captured in the Eta analysis even though the model shows a linear flow through the center of it. However, the model was unable to capture any of the series of trailing vortices, many of which became more evident in the satellite imagery at later times. The next derecho event initiated later that afternoon over South Dakota inside the core of the rotor vortex, shown at 1200 UTC over southern Montana. Subsequent severe convective systems initiated in conjunction with the other trailing vortices.

In the future, while narrow features such as these may be better incorporated into the model analysis, an opposite type of problem will occur in high-resolution models with excellent physical detail: spurious, physically self-consistent features in the model first guess that remain in the analysis due to insufficient observation density and insufficiently detailed background error covariance structure to correct them. Forecasters will then be able to improve upon the model by recognizing not only real features that the model may have poorly included, but incorrect features that the analysis does have.

5. PROBLEMS WITH CONVECTIVE INITIATION AND CONVECTIVE EVOLUTION

Convective initiation and convective evolution are well known difficult forecast challenges for which model skill levels are poor compared, for instance, to skill of synoptic overrunning precipitation events. Forecasters can do better by utilizing more accurate aspects of the model forecast of the convective environment and mechanisms that could supply lift to sustain convection. Additionally, a forecaster can make specific types of adjustments by knowing the characteristics of different parameterizations, such as that the BMJ convective parameterization in the operational Eta model generates its precipitation entirely from precipitable water in the cloud-bearing layer (none from the subcloud layer) and typically triggers too early in a moist environment but not at all in a dry mid-level environment characteristics of plains severe weather.

However, just to put in perspective the challenge of predicting precipitation on the 5-km National Digital Forecast Database, consider the differences in precipitation “observed”! Figure 10 shows precipitation estimates from quality-controlled, mosaiced Stage IV multisensor data, from RFC gauge data, and the latter on a coarser grid.

Convective parameterizations will be needed for the foreseeable future for large-domain real-time NWP. Small-domain runs with explicit convection require detailed mesoscale observations (Xue et..al. 2003) and may require grid spacing on the order of hundreds of meters or less for physically realistic triggering (Bryan, et.al., 2003). Experiments at 6 km (Fowle and Roebber, 2003) showed surprisingly good skill in overall convective forecast scenario over a region but of course not for individual cells or precise locations.

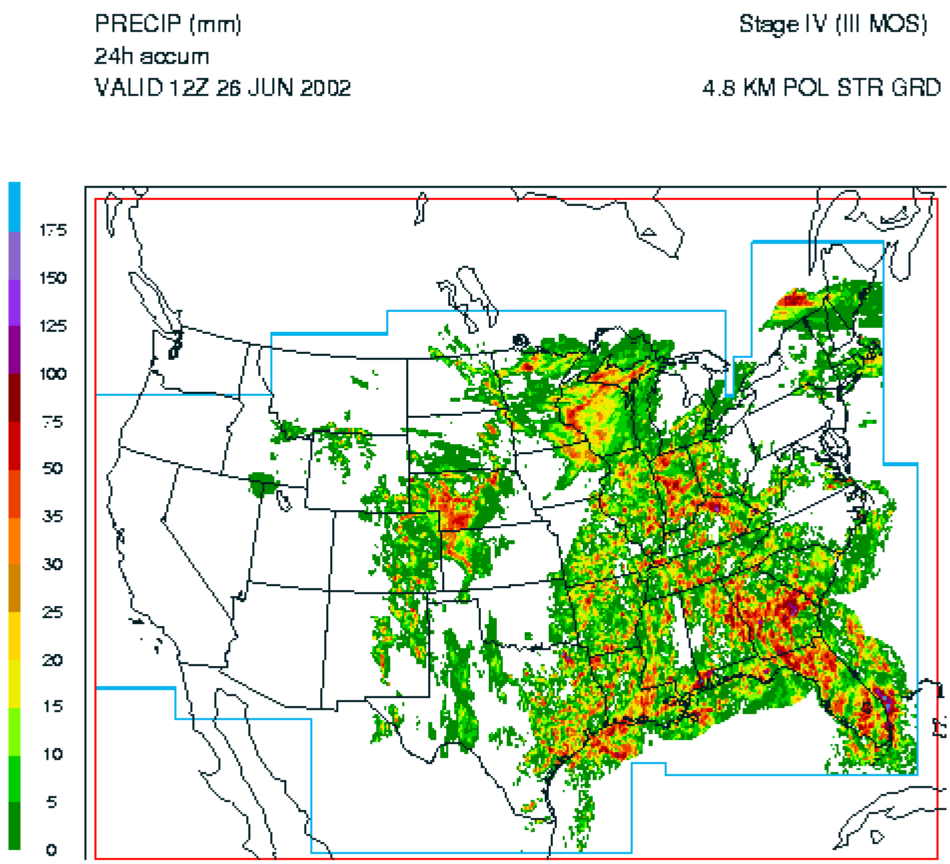
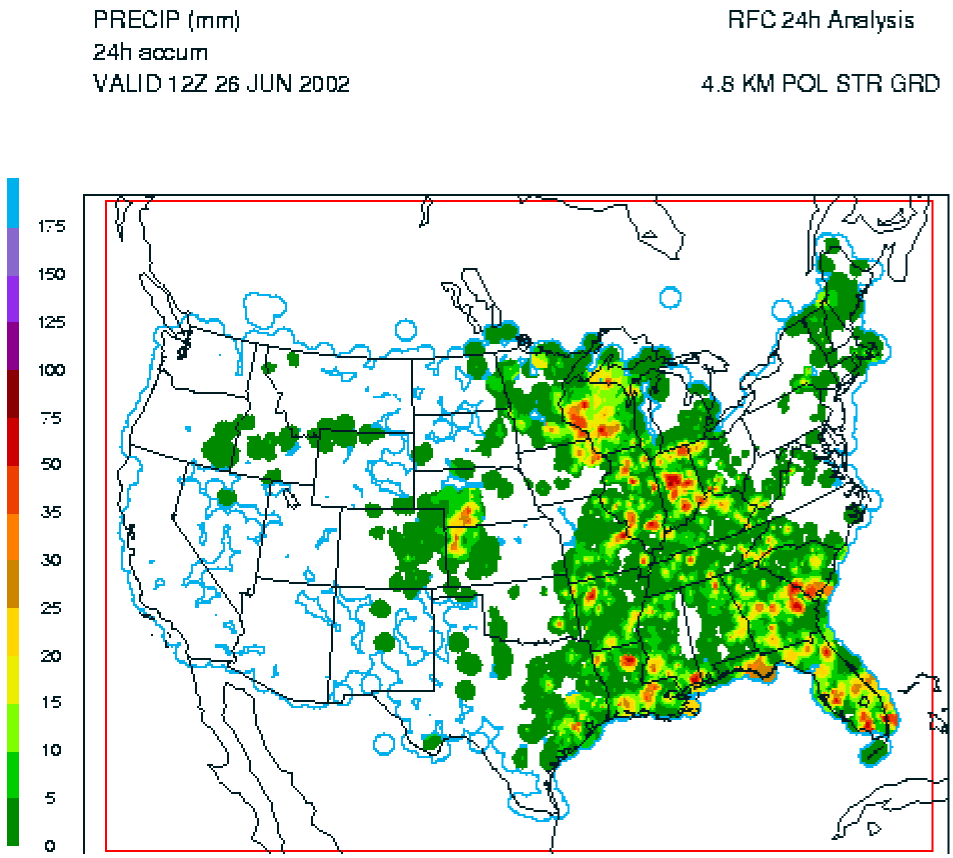
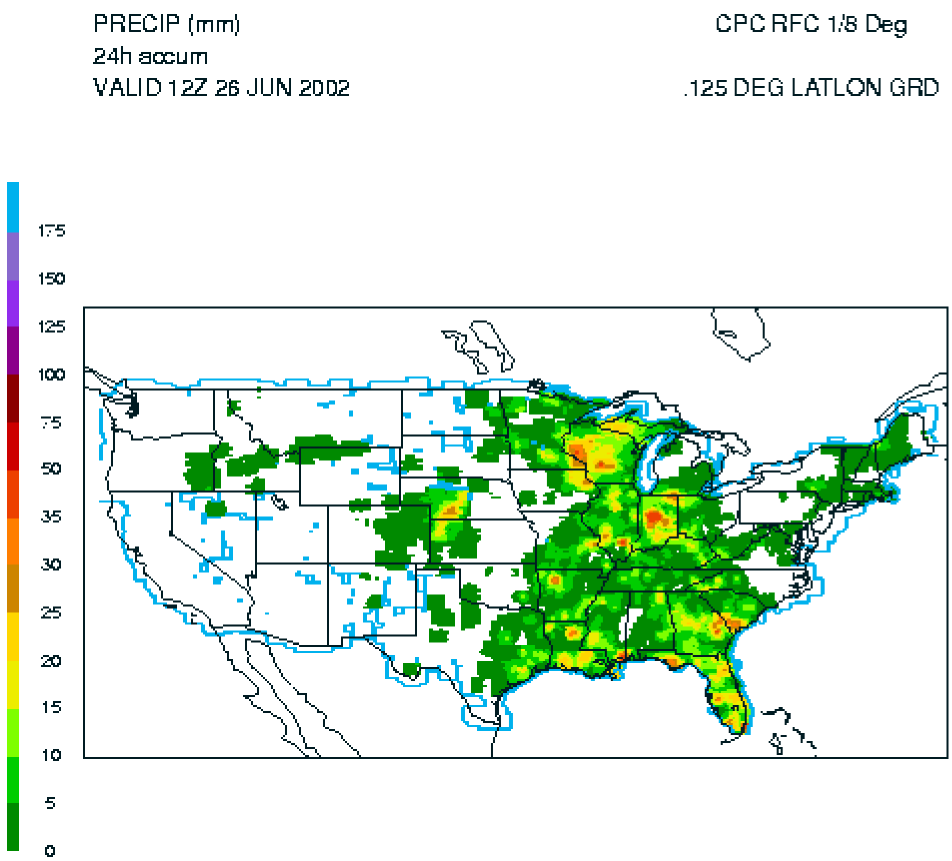


Figure 10. Three different analyses of *observed* precipitation accumulated *over 24 hours*, no model forecast involved. With these *"observations"* so different, how can we possibly expect a model forecast to predict the precise location and timing of heavy convective precipitation on the 5-km National Forecast Digital Database grid?





During the “verification” period shown in Figure 10, one run of the Eta model predicted a precipitation blow-up near Omaha, where none was observed. It even predicted a transition toward a stratiform region, along with a mesohigh and wake low, as shown in Figures 11-13. Forecasters will be challenged with needing to identify which physically realistic, coherent features will verify and which will be spurious.

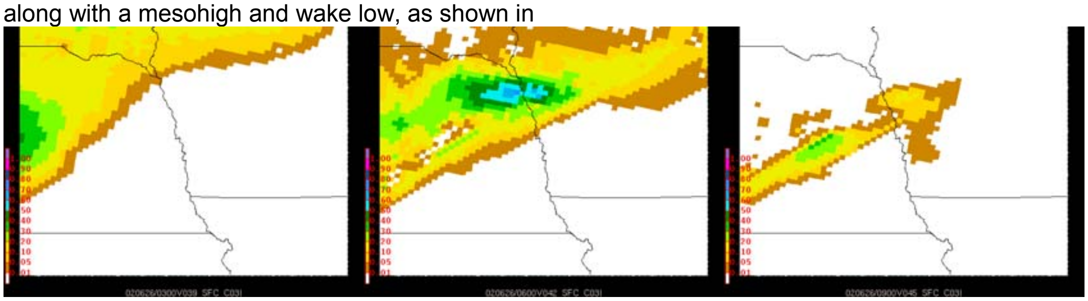


Figure 11. 3-hour accumulated precipitation from the convective parameterization ending at forecast hours 39, 42, and 45.

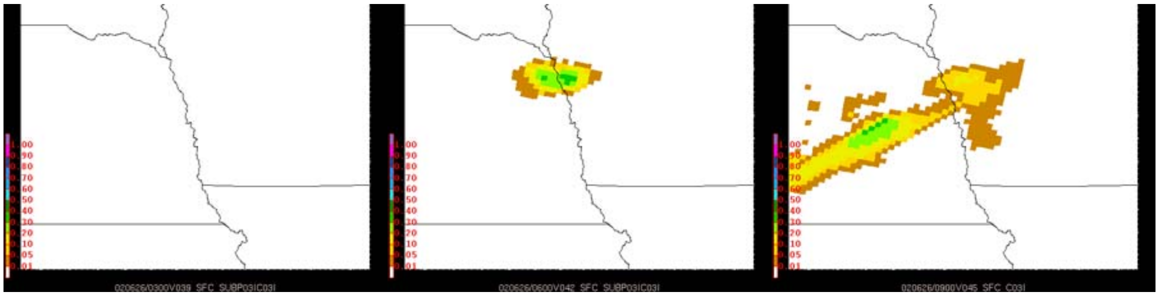
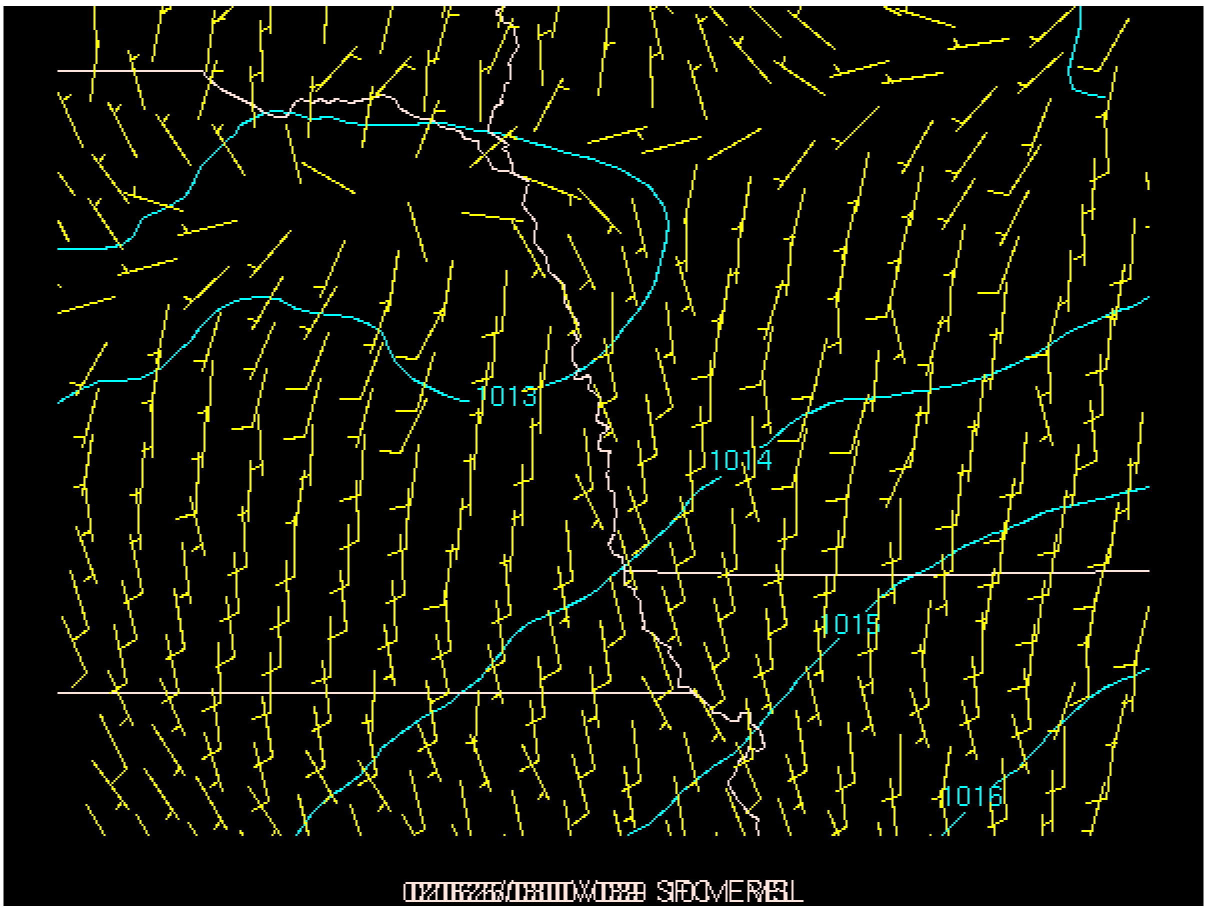
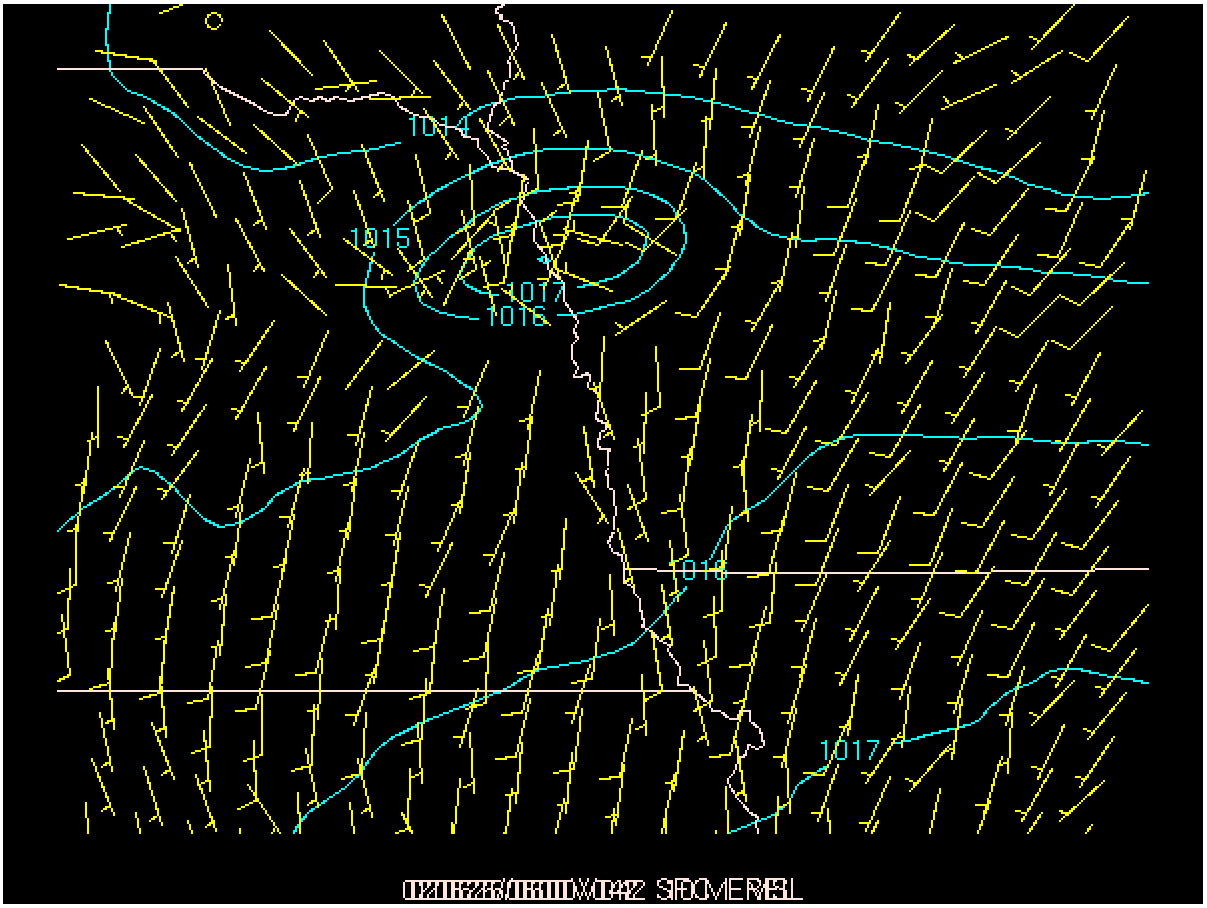


Figure 12. 3-hour accumulated precipitation from the grid-scale parameterization ending at forecast hours 39, 42, and 45.




1014
1015
1017
1016
1018
1017

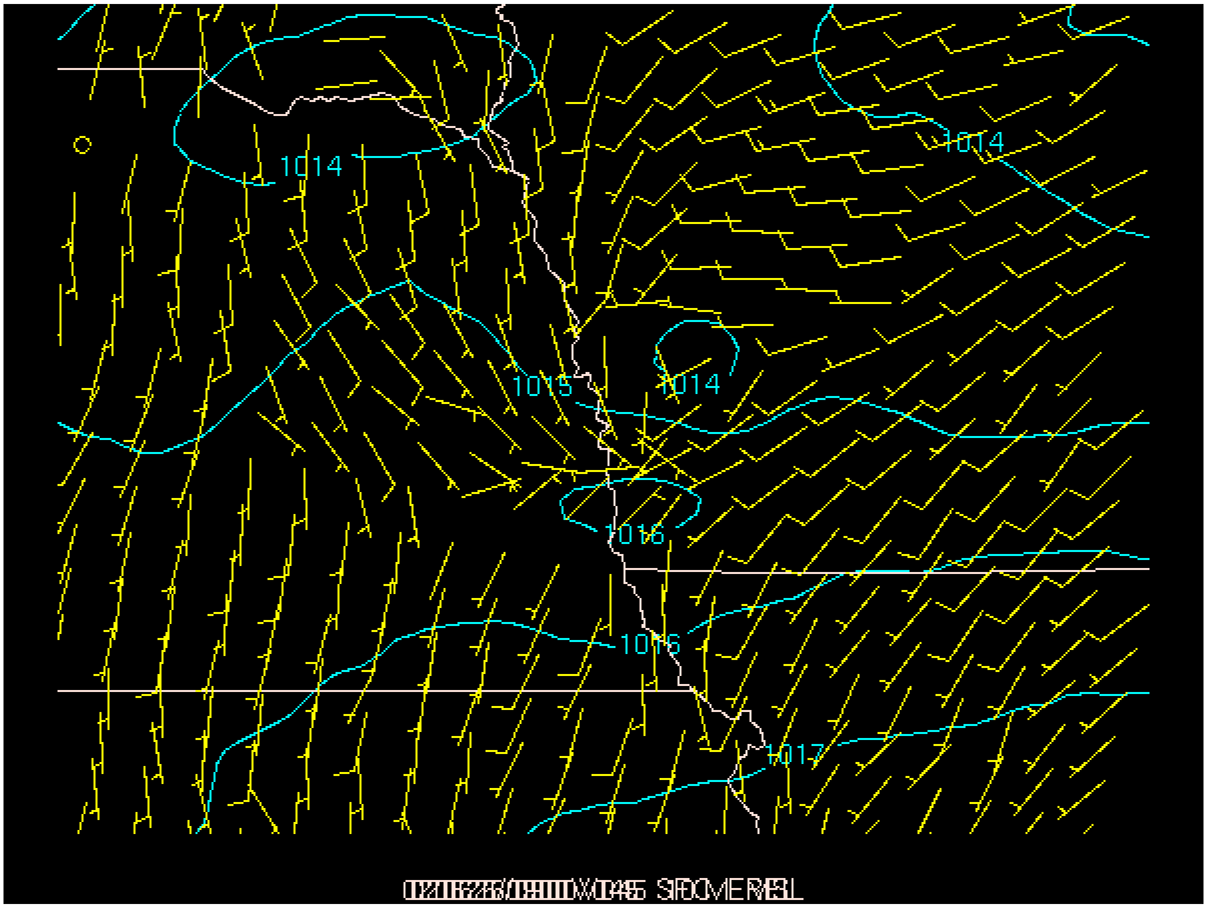


Figure 13. Winds at 10-meters and sea-level pressure from forecast hours 39, 42, and 45 showing synoptic low over northeast Nebraska, mesohigh created by the spurious convective system event, and finally even a wake low it produced over western Iowa.

6. DOWNSCALING

Models run at finer resolution over limited domains are being increasingly used to downscale forecasts for local effects. However, as shown by Mass (2002) and others, the skill improvement for resolutions finer than 10 km is questionable. NCEP has downscaled the Eta model from 12 km to 8 km using the NCEP Nonhydrostatic Mesoscale Model (NMM) with error scores from fits to radiosonde observations showing mixed results.

Some of the challenges for downscaling include:

- Different vertical coordinate than parent model
- More detailed topography with deeper valleys requires filling part of the column – what lapse rate should be used?
- Initial balance issues

Verification of these runs also presents challenges, both from lack of sufficient observations and from the fact that a useful, detailed forecast at full observed amplitude of some weather feature but which is off in placement or timing will yield worse pointwise error statistics than a watered-down forecast of lower amplitude. More meaningful assessment tools are needed.

The question is, how is the forecaster to use this higher-resolution nested model? Sometimes the forecast clearly adds value, but sometimes it does not or it is difficult to assess. For instance, the NMM and Eta topography over northwest Wyoming and adjacent portions of Montana and Idaho are shown in Figure 14 and their corresponding forecasts of 10-meter winds, one during the day, the other at night, are shown in Figure 15 and Figure 16 along with observed winds. The errors at the station locations are similar during the daytime, indeed more error appears in the NMM winds at some stations and overall it is not clearly a better forecast, but at night it performed much better with cold air drainage flows.

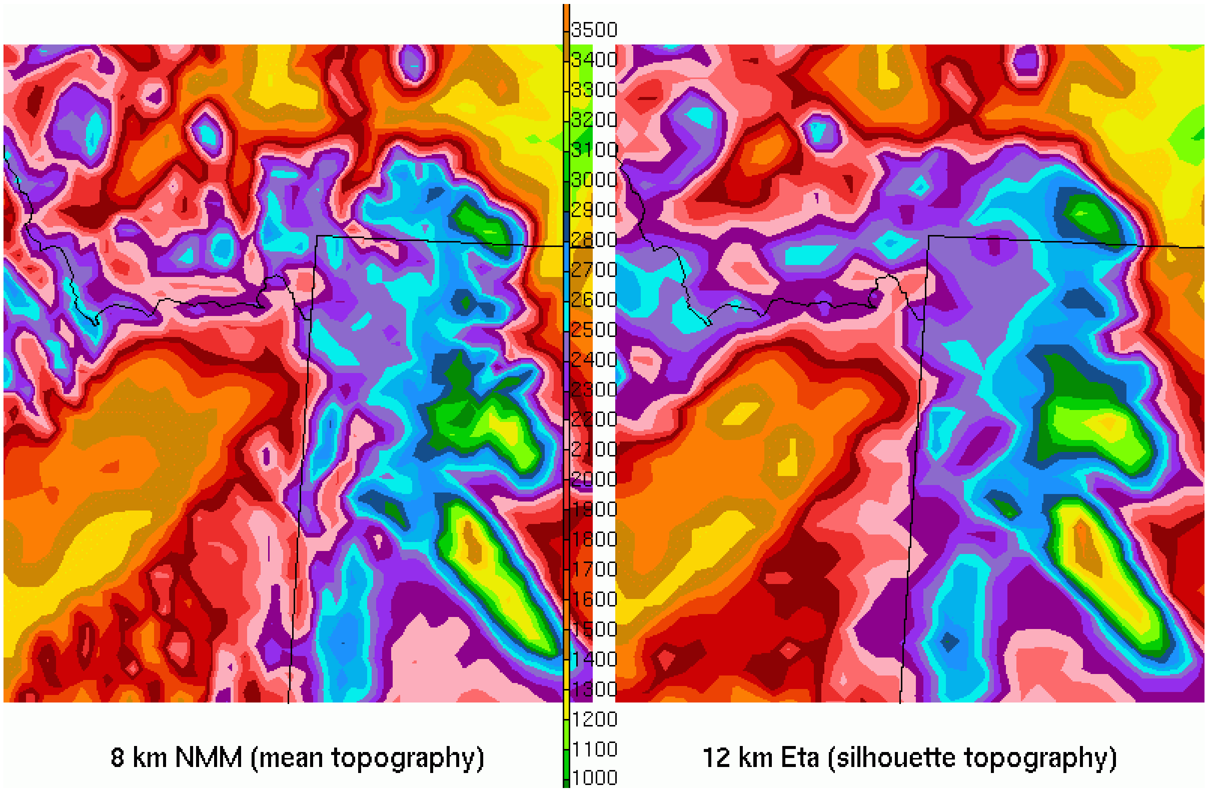


Figure 14.

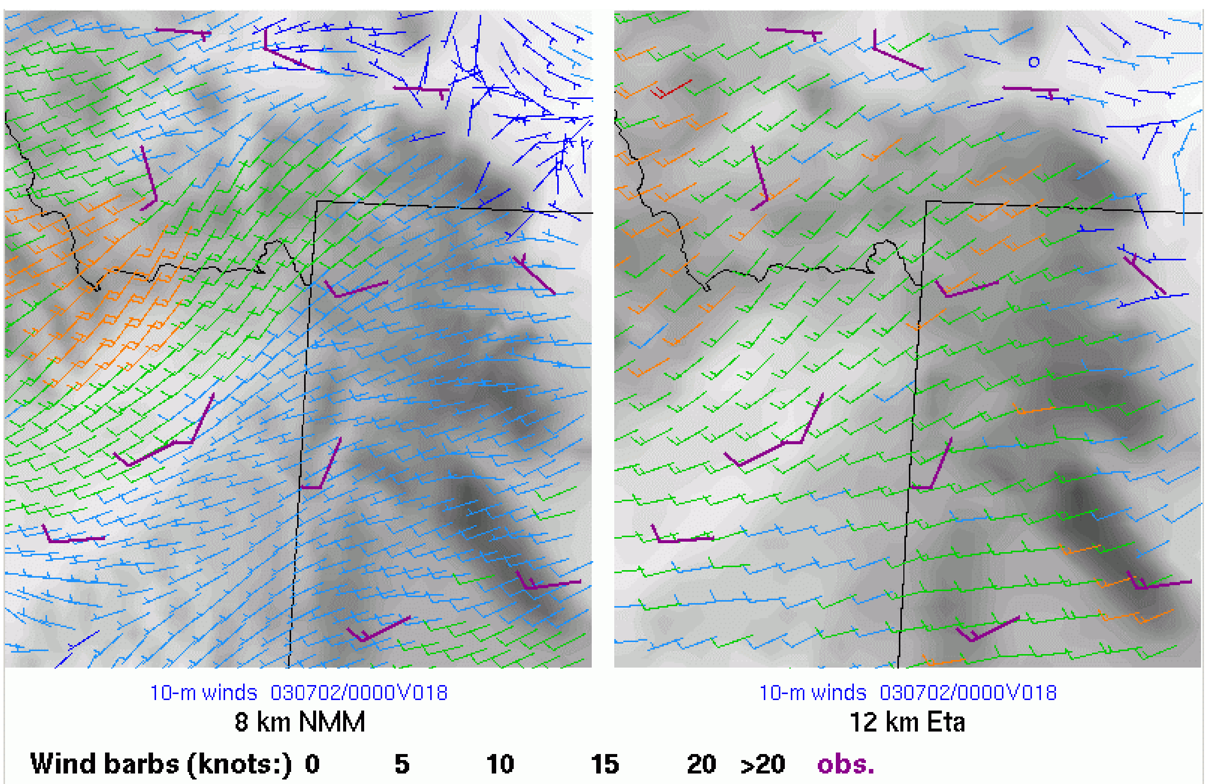


Figure 15. Model 10-meter forecast winds color-coded by speed. Observed winds shown in enlarged purple barbs.

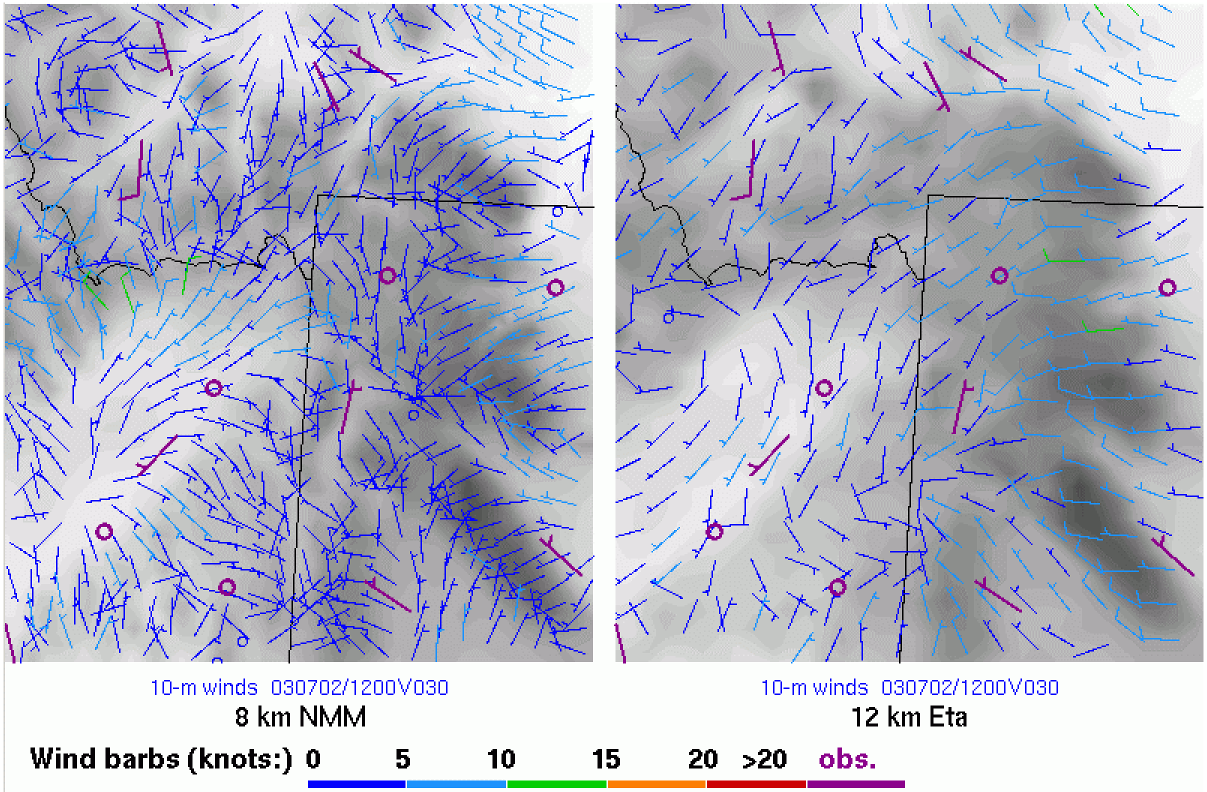


Figure 16. Model 10-meter forecast winds color-coded by speed. Observed winds shown in enlarged purple barbs.

7. CONCLUSION

Now and even into the foreseeable future, models and their analyses will have specific shortcomings that a human forecaster can identify and correct for to improve upon the model forecast.

8. ACKNOWLEDGEMENTS

This paper is funded by cooperative agreement #NA17WD2383 from the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

9. REFERENCES

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep convection. *Mon. Wea. Rev*., **131**, 2394-2416

Fowle, M. A. and P. J. Roebber, 2003: Short-range (0-48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782-794

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc*., **83**, 407-430

Xue, M., D. Wang, J. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather predication and data assimilation. *Meteorology and Atmos. Phys*., **82**, 139-170