<u>The Top Ten Misconceptions About NWP Models</u>

(Teletraining session given on 4/13/01)

Talking Points

<u>Slide #</u>

1) Title page

2) Introducing presenters and ways in which we will be acting as sort of a liason between NCEP and the field

3) Purpose of teletraining:

   1. point out some common but operationally significant misunderstandings of how NWP models work
   2. explain a little to dispel these misconceptions
   3. motivate use of the COMET NWP PDS

   Please refer to the NWP PDS for more information on the topics discussed in this teletraining and for other questions you have about how NWP models work!

4) Categories of misconceptions covered here

   - Model physics, to a large extent, is a crude estimation of what is going on meteorologically within a grid column and inside the grid square at the surface. This is NOT done well. As we get down to finer and finer resolutions, model physics will become more and more important.
   - Better resolution does not always mean better forecasts. Physics is one of the big reasons why.

   - Data assimilation is seen as an even more mystical black box within the perceived black box of NWP. We hope to remove at least one misconception about DA with this presentation.
   - Dynamics misconceptions: Misunderstandings about how well a model can resolve and predict atmospheric features at specific resolutions fall under this category.

   - Post-processing misconceptions: How does the model come up with its forecast output? We will look at three cases: information content in AWIPS grids coarser than actual model resolution, model output statistics (MOS), and the "predicted" 2-m temperature output from NWP models.

5) Question to engage participants

6) Common misconception is that AWIPS grids at less than full model resolution are missing valuable forecast information. Actually the full-resolution grids contain considerable noise as well. See page 58 for explanation of issues about AWIPS grids

7) Legend for pictures in page 8. Note that issue isn't whether the grid is 20, 40, or 80 km but how that compares with the native grid the model is run on (i.e. 1x, 2x, 4x for this case)

8) Frame 1 is on 20 km AWIPS grid#215:

   - Differences mostly due to channeling in valleys and vertical shear low level winds are picking up influence of winds at different heights. Circulations such as mountain-valley and sea breeze of small scale topographic features included in model terrain and San Francisco Bay (around 3 or 4 grid boxes by one grid box) cannot be resolved even at this scale and should not be relied upon for accuracy.

- Vorticity field very noisy but main bulls eyes features around the peaks and bends in the Sierras and the San Gabriels and edges of the Central Valley are probably correct, if the large-scale flow is correct.

- Notice noisy low-level wind variations over Nevada where terrain is not as great a factor but there may be waves, the specifics of which will be poorly forecast, propagating off the Sierras which are influencing these details in model forecast. Frame 2 is on 40 km AWIPS grid#212:

- Main features such as bulls eyes features around the peaks and bends in the Sierras and the San Gabriels and edges of the Central Valley are the same as on the 20 km grid

- Still noisy over Nevada but overall not quite as bad as on 20 km grid. The details creating the noisy vorticity pattern is still numerical noise at 2 times the model grid spacing.

- Valley circulation and flow channeling and winds on peaks not as sharp but still present Frame 3 is on 80 km AWIPS grid#211

- All fields are washed out, all useful detail and all artificial noise is lost

- The washed out nature is due to applying smoothing filters. For absolute vorticity on the 211 grid and 104 grid, two passes of a 4th order filter are applied to the native grid, interpolation is done to the AWIPS grid, and 4 passes of a 25 point filter are done on the AWIPS grid. Other variables have different but similar filtering. This smoothing is a legacy of the 80 km Eta which replaced the LFM years ago. Forecasters said they needed the smoothing to use QG diagnostics on the model fields which otherwise were too noisy.

9) Question to engage participants

10) A common misconception is that as the models are upgraded, MOS should improve. Actually, MOS applies a statistical relationship between model forecast fields and observed weather; if the model improves, these relationships may change. The more cases that go into the statistical relationship, the more reliable the relationship is likely to be. Therefore, frequent model changes tend to produce less robust MOS equations.

11) MOS finds the best statistical relationships between model synoptic forecast variables and observed weather.

- Scatter is inherent in any statistical relationship

- Less scatter means the relationship holds better.

- The relationship removes overall bias but not regime-dependent or situation-dependent bias

- MOS applies smoothing to model fields used for predictor variables to avoid noise and use the synoptic-scale signal. Therefore, MOS is designed to not "see" smaller-scale features predicted by the models as resolution improves. NGM MOS uses 5, 9, or 25 point smoothers, different for different predictor variables, applied to a 190 km gridded output product. AVN/MRF MOS uses a 25-point (5x5) smoother on a 95 km grid.

12) An adequate sample size is required to make the MOS statistical relationships

● For rare events such as low visibility and precipitation (all the more so for thunderstorms!), data at many stations are combined into one sample.

● Regions for creating samples for MOS visibility equations are shown on these two frames for winter and summer.

● Regions group stations with similar climatology, where the statistical relationship may be similar. Note that the regions in winter and summer are divided differently.

13) MOS forecasts will be poor if the statistical relationship isn't appropriate for today's weather, such as if

● today's regime was not common during the period used for creating the MOS equations

● the model changed since the MOS equations were developed.

● particular circumstances are handled differently than normal for similar cases such as if you usually have a snowcover with an arctic outbreak but not today.

14) Question to engage participants

15) A common misconception is that the model surface conditions (2-m T, Td, 10-m winds) are directly predicted by the model. Actually the model approximates these based on its forecast of the skin condition and the forecast of the lowest model layer, which may be 60 meters or thicker.

16) 2-m air temperature is approximated based on skin temperature and lowest model atmospheric layer temperature

● Skin temperature is predicted from surface energy budget involving radiation, soil, and vegetation and is subject to errors in radiation and surface physics parameterizations and limitations in assignment of ground conditions (vegetation cover, soil moisture, etc)

● In the AVN, a logarithmic weighting is used, following the dashed curve, accounting for the ground during the daytime/nighttime being much warmer/cooler than the air. An example value using a linear interpolation is shown for comparison.

17) Depth of model's lowest layer affects how representative the layer air temperature is used in diagnosing the 2-m temperature.

● Lowest layer thickness in 50-layer Eta model varies from 2 hPa at sea level to around 10 hPa for model surface elevations of around 400-500 meters to more than 20 hPa for grid boxes containing model terrain above 3000 meters. Thus the representativeness for diagnosing 2-m temperatures is very good over low terrain and very poor over high terrain.

● 42-layer AVN/MRF uses terrain-following sigma coordinates with a lowest layer thickness of around 8 hPa, or around 70 meters

18) What can the model use for 2-meter temperatures in areas of locally steep terrain like in the illustration?

- Height of the model 2-meter temperature may be very different than the height of any station(s) in that model grid box, especially for valley stations typically located well-below the model terrain.

- If the model fields were extrapolated to an actual station elevation, that might be inconsistent with the model calculations of surface energy balance and overlying airmass stratification for instance, cold air trapped in a valley and valley fog may not be able to be represented in the model with its topography.

19) Question to engage participants

20) A common misconception is that models have good equations for radiation, suffering errors in incoming radiation forecasts due to errors in predicting cloud cover. Actually, radiation is a very complex molecular-scale process, and properly parameterizing it would require more computer time than running the whole rest of the model combined.

21) Clear sky (no forecast or observed clouds) radiation bias in Eta model from July 1999 over northeast U.S.

- Horizontal axis is solar insolation reaching the surface predicted by the Eta model

- Vertical axis is solar insolation reaching surface as measured by GOES.

- Model error is the horizontal displacement of the data points from the diagonal line.

- The average bias is 78 Wm-2 corresponding to a skin temperature bias of around 6^{o} C. A few points even have an error of up to 300 Wm-2!

22) Eta forecast (horizontal axis) and GOES observed (vertical axis) skin temperature, for cases for which radiation plotted in previous picture.

- Skin temperature indeed has a warm bias, as expected with too much solar radiation reaching the surface.
- However, the warm bias is only 4.3^{o} C, smaller than would be caused by the radiation bias
- This indicates a compensating error in surface physics. If the surface physics alone were improved, the skin temperature would be even warmer, degrading the model forecast!

23) Question to engage participants

24) A common misconception is that the model will make a reasonable forecast of the time and location of convection if it gets the large-scale forcing and pre-convective forecast soundings correct. Actually, different convective parameterization schemes will often produce wildly different patterns, timing, and amounts of convection given the same conditions, because the schemes have different trigger functions, different links to the large-scale, and different ways of handling redistribution of water including how much falls out as precipitation.

25) Two model runs using identical initial and boundary conditions:

- Lower panels show precipitation predicted by Eta model runs using the Betts-Miller-Janjic (BMJ) scheme and Kain-Fritsch (KF) scheme.

- Upper two panels show verification, based on RFC gauge analysis and based on gauge-adjusted radar estimates (multisensor analysis).

- Note the remarkably different patterns in the two Eta forecasts and that each performed better than the other in some locations.

26) These are grid-point soundings from the western Florida panhandle, where both the BMJ (green sounding) and KF (blue sounding) schemes produced convective precipitation in the Eta runs shown on the previous page. The model integrations began at 00 UTC.

- Frame one: 18 UTC (18 hour forecast). BMJ scheme is already convecting in the model, producing a sounding resembling a BMJ reference sounding. It has a deep layer near moist-adiabatic but not saturated, eliminates saturated layers in cloud and cools the lower part of the cloud, in this case down to cloud base around 950 hPa. The KF scheme is capped and not yet convecting, with a surface-based mixed layer, so it is warmer.

- Frame two: 21 UTC. (21 hour forecast) KF is now convecting too. The KF scheme should cause deep drying, however the model dynamics are responding to the heating distribution, pumping up moisture to make a deep moist sounding above 700 hPa, leading to some grid-scale precipitation in addition to the convective precipitation. Also, the KF scheme has downdrafts, which have stabilized the boundary layer, though after earlier sunshine, it remains warmer than in the BMJ run. And the winds are noticably different as a result of different dynamic responses to different horizontal and vertical distributions of heating. Nonetheless, note the remarkably similar temperature and dewpoint predictions in the 500-300 hPa layer.

- Frame three: 00 UTC (24 hour forecast) Convection has ended in both and synoptic advection of upstream conditions is dominant. KF remains warmer in the lower troposphere, while BMJ has dried out more aloft due to stronger advection (much stronger winds).
- Frame four: 01 UTC (25 hour forecast) Synoptic advection continues. Notice how similar the temperature profiles are above 600 hPa yet how different the dewpoint and wind profiles are.

27) 30 km x 30 km subset of a model run at 1 km resolution (not same case as in previous two pages), showing boundary layer rolls

- Observational studies have shown boundary layer rolls sometimes are the determining factor in convective initiation

- Convective parameterizations have to infer information about such subgrid-scale processes which initiate convection in real life. However, there is little generally reliable basis possible for this inference using only information from one sounding representing the average over this whole area!

- How can a 30-km resolution model possibly "know" the effect of this level of detail, which it needs to predict where and when convection will occur?

- Convective nowcasting is difficult even with a proximity sounding!

28) Question to engage participants

29) A common misconception is that convective schemes are designed to make good QPF forecasts. Actually, convective schemes are used to remove instability in order to prevent the model from making a grid-point thunderstorm (the size of an entire grid box). Precipitation is produced by whatever water the scheme happens to squeeze out in the process.

30) Frame 1: Convection in nature.

- Buoyant updrafts strong and cells develop rapidly

- Updraft rapidly moistens entire troposphere up to cloud top but covers only a fraction of the grid box, with subsidence warming/drying outside the cloud in remainder of the grid box

- End result, after some downdrafts and convective dissipation, is a stabilized sounding, typically moister in the upper troposphere and drier in the low-mid troposphere than previously.

Frame 2: Convection in a model without a convective parameterization (or with a convective parameterization which is not doing enough to prevent grid-point thunderstorms) ["PCP"=Precipitation and Cloud Parameterization, same as grid-scale precipitation scheme]

- Cloud builds up slowly from small grid-scale vertical velocities. In a hydrostatic model, this vertical motion is only indirectly forced by buoyancy .

- The cloud fills the entire grid box, which becomes saturated through the depth of the slowly ascending cloud

- Heavy rain occurs over the entire grid box

- End result is a nearly saturated moist-adiabatic sounding

- Too much latent heating is deposited in the lower troposphere, creating low-level cyclogenesis, which feeds back to make further grid-scale convection. The grid-scale vertical motion may also have generated a mid-level vorticity maximum and advected low-level vorticity upward. The result looks like "convective feedback" but is caused by the model convecting on the grid scale instead of through the convective parameterization. The primary purpose of convective parameterizations is to ensure that this scenario does not occur!

31) The BMJ scheme is an adjustment scheme, rearranging heat and moisture to relieve the instability, with convective precipitation as the excess moisture after the adjustment.

- It imposes a (case-dependent) reference temperature and dewpoint profile (blue curves in skew-T diagram).

- Moisture is transported upward by changing the original sounding to the reference sounding.

- The precipitable water is reduced (higher RH in colder layers, lower RH in warmer layers),

- This change in precipitable water is the convective precipitation produced.

32) A mass flux scheme such as KF, Arakawa-Schubert, and Grell, assumes an exchange of air between the unstable layer and the middle to upper troposphere, resulting in rearranging some heat and moisture in the grid column. Precipitation is squeezed out through parcels assumed to ascend from the unstable layer, with some falling precipitation possibly evaporating.

- The grid column sounding is assumed to be affected by detrainment (green arrows) from sub-grid scale updrafts originating at low levels (small red arrows), by sub-grid scale downdrafts stabilizing the low levels (blue arrows), and by compensating environmental subsidence (thick red arrows).

- The amount of convective overturning depends on how much is needed to stabilize the environment (such as eliminating CAPE, reducing CAPE to a certain amount, or counteracting grid-scale destabilization different formulations for different schemes)

- A 1-dimensional cloud model is used to calculate properties of ascending parcels, including water content available for precipitation. It may include some entrainment and it may include some simple microphysics considerations.

- Precipitation per unit of updraft mass flux is whatever is squeezed out by an ascending low-level parcel and not evaporated on the way down.

- The total amount of convective precipitation depends on how much convective overturning (updraft mass flux) is needed for stabilizing the sounding.

33) A common misconception is that surface conditions are updated daily and reflect the actual ground condition. Actually, for the Eta model in early 2001:

- Soil moisture is based on cycling of model forecast precipitation rather than soil moisture measurements

- Vegetation type is prescribed to the prevailing type based on remote sensing. Suburban sprawl may cause massive clearing of previously vegetated areas, and other human and natural forces may be changing the landscape since the dataset was created.

- Greenness fraction is the fraction of a grid box covered by active vegetation. For deciduous forest, it is high in summer and low in winter. The Eta prescribes greenness fraction based on monthly climatology derived from remote sensing rather than the current conditions.

- Snowcover is updated daily but the 12 UTC daily NESDIS snowcover and Air Force snowdepth data arrive in time for the following 0 UTC model run. So a 12 UTC run will be using yesterday's snow analysis any snow falling since then is cycled from the model forecast. The snowpack water content is derived from the snowdepth analysis, not observed directly.

- SST are updated daily using the new Ocean Modeling Branch Real-Time Global Analysis, which is a 2D-VAR analysis incorporating satellite, ship and buoy data.

34) Annual cycle of monthly greenness fraction used in Eta model.

- Winter wheat belt in central KS/OK greens up by April and is harvested, leaving brown conditions by mid-summer

- Corn belt IA/IL/IN/OH shows bare ground in May and high greenness fraction in July.

- If the harvest cycle is earlier or later than usual, then this greenness fraction will not reflect current conditions.

35) Illustration showing different surface energy balance before and after harvest, with more solar energy being used for latent (sensible) heat flux before (after) harvest.

36) Results from one-dimensional column model using AVN/MRF physics parameterizations for the same conditions except differences in greenness fraction, effectively simulating preharvest and postharvest conditions.

- Note the very large effect on surface temperature! The daily maximum skin temperature differs by 6ºC and the lowest atmospheric layer temperature differs by 5ºC, with corresponding differences in boundary layer depth and mixing out of low-level moisture.

37) Question to engage participants

38) A common misconception is that the data assimilation system should produce an analysis that looks like a best fit to the data or how you would draw a hand analysis. Actually, the model analysis needs to be consistent with the model resolution, dynamics, physics. Therefore, instead of directly analyzing the observations, it uses observations to make small corrections to a short-range forecast. Also, the model ingests many times more data than just radiosonde and surface observations and has to reconcile all of these observations, which may not all agree. Finally, the model has particular difficulty being able to make effective use of single level data such as surface observations.

39) The purpose of the model analysis is to produce initial conditions leading to the best possible model forecast rather than to represent the current state of the atmosphere as accurately or in as much detail as possible. This has important ramifications on how data must be used. Following the illustration containing three grid columns and many layers,

- Observations may sample features the model cannot resolve, such as the radiosonde balloon ascending through the snow shower and the surface observations in and around the cumulonimbus. Perfectly fitting these would degrade the model forecast.

- Observations may sample deep layers coarser than what the model can resolve, such as the satellite shown receiving radiation attenuated by and emitted from many grid boxes in the column

- Problems are created with inconsistency with model physics limitations, such as good humidity observations causing the model's imperfect cloud parameterization to produce too much or too little cloud cover.

- Many grid boxes have no data but may be influenced in the analysis by many data at other levels and locations.

- The analysis must combine and reconcile all of the data, including the three surface observations in the right grid column, one pre-convective, one post-convective, and one under the rain shaft.

- Good observations all have some error, with some observation platforms more accurate than others

- Vertical structure is essential to retaining a weather feature in the forecast. Vertical structure has to be assumed instead of observed in the case of single level data such as surface observations and cruising-level aircraft data. This greatly reduces the model forecast benefit of using such data.

- Both mass and wind information are needed together because ageostrophic flow which

  requires knowing the mass and the wind fields is central to the dynamics of any weather feature of forecast interest and required by the model to make a good forecast. A large region of satellite winds with no corresponding temperature data or of satellite radiances with no corresponding wind data has far less value than a radiosonde network sampling both winds and temperatures. These factors all require some treatment other than precisely fitting each observation. More discussion of the limitations of surface observations is given on slide 68.

40) In order to address this range of issues in handling data, the model actually makes corrections to
a short-range (3-hour for Eta, 1-hour for RUC, 6-hour for AVN) forecast instead of simply analyzing the data. The illustration shows the analysis correcting a series of short range forecasts back toward reality as best as can be determined from the observations. This has the advantages listed on the slide, if the short-range forecast is decent.

41) This shows how analysis corrections to the short range forecast ("first guess") compare to what the observations show.

- In most locations, the corrections indicated by the observations (red) and the corrections actually applied in the analysis (black) are in the same direction but the analysis corrects only around half as much as necessary to match the radiosonde observations.

- Where the analysis corrected in a different direction than the corrections indicated by raobs, the corrections are mostly under 5 knots, so overall analysis and raob disagreement in those places is not large

- This is a worst-case scenario (from 12 UTC January 24, 2000 east-coast forecast bust) because the first guess was particularly bad indeed so bad that the Peachtree City observation was rejected, resulting in a "correction" worsening what was already a 50-knot discrepancy!

- A more typical analysis would have smaller corrections and smaller discrepancies but would still undercorrect.

- Model analyses and model forecasts are subject to largest errors when the previous short-range forecast is poor and in rapidly changing situations, because the analysis does not fully catch up to the observations.

42) Question to engage participants

43) A common misconception is that resolution is a panacea for NWP models. Actually the models require a synergistic interaction between all parts of the model, including physics, numerics, resolution, and data assimilation. If any of these are not compatible with the others, forecast quality suffers. Improving resolution and corresponding model topography alone, while benefitting prediction of orographically forced precipitation and local terrain-induced circulations, may otherwise help far less than you might have expected. This is why modeling centers like NCEP invest resources in all parts of the model together.

44) Example where better resolution mostly buys stronger sensitivity to physics

- AVN made much better QPF prediction than operational Eta, despite Eta having around 4 times better resolution.
- Experimental Eta with better SST analysis also made good forecast
- As models go toward higher resolution, they will be more sensitive to lower boundary treatment and other physics. This may be closer to nature's sensitivity, but it may also degrade forecasts unless physics are upgraded to keep pace with resolution changes.

45) Example where high resolution buys a very useful, immensely improved forecast but not perfect

- Best case situation precipitation produced by large-scale flow forced over fine-scale topography
- Forecast useful because it correctly indicates flood potential and even which basins may flood
- Actual amounts well short and rainfall peaks displaced to windward side, as well as missing the peak in southwest Los Angeles County.

46) Example where 3 km resolution buys good sense of variability and deceiving realism, not accuracy!

- In big picture sense, forecast of convection patches and coverage is amazingly good! This could be very useful for making zone forecasts if there were more lead time.
- On a storm-by-storm basis, cell speed and longevity, clustering, and new development are almost all wrong. This would not be useful for warnings and of limited use for severe weather statements.
- This model simulation used radar data once a cell already existed. Forecasts from earlier times were unable to correctly predict convective initiation.

47) Just what it says!

48) A common misconception is that the model can resolve weather features of twice or four times

the model grid spacing. Actually, the model requires at least around five grid boxes to capture the shape of a feature and at least around ten to make a good prediction.

49) A wave 4 grid boxes across does not have its shape OR amplitude captured.

50) The wave is advected a distance of one wavelength, getting worse during the forecast.

- Amplitude may be affected
- Phase is badly affected in underresolved features

- Numerical dispersion creates leaves a trailing wake of little waves

- Exact details of phase shift, amplitude retention, and dispersion will vary by numerical scheme. However, the 4Dx wave will start off poorly and degrade during the forecast with any scheme, while the 10Dx wave will be well-represented and well-predicted.

51) The wave depiction deteriorates in the forecast because the gradient is not adequately represented

- Forecast equation for simple linear advection is wind speed * gradient

- Undersampling (as when feature not adequately resolved) usually causes the gradient to be too weak in the model for instance, compare slope of green line (numerical gradient) to slop of red curve (actual gradient)

- If the gradient is too weak, the tendency will be too small

- If the tendency is too small, the wave will lose amplitude during the forecast

52) Here's what an 18 hour forecast of a hurricane looks like in the GFDL hurricane model inner nest, which has 1/6 degree resolution.

53) Here's the same hurricane on the outer nest of the two-way nested model, at 1 degree resolution same resolution as AVN/MRF. This is the same sampling problem as dealt with in misconception #10 about AWIPS grids. Suppose instead we were to look at this on the AVN, which does not interact with a finer nest. How would we expect it to be different?

- The hurricane would be poorly resolved, resulting in loss of amplitude during the forecast. Interactions with model physics may cause it to intensify, but not nearly as much as the same amount of latent heating, sea fluxes, etc. would if it were adequately resolved

- The hurricane development would be further stymied because of the nonlinear feedback between the radial circulation/latent heating/eye warming and the tangential circulation and sea fluxes. If these are underforecast, then their interaction and intensification will be underforecast.

- Thus, a 50-knot tropical cyclone in the AVN/MRF at 3 or 5 days corresponds to a pretty intense hurricane! If the model is making an excellent forecast for its resolution, you should not expect to see it predicting hurricane wind speeds!

54) CONCLUSIONS: What did you learn? Reflect on the lesson and let us know what was most

useful and least useful. Note link to NWP course!

55) Resource for keeping up to date on NCEP models

56) Another resource: mini case studies highlighting specific aspects of the models

57) Our emails: be sure to let us know if you have questions/concerns about the models or feedback on this teletraining!

END OF PRESENTATION
extra slides for extra info (Q/A, etc) below

58) AWIPS grids issues (text)

59) Same as for frames on page 8, but for midwest and only shows 20 km and 40 km grids. Again, information content essentially same at 20 km and 40 km.

60) 500 hPa absolute vorticity on 20 km, 40 km, and 80 km grids. Same message for mid-troposphere fields as for surface fields on pages 8 and 59: 20 km contains signal and spurious noise, 40 km has full detail, 80 km AWIPS fields are too smoothed, losing important gradient information.

61) MOS assumptions (text)

62) What violates MOS assumptions (text)

63) Nearest grid point may not be representative of local conditions due to model surface type (such as land vs. water) and topography. A neighboring point might be better, a consideration when selecting a grid point to view model soundings.

64) Text for points to consider with pages 15-18

65) Text for points to consider with page 20

66) Text for points to consider with pages 21-22

67) Text for points for page 33

68) Why surface observations have limited usefulness for models (text)