Verification testbed

- Would have a lot of benefit for the community
- How this would work is not yet clear
- Need to determine whether verification developers can implement things in MET on their own, or need to work with DTC

Spatial methods

- A number of spatial methods are ready for inclusion in MET, including
  - Composite
  - FQM
  - Scale separation methods
  - CRA
  - Cluster?
  - SAL
  - Other?
- MET already includes MODE and the neighborhood methods; intensity-scale has been implemented recently

Ensemble and probability forecasts

- Initially implement standard scores.
  - Including Wilson score, ROC, reliability.
- Include information for quantifying spread-skill relationships, but give a lot of warnings about interpretation
- Spatial methods for ensemble forecasts generally are not ready to be implemented – more research is needed
  - Can help with arguments vis a vis use of ensemble mean

Inference

- Confidence intervals may be preferable (but essentially equivalent) to hypothesis tests
- Application of many of the methods is a fairly complex process, and requires users to understand a number of assumptions
  - If possible, find a way to give users warnings about making “bad” decisions

Inference cont.

- Add the capability to compute CIs of differences
- Education and training are needed on verification methods and interpretation (not just how to turn the crank with tools like MET)
  - Disseminate this kind of information to operational and other users

Observation uncertainty etc.

- Can have large impacts on verification statistics and should be taken into account; this uncertainty arises from many different sources
- Although a number of ideas have been proposed regarding how to do cope with this error, a coordinated approach has not been established
  - There is a great need for good information about observational errors
- Data assimilation can provide limited information about observational uncertainty, relevant for a particular model

Observation uncertainty etc.

- Satellite data have promise for helping with verification of a wide variety of phenomena
  - We need to understand the error characteristics of these before we use them extensively (on the other hand, AFWA has been using these data for a long time!)
- Ian's question (to keep in mind)

  Do you really want to forecast the truth, or the observation?
- Wavelet methods can incorporate observation (analysis) uncertainty information directly in the evaluation of the forecasts

Cyclone tracks, extremes, and standards of comparison

- Automatic tools are available for identifying cyclone tracks – NCEP applies these to model fields and verifies using the analysis field
- Extremes
  - Special approaches are required for bivariate extremes – require adherence to assumptions
  - Extreme dependency score should be implemented in MET
  - Also implement other more traditional methods (already there…)
  - EVT has promise for leading to methods for evaluating forecasts of extremes
  - Also consider extreme *errors*

Cyclone tracks, extremes, and standards of comparison

- Standards of comparison
  - Skill scores are essential part of what we do
  - Appropriate standards depend on user and application
    - Skill is highly dependent on accuracy of standard
  - New methods (e.g., spatial forecast verification methods) also need some kind of standard
  - Need a reference for how difficult a day was

Thanks to all of you for your participation and help!