

Ensembles - roundtable

1. Glahn - to see whether there is anything there with ensembles to help build MOS system. To compare with standard MOS?
2. Joliffe matter of curiosity.
3. Claudia not interested in verification. Make sense of output. Understand pdf and use them for forecast.
4. Multianalysis, skill score on individual and how well each member does as compared to the whole
5. Ensemble verification from tom Hamill. Theoretical framework. What is the EPS capable of and what we want
6. Stream flow forecast. Reliability and roc. Lack of historical observations, how can you use limited amount of obs and verify
7. Pdf forecast, different ways to produce pdf how different methods compare → verification issues?

Questions:

Question 2:

Are ensemble members representative of a pdf of the variable to be considered. Model error is not considered when define the members initial condition.

Verification in situation when bimodal situations are present.

Verification

quality
value

Predictability should be included in the verification

Having standard to define what the predictability.

Do we know the random variable pdf?

Small spread → calibrated centered on the observation

Modelers are interested I uncalibrated forecasts

Users are better off when given calibrated forecast.

Multivariate calibration: taking more variables into consideration and their interconnection.

Stratify dataset according to distribution. Talagrand diagram → what may be interesting is what happens at the either end of diagram.

Extreme events would be in the envelope but they are extreme members.

Question 3:

Individuals member verification useful is it a multi model ensemble.

Methods should differ → we are looking at single deterministic forecast.

Question 4:

Approaches to verify predicted pdfs:

Rank histogram

Marginal calibration: comparison unconditional forecast and observation distributions.

Spatial distribution of ensemble distribution. Events based verification.

Event based approaches on single members

Number of members in ensemble is important is verification. Small ensembles do not have enough variance.

Comparative verification needs ensemble size similar or large size.

You need many independent realizations → in short range poor man ensemble better than EPS

Question 5:

Predictability → when should deterministic be switched to probabilistic forecast

Spread large enough? Question unanswerable. Unless you look at long time or large sample → we go back to calibration. If you need large calibration then the spread is not good enough.

Training of users in using pdf and understanding verification of probabilistic products.

PANEL DISCUSSION

Question 6. How can the goodness of a product be expressed simply and in a manner that is intuitive to pilots? There is a request for confidence level information (for pilots) mentioned d' from SDT. Could we use a modified version of that?

- A. Mason: Could be done, but problem is different aspects of the forecast are more or less confident at different times, and what is important changes.
- B. Prob forecasts for FAA will not happen very soon.
- C. Could use box and whisker plots to convey confidence.
- D. Idea to quantify uncertainty using time-perturbed ensembles.
- E. Make recent verification information available graphically.

Question 21: Statistical significance in met verification - is it used or abused?

- A. (Schwartz) Both. Abused because of questions about validity underlying assumptions. Also underused.
- B. Nicholls paper: The insignificance of significance tests. Better to go to confidence intervals.
- C. Also Tressa's experiments.
- D. Confidence intervals more informative than hypothesis tests. Data tells one whether assumptions are appropriate. No null hypothesis is true. If you collect enough data, one will eventually reject it. Should say "doing tests at 5% level" rather than at "95% level"

Question 13: How can trust be built between forecasters/developers and verifiers?

- A. Need more guidance on feedback to developers.
- B. Be careful to explain how verification results were obtained.
- C. Involve forecasters in the design process.
- D. Have to talk with users before start developing.
- E. Collect subjective evaluation information along with objective results, compare - if trends similar, will build trust.
- F. Should add end user to question.

Question 4. How should verification of operational use of a forecast be related to meteorological verification of the forecast? Do users need both types of information?

- A. Yes. Should be related to verification of model. Question about how to do it.
- B. Paper at WAF conference on this topic. Should be on FSL website, and on AMS site.
- C. Previous work done on how forecasters make decisions. Some projects about this in Washington.
- D. Problem of finding out about related work going on in other fields.

Question 5. Allan Murphy that developing appropriate verification tools is just as hard as developing forecasting tools. Why, in the real world, is verification development always the "poor cousin" of forecast development?

- A. People think verification is easy. Also fear of verification.
- B. Much of what is needed isn't archived.
- C. Idea that verification is solved - RMSE etc. is good enough.
- D. Is verification important in Finland? A. Not really.
- E. Program manager for 8 years - asked for proposals on verification, didn't get any.
- F. Scary and painful for model developers to get hard cold verification information about their model. Goes back to question of building trust. Should also tell why model is lousy.

- G. Verification is not really a decision-making process - should be able to cut something. More complex than that, verification info can be misleading.
- H. Developers build verification into development process. Not fair to ignore this aspect.
- I. Lack of knowledge of statistics in verification community.
- J. Forecasts have to be issued operationally, and verification of usually not.

Question 12. There seems to be an overkill in verification of precipitation. Why not look at state variables? Is it useful? To whom?

- A. There isn't overkill: Precipitation is hard to analyze and to verify. Precipitation integrates all processes in model.
- B. User community is sensitive to other variables.
- C. Are other variables easy to verify? No. e.g. wind.
- D. Some disagreement that pcpn is easier or harder to verify than other variables.
- E. How to translate verification information back into changes in model?
- F. In Midwest, radar data useless in winter, data network not adequately supported. No way is there overkill of precip verification.