

# Working Group D

## Ensemble Verification Methods

## Interests of Participants

- Use of ensembles to build MOS systems/ compare with standard MOS
- Make sense of eps output. Know whether pdf from model is similar to observed. How to compare pdf's
- To score individual members/ how to compare to whole ensemble fcst performance
- Theoretical framework. Matter of curiosity. What is the EPS capable of and what we want from it.
- How to verify with small samples

**Question 1.** How do ensemble verification approaches differ from traditional approaches? Are there lessons from ensemble verification that could be applied in other areas?

- We rejected this question

## Question 2. What are appropriate approaches for measuring the adequacy of the spread in the ensemble forecasts? How can we evaluate whether the ensemble distribution appropriately represents the pdf of the random variable being forecast?

- Is ensemble pdf representative of true pdf of variable? Limits of perturbation strategy (initial state and/ or model) (Can't assess on one realization.
- Verification of bimodal and other non-normal distributions
- Verification at different levels: For modellers and for forecasters. Need different methods
- Should verification take into account the predictability (as standard)? How to quantify? (we didn't answer this)
- Definition of perfect ensemble: calibration first, then sharpness.
- Verify before and after calibration. Ideal to have multi-dimensional calibration over several interacting variables. (don't know how to do this)
- We decided not feasible to assess consistency directly. RH best we can do.
- THE ANSWER: Look at extremes of rank histogram.
- Another idea: composite ensembles over long period, compare with obs distn. Should they agree? (yes and no)

**Question 3.** Should the independent components of the ensemble runs be verified as well as the resultant run? If so, should the verification approaches differ between two?

- Yes, especially if a multi-model ensemble or ensemble where model is perturbed.
- Use for assessment of different model versions.
- Methods should differ - verifying a deterministic forecast.

Question 4. What are appropriate approaches for verifying predicted pdfs?

- rank histogram
- unconditional ensemble distributions compared to observation distribution
- possibility of extending spatial verification ideas to pdf's - event-based approaches
- Effects of different ensemble sizes - important to consider when comparing results on different ensembles
  - independent realizations
  - comparative verification on individual members

## Question 5. What are the main verification-related issues in this area? What are some approaches to solving them?

- Predictability:
  - at what projection to switch between deterministic and probabilistic forecasts; when to switch to climatology
  - use skill scores
  - caveat about value - may still be + when skill -
  - work in progress - Baumhaufner (error variance vs. ens spread)
- Issue of ens. mean use and verification
  - on giving information to fcstrs: box and whisker plots
  - extreme members to give range
  - ens mean verifies better on std scores, but not likely on spatial and/or diagnostic methods because of smoothing
- Training issue:
  - interpretation and meaning of verification methods
  - use of pdfs in forecasting