2020 IVMWO Metaverification Session

Time: 19 November 2020. 00:00Z

Chair: Babara Casati

Co-chairs: Raghu Ashrit, Harold Brooks.

Note taker: Tim Hume

Introduction
Barbara Casati gave a brief introduction to the session.

Presentation 1: How can we check the reliability of ensemble flood forecasts?
Presenter: James Bennett
Organisation: CSIRO (Australia)

- Question about why the statistics are being compared to the standard uniform variate. Presenter described why this is done.
- Interest in how presenter is verifying probabilistic timing forecasts. Suggestion that there are many other applications (e.g. timing of fog, thunderstorms etc). Another comment that verifying the end time for events would be very valuable. Quite active discussion about verifying timing forecasts.
- Question asking if a weighted CRPS has been trialled for verification? Suggestion it might help with issues presenter noted with forecast based selection where comparing two different forecast systems. Presenter said weighted CRPS was investigated. The problem was that once weighted, it was not possible to compute things like sharpness and other measures.

Presentation 2: Point-Biserial Correlation-Based Skill Scores for Probabilistic Forecasts
Presenter: Nachiketa Acharya
Organisation: International Research Institute for Climate and Society (Columbia University)

- The presenter is presenting some ideas regarding biserial correlation-based skill scores and seeking feedback from the audience. There was quite a bit of feedback in the chat.
- A question asking if IRI do experimental forecasts with users.
- Another question if using biserial correlation for probabilistic forecasts provides the same sorts of results/messages as BS/BSS/CRPS. Presenter says the answer is no – it measures something different. Sounds like (to the notetaker) that it is intended to complement the other scores rather than replace them.
- A question how this score compares to the generalised discrimination score (GDS).
- A comment than if bias and spread is corrected in a forecast, then RMSE and Pearson's correlation coefficient are one to one. Question that maybe a relation could be found between Brier score and biserial-correlation.

Presentation 3: Verification of Quantile Forecasts. A journey

Presenter: Deryn Griffiths *et al*.

Organisation: Australian Bureau of Meteorology.

- A question as to whether Bureau of Meteorology forecasts have other restrictions (e.g. forecasts are reported as integer millimetres rather than floats). Forecasts are rounded to integers for the public. Underlying data (in database) is to the nearest 0.1 mm. Percentages are rounded to the nearest one percent.
- A question about the Brier score. It is popular. Asking why the community does not use the log score. One response was that there are in fact a range of scores which are used; Brier score is just one of these.
- One comment that Brier score has been around for ages. Answer was that other scores have been around a long time, but Brier score is easy to use and calculate.

Presentation 4: Huber loss as a scoring function

Presenter: Rob Taggart et al.

Organisation: Australian Bureau of Meteorology

- Comment on how having a score (e.g. Huber mean) between the mean and median is useful. Commenter was very enthusiastic about this.
- Question about how the Huber metric is accepted by operational forecasters? Rob has found no real resistance to its use, provided it is explained properly.
- Rob suspects there are more applications of the Huber loss.

General Discussion

- An Australian member of the audience asked if some of the interesting previous talks could be given at a better time (for Australians) ☺.