# Verification of numerical models – what are the biggest challenges?

**Elizabeth Ebert[a] and Barbara Brown[b]**

[a] *Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, Melbourne*
[b] *National Center for Atmospheric Research (NCAR), Boulder, USA*
*Email: e.ebert@bom.gov.au*

Forecasts from numerical weather prediction (NWP) have been verified since the 1950s when they first started providing reasonable predictions, and beginning in 1985 modelling centres have exchanged standard scores for forecast fields on selected pressure levels. Operational verification hasn't really kept up with recent improvements in NWP, particularly as models have increased in resolution to <20 km for some global models and 1-2 km (convection-permitting) for mesoscale models. Ensemble prediction is overtaking deterministic NWP as the most important source of numerical guidance, and it too is being run at very high resolution. Coupling of atmospheric models to ocean and land surface models is being brought forward from the seasonal range to shorter ranges. Numerical weather and climate predictions are also being used to drive downstream impact models for emergency management, hydrology, agriculture, energy, and many other applications.

With the improvement in NWP now enabling the prediction of surface weather and its impacts, improved approaches are needed to evaluate these forecasts. This is an active area of research. To measure the performance of high resolution forecasts in a way that is more consistent with how they are used, several new spatial verification approaches have been proposed that consider coherent spatial structures in various ways. As these methods become mature for high resolution deterministic forecasts, they must now be adapted to verify high resolution ensemble predictions. A significant challenge in verifying predictions for extremes is accurately observing them and collecting enough forecast-observation pairs to compute meaningful and robust statistics. In general it is no longer appropriate to ignore observation error; in fact, as models continue to improve, the apparent error in forecasts will become closer and closer to the error in the observations. While it is possible to remove biases in observations when they are known, it is much more difficult to account for random errors in observations which lead to poorer verification scores.

Evaluation of seamless (weather-climate) numerical prediction requires verification approaches that allow for consistent interpretation across time scales. This is tricky because short- and medium- range forecasts tend to be deterministic or ensemble predictions of instantaneous "absolute" weather variables at fine spatial and temporal scales, whereas extended range forecasts are based on coarser resolution ensembles, are typically given as probabilistic predictions of weekly or fortnightly anomalies being in a particular category (e.g., highest tercile), and rely on large hindcast datasets for forecast calibration.

Verification of warnings of impacts associated with extreme weather brings extra challenges. Observations of the impacts may be difficult to obtain for a variety of reasons relating to how they are collected, and by whom, how they are stored and disseminated, and whether they measure something that can be predicted and verified or are only indirectly related to the impact. Communication between the meteorological and various downstream communities is also challenging, with each sector "speaking their own language" and having their own priorities for what makes a forecast useful to them. To enable the benefits of improved NWP and weather forecasting to be translated into improvements in downstream impact forecasts, it is necessary to develop and apply verification metrics that are meaningful to the downstream users.

Although this talk will draw on weather prediction for most of its examples, many of the issues apply to land surface and water prediction as well.

***Keywords:*** *Forecast verification, model evaluation*