Evaluating reliability and resolution of ensemble forecasts using information theory

Steven Weijs (1) and Nick van de Giesen (2)

(1) Delft University of Technology, Water Resources section, Delft, Netherlands (s.v.weijs@tudelft.nl), (2) Delft University of Technology, Water Resources section, Delft, Netherlands (n.c.vandegiesen@tudelft.nl)

Ensemble forecasts are increasingly popular for the communication of uncertainty towards the public and decision makers. Ideally, an ensemble forecast reflects both the uncertainty and the information in a forecast, which means that the spread in the ensemble should accurately represent the true uncertainty. For ensembles to be useful, they should be probabilistic, as probability is the language to precisely describe an incomplete state of knowledge, that is typical for forecasts. Information theory provides the ideal tools to deal with uncertainty and information in forecasts.

Essential to the use and development of models and forecasts are ways to evaluate their quality. Without a proper definition of what is good, it is impossible to improve forecasts. In contrast to forecast value, which is user dependent, forecast quality, which is defined as the correspondence between forecasts and observations, can be objectively defined, given the question that is asked. The evaluation of forecast quality is known as forecast verification.

Numerous techniques for forecast verification have been developed over the past decades. The Brier score (BS) and the derived Ranked Probability Score (RPS) are among the most widely used scores for measuring forecast quality. Both of these scores can be split into three additive components: uncertainty, reliability and resolution. While the first component, uncertainty, just depends on the inherent variability in the forecasted event, the latter two measure different aspects of the quality of forecasts themselves. Resolution measures the difference between the conditional probabilities and the marginal probabilities of occurrence. The third component, reliability, measures the conditional bias in the probability estimates, hence unreliability would be a better name.

In this work, we argue that information theory should be adopted as the correct framework for measuring quality of probabilistic ensemble forecasts. We use the information-theoretical measures of entropy (uncertainty) and relative entropy (Kullback-Leibler divergence) to formulate scores analogous to the BS and RPS, that allow for a similar decomposition into uncertainty, resolution and reliability. The BS and it's components are shown to be second order approximations their information-theoretical counterparts. In the new score, named the divergence score, the resolution can be seen as the information gained from the conditioning data on which the forecast is based, while the reliability measures to which degree the information gain is annihilated by the information loss that occurs due to inadequate processing of this data. In other words, resolution measures the right information in the forecast, while reliability measures the wrong information in the forecast.

Furthermore, the new decomposition allows for a precise definition of the difference between estimated uncertainty (depending only on the ensemble) and actual uncertainty (depending on how the ensembles compare to the observations). It follows that a minimum actual uncertainty can only be achieved when the estimated uncertainty equals the actual. Because deterministic forecasts (unless perfect) do not satisfy this condition, they necessarily lead to a loss of information. This provides a clear case for the use of ensemble forecasts, which can now be evaluated using tools from information theory.