Geophysical Research Abstracts
Vol. 14, EGU2012-8991, 2012
EGU General Assembly 2012

Constraints on Rational Model Weighting, Blending and Selecting when Constructing Probability Forecasts given Multiple Models

S. M.W. Higgins (1), H.L. Du (1), L.A. Smith (1,2)

(1) Centre for the Analysis of Time Series, London School of Economics, United Kingdom (s.m.higgins@lse.ac.uk), (2) Oxford Centre for Industrial and Applied Mathematics, University of Oxford, United Kingdom

Ensemble forecasting on a lead time of seconds over several years generates a large forecast-outcome archive, which can be used to evaluate and weight “models". Challenges which arise as the archive becomes smaller are investigated: in weather forecasting one typically has only thousands of forecasts however those launched 6 hours apart are not independent of each other, nor is it justified to mix seasons with different dynamics. Seasonal forecasts, as from ENSEMBLES and DEMETER, typically have less than 64 unique launch dates; decadal forecasts less than eight, and long range climate forecasts arguably none.

It is argued that one does not weight “models" so much as entire ensemble prediction systems (EPSs), and that the marginal value of an EPS will depend on the other members in the mix.

The impact of using different skill scores is examined in the limits of both very large forecast-outcome archives (thereby evaluating the efficiency of the skill score) and in very small forecast-outcome archives (illustrating fundamental limitations due to sampling fluctuations and memory in the physical system being forecast). It is shown that blending with climatology (J. Bröcker and L.A. Smith, Tellus A, 60(4), 663-678, (2008)) tends to increase the robustness of the results; also a new kernel dressing methodology (simply insuring that the expected probability mass tends to lie outside the range of the ensemble) is illustrated.

Fair comparisons using seasonal forecasts from the ENSEMBLES project are used to illustrate the importance of these results with fairly small archives. The robustness of these results across the range of small, moderate and huge archives is demonstrated using imperfect models of perfectly known nonlinear (chaotic) dynamical systems. The implications these results hold for distinguishing the skill of a forecast from its value to a user of the forecast are discussed.