

Quantifying Predictability using Multiple Ensembles under different Models: Limitations on the value of Probabilistic Forecasting

R. L. Machete (1), J. Broecker (2), D. Kilminster (3), L. A. Smith (1,2), I. M. Moroz (1)

(1) Mathematical Institute, University of Oxford, (2) Centre for Analysis of Times Series, Department of Statistics, London School of Economics and Political Science, (3) METRA, New Zealand Met Service

Examining multi-model ensemble forecasts of a simple physical system suggests new insights into probabilistic forecasting. The limitations of initial condition ensembles from imperfect models in the case where arguably each model provides high quality forecasts, are illustrated using observational data from a nonlinear circuit. Methods for interpreting agreement between model distributions are discussed, as well as the implications for forecasting when the ensemble distributions are not very similar. Kernel dressing methods (Roulston and Smith) and Bayesian Model Averaging (Raftery et. al.) methods are discussed in this context. Fundamental limitations on interpreting forecasts based upon ensembles of simulations as a (decision-relevant) probability forecast are highlighted and more general implications for probabilistic forecasting are considered.

M. S. Roulston and L. A. Smith, Combining Dynamical and Statistical Ensembles, *Tellus*, Volume 55 A, pp 16-30, 2003.

A. E. Raftery, T. Gneiting, F. Balabdaoui and M. Polakowski, Using Bayesian Model Averaging to Calibrate Forecast Ensembles, *Monthly Weather Review*, Volume 133 (5), pp 1155-1174, 2005