Geophysical Research Abstracts, Vol. 6, 06325, 2004
SRef-ID: 1607-7962/gra/EGU04-A-06325


# ON CONTRASTING MEASURES OF SKILL FOR PROBABILITY FORECASTS: THE CASE OF RARE EVENTS

L.A. Smith and C. Ziehmann
Center for Analysis of Time Series, LSE, London UK WC2 A 2AE

Currently, a wide variety of measures are available for evaluating the skill of probability forecasts. Some, like the Brier score, are fairly easily computed, while the computation and interpretation of others is less intuitive (as with the area under the ROC curve). Other scalar measures include Ignorance, which is based in information theory (Roulston and Smith, 2002), and the likelihood measure of classical statistics (Jewson et al, 2003). This contribution aims to clarify the strengths and weaknesses of these measures of skill, since the shear number of options poses a significant obstacle for the user of probability forecasts when deciding which forecast (or forecasts) to purchase and use. It is often suggested that the best measure depends on the user's utility function; we consider this suggestion in light of the fact that many users wish to put one forecast to multiple uses.

These issues are addressed in the context of nonlinear systems using the full range of joint forecast model and ensemble scenarios: perfect model - perfect ensemble, perfect model - imperfect ensemble, imperfect model. The relevance of each skill score and its weaknesses are discussed, both in day to day usage and when there is particular interest in relatively rare events, as is their value to the user who does not have a single utility function, but wishes to use the forecast for different applications.

M. Roulston and L.A. Smith (2002) Evaluating probabilistic forecasts using information theory, *Monthly Weather Review* **130** 6, 1653–1660.

S. Jewson, A. Brix, and C. Ziehmann (2003) A New Framework for the Assessment of Medium Range Ensemble Temperature Forecasts (in review *Atmos. Sci. Lett.*)