Geophysical Research Abstracts
Vol. 14, EGU2012-11444-1, 2012
EGU General Assembly 2012

Insightful Measures of Predictive Skill in Seasonal Forecasts

H. Du (1), F. Niehorster (1), L.A. Smith (1,2)

(1) CATS, London School of Economics, Statistics, London, United Kingdom (h.l.du@lse.ac.uk), (2) Pembroke College, University of Oxford, United Kingdom (lenny@maths.ox.ac.uk)

Doubt is cast on the common claim that an ensemble of models outperforms a single best model; the source of this misperception being the use of inappropriate measure of skill. The use of such measures can lead to a loss of information presented to the decision maker. The skill of probability forecasts of the Nino 3.4 index & the Sea Surface Temperature (SST) in the Main Development Region (MDR) based upon the ENSEMBLES seasonal simulations are considered. Issues in the interpretation of probability forecasts based on these multi-model ensemble simulations are addressed. The predictive distributions considered are formed by kernel dressing the ensemble and blending with the climatology; hindcasts from 1960 to 2005 are constructed. The sources of apparent skill (typical linear skill measures like Root Mean Squares and correlation skill) in distributions based on multi-model simulations is discussed, and it is demonstrated that the inclusion of "zero-skill" models in the long range can improve such scores. This casts doubt on one common justification for the claim that all models should be included in forming an operational PDF. The sources of skill from multi-model ensemble are discussed. True cross validation is also shown to be important given the small sample size available in seasonal forecasting. Probabilistic skill is shown to be robust out to 8 months for the Nino 3.4 index and out to month 2 for MDR SSTs. Are ensembles of models more fit for decision making than an initial condition ensemble of the "best" model? Results using a proper skill score show the multi-model ensembles do not significantly outperform a single model ensemble for Nino 3.4. Situations in which ensemble over model structures outperforms comparable ensemble using the "best" model structure are requested.