Fourth International Conference on Earth System Modelling
4ICESM-166, 2017

Interrelationships beween different scores in decadal prediction evaluation

A. Hense and R. Glowienka-Hense
Universität Bonn, Meteorologisches Institut, Bonn, Germany (ahense@uni-bonn.de)

For standardized variables the ensemble forecast verification scores "ensemble spread", "mean square error" and "ensemble spread score" can be expressed as analytic functions of the correlation coefficient and the common variance of the ensemble members ANOVA calculated from the analysis of variance taking the forecast years as treatments. ANOVA is a measure of the model sharpness and the correlation is calculated from the ensemble mean. An ensemble spread score of 1 indicates an ideal ensemble with a flat analysis rank histogram with the correlation being equal to the square root of ANOVA. Values less than 1 show that the ensemble is too sharp which might be due to ensemble or model deficiencies. A reduction in sharpness must finally lead to a reduction of correlation. The ensemble spread score does not depend only on the ratio of correlation and ANOVA but also on the absolute value of ANOVA. This suggests that the three scores mentioned should be interpreted together and that the analysis of variance of a forecast ensemble should be considered additionally. Kleeman (2002) suggested to use the utility function as a measure of the information of a single forecast ensemble calculated from ensemble spread and a standardized ensemble mean anomaly. We will show that the longterm mean of the utility function depends approximately only on ANOVA stressing again the advantage of an analysis of variance when issueing ensemble forecasts.