

Statistical Tools to Evaluate Skill of Probabilistic Forecasts

Maia AHN^{1*}, Meinke H²

1. Embrapa Meio Ambiente, Jaguariúna, Brazil
2. Department of Primary Industries, Toowoomba, Australia

Climate variability forecasting based on probabilistic data is becoming an increasingly important risk management tool for many climate-dependent industries and sectors. Although the debate about appropriate, statistical skill measures of such probabilistic systems has been going on for many decades, it has now reached a new sense of urgency due to the profound consequences that such decisions might have for our future. Skill measures are used to investigate the relevance of classification factors effect on rainfall probability distributions.

Here, we evaluated that the relative performance of skill measures using Monte Carlo methods, considering three scenarios. Gamma distributions with (a) shifts in means, (b) shifts in variance and, (c) simultaneous shifts in mean and variance.

The following measures were evaluated: Snedecor's F (SNF), Log-rank (LGR), Gamma likelihood ratio test (GLRT) for the hypothesis of no difference among means (LRM) and GLRT test for the hypothesis of no difference among means or among variances (LRMV).

The skill measures performance was evaluated by the frequency of rejection of the null hypothesis of no classification factor effect. The measures' adequacy depends on the scenario. For scenario (a), a simple skill measure based only on shifts in means is adequate. As expected, LRM and SNF tests do not perform well, if variances of conditional distributions are quite unlike. However, the LRMV and LGR tests, designed to capture shifts in means and/or variances are adequate for any scenario. We will outline implications of our findings for the on-going debate regarding the skilfulness of probabilistic forecasting approaches.