Geophysical Research Abstracts
Vol. 15, EGU2013-10196-1, 2013
EGU General Assembly 2013

Sensitivity of RCM evaluation to the choice of performance metrics

Christopher May and Clare Goodess
Climatic Research Unit, University of East Anglia, Norwich, United Kingdom (C.May1@uea.ac.uk)

Performance metrics allow the assessment of regional climate model skill through direct comparison with observations. While it is widely recognized that model performance depends on the variable, season and region considered, rather less consideration has been given to the choice of performance metric(s). Here, a 'metric' is taken to be a scalar measure and four different types of metric are identified: temporal variability metrics (such as Annual Cycle Skill Score and Linear Trends); standard error/variability statistics (such as RMSE); spatial pattern metrics (such as Spatial Skill Score and Correlation); and, event frequency metrics (such as PDF Overlap Skill Score). Currently there is no consensus on what a standard suite of performance metrics might constitute, thus an investigation into the robustness of individual metrics is being undertaken as a first step in identifying potential candidates.

The analysis undertaken uses a number of metrics from the four categories identified above, ranging from standard statistical measures such as standard deviation, through measures assessing spatial errors such as correlation coefficients, through to more complex PDF/CDF scores to assess event occurrence frequencies. The focus is on mean temperature, diurnal temperature range, precipitation, sea level pressure as well as a subset of indices of extremes based on those defined by the Expert Team on Climate Change Detection and Indices (ETCCDI). Initially work is focusing on the reanalysis-driven ENSEMBLES simulations, but will be extended to CORDEX simulations for Europe, Africa and other regions as these become available.

In this first step, the sensitivity of evaluation outcome to choice of metric and to variations in their construction (such as the domain over which they are calculated, the time frame used, be it seasonal compared to annual evaluation, and individual metric constructions, such as the number of bins used in PDF methods) is assessed. If metric scores are to be used for simple ranking of members of a multi-model ensemble or as input in probabilistic climate change projections using performance based weighting, for example, it is important to have a clear understanding of the uncertainty inherent in the metrics themselves. The aim of this first step is to identify a robust set of metrics which is sufficiently comprehensive for use in a range of applications, but which does not include redundant or repetitive information. Further work will focus on the evaluation of methods for combining multiple metrics and on assessment of the stationarity of model errors.