



An intercomparison of a large ensemble of statistical downscaling methods for Europe: Overall results from the VALUE perfect predictor cross-validation experiment

Jose Manuel Gutiérrez (1), Douglas Maraun (2), Martin Widmann (3), Radan Huth (4), Elke Hertig (5), Rasmus Benestad (6), Ole Roessler (7), Joanna Wibig (8), Renate Wilcke (9), and Sven Kotlarski (10)

(1) Institute of Physics of Cantabria (IFCA), CSIC – University of Cantabria, Santander, Spain., (2) Wegener Center for Climate and Global Change, University of Graz, Graz, Austria., (3) School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK., (4) Department of Physical Geography and Geoecology, Faculty of Science, Charles University and Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic. , (5) Institute of Geography, University of Augsburg, Augsburg, Germany., (6) The Norwegian Meteorological Institute, Oslo, Norway., (7) Institute of Geography, University of Bern, Bern, Germany., (8) Department of Meteorology and Climatology, University of Lodz, Lodz, Poland. , (9) Rossby Centre, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden., (10) Federal Office of Meteorology and Climatology MeteoSwiss, Switzerland.

VALUE is an open European network to validate and compare downscaling methods for climate change research (<http://www.value-cost.eu>). A key deliverable of VALUE is the development of a systematic validation framework to enable the assessment and comparison of both dynamical and statistical downscaling methods. This framework is based on a user-focused validation tree, guiding the selection of relevant validation indices and performance measures for different aspects of the validation (marginal, temporal, spatial, multi-variable). Moreover, several experiments have been designed to isolate specific points in the downscaling procedure where problems may occur (assessment of intrinsic performance, effect of errors inherited from the global models, effect of non-stationarity, etc.). The list of downscaling experiments includes 1) cross-validation with perfect predictors, 2) GCM predictors –aligned with EURO-CORDEX experiment– and 3) pseudo reality predictors (see Maraun et al. 2015, Earth's Future, 3, doi:10.1002/2014EF000259, for more details). The results of these experiments are gathered, validated and publicly distributed through the VALUE validation portal, allowing for a comprehensive community-open downscaling intercomparison study.

In this contribution we describe the overall results from Experiment 1), consisting of a European wide 5-fold cross-validation (with consecutive 6-year periods from 1979 to 2008) using predictors from ERA-Interim to downscale precipitation and temperatures (minimum and maximum) over a set of 86 ECA&D stations representative of the main geographical and climatic regions in Europe. As a result of the open call for contribution to this experiment (closed in Dec. 2015), over 40 methods representative of the main approaches (MOS and Perfect Prognosis, PP) and techniques (linear scaling, quantile mapping, analogs, weather typing, linear and generalized regression, weather generators, etc.) were submitted, including information both data (downscaled values) and metadata (characterizing different aspects of the downscaling methods). This constitutes the largest and most comprehensive to date intercomparison of statistical downscaling methods. Here, we present an overall validation, analyzing marginal and temporal aspects to assess the intrinsic performance and added value of statistical downscaling methods at both annual and seasonal levels. This validation takes into account the different properties/limitations of different approaches and techniques (as reported in the provided metadata) in order to perform a fair comparison. It is pointed out that this experiment alone is not sufficient to evaluate the limitations of (MOS) bias correction techniques. Moreover, it also does not fully validate PP since we don't learn whether we have the right predictors and whether the PP assumption is valid. These problems will be analyzed in the subsequent community-open VALUE experiments 2) and 3), which will be open for participation along the present year.