Automatic input variable selection with genetic algorithms for statistical precipitation prediction

P. Horton[1], O. Martius[1], S. L. Grimm[2]

[1] Institute of Geography, Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[2] Center for Space and Habitability, University of Bern, Bern, Switzerland

Precipitation prediction remains a challenge for meteorological models, whether for weather forecasting or climate change impact studies. To this end, statistical precipitation prediction methods are used to post-process or downscale the precipitation fields provided by the models or even build a prediction based on other predictors. One of the latter methods is the analog method (AM) that relies on the hypothesis that similar atmospheric conditions are likely to result in similar local effects. AMs are non-parametric methods similar to k-nearest neighbors approaches. They usually rely on predictors describing the atmospheric circulation and the moisture content of the atmosphere to sample similar meteorological situations in the past and establish a probabilistic prediction for a target date. AMs can be based on outputs from numerical weather prediction models in the context of operational forecasting or outputs from climate models in climatic applications.

AMs are constituted of multiple predictors and can be structured in stepwise subsampling steps, named levels of analogy, that subsample analog situations from a pool of candidate situations previously selected with other predictors. The predictors are meteorological 2D fields considered on spatial domains and compared with specific metrics (analogy criteria), such as the Root Mean Square Error or a criterion based on gradients in the predictor field.

The elaboration of AMs is commonly a semi-manual process, where input variables and analogy criteria are explicitly chosen, and the spatial domains are optimized. The selection of predictors in AM applications often builds on previous works and does not evolve much. However, the climate models providing the predictors evolve continuously, and new variables might become relevant to be integrated into AMs. Moreover, the best predictors might change from one region to another or for another predictand of interest.

We thus introduced genetic algorithms (GAs) to select the predictor variables and all other AM parameters automatically. We also let the GAs pick the best analogy criteria, i.e., the metric quantifying the analogy between two situations. The whole approach, including the GAs, is coded in the AtmoSwing software (C++) developed by the authors. The computing was done on UBELIX, and due to the high number of calculations required by the GAs for this application, a more efficient GPU-based implementation of the predictor fields comparison has been implemented.

We applied this approach to the prediction of precipitation for 25 catchments in Switzerland. The results showed an informative variability in the selected variables but a great consistency regarding the underlying meteorological processes. The resulting AMs showed better skills than reference methods, and two unexpected results arose: (1) fewer levels of analogy with more predictors showed better results than more complex structures, and (2) a new analogy criterion, developed for this study, showed great success and was the most selected by the GAs. We also found out that a fully auto-adaptive mutation operator composed of three chromosomes, developed by the authors, worked considerably better than other alternatives employed.

This approach opens new perspectives for a more straightforward adaptation of AMs to new regions or new predictands, or different datasets, by providing an automatic input variable selection. It also provides insight into the relevance of specific predictors for certain meteorological phenomena.