Statistical downscaling: Obtaining local precipitation from large-scale flow fields

Precipitation varies in space, in scales that are often much shorter than that of climate model simulations. Part of this local variability is related to orographic effects, and it may influence precipitation under some weather types but not under other types. Coarse resolution models are not supposed to resolve such topographic effects on precipitation, but they may well depict the general flow situation.

Based on station data, a relation between local precipitation and the large-scale flow can be established, and this relation can then be applied to statistically "downscale" large-scale flow from another source, e.g., reanalyses or climate model simulations, to local precipitation.

A simple statistical downscaling approach is analog resampling. The assumption here is that similar large-scale flow situation will lead to similar local precipitation. The method requires a large pool of potential analogs, ideally thousands of days. If we have the large-scale flow, e.g., from future climate simulations, then we can search, for each day in the future, the closest analog in the past according to the large-scale flow. The local precipitation on these days then constitutes a statistical downscaling.

In the exercise, we have three large-scale predictors: local 500 hPa geopotential height (Z500) as well as the local zonal and meridional gradients in the 500 hPa geopotential height field (dZ500/dx and dZ500/dy, respectively). We have this information for the present (we use one year of data as our pool of potential analogs – which is far too little but serves the purpose) as well as for a future period (here we use one month). From these three variables we have subtracted the mean annual cycle and have standardized the anomalies.

Your task now is to find, for each day in the future, the most similar analog day in the pool of analogs. The first question is to define a distance measure. For a single variable the distance is just the difference *d*. Over all three variables we have different options. The Eucledian distance:

$$D = \sqrt{d_1^2 + d_2^2 + d_3^2}$$

is the square root of the sum of the squared distances. It is however only defined if the all differences *d* have the same unit. This is not true in our case. One solution is to standarize the distances. We can write this as:

$$D = (\mathbf{d}^{\mathbf{T}}\ \mathbf{C}^{\mathbf{-1}}\ \mathbf{d})^{0.5}$$

With **C** being a diagonal matrix with the variances in the diagonal. Relatide to this measure is the Mahalanobis distance:

$$D = (\mathbf{d}^{\mathbf{T}}\ \mathbf{B}^{\mathbf{-1}}\ \mathbf{d})^{0.5}\,.$$

It additionally takes the covariance between the series into account; **B** is the (non-diagonal) covariance matrix. There are further distance measures; in the exercise here we use the Eucledian distance.

Once a distance measure is defined, distances between each target day and all days in the pool of analog must be calculated. For each target day, all distances are compared and the shortest distance is searched. Once the closest analog day is found, the procedure is trivial: Select precipitation from that day. A schematic figure is given below.

The analog approach can be powerful for precipitation downscaling for small catchments. Often it is implemented in a hierachical way, i.e., the selection proceeds over several steps. The resulting downscaled data can be further post-processed according to the observations. The closest *n* analogs can be used to obtain a measure of robustness.

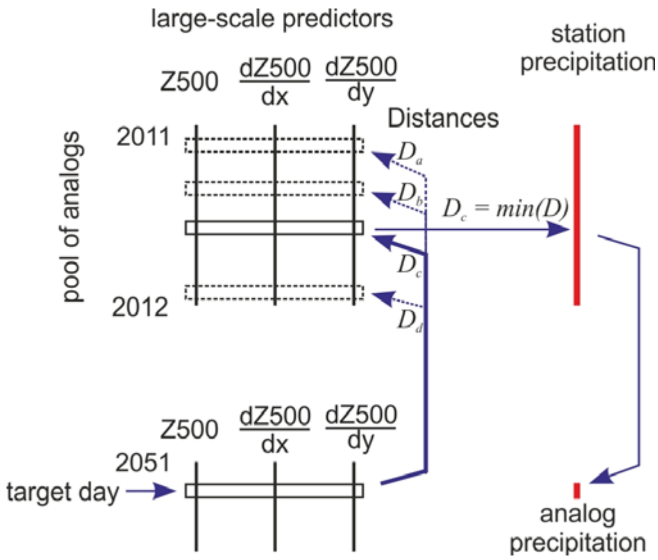

Figure: Statistical downscaling through analog resampling.