ANALOGUES AND
WEATHER TYPING

NICOLA CORTESI

CERFACS
Centre Européen de Recherche et
de Formation Avancée en Calcul Scientifique
TOULOUSE, FRANCE
The AM was first developed as a method for weather predicting more than 60 years ago and later also for seasonal forecasting. Only in 1995 Zorita et al. introduced it also for a downscaling purpose. A few years later they compared the AM with the other downscaling methods available at that time and soon the AM became one of the more commonly used Downscaling methods. Nowadays, there are hundreds of different variants of the AM. It is also used as a benchmark method to assess more complicated downscaling techniques, and it has been also coupled with other Downscaling Methods to overcome some of its limitations, as we'll see later.
In spite of its simplicity, it is still one of the best methods to downscale daily precipitation, because it is not a parametric approach, I mean, that no assumptions are made about the statistic distribution of the predictors, and also because it is a non-linear method, so it can take into account the non-linearity of the relationships between the predictors and the predictand; it is one of the few methods that doesn’t underestimate the variance of the predictand variable, it is spatially coherent (i.e: preserves the \textit{spatial covariance structure} of local predictand), an important requirement for the hydrological studies, because there must not be any spatial discontinuities of the down scaled precipitation inside the same water basin (it is a consequence of using only one predictor domain for all target locations, so the same analog day is selected for the whole study region). It is also easy to implement and has a low computational cost, like almost all the statistical downscaling methods.
In case the predictors are only 2 scalar variables, then it is possible to visualize how the AM works plotting the historical data of the two predictors, more the data of the day to downscale, here in red. Data can be daily or monthly, if the downscaling is monthly. We also have to remove the days belonging to the same year of the target downscaled day to avoid including also the same day of the day to downscale.
First, the data outside the search radius of the AM is removed.
Usually a search radius of about 60 days is used, two months, but it is also possible to choose only the days belonging to the same season of the day to downscale instead of using a search radius. Of course, the number of predictors usually is much higher than two. ..
... because each predictor is represented by a large-scale field from a reanalysis dataset that can assume hundreds of different values for each day. So, the predictor fields are converted to normalized anomalies and the dimensionality of the fields is reduced with a Principal Component analysis, stopping for instance at 95% of the explained variance to determine the optimal number of principal components, ...
… that are used as predictors instead of the original ones. Typically from 10 to 30 principal components are retained at the end of the analysis.
The AM consists in searching in the (remaining) historical data (of the predictand variable) when the predictor fields are more similar to the predictor fields of the day to downscale, or the month if the downscaling is monthly.

The relationship between the large-scale predictor fields and the local predictand for the AM depends on how such a distance between the predictor fields of any 2 days or months is measured. Such a measure is called “Similarity measure”; it is a scalar variable, so all the days can be ordered in a sequence from the more similar to the downscaled day to the less similar one.

While many downscaling methods have automated procedures to select the best combination of predictors, the AM instead needs testing manually all the possible combinations of predictors and Similarity measures (and different predictors can also be associated to different Similarity measures within the same AM). So, especially in the case of the AM, it is better to select the predictors and the Similarity measure on the base of theoretical considerations, to capture the physical forcings between the predictors and the predictand (and also to reduce the stationarity problem).
The performance of the AM is greatly influenced by the type of Similarity measure employed. The more commonly used Similarity Measures are: the Euclidean distance; here \( z \) is the difference between \( x \) and \( y \), the two vectors with the principal components of the couple of days that we want to measure their distance; \( z \) is eventually weighted with the explained variance of each principal component; or we can measure the sum of the absolute values of each component of \( z \); we can measure the cosine of the angle between the two vectors (that is proportional to their scalar product); the Mahalanobis distance (less performant), is a kind of scalar product that takes into account also the eigenvalues of the principal components, the lambda (that are used to weight the scalar product of the two vectors); while the Pattern correlation is simply the Pearson correlation coefficient between the two predictor fields of the two days; and the Teweles-Wobus score, which is proportional to the difference between the horizontal pressure gradients of the two predictor fields and do not rely on the principal components. It is often used for downscaling daily precipitation because the horizontal gradient is proportional to the strength of the geostrophic winds. You can also choose to minimize the similarity measure for a sequence of days instead, in this case up to a week before the day to downscale. Naturally, you give a higher weight to the couples of days closer to the downscaled day.
The day that minimize the Similarity Measure is called “analog day”…
...and it is specific of that downscaled day only.

The downscaled value then is the same observed value of the predictand of the Analog day.
For example, if the predictand is the daily precipitation, the downscaled value would be 3.7 mm in this case.

However, if you always select the closer day, the variance of the predictand will be underestimated because the predictor fields are not the only source of variability of the predictand: part of its variability is not captured by the predictors, so the same predictor field (synoptic configuration) is associated to several different values of the local predictand.
To overcome this problem, you can select not only the closer day, but an ensemble of the $n$ closer days, 4 in this example, and then you select one of these days. This kind of Analogue Method is called Nearest Neighbour Resampling or Analogue Resampling.

In this way, the AM becomes a stochastic method, because each time you downscale a variable, its time serie will be different.

The AMs are classified as Deterministic, if there is no resampling, or Stochastic, if there is resampling.

The higher the value of $n$, the less the variance is underestimated.
Now that you know how the AM works in detail, you are also able to understand its limitations. The first one is that the AM cannot extrapolate the values outside the range of the historical data, extremes in particular, and since their frequency and intensity are often projected to increase in future, this is quite a serious limit.
Also for this reason, the AM needs a large calibration sample. The more data, the better. This limitation could be somewhat mitigated extending the Search Radius to the whole year instead of working with the same season, even if the performance of the AM decreases a little. On the other hand, the future seasonal climate might not correspond to the present seasonal climate, so the calibration should be applied (not separately for each season but) to the whole training period (without using a search radius). But this is difficult to achieve, because the best combination of predictors is often different from season to season. So, this is still an open issue.
Caveats of the AM

- It doesn’t preserve the *spatial autocorrelation structure* of the observed variable or its observed temporal *frequency distribution* or its *persistence* (for daily precipitation).

Recently, a number of AMs that succesfully deal with these issues have been presented:

- **Ribalaygua et al. (2013):** a two step Analogue/Regression method
- **Benestad (2009):** a two step Analogue/Quantile mapping method using the Extended EOFs instead of the classic EOFs.
- **Hwang and Graham (2013):** a two step Bias-correction/Stochastic AM
- **Matulla et al. (2007):** an AM applied to sequences of 2-3 days

It also doesn’t preserve the spatial autocorrelation structure of the observed variable or its observed temporal frequency distribution or its persistence, which are important for many studies. (for daily precipitation: also the sequence of wet/dry days and the length of wet/dry spells, and the probability of a rain day following a dry day.)

To overcame this problem and also the previous ones, Ribalaygua et al. recently presented a two step model. The first step is an analogue approach and then the n most similar days are used in a multiple regression model to extrapolate the daily temperature. While for daily precipitation, Dr. Benestad proposed a two step Analogue/quantile-quantile mapping technique using the extended empirical ortogonal functions instead of the classic ones. Hwang and Graham also presented a 2-step method based on a bias correcting technique and then on a stochastic AM, while Matulla et al found good results using sequences of 2-3 days.

These are the "state of the art" AMs, because they lessen some of the limitations of the classic AM. So, it seems that the AM is not a dead method, but there is still plenty of room for improvement in the near future.
Here you have a few selected readings on the AM.

- **Introduction to the AM:**

- **Robustness Analysis:**
Gutierrez et al. (2013) Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions. *J. Climate*, 26, 171-188. DOI: 10.1175/JCLI-D-11.00687.1

- **Similarity Measures:**

- **Two step Analogue/Regression method:**
The Weather Typing method is more complicated than the AM; it is composed by two different steps; in the first one, the continuum of the atmospheric circulation at the synoptic scale (more or less 1000 kilometres), represented by the surface pressure fields or by the geopotential height, is classified into a small number of (circulation or weather) classes, called circulation types or weather types, usually from 4 to 43. The idea of the Weather Typing method is that the same type of weather pattern will also generate the same precipitation or temperature pattern. For example, in the Iberian Peninsula the western winds bring moist air from the ocean to the land and as a consequence, precipitation is abundant over the peninsula (up to the Spanish side of the Pyrenees, also because of the orographic effect). These westerly winds can be associated to a Weather Type, as we’ll see later.

In the second step, you select one of the other Perfect Prog Statistical Downscaling Methods and apply it to model the statistical relationship between the Weather Types and the observed values of the local predictand (stations or grid points) during the same historical calibration period of the reanalysis field.
So, if you use the output of a GCM instead of a reanalysis dataset, you can select the same WT classification and the same downscaling method of before, in this way you can also employ the same statistical relationship of before to downscale the values of the predictand for the future climate, during the same period of the GCM output, assuming the statistical relationships are constant in time, so that they can be applied also for the future climate (the so-called “stationarity hypotheses”).
Advantages of the WTM

1. *Greater understanding* of the problems involved
2. Weather Types useful for downscaling *Extreme Indices*
3. *Spatially coherent*
4. *Low computational cost*
5. It can be *non-parametric*
6. It can be *non-linear*
7. It can *not to underestimate the variance*

The strengths of this method are that there is a greater understanding of the problems involved because the predictors, the Weather Types, are linked to the predicand in a sensible way, and sometimes also in a physical way, as we’ll see later with the Lamb classification.

They have been used successfully also to downscale Extreme indices, including extreme precipitation.

They are spatially coherent, like the AM, and have a low computational cost.

The other properties, such as that of being parametric or not, linear or not, or to underestimate the variance or not, depends on the type of downscaling method applied in the second step.
Almost every Statistical Downscaling method can be applied to a Weather Classification, but the most commonly used methods are the Regression Models for temperature, that for daily precipitation are usually extended to the Generalized Linear Models, to replace the Gaussian distribution with the Gamma distribution, or to the Generalized Additive Models to be also non-linear (it works well for downscaling extremes); sometime a Logistic (logit) Regression Model is first applied to model the prob. of occurrence or daily rainfall, and then the non-zero daily prec. values are determined (separately) with one of the above-mentioned methods. The Weather Generators are widely used too: they are stochastic models that generate random time series (also of spatial fields) with the same statistical properties of the observed predictand, such as the prob. of occurrence, the sequence of wet and dry intervals, the temporal and spatial correlation, ecc. (it is usually modelled with a Markov chain).

Also the Analogue Method can be applied, for example if you select only the analog days that belongs to the same Weather Type.

Canonical Correlation Analysis: to determine the linear relationships between the predictors and the predictand.
So, the novelty of the WT approach really lies in the first step, in how the WTs are classified. The two oldest classifications are the Grosswetterlagen classification for Central Europe and the Lamb classification for the British Isles. They were developed before the computer age, so they require a subjective determination of daily weather charts. (With the advent of computers, the classifications became more objective and could be automatically applied to other regions, even if they should always be reviewed by an expert to check their suitability and coherence for the study area).

A few years later, Jenkinson and Collison presented the first almost objective version of the Lamb classification introducing indices (based on the direction and vorticity of the geostrophic winds), and later Jones et al. compared it with the original Lamb WTs. In the year 2000, Spellman applied the Lamb classification to Spain and Trigo and DaCamara to Portugal. Subsequently, it was applied to almost every European country. In 2007 also the Grosswetterlagen classification was converted to an objective classification.

In 2010, a catalogue with all the WT classifications was published within the framework of the COST733 project (in the quest for the most suitable classification and more than 70 different classifications were described only for Europe. This catalogue groups all the classifications in 5 categories:
The first two categories are based on a top-down approach, in which the WTs are defined a priori (often on the base of physical arguments): you already know how the weather types look like before classifying the predictor fields.

All the 7 subjective methods fall in this category, and also the category of the threshold based methods, that are almost objective methods (because they are automated), only the thresholds (used to separate one WT from another) are subjective. Both the Grosswetterlagen classification and the Lamb Classification modified by Jenkinson and Collison fall in this category. The third column shows the number of weather types for each classification. For example, there are three Lamb classifications: one with 10 WTs, one with 18 and one with 26 (the last column shows the predictors).
The other three categories follow a bottom-up approach in which there isn’t any a priori knowledge of the WTs, so they are derived searching for any structure in the dataset itself. The classifications of the third category use (the multivariate analysis, in particular) the principal component analysis so they are not fully objective because there are still some subjective choices such as the number of principal components (the % of variance explained, T or S mode of the PCA, rotation of the component or not, ecc.).

The classifications based on the Leader Algorithm were developed at a time when computing capabilities were still limited and I won’t enter into the details even because nowadays the more used classifications belongs to the category of the optimization methods. Some of them are based on the k-means clustering analysis, that minimize the variability within the same Weather type and some classifications also filter the data before with a PCA. The main drawback of the k-means clustering is that the number of clusters has to be chosen a priori, so some classifications rely instead on the hierarchical clustering analysis that doesn’t need to know the number of clusters beforehand. Recently, the SANDRA classification improved the k-means clustering avoiding the convergence in the various local minimum of the optimization function. The last classifications are based on Neural Networks but have a very high computational cost.
Now let’s apply for example the Lamb classification in its automated version of Jenkinson and Collison, that is still widely used, not only for precipitation and hydrological studies but also for air quality studies (based on the ozone concentration and the pm5 and 10 particles), erosion studies, and even to diagnosticate the insurgence of respiratory diseases. You only need to extract the time series of SLP at these 16 grid points p1, ..., p16; as you can see they are regularly distributed over the region to downscale, in this case the Iberian Peninsula. Then you apply this set of rules that define 7 geostrophical indexes; the first two indexes are the two components of the wind flow, the third index is the vectorial sum of the two components, D is the direction of the flow in degrees; then there are the two component of the vorticity and the total vorticity is simply the sum of its two components. All these constants depends only by the latitude of the grid; for instance, 1.305 is the inverse of the sinus of the latitude; and finally, you select how many WTs you want to introduce in the classification, 26 in our case, and assign each day to one weather types using these thresholds: if the absolute value of the total vorticity is lower than the total flow, the Weather Type will be a directional WT depending on the value of D; if not, it will be a cyclonic or an anticyclonic WT, depending on the sign of the total vorticity. The direction of the flow determines the final WT. For instance, the westerly WT is characterized by a low vorticity and a flow direction coming from west.
For example, here you can see the mean SLP fields of the 8 directional WT more the pure Cyclonic WT and the pure anticyclonic WT when using the 20th century reanalysis dataset.
Of course, not everything is perfect and even the WTM has its limitations, in particular using only the pressure or the geopotential height as predictors to classify the WTs cannot take into account the projected changes for predictands such as temperature and precipitation.

Only the WT classifications that include also additional predictors such as large-scale temperature or atmospheric humidity can deal with this limit.

Second, there is a limit on the maximum number of WTs for each classification because the higher is the number of WTs, the lower is their frequency of occurrence, and in this case the Statistical Downscaling Methods applied in the second step don’t work very well, especially the regression models (in the classification process there is always a loss of information, thus to capture the most part of the variability of the predictand there is the need to select the higher number possible of WTs).

But if you select too few WTs, you might miss some important low-frequency synoptic patterns, for example those associated to extreme events.

So, a compromise must be done to select the optimal number of WTs for a certain application.

Third, many WTM under estimate the temporal climate variability, especially the linear models. Two widely used solutions are that of inflating the downscaled variance artificially, after the downscaling, or an even better solution is that of adding an noise term to the predictand, to increase its variability.
There are also some limitations that are shared by both the AM and the WTM, and in general by all the statistical downscaling methods:

First, the relationships between the predictors and the predictand must be constant in time, so such hypothesis must be tested each time you employ a new dm, usually with a perfect model experiment to verify the stationarity in the "pseudo-reality" world simulated by the RCMs or the GCMs or a Kolmogorov-Smirnov test with the null hypothesis of equal statistical distribution of the downscaled series and the observed series. For temperature, you can compare the bias of the method in an historical warm period with that obtained in normal conditions to test the robustness of the bias to warmer climatic conditions.

For example, Dr. Gutierrez has recently found that the AM underestimates the temperature anomalies in Spain during the warm periods, so the AM doesn’t seem really well suited for daily temperature.

Convective structures such as those responsible of heavy autumn precipitation along the Mediterranean coast of Spain and France are still not well simulated, mainly because the current generation of GCMs cannot resolve small or short atmospheric structures (because of its insufficient spatial and temporal resolution), so rely on parametrizations instead.

Domain Issue: each Statistical Downscaling method requires to optimize also the predictor domains, (and different predictors can also have different domains) because it has been demonstrated that the variation in performance between different domains (applied to the same Downscaling method) has the same magnitude of the variation in performance between different Downscaling methods.
Suggested Readings for the WTM:

- *Weather Types Catalogue:*

- *Lamb Classification:*

- *SANDRA Classification:*

- *A Classic:*
Thanks for your attention!

Any comments and questions are welcome

cortesi@cerfacs.fr