Geoscientific
Model Development

Open Access

AnaWEGE: a weather generator based on analogues of atmospheric circulation

P. Yiou

Laboratoire des Sciences du Climat et de l'Environnement, UMR8212 CEA-CNRS-UVSQ, CE Saclay, l'Orme des Merisiers, 91191 Gif-sur-Yvette cedex, France

Correspondence to: P. Yiou (pascal.yiou at lsce.ipsl.fr)

Abstract. This paper presents a stochastic weather generator based on analogues of circulation (AnaWEGE). Analogues of circulation have been a promising paradigm to analyse climate variability and its extremes. The weather generator uses precomputed analogues of sea-level pressure over the North Atlantic. The stochastic rules of the generator constrain the continuity in time of the simulations. The generator then simulates spatially coherent time series of a climate variable, drawn from meteorological observations. The weather generator is tested for European temperatures, and for winter and summer seasons. The biases in temperature quantiles and autocorrelation are rather small compared to observed variability. The ability of simulating extremely hot summers and cold winters is also assessed.

1 Introduction

Weather generators are tools to generate random time series of climate variables (generally precipitation, temperature or wind speed) with realistic statistics. Such statistics can be the mean, variance or quantiles of the variables. More sophisticated statistical quantities can be evaluated (e.g. persistence, power spectra, skewness, etc.), depending on the application of the weather generator. Their realism is tested on meteorological observations. Their use is mainly to simulate long series at local spatial scales that are not accessible to general circulation models (GCMs) or even regional climate models (RCMs). Empirical probability distributions of relevant climate variables can hence be estimated from those weather generator simulations.

Several random weather generators have been developed, with the aim of simulating long sequences of precipitation, for agronomical applications (Mavromatis and Hansen, 2001; Huth et al., 2001; Hansen et al., 2006; Semenov and Barrow, 1997; Flecher et al., 2010; Busuioc and von Storch, 2003), by using an empirical relation between large and small scale variability. Such tools can simulate thousands of years with daily increments in just a few minutes. This ease of use has been an incentive for the development of such applications.

One of the limitation of many random weather generators is their lack of spatial coherence, unless it is imposed on the marginal distributions of a variable at two locations or more (Naveau et al., 2009). Such a spatial constraint is technically difficult to impose, because there is an infinity of choices for models of spatial covariance (Schölzel and Friederichs, 2008). Empirical methodologies considering spatial coherence have been tested for precipitation in the USA (Wilks, 1999; Schoof and Robeson, 2003) or Australia (Westra and Sharma, 2009). Stochastic models that use a decomposition of the atmospheric circulation into weather types and their empirical relation with surface variables have also been developed (Bissolli and Dittmann, 2001; Kreienkamp et al., 2013) for European climate.

This paper presents a random weather generator (AnaWEGE) based on circulation analogues (Vautard and Yiou, 2009; Yiou et al., 2012). This weather generator preserves spatial constraints by construction and can be used to generate time series of climate variables such as temperature, precipitation or wind speed, distributed over a continent. The principle is similar to the one derived from weather types (Bissolli and Dittmann, 2001; Kreienkamp et al., 2013), although the constraint on the existence of clusters is relaxed.

The paper first presents the method of analogues of circulation. Two stochastic methodologies are proposed for a stochastic weather generator based on analogues of circulation. The first one is a perturbation of an observed sequence of climate variables. The second one allows one to explore several likely sequences of a climate variable. This weather generator is tested for temperature observations over Europe, with a focus on time autocorrelation, and quantile properties. The ability of the weather generator to simulate hot summers and cold winters is also tested. The motivation stems from an application to the energy sector, because peaks of energy consumption occur during cold winters (for heating) and hot summers (for air-conditioning in buildings).

2   Analogues of circulation

The preliminary step of the weather generator is to compute analogues of circulation (Lorenz, 1969; Van den Dool, 1994; van den Dool, 2007; Zorita and von Storch, 1999). This computation is done once (before using the weather generator), and the resulting analogues are used to generate weather time series. Here, the terminology of Yiou et al. (2012) for computing analogues of circulation is recalled. The analogues of circulation are computed from daily sea-level pressure (SLP) data. The data is extracted from the National Centers for Environmental Prediction (NCEP) reanalysis data (Kalnay et al., 1996) for daily SLP between 1 January 1948 and 31 December 2012. The SLP data have a horizontal resolution of $2.5 \times 2.5°$. We focus on the North Atlantic region ($80°$ W–$30°$ E; $30°$ N–$70°$ N). This region is chosen because it encompasses the atmospheric patterns that influence surface temperature and precipitation over Europe (Hurrell et al., 2003; Cassou et al., 2005). Of course other reanalysis datasets or climate model simulations could be substituted to the NCEP reanalysis, for instance CMIP5 simulations (Taylor et al., 2012).

Each day j between 1 January 1948 and 31 December 2012 can be written:

$$j = y10^4 + m10^2 + d, \qquad (1)$$

where y is the year (between 1948 and 2012), m is the month (between 1 and 12) and d is the day (between 1 and 31). In the sequel, days of the year are encoded in this manner. By convention, the *calendar day* of j is:

$$\kappa(j) = m10^2 + d. \qquad (2)$$

For each "target" day j, the set S_j of days $j' = y'10^4 + m'10^2 + d'$ is determined, where $y' \neq y$, and the *calendar distance* $\delta(j, j')$ between j and j' is less than 30 days. The calendar distance is the number of days that separate $\kappa(j)$ and $\kappa(j')$. The set S_j is:

$$S_j = \{j' = y'10^4 + m'10^2 + d', y' \neq y, \delta(j, j') \leq 30\}. \qquad (3)$$
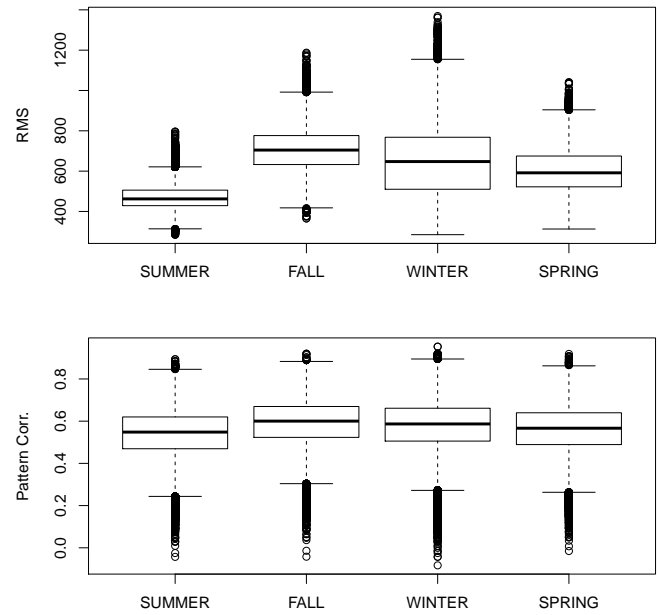
Fig. 1. Scores of daily analogues for each season. Upper panel: RMS values (in Pa). Lower panel: pattern rank correlation between RMS analogues and actual SLP.

The analogues are computed by minimising the root mean square (RMS) distance between SLP[j] and SLP[j'] over S_j:

$$D(j, j') = \left[\sum_x \left(\text{SLP}[x, j] - \text{SLP}[x, j'] \right)^2 \right]^{1/2}, \qquad (4)$$

where x is the spatial dimension. Here the first $K = 20$ analogues of the target SLP of day j are considered, i.e. the ones achieving the $K = 20$ smallest RMS values over S_j. The choice of the RMS as a distance to minimise is debatable. Other distances can be chosen (e.g. Mahalanobis or "taxicab"). The Mahalanobis (1936) distance is potentially interesting because it normalizes the data by their spatial covariance structure. But it is a computational burden that makes it almost ten times slower than RMS, due to repeated products of large matrices.

For each day $j \in$ [1 January 1948, 31 December 2012], $K = 20$ analogues of SLP are obtained, with days J_j^k ($k \in$ [1, K]) in years different than those of j. For all analogues, the spatial rank correlation between SLP[j] and SLP[J_j^k] is computed. This score is used to provide an objective degree of similarity between the target SLP and its analogues: correlation values lie between -1 and 1 and p values can be obtained for statistical significance. The (Spearman) rank correlation measures the pattern similarity, rather than the average field proximity captured by the distance, although correlation is not a *distance* in the mathematical sense.

The distribution of RMS and correlation values for the 20 first analogues are indicated in Fig. 1. They are computed for the four seasons (winter, spring, summer and fall). The
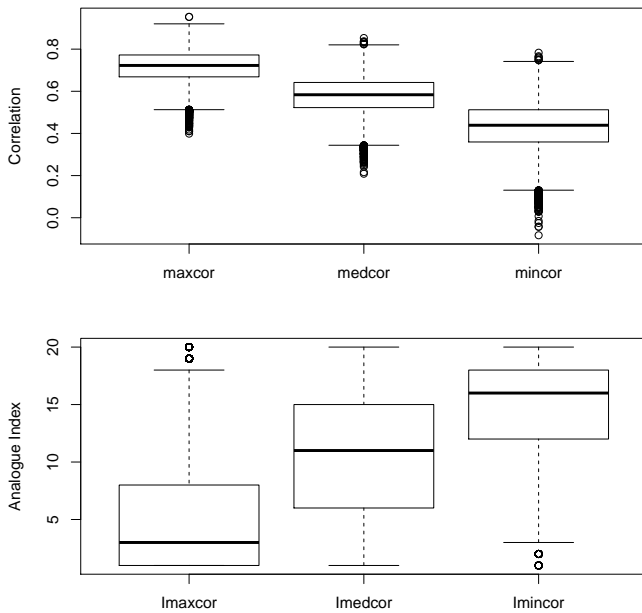
Fig. 2. Upper panel: distribution of maximum (maxcor), median (medcor) and 20th (mincor) rank spatial correlations of the 20 analogues. Lower panel: analogue rank (between 1 and 20) for maximum (lmaxcor), median (lmedcor) and 20th (lmincor) spatial rank correlations.

RMS values exhibit a seasonal cycle, with higher values in the Fall and Winter. This is due to a higher variance of SLP in the cold seasons than in the warm seasons. The correlation values also yield a seasonal cycle, albeit with a smaller relative amplitude. High correlations are found in the winter, and lower correlations appear in the summer. This is explained by less contrasted spatial patterns and a lower signal to noise ratio in the summer than in the winter, so that the average RMS can have a small value, but the spatial patterns can be shifted for the analogues. This was explained by Yiou et al. (2012).

In the following, for each day j, the set J of $K = 20$ analogues days is written:

$$J = \{\hat{j}^k, k \in [1, K]\}, \tag{5}$$

with (decreasing) RMS values:

$$D = \{d^k, k \in [1, K]\}, \tag{6}$$

and spatial correlation values:

$$C = \{c^k, k \in [1, K]\}. \tag{7}$$

The distribution of the maximum, median and minimum correlations of the 20 analogues is shown in Fig. 2 (upper panel). This shows that the highest correlation among the 20 analogues exceeds $r = 0.6$ in 75 % of cases, and the minimum correlation is significantly positive in 75 % of cases.

For each day, the analogues with the maximum, median and minimum correlations are determined. Their ranking according to RMS value is shown in Fig. 2 (lower panel). This shows that the analogues with high, median or low correlations roughly yield low, medium and high RMS values, although correlations and RMS values are generally not correlated in time, because their analytical relationship yields time-varying factors ((e.g. Murphy and Epstein, 1989). This ranking relation does not depend on the season (not shown).

In practice, the computation of circulation analogues is done once. The weather generator is written in the R language. It can be accelerated by parallelization, because each analogue computation is done independently from the others. It produces a multi-column text file. Each line represents a day between 1 January 1948 and 31 December 2012. The first column is the date of the target day. The 20 following columns are for the dates of analogues. The 20 following columns are for the RMS values. The last 20 columns are for the Spearman spatial correlation values. It serves as input to the weather generator.

## 3 Generating random sequences from SLP analogues

The goal of the weather generator is to produce a random sequence of dates (between 1 January 1948 and 31 December 2012) with a temporal coherence. A random resampling of the calendar would not be sufficient because time continuity of the SLP field would be lost. Thus two methodologies are presented for creating random samples of dates from analogues in order to preserve time continuity of SLP. The rationale of those methodologies stems from dynamical system theory and ensemble weather prediction. The first methodology (called "static") samples an observed trajectory of SLP and shadows it from a random selection of analogues. The second methodology (called "dynamic") computes a new trajectory from a selected initial condition, with a constraint of staying on the underlying attractor.

3.1 Static weather generator

The weather generator selects random years (between 1948 and 2012). The goal is to generate ensembles of seasons of typically 90 days. The season to be simulated is written $\mathcal{S}$ (e.g. winter, spring, summer or fall). For each selected random year (y), the dates $j_{\mathcal{S}}$ in the season $\mathcal{S}$ are considered. Each day $j_{\mathcal{S}}$ is replaced by a random sample of $(j_{\mathcal{S}}, \hat{j}^1, \ldots, \hat{j}^K)$ with probabilities:

$$\boldsymbol{p} = (p^0, p^1, \ldots, p^K). \tag{8}$$

The value of $p_0 = \beta \alpha_1$ gives a probability of choosing $j_{\mathcal{S}}$, i.e. not perturbing the trajectory by an analogue. α_1 is a positive number controling the time persistence of the weather generator. The probabilities $\{p^1, \ldots, p^K\}$ are chosen to be proportional to the spatial correlation between analogue and
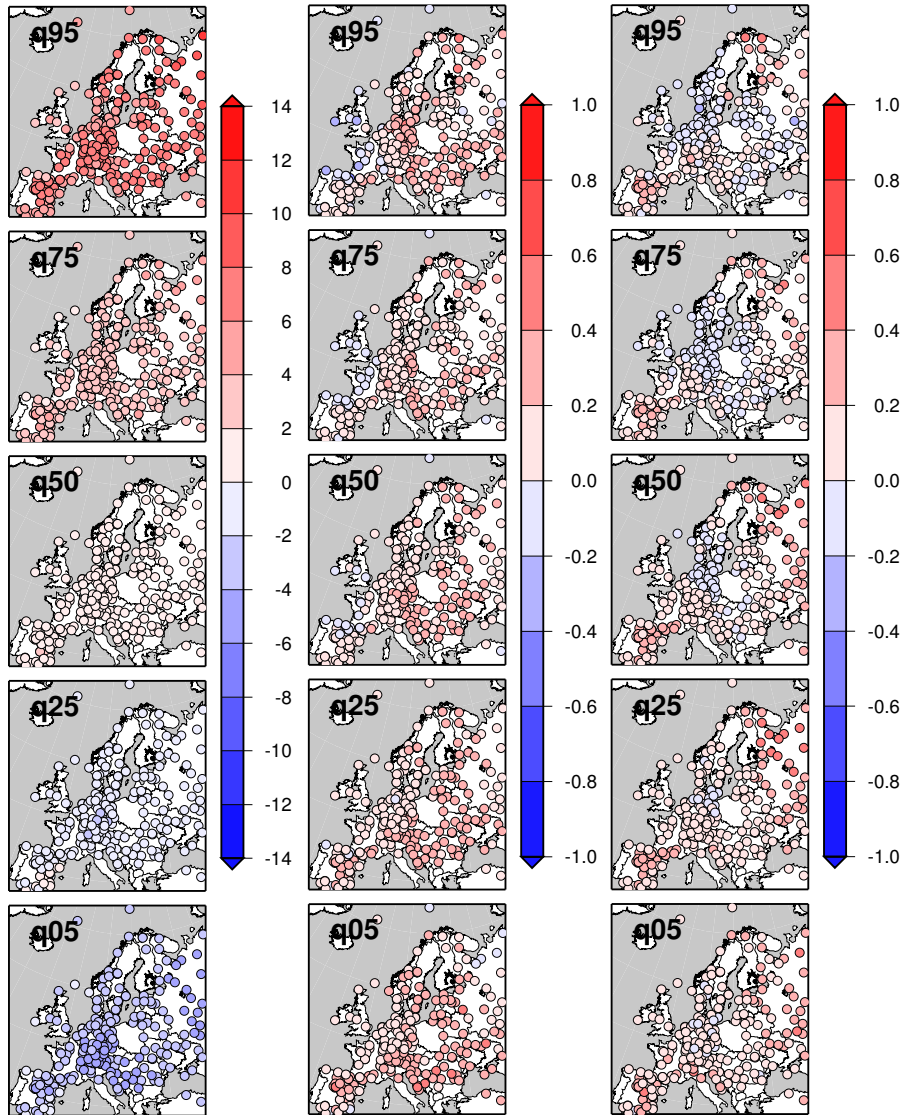
Fig. 3. Left column: 5th, 25th, 50th, 75th and 95th quantiles of summer temperature anomalies in Europe from the ECA&D set (in °C). Central column: differences of quantiles between observed and simulated temperature anomalies with "static" weather generator. Right column: differences of quantiles between observed and simulated temperature anomalies with "dynamic" weather generator.

observed SLP C:

$$p^k = \beta(1 + c^k)/2. \tag{9}$$

β is a normalization factor so that the sum of probabilities equals 1:

$$\sum_{k=0}^{K} p^k = \beta \left(\alpha_1 + \sum_{k=1}^{K} (c^k + 1)/2 \right) = 1. \tag{10}$$

This procedure randomly transforms observed trajectories with weights on "resembling" analogues: each day is perturbed independently of other days on the reference trajectory. It is called *static* because the transformed trajectory does not have the possibility of jumping to a very different

trajectory of the underlying climate attractor. This method can be used when one wants to "replay" a particular season or event, and assess uncertainties on temperature estimates during this season or event by computing random but likely surrogates of trajectories.

3.2 Dynamic weather generator

For each day (or initial condition), the next step of the trajectory is estimated, knowing that there is an uncertainty in the observation of the initial condition. Hence, the weather generator looks at the nearest neighbours (i.e. the analogues) of the initial condition and examine the trajectories emerging from those nearest neighbour initial conditions. The

proposed methodology assigns probability distributions to the nearest neighbours in order to compute random (but likely) trajectories of the system.

The generator is initialised by a random day $j_0 = 10^4 y_0 + 10^2 m_0 + d_0$. Let the day coming after j_0 be $\hat{j}$. This day has $K = 20$ analogues:

$$J = \{\hat{j}^k, k \in [1, K]\}, \tag{11}$$

with spatial correlation values:

$$C = \{c^k, k \in [1, K]\}. \tag{12}$$

The weather generator chooses a random "next day" for j_0 among $\hat{j}$ and its analogues J. Hence a probability vector is assigned:

$$\boldsymbol{p} = (p^0, p^1, \ldots, p^K) \tag{13}$$

to those potential "next days". The most likely candidate should certainly be $\hat{j}$, so that p^0 is proportional to a high value α_1. The value of α_1 controls the persistence of the generator: if α_1 is too high, the generated sequence will mostly be consecutive days in a deterministic fashion. The probabilities $\{p^1, \ldots, p^K\}$ are chosen to be proportional to the spatial correlation and the calendar distance between $\hat{j}^k$ and j_0. This condition ensures an average seasonal cycle in the simulated series of dates. Hence, the probabilities $\{p^1, \ldots, p^K\}$ are taken as:

$$p^k = \beta(c^k + 1) \exp\left(-\alpha_2 \delta(\hat{j}^k, j_0)\right). \tag{14}$$

The α_2 parameter controls the weight given to the calendar proximity of the analogues. If α is large, only analogues that have calendar dates close to the one of j_0 will be chosen. The β parameter is determined so that the sum of probabilities equals 1:

$$\sum_{k=0}^{K} p^k = \beta\left(\alpha_1 + \sum_{k=1}^{K} (c^k + 1) \exp\left(-\alpha_2 \delta(\hat{j}^k, j_0)\right)\right) = 1. \tag{15}$$

From the vector of probabilities $\boldsymbol{p}$, one "next" date j_1 for j_0 is sampled. The operation is then repeated for the desired number of iterations.

The free parameters of the weather generator are α_1 (persistence) and α_2 (seasonality). By default, the values are $\alpha_1 = 0.5$ and $\alpha_2 = 4$. By construction, a positive α_2 ensures that a seasonal cycle in the simulations if one is interested in simulating long time series (and not just a large ensemble of seasons). This parameter also constrains the dynamic weather generator to flow "forward" in time, because analogue dates occurring far away from the desired calendar date have a very low probability of being drawn.

This operation can be repeated for an arbitrary number of time steps. The outcome of this simulation is a sequence $\boldsymbol{j}$ of dates of analogues:

$$\boldsymbol{j} = \{j_0, \ldots, j_N\}. \tag{16}$$

If one is interested in simulating weather conditions for a given season, one can initialise j_0 with a random year and a calendar day starting the season (e.g. 21 March, June, September or December) and let the weather generator run for 90 days and an arbitrary number of seasons.

This type of Monte Carlo simulation (simulating a large number of seasons) can be done in parallel, in order to increase the efficiency of the computation. The weather generator code has been tested on a computing server with 2 to 8 CPUs.

The goal of this method for exploiting analogues of circulation is to generate potentially new sequences of (already observed) weather patterns. For example, if the weather generator is initialised with SLP conditions at the beginning of the summer of 2003, it is possible to assess the probability of observing a major European heatwave by repeating simulations. The weather generator hence works like a seasonal climate prediction, with a very large ensemble. The provided code in R is not configured to do an actual seasonal prediction.

4 Simulation of European temperatures

In this section, we are interested in simulating mean daily temperature anomaly variations at a given location, or a set of locations in western Europe. The goal is to combine existing observations and the sequence of dates produces by the random analogues.

4.1 Composites of temperatures from analogues

We want simulate random temperature *anomalies* T with respect to a seasonal cycle, at a given location that are coherent with the large scale information given by the sequence $\boldsymbol{j} = (j_0, \ldots, j_N)$ obtained in Sect. 3. It is assumed that there are daily observations T_j during the reanalysis period ($j \in [1$ January 1948, 31 December 2012]). The simulation of temperature variations simply considers the set of temperatures $\hat{T}$:

$$\hat{T} = (T_{j_0}, \ldots, T_{j_N}). \tag{17}$$

Therefore, composite temperatures for a random selection of analogues are determined.

The advantage of this approach appears when one wants to simulate temperature at several stations. By construction, the local temperature simulations are consistent with large scale SLP on daily time scale. This implies that for each day, the simulated temperatures at two or more locations are coherent with each other. This can be achieved with models of the multivariate dependence of several time series (Schölzel and Friederichs, 2008; Naveau et al., 2009; Bonazzi et al., 2012). But such a model needs to be re-evaluated if one set of observations is added or subtracted. Here, the spatial dependence structure is provided by the SLP analogues.
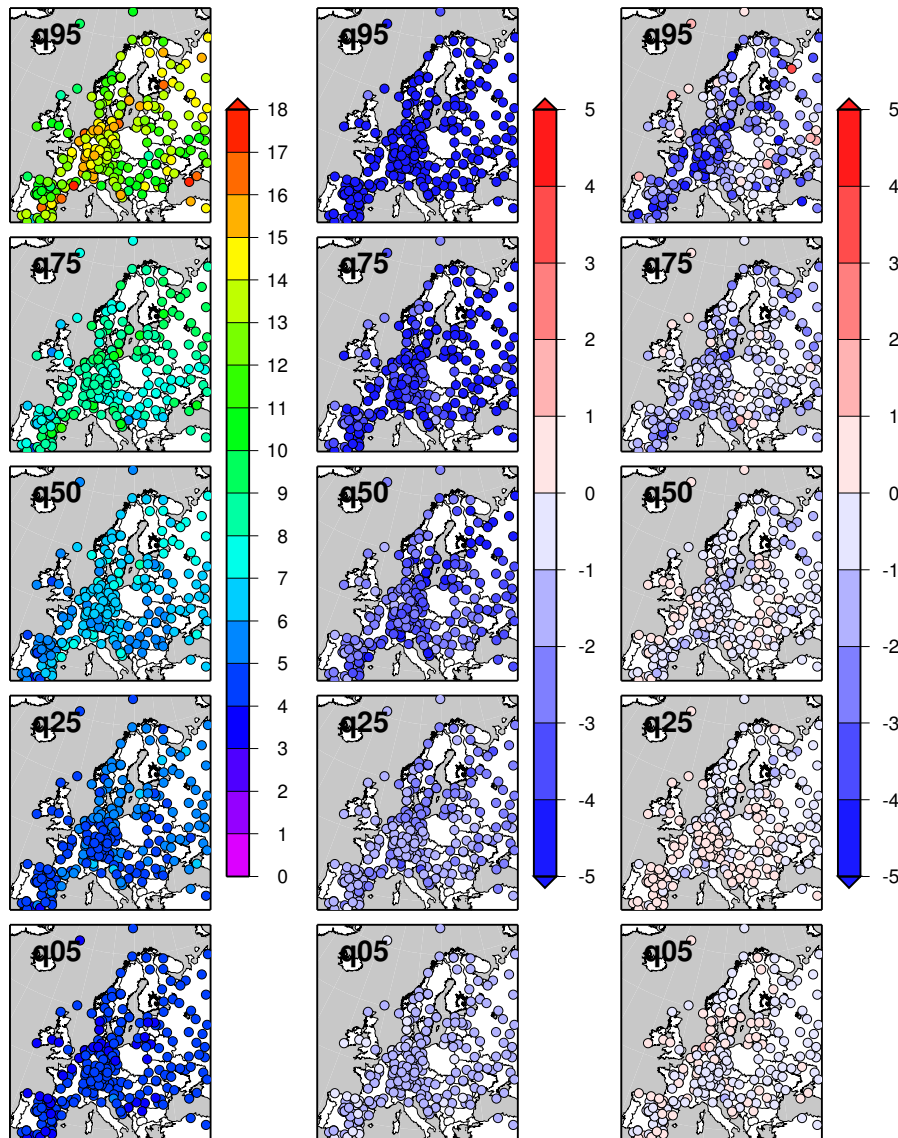
Fig. 4. Left column: 5th, 25th, 50th, 75th and 95th quantiles of summer τ (first non-positive autocorrelation lag) in ECA&D observations (in days). Central column: differences of quantiles of τ between observed and simulated temperature anomalies with "static" weather generator. Right column: differences of quantiles of τ between observed and simulated temperature anomalies with "dynamic" weather generator.

With this simple first procedure, the values of $\hat{T}$ are drawn from the values of the observations. What changes is the sequence of values. For an $N = 90$ day season, the number of possibilities for a simulated trajectory is of the order of $K^N > 10^{117}$ if $K = 20$ analogues are used. If persistence constraints of a few days are imposed, with large values of α_1, this still leaves a large number of possibilities.

In summary, the weather generator for temperature proceeds in four steps:

1. Read SLP analogues and Pareto parameters for temperature at selected locations.

2. Simulation of random dates from SLP analogues (static or dynamic).

3. Computation of temperatures for simulated dates for selected locations.

Time series of climate variables can be saved in various formats. By default, the native R binary format is used for output.

4.2 Data

In principle, the weather generator can simulate temperatures for any location, provided that it yields observations.
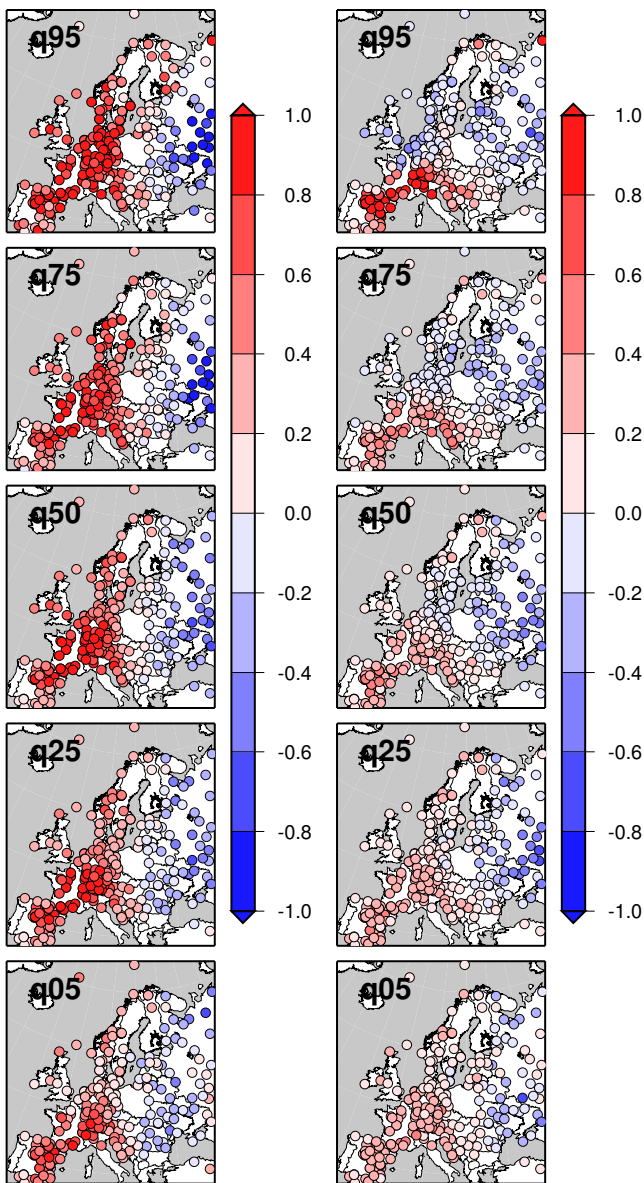
Fig. 5. Left column: differences of quantiles between observed and simulated temperature anomalies with "static" weather generator initiated on 21 June 2003 (in °C). Right column: differences of quantiles between observed and simulated temperature anomalies with "dynamic" weather generator with a 21 June 2003 initialisation.

We opted to focus on European temperatures. The European Climate Assessment and Data (ECA&D) dataset provides a regularly updated set of observations done by meteorological services over Europe (Klein-Tank et al., 2002). Time series are provided on a daily timescale. The data have been homogenized and quality checks were performed by the data providers. Data and metadata available at http://www.ecad.eu.

A subset of 291 series from the 1872 average temperature time series of ECA&D (TG) was selected. Time series starting before 1948 and ending after 2012 (hence covering the NCEP reanalysis period) were chosen. Stations for which more than 10 % of data are missing or doubtful were removed. This left 291 time series over Europe, with a high density of stations in Germany. For each time series, a seasonal cycle was computed by averaging over calendar days between 1971 and 2000. The seasonal cycle was then smoothed by a spline (smooth.spline function in R) with 9 degrees of freedom. The seasonal cycle was removed to daily temperature values in order to obtain temperature anomalies.

5 Metrics and bias estimates

Here the weather generator for summer and winter temperatures in Europe is tested. The weather generator is run for 100 winters and summers of 90 days. Two sets of experiments were performed for each season. The first one set is initialised with a random year between 1948 and 2011. Such an experiment tests the climatological features of the weather generator. The second type of experiment initialises the weather generator from years that have experienced extreme temperatures, with hot summers and cold winters. The prototype year for hot summer is 2003 (Schaer et al., 2004). The prototype years for cold winters is 2009. Such experiments test the ability to simulate extreme temperatures (Cattiaux et al., 2010).

The average daily mean temperature (TG) anomaly was simulated for all 291 stations. The 5th, 25th, 50th, 75th and 95th quantiles of temperature were computed for the observed time series (between 1948 and 2011) and the simulated time series. The comparison of quantiles allows one to verify the probability distribution induced by the weather generator.

The autocorrelation function was also computed for the observed and simulated time series of temperature for each 3 month season. At lag 0, the autocorrelation function is 1 (by construction). It tends to 0 when the lag tends to infinity. The first lag time τ for which the autocorrelation is no longer significantly positive is considered. This lag τ provides a measure of the persistence of temperature variations. The value of τ is computed for each simulated season. It is then possible to compare the quantiles of τ for the weather generator simulations and the observed time series.

5.1 Summer temperatures

In those sets of experiments, the summers were initiated on the 21st of June. The five quantiles (5, 25, median, 75 and 95) of temperature anomalies from the ECA&D database are shown in Fig. 3 (left column), for reference. 90 % of temperature anomaly values range between −6 and 10 °C. For each quantile, the differences between observations and
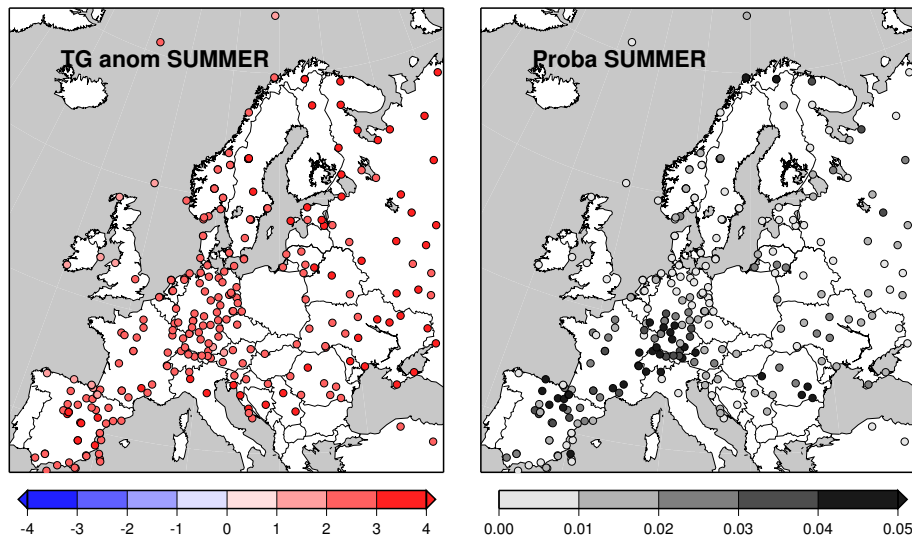
Fig. 6. Left panel: observed mean summer (June to August: JJA) temperature anomaly (in °C) in Europe for the 6 hottest summers between 1948 and 2012 (upper 90th quantile of mean JJA temperature). Right panel: probability of exceeding the 90th quantile of mean JJA temperature in 100 simulations of the "dynamic" weather generator.

simulations are shown in Fig. 3 (central and right column) for the static and dynamic weather generators.

The static weather generator has a generally slight warm bias (< 0.6 °C). The bias is less than 0.2 °C in France or Great Britain.

The bias for the dynamic weather generator is slightly positive for the lower quantiles (< 0.4 °C). It yields small negative values (< 0.2 °C), especially in Germany, for the upper quantiles.

The extreme summer conditions were simulated with initialisations on 21 June 2003. The quantile differences are shown in Fig. 5. The static weather generator, by construction, simulates high temperature differences for all quantiles, especially for France. This is to be expected because such simulation only alters each day of summer 2003 with analogue SLP. During the summer of 2003, the weather patterns were mostly anticyclonic, and caused the major observed heatwave in Western Europe (Cassou et al., 2005).

The dynamic simulations yield more moderate temperature positive anomalies in Western Europe, although the anomalies have higher values for the upper quantiles. This means that not all synoptic conditions resembling those at the beginning of the summer 2003 lead to a major heatwave. This was the case, for instance, for the year 2005 in Europe, which had similar weather patterns as 2003 at the end of June, but did not reach a heatwave at the middle of the summer. This result however suggests that if a summer starts like the one of 2003, it is likely that it will be warmer than usual, although cool conditions can occur too.

From the set of 100 dynamic experiments starting on 21 June 2003, the stationwise probability of simulating a summer with an average temperature exceeding the 90th quantile of observed mean temperature between June and August over Europe since 1948 is computed (this corresponds to the 6th hottest summer). It is found that this probability lies around 1 % in Europe. This probability exceeds 3 % in Spain, Eastern France, Switzerland and Germany (Fig. 6). This means that, although the simulated temperatures are on average warmer than usual, the probability of obtaining an extremely warm summer is small. This test is very conservative, because the radius of European heatwaves is less than 1000 km, and the criterion used here considered the whole of Europe.
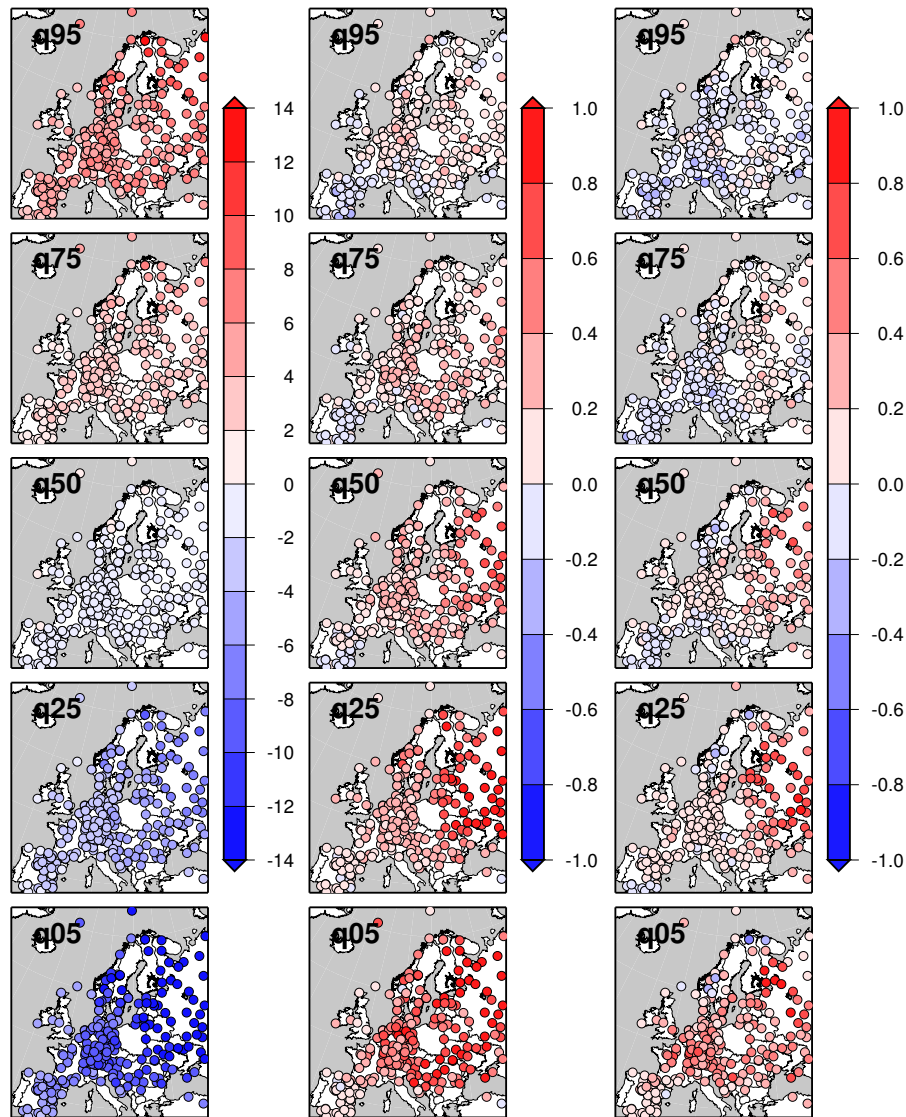
5.2 Winter temperatures

In those sets of experiments, the summers were initiated on the 21st of December. The five quantiles (5, 25, median, 75 and 95) of temperature anomalies from the ECA&D database are shown in Fig. 7 (left column), for reference. 90 % of temperature anomaly values range between -14 and 10 °C. For each quantile, the differences between observations and simulations are shown in Fig. 7 (central and right column) for the static and dynamic weather generators.

The static weather generator has a generally warm bias (< 1 °C) (Fig. 7, central column). The bias is less than 0.6 °C in France or Great Britain. This warm bias is larger for the extremely low quantiles (5th quantile), especially in central Europe. The bias over Western Europe for quantiles above the 25th are generally smaller than 0.2 °C.

The dynamic weather generator also yields a positive bias for temperature under the 25th quantile (Fig. 7, right column). This bias is lower than 0.2 °C above the 25th quantile.
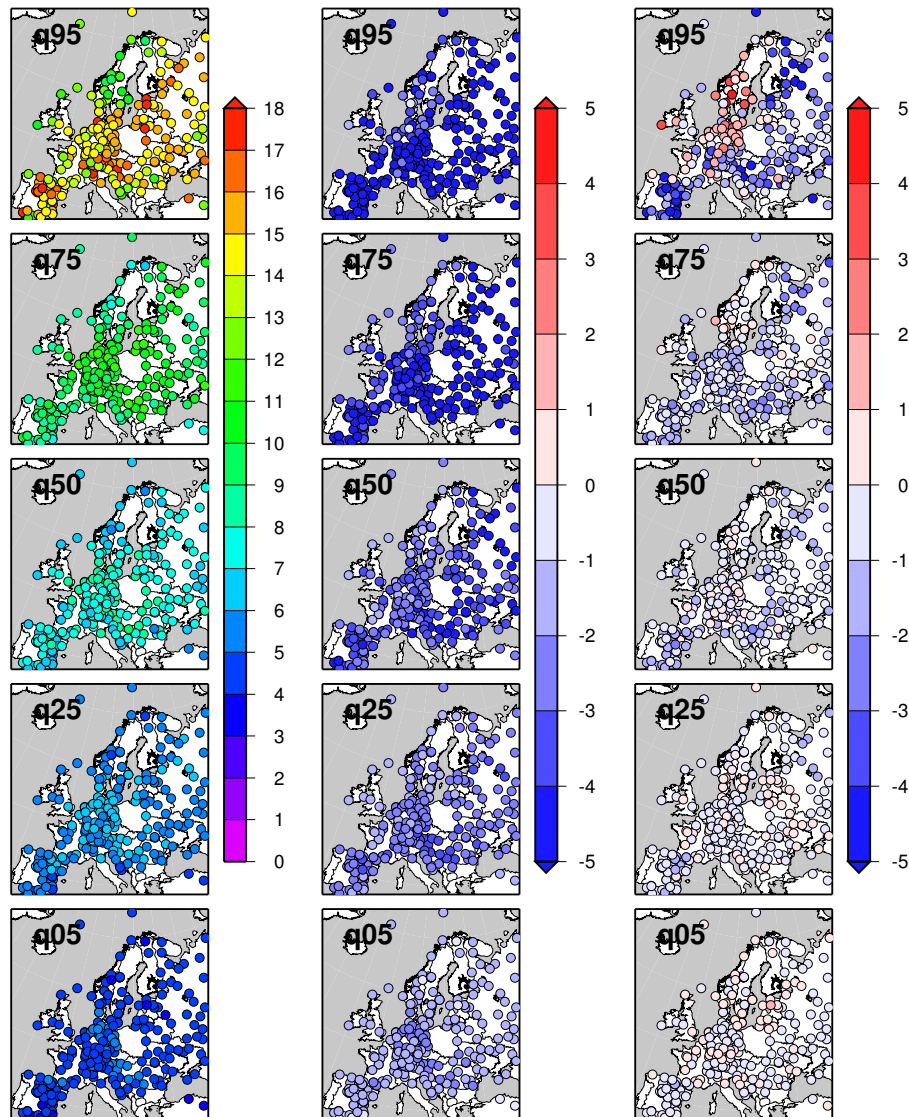
Fig. 7. Left column: 5th, 25th, 50th, 75th and 95th quantiles of winter temperature anomalies in Europe from the ECA&D set (in °C). Central column: differences of quantiles between observed and simulated temperature anomalies with "static" weather generator. Right column: differences of quantiles between observed and simulated temperature anomalies with "dynamic" weather generator.

The extreme winter conditions were simulated with initialisations on 21 December 2009. The quantile differences are shown in Fig. 9. The static weather generator, by construction, simulates highly negative temperature differences for all quantiles, especially for France. This is to be expected because such simulation only alters each day of winter 2009/2010 with analogue SLP. During the winter of 2009/2010, the weather patterns were locked to a negative phase of the North Atlantic Oscillation, and caused the major observed cold spell in Western Europe (Cattiaux et al., 2010; Cohen et al., 2010).

The dynamic simulations yield more moderate temperature negative anomalies in Western Europe (Fig. 9, right column). The temperature differences are more negative for

northern Europe (incl. Germany and Great Britain). When they are positive (e.g. in France for the lower quantiles), the quantile differences are smaller than for the climatological simulations in Fig. 7, right column. This implies that the simulated temperatures starting in December 2009 are colder than the ones obtained from a random year. This also suggests that a winter that starts like the 21 December 2009 is likely to be colder than usual.

From the set of 100 dynamic experiments starting on 21 December 2009, the stationwise probability of simulating a winter with an average temperature below the 10th quantile of observed mean temperature between December and February over Europe since 1948 is computed (this corresponds to the 6th coldest winter). It is found that this

Fig. 8. Left column: 5th, 25th, 50th, 75th and 95th quantiles of winter τ (first non-positive autocorrelation lag) in ECA&D observations (in days). Central column: differences of quantiles of τ between observed and simulated temperature anomalies with "static" weather generator. Right column: differences of quantiles of τ between observed and simulated temperature anomalies with "dynamic" weather generator.

probability lies around 1 % in Europe. This probability lies between 1 and 3 % in northern Spain, France, Switzerland and Germany (Fig. 10). This means that, although the simulated temperatures are on average colder than usual, the probability of obtaining an extremely cold winter is small.

6 Conclusions and perspectives

A weather generator based on analogues of atmospheric circulation is presented in this paper. The main feature of this weather generator (AnaWEGE) is that it can simulate meteorological variables at a set of locations (in Europe) and achieve a natural spatial coherence due to physical

relationships between large scale and small scale variability. AnaWEGE is well adapted to simulate seasons and was tested for winter and summer. The constraints of the dynamical generator (the α_1 and α_2 parameters) ensure that a seasonal cycle is obtained if a long continuous time series is desired.

AnaWEGE yields static and dynamic modes, and can serve two different purposes:

– the generation of ensembles of random perturbations of observed climate trajectories. This is useful for generating large catalogues of events (e.g. heatwaves or coldspells). In terms of regional climate simulation,
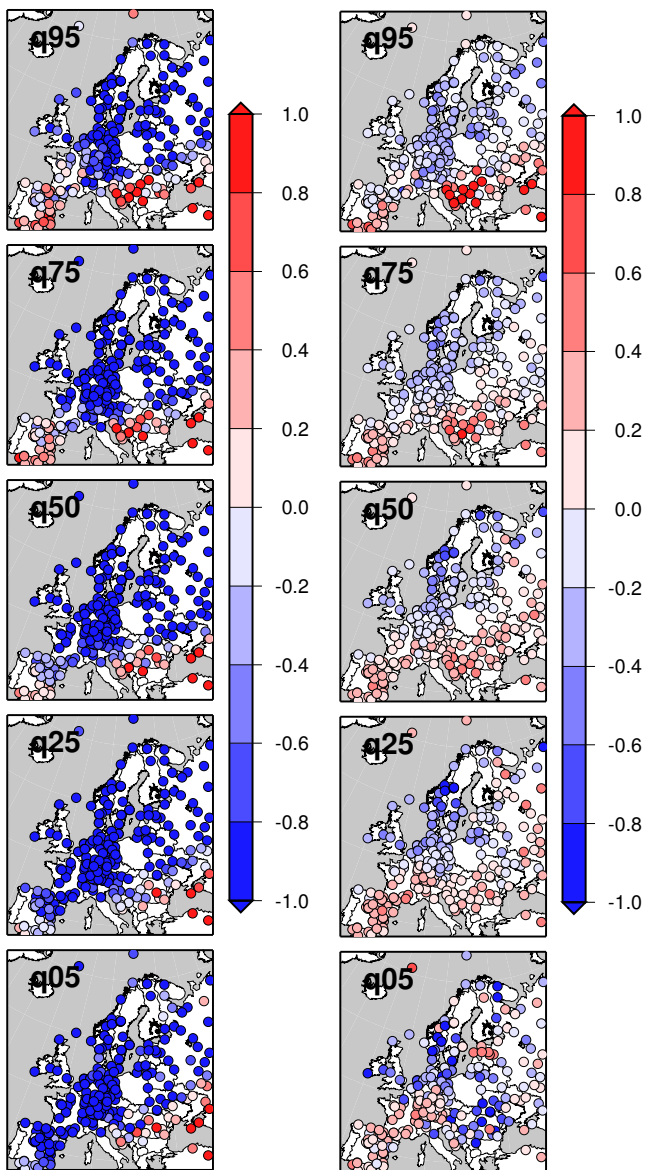
Fig. 9. Left column: differences of quantiles between mean observed and simulated temperature anomalies with "static" weather generator for the winter of 2009/2010 (initiated on the 21 December 2009). Right column: differences of quantiles between mean observed and simulated temperature anomalies with "dynamic" weather generator with a 21 December 2009 initialisation.

this corresponds to a *nudging* procedure with observed large scale conditions.

– the generation of ensembles of trajectories from given initial conditions. This is useful for assessing probability distributions of events, for instance by choosing initial conditions preceding the events. This feature is similar to a numerical weather forecast, although it uses an already computed reanalysis dataset as a basis (or could use any model simulation). Such an option

is a very cheap alternative (albeit with no physical constraint) to real atmospheric model simulations, although large ensembles can be achieved without the use of a supercomputer.

AnaWEGE was tested on European surface temperatures, from the ECA&D data (Klein-Tank et al., 2002), for which a dedicated computation of daily anomalies is provided. The rationale for focusing on temperature was to provide a tool to estimate background temperature extremes (especially in winter and summer), for European energy providers. The weather generator can be extended to simulate other climate variables (such as precipitation or wind speed), provided that time series of observations on the same time span as the set of circulation analogues is available. The analogues of circulation yield good skill for European precipitation (Vautard and Yiou, 2009) and geopotential height (Yiou et al., 2012).

This weather generator can serve as a basis for more sophisticated weather generators, which can add layers of randomness over the values that it generates. For example, one can rectify the values of a simulated climate variable by the weather generator by a random variable which yields similar statistical properties. For example, values of a climate variable exceeding a chosen threshold can be replaced by a simulation of a Pareto distribution (e.g. Vrac and Naveau, 2008; Bonazzi et al., 2012).

An underlying hypothesis of the weather generator is a stationary climate, in order to simulate stationary time series, which is certainly not true for observed European temperature, although the temperature trend ($\approx 0.5\,°C$ in 50 yr) is lower than the intra-seasonal and interannual variability.

Weather generators have been used to downscale climate variables in simulations of future climates (Carter, 1996; Iizumi et al., 2012). The weather generator presented here can be used in such a configuration once circulation analogues are computed for scenario simulations (Taylor et al., 2012), provided that their SLP output is available on daily time increments.

The computer performance of AnaWEGE might not be as high as already existing ones (Mavromatis and Hansen, 2001; Huth et al., 2001; Hansen et al., 2006; Semenov and Barrow, 1997; Flecher et al., 2010). It takes $\approx 2\,min$ to make 100 simulations of an 90 day season on a computer with two processors, with the parallel option. AnaWEGE requires packages (snowfall for parallel computing and evd for optional Pareto distributions) that are available on the R web site (cran.r-project.org/). The source code, input data files and a rudimentary user manual of version 1.0 can be downloaded at: http://www.lscedods.cea.fr/AnaWEGE/.

It is designed for scientific research (no gui interface) and the parameters can be changed easily. The season_sim_v1.R file is a wrapper to initialise and run the weather generator. Computer system path parameters need to be adapted for each user. The data files (analogues and mean daily temperature anomalies over Europe) are provided for an immediate
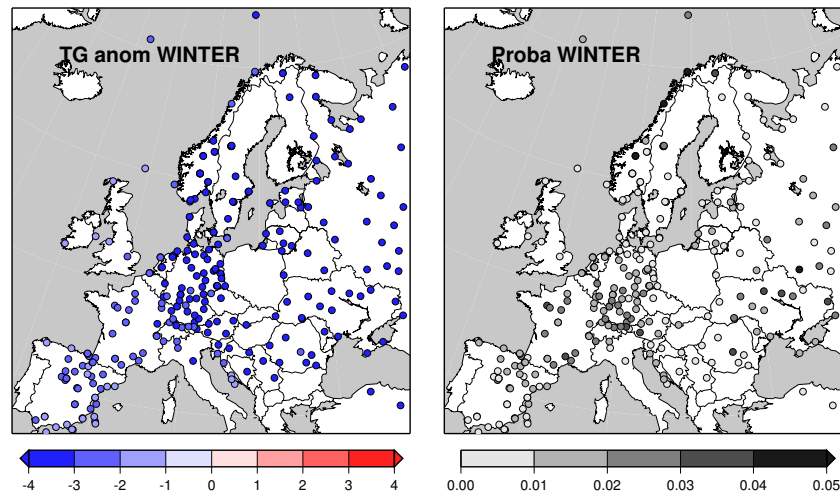
Fig. 10. Left panel: Observed mean winter (December to February: DJF) temperature anomaly (in °C) in Europe for the 6 coldest winters (lower 10th quantile of mean DJF temperature). Right panel: Probability of being below the 10th quantile of mean DJF temperature in 100 simulations of the "dynamic" weather generator.

use of the weather generator. The weather generator is hence a very versatile tool, especially if one generates files of analogues from other reanalyses or model simulations, and other climate variables (from other sources).

Edited by: W. Hazeleger

References

Bissolli, P. and Dittmann, E.: The objective weather type classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations, Meteorologische Zeitschrift, 10, 253–260, doi:10.1127/0941–2948/2001/0010–0253, 4th Mettools Conference, Stuttgart, Germany, APR 03-05, 2000, 2001.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S.: The spatial structure of European wind storms as characterized by bivariate extreme-value Copulas, Nat. Hazards Earth Syst. Sci., 12, 1769–1782, doi:10.5194/nhess-12-1769-2012, 2012.

Busuioc, A. and von Storch, H.: Conditional stochastic model for generating daily precipitation time series, Clim. Res., 24, 181–195, 2003.

Carter, T. R.: Developing scenarios of atmosphere, weather and climate for northern regions, Agr. Food Sci. Finland, 5, 235–249, 1996.

Cassou, C., Terray, L., and Phillips, A. S.: Tropical Atlantic influence on European heat waves, J. Climate, 18, 2805–2811, 2005.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: A cold extreme in a warming climate, Geophys. Res. Lett., 37, L20704, doi:10.1029/2010gl044613, 2010.

Cohen, J., Foster, J., Barlow, M., Saito, K., and Jones, J.: Winter 2009-2010: A case study of an extreme Arctic Oscillation event, Geophys. Res. Lett., 37, L17707, doi:10.1029/2010gl044256, 2010.

Flecher, C., Naveau, P., Allard, D., and Brisson, N.: A stochastic daily weather generator for skewed data, Water Resour. Res., 46, W07519, doi:10.1029/2009wr008098, 2010.

Hansen, J., Challinor, A., Ines, A., Wheeler, T., and Moron, V.: Translating climate forecasts into agricultural terms: advances and challenges, Clim. Res., 33, 27–41, 2006.

Hurrell, J., Kushnir, Y., Ottersen, G., and Visbeck, M. (Eds.): The North Atlantic Oscillation : Climatic Significance and Environmental Impact, vol. 134 of Geophysical monograph, American Geophysical Union, Washington, DC, 2003.

Huth, R., Kysely, J., and Dubrovsky, M.: Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series, J. Climate, 14, 4047–4061, 2001.

Iizumi, T., Takayabu, I., Dairaku, K., Kusaka, H., Nishimori, M., Sakurai, G., Ishizaki, N. N., Adachi, S. A., and Semenov, M. A.: Future change of daily precipitation indices in Japan: A stochastic weather generator-based bootstrap approach to provide probabilistic climate information, J. Geophys. Res.-Atmos., 117, D11114, doi:10.1029/2011jd017197, 2012.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77, 437–471, 1996.

Klein-Tank, A., Wijngaard, J., Konnen, G., Bohm, R., Demaree, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-

Hansen, C., Heino, R., Bessemoulin, P., Muller-Westermeier, G., Tzanakou, M., Szalai, S., Palsdottir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Lopez, J., Dahlstrom, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, Int. J. Climatol., 22, 1441–1453, 2002.

Kreienkamp, F., Spekat, A., and Enke, W.: The Weather Generator Used in the Empirical Statistical Downscaling Method, WET-TREG, Atmosphere, 4, 169–197, doi:10.3390/atmos4020169, 2013.

Lorenz, E. N.: Atmospheric Predictability as Revealed by Naturally Occurring Analogues, J. Atmos. Sci., 26, 636–646, 1969.

Mahalanobis, P. C.: On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India, 2, 49–55, 1936.

Mavromatis, T. and Hansen, J.: Interannual variability characteristics and simulated crop response of four stochastic weather generators, Agr. Forest Meteorol., 109, 283–296, 2001.

Murphy, A. and Epstein, E.: Skill scores and correlation-coefficients in model verification, Mon. Weather Rev., 117, 572–581, doi:10.1175/1520–0493(1989)117<0572:SSACCI>2.0.CO;2, 1989.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J.: Modelling pairwise dependence of maxima in space, Biometrika, 96, 1–17, 2009.

Schaer, C., Vidale, P., Luthi, D., Frei, C., Haberli, C., Liniger, M., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, Nature, 427, 332–336, 2004.

Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, Nonlin. Processes Geophys., 15, 761–772, doi:10.5194/npg-15-761-2008, 2008.

Schoof, J. T. and Robeson, S. M.: Seasonal and spatial variations of cross-correlation matrices used by stochastic weather generators, Clim. Res., 24, 95–102, 2003.

Semenov, M. A. and Barrow, E. M.: Use of a stochastic weather generator in the development of climate change scenarios, Climatic Change, 35, 397–414, 1997.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bull. Am. Meteorol. Soc., 93, 485–498, 2012.

Van den Dool, H.: Empirical Methods in Short-Term Climate Prediction, Oxford University Press, Oxford, iSBN-10: 0-19-920278-8 ISBN-13: 978-0-19-920278-2, 2007.

Van den Dool, H. M.: Searching for analogs: how long must we wait?, Tellus A, 46, 314–324, 1994.

Vautard, R. and Yiou, P.: Control of recent European surface climate change by atmospheric flow, Geophys. Res. Lett., 36, L22702, doi:10.1029/2009GL040480, 2009.

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfall (vol 43, art no W07402, 2007), Water Resour. Rese., 44, W05702, doi:10.1029/2008wr007083, 2008.

Westra, S. and Sharma, A.: Probabilistic Estimation of Multivariate Streamflow Using Independent Component Analysis and Climate Information, J. Hydrometeor., 10, 1479–1492, doi:10.1175/2009JHM1121.1, 2009.

Wilks, D. S.: Multisite downscaling of daily precipitation with a stochastic weather generator, Clim. Res., 11, 125–136, 1999.

Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R., and Vrac, M.: Ensemble reconstruction of the atmospheric column from surface pressure using analogues, Clim. Dynam., 41, 1333–1344, doi:10.1007/s00382-012-1626-3, 2013.

Zorita, E. and von Storch, H.: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, J. Climate, 12, 2474–2489, 1999.