The replies to referees are indicated in red. A track change version of the manuscript follows the point by point response to reviewers. This file was created with latexdiff. It does not include the new figures that appear in the revised version. We thank the reviewers for pointing out unclear points in the manuscript.

Referee#1

In this paper, authors use a low-cost stochastic analogue forecasting method to predict the NAO index and ground temperatures in specific locations. The idea is the following: find 20 analog situations using the sea level pressure at time t, randomly choose 1 of the 20 analogs (using a proper distance), take the corresponding successor to make the prediction at t+1, apply the same procedure until lead time t+T. Authors repeat this statistical forecast and obtain a stochastic ensemble forecast of 100 simulated trajectories. The method is original and have good performance compared to classic ones, using persistence or climatology. The introduction is very clear and is a good summary of stochastic weather generators and analog methods. However, quality of the figures needs to be improved.

Specific comments:

• The stochastic analog forecast presented here is a nonparametric approach (in a statistical sense). At some points, the reader would like to have a comparison with simple parametric methods like an autoregressive model, building a linear regression between the SLP at time t and NAO index or ground temperature at time t+1. Another option is to build a local linear regression between the 20 analogs and 20 successors. In that case, the biases given highlighted in the q-q plots should be reduced and quality of the prediction should be improved. But the use of low-rank methods (like Partial Least Squares method) must be used. Note that using such parametric methods can also lead to stochastic forecasts, when randomly sampling on the distribution function (e.g., Gaussian with the estimated mean and covariance) of the successors.

This is an interesting suggestion, albeit quite unusual (with respect to the available literature). We built a multivariate autoregressive (mAR1) model on SLP. To simplify numerical problems, the mAR1 model is done on the first ten principal components of North Atlantic SLP (representing approx. 80% of the variance):

$$R_{t+1} = AR_t + B_t,$$

where R_t is a vector of 10 PCs, A is a 10×10"persistence" matrix, and B_t is a 10-variate Gaussian centered white noise with covariance matrix Σ.

The mAR1 coefficients A and Σ are determined from the covariance $C(0)$ and lag-1 covariance $C(1)$ matrices of SLP:

$$A = C(1)^t C(0)^{-1},$$
$$\Sigma = C(0) - C(1)^t A.$$

This procedure is similar to what was done by Michelangeli et al. (Weather regimes: Recurrence and quasi stationarity. J. Atmos. Sci., 52(8), 1237-1256. 1995) to simulate a multivariate AR1 process that mimics atmospheric geopotential heights.

Ensembles of mAR1 simulations can be performed, with initial conditions from observed values of SLP, at incremental times. This is similar to the analogue weather generator of SLP presented in the paper.

We performed a multilinear regression between the five temperature series and NAO index and the preceding values of SLP principal components:

$$X_t = [T_t^1, \ldots, T_t^5, NAO_t] = aSLP_t + b + e_t.$$

This multivariate regression is applied to the mAR1 model to perform ensembles of forecasts of temperatures and NAO index. For each realization, averages over lead times between 5 and 80 days are then performed. Such a simple model cannot reproduce a seasonal cycle

of temperature (unless it is explicitly added, which we did not do). Therefore, only comparisons on the warmest (July) or coldest (January) months would be meaningful (although weakly more meaningful). Such a problem does not occur with the NAO index, which does not yield a clear seasonality.
We computed the CRPSS and correlation of this stochastic model (mAR1). The skill scores always give negative (or non significantly positive) values with respect to references for temperature, due to the absence of seasonality in such a model. The skill scores for the NAO index are positive, and give an interesting background for the analogue model.
In addition (we had not mentioned it but we will in the revised text), a stochastic IID perturbation is always added to the reference (climatological, persistence) forecasts. This is necessary because we compare probability distributions. This is an even simpler first order parametric stochastic model.
This is now discussed in the manuscript, and the results with the mAR1 model are reported in the results section (new Figure 6)

• The quality of the figures needs to be significantly improved:
–Fig. 1, can you remove the 2nd map and put only the 5 points of interest in the 1st map?
OK. The 2nd panel was removed and the 5 stations were added on the bigger map (new Figure 1).

–Fig. 2, what do you mean by observed average. Is it really useful? Where are the median analog forecasts? Please use T instead of N in the legend.
It should have been “the average of observed temperatures TG between Jan. 1st 2007 to the lead time T”. This is the values that we try to forecast. The legend of the new figure 3 is changed (T rather than N). Thank you for pointing this out.

–Fig. 3, plot only 1 legend (for instance in the bottom left sub-figure)? Be careful with the y-label on the right sub-figures.
OK. The legends were removed, and grouped in an additional panel (new figure 4).

–Fig. 4-5, authors should separate Jan and Jul in 2 sub-figures (not necessarily to plot "all"). Please connect the [squares, dots, triangles] between different lead times. Use a classic boxplot to represent error bars.
OK. The figures are splitted in two panels. The error bars represented the 95% confidence interval obtained from a usual formula on uncertainty on the correlation (see H. von Storch and F. Zwiers, Statistical Analysis in Climate Research, 1999, Cambridge University Press, sec. 8.2.3) between the median of forecasts and observed values. The new figures now represent the spread of correlations between realization members and observations with boxplots. The interpretation of confidence intervals is hence different (but the mean values are the same).

Technical corrections:

• Avoid the use of "dynamical" and use "dynamic" instead.
We keep the adjective “dynamical” when referring to “dynamical systems”. This is how it is used in textbooks, journal names, etc. The adjective was changed to “dynamic” when referring to the simulation mode of the stochastic weather generator.

• Can you explain the difference between "predictand" and "predictor"? Avoid the use of predictand?
Predictand is the variable that we want to predict. Predictor is the variable that is used to predict the predictand. There was a confusion p. 12, l. 32, which is now corrected.

• Can you remind the difference between positive and negative values of the NAO index?
A sentence is added to explain the pressure features during high and low values of the NAO index (p. 2, near l. 30).

Referee#2

In this manuscript the authors develop an ensemble forecasting system using an analog-based weather generator. They test this ensemble forecasting system for NAO and the temperature at several weather stations. They focus on the forecasts of temporal averages from 5 to 80 days. The forecast is made for each averaging period at the first corresponding lead time. The skill of these forecasts are evaluated through skill scores (the correlation and the continuous rank probability score, CRPS, the latter being well adapted to ensemble forecasts). The authors claim that there is some skill of the temperature and NAO up to seasonal time scales. I am not convinced by the system they propose, nor by the skill they found, for three important reasons:

(i) The system they propose suffers from a very important drawback, which is the progressive convergence toward the climatological mean as illustrated on the right column of Figure 3. There is only little variability of the forecasts for long time averages, indicating that the ensemble forecast is unreliable. This makes of this system a very poor probabilistic forecasting system, since the forecasts do not span the set of possible values of the observed variable. Reliability is one essential ingredient of ensemble forecasts that can also be easily checked with the decomposition of the CRPS in reliability and resolution. I therefore do not consider this ensemble system appropriate.
We never hide the fact that there is a convergence towards climatology (this is mentioned in the text). But long term forecasts with full scale climate models yield the same feature of convergence to climatology (as outlined by Hersbach (2000) and others). Our claim is that this system does a better job than usual references (Climatology or Persistence) or AR1 models (see response to referee #1). The ease of use of this system makes it possible at low cost to investigate the limit for large lead times. We consider that the fact that the scores are positive for shorter lead times (20 days ahead) is interesting. We now mention (and use) the CRPS decomposition of Hersbach (2000) in terms of reliability and potential CRPS.

(ii) It is not clear at all to me why the authors are looking at the first lead time of the 5, … 80 days averages. Using this approach, one can certainly expect that if one start from an initial state close to the reality, the forecast of the averages will always be better than the climatological average (provided we have access to an infinite sample). In other words some positive correlation will always be present, even if it is very small. This skill is artificial (due to averaging from the initial state) and I am wondering why the authors did not have looked at the skill of the daily values of NAO or temperatures. My guess is that there is no skill beyond a month or so.
We never claim the contrary and discussed it in the text. Starting from an observed state, daily trajectories tend to diverge from each other. The computing T-averages for various lead times allows accessing to the limit of predictability of our system.
This analogue method is also better than an autoregressive model (mAR1, see response to referee#1) that is initialized from observations. As stated in the text, we do not consider that the system has any skill beyond a month.

(iii) The analysis of the skill of ensemble forecasts should be done with appropriate tools. The CRPSS is one of them, but it is much more important to look at its decomposition in reliability, resolution and uncertainty. These are standard tools that can be found in classical books or papers (e.g. H. Hersbach, 2000, Weather and Forecasting, 15, 559-570).

Thank you for this suggestion. We added a discussion on the decomposition of CRPS (citing the paper of Hersbach 2000) in terms of reliability and potential CRPS. The problem with reliability is that its magnitude depends on the unit of the variable to be predicted (as discussed by Hersbach 2000). The results reported by Hersbach give very small values of reliability (for precipitation forecast) when the ECMWF analysis is used. But those numbers are small because the variable values to predicted are small.
We used the R package "verification" (by E. Gilleland) to compute this decomposition. The relative of variations of reliability that we obtain for temperature or NAO forecasts is in the same range of what is reported in the paper of Hersbach (2000) for lead times of 5 to 10 days. We now discuss the values of reliability, which appears in Figures 4-5. The reliability values for NAO are small ($\approx 8\ 10^{-3}$), and the ratio to the CRPS value is in the same range of what is reported in Hersbach's paper.

Some additional (less important) points

1. The algorithm of page 4 (section 3.2) is far from clear. It would be nice to visualize the algorithm, together with the relations that are used for evaluating the weights.
OK. A graphical illustration is added (new Figure 2) to visualize the iteration procedure and the choice of weights to sample analogues.

2. Page 6, line 5. Is S=N? This is not clear to me.
We compute the N=20 best analogues for each day. At each time increment, we simulate S=100 trajectories, sampled from those 20 best daily analogues. For a lead time of, say, 10 days, there are 20^{10} possible trajectories, which is far larger than S. This is emphasized in the text.

3. An additional concern I have is the comparison with the persistence in Figs 4 and 5. It seems to me that the observables based on persistence display a higher variability than the forecasts constructed here (that are converging to the climatology). I therefore suspect that the reliability of the persistent forecast is better than the one of the stochastic forecasts (the reliability term in the CRPS decomposition should be smaller for the persistence case), which is not reflected here in the analysis of the CRPSS. I would be very useful to evaluate the different terms of the CRPS to clarify the difference between the two systems. This will allow in particular to clarify why one gets 0.45
for all averages for NAO and why the skill increases for temperature.
A discussion on the CRPS decomposition for the different forecasts is added. The reliability value of CRPS for the persistence or the climatology give higher values (roughly twice larger) than for our model.

Stochastic Ensemble Climate Forecast with an Analogue Model

Pascal Yiou[1] and Céline Déandréis[2]

[1]Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL and Université Paris-Saclay, CE l'Orme des Merisiers, 91191 Gif-sur-Yvette, France
[2]ARIA Technologies, 8-10 Rue de la Ferme, 92100 Boulogne-Billancourt, France

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

Abstract. This paper presents a system to perform large ensembles climate stochastic forecasts. The system is based on random analogue sampling of sea-level pressure data from the NCEP reanalysis. It is tested to forecast ~~an NAO~~a North Atlantic Oscillation (NAO) index and the daily average temperature in five European stations. We simulated 100 member ensembles of averages over lead times from 5 days to 80 days in a hindcast mode, i.e. from a meteorological to a seasonal forecast. We tested the hindcast simulations with usual forecast skill scores (~~CRPSS~~ CRPS or correlation), against persistence and climatology. We find significantly positive skill scores for all time scales. Although this model cannot outperform numerical weather prediction, it presents an interesting benchmark that could complement climatology or persistence forecast.

1 Introduction

Stochastic weather generators (SWG) have been devised to simulate many and long sequences of climate variables that yield realistic statistical properties ~~Semenov and Barrow (1997)~~(Semenov and Barrow, 1997). Their main practical use has been to investigate the probability distribution of local variables such as precipitation, temperature or wind speed, and their impacts on agriculture (Carter, 1996; Semenov, 2006), energy (Parey et al., 2014) or ecosystems (Maraun et al., 2010). Such systems can simulate hundreds or thousands of trajectories on desktop computers and propose cheap alternatives to climate model simulations.

There are many categories of SWGs (Ailliot et al., 2015). Some SWGs are explicit random processes, whose parameters are obtained from observations of the variable to be simulated (Parey et al., 2014). Some SWGs are based on a random resampling of the observations (Iizumi et al., 2012). Some SWGs simulate local variables from their dependence to large-scale variables such as the atmospheric circulation (Kreienkamp et al., 2013). This allows to simulate spatially coherent multivariate fields (Yiou, 2014; Sparks et al., 2018) and can be used for downscaling (Wilks, 1999).

SWGs that use observations as input could in principle be used to forecast variables. This is the case for analogue weather generators (Yiou, 2014). Methods of analogues of atmospheric circulation were first devised for weather forecast (Lorenz, 1969; van den Dool, 1989). They were abandoned when numerical weather prediction was developed and implemented, be-

cause their performance was deemed inadequate (van den Dool, 2007). However, recent studies on *nowcasting* have shown that analogue based methods could outperform numerical weather prediction for precipitation (Atencia and Zawadzki, 2015). Yiou (2014) showed some skill for temperature simulations in Europe of an analogue SWG.

Due to uncertainties in observations and the high sensitivity to initial conditions (van den Dool, 2007) weather forecasts estimate probability density functions rather than deterministic meteorological values. Therefore, weather forecasts examine the properties of all possible trajectories of an atmospheric system from an ensemble of initial conditions. Such properties include the range and the median, for example. Then one can compare how the ensemble of trajectories compares to observations, and other reference forecasts. Numerical weather forecasts rely on large ensembles of model simulations and require a massive use of supercomputers in order to provide estimates of the probability density function (pdf) of variables of interest, for various lead times. Being able to increase the ensemble size of weather forecast systems in order to lower the bias of the forecast skill has been a challenge of major centers of weather prediction (Weisheimer and Palmer, 2014).

The most trivial prediction systems are based on either climatology (i.e. predicting from the seasonal average) or persistence (i.e. predicting from the past observed values) (Wilks, 1995). Probabilistic and statistical models can provide more sophisticated benchmarks for weather forecast systems, still without simulating the underlying primitive hydrodynamic equations and using supercomputers. For example, statistical models of forecast for precipitation based on analogues (of precipitation) were tested for North America (Atencia and Zawadzki, 2015). Such systems tend to outperform numerical weather forecast systems, although their computing cost is steeper than most SWGs. Therefore the potential of analogue based methods can be useful to assess probability distributions, rather than a purely deterministic forecast.

Machine learning algorithms were recently devised to simulate complex systems (Pathak et al., 2018a) with surprising performances. Such algorithms are sophisticated ways of computing analogues of observed trajectories in a learning step, and simulating potentially new trajectories from this learning. The main drawback is that such algorithms generally require a tricky tuning of parameters that might not be based on a physical intuition. From the inspiration of machine learning algorithms, we propose to devise a weather forecast system based on a stochastic weather generator that uses analogues of circulation to generate large ensembles of trajectories. The rationale for using analogues, rather than more sophisticated machine learning, is that they correspond to a physical interpretation of relations between large scale and regional scales. Moreover, mathematical results in dynamical system theory (Freitas et al., 2016; Lucarini et al., 2016) suggest that properties of recurring patterns is asymptotically independent on the distance that is used to compute analogues.

This paper presents tests of such a system to forecast temperatures in Europe and an index of the North Atlantic Oscillation (NAO). The NAO controls the strength and direction of westerly winds and location of storm tracks across the North Atlantic in the winter (Hurrell et al., 2003). ~~It is therefore~~ Positive values of the index indicate a strengthened Azores anticyclone and a weaker Icelandic low. Negative values indicate a weak Azores anticyclone and a strong Icelandic low. The North Atlantic Oscillation is strongly tied to temperature and precipitation variations in Europe (Slonosky and Yiou, 2001).

Since the set up of such a system is fairly light, it is possible to test it for time leads from a meteorological forecast (5 days ahead) to a seasonal forecast (80 days ahead). We test this system in hindcast experiments to forecast climate variables between 1970 and 2010. The tests are performed with usual skill scores (continuous rank probability score and correlation).

The paper is organized as follows. Section 2 presents the datasets that are used as input of the system. Section 3 presents the forecast system based on analogues, the skill scores and the experimental protocol. Section 4 presents the results on simulations of the NAO index and European temperatures.

2 Data

We used data from different sources for sea-level pressure (SLP), NAO index and temperatures. SLP data are used for analogue computations as predictor. The NAO index and temperatures are the predictands (i.e. variables to be predicted). It is important that they share a common chronology, in order to allow their simulation because the NAO index and temperatures are simulated from from SLP analogues.

2.1 Sea-level pressure

We use the reanalysis data of the National Centers for Environmental Prediction (NCEP) (Kistler et al., 2001). We consider the sea-level pressure (SLP) over the North Atlantic region. We used SLP daily averages between January 1st 1948 and April 30th 2018. The horizontal resolution is 2.5° in longitude and latitude. The rationale of using this reanalysis is that it covers more than 60 years and is regularly updated, which makes it a good candidate for a continuous time forecast exercise.

One of the caveats of this reanalysis dataset is the lack of homogeneity of assimilated data, in particular before the satellite era. This can lead to breaks in pressure related variables, although such breaks are mostly detected in the southern hemisphere and the Arctic regions (Sturaro, 2003). We are not interested in the evaluation of SLP trends, therefore breaks should only marginally impact our results.

2.2 NAO index

The North Atlantic Oscillation (NAO) is a major mode of atmospheric variability in the North Atlantic (Hurrell et al., 2003). Its intensity is determined by an index that can be computed as the normalized sea-level pressure difference between the Azores and Iceland (Hurrell, 1995). The NAO index is related to the strength and direction of the westerlies, so that high values correspond to zonal flows across the North Atlantic region, stormy conditions and rather high temperatures in Western Europe (Slonosky and Yiou, 2001; Hurrell et al., 2003).

We retrieved the daily NAO index from the NOAA web site:
http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml.

The procedure to calculate the daily NAO teleconnection indice is detailed on the NOAA web site. In short, a Rotated Principal Component Analysis (RPCA) is applied to monthly averages of geopotential height at 500 hPa (Z500) anomalies (Barnston and Livezey, 1987) in the 20N–90N region, between January 1950 and December 2000, from the NCEP reanalysis. The empirical orthogonal functions (EOFs) provide climatological monthly teleconnection patterns (Wilks, 1995). Those monthly teleconnection patterns are interpolated for every day in the year. Then daily Z500 anomaly fields are projected onto the interpolated climatological teleconnection patterns in order to obtain a daily NAO index.

Figure 1. Upper panel: North Atlantic region (blue rectangle) and Western European region (red rectangle) on which analogues are computed.~~Lower panel: European stations used for daily mean temperature.~~

The geographical domain on which this NAO index is computed is larger than the one for SLP data. Scaife et al. (2014) used an NAO index to test the UKMO seasonal forecast system. The index they used is based on monthly SLP differences between the Azores and Iceland, and is therefore different from ours.

2.3 European temperatures

We took daily averages of temperatures from the ECAD project (Klein-Tank et al., 2002). We extracted data from Berlin, De Bilt, Toulouse, Orly and Madrid (Fig. 1). Those five stations cover a large longitudinal and latitudinal range in western Europe. These datasets were also chosen because

- they start before 1948 and end after 2010. This allows the computation of analogue temperatures with the SLP from the NCEP reanalysis, which includes that period,
- they contain less than 10% of missing data.

These two criteria allow keeping 528 out of the 11422 ECAD stations that are available in 2018.

3 Methods

3.1 Analogues of circulation

Analogues of circulation are computed on SLP data from NCEP (Sec. 2.1). For each day between Jan. 1st 1948 and Dec. 31st 2017, the best 20 analogues (with respect to a Euclidean distance) in a different year are searched. This follows the procedure of (Yiou et al., 2013). The analogues are computed over two regions (large region: North Atlantic region (80W–30E; 30–70N); small region: Western Europe (30W–20E; 40–60N)). The large region is used to simulate/forecast the NAO index. This choice is justified by the fact that the North Atlantic atmospheric circulation patterns are well defined over that region (Michelangeli et al., 1995). The small region is used to simulate/forecast continental temperatures, following the domain recommendations of the analysis of Jézéquel et al. (2018).

3.2 Forecast with analogue stochastic weather generator

Ensembles of simulations of temperature or the NAO index can be performed with the rules illustrated by Yiou (2014), with an analogue-based stochastic weather generator. This stochastic weather generator can be run in so called ~~*dynamical*~~*dynamic* mode. For each ~~day t~~initial day $t^{(1)}$, we have N best SLP analogues. We randomly select one (k) of those N analogues and time $\tilde{t}t^{(1)}_k$, with a probability weight that is

Figure 2. Schematic of the iteration procedure to simulate one random trajectory of temperature (TG) from SLP analogues. The values of $t^{(k)}$ are the days to be simulated by the system. The values of $t_1^{(k)},\dots,t_N^{(k)}$ are the analogue days for $t^{(k)}$. The red SLP rectangles are the randomly selected analogues, according to the rule defined in the lower box. This procedure is repeated S times to generate an ensemble of trajectories.

1. inversely proportional to the calendar distance of the analogues dates ~~$\tilde{t}$ to t~~ $t_k^{(1)}$ to $t^{(1)}$. This constrains the time of analogues to move forward.

2. inversely proportional to the correlation of the analogue with the SLP pattern at time ~~t~~$t^{(1)}$. This constraint favors analogues with the best patterns, among those with the closest distance.

3. a zero weight if ~~$\tilde{t}$~~ $t_k^{(1)}$ is larger than ~~t~~$t^{(1)}$. This ensures that no information coming from times beyond ~~t~~ $t^{(1)}$ is used in the simulation process.

The simulated SLP at ~~$t+1$~~ the next day $t^{(2)}$ is then the next day of the selected analogue ($t^{(2)}=t_k^{(1)}+1$). We repeat this operation on $t^{(2)},\dots t^{(t)}$ until a lead time T. This generates one random ~~trajectory between t and $t+T$~~ daily trajectory between $t^{(1)}$ and $t^{(1)}+T$. The random sampling procedure is repeated S times to generate an ensemble of trajectories. Here, $S=100$. This procedure is summarized in Fig. 2.

If we want to simulate a daily sequence starting at time t and until $t+T$, we have excluded all analogues whose date falls in $[t,t+T]$ in the random analogue selection. This provides a simple way of performing hindcast forecast for temperature or NAO index.

In this paper, the lead time T is 5, 10, 20, 40 and 80 days ahead. The latter two values are meant to illustrate the limits of the system. For each daily trajectory starting at t, we compute the temporal average between t and $t+T$. Therefore, we go from ~~a~~ an ensemble meteorological forecast (5 days) to a seasonal forecast (80 days) of averaged trajectories.

The $S=100$ simulations at each time steps allow computing medians and quantiles of the averaged trajectories.

For comparison purposes, climatological and persistence forecasts are also computed. The climatological forecast for a lead time T is determined from the seasonal cycle of T averages of the variable we simulate. For each time t, the *climatological* forecast for $t+T$ is the mean seasonal cycle of T averages at the calendar day of t. The *persistence* forecast at time t for a lead time of T is the observed average between $t-T$ and t. Those two types of forecasts are illustrated in Fig. 3. Ensembles of reference forecasts are performed by adding a Gaussian random noise (independent and identically distributed), whose variance is the variance of the observed T averages. These two definitions ensure a coherence between the predictand (averages over T values ahead) and predictors for references (mean of averages over T values for climatology, or average over T preceding values for persistence).

3.3 Alternative autoregressive weather generator

Figure 3. Illustration of average forecast for daily mean temperature (TG) in Toulouse, for Jan. 1st 2007. The continuous black line indicates the observations of TG for the first 90 days of 2007. The colors indicate lead times T. The continuous arrows are for ~~observed~~ averages of observed TG ~~on~~ from Jan. 1st ~~2007.~~ 2007 to the lead time T. The dashed lines are for the persistence forecast of TG and the dotted lines are for the climatology forecast of TG on Jan. 1st 2007.

This non parametric weather generator (based on data resampling) is compared to a parametric autoregressive simple model, based on a similar principle of a relation between SLP and variables like temperature and NAO index. We build a multi-variate autoregressive model of order 1 R_t for SLP by expressing:

$$R_{t+1} = A \cdot R_t + B_t, \tag{1}$$

where A is a "memory" matrix and B_t is multivariate random Gaussian noise with covariance matrix Σ. We impose that the multivariate process R_t yields the same covariance matrix $C(0)$ and same lag-1 covariance matrix $C(1)$ as SLP. The matrices A and Σ can be estimated by:

$$A = C(1)^t \cdot C(0)^{-1} \tag{2}$$

and

$$\Sigma = C(0) - C(1)^t \cdot A. \tag{3}$$

The superscript ^{t} is matrix transposition. In order to avoid numerical problems in the estimation of $C(0)^{-1}$, the model in Eq. (1) is formulated on the first 10 principal components (von Storch and Zwiers, 2001) of North Atlantic SLP (80W–30E; 30–70N: blue rectangle in Figure 1), which account for ≈ 80 % of the variance. In this way, $C(0)$ is a diagonal matrix whose elements are the variances of the 10 principal components of SLP. Such a parametric model has been used as a null hypothesis for weather regime decomposition by Michelangeli et al. (1995).

We then perform a multilinear linear regression between the five mean daily temperature records (TG at Berlin, Toulouse, Orly, Madrid, De Bilt) and the NAO index:

$$X_t = a\mathrm{SLP}_t + b + \epsilon_t \tag{4}$$

where $X = (\mathrm{TG}_{\mathrm{Berlin}}, \ldots, \mathrm{TG}_{\mathrm{DeBilt}}, \mathrm{NAO})$, a is a 6×10 matrix, b is a 10-dimensional vector, and ϵ_t is a 10 dimensional residual term. We simulate Eq. (1) with the same observed initial conditions as for the analogue forecast. Then Eq. (4) is applied to simulate an ensemble of forecasts of temperatures and NAO index. In this multivariate autoregressive model (mAR1), the temporal atmospheric dynamics is contained in the matrix A. The major caveat of the parametric model in Eqs. (1 and 4) is that it does not contain a seasonal cycle. Introducing a seasonal dependence on the matrices A and a would require many tests that are beyond the scope of this paper.

3.4 Forecast skill

The simplest score we use is the temporal correlation between the median of the ensemble forecast and the observations. Due to the autocorrelation and seasonality of the variables we try to simulate (temperature and NAO index), we consider the correlations for the forecast in the months of January and July.

The continuous rank probability score (CRPS) compares the cumulated density functions of a forecast ensemble and observations y_t, for all times t (Ferro, 2014).

$$\mathrm{CRPS}(t) = \int_{-\infty}^{\infty} \left(F_t(x) - \mathbf{1}(x \geq y_t)\right)^2 dx. \tag{5}$$

F_t is the cumulated density function of the ensemble forecast at time t. It is obtained empirically from the ensemble of simulations of the model. $\mathbf{1}(x \geq y_t)$ is the empirical cumulated density function of the observation y_t.

The score is the average over all times:

$$\mathrm{CRPS} = \frac{1}{N}\sum_{t=1}^{N} \mathrm{CRPS}(t). \tag{6}$$

The CRPS is a fair score (Ferro, 2014; Zamo and Naveau, 2018) in that it compares the probability distributions of forecasts and observations and it is optimal when they are the same. Discrete estimates of CRPS can yield a bias for small ensemble sizes ~~M~~S. We simulate ~~$M = 100$~~ $S = 100$ trajectories for each forecast. This is more than most ensemble weather forecasts (typically, ~~$M = 51$~~ $S = 51$ for the European Center for Medium Range Forecast (ECMWF) ensemble forecast) and guaranties that the bias due to the number of samples is negligible.

A perfect forecast gives a CRPS value of 0.

The CRPS can be decomposed into a reliability, resolution and uncertainty terms (Hersbach, 2000, Eq. (35)):

$$\mathrm{CRPS} = \mathrm{Reli} - \mathrm{Resol} + \mathrm{U}. \tag{7}$$

The reliability Reli term measures whether events that are forecast with a certain probability p did occur with the same fraction p from the observations (Hersbach, 2000). The remaining terms of the right hand side of Eq. (7) are called the potential CRPS, i.e. it is the CRPS value one would obtain if the forecast were perfectly reliable ($\mathrm{Reli} = 0$). Hersbach (2000) argues that the potential CRPS is sensitive to the average spread of the ensemble.

The units of CRPS are those of the variable to be ~~forecasted~~ forecast, therefore its interpretation is not universal, and comparing the CRPS values for NAO index and temperatures is not directly possible. Therefore, it is useful to compare the CRPS of the forecast with the one of a reference forecast. A normalization of CRPS provides a skill score with respect to that reference:

$$\mathrm{CRPSS}_{\mathrm{ref}} = 1 - \frac{\mathrm{CRPS}}{\mathrm{CRPS}_{\mathrm{ref}}}. \tag{8}$$

The CRPSS indicates an improvement over the reference forecast. A perfect forecast has a CRPSS of 1. A positive improvement over the reference yields positive a CRPSS value. A value of 0 or less indicates that the forecast is worse than the reference.

We compute CRPSS for the climatological and persistence references. We used the ~~package~~ packages "SpecsVerification" and "verification" in R to compute CRPS decomposition and CRPSS scores. Hence we compare our stochastic forecasts with forecasts made from climatology and persistence. By construction, the persistence forecast shows an offset with the actual value ahead, because the persistence is the value of the average of observations between $t-T$ and t. The variability of the climatological forecast is low because it is an average of T long sequences.

3.5 Protocole

We tested the ensemble forecast system on the period between 1970 and 2010. We simulate $N = 100$ trajectories of ~~length~~ lengths $T \in \{5, 10, 20, 40, 80\}$ days for a given date t, and average each trajectory over T. The dates t are shifted every $\delta t \in \{2, 5, 10, 10, 20\}$ days, respectively for each different value of lead times T.

We recall that the tests we perform are on the *average* of the forecast between t and $t+T$, not on the *value at* time $t+T$.

The CRPS and CRPSS is computed for each value of lead times T, with references of climatology and persistence. We determine the CRPS reliability and plot quantile-quantile plots for observed and forecast values of the averages. This allows assessing biases in simulating averages. Variables such as temperature yield a strong seasonality, which is larger than daily variations. It is hence natural to have very high correlations or skill scores if one considers those scores over the whole year. Therefore we compare the skill scores for January and July, in order to avoid obtaining artificially high scores.

4 Results

We performed our stochastic forecasts on the NAO index and European temperatures with the analogue stochastic weather generator and the mAR1 model. The two datasets (NAO and temperature) are treated separately because the simulations are done with two different analogue computations (Sec. 3.1).

4.1 ~~NAO index~~

4.1 NAO index

For illustration purposes, we ~~crepommenrt~~ comment on the skill on simulations of 2007. Fig. 4 shows the simulated and observed values of averages of the NAO index, for five values of T (5, 10, 20, 40 and 80 days). This example suggests a good skill to forecast the NAO index from SLP analogues, especially at lead times of $T = 5$ to 10 days.

The q-q plots of the median of simulations versus observations show a bias that reduces the range of variations (Fig. 4, right column). There are two reasons for this reduction of variance, which is proportional to the lead time T:

1. individual simulated trajectories tend to "collapse" toward a climatological value after ≈ 10 days,

Figure 4. Left column: time series of analogue ensemble forecasts for 2007, for lead times $T \in \{5, 10, 20, 40, 80\}$ days. Red lines represent the median of 100 simulations; pink lines represent the 5th and 95th quantiles of the 100 member ensemble. Right column: q-q plots of NAO forecasts vs. observed values for all years in 1970–2010 for all lead times. The dotted line is the first diagonal.

Figure 5. Skill scores for NAO index for lead times T of 5, 10, 20, 40 and 80 days for ~~all days (black),~~ January (left: blue) and July (right: red). ~~Square~~ Squares indicate $\mathrm{CRPSS}_{\mathrm{pers}}$, triangles $\mathrm{CRPSS}_{\mathrm{clim}}$ and ~~circles~~ boxplots are ~~fore~~ for correlation. The diamonds indicate the reliability of CRPS (on the same scale as CRPSS). Triangles are identical for all days, January and July. The ~~error bars~~ boxplots for the correlation indicate ~~a 95% confidence interval (von Storch and Zwiers, 2001)~~the spread across the 100 member ensemble forecasts.

2. taking the median of all simulations also naturally reduces the variance.

The q-q plots are almost linear. This means that the bias could in principle be corrected by a linear regression. We will not perform such a correction in the sequel.

The correlation, CRPS reliability and CRPSS values for NAO index forecast are shown in Fig. 5. The values of $\mathrm{CRPSS}_{\mathrm{pers}}$ (for a persistence reference) are rather stable (with a slight increase) near 0.45, and the climatology score slightly decreases with T although positive.

The CRPS reliability values range from $5 \cdot 10^{-3}$ (10 days) to 0.01 (40 days). If they are normalized to the CRPS value (or the variance of the NAO index), this is in the same range as the results of Hersbach (2000) for the ECMWF forecast system up to 10 days.

One the one hand, ~~tThe~~ the $\mathrm{CRPSS}_{\mathrm{clim}}$ values do not depend on the season (identical triangles in Fig. 5). On the other hand, $\mathrm{CRPSS}_{\mathrm{pers}}$ values are higher in July than January for lead times $T \leq 10$ days, and lower for lead times $T \geq 40$ days (squares in Fig. 5). This means that the climatology forecast tends to be better than the persistence forecast for $T > 5$ days (squares higher than triangles in Fig. 5), which can be anticipated because of the inherent lag of the persistence forecast.

The correlation scores decrease with lead time T. The correlation skill is higher in January than in July. It is no longer significantly positive for T larger than 40 days (~~confidence~~ 25 to 75th quantile intervals contain the 0 value). The correlation score values range between 0.65 and 0.82 for $T = 5$ day forecasts, and 0.45 and 0.77 for $T = 10$ day forecasts, depending on the season. This is consistent with the NAO forecast of the Climate Prediction Center ($r = 0.69$ for a 10 day forecast). The correlation score is still significantly positive for $T = 20$ days. The higher correlation scores over the whole year (~~black circles~~not shown) reflect a (small) seasonal cycle of the NAO index. This artificially enhances the score for those lead times because SLP analogue predictands tend to reproduce the seasonality of the SLP field (by construction of the simulation procedure), and the NAO index and SLP variations are closely linked on monthly time scales (by construction of the NAO index).

For comparison purposes, the multivariate autoregressive model NAO time series are shown in Fig. 6. The skill scores (correlation and CRPSS) for the NAO index give positive values, but not as high as for the analogue forecast. Since this weather generator is designed to yield stationary statistical properties, the score values do not depend on the season. CRPSS values for climatology range from 0.36 ($T = 5$ days) to 0.23 ($T = 80$ days). Those values are lower than for the analogue

Figure 6. Multivariate autoregressive (mAR1) model time series of ensemble forecasts for 2007, for lead times $T \in \{5, 10, 20, 40, 80\}$ days. Black lines represent observed averages over lead times T. Red lines represent the median of 100 simulations; pink lines represent the 5th and 95th quantiles of the 100 member ensemble.

system for lead times lower than 20 days (Fig. 5, triangles). The correlation values decrease from 0.58 ($T = 5$ days) to 0.05 ($T = 80$ days), which is lower than for the analogue system (Fig. 5, boxplots).

4.2 European temperatures

The correlation and CRPSS values for ~~temperature~~ daily mean temperature (TG) forecast are shown in Fig. 7. The values of $\text{CRPSS}_{\text{pers}}$ (for a persistence reference) increases with lead time T. This is not surprising because the forecast for the next T days is based on the average of the past T days. Therefore, the persistence forecast is always "late" due to the strong seasonality of temperature variations.

The $\text{CRPSS}_{\text{clim}}$ values decrease with T and plateau near ≈ 0.2. This skill score is still positive (albeit small) for a seasonal forecast. This positive average skill ($\text{CRPSS} > 0.2$) illustrates that the stochastic weather generator follows the seasonality of temperature variations. We note that the CRPSS values for temperature are higher than for the NAO index. This is explained by the seasonality of temperatures, which is more pronounced than in the daily NAO index.

The CRPSS values are rather consistent for the four of the stations (Toulouse, De Bilt, Berlin and Orly). The stochastic model CRPSS fares slightly worse at Madrid station.

The CRPS reliability values are shown in Fig. 7. Their absolute values are larger than for the NAO index, and need to be normalized by the variance of temperature (or the CRPS value itself), as the units of TG are tenths of degrees. The average relative reliability values for lead times lower than 10 days are also similar to what is reported by Hersbach (2000). The reliability values seem to decrease with lead times in winter. They peak at lead times of 40 days in the summer (except for Berlin, where the peak is at 20 days in the summer), then decrease.

The correlation scores for January and July decrease with lead time T. The correlation score values for all days are above 0.97 due to the seasonality of temperatures and forecasts. Since this is not informative, this is not shown in Fig. 7. The correlations are always significantly positive for Toulouse, De Bilt, Berlin and Orly. The summer correlation intervals contain the 0 value at Madrid. This is probably due to the fact that temperature is not linked to the atmospheric circulation in the summer, but rather to local processes of evapo-transpiration (Schaer et al., 1999; Seneviratne et al., 2006). The distribution of the correlation scores (boxplots in Fig. 7) is significantly positive for lead times up to 20 days. It becomes stable near values of 0.2 (or increases) for lead times larger than 40 days. This indicates that there is certainly an artificial predictability beyond those lead times, that show an upper limit of forecasts for this system.

The mAR1 system for temperature is not designed to yield a seasonal cycle (contrary to the analogue system). Therefore, the skill scores of this system for temperatures are negative (for CRPSS) or with non significant correlations.

Figure 7. Skill scores for mean daily temperature in Toulouse, De Bilt, ~~Madrid and Berlin,~~ for lead times T of 5, 10, 20, 40 and 80 days. Square indicate $\mathrm{CRPSS}_{\mathrm{pers}}$, triangles $\mathrm{CRPSS}_{\mathrm{clim}}$ and circles are fore correlation. ~~Black symbols are for all days, blue~~ The diamonds indicate the reliability of CRPS (on the same scale as CRPSS). Blue symbols (left) are for January and red symbols (right) are for July. Triangles are identical for ~~all days,~~ January and July. The ~~correlations for "all days" are very close to 1, due to the seasonal cycle and do not appear on the figure. The error bars~~ boxplots for the correlation indicate ~~a 95% confidence interval (von Storch and Zwiers, 2001)~~the spread across the 100 member ensemble forecasts.

Figure 7. Skill scores for temperature in Madrid and Berlin (continued).

5 Conclusions

We have presented a system to generate ensembles of stochastic simulations of the atmospheric circulation, based on pre-computed analogues of circulation. This system is fairly light in terms of computing resources as it can be run on a (reasonably powerful) personal computer. The most fundamental assumption of the system is that the variable to be predicted is linked to the atmospheric circulation. The geographical window for the computation of analogues needs to be adjusted to the variable to be predicted, so that a prior expertise is necessary for this analogue forecast system. This implies that this approach would not be adequate for variables that are not connected in any way to the atmospheric circulation (here approximated by SLP). The use of other atmospheric fields (e.g. geopotential heights) might increase the skill of the system. The computation of analogues with other parameters (geographical zone, atmospheric ~~predictand,~~ predictor, type of reanalysis, climate model output, etc.) can be easily performed with a web processing service (Hempelmann et al., 2018).

We have tested the performance of the system to simulate an NAO index and temperature variations in five European stations. The performance of such a system cannot beat a meteorological or seasonal forecast with a full-scale atmospheric model (Scaife et al., 2014), but its skill is positive, even at a ~~seasonal~~ monthly time scale, with a rather modest computational cost. From the combination of several skill scores (from CRPS and correlation), we obtain a forecast limit of 40 days, beyond which the interpretation of score values is artificial. We emphasize that the forecast is done on averages over lead times, not on the last value of the lead time.

The reason for the positive skill (especially against climatology) remains to be elucidated, especially for ~~long lead times~~ lead times longer than 20 days. We conjecture that the information contained in the initial condition (as done with regular weather forecasts) actually controls the mean behavior of the trajectories from that initial condition. But such a skill is actually "concentrated" in the first few days, because the trajectories tend to converge to the climatology after ~~15 days.~~ 20 days. The

Figure 7. Skill scores for temperature in Orly (continued).

combination of several skill scores shows that such a system is not appropriate for ensemble forecasts beyond lead times of 40 days, which is lower than what is reported by Baker et al. (2018) for a meteorological forecast of the NAO.

Although the forecast system is random, it contains elements of the dynamics of the atmosphere, from the choice of the analogues. This system is consistently better than a simple multivariate autoregressive (mAR1) model for lead times shorter than 20 days. Since the seasonal cycle is naturally embedded in the analogues simulations, there is no need to parameterize it, contrary to the mAR1 model.

Recent experimental results in chaotic systems have shown that a well tuned neural network algorithm could simulate efficiently the trajectories of a chaotic dynamical system (Pathak et al., 2018b). Our system is an extreme simplification of an artificial intelligence algorithm, but it does demonstrate the forecast skill of such approaches. The advantage here is the physical constraint between the atmospheric circulation and the variables to be simulated.

This system was tested on temperature for five European datasets. This could be extended to precipitation or wind speed. If a real-time forecast is to be performed, we emphasize that only the ~~predictand~~ predictor (here, SLP) needs to be regularly updated for the computation of analogues.

The goal of such a system is not to replace ensemble numerical weather/seasonal forecast. Rather, it can refine the usual references (climatology and persistence) for the evaluation of skill scores. This would create a third "machine learning" reference for CRPSS that might be harder to beat than the classical references.

Code and data availability. The code for the computation of analogues is available at (free CeCILL license):
https://a2c2.lsce.ipsl.fr/index.php/deliverables/101-analogue-software
The temperature data are available at: https://www.ecad.eu
The NAO index data are available at: http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml
The NCEP reanalysis SLP data is available at: https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html

Author contributions. PY wrote the codes, designed the experiments. CD participated to the writing of the manuscript.

Competing interests. The authors declare no competing interest.

Disclaimer. TEXT

Acknowledgements. This work was supported by a grant from the Labex-IPSL and ERC grant No. 338965-A2C2. We thank Mariette Lamige and Zhongya Liu who performed preliminary analyses during their training periods. We thank the two anonymous reviewers for their suggestions to use the CRPS decomposition and a simple parametric stochastic model.

References

Ailliot, P., Allard, D., Monbet, V., and Naveau, P.: Stochastic weather generators: an overview of weather type models, Journal de la Société Française de Statistique, 156, 101–113, 2015.

Atencia, A. and Zawadzki, I.: A comparison of two techniques for generating nowcasting ensembles. Part II: Analogs selection and comparison of techniques, Monthly Weather Review, 143, 2890–2908, 2015.

Baker, L. H., Shaffrey, L. C., and Scaife, A. A.: Improved seasonal prediction of UK regional precipitation using atmospheric circulation, International Journal of Climatology, 38, e437–e453, 2018.

Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, Monthly Weather Review, 115, 1083–1126, <GotoISI>://A1987H976400002, 1987.

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

Ferro, C. A. T.: Fair scores for ensemble forecasts, Quarterly Journal of the Royal Meteorological Society, 140, 1917–1923, 2014.

Freitas, A. C. M., Freitas, J. M., and Vaienti, S.: Extreme Value Laws for sequences of intermittent maps, arXiv preprint arXiv:1605.06287, 2016.

Hempelmann, N., Ehbrecht, C., Alvarez-Castro, C., Brockmann, P., Falk, W., Hoffmann, J., Kindermann, S., Koziol, B., Nangini, C., Radanovics, S., Vautard, R., and Yiou, P.: Web processing service for climate impact and extreme weather event analyses. Flyingpigeon (Version 1.0), Computers & Geosciences, 110, 65–72, https://doi.org/10.1016/j.cageo.2017.10.004, http://www.sciencedirect.com/science/article/pii/S0098300416302801, 2018.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and Forecasting, 15, 559–570, 2000.

Hurrell, J.: Decadal trends in the North-Atlantic Oscillation - regional temperatures and precipitation, Science, 269, 676–679, <GotoISI>://A1995RM70200029, 1995.

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.

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.-Atmospheres, 117, Doi 10.1029/2011jd017 197, 2012.

Jézéquel, A., Yiou, P., and Radanovics, S.: Role of circulation in European heatwaves using flow analogues, Climate Dynamics, 50, 1145–1159, 2018.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bulletin of the American Meteorological Society, 82, 247–267, <GotoISI>://000166742900003, 2001.

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, WETTREG, Atmosphere, 4, 169–197, doi:10.3390/atmos4020 169, 2013.

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

Lucarini, V., Faranda, D., Freitas, A. C. M., Freitas, J. M., Holland, M., Kuna, T., Nicol, M., Todd, M., and Vaienti, S.: Extremes and recurrence in dynamical systems, John Wiley & Sons, 2016.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation Downscaling under Climate Change: Recent Developments to Bridge the Gap between Dynamical Models and the End User, Reviews of Geophysics, 48, https://doi.org/Artn Rg3003 Doi 10.1029/2009rg000314, <GotoISI>://000282328000002, 2010.

Michelangeli, P., Vautard, R., and Legras, B.: Weather regimes: Recurrence and quasi-stationarity, J. Atmos. Sci., 52, 1237–1256, 1995.

Parey, S., Hoang, T. T. H., and Dacunha-Castelle, D.: Validation of a stochastic temperature generator focusing on extremes, and an example of use for climate change, Climate Research, 59, 61–75, 2014.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E.: Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach, Physical Review Letters, 120, 024 102, 2018a.

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., and Ott, E.: Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model, Chaos: An Interdisciplinary Journal of Nonlinear Science, 28, 041 101, 2018b.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., and Gordon, M.: Skillful longrange prediction of European and North American winters, Geophysical Research Letters, 41, 2514–2519, 2014.

Schaer, C., Luthi, D., Beyerle, U., and Heise, E.: The soil-precipitation feedback: A process study with a regional climate model, J. Clim., 12, 722–741, <GotoISI>://000079181700004, 1999.

Semenov, M. A.: Using weather generators in crop modelling, Proceedings of the VIIth International Symposium on Modelling in Fruit Research and Orchard Management, pp. 93–100, <GotoISI>://000239370200011, 2006.

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.

Seneviratne, S. I., Luthi, D., Litschi, M., and Schaer, C.: Land-atmosphere coupling and climate change in Europe, Nature, 443, 205–209, <GotoISI>://000240467000045, 2006.

Slonosky, V. and Yiou, P.: The North Atlantic Oscillation and its relationship with near surface temperature, Geophys. Res. Lett., 28, 807–810, <GotoISI>://000167229700016, 2001.

Sparks, N. J., Hardwick, S. R., Schmid, M., and Toumi, R.: IMAGE: a multivariate multi-site stochastic weather generator for European weather and climate, Stochastic Environmental Research and Risk Assessment, 32, 771–784, 2018.

Sturaro, G.: A closer look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data, Climate dynamics, 21, 309–316, https://doi.org/10.1007/s00382-003-0334-4, 2003.

van den Dool, H. M.: A new look at weather forecasting through analogs, Monthly Weather Review, 117, 2230–2247, https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2, 1989.

van den Dool, H. M.: Empirical Methods in Short-Term Climate Prediction, Oxford University Press, Oxford, 2007.

von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 2001.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, Journal of the Royal Society Interface, 11, 20131 162, 2014.

Wilks, D.: Statistical Methods in the Atmospheric Sciences: An Introduction, Academic Press, San Diego, 1995.

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

Yiou, P.: AnaWEGE: a weather generator based on analogues of atmospheric circulation, Geoscientific Model Development, 7, 531–543, https://doi.org/10.5194/gmd-7-531-2014, http://www.geosci-model-dev.net/7/531/2014/, 2014.

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. Dyn., 41, 1333–1344, https://doi.org/10.1007/s00382-012-1626-3, 2013.

Zamo, M. and Naveau, P.: Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts, Mathematical Geosciences, 50, 209–234, 2018.