Ref. Ares(2021)1609865 - 03/03/2021

CAFE
Climate Advanced Forecasting of sub-seasonal Extremes

D3.3 Predictability of European precipitation and temperature from analogue SWG

Lead Beneficiary: ARIA Technologies (AT)
Work Package: WP3
Dissemination level: PUBLIC
Type: Report
Due date: 28.02.2021
Submission date: 01.03.2021
Authors: Meriem Krouma (ARIA), Cèline Deandreis (ARIA); Pascal Yiou (CEA)

Summary

Figure and Table

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Abstract

Accurate predictions of precipitation are essential for many life activities. In this study, we aim to assess the skill of a stochastic weather generator (SWG) to forecast precipitation in several cities of Western Europe. The SWG is based on random sampling of analogues of the geopotential height at 500 mb. The SWG is evaluated for two reanalyses (NCEP and ERA5).

We simulate 100-member ensemble forecasts on a daily time increment. We evaluate the performance of SWG with forecast skill scores. Results show significant positive skill score (CRPSS and correlation) for lead times of 5 and 10 days for different areas in Europe.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Introduction

Weather forecast was made to provide useful information to decision-makers. Providing weather forecasting at different time scales is a key component to mitigate the impacts of extreme weather events. It allows to elaborate adequate strategies for mitigation and development in several sectors such as agriculture, water management and urban construction. Short and medium-term forecasts can be useful for decisions, nevertheless, information derived from the sub-seasonal range could help to implement additional measures in advance. The sub-seasonal time range lies between the medium term daily weather forecast and the seasonal. For this purpose, we aim through this study to assess the skill of a stochastic weather generator (SWG) to forecast precipitation in several locations in Europe. The SWG is based on random sampling of analogues of the geopotential height at 500 mb. In fact, combining stochastic generator and analogs has shown results in the field of ensemble weather forecast (Horton, 2019; Yiou and Déandréis, 2019). Indeed, the influence of large-scale circulation on local climate variables has been confirmed in previous studies such as the influence of atmospheric circulation on precipitation in the Eastern Mediterranean Basin and in Greece (Türkes et al., 2002; Xoplaki et al., 2000). Similar influences have been found on precipitation and temperature in the North Atlantic region (Jézéquel et al., 2018b). This has motivated the use of analogs of atmospheric circulation to forecast ensemble weather. In fact, analog forecasting is a non-parametric technique introduced by Lorenz (1969) and applied for weather forecasting (Delle Monache et al., 2013). This method assumes that similar situations in atmospheric circulation may lead to similar local weather conditions (Lorenz, 1969). It consists of identifying similarities of atmospheric circulation conditions from the past to the current observations. Weather analog has been applied to forecast precipitation (Blanchet et al., 2018), temperature (Yiou and Déandréis, 2019), and snowfall (Nicolet et al., 2019).

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Thus, SWGs are statistical tools to simulate ensembles of trajectories. They were developed and used to estimate the probability distributions of climate variables such as temperature, solar radiation, and precipitation through extensive simulations (Richardson, 1981). They represent a good solution for weather prediction as they can be calibrated for small spatial scales comparing to numerical models (Ailliot et al., 2015). This explains their wide applications in impact studies.

In this report, we explain further the methodology applied and the results found. This deliverable is part of a paper that was submitted recently (Krouma et al., submitted to Geosci. Model Dev.).

Methodology

The methodology is based on computing analogs of the atmospheric circulation and simulating precipitation using the stochastic weather generator. We started first by collecting and assessing data.

1. **Data.**

The daily precipitation data was obtained from the European Climate Assessment and Data (ECAD) project (Klein Tank et al., 2002) for different locations in western Europe (Berlin, Madrid, Orly, Toulouse, De Bilt, Brest), which are subject to various meteorological influences (Figure 1). The data were available at a daily time step from 1948 to 2019. The choice of those stations was based on the availability of a large and common period of observations with a low rate of missing data (less than 10%).

We recovered the geopotential height at 500 hPa fields from the reanalysis of the National Centers for Environmental Prediction (NCEP: (Kistler et al., 2001)) with a spatial resolution of 2.5° x 2.5° from 1 January 1948 to 31 December 2019.

We also used the atmospheric reanalysis (version 5) of the European Centre for Medium-Range Weather Forecasts (ECMWF) (ERA5; Hersbach et al. (2020)). ERA5 data are available from 1979 to present with a horizontal resolution of 0.25° x 0.25°. We used different sources of reanalyses for verification purposes. Moreover, we computed analogs

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

over different geographic domains as showed in Figure 1 for the computation of analogs, and we choose an optimized region to compute circulation analogs for the different studied areas, which covers 30°W – 20°E and 40°– 60°N.

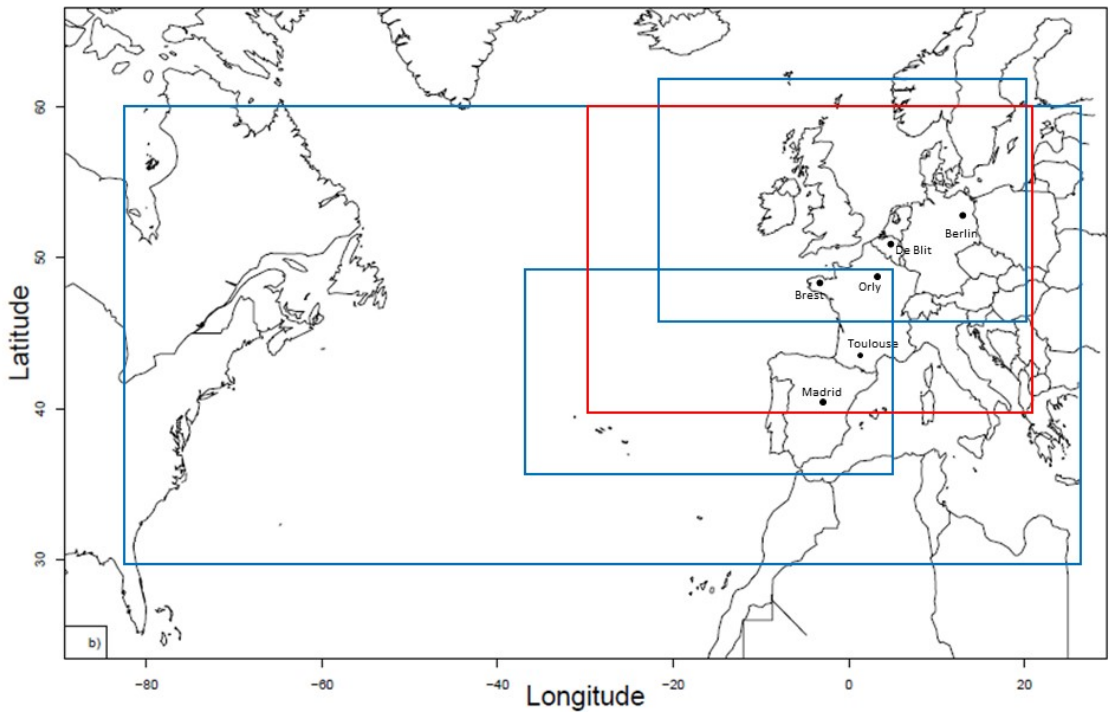

Figure 1. Domains of computation of analogs, we computed analogs over different domains, each one includes a part of the Atlantic and focus on a part of Western Europe, in order to test the sensitivity of our model to different geographic areas.

2. Computing analogs of the atmospheric circulation

In this work, we start first by building a database of atmospheric circulation analogs. The procedure of Yiou and Déandréis (2019) is used as follow: For a given day t, the similarity of the average of Z500 is determined for all days t0 that are in a different year than t, but within 30 calendar days of t. This similarity is quantified by a Euclidean distance (or square root) between the daily maps of Z500. For each day t, we have chosen K = 20 best analogues with the smallest distance. We calculate the spatial rank correlation between the best analogs of the Z500 and the Z500 at time t for post-verification purposes.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

We consider a time embedding of 4 days to search of analogs dates. This means that the field X(t) for which we compute analogs is X(t) = (Z500(t) ; Z500(t+1) ; … ; Z500(t+3)). This ensures that temporal derivatives of the atmospheric field are preserved (Yiou et al., 2013). Hence the distance that is optimized to find analogs of the Z500(x; t) field is:

$$D(t,t') = [\sum_x (\sum_{i=0}^{3} |Z500(x,t+i) - Z500(x,t'+i)|)^2]^{1/2} \quad (1)$$

where x is a spatial index.

3. Configuration of stochastic weather generator

As detailed by (Yiou and Déandréis, 2019), the stochastic weather generator (SWG) we use is based on a random reshuffling of circulation analogs. For a given day t_0 we perform an ensemble of simulations until a lead time $t_0 + T$, with T in {5, 10, 20} days. In order to go from t in $[t_0, t_0 + T]$ to $t + 1$, we sample one analog (out of K=20) at day t, with weights on the calendar day of the analogs and excluding samples that are in $[t_0, t_0 + T]$, so that this procedure reflects a hindcast forecast from t_0. This procedure is iterated from $t = t_0$ to $t = t_0 + T$ to generate one trajectory. It is then repeated $N=100$ times to generate an ensemble of daily trajectories starting at t_0 . Each daily trajectory is time averaged between t_0 and $t_0 + T$.

For each day t_0, we also compute persistence and climatological forecasts for the average between t_0 and $t_0 + T$. The persistence forecast consists in using the average value between $t_0 - T$ and t_0. The climatological forecast takes the climatological average between t_0 and $t_0 + T$. Both "control" forecasts are randomized by adding a small Gaussian noise, whose standard deviation is estimated by bootstrap over T long intervals. Hence we generate ensembles of persistence and climatological forecasts that are consistent with observations (Yiou and Déandréis, 2019).

4. Forecast Verification

Forecast verification is the process of determining the statistical quality of the forecasts. A wide variety of ensemble forecast verification procedures exists. They involve measures of the relationship between a set of forecasts and corresponding observations.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

To assess the quality of precipitation forecasts, we compute indicators such as the Correlation and Continuous Rank Probability Skill Score (CRPSS) for each lead time T, for different seasons and months. The temporal rank correlation is calculated between the precipitation observations and the median of 100 simulations. This simple diagnostic is often used to assess forecast skills of indices (Scaife et al., 2014). The continuous ranked probability score (CRPS) represents the most used score for probability forecast verifications (Ferro, 2007). It is sensitive to the distance between forecast and observation probability distributions. If the ensemble forecast yields a probability distribution P(x), the CRPS measures how is the probability distribution of x (Hersbach, 2000).

The CRPS is computed as:

$$CRPS\ (P, x_a) = \int_{-\infty}^{+\infty} (P(x) - P_a(x))^2 dx \qquad (2)$$

where Pa represents the Heaviside function of the occurrence of x.

The decomposition and properties of the CRPS have been investigated by (Ferro, 2007; Hersbach, 2000; Zamo and Naveau, 2018). A perfect forecast would have a CRPS equal to 0, but the CRPS value obviously depends on the units of the variable to forecast. It is hence difficult to compare CRPS values for temperature and precipitation, within the same ensemble forecast. This issue is also acute for non-Gaussian variables with heavy tails (Zamo and Naveau, 2018), so that the interpretation of a given CRPS value might not always be informative.

One way of circumventing this difficulty is to compare CRPS values to reference forecasts, such as persistence or seasonality. The continuous rank probability skill score (CRPSS) is a normalization of Eq. (2) with respect to such a reference.

The CRPSS is hence computed by:

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}} \qquad (3)$$

where the $CRPS_{ref}$ is the CRPS of reference forecast (climatology or persistence). The CRPSS is interpreted as a percentage of improvement over a reference forecast. The

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

values of the CRPSS varies between $-\infty$ and 1. The forecast is considered to be an improvement over the reference when the CRPSS value is close to 1 (i.e. when the CRPS is 0). Values of CRPSS equal to 0 indicates no improvement over the reference. Values inferior to 0 mean that the forecast is worse than the reference.

We use the CRPSS values to determine the maximum lead time T for which the SWG forecast is better than references. Then the SWG assessments will use the CRPS and directly compare the probability distributions of precipitation ensemble forecasts.

Results

Figure 2 shows the observed and simulated values of precipitation for lead times of 5 and 10 days for summer (June–July–August: JJA) and winter (December–January–February: DJF), for Orly precipitation data. We observe significantly positive correlations between observed values and the median of the forecasts, for the different data sets as represented in table 1. The correlation is generally smaller in the summer than in the winter and - and the forecast misses some extreme values of precipitation.

Table 1. Correlation between observations and the median of 100 simulations for winter and summer for a lead time of 5 days

Location	Correlation DJF	p-values	Correlation JJA	p-values
Berlin	0.42	$< 2.2^{-16}$	0.22	$< 2.2^{-16}$
Orly	0.58	$< 2.2^{-16}$	0.23	$< 2.2^{-16}$
De Bilt	0.45	$< 2.2^{-16}$	0.20	$< 2.2^{-16}$
Toulouse	0.43	$< 2.2^{-16}$	0.18	$< 2.2^{-16}$
Madrid	0.58	$< 2.2^{-16}$	0.29	$< 2.2^{-16}$

For a lead time of 10 days, SWG simulation still show capacity to predict precipitation especially for winter with a correlation equal to 0.23 (Orly), 0.30 (Berlin), 0.43 (Madrid). We observe that the 5th and 95th quantiles of simulations include the different values of observations. This heuristically confirms the good skill of SWG to forecast precipitation

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

from Z500 for several seasons (winter and summer) in several locations for T = 5 and T = 10 day lead times.

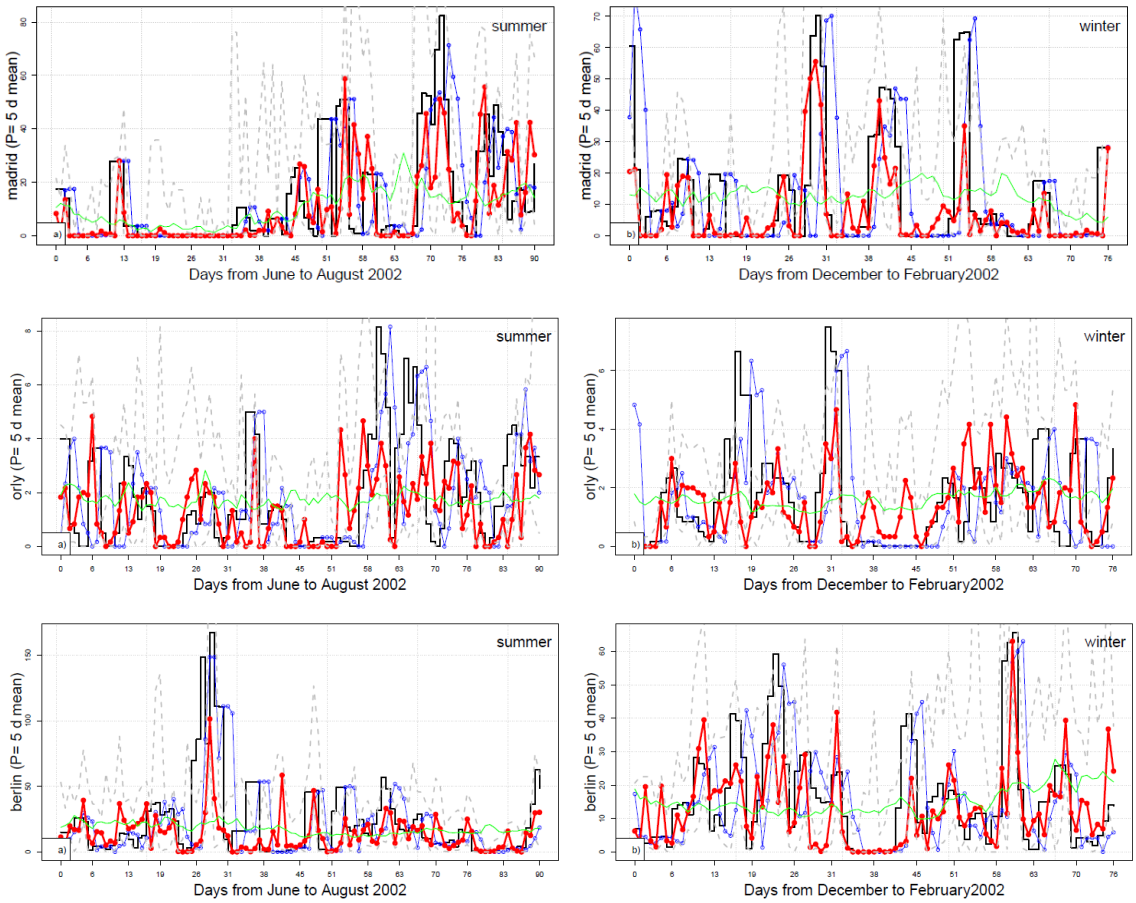

Figure 2. Time series of analogue ensemble forecasts for 2002, for lead times of 5 days a) for summer (June to August) and b) winter (December to February) for Orly, Berlin and Madrid. The median of 100 simulations is represented by red line. Black lines represent observations values. Dash lines represent the 5th and 95th quantiles. Blue lines represent persistence forecasts and green lines represent the climatology forecasts.

The difference of the forecast correlation skills between the six studied locations may be related to the variation of the local climate from one region to another. The studied areas are in different climate types according to Köppen-Geiger's climate classification map (Peel et al., 2007). From the south western side of Europe, Madrid is in the arid zone (Peel

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

et al., 2007), which indicates that convective rains are less significant, so that the origin of precipitation might be the result of humidity coming from the Atlantic. Conversely, Berlin is located in a cold zone characterized by warm summer and the absence of a dry season (Peel et al., 2007), so that the precipitation could be the result of both convective rains and Atlantic humidity.

Moreover, we computed the CRPSS for precipitation for lead times from 5 to 20 days (Figure 3) for the different studied areas. We represent skill scores for January and July in order to evaluate the skill of the SWG to predict precipitation in both seasons (winter and summer). For comparison purposes, SWG simulations are obtained using analogs computed from reanalyses on the NCEP and ERA5 reanalyses. Comparing their skill scores, we found that CRPSS and correlation between observations and simulations are positive in both cases, and showing positive improvement comparing to persistence and climatology forecasts. The CRPSS and correlation for simulations with analogs of NCEP are slightly higher than with ERA5, due to the longer length of the NCEP reanalysis, which has a better potential to find good analogs.

We determined that the SWG simulations showed better skills for the geographic domain outlined in red, in Figure 1 as it allows to make forecasts for all the studied areas and we find that the skill scores over this geographic domain remained the highest ones. Therefore, we focus on SWG simulations with analogs from the NCEP reanalysis in the sequel. The CRPSS for persistence and climatology references show positive values for lead times of up to 20 days (Figure 3).

The values of CRPSS with persistence reference (represented by squares) decrease with lead times, showing high values over 5 days. The CRPSS for climatology (triangles) show lower values, although positive. The correlation skill is positive for both seasons but higher in winter (January) than in summer (July). For a lead time of 5 days, the correlation is equal to 0.59 for Madrid, and to 0.50 for Berlin. For a lead time of 10 days, it is equal to 0.42 for Madrid, and to 0.30 for Berlin.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

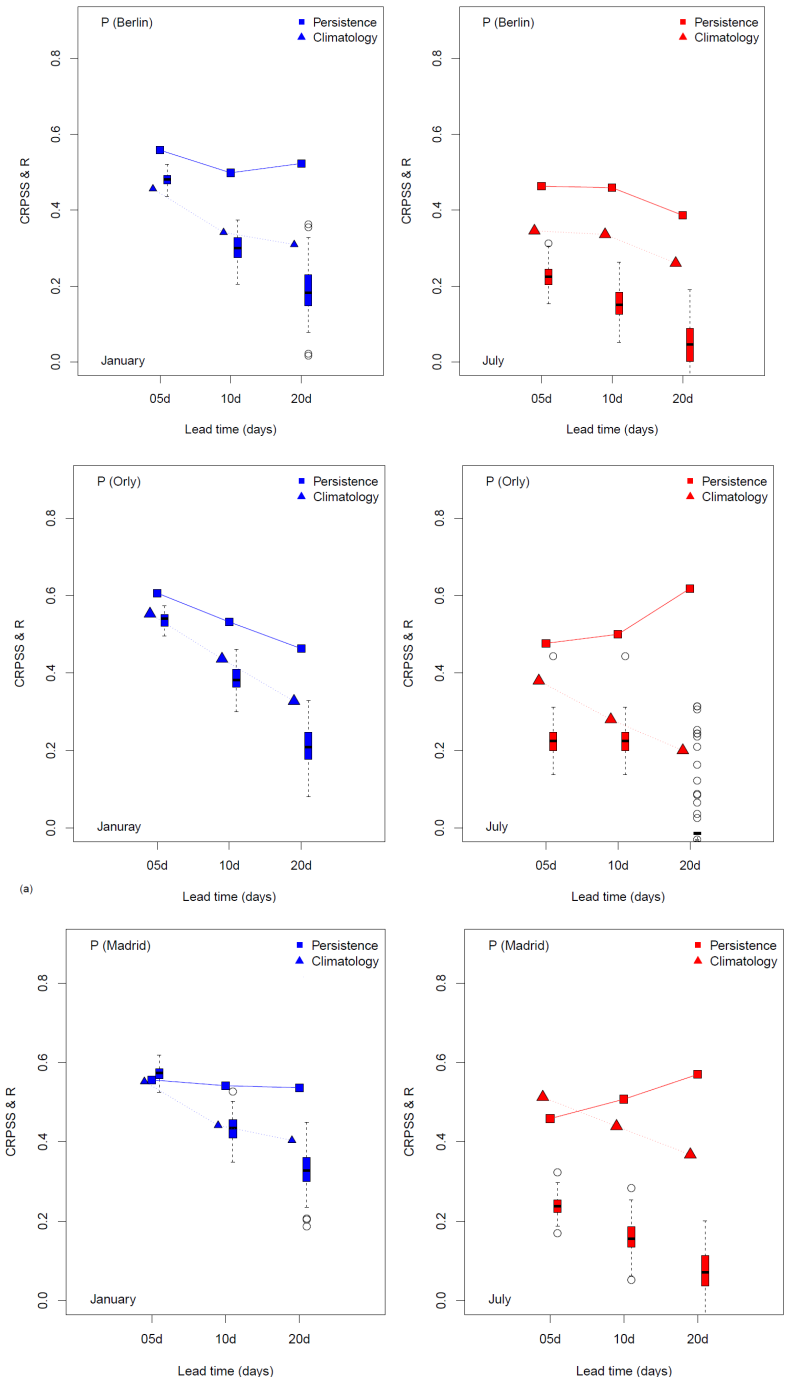

Figure 3. Skill scores for the precipitation of Berlin, Orly and Madrid for lead times of 5, 10, 20 days for January (blue) and July (red) for analogs computed from reanalyses of

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

NCEP. Squares indicate CRPSS where the Persistence is the baseline, triangles indicate CRPSS where the climatology is the reference, and box-plots indicates the correlation between observation and median of 100 simulations.

The SWG was tested in previous work of (Yiou and Déandréis, 2019) to forecast North Atlantic oscillation (NAO) and temperature in western Europe. Comparing the performance of the SWG to forecast those different meteorologic variables, we notice that the model shows good performance to forecast the temperature and NAO in the winter, also the best performance of the model is at a lead time of 5 days. We find that the skill scores (CRPSS and correlation) decrease with lead of times. The forecast skill of the SWG shows variability from one location to another. However, the model was able to forecast temperature until 40 days in Berlin, Orly, Toulouse and De Bilt with positive skill scores. From a visual inspection of the CRPSS and correlations, we chose to focus on lead times of T = 5 days, for which the correlation exceeds 0.5 in the winter. It is rather low in the summer, due to convective events leading to a high precipitation variability (from no rain to very high values). Correlation scores become barely significant for lead times of 20 days, so that, like temperature, the SWG should not be used beyond that horizon.

Conclusion

In this work, we showed the performance of a stochastic weather generator (SWG) to simulate precipitation over different locations in western Europe and for various times scales from 5 to 20 days. The input of our model was analogs of geopotential heights at 500 hPa (Z500). The choice of such input was made in order to evaluate the impact of large scale circulation on local weather variables. SWG showed a good skill to predict precipitation for a lead time of 5 and 10 days from analogues of Z500. This study complements the work of (Yiou and Déandréis, 2019), for precipitation. We explored the sensitivity of the SWG model on analogs computed from different geographical areas and from different reanalyses (ERA5 and NCEP). We found that the analogs from the NCEP

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

reanalyses provide better performances for simulations, due to its larger length (70 years in NCEP vs. 40 years in ERA5). Therefore the length of the analog database does make a difference, as already suggested by (Jézéquel et al., 2018a).

Bibliography

Ailliot, P., Allard, D., Monbet, V., Naveau, P., n.d. Stochastic weather generators: an overview of weather type models 15.

Blanchet, J., Stalla, S., Creutin, J.-D., 2018. Analogy of multiday sequences of atmospheric circulation favoring large rainfall accumulation over the French Alps. Atmospheric Sci. Lett. 19, e809. https://doi.org/10.1002/asl.809

Delle Monache, L., Eckel, F.A., Rife, D.L., Nagarajan, B., Searight, K., 2013. Probabilistic Weather Prediction with an Analog Ensemble. Mon. Weather Rev. 141, 3498–3516. https://doi.org/10.1175/MWR-D-12-00281.1

Ferro, C.A.T., 2007. A probability model for verifying deterministic forecasts of extreme events. Weather Forecast. 22, 1089–1100.

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. WEATHER Forecast. 15, 12.

Horton, P., 2019. AtmoSwing: Analog Technique Model for Statistical Weather forecastING and downscalING (v2.1.0). Geosci. Model Dev. 12, 2915–2940. https://doi.org/10.5194/gmd-12-2915-2019

Jézéquel, A., Cattiaux, J., Naveau, P., Radanovics, S., Ribes, A., Vautard, R., Vrac, M., Yiou, P., 2018a. Trends of atmospheric circulation during singular hot days in Europe. Environ. Res. Lett. 13, 054007.

Jézéquel, A., Dépoues, V., Guillemot, H., Trolliet, M., Vanderlinden, J.-P., Yiou, P., 2018b. Behind the veil of extreme event attribution. Clim. Change. https://doi-org.insu.bib.cnrs.fr/10.1007/s10584-018-2252-9

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., Fiorino, M., 2001. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bull. Am. Meteorol. Soc. 82, 247–267.

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V., Petrovic, P., 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment:

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

EUROPEAN TEMPERATURE AND PRECIPITATION SERIES. Int. J. Climatol. 22, 1441–1453. https://doi.org/10.1002/joc.773

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

Nicolet, G., Eckert, N., Morin, S., Blanchet, J., 2019. Inférence et modélisation de la dépendance spatiale des extrêmes neigeux dans les Alpes françaises par processus max-stables. Houille Blanche 150–158.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Ko¨ppen-Geiger climate classification. Hydrol Earth Syst Sci 12.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190. https://doi.org/10.1029/WR017i001p00182

Scaife, A.A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M., Dunstone, N., Knight, J., MacLachlan, C., Manzini, E., Müller, W.A., 2014. Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales. Geophys. Res. Lett. 41, 1752–1758.

Türkes, M., Sümer, U., Kiliç, G., 2002. Persistence and periodicity in the precipitation series of Turkey and associations with 500 hPa geopotential heights. Clim. Res. 21, 59–81. https://doi.org/10.3354/cr021059

Xoplaki, E., Luterbacher, J., Burkard, R., Patrikas, I., Maheras, P., 2000. Connection between the large-scale 500 hPa geopotential height fields and precipitation over Greece during wintertime. Clim. Res. 14, 129–146. https://doi.org/10.3354/cr014129

Yiou, P., Déandréis, C., 2019. Stochastic ensemble climate forecast with an analogue model. Geosci. Model Dev. 12, 723–734. https://doi.org/10.5194/gmd-12-723-2019

Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R., Vrac, M., 2013. 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

Zamo, M., Naveau, P., 2018. Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. Math. Geosci. 50, 209–234. https://doi.org/10.1007/s11004-017-9709-7

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.