1 Introduction

Extreme weather events, such as heatwaves, can be exacerbated by climate change and have severe impacts on human health, agriculture, and infrastructure. The study of heatwaves is a major challenge in the field of climate informatics due to the scarcity of data for these rare events and the computational complexity associated with their analysis. The important questions that remain are how heatwaves can be forecast on intermediate or subseasonal and seasonal scales as well as how the frequency and the associated teleconnection patterns of these events would be impacted by climate change.

The first goal of this study is to compare the performance of two data-driven probabilistic forecasting approaches for extreme heatwaves in two European regions: France and Scandinavia, both trained on an output of a General Circulation Model (GCM). We aim to explore the use of Stochastic Weather Generator (SWG), that rely on the method of analog Markov chain, optimize it for our task and compare or combine it with newer deep learning approaches (while also discussing dimensionality reduction techniques). Our second goal will consist of benchmarking the SWG trained on 90 years and comparing statistics of its synthetic trajectories to a 7200 years long control run.

2 Forecasting extreme heatwaves

We use two methods for heatwave forecasting: Convolutional Neural Network (CNN) **?** and analog forecasting technique inspired by SWG Yiou (2014). To validate the methods, we split the data into training and validation sets using the 5-fold cross-validation method. This allows us to estimate uncertainty of the forecasting skill. Both methods are trained on climatic fields from an intermediate complexity climate model (Plasim). The fields include 2-meter temperature $\mathcal{T}$, soil moisture $\mathcal{S}$, and 500 hPa geopotential height $\mathcal{Z}$ anomalies. The soil component of Plasim allows us to study the impact of $\mathcal{S}$ on heatwaves in Europe. The heatwave definition we use depends on the time and space-average temperature anomaly

$$A(t) = \frac{1}{T}\int_t^{t+T} \frac{1}{|\mathbb{D}|}\int_{\mathbb{D}} \left(\mathcal{T} - \mathbb{E}\left(\mathcal{T}\right)\right)(\vec{r}, t')\, \mathrm{d}\vec{r}\, \mathrm{d}t', \tag{1}$$

so that precise identification of a heatwave will depend on a specific choice of the threshold $A(t) = a$ and a period T as well as the region of interest $\mathbb{D}$ and its area $|\mathbb{D}|$.

CNN **?** stacks masked fields and passes them through subsequent convolutional, maxpooling and dense layers and a soft-max function. Gradient descent optimizes the weights of the neural network so that optimal prediction is achieved. Early stopping and dropout reduce potential overfitting.

SWG is based on the method of analog forecasting which involves computation of the proximity between two states with a given metric which accepts the fields $\mathcal{X}$ labeled by index i

$$d(\mathcal{X}_1, \mathcal{X}_2) := \sqrt{\int d\mathbf{r} \sum_i \frac{\alpha_i}{dim(\mathcal{X}^i)} \left(\mathcal{X}_1^i - \mathcal{X}_2^i\right)^2}, \tag{2}$$

The potential synthetic trajectory is constructed by iterating over states in the validation set, finding their best analogs in the training set and then constructing the remainder of the trajectory by successively choosing a random analog among a set of nearest neighbors obtained from the training set and updating the state in time using historical evolution observed in the same training set. This allows Monte Carlo sampling of the conditional probability of the extreme event of interest (due to large number of trajectories we parallelize our python code using numba package). SWG requires computation of a large matrix of nearest neighbors, for which we use kDTree from scikit-learn package and we run it in parallel on 16 CPU cores of Intel Xeon Gold 6142. Because the feature space resulting from $\mathcal{Z}$ field is 1152 dimensional (number of cells) and we run a grid search to find optimal α_0 coefficient in Eq. (1) for each of the 5 folds the procedure takes approximately 18 hours.

We also investigate ways to accelerate the computation of analogs using dimensionality reduction techniques, such as training a Variational Autoencoder (VAE) to project the state of the system to a small-dimensional latent space. For this aim we use 8 layer U-Net type network with residual connections within encoder and decoder respectively, which allows projecting $\mathcal{Z}$ to the 16 dimensional latent space. In this case, the input corresponding to $\mathcal{Z}$ field contribution in Eq. (2) is provided from the latent space. The training takes approximately 50 epochs for each fold on the corresponding training set using TU102 [RTX 2080 Ti Rev. A] GPU which results in total training time of approximately one hour. Such large degree of freedom reduction naturally leads to significant gains (2 orders of magnitude) in subsequent kDTree computation speed. The forecast skill of this VAE SWG, however, remains the same compared to regular SWG.

3 Extreme heatwave sampling

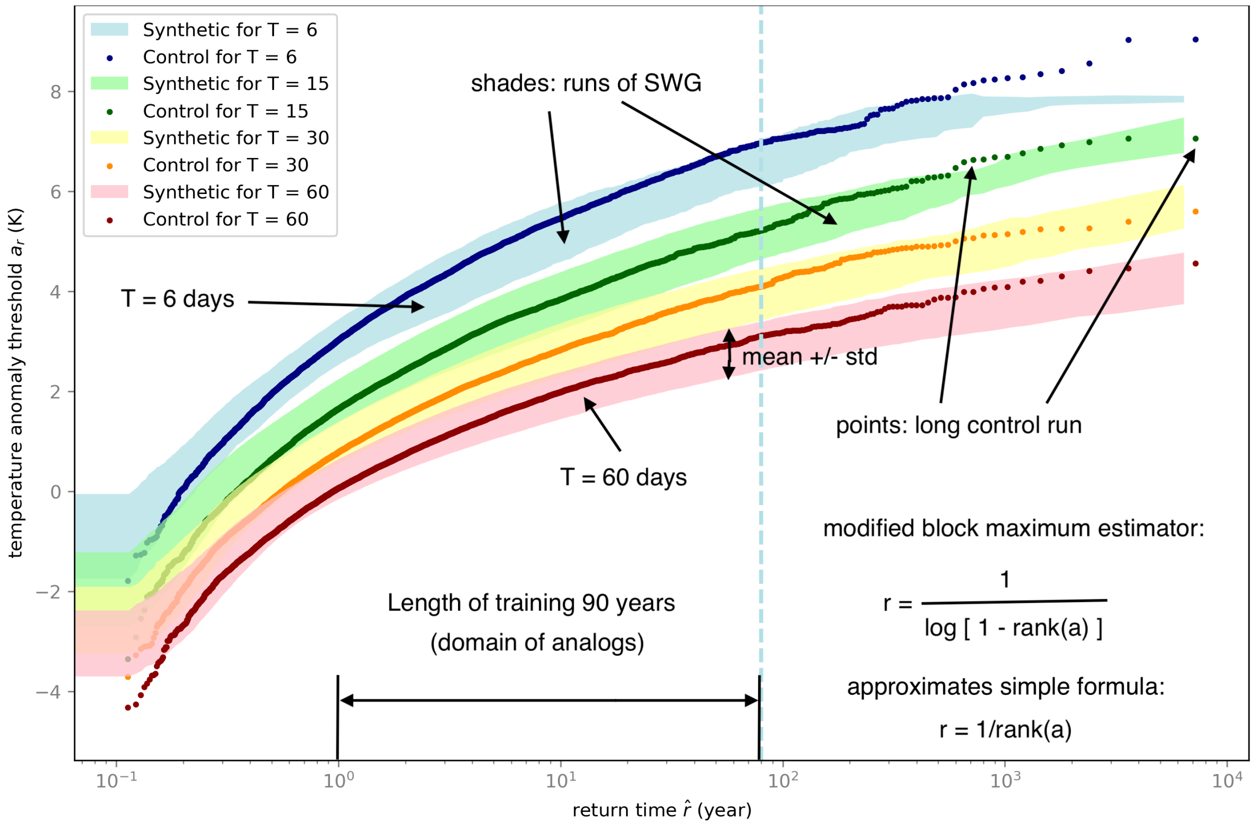
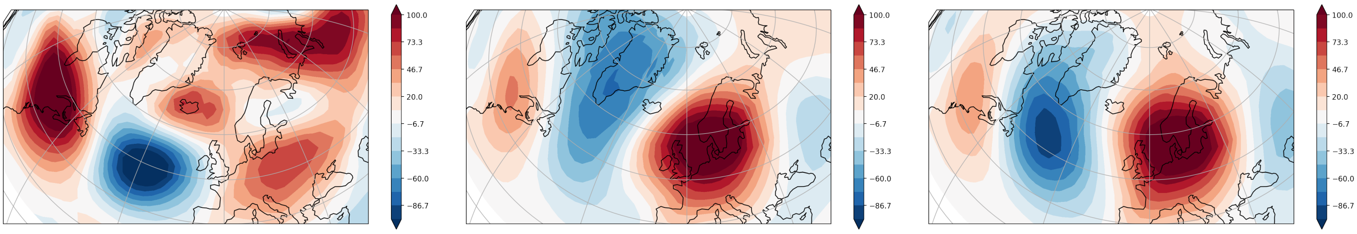


Figure 1: Return time plot generated using modified block maximum estimator: the procedure consists of ranking the yearly summer extremes of $A(t) = a_i$, e.g. if the full dataset has length of 7200 years, a_1 is considered as one in 7200 years event, a_2 as one in 3600 years event, etc. On the y axis we show the magnitude of the extreme a for heatwaves of four different lengths (periods $T = \{6, 15, 30, 60\}$ days). On the x axis we plot the expected return time for such an extreme. The dots correspond to 7200 year-long control run. The shaded areas are obtained by running SWG 10 times and shading mean plus or minus one standard deviation.

In addition to forecasting heatwaves, SWG can be used as a climate model emulator to generate long synthetic series Yiou u.a. (2023) and also estimating rate of returns (waiting time until an extreme value $A(t) = a$ is attained). We validate the use of the SWG model whose analogs are restricted to 90 years training set by comparing the statistical properties of synthetic heatwave sequences it generates to a 7200-year-long control run. The property we will concentrate in this abstract is return time plots (see Lestang u.a. (2018) for the plotting method).

This training dataset, although coming from a GCM and longer than ERA5 reanalysis, is intended to model the lack of data regime one faces when working with (pseudo-)observations. During the validation stage we run SWG to produce 10 sequences, each 7200 years long. We compare resulting return times of extreme heatwaves of various lengths such as $T \in \{6, 15, 30, 60\}$ days on Figure 1 (see equation (1)) providing *quantitative* assessments of the performance of the SWG method. Return times of the synthetic trajectories match the control run mostly within the error bars, except for extremes of short-lasting heatwaves.



(a) 90 year extreme (b) France 7200 years (c) France synthetic SWG

Figure 2: Benchmarking SWG for heatwaves in France. The plots show 500 hPa geopotential height anomalies ($\mathcal{Z}$ in meters) (a) The most extreme heatwave from 90 year long dataset, daily map at the onset of the heatwave with a threshold of 5K. (b) Control run composite which includes heatwaves surpassing 5K threshold. (c) Synthetic run composite which includes heatwaves surpassing 5K threshold. We see that it reproduces better the teleconnection patterns of the control run.

Another interesting question to address is how reliable are the teleconnection patterns computed from the synthetic series generated by SWG. Here we will provide *qualitative* comparisons. First, we show the teleconnection pattern one obtains from the 90 year-long training dataset when conditioning to the onset of the most extreme heatwave in that dataset (see panel (a) in 2). In some sense, this panel is a single member ensemble realization of such a heatwave. We would like to know what are the typical conditions one would expect at the onset of a heatwave of this magnitude. Unfortunately, in reanalysis we may lack the data to answer such questions, and GCMs could be biased or expensive to run to generate sufficient statistics. Could this problem be overcome by SWG? On panel (c) of Figure 2 we plot $\mathcal{Z}$ anomalies obtained from long synthetic trajectories of SWG when conditioned to the heatwave of that magnitude (so composite over many synthetic heatwaves), and compare it with panel (b) — the control run (composite over many modelled heatwaves). There are qualitative similarities between the patterns, the greatest difference being that SWG has a southerly bias for the cyclonic anomaly (in blue) over Greenland, while the direct baseline method (a) yields teleconnection patterns that look substantially different. This approach illustrates how we can improve sampling of extremes using SWG. For further reference please consult our GitHub page: Miloshevich (2023).

References

Yiou, Pascal(2014): *Anawege: a weather generator based on analogues of atmospheric circulation*, 2: 531–543.

Lestang, Thibault / Ragone, Francesco / Bréhier, Charles Edouard / Herbert, Corentin / Bouchet, Freddy(2018): *Computing return times or return periods with rare event algorithms*, 4: 043213.

Yiou, Pascal u.a.(2023): *Worst case climate simulations show that heatwaves could cause major disruptions in the Paris 2024 Olympics.*

Miloshevich, George (2023): *Climate-Learning* https://github.com/georgemilosh/Climate-Learning.