Stochastic weather generator and deep learning approach for predicting and sampling extreme European heatwaves

George Miloshevich, *Dario Lucente, Freddy Bouchet, Pascal Yiou*

[1]Laboratoire des sciences du climat et de l'environnement

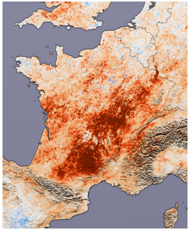
[2]Departement de Physique Ecole Normale Superieure de Lyon

Outline

1. Introduction
   - Problem: prolonged heatwaves
   - Validation: probabilistic score
2. Data-driven probabilistic forecasting of heatwaves
   - Masked Convolutional Neural Network
   - Stochastic Weather Generator: Markov chain of analogs
   - Stochastic Weather Generator on latent space of Variational Autoencoder
3. Extreme event sampling
   - Return time plots
   - Teleconnection patterns
   - Conclusions

Outline

Forecasting heatwaves and validating with Normalized Log Score



European heatwave 2003
Unprecedented event

- HW: extreme of space-time averaged temperature anomalies:

$$A_T(t) = \frac{1}{T}\int_t^{t+T} \frac{1}{|\mathcal{D}|}\int_D \left(T_{2m} - \mathbb{E}\left(T_{2m}\right)\right)(\vec{r},u)\,\mathrm{d}\vec{r}\mathrm{d}u \qquad (1)$$

$T = 14$ days

Area D - "France" / "Scandinavia"

- The goal: find $P(A(t) > a|X(t-\tau),\tau)$ with lead time τ
- NNs trained to optimize MSE do not reconstruct extremes
- Logarithmic score (cross-entropy) suitable for rare events[1]
- Threshold $a \to Y_1 = 1$ above 95 percentile, 3 yrs return time

$$S\left[\hat{p}_Y(X)\right] = -\sum_{k=0\ldots K} Y_k \log\left[\hat{p}_k(x)\right], \quad K = 2 \text{ for binary} \qquad (2)$$

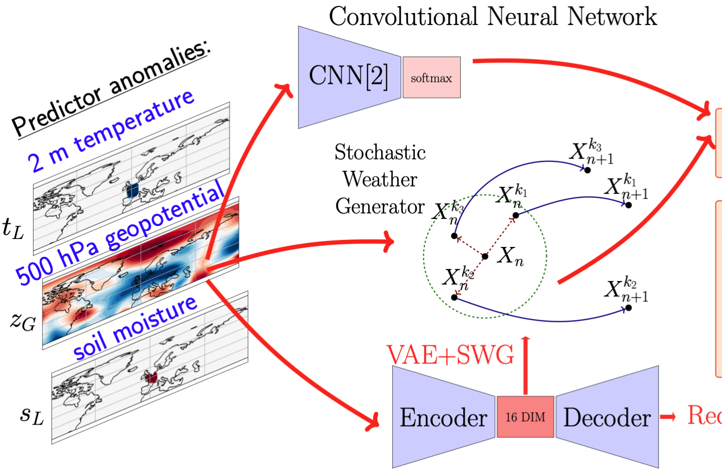
Normalized Log Score (NLS): subtract climatological prediction

$$\mathrm{NLS} = \frac{-\sum_i \overline{p}_i \log \overline{p}_i - \mathbb{E}\left\{S\left[\hat{p}_Y(X)\right]\right\}}{-\sum_i \overline{p}_i \log \overline{p}_i} \qquad (3)$$

[1] R. Benedetti, Monthly Weather Review **138**, 203 –211 (2010)

Outline

Data-driven methods trained on long Plasim climate model run



Reconstruction of z_G, Variational AutoEncoder

committor function: probability of an event

$$\hat{P}(A_T(t) > a | X = x, \tau) \quad (6)$$

Normalized Logarithmic Score (NLS)

- Suitable for rare events:

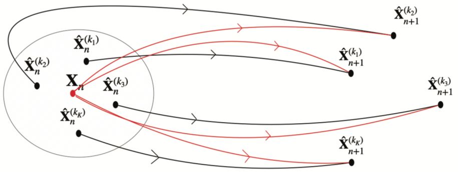
$$\text{NLS} := 1 - \frac{\mathbb{E}\left\{\Sigma_{k=0}^{1} y_k \log\left[\hat{P}_k(x,\tau)\right]\right\}}{\Sigma_i \overline{p}_i \log \overline{p}_i} \quad (7)$$

Figure: Possible field inputs $X = \{t_M, z_G, s_M\}$ in stacked architecture which works better for CNN

[2] https://www.cnrs.fr/fr/changements-climatiques-une-meilleure-prediction-des-canicules-grace-lia
[3] G. Miloshevich et al., Phys. Rev. Fluids (Apr. 2023)

Stochastic Weather Generator: Markov chain of analogs

Analogs are sought using $$X_{n_\star} = \underset{\{X_n\}}{\operatorname{argmin}} \{d\,(x, X_n)\} \qquad (4)$$



Analog method

Stochastic Weather Generator

- Markov chain $\longrightarrow$ surrogate trajectory

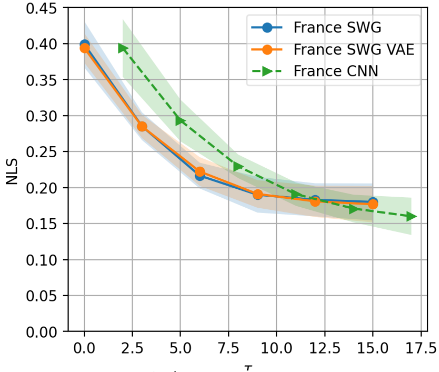
$$P(X(t+\delta t)|X(s)) \propto \begin{cases} 1, & \text{if } X(t) \sim X(s) \\ 0, & \text{otherwise} \end{cases}$$

- Problem (1): how to combine different vars in Euclidean d? Global (G) vs Local (L)
- Problem (2): in big models: curse of high dimensionality (z_G)

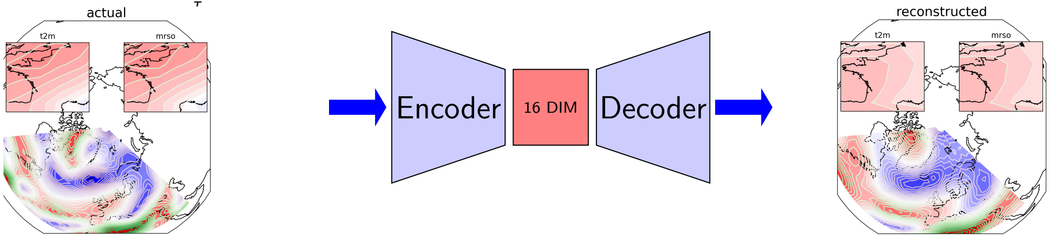
$$d\,(X_1, X_2) = \left[\frac{\alpha}{\sigma_Z^2 \dim(Z_G)} \sum_{i=1}^{\dim(Z)_G} \left(\Delta Z_G^I\right)^2 + \frac{1}{\sigma_T^2}\left(\Delta T_L^I\right)^2 + \frac{1}{\sigma_S^2}\left(\Delta S_L^I\right)^2\right]^{\frac{1}{2}} \qquad (5)$$

[3] P. Yiou et al., https://hal.archives-ouvertes.fr/hal-03921111 (Jan. 2023)

Predicting heatwaves with MC sampling of SWG vs CNN



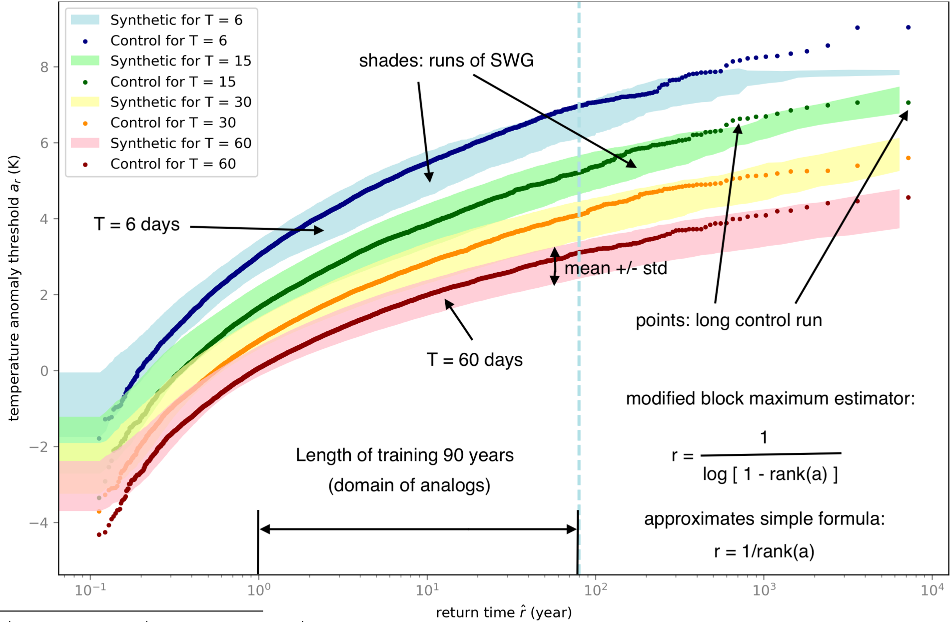
- 400 years to train and 100 to validate
- k-fold cross-validation for uncertainty estimation
- For SWG 10000 trajectories per validation day are launched (numba parallelism)
- Real space SWG (dim z $\approx 1000 + 2$),
- VAE+SWG (dim z $= 16 + 2$),
- Convolutional Neural Network from [4]



[4] G. Miloshevich et al., Phys. Rev. Fluids (Apr. 2023)

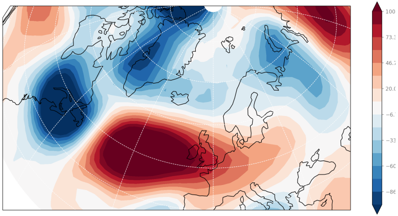
Outline

Return times of long series generated by SWG vs control run

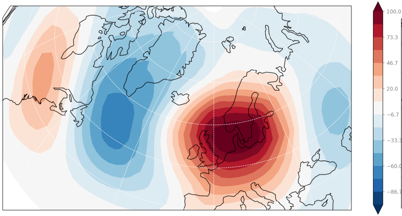


[5] https://github.com/georgemilosh/Climate-Learning, a public repo of our project

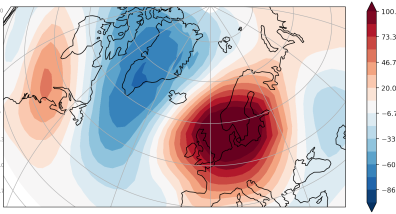
Synthetic teleconnections vs long control run and conclusions



most extreme $(A(\tau = 0) = 4.5K)$ from 80 years "training" dataset



composites of $(A(\tau = 0) \geq 4.5K)$ from synthetic SWG series



composites of $(A(\tau = 0) \geq 4.5K)$ from control run with 7200 years

- Probabilistic prediction occurs in the regime of lack of data requiring long datasets [6]
- SWG is a simple to implement and can be used to sample and predict unobserved extremes.
- Such data-driven algorithms have a potential to sample rare events: $\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow\Rightarrow$

scan for: Climate-Learning@GitHub

[7] https://www.cnrs.fr/fr/changements-climatiques-une-meilleure-prediction-des-canicules-grace-lia

[8] G. Miloshevich et al., Phys. Rev. Fluids (Apr. 2023)