Sampling and forecasting extreme heatwaves using analogs and neural networks

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

ENS DE LYON

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

[2]Departement de Physique Ecole Normale Superieure de Lyon

XAIDA: Artificial Intelligence for Detection and Attribution of Climate Extremes, 2022

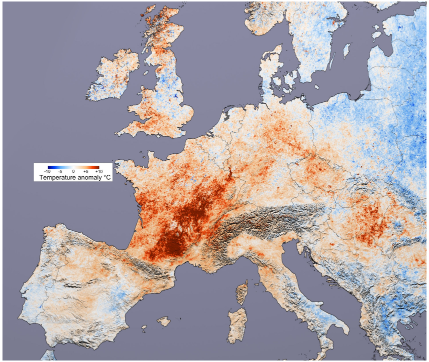
Outline

1. Introduction
2. Data-driven probabilistic forecasting of heatwaves
   - Serving calibrated probabilistic predictions
   - Convolutional Neural Networks
   - A regime of lack of data
   - Stochastic Weather Generator
   - Dimensionality reduction
3. Data-driven extreme event sampling
   - Return time plots
   - Teleconnection patterns
4. Intra-model comparisons
5. Future prospects

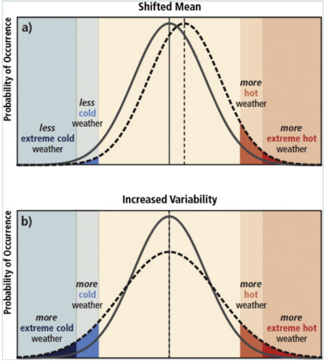
Outline

Studying extremes with models vs ML

- General Circulation Models (GCMs) when used for extremes of : [1]
  - At the regional scale, are still limited by the rarity of events
  - To capture processes requires running expensive simulations
  - Can machine learning be used to extract useful information from smaller datasets?



European heat wave 2003



Changes in temperatures[2]

[1] S. Seneviratne et al., A Special Report of Working Groups I and II of the IPCC (2012)
[2] S. E. Perkins, Atmospheric Research (2015)

Outline

1. Introduction
2. Data-driven probabilistic forecasting of heatwaves
   - Serving calibrated probabilistic predictions
   - Convolutional Neural Networks
   - A regime of lack of data
   - Stochastic Weather Generator
   - Dimensionality reduction
3. Data-driven extreme event sampling
   - Return time plots
   - Teleconnection patterns
4. Intra-model comparisons
5. Future prospects

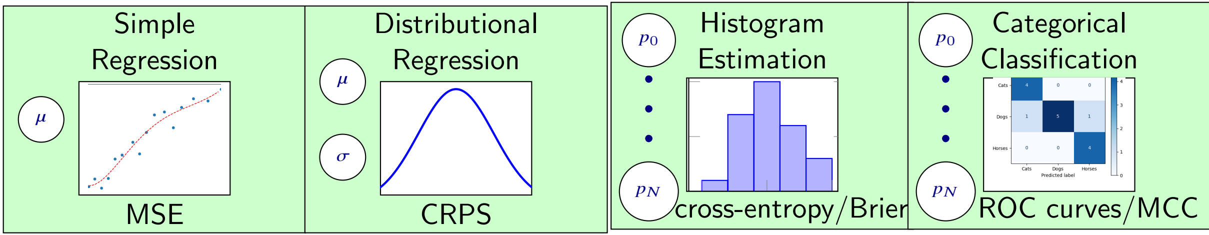
From pattern recognition to physical models and back

Bjerknes Analogs [3] → Physical Models → CPUs bottleneck → ML in modeling?[4]

- Recent success in deterministic intermediate range forecast with GraphCast [5]
- Recent papers report advances in predicting extreme heatwaves with ML [6] [7] [8]
- These Neural Networks are NOT trained for probabilistic prediction of extremes
- This is because the methods often used MSE or MCC as the target
- This is not optimal for UQ and probabilistic extreme event forecasting

[4] E. N. Lorenz, Journal of Atmospheric Sciences (1969)
[5] V. Balaji, Phil. Trans.of the Royal Soc.A: Math., Phys.and Eng. Sciences (2021)
[6] R. Lam et al., (Dec. 24, 2022)
[7] A. Chattopadhyay et al., Journal of Advances in Modeling Earth Systems (2020)
[8] V. Jacques-Dumas et al., Frontiers in Climate (2022)
[9] I. Lopez-Gomez et al., Artificial Intelligence for the Earth Systems (Dec. 19, 2022)

Probabilistic scores: what remains to be done for heatwaves



- Probabilistic forecasting of heatwaves using Brier Score with Random Forest[10]
- BS is a strictly proper score but depends on never occurred events

$$\mathrm{BS} = \frac{1}{n}\sum_{k=1}^{n} |p_k - \hat{e}_k|^2 \tag{1}$$

Logarithmic (a.k.a, cross-entropy) score is suitable for rare events[11]

[10] C. v. Straaten et al., Monthly Weather Review (May 1, 2022)
[11] R. Benedetti, Monthly Weather Review (2010)

Defining heatwaves and Normalized Log Score

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

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

Duration: $T = 14$ days

Area D - "France" / "Scandinavia"

- The goal: find $P(A(t) > a|X(t-\tau), \tau)$ with lead time τ
- Logarithmic (cross-entropy) score suitable for rare events[12]
- Threshold α is chosen so that $Y = 1$ is above 95 percentile

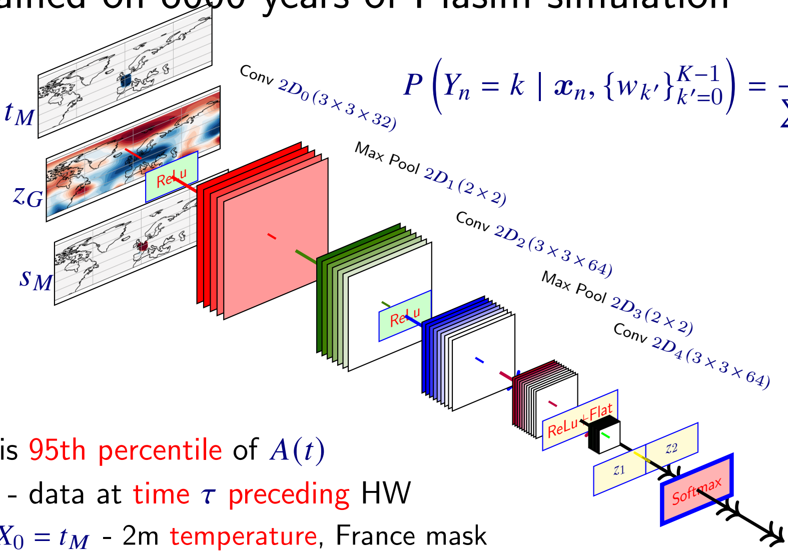
$$S\left[\hat{p}_Y(X)\right] = -\sum_{k=0}^{K-1} Y_k \log\left[\hat{p}_k(x)\right], \quad K = 2 \text{ for binary} \qquad (3)$$

Normalized Log Score (NLS): subtract climatological prediction

$$\text{NLS} = \frac{-\sum_i \overline{p}_i \log \overline{p}_i - \mathbb{E}\left\{S\left[\hat{p}_Y(X)\right]\right\}}{-\sum_i \overline{p}_i \log \overline{p}_i} \qquad (4)$$

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

CNN trained on 8000 years of Plasim simulation



$$P\left(Y_n = k \mid \boldsymbol{x}_n, \{w_{k'}\}_{k'=0}^{K-1}\right) = \frac{e^{-x_n^T w_k}}{\sum_{k'=0}^{K-1} e^{-x_n^T w_{k'}}},$$

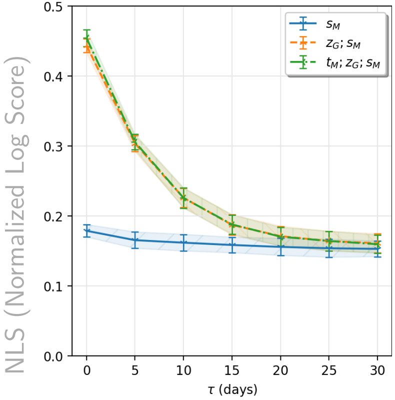
Def: HW is 95th percentile of $A(t)$

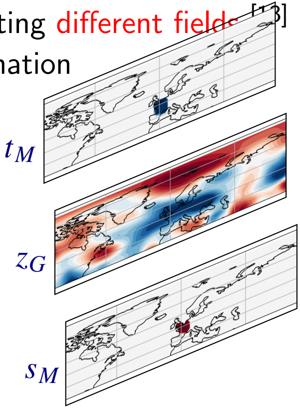
- $X(\tau)$ - data at time τ preceding HW
  - $X_0 = t_M$ - 2m temperature, France mask
  - $X_1 = z_G$ - 500hPa geopotential height
  - $X_2 = s_M$ - soil moisture, France mask

Training performed with Tensorflow-GPU 2.4 on 554400 samples that are 22 by 128 by 3

Geopotential/soil moisture contributions

- k-Fold cross-validation is used to assess the variance of the skill with $k = 10$ folds
- The CNN was optimized using cross-validation tuning hyperparameters
- We present the plots of NLS vs lead time τ selecting different fields [13]
- s_M has long-term, while z_G has short-term information



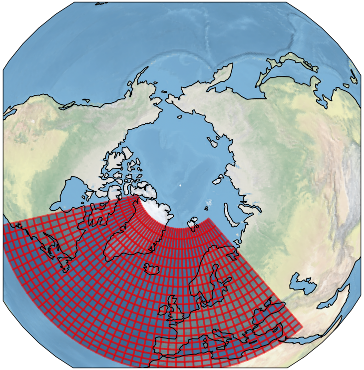


Possible field inputs in stacked architecture which works better for heatwave classification

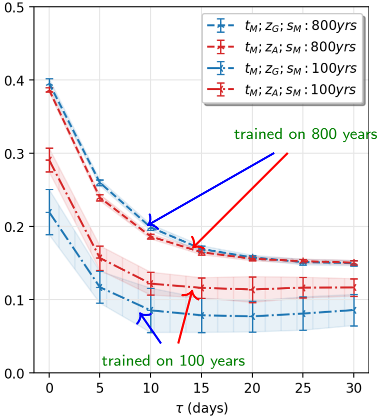
[13] G. Miloshevich et al., Phys. Rev. Fluids (Apr. 2023)

Learning regional correlations vs data length

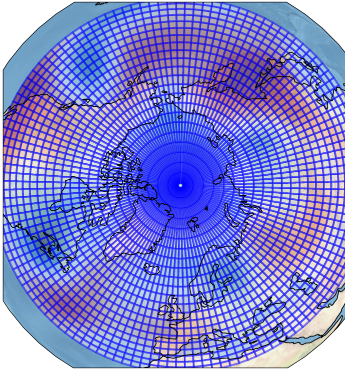
- We present the plots of NLS vs lead time time τ
- Having less data, some global teleconnections not represented well [14]
- In reanalysis only the data from 1950 to present is available



z_A, North Atlantic



NLS data reduction

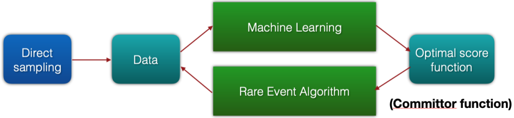


z_G North Hemisphere

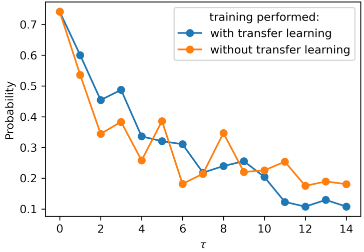
[14] G. Miloshevich et al., Phys. Rev. Fluids (Apr. 2023)

Smoothness of the committor & transfer learning

- $q = q(\tau)$ is expected to be a smoothly increase closer to the heat wave
- This property is epected to play a role in rare event algorithm
- We achieve this by transfer learning applied to successive τ [15]



Training pipeline

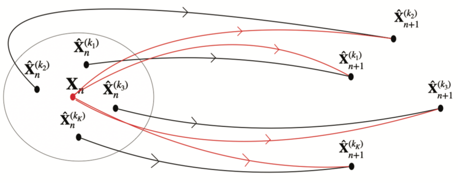


q_{t_M, z_G, s_M} vs transfer learning

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

Stochastic Weather Generator a.k.a. Analog Markov chain

Analogs are sought using $$X_{n_\star} = \underset{\{X_n\}}{\operatorname{argmin}} \{d\,(x, X_n)\} \qquad (5)$$



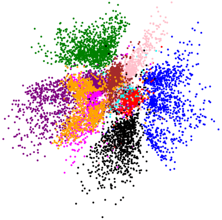
Analog method

- SWG is used often to estimate the probability of circulation models
- 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)

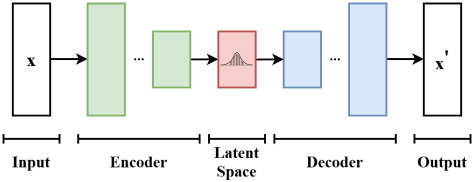
$$d\,(X_1, X_2) = \left[\frac{\alpha}{\sigma_Z^2 \dim(Z_G)} \sum_{i=1}^{\dim(Z)_G} \left(\Delta Z_G^I\right)^2 + \frac{1}{\sigma_T^2}\left(\Delta T_L^I\right)^2 + \frac{1}{\sigma_S^2}\left(\Delta S_L^I\right)^2\right]^{\frac{1}{2}} \qquad (6)$$

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

Alternative solution: Variational Autoencoder



MNIST latent space



Schematics of a (variational) autoencoder

$$p(z \mid x) = \frac{p(x \mid z)p(z)}{\int p(x \mid u)p(u)du}$$

$$p(z) \equiv \mathcal{N}(0, I)$$

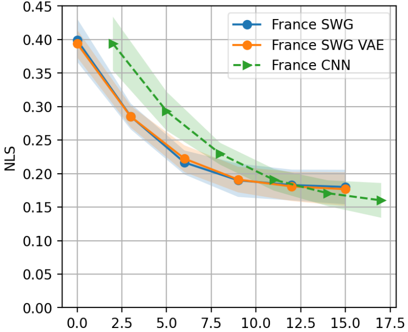
$$p(x \mid z) \equiv \mathcal{N}(f(z), cI)$$

$$p(z \mid x) \sim q_x(z) \equiv \mathcal{N}(g(x), h(x))$$

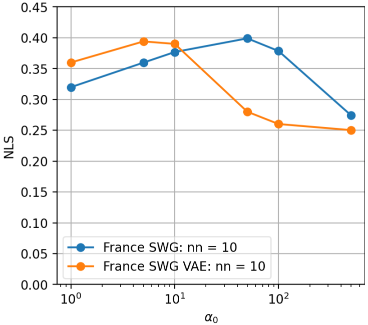
$$(f^*, g^*, h^*) = \underset{(f,g,h) \in F \times G \times H}{\arg\max} \left(\mathbb{E}_{z \sim q_x} \left(-\frac{\|x - f(z)\|^2}{2c} \right) - KL\left(q_x(z), p(z)\right) \right)$$

Stochastic Weather Generator vs CNN

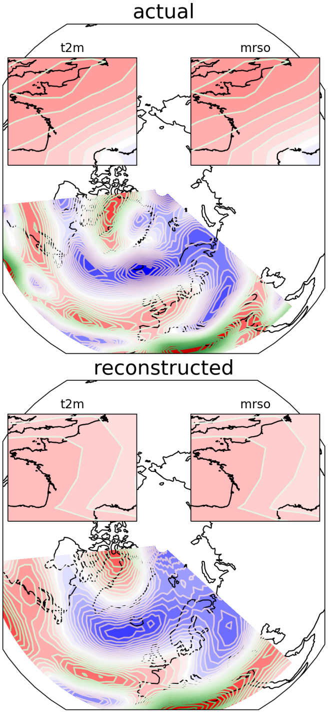
- Variational AutoEncoder (VAE) reduces dimension
- We SWG to the latent space of VAE and optimize
- 10000 trajectories per validation day are launched



Markov Chain ($\dim z = 16$), CNN



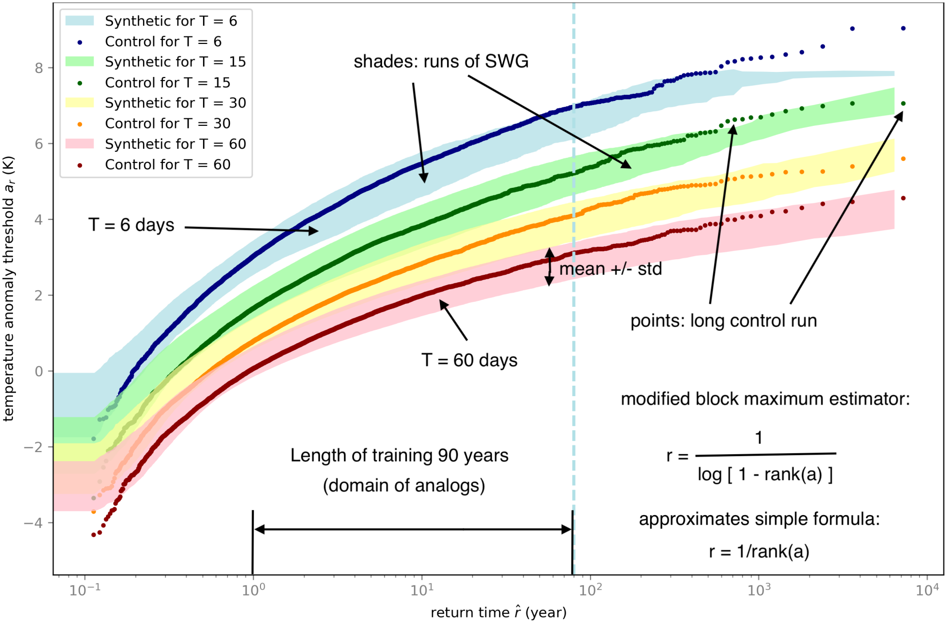
Optimizing weights



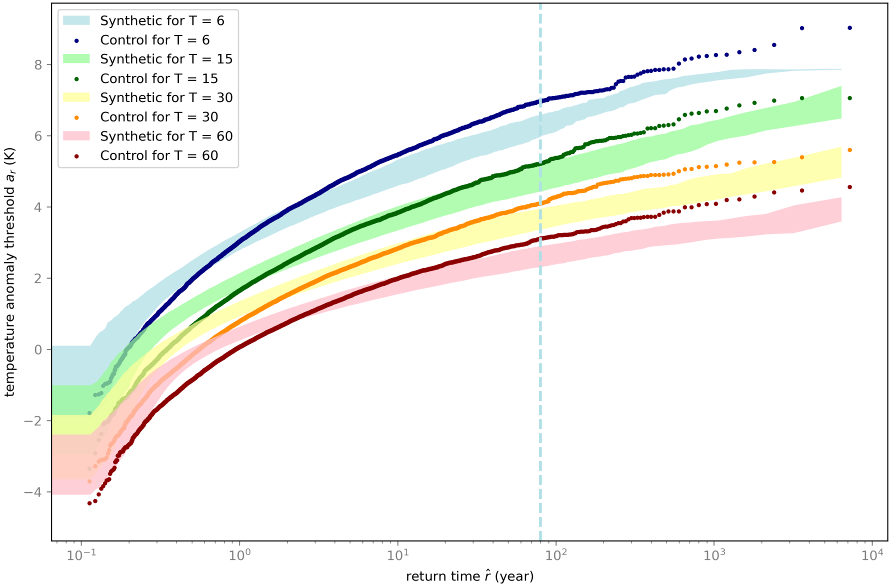
[15] G. Miloshevich et al., In preparation (2022)

Outline

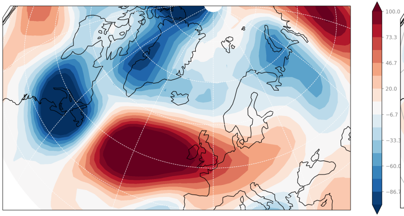
Return times of long time series generated by SWG ($\alpha_0 = 1$)



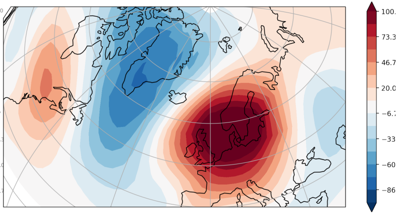
Return times of long time series generated by SWG ($\alpha_0 = 50$)



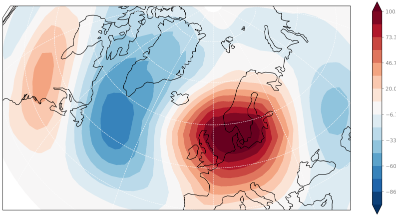
Generating synthetic SWG teleconnections



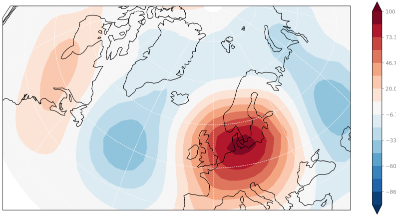
90 years extreme
$(A(\tau = 0) = 4.5K)$



Control run composite
$(A(\tau = 0) \geq 4.5K)$



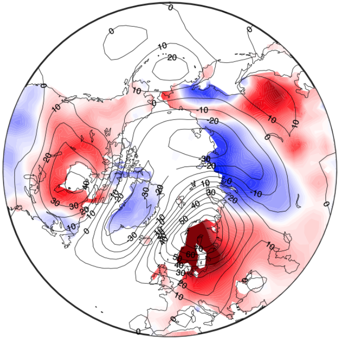
SWG composite trained on 90 yrs
$(A(\tau = 0) \geq 4.5K)$, $\alpha_0 = 50$



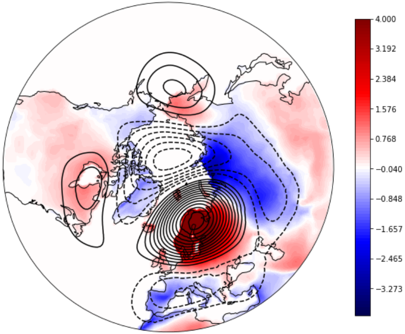
SWG composite trained on 90 yrs
$(A(\tau = 0) \geq 4.5K)$, $\alpha_0 = 1$

Outline

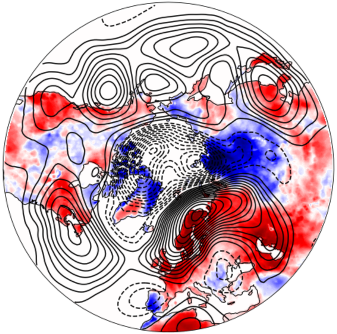
Plasim/CESM/ERA5 teleconnections (Scandinavian heatwave)



Plasim rare event[16]



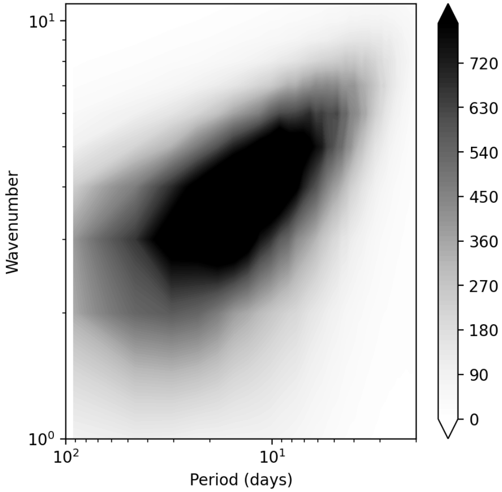
CESM composite[17]



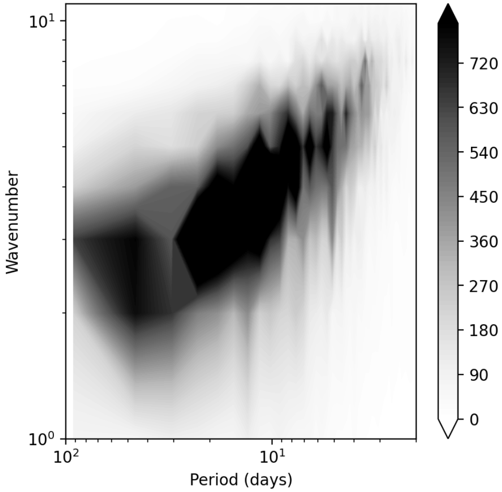
ERA5 July 2018

[16] F. Ragone et al., Proceedings of the National Academy of Sciences (2018)
[17] G. Miloshevich et al., “Drivers of midlatitude extreme heat waves revealed by analogues and machine learning”, in Egu general assembly conference abstracts, EGU General Assembly Conference Abstracts (Apr. 2021), EGU21–15642

Hayashi spectra

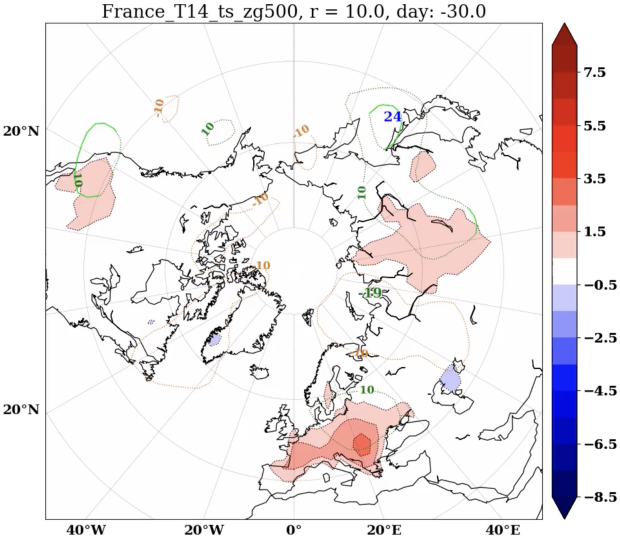


$H_E(k,\omega)$, th $= \infty$.



$H_E(k,\omega)$, th $= -4.5$.

A composite of 10 most intense heatwaves



- 1000 year long PlaSim (*Planet Simulator*)
- No daily cycle: stronger land-atmosphere coupling
- The maps are conditioned to 10 most extreme heatwaves
- The CNN will be trained on 8000 years of PlaSim simulation with daily cycle

Outline

Concluding remarks

- Summary:
  - Intra-model spatio-temporal correlations of heatwaves reveal robust teleconnections
  - Probabilistic prediction occurs in the regime of lack of data requiring long datasets
  - How fields are encoded in the CNN or SWG affects the skill significantly
  - SWG sampling of extreme heatwaves validated on a very long GCM run
- Possible future steps:
  - Improve sampling/prediction applying rare event algorithms to high res models
  - Improve prediction of extreme events using transfer learning: across datasets
  - Study other extremes and Use explainable AI to reveal the precursors
  - Use causal inference and/or physical losses/architectures, e.g. GNNs
- Thank you for your attention!

Climate-Learning@GitHub