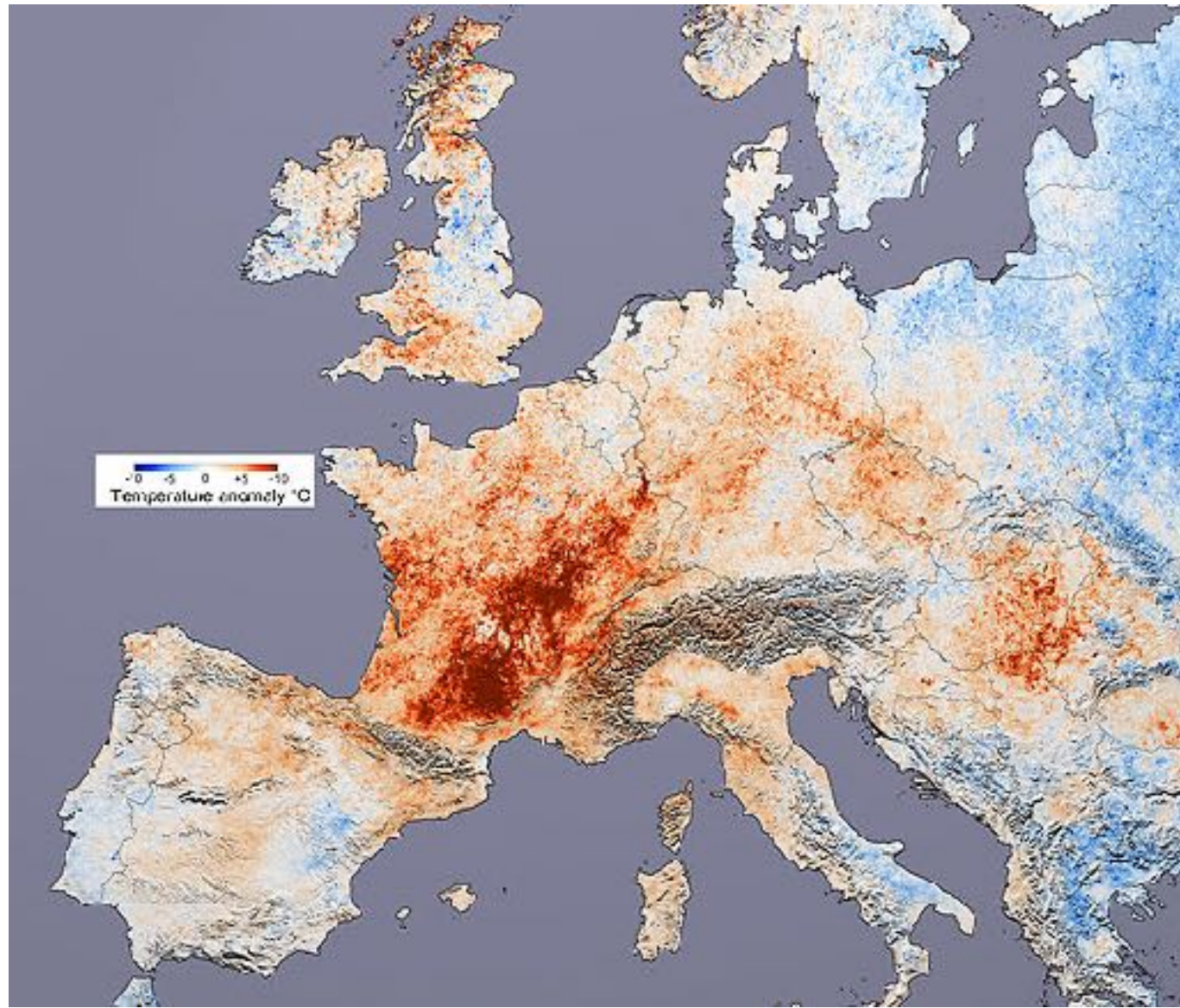




Drivers of midlatitude extreme heat waves revealed by analogues and machine learning

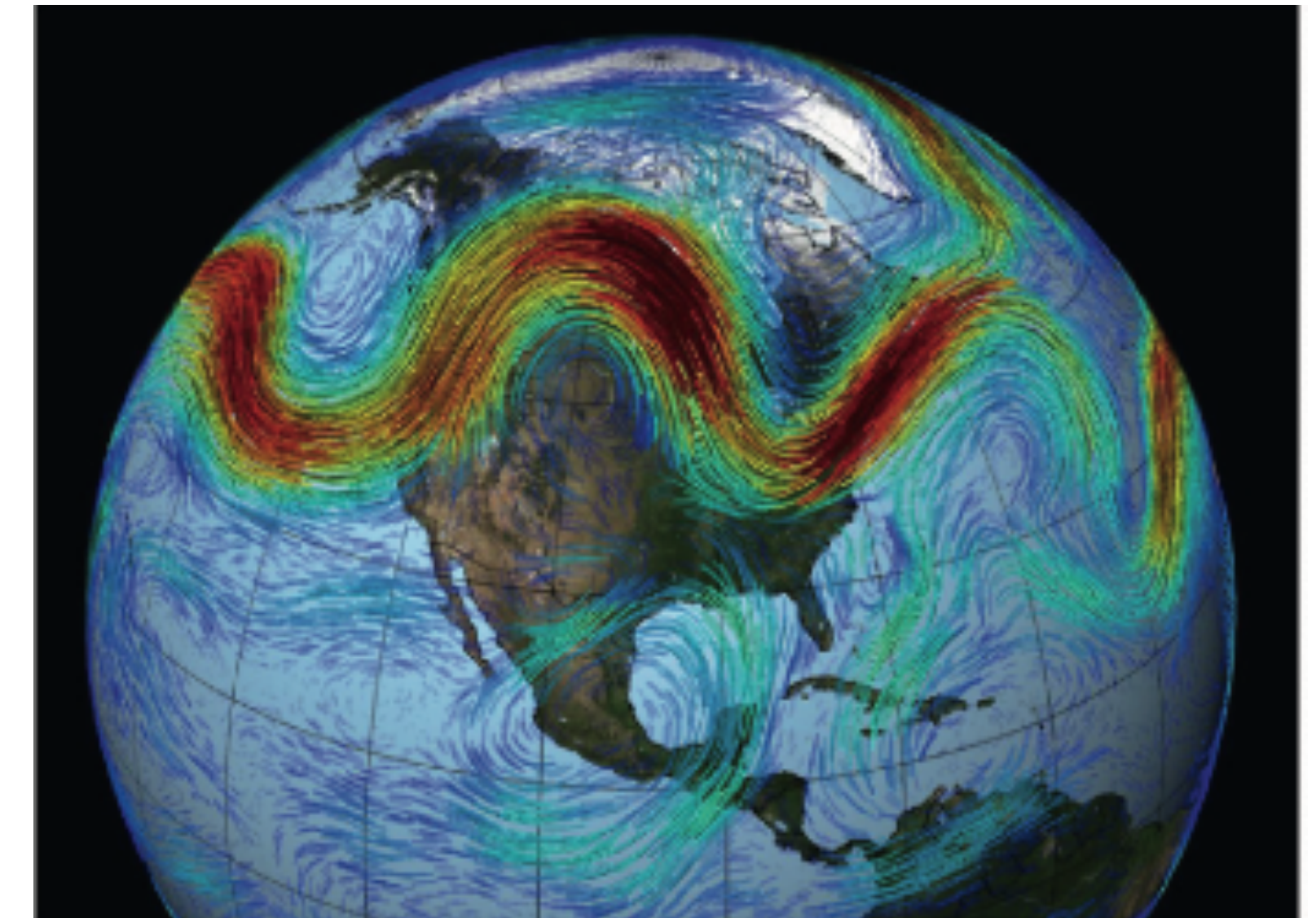
George Miloshevich, Dario Lucente, Corentin Herbert, and Freddy Bouchet



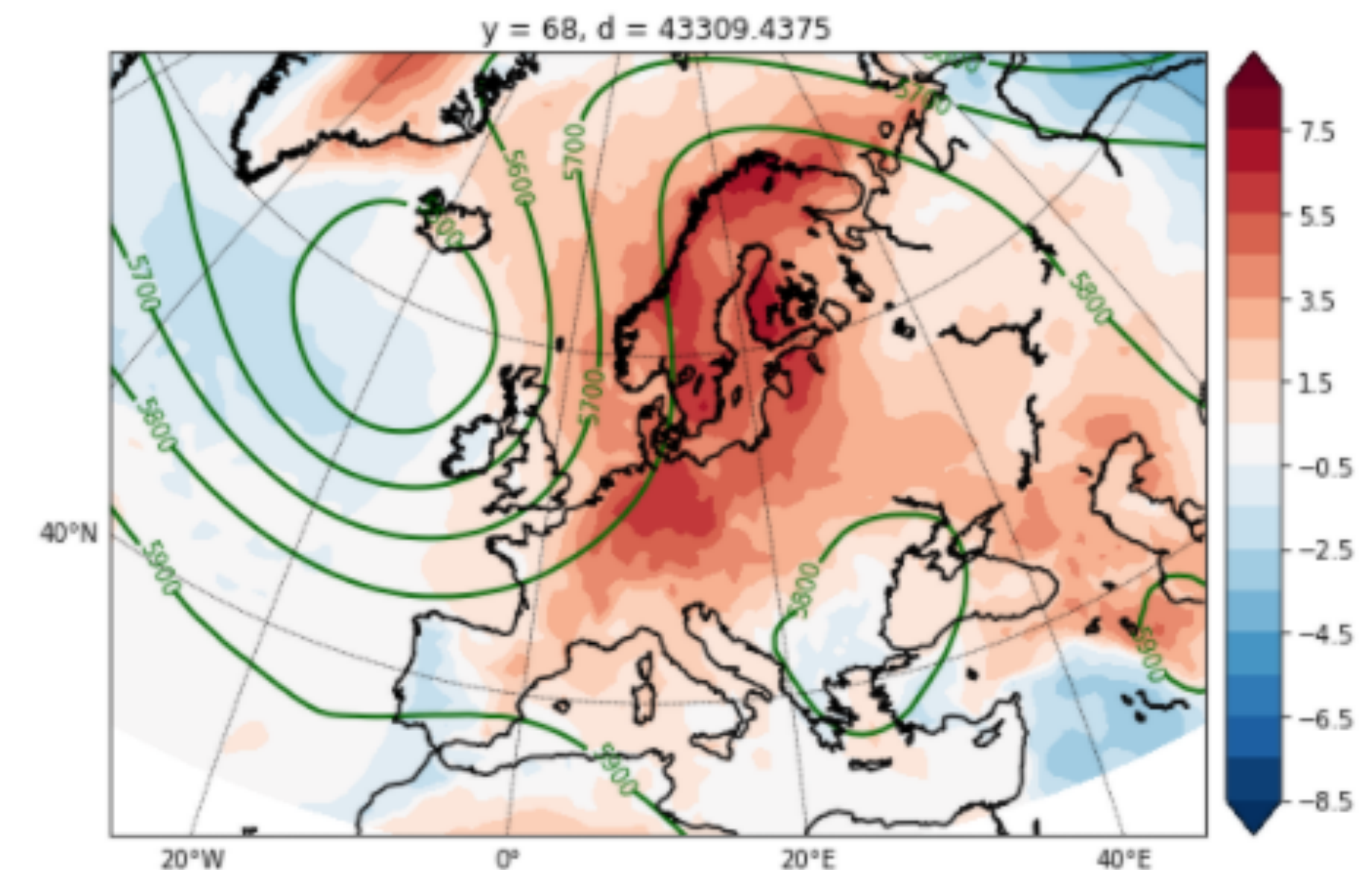
- Persistent (from several weeks up to a month) anomalies associated with strongest heat waves
 - Western European heat wave 2003
 - Russian heatwave 2010
- Impacts involve: excess mortality, crop failure
- IPCC predicts increase in frequency and severity

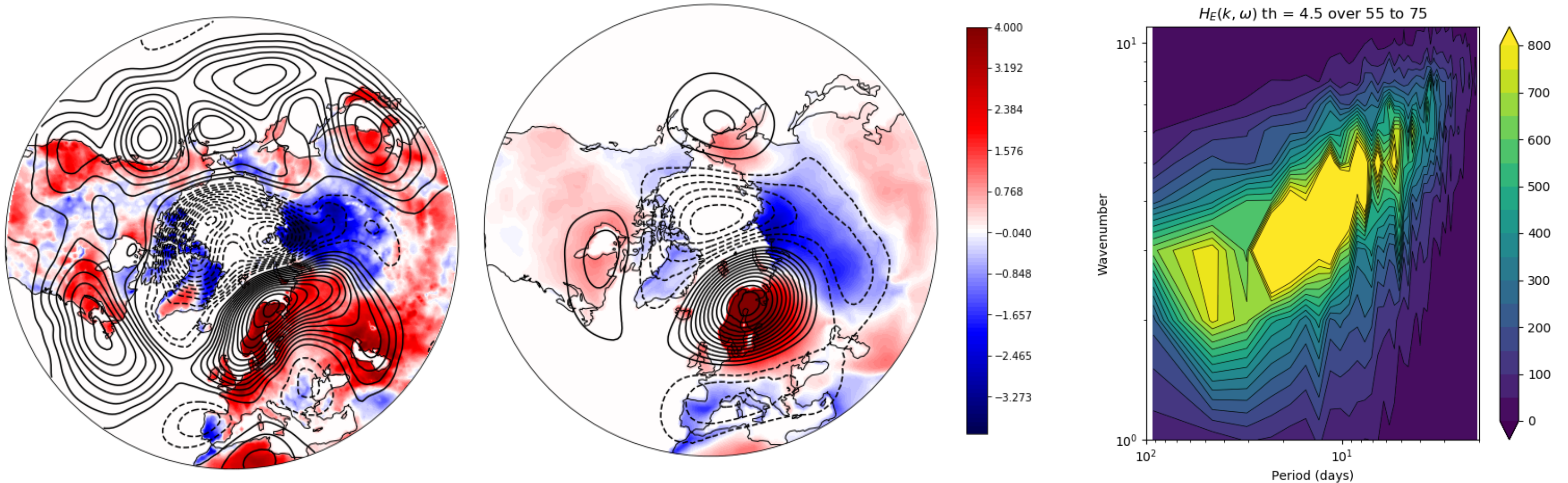


- Midlatitudes dominated by **jet streams**
- Strong narrow westward currents at 45 N, ~ 40 m/s
- Jet stream meandering and nonlinear Rossby waves
- Cyclonic/anticyclonic anomalies engender weather
- Anomalies may arise due to
 - Rossby wave breaking (**blocking**)?
 - Shifts of the jet stream?
 - Precursors?
- Extreme events are rare thus are costly to study
- **Rare event algorithm** to improve model statistics



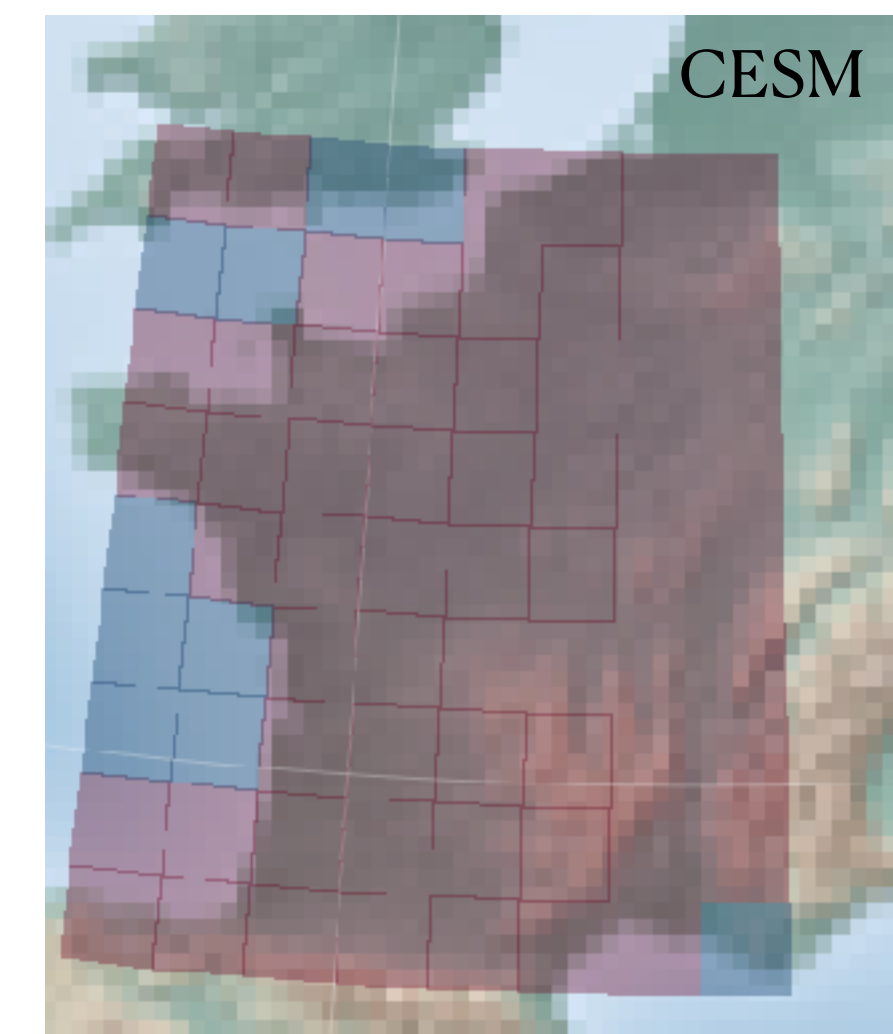
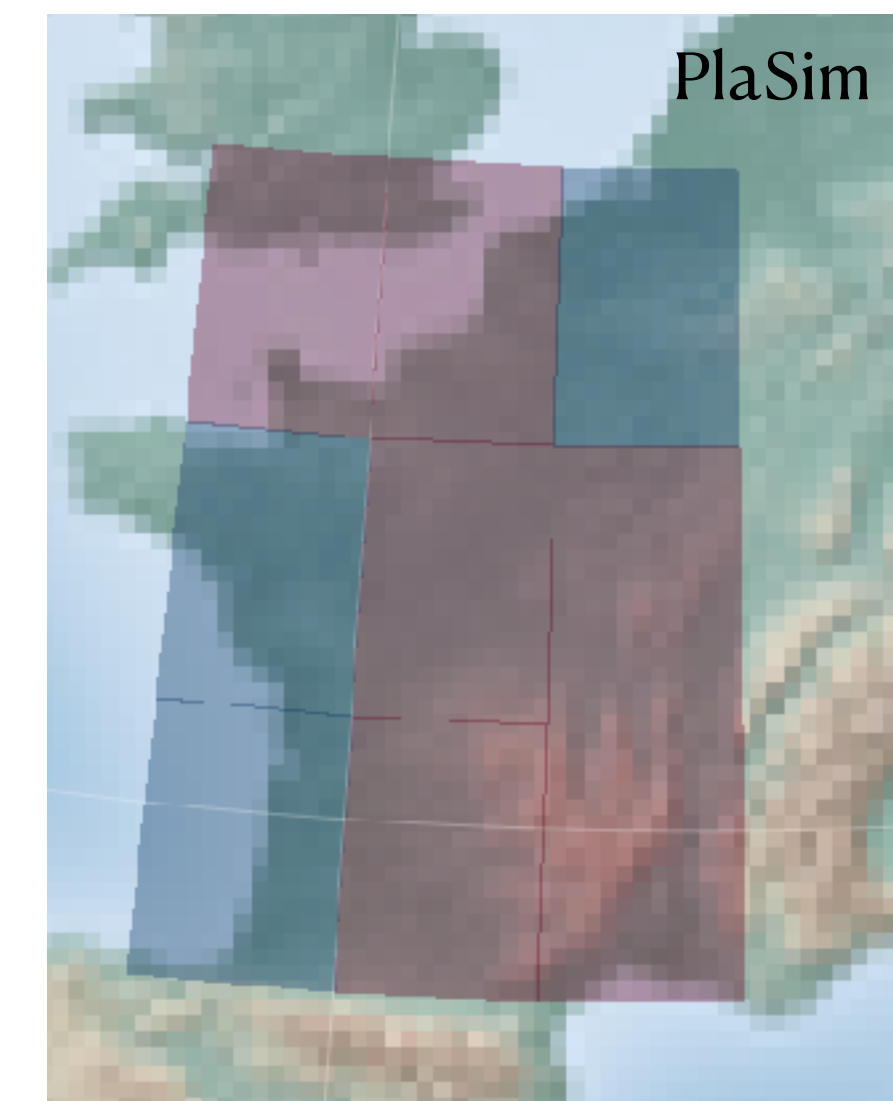
Snapshot of wind speed velocity (image courtesy of NASA)





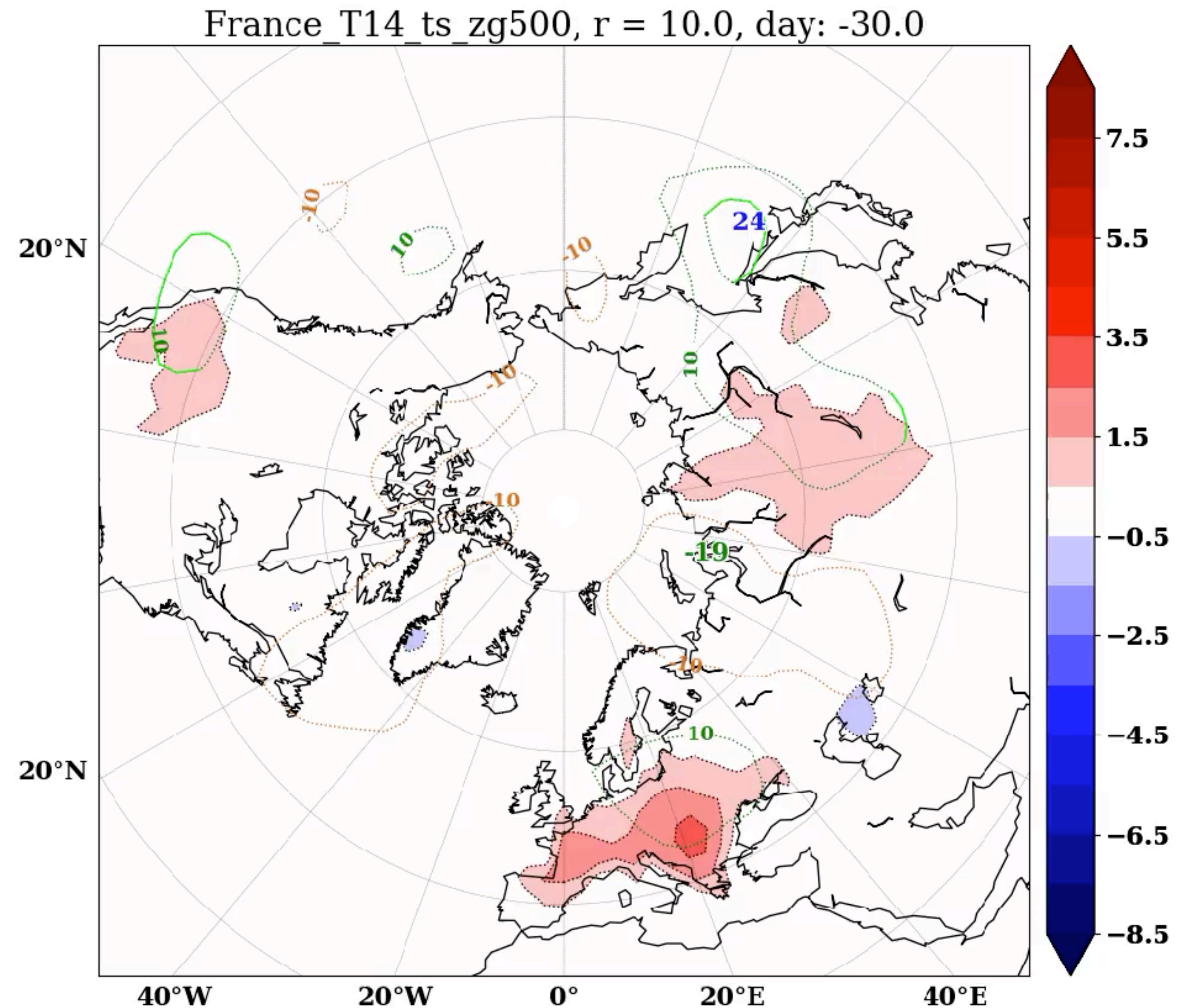
- ERA5 detrended July 2018
- CESM 4 K composite on Scandinavia
- Hayashi spectrum
- (Eastward/Westward/Stationary) decomposition
- on 55 N to 75 N conditioned to a heatwave
- Some of the features of both 2m temperature and 500 mbar geopotential captured
- Hayashi spectrum shows that in addition to regular Rossby waves we find an island at mode 3, 50 days long period

- We use a **GCM** which models the atmosphere, soil, ice etc
- Static ocean with statistically **stationary** climate (fixing CO₂)
- Slower variability of SST is not included (such as ENSO)
- Model 1: **PlaSim** - a trade-off between complexity and speed.
 - Resolution: 2.8 by 2.8 degrees. 10 vertical atmospheric levels, 5 soil
 - Parametr.: specific and latent heat, cloud formation and radiation
 - The SST and the ice cover is repeated cyclically every year
 - We are suppressing daily cycle but maintain seasonal cycle
- Model 2: **CESM** . Resolutions: about 1 by 1 degrees
 - More precise land-atmosphere interactions, vegetation, radiation
 - CAM4 atmospheric model with 26 vertical levels, we include daily cycle
 - More precise hydrology. Soil has 10 soil levels and 5 bedrock levels



$$X^r(\mathbf{x}, \tau) = \frac{1}{N_a} \sum_{\{i|a_i \geq a^r\}} X(\mathbf{x}, t_i - \tau).$$

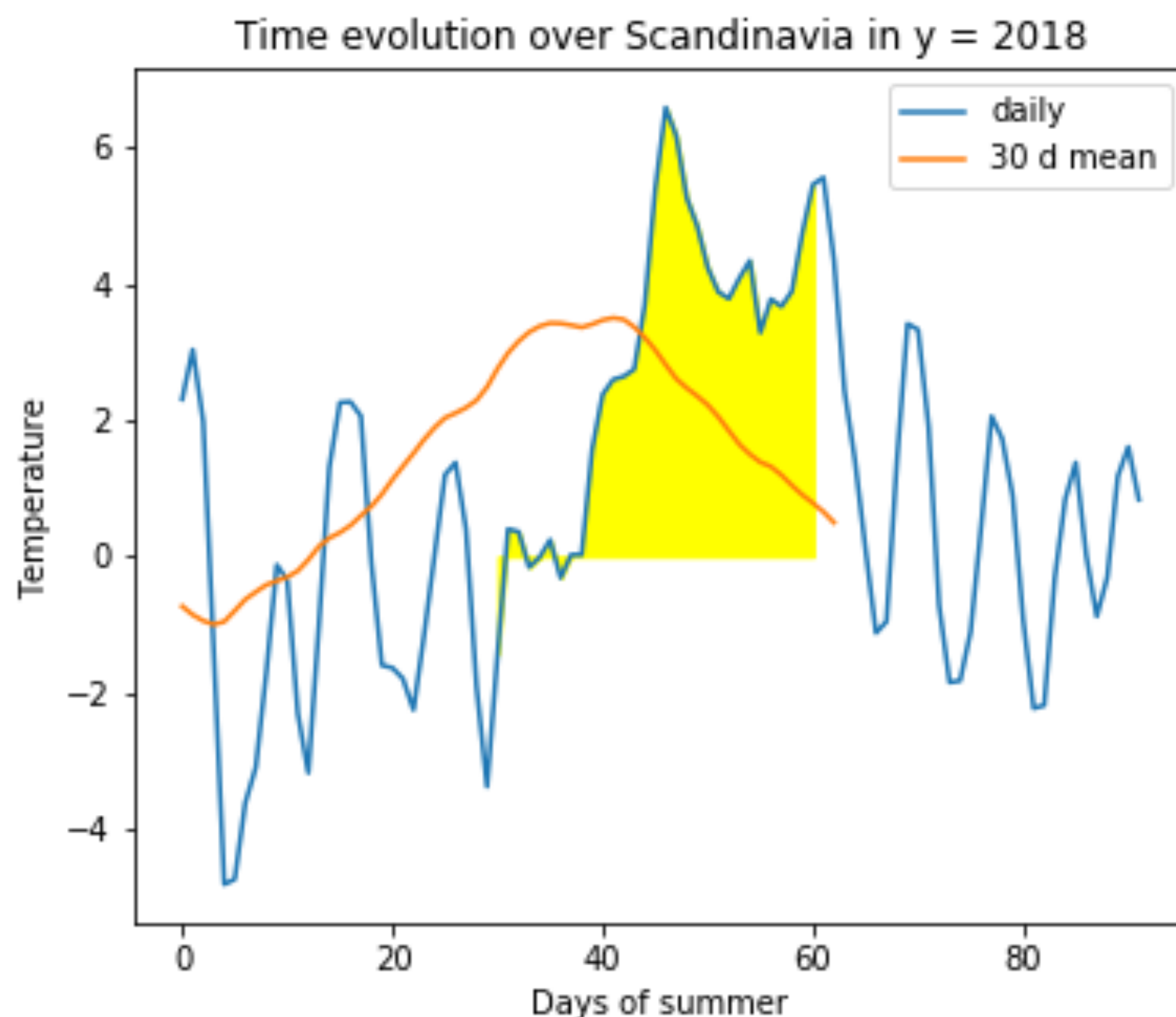
- -30 days: 1 degree preheating (related to PlaSim)
- -10 days: Arctic cyclonic anomaly emerges
- -5 days: Canada-Greenland dipole shifts eastwards
- 0 days: Formation of a **tripole** system with a heat wave
- 14 days: structure dissolves



- Running mean anomaly:

- One heat wave per year

$$a_i := A(t_i) = \max_{t \in \text{summer}(i)} \{A(t)\},$$



$$A(t) := \frac{1}{T} \int_t^{t+T} \frac{1}{|\mathcal{D}|} \int_D (T_s - \mathbb{E}(T_s))(\vec{r}, u) d\vec{r} du$$

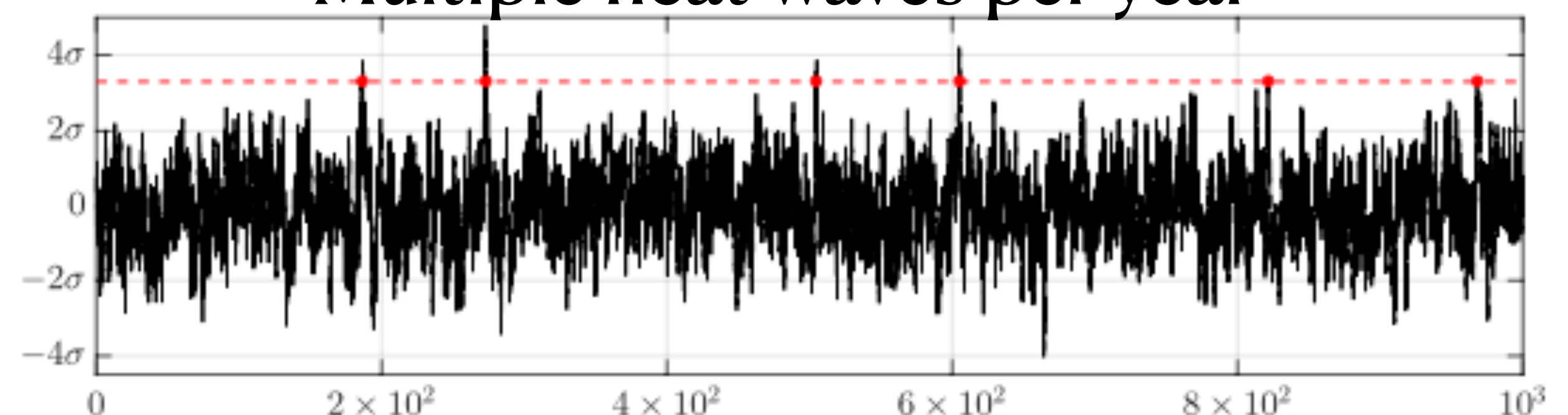
- Using modified definition for the return time

$$\hat{r}'_B(a) = - \frac{\Delta T}{\ln \left(1 - \frac{1}{M} \sum_{m=1}^M s_m(a) \right)}.$$

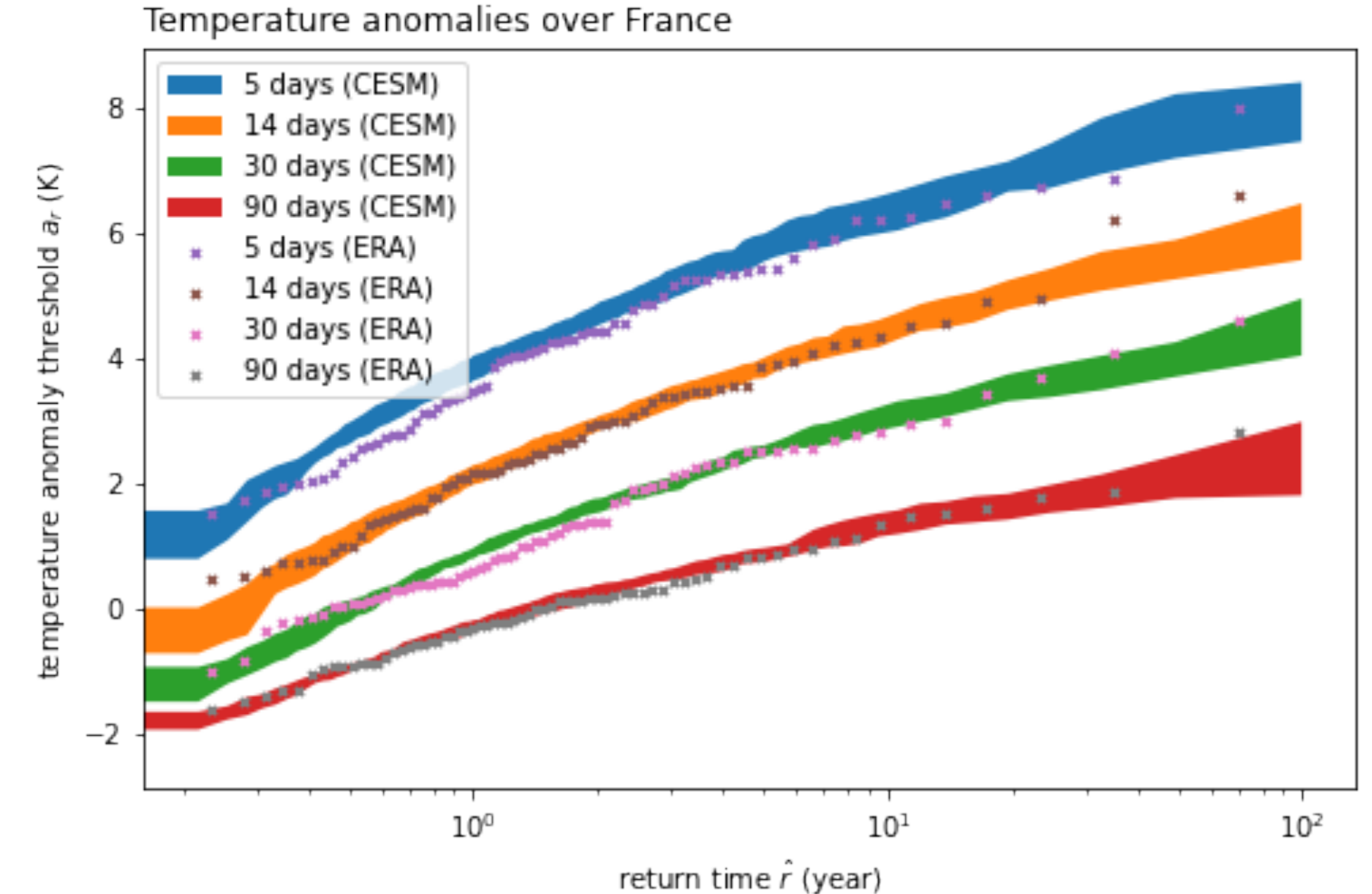
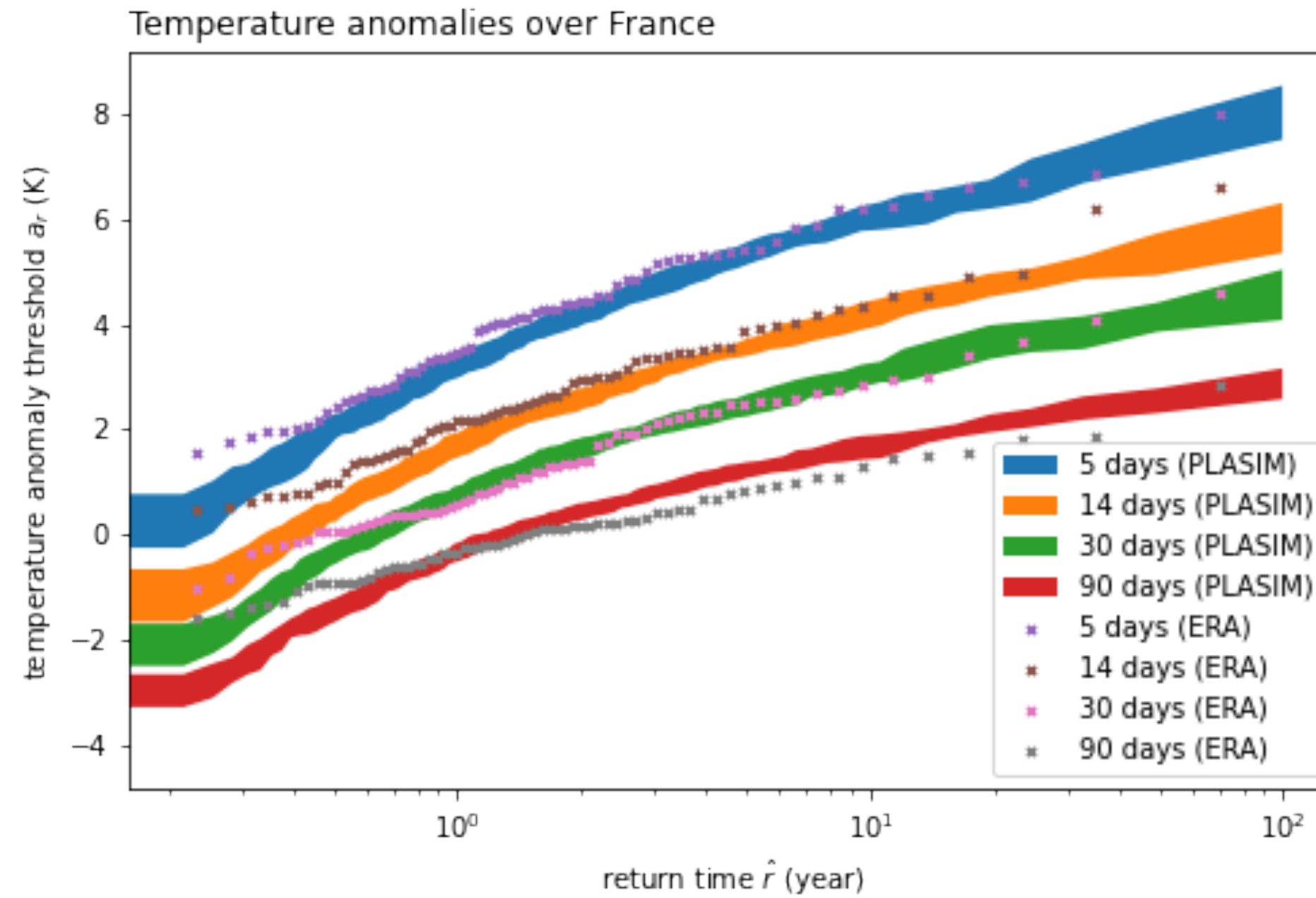
Using Poisson statistics

of heatwaves for a threshold a

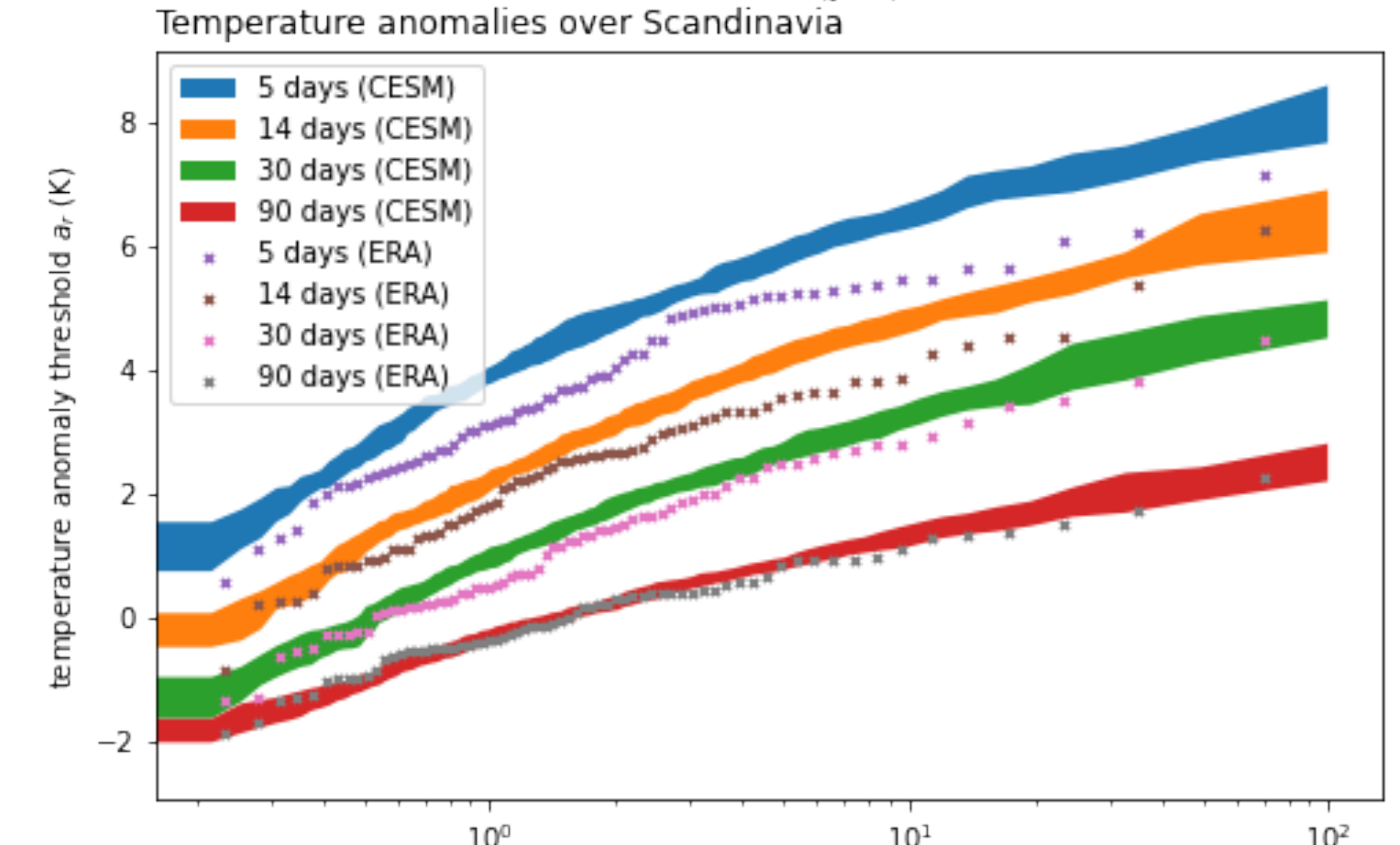
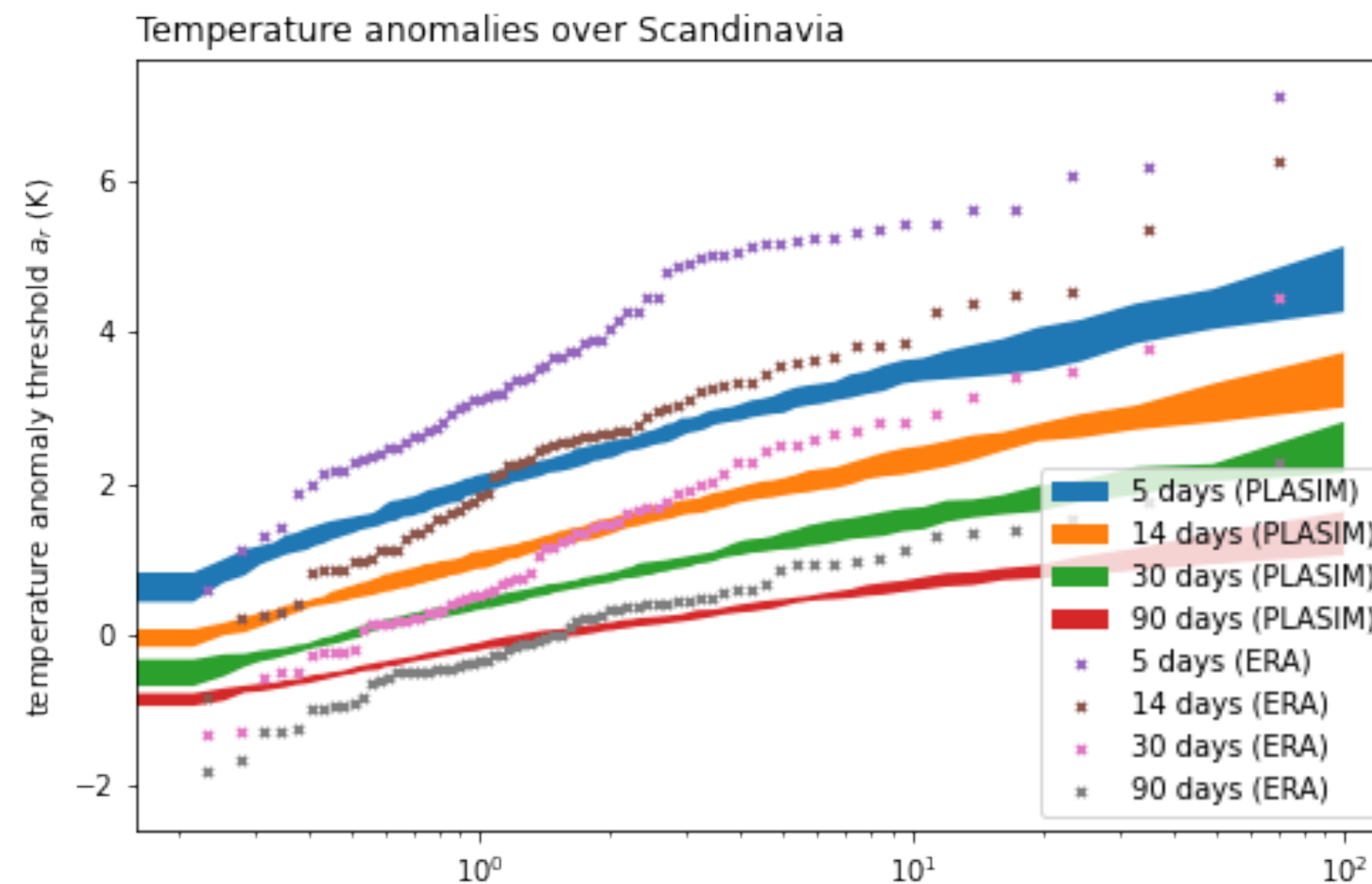
- Multiple heat waves per year



- France:
- PlaSim
- CESM



- Scandinavia:
- PlaSim
- CESM



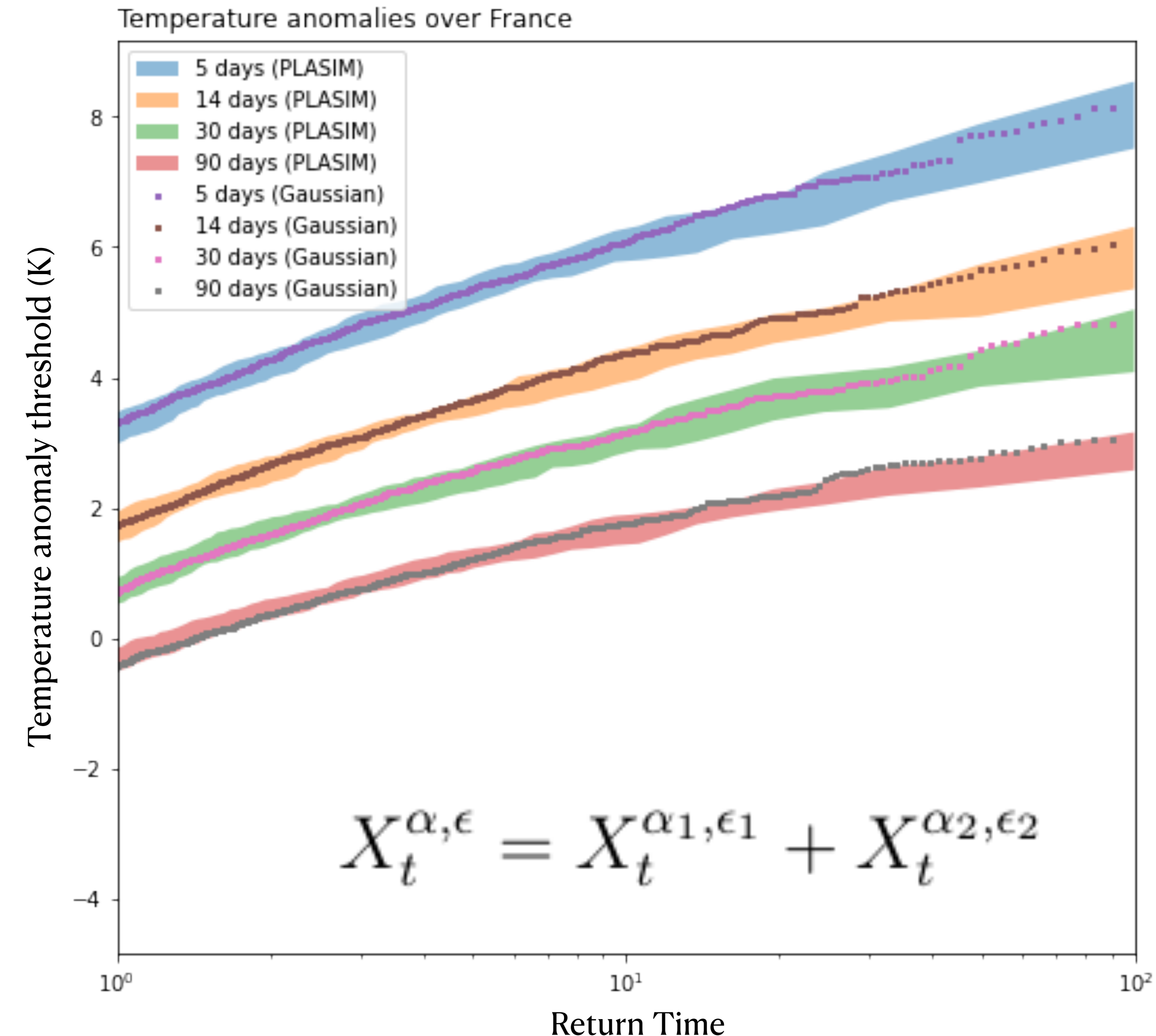
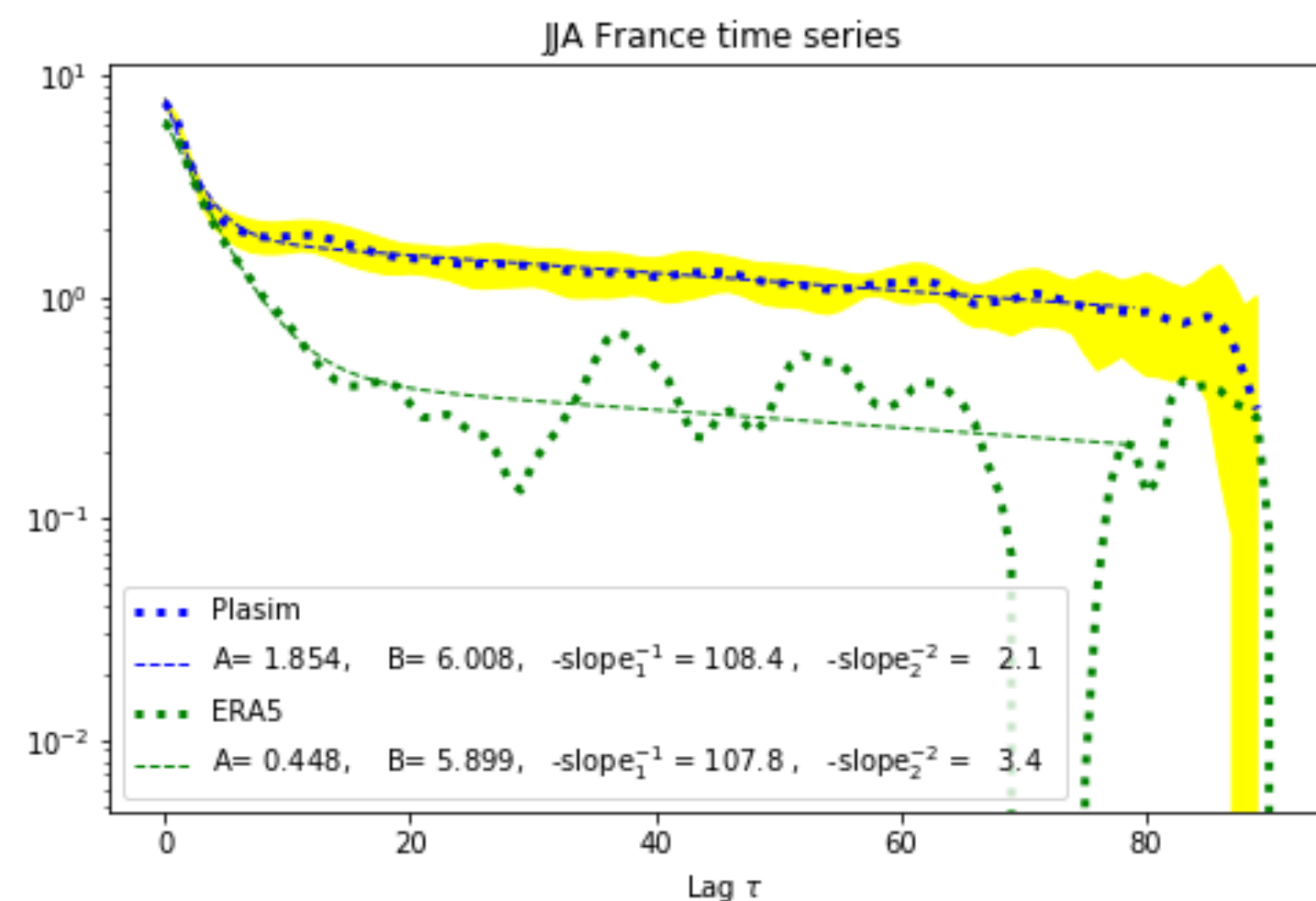
- Given time series analysis we may attempt modelling via a simple stochastic process

$$dX_t^{\alpha, \epsilon} = \underbrace{-\alpha X_t}_{\text{Drift vector}} dt + \underbrace{\sqrt{2\epsilon} dW_t}_{\text{Gaussian white noise}}$$

$$P_s(x) = \sqrt{\frac{\alpha}{\pi\epsilon}} \exp\left(-\frac{\alpha x^2}{\epsilon}\right) \quad \sigma = \sqrt{\epsilon/\alpha}.$$

- OU has one time scale:

$$\langle X(t)X(s) \rangle_s = \frac{\epsilon}{2\alpha} \exp(-\alpha|t-s|)$$



- Markov process satisfies: $p(\mathbf{x}_1, t_1 | \mathbf{x}_3, t_3) = \int d\mathbf{x}_2 p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3),$

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}(t)) + \sqrt{2D}\sigma(\mathbf{x})\boldsymbol{\xi}(t),$$

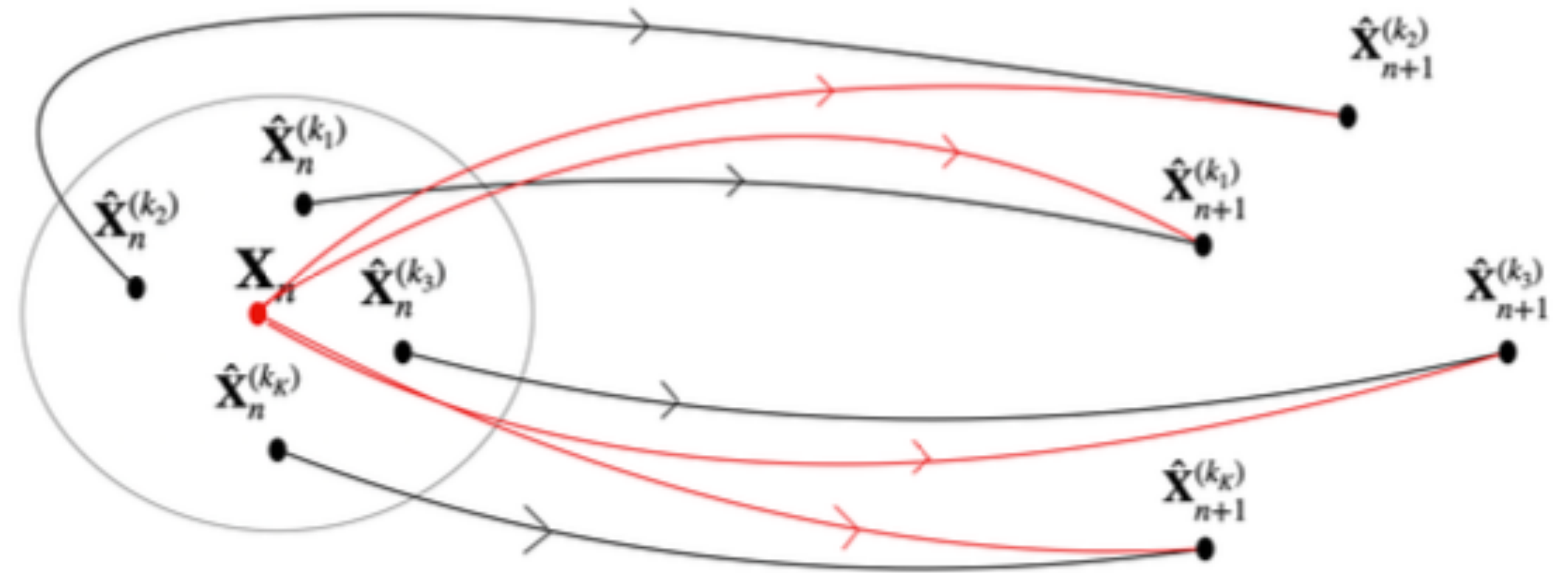
Drift vector Gaussian white noise

- Following Lorentz (1969) we employ **method of analogues**

- “Closest” analogues are searched

$$\mathbf{X}_{n*} = \underset{\{\mathbf{X}_n\}}{\operatorname{argmin}} \{ \mathcal{D}(\mathbf{x}, \mathbf{X}_n) \},$$

$$T \in \mathcal{M}^{N_t \times K}$$

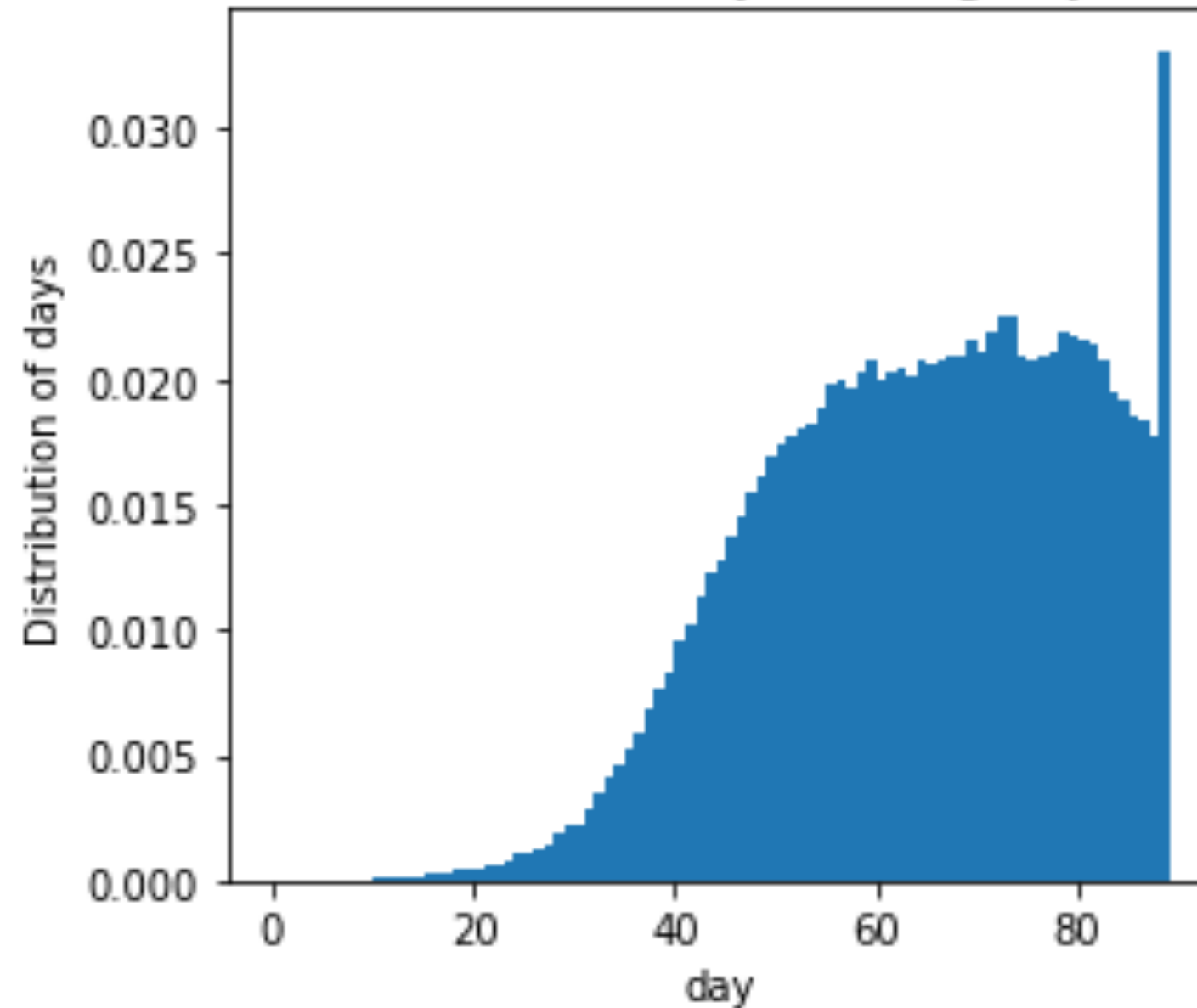


Among K nearest neighbours analogues chosen randomly

- Distance** defined:

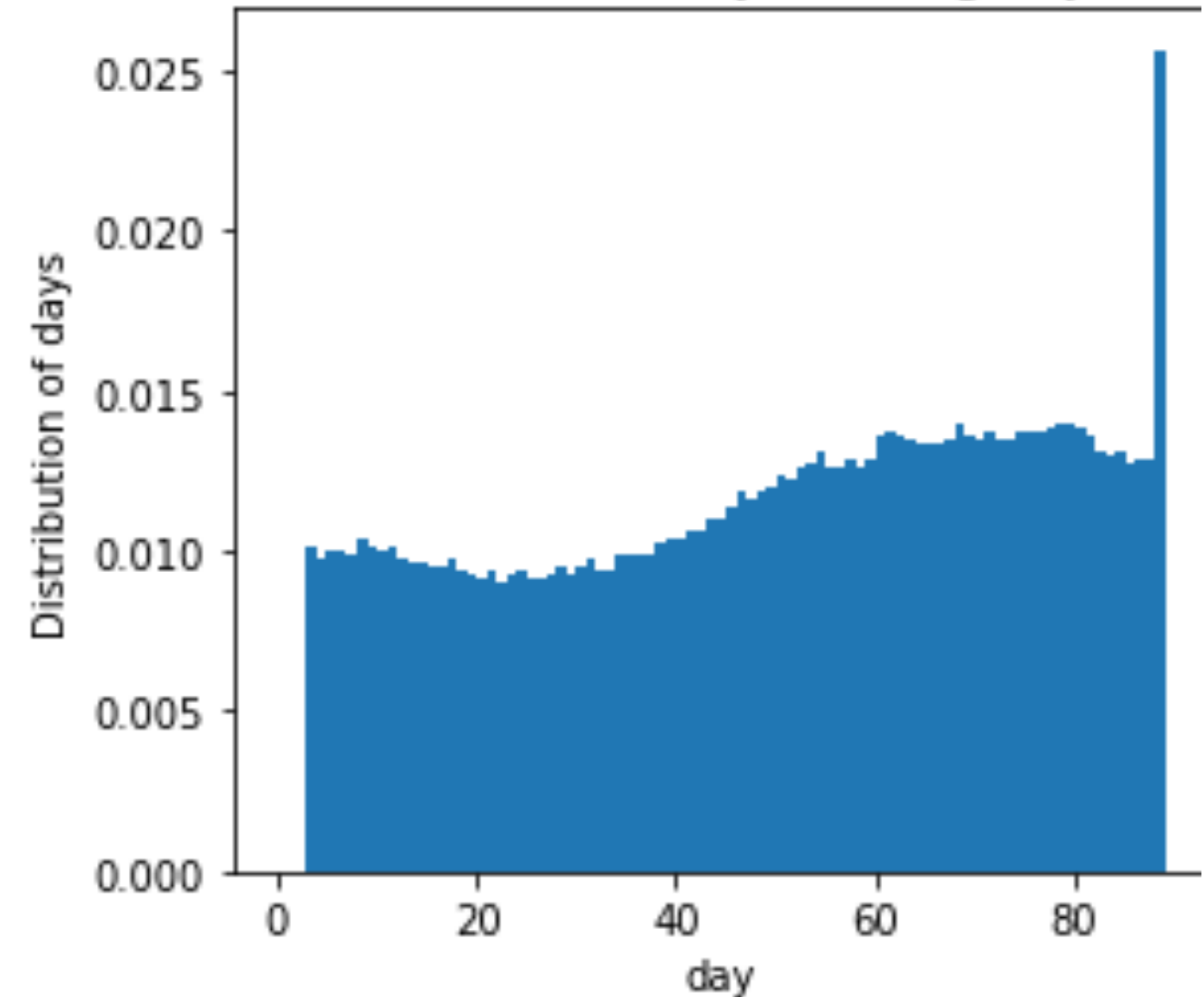
$$d^l(X_{n_1}, X_{n_2}) = \left[\int d\mathbf{r} \left(\frac{(\tilde{T}_a(\mathbf{r}, t_{n_1}) - \tilde{T}_a(\mathbf{r}, t_{n_2}))^2}{\sigma_{\tilde{T}_a}^2(\mathbf{r})} + \frac{(\tilde{Z}(\mathbf{r}, t_{n_1}) - \tilde{Z}(\mathbf{r}, t_{n_2}))^2}{\sigma_{\tilde{Z}}^2(\mathbf{r})} \right) \right]^{\frac{1}{2}}.$$

Occurrence of each day in a long trajectory



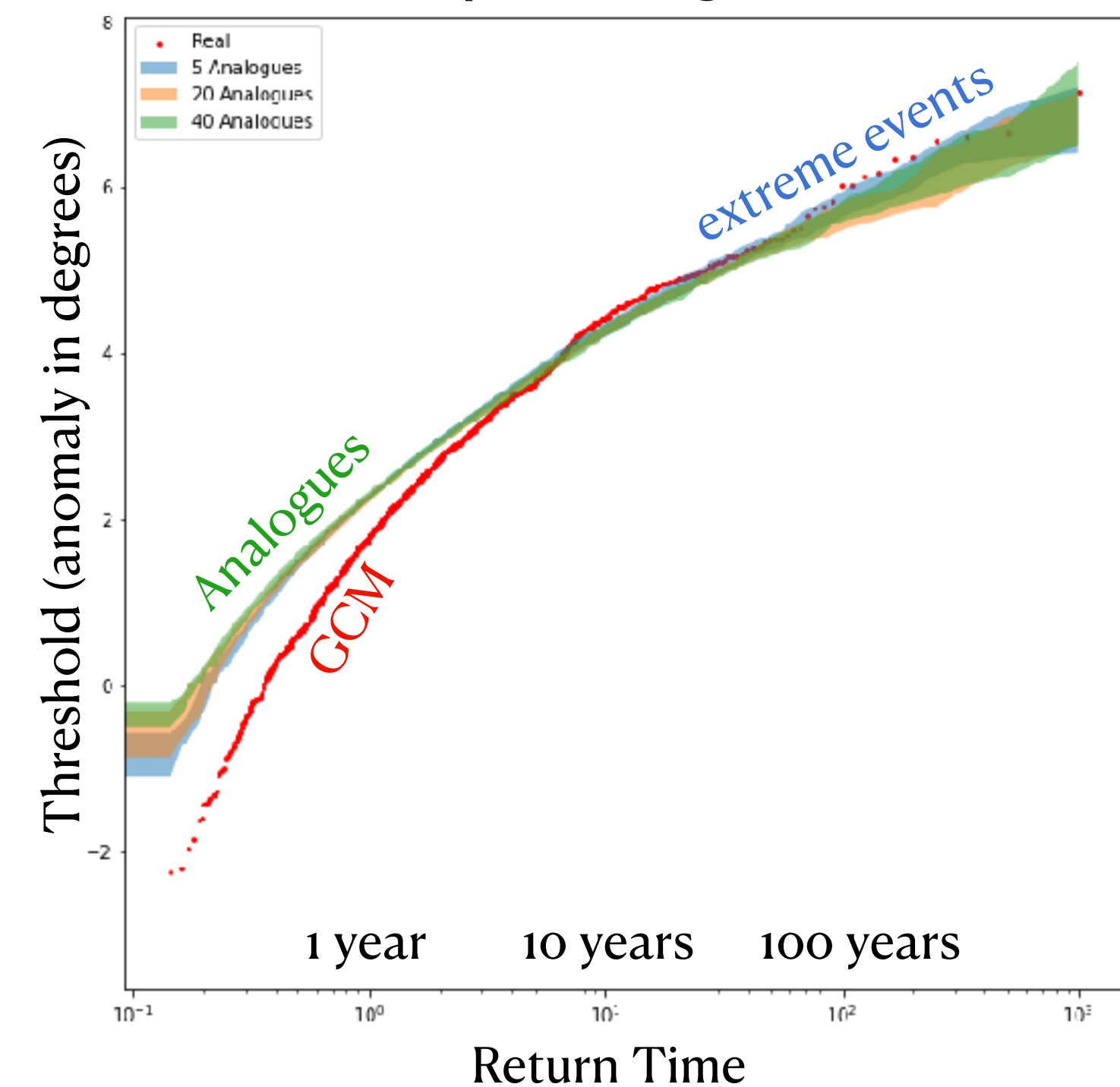
Analogues of geopotential

Occurrence of each day in a long trajectory



Analogues of surface temperature anomaly

Return Time plot Analogues vs GCM



Analogues of temperature anomaly

- Learning from geopotential rather than its anomalies -> systematic seasonal drift
- Using the method of analogues return time plot can be extended

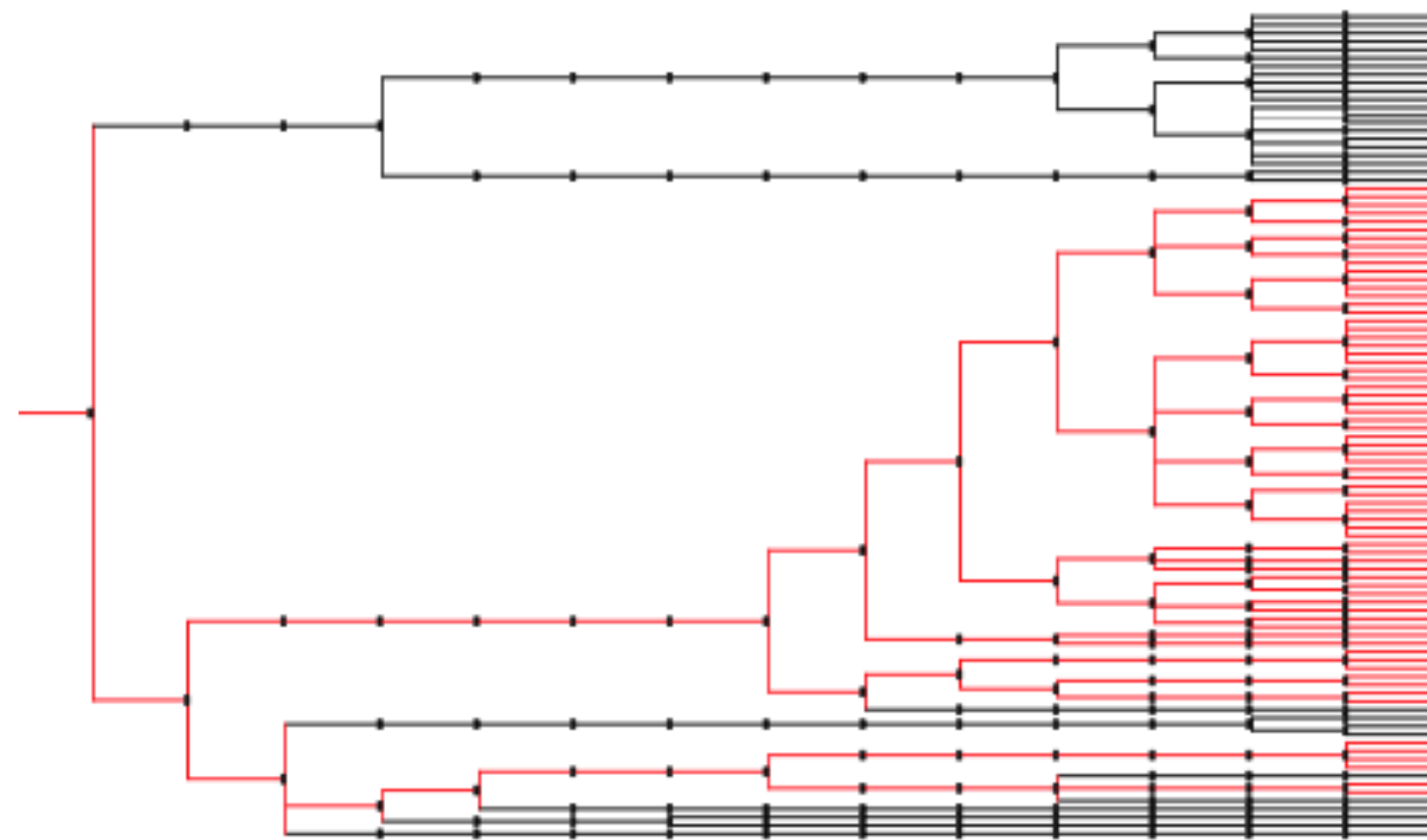
- Goal: construct more efficient **rare event algorithm** for extreme events in CESM
- Rare event algorithms require a score function
- **Committor Function** (conditional probability) can serve as optimal score
- **Committor Function** in transition path theory

$$q(\mathbf{x}) = \mathbb{P}(\tau_B(\mathbf{x}) < \tau_A(\mathbf{x})).$$

- In our case:

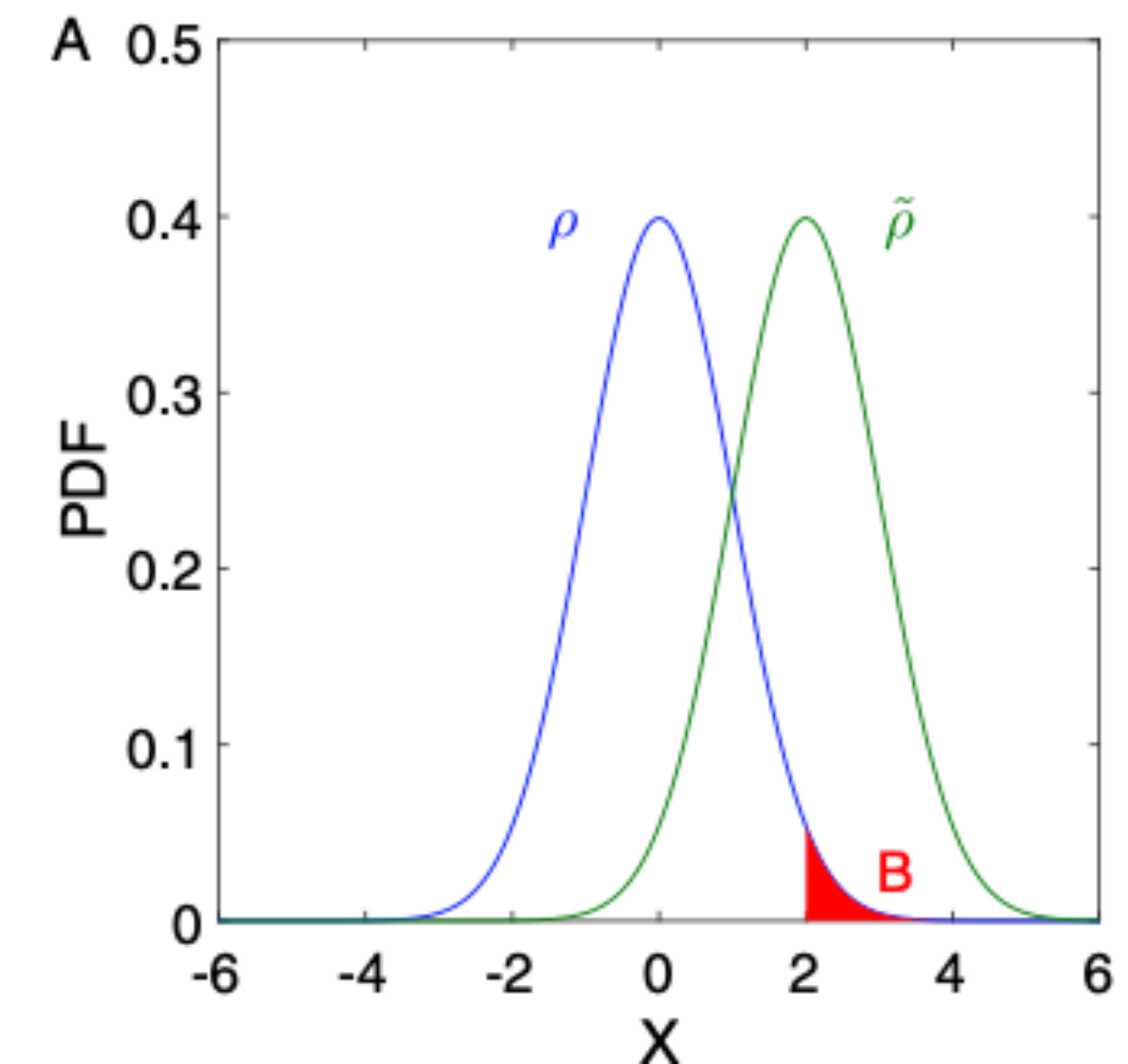
$$q_d = \mathbb{P}(T_{France}^{daily}(t_0 + 15) > a_d | X(t_0) = X_0),$$

$$q_a = \mathbb{P}(A(t_0)_{France} > a_a | X(t_0) = X_0),$$

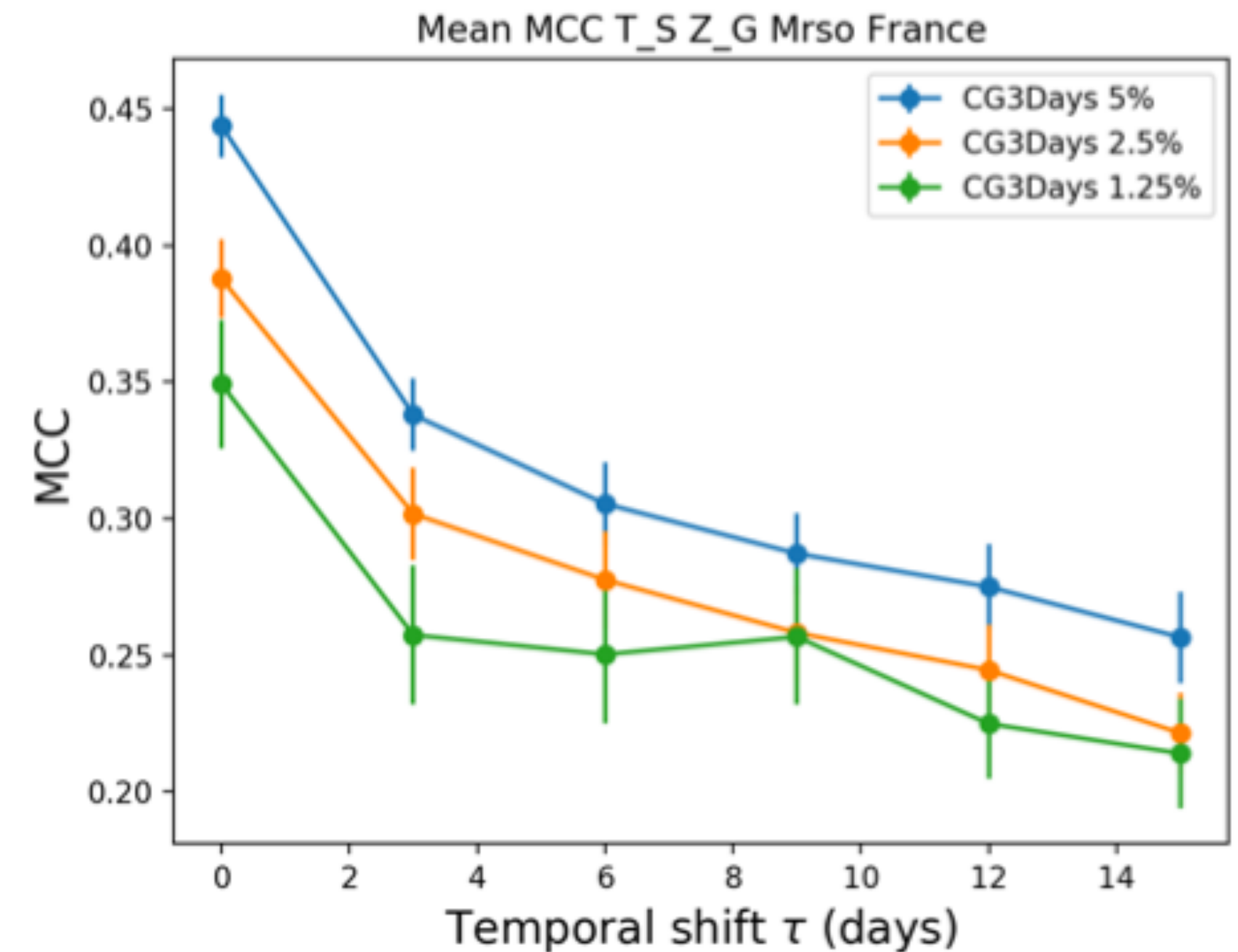
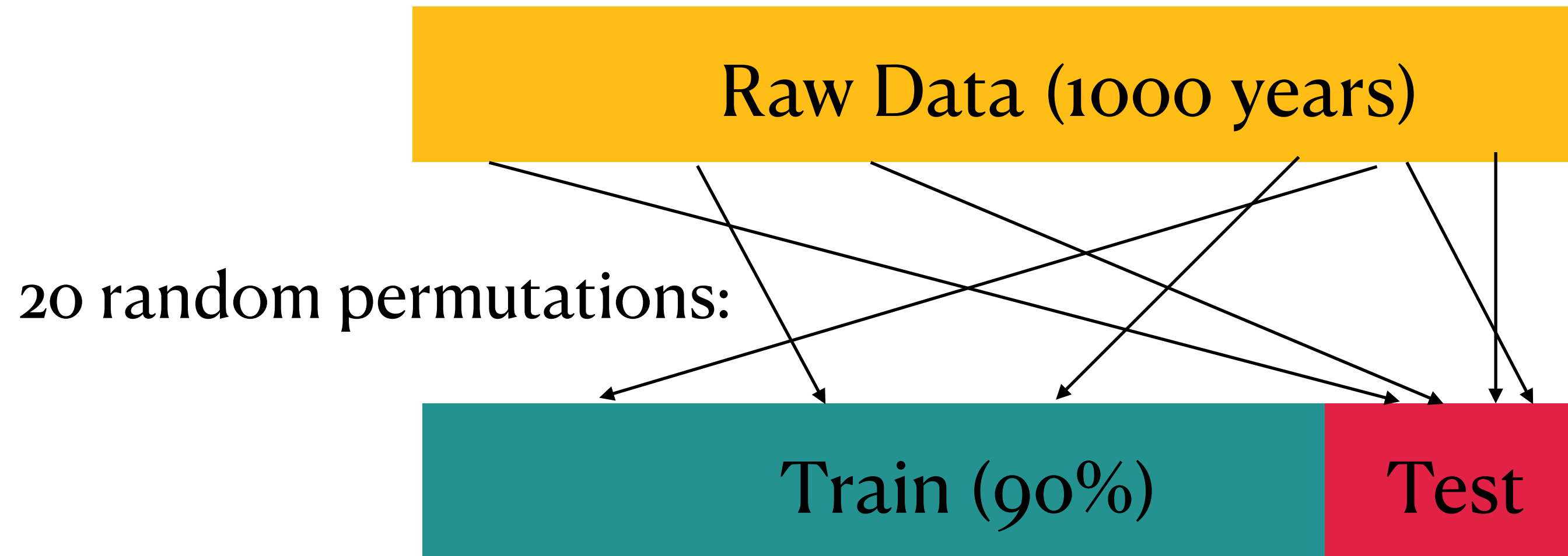


Tree of a rare event algorithm

Ragone et al 2020



We want to estimate the probability to be in the set B , for the model PDF

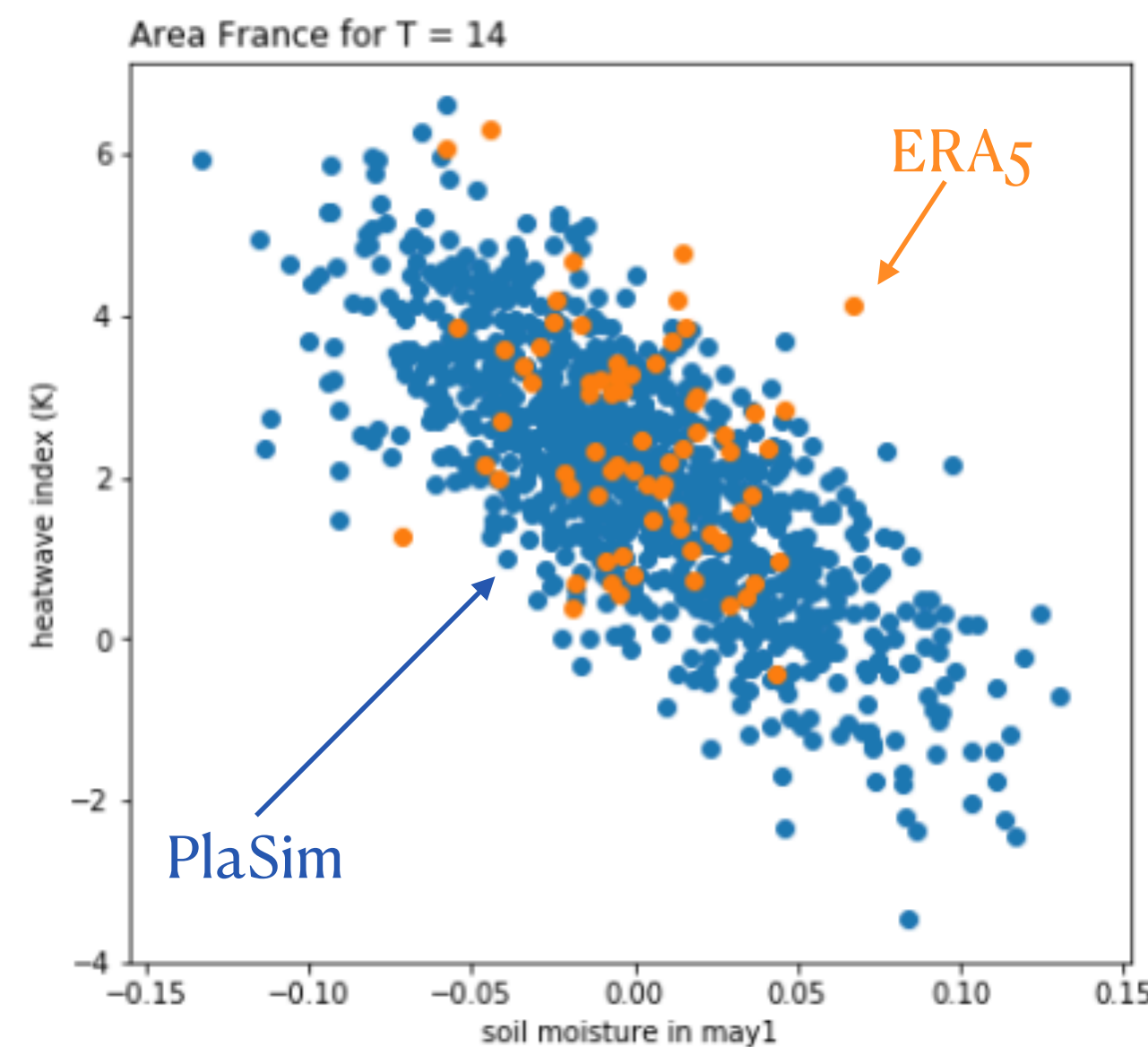


- Quality of binary classification of evenly weighted labels with **MCC**

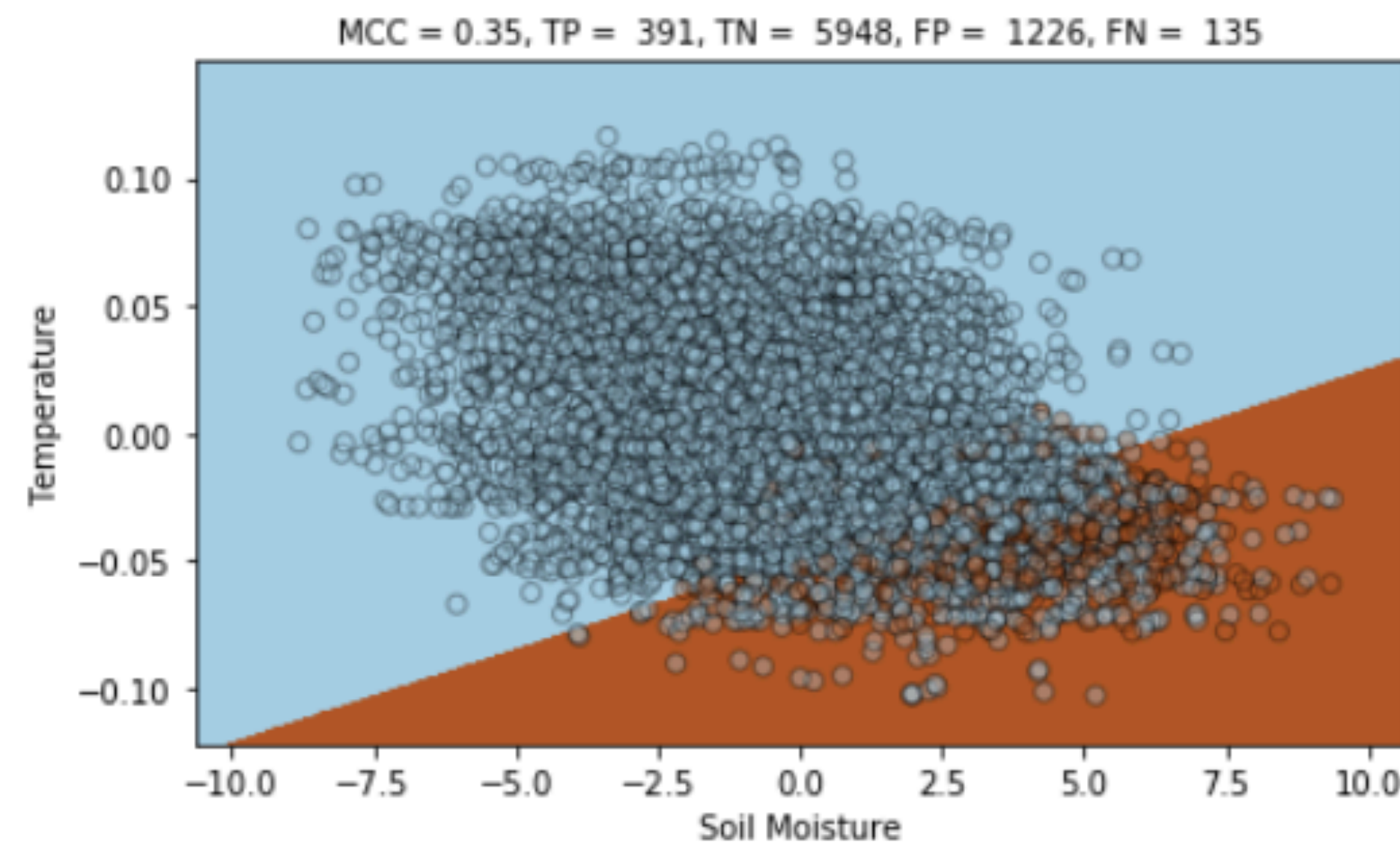
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

- We experiment with different areas in determining the distance to analogues D
- Smaller areas generally perform better (dimensionality reduction?)

- **Soil moisture** affects the sensible/latent heat flux (Bowen ratio)
- We attempt to incorporate it in our learning
- First, we perform **logistic regression** on temperature and soil moisture
- We find that these two scalars contain a lot of information that should be used



Summer 14 day extreme vs
soil water in France



Logistic regression of 14 day extreme vs
instantaneous temperature and soil water

- Compared between 1000 year **PlaSim**, **CESM** and 70 years of detrended **ERA5**
- We observe a **tripole** structure in the geopotential field in France and Scandinavia
- Used Gaussian process to model return time plots
- Used **analogue method** to compute the **committor** function
- Computed **Matthews correlation coefficient** for the analogue method
- Include soil moisture using the analogue method or deep CNN
- Obtain committor for **rare event algorithm** to better sample rare events

