Coupling rare event algorithms and machine learning to predict extreme heat waves

Freddy Bouchet (CNRS and ENS de Lyon)

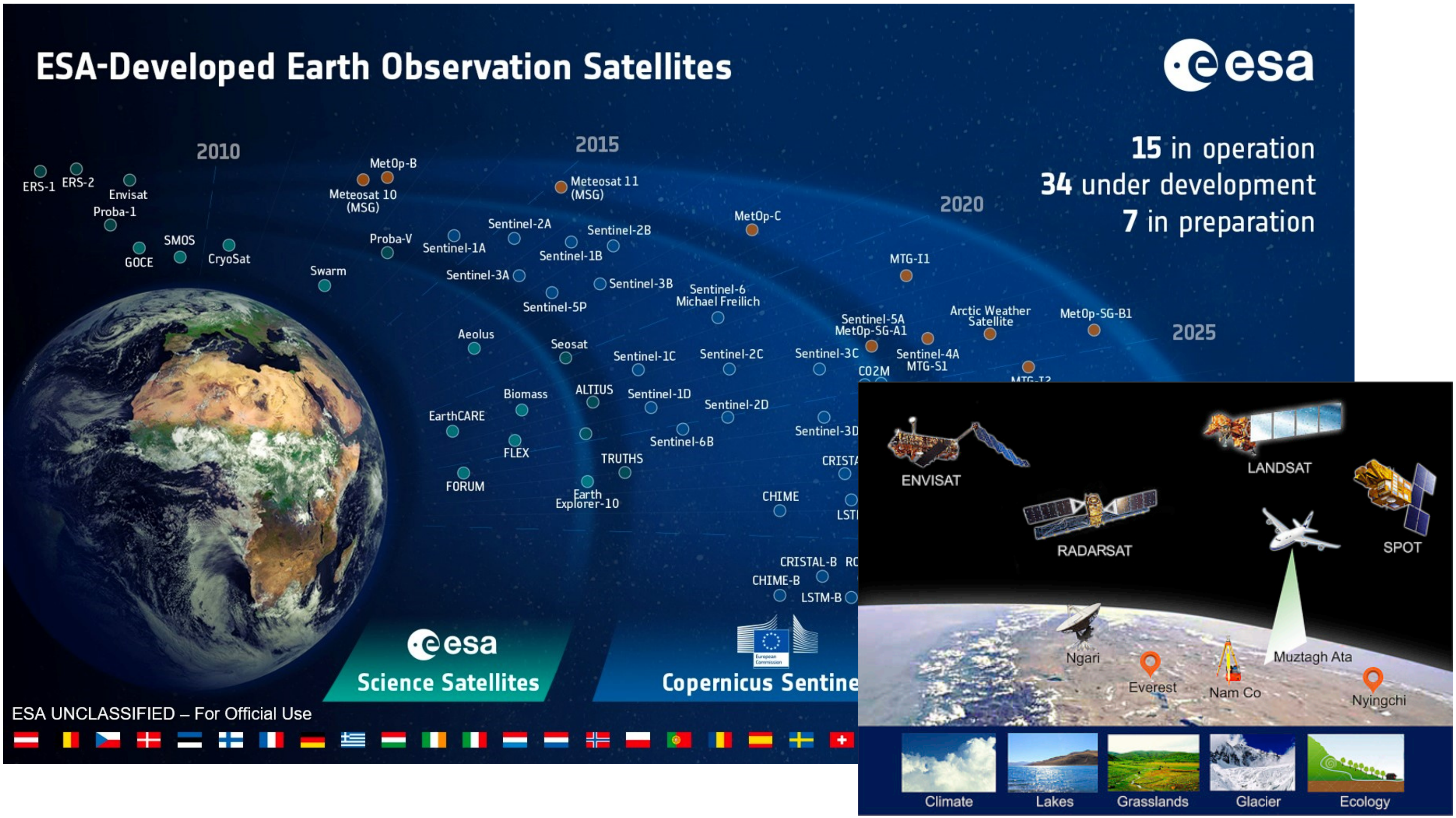
With:

a) Francesco Ragone and Jeroen Wouters (rare event algorithms in climate models)
b) Dario Lucente, George Miloshevich, and Francesco Ragone (extreme heat waves)
c) Patrice Abry, Pierre Borgnat, Valerian Jacques-Dumas, George Miloshevich, and Francesco Ragone (prediction of extreme heat waves with deep neural networks)
d) Dario Lucente, Joran Rolland, and Corentin Herbert (coupling rare event algorithms and machine learning)

CIRM, September 2021

I-a) Introduction to machine learning for climate dynamics and weather forecast

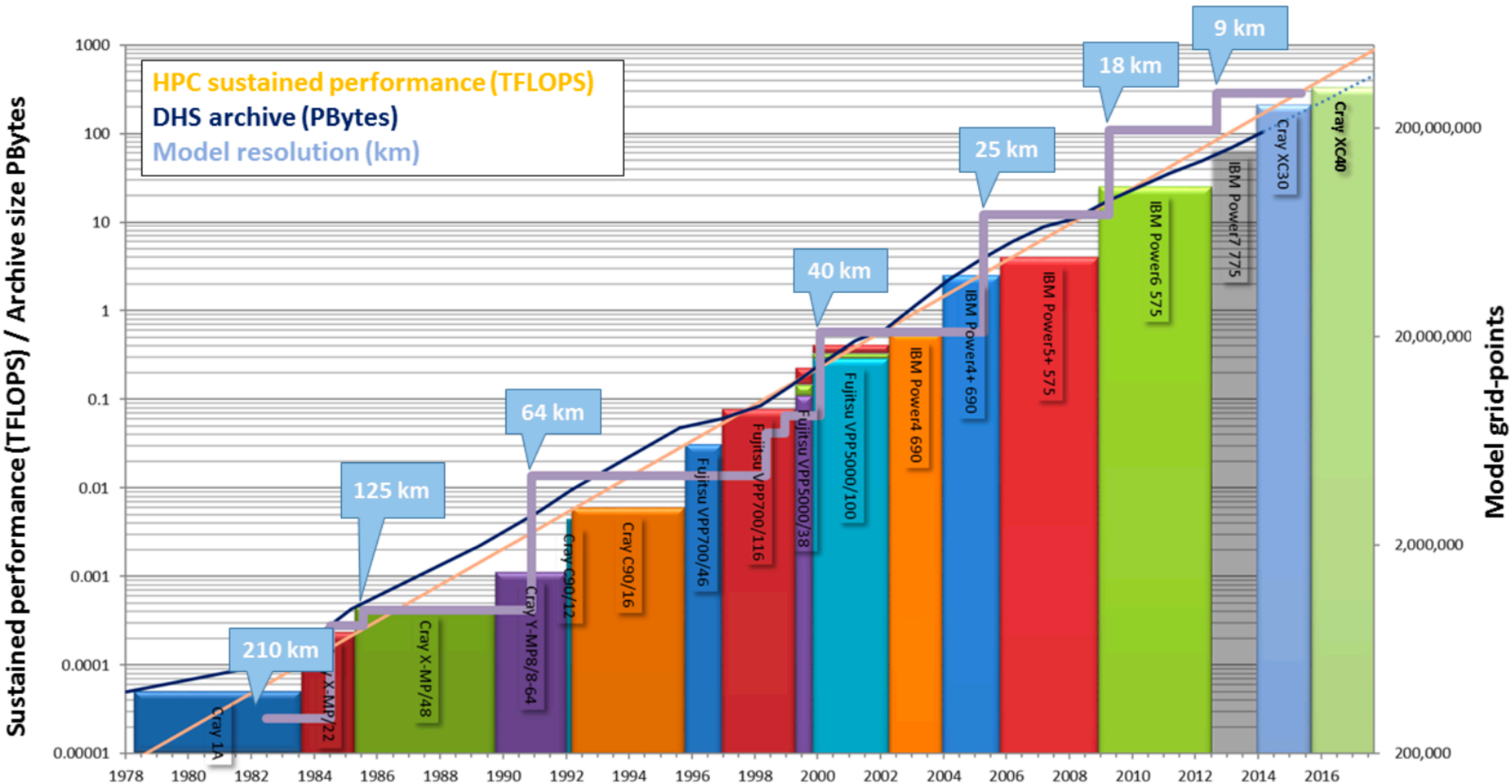
Earth observation satellites



Atmosphere, ocean and land are the most observed physical systems

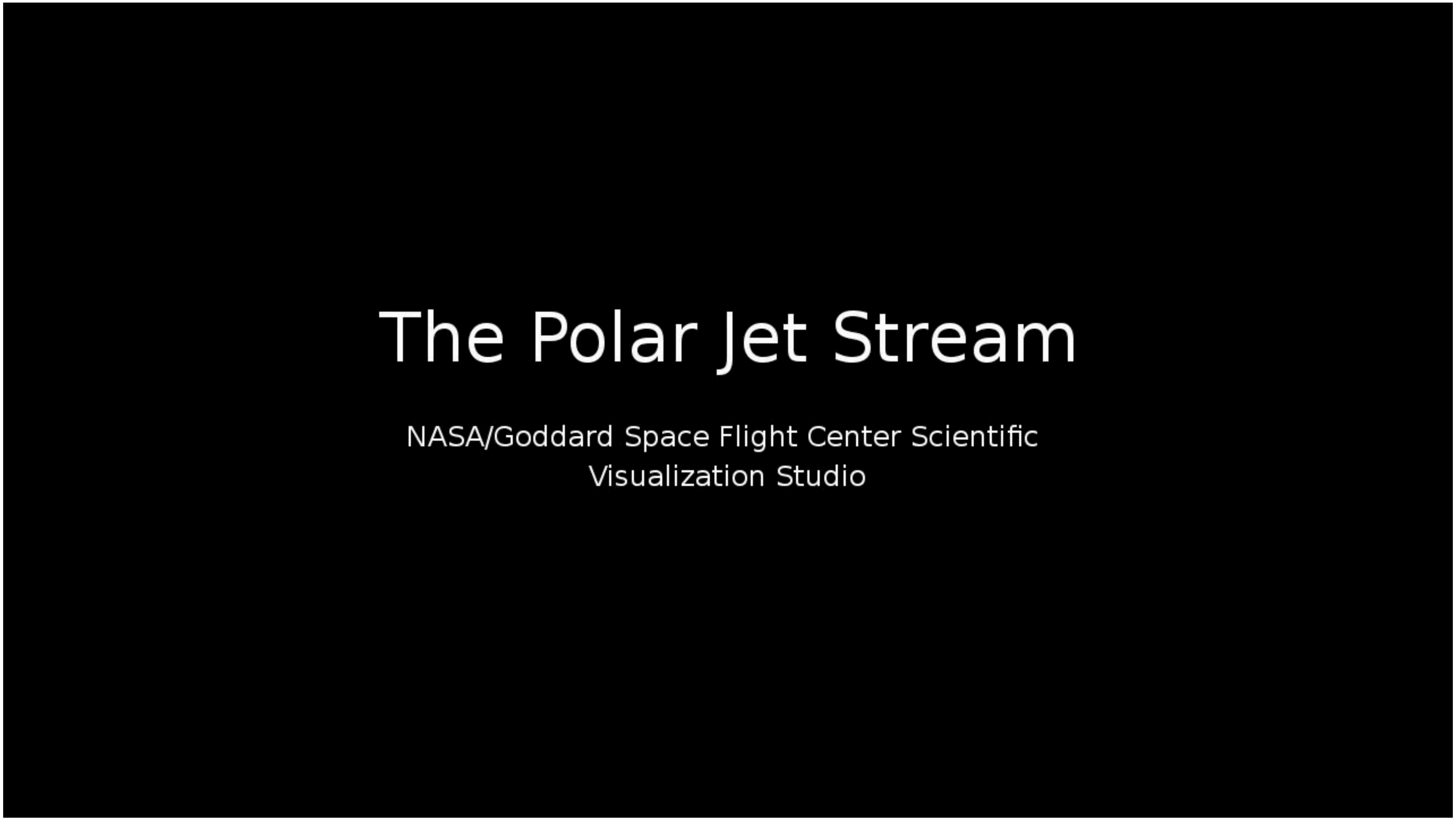
Example of ECMWF

ECMWF = European Center for Medium-Range Weather Forecasts



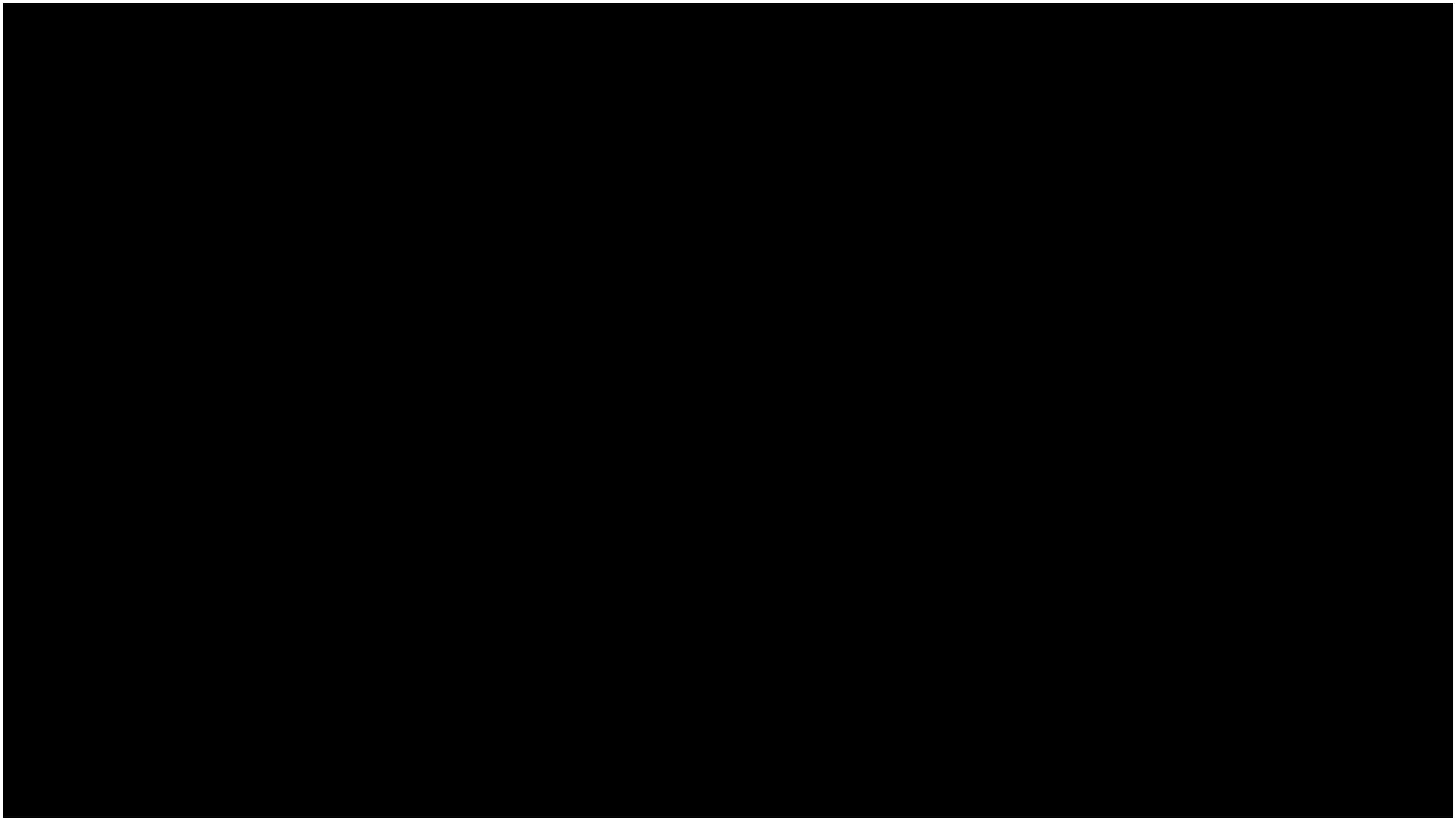
P. Bauer et al, 2020, the ECMWF scalability programme

Jet stream dynamics



Higher troposphere wind speed. (NASA/Goddard Space Flight Center Scientific Visualization Studio, MERRA reanalysis dataset)

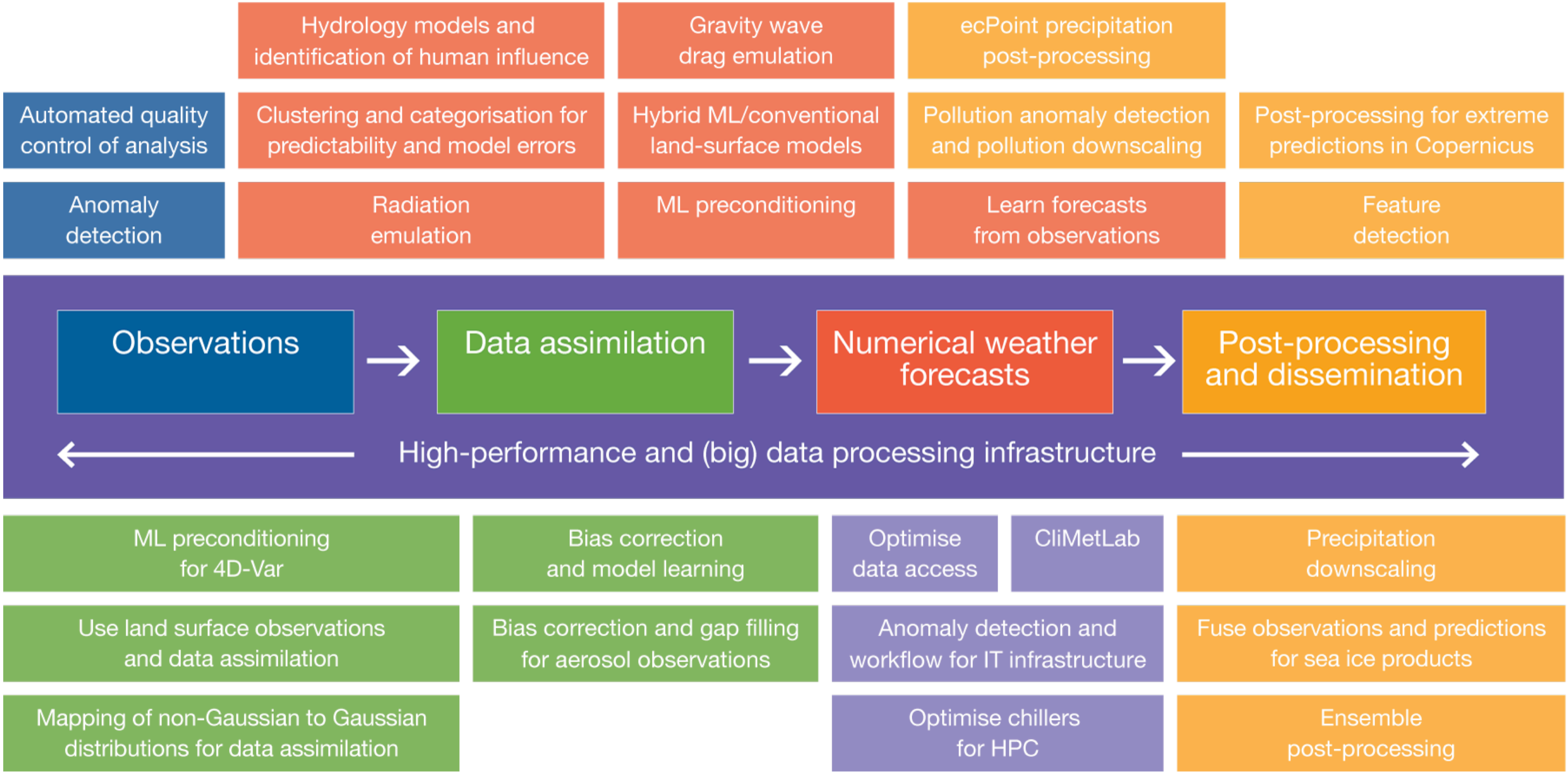
Ocean dynamics



Ocean surface flows.

(Perpetual ocean, NASA Goddard visualization, data from ECCO2 MIT/JPL project)

Many application areas for machine learning across ECMWF



Machine learning at ECMWF: A roadmap for the next 10 years, P. Dueben et al, 2021

Machine learning and deep neural networks enter in many different ways for both weather forecast and climate dynamics.

High level objectives of the roadmap for machine learning at ECMWF

Objective 1
Explore machine learning applications across the weather and climate prediction workflow and apply them to improve model efficiency and prediction quality.

Objective 2
Expand software and hardware infrastructure for machine learning.

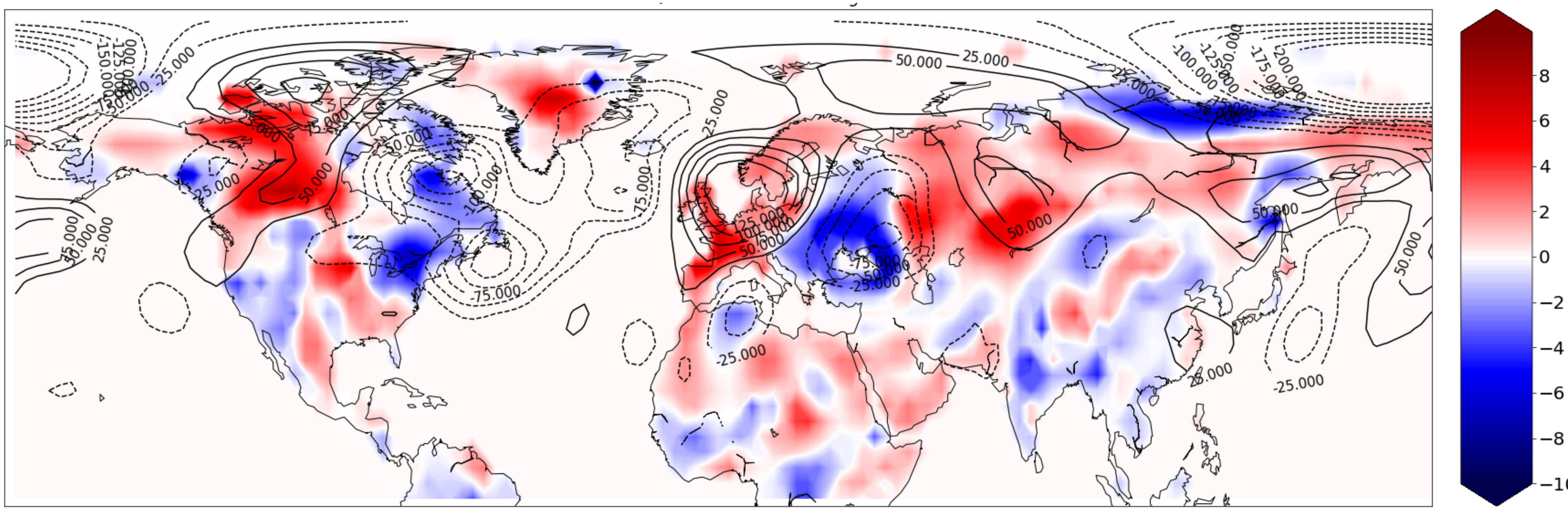
Objective 3
Foster collaborations between domain and machine learning experts with the vision of merging the two communities.

Objective 4
Develop customised machine learning solutions for Earth system sciences that can be applied to various applications and at scale on current and future supercomputing infrastructure.

Objective 5
Train staff and Member and Co-operating State users and organise scientific meetings and workshops.

Machine learning at ECMWF: A roadmap for the next 10 years, P. Dueben et al, 2021

Predicting heat waves with a deep neural network



Surface temperature (T_s, colors) and 500 hPa geopotential height (Z_g, lines) anomalies

- **Plasim and CESM climate models.**
- We use summer (JJA) data: 8 maps/day, 90 days/year, 1000 year = 720 000 maps.
- For Plasim data, each field has a resolution 64×128, restricted to 25×128 above 30^o North.

Heat wave definition

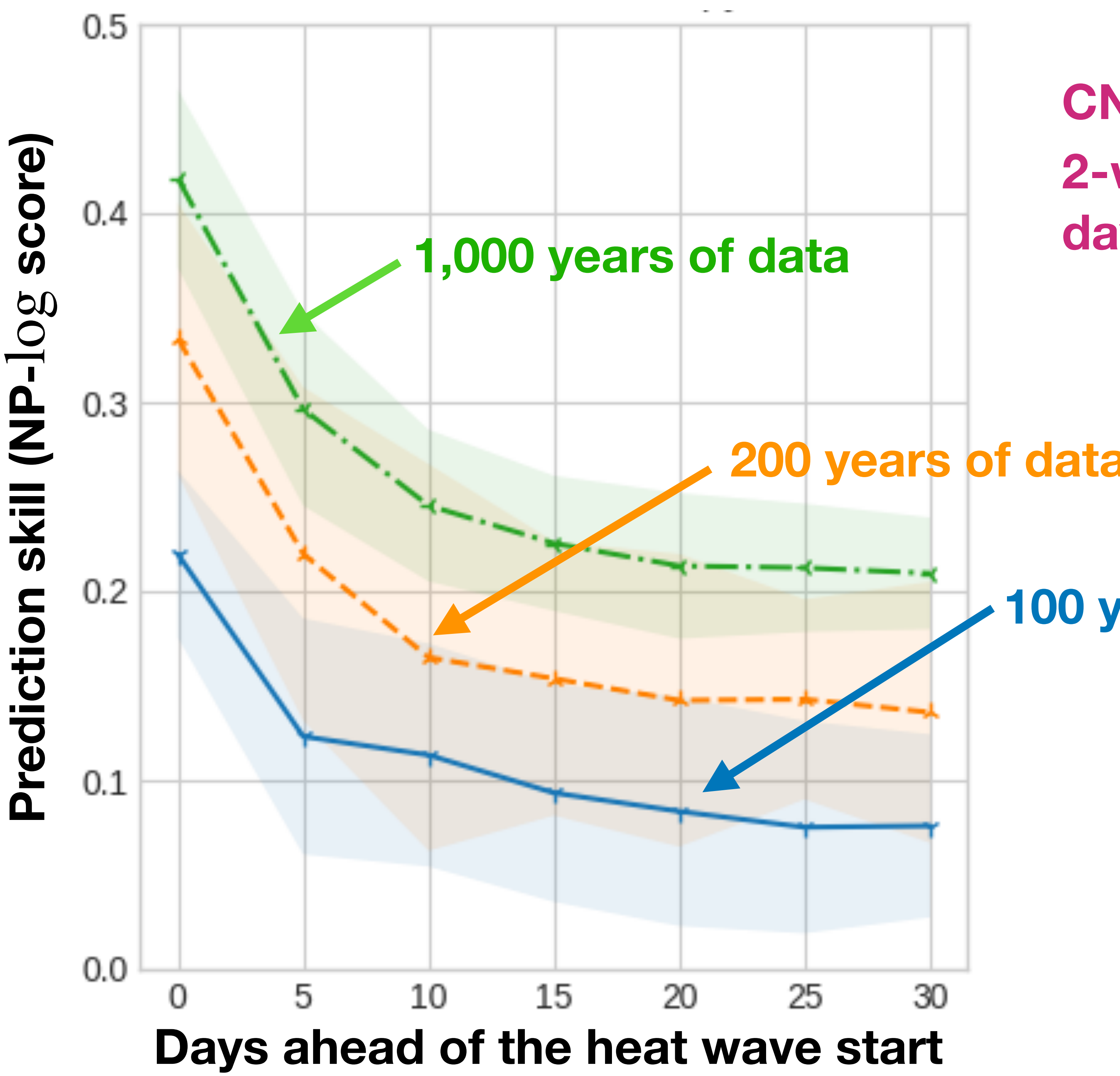
- $X(t) = T_s$ field at time t, or $X(t) = (T_s, Z_g)$ fields at time t.
- $Y(t)$: **time and space averaged surface temperature anomaly within τ days:**

$$Y(t) = \frac{1}{D}\int_{t+\tau}^{t+\tau+D} \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} T_s(\vec{r}, u)\ \mathrm{d}\vec{r}\ \mathrm{d}u,$$

$$\text{and } Z(t) = 1 \text{ if } Y(t) > a, \text{ and } Z(t) = 0 \text{ otherwise}$$

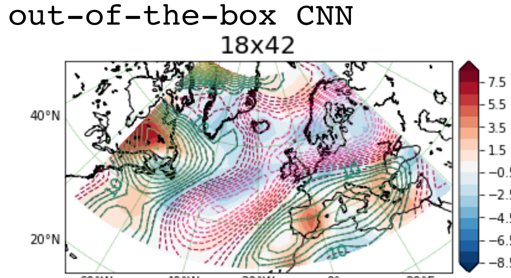
- $Z(t) \in \{0,1\}$. A heat wave occurs if $Z = 1$.
- We have a classification problem for the data (X, Z). We want to learn the probability $q(x)$ that $Z = 1$ given that $X = x$ (committor function).
- 5% most extreme events: $a = a_5 = 3.08\ \mathrm{K}$. 2.5% most extreme events: $a = a_{2.5} = 3.7\ \mathrm{K}$. 1.25% most extreme events: $a = a_{1.25} = 4.23\ \mathrm{K}$.

Machine learning for climate applications: a regime of lack of data



CNN classification of 2-week heatwaves τ days ahead of time

a_5 . The input consists of (t2m,zg500,mrso) with resolution (18,42,3). filtering applied to mrso (no part of the map is set to zero. We were using out-of-the-box CNN



The observation dataset is way too small for good machine learning predictions

Machine learning, climate, and weather forecast models

- The Earth (atmosphere, ocean, land, etc.) is the most observed system with an exponentially growing dataset.

- Those observations are coupled to physical models through data assimilation techniques (a very old and very smart machine learning scheme for physically based data integration).

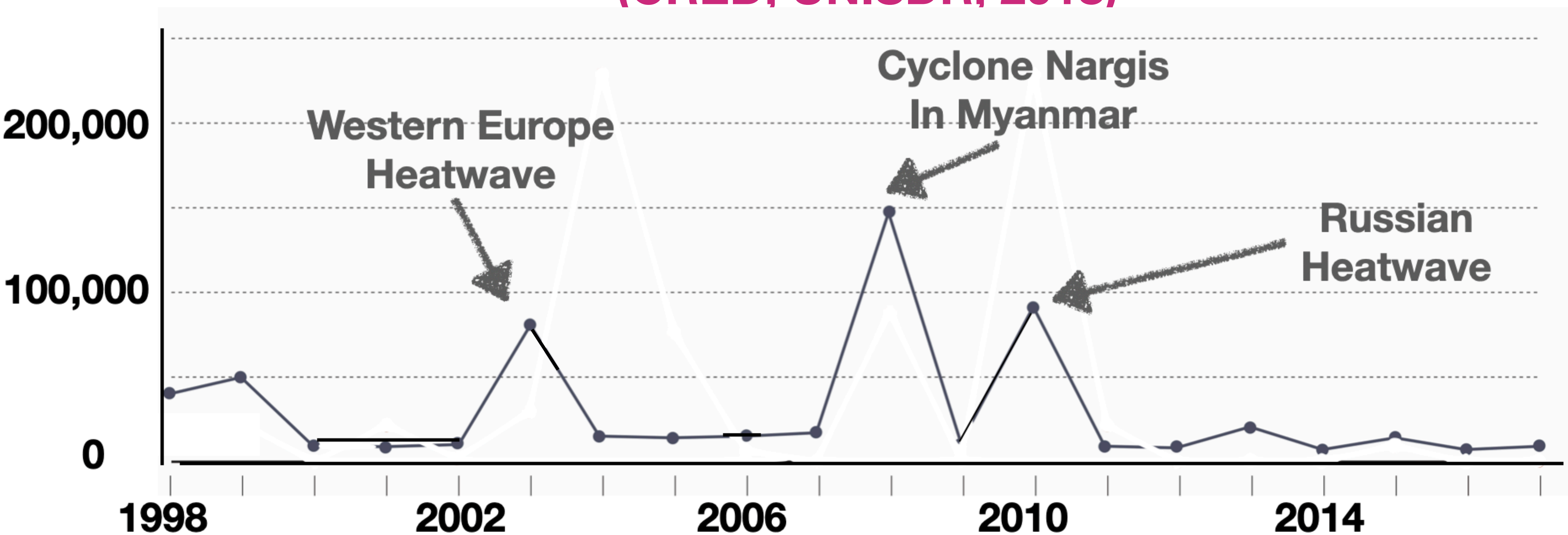
- Machine learning and deep neural networks enter in many different ways for both weather forecast and climate dynamics.

- For many (not all) of these problems, machine learning should be performed in a regime of lack of data. This is key for understanding.

I-b) Introduction to climate extreme events and rare event algorithms

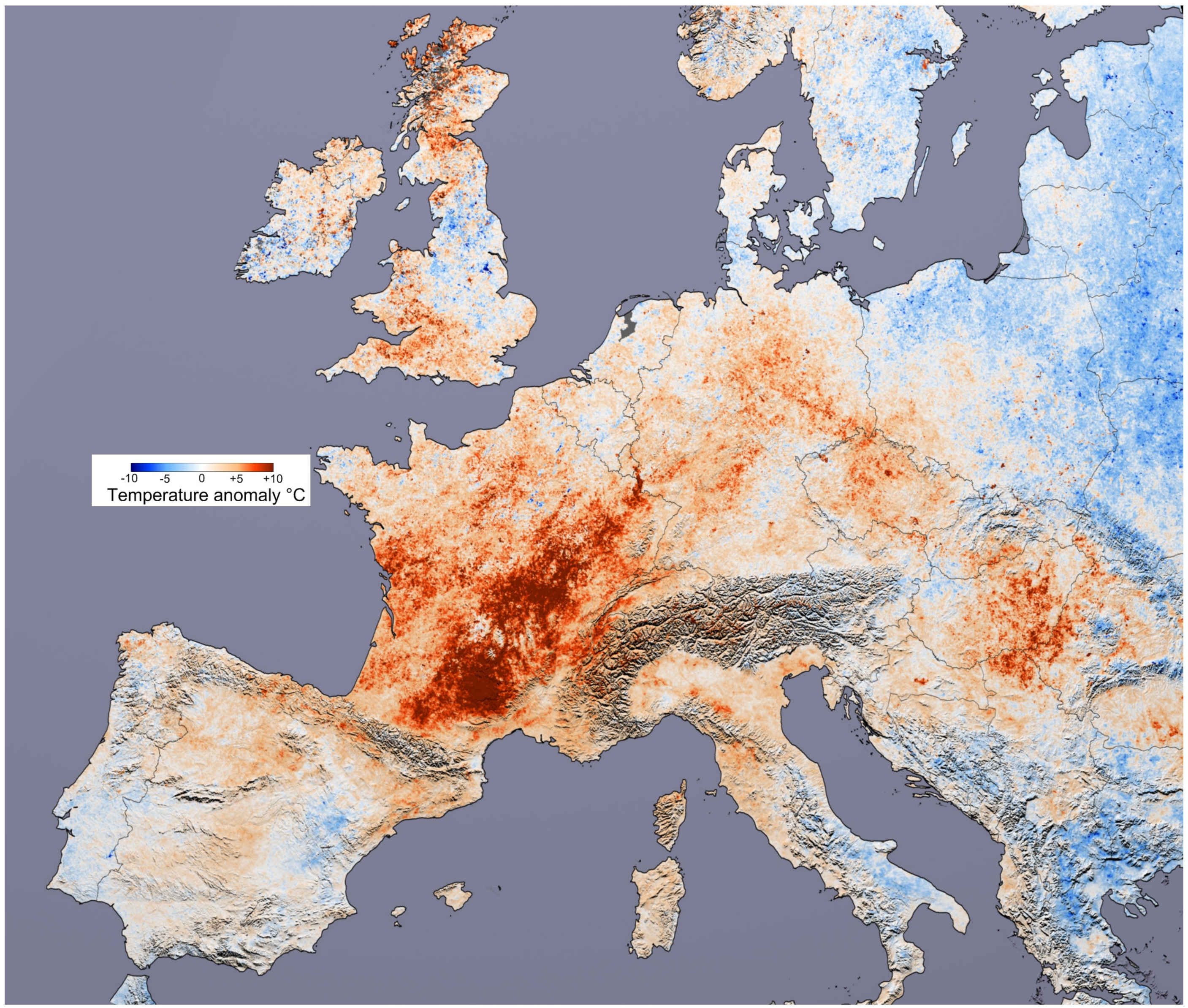
The few most extreme climate events have more impact than all the others

Annual deaths by major climate related disaster
(CRED, UNISDR, 2018)



We need to study extremely rare events. This is a serious scientific challenge.

What is the probability (return time) of the 2003 Europe heatwave ?



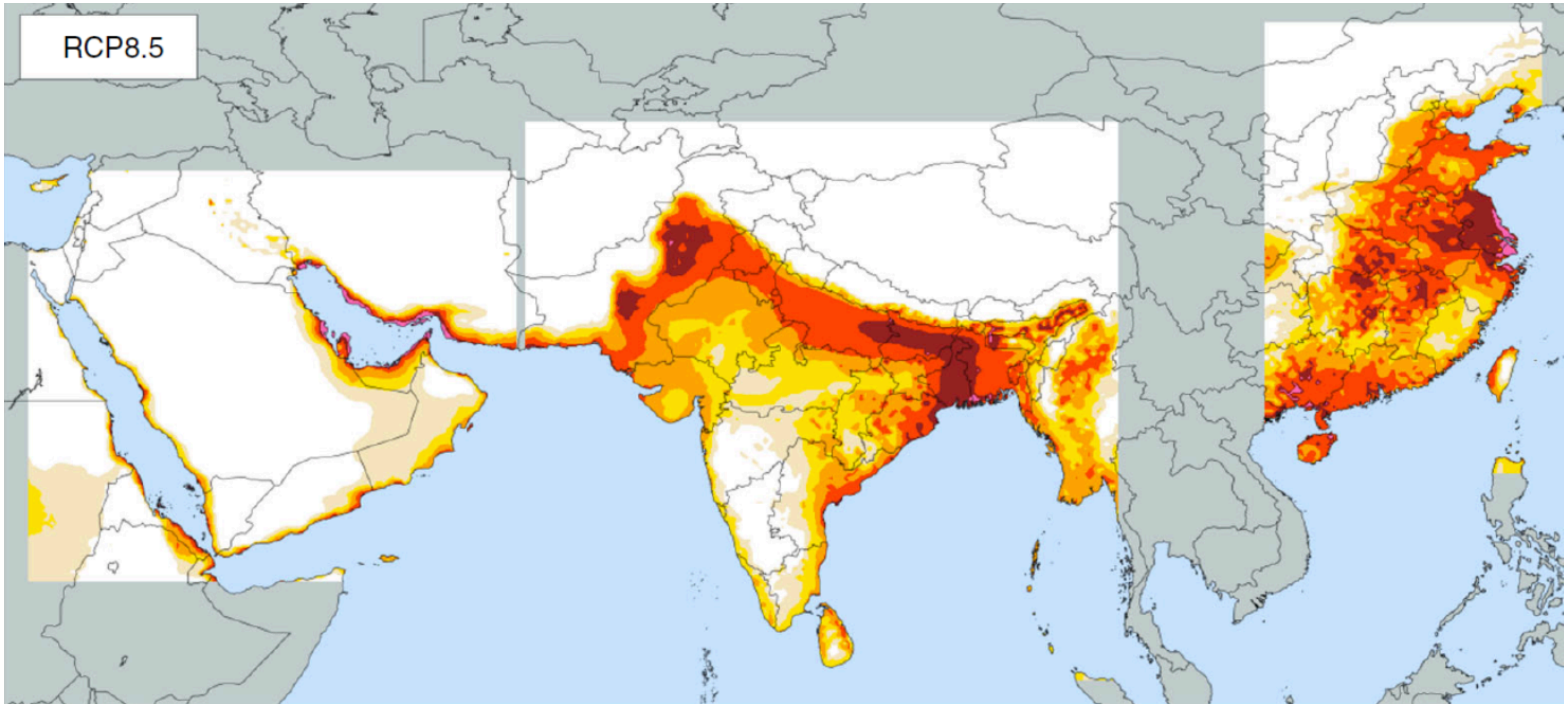
July 20 2003-August 20 2003 land surface temperature minus the average for the same period for years 2001, 2002 and 2004 (TERRA MODIS).

Why are return times so hard to estimate?

i) lack of observation data, ii) model biases,

iii) because of rareness, gathering good model statistics is too costly.

Potential impacts of global warming and extreme events



Maximal wet bulb temperature (red color =31-32°C), in 2070, with the RCP8.5 scenario.

(Kang, Elfatih and Eltahir, 2018)

Hundreds of thousands of people leave now in areas of the world that will become inhabitable before the end of the century if we do not halt global warmings. Thinking of these phenomena in a classical economic framework does not make any sense.

Three key problems in the study of climate extreme events

- The historical records are way too short to make any meaningful predictions for the rarest events (those that matter the most).

- Climate models are wonderful tools, but they have biases. The more precise, the more computationally costly.

- Because they are too rare, the most extreme events cannot be computed using direct numerical simulations (the needed computing times are often unfeasible).

The practical questions: How to sample the probability and dynamics of rare events in complex models? How to build effective models which are relevant for estimating the probability of rare events?

How to study a 10 000 year heat wave with a 200 year simulation ?

- **Because they are too rare, extreme events cannot be computed using direct numerical simulations** (the needed computing times are often unfeasible).

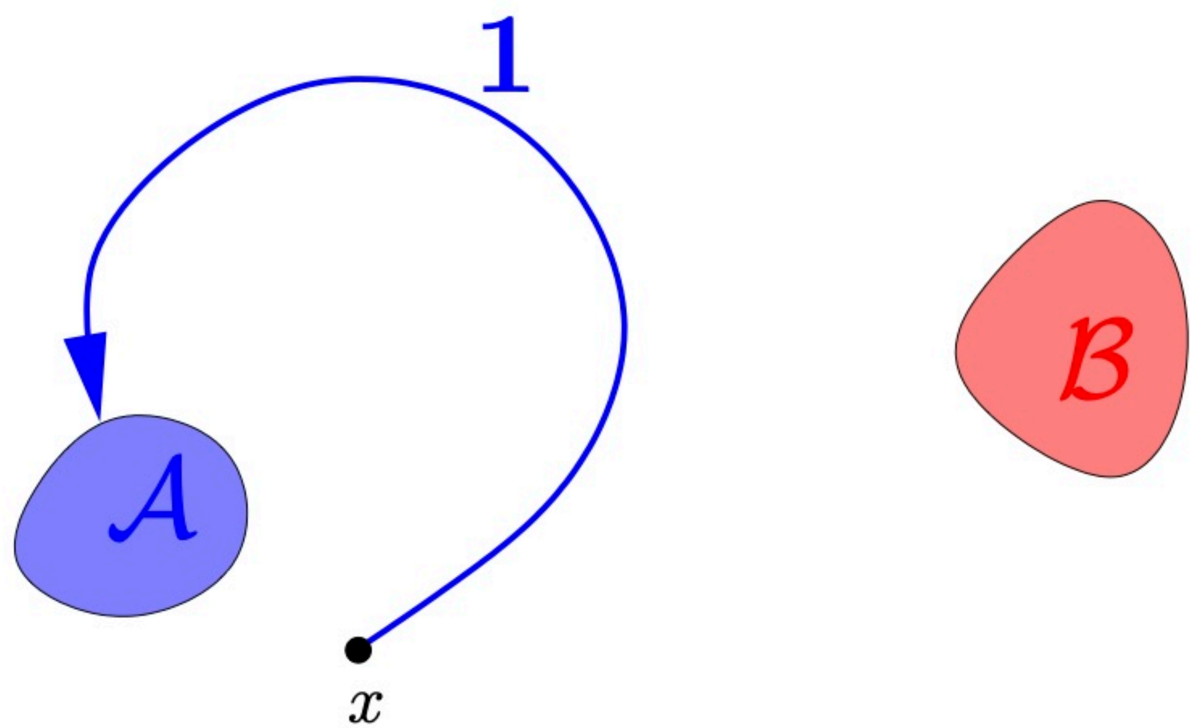
- **Rare event algorithms:** Kahn and Harris (1953).

- **Statistical mechanics:** diffusion Monte-Carlo, Wang Landau algorithms, go with the winners, etc.

- **Applied Mathematics:** Chandler, Vanden-Eijnden, Schuss, Del Moral, Dupuis, Lelièvre, Guyader, etc.

- **For turbulence and climate applications:** J. Weare and D. Abbot, R. Grauer and T. Grafke, E. Vanden-Eijnden, Lyon group, etc.

How to compute extremely rare trajectories ?

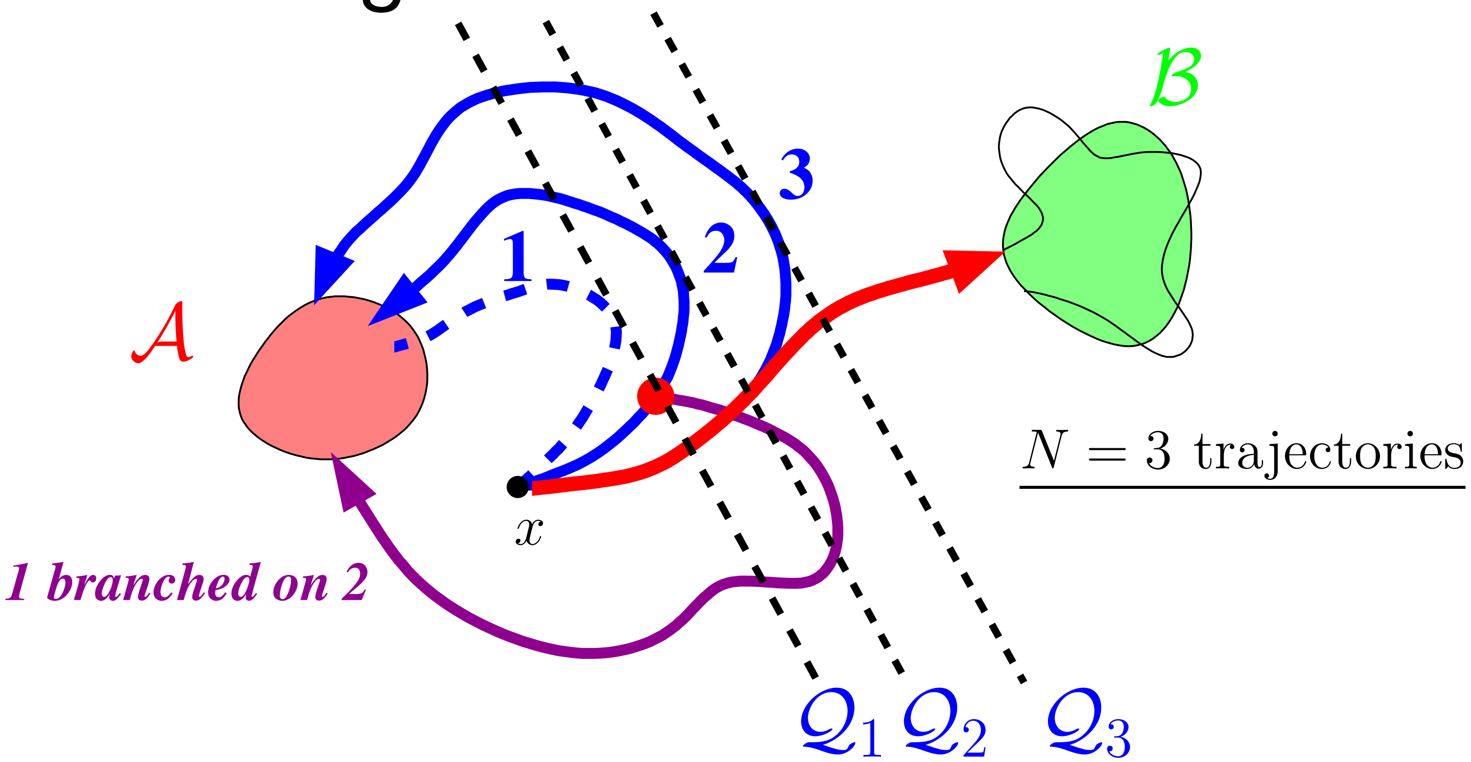


Aim: compute **extremely rare trajectories** from the point x to the rare event set $\mathscr{B}$.

Most of the times, trajectories that start from x end in $\mathscr{A}$.
The probability to reach $\mathscr{B}$ may be 10^{-3} or 10^{-20}.

The Adaptive Multilevel Splitting (AMS) rare event algorithm

Strategy: ensemble computation, selection, pruning and cloning.



Probability estimate:

$\hat{p} = (1 - 1/N)^K$,

where N is the clone number and K is the iteration number.

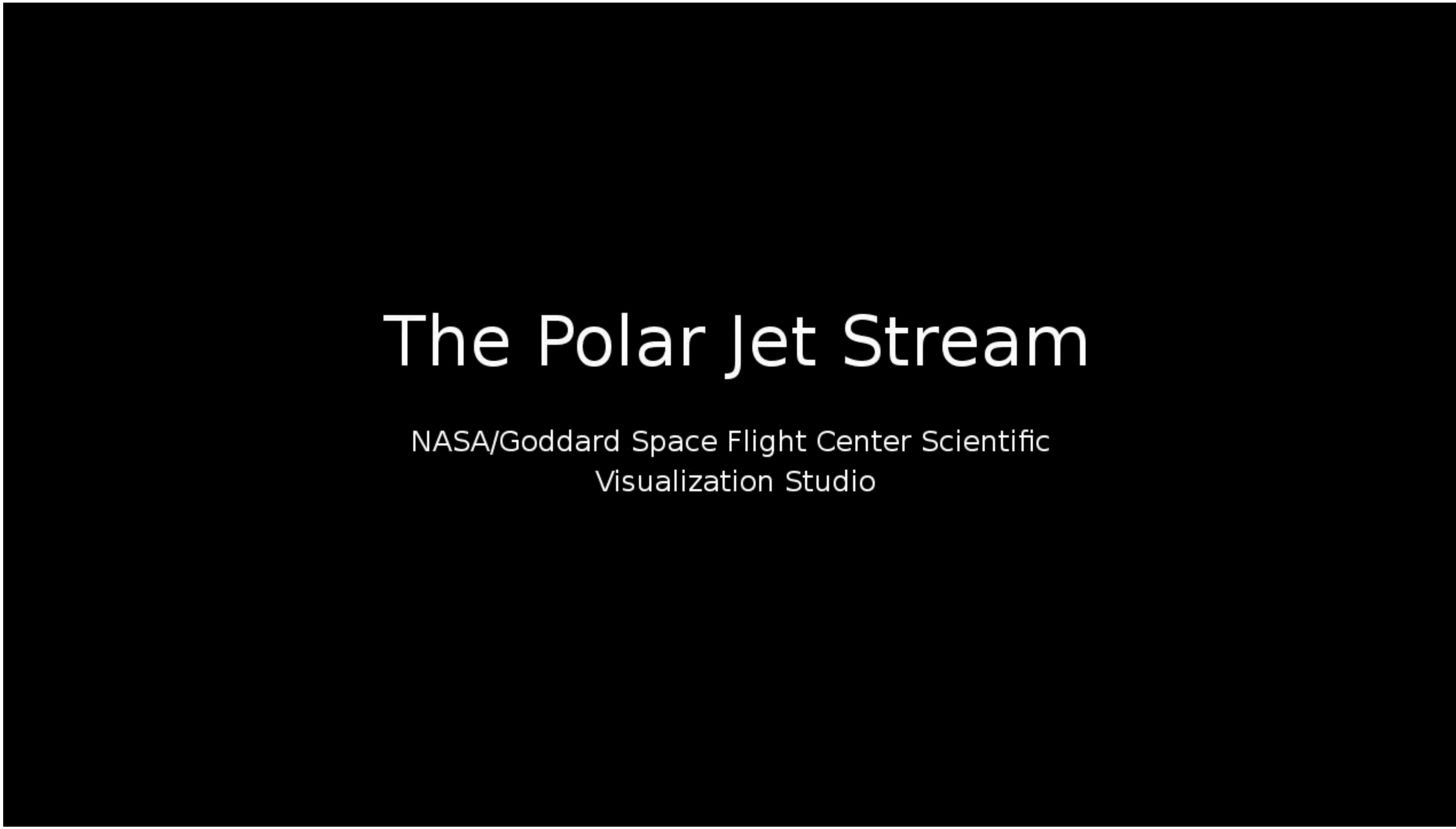
Cérou, Guyader (2007). Cérou, Guyader, Lelièvre, and Pommier (2011).
PDEs: Rolland, Bouchet et Simonnet (2016) - **TAMS:** Lestang et al (2018)
Atmosphere turbulent jets: Rolland, Bouchet et Simonnet (2019 and 2021).

II-a) Rare event algorithms to study extreme heat waves with climate models

Francesco Ragone
RMI, Bruxelles, Belgium

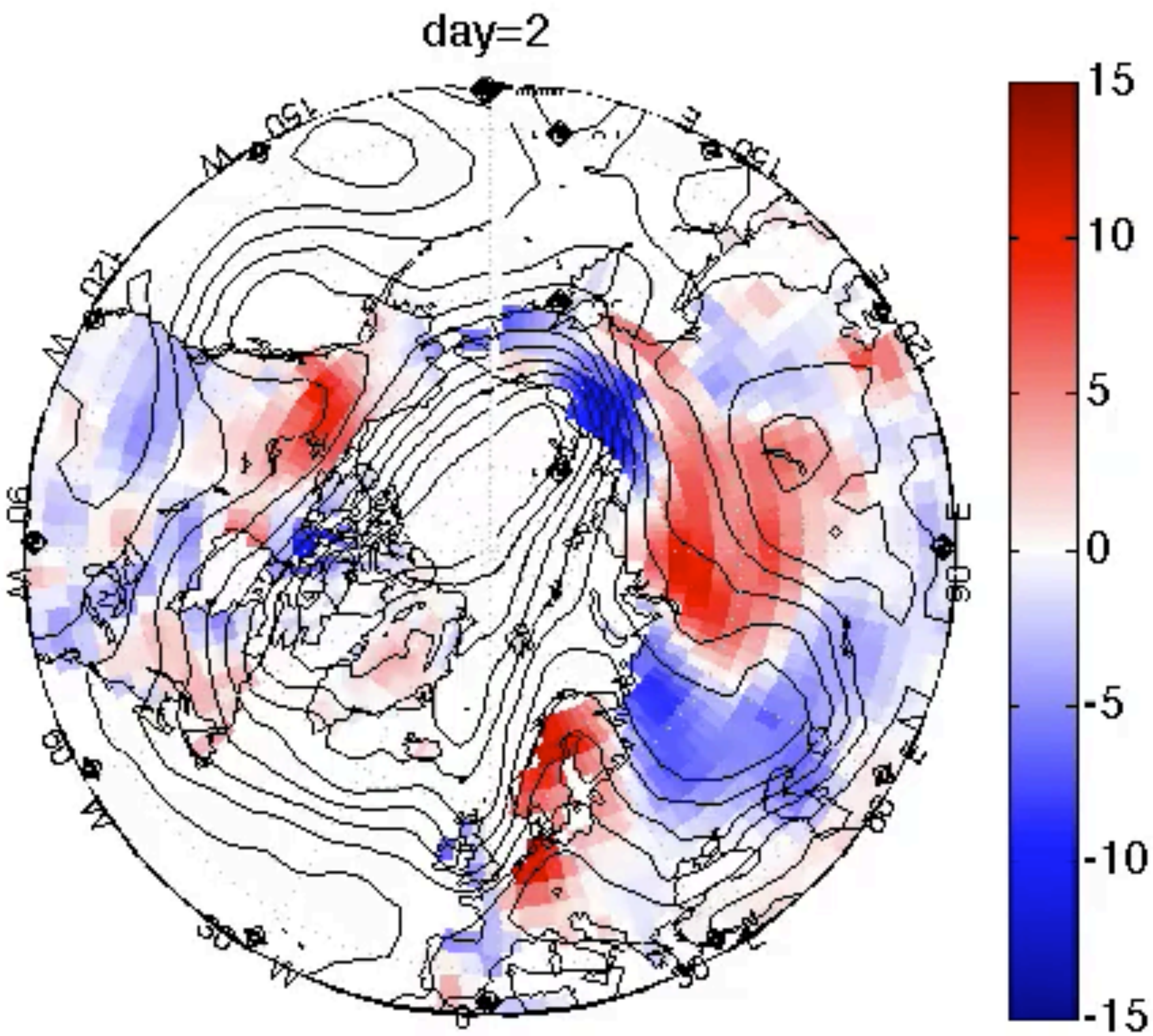
Jeroen Wouters
University of Reading, UK

Jet stream dynamics



Higher troposphere wind speed. (NASA/Goddard Space Flight Center Scientific Visualization Studio, MERRA reanalysis dataset)

General Circulation Model



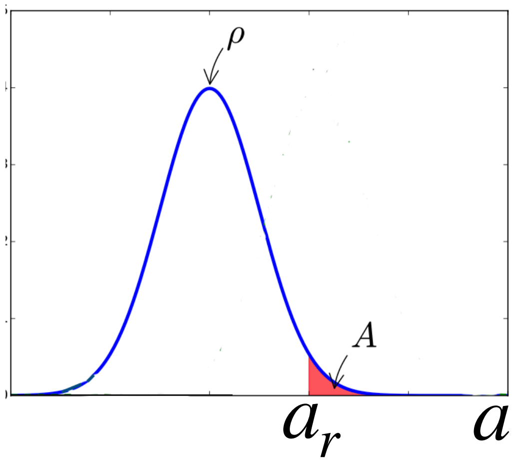
Surface temperature (T_s, colors) and 500 hPa geopotential height (Z_g, lines) anomalies

- Plasim and CESM climate models.
- Global. Coupled atmosphere/land/ocean/vegetation.

Long lasting summer heat waves

We will study extremes of the time averaged temperature:

$$a = \frac{1}{T}\int_0^T \mathrm{d}t \int_{\mathscr{D}} \mathrm{d}\mathbf{r}\ \mathrm{Temp}(\mathbf{r}, t)$$



- $T =$ one week, a few weeks, a month, or a season.
- $\mathscr{D}$ = Scandinavia, Europe, France, Alberta, Russia, …
- **Climate models** (CESM or PLASIM) or reanalysis datasets.

The Giardina—Kurchan (Del-Moral —Garnier) rare event algorithm

- With $A[X](t) = \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} \mathrm{d}\mathbf{r}\ \mathrm{T_S}(\mathbf{r},t)$, we sample the tilted path-distribution

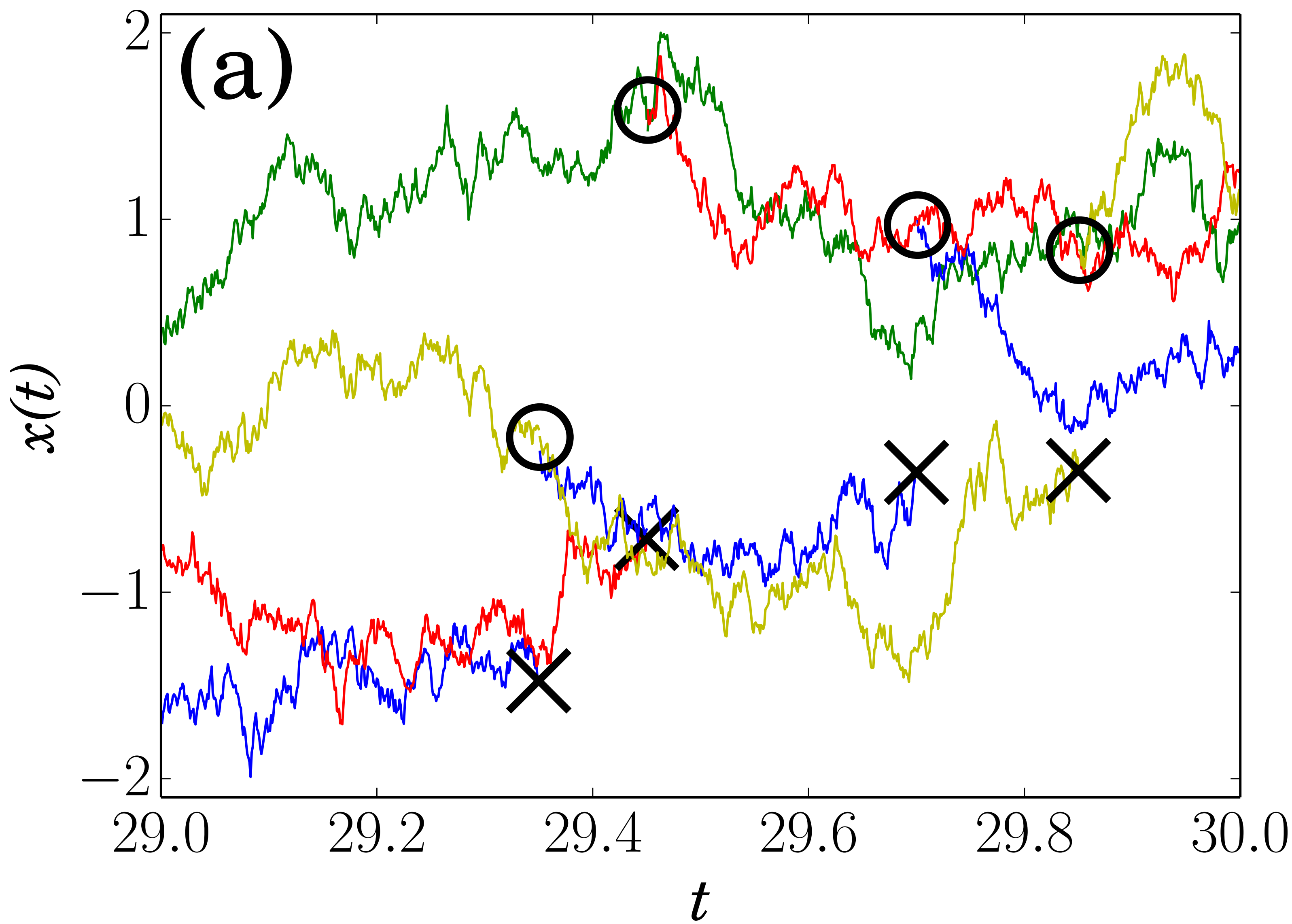
$$\tilde{P}_k\left(\left\{X(t)\right\}_{0\le t\le T}\right) = \frac{1}{\exp(T\lambda(k))} P_0\left(\left\{X(t)\right\}_{0\le t\le T}\right) \exp\left[k\int_0^T A[X](t)\,\mathrm{d}t\right].$$

- We simulate an ensemble of N trajectories $x_n(t)$. At each time step $t_i = i\tau$, each trajectory can be killed or cloned according to the weights

$$\frac{1}{W_i(k)}\exp\left(k\int_{t_{i-1}}^{t_i} A[x_n](t)\,\mathrm{d}t\right) \quad \text{with} \quad W_i(k) = \sum_{n=1}^{N}\exp\left(k\int_{t_{i-1}}^{t_i} A[x_n](t)\,\mathrm{d}t\right).$$

- Algorithm: Giardina et al. 2006. Mathematical aspects: Del Moral's book (2004).

Genealogical algorithm: selecting, killing and cloning trajectories

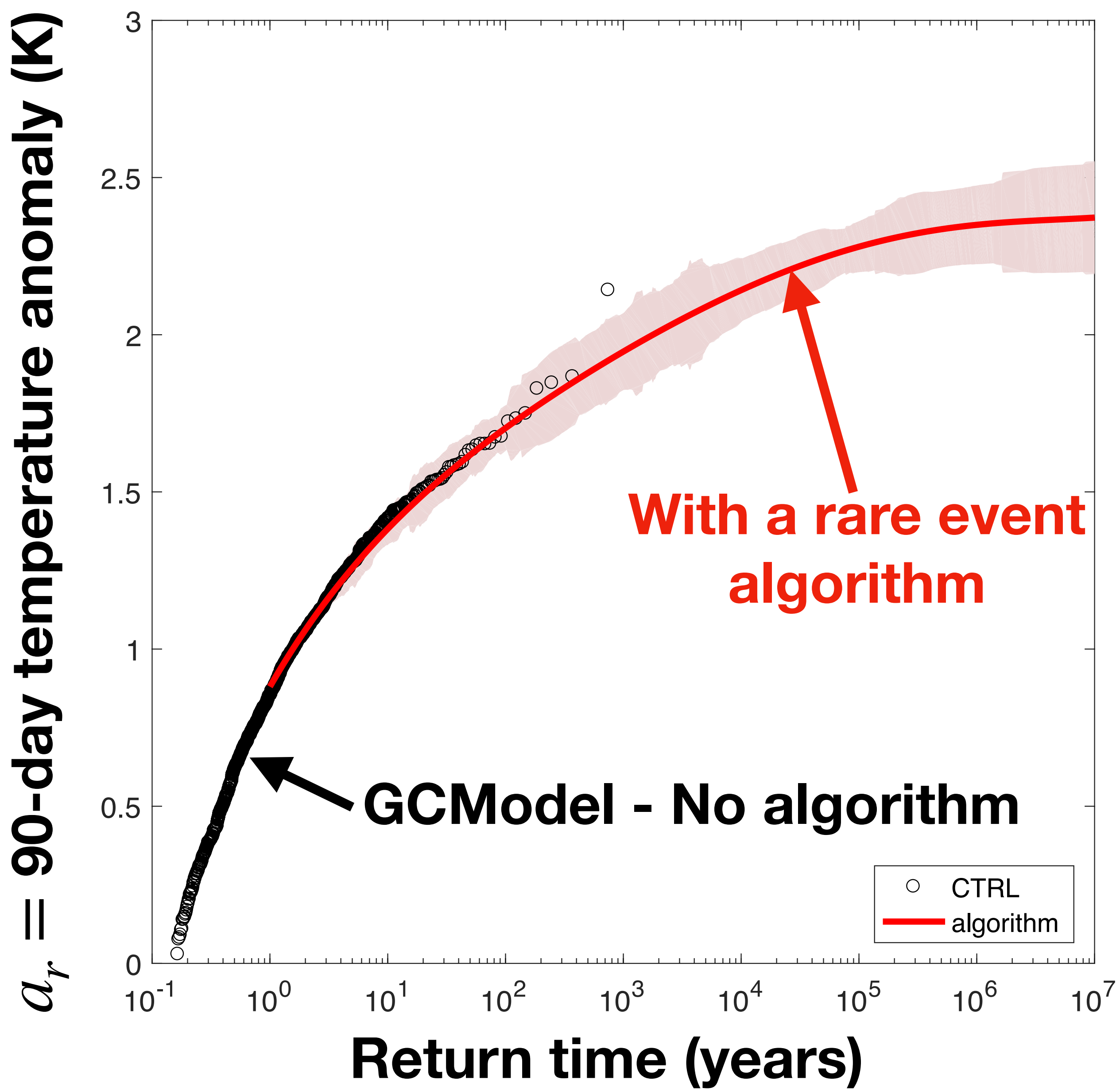


The trajectory statistics is tilted towards the events of interest.

Sample paths of the Giardina Kurchan algorithm

(from Bouchet, Jack, Lecomte, Nemoto, 2016)

Return time plot computed using a rare event algorithm (PLASIM)



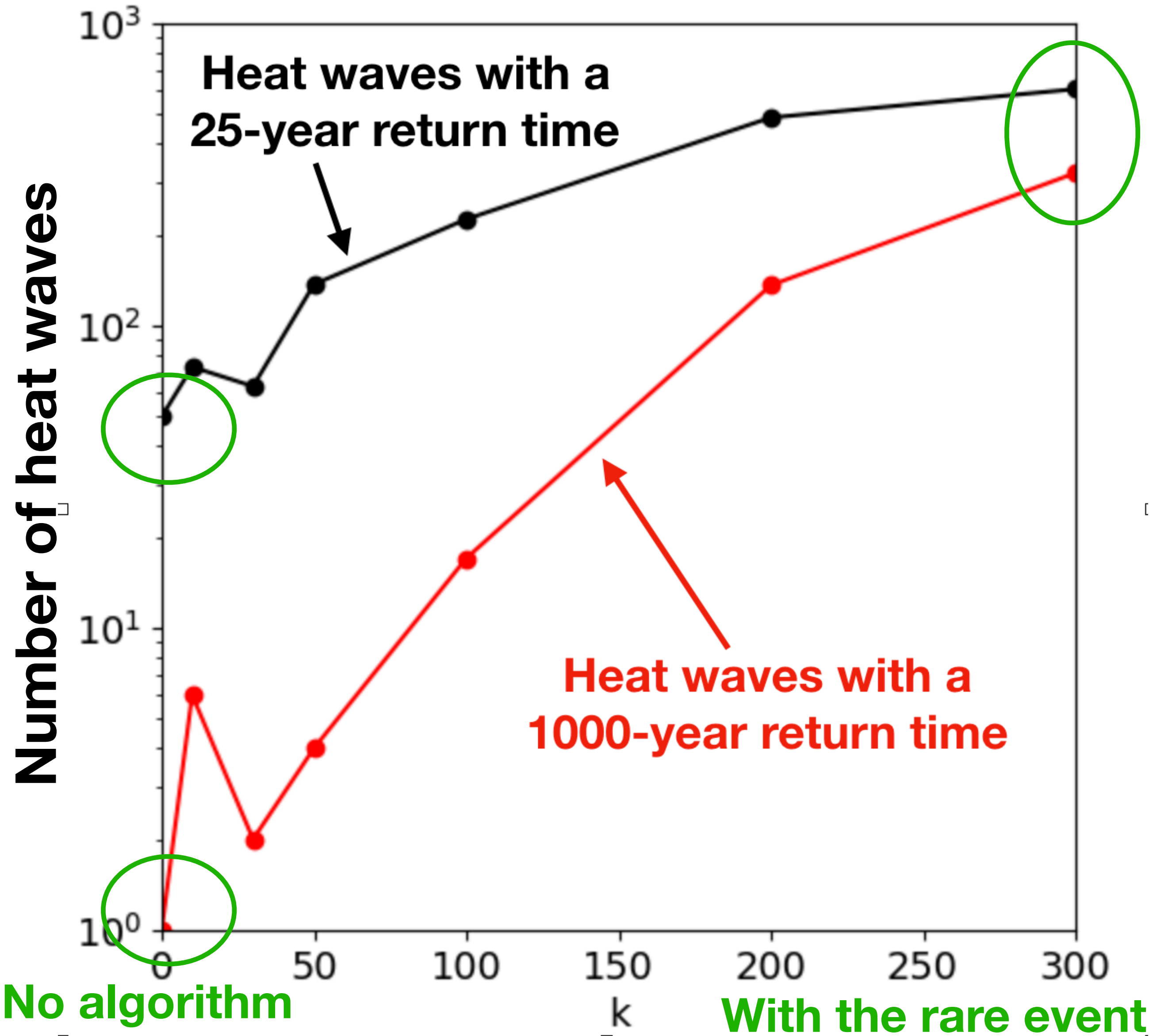
Extremes of 90-day Europe heat waves

PLASIM model.
No seasonal cycle.
Del-Moral—Garnier (or Giardina—Kurchan) algorithm.

F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

At a fixed numerical cost, we can study events which are several orders of magnitude rarer.

Oversampling of extreme events using a rare event algorithm (CESM)



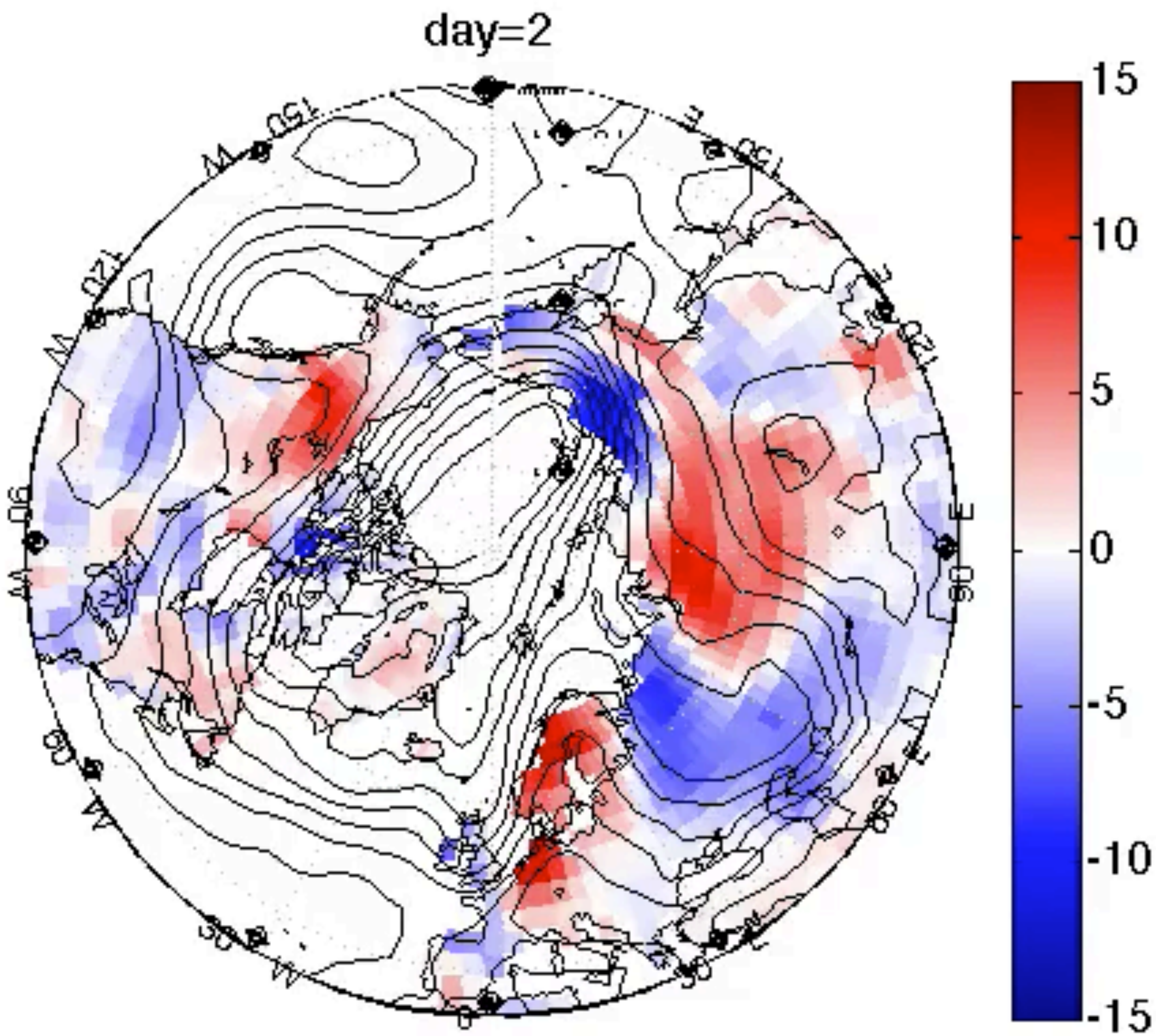
CESM model
With seasonal cycle.
Del-Moral—Garnier (or Giardina—Kurchan) algorithm.

F. Ragone, and F. Bouchet, GRL, 2021.

Number of observed heat waves for 1,000 of simulations

With a rare event algorithm, we get several hundreds more heat waves, for a fixed return times

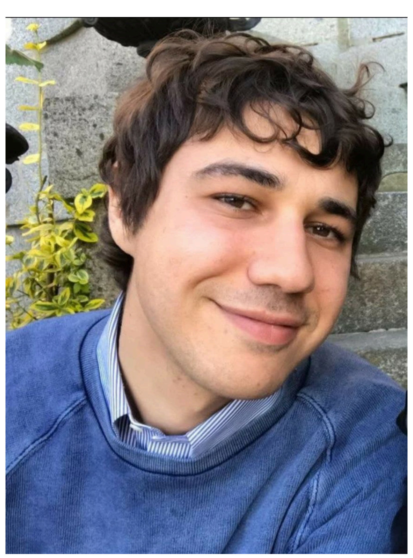
Heat wave dynamics



Plasim heat wave over Scandinavia

500 hPa geopotential height and temperature anomalies

II-b) Heat wave dynamics and global teleconnection patterns for extremes



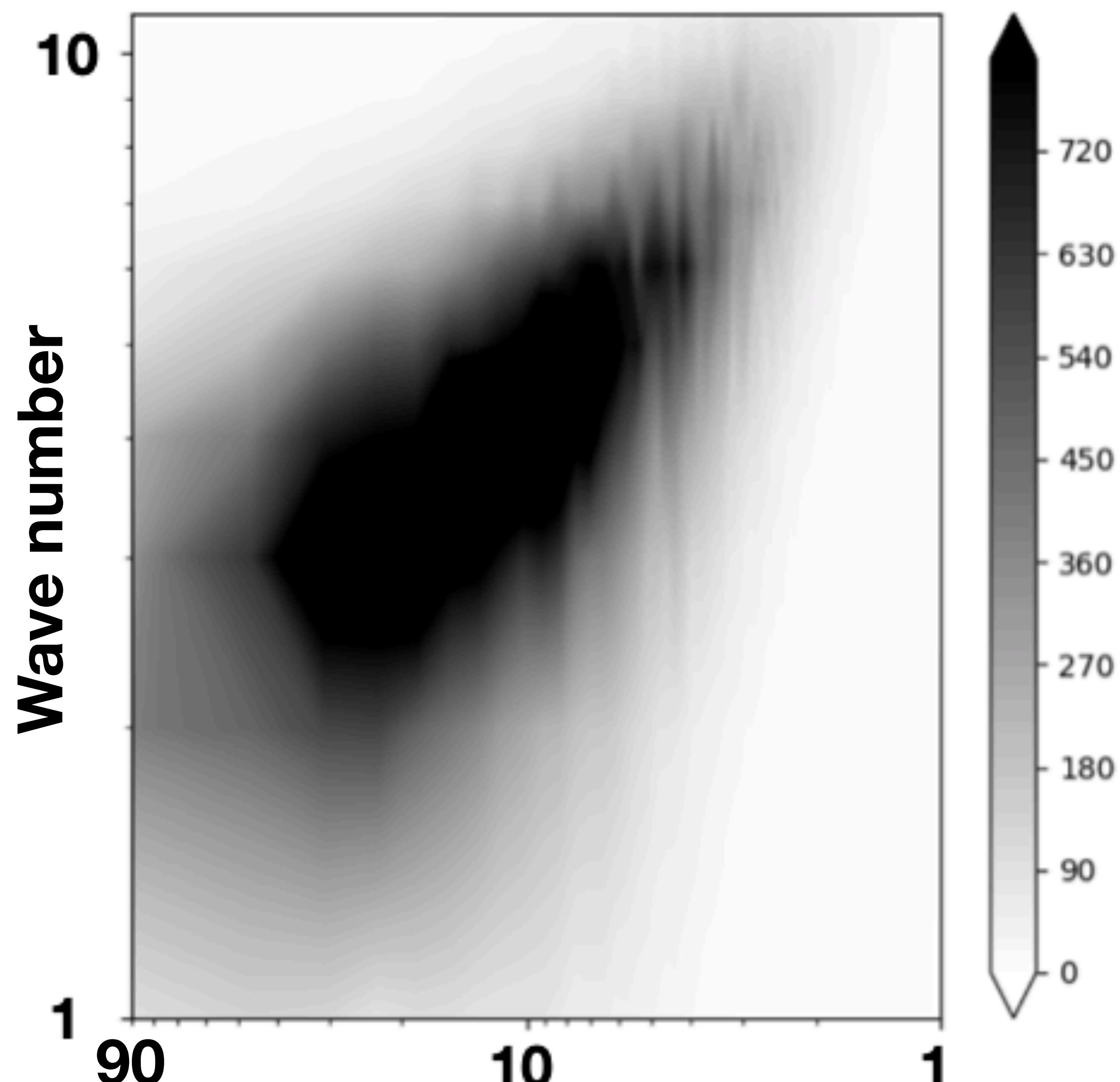
Dario Lucente

George Miloshevich

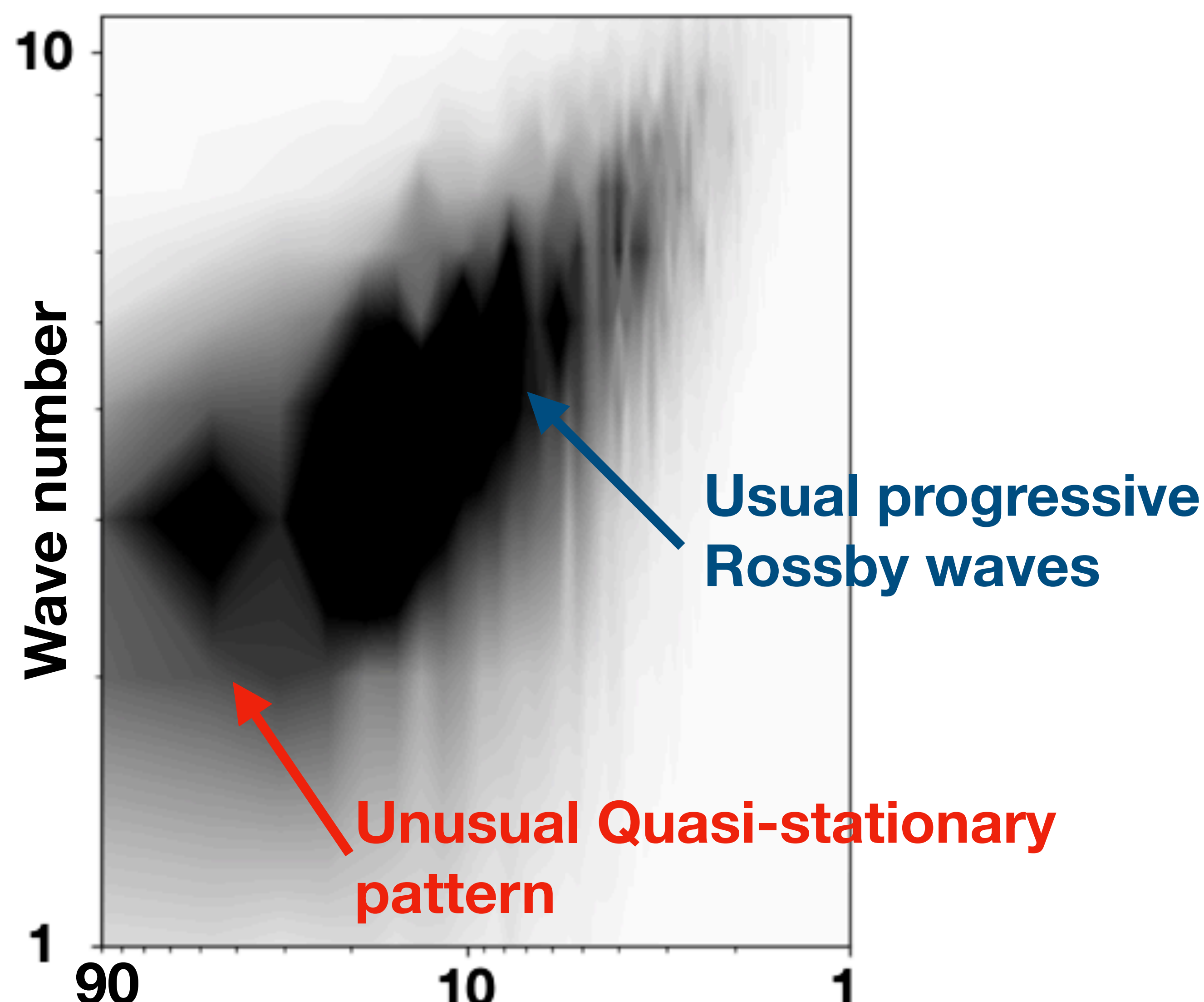
Francesco Ragone

Heat wave = unusual quasi stationary pattern + progressive Rossby wave

Hayashi spatio-temporal spectrum for eastward waves - CESM model (from the 500 hPa geopotential height over a latitudinal band $55^o - 75^o N$)

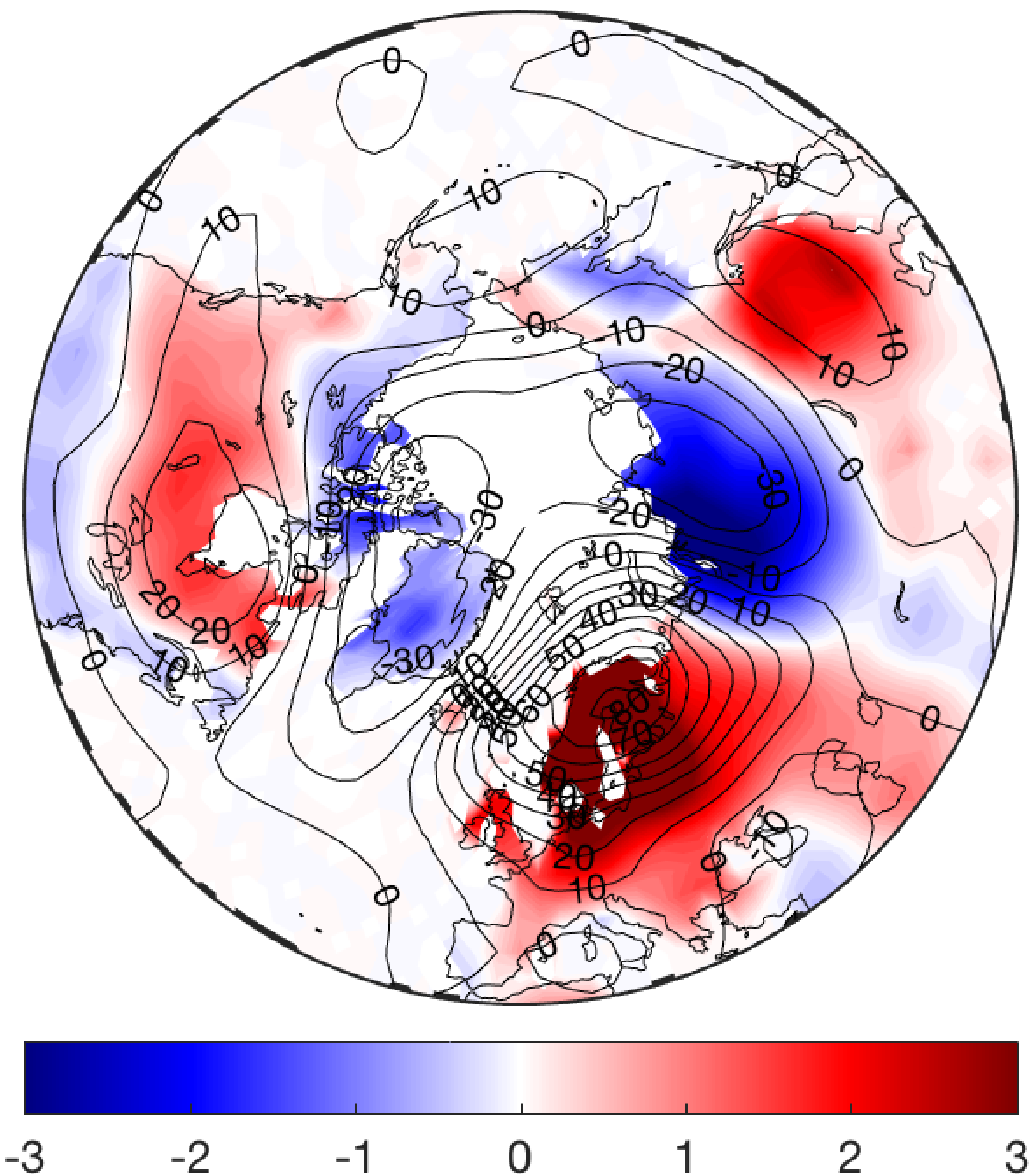


CESM Climatology



CESM $T = $ 30-day extreme heatwaves

Extreme teleconnection pattern



500 hPa geopotential height and temperature anomalies

Extreme teleconnection patterns = conditional averages with

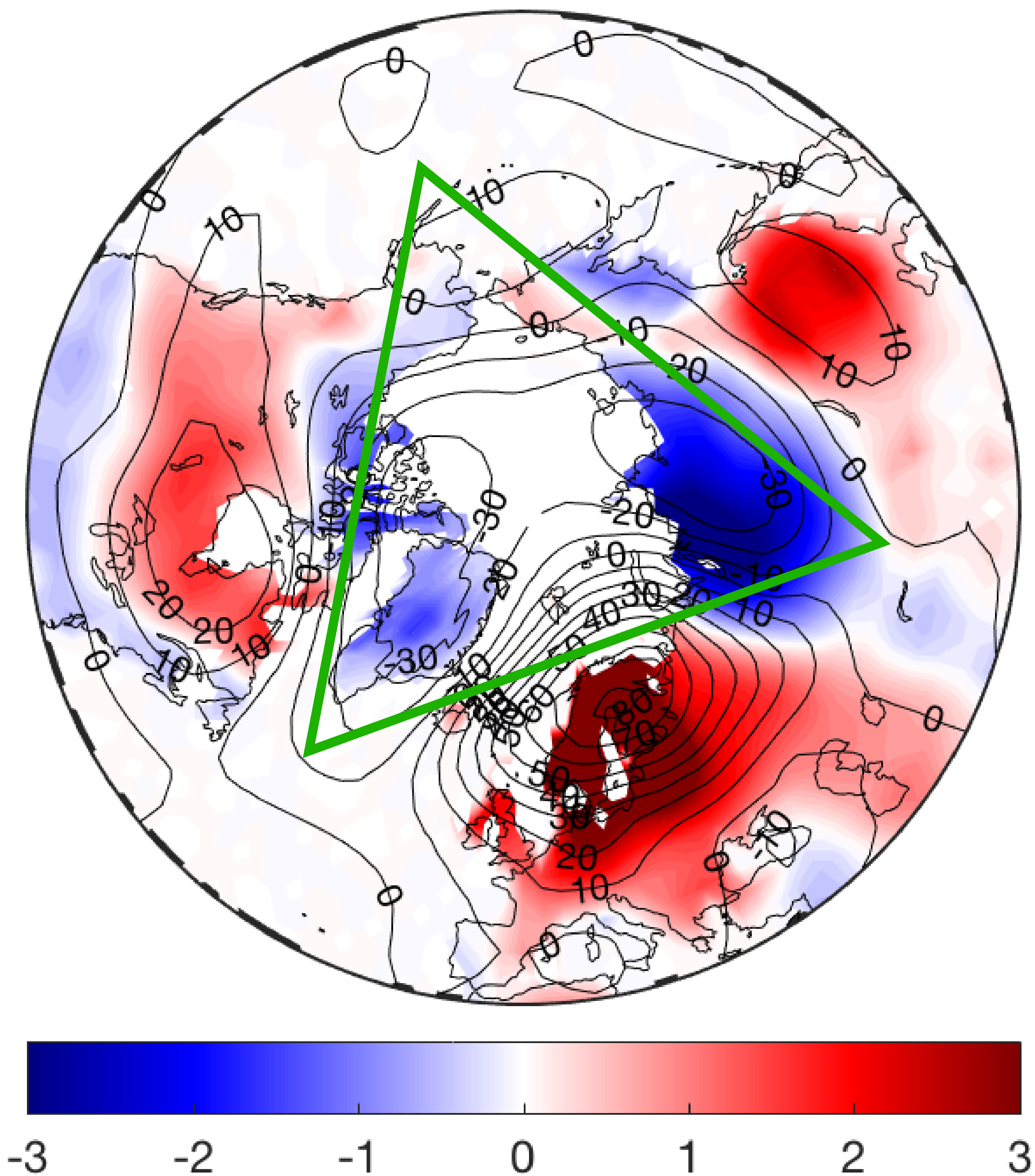
$$\frac{1}{D}\int_0^D \mathrm{d}t \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} \mathrm{d}\mathbf{r}\ \mathrm{T}_S(\mathbf{r},t) > 2\ \mathrm{K}$$

and $D = 40$ days.

Plasim model. Summer Scandinavian heat waves.

F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

Extreme teleconnection pattern



500 hPa geopotential height and temperature anomalies

Extreme teleconnection patterns = conditional averages with

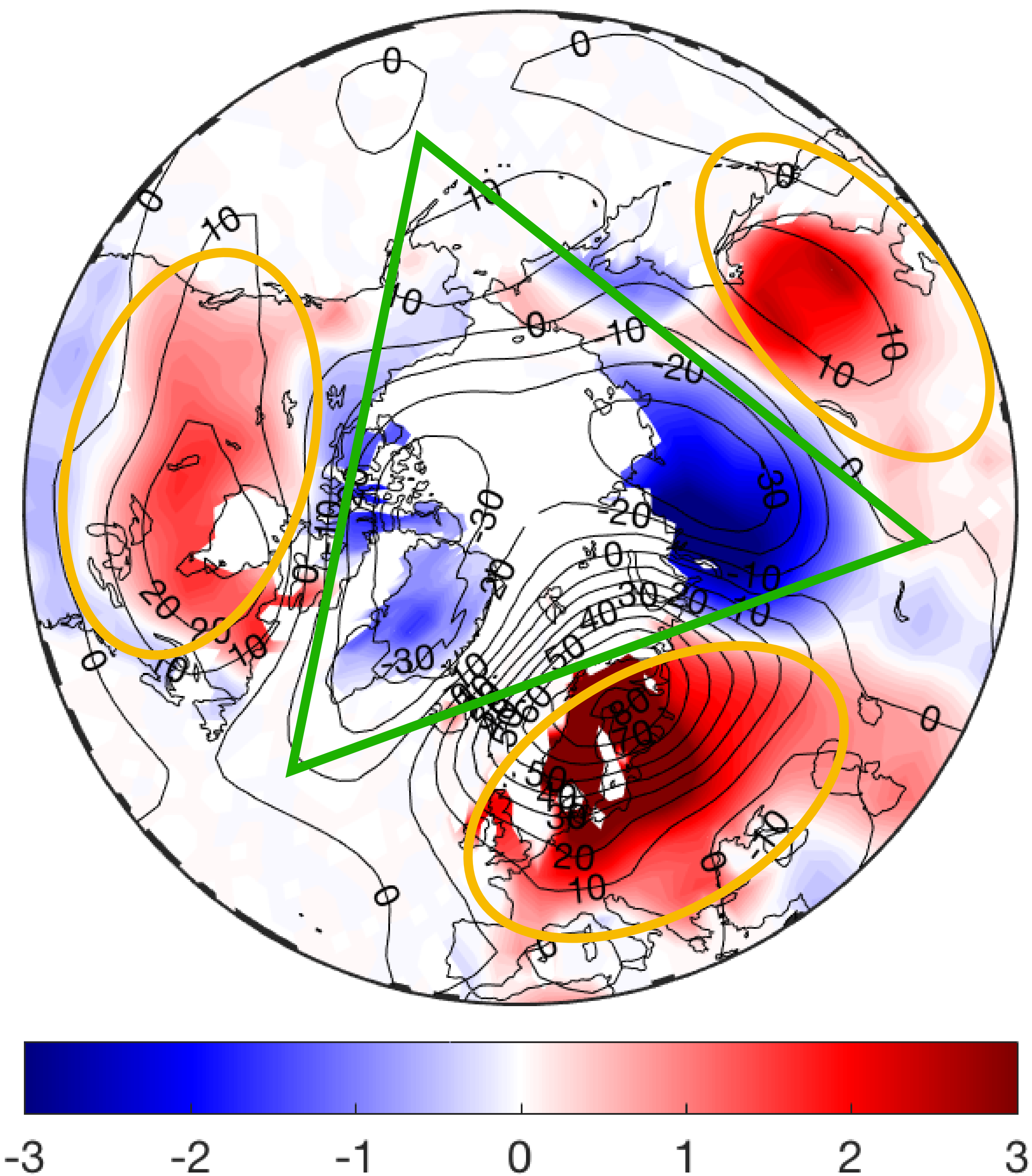
$$\frac{1}{D}\int_0^D \mathrm{d}t \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} \mathrm{d}\mathbf{r}\ \mathrm{T_S}(\mathbf{r},t) > 2\ \mathrm{K}$$

and $D = 40$ days.

Plasim model.
Summer Scandinavian heat waves.

F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

Extreme teleconnection pattern



500 hPa geopotential height and temperature anomalies

Extreme teleconnection patterns = conditional averages with

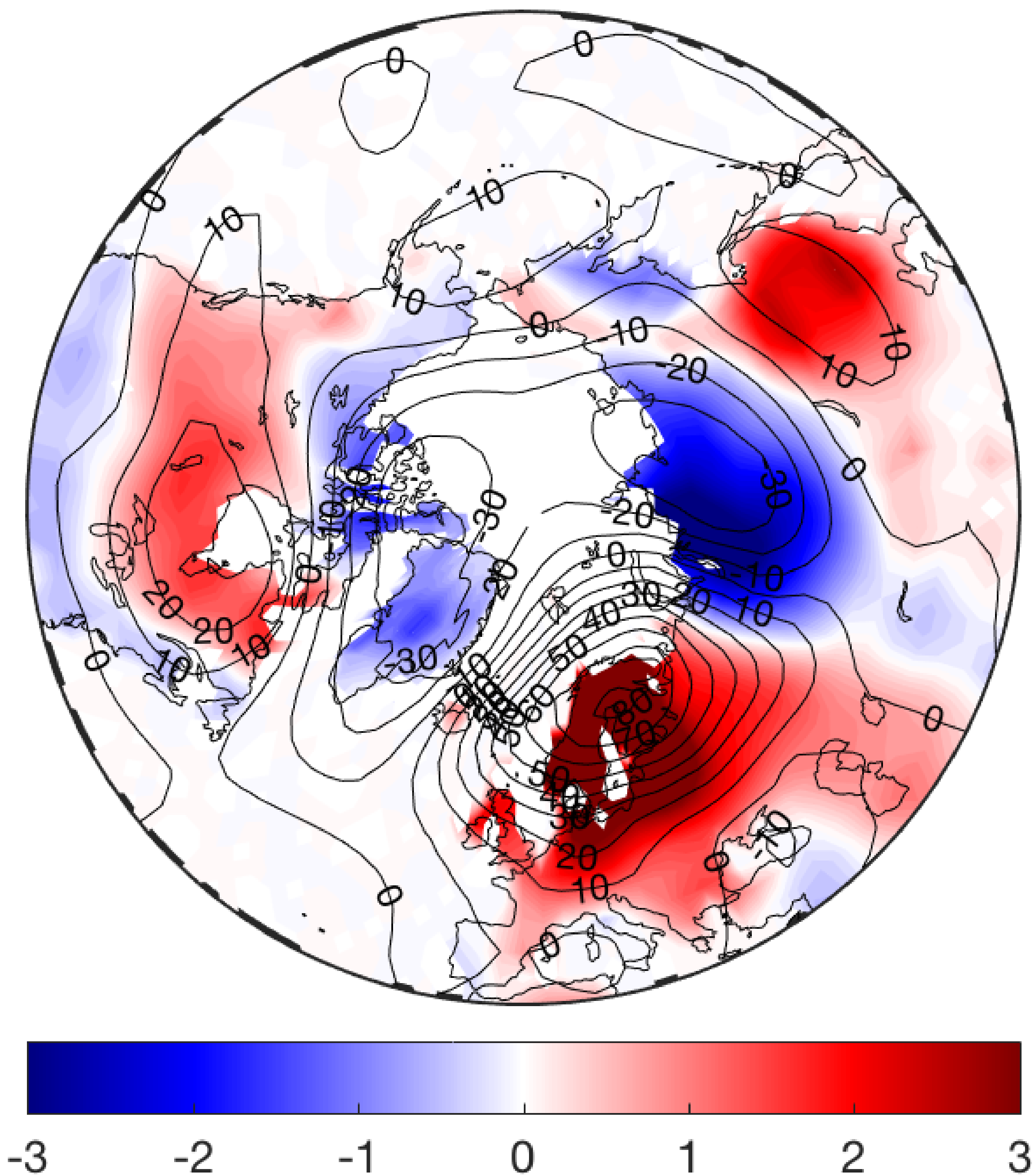
$$\frac{1}{D}\int_0^D \mathrm{d}t \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} \mathrm{d}\mathbf{r}\ \mathrm{T_S}(\mathbf{r},t) > 2\ \mathrm{K}$$

and $D = 40$ days.

Plasim model.
Summer Scandinavian heat waves.

F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

Extreme teleconnection pattern



500 hPa geopotential height and temperature anomalies

Extreme teleconnection patterns = conditional averages with

$$\frac{1}{D}\int_0^D \mathrm{d}t \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} \mathbf{dr}\ \mathrm{T_S}(\mathbf{r}, t) > 2\ \mathrm{K}$$

and $D = 40$ days.

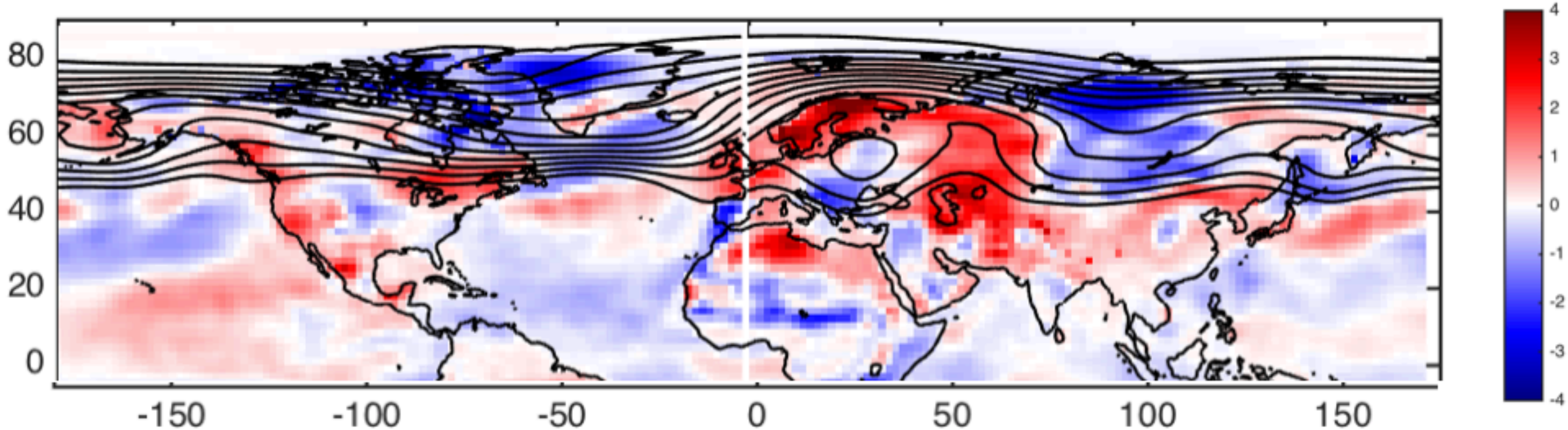
Plasim model.
Summer Scandinavian heat waves.

F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

Extreme teleconnection patterns differ from teleconnections for typical fluctuations and are not characterized by a single wavenumber but are much constrained by geography.

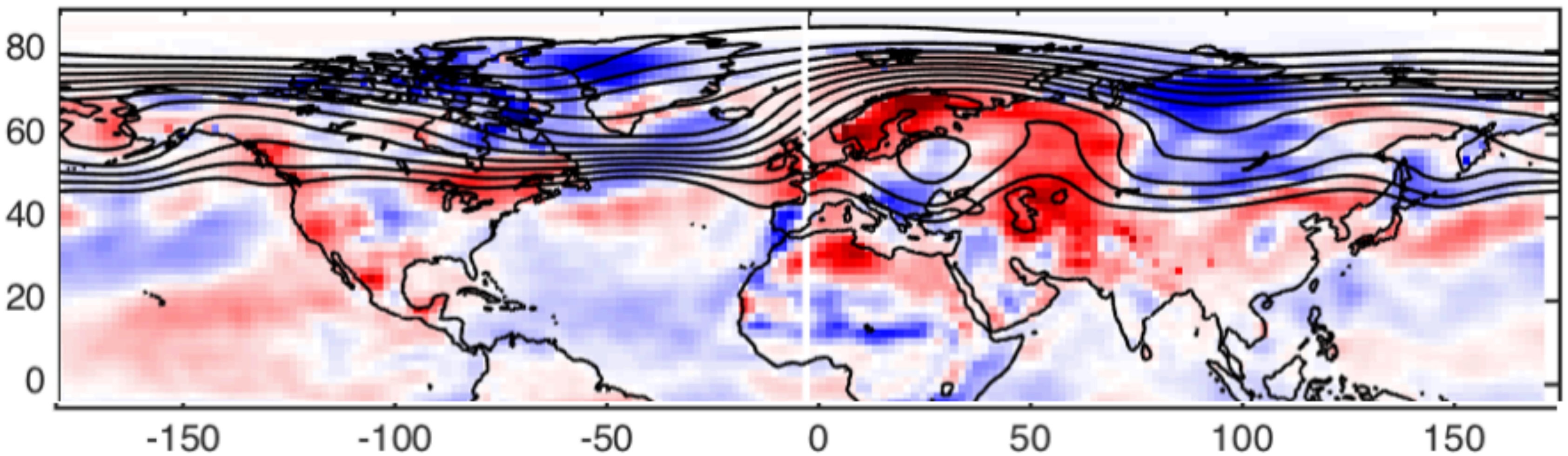
2018 heat wave over Scandinavia

Era 5 reanalysis dataset



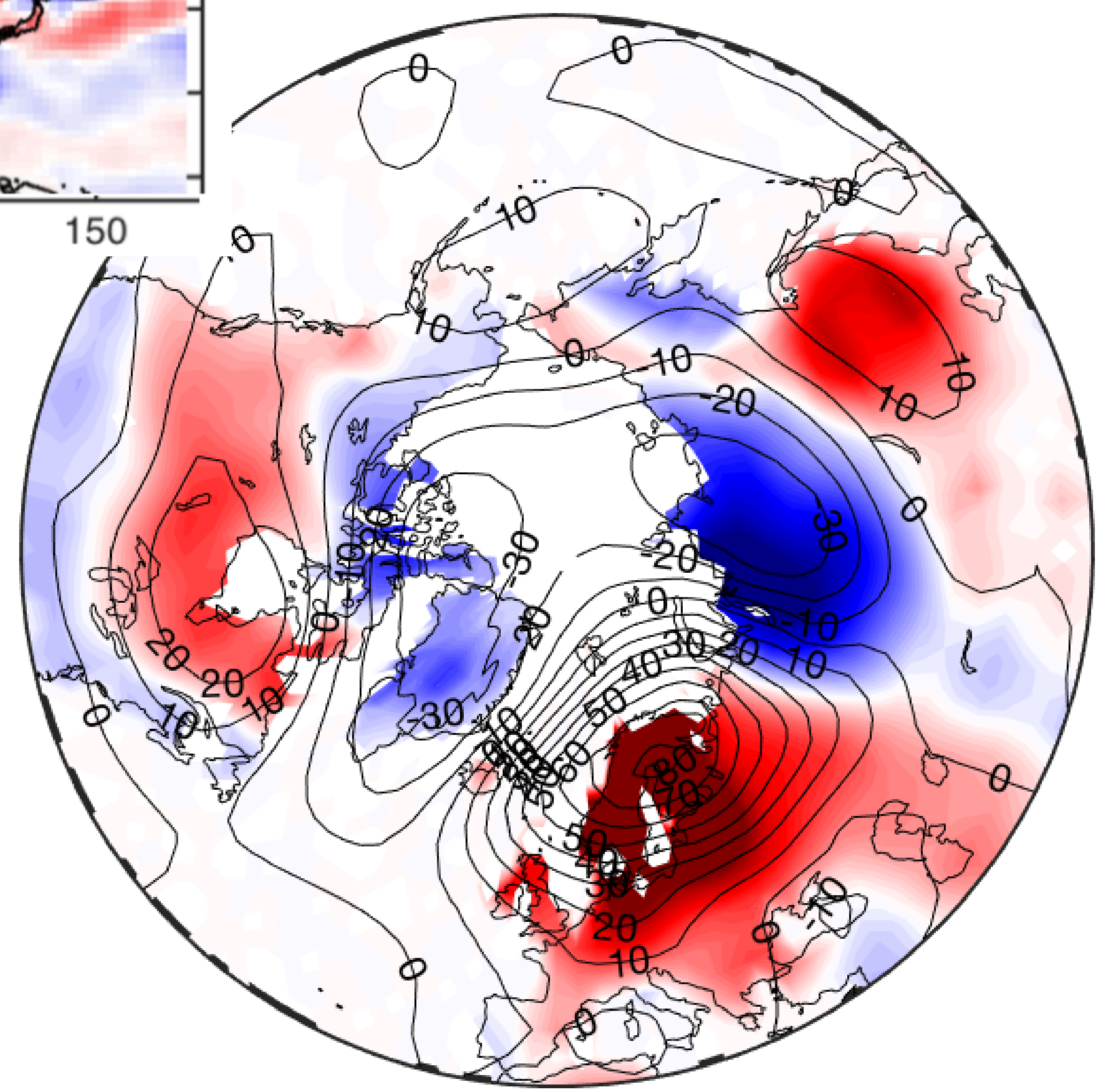
Temperature anomalies and 500 hPa geopotential height in July 2018

2018 heat wave over Scandinavia



Observed in July 2018

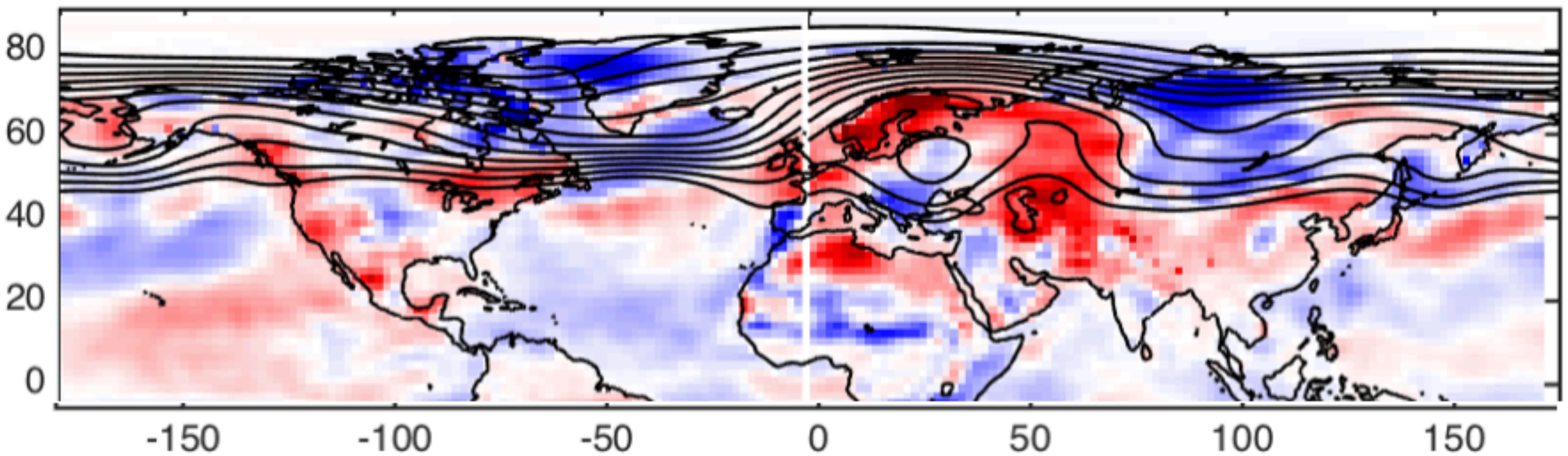
Published in January 2018



Is this just by chance?

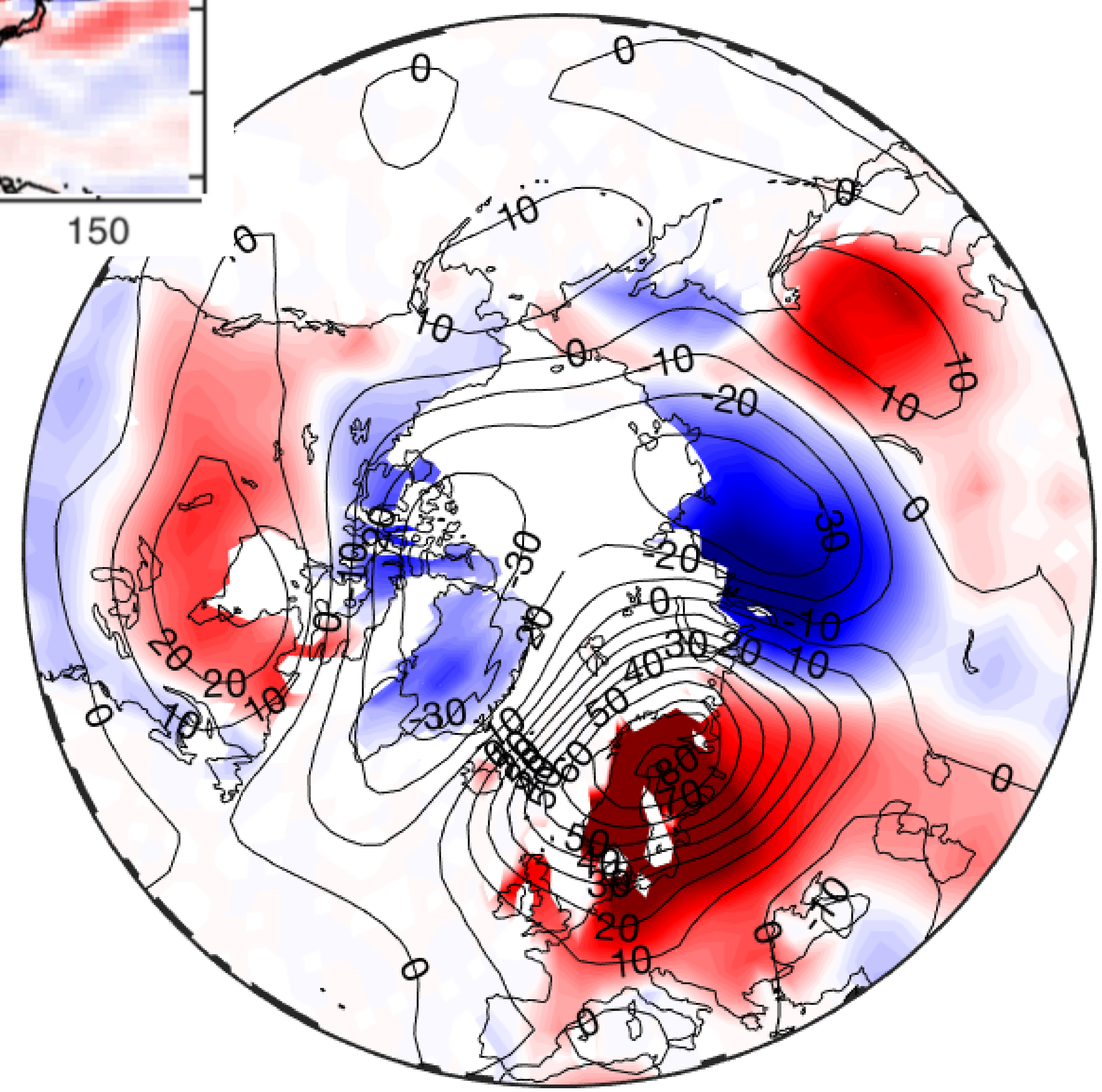
How good are the models for predicting those patterns?

2018 heat wave over Scandinavia



Observed in July 2018

Published in January 2018

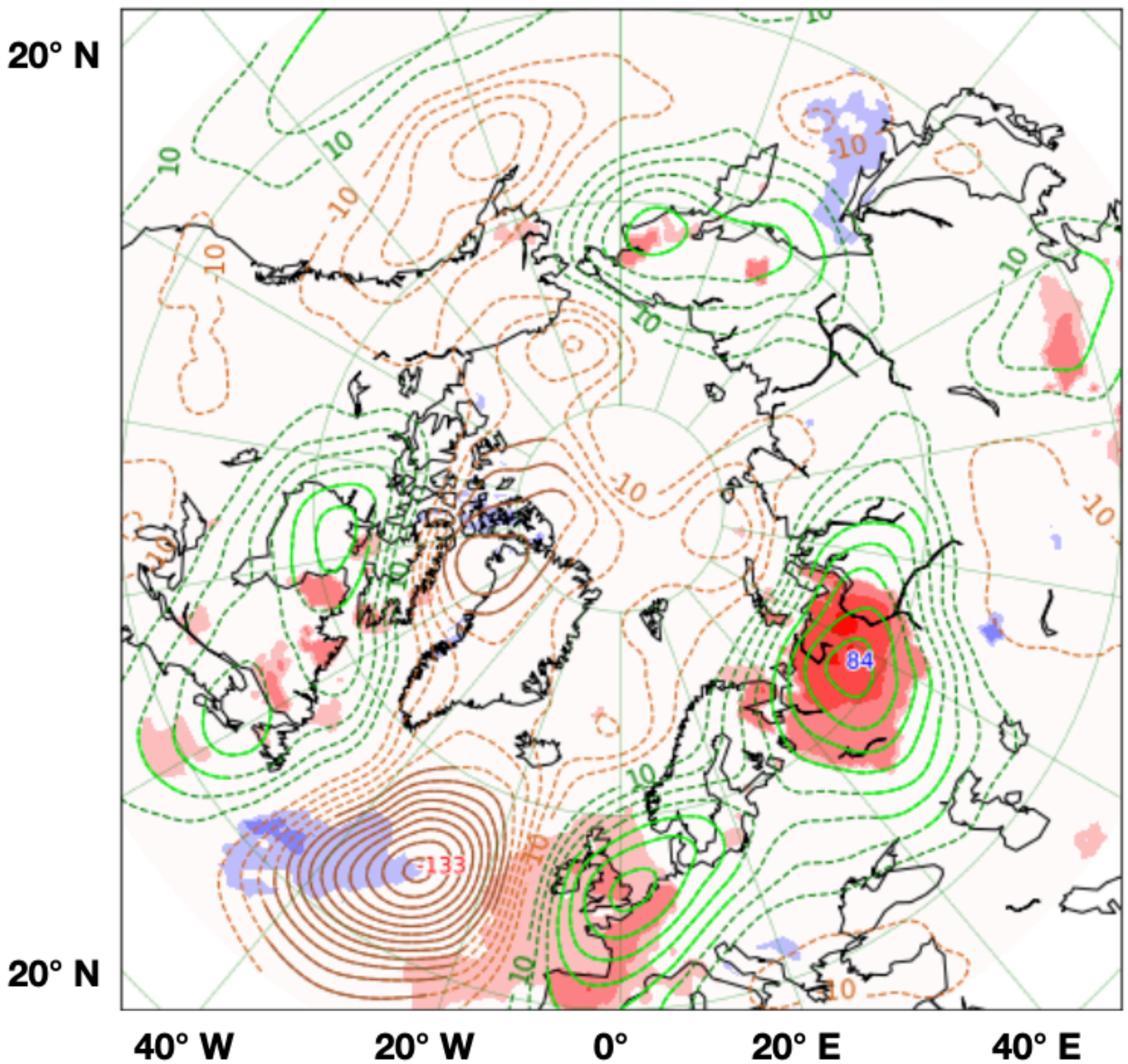


Is this just by chance?

How good are the models for predicting those patterns?

CESM patterns (2020) are similar to Plasim ones (2018).

Teleconnection patterns for moderate heat waves over France - ERA5



ERA5 reanalysis

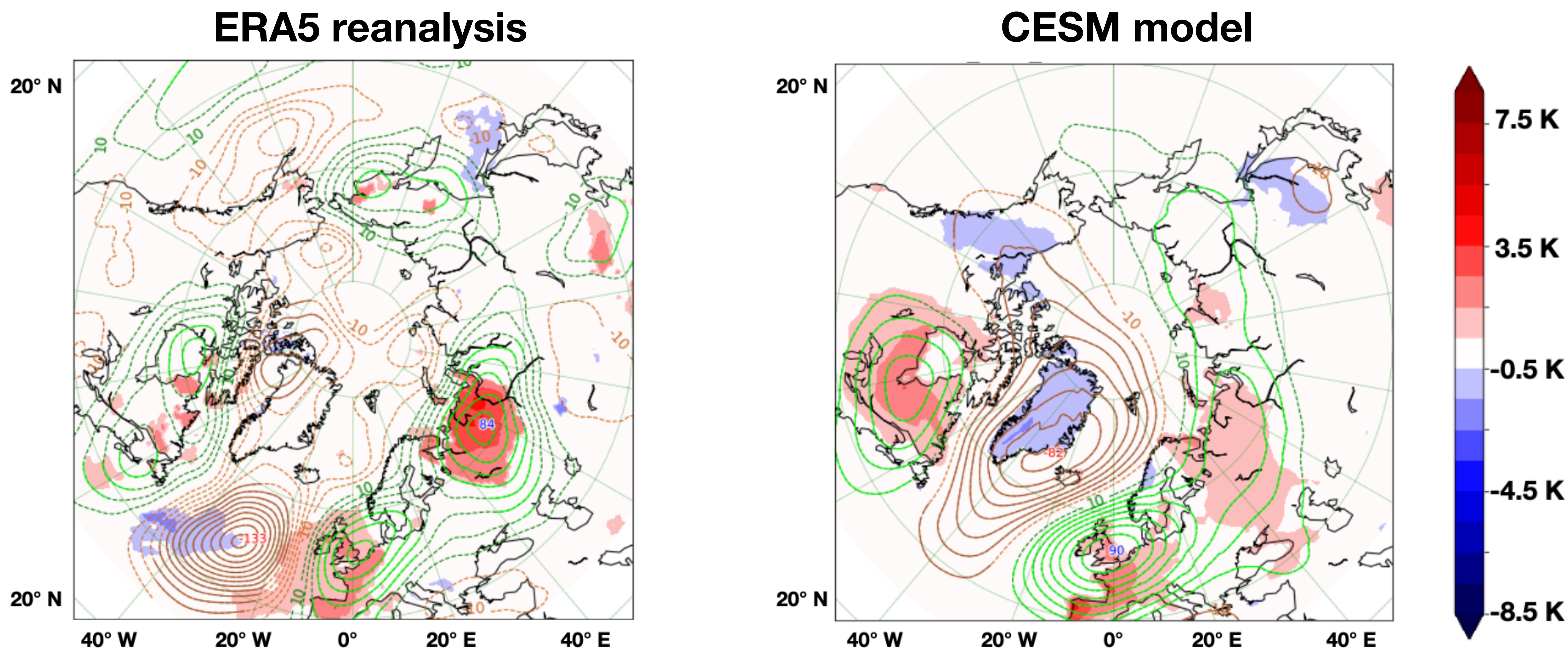
4-year return time.
$D = 14.$
At day $\tau = 0.$

Dashed = not statistically significant for a t>2 student test.

Temperature anomalies and 500 hPa geopotential height conditioned on moderate heat waves

It is extremely difficult to have statistically significant patterns from reanalysis datasets.

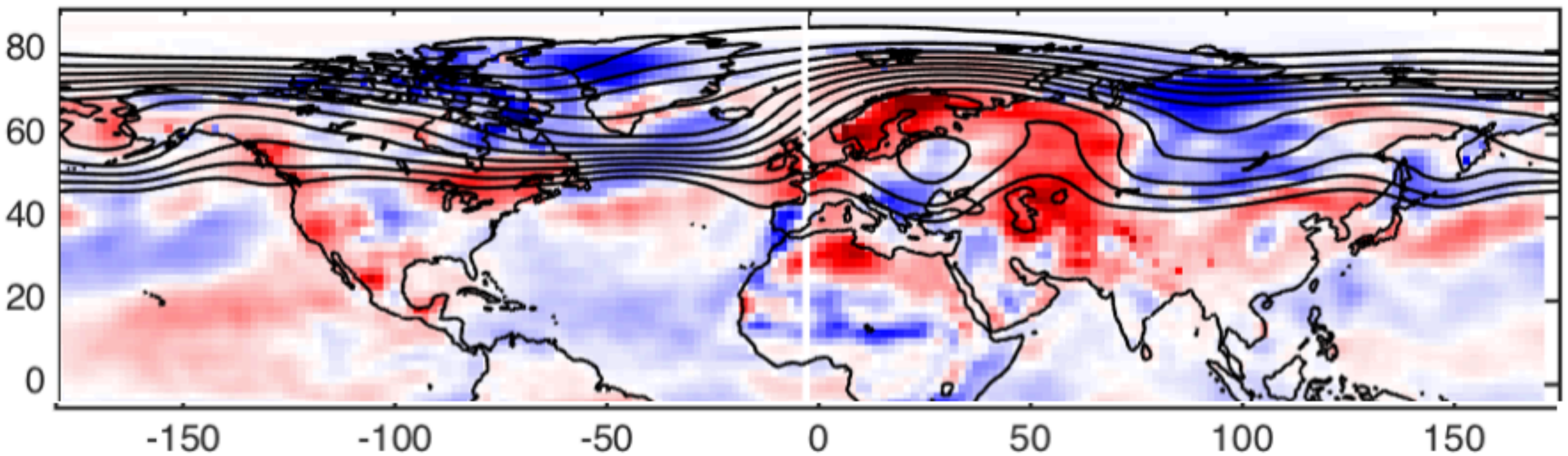
Teleconnection patterns for moderate heat waves over France



Temperature anomalies and 500 hPa geopotential height conditioned on extreme heat waves with a 4-year return time, with $D = 14$, at day $\tau = 0$

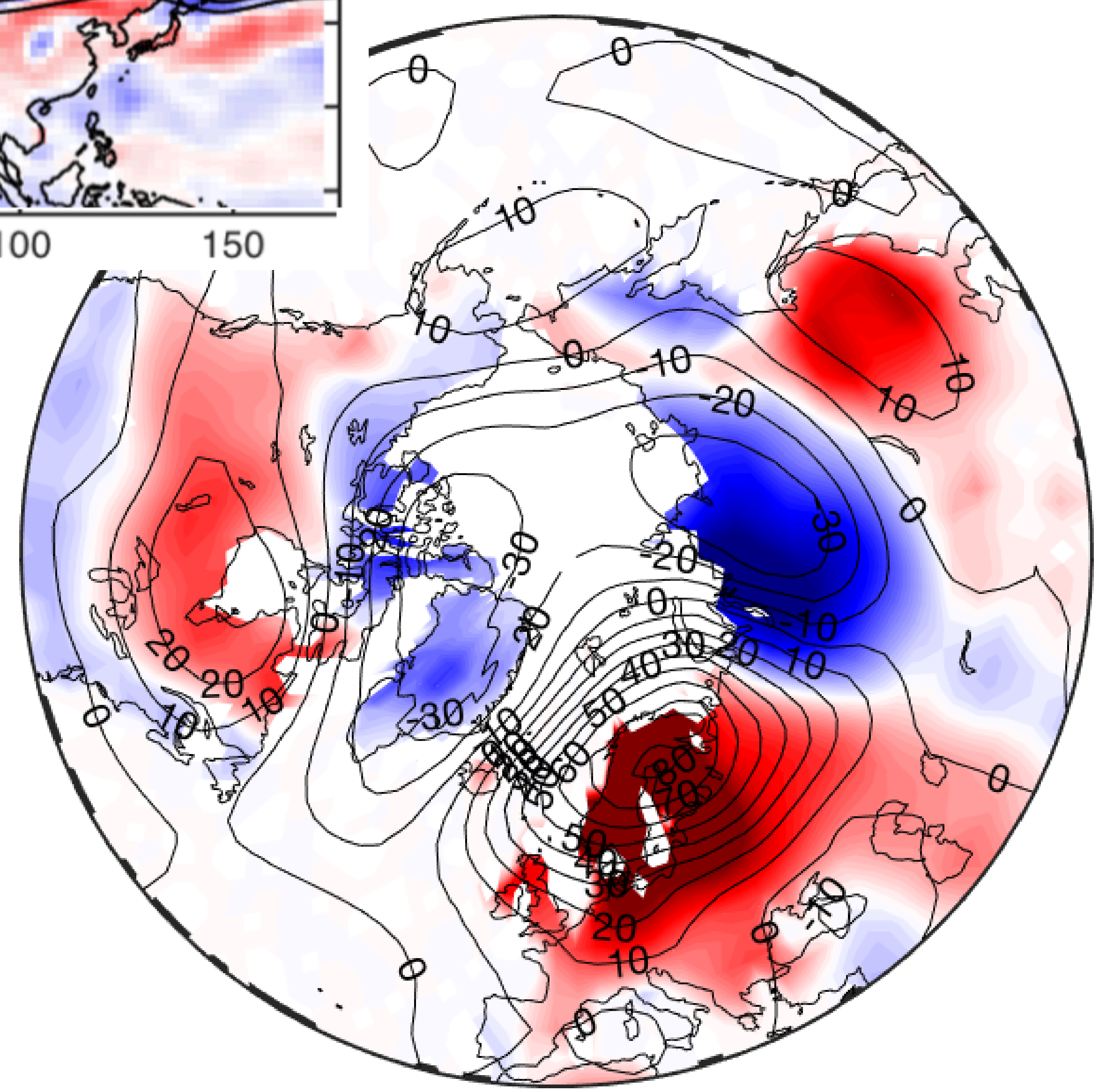
For moderately extreme heat waves, CESM and ERA5 reanalysis dataset are qualitatively consistent.

2018 heat wave over Scandinavia



Observed in July 2018

Published in January 2018



Conclusions:

Climate models predict well moderately extreme teleconnection patterns.

This is a hint that they might predict correctly more extreme teleconnection patterns and that what we observed is not a coincidence.

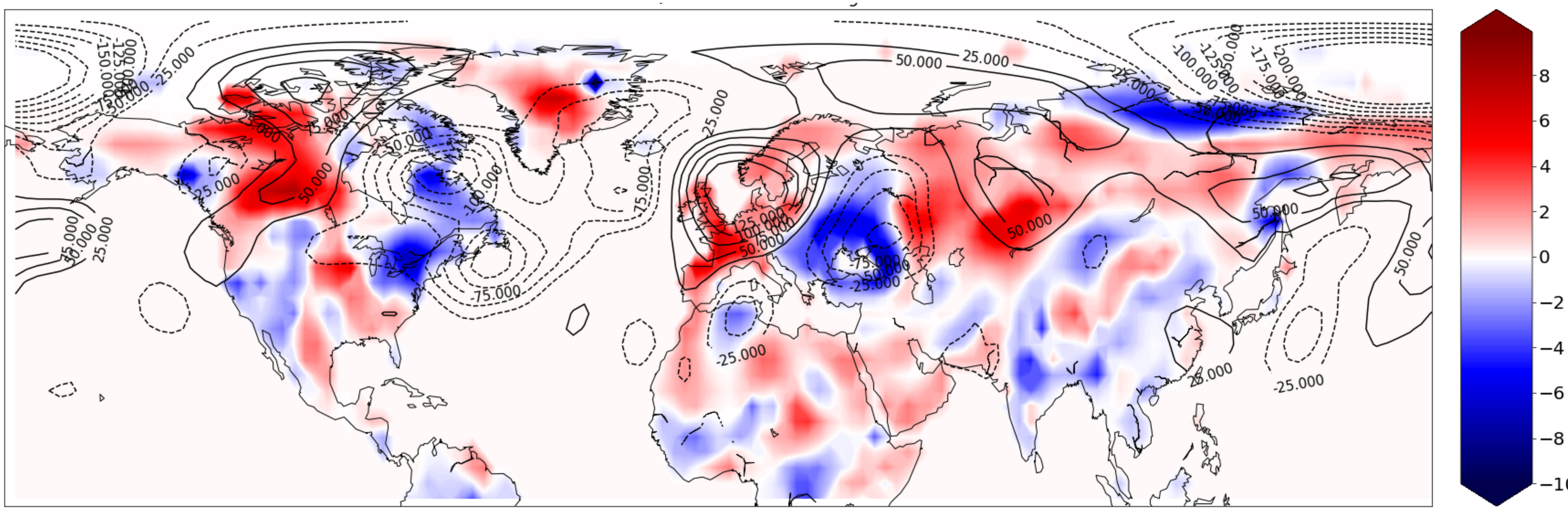
III) Predicting heat waves (committor functions) with deep neural networks

With P. Abry, P. Borgnat, V. Jacques-Dumas, G. Miloshevich, and F. Ragone

Valerian Jacques-Dumas

George Miloshevich

Predicting heat waves with a deep neural network - 1) Data



Surface temperature (T_s, colors) and 500 hPa geopotential height (Z_g, lines) anomalies

- **Plasim and CESM climate models.**
- We use summer (JJA) data: 8 maps/day, 90 days/year, 1000 year = 720 000 maps.
- For Plasim data, each field has a resolution 64×128, restricted to 25×128 above 30^o North.

Heat wave definition

- $X(t) = T_s$ field at time t, or $X(t) = (T_s, Z_g)$ fields at time t.

- $Y(t)$: **time and space averaged surface temperature anomaly within τ days:**

$$Y(t) = \frac{1}{D}\int_{t+\tau}^{t+\tau+D} \frac{1}{|\mathscr{A}|}\int_{\mathscr{A}} T_s(\vec{r}, u)\ \mathrm{d}\vec{r}\ \mathrm{d}u,$$
$$\text{and } Z(t) = 1 \text{ if } Y(t) > a, \text{ and } Z(t) = 0 \text{ otherwise}$$

- $Z(t) \in \{0,1\}$. A heat wave occurs if $Z = 1$.

- We have a classification problem for the data (X, Z). We want to learn the probability $q(x)$ that $Z = 1$ given that $X = x$ (committor function).

- 5% most extreme events: $a = a_5 = 3.08\ \mathrm{K}$. 2.5% most extreme events: $a = a_{2.5} = 3.7\ \mathrm{K}$. 1.25% most extreme events: $a = a_{1.25} = 4.23\ \mathrm{K}$.

Predicting heat waves with a deep neural network

Observing the temperature and geopotential height at 500 hPa today, what is the probability to observe a D-day heat wave starting τ days from now?

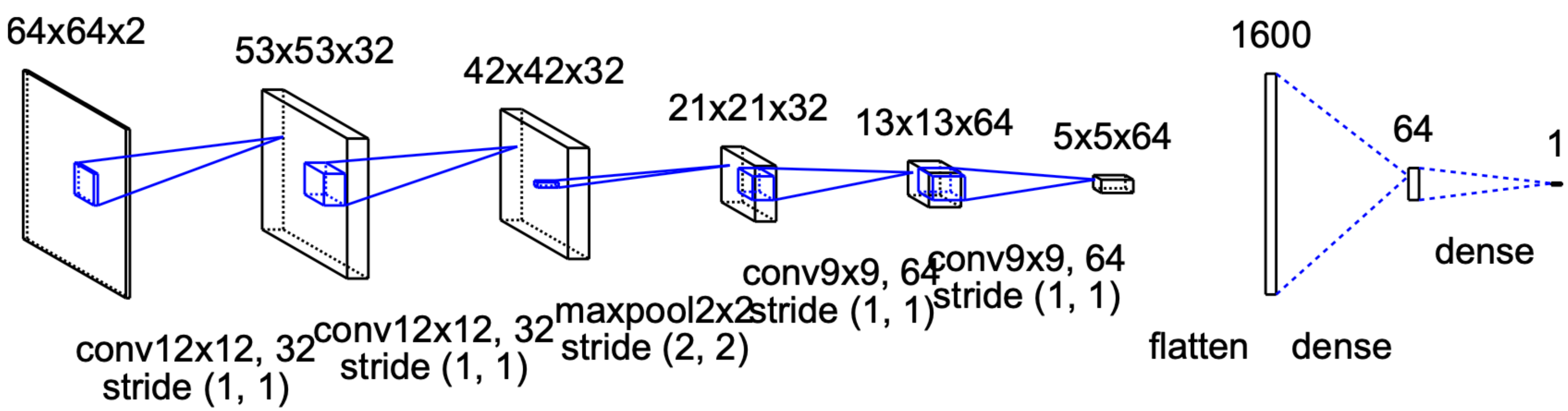
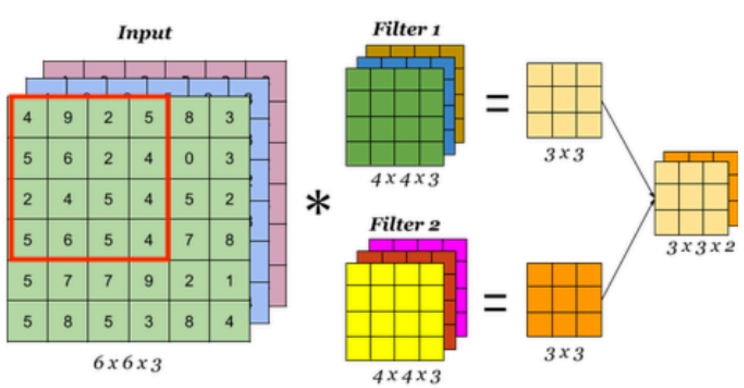
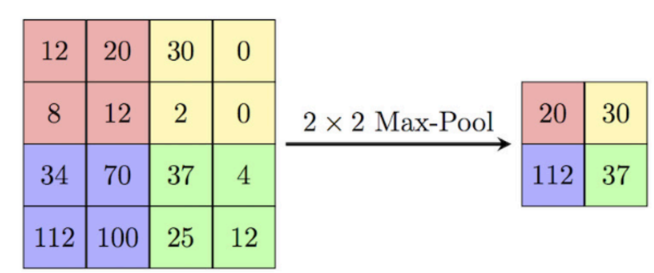


Figure 2: Architecture of the CNN used to forecast extreme heatwaves.



(a) A Convolution layer



(b) A MaxPool layer

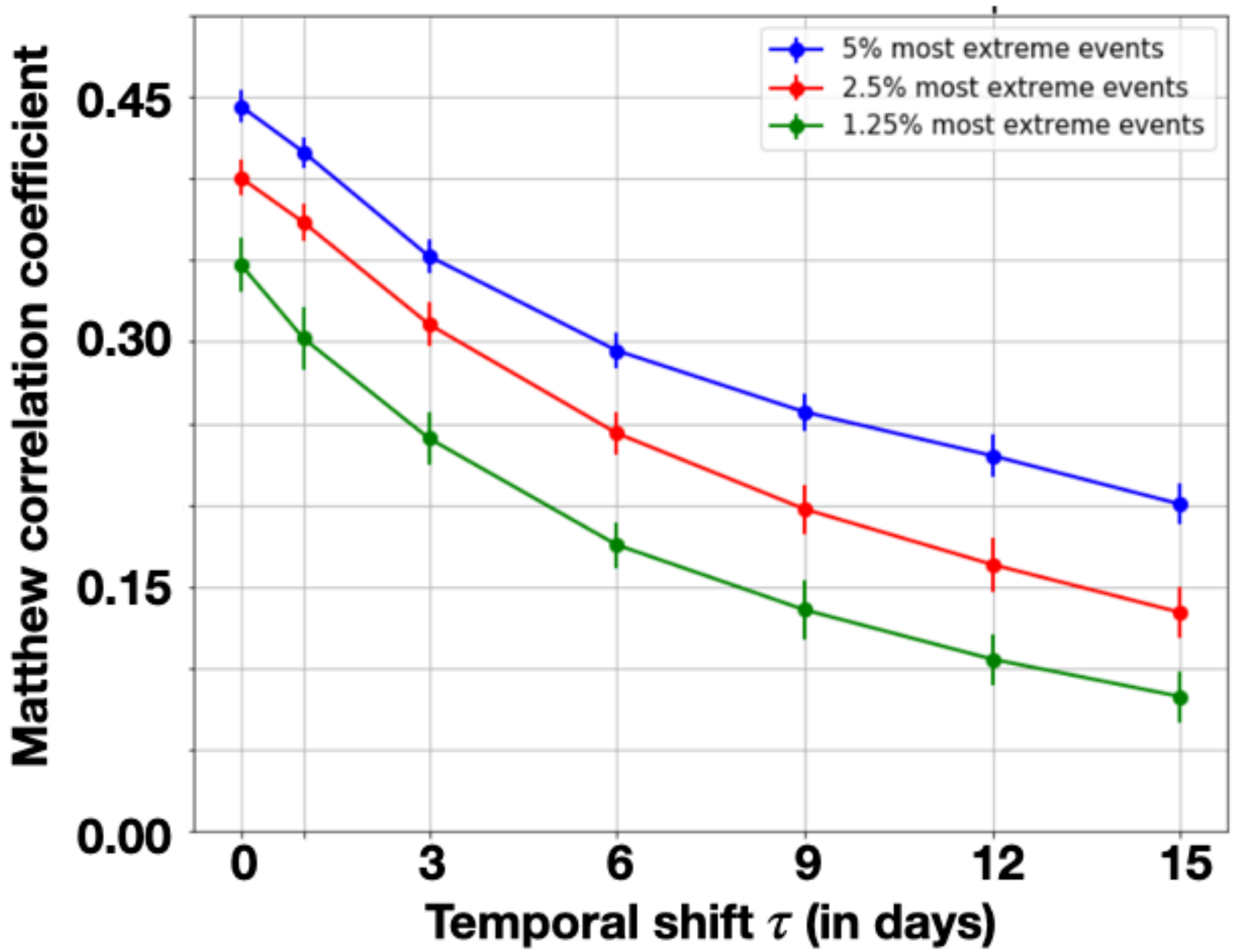
Machine learning for extreme heat waves

- **Supervised learning** from 1,000 years of climate model data (720 000 couples (X, Z)).

- We use **undersampling** to deal with class imbalance.

- We use **transfer learning** between return levels a, first training a deep neural network for less rare events, and then transferring to learn rarer events with less data.

Predicting heat waves



Heat waves over France
$D = 14$

Predictability, τ day ahead, for a 14-day heatwave from the temperature and GPH fields

We have very interesting prediction capabilities up to 15 days ahead of time for $D = 14$-day heatwaves

V. Jacques-Dumas, F. Ragone, F. Bouchet, P. Borgnat and P. Abry, 2021, sub. to IEEE TPAMI + ArXiv

The Normalized and Positively oriented logarithmic score

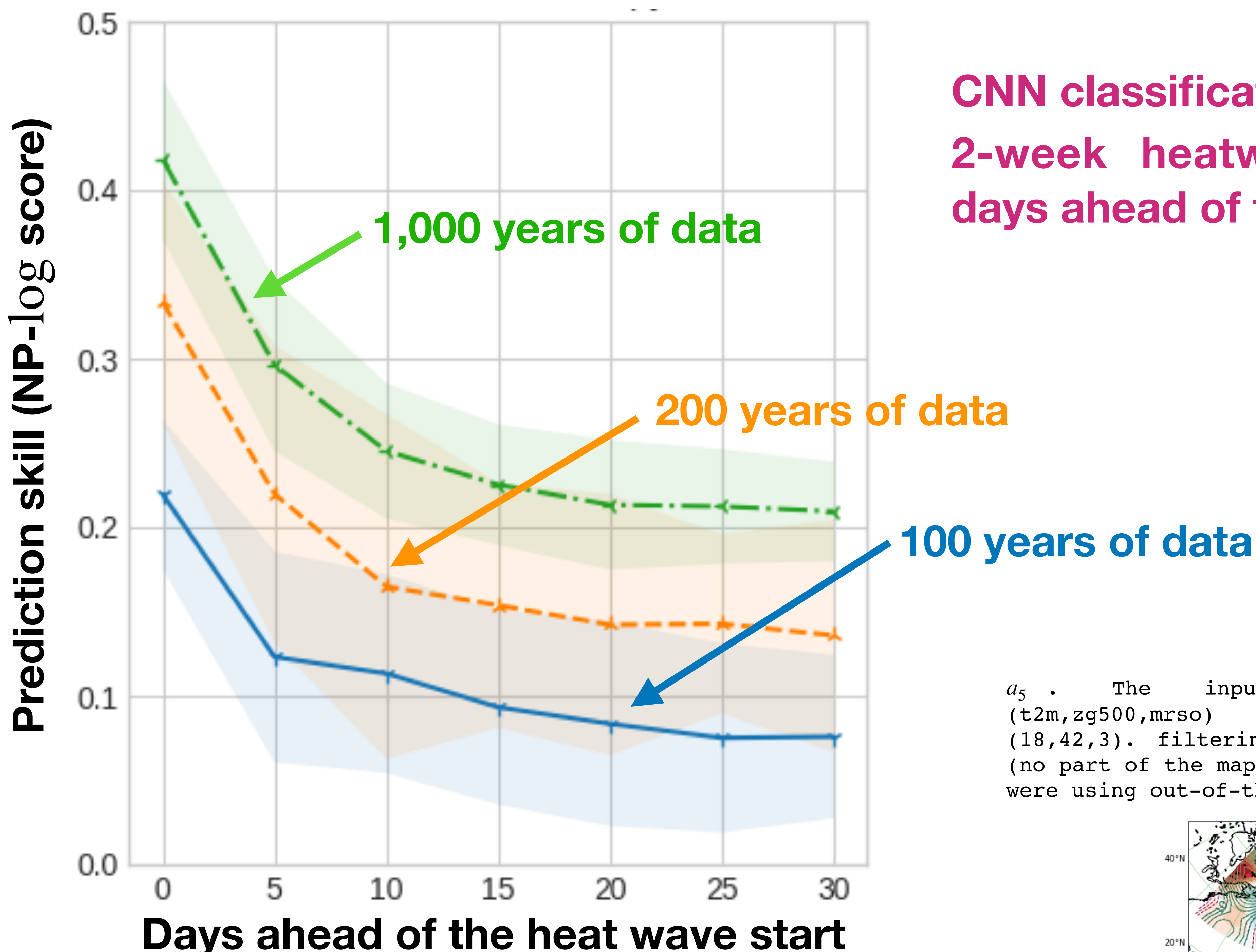
- In order to test the efficiency of the probabilistic prediction of the probabilistic classification, we use the logarithmic score $\mathbb{E}\left\{\log\left[p_{Y_n}\left(X_n\right)\right]\right\}$.

- We define a normalized and positively oriented logarithmic score

$$NP\log = a\mathbb{E}\left\{\log\left[p_{Y_n}\left(X_n\right)\right]\right\} + b,$$

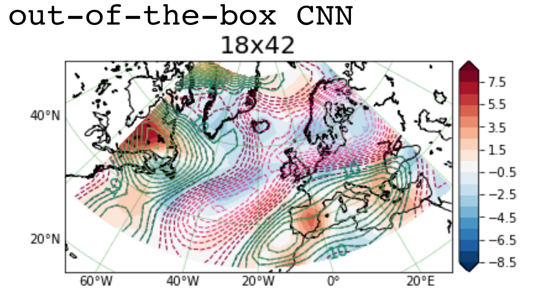
where a and b are such that $NP\log = 0$ for the prediction according to the climatology (prediction using no information on the state X) and $NP\log = 1$ for perfect prediction.

Machine learning for climate applications: a lack of data regime



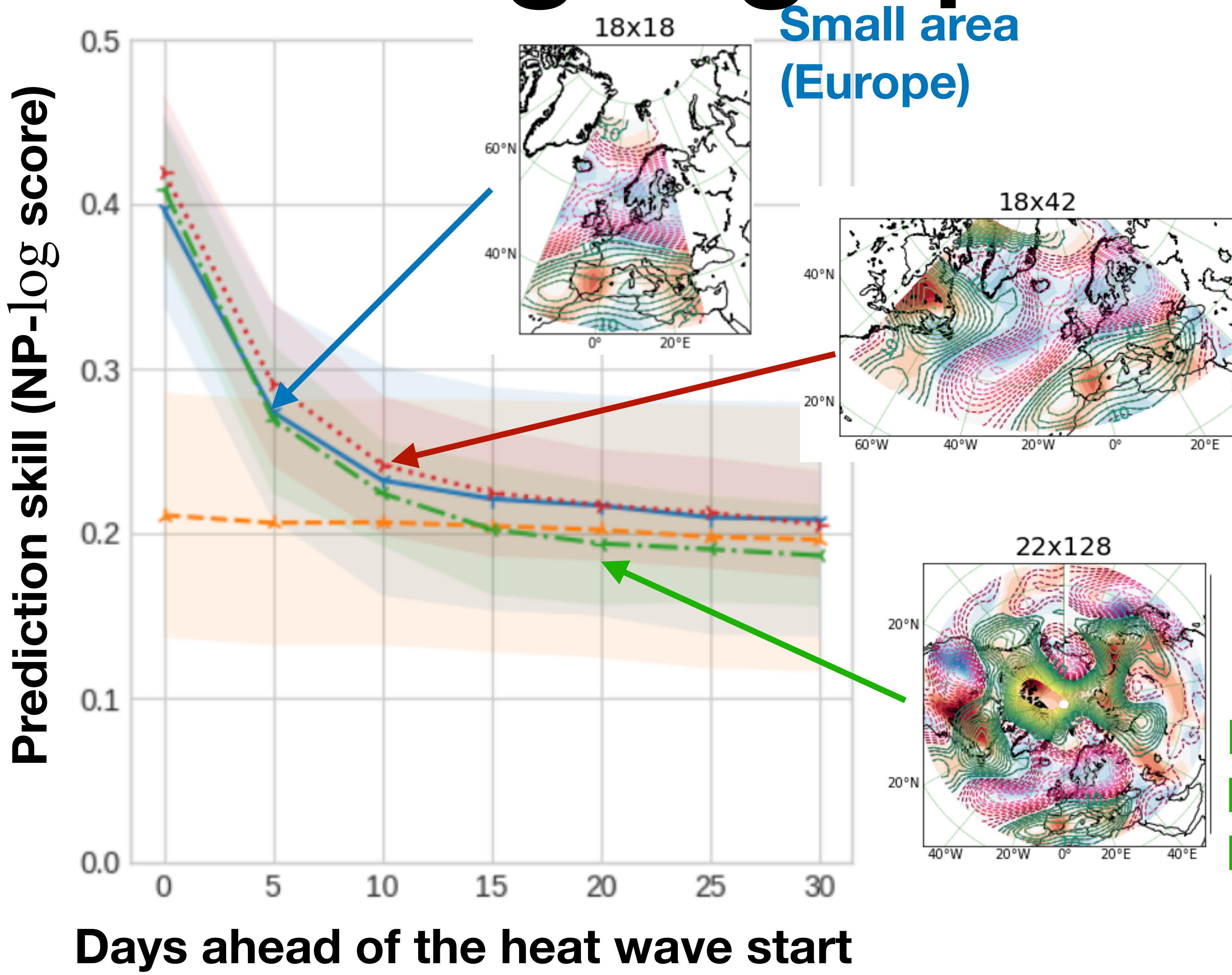
CNN classification of 2-week heatwaves τ days ahead of time

a_5 . The input consists of (t2m,zg500,mrso) with resolution (18,42,3). filtering applied to mrso (no part of the map is set to zero. We were using out-of-the-box CNN



The observation dataset is way too small for good machine learning prediction

Which is the optimal dataset geographical area?



CNN classification of 2-week heatwaves τ days ahead of time

Intermediate size area (Europe+North Atlantic)

Large area (North hemisphere mid-latitudes)

The best performance is obtained for an area of an intermediate size.
This also points to a regime of lack of data for optimal learning.

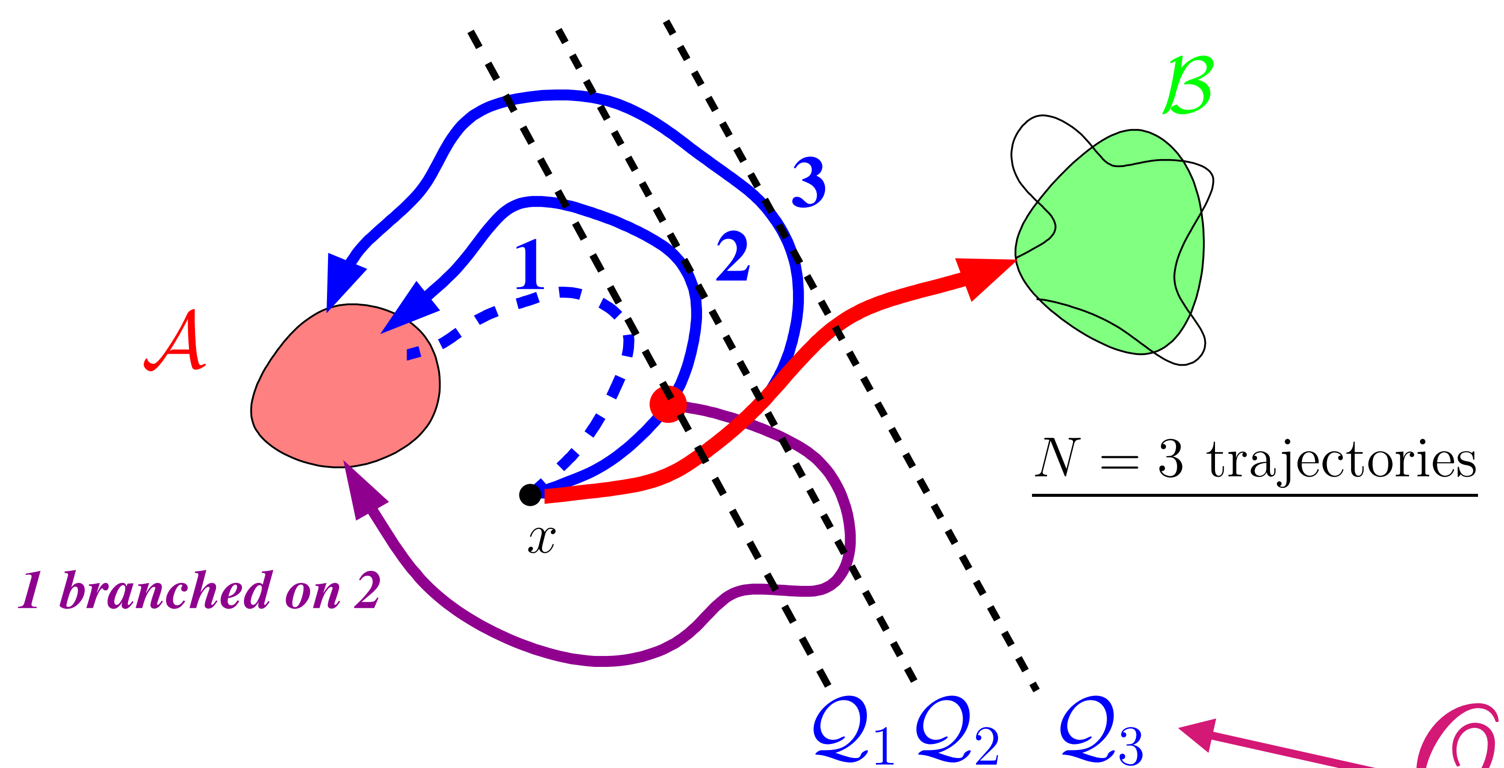
Conclusions: predicting heat waves with deep neural networks

- Prediction of heat waves is an example of a probabilistic classification problem.
- We use off-the-shelf CNN algorithms, adapted to this situation (probabilistic scores, undersampling, transfer learning).
- Two-week heat waves can be efficiently predicted up to 15 days ahead.
- We are clearly in a regime of lack of data for an optimal prediction.

IV) Coupling Rare Event Algorithms with Machine Learning

The Adaptive Multilevel Splitting (AMS) rare event algorithm

Strategy: selection, pruning and cloning.



Probability estimate:

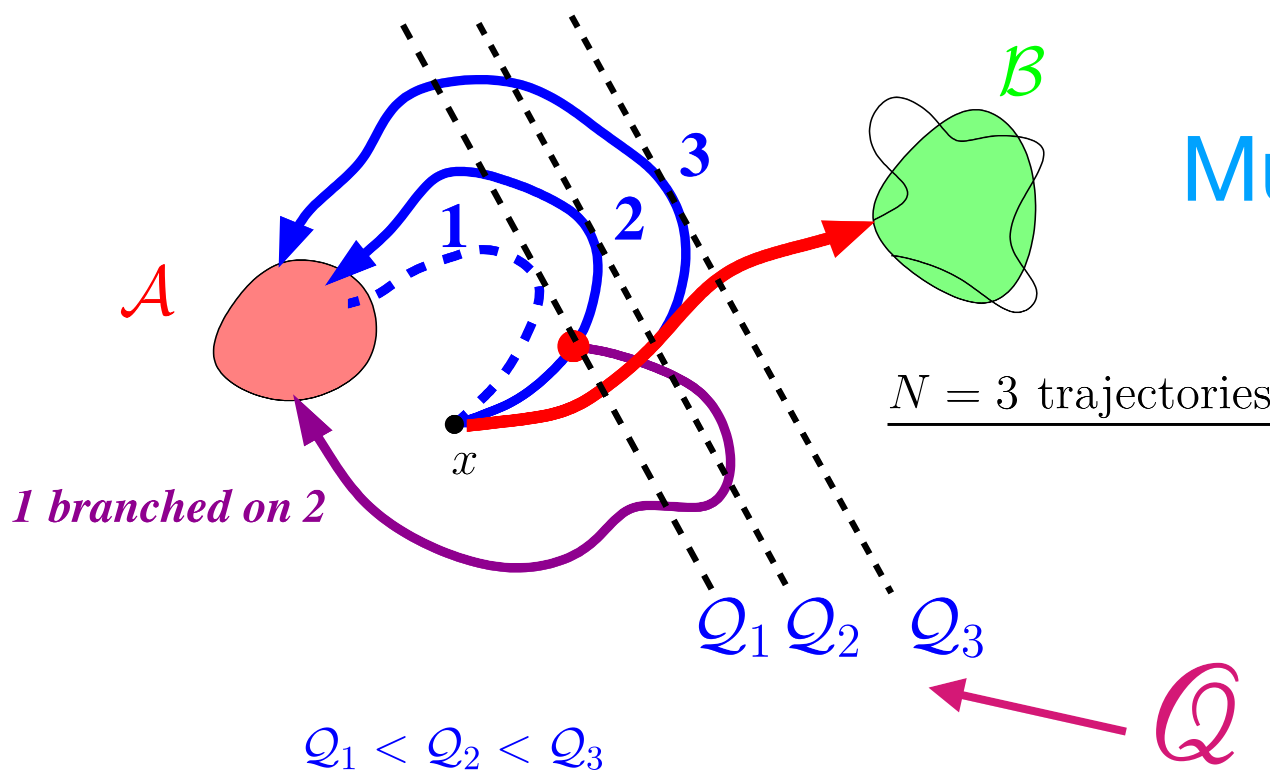
$\hat{p} = (1 - 1/N)^K$,

where N is the clone number and K is the iteration number.

$\mathcal{Q}$ is the score (sélection) function

Cérou, Guyader (2007). Cérou, Guyader, Lelièvre, and Pommier (2011).
PDEs: Rolland, Bouchet et Simonnet (2016) - **TAMS:** Lestang et al (2018)
Atmosphere turbulent jets: Rolland, Bouchet et Simonnet (2019 and 2021).

Committor functions are optimal score functions for rare event algorithms



The Adaptive Multilevel Splitting algorithm

$\mathcal{Q}$ is the score (sélection) function

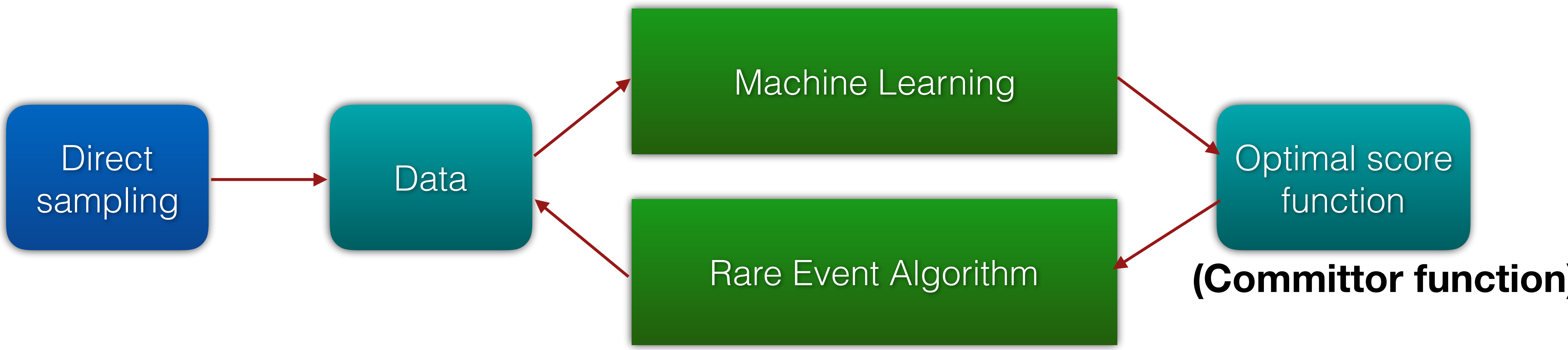
The efficiency of the algorithm depends on the choice of the score function.

The optimal score function is the committor function.

The score function is the key practical problem

- With a poor score function, rare event algorithms are useless.
- How to build good score functions?
- Running a rare event algorithm !!
- One needs one algorithm to improve the algorithm efficiency: use an adaptive strategy.
- Examples: Wang Landau algorithm, multi canonical methods, adaptive importance sampling, etc.

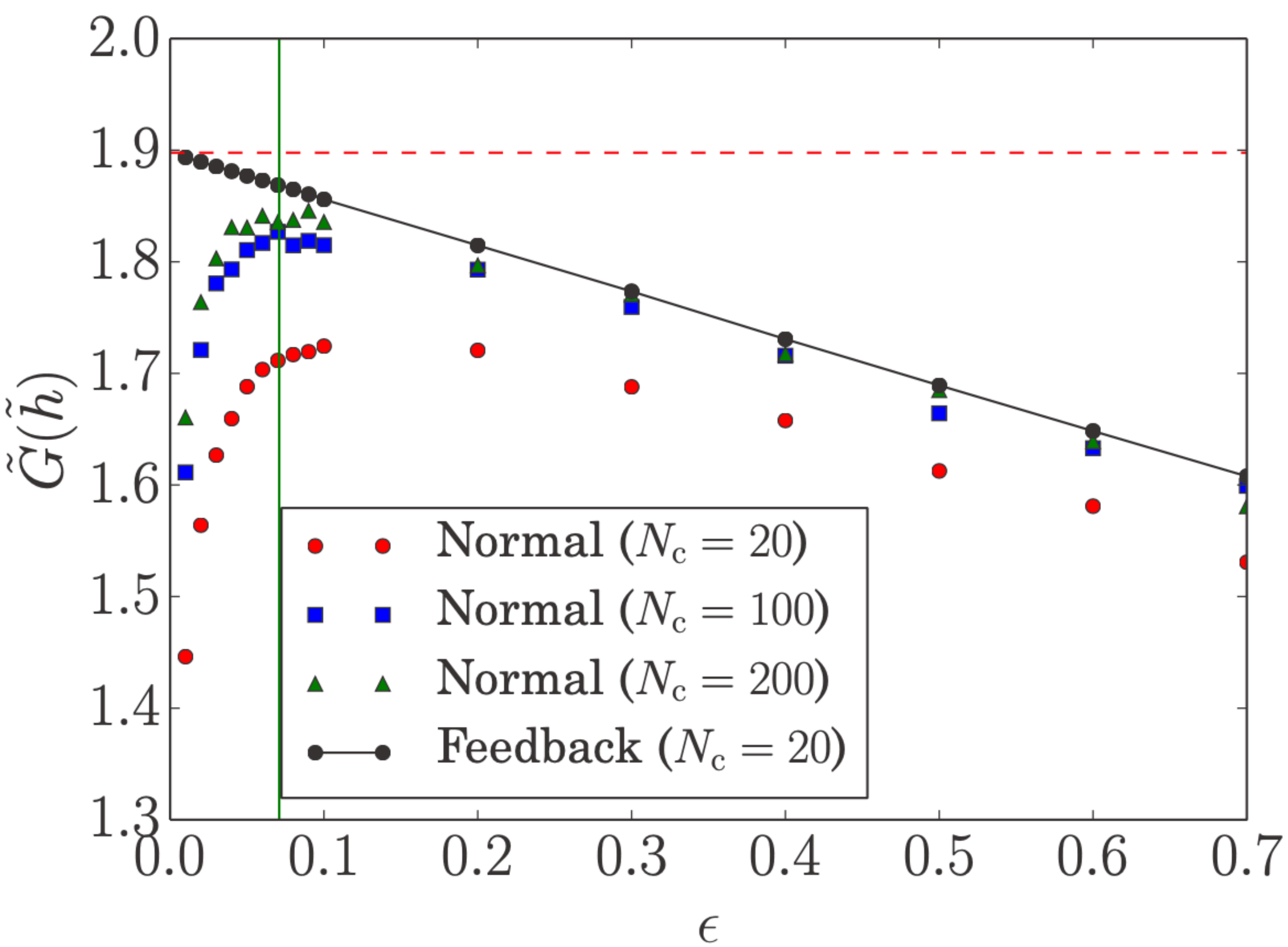
Coupling rare event algorithms with data based learning of committor functions

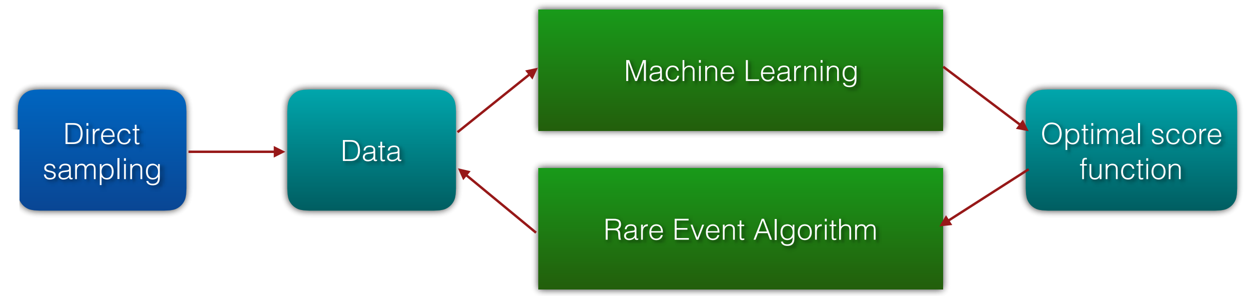


One example: Bouchet, Jack, Lecomte, Nemoto, PRE, 2016

(For X in dimension 1)

Coupling rare event algorithms with data-based learning of committor functions





$$\mathrm{d}X_t = -X_t^3 \mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}W_t,$$

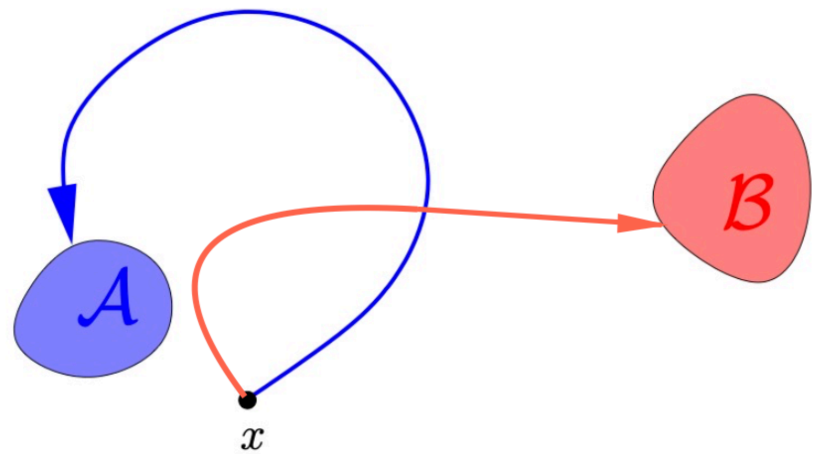
with X_t in dimension 1.

$$G(h) = \lim_{T\to+\infty} \frac{1}{T} \log \mathbb{E}\left[\exp\left(\int_0^T \lambda(X_t)\mathrm{d}t\right)\right]$$

with $\lambda(x) = x(x+1)$

Bouchet, Jack, Lecomte, Nemoto, PRE, 2016

Committor function



- $\{X(t)\}_{-\infty \leq t < +\infty}$ is a Markov process. A, B are subsets of the phase space.

- For a given sample path $\{X(t)\}_{-\infty \leq t < +\infty}$, the first hitting time τ_A is $\tau_A = \inf\{t \,|\, X(t) \in A\}$.

- The committor function *q(x)* of the sets A and B is defined as the probability that a trajectory starting at the point *x* reaches the set *B* before the set *A*

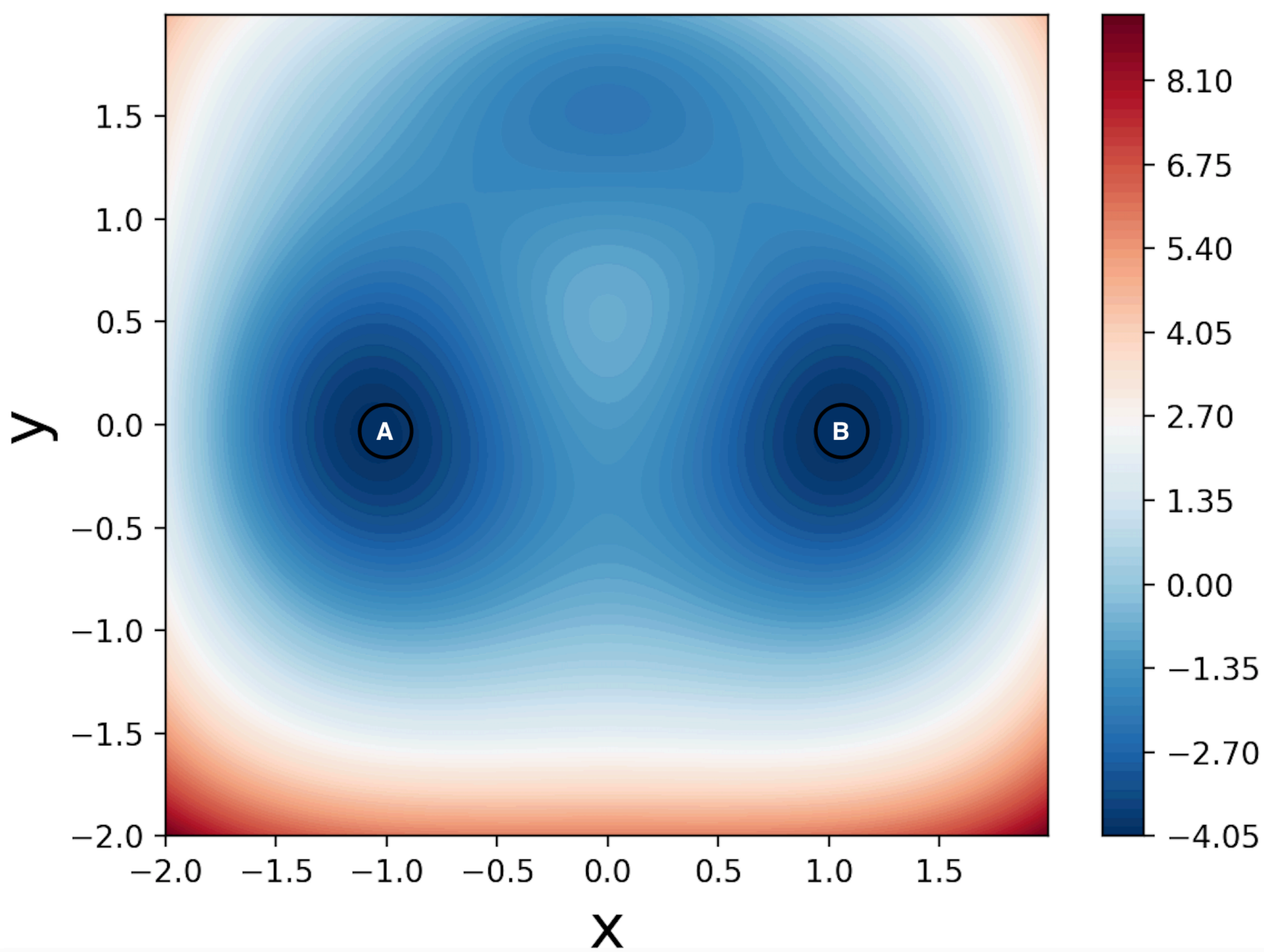
$$q(x) = \mathbb{P}_x\left(\tau_B < \tau_A\right).$$

- How to estimate the committor function? With a rare event algorithm!

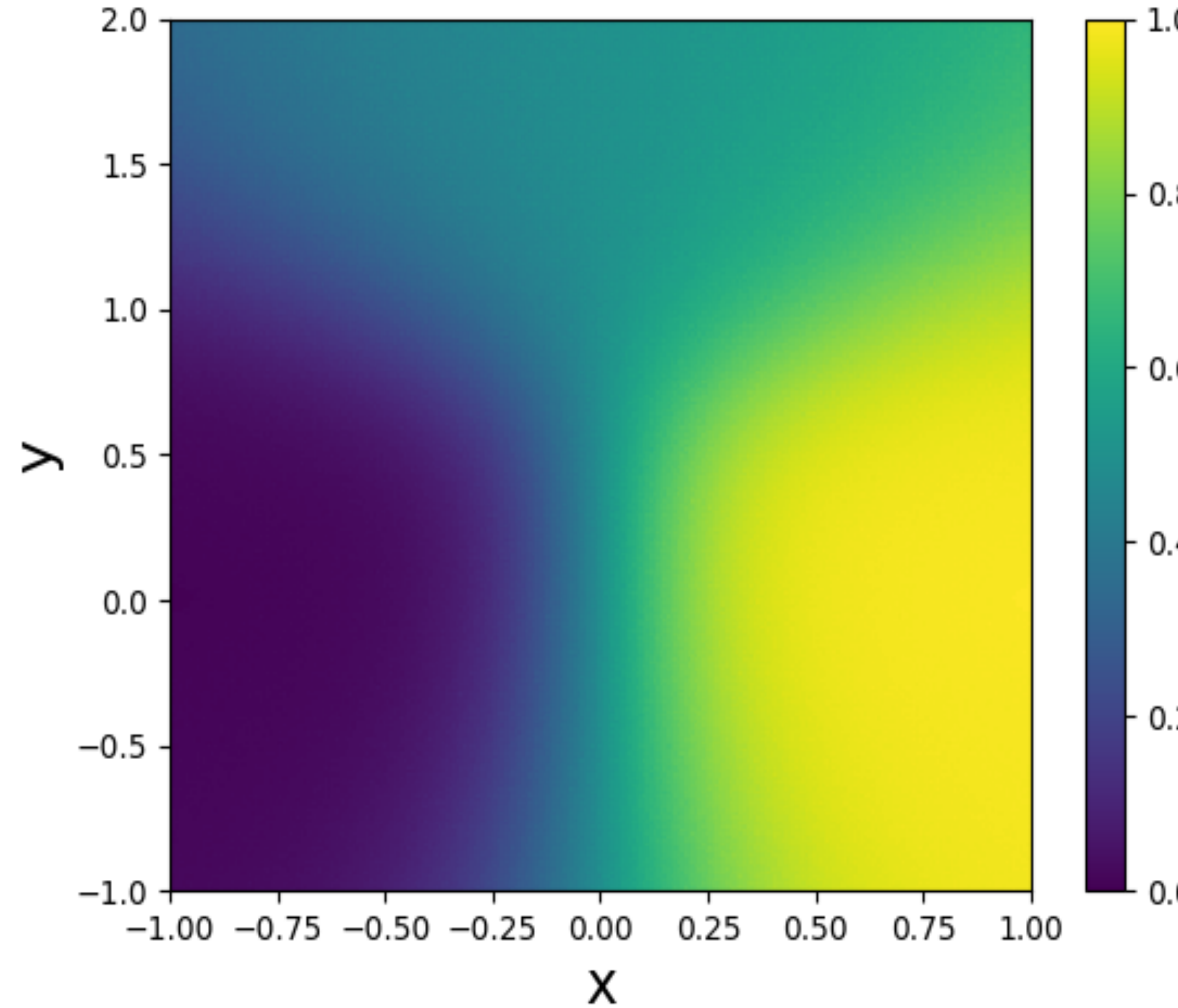
Committor function for a simple gradient dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}W_t \quad \text{and} \quad X = (x, y)$$

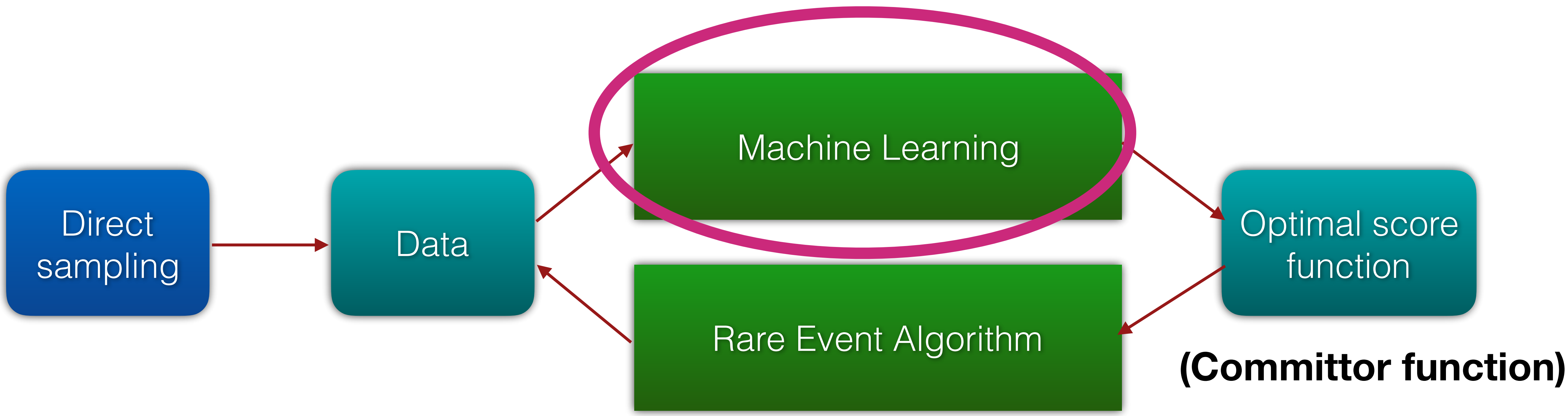
Potential V



Committor function q



Coupling rare event algorithms with data based learning of committor functions



Data-based learning of a committor function

$$q(x) = \mathbb{P}_x\left(\tau_B < \tau_A\right).$$

- Our dataset is a trajectory $\{X_n = X(t_n)\}_{1\leq n\leq N}$ (or a set of trajectories)

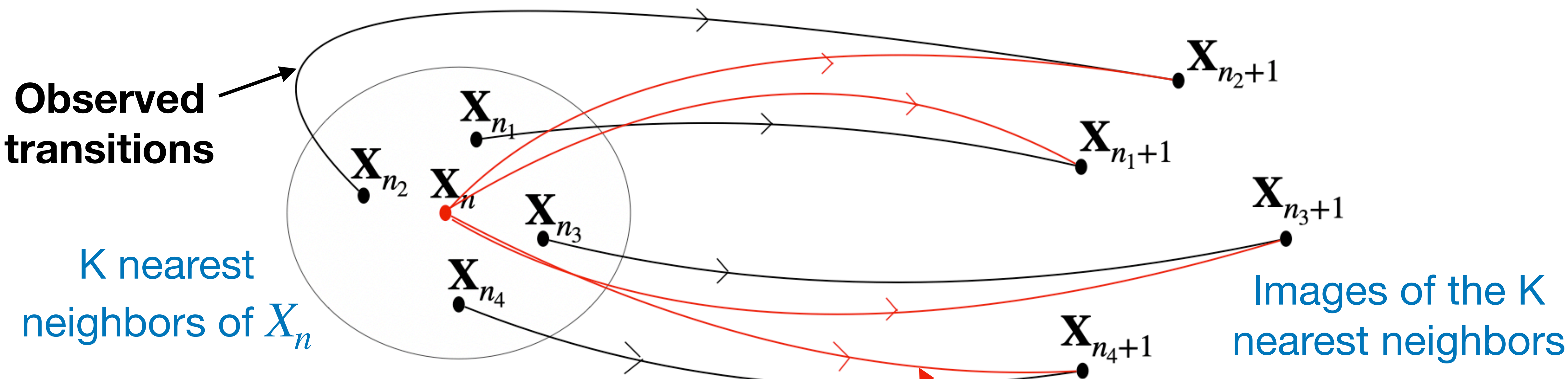
- Approach 1: direct learning as a probabilistic classification problem. We set $Y_n = 1$ if $\tau_B < \tau_A$, $Y_n = 0$ otherwise. Then we have a probabilistic classification problem where X are the data and Y the classes (see previous discussion using CNN).

- Approach 2: learn first a Markov chain from the data and then compute the committor function.

The analogue Markov chain

- We learn an approximate Markov chain on the set of the observed states $\{X_n\}_{1\leq\, n\leq N}$.

- How to learn from the data an approximate transition probability from an observed state X_n to one of the other observed states ?



$P(X_n, X_m) = 1/K$ if X_m is one of the images of the nearest neighbors, 0 otherwise

Transitions in the approximate Markov chain

Estimating committor functions from the analogue Markov chain

- G is the propagator of the analogue Markov chain:

$$\begin{cases} G_{nj} = \frac{1}{K} \text{ if } j-1 \text{ is one of the analogues of } n\,, \\ G_{nj} = 0 \text{ otherwise}\,. \end{cases}$$

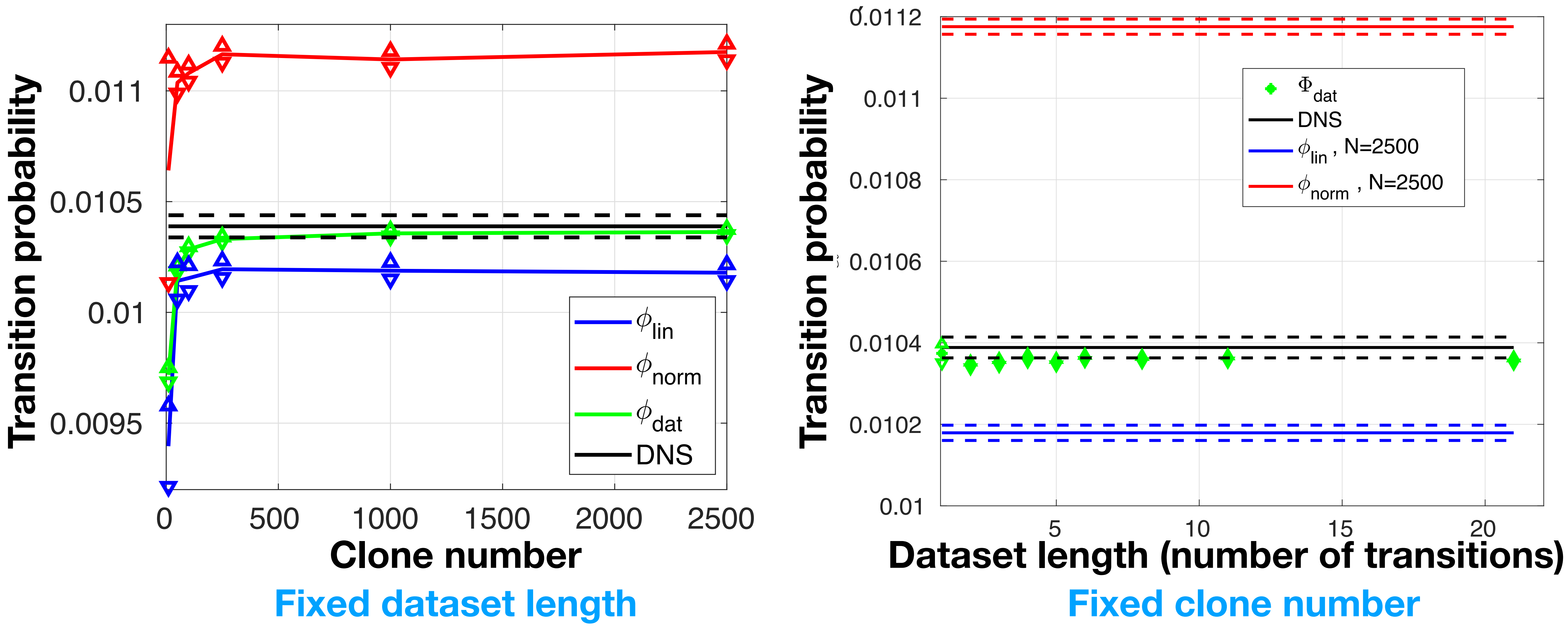
- q is the solution of the linear problem:

$$q_n = \sum_j G_{nj} q_j \text{ if } n \in (A \cup B)^c,\ q_n = 1 \text{ if } n \in B\ , q_n = 0 \text{ if } n \in A\,.$$

We can estimate the committor function, based on data using dynamical informations.

With the learned score function the AMS algorithm is extremely efficient

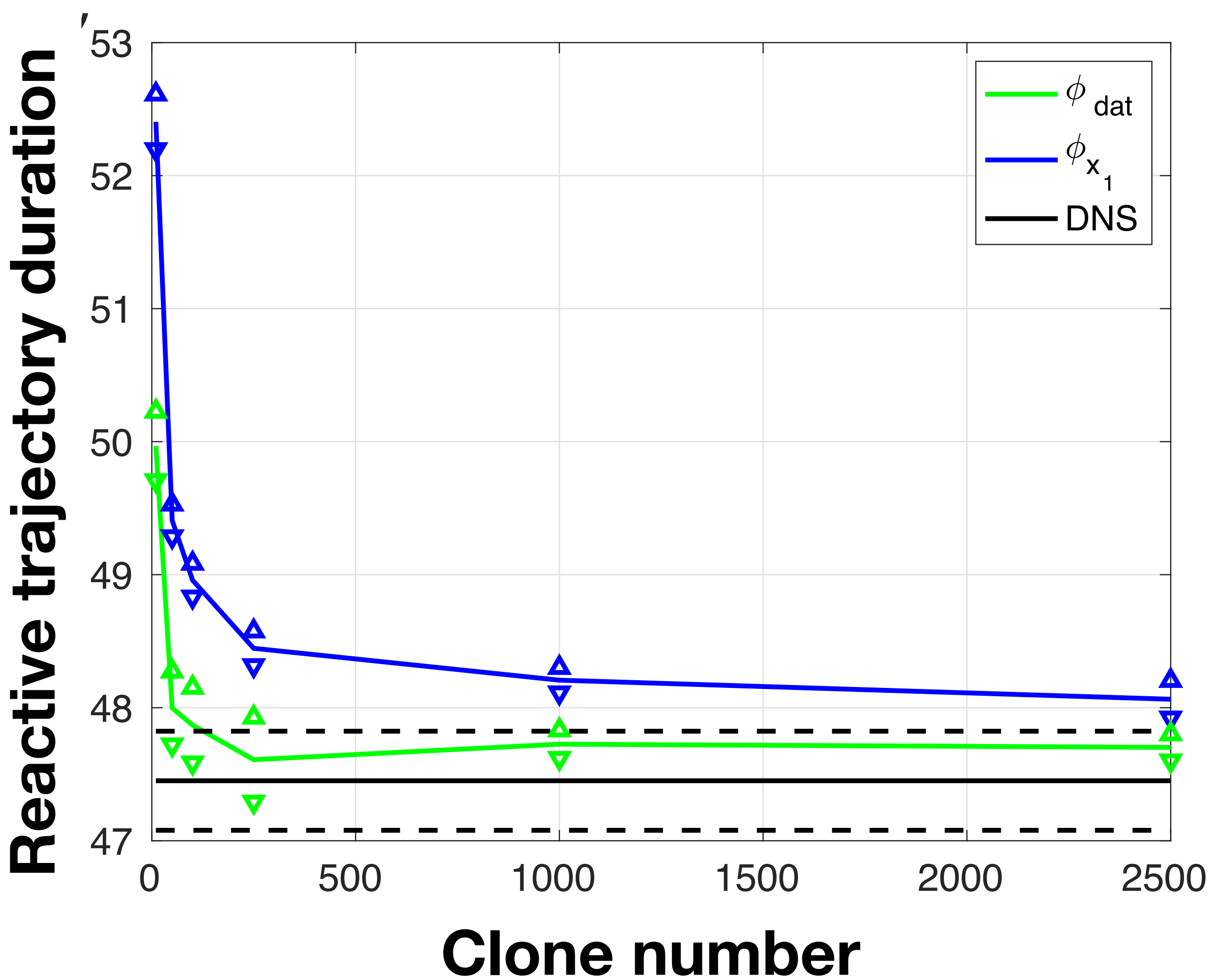
The three-well gradient dynamics: $\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}W_t$ and $X = (x, y)$



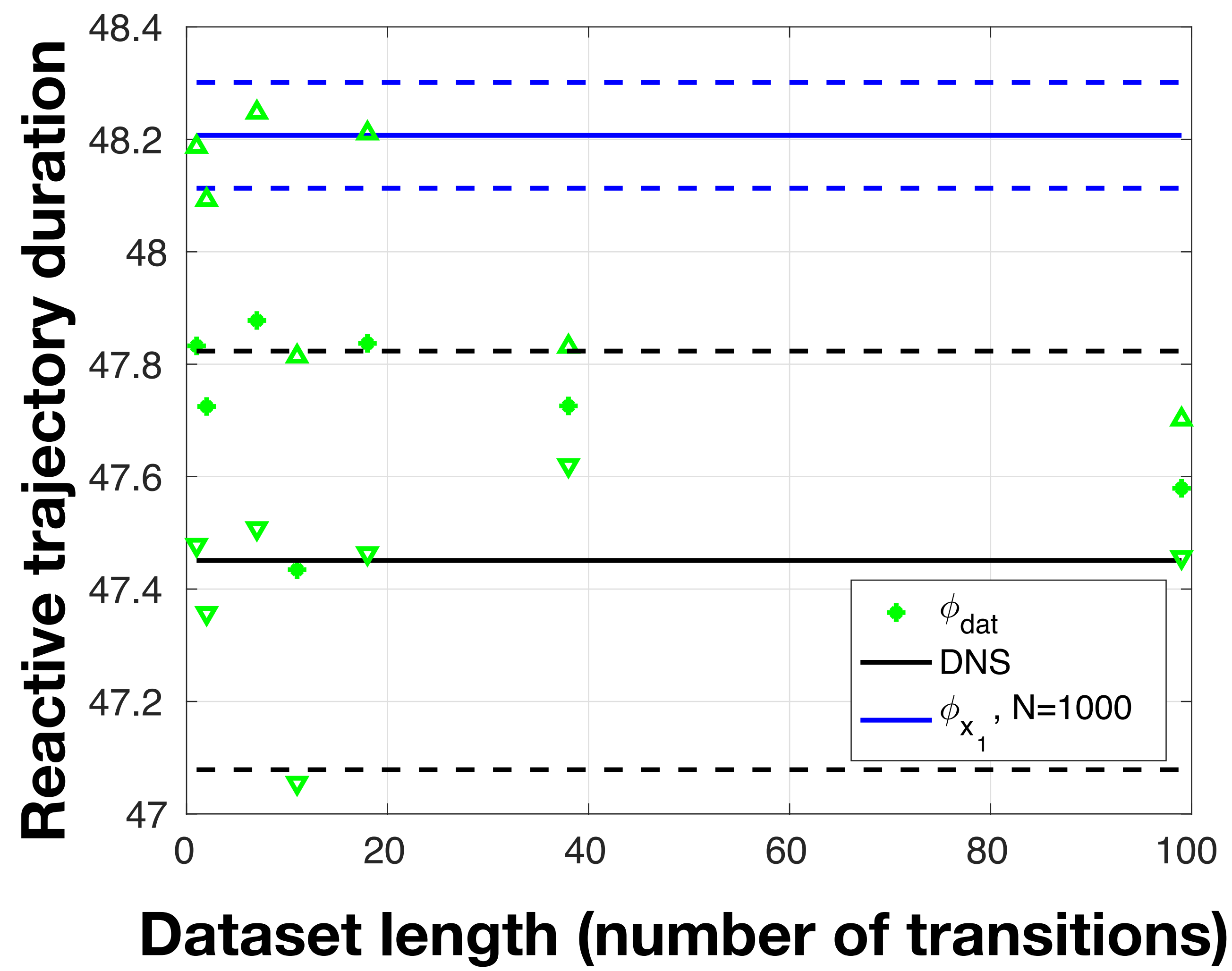
We have an impressive efficiency when learning from a dataset **with only a few observed transitions**

With the learned score function the AMS algorithm is extremely efficient

The Charney—DeVore model (6 dimension chaotic dynamics + noise)



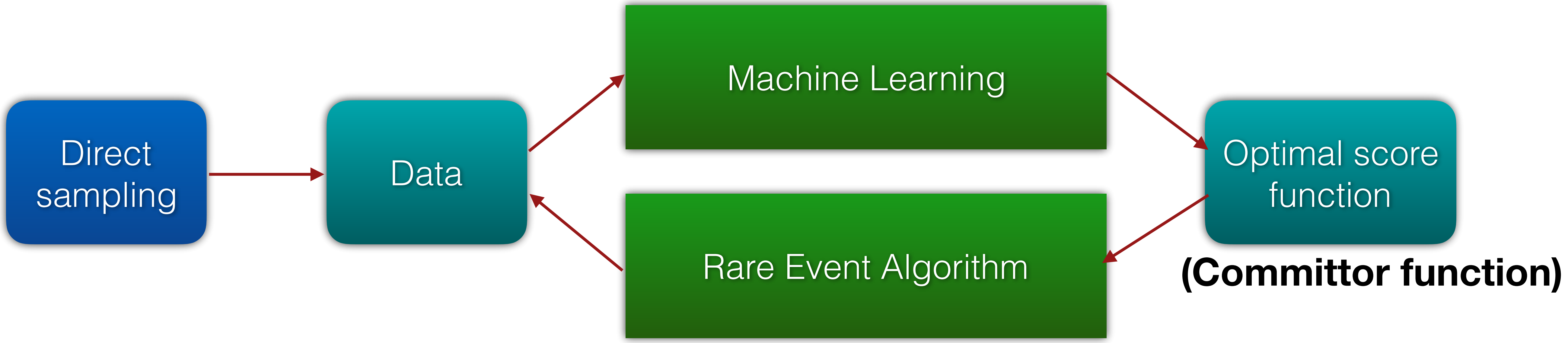
Fixed dataset length



Fixed clone number

We have an impressive efficiency when learning from a dataset **with about 40 to 100 observed transitions**

Coupling rare event algorithms with data based learning of committor functions



Work in progress for climate models!

Conclusion: Coupling machine learning with rare event algorithms

- We can learn committor functions from dynamical datasets either using the definition, or first learning an approximate Markov dynamics.

- The analogue Markov chain does not require an impossible discretization of the phase space, and can use any kind of dynamical data, including short trajectories.

- Using learned committor functions is much more efficient than using user-defined score functions with the AMS rare event algorithm.

- The range of applicability of this approach, in terms of system dimension and complexity, is a key question for the future.

D. Lucente, J. Rolland, C. Herbert and F. Bouchet, to be posted on ArXiv this week

II) Abrupt climate changes

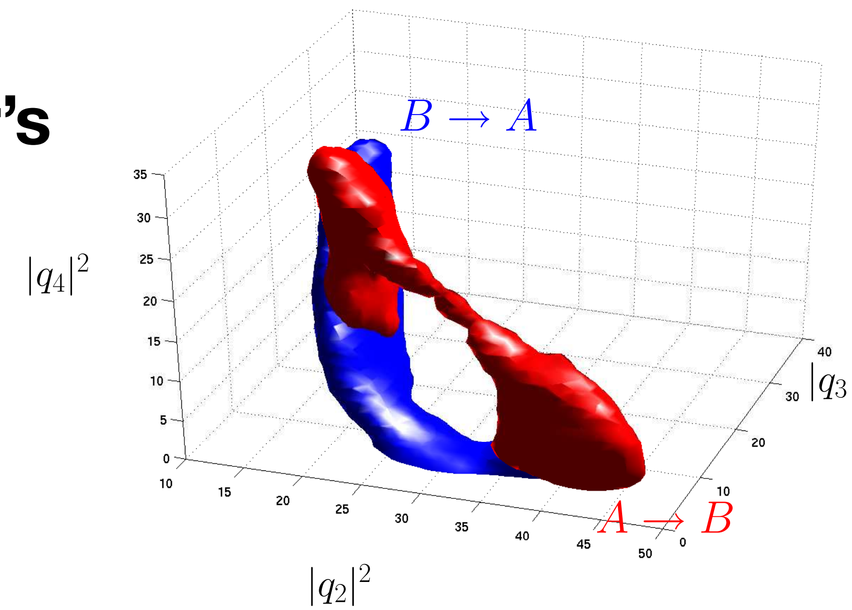
- **II a) Instantons and Arrhenius law for Jupiter's troposphere abrupt transitions.**

F. Bouchet, J. Rolland, and E. Simonnet, PRL, 2019: instantons and Arrhenius law.

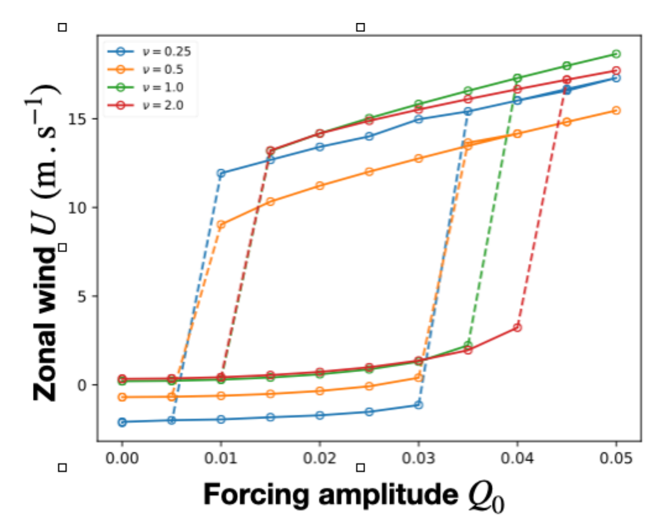
F. Bouchet, J. Rolland, and E. Simonnet, JAS, 2021: multistability with symmetry breaking and instantons.

- **II b) Discontinuous transitions to superrotation on planetary atmospheres.**

C. Herbert, R. Caballero and F. Bouchet, JAS, 2019: study of abrupt transitions, negative feedbacks and the robustness of the bistability range.



Atmosphere jet instantons



Superrotating atmosphere hysteresis

V) Large deviation theory and other applications (more theoretical works)

- **III a) Path large deviations for kinetic theories.**

F. Bouchet, J. Stat. Phys., 2020: path large deviations for the Boltzmann equation and the irreversibility paradox.

O. Feliachi and F. Bouchet, sub. to J. Stat. Phys., 2020: path large deviations for the plasma and the Vlasov equation.

- **III b) Rare events for the Solar System (planet collisions).**

F. Bouchet and E. Woillez, PRL, 2020.

RESEARCH HIGHLIGHTS

Nature Reviews Physics | **The path to the Solar system's destabilization**

16 July 2020

GDR « Theoretical challenges for climate sciences »

- **Identify and work on key theoretical issues that need to be solved for improving the quantitative predictions in climate sciences.**

- **A multidisciplinary consortium: climate sciences, mathematics, physics, computer sciences, statistical physics, data sciences.**

- **Examples :** i) How to reduce the uncertainty about climate sensitivity? ii) How to reduce uncertainty when quantifying probabilities of climate extreme events? iii) How to integrate data and theoretical constraints, using machine learning, to build the next generation of climate models? iv) How to make quantitative the study of future and past climate? v) How to build effective coarse-grained descriptions of climate processes?

Conclusions

- We can use **rare event algorithms** to gather an amazing statistics for extreme heat waves with Plasim (PNAS, 2018), and CESM (GRL, 2021).

- The dynamical mechanism is a **quasi-stationary non zonal global patterns,** which are much affected by topography and oceans (PNAS, 2018, GRL 2021).

- **Models reproduce correctly those extreme teleconnection patterns** for moderate extremes. **We need model and rare event algorithms to study more extreme heat waves teleconnection patterns.**

- **Machine learning** has the potential to give **meaningful statistical predictions for long-lasting heat wave up to 2 weeks ahead of time** (Sub. to IEEE TPAMI, 2021).

- **The coupling of learned committor functions with rare-event algorithms is extremely efficient for toy model** (Arxiv, 09-2021). Work in progress for climate models.

Please join us to study climate extreme events
The scientific questions are fascinating!