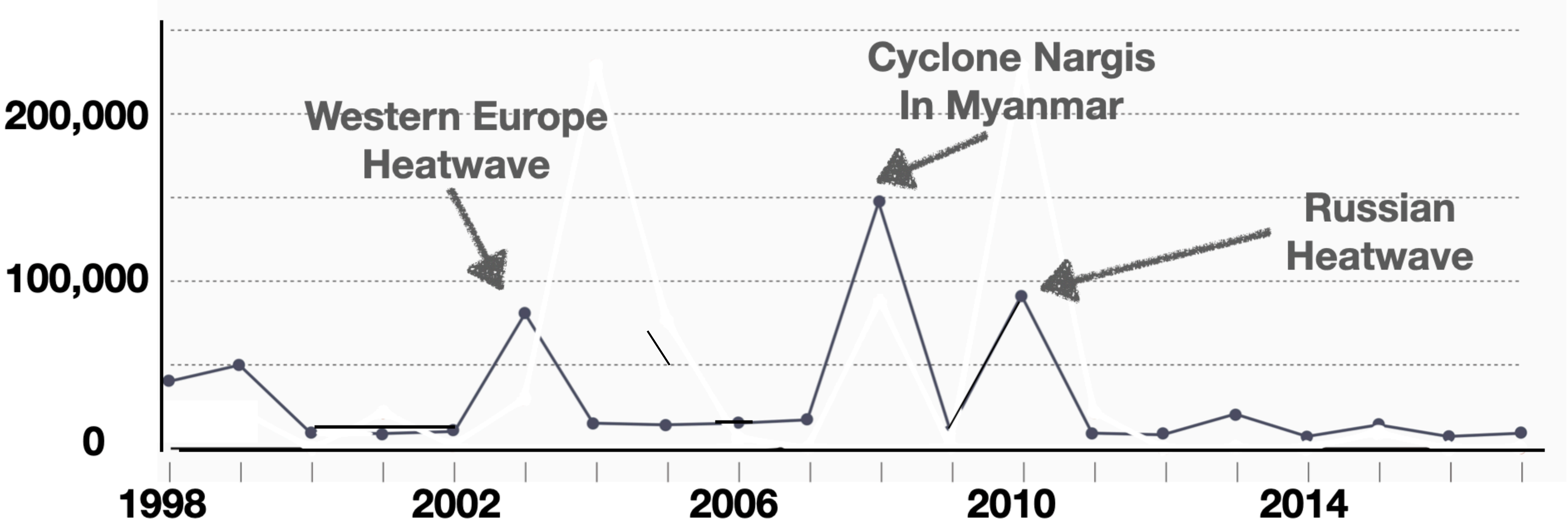
ADVANCES IN RARE EVENT SIMULATIONS USING DATA-BASED ESTIMATION OF COMMITTOR FUNCTIONS

D. Lucente, J. Rolland, C. Herbert and F. Bouchet

EGU General Assembly 2022 — 23 May 2022
Session NP2.2 — Extremes in geophysical sciences: drivers, methods and impacts quantification

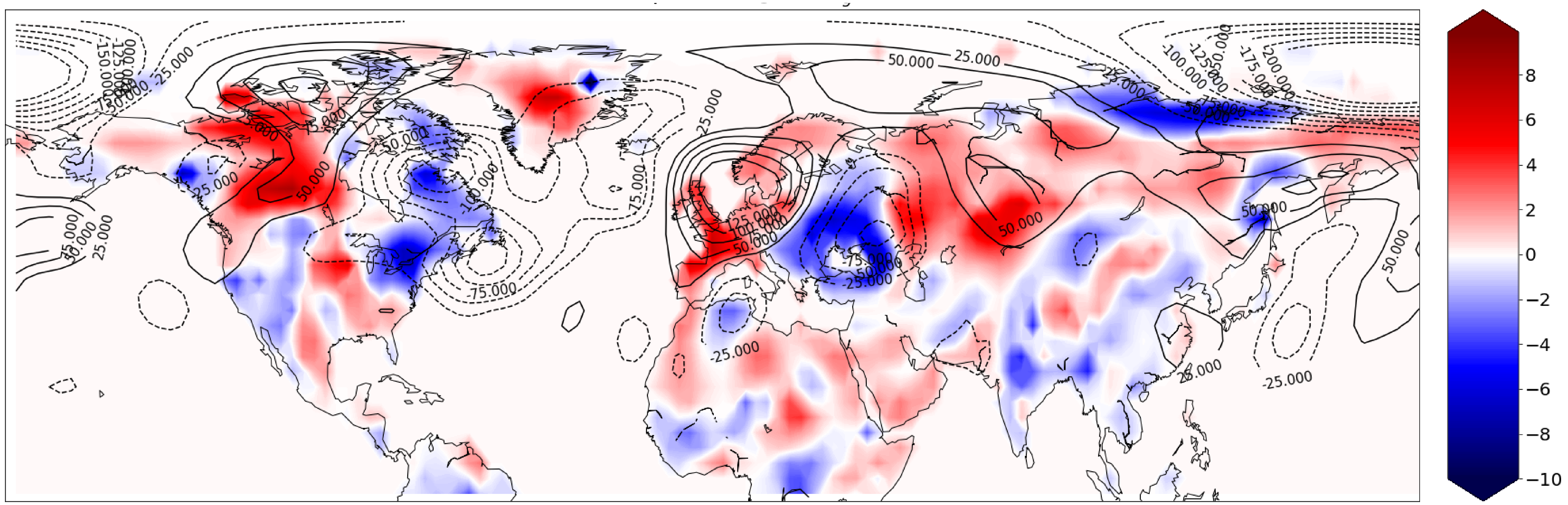
WHY IS IT IMPORTANT TO STUDY EXTREME CLIMATE EVENTS?

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



The extreme climate events have a huge impact.

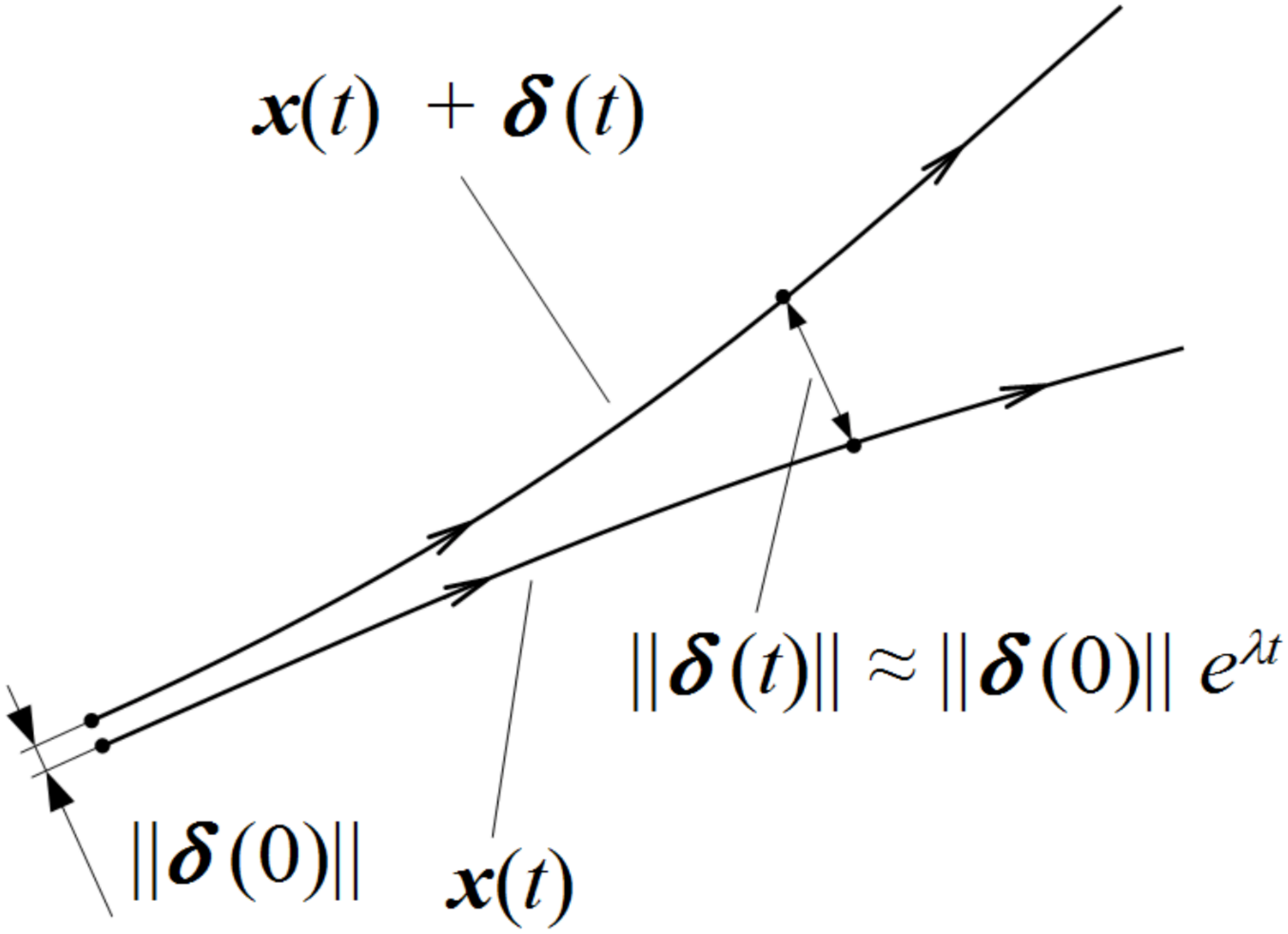
PREDICTING HEAT WAVES



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

- $X(t) = T_s$ **field at time** t**, or** $X(t) = (T_s, Z_g)$ **fields at time** t**.**
- $A(t_0)$**: time and space averaged surface temperature anomaly.**
- $Y(t_0) \in \{0,1\}$. $Y(t_0) = 1$ if $A(t_0) > a$, and $Y(t_0) = 0$ otherwise. **A heat wave occurs if** $Y = 1$.
- **Aim: Predict** $Y(t_0 + \tau)$ **given** $X(t_0)$ **(occurence of not of an heat waves within** τ **days).**

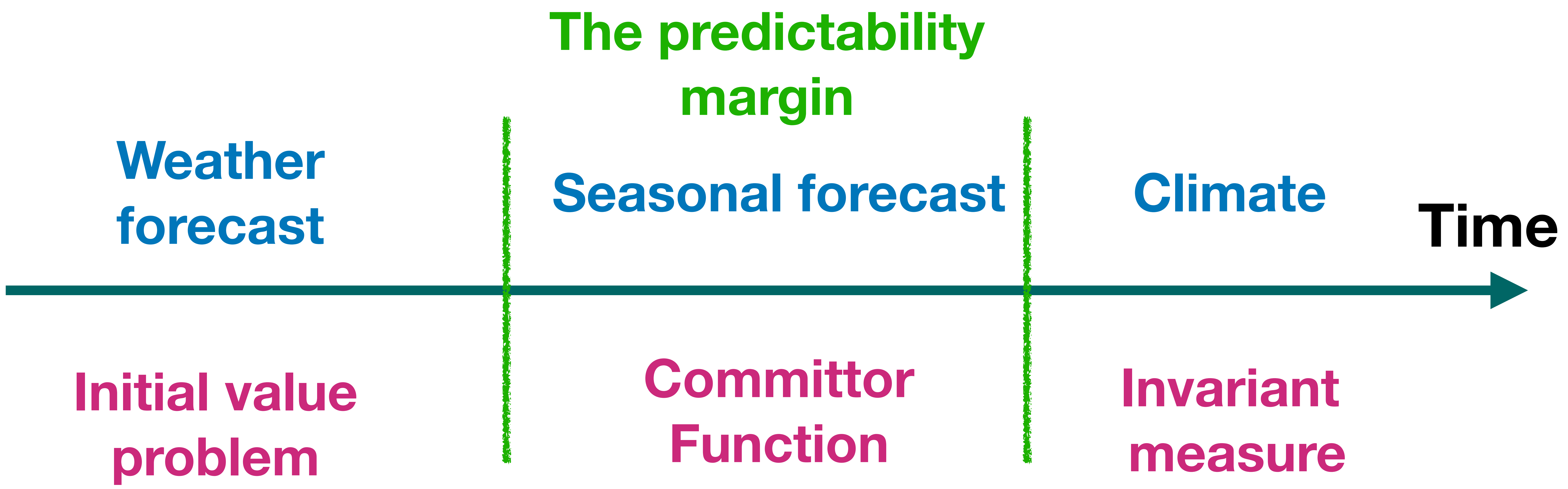
DETERMINISTIC PREDICTABILITY FOR A CHAOTIC DYNAMICAL SYSTEM



λ **Lyapunov Exponent**

λ^{-1} **Lyapunov Time: It is of the order of time needed to double the initial separation $\delta(0)$ of two trajectories**

WHAT IS THE MATHEMATICAL CONCEPT OF THE PREDICTABILITY MARGIN?



The relevant mathematical concept for prediction problems at the predictability margin is the **committor function**.

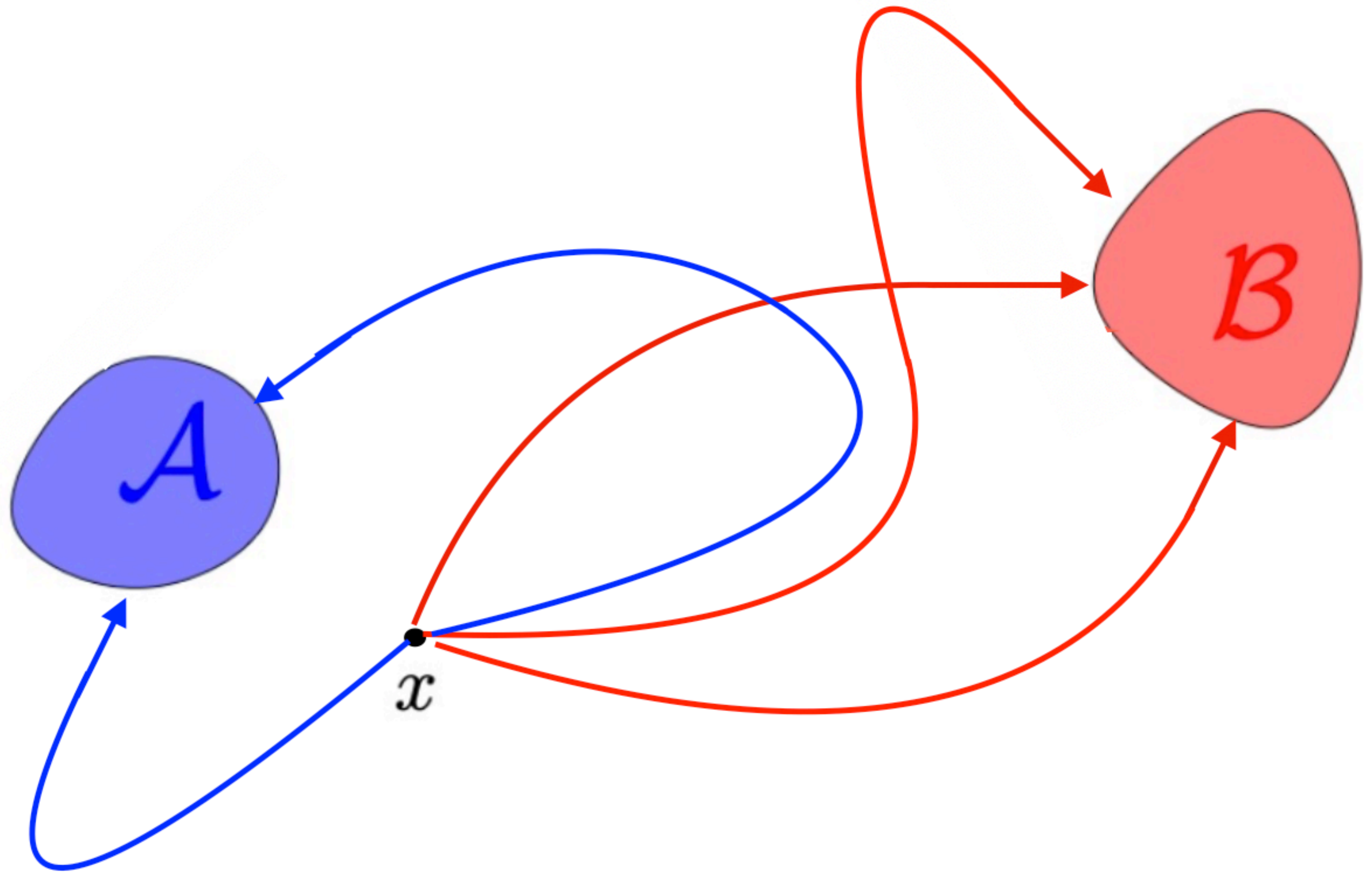
OUTLINE

- **Committor functions**
- **Computing committor functions with the analogue Markov chain**
- **Coupling rare events algorithms and learned committor functions**
- **Computing committor function for heat waves**

COMMITTOR FUNCTION

COMMITTOR FUNCTION

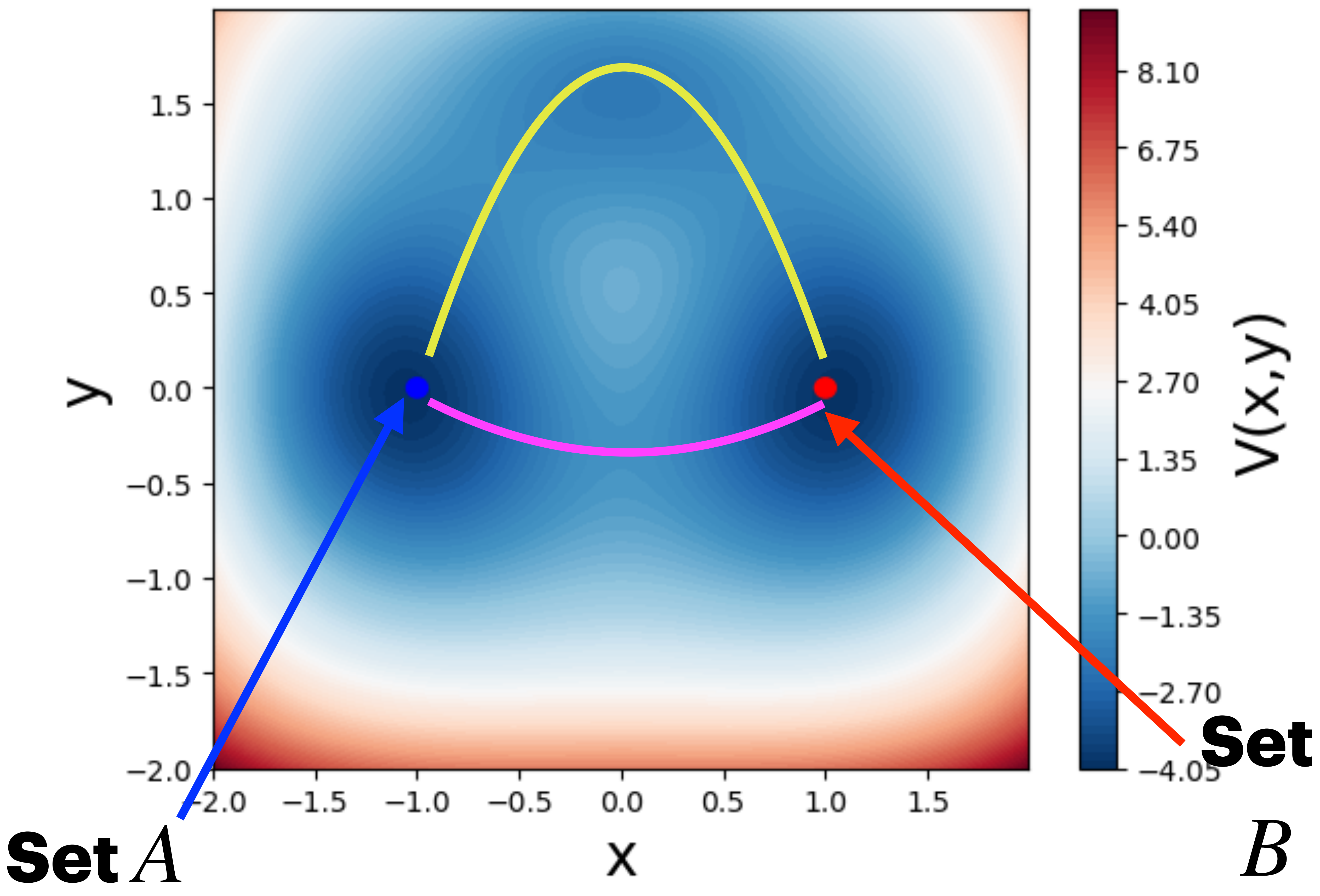
The committor function $q(x)$ of the sets A and B is defined as the probability that a trajectory starting at the point x at the time 0 reaches the set B before the set A.



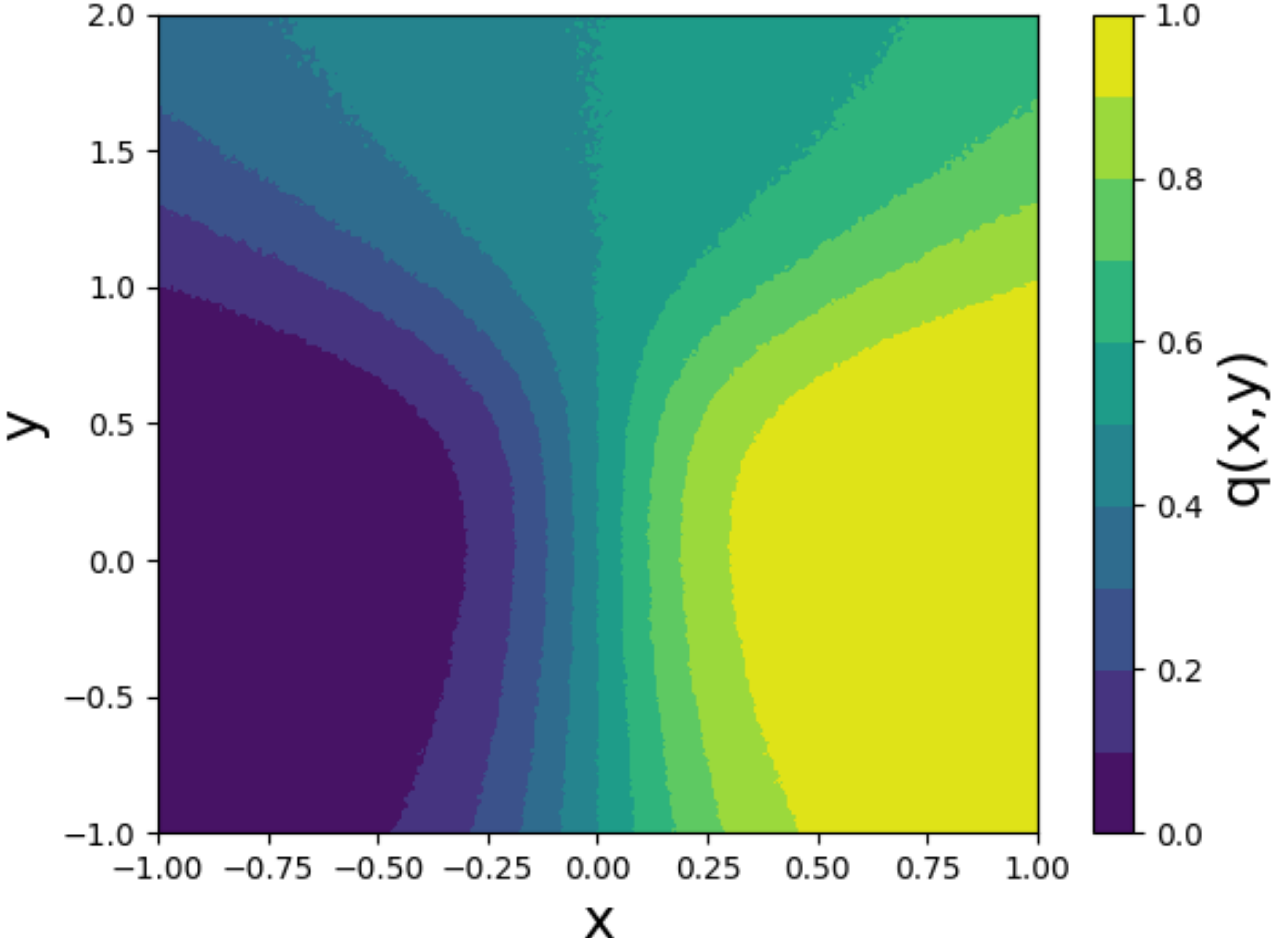
EXAMPLE: 2-DIMENSIONAL STOCHASTIC DIFFERENTIAL EQUATION

The three-well gradient dynamics: $dX_t = -\nabla_X V(X_t)dt + \sqrt{2D}dW_t$ **and** $X = (x, y)$

Potential $V(x, y)$



Committor Function $q(x, y)$



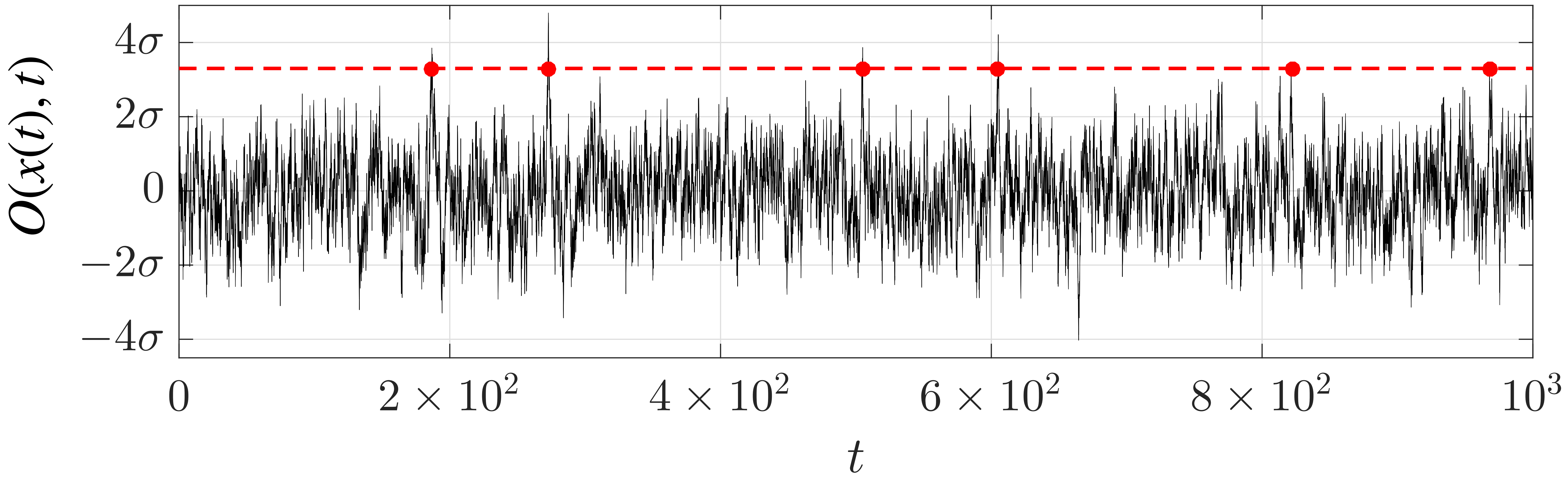
TIME-DEPENDENT COMMITTOR FUNCTION

- **The committor function can also be the probability that a given observable $O(x(t), t)$ becomes greater than a threshold ϵ before a certain lag time T.**

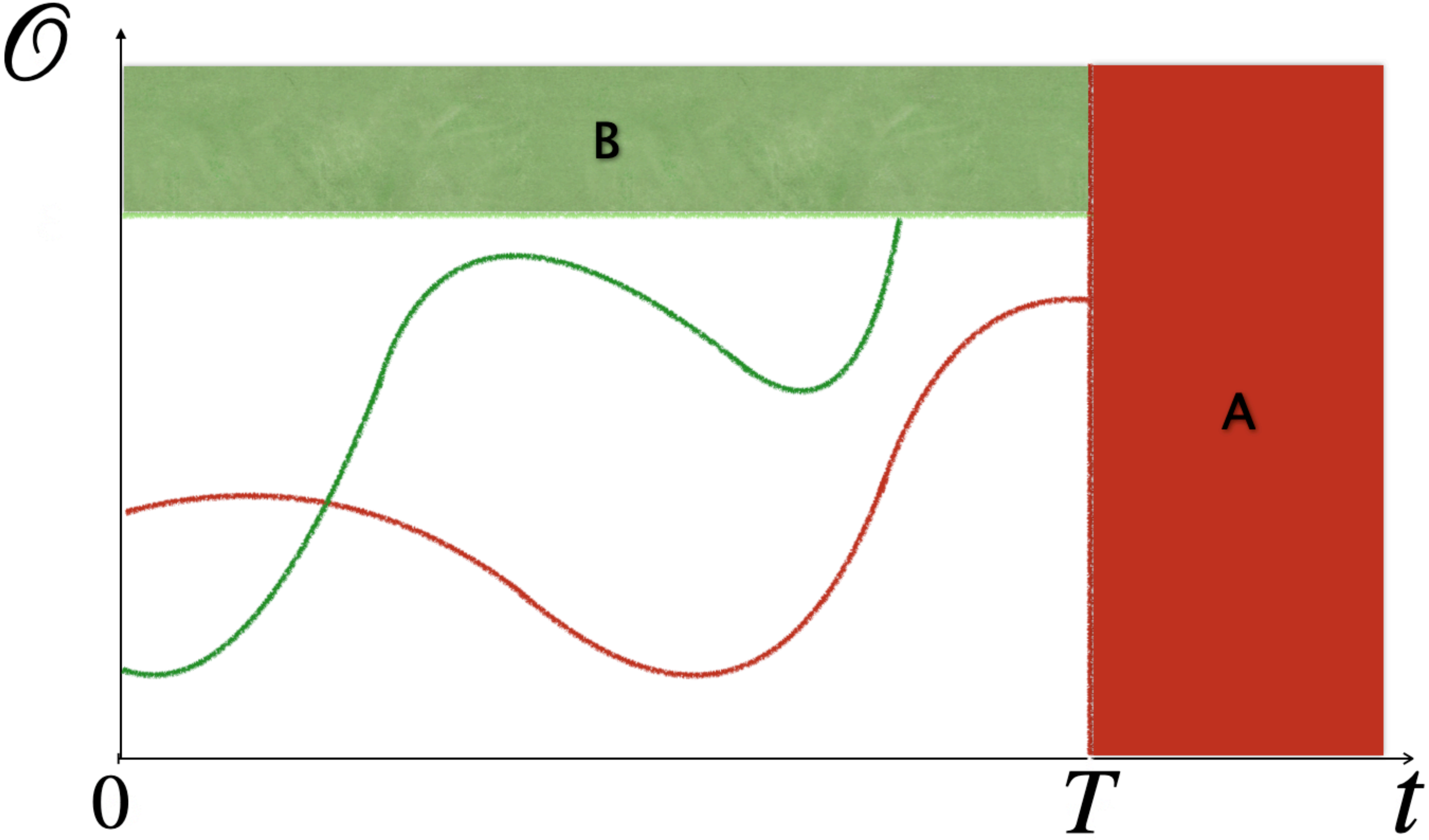
- **We can define an auxiliary process $Y = (O, t)$ such that**

$$B = \{(z, t) : 0 \leq t \leq T, z > \epsilon\}, \qquad A = \{(z, T) : z < \epsilon\}.$$

- $\mathbb{P}(max\{O(x(t), t)\} > \epsilon \,|\, x(0) = x) = \mathbb{P}(\tau_B < \tau_A) = q(x, T)$



PHASE SPACE OF THE AUXILIARY VARIABLE



DIRECT COMPUTATION OF COMMITTOR FUNCTION

- **Monte Carlo experiments.**

- **Ergodic process: the committor function $q(x)$ and the stationary distribution $\rho(x)$ can be computed from the formulas**

$$\rho(x)q(x) = \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} \mathrm{d}t\, \delta\left(X_t - x\right) 1_{\{\tau_B \le \tau_A\}}, \quad \rho(x) = \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} \mathrm{d}t\, \delta\left(X_t - x\right).$$

- **Stochastic process: $q(x)$ solves $L\left[q\right] = 0$, where L is the infinitesimal generator, with boundary conditions $q(x) = 1$ if $x \in B$ and $q(x) = 0$ if $x \in A$.**

THE ANALOGUE MARKOV CHAIN

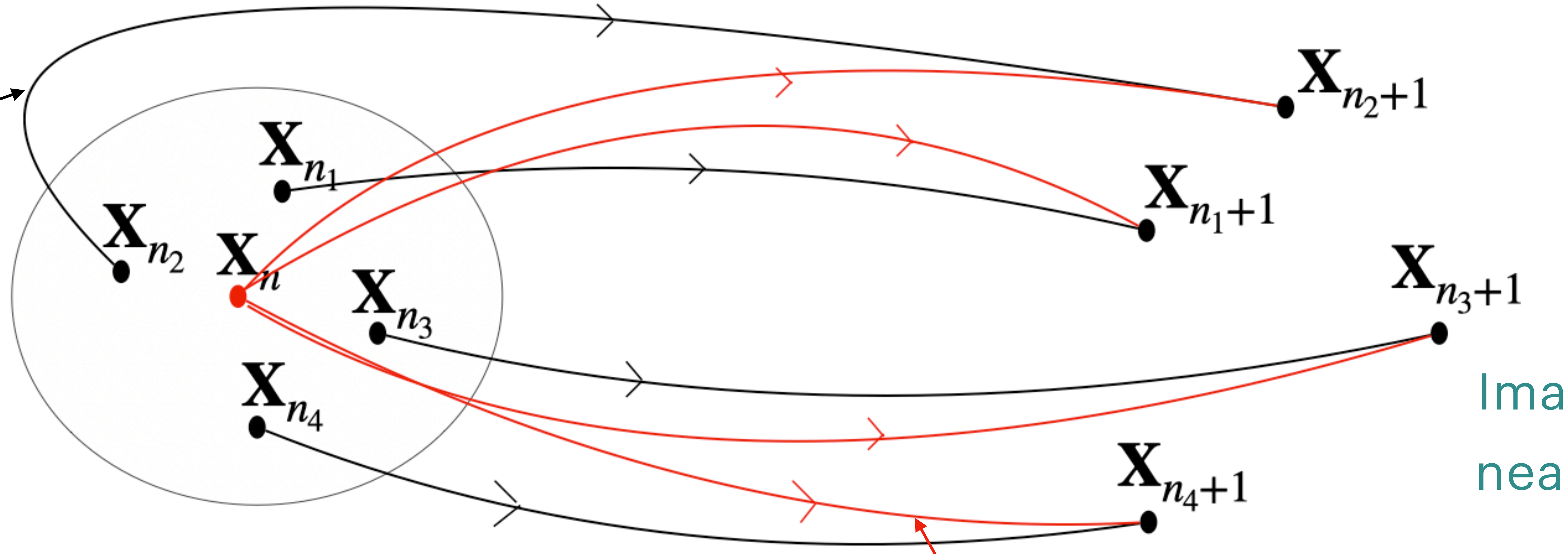
THE ANALOGUE MARKOV CHAIN

- **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 ?**

Observed transitions

K=4 nearest neighbors of X_n



Images of the K=4 nearest neighbors

Transitions in the approximate Markov chain

COMPUTING COMMITTOR FUNCTIONS FROM THE ANALOGUE MARKOV CHAIN

- G **is the propagator of the analogue Markov chain:**

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

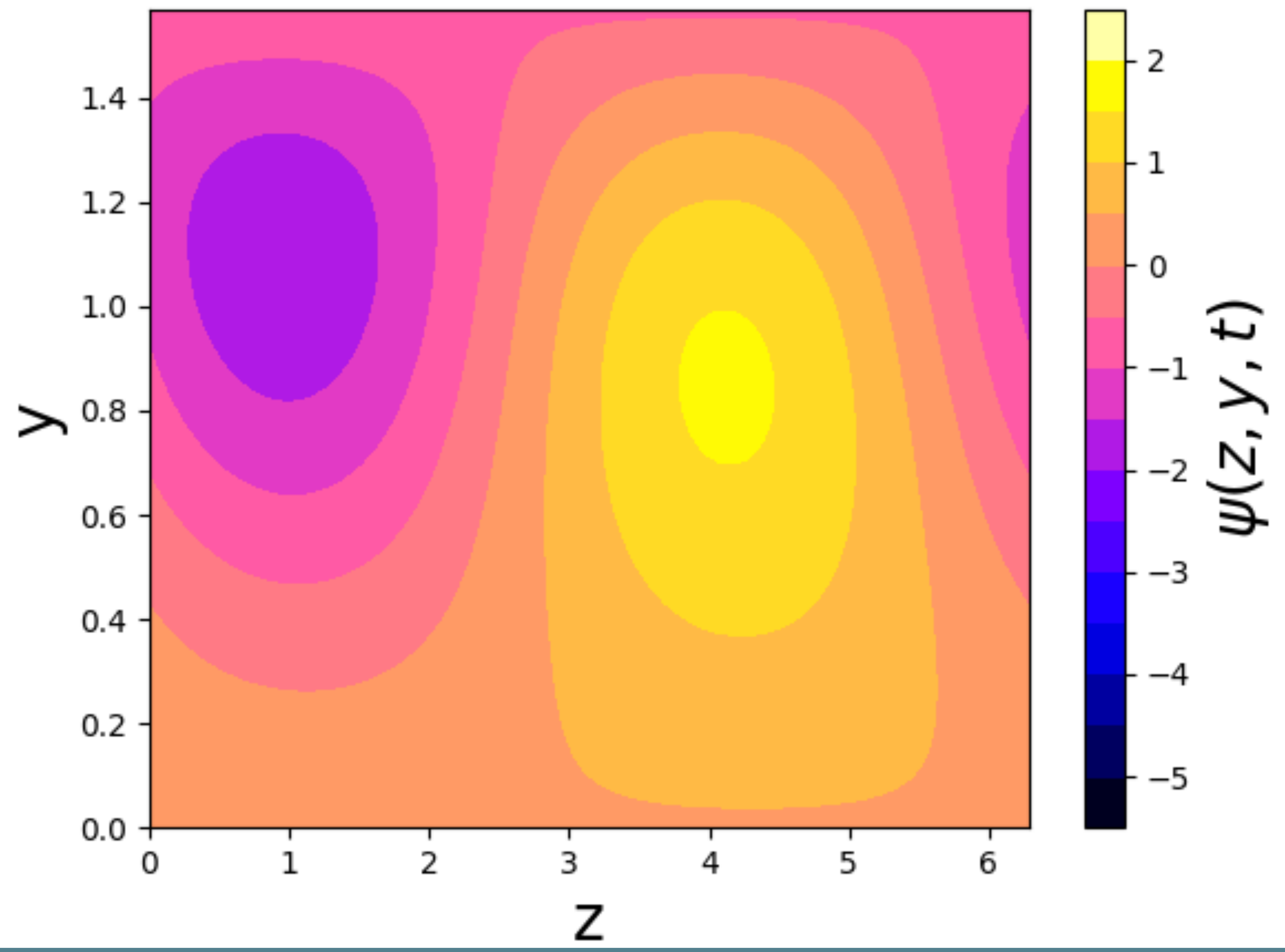
- q **is the solution of the linear problem:**

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

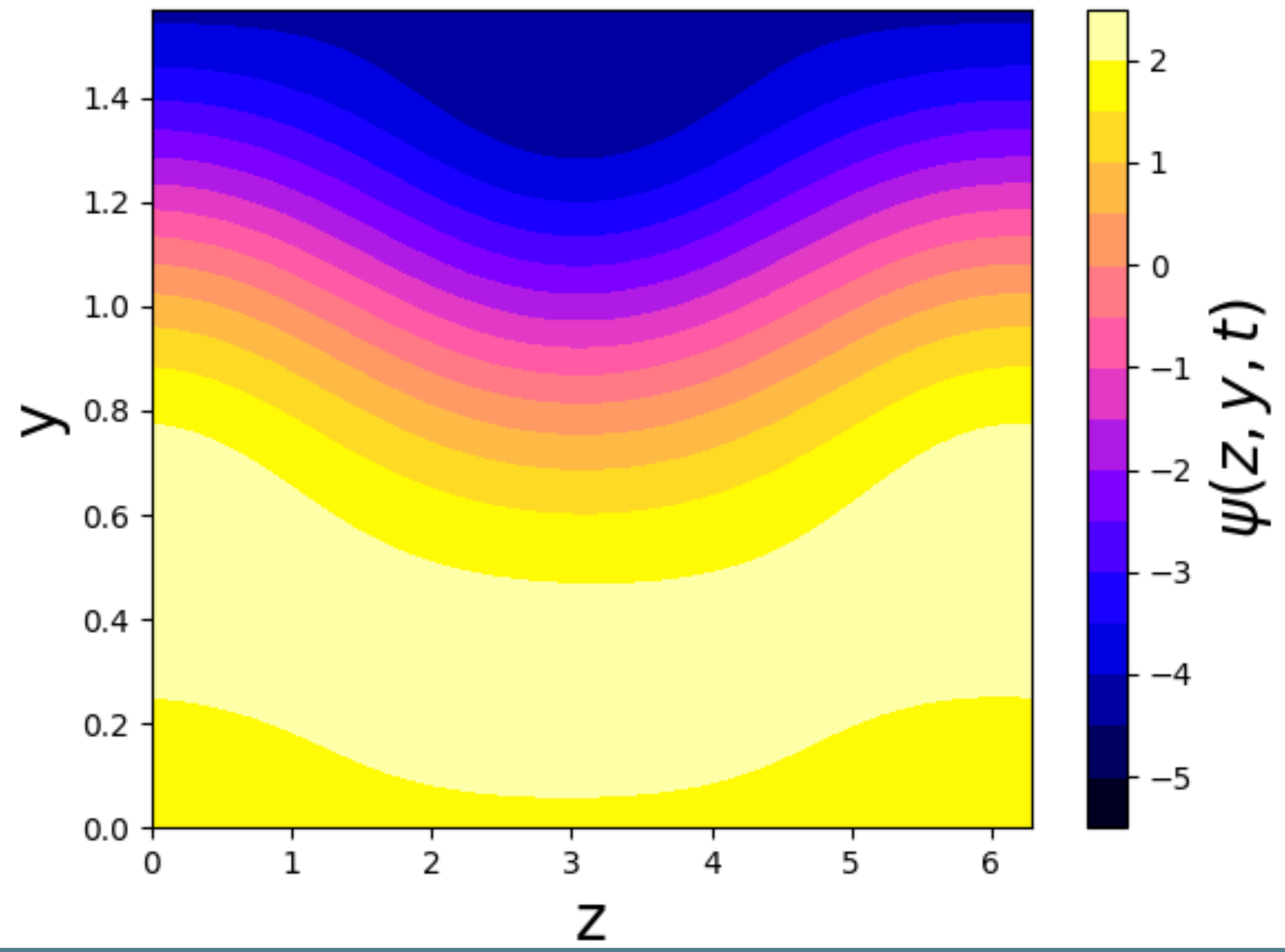
THE CHARNEY-DEVORE MODEL

$$dX_t = F(X_t)dt + \sqrt{2D}dW_t \quad \textbf{and} \quad X \in \mathbb{R}^6$$

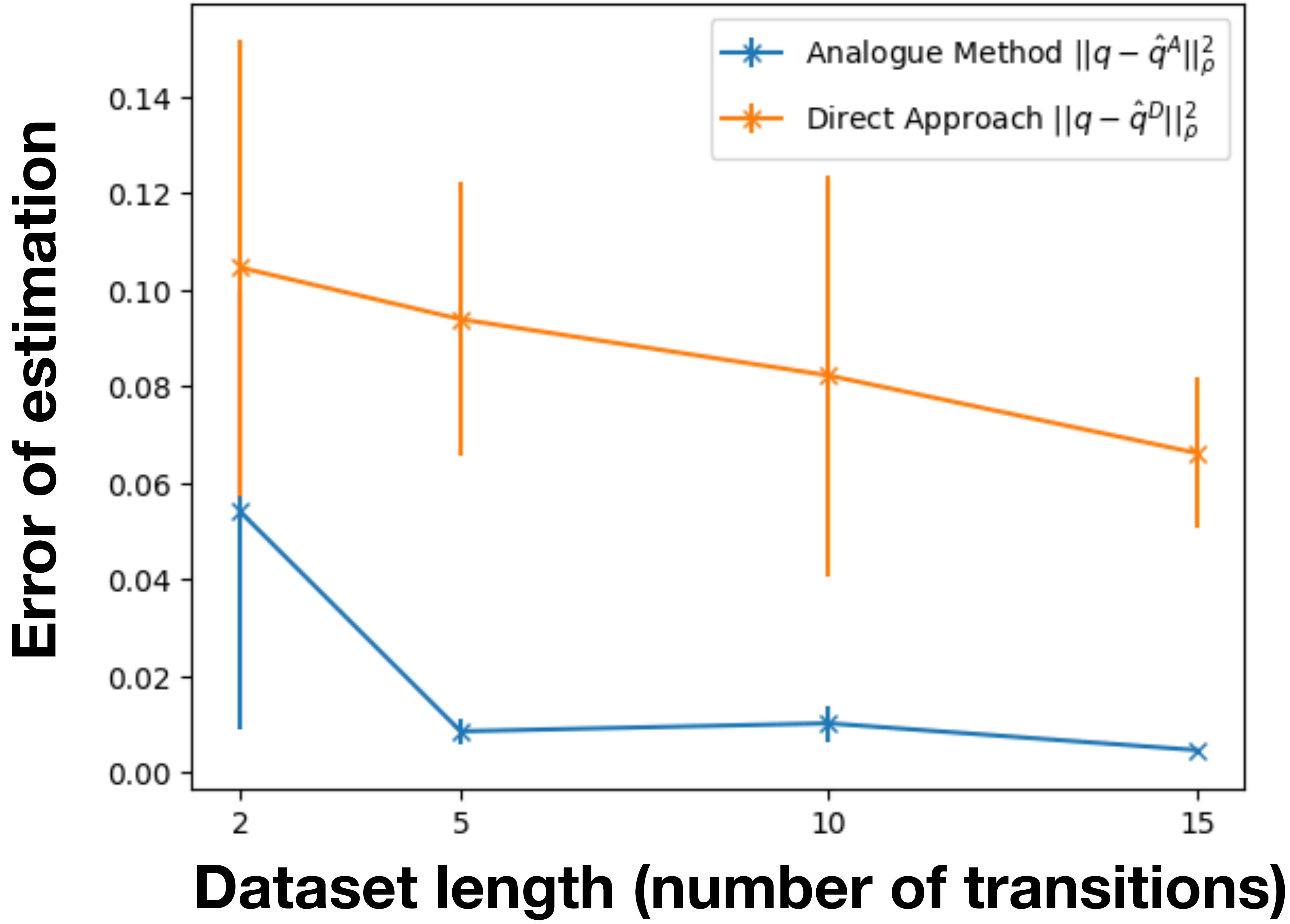
Blocked stream function $\psi(z, y, t)$



Zonal stream function $\psi(z, y, t)$



THE ANALOGUE MARKOV CHAIN PROVIDES MORE PRECISE ESTIMATION OF THE COMMITTOR THAN THE DIRECT APPROACH



The Charney-DeVore model:

$$dX_t = F(X_t)dt + \sqrt{2D}dW_t$$

- **The convergence towards the reference committor is faster by using the analogue Markov chain.**

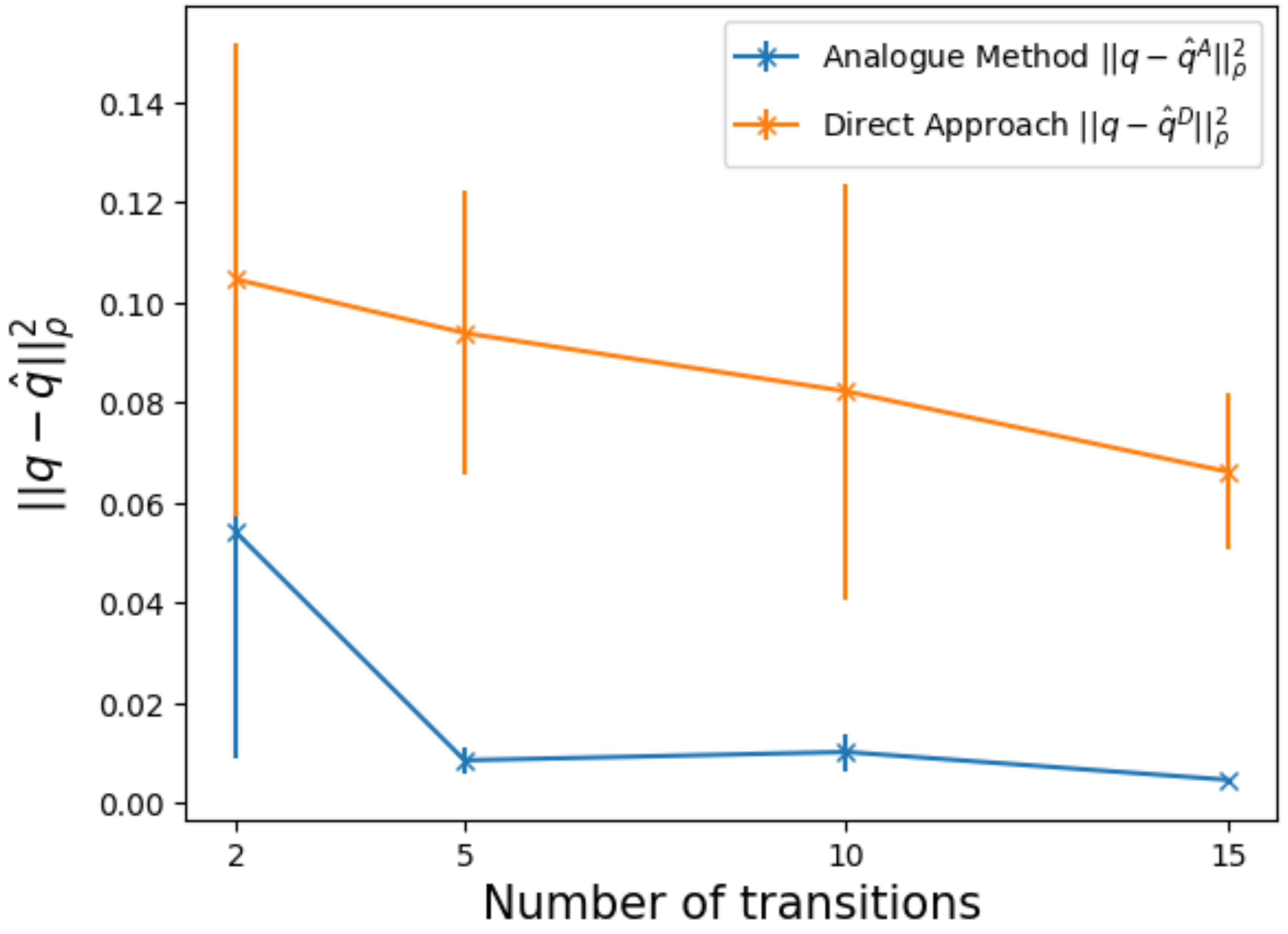
D. Lucente, J. Rolland, C. Herbert and F. Bouchet, Submitted to "Journal of Statistical Mechanics: Theory and Experiment" arXiv preprint *arXiv: 2110.05050.*

CONCLUSIONS ANALOGUE MARKOV CHAIN

1. **The committor function can be accurately estimated from few observations.**

2. **These approximations are more precise than those provided by a more naive data-driven approach.**

3. **The approximations converge faster to the exact committor when the number of observations increases.**



D. Lucente, J. Rolland, C. Herbert and F. Bouchet, Submitted to "Journal of Statistical Mechanics: Theory and Experiment" arXiv preprint *arXiv: 2110.05050.*

COUPLING RARE EVENT ALGORITHMS WITH THE ANALOGUE MARKOV CHAINS

LACK OF DATA

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

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

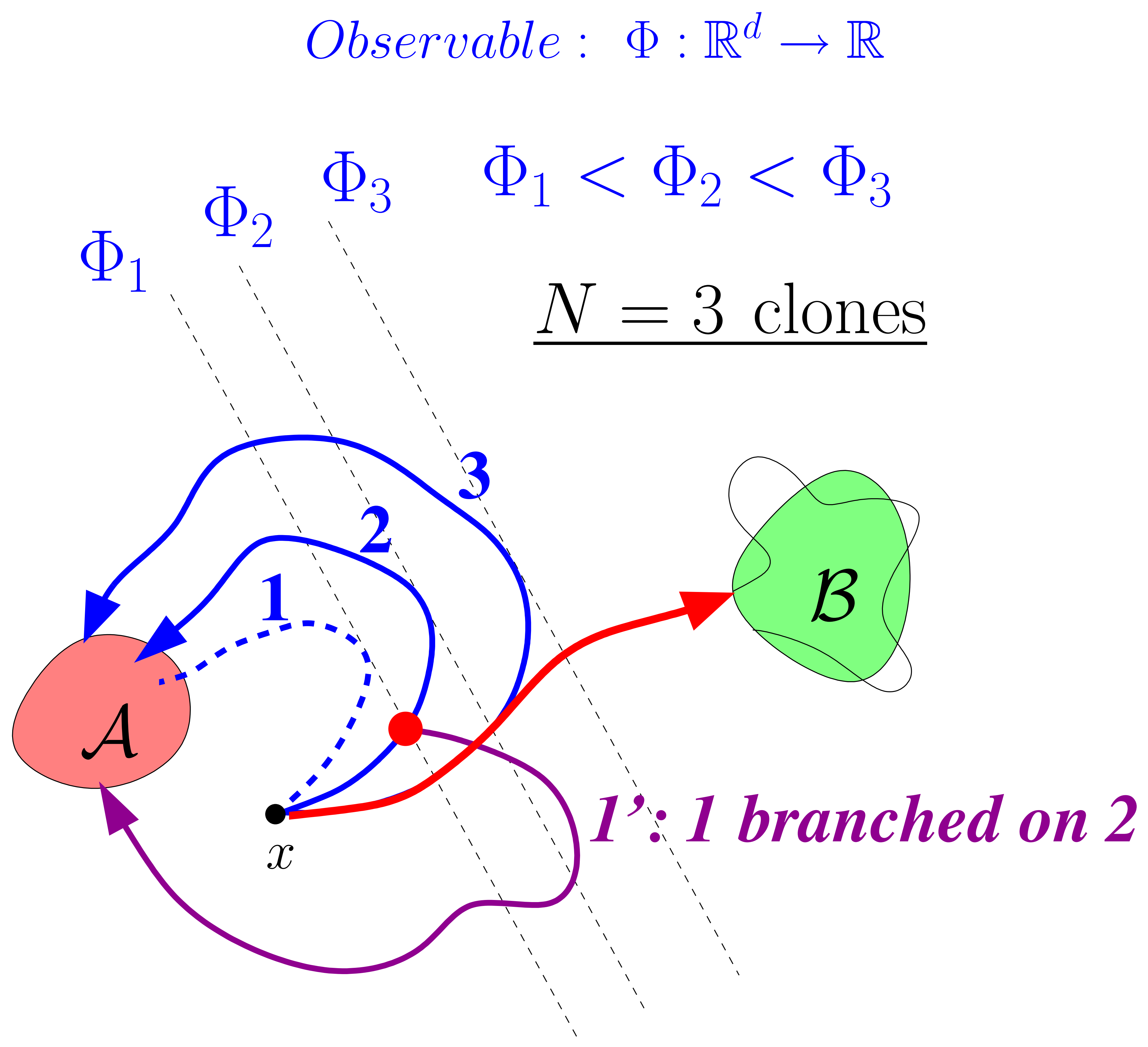
Practical question: **How to sample the probability and dynamics of rare events in complex models?**

Answer: By using rare event algorithms!

THE ADAPTIVE MULTILEVEL SPLITTING (AMS) RARE EVENT ALGORITHM

- **System of interacting replicas: copies of the Markov trajectories are simulated.**
- **Adaptive Splitting: selection of replicas using a score function $\phi(x)$.**
- **Partial resampling: new copies are simulated, by branching replicas (mutation).**

Aim: obtain the statistic of reactive paths.



COMMITTOR FUNCTIONS ARE OPTIMAL SCORE FUNCTIONS FOR RARE EVENT ALGORITHMS

The Adaptive Multilevel Splitting algorithm

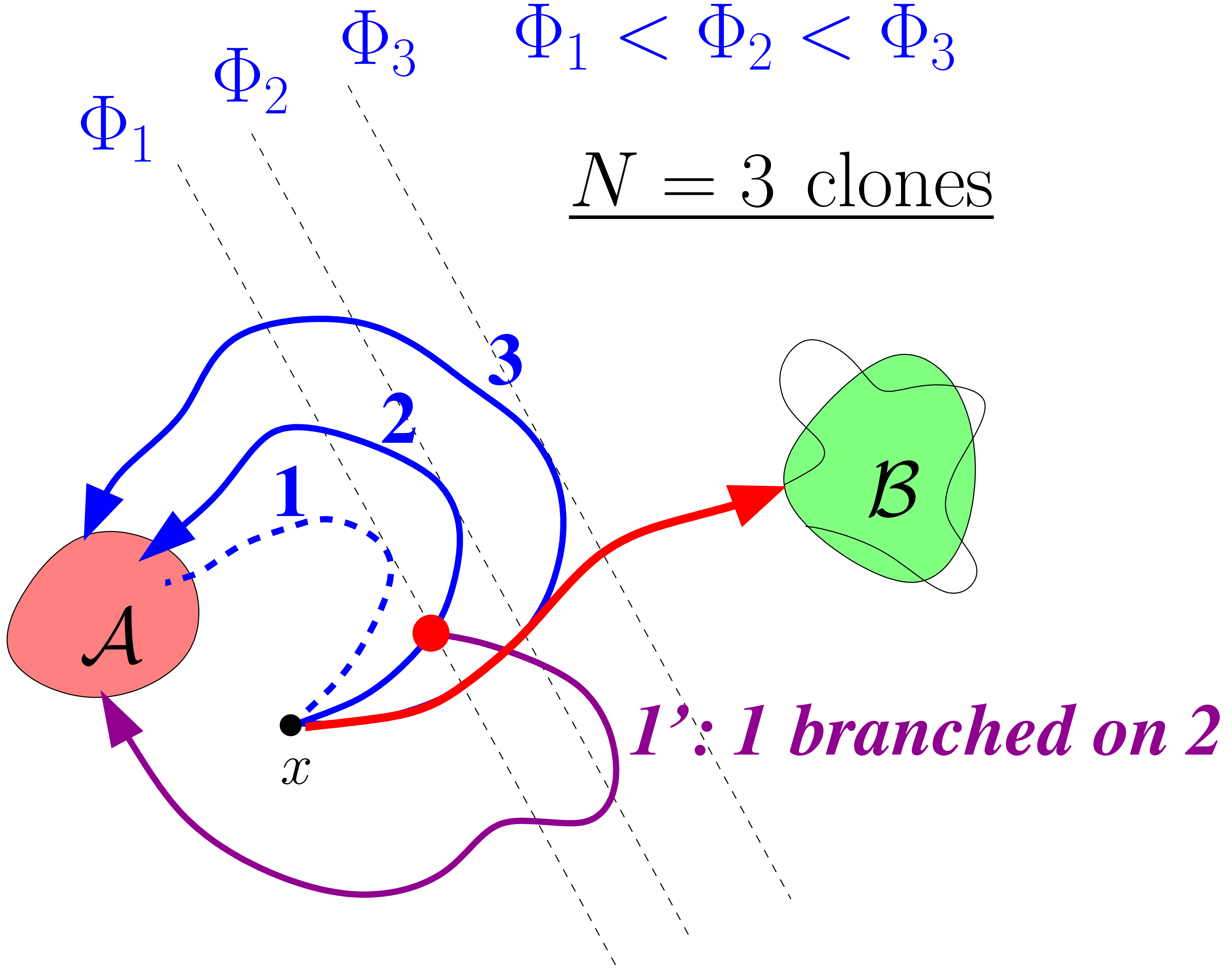
The efficiency of the algorithm depends on the choice of the score function $\phi(x)$.

The optimal score function is the committor function.

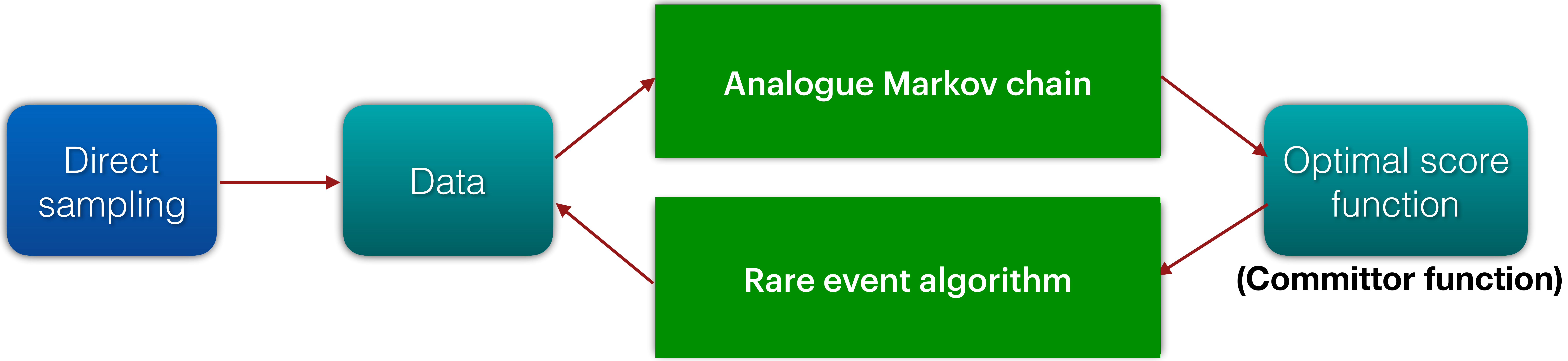
$Observable: \ \Phi : \mathbb{R}^d \to \mathbb{R}$

$\Phi_1 < \Phi_2 < \Phi_3$

$N = 3$ clones

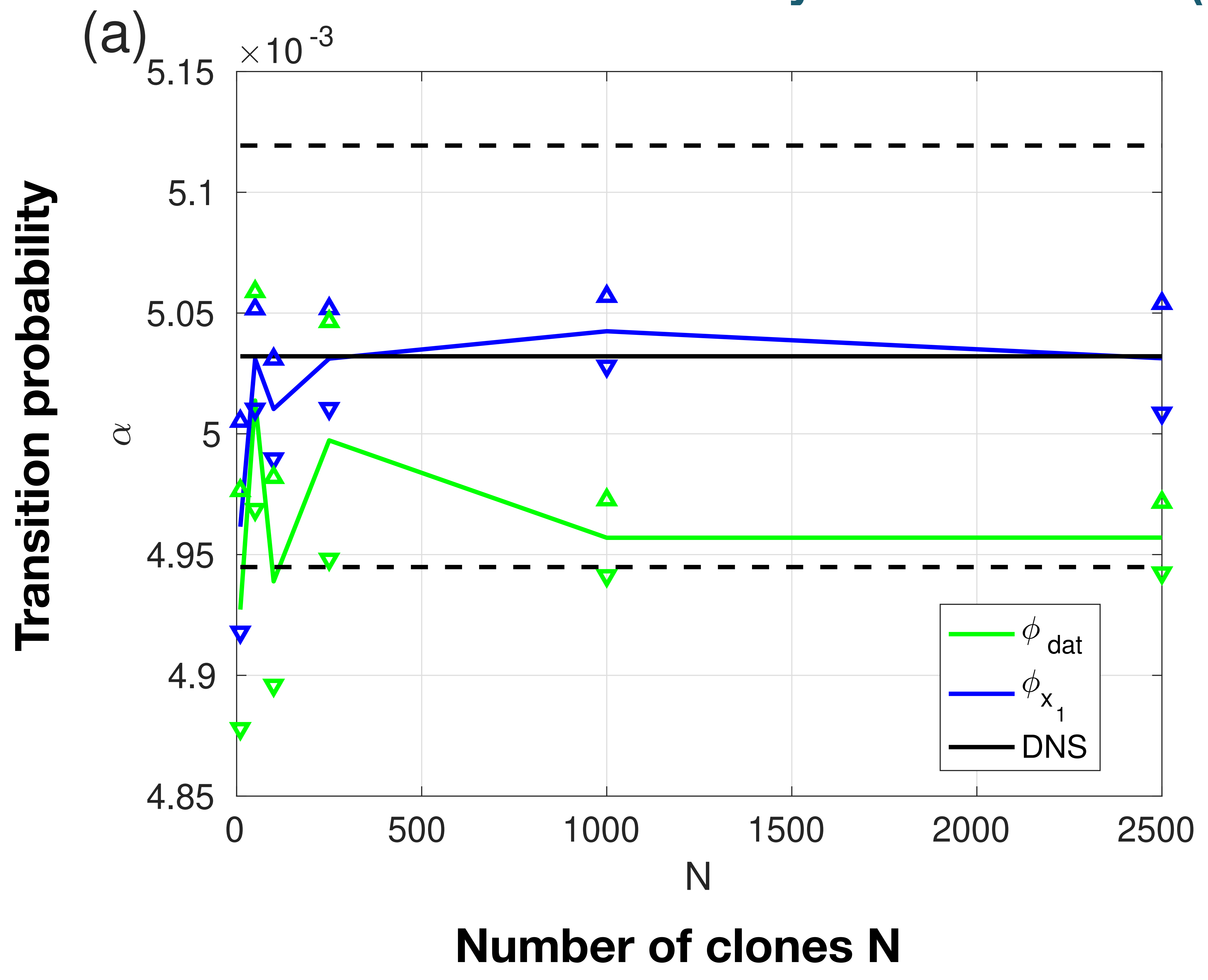


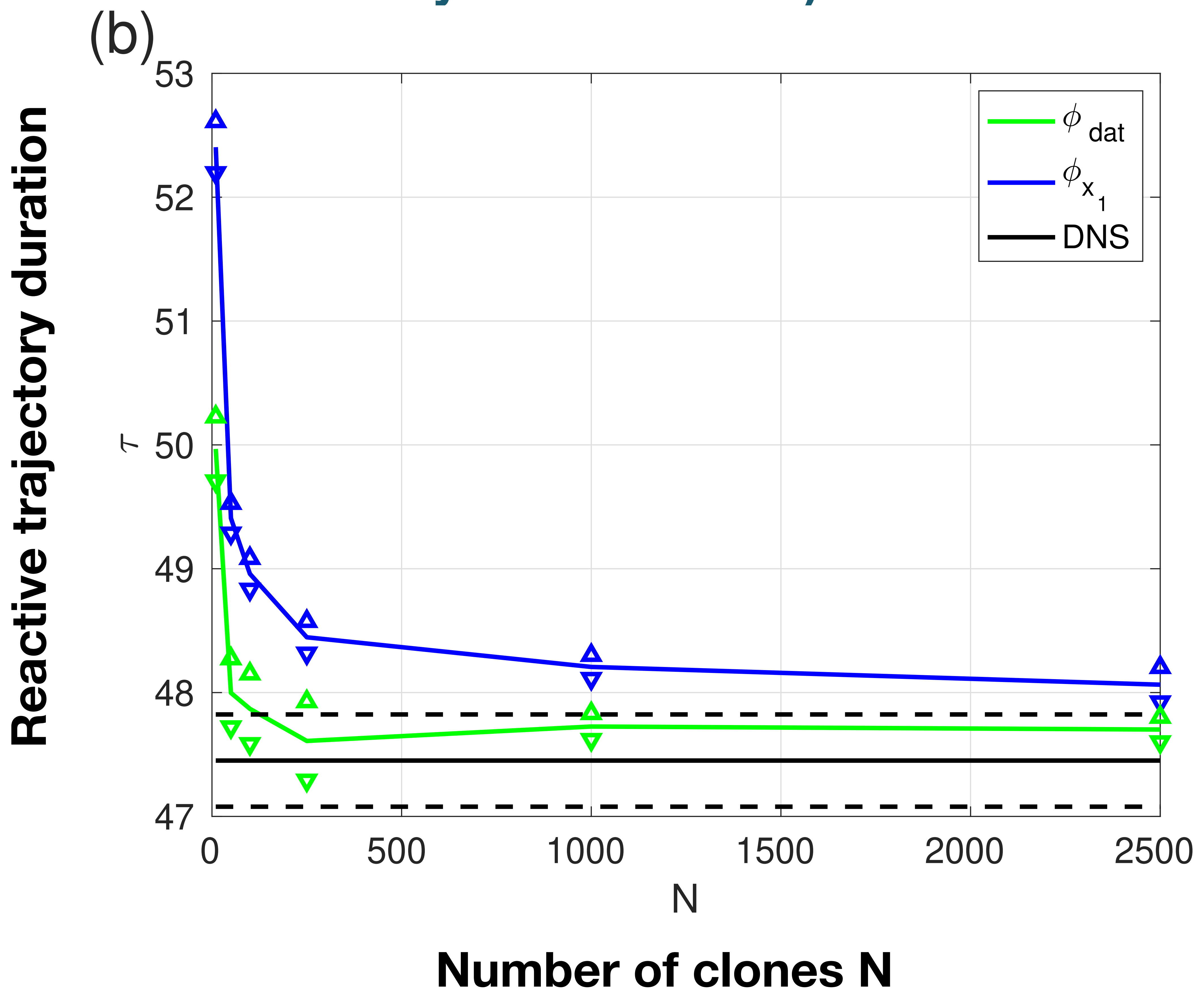
COUPLING RARE EVENT ALGORITHMS WITH DATA BASED LEARNING OF COMMITTOR FUNCTIONS



WITH THE LEARNED SCORE FUNCTION THE AMS ALGORITHM IS EXTREMELY EFFICIENT

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

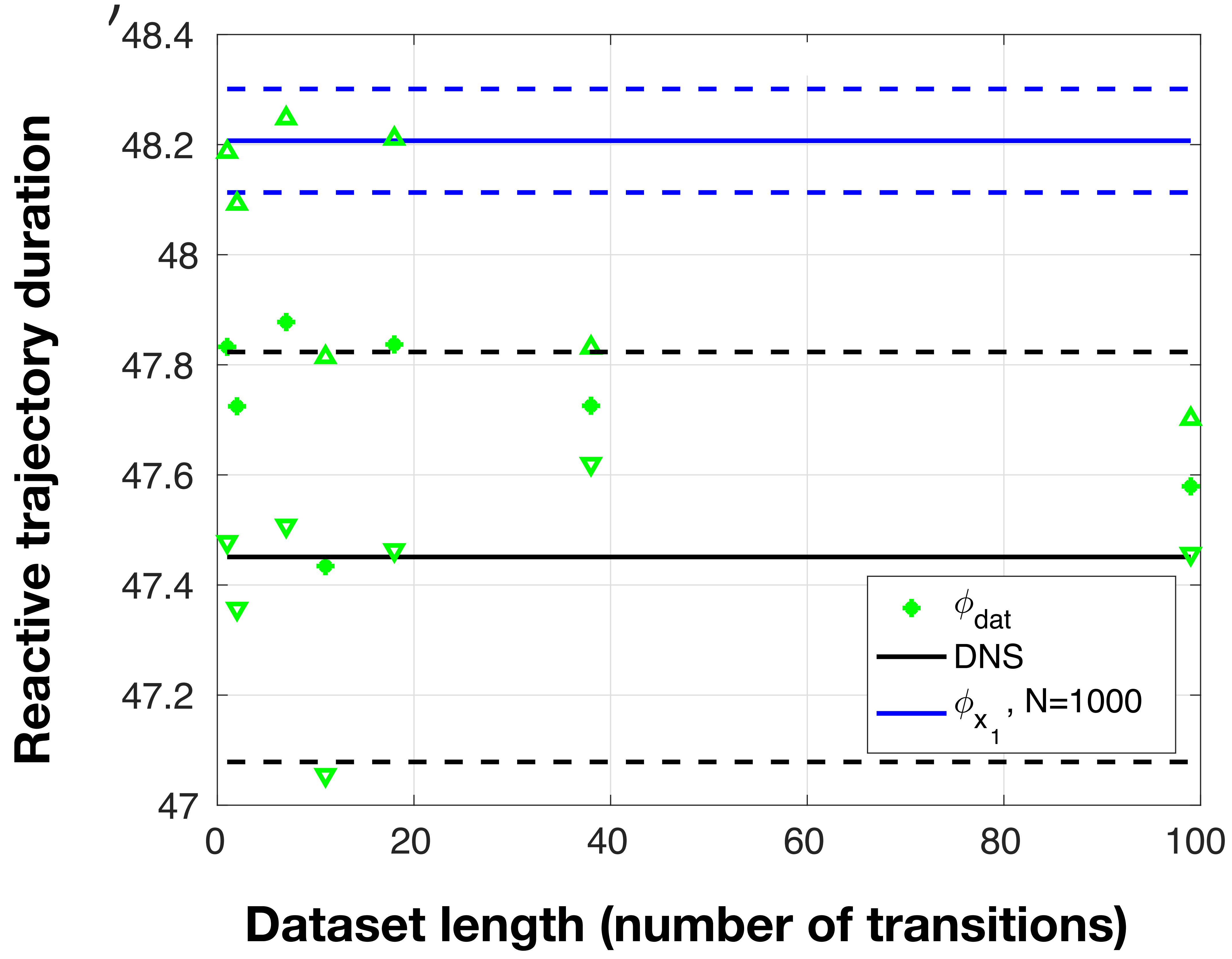




The learned committor function greatly improves the AMS performance.

WITH THE LEARNED SCORE FUNCTION THE AMS ALGORITHM IS EXTREMELY EFFICIENT

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



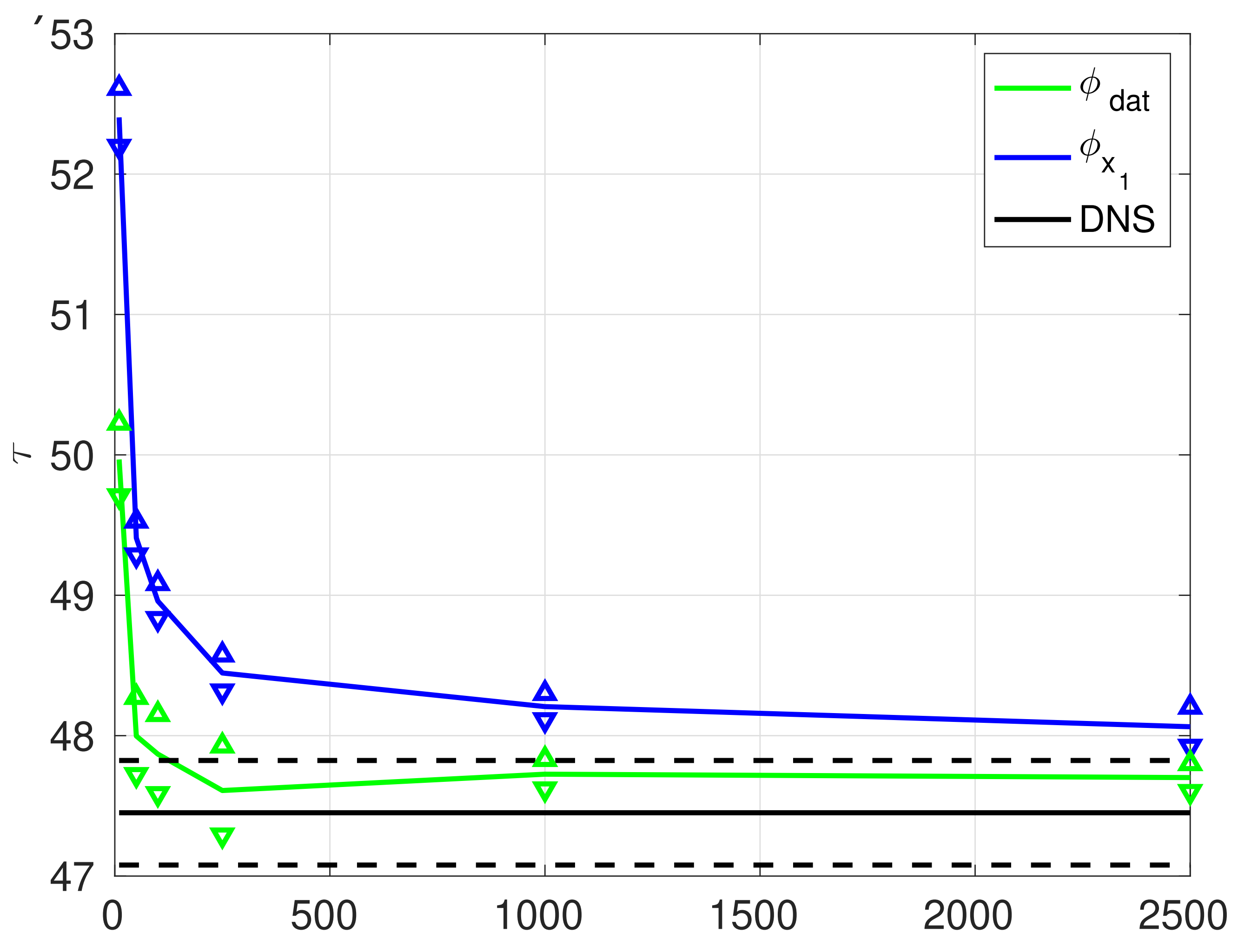
D. Lucente, J. Rolland, C. Herbert and F. Bouchet, Submitted to "Journal of Statistical Mechanics: Theory and Experiment" arXiv preprint *arXiv: 2110.05050.*

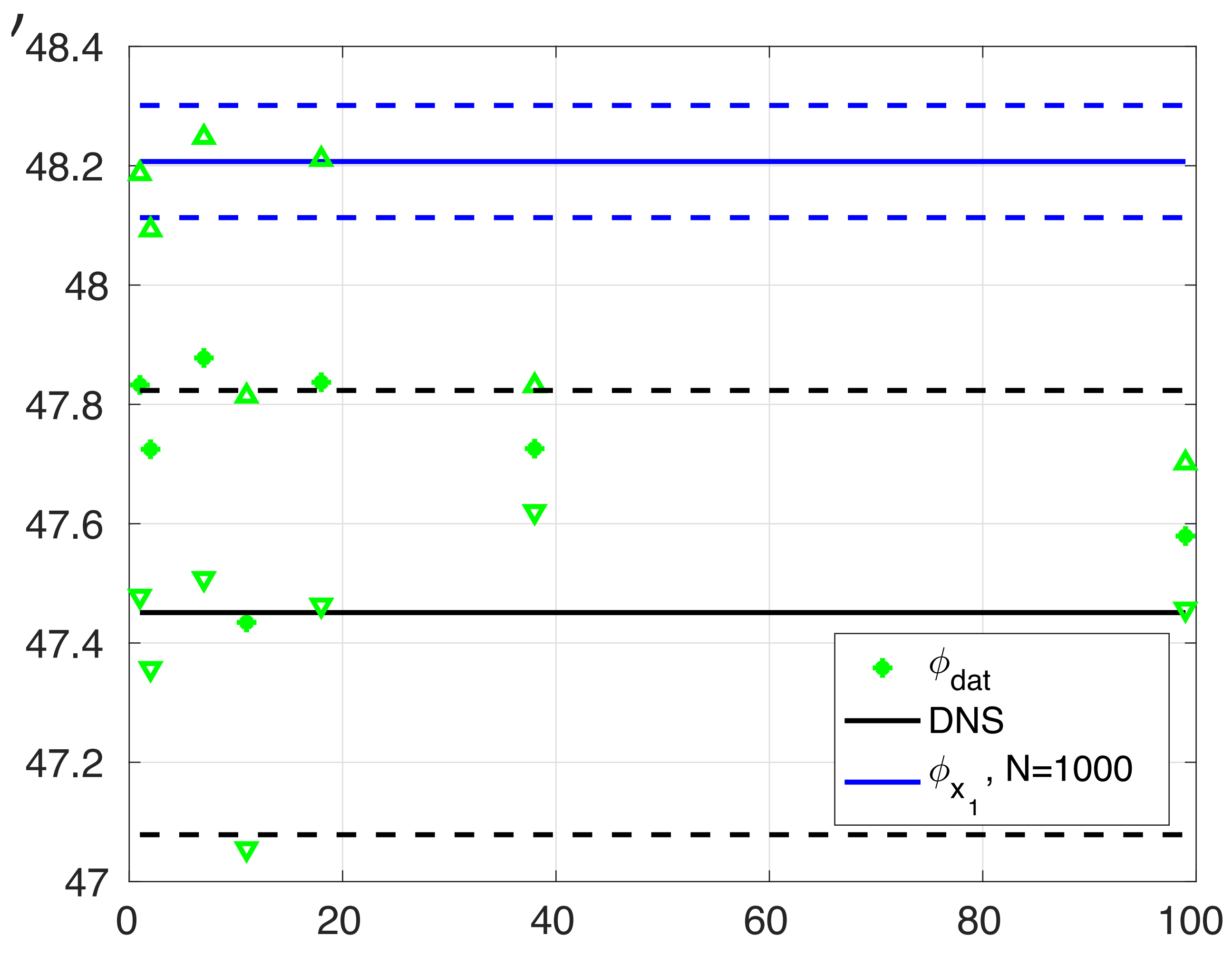
We have an impressive efficiency when learning from a dataset with about 40 to 100 observed transitions

CONCLUSIONS COUPLING RARE EVENT ALGORITHMS AND ANALOGUE MARKOV CHAINS

1. The learned approximate **committor functions are extremely efficient score functions** for the AMS algorithm.

2. The rare events can be simulated with a **minimal prior knowledge on the system.**





D. Lucente, J. Rolland, C. Herbert and F. Bouchet, Submitted to "Journal of Statistical Mechanics: Theory and Experiment" arXiv preprint *arXiv: 2110.05050.*

THE ANALOGUE MARKOV CHAIN FOR HEAT WAVES

HEAT WAVE DEFINITION

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

- **Long term fluctuations of temperature:**

$$A(t_0) = \frac{1}{\mathcal{T}} \int_{t_0}^{t_0+\mathcal{T}} \mathrm{d}t \frac{1}{\mathcal{D}} \int_{\mathcal{D}} \mathrm{d}\mathbf{r}[T(\mathbf{r}, t) - \mathbb{E}[T](\mathbf{r}, t)]$$

$\mathcal{T} = 15$ **days and** $\mathcal{D} =$ **France.**

- $Y(t_0) \in \{0,1\}$. $Y(t_0) = 1$ if $A(t_0) > a$, and $Y(t_0) = 0$ otherwise. **A heat wave occurs if** $Y = 1$**.**

- **Committor function:**

$$q(X, \tau) = \mathbb{P}(A(t_0 + \tau) > a \,|\, X(t_0) = X) = \mathbb{E}_X[Y(t_0 + \tau)].$$

COMPUTING ANALOGUE MARKOV CHAIN FOR STUDYING MID-LATITUDES HEAT WAVES

- **High-dimensional data**

- **Presence of a strong seasonal cycle**

- **Non-stationary dynamics**

COMPUTING ANALOGUE MARKOV CHAIN FOR STUDYING MID-LATITUDES HEAT WAVES

- **How to validate or invalidate an analogue approach? What is it good for ?**
- **Which variables should be used? In which domain?**
- **Which metric?**
- **How to choose the time step of the Markov chain or the coarse-graining time of the variables?**

SEASONAL VS ANNUAL MARKOV CHAIN

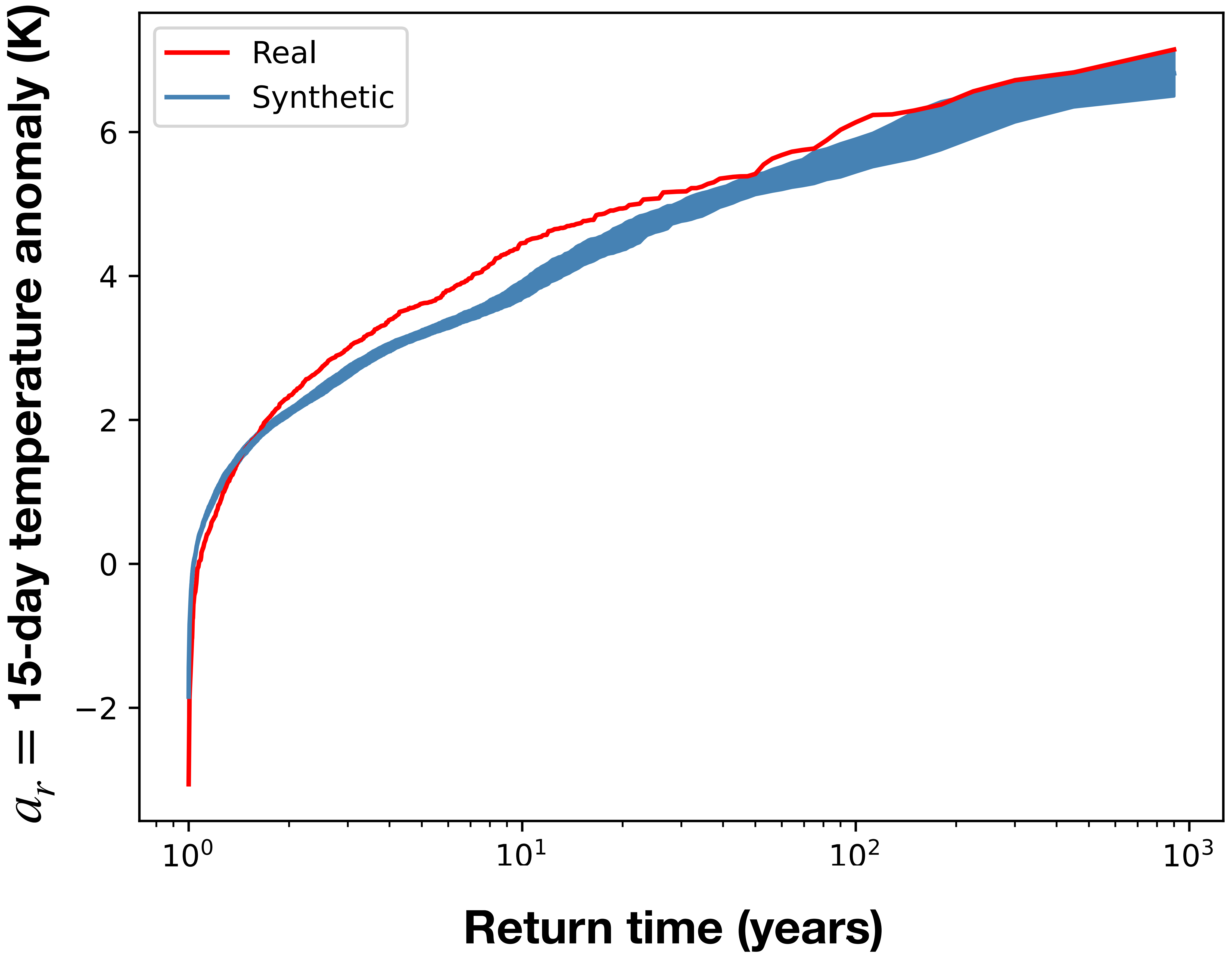
- **SEASONAL ANALOGUE MARKOV CHAIN:**

1. **The analogues are sought in a given season (e.g summer)**
2. **The trasition matrix G does not depend on time**

- **ANNUAL TIME-PERIODIC ANALOGUE MARKOV CHAIN:**

1. **One should keep track of the calendar date of the Markov chain**
2. **The analogues are sought in a time window of 2 months centered around the calendar date of the Markov chain**
3. **The trasition matrix G is periodic in time**

RETURN TIME PLOT COMPUTED USING THE ANALOGUE MARKOV CHAIN (PLASIM)



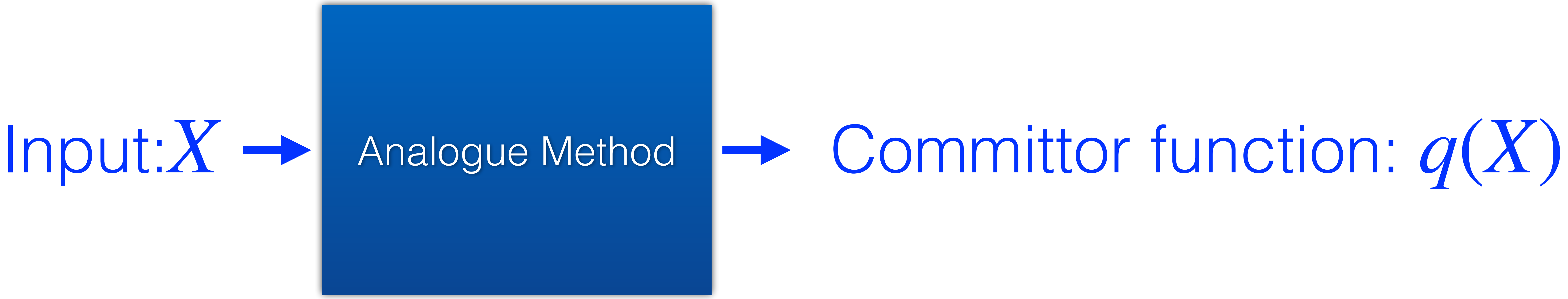
Annual time-periodic Markov chain.

$X = \tilde{Z}_g^a$ **North-Atlantic.**

K=5 analogues.

By using 100 years of data, the return period of extreme events up to 900 years can be computed.

HEAT WAVES PREDICTION: ANALOGUE BASED COMMITTOR



- The Logarithmic score is used to test the committor function prediction.

COMBINATION OF DIFFERENT PHYSICAL INFORMATION

Normal Euclidean distance: $X = (Z, T, M)$ **and**

$$d(X_1, X_2) = \left[\frac{1}{\sigma_Z^2} \sum_{i=1}^{\dim(Z)} (Z_1[i] - Z_2[i])^2 + \frac{1}{\sigma_T^2} (T_1 - T_2)^2 + \frac{1}{\sigma_M^2} (M_1 - M_2)^2 \right]^{\frac{1}{2}}$$

Normalized Euclidean distance: $X = (Z, T, M)$ **and**

$$d(X_1, X_2) = \left[\frac{\alpha}{\sigma_Z^2 \dim(Z)} \sum_{i=1}^{\dim(Z)} (Z_1[i] - Z_2[i])^2 + \frac{\beta}{\sigma_T^2} (T_1 - T_2)^2 + \frac{\gamma}{\sigma_M^2} (M_1 - M_2)^2 \right]^{\frac{1}{2}}$$

PREDICTION OF HEAT WAVES BY USING THE ANALOGUE MARKOV CHAIN

$$X = (T_s, Z_g, Mrso)$$

where

$T_s =$ **temperature anomaly averaged over France,**

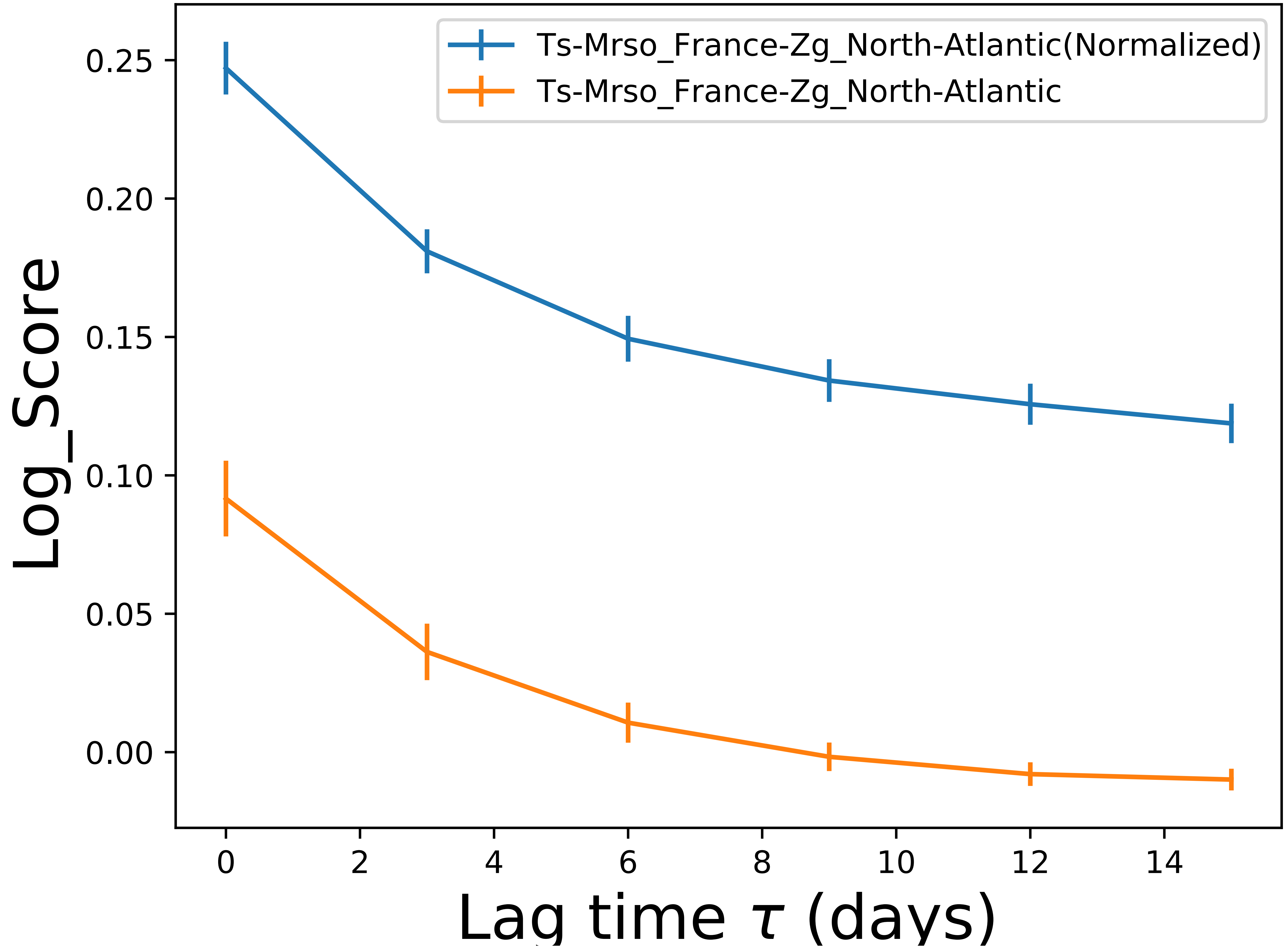
$Z_g =$ **geopotential height anomaly over North-Atlantic,**

$Mrso =$ **soil moisture anomaly averaged over France.**

$$q(X, \tau) = \mathbb{P}(A(t_0 + \tau) > a \,|\, X(t_0) = X)$$

$a = 3.08K$ (5 % most extreme events)

It is better to use the **normalized Euclidean distance**



PREDICTION OF HEAT WAVES BY USING THE ANALOGUE MARKOV CHAIN

$$X = (T_s, Z_g, Mrso)$$

where

T_s = **temperature anomaly averaged over France,**

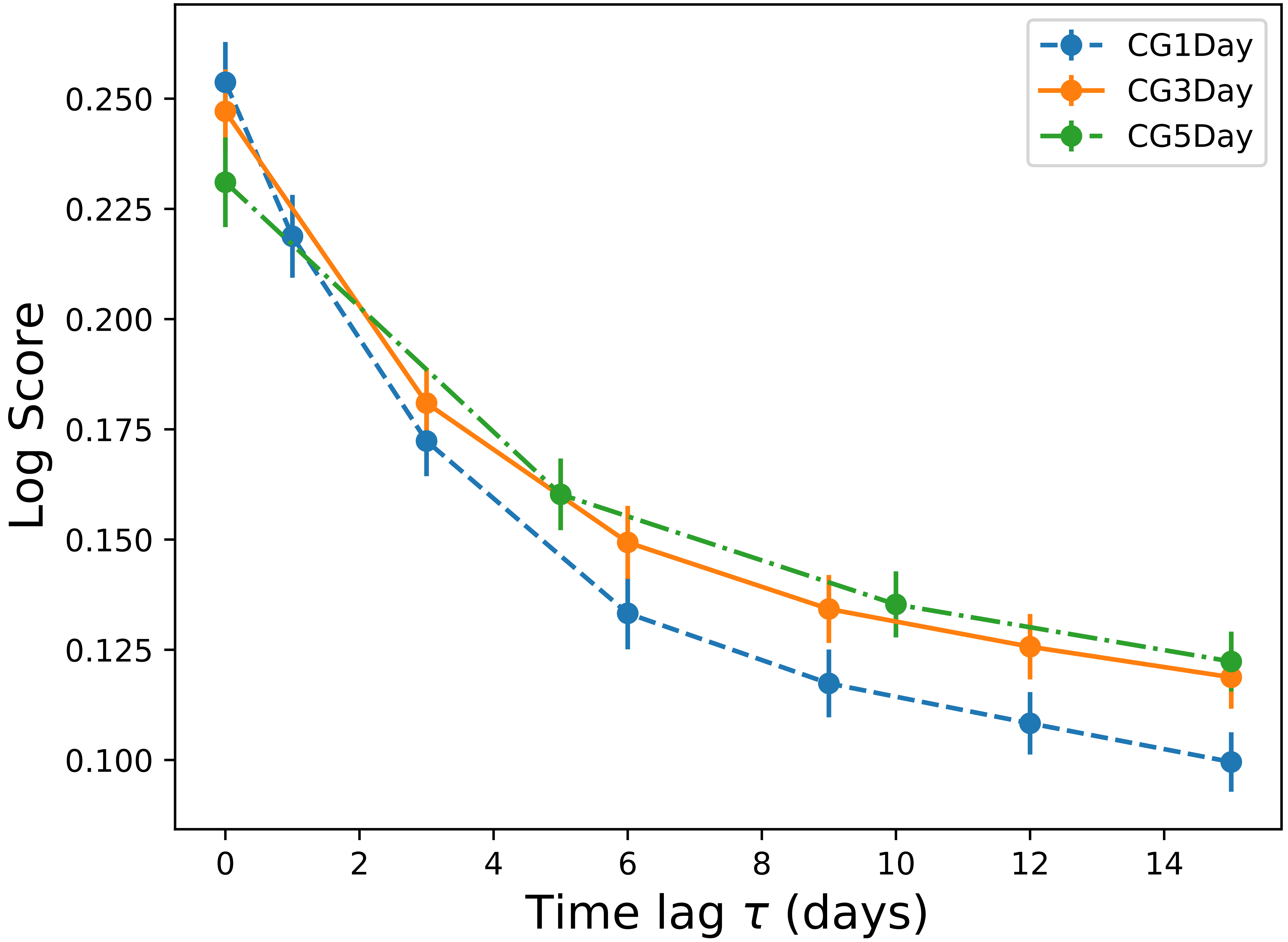
Z_g = **geopotential height anomaly over North-Atlantic,**

$Mrso$ = **soil moisture anomaly averaged over France.**

$$q(X,\tau) = \mathbb{P}(A(t_0+\tau) > a \,|\, X(t_0) = X)$$

$a = 3.08K$ (5 % most extreme events)

It is better to use a **coarse graining time of 3-days**



PREDICTION OF HEAT WAVES BY USING THE ANALOGUE MARKOV CHAIN

$$X = (T_s, Z_g, Mrso)$$

T_s = **temperature anomaly averaged over France,**

Z_g = **geopotential height anomaly over North-Atlantic,**

$Mrso$ = **soil moisture anomaly averaged over France.**

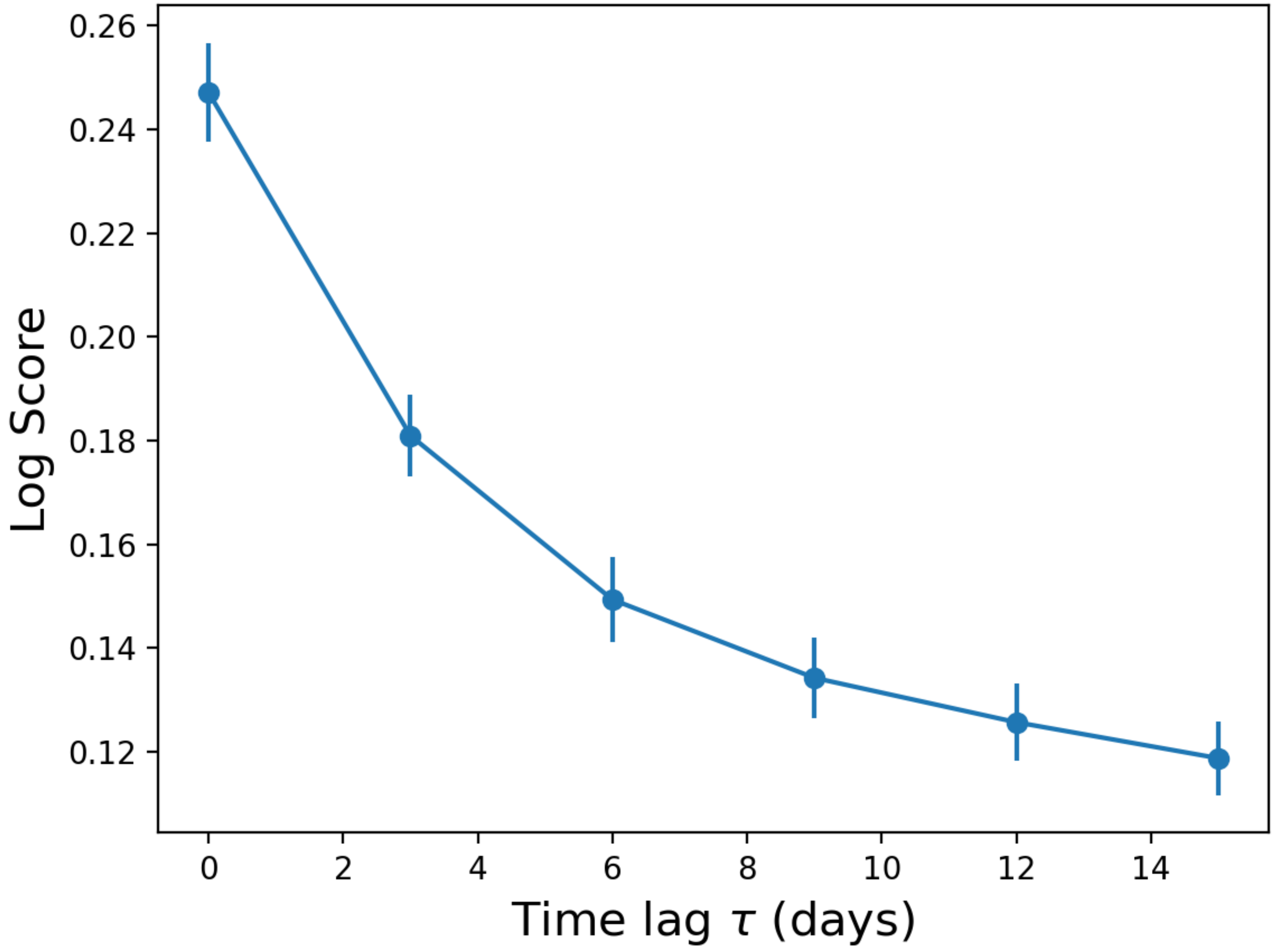
$$q(X, \tau) = \mathbb{P}(A(t_0 + \tau) > a \,|\, X(t_0) = X)$$

$a = 3.08K$ (5 % most extreme events)

Logaritmic score far from zero

↓

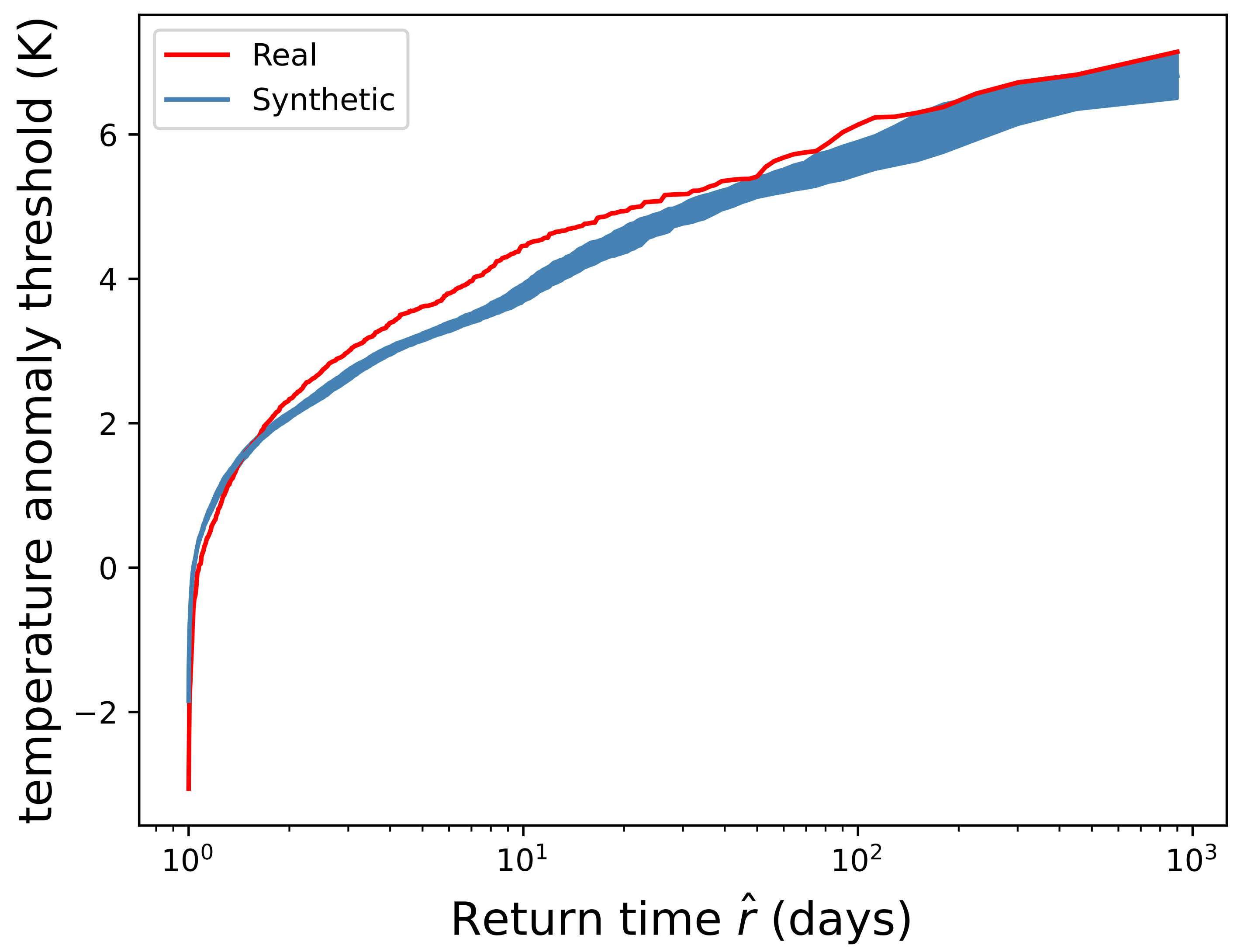
We can predict heat waves 15 days in advance.

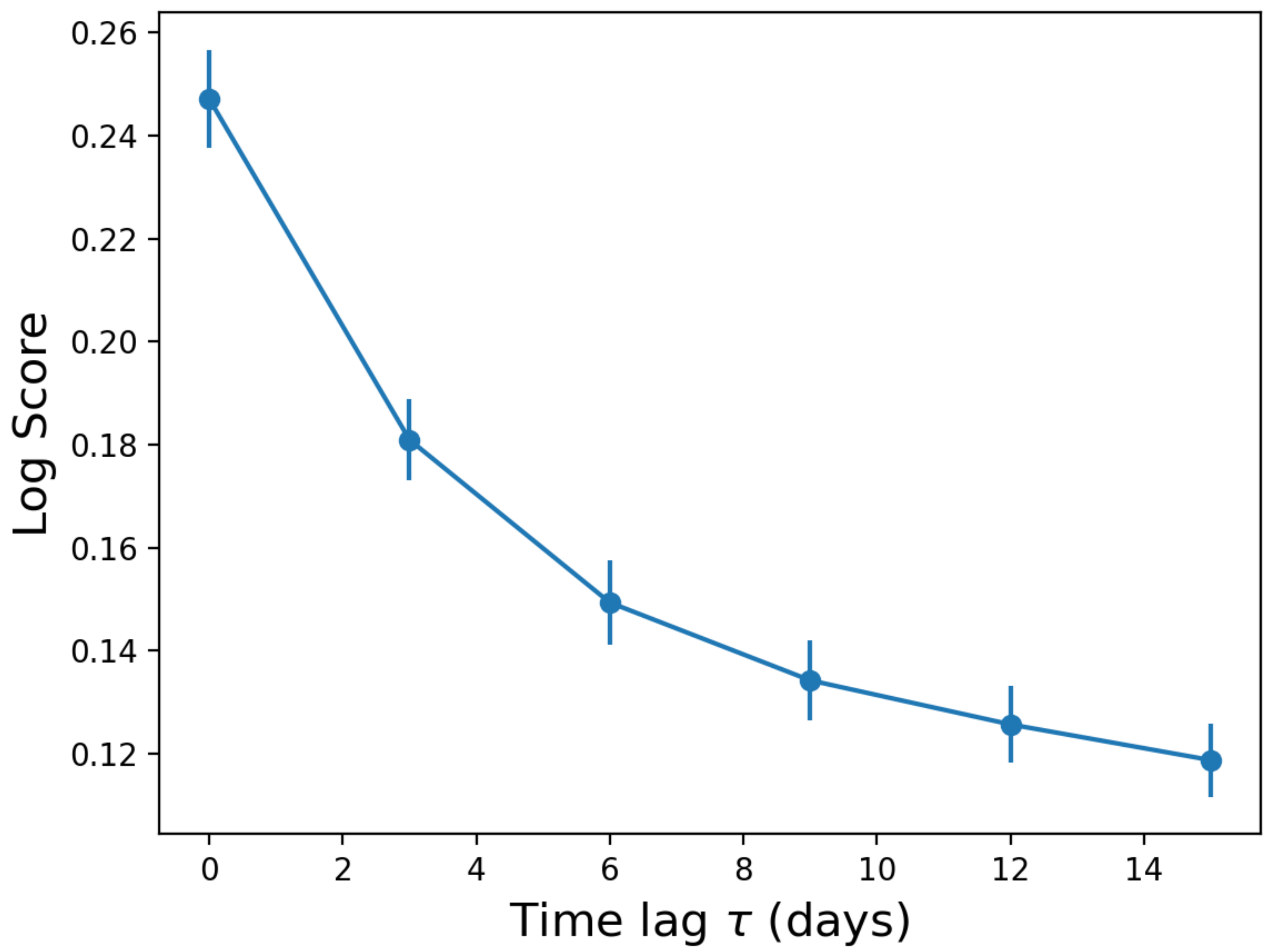


CONCLUSIONS ANALOGUE MARKOV CHAIN FOR HEAT WAVES

1. **The annual time-periodic Markov chain correctly reproduce the long-term statistics of the heat waves. By using 100 years of data, the return period of extreme events up to 900 years can be computed.**

2. **The analogue method allows the computation of the committor function. The probabilities of occurence of heat waves are correctly estimated up to 15 days in advance.**





CONCLUSIONS AND PERSPECTIVES

CONCLUSIONS

1. The committor function quantifies the probability of occurence of the event. It is the relevant mathematical object for prediction problem at the predictability margin.

2. The AMS algorithm performs extremely well when it is coupled to the analogue Markov chain.

3. The analogue Markov chain allows the computation of the committor function for heat waves.

PERSPECTIVES

- Learn more informative distance that takes into account the structure of the data.
- Learn most informative physical variables for computing committor functions.
- Coupling rare events algorithms and the analogue Markov chains for climate models.

REFERENCES

1. D. Lucente, C. Herbert and F. Bouchet: arXiv:2106.14990

2. D. Lucente, J. Rolland, C. Herbert and F. Bouchet: arXiv:2110.05050

Contact: dariolucente32@gmail.com