FlarePredict:
Predicting solar flares, sandpile models, and machine learning

Antoine Strugarek

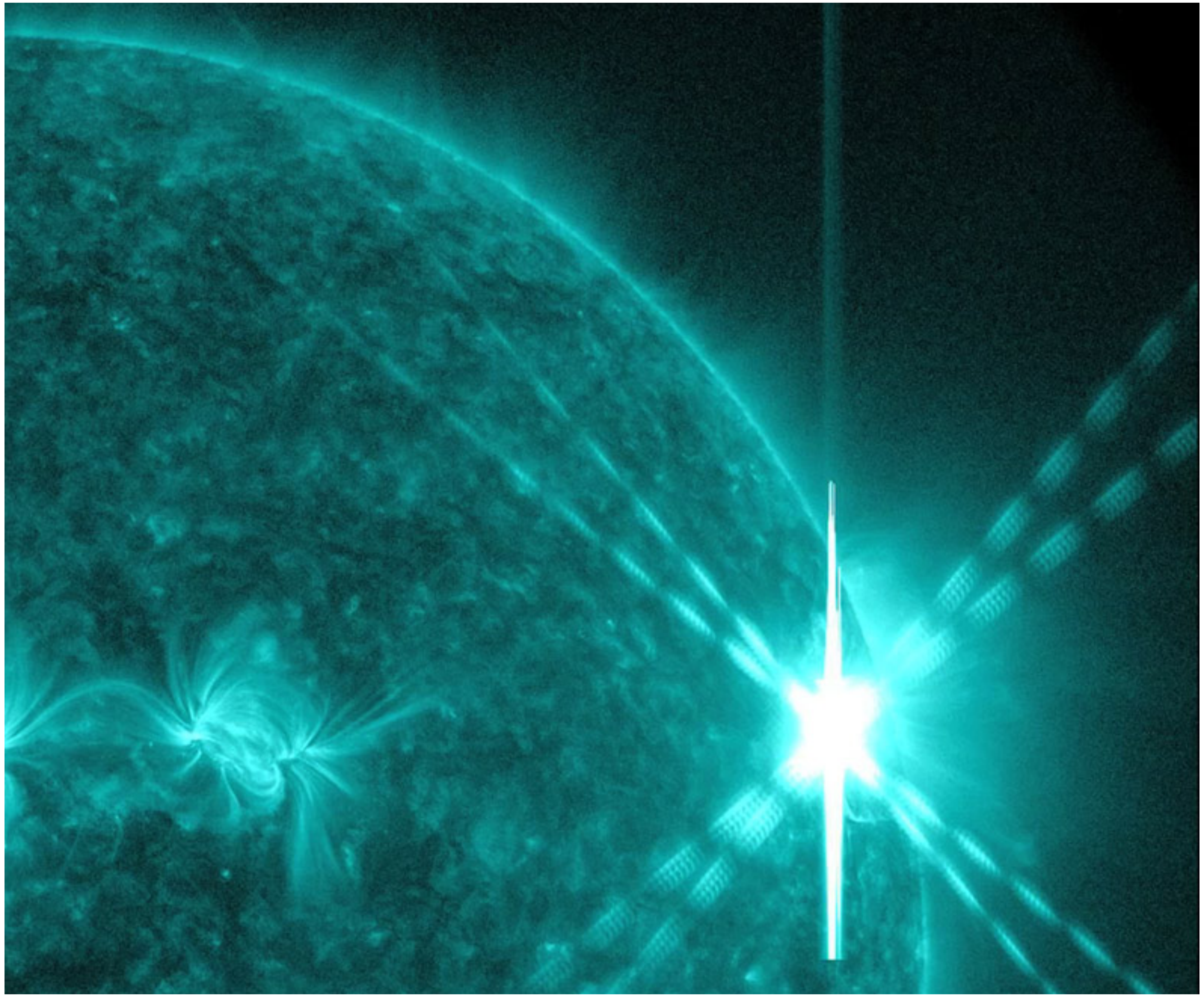
CEA/Saclay, DAp-AIM

With A.S. Brun, P. Charbonneau, H. Lamarre, C. Thibeault, B. Tremblay

Projet Emergent

AIM cea P2IO Physique des 2 Infinis et des Origines INVESTIR L'AVENIR Université de Montréal

What are solar flares?

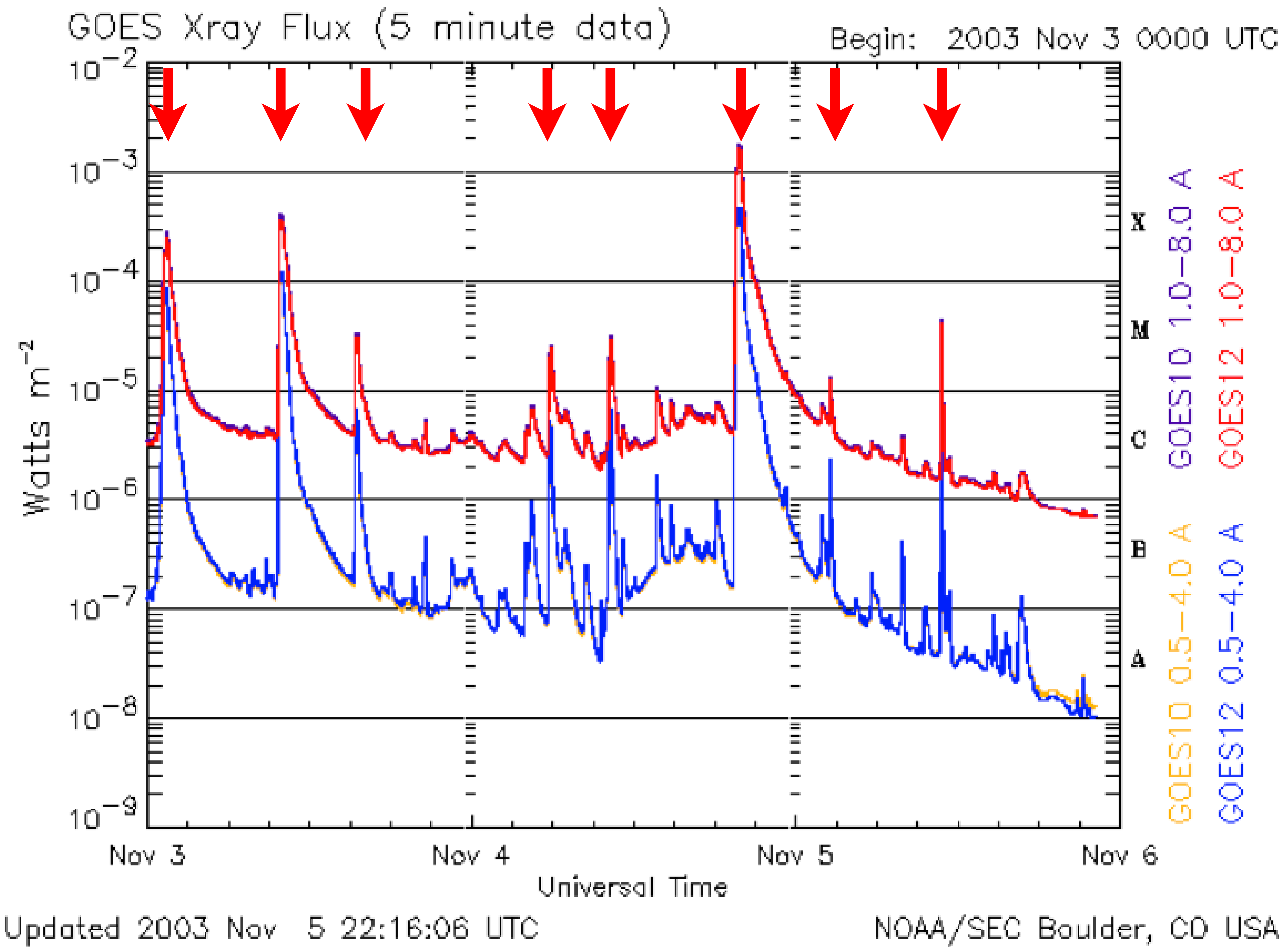


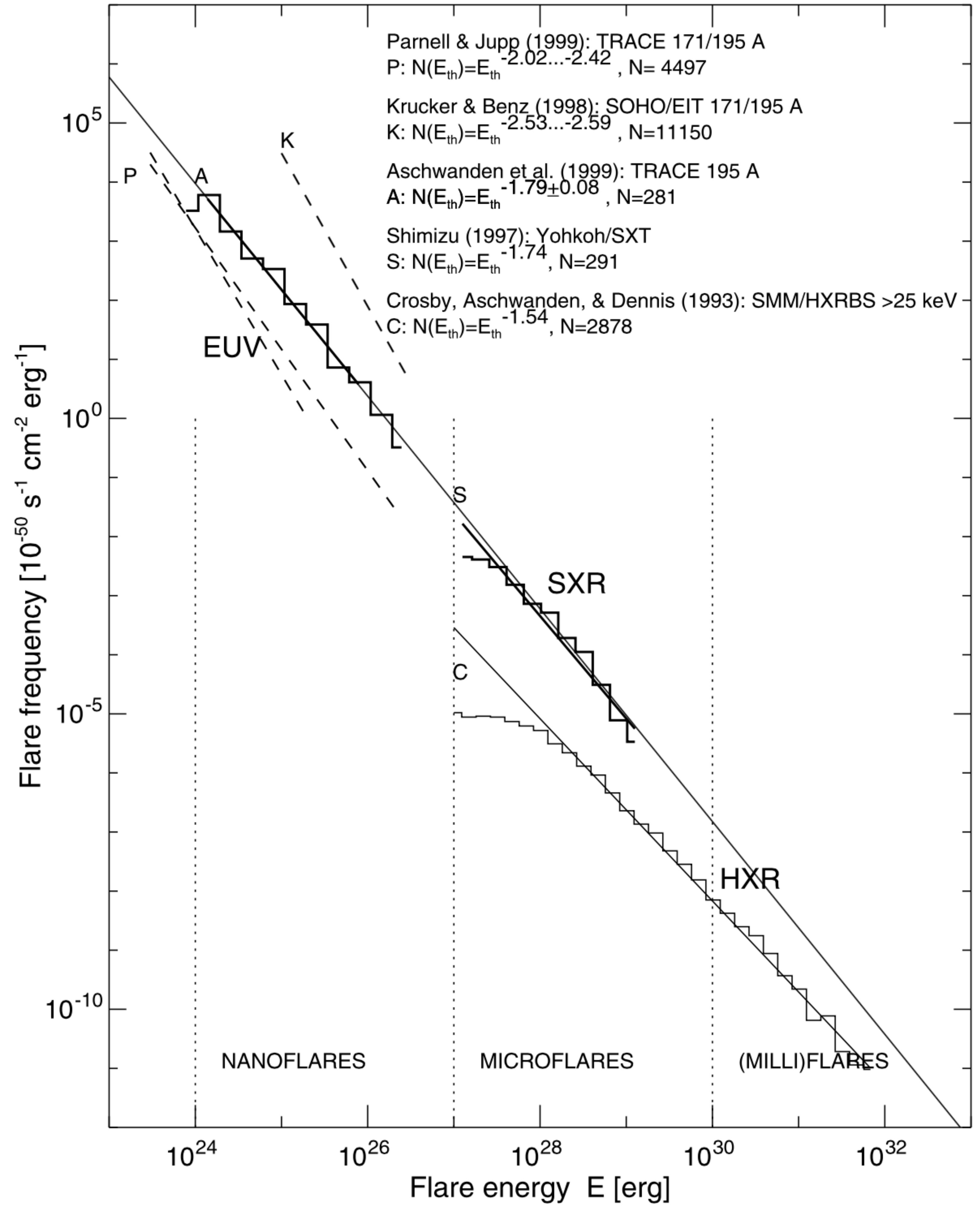
Impulsive (< 1s) solar events, visible in X-rays, UV, extreme UV…

Originate from magnetic structures in the low atmosphere of the Sun

Can reach energies of typically a few 10^{32} erg

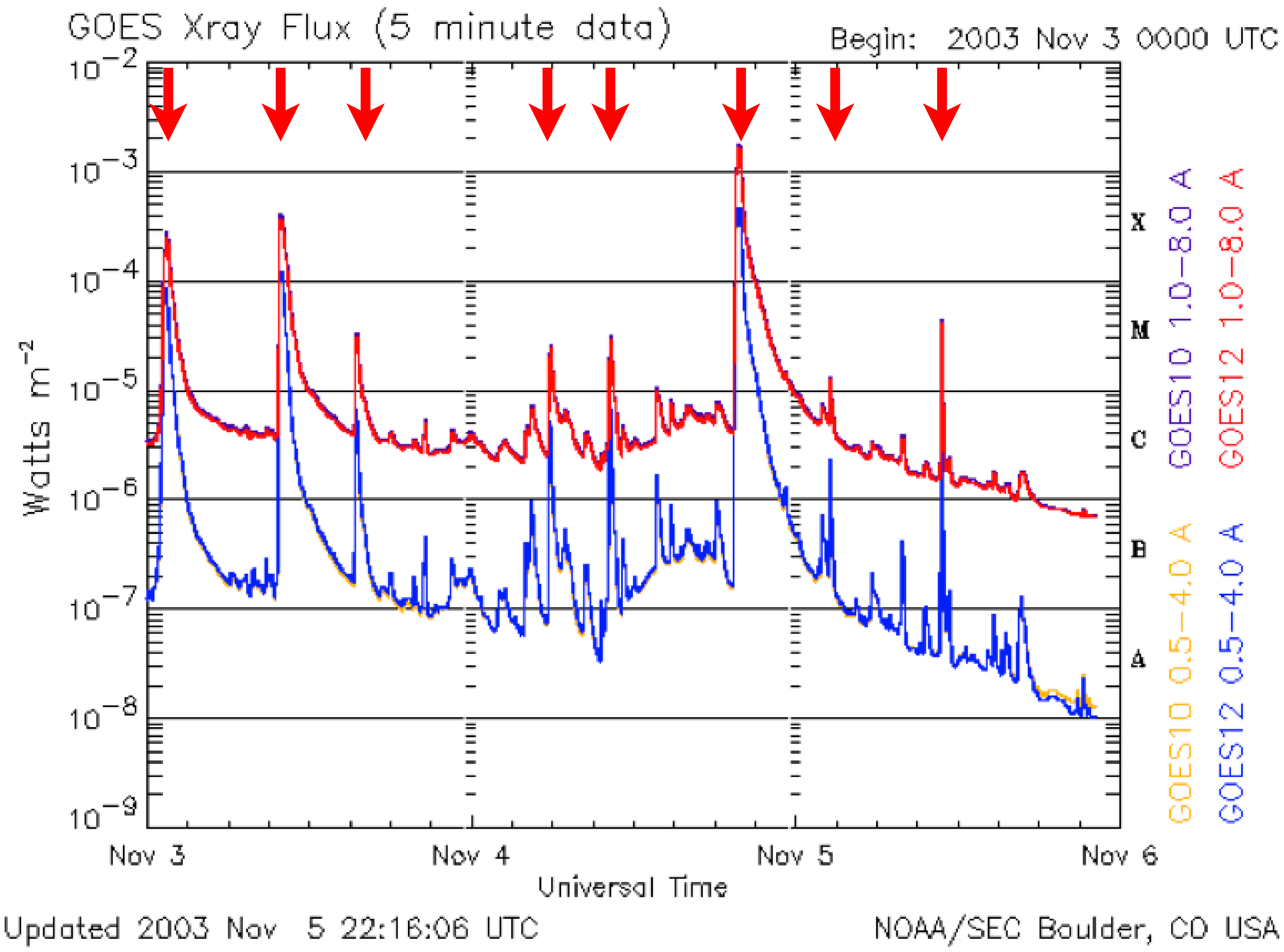
Detection and classification of solar flares in X-rays

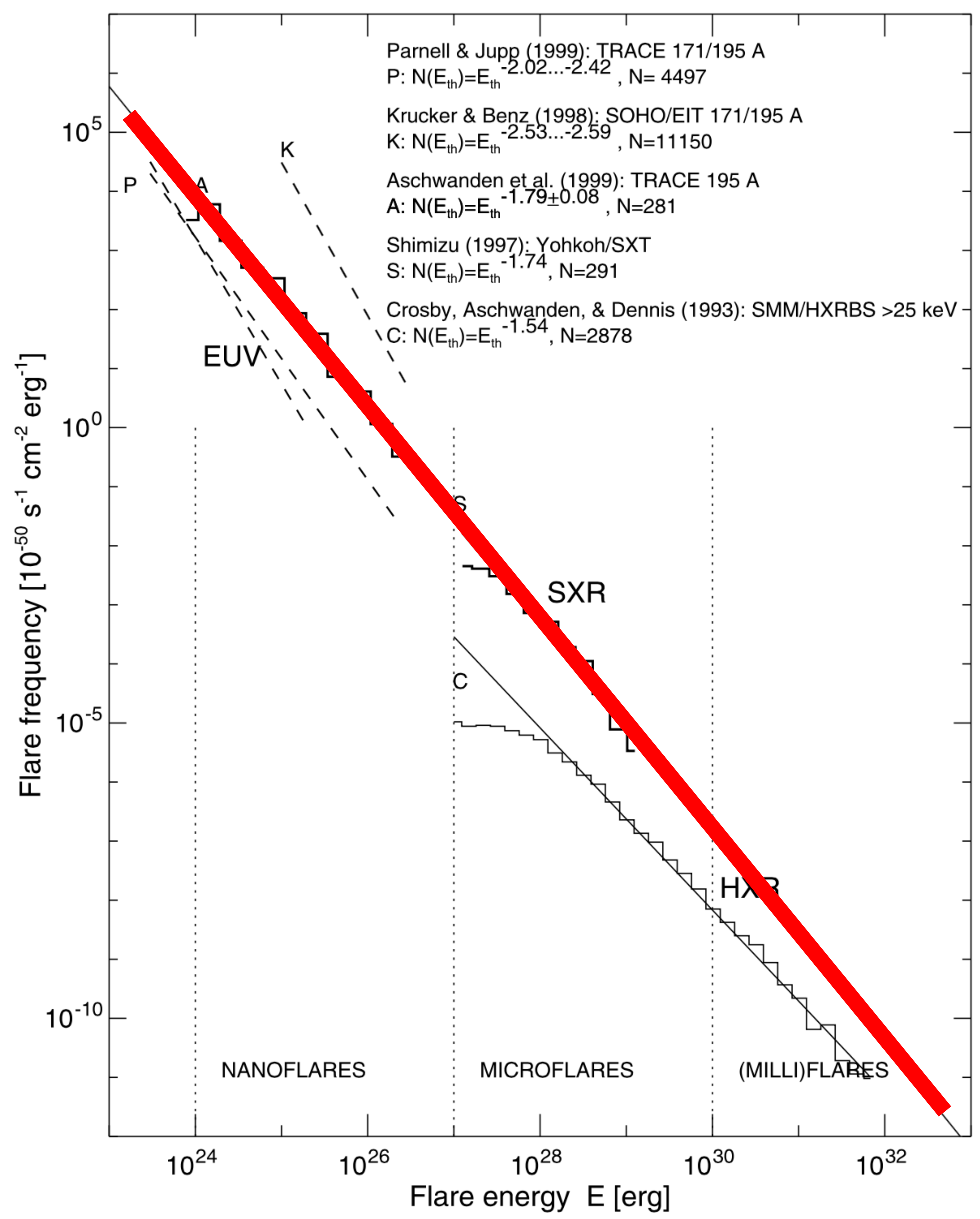




[Aschwanden+, 2014]

Detection and classification of solar flares in X-rays





[Aschwanden+, 2014]

Motivation: predicting solar flares still gives us a hard time

The Astrophysical Journal, 829:89 (32pp), 2016 October 1
doi:10.3847/0004-637X/829/2/89

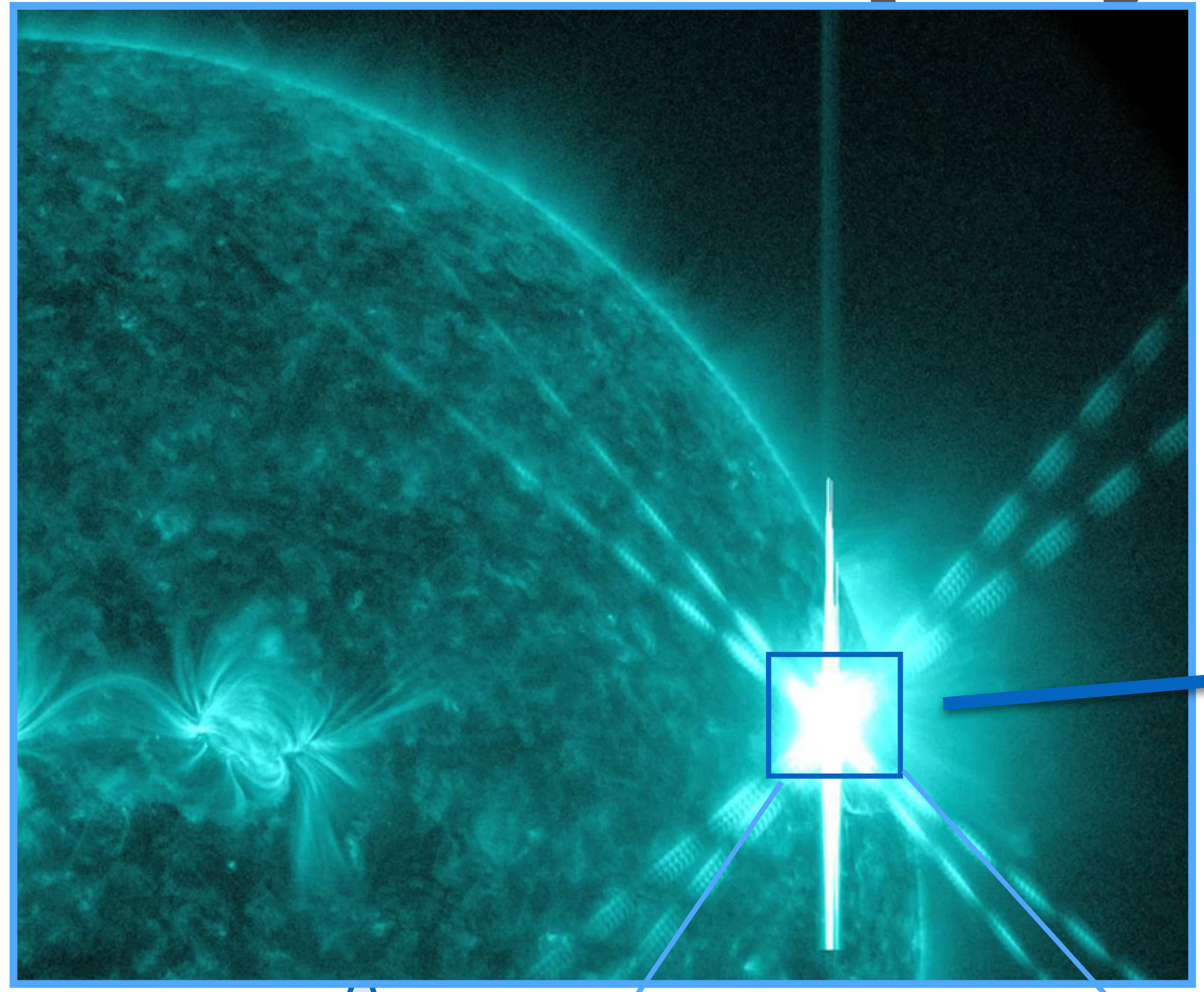
A COMPARISON OF FLARE FORECASTING METHODS. I. RESULTS FROM THE "ALL-CLEAR" WORKSHOP

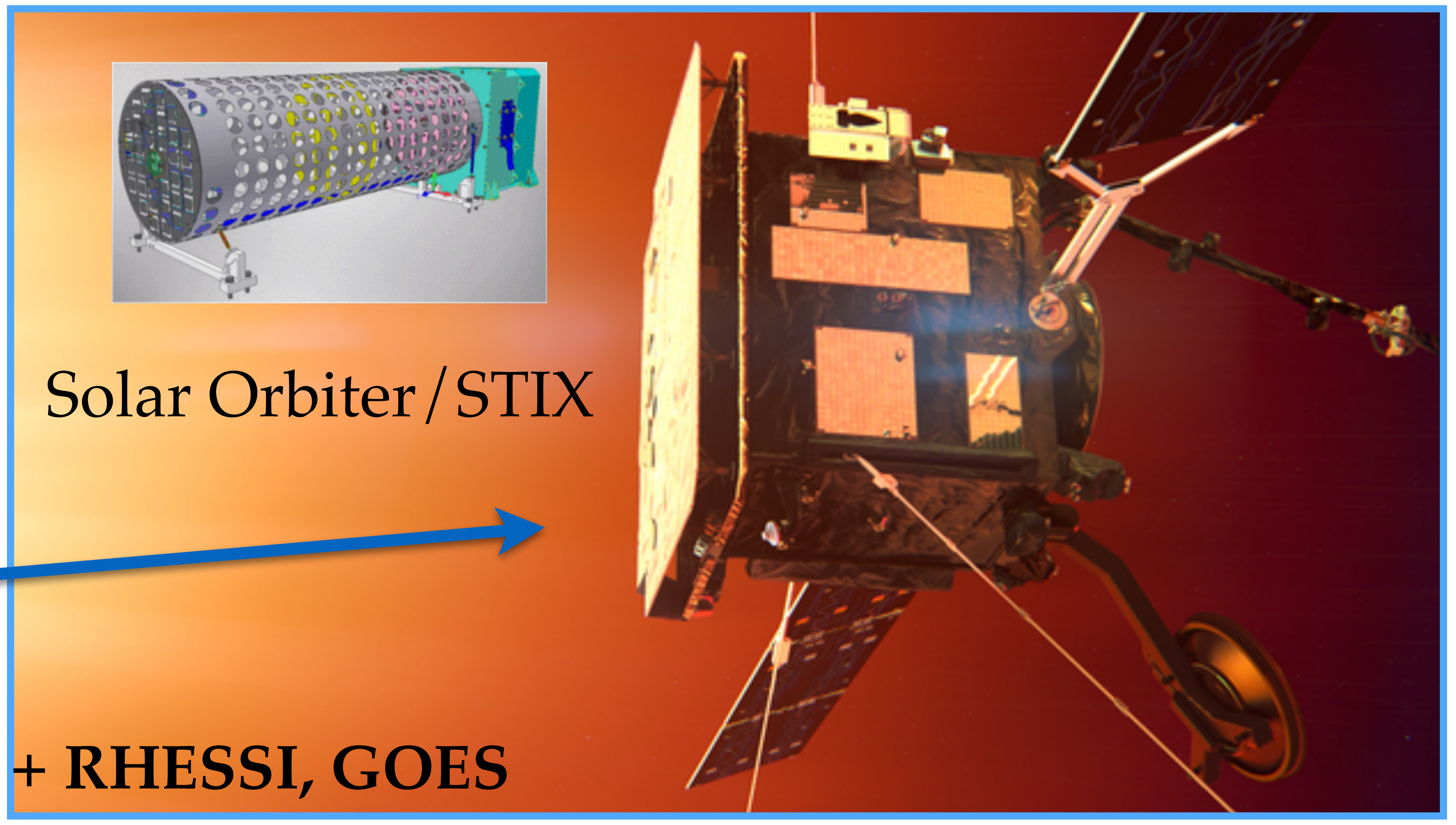
G. Barnes[1], K. D. Leka[1], C. J. Schrijver[2], T. Colak[3], R. Qahwaji[3], O. W. Ashamari[3], Y. Yuan[4], J. Zhang[5], R. T. J. McAteer[6], D. S. Bloomfield[7,14], P. A. Higgins[7], P. T. Gallagher[7], D. A. Falconer[8,9,10], M. K. Georgoulis[11], M. S. Wheatland[12], C. Balch[13], T. Dunn[1], and E. L. Wagner[1]

When a comparison was made in this fashion, no one method clearly outperformed all others, which may in part be due to the strong correlations among the parameters used by different methods to characterize an active region. For M-class flares and above, the set of methods tends toward a weakly positive skill score (as measured with several distinct metrics), with no participating method proving substantially better than climatological forecasts.

Parameter/	Statistical	C1.0+, 24 hr		M1.0+, 12 hr		M5.0+, 12 hr	
Method	Method	ApSS	BSS	ApSS	BSS	ApSS	BSS
B_{eff}	Bayesian	0.12	0.06	0.00	0.03	0.00	0.02
ASAP	Machine	0.25	0.30	0.01	−0.01	0.00	−0.84
BBSO	Machine	0.08	0.10	0.03	0.06	0.00	−0.01
WL_{SG2}	Curve fitting	N/A	N/A	0.04	0.06	0.00	0.02
NWRA MAG 2-VAR	NPDA	0.24	0.32	0.04	0.13	0.00	0.06
$\log(\mathcal{R})$	NPDA	0.17	0.22	0.01	0.10	0.02	0.04
GCD	NPDA	0.02	0.07	0.00	0.03	0.00	0.02
NWRA MCT 2-VAR	NPDA	0.23	0.28	0.05	0.14	0.00	0.06
SMART2	CCNN	0.24	−0.12	0.01	−4.31	0.00	−11.2
Event Statistics, 10 prior	Bayesian	0.13	0.04	0.01	0.10	0.01	0.00
McIntosh	Poisson	0.15	0.07	0.00	−0.06	N/A	N/A

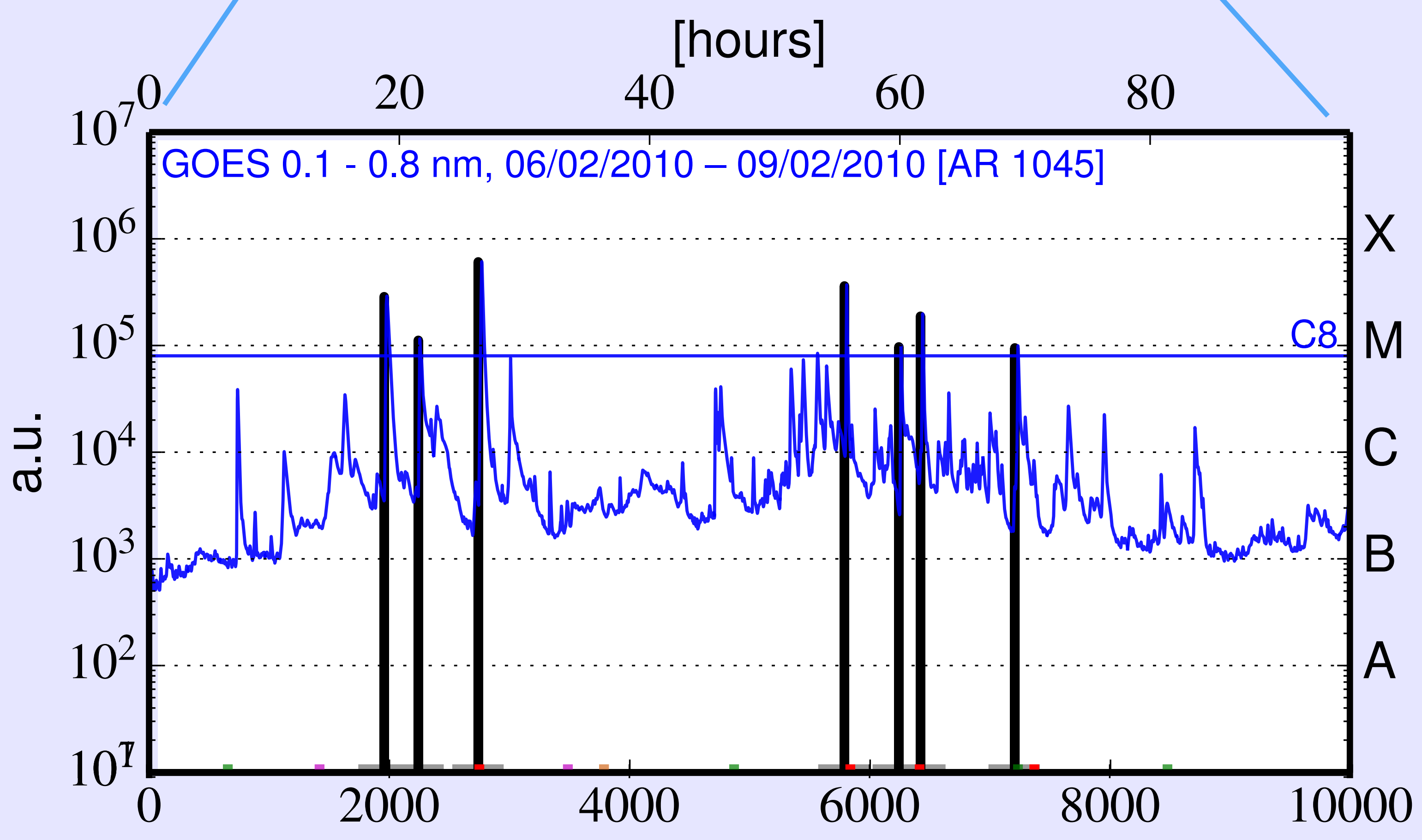
The FlarePredict project

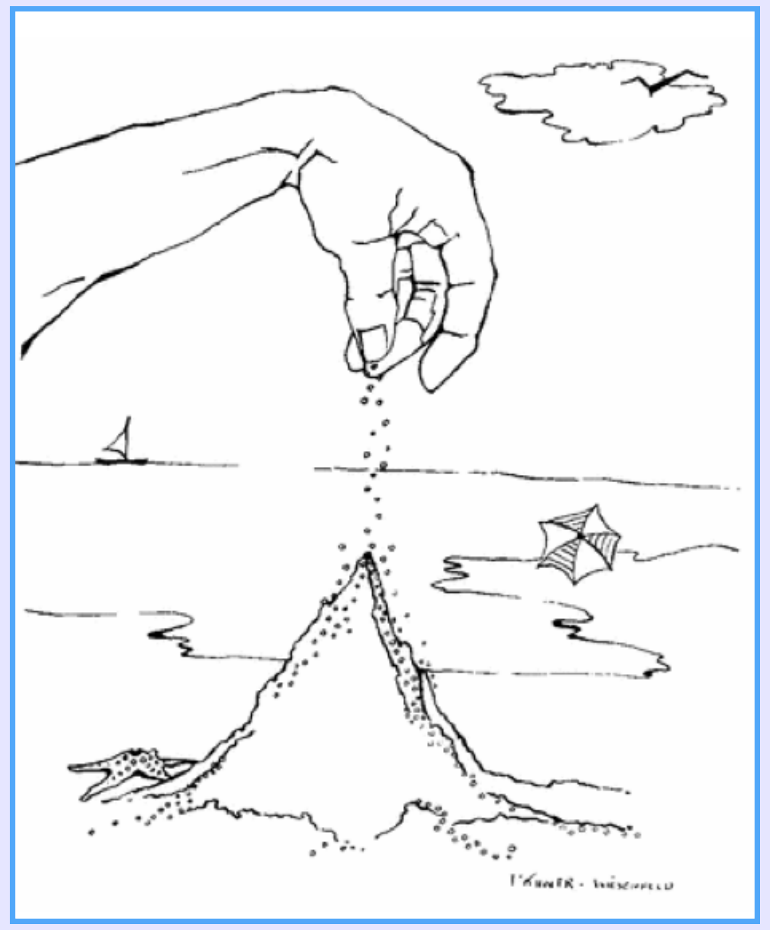




Towards real time prediction

Data assimilation + machine learning





DamieNN

(basé sur TensorFlow)

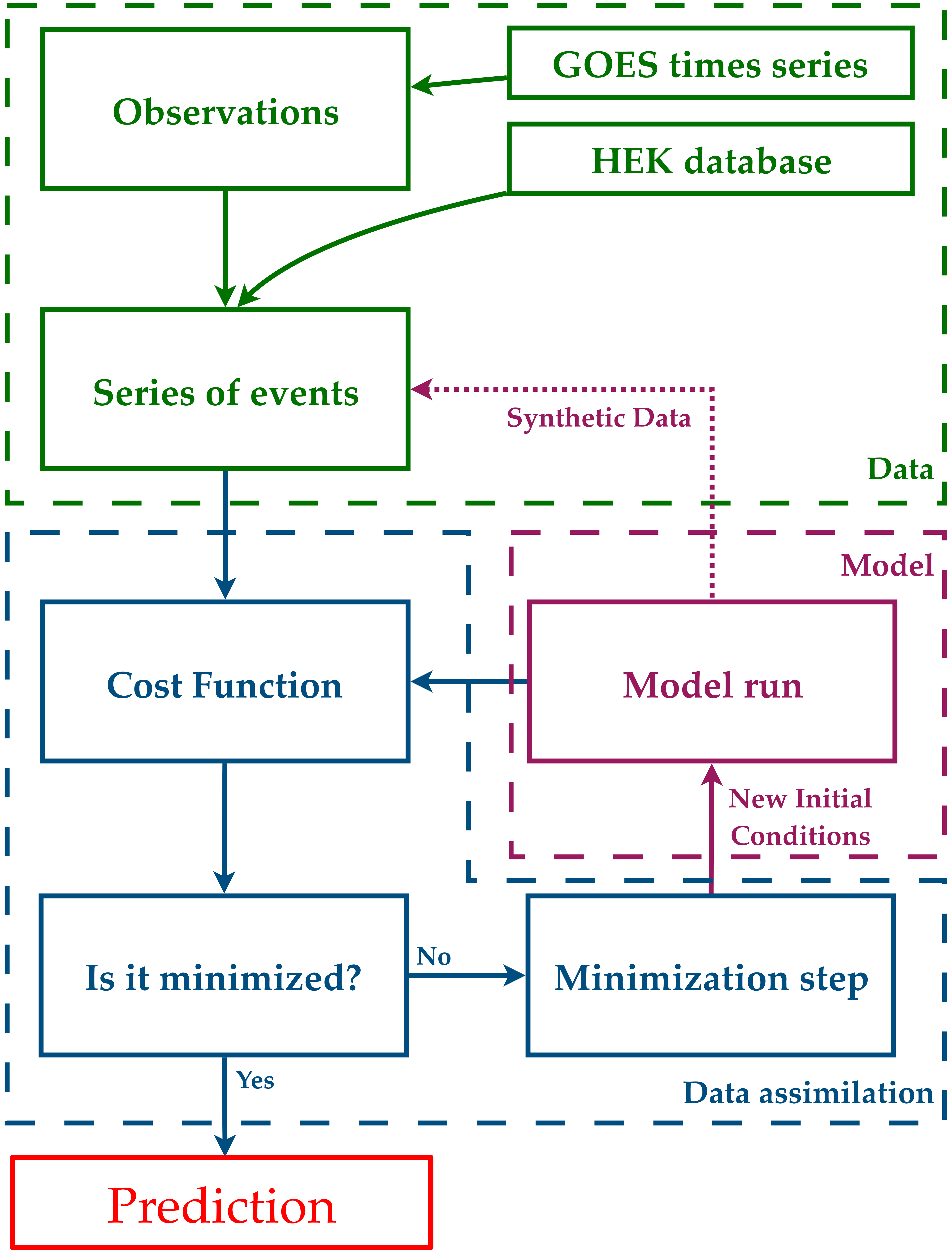
Present FlarePredict workflow

Data: construction of a discrete time-series of events

Model: sandpile models

Data assimilation: simulated annealing

Prediction of large events



Present FlarePredict workflow

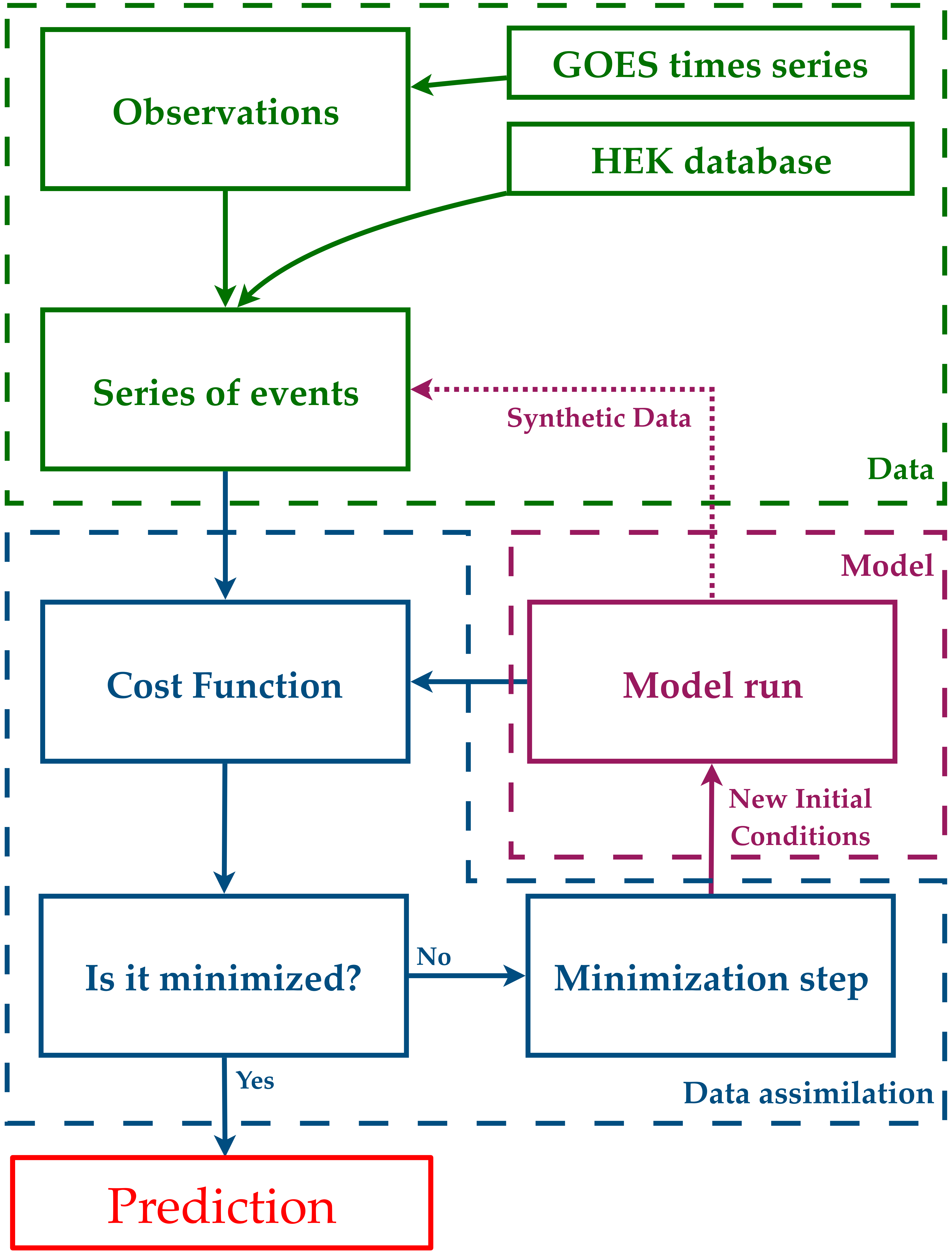
Data: construction of a discrete time-series of events

Model: sandpile models

Data assimilation: simulated annealing

Too slow for very large parameter exploration

Prediction of large events



Present FlarePredict workflow

Data: construction of a discrete time-series of events

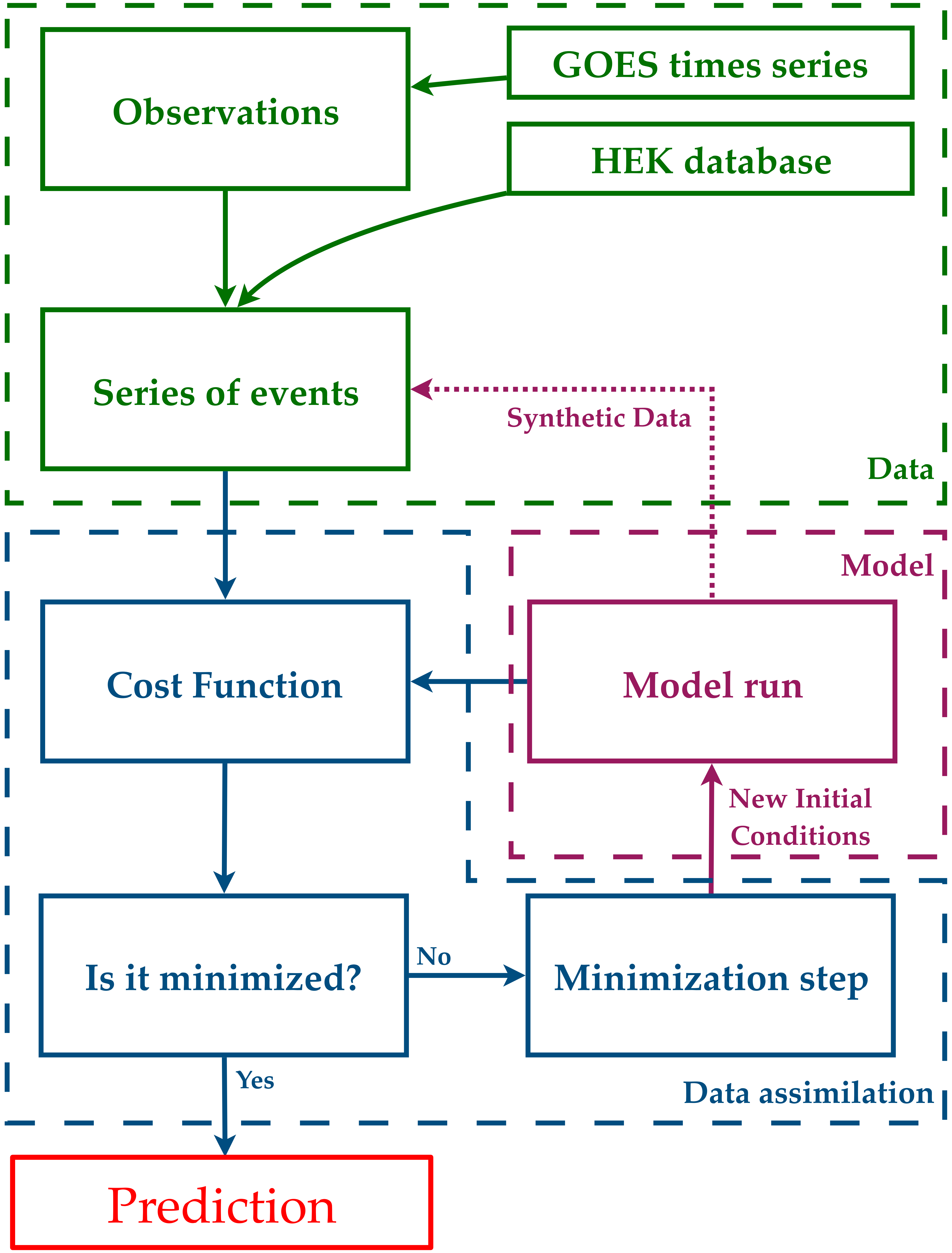
Model: sandpile models

Data assimilation: simulated annealing

Too slow for very large parameter exploration

Prediction of large events

Not proved to work so far



FlarePredict tasks

Task I

Goal: speed-up assimilation process

Mean: machine learning on GPUs

Status:

- workflow redeveloped in python/keras
- Arrival of GPU server very late (received on November 2022, but machine is not booting, waiting for replacement parts…)
- Visit from H. Lamarre (Spring 2022): tests with simpler approach (not successful)

Task II

Goal: proof-of-concept for real prediction

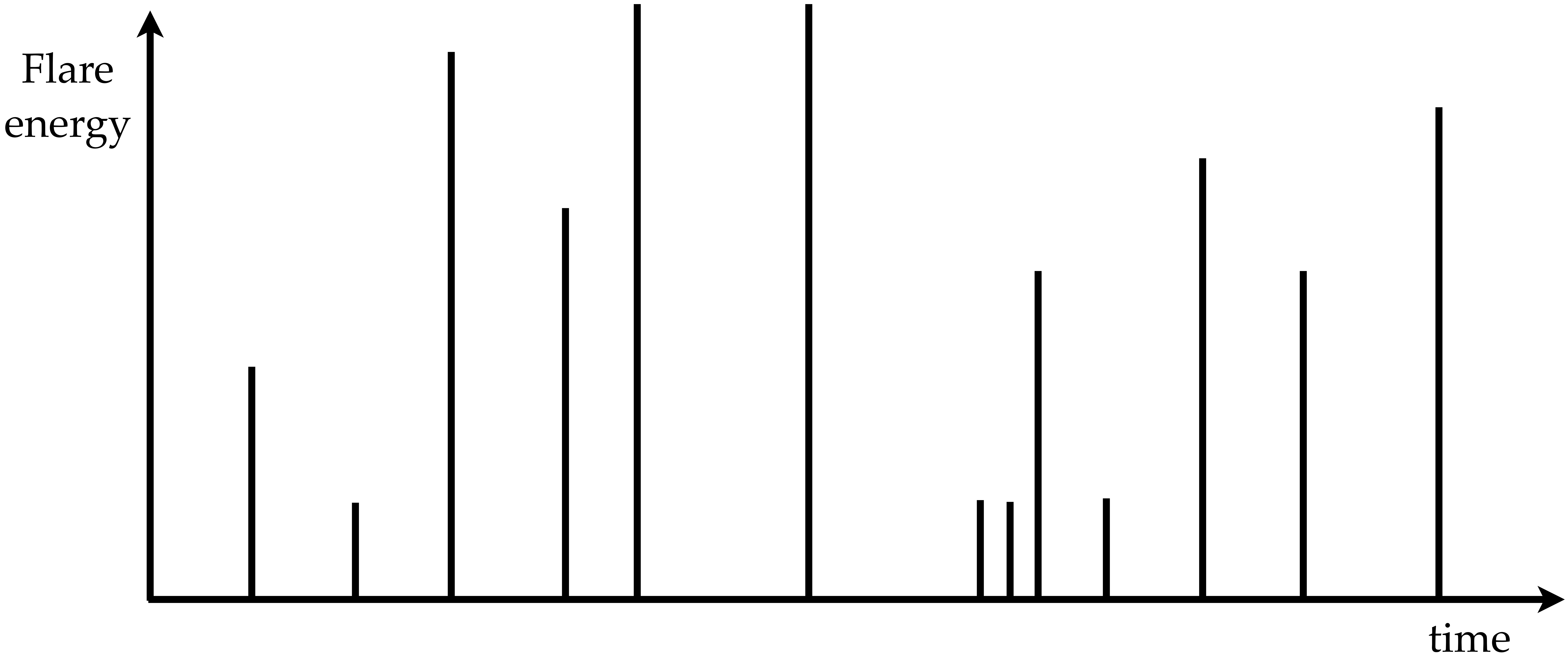
Mean: ensemble forecasts

Status:

- Predictions proved to be possible for both synthetic and real data!
- Current tests (H. Lamarre): largest sample of real solar flares

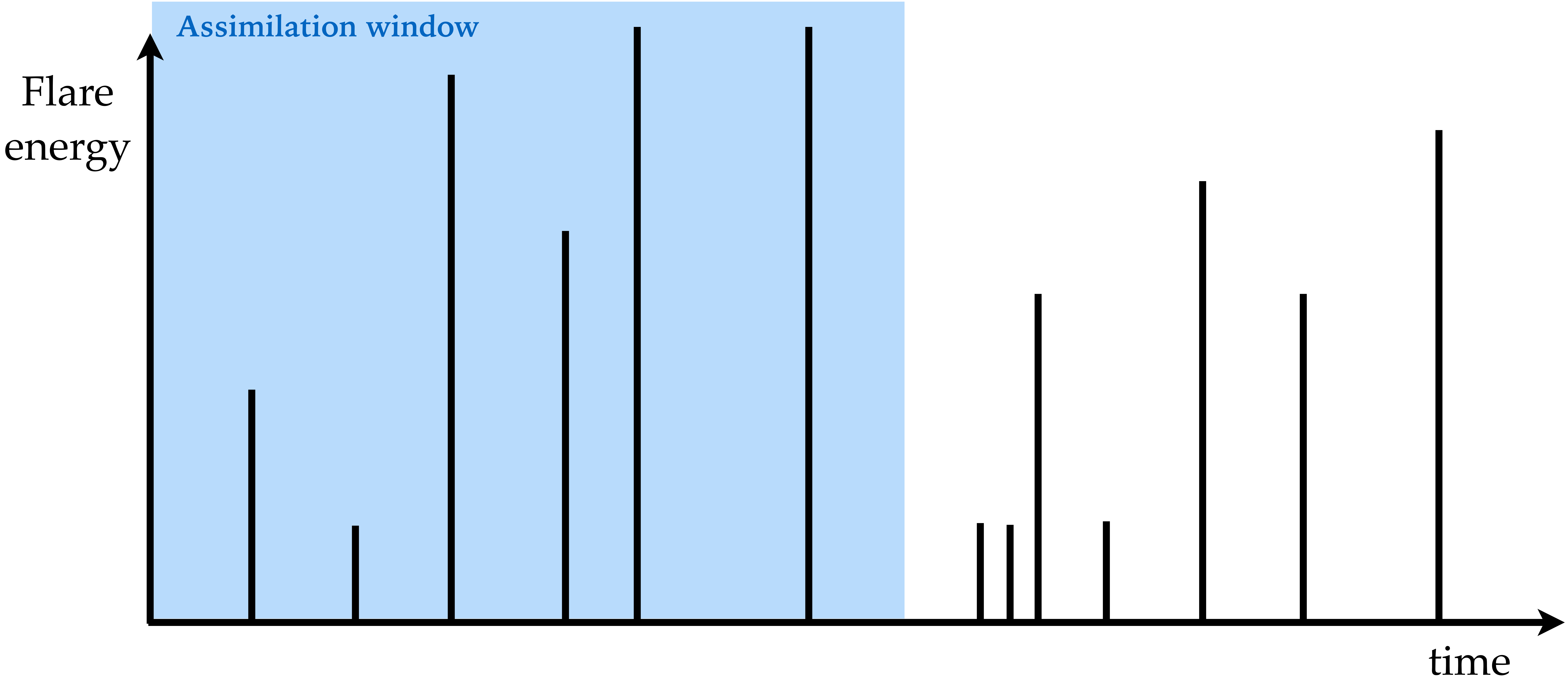
Thibeault+ 2022, Solar Physics, 297
Lamarre+ 2022, in prep

Task II: the All-Clear Forecast (ACF) test

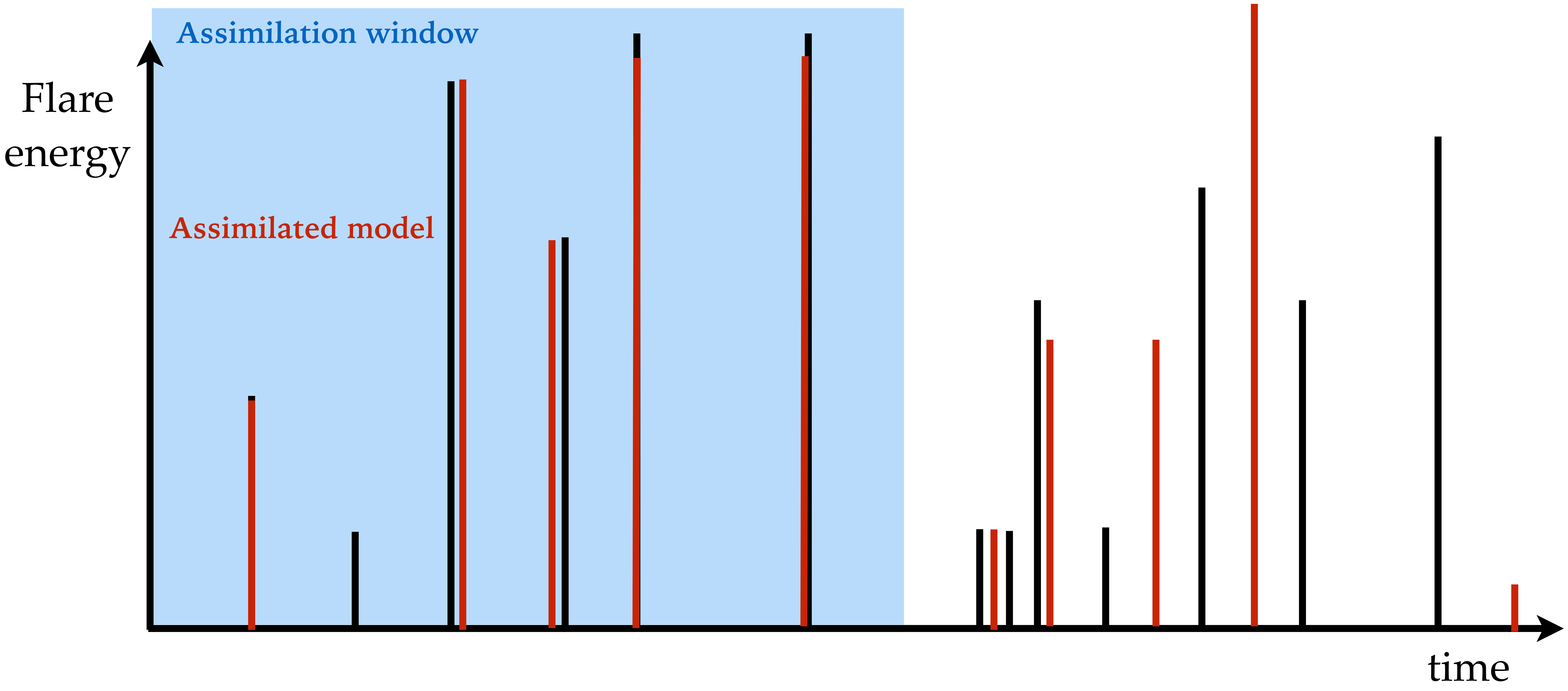


[Thibeault+ 2022]

Task II: the All-Clear Forecast (ACF) test



Task II: the All-Clear Forecast (ACF) test

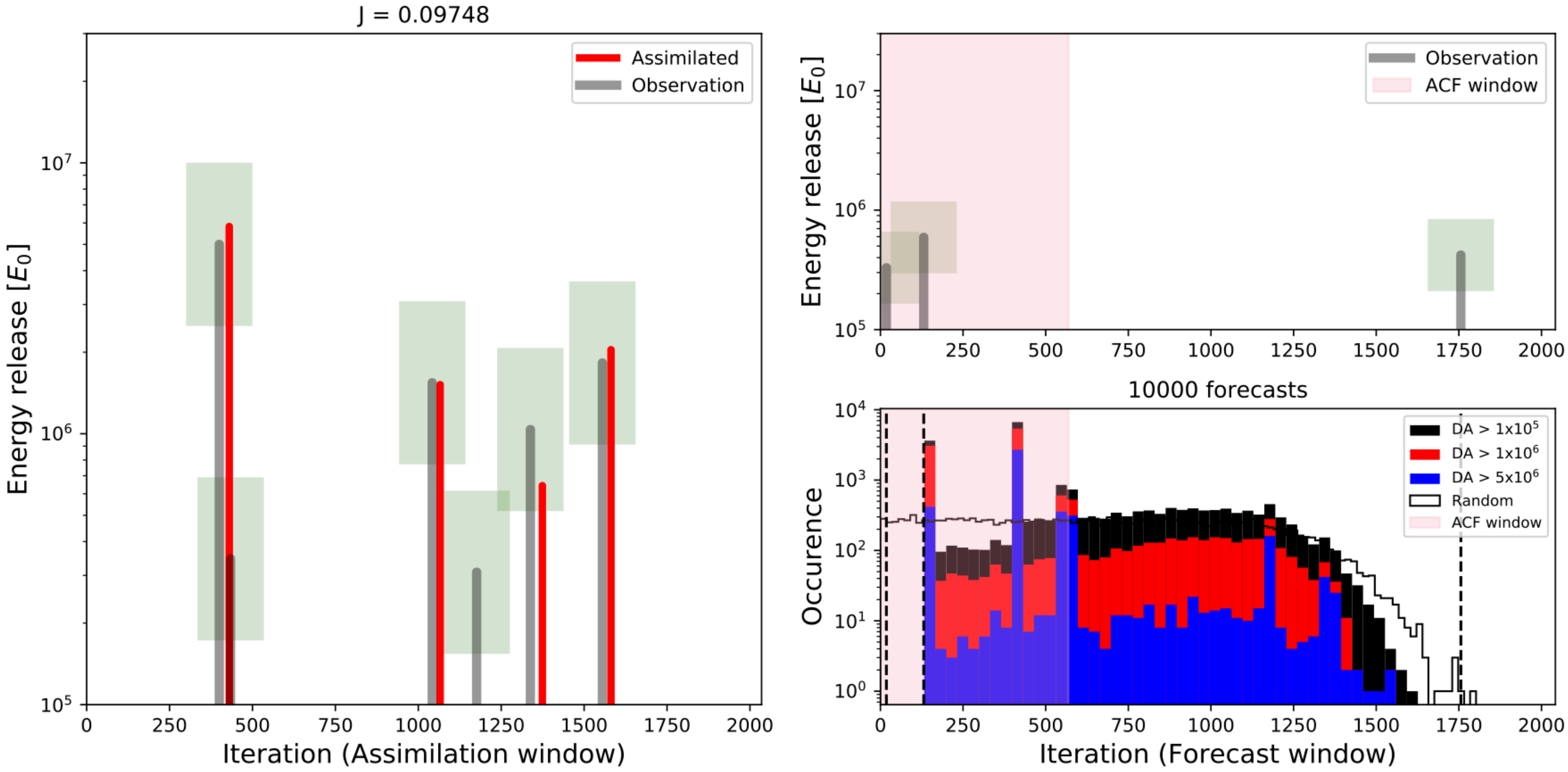


Task II: the All-Clear Forecast (ACF) test

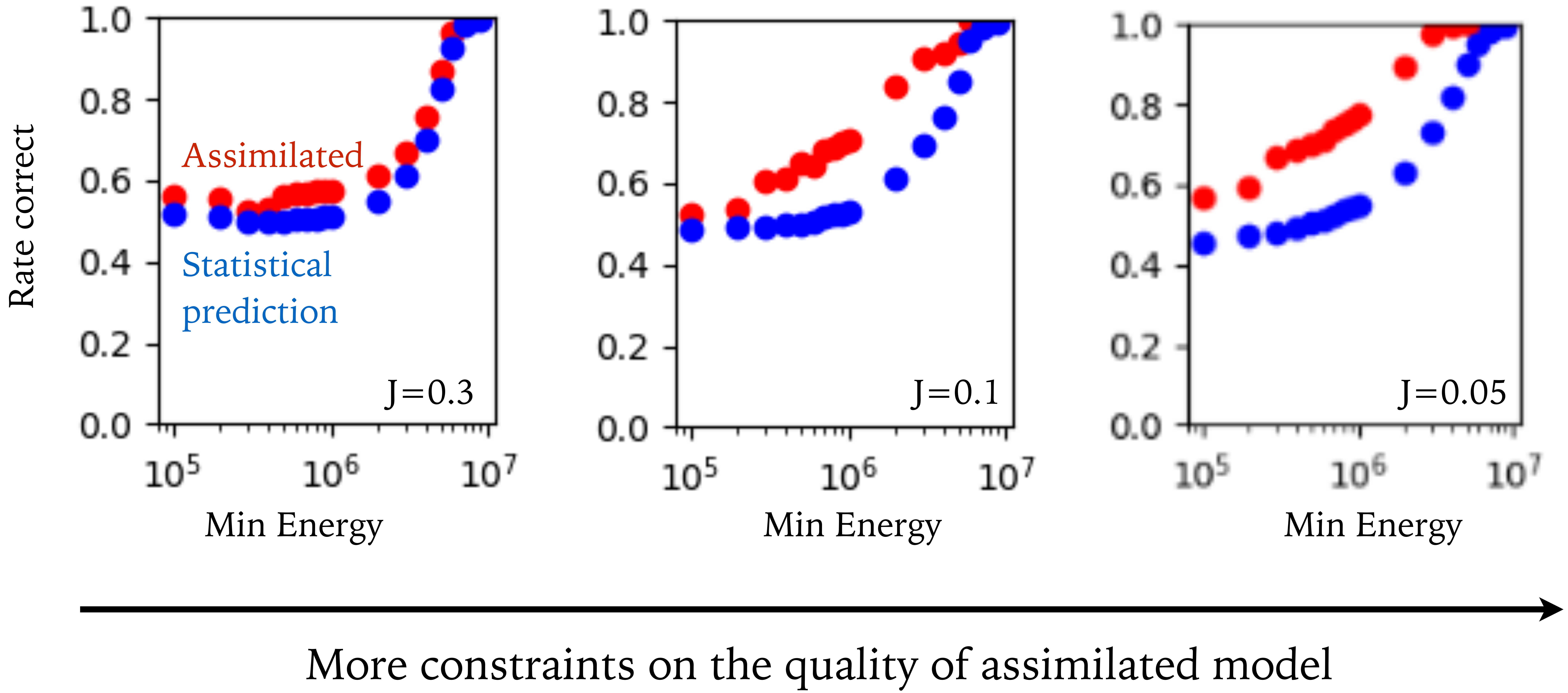
Task II: the All-Clear Forecast (ACF) test

All-Clear forecast: is there an event larger than a given threshold in the AC window?

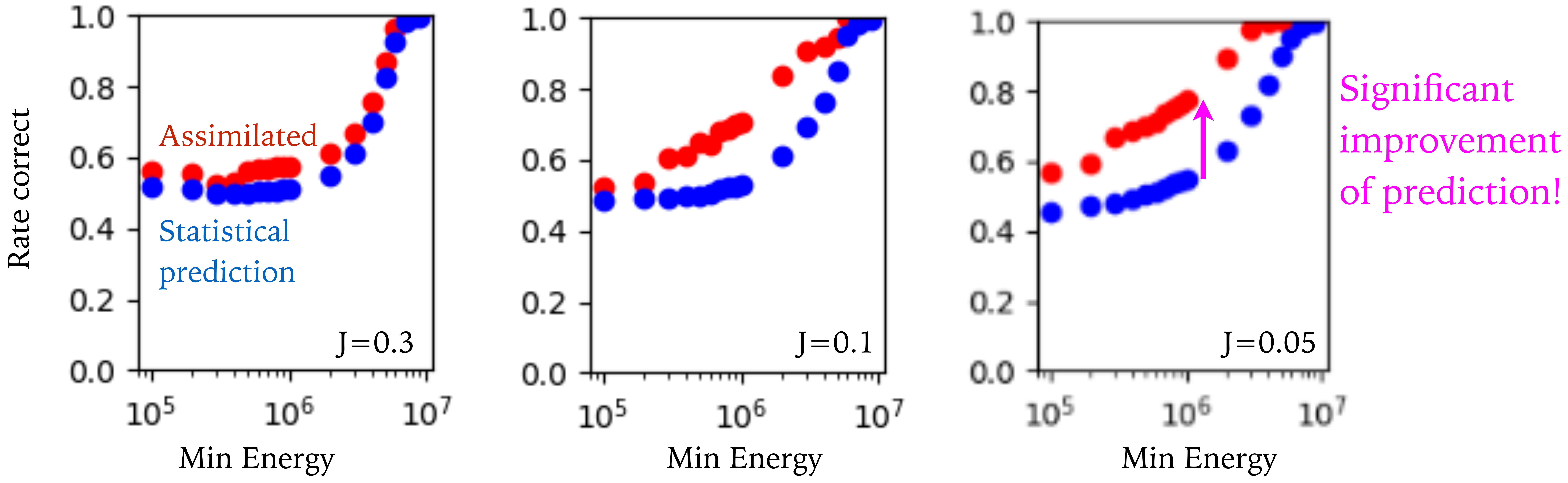
Task II: results for the All-Clear Forecast (ACF) test



Task II: improvements thanks to data-assimilation

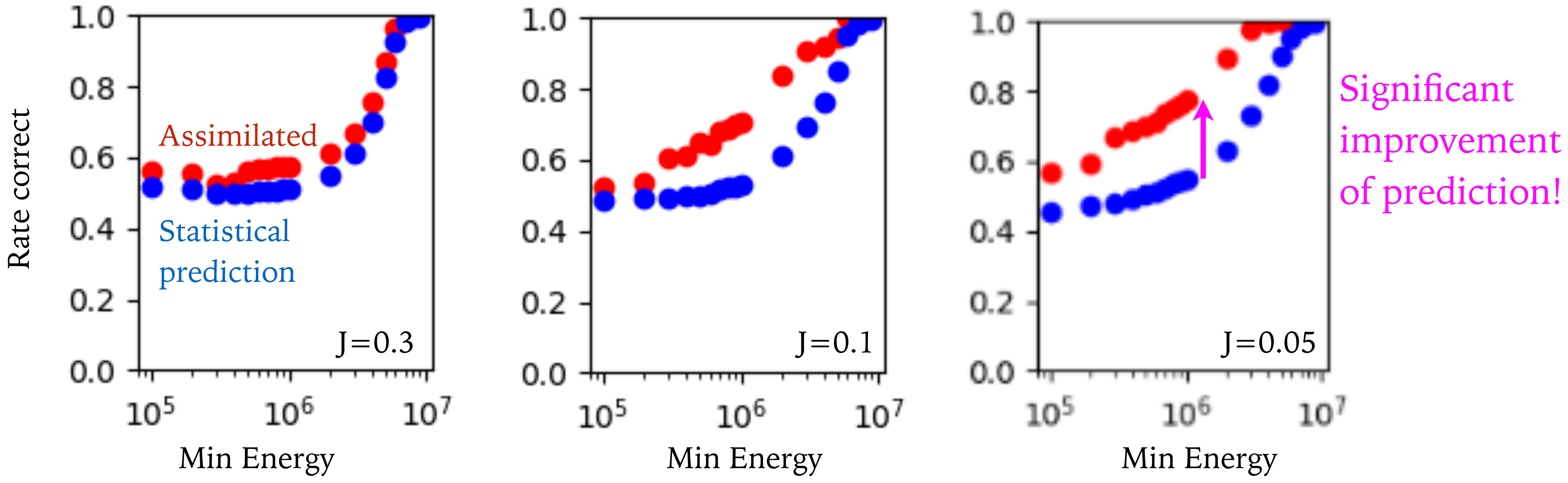


Task II: improvements thanks to data-assimilation



More constraints on the quality of assimilated model

Task II: improvements thanks to data-assimilation

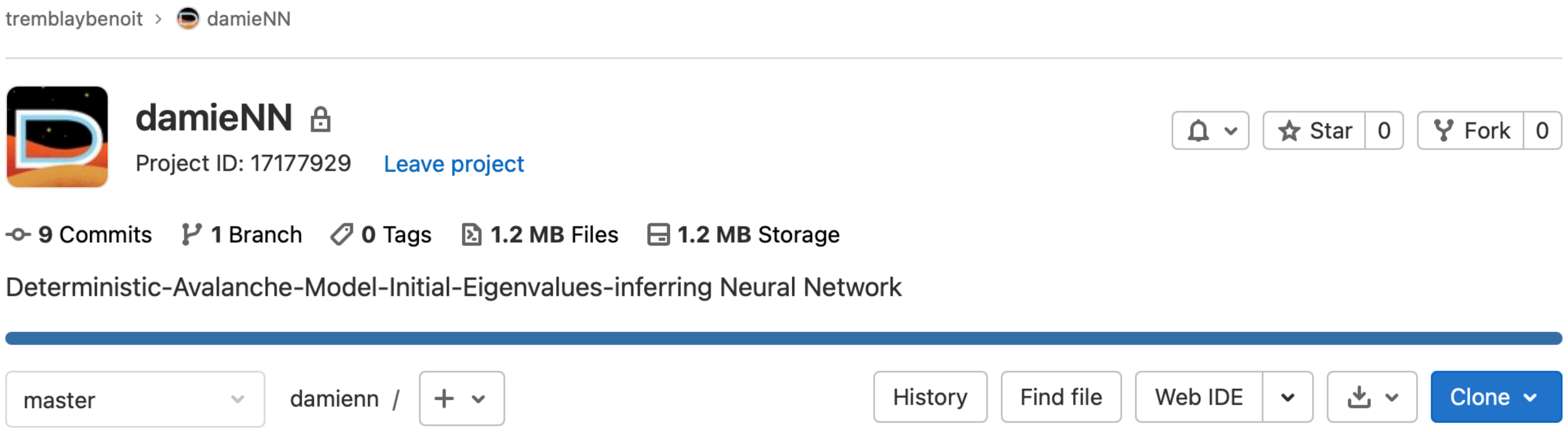


More constraints on the quality of assimilated model

Validated on a syntethic sample (100 independent time series) and 10 GOES time series

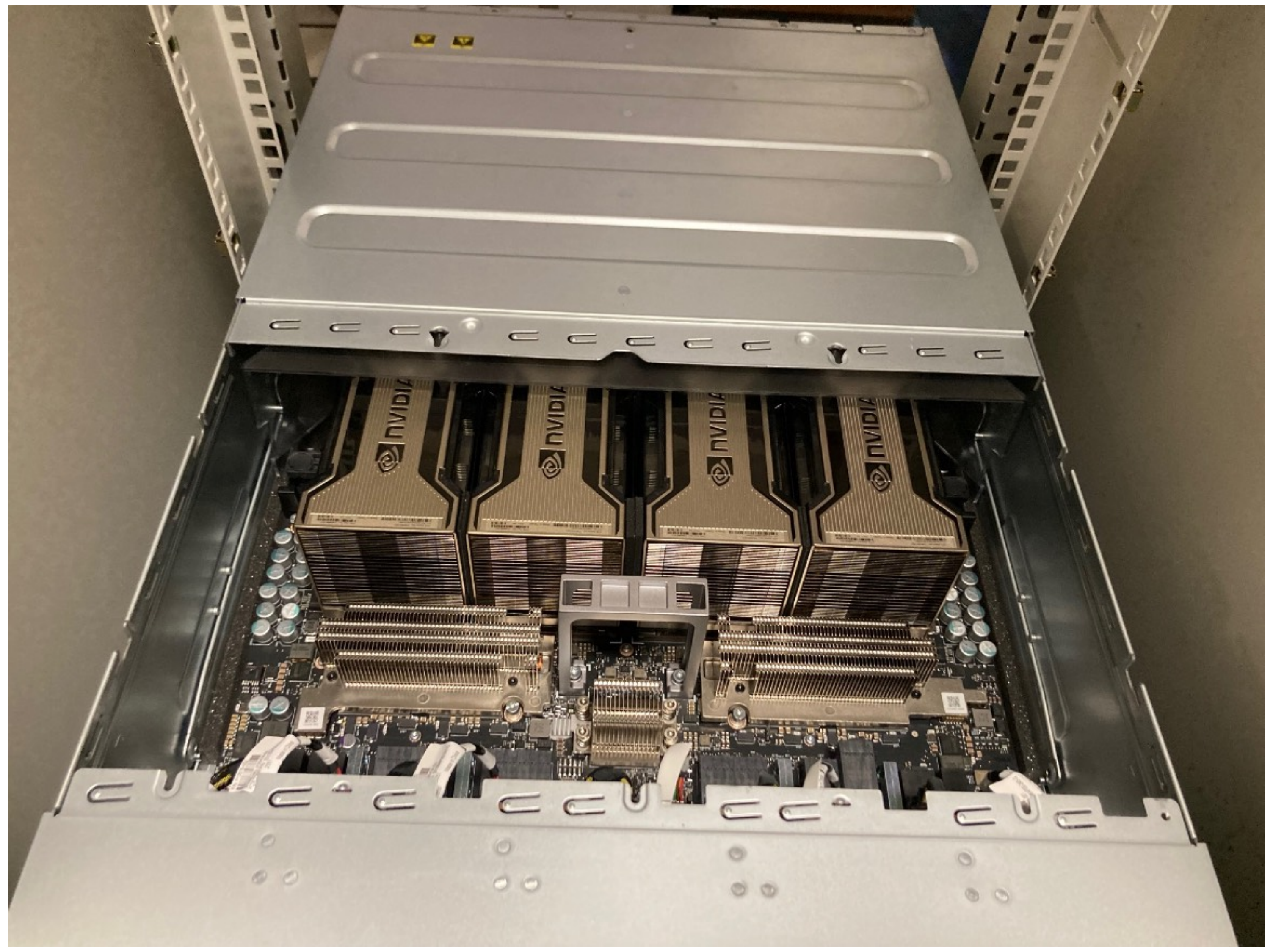
Task I: status and plan

Delayed due to COVID



First operational version developed, fully rewritten with keras (TensorFlow-based, in full python)

Visit of B. Tremblay in march 2021, with whom I developed this new version



Acquisition of GPU server (partially financed through this project):

- 8 GPUs A100 (80Gb) connected through NVlink
- Platform Apollo (HPE vendor) selected after consulting various vendors

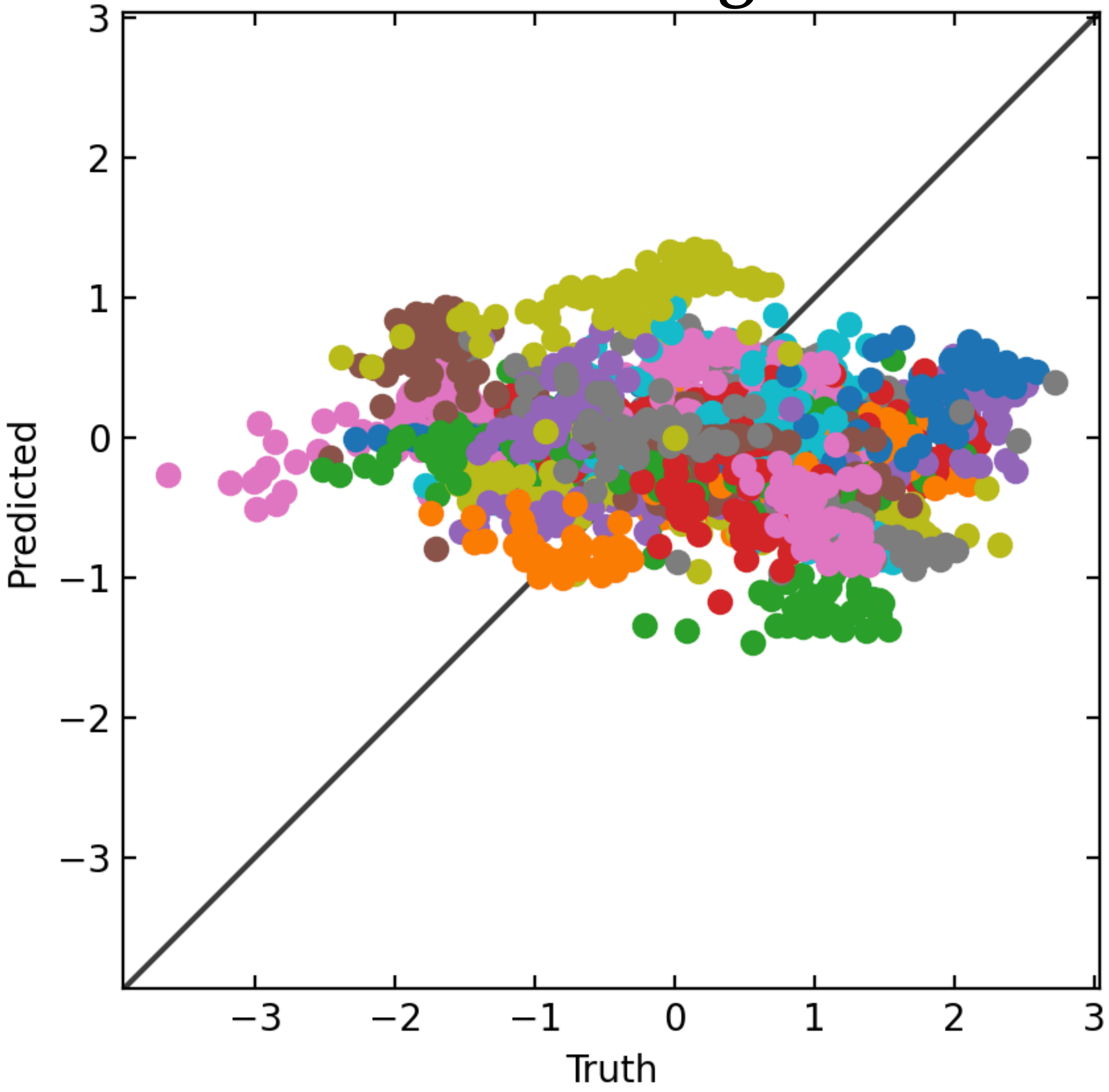
Machine delivered but it does not boot… waiting for replacement parts from HPE

Task I: simpler approach waiting for the GPU server

Generic approach followed:

- Generate two independent set of model realisation, one used for training, the other for validation
- Generic ML algorithms tested on the sample (Random Forests, Regression Trees, Multi-layer Perceptron)
- **None of the tested approach present promising results**: the initial plan is still the most promising, we hope to complete it next year as a follow-up to the P2IO FlarePredict project.

Random Forest Regressor



Conclusions and future for FlarePredict

We have developed **stochastic models** with enough memory to have **predictive capabilities**

These models are **tailored to reproduce solar flare statistics**, and can **leverage data assimilation**

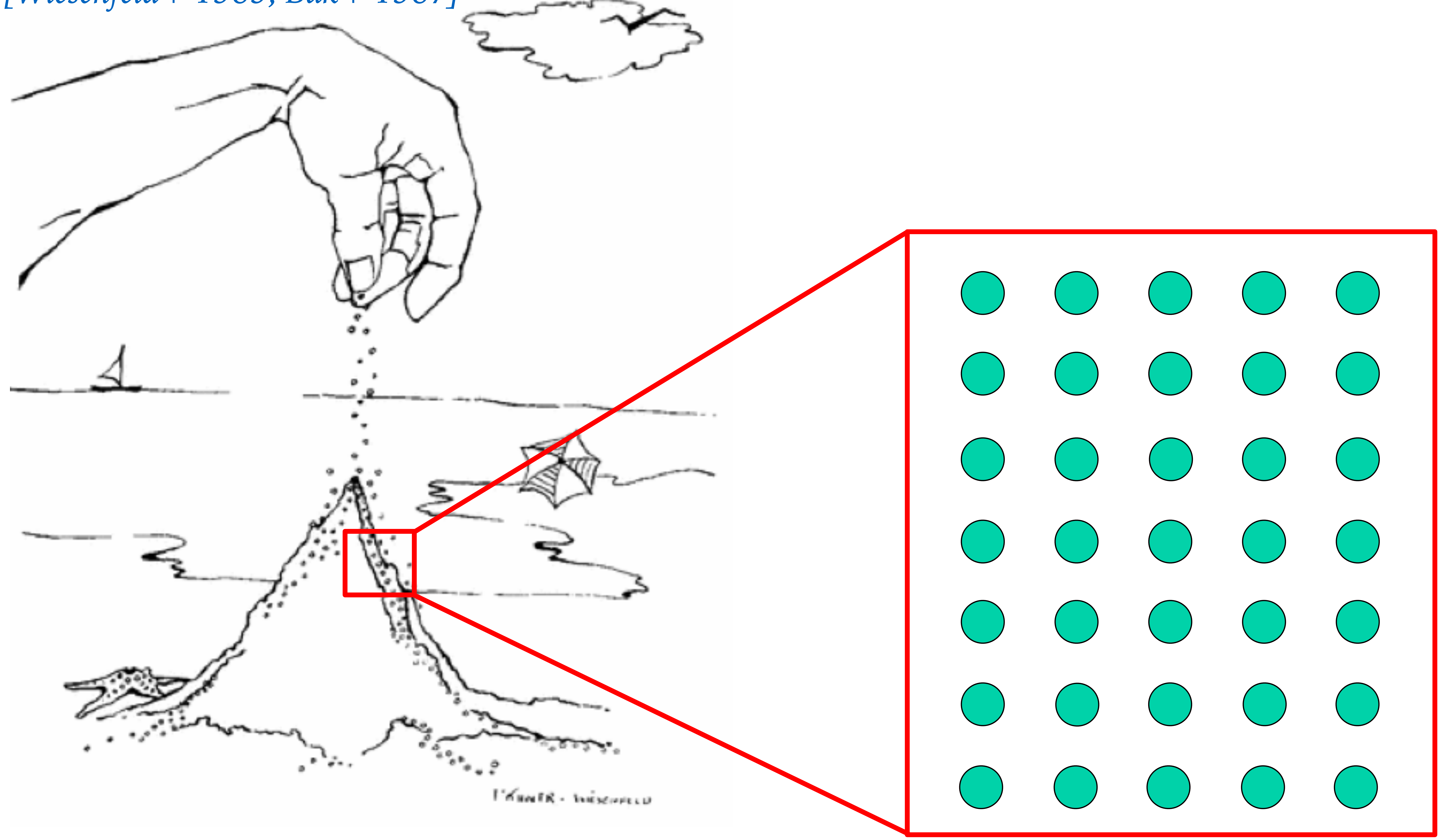
With an inefficient (in computing time) assimilation, we have shown that **we can predict large and rare synthetic events** on a large sample. We have found the same evidence for a limited real-data sample of solar flares, which we need to extent

The next phase of the project is devoted to the **data assimilation leveraging machine-learning where the model is exposed itself to the full convolutional neural network**, which should help exploring a larger sample and validate our original methodology

Basics of sandpile models

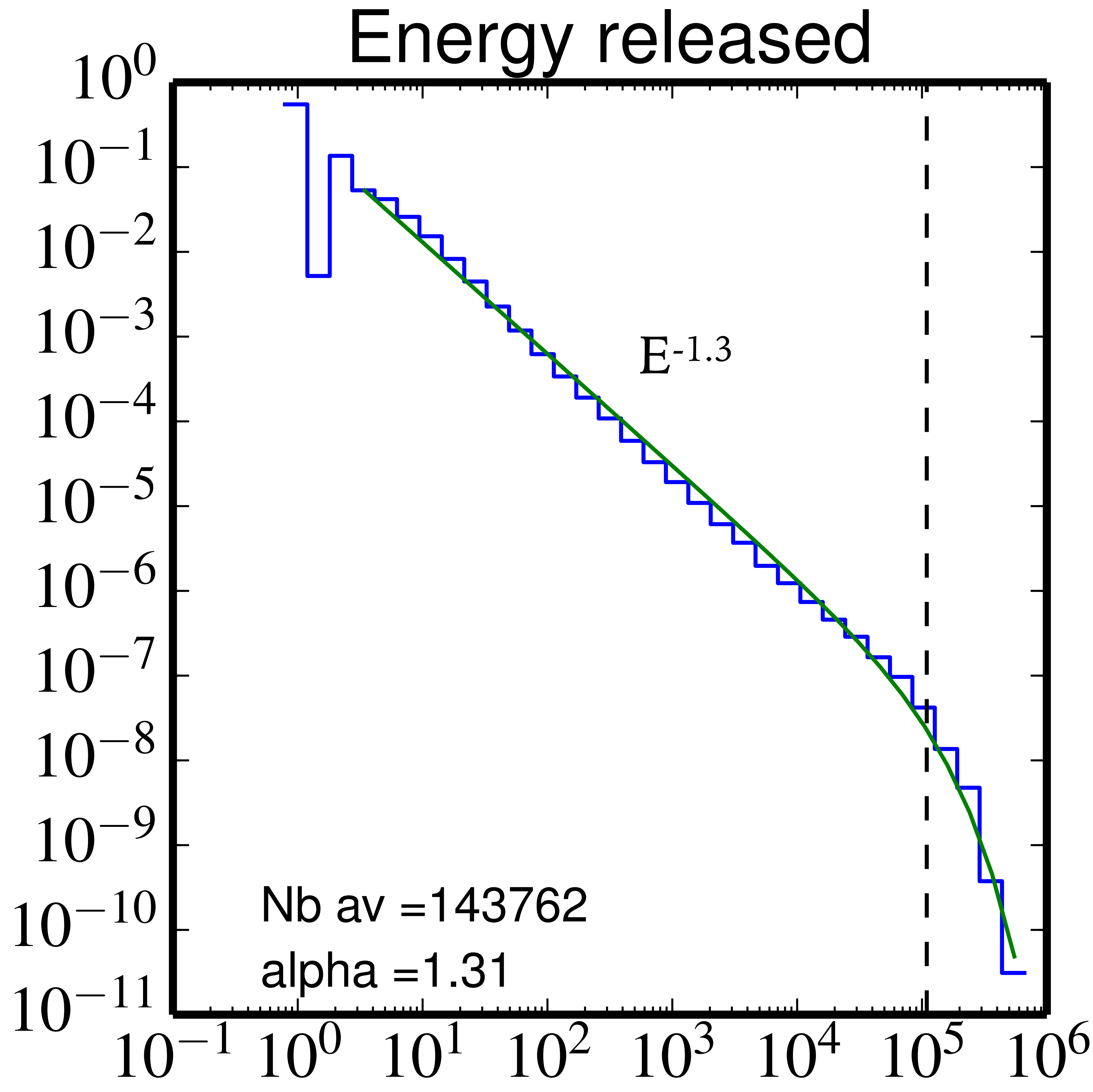
Driver

[Wiesenfeld+ 1989; Bak+ 1987]



‘Self-Organized Criticality’

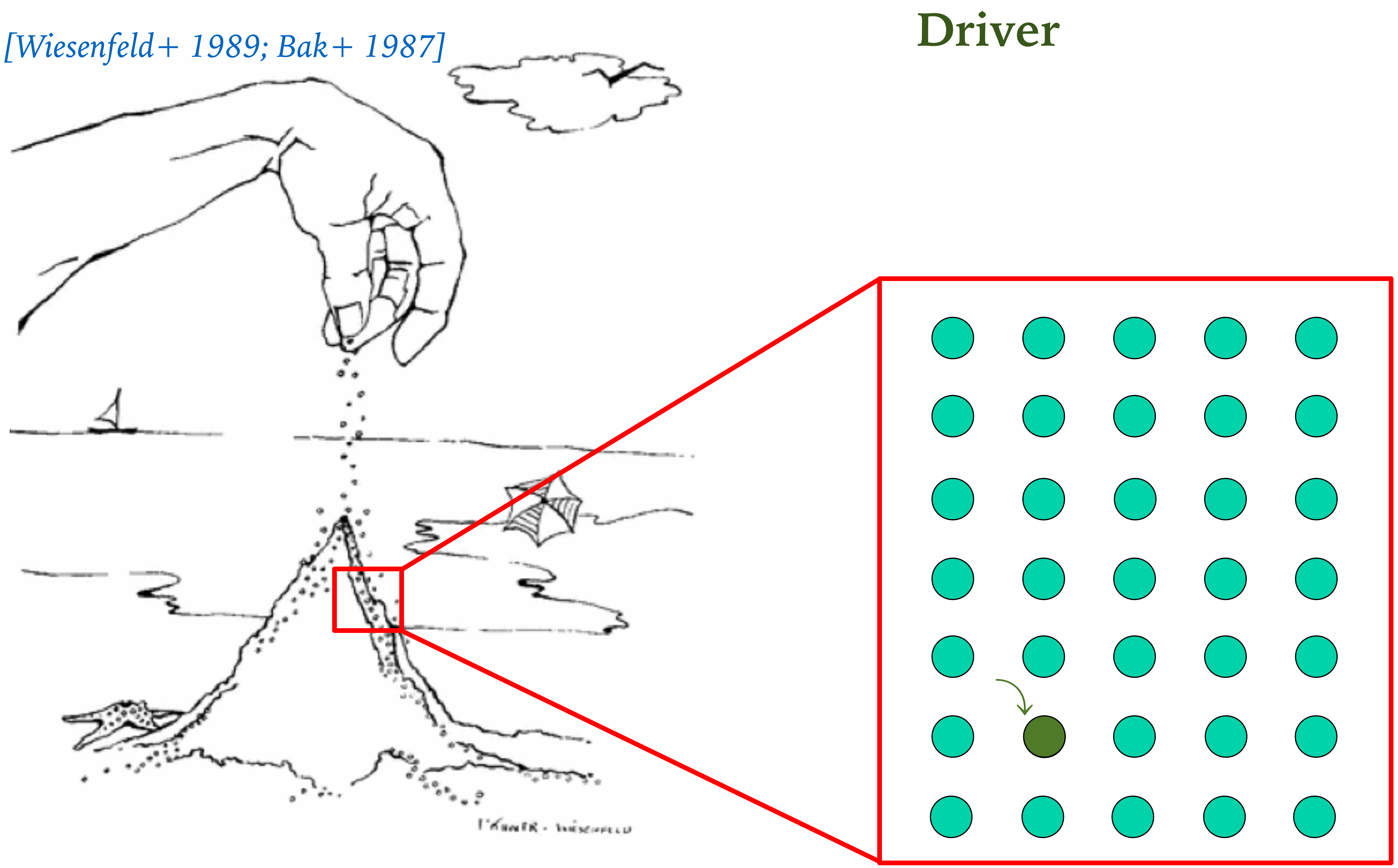
[Lu & Hamilton 1993]



[Strugarek+ 2014a,b]

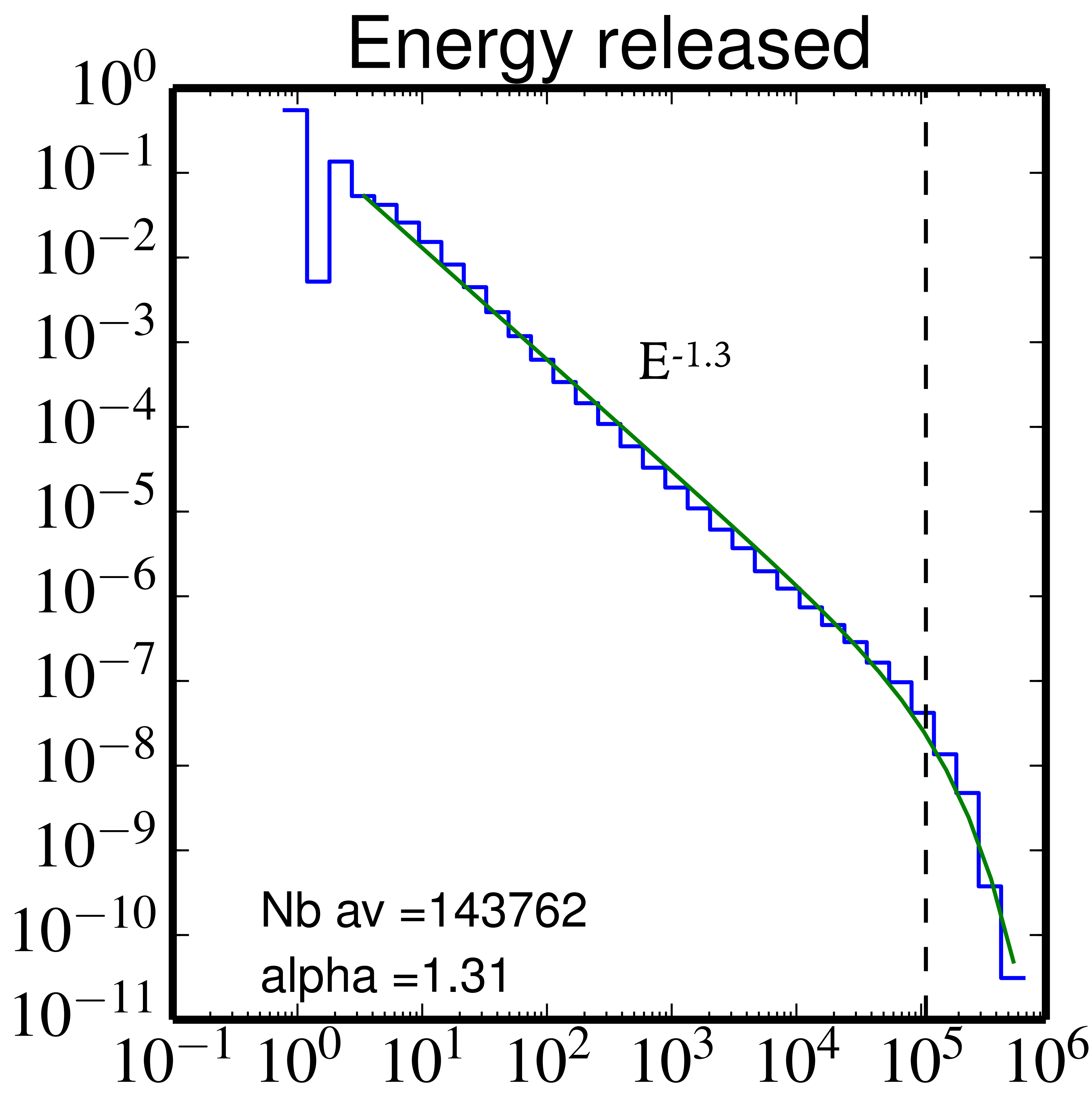
Basics of sandpile models

[Wiesenfeld+ 1989; Bak+ 1987]



'Self-Organized Criticality'

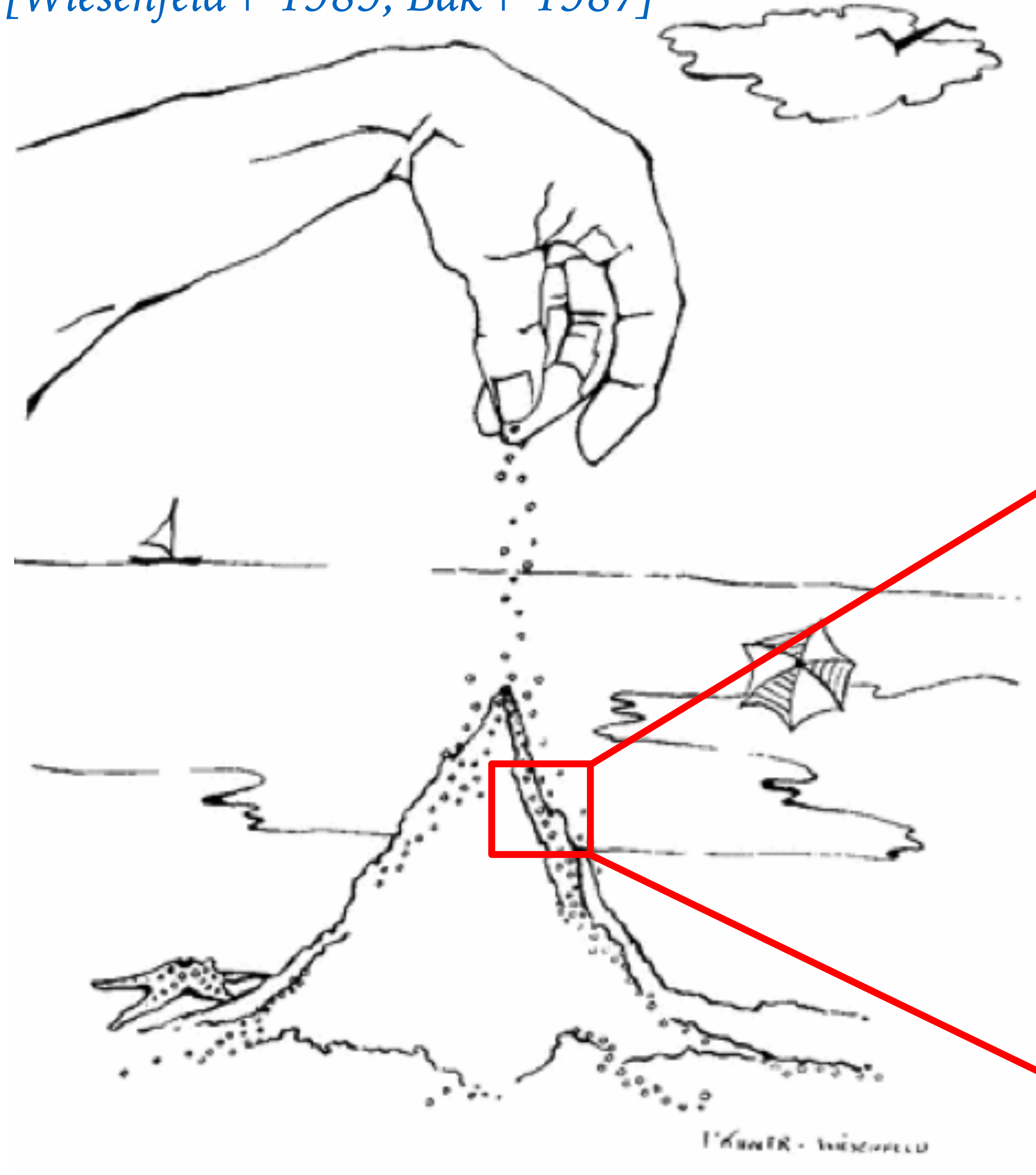
[Lu & Hamilton 1993]



[Strugarek+ 2014a,b]

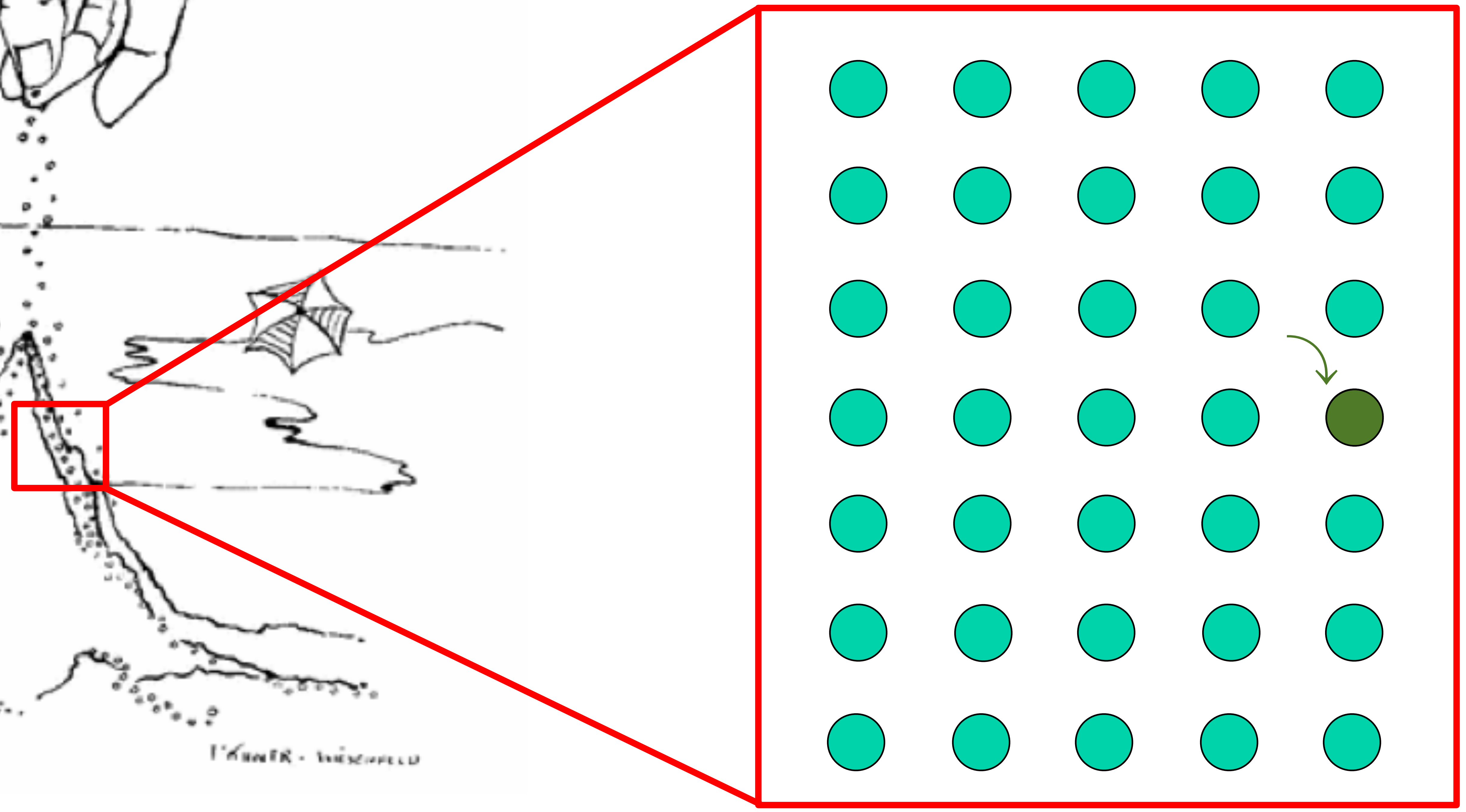
Basics of sandpile models

[Wiesenfeld+ 1989; Bak+ 1987]

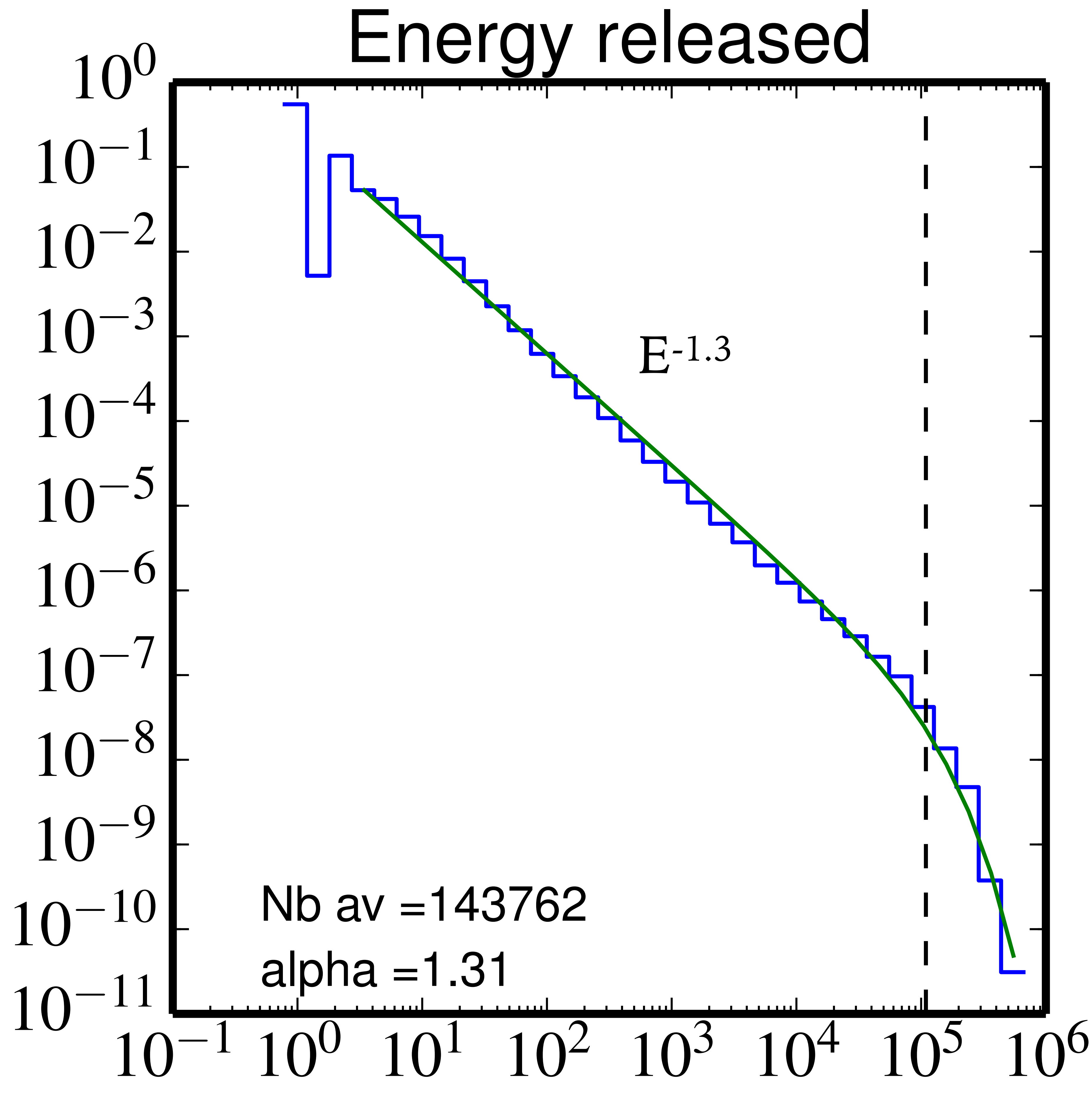


‘Self-Organized Criticality’

Driver



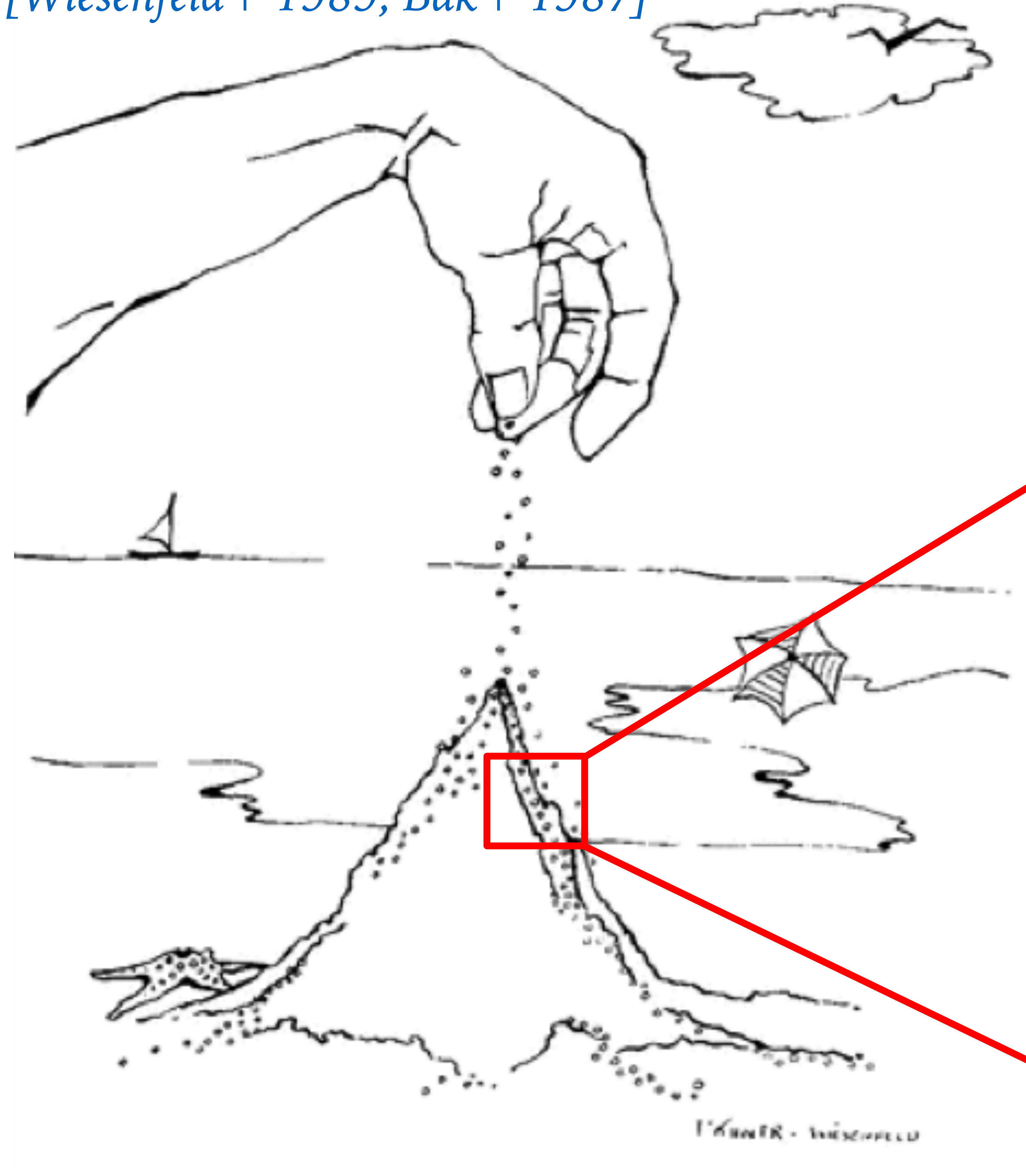
[Lu & Hamilton 1993]



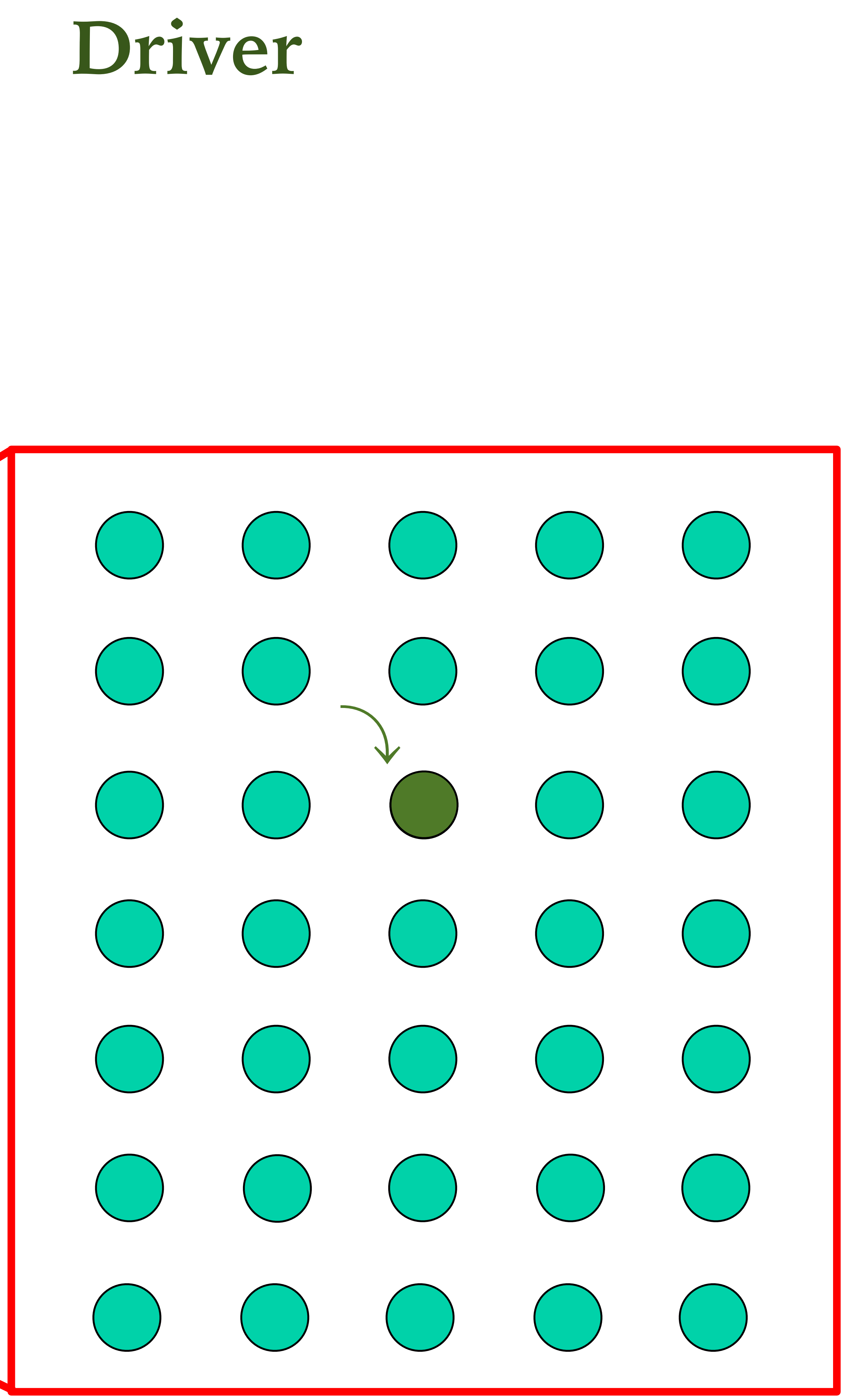
[Strugarek+ 2014a,b]

Basics of sandpile models

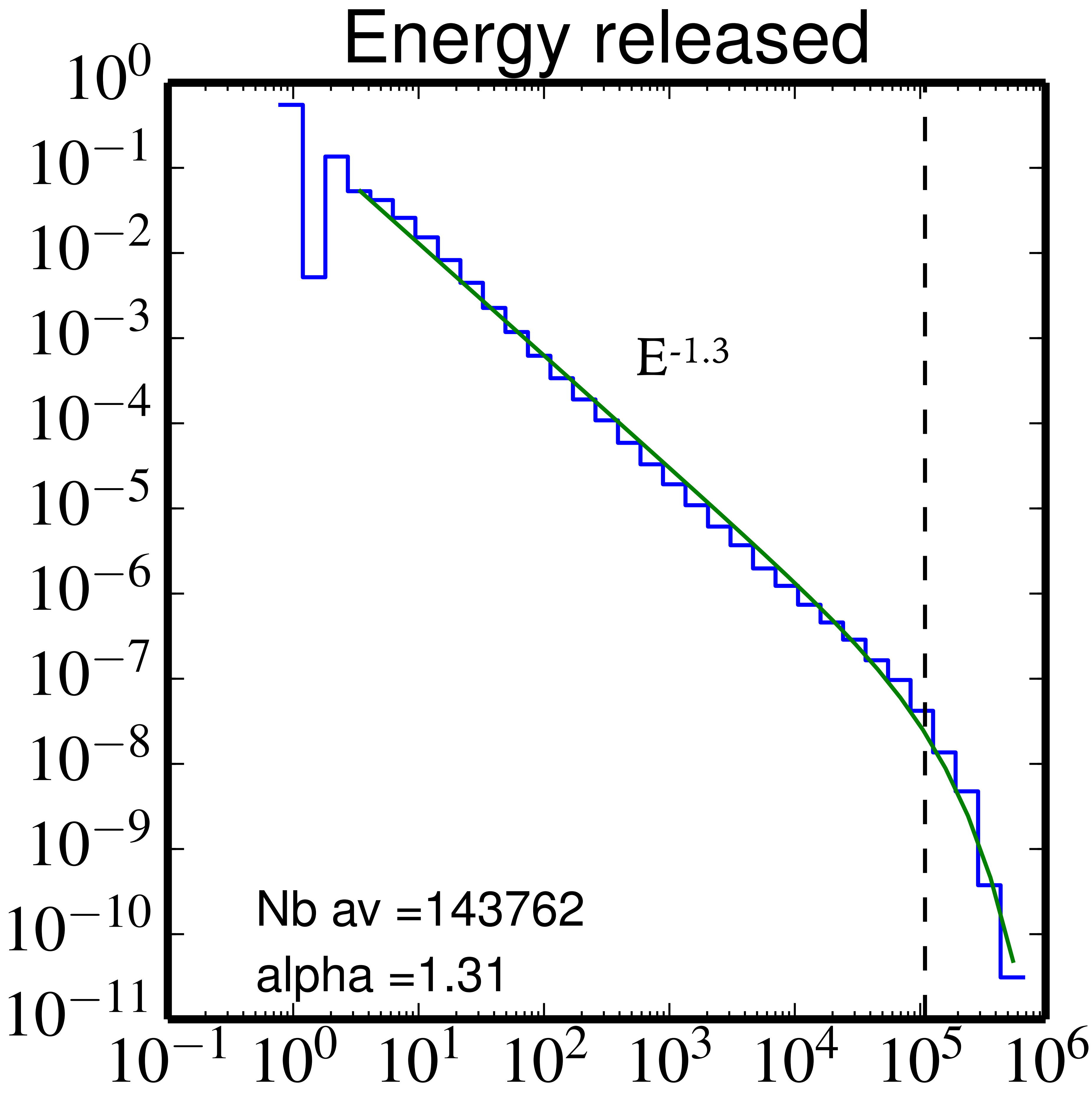
[Wiesenfeld+ 1989; Bak+ 1987]



‘Self-Organized Criticality’



[Lu & Hamilton 1993]



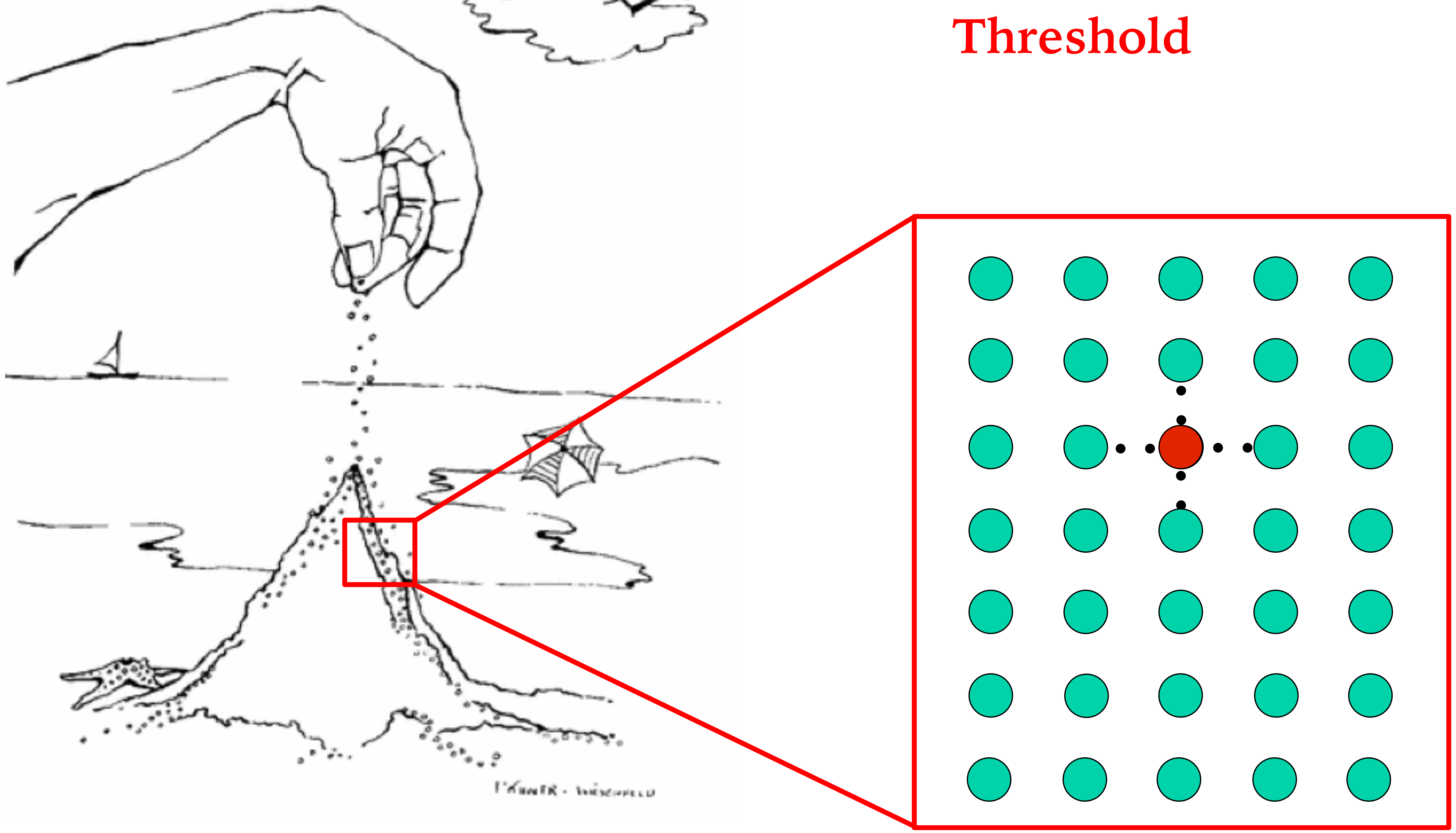
[Strugarek+ 2014a,b]

Basics of sandpile models

[Wiesenfeld+ 1989; Bak+ 1987]

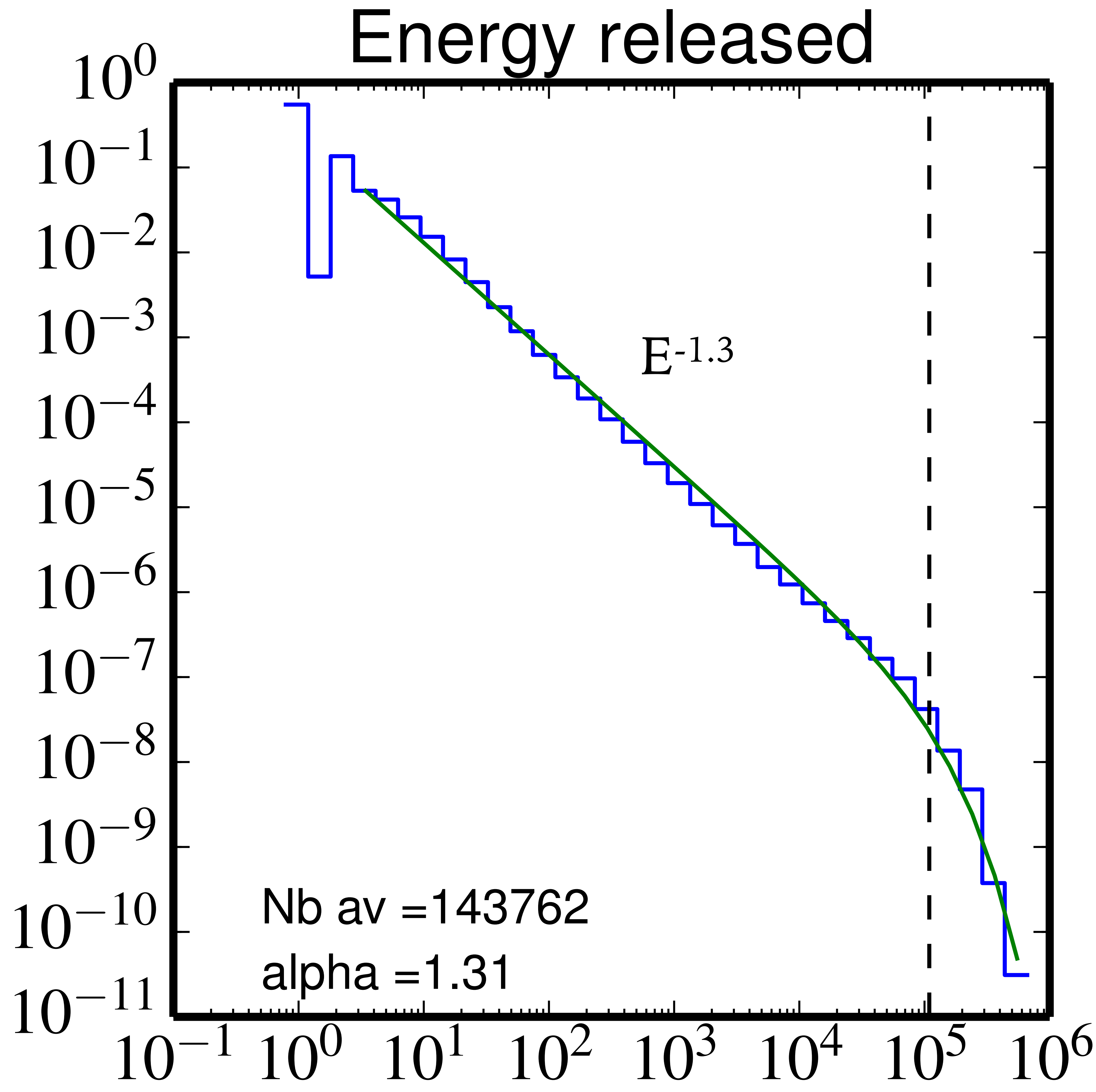
Driver

Threshold



‘Self-Organized Criticality’

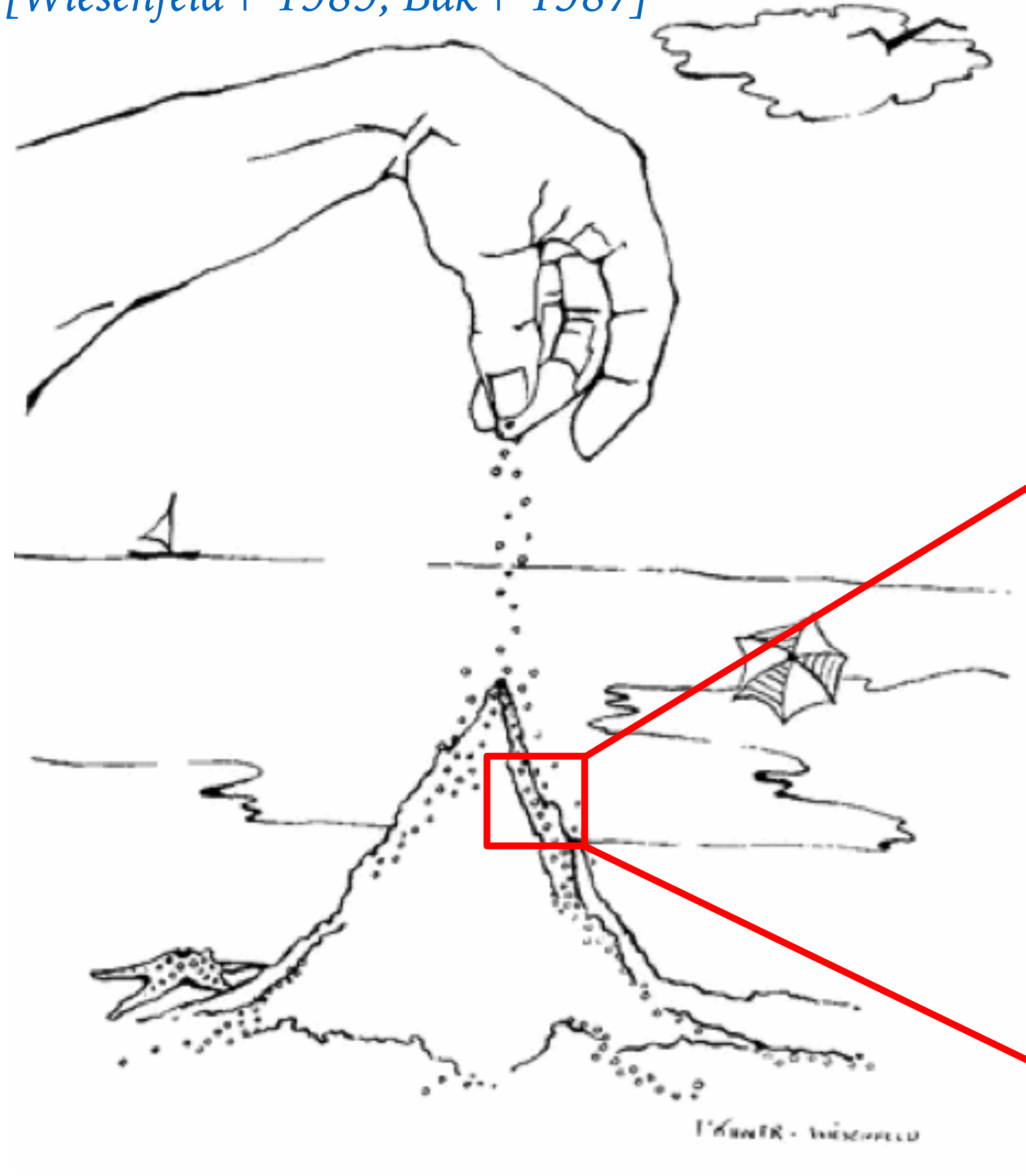
[Lu & Hamilton 1993]



[Strugarek+ 2014a,b]

Basics of sandpile models

[Wiesenfeld+ 1989; Bak+ 1987]

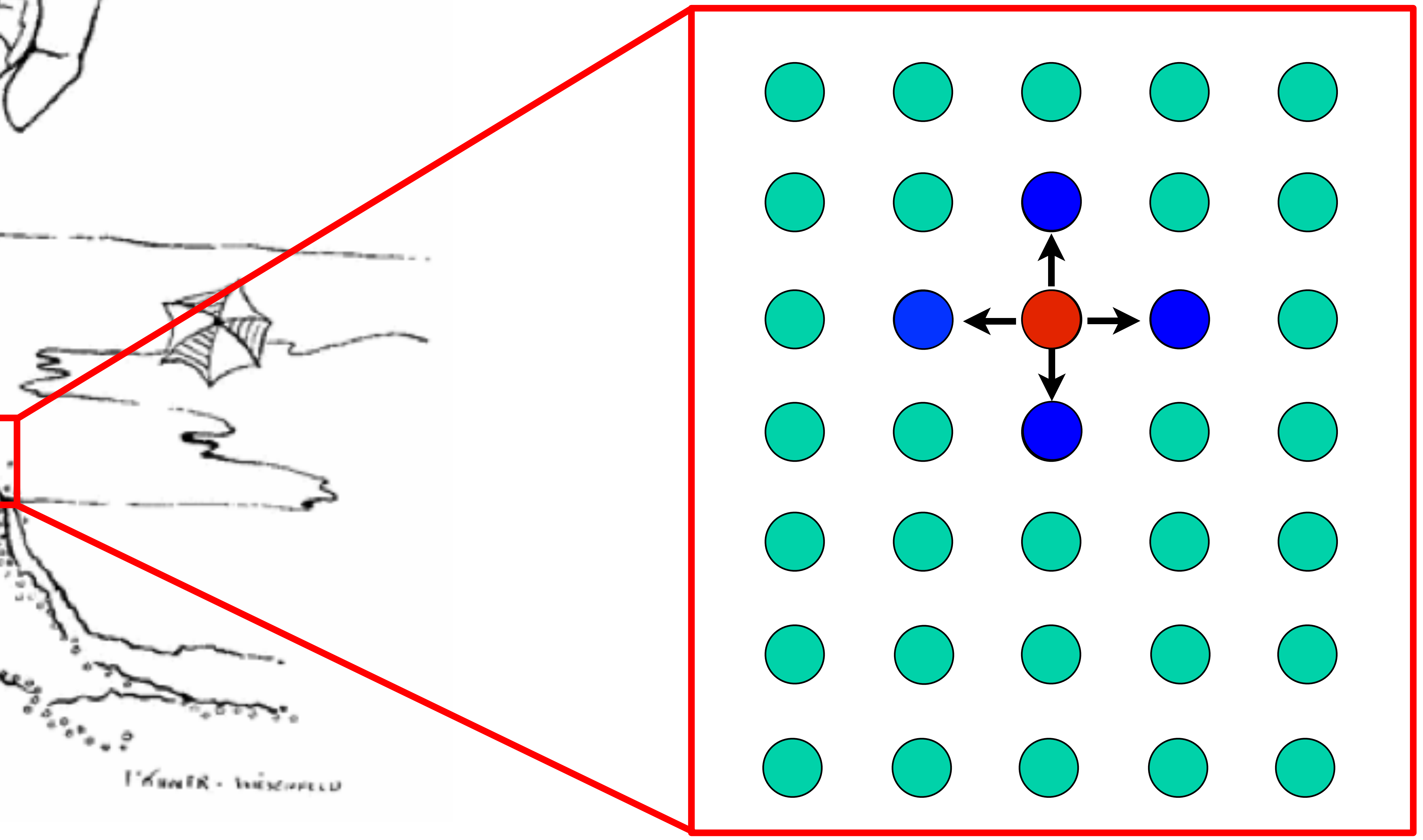


'Self-Organized Criticality'

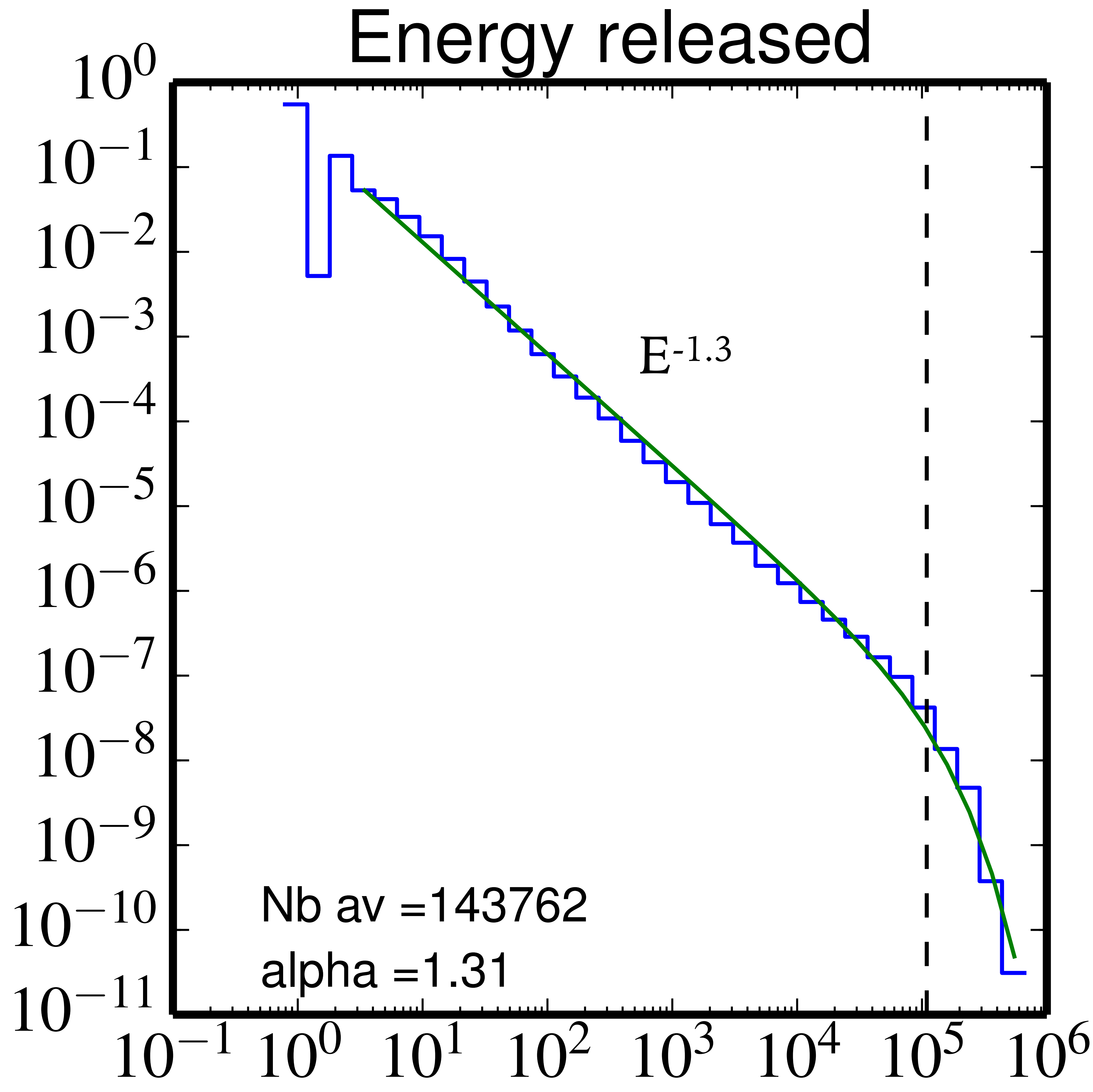
Driver

Threshold

Redistribution rule

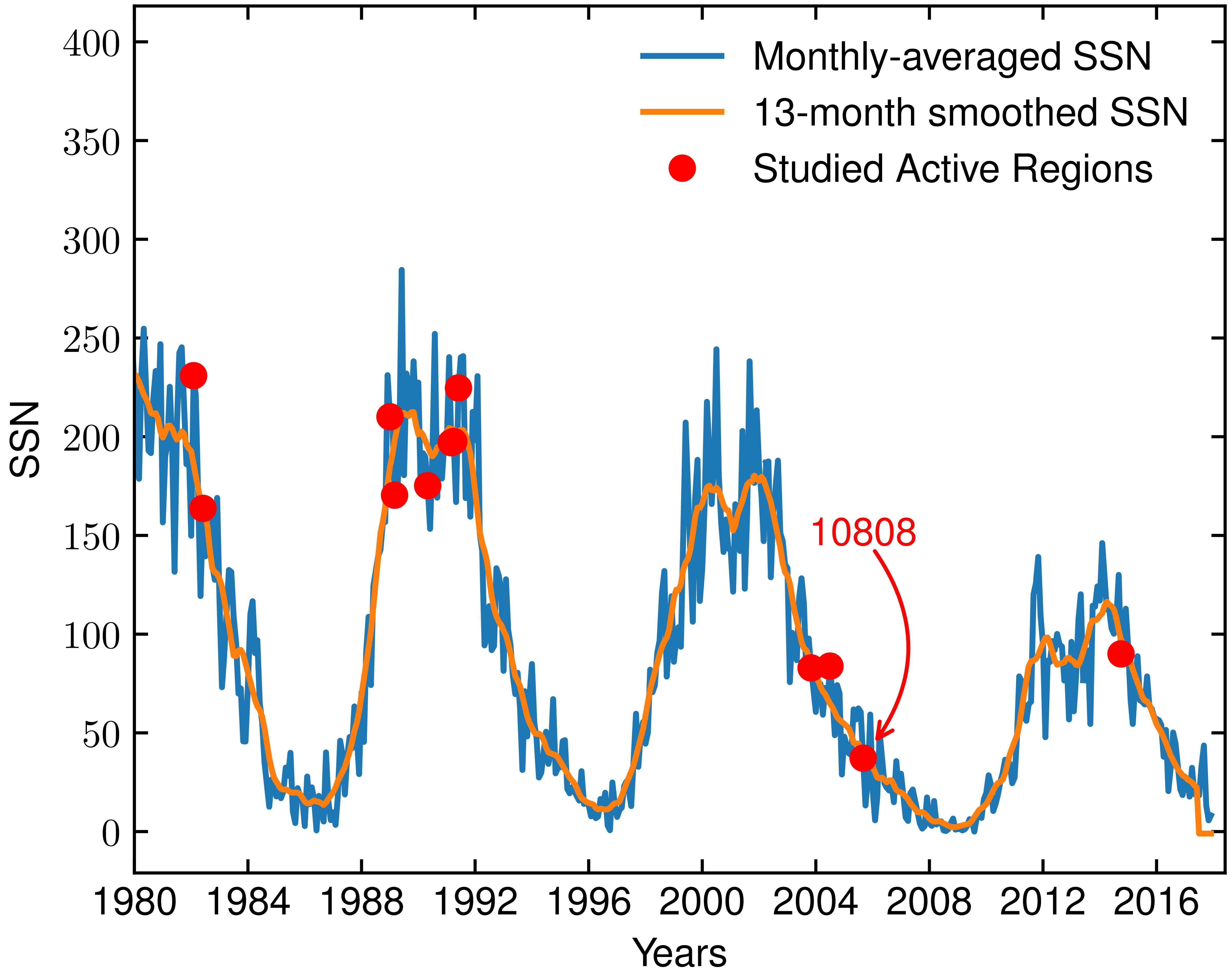


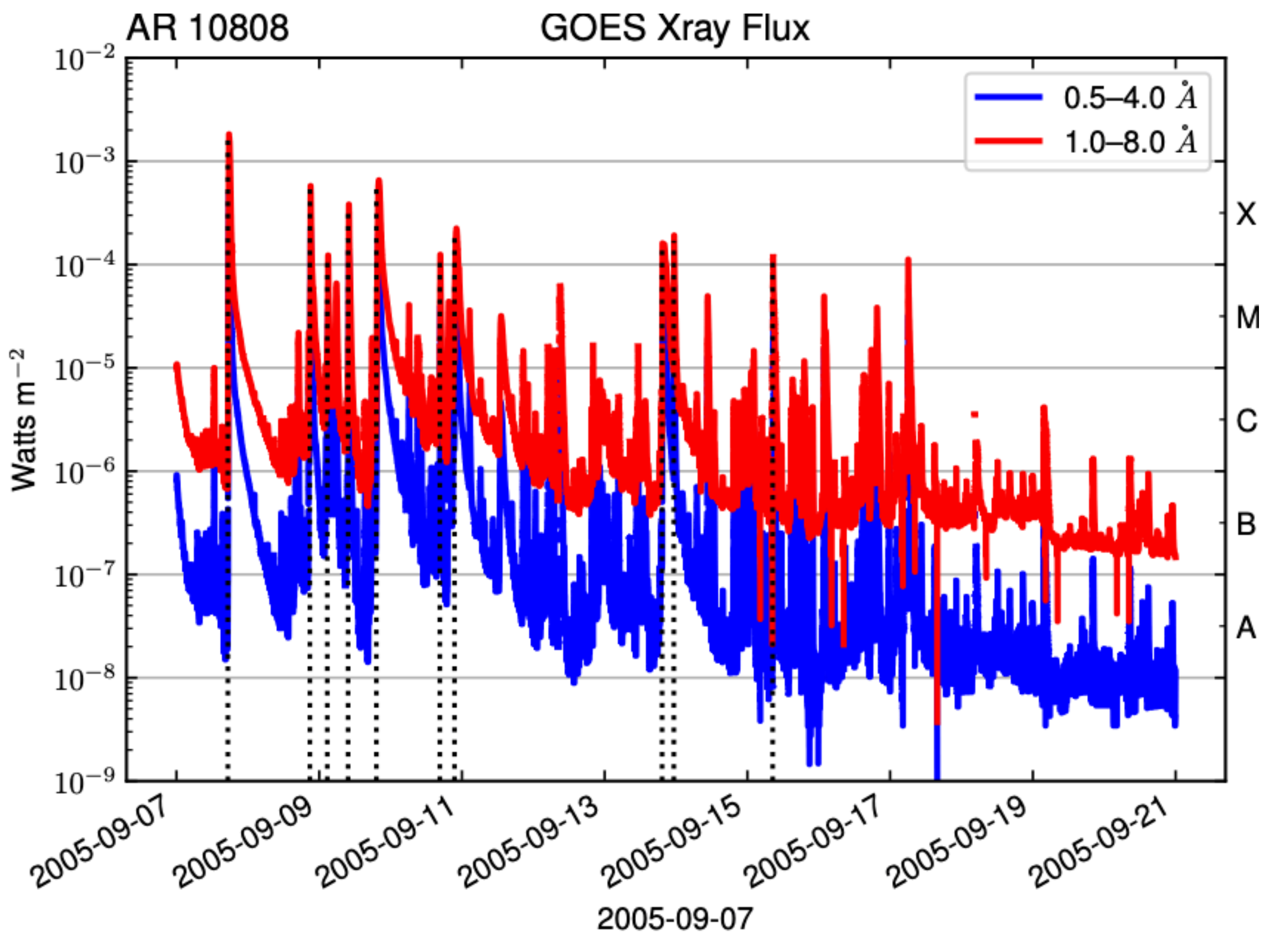
[Lu & Hamilton 1993]



[Strugarek+ 2014a,b]

Avancées de la tâche II (Thibeault et al. 2021): application à des données GOES

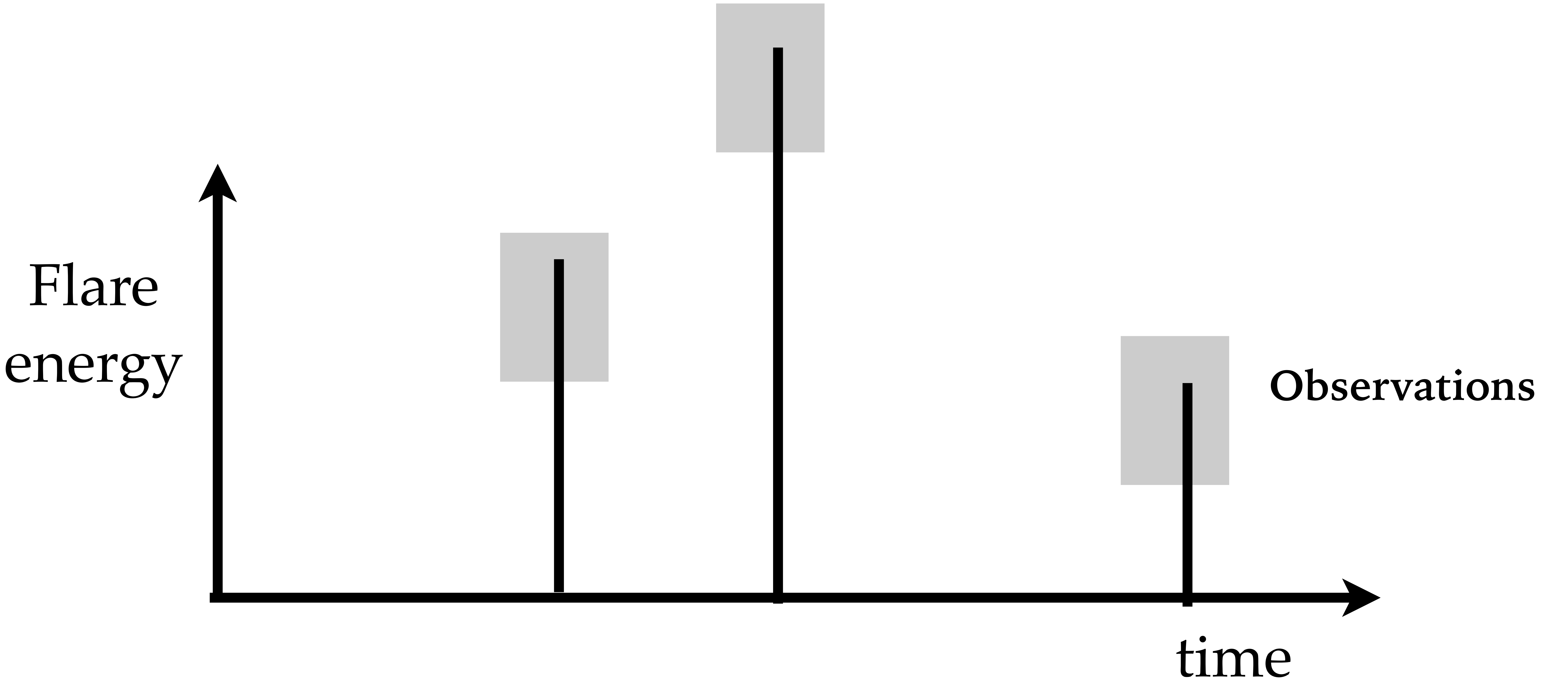




A) Compare observational data with the sandpile model (III)

Least squares not efficient because of the discrete character of the events

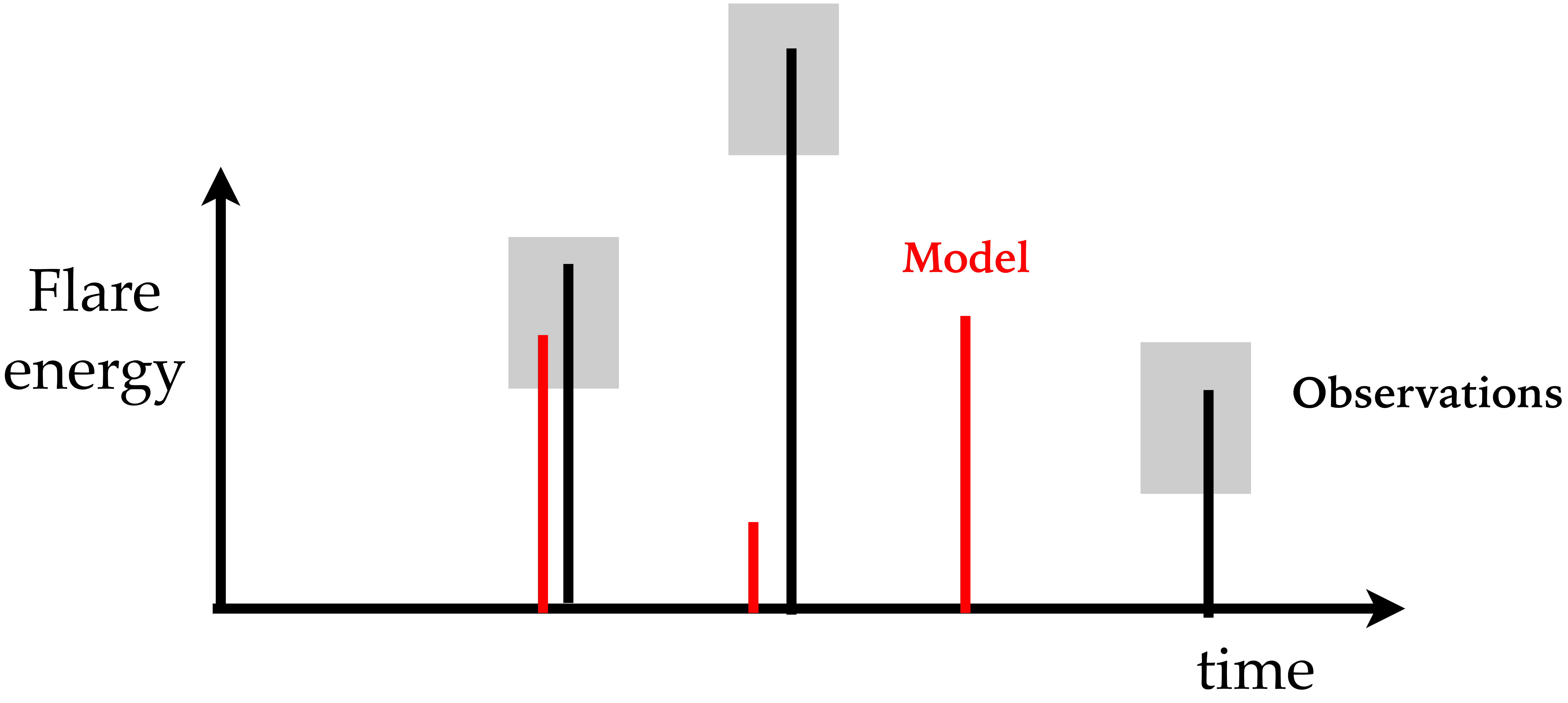
Defining a versatile **cost function**:



A) Compare observational data with the sandpile model (III)

Least squares not efficient because of the discrete character of the events

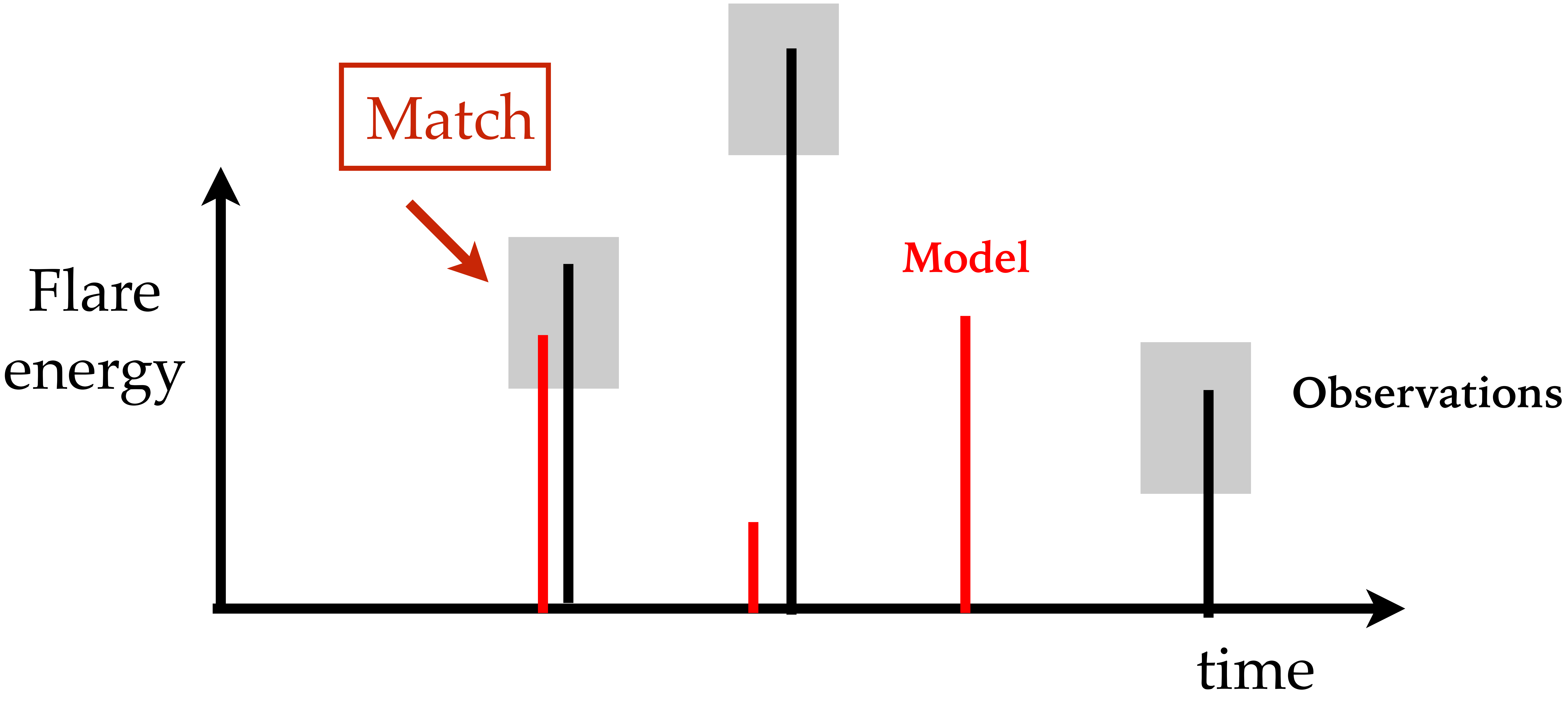
Defining a versatile **cost function**:



A) Compare observational data with the sandpile model (III)

Least squares not efficient because of the discrete character of the events

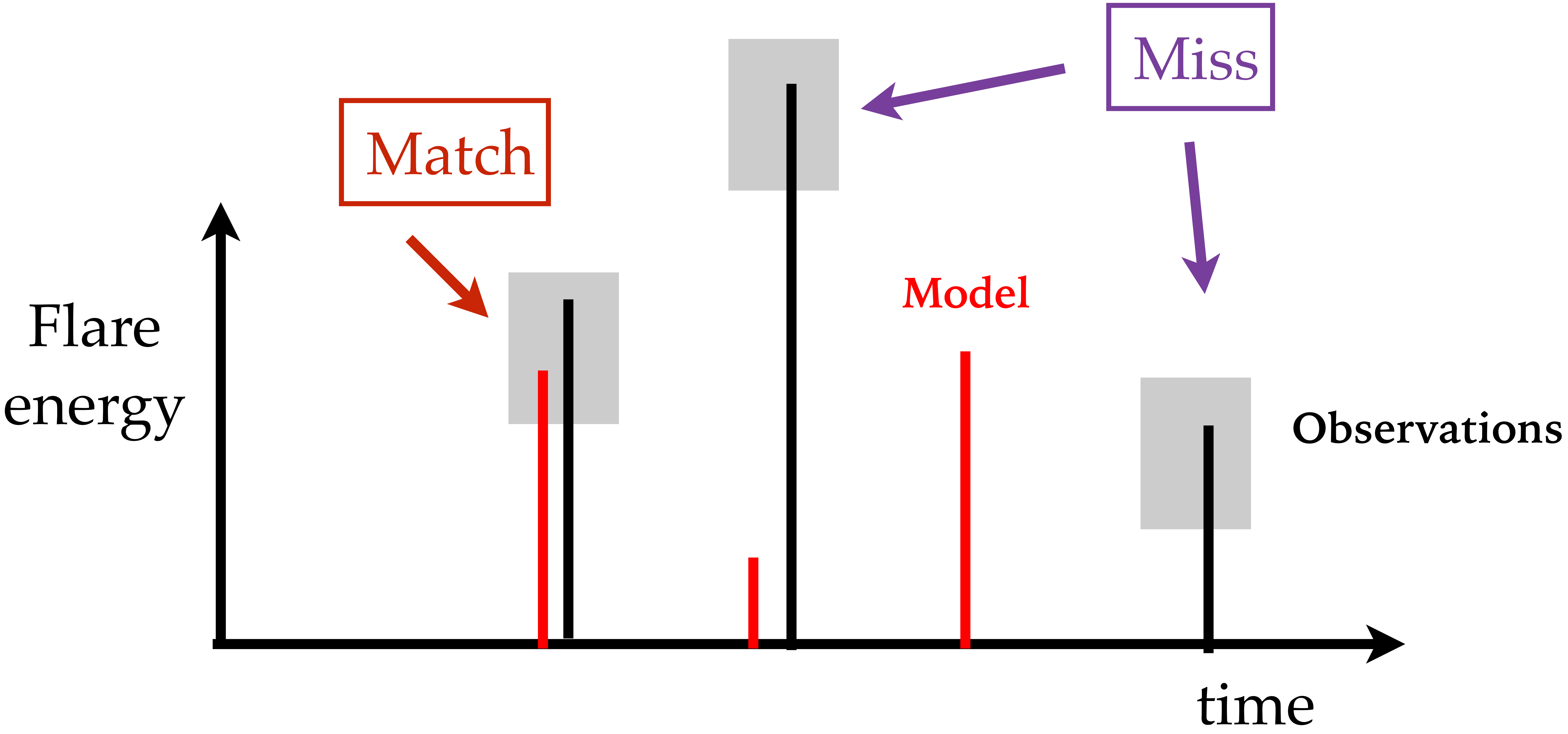
Defining a versatile **cost function**:



A) Compare observational data with the sandpile model (III)

Least squares not efficient because of the discrete character of the events

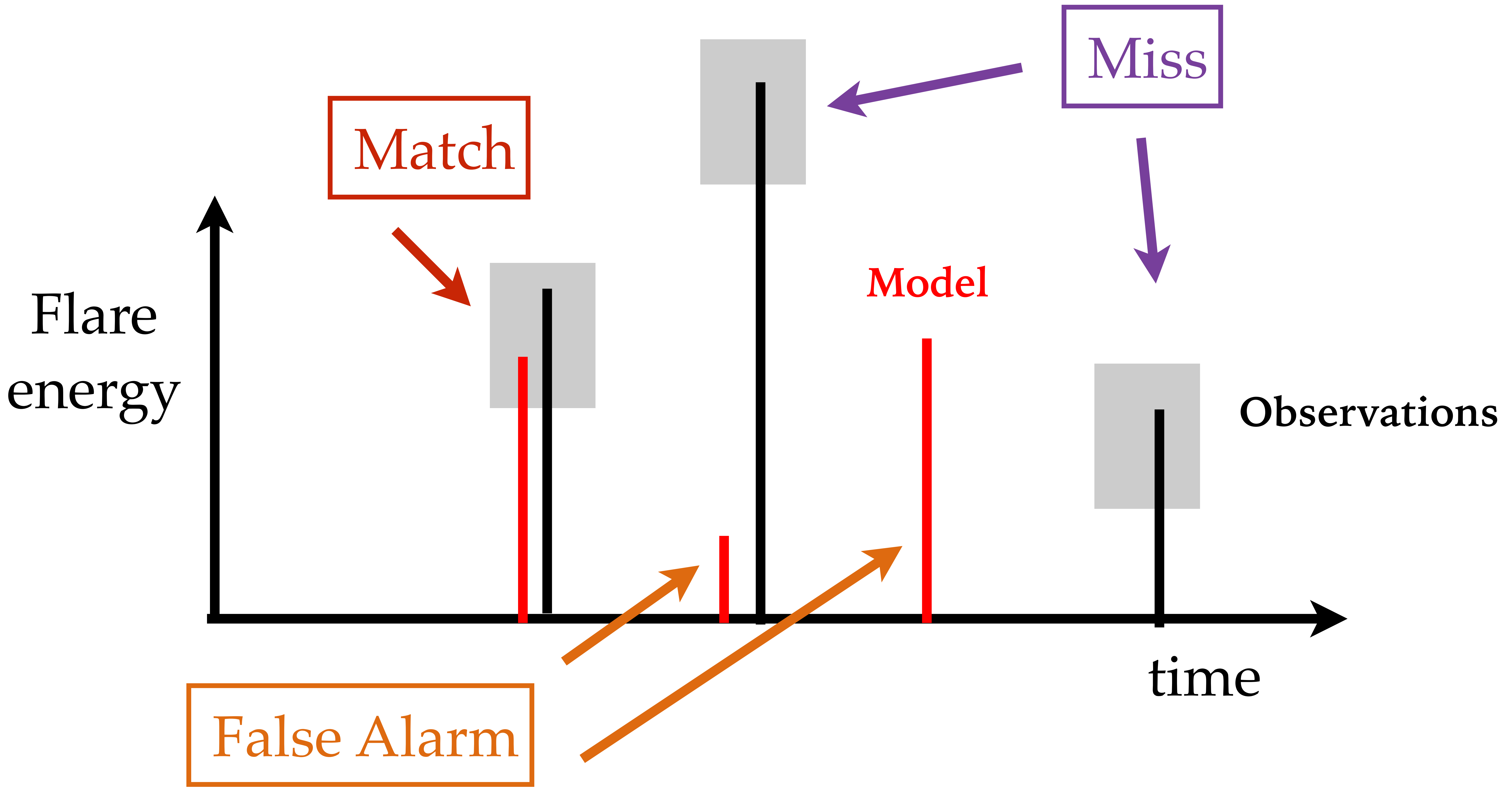
Defining a versatile **cost function**:



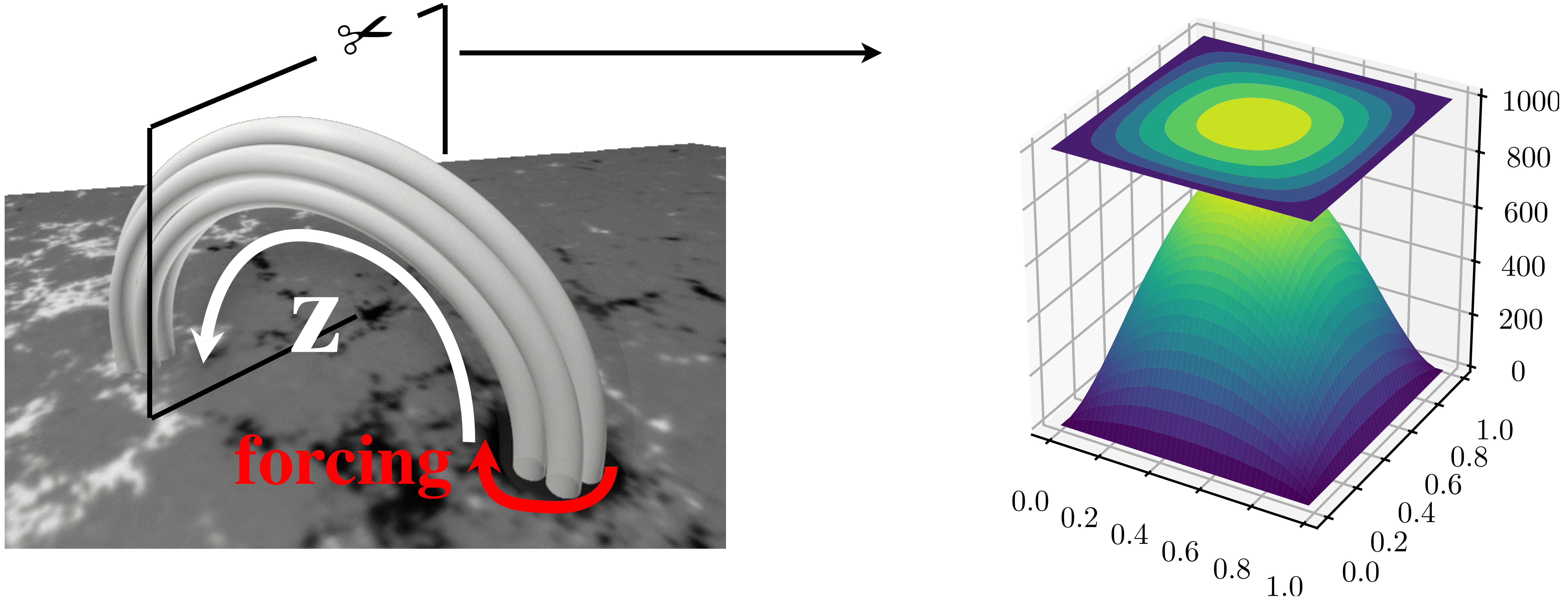
A) Compare observational data with the sandpile model (III)

Least squares not efficient because of the discrete character of the events

Defining a versatile **cost function**:



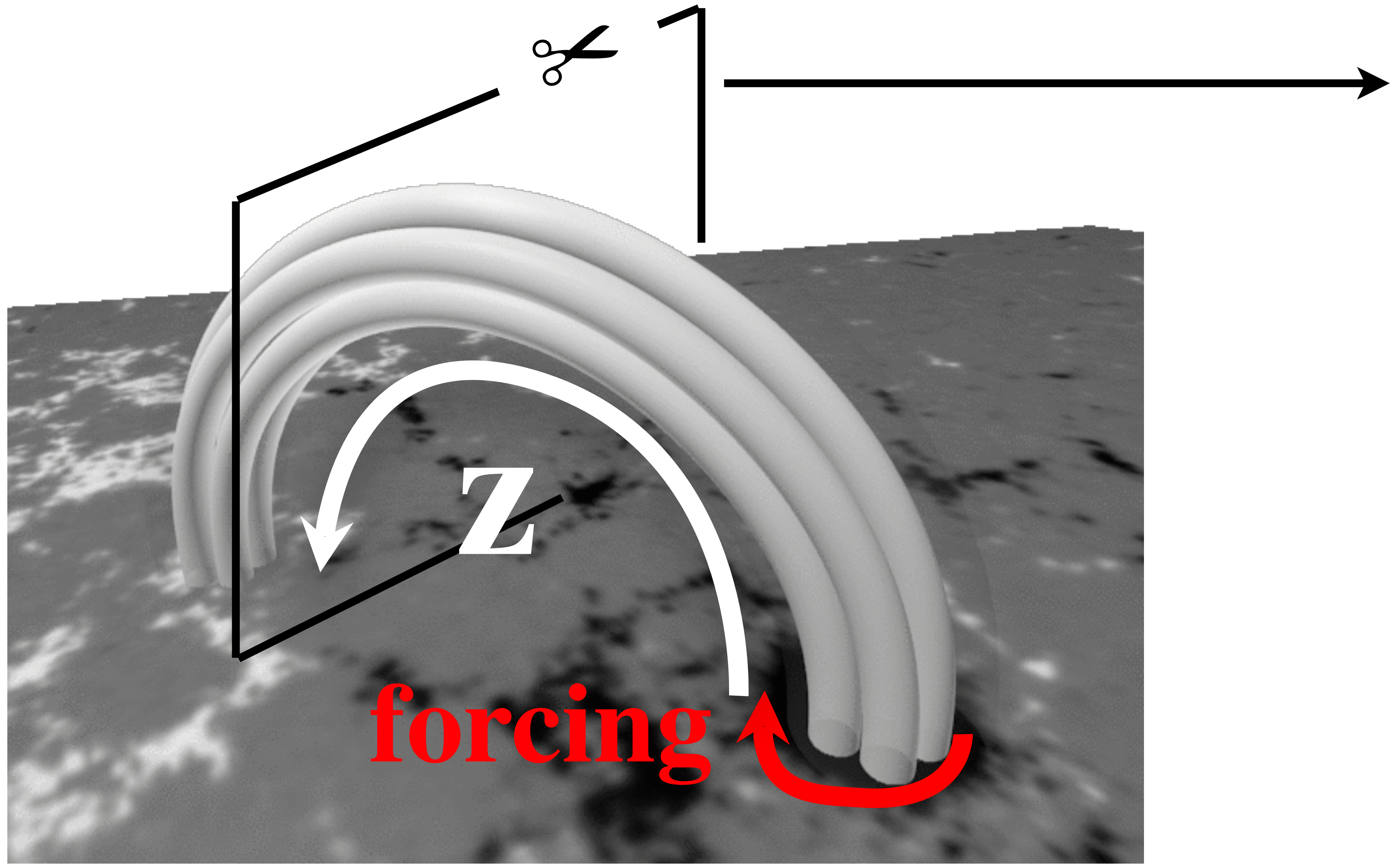
Physical interpretation(s) of sandpiles

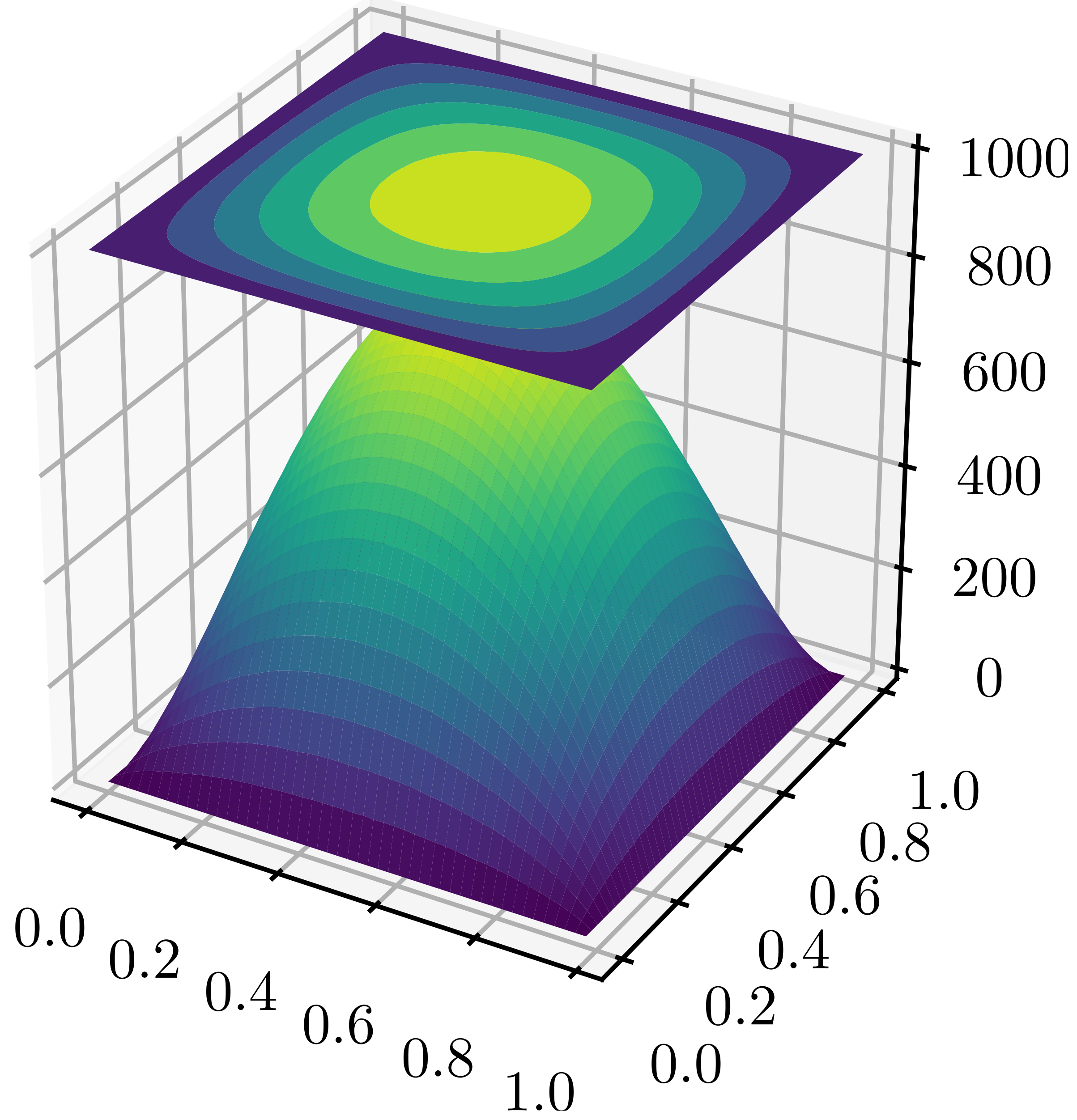


Coronal Loop	Sandpile
Magnetic potential A_z	Height
Turbulent twisting of loop	Homogenous forcing
Currents	Curvature
Magnetic reconnection	Stochastic redistribution

[Strugarek+ 2014]

Physical interpretation(s) of sandpiles

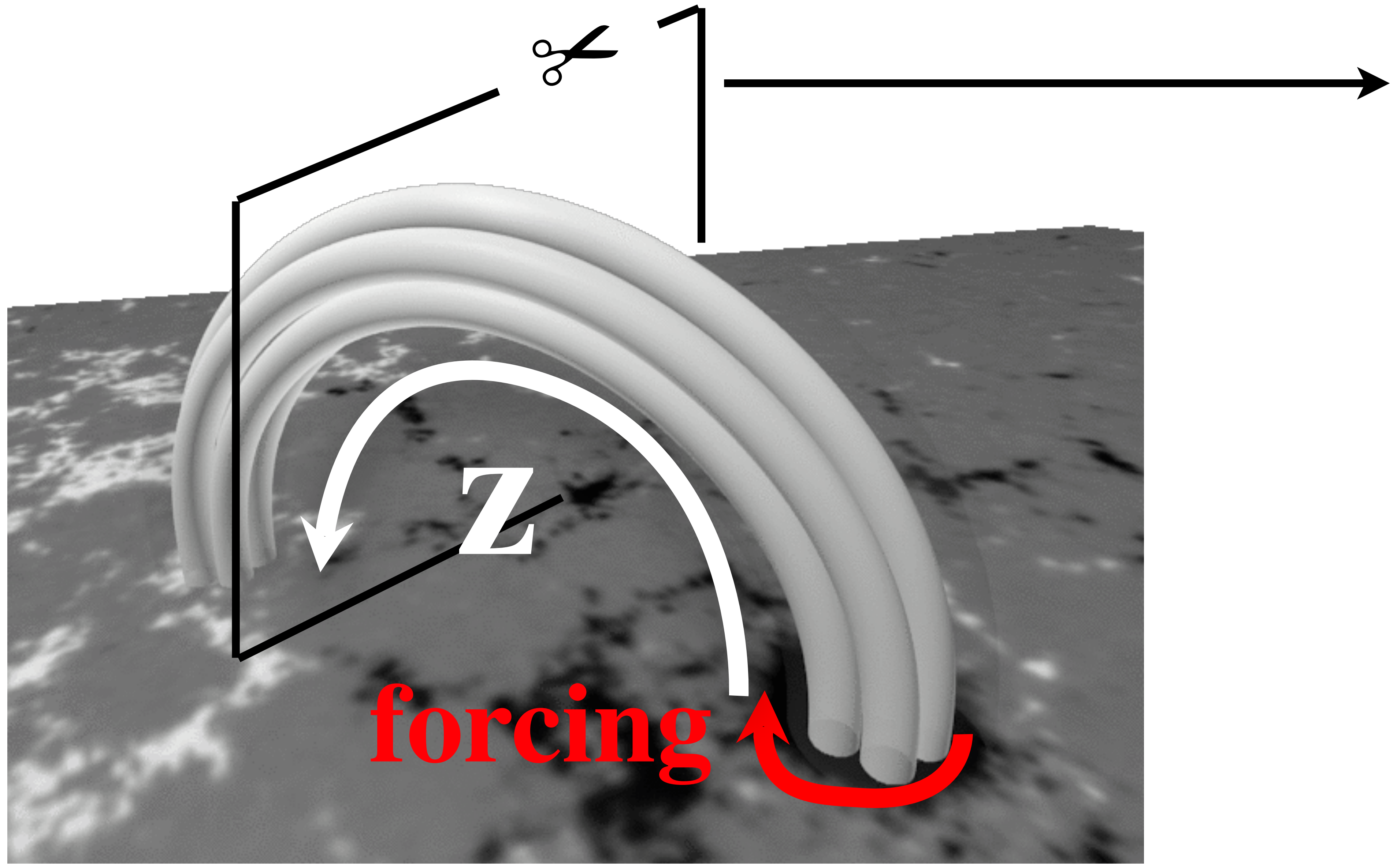


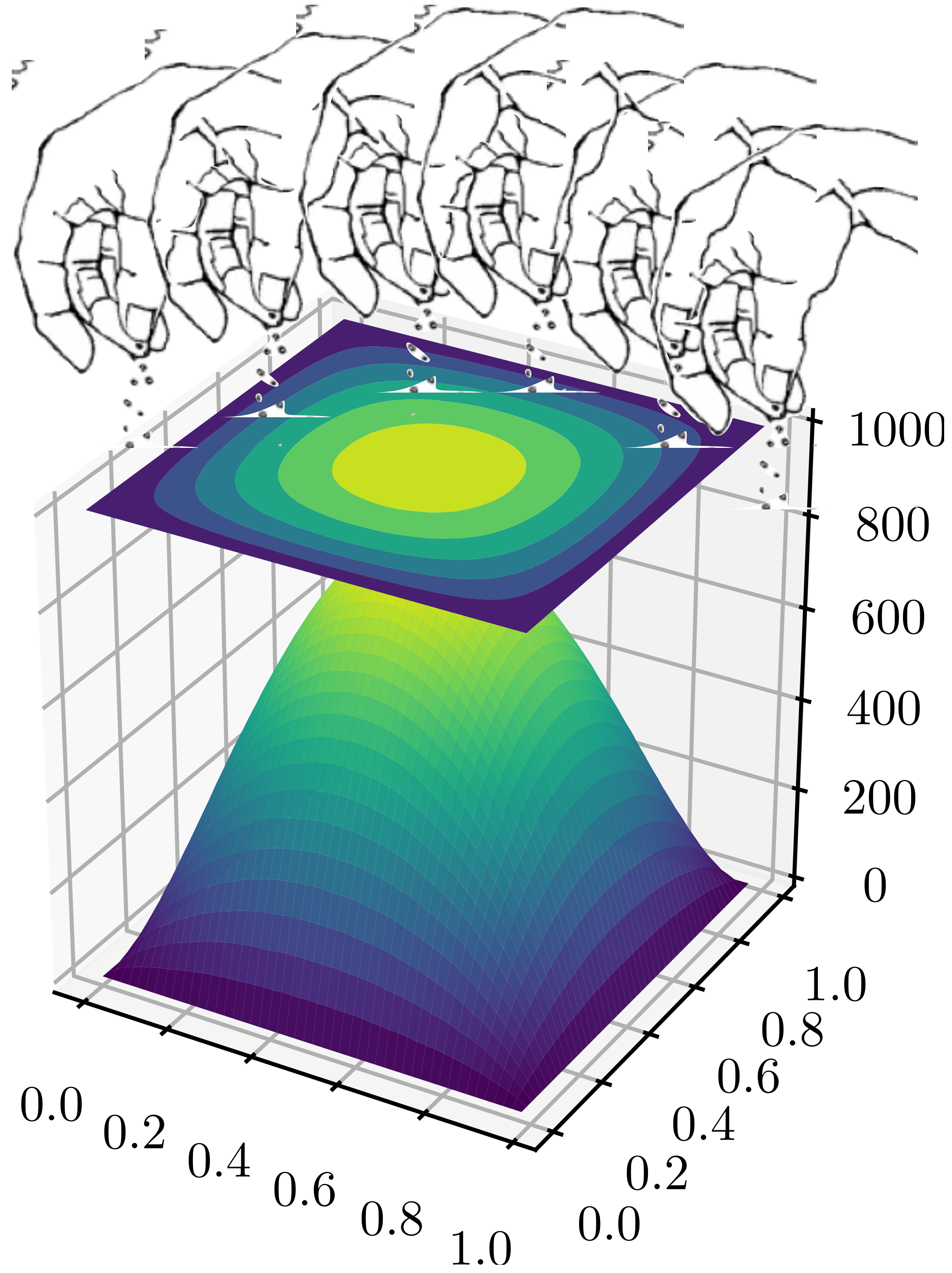


Coronal Loop	Sandpile
Magnetic potential A_z	Height
Turbulent twisting of loop	Homogenous forcing
Currents	Curvature
Magnetic reconnection	Stochastic redistribution

[Strugarek+ 2014]

Physical interpretation(s) of sandpiles





Coronal Loop	Sandpile
Magnetic potential A_z	Height
Turbulent twisting of loop	Homogenous forcing
Currents	Curvature
Magnetic reconnection	Stochastic redistribution

[Strugarek+ 2014]

Data assimilation in sandpile models: FlarePredict

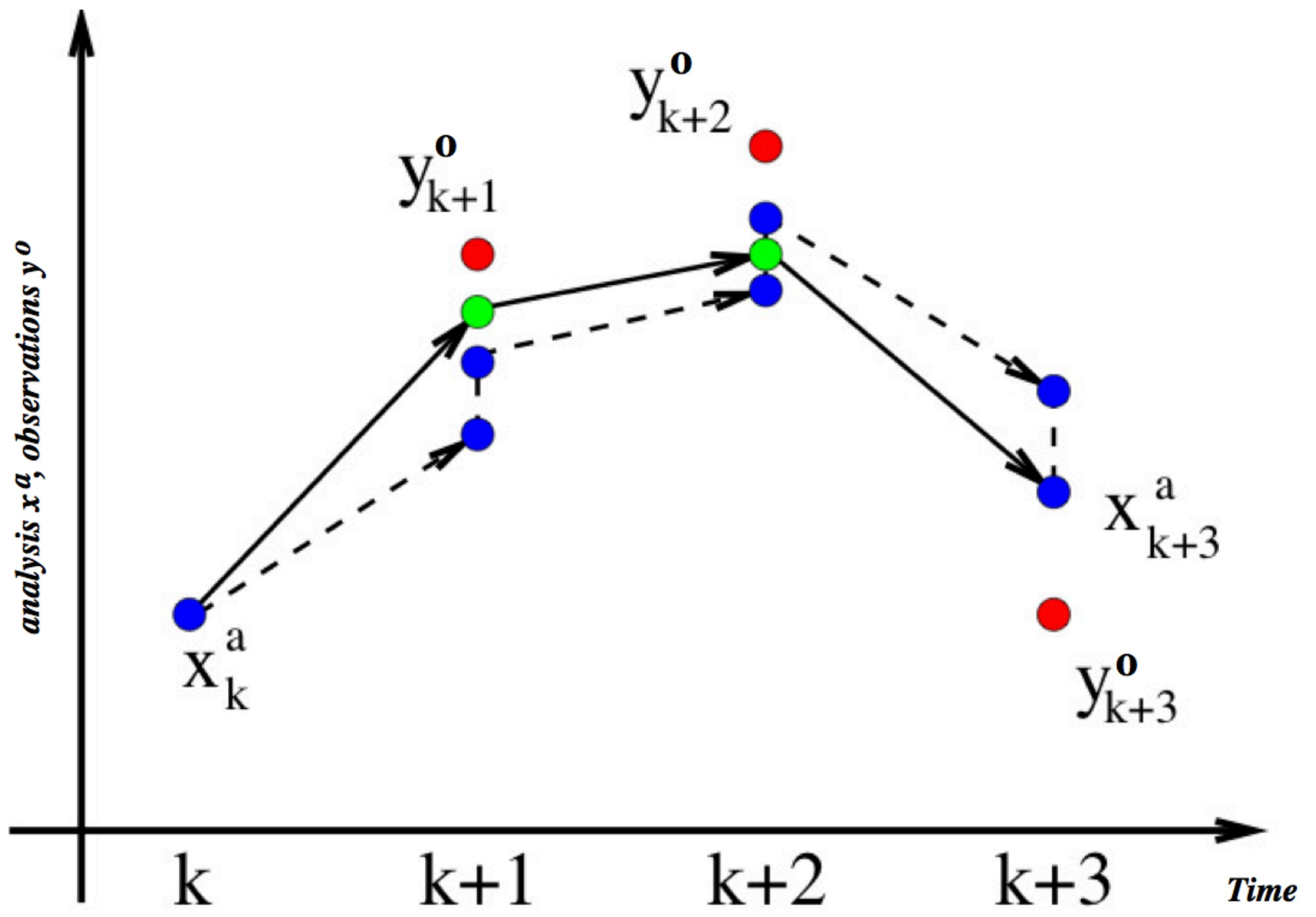
Initial condition (+model parameters)

A) Compare model *vs* observations

B) Observations are **not** reproduced

C) Observations are reproduced

Variational data assimilation

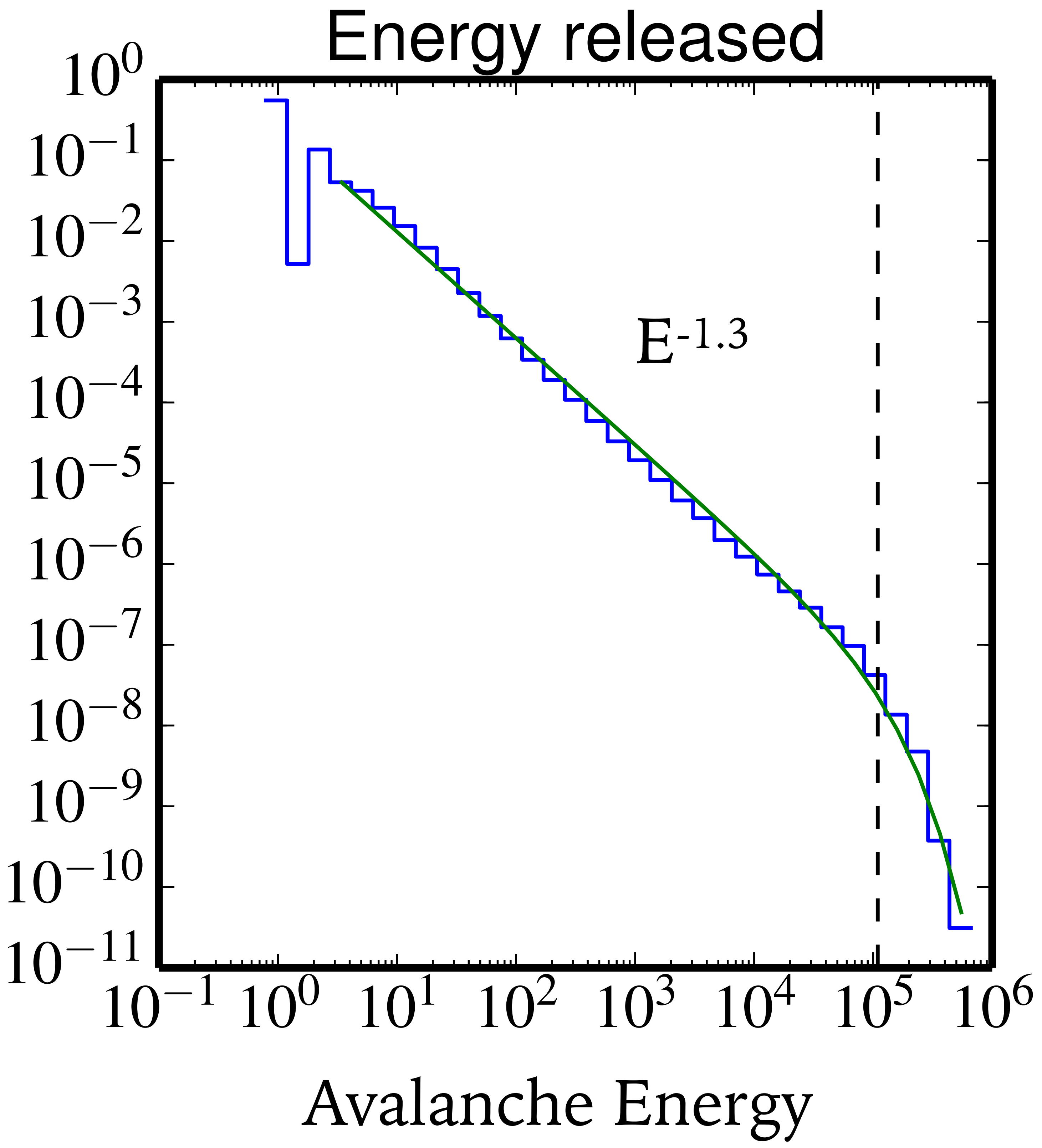


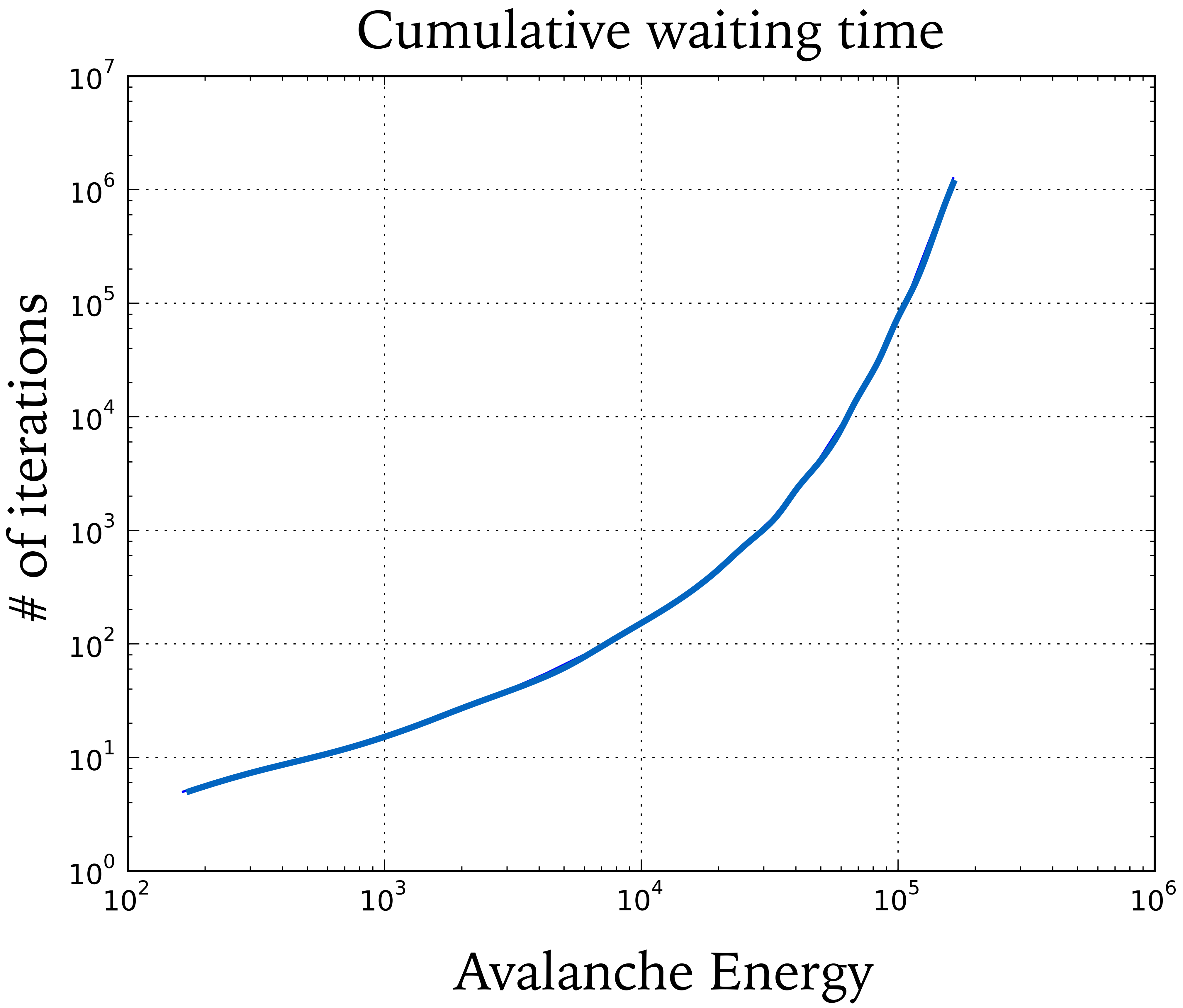
Need **efficient minimization** technique

Determine eruptive probability (prediction)

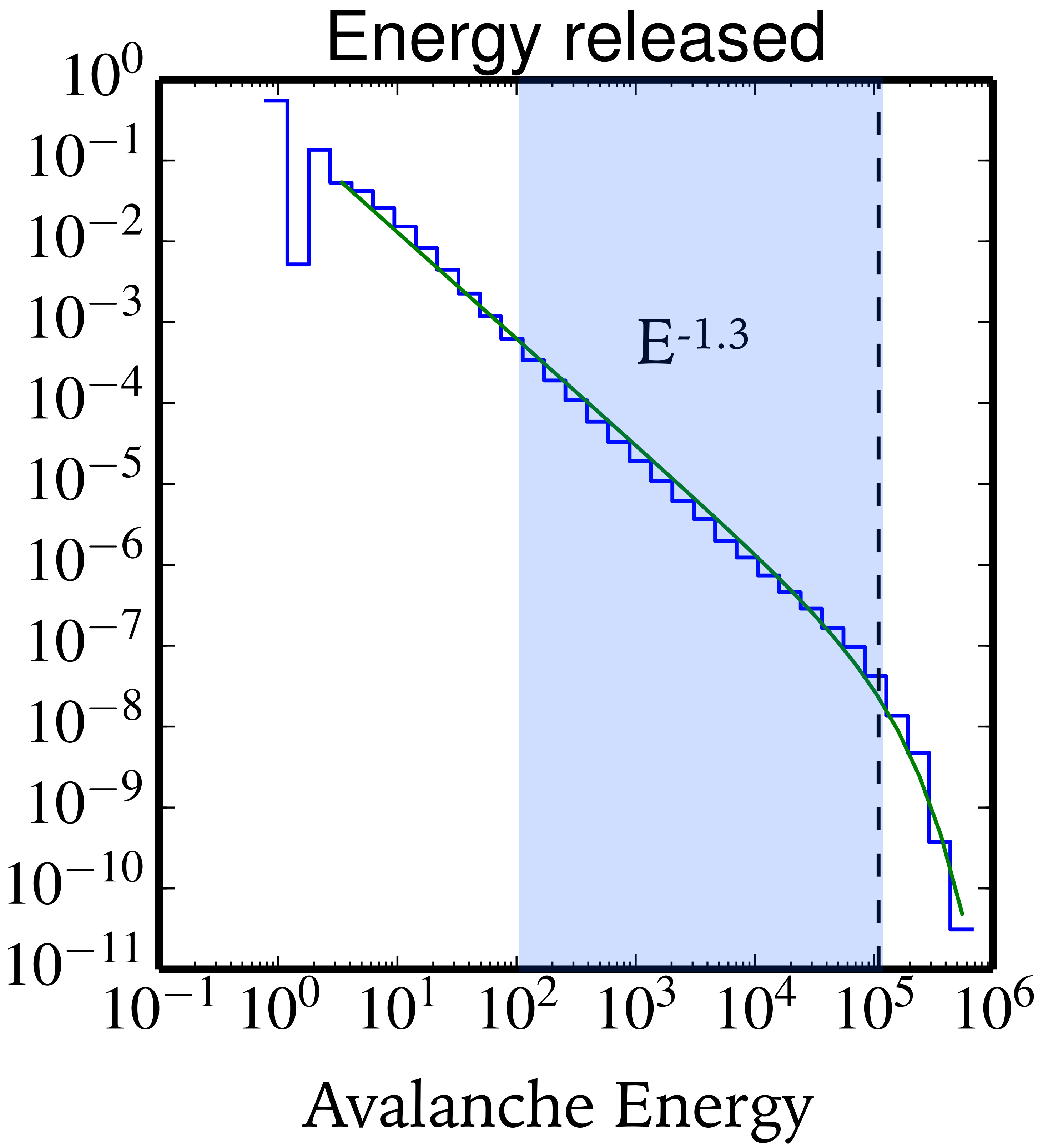
'Hindcasting' on past records first

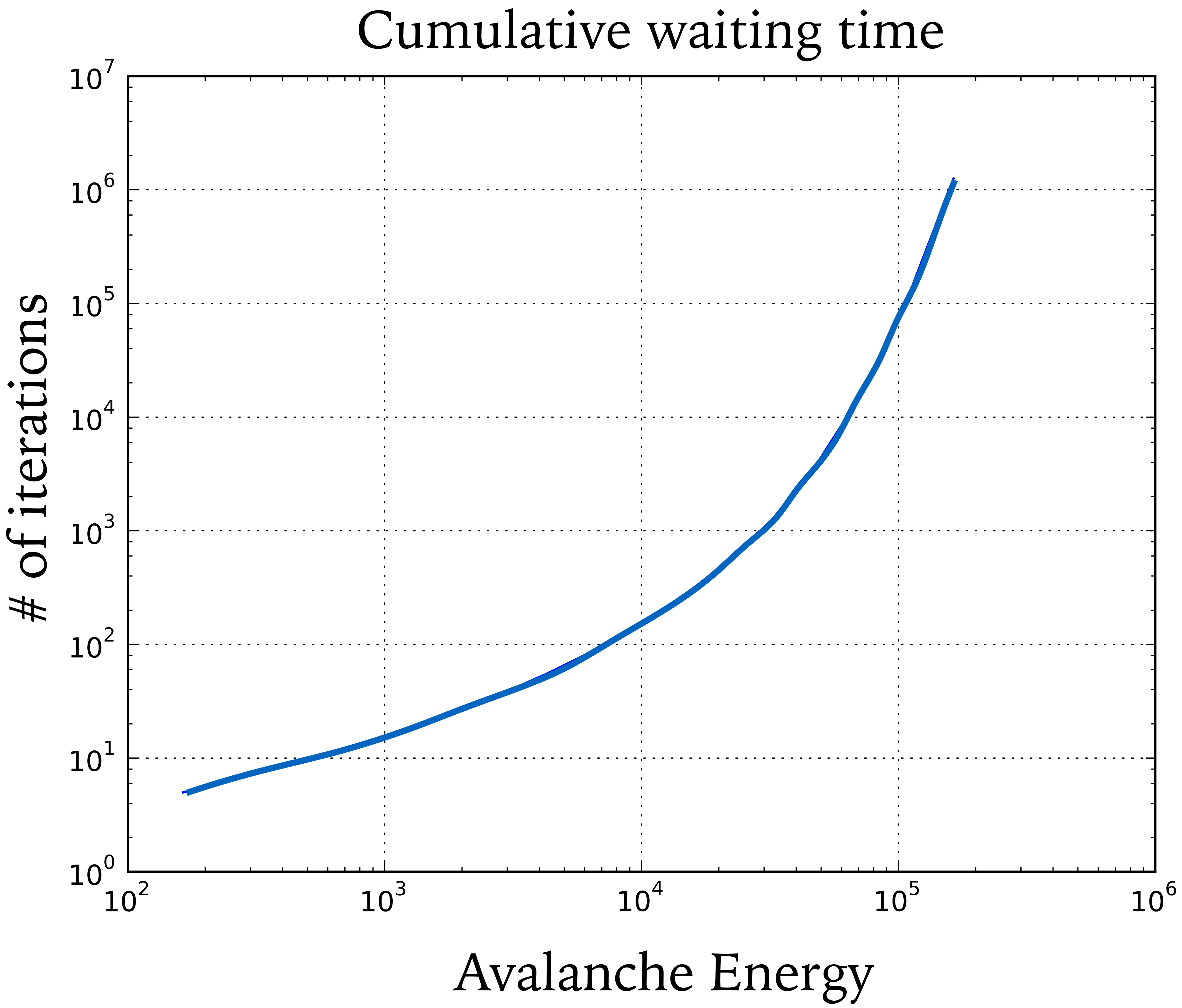
A) Compare observational data with the sandpile model (I)





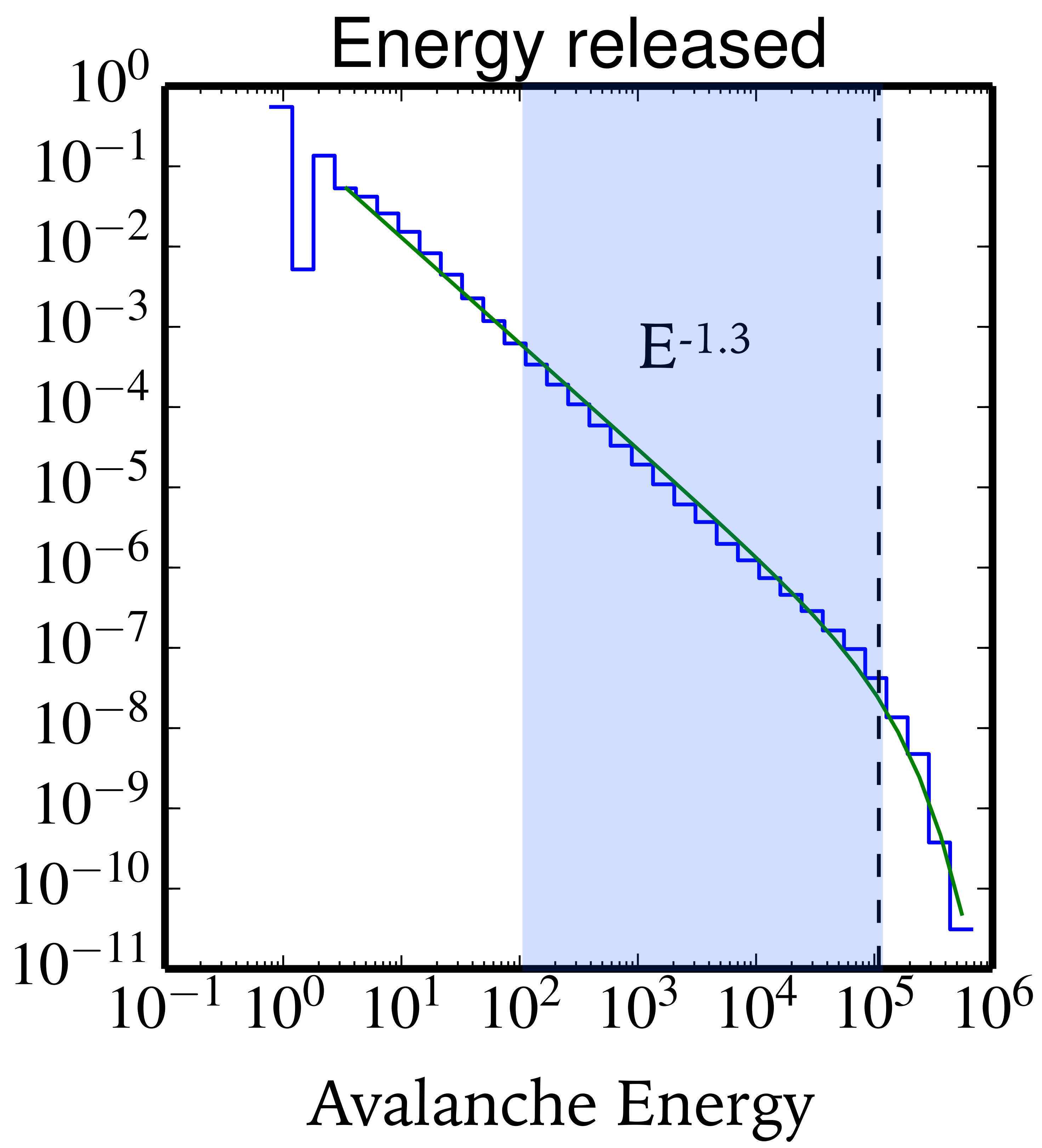
A) Compare observational data with the sandpile model (I)

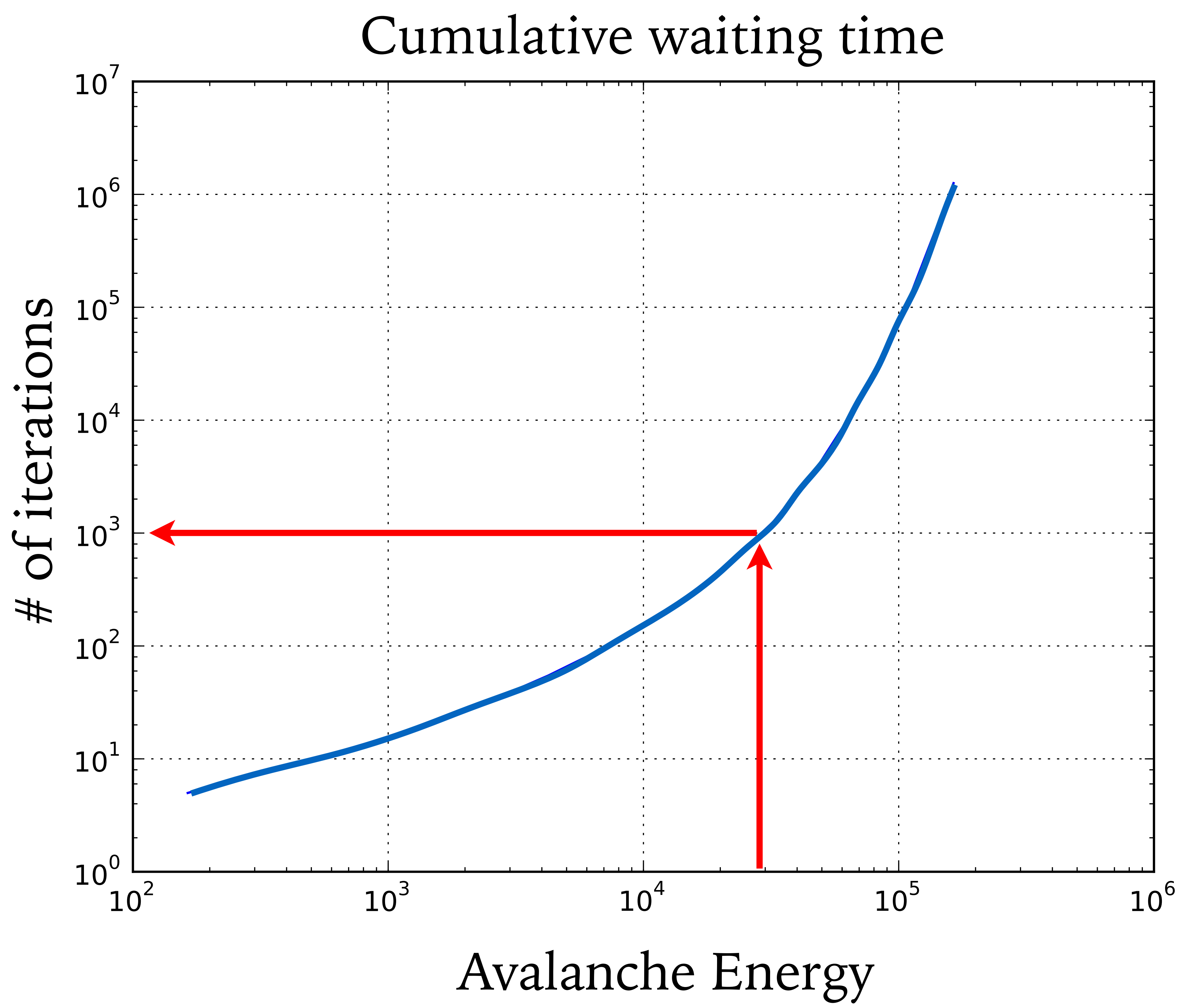




1) Match range of model energies to Flare energy from C to X

A) Compare observational data with the sandpile model (I)

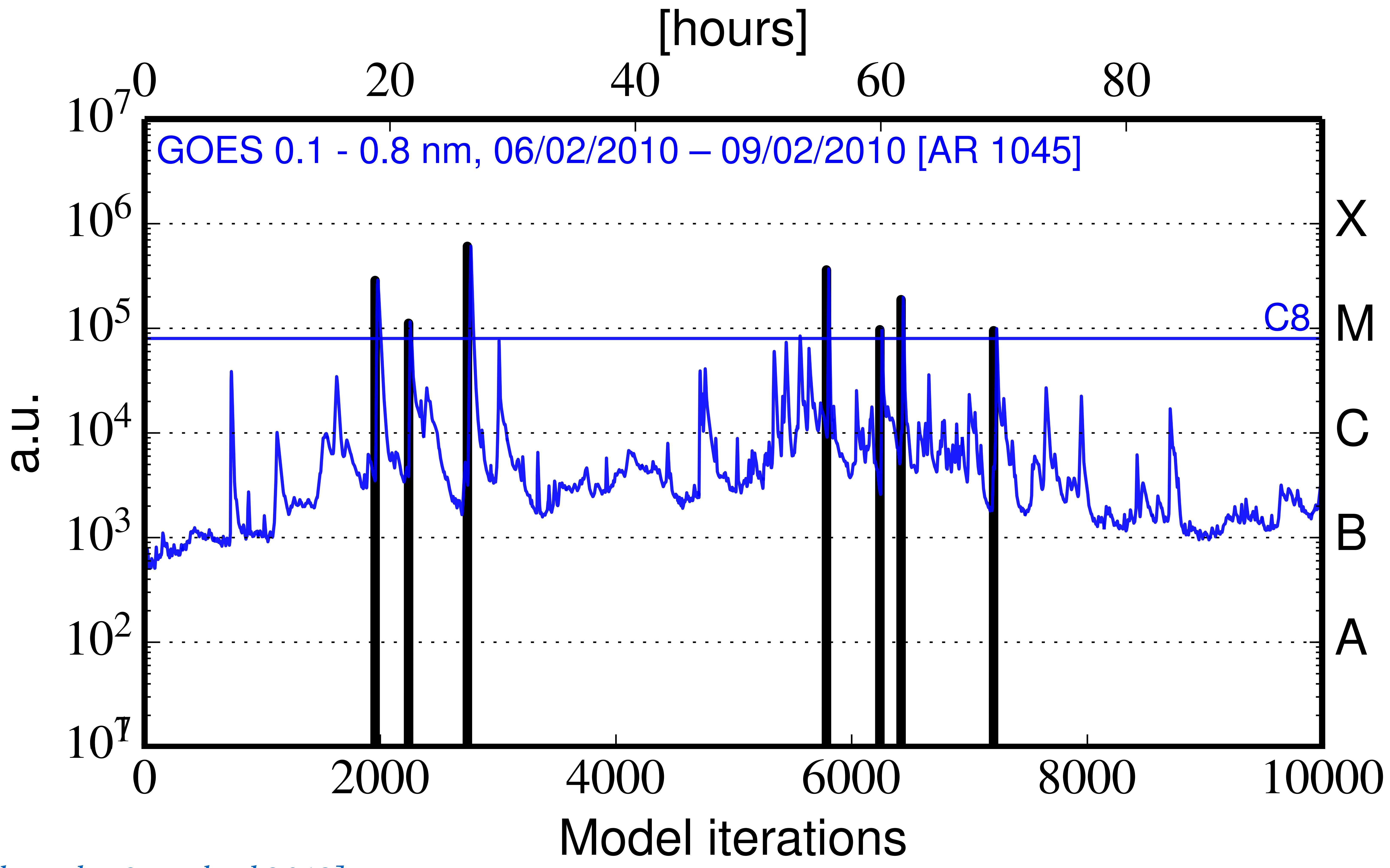




1) Match range of model energies to Flare energy from C to X

2) Deduce time normalization from the cumulative waiting time distributions

A) Compare observational data with the sandpile model (II)

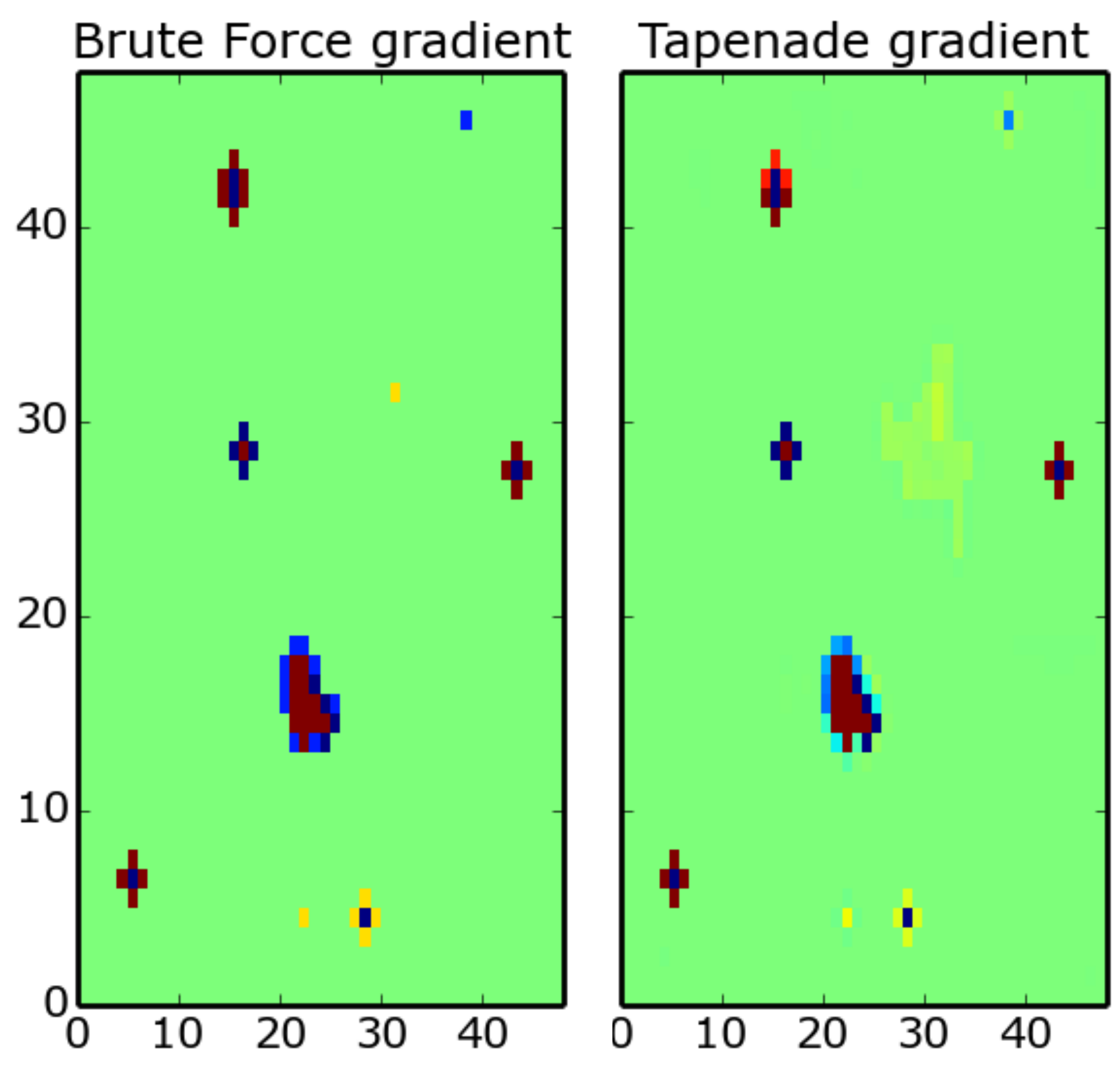


[Aschwanden & Freeland 2012]

B) Find new initial conditions reproducing the data (I)

B) Find new initial conditions reproducing the data (I)

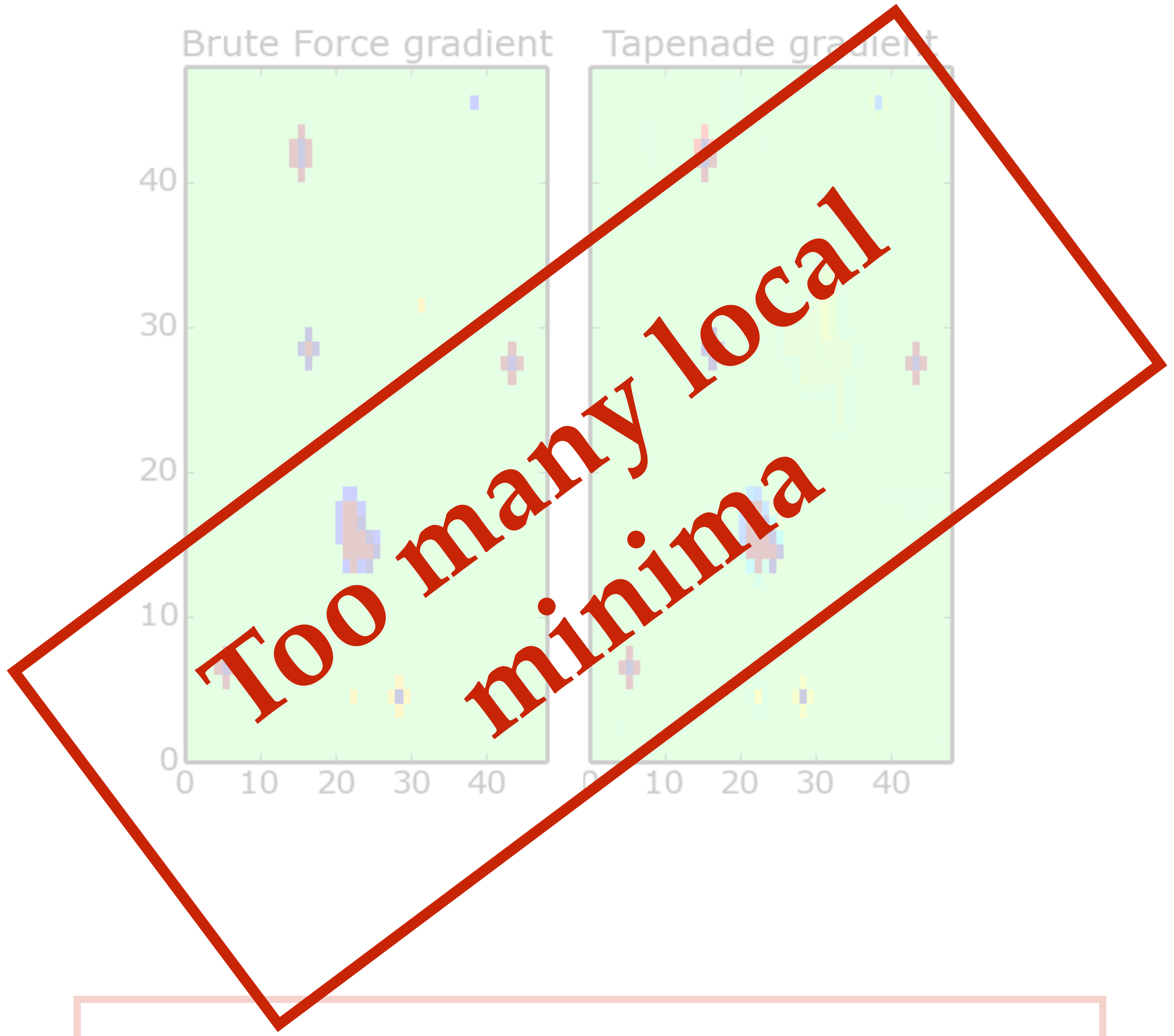
Using the gradient of the cost function



Adjoint method

B) Find new initial conditions reproducing the data (I)

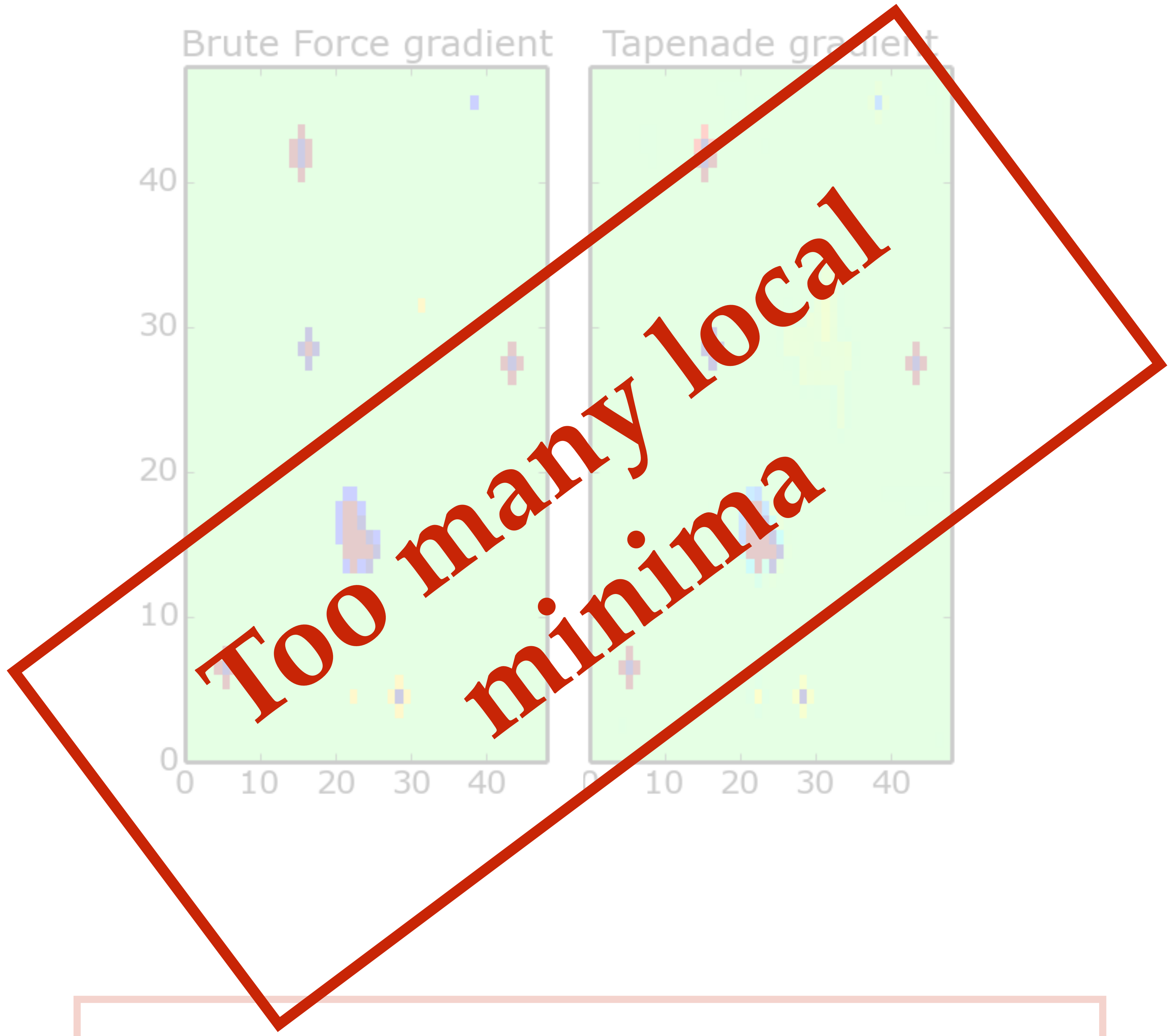
Using the gradient of the cost function



Adjoint method

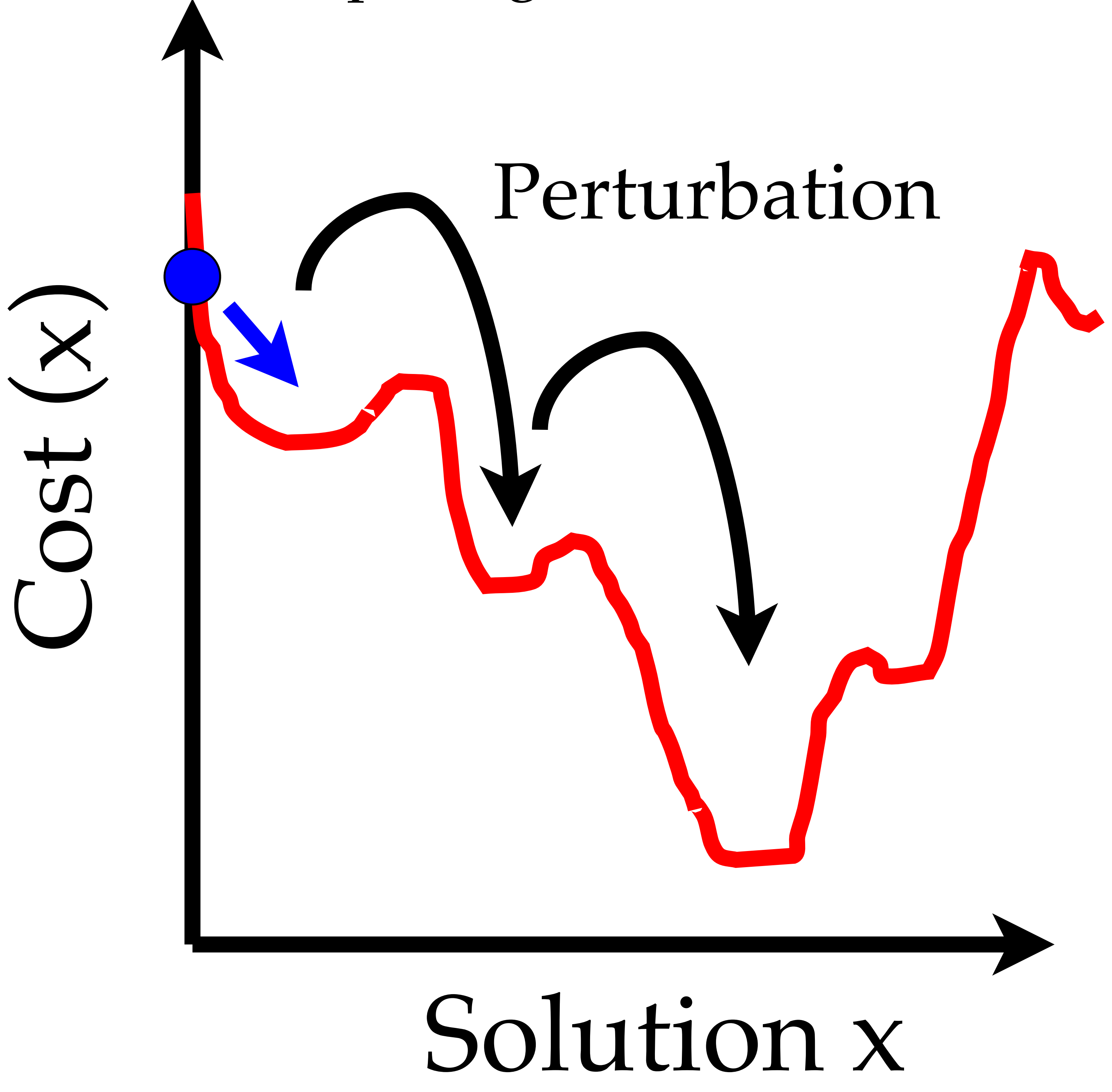
B) Find new initial conditions reproducing the data (I)

Using the gradient of the cost function



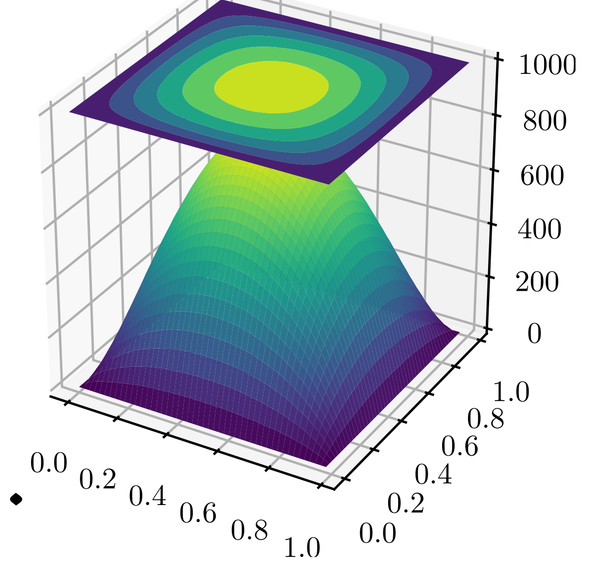
Adjoint method

Using more robust but more expensive methods (w/o explicit gradient calculation)



Simulated annealing

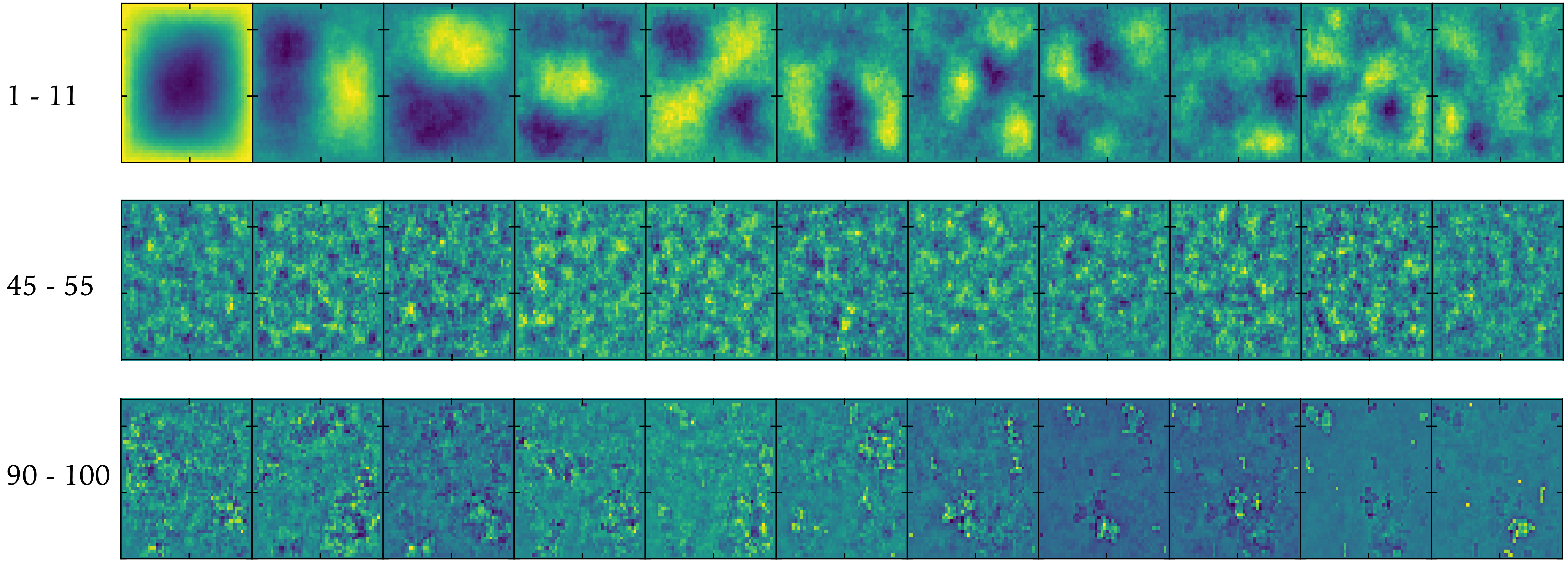
B) Find new initial conditions reproducing the data (II)



Simulated annealing is not very efficient: N^2 realizations of the model are used…

Idea: reduce this number by projecting the sandpile model on its **eigenvectors**

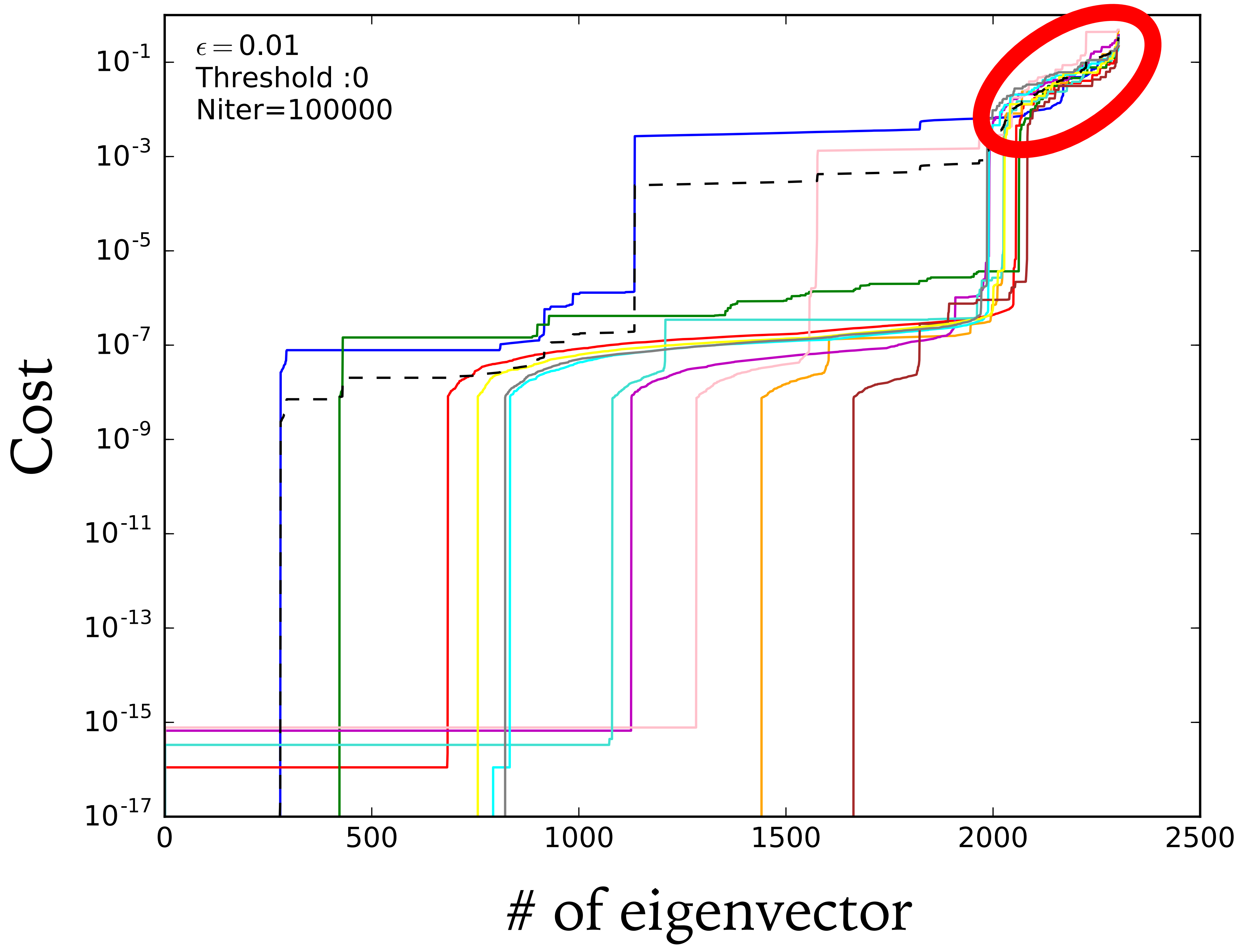
☛ Diagonalize co-variance matrix $$\mathcal{C}(\mathbf{x}) = \overline{(Y(\mathbf{x},t) - Y_0(\mathbf{x}))^T \cdot (Y(\mathbf{x},t) - Y_0(\mathbf{x}))}$$



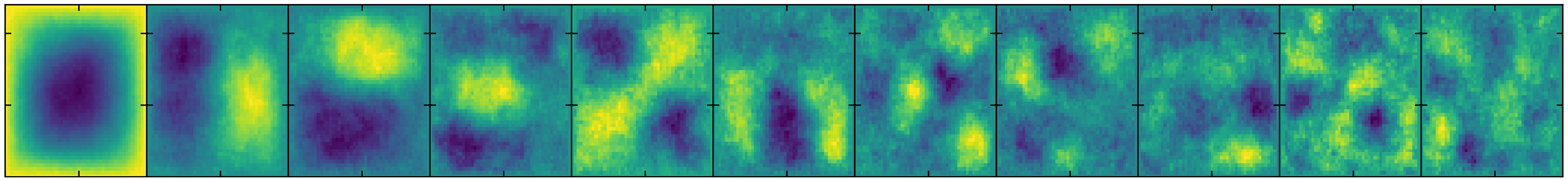
In this case, **N^2 = 2304 eigenvectors**

B) Find new initial conditions reproducing the data (II)

Sensitivity of solution to specific eigenvector

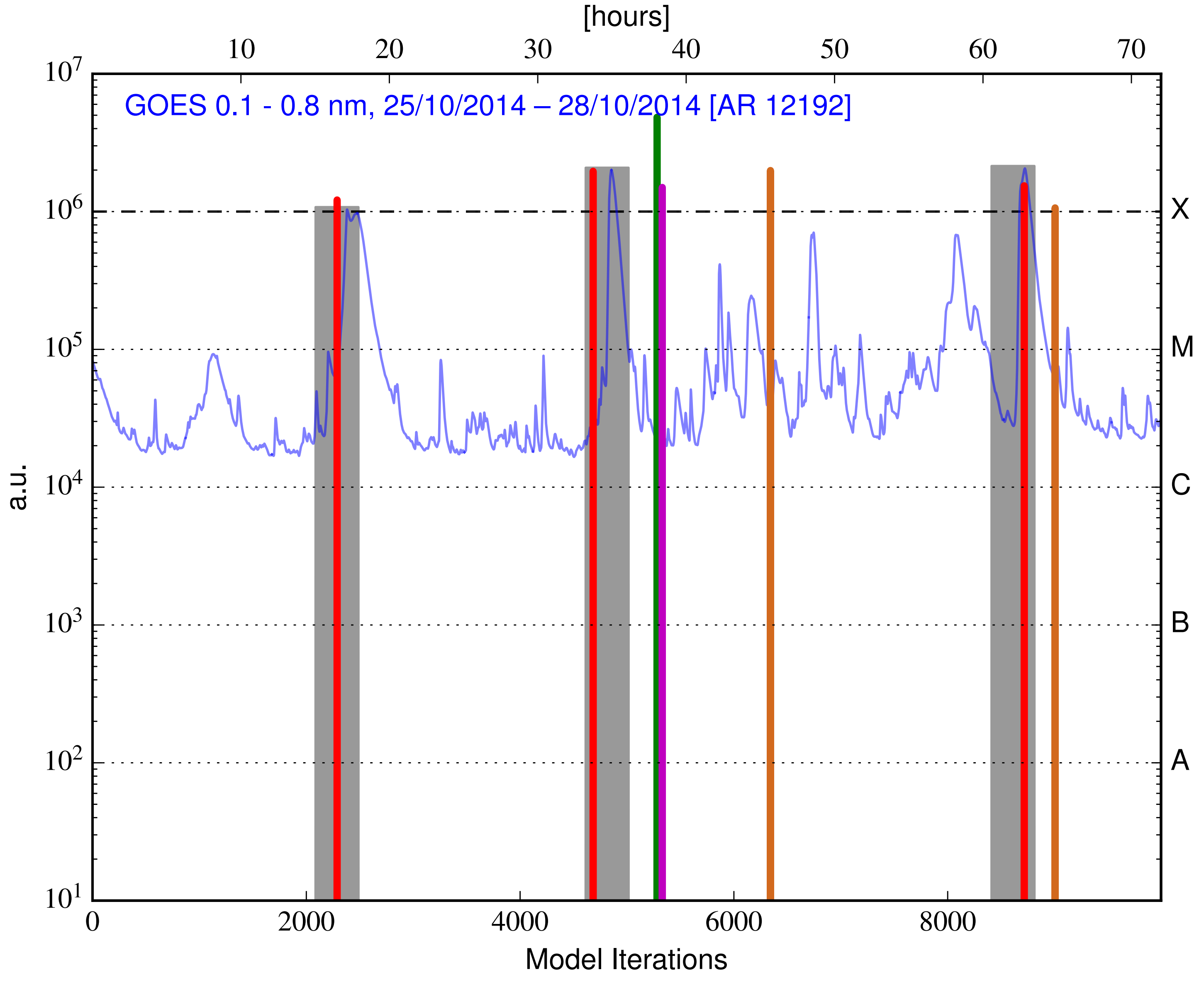


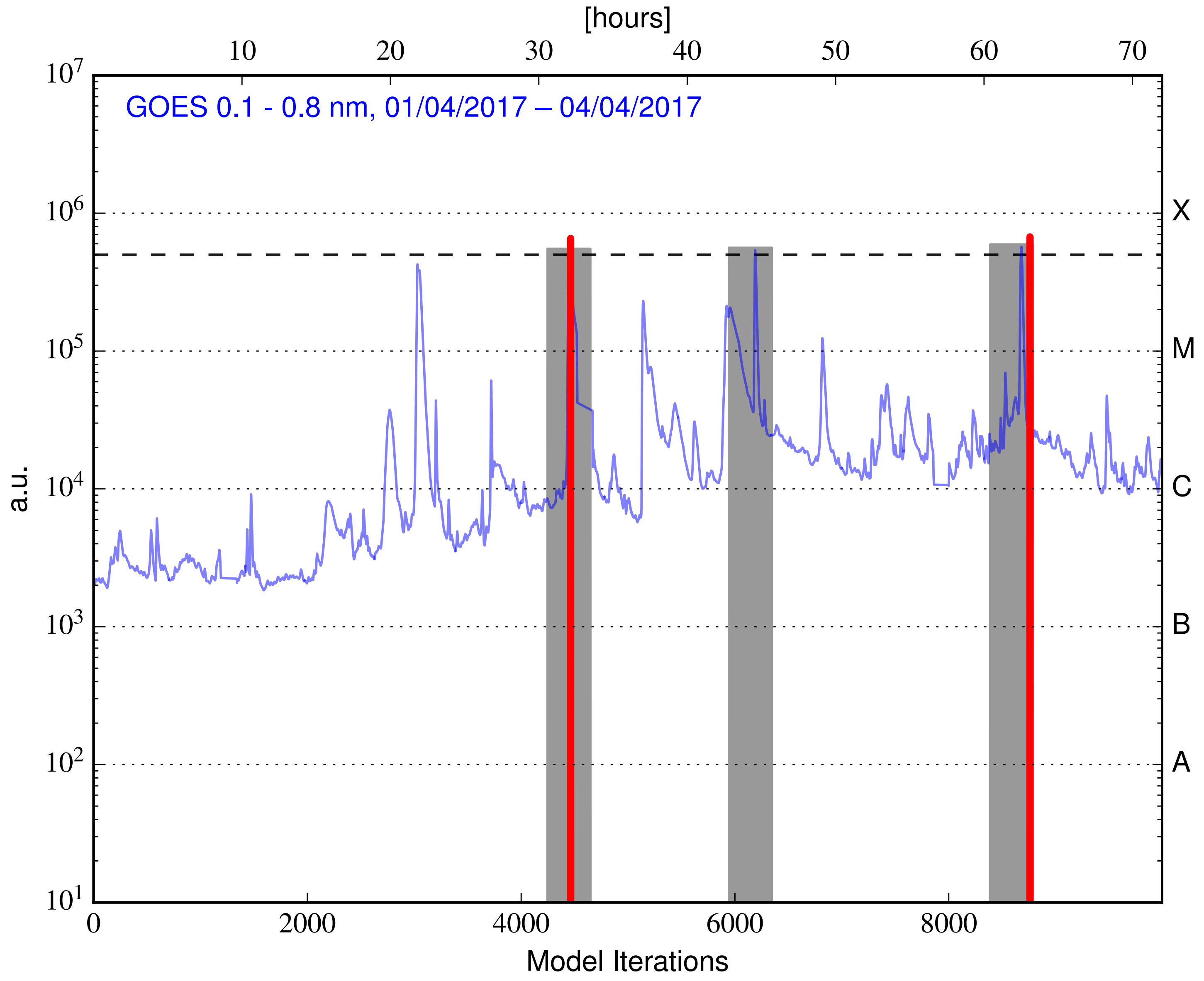
It seems to be possible to use a reduced set of eigenvectors to perform data assimilation…
work in progress!



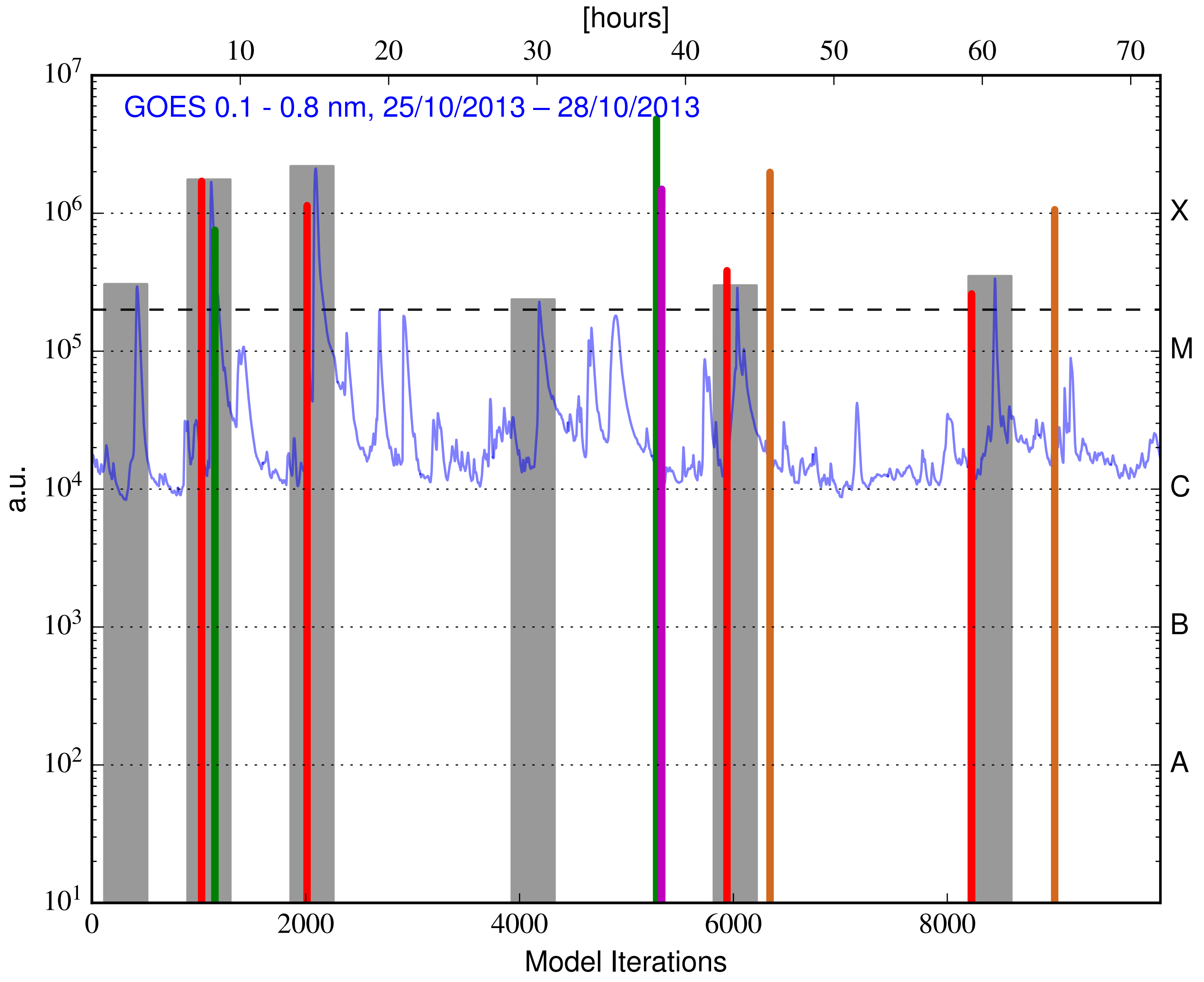
C) Towards a predictive tool…

3 X-flares



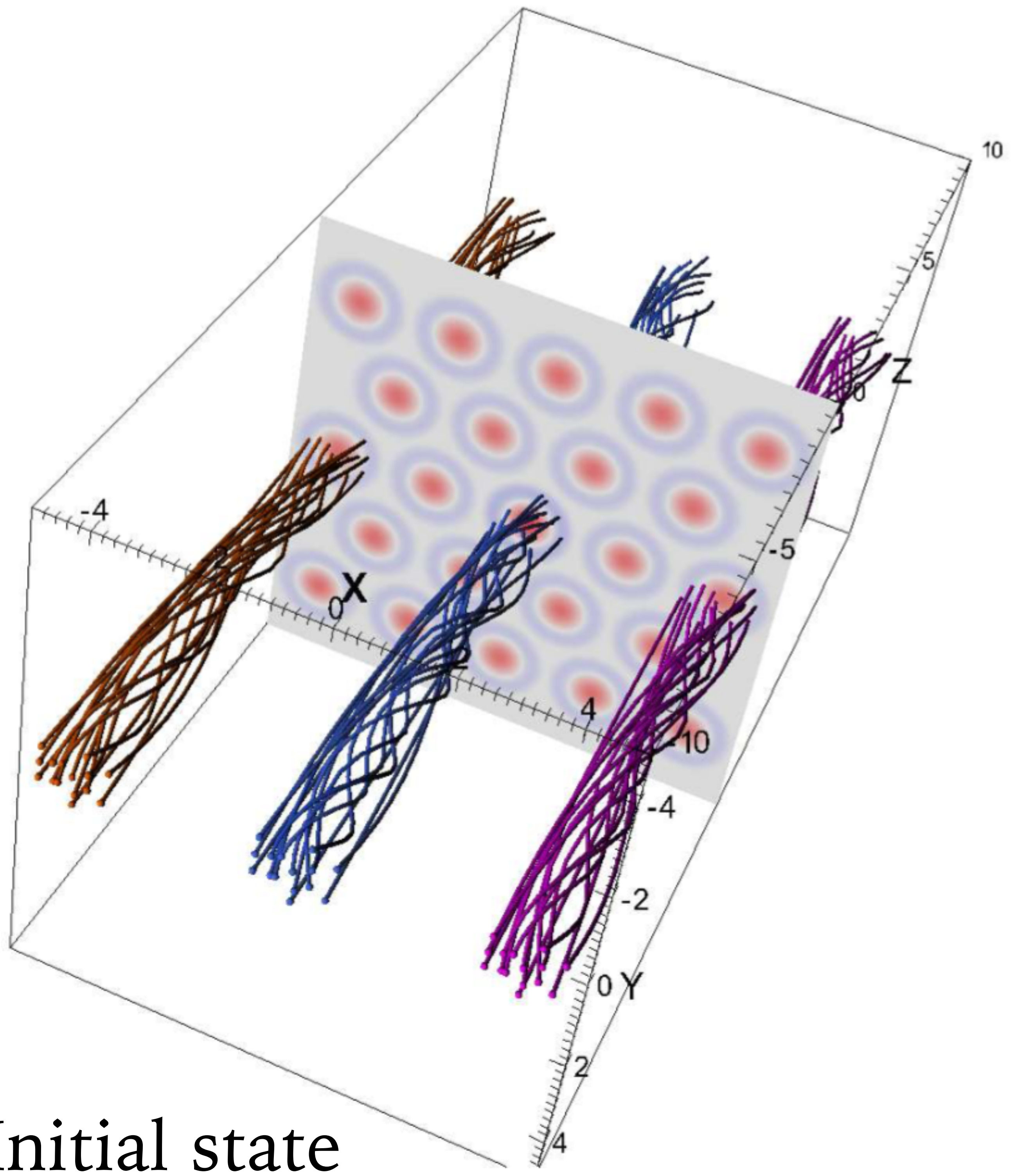


3 M-flares



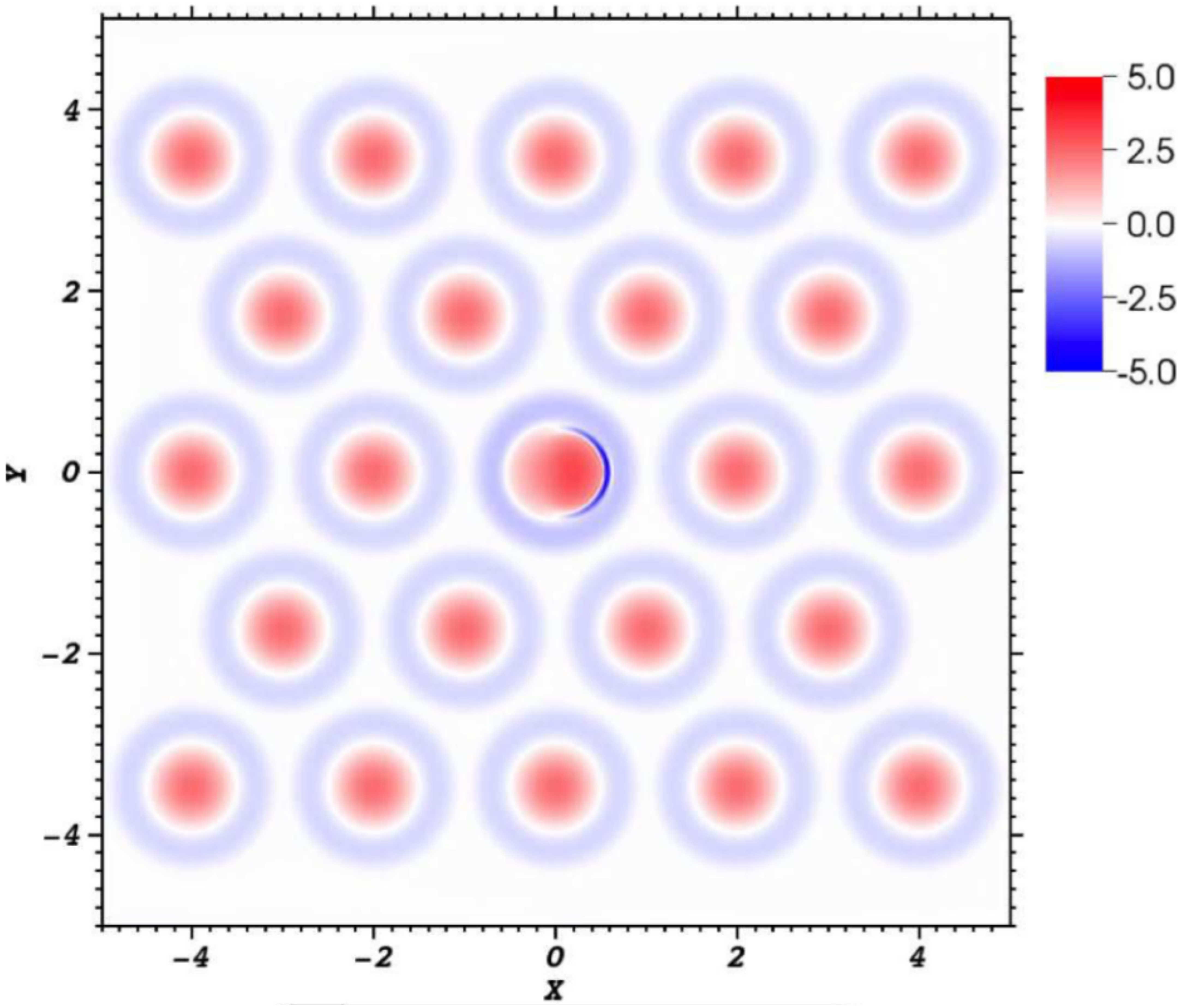
2 X-flares
+
4 M-flares

Physical interpretation(s) of sandpiles



Initial state

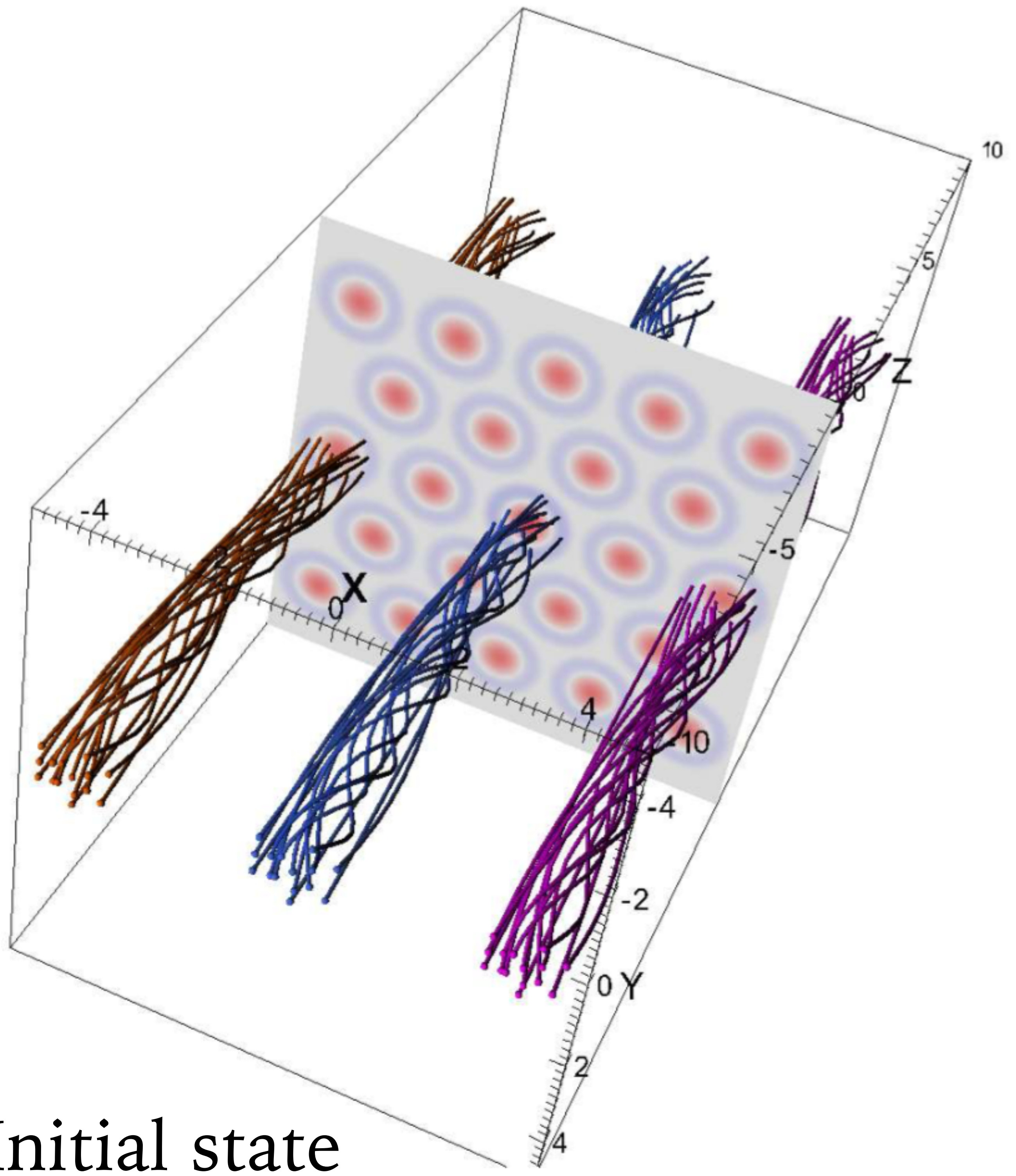
A first resistive MHD realization of avalanche-type behaviour in a coronal loop composed of 23 distinct magnetic threads



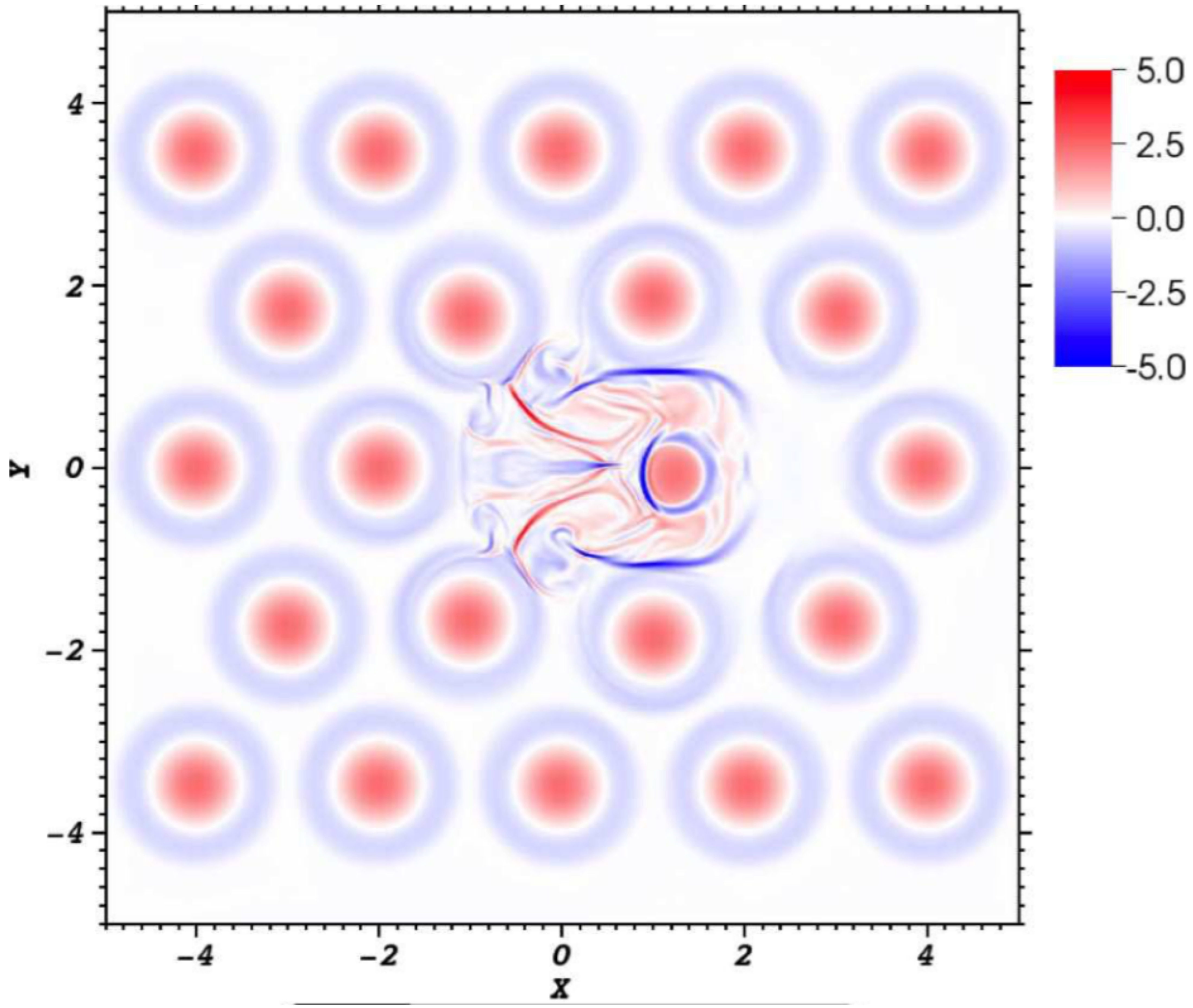
[Hood+ 2016]

Physical interpretation(s) of sandpiles

A first resistive MHD realization of avalanche-type behaviour in a coronal loop composed of 23 distinct magnetic threads



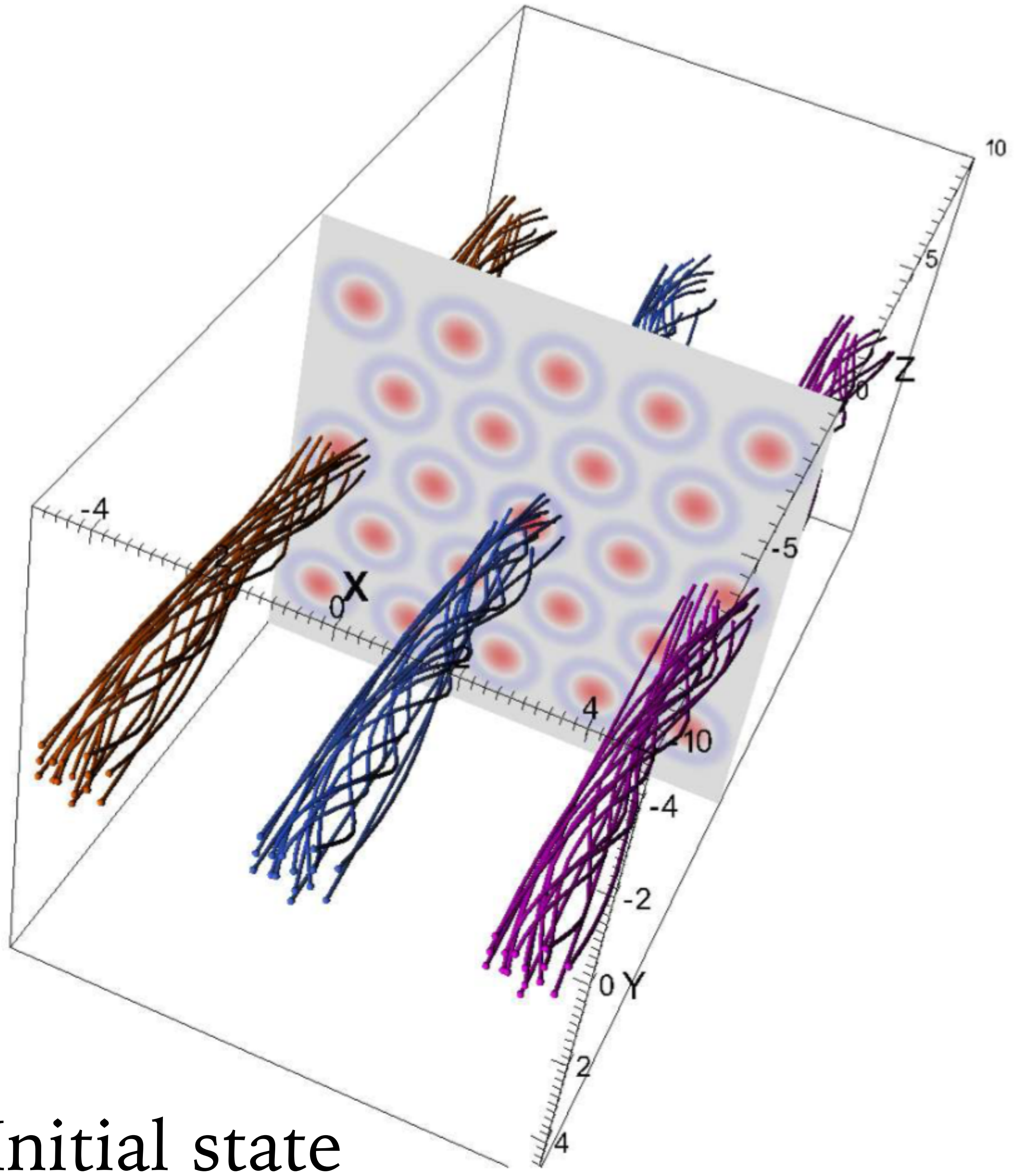
Initial state



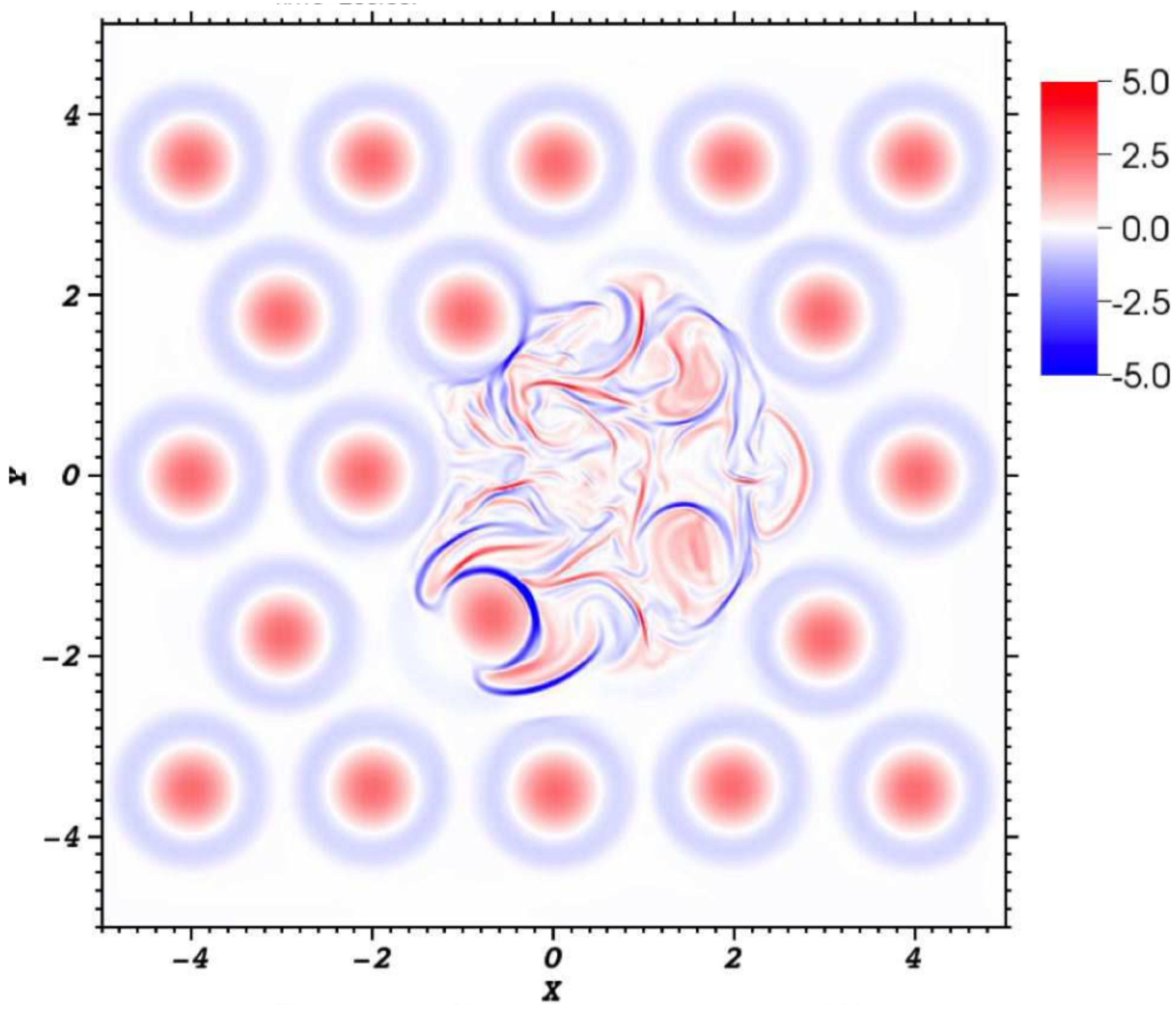
[Hood+ 2016]

Physical interpretation(s) of sandpiles

A first resistive MHD realization of avalanche-type behaviour in a coronal loop composed of 23 distinct magnetic threads



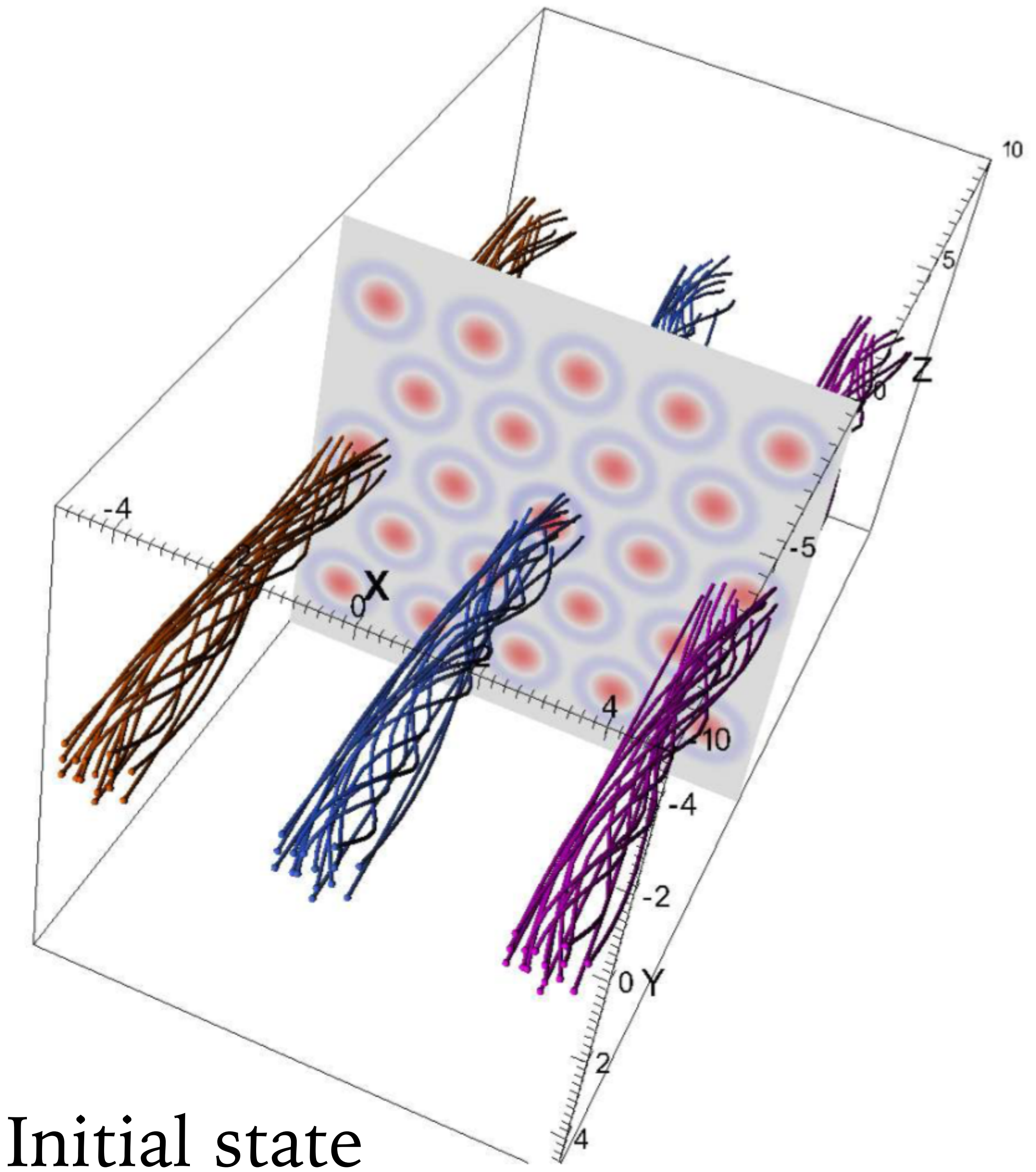
Initial state



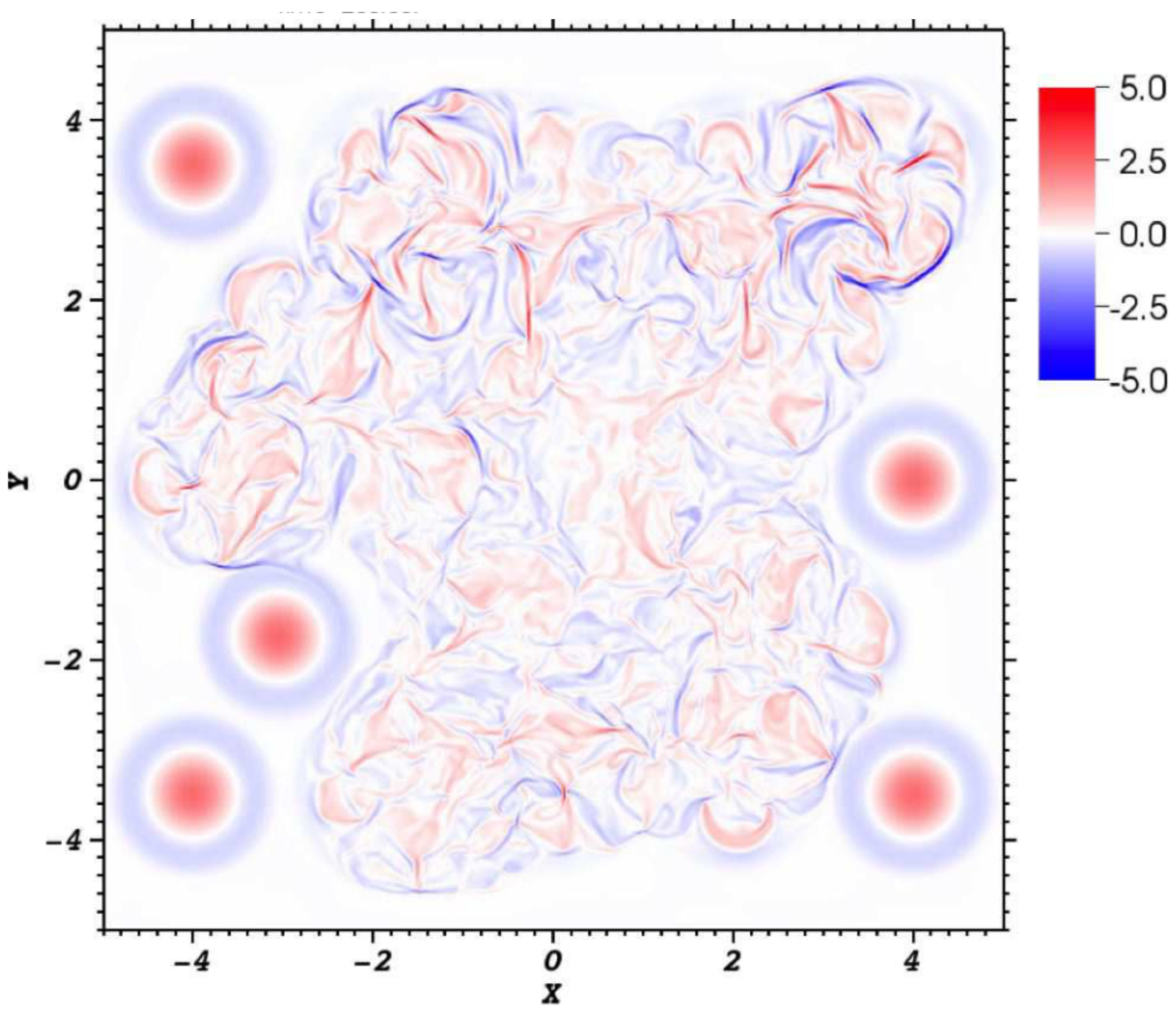
[Hood+ 2016]

Physical interpretation(s) of sandpiles

A first resistive MHD realization of avalanche-type behaviour in a coronal loop composed of 23 distinct magnetic threads

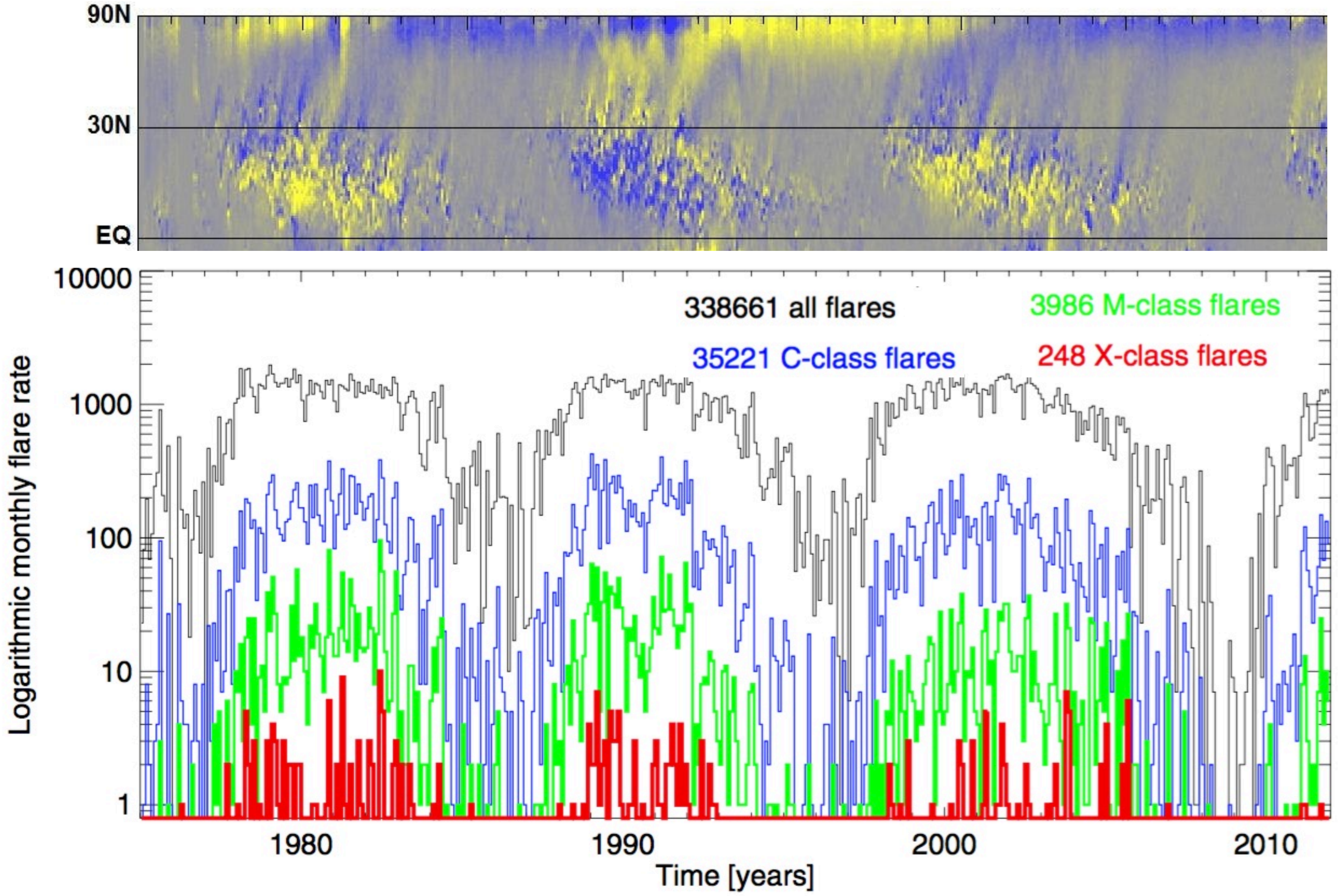


Initial state



[Hood+ 2016]

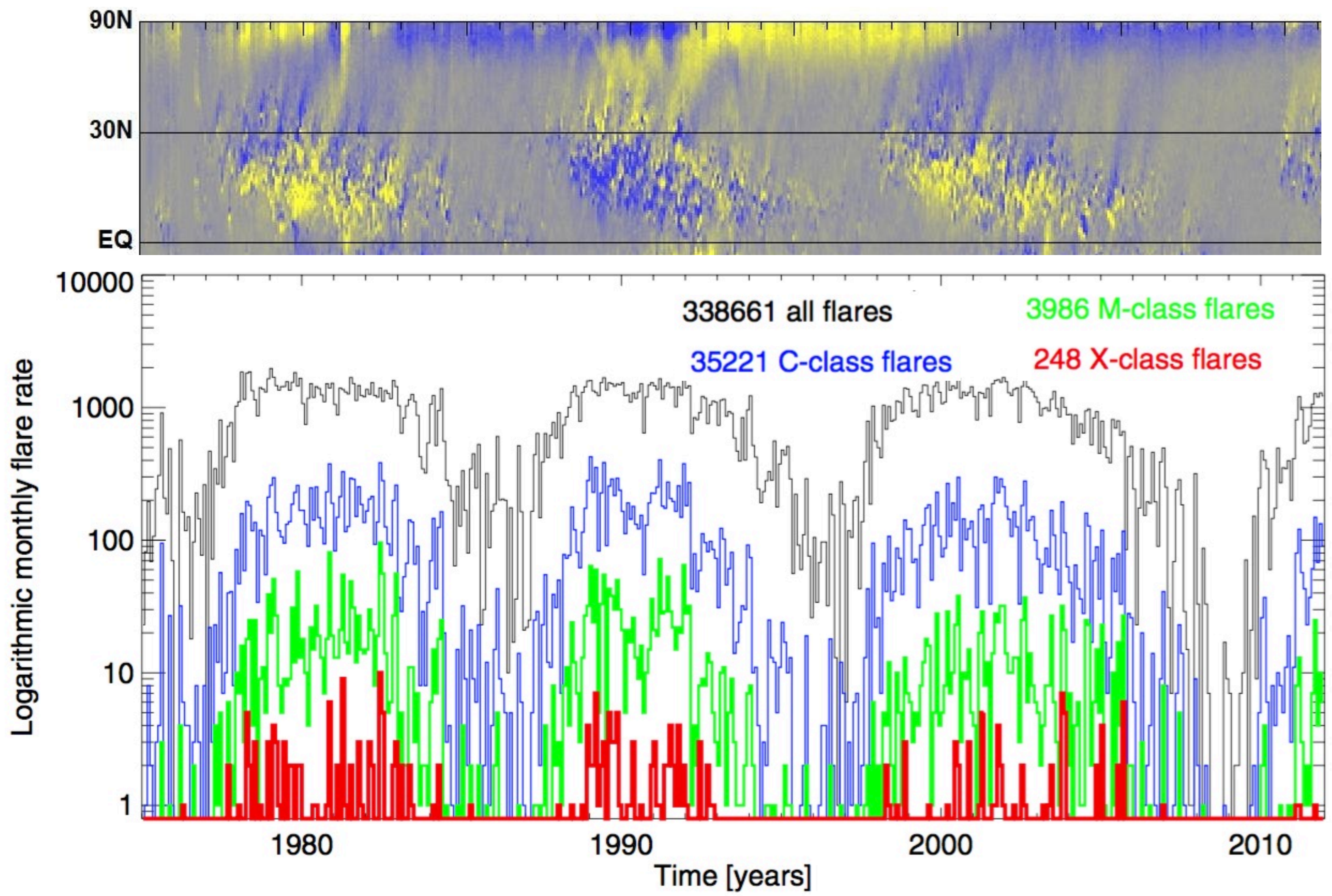
The number of flares strongly varies during the solar cycle



Line of sight magnetic field in the north hemisphere of the Sun

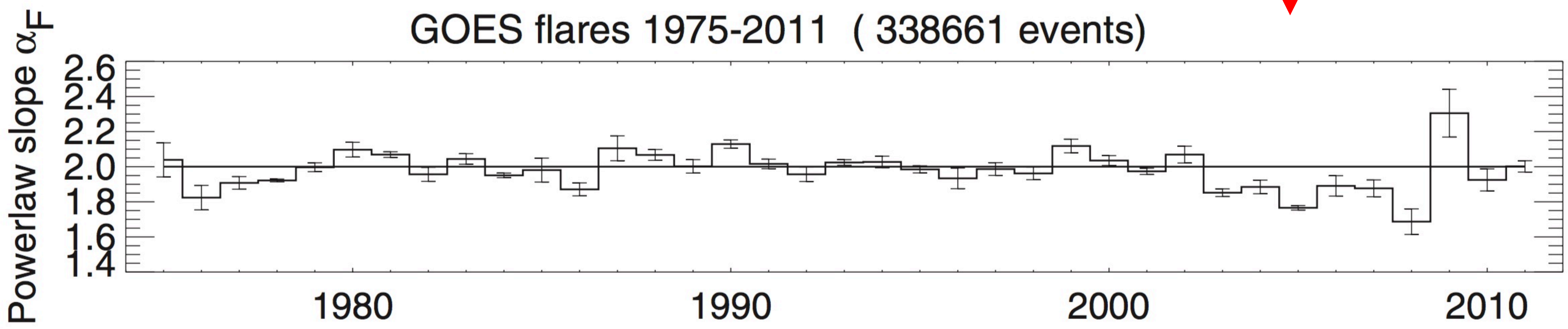
[Aschwanden & Freeland 2012]

The number of flares strongly varies during the solar cycle



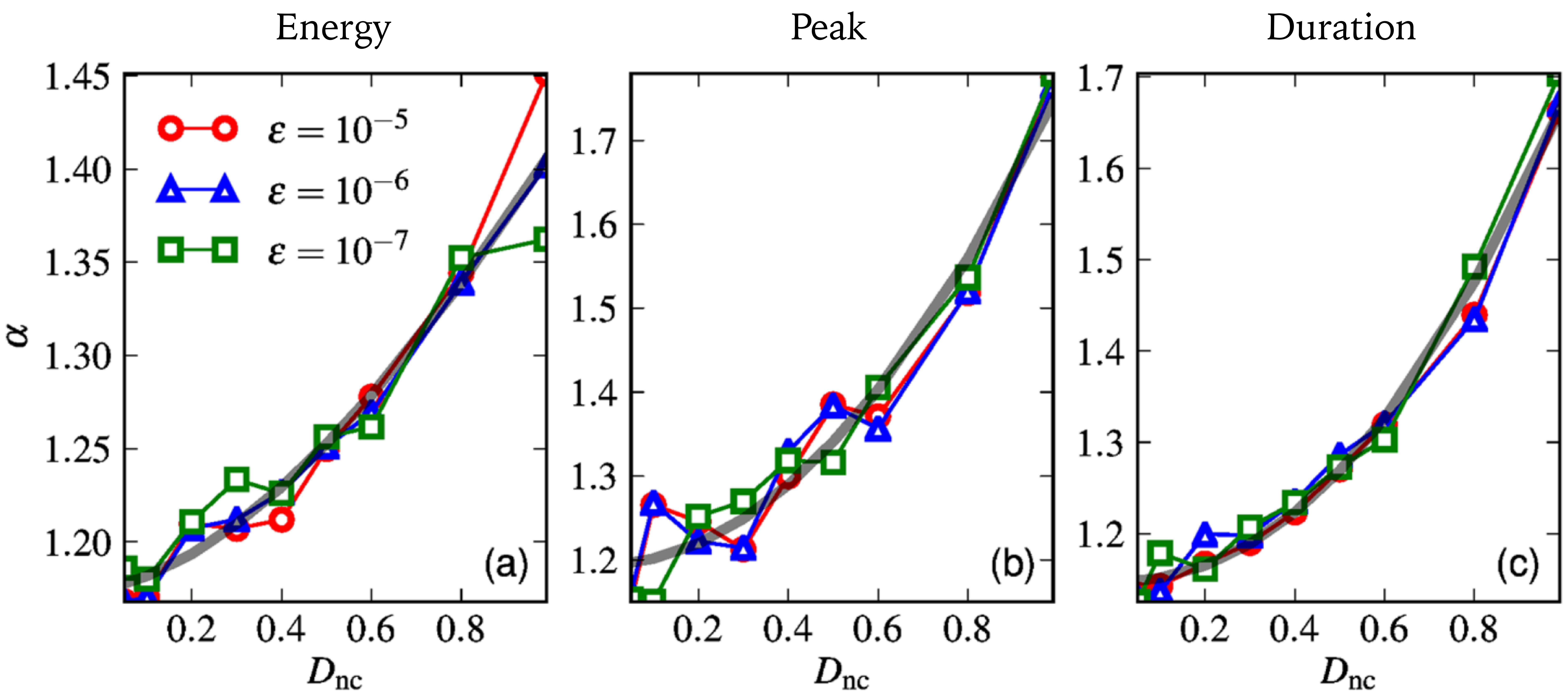
Line of sight magnetic field in the north hemisphere of the Sun

But no correlation of the power-law slope of the peak X-ray flux during flares



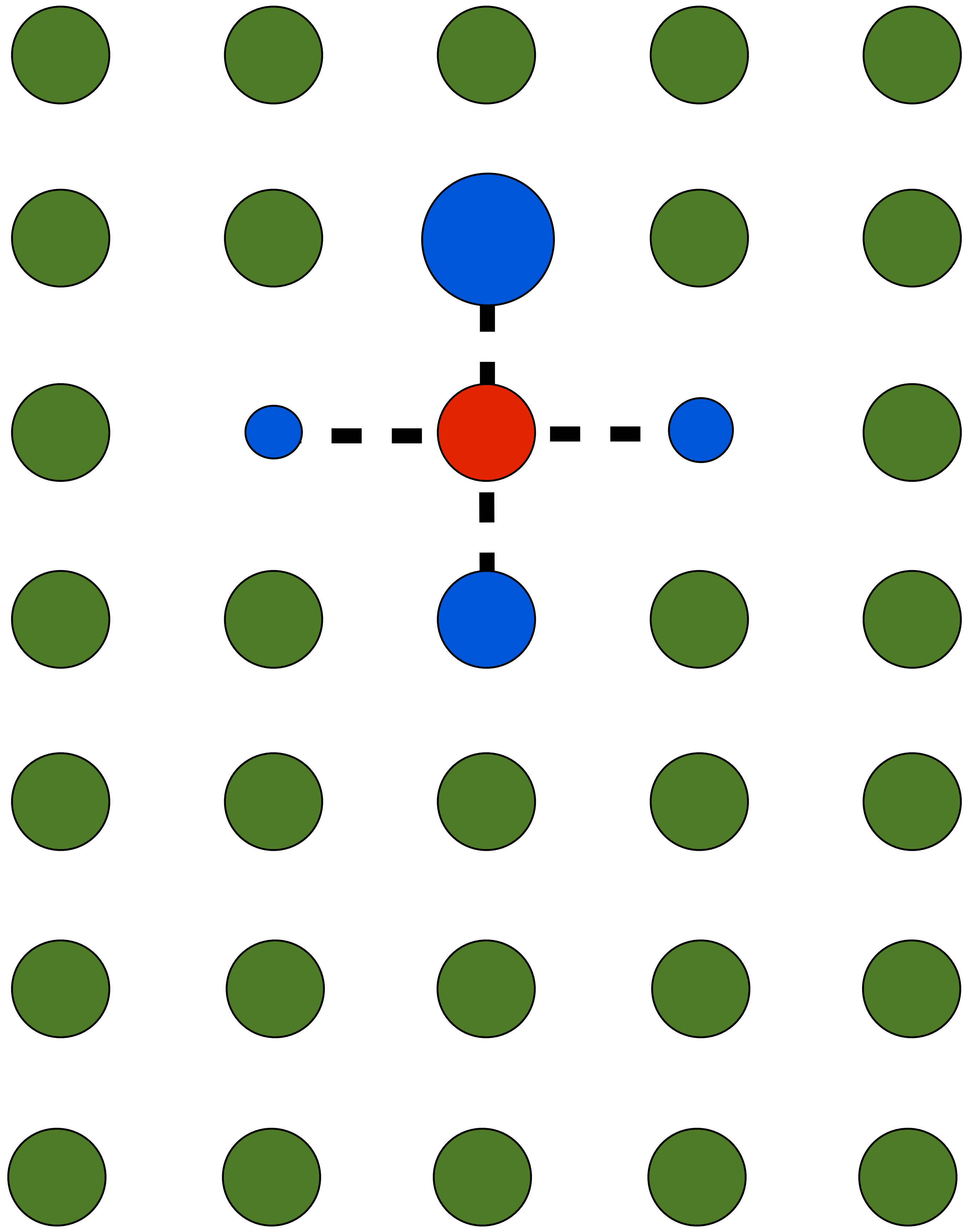
[Aschwanden & Freeland 2012]

Power-law exponents in the D-models



[Strugarek+ 2014]

Deterministically-driven sandpile models



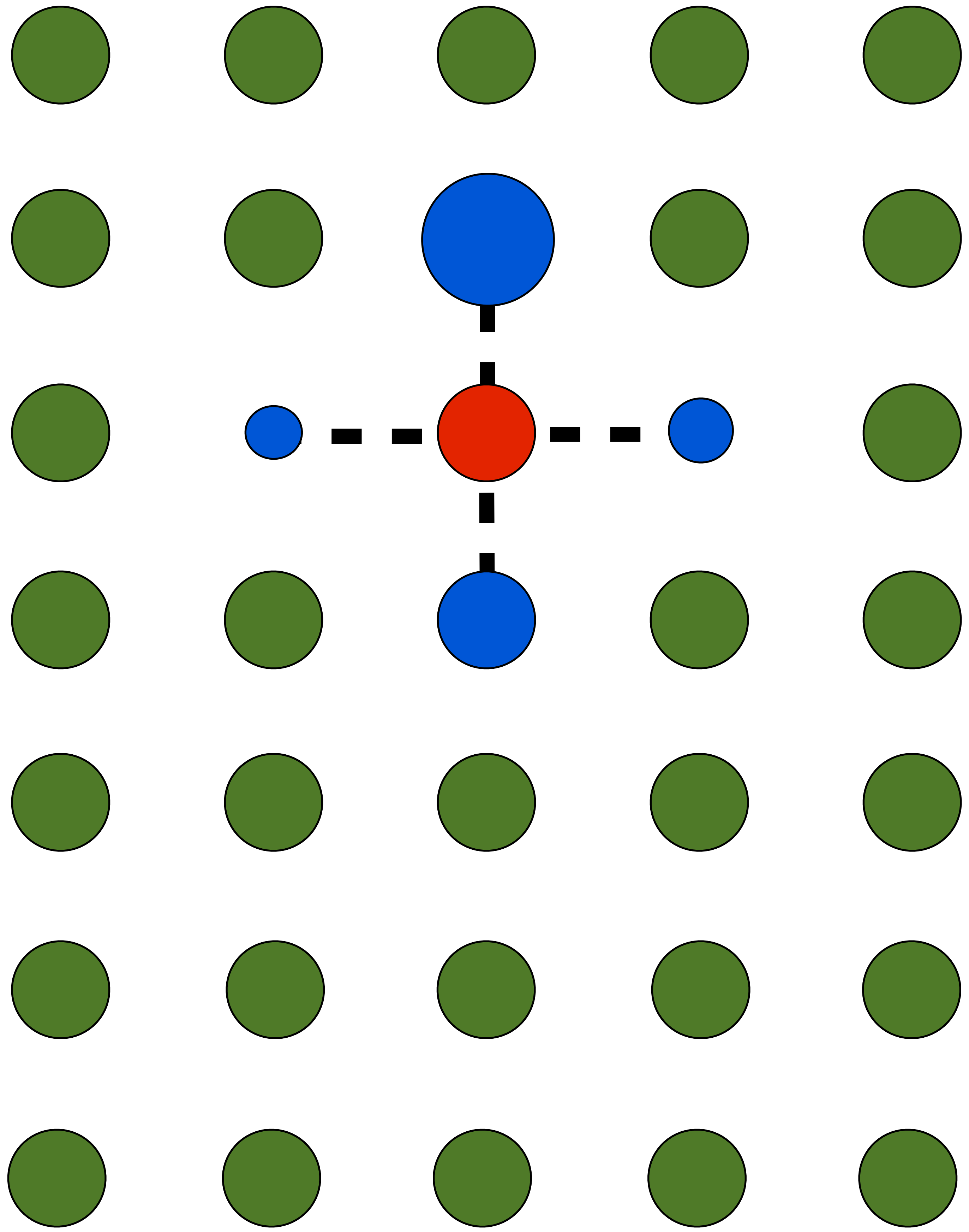
Deterministic driving on **all nodes**

(**Non)-Conservative** redistribution rule

Random process in:

[Strugarek+ 2014]

Deterministically-driven sandpile models



Deterministic driving on **all nodes**

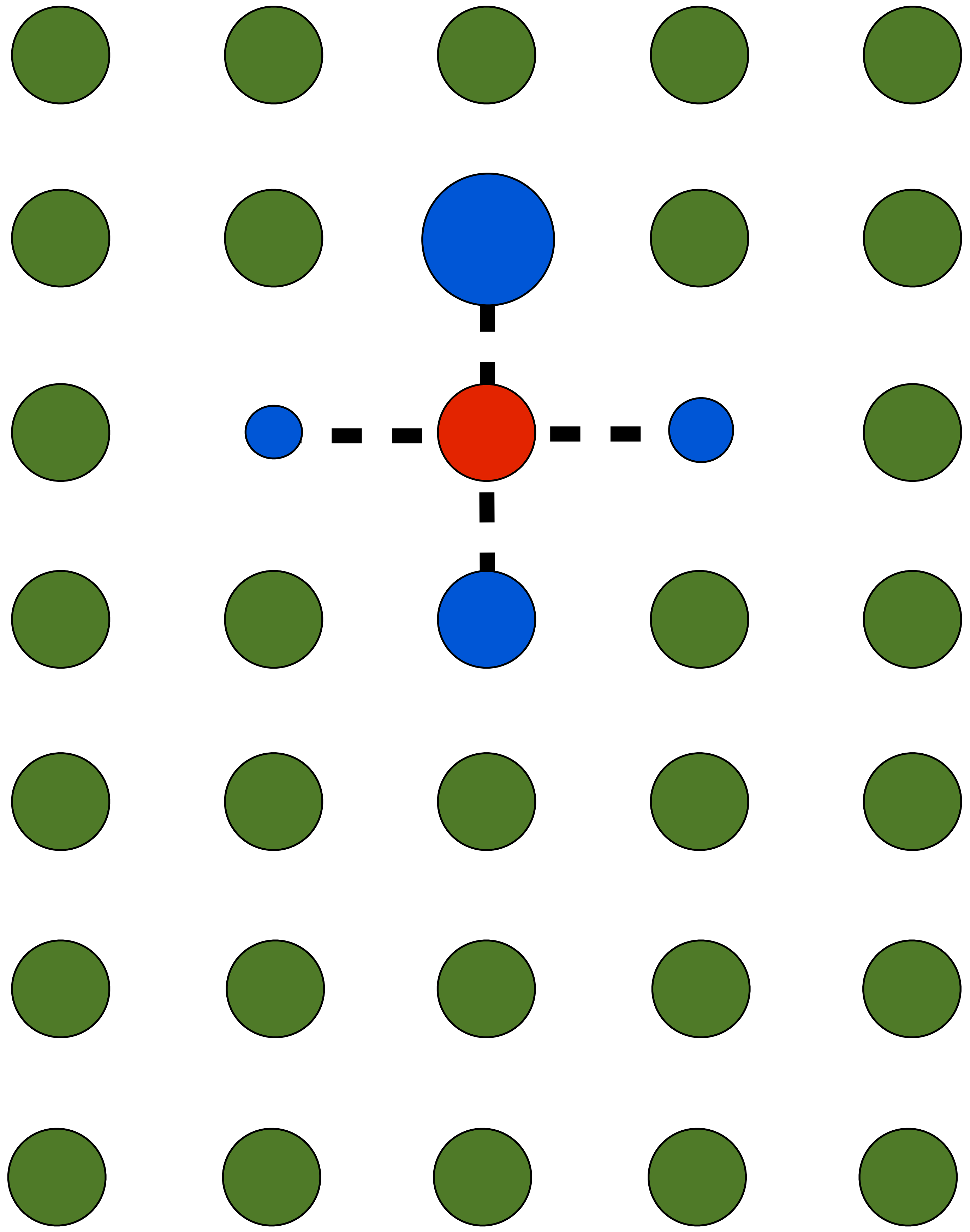
(**Non)-Conservative** redistribution rule

Random process in:

- **Threshold**

[Strugarek+ 2014]

Deterministically-driven sandpile models



Deterministic driving on **all nodes**

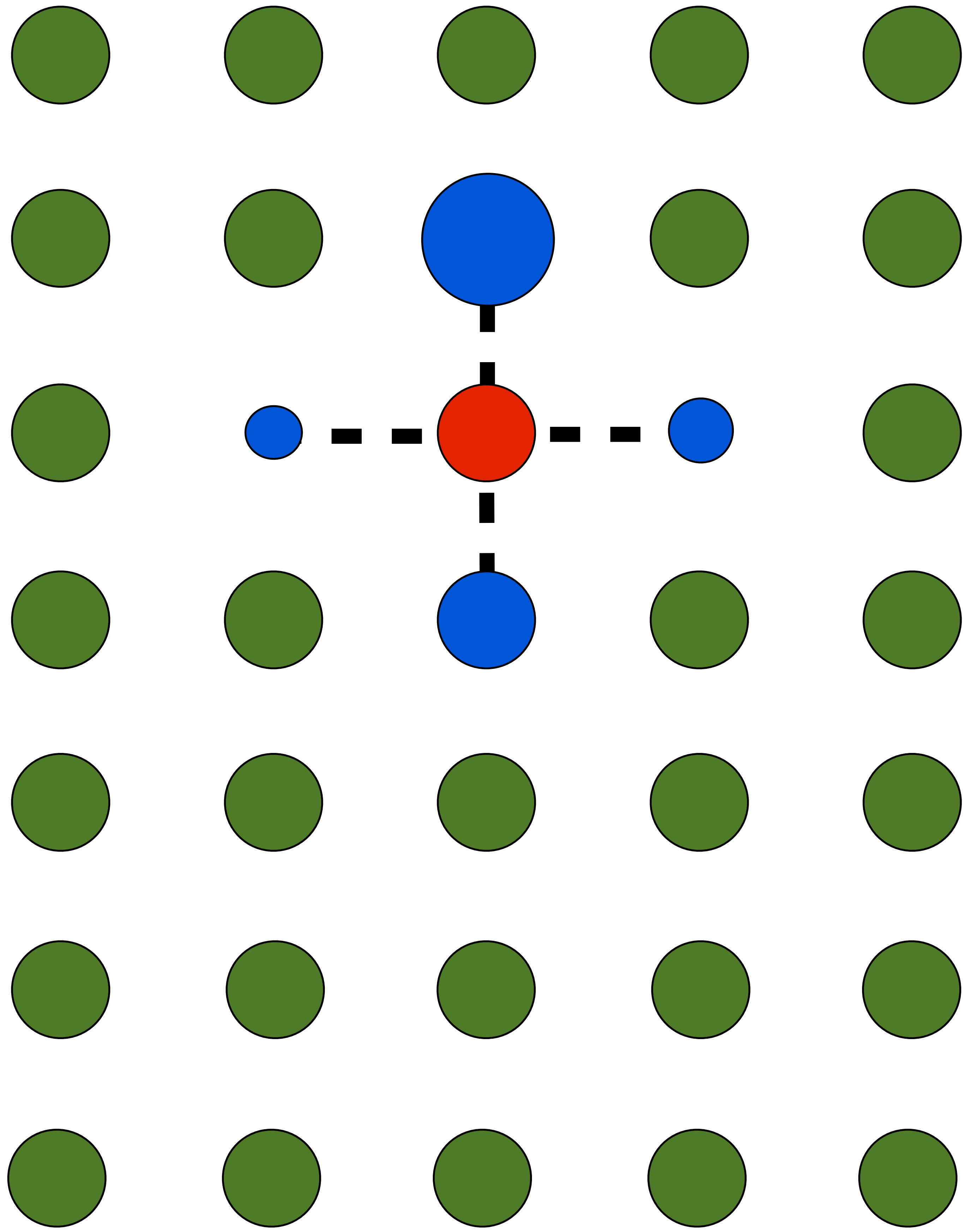
(Non)-Conservative redistribution rule

Random process in:

- **Threshold**
- **Redistribution**

[Strugarek+ 2014]

Deterministically-driven sandpile models



Deterministic driving on **all nodes**

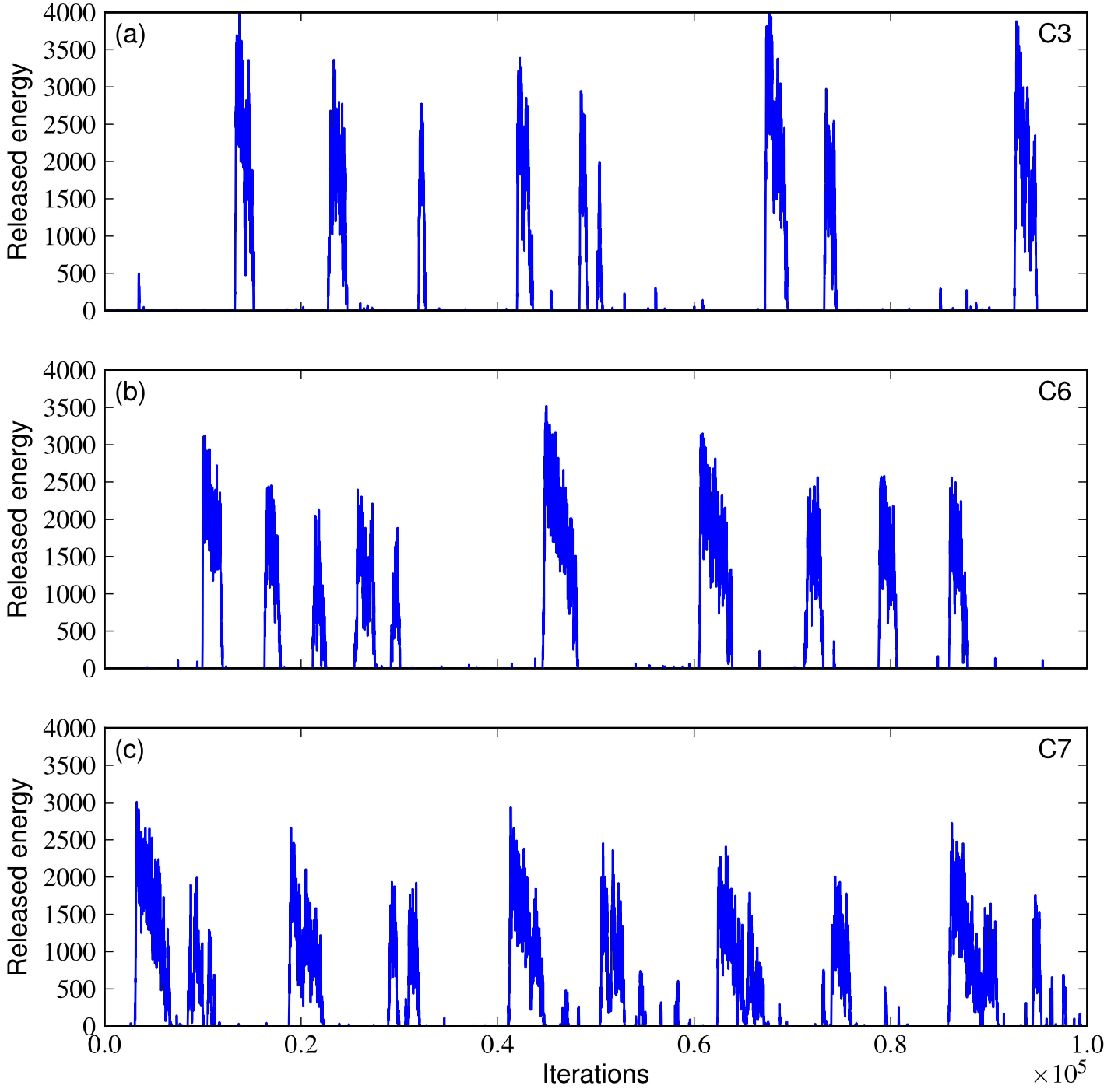
(Non)-Conservative redistribution rule

Random process in:

- **Threshold**
- **Redistribution**
- **Extraction**

[Strugarek+ 2014]

Conservative models do not reach the 'SOC' state



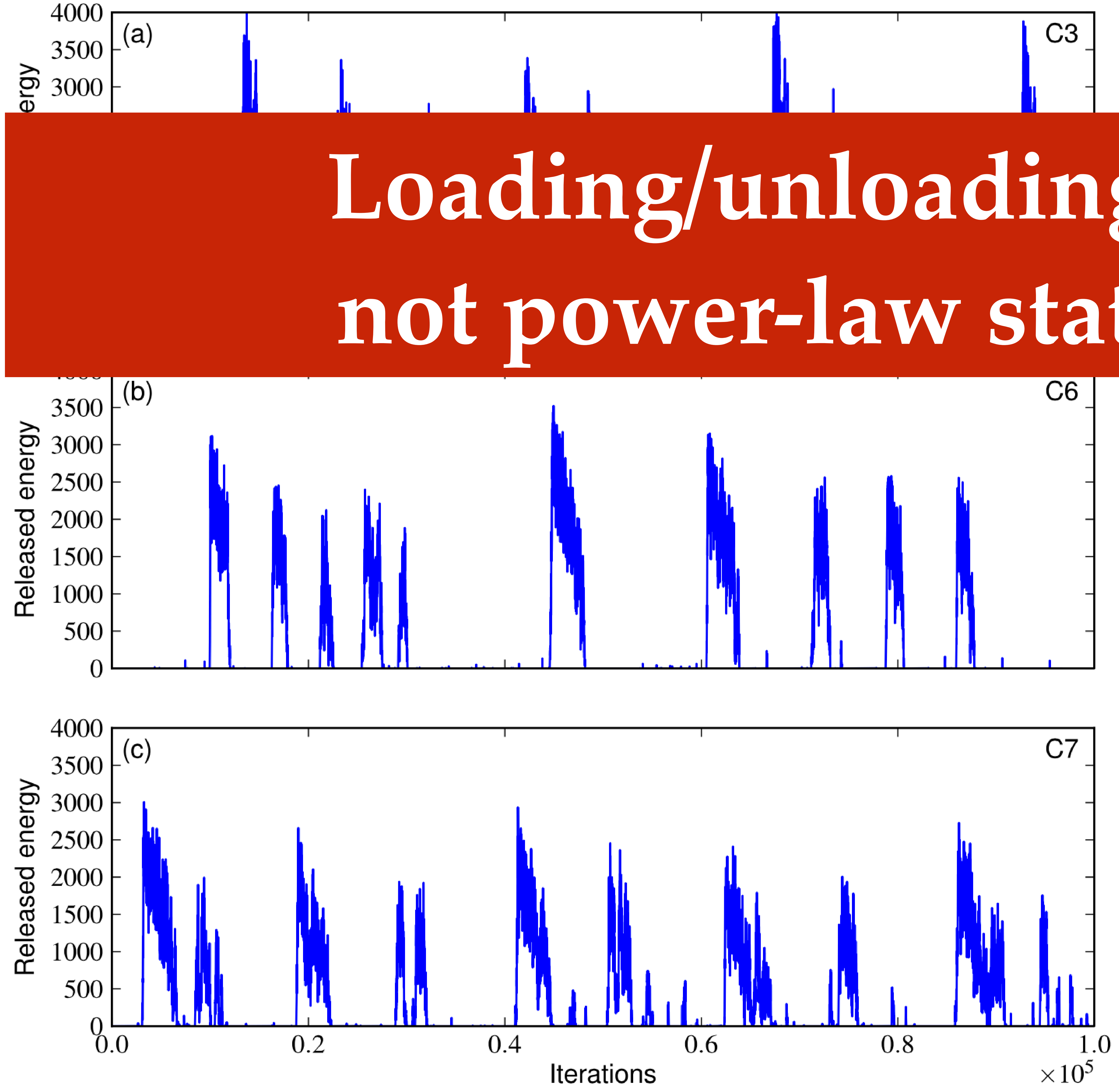
Random extraction

Random extraction
+
Random redistribution

Random extraction
+
Random redistribution
+
Random threshold

[Strugarek+ 2014]

Conservative models do not reach the ‘SOC’ state



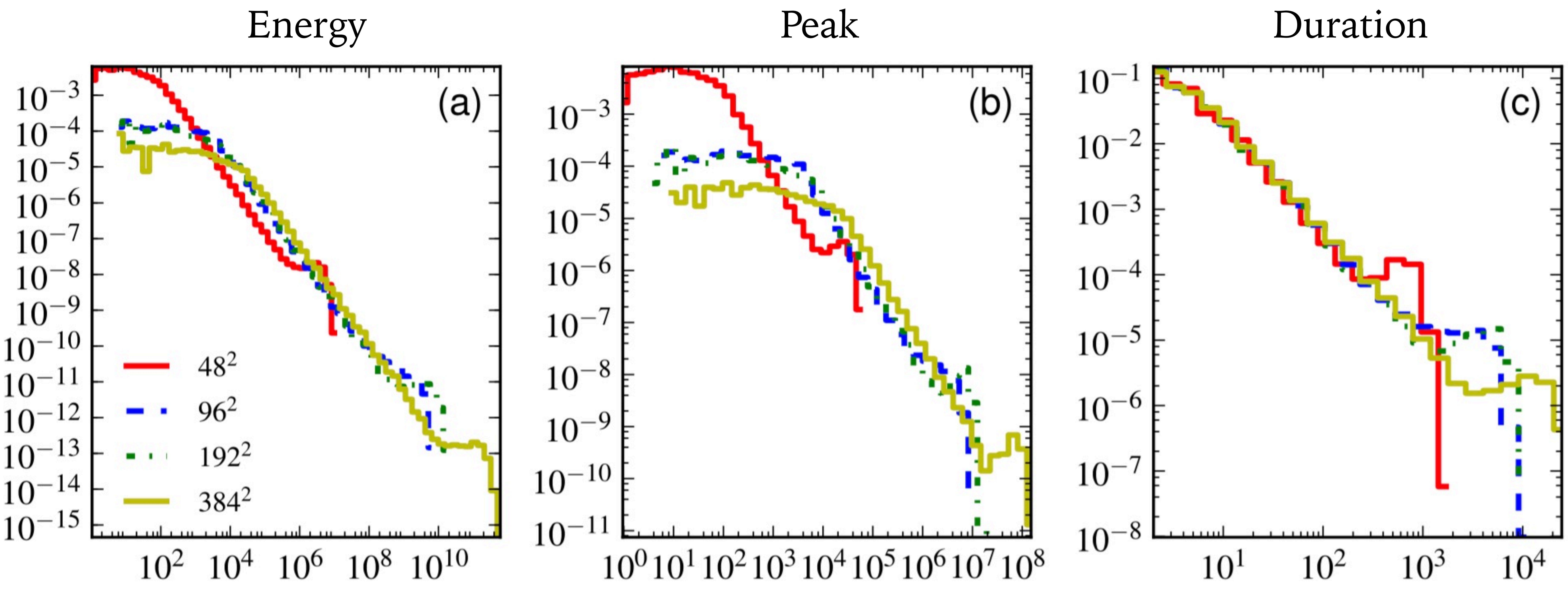
Loading/unloading cycles: not power-law statistics…

Random extraction
+
Random redistribution

Random extraction
+
Random redistribution
+
Random threshold

[Strugarek+ 2014]

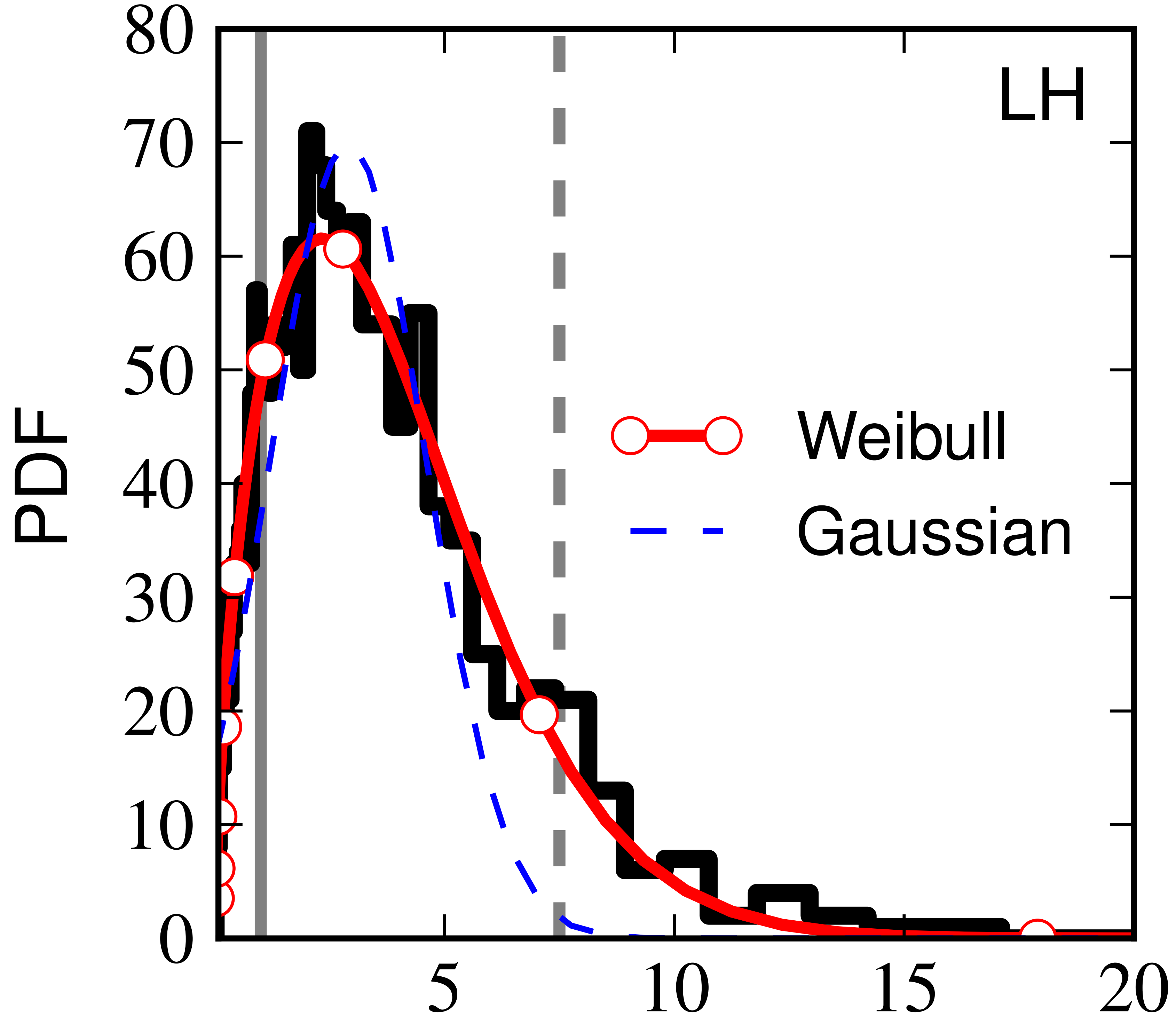
Lattice size: it only affects the accessible energy range



[Strugarek+ 2014]

Robustness of one large event (standard model)

Energy of the largest event
in the time window

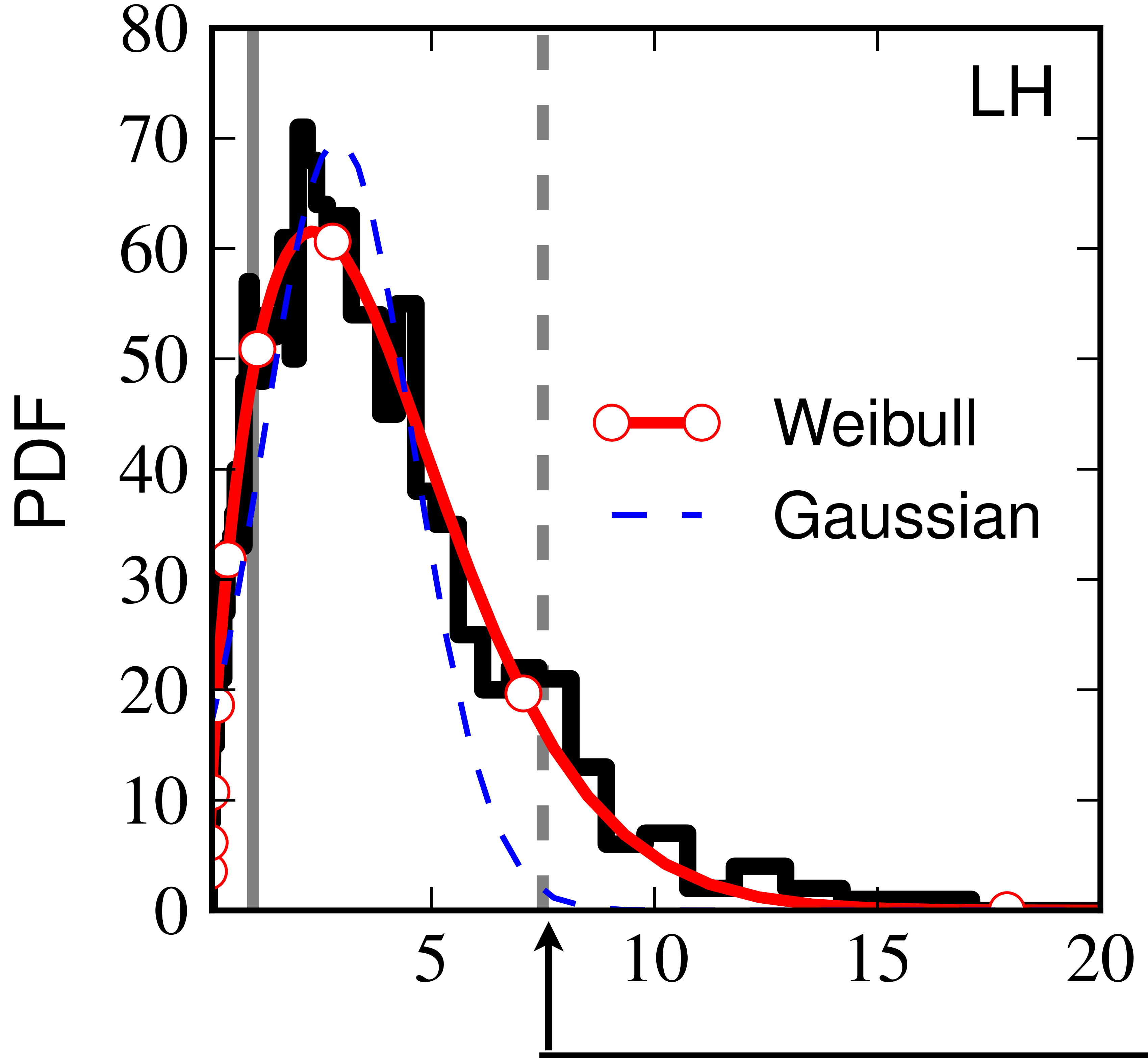


2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (standard model)

Energy of the largest event
in the time window



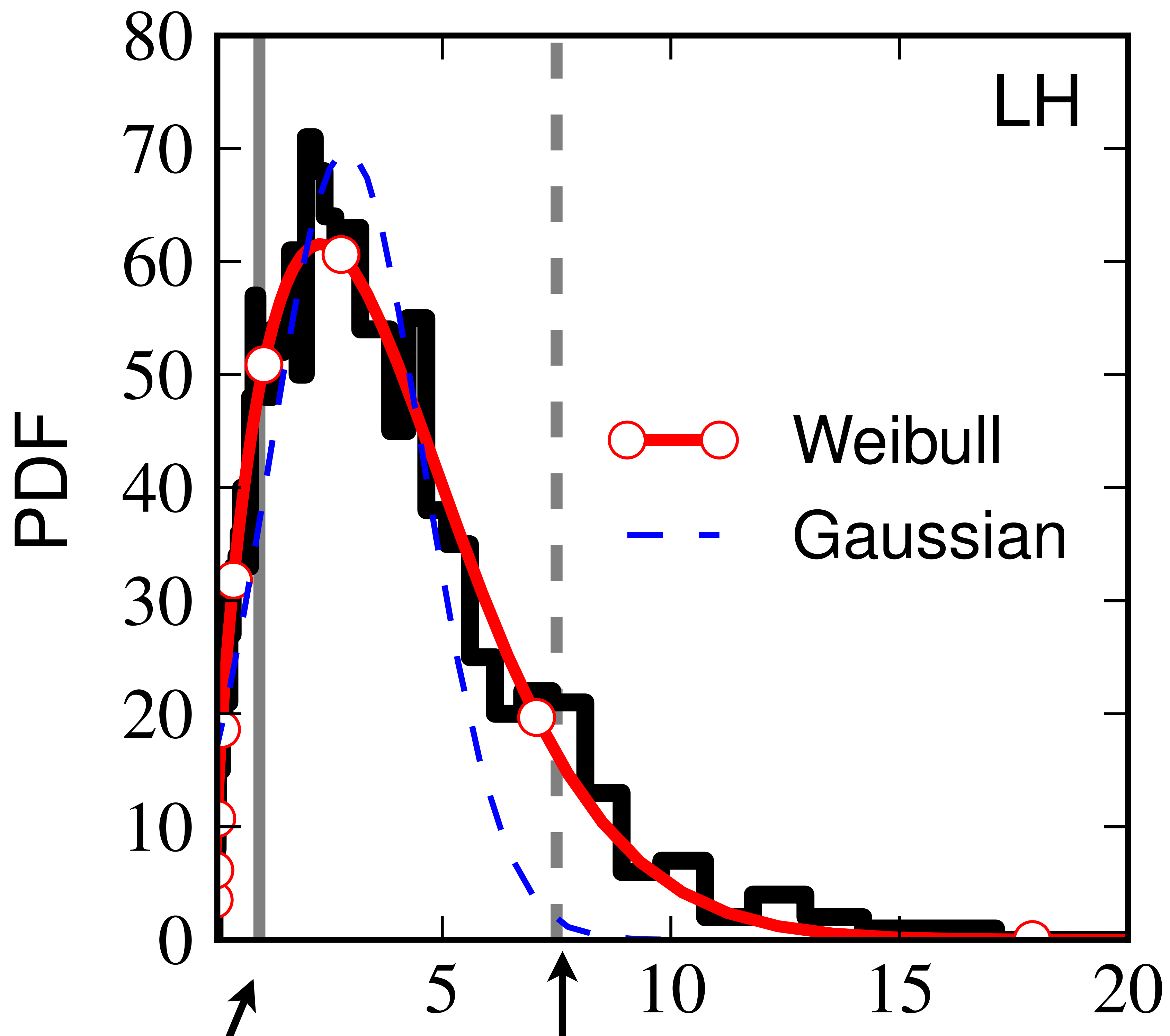
Targeted event energy

2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (standard model)

Energy of the largest event
in the time window



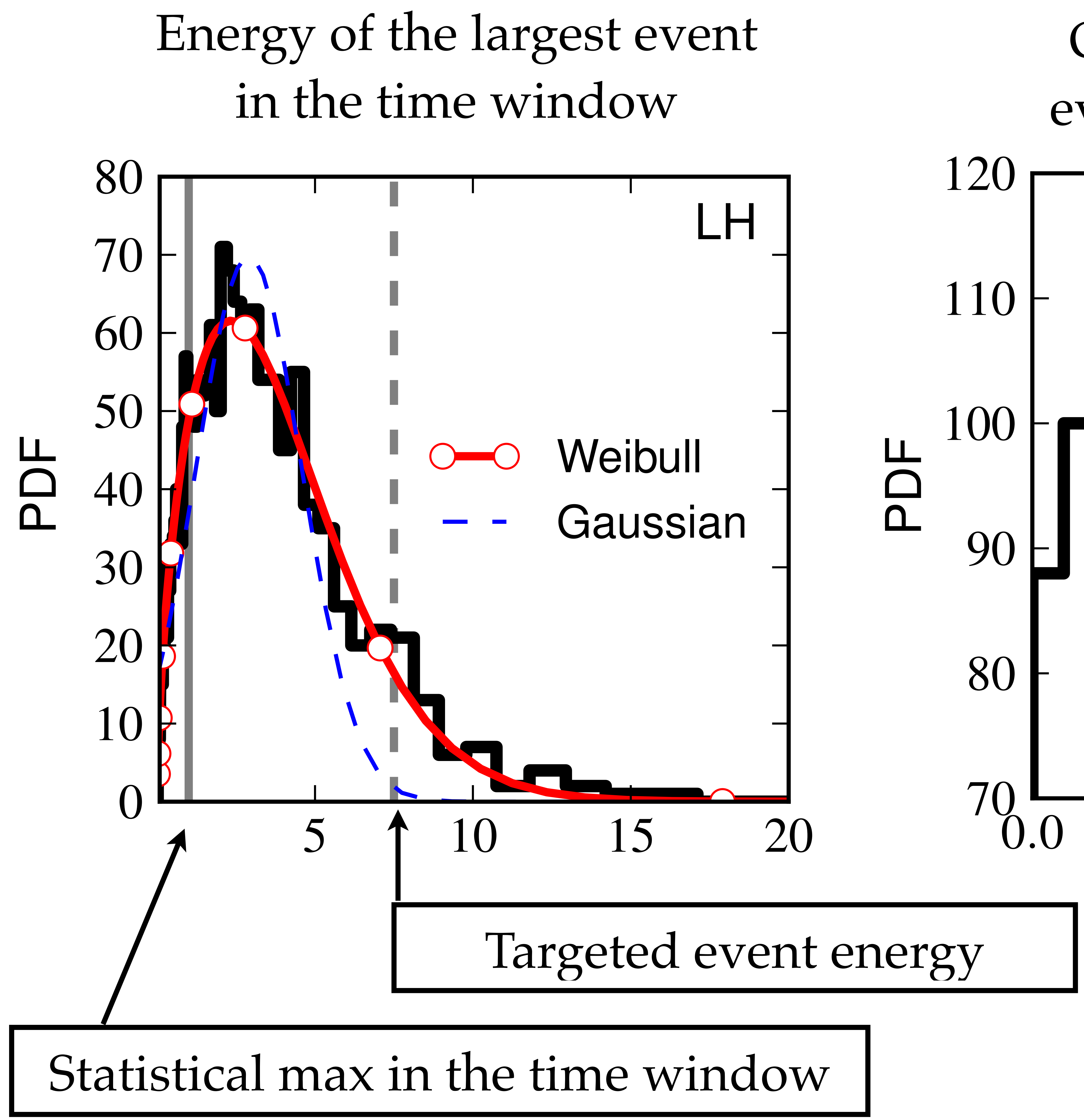
Targeted event energy

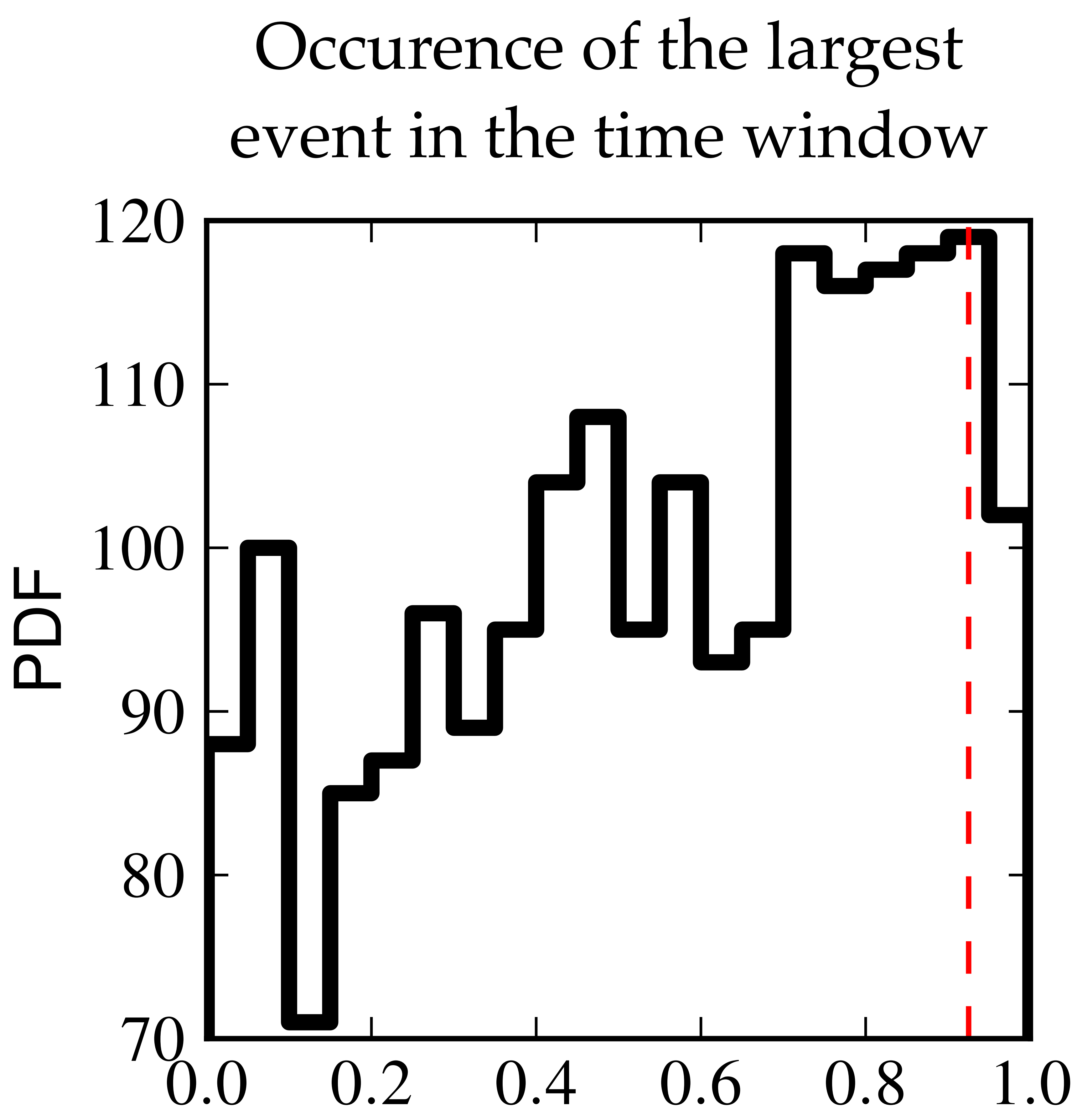
Statistical max in the time window

2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (standard model)



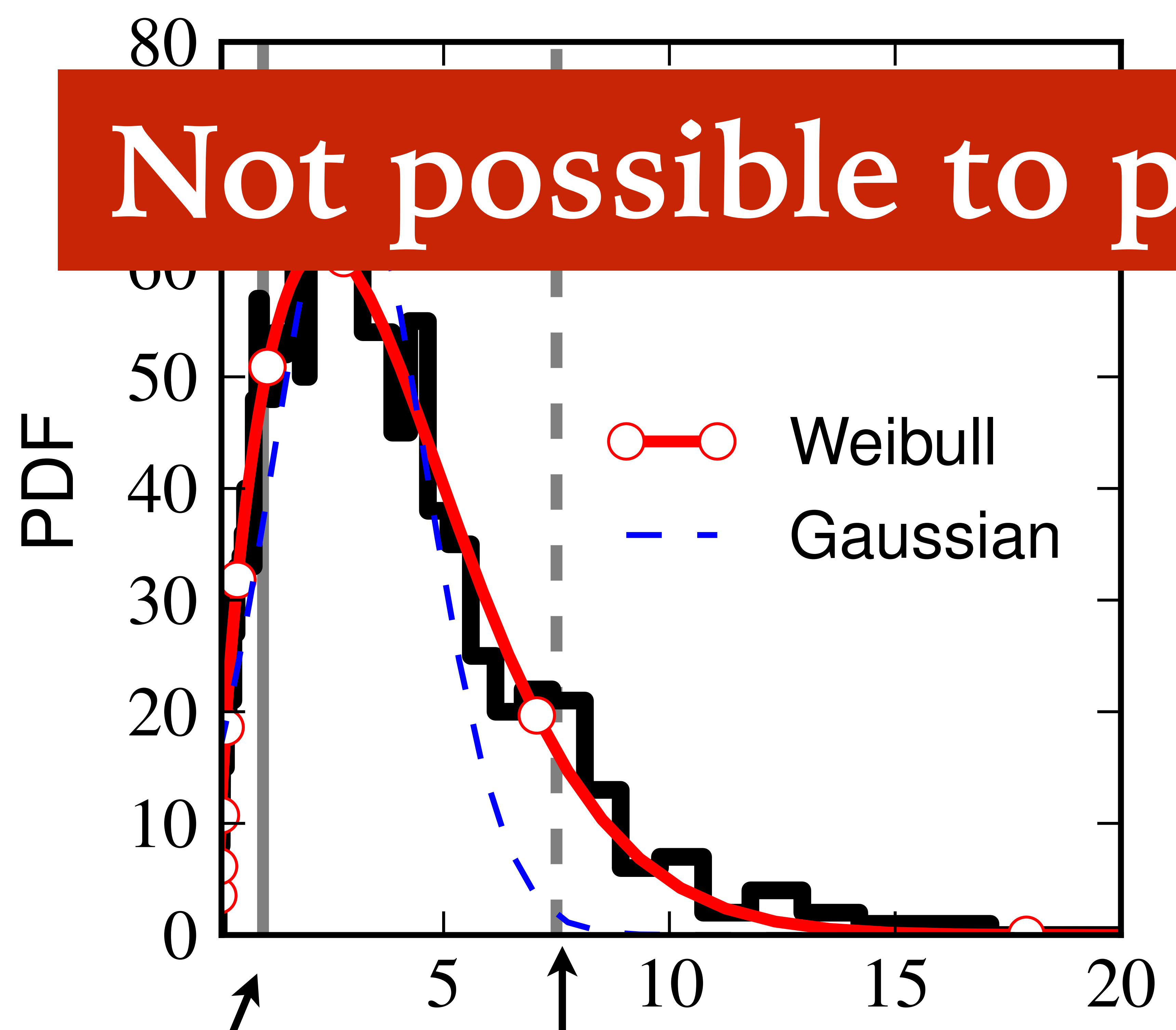


2000 stochastic realisations

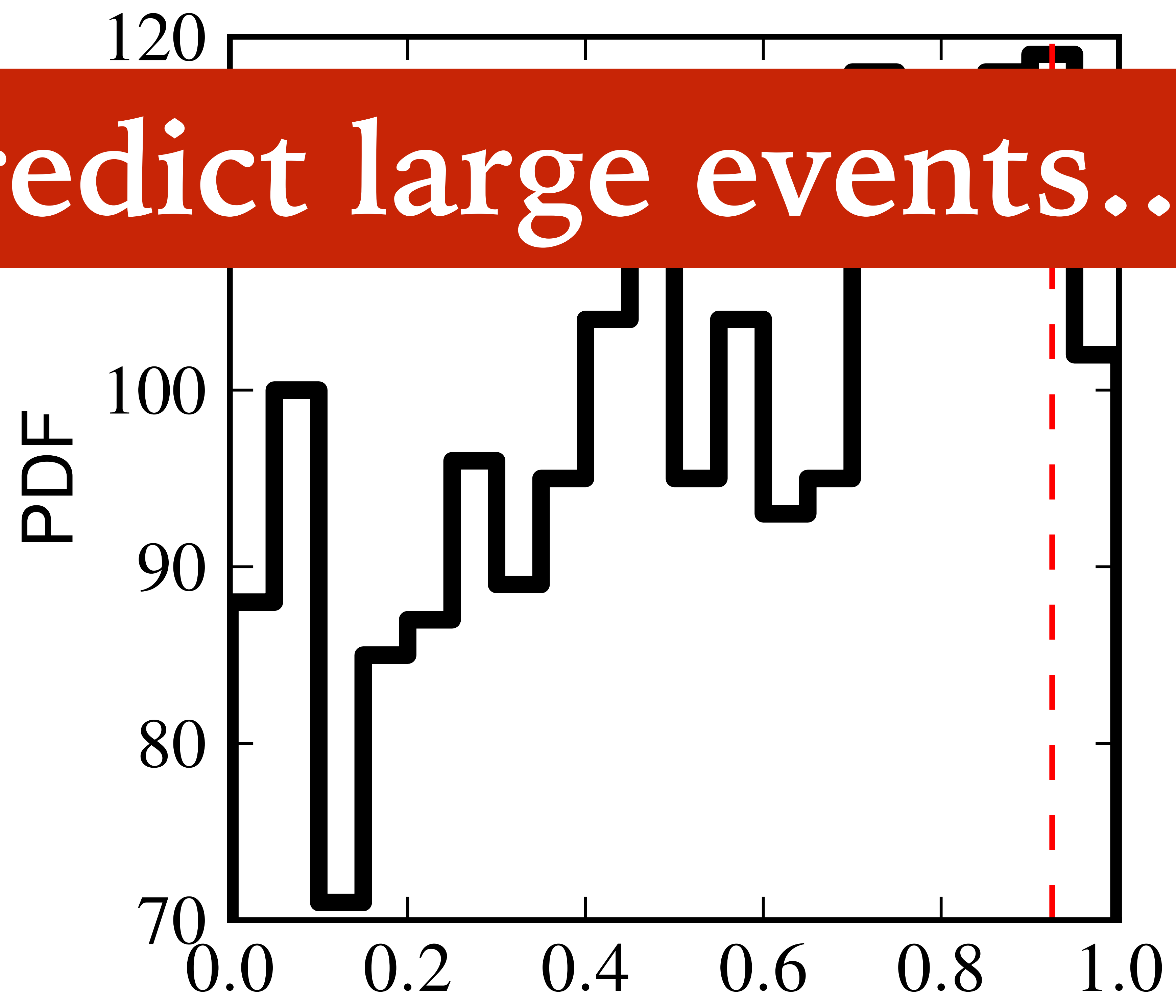
[Strugarek & Charbonneau 2014]

Robustness of one large event (standard model)

Energy of the largest event in the time window



Occurence of the largest event in the time window



Not possible to predict large events…

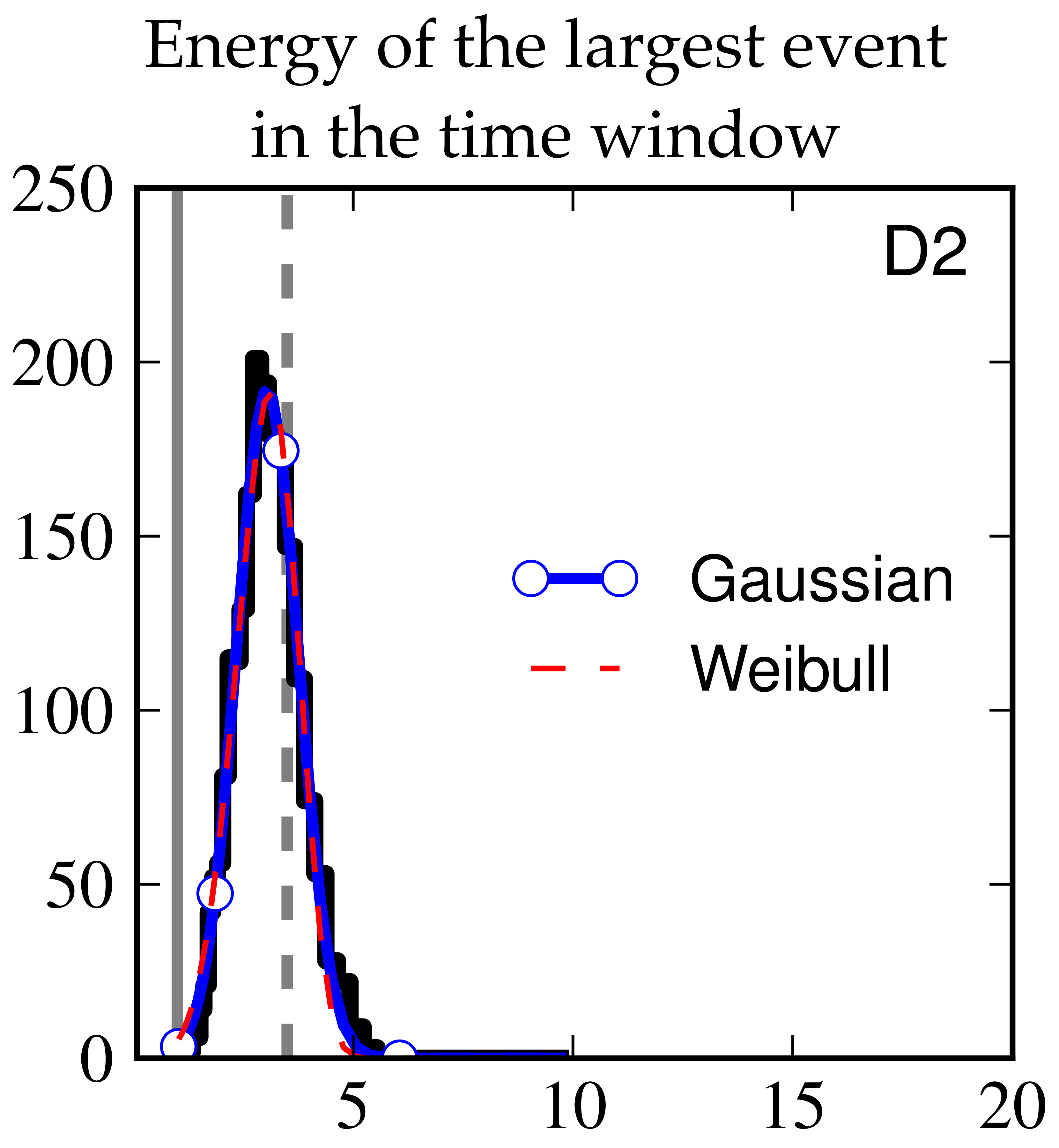
Targeted event energy

Statistical max in the time window

2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (D model)

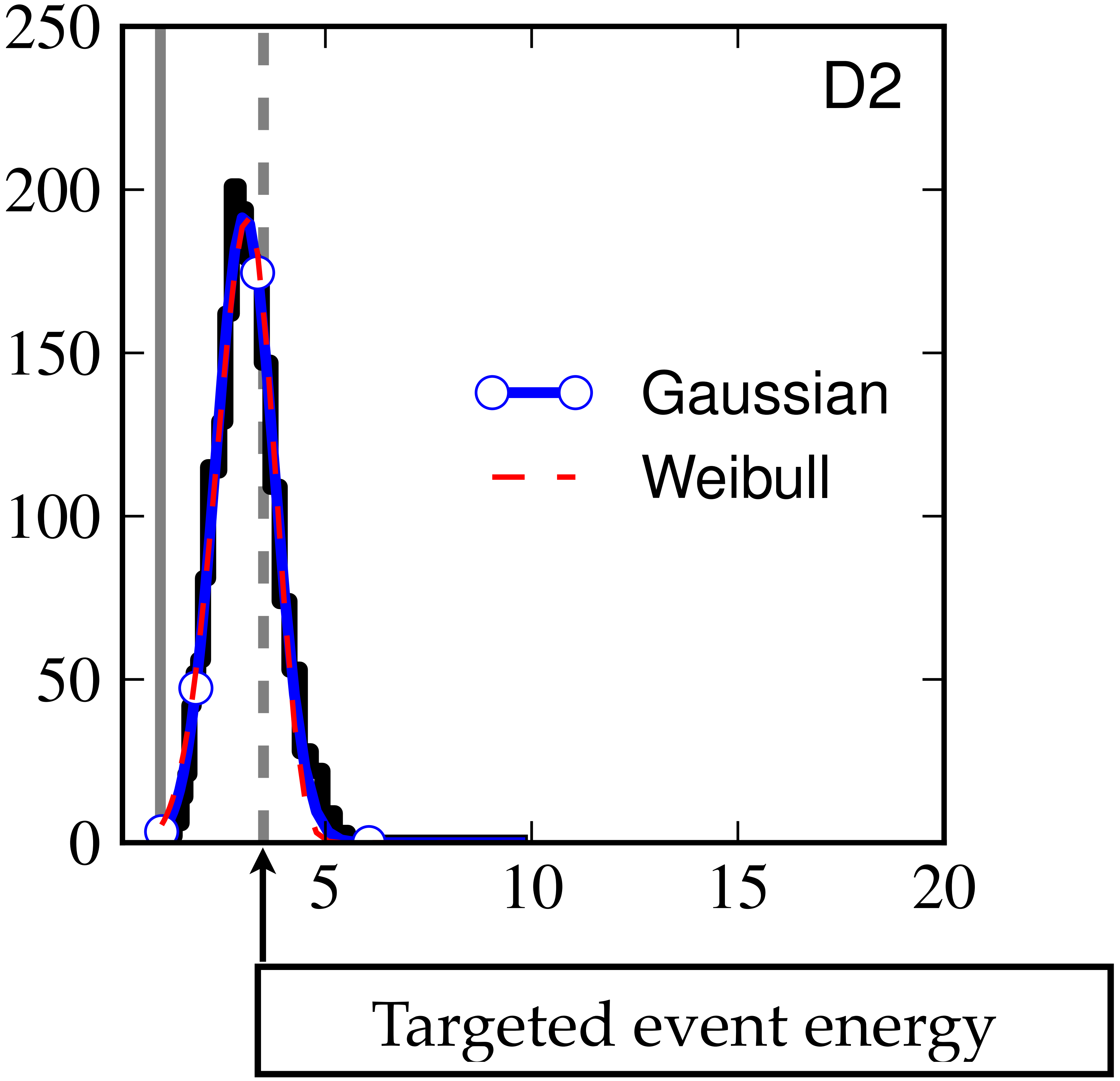


2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (D model)

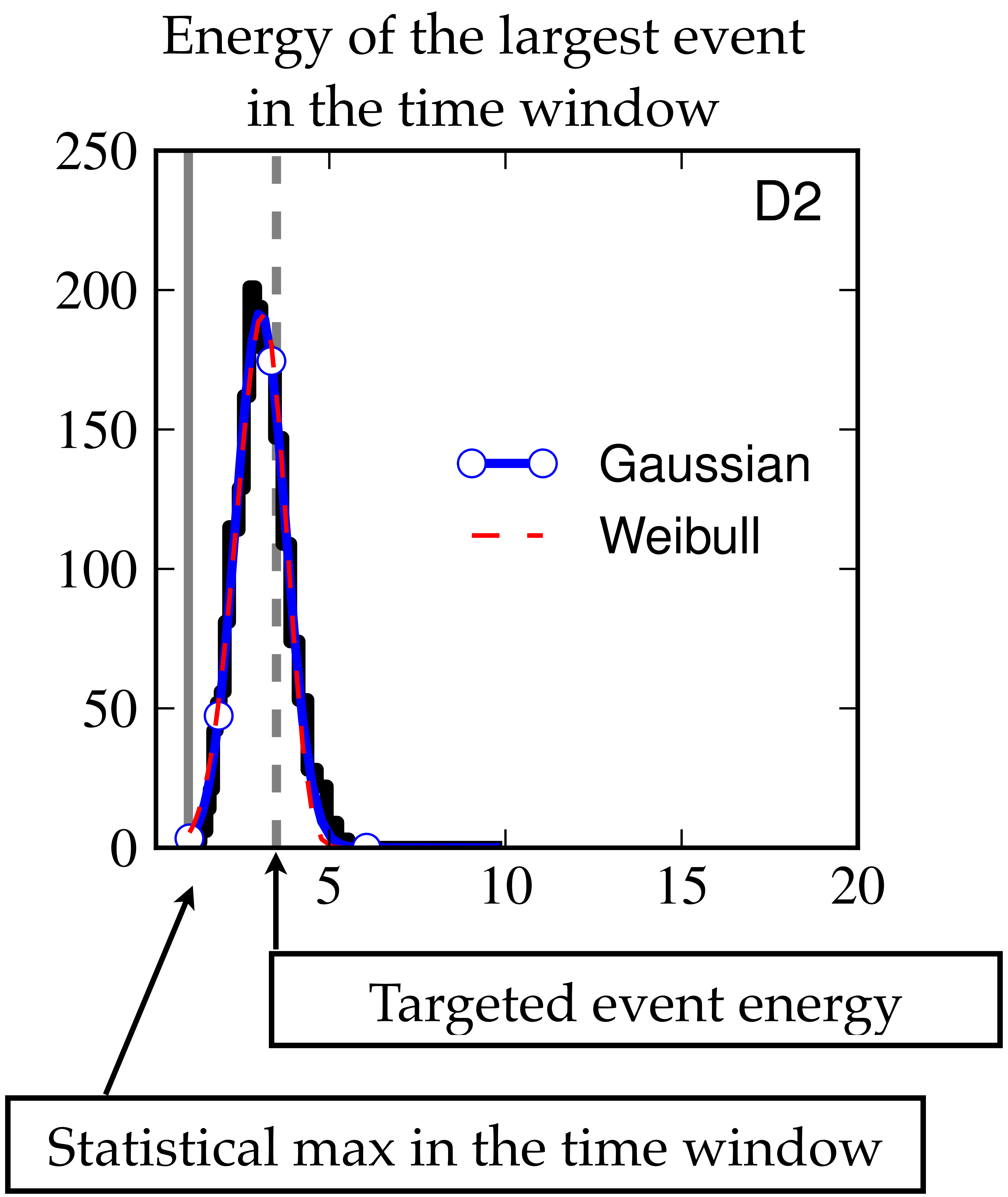
Energy of the largest event
in the time window



2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (D model)

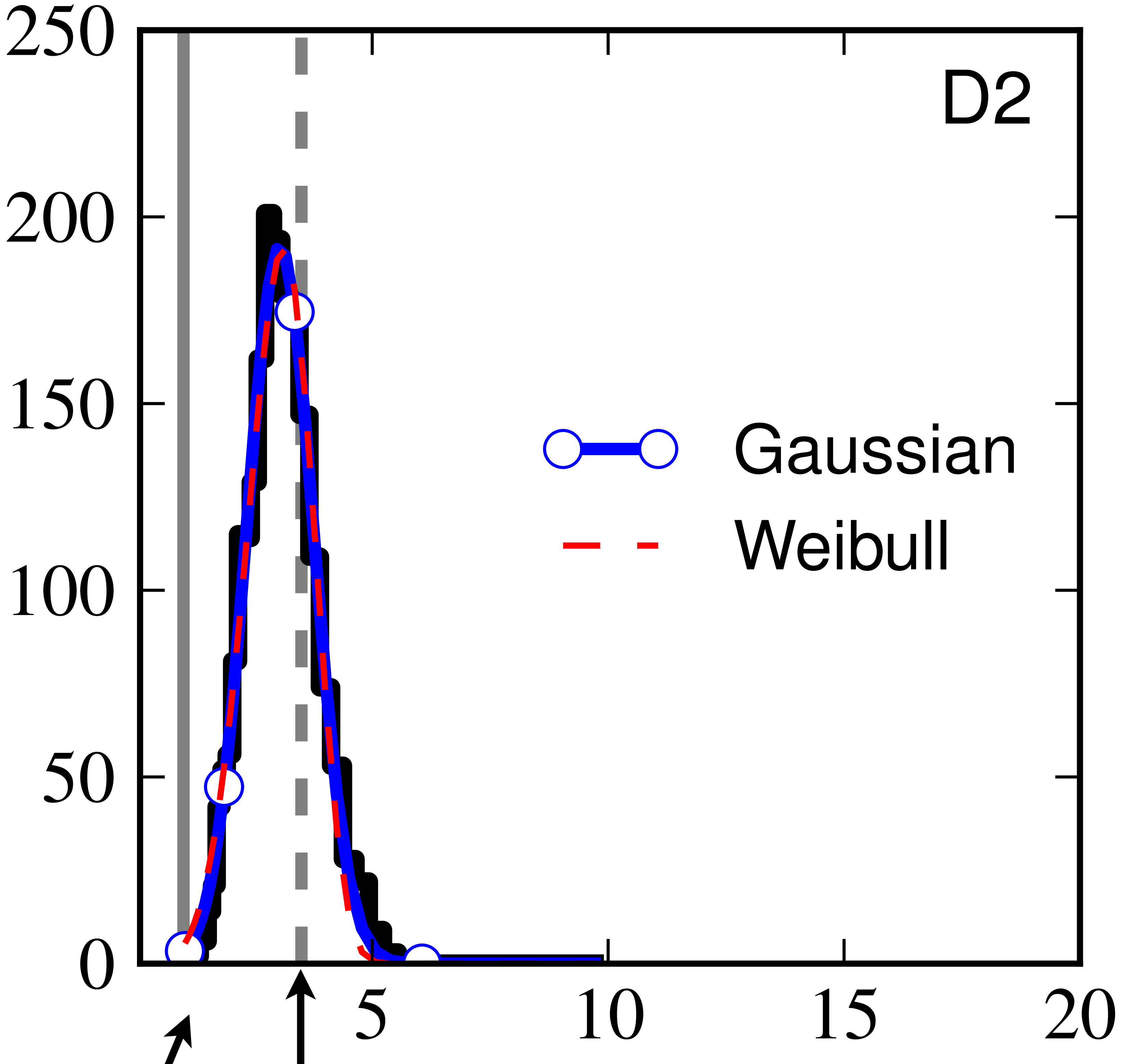


2000 stochastic realisations

[Strugarek & Charbonneau 2014]

Robustness of one large event (D model)

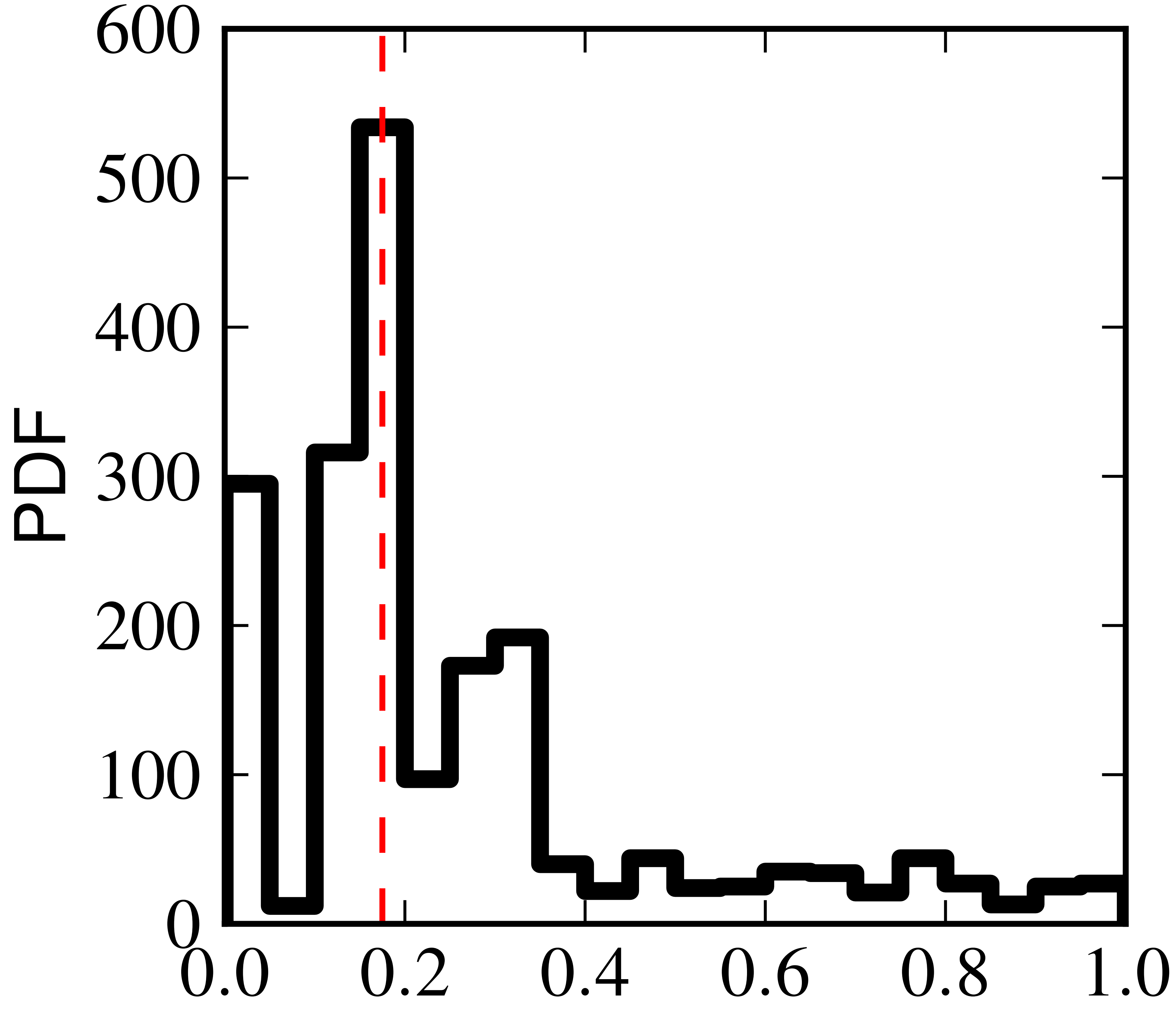
Energy of the largest event in the time window



Targeted event energy

Statistical max in the time window

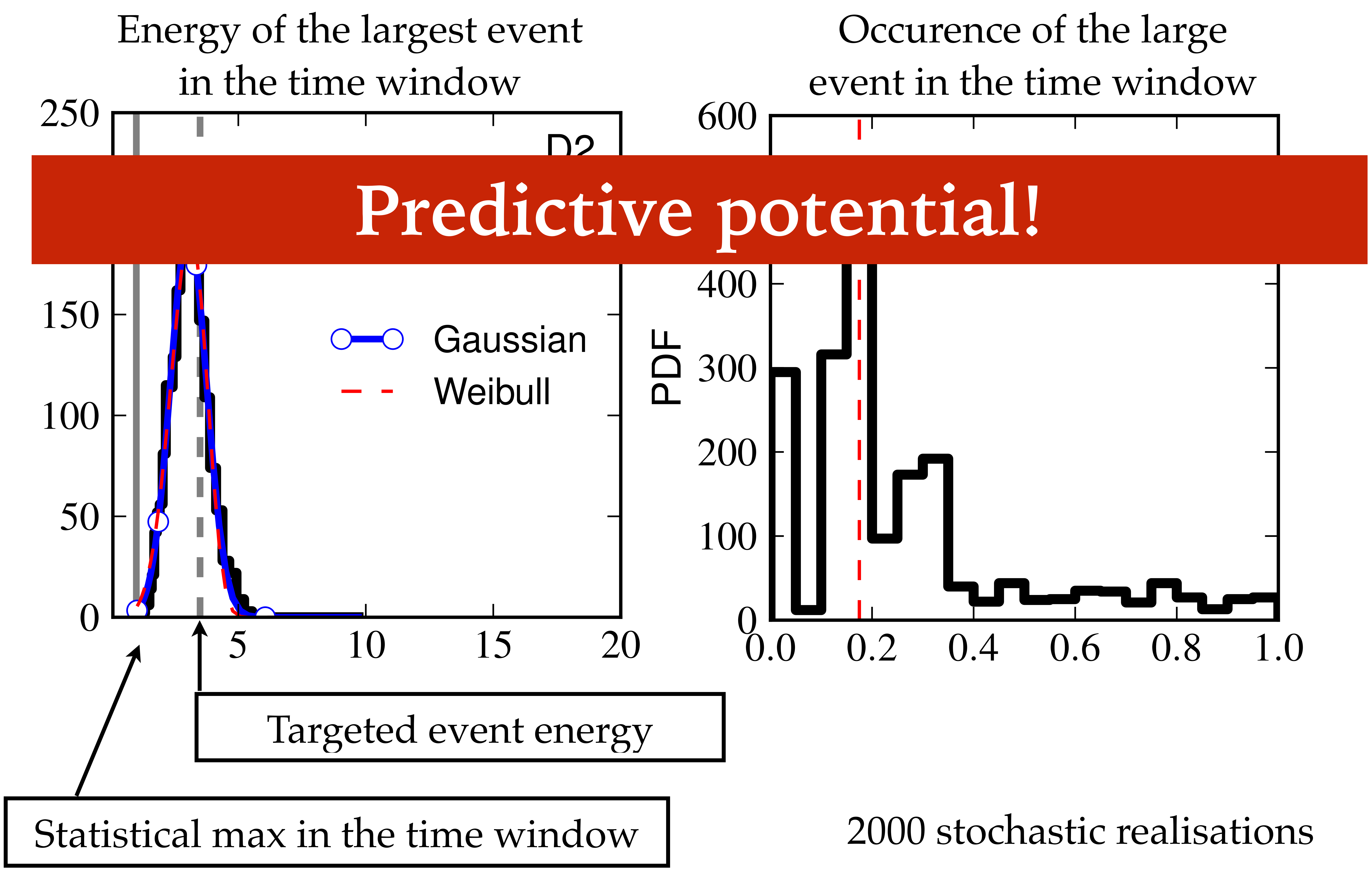
Occurence of the large event in the time window



2000 stochastic realisations

[Strugarek & Charbonneau 2014]

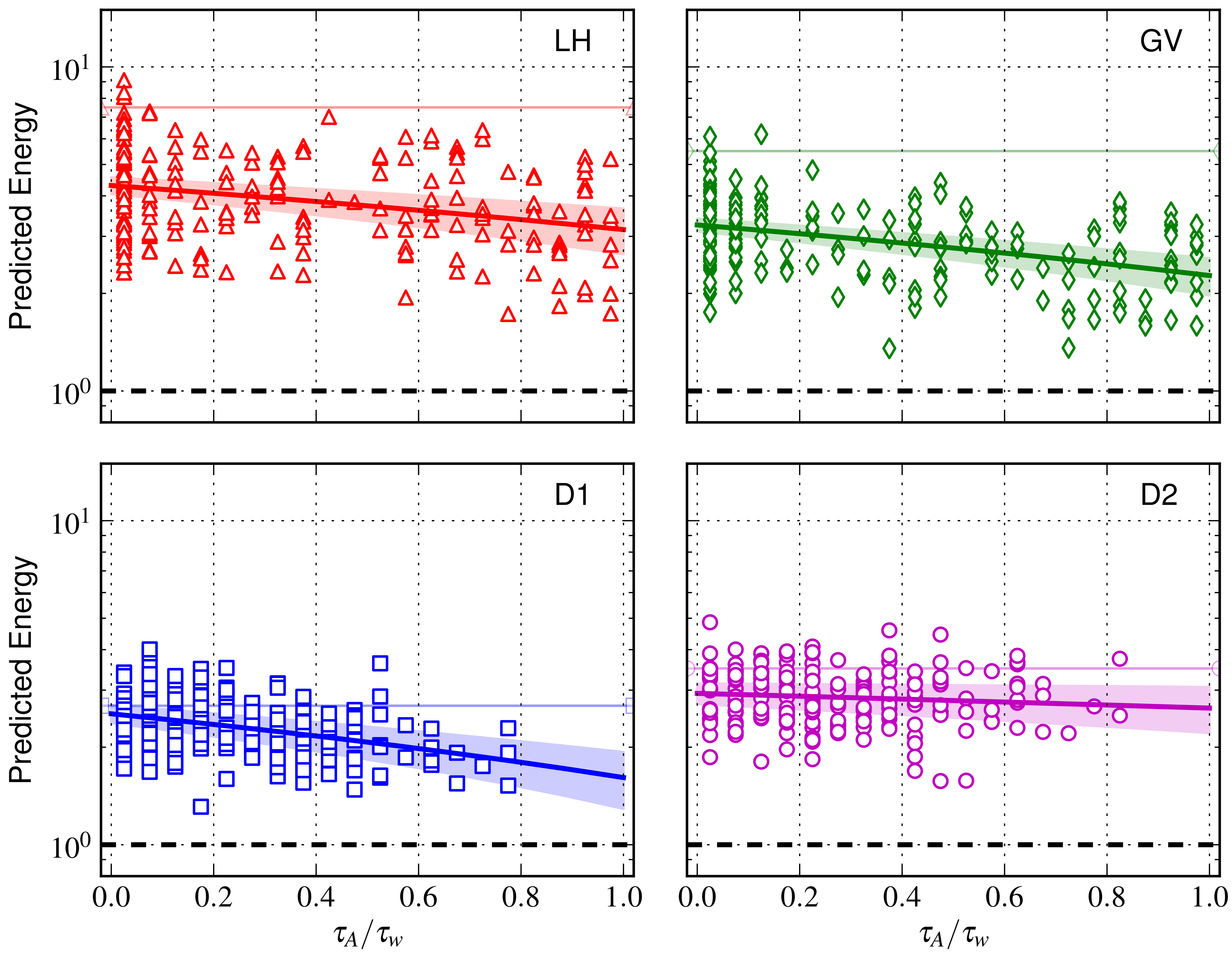
Robustness of one large event (D model)



2000 stochastic realisations

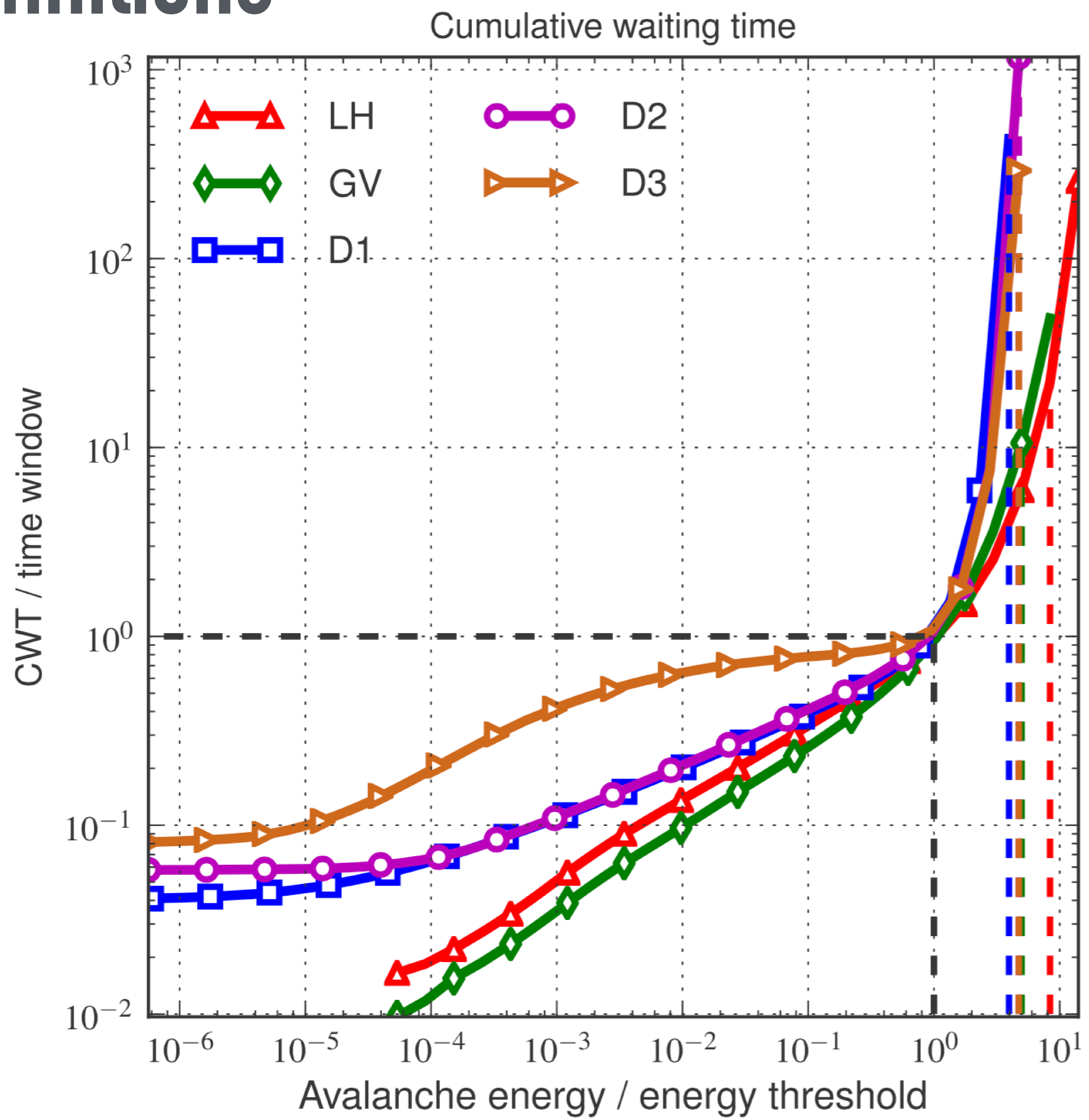
[Strugarek & Charbonneau 2014]

Bias with occurence in the time window



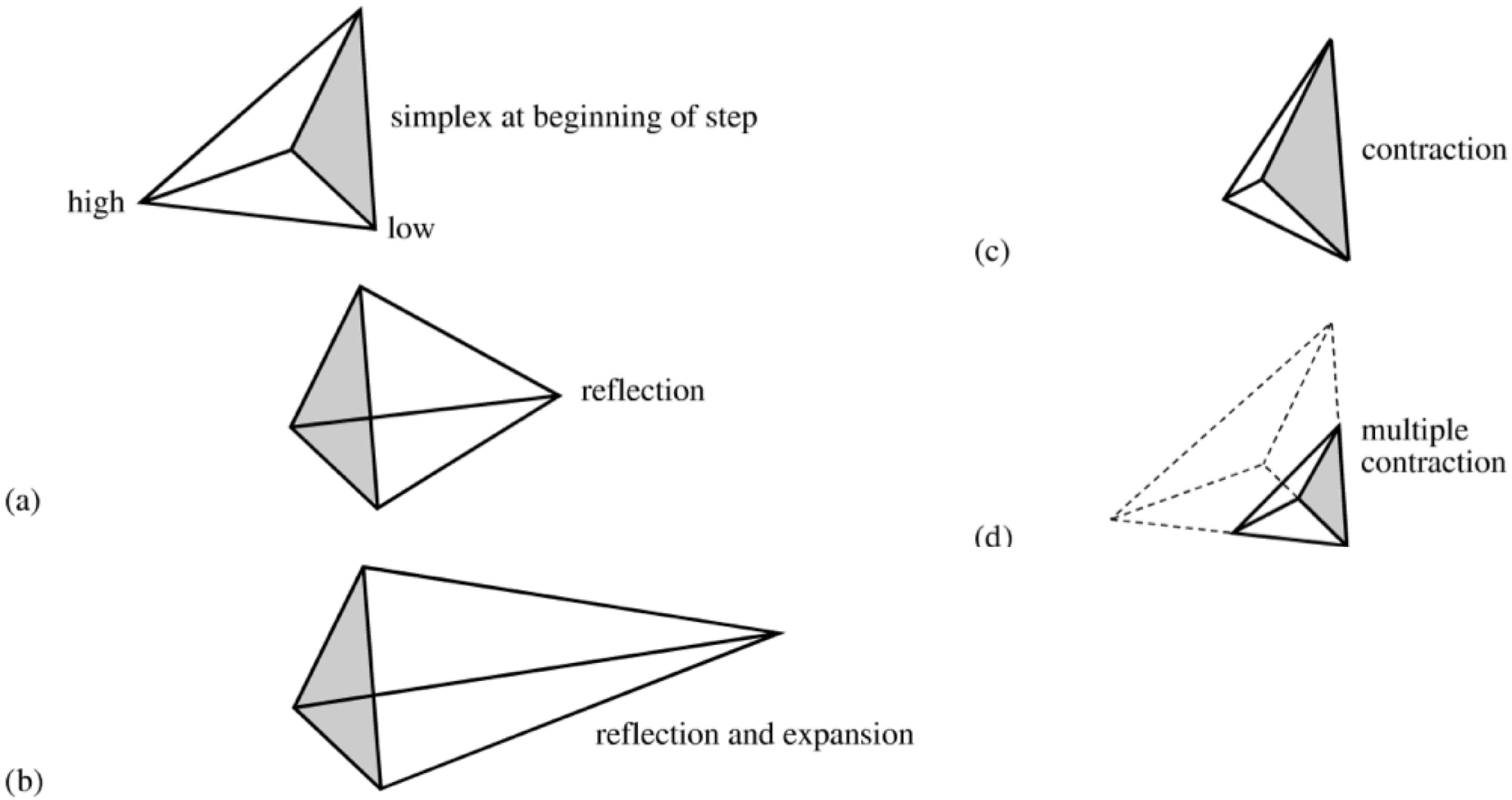
[Strugarek & Charbonneau 2014]

Time window definitions



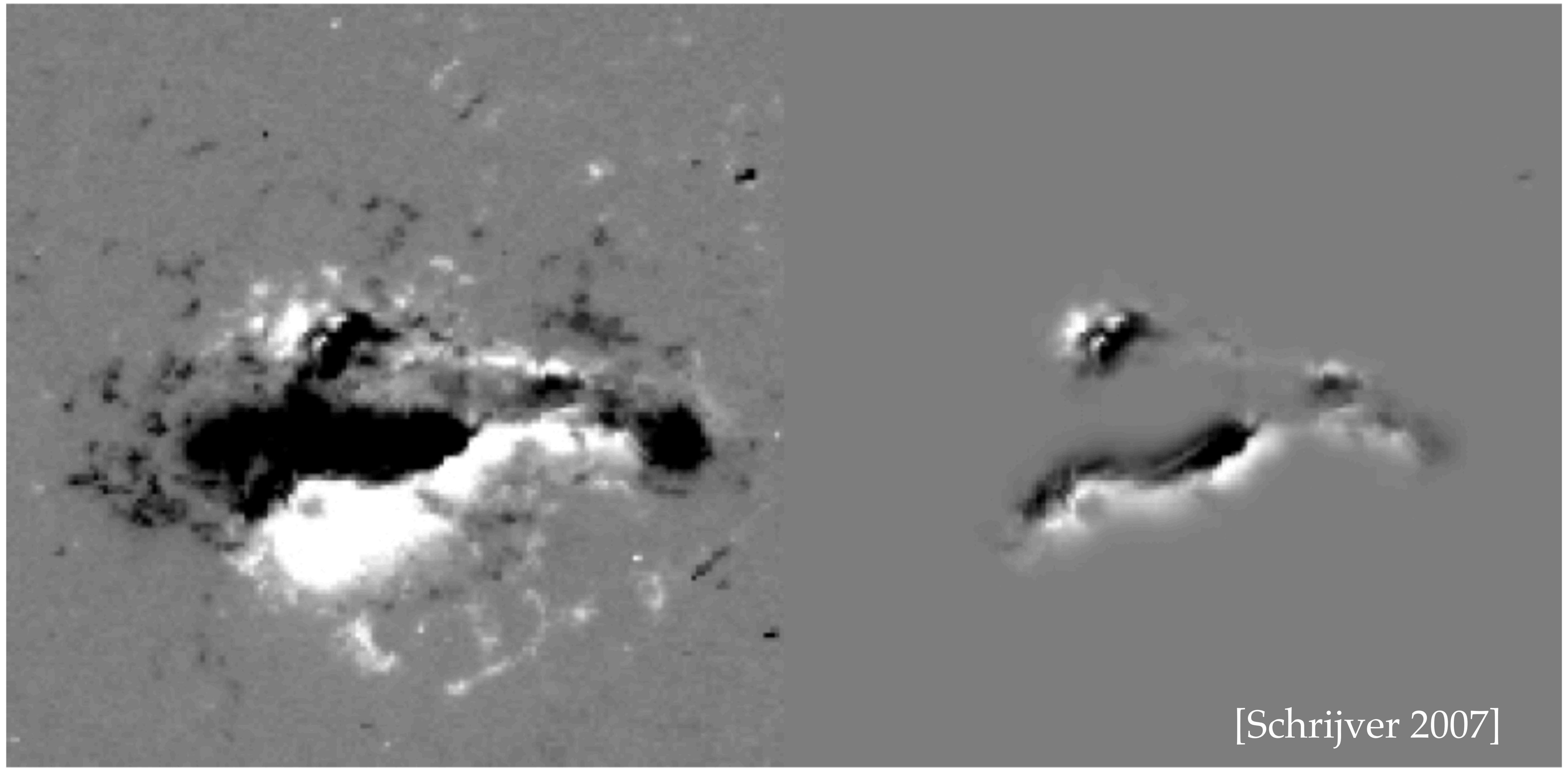
[Strugarek & Charbonneau 2014]

Simulated annealing with the simplex method



The largest flares are extremely hard to predict

Parameter	Success Rate
Climatology	0.908
Φ_{tot}	0.922
E_e	0.916
R	0.922
B_{eff}	0.913

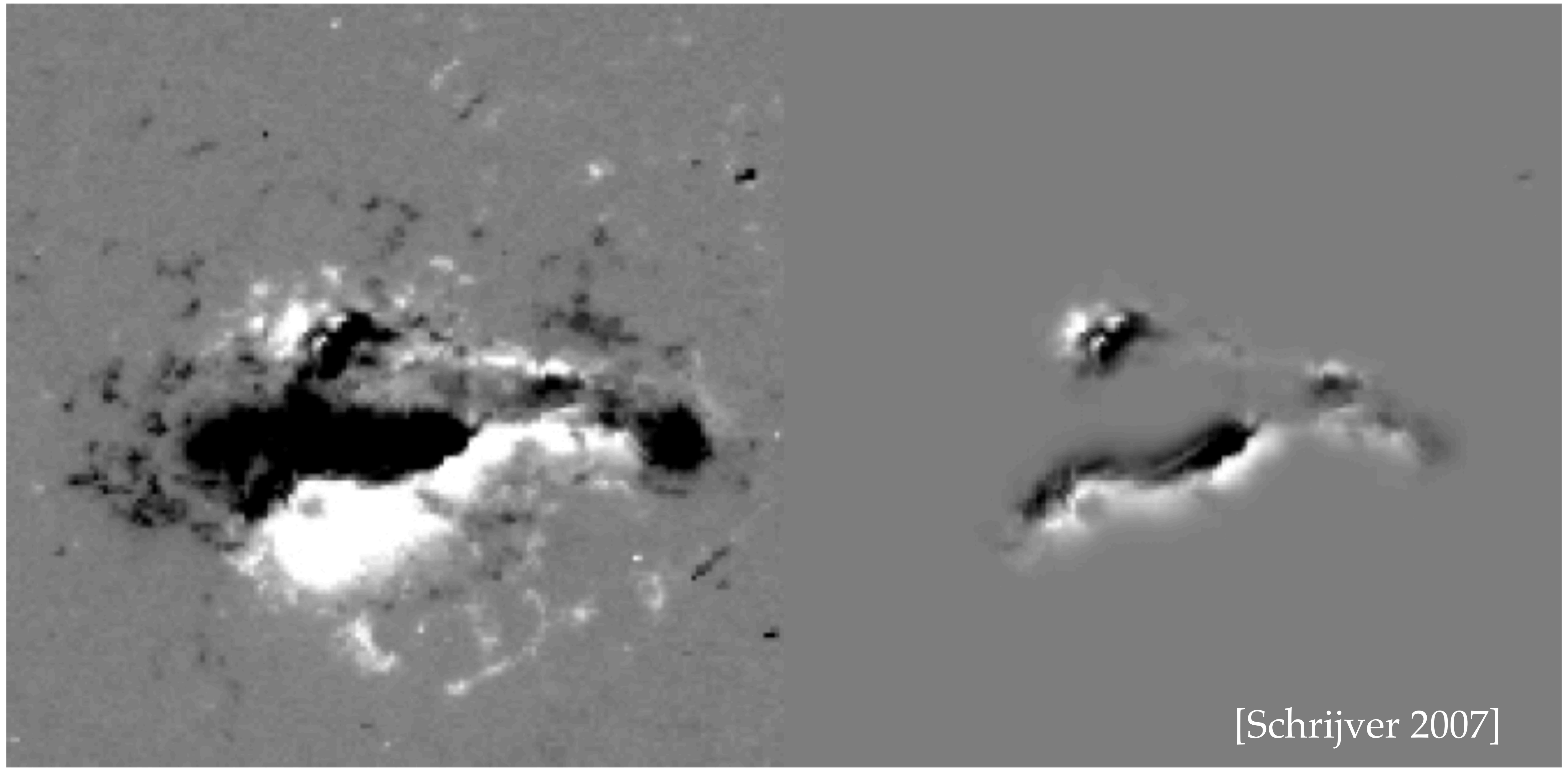


[Barnes & Leka, ApJ 2008]

THE LARGEST FLARES ARE EXTREMELY HARD TO PREDICT

Parameter	Success Rate	Heidke Skill Score	Climatological Skill Score
Climatology	0.908	0.000	0.000
Φ_{tot}	0.922	0.153	0.197
E_e	0.916	0.081	0.231
R	0.922	0.144	0.242
B_{eff}	0.913	0.072	0.220

Improvement factor compared to an assumed statistics of events



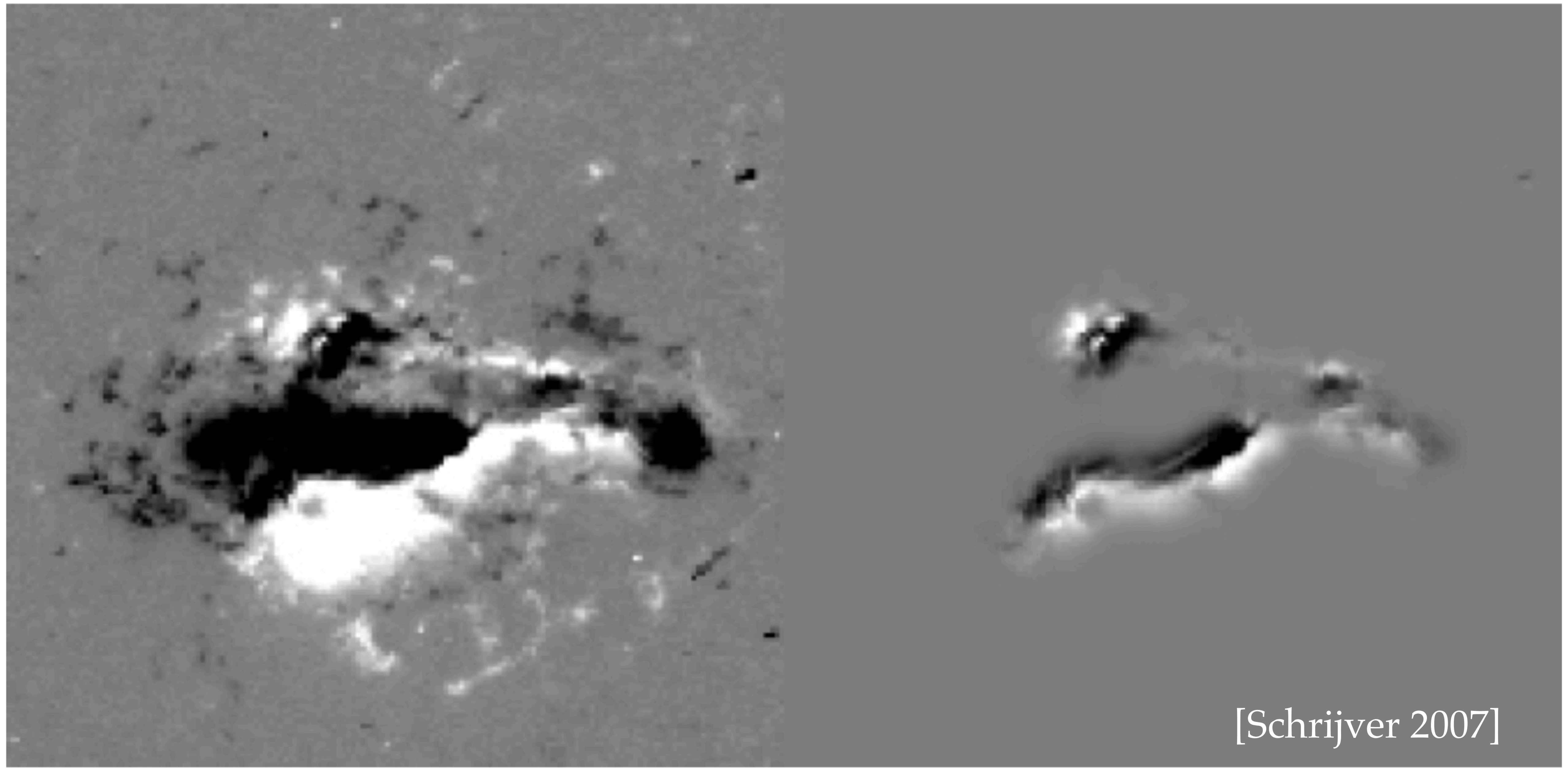
All the estimates perform quite poorly for the large events

[Barnes & Leka, ApJ 2008]

THE LARGEST FLARES ARE EXTREMELY HARD TO PREDICT

Parameter	Success Rate	Heidke Skill Score	Climatological Skill Score
Climatology	0.908	0.000	0.000
Φ_{tot}	0.922	0.153	0.197
E_e	0.916	0.081	0.231
R	0.922	0.144	0.242
B_{eff}	0.913	0.072	0.220

Improvement factor compared to an assumed statistics of events



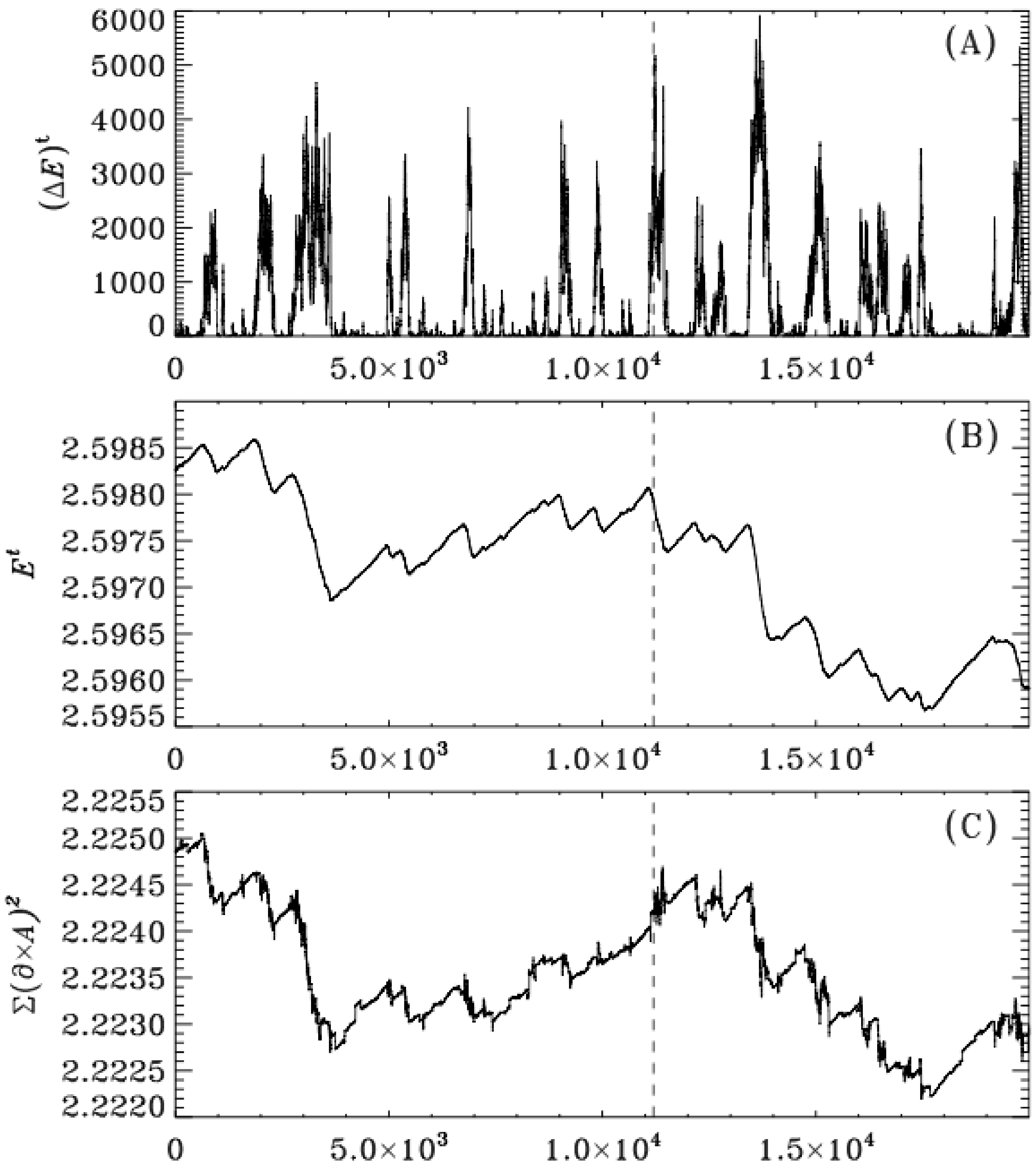
All the estimates perform quite poorly for the large events

Estimates using the statistical distribution of flares [Wheatland 2005] give *slightly* better results

[Barnes & Leka, ApJ 2008]

Physical interpretation of the nodal variable

[Charbonneau 2013]



STOCHASTICITY IN THE D MODELS

★ Where to put the random process?

✤ Random **extraction**

$$(Z_{i,j} > Z_c) \rightarrow \begin{cases} B_{i,j} & -= \quad 4\delta B_r \\ B_{i\pm1,j\pm1} & += \quad \delta B_r \end{cases}$$

✤ Random **threshold**

$$(Z_{i,j} > Z_c^r) \rightarrow \begin{cases} B_{i,j} & -= \quad 4\delta B \\ B_{i\pm1,j\pm1} & += \quad \delta B \end{cases}$$

✤ Random **redistribution**

$$(Z_{i,j} > Z_c) \rightarrow \begin{cases} B_{i,j} & -= \quad 4\delta B \\ B_{i\pm1,j\pm1} & += \quad \frac{r_k}{R}\delta B \end{cases}$$

$$r_k \text{ random deviate } \in [0,1] \; (k \in \{1,4\}) \qquad \sum_k r_k = R$$