Two Projects in Machine Learning for Climate Dynamics

Journées Calcul et Apprentissage - April 2019

I) Machine Learning and rare event algorithms

Commitors are often optimal score functions for rare event algorithms

- A good choice of a selection rule (the score function) is crucial for the effectiveness of rare event algorithms.

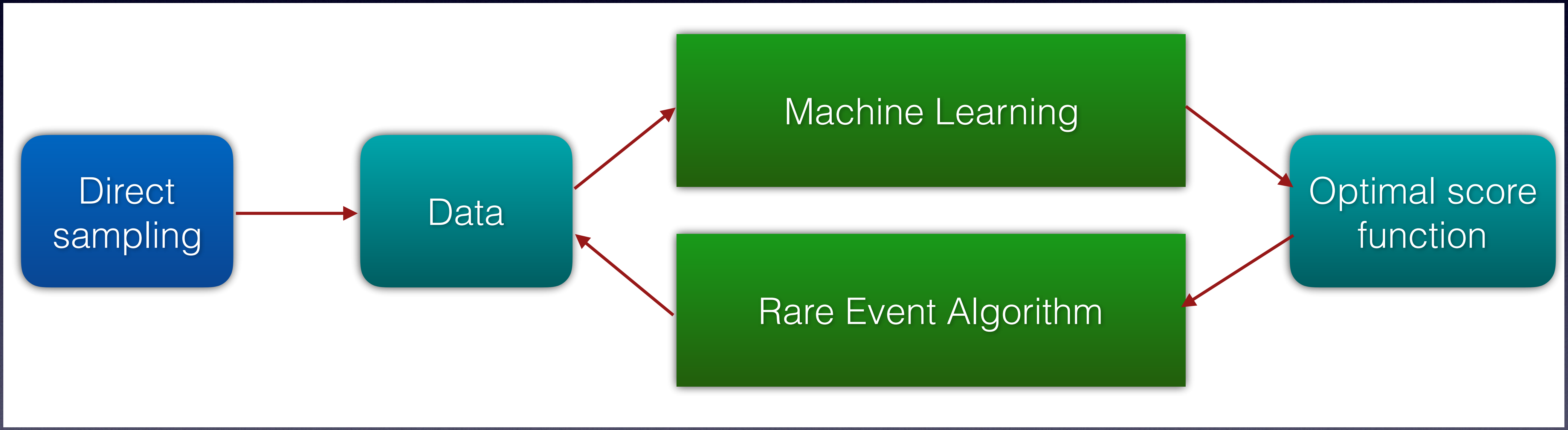
- The commitor q(x), for two sets A and B, is the probability that a trajectory starting at the point x reaches the set B before the set A:

$$q(x) = \mathbb{P}(t_B(x) < t_A(x))$$

$$\text{where} \quad t_C(x) = \inf\{t : X(t) \in C \,|\, X(0) = x\}$$

- The commitor is the optimal score function.

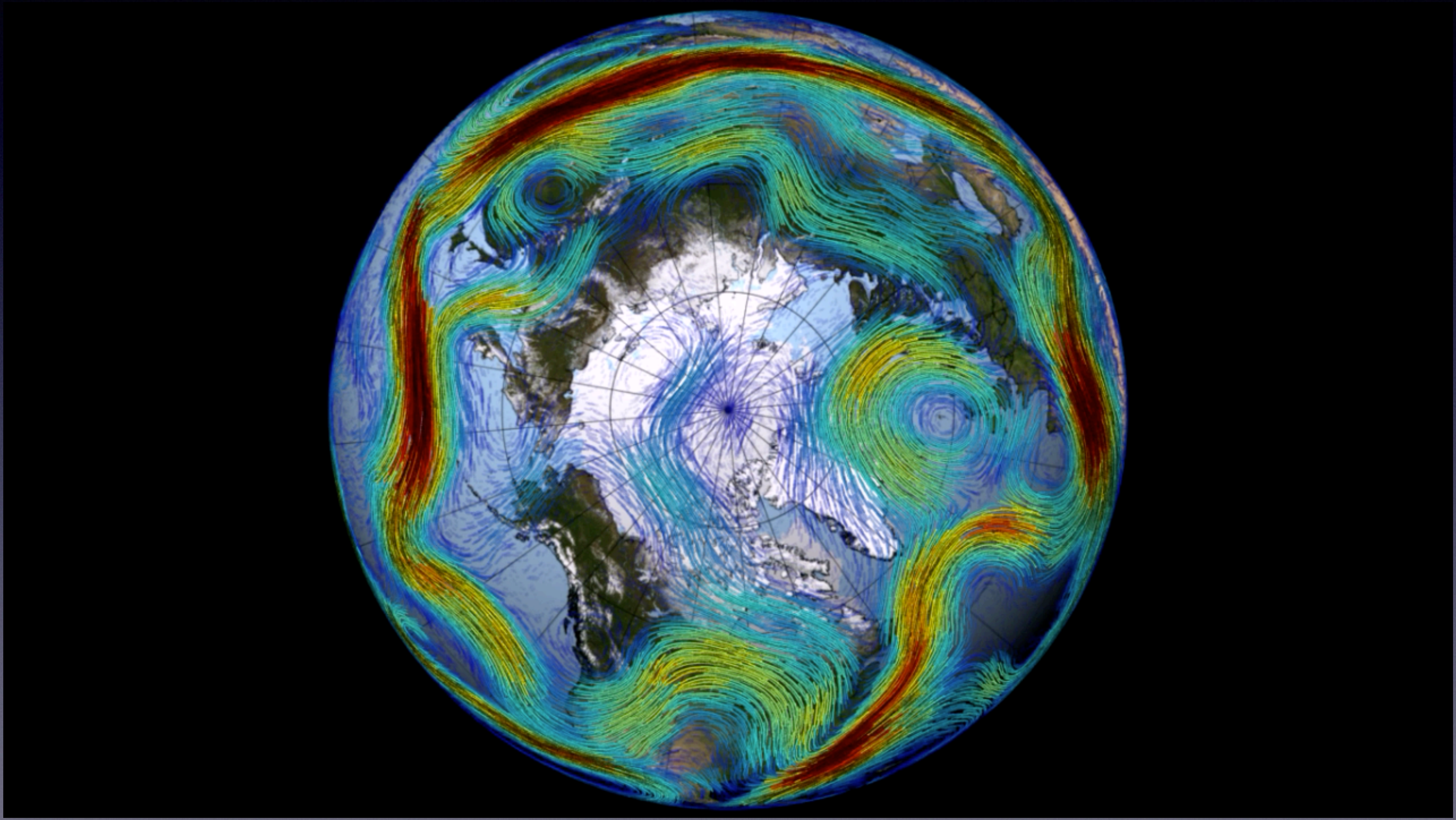
Machine learning of commitor functions and feedback loop control for rare event algorithms



Bouchet, Jack, Lecomte, Nemoto, PRE, 2016

II) Machine learning and dynamical model reduction

The jet stream

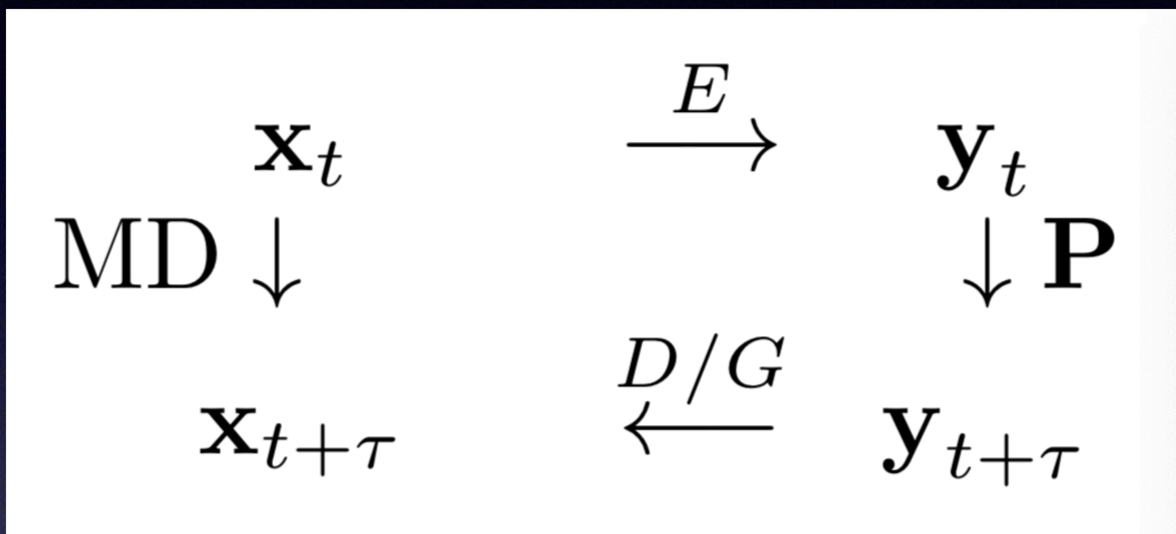


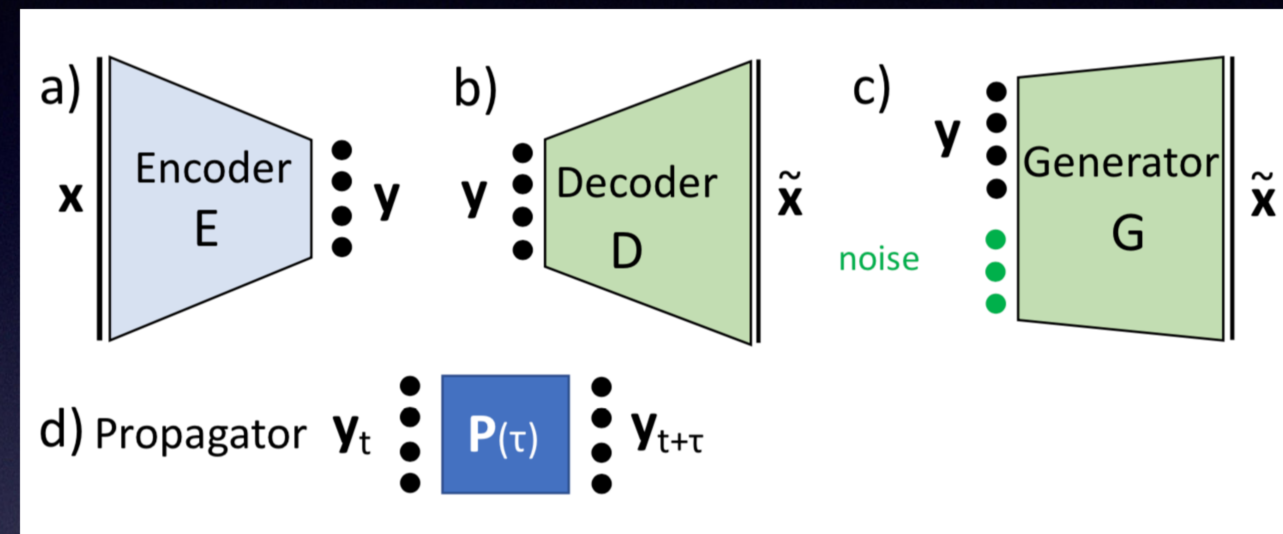
Markov model reduction

$$
\begin{array}{ccc}
\mathbf{x}_t & \xrightarrow{E} & \mathbf{y}_t \\
\mathrm{MD}\downarrow & & \downarrow\mathbf{P} \\
\mathbf{x}_{t+\tau} & \xleftarrow{D/G} & \mathbf{y}_{t+\tau}
\end{array}
$$

- x_t is the state variable of the initial Markov process, MD=P* its temporal evolution.
- $y_t = E(x_t)$ is the reduced variable. P its dynamics (a linear operator).
- $D/G(x, y_t) = \mathbb{P}(x_t = x \,|\, y_t)$ a decoder.
- We want to learn $E, P,$ and possibly D/G

Machine learning for model reduction





What criteria? Which loss function?

Mardt, Pasquali, Wu and Noé, Nature Communication, 2018

Generator of a Markov process and singular value decomposition

- Singular value decomposition of the transfer operator

$$P^*\left(x_{t+\tau}|x_t\right) = \sum_{n=1}^{\infty} \sigma_n^* \psi_n^*\left(x_{t+\tau}\right)\phi_n^*\left(x_t\right)$$

- We look for a finite dimensional approximation of the transfer operator (or of the infinitesimal generator)

$$P^*\left(x_{t+\tau}|x_t\right) = \sum_{n=1}^{N} \sigma_n^* \psi_n^*\left(x_{t+\tau}\right)\phi_n^*\left(x_t\right)$$

Finite dimensional approximation of the Markov chain

$$P^*\left(x_{t+\tau} | x_t\right) = \sum_{n=1}^{\infty} \sigma_n^* \psi_n^* \left(x_{t+\tau}\right) \phi_n^* \left(x_t\right)$$

- We consider approximation of the right singular-vectors $\{\phi_n\}_{1 \le n \le N}$ in the span of $y = E(x)$.
- The approximated evolution is computed from the data.
- The projected transfer operator is

$$P\left(y_{t+\tau} | y_t\right) = \sum_{n=1}^{N} \sigma_n u_n v_n^T,$$

where $\{u_n\}, \{v_n\},$ and σ_n are computed such that this is the SDV of the projected operator.

Optimal truncation

For any $y = E(x)$, we have:

$$\text{For } 1 \leq n \leq N: \; 0 \leq \sigma_n < \sigma_n^*$$

And

$$\text{For } 1 \leq n \leq N: \sigma_n = \sigma_n^* \iff E_n = \phi_n^*$$

- This is the optimal truncation in the sense that it is a perfect N dimensional approximation of the SVD of the Markov chain.

Optimal truncation and loss function

We have:

$$\text{For } 1 \leq \text{n} \leq \text{N}: \;\; 0 \leq \sigma_\text{n} < \sigma_\text{n}^*$$

And

$$\text{For } 1 \leq \text{n} \leq \text{N}: \sigma_\text{n} = \sigma_\text{n}^* \iff \text{E}_\text{n} = \phi_\text{n}^*$$

- Then a good loss function could be any of the functions

$$L_k = -\sum_{n=1}^{N} (\sigma_n)^k$$

Conclusion

- For complex dynamical systems and Markov processes, several fundamental problems should be approached from a machine learning perspective (model reduction, parameterisation, prediction problems, commitor function sampling, etc)
- Machine learning should help us tackle many fundamental problems in climate dynamics or astronomy (extreme event probability, effective dynamics, turbulence parameterisation, etc)