Supporting Information for "Rare event algorithm study of extreme warm summers and heatwaves over Europe"

Francesco Ragone[1]and Freddy Bouchet[1]

[1]Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France

Introduction In these Supporting Informations we present a more detailed description of the theory and the implementation of the algorithm, a more detailed description of the climate model we use and of the experiments we have performed, and some comments on data analysis. We also show additional figures for the definition of the areas we apply the rare event algorithm to, and for the importance sampling for subseasonal fluctuations.

Corresponding author: Francesco Ragone, Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France (francesco.ragone@uclouvain.be)

Description of the algorithm Rare event algorithms have been devised long ago (Kahn & Harris, 1951; Del Moral, 2004) in order to lower the numerical cost of rare events. In a dynamical context, they have been first applied for complex and bio-molecules, see for instance (Metzner et al., 2009; Hartmann et al., 2013). More recently, in order to progressively go towards genuine geophysical applications, they have been applied to Lorenz models (Wouters & Bouchet, 2016), partial differential equations (Rolland et al., 2016), turbulence problems (Grafke et al., 2015; Laurie & Bouchet, 2015; Ebener et al., 2019; Bouchet et al., 2019; Lestang et al., 2018, 2020), geophysical fluid dynamics (Bouchet et al., 2019), and climate applications (Ragone et al., 2018; Webber et al., 2019; Ragone & Bouchet, 2020; Plotkin et al., 2019). Some approaches through minimum action methods, related to large deviation theory, are reviewed in (Grafke & Vanden-Eijnden, 2019).

Based on the phenomenology of the dynamics for each family of extreme events, an appropriated type of rare event algorithm should be carefully chosen. In this paper, in order to study long-lasting events, we use an algorithm of the family of genealogical algorithms, very close in spirit to the initial one (Kahn & Harris, 1951), and the one described by (Del Moral et al., 2005) and mathematically comprehensively studied in (Del Moral, 2004). The idea is to perform an ensemble simulation, stop regularly every resampling time, and based on a score function select some trajectories to clone them and discard some others. The ensemble statistics is then tilted towards the events of interest.

Choosing well the score function in order to define the selection criteria is critical. For this algorithm, we use a score function which is the time integral of an observable, as for instance in (Giardina et al., 2011). In (Giardina et al., 2011), the algorithm was specifically designed for studying long time large deviations. This is suited for long-lasting events. In this letter, we use a setup similar to the one in (Giardina et al., 2011), with the difference that our simulations, like in (Del Moral et al., 2005) are transient ones (they depend from the initial state and extend over finite time) rather than statistically stationary. We call this algorithm the Del-Moral Garnier, or Giardina-Kurchan-Tailleur-Lecomte algorithm, or simply the rare event algorithm.

We have first proposed this rare event algorithm to study heatwaves and generally long-lasting extreme events in (Ragone et al., 2018). Briefly, the algorithm consists in running an ensemble simulation of N ensemble members or trajectories with a climate model starting from different initial conditions. Let us indicate with $\vec{X}_n(t)$ the state vector of trajectory n at time t, and with $A(t) = A(\{\vec{X}(t)\})$ an observable of interest function of the state of the system, of whose time average we want to study the extremes (e.g. the surface temperature over a region). At constant intervals of a fixed resampling time τ, each trajectory is assigned a weight, which determines if that trajectory is killed or if it continues its evolution, possibly spawning copies of itself. The weight w_i^n of trajectory n at time $t_i = i\tau$ is computed as

$$w_i^n = \frac{e^{k\int_{t_{i-1}}^{t_i} A(\{\vec{X}_n(t)\})\,\mathrm{d}t}}{\frac{1}{N}\sum_{n=1}^{N} e^{k\int_{t_{i-1}}^{t_i} A(\{\vec{X}_n(t)\})\,\mathrm{d}t}}, \tag{1}$$

where k is a parameter that control how stringent is the selection. At times t_i the simulation is stopped, the weights are computed, and each trajectory generates a number of copies of itself proportional to the weight assigned to it. A small random perturbation is added to each copied trajectory, so that they evolve differently from the original trajectory and from each other. Trajectories with weights smaller than 1 are killed, and do not continue their evolution. Killed trajectories are substituted in the ensemble by the copies of the surviving trajectories. In practice this resampling is done in such a way that the the number of trajectories in the ensemble always remains constant and equal to N. See (Ragone et al., 2018; Ragone & Bouchet, 2020) for details about the technical implementation.

The weights are defined in such a way that trajectories featuring large values of the time average of the observable are the ones that will populate more densely the ensemble. After a total running time T_a, the probability $\mathbb{P}_k(\{\vec{X}(t)\})$ of observing a trajectory in the ensemble generated by the algorithm is related to the probability $\mathbb{P}_0(\{\vec{X}(t)\})$ of observing the same trajectory in a normal ensemble simulation with no resampling as

$$\mathbb{P}_k(\{\vec{X}(t)\}) = \frac{e^{k \int_{t_a}^{t_a+T_a} A(u)\,\mathrm{d}u}}{Z} \mathbb{P}_0(\{\vec{X}(t)\}), \tag{2}$$

where Z is a normalization term, and the equation is valid in the limit of large N with relative errors on the computation of expectation values of the order of $1/\sqrt{N}$ (Del Moral, 2004). In ensemble simulations performed with the rare event algorithm, trajectories featuring a large value of the time average of the control observable over the entire simulation period are thus much more likely to be observed than in a normal simulation, and viceversa. The value of k determines how strong is the tilt of the probability distribution. This is called *importance sampling*. More details on the theory behind the method and the reconstruction of the effective ensemble of trajectories, as well as on the choice of N, k and τ, can be found in (Ragone et al., 2018; Ragone & Bouchet, 2020).

Description of the climate model and of the control run The simulations are performed with the Community Earth System Model (CESM) version 1.2.2 (Hurrell et al., 2013). We use an atmosphere and land only setup, whose active components are the Community Atmospheric Model version 4 (CAM4) and the Community Land Model version 2 (CLM2). CAM4 is a widely used model of the Earth's atmosphere. The dynamical core is a finite volume discretisation of the Navier-Stokes equations for a planetary fluid envelope in hydrostatic approximation. The physics suite consists of parameterisations of moist precipitation processes, clouds and radiation, turbulent mixing, and surface fluxes. CLM2 is a model of the Earth's land surface. It represents several aspects of the land surface, including surface heterogeneity, and consists of components related to land biogeophysics, hydrologic cycle, biogeochemistry, human dimensions, and ecosystem dynamics. The horizontal resolution is 0.9 and 1.25 degrees in latitude and longitude respectively, and we use 26 vertical layers in hybrid pressure coordinates. The output is sampled every 3 hours. The model is run at statistically stationary state, with sea surface temperatures (SST), sea ice cover, and the concentration of atmospheric CO2 and other greenhouse gases fixed at values representative of present day climate (year 2000). Contrary to our previous proof of concept (Ragone et al., 2018), the model features daily and seasonal cycles. We have performed a control run of 1000 years, as a set of K=10 independent 100 years long simulations. The control run is used as a benchmark for the statistics, in order to evaluate the performances of the algorithm.

Description of the experiments with the rare event algorithm We perform two sets of simulations, setting D as the area over France and Scandinavia. For both cases we perform K=10 ensemble simulations, each with N=100 trajectories, biasing parameter set to k=30, and running for T_a=90 days from June 1st to August 29th. The resampling time τ in general should be of the order of magnitude of the decorrelation time of the dynamics (see the discussion in (Ragone et al., 2018; Ragone & Bouchet, 2020)). In this case the dominant dynamics relevant for heatwaves acts at synoptic timescale, thus we set τ=5 days. In order to allow copies of the same trajectory to evolve differently, we add a small perturbation to each trajectory after the resampling. The perturbation is added multiplying at each vertical level the amplitudes of the spherical harmonics of order (m, n) of the potential temperature field by $\gamma = 1 + \epsilon r_{m,n}$, where $\epsilon = 10^{-4}$ and the $r_{m,n}$ are uniform random numbers extracted between -1 and 1. The initial conditions for the 1000 trajectories are taken from the 1st of June of the 1000 years of the control run. For a choice of D each ensemble starts therefore from a set of 100 different initial conditions.

Computation of error bars In order to compute error bars for a quantity f, except otherwise stated, we perform a set of K independent experiments and compute an empirical standard deviation, denoted $\sigma(f)$, based on these K samples. For instance, from the control run, we divide the 1000-year data in $K = 10$ samples of 100-years. For the rare event algorithm, we have $K = 10$ independent realizations. We plot composite maps which are the averages of a quantity f, conditional on a heat wave to occur (f will be either surface temperature or geopotential height). We estimate the statistical significance of the average of a quantity f by computing its t value: $t = \sqrt{K}\mathbb{E}(f)/\sigma(f)$, where $\mathbb{E}(f)$ is the empirical average. t is a test of the null hypothesis that $\mathbb{E}(f) \neq 0$. For instance if $t > 2$, our interpretation of the test is that the probability that $\mathbb{E}(f) \neq 0$ and positive is larger or equal to 97%.

Plateau on return time plots. The plateau observed on return time plots for return times larger than 10^6 years are due to undersampling by the algorithm. Such a plateau is caused by the increasing multiplicity of trajectories with the same ancestor, as the amplitude a increases. Indeed, because of the selection procedure involved in the rare event algorithm, a subset of trajectories can share the same ancestor. Henceforth, they are likely to differ only by a small time-interval at the end of their whole duration. As a consequence, this subset of trajectories will contribute the same value to the set of observable maxima from which return times are computed, leading to a plateau. This feature is further discussed in (Lestang et al., 2018).

Lack of independence. As said, a consequence of the cloning mechanics is that the trajectories tend to have common ancestors in the first part of the simulation. This is visible in figure 4 of the main text. From a statistical point of view, the efficiency of the algorithm comes from a trade-off between the reduction of the independence among the trajectories and the shift of the distribution. In this example all the trajectories come from a common ancestor, and differentiation starts after the first 20 days. All the 10 sets of experiments show a similar behaviour, with the number of original ancestors varying between 1 and 3. Since the 10 sets of experiments start from pools of independent initial conditions, this means that even at the beginning of the simulation we always have at least 10-20 independent trajectories when we take composite averages.

References

Bouchet, F., Rolland, J., & Simonnet, E. (2019). Rare event algorithm links transitions in turbulent flows with activated nucleations. *Physical Review Letters*, *122*(7), 074502.

Del Moral, P. (2004). *Feynman-Kac Formulae Genealogical and Interacting Particle Systems with Applications*. New York, NY: Springer New York.

Del Moral, P., Garnier, J., et al. (2005). Genealogical particle analysis of rare events. *The Annals of Applied Probability*, *15*(4), 2496–2534.

Ebener, L., Margazoglou, G., Friedrich, J., Biferale, L., & Grauer, R. (2019). Instanton base importance sampling for rare events in stochastic pdes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*.

Giardina, C., Kurchan, J., Lecomte, V., & Tailleur, J. (2011). Simulating Rare Events in Dynamical Processes. *Journal of Statistical Physics*, *145*(4), 787-811. doi: 10.1007/s10955-011-0350-4

Grafke, T., Grauer, R., & Schäfer, T. (2015). The instanton method and its numerical implementation in fluid mechanics. *J.Phys.A:Math.Theor.*, *48*, 333001.

Grafke, T., & Vanden-Eijnden, E. (2019). Numerical computation of rare events via large deviation theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(6), 063118.

Hartmann, C., Banisch, R., Sarich, M., Badowski, T., & Schütte, C. (2013, December). Characterization of Rare Events in Molecular Dynamics. *Entropy*, *16*(1), 350–376. Retrieved 2015-11-18, from http://www.mdpi.com/1099-4300/16/1/350 doi: 10.3390/e16010350

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., ... Marshall, S. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339-

1360. doi: 10.1175/BAMS-D-12-00121.1

Kahn, H., & Harris, T. (1951). Estimation of particle transmission by random sampling. *Natl. Bur. Stand., Appl. Math. Ser.*, *12*, 27–30.

Laurie, J., & Bouchet, F. (2015). Computation of rare transitions in the barotropic quasi-geostrophic equations. *New J. of Phys.*, *17*, 015009.

Lestang, T., Bouchet, F., & Lévêque, E. (2020). Numerical study of extreme mechanical force exerted by a turbulent flow on a bluff body by direct and rare-event sampling techniques. *Journal of Fluid Mechanics*, *895*.

Lestang, T., Ragone, F., Bréhier, C.-E., Herbert, C., & Bouchet, F. (2018). Computing return times or return periods with rare event algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, *2018*(4), 043213. doi: 10.1088/1742-5468/aab856

Metzner, P., Schütte, C., & Vanden-Eijnden, E. (2009). Transition path theory for markov jump processes. *Multiscale Modeling & Simulation*, *7*(3), 1192–1219.

Plotkin, D. A., Webber, R. J., O'Neill, M. E., Weare, J., & Abbot, D. S. (2019). Maximizing simulated tropical cyclone intensity with action minimization. *Journal of Advances in Modeling Earth Systems*, *11*(4), 863–891.

Ragone, F., & Bouchet, F. (2020). Computation of extreme values of time averaged observables in climate models with large deviation techniques. *J Stat Phys*, *179*, 1637–1665. doi: 10.1007/s10955-019-02429-7

Ragone, F., Wouters, J., & Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, *115*(1), 24-29. doi: 10.1073/pnas.1712645115

Rolland, J., Bouchet, F., & Simonnet, E. (2016). Computing transition rates for the 1-D stochastic Ginzburg–Landau–Allen–Cahn equation for finite-amplitude noise with a rare event algorithm . *J. Stat. Phys.*, *162*, 277–311.

Webber, R. J., Plotkin, D. A., O'Neill, M. E., Abbot, D. S., & Weare, J. (2019). Practical rare event sampling for extreme mesoscale weather. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(5), 053109. doi: 10.1063/1.5081461

Wouters, J., & Bouchet, F. (2016). Rare event computation in deterministic chaotic systems using genealogical particle analysis. *Journal of Physics A: Mathematical and Theoretical*, *49*(37), 374002.

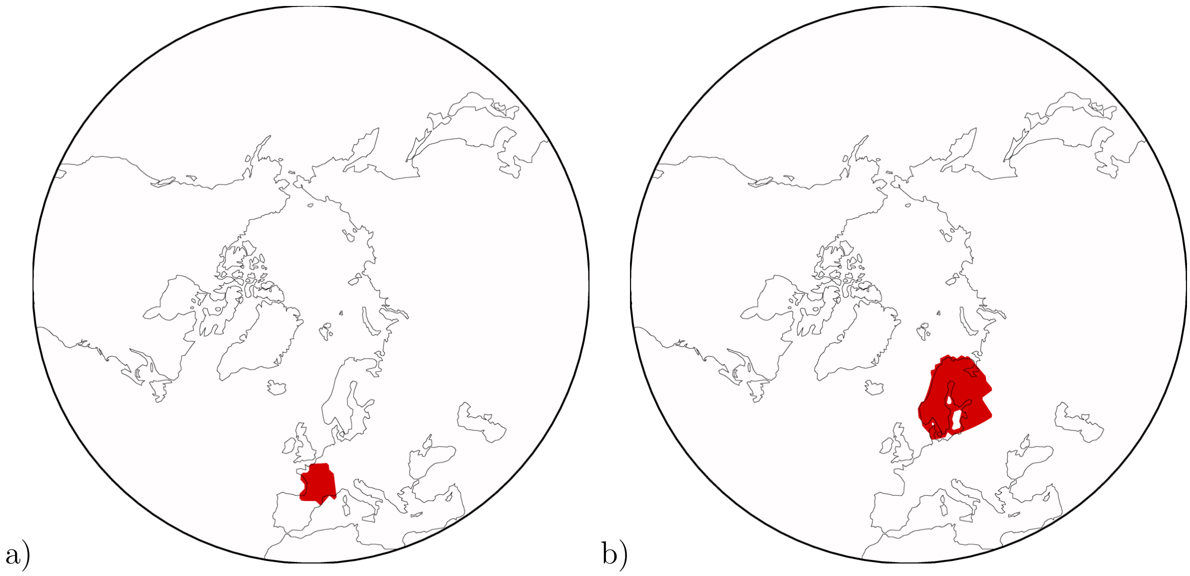

Figure S1. Domain D used to define the area integral in the selection function of the algorithm, for France (a) and Scandinavia (b).

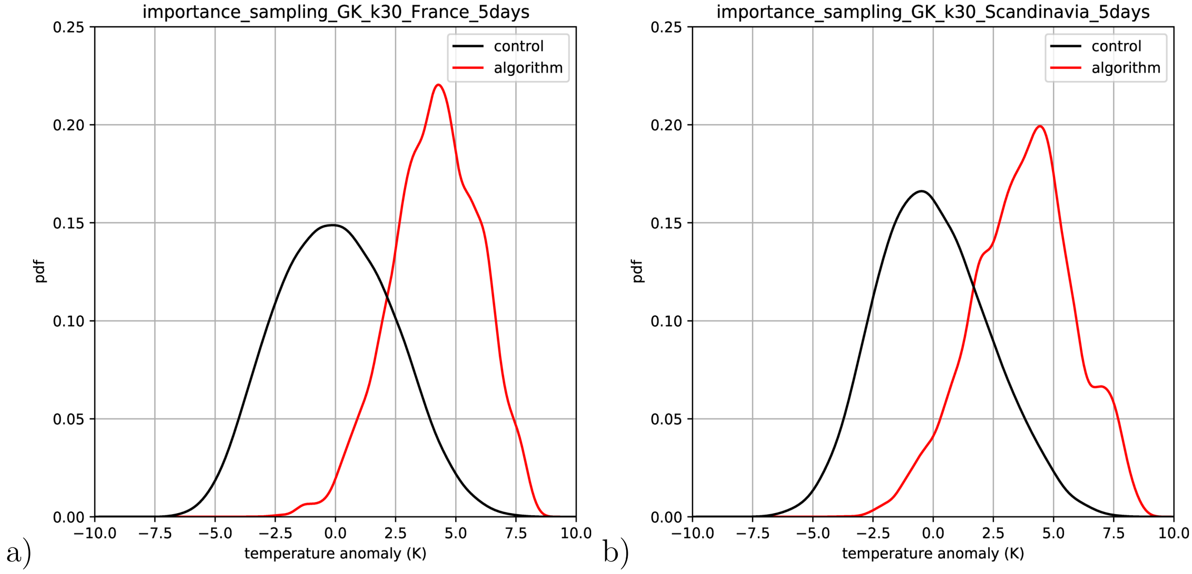

Figure S2. Distribution of 5 days running mean temperatures anomalies averaged over France (a) and Scandinavia (b) from a control run of 1000 years (black) and a set of 10 experiments with the algorithm with target domain the corresponsing region, with k=30, trajectories N=100, resampling time 5 days, and running time 90 days starting from June 1st (red).