Rare event computation in deterministic chaotic systems using genealogical particle analysis

J Wouters$^{1,2,3,4}$ and F Bouchet$^3$

$^1$ School of Mathematics and Statistics, University of Sydney, Sydney, Australia
$^2$ Meteorological Institute, University of Hamburg, Hamburg, Germany
$^3$ Laboratoire de Physique, École Normale Supérieure de Lyon, Lyon, France

E-mail: jeroen.wouters@uni-hamburg.de and freddy.bouchet@ens-lyon.fr

Received 31 October 2015, revised 17 May 2016
Accepted for publication 18 July 2016
Published 24 August 2016

Abstract

In this paper we address the use of rare event computation techniques to estimate small over-threshold probabilities of observables in deterministic dynamical systems. We demonstrate that genealogical particle analysis algorithms can be successfully applied to a toy model of atmospheric dynamics, the Lorenz 96 model. We furthermore use the Ornstein–Uhlenbeck system to illustrate a number of implementation issues. We also show how a time-dependent objective function based on the fluctuation path to a high threshold can greatly improve the performance of the estimator compared to a fixed-in-time objective function.

Keywords: rare event simulation, deterministic dynamical systems, genealogical particle analysis, large deviation theory

1. Introduction

Rare events may have a large impact on the dynamics of geophysical turbulent flows and the climate. In bistability situations, a rare transition can drastically change the structure of the flow, like for instance the bistability of the Kuroshio current [1, 2] or a change of polarity of the Earth’s magnetic field due to the turbulent dynamics of the Earth metal core [3]. Rare events can also be extremely important because of their impact on society, ecosystems or the economy. There are many such examples in climate dynamics, for example extreme droughts, heat waves, rainfalls and storms [4]. The probability and the impact of these events is likely to
change in the future due to a changing climate [4]. The magnitudes of possible changes are however still uncertain [5].

On the one hand, for climate dynamics, there is a lack of sufficient reliable empirical data [6]. How could one assess faithfully the probability of events with recurrence times longer than one decade with only one or two century long reliable data? In the last decade, many methods have been developed to extract the most information possible from this too short time-series. For instance extreme value statistics [7, 8] has interestingly allowed to extrapolate from the information available from empirical observation [9, 10].

Another approach would be to critically study and understand the dynamics of rare events produced by complex climate models. This second approach seems to be the only available one for events with a recurrence time longer than decades or centuries. However, the current scientific state of the art does not yet allow to obtain many results by following this route. The first critical issue is a sampling problem. Indeed, if one wants to study events with centenary or millenial recurrence times and assess the reliability of the model dynamics to produce those events, a direct numerical simulation would require to have model runs of at least hundreds of thousands of years long in order to get reliable statistics on both the probability of and the dynamics leading to these events. As it is not always reasonable to trade computational length with model complexity, especially for the turbulent part which is responsible for most of the fluctuations, it is clear that we are facing an extremely difficult scientific challenge.

Is there a way to produce reliable statistics of specific rare events of a given model, without having to rely on prohibitively long direct numerical simulations? The same issue has been faced in many other scientific fields and has led to the development of some interesting approaches. Indeed, many of the complex systems studied in different branches of science feature events that are very rare but nevertheless very relevant due to their high impact. Take for example buffer overflows in digital communication networks, the insolvency of an insurer or bank, collisions in planetary systems, the dynamics of phase transitions in condensed matter, the long time dynamics of complex molecules in chemistry or biology, to name but a few. In recent years a number of promising algorithms have been developed to tackle these problems [11–14]. These rare event simulation algorithms can drastically reduce the error made on the estimation of small probabilities.

Generally speaking, the objective of the algorithms is to make rare events less rare, either by altering the dynamics (importance sampling) or by targeted killing and cloning of an ensemble of realizations (genealogical particle analysis or interacting particle algorithms). Upon estimation the intervention of the algorithm is then taken into account to obtain an estimate for quantities of the original system.

Within the class of genealogical particle analysis algorithms, a number of different strategies exist. A first crucial difference is the type of quantity one aims at estimating. One can be interested in the distribution of first entrance time to a set or to sample transition paths [15–18], the probability of a rare event [19–21] or expectation values of long time averages such as the scaled cumulant generation function [22–27] as it appears in large deviation theory. For these different aims again different algorithms exist, for example genealogical algorithms with fixed [19, 20] versus variable [21] particle numbers, minimum action algorithms [18], or milestoning [28]. These algorithms have already been applied to a wide range of systems, for example percolation problems [29], in complex chemistry [30], polymer and biomolecule dynamics [31–34], magnetism [35, 36], Burger turbulence [37, 38].

The aim of this work is to make a first step in the application of those approaches to problems in climate dynamics. Climate dynamics has specificities that make past approaches not directly adaptable. First of all, the climate is clearly out of equilibrium (without time reversal symmetry or detailed balance), therefore only non-equilibrium approaches can be
considered. Second, the phenomenology of geophysical turbulent flows is dominated by large scale synoptic scales and is rather different from other complex dynamics, for instance molecular dynamics. And third, most climate models are deterministic models, or sometimes include a stochasticity that does not affect directly the synoptic scales.

The aim of this paper is to consider the latter specificity of many climate model. We address the following question: can rare event algorithms based on genealogical particle analysis be used effectively and efficiently for deterministic dynamics? Most algorithms rely on a Markov assumption, which is verified for deterministic models. However at the cloning stage, a new trajectory is branched from another one in order to produce a new ensemble member. For a strictly deterministic system, the offspring trajectory will not be different from its parent. To ensure separation of the two trajectories, one has therefore to add either a very small noise on the overall dynamics or a small change on the initial condition of the offspring trajectory and rely on the chaoticity of the dynamics. A key issue is then to verify a posteriori that the noise is small enough so as not to distort the measured statistics and probabilities. A test using different decreasing noise strengths and checking for stability should therefore be used.

In order to perform the first study of the effectiveness of these approaches for chaotic deterministic dynamics, we have chosen to study a simple chaotic system with many degrees of freedom, and of relevance for climate dynamics: the Lorenz’96 model [39, 40]. We have also chosen the conceptually simplest and most robust genealogical algorithm that allows to sample invariant measure or transition probabilities: the genealogical particle analysis algorithm. We give a detailed heuristic presentation of the algorithm and a benchmark on the Ornstein–Uhlenbeck process in section 4.

The genealogical particle analysis algorithms explore the statistics of solutions of the dynamical system by running an ensemble of realizations, interrupting the ensemble simulation at given times and killing ensemble members that do not perform well as measured by a weight or objective function and cloning the ones with a high weight. This selective procedure explains the terminology genealogical particle analysis, since the individual realizations are sometimes referred to as particles. The design of a good objective function is then arguably the main design issue one faces when using genealogical particle analysis algorithms. Other choices that have to be made are the number and timing of interactions and the number of particles to use. We will address these practical issues in a detailed study of the genealogical particle analysis algorithms on the Ornstein–Uhlenbeck process. This process is easy to simulate numerically and allows for analytic expressions to be derived; it is therefore well suited for the purpose of illustration and testing.

Another aim of this paper is to propose a systematic approach and procedure to get reliable results and error estimates. We propose to build the tail of the cumulant distribution function of interest by gluing together pieces of results obtained for different cloning parameter by a systematic study of the most reliable one, through an empirical estimate of the algorithm variance. Moreover, we propose a procedure to test empirically this class of algorithms against the real dynamics. Indeed, for a model like the Lorenz’96 model, we have no theoretical results that can serve as a benchmark.

The paper is organized as follows. In section 2 we discuss how the need for rare event simulation techniques arises, what the objective of such algorithms is (making rare events typical) and how this goal can be achieved for stochastic processes by implementing a genealogical particle analysis simulation. In section 3 we present a brief discussion of the theory of large deviations and what it can say about the way in which rare events are reached by a process. This theory can be used to implement a more efficient rare event sampling method. In section 4 we proceed by implementing the genealogical particle analysis simulation to the Ornstein–Uhlenbeck system. We discuss in depth the selection of the parameters in the algorithm. In section 5 we then present the implementation of the genealogical particle
analysis simulation on a chaotic deterministic dynamical system. Finally, we present our conclusions in section 6.

2. Rare event computation for Markov dynamics

In sections 2.1 and 2.2 we present a classical discussion of the inefficiency of brute force Monte Carlo simulation for estimating small probabilities. This motivates the need for rare event computation techniques. We introduce the genealogical particle analysis algorithm and the related theory in section 2.3.

2.1. Motivation

The goal of rare event computation techniques is to make the numerical estimation of small probabilities more efficient. The necessity of using such techniques is demonstrated by the sampling of the tail of a distribution \( P \) using independent samples identically distributed according to the distribution \( P \). Say one wants to estimate a small probability \( \gamma_A = P(X \in A) \ll 1 \) by means of a brute force Monte Carlo estimate

\[
\hat{\gamma}_A = \frac{1}{N} \sum_{i=1}^{N} I_A(X_i) ,
\]

where \( I_A \) is the indicator function on the set \( A \). The estimator \( \hat{\gamma}_A \) is an unbiased estimator of \( \gamma_A \) since the expectation value of \( \hat{\gamma}_A \) is clearly \( \gamma_A \). When the number of samples \( N \) is large enough for \( \hat{\gamma}_A \) to follow a central limit theorem, the statistical error of the estimator can be quantified by its variance

\[
\text{Var}(\hat{\gamma}_A) = \frac{1}{N} \text{Var}(1_A(X))/N ,
\]

where

\[
\text{Var}(1_A(X)) = E((1_A(X) - \gamma_A)^2) = E(1_A(X)) - \gamma_A^2 = \gamma_A - \gamma_A^2 \\
\approx \gamma_A
\]

when \( \gamma_A \) is small. The relative error (RE) of the estimator being proportional to the standard deviation divided by the estimated quantity, we have \( \text{RE} \sim \frac{1}{\sqrt{\gamma_A}} \). The relative error quickly becomes large as \( \gamma_A \) goes to zero for fixed sample size \( N \). Fortunately there exist methods for estimating small probabilities more efficiently.

2.2. Importance sampling

The main ingredient of rare event computation techniques is a sampling from a modified distribution together with an adapted estimator to counteract this change of measure. This method to lower the estimator variance of a rare event probability is termed importance sampling. Again the example of the sampling from independent identically distributed random variables provides valuable insights.

Say we want to estimate

\[
\gamma_A = \int \text{d}X \rho(X) I_A(X) \ll 1 ,
\]

where \( \rho \) is the density for our random variable \( X \). Instead of doing a straightforward sampling of \( X \) as in (1), assume we can sample from a modified measure \( \tilde{\rho} \) for which \( \tilde{\rho}(X) > 0 \)
whenever \( X \in A \) and \( \rho(X) > 0 \). In such a case, the probability we want to estimate can be rewritten as

\[
\gamma_A = \int \rho(X) \frac{\rho(X)}{\rho(X)} 1_A(X) = \hat{E}(L(X) 1_A(X))
\]

and \( L(X) \equiv \frac{\rho(X)}{\rho(X)} \) whenever \( 1_A(X) \rho(X) > 0 \)

and we can therefore estimate \( \gamma_A \) using the estimator

\[
\tilde{\gamma}_A = \frac{1}{N} \sum_{i=1}^{N} L(\tilde{X}_i) 1_A(\tilde{X}_i)
\]

on samples \( \tilde{X}_i \) distributed according to \( \tilde{\rho} \).

The variance for such an estimator is

\[
\tilde{\text{Var}}(L(X) 1_A(X)) = \hat{E}(L^2(X) 1_A(X)) - \gamma_A^2 = E(L(X) 1_A(X)) - \gamma_A^2.
\]

If we could take \( \tilde{\rho} \) to be the conditional measure with \( \tilde{\rho}(X) = \rho(X)/\gamma_A \) for \( X \in A \) and zero elsewhere, such that \( L(X) = 1_A(X) \gamma_A \), this would result in a zero variance estimator. This estimator is however not practically implementable, since for this we would need to know the value of \( \gamma_A \), which is the value we seek to calculate.

This calculation demonstrates some important points however. First of all, it shows that a change of measure can indeed reduce the variance of the estimator. Although the ideal change
of measure is not feasible in practice, a change of measure that is in a sense close to it should also give a substantial variance reduction. This modified measure should therefore have most of its weight on the set of interest \( A \). On the other hand, this also implies that one needs to have some understanding of the shape of the set \( A \) and the distribution on it to construct an efficient importance sampling.

2.2.1. Tilting a normal distribution. To illustrate how a change of measure can provide significant variance reductions, even if the modified measure is not the ideal conditional measure, we discuss an example for normally distributed random variables. This example will also be useful to illustrate and validate our rare event algorithm for dynamical systems.

Say we want to estimate the probability of the rare event \( \{ x > a \} \) for a normally distributed random variable \( x \sim N_{0,1} \) with zero average and standard deviation equal to one. Assume that we can tilt the distribution with an exponential function

\[
\rho(X) = \rho(X) \exp(CX)/E(\exp(CX)) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-C)^2}{2}\right]
\]

which constitutes of a shift of the average by \( C \). Since \( L(X) = E(e^{CX})/\exp(CX) = \exp(-CX + C^2/2) \), the variance of the terms in the importance sampling estimator is now

\[
\hat{\text{Var}}(L(X)1_A(X)) = P_{-c,i}(x > a) e^{C^2} - \gamma_{\lambda'}^2
\]

where \( P_{\mu,\sigma} \) denotes probabilities under a normal distribution with mean \( \mu \) and variance \( \sigma \). The standard deviation, the square root of the variance (4), is plotted for \( a = 2 \) in figure 1. The standard deviation has a single minimum, which is obtained for a value of \( C \) which is close to \( C = 2 \), for which the mean of the tilted value coincides with the threshold. This basic example illustrates how importance sampling can lower dramatically the estimator variance.

As figure 1 shows, the relative error, which is proportional to the plotted quantity \( \text{Std}(\gamma_{\lambda'})/\gamma_{\lambda'} \), can be reduced by a factor of more than 4 for the case where \( \lambda = [2, +\infty) \). Since the error decreases as \( 1/\sqrt{N} \), this means a at least 16-fold longer brute force simulation would be necessary to obtain a similarly accurate result. For graphical purposes a relatively low threshold 2 was chosen here. For higher thresholds, the performance gains increase drastically, with a reduction of computational effort by a factor \( 6 \times 10^5 \) when \( \lambda = [5, +\infty) \).
2.3. Genealogical particle analysis algorithms

The motivation for rare event simulation and the discussion of importance sampling have shown that it is necessary to make rare events less rare. This concept can also be applied to processes such as the paths followed by either stochastic dynamics or chaotic deterministic dynamics. In those cases the objective is to alter the probability of certain paths that are connected to the rare event one wants to study.

Two different strategies are employed to alter the path sampling in stochastic dynamical systems. The first one is to alter the dynamical equations of the system by introducing a forcing term \( \mathbf{41} \). By tuning a parameter of the added forcing term, one can then attempt to decrease the variance of the rare event estimator. A second strategy consists in calculating an ensemble of realizations of the stochastic system in parallel and manipulating the ensemble members by performing selections at a finite number of selection times so as to bias the population.

Here we will use the second strategy, by employing a variant of the so called genealogical particle analysis algorithms. The selections applied to the ensemble consist of dynamical trajectories, called particles, being copied and killed depending on weight factor asigned to every ensemble member. The procedure applied in the algorithm is illustrated in figure 2. This strategy has the advantage of not altering the dynamical trajectories themselves, such that their dynamics can be studied \textit{a posteriori}.

Extensive analysis of the convergence of genealogical particle analysis algorithms can be found for example in \[13, 14\]. In the following sections, we perform a simpler calculation, assuming a mean field approximation, to demonstrate the evolution of the expected particle distributions in a genealogical particle analysis. The validity of this mean field approximation for large particle number is the subject of the complete proofs given in \[13, 14\]. Before going to a truly interacting genealogical particle analysis in section 2.3.3, we first get some insight by looking at an algorithm where particles are reweighted, but by a factor depending only on the evolution of the particle itself, in section 2.3.1.

2.3.1. A non-interacting genealogical particle analysis. We calculate rare events of a continuous time Markov chain. The transition probabilities from configuration \( x \) to \( y \) over a time interval \( \Delta t \) are denoted by \( P^{(2)}(y|x, \Delta t) \). We are interested in the probability of being in a set of configurations \( A \) at a time \( t = t_n \), given that the process started in configuration \( x_0 \) at time \( t = 0 \). The number of particles at time \( t \) is denoted by \( N_t \), whereas \( \{ \xi_i \}_{i \in N} \) denote the particle configuration at time \( t \). Expectation values under the original Markov dynamics \( P^{(2)} \) at time \( t \) are denoted by \( E_{0,t} \). The algorithm to generate the particles is described in the box Algorithm 1.

The algorithm can be summarized as follows: after initialization (step (i)) the ensemble members are evolved forward in time (step (ii-a)) and weight values are calculated from the previous and current configurations of the particle, \( \xi_{i,k} \) and \( \xi_{i,k-1} \) respectively (step (ii-b)). Based on these weight values, ensemble members are killed or cloned (step (ii-c)). Repeating this procedure results in a reweighted sample of paths, from which expectation values of the unweighted path distribution can be estimated (step iii). The rare event probability \( \gamma_A \) for a set \( A \) can be obtained by taking as observable \( F(x) = F_A(x) = 1_A e^{-V(x)\Delta t} \) such that as the end result we obtain \( E_{0,t}(F \omega) e^{-V(x_0)} = E_{0,t}(1_A) \).

Note that the random number \( N_{i,k} \) generated in step (ii-c) can be zero, such that particles can be killed as well as cloned (when \( N_{i,k} > 1 \)). A way to generate the random number
described in step (ii-c) is to take $N_{ik} = \lfloor W_k u \rfloor$, where $u$ is uniformly distributed on $[0, 1]$ and $\lfloor x \rfloor$ is the floor of $x$ (the largest integer smaller than $x$).

**Algorithm 1.** Non-interacting weighted particle system

(i) Initiate $M$ particles in configuration $x_0$: $\zeta_{i,0} = x_0$ for $1 \leq i \leq N_0 = M$.

(ii) For every time step $k \in \{1, \ldots, n\}$:

(a) Propagate $\zeta_{i,k-1}$ under the dynamics, resulting in $\zeta_{i,k}$ distributed according to $P^{\Delta t}(\zeta_{i,k} | \zeta_{i,k-1}, \Delta t)$ with $\Delta t = t_k - t_{k-1}$.

(b) Calculate weights $W_k$ for particle $i$:

$$W_k = W(\zeta_{i,k}, \zeta_{i,k-1}) := \exp(V(\zeta_{i,k}) - V(\zeta_{i,k-1}))$$

for a suitably chosen function $V$.

(c) Generate a new particle distribution $\xi_{i,k}$ consisting of $N_{ik}$ copies of the particle with configuration $\zeta_{i,k}$ where $N_{ik}$ is chosen at random such that $E(N_{ik}) = W_k$ (note that $N_k = \sum_i N_{ik}$).

(iii) Finally, for any $F$, calculate $\tilde{F} = \frac{1}{M} \sum_{i=1}^{N_k} F(\xi_{i,k})$ to estimate $E_{0,t_k}(F^{\gamma_k})$ (to estimate $\gamma_k$ take $F(x) = F_t(x) = 1_i(x)\exp(V(x_0) - V(x))$).

2.3.2. Unbiased estimator. We first show that algorithm 1 provides an unbiased estimator for the quantity $E_{0,t_k}(F^{\gamma_k})e^{-V(x_0)}$, i.e. the algorithm results in a random estimate whose expectation value equals the quantity to be estimated:

$$E_1 \left( \frac{1}{M} \sum_{i=1}^{N_k} F(\xi_{i,k}) \right) = E_{0,t_k}(F^{\gamma_k})e^{-V(x_0)},$$

where $E_1$ is the expectation over the random variables in the algorithm.

Write $N(x, t)$ the particle number at configuration $x$, i.e. $N(x, t)\, dx$ is the number of particles with $x \leq x_t < x + dx$:

$$N(x, t_{k-1}) = \sum_{i=1}^{N_{ik}} \delta(x - \zeta_{i,k-1}).$$

According to algorithm 1, if a particle sits at $\zeta_{i,k-1}$ at time step $k - 1$, $N_{ik}$ copies are created of $\zeta_{i,k}$ at the next time step. Hence, the particle number at the next time step will be

$$N(x, t_k) = \sum_{i=1}^{N_{ik}} N_{ik} \delta(x - \zeta_{i,k}). \quad (5)$$

One step in the algorithm therefore involves the generation of two sets of random variables, the updated particle configurations $\zeta_{i,k}$, which is conditioned on $\zeta_{i,k-1}$, and the number of particle copies $N_{ik}$, which depends on both $\zeta_{i,k}$ and $\zeta_{i,k-1}$. The expectation value of functions depending on the particle configurations $\zeta_{i,k}$ at step $k$ can then be expressed as the expectation value

$$E_{\zeta_{i,k}}(\bullet) = E_{\zeta_{i,k-1}}(E_{\zeta_{i,k}}|\zeta_{i,k-1}(E_{N_{ik}}|\zeta_{i,k-1}(\bullet))).$$

Applying this expression to equation (5) and using the probabilities for the updated particle configurations $P(\zeta_{i,k} | \zeta_{i,k-1}) = P^{\Delta t}(\zeta_{i,k} | \zeta_{i,k-1}, \Delta t)$ and that the number of particle copies $E_{N_{ik}}(\zeta_{i,k} | \zeta_{i,k-1}) = W_k(\xi_{i,k}, \zeta_{i,k-1})$, we have
\[ E_1(N(x, t_k)) = E_{\xi_{i,t_k}}(E_{\xi_{i,t_k}}(\sum_{j=1}^{N_t} W_{j,k}(\xi_{i,t_k}, \xi_{j,t_k}) \delta(x - \xi_{j,t_k}))) \]

\[ = E_{\xi_{i,t_k}}(\sum_{j=1}^{N_t} P^{(2)}(y|\xi_{i,t_k}, \Delta t_k) W_{j,k}(y, \xi_{i,t_k}) \delta(x - y)) \]

\[ = E_{\xi_{i,t_k}}(\sum_{j=1}^{N_t} P^{(2)}(x|\xi_{i,t_k}, \Delta t_k) W_{j,k}(x, \xi_{i,t_k})) \]

\[ = E_{\xi_{i,t_k}}(\int dz P^{(2)}(x|z, \Delta t_k) W(x, z) \delta(z - \xi_{i,t_k})) \]

\[ = \int dz P^{(2)}(x|z, \Delta t_k) W(x, z) E_1(N(z, t_{k-1})) \]

\[ = \int dz P^{(2)}(x|z, \Delta t_k) e^{V(x) - V(z)} E_1(N(z, t_{k-1})). \]

This equation relates the expected particle density at step \( k \) to the density at step \( k - 1 \). By iteration we can relate the density at step \( k \) to the density at the start of the algorithm, which is \( M\delta(x - x_0) \):

\[ E_1(N(x, t_k)) = \int dx_{k-1} \ldots dx_1 dz P^{(2)}(x|x_{k-1}, \Delta t_k) \ldots P^{(2)}(x_1|z, \Delta t_k) \]

\[ \times e^{V(x) - V(z)} M\delta(z - x_0) \]

\[ = MP^{(2)}(x|x_0, t_k - t_0) e^{V(x) - V(x_0)}. \]

The expectation value of the quantity calculated at the end of the algorithm in step (iii) is therefore

\[ E_1\left(\frac{1}{M} \sum_{i=1}^{N_t} F(\xi_{i,t_k})\right) = \frac{1}{M} E_1\left(\int dx \sum_{i=1}^{N_t} \delta(x - \xi_{i,t_k}) F(x)\right) \]

\[ = \frac{1}{M} \int dx E_1(N(x, t_0)) F(x) \]

\[ = \int dx F(x) e^{V(x) - V(x_0)} P^{(2)}(x|x_0, t_k - t_0) \]

\[ = E_{0,t_k}(F e^{V(x)}). \]

Note that the expected total particle number

\[ E_1(N(t_k)) = \int dx E_1(N(x, t_k)) = M \int dx F^{(2)}(x|x_0, t_k - t_0) e^{V(x) - V(x_0)} \]

is in general not conserved over time. The particle number can strongly increase, which entails a large numerical cost. The solution to this problem is to renormalize the weights calculated in step (ii-b) of the algorithm, hence introducing an interaction between the ensemble members. We discuss this new algorithm in the next section. As we will see, the interaction complicates the algorithm analysis.

2.3.3. Interacting particles. We now add interaction to the weights of the particle system, so as to control the particle number. A similar analysis as in the previous section can still be carried out, if one assumes that the number of particles used in algorithm is large enough, such that averages over particle configurations can be replaced by an expectation value under the law of large numbers (mean field approximation). The corresponding algorithm is described in the box Algorithm 2. By applying the algorithm to a function \( F(x) = F_\lambda(x) = 1_\lambda(x) \exp(V(x_0) - V(x)) A \), estimates \( \gamma_\lambda \) of the probability \( \gamma_\lambda \) can be obtained.
Algorithm 2. Interacting genealogical particle analysis

(i) Initiate $M$ particles in configuration $x_0$, $\xi_i^0 = x_0$ for $1 \leq i \leq N_0 = M$.
(ii) For every time step $k \in \{1, \ldots, n\}$:
   (a) Propagate $\xi_{i,k-1}$ under the dynamics, resulting in $\xi_{i,k}$ distributed according to $P^{2i}(\xi_{i,k}|\xi_{i,k-1}, \Delta t_k)$ with $\Delta t_k = t_k - t_{k-1}$.
   (b) Calculate weights $\tilde{W}_{i,k}$ for particle $i$:
       \[
       \tilde{W}_{i,k} = \frac{W(\xi_{i,k}, \xi_{i,k-1})}{Z_k},
       \]
       \[
       Z_k = \frac{1}{N_k} \sum_i W(\xi_{i,k}, \xi_{i,k-1}).
       \]
   (c) Store the value of the normalizing factor $Z_k$.
   (d) Generate a new particle distribution $\xi_{i,k}$ consisting of $N_{i,k}$ copies of the particle with configuration $\xi_{i,k}$, where $N_{i,k}$ is chosen at random such that $E(N_{i,k}) = \tilde{W}_{i,k}$.

(iii) Finally, for any $F$, calculate $\frac{1}{M} \sum_{k=1}^{n} F(\xi_{i,k}) \prod_{k=1}^{n} Z_k$ to estimate $E_0(F(x)) e^{-\gamma(x)}$ (to estimate $\gamma(x)$ take $F(x) = F_{\lambda}(x) = 1_{\lambda}(x) \exp(V(x_0) - V(x))$).

We again perform an analysis of the evolution of the expected particle distribution in algorithm 2. For simplicity of derivation we assume that the number of particles $N_{i,k}$ in the algorithm is large, such that by the law of large numbers

\[
Z_k = \frac{1}{N_k} \sum_{i} W(\xi_{i,k}, \xi_{i,k-1}) \approx E_1(W(\xi_{i,k}, \xi_{i,k-1})), \tag{6}
\]

\[
\approx \int dx \int dy \frac{E_1(N(x, t_{k-1}))}{E_1(N_{i,k-1})} P^{2i}(y|x, \Delta t_k) W(y, x). \tag{7}
\]

Using this estimate and the same reasoning as for the non-interacting particle algorithm, we have for the expected particle distribution that

\[
E_1(N(x, t_k)) \approx \int dz P^{2i}(z|x, \Delta t_k) \frac{W(x, z)}{E_1(W)} E_1(N(z, t_{k-1})). \tag{8}
\]

The expectation value of $W$ in the denominator can be substituted using equation (6). The particle number is now constant:

\[
E_1(N_{i,k}) = \int dx E_1(N(x, t_k)) = E_1(N_{i,k-1}) = \cdots = E_1(N_{i,0}) = M.
\]

We therefore have that

\[
E_1(W(\xi_{i,k}, \xi_{i,k-1})) \approx \int dx \int dy P^{2i}(y|x, \Delta t_k) W(y, x) \frac{E_1(N(x, t_{k-1}))}{M}.
\]

Inserting this into (8), we have

\[
E_1(N(x, t_k)) \approx M \int \frac{dy P^{2i}(y|x, \Delta t_k) W(x, y) E_1(N(y, t_{k-1}))}{\int \frac{dy P^{2i}(y|x, \Delta t_k) W(x, y) E_1(N(y, t_{k-1}))}. \tag{9}
\]
By iteration it follows that
\[
\frac{1}{M} \sum_{i=1}^{N} F(x_{i,t}) \approx \frac{E_0(F(x_n)W(x_n, x_{n-1}) \ldots W(x_1, x_0))}{Z_n} = \frac{E_0(F(x_n)W(x_n, x_{n-1}) \ldots W(x_1, x_0))}{Z_n},
\]

where \( Z_n = E_0(W(x_n, x_{n-1}) \ldots W(x_1, x_0)) \). From equations (6) and (8) we see that \( Z_k \approx \frac{Z_{k-1}}{Z_k} \) and therefore
\[
\frac{1}{M} \sum_{i=1}^{N} F(x_{i,t}) \prod_{k=1}^{n} Z_k \approx E_0(F(x_n)W(x_n, x_{n-1}) \ldots W(x_1, x_0)) = E_0(Fe^{V(x)})e^{-V(x_0)}.
\]

The above reasoning can also be extended to show that path dependent quantities (such as \( E[x(\tau) \mid x(T) > a] \) for \( \tau < T \)) can be estimated from the ancestral paths of the particle system.

2.3.4. Time-dependent weighting. The weighting function \( W(x, y) = \exp(V(x) - V(y)) \) results in a particle distribution tilted by \( \exp(V(x)) \) at all selection times \( t_k \). More flexibility can be obtained by using time-dependent weighting, for example with a weighting function of the form
\[
W(t_k, x, y) = \exp(V_h(x) - V_{h-1}(y)).
\]

This way the telescoping canceling of \( V_h \) is preserved in products of weights that appear in the calculation of the tilted measure. For example
\[
W(t_k, x, y)W(t_{k-1}, y, z) = e^{V_h(x) - V_{h-1}(y)}e^{V_{h-1}(y) - V_{h-2}(z)} = \exp(V_h(x) - V_{h-1}(z)).
\]
The result is again a particle distribution tilted by \( \exp(V_h(x)) \) at time \( t_k \), as with the time-independent weight function. However, paths up to the final time will have different weights, which can make a large difference in the algorithm performance, as we will demonstrate in section 4.2.4.

3. Fluctuation paths and the weighting function

The ideal change of measure discussed in section 2.2 suggests to make the rare event that is the least rare the most probable one under the reweighted dynamics. This rationale extends not only to the distribution of the system at the final time, but also to the entire path up to the final time. This means that variance can be reduced if the least unlikely path leading to a high threshold is made more likely under the particle system dynamics.

For stochastic differential equations in the weak noise limit, the least unlikely path from an attractor can be calculated from Freidlin–Wentzell type large deviation theory and is called a fluctuation path (also sometimes an instanton). The particle system dynamics can be made to more closely follow the fluctuation path by using the time-dependent weighting discussed in section 2.3.4. Even if the particle distribution at the final time is the same as the particle distribution obtained with a constant weighting, there is still a variance reduction since less particles are killed, increasing the independence of the particle and thus the effective particle number.
3.1. Fluctuation paths

The probability of a given path in a stochastic differential equation with small noise

\[ dX = b(X')dt + \sqrt{\epsilon} dW, \]

where \( W \) is a Brownian motion, can be estimated using the Freidlin–Wentzell large deviation theory. The theory determines the probability of seeing a path that is close to a specified continuous function in the limit of \( \epsilon \) going to zero. It roughly states that

\[ \lim_{\epsilon \to 0} \epsilon \log P[X(t) \in F] = -\inf_{\omega \in F} I(\omega), \]

where \( F \) is any closed subset of the set of continuous trajectories and the rate functional \( I \) is called the action. The action is given by

\[ I(\omega) = \frac{1}{2} \int_0^T dt (\dot{\omega}(t) - b(\omega(t)))^2 = \int_0^T dt \mathcal{L}\left[ \omega, \frac{\partial \omega}{\partial t} \right], \]

(10)

\[ \mathcal{L}\left[ \omega, \frac{\partial \omega}{\partial t} \right] = \frac{1}{2}(\dot{\omega}(t) - b(\omega(t)))^2. \]

(11)

The distribution of paths leading to rare fluctuations then concentrates around action minima as \( \epsilon \) decreases, with given constraints. If the set of paths \( F \) contains the evolution along the deterministic dynamics \( \dot{x} = b(x) \), this path will obviously minimize the above action, hence the need for constraints to obtain more interesting results. For example, in the simple case were the deterministic dynamics \( \dot{x} = b(x) \) has a single attractor \( x_0 \), the distribution of the paths conditioned on \( X(0) = X_0, X_0 = x_0 \) evolve according to the equations

\[
\dot{m} = -\lambda m \\
\dot{v} = \sigma^2 - 2\lambda v
\]

which can be solved explicitly.

4. Rare event simulation for a stochastic process: the Ornstein–Uhlenbeck process

We now illustrate some of the practical issues arising when implementing a genealogical particle analysis algorithm for rare event estimation. We start off with a stochastic process for which we can calculate explicitly all of the probabilities that we want to estimate, for pedagogical reasons, and so that we can compare the numerical results to the analytic expressions.

4.1. Description of the Ornstein–Uhlenbeck process

We consider the Ornstein–Uhlenbeck process

\[ dx = -\lambda x dt + \sigma dW. \]

(12)

As the transition probabilities \( P(x(t)|x(0)) \) are Gaussian, the Ornstein–Uhlenbeck process preserves Gaussianity. Using the Itô formula, one can derive that the mean \( m(t) = E(x(t)) \) and the variance \( \nu(t) = E((x(t) - m(t))^2) \) evolve according to the equations

\[
m = -\lambda m \\
\nu = \sigma^2 - 2\lambda \nu
\]

which can be solved explicitly.
The probability that \( x \) exceeds a certain threshold \( a \) at a time \( t \), given that the process started at \( x(0) = x_0 \) at time zero can be calculated explicitly as

\[
P(x(t) > a | x(0) = 0) = \int_a^\infty \mathcal{N}(m(t),\nu(t))(x) \, dx,
\]

where \( m \) and \( \nu \) solve (13) with initial conditions \( m(0) = \nu(0) = 0 \) and \( \mathcal{N} \) is the probability density function of the normal distribution with mean \( m \) and variance \( \nu \). We consider in the following the estimation of this probability through a genealogical particle analysis algorithm. Below we will use the parameter values \( \lambda = 1 \) and \( \sigma = 1 \).

4.2. Algorithm implementation

Let us assume that we seek to estimate a small probability \( \gamma_A \), for instance \( \gamma_A = P(x(t) \in A | x(0) = 0) \). We denote by \( M \) the average number of particles for each realization of the algorithm. Then each independent realization \( i \) of the algorithm, each with \( M \) particles on average, will give an estimate \( \hat{\gamma}_A \). According to a theorem discussed in [19], asymptotically for large \( M \), the random number \( \gamma_A \) is distributed according to a Gaussian distribution with standard deviation \( \sigma_A(M) = \sigma_A / \sqrt{M} \) and a corresponding relative error \( \text{RE}(M) = \sigma_A(M) / \gamma_A \). The value of the estimator relative error \( \text{RE}(M) \) is essential as it quantifies the relative error one should expect for each realization of the algorithm, and thus the quality of the result. How the estimator relative error \( \text{RE}(M) \) depends on the number of selections, on their timing, and the type observables are critical questions that we analyze in this section.

Figure 3. (a) The estimator variance \( \text{RE} = \sqrt{\frac{1}{K} \sum_{i=1}^K (\hat{\gamma}_A - \gamma_A)^2 / \gamma_A} \) where \( A = [a, +\infty) \) for different values of the threshold \( a \) and number of particles \( M \). \( \text{RE} \) is estimated from \( K = 50 \) independent runs of the genealogical particle analysis algorithm. The weight function is \( \exp(C \Delta x) \) with \( C = 4 \). For comparison: the brute force Monte Carlo estimator variance with 512 particles for a threshold \( a = 2 \) is 0.95. (b) The estimator variance for a fixed threshold \( a = 2 \) for different numbers of particles \( M \). A \( 1 / \sqrt{M} \) function is fitted and shown as the dashed line.
4.2.1. Number of particles. The result in [19] proves the existence of the central limit theorem. In order to get an estimate of \( \text{RE}(M) \), we compute it empirically by performing \( K \) independent algorithm realizations and using the estimator

\[
\text{RE}(M) \approx \frac{1}{K} \sum_{i=1}^{K} \left( \hat{\gamma}_{A,i} - \gamma_A \right)^2 / \gamma_A. \tag{15}
\]

In this formula the value of \( \gamma_A \) will be either the theoretical value when it is available, for instance for the Ornstein–Uhlenbeck process, or the estimated value of the probability by averaging \( \hat{\gamma}_{A,i} \) over \( K \) realizations. In the following, by an abuse of notation, \( \text{RE} \) denotes either the theoretical estimator variance of the estimator variance evaluated from (15), which one will be clear from the context.

We first study the estimator variance \( \text{RE} \) for the Ornstein–Uhlenbeck case. We first test whether or not the regime of the central limit theorem has been reached by changing the number of particles \( M \) and verifying whether \( \text{RE} \) (15) reduces by the corresponding \( \sqrt{M} \) factor. Figure 3(a) shows, as expected, a decrease in relative error as the number of particles is increased for a range of thresholds \( a \) (see equation (14)). The inverse square root behavior of the error with increasing number of particles is demonstrated for a fixed threshold in figure 3(b). The parameters used in the simulation are specified in the figure captions.

4.2.2. Number of selections and their timing. Little theoretical analysis has been performed on the optimal number of selection steps, making this the most heuristic choice to be made. Some numerical analysis of this issue has been performed in [42]. For the problem they investigate, changing the number of selections, with equidistant in time selections, the estimator variance clearly shows a minimum for a certain number of selections.

This behaviour can be interpreted as follows. Selections should not be performed too frequently, as cloning increases correlations between the particles and therefore reduces the
effective number of independent particles, increasing the estimator variance. If not performing selections frequently enough however, the particle distribution relaxes to the unbiased particle measure, leading to the poor brute force Monte-Carlo variance. This can be seen in figure 4:

for low thresholds \( a \), importance sampling is useless and estimations with a small number of selections have the lowest estimator relative error \( \text{RE} \). Due to the large time between selections, the particles have relaxed to the particle measure of brute force Monte Carlo simulations and therefore have a similar estimator variance. For higher thresholds, for instance for \( a > 1.7 \), it becomes advantageous to kill a larger number of particles to obtain a more tilted final particle distribution, in order to lower the variance. For the threshold value \( a = 2 \) the optimal number of interactions among the values in the figure is \( N = 16 \). For higher thresholds there is a small reduction in error by increasing the number of selections, although increasing the number of selections beyond \( N = 64 \) results in an overall increase of error.

Figure 4 also illustrates the large improvement in estimator variance for the genealogical particle analysis algorithm compared to Monte-Carlo sampling, as soon as \( a \geq 2 \).

Besides the number of selections, there also seems to be little theoretical understanding of the optimal timing of selections. One strategy to selection timing is to calculate on-the-fly a criterion on the distribution of particle weights (such as the squared coefficient of variation or entropy) and only perform selection if a fixed threshold is exceeded. The convergence of such adaptive selection strategies is discussed in [43].

**4.2.3. Estimating a range of over-threshold probabilities.** From the point of view of the estimator variance \( \text{RE} \) (15), to each value of the threshold \( a \) corresponds an optimal value of the parameter \( C \) in the weight function \( W(x, y) = \exp(C(x-y)) \), denoted \( C^*(a) \), or equivalently for each value of \( C \) the estimator variance has a minimum for a given value of \( a \), denoted

![Figure 5. (a) The estimator variance \( \text{RE} \) for different weight factors \( C \) for a range of thresholds. The brute force error is computed as \( \sqrt{\gamma_A - \gamma_A^2} / \sqrt{\gamma_M} \), where \( A = [a, +\infty) \). (b) The estimated over-threshold probability \( P(a) \) compared to the analytic result. For each value of the threshold \( a \), the estimate corresponding to the value of \( C \) with lowest estimator variance is chosen.](image-url)
For instance, Figure 4 shows that the value \( C = 4 \) is optimal for \( a \approx 2.5 = a^* \). In simple cases, we expect \( C^*(a) \) to increase monotonically with \( a \).

There is an optimal value of \( a \) for each value of \( C \), however the estimate is good for a range of thresholds around this optimum. When instead of a particular over-threshold probability one is interested in the tail of the complete distribution probability, one can perform a number of genealogical particle analysis simulations each with different value of \( C \), and select for each threshold the value corresponding to the lowest estimator variance RE. Figure 5 illustrates how the tail of \( P(a) = P(x(t) \geq \alpha \mid x(0) = 0) \) can be estimated this way, with \( x(t) \) an Ornstein–Uhlenbeck process. For large values of the threshold (above \( a \approx 4.6 \)) all estimates have a high error and the highest value of \( C \) is chosen by default. As can be seen on this figure there is very good agreement with the theoretical value up to probabilities as low as \( 10^{-10} \). Using this strategy, we can accurately estimate the tail of the over-threshold distribution down to probabilities as small as \( 10^{-10} \), with relative error lower than one.
4.2.4. Selections along the fluctuation path. We have discussed in section 2.3.4 how a time-dependent weighting function can be used. In this way the particle distribution can be weighted with different exponential factors $C(t_k)$ at different selection times $t_k$, but still lead to the same exponentially tilted final particle distribution. Furthermore, in section 3.1 we have discussed how for small noises, most of the paths leading to a rare event will concentrate around fluctuation paths that minimize the action functional. The aim of this section is to demonstrate the interest of using fluctuation paths to construct time-dependent weighting functions in order to increase the efficiency of the genealogical particle analysis algorithm.

Since the Ornstein–Uhlenbeck process is linear, taking limits of higher thresholds is equivalent to taking a weak-noise limit through a rescaling of the $x$ coordinate. Hence, for fixed noise intensity $\sigma$, paths starting at $x_0 = 0$ conditioned on reaching the final threshold $a$ will concentrate around the fluctuation paths in the limit $a \to \infty$. The action (10) for the Ornstein–Uhlenbeck process (12) is given by $I[X] = \int_0^T d\tau (\dot{X} + \lambda X)^2$. Taking as boundary conditions $X(0) = 0$ and $X(T) = a$ the fluctuation paths are easily computed to be $X(t) = a \tanh(\frac{\lambda t}{a})$.

By using the potential function $W(t, x, y) = \exp(C(t_k)x - C(t_{k-1})y)$ with a weight parameter $C(t_k)$ dependent on the selection time $t_k$, we can control $\bar{\mu}(t_k)$, the mean particle position at $t_k$, by varying $C(t_k)$. The expected particle distribution for the Ornstein–Uhlenbeck process tilted with this weighting function at the selection at $t_k$ is
as discussed in section 2.3.4. The corresponding expected mean particle position is therefore
\[ \tilde{\mu}(t) = C(t) v(t), \]
where \( v(t) = (1 - \exp(-2t))/2 \) is the variance of the Ornstein–Uhlenbeck process at time \( t \) (the solution of equation (13) with \( v(0) = 0 \)). Choosing \( C(t) = X(t)/v(t) \), \( \tilde{\mu}(t) \) equals \( X(t) \) and the algorithm particle distribution follows the fluctuation path leading to the threshold \( a \).

Figure 6 shows the effect of using a weight function based on a fluctuation path versus an exponential weight function. The bottom two plots show how using the fluctuation path significantly decreases the fraction of particles that are killed during the selection steps \( N_{1/k} \) to the total number of particles at that time step \( N_k \). This is also illustrated in the plots in the top and middle rows. The top plots show the paths from the initial state for all surviving particles at the final time. Paths that have been killed during the process are not shown. We call these paths the ancestral paths. As can be seen on the top left plot, only few trajectories from the initial stage of the algorithm are ancestors of the final positions. This is not the case for the top right plot. The algorithm using a weight based on a fluctuation path has a much larger number of ancestors. This richer ancestral tree results in a decreased estimator error for the over-threshold probabilities, as is demonstrated in figure 7.

Note that for both the exponential weighting function and the weighting based on the fluctuation path, the paths reaching the threshold follow the fluctuation path. Other paths reaching the threshold are so rare that few of them are generated, even if they are more likely to survive selection in the case of exponential weighting. Note that the killed paths, partially shown in the middle row of figure 6, tend to have a negative change in position before being killed. The higher target path in the exponential makes for a higher average dissipative force \(-x \) on the particles, leading to a large discrepancy between the actual particle distribution and the target distribution at selection times.

5. Genealogical particle analysis algorithm for a deterministic dynamical system: the Lorenz '96 model

The Lorenz '96 model is a deterministic dynamical system that is often used as a toy model in the meteorology community. It was proposed by Lorenz as part of a study on error growth and predictability for chaotic dynamical systems \([39, 40]\).

A crucial difference between the famous Lorenz '63 model and the less well-known Lorenz '96 model is that the latter has a large number of degrees of freedom. Indeed macroscopic variables of deterministic systems with a large number of degrees of freedom often behave qualitatively similar to solutions of stochastic differential equations with much less degrees of freedom. Such results can be proven for some specific types of models (featuring separation of time scale, independence, ...).

It is believed however that similar results remain true for a wide range of models and observables even though mathematical proofs are out of reach. If this conjecture is correct, then the sampling of rare events through genealogical particle analysis algorithms should be applicable to macroscopic variables of deterministic systems with a large number of degrees of freedom. In this section, we demonstrate empirically, through numerical simulation, that genealogical particle analysis algorithms can indeed efficiently sample the tail of the energy distribution of the Lorenz '96 model.
The Lorenz '96 model consists of \( L \) variables \( x_i \) on a ring \( \{ \ldots, 0, 1 \} \), with dynamics

\[
\dot{x}_i = x_{i-1}(x_{i+1} - x_{i-2}) + R - x_i,
\]

where indices \( i \) are in \( \mathbb{Z}_L \), i.e. the index \( i \) is identified with \( i \mod L \) if \( i \not\in \{ 0, \ldots, L-1 \} \). The nonlinear part of the dynamics \( x_{i-1}(x_{i+1} - x_{i-2}) \) conserves the energy \( E(x) = \frac{1}{L} \sum_{i=1}^{L} x_i^2 \), while \( R \) is a forcing and \( -x_i \) a linear dissipation. The dynamics is chaotic for \( R \geq 8 \) [39, 40].

We will estimate the probability of reaching a certain energy threshold after a time \( t \), starting from the zero vector \( x_{0,i} = 0 \forall i \). A small perturbation \( \epsilon \mathcal{N}_{0,1} \) is added to the initial particle configuration to make the trajectories diverge. For large enough times \( T \), the system will therefore relax to its physical invariant measure and it makes sense to determine probabilities of exceedances of macroscopic observables \( \gamma_{E_i} = P(E(x(T)) > E_i) \). Throughout the article we will use a number of variables \( L = 32 \) and a forcing \( R = 2^8 = 256 \).

Figure 8 shows a plot of the over-threshold probabilities of the energy of the Lorenz '96 system, estimated through a brute force Monte-Carlo simulation from randomly perturbed initial conditions. Given that we have finite computer resources at our disposal, assume we can generate at most \( M = 10^5 \) independent measurements of the energy. If the maximal relative error that we are willing to tolerate is for example 0.5 then since

\[
\text{RE} = \sqrt{\gamma_{E_i} - \gamma_{E_i}^2}/(M\gamma_{E_i}) \approx 1/(M\gamma_{E_i})
\]

the lowest probability that we can estimate is approximately \( \gamma_{E_i} = 1/M(\text{RE})^2 = 4.10^{-5} \). From figure 8 we can deduce that the
corresponding highest energy threshold obtainable lies around $E_t = 1785$. Beyond this threshold the use of rare event algorithms becomes necessary.

5.2. Algorithm implementation

We use the following settings for the genealogical particle analysis simulation. The initial condition is set to $\mathbf{x}_i \sim \mathcal{N}_0$. The total integration time per realization is $T = 1.27$. This time interval corresponds to roughly five times the decorrelation time of the energy observable.

The standard deviation of the estimator $\mathbf{\mu} = \mathbf{\sqrt{\sum_{i=1}^{K} (x_{A,i} - \gamma_A)^2}}$ is estimated from $K = 10$ independent runs of the algorithm and the truth $\gamma_A$ is taken from a long brute force Monte Carlo simulation. The number of interactions is set to 64.

5.2.1. Weight function. For simplicity, we have employed an exponential weight function $W = \exp(C\Delta E)$, where $\Delta E$ is the change in energy between two interactions. This choice doesn’t require any a priori knowledge of the dynamics and is easy to implement. This weight function is not optimal, but, as we will show, it already gives good results.

For the value of the forcing parameter $R = 2^8$ the distribution of the energy values is roughly Gaussian. One can therefore estimate the mean $\mu_E$ and the variance $\sigma_E^2$ from a brute force Monte Carlo simulation and use these values along with the reasoning of section 2.2.1 to determine an appropriate value of the exponential weighting factor $C$ in the weighting function $W = \exp(C\Delta E)$. One can then choose a value $C = \Delta\mu_E/\sigma_E^2$, where $\Delta\mu_E$ is the desired change of the mean energy of the final particle distribution.

The values of $C$ in the weighting function $W = \exp(C\Delta E)$ for the calculations presented in this section are taken as $C_r = r/(2\sigma_E)$ with $r \in \{1, 2, 3, 4\}$ and $\sigma_E$ being the standard deviation of the energy so as to increase the mean energy $\Delta\mu_E$ by steps of size $\sigma_E/2$.

5.2.2. Noise perturbation. For deterministic dynamical systems, in order for two trajectories to have different dynamics after selection, a small perturbation can be added. This can be achieved by adding for example a weak Brownian perturbation at all times, or by adding a small instantaneous perturbation to offspring at the selection times. The former approach provides a simpler mathematical framework. Indeed, the study of the noise effect would amount to the study of the stochastic differential equation properties in the weak noise limit, independent of the genealogical particle analysis algorithm. By contrast the latter approach intertwines the random perturbation with the genealogical particle analysis algorithm effects and is therefore more complicated to analyze. The latter approach, however, has the practical advantage of being computationally simpler. In this study, as we will proceed purely empirically, we have opted for the latter approach. The clones are perturbed by $\mathbf{\epsilon N}_{0,1}$, where $\mathbf{N}_{0,1}$ is a standard $L$-dimensional Gaussian random variable, i.e. the noise acts independently on all of the variables.

The small noise perturbation invariably adds an error to the estimates of the tail probabilities. To obtain a rough upper bound on the strength of the perturbation that can be added without significantly perturbing the tail, we first perform a brute force simulation with the added noise for different noise strengths and verify that the tail probabilities do not change significantly compared to the sampling error of the brute force calculation. A set of independent realizations is performed like in the brute force Monte Carlo approach, the only difference being that at the selection times $t_k$ noise perturbations are added. No selection is performed however in these runs. This way we can estimate the effect of the noise on the final time particle distribution. Figures 9(a) and (b) show that below a perturbation strength of
Figure 9. (a) The over-threshold probability \( P(E > a) \) of the energy \( E \) of the Lorenz '96 system with perturbations of varying strengths \( \epsilon \) at times \( t_k \), without performing killing and cloning, with 10 000 independent realizations. The dotted line shows the estimated \( 2\sigma \) interval for over-threshold probability of the energy of the Lorenz '96 system, as estimated from two realizations of the genealogical particle analysis algorithm with different perturbing noise strength \( \epsilon \) upon cloning.

Figure 10. (a) The empirical relative error \( RE \) for different weight factors \( C \), for a range of energy thresholds \( a \), for the Lorenz '96 model. The number of particles is \( M = 1000 \). The brute force Monte Carlo relative error (in blue) is estimated with the same number of realizations \( M \) as \( \sqrt{\overline{\epsilon}_m^2 - \epsilon^2_j / (\overline{\epsilon}_m \sqrt{M})} \) (b) The over-threshold probability tail as estimated from the genealogical particle analysis algorithm compared to a brute force reconstruction. The number of particles used is 1000 for the genealogical particle analysis simulation and 10 000 for the brute force simulation.
\( \epsilon = 0.87 \) and for thresholds higher than \( a = 1600 \), the noise does not have a significant effect on the over-threshold probabilities. More complex schemes of noise perturbation could be implemented to assure that the perturbed trajectory remains close to the attractor, for example by storing a configuration at a time point before \( t_k \), adding a small perturbation to it and evolve it up to \( t_k \) to have the perturbation relax towards the attractor.

Furthermore, after performing the genealogical particle analysis algorithm, we check that the perturbing noise intensity \( \epsilon \) is small enough by decreasing \( \epsilon \) and checking that the estimates of the over-threshold tail statistics are consistent. Figure 9(b) shows that for \( \epsilon = 0.1 \) the results remain stable upon halving the noise intensity.

### 5.2.3. Estimating the tail of the energy distribution

We use for the Lorenz ’96 model the procedure described in section 4.2.3: we increase the values of the weight parameter \( C \) and use for each threshold value the best estimate from the point of view of the empirical estimator variance. The result is shown in figure 10. As there is no analytic expression for the energy distribution tail of the Lorenz ’96 system, we use a long brute force Monte Carlo estimation as comparison. The estimator variance markedly decreases when using the genealogical particle analysis algorithm. When constructing the over-threshold probability, we see that the tail can be reliably reproduced when compared to the longer brute force calculation.

The improvements in efficiency from using a rare event simulation scheme can be quantitatively estimated from figure 10. The plot of the empirical relative error shows how for a threshold around \( a = 1800 \) a brute force Monte Carlo calculation yields a relative error of 0.5, whereas the genealogical particle analysis simulation yields a relative error of approximately 0.05. A reduction in relative error by a factor 10 is achieved. Since the brute force Monte Carlo error scales as \( 1/\sqrt{M} \), a similar reduction by a raw increase of processing power would require \( M \) to increase by a factor of 100. For higher thresholds and with more fine tuning of the selection process, a larger reduction is likely to be achievable.

### 6. Conclusion

In this paper we have addressed the use of rare event computation techniques to estimate small over-threshold probabilities of observables in deterministic dynamical systems. We have demonstrated that the genealogical particle analysis algorithms can be successfully applied to a toy model of atmospheric dynamics, the Lorenz ’96 model, as presented in section 5.1. We have furthermore used the Ornstein–Uhlenbeck system to illustrate a number of implementation issues.

The example of the Ornstein–Uhlenbeck has illustrated the importance of the choice of the objective function for the performance of the genealogical particle analysis algorithm estimator. We have shown how a time-dependent objective function based on the fluctuation path to a high threshold can greatly improve the performance of the estimator compared to a fixed-in-time objective function. Furthermore we have discussed how the number of particles and the number of selection steps influence the performance of the estimator.

For the deterministic chaotic system a complication arises in that a stochastic perturbation needs to be added to the system to make identical clones of one parent diverge and explore the system’s path space. We have demonstrated in this example how the estimates of the rare event simulation are stable for small perturbations and agree with results from brute force Monte Carlo estimations. We therefore can have confidence in the correctness of these estimates.
For the example of a deterministic chaotic system that we have studied we have not yet used the fluctuation path approach, since this would require information on the dynamics to the rare event that we a priori do not possess. This lack of knowledge can be improved by iterating the estimation procedure, where one uses estimates of an initial brute force simulation to estimate the fluctuation path, after which a genealogical particle analysis simulation based on this path can be used to estimate a higher fluctuation path, which can be used for a next iteration of the algorithm. However, the results of the straightforward implementation of the rare event simulation already shows significant improvements compared to brute force estimation.

Acknowledgments

The authors would like to thank Eric Vanden-Eijnden, Eric Simonnet, Joran Rolland, Pascal You, Bérengère Dubrulle, Francesco Ragone and Takahiro Nemoto for fruitful discussions. J W acknowledges the support of the AXA Research Fund. The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n PIOF-GA-2013-626210.

References

[1] Qiu B and Miao W 2000 J. Phys. Oceanogr. 30 2124–37
[2] Schmeits M J and Dijkstra H A 2001 J. Phys. Oceanogr. 31 3435–56
[3] Glatzmaier G A and Roberts P H 1995 Phys. Earth Planet. Inter. 91 63–75
[4] Intergovernmental Panel on Climate Change 2012 Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption: Special Report of the Intergovernmental Panel on Climate Change (New York, NY: Cambridge University Press) ISBN 978-1-107-02506-6
[5] Barrio-pedro D, Fischer E M, Luterbacher J, Trigo R M and García-Herrera R 2011 Science 332 220–4
[6] Wetter O and Pfister C 2013 Clim. Past 9 41–56
[7] de Haan L 2006 Extreme Value Theory: An Introduction (Springer Series in Operations Research) (New York; London: Springer) ISBN 0-387-23946-4
[8] Leadbetter M R 1983 Extremes and Related Properties of Random Sequences and Processes (Springer Series in Statistics) (New York: Springer)
[9] Kharin V V and Zwiers F W 2005 J. Clim. 18 1156–73
[10] Kyselý J 2002 Stud. Geophys. Geod. 46 93–112
[11] Rubino G and Tuffin B (ed) 2009 Rare event Simulation using Monte Carlo methods (Chichester, UK: Wiley)
[12] Bucklew J A 2004 An Introduction to Rare Event Simulation (New York: Springer) ISBN 0-387-20078-9 978-0-387-20078-1 1-4419-1893-0 978-1-4419-1893-2
[13] Del Moral P 2013 Mean field simulation for Monte Carlo integration (Boca Raton, FL: CRC Press) ISBN 978-1-4665-0417-2 1-4665-0417-X
[14] Del Moral P 2004 Feynman-Kac Formulae Genealogical and Interacting Particle Systems with Applications (New York, NY: Springer) ISBN 978-1-4684-9393-1 1-4684-9393-0
[15] Cérou F and Guyader A 2007 Stoch. Anal. Appl. 25 417–43
[16] Rolland J, Bouchet F and 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
[17] Rolland J and Simonnet E 2015 J. Comput. Phys. 283 541–58
[18] Heymann M and Vanden-Eijnden E 2008 Commun. Pure Appl. Math. 61 1052–117
[19] Del Moral P and Garnier J 2005 Ann. Appl. Probab. 15 2496–534
[20] Garnier J and Moral P D 2006 Opt. Commun. 267 205–14
[21] Hairer M and Weare J 2014 Commun. Pure Appl. Math. 67 1995–2021
[22] Tailleur J and Kurchan J 2007 Nat. Phys. 3 203–7
[23] Laffargue T, Lam K D N T, Kurchan J and Tailleur J 2013 J. Phys. A: Math. Theor. 46 254002
[24] Giardinà C, Kurchan J, Lecomte V and Tailleur J 2011 J. Stat. Phys. 145 787–811
[25] Lecomte V and Tailleur J 2007 J. Stat. Mech. P03004
[26] Giardinà C, Kurchan J and Peliti L 2006 Phys. Rev. Lett. 96 120603
[27] Kurchan J 2015 Physica A 418 170–88
[28] Vanden-Eijnden E and Venturoli M 2009 J. Chem. Phys. 131 044120
[29] Adams D A, Sander L M and Ziff R M 2008 Phys. Rev. Lett. 101 144102
[30] Chandler D 2005 Nature 437 640–7
[31] Noé F, Schütte C, Vanden-Eijnden E, Reich L and Weikl T R 2009 Proc. Natl Acad. Sci. 106 19011–6
[32] Metzner P, Schütte C and Vanden-Eijnden E 2006 J. Chem. Phys. 125 084110
[33] Bolhuis P G, Chandler D, Dellago C and Geissler P L 2002 Ann. Rev. Phys. Chem. 53 291–318
[34] Wolde P R t and Frenkel D 1997 Science 277 1975–8
[35] E W, Ren W and Vanden-Eijnden E 2003 J. Appl. Phys. 93 2275–82
[36] Kohn R V, Reznikoff M G and Vanden-Eijnden E 2005 J. Nonlinear Sci. 15 223–53
[37] Grafke T, Grauer R and Schäfer T 2013 J. Phys. A: Math. Theor. 46 062002
[38] Grafke T, Grauer R, Schäfer T and Vanden-Eijnden E 2014 Multiscale Model. Simul. 12 566–80
[39] Lorenz E N 1996 Predictability: A Problem Partly Solved (GARP Publication Series vol 16) (Geneva, Switzerland: WMO) pp 132–6
[40] Lorenz E N 2005 J. Atmos. Sci. 62 1574–87
[41] Vanden-Eijnden E and Weare J 2012 Commun. Pure Appl. Math. 65 1770–803
[42] El Makrini M, Jourdain B and Lelièvre T 2007 ESAIM: Math. Model. Numer. Anal. 41 189–213
[43] Del Moral P, Doucet A and Jasra A 2012 Bernoulli 18 252–78