Practical rare event sampling for extreme mesoscale weather

Robert J. Webber, David A. Plotkin, Morgan E O'Neill, et al.

COLLECTIONS

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Practical rare event sampling for extreme mesoscale weather

View Online   Export Citation   CrossMark

Robert J. Webber,[1,a] David A. Plotkin,[2] Morgan E O'Neill,[3] Dorian S. Abbot,[2] and Jonathan Weare[1]

AFFILIATIONS

[1] Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA
[2] Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA
[3] Department of Earth System Science, Stanford University, Stanford, California 94305, USA

Note: This article is part of the Focus Issue on "Rare Event Sampling Methods: Development, Analysis and Application."
[a] **Electronic mail:** rw2515@nyu.edu

ABSTRACT

Extreme mesoscale weather, including tropical cyclones, squall lines, and floods, can be enormously damaging and yet challenging to simulate; hence, there is a pressing need for more efficient simulation strategies. Here, we present a new rare event sampling algorithm called quantile diffusion Monte Carlo (quantile DMC). Quantile DMC is a simple-to-use algorithm that can sample extreme tail behavior for a wide class of processes. We demonstrate the advantages of quantile DMC compared to other sampling methods and discuss practical aspects of implementing quantile DMC. To test the feasibility of quantile DMC for extreme mesoscale weather, we sample extremely intense realizations of two historical tropical cyclones, 2010 Hurricane Earl and 2015 Hurricane Joaquin. Our results demonstrate quantile DMC's potential to provide low-variance extreme weather statistics while highlighting the work that is necessary for quantile DMC to attain greater efficiency in future applications.

When rare events are studied using simulation, it can take a long time to gather sufficient data through direct sampling. As an alternative to direct sampling, specialized rare event sampling algorithms provide data more quickly, thus reducing computational costs. Here, we present a new rare event sampling method, quantile diffusion Monte Carlo (quantile DMC), that is simple to use. Quantile DMC performs extremely well on a one-dimensional test case, accurately estimating rare event probabilities with less than one thousandth the computational cost of direct sampling. Quantile DMC could potentially be of use in complex rare event simulations, for example, simulating the frequency of tropical cyclones (TCs), mesoscale convective systems, or floods under different climate conditions. When we apply quantile DMC to simulate intense tropical cyclones, we obtain promising results: storms at high intensities are more reliably simulated using quantile DMC compared to direct sampling.

I. INTRODUCTION

A common strategy for estimating rare event probabilities is *direct sampling*.[1–3] The direct sampling approach is to repeatedly simulate data from a model and then calculate the frequency of a rare event over all the simulated data. This approach can be effective in some contexts but can also be computationally expensive. For the rarest probabilities, an exorbitant amount of computational effort might be required before the event occurs even once in the simulations. Responding to these concerns, researchers as early as the 1950s[4,5] developed specialized rare event sampling algorithms to improve computational efficiency.

Today, a diverse community of scientists uses rare event sampling and analysis tools to study processes that take place infrequently, are too complex to be described analytically, and can be simulated on a computer. For example, the following extraordinary events have all been simulated using rare event sampling: a high-energy particle penetrating a nuclear shield,[6] a life-sustaining protein-protein reaction,[7] and an extreme loss of portfolio value.[8]

A burgeoning field of research explicitly links rare event simulation and analysis tools with geophysical applications.[9–14] In a recent paper, Ragone *et al.*[12] showed that rare event sampling methods can be used to study the probability of extreme weather occurring. They sampled intense 90-day heat waves over Western Europe at a fraction of the computational expense of direct sampling. Their simulations led to the surprising insight that extreme heat waves over Western

Europe are associated with a stationary wavenumber 3 anomaly in the jet stream.

The work of Ragone and coauthors concerns *synoptic scale* weather, weather that occurs on a length scale of 1000 km or greater.[1] A more challenging question is how to apply rare event sampling techniques to *mesoscale* weather, which occurs on a smaller length scale of 10–1000 km. Extreme mesoscale weather, including tropical cyclones and floods, accounts for many of the world's most destructive natural disasters.[15,16] Yet, simulations of mesoscale weather can demand enormous computational resources due to the need for high spatial resolution.[17,18] Mesoscale weather simulation presents a unique challenge for rare event sampling and analysis, where the need for efficient simulation strategies is great and yet sample size is highly limited due to computational expense.

In Plotkin *et al.*,[14] we present a rare event analysis strategy for potential use in extreme mesoscale weather simulations. Using a computationally efficient algorithm, we identify maximum likelihood perturbations that lead to the occurrence of an intense tropical cyclone in a high-resolution weather model. In particular, we identify key changes in wind, temperature, and relative humidity fields that help explain the rapid intensification process in modeled tropical cyclones.

In contrast to Plotkin *et al.*,[14] which analyzes the single most likely path toward extreme mesoscale weather, the present paper analyzes *statistics* of extreme mesoscale weather. Accurate estimation of statistics can require sampling numerous possible paths. To achieve this goal, therefore, it is appropriate to use a rare event sampling algorithm.

Here, we present a new rare event sampling algorithm called quantile diffusion Monte Carlo (quantile DMC) that is suited for complex real-world applications such as extreme weather simulations. The algorithm is simple to implement, yet suitable for a large class of nonlinear processes. When we apply quantile DMC to a simple example, the algorithm is more efficient than direct sampling by a factor of more than a thousand.

Quantile DMC is a "splitting" algorithm, an algorithm in which some simulations are split into multiple replicas to promote progress toward the rare event of interest and other simulations are "killed." A key advantage of splitting algorithms is that they are practical to implement. Our simulations with quantile DMC are simple to code and require the same computational cost as direct sampling from the dynamical model. In contrast, alternative rare event sampling approaches can be more challenging to implement, because they require modifying the underlying dynamical model or frequently starting and stopping the dynamics. For mesoscale weather models like the one simulated in the current paper, these manipulations would require extensive code development or substantial additional computational cost.

Quantile DMC is inspired by a previous splitting algorithm called diffusion Monte Carlo,[4,5,19–22] but it incorporates two new features. First, in diffusion Monte Carlo, splitting is typically uniform in time, but in quantile DMC the intensity and frequency of splitting increase over time, thereby improving efficiency. Second, quantile DMC adaptively makes use of data from simulations, so that the algorithm requires less tuning as compared to diffusion Monte Carlo.

We envision that quantile DMC could be used to study the frequency of extreme mesoscale weather under different climate conditions. While a full application of rare event sampling techniques to study extreme weather is beyond the scope of the current paper, we test the feasibility of our approach using a high-resolution weather model to simulate tropical cyclones. We apply quantile DMC to study the upper tail of the intensity distribution for numerical simulations of two historical tropical cyclones: 2010 Hurricane Earl and 2015 Hurricane Joaquin. Using an ensemble of $N = 100$ simulations, quantile DMC produces more than seven times as many high-intensity Category 5 realizations for both of the storms compared to direct sampling. Moreover, the variance of important rare event statistics is improved by a factor of two to ten. Building on this success, we anticipate that we can improve the performance of this method in the future.

This paper is organized into two major sections. Section II presents various approaches to estimating rare event probabilities: direct sampling, diffusion Monte Carlo, and the new method quantile DMC. Section III examines the potential role of rare event sampling in extreme weather and presents tropical cyclone simulations.

II. ESTIMATING RARE EVENT PROBABILITIES

In this section, we introduce several approaches to estimating rare event probabilities for potential use in climate and weather applications. First, we discuss the well-known method of direct sampling. Then, we describe a rare event sampling algorithm, called diffusion Monte Carlo (DMC). We illustrate the advantages and disadvantages of DMC on a simple example. Then, we introduce quantile DMC and explain why it gives more robust performance compared to standard DMC. Lastly, we discuss the implementation of quantile DMC in practical settings.

A. Direct sampling

While direct sampling is a useful tool for studying the typical behavior of complex or high-dimensional systems, it is not an ideal approach for investigating unlikely or infrequent phenomena. In particular, as we will demonstrate, direct sampling can give very high error when estimating statistics of rare events.

Direct sampling uses a straightforward approach to estimate probabilities. We assume that X is a random process, A is an important event (e.g., the occurrence of an intense tropical cyclone), and we can draw independent samples of X, labeled as $\xi^{(1)}, \xi^{(2)}, \xi^{(3)}, \ldots, \xi^{(N)}$. To estimate the probability $p = \mathrm{P}\{X \in A\}$, direct sampling uses

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{\xi^{(i)} \in A\right\}. \tag{1}$$

This estimator is called the *sample average*.

To assess the error in direct sampling, we calculate the mean and variance of $\hat{p}$

$$\begin{cases} \mathrm{E}\left[\hat{p}\right] = p, \\ \mathrm{Var}\left[\hat{p}\right] = \frac{1}{N} p \left(1 - p\right). \end{cases} \tag{2}$$

On the surface, the variance of the sample average $\hat{p}$ would appear to be quite good. In particular, the variance depends only on p

and not on the process X. The process X can have millions of dimensions or even infinite dimensions and still the variance of $\hat{p}$ converges to zero at a $1/N$ rate as $N \to \infty$.

Surprisingly then, direct sampling estimates $\hat{p}$ can have unacceptably high error in rare event calculations. For example, suppose $p = 0.01$ is the probability of a rare event A, and $N = 100$ is the sample size of simulations. Then, $\hat{p}$ may take the value $\hat{p} = 0$ with probability 0.37, the value $\hat{p} = 0.01$ with probability 0.37, the value $\hat{p} = 0.02$ with probability 0.18, the value $\hat{p} = 0.03$ with probability 0.06, and higher values with probability 0.02. The error in $\hat{p}$ is overwhelming. If the estimate $\hat{p}$ is used for risk analysis, then the error in $\hat{p}$ might have harmful practical consequences.

How can direct sampling produce estimates that are simultaneously so good and so bad? In most applications, what is important is not absolute error $\hat{p} - p$, but rather *relative error* $(\hat{p} - p)/p$. To assess the relative error in direct sampling, we calculate the mean and variance of $\hat{p}/p$

$$\begin{cases} \mathrm{E}\left[\hat{p}/p\right] = 1, \\ \mathrm{Var}\left[\hat{p}/p\right] = \frac{1-p}{Np}. \end{cases} \qquad (3)$$

To estimate a probability p with even one digit of precision, $\mathrm{Var}\left[\hat{p}/p\right]$ must be many times smaller than 1, and this requires a sample size N that is many times larger than $1/p$. When models are expensive to run and probabilities p are small, obtaining these large sample sizes is not a practical option.

Rare event sampling methods can help address the deficiency of direct sampling in estimating rare probabilities. For example, in carefully designed rare event splitting algorithms similar to diffusion Monte Carlo and quantile DMC it can suffice to increase sample size N as slowly as $N \sim \log\left(1/p\right)$ as $p \to 0$ to achieve fixed relative error.[23] This is an exponential improvement compared to direct sampling, where it is necessary to increase sample size N at a rate of $N \sim 1/p$ to achieve fixed relative error. The exponential improvement due to rare event sampling methods can make possible very precise calculations of rare event probabilities even with a limited ensemble size N.

B. Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) is a sampling algorithm that causes simulations to explore regions of state space that would rarely be accessed under typical conditions. The earliest antecedents of DMC were splitting algorithms invented in the 1950s.[4,5] In the 1960s, DMC was popularized in the quantum chemistry community where researchers used DMC to obtain information about the ground state energy of the Schrödinger equation for chemical systems.[19,20] In the 2000s, the tools of DMC were increasingly applied to rare event sampling, and the algorithm became known in some circles as "genealogical particle analysis."[24,25] Recently, DMC has been the subject of a series of mathematical analyses, which describe the convergence and asymptotic behavior of DMC as the ensemble size N approaches infinity.[26,27] Hairer and Weare[22] provide a more detailed history of DMC.

For a simple example of DMC, assume $(X_t)_{t\geq 0}$ is a Markov process in $\mathbb{R}^d$, and $\xi_t^{(1)}, \xi_t^{(2)}, \ldots, \xi_t^{(N)}$ are simulations of X_t, which are

called "particles." To estimate the probability of a rare, important event A, the DMC algorithm iterates the following steps:

1. Evolve particles $\left(\xi_t^{(i)}\right)_{1\leq i\leq N}$ forward from time t to a later time t'.
2. Using a consistent set of rules, randomly "split" particles $\xi_{t'}^{(i)}$ that have moved much closer to A and randomly "kill" particles $\xi_{t'}^{(i)}$ that have moved much farther from A, making sure that the total number of particles N remains unchanged.

DMC uses splitting and killing to cause a greater number of particles to reach the rare event state A, compared to direct sampling.

In greater generality, the DMC algorithm is guided by one-dimensional coordinate $\theta \colon \mathbb{R}^d \to \mathbb{R}$ that is high in some regions of the state space $\mathbb{R}^d$ and low in other regions of the state space. Where θ is high, DMC exhibits a greater propensity toward splitting. Where θ is low, DMC exhibits a greater propensity toward killing. The coordinate θ is often known as an *order parameter* or *reaction coordinate*. The particular choice of reaction coordinate can be crucial to the efficiency of DMC for computing rare event statistics.

A basic schematic of DMC is given in Fig. 1. In this schematic, the reaction coordinate is the position $\theta(x) = x$. Therefore, splitting and killing of simulations drives the process toward high values of x.

To implement DMC on a computer, a sequence of actions are required. The user defines a series of resampling times $0 = t_0 < t_1 < t_2 < \cdots$. For each resampling time t_k, the user specifies $V_k(x)$, a splitting function that increases with the reaction coordinate $\theta(x)$. DMC begins with an initialization step and then iterates over reweighting, resampling, and mutation steps according to the following definition:

Definition II.1 (Diffusion Monte Carlo).

1. *Initialization*: Independently sample initial particles $\xi_0^{(i)} \sim Law(X_0)$ for $1 \leq i \leq N$
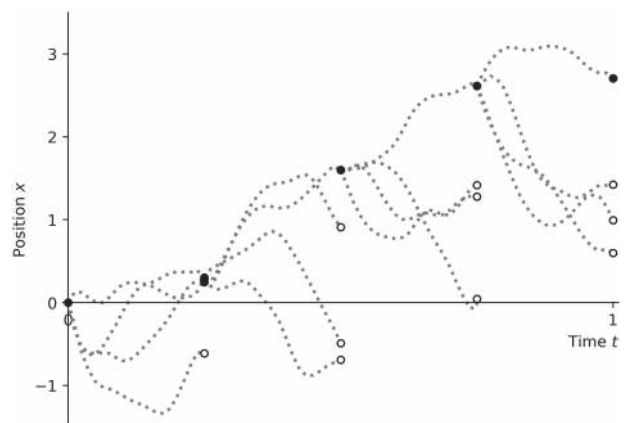2. *For $k = 0, 1, 2, \ldots,$*

FIG. 1. Illustration of diffusion Monte Carlo. At fixed times t, some simulations are killed (white circles). Simulations that are not killed (black circles) are possibly replicated, and all simulations are run forward in time. Splitting and killing create a net flux, driving simulations toward high values of x.

(a) *Reweighting*: If $k = 0$, define initial weights

$$w_0^{(i)} = \exp\left\{V_0\left(\xi_0^{(i)}\right)\right\}.\qquad(4)$$

If $k > 0$, define weights

$$w_k^{(i)} = \overline{w}_{k-1}\exp\left\{V_k\left(\xi_k^{(i)}\right) - V_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}.\qquad(5)$$

Define the average weight $\overline{w}_k = \frac{1}{N}\sum_{i=1}^{N} w_k^{(i)}$.

(b) *Resampling*: By splitting and killing particles $\left(\xi_k^{(i)}\right)_{1\leq i\leq N}$, create an ensemble of updated particles $\left(\hat{\xi}_k^{(i)}\right)_{1\leq i\leq N}$ consisting of $N_k^{(i)}$ copies of each particle $\xi_k^{(i)}$. The numbers $N_k^{(i)}$ are randomly chosen to satisfy

$$\begin{cases}\displaystyle\sum_{i=1}^{N} N_k^{(i)} = N,\\[2mm] E\left[N_k^{(i)}\right] = w_k^{(i)}/\overline{w}_k.\end{cases}\qquad(6)$$

(c) *Mutation*: Independently sample $\xi_{k+1}^{(i)} \sim Law(X_{t_{k+1}}|X_{t_k} = \hat{\xi}_k^{(i)})$ for $1 \leq i \leq N$.

3. *Estimation*: To approximate $E\left[f\left(X_{t_k}\right)\right]$, DMC uses the estimate

$$E\left[f\left(X_{t_k}\right)\right] \approx \frac{\overline{w}_{k-1}}{N}\sum_{i=1}^{N}\frac{f\left(\xi_k^{(i)}\right)}{\exp\left\{V_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}}.\qquad(7)$$

The intialization and mutation steps in DMC are straightforward, but the reweighting and resampling steps require further elaboration. In the reweighting step, splitting/killing weights $\left(w_k^{(i)}\right)_{1\leq i\leq N}$ are defined by means of the splitting functions V_k. The simplest example of a splitting function is $V_k(x) = C\theta(x)$, where $C > 0$ is a splitting parameter that controls how many times a single particle can be split to create new copies.

In the resampling step, particles are split and killed according to the weights $\left(w_k^{(i)}\right)_{1\leq i\leq N}$. Random numbers $N_k^{(i)}$ indicate how many times each particle $\xi_k^{(i)}$ is copied. The random numbers $N_k^{(i)}$ have expectation

$$E\left[N_k^{(i)}\right] = w_k^{(i)}/\overline{w}_k.\qquad(8)$$

In this formula, $w_k^{(i)}$ is divided by $\overline{w}_k$ to ensure that the total number of particles satisfies $E\left[\sum_{i=1}^{N} N_k^{(i)}\right] = N$. To define the particular distribution for the random numbers $\left(N_k^{(i)}\right)_{1\leq i\leq N}$, we use a low-variance resampling scheme called *sorted stratified resampling*,[27,28] the details of which we describe in the Appendix. While a good choice of resampling scheme can slighly reduce DMC error, other factors determine a greater share of the error in DMC estimates. The splitting functions V_k are the most important parameters for determining the dynamics of DMC.

The validity of DMC estimates is supported by mathematical analyses.[26,27] DMC estimates are unbiased and converge as the number of particles N tends to infinity under mild integrability conditions. These theoretical results are very general, holding true for systems with arbitrarily high dimension d. Any quantity that can be estimated by direct sampling can also be estimated by DMC. Estimates can include functions that depend on the entire path from time 0 until a later time t_k.

Analysis of DMC supports the conclusion that DMC oversamples regions, where the reaction coordinate θ is large and undersamples regions where θ is small. In particular, the distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta\left(\hat{\xi}_k^{(i)}\right)$ converges weakly as $N \to \infty$ to the distribution of X_{t_k} weighted by a likelihood ratio of

$$L_k(x) = \frac{\exp\{V_k(x)\}}{E\left[\exp\left\{V_k\left(X_{t_k}\right)\right\}\right]}.\qquad(9)$$

Since $V_k(x)$ increases with the reaction coordinate $\theta(x)$, more particles occupy regions where θ is high, compared to direct sampling.

In summary, we have defined diffusion Monte Carlo and presented two key facts. First, DMC provides unbiased, convergent estimates. Second, at each resampling step t_k, DMC moves particles to regions where the reaction coordinate θ is large.

C. Strengths and weaknesses of DMC: The Ornstein-Uhlenbeck example

The Ornstein-Uhlenbeck (OU) process is a one-dimensional, linear process, which we use to illustrate the strengths and weaknesses of DMC. On the one hand, DMC can effectively sample rare extreme deviations of the OU process. On the other hand, when sampling transformations of the OU process, DMC can require delicate tuning, which limits the practical effectiveness of the algorithm.

The OU process evolves under the dynamics

$$dX_t = -\alpha X_t dt + \sqrt{2\alpha}dW_t,\qquad(10)$$

where $\alpha > 0$ is a constant. We present three properties of the OU process for later reference:

1. From any starting distribution, the OU process converges geometrically to an equilibrium distribution of $N(0,1)$. Therefore, it is a rare event for the OU process to reach a position much larger than the standard deviation of 1 at a large time T.
2. From any starting position $X_0 = x$, the mean of the OU process converges geometrically to 0 at a time scale of $1/\alpha$; that is, $E[X_t] = xe^{-\alpha t}$.
3. From any starting position $X_0 = x$, the variance of the OU process converges geometrically to 1 at a time scale of $1/(2\alpha)$; that is, $\mathrm{Var}[X_t] = 1 - e^{-2\alpha t}$.

We consider using DMC to estimate the probability that the OU process starting from $X_0 = 0$ exhibits a rare extreme deviation $X_1 \geq U$. To set up the DMC algorithm, a natural choice of reaction coordinate is position $\theta(x) = x$. We must also choose a series of resampling times $0 < t_1 < \cdots < t_{K-1} < t_K = 1$ and splitting functions V_k. The simplest choice is to set $t_k = k/K$ for $k = 1, 2, \ldots, K$ and $V_k(x) = C\theta(x)$, where $C > 0$ is a positive number. We refer to this strategy as "time-homogeneous" DMC, since the splitting intensity and splitting frequency are uniform in time.
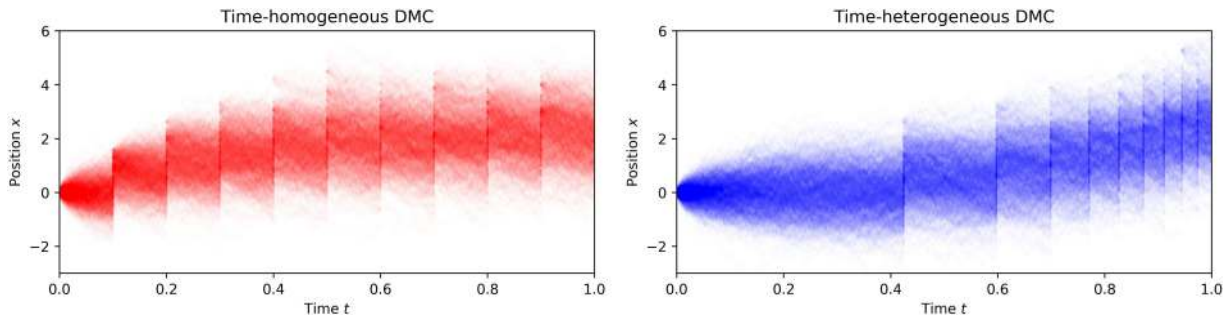
FIG. 2. In time-homogeneous DMC (left) particles reach high positions x more quickly than in time-heterogeneous DMC (right). Both methods are implemented with $N = 1000$ particles, tilting constant $C = 2.5$, and dynamics $dX_t = -2X_t dt + 2dW_t$.

As an alternative to time-homogeneous DMC, we also consider a "time-heterogeneous" resampling strategy. Since the OU process loses its memory exponentially quickly, we observe that random motion of particles at early times is not as important as random motion of particles at later times for determining final locations at time 1. Motivated by this observation, we can define resampling times $0 < t_1 < \cdots < t_{K-1} < t_K = 1$ using the formula

$$\int_0^{t_1} e^{2\alpha t} dt = \int_{t_1}^{t_2} e^{2\alpha t} dt = \cdots = \int_{t_{K-1}}^{t_K} e^{2\alpha t} dt. \tag{11}$$

We can also define splitting functions $V_k(x) = Ce^{\alpha(t_k-1)}\theta(x)$. In this time-heterogeneous resampling strategy, strength and frequency of splitting increase exponentially with time. Splitting strength increases at the $1/\alpha$ time scale with which the OU process mean reverts to zero. Splitting frequency increases at the $1/(2\alpha)$ time scale with which the OU process variance reverts to 1. We note that a similar suggestion to increase the strength of splitting appears in the work of Wouters and Bouchet.[25] The suggestion to increase the frequency of splitting is newly presented here.

Figure 2 contrasts the different qualitative behavior of time-homogeneous DMC, shown in red, and time-heterogeneous DMC, shown in blue. In time-homogeneous DMC, the distribution of particles is immediately shifted toward high positions x in the time interval $[0, 1/(2\alpha)]$; in time-heterogeneous DMC, on the other hand, the distribution of particles is shifted toward high positions x at a later time interval $[1 - 1/\alpha, 1]$.

DMC can be used to estimate the probability $p = P\{X_1 \geq U\}$ that the Ornstein-Uhlenbeck process exceeds a threshold U at time 1. Following Definition II.1, estimates take the form

$$\hat{p} = \frac{\overline{w}_{K-1}}{N} \sum_{i=1}^{N} \frac{\mathbb{1}\left\{\xi_K^{(i)} \geq U\right\}}{\exp\left\{V_{K-1}\left(\hat{\xi}_{K-1}^{(i)}\right)\right\}}. \tag{12}$$

We simulated rare extreme deviations of the Ornstein-Uhlenbeck process ten thousand times using DMC with a splitting intensity of $C = 2.5$. We then computed estimates $\hat{p}$ and assessed error using the relative standard error $\sqrt{\text{Var}\left[\hat{p}/p\right]}$. The results are shown in Table I. The table shows for a range of U values that time-heterogeneous DMC is more accurate than time-homogeneous DMC. Moreover,

when studying the most extreme rare events, time-heterogeneous DMC is more than 50 times more accurate than direct sampling. Thus, with direct sampling, it would be necessary to increase the sample size by a factor of more than a thousand to obtain comparable error to time-heterogeneous DMC.

Having discussed the strengths of diffusion Monte Carlo, we now turn to a discussion of the method's shortcomings. Under the best of conditions, diffusion Monte Carlo is a highly effective rare event sampling strategy. However, to sample a transformation of the OU process, DMC can require delicate tuning, which limits the method's practical appeal.

Consider using DMC to sample extreme deviations of the process

$$\begin{cases} d\log(Y_t/4) = -\alpha \log(Y_t/4)\, dt + \sqrt{\alpha/8}\, dW_t, \\ Y_0 = 4. \end{cases} \tag{13}$$

The process Y_t is a nonlinear transformation of the OU process, with $Y_t = 4\exp\{X_t/4\}$. Figure 3 illustrates what can happen when DMC is used to sample Y_t without a careful tuning of parameters. We apply DMC with a reaction coordinate $\theta(y) = y$ and splitting functions $V_k(y) = 2.5e^{\alpha(t_k-1)}\theta(y)$. The resulting DMC scheme performs well at the first resampling time, but as soon as the first particles reach positions $y > 10$, the algorithm becomes unbalanced. Particles with the highest positions y are split into dozens or hundreds of replicas. Thus, particle positions y become highly correlated, leading to volatile and error-prone estimates for rare event probabilities.

TABLE I. Relative standard errors. With 1000 particles, time-heterogeneous DMC gives better estimates for tail probabilities than time-homogeneous DMC or direct sampling.

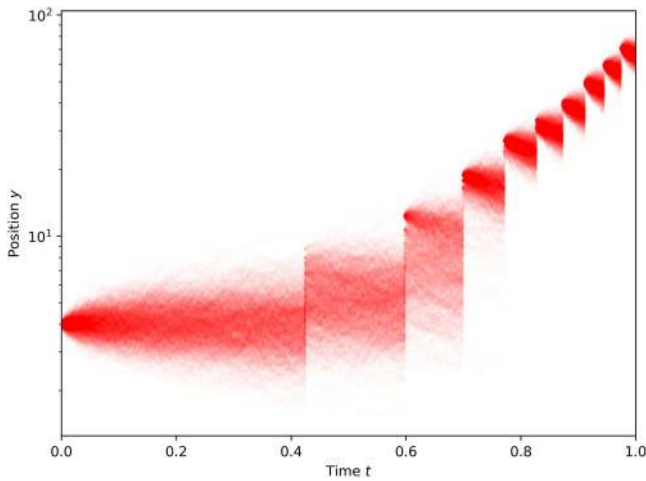|	Direct sampling	Time-homogeneous DMC	Time-heterogeneous DMC
$P\{X_1 \geq 1\}$	0.073	0.26	0.096
$P\{X_1 \geq 2\}$	0.21	0.25	0.11
$P\{X_1 \geq 3\}$	0.90	0.32	0.17
$P\{X_1 \geq 4\}$	6.1	0.69	0.37
$P\{X_1 \geq 5\}$	67	2.5	1.3

FIG. 3. With the wrong splitting functions V_k, DMC can experience a catastrophic instability. Extreme splitting leads to high correlations between particles and error-prone estimates.

While DMC can potentially be tuned to efficiently sample the process Y_t, the tuning process requires great care and flexibility. In particular, no splitting function of the form $V_k(y) = Ce^{\alpha(t_k-1)}\theta(y)$ efficiently samples extreme deviations of Y_t. A splitting function of a different parametric form is required. Moreover, tuning becomes much more difficult when sampling a process with unknown dynamics. Large amounts of data are necessary to tune DMC, and gathering the necessary data from a complex model can be computationally expensive.

We draw two essential observations from the Ornstein-Uhlenbeck example. First, DMC is most effective when the strength and frequency of splitting increase over time. Second, DMC is quite sensitive to the particular splitting functions V_k that are used. This sensitivity potentially compromises the real-world performance of DMC and prompts the development of a more robust version of DMC.

D. Quantile DMC

Quantile DMC is an elaboration of the DMC algorithm algorithm, with additional adaptation steps that make the scheme more robust. The current section describes how adaptation steps are performed and explores the specific theoretical properties that explain the robustness of quantile DMC.

The key difference between quantile DMC and standard DMC is that quantile DMC adaptively rescales the reaction coordinate θ to match a target distribution ν_k. It is the rescaled reaction coordinate θ'_k that is used for splitting and killing of simulations.

To perform quantile DMC, first define a series of resampling times $0 = t_0 < t_1 < t_2 < \cdots$. For each resampling time t_k, specify a target distribution ν_k for the rescaled reaction coordinate θ'_k and specify $V'_k(x)$, a splitting function that increases with $\theta'_k(x)$. Quantile DMC begins with an initialization step and then iterates over adaptation, reweighting, resampling, and mutation steps according to the following definition:

Definition II.2 (Quantile DMC).

1. *Initialization*: Independently sample initial particles $\xi_0^{(i)} \sim Law(X_0)$ for $1 \le i \le N$
2. *For $k = 0, 1, 2, \ldots$,*
   (a) *Adaptation*: If $k = 0$, let γ_0 be a transport function from $\frac{1}{N}\sum_{i=1}^N \delta\left(\theta\left(\xi_0^{(i)}\right)\right)$ to ν_0. If $k > 0$, let γ_k be a transport function from

$$\frac{\sum_{i=1}^N \exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}\delta\left(\theta\left(\xi_k^{(i)}\right)\right)}{\sum_{i=1}^N \exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}} \quad (14)$$

   to ν_k. Define the rescaled reaction coordinate $\theta'_k = \gamma_k(\theta)$.
   (b) *Reweighting*: If $k = 0$, define initial weights

$$w_0^{(i)} = \exp\left\{V'_0\left(\xi_0^{(i)}\right)\right\}. \quad (15)$$

   If $k > 0$, define weights

$$w_k^{(i)} = \overline{w}_{k-1}\exp\left\{V'_{t_k}\left(\xi_k^{(i)}\right) - V'_{t_{k-1}}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}. \quad (16)$$

   Define the average weight $\overline{w}_k = \frac{1}{N}\sum_{i=1}^N w_k^{(i)}$.
   (c) *Resampling*: By splitting and killing particles $\left(\xi_k^{(i)}\right)_{1\le i\le N}$, create an ensemble of updated particles $\left(\hat{\xi}_k^{(i)}\right)_{1\le i\le N}$ consisting of $N_k^{(i)}$ copies of each particle $\xi_k^{(i)}$. The numbers $N_k^{(i)}$ are randomly chosen to satisfy

$$\begin{cases} \sum_{i=1}^N N_k^{(i)} = N, \\ E\left[N_k^{(i)}\right] = w_k^{(i)}/\overline{w}_k. \end{cases} \quad (17)$$

   (d) *Mutation*: Independently sample $\xi_{k+1}^{(i)} \sim Law(X_{t_{k+1}}|X_{t_k} = \hat{\xi}_k^{(i)})$ for $1 \le i \le N$.
3. *Estimation*: To approximate $E\left[f\left(X_{t_k}\right)\right]$, Quantile DMC uses the estimate

$$E\left[f\left(X_{t_k}\right)\right] \approx \frac{\overline{w}_{k-1}}{N}\sum_{i=1}^N \frac{f\left(\xi_k^{(i)}\right)}{\exp\left\{V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}}. \quad (18)$$

Quantile DMC is distinguished from standard DMC by an adaptation step. The adaptation step begins by estimating the distribution of $\theta\left(X_{t_k}\right)$ using data from simulations. At time t_0, the distribution of $\theta(X_0)$ is estimated using $\eta_0 = \frac{1}{N}\sum_{i=1}^N \delta\left(\theta\left(\xi_0^{(i)}\right)\right)$. At later times, the distribution of $\theta\left(X_{t_k}\right)$ is estimated using

$$\eta_k = \frac{\sum_{i=1}^N \exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}\delta\left(\theta\left(\xi_k^{(i)}\right)\right)}{\sum_{i=1}^N \exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}}. \quad (19)$$

After estimating the distribution of $\theta\left(X_{t_k}\right)$, quantile DMC builds a transformation $\theta'_k = \gamma_k(\theta)$ so that the distribution of $\theta'_k\left(X_{t_k}\right)$

approximates a target distribution ν_k. In particular, quantile DMC builds a transport function[29] from η_k to ν_k of the form

$$\gamma_k(x) = F_{\nu_k}^{-1}\left(F_{\eta_k}(x)\right). \tag{20}$$

Here, F_{η_k} is a distribution function for η_k, defined by

$$F_{\eta_k}(x) = \eta_k((-\infty, x)) + \frac{1}{2}\eta_k(\{x\}) \tag{21}$$

and $F_{\nu_k}^{-1}$ is a quantile function for ν_k, defined by

$$F_{\nu_k}^{-1}(\alpha) = \inf\left\{x \in \mathbb{R}: F_{\nu_k}(x) \geq \alpha\right\}. \tag{22}$$

Since the transport function γ_k maps the quantiles of η_k to the quantiles of ν_k, we call this algorithm quantile DMC.

An explicit example of quantile DMC helps illustrate the main features of this new algorithm. Consider using quantile DMC to estimate the probability of extreme deviations of the process

$$\begin{cases} d\log(Y_t/4) = -\alpha \log(Y_t/4)\,dt + \sqrt{\alpha/8}\,dW_t, \\ Y_0 = 4. \end{cases} \tag{23}$$

This is the same nonlinear transformation of the OU process that was responsible for a catastrophic failure of DMC in Sec. II C. To sample extreme deviations of the process Y_t, we define a reaction coordinate $\theta(y) = y$ and target distributions $\nu_k = N(0, 1)$. We use splitting functions

$$V_k'(x) = 2.5e^{\alpha(t_k-1)}\theta_k'(x). \tag{24}$$

Figure 4 presents results of these quantile DMC simulations. The behavior of quantile DMC is highly stable. Particles are nudged gently but forcibly in the direction of high θ values. Explicit error calculations confirm that quantile DMC is just as effective at computing tail probabilities for this nonlinear transformation of the OU process as for the OU process itself.
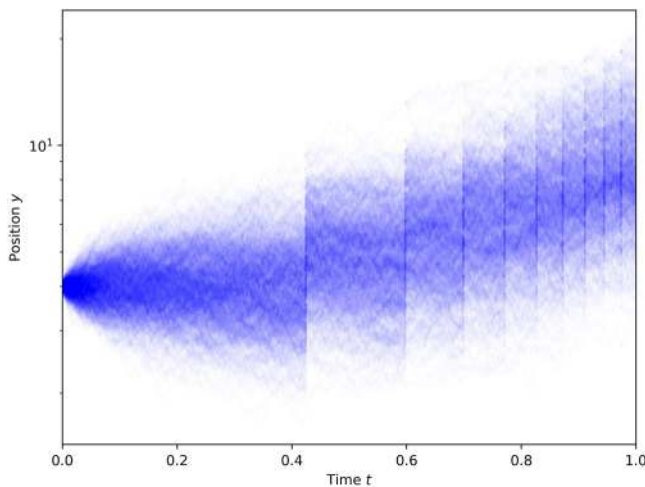
The main advantage of quantile DMC, compared to DMC, is that quantile DMC requires less tuning when it is applied to a wide class of nonlinear processes. This robustness is due to the fact that quantile DMC estimates have the same distribution if the reaction coordinate θ is replaced with any other reaction coordinate $\tilde{\theta}$ that is a monotonic, one-to-one transformation of θ. Thus, for example, quantile DMC is equally effective when sampling from the OU process or from a nonlinear, montonic transformation of the OU process. We note that this property of *invariance under monotonic transformations* is also shared by Adaptive Multilevel Splitting[30] and Steered Transition Path Sampling[31] and thus appears to be an important property underlying the success of a variety of effective rare event sampling methods. We refer the reader to the Appendix for a proof of this theoretical property and further discussion of convergence properties for quantile DMC estimates.

E. Implementation of quantile DMC

Quantile DMC provides a method to estimate rare event probabilities with much reduced computational effort compared to direct sampling. For example, Fig. 5 compares tail probabilities estimated using quantile DMC with $N = 1000$ particles to the same tail probabilities estimated using direct sampling with $N = 1\,000\,000$ particles. When sampling from the OU process, quantile DMC achieves better accuracy than direct sampling but uses one thousand times less computational power.

Quantile DMC is straightforward to implement. Whereas some rare event sampling algorithms can require additional simulation time, additional storage, or additional manipulations of the underlying dynamical model compared to direct sampling, this is not the case with quantile DMC. The only additional cost of using quantile DMC is the cost of resampling, and the cost of resampling is often negligible compared to the cost of running complex simulations forward in time. For the tropical cyclone simulations presented in Sec. III,
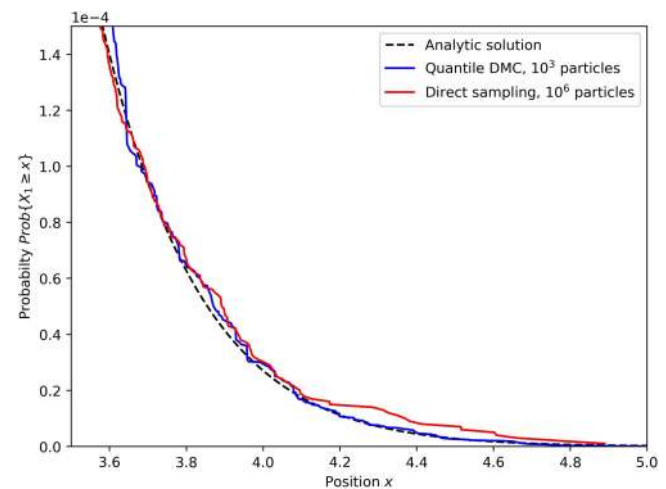

FIG. 4. Quantile DMC efficiently samples extreme deviations of the nonlinear process $d\log(Y_t/4) = -2\log(Y_t/4)\,dt + 1/2dW_t$.

FIG. 5. Extreme tail probabilities obtained using quantile DMC with $N = 1000$ particles are more accurate than tail probabilities obtained using direct sampling with $N = 1\,000\,000$ particles.

resampling was completed with ten lines of code and a few seconds of processing time.

To maximize the efficiency of quantile DMC, parameters should be adjusted depending on the particular rare event sampling problem being investigated. To illustrate this tuning process, we investigate optimal parameter choices for sampling from the OU process with resampling times, target distributions, and splitting functions

$$\begin{cases} \int_0^{t_1} e^{2\alpha t}dt = \int_{t_1}^{t_2} e^{2\alpha t}dt = \cdots = \int_{t_{K-1}}^{t_K} e^{2\alpha t}dt, \\ \nu_k = N(0,1), \\ V_k'(x) = Ce^{\alpha(1-t_k)}\theta_k'(x). \end{cases} \quad (25)$$

With this resampling schedule and this approach to splitting and killing particles, four parameters can impact the quality of quantile DMC estimates: the time scale parameter α, the tilting constant C, the number of resampling times K, and the number of particles N.

The simplest parameter to analyze is the time scale parameter α. Ideally, the parameter α should be the same time scale as the underlying Ornstein-Uhlenbeck process. However, if the parameter α is overestimated or underestimated by a factor of two, we found in our experiments that error increases by less than 20%.

To ensure quantile DMC's effectiveness, the tilting constant C must be adjusted depending on the rareness of the probabilities being investigated. When estimating a tail probability $P\{X_1 \geq U\}$ for the OU process, the optimal tilting constant C lies within one or two units of U. This numerical result is consistent with the fact that particles at time $t = 1$ are approximately normally distributed with mean C and variance 1, and the $N(C,1)$ distribution is a suitable importance sampling distribution for estimating $P\{X_1 \geq U\}$ when U lies close to C.

Parameters K and N should also be increased as probabilities being investigated become rarer. For example, when estimating $P\{X_1 \geq 2\}$, 100 particles and 10 resampling times yield near-maximal efficiency. When estimating $P\{X_1 \geq 4\}$, 1000 particles and 100 resampling times are required for near-maximal efficiency. In the second situation, the probability being estimated is rarer and consequently more computational power is required.

The efficiency gains from using quantile DMC instead of direct sampling become most dramatic when the number of particles N exceeds a critical threshold. Figure 6 shows how quantile DMC error decays quickly, at a faster than $N^{-1/2}$ rate as N is increased from 10 to 1000. Once the ensemble size reaches $N = 1000$ particles, then error decreases less quickly, at an asymptotic $N^{-1/2}$ rate.

The tuning of parameters α, C, K, and N must be coupled with a careful selection of reaction coordinate θ to guarantee the effectiveness of quantile DMC. The goal of quantile DMC is to emphasize paths leading to a rare event A via splitting and to deemphasize paths leading away from A through random killing. Thus, the ideal reaction coordinate θ should anticipate what paths lead to the rare event A. One specific reaction coordinate that is appropriate for this goal is therefore the conditional probability function

$$\theta_k(x) = P(A \mid X_{t_k} = x). \quad (26)$$

Here, the reaction coordinate θ_k changes at each resampling time t_k. This reaction coordinate is proven to be optimal for a splitting algorithm similar to quantile DMC.[23]
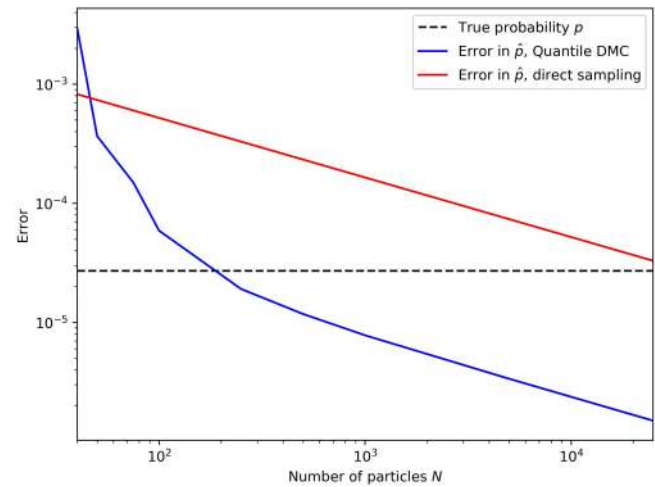

FIG. 6. Error in calculating $p = P\{X_1 \geq 4\}$. In quantile DMC, error decays quickly for small N and then levels off asymptotically to a $N^{-1/2}$ scaling. In direct sampling, error decays with a perfect $N^{-1/2}$ scaling.

With quantile DMC, identifying a reaction coordinate can be easier than with DMC because any monotonic, time-dependent transformation of $P(A \mid X_{t_k} = x)$ serves equally well as a reaction coordinate. However, in a complex, high-dimensional system, it can be challenging even to approximate a monotonic transformation of $P(A \mid X_{t_k} = x)$. In our analysis of tropical cyclone simulations in Sec. III, we conclude that identifying an appropriate reaction coordinate requires careful data analysis and scientific insight into the model being simulated.

With an imperfect choice of reaction coordinate, quantile DMC users should take care not to resample too often. In our experiments with the OU process, we can resample more than $K = 100$ times without any adverse effects because the process of splitting/killing particles is carefully tuned to the underlying dynamics. However, for more complicated problems in which an ideal choice of θ is not available, resampling should be performed as little as possible, while sustaining the necessary particle dynamics.[32] When testing a new reaction coordinate, it is a good strategy to start with just a few resampling times and increase resampling frequency once the coordinate is proven to be effective.

In a complex, high-dimensional system, where the reaction coordinate θ is imperfect, it is essential to gauge the quality of DMC estimates by providing error bars. For estimates

$$P\{X_{t_K} \in A\} \approx \frac{\overline{w}_{K-1}}{N} \sum_{i=1}^{N} \frac{\mathbb{1}\left\{\xi_K^{(i)} \in A\right\}}{\exp\left\{V_{K-1}'\left(\hat{\xi}_{K-1}^{(i)}\right)\right\}}, \quad (27)$$

it is possible to estimate variance σ^2 using[33]

$$\frac{1}{N^2} \sum_{i=1}^{N} \left| \sum_{\text{anc}\left(\xi_K^{(j)}\right)=i} \frac{\overline{w}_{K-1}\mathbb{1}\left\{\xi_K^{(j)} \in A\right\}}{\exp\left\{V_{K-1}'\left(\hat{\xi}_{K-1}^{(j)}\right)\right\}} - \hat{p} \right|^2. \quad (28)$$

Here, anc $\left(\xi_K^{(j)}\right)$ denotes the "ancestral index" of particle $\xi_K^{(j)}$. Tracing the ancestry of $\xi_K^{(j)}$ back to an initial particle $\xi_0^{(i)}$, anc $\left(\xi_K^{(j)}\right)$ is the index i for the initial particle. We note that the variance estimator $\hat{\sigma}^2$ is most accurate when the sample size N is very high and when the number of resampling times K is small, which is not always the case in practical simulations. Properties of the variance estimator $\hat{\sigma}^2$ are discussed at more length in the Appendix.

In summary, quantile DMC is straightforward to implement and becomes increasingly effective relative to other sampling methods as parameters of splitting/killing are tuned and as the number of particles N increases. However, quantile DMC can perform poorly when the reaction coordinate θ fails to anticipate paths leading to rare event states. Lastly, we have shown how to provide rough error bars that assess the accuracy of quantile DMC estimates.

III. EXTREME MESOSCALE WEATHER

In this section, we discuss why it is difficult to estimate the frequency of extreme mesoscale weather and how rare event sampling methods like quantile DMC can potentially assist in calculations. Then, we focus our attention on the frequency of intense tropical cyclones and present simulations, which illustrate both the potential benefits of rare event sampling and the work that remains to be done.

A. Frequency of extreme weather

While extreme weather events such as heat waves, floods, and tropical cyclones are rare, they can cause immense damage and fatalities.[15,16,34] Understanding the frequency of extreme weather is therefore an essential task, both for real-world disaster preparedness and for assessing weather's impact on society. The study of weather extremes is more relevant than ever, since evidence points to changing frequencies of extreme weather events with climate change.[34–36]

To understand the frequency of extreme weather, observations provide the most fundamental data source, but these data have important limitations. For many potential extreme weather events, no historical analog exists. Storms can occur in surprising places. Droughts can afflict new areas. Even when historical data are available, measurements can be sparse and sometimes corrupted.[37] Most critically, as the climate changes, the frequency and intensity of tropical cyclones, of heat waves, and of flooding are expected to change, so that historical measurements will become less relevant.[34–36] For all these reasons, climate simulations provide essential additional insight into extreme weather in historical, current, and potential future climates.

When modeling extreme weather, there is a trade-off between bias and variance. While inexpensive models can be run many times, leading to low-variance estimates, these estimates can be highly biased. More computationally-intensive models potentially provide a less biased climatology of extreme weather[17,18,36] yet these models cannot be run for as long or with as many ensemble members, due to limited computational resources. Because of these practical limitations, estimates can have high variance, particularly for the rarest, most extreme weather events.

Mesoscale extreme weather, such as floods and tropical cyclones, provides a stark example of the need for increased statistical accuracy without sacrificing model fidelity. Because mesoscale extreme weather occurs on a fine spatial scale (10-1000 km) at which simplifying assumptions for thermodynamics and dynamics begin to fail, running accurate mesoscale weather models can be enormously expensive. It is here that the computational burden is the greatest. It is here, therefore, that rare event sampling methods stand to provide the greatest benefit.

In extreme mesoscale weather simulations where rare event sampling could potentially provide a benefit, a process known as "dynamical downscaling"[38] is now common. First, a Global Climate Model (GCM) simulates a coarse-resolution version of an extreme weather event. Using initial and boundary conditions from the GCM, a high-resolution regional model with more complex physics then enhances the GCM output, simulating the local details of the extreme weather event.

The dynamical downscaling approach is necessary because current GCMs cannot simulate the details of mesoscale weather that are essential for damage assessment. For example, the peak winds of a tropical cyclone are underestimated by a GCM,[2] and a GCM cannot resolve the overflowing riverbeds that lead to flooding.[18,36] Versions of dynamical downscaling have become standard in storm surge modeling,[3] flood modeling,[36] and tropical cyclone modeling.[2]

A simple probabilistic interpretation helps clarify how dynamical downscaling can estimate the probability of an extreme weather event occurring. Let A denote an extreme event and let B denote the coarse-scale meteorological conditions that are necessary for event A to occur. We can then write

$$P\{\text{event } A\} = P\{\text{conditions } B\}$$
$$\times P\{\text{event } A \mid \text{conditions } B\}. \quad (29)$$

A GCM is used to evaluate the first probability $P\{\text{conditions } B\}$, whereas a regional model is used to evaluate the second probability $P\{\text{event } A \mid \text{conditions } B\}$ using output from the GCM. For example, Bender and co-workers[2] identified protocyclones in a GCM and then used a high-resolution regional model to simulate the intensity evolution of protocyclones into full-fledged tropical cyclones.

Applying rare event sampling in a dynamical downscaling context is a multitiered process. A GCM can be run either directly or with a splitting method such as quantile DMC. Then, starting from the intial conditions selected from GCM output, a regional weather model can be run either directly or with a splitting method such as quantile DMC. At multiple stages of the dynamical downscaling process, statistics can be potentially improved by a judicious application of rare event sampling.

In summary, there is a pressing need for rare event sampling of mesoscale extreme weather since mesoscale simulations are enormously expensive. The incorporation of rare event sampling into extreme weather calculations can enable higher-resolution, more computationally-intensive models, ultimately leading to more accurate extreme weather risk assessment.

B. Tropical cyclone test case

1. Motivation for simulations

Tropical cyclones rank among the deadliest natural disasters in human history. Approximately 300 000 died in the 1970

Bhola cyclone, and an estimated 138 000 died in the 2008 cyclone Nargis.[39,40] While high-intensity tropical cyclones (TCs) are rare, they are the most destructive and fatal TCs.[16,41] Moreover, the frequency of the most intense storms is expected to increase with climate change, the precise rate of change being an open area of research.[35] Understanding the upper tail of intensity for TCs is therefore of paramount societal concern.

Reducing computational cost is a central priority for the TC modeling community. TCs are most accurately simulated using high-resolution weather models with 1–10 km horizontal resolution.[17] High spatial resolution is required to resolve the storm eyewall where winds are the strongest. Increasing horizontal resolution can lead to more accurate simulations; however, increased resolution comes at a steep computational cost. Doubling horizontal resolution requires an eight-fold increase in computational expense because resolution must be doubled in the zonal and meridional directions, and the time step must be cut in half to ensure numerical stability.

Helping to alleviate the computational burden of TC modeling, rare event sampling potentially provides a means to accurately estimate TC statistics with a reduced sample size of high-resolution simulations. Here, as a proof of concept, we apply quantile DMC to estimate statistics for two high-intensity TCs. For initial and boundary conditions, we use reanalysis data for two storms that achieved Category 4 status in the real world but did not achieve the highest intensity level, Category 5 status. We model these storms using a stochastic model that predicts a range of possible intensities, different from the real-world intensities of Hurricane Earl and Hurricane Joaquin. Starting from coarse-scale protocyclones, we nudge the evolution of storms toward high intensities using the quantile DMC algorithm and estimate the probability of high-intensity manifestations of these storms. We then compare the efficiency of quantile DMC to the efficiency of direct sampling for this estimation problem.

These simulations are envisioned as a first step toward the goal of using quantile DMC to study the probability of intense TCs in historical, present, and future climates. In our simulations, we use quantile DMC to provide statistics for just two storms. In the future, however, as quantile DMC is applied to study the frequency of intense TCs in different climates, it will be necessary to start simulations from an ensemble of hundreds of different protocyclones. For these initial investigations, we use a small sample size $N = 100$, whereas a sample size of $N = 1000$ would be more appropriate for a full implementation of quantile DMC in the future.

2. Simulation details

The 2010 storm Earl was a long-lived hurricane of tropical origin that came very close to the Eastern seaboard of the United States but ultimately did not make landfall, passing 150 km off the coast of Massachusetts. The 2015 storm Joaquin was a hurricane of extratropical origin that intensified more rapidly than expected, leading to the worst U.S. maritime disaster in decades, the sinking of the cargo ship El Faro with all 33 sailors aboard.[42]

We model Hurricanes Earl and Joaquin using the Advanced Research Weather Research and Forecasting Model[43] (ARW, version 3.9.1.1), which has been applied extensively for hurricane research in the past.[17] ARW is a finite-difference model whose governing equations include energy conservation, mass conservation, a thermodynamic law, and an equation of state. ARW incorporates parametrized physics schemes for precipitation processes, heat/moisture fluxes over land, radiation, and mixing in atmospheric columns. ARW uses artificial dissipation and filters to achieve numerical stability.

We have previously used the ARW model to simulate intense tropical cyclones;[14] however, our previous simulations did not incorporate the quantile DMC algorithm. In Plotkin et al.,[14] we used an optimization strategy to identify a maximum likelihood pathway for a simulated tropical cyclone to achieve a high intensity at a terminal time. While the optimization approach is useful for identifying factors that can cause tropical cyclones to intensify, it is less suitable for computing tropical cyclone statistics. Statistics of TCs can depend on myriad possible paths, and sampling these paths is a necessary requirement for accurate estimation of statistics. Thus, in the current work, we present quantile DMC as an additional tool, uniquely suited to low-variance calculation of tropical cyclone statistics.

Among weather models, ARW has the advantage that it supports vortex-following nested domains. Three domains of different resolutions are often used in TC simulations. The outer domain is static, while the inner domains follow a local minimum in the 500 hPa geopotential height field, indicating a TC's location. Vortex-following domains enable high resolution around the eye of the storm without the computational expense of high resolution across the entire storm path.

Our simulations use a time step of 6.7 s and horizontal resolution of 2 km in the inner domain, a time step of 20 s and resolution of 6 km in the middle domain, and a time step of 60 s and resolution of 18 km in the outer domain. The inner domain stretches $468 \times 468 \, \text{km}^2$, the middle domain $1404 \times 1404 \, \text{km}^2$, and the outer domain $5382 \times 5382 \, \text{km}^2$. All domains use 40 vertical levels. The physics parametrizations are the same as used by Judt et al.[44] Convection is explicitly simulated in the two inner domains, whereas convection is parametrized in the outermost domain.

Simulations are randomly perturbed using the Stochastic Kinetic Energy Backscatter (SKEB) scheme.[45] Perturbations from this physics scheme are smooth in space and time, but they change rapidly, modeling the effects of small-scale turbulent processes. Potential temperature and the non-divergent component of horizontal wind are both independently perturbed with forcing terms

$$F(x, y, z, t) = \sum_{j,k} \Re \left\{ F_{j,k}(t) \, H_{j,k}(x, y) \, e^{iC_{j,k,z}} \right\}, \quad (30)$$

where $H_{j,k}$ are the Fourier modes for the domain and $C_{j,k,z}$ are constants that produce a westward phase tilt in the perturbation field. $F_{j,k}$ terms evolve randomly, according to a complex-valued Ornstein-Uhlenbeck process with decorrelation time scale $\alpha^{-1} \approx 0.5$ h, that is,

$$dF_{j,k} = -\alpha F_{j,k} dt + \sqrt{2\alpha \sigma_{j,k}^2} dW_{j,k}, \quad (31)$$

where $W_{j,k}$ denote independent complex-valued Brownian motions and the noise amplitude $\sigma_{j,k}^2$ decreases as a power law of $\sqrt{j^2 + k^2}$. The SKEB scheme is implemented as a physics module within the ARW software. The paper of Berner et al.[45] describes how this physics module discretizes the underlying OU process dynamics as a first-order autoregressive process. At every ARW time step, the perturbation

field is updated, and perturbations provide a series of small, frequent changes influencing model dynamics.

In our previous simulations of tropical cyclones,[14] we used an alternative probabilistic model for ARW perturbations, different from the SKEB scheme. In Plotkin et al.,[14] a Gaussian perturbation is applied once per hour to the ARW model. The optimization strategy used in Plotkin et al.[14] would not have been practical to apply to the SKEB scheme. The SKEB scheme perturbs simulations many times per minute, making a derivative-based optimization challenging, but leading to small and physically realistic perturbations.

We use direct sampling and quantile DMC to sample extreme intensities of Earl and Joaquin with an ensemble size of $N = 100$ simulations. Direct sampling runs are seven-day forward runs of the ARW model with SKEB. Quantile DMC runs are seven-day forward runs, which incorporate splitting and killing of trajectories. Since the ARW model with SKEB is a Markovian model, quantile DMC provides unbiased estimates of a wide range of statistics of tropical cyclones, including pathways, characteristics, and frequencies. For simplicity, however, this paper discusses only the probability of intense tropical cyclones occurring.

Table II describes the specific start and end times for the simulations. Start times are selected near the beginning of the hurricane life cycle. End times are selected 7 days after start times, which gives hurricanes sufficient time to reach peak intensity and then recede for at least two days. For the quantile DMC runs, we select a time T when we expect each storm to achieve peak intensity. We resample at times $t_1 = T - 48\,\text{h}$, $t_2 = T - 24\,\text{h}$, $t_3 = T - 12\,\text{h}$, and $t_4 = T$. Thus, resampling increases in frequency in an attempt to maximize quantile DMC efficiency. Resampling stops once hurricanes are expected to achieve peak intensity since resampling too often can increase the variance of quantile DMC estimates. The specific time T for each storm is identified using official best track historical data.[46]

The parameters for our quantile DMC simulations are defined as follows:

1. The reaction coordinate θ is the deviation of sea surface pressure from a hydrostatically-balanced reference state[43] at the storm core.
2. We assume the reaction coordinate θ can be modeled as a monotonic, time-dependent transformation of an OU process. Since $N(0, 1)$ distributions work well as target distributions when sampling from the OU process, our target distribution is $N(0, 1)$ at each adaptation step.

TABLE II. Start times, end times, and resampling times for Earl and Joaquin (all time zones are UTC).

		Earl	Joaquin
t_0	Start	August 27 00:00	September 29 00:00
t_1	Resample 1	August 30 00:00	October 1 00:00
t_2	Resample 2	August 31 00:00	October 2 00:00
t_3	Resample 3	August 31 12:00	October 2 12:00
t_4	Resample 4	September 1 00:00	October 3 00:00
t_5	End	September 3 00:00	October 6 00:00

3. Splitting functions take the form $V'_k = Ce^{\alpha(t-T)}\theta'_k$. The decorrelation time scale $\alpha^{-1} = 3d$ is the appropriate time scale for large-scale differences to emerge in TC development in the ARW model.[44] The splitting constant $C = 1$ is appropriate for estimating intensity quantiles up to the 99.9th percentile of intensity. Moreover, when sampling an OU process with a splitting constant of $C = 1$, four rounds of resampling and a sample size of $N = 100$ simulations are sufficient to ensure the effectiveness of quantile DMC compared to direct sampling.

For both quantile DMC and direct sampling, equivalent computing resources are required on the University of Chicago Research Computing Center high-performance cluster: 100 nodes ran continuously for 2 days with 28 CPUs per node and 2 gigabytes of RAM per CPU.

3. Simulation results

In Fig. 7, we present intensity trajectories for direct sampling and quantile DMC runs for Hurricane Earl and Hurricane Joaquin. Direct sampling trajectories typically occupy the middle quantiles of intensity, whereas quantile DMC trajectories are more likely to occupy the upper quantiles of intensity. For example, direct sampling produces zero Category 5 realizations of Earl and only four of Joaquin. In contrast, quantile DMC produces three Category 5 realizations of Earl and 22 of Joaquin. Therefore, quantile DMC is successful at simulating more intense storms compared to direct sampling.

When presenting results, we measure TC intensity using minimum sea surface pressure. While it is also common to see TC intensity reported using maximum wind speed, econometric analysis finds that minimum sea surface pressure is a better predictor of TC damage and fatalities than is wind speed.[41] Pressure combines information on wind speed and storm size,[47] thereby giving a more holistic indication of TC damage.[48] We note that historically the Saffir-Simpson hurricane scale combined maximum wind speed, minimum pressure and maximum storm surge information to classify TCs into Categories 1 (least intense) to 5 (most intense). More recently, the scale was renamed the Saffir-Simpson hurricane wind scale, and Categories 1–5 are defined by maximum wind speeds alone.[49] When presenting results, we use the historical Saffir-Simpson hurricane scale to define Categories 1–5 in terms of minimum sea surface pressure.

We can use data from quantile DMC and direct sampling runs to estimate a range of statistics associated with storm intensity. In particular, we estimate cumulative distribution functions for random variables P_T and P_{life}. P_T is the TC intensity at the particular time T when each storm is expected to reach maximum strength, namely, September 1 00:00 UTC for Earl and October 3 00:00 UTC for Joaquin. P_{life} is the strongest TC intensity over the entire seven days of simulated time. The cumulative distribution function for P_T is defined by $F(U) = P\{P_T \leq U\}$, where U ranges over all possible pressures. To estimate $P\{P_T \leq U\}$ from Quantile DMC data, we use

$$P\{P_T \leq U\} \approx \frac{\overline{w}_{K-1}}{N} \sum_{i=1}^{N} \frac{\mathbb{1}\left\{P_T\left(\xi_K^{(i)}\right) \leq U\right\}}{\exp\left\{V'_{K-1}\left(\hat{\xi}_{K-1}^{(i)}\right)\right\}}. \quad (32)$$
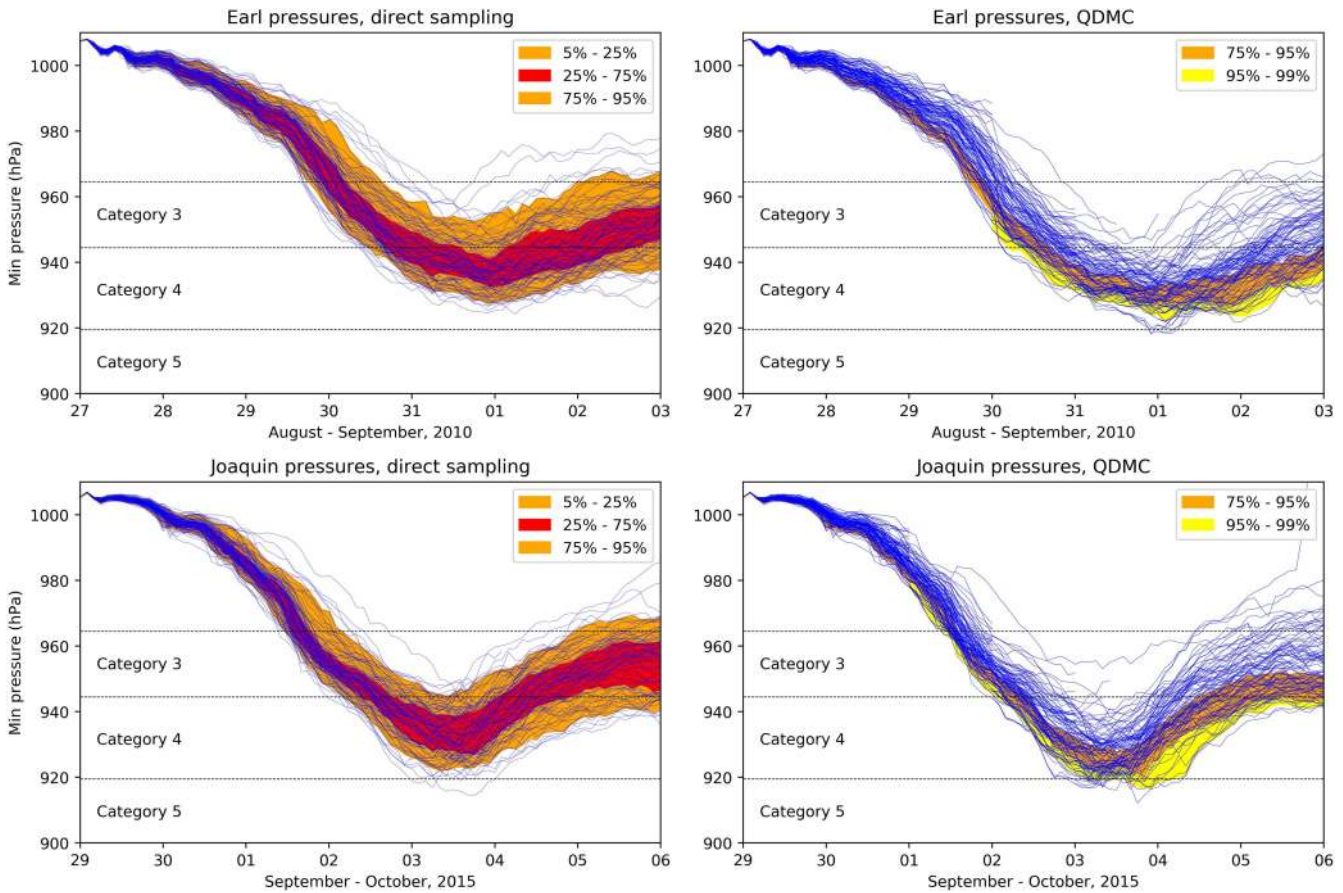
FIG. 7. Intensity trajectories for $N = 100$ direct sampling and quantile DMC simulations with intensity quantiles estimated from data. In direct sampling (left), trajectories occupy the middle quantiles of intensity (orange and red); in quantile DMC (right), trajectories are more likely to occupy extreme quantiles of intensity (yellow). While direct sampling trajectories span from the beginning to the end of simulations, quantile DMC trajectories can end when killing occurs.

To estimate $\mathrm{P}\{P_T \leq U\}$ from direct sampling data, we use

$$\mathrm{P}\{P_T \leq U\} \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{P_T\left(\xi^{(i)}\right) \leq U\right\}. \quad (33)$$

We apply analogous formulas when estimating the cumulative distribution function for P_{life}.

Figure 8 provides side-by-side comparisons between the direct sampling and quantile DMC estimates. Intensity estimates from quantile DMC are statistically consistent with intensity estimates from direct sampling. For example, the estimated probability for Earl to reach Category 5 status is 0% from direct sampling and 0.2% from quantile DMC. The estimated probability for Joaquin to reach Category 5 status is 4% from direct sampling and 6% from quantile DMC. The disagreement between estimates falls well within the range of random error, particularly if one or the other estimate has high relative variance. The cumulative distribution functions estimated from direct sampling exhibit sharper jump discontinuities compared to the relatively smooth behavior of the CDFs from quantile DMC.

This jumpy behavior may reflect a greater degree of error in the direct sampling estimates if it can be assumed that the distribution of hurricane pressures is smooth.

To check the hypothesis that quantile DMC estimates are more accurate, we can use data from quantile DMC and direct sampling runs to gauge the variance in our estimates. For quantile DMC estimates of $\mathrm{P}\{P_T \leq U\}$, we assess variance using

$$\hat{\sigma}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \left| \sum_{\mathrm{anc}\left(\xi_K^{(j)}\right)=i} \frac{\overline{w}_{K-1}\mathbb{1}\left\{P_T\left(\xi_K^{(j)}\right) \leq U\right\}}{\exp\left\{V_{K-1}'\left(\hat{\xi}_{K-1}^{(j)}\right)\right\}} - \hat{p} \right|^2. \quad (34)$$

For direct sampling estimates of $\mathrm{P}\{P_T \leq U\}$, we assess variance using $\hat{\sigma}^2 = \frac{1}{N}\hat{p}\left(1 - \hat{p}\right)$.

Figure 9 provides side-by-side comparisons between direct sampling and quantile DMC relative variances $\hat{\sigma}^2/\hat{p}^2$. For many important rare event sampling estimates, quantile DMC provides substantial variance reduction compared to direct sampling. When
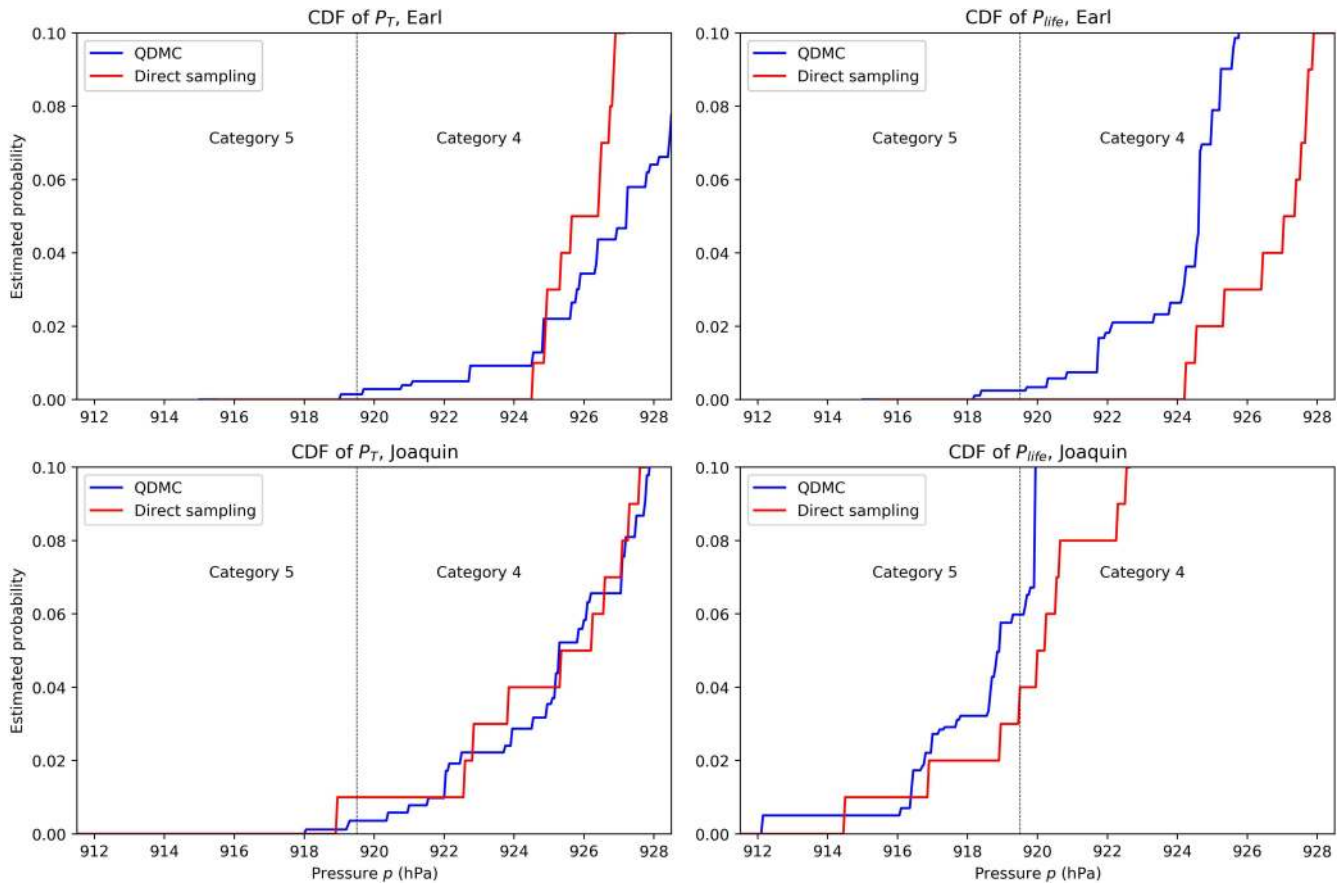
FIG. 8. Estimates of cumulative distribution functions P $\{P_T \leq p\}$ and P $\{P_{life} \leq p\}$ are roughly consistent between quantile DMC and direct sampling. However, estimates from direct sampling exhibit large jumps, which may be a sign of higher error.

estimating the P_T distribution, quantile DMC gives reduced variance for all pressures lower than 925 hPa for both Earl and Joaquin. When estimating the P_{life} distribution, quantile DMC gives reduced variance for all pressures lower than 925 hPa for Earl and 916 hPa for Joaquin. At the lowest pressures, the variance of quantile DMC is two to ten times lower than the variance of direct sampling.

An important reason for presenting the P_T distribution is to show that quantile DMC is highly effective at sampling intense storms at the reference time T. More generally, quantile DMC is effective at sampling intense storms in the twelve hours leading up to the reference time T and the twelve hours following the reference time T. Outside this window of time, quantile DMC may be less effective at sampling intense storms. It is therefore of central importance to make sure the reference time T aligns with the time of maximum intensity. To achieve this goal, it may be necessary to run the model in advance of quantile DMC simulations or predict based on initial and boundary conditions when the model will achieve peak intensity.

The one anomalous result in the pattern of variance reduction due to quantile DMC is quantile DMC's limited benefit when estimating the P_{life} distribution for Hurricane Joaquin. To shine light on the

limitations of quantile DMC for this particular estimation problem, we can examine the family weights

$$W^{(i)} = \sum_{\text{anc}\left(\xi_K^{(j)}\right)=i} \frac{\overline{w}_{K-1} \mathbb{1}\left\{P_{life}\left(\xi_K^{(j)}\right) \leq U\right\}}{\exp\left\{V'_{K-1}\left(\hat{\xi}_{K-1}^{(j)}\right)\right\}} \tag{35}$$

contributing to the variance estimator $\hat{\sigma}^2$. When estimating the probability for Joaquin to reach Category 5 status, the eight nonzero family weights $W^{(i)}$ are 0.2, 0.2, 0.3, 0.3, 0.5, 0.6, 0.9, and 3.1. The largest family weight of $W^{(i)} = 3.1$ accounts for 85% of the variance $\hat{\sigma}^2$. Without this particular $W^{(i)}$ value, $\hat{\sigma}^2$ would be 3.5 times smaller for quantile DMC compared to direct sampling. Because of this particular $W^{(i)}$ value, $\hat{\sigma}^2$ is instead 1.9 times larger for quantile DMC compared to direct sampling.

The largest weight of $W^{(i)} = 3.1$ belongs to a family of late-developing storms. The initial simulation in this family exhibited low intensity at the first resampling time, but it was not killed during the resampling step due to random chance. Over time, the simulation increased in intensity. The single particle was eventually split into
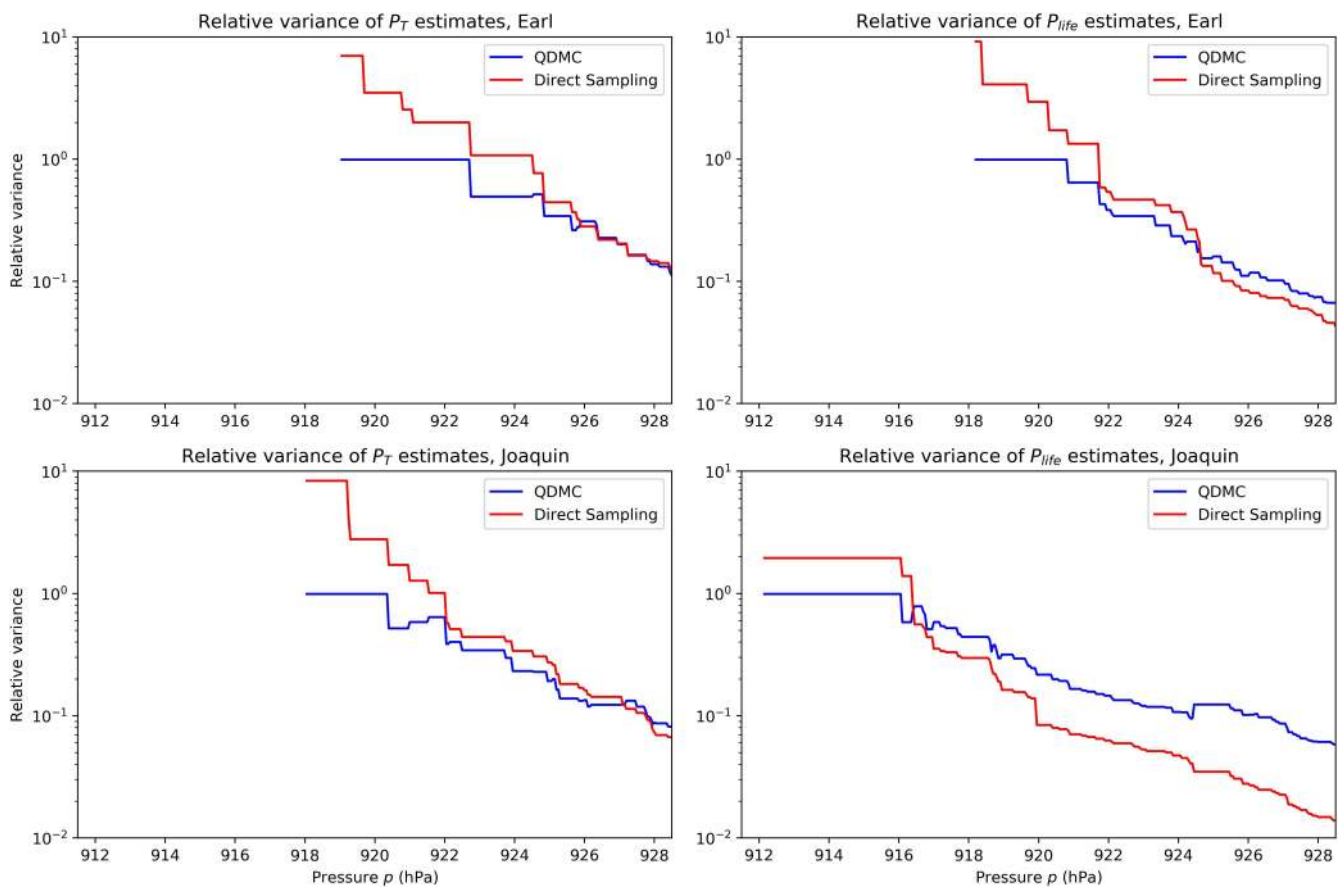
FIG. 9. Variances are lower for quantile DMC than direct sampling at extreme tail pressures p. Quantile DMC provides the greatest benefit when estimating P_T statistics for Earl and Jaquin and P_{life} statistics for Earl. Quantile DMC provides less of an improvement when estimating P_{life} statistics for Joaquin.

six family members, and ultimately all six family members achieved Category 5 status. The high variance of Joaquin P_{life} estimates can be wholly attributed to this single family of late-developing storms.

The story of the largest family weight $W^{(i)} = 3.1$ illustrates key areas in which our design of quantile DMC simulations of tropical cyclones was not optimal. First, the reaction coordinate θ failed to anticipate which trajectories would lead to high intensities at later times. At the first resampling time, the value of $\theta\left(\xi_1^{(i)}\right)$ was low, but all six descendents of $\xi_1^{(i)}$ would go on to achieve Category 5 status. This rapid intensification was not predictable using sea surface pressure as our reaction coordinate. However, an improved reaction coordinate θ could predict lifetime intensity based on additional variables such as steering flow, vertical wind shear, and relative humidity. With an improved reaction coordinate, future intensity could be more accurately identified, thereby reducing quantile DMC variance.

A second shortcoming in the design of simulations was the poor identification of the time T for Joaquin to reach peak intensity. For Earl, the reference time T was correctly identified as September 1 00:00 UTC, the approximate time when storms achieved peak

intensity in direct sampling and quantile DMC runs. For Joaquin, on the other hand, the reference time T was incorrectly identified as October 3 00:00 UTC, nearly 24 h before peak intensity occurred in direct sampling and quantile DMC runs. With a later reference time T, late-developing storms could be more appropriately sampled, reducing the variance of P_{life} estimates for Hurricane Joaquin.

We report three takeaway messages for future applications of quantile DMC to study intense tropical cyclones.

First, with an increased ensemble of $N = 1000$ simulations, we are optimistic that quantile DMC can provide low-error estimates of extreme tail probabilities. Already with a small sample of $N = 100$ simulations, we see signs of reduced variance using quantile DMC. But with increased sample size N, we expect the error of quantile DMC to shrink with a faster-than-$N^{-1/2}$ scaling, enhancing quantile DMC's advantages over direct sampling. When estimating the most extreme quantiles of intensity, a splitting method such as quantile DMC truly excels.

Second, when simulating cyclones, quantile DMC is a more convenient algorithm to use than standard DMC. Using quantile DMC, the decision to select a splitting constant of $C = 1$ is straightforward.

In contrast, using DMC, similar results can only be obtained with the foresight to select a splitting constant C with C^{-1} being the standard deviation of intensity, namely, 7 hPa–8 hPa. Moreover, in future simulations of tropical cyclones, there is the possibility of highly skewed or bimodal intensity distributions.[50] In these contexts, the additional robustness of quantile DMC over DMC may provide a further benefit.

Lastly, to maximize the potential benefit of quantile DMC in future simulations, it is of paramount importance to improve the reaction coordinate θ and estimates of the time of maximum intensity. In our simulations, the poor identification of these parameters led to added variance in some quantile DMC calculations. Improved parameters θ and T would better predict lifetime storm intensities, alleviating the problem of variance inflation.

The search for improved predictions of lifetime intensity is challenging in part due to a limited understanding of the precursors and dynamics of rapid intensification.[51] In complementary work, we develop a rare event analysis tool that offers insight into the physics of TC rapid intensification.[14] Such a technique could potentially identify a more suitable and predictive reaction coordinate for future quantile DMC applications. By incorporating careful data analysis and scientific insight into the model being simulated, future work can potentially improve the efficiency of quantile DMC in tropical cyclone simulations.

IV. CONCLUSION

Efficient sampling of extreme mesoscale weather remains one of the outstanding computational challenges of the 21st century. Extreme weather, such as tropical cyclones, squall lines, and floods, has a tremendous impact on human society, yet assessing the frequency of extreme weather in past, current, and projected future climates is extremely difficult. Responding to this challenge, we have introduced a new rare event sampling algorithm, quantile DMC. Combining quantile DMC with dynamical downscaling provides a new paradigm for calculating extreme weather statistics. This approach potentially enables high-accuracy, computationally-intensive models to be run with reduced computational cost, raising the quality of extreme weather statistics.

In Secs. II D and II E, we have provided a practical guide to using quantile DMC. In particular, we offer specific recommendations for the parameters to be used in the algorithm. When computing tail probabilities for the Ornstein-Uhlenbeck process, quantile DMC is over a thousand times more efficient than direct sampling and is more stable than diffusion Monte Carlo. When computing tail probabilities for intense tropical cyclones, quantile DMC is two to ten times more efficient than direct sampling, with the possibility for greater efficiency in future simulations.

There remain important challenges in applying quantile DMC to simulate extreme weather events. In our simulations of tropical cyclones, we observe that quantile DMC's performance depends on a *reaction coordinate*, a one-dimensional coordinate that anticipates the occurrence of high-intensity weather. The reaction coordinate that we used in our simulations was not optimal, and we anticipate using an improved reaction coordinate in future TC sampling. Fortunately, even with an imperfect choice of reaction coordinate, it is possible for quantile DMC to provide a reduction of variance.

We acknowledge two issues that affect the future of rare event sampling of extreme weather. First, splitting methods like quantile DMC can only be successful if extreme weather is simulated stochastically. However, stochastic models are easily available and increasingly used in many state-of-the-art geophysical computations,[52] so this does not present a major limitation in practice. Second, rare event sampling is unnecessary if there exist models that are both accurate and inexpensive to run. While we acknowledge some statistical models and simplified physics models perform remarkably well in today's extreme weather calculations,[38,53] our outlook is to the future. We believe high-resolution three-dimensional models will eventually provide the greatest accuracy in all areas of extreme weather inference. Developing rare event sampling methods is therefore essential preparation for the future as computationally-intensive ensembles become the authoritative source of insight in extreme weather inference.

APPENDIX: FURTHER ALGORITHMIC DETAILS

1. Resampling schemes

There are many possible *resampling schemes* that can be used during DMC's resampling step. One scheme that works well in practice is *sorted stratified resampling*.[27,28] The scheme first sorts particles $\left(\xi_k^{(i)}\right)_{1\leq i\leq N}$ based on the values of $\left(\theta\left(\xi_k^{(i)}\right)\right)_{1\leq i\leq N}$ and then selects new particles $\left(\hat{\xi}_k^{(i)}\right)_{1\leq i\leq N}$ using stratified resampling.

Definition A.1 (Sorted stratified resampling).

1. *Sorting*: Reindex the particles and weights $\left(w_k^{(i)}, \xi_k^{(i)}\right)_{1\leq i\leq N}$ so that

$$\theta\left(\xi_k^{(1)}\right) \leq \theta\left(\xi_k^{(2)}\right) \leq \cdots \leq \theta\left(\xi_k^{(N)}\right). \tag{A1}$$

2. *Stratified resampling*: Construct the empirical quantile function $Q_t\colon [0,1) \to \mathbb{R}^d$ for the ensemble $\left(w_k^{(i)}, \xi_k^{(i)}\right)_{1\leq i\leq N}$ as follows:

$$Q_k(x) = \xi_k^{(i)}, \quad \frac{\sum_{j=1}^{i-1} w_k^{(j)}}{\sum_{j=1}^{N} w_k^{(j)}} \leq x < \frac{\sum_{j=1}^{i} w_k^{(j)}}{\sum_{j=1}^{N} w_k^{(j)}}. \tag{A2}$$

Select updated particles $\hat{\xi}_k^{(i)} = Q_k\left(\frac{j-1+U_k^{(i)}}{N}\right)$ *for* $1 \leq i \leq N$,

where $U_k^{(i)}$ *are independent Unif* $(0, 1)$ *random variables.*

It can be checked that each particle $\xi_k^{(i)}$ is duplicated an expected number of $w_k^{(i)}/\overline{w}_k$ times. Therefore, sorted stratified resampling is a valid resampling scheme.

2. Invariance under monotonic transformations

Quantile DMC is unchanged if the reaction coordinate θ is replaced with a reaction coordinate $\tilde{\theta}$ that is a monotonic, one-to-one transformation of θ. To show this, we provide an alternate description of quantile DMC's process for splitting/killing particles that makes clear the property of invariance under monotonic transformations.

First, we introduce an order relation $x \prec y$ that indicates $\theta(x) < \theta(y)$ and an equivalence relation $x \sim y$ that indicates $\theta(x) = \theta(y)$. We observe that the order relation $x \prec y$ and equivalence relation $x \sim y$ remain unchanged if θ is replaced by $\tilde{\theta}$.

Second, define the quantiles $p_k^{(i)}$ with the formula

$$\sum_{j=1}^{N} z_k^{(j)} \left[\mathbb{1}\left\{\xi_k^{(j)} \prec \xi_k^{(i)}\right\} + \frac{1}{2}\mathbb{1}\left\{\xi_k^{(j)} \sim \xi_k^{(i)}\right\} \right], \qquad (A3)$$

where

$$\begin{cases} z_k^{(i)} = \dfrac{1}{N}, & k = 0, \\[2mm] z_k^{(i)} = \dfrac{\exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(i)}\right)\right\}}{\sum_{j=1}^{N}\exp\left\{-V'_{k-1}\left(\hat{\xi}_{k-1}^{(j)}\right)\right\}}, & k > 0. \end{cases} \qquad (A4)$$

The quantitities $p_k^{(i)}$ are approximate quantiles for the distribution $\theta(X_{t_k})$. Thus, if $p_k^{(i)} = 0.9$ there is an approximate one-in-ten chance that $\theta(X_{t_k})$ takes a value as high as $\theta\left(\xi_k^{(i)}\right)$.

Third, the splitting function takes the values

$$V'_k\left(\xi_k^{(i)}\right) = V'_k\left(F_{\nu_k}^{-1}\left(p_k^{(i)}\right)\right). \qquad (A5)$$

In this description, quantile DMC only relies on the reaction coordinate θ through the order relation $x \prec y$ and equivalence relation $x \sim y$. These two relations are unchanged under monotonic, one-to-one transformations of the reaction coordinate.

In addition to the property of invariance under monotonic transformations, quantile DMC also has the property that estimates are unbiased, which can be established following standard martingale arguments.[26,27] Numerical evidence indicates that estimates converge with an asymptotic $1/\sqrt{N}$ error rate, as $N \to \infty$. However, a rigorous mathematical analysis of quantile DMC's error remains a task for future research.

3. Variance estimation for quantile DMC

The variance estimator (28) for quantile DMC was originally developed assuming a different resampling scheme called Bernoulli resampling is used.[33] Indeed, when the DMC algorithm is performed using Bernoulli resampling, the variance estimator is asympotically

consistent as $N \to \infty$. We have chosen to use a different resampling scheme that gives better performance than the Bernoulli resampling scheme. Consequently, the variance estimator $\hat{\sigma}^2$ is biased toward overestimating the variance. Numerical experiments with the Ornstein-Uhlenbeck process suggest that the variance is inflated by less than 10%, at least for $K \leq 10$ resampling times. Another potential concern is the natural variability in the estimator $\hat{\sigma}^2$. We find that $\hat{\sigma}^2$ is most reliable when family weights

$$W^{(i)} = \sum_{\text{anc}\left(\xi_k^{(j)}\right)=i} \frac{\overline{w}_{K-1}\mathbb{1}\left\{\xi_K^{(j)} \in A\right\}}{\exp\left\{V'_{K-1}\left(\hat{\xi}_{K-1}^{(j)}\right)\right\}} \qquad (A6)$$

are nonzero for many indices i. Consequently, $\hat{\sigma}^2$ is most reliable when sample size N is high and the number of resampling times K is low. We find that the variance estimator $\hat{\sigma}^2$ provides a valuable tool for gauging the quality of quantile DMC estimates. We caution the reader, however, there may be situations outside the current context where the variance estimator $\hat{\sigma}^2$ behaves poorly.

REFERENCES

[1] G. A. Meehl and C. Tebaldi, "More intense, more frequent, and longer lasting heat waves in the 21st century," Science **305**, 994–997 (2004).
[2] M. A. Bender, T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. M. Held, "Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes," Science **327**, 454–458 (2010).
[3] N. Lin and K. Emanuel, "Grey swan tropical cyclones," Nat. Clim. Change **6**, 106–111 (2016).
[4] H. Kahn and T. E. Harris, "Estimation of particle transmission by random sampling," Natl. Bureau Stand. Appl. Math. Ser. **12**, 27–30 (1951).
[5] M. N. Rosenbluth and A. W. Rosenbluth, "Monte Carlo calculation of the average extension of molecular chains," J. Chem. Phys. **23**, 356–359 (1955).
[6] A. Dubi, T. Elperin, and D. J. Dudziak, "Geometrical splitting in Monte Carlo," Nucl. Sci. Eng. **80**, 139–161 (1982).
[7] G. A. Huber and S. Kim, "Weighted-ensemble Brownian dynamics simulations for protein association reactions," Biophys. J. **70**, 97 (1996).
[8] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Variance reduction techniques for estimating value-at-risk," Manage. Sci. **46**, 1349–1364 (2000).
[9] R. N. Hoffman, J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehrkorn, "The response of damaging winds of a simulated tropical cyclone to finite-amplitude perturbations of different variables," J. Atmos. Sci. **63**, 1924–1937 (2006).
[10] J. Weare, "Particle filtering with path sampling and an application to a bimodal ocean current model," J. Comput. Phys. **228**, 4312–4331 (2009).
[11] E. Vanden-Eijnden and J. Weare, "Data assimilation in the low noise regime with application to the Kuroshio," Mon. Weather Rev. **141**, 1822–1841 (2013).
[12] F. Ragone, J. Wouters, and F. Bouchet, "Computation of extreme heat waves in climate models using a large deviation algorithm," Proc. Natl. Acad. Sci. U.S.A. **115**, 24–29 (2018).
[13] G. Dematteis, T. Grafke, and E. Vanden-Eijnden, "Rogue waves and large deviations in deep sea," Proc. Natl. Acad. Sci. U.S.A. **115**, 855–860 (2018).
[14] D. A. Plotkin, R. J. Webber, M. E. O'Neill, J. Weare, and D. S. Abbot, "Maximizing simulated tropical cyclone intensity with action minimization," J. Adv. Model. Earth Syst. (published online).
[15] R. A. Pielke, Jr. and M. W. Downton, "Precipitation and damaging floods: Trends in the United States, 1932–97," J. Clim. **13**, 3625–3637 (2000).
[16] R. A. Pielke, Jr., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, "Normalized hurricane damage in the United States: 1900–2005," Nat. Hazards Rev. **9**, 29–42 (2008).
[17] C. Davis, W. Wang, S. S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes *et al.*, "Prediction of landfalling hurricanes with the Advanced Hurricane WRF model," Mon. Weather Rev. **136**, 1990–2005 (2008).

[18]Y. Hirabayashi, S. Kanae, S. Emori, T. Oki, and M. Kimoto, "Global projections of changing risks of floods and droughts in a changing climate," Hydrol. Sci. J. **53**, 754–772 (2008).

[19]M. Kalos, "Monte Carlo calculations of the ground state of three- and four-body nuclei," Phys. Rev. **128**, 1791–1795 (1962).

[20]R. Grimm and R. Storer, "Monte-Carlo solution of Schrödinger's equation," J. Comput. Phys. **7**, 134–156 (1971).

[21]C. Giardina, J. Kurchan, and L. Peliti, "Direct evaluation of large-deviation functions," Phys. Rev. Lett. **96**, 1–4 (2006).

[22]M. Hairer and J. Weare, "Improved diffusion Monte Carlo," Commun. Pure Appl. Math. **67**, 1995–2021 (2014).

[23]F. Cérou, P. Del Moral, F. LeGland, and P. Lezaud, "Genetic genealogical models in rare event analysis," Alea **1**, 181–203 (2006).

[24]P. Del Moral and J. Garnier, "Genealogical particle analysis of rare events," Ann. Appl. Probab. **15**, 2496–2534 (2005).

[25]J. Wouters and F. Bouchet, "Rare event computation in deterministic chaotic systems using genealogical particle analysis," J. Phys. A Math. Theor. **49**, 1–24 (2016).

[26]P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications* (Springer Science & Business Media, 2004).

[27]R. J. Webber, "Unifying Sequential Monte Carlo with resampling matrices," e-print arXiv:1903.12583 [math.NA] (2019).

[28]G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," J. Comput. Graph. Stat. **5**, 1–25 (1996).

[29]S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory* (Springer Science & Business Media, 1998), Vol. 1.

[30]F. Cérou and A. Guyader, "Adaptive multilevel splitting for rare event analysis," Stoch. Anal. Appl. **25**, 417–443 (2007).

[31]N. Guttenberg, A. R. Dinner, and J. Weare, "Steered transition path sampling," J. Chem. Phys. **136**, 1–11 (2012).

[32]J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer Science & Business Media, 2008).

[33]H. P. Chan, T. L. Lai *et al.,* "A general theory of particle filters in hidden Markov models and some applications," Ann. Stat. **41**, 2877–2904 (2013).

[34]D. Coumou and S. Rahmstorf, "A decade of weather extremes," Nat. Clim. Change **2**, 491–496 (2012).

[35]T. R. Knutson, J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. Srivastava, and M. Sugi, "Tropical cyclones and climate change," Nat. Geosci. **3**, 157–2010 (2010).

[36]P. Pall, T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. Hilberts, D. Lohmann, and M. R. Allen, "Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000," Nature **470**, 382–385 (2011).

[37]S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer *et al.,* "The NCEP climate forecast system reanalysis," Bull. Am. Meteorol. Soc. **91**, 1015–1058 (2010).

[38]D. Maraun, F. Wetterhall, A. Ireson, R. Chandler, E. Kendon, M. Widmann, S. Brienen, H. Rust, T. Sauter, M. Themeßl *et al.,* "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," Rev. Geophys. **48**, 1–34 (2010).

[39]K. A. Emanuel, "Toward a general theory of hurricanes," Am. Sci. **76**, 370–379 (1988).

[40]H. M. Fritz, C. D. Blount, S. Thwin, M. K. Thu, and N. Chan, "Cyclone Nargis storm surge in Myanmar," Nat. Geosci. **2**, 448 (2009).

[41]L. A. Bakkensen and R. O. Mendelsohn, "Risk and adaptation: Evidence from global hurricane damages and fatalities," J. Assoc. Environ. Res. Econ. **3**, 555–587 (2016).

[42]W. Langewiesche, "The last words on the bridge," Vanity Fair **60**, 92–131 (2018).

[43]W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X. Huang, W. Wang, and J. Powers, "A description of the advanced research WRF version 3: NCAR/TN-475," STR, NCAR Technical Note, NCAR, Boulder (2008).

[44]F. Judt, S. S. Chen, and J. Berner, "Predictability of tropical cyclone intensity: Scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles," Q. J. R. Meteorol. Soc. **142**, 43–57 (2016).

[45]J. Berner, S.-Y. Ha, J. Hacker, A. Fournier, and C. Snyder, "Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations," Mon. Weather Rev. **139**, 1972–1995 (2011).

[46]C. W. Landsea and J. L. Franklin, "Atlantic hurricane database uncertainty and presentation of a new database format," Mon. Weather Rev. **141**, 3576–3592 (2013).

[47]D. R. Chavas, K. A. Reed, and J. A. Knaff, "Physical understanding of the tropical cyclone wind-pressure relationship," Nat. Commun. **8**, 1–11 (2017).

[48]A. R. Zhai and J. H. Jiang, "Dependence of US hurricane economic loss on maximum wind speed and storm size," Environ. Res. Lett. **9**, 1–9 (2014).

[49]Office of the Federal Coordinator for Meteorological Services and Supporting Research: National hurricane operations plan: FCM-P12-2012, Technical Report, National Oceanic and Atmospheric Administration, 2010.

[50]C.-Y. Lee, M. K. Tippett, A. H. Sobel, and S. J. Camargo, "Rapid intensification and the bimodal distribution of tropical cyclone intensity," Nat. Commun. **7**, 1–5 (2016).

[51]M. T. Montgomery and R. K. Smith, "Paradigms for tropical cyclone intensification," Aust. Meteorol. Oceanographic J. **64**, 37–66 (2014).

[52]J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. de la Cámara, H. M. Christensen, M. Colangeli, D. R. Coleman, D. Crommelin, S. I. Dolaptchiev *et al.,* "Stochastic parameterization: Toward a new view of weather and climate models," Bull. Am. Meteorol. Soc. **98**, 565–588 (2017).

[53]K. Emanuel, R. Sundararajan, and J. Williams, "Hurricanes and global warming: Results from downscaling IPCC AR4 simulations," Bull. Am. Meteorol. Soc. **89**, 347–368 (2008).