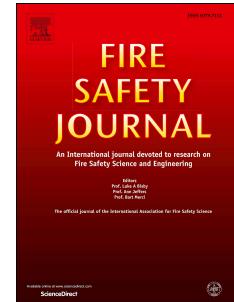


Journal Pre-proof

Quantifying rare events in spotting: How far do wildfires spread?

Alexander Mendez, Mohammad Farazmand



PII: S0379-7112(22)00108-4

DOI: <https://doi.org/10.1016/j.firesaf.2022.103630>

Reference: FISJ 103630

To appear in: *Fire Safety Journal*

Received Date: 2 May 2022

Revised Date: 17 June 2022

Accepted Date: 28 June 2022

Please cite this article as: A. Mendez, M. Farazmand, Quantifying rare events in spotting: How far do wildfires spread?, *Fire Safety Journal* (2022), doi: <https://doi.org/10.1016/j.firesaf.2022.103630>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd.

Quantifying rare events in spotting: How far do wildfires spread?

Alexander Mendez^a, Mohammad Farazmand^{a,*}

^aDepartment of Mathematics, North Carolina State University, Raleigh, NC, 27606, USA

ARTICLE INFO

Keywords:

Firebrand transport
Spotting
Monte Carlo method
Rare event simulation
Large deviation theory

ABSTRACT

Spotting refers to the transport of burning pieces of firebrand by wind which, at the time of landing, may ignite new fires beyond the direct ignition zone of the main fire. Spot fires that occur far from the original burn unit are rare but have consequential ramifications since their prediction and control remains challenging. To facilitate their prediction, we examine three methods for quantifying the landing distribution of firebrands: crude Monte Carlo simulations, importance sampling, and large deviation theory (LDT). In particular, we propose an LDT method that accurately and parsimoniously quantifies the low probability events at the tail of the landing distribution. In contrast, Monte Carlo and importance sampling methods are most efficient in quantifying the high probability landing distances near the mode of the distribution. However, they become computationally intractable for quantifying the tail of the distribution due to the large sample size required. We also show that the most probable landing distance grows linearly with the mean characteristic velocity of the wind field. Furthermore, defining the relative landed mass as the proportion of mass landed at a given distance from the main fire, we derive an explicit formula which allows computing this quantity as a function of the landing distribution at a negligible computational cost. We numerically demonstrate our findings on two prescribed wind fields.

1. Introduction

As the amount of human dwellings near forest fire danger zones increases and as climate change results in more amenable conditions to the creation of fires, there is a greater urgency to predict wildfire dynamics and in particular spot fires [1, 2, 3, 4]. Forest fires generate burning pieces of vegetation, called firebrands, and launch them into the air through columns of gas produced by the fire. After lofting into the air, firebrands are taken away by the ambient wind. Once landed, these burning firebrands can start separate fires away from the original fire. The secondary fires are called spot fires and the entire process is referred to as spotting [5, 6, 7].

The process of spotting consists of three broad phases [6]. The first is the generation of the firebrands by the main fire. The second is the transport of the firebrands by wind. The third stage occurs after the firebrands land on the ground hence possibly igniting fuels, such as shrubs, dead leaves, and branches, at landing sites (see figure 1).

In this paper, we focus on the transport stage of the spotting process, particularly when the wind carries the firebrand to its landing location. In particular, given any firebrand-producing fire, we quantify the probability that a firebrand will land at a certain distance from the primary burn unit. We are particularly interested in accurately approximating the probability that a firebrand will land at a distance far away from the original fire, i.e., the tail of the spotting distribution. These are rare but consequential events since spot fires that start far away from the main fire can damage ostensibly safe areas and catch emergency personnel by surprise [6, 8].

Firebrand transport follows complex nonlinear dynamics arising from the coupling between the firebrand trajectory and the atmospheric flow. Computational fluid dynamics (CFD) models generate the atmospheric flow that transports the firebrands. The firebrand motion is then determined by solving the ordinary differential equations (ODEs) which govern the motion of inertial particles. Monte Carlo simulations of this coupled model can in principle be used to quantify the spotting distribution. However, it is well-known, and we show here in the context of spotting, that Monte Carlo methods with a moderate sample size return only a crude approximation of rare low probability events corresponding to the tail of the distribution [9, 10]. In order to obtain a reasonable approximation of the tail an extremely large sample size is required. Given that CFD models are computationally expensive, such Monte Carlo simulations are not practical.

*Corresponding author.

 amendez6@ncsu.edu (A. Mendez); farazmand@ncsu.edu (M. Farazmand)

ORCID(s): 0000-0003-1764-055X (A. Mendez); 0000-0001-5836-6425 (M. Farazmand)

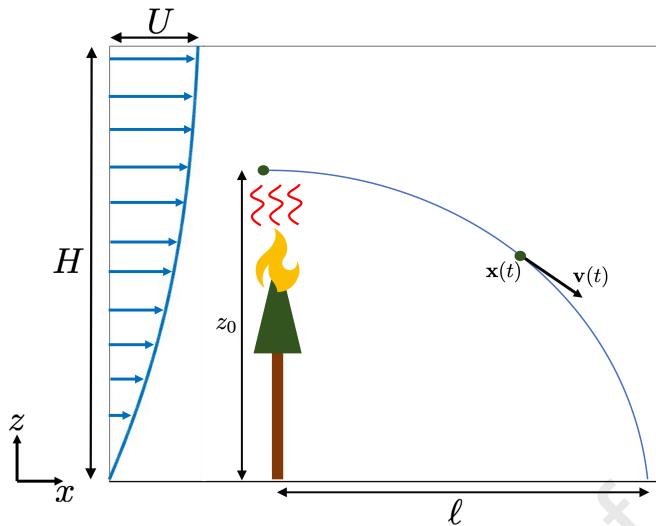


Figure 1: Schematic of firebrand transport from a crown fire. The wind field has asymptotic wind velocity U , which is roughly reached at boundary layer height H . The firebrand is carried up by the fire's plume to initial height z_0 , where it then travels due to the wind with position $\mathbf{x}(t)$ and velocity $\mathbf{v}(t)$. The firebrand eventually lands at landing location ℓ .

Here, we explore two alternatives: Importance sampling and large deviation theory (LDT). Importance sampling biases the sampling distribution in favor of rare events, hence quantifying the tail of the distribution more accurately while using fewer samples as compared to Monte Carlo methods [9]. In contrast, LDT does not require sampling at all; instead, it approximates the tail distribution by an asymptotic expansion determined based on an appropriately defined rate function. To evaluate the rate function, one needs to solve an optimization problem which constitutes the main computational cost of LDT. Nonetheless, since LDT does not require sampling, its computational cost is much lower than Monte Carlo and importance sampling methods. Using two prescribed wind fields, we carry out an extensive study, investigating the strengths and drawbacks of Monte Carlo simulation, importance sampling and LDT. While we use classic Monte Carlo and importance methods, the appropriate LDT method was only recently developed [11, 12, 13] with its formulation and application to spotting presented here for the first time.

1.1. Related work

Much of wildfire research has focused on the physics of coupling between the atmosphere and the fire which involves fluid dynamics, heat transfer and combustion [14, 15, 16, 17, 18]. The resulting CFD models have culminated in comprehensive software packages such as HIGRAD/FIRETEC [19], QUIC-FIRE [20], and WRF-SFIRE [21].

The study of firebrand trajectories can be traced back to the work of Tarifa et al. [7, 22] on the maximum spotting distance of a single firebrand. The maximum spotting distance is the measurement of how far a firebrand can travel between its initial lofting to when it fully burns out. Using a force balance argument, Tarifa et al. [7, 22] used two-dimensional equations of firebrand transport for disk-shaped and cylindrical firebrands.

In particular, Tarifa et al. [22] observed that, in a steady and laminar wind field, firebrands reach their terminal velocity after a very short period of time. Consequently, the authors derived an approximation of maximum spotting distance by assuming that a firebrand always moves with its terminal velocity. This assumption subsequently became a common approximation in firebrand research.

Ever since the pioneering work of Tarifa et al. [7, 22], many similar studies have been conducted which improve or build upon it [23, 24, 25, 26, 27]. In particular, Koo et al. [19] cast doubt on the terminal velocity assumption. They point out that, although this simplifying assumption is reasonably valid for laminar and steady wind fields, it fails to hold true for more realistic unsteady and inhomogeneous turbulent flows. More specifically, Koo et al. [19] find that, in turbulent flows, firebrands without the terminal velocity assumption travel significantly farther than those with this simplifying approximation.

Furthermore, we note that the maximum spotting distance only quantifies the farthest distance from the original burn unit where a secondary fire can potentially start. In contrast, here we study the distributions of landing distance

and relative landed mass which quantify the probability of spot fires at any distance from the original burn unit. As such, our approach is more comprehensive than the maximum spotting distance.

To avoid the computationally expensive firebrand evolution, alternative modeling methods have been proposed [25]. For instance, cellular automaton (CA) models [28, 29, 30, 31] seek to estimate the fire propagation on a specially discrete grid. CA models specify local rules for evolving the fire in discrete time. A similar but continuum approach was proposed by Hillen et al. [32] who developed a non-local transport equation for spotting. Their model expresses the likelihood of fire at a particular time and location in terms of a partial integrodifferential equation (also see Ref. [33]).

Both CA and non-local transport models make several simplifying assumptions, most notably about the wind field and the firebrand trajectories. Here instead, we estimate the landing distribution of firebrands by directly solving for their trajectories in a given wind field. The main objective of the present work is to investigate efficient statistical quantification methods for spotting distribution which can ultimately be integrated into existing high-fidelity software packages, such as HIGRAD/FIRETEC, QUIC-FIRE, and WRF-SFIRE. Our main focus is on the tail of the distribution whose statistical quantification presents a major challenge. We expect that our findings, specially the accurate quantification of the tail, will also inform and improve existing cellular automaton and non-local transport models.

1.2. Outline

This paper is organized as follows. In section 2, we review the equations of motion for a spherical firebrand, along with the characterization of the surrounding wind field. Section 3 compares the methods of Monte Carlo simulation, importance sampling, and large deviation theory in approximating the tail end of the firebrand landing distribution. In section 4, we define the relative landed mass and derive a formula relating its distribution to the landing distribution. Section 5 examines the numerical results found by using each method, along with a discussion of their computational cost. We conclude and summarize our results in section 6.

2. Spotting model

2.1. Firebrand transport model

The motion of a firebrand is governed by the interaction of gravitational and aerodynamic forces acting on it. To describe this motion, we denote the position of a firebrand at time t by $\mathbf{x}(t) = (x(t), z(t))$ and its velocity by $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$. We denote the wind velocity field by $\mathbf{u}(\mathbf{x}, t)$. For simplicity, here we assume that the firebrand moves in a two-dimensional plane, but the transport model is also valid in three dimensions. Then the equations of motion for a spherical firebrand are given by [34, 35, 36, 37, 38, 39],

$$\begin{aligned}
 m(t)\dot{\mathbf{v}} - \dot{m}(t)\mathbf{v}_{rel} &= \frac{1}{2}\rho_f A_c C_d |\mathbf{u} - \mathbf{v}|(\mathbf{u} - \mathbf{v}) && \text{(Quadratic drag)} \\
 &\quad - (m(t) - \rho_f V)\mathbf{g} && \text{(Gravity & Buoyancy)} \\
 &\quad + \rho_f V \frac{D\mathbf{u}}{Dt} && \text{(Pressure gradient)} \\
 &\quad - \frac{1}{2}\rho_f V(\dot{\mathbf{v}} - \frac{D\mathbf{u}}{Dt}). && \text{(Added mass)} \tag{1}
 \end{aligned}$$

We denote the fluid density by ρ_f , the drag coefficient by C_d , cross sectional area of the firebrand by A_c , firebrand mass by $m(t)$, and firebrand volume by $V = 4\pi r^3/3$ where r denotes the firebrand radius. Table 1 contains all parameters, their units, and numerical values used here. The right-hand side represents various forces exerted on the firebrand. The first term represents the empirical law of quadratic drag force. The second term represents gravitational and buoyancy forces. The third term accounts for the pressure gradient exerted by the undisturbed fluid. Finally, the fourth term is the added mass effect as a result of the acceleration of the firebrand with respect to the fluid.

The left-hand side of equation (1) represents the rate of change of momentum for a combusting firebrand [40, 41]. The vector $\mathbf{v}_{rel}(t)$ represents the velocity of matter leaving the firebrand relative to its center of mass. Following [18], we assume that, as the firebrand combusts, burnt matter leaves it isotropically in all directions. Therefore, it is reasonable to assume that $\mathbf{v}_{rel}(t) = 0$ for all time t . The fluid density ρ_f is typically much smaller than the firebrand density ρ_p , which implies that the buoyancy, pressure gradient, and added mass terms are negligible.

Variable	Physical meaning	Value	Units
ρ_f	Fluid density	1.204	kg/m^3
ρ_p	Firebrand density	513	kg/m^3
A_c	Cross sectional area	—	m^2
C_d	Drag coefficient	0.45	—
V	Firebrand volume	—	m^3
U	Boundary layer velocity	5	m/s
H	Boundary layer height	25	m
v_*	Friction velocity	0.7	m/s
κ	von Karman's constant	0.4	—
ϵ	Surface vegetation height	0.05	m
η	Combustion constant	2.86×10^{-4}	s^{-2}
$m(t)$	Firebrand mass	—	kg
$p(r)$	Nominal size distribution	—	—
$q(r)$	Proposal size distribution	—	—
$f_L(\ell)$	Landing distribution	—	—
$f_M(\ell)$	Landed mass distribution	—	—
$g(\ell)$	Relative landed mass distribution	—	—

Table 1
Model parameters and their physical dimensions.

The simplified equations of motion are then given by

$$\dot{\mathbf{x}} = \mathbf{v}, \quad m(t)\dot{\mathbf{v}} = \frac{1}{2}\rho_f A_c C_d |\mathbf{u}(\mathbf{x}, t) - \mathbf{v}|(\mathbf{u}(\mathbf{x}, t) - \mathbf{v}) - m(t)\mathbf{g}, \quad (2)$$

supplied with the initial conditions $\mathbf{x}(0) = \mathbf{x}_0$ and $\mathbf{v}(0) = \mathbf{v}_0$. Here, $\mathbf{g} = (0, 1)^\top$ where $g = 9.8 \text{ m/s}^2$ is the constant gravitational acceleration. As the firebrands are lofted into the air, they create a non-localized distribution of initial positions \mathbf{x}_0 and velocities \mathbf{v}_0 . This distribution depends on the height of the canopy, the convective plume, and the size of the firebrand. However, following Bhutia et al. [15], we make the simplifying assumption that firebrand transport begins from a point source located 50 meters above the origin, and the firebrand is initially motionless. This corresponds to initial position $\mathbf{x}_0 = (0, 50)$ and initial velocity $\mathbf{v}_0 = (0, 0)$.

Of course, the wind velocity field $\mathbf{u}(\mathbf{x}, t)$ in equation (2) also needs to be supplied. In CFD packages such as HIGRAD/FIRETEC, the wind is obtained by solving the relevant Navier–Stokes equation. This constitutes the most computationally expensive part of firebrand trajectory computation, and by extension, the most expensive part of spotting distance estimation. To avoid this computational cost and focus our attention on quantifying the spotting distribution, we use analytically prescribed wind fields. A common choice in firebrand research is a logarithmic wind profile [24, 15, 16, 33], given by

$$\mathbf{u}(\mathbf{x}) := \begin{pmatrix} \frac{v_*}{\kappa} \ln(\frac{z}{\epsilon}) \\ 0 \end{pmatrix}, \quad (3)$$

where v_* denotes the friction velocity, κ is the von Karman's constant, and ϵ denotes surface roughness length scale, describing the surface vegetation height. In addition to this logarithmic velocity, we also use the bounded wind field,

$$\mathbf{u}(\mathbf{x}) := \begin{pmatrix} U \left(1 - \left(\tanh \left(\frac{z}{H} \right) - 1 \right)^2 \right) \\ 0 \end{pmatrix}. \quad (4)$$

The horizontal component of this velocity field approaches the free stream velocity U as the height z approaches the boundary layer thickness H . For $z > H$, the horizontal wind speed remains approximately constant at U . We report all our results for both logarithmic and hyperbolic tangent profiles described above. The numerical value of all parameters are reported in Table 1.

2.2. Combustion model

While firebrands are in flight, they simultaneously burn, resulting in a time-varying mass $m(t)$. Martin [42] gives a detailed overview of combustion models of firebrands in flight.

The simplest combustion model assumes that the loss of mass is linear in time, $m(t) = m_0 - ct$, where m_0 is the initial mass of the firebrand and c is the combustion rate. This model has a major flaw in that the mass of the firebrand becomes negative for a large enough value of time t . Tarifa et al. [22] proposed the empirical model,

$$\rho_p(t) = \frac{\rho_p(0)}{1 + \eta t^2}, \quad (5)$$

where the firebrand density at time t is given by $\rho_p(t)$ and $\eta = 2.86 \times 10^{-4}$ is a constant determined by analyzing firebrand experiments. Assuming that the shape of the firebrand remains the same during combustion [42], equation (5) can be written in the equivalent form,

$$m(t) = \frac{m(0)}{1 + \eta t^2}, \quad (6)$$

using the fact that $m(t) = V \rho_p(t)$ where V denotes the firebrand volume.

There exist more complex models for combustion. For instance, Tse and Fernandez-Pello [24] developed a model that used Nusselt's shrinking drop theory to compute the change in particle diameter of a burning firebrand. Their model is in agreement with the experimental results of Tarifa et al [22].

Albini [5] used a model that assumed that the mass loss rate due to combustion is proportional to the rate of the supply of air to the surface of the firebrand. In this paper, we use the empirical model (6), although more complex combustion models can be used with no significant change to the methods introduced in section 3.

2.3. Assumptions

We made a number of simplifying assumptions in sections 2.1 and 2.2. Most of these assumptions are justified and do not significantly alter the results. For instance, we neglected the buoyancy, added mass and pressure gradient effects in equation (1) based on the fact that the density of the firebrand is much larger than the fluid density. This assumption is routine and it is justified since inclusion of the neglected forces does not significantly change the firebrand trajectories.

However, there are some crucial assumptions made to reduce the computational cost. Although, these assumptions are made for this first study of rare events in spotting, they should ultimately be relaxed in future studies. We list these crucial assumptions below.

1. Point sources: We assume that the firebrands are released from a point source at \mathbf{x}_0 and with the deterministic velocity \mathbf{v}_0 . In truth, the initial position and velocity of the firebrands themselves are random variables.
2. Spherical firebrands: We assume that all firebrands are spherical with variable radius size. In reality, firebrands come in various complex shapes. Earlier studies have focused primarily on spherical, cylindrical, and disk-shaped firebrands [37].
3. Prescribed steady two-dimensional wind: We assumed that the firebrands move in a two-dimensional plane aligned with the predominant direction of the wind. Furthermore, we consider two prescribed steady and laminar velocity fields. In reality, the wind velocity is turbulent and three-dimensional, occurring in areas with obstacles and complex topography.

Although we make the above simplifying assumptions, the methodology and our main findings are applicable to more complex flows with turbulent fluctuations and random shape and initial conditions of the firebrands.

3. Rare event quantification

In this section, we review three methods for quantifying the spotting distribution with a special focus on the tail of the distribution where the firebrands land farthest from the original fire. One of the main factors that determines the landing distance is the size of the firebrand, i.e., the radius r of a spherical firebrand. We treat this radius as a random variable R which is distributed according to a known probability density $p(r)$.

Given firebrands whose radii are a random variable, our goal is to determine the probability distribution of the landing locations and to find the distribution of the relative landed mass as a function of space.

Algorithm 1: MC approximation

```

1 Inputs: Sample size  $N_{MC}$  and number of intervals  $K$ .
2 for  $i = 1, \dots, N_{MC}$  do
3   | Generate firebrand sizes  $R_i$  from distribution  $p$ .
4   | Advect each particle using equation (2) to obtain the corresponding landing distances  $L_i$ .
5 end
6 Divide the landing interval  $[0, \max_i L_i]$  into  $K$  equisized intervals  $D(\ell_j)$  with centers  $\{\ell_j\}_{j=1}^K$  and widths  $\Delta\ell$ .
7 for  $j = 1, \dots, K$  do
8   | For landing distance of interest  $\ell_j$ , use equation (9) to approximate  $\mathbb{P}(L(R) \in D(\ell_j))$ .
9   | Estimate the landing PDF  $f_L(\ell_j)$  using approximation (7).
10 end
11 Output: Landing distribution  $f_L$ 

```

3.1. Monte Carlo method

The most straightforward method for estimating the spotting distribution is the crude Monte Carlo (MC) method. To describe this method, we first define the map $L : \mathbb{R} \rightarrow \mathbb{R}$ that maps the firebrand radius R to a landing distance $L(R)$ obtained by using equation (2) to advect firebrands. Note that $L(R) = x(t_*)$ for a firebrand with initial mass $m(0) = (4\pi R^3/3)\rho_p$, where t_* is the time it takes for the firebrand to land, so that $z(t_*) = 0$.

Given the probability distribution of the radii R , we want to estimate the probability that a firebrand lands at a distance ℓ from the source. More precisely, consider the interval $D(\ell, \Delta\ell) := [\ell - \Delta\ell/2, \ell + \Delta\ell/2]$ centered at ℓ with a small length $\Delta\ell$. We would like to estimate the probability $\mathbb{P}(L(R) \in D(\ell, \Delta\ell))$, which is the probability that a firebrand lands in the interval $D(\ell, \Delta\ell)$. In the following, we use the shorthand $D(\ell)$ in place of $D(\ell, \Delta\ell)$. If f_L is the probability density function (PDF) associated with the random variable $L(R)$, we have

$$f_L(\ell) \simeq \frac{\mathbb{P}(L(R) \in D(\ell))}{\Delta\ell}. \quad (7)$$

We express the probability in terms of the integral,

$$\begin{aligned} \mathbb{P}(L(R) \in D(\ell)) &= \int_{L^{-1}(D(\ell))} p(r) dr \\ &= \int_0^\infty \mathbb{1}_\ell(L(r)) p(r) dr, \end{aligned} \quad (8)$$

where $\mathbb{1}_\ell(\cdot)$ is shorthand for the indicator function of the set $D(\ell)$,

$$\mathbb{1}_\ell(\hat{\ell}) := \begin{cases} 1, & \hat{\ell} \in D(\ell), \\ 0, & \text{otherwise.} \end{cases}$$

The MC method is a straightforward method for estimating integral (8). Consider N independent, identically distributed (i.i.d.) realizations of the firebrand radii, denoted by R_i for $i = 1, \dots, N$. The corresponding landing distances are given by $L_i := L(R_i)$. We can approximate (8) with the MC estimator,

$$P_{MC}(\ell) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_\ell(L_i) \simeq \mathbb{P}(L(R) \in D(\ell)), \quad (9)$$

where radius realizations R_i are drawn from the distribution p and the firebrands are advected using equation (2) to obtain their corresponding landing distances L_i . The MC estimator P_{MC} computes the ratio of the firebrands that land in the interval $D(\ell)$ to the total number of firebrands.

As seen in its implementation in Algorithm 1, direct MC simulation is straightforward. A sample of N_{MC} firebrands with radii R_i are drawn from the distribution $p(r)$. Each firebrand is evolved separately under equation (2) to obtain its

corresponding landing location $L_i = L(R_i)$. Then the landing interval is divided into K bins with width $\Delta\ell$. Finally, equations (7) and (9) are used to estimate the landing distribution f_L .

The MC estimator is unbiased in the sense that $\mathbb{E}_p[P_{MC}] = \mathbb{P}(L(R) \in D(\ell))$, where \mathbb{E}_p denotes the expected value taken with respect to the probability density p . It is also straightforward to show that the variance of σ^2 of the estimator P_{MC} is given by $P_\ell(1 - P_\ell)/N_{MC}$ [43], where we denoted $\mathbb{P}(L(R) \in D(\ell))$ by P_ℓ . The relative error, defined as the ratio of the standard deviation to the mean, is given by $\sigma/P_\ell = \sqrt{(1 - P_\ell)/N_{MC}P_\ell}$. For rare events where the probability P_ℓ is very small, the relative error is approximately $\sigma/P_\ell \simeq 1/\sqrt{N_{MC}P_\ell}$. In order to obtain a small relative error, one needs to use an exceedingly large sample size N_{MC} . For instance, if the probability is 10^{-6} one needs a sample size of 100 millions to obtain a 10% relative error.

As a result, MC method is not practical for quantifying rare spotting events, i.e., spot fires forming far away from the primary fire. Importance sampling, as reviewed in section 3.2, seeks to alleviate this computational cost.

3.2. Importance sampling

Importance sampling (IS) is a variance reduction method [44]. The basic idea behind IS is to modify the sampling distribution so that more samples are obtained from the low probability tail of the landing distribution f_L . More precisely, we draw samples from a *proposal distribution* q instead of the nominal distribution p of firebrand radii. We seek the proposal distribution which minimizes the variance in estimating $\mathbb{P}(L(R) \in D(\ell))$.

Before specifying the optimal distribution q , note that

$$\begin{aligned}\mathbb{P}(L(R) \in D(\ell)) &= \int_0^\infty \mathbb{1}_\ell(L(r))p(r)dr \\ &= \int_0^\infty \frac{\mathbb{1}_\ell(L(r))p(r)}{q(r)}q(r)dr,\end{aligned}$$

for any proposal distribution q . Of course, for the integrals to be well-defined, we must have $\mathbb{1}_\ell(L(r))p(r) = 0$ when $q(r) = 0$. Then the importance sampling estimator is

$$P_{IS}(\ell) := \frac{1}{N} \sum_{i=1}^N \frac{\mathbb{1}_\ell(L_i)p(R_i)}{q(R_i)}, \quad (10)$$

where radius random variables R_i are now drawn from the proposal distribution q . As in the MC case, the new quantity P_{IS} is an unbiased estimator since $\mathbb{E}_q[P_{IS}] = \mathbb{P}(L(R) \in D(\ell))$, where \mathbb{E}_q is the expected value with respect to the proposal distribution.

The question remains on how to choose the optimal proposal distribution q which minimizes the variance of the estimator $\text{Var}[P_{IS}]$. In general, determining this optimal distribution is laborious. However, if we restrict the admissible class of proposal distributions q to the same type of distribution as the nominal distribution p , the optimal distribution can be identified more easily [43, 45].

More precisely, let q be a lognormal distribution with mean μ_q and variance σ_q^2 , which are potentially different from the mean μ_p and variance σ_p^2 of the firebrand radius distribution p . We would like to determine $\theta := \{\mu_q, \sigma_q\}$ such that the variance of the estimator $\text{Var}[P_{IS}]$ is minimized. The problem is that $\text{Var}[P_{IS}]$ is a priori unknown. To rectify this issue, we estimate this variance by running a relatively small MC simulation to compute

$$V(\theta; \ell) := \frac{1}{N_{IS}} \sum_{i=1}^{N_{IS}} \mathbb{1}_\ell(L_i) \frac{p(R_i)}{q(R_i; \theta)} \simeq \text{Var}[P_{IS}], \quad (11)$$

where $R_i \sim p$ and N_{IS} is the MC sample size. Note that unlike the mean estimator (10), $V(\theta)$ is computed by sampling the radii R_i from the distribution p not the proposal distribution q . Also note that the *likelihood ratio* is given by

$$\frac{p(r)}{q(r; \theta)} = \frac{\sigma_q}{\sigma_p} \exp \left\{ -\frac{1}{2} \left(\frac{(\ln(r) - \mu_p)^2}{\sigma_p^2} \right) + \frac{1}{2} \left(\frac{(\ln(r) - \mu_q)^2}{\sigma_q^2} \right) \right\}, \quad (12)$$

where μ_p and σ_p are the known mean and variance of the firebrand size distribution p , respectively.

Algorithm 2: IS approximation

```

1 Inputs: Sample sizes  $N_{IS}$  and  $\hat{N}_{IS}$ , and number of intervals  $K$ .
2 for  $i = 1, \dots, N_{IS}$  do
3   Generate firebrand sizes  $R_i$  from distribution  $p$ .
4   Advect each particle using equation (2) to obtain corresponding landing distances  $L_i$ .
5 end
6 Divide the landing interval  $[0, \max_i L_i]$  into  $K$  equisized intervals  $D(\ell_j)$  with centers  $\{\ell_j\}_{j=1}^K$  and widths  $\Delta\ell$ .
7 for  $j = 1, \dots, K$  do
8   For landing distances of interest  $\ell_j$  solve problem (13) to obtain  $q$ . for  $k = 1, \dots, \hat{N}_{IS}$  do
9     Generate firebrand sizes  $R_k$  from distribution  $q$ .
10    Advect each particle using equation (2) to obtain corresponding landing distances  $L_k$ .
11  end
12  Calculate (10) to approximate  $\mathbb{P}(\ell_j - \Delta\ell/2 \leq L(R) \leq \ell_j + \Delta\ell/2)$ .
13  Estimate the landing distribution  $f_L(\ell_j)$  using the approximation (7).
14 end
15 Output: Landing distribution  $f_L$ 

```

Then, we solve the optimization problem,

$$\theta_*(\ell) := \arg \min_{\theta} V(\theta; \ell), \quad (13)$$

for each ℓ to obtain the corresponding optimal proposal distribution $q(r; \theta_*)$. Finally, we draw \hat{N}_{IS} samples from this optimal distribution in order to compute the importance sampling estimate (10). Algorithm 2 summarizes the entire IS method.

Note that since the optimal θ_* depends on ℓ , the optimization problem (13) must be solved for each spatial interval $D(\ell)$. If there are K intervals, the IS algorithm requires $N_{IS} + K\hat{N}_{IS}$ samples. Since the variance of the estimator $\text{Var}[P_{IS}]$ is reduced compared to the crude MC method, accurate approximations can be obtained with even small sample sizes N_{IS} and \hat{N}_{IS} . We show this with numerical examples in section 5.

3.3. Large deviation theory

Recall that estimating the landing distribution $f_L(\ell)$ is most demanding for rare events where ℓ is large, or theoretically when $\ell \rightarrow \infty$. Large deviation theory refers to a collection of methods that focus on this asymptotic limit of probability distributions [46, 47, 48, 9]. An LDT method that is best suited for application to spotting was only recently developed by Dematteis et al. [12] (also see [11, 13]). The theory is quite technical and therefore here we only review its essential aspects and formulate it for its application to spotting.

Unlike crude MC and importance sampling, LDT does not rely on sampling. Instead, LDT provides an asymptotic expression for evaluating the probability $\mathbb{P}(L(R) \geq \ell)$ for large values of ℓ . Evaluating the LDT estimate requires solving an optimization problem, but not sampling.

The LDT theory is best described for Gaussian random variables. Therefore, we define log radius $Z := \ln R$ which is a Gaussian random variable since the firebrand radius R is lognormal. The landing distance for a firebrand with log radius Z is given by $L(\exp(Z))$ which, for notational simplicity, we denote by $L(Z)$. LDT predicts that, for large enough ℓ , the probability $\mathbb{P}(L(Z) \geq \ell)$ is approximately given by

$$P_{LD}(\ell) := (2\pi)^{-1/2} \frac{1}{\sqrt{2I(Z^*(\ell))}} \exp(-I(Z^*(\ell))) \simeq \mathbb{P}(L(Z) \geq \ell). \quad (14)$$

where $I : \mathbb{R} \rightarrow \mathbb{R}$ is the so-called rate function,

$$I(z) := \max_{\eta \in \mathbb{R}} [\eta z - \ln T(\eta)], \quad (15)$$

Algorithm 3: LDT approximation

```

1 Inputs: Positive increasing sequence  $\lambda_1 < \lambda_2 < \dots < \lambda_{N_\lambda}$ .
2 for  $i = 1, \dots, N_\lambda$  do
3   Solve optimization problem (18) with  $\lambda = \lambda_i$ .
4   Advect a firebrand of size  $Z^*(\lambda_i)$  to obtain the landing distance  $\ell_i = L(Z^*(\lambda_i))$ .
5   Calculate  $P_{LD}(\ell_i)$  using equation (14) with  $Z^*(\ell_i) = Z^*(\lambda_i)$ .
6 end
7 for  $i = 1, \dots, N_\lambda - 1$  do
8   Estimate the landing PDF at landing distance  $\ell_i$  using

$$f_L(\ell_i) \simeq \frac{P_{LD}(\ell_i) - P_{LD}(\ell_{i+1})}{\ell_{i+1} - \ell_i}.$$

9 end
10 Output: Landing distribution  $f_L$ 

```

and $T(\eta) = \mathbb{E} [\exp(\eta Z)]$ is the moment generating function for the Gaussian random variable Z . The log radius Z^* in equation (14) is the solution to the optimization problem,

$$Z^*(\ell) := \arg \min_{Z \in \Omega(\ell)} I(Z), \quad (16)$$

where $\Omega(\ell) := \{Z \in \mathbb{R} : L(Z) \geq \ell\}$ is the set of all log radii such that the corresponding landing distance $L(z)$ exceeds ℓ .

A few remarks are in order here. First, since Z has a Gaussian distribution, the rate function $I(Z)$ can be computed explicitly. Note that the moment generating function for a Gaussian random variable is given by $T(\eta) = \mathbb{E} [\exp(\eta Z)] = \exp(\eta\mu_0 + \sigma_0^2\eta^2/2)$ where μ_0 and σ_0^2 are the mean and variance of Z , respectively. As a result, the rate function becomes

$$I(z) = \frac{1}{2\sigma_0^2}(z - \mu_0)^2. \quad (17)$$

Therefore, the only optimization required is when finding $Z^*(\ell)$, which involves solving the constrained optimization problem (16). But, as shown by Tong et al. [13], the optimizer lies on the boundary of $\Omega(\ell)$. Therefore, equation (16) can be rewritten as the unconstrained optimization problem,

$$Z^*(\lambda) := \arg \min_{Z \in \mathbb{R}} [I(Z) - \lambda L(Z)], \quad (18)$$

with the Lagrange multiplier $\lambda > 0$.

In practice, we choose a sequence of N_λ Lagrange multipliers $0 < \lambda_1 < \lambda_2 < \dots < \lambda_{N_\lambda}$ and, for each λ_i , solve optimization (18). This then determines a corresponding sequence of log radii $Z^*(\lambda_i)$ and landing distances $\ell_i = L(Z^*(\lambda_i))$. Since the optimizer $Z^*(\ell_i)$ of (16) lies on the boundary of $\Omega(\ell_i)$, it coincides with the optimizer $Z^*(\lambda_i)$ of (18) with $\ell_i = L(Z^*(\lambda_i))$. Larger values of λ_i correspond to more extreme landing locations.

With the optimizers $Z^*(\ell_i)$ at hand, the LDT approximation $P_{LD}(\ell_i)$ can be computed from (14). Note that $P_{LD}(\ell_i) - P_{LD}(\ell_{i+1})$ estimates the probability that a firebrand lands in the interval $[\ell_i, \ell_{i+1}]$. Therefore, the probability density $f_L(\ell_i)$ can be approximated by

$$f_L(\ell_i) \simeq \frac{P_{LD}(\ell_i) - P_{LD}(\ell_{i+1})}{\ell_{i+1} - \ell_i}, \quad (19)$$

as long as $\ell_{i+1} - \ell_i$, or equivalently $\lambda_{i+1} - \lambda_i$, is sufficiently small. Note that $P_{LD}(\ell_i) \geq P_{LD}(\ell_{i+1})$ since $\ell_{i+1} \geq \ell_i$. The entire LDT approximation is summarized in Algorithm 3.

We recall that the LDT method does not require sampling and thus it is computationally less expensive than Monte Carlo and importance sampling methods. However, this speed up comes at a cost. First, the LDT approximation (14) is

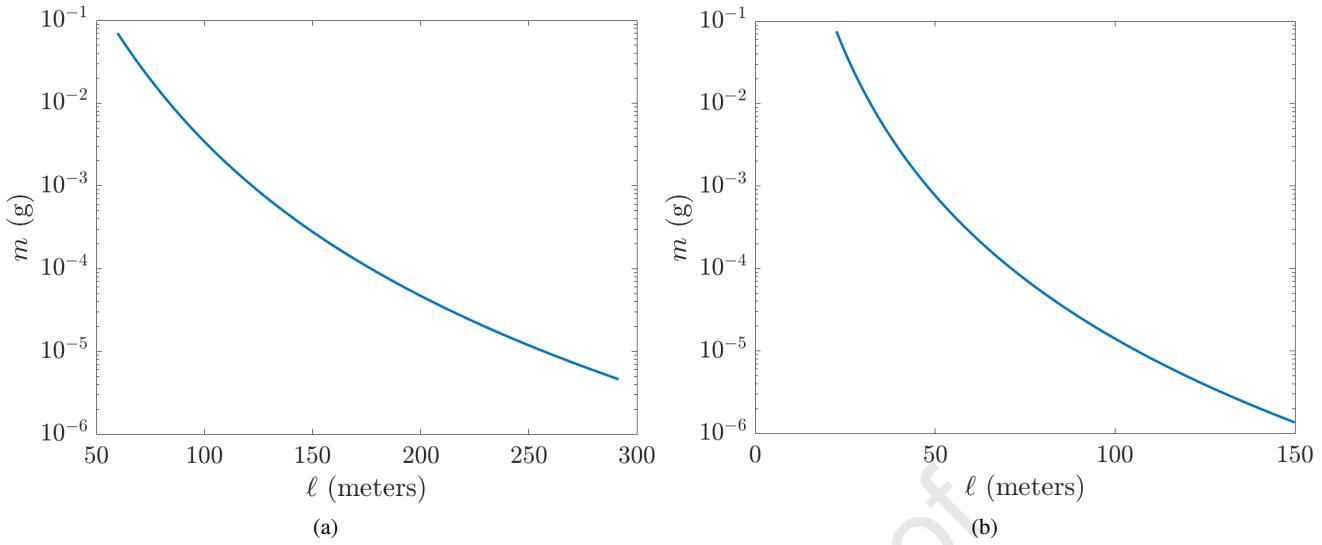


Figure 2: Comparison of final mass functions ϕ for each wind field. (a) Final mass function ϕ corresponding to wind field (3). (b) Final mass function ϕ corresponding to wind field (4).

only valid asymptotically, i.e., for large enough ℓ . As a result, LDT can only be applied to quantifying spotting probabilities at large distances. Furthermore, we are unaware of any estimates on the accuracy of the LDT approximation, e.g., the variance of the LDT approximation.

Nonetheless, as we show in section 5, in our numerical experiments LDT approximation agrees very well with the more costly MC and IS estimates. Furthermore, in our experiments, the range of validity of LDT is not excessively small. In fact, it is accurate for landing locations within one standard deviation from the mode of the distribution.

4. Relative landed mass distribution

The landing distribution f_L , that was estimated in section 3, quantifies the proportion of the firebrands landing at a distance ℓ , regardless of their size or mass. However, burning firebrands with larger mass are more likely to start a fire at their landing location. As a result, it is perhaps more relevant to quantify the proportion of firebrand mass landed in a small interval at distance ℓ from the main fire. We refer to this quantity as the *relative landed mass distribution* and derive an equation that enables us to compute this quantity from the landing distribution f_L .

We denote the probability density associated with the relative landed mass distribution by $g : \mathbb{R} \rightarrow \mathbb{R}^+$ so that $g(\ell)\Delta\ell$ estimates the ratio of the mass landed in the interval $D(\ell)$ to the total mass landed anywhere. More precisely, the density g is given by

$$g(\ell) := \lim_{\Delta\ell \rightarrow 0^+} \frac{1}{\Delta\ell} \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N \mathbb{1}_\ell(L_i) M_i}{\sum_{i=1}^N M_i}, \quad (20)$$

where $L_i = L(R_i)$ and the firebrand radii R_i are drawn from the probability density p . The random variable M_i is the mass of the firebrand with radius R_i at the time of landing, which is computed using the combustion model (6).

We define $\phi : \mathbb{R} \rightarrow \mathbb{R}^+$ as the map between the landing distance and its associated landing mass so that $M_i = \phi(L_i)$. As shown in figure 2, this is a one-to-one decreasing function. This function is monotonically decreasing because the firebrands that land farther are airborne for a longer time and therefore have more time to burn and lose mass.

We note that relative landed mass density $g(\ell)$ should not be confused with the landed mass density $f_M(m)$ of M_i . The quantity $f_M(m)\Delta m$ measures the probability that the landed firebrand mass falls in the mass interval $[m - \Delta m/2, m + \Delta m/2]$. As such, f_M contains no information about the landing location. In contrast, $g(\ell)$ estimates the proportion of mass landed in the spatial interval $[\ell - \Delta\ell/2, \ell + \Delta\ell/2]$.

In principle, the relative landed mass density (20) can be approximated using the MC method by sampling a large number of firebrands, computing their landing locations L_i and the corresponding mass M_i at the landing time.

However, as in the case of the landing distribution f_L , this estimate will be inaccurate for large ℓ unless we use a prohibitively large sample size. To avoid this problem, here we show that

$$g(\ell) = \frac{\phi(\ell)f_L(\ell)}{\mathbb{E}[M]}, \quad (21)$$

where $\phi(\ell)$ is the landing mass of firebrands that land at the distance ℓ (see figure 2), f_L is the landing distribution computed in section 3, and $\mathbb{E}[M]$ is the expected value of the landed mass. The remainder of this section is devoted to proving the estimate (21).

We first rewrite equation (20) by multiplying $\frac{1}{N}$ in the numerator and the denominator which yields

$$g(\ell) = \lim_{\Delta\ell \rightarrow 0^+} \frac{1}{\Delta\ell} \lim_{N \rightarrow \infty} \frac{\frac{1}{N} \sum_{i=1}^N \mathbb{1}_\ell(L_i) \phi(L_i)}{\frac{1}{N} \sum_{i=1}^N \phi(L_i)}, \quad R_i \sim p. \quad (22)$$

The denominator of this expression is the expected value of the landed mass, so that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \phi(L_i) = \mathbb{E}[M] := \int_0^\infty m f_M(m) dm. \quad (23)$$

The numerator is the expected value of the mass that lands in the interval $D(\ell)$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{1}_\ell(L_i) \phi(L_i) = \int_{\phi(D(\ell))}^\infty m f_M(m) dm. \quad (24)$$

Recall that by definition $m = \phi(\ell)$. Therefore, by the change of variable formula, we have

$$\int_{\phi(D(\ell))}^\infty m f_M(m) dm = \int_{D(\ell)} -\phi(\ell) f_M(\phi(\ell)) \phi'(\ell) d\ell, \quad (25)$$

where the minus sign is due to the fact that ϕ is monotonically decreasing.

On the other hand, the probability that a firebrand lands in $D(\ell)$ is equal to the probability that the firebrand mass at the time of landing is in $\phi(D(\ell))$. More precisely, we have

$$\int_{D(\ell)} f_L(\ell) d\ell = \int_{\phi(D(\ell))} f_M(m) dm = - \int_{D(\ell)} f_M(\phi(\ell)) \phi'(\ell) d\ell, \quad (26)$$

where we again used change of variables for the last identity. Since (26) must hold for any arbitrary interval $D(\ell)$, we obtain $f_L(\ell) = -f_M(\phi(\ell)) \phi'(\ell)$. Substituting this in equation (25) and using equation (24), we obtain

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{1}_\ell(L_i) \phi(L_i) = \int_{D(\ell)} \phi(\ell) f_L(\ell) d\ell \simeq \Delta\ell \phi(\ell) f_L(\ell). \quad (27)$$

Finally, combining equations (22), (23) and (27) gives the desired result (21).

5. Numerical results and discussion

In this section, we report our numerical results, comparing MC, IS, and LDT methods as described in section 3. We particularly focus on the trade-off between accuracy and computational cost. We assume that the firebrand radius R is a lognormally distributed random variable, with a mean radius of 0.75 millimeters and a variance of 0.125 millimeters. The corresponding probability density function p is shown in figure 3.

For each method, we present our results corresponding to the logarithmic wind profile (3) and the hyperbolic wind profile (4). Figure 4 shows the wind profiles along with a set of corresponding firebrand trajectories. Smaller firebrands travel farther before landing.

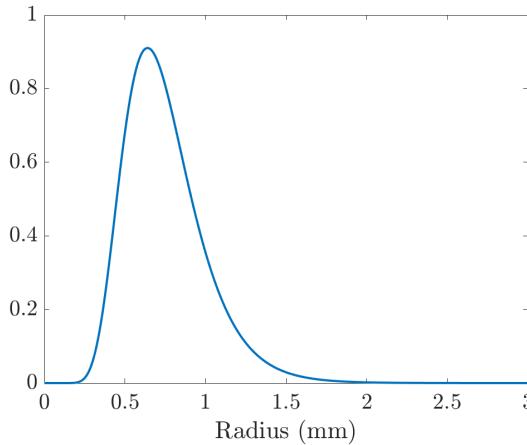


Figure 3: Probability density function of a lognormally distributed firebrand radius R with mean 0.75 millimeters and variance 0.125 millimeters.

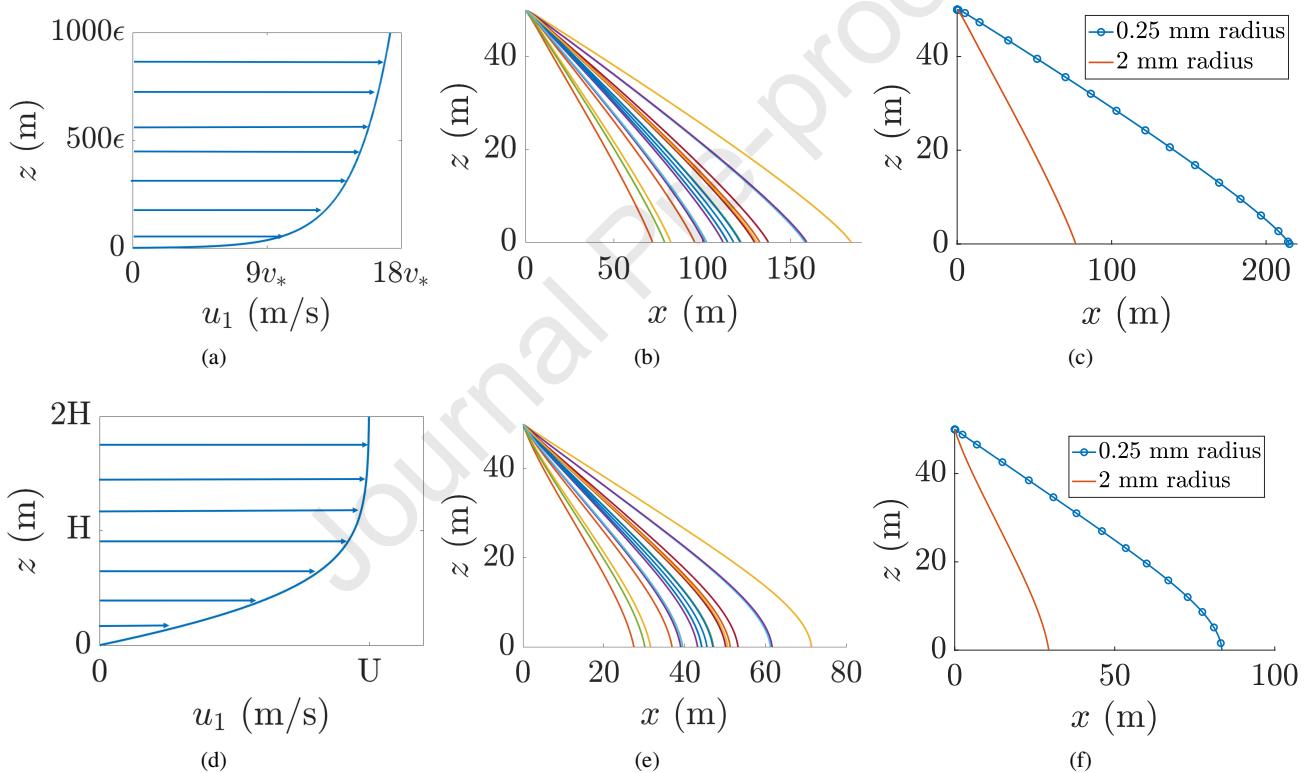


Figure 4: Wind profile and the corresponding firebrand trajectories. (a,b,c) The logarithmic velocity field (3). (d,e,f) The hyperbolic tangent velocity field (4). Panels (b) and (e) show trajectories corresponding to different firebrand radii drawn randomly from the lognormal distribution p .

We begin by comparing the MC, IS, and LDT methods for quantifying the landing distance PDF f_L . Figure 5(a) shows the estimated PDF for the logarithmic wind profile. For landing distances with relatively high probability, roughly between 100 – 180 meters, crude MC provides slightly more accurate approximations as compared to IS. The inset of figure 5(a) shows a closeup view of this high-probability region showing the smaller variance associated with the MC method. This is due to the large number of MC realizations that land in this interval. However, when approximating low probability events at the tails of the distribution, the IS method is more accurate as it exhibits lower variance compared to MC. Both MC and IS methods fail beyond landing distance $\ell = 230$ meters, since no samples reach beyond that landing distance. Since LDT is not based on sampling, it avoids this issue and gives approximations

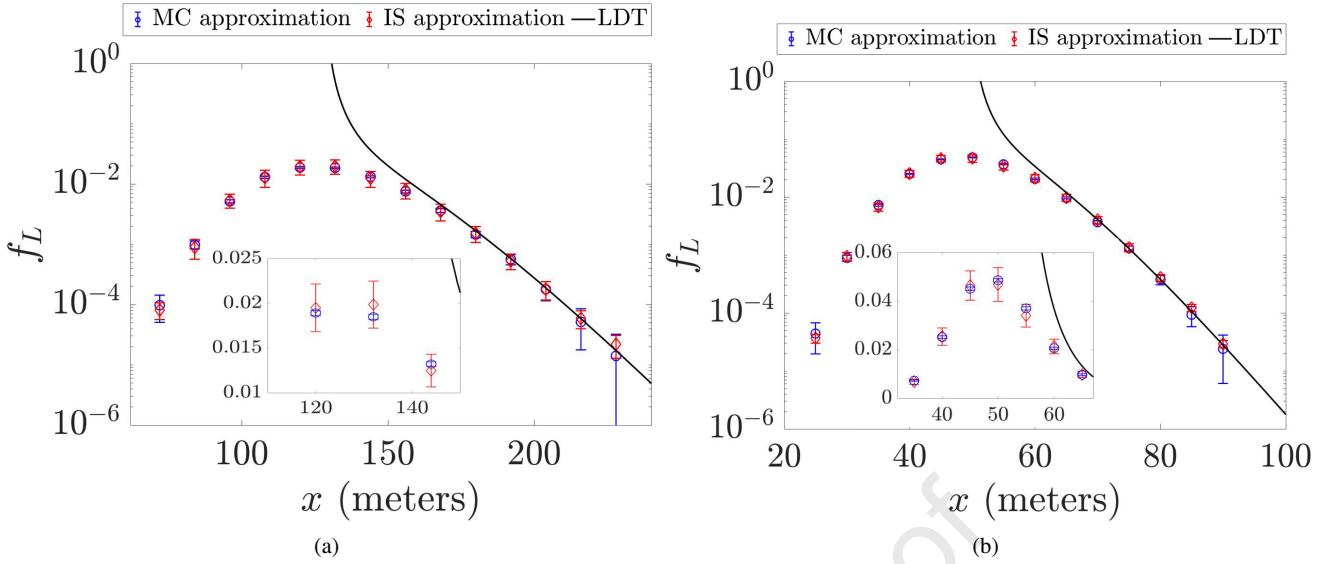


Figure 5: Landing PDF approximated with MC, IS and LDT methods. Bars represent 95 percent confidence intervals. (a) Logarithmic wind. (b) Hyperbolic tangent wind.

past the point where MC and IS fail. However, as noted in section 3.3, LDT is an asymptotic estimate and therefore it is only valid for large enough landing distance ℓ . As a result, it fails near the mode (i.e., high probability region) of the PDF but accurately estimates the tail for $\ell > 170$ meters.

Similar observations are made for landing PDF f_L corresponding to the hyperbolic tangent wind profile as shown in figure 5(b). Again, we see that MC produces an approximation with a smaller variance than IS at landing locations with high probability, but IS approximations have smaller variance towards the tail end of the landing distribution. For $\ell > 90$ meters, there are no samples and therefore both MC and IS methods fail to produce an approximation for the PDF. In contrast, the LDT approximation can easily be extended beyond $\ell = 90$ meters. The asymptotic range over which LDT is valid is over $\ell > 70$ meters.

In order to compare the computational cost of the methods, we count the number of firebrand advections and record the total computational time. The number of advections accounts for the number of samples as well as the advections needed for the optimization steps inside IS and LDT algorithms. All methods were implemented in MATLAB, version 2018a, and executed on 12 cores of Intel Xeon CPU E5-2690 8-core CPUs with 2.90GHz DDR3 RAM. Firebrand trajectories are found by numerically integrating the equations of motion (2) using the Runge–Kutta scheme of ODE45. The optimization problems in the IS and LDT methods are solved using MATLAB’s built-in package fmincon.

MC approximation requires solving equation (2) for N_{MC} realizations of the firebrand size, which are then used in approximation (9). For IS, the first step is to integrate N_{IS} realizations of the firebrand size in order to set up the optimization problem (13). Recall that N_{IS} is much smaller than the MC sample size N_{MC} . Then we choose K landing distances ℓ_i , $i = 1, 2, \dots, K$. For each landing interval $D(\ell_i)$, we solve the optimization problem (13) to obtain a proposal distribution $q(r; \theta_*)$. The average number of optimization iterations are denoted by O_{IS} . Then \hat{N}_{IS} realizations are taken from each proposal distribution q , which are then used to compute the IS approximation (10). For LDT, we need N_λ values of the Lagrange multiplier λ . For each value of λ , we solve the optimization problem (18) which takes an average of O_{LD} iterations to converge. This yields $N_\lambda \times O_{LD}$ advections on average.

Table 2 compares the computational cost of MC, IS and LDT for the hyperbolic tangent wind field (4). The results are similar for the logarithmic wind. For the MC method, we use $n_{MC} = 10^6$ samples width $K = 15$ spatial intervals to estimate the landing PDF. In this sample, firebrands do not travel much farther than $\ell \simeq 90$ meters. For IS method, we use $N_{IS} = 2 \times 10^5$ samples from the nominal distribution p . As in MC, we use $K = 15$ spatial intervals. The corresponding optimization problem takes an average of $O_{IS} = 35$ iterations to converge. Once the proposal distribution q is found, we take $\hat{N}_{IS} = 10^3$ samples for each spatial interval. For the LDT approximation, the number of λ values used in the sequence is $N_\lambda = 296$ and it takes an average of $O_{IS} = 45$ iterations to solve the optimization problem (18) for each value of λ .

Method	Formula for number of advects	Number of advects	Compute time
MC	N_{MC}	1,000,000	1084 sec
IS	$N_{IS} + K(\hat{N}_{IS} + O_{IS})$	215,525	204 + 19 sec
LDT	$N_\lambda \times O_{LD}$	13,294	23 sec

Table 2

Computational cost of the methods for the hyperbolic tangent wind field. The number of samples are denoted by N_{MC} , N_{IS} and \hat{N}_{IS} . The number of intervals is denoted by K , and the number of optimization iterations by O_{IS} and O_{LD} . For LDT, the number of Lagrange multipliers is given by N_λ .

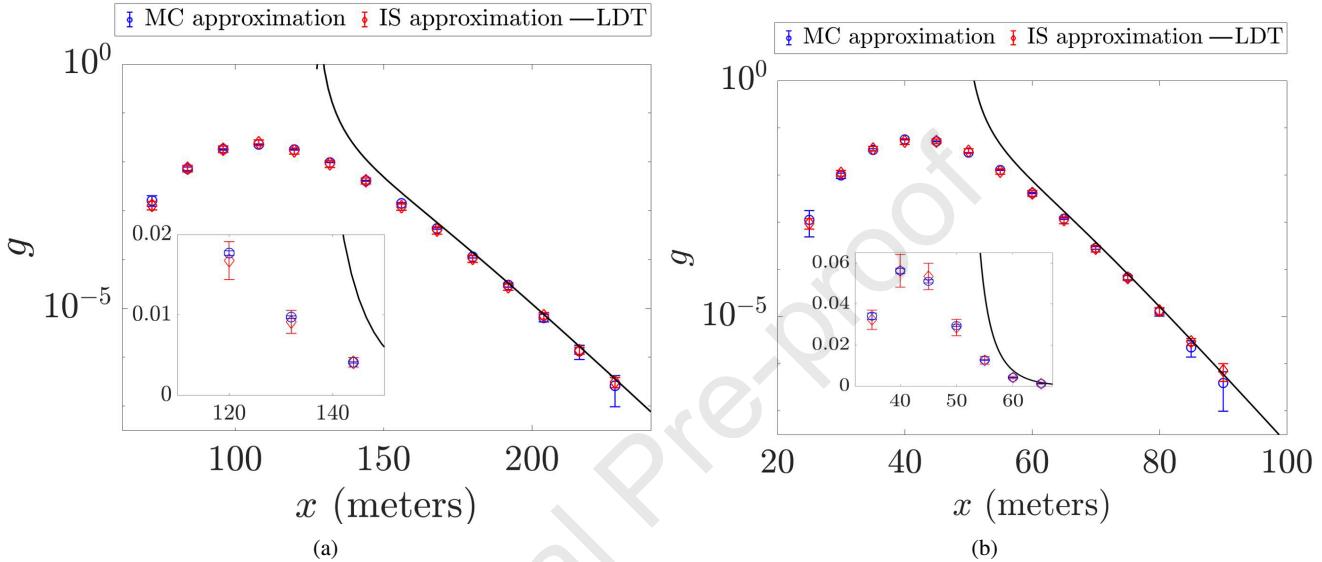


Figure 6: The relative landed mass distribution approximated by MC, IS and LDT methods. Bars represent 95 percent confidence intervals. (a) Logarithmic wind. (b) Hyperbolic tangent wind.

The last column of Table 2 shows the total computational time for each method. The MC simulations are most expensive and take about 18 minutes. We note that this computational time will increase significantly in the realistic situation where the wind field is not available analytically, and a CFD simulation is needed to obtain it. The computational time of IS is an order of magnitude smaller around 3.7 minutes. Note that the main computational cost of IS is associated with advecting the sample from the nominal distribution p which takes about 204 seconds. LDT takes only 23 seconds and therefore is computationally most efficient. However, we reiterate that LDT is only valid for rare events at the tail of the distribution and fails to quantify the most probable events.

As mentioned in section 4, the relative landed mass distribution g can also be computed from the landing distribution f_L ; see equation (21). Figure 6 shows the results for both wind fields. As expected, the relative landed mass distribution g is slightly different from the landing distribution f_L . In particular, the tail of the relative landed mass distribution g decays more rapidly and its mode occurs at a slightly smaller distance as compared to the landing distribution f_L . Both these features are associated with the fact that firebrands which travel a longer distance are airborne for a longer period of time and therefore burn more mass before landing. In terms of the approximation methods, we reach the same conclusions as when approximating the landing distribution. Namely, near the mode of the distribution, MC is more accurate than IS owing to the large number of firebrands that land there. However, near the tail where low probability events occur, IS becomes more accurate than MC. Near the tail, LDT is in excellent agreement with IS results, but it fails near the mode of the distribution.

Finally, we examine the most likely landing location as a function of the characteristic velocity of the wind fields (3) and (4). So far, we have mainly focused on the tail of the distributions since they correspond to rare but consequential spotting events. Although quantifying these rare events is important, the most likely place where a spot fire can form is near the mode of the landing or relative landed mass distributions. Figure 7 shows the most probable landing location

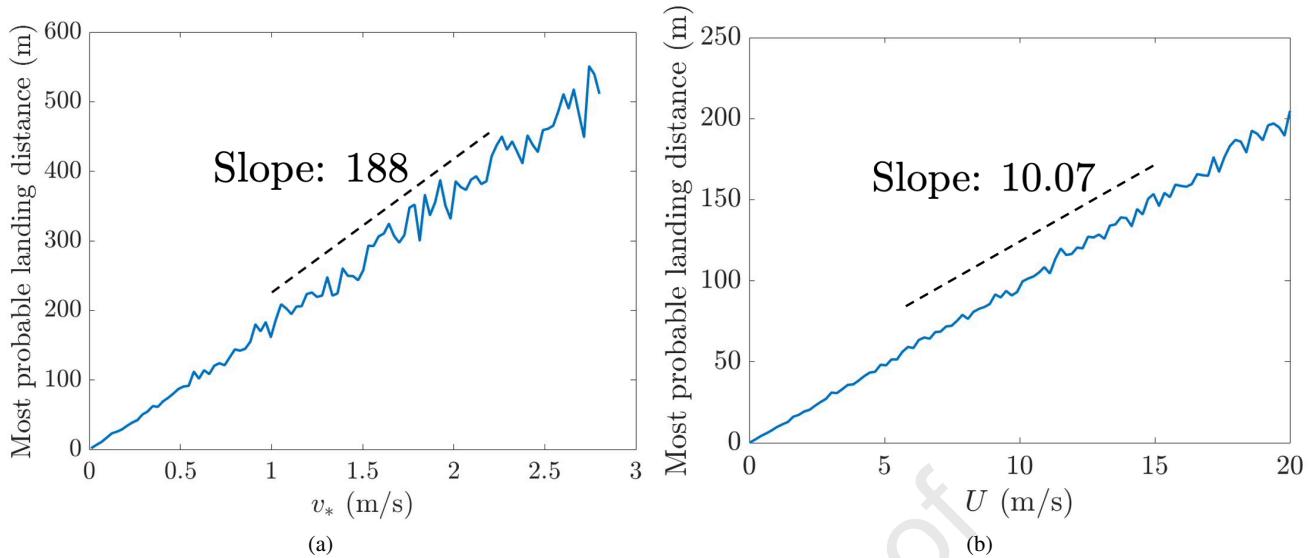


Figure 7: Most probable landing distance as a function of the characteristic velocity. (a) Logarithmic wind where v_* denotes the friction velocity. (b) Hyperbolic tangent wind where U denotes the free stream velocity.

L_m , i.e. mode of the distribution, as a function of the characteristic velocity of the logarithmic and hyperbolic tangent winds. For the logarithmic wind, L_m grows linearly with the friction velocity v_* such that $L_m \simeq 188v_*$. Similarly, for the hyperbolic tangent wind, we see a linear relationship $L_m \simeq 10U$. In both cases, the mode grows linearly with the characteristic velocity which is notable because the equations of motion (2) depend quadratically on the relative velocity $\mathbf{u} - \mathbf{v}$.

6. Conclusions

The main objective of this study was to quantify the fire spotting events by approximating the landing distribution and relative landed mass distribution of firebrands. This is a challenging problem because the resulting distributions have heavy tails, corresponding to rare but consequential spotting events far away from the main fire. We compared three different methods: crude Monte Carlo simulations, importance sampling, and large deviation theory.

The MC method is most expensive and returns the least accurate tail approximations. Since large landing distances are rare, a very large sample size is required to approximate the tail with a reasonable accuracy. Here, for simplicity, we used analytically prescribed wind fields. But in realistic situations, where the wind field is obtained from a CFD model, such large sample sizes are not practical.

We then turned to the variance reduction method of importance sampling. IS uses a small initial MC sample to obtain the nominal distribution. Then the IS proposal distributions are computed by solving an optimization problem. The total number of realizations used for IS was an order of magnitude smaller than MC, yet IS approximations were more accurate in quantifying the tail of the distribution. More precisely, as the landing distance grows, the variance of the MC approximation also grows. However, the variance of IS remains bounded and relatively small. We note that, although IS is more accurate at the tails, the MC method is slightly more accurate near the mode of the distribution owing to the large number of firebrands that land there.

The third method we considered is based on recently formulated large deviation theory for ordinary and stochastic differential equations [12]. LDT offers an asymptotic approximation for the distribution in terms of the so-called rate function. To evaluate the rate function, one needs to solve an associated optimization problem. Unlike MC and IS, LDT does not require any sampling. As a result, LDT is computationally more efficient than both MC and IS, being two and one order of magnitude faster, respectively. Furthermore, MC and IS are both limited by the farthest landing firebrand from a sample such that the tail distribution cannot be approximated beyond this point. But since LDT does not rely on sampling, it can approximate the tail at arbitrarily large distances.

The LDT method has two notable drawbacks. First, being an asymptotic theory, it can only quantify the tail of the distribution and fails to approximate the mode. Second, to the best of our knowledge, the current large deviation theory

is not equipped with error bars to quantify the accuracy of the approximation. Although our results show that LDT results are in excellent agreement with MC and IS results, the accuracy cannot be known a priori.

Given our observations, we recommend a hybrid approach for quantifying spotting distributions where the MC or IS methods are used to quantify the high-probability events near the mode of the distribution. Accurate results can be obtained here even with a relatively small sample size since most firebrands land near the mode. In contrast, LDT method should be used to quantify low probability events at the tails. Since LDT does not require any sampling, it quantifies the tails accurately at a fraction of the computational cost.

We also considered the relative landed mass distribution, which quantifies the proportion of the firebrand mass landed at a distance. Since firebrands burn throughout their flight, the relative landed mass is generally different from the landing distribution. Nonetheless, in section 4, we derived a formula which allowed us to compute the relative landed mass distribution from the landing distribution, at no significant computational cost.

In addition to approximating probability distributions, we also examined the effect of the wind field on the landing distance. Specifically, the relationship between the most probable landing distance and the characteristic velocity of the wind was observed to be linear. This is despite the nonlinearities in the equations of motion.

Future work will focus on relaxing the simplifying assumptions made in this paper (see section 2.3), with the ultimate goal of implementing an efficient spotting quantifier in high-fidelity fire simulators such as HIGRAD/FIRETEC, QUIC-FIRE, and WRF-SFIRE. Existing work on inertial particle transport [6, 49, 50] suggests that, in turbulent flows, firebrands get trapped inside vortices and therefore quantifying their landing distribution may present new challenges that are absent in laminar steady flows. Furthermore, although IS and LDT are in principle applicable to these more complicated situations, the corresponding optimization problems need to be solved on a higher dimensional parameter space involving the firebrand's initial conditions, shape, and turbulent fluctuations. Nonetheless, we still expect that these methods will significantly outperform MC simulations.

CRediT authorship contribution statement

Alexander Mendez: Carried out the research, prepared the results for publication and wrote the first draft of the paper. **Mohammad Farazmand:** Conceptualized and supervised the research, and edited the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to Dr. Joseph O'Brien (USDA Forest Service) for fruitful conversations.

Funding

This work was partially supported by the National Science Foundation grant DMS-1745654.

References

- [1] F. Tedim, V. Leone, M. Amraoui, C. Bouillon, M. R. Coughlan, G. M. Delogu, P. M. Fernandes, C. Ferreira, S. McCaffrey, T. K. McGee, J. Parente, D. Paton, M. G. Pereira, L. M. Ribeiro, D. X. Viegas, and G. Xanthopoulos. Defining extreme wildfire events: Difficulties, challenges, and impacts. *Fire*, 1(1), 2018.
- [2] A. Tohidi and N.B. Kaye. Aerodynamic characterization of rod-like debris with application to firebrand transport. *Journal of Wind Engineering and Industrial Aerodynamics*, 168:297–311, 2017.
- [3] A. Mendez and M. Farazmand. Investigating climate tipping points under various emission reduction and carbon capture scenarios with a stochastic climate model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2256):20210697, 2021.
- [4] M. Burke, A. Driscoll, S. Heft-Neal, J. Xue, J. Burney, and M. Wara. The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences*, 118(2):e2011048118, 2021.
- [5] F. A. Albini. Spot fire distance from burning trees - A predictive model. Technical report, US Forest Service, 1979.
- [6] E. Koo, P. J. Pagni, D. R. Weise, and J. P. Woycheese. Firebrands and spotting ignition in large-scale fires. *International Journal of Wildland Fire*, 19(7):818–843, 2010.

- [7] C. S. Tarifa, P. P del Notario, A. R. Villa, M. L. Martinez, and O. Perez. Open fires and transport of firebrands. Technical report, Instituto Nacional De Tecnica Aeroespacial, 1965.
- [8] J. R. Weir. Using relative humidity to predict spotfire probability on prescribed burns. In: Sosebee, Ronald E.; Wester, David B.; Britton, Carlton M.; McArthur, E. Durant; Kitchen, Stanley G., comps. *Proceedings: Shrubland dynamics–fire and water; 2004 August 10-12; Lubbock, TX. Proceedings RMRS-P-47. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.* p. 69-72., 47, 2007.
- [9] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer Series in Statistics. Springer, New York, USA, 2004.
- [10] M. Farazmand and T. P. Sapsis. Extreme Events: Mechanisms and Prediction. *Applied Mechanics Reviews*, 71(5), 2019.
- [11] G. Dematteis, T. Grafke, and E. Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, 115(5):855–860, 2018.
- [12] G. Dematteis, T. Grafke, and E. Vanden-Eijnden. Extreme event quantification in dynamical systems with random components. *SIAM/ASA Journal on Uncertainty Quantification*, 7(3):1029–1059, 2019.
- [13] S. Tong, E. Vanden-Eijnden, and G. Stadler. Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis. *Communications in Applied Mathematics and Computational Science*, 16(2):181–225, 2021.
- [14] N. Sardoy, J-L. Consalvi, B. Porterie, and A. Fernandez-Pello. Modeling transport and combustion of firebrands from burning trees. *Combustion and Flame*, 150:151–169, 08 2007.
- [15] S. Bhutia, M. Jenkins, and R. Sun. Comparison of firebrand propagation prediction by a plume model and coupled-fire/atmosphere large-eddy simulator. *Journal of Advances in Modeling Earth Systems*, 2(1):4, 2010.
- [16] R. Sun, S. Krueger, M. Jenkins, M. Zulauf, and J. Charney. The importance of fire-atmosphere coupling and boundary-layer turbulence to wildfire spread. *International Journal of Wildland Fire*, 18(50-60), 01 2009.
- [17] C. Anand, B. Shotorban, and S. Mahalingam. Dispersion and deposition of firebrands in a turbulent boundary layer. *International Journal of Multiphase Flow*, 109:98–113, 2018.
- [18] S. L. Manzello, S. Suzuki, M.J. Gollner, and A. C. Fernandez-Pello. Role of firebrand combustion in large outdoor fire spread. *Progress in Energy and Combustion Science*, 76:100801, 2020.
- [19] E. Koo, R. R. Linn, P. J. Pagni, and C. B. Edminster. Modelling firebrand transport in wildfires using HIGRAD/FIRETEC. *International Journal of Wildland Fire*, 21:396–417, 2012.
- [20] R. R. Linn, S. L. Goodrick, S. Brambilla, M. J. Brown, R. S. Middleton, J. J. O'Brien, and J. K. Hiers. QUIC-fire: A fast-running simulation tool for prescribed fire planning. *Environmental Modelling and Software*, 125:104616, 2020.
- [21] J. Mandel, J. D. Beezley, and A. K. Kochanski. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geosci. Model Dev.*, 4:591–610, 2011.
- [22] C. S. Tarifa, P. P. del Notario, and F. G. Moreno. On flight paths and lifetimes of burning particles of wood. *Symposium (International) on Combustion*, 10(1):1021–1037, 1965.
- [23] S. L. Lee and J. M. Hellman. Study of firebrand trajectories in a turbulent swirling natural convection plume. *Combustion and Flame*, 13(6):645–655, 1969.
- [24] S. D. Tse and A. C. Fernandez-Pello. On the flight paths of metal particles and embers generated by power lines in high-winds - A potential source of wildland fires. *Fire Safety Journal*, 30:333–356, 1998.
- [25] Albini F. A., Alexander M. E., and Cruz M. G. A mathematical model for predicting the maximum potential spotting distance from a crown fire. *International Journal of Wildland Fire*, 21:609–627, 2012.
- [26] G. Zheng, A. J. Sedlacek, A. C. Aiken, Y. Feng, T. B. Watson, S. Raveh-Rubin, J. Uin, E. R. Lewis, and J. Wang. Long-range transported North American wildfire aerosols observed in marine boundary layer of eastern North Atlantic. *Environmental International*, 139:105680, 2020.
- [27] J. P. Woycheese, P. J. Pagni, and D. Liepmann. Brand propagation from large-scale fires. *Journal of Fire Protection Engineering*, 10(2):32–44, 1999.
- [28] K. C. Clarke, J. A. Brass, and P. J. Riggan. A cellular automaton model of wildfire propagation and extinction. *Photogrammetric Engineering and Remote Sensing*, 60(11):1355–1367, 1994.
- [29] I. Karafyllidis and A. Thanailakis. A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, 99(1):87–97, 1997.
- [30] D. Boychuk, W. J. Braun, R. J. Kulperger, Z. L. Krougly, and D. A. Stanford. A stochastic model for forest fire growth. *INFOR: Information Systems and Operational Research*, 45(1):9–16, 2007.
- [31] A. Alexandridis, D. Vakalis, C.I. Siettos, and G.V. Bafas. A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through Spetses Island in 1990. *Applied Mathematics and Computation*, 204(1):191–201, 2008.
- [32] T. Hillen, B. Greese, J. Martin, and G. de Vries. Birth-jump processes and application to forest fire spotting. *Journal of Biological Dynamics*, 9(sup1):104–127, 2015.
- [33] J. Martin and T. Hillen. The spotting distribution of wildfires. *Applied Sciences*, 6(6):177, 2016.
- [34] M. R. Maxey and J.J. Riley. Equation of motion for a small rigid sphere in a nonuniform flow. *Phys. Fluids*, 26:883–889, 1983.
- [35] I. Kim, S. Elghobashi, and W. A. Sirignano. On the equation for spherical-particle motion: effect of reynolds and acceleration numbers. *Journal of Fluid Mechanics*, 367:221–253, 1998.
- [36] A. Babiano, J. H. E. Cartwright, O. Piro, and A. Provenzale. Dynamics of a small neutrally buoyant sphere in a fluid and targeting in Hamiltonian systems. *Phys. Rev. Lett.*, 84:5764–5767, 2000.
- [37] R. R. Linn, E. Koo, C. B. Edminster, and J. A. Perry. Using a physics-based model to characterize spotting potential for protection of the wildland urban interface. Technical Report 07-1-5-01, Joint Fire Science Program, 2007.
- [38] M. Farazmand and G. Haller. The Maxey–Riley equation: Existence, uniqueness and regularity of solutions. *Nonlinear Analysis: Real World Applications*, 22:98–106, 2015.

- [39] G. P. Langlois, M. Farazmand, and G. Haller. Asymptotic dynamics of inertial particles with memory. *Journal of nonlinear science*, 25(6):1225–1255, 2015.
- [40] A. Sommerfeld. *Mechanics: Lectures on theoretical physics*, volume 1. Academic press, 1952.
- [41] A. R. Plastino and J. C. Muzzio. On the use and abuse of Newton's second law for variable mass problems. *Celestial Mechanics and Dynamical Astronomy*, 53(3):227–232, 1992.
- [42] J. Martin. *Derivation and investigation of mathematical models for spotting in wildland fire*. PhD thesis, University of Alberta, 2013.
- [43] M. Bee. Importance sampling for sums of lognormal distributions with applications to operational risk. *Communications in Statistics - Simulation and Computation*, 38(5):939–960, 2009.
- [44] T. Kloek and H. K. van Dijk. Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica*, 46(1):1–19, Jan 1978.
- [45] D. L. McLeish and Z. Men. Extreme value importance sampling for rare event risk measurement. *Innovations in Quantitative Risk Management*, 99, 2015.
- [46] S. R. S. Varadhan. *Large deviations and applications*. SIAM, Philadelphia, Pennsylvania, 1984.
- [47] F. den Hollander. *Large Deviations*. Fields Institute Monographs. American Mathematical Society, Providence, Rhode Island, 2000.
- [48] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer, Berlin, Germany, second edition, 1998.
- [49] G. Haller and T. Sapsis. Where do inertial particles go in fluid flows? *Physica D*, 237(5):573–583, 2008.
- [50] F. J. Beron-Vera, M. J. Olascoaga, G. Haller, M. Farazmand, J. Triñanes, and Y. Wang. Dissipative inertial transport patterns near coherent lagrangian eddies in the ocean. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087412, 2015.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: