4-30-2021

Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis

Mario Miguel Valero
San Jose State University, mm.valero@sjsu.edu

Lluís Jofre
Stanford University

Ricardo Torres
Technical University of Catalonia - BarcelonaTech

Recommended Citation
Mario Miguel Valero, Lluís Jofre, and Ricardo Torres. "Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis" *Environmental Modelling & Software* (2021). https://doi.org/10.1016/j.envsoft.2021.105050

Environmental Modelling and Software 141 (2021) 105050

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis

Mario Miguel Valero [a,*], Lluís Jofre [b,c], Ricardo Torres [c]

[a] Department of Meteorology and Climate Science, San José State University, San José, CA, 95192, USA
[b] Center for Turbulence Research, Stanford University, Stanford, CA, 94305, USA
[c] Department of Fluid Mechanics, Technical University of Catalonia - BarcelonaTech, Barcelona, 08019, Spain

ARTICLE INFO

Keywords:
Forest fire
Multifidelity Monte Carlo
Predictive science & engineering
Sensitivity analysis
Uncertainty quantification
FDS

ABSTRACT

Wildfire behavior predictions typically suffer from significant uncertainty. However, wildfire modeling uncertainties remain largely unquantified in the literature, mainly due to computing constraints. New multifidelity techniques provide a promising opportunity to overcome these limitations. Therefore, this paper explores the applicability of multifidelity approaches to wildland fire spread prediction problems. Using a canonical simulation scenario, we assessed the performance of control variates Monte-Carlo (MC) and multilevel MC strategies, achieving speedups of up to 100x in comparison to a standard MC method. This improvement was leveraged to quantify aleatoric uncertainties and analyze the sensitivity of the fire rate of spread (RoS) to weather and fuel parameters using a full-physics fire model, namely the Wildland-Urban Interface Fire Dynamics Simulator (WFDS), at an affordable computation cost. The proposed methodology may also be used to analyze uncertainty in other relevant fire behavior metrics such as heat transfer, fuel consumption and smoke production indicators.

1. Introduction

The number of uncertainties involved in the study of wildfire spread is typically large due to (i) the modeling assumptions required to mathematically describe the different physics and their couplings, i.e., epistemic uncertainty, and (ii) the aleatoric incertitude resulting, for instance, from the lack of detailed evidence regarding initial and boundary conditions. Therefore, numerical analyses based on a single deterministic realization for a particular set of input parameters are typically not conclusive and neither truly predictive (Arca et al., 2007; Alexander and Cruz, 2013; Cruz and Alexander, 2013; Filippi et al., 2014; Sá et al., 2017; Cruz et al., 2018). A solution to this problem is to consider the system under study stochastic and analyze the relationship between input and output probability distributions by means of efficient statistical methods. In this regard, the fields of uncertainty quantification (UQ) and sensitivity analysis (SA) have remarkably grown over the last decades within the computational fluid dynamics (CFD) community (Najm, 2009; Masquelet et al., 2017; Jofre et al., 2018, 2019), and it is now extensively accepted that the potential of estimating and minimizing uncertainties, in combination with numerical verification and physics validation (V&V), is crucial for augmenting the confidence in the numerical predictions.

In the specific case of wildfire spread modeling and prediction, uncertainty quantification is essential to prioritize the needs in data collection, inform decisions about future research investments and improve communication between modelers and managers, leading to more informed fire management decisions (Riley and Thompson, 2016). The benefits of quantifying uncertainty in wildfire spread studies have already been demonstrated — see, for instance, Cruz (2010)—, and multiple authors have emphasized the need to improve the statistical characterization of fire model outputs (Benali et al., 2016; Pinto et al., 2016; Ramirez et al., 2019). However, previous attempts to quantify uncertainty in forest fire modeling suffer from significant limitations, primarily caused by unattainable computing requirements. Full-physics models, which carefully resolve and/or model the physical and chemical phenomena involved in fire spread, are too computationally expensive to be used within Monte Carlo (MC) analyses. Consequently, most of the previous studies on fire spread uncertainty have been restricted to Rothermel's semi-empirical model (Rothermel, 1972) and a few other completely empirical formulations (e.g. Cheney and Gould (1997); Cruz et al. (2008); Liu et al. (2014)).

One of the first attempts to apply Global Sensitivity Analysis (GSA) techniques to Rothermel's model was carried out by Salvador et al. (2001). Since then, several authors have extended their work. Anderson et al. (2007) quantified aleatoric uncertainties by applying random

* Corresponding author.
E-mail addresses: mm.valero@sjsu.edu (M.M. Valero), lluis.jofre@upc.edu (L. Jofre), ricardo.torres@upc.edu (R. Torres).

https://doi.org/10.1016/j.envsoft.2021.105050
Accepted 30 March 2021
Available online 13 April 2021

Glossary of acronyms

CFD	Computational Fluid Dynamics
CoV	Coefficient of Variation
CV	Control Variates
CSIRO	Commonwealth Scientific and Industrial Research Organisation
FDS	Fire Dynamics Simulator
GSA	Global Sensitivity Analysis
HF	High Fidelity
LF	Low Fidelity
MC	Monte Carlo
MF	Multifidelity
ML	Multilevel
MSE	Mean Squared Error
PCE	Polynomial Chaos Expansion
PDE	Partial Differential Equation
QoI	Quantity of Interest
RoS	Rate of Spread
SA	Sensitivity Analysis
SAV	Surface Area to Volume ratio
UQ	Uncertainty Quantification
V&V	Verification and Validation
WFDS	Wildland-Urban Interface Fire Dynamics Simulator

perturbations to weather variables (namely, temperature, relative humidity, wind speed and wind direction) in a Monte Carlo fashion. A similar approach was followed by Finney et al. (2011), who proposed a probabilistic framework to account for uncertainties in weather variables based on ensemble simulations of the fire spread model described by Finney (2002). Cruz (2010) studied the response of another fully empirical model to 10-m open wind speed and dead fuel moisture content. Later, Cruz and Alexander (2017) extended the work of Cruz (2010) to account for crown fire ignition and spread. They continued to use fuel moisture content and wind speed as uncertain inputs, and they added uncertainty ranges to fuel structure parameters such as fuel load, surface fuel depth, canopy base height, stand height, and canopy bulk density.

Alternatively to the MC approach, which may become computationally prohibitive for large sample sizes, Bachmann and Allöwer (2002) assessed uncertainty propagation in Rothermel's model using a first-order approximation based on a Taylor series expansion. In a similar attempt to reduce the computational cost of the original MC method, Jimenez et al. (2008) proposed a sampling algorithm to accelerate convergence using sensitivity derivative information. Such algorithms prioritize the input variables that have a larger impact on the model outputs to reduce the dimensionality of the problem. However, that approach can only be applied to differentiable formulations that meet certain smoothness conditions. Liu et al. (2015) extended the work of Jimenez et al. (2008) by using GSA techniques to simplify the fuel representation required in Rothermel's fire spread model. Afterwards, they combined a Sensitivity Derivative Enhanced Sampling (SDES) algorithm with a control variates scheme to accelerate the convergence of a randomized quasi-MC method to quantify aleatoric uncertainties. Recently, Yuan et al. (2020) presented a first attempt to characterize epistemic uncertainty in a semi-physical fire spread model developed by Liu et al. (2014). Specifically, they studied the effect on fire rate of spread of model parameters that describe energy transfer, including flame temperature and emissivity, flame length and tilt angle, fuel temperature of ignition, fuel consumption efficiency and the heat convection coefficient.

Despite the recent improvements in uncertainty quantification studies focused on wildfire spread, no study so far has included full-physics models in the analysis. While empirical and semi-empirical fire spread models are of most value for rapid operational purposes, they do not resolve physical and chemical phenomena essential to accurately describe fire behavior. This fact has two important implications. First, not resolving fundamental physics mechanisms prevents the study of epistemic uncertainties related to such phenomena. Second, because empirical models are built upon statistical relationships between macroscopic variables rather than physical principles, their application cannot be easily extrapolated beyond the conditions used to develop the models; acquiring experimental data in a range of conditions wide enough to generalize empirical models is practically unfeasible, which also limits the study of aleatoric uncertainties.

Conversely, physics-based CFD simulators, which are intended to resolve and/or model all relevant physical and chemical mechanisms that characterize wildfire spread, are exceedingly expensive to be used within uncertainty quantification studies based on standard MC methodologies. The resulting high-dimensional parameter space, in conjunction with the large computational demands of the simulation runs required, necessitates cost-efficient, non-intrusive — i.e., sampling-based — UQ methods that accurately estimate the statistics of the quantities of interest (QoI). Many widely-used non-intrusive methods, such as stochastic collocation (Mathelin and Hussaini, 2003; Xiu and Hesthaven, 2005) and polynomial chaos expansions (PCE) (Ghanem and Spanos, 2003; Xiu and Karniadakis, 2002; Doostan and Owhadi, 2011), suffer from a rapid (up to exponential) growth of computational cost as a function of the number of input variables characterizing the uncertainty. On the other hand, MC methods are popular and powerful approaches for the estimation of statistical parameters due to their robust convergence behavior independent of the number of uncertainties. However, it is well-known that the mean square error (MSE) of the MC estimator converges slowly as a function of the sample size N. This slow convergence may thus become a critical issue, or even prohibitive, especially if sampling involves computationally expensive operations, such as solving a (discretized) partial differential equation (PDE). Recent research has targeted the development of cost reduction techniques to improve MC sampling methods, which are usually based on multifidelity (MF) methodologies. In MF frameworks, high-fidelity (HF) models are exploited to provide the required level of accuracy and insight into detailed physical phenomena, whereas low-fidelity (LF) simulations are leveraged to economically improve the statistical characterization of modeled QoIs. At present, the performance of these strategies has been mainly assessed on canonical PDEs and rather simplified flow problems. However, to the best of the authors' knowledge, very few UQ studies of complex, large-scale, high-dimensional systems have been conducted and published in the literature. Selected recent examples include shock/turbulent-boundary-layer interaction in scramjet engines (Bermejo-Moreno et al., 2012), analysis and optimization of high-performance aircraft nozzles (Alonso et al., 2017), multiphase flow simulations of cloud cavitation collapse (Sukys et al., 2017), and the study of radiative heat transfer in particle-laden turbulence (Jofre et al., 2020).

Therefore, the objective of this work is to demonstrate the potential of MF techniques to reduce the computing requirements of performing uncertainty quantification and sensitivity analysis in physics-based wildfire spread computational predictions, allowing in this manner a more complete characterization of aleatoric and epistemic uncertainties. The paper is organized as follows. First, a description of the considered MF strategies is provided in Section 2. Next, Section 3 introduces the wildfire models utilized. Section 4 presents the canonical simulation problem analyzed in this paper, and Section 5 describes and discusses the results obtained. Finally, Section 6 summarizes the work, provides conclusions, and proposes future research.

2. Multifidelity sampling strategies

In computational science and engineering, multiple physical/mathematical/numerical models with different features can be constructed to characterize a system of interest. Typically, computationally expensive HF models are designed to describe the system with the degree of accuracy required by the problem under study, while LF models are formulated as cheaper representations, usually at the expense of lower accuracy, robustness or generality. Outer-loop problems, such as inference, UQ and optimization, require large numbers of model evaluations for different input values, resulting in unaffordable computational requirements in the case of large-scale, multiphysics calculations. The objective of MF methods, therefore, is to reduce the cost of the outer-loop problem by combining the accuracy of the HF models with the speedup achieved by the LF representations. Different MF UQ strategies exist in the literature; see, for example, the reviews by Peherstorfer et al. (2018) and Fernandez-Godino et al. (2016, 2019). However, due to the high-dimensional input space and the complexity of the conservation equations involved, this study is restricted to a reduced subset of acceleration strategies appertaining to surrogate-based MC type sampling approaches.

As its name indicates, MC-type approaches are derived from the original Monte Carlo method, in which the expectation of the QoI as a function of the stochastic inputs $\boldsymbol{\xi}$, $Q = Q(\boldsymbol{\xi})$, is estimated via a sample average. Let $\mathbb{E}[Q]$ and $\mathbb{V}[Q]$ denote the mean and variance of Q. Given N independent realizations of the stochastic input, denoted $\boldsymbol{\xi}^{(i)}$, the MC estimator of $\mathbb{E}[Q]$ is defined as $\widehat{Q}_N^{\mathrm{MC}} = N^{-1}\sum_{i=1}^{N} Q^{(i)}$, where $Q^{(i)} = Q(\boldsymbol{\xi}^{(i)})$. Although unbiased, the precision of $\widehat{Q}_N^{\mathrm{MC}}$, measured by its standard deviation $\sqrt{\mathbb{V}[Q]/N}$, decays slowly as a function of N. Therefore, for a fixed computational budget ($\propto N$), a viable alternative to increase the MC precision is to possibly replace Q with other quantities with smaller variances.

2.1. Multilevel Monte Carlo

One of the most popular acceleration strategies is the multilevel (ML) method (Giles, 2008; Adcock et al., 2020). This technique, inspired by the multigrid solver idea in linear algebra, is based on evaluating realizations of Q from a hierarchy of models with different levels ℓ, $\ell = 0, \ldots L$, with L the most accurate model, in which Q is replaced by the sum of differences $Y_\ell = Q_\ell - Q_{\ell-1}$; to simplify the notation for level 0, the expression is redefined to $Y_0 = Q_0$. As a result, the QoIs of the original and new ML problems have the same mean $\mathbb{E}[Q]$. An example of a level is the grid resolution considered for solving the system of equations, so that a LF (or HF) model can be established by simulating Q on a coarse (or fine) grid. Then, $\mathbb{E}[Q]$ can be computed using the ML QoI and an independent MC estimator on each level ℓ as

$$\widehat{Q}^{\mathrm{ML}} = \sum_{\ell=0}^{L} \widehat{Y}_\ell^{\mathrm{MC}} = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} Y_\ell^{(i)}. \tag{1}$$

This approach is referred to as multilevel Monte Carlo (MLMC), or simply ML, and the resulting estimator has a variance equal to

$$\mathbb{V}\left[\widehat{Q}^{\mathrm{ML}}\right] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Y_\ell]}{N_\ell}. \tag{2}$$

Consequently, if the level definition is such that $Q_\ell \to Q$ in mean square sense, then $\mathbb{V}[Y_\ell] \to 0$ as $\ell \to \infty$, and therefore fewer samples are required on the finer level L. In particular, it is possible to show that the optimal sample allocation across levels N_ℓ is obtained in closed form given a target variance of the ML estimator equal to $\varepsilon^2/2$, and resulting in (Giles, 2008)

$$N_\ell = \frac{\sum_{k=0}^{L} \sqrt{\mathscr{C}_k \mathbb{V}[Y_k]}}{\varepsilon^2/2} \sqrt{\frac{\mathbb{V}[Y_\ell]}{\mathscr{C}_\ell}}, \tag{3}$$

where the computational cost of the individual Y_ℓ evaluations is denoted by $\mathscr{C}_\ell$, and ε^2 represents the MSE of the estimator. It is important to note that the variance decay can be proven to be satisfied only for levels based on a numerical discretization (spatial/temporal meshes), and not for general hierarchies of models, such as 2-D versus 1-D, large-eddy simulation (LES) versus Reynolds-Averaged Navier Stokes (RANS), etc.

2.2. Control variates Monte Carlo

To accommodate LF representations that are not obtained directly from coarsening the HF models, a common approach is to utilize LF realizations as a control variate (Peherstorfer et al., 2018; Pasupathy et al., 2014; Geraci et al., 2017). In statistics, the control variates approach replaces a generic quantity q by $q + \alpha(\mathbb{E}[g] - g)$, where g is a function chosen for its high correlation with q and for which the value of $\mathbb{E}[g]$ is readily available. However, in the problem of interest here the LF correlations and expected values are not available *a priori*, and consequently need to be established during the computations along with the HF calculations. As a consequence, the expected values of the LF models are generally approximated by means of MC estimators requiring a set of additional (independent) LF computations. The control variates (CV) MC estimator is defined as

$$\widehat{Q}^{\mathrm{CV}} = \widehat{Q}_{\mathrm{HF}}^{\mathrm{MC}} + \alpha\left(\mathbb{E}[Q_{\mathrm{LF}}] - \widehat{Q}_{\mathrm{LF}}^{\mathrm{MC}}\right), \tag{4}$$

where $\widehat{Q}_{\mathrm{HF}}^{\mathrm{MC}} = N_{\mathrm{HF}}^{-1}\sum_{i=1}^{N_{\mathrm{HF}}} Q_{\mathrm{HF}}^{(i)}$, $\mathbb{E}[Q_{\mathrm{LF}}] \approx (N_{\mathrm{LF}} - N_{\mathrm{HF}} + 1)^{-1}\sum_{i=N_{\mathrm{HF}}+1}^{N_{\mathrm{LF}}} Q_{\mathrm{LF}}^{(i)}$, $\widehat{Q}_{\mathrm{LF}}^{\mathrm{MC}} = N_{\mathrm{HF}}^{-1}\sum_{i=1}^{N_{\mathrm{HF}}} Q_{\mathrm{LF}}^{(i)}$, N_{HF} and N_{LF} are the number of HF and LF samples, respectively, and $\alpha = \mathbb{C}[Q_{\mathrm{HF}}, Q_{\mathrm{LF}}]/\mathbb{V}[Q_{\mathrm{LF}}]$ is the control variates coefficient chosen to minimize the variance of $\widehat{Q}^{\mathrm{CV}}$. $\mathbb{C}[Q_{\mathrm{HF}}, Q_{\mathrm{LF}}]$ denotes the covariance between Q_{HF} and Q_{LF}. The optimal α selection leads to

$$\mathbb{V}\left[\widehat{Q}^{\mathrm{CV}}\right] = \mathbb{V}[Q_{\mathrm{HF}}]\left(1 - \rho^2 \frac{r}{r+1}\right), \tag{5}$$

with $-1 \le \rho = \mathbb{C}[Q_{\mathrm{HF}}, Q_{\mathrm{LF}}]/\sqrt{\mathbb{V}[Q_{\mathrm{HF}}]\mathbb{V}[Q_{\mathrm{LF}}]} \le 1$ the Pearson correlation coefficient between the HF and LF models, and r is used to parameterize the additional rN_{HF} LF realizations with respect to HF. As described by Geraci et al. (2017), the optimal control variates is obtained for

$$r = \sqrt{\frac{\mathscr{C}_{\mathrm{HF}}}{\mathscr{C}_{\mathrm{LF}}} \frac{\rho^2}{1-\rho^2}} - 1, \tag{6}$$

where $\mathscr{C}_{\mathrm{HF}}$ and $\mathscr{C}_{\mathrm{LF}}$ are the costs of a HF and LF sample, respectively. A comprehensive description of the control variates MC estimator, like for example the derivation of optimal coefficients and number of samples per fidelity, is detailed in Peherstorfer et al. (2018).

3. Wildfire spread models

Over the past decades, numerous approaches have been proposed to model and predict wildfire behavior. These strategies range from completely empirical correlations to detailed physics-based simulation frameworks; see, for instance, Sullivan (2009a, 2009b, 2009c) for a comprehensive review. Tools designed to be used operationally for decision support need to be computationally fast and provide easy-to-interpret information regarding macroscopic variables, such as rate of spread, flame height and fire line intensity. Therefore, operational simulators are frequently built upon empirical or semi-empirical models. Conversely, the detailed study of fire dynamics requires additional insight into the physical and chemical phenomena involved in fire spread, including, among others, pyrolysis, combustion, heat transfer and turbulence. Such level of detail can only be achieved by means of CFD approaches, which typically require intense computing resources that are several orders of magnitude larger than what operational

simulators demand.

One of the most popular solutions among operational wildfire simulators is the Rothermel model (Rothermel, 1972). Based on semi-empirical relationships between the parameters that determine the heat emitted by the fire and the energy needed by the unburned fuel to ignite, the Rothermel model provides a well-balanced combination of physical insight and operational capabilities. While its application requires empirical adjustment of fuel and wind parameters, the fact that such parameters are defined following physical principles facilitates extrapolating the use of the model to new vegetation types and diverse weather conditions. Nonetheless, the Rothermel model has important limitations, and consequently it is usually coupled with additional models that account for phenomena not considered initially, such as crown ignition and spotting.

From a full-physics modeling perspective, several CFD simulation frameworks have been designed specifically for wildfire applications. Some of the most popular solutions include the Wildland-Urban Interface Fire Dynamics Simulator (WFDS) (Mell et al., 2007), FIRETEC (Linn et al., 2002), and FIRESTAR (Morvan et al., 2006, 2018). WFDS was selected in this study due to its widespread adoption by the scientific community. WFDS is currently part of the broader Fire Dynamics Simulator (FDS) (Forney et al., 2003; Floyd and McGrattan, 2008), a CFD fire and smoke simulation tool widely used for fire safety applications. FDS is an open-source solver,[1] maintained by the US National Institute of Standards and Technology (NIST) and jointly developed by several academic institutions in multiple countries.

Finally, there exist a variety of fully empirical fire models that have demonstrated high accuracy and applicability when used within specific parameter ranges. These models were not included in the present analysis due to the wide range of alternatives available, but they constitute an important pool of LF candidates to be considered in the MF strategies introduced in Section 2. The subsections below provide further details on the wildfire models used in this study.

3.1. The rothermel model

Rothermel's original formulation, initially published in 1972 (Rothermel, 1972), is based on an energy balance between the heat emitted by a flaming fire front and the energy required to ignite the fuel ahead of it. The model is given by (Andrews, 2018)

$$R = \frac{I_R \xi (1 + \varphi_w + \varphi_s)}{\rho_b \varepsilon Q_{ig}}, \tag{7}$$

where R is the fire rate of spread, I_R is the reaction intensity, ρ_b is the fuel bulk density, and Q_{ig} is the heat required to ignite the fuel. The remaining variables are empirical parameters that allow matching observed fire behavior to this simple formulation. ξ is the propagating flux ratio and represents the portion of reaction intensity that propagates toward the unburned fuel. φ_w and φ_s are named wind and slope factors, respectively, and account for the increase in heat transfer that occurs when the fire spreads in the direction of the wind and/or upslope. Finally, ε is an efficiency parameter dependent on the fuel particle size and shape. Rothermel developed empirical functions so that the value of these variables could be derived directly from fuel and environmental parameters, providing in this manner physical intuition to their calibration. Through such empirical relationships, the fire rate of spread can ultimately be expressed as a function of fuel heat content (h), fuel total mineral content (S_T), fuel effective mineral content (S_e), oven-dry fuel particle density (ρ_p), fuel particle surface-area-to-volume (SAV) ratio (σ), oven-dry fuel load (w_0), fuel bed depth (δ), dead fuel moisture of extinction (M_x), fuel moisture content (M_f), wind velocity at midflame height (U), and terrain slope (φ) in the form

$$R = F\left(h, S_T, S_e, \rho_p, \sigma, w_0, \delta, M_x, M_f, U, \varphi\right). \tag{8}$$

Following the formulation outlined in Eq. (8), the fuel-related parameters are usually grouped into standard categories that broadly represent the most common vegetation distributions. Such categories are known as fuel models and have experienced substantial development since the first version proposed by Rothermel. Multiple authors have highlighted the importance of the fuel input parameters on the predictive capability of fire spread achieved by the Rothermel model (e.g., Arca et al. (2007)). For this reason, ad-hoc fuel models are usually developed prior to applying the Rothermel model to a new geographical area and/or vegetation type.

The original Rothermel model is one-dimensional (1D), and consequently its practical application requires coupling it with a two-dimensional (2D) propagation scheme. Several propagation methodologies have been proposed following Eulerian and Lagrangian approaches (Bova et al., 2016). Furthermore, the Rothermel model has been incorporated into a significant number of 1D and 2D simulators, such as BehavePlus (Heinsch and Andrews, 2010), FARSITE (Finney, 1998) and WRF-SFIRE (Mandel et al., 2011), which are at present extensively used internationally by fire managers and researchers.

3.2. Wildland-Urban Interface Fire Dynamics Simulator

WFDS was designed as an extension of FDS to include fire spread in vegetative fuels (Mell et al., 2007). FDS is a CFD model of fire-driven fluid flow, which numerically solves a form of the Navier-Stokes equations appropriate for low-speed (Mach numbers below $Ma < 0.3$), thermally-driven flow, with an emphasis on smoke and heat transport from fires. It is widely used in fire protection engineering problems and its applicability to study fundamental fire dynamics and combustion phenomena is gaining increasing attention.

The core hydrodynamic formulation is solved by an explicit predictor-corrector scheme, which is second-order accurate in space and time, and spatially discretized on a rectilinear mesh. By default, the small-scale turbulent fluctuations are modeled by means of LES strategies, although it is possible to perform Direct Numerical Simulations (DNS) if the underlying numerical mesh is fine enough. For most applications, combustion is modeled through a single step, mixing controlled chemical reaction. Radiative heat transfer is included in the model and solved via a Discrete Ordinates Method (DOM) with a default number of 100 discretization angles. The radiation transport equation is solved for a grey gas, and in some limited cases using a wide-band model. In addition, FDS includes Langrangian particle tracking capabilities to represent elements that are not captured by the Eulerian grid (McGrattan et al., 2019a), like for example fuel and soot particles. The WFDS extension, now merged into FDS, added the functionalities needed to (i) describe vegetative fuels, (ii) resolve convective and radiative heat transfer within those fuels, and (iii) calculate their thermal degradation, pyrolysis and combustion.

4. Description of the wildfire spread scenario

The MF techniques presented in Section 2 were applied to the canonical wildfire spread simulation problem described below. The analyzed scenario belongs to a set of medium-scale field fire experiments conducted by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in the Northern Territory of Australia in 1986 (Cheney et al., 1993). These tests were monitored to measure the fire rate of spread and evaluate its correlation with fuel and weather variables such as fuel height, fuel moisture content, fuel load, fuel bulk density, and wind speed. Two of the experiments had been simulated using WFDS (Mell et al., 2007) and incorporated into the FDS test suite as validation examples (McGrattan et al., 2019b). Furthermore, the fact that the experiments were conducted on horizontal grassland fields

[1] https://github.com/firemodels/fds.

facilitates the application of Rothermel's model for surface fire spread. Particularly, this study considers the CSIRO grassland F19 experiment, which burned a square field of $200m \times 200m$ using a 175-m long ignition line. This type of experiment design is very frequent in wildfire behavior studies.

4.1. Problem setup

The setup of the problem is depicted in Fig. 1. This configuration was generated with WFDS by reproducing the validation study performed by Mell et al. (2007). For the HF samples, the computational domain of size $240m \times 240m \times 20m$ was discretized using a uniform Cartesian grid with $\Delta = 0.5m$ resolution. Turbulence was solved by means of a LES strategy based on Deardorff's turbulent viscosity model. Time integration was performed by limiting the Courant-Friedrichs-Lewy (CFL) value to CFL < 1, and radiation transport was calculated utilizing a DOM method with 100 discrete angles (McGrattan et al., 2019a).

Vegetation (grass) was represented as a homogeneous bed of Lagrangian combustible particles, whose main physical and chemical properties are summarized in Table 1. Grass particles were arranged in a rectangular prismatic space with a predefined bulk density and fuel bed depth. These two parameters were included into the list of uncertainties analyzed, together with the fuel moisture fraction and the particle SAV ratio. See Section 4.2 for a detailed discussion of their values.

4.2. Uncertainties and quantities of interest

Wildfire spread modeling requires the selection of a large number of parameters. However, most of the aleatoric uncertainty is usually concentrated in a limited subset of the input variables. Previous works identified the following variables as significant sources of aleatoric uncertainty in wildfire spread: 1-h fuel moisture content (Salvador et al., 2001; Clark et al., 2008; Finney et al., 2011; Ervilha et al., 2017), 1-h fuel load (Liu et al., 2015; Cai et al., 2019), fuel bed depth (Salvador et al., 2001; Jimenez et al., 2008; Liu et al., 2015; Ervilha et al., 2017; Cai et al., 2019), fuel particle SAV ratio (Jimenez et al., 2008; Liu et al., 2015; Ervilha et al., 2017; Cai et al., 2019), and wind speed (Salvador et al., 2001; Anderson et al., 2007; Jimenez et al., 2008; Clark et al., 2008; Finney et al., 2011; Liu et al., 2015; Benali et al., 2016; Ervilha et al., 2017). Consequently, this list of variables were considered as uncertain inputs (ξ_i) in this work, with their stochastic ranges characterized from studies and experiments available in the scientific literature. Scott and Burgan (2005) summarized a series of standardized fuel models, which are extensively used at present to model fire behavior. Their list of models contains 9 grass-type options, each of them with a characteristic value of fine fuel load, SAV ratio and packing ratio. Additionally, they suggested a range of dead fuel moisture and wind conditions to evaluate these fuel models. The data variability considered by Scott and Burgan (2005) was used in the present study to limit the sampling space as detailed in Table 2. In the case of wind speed, mid-flame estimations were converted to values at a 20-feet height using a conversion factor of 0.4 as recommended by the US National Wildfire Coordinating Group (Estimating Winds for Fire Behavior, 2019).

Table 1
Main physical and chemical properties used to represent grass particles in the simulation experiments; see (Mell et al., 2007) and (McGrattan et al., 2019b) for additional details.

Property	Units	Value
Chemical composition	—	$C_6H_{10}O_5$
Oven-dry density	kg/m^3	512
Conductivity	$W/(m \cdot K)$	0.1
Specific Heat	$kJ/(kg \cdot C)$	$= 1.11 + 0.0037T$
Heat of combustion	kJ/kg	15,600
Soot yield	—	0.02
Heat of pyrolysis	kJ/kg	416
Char yield	—	0.02
SAV ratio	—	stochastic (see Table 2)
Moisture fraction	—	stochastic (see Table 2)

Amidst the large number of output statistics that can be obtained from a wildfire spread calculation, the fire rate of spread (RoS) is one of the most critical quantities to understand fire behavior and its effects on the environment (O'Brien et al., 2016). Among other applications, RoS provides a quantitative representation of fire evolution, and it can be used to analyze the effect of varying weather conditions or possible fuel treatments in a specific area. Consequently, this study uses the rate of spread, averaged in the simulation domain, as QoI. Simulated RoS was measured by recording the longitudinal coordinate of the point with maximum gas temperature at every time step. The search for maximum temperatures was restricted to a narrow longitudinal band placed along the domain symmetry plane and close to the ground. The resulting time-distance distribution was fitted to a linear function in order to estimate the average RoS.

4.3. Multifidelity modeling strategy

The MF strategies described in Section 2 rely on the construction of levels with different ratios of fidelity and computational cost to accelerate the process of reducing the variance of the estimators. Diverse approaches can be pursued to develop such cheaper models, with the only requirement that these lower-fidelity representations need to be orders of magnitude faster to compute while maintaining some degree of interdependence, the higher the better, with the QoI calculated at the different levels. In this work, two different methods to build coarser

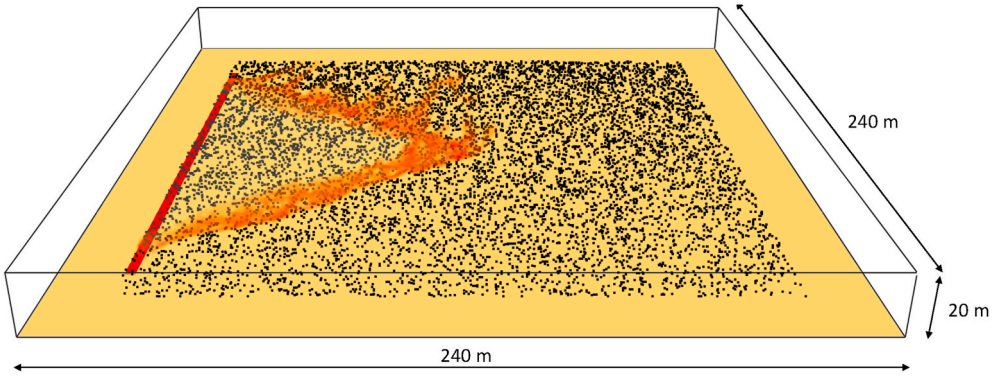


Fig. 1. Computational domain of size $240m \times 240m \times 20m$ used to reproduce the CSIRO grassland F19 experiment (Cheney et al., 1993; Mell et al., 2007; McGrattan et al., 2019b). The image is a snapshot outputted from WFDS for a sample run and displayed using Smokeview (Forney, 2016). The yellow surface represents the field, the black dots correspond to the grass particles, the red line is the ignition line, and the orange region indicates the distribution of the fire as it spreads.

representations of the HF simulation were tested.

The first strategy is based on the ML method in which the HF CFD scenario is resolved using decreasing levels of resolution in (i) flow cell size, (ii) integration time-step (Δt), and (iii) number of discrete angles utilized to resolve the radiative heat transport. Based on preliminary tests, we designed three different levels named LF-2x, LF-5x and LF-10x, which are 25 $\times$, 250$\times$ and 2500$\times$ cheaper to compute than the HF level, respectively. The different resolution levels for (i) and (iii) are specified in Table 3. The integration time-step was automatically adjusted at every iteration based on the CFL constraint; approximately, a cell-size increase of 2$\times$ corresponds to a two-fold increase of Δt.

The second approach explores the benefits of introducing the output of the Rothermel fire spread model as a control variate. This was accomplished through the execution of FARSITE simulations for a similar scenario and with the same input parameters. This fidelity is referred to as LF-CV, and is 2500$\times$ faster to compute than the HF level. A visualization of a sample output obtained from FARSITE is depicted in Fig. 2. The map of fire time of arrival was used to compute an average value for the RoS.

5. Results and discussion

This section presents and analyzes the data and results for the problem described in Section 4, which were obtained by running HF and LF simulations based on the modeling approaches described in Section 3, and provides a discussion of the statistics calculated by utilizing the methodologies introduced in Section 2. First, Section 5.1 presents a concise statistical characterization of the RoS — the QoI considered in this work — for the different fidelities. Next, in Section 5.2, the performance of various MF estimators is studied to select the optimal strategies in terms of balancing variance reduction/correlation and computational costs. The list of MF designs considered is visually summarized in Table 4. In Section 5.3, the uncertainty of the problem is propagated through the model and quantified by making use of the acceleration techniques chosen in the previous subsection. Finally, Section 5.4 discusses a variance-based sensitivity analysis performed to rank the effect of the uncertainties on QoI variability.

5.1. Pilot sampling of the quantity of interest

As a first step to construct efficient MF estimators, exploratory data was collected by running the same 32 pilot samples for the 5 fidelities designed. This set of samples were generated following a design of experiment (DoE) based on the KDOE approach (Roy et al., 2020). KDOE is an iterative method that introduces stochasticity in the sampling process by means of a variable kernel density estimation to optimize the uniformity of the DoE. This approach provides a more homogeneous exploration of the input parameter space, especially when the number of samples is relatively small. For example, instantaneous snapshots of the vertical velocity field for the first sample generated by the HF, LF-2x, LF-5x and LF-10x fidelities are depicted in Fig. 3. It can be inferred from these snapshots that (i) the lowest-resolution fidelities (e.g., LF-10x) tend to underresolve eddies located at the fire front, which could affect the rate of convective heat transfer, and consequently the predicted fire rate of spread; (ii) nonetheless, larger-scale turbulent motions are sufficiently resolved, with buoyancy pulses present in all fidelity levels; (iii) velocity values maintain similar absolute ranges across all fidelity levels, although (iv) the velocity field becomes increasingly diffused in space as spatial resolution decreases. Conclusive arguments, however, cannot be obtained by solely analyzing Fig. 3 as (i) the QoI targeted in this work is the average RoS, and (ii) the performance of the MF estimators is not directly related (to a first-order approximation) to the accuracy of the LF results with respect to the HF data.

The data generated from the pilot samples for the different fidelities were collected and visually summarized in Fig. 4 by means of boxplots. The fidelities are sorted (left to right) in decreasing computational cost starting from HF, ending with LF-10x, and spanning a total of 4 orders of magnitude. As stated in the paragraph above, the accuracy of the LF data with respect to the HF results is not the principal component for the performance of the ML and CV estimators as they are, respectively, formulated in terms of variance reduction and correlation. Nevertheless, it is instructive to analyze the relations between the HF and LF data from a statistical perspective to gain insight and facilitate the effective construction of the MF estimators. The distributions in Fig. 4 show that the data are organized in 2 main blocks across fidelities: (i) HF, LF-2x, LF-5x, and LF-10x generate data distributed around $Q \approx 2.5$ and displaying

Table 2
Sources of aleatoric uncertainty (stochastic variables, ξ_i) considered in this study. Ranges were characterized based on Scott and Burgan (2005).

No.	Parameter	Units	Uncertainty range
1	Fuel moisture fraction	–	[0.03 : 0.12]
2	Fuel bed depth	m	[0.122 : 1.5]
3	Fine fuel load	kg/m^2	[0.09 : 2.25]
4	Fuel particle SAV ratio	$1/m$	[4265 : 7218]
5	Wind speed at 20-feet height	m/s	[0 : 25]

Table 3
ML levels designed for the computation of the canonical fire spread scenario.

Level	Cell size (m)	# Discrete angles	HF-equivalent sample cost
HF	0.50	100	1
LF-2x	1.00	50	1/25
LF-5x	1.25	20	1/250
LF-10x	2.00	10	1/2500

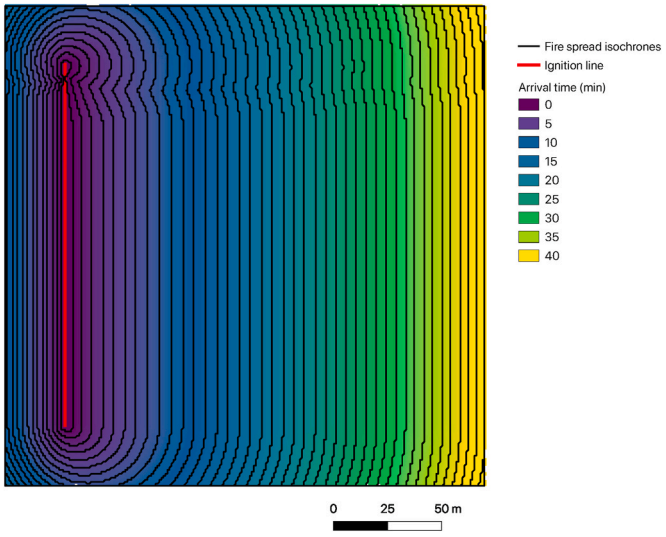


Fig. 2. Example of FARSITE's output corresponding to a snapshot of a sample for the canonical fire spread scenario.

Table 4
MF designs studied based on the CV and ML strategies and fidelities LF-CV, LF-2x, LF-5x and LF-10x.

Estimator type \ Fidelity levels	HF	LF-CV	LF-2x	LF-5x	LF-10x
Monte-Carlo (MC)	×				
Control Variates MC (CV)	×	×			
Multilevel MC (ML)	×		×		
Multilevel MC (ML)	×		×	×	
Multilevel MC (ML)	×		×	×	×

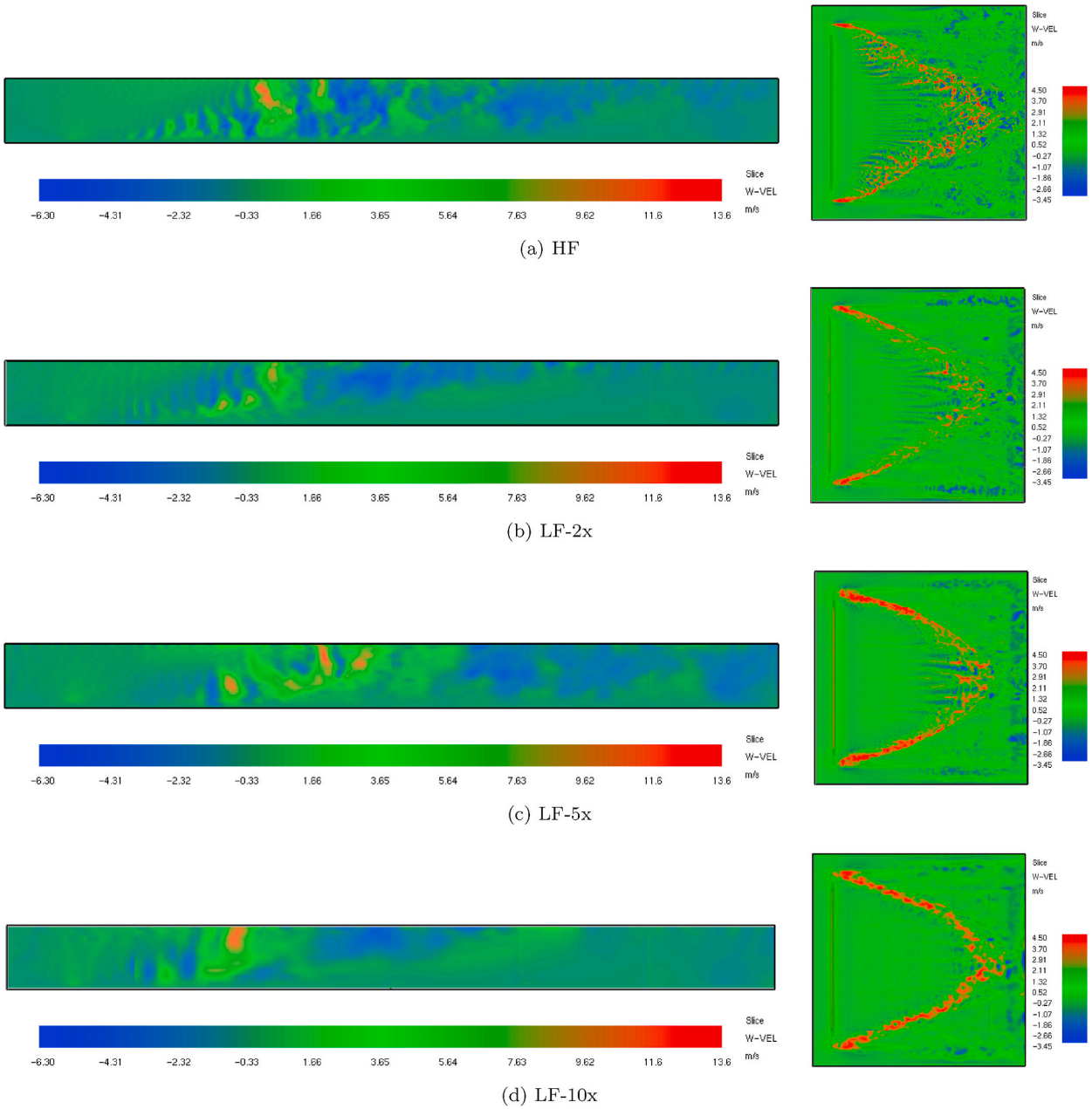


Fig. 3. Snapshots of vertical velocity for the first sample from the different fidelities: lateral view of the longitudinal symmetry plane (left), and top view of the horizontal plane at a 2-m height (right).

coefficients of variation with values CoV $\equiv \sqrt{\mathbb{V}[Q]}/\mathbb{E}[Q] \approx 0.4$, whereas (ii) LF-CV tends to underestimate Q by approximately $1.6\times$ and presents a significantly larger CoV ≈ 0.7.

5.2. Performance of the multifidelity estimators

The performance of various candidate MF estimators constructed by means of CV and ML strategies was analyzed by utilizing the pilot data described in the previous subsection. As discussed in Section 2, the speedup obtained by the ML and CV approaches is function, respectively, of the variance of Y_ℓ ($\mathbb{V}[Y_\ell]$) in Eq. (2) and the Pearson correlation coefficient (ρ) between fidelities in Eq. (5). Consequently, $\mathbb{V}[Y_\ell]$ and ρ for all potential combinations are listed in Table 5. In the case of CV-based MF estimators, which are constructed utilizing HF information and LF samples as a control variate, the best LF candidates are LF-2x and LF-5x as they present correlations of $\rho = 0.995$ and $\rho = 0.990$ with speedups of approximately $25\times$ and $250\times$, respectively. Instead, LF-CV and LF-10x are slightly less correlated with HF as their values are $\rho = 0.884$ and $\rho = 0.869$. If considering ML strategies, in which HF and different LF are combined forming a telescopic sum, a good hierarchical structure is composed by the lower fidelities LF-2x, LF-5x and LF-10x as their $\mathbb{V}[Y_\ell]$ values decay orders of magnitude, 0.002, 0.017 and 0.224 specifically, while becoming $25\times$, $250\times$ and $2500\times$ faster to compute than HF.

The extrapolated performances of a straightforward MC approach and the CV and ML estimators proposed above are reported in Fig. 5. The horizontal logarithmic-scale axis corresponds to the total cost of each estimator normalized by the cost of a HF sample. The total costs are evaluated as $\mathscr{C}^{\mathrm{MC}} = N_{\mathrm{HF}}\mathscr{C}_{\mathrm{HF}}$, $\mathscr{C}^{\mathrm{CV}} = N_{\mathrm{HF}}(\mathscr{C}_{\mathrm{HF}} + \boldsymbol{r}\mathscr{C}_{\mathrm{LF}})$ and $\mathscr{C}^{\mathrm{ML}} = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell$ for the MC, CV and ML, respectively. On the vertical logarithmic-scale axis, target estimators' MSE, $\varepsilon^2 \equiv \mathbb{V}[\widehat{Q}]$, normalized by a reference MC value $\varepsilon_0^{\mathrm{MC2}}$ obtained from the 32 pilot samples, are shown

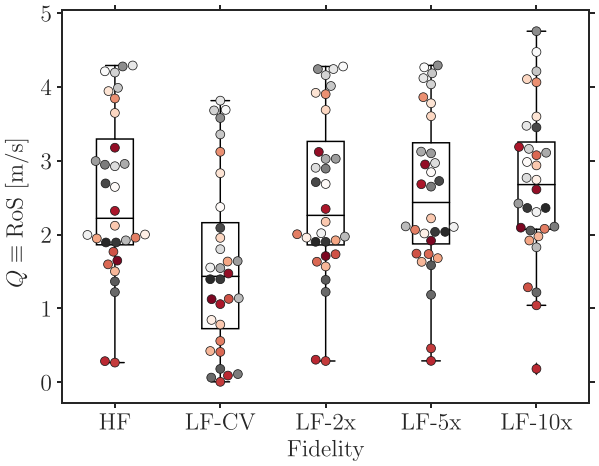


Fig. 4. RoS data distribution of the pilot samples for the different fidelities. Boxplots display the minimum (small horizontal lines at Q1-1.5 × IQR), maximum (small horizontal lines at Q3+1.5 × IQR), whiskers (vertical lines), interquartile range (boxes spanning IQR = Q3-Q1), median (large horizontal lines), outliers (diamonds), and data points (colored circles, with one color per sample across fidelities) of the distributions.

Table 5
Pearson correlation coefficient ρ (elements below diagonal) and variance of levels $\mathbb{V}[Y_\ell]$ (elements above diagonal) for all the potential combinations of fidelities.

$\rho \setminus \mathbb{V}[Y_\ell]$	HF	LF-CV	LF-2x	LF-5x	LF-10x
HF	1.0 \ 0.0	0.303	0.002	0.024	0.312
LF-CV	0.884	1.0 \ 0.0	0.300	0.284	0.539
LF-2x	0.995	0.885	1.0 \ 0.0	0.017	0.291
LF-5x	0.990	0.889	0.993	1.0 \ 0.0	0.224
LF-10x	0.869	0.784	0.877	0.901	1.0 \ 0.0

for MC and ML estimators through evaluating Eq. 3, and for the CV estimators utilizing the expression (Peherstorfer et al., 2018)

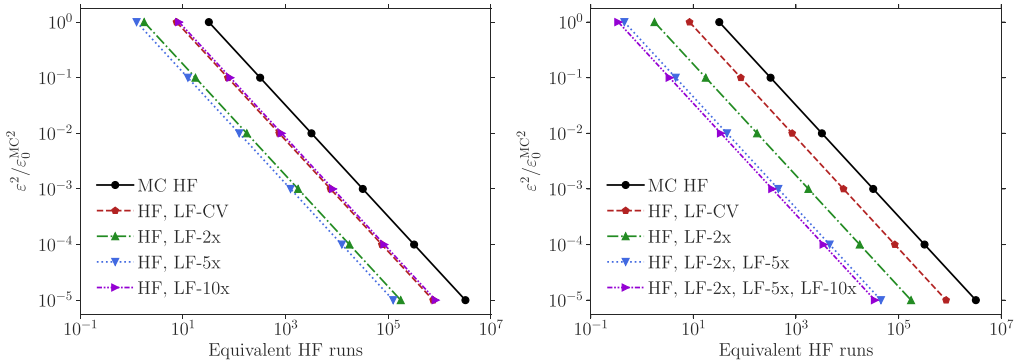
$$\frac{\varepsilon^2}{\varepsilon_0^{\mathrm{MC2}}} = \left[\sqrt{1-\rho^2} + \sqrt{\frac{\mathscr{C}_{\mathrm{LF}}}{\mathscr{C}_{\mathrm{HF}}}\rho^2}\right]^2. \tag{9}$$

The results depicted in Fig. 5 show that the speedups of the MF estimators with respect to MC are in the order of 100× to 1000× for the best-performant CV (constructed with HF and LF-5x) and ML (generated by levels Y_0 : LF-10x, Y_1 : LF-5x − LF-10x, Y_2 : LF-2x − LF-5x and Y_3 : HF − LF-2x) strategies, respectively. For example, to reduce ε^2 by an order of magnitude with respect to $\varepsilon_0^{\mathrm{MC2}}$, the MC approach requires computing 320 HF runs, while the CV (composed by 1 HF and 493 LF-5x samples) and ML (composed by 1 HF, 19 LF-2x, 222 LF-5x and 1556 LF-10x samples) demand only 3.0 and 2.8 equivalent HF runs.

Studying in detail the differences in performance between the candidate CV estimators considered at the beginning of this subsection, it is clear from Fig. 5 that ρ is a very important parameter for maximizing the speedup. In particular, the lower fidelities LF-CV and LF-10x, which present correlation coefficients slightly smaller than LF-2x and LF-5x, performed only 10× faster than MC (instead of 100 ×) even in the case of using LF-10x that is 10× and 100× faster than LF-2x and LF-5x, respectively. This observation is corroborated by the mathematical expression of the CV estimator's variance given in Eq. (5), in which the impact of ρ on the reduction of $\mathbb{V}[\widehat{Q}^{\mathrm{CV}}]$ is quadratic, while sublinearly proportional to the ratio $\mathscr{C}_{\mathrm{HF}}/\mathscr{C}_{\mathrm{LF}}$ through the parameter r. However, if two or more fidelities have similar correlation with HF, the computational cost becomes the dominant parameter, and the cheapest LF generates the most efficient CV estimator as in the case of LF-2x and LF-5x, in which the latter is the best-performant option.

Carefully analyzing the speedups of the ML estimators considered in Fig. 5, two observations can be made. The first observation is that ML strategies based on utilizing LF-CV realizations, which are not generated from coarsening the HF model, notably underperformed with respect to standard ML estimators created by coarsening the HF resolution as in the case of LF-2x, LF-5x and LF-10x. For instance, the 2-level ML estimator generated by HF and LF-2x outperformed by 10× the same-cost ML estimator constructed with HF and LF-CV. This observation agrees with the statement made at the end of Section 2.1 that the variance decay can be proven to be satisfied only for levels based on numerical discretizations (spatial/temporal meshes), and not for general hierarchies of models. The second observation is that the combination of variance decay and cheaper calculations across levels provides the optimal recipe for constructing efficient ML estimators as the 4-level one generated in this work with fidelities HF, LF-2x, LF-5x and LF-10x (levels Y_0 : LF-10x, Y_1 : LF-5x − LF-10x, Y_2 : LF-2x − LF-5x and Y_3 : HF − LF-2x). The mathematical explanation of this behavior can be inferred from the ML estimator's variance provided in Eq. (2) as it is composed by a sum of $\mathbb{V}[Y_\ell]/N_\ell$, and therefore reducing each of the addends results in an decrease of the total sum.

The results shown in Fig. 5 highlight the better performance of the ML with respect to the CV estimators in the case of LF models/levels that present small bias and moderate CoV as revealed in Fig. 4. However, in a more general problem involving very complex, non-linear fire spreads, in which such "good" LF models are more challenging to design and/or discover, the CV approach may be a more robust option. Thus, hybridization strategies, like for example the bi-fidelity (BF) approximation (Fairbanks et al., 2017, 2020; Jofre et al., 2017) and the multilevel



(a) CV estimators with different LF models (b) ML estimators considering 2, 3 and 4 levels

Fig. 5. Extrapolated MSE (normalized by the pilot $\varepsilon_0^{\mathrm{MC2}}$ value) of the MC and potential combinations of MF estimators as function of the overall computational cost in terms of equivalent number of HF runs. Solid black lines correspond to plain MC with HF samples.

multifidelity (MLMF) approach (Geraci et al., 2017; Gorodetsky et al., 2019), are promising extensions of the standard ML and CV methods for accelerating the convergence rate of statistical estimators in challenging wildfire spread scenarios.

5.3. Uncertainty propagation and quantification

As discussed in the previous subsection, the best-performant MF estimators correspond to the CV constructed with HF and LF-5x, and the ML generated by levels Y_0 : LF-10x, Y_1 : LF-5x − LF-10x, Y_2 : LF-2x − LF-5x and Y_3 : HF − LF-2x. The next step, therefore, is to utilize these MF estimators to propagate the uncertainty and quantify its impact on the QoI for the problem described in Section 4.

The precision in the prediction of the QoI by means of the MF estimators is given in terms of a coefficient of variation defined as $\mathrm{CoV}^{\mathrm{MF}} \equiv \sqrt{\mathbb{V}[\widehat{Q}^{\mathrm{MF}}]}/\widehat{Q}^{\mathrm{MF}}$. In the case of targeting $\mathrm{CoV}^{\mathrm{MF}} \lesssim 2\%$, the best-performant CV estimator requires (approximately) 4 HF and 1540 LF-5x realizations with an equivalent cost of 10 HF runs, and the optimal ML estimator demands (approximately) 4 HF, 61 LF-2x, 695 LF-5x and 4862 LF-10x samples with an equivalent cost of 9 HF runs. In contrast, the same CoV for a MC strategy based on HF, i.e., $\mathrm{CoV}^{\mathrm{MC}} \lesssim 2\%$, would require 1000 HF runs, which is approximately 100× more expensive than using the MF estimators selected.

The estimation of RoS under uncertainty using the MF and MC estimators for different HF-equivalent costs is shown in Fig. 6. For all three estimators, the predictions were not accurate when utilizing 2 HF-equivalent runs. However, starting from 4 HF-equivalent runs, the CV and ML predictions rapidly improved by approaching the "expected" estimation (approximated from 128 HF samples) and reducing their variance, whereas the MC estimator converged very slowly as revealed by the wide boxplots. In fact, even when utilizing the total 128 HF samples available, the coefficient of variation of the MC prediction was $\mathrm{CoV}^{\mathrm{MC}} \approx 9\%$, while it was $\mathrm{CoV}^{\mathrm{MF}} \lesssim 2\%$ (approximately 5× smaller) for the MF strategies with less than 10 HF-equivalent runs. Finally, as indicated by the MC and MF approaches, the prediction of RoS tends to $Q \simeq 3.1$ m/s when increasing the precision of the three estimators.

This estimation of RoS convergence can be understood as a conservative approximation due to the wide uncertainty ranges defined for the input parameters (Table 2). Such a broad sampling space was used in this study to demonstrate the potential of MF approaches even when knowledge about critical fuel and weather variables is extremely scarce. However, at least a rough characterization of the wind field and easy to measure fuel properties such as fuel load and fuel bed depth will frequently be available in field experiments. This knowledge allows narrowing the uncertainty input distribution, thus reducing spread in the simulated QoIs and accelerating convergence. In these cases, the convergence trends shown in Fig. 6 are expected to improve for all MF estimators. The relative efficiency of MF estimators, however, is not likely to change.

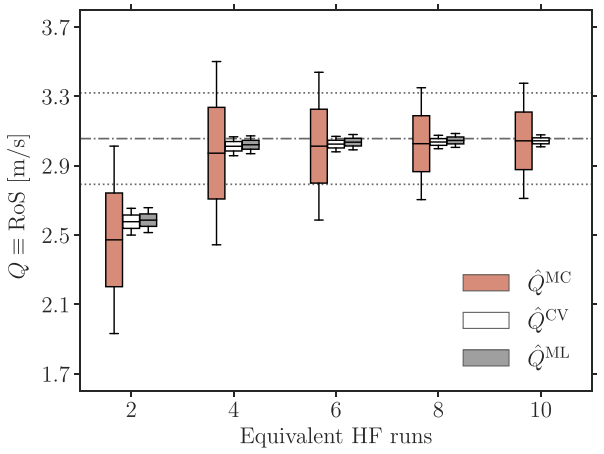


Fig. 6. Estimation of RoS using the MC (red), CV (white), and ML (grey) strategies as a function of equivalent number of HF runs. Boxplots display the minimum (small horizontal lines at Q1-1.5 × IQR), maximum (small horizontal lines at Q3+1.5 × IQR), whiskers (vertical lines), interquartile range (boxes spanning IQR = Q3-Q1), and median (large horizontal lines). The dashed grey lines correspond to the MC estimation (minimum/median/maximum) using 128 HF samples.

5.4. Variance-based sensitivity analysis

The final step is to characterize how the variability of the QoI is divided and allocated to the different sources of uncertainty present in the inputs of the problem by means of SA. The classification of uncertainties is of main importance as it clearly indicates the modeler where to concentrate resources to decrease the input incertitude by means of, for example, improved physics modeling or by acquiring additional experimental data with the objective to maximize the reduction of output's variability. There are many methods to perform SA (Ghanem et al., 2017), for example regression analysis, derivative-based local methods, screening, variogram analysis of response surfaces, variance-based methods, and scatter plots. In this work, the variance-based approach based on Sobol' indices (Sobol, 1993) was selected as it allows full exploration of the stochastic input space, accounting for interactions and nonlinear responses. The underlying principle of the methodology is to quantify the input and output uncertainties as probability distributions, and decompose the output variance into parts attributable to input variables and their combinations. The sensitivity of the output to an input variable is therefore measured by the amount of variance in the output caused by that input.

For the SA discussion, focus was placed on the ML estimator as it performed moderately better than the MF CV strategy; see Sections 5.2 and 5.3 for details. Originally designed for the estimation of expectations, as presented in Section 2.1, ML has been recently extended to the estimation of higher-order statistical moments, such as variances (Bierig and Chernov, 2015) and covariances (Mycek and De Lozzo, 2020). The combination of these extensions allows to formulate Sobol' indices within a ML framework. Following the work by Mycek & De Lozzo (2020), the first-order Sobol' index S_i associated to the i-th random input ξ_i, and corresponding to the share of output variance attributable to ξ_i individually, can be written in "pick-and-freeze" formulation (Janon et al., 2014) as

$$S_i = \frac{\mathbb{V}_{\xi_i}[\mathbb{E}_{\boldsymbol{\xi}_{-i}}[Q|\xi_i]]}{\mathbb{V}[Q]} \simeq \frac{\mathbb{V}\left[\widehat{Q}^{\mathrm{ML}}_{\xi_i}\right]}{\mathbb{V}\left[\widehat{Q}^{\mathrm{ML}}\right]}, \tag{10}$$

where subindex ξ_i indicates the stochastic input variable picked and held frozen.

Based on the best-performant ML estimator (levels Y_0 : LF-10x, Y_1 : LF-5x − LF-10x, Y_2 : LF-2x − LF-5x and Y_3 : HF − LF-2x) characterized in Sections 5.2 and 5.3 for a $\mathrm{CoV}^{\mathrm{MF}} \lesssim 2\%$, a SA study for $Q \equiv$ RoS is depicted in Fig. 7 utilizing Sobol' indices accelerated by means of the MF strategy described in the paragraph above. Computationally, the savings with respect to a straightforward MC-based SA are proportional to the speedup achieved by the MF estimator, and it scales linearly with the number of stochastic input parameters. In terms of impact on the QoI, the SA results indicate that the fuel moisture fraction (ξ_1), the fine fuel load (ξ_3) and the speed of wind (ξ_5) are the uncertainties responsible for most of the variation (approximately 20% − 30% each), followed in decreasing S_i^{ML} value by the depth of the fuel bed (ξ_2) and the fuel particle SAV ratio (ξ_4), each of which account for a slightly smaller share of the total variance (approximately 10% − 15%).

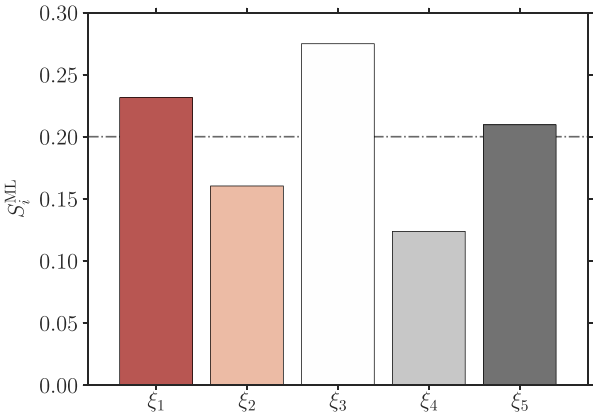


Fig. 7. Sensitivity analysis for $Q \equiv$ RoS based on Sobol' indices accelerated by means of the ML strategy. The stochastic input parameters, i.e., uncertainties of the problem, correspond to fuel moisture (ξ_1), depth of the fuel bed (ξ_2), fuel load (ξ_3), SAV (ξ_4), and wind speed (ξ_5). The horizontal dashed line indicates a level where all the uncertainties would have the same share of variability on the QoI.

These results are in line with previous findings obtained utilizing empirical and semi-empirical fire spread models. For example, Salvador et al. (2001) identified fuel depth and dead fuel moisture as the most influential parameters on fire rate of spread under moderate winds, with the importance of wind speed increasing in high-wind scenarios. Similarly, Anderson et al. (2007) found that variations in wind speed and ambient humidity have the largest impact on the variability of the extensions burned. Moreover, Clark et al. (2008) reported that wind speed and dead fuel moisture accounted for the majority of the variation on the rate of spread, followed by fuel model parameters. More recently, the analysis performed by Ervilha et al. (2017) using PCE identified the dead fuel moisture content as the most influencing parameter in the variability of Rothermel's rate of spread, followed in decreasing order by wind speed, fuel bed depth and fuel particle SAV ratio. Finally, Liu et al. (2015) reported a slightly different order in parameter importance, presenting highest Sobol' indices for wind speed, followed by fuel bed depth and fuel particle SAV ratio, and lower contributions to RoS variance by fuel load and fuel moisture.

6. Summary, conclusions and future work

Performing uncertainty quantification in wildfire modeling studies is usually challenging due to the expensive high-fidelity calculations required and the large number of uncertainties typically encountered. Consequently, this work explored the applicability of state-of-the-art multifidelity techniques to quantify uncertainty and conduct sensitivity analyses in wildland fire simulations. By constructing and using multifidelity estimators, this study was able to notably accelerate the propagation of aleatoric uncertainty through a CFD framework and quantify the sensitivity of the fire rate of spread to different weather and fuel variables.

On the basis of the problem of interest and methods considered, the multifidelity estimators achieved speedups larger than 100× with respect to straightforward Monte Carlo methods. Particularly, the multilevel method performed slightly better than the control variates due to the small bias and moderate variability of the low-fidelity data generated. However, in a more general problem involving very complex fire behavior, in which "good" low-fidelity models are challenging to design and/or discover, the control variates approach may be a more robust option. The acceleration of propagating uncertainty through the problem was leveraged to perform a sensitivity analysis. The results indicate that the fuel moisture fraction, the fine fuel load and the wind speed are the uncertainties responsible for most of the variation in fire rate of spread, followed in decreasing order by the depth of the fuel bed and the fuel particle SAV ratio.

To the best of the authors' knowledge, this work is the first study in which wildfire spread uncertainty is propagated and quantified using full-physics CFD models. The notable reduction in computing requirements achieved opens an avenue of further research. In addition to the analysis of other QoIs, this study may be extended to quantify aleatoric uncertainty emanating from additional weather, fuel and terrain dependent variables that are not resolved in empirical and semi-empirical models. Furthermore, multifidelity strategies may facilitate, for the first time, a detailed analysis of parametric uncertainties related to specific physical and chemical phenomena, such as pyrolysis and combustion reactions.

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

This research was financially supported by the Beatriz Galindo Program (Distinguished Researcher, BGP18/00026) of the Ministerio de Ciencia, Innovación y Universidades, Spain. The authors thank the anonymous reviewers for their valuable comments and suggestions that helped improve the quality of the paper.

The work presented in this paper used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (NSF), Grant No. ACI-1548562. In particular, the authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin through allocation TG-CTS190073 for providing high-performance computing resources that have contributed to the results presented in this paper.

References

Adcock, C., Ye, Y., Jofre, L., Iaccarino, G., 2020. Multilevel Monte Carlo sampling on heterogeneous computer architectures. Int. J. Uncertain. Quantification 10, 575–594.

Alexander, M.E., Cruz, M.G., 2013. Are the applications of wildland fire behaviour models getting ahead of their evaluation again? Environ. Model. Software 41, 65–71.

Alonso J. J., Eldred M. S., Constantine P., Duraisamy K., Farhat C., Iaccarino G., Jakeman J., 2017. Scalable environment for quantification of uncertainty and optimization in industrial applications (SEQUOIA), in: 19th AIAA Non-deterministic Approaches Conference, p. 1327.

Anderson, K., Reuter, G., Flannigan, M.D., 2007. Fire-growth modelling using meteorological data with random and systematic perturbations. Int. J. Wildland Fire 16, 174–182.

Andrews, P.L., 2018. The Rothermel Surface Fire Spread Model and Associated Developments: a Comprehensive Explanation, RMRS-GTR-371. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 1–121. Technical Report 371.

Arca, B., Duce, P., Laconi, M., Pellizzaro, G., Salis, M., Spano, D., 2007. Evaluation of FARSITE simulator in Mediterranean maquis. Int. J. Wildland Fire 16, 563–572.

Bachmann, A., Allöwer, B., 2002. Uncertainty propagation in wildland fire behaviour modelling. Int. J. Geogr. Inf. Sci. 16, 115–127.

Benali, A., Ervilha, A.R., Sá, A.C., Fernandes, P.M., Pinto, R.M., Trigo, R.M., Pereira, J. M., 2016. Deciphering the impact of uncertainty on the accuracy of large wildfire spread simulations. Sci. Total Environ. 569–570, 73–85.

Bermejo-Moreno, I., Larsson, J., Campo, L., Bodart, J., Emory, M., Palacios, F., Helmer, D., Iaccarino, G., Eaton, J., 2012. Multi-fidelity numerical simulations of shock/turbulent-boundary-layer interaction in a duct with uncertainty quantification. CTR Annu. Res. Briefs 67–79.

Bierig, C., Chernov, A., 2015. Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems. Numer. Math. 130, 579–613.

Bova, A.S., Mell, W.E., Hoffman, C.M., 2016. A comparison of level set and marker methods for the simulation of wildland fire front propagation. Int. J. Wildland Fire 25, 229–241.

Cai, L., He, H.S., Liang, Y., Wu, Z., Huang, C., 2019. Analysis of the uncertainty of fuel model parameters in wildland fire modelling of a boreal forest in north-east China. Int. J. Wildland Fire 28, 205–215.

Cheney, N.P., Gould, J.S., 1997. Fire growth and acceleration. Int. J. Wildland Fire 1, 1–5.

Cheney, N.P., Gould, J.S., Catchpole, W.R., 1993. The influence of fuel, weather and fire shape variables on fire-spread in grasslands. Int. J. Wildland Fire 3, 31–44.
Clark, R.E., Hope, A.S., Tarantola, S., Gatelli, D., Dennison, P.E., Moritz, M.A., 2008. Sensitivity analysis of a fire spread model in a chaparral landscape. Fire Ecology 4, 1–13.
Cruz, M.G., 2010. Monte Carlo-based ensemble method for prediction of grassland fire spread. Int. J. Wildland Fire 19, 521–530.
Cruz, M.G., Alexander, M.E., 2013. Uncertainty associated with model predictions of surface and crown fire rates of spread. Environ. Model. Software 47, 16–28.
Cruz, M.G., Alexander, M.E., 2017. Modelling the rate of fire spread and uncertainty associated with the onset and propagation of crown fires in conifer forest stands. Int. J. Wildland Fire 26, 413–426.
Cruz, M.G., Alexander, M.E., Fernandes, P.A., 2008. Development of a model system to predict wildfire behaviour in pine plantations. Aust. For. 71, 113–121.
Cruz, M.G., Alexander, M.E., Sullivan, A.L., Gould, J.S., Kilinc, M., 2018. Assessing improvements in models used to operationally predict wildland fire rate of spread. Environ. Model. Software. 105, 54–63. https://doi.org/10.1016/j.envsoft.2018.03.027, 1364-8152. https://www.sciencedirect.com/science/article/pii/S1364815218300161.
Doostan, A., Owhadi, H., 2011. A non-adapted sparse approximation of PDEs with stochastic inputs. J. Comput. Phys. 230, 3015–3034.
Ervilha, A.R., Pereira, J.M., Pereira, J.C., 2017. On the parametric uncertainty quantification of the Rothermel's rate of spread model. Appl. Math. Model. 41, 37–53.
Estimating Winds for Fire Behavior, Fire Behavior Field Reference Guide, PMS 437, 2019. National Wildfire Coordinating Group.
Fairbanks, H.R., Doostan, A., Ketelsen, C., Iaccarino, G., 2017. A low-rank control variate for multilevel Monte Carlo simulation of high-dimensional uncertain systems. J. Comput. Phys. 341, 121–139.
Fairbanks, H.R., Jofre, L., Geraci, G., Iaccarino, G., Doostan, A., 2020. Bi-fidelity approximation for uncertainty quantification and sensitivity analysis of irradiated particle-laden turbulence. J. Comput. Phys. 402, 108996.
Fernández-Godino, M.G., Park, C., Kim, N.H., Haftka, R.T., 2016. Review of Multi-Fidelity Models arXiv preprint arXiv:1609.07196.
Fernández-Godino, M.G., Park, C., Kim, N.H., Haftka, R.T., 2019. Issues in deciding whether to use multifidelity surrogates. AIAA J. 57, 2039–2054.
Filippi, J.B., Mallet, V., Nader, B., 2014. Evaluation of forest fire models on a large observation database. Nat. Hazards Earth Syst. Sci. 14, 3077–3091.
Finney, M.A., 1998. FARSITE : Fire Area Simulator — Model Development and Evaluation, Technical Report. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
Finney, Mark A, 2002. Fire growth using minimum travel time methods. Can. J. For. Res. 32 (8), 1420–1424. https://doi.org/10.1139/x02-068.
Finney, M.A., Grenfell, I.C., McHugh, C.W., Seli, R.C., Trethewey, D., Stratton, R.D., Brittain, S., 2011. A method for ensemble wildland fire simulation. Environ. Model. Assess. 16, 153–167.
Floyd, J., McGrattan, K.B., 2008. Validation of A CFD Fire Model Using Two Step Combustion Chemistry Using the NIST Reduced-Scale Ventilation-Limited Compartment Data. Fire Safety Science 9, 117–128. https://doi.org/10.3801/IAFSS.FSS.9-117.
Forney, G.P., 2016. Smokeview, A Tool for Visualizing Fire Dynamics Simulation Data. Volume I: User's Guide. National Institute of Standards and Technology, U.S. Department of Commerce, Washington, DC.
Forney, G.P., Madrzykowski, D., McGrattan, K.B., Sheppard, L., 2003. Understanding fire and smoke flow through modeling and visualization. IEEE Computer Graphics and Applications 23, 6.
Geraci, G., Eldred, M., Iaccarino, G., 2017. A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications. In: 19th AIAA Non-deterministic Approaches Conference. AIAA 2017–1951.
Ghanem, R., Spanos, P., 2003. Stochastic Finite Elements: a Spectral Approach. Dover.
Ghanem, R., Higdon, D., Owhadi, H., 2017. Handbook of Uncertainty Quantification. Springer.
Giles, M.B., 2008. Multi-level Monte Carlo path simulation. Oper. Res. 56, 607–617.
Gorodetsky, A.A., Geraci, G., Eldred, M.S., Jakeman, J.D., 2020. A generalized approximate control variate framework for multifidelity uncertainty quantification. J. Comput. Phys. 408, 109257 https://doi.org/10.1016/j.jcp.2020.109257. ISSN 0021-9991.
Heinsch, F.A., Andrews, P.L., 2010. BehavePlus Fire Modeling System, Version 5.0: Design and Features, Technical Report. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO: U.S.
Janon, A., Klein, T., Lagnoux, A., Nodet, M., Prieur, C., 2014. Asymptotic normality and efficiency of two Sobol index estimators. ESAIM P. S. 18, 342–364.
Jimenez, E., Hussaini, M.Y., Goodrick, S., 2008. Quantifying parametric uncertainty in the Rothermel model. Int. J. Wildland Fire 17, 638–649.
Jofre, L., Geraci, G., Fairbanks, H.R., Doostan, A., Iaccarino, G., 2017. Multi-fidelity uncertainty quantification of irradiated particle-laden turbulence. CTR Annu. Res. Briefs 21–34.
Jofre, L., Domino, S.P., Iaccarino, G., 2018. A framework for characterizing structural uncertainty in large-eddy simulation closures. Flow, Turbul. Combust. 100, 341–363.
Jofre, L., Domino, S.P., Iaccarino, G., 2019. Eigensensitivity analysis of subgrid-scale stresses in large-eddy simulation of a turbulent axisymmetric jet. Int. J. Heat Fluid Flow 7, 314–335.
Jofre, L., Papadakis, M., Roy, P.T., Aiken, A., Iaccarino, G., 2020. Multifidelity modeling of irradiated particle-laden turbulence subject to uncertainty. Int. J. Uncertain. Quantification 10, 499–514.
Linn, R., Reisner, J., Colman, J.J., Winterkamp, J., 2002. Studying wildfire behavior using FIRETEC. Int. J. Wildland Fire 11, 233.
Liu, N., Wu, J., Chen, H., Xie, X., Zhang, L., Yao, B., Zhu, J., Shan, Y., 2014. Effect of slope on spread of a linear flame front over a pine needle fuel bed: experiments and modelling. Int. J. Wildland Fire 8, 1087–1096.
Liu, Y., Jimenez, E., Hussaini, M.Y., Ökten, G., Goodrick, S., 2015. Parametric uncertainty quantification in the Rothermel model with randomised quasi-Monte Carlo methods. Int. J. Wildland Fire 24, 307–316.
Mandel, J., Beezley, J.D., Kochanski, A.K., 2011. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. Geosci. Model Dev. (GMD) 4, 591–610.
Masquelet, M., Yan, J., Dord, A., Laskowski, G., Shunn, L., Jofre, L., Iaccarino, G., 2017. Uncertainty quantification in large eddy simulations of a rich-dome aviation gas turbine. Proc. ASME Turbo Expo GT2017–64835, 1–11.
Mathelin, L., Hussaini, M.Y., 2003. A Stochastic Collocation Algorithm for Uncertainty Analysis, Technical Report NASA 1.26:212153; NASA/CR-2003-212153. NASA Langley Research Center.
McGrattan, K.B., Hostikka, S., Floyd, J., McDermott, R.J., Vanella, M., 2019a. Fire Dynamics Simulator User's Guide.
McGrattan, K.B., McDermott, R.J., Vanella, M., Hostikka, S., Floyd, J., 2019b. Fire Dynamics Simulator Technical Reference Guide, vol. 3. Validation.
Mell, W., Jenkins, M., Gould, J., Cheney, P., 2007. A physics based approach to modeling grassland fires. Int. J. Wildland Fire 16, 1–22.
Morvan, D., Dupuy, J., Rigolot, E., Valette, J., 2006. FIRESTAR: a Physically based model to study wildfire behaviour. For. Ecol. Manag. 234S, S114.
Morvan, D., Accary, G., Meradji, S., Frangieh, N., Bessonov, O., 2018. A 3D physical model to study the behavior of vegetation fires at laboratory scale. Fire Saf. J. 101, 39–52.
Mycek, P., De Lozzo, M., 2020. Multilevel Monte Carlo covariance estimation for the computation of Sobol' indices. SIAM-ASA J. Uncertain. 7, 1323–1348.
Najm, H.N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Annu. Rev. Fluid Mech. 41, 35–52.
O'Brien, J.J., Loudermilk, E.L., Hornsby, B., Hudak, A.T., Bright, B.C., Dickinson, M.B., Hiers, J.K., Teske, C., Ottmar, R.D., 2016. High-resolution infrared thermography for capturing wildland fire behaviour: RxCADRE 2012. Int. J. Wildland Fire 25, 62–75.
Pasupathy, R., Taaffe, M., Schmeiser, B.W., Wang, W., 2014. Control-variate estimation using estimated control means. IIE Trans. 44, 381–385.
Peherstorfer, B., Willcox, K., Gunzburger, M., 2018. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM Rev. 60, 550–591.
Pinto, R.M., Benali, A., Sá, A.C., Fernandes, P.M., Soares, P.M., Cardoso, R.M., Trigo, R. M., Pereira, J.M., 2016. Probabilistic fire spread forecast as a management tool in an operational setting. SpringerPlus 5.
Ramirez, J., Monedero, S., Silva, C.A., Cardil, A., 2019. Stochastic decision trigger modelling to assess the probability of wildland fire impact. Sci. Total Environ. 694, 133505.
Riley, K., Thompson, M., 2016. An uncertainty analysis of wildfire modeling. In: Natural Hazard Uncertainty Assessment: Modeling and Decision Support, pp. 193–213.
Rothermel, R.C., 1972. A Mathematical Model for Predicting Fire Spread in Wildland Fuels. USDA Forest Service Research Paper INT USA, p. 40.
Roy, P.T., Jofre, L., Jouhaud, J.-C., Cuenot, B., 2020. Versatile sequential sampling algorithm using Kernel Density Estimation. Eur. J. Oper. Res. 284, 201–211.
Salvador, R., Piol, J., Tarantola, S., Pla, E., 2001. Global sensitivity analysis and scale effects of a fire propagation model used over Mediterranean shrublands. Ecol. Model. 136, 175–189.
Sá, A.C.L., Benali, A., Fernandes, P.M., Pinto, R.M.S., Trigo, R.M., Salis, M., Russo, A., Jerez, S., Soares, P.M.M., Schroeder, W., Pereira, J.M.C., 2017. Evaluating fire growth simulations using satellite active fire data. Remote Sens. Environ. 190, 302–317. https://doi.org/10.1016/j.rse.2016.12.023, 0034-4257. https://www.sciencedirect.com/science/article/pii/S0034425716305028.
Scott, J.H., Burgan, R.E., 2005. Standard Fire Behavior Fuel Models: a Comprehensive Set for Use with Rothermel's Surface Fire Spread Model. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 1–72. General Technical Report RMRS-GTR-153.
Sobol, I., 1993. Sensitivity estimates for nonlinear mathematical models. Math. Modeling Comput. Experiment 4, 407–414.
Sukys, J., Rasthofer, U., Wermelinger, F., Hadjidoukas, P., Koumoutsakos, P., 2017. Optimal Fidelity Multi-Level Monte Carlo for Quantification of Uncertainty in Simulations of Cloud Cavitation Collapse arXiv preprint arXiv:1705.04374.
Sullivan, A.L., 2009a. Wildland surface fire spread modelling, 1990–2007. 1: physical and quasi-physical models. Int. J. Wildland Fire 18, 349–368.
Sullivan, A.L., 2009b. Wildland surface fire spread modelling, 1990–2007. 2: empirical and quasi-empirical models. Int. J. Wildland Fire 18, 369–386.
Sullivan, A.L., 2009c. Wildland surface fire spread modelling, 1990–2007. 3: simulation and mathematical analogue models. Int. J. Wildland Fire 18, 387–403.
Xiu, D., Hesthaven, J.S., 2005. High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. 27, 1118–1139.
Xiu, D., Karniadakis, G.M., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. 24, 619–644.
Yuan, X., Liu, N., Xie, X., Viegas, D.X., 2020. Physical model of wildland fire spread: parametric uncertainty analysis. Combust. Flame 217, 285–293.