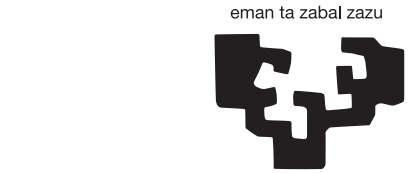
Front Propagation in Random Media



Universidad del País Vasco
Euskal Herriko Unibertsitatea

Author: Andrea Trucchia
Supervisor: Dr. Gianni Pagnini

A thesis submitted for the degree of
Doctor of Philosophy
Bilbao, September 2019

Front Propagation in Random Media

Author: Andrea Trucchia
Supervisor: Dr. Gianni Pagnini

Basque Center for Applied Mathematics
Universidad del País Vasco – Euskal Herriko Unibertsitatea

A thesis submitted for the degree of
Doctor of Philosophy

Bilbao, September 2019

Contributions

This research was carried out at the Basque Center for Applied Mathematics (BCAM) within the Statistical Physics (SP) Group and supported by the Basque Government through the BERC 2014-2017 and BERC 2018-2021 programs. It was also funded by the Spanish Ministry of Economy and Competitiveness MINECO via the BCAM Severo Ochoa SEV-2013-0323 and SEV-2017-0718 accreditations, the MTM2016-76016-R "MIP" project, and the PhD grant "La Caixa 2014". The support of MINECO and Caixa Grants is also acknowledged for the six months research stay at the research facilites of Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS). Funding from "Programma Operativo Nazionale Ricerca e Innovazione" (PON RI 2014-2020) , "Innotavive PhDs with Industrial Characterization" is kindly acknowledged for a research visit at the department of Mathematics and Applications "Renato Caccioppoli" of University "Federico II" of Naples. This work has been performed using the computing infrastructure of the BCAM groups clusters GAUSS and HIPATIA, and the cluster NEMO at CERFACS.

Abstract

This PhD thesis deals with the problem of the propagation of fronts under random circumstances. While front propagation is a physical phenomenon with countless applications, it is notoriously hard to study the advancing of a front when the underlying dynamics is affected by random phenomena that affect the topology and the shape of the studied contour. In this thesis, a statistical model to represent the motion of fronts when are evolving in a media characterized by microscopical randomness is discussed and expanded, in order to cope with three distinct applications: wild-land fire simulation, turbulent premixed combustion, biofilm modeling. In the studied formalism, the position of the average front is computed by making use of a sharp-front evolution method, such as the level set method. The microscopical spread of particles which takes place around the average front is given by the probability density function linked to the underlying diffusive process, that is supposedly known in advance, say in a parametric formulation; the effective front is computed by taking weighted averages of the average fronts adopting the previously chosen probability density as a weight function. The front position is obtained as the weighted mean of fronts calculated by means of the selected sharp front technique, using as weight-function the probability density function which characterizes the underlying microscopic diffusion process.

The adopted statistical front propagation framework allowed a deeper understanding of any studied field of application. (i) In the case of wild-land fire, the proposed splitting allowed to study separately the rate of spread of the fire and the turbulence and fire spotting effects (the latter involving random trajectories of fire brands). This work put the basis for a multi scale analysis of the different macroscopic and meso-scopic factors in fire-spotting driven fires. (ii) In the case of turbulent premixed combustion, the use of the probability density of the front fluctuations allowed for physical insights of flame instabilities. (iii) In the case of biofilm spread modeling, the proposed splitting allowed for the study of the seeding and dispersal phenomenon, studying the generation of satellite biofilm spots generated by an active colony through planktonic cell migration.

The application of this model introduced eventually parameters whose impact on the physical observables of the front spread have been studied with Uncertainty Quantification and Sensitivity Analysis tools. In particular, metamodels for the front propagation system have been constructed in a non intrusive way, by making use of generalized Polynomial Chaos expansions (gPC) and Gaussian Processes (GP). Such techniques have been transversal to nearly any application of the presented thesis. Different techniques for gPC have been compared for the study case of the studied statistical front propagation model in wild-land fire with the occurrence of fire-spotting, keeping GP as a benchmark. On the other hand, the UQ and SA techniques that have proven to be successful for the latter application have been applied to a recent 1D biofilm model that has recently been added a multi species invasion module. The use of metamodel-based UQ and SA techniques had a two-fold purpose: on the one hand, it allowed to shed some light on the underlying physics, ranking the parameters according to their relevance, measuring their interaction in affecting the observables, and quantifying the uncertainty in the outputs given random input parameters. On the other hand, the given front propagation models have been a chance to rank metamodels and explore their suitability to a large spectrum of applications.

Contents

List of Figures

Chapter 0

Introduction

This thesis deals with the propagation of fronts (interfaces) in either a random medium or in a deterministic one, where the dynamics pushing the propagating front affected by any kind of randomness. This randomness is usually associated with phenomena occurring at scales that are smaller than the scales of the front object of the study, i.e. we are dealing with microscopical random fluctuations of the front.

Front propagation is a discipline that stands in between several scientific fields such as engineering, computer science and mathematics.

Letting the modeled front embrace a random environment or a random propagating law can help the bettering of different models in their task of adhering as close as possible to reality. To name a first example,in the setting of wild-land fire propagation, a simple propagation law that neglects the rise and fall of burning embers, as well as the turbulent transport of hot air, may be not as accurate as expected in predicting the topology and the spatial extension of a burning front, this being detrimental to overall forecasting potential of the model itself. Shifting the attention to biological modeling, a standard front propagation approach for a planar colony of bacteria cannot grasp easily the seeding and dispersal put in place by bacteria switching from sessile to planktonic activity, that would eventually determine the formation of mature colonies in places richer in resources.

In order to describe the motion of random fronts, in this thesis we shall adopt a modeling framework initially thought for studying turbulent premixed combustion, that consists of a splitting between a deterministic front and the super-imposed random fluctuations. The deterministic front can be computed effectively by making use of the algorithms and know-how already present in literature. On the other hand, the random fluctuations are deeply interconnected with the microscopical features of the modeled phenomenon. The description of the aleatory phenomena at the small scale can be modeled via the use of random variables. In the case of the approach adopted in the present Thesis, we limit ourselves to the probability density function of random displacements. The more physics is known about the small scale phenomena, the more accurate the probability density function of the random effects can be conceived. The underlying dynamics is allegedly well described by

a suitable distribution, that will depend on a set of parameters. The final model will then comprise of the standard parameters of a front propagation study at the macro-scale (external velocity field, curvature effects...) while at the micro-scale an accurate choice of the parameters of the probability density is bound to play a paramount role. Most of the models obtained in this framework exhibits thus a *concurrent multi-scale* behavior, with different scales contributing at the same time to the front topology (i.e. number of isolated fires in a wild-land fire model) and front position. Even though the formal structure of the model is common to any application, each particular scenario needs *ad-hoc* functions to express this connection between the microscopic scale and the macroscopic front advancing. It is remarked that even if we are *de facto*splitting the problem of front advancing in two separate problems, the introduced functions are relatively straightforward speaking in mathematical terms, and a clear advantage of this approach is that as long as there is enough information to provide a good modeling, the totality of the small scale modeling is encapsulated inside of the prescribed probability of microscopic displacements, and there is no need to recur to a possibly stiff system of coupled PDEs.

As we mentioned earlier, this modeling framework entails the introduction of parameters related to the microscopic scale. Since this approach is quite new in the mathematical landscape, the influence of each parameters on the output of the model runs (i.e., on a set of physical observable representative of the quantity of information forecast by a model execution) needs to be evaluated. This is useful because ranking correctly the input parameters could on the one hand give important insights on the leading physical processes in the modeled phenomena, and on the other hand give important information to the experimentalists about which parameter should be measured with more care. This problem is addressed by the Sensitivity Analysis (SA). Another problem is to quantify how the uncertainty on the input parameters affects the overall uncertainty on the physical quantities of interest of the model forecast. An answer to this problem, in the case of the proposed approach, would also quantify the uncertainty on the parameters of the prescribed probability density function of the fluctuating part. This issue is answered by the so called Uncertainty Quantification (UQ). In this thesis, both Sensitivity Analysis and Uncertainty Quantification are applied to almost any declination of the model, with a particular stress on wild-land fire application. Despite the fact that in this thesis are used only synthetic settings, that is, the analyzed data for UQ and SA is created by the means of simulations, reaching the desired level of accuracy in both UQ and SA would have implied trespassing the computational budget. To circumvent this limitation, a computationally cheap replica (known as *metamodel*, or *surrogate* of the front propagation model is sought. Instead of modifying the model equations, a so called *ensemble method* is pursued. In other words, databases are built via an opportune sampling of the uncertain parameters. All the information to build the surrogate of the model is then inferred from the database, and the computationally cheap surrogate is launched the necessary number of times in order to retrieve the needed information about UQ and

SA. In order not to get algorithm dependent results, two different families of surrogate models have been applied. The first one is the Gaussian Process one, a technique used in unsupervised machine learning. The second one is a spectral method, generalized Polynomial Chaos (gPC). Due to the non-linearities exhibited by the processes involved, we investigate the impact of different choices of the gPC polynomial basis (full or sparse) on the surrogate performance for a fixed computational budget. Using a sparse polynomial basis helped in reducing the size of the stochastic problem by only choosing the most significant polynomial basis components, and help to better capture a complex model response to variations in the input factors.

Structure of the Thesis This thesis is structured as follows:

- In Chapter 1, some insights are given on the broad topic of front propagation. More specifically, the two main families of methods for front propagation modeling, namely reaction diffusion equations and sharp front methods, are described. In the latter group, the one actively adopted throughout the thesis, the main features of Level Set Method (LSM), a well known Eulerian front propagation scheme, and DEVS, a Lagrangian front tracking scheme, are pointed out.

- In Chapter 2, the reader is introduced to the key concepts of Uncertainty and Sensitivity Analysis. Several UQ and SA techniques (especially the ones connected to surrogate modeling) do constitute an essential expertise cutting across nearly any application of the presented front propagation model. Therefore, the ultimate aim of the Chapter is to cover the main ideas and techniques adopted in the rest of the Thesis.

- Chapter 3 begins with a description of the modeling framework for statistical front propagation adopted in this Thesis. In particular, it is shown how the proposed formalism reconciles two approaches considered alternative to each other that is the ones based on moving interface schemes (sharp front schemes such as LSM and DEVS) and the ones based on reaction-diffusion equations. In the last part of the Chapter, an example of application of this framework is given. More specifically, the simulation of a front of complex initial geometry moving in a random anomalously diffusive media is performed. UQ and SA techniques described in Chapter 2 are adopted in order to study two parameters connected to the random particle displacement.

- Chapter 4 is about the application of the model to Premixed Turbulent Combustion. After an introductory part, some thoughts deriving from the application of the proposed approach to combustion instabilities are delivered.

- Chapter 5 deals with the wild-land fire modeling application field. At the beginning of the Chapter, the main model and techniques adopted in literature to model and simulate wild-land fire spread are reviewed. Afterwards, the phenomenon of fire-spotting is described and

the main model available for its prediction are listed. The mathematical adjustments necessary for the model to be applied in the prediction of wild-land fires including turbulence and fire spotting effects. Its parametrization, named `RandomFront`, is presented as well. The output of the model in two different simulation environment (LSFire and WRF-SFire) is then discussed with a response analysis. UQ and SA of a simple test case for the model with `RandomFront` parametrization is discussed. The Chapter concludes with the latest advances in this research, regarding multi-scale modeling of fire spotting.

- In Chapter 6 the application of the front propagation model to microbial biofilm is described. At first, the biological phenomenon is shortly described. After that, the presented framework is adapted to model seeding and dispersal in planar biofilms. Synthetic experiments are initially performed. The capabilities of the model in reproducing reality are then tested against a simple *in vitro* test case. The chapter concludes with the application of surrogate-based UQ and SA to a detailed 1D biofilm model oriented to engineering applications, to study the parameter of a species invasion submodel. The extension of the biomass front is studied under random bacterial invasion parameters.

- At last, in Chapter 7 the future perspectives are listed, and the main conclusions are drawn.

List of Publications

In this Chapter are reported all the papers (published and in preparation), conference proceedings, given talks and presented posters which are the result of the PhD course.

Use of Symbols

In the following, we will connect each result to its corresponding Chapter(s).

- Chapter 2 is related to the use of Uncertainty quantification and Sensitivity Analysis techniques.
- Chapter 3 is related to the framework of statistical front propagation object of this thesis.
- Chapter 4 is related to the application of the statistical front propagation model to Turbulent Premixed Combustion. Its symbol is (Ch.4).
- Chapter 5 is related to the application of the statistical front propagation model to wildland fires.
- Chapter 6 is related to the study of bacterial biofilms, with or without the use of the proposed framework of statistical front propagation.

List of Publications

Publications in Peer Reviewed Journals

[(Ch.3) (Ch.5)] Ref. [283]: **A. Trucchia,** V. Egorova, A. Butenko, I. Kaur, and G. Pagnini. *Randomfront 2.3: a physical parameterisation of fire spotting for operational fire spread models – implementation in WRF-Sfire and response analysis with LSFire+.* Geoscientific Model Development, 12(1):69–87, 2019.

[(Ch.2) (Ch.3) (Ch.5)] Ref. [284]: **A. Trucchia,** V. Egorova, G. Pagnini, and M.C. Rochoux. *On the merits of sparse surrogates for global sensitivity analysis of multi-scale nonlinear problems: Application to turbulence and fire-spotting model in wildland fire simulators.* Communications in Nonlinear Science and Numerical Simulation, 73:120 – 145, 2019.

[(Ch.2) (Ch.6)] Ref. [285]: **A. Trucchia,** M.R. Mattei, V. Luongo, L. Frunzo, and M.C. Rochoux. *Surrogate-based uncertainty and sensitivity analysis for bacterial invasion in multi-species biofilm modeling.* Communications in Nonlinear Science and Numerical Simulation, 73:403 – 424, 2019.

[(Ch.3) (Ch.5)] Ref. [77]: Vera N. Egorova, **A. Trucchia,** and Gianni Pagnini. *Fire-spotting generated fires. part i: The role of atmospheric stability.* Applied Mathematical Modelling (in press), 2019.

[(Ch.3) (Ch.4)] Ref. [287]: **A. Trucchia** and Gianni Pagnini. Restoring property of the Michelson-Sivashinsky equation. Combustion Science and Technology, 2019. Accepted (2019).

Submitted Papers

[(Ch.3) (Ch.5)] Ref. [78]: Vera N. Egorova, **A. Trucchia,** and Gianni Pagnini. *Fire-spotting generated fires. part ii: The role of flame geometry.* Applied Mathematical Modelling, 2019. Submitted.

Papers in preparation

[(Ch.3)] **A. Trucchia**, A. Mentrelli, G. Pagnini, *The role of diffusion and curvature in front propagation*, manuscript in preparation

[(Ch.3) (Ch.6)] **A. Trucchia**, F. Villa, L. Frunzo, G. Pagnini, *Seeding Dispersal Modeling For Systems of Planar Microbial Biofilms*, arXiv preprint, arXiv:1712.04832, `https://arxiv.org/abs/1712.04832`

Outreach: proceedings, talks and posters

Proceedings

[(Ch.3) (Ch.5)] Egorova V. N.; Pagnini G.; **Trucchia A.** *Wildland fire propagation modeling: fire-spotting parametrisation and energy balance.* Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2017, pp. 805 - 813, 2017-07-04

[(Ch.3) (Ch.5)] Egorova V. N.; Pagnini G.; **Trucchia A.** *Wildland fire propagation modelling.* MODELLING FOR ENGINEERING AND HUMAN BEHAVIOUR 2017 Extended abstract, December 2017

[(Ch.3) (Ch.4)] Pagnini G.; **Trucchia A.** *Darrieus-Landau instabilities in the framework of the G-equation.* Digital proceedings of the 8th European Combustion Meeting, 18-21 April 2017, Dubrovnik, Croatia, April 2017

[(Ch.3) (Ch.4)] G. Pagnini, **A. Trucchia**, *Quasi-probability Approach for Modelling Local Extinction and Counter-gradient in Turbulent Premixed Combustion* . Proceedings Joint Meeting the German and Italian Sections of the Combustion Institute, 23-26/05/2018, Sorrento, Italy

[(Ch.2) (Ch.3)] **A. Trucchia**, G. Pagnini, *The role of the environment in front propagation*, Proceedings of the 18th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2018 July 914, 2018, 2018-07-09

[(Ch.3) (Ch.5)] Egorova V. N., **Trucchia A.** , Pagnini G.; *Concurent multi-scale physical parametrization of fire-spotting: A study on the role of macro- and meso-scale characteristics of the system* , Advances in Forest Fire Research, 2018

[(Ch.3) (Ch.5)] Pagnini G.; Egorova V.; **Trucchia A.**; Mentrelli A.; Kaur I. , *Wildfire Propagation Modelling*, Geophysical Research Abstracts Vol. 20, 2018

[(Ch.3) (Ch.4)] Pagnini G., **Trucchia A.**, *Front Curvature Evolution and Hydrodynamics Instabilities*, Proceedings/Extended Abstract Book (6 pages) of the XXXX Meeting of the Italian Section of the Combustion Institute, Rome, Italy, 2017-06-07

Outreach

Invited Talks

[(Ch.3)] Presentation of seminar *Front propagation in random media* at the *GNFM XLI Mathematical Phyisics Summer School*, Ravello, Italy, 15/09/2016

[(Ch.2) (Ch.3)] Presentation of seminar *"The role of the environment in front propagation"* at the *International Conference on Computational and Mathematical Methdos in Science and Engineering, CMMSE 2018* (July 9-14 2018, Rota, Cadiz, Spain)

[(Ch.2) (Ch.3),(Ch.5)] Seminar *"Surrogated - based Uncertainty Quantification and Sensitivity Analysis of a Wild-land fire model"* at *BCAM LIGHT SEMINAR*, May 3rd 2018, BCAM, Bilbao, Spain

[(Ch.2) (Ch.6)] Seminar *"Surrogate-based Uncertainty Quantification and Sensitivity Analysis for Bacterial Invasion in Multi-species Biofilms"* at *BCAM LIGHT SEMINAR*, February 19th 2019, BCAM, Bilbao, Spain

[(Ch.2) (Ch.3),(Ch.5)] Talk *Surrogate Analysis of turbulence and fire spotting in wild-land fire modeling* delivered at *FIRST BYMAT CONFERENCE: BRINGING YOUNG MATHEMATICIANS TOGETHER* May 07-09 2018 , ICMAT, Universidad Autonoma de Madrid, Madrid, Spain.

[(Ch.2) (Ch.3),(Ch.5)] Seminar *"Uncertainty Quantification and Sensitivity Analysis of a Wild-land fire model"* delivered at Mathematics Department "Renato Caccioppoli" at Federico II University, Naples, Italy, November 2019

[(Ch.3) (Ch.6)] Seminar *"Seeding and Dispersal of Planar Microbial Biofilms: a chance for modelling"* delivered at workshop *QBIO2019 - Quantitative Biomedicine for Health and Disease*, Bilbao, February 13th, 2019

Research visits

1. Six months research visit at CERFACS (Toulouse, France) under the supervision of Dr Mélanie Rochoux, from 01/09/2017 to 28/02/2018.

2. Short research visit at Mathematics Department of Federico II University (Naples, Italy) under the superviosion of Prof. Luigi Frunzo, from 05/11/2018 to 12/11/2018

Posters

[(Ch.3) (Ch.5)] V.N. Egorova, **A. Trucchia** and G. Pagnini, *Numerical simulations of fire-spotting: flame characteristics formulation,* XVIII Spanish-French School Jacques-Louis Lions about Numerical Simulation in Physics and Engineering, Las Palmas de Gran Canaria, 25-29 June 2018

[(Ch.2) (Ch.3), (Ch.5)] **A. Trucchia**,V.N. Egorova, M.C. Rochoux and G. Pagnini, *Surrogate-based analysis of turbulence and fire spotting in wild-land fire modelling*, XVIII Spanish-French School Jacques-Louis Lions about Numerical Simulation in Physics and Engineering, Las Palmas de Gran Canaria, 25-29 June 2018

[(Ch.3) (Ch.4)] G. Pagnini, **A. Trucchia**, *Darrieus-Landau instabilities in the framework of the G-equation*, 8th European Combustion Meeting held in Dubrovnik, Croatia April 18-21 2017

[(Ch.3) (Ch.6)] **A. Trucchia**, L. Frunzo, G. Pagnini, *A Front Propagation Model For Phototropic Microbial Biofilms Dynamics*, at EUROBIOFILMS 2017: FITH EUROPEAN CONGRESS ON MICROBIAL BIOFILMS–BASIC AND CLINICAL ASPECTS. AMSTERDAM ,THE NETHERLANDS, 19-22 SEPTEMBER 2017.

[(Ch.5)] Gianni Pagnini, Vera N. Egorova, **Andrea Trucchia**, Andrea Mentrelli and Inderpreet Kaur, *Wildfire propagation modelling* EGU 2018 (European Geosciences Union General Assembly 2018), Vienna, Austria, 8-13 Abril 2018.

If the path be beautiful, let us not ask where it leads.

Anatole France

Chapter 1

Introduction to Front Propagation

The main aim of the thesis is to study a novel method for front propagation, and its applications to different scenarios. In order to do so, an accurate description about what front propagation is appears to be necessary.

The first question that shall be addressed in this introductory Chapter, is surely "What is a front?". The reader surely shall bear in mind some intuitive notion of front, as a line, or a surface, that more or less sharply separates two parts of the space that differ in something easy to detect at a first glance. For example, in a wild land fire seen from above, it is clear that a fire line (one-dimensional front) neatly separates two parts of the territory, the burnt (or still burning) one and the untouched one.

Speaking the language of mathematics, a front is a solution of a spatially distributed system connecting two steady states. [221, 189]. From the dynamical systems point of view, a front can be defined as a heteroclinic orbit of a dynamical system in a co-moving frame of reference.

Front propagation and interface motion occur in many scientific areas such as chemical kinetics, combustion, biology, transport in porous media, and industrial deposition processes. It is quite an ubiquitous phenomenon, and its acquired thus a scientific and technical relevance.

The modeling of such varied spectra of situations can be ascribed to two main families of equations:

- (Systems of) non-linear Parabolic Partial Differential Equations (PDEs), belonging to the case of Reaction Diffusion (R-D) systems;
- A so called *Sharp Interface*, given e.g. by (systems of) hyperbolic Hamilton Jacobi equations, or with the aid of *Lagrangian Techniques*.

In the following we will briefly describe the two main families, bearing in mind that the formulation of Chapter 3 will set a bridge between the two approaches.

1.1 Scalar R-D equations in Homogeneous media

One of the simplest model of scalar R-D equation is the well known model

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(u) \tag{1.1}$$

where $x \in \mathbb{R}$,. In spite of its simplicity it has been used in a wide number of applications, ranging from chemical kinetics to combustion.

The quantity u in (1.1) may stand for the concentration of a particular reactant, the population density of a biological species, temperature of a mixture, or whatever the practitioner needs to represent in order to compute a travelling front.

The term f is a (possibly non-linear) reaction term, that couples the diffusive $u_t = u_{xx}$ part of (1.1). It is this specific term that governs the kinetics of the front propagation. In Ref. [311] a short list of nonlinearities, arising in literature and applications, are presented. We report here this selection:

1. $f(u) = u(1-u)$ the KPP (Fisher) non-linearity;
2. $f(u) = u^m(1-u)$, $m \geq 2, m \in \mathbb{N} : m^{th}$ order Fisher non-linearity (Zeldovich nonlinearity is the case when $m = 2$);
3. $f(u) = u(1-u)(u-\mu)$ is the bistable nonlinearity;
4. $f(u) = e^{-E/u}(1-u)$ is the *Arrhenius combustion nonlinearity* or *combustion nonlinearity with activation energy E but no temperature cutoff*;
5. $f(u) = 0 \forall u \in [0, \theta], f(u) > 0 \forall u \in (0,1)$, f *Lipschitz continous* is the *combustion nonlinearity with ignition temperature θ* .

The first two types are borrowed from chemical kinetics (e.g., from autocatalytic reactions). The second type is the higher order generalization of the type 1. Type 3 has been introduced in a biological framework (Fitz-Hugh-Nagumo , FHN systems) and from phase-field models for solidification processes. The last two types come from premixed combustion studies, type 4 being the $\sigma \to 0$ limit case.

The isotropic scalar R-D equation in d dimension, $d > 1$ is given by

$$\frac{\partial u}{\partial t} = \nabla_x u + f(u) \tag{1.2}$$

where ∇_x is the operator

$$\nabla_x(\cdot) \equiv \sum_{i=1}^{d} \frac{\partial^2(\cdot)}{\partial {x_i}^2} \tag{1.3}$$

1.2 Scalar R-D equation in non-homogeneous media: going towards Complexity

1.2.1 R-D Fronts in Periodic Media

This section will briefly address the problem of front propagation in a medium that is not homogeneous, described by the means of a scalar R-D equation.

In literature we can see that this problem has been approached by adding increasing layers of complexity. To be more specific, it has been addressed at first analyzing periodic media, and then looking at the more realistic, but also more challenging, random media setting. Periodic media assumption comes in handy when multi-scale problems are studied. Problems characterized by two or more scales are common in applications, e.g. when the effective permeability needs to be determined for flows in porous media. In the example of the flow in porous media, the two scales are the sample one and the one related to pores. These two scales are usually different in magnitude, and this makes the solution of the problem difficult. Some mathematical tool to express at the large scale the contribution of the collective effect at the small scale. When the small scale can be assumed to be organized in a periodic structure, this up-scaling problem can rely on the framework of homogenization theory. We recall that in this setting, homogenization is an Asymptotic Analysis Theory that originates from material engineering, or more precisely, from understanding the way the constitutive equation of composite material can be retrieved from the constitutive equation of each component of the given material and from their topological and geometrical distributions. With the ansatz of two-scale homogenization, the general form of traveling fronts can be derived for some of the cases from 1 to 5 of the previous section's list, and analytical results have been found for stability of the fronts as well as estimates on the front speeds (see Ref. [311]).

1.2.2 RD Fronts in Random Media

A random setting is in a plethora of cases a way more practical assumption while modeling a physical phenomenon. In engineering, sometimes the collect of information on the medium is a costly procedure and thus some estimation that makes uses of a random, uncertain setting has to be adopted, rather than a precise and analytical periodic expected behavior.

As can be found in any book of fluid dynamics, see e.g. Ref. [22], turbulent flows are known to be intrinsically random in nature, and one of the main approach to be considered in describing the motion of a turbulent fluid is by the means of statistical quantities such as energy spectrum and two point velocity correlation functions.

In literature two broad categories of random media can be found, namely the tame media and the wild ones.

Tame Random Media

A tame random medium can by defined by a stochastic process $X(\mathbf{x}, \omega)$ with finite moments (at least up to the second one) and short range correlations. An example is $X(\mathbf{x}, \omega) = \xi_n(\omega)$, for $x \in [n-1, n), n = 1, 2, 3 \dots$, where the ξ_n are independent identically distributed (i.i.d.) random variables with finite first and second moments,

$$\mathbf{E}[\xi_n] = \mu, \qquad \mathbf{E}[\xi_n^2] = \sigma^2 + \mu^2.$$

We denoted by the positive constant σ^2 the variance of the stochastic process where σ is the standard deviation. Applying the classical Central Limit Theorem , we have that the sum $S_n = \xi_1 + \ldots + \xi_n$ is characterized by the limit

$$\lim_{n\to\infty} \frac{(S_n n\mu)}{\sigma\sqrt{n}} = W_1$$

where W_1 is a unit Gaussian random variable (that is, a random variable with a normal distribution) . The central limit theorem is a robust result and can be extended to cases with short correlations. The product expectation $\mathbb{E}[(X(s) - \mathbb{E}[X(s)])(X(t) - \mathbb{E}X(t))]$ is generally called the covariance function, and the correlation is the ratio of the covariance to $\sigma_s\sigma_t$, where $\sigma_s^2 = \mathbb{E}[X(s)^2] - (\mathbb{E}[X(s)])^2$. The covariance or correlation is a measure of the degree of independence. For a stationary random process, the distribution of X is invariant under translation in x, so σ_s is a constant, and we may just use covariance for correlation if the variance is finite.

This framework of random processes can be effectively linked to front propagation in a random medium. We may in fact model the location of a one-dimensional random front as a time-dependent random process $S(t, \omega)$. In a scalar R-D setting, the front can have the interpretation of the iso-contour of the quantity of interest with respect to a fixed value, i.e. $\frac{1}{2}$.

The main concept of the R-D approach to random front propagation is that the position of the front at a given time can be computed by the sum of the space elapsed by a particle that is advected by an average constant speed, and a noise term.

That will translate into formulas as

$$S(t) \approx c^* t + W \tag{1.4}$$

with W an unknown noise term. This writing can be further exploited and made rigorous by the means of the Central Limit Theorem, at least in the case of one dimensional Burgers fronts [311]:

$$\lim_{t\to\infty} \frac{S(t) - ct}{\sigma\sqrt{t}} = W_1 \tag{1.5}$$

As stated in [311], this result holds for a much more general class of equations than the conservative PDEs of Burgers type. However, PDE of conservative type allow for an interpretation of (1.5) by the means of mass conservation arguments.

The homogenization process that in the periodic case allowed the derivation of front speeds , sees his alias in this random settings as the law of large numbers for a sum of i.i.d. random variables. However, even for tame random media, there is more to be added to represent realistically the front position. The term $\sqrt{t}W$ is the new phenomenon taking place, *front fluctuation*, which has no analogue in deterministic front problems (being them in a homogeneous or periodic medium). In principle, one has to understand both the mean field phenomenon (a homogenization result on the average front speed) and the statistics of the front fluctuation (Gaussian or otherwise) in order to completely describe a random front. This is certainly a more challenging task. The modeling framework adopted in this Thesis takes into account this intrinsic random fluctuations by the means of the random process probability density function, without having to model explicitly the random contributions like in 1.5.

Wild Random Media

A wild random medium is one in which the second or first moment of the random variable which characterizes the front motion in the medium is infinite.

In the case of a sum of i.i.d. random variables, one encounters non-Gaussian stable laws (see e.g. Reference [100]). If the first moment is still finite, then the following weak convergence holds:

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}[\xi_1]n}{A_n} = Y \tag{1.6}$$

where $\lim_{n\to\infty} \frac{A_n}{\sqrt{n}} = +\infty$.

The random variable Y has a stable probability distribution of exponent $\alpha \in (1,2)$ whose characteristic function can be derived explicitly. Such a random variable can be viewed as having a so called *fractional moment* α. One still has a law of large numbers, but the scaling of the fluctuation (if it exists) is a power larger than $\frac{1}{2}$, and so it falls below the definition of anomalous. If the probability density function has an even slower tail, so that the first moment is infinite but a fractional moment $\alpha \in (0,1)$ is finite, then

$$\lim_{n\to\infty} \frac{S_n}{A_n} = Z \tag{1.7}$$

with Z is a stable random variable with exponent $\alpha \in (0,1)$. If in (1.7) the front location is represented by S_n, the fluctuations may provoke front acceleration up to leading (macroscopic) order [311].

1.3 Sharp Interface Fronts

Another way of looking at the challenging problem of front propagation is a formulation where the observable of the front is not given by a smooth function that evolves through a set of R-D PDEs,

but rather a so-called sharp interface, that is a discontinuity that propagates in space and time. Such method give as an output two distinct sub-domains, the inner and the outer, and for a given point of the interface, only a front speed is given, not a spectrum of speeds like the R-D case. There is, therefore, an hyper-surface to be analytically studied or to be discretized numerically. In some cases, the literature will refer to the method as "front tracking" rather than "front propagation scheme", e.g. when the front is represented a series of markers that evolve following Lagrangian laws.

In this thesis we will deal with at least two methods that give as an output a propagating interface.

The first is the so called Level Set Method, an Eulerian formulation that follows an iso-contour of a higher dimensional scalar function that evolves through an Hamilton Jacobi (HJ) set of equations. This method has been introduced in the seminal work of *Osher and Sethian* [196] in 1988, and has reached an impressive amount of application fields, ranging from multi-phase flows to material science [44], with combustion [136] being the first notable field of application of such method. The combustion community however has historically adopted the terminology G-equation addressing formally to the same equation.

The second strategy is a front tracking one, shifting the problem from an Eulerian Setting (analyzing some field in a position $\mathbf{x}$ at time t, $\quad \forall(\mathbf{x}, t) \in (\mathbf{R}^d \times \mathbf{R})$ to a Lagrangian one, following the trajectories of a finite set of point (which discretizes the front position at time t) characterized by an initial position and a according to a speed function.

Lagrangian front tracking is also known under the name of marker method (see Ref. [138]); From a numerical point of view, this Lagrangian technique with computational cost simulates the evolution of an interface without the need of any underlying grid to depict the state of the system. It is mandatory to specify that topology change of a front is not, in principle, easy to be dealt with in this setting, while with Level Set Method formulation the merging or detaching of fronts come "for free" already built-in in the analytical formulation.

1.4 Level Set Method

The level set method is in a few introductory words an initial value problem for a scalar field $\varphi(\mathbf{x}; t) \in \mathbb{R}$ whose level surfaces represent interfaces. An equation for such field has been formulated in [307] , and [249]. Despite its simplicity, its formulation can be used in a wide range of situation, in both homogeneous and random (turbulent) environment.

The field equation of the Level Set Method (LSM from now on) is the following:

$$\frac{\partial \varphi}{\partial t} + \mathbf{v}\nabla\varphi = u_f \left|\nabla\varphi\right| \tag{1.8}$$

where $\mathbf{v}$ is a given (macroscopic) flow field, possibly turbulent, and u_f is a *contact speed*, that in the application field of combustion is the chemical velocity of the flame. Its left side is a Total

(material, convective) derivative $D\varphi/Dt$. The right hand side is a nonlinear source term depending on the gradient of the advected field.

Its derivation is quite straightforward, not dissimilar in its nature from many results of classical Hydrodynamics.

Let us start with a point $\mathbf{x}_0$ such that $\varphi(\mathbf{x}_0, t_0) = c$. We shall then expand in Taylor series up to the first order the field $\varphi(\mathbf{x}, t)$ in a neighborhood of $(\mathbf{x}_0, t_0)$:

$$\varphi(\mathbf{x}, t) \approx \varphi(\mathbf{x}_0, t_0) + (\mathbf{x} - \mathbf{x_0}\nabla\varphi(\mathbf{x}_0, t_0) + (t - t_0)\frac{\partial\varphi(\mathbf{x}_0, t_0)}{\partial t} \tag{1.9}$$

For $t \to t_0$ we shall get to the formulation for $\mathbf{x}(t)$:

$$\frac{\mathbf{x}(t) - \mathbf{x_0}}{t - t_0} = \mathbf{v}(\mathbf{x}_0, t_0) - u_f\frac{\nabla\varphi_0}{|\nabla\varphi_0|} \tag{1.10}$$

where $\varphi_0 \equiv \varphi(\mathbf{x}_0, t_0) = c$ because of the initial assumption.

The velocity of a point on an iso-contour (that can be, for example, the burning surface of a flame or the front between two distinct phases in a multi-phase flow) is thus the sum of an advection velocity and of the (chemical, in the case of combustion) propagation of the iso-contour in the normal direction with respect to the surface. In this formulation, by making good use of the notion of iso-contour, we can get an useful analytical formula for the normal to the interface:

$$\mathbf{n} = \frac{\nabla\varphi}{|\nabla\varphi|} \tag{1.11}$$

The (1.8) equation can be formulated also as

$$\frac{\partial\varphi}{\partial t} + \mathbf{v}(\mathbf{x}, t) + u_f(\mathbf{x}, t)\mathbf{n}\nabla\varphi = 0 \tag{1.12}$$

With the desired iso-contour φ_0 advected by the external velocity field $v(\mathbf{x}, t)$ and by the problem-dependent *contact* growth $u_f(\mathbf{x}, t)$.

1.4.1 Normal, Curvature, Signed Distance Setting

The level set method, by augmenting the dimension of the problem (i.e., trading the problem of tracking a curve in a plane with the problem of computing the evolution of a moving surface $\varphi = F(x, y; t)$ gives us explicit representations for the normal vector and the curvature of the front. The normal to the evolving front at each point $\mathbf{x} \in \Gamma(t)$ is given by Equation (1.11); note that in this case the normal vector field is given for any point in the domain, not only in $\Gamma(t)$. Anyway, the notion of normal to a surface looses its significance outside of the surface.

Mean curvature κ of the interface is defined as the divergence of the normal $\mathbf{n}$,

$$\kappa(\mathbf{x}, t) = \nabla \cdot \left(\frac{\nabla\varphi(\mathbf{x}, t)}{\|\nabla\varphi(\mathbf{x}, t)\|}\right), \tag{1.13}$$

Equation (1.12) with a curvature-dependent speed reads (after substituting Equation (1.11)):

$$\frac{\partial \varphi}{\partial t} = \mathcal{V}_0(1 - \lambda_M \kappa(\mathbf{x}, t)) \left\| \nabla \varphi \right\|, \tag{1.14}$$

$\mathcal{V}_0$ is a drifting speed here assumed constant and λ_M is a parameter that regulates the influence of the curvature on the front speed. In combustion literature, λ_M is called Markstein length.

By including the curvature, the mathematical problem expressed by (1.8) shiftes from hyperbolic to parabolic.

Signed distance function

Given a set $\Omega(t)$, we call a *signed distance function* a function φ_s that for any given point $\mathbf{x} \in \mathcal{S} \subseteq \mathbb{R}^d$ gives

$$\varphi_\Gamma(\mathbf{x}) = \begin{cases} dist(\mathbf{x}, \partial\Omega(t) & \text{if } \mathbf{x} \in \Omega(t) \\ -dist(\mathbf{x}, \partial\Omega(t) & \text{if } \mathbf{x} \notin \Omega(t) \end{cases} \tag{1.15}$$

From Equation (1.15), we have that if the boundary of $\Omega(t)$, $\partial\Omega(t) \equiv \Gamma$, is piecewise smooth, the signed distance function φ_Γ is differentiable almost everywhere, its gradient satisfying the so called *Eikonal Equation*

$$|\nabla \varphi_s| = 1 \tag{1.16}$$

Please note that the signed distance function is a Level Set function since its $0-$level iso-contour is nothing but the interface $\Gamma(t) = \partial\Omega(t)$. It is not only one of the infinite variation of level set functions for describing such interface, but it is one of the preferred ones in applications due to numerical stability issues and its access to useful formulas to describe normal vector to interface and its mean curvature [252].

In this formulation, choosing $\varphi = \varphi_\Gamma$, we shall have

$$\mathbf{n} = \nabla(\varphi) \tag{1.17}$$

$$\kappa = \Delta(\varphi) \tag{1.18}$$

Arrival time formulation

It is worth pointing out that an alternative representation of the front propagation is possible whenever the velocity $\mathcal{V}(\mathbf{x}, t)$ is constant in time and strictly positive (or strictly negative). In fact, let $\tau(\mathbf{x})$ be the *arrival time* function that represents the temporal instant at which the front reaches the point $\mathbf{x}$, then $\|\nabla\tau\|$ is the rate of change of the arrival time with respect to the change in the front distance, i.e.,

$$\|\nabla\tau\| = \frac{1}{\mathcal{V}}, \quad \tau(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in \Gamma_0. \tag{1.19}$$

Equation 1.19 is known as eikonal equation and it is a time-independent version of the level set equation (3.5). It has the important advantage of allowing – when applicable – for a great reduction of the computational costs.

1.5 Insights of Level Set Method

We can be slightly informal and describe the formulation of (1.8) as the following so called Hamilton Jacobi partial differential equation:

$$\frac{\partial \varphi}{\partial t} + H(\nabla \varphi, \mathbf{x}) = 0 \tag{1.20}$$

One of the main subtleties that arises in solving this equation is that the solution need not be differentiable, even with arbitrarily smooth boundary data. It is common knowledge, indeed, that solutions of Hamilton Jacobi Equations such as (1.8) may develop discontinuities as t grows larger, depending on the configuration for u and the initial level set field $\varphi(\mathbf{x}, 0)$ (and thus the initial shape γ corresponding to the desired iso-contour of φ at time $t = 0$. See e.g.[195, 252]. This non-differentiability is intrinsically connected to the notion of *weak solutions*. Since Level Set Method is a mature scheme from the numerical point of view, the modelist has a wide set of numerical techniques which naturally account for this non-differentiability in the construction of accurate and efficient approximation schemes and do not exclude physically correct non-smooth solutions [195, 252].

1.5.1 Breaking the differentiability: a key example

We shall consider a well known example from the literature, in order to show how a Level Set Method behaves when differentiability is lost. Let us consider the periodic initial cosine curve

$$\gamma(t = 0) = (1 - s, [1 + cos(2\pi s)/2]) \tag{1.21}$$

and propagate it in a Level Set Method formulation with $u_f = 1$ and $\mathbf{v} = \mathbf{0}$. In this rather fortunate case, we may construct the solution at time t *analytically* in an explicit form. This is a common test case for the sake of illustrating the potentialities and the theoretical foundations (as well as the robustness) of Level Set Method for front tracking.

With a little amount of computation the reader can get to this expression for the parametrization of the front at time t and with parametrization parameter term s:

$$x(s,t) = \frac{y_s(s, t = 0)}{(x_s^2(s, t = 0) + y_s^2(s, t = 0))^{1/2}} t + x(s, t = 0) \tag{1.22}$$

$$y(s,t) = \frac{-x_s(s, t = 0)}{(x_s^2(s, t = 0) + y_s^2(s, t = 0))^{1/2}} t + y(s, t = 0) \tag{1.23}$$

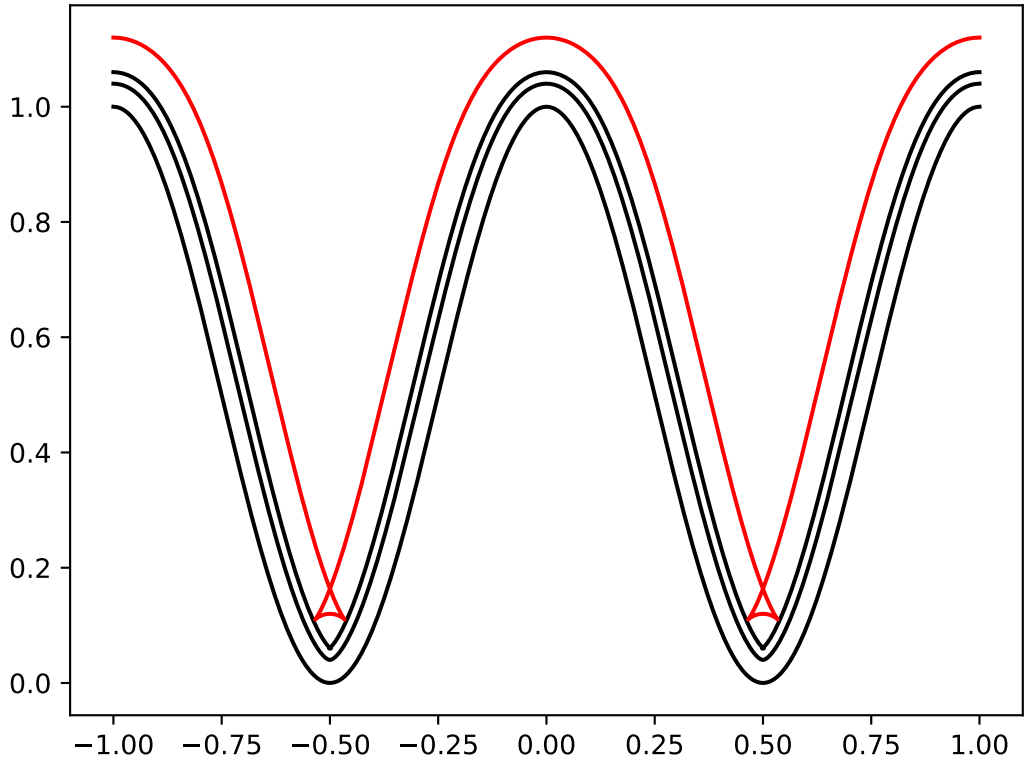
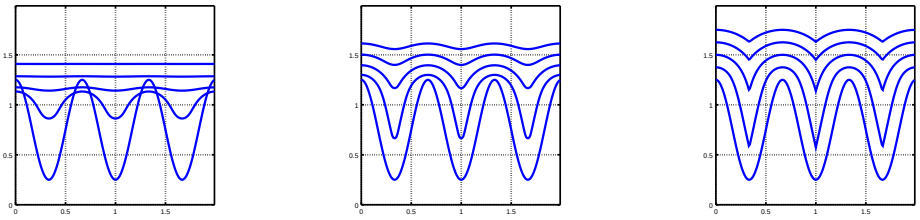


Figure 1.1: Four different time steps (from lowest to highest plot, respectively) of the motion described by Equation 1.22. The last time step (red plot) corresponds to a time where the swallowtail structure is already formed.

As can be seen in Figure 1.1, the front develops a sharp corner in finite time (exhibited from the red contour). Once this corner develops, the normal is ambiguously defined, and it is not clear how to continue the evolution. Hence, beyond the formation of the discontinuity in the derivative, we will need a so called *weak solution*, i.e. a solution that satisfies only weakly the definition of differentiability, see e.g. [81]. How can a solution be continued beyond the formation of a singularity in the curvature corresponding to a corner in the front? A reasonable answer depends on the nature of the interface under discussion. If the interface is viewed as a geometric curve evolving under the prescribed speed function, then one possible weak solution is the "swallowtail" solution formed by letting the front pass through itself, that is leaving everything as it is in Fig. 1.1. This solution is in fact the one given by equations (1.22); the lack of differentiability at the point where the swallowtails crosses itself does not cripple the solution, since the latter is written only in terms of the initial data. However, this solution after the developing of sharp cusps is not anymore a graph in the strict sense. Furthermore, if we suppose the moving curve is to be regarded as an interface separating two regions (say, burnt/burning surface of a forest fire and unburt part of the forest), we cannot accept an output such as Fig. 1.1. From a geometrical argument, if we set $u_f = 1$, the the front at time t should consist of only the set of all points located a distance t from the initial curve. In other words, the solution is developed by imagining wave fronts emanating with unit speed from each point of the boundary data and the envelope of these wave fronts always corresponds to the "first arrivals". A careful adherence to this statement produces the so called *Huygens' principle* construction. This will mantain the solution of (1.8) a graph for all times. Several theoretical arguments are described in literature, such as the Entropy Solution.The first mention to Entropy Solution for the LSM was posed by Sethian in [251, 249]. If we imagine the boundary curve as a source for a propagating fiame, then the expanding flame satisfies the requirement that once a point in the domain is ignited by the expanding front, it stays burnt. This construction is compatible with the Huygens' principle one, as shown as in [252]. We should remark what is the link between the "entropy condition" and the classical notion of entropy. An intuitive explanation is that an entropy condition stipulates that no new information can be created during the evolution of the problem. Once an entropy condition is invoked, some information about the initial data is lost, because different previous configuration could lead to the same result. Indeed, the entropy condition used in the context of LSM states that once a particle is burnt, it remains burnt, that is, once a corner has developed, the solution is no longer reversible. The problem cannot be run backwards in time; we will not retrieve the initial data in any case. Thus, some information about the solution is definitely discarded. Figure 4.2 shows this alternate weak solution, with no "swallow-tail" formation and a clear distinction between the two areas the front is dividing; in a combustion setting, a we shall say a clear distinction between the burnt domain and the unburnt one.



(a) (b) (c)

Figure 1.2: Simulations performed by integrating the LSM with Ian Mitchell's toolbox [181]. Each panel shows 5 time steps of a simulation with a constant normal velocity of $1ms^{-1}$ which pushes the contours upwards with time. The simulated window is $t \in [0, 0.5]$. The parameter λ of Equation 1.14 is decreasing for each panel, respectively 1 in Figure 4.2, 0.5 in 4.2 and 0.05 in 4.2

1.5.2 Link with Conservation Laws

Let us imagine that we are so fortunate that we can put the initial front as a graph of a function $f(x)$. That is, $\Gamma = (x, f(x)$, with f periodic on $[0, 1]$. Let us also suppose that the propagating interface remains a function for all the time. Let γ be the height of the propagating function at time t, that is $\gamma(x, 0) = f(x)$. In the level set formalism we have the following equation of motion, given by Equation (1.14):

$$\gamma_t = u_f \mathbf{n} \sqrt{(1 + \gamma_x^2)} \tag{1.24}$$

Using $u_f = \mathcal{V}_0 * (1 - \lambda_M \kappa)$ and setting $\mathcal{V}_0 = 1$, we get to

$$\gamma_t - \sqrt{(1 + \gamma_x^2)} = \lambda \frac{\gamma_{xx}}{1 + \gamma_x^2} \tag{1.25}$$

Substituting in the last equation the slope $c = \gamma_x$, we obtain

$$c_t + \left[-\sqrt{(1 + c^2)} \right]_x = \lambda \left[\frac{c_x}{1 + c^2} \right]_x \tag{1.26}$$

In the setting of hyperbolic conservation laws, we have in the last equation an equivalent of the well known Burgers Equation with vanishing viscosity (viscous hyperbolic conservation law). From this example it is evident the role of the curvature in switching from the hyperbolic setting ($\lambda = 0$, possibilities of cusps in the graph given by $\gamma(x, t)$ given by shocks in the Equation (1.26) that regulates its slope c) to parabolic setting ($\lambda > 0$).

From the work of J.A. Sethian [249] we have the following Theorem, here readapted in the context of the equation written so far:

For the constant speed case of Equation (1.8) *setting* $u_f = 1$ *, the entropy solution which respects the Huygens' principle is given by the limit* $\lambda_M \to 0$ *of* (1.14) *, with* $\mathcal{V}_0 = 1$

In Figure 1.2 some simulations with Ian Mitchell's toolbox [181] in `OCTAVE` programming environment [74]. For the same normal velocity, the multiplier for the curvature dependent term is lowered from 1 to 0.05 from the leftmost panel to the rightmost one. The simulation is performed from time $t = 0$ to time $t = 0.5$ and 5 equally spaced time steps are reported by plotting iso-contours corresponding to $\varphi = 0$ Such contours are represented by blue lines, advected by the normal velocity that pushes the front upwards. It is clear that a higher multiplier for the curvature term smears out the geometrical inhomogeneities and the front tends to a planar contour pushed vertically by a constant speed. By lowering the multiplier for the curvature term λ, we have that the profile tends to the one devised by the entropy solution of the level set equation (vanishing viscosity).

To summarize the discussion so far concerning the theoretical foundation of LSM:

- A front propagating at a constant speed may form corners as it evolves, since it is linked to Hamilton Jacobi equations. At such points, the front is no longer differentiable and a weak solution has to be constructed to continue the solution retaining physical meaning.
- The correct weak solution comes by means of an entropy condition. The choice of weak solution given by our entropy condition rests on the perspective that the front is an evolving interface separating two regions, and the assumption that one is interested in tracking the progress of one region into the other (as in the combustion setting).
- A front propagating at a normal speed $1 - \lambda_M \kappa$ for $\lambda_K > 0$ does not form corners and stays smooth for all time; Mathematically speaking, it is a parabolic equation not anymore prone to shocks. Furthermore, as the dependence on curvature vanishes ($\lim_{\lambda_M \to 0}$) , the limit of this motion is the entropy-satisfying solution obtained for the constant speed case.
- In the case that the propagating curve remains a graph during its evolution, there is a direct connection between the equation of motion and a one-dimensional hyperbolic conservation law for the front slope. The role of curvature in a propagating front is analogous to the role of viscosity in this hyperbolic conservation law (from this the use of the term "viscosity solution").
- By embedding the motion of a curve as the zero level set of a higher dimensional function, an initial value partial differential equation can be obtained whose treatment can be way simpler in many cases.

1.6 Lagrangian Methods

Purely Lagrangian methods for modeling moving interfaces consist of seeding the interface with marker particles and then moving them according to a the velocity field accounting for the underlying physics. These methods have proved to be capable of reaching high level of accuracy in literature [289, 66, 102].

However, as also pointed out in [102] they share three drawbacks:

- It is difficult to simulate breakup and merging processes of the surfaces. Therefore, topology changes are quite hard to deal with. See in this respect Figure 1.3 and Figure 1.4, where

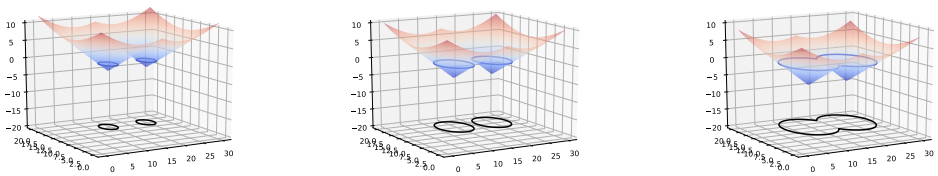


Figure 1.3: The three figures show the merging of two (initially separate) radially expanding fronts in 2 dimensions, according to the formalism of LSM. They correspond to three separate time steps, from the initial setting (left) to the last time step of the simulation (right panel).$\varphi(\boldsymbol{x}, t)$ is here represented as a signed distance function, and plotted as a colored surface assuming values in the z axis. The contour of the front, namely the 0 iso-contour of $\varphi(\boldsymbol{x}, t)$, is projected in the plane $z = -18$ for clarity purposes.

two fronts are merging in LSM formalism and a Lagrangian formalism, respectively. It is clear that when LSM includes merging of fronts in a straightforward way, some other front detection algorithm is needed for a Lagrangian scheme.

- It is not straightfoward to keep particle density consistent with the desired level of discretization. They tipically accumulate at some areas while leaving other areas few particles. Effective particle creation and deletion strategies are needed to avoid such situation.

- Most purely Lagrangian methods maintain the connectivity of the particles see e.g. [292]. Such connectivity is needed to reconstruct the free surface from the scattered particles that are given as a result of the numerical lagrangian approach. This connectivity defines a mesh of the moving interface. Unfortunately, the latter mesh may get too distorted by the action of the underlying velocity field, corrupting the reconstruction process, and ultimately leading to collapse of the simulation.

1.6.1 Lagrangian front tracking: DEVS algorithm

While this Thesis will focus on the application of LSM for the drifting part of the model described in Chapter 3, the studied model modularity allowed for its implementation in a Lagrangian setting in a study concerning wild-land fire applications [126]. In that setting a specific Lagrangian scheme for front tracking has been adopted, the so-called DEVS [316], that was included in the operational code for wild-land fire simulation ForeFire [88, 85].

DEVS handles the time advancement in terms of the increment of physical quantities instead of a discrete time step. The resulting front is a polygon (as a common output for Lagrangian algorithms) whose marker points have real (i.e.. non-discrete) coordinates instead of being located on nodes of a regular mesh or grid. This computationally in-expensive Lagrangian technique simulates the

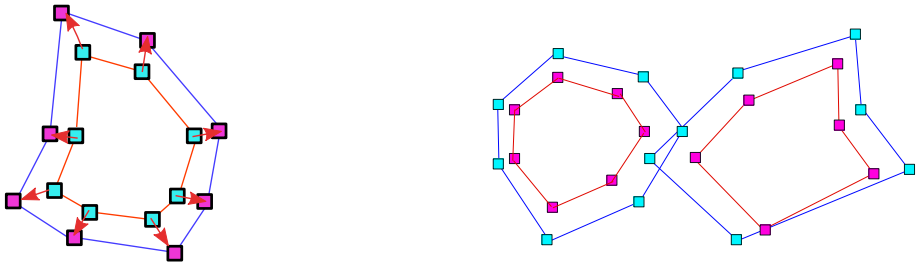


Figure 1.4: Left panel: interface represented in a Lagrangian way at two different timesteps: t_0, with light blue markers and red contour, and t_1, with fuchsia markers and blue contour. The interfaces evolves trough a prescribed velocity field represented buy the arrows. Right panel: the merging of the fronts without any ad-hoc algorithm that accounts for the change in topology of the front, showing thus the problem addressed in [102]. Front at $t = t_0$ is represented by two fronts of magenta markers with red contour, while front at $t = t_1$ is represented by two entwined fronts of blue markers with blue contour, while the correct physical result would rather be a unique front.

evolution of an interface *without making use of any underlying grid* representing the state of the system. Note that in the case of Eulerian front advancing schemes such as LSM, we had to rely on a mesh where to discretize the PDEs. Each contour is discretised into a set of markers. Each marker is connected to the next one through a piecewise linear segment.

Each marker is associated with a propagation speed and direction; and each time a marker moves, the intersection with the neighborhood is checked to take care of collision and topological changes. In order for the markers to be advected at each time-step, a velocity field $\mathbf{V}(\mathbf{x}, t)$ need to be defined The propagation speed of the marker is defined by the ROS, while the propagation direction is defined by a front normal function. While spline interpolation may be used to estimate this normal, the bisector angle made by the marker with its immediate left and right neighbors is used as a computationally effective approximation of this outward normal.

In DEVS, time is treated as a continuous parameter and each marker evolves according to its own independent time step. The time advancement of the markers is event based and all markers do not share the same time step. Due to different time advances for each marker, the CFL condition applies only locally and the markers move asynchronously. All events, triggering marker movement, are time sorted in an event list and processed by a scheduler. This self adjusting temporal resolution in Lagrangian formulation is computationally efficient and provides an efficient way to simulate the spatially in-homogeneous problem. The simulation in DEVS advances as new events are generated; the generation of new events is managed by the following two criteria:

Collision criterion: A collision happens when a marker moves into a different area, e.g., from an unburned to a burned area or from an active fire to a fuel break. Each collision generates a modification of the shape.

Quantum distance criterion: Quantum distance Δq is defined as the maximum distance each marker is allowed to cover during advancement. The actual resolution of the simulation is limited by this parameter and details smaller than the quantum distance Δq may not be accounted for.

Three type of events can be defined to control the front propagation: *decomposition*, *regeneration* and *coalescence*. The *decomposition* function is activated when a marker enters a different area (e.g., while approaching a fuel break zone). As soon as the marker enters a new area, two new markers are created on the boundary of the new area. In *regeneration*, the markers are redistributed to refine the shape; if two markers are separated by a distance greater than the perimeter resolution Δc, a new marker is generated between the two markers. The *coalescence* function reconstructs the fire-perimeter by merging the markers. All markers with separation less than Δr are merged together. For stability and to avoid cross over of two markers, $\Delta r = \Delta c/2$ is assumed and the condition $\Delta c \geq 2\Delta q$ should be respected. The precision of the method is highly dependent on the choice of Δq and Δc. Quantum distance Δq, should be of a much higher resolution than the wind data for minimal error. A detailed description of the DEVS front tracking method can be found in Ref. [87].

An example of the output of a ForeFire run is available in Section 5.6.2 of Chapter 5.

Data is the sword of the 21st century,
those who wield it the samurai.

Jonathan Rosenberg

Chapter 2

Uncertainty Quantification and Sensitivity Analysis

This Chapter contains:

2.1 Introduction

This chapter introduces the reader to the key concepts of Uncertainty and Sensitivity Analysis, with the aim of covering the main ideas and techniques adopted.

The main keywords of the chapter are here resumed below. All the models presented in this thesis can be thought as a black-box, connecting the uncertain inputs to some observable, that is some quantity of interest that can be measured with ease and results from the integration of the model with selected input parameters. Uncertainty Quantification on the outputs (treated in Section 2.2) and Sensitivity Analysis on the model parameters (treated in Section 2.3) are then useful techniques, mandatory to fully understand the behavior of any computational model. Non intrusive methods

are adopted to avoid modifying the forward model and call it as a black box to carry out UQ and SA. Stochastic approach to run the model multiple times to cover the space of uncertainties referred to as the *hypercube*: this method is also called *ensemble approach*. In order to compute the statistics for UQ and SA, the practitioner could be either adopt Monte Carlo random sampling approach, or use a surrogate of the physical model to avoid the high numerical cost of the former method. Surrogate models are treated in Section 2.4.

Computer models, or simulators, are nowadays widespread over every field of science. They allow us to study a (possibly complex) physical model without the need to perform experiments (be they computer models or real experimental settings) which may prove to be very expensive, if not impossible, in terms of time and resources. Just to name a few examples, CAE (Computer Aided Engineering) that may adopt Finite Element Modeling to compute mechanical and thermal stresses over complex structures, or a Computational Fluid Dynamics simulation of the wind around an aircraft wing. Most simulations related to real world problems involve the discretization of a set of Partial Differential Equations via some numerical scheme (Finite Differences, Finite Volumes, Finite Elements...) or Ordinary Differential Equation if the spatial variability is not object of study. Nonetheless, some of them do not depend explicitly on such a set of coupled Differential Equations, like Pott's models or Cellular Automatas.

In this work, the simulated models are obviously the fire spread model of Chapter 5 and the front propagation model for complex media of chapter 3, which rely on WENO finite-difference schemes to discretize the Hamilton Jacobi Level Set Equation, and to numerical kernel applied to their solution to represent underlying random effects. In both cases, as stated before, experiments may be expensive, not practical, or even dangerous (if involved e.g. some kind of wild-land fire run amok), or even impossible to be projected at all. Computer models can be therefore a much more viable and safer alternative.

Simulators are usually made of two parts. The first one is the so called Mathematical Model, which depends on the aforementioned set of PDEs or ODEs and makes an effort to describe the analyzed process totally or partially, according to the state of theoretical knowledge reached by the scientific field involved by the modelled process. There may be a tradeoff between the complexity of such mathematical model and the level of accuracy the modelist expects from the latter. The output it gives may reflect spatial or temporal trends, that should be refined in most cases to give answers to the modelist's questions. The mathematical models written for practical purposes involves typically large sets of inputs and outputs, that may represent either real world quantities or a parametrization of them.

The second part of a simulator is its numerical implementation in a computational architecture.

In the field of applied mathematical modeling, most models cannot be resolved analytically, being thus the coding of algorithms into computer programs the only way to achieve a result. Such programs can consists of thousand of low level and/or high level computer code lines, and to perform a simulation of the mathematical model under consideration expensive architectures and huge amount of time has to be employed, due to the complexity of most useful models.

Since simulators typically depend on a large set of parameters, they may be treated as "black-box" models, that is considering them just a way to connect inputs to outputs, neglecting their inner mechanics. Even in the case where all the mathematical features are known analytically, its complexity may be so severe that considering the model other than a "black-box" would be challenging. The input is usually made of a vector of input parameters $\boldsymbol{\theta}$ (physical parameters, numerical parameters, initial conditions, external forcing...), while the output is usually coded into a set of Quantities of Interest (QoI), $\mathbf{y}$.

Such quantities of interests may be scalars, vectors or fields and ad-hoc techniques are adopted to compress an entire set of spatial and temporal fields, typical output of a computer simulation, into such vector of observables [168], [16]. Some of the main techniques adopted for data compression of QoIs are spatial aggregation, modal decomposition (POD Proper Orthogonal Decomposition, DMD Dynamic Mode Decomposition, SINDY Sparse identification of nonlinear dynamics).

The model equations (e.g. system of partial differential equations) can be written in a compact form using the operator $\mathcal{M}$ mapping the input parameters $\boldsymbol{\theta}$ onto the set of quantities of interest $\boldsymbol{y}$ such as

$$\mathcal{M}(\boldsymbol{\theta}) = \mathbf{y}, \tag{2.1}$$

being $\mathcal{M}$ the considered black-box function, $\boldsymbol{\theta} \in \mathbb{R}^d$ the input parameter vector, $\mathbf{y} \in \mathbb{R}^n$ the vector of quantities of interests, $d, n \in \mathbb{N}$.

2.2 Uncertainty Quantification

Each mathematical model computed with the aid of a numerical scheme is affected by several layers of uncertainty. The field of Uncertainty Quantification (UQ) aims at quantifying with the best accuracy possible such sources of uncertainty. This is of particular relevance bearing in mind that the output of a computational model may be an important part of a decision making process.

Kennedy and O'Hagan, in Ref. [127], formulated an exhaustive list of the main sources of uncertainty in computer experiments:

- *Parameter (input) Uncertainty:* The model is run following a set of physical and modeling parameters, often subject to epistemic uncertainty. Some of them, like e.g. the gravitational

acceleration, may be declared fixed being thus subject to no appreciable fluctuation. Nonetheless, the rest of the input set can be considered as random quantities (even if sometimes just a rough estimate of their support - lower and upper bound - can be given).

- With the term *Model Discrepancy* Kennedy and O'Hagan address the fact that even when setting to zero the uncertainties on the parameters, hardly ever any useful computational model will reproduce the physical phenomena it intends to represent. We should then pay attention since this epistemic uncertainty lies under such Model Discrepancy. In the framework adopted in this thesis, this corresponds to the representation error of the operator $\mathcal{M}$ (*model error*). In the field of Data Assimilation it is also called *representativeness error*.

- *Residual Variability* stands for the fact that either true (experimental) realizations of the observed phenomena may have a stochastic nature not grasped by the model and its computational implementation, or the real world phenomena for each experimental realization varies according to a set of parameters larger than the ones used by the model. In other words, the phenomenon to be modeled is either too unpredictable or too complex to have a computational model adopting a limited set of parameters capable of mimicking perfectly every realization.

- Another layer of uncertainty (strongly connected to the previous one) is represented by the so called *Parametric variability*, which accounts for the fact that since the inputs cannot be specified properly due to lack of information, and the output depends on uncontrolled conditions.

  It is also known as *internal variability* linked to the intrinsic variability of the system. As an example, some systems exhibit space and time variability of the parameters. This kind of uncertainty is also referred to as aleatory uncertainty.

- If the model is being calibrated against some data coming from real world experiments, the measurements of such physical process are bound to be affected in some way by the so called *Observational error*, which constitutes a source of uncertainty, whose magnitude varies from phenomenon to phenomenon.

- With *Code Uncertainty* Kennedy and O'Hagan express the fact that not every configuration can be tested by the code, since it would be not achievable at all in acceptable times. Some kind of interpolation between selected code runs should be employed (like the Surrogate Models that are described in Section 2.4).

In the end, we can classify the sources of uncertainty into two categories: *epistemic uncertainty* due to the modeling choices and the available inputs that can be theoretically reduced if we have more information versus the *aleatory uncertainty* due to the inner variability of the physical processes at play that cannot be reduced, but does need to be quantified.

In this work, we focus on the Parameter Uncertainty, i.e. in the epistemic uncertainty in the vector $\boldsymbol{\theta}$ composed by input parameters of a given model $\mathcal{M}$. In this framework, $\boldsymbol{\theta}$ is represented as a random vector.

Even if Sensitivity Analysis has been historically considered a part of UQ, in this work we neatly divide the two subjects.

SA is a way to analyze the connection between the input parameters and the quantities of interest, and more specifically to analyze how the variability in the inputs affect the variability in the QoI. SA is treated in detail in Section 2.3.

On the other hand, UQ is a way to quantify the uncertainty in the quantities of interest, to characterize the associated PDF or its statistical moments. Here, the PDF of the QoI will directly depend on the PDF of the input parameters subject to uncertainties.

The main purpose of UQ as presented here is then the Uncertainty Propagation: the investigation on the way the uncertainty on input parameters propagates (possibly in a non-linear way) through the computational model $\mathcal{M}$ and affect some quantities of interest (QoI). Such QoIs, due to the randomness/uncertainty on the inputs, may be treated in a statistical framework, studying their probability density function and statistical moments (mean, variance...) as well as obtaining the probability of exceeding certain thresholds. Another problem of paramount relevance addressed by UQ is the *Inverse Problem*: help the simulator to match reality, for instance connecting some range of the quantity of interest of the real world phenomenon to to some specific subsets of the input parameter space, thus honing the computational model in its capabilities.

Techniques for propagating uncertainties can generally be classified as intrusive or non-intrusive, see Reference [69]. Intrusive UQ methods need a reformulation of the governing equations of the mathematical models describing the phenomena of interest [51]. Non-intrusive UQ methods, on the other hand, uses ensembles of model runs. The elements of such simulation ensemble are retrieved by sampling the uncertain inputs according to various sampling schemes. If we adopt a probabilistic interpretation of UQ, we are exploring the support of the PDF of each parameter. The impact of the input uncertainties can then be analyzed for the selected model output quantities of interest adopting some kind of statistical strategy. In this thesis we shall work exclusively with non-intrusive ensemble methods.

Design of Experiments In the framework of ensemble methods, a *design of experiments* (also known as *experimental* design) refers to the way of discretizing the uncertainty (or "hypercube") $Z_\Theta \in \mathbb{R}^d$, in which the parameters vary. It is a way to encode the N realizations of parameters $\boldsymbol{\theta}$, for which the mathematical model (the Forward Model) is integrated as a "black-box" to obtain the ensemble of N functional outputs $\mathbf{y}$ from which statistics can be performed. In practice, this

ensemble forms a database $\mathcal{D}_N$:

$$\mathcal{D}_N = \left\{ \left(\boldsymbol{\theta}^{(l)}, \mathbf{y}^{(l)} \right)_{1 \leq l \leq N} \right\}, \tag{2.2}$$

where $\mathbf{y}^{(l)} = \mathcal{M}\left(\boldsymbol{\theta}^{(l)}\right)$ stands for the integration of numerical model $\mathcal{M}$ associated with the lth set of input parameters $\boldsymbol{\theta}^{(l)}$.

Some of the classical methods to build these experimental designs are quasi-random low discrepancy sequences [107, 82], good lattice points [23], Latin Hypercubes [42], orthogonal Latin Hypercube [17], that have generally proven to behave well for spaces of dimensionality $d \leq 10$.

As we shall see in the further sections, the mathematical formulation of DoE given above will be the key point to understand how Uncertainty Quantification and Sensitivity Analysis are linked together.

2.3 Sensitivity Analysis

As stated in [218], Sensitivity Analysis (SA) investigates how the variability of the output of a computational model can be attributed to the variation of the input parameters in the space where they are defined. It is a very vast subject of Applied Mathematics and Statistics, historically considered entwined with UQ. The use of SA techniques is becoming year after year a consolidated practice, being considered nowadays a fundamental tool in the construction, use and understanding of mathematical models of all sorts [274] and there is thus a vast literature about the application of SA techniques. To name just a few examples of application area, in [290] local sensitivity analysis is used in Combustion Chemistry, where models of homogeneous hydrogen explosion and of premixed laminar hydrogen-air flame are studied and the kinetic parameters ranked via SA; *Confalonieri et al.* in [58] applied several routines of SA for the sake of ranking the parameters of an ecological model for the prediction of rice model, WARM; as a last example, in nuclear engineering simulations, *Auder et al.* in [16] studied the sensitivity to inputs for a nuclear reactor application (coping with uncertainties in the pressurized thermal shock analysis).

2.3.1 Purposes of a SA procedure

Saltelli et al. in Ref. [243] referred to the main purposes of SA as the three *Settings* of a SA procedure.

They are reported below:

1. *Ranking - (Factor Priorisation)*: it is the objective of ordering the input parameters by their respective relevance, i.e. their impact on the vector of QoI $\mathbf{y}$;

2. *Screening - (Factor Fixing)*: it consists on the identification of the input parameters (factors) that share a negligible effect on the vector of QoI $\mathbf{y}$; those parameters may be set to their

nominal values and no more modeling effort, experimental characterization or sampling should be performed, hence simplifying the analysis of the computational or mathematical model (dimensionality reduction).

3. *Mapping*: it is the most ambitious objective of a SA procedure, and it is closely related to the *inversion* of the forward model $\mathcal{M}$. It consists on the determination of the region of input parameters $\mathcal{Z}$ that produces a-priori selected effects on the output vector $\mathbf{y}$. More formally, it aims at defining the input set

$$\mathcal{Z}^* \subset \mathcal{Z} \quad s.t. \quad f_m(\theta) > 0 \quad \forall \theta \in \mathcal{Z}^*,$$

for a selected functional f_m.

2.3.2 Types of Sensitivity Analysis

Since the SA is a rather vast area of applied Mathematics and Statistics, pursuing different objectives, the algorithms and methodologies evolved according to different guidelines. There is a extensive literature about review and comparison between different SA approaches (to name a few examples, the review of *Iooss and Lematre* in [119] and the review of *Pianosi et al.* [218], the latter with a focus on models for environmental predictions).

In this section a brief description of the main types of SA techniques currently available for the practitioners are listed.

Three important subdivisions of SA

Before we start listing the main techniques for SA, a main subdivision of the features that a SA analysis may exhibit is the following:

- *Local or Global SA*: Local sensitivity analysis considers the output variability against variations of the vector of input factors θ in the neighborhood of a certain fixed value θ_0. Examples of Local SA are the derivative methods based on Taylor decomposition, described with more detail in the next paragraph. On the contrary, *Global Sensitivity Analysis (or GSA)* considers variations within the entire space of variability of the input factors,. The application of local SA obviously requires the user to specify a nominal value θ_0 for the input factors. While GSA overcomes this possible limitation, it still requires specifying the input variability space $Z_\Theta \in \mathbb{R}^d$ introduced in Section 2.2, in which the parameters may vary. If the latter space of feasibility of the parameters is poorly known, and so built by the means of more or less severely rough estimates, the results from GSA should be handled with care.

- *One at time SA or All at Time SA*: In *sampling-based* SA, that is a SA analysis that involves the construction of a Design of Experiment (DoE) the sampling may be *One At Time (OAT)*

or All At Time (AAT). In OAT sampling, the input factors of the simulation model θ_i will be perturbed from their nominal values one at a time while keeping all others fixed. In AAT sampling instead, the variability of the output QoI are induced by varying all the input factors at the same time. This allows the sensitivity to each factor to account for the direct influence of that factor as well as the joint influence with the other factors due to interactions.

- *Quantitative or Qualitative SA* The term *quantitative SA* refers to methods where each input factor θ_i is associated with a quantitative and reproducible evaluation of its relative influence. In practice this is achieved through a set of sensitivity indices (or *importance measures*). An example of them are the *Sobol' Indices* described in the following. In qualitative SA, instead, the sensitivity of the QoI on the variability of inputs is assessed in a qualitative way, visualizing plots of model outputs or via the aid of specific visualization tools like, for instance, scatter plots (e.g. [27]), tornado plots (see e.g. [115]), visualization of the posterior distributions of the input parameters (e.g. [98]). Qualitative SA tools can be used to help visualize in a more complete fashion results carried out by a Quantitative SA.

Perturbation - derivatives methods and multiple-starts perturbation methods

The most basic technique for SA varies the input parameters (input factors) of the simulation model from their nominal values (that is the vector θ_0 one $\theta_i \in \theta$ at a time (OAT) and assesses the effects on the simulation results in a qualitative (e.g. via visual inspection) or quantitative way. If a scalar or vectorial QoI $\mathbf{y}$ can be defined, a quantitative approach is then to measure the output sensitivity of the desired QoI with respect to the θ_i , i-th input parameter via partial derivative evaluated at $\theta = \theta_0$ or via finite-difference techniques if the QoI cannot be expressed trough analytic functions (as it is in the vast majority of the cases). The analytic measure of relevance of y output with respect to $i-th$ input θ_i is then in analytic form

$$S_i(\theta_0) = c_a \frac{\partial y(\theta)}{\partial \theta_i} \tag{2.3}$$

while in finite-difference form it reads

$$\widehat{S}_i(\theta_0) = c_{fd} \frac{y(\theta_1, \dots \theta_i + \delta_i, \dots, \theta_d) - y(\theta_1, \dots \theta_i, \dots, \theta_d)}{\delta_i} \tag{2.4}$$

with $c_{\mathrm{a}}, c_{\mathrm{fd}}$ opportune scaling/normalizing factors.

These so called Derivative-based SA methods are motivated theoretically by the means of the Taylor series expansion of the QoI $y_i \in \mathbf{y}$, see e.g. [112, 38]. When no analytical formula is available, the computation of the sensitivity measures for d input parameters requires $(d+1)$ model evaluations. Derivative-based sensitivity measures are therefore computationally very cheap, but their range of applicability is somewhat crippled by the fact that they explore solely a local neighborhood of $\boldsymbol{\theta}_0$, providing only information about local sensitivity. Second derivatives (with particular focus on the

mixed ones), which may be estimated with a relatively small number of additional model evaluations, can provide information about local interactions between input parameters.

A global extension of the perturbation approach is to compute output perturbations from multiple points θ_k within the feasible input space Z, in a *Design-Of-Experiment* fashion. A global sensitivity measure will come by aggregating these individual sensitivities. There are several methods belonging to this category, differing one from each other according to the way they tackle the following aspects: (i) whether they use finite differences directly, or they apply a transformation, such as norms or squared values; (ii) how they select the fixed points of the DoE; (iii) how the individual sensitivities are aggregated.

The most established method of this type is the method of Morris [185], also called the Elementary Effect Test (EET). Here, the mean of r finite differences (also called Elementary Effects or EEs) is taken as a measure of global sensitivity:

$$S_i = \mu_{\text{Morris}} = \frac{1}{n}\sum_{j=1}^{n} EE^j = \frac{1}{n}\sum_{j=1}^{n} \frac{y(\theta_1^j, \dots \theta_i^j + \delta_i^j, \dots, \theta_d^j) - y(\theta_1^j, \dots \theta_i^j, \dots, \theta_d^j)}{\delta_i^j} c_i \tag{2.5}$$

where c_i stands for a normalization coefficient or a weight and the apex j stands for the $j-th$ "finite difference gradient" explored in the feasibility space $Z_\Theta \in \mathbb{R}^d$. An alternative measure proposed by Campolongo and Saltelli [46] consists of taking the absolute value of the finite differences to avoid that differences of different signs would cancel out,

$$S_i = \mu^*_{\text{Morris}} = \frac{1}{n}\sum_{j=1}^{n} EE^j = \frac{1}{n}\sum_{j=1}^{n} \left| \frac{y(\theta_1^j, \dots \theta_i^j + \delta_i^j, \dots, \theta_d^j) - y(\theta_1^j, \dots \theta_i^j, \dots, \theta_d^j)}{\delta_i^j} \right| c_i \tag{2.6}$$

Besides the above sensitivity measure, it is common practice to also compute the standard deviation of the EEs, which provides information on the degree of interaction of the i-th input factor with the others, and on the non-linearity of the forward model $\mathcal{M}$. A high standard deviation indicates that a factor is interacting with others because its sensitivity changes across the variability space due to the different values assumed buy the other θ_is.

Importance Measures:Methods based on the analysis of linear models

In a DoE based setting for SA, a sample of inputs and outputs $\mathcal{D}_N = \left\{ \left(\boldsymbol{\theta}^{(l)}, \mathbf{y}^{(l)}\right)_{1 \le l \le N} \right\}$, is available. it is possible to fit a linear model explaining the behaviour of $\mathbf{y}$g iven the values of θ, provided that the sample size N is sufficiently large. Adopting such linear models, several indices of importance measure to rank variable and quantify their inpacts on the outputs are given in the following.

- Pearson correlation coefficient:

$$\rho(\theta_j, y) = \frac{\sum_{i=1}^{N} (\theta_j^{(i)} - \mathbb{E}(\theta_j))(Y_i - \mathbb{E}(y))}{\sqrt{\sum_{i=1}^{N} \left(\theta_j^{(i)} - \mathbb{E}(\theta_j)\right)^2} \sqrt{\sum_{i=1}^{N} \left(Y_i - \mathbb{E}(Y)\right)^2}} \tag{2.7}$$

It may be seen as a linearity measure between the input parameter θ_i and the output y (here to be interpreted it it is the case as a component of the vector of QoI $\mathbf{y}$. It equals ± 1 if the tested input variable share a linear relationship with the output. If the input variable and the output are independent one from each other, the index equals 0.

- Standard Regression Coefficient (SRC):

$$SRC_j = \beta_j \sqrt{\frac{Var(\theta_j)}{Var(y)}} \tag{2.8}$$

where β_j is the linear regression coefficient associated to θ_j. In the case when the linearity hypothesis is confirmed, the square of SRC_j is a share of the total variance.

- Partial Correlation Coefficient (PCC):

$$PCC_i = \rho(\theta_j - \widehat{\theta_{-j}}, Y - \widehat{Y_{-j}}) \tag{2.9}$$

where $\widehat{\theta_{-j}}$ is the prediction of the linear model expressing the input θ_j as a function of the other inputs and $\widehat{Y_{-j}}$ is the prediction of the linear model where θ_j is absent.

The estimation of these sensitivity indices is subject to an uncertainty estimation, due to the fact that in most application the sample has a limited size due the cost of data retrieval or CPU time for simulations. Such uncertainty can be estimated through analytical formulas ([52]).

Importance Measures: Methods based on the variance decomposition

Following *Pianosi et al.* in [218], Variance-based SA methods rely on three fundamental principles: (i) input factors are regarded as stochastic variables so that the variability on random input parameters induce a variance in the output space; (ii) the variance of the output distribution is a good indicator of output uncertainty; (iii) the contribution from a given input factor θ_i to the output variance should be considered an acceptable measure of its sensitivity.

Sobol' indices [257, 243] are a well known mathematical tool for global sensitivity analysis based on variance decomposition. They provide the quantification of how much of the total variance in the quantity of interest is due to the spread in the selected uncertain input parameter, assuming the input parameters to be independent random variables. The variance of the output random variable Y denoted by $\mathbb{V}[Y]$ can be decomposed as

$$\mathbb{V}[Y] = \sum_{i=1}^{d} \mathbb{V}_i(Y) + \sum_{j=i+1}^{d} \mathbb{V}_{ij}(Y) + \cdots + \mathbb{V}_{1,2,\ldots,d}(Y), \tag{2.10}$$

where $\mathbb{V}_i(Y) = \mathbb{V}\left[\mathbb{E}(Y|\Theta_i)\right]$, $\mathbb{V}_{ij}(Y) = \mathbb{V}\left[\mathbb{E}(Y|\Theta_i, \Theta_j)\right] - \mathbb{V}_i(Y) - \mathbb{V}_j(Y)$ and more generally,

$$\mathbb{V}_I(Y) = \mathbb{V}\left[\mathbb{E}(Y|\Theta_I)\right] - \sum_{J \subset I \text{ s.t. } J \neq I} \mathbb{V}_J(Y), \ \forall I \subset \{1, \ldots, d\} \tag{2.11}$$

Based on this variance decomposition, the first-order Sobol' index S_i associated with the ith parameter of $\mathbf{\Theta}$ reads

$$S_i = \frac{\mathbb{V}_i(Y)}{\mathbb{V}(Y)}. \tag{2.12}$$

It corresponds to the ratio of the output variance $\mathbb{V}(Y)$ that is *uniquely due* to the ith input factor; S_i ranges between 0 and 1. The corresponding total Sobol' index S_{T_i} instead measures the whole contribution of the ith input parameter (including this time interactions with other parameters of $\mathbf{\Theta}$) on the output variance. It is defined as:

$$S_{T_i} = \sum_{\substack{I \subset \{1,\dots,d\} \\ I \ni i}} S_I. \tag{2.13}$$

By definition, $S_{T_i} \geq S_i$. If both first-order and total indices are not equal, that means that the input parameter Θ_i share some interactions with other parameters of $\mathbf{\Theta}$ to explain the output variance.

Sobol in [257] proposed the following Monte Carlo algorithm for the estimation of $S_i = \frac{\mathbb{V}_i(Y)}{\mathbb{V}(Y)}$: given $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ two independent sample points, $\mathbb{V}_i(Y)$ is calculated with the formula

$$\mathbb{V}_i(Y) = \int Y(\boldsymbol{\theta}) Y(\eta_1, \dots, \theta_i, \dots \eta_d) d\boldsymbol{\theta} d\eta_1 \dots d\eta_{i-1}, d\eta_{i+1}, \dots d\eta_d - Y_0^2 \,, \tag{2.14}$$

where Y_0 is the mathematical expectation of the quantity Y.

In this case, Monte Carlo estimator for the latter formula reads

$$\mathbb{V}_i(Y) \approx \frac{1}{N} \sum_{k=1}^{N} Y(\boldsymbol{\theta}_k) Y(\eta_{1k} \dots \eta_{i-1k}, \theta_{ik}, \eta_{i+1k}, \dots \eta_{dk}) - \left[\frac{1}{N} \sum_{k=1}^{N} Y(\boldsymbol{\theta}_k) \right]^2 , \tag{2.15}$$

where $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k = \{\eta_{1k} \dots \eta_{dk}\}$ are two independent samples of realizations the d-dimensional set of random input parameters. Formulas 2.14 and 2.15 can be straightforwardly generalized for the case when $\mathbb{V}_{\mathcal{I}}(Y)$ is sought, with $\mathcal{I} \subset \{1 \dots d\}$ a set of indices.

The sample size N in 2.15 required to reach a good level of precision in estimating variance based indices can be quite high in most of the cases, this constituting a problem for simulations that are expensive in terms of computational time.

In order to overcome this issue, several techniques are currently available in literature. The methodology known as FAST (Fourier Amplitude Sensitivity Test) [117] adopts a Fourier Series expansion in order to approximate first order Sensitivity Indices, while extended-FAST (eFAST) [242] improves the latter technique to cope with the determination of total order indices.

Another technique to circumvent the size of the DoE is to develop a surrogate of the forward model, in order to make the reiterate model evaluations needed for the estimation of the desired sensitivity index a simple computation of an analytical function that comes from a training of the surrogate.

Importance Measures: other methods and final thoughts

The methods illustrated so far are nothing but a part of the full asset of algorithms and techniques available in literature to perform a SA. While the reader is referred to complete and exhaustive textbooks on the topic [243, 142] , we shall name here a few more examples that do not fold in the aforementioned categories.

Density Based Methods The limits of Variance-Based methods is that they rely solely on the variance output to rank and screen input parameters. Density - Based methods, nevertheless, focus on exploiting all the information contained in the output PDF of the forward model, rather than concentrating just on some moment of the latter.

The main concept of Density Based Methods is to measure the sensitivity to the $i-th$ parameter θ_i through the variations in the ouput PDF that are induced while such parameter is fixed to a certain value. This can be carried out by measuring some kind of divergence between the unconditional output PDF (generated by varying all parameters) and the conditional PDF obtained by fixing the considered input factor to a fixed nominal value or to a set of prescribed values. If the analyzed parameter is fixed to a set of values, some kind of statistics of the output is needed.

In formulas, we have

$$S_i = \text{stat}\left[\quad \text{divergence}\left(f_{\boldsymbol{y}}, f_{\boldsymbol{y}|\theta_i}\right)\right] \tag{2.16}$$

where $f_{\boldsymbol{y}}, f_{\boldsymbol{y}|\theta_i}$ are the unconditional and conditional output PDF, respectively, 'stat' denote some statistics and 'divergence' some divergence measure between the two PDFs.

As a measure of divergence between the two PDFs Shannon Entropy has been adopted in literature [135], while Borgonovo [37] adopted the so-called δ-sensitivity approach, measuring the area between the cumulative distribution functions (CDFs) of the two PDFs.

In Entropy based methods there is no statistic adopted on the choice of θ_i, while in δ-sensitivity approach the Expected Value operator is adopted as a statistics of a sample of values for θ_i.

A rather recent approach related to Density Based Methods is the so-called PAWN algorithm, introduced by Pianosi and Wagener [219]. The key idea that characterizes such approach is to analyze output distributions by their Cumulative Distribution Functions, which are easier to derive than PDFs, and as a measure of distance between unconditional and conditional CDFs, the Kolmogorov-Smirnov statistic is adopted.

Mapping through SA: CART and Regional Sensitivity The difficult task of mapping model output ranges and possible regions of the input hyperspace may be performed through ad-hoc algorithms, that shall be only briefly addressed in this Chapter. To begin with, a well established method is Regional Sensitivity Analysis (RSA), also known as Monte Carlo Filtering [317]. Its key

idea is the assignation of the label "behavioral" or "not behavioral" to each input parameter according on whether the associated forward model evaluation follows or not the expected pattern. One advantage of this approach is that it can be applied to any type of model output, including non-numerical ones, as long as a splitting condition can be verified, possibly also by recurring to qualitative evaluation. RSA has been widely used in applications where the model output is an objective function (i.e. a measure of the model accuracy against observations) and the splitting criterion reflects the achievement of a minimum requirement of model performance.

Another interesting technique that allows for the mapping between input and outputs of a given model (as well as ranking) is that of Classification And Regression Trees (CART), see e.g. [110]. It belongs to the group of Correlation and Regression analysis SA techniques, and can also work with non-numerical inputs and outputs.

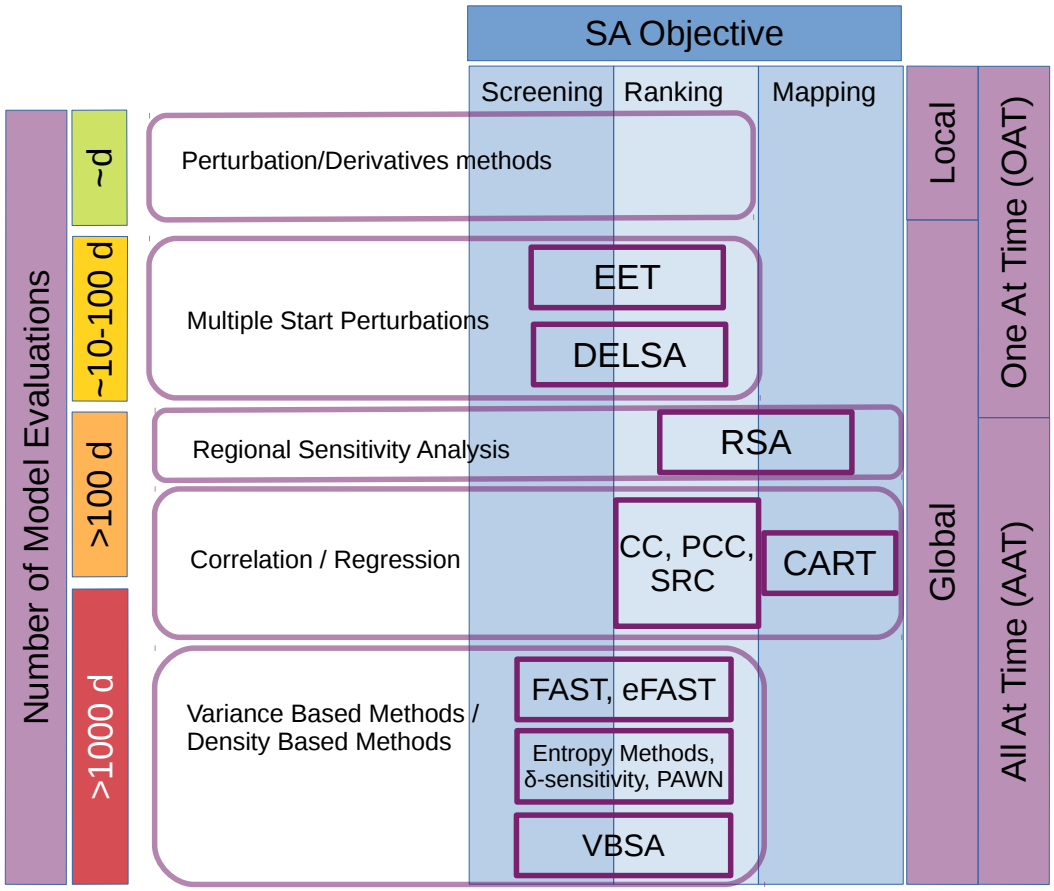


Figure 2.1: Schematic view of the different techniques for SA described in Subsection 2.3.2. In this table all the methods described in this Chapter are catalogued by the right hand side lines following their nature (local or global), the nature of their sampling (OaT or AaT). The left hand side labels inform us about the computational cost of each method, being d the cardinality of $\boldsymbol{\theta}$. The label of the columns indicate the tasks fulfilled by each method. Please note that even if they are Global Methods useful for ranking and screening, Variance Based Sensitivity Analysis techniques are very expensive in terms of sample size. This motivates the construction of a surrogate model. Adapted from [218]

2.4 Surrogate Models

A *Surrogate Model* (SM) (also known as *emulator*, or *metamodel*) is a technique employed in engineering and applied sciences when an outcome of interest (the QoI encoded in the vector $\mathbf{y}$) cannot be found straightforwardly. In the framework of this thesis, it corresponds to the solution of the forward model $\mathcal{M}$ for a given set of input parameters θ, that may also represent an engineering design problem.

In our sampling-based (DoE based) UQ and SA setting, we can see from Figure 2.1 that for an all-at-time variance-based global SA, $\approx 1000d$ model evaluation are needed to effectively rank the input parameters. This would constitute a challenge even for the evaluation of a model with acceptable computational time consumption.

Moreover, the dimension $d > 1$ adopted in this thesis is always ≤ 3, but further works could consider wider set of parameters, that could give rise to the "curse of dimensionality" as the number of samples required to cover in an efficient way the parameter space in uncertainty increases exponentially with the number of model parameters.

In general, when the forward operator $\mathcal{M}$ is a computational model, the compilation of the simulations database $\mathcal{D}$ may not be affordable, and this may indicate that numerical resolution needs to be reduced or physical processes ignored to decrease runtime and make many model runs computationally viable.

Surrogate models have the purpose to speed up complex models while retaining a reasonable level of accuracy and detail in reproducing the QoI. In practice, the forward operator $\mathcal{M}$ is then replaced by an emulator, $\hat{\mathcal{M}}$. In a metamodel approach, DoE of equation 6.29 presented in Section 2.2, $\mathcal{D}_N$ from where information for UQ and SA can be extracted, is made of metamodel execution rather than real model ones. The metamodel is trained on a smaller database of (numerically expensive) real model execution, $\mathcal{D}_M$ with $M << N$.

2.4.1 Main categories of Surrogate Models

In order to better locate the techniques adopted in this thesis, we present in the following list the main categories of Surrogate Models (see e.g. the reviews [15, 314]):

- *Data-Driven Surrogates*, where an analytical function, rather cheap to call, is constructed based on a database $\mathcal{D}$ of simulations. Data Driven Surrogates are a broad family, including Polynomial interpolation, Gaussian Processes (and the Gaussian Process related Kriging scheme), polynomial chaos expansions, Bayesian Networks, support vector machines.

  The key idea of many Data-Driven Surrogates such as Polynomial Chaos and Gaussian Process (both treated in detail in the following sections) is to replace the forward model $\mathrm{y} = \mathcal{M}(\boldsymbol{\theta})$ by

a weighted finite sum of basis functions that can be generally expressed as

$$\widehat{\mathrm{y}}\left(\boldsymbol{\theta}\right) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \gamma_{\boldsymbol{\alpha}}\, \Psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right), \tag{2.17}$$

where the coefficients $\gamma_{\boldsymbol{\alpha}}$ and the basis functions $\Psi_{\boldsymbol{\alpha}}$ are to be determined, $\mathcal{A}$ being the set of indices that defines the basis size. The coefficients and basis functions are calibrated by the training set (or "database") compiled through a DoE, $\mathcal{D}_N$. The latter corresponds thus to a limited number N of forward model integrations (or "training set") such that

$$\mathcal{D}_N = (\Theta, \mathcal{Y}) = \left\{ \left(\boldsymbol{\theta}^{(k)}, \mathrm{y}^{(k)} \right)_{1\leq k\leq N} \right\}, \tag{2.18}$$

where $\mathrm{y}^{(k)} = \mathcal{M}(\boldsymbol{\theta}^{(k)})$ corresponds to the integration of the forward model for the kth set of input parameters $\boldsymbol{\theta}^{(k)}$.

- *Projection Based Surrogates*, *also called reduced order, reduced basis models*, where the set of equations that constitute the forward model $\mathcal{M}$ are projected onto a subspace of reduced dimension, with the use of a ad-hoc basis of orthonormal vectors. The two main sub-categories are singular value decomposition (SVD) and Krylov-based methods. They include Proper Orthogonal Decomposition, Krylov Subspace Methods, and Fourier mode reduction, to name a few examples.

- *Multi-fidelity based surrogates* , also called *multi-scale, hierarchical methods* are a category of SM quite different from the previous two. In order to obtain such surrogates the underlying physics is simplified by physical reasoning, or ad-hoc techniques to extract information using different level of coarseness of the numerical grid are employed. Multi-grid methods, multi-scale finite element methods and variational multi-scale method fall in this last category.

In this Thesis we shall adopt two types of SMs, both of them belonging to the first category (Data Driven SMs): generalized Polynomial Chaos (gPC) and Gaussian Process (GP). They will be described in detail in the next sections.

While the gPC expansion retrieves the global forward model behavior, the GP regression is a local interpolator of the forward model behavior at the training points. For gPC-expansion, the user needs to determine the appropriate total polynomial order of the expansion as well as the appropriate type and number of basis functions $\Psi_{\boldsymbol{\alpha}}$. There are also different projection strategies to compute the coefficients $\gamma_{\boldsymbol{\alpha}}$. For GP regression, the user needs to choose the type of correlation structure and to estimate its associated hyperparameters.

2.4.2 Polynomial Chaos Expansion

$\boldsymbol{\theta}$, the vector of input parameters, is defined in the input physical space. We denote its counterpart in the standard probabilistic space with $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_d)$, where ζ_i the random variable associated with the ith uncertain parameter θ_i in $\boldsymbol{\theta}$ characterized by its marginal PDF ρ_{θ_i}. $\boldsymbol{\theta}$ is thus rescaled in the standard probabilistic space to which the gPC formulation applies.

Polynomial Basis

$\boldsymbol{\theta}$ is projected onto a stochastic space spanned by the orthonormal polynomial functions $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}_{\boldsymbol{\alpha}\in\mathcal{A}}$. The basis functions are orthonormal with respect to the joint PDF $\boldsymbol{\rho_\zeta}(\boldsymbol{\zeta})$, i.e.

$$\langle \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}), \Psi_{\boldsymbol{\beta}}(\boldsymbol{\zeta}) \rangle = \int_Z \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\, \Psi_{\boldsymbol{\beta}}(\boldsymbol{\zeta})\, \boldsymbol{\rho_\zeta}\, d\boldsymbol{\zeta} = \delta_{\boldsymbol{\alpha\beta}}, \tag{2.19}$$

with $\delta_{\boldsymbol{\alpha\beta}}$ the Kronecker delta-function and $Z \subseteq \mathbb{R}^d$ the space in which $\boldsymbol{\zeta}$ evolves. In practice, the orthonormal basis is built using the tensor product of one-dimensional polynomial functions, $\Psi_{\boldsymbol{\alpha}} = \phi_{\alpha_1} \dots \phi_{\alpha_d}$ with ϕ_{α_i} the one-dimensional polynomial function. The choice for the basis functions shall be formulated depending on the probability measure of the random variables. According to Askey's scheme, the Jacobi polynomials form the optimal basis for random variables following Beta-distribution,the Hermite Polynomials for Gaussian distributions, Laguerre Polynomials for Gamma distributed parameters and finally the Legendre polynomials are the counterpart for uniform distribution [313].

Assuming that the considered output of the Forward Model is of finite variance, each quantity of interest y (see Sec. 2.2) can be considered as a random variable for which there exists a gPC expansion of the form (2.20)

$$\widehat{\mathrm{y}}\,(\boldsymbol{\theta}) = \mathcal{M}_{\mathrm{pc}}(\boldsymbol{\theta}(\boldsymbol{\zeta})) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \gamma_{\boldsymbol{\alpha}}\, \Psi_{\boldsymbol{\alpha}}\,(\boldsymbol{\zeta})\,. \tag{2.20}$$

$\Psi_{\boldsymbol{\alpha}}$ is the $\boldsymbol{\alpha}$th multivariate basis function chosen in adequacy with the PDF $\boldsymbol{\rho_\theta}$ associated with the parameters $\boldsymbol{\theta}$ (we assume that all random variables in $\boldsymbol{\theta}$ are independent so that $\boldsymbol{\rho_\theta}$ is the product of the marginal PDFs $\{\rho_{\theta_i}\}_{i=1,\cdots,d}$). $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_d)$ is a multi-index in $\mathcal{A}$, which identifies the components of the multivariate polynomial $\Psi_{\boldsymbol{\alpha}}$.

Once the PDF $\boldsymbol{\rho_\theta}$ is chosen, $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathcal{A}}$ are the unknowns to be determined to build the surrogate $\mathcal{M}_{\mathrm{pc}}$.

Truncation Strategy

For computational purposes, the sum in Eq. (2.20) is truncated to a finite number of terms r that is associated with the total polynomial order P of the gPC-expansion. There are several ways of choosing the r terms. We shall refer to such ways as "truncation strategies". We will also use the concept of "enumeration functions" in the following: a linear (or hyperbolic) enumeration function

is a mapping $\mathfrak{I}$ from $\mathbb{N}$ to $\mathbb{N}^d$, which establishes a bijective mapping between a given integer i and a multi-index $\boldsymbol{\alpha}$.

Linear Truncation Strategy. The standard truncation strategy (referred to in the following as "linear") consists in retaining in the gPC-expansion all polynomials involving the d random variables of total degree less or equal to P. As a consequence, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_d) \in \{0, 1, \cdots, P\}^d$. The number of terms r is thus constrained in this linear case by the number of random variables d and by the total polynomial order P so that

$$r_{\text{lin}} = (d + P)!/(d!\ P!). \tag{2.21}$$

The set of selected multi-indices for the multi-variate polynomials $\mathcal{A}$ is defined as

$$\mathcal{A}_{\text{lin}} \equiv \mathcal{A}_{\text{lin}}(d, P) = \{\boldsymbol{\alpha} \in \mathbb{N}^d : |\boldsymbol{\alpha}| \leq P\} \subset \mathbb{N}^d, \tag{2.22}$$

where $|\boldsymbol{\alpha}| = ||\boldsymbol{\alpha}||_1 = \alpha_1 + \cdots + \alpha_d$ is the "total order" of the multi-index. In this case, we refer to the basis as the so called "full basis" for a given total polynomial order P.

Hyperbolic Truncation Strategy. Following the sparsity-of-effects principle, high-order interaction terms (i.e. polynomial terms involving several uncertain parameters of $\boldsymbol{\theta}$) are often less important in physical problems and can be neglected when compared to main effects (i.e. polynomial terms involving a single uncertain parameter of $\boldsymbol{\theta}$) and low-order interaction terms. As an alternative to the linear truncation strategy, the "hyperbolic" truncation strategy consists in eliminating a priori high-order interaction terms. A more general approach than Eq. (2.22) to define the number of terms r in the gPC expansion consists in introducing q-quasi-norms:

$$\mathcal{A}_{\text{hyp}} \equiv \mathcal{A}_{\text{hyp}}(d, P, q) = \left\{\boldsymbol{\alpha} \in \mathbb{N}^d : ||\boldsymbol{\alpha}||_q \leq P\right\}, \tag{2.23}$$

where the q-semi-norm is defined as

$$||\boldsymbol{\alpha}||_q \equiv \left(\sum_{i=1}^{d} (\alpha_i)^q\right)^{1/q}. \tag{2.24}$$

The number of terms in the gPC expansion is expressed by the cardinality of $\mathcal{A}$, which varies according to P and q for a given dimension d. The adoption of such semi-norms penalizes high-rank indices and high-order interactions. That is, the lower the value of q, the higher the penalty in the determination of $\mathcal{A}$. As a limiting case, setting $q = 1$ we retrieve the linear truncation strategy and thus a full basis of cardinality $\mathcal{A}_{\text{lin}}(d, P)$.

Sparse Truncation Strategies. There are alternatives to reduce the number of terms in the gPC-expansion. We will now schematically represent three of them, ordered by their growing complexity: 1-"sequential strategy", 2- "cleaning strategy", 3- "least angle regression".

1- The sequential strategy [25] consists in constructing the gPC-expansion in an incremental way, starting from the first term Ψ_0 ($K_0 = \{0\}$) and adding one term at a time in the basis ($K_{i+1} = K_i \cup \{\Psi_{i+1}\}$). The terms that are sequentially added to the basis are ordered according to the adopted enumeration strategy (linear or hyperbolic). The response surface is therefore of increasing complexity, since the enumeration functions in both cases increase the polynomial complexity when increasing the index. In the applications in this Manuscript, the construction process is stopped when a given accuracy is achieved, or when the number of terms in the gPC-expansion reaches the maximum size of the basis $r_{\max}$ specified by the user.

2- An alternative to the sequential strategy is the cleaning strategy [25], which builds a gPC-expansion containing at most $r_{\max}$ significant coefficients, i.e. at most $r_{\max}$ significant basis functions, starting from the full basis (still retaining the constraint of hyperbolic truncation if selected). The main idea of the cleaning strategy is to discard from the active basis the polynomials $\Psi_{\boldsymbol{\alpha}}$ that are associated with low magnitude coefficients, i.e. satisfying

$$|\gamma_{\boldsymbol{\alpha}}| \leq \epsilon \cdot \max_{\boldsymbol{\alpha}' \in \mathcal{A}'} |\gamma_{\boldsymbol{\alpha}'}| \tag{2.25}$$

where ϵ is the significance factor (set in this work to 10^{-4}), and where $\mathcal{A}'$ represents the current active basis. This selection procedure means that the terms in the gPC-expansion are not ordered following the degree of the polynomial functions but instead according to the magnitude of the coefficients.

3- In complement to the aforementioned strategies, there is a more advanced approach called least-angle regression (LAR) to select the active polynomial terms. The key idea of the LAR algorithm is to select at each iteration a polynomial among the r terms of the full basis (or eventually the hyperbolic-truncated basis) based on the correlation of the polynomial term with the current residual. The selected term is then added to the active set of polynomials. The coefficients of the active basis are computed so that every active polynomial is equi-correlated with the current residual until convergence is reached. Hence, LAR builds a collection of surrogates that are less and less sparse along the iterations. Iterations stop either when the full basis has been looked through or when the maximum size of the training set has been reached. When the iterations stopped, the polynomial coefficients are computed via the least-square algorithm presented in the following paragraphs. Further details can be found in Refs. [32, 31, 76].

Projection strategy

In this manuscript, we focus on non-intrusive approaches based on ℓ_2-minimization methods to numerically compute the coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathcal{A}}$ using the N snapshots from the training set $\mathcal{D}_N$.

Galerkin Pseudo-Spectral Projection. This Galerkin-type projection is based on the orthonormality property of the polynomial basis. Adopting this approach, the $\boldsymbol{\alpha}$th coefficient $\gamma_{\boldsymbol{\alpha}}$ is computed using the definition of the inner product. The latter is numerically approximated using tensor-based Gauss quadrature (referred to as "quadrature" in the following) as follows:

$$\gamma_{\boldsymbol{\alpha}} = \langle \mathrm{y}, \Psi_{\boldsymbol{\alpha}} \rangle \cong \sum_{k=1}^{N} \mathrm{y}^{(k)} \, \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}^{(k)}) \, w^{(k)}, \tag{2.26}$$

where $\mathrm{y}^{(k)} = \mathcal{M}(\boldsymbol{\theta}^{(k)})$ is the kth snapshot of the $\mathcal{D}_N$-database corresponding to the Forward Model evaluations for the kth quadrature root $\boldsymbol{\theta}^{(k)}$ of $\Psi_{\boldsymbol{\alpha}}$, and where w^k is the weight associated with $\boldsymbol{\theta}^{(k)}$. When considering a full basis, $(P+1)$ is the number of quadrature roots required in each uncertain direction to ensure an accurate calculation of the integral $\langle \mathrm{y}, \Psi_{\boldsymbol{\alpha}} \rangle$. Hence, in our problem, we have $N = (P+1)^3$ simulations in the training set to build the PC surrogates through Galerkin pseudo-spectral projection.

Least-Square Minimization. With this approach, the estimation of the coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ is done by solving a least-square minimization problem, i.e. by minimizing the approximation error between the (exact) `LSfire+` model evaluations and the PC-surrogate estimations at the points of the training set $\mathcal{D}_N$. The least-square projection solves a minimization problem over the given basis as follows:

$$\widehat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^r}{\operatorname{argmin}} \sum_{k=1}^{N} \left(\mathrm{y}^{(k)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}^P} \gamma_{\boldsymbol{\alpha}} \, \Psi_{\boldsymbol{\alpha}} \left(\mathbf{x}^{(k)} \right) \right)^2 \tag{2.27}$$

which is achieved by the means of classical linear algebra algorithms. Note that the sample size N required by this strategy for the problem to be well posed is at least equal to $(r+1)$, where r is the total number of gPC coefficients to be computed (i.e. the cardinality of the set $\mathcal{A}$). Note also that in the present work this is the projection method used to compute the coefficients selected by the sparse truncation methods (sequential, cleaning or LAR); when using non-sparse truncation strategies, this is referred to as the standard least-square (SLS) approach.

Workflow scheme for constructing the gPC-expansion

The algorithm relative to the construction of the gPC-surrogate is here summarized as follows:

1. choice of the polynomial basis $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ following the assumed marginal PDFs of the inputs $\boldsymbol{\theta}$;
2. choice of the total polynomial degree P (problem-dependent choice);
3. truncate the expansion to r_{lin} (or r_{hyp}) prior terms terms corresponding to the multi-index set $\mathcal{A}_{\text{lin}} or \mathcal{A}_{\text{hyp}}$ to keep the predominant information given by the forward model using linear or hyperbolic truncation strategy (r_{lin} depends on d, P; r_{hyp} depends on d, P and q with q the hyperbolicity factor so that $0 < q \leq 1$);

4. if a sparse strategy is used (sequential, cleaning or LAR), find a suitable set of multi-indices $\mathcal{A} \subset \mathcal{A}_{\text{lin,hyp}}$ with a cardinality $r \leq r_{\text{lin, hyp}}$, otherwise skip this step;

5. projection strategy (quadrature or least-square) to compute the coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathcal{A}\subset\mathbb{N}^d}$ using $N = (P+1)^d$ snapshots from the simulation database $\mathcal{D}_{N_{\text{ref}}}$;

6. formulate the surrogate model $\mathcal{M}_{\text{pc}}$, which can now be evaluated for any new pair of parameters $\boldsymbol{\theta}^*$.

2.4.3 Gaussian Process surrogate

As stated by Ref. [229], a GP is a random process (here the QoI of the forward model $\boldsymbol{y}$) indexed over a domain (here $\mathbb{R}^d$), for which any finite collection of process values (here $\left\{\mathrm{y}(\boldsymbol{\theta}^{(k)})\right\}_{1\leq k\leq N}, \boldsymbol{\theta}^{(k)} \in \Theta$) is characterized by a joint Gaussian distribution. More specifically, let $\widetilde{\mathrm{y}}$ be a Gaussian random process completely described by its zero mean and its correlation π:

$$\widetilde{\mathrm{y}}(\boldsymbol{\theta}) \sim \mathrm{GP}\left(0, \sigma_{\text{gp}}^2\, \pi(\boldsymbol{\theta}, \boldsymbol{\theta}')\right), \tag{2.28}$$

with $\pi(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{E}\left[\widetilde{\mathrm{y}}(\boldsymbol{\theta})\widetilde{\mathrm{y}}(\boldsymbol{\theta}')\right]$. In the present case, the correlation function π (or kernel) is chosen as a squared exponential (also known as "RBF kernel"):

$$\pi(\boldsymbol{\theta}, \boldsymbol{\theta}') = \exp\left(-\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2}{2\,\ell_{\text{gp}}^2}\right), \tag{2.29}$$

where ℓ_{gp} is a length scale which represents the model output dependency between two input vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, and where σ_{gp}^2 is the variance of the output signal. The surrogate model of the QoI is thus the mean of the GP, resulting from the conditioning of $\widetilde{\mathrm{y}}$ on the training set $\mathcal{Y} = \left\{\mathrm{y}\left(\boldsymbol{\theta}^{(k)}\right)\right\}_{1\leq k\leq N}$. The quantity of interest provided by the GP-surrogate for any given $\boldsymbol{\theta}^* \in \mathbb{R}^d$ satisfies

$$y_{\text{gp}}(\boldsymbol{\theta}^*) = \sum_{k=1}^{N} \beta_k\, \pi\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(k)}\right), \tag{2.30}$$

where

$$\beta_k = \left(\boldsymbol{\Pi} + \tau_{\text{gp}}^2\, \mathbf{I}_N\right)^{-1} \left(\mathrm{y}(\boldsymbol{\theta}^{(1)}) \dots \mathrm{y}(\boldsymbol{\theta}^{(N)})\right)^T, \tag{2.31}$$

$$\boldsymbol{\Pi} = \left(\pi(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(k)})\right)_{1\leq j,k\leq N}, \tag{2.32}$$

and where τ_{gp} (referred to as the "nugget effect") is used to circumvent ill-conditioning issues for the matrix $\boldsymbol{\Pi}$. The hyperparameters $\{\ell_{\text{gp}}, \sigma_{\text{gp}}, \tau_{\text{gp}}\}$ are optimized through maximum likelihood applied to the dataset $\mathcal{D}_N$ adopting a basin hopping technique [300].

Workflow

The algorithm relative to the construction of the GP-model is resumed below:

1. choose the kernel function π suitable for the input vector $\boldsymbol{\theta} = (\theta_1 \dots \theta_d) \in \mathbb{R}^d$ – we consider RBF throughout this Thesis, see Eq. (2.29);

2. optimize the hyperparameters of GP, $\{\ell_{\boldsymbol{gp}}, \sigma_{gp}, \tau\}$ associated with the kernel π using maximum likelihood;

3. formulate the surrogate $\mathcal{M}_{\text{gp}}$, which can be evaluated for any new vector of parameters $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_d^*)$ with Eq. (2.30) and Eq. (2.31).

2.4.4 Retrieving Results from Surrogates: GP and gPC

In this thesis manuscript, the adopted Surrogate Methods are gPC and GP. In practice, independently of the different model, the objective is to retrieve statistical moments or PDF of the QoI (Uncertainty Quantification) or to retrieve Sobol' Indices (variance-based SA). Below are listed the methods adopted to infer that relevant information from the built Surrogate Model.

Obtaining Information from Gaussian Process

Sensitivity Analysis: For GP approach, the Sobol indices can be computed via the Fourier Amplitude Sensitivity Testing (FAST) method [117] and its evolution eFAST [242], using the surrogate model to perform the required computations instead of time-consuming forward model evaluations. Sobol' indices may also be stochastically estimated using the Martinez' formulation. The latter estimator is stable and gives asymptotic confidence intervals for first-order and total-order indices [24, 171].
Uncertainty Quantification: After building a database $\mathcal{D}$ with the aid of Monte Carlo Sampling over the domain of the input parameter, Z_θ, and computing the surrogate on that set of samples, the mean and STD of the QoI will read

$$\mu_{\widehat{\mathrm{y}}} = \frac{1}{N_{\text{sample}}} \sum_{k=1}^{N_{\text{sample}}} \widehat{\mathrm{y}}^{(k)}, \tag{2.33}$$

$$\sigma_{\widehat{\mathrm{y}}} = \sqrt{\frac{1}{N_{\text{sample}} - 1} \sum_{k=1}^{N_{\text{sample}}} \left(\widehat{\mathrm{y}}^{(k)} - \mu_{\widehat{\mathrm{y}}}\right)^2}, \tag{2.34}$$

with $\widehat{\mathrm{y}}^{(k)}$ the kth element of the dataset $\mathcal{D}_{N_{\text{sample}}}$ containing the surrogate evaluations over the aforementioned Monte Carlo sampled points.

The PDF of the QoI can be obtained through kernel smoothing techniques applied to the database $\mathcal{D}_{GP,MC}$.

Obtaining Information from gPC

Sensitivity Analysis: For the PC approach, Sobol' indices can be directly derived from the PC coefficients. For the ith component of the input random variable $\mathbf{x}$, the Sobol' index reads:

$$S_{\mathrm{pc},i} = \frac{1}{\sigma^2_{h_{\mathrm{pc}}}} \sum_{\substack{j \in \mathcal{A} \subset \mathbb{N}^d, \\ j_i > 0 \text{ and } j_{k \neq i} = 0}} \gamma_j^2, \tag{2.35}$$

with $\sigma_{h_{\mathrm{pc}}}$ the STD computed in Eq. (6.33), shown in the following paragraph.

Uncertainty Quantification: Using the gPC metamodel, the statistical moments can be derived analytically with relative ease from the coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{N}^d}$ such that the mean and the STD will read:

$$\mu_{\widehat{\mathrm{y}}_{\mathrm{pc}}} = \gamma_0, \tag{2.36}$$

$$\sigma_{\widehat{\mathrm{y}}_{\mathrm{pc}}} = \sqrt{\sum_{\substack{\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{N}^d \\ \boldsymbol{\alpha} \neq 0}} \gamma^2_{\boldsymbol{\alpha}}}. \tag{2.37}$$

The PDF of the QoI can be obtained through Kernel Smoothing Techniques applied to a Montecarlo database of Surrogate Model evaluations $\mathcal{D}_{PC,MC}$, in a similar fashion to GP case.

2.4.5 Error Metrics

In the present thesis, three error metrics are used to assess the quality of the surrogate predictions: the empirical error between the surrogate prediction and the forward model $\mathcal{M}$ prediction (also known as "training error"), the Leave One Out (LOO) error and the Q_2 predictive coefficient [167].

Empirical Error ϵ_{emp}

The truncation of the gPC-expansion may introduce an approximation error at the training points, which can be computed posterior to the surrogate construction. This empirical error denoted by ϵ_{emp} reads

$$\epsilon_{\mathrm{emp}} = \frac{1}{N} \sum_{k=1}^{N} \left(\mathrm{y}^{(k)} - \widehat{\mathrm{y}}^{(k)} \right), \tag{2.38}$$

with $\mathrm{y}^{(k)}$ the kth element of the training set $\mathcal{D}_N$ and $\widehat{\mathrm{y}}^{(k)}$ its corresponding value, predicted this time by the surrogate, for the same element of the training set.

However, this error estimator has several drawbacks. First, the GP metamodel (when noise is not introduced in the kernel) is *de facto* an interpolator. Because of this, the approximation error is expected to be $\epsilon_{\mathrm{emp}} = 0$. Secondly, such estimator may severely underestimate the magnitude of the mean square error. When the size of the training set N comes closer to the cardinality of the gPC-expansion $\mathcal{A}$, ϵ_{emp} may tend to zero, while the actual mean square error does not; this problem is addressed to as "overfitting".

Leave-one-out Error

Since a more robust estimator is needed, the Leave-one-out (LOO) predictive coefficient based on *cross-validation* is often adopted in literature. LOO error is sometimes referred to as PRESS (Predicted Residual Sum of Squares) or *jacknife error*. It is built by excluding each time a different m^{th} collocation point in the building of the surrogate model $\widehat{\mathcal{M}}$ and computing the value of such surrogate model, addressed to as $\widehat{\mathcal{M}}_{-m}$, on the DOE collocation points. Such error estimator is expressed by the following relation

$$\epsilon_{\text{LOO}} = \frac{\frac{1}{N_{DOE}} \sum_{m=1}^{N_{DOE}} \left(\mathcal{M}(P_m) - \widehat{\mathcal{M}}_{-m}(P_m) \right)^2}{Var(\mathbb{P}_{N_{DOE}})}, \tag{2.39}$$

where $\mathbb{P}_{N_{DOE}}$ is the vector of DoE model evaluations.

Predictive coefficient Q_2

We require a more robust error estimator suitable for both gPC-expansion and GP-model. In this thesis, we use the Q_2 predictive coefficient based on cross validation. The computation of Q_2 relies on two distinct datasets: the current training set $\mathcal{D}_N$ and a *reference* sample $\mathcal{D}_{N_{\text{ref}}}$ that is independent of the surrogate construction and that is therefore referred to as the "validation dataset". Q_2 will then read

$$Q_2 = 1 - \frac{\sum_{k=1}^{N_{\text{ref}}} \left(\mathrm{y}^{(k)} - \widehat{\mathrm{y}}^{(k)} \right)^2}{\sum_{k=1}^{N_{\text{ref}}} \left(\mathrm{y}^{(k)} - \overline{\mathrm{y}}_{\text{ref}} \right)^2}, \tag{2.40}$$

with $\mathrm{y}^{(k)}$ the kth element of the validation dataset sample $\mathcal{D}_{N_{\text{ref}}}$, $\widehat{\mathrm{y}}^{(k)}$ the surrogate prediction for the same element of $\mathcal{D}_{N_{\text{ref}}}$ and $\overline{\mathrm{y}}_{\text{ref}}$ the empirical mean over the Validation Dataset $\mathcal{D}_{N_{\text{ref}}}$. Note that computing Q_2, the training set $\mathcal{D}_N$ is only used to construct the surrogate model and to obtain the estimation $\widehat{\mathrm{y}}$ of the quantity of interest y. The target value for Q_2 is 1.

2.4.6 Numerical Implementation

The GP implementation relies on the Python package *scikit-learn* [211] (see http://scikit-learn.org/). The gPC-implementation relies on the Python package *OpenTURNS* [25] (see www.openturns.org). The *batman* [239] Python package is used to build datasets and perform statistical analysis.

Philosophy is written in this grand book, which stands continually open before our eyes (I say the 'Universe'), but can not be understood without first learning to comprehend the language and know the characters as it is written. It is written in mathematical language, and its characters are triangles, circles and other geometric figures, without which it is impossible to humanly understand a word; without these one is wandering in a dark labyrinth.

Galileo Galilei

Chapter 3

Model Introduction

This Chapter contains:

In this chapter the model that will be the core of this thesis will be presented. Its main aim is to represent the motion of fronts when are evolving in a media characterized by microscopical randomness, here modeled through its Probability density Function (PDF). In the following, after a short introduction the similarities and links with other models already available in literature are presented. In Section 3.1 the mathematical model is described thoroughly. Later, in Section 3.2 an example of application of the model is illustrated. In particular, a model for a front characterized

by complex initial geometry subject to curvature, deterministic propagation and underlying anomalously diffusive fluctuations is derived in the framework of the proposed approach. Some thoughts about the main actors in the possible quenching of the front are given in 3.3. Subsequently, Section 3.4 describes an *ad hoc* setting of complex initial geometry to see the effects of the parameters related to both microscopic and macroscopic physics with a simple response analysis. Finally, Section 3.5 presents a Surrogate-based UQ and SA of the model for a selection of parameters and physical observables (QoI).

As introduced in Chapter 1, the tracking of fronts plays an important role in many fields of applied science. The model that is object of investigation in this manuscript focuses on the case when the propagating interface is embedded in a medium characterized by a random motion which implies some diffusion process. In classical diffusion processes (Gaussian diffusive process) the particle mean square displacement grows linearly in time. On the other hand, in complex media diffusion typically shows a non-linear growth in time. Some example of such phenomena can be found in biological systems [20, 114, 178], in fluids and plasmas [67] and in other systems [179].

In the latter case, it can be stated that *anomalous diffusion* takes place, instead of classical diffusion [34, 114, 20, 201].

The main bricks the proposed method is built with are listed below:

- the position of the *average front* is computed by making use of a sharp-front evolution method, such as the level set method (LSM) [250];
- the microscopical spread of particles which takes place around the average front is given by the probability density function (PDF) linked to the underlying diffusive process, *that is supposedly known in advance, say in a parametric formulation*;
- the *effective front* is computed by taking weighted averages of the average fronts adopting the previously chosen PDF as a weight function.

This methodology to depict random fronts has been introduced first to study the evolution of the burnt mass fraction in turbulent premixed combustion [202], which shall be the object of Chapter 4.

Later on, it has been applied successfully to wild-land fire propagation simulation [203, 204, 206, 199, 200], through an adaptation of the modeling approach that shall be treated in detail in Chapter 5.

In these two application fields, the chosen PDF has been chosen either as the Gaussian one [202, 203, 204, 206, 199, 200], or a convolution of the latter with another PDF closely related to the problem physics [206, 199, 200].

In the recent work of Mentrelli and Pagnini [2], anomalous diffusion has been first introduced in this front tracking framework adopting the time-fractional diffusion equation, with the required PDF

being the fundamental solution, or Green's function, of this family of fractional partial differential equations.

Similarities and difference with other existing models

The formalism proposed by this formulation is a very general one, and could in principle share some features with some well established algorithms. In [2] a thorough description of similarities between this and other approaches has been undertaken. More specifically, it emerged that the formalism that will be described in the next Section will have something in common with Smoothed Particle Hydrodynamics (SPH) [183]. This framework is characterized by the robustness of its numerical implementation and it is well suited for many practical problems as fluid-structure interactions. However, the typical problems of SPH, namely the choice of the kernel function and of the smoothing length, are here avoided because they are directly determined by the particle PDF. This can be considered a positive feature of the present approach.

Moreover, presented formalism could be linked with relative ease to procedures adopted in turbulent premixed combustion, where the LSM/G-equation [216] is widely utilized and the *turbulent flame closure model* introduced by Zimont [319, 241] can be taken into consideration. An analysis of the applicability of this approach in that specific combustion science area can be found in [202].

In addition to all this, the present approach is also connected to the so-called Stochastic Level Set Method (SLSM), developed and adopted in computer vision. Typical problems in this research area consist in recovering a certain surface or region through a shape optimization framework. However, classical methods suffer from being sometimes stuck in local minima. To overcome this difficulty, the SLSM has been developed. This method combines stochastic motion and the classical LSM [194, 259, 123] and adopts a decision mechanism. The SLSM is quite close to the approach here proposed, as in both models the interface location is obtained by solving the level set equation with a suitably randomized front speed. However, the final aims of the two methods are different, and so is the mathematical construction. In fact, in typical computer vision problems the final aim is not to compute the average shape after a large number of independent realizations, but to recover a certain shape with the frontline upon the derivation of a proper stochastic differential equation. In the present study, the interest is indeed focused on the average properties of the processes for applications in physics and engineering. Even if it could be possible to have average properties of the process obtained with the SLSM by performing the ensemble average of many independent realizations, the approach here proposed uses physical arguments to derive the PDF of the interface particle displacement instead of focusing on the selection of the correct stochastic differential equation, so it can be regarded as the averaged counterpart of the method adopted in computer vision.

3.1 Modelling approach

The core model that is declinated in other forms for each kind of specific application is the same discussed in Refs. [206, 126]. Its modularity allow for straightforward adaptations towards each specific modeling target.

This approach is based on the idea to split the motion of the front into a *drifting* part and a *fluctuating* part and the front position is also split correspondingly. This splitting allows specific numerical and physical choices that can improve the algorithms and the models. In particular, the drifting part can be obtained by existing methods for simulate front propagation, for example the Eulerian LSM (or G-equation in the combustion community) or the Lagrangian DEVS. The fluctuating part is the result of a comprehensive statistical description of the phenomenon which includes the random effects in agreement with the physical properties of the system. As a consequence, the fluctuating part can have a non-zero mean (for example, due to ember jump lengths in wildland fire propagation [206, 126]), which means that the drifting part does not correspond to the average motion. This last fact makes a distinction between the present splitting and the well-known Reynolds decomposition adopted in turbulence studies.

The random front line is obtained as follows. First a front contour $\Gamma(t)$ is computed by using an opportune existing method (e.g., LSM or DEVS) and this step gives the drifting part of the front position. Later a *deterministic* indicator function is introduced which takes the values 0 and 1 outside and inside the domain $\Omega(t) \subset \mathbb{R}^n$, $n \geq 1$, surrounded by the computed front contour $\Gamma(t)$. Since the front line can be re-located in a different position by using the translation or *sifting property* of the Dirac delta function, by using a delta function peaked on a random position, a *random front contour* is constructed. A *stochastic* indicator function can be determined by an integral formula involving the product of the deterministic indicator function times the Dirac delta function peaked along the stochastic trajectory. Finally, by applying the ensemble average, the averaged stochastic indicator function states the effective front position by a smooth function $\varphi_e : \mathcal{S} \times \mathbb{R}_0^+ \to [0,1]$, where $\mathcal{S} \subseteq \mathbb{R}^n$ is the domain of interest. A selected isoline/isosurface of this smooth function represents the effective front line emerging from the proposed method, i.e., $\Gamma_e(t) = \{\mathbf{x} \in \mathcal{S} | \varphi_e(\mathbf{x}, t) = \varphi_*\}$.

Let $\mathbf{x} \in \mathcal{S} \subseteq \mathbb{R}^n$ and $t \in \mathbb{R}_0^+$, in formulae we have the following indicator function $I_\Omega : \mathcal{S} \times \mathbb{R}_0^+ \to \{0,1\}$ defined as

$$I_\Omega(\mathbf{x}, t) = \begin{cases} 1\,, & \mathbf{x} \in \Omega(t) \\ \\ 0\,, & \text{elsewhere} \end{cases}\,, \tag{3.1}$$

where $\Omega(t)$ is the domain surrounded by the frontline generated at time t by the adopted operational code. Let $\mathbf{X}^\omega(t, \overline{\mathbf{x}}) = \overline{\mathbf{x}} + \xi_t^\omega$ be the ω-realization of a stochastic trajectory where $\overline{\mathbf{x}}$ is a point inside the domain $\Omega(t)$ established at time t by the adopted operational code and ξ_t is the random noise. The one-particle density function of $\mathbf{X}^\omega(t, \overline{\mathbf{x}})$ is $p^\omega(\mathbf{x}; t|\overline{\mathbf{x}}) = \delta(\mathbf{x} - \mathbf{X}^\omega(t, \overline{\mathbf{x}}))$ and, after the ensemble averaging denoted here by $\langle \cdot \rangle$, the PDF of $\mathbf{X}^\omega(t, \overline{\mathbf{x}})$ results to be $p(\mathbf{x}; t|\overline{\mathbf{x}}) = \langle \delta(\mathbf{x} - \mathbf{X}^\omega(t, \overline{\mathbf{x}})) \rangle$. The

indicator function $I_\Omega(\mathbf{x},t)$ turns to be stochastic by using the sifting property of the delta function as explained above, i.e.,

$$I^\omega_\Omega(\mathbf{x},t) = \int_{\mathcal{S}} I_\Omega(\overline{\mathbf{x}},t)\delta(\mathbf{x}-\mathbf{X}^\omega(t,\overline{\mathbf{x}}))\,d\overline{\mathbf{x}}\,. \tag{3.2}$$

Finally, after ensemble averaging of $I^\omega_\Omega(\mathbf{x},t)$ it follows

$$\begin{aligned}\langle I^\omega_\Omega(\mathbf{x},t)\rangle &= \langle \int_{\mathcal{S}} I_\Omega(\overline{\mathbf{x}},t)\delta(\mathbf{x}-\mathbf{X}^\omega(t,\overline{\mathbf{x}}))\,d\overline{\mathbf{x}}\rangle \\ &= \int_{\mathcal{S}} I_\Omega(\overline{\mathbf{x}},t)\langle\delta(\mathbf{x}-\mathbf{X}^\omega(t,\overline{\mathbf{x}}))\rangle\,d\overline{\mathbf{x}} \\ &= \int_{\mathcal{S}} I_\Omega(\overline{\mathbf{x}},t)p(\mathbf{x};t|\overline{\mathbf{x}})\,d\overline{\mathbf{x}} \\ &= \int_{\Omega(t)} p(\mathbf{x};t|\overline{\mathbf{x}})\,d\overline{\mathbf{x}} = \varphi_e(\mathbf{x},t)\,. \end{aligned} \tag{3.3}$$

The averaging procedure can pass into the integral because the domain of integration is not affected by the stochastic motion and the indicator function $I_\Omega(\overline{\mathbf{x}},t)$ comes out of the averaging brackets because it is a deterministic quantity.

Applying the Reynolds transport theorem and subsequently the divergence theorem to formula (3.3), the following reaction-diffusion equation is obtained

$$\frac{\partial\varphi_e}{\partial t} = \int_{\Omega(t)} \frac{\partial p}{\partial t}\,d\overline{\mathbf{x}} + \int_{\Omega(t)} \nabla_{\overline{\mathbf{x}}}\cdot\left[\mathbf{V}(\overline{\mathbf{x}},t)\,p(\mathbf{x};t|\overline{\mathbf{x}})\right]d\overline{\mathbf{x}}\,, \tag{3.4}$$

where $\mathbf{V}(\mathbf{x},t)$ is the expansion velocity of the domain $\Omega(t)$. With reference to the right-hand side of equation (3.4), the first term gives the spreading and represents the diffusive part. In particular, in the case when $p(\mathbf{x};t|\overline{\mathbf{x}})$ evolves locally in time, this term reduces to $\mathcal{L}\varphi_e$, where $\mathcal{L}$ is a generic spatial operator that can be for example the classical or the fractional Laplacian. The second term on the right-hand side is indeed the source term. In fact, this term depends on the velocity of the temporal expansion of the domain $\Omega(t)$, and for this it is responsible for the insertion of further amount of a proper quantity into the simulation domain.

In the family of reaction-diffusion equations (3.4), the source term includes the front velocity of the moving interface scheme $\mathbf{V}(\mathbf{x},t)$ and the fundamental solution $p(\mathbf{x};t|\overline{\mathbf{x}})$ of equation (3.4) in the non-reactive case, i.e., null source term. Hence, this approach reconciles two approaches considered alternative to each other that is the ones based on moving interface schemes (which generate a sharp function that is non zero inside a bounded domain and zero outside) and the ones based on reaction-diffusion equations (whose solution is generally a continuous smooth function that has an exponential decay, and it is non zero in an infinite domain). In fact, even if equation (3.4) is a reaction-diffusion equation it is strongly based on the velocity of growth of the domain $\Omega(t)$ that is indeed established by a moving interface scheme.

It is known that the ensemble average of Dirac delta functions with stochastic arguments representing independent realizations of a stochastic process can be understood as the PDF corresponding

to the so-called *empirical* distribution function [306]. This means that the effective front position emerges to be described by a weighting procedure of the drift-generated front where the weight-function is the PDF of the fluctuating part. Due to the presence of the deterministic indicator function in the integrand, the above weighting procedure is equivalent to compute the integral of the PDF of fluctuations for the values of drifting component belonging to the internal region of the front contour computed from a literature method (e.g. LSM, DEVS or G-equation). Any application of the method is therefore strongly dependent on the correct determination of the statistics of the fluctuating part because it describes the physics of the specific underlying system and, in particular, on the determination of the PDF of the fluctuations.

Some remarks on the above method are:

i) if a non-random initial front contour is considered, the initial condition of the PDF of fluctuations is a Dirac delta function peaked on internal point of the initial front line.

ii) if no randomness occurs in the process, the PDF of fluctuations remains a Dirac delta function so the result of the proposed method reduces to the chosen literature method (e.g., LSM or DEVS);

iii) the proposed method avoids to use the Reynolds decomposition of the value of the observable function, as it was erroneously assumed sometimes in literature in the context of turbulent premixed combustion [214, 215, 148]. In fact, by definition, the curve representing the frontline does not have fluctuations in its value in any adopted method, but it has fluctuations in its position and these fluctuations are exactly what is modelled in the present method by the fluctuating part and its PDF. It should be remarked here that this approach avoids to assume explicit fluctuations of the level-set function, see also [193], as in other literature approaches [214, 215, 148, 240].

iv) the resulting effective front differs from an average procedure proposed by Oberlack et al. [193], because in the method proposed here the full PDF is considered, which means that the variance of interface position is also included, along with the average position. This difference is fundamental to take into account the diffusive characteristics, the displacement variance being closely related to velocity correlation, i.e., Green–Kubo–Taylor formula, which is the most important element to characterize the structure of stochastic processes. If the displacement variance tends to zero, then the PDF tends to the Dirac delta function and the average G-equation derived by Oberlack et al. [193] is recovered.

3.2 Example of model application: complex front kinetics in random media

To conclude this Chapter, an experiment combining salient features from both drifting part and fluctuating one is presented.

It is an extension of the multidimensional front evolution in anomalous diffusive media proposed in [2], considering more complex geometry and extending the underlying microscopical fluctuations in order to model the so called *Erdélyi–Kober fractional diffusion*.

In the one–dimensional case of such setting, the Green's function of the PDF that appear in formula (3.3) is related to the $\mathcal{M}$-Wright/Mainardi function [155, 153, 198], which reduces to the Gaussian distribution as the fractional order of the time derivative tends to one. The $\mathcal{M}$-Wright/Mainardi function is a well-established function, and with state-of -art robust algorithm available for its computation [223].

To deal with multi-dimensionality, the one–dimensional case analyzed in [176] has been extended to multiple dimension by [108] and put in a more simple formulation by [2].

The latter work showed specifically that using an integral representation formula of the $\mathcal{M}$-Wright/Mainardi function [156], Green's function for multi-dimensional cases can be represented by the same integral formula of the one-dimensional one with the help of a simple multi-dimensional Gaussian PDF.

The deterministic evolution of a single mean front is analyzed with the well-known Level Set Method as the selected sharp front technique, already described in detail in Chapter 1.

We remind that in the framework of the LSM, the position of the interface identifies with an isocontour (for the sake of simplicity the zero isocontour) of a certain a scalar field which is obtained solving the so called *level set equation*, an equation of the Hamilton–Jacobi type.

For this experiment we chose such framework for the computation of the drifting part since it has been proven useful to depict a plethora of geometrical configurations, and gives simple recipes for interface kinetics driven by normal vector, curvature, or external velocity fields; sharp gradients, cusps and even change of topology of the fronts (merging and splitting of sub-domains) can be described with relative ease.

In addition to these advantages, one key-feature of the level set method framework is that it allows to easily improve the physical model by progressively including more detailed interface effects that might influence the front propagation, as the dispersion effects of interface particles due to diffusive phenomena considered in this experiment.

Given the recent extension to the multi–dimensional case of the well established one–dimensional model useful for the study of the propagation of fronts in anomalous diffusive media [176], the experiment highlighted in this Chapter provides the latter model a time stretching effect and a

curvature driving term, in order to analyze a complex kinetic where there is an interplay between *microscopic, stochastic* and *macroscopic, geometric* features.

3.2.1 Computation of the drifting part

Let $\gamma : \mathcal{S} \times \mathbb{R}_0^+ \to \mathbb{R}$, then its evolution in the framework of the LSM is governed by the following Hamilton–Jacobi equation

$$\frac{\partial \varphi}{\partial t} = \mathcal{V}(\mathbf{x}, t) \, \|\nabla \varphi\|, \tag{3.5}$$

that corresponds to formula (1.8) of Chapter 1 with null external velocity field $\mathbf{v}$ and contact speed $\mathcal{V}(\mathbf{x}, t)$. The domain $\Omega(t)$ is that one bounded by the front contour $\Gamma(t)$ now established by the level set value $\varphi(\mathbf{x}, t) = \varphi*$, i.e., $\Gamma(t) = \{\mathbf{x} \in \mathcal{S} | \varphi(\mathbf{x}, t) = \gamma_*\}$.

Front curvature is connected to the surface density per unit volume, and this affects the speed of the propagation when the propagation is due to a chemical reaction, as for example in combustion [] and material science []. Equation (3.5) with a curvature-dependent speed reads

$$\frac{\partial \varphi}{\partial t} = \mathcal{V}_0(1 - \lambda \kappa(\mathbf{x}, t)) \, \|\nabla \varphi\| \, , \tag{3.6}$$

where κ is the curvature of the front, i.e.,

$$\kappa(\mathbf{x}, t) = \nabla \cdot \left(\frac{\nabla \varphi(\mathbf{x}, t)}{\|\nabla \varphi(\mathbf{x}, t)\|} \right) , \tag{3.7}$$

$\mathcal{V}_0$ is a drifting speed here assumed constant and λ is a parameter that regulates the influence of the curvature on the front speed. In combustion literature, λ is called Markstein length.

By including the curvature, the mathematical problem passes from hiperbolic to parabolic. This transition could affect the evolution of the front by provoking for example a quenching behavior, as discussed in Section 3.3.

3.2.2 Computation of the fluctuating part

Erdélyi–Kober fractional diffusion

The PDF of fluctuations can be derived from the analysis of the physics of the system. So in the case of classical diffusive environments it is Gaussian while it can be a more general PDF when anomalous diffusion occurs, e.g., [154, 208, 176, 2].

The generalized diffusion process here considered is the isotropic multi-dimensional Erdélyi–Kober fractional diffusion [208], and its Green function is the PDF $p(\mathbf{x}; t)$ to plug into in Eq. 3.3 .

The Erdélyi–Kober fractional diffusion equation is obtained from the classical diffusion equation by introducing a power-law memory kernel togeher with a time-stretching change of variable. As it will be shown in detail, one of the main outcomes of this modelling choice is the clear separation

between the shape of the PDF and the variance of the underlying stochastic process because they are controlled by two independent parameters.

Consider the standard Gaussian (or Normal) diffusion, the evolution of the process is driven by the well-know heat equation

$$\frac{\partial f}{\partial t} = \nu\,\Delta f\,, \quad f(\mathbf{x};0) = \delta(\mathbf{x})\,, \tag{3.8}$$

whose solution is

$$f(\mathbf{x};t) = \frac{1}{(4\pi\nu t)^{n/2}} \exp\left\{-\frac{\|\mathbf{x}\|^2}{4\nu t}\right\}, \tag{3.9}$$

that is characterised by a variance which grows linearly in time, i.e.,

$$\sigma_n^2 = \int_{\mathbb{R}^n} \|\mathbf{x}\|^2\, f(\mathbf{x};t)\, d\mathbf{x} = 3\sigma_1^2\,, \quad \sigma_1^2 = 2\nu\, t\,. \tag{3.10}$$

It is here reminded that the density function $P(x,t)$ with general initial condition $P(x,0) = P_0(x)$ is connected to the fundamental solution $f(x,t)$ by a convolution integral, i.e.

$$P(x,t) = \int_{-\infty}^{+\infty} f(\xi,t) P_0(x-\xi)\, d\xi\,. \tag{3.11}$$

Following Mura, Taqqu and Mainardi [186], this classical setting of diffusion can be generalized by considering a PDF of the form

$$P(\mathbf{x};t) = \int_0^{\infty} f(\mathbf{x};\tau) h(\tau;g(t)) d\tau\,, \tag{3.12}$$

where $f(\mathbf{x},t)$ is the Gaussian density in (3.9) and $h(\tau,g(t))$ introduces both a memory kernel $K(t) \geq 0$ and a change of the time variable $t \to g(t) \geq 0$, where g is smooth and increasing with $g(0) = 0$, because it is the solution of the equation

$$\begin{aligned} h(\tau;g(t)) &= h(\tau;0) - \int_0^{g(t)} K[g(t)-g(s)] \frac{\partial h(\tau;g(s))}{\partial \tau}\, dg(s) \\ &= h(\tau;0) - \int_0^{t} \frac{\partial g}{\partial s} K[g(t)-g(s)] \frac{\partial h(\tau;g(s))}{\partial \tau}\, ds\,. \end{aligned} \tag{3.13}$$

Then the PDF $P(\mathbf{x};t)$ defined in (3.12) solves the equation

$$P(\mathbf{x};t) = P_0(\mathbf{x}) + \nu \int_0^t \frac{\partial g(s)}{\partial s} K[g(t)-g(s)]\, \Delta P(\mathbf{x};s)\, ds\,, \tag{3.14}$$

whose explicit derivation can be found in Ref. [186]. Consider the following case

$$K(t) = \frac{t^{\beta-1}}{\Gamma(\beta)}\,, \quad g(t) = t^{\alpha/\beta}\,, \quad 0 < \alpha \leq 2\,, \quad 0 < \beta \leq 1\,, \tag{3.15}$$

from equation (3.14) we obtain

$$P(x,t) = P_0(x) + \nu \frac{\alpha}{\beta} \frac{1}{\Gamma(\beta)} \int_0^t s^{\alpha/\beta-1} (t^{\alpha/\beta} - s^{\alpha/\beta})^{\beta-1}\, \Delta P(x;s)\, ds\,. \tag{3.16}$$

Equation (3.16) can be re-written in the framework of Erdélyi–Kober fractional integrals, i.e.,

$$P(x,t) = P_0(x) + \nu t^{\alpha} \left[\mathbb{I}^{0,\beta}_{\alpha/\beta} \, \Delta P \right] \tag{3.17}$$

where the Erdélyi–Kober fractional integral operator $\mathbb{I}^{\ell,m}_{q}$ is defined in the following.

Let μ, η and γ be $\mu > 0$, $\eta > 0$ and $\gamma \in \mathcal{R}$, the Erdélyi–Kober fractional integral operator $\mathbb{I}^{\gamma,\mu}_{\eta}$, for a sufficiently well-behaved function $\varphi(t)$, is defined as [128, formula (1.1.17)]

$$\begin{aligned} \mathbb{I}^{\gamma,\mu}_{\eta} \, \varphi(t) &= \frac{t^{-\eta(\mu+\gamma)}}{\Gamma(\mu)} \int_0^t \tau^{\eta\gamma} \, (t^{\eta} - \tau^{\eta})^{\mu-1} \varphi(\tau) \, d(\tau^{\eta}) \, , \\ &= \frac{\eta}{\Gamma(\mu)} t^{-\eta(\mu+\gamma)} \int_0^t \tau^{\eta(\gamma+1)-1} (t^{\eta} - \tau^{\eta})^{\mu-1} \varphi(\tau) \, d\tau \, . \end{aligned} \tag{3.18}$$

As also remarked by Pagnini in [198], in the particular case $\gamma = 0$ and $\eta = 1$, the Erdélyi–Kober fractional integral operator (3.18) and the so called Riemann–Liouville fractional integral of order μ, here noted by J^{μ}, are related by the formula

$$\mathbb{I}^{0,\mu}_{1} \, \varphi(t) = \frac{t^{-\mu}}{\Gamma(\mu)} \int_0^t (t-\tau)^{\mu-1} \varphi(\tau) \, d\tau = t^{-\mu} J^{\mu} \, \varphi(t) \, . \tag{3.19}$$

Deriving both sides of Equation 3.17 we obtain its diffusive form, which reads

$$\frac{\partial P}{\partial t} = \nu t^{\alpha-1} \left[\alpha + t \frac{\partial}{\partial t} \right] \left(\mathbb{I}^{0,\beta}_{\alpha/\beta} \, \Delta P \right) \, , \tag{3.20}$$

and finally, introducing Erdélyi–Kober fractional differential operator $\mathbb{D}^{\ell,m}_{q}$,

$$\frac{\partial P}{\partial t} = \nu \frac{\alpha}{\beta} t^{\alpha-1} \, \mathbb{D}^{\beta-1,1-\beta}_{\alpha/\beta} \, \Delta P \, , \tag{3.21}$$

The definition of the Erdélyi-Kober fractional differential operator is here reported. Let $n-1 < \mu \leq n$, $n \in \mathcal{N}$, the Erdélyi–Kober fractional derivative is defined as [128, formula (1.5.19)]

$$\mathbb{D}^{\gamma,\mu}_{\eta} \, \varphi(t) = \prod_{j=1}^{n} \left(\gamma + j + \frac{1}{\eta} t \frac{d}{dt} \right) \left(\mathbb{I}^{\gamma+\mu,n-\mu}_{\eta} \, \varphi(t) \right) . \tag{3.22}$$

Such differential operator shares some connection with other fractional operators, e.g. Riemann–Liouville fractional derivative. We recall that Riemann–Liouville fractional derivative of order μ, $m-1 < \mu \leq m$, $m \in \mathcal{N}$ is defined as $\mathbb{D}^{\mu}_{RL} \, \varphi(t) = \dfrac{d^m}{dt^m} J^{m-\mu} \, \varphi(t)$, and it is linked with the Erdélyi–Kober fractional derivatives by the relation

$$\mathbb{D}^{-\mu,\mu}_{1} \, \varphi(t) = t^{\mu} \mathbb{D}^{\mu}_{RL} \, \varphi(t) \, . \tag{3.23}$$

A further important property of the Erdélyi–Kober fractional derivative is the reduction to the identity operator when $\mu = 0$, i.e.,

$$\mathbb{D}^{\gamma,0}_{\eta} \varphi(t) = \varphi(t) \, . \tag{3.24}$$

Mura, Taqqu and Mainardi [186] pointed out that with the choice (3.15), $h(\tau; g(t))$ turns out to be:

$$h(\tau; g(t) = t^{\alpha/\beta}) = \frac{1}{g(t)^{\beta}} \mathcal{M}_{\beta} \left(\frac{\tau}{g(t)^{\beta}} \right) = \frac{1}{t^{\alpha}} \mathcal{M}_{\beta} \left(\frac{\tau}{t^{\alpha}} \right) \, , \tag{3.25}$$

where $\mathcal{M}_\eta(z)$, $z \in \mathbb{C}$ and $0 < \eta < 1$, denotes the so called $\mathcal{M}$-Wright/Mainardi function [151, 152, 155, 153]

$$\mathcal{M}_\eta(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!\,\Gamma\left(-\eta n + (1-\eta)\right)} = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{(-z)^{n-1}}{(n-1)!}\,\Gamma(\eta n)\,\sin\left(\pi\eta n\right)\,. \tag{3.26}$$

The $\mathcal{M}_\eta$ function is a special case of the Wright function [155, 208] and it is an entire function of order $1/(1-\eta)$. It provides a generalization of the Gaussian density and the following special cases hold:

$$\mathcal{M}_0(z) = \mathrm{e}^{-z}\,, \quad \mathcal{M}_{1/3}(z) = 3^{2/3} Ai(z/3^{1/3})\,, \quad \mathcal{M}_{1/2}(z) = \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-z^2/4}\,, \tag{3.27}$$

where Ai is the Airy function.

We also remind the following integral representation [156]:

$$\frac{1}{t^{\eta/2}}\mathcal{M}_{\eta/2}\left(\frac{r}{t^{\eta/2}}\right) = 2\int_0^\infty \frac{\mathrm{e}^{-r^2/(4\tau)}}{\sqrt{4\pi\tau}}\,\mathcal{M}_\eta\left(\frac{\tau}{t^\eta}\right)\frac{d\tau}{t^\eta}, \quad 0 < \eta < 1\,, \quad r \in \mathbb{R}\,. \tag{3.28}$$

By setting $h(\tau;0) = \delta(\tau)$ we have $P_0(\mathbf{x}) = \delta(\mathbf{x})$.

Finally, the desired PDF which is the Green function of (3.16) is

$$\begin{aligned} p(\mathbf{x};t) &= \int_0^\infty \frac{\mathrm{e}^{-\|\mathbf{x}\|^2/(4\nu\tau)}}{(4\pi\nu\tau)^{n/2}}\,\mathcal{M}_\beta\left(\frac{\tau}{t^\alpha}\right)\frac{d\tau}{t^\alpha} \\ &= \int_0^\infty \prod_{i=1}^{n} \frac{\mathrm{e}^{-x_i^2/(4\nu\tau)}}{(4\pi\nu\tau)^{1/2}}\,\mathcal{M}_\beta\left(\frac{\tau}{t^\alpha}\right)\frac{d\tau}{t^\alpha} \\ &= \int_0^\infty \prod_{i=1}^{n} \frac{\mathrm{e}^{-x_i^2/(4\nu\lambda t^\alpha)}}{(4\pi\nu\lambda t^\alpha)^{1/2}}\,\mathcal{M}_\beta(\lambda)\,d\lambda \\ &= \frac{1}{(\nu t^\alpha)^{n/2}}\,\mathbb{M}_{\beta/2}\left(\frac{\|\mathbf{x}\|}{\sqrt{\nu t^\alpha}}\right)\,, \end{aligned} \tag{3.29}$$

where the notation $\mathbb{M}_\eta(z)$ is used for the multi-dimensional extension of the $\mathcal{M}$-Wright/Mainardi function $\mathcal{M}_\eta(z)$.

As the reader had seen in the previous derivation, the marginal *pdf* of the non-Markovian diffusion process ggBm emerges to be related to the Mainardi function M_ν and the shape of the pdf, and its possible resemblance with a Gaussian, is managed by the means of this versatile yet quite recent mathematical tool; another thing worth to be noticed is that in this Time Stretched Fractional Diffusion framework, (or, talking in terms of the stochastic particle dispersion, in the ggBm framework) let the model distinguish between the shape of the PDF and the sub- super diffusive regime, acting on the parameters β and α, respectively. The fact that the diffusive regimes (wheter it happens to be super or sub-diffusive) is up to the α parameter is highlighted by the particle variance σ_d^2 which in one–, two– and three–dimensions ($d = 1, 2, 3$) is [154, 108]:

$$\sigma_1^2 = \int_{-\infty}^{+\infty} x^2\,p_1(x;t)\,dx = \frac{2\nu\,t^\alpha}{\Gamma(\beta+1)}, \qquad \sigma_2^2 = 2\,\sigma_1^2, \qquad \sigma_3^2 = 3\,\sigma_1^2. \tag{3.30}$$

Differently from the analysis performed in [2], the PDF stated in (3.29) allows for studying independently the spreading (driven by α) and the shape of the PDF (driven by β).

here the diffusive regime and the gaussianity are well distinguished and it is in principle possible to simulate even the case of a linear variance law, but with non-Gaussian PDF, that is the case when $\beta \neq \alpha = 1$ and a Gaussian PDF with non linear variance when $\beta = 1$ and $\alpha \neq 1$.

A microscopical insight

Erdélyi–Kober fractional diffusion equation is related to the class of stochastic processes called *generalized grey Brownian motion* (ggBm) [208]. In fact, the one-point one-time particle PDF of the ggBm is the Green function of the Erdélyi–Kober fractional diffusion equation.

We recall that the ggBm [187, 62] is a special class of *H-self-similar-stationary-increments* (H-sssi) of order $H = \alpha$, or Hurst exponent $H = \alpha/2$. The ggBm provides non-Markovian stochastic models for anomalous diffusion, and it can generate sub- $(0 < \alpha < 1)$ and super- $(1 < \alpha < 2)$ diffusion processes. Special cases of the ggBm are the fractional Brownian motion $(\beta = 1)$, the grey Brownian motion $(0 < \alpha = \beta < 1)$ and the standard Brownian motion $(\alpha = \beta = 1)$.

Stochastic trajectories of the ggBm can be generated by the product of a fBm with Hurst exponent H times and positive constant independent random variable ℓ, i.e.,

$$X = \ell\, X_H\,. \tag{3.31}$$

The ggBm considered in Ref. [187] is established by imposing $H = \alpha/2$ and $\Lambda = \ell^2$ be distributed according to $\mathcal{M}_\eta$.

This approach has been recently propose to model diffusion in biological cells [182], but also to derive the stochastic solution of the space-time fractional diffusion equation [210] or two-particle anomalous diffusion [209], with the proper distribution of Λ.

3.3 Quenching

Depending on the physical interpretation of the two evolving unknowns of the model,$\varphi(\mathbf{x}, t)$ and $\varphi_e(\mathbf{x}, t)$, it may be useful to provide a description of the possible cases when one of the two domains, the one embedded by one (the zero - level one, for the sake of simplicity) iso-contour of φ or the other by φ_{eff}, eventually shrinks to one point and vanish.

The proposed model could thus in principle lead to

1. A *Geometrical Quenching*, due to the curvature term in the equation (3.6), that involves the deterministic front computed by the means of φ and does not receive any feedback by the effective front, being moreover independent of any choice of the probability density function;

2. A *Diffusive quenching*, due to the dynamics of the effective front evolution law (3.4), which depends on the choice of the probability density function and is moreover closely related to the deterministic front advancing.

3.3.1 Geometrical quenching

The Geometrical Quenching, in the formulation of equation (3.6) , is due to the curvature term $\lambda\kappa$: we remind the reader that in our formulation the normal velocity is always pointing outwards and cannot be responsible for such a recoiling of the front. To be more specific, the Level Set equation (3.6) sees in its curvature operator a dissipative term , as it is pointed out in many theoretical works on front propagation (see e.g. Ref. [101]). We also remind that in the case of null normal velocity term, such equation becomes

$$\varphi(x,t)_t = \xi\kappa(\varphi(x,t))\,|\nabla\varphi(x,t)| = \xi\nabla\cdot\left(\frac{\nabla\varphi(x,t)}{|\nabla\varphi(x,t)|}\right)|\nabla\varphi(x,t)| \tag{3.32}$$

Equation (3.32) gives rise to a geometric flow called *mean curvature flow*, that in its one-dimensional characterization (that is, regarding the evolution of a closed curve in a plane) has been called in the literature *curve shortening flow*, see again [101]. Each surface (with suitable hypothesis of regularity) or curve will shrink to a sphere (or a circle) and then collapse into a single point. This situation shall be regarded as a limiting case for the proposed model, i.e. for the case when

$$|\mathcal{V}(x,t)| \to 0. \tag{3.33}$$

This behavior is portrayed in Figure 3.2c.

Apart from the limiting case of (3.33) , the intermediate case of the Level Set equation shall be discussed making use of its *convection diffusion formulation* valid as long as the condition $|\nabla\varphi(x,t)| = 1$ is enforced :

$$\varphi(x,t)_t + \mathbf{V}\cdot\nabla\varphi(x,t) = \xi\nabla^2\varphi(x,t) \tag{3.34}$$

with $\mathbf{V} = \mathcal{V}\mathbf{n}$.

This formulation of the curvature / convection driven motion of the deterministic front suggests, for a fixed initial condition of the unitary gradient level set equation , an interplay between the two quantities $\mathcal{V}$ and ξ.

In the case $\mathcal{V}/\xi \ll 1$ can be observed a dissipative kinetic that takes over the entire process and drive inwards the level set curves, causing the deterministic front to extinguish. This case describes of course even the previous limiting case of curvature driven flow. On the other hand, when $\mathcal{V}/\xi \gg 1$ we have the deterministic front that will advance with the area of the enclosed domain augmenting with respect to time. The case without the effects of curvature considered (i.e. $\xi = 0$) falls obviously in this case.

The authors have seen a striking analogy of this dynamic with the results of [202], where with Lagrangian arguments this two-parameter behaviour has been depicted even though with a totally different physical interpretation.

3.3.2 Diffusive quenching

Shifting the attention towards the diffusive quenching, a thorough description (regarding the model that neglects curvature effects) can be found in [2]. We shall nonetheless remark the fact that due to different levels of the diffusion parameter ν (and adopting different threshold values φ_{th}) the quenching can either take place or not, and that a quenching of the effective front can be followed by a reappearance of the front due to diffusive phenomena, with the latter event that can acquire some sense depending on the physical interpretation of the model. For example, in the context of the modeling of turbulent premixed combustion, if the effective front modeled by the $\varphi_{th}-$ isocontour of the field $\varphi_e(\mathbf{x};\mathbf{t})$ vanishes, the corresponding premixed flame can be considered extinguished and the system can be considered from that moment on quiescent.

3.3.3 Feedback mechanisms

The remarkable fact that a front propagation phenomenon in a complex medium is described through the study of two different fields φ and φ_e , gives credit to a straightforward question: how should the two fields interact with each other? The first interaction worth noting is that the deterministic front, can itself exhibit a shrinking behavior and in this case the effective front (built up as an integral of positive quantities on the domain given by the deterministic advancing, as shown in equation (3.3)) will eventually extinguish itself, this time without any chance of coming back. The inclusion of the curvature term in the deterministic front evolution law allowed this one-way feedback between the two descriptions of the moving interface.

The other way round, that is the influence of φ_e on the level-set field φ is, in the framework of the model depicted by this paper, not observable. Nonetheless, there are works where this link has been successfully established and thoroughly studied, such as [203] and [205]. In the latter works, an auxiliary function it is utilised in order to change at a time τ the value of (3.1) from 0 to 1 according to the behaviour of $\varphi(x;t)$ in the whole interval of time $[t_0,\tau]$. Such methodology has been applied in the modelling of wild-land fire spread, the probability distribution function representing in that case the probability of burning embers fall; the topic is covered in Chapter 5.

3.4 Numerical results

In this Section the profiles of the isocontours of φ and φ_e defined in the previous sections are presented and discussed. It has to be reminded that the profile of the isocontours of φ_e are being computed by using formula (3.3) depending on the level set equation (3.5), and that the PDF $p_d(\mathbf{x};t|\overline{\mathbf{x}}) = p_d(\mathbf{x}-\overline{\mathbf{x}};t)$ is the fundamental solution of the d-dimensional stretched time-fractional diffusion given in (3.29).

3.4.1 Insight on the numerical tools and recipes

The motion of the fronts that propagate in random diffusive and sub-diffusive media, under the effect of normal velocity and curvature, has been computed by the means of a software package developed ad-hoc for this goal, that is actually an evolution of the one described in [2]. The latter code, written in C and Fortran2008 utilizes a library (`liblsm90`) which is capable of computing the evolution of a co-dimension one front in one, two and three dimensions, with the use of robust finite-difference algorithms (belonging to the family of Weighted ENO schemes) and a Runge-Kutta scheme for the time-advancing. In order to make advantage of the parallel architecture were the code has been run we adopted an OpenMP parallelization paradigm.

All the results that will be shown in the following part are the result of a post-processing work done by the means of open-source software as `SciPy` [122] and `Matplotlib` [116], in the IPython framework [212].

The computations have been performed by the means of the computational server HIPATIA, made available at the Basque Center for Applied Mathematics (BCAM) in Bilbao, Basque Country – Spain.

3.4.2 Geometrical asset

To reproduce the effects of motion under curvature and normal velocity subject to random fractional fluctuations, a series of numerical experiments are conducted. To show the different behavior depending on the geometry of the front, a particular initial front $\varphi^*(t=0)$ is adopted. Such initial front is given by a circle from which a series of rectangles have been carved only in the upper half. The motion will suddenly smear out the sharp angles and the front will exhibit a low uniform curvature in the lower half and a heterogeneous curvature in the upper one.

For the sake of establishing a suitable initial condition, and for its good numerical proprieties (again, see [252]) the adopted level set is a signed distance function, i.e. a function that for each point of space will take the value of its distance from the interface. See Figure 3.1 for an example of model run when φ is a signed distance function.

The signed distance function ([252]) for the initial circle, of radius r and centered in $\mathbf{x}_0$ has been obtained by the well known formulation

$$\varphi C(x) = ||\mathbf{x} - \mathbf{x}_0|| - r \tag{3.35}$$

The rectangles we have implemented have the signed distance function

$$\varphi R(x) = \begin{cases} max(|x - x_C| - l, |y - y_C| - w) & \text{if } \mathbf{x} = (x, y) \in \Omega^+ \\ min(|x - x_C| - l, |y - y_C| - w) & \text{if } \mathbf{x} = (x, y) \in \Omega^- \end{cases} \tag{3.36}$$

where in the last equation l and w are the semi axis of the rectangle, the symbol Ω^+ denotes the exterior of the domain while Ω^- its interior, and (x_C, y_C) is the center of the rectangle. The carving

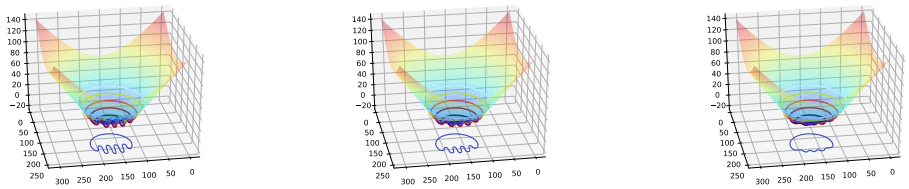


Figure 3.1: Three different time steps of the evolution of the iso contour $\varphi = 0$ with the complex initial geometry introduced in this Section. The colored surface is the level set field φ while the iso-contour of interest is projected in the plane $z = -20$ for clarity purposes.

of the circle has been performed by using the set subtraction

$$\varphi(x) = max(\varphi(x), -\varphi RT(x)) \tag{3.37}$$

where φRT is the set theoretic sum of the rectangles, that is, using the signed distance function algebra (as in [195])

$$\varphi RT(x) = min(\varphi R_1(x), ..., \varphi R_n(x)) \tag{3.38}$$

Since the signed distance function embedding was adopted for the reconstruction of the advecting front, to maintain numerically the analytics properties of the latter a reinitialization procedure had to be undertaken. A simple time dependent approach was adopted ; i.e. the steady solution of

$$\frac{\partial \phi}{\partial t} + sgn(\phi)\hat{\mathbf{n}} \cdot \nabla \phi \tag{3.39}$$

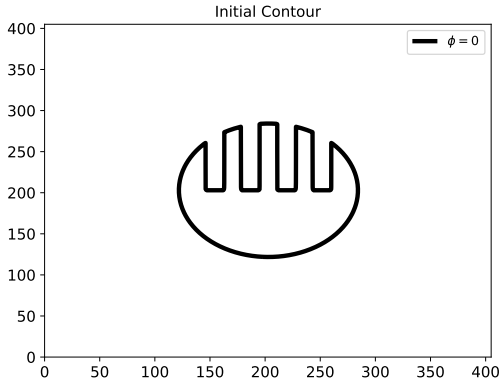
has been sought.

The interval of time steps between each reinitialization procedure was not fixed, since in the initial phases of the evolution there were sharp edges that required extra attention in preserving the signed distance function structure of the scheme, and so a reinitialization had to be performed each time step. In the later part of the simulation, when the geometrical structure of the front was generally more easy to deal with, the reinitialization could be performed each 20 time steps without degrading dramatically the numerical accuracy.

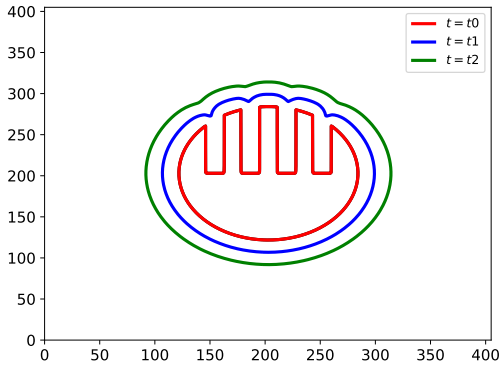
The computational domain was a square box, with side of $1m$, and the radius of the initial circle was $0.2m$. For the undertaken simulations, 4 equally spaced rectangles have been carved out, starting from the horizontal diameter of the circle. The horizontal length of each rectangle was $0.2 \times R$, with R the radius of initial circle.

3.4.3 Parameters and ranges

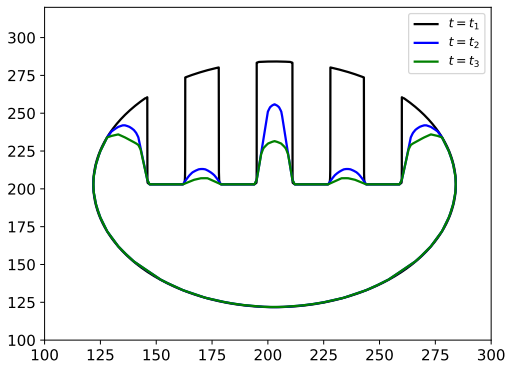
Equation (3.29) as well as (3.6) regulate the motion of the front, and do depend actually on various parameters, related to microscopical and macroscopical features. This dynamic is prone to exhibit



(a)



(b)



(c)

Figure 3.2: Setting for the example of model application. Figure 3.2a shows the initial contour $\varphi(\boldsymbol{x}, t = 0) = 0$. Figure 3.2b shows the loss of spatial information on the front $\varphi(\boldsymbol{x}, t = t_{1,2,3}) = 0$ due to a kinetics driven only by the normal velocity. Figure 3.2c shows the shrinking of a contour when its LSM equation is driven just by the curvature term, that is when Equation 3.32 is enforced.

very complex behaviour for it depends at least on the subsequent parameters, summed up for the sake of completeness:

- Normal velocity of the macroscopic front, V_0 ;
- Markstein Length related to the curvature effect
- Parameter of the diffusion of space/time fractional diffusion, ν ;
- Parameter β of the Mainardi Function (that accounts for the Gaussianity of the stochastic front);
- Parameter α of the time exponent (that governs super and sub diffusivity).

When the computation of φ_e has been performed, all the possibilities have been tried among the combinations of the subsequent parameters:

- $V_0 \in \{0.01, 0.1\}[\frac{m}{s}]$
- $\lambda \in \{0.05, 0.1\}[\frac{cm}{s}]$
- $\nu \in \{0.001, 0.01\}[\frac{m^2}{s}]$
- $\alpha \in \{0.5, 1.0, 1.5\}$
- $\beta \in \{0.1, 0.3, 0.5\}$

3.4.4 A glimpse at the kinetic

In order to understand the plots of the numerical results of the simulation, we remark a few characteristics of the motion that our surface have exhibited:

1. The normal velocity in the lower part of the hand-shaped contour will just advect the $\varphi = 0$ iso-contour along a radial direction. In the upper part, it would do the same for the outer part of the fingers. The inner sides of the fingers will be advected by the normal velocity one towards the other, and the bigger the value of V_0, the less the deterministic shape of the hand will mantain its carachteristic form before collapsing into a circle. See Figure 3.2b for an example of a purely normal-driven flow that converges to an expanding circle.

2. The curvature effect (governed in this case by λ will contribute (as remarked in the previous Section) to smear out the dis-uniformities of the initial shape, contributing to making it converge to a circle. Its effect is of paramount importance in the upper part of the hand-shaped profile, since the radius of curvature will be significantly smaller in the corners of the fingers, while in the lower side the radius of curvature the radius of curvature will be large and uniform. We did not appreciate in the explored range of Markstein numbers an evident change in the kinetics of the front, and therefore only the plots for $\lambda = 0.1$ are presented. In Figure 3.2c a shrinking profile for $\varphi = 0$ under a motion regulated by Equation 3.32 is presented.

3. The diffusion coefficient ν of the PDF will regulate the sharpness of the φ_e front (the higher ν, the less sharp the ϕ_e profile). Following the results of [2] , it will play an important role on the eventual shrinking of the $\varphi_e = \varphi_{th}$ iso-contour we are modelling. We remark the limit

$$\lim_{\nu \to 0} p(\mathbf{x}; t) = \delta(\mathbf{x}; t) \tag{3.40}$$

that gives an intuitive explaining to how the iso-contours of ϕ_e stick to the deterministic front when ν is low.

4. The cofficient β shifts the motion to sub-diffusive to diffusive. This is quite evident from the plots, since the contour of a low beta diffuses less and stick more to the LSM profile; this is evident also for the quenching regime captured by the choice of parameters $\nu = 100$ and $V_0 = 1$, when in the sub-diffusive regime the shape of the front is more similar to LSM than in the $\beta = 0.5$ case, thus limiting the quenching towards a vanishing circle.

5. The coefficient α exhibits the particular capability to govern in some sense the shape memory of φ_e: in each of the right hand side columns, the central figure of the left hand side column has been proposed again, this time varying the coefficient α. In this way it can be appreciated how the profiles are more close to the deterministic LSM contour as α increases. This is particularly evident e.g. in the plot with $\nu = 100$ and $V_0 = 1$, where the $\beta = 0.5$ iso-contour surpasses the black-colored line of the standard Gaussian φ_e contour when α passes from 0.5 to 1.5.

Table 3.1: Left column, from top to bottom: contours corresponding to $D = 10m^2s^{-1}, V_0 = 1m/s, \alpha = 1$, $\beta = 0.1$ (red), 0.3 (blue) ,0.5 (green), at $t = 0.3, 0.6, 0.9s$ respectively. Right column is the same as the central plot of left column but for increasing α, from top to bottom $\alpha = 0.5, 1, 1.5$, in order to show effect of changing α. The corresponding plots with Gaussian PDF are outlined by black solid contours; the ones corresponding with plain LSM advancing (PDF $= \delta(\boldsymbol{x}, t)$ are represented by pink dash-dotted contours.

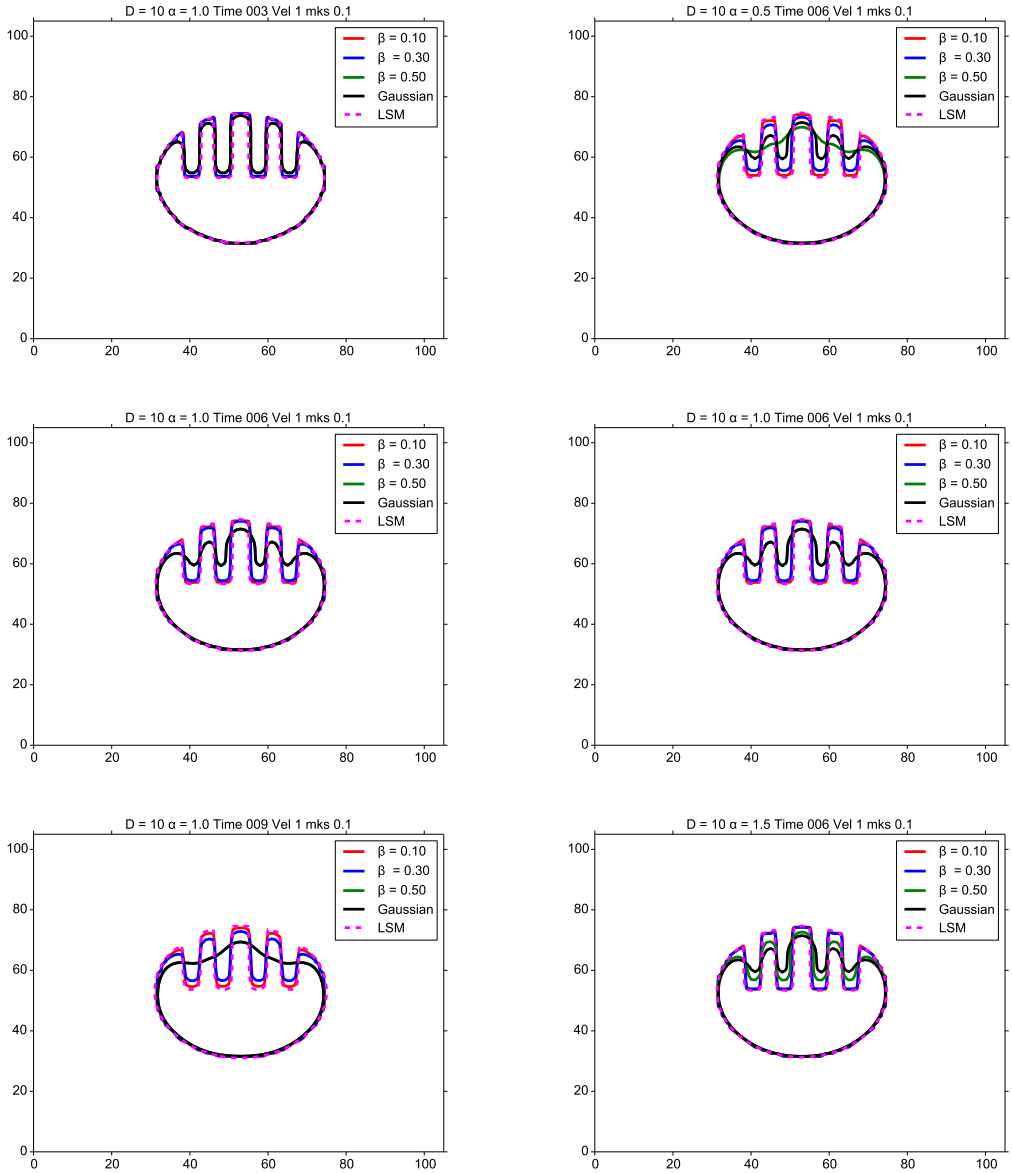


Table 3.2: Left column, from top to bottom: contours corresponding to $D = 10m^2s^{-1}, V_0 = 10m/s, \alpha = 1$, $\beta = 0.1$ (red), 0.3 (blue) ,0.5 (green), at $t = 0.3, 0.6, 0.9s$ respectively. Right column is the same as the central plot of left column but for increasing α, from top to bottom $\alpha = 0.5, 1, 1.5$, in order to show effect of changing α. The corresponding plots with Gaussian PDF are outlined by black solid contours; the ones corresponding with plain LSM advancing (PDF = $\delta(\boldsymbol{x}, t)$ are represented by pink dash-dotted contours.

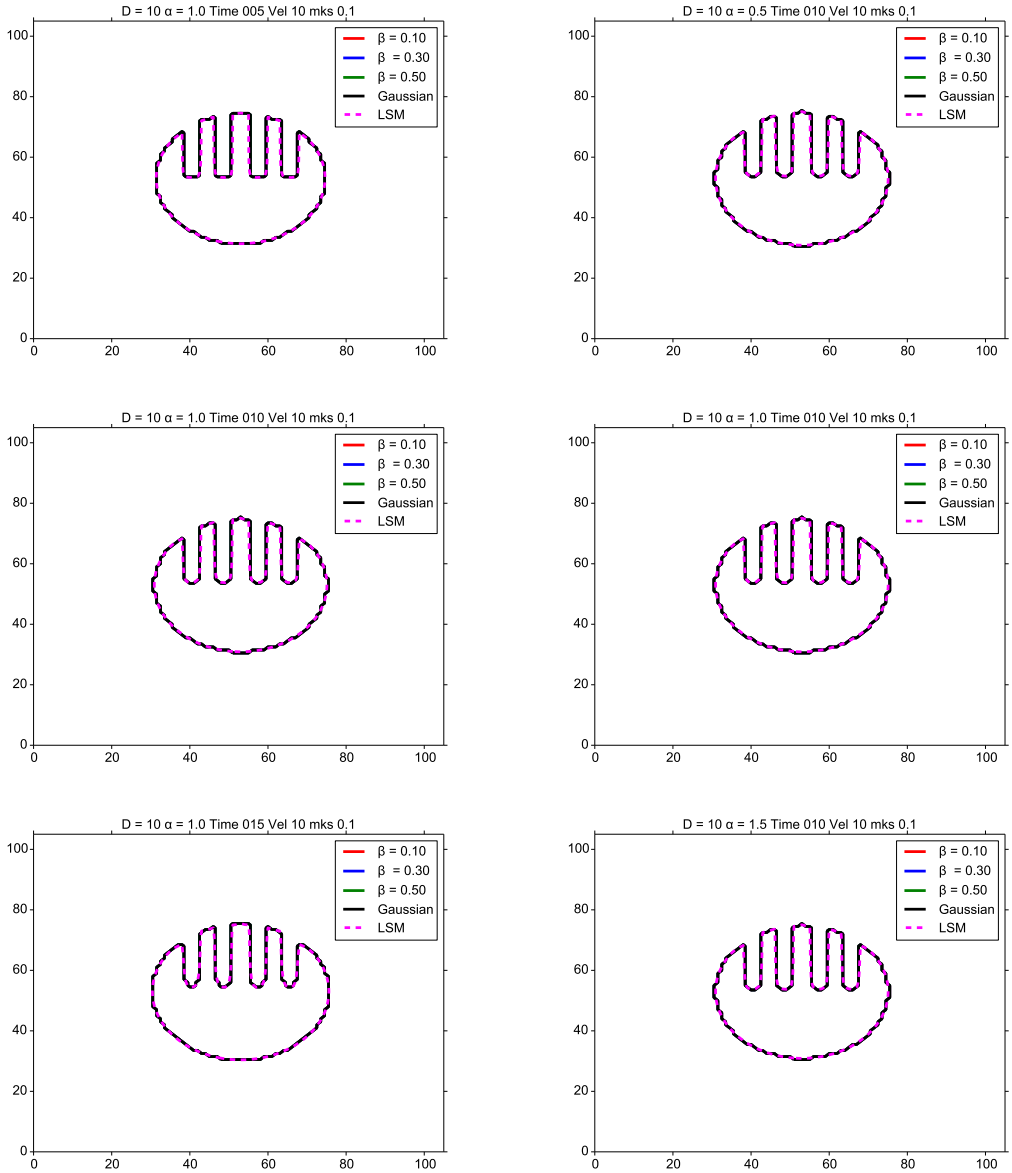


Table 3.3: Left column, from top to bottom: contours corresponding to $D = 100m^2s^{-1}, V_0 = 1m/s, \alpha = 1$, $\beta = 0.1$ (red), 0.3 (blue) ,0.5 (green), at $t = 0.3, 0.6, 0.9s$ respectively. Right column is the same as the central plot of left column but for increasing α, from top to bottom $\alpha = 0.5, 1, 1.5$, in order to show effect of changing α. The corresponding plots with Gaussian PDF are outlined by black solid contours; the ones corresponding with plain LSM advancing (PDF $= \delta(\boldsymbol{x}, t)$ are represented by pink dash-dotted contours.

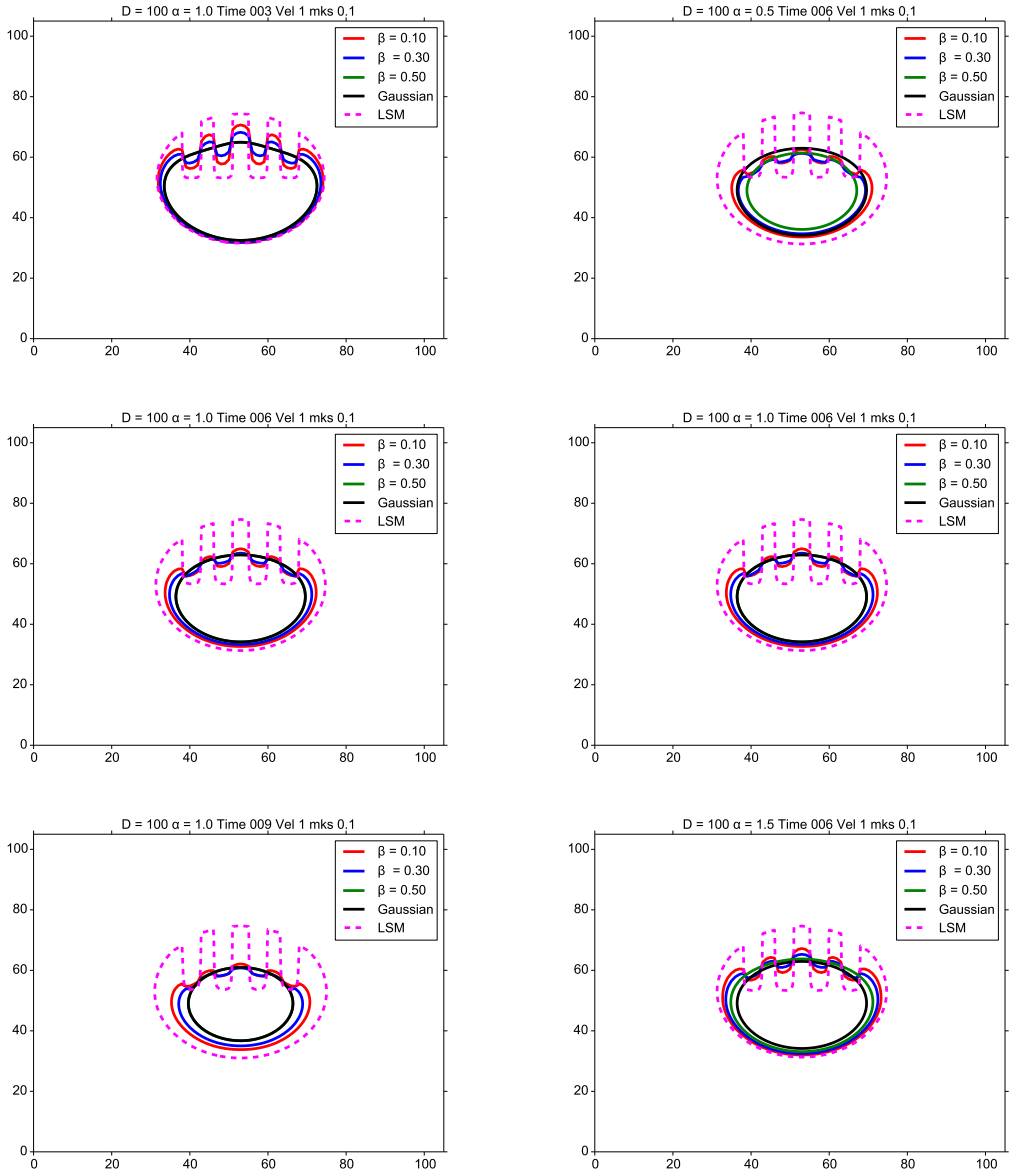
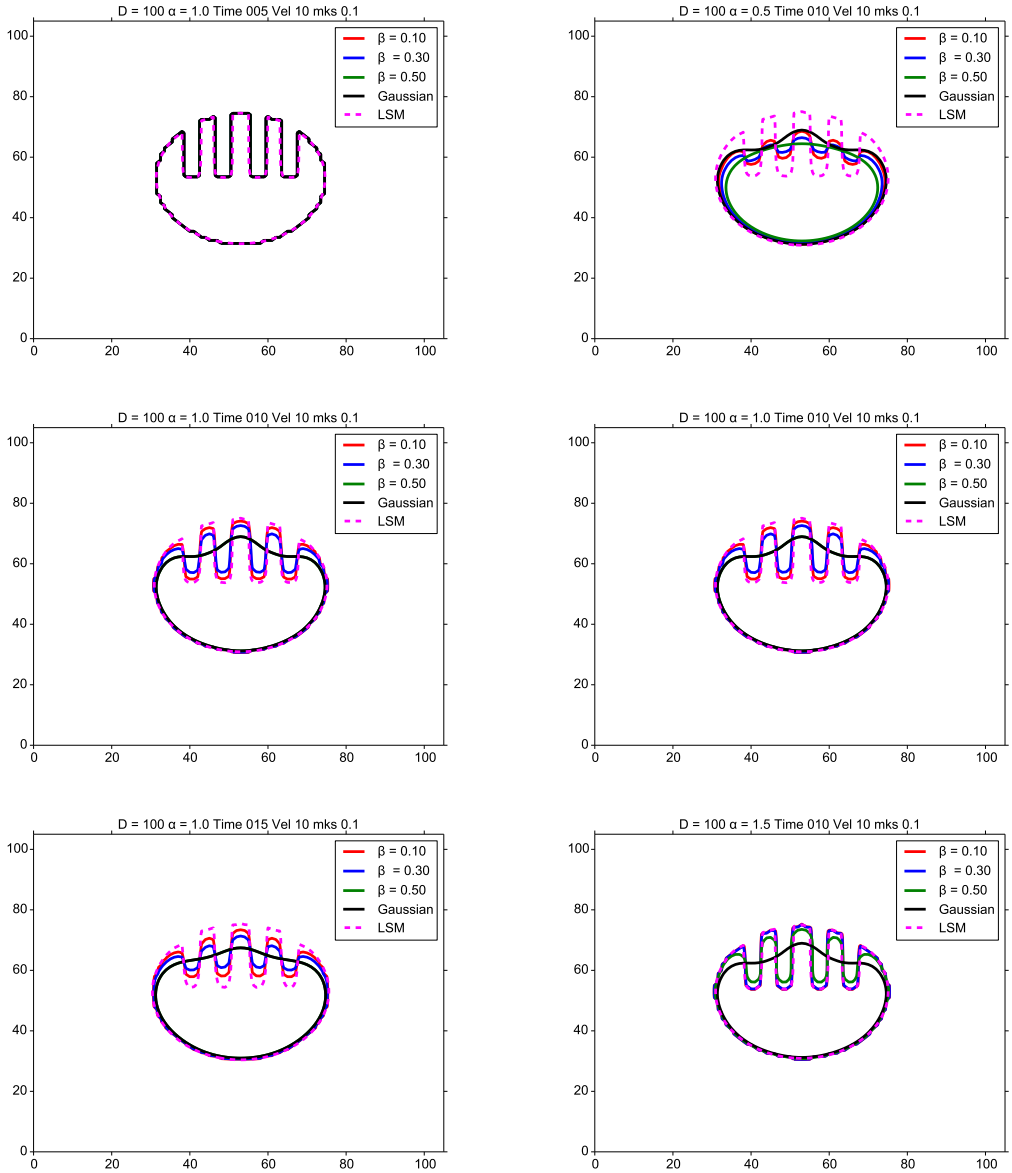


Table 3.4: Left column, from top to bottom: contours corresponding to $D = 100m^2s^{-1}, V_0 = 10m/s, \alpha = 1$, $\beta = 0.1$ (red), 0.3 (blue) ,0.5 (green), at $t = 0.3, 0.6, 0.9s$ respectively. Right column is the same as the central plot of left column but for increasing α, from top to bottom $\alpha = 0.5, 1, 1.5$, in order to show effect of changing α. The corresponding plots with Gaussian PDF are outlined by black solid contours; the ones corresponding with plain LSM advancing (PDF $= \delta(\boldsymbol{x}, t)$ are represented by pink dash-dotted contours.



3.5 Surrogate Modeling of the problem

The example of model application of Section 3.2 is studied via some of the UQ and SA techniques illustrated in Chapter 2, In particular,the approach of Variance-Based Sensitivity Analysis (with Sobol' indices) is chosen for SA. Due to the limited computational budget, that make a

Montecarlo-based analysis prohibitive, a surrogate model is constructed via the means of Polynomial Chaos with LAR strategy (see Section 2.4.2 of Chapter 2). This example of application has two objectives. The first is the ranking between the two parameters α, β of the PDF of random fluctuation in complex media and given in Equation 3.29, under complex initial geometry. The other objective is the quantification on the uncertainty in the output of such complex kinetic given uncertainty in the two parameters α, β that characterize the medium, governing diffusive regime and the gaussianity respectively.

3.5.1 Parameters, databases and observables

For this analysis, the random input parameters is defined as $\boldsymbol{\theta} = \{\alpha, \beta\}$.
To see the effect of the parameters range of variation on the output shape, we need to think of some kind of observable to better quantify the effect of the parameter choice. The following quantities of interest (QoI) are chosen:

1. The first one is the Area inside of the statistical contour, i.e. inside of a selected iso-contour of the smooth indicator function φ_e. Setting the threshold value $\varphi_{th} = 0.5$, the first QoI reads
$$Y_1 = \int_{\varphi(\mathbf{x})>\varphi_{th}} \mathbf{1} d\mathbf{x} \tag{3.41}$$
2. The second QoI is related to the shape of the figure. N_M couples of markers are put on the $\varphi 0.5$ iso-contour and the $\gamma 0$ isocontour so that at time $t = 0$ they coincide. For $t > 0$, the two iso-contour shall exhibit some difference each couple of marker is then characterized by their euclidean distance d_i; That difference encodes the effect of diffusion on the macroscopic shape configuration. See Figure for a representation of this setting. For the limiting case $p(\mathbf{x}; t) = \delta(\mathbf{x}; t)$ we get to the outcome $d_i = 0 \ \forall i$. The second QoI is then defined as
$$Y_2 = \frac{\sum_{i=1}^{N_M} d_i}{N_M} \tag{3.42}$$

The values Y_i may depend on time. We analyze two different time steps $Y_i(t = \{t_1, t_2\})$ in the following.
For this purpose, two databases of $N_{DOE} = 216$ are compiled with Halton sampling, one with the information of Y_1, Y_2 at time $t = t_1$, while the other at $t = t_2$. A scatter plot of a database used for this exercise is portrayed in Figure 3.4.
A schema of the adopted work-flow for a non-intrusive surrogate analysis is outlined below, in line with what reported in Chapter 2.
For $Y_i \in \{Y_1, Y_2\}$:

1. Determine the PDF for each variable. By using the support of those PDFs construct the hypercube
$$\mathbb{H} \equiv [\theta_{1_{min}}, \theta_{1_{max}}] \times [\theta_{2_{min}}, \theta_{2_{max}}] \tag{3.43}$$

2. Sample Design Of Experiment (DOE) of $N_{DOE} = 216$ samples $\boldsymbol{\theta}_k = \{\theta_1, \theta_2\}_i \in \mathbb{H}$. This sampling used the PDFs of each parameter $\theta_i \in \boldsymbol{\theta}$. In our specific case, the Halton Method, a low-discrepancy scheme, is employed.

3. Run the simulation for each choice of parameters $1 \dots N_{DOE}$ and collect the QoIs Y_i into the training database $\mathcal{D}_{N_{DOE}}$.

4. Compute the surrogate model $\mathcal{M}$ such that $\mathcal{M}(\boldsymbol{\theta}_i^*) = \widehat{Y^*}_i$ is the surrogate representation of the value Y_i associated at the evolution of the analyzed system subject to the choice $\boldsymbol{\theta}_i^*$ of the parameters. The adopted surrogate model is gPC with LAR projection strategy, see Chapter 2.

5. With the computationally cheap surrogate model perform Sensitivity Analysis and Uncertainty Quantification obtaining Partial and Total Sobol Indices, PdF and moments of the QoI Y_i.

6. Perform convergence and error analysis with some error estimator. For this specific application we chose LOO indicator.

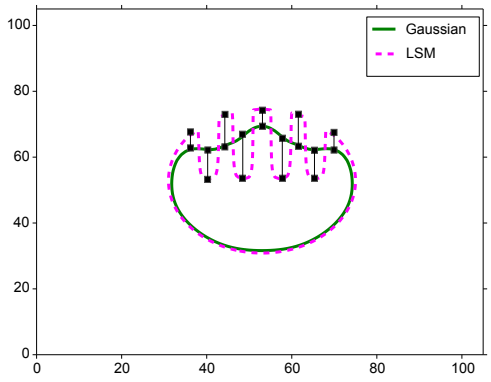


Figure 3.3: Representation of the Markers used to compute observable Y_2. The pink dashed contour is the LSM contour, in our case the iso-contour $\varphi = 0$. The green contour is a realization of the process with selected α, β (in the picture $(\alpha, \beta) = (1, 0.5)$. In particular, the green contour corresponds thus to the iso-contour $\varphi_e = 0.5$.

Run settings

Since we do not want to characterize a specific medium or regime of front propagation, we span the hypercube $\mathbb{H}$ with Uniform Distributions. Those distributions will be used by the Halton Sampling of the DOE and by the Polynomial Basis construction for the PC routines.
The two analyzed times are $(t_1, t_2) = (0.2, 0.4)$ While $\boldsymbol{\theta} = (\alpha, \beta)$ is varying, we fix
$D = 100, V_0 = 0.2, \lambda = 0.1$.

Input Parameter	Distribution
α	$\mathcal{U}(0, 2)$
β	$\mathcal{U}(0, 0.5)$

3.5.2 Surrogate Model Results

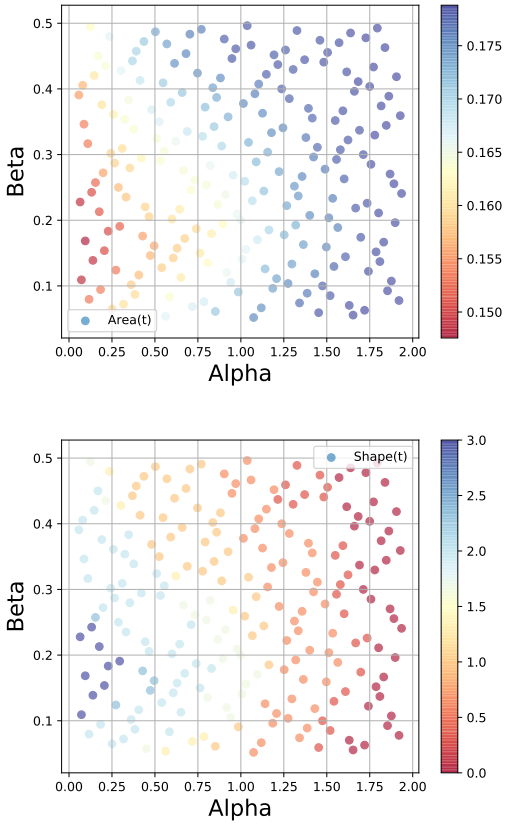
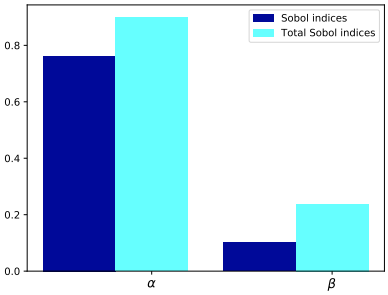
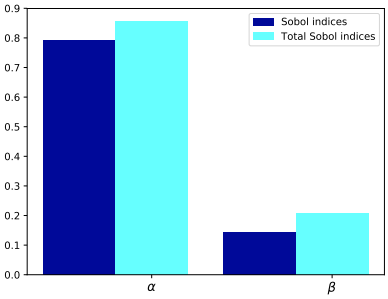


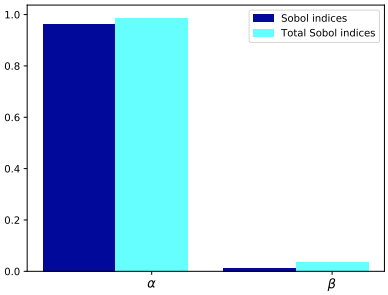
Figure 3.4: Panel: DOE representation of the $\boldsymbol{\theta} = \{\alpha, \beta\}$ case for Y_1 and Y_2.



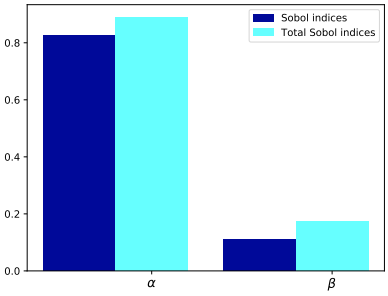
(a)



(b)



(c)



(d)

Figure 3.5: Panel: Y_1 (covered area), Sobol Indices at two different times (rows), $t_1 = 0.2$ (a), $t_2 = 0.4$ (b) ; Y_2 (shape discrepancy), Sobol Indices at two different times (rows), $t_1 = 0.2$ (c) , $t_2 = 0.4$ (d).

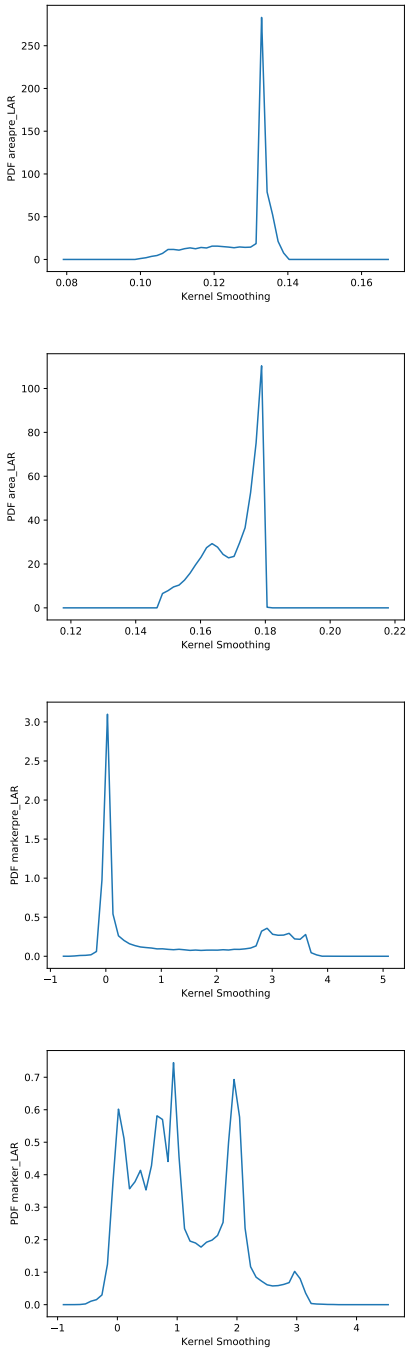


Figure 3.6: Panel: Y_1 (covered area), PDF at two different times (rows), $t_1 = 0.2$ (a), $t_2 = 0.4$ (b) ; Y_2 (shape discrepancy), PDF at two different times (rows), $t_1 = 0.2$ (c) , $t_2 = 0.4$ (d).

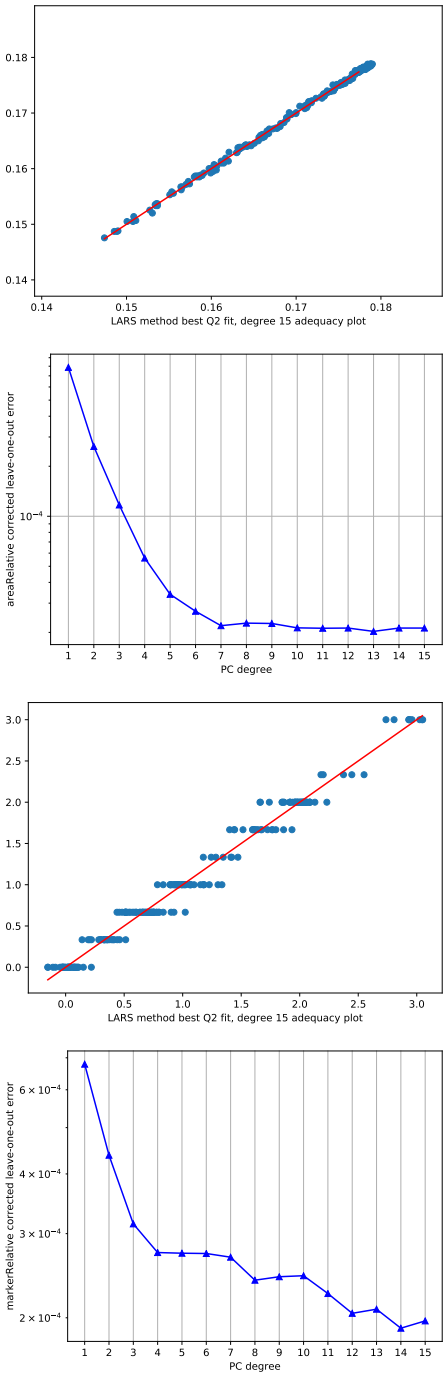


Figure 3.7: Panel: Y_1 (covered area) at time $T2 = 0.4$, adequacy plot for empirical error (a) and LOO error versus maximum degree of employed Polynomial Chaos (b) ; Y_2 (shape discrepancy) at time $T2 = 0.4$, adequacy plot for empirical error (c) and LOO error versus maximum degree of employed Polynomial Chaos (d) ;

3.5.3 Discussion of Surrogate Model results

The surrogate model of the test case 3.2, given by gPC with LAR regression strategy, gives out interesting results concerning both SA and UQ. To begin with, we have from Figure 3.5 the First order and total order Sobol' indices for both Y_1 at time t_1, t_2 (Fig. 3.5a,3.5b) and Y_2 at time t_1, t_2 (Fig. 3.5c,3.5d).

α, that governs the Hurst exponent of the underlying fluctuating dynamic, is the leading term for both observables Y_1, Y_2. However, the effects of Beta are not entirely negligible. Since first order Sobol' index (dark blue) is different from total Sobol index (light blue), the two parameters do interact one with each other in determining change in the covered area (Y_1) or shape modifications (Y_2). We shall now shift the attention to Figure 3.6, which portrays the different PDF of the observables Y_i at the two analyzed times. The PDF for Y_1 shows the fact that the contour given by LSM (that corresponds to the statistical contour given by $p = \delta(\boldsymbol{x}, t)$ cannot be trespassed. In fact, a barrier-like peak can be observerd at the right of the PDF for both analyzed times t_1, t_2. This can deliver the message that in a random medium characterized by unknown Hurst exponent (and unknown particle displacement variance) and unknown Gaussianity in the shape of the particle PDF, an upper estimate over the covered the area may be furnished. On the other hand, shape discrepancy (last two figures of the panel of Figure3.6) is way more noisy; a somewhat bimodal trend can be discerned. This can mean that the statistical contour $\varphi_e = 0.5$ can stick to LSM (the peak on 0) or either provoke a remarkable shape loss (peak over 3, on the plot relative to $t = t_1$). The gPC with LAR projection strategy has been applied and its trustworthiness tested via error metrics. LOO error and Empirical error behaved well.

As soon go kindle fire with snow, as seek
to quench the fire of love with words.

W. Shakespeare

Chapter 4

Applications to Turbulent premixed combustion

Turbulent premixed combustion (TPC) is a vast scientific field involving nonequilibrium phenomena, which plays a role of paramount importance in important industrial issues such as engine design and energy production. TPC constitutes a challenging scientific field which involves nonequilibrium phenomena. It requires a set of governing equations and a rich phenomenology follows. The set of governing equation includes: mass and momentum conservation, equation of state for gases, energy and species conservation. The nonlinearity of the problem requires equation closures and modeling. While a description of the physical phenomena involved, the mathematical formulation and the computational state of art to study TPC is beyond the scope of this Chapter, the application of the proposed modeling framework described in Chapter 3 is here summarized. We refer the interested reader to [147] for a comprehensive treatise of TPC under a physical and engineering perspective.

The application to TPC that is described in this Chapter is a particular case of a more general framework, already discussed in Chapter 3, about reacting fronts in random environments whose formulation can be applied to processes in which a closed interface netly separates what is inside from what is outside, TPC being of course one of the processes to which such formulation can be applied. With respect to TPC, the main aim is to provide novel methods able to lead to new theoretical insights in combustion science. The derivation of such framework is sparse in a number of papers and formulated under different point of views relative to the different considered application [202, 204, 203, 206]. The present formulation is intended for the family of processes that are studied in literature by using, for example, the so-called level-set method (LSM) [250], discussed in Chapter 1. We recall that LSM has been adopted successfully in many different applications. wildland fire propagation [157], groundwater infiltration [121], biology [150], material science [104] and TPC [147] (for solving the so-called G-equation).

As discussed in Chapter 3, the level-set method and reaction-diffusion equations can be considered

approaches that are alternatives to each other because the solution of reaction-diffusion equations is generally a continuous smooth function that has an exponential decay, and it is non-zero in an infinite domain, while the level-set method provides a sharp function that is non-zero on a compact domain. However, one of the main results of the present formulation is that these two approaches are indeed complementary and they can be reconciled. Adopting the TPC terminology, we start from the G-equation that describes a level surface propagating with a random motion, we show that it is possible to derive a reaction-diffusion equation that describes the effective volume of the reacted mixture fraction under the assumption that the probability density function (PDF) of the random diffusive process underlying the front motion is known.

Thus, the evolution equation for the resulting observable can be reduced to equations for the progress variable by choosing the PDF of the front fluctuations. This approach reduces to the G-equation when there are no fluctuations [202] and to the Zimont equation for a planar flame and Gaussian fluctuations [320]. This promising approach has been directed lately towards the challenging problem of hydrodynamic instabilities described by the MS equation [253, 180].

The G-equation represents the motion of an iso-surface that is assumed to be the burned/unburned interface and its propagation can be assumed to be linked with the flame propagation. The G-equation provides the motion of an iso-surface governed by a flow field and by a reaction speed in the outward normal direction to the interface. The flow field is generally governed by the hydrodynamical equations providing a random velocity field. As a consequence of this random velocity field the position of the iso-surface turns to be random, too. After averaging the random locations of the iso-surface, the effective front profile emerges to be governed by a reaction-diffusion equation. This reaction-diffusion equation is different for different PDFs of the random fluctuations of the front positions.

In Section XX, We provide the density function of the fluctuations of the front positions such that the corresponding reaction-diffusion equation is the Michelson Sivashinsky equation. That would allow the derivation of the MS equation as the average of random fronts propagating according to their G-equation. Anyway, it happens that their fluctuations in position emerge to be distributed according to a quasi-probability distribution, exhibiting negative values that need special care under both physical interpretations and aspects. In particular, we highlight through the study of the information entropy of the fluctuation density that the MS equation describes a dynamic that includes the restoring of the system condition. Moreover, we propose such negative values as a mere *ad hoc* modeling approach for local extinction and counter-gradient phenomena.

4.1 Model Description

The process of turbulent premixed combustion is mainly characterized by flame propagation towards the unburned region and turbulent dispersion of the resultant product particles. The

combustion process is described by a single dimensionless scalar observable, denoted as *average* progress variable, $0 \leq c(\mathbf{x}, t) \leq 1$, and representing the burned-mass fraction, i.e. the fraction of burned particles which are located in $\mathbf{x}$ at time t. The value $c(\mathbf{x}, t) = 1$ describes the presence of only products and the value $c(\mathbf{x}, t) = 0$ describes the presence of only reactants. To avoid unnecessary mathematical difficulties, we consider a constant-density mixture and a zero-mean turbulent velocity field.

Let $\Gamma(t)$, $t \geq 0$, be the front under investigation, corresponding to a certain iso-line of the progress variable $c(\mathbf{x}, t) = 1$, say $c = 0.5$. We assume that such front can be represented by a level surface of the scalar field $G(\mathbf{x}, t)$, $\mathbf{x} \in \mathbb{R}^n$. The front velocity $\mathbf{v}$ is assumed to be composed as

$$\mathbf{v} = \mathbf{u} - s\,\mathbf{n}\,, \tag{4.1}$$

where $\mathbf{u}$ is the flow velocity and s the consumption speed corresponding to the chemical reaction such that the front propagates relative to the mixture element in the normal direction with velocity $s\,\mathbf{n} = s\nabla G/\|\nabla G\|$. The front equation results to be

$$\frac{\partial G_\Gamma}{\partial t} + \mathbf{u}_\Gamma \cdot \nabla G_\Gamma = s\,\|\nabla G_\Gamma\|\,, \tag{4.2}$$

where index Γ means that only the points $\mathbf{x} \in \Gamma(t)$ are considered. Equation (4.2) is the so-called G-equation originally introduced by Markstein in 1964 [165].

Since the chemical reaction occurs only on the selected level surface $G = c$, it is remarked that a field equation for $G(\mathbf{x}, t)$ is indeed uniquely defined only on such iso-surface [215]. In other words, equation (4.2) is defined solely for points $\mathbf{x}$ belonging to $\Gamma(t)$ such that $G(\mathbf{x}, t) = c$, from which the restriction to $\mathbf{x} \in \Gamma(t)$ in (4.2) follows.

Introducing $\check{G}(\overline{\mathbf{x}}, t)$ as implicit formulation of the mean front position $\overline{\mathbf{x}}$, where the symbol $\check{\cdot}$ represents the ensemble average, and remembering that we are considering the iso-surface of value equal to c, i.e., $\check{G} = c$ without any fluctuation in its value, the ensemble averaging procedure leads to

$$\frac{\partial \check{G}}{\partial t} + \hat{\boldsymbol{u}} \cdot \nabla \check{G} = \widehat{s\mathbf{n}} \cdot \nabla \check{G}\,. \tag{4.3}$$

Equation (4.3) was derived and studied by Oberalck *et al.* [193]. According to an analysis on its symmetries, there is a unique model for $\widehat{s\mathbf{n}}$ that is [193]:

$$\widehat{s\mathbf{n}} = s_T\,\check{\mathbf{n}}\,, \quad \check{\mathbf{n}} = \frac{\nabla \check{G}}{\|\nabla \check{G}\|}\,. \tag{4.4}$$

To conclude, the level surface that propagates following the mean front position $\overline{\mathbf{x}}(t)$ solves the equation

$$\frac{\partial \check{G}}{\partial t} + \hat{\boldsymbol{u}} \cdot \nabla \check{G} = s_T\,\|\nabla \check{G}\|\,, \tag{4.5}$$

where s_T is the turbulent burning velocity in opposition to s, that is the laminar burning velocity. The LHS of (4.5) represents the convective derivative and the RHS describes a source term where the front surface propagates relative to a mixture element in the normal direction with velocity s_T.

It is reported that Sabelnikov & Lipatnikov [240] questioned equation (4.5) derived by Oberalck *et al.* [193]. Their main question concerns the difference between the mean contour and the contour of the mean. In fact, in general the average values denoted by $\hat{}$, or by $\langle\ \rangle$, of observables computed along the random instantaneous front line depicted by $\mathbf{X}_c^\omega(t)$, and the values of observables computed along the mean trajectory $\langle \mathbf{X}_c^\omega(t)\rangle = \overline{\mathbf{x}}(t)$, and here denoted by $\check{}$, differ. In particular, the mean normal direction, i.e., $\hat{\mathbf{n}}$, and the normal direction of the front in the mean position, i.e., $\check{\mathbf{n}}$, are different and, because of the key role of curvature in the front propagation, the evolution of the corresponding fronts strongly differ.

The iso-surface $\check{G} = c$ corresponds to the front contour $\check{\Gamma}(t)$, which is depicted by the mean motion of $\mathbf{X}_c^\omega(t)$, that encloses the domain $\check{\Omega}(t)$. Equation (4.5) is defined only on the level surface corresponding to the front line $\check{\Gamma}(t) = \{\mathbf{x} = \overline{\mathbf{x}}(t) \in \mathcal{S} | \check{G}(\overline{\mathbf{x}}, t) = c\}$. All points inside $\check{\Omega}(t)$ can be considered already reacted and those outside as still unreacted. Then the following indicator function is introduced

$$I_{\check{\Omega}}(t) = \begin{cases} 1\,, & \mathbf{x} \in \check{\Omega}(t) \\ \\ 0\,, & \mathbf{x} \notin \check{\Omega}(t) \end{cases} . \tag{4.6}$$

Indicator $I_{\check{\Omega}}(t)$ allows to estimate the propagation of the chemical reaction in terms of the volume of the reacted mixture fraction.

Random fluctuations of the front contour define also random fluctuations of the volume of the reacted mixture fraction, that are stated by using the sifting property of the Dirac delta-function peaked in the random points $\mathbf{X}_c^\omega(t)$, i.e.,

$$\begin{aligned} I_{\check{\Omega}}^\omega(\mathbf{x}, t) &= \int_{\mathbb{R}^n} I_{\check{\Omega}}(\overline{\mathbf{x}}, t) \delta(\mathbf{x} - \mathbf{X}_c^\omega(\overline{\mathbf{x}}, t))\, d\overline{\mathbf{x}} \\ &= \int_{\check{\Omega}} \delta(\mathbf{x} - \mathbf{X}_c^\omega(\overline{\mathbf{x}}, t))\, d\overline{\mathbf{x}}\,, \end{aligned} \tag{4.7}$$

and after averaging we have

$$\begin{aligned} \langle I_{\check{\Omega}}^\omega(\mathbf{x}, t)\rangle &= \int_{\check{\Omega}} \langle \delta(\mathbf{x} - \mathbf{X}_c^\omega(\overline{\mathbf{x}}, t))\rangle\, d\overline{\mathbf{x}} \\ &= \int_{\check{\Omega}(t)} \mathcal{P}_c(\mathbf{x}; t|\overline{\mathbf{x}})\, d\overline{\mathbf{x}} = \mathcal{V}_e(\mathbf{x}, t)\,. \end{aligned} \tag{4.8}$$

The averaging procedure introduced for the random indicator $I_{\check{\Omega}}^\omega$ takes into account the fluctuations of the random front points $\mathbf{X}_c^\omega$ around their mean position $\overline{\mathbf{x}}$.

To conclude, applying Reynolds transport theorem [105] we obtain the following evolution equation:

$$\frac{\partial \mathcal{V}_e}{\partial t} = \int_{\check{\Omega}(t)} \frac{\partial \mathcal{P}_c}{\partial t}\, d\overline{\mathbf{x}} + \int_{\check{\Omega}(t)} \nabla_{\overline{\mathbf{x}}} \cdot \left[s_T\, \check{\mathbf{n}} \mathcal{P}_c(\mathbf{x}; t|\overline{\mathbf{x}}) \right] d\overline{\mathbf{x}}\,. \tag{4.9}$$

Let the turbulent dispersion be represented by the general evolution equation

$$\frac{\partial \mathcal{P}_c}{\partial t} = \mathcal{E}_{\mathbf{x}}[\,\mathcal{P}_c\,]\,, \quad \mathcal{P}_c(\mathbf{x}; 0|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0)\,, \tag{4.10}$$

where the spatial operator $\mathcal{E}_{\mathbf{x}}[\cdot]$ includes the particle displacement statistics such as the variance $\sigma^2(t) = \langle ||\mathbf{x} - \mathbf{x}_0||^2 \rangle / 3$.

Equation (4.9) is also governed by the volumetric expansion of $\Omega(t)$. This expansion is connected with the *consumption rate* that in a general form is set to be

$$\mathbf{u}(\mathbf{x}, t) = s_T(\kappa, t)\, \hat{n}\,, \tag{4.11}$$

where κ denotes the local mean *curvature*. If molecular processes are neglected, the initial burning speed must be zero, i.e. $s_T(\kappa, 0) = 0$, see Ref. [202]. From (4.11) the location of the flame surface from a Lagrangian point of view can be expressed as

$$\mathcal{L}_f(t) = \mathcal{L}_0 + \int_0^t \mathbf{u}(\mathbf{x}, \tau)\, d\tau\,. \tag{4.12}$$

Finally, inserting (4.10) and (4.11) in (4.9) gives

$$\begin{aligned} \frac{\partial \mathcal{V}_e}{\partial t} = \mathcal{E}_{\mathbf{x}}[\,\mathcal{V}_e\,] - \nabla \cdot \int_{\check{\Omega}(t)} & s_T\, \check{\mathbf{n}} \mathcal{P}_c(\mathbf{x} - \overline{\mathbf{x}}; t)\, d\overline{\mathbf{x}} \\ & + \int_{\check{\Omega}(t)} \mathcal{P}_c \left\{ \frac{\partial s_T}{\partial \kappa} (\nabla_{\overline{\mathbf{x}}}\, \kappa) \cdot \check{\mathbf{n}} + s_T \kappa \right\} d\overline{\mathbf{x}}\,, \end{aligned} \tag{4.13}$$

As a result, it can be observed that the evolution of the progress variable is led by three main factors: turbulent motion, displacement speed of the contours of $c(\mathbf{x}, t)$ and their mean curvature. It is worth remarking that (4.13) cannot be reduced to the most widely used front propagation equations [311], and, since (4.13) follows from the exact Lagrangian definition (4.8), none of them is physically correct to model turbulent premixed combustion.

Another formulation for equation (4.9), taking into account the probability flux $\boldsymbol{J}$, as well as the following general kinetic equation for $\mathcal{P}_c$

$$\frac{\partial \mathcal{P}_c}{\partial t} = -\nabla \cdot \boldsymbol{J}\,, \tag{4.14}$$

is

$$\frac{\partial \mathcal{V}_e}{\partial t} = -\nabla \cdot \int_{\check{\Omega}(t)} \boldsymbol{J}(\mathbf{x}; t|\overline{\mathbf{x}})\, d\overline{\mathbf{x}} + \int_{\check{\Omega}(t)} \nabla_{\overline{\mathbf{x}}} \cdot \left[s_T\, \check{\mathbf{n}} \mathcal{P}_c(\mathbf{x}; t|\overline{\mathbf{x}}) \right] d\overline{\mathbf{x}}\,, \tag{4.15}$$

that is again a reaction-diffusion equation.

According to classical properties of diffusion, the PDF $\mathcal{P}_c$ of the stochastic process $\mathbf{X}_c^\omega$ is assumed to be unimodal and its mean and median are coincident. This means that it is symmetric and normalizes with respect to both $\mathbf{x}$ and $\overline{\mathbf{x}}$. Consequently, $\mathcal{V}_e(\mathbf{x},t)$ ranges in the compact interval $[0,1]$,as we expect by our equivalent for the consumption rate that comes from intrinsic Lagrangian arguments. In applications, the effective front contour is given by selecting a threshold value for $\mathcal{V}_e(\mathbf{x},t)$, i.e., $\Gamma_e(t)=\{\mathbf{x}\in\mathcal{S}|\mathcal{V}_e(\mathbf{x},t)=\mathcal{V}_e^{th}\}$.

We thus remark that all formalism derived so far can be applied in TPC, after recognizing that equation (4.5) is related to the G-equation [216] and $\mathcal{V}_e(\mathbf{x},t)$ to the so-called progress variable.

Moreover, if no curvature effects is considered, the consumption speed s_T results to be the turbulent burning velocity and it can be stated [202]

$$s_T = s_L + \frac{\mathcal{D}}{\lambda}\,, \tag{4.16}$$

where s_L is the laminar burning velocity and λ a length scale related to the equivalence ratio of the mixture [202] and $\mathcal{D}$ is the turbulent diffusion coefficient.

Finally, for homogeneous isotropic and stationary turbulence classical flux-gradient relation can be assumed:

$$\boldsymbol{J}(\mathbf{x},t) = -\mathcal{D}\nabla\mathcal{P}_c\,, \tag{4.17}$$

that is the equivalent of choosing the simple non-Markovian parabolic model for the turbulent dispersion in Equation 4.10:

$$\mathcal{E}_{\mathbf{x}}[\,\mathcal{P}_c\,] \equiv \mathcal{D}\nabla^2\mathcal{P}_c. \tag{4.18}$$

and equation (4.15) becomes

$$\frac{\partial\mathcal{V}_e}{\partial t} = \mathcal{D}\nabla^2\mathcal{V}_e + \int_{\check{\Omega}(t)} \nabla_{\overline{\mathbf{x}}}\cdot\left[s_T\,\check{\mathbf{n}}\mathcal{P}_c(\mathbf{x}-\overline{\mathbf{x}};t)\right]d\overline{\mathbf{x}}\,. \tag{4.19}$$

that reduces to the G-equation when no diffusion is assumed, i.e., $\mathcal{P}_c(\mathbf{x}-\overline{\mathbf{x}};t)=\delta(\mathbf{x}-\overline{\mathbf{x}})$.

Moreover, when the front curvature $\kappa=\nabla\cdot\check{\mathbf{n}}$ in the mean front position is taken into account, we have that $s_T(\mathbf{x},t)=s_T(\kappa,t)$ and equation (4.19) can be re-written as

$$\begin{aligned}\frac{\partial\mathcal{V}_e}{\partial t} = \mathcal{D}\nabla^2\mathcal{V}_e - \nabla\cdot\int_{\check{\Omega}(t)} s_T\,\check{\mathbf{n}}\mathcal{P}_c(\mathbf{x}-\overline{\mathbf{x}};t)\,d\overline{\mathbf{x}} \\ + \int_{\check{\Omega}(t)} \mathcal{P}_c\left\{\frac{\partial s_T}{\partial\kappa}(\nabla_{\overline{\mathbf{x}}}\,\kappa)\cdot\check{\mathbf{n}} + s_T\kappa\right\}d\overline{\mathbf{x}}\,,\end{aligned} \tag{4.20}$$

where again, because $\mathcal{P}_c=\mathcal{P}_c(\mathbf{x}-\overline{\mathbf{x}};t)$, the formula $\nabla_{\mathbf{x}}\mathcal{P}_c=-\nabla_{\overline{\mathbf{x}}}\mathcal{P}_c$ is used.

If a plane flame is considered then $\kappa=0$ such that $s_T=s_T(t)$ and $\check{\mathbf{n}}=\check{\mathbf{n}}(t)$, and equation (4.20) becomes

$$\frac{\partial\mathcal{V}_e}{\partial t} = \mathcal{D}\nabla^2\mathcal{V}_e + s_T(t)\|\nabla\mathcal{V}_e\|\,, \tag{4.21}$$

that is the same equation derived by Zimont & Lipatnikov [320] and discussed also in Ref. [319]. Hence equation (4.20) can be considered as the natural extension to the case with non null curvature of the model equation presented in [320, 319]. This application is discussed in Ref. [202].

This method suggests that for different choices of the PDF of the fluctuation of the front position $\mathcal{P}_c(\mathbf{x};t)$ different equations can be recovered. In this paper we are interested in deriving through this method the MS equation, and some properties of this equation are studied.

4.2 Statistical fronts in Michelson–Sivashinsky Equation

In addition to the effect of flow interactions [147], premixed flames are also subject to various instabilities, which can be thermal diffusive or hydrodynamic in nature. The thermal diffusive instability [254] is caused due to non-unity Lewis number, implying non equi-diffusion and leads to the formation of cellular structures on the flame surface and/or oscillations of the flame surface. The hydrodynamic instability, also known as the Darrieus-Landau instability [139] is caused by the gas expansion produced by heat release in a flame, causing the flow lines across the front to deviate towards the normal to the flame which leads to wrinkling of the flame. Large scale flames, often found in practical applications, are susceptible to this instability leading to the formation of sharp crests on the flame pointing towards the burned side or the roughened surface observed on expanding flames in turbulent flow fields [96].

This instability, acting together with turbulence in the flow, is known to play an important role in turbulent flame propagation.

One such simplified model describing the evolution of a weakly perturbed flame front is the Michelson-Sivashinsky (MS) equation [253, 180]. The latter model is based on a nonlinear integro-differential equation obtained via asymptotic analysis. Such equation holds in the limit of weak thermal expansion, or small density changes along the flame. M-S equation sheds some light in the development of the hydrodynamical instability by adequately describing the corrugated shapes that form beyond the occurrence of instability and the subsequent increase in propagation velocity of the flame. However, its application is restricted by the fact that density changes in combustion are rarely small.

The potentiality of the approach is investigated further in this Section focusing on the Darrieus-Landau (hydrodynamic) instabilities. In particular, this model formulation is set to lead to the Michelson-Sivashinsky (MS) equation.

Hereinafter, we consider the one-dimensional MS equation in the following form:

$$\frac{\partial \phi}{\partial t} = \frac{\partial^2 \phi}{\partial x^2} - \left(\frac{\partial \phi}{\partial x}\right)^2 - \mathbb{D}_x^1 \phi \,, \tag{4.22}$$

where $\mathbb{D}_x^1$ is the space-fractional derivative in the sense of Riesz–Feller [222] of order 1.

Let the Fourier transformation be defined by

$$\tilde{f}(\xi) = \int_{-\infty}^{+\infty} \mathrm{e}^{+i\xi x} f(x)\, dx\,, \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{e}^{-i\xi x} \tilde{f}(\xi)\, d\xi\,, \tag{4.23}$$

then the space-fractional derivative in the Riesz–Feller sense $\mathbb{D}_x^\alpha$, with $\alpha \in (0,2)$, is defined by its Fourier symbol $-|\xi|^\alpha$. For the integer value $\alpha = 2$ the classical second derivative is recovered, i.e., $\mathbb{D}_x^2 = \dfrac{\partial^2}{\partial x^2}$, and for the integer value $\alpha = 1$, the operator $\mathbb{D}_x^1$ is the fractional derivative of order 1 with Fourier symbol $-|\xi|$ and differs substantially from the standard first derivative. Actually, the fractional derivative of order 1 is related to the Hilbert Transform via the relation

$$\mathbb{D}_x^1 w(x) = \frac{1}{\pi} \frac{d}{dx} \int_{-\infty}^{+\infty} \frac{w(y)}{(y-x)}\, dy\,. \tag{4.24}$$

According to the standard derivation of MS equation in cylindrical symmetry [95], function $\phi(x)$ represents the height of the flame in point x of the axis directed along the size of the cylinder.

We consider the front $\check{G}$ described by (4.5) as the reference front. We denote by the vector $\boldsymbol{y}$ the points of he multidimensional domain where the iso-surface $\check{G}$ is embedded.

By adopting the model described in Section 4.1, we introduce the field $g(\boldsymbol{y}, t)$ defined according to the second line of formula (4.8), i.e,

$$g(\boldsymbol{y}, t) = \int_{\check{\Omega}(t)} \mathcal{P}_c(\boldsymbol{y} - \hat{\boldsymbol{y}}; t)\, d\hat{\boldsymbol{y}}\,. \tag{4.25}$$

In [287], this field is analyzed through the equation 4.20 and arranged in order to provide an alternative solution of the M-S equation.

In this case, to apply the formalism discussed in this Chapter to this problem, a different turbulent dispersion law must be considered.

In formulas 4.10 reads

$$\frac{\partial \mathcal{P}_c}{\partial t} = \mathcal{E}_{\mathbf{x}}[\,\mathcal{P}_c\,] = \frac{\partial^2 \mathcal{P}_c}{\partial r^2} - \mathbb{D}_r^1 \mathcal{P}_c\,, \quad \mathcal{P}_c(r;0) = \delta(r)\,, \tag{4.26}$$

We recall that the Green functions of the space-fractional diffusion equation

$$\frac{\partial \mathcal{P}_c}{\partial t} = \mathbb{D}_r^\alpha \mathcal{P}_c\,, \quad \mathcal{P}_c(r;0) = \delta(r)\,, \tag{4.27}$$

are Lévy stable densities. In particular, when $\alpha = 2$ the Green function is the Gaussian density and when $\alpha = 1$ it is the Lorentzian density. The Gaussian density is associated to classical diffusion and the Lorentzian density can be associated to a lightly damped linear oscillator.

We shift the attention towards the fractional differential equation involved in 4.26

$$\frac{\partial \mathcal{P}_c}{\partial t} = \frac{\partial^2 \mathcal{P}_c}{\partial r^2} - \mathbb{D}_r^1 \mathcal{P}_c\,, \quad \mathcal{P}_c(r;0) = \delta(r)\,, \tag{4.28}$$

and, by using Fourier inverse transformation, the quasi-probability that solves (4.28) can be written in the following integral form

$$\mathcal{P}_c(r,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-i\xi r}\,\mathrm{e}^{-\xi^2 t+|\xi|t}\,d\xi\,, \tag{4.29}$$

such that it results to be similar to the integral form of the Wigner quasi-distribution for quantum optics. The second term on the RHS of (4.28), because of the sign minus and $\alpha = 1$, may be understood as a counter-damping effect of an harmonic oscillator.

4.3 Discussion

It is here highlighted that the PDF of fluctuations which solves (4.28) emerges to be a quasi-probability distribution showing negative values that requires high care, see Figures 4.1 and 4.2. This fact addresses a non-complete correspondence between the modelling approach for random front propagation described in Section 4.1 and the MS equation. We study this failure with the aim to provide an understanding of the dynamic inside the MS equation and some suggestions for future model developments.

In particular, we argue that this failure is mainly due to a restoring property of the nonlocal term related to the velocity v^* provided by the flow field at the interface, and, as a mere modelling approach, these negative values can be interpreted as due to local reversibility that can be ascribed to the so-called counter-gradient phenomenon.

4.3.1 Restoring property of the MS equation

The nonlocal term of the MS (4.22), i.e.,

$$\frac{1}{2}\mathbb{D}_x^1\phi = \frac{1}{4\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\xi|\,\mathrm{e}^{i\xi(x-\mu)}\phi(\mu,t)\,d\xi d\mu = v^*\,, \tag{4.30}$$

where v^* is the velocity of the flow field at the interface, embodies a restoring characteristic of the MS equation.

Actually, this deviation from the Gaussian behaviour causes the negative values in the PDF $\mathcal{P}_c$ (4.29). To study this restoring we consider the information entropy of the modulus of the PDF $\mathcal{P}_c(\mathbf{x};t)$, namely:

$$S(t) = -\int |\mathcal{P}_c| \ln|\mathcal{P}_c| dx\,, \tag{4.31}$$

and the plots of the entropy and the entropy production rate dS/dt are displayed in Figure 4.3. The entropy production rate is a measure for the irreversibility of a process. The diffusion equation represents the canonical irreversible process and has positive entropy production. The wave propagation corresponds to the reversible process and has zero entropy production. In Figure 4.3, we observe that the rate of the information entropy decays faster, i.e., $\sim t^{-2.6}$, than in the standard diffusive case, i.e., $\sim t^{-1}$.

In order to reproduce a similar decay of the rate of information entropy, we look for a simpler equivalent to $|\mathcal{P}_c|$ that share similarities with respect to the spatial and temporal variations observed in Figure 4.2. In particular, we chose a function that expresses temporal and spatial variations in a first window of time, and then decayed to a standing spatial wave depending only on x. The selected wave-like function consists of the superposition of a wave and decaying wave-like disturbances with high frequency. In formulas it reads:

$$\eta(x,t) = \cos\left[\frac{x}{2}\right] + \cos[30\,x]t^{-0.6}\,. \tag{4.32}$$

In Figure 4.4, the information entropy and its rate of function $|\eta(x,t)|$ are displayed against the scaling t^{-1} and $t^{-2.6}$, and the fast decay $t^{-2.6}$ is plotted as reference for the entropy. The fact that the same decay of entropy rate of Figure 4.3 is expressed by making use of a simpler analytical representation, confirms that the salient features of the Green function of the fractional differential equation (4.28) are a spatial (traveling) wave with initial spatio-temporal perturbations.

To conclude, the solution provided by formula (4.25) highlights through the function $\mathcal{P}_c$ the existence of a restoring wave-like motion inside the MS equation. Such behaviour cannot be reproduced by a pure diffusive process, and from this the emergence of negative values of $\mathcal{P}_c$ and the failure of the proposed approach described in Section 4.1. This wave-like motion is introduced in the MS equation through the nonlocal velocity v^* defined in (4.30). This nonlocal term, together with its sign, provides a restoring force that poses a challenge for modelling the MS dynamic by random fronts.

4.3.2 Quasi-probability for modelling local extintion and counter-gradient phenomena

The emergence of a quasi-probability suggests the idea to introduce, in analogy with quantum mechanics, a probability amplitude whose squared modulus provides the observable for the purely propagating front. In formulas, the following can be stated.

Let $C = |C|\,\mathrm{e}^{iQ} = C_R + iC_I$ be a probability amplitude with real and imaginary part. If we assume that function $g(\boldsymbol{y},t)$ defined in (4.25), and related to the MS equation (4.22) as previously derived, corresponds to the square of the real part of C, then we have

$$g = C_R^2 = |C|^2 - C_I^2\,, \tag{4.33}$$

where accordingly to definition (4.25)

$$|C|^2 = \int_{\mathcal{P}_c \geq 0} I_{\tilde{\Omega}}(t)\mathcal{P}_c(x-\overline{x},t)\,d\overline{x}\,, \tag{4.34}$$

and

$$C_I^2 = -\int_{\mathcal{P}_c < 0} I_{\tilde{\Omega}}(t)\mathcal{P}_c(x-\overline{x},t)\,d\overline{x}\,. \tag{4.35}$$

Hence, the negative values of the quasi-probability result to be associated to the imaginary part of the probability amplitude C and they have a role in slowing the propagation because $C_I^2 \geq 0$. In particular, if we integrate (4.25) in space we have that in correspondence of the negative values of the quasi-probability there is a reduction of the mass amount. Actually, the negative values highlight where statistically the fraction of burned mixture is replaced by unburned mixture.

From this reasoning, we propose to use the negative intervals of the quasi-probability for modelling the non-propagating features of the process as local extintion and counter-gradient phenomena.

In fact, the negative values refer to a Eulerian local reversibility of the progress variable (and not to a reversibility of the Lagrangian volume of mixture), that occurs because of the entering of fresh mixture into a volume just now fully burned. This effect can be ascribed first to the local extinction, that stops the propagation of the combustion, and then to the so-called counter-gradient, which is generated by the density difference between reactants and products, that pushes back the front of the burned mixture.

4.4 Summary and conclusions

A model of the flame front propagation based on the G-equation and on stochastic fluctuations imposed to the mean flame position is considered. Stochastic fluctuations described by an adequate PDF are thus imposed to the average flame position in order to give a proper formulation of the flame surface propagation. An evolution equation of reaction-difffusion type is derived for an observable that can be understood as the effective burned fraction, or progress variable.

This modelling approach allows for a number of observations on TPC, in general, and on the MS equation, in particular. In general, when stochastic fluctuations are removed, the G-equation along the motion of the mean flame position is recovered suggesting that approaches based on reaction-diffusion equations and G-equation are indeed complementary and they can be reconciled.

Moreover, it is known that hydrodynamic instabilities in TPC are described by the MS equation. This model could in principle lead to an alternative derivation of the MS equation allowing for a number of findings of its properties that are complementary to, for example, the skewness of the flame curvature [61] or the stability of the front [295]

The PDF of the fluctuations of the front position results to be governed by the nonlocal term associated to the MS equation, and from this it follows that such PDF is a quasi-probability, i.e., a density function with negative values. We showed that these negative values are a footprint of a restoring property related to the nonlocal term. Actually, by studying the information entropy and its rate, it emerges that the fluctuations of the front position are not driven by a diffusive motion only, but also and more by a wave-like motion. This wave-like motion restores the system configuration and opens a challenge for its statistical and stochastic modeling.

However, as a mere modeling approach, the negative values of a quasi-probability can be used to model the local reversibility of the Eulerian progress variable, which means that if a fixed point is occupied by a burned volume it turns to be occupied by an unburned one and then the modelization of the local extinction follows, together with the counter-gradient.

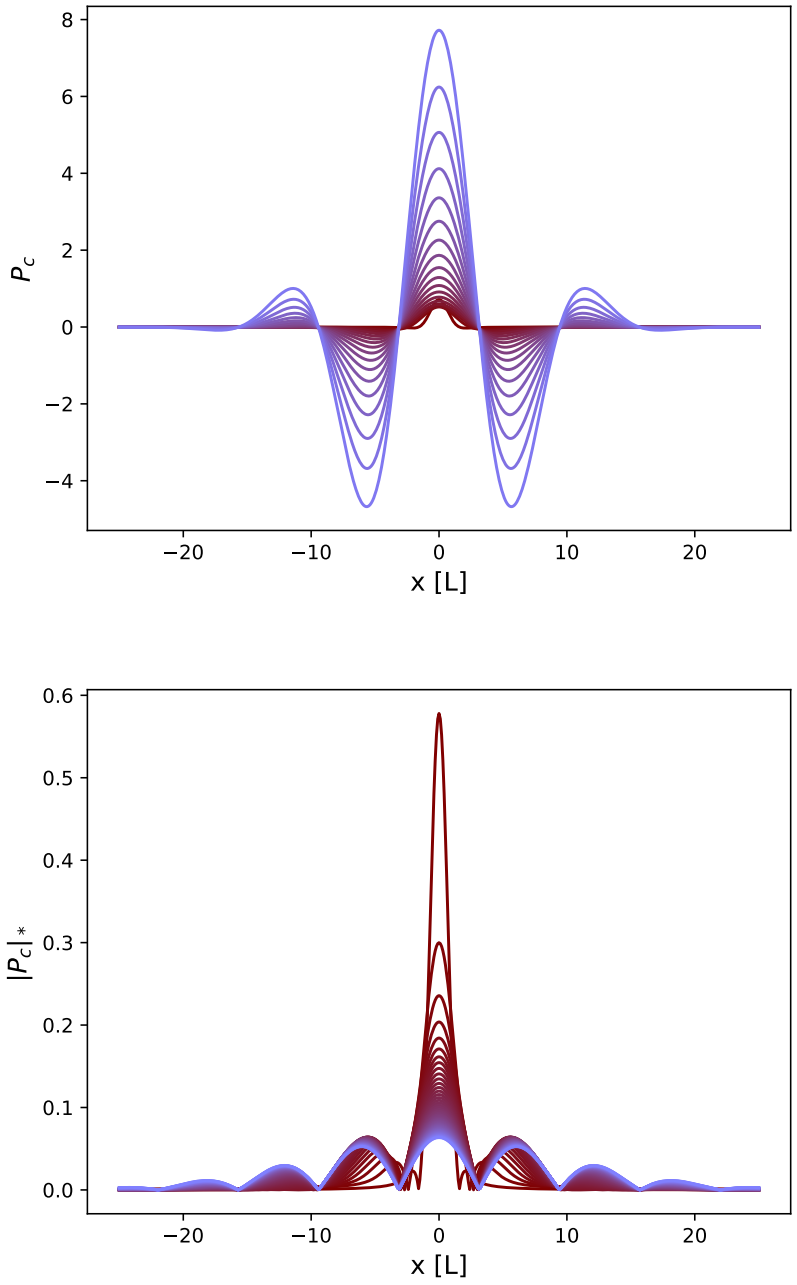


Figure 4.1: Top: evolution in time of the quasi-probability $\mathcal{P}_c$ (4.29). The colors from brown to purple stay for time since $t = 0.3$ to $t = 18$. Bottom: the same as in the Top panel but for the function $|\mathcal{P}_c|$ normalized in order to represent a probability density function.

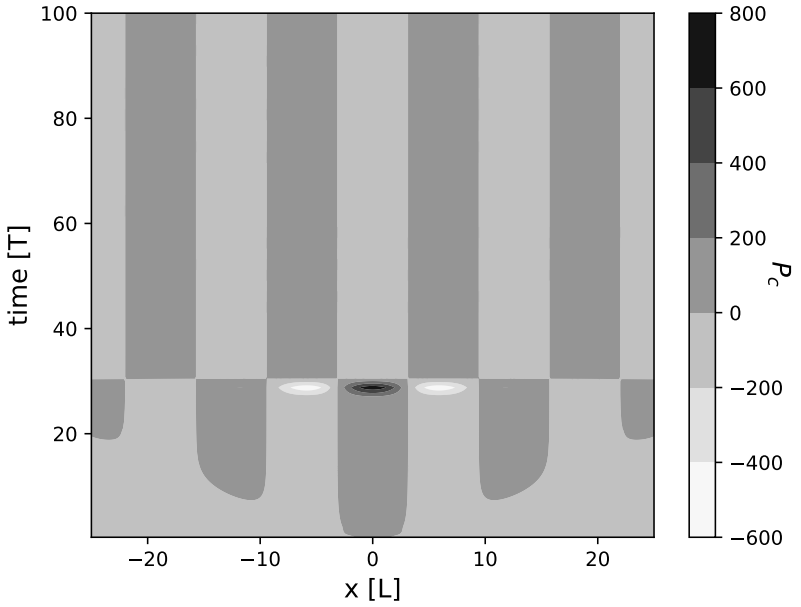


Figure 4.2: Surface plot of $\mathcal{P}_c$. Space is reported in the horizontal axis and time in the vertical axis. The appeareance of the two new modes around the x-t point $(\pm 10, 7)$ establihes a connection with the peak in the information entropy, see Figure 4.3. We hypothesize that the solution of the considered Fractional Differential Equation (4.28) due to its nonlocal nature creates new information (two new maximums appear in the quasi Probability $\mathcal{P}_c$) thus decreasing the entropy rate of the PDF portrayed in figure 4.1.

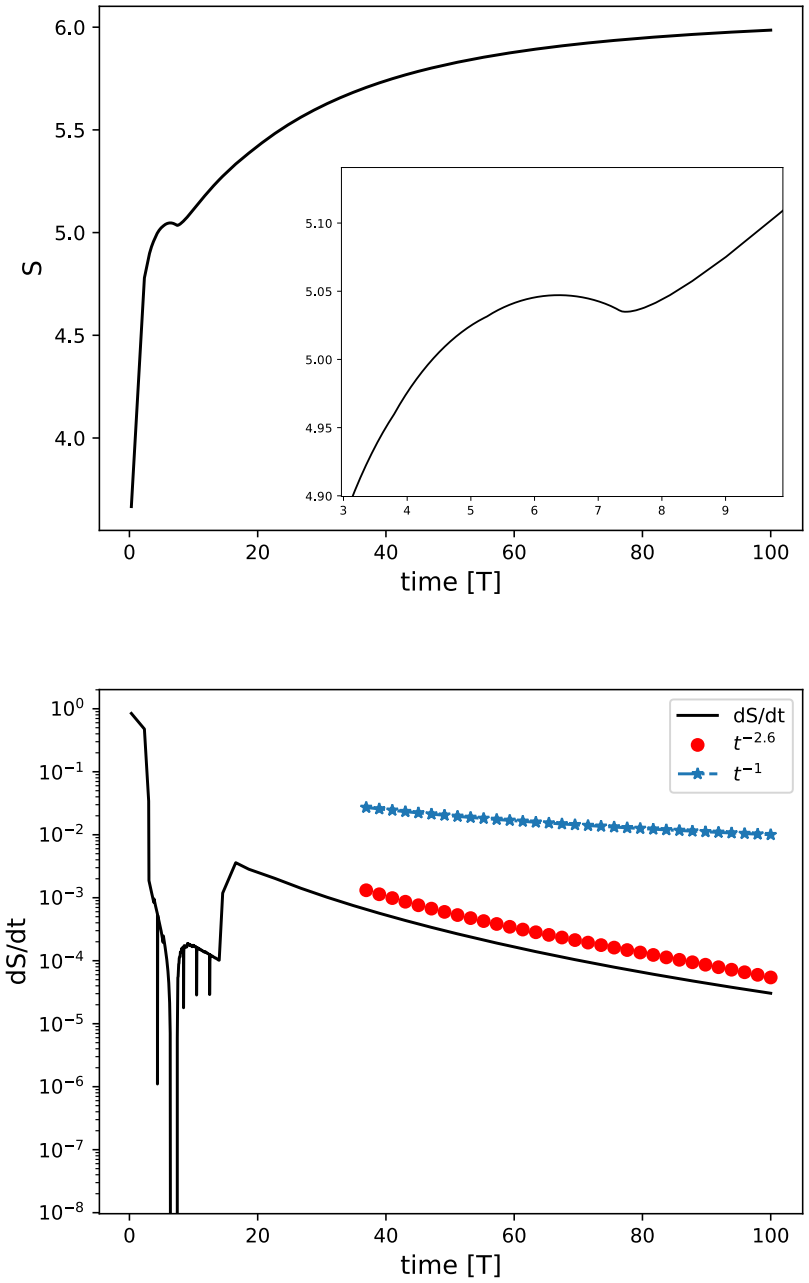


Figure 4.3: Information entropy (Top) and its rate (Bottom) of the PDF $|\mathcal{P}_c|$ versus time. Blue stars and red dots decay as t^{-1} and $t^{-2.6}$, respectively.

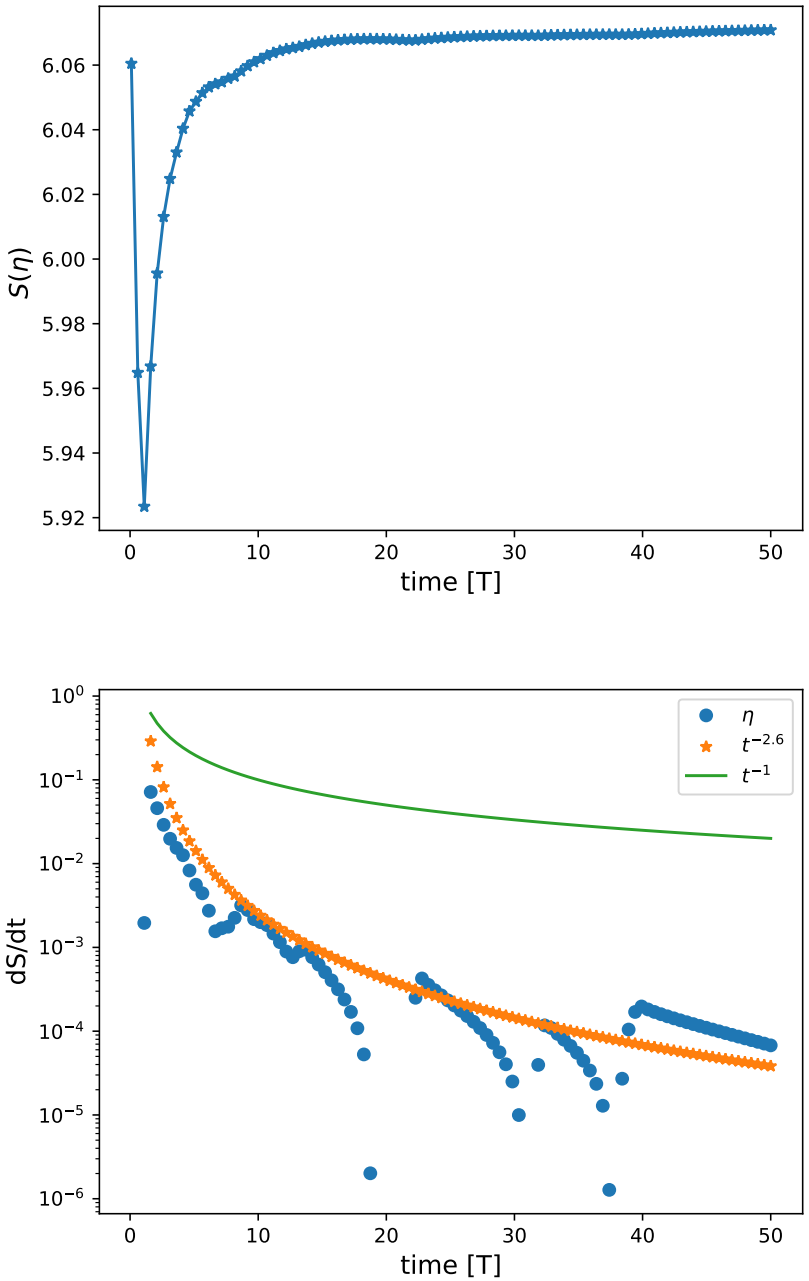


Figure 4.4: Information entropy (Top) and its rate (Bottom) of function $|\eta(x,t)|$ defined in (4.32) versus time. Green line and orange stars decay as t^{-1} and $t^{-2.6}$, respectively.

The No. 1 cause of forest fires is trees.

Pat Paulsen

Chapter 5

Applications to Wildland forest Fires

This Chapter contains:

5.1 Introduction

Wild-land fire (also known as *wildfire*) is a worldwide phenomenon that occurs in any vegetated area independently of national fire-fighting management strategies, causing significant damages to the environment, properties and human lives (see e.g. [39]).
Wildfires are characterized then by a strong social and economical impact, calling for a more profound understanding of their dynamics for controlling the risk and managing their suppression [277].
Mathematical and physical Modeling of wild-land fire propagation serves as a crucial (and relatively cheap) tool in order to contrast the high economic and environmental damage provoked by wild-land fires.
An effective and swift modeling can aid in building up an efficient plan for fire suppression and restrict the life and property damage to a minimum, helping in gathering data to orient the practitioners in their decision-making processes.
Unfortunately, producing easy-to-use models that are accurate from a physical perspective is not an easy task at all. Modeling the propagation of wild-land fire has proven to be a complex, multi-scale, multi-physics and multi-discipline process.
Wildfire propagation involves processes from the scale of the combustion chemistry to the fire-atmosphere coupling including interactions with rather different factors such as the topography and flame geometry [270, 271].
The core problem lies in delivering either a versatile or a specialized model which is relatively easy to implement and may provide the requested previsions in a reasonable amount of time.
With the growth in available computational power, various new and improved modeling approaches have been introduced and analyzed over the last years.
We address for this purpose the quite complete reviews by Sullivan [267, 268, 269], which provide a comprehensive summary of the large set of physical, empirical, statistical and mathematical analogue models that are currently available in literature to tackle this difficult problem.

Structure of the Chapter This Chapter is structured as follows. In Section 5.2 the main model and techniques adopted in literature to model and simulate wild-land fire spread are reviewed. In Section 5.3 the phenomenon of fire-spotting is described and the main model available for its prediction are listed. Section 5.4 describes how to model derived in Chapter 3 has been adapted to model random effects in forest fires (such as turbulence and fire spotting) while its parametrization, `RandomFront`, is presented in Section 5.5.
The computational tools used for this application of the model are listed in Section 5.6. The output of the model in two different simulation environment (LSFire and WRF-SFire) is discussed in Section 5.7. UQ and SA of a simple test case for the model with `RandomFront` parametrization is given in Sections 5.8 and 5.9. The latest advances in this research, regarding multi-scale modeling of fire spotting, are described in Section 5.10.

5.2 State of Art for Wild-land fire modeling

The different models and analogues for wild-land fire can be categorized in several ways. In the following, a first distinction concerning their complexity is furnished. After that, another subdivision of the present models is outlined, this time in the context of the dichotomy between R-D fronts and sharp fronts, in line with the contents of Chapter 1.

5.2.1 Wildfire models: growing complexity

- *Empirical models*: they are those models which describe the evolution of a fire by making use of approximations derived from statistics of real fire scenarios. Those models share the feature of depicting an advancing front who propagates according to the normal vector of the front calculating for each point the rate of spread (ROS) as a scalar function of *local parameters*: fuel, moisture, terrain steepness (orography), wind direction and magnitude, etc. They are by far the most accessible models in term of model complexity and use of computational resources. As a drawback, they do exhibit some limitations when they are used in contexts different from the ones they were conceived for since they rely on empirical or quasi-empirical laws. Some models that folds in this category are the ones based on Rothermel's Rate of Spread [237], like Behave [10], BehavePlus [11], FARSITE [92] and Anderson model [8].

- *Quasi-physical* models, based on Reaction-Diffusion-Convection equations: they are typically derived from mass, energy and momentum conservation arguments. They consists on (possibly large) sets of coupled PDEs that model the physical processes that give rise to the fire spread. The computational cost is usually more demanding compared with the empirical models, but they adapt well to a large spectrum of different settings since they rely less on

empirical assumptions. Belong to this category Albini's model [5], Butler's model [41] and PhyFire [14].

- *Full fledged physical models*: more sophisticated models which describe in depth and with high resolution both physical and chemical aspects of wild-land fires. They tend to be numerically demanding, and the real-time simulations of real case scenarios is not feasible even if massively parallel supercomputers are employed. To name a couple of examples, we cite FIRETEC [145, 146] and WFDS Wildland-Urban Interface Fire Dynamics Simulator) [175].

5.2.2 Cathegorizing wildfire models: R-D fronts and Sharp Front approaches

Reaction-Diffusion based models - brief overview

It rather intuitive that an important observable in regions where a wild-land fire takes place is the temperature field. Given that the mechanisms of energy transport involve molecular processes in turbulent flows, on a macroscopic level the transport of energy (that is of paramount importance in a wild-land fire simulation) can be reproduced analytically as a diffusion process, where the fuel combustion is considered as a source term.
A two-equation model involving the average temperature field, $T(\mathbf{x}, t)$ and the fuel mass fraction $F(\mathbf{x}, t)$, with $F \in [0, 1]$ has been studied and has given good results, see e.g. [14, 184, 248, 18, 160]. In its most simple formulation, this two-equations model can be formulated as:

$$\frac{\partial T}{\partial t} + \mathbf{U}\nabla T = \nabla(k\nabla T) + RF - (T - T_a), \tag{5.1}$$

$$\frac{\partial F}{\partial t} = -RF, \tag{5.2}$$

where $\mathbf{U}$ stands for the mean wind velocity at fire height, K is a thermal diffusion coefficient, R is the reaction rate and T_a is the ambient temperature.
In models based on sets of reaction-diffusion equations, the typical solution features large gradients in the neighbourhood of the fire front, while it remains nearly constant elsewhere. The fire evolution can then be tracked following this sharp gradients, that are to be regarded then as an important asset.

Sharp front models

Even if a Reaction Formulation of the propagating front is feasible, a large part of the statistical and mathematical models available in literature comprise of a Rate of Spread (ROS) formulation and a moving interface technique to model the interface between burnt and unburnt regions. The analytical formulation of the ROS is developed independent of the moving interface scheme and

can be characterised in terms of the wind speed, slope, fuel characterisation, combustion properties, along with other experimental data. Various formulations for the ROS are available in literature [237, 19, 157, 90, 91], and they are usually versatile enough to be used with most moving interface schemes. Among the different moving interface methods available, Eulerian Level Set Method (LSM) described in Chapter 1 is used extensively in wildfires to represent the propagating fire-line [250, 157, 159]. As already pointed out in Chapter 1, it is a scheme which represents a moving interface on a simple Cartesian grid. This method is particularly appropriate for wildfire simulations, as it permits an accurate computation of the front normal vector which is used to propagate the front. Other approaches available in literature which focus on wildfire propagation modelling include the Lagrangian Discrete Event System Specification (DEVS) [124, 87, 86] and Huygens' principle [8, 231, 232, 233, 92]. The Lagrangian DEVS approach accounts for the propagation of the wildfire without defining any underlying mesh to represent the burning state. Since DEVS folds into the cathegory of event driven simulations, it considers time as a continuous variable and allows for faster simulations over higher resolution. On the other hand, wildfire models based on Huygens' principle utilize an elliptical spread at each point of the fire-front and in some cases also benefit from some analytical results [8, 231, 232, 233].

5.2.3 Rate of Spread Formulation

Since the semi-emprical model of Rothermel is adopted in the following part of the Chapter, it surely deserves a detailed explanation. The interested reader is invited to read the recent technical report [12].

Rothermels model originated from the work provided by Frandsen [97] where the principle of energy conservation was applied to a unit volume of fuel ahead of a steadily advancing fire in a homogeneous fuel bed. In the latter analysis, the fuel-reaction zone is considered as fixed and the unit volume moves as a constant depth toward the interface. The unit volume ignites then at the interface. Rate of spread can be considered thus a ratio between the heat flux received from the source and the heat required for ignition by the potential fuel.

Rothermel's model is flexible in the sense that can be adapted to several tipology of fuels. To be more specific, the practitioner can change the number of fuels, the distribution of fuel size (homogeneous or heterogeneus size distributions), and even include the distinction between dead and live fuel, for the simulator to fit real wild-land fires best.

While the details of such models can be found in [12], we report here the seminal case, where a single-sized dead fuel is considered.

$$ROS = \frac{I_R \xi \left(1 + \phi_W + \phi_S\right)}{\rho_b \epsilon Q_{i,g}} \tag{5.3}$$

Where ξ is the propagating flux ratio, $\phi_{W,S}$ are corrective terms due to wind and slope respectively, ρ_b is the bulk density (lb/ft^3), that is the amount of oven-dry fuel per unit volume of fuel bed, ϵ is the Effective Heating Number (i.e. the proportion of a fuel particle that is heated to ignition temperature at the time flaming combustion starts, that tends to unity for fine fuels), and Q_{ig} is the Heat of preignition ($Btu/ft^2/min$), that is the amount of heat required to ignite one mass unit of fuel.
In the last equation I_R is the reaction intensity of the fire ($Btu/ft^2/min$), that is the energy release rate per unit aera of fire front.
It is derived through the following relation

$$I_R = \Gamma' w_\eta h \eta_M \eta_S \tag{5.4}$$

where Γ' is the optimum reaction rate (min^{-1}) (reaction velocity that would exist in the case of mosture-free fuel that contained minerals and aplpha cellulose), η_M is the Moisture damping coefficient and η_S is the Mineral Damping Coefficient.
The main concepts that this simple yet effective equation help describing are enlisted in Figure 5.1
The input parameters to correctly generate a ROS field $ROS(\boldsymbol{x}, t)$ computing one after one every factor of Equation 5.3 (that tipically depend on other subparameters, not shown here for the sake of simplicity) can be divided in three main areas:

- *Properties of the fuel particle*: heat content, mineral content, particle density;
- *fuel array settings*: fuel load, surface-area-to-volume ratio (SAV), mean depth of fuel bed, dead fuel moisture of extinction (the content in moisture of the fuel, weighed over all the fuel classes, at which the fire is not spreading).;
- *environmental factors*: fuel moisture content, wind velocity, slope steepness (orography).

5.3 Introduction to fire spotting

One of the key aspects of fire propagation is the so-called fire-spotting, whose modeling constitutes a difficult challenge [83]. This phenomena occurs when burning embers tear off from the main fuel source (that is, from the main propagating wild-land fire) and cause new independent ignitions thanks to the horizontal wind which may transport them far beyond the zone of direct ignition. It is documented as a dominant phenomenon contributing towards a conspicuous spread of fire in many devastating recorded historical fires [130].
It is a hazardous phenomenon because it may lead to a drastic acceleration of the overall process of fire spread, bringing harmful consequences.

Moreover, fire-spotting has an unpredictable, random and non-linear nature and it constitutes thus a challenging issue in wildfire science.

5.3.1 State of Art for Fire-spotting Modeling

Researchers have studied the phenomenology of fire-spotting by the means of both experimental and theoretical insights in order to update the existing wildfire management decision support systems.

Most of the experimental procedures and techniques for studying the fire-spotting phenomenon focus on characterization of firebrands generation (see e.g. [163, 79]), drag forces and ignition processes [162], size and shape of the firebrands [164, 279]. However, the experimental settings shared a few drawbacks. Indeed, the short temporal and spatial scales of the experimental settings limited a detailed description of the flight paths of the firebrands and thus the landing distributions.

On the other hand, shifting the attention to analytical and *in silico* models, mathematical and computational modeling of firebrands can provide an estimate of the maximum landing distance and flight paths of the firebrands through a simplified overview of the physical dynamics of plume characteristics, the overall fire behavior, and the atmospheric conditions around the fire.

Tarifa et al. in [275, 276] and Albini et al. in [3] were the first to develop simplified plume models to estimate firebrand lifetimes, flight paths and the potential fire-spotting distance. Starting with their works, there has been a paradigm shift in the development of the firebrand models, with the latest models making use of from the advanced computational techniques and resources.

Woycheese and coauthors in [310] provide a model for the lofting of spherical and cylindrical firebrands by using the plume model proposed by [26]. They also provide analytical functions for the maximum loftable diameter and the maximum loftable height in terms of the fire intensity, atmospheric wind and the fuel characteristics. Numerical experiments performed by Sardoy and co-workers [245, 244] considered also the effects of atmospheric conditions, fire properties and fuel properties on the overall firebrand behavior, providing a statistical estimate of the ground level distributions of the disk shaped firebrands. Their results highlighted that firebrands landing at *short distances* (i.e., up to 1000 m from the firebrand source) are distributed according a *lognormal* distribution.

The work [302] furnished another mathematical model to quantify the distribution and the mass of the firebrands through a Rayleigh distribution function. In an another paper, [131] present a multiphase transport model for wildfires (FIRETEC) that relied on solid physical grounds, in order to study the firebrand transport. In a more recent study, also *Martin et al.* [170] discussed in detail the underlying physical processes for firebrands generation and transport, deriving a landing distribution based on these physical processes.

Besides these statistical approaches, other kind of numerical attempts have been tried. To name a few example, some numerical models based on Large Eddy Simulation (LES) [113, 278, 280] or more in general Computational Fluid Dynamics (CFD) [299] have been produced and analyzed. Totally different approaches such as graph-based ones (the small world networks -based work of Porterie et al in [225])and individual based cellular automata models [213] also exist in the literature.

[29] present one such study based on coupled fire/atmosphere LES for predicting the short range fire-spotting. They simulate different firebrand trajectoriesto analyze the sensitivity of the flight path to different sizes of lofting particles, different release heights and wind conditions. However, they do mention the limited applicability of such models to operational use because of their high the computational costs.

Despite the presence of multiple studies focusing on the detailed aspects of the firebrand landing distributions, none of the them is able to provide a comprehensive yet versatile approach for an application to operational fire spread models. The continuing demand for the operational management tools is to provide a quick and efficient output with simple inputs but at the same time taking the most important parameters into consideration. Few operational fire spread models like FARSITE [89], BEHAVEPLUS [13] and Prometheus [291] incorporate the phenomenon of fire-spotting through the Albini's model [3, 4]. However, Albini's model provides only an estimate of the maximum distance for a spot fire and does not include any function for the ignition probability to model the spread of spot fires. The Australian wildfire simulator PHOENIX Rapidfire [281] is designed to model large fast moving fires and also includes a fire-spotting module, but the formulations for fire spread in PHOENIX are calibrated for eucalyptus forests and a generic application to other types of fuels requires a possibly resource-consuming re-calibration [226]. The new operational models like WRF-SFIRE [159] and FOREFIRE [87] are fast and allow coupling with the atmospheric models for a better representation of the initial and concurrent atmospheric conditions (generally speaking, they are intended to be modular); but lack so far any specific module to explicitly model the fire-spotting behavior.

5.4 A novel family of Wild-land Fire models

The adopted modeling approach adopted in the present work is based on the idea to split the motion of the front into a drifting part and a fluctuating part, since it relies on the core model described in detail in Chapter 3 with major adjustment for the specific wild-land fire setting [206, 126].

Drifting Part The drifting part, that can be provided by choosing an existing operational model, is here based on sharp front description of the advancing interface, that may result from the integration of several models. However, in the precise framework of this rather new model, the level-set method [250, 157] (which is also used in WRF-SFIRE model [159]) has been adopted in every work related to the application of the Model to wild-fire spread. The only notable exception is the implementation of DEVS Lagrangian Model [126].
In formulas, temporal evolution of the level set $\phi(\mathrm{x}, t) = \phi^*$ is governed by the Eikonal equation

$$\frac{\partial \phi}{\partial t}(\mathrm{x}, t) = \mathcal{V}(\mathrm{x}, t) \, \|\nabla \phi(\mathrm{x}, t)\| \,, \quad \phi(\mathrm{x}, 0) = \phi_0(\mathrm{x}), \quad \mathrm{x} \in \Omega, \; t \geq 0, \tag{5.5}$$

where $\mathcal{V}$ corresponds to the ROS parameterization that is a function of the wind field $\mathrm{U}(\mathrm{x}, t)$, orography and biomass fuel conditions, and where $\phi_0(\mathrm{x})$ is the initial condition at time 0. Fireline propagation is then assumed to be directed towards the normal direction to the front. Equation 5.5 can be derived directly from the Level Set Formulation of Chapter 1 (for instance from Equation 1.12 by setting a null external velocity field, imposing u_f to be equal to the ROS field and substituting the formula for the normal $\boldsymbol{n} = \frac{\nabla \varphi}{|\nabla \varphi|}$.

Fluctuating Part The fluctuating part, that is the result of a comprehensive statistical description of the physics of the system and includes the random effects, can be opportunely parametrized according to the underlying physics of wild-land fires to include turbulent hot-air transport and firebrand landing distance.
The statistical formulation and parametrization adopted in this Chapter to adapt the model of 3 is labeled as `RandomFront`. The chronology of this approach refers to the following papers: `v1.0` , the simplest one, which accounted for the sole turbulence without any parametrization [203, 204]; `v2.0` included turbulence and fire-spotting with literature parametrization for fire-spotting [206], while `v2.1` included turbulence and fire-spotting along with a parametrization for turbulence effects [126] using nominal values for fire spotting. In [125] `v2.2` included turbulence and fire-spotting with a first physical parametrization of fire-spotting .
Finally, in the recent [283] version `v2.3` has modified and corrected the parametrization of fire spotting with respect to the previous version.

5.4.1 Enhancing the Front Propagation model for wild-land fire applications

The model described so far in this Thesis, that is described in Chapter 3, Eqs XX – XX, gives information about a *statistical* front, given by the smooth function φ_e, that ranges in the compact interval $[0, 1]$. Is here remarked that such function results from an application of a kernel supposedly known *a priori* that includes the randomness not modeled by the macroscopic front

evolution speed $\mathbf{V}$ that otherwise would not be caught by the modeling process. Such a smooth function can be used in many ways; an example can be tracking a selected iso-contour, or using this smooth indicator function as an index about how microscopical randomness can shed uncertainty into a contour deriving from a deterministic front evolution.

Still, this formulation is somewhat incomplete in modeling complex dynamics since at each time $t > 0$, it neglects the history of the front evolution in the time interval $[0, t)$. Chronologically, the first field of application which necessitated as a new feature the history of the evolution up to time t has been wild-land fire simulation [206, 199, 200].

This has been possible thanks to a *storing function* $\psi(\mathbf{x}; t) : \mathcal{S} \to \mathbb{R}^+$ that will be described in detail in this Section.

For the sake of readability, the whole derivation of the model will be re-proposed here.

Let the domain be denoted by S, and let $\Omega \subseteq S$ represent the burnt - burning area, while $X^\omega = X + \eta^\omega$ stands for the trajectory of each active fire point as the sum of a drifting part X and the superimposed fluctuating part η^ω. The drifting part X is obtained from the output of a sharp-front wildfire propagation model, while the fluctuations in the fire-line are included through a probability density function (PDF) corresponding to the type of random process under consideration. Let the area enclosed by the drifting part be described through an indicator function $I_\Omega(\boldsymbol{x}, t) = 1$ when $\boldsymbol{x}$ is inside the domain Ω, and $I_\Omega(\boldsymbol{x}, t) = 0$ when $\boldsymbol{x}$ is outside. Using the sifting property of the Dirac Delta function, the time evolution of each active burning point can be randomized as:

$$I_\Omega^\omega = \int_{\Omega(t)} \delta(\boldsymbol{x} - \boldsymbol{X}^\omega(t, \overline{\boldsymbol{x}})) d\overline{\boldsymbol{x}}\,. \tag{5.6}$$

Considering the ensemble average of the active burning points, a new effective indicator function is defined as seen in Chapter 3:

$$\phi_e(\boldsymbol{x}, t) = \int_S I_\Omega(\overline{\boldsymbol{x}}, t) f(\boldsymbol{x}; t|\overline{\boldsymbol{x}})\, d\overline{\boldsymbol{x}}\,. \tag{5.7}$$

As already explained in detail in Chapter 3 in the last equation $f(\boldsymbol{x}; t|\overline{\boldsymbol{x}})$ represents the PDF which accounts for the random effects fluctuations. The effective indicator $\phi_e \in [0, 1]$ and an arbitrary threshold is fixed to mark points as burned, i.e., $\Omega_e(\boldsymbol{x}, t) = \{\boldsymbol{x} \in S \mid \phi_e(\boldsymbol{x}, t) > \phi_e^{th}\}$. The ignition of the fuel caused by the firebrands involves heat exchange over a sufficient period of time, hence, a sufficient delay is also incorporated in the model through an other function ψ, that is used as a *storing function.* This addition helps makes the model capable of accounting fot the history of the process, as anticipated at the beginning of this Subsection. The function ψ simulates thus the ignition of fuel by hot air and burning embers as an accumulative process over time, by the means of a *heating-before-burning* mechanism.

Since the fuel can burn following two pathways, namely hot-air heating and firebrand landing, the *resistance analogy* suggests that the resulting ignition delay can be computed as a first guess as resistances acting in parallel.

Specifically, let τ_h and τ_f be the ignition delay due to hot air and firebrands landing, respectively. Hence, their joint ignition delay τ reads

$$\frac{1}{\tau} = \frac{1}{\tau_h} + \frac{1}{\tau_f} = \frac{\tau_h + \tau_f}{\tau_h \tau_f} \,. \tag{5.8}$$

The heating-before-burning mechanism is depicted as the persistence in time of the effective fire front, i.e.,

$$\psi(\boldsymbol{x}, t) = \int_0^t \phi_e(\boldsymbol{x}, \eta) \frac{d\eta}{\tau} \,, \tag{5.9}$$

where $\psi(\mathbf{x}, 0) = 0$ corresponds to the unburned initial condition.

At each time t, all points $\boldsymbol{x}$ that satisfy the conditions $\psi(\boldsymbol{x}, t) > 1$ or $\phi_e(\boldsymbol{x}, t) > \phi_e^{th}$ are labelled as burned.

The main components of the model are portrayed in Figure 5.2

The choice of PDF $f(\boldsymbol{x}; t|\overline{\boldsymbol{x}})$ is a crucial part of the model and it is application dependent. In the case of turbulence and fire spotting modelling for wild-land fires, it is assumed that fire-spotting can be considered a downwind phenomenon which takes place in turbulent atmosphere.

The shape of the PDF is defined as follows:

$$f(\boldsymbol{x}; t|\overline{\boldsymbol{x}}) = \begin{cases} \displaystyle\int_0^\infty G(\boldsymbol{x} - \overline{\boldsymbol{x}} - l\,\hat{\boldsymbol{n}}_U; t) q(l)\, dl \,, & \text{downwind} \,, \\ \\ G(\boldsymbol{x} - \overline{\boldsymbol{x}}; t) \,, & \text{upwind} \,. \end{cases} \tag{5.10}$$

The distribution function $G(\boldsymbol{x} - \overline{\boldsymbol{x}}; t)$ is an isotropic bi-variate Gaussian which models the effect of the turbulent heat fluxes in fire propagation. On the other hand, the distribution function $q(l)$ represents the firebrand landing distribution. The strength of the turbulence around the burning fire can be calibrated through the turbulent diffusion coefficient $\mathscr{D}$, whose physical characterisation is outlined in the next Section. A precise description of the landing distributions through experimental observations is difficult due to temporal and spatial constraints. It should also be remarked that experimenting real settings may not be so feasible for forest fire in the case of strong wind-driven fire spotting, due to the related hazards to the surrounding environment. Nevertheless, the experimental results analysing the flight paths, shape and landing distributions of the firebrands have shown that the frequency of the firebrands landing in the positive direction (along the mean direction of the wind) from the source increases with distance to a maximum value and then gradually decays to zero (see [106]). The landing distributions of the firebrands have also been studied employing numerical solution of the energy balance equations [244, 113, 133]. Among the several transport models proposed in literature, both [113] and [244]

choose the lognormal density function as an approximate fit to the firebrands landing distribution. Howbeit, [302] adopt a Rayleigh distribution for the same purpose. In the model parametrized via `RandomFront`, the shape of $q(l)$ is defined by a lognormal distribution to represent the frequency profile of the fallen firebrands:

$$q(l) = \frac{1}{\sqrt{2\pi}\sigma l} \exp \frac{-(\ln l/\mu)^2}{2\sigma^2}\,, \tag{5.11}$$

where μ is the ratio between the square of the mean of landing distance l and its standard deviation, while σ is the standard deviation of $\ln l/\mu$.

5.5 Physical parametrisation of fire-spotting (`RandomFront v2.3`)

In this section, the parametrization `RandomFront v2.3` is described. This description follows the one on the recently published work by *Trucchia et al.* [283]. We recall that two main physical phenomena with random nature are investigated by the model studied in this thesis, that is turbulence effects and fire spotting. These two phenomena, following the framework of Section 5.4, depend on three parameters, namely D (turbulence effects), μ and σ (fire spotting phenomenon). In particular, turbulent heat fluxes in fire propagation is accounted for by the means of an isotropic bi-variate Gaussian function $G(\boldsymbol{x} - \overline{\boldsymbol{x}}; t)$, which appears in Equation 5.10. Turbulence is then parametrized only by the turbulent diffusivity D. Physical based values for D are given by `RandomFront`. Details can be found in Subsection 5.5.

We shift now the attention towards fire spotting, that in the model described in Section 5.4 is modeled by the lognormal distribution of the firebrands fall, in Equation 5.10. The latter distribution is specified by the parameters σ and μ.

In the following, a physical interpretation to the different aspects of the fire spotting part of the mathematical model is provided.

We recall that the firebrands generated from the vegetation face strong buoyant forces and the ones characterized by a size less than the maximum loftable one are uplifted vertically in the convective column. These firebrands will eventually rise to a maximum height until a balance between the buoyant and the gravitational forces is reached.

The firebrands are then are expelled from the column, and are successively steered by the atmospheric wind and they will travel at their terminal velocity of fall. A simplified model for the landing distance is adopted, where it is assumed that the ejection of the firebrands from the vertical convective column is a random process affected by the turbulence in the environment around the fire. The strength of the convective column, the atmospheric conditions and the dimensions of the firebrands play a vital role in governing the trajectory of the firebrands.

In this section, the landing distribution of the firebrands based on a lognormal probability function is combined with the physical characterization of the firebrand transport.

The parametrization `RandomFront 2.3` is of course simplified, in the sense that it includes only the essential ingredients to describe the firebrand transport. Firebrands are assumed to be spherical and for a particular set of concurrent atmospheric conditions and fuel characteristics (that is, for a given choice of the set of parameters) the size is assumed to be constant. This parametrization also neglects any modification in the flight of the firebrand due to rotation of firebrand or collision with other firebrands.

In literature, the maximum spotting distance is often employed as a numerical measure to assess the severity of the fire-spotting danger under different circumstances, see e.g.

References [3, 275, 276]. Given the importance of maximum-spotting distance, the mathematical model is parametrized in terms of the p^{th} percentile of a lognormal distribution as a measure of the maximum landing distance. The p^{th} percentile for a lognormal distribution is described by its location and shape parameters μ and σ respectively:

$$\mathcal{L} = \mu \exp(z_p \sigma)\,, \tag{5.12}$$

where the value of z_p corresponding to the p^{th} percentile can be estimated from the z-tables (see, for example, `http://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm`). A further hypothesis is made, that the p^{th} percentile represents the maximum landing distance for firebrands under different situations and no ignition is possible beyond this cut-off. To ascertain the value of "cut-off" percentile, it is assumed that the effective contribution of the firebrands stops to be meaningful when its probability decreases to 20 times its peak value. After that, the ability of the firebrands to cause an ignition is neglected. While this cut-off criteria is chosen rather empirically, it is remarked that a sufficiently large number (like 20) ensures that no considerable fire-spotting behavior is missed out.

For this particular distribution, the cut-off for 50th percentile lies way beyond the point denoting the 1/20th of the maximum probability. In order to define a generalized value of the cut-off percentile for all the simulation cases presented in this article, the value of z_p is chosen to be 0.45, which corresponds to the 67^{th} percentile point.

The process of fire-spotting can be roughly subdivided into generation, lofting and transport of firebrands. The generation of firebrands from a burning canopy is a process characterized by random and dynamic nature. On the other hand, the lofting and transport of the firebrands is regulated by the firebrand geometry, fuel combustion rates, plume dynamics and of course ambient wind conditions. Strong buoyant forces act over the firebrands generated from the vegetation. The firebrands with size less than the maximum loftable size are uplifted vertically in the convective column. Such firebrands rise to a maximum height until the buoyant and the gravitational forces

counterbalance each other. Once the firebrands are expelled from the column, they are transported by the atmospheric wind and they fly at their terminal velocity of fall. The several sub-processes involved in the firebrand lofting and transport may share interactions and affect the maximum spotting distances. However, an explicit modeling of the coupled processes is mnot an easy task and in most cases approximations and assumptions are adopted in order to simplify the physical processes. Some relevant progresses on fire-spotting distributions and maximum spotting distances are provided by *Tarifa et al.* in [275, 276]. In their different works the spotting distributions and the maximum spotting distances are identified with the combination of experimental and theoretical approaches. In the formulation of `RandomFront`, these existing approaches are followed. In the following a brief discussion of the different processes which are considered in the physical parametrization is provided, following [283]:

1. *Firebrand lofting*:

   (a) Vertical gas flow: In the convective column, the updraft introduced by fire lifts the firebrands. The strength of vertical gas flow U_{gas} increases with the fire intensity I and is empirically expressed as in [188]:

   $$U_{\text{gas}} = 9.35 \left(\frac{I}{H_c}\right)^{1/3} , \tag{5.13}$$

   where H_c stands for the heat of combustion of wildland fuels.

   (b) *Size of firebrands*: The convective activity inside the plume regulates the maximum size of the firebrand that can be lofted. The terminal velocities of the loftable firebrands can not trespass the vertical gas flow rate. As the vertical gas flow increases for increasing fire rate, heavier firebrands may be uplifted into the plume. From literature,we have that the maximum loftable radius for spherical firebrands can be expressed as [275, 3, 302]:

   $$r_{\text{max}} = \frac{3}{2} C_d \frac{\rho_a}{\rho_f} \frac{U_{\text{gas}}^2}{g} , \tag{5.14}$$

   where ρ_a and ρ_f are the density of the ambient air and wild-land fuels respectively, C_d is the drag coefficient and g is the gravitational acceleration.

   (c) *Maximum loftable height*: From References [302] and [310] we retrieve a parametration of the maximum loftable height for spherical firebrands in terms of the radius of firebrand r, constrained by the maximum radius of the firebrands r_{max}:

   $$\mathcal{H} = 1.46 \left(\frac{\rho_f}{\rho_a C_d}\right) \frac{r_{\text{max}}^{5/2}}{r^{3/2}} . \tag{5.15}$$

2. *Horizontal transport*:

(a) *Maximum landing distance*: Assuming the shape of the firebrands to be spherical, *Tarifa et al.* in [275] combine both experimental and theoretical approaches to describe the firebrands maximum landing distances. Based on this framework, *Wang* in [302] provides an approximation of the maximum travel distance from a vertical convective column for spherical firebrands, using the maximum loftable height $\mathcal{H}$, the meteorological mean wind $\mathcal{U} = |\mathbf{U_h}|$, where $\mathbf{U_h}$ is the horizontal wind vector field at fire-height, and the radius of the firebrands r:

$$\mathcal{L} = \mathcal{H}\,\frac{\mathcal{U}}{U_{\text{gas}}}\left(\frac{r_{\max}}{r}\right)^{1/2}. \tag{5.16}$$

3. *Ignition probability*: As described in the previous section, the probability that the fuel is ignited by burning embers is modeled by the means of the storing function ψ (5.9). Until now, it has been always assumed that the fuel conditions are homogeneous and the ignition probability depends only on an ignition delay τ. No other local variables are taken into account in this respect.

4. *Secondary fire-lines*: The secondary emissions generated during the fire-spotting are assumed as new sources of fire with a proper fire intensity. These new fires act as additional input along with the primary fire towards generation of other secondary fired. It is assumed that these new sources are capable of generating firebrands of the same size as the primary source.

Small-scale processes, like the progressive mass loss of a firebrand due to combustion, influence the fire-spotting phenomenon, by generating fluctuations of random nature in the firebrand trajectory. This fluctuations are embodied in the parametrized model by the presence of probability density function for for the landing distance.

Finally, the above mentioned large-scale subprocesses under lofting and transport mechanisms are linked through Eq. (5.12) to obtain the physical parametrization of μ and σ of the lognormal distribution. Using Eq. (5.16) and Eq. (5.15), we express the shape and location parameters as follows:

$$\sigma = \frac{1}{2z_p}\ln\left(\frac{\mathcal{U}^2}{rg}\right), \tag{5.17}$$

$$\mu = \mathcal{H}\left(\frac{3}{2}\frac{\rho_a}{\rho_f}C_d\right)^{1/2}. \tag{5.18}$$

With the adoption of such parametrization of μ and σ the governing parameters for lofting are represented by μ, while transport mechanisms are represented by σ. Insights on the parametrization of μ are given in next Subsection 5.5, while other observations on the parametrization of σ are given in Subsection 5.5.

Parametrization of μ

We hypothesize that the definition of μ is sufficiently elaborated to cover all the essential input parameters that are needed *to describe the lofting mechanism of the firebrands inside of the convective column*. The relative density ρ_a/ρ_f as well as atmospheric drag do quantify the buoyant forces experienced by the firebrand. For this reason, it is appropriate to include these quantities into the definition of μ for varying fireline intensity. Substituting maximum loftable height from Eq. (5.15) in μ gives:

$$\mu = 3.52 \times 10^5 \left(\frac{\rho_a}{\rho_f}\, C_d\right)^2 \left(\frac{I}{H_c}\right)^{5/3} r^{-3/2} g^{-5/2}\,. \tag{5.19}$$

The radius of the firebrand r and the fuel density are important ingredients to determine the height of the lofted firebrands.

In [284] another equivalent formulation for μ has been used, that is

$$\mu = H_p \left(\frac{3\,\rho_{\mathrm{a}}\, C_{\mathrm{d}}}{2\,\rho_{\mathrm{f}}^*}\right)^{1/2}, \tag{5.20}$$

where H_p [m] represents the plume height, which is related to the maximum loftable height $\mathcal{H}$ via the relation $\mathcal{H} = \lambda\, H_p$, and where $\rho_f^* = \rho_f/\lambda^2$ [kg m^{-3}] is the biomass fuel density that accounts for the correlation factor λ between smoke plume height and maximum allowable height for firebrands. We adopt the analytic formulation of H_p with respect to the fire-line intensity I used in [258], i.e.

$$H_p = \alpha_H\, H_{\mathrm{ABL}} + \beta_H \left(\frac{I}{dP_{\mathrm{f0}}}\right)^{\gamma_H} \exp\left(\delta_H\, \frac{N_{\mathrm{FT}}^2}{N_0^2}\right), \tag{5.21}$$

where α_H, β_H, γ_H and δ_H are empirical constant parameters, P_{f0} [W] is the reference fire power ($P_{\mathrm{f0}} = 10^6\, W$), H_{abl} [m] is the height of the atmospheric boundary layer (ABL). The subscript FT stands for *free troposphere*.

Parametrization of σ

The second parameter of the lognormal distribution σ, given by 5.17 is hypothesized to define the transport of firebrands under the effect of horizontal wind after ejection from the convective column. The definition of σ includes a dimensionless ratio $\mathcal{F} = \mathcal{U}^2/(rg)$ which is analogous to the Froude number, which quantifies the balance between inertial and gravitational forces experienced by the firebrand. All firebrands with $r \leq \mathcal{U}^2/g$ can be transported by the horizontal wind.

Parametrization of D

We assume that all turbulent processes are represented in the forward model through the standalone turbulent diffusion coefficient D. We only consider turbulent fluctuations, implying that the estimation of D is independent of the wind U. Since we consider a flat terrain and an extension of the wildland fire that is not limited to the computational domain Ω under

consideration, we assume horizontal isotropy. Even though an exact estimation of D is beyond the scope of the present study, a quantitative estimation of D is required to carry out sensitivity analysis related to turbulence and fire-spotting. D corresponds to the turbulent heat convection generated by the fire. The ratio between the total heat transfer and the heat molecular conduction is widely known as the Nusselt number, $\mathrm{Nu} = (D + \chi)/\chi$, where χ is the air thermal diffusivity. The relation between the Nusselt number Nu and the Rayleigh number Ra (i.e. ratio between convection and heat conduction) is given by the experimental correlation $\mathrm{Nu} \simeq 0.1\,\mathrm{Ra}^{\beta_{\mathrm{RA}}}$ with $\mathrm{Ra} = \gamma\,\Delta T\,g\,h^3/(\nu\chi)$ (γ is the thermal expansion coefficient, ΔT is the temperature difference in the convective cell, h is the dimension of the convective cell, g is the gravity constant and ν is the kinematic viscosity). Thus, the turbulent diffusion coefficient D is computed in this work as

$$D \simeq 0.1\,\chi \left[\frac{\gamma\,\Delta T\,g\,h^3}{\nu\chi}\right]^{1/3} - \chi, \tag{5.22}$$

with $\chi = 2 \cdot 10^{-5}\,\mathrm{m^2\,s^{-1}}$, $\gamma = 3.4 \times 10^{-3}\,\mathrm{K^{-1}}$, $g = 9.8\,\mathrm{m\,s^{-2}}$ and $\nu = 1.5 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$. To define the range of variation of D, we introduce some assumptions. The heat transfer is considered in the horizontal plane, perpendicular to the vertical "heating wall" embodied by the fire; the length scale of the convective cell is assumed to be $h = 100\,\mathrm{m}$ [126], and ΔT varies from 800 to 923 K. Note that the relation between the Rayleigh number and the Nusselt number is highly sensitive to the scaling exponent β_{RA} due to the power-law. Libchaber's experiments found $\beta_{\mathrm{RA}} \approx 2/7$ instead of 1/3. In [191], the relation $\mathrm{Nu} = 0.146\,\mathrm{Ra}^{0.299}$ is proposed for $\mathrm{Ra} > 5 \cdot 10^7$; for higher values of Ra, it is recommended to use $\beta_{\mathrm{RA}} = 0.3$.

For all the simulations presented as a response analysis of the model in this article, the value of the turbulent diffusion coefficient $\mathscr{D}$ is assumed to be $0.15\,\mathrm{m^2 s^{-1}}$.

The simple design of the physical parametrisation makes the model computationally less expensive and the requirement of defining only few vital parameters to execute any simulation also serves as an added advantage to the operational users. Static and dynamic input parameters of the model are reported in Table 5.1.

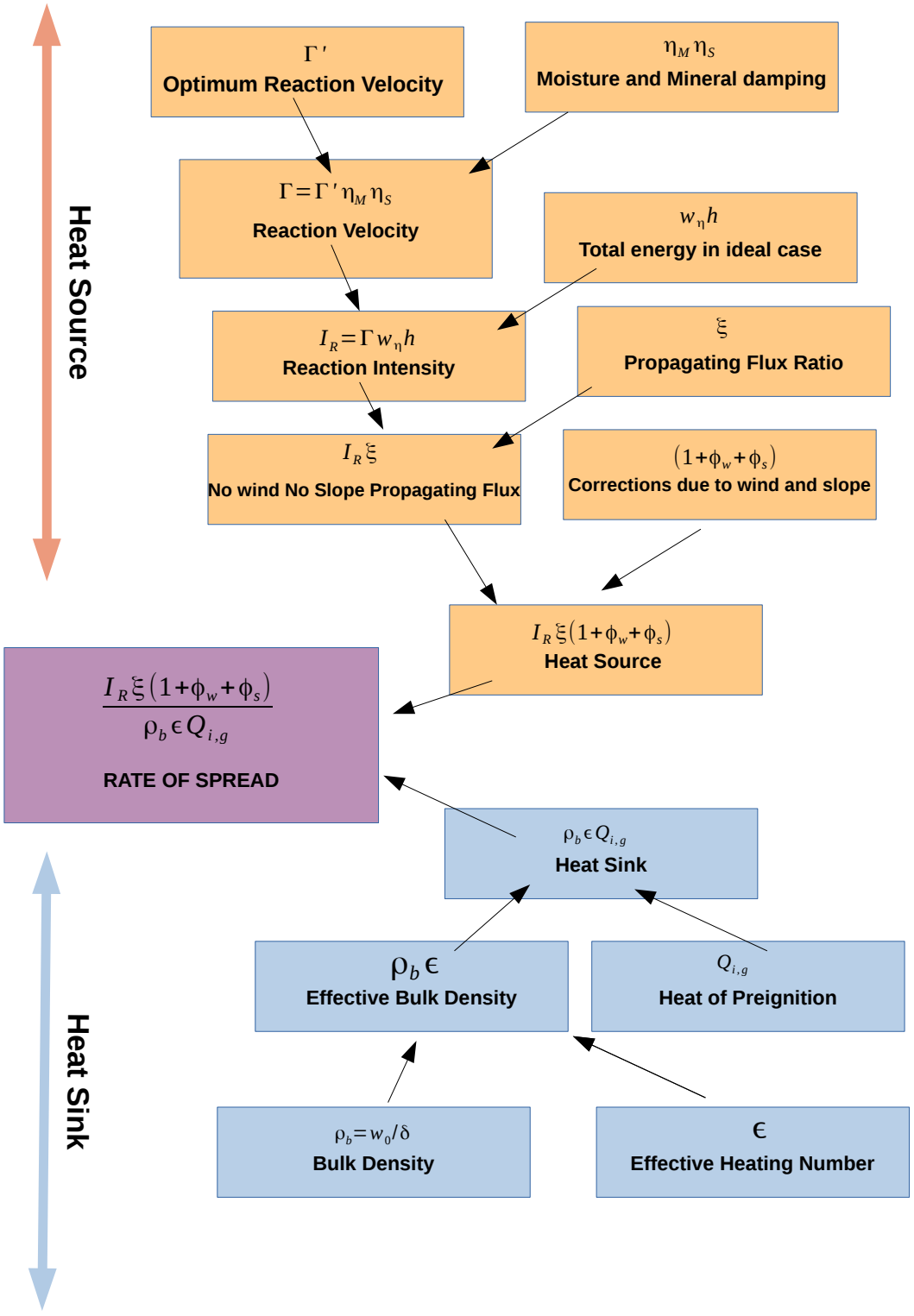


Figure 5.1: Adaptated from [12]

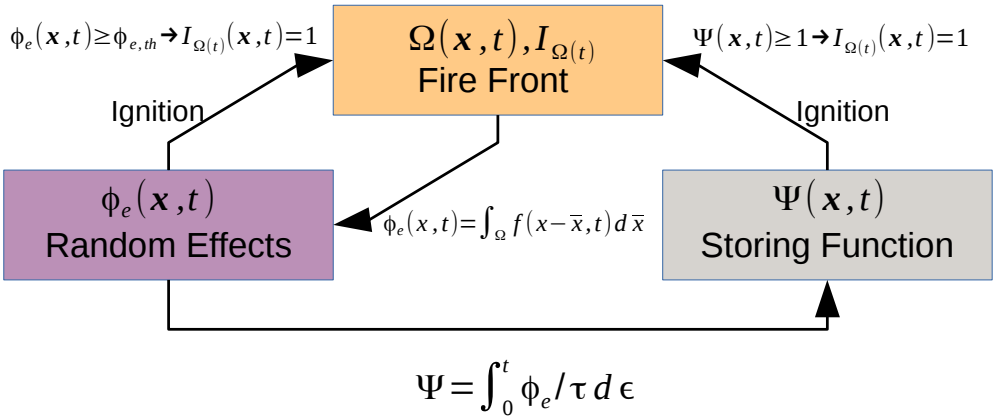


Figure 5.2: Schematic view of the front propagation model for wild-land fire propagation application.

Table 5.1: List of symbols.

Model quantities	**Units**
ϕ_e, effective indicator	-
ψ, ignition function	-
$\boldsymbol{x} = (x, y)$, horizontal space variable	m
t, time	s
f, probability density function of the random processes	m^{-2}
$G(\boldsymbol{x}; t))$, isotropic bi-variate Gaussian probability density	m^{-2}
$q(l)$, lognormal distribution of firebrand landing	m^{-1}
Parameters	
Static Parameters	**Value**
μ , parameter of $q(l)$	–
σ, parameter of $q(l)$	–
$\mathscr{D}$ turbulent diffusion coefficient of G	$0.15\mathrm{m}^2\mathrm{s}^{-1}$
ρ_a, density of air	1.2 kgm^{-3}
ρ_f, Density of wildland fuel (*Pinus Ponderosa*)	542 kgm^{-3}
C_d, drag coefficient	0.45
z_p, p-th percentile	0.45
g, acceleration due to gravity	9.8 ms^{-1}
H_c, heat of combustion of wildland fuels	18620 kJkg^{-1}
ω_0, oven-dry mass of fuel	2.243 kgm^{-2}
H, fuel low heat of combustion	22000 kJkg^{-1}
Dynamic Parameters	**Units**
$\mathbf{U} = (U, V, W)$, wind vector at fire-height	ms^{-1}
$\mathbf{U_h} = (U, V)$, horizontal wind vector field at fire-height	ms^{-1}
τ, ignition delay of firebrands	s
I, fire-line intensity	MWm^{-1}
U_{gas}, Vertical gas flow	ms^{-1}
r, radius of spherical firebrand	m
$r_{\max}$, maximum loftable radius for spherical firebrand	m
$\mathcal{H}$, maximum loftable height for spherical firebrands	m
$\mathcal{V}_{ros}$, rate of spread	ms^{-1}

5.5.1 Tables of Symbols

5.6 Computational Implementation of the Model

Three distinct computational frameworks have been employed so far in the simulations of the mathematical model of random front propagation, along with the several stages of model development (and of RandomFront parametrization growth in complexity). They will be listed here with a short description.

5.6.1 LSFire+

The code `LSFire+` is developed in C and Fortran where the model `RandomFront 2.3` acts as a post-processing routine at each time step in a LSM code for the front propagation implemented by the means of `LSMLIB` [53] and the ROS is computed by using the library `FireLib` [28]. The

numerical library `LSMLIB` is written in Fortran2008/OpenMP and propagates the fire-line through standard algorithms for the LSM, including also Fast Marching Method algorithms.

5.6.2 ForeFire

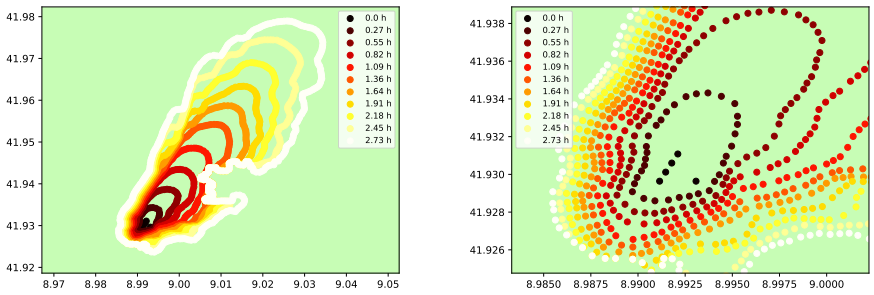
`ForeFire` (`https://github.com/forefireAPI/firefront`) [87] is a Lagrangian DEVS based wild-land fire simulation model; for technicalities on DEVS see Chapter 1.
`ForeFire` is an open-source wildfire simulator developed by University of Corsica, France [84] to serve as a basis for an operational simulator. It is flexible in the sense that it may include new physical feature with relative ease. ForeFire is based on a Lagrangian front tracking technique. Shifting the attention towards fire models, it integrates the so-called Balbi model for the evaluation of the RoS (see [19]).
We recall that the temporal scheme used to simulate the fire perimeter in wild-land fire simulator ForeFire is based on the Lagrangian DEVS scheme [315]. DEVS handles the time advancement in terms of the increment of physical quantities instead of a discrete time step. The front that constitutes the output at each time step is a polygon whose marker points have real (that is, non-discrete) coordinates instead of being necessarily located on nodes of a regular mesh. Also fire-break zones (e.g. rivers, roads ...) are represented as arbitrary shaped polygons.
Coupling the front tracking method with DEVS allows for the utilization of time as a continuous variable; by doing so, the limitation of establishing a trade off between the temporal resolution and the scale of the simulation according to the Courant–Friedrichs–Lewy (CFL) condition is removed. ForeFire original model is developed without any description of underlying random processes such as turbulent heat transfer and fire-spotting phenomena. Without further modeling or coupling with ad-hoc sub-models, it cannot simulate the evolution of the fire across fuel break zones.
In the synopsis of the random front model object of this Thesis, ForeFire has been adopted to show the performance and the capabilities of the front propagation model which is the object of this Thesis in a full-fledged Lagrangian wild-land fire simulator. It has been tested after the preliminary attempts with Eulerian LSM simulators (LSFire+) [203, 204, 199, 200, 206, 207] and the main output of the related research lies in the work of *Kaur et al.* [126].
The main highlight of the work of *Kaur et al.* is the comparison of the performance of a Lagrangian (ForeFire) and an Eulerian (LSFire+) moving interface method for wild-land fire propagation. Through rather simple and idealized numerical experiments the potential applicability of the proposed formulation to DEVS has been explored. The result obtained with DEVS-based ForeFire have been compared with the ones provided by LSFire+. The RandomFront implementation on Forefire highlights that DEVS based wildfire propagation model qualitatively improves its performance (e.g., reproducing flank and back fire, increase in fire spread due to pre-heating of the fuel by hot air and firebrands, generation of secondary fires, fire propagation

through no fuel zones, ...) when random effects are included through the model parametrized with RandomFront. The resulting comparison showed that DEVS and LSM based wildfire models are comparable, while they do exhibit some technical differences in the geometrical construction of the direction of propagation (mainly due to the different construction of the normal vector from LSM to Lagrangian formalism)

ForeFire library is written in C/C++/Java and Fortran bindings for a UNIX compatible environment. It is compiled using the build platform SWIG (`http://swig.org/`) . NetCDF library with legacy C++ interface is necessary in order to build ForeFire. Moreover, the Python tool SCons (`http://www.scons.org/`) is used while making the library. A python based interface is available to perform the simulations with software calls directed to the main code. In addition, a command-line interpreter to launch simulation is available. An on-line version of `ForeFire` is available at `http://forefire.univ-corse.fr/`

Example of ForeFire simulation

To exemplify a launch of ForeFire a simple simulation is performed. The fire is started at 10:00 a.m. and the simulation window ends at 13:00. The fire starts at 41°55'48.0"N 8°59'24.0"E, Bastelica, Corsica Island (France). The wind direction has been uniformly set to 220° and with constant magnitude of 32 km/h.



(a) (b)

Figure 5.3: Output of `ForeFire` simulation in Bastelica, Corsica Island, France. The different contours of the fire perimeter are represented with different colors from 0 to 2.73 elapsed hours. Panel 5.3b portrays a zoom of 5.3a in order to better appreciate the discrete markers related to each fire contour.

5.6.3 WRF-Sfire

In order to prove the viability of the proposed formulation within an operational code, we have implemented `RandomFront 2.3` in the framework of the `WRF-Sfire` simulator ([56, 158]).

`WRF-Sfire` is a coupled fire-atmosphere model, which operates in the computational environment

of a well known public domain numerical weather prediction model : `WRF v3.4` (Weather Research and Forecasting; [255]; `http://wrf-model.org/users/users.php`)

The fire module embedded into `WRF` simulates a surface fire and takes into account a two-way coupling with the atmospheric model. The near-surface winds from the atmospheric model are interpolated on the fire grid and are used along with fuel properties and local terrain gradients to compute both the ROS and the outward front-direction, that are further used as an input to the front propagation routines through a LSM scheme. Fuel consumption is responsible for the release of sensible and latent heat into the lowest layers of the atmosphere and this has a role in the computation of the boundary-layer meteorology. Recently, the model has been equipped with a fuel-moisture sub-model and a chemistry sub-model (`WRF Chem`), which contribute towards reproducing and investigating the effects of the fire-atmosphere coupling. A schematic representation of WRF Sfire modular structure is given in Figure 5.4.

Coen et al., in [56] point out that fires generally start from a horizontal extent much smaller than the size of the fire mesh-cell. The same may be argued for the secondary ignitions related to fire-spotting phenomenon. In this respect, [56] propose and explain in detail an algorithm for a punctual or line ignition that actually runs on `WRF-Sfire`. The purpose of this algorithm is two-fold: *i*) it guarantees from a physical point of view that the ignition starts at sub-grid scale without generating unrealistically large initial heat flux and an accelerated ignition; *ii*) this procedure is numerically robust because it is fully integrated into the representation in terms of a *signed distance function* of the LSM [250].

In the formulation described in this Chapter, a punctual ignition occurs whenever the condition $\psi(\mathbf{x}, t) \geq 1$ holds true. This procedure is not computationally viable, as pointed out by *Trucchia et al.* in [283]. In practice, a threshold distance $R_{\text{th}} = 200\,\text{m}$ is set for the sake of separating every pair of punctual ignitions. Let $\mathcal{P}$ be the set of point-wise fire-spotting ignitions, the actual algorithm performed at each time-step within `WRF-Sfire` model is reported in Algorithm 1.

Algorithm 1 Algorithm for Point-Wise uniformly ignition due to Fire-Spotting, adapted from [283]

for $\mathbf{x}_i \in$ Grid **do**
  if $\psi(\mathbf{x}_i, t) > 1 \wedge \mathbf{x}_i \notin \Omega(t)$ **then**
    if dist$(\mathcal{P}\ , \mathbf{x}_i) > R_{\text{th}}$ **then**
      New fire-spotting ignition in $\mathbf{x}_i$ **and**
      $\mathcal{P} \leftarrow \mathcal{P} \cup \{\mathbf{x}_i\}$
    end if
  end if
end for

5.6.4 Computational resources

The computational resources used for the simulations related to the topics presented in the present Chapter are listed below. Simulations with `WRF-Sfire` have been perfomed on a Intel(R)

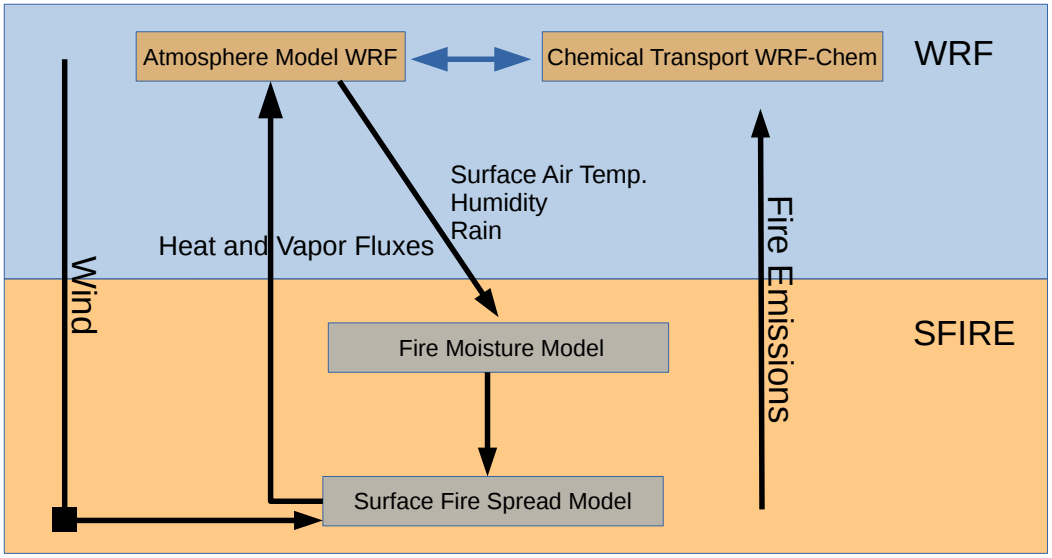


Figure 5.4: Scheme of WRF-Sfire modular structure

Core(TM) i5-4310M 2.70GHz CPU laptop with 8 GB of RAM. Each simulation that spanned 20 physical minutes took about 100 minutes of computational time.

Simulations with `LSFire+` are performed over the cluster HYPATIA of BCAM, Bilbao, using `OpenMP` shared memory parallelism, running over 24 cores inside of an `Intel(R) Xeon(R) CPU E5-2680 v3 2.50GHz` node with 128GB of RAM. The computational time for each simulation, that spanned 140 minutes of physical time, was about 45 minutes.

5.7 Response analysis of `LSFire+` and outputs of `WRF-SFire`

Quantity	Unit of measurement	First Test Case	Second Test Case
$\mathscr{D}$	m^2s^{-1}	0.15	0.15
U	ms^{-1}	10	$2 \div 26$
I	MWm^{-1}	$5 \div 100$	50
r	m	0.015	$0.015 \div 0.03$
τ	s	1	1

Table 5.2: Values of the main parameters for numerical simulations performed with `LSFire+`.

5.7.1 Discussion for `LSFire+`

Simulation Setting The simulated domain for the response analysis in `LSFire+` isa rectangle of dimensions $[0\text{m}, 6000\text{m}] \times [0\text{m}, 6000\text{m}]$. The simulations start at time $t = 0\,\text{min}$ and end at time $t = 140\,\text{min}$. The grid spacing is $\Delta x = \Delta y = 20\text{m}$. At time $t = 0\,\text{min}$ the initial fire-line is a circle of radius 180 m centered at $\mathbf{x}_c = (720\text{m}, 3000\text{m})$. The horizontal wind has been assumed in this simulation set-up as a constant field parallel to the vector $\mathbf{j} = (1, 0)$ and with modulus $|\mathbf{U_h}| = |(U, V)|$.

Quantifying randomness: β_e parameter For the response analysis separate set of simulations are carried and the response is evaluated through a parameter β_e, which describes the effective increase in the burned area:

$$\beta_e = (\mathbf{x}_{\text{random}} - \mathbf{x}_{\text{no-random}})/\mathbf{x}_{\text{no-random}}\,. \tag{5.23}$$

β_e thus stands for *the increase in the number of burned grid points with respect to the simulation when no random effects are considered.*

Fire Spotting vs Turbulence effects

In `LSFire+` the effects of fire spotting occur in conjunction with turbulence and both processes contribute towards the fire propagation, as they are both included in the formulation of the PDF of Equation 5.10 It is therefore difficult to separate the effects of both processes individually. However, a comparison of the increase in burned area due to turbulence and turbulence plus firespotting is portrayed in Fig. 5.5a.

In that plot, the number of burned grid points is plotted versus the elapsed simulation time for two simulations corresponding to different setting of the PDF of Equation 5.10. The first case accounts only for turbulence (that is, $f = G$, while the second is the turbulence plus fire-spotting (that is, f is the one expressed in 5.10). All the simulation parameters remain the same in both the simulations. In the first part of the simulations, the line plots for both the simulations overlap significantly. This means that fire-spotting has no visible contribution in the initial stages of the fire. However, after 50 minutes, the effect of fire-spotting picks up and the burned area increases

rapidly with respect to the turbulence-only setting. At $t = 130 - 140$ min. the number of burnt pixel due to fire spotting plus turbulence is almost three times the effect of turbulence alone.

Response analysis to fire intensity

An increase in the fire intensity causes an increase in the burned area also on the non-random model (such parameter is involved in Eqn (5.24), accounting for the definition of the ROS). At the same time,changes in I also affects the fire-spotting behavior due to the definition of μ (Equations 5.20, 5.21) . The parameter β_e allows us to identify the contribution of fire-spotting towards the fire-propagation. Figure 5.5b shows the change in the burned area under the combined effect of turbulence and fire-spotting when increasing the fire-intensity. The two plots correspond to the same wind speed ($10\,\mathrm{ms}^{-1}$) with two different choices for the firebrand radii, i.e., 0.015 m and 0.030m. According to the physical parametrization of the lognormal shape parameters μ and σ shown in detail in Section 5.5 for the set of simulations of Figure 5.5b, (increase in fire intensity I), the parameter μ varies while parameter σ remains constant. For both radii, an increase in the fire intensity is followed by a sharp increase in burnt area for low fire intensities. A zoom-in of this sharp rise is also shown in Figure 5.5c. For smaller firebrand radius, the fire-spotting effect exhibits a slight saturation between 15-25 MWm^{-1}. However, after further increase in the fire intensity, the contribution of fire-spotting remains positive until $60\,\mathrm{MWm}^{-1}$ when then it saturates again. Any further increase in the fire-intensity causes a decreasing importance of the fire-spotting. When the larger firebrand radius is considered, a similar behavior is reproduced for fire intensities less than $15\,\mathrm{MWm}^{-1}$, but with further increase in I, the contribution from fire-spotting takes a dip before it start increasing again.

The zoom-in of the response analysis given by Figure 5.5c is important as literature shows that for fire intensities around $8\mathrm{MWm}^{-1}$ for vegetation type *Pinus ponderosa* are already classified as high "severity class" [50]. For fire intensities up to 10 MWm^{-1} a rapid increase in the fire-spotting behaviour up to $4\mathrm{MWm}^{-1}$ is observed. Any further increase in the fire intensity has a positive effect on the ROS and on the propagation of the main fire. That is, deterministic ignition prevails on ignition by random effects. Because of that, a lower effect on the fire-line due to fire-spotting is observed. It is also worth attention the fact that for weak fires (less than $1\mathrm{MWm}^{-1}$), the fire-spotting mechanism seems to be independent of the firebrand radius.

To better explain the observation made in the previous paragraph, the lognormal distribution for selected values of I are plotted in Figure 5.6 . These lognormal distribution plots show general trend in the distribution with varying μ but keeping a constant value for σ. With an increase in I (or μ, given the parametrization of Section 5.5), the maximum probability increases, but at the same time the distribution becomes increasingly skewed. For these particular choices of parameters μ and σ, the skewness is more pronounced for fire intensities greater than $20\,\mathrm{MWm}^{-1}$. For lower

values of I (less than 20MWm^{-1}), the lognormal distribution exhibits a slow decay and the probability of a "long-range" ignition increases. This may explain the large initial increase in the contribution of the fire-spotting behavior for low range of fire-intensities. As the magnitude of the corresponding peak value is also decreasing with increasing μ (or I), this "long-range" ignition probability might have a positive influence only until a threshold. This threshold is the range where parameter β_e in Figure 5.5b takes a dip or exhibits a saturation. Beyond this point, the effective contribution of "long-range" probability is not relevant anymore, while the contribution of the"short-range" ignitions becomes more and more important. The gradual increase in the effective burned area for both firebrand sizes (for fire-intensities greater than 30 MWm^{-1}) can be explained by such reasoning. For large values of μ the lognormal distributions tend to be quite similar, even if still become increasingly skewed (see Figure 5.6b). In the limit of increasing skewness, the "short-range" probability will lie so close to the main fire-line that the real contribution of fire-spotting to the fire profile should decrease. In [283] it is observed a similar trend only for the small firebrands. It has been hypothesized that such behavior may be replicated by larger firebrands too, but only outside of the specific range of simulations of such paper.
The effect of the fire intensity on the fire-spotting behavior can be explained also by analyzing the physical parametrization provided in Section 5.5. According to `RandomFront` parametrization, an increase in the fire intensity increases the maximum loftable height. The firebrands are therefore ejected from more elevated heights. Higher release height contributes to the increase in the firebrand activity at longer distances and the initial increase in the fire perimeter may thus follow this observation. At the same time, the increase in the firebrand ejection height over constant wind conditions causes the firebrands to travel longer in the atmosphere before reaching the ground. The growing travel time for a firebrand favors its combustion and then firebrands reach the ground with a lower temperature (less "long-range" ignition probability) compared to its counterpart ejected at lower heights.
Lower temperature of the firebrands leads to a lower heat exchange with the unburned fuel, compromising thus the successful ignition by the falling ember. Because of that, after reaching an area of maximum activity, the effective contribution of the "long-range" firebrands under same atmospheric conditions decreases with increasing fire activity. This physical reasoning may thus explain the initial dip/saturation in the fire-brand activity. Moreover, it can be stated that the "short-range" firebrands have larger energy and becomes the dominant cause of the fire-spread. This range can be considered as the transition time when the "long-range" firebrands become less important but the "short-range" activity becomes more noticeable. For the heavier fire-brands a similar behavior is expected, though the maximum loftable height under identical wind and fire conditions is lower due to its weight. Lower loftable height decreases the maximum landing

distance and the magnitude of the burned area is less. This is also evident from the lower magnitude of β_e in the results.

Response analysis to wind speed

Figure 5.5d highlights the simulation results with an increasing value of the wind velocity over constant fire intensity ($50\,\mathrm{MWm^{-1}}$). Similarly to the previous analysis for I, results for two different radii (0.015 m and 0.030 m) are presented. The fire-spotting mechanism over varying wind speeds shows similar behavior for the different sizes of the firebrands. For both radii indeed, the effective burnt area does increase with the increasing wind speed but after a threshold value, an increase in the wind speed leads actually to a decline in the effective burnt area. The plot for $r = 0.015\,\mathrm{m}$ shows that after a value around $10 - 11\,\mathrm{ms^{-1}}$, the contribution of the firebrands decreases. In a similar way, in the red plot for radius $r = 0.030\,\mathrm{m}$, the effective increase in the burned area follows an identical pattern but the total increase in the burned area is lesser in magnitude and shows a saturation around the maximum before it start decreasing (at around $22\mathrm{ms^{-1}}$). It is understood that for bigger firebrands the maximum value of fire spotting activity occurs at higher wind speeds and such strong activity is supported by a broader range of wind speed values.

The lognormal distributions for a selection of wind speed values, fixing r and I, are plotted Figure 5.7. These two plots show two different aspects of the response behavior of the lognormal distribution when parameter σ is varying but parameter μ is constant. Firstly, from the Figure 5.7a it emerges that with increase in the wind speed (increasing σ, constant μ) the lognormal distribution shifts towards the left but the tails decrease slower, and the distribution is wider around the maximum. The increase in the width of the lognormal distribution leads to a larger area of "long-range" and "short-range" probability of ember fall. This explains the initial rise of the burned area in Figure 5.5d. At the same time, the increasing wind speed is also responsible for a decrease in the magnitude of the peak probability, hence beyond a certain threshold, the overall contribution from fire-spotting for a localized area starts decreasing. The saturation in the fire-spotting behaviour can be explained by the second aspect of the lognormal response. In fact, Figure 5.7b reports that after a certain threshold of U (that, being r fixed, uniquely determines σ in a monotonic fashion), the lognormal distributions become more and more similar. This threshold depends upon the value of parameter μ, and for smaller values of μ (smaller r or I), we have an early appearance of such threshold. As the lognormal distributions tend to have similar probability distributions with increasing wind speed (Figure 5.7b), the contribution from fire-spotting also becomes similar and it may explain the saturation effect of bigger firebrand radius reported in Figure 5.5d.

As we discussed for the effects of fire intensity, the response of the model over different wind velocities can be interpreted with physical reasoning following the `RandomFront` parametrization of σ given in Section 5.5. In terms of the physical quantities used in the parametrization, it can be stated that strong winds can carry away the firebrands at longer distances from the main ignition source and thus result in a larger fire perimeter (increase in "long-range probability"). Historically it has been reported that strong winds coupled with very dry conditions made the perfect setting for long range fire-spotting. Strong wind speeds can loft the smaller firebrands to longer distances but with an increasing wind speed the combustion process quickens and the firebrands reach the ground with less temperature. These physical arguments may therefore explain the reduced effect of fire-spotting on the burned area over very high wind conditions. Contrarily, a larger firebrand size can sustain longer in the atmosphere. Their "long-range" probability is hence relatively higher than the one for smaller firebrands. This explains the occurrence of maximum burned area at 15ms^{-1} instead of 12 ms^{-1} for 0.015m radius. The heavier mass of bigger firebrands confines their flight to shorter distances compared to the lighter firebrands and thus a lower magnitude of the burned area can be seen. The longer saturation in the fire-brand activity for the larger firebrand may be related to the fire-brand's ability to stay airborne longer without burning out.

5.7.2 Discussion of the test case with `WRF-Sfire`

Experimental setting

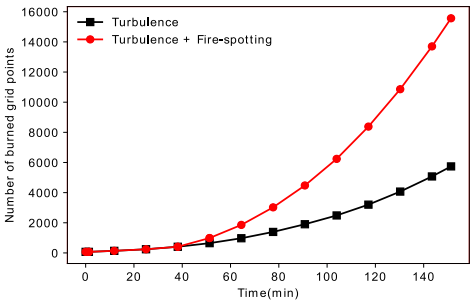
The implementation of `RandomFront` in `WRF-Sfire` has been described in the recent work [283]. The outputs of this Section mainly refer to the that work. In that numerical experiment, a slight modification of the `hill` test case is considered (`https://github.com/openwfm/wrf-fire/blob/master/wrfv2_fire/test/em_fire/hill/namelist.input.hill`). To simplify the underlying dynamics,retaining at the same time the fire-atmosphere coupling, the hill is removed (by setting `fire_mountain_type` $= 0$) and we have therefore a square-grid simulation over a flat domain with side 2.5 km. The horizontal atmospheric grid-spacing at terrain-height is 60 m, and the fire spread grid-spacing is 15 m. The initial wind field at the fire height is
$(U(\mathbf{x}; t = 0), V(\mathbf{x}; t = 0)) = (-6.4\,\text{ms}^{-1}, -3.6\,\text{ms}^{-1})$ in all the points of the domain. The simulation starts at the instant $t = 0$ min and ends at $t = 20$ min. The fire-line is located initially along the segment given by the extrema $(1900\,\text{m}, 1500\,\text{m})$ and $(1900\,\text{m}, 1800\,\text{m})$.
The fuel has been set equal to fire *Type 9*, i.e., *FM 9 Hardwood litter* according to Anderson classification [9]. This fuel type may model a terrain covered by *Pinus ponderosa* trees. The radius of the spherical embers has been fixed to $r = 12.5\,\text{mm}$ following a size considered by [161] with the same vegetation.
For what concerns the integration in `WRF-Sfire` of the formulation proposed in this Chapter, the fire-line intensity I and the wind field are computed by means of the `WRF-Sfire` model. This

allows for a space and time varying field of both parameters σ and μ according to formulas (5.17) and (5.19), respectively. In particular, the latest advancements of `WRF-Sfire` documented in [158], furnish a spatial representation of the potential-fire characteristics from which a field extension of the fire-line intensity I is available. This allows to have a space and time varying field of μ for the fire-spotting algorithm. Parameters $\mathscr{D}$ and τ are set as $\mathscr{D} = 0.15\,\mathrm{m^2 s^{-1}}$ and $\tau = 8\,\mathrm{s}$, without using specific estimations given by `WRF-Sfire`.

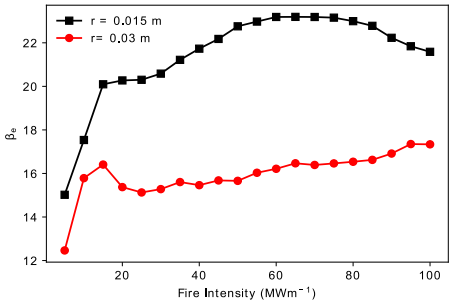
Discussion of results

Figures 5.8-5.10 display the simulation results. In each figure the fire front is reported by a dashed line at the time steps corresponding to $t = 6\,\mathrm{min}$, $t = 10\,\mathrm{min}$ and $t = 20\,\mathrm{min}$. The selected instants permit us to track the main fire alone, the generation of a secondary fire and the multi-generation of secondary fires, respectively.
In Fig. 5.8 the evolution of the fire-line is shown reporting the three components of the (tridimensional)wind field; Fig. 5.9 shows the relationship between parameter μ and the fire intensity field; Fig. 5.10 shows the relationship between parameter σ and the squared norm of the horizontal wind. The latter two relationships follow straightly from `RandomFront` parametrization.
In general, it can be observed that the fire-line propagation is "pulled" in the direction of the maximum value of the squared norm of the horizontal wind (see right column in Fig. 5.10). This direction is induced by the fire itself as a feedback on the weather as it is shown by the patterns of the atmospheric observables when secondary fires are generated by fire spotting effects.
The geometrical profile of the fire perimeter plays an important role in determining the behavior of fire-spotting. In particular, the asymmetry in the fire perimeter at 20 min along the prominent direction of propagation, makes the first secondary fire show up in the top-left part of the domain. As time (and fire-activity as well) increases, the differences between the maximum value of the squared norm of the horizontal wind and its surroundings increase and the fire-line becomes symmetric with respect to the main direction of the wind.
This has direct implications in the fire-spotting action, as the new secondary fires appear increasingly aligned along the main direction of propagation.
The secondary fires are equally important as the primary fire in influencing the weather around the fire. The plots clearly show the influence of fire-atmosphere coupling, and a feedback dynamics from secondary fires to primary fire can be also observed. The secondary fires affect the wind (see Fig. 5.8) and also the parameter σ (Fig. 5.10), which implies a change in fire-spotting characteristics for further ignitions. Another thing worth interest is that the first secondary fire occurs at a distance of almost 1500 m from the main fire. This observation supports the viability of the proposed formulation to simulate fire-spotting mechanism in studies of medium-to-large scale fire events.

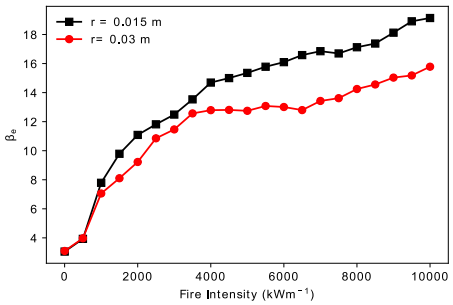
Nonetheless, even if the implementation in `WRF-Sfire` allows for a more comprehensive picture including the physical features of a multi-scale and multi-physics process, the complexity of the model, the number of parameters and the numerical cost increase as well.



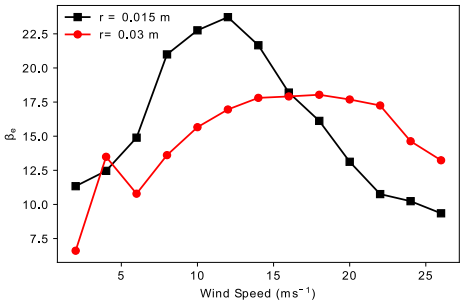
(a) Comparison of Turbulence and Turbulence + Fire Spotting effects on the number of burnt grid points versus elapsed simulation time. Adapted from [283].



(b) Gain factor β_e versus Fire Intensity

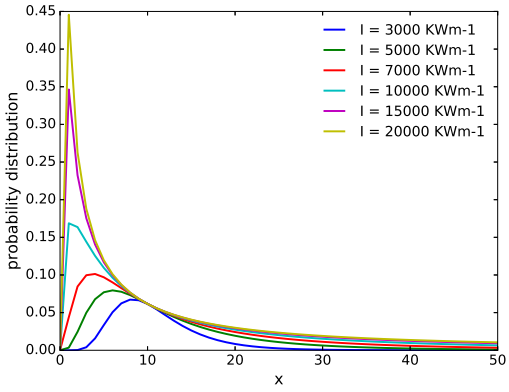


(c) Gain factor β_e versus Fire Intensity (detail)

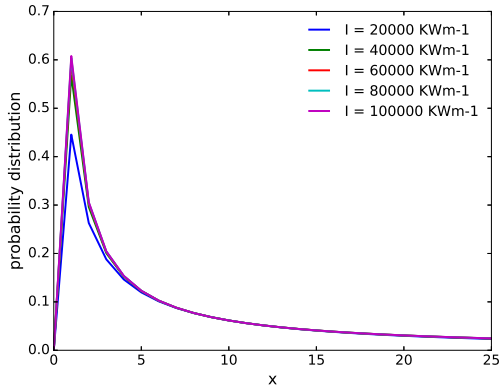


(d) Gain factor β_e versus horizontal wind speed magnitude.

Figure 5.5: Results of the response analysis of `LSFire+`.

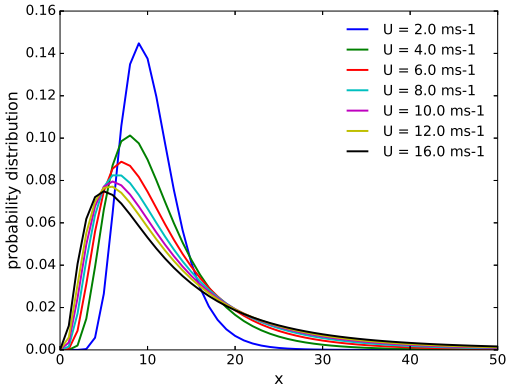


(a) $I \in [3000, 20000] KWm^{-1}$; Picture adapted from [283].

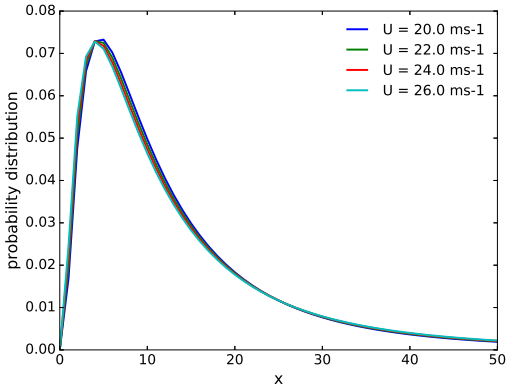


(b) $I \in [20000, 100000] KWm^{-1}$; Picture adapted from [283].

Figure 5.6: Lognormal distributions for various values of I while fixing σ.

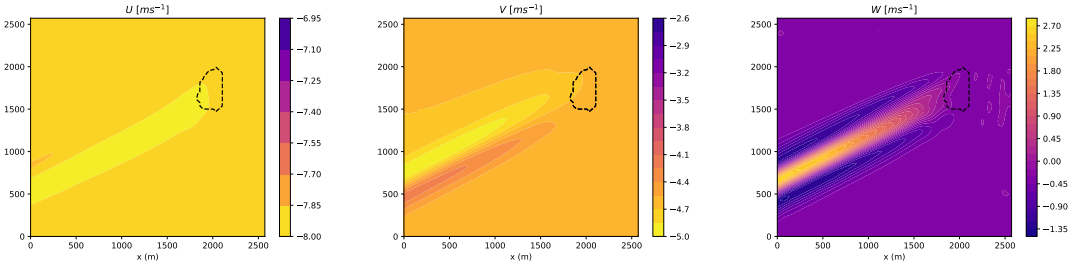


(a) $U \in [2, 16]\, ms^{-1}$; Picture adapted from [283].

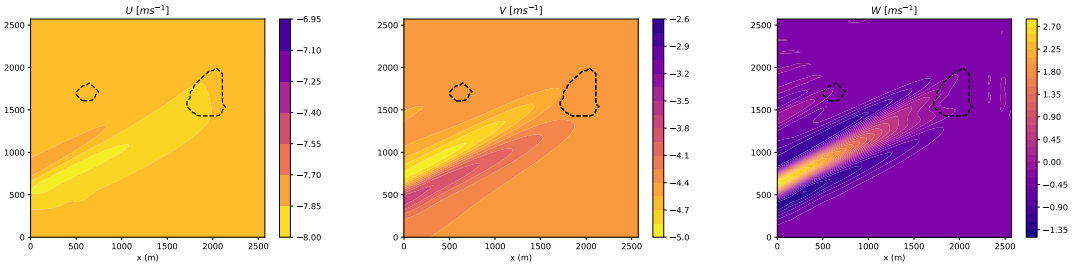


(b) $U \in [20, 26]\, ms^{-1}$; Picture adapted from [283].

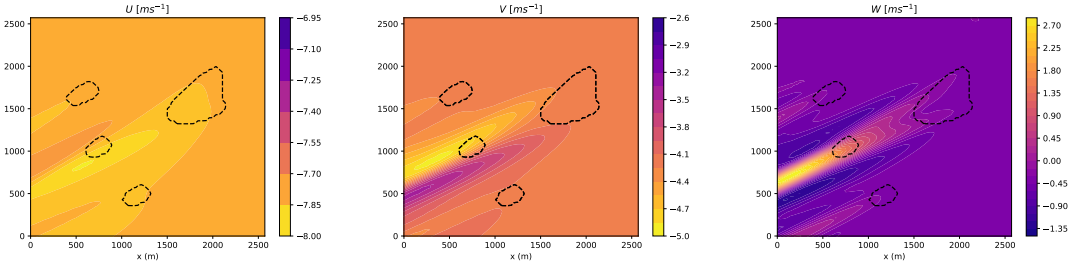
Figure 5.7: Lognormal distributions q for various values of U while fixing μ.



(a) $t = 6$ min.

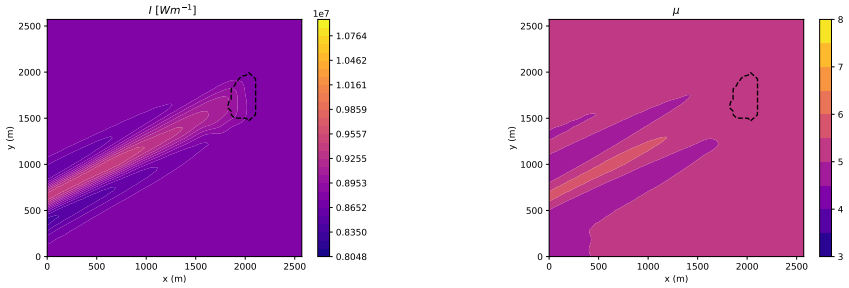


(b) $t = 10$ min.

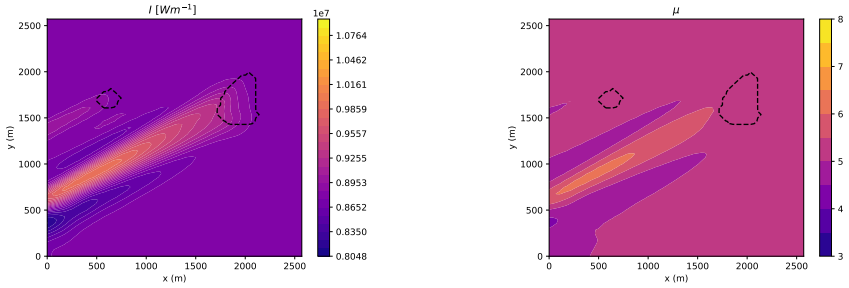


(c) $t = 20$ min.

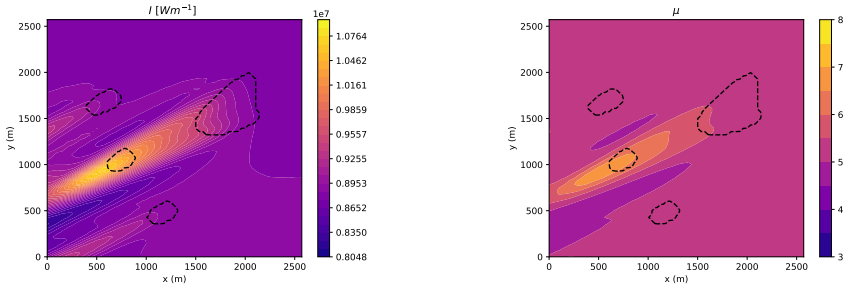
Figure 5.8: Wind vector components (U, V, W) performed with `WRF-Sfire` at times $t = 6, 10, 20$ min. Firefront is reported by a dashed line.



(a) $t = 6$ min.

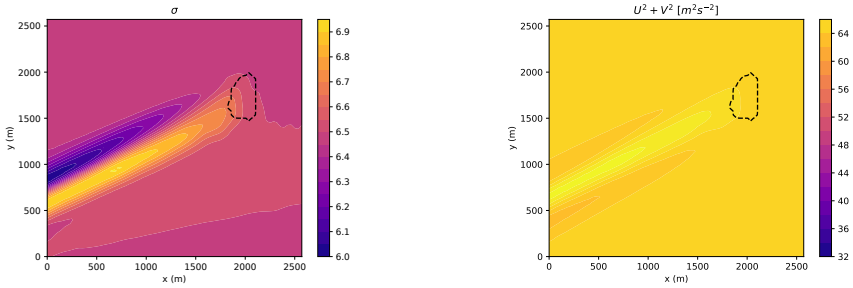


(b) $t = 10$ min.

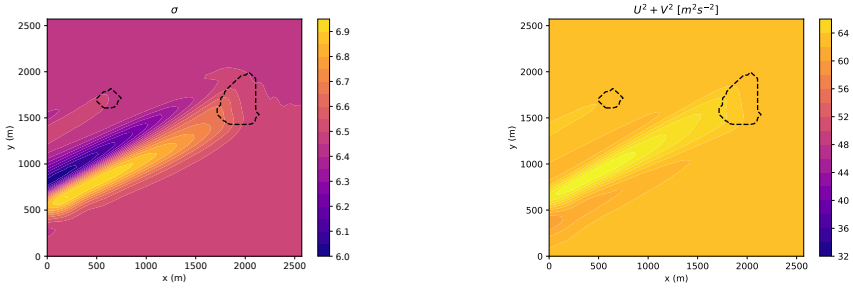


(c) $t = 20$ min.

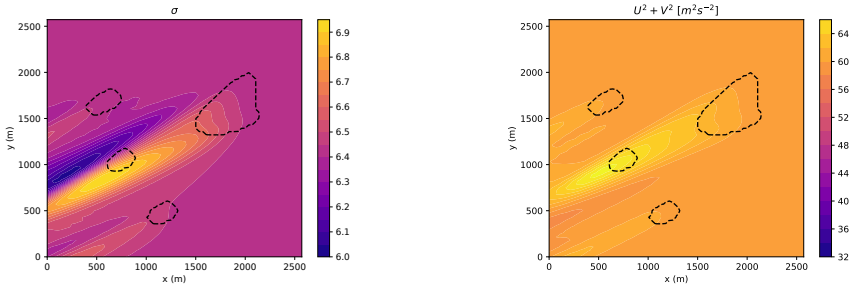
Figure 5.9: Fire intensity I and PDF shape parameter μ performed with `WRF-Sfire` at times $t =$ 6, 10, 20 min. Firefront is reported by a dashed line.



(a) $t = 6$ min.



(b) $t = 10$ min.



(c) $t = 20$ min.

Figure 5.10: PDF shape parameter σ and horizontal wind squared magnitude ($|\mathbf{U_h}|^2$) performed with `WRF-Sfire` at times $t = 6$, 10, 20 min. Firefront is reported by a dashed line.

5.8 UQ and SA of Wild-land fire Model

In order to spot the principal parameters for further model calibration (and with the perspective of integrating this model into Data Assimilation routines), and to quantify the uncertainties on the fire model output as well, a recent work of Sensitivity Analysis and Uncertainty Quantification has been carried out. The main framework is outlined in this Section, while the discussion of the main results is given in Section 5.9 Regarding the aim of UQ and SA, the adopted methodologies, e.g. Surrogates models, we refer to Chapter 2. The two following chapters are based on the recent publication [284].

Rate of Spread Submodel and Test Case Study

The focus of this Section is on sensitivity analysis methodology. Because of that, we consider a simplified version of the ROS parameterization required in Eq. (5.5). The maximum value of the ROS, $\mathcal{V}(\mathrm{x}, t)$, is specified by means of Byram's formula [43, 6]:

$$\mathcal{V}_0 = \frac{I}{\Delta h_c \, \omega_0}, \tag{5.24}$$

where I [kW m^{-1}] is the fireline intensity, Δh_c [kJ kg^{-1}] is the fuel heat of combustion and ω_0 [kg m^{-2}] is the oven-dry mass of fuel consumed per unit area in the active flaming zone. In analogy with the methodology adopted in [206], the action exerted by the near-surface wind field U on the ROS field is quantified using a corrective factor f_{w}, that reads:

$$\mathcal{V} = \mathcal{V}_0 \, \frac{(1 + f_{\mathrm{w}})}{\alpha_{\mathrm{w}}}. \tag{5.25}$$

In the equation above, f_{w} is computed adopting `fire-Lib` and `Fire Behaviour SDK` libraries (`http://fire.org`; see also [159], in the case of the NFFL – Northern Forest Fire Laboratory – Model 9). The parameter α_{w} is a suitable angle factor to guarantee that the maximum ROS in the upwind direction equals the ROS prescribed by Byram's formula (5.24). This choice makes the ROS dependent on the wind direction rather than on its magnitude to constrain the well-known dominant role of the wind in the fire propagation and to allow for the emergence, if they exist, of second-order effects due to other factors.

In this UQ and SA attempt, an idealized test case is considered. The computational domain is $7,200$ m $\times$ $6,000$ m. The wind field is uniform and constant. Vegetation is homogeneous and the terrain is flat. The fire front at time $t = 0$ is a circle of radius $r_c = 130$ m, whose center is located at $\mathrm{x}_c = (1,500 \text{ m}; 3,000 \text{ m})$.

5.8.1 Model Input Description

By adopting the same formalism of Chapter 2, the set of uncertain parameters is defined as $\boldsymbol{\theta} \in \mathbb{R}^d$, where d is the number of parameters to consider for the following sensitivity analysis. In this Section, two different sets of uncertain model parameters with $d = 3$ are studied. SA requires for each parameter a PDF representing its statistics and thereby its variability. This is useful for two main reasons. First, by knowing a fairly good prior of the input parameters, sampling for compiling the databases will be more meaningful, reproducing a more faithful statistical population. On the other hand, some statistical algorithm such as generalized Polynomial Chaos (see Chapter 2) can be specifically tuned according to the prior distribution of the studied parameters.

Sensitivity analysis for macroscopic/microscopic quantities

The first set of parameters mixes macroscopic and microscopic quantities: the wind speed magnitude $\|U\|$, the fireline intensity I and the ignition delay time-scale τ. Sensitivity analysis with $\boldsymbol{\theta} = (\|\mathrm{U}\|, I, \tau)^T$ corresponds to a preliminary step: in [284], uniform marginal distributions that spanned around the mean values adopted in previous work [206, 177, 126]. These uniform PDFs are outlined in Table 5.3.

Table 5.3: Uniform marginal PDFs for $\boldsymbol{\theta} = (\|\mathrm{U}\|, I, \tau)^T$. The uniform distribution is here reported as $\mathcal{U}(a; b)$ with a the minimum value and b the maximum value of the parameter.

Parameter	Uniform distribution
Wind $\|\mathrm{U}\|$ $[\mathrm{m\,s^{-1}}]$	$\mathcal{U}(6; 14)$
Fireline intensity I $[\mathrm{kW\,m^{-1}}]$	$\mathcal{U}(15,000; 25,000)$
Ignition delay reference time τ [s]	$\mathcal{U}(0.6; 1.4)$

Sensitivity analysis for microscopic parameters

The dependence of the wildfire spread model on a set of microscopic variables is also explored. This time, instead of using Uniform priors, we determine a suitable Bayesian description for the uncertain parameters $\boldsymbol{\theta} = (\mu, \sigma, D)^T$. Recall that this choice of $\boldsymbol{\theta}$ is linked exclusively to the fluctuating part of the forward model. As shown in Section 5.4, μ and σ are the two parameters of the log-normal PDF $q(\ell; t)$ (Eq. 5.11) that models the ember landing position. We remind that D (given by Eq. 5.22) is the diffusive coefficient of turbulent hot air, acting in the Gaussian PDF $G(\mathrm{x} - \overline{\mathrm{x}}; t)$, which describes turbulent diffusion. The fire-spotting parameterization introduced in [126], corresponding to `RandomFront v2.2`, described in Section 5.5 is adopted in this work. For each parameter, their dependence on subparameters is given by parametrization `RandomFront`. This allows for the determination of marginal PDFs by using a Monte Carlo random sampling on the subparameters. The resulting Beta-distributions are summarized in Table 5.4.

Statistical Description. The following methodology is carried out to obtain a statistical description of the three parameters $\{D, \sigma, \mu\}$, which depend on several subparameters.
Such subparameters are perturbed around their nominal values found in the literature following uniform PDFs. We recall that D is computed following Eq. (5.22). To obtain a range of variation for D, we modify the temperature difference in the convective cell ΔT and the dimension of the convective cell h. As for parameters σ and μ, they can be computed following Eqs. (5.20)–(5.17). We perturb the following parameters: α_H, β_H, γ_H , δ_H, H_{abl} in Eq. (5.21); ρ_{a}, ρ_{f} in Eq. (5.19); z_{p} and r in Eq. (5.17). All the identified subparameters are associated with a uniform PDF. For the parameters α_H, β_H, γ_H and δ_H, the extrema of the uniform PDF correspond to the highest and lowest values found in all the possible configurations outlined in [258], where both ABL and FT

Table 5.4: Range of variations and Beta-distribution for $\boldsymbol{\theta} = (\mu, \sigma, D)^T$. Note that the parameters of the Beta-distribution (Eq. 5.26) are given in the following order: shape parameters a and b, location and scale.

Parameter	Minimum/maximum values	Beta-distribution parameters
Log-normal parameter σ	5.49–12.69	1.37 1.99 5.94 4.93
Log-normal parameter μ	7.25–98.16	3.18 7.49 7.43 94.73
Turbulent diffusion coef. D [$\mathrm{m^2\,s^{-1}}$]	0.23–0.47	1.19 1.20 0.23 0.23

regimes are accounted for. ΔT lays within the range $[800; 923\ \mathrm{K}]$. For all other factors, we adopt a uniform PDF, whose extrema are defined adding a perturbation of 20 % to the values adopted in the work [126].

Once uniform PDFs are defined for each subparameter, they are sampled through a Monte Carlo random sampling. The size of the sample (or "ensemble") is 10,000 to obtain converged statistics. Starting with Eqs. (5.17 –5.22), we obtain 10,000 realizations of the three parameters of interest $\{D, \sigma, \mu\}$. We can then analyze their empirical statistical distribution by fitting the resulting histograms with different types of PDF. Figure 5.11 portrays the fits obtained when using a Beta-distribution for each sample. We adopt such distribution due to the subsequent requirements: positiveness, limited support, and compatibility with the adopted surrogate models (again, see Chapter 2). In Table 5.4 the characteristics of each Beta-distribution and the associated range of variation for each parameter in $\boldsymbol{\theta} = (\mu, \sigma, D)^T$ are presented. The analytic formulation for the Beta-distribution, with a and b ($a, b > 0$) the "shape parameters", reads:

$$\mathrm{Beta}(\mathrm{x}; a, b) = \frac{\Gamma(a+b)\,\mathrm{x}^{a-1}\,(1-\mathrm{x})^{b-1}}{\Gamma(a)\Gamma(b)}, \tag{5.26}$$

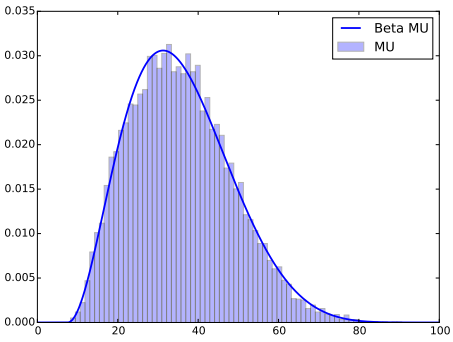
for $\mathrm{x} \in (0, 1)$, with $\Gamma(\mathrm{x})$ the Gamma function. To shift and/or scale the distribution, the "location" and "scale" parameters are introduced. $\mathrm{Beta}(\mathrm{x}, a, b, \mathrm{location}, \mathrm{scale})$ is equivalent to $\mathrm{Beta}(\mathrm{y}, a, b)/\mathrm{scale}$ with $\mathrm{y} = (\mathrm{x} - \mathrm{location})/\mathrm{scale}$.

5.8.2 Simulated Quantities of Interest

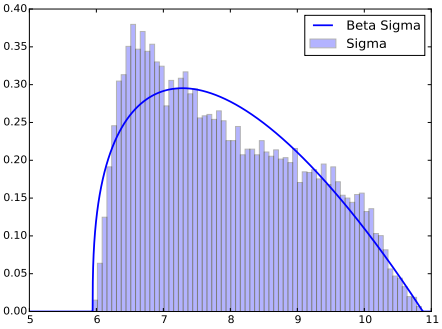
To quantify the evolution of a fire over a time period $[0; T]$, two different indices are defined. We consider first the percentage of the domain Ω that is burnt at a given time t:

$$A_t = \frac{\int_\Omega \mathcal{I}_{\mathcal{B}(t)}(\mathrm{x}, t)\, d\mathrm{x}}{|\Omega|}, \tag{5.27}$$

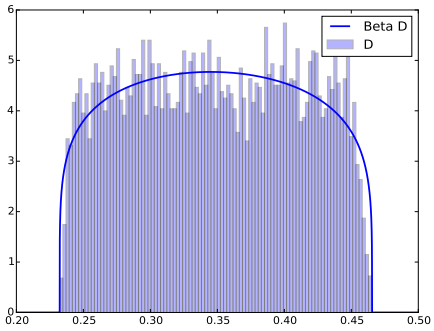
where $|\Omega|$ [$\mathrm{m^2}$] corresponds to the area of the computational domain and $\mathcal{I}_{\mathcal{B}(t)}$ is the indicator function of the burnt area, which returns 1 inside of the burnt area and 0 elsewhere. A_t corresponds to a normalized burnt area. However, this quantity does not give information on the topology of the fire, which can be complex in the case of fire-spotting. To overcome this limitation, we also consider an indicator S_t that describes the minimum spanning rectangle (from now on,



(a) Fire-spotting parameter μ.



(b) Fire-spotting parameter σ.



(c) Turbulent diffusion parameter D [$\mathrm{m}^2\,\mathrm{s}^{-1}$].

Figure 5.11: Histograms and corresponding fits with Beta-distribution (solid lines) for the three parameters μ, σ (fire-spotting effects) and D (turbulence effect) following a Monte Carlo random sampling with 10,000 realizations in the sample.

MSR) of the burnt area over the area of the domain $|\Omega|$ at a given time:

$$S_t = \frac{|\mathrm{MSR}(t)|}{|\Omega|}. \tag{5.28}$$

The MSR is a geometrical quantity that corresponds to the smallest rectangle within which all burnt grid points lie at a given time t. So $|\mathrm{MSR}(t)|$ [m^2] measures the area of this rectangle.

Fig. 5.12 portrays an ensemble of 100 firelines at time 50 min. Each plotted fireline corresponds to

a different set of parameters D, μ and σ (that is, to a different realization of $\boldsymbol{\theta} = (\mu, \sigma, D)^T$) obtained by sampling the Beta-distributions given in Table 5.4. For each fireline, Fig. 5.12 shows the corresponding normalized MSR as defined in Eq. (5.28) at time 50 min. Low MSR values (cyan colors) indicate simple topology of the fireline, while for high MSR values (rose colors) the fireline presents more irregularities and a more complex propagation induced by turbulence and fire-spotting, spanning thus a wider area.

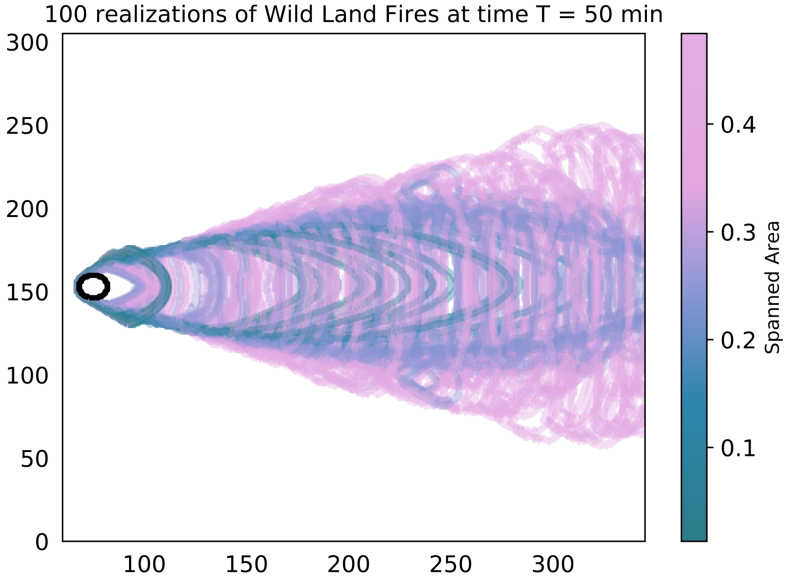


Figure 5.12: Ensemble of 100 fireline positions over the 2-D computational domain Ω after 50 min of `LSFire+` model integration obtained when varying D, μ and σ as presented in Table 5.4. The black circle is the initial fireline that is the same for all simulations. The colormap corresponds to the normalized MSR S_t at time $t = 50$ min (Eq. 5.28).

In this Chapter, in analogy with [284], we analyze the time dependency of the quantities A_t and S_t by comparing them at two different times, $t_1 = 26$ min and $t_2 = 34$ min. The resulting scalar quantities (or "observables", "Quantities of Interest", "QoI") are noted A_1, A_2, S_1 and S_2.

5.8.3 Design of Experiments

We build several datasets to analyze the performance of the gPC- and GP-surrogates in an extensive way in Section 5.10.2; these datasets are summarized in Table 5.5. Note that estimating the generalization error of the surrogate model requires the use of an independent dataset, that is why we use a Monte Carlo random sampling including $N = 216$ members for validation. Note also that the Halton's low-discrepancy sequence is involved in this work in order to explore the hypercube defined by the distribution of the uncertain parameters. This design of experiment will be compared to a tensor-based Gauss quadrature in terms of performance of the surrogate model.

Table 5.5: Datasets $\mathcal{D}_N$ of `LSfire+` simulations used in this work for building surrogates ("training") or for validating them ("validation").

Sampling Strategy	Purpose	Sample size
	$\boldsymbol{\theta} = (\|\mathrm{U}\|, I, D)^T$	
Halton's sequence	Training	216
Monte Carlo random sampling	Validation	216
	$\boldsymbol{\theta} = (\mu, \sigma, D)^T$	
Halton's sequence	Training	216
Quadrature rule	Training	216
Monte Carlo random sampling	Validation	216

5.9 Results of UQ and SA

The presented UQ and SA procedure has at least two objectives.

i) Firstly, an extensive comparison of the performance of different surrogate strategies for a limited budget (the size N of the training set $\mathcal{D}_N$, $N = 216$). See Table 5.6 forthe different types of surrogate adopted. The effect of the different surrogate model techniques on the predicted Quantities of Interest A_t and S_t in terms of mean value and STD are analyzed, as well as the effect on the predicted Sobol' sensitivity indices. This extensive analysis is carried out for the case $\boldsymbol{\theta} = (\mu, \sigma, D)^T$, meaning that we only consider uncertainty in the fluctuating part of the forward model `LSFire+`.

ii) The second objective is to rank by order of relevance the uncertain parameters, either $\boldsymbol{\theta} = (\|\mathrm{U}\|, I, \tau)^T$ or $\boldsymbol{\theta} = (\mu, \sigma, D)^T$, . On physical grounds, this allows to isolate the most influential input parameters for the problem of turbulence and fire-spotting.

5.9.1 Comparison of surrogate performance

Error assessment

The error indices here adopted follow directly from Section 2.4.5 of Chapter 2. In Table 5.7 the error metrics (i.e. the ϵ_{emp} empirical error and the Q_2 predictive coefficient) are presented. They are obtained for different types of surrogate (gPC on the one hand, and GP on the other hand) with respect to $\boldsymbol{\theta} = (\mu, \sigma, D)^T$ with a fixed size of the training set, $N = 216$. Concerning the gPC-surrogate, its performance is analyzed in details varying truncation and projection schemes, summarized in Table 5.6. The GP-surrogate is obtained using a standard RBF kernel (see Section 2.4.3). The GP surrogate is in this context a useful benchmark order to evaluate the quality of the gPC-surrogates. For every surrogate approach, one surrogate model is built for each of the four observables $\{A_1, A_2, S_1, S_2\}$ corresponding to the two observables A_t and S_t at times $t_1 = 26$ min and $t_2 = 34$ min,respectively.

In Table 5.7 we first focus on the results obtained with linear truncation ($q = 1$), meaning that the basis of polynomial functions is full for a given total polynomial order P. Table 5.8 (right column) presents corresponding scatter plots (referred to as "adequacy plots") of the surrogate predictions with respect to the physical model predictions. These plots quantify the adequacy of the surrogate to the physical model at the training points in terms of predicted burnt area ratio A_2. It is found that the Q_2 predictive coefficient is over 0.9 only for the LAR and cleaning sparse methods for all observables. The empirical error is of the same order of magnitude, varying between 10^{-3} for the MSR ratio S_t and 10^{-4} for the burnt area ratio A_t. Note that for a given observable at a given time, there is no significant difference among the surrogate strategies in terms of empirical error. We therefore focus the following analysis on the standalone Q_2 predictive coefficient. Note also that the performance of each surrogate is time independent since for a given observable, the Q_2 predictive coefficient is similar at times t_1 and t_2. We therefore focus on results at time t_2 in the following.

When moving to hyperbolic truncation schemes ($q = 0.75$ or $q = 0.5$), we reduce a priori the number of coefficients to compute in the gPC-expansion, while the size of the training set remains the same ($N = 216$). The lower the value of q, the smaller the number of gPC-coefficients r. Table 5.9 (right column) presents adequacy plots for hyperbolic truncation with $q = 0.5$; this is to compare to the adequacy plots obtained for linear truncation in Table 5.8 (right column). Results show that the performance of the quadrature approach does not improve when q decreases. In the opposite, the performance of the SLS approach improves and features a Q_2 predictive coefficient over 0.9 for A_2 and over 0.8 for S_2 when using hyperbolic truncation. This improvement is also noticeable in Table 5.9 (right column), where hyperbolic truncation allows to better represent the model response for low values of the burnt area ratio ($A_2 < 0.03$). The sequential sparse method also provides better results for a hyperbolic coefficient $q = 0.5$. The performance of LAR and cleaning sparse methods remains similar as in the linear case $q = 1$.

LAR appears as the most accurate gPC strategy and has a Q_2 predictive coefficient that is similar to that obtained with the GP-model based on RBF kernel. Hyperbolic truncation does not add much value to the results compared to linear truncation, except for the SLS strategy. This may be explained by the fact that the terms that are important to retain in the gPC-expansion are not located in an isotropic way in the three dimensions ($d = 3$). It is therefore of interest to identify which polynomial terms are important to keep in the basis in order to obtain a good performance of the surrogate in each of the three dimensions.

Sensitivity of gPC-surrogates to total polynomial order P

In Table 5.7, the results for SLS and LAR methods are obtained by choosing the optimal value of the total polynomial order P in the sense that the surrogate was obtained by finding the value of

Table 5.6: Types of surrogate used in this work. Recall that q is the hyperbolic parameter for truncation ($q = 1$ corresponds to linear truncation) and N is the size of the training set.

Name	Truncation	Sparse	Training set
Quad. (Quadrature)	$q = 1, 0.75, 0.5$	No	Gauss quadrature, $N = 216$
SLS (Standard Least-Squares)	$q = 1, 0.75, 0.5$	No	Halton, $N = 216$
LAR (Least-Angle Regression)	$q = 1, 0.75, 0.5$	Yes	Halton, $N = 216$
Cleaning	$q = 1, 0.75, 0.5$	Yes	Halton, $N = 216$
Sequential	$q = 1, 0.75, 0.5$	Yes	Halton, $N = 216$
RBF kernel	–	–	Halton, $N = 216$

P that maximizes the Q_2 predictive coefficient; P varying between 1 and 14. Recall that the total polynomial order P determines the size of the full basis used to construct the surrogate when using linear truncation. The SLS method considers the full basis, while the LAR method selects the most influential terms among the full basis. Since the size of the training set is fixed to $N = 216$ and since $(P + 1)^3 = 216$ for $P = 5$, we know that the problem becomes ill-posed for a full basis when the total polynomial order is over 5. This is not an issue for LAR since it selects inline the influential coefficients in the basis. It is therefore of interest to investigate if the LAR method features an improved performance when $P > 5$.

Figure 5.13 presents the Q_2 predictive coefficient for P varying between 1 and 14 for SLS and LAR surrogates obtained for the burnt area ratio A_2. As expected, Fig. 5.13a shows that the best performance of the SLS method with linear truncation is obtained for $P = 5$ and that it degrades very fast when increasing P (the Q_2 predictive coefficient is below 0.4 for $P > 6$). When moving to hyperbolic truncation with $q = 0.5$, Fig. 5.13c shows that the Q_2 predictive coefficient remains over 0.4 for $P > 5$. The resulting surrogate is therefore improved in this configuration as already pointed out in Table 5.7. Hyperbolic truncation allows the SLS approach to include high-order polynomials in the basis without generating an ill-posed problem (i.e. without having more coefficients to compute than the size N of the training set). Still, results show that the Q_2 predictive coefficient does not follow a monotonically increasing function toward the target value 1 in this hyperbolic configuration; this configuration is therefore not robust. In the opposite, the LAR method shows a monotonic convergence towards the target value 1 when increasing P in Figs. 5.13b–d. A good performance of LAR is obtained for $P = 10$ for both linear and hyperbolic truncation schemes. This sensitivity study shows that a total polynomial order P higher than 5 is required to build the response surface of the burnt area ratio. Similar results are obtained for the MSR ratio (not shown here). This demonstrates the benefits from sparse schemes when having a fixed and limited training set $\mathcal{D}_N$. Improving the performance of the SLS approach using linear truncation would require a higher total polynomial order P and therefore a larger training set.

Table 5.7: Error metrics ϵ_{emp} and Q_2 for gPC-expansions and GP-model detailed in Table 5.6. The size of the training set is $N = 216$. One type of surrogate is built for each of the four observables, A_1, A_2, S_1 and S_2.

	gPC expansion – Linear truncation ($q = 1$)							
	A_1		A_2		S_1		S_2	
	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2
Quad.	$1.4 \cdot 10^{-4}$	0.84	$2.7 \cdot 10^{-4}$	0.86	$5.5 \cdot 10^{-4}$	0.77	$4.6 \cdot 10^{-4}$	0.83
SLS	$3.0 \cdot 10^{-4}$	0.83	$6.3 \cdot 10^{-4}$	0.88	$1.0 \cdot 10^{-3}$	0.74	$2.3 \cdot 10^{-3}$	0.75
LAR	$1.0 \cdot 10^{-4}$	0.99	$4.2 \cdot 10^{-4}$	0.970	$5.0 \cdot 10^{-4}$	0.96	$2.3 \cdot 10^{-3}$	0.95
Cleaning	$1.0 \cdot 10^{-4}$	0.96	$4.1 \cdot 10^{-4}$	0.95	$5.5 \cdot 10^{-4}$	0.96	$1.2 \cdot 10^{-3}$	0.95
Sequential	$3.3 \cdot 10^{-4}$	0.85	$6.7 \cdot 10^{-4}$	0.89	$1.1 \cdot 10^{-3}$	0.77	$2.5 \cdot 10^{-3}$	0.85
	gPC expansion – Hyperbolic truncation ($q = 0.75$)							
	A_1		A_2		S_1		S_2	
	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2
Quad.	$3.7 \cdot 10^{-4}$	0.76	$8.6 \cdot 10^{-4}$	0.77	$1.6 \cdot 10^{-3}$	0.67	$3.7 \cdot 10^{-4}$	0.66
SLS	$1.5 \cdot 10^{-4}$	0.93	$1.8 \cdot 10^{-4}$	0.93	$1.0 \cdot 10^{-3}$	0.84	$2.5 \cdot 10^{-3}$	0.84
LAR	$2.0 \cdot 10^{-4}$	0.94	$5.6 \cdot 10^{-4}$	0.95	$1.0 \cdot 10^{-3}$	0.84	$2.6 \cdot 10^{-3}$	0.86
Cleaning	$9.9 \cdot 10^{-5}$	0.94	$3.3 \cdot 10^{-4}$	0.90	$5.0 \cdot 10^{-4}$	0.96	$1.1 \cdot 10^{-3}$	0.96
Sequential	$1.9 \cdot 10^{-4}$	0.94	$4.7 \cdot 10^{-4}$	0.94	$8.7 \cdot 10^{-4}$	0.86	$1.9 \cdot 10^{-3}$	0.92
	gPC expansion – Hyperbolic truncation ($q = 0.5$)							
	A_1		A_2		S_1		S_2	
	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2	ϵ_{emp}	Q_2
Quad.	$1.8 \cdot 10^{-4}$	0.83	$2.0 \cdot 10^{-4}$	0.87	$6.2 \cdot 10^{-4}$	0.74	$3.6 \cdot 10^{-4}$	0.83
SLS	$1.4 \cdot 10^{-4}$	0.96	$9.6 \cdot 10^{-5}$	0.95	$7.4 \cdot 10^{-4}$	0.86	$1.9 \cdot 10^{-3}$	0.86
LAR	$1.5 \cdot 10^{-4}$	0.97	$4.3 \cdot 10^{-4}$	0.97	$6.5 \cdot 10^{-4}$	0.93	$1.6 \cdot 10^{-3}$	0.94
Cleaning	$8.8 \cdot 10^{-5}$	0.95	$3.3 \cdot 10^{-4}$	0.94	$4.5 \cdot 10^{-4}$	0.92	$9.2 \cdot 10^{-4}$	0.98
Sequential	$1.3 \cdot 10^{-4}$	0.97	$4.2 \cdot 10^{-4}$	0.96	$6.4 \cdot 10^{-4}$	0.93	$1.5 \cdot 10^{-3}$	0.95
	GP model							
RBF	$--$	0.99	$--$	0.98	$--$	0.88	$--$	0.99

Identification of the influential gPC-coefficients

Table 5.8 (left column) presents a three-dimensional schematic (referred to as "sparsity plot") of the coefficients retained in the gPC-expansion using linear truncation, each dimension corresponding to one stochastic/uncertain dimension. The three dimensions are here the turbulent diffusion coefficient D and the lognormal parameters μ and σ. This is useful to visualize the polynomial degree associated with the active coefficients as well as the magnitude of the coefficients given by the colormap (recall that there is a direct link between the coefficients and the statistical moments of the predicted quantity of interest for gPC-expansion).

Quadrature and SLS methods have the same full basis for a given polynomial order P (here $P = 5$ since the size of the training set is $N = 216$); they are associated with a typical "pyramidal" sparsity plot, where the first coefficient corresponding to the mean estimate of the burnt area ratio A_2 has the highest magnitude (approximately equal to 0.04). For sparse methods (LAR, cleaning, sequential), the number of coefficients is significantly reduced since the terms with the least impact are automatically filtered out of the sparse basis. The sparsity plot has no longer a "pyramidal" shape. LAR and sequential strategies feature instead a two-dimensional structure (along the

Table 5.8: Comparison between quadrature, SLS and sparse (LAR, cleaning, sequential) methods to build the gPC-expansion for the burnt area ratio A_2 using linear truncation. Left: sparsity plots representing the magnitude of the coefficients with respect to the three-dimensional input space ($d = 3$). Right: adequacy scatter plots comparing surrogate (x-axis) and model (y-axis) predictions at the training points. For SLS and LAR, results are obtained with the best fit obtained for varying P.

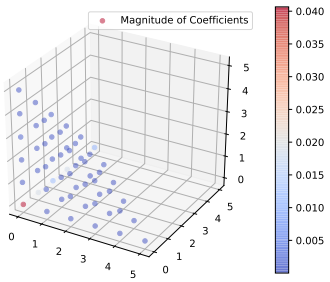
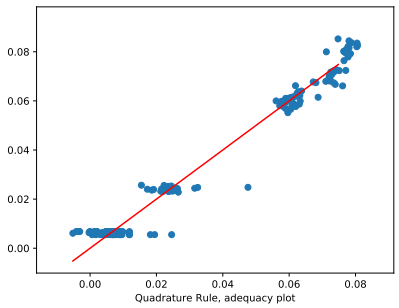
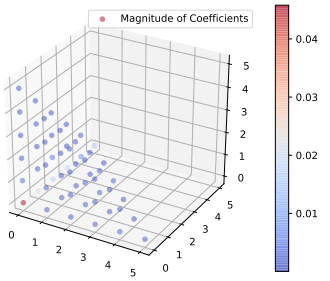
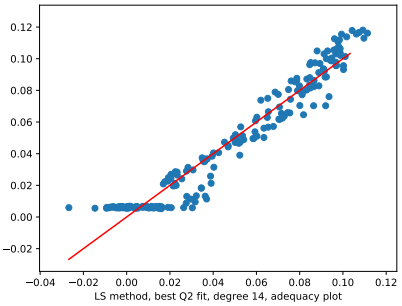
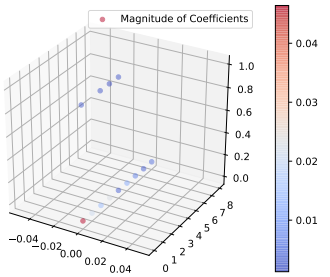
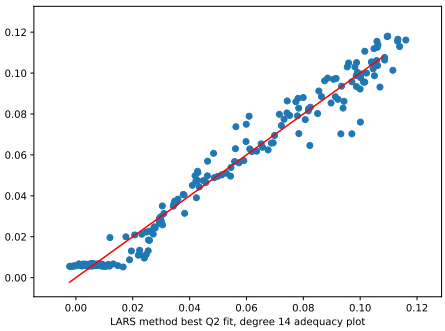
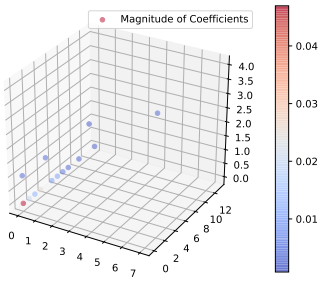
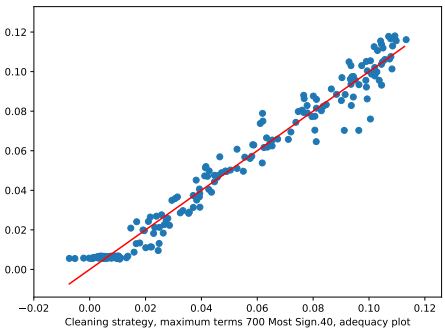
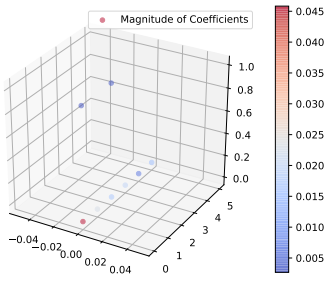
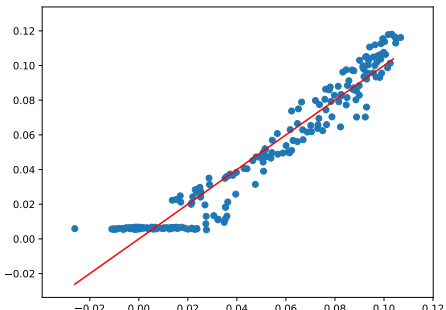
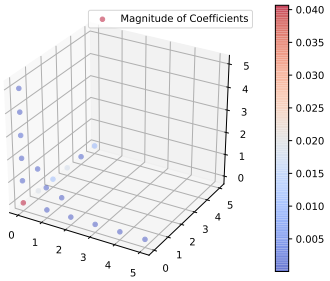
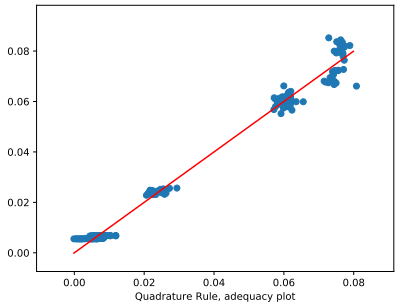
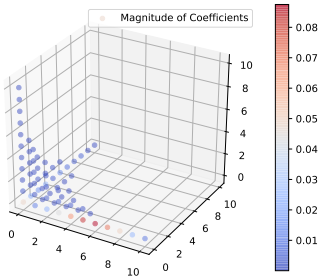
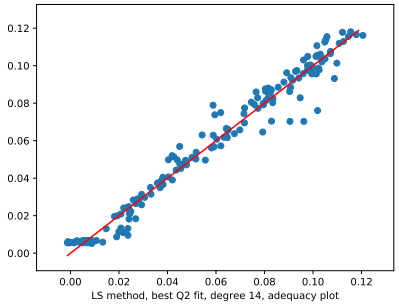
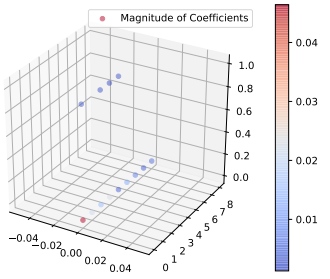
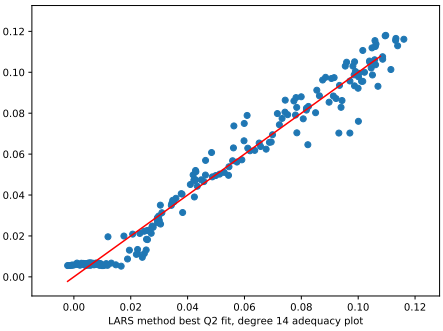
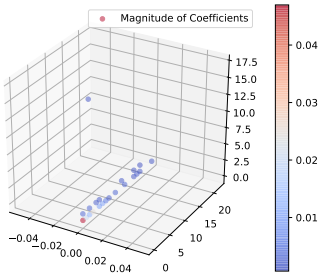
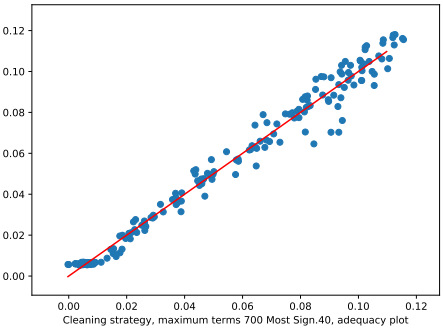
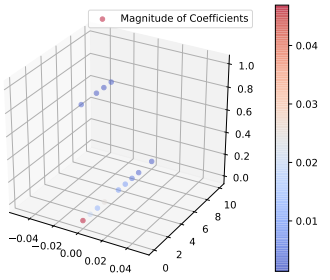
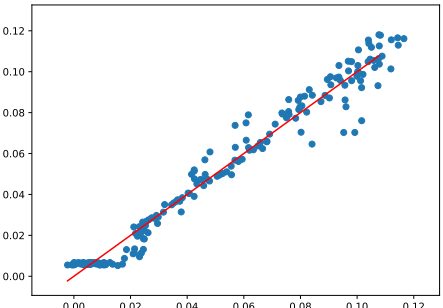
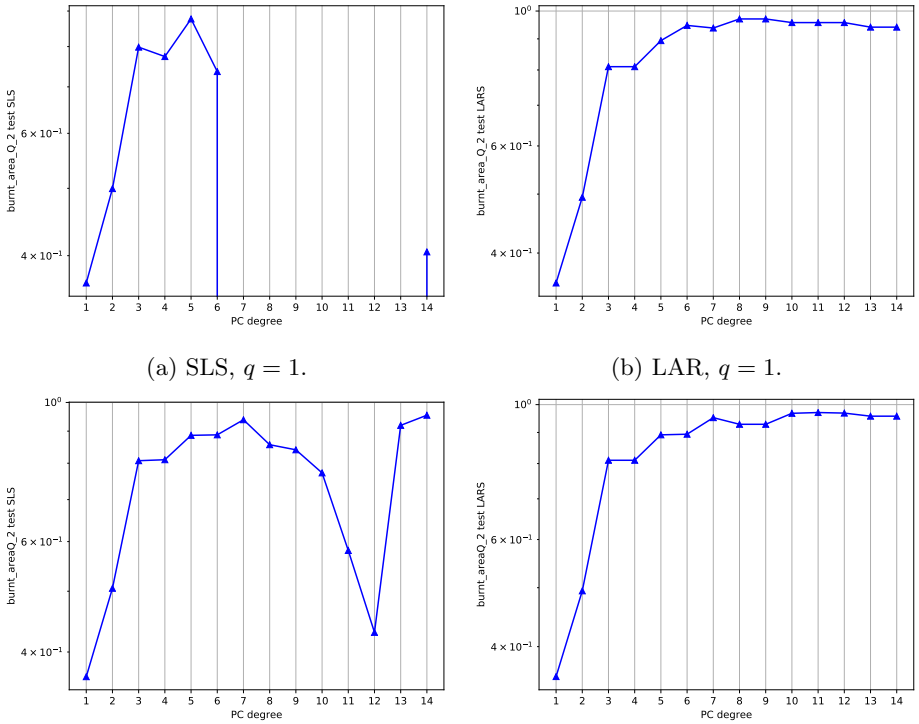
Quad.		
SLS		
LAR		
Cleaning		
		

Table 5.9: Same caption as Fig. 5.8 but for hyperbolic truncation with $q = 0.5$.

Quad.		
SLS		
LAR		
Cleaning		
Sequential		



(a) SLS, $q = 1$. (b) LAR, $q = 1$.

(c) SLS, $q = 0.5$. (d) LAR, $q = 0.5$.

Figure 5.13: Sensitivity of the Q_2 predictive coefficient with respect to the total polynomial order P. Comparison of the SLS (a)–(c) and LAR (b)–(d) surrogate methods for linear truncation (top panels) and hyperbolic truncation with $q = 0.5$ (bottom panels) for $1 \leq P \leq 14$.

vertical plane) indicating that the burnt area ratio A_2 is not sensitive to the third dimension, here the lognormal parameter μ, but only to the lognormal parameter σ and to the turbulent diffusion coefficient D. Only the cleaning strategy retains a three-dimensional structure by accounting for interaction terms involving the lognormal parameter μ. This highlights the presence of influential interaction terms involving several parameters. However, all sparse strategies indicate that one direction is dominant since the number of coefficients in this direction is high and the basis terms can go up to a total polynomial order $P = 12$ in the case of cleaning and $P = 8$ in the case of LAR (instead of the constrained $P = 5$ for quadrature and SLS). This dominant direction corresponds to the lognormal parameter σ.

Note that Table 5.9 (left column) presents similar plots as Table 5.8 (left column) but for hyperbolic truncation with $q = 0.5$. The magnitude of the coefficients does not change for quadrature, explaining why hyperbolicity does not improve the performance of the surrogate based on quadrature. This is not the case of SLS, which now features high magnitude for the coefficients along the direction D for polynomial terms having a degree between 4 and 8. This highlights the need to have polynomials of higher degree to capture underlying physical processes. Still, SLS with hyperbolicity is not sufficient to capture the same structure as sparse methods. Note that sparse

methods converge to the same structure using linear or hyperbolic truncation schemes, indicating the robustness of these methods.

The influence of the three parameters on the behavior of the burnt area ratio A_2 can be quantified using Sobol' sensitivity indices. Table 5.10 presents the Sobol' indices using sparse methods and linear truncation for the burnt area ratio A_2 (same results are obtained using hyperbolic truncation with $q = 0.5$ – not shown here). Table 5.11 presents similar quantities for the MSR ratio S_2. Results confirm that the lognormal parameter σ is the most influential one for both quantities of interest A_2 and S_2 with a first-order sensitivity index above 0.98 for A_2 and above 0.92 for S_2. This means that more than 90 % of the variance in A_2 and S_2 is explained by uncertainties in the lognormal parameter σ. Results also show interaction effects are limited but still present between the lognormal parameter σ and the turbulent diffusion parameter D as foreseen in sparsity plots. Note that all sparse gPC-surrogates as well as the GP-model exhibit the same global trend. The main differences lie in the relevance of the lognormal parameter μ. LAR and sequential strategies cut out any contribution of μ in the variability of the predicted quantities of interest. This is not the case of the cleaning strategy that has a non-zero total Sobol' index for μ as the GP-model.

We can evaluate the impact of the choice in the surrogate strategy on the predicted mean and STD estimates of the quantities of interest. Note that the coefficients of the gPC-expansion can be interpreted in a statistical way with the first coefficient being the mean estimate and the squared sum of the other coefficients being its corresponding variance estimate (see Section 2.3). Table 5.12 presents the mean and STD estimate of the burnt area ratio A_2 and of the MSR ratio S_2 obtained for different gPC- and GP-surrogates. Results show the consistency of the statistical moments obtained using sparse gPC-expansions and GP-model for both A_2 and S_2. The SLS approach using linear truncation is able to retrieve accurate mean and STD estimates (about 1 % deviation with respect to GP-model predictions). In the opposite, the quadrature approach provides mean and STD estimates with more than 10 % deviation with respect to GP-model predictions.

This highlights the importance of having high-order polynomial terms in some uncertain directions to build an accurate gPC-expansion and have accurate estimate of the statistical moments in the present study. These directions can be identified using Sobol' sensitivity indices. Sparse gPC-strategies are relevant to address such issues due to the flexibility of selecting the most influential polynomial terms during the construction of the surrogate (linear and hyperbolic schemes are defined a priori).

Sensitivity to the size of the training set

So far the analysis was obtained for a fixed training set of size $N = 216$ (generated using Halton's low discrepancy sequence or, in case of quadrature rule, tensor-based Gauss quadrature sampling). A convergence analysis is carried out, to discover whether the same level of accuracy could be

Table 5.10: Comparison of Sobol' sensitivity indices associated with the burnt area ratio A_2 and obtained for Halton's low discrepancy sequence.

	S_μ	S_σ	S_D	$S_{T,\mu}$	$S_{T,\sigma}$	$S_{T,D}$
	gPC expansion – Linear truncation $q = 1$					
LAR	0.	0.986	$5.67 \cdot 10^{-3}$	0.	0.994	$1.35 \cdot 10^{-2}$
Cleaning	0.	0.984	$5.89 \cdot 10^{-3}$	$4.70 \cdot 10^{-3}$	0.994	$1.62 \cdot 10^{-2}$
Sequential	0.	0.987	$4.84 \cdot 10^{-3}$	0.	0.995	$1.33 \cdot 10^{-2}$
	GP model					
RBF kernel	$4.59 \cdot 10^{-4}$	0.982	$5.97 \cdot 10^{-3}$	0.001	0.992	0.012

Table 5.11: Same caption as Table 5.10 but for the MSR ratio S_2.

	S_μ	S_σ	S_D	$S_{T,\mu}$	$S_{T,\sigma}$	$S_{T,D}$
	gPC expansion – Linear truncation $q = 1$					
LAR	0.	0.948	$1.49 \cdot 10^{-2}$	0.	0.985	$5.22 \cdot 10^{-2}$
Cleaning	0.	0.925	$1.66 \cdot 10^{-2}$	$2.66 \cdot 10^{-3}$	0.983	$7.18 \cdot 10^{-2}$
Sequential	0.	0.954	$1.45 \cdot 10^{-2}$	$7.15 \cdot 10^{-3}$	0.978	$4.63 \cdot 10^{-2}$
	GP model					
RBF kernel	$5.43 \cdot 10^{-4}$	0.941	$9.89 \cdot 10^{-3}$	0.002	0.975	0.047

obtained or not, for sparse gPC-surrogates built with a reduced training set ($N < 216$). To answer this question, we vary the size of the training set N between 10 and 216 with respect to the observable S_2. For each size of the training set, a LAR gPC-surrogate is built and the Q_2 predictive coefficient provided its cross-validation using the available Monte Carlo validation database (Table 6.3.9). This convergence test is performed for different truncation strategies, i.e. for different levels of hyperbolicity $q \in \{1, 0.75, 0.5\}$. Figure 5.14 presents the evolution of Q_2 with respect to the size of the training set N. Results show the convergence of Q_2 to a constant value when $N > 100$. Linear truncation and hyperbolic truncation ($q = 0.5$) provided a similar performance for $N > 100$. It can be noted that the hyperbolic solution obtained using $q = 0.75$ is not the best option.

5.9.2 Analysis of the physical model predictions

Results show that the LAR gPC-strategy features a good overall performance.
For this reason, the following analysis is restricted to the latter Surrogate Model. It will be used to further analyze the fire-spotting and turbulence submodel included in `LSFire+`, retrieving Sobol Indices for each input parameter.
In Table 5.14 and Table 5.16 are summarized the error metrics as well as the mean and STD estimate of the burnt area ratio A_2 and of the MSR ratio S_2 at time t_2 for the two sets of uncertain parameters $\boldsymbol{\theta} = (\|\mathrm{U}\|, I, \tau)^T$ and $\boldsymbol{\theta} = (\mu, \sigma, D)^T$, respectively. In Table 5.13 and Table 5.15 the corresponding Sobol' Indices can be found. Note that the empirical error ϵ_{emp} and the Q_2 predictive coefficient are in acceptable range for all tested configurations. Anyway, in this Subsection the focus lies upon the physics of the problem.

Table 5.12: Mean and STD estimate of the burnt area ratio A_2 (left column) and of the MSR ratio S_2 (right column) using linear truncation scheme ($q = 1$), Halton's low discrepancy sequence and gPC or GP surrogate approach.

	A_2	S_2
	gPC expansion – Linear truncation ($q = 1$)	
	mean ± STD	mean ± STD
Quad.	0.0406 ± 0.175	0.102 ± 0.322
SLS	0.0458 ± 0.198	0.114 ± 0.333
LAR	0.0464 ± 0.194	0.114 ± 0.324
Cleaning	0.0469 ± 0.194	0.115 ± 0.327
Sequential	0.0458 ± 0.196	0.113 ± 0.319
	GP model	
	mean ± STD	mean ± STD
RBF kernel	0.0463 ± 0.194	0.114 ± 0.327

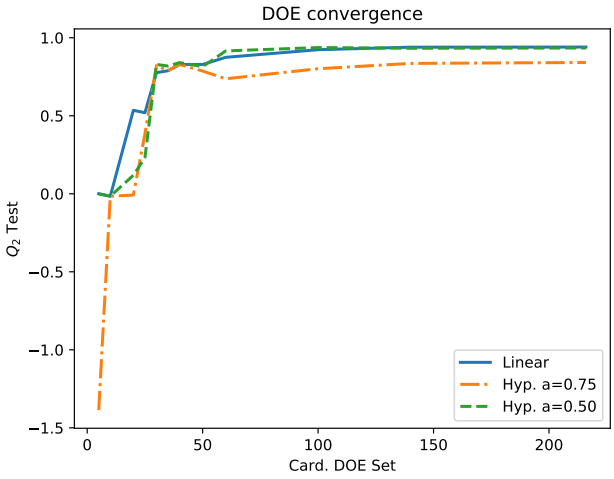


Figure 5.14: Convergence test with respect to Q_2 predictive coefficient for the LAR gPC-surrogate built using Halton's low discrepancy sequence (cross-validated using the Monte Carlo random sampling). Solid blue line corresponds to linear truncation; dash-dotted orange line corresponds to hyperbolic truncation with $q = 0.75$; and dashed green line corresponds to hyperbolic truncation with $q = 0.5$.

As shown in Section 2.3.2 of Chapter 2, Sobol' sensitivity indices order by relevance each input factor. In the case $\boldsymbol{\theta} = (\|\mathrm{U}\|, I, \tau)^T$, a clear predominance of the wind speed $\|\mathrm{U}\|$ is reported, given the considered range of the fireline intensity I. This is a rather interesting result, because some normalization has been applied on the wind effects on the "deterministic" ROS model (i.e. parameter α_{w} in Eq. 5.24). The latter normalization made the propagation of the deterministic fireline depend solely on the orientation of the wind vector and not on its magnitude. This means that the wind has a more general and fundamental role, reflected e.g. in fire-spotting and secondary fire generation (the so called "fluctuating part" of the model described in Section 5.4. The ballistic term σ in Eq. (5.11) gjven by the parametrization `RandomFront` depends heavily on the value of $\|\mathrm{U}\|$. This is in line with the results of the second set of input factors. In the case

$\boldsymbol{\theta} = (\mu, \sigma, D)^T$, σ is the most influential parameter when considering Sobol' indices, way above D and μ (in order of relevance). The Sobol Indices for both A_t and S_t are comparable. Nevertheless, S_t gives slightly more relevance to μ and D inputs with respect to A_t. For both parameter sets, the mean of the S_2-observable is larger than that of A_2, as we could expect from their respective definition. Its value of STD is also larger. Uncertainties in $\{\|\mathrm{U}\|, I, \tau\}$ cause a more significant spread of the fire-line shape and position, when compared to uncertainties in $\{\mu, \sigma, D\}$. This can be explained by the fact that in the first case the ember ignition time scale is also varied.
To sum up, the results of this UQ and SA highlighted the importance of the mean wind factor, on both main fire propagation and the generation of secondary fires. This result is consistent with the phenomenology of wild-land fires and with the fire-spotting phenomenon. We recall that fire-spotting refers to independent ignitions located far away from the main fireline. This occurs when the convective column lofts firebrands, the wind transports them up to their falling into the downwind fuel and the firebrands may ignite. The stronger the wind, the larger distance firebrands can be transported. This process is accounted in the model via the lognormal parameter σ. The importance of σ is also, put into mathematical terms, a feature of the adopted lognormal PDF for firebrand landing distance. It controls in facts the tail of the density function, the kurtosis of the lognormal density being equal to $\mathrm{e}^{4\sigma^2} + 2\mathrm{e}^{3\sigma^2} + 3\mathrm{e}^{2\sigma^2} - 3$. We remark that these results, reported also in [284], are not neglecting the "response saturation" due to very strong winds and the eventual burn-out of the airborne firebrands seen in Section 5.7.1 (see Figure 5.5d). It should be pointed out that in the setting of this SA and UQ procedure, the wind varied according to Table 5.3, in a range way smaller than the one explored by the simpler response analysis summarized in Figure 5.5d. Similar arguments can be stated for the selected range of I, that in [284] is a fraction of the whole range analyzed in [283], which is portrayed in Figure 5.5b. Nevertheless, this UQ and SA study thus showed that the studied turbulence and fire-spotting sub-model correctly includes the double role of the mean wind,on the one hand enhancing the propagation of the main fire-line , and on the other hand transporting firebrands for secondary ignitions.

5.10 Multi-scale modelling of fire spotting

This Section, which concludes the wild-land fire-themed Chapter, concerns the two last works in chronological sense related to the Wild-land fire model of Section 5.4. These works used the presented Model to study the role for the emergence of fire-spotting phenomena and the ignition of secondary fires of both a macroscale factor, as the atmospheric stability ([77]) and of a mescoscale factor, as the flame length ([78]).
Such works focused on the interactions among scales of the whole wild-land fire phenomenon, following a *concurrent multiscale modelling*: that is, estimating parameters related to aspects

Table 5.13: Sobol' indices (first-order in dark blue and total-order in light blue) using LAR gPC-surrogate and linear truncation; $\boldsymbol{\theta} = (U, I, \tau)^T$; $N = 216$. Left: Sobol' indices associated with the burnt area ratio A_2. Right: Sobol' indices associated with the MSR ratio S_2.

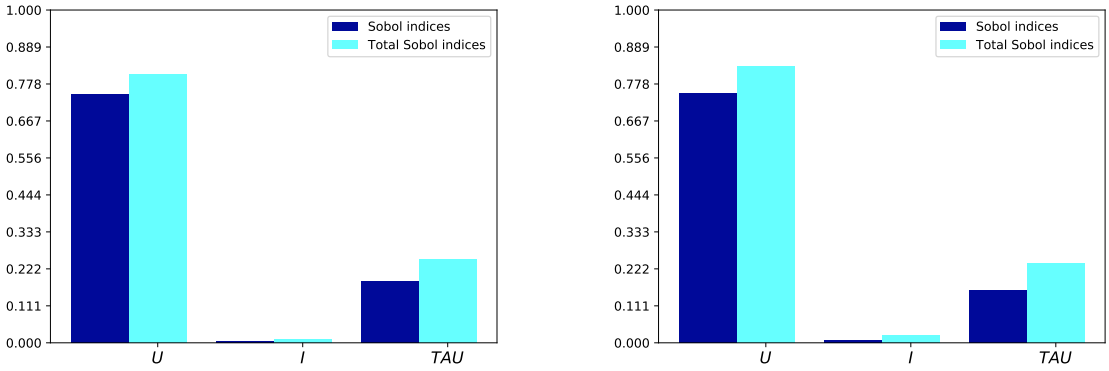


Table 5.14: Mean and STD of observables A_2 and S_2 as well as error metrics ϵ_{emp} and Q_2 using LAR gPC-surrogate and linear truncation; $\boldsymbol{\theta} = (U, I, \tau)^T$; $N = 216$.

Quantity of interest	Mean	STD	ϵ_{emp}	Q_2
A_2	0.07	0.06	$9 \cdot 10^{-4}$	0.95
S_2	0.19	0.13	$2 \cdot 10^{-3}$	0.96

Table 5.15: Same caption as in Table 5.13 but for $\boldsymbol{\theta} = (\mu, \sigma, D)^T$.

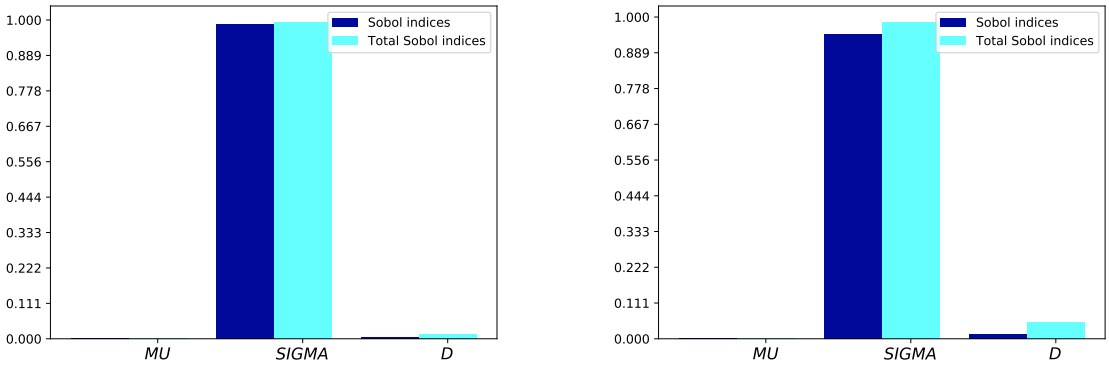


Table 5.16: Same caption as in Table 5.14 but for $\boldsymbol{\theta} = (\mu, \sigma, D)^T$.

Quantity of interest	Mean	STD	ϵ_{emp}	Q_2
A_2	0.05	0.04	$4 \cdot 10^{-4}$	0.97
S_2	0.11	0.11	$2 \cdot 10^{-3}$	0.95

occurring in a very large range of scales and implementing them into the model for the macroscopic fire front. This concept is delineated in Figure 5.15.

- At the macroscopic scale, fire-spotting phenomenon is affected by atmospheric conditions. In

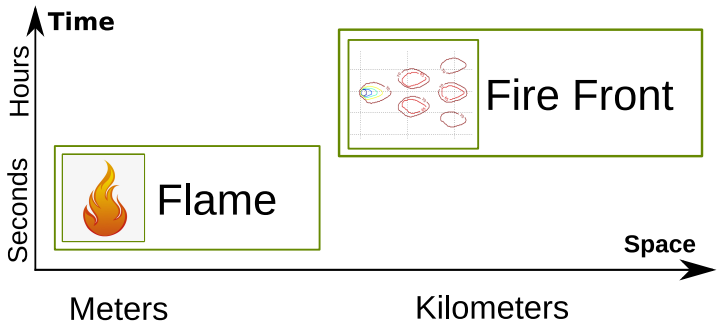


Figure 5.15: Sketch of the multiple scales occurring in a wild-land fire. Adapted from [77]

`RandomFront` parametrization, the depth of the atmospheric boundary layer (ABL) is used as macroscopic factor. Such quantity is related to the atmospheric stability, into the estimation of the smoke-injection height including the uplift against the atmospheric stratification and the plume widening due to entrainment of the surrounding air [258]. In `RandomFront` the firebrand-injection height is estimated as an approximated lift for inertial particles that are flowing into the fire plume. In [282] a simple formulation of the stable boundary layer is derived by making use of the bulk Richardson number. The refinement of this formulation is proposed by Serafin in [247]. More general study, that includes formulations for stable, neutral and unstable boundary layers may be found in [266].

- At the mesoscopic scale, fire-spotting is affected strongly by the fire intensity and by the flame characteristics, fire intensity being related to many aspects of the flame geometry, see [6] and references therein. As shown in [296], fire spread is strongly affected by the geometrical characteristics of the flame . An accurate estimation of these parameters can help to determine how a wildfire may be controlled. In practice, the flame length is used to determine the size of fire control lines [1], while the flame height is used to predict the exposure to heat flux [236]. Also fire-spotting is affected by fire intensity. Such interaction is used by practitioners to forecast house survival probability in bush fires, see e.g. [308].

To sum up, the objective of this last research effort is then to highlight the role of a macroscopic factor and the one of that of a mesoscopic factor for characterizing fire-spotting. With such characterization, information on favorable configurations for the occurrence of fire-spotting and the associated risk can be provided. Implicit connection between the atmospheric stability and fire propagation allows the modeling of several scenarios for different times of day of wildfire. Another important result of this research is the establishment of the range of unburned terrain where the secondary fires may occur, which models different scenarios of firebrand showers.

The papers [77] and [78] are the last two works, chronologically speaking, which make use of the Model of Section 5.4 and of parametrization `RandomFront 2.3`. In these works some slight modifications are brought to the parametrization, giving rise to the version `RandomFront 2.3b`. The simulations needed from the two above cited papers are carried out in `LSFire+`. The results presented in this thesis refer exclusively to the macroscopic scale effects, with the model upgrading presented in the next subsection. Such results have been published in the recent paper [77].

5.10.1 Upgrading the Model: `RandomFront2.3b`

In this subsection some slight modifications to the model adopted in the last works, that converged into version 2.3*b* of `RandomFront` are presented. To begin with, it is remarked that the injection height of firebrand is a fraction of the injection height of the smoke, H_{smoke}, i.e., $H = \lambda H_{\text{smoke}} < H_{\text{smoke}}$, as already seen in Equation 5.20 of Section 5.5. In the following the choice $\lambda = 0.4$ is adopted. This estimation is explained in detail in the Appendix of [77].
It is recalled that the injection height can be described in terms of buoyancy frequency, or Brünt-Väsäla frequency N^2 [258]. This quantity is used in meteorology to measure atmospheric stratification and atmospheric stability. Positive values of N^2 refer to a stable boundary layer (SBL). If the potential temperature is uniform with height, a displaced air parcel experiences no buoyancy force and will thus remain stable at its new location. Such a layer of air is described as *neutrally stable*. If the potential temperature increases with height, a parcel of air which is displaced upwards (downwards) experiences this time a negative (positive) restoring force and shall tend to return to its equilibrium position. It can usually be observed at nighttime. For this reason, it is also known as *nocturnal boundary layer*. On the contrary, atmosphere is considered to be unstable, if the potential temperature decreases with height, and a displaced parcel would experience a force in the direction of the displacement. Unstable atmosphere is described by negative values of N^2 that may cause the appearance of complex numbers in the formulation of the injection height. In order to avoid such problems, general formulation has been proposed in [258] by using the free troposphere value N^2_{FT} [258] and letting a part of the ABL ($\alpha < 1$) be considered as "freely" passed. In `RandomFront` it is used the following generic formula for the injection height of the smoke of Equation 5.21, here reported for clarity:

$$H_{\text{smoke}} = \alpha H_{\text{ABL}} + \beta \left(\frac{I_{\text{f}}}{d\, P_{f_0}} \right)^{\zeta} \exp\left(-\frac{\delta_{\text{FT}} N^2_{\text{FT}}}{N_0^2} \right) , \tag{5.29}$$

where I_{f} is the fire intensity, H_{ABL} is the height of ABL. Parameters are constrained by: $\alpha < 1$, $\beta > 0$, $\zeta < 0.5$ and $\delta_{\text{FT}} \geq 0$ [258]. The stability conditions of the atmosphere enter into such parametrization with the height of the ABL. Note that H_{ABL} has an impact not only in the fire-spotting parametrization, but also for turbulent diffusion, due to Parametrization of D, in Section 5.5.

On the basis of the maximum loftable height (5.21), the maximum travel distance can be defined as follows [302]

$$\ell_{\max} = H \left(\frac{3\rho C_{\mathrm{d}} Fr}{2\rho_{\mathrm{f}}} \right)^{1/2} , \tag{5.30}$$

Such equation is of course compatible with Equation 5.16. Following the definition of μ and σ, we report for clarity from Section 5.5 the expression for $\ell_{\max}$ in terms of σ, μ:

$$\ell_{\max} = \mu \exp(z_{\mathrm{p}} \sigma) . \tag{5.31}$$

Formulation (5.21), as well as (5.30), may take into account stable, neutral and unstable atmosphere. The model is therefore versatile enough to describe wildfire behavior at any daytime. The wind velocity in (5.17) is the projection of the vector of wind to the vector Φ from some point of the computational domain to the point where the PDF is computed, see Figure 5.16. Denoting by ω the angle between the wind direction and the vector Φ, the expression of σ in 5.17 reads

$$\sigma_\omega = \frac{1}{2 z_{\mathrm{p}}} \ln \left[\frac{(U \cos\omega)^2}{rg} \right] . \tag{5.32}$$

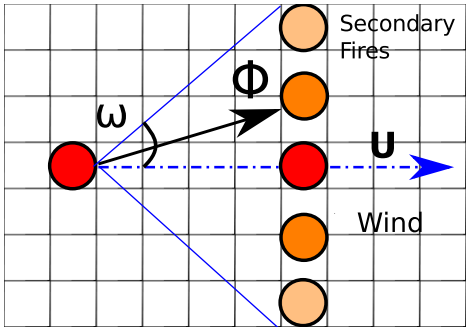


Figure 5.16: Sketch of the vector Φ and the angle ω in the generation of secondary fires. Adapted from [77].

This upgrade of the model allows for σ_ω to depend not only on the wind velocity, but also on the position of the considered point. For physical reasons, it is assumed that fire-spotting is a downwind phenomenon, that translates into $-\pi/2 \leq \omega \leq \pi/2$. If angle ω tends to $\pi/2$, σ becomes negative: that would eventually lead to small value of the maximum landing distance. We observe that in the limit case, firebrands would eventually fall down to already burned terrain and there is thus no fire-spotting effect. Therefore, negative values of σ_ω are not considered. On the other side, when ω is close to zero, the position vector Φ has approximately the same direction as the mean wind vector (see Figure 5.16) and the travel distance reaches its maximum. We define a limit angle ω_0 in order to reduce computational cost in the simulations, such that the fire-spotting distribution is evaluated only when $\cos\omega > \cos\omega_0$. Such angle can be determined with a few computations, as shown in [77].

The minimum travel distance expressed in in terms of the parameters μ and σ and the limit angle ω_0 reads:

$$\ell_0 = \mu \exp(z_{\mathrm{p}}^0 \sigma) = H \left(\frac{3\rho C_{\mathrm{d}}}{2\rho_{\mathrm{f}}} \right)^{1/2} \frac{U \cos\omega_0}{\sqrt{rg}}\,, \tag{5.33}$$

after some computations we get to

$$\cos\omega_0 = \left(\frac{U}{\sqrt{rg}} \right)^{\frac{z_{\mathrm{p}}^0}{z_{\mathrm{p}}}-1} = Fr^{\frac{z_{\mathrm{p}}^0}{z_{\mathrm{p}}}-1}, \tag{5.34}$$

where z_{p}^0 is the the lognormal distribution percentile corresponding to the equality:

$$q(\ell_{\max}) = q(\ell_0)\,. \tag{5.35}$$

By plugging-in (5.30) and (5.33) into (5.35) it follows

$$(z_{\mathrm{p}}^0)^2 + 2\sigma z_{\mathrm{p}}^0 = z_{\mathrm{p}}^2 + 2\sigma z_{\mathrm{p}}\,. \tag{5.36}$$

The last equation has two solutions: one is z_{p}, that gives the maximum travel distance. The other one is

$$z_{\mathrm{p}}^0 = -z_{\mathrm{p}} - 2\sigma\,, \tag{5.37}$$

that corresponds to the minimum travel distance $\ell_0 > 0$. In this new framework, all the jumps of the firebrands fall eventually inside of the interval $[\ell_0, \ell_{\max}]$.
Since $\ell_0 > 0$, it is always possible to estimate the ratio $\ell_{\max}/\ell_0 = \kappa$. Thus, taking into account this ratio and repeating the computations above, the following expression for κ is found:

$$\kappa = \exp\left(2\sigma(z_p + \sigma)\right), \tag{5.38}$$

that leads to the expression of $\cos\omega_0$ as a function of κ:

$$\cos\omega_0 = Fr^{-\ln\kappa/\ln Fr} = \frac{1}{\kappa}. \tag{5.39}$$

The last equation found (5.39 is a connection between the critical angle ω_0 and the range of the possible travel distances represented by κ. It can also be interpreted as the inverse proportionality of the arc and the radius in a sector of the fixed area. In the limit case when ω tends to $\pi/2$, the projection of the wind tends to zero, that corresponds to the situation when the firebrand falls down into the burning area. In practice, the constrain to reduce computational efforts is [77]

$$\cos\omega \geq \max\left\{ \frac{1}{\kappa},\ Fr^{-1/2} \right\}. \tag{5.40}$$

5.10.2 Results and discussion of the Macroscale study

In this section we present the numerical results with special attention to the effects of varying the height of ABL. The results are based on the ones reported on [77].

Please note that all of the results are backed by statistical data of wildfires in Spain from 2012 to 2017 (documented in [77]). The key idea conveyed is that wildfires develop differently under different boundary layer conditions.

In all the simulations we consider a constant wind $U = 4.47$ ms^{-1}, and fire intensity $I_{\rm f} = 20$ MWm^{-1}. The changes in the values of $H_{\rm ABL}$ influence the values of μ, involved in the determination of maximum travel distance of firebrands $\ell_{\max}$ as well as the values of the turbulent diffusion coefficient D. See in this respect Table 5.17 and Figure 5.17.

The PDF of firebrand landing distance for different values of $H_{\rm ABL}$ is portrayed in Figure 5.20. From formula (5.30) we have that $\ell_{\max}$ depends linearly on $H_{\rm ABL}$ and, due to the lognormal distribution of landing distance, the resulting changes in $\ell_{\max}$ do not strongly affect fire-spotting. From (5.22) we have that the turbulent diffusion coefficient depends on $H_{\rm ABL}$ with the third power and the resulting changes affects strongly the propagation of the fire with an important effect on the merging of the secondary fires generated by the fire-spotting. In fact, during stable conditions a larger number of fires with respect to unstable conditions is observed but at the same time a lower burned area is computed, showing that in unstable conditions the turbulent heat transport is stronger with the double effects of a faster fire propagation and a more efficient merging.

	Case 1: Stable			*Case 2: Unstable*			
$H_{\rm ABL}$ [m]	100	200	400	800	1000	1200	2000
μ [m]	4.0325	4.3879	5.0985	6.5198	7.2304	7.9411	10.7836
D [m^{2}s^{-1}]	0.0533	0.0995	0.1858	0.3467	0.4238	0.4993	0.7908

Table 5.17: Effects of atmospheric stability on the fire-spotting parameters.

Stable atmosphere: $H_{\rm ABL}$ varies from 100 m to 500 m.

If an air parcel displaced from the original height returns to the original height, then the the atmosphere is considered in *stable condition*. The so called *stable boundary layer* forms at night over land (it is also known as *nocturnal boundary layer*). It grows to depths of about 100 to 500 m, see for further information [264]. Due to the stability condition, clouds lire in layers and winds are steady and light. The smoke drops back down to the ground, and the convective column is not high, and this affects both fire-spotting and turbulent transport of heat. The contour lines of the effective indicator function is plotted in Figure 5.18. Please note that when the corresponding value of the effective indicator function reaches the threshold value $\phi_{\rm e}^{th} = 0.5$, the corresponding grid point is marked as burned, as shown in the scheme of Figure 5.2. The plots of 5.18 show the growth of the burning area after 37, 70 and 119 minutes after the ignition, respectively.

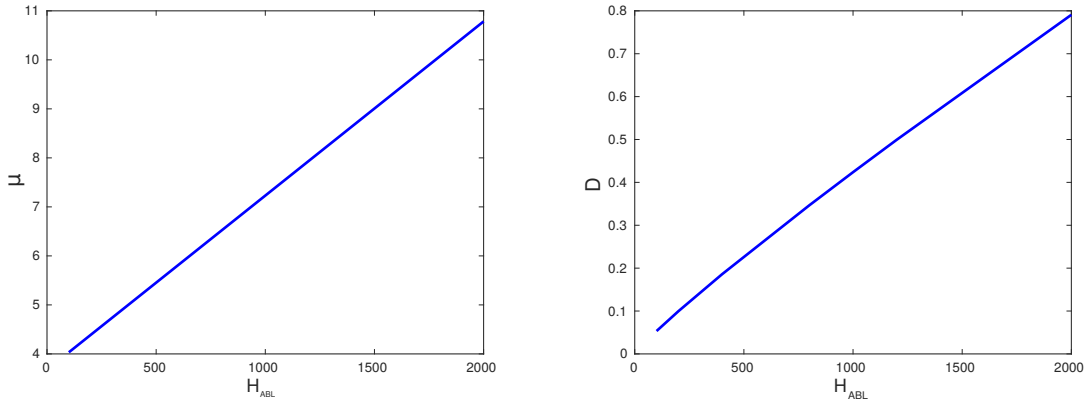


Figure 5.17: Plots of the fire-spotting parameter μ (left) and the diffusion coefficient D (right) varying H_{ABL}.

Unstable atmosphere: H_{ABL} varies from 800 m **to** 2000 m**.** When an air parcel after displacement from the original height accelerates upward due to buoyancy effects, the atmosphere is called *unstable*. It is usually observed during the day and afternoon. The depth of the boundary layer varies according to clime classes and seasons. The well known Köppen-Geiger climate classes are used to categorize the world-wide characteristics of the boundary layer, see e.g. [174]. In such a way, analyzing seasonal mean diurnal cycle of ABL depth for climate class Dfb (Cold with warm summers and no dry season, in summer and winter) given in [174], it is found that the diurnal H_{ABL} takes value in the range between 1000 m and 1800 m during the summer, and between 600 m and 800 m during winter. For other climate classes, H_{ABL} differs substantially (it may rise up to 3000m). Fire contours under the unstable atmospheric conditions is portrayed in Figure 5.19.
By comparing Figures 5.18 and 5.19, it can be observed that the generation of secondary fires is faster in the unstable case (see both panels (b)), and the correspondent burned area is also larger. Likewise, the merging of secondary fires in the unstable case is more rapid (see both panels (c)). It is worth noting the emergence of patterns of not burned areas surrounded by the fire, see 5.19(c). The plots in Figure 5.20 display the lognormal distribution for two different values of the height of atmospheric boundary layer H_{ABL}, following the proposed `RandomFront` parametrization. Larger values for H_{ABL} correspond to a growth in the probability of long ember jumps.
Figure 5.21 shows the total number of burned points simulated with several values of H_{ABL} at different times. After an initial interval of time of dormancy with limited growth, the burned area starts growing rapidly according to a power law.
It can be observed that the dormancy interval ends first under unstable atmosphere, and that the burned area during unstable atmospheric conditions is always larger than during stable atmospheric conditions.

To detect the main reason behind the rapid boost of the burned area, whether it is due to turbulence or fire-spotting, a further series of simulations with fixed diffusion coefficient D is performed. By doing so,D is independent of the atmospheric conditions and does not adopt `RandomFront` parametrization of Section 5.5. The results for stable ($H_{\mathrm{ABL}} = 100$ m) and unstable ($H_{\mathrm{ABL}} = 1000$ m) atmospheric conditions are shown in Figures 5.22 and 5.23. All the parameters are set equal to the previous simulations, but with fixed $D = 0.0387$ m^{2}s^{-1}. Note that this fixed value of D is less than any other value followed by the dependency on the H_{ABL}. Fire front propagates in the same way in both atmospheric conditions, displaying an equal number of secondary fires and equal burned area (see Figure 5.24). Thus, the atmospheric stability conditions affects the fire propagation mainly through the turbulent diffusion coefficient. Hence, comparing Figures 5.19 and 5.23 we observe that the number of independent fires is less during unstable conditions and by taking into account also Figures 5.21 and 5.24, we conclude that the atmospheric conditions affect the propagation of wildfires through the heat turbulent transfer. In particular, during unstable conditions we observe that turbulence pushes the front resulting into a faster propagation causing an increasing of the burned area and a more rapid merging of independent fires such that during unstable conditions the number of independent fires is less than during stable conditions in spite of the fact that the burned area is larger.

The numerical results show that the inclusion of the atmospheric stability conditions in terms of the ABL height in the parametrization changes substantially the form and the speed of the fire front. Nonetheless, the effect of atmospheric stability is significantly stronger in the fire spread phenomenon when it acts on turbulent heat transfer than when it acts on fire-spotting. Speaking in terms of the adopted physical model, H_{ABL} determines more noticeable effects due to its inclusion in D (following Equation 5.22) than due to its inclusion in μ (following Equation 5.19) This result is in line with the main conclusions of Section 5.9, where the Sobol' importance measure related to D outclassed the ones related to μ for both observables, burnt area and MSR.

Let us now consider historical data on fires in Spain from 2012 to 2017 from The Earth Observing System Data and Information System (`https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data`). The database contains information about the acquisition time but not about the corresponding atmospheric stability. In order to distinguish between stable and unstable conditions we take into account two extremely different cases [264]: diurnal ignitions during the summer months and nocturnal ignitions in winter months. The historical data is retrieved from two satellites characterized by different resolution: MODIS C6 with resolution 1 km and VIIRS with resolution 375 m (see Table 5.18). The more precise satellite VIIRS is able to capture smaller wildfires and therefore it reports a higher number of ignitions, while MODIS C6 identifies wildfires larger than the resolution threshold.

Comparing the results in Table 5.18 one can find that the number of fires in unstable conditions captured by VIIRS satellite is almost 3 times larger than the observations by MODIS C6. In stable atmosphere the number of fires captured by VIIRS is usually more than 20 times larger. Taking into account the different resolution of the two satellites one can conclude that wildfires under unstable conditions are characterized by larger burning area, while under stable conditions the burning area is smaller with a significant increase in the number of wildfires.

Year	MODIS C6 Data (1km)		VIIRS Data (375 m)	
	Stable	Unstable	Stable	Unstable
2012	138	1150	1170	3302
2013	34	639	960	1580
2014	16	361	803	1243
2015	30	769	664	2167
2016	38	672	1029	1895
2017	64	671	1106	1866

Table 5.18: Comparison of number of wildfires in stable that corresponds to winter night time, and unstable that corresponds to summer daylight time, atmosphere observed by satellites with different resolution. Historical data: Spain, 2012-2017 years.

As it is mentioned above, stable boundary layer is also known as nocturnal boundary layer [264], that allows to suppose that, in general, the stable atmosphere is observed during the night and, consequently, the diurnal boundary layer is more unstable. Thus, the data collected from the satellites during the whole year is considered distinguishing the diurnal and nocturnal fires (see Table 5.19).

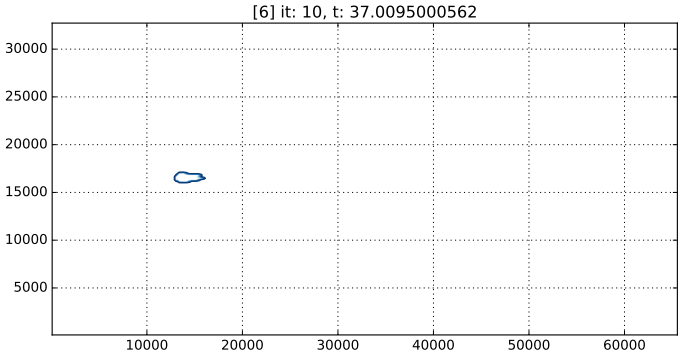
Bearing in mind that MODIS C6 satellite can capture only large enough burning areas, while VIIRS is capable of identifying smaller fires, data from Table 5.18 can be explained as follows. In the daytime wind is more erratic and strong, causing thus more unpredictable behavior of the fire propagation. On the other hand, at night the main front may propagate slower but with a larger number of new ignitions.

Year	MODIS C6 Data (1km)		VIIRS Data (375 m)	
	Day	Night	Day	Night
2012	4240 (68.5%)	1946 (31.5%)	8430 (39.4%)	12945 (60.6%)
2013	2631 (76.2%)	821 (23.8%)	6261 (45.3%)	7575 (54.7%)
2014	2181 (83.8%)	422 (16.2%)	5400 (42.8%)	7228 (57.2%)
2015	2317 (73.0%)	857 (27.0%)	5547 (41.8%)	7723 (58.2%)
2016	2328 (72.5%)	884 (27.5%)	7906 (42.9%)	5942 (57.1%)
2017	3578 (69.0%)	1606 (31.0%)	8288 (42.0%)	11470 (58.0%)

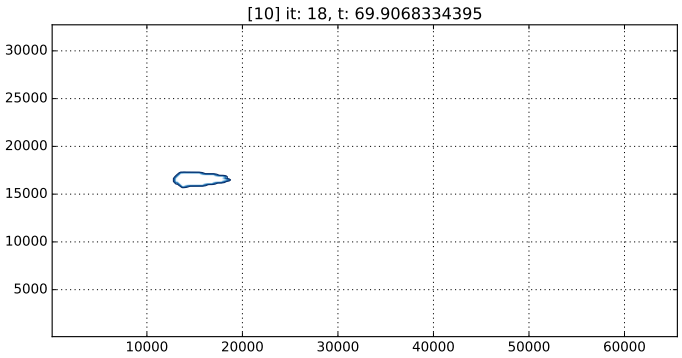
Table 5.19: Number and % of day and night fires in Spain. Historical data refers to 2012-2017 period.

Basing ourselves on official statistical data (`http://www.mapama.gob.es`), more than 50% of new ignitions in Spain are deliberate. We expect that the most of them occur during the night, that also explains the data presented in Table 5.19.

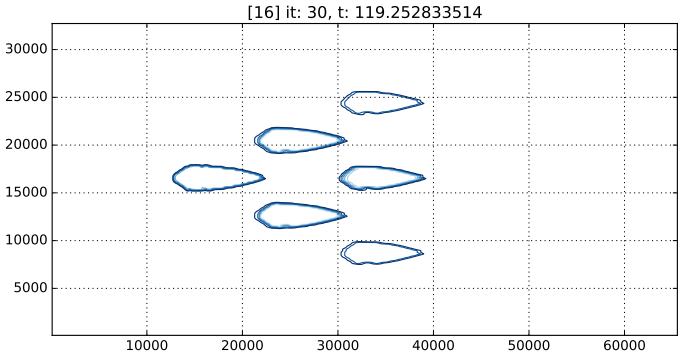
During the day the number of ignitions is more reduced, while the burning area is larger, since the percentage of the daytime fires obtained by MODIS C6 is higher than the percentage of night-time ones, and the contrary is observed for the VIIRS data. Hence, these satellite data may be interpreted in a similar way to the numerical results obtained through the model simulations.
The main physical mechanism that constitutes the claim of this data-supported numerical investigation is recalled in the following. During daytime, an higher ABL layer causes an higher convection cell, that besides augmenting the firebrand travel distance , increases the turbulent transport. Due to this phenomenon, the fire front propagation is faster and the merging of the spotting fires is accelerated. As a consequence, during the daytime the burning area is larger and the number of independent secondary fires is lower.
As it has been already remarked, such scenario is reproduced by the proposed model varying the H_{ABL} (that acts on D and μ in the physical parametrization). The feedback of the model on different H_{ABL} allows the model to be adapted to different climatic classes.



(a) $t = 37$ min.

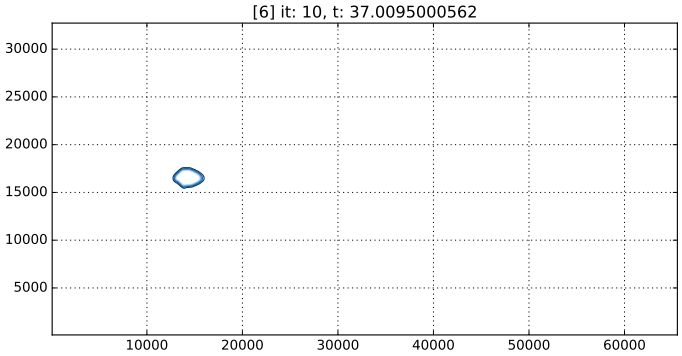


(b) $t = 70$ min.

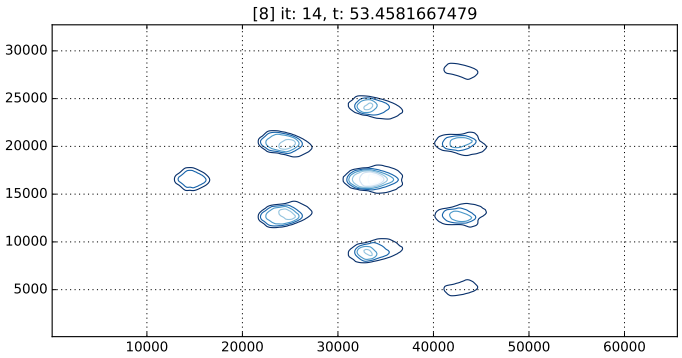


(c) $t = 119$ min.

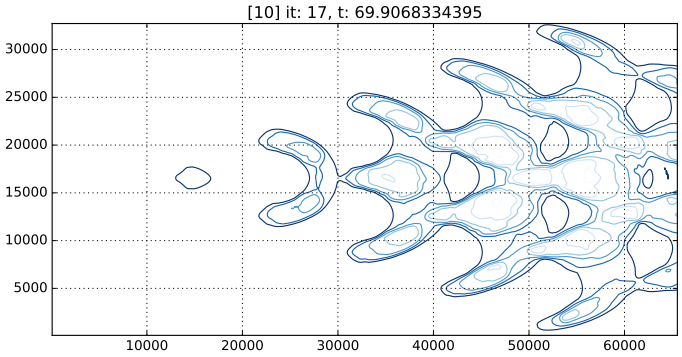
Figure 5.18: Fire front propagation during stable atmospheric boundary conditions: $H_{\mathrm{ABL}} = 100$ m, $\mu = 4.0325$ m, $D = 0.0533$ m^{2}s^{-1}.



(a) $t = 37$ min.

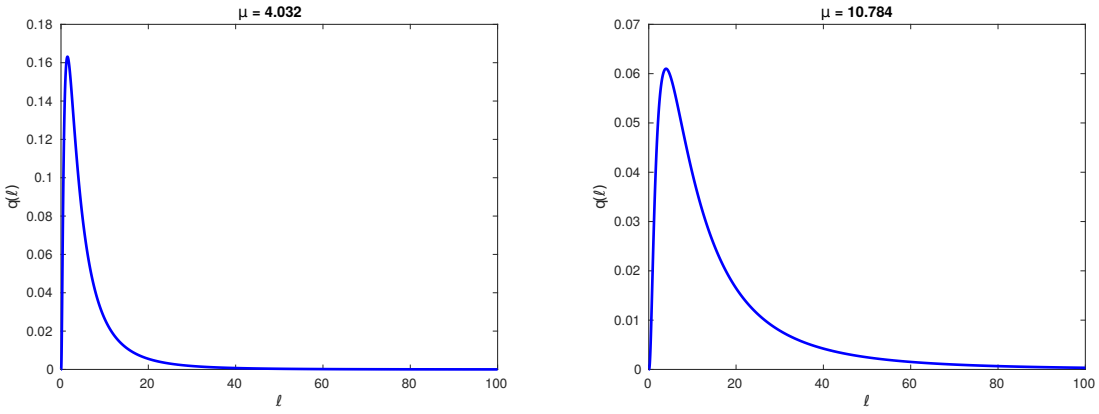


(b) $t = 53$ min.



(c) $t = 70$ min.

Figure 5.19: Fire front propagation during unstable atmospheric boundary conditions: $H_{\mathrm{ABL}} = 1000$ m, $\mu = 7.9411$ m, $D = 0.4238$ m^{2}s^{-1}.



(a) $H_{\mathrm{ABL}} = 100$ m. (b) $H_{\mathrm{ABL}} = 1000$ m.

Figure 5.20: Lognormal distributions for various values of H_{ABL}. The adopted values correspond to the stable and unstable atmospheric conditions, respectively.

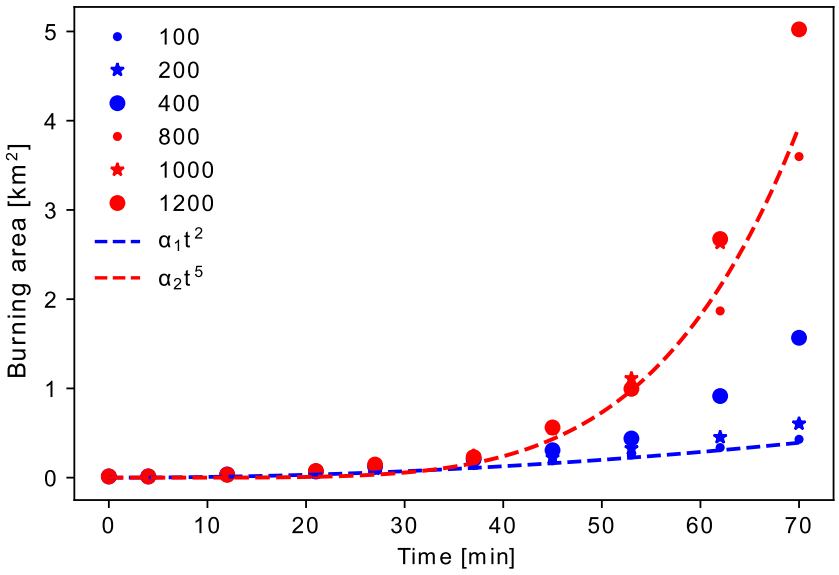
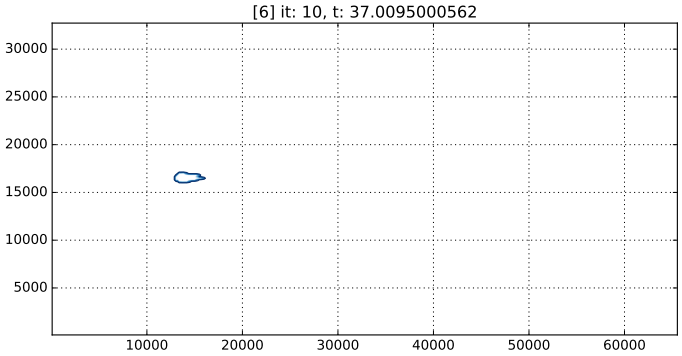
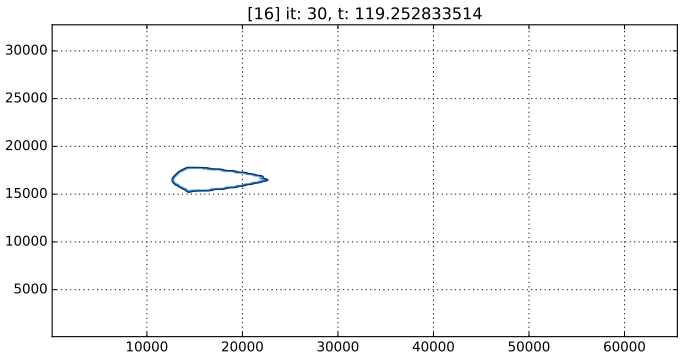


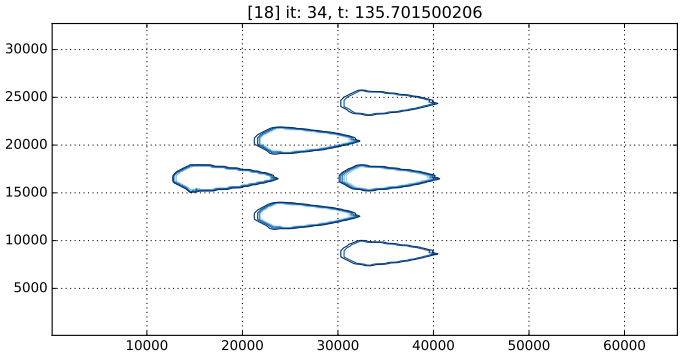
Figure 5.21: Burned area versus elapsed simulated time under stable (blue) and unstable (red) atmospheric boundary conditions. H_{ABL} varies from 100 m to 1200 m.



(a) $t = 37$ min.

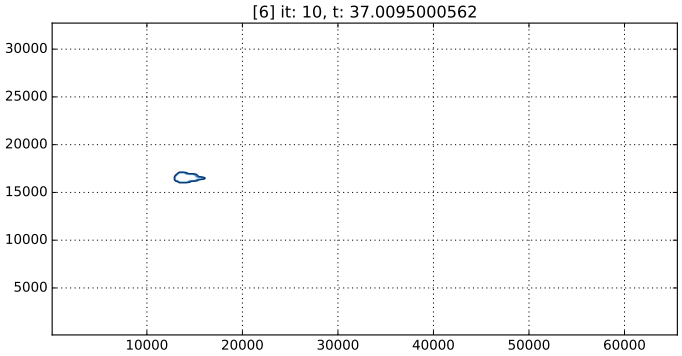


(b) $t = 119$ min.

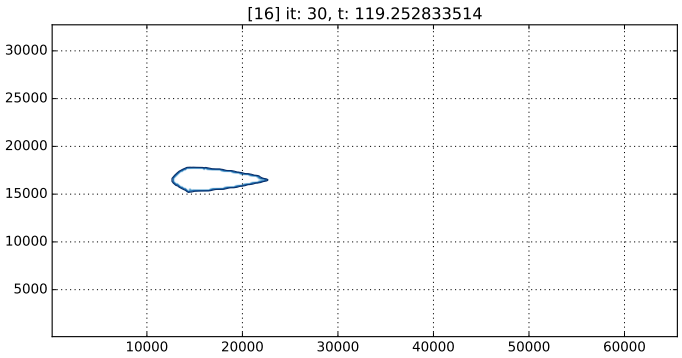


(c) $t = 135$ min.

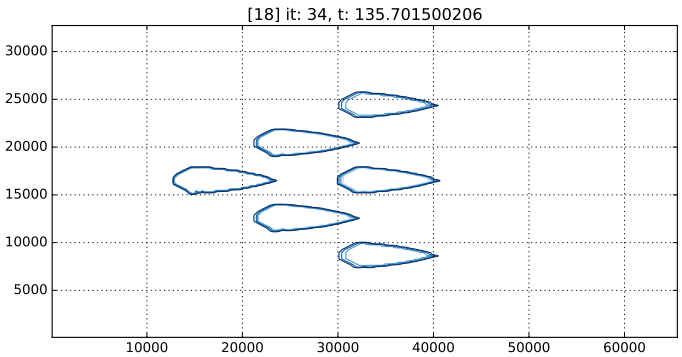
Figure 5.22: Fire front propagation during the night (Stable atmospheric boundary conditions: $H_{\mathrm{ABL}} = 100$ m, $\mu = 4.0325$ m, $D = 0.0387$ m^{2}s^{-1}).



(a) $t = 37$ min.



(b) $t = 119$ min.



(c) $t = 135$ min.

Figure 5.23: Fire front propagation during the daylight (Unstable atmospheric boundary conditions: $H_{\mathrm{ABL}} = 1000$ m, $\mu = 7.2304$ m, $D = 0.0387$ m^{2}s^{-1}).

We chose such parametrization of μ and σ in order to delineate the governing parameters for lofting (represented by μ) and transport mechanisms (represented by σ).

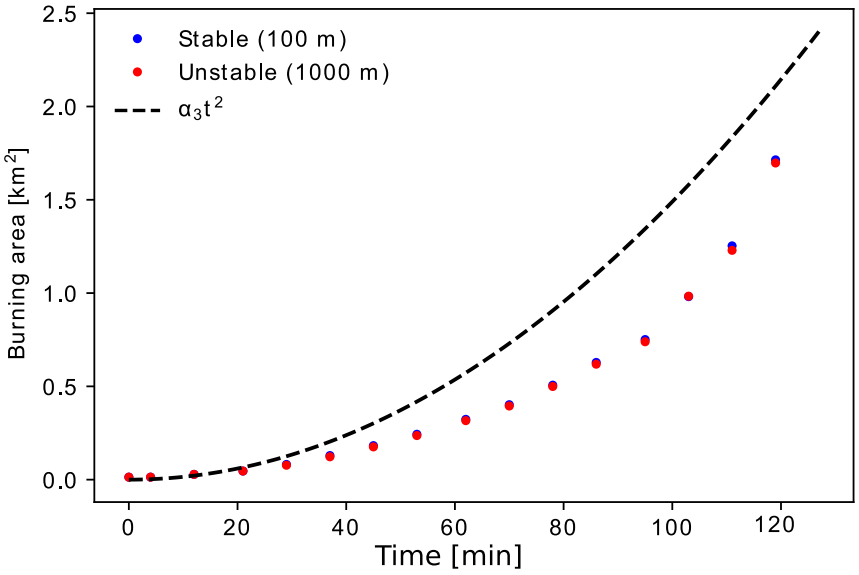


Figure 5.24: A comparison of the total burning area in time in stable (blue) and unstable (red) atmospheric boundary conditions in simulations with fixed diffusion coefficient $D = 0.0387$ m^{2}s^{-1}.

A biofilm is a community of microbes
bound together in a viscous, gooey blob.

Michael Lemonick, 2014.

Chapter 6

Applications to Biofilm Spread

This Chapter contains:

In this Chapter, the work developed on applications related to Biofilm modeling is reported. In Section 6.1, the biological phenomena is described. In Section 6.2, the studied model of statistical front propagation is adapted to a particular case of biofilm spread, that is, the case of planar

microbial biofilm. In Section 6.3, some of the UQ and SA techniques of Chapter 2 are adopted to analyze a model of biofilm growth accounting for species invasion. This last work is the same that is described in the recent paper of *Trucchia et al.* [285].

6.1 What is a biofilm?

It is well accepted that microorganisms lead social lives and exhibit complex behavior responding to the extracellular environment and other organisms. Experimental activity in fact highlighted that both in natural and artificial environments, microorganisms preferentially exist in the form of self-organized assemblages termed "biofilms", consisting of surface-associated communities embedded in an exopolysaccharide matrix and organized into microcolonies [94, 262]. Such exopolysaccharide matrix corresponds to extracellular polymeric substances that are secreted by microorganisms into their environment and that play a crucial role in the cell attachment to a given surface and in the biofilm formation overall. *Wilderer and Characklis* in [49] defined biofilms "a layer of prokaryotic and eukaryotic cells anchored to a substratum surface and embedded in an organic matrix of biological origin". The importance of biofilms has emerged since their first scientific description in 1936 [321] and the recognition of their omnipresence in the 1970s ([169, 59]). Biofilms are complex microbial communities that drive biogeochemical cycling processes of most elements in nearly any environment(soil, water, sediment and subsurface). *Costerton et al.* in [60] estimated that planktonic microorganisms (that is, microorganisms that do not organize themselves into biofilms) constitute less than 0.1% of the total aquatic microbial life. Attached to a surface (*substratum*), microorganisms inside of biofilms adopt coordinated chemical and physical interactions, enabling cells to develop efficient survival strategies [190]. As a matter of fact, bacteria in biofilms (sessile bacteria) differ substantially from free-living bacterial cells (also called planktonic bacteria) by the means of a set of emerging properties, such as the formation of physical and social interactions, the increased tolerance to antimicrobials and the enhanced rate of gene exchange [94]. The key idea is that the ability of bacteria to form biofilms attaching to surfaces can thus become an important competitive advantage over bacteria growing in suspension. As an example, bacteria in suspension can be washed away with the water flow, while attached to a biofilm microorganisms are protected from the hydrodynamics of the bulk liquid and can then keep growing in locations characterized by the presence of nutrients supply. The physical structure of biofilm itself encourages distinct biological niches that allow the growth and survival of microorganisms that could not compete successfully in a completely homogeneous system. Furthermore, specific microbial activity in biofilms can modify the internal environment (e.g., pH, O_2 , metabolic products, or disinfectant concentration) to make the biofilm more hospitable than the bulk liquid [234]. A single biofilm structure can host different species of microorganism that can contribute to the conversion of different organic and inorganic dissolved substrates. For

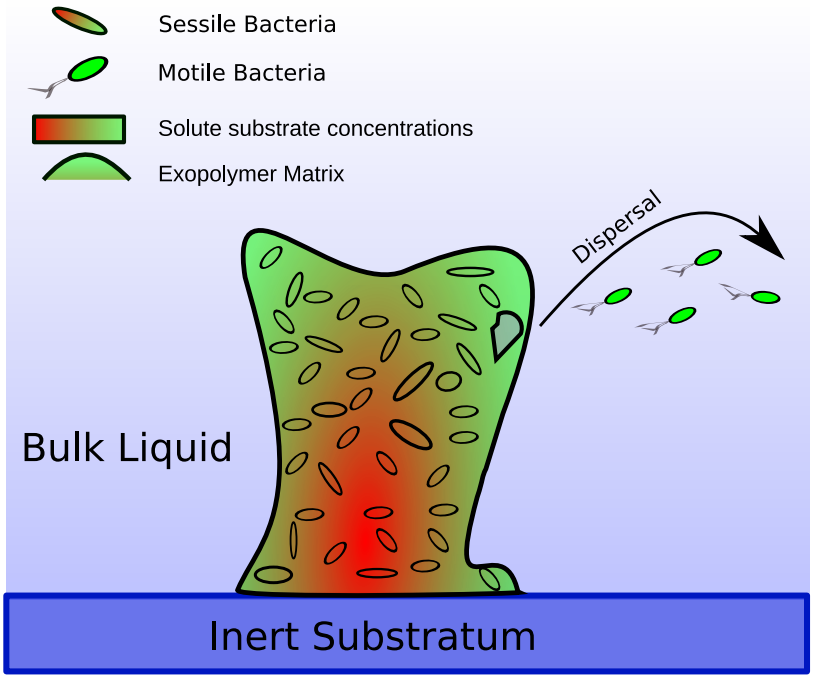


Figure 6.1: Schematic representation of a biofilm. A micro colony of bacteria is embedded in a self produced exopolysaccharide matrix, anchored to an inert substratum and submerged in a bulk liquid (even though sub-aerial biofilm also exist). While bacteria are rather rigid bodies (particulate phase of the biofilm),the several dissolved substances that play an active role in the metabolism of the bacteria which compose the biofilm (dissolved substrates) can diffuse through the ECM.

example, when a wastewater contains a mixture of conventional and xenobiotic organic pollutants, biodegradation of the xenobiotics requires a population of slow-growing bacterial organisms - those capable of degrading the xenobiotics. The slow growing ones could be washed out of a suspended-growth process, since all the biomass has the same growth rate, which normally is controlled for the benefit of the bacteria which degrade the conventional organic pollutants. However, when bacteria organize themselves as a biofilm, the slow-growing bacteria can establish themselves deeper inside the biofilm, protected thus from loss, while the conventional pollutants are removed near the interface between biofilm and the bulk liquid. [235].

Impact of biofilms on society

In general, biofilms play a significant role in various natural and engineered systems. Their action ranges from being deleterious to human activities and health to being a strong ally in several situations. [303].

To name a few example of detrimental biofilms, we cite the biofilm-related problems in dental hygiene (dental plaque), infectious diseases (e.g., cystic fibrosis), and infections related to the presence of bacteria in medical implants (e.g., catheters, contact lenses, heart valves). Biofilms are

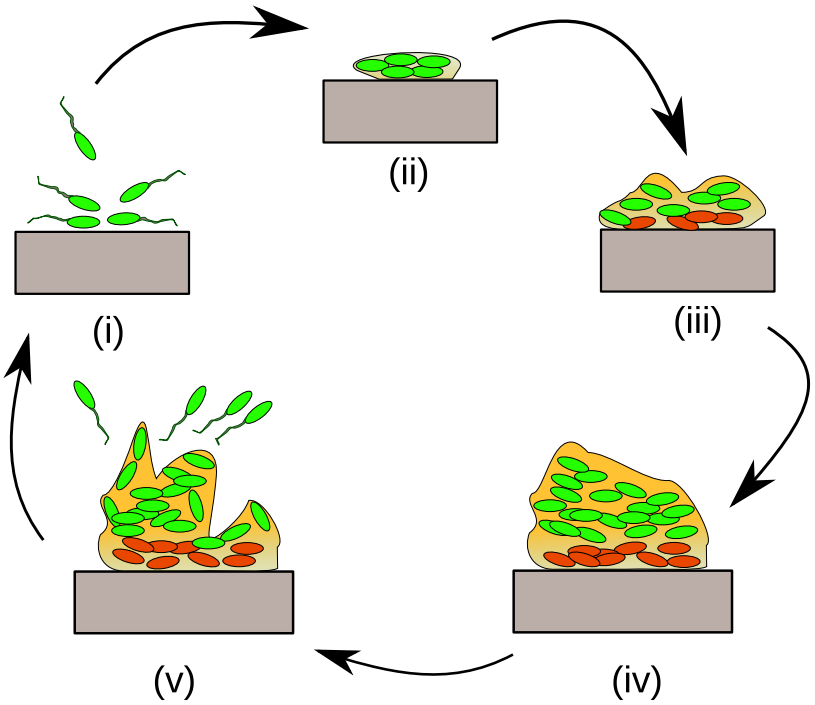


Figure 6.2: Schematic representation of the biofilm life cycle stages ([260]): (i) Reversible attachment of planktonic bacteria. (ii) Irreversible attachment. (iii) Micro colonies formation. (iv) Mature biofilm development. (v) Biofilm detachment/seeding and dispersal.

also responsible for biofouling and process water contamination, microbially influenced corrosion and quality deterioration of drinking water.

Conversely, there are many examples of biofilms which are good for society.

For example, biofilms have been extensively used in biotechnological applications such as waste-water and solid waste treatment, drinking water filtration, biofuel production. In coastal areas, rivers and lakes, a large fraction of bacterial activity is located in biofilms present in stones and sediments. Biofilms also occur in nature in soils and on the roots of plants. All these above mentioned naturally occurring biofilms are of paramount importance for cycling nutrients in the Biosphere.

Biofilm life cycle and dispersal

Cells in biofilms experience developmental programs which result in an ordered and predictable transition through several distinct stages. Each of these stages is based on stage-specific expression of genes [93]. Such biofilm developmental program culminates with the release of free-living cells that are able to colonize new habitats, possibly richer in resources [173]. There are a total of five stages [260], which are represented in Figure 6.2.

In the first stage, *reversible attachment* the formation of microbial biofilms begins with the reversible adhesion of a small number of cells to a surface. The initial attachment between bacteria

and the surface is governed by non-specific interactions such as electrostatic, hydrophobic, or van der Waals forces.

After binding to the surface through the glue-like matrix, bacterial cells begin the process of irreversible adhesion, proliferation, and accumulation as multi-layered cell clusters. This is the second stage of biofilm life cycle, i.e. *irreversible attachment*.

The attached cells then synthesize new exopolysaccharide material in order to cement their adhesion to the surface, and to other cells in the developing biofilm. This extracellular matrix, composed of a mixture of biomaterials such as polysaccharides, proteins, nucleic acids, and other substances, is considered essential in holding bacterial cells together in the biofilm structure, in helping to retain nutrients for cell growth, and in protecting cells from dehydration and the effects of stress. Once having irreversibly attached to a subtratum, bacterial cells undergo phenotypic changes. This phase, named *micro colonies formation*, is characterized by active binary division of attached cells and cell recruitment.

The fourth stage is the one of *mature biofilm development*. Biofilms in fact typically acquire a three-dimensional structure when they reach maturity. The cell accumulation requires coordinated efforts from the microbial community to produce a well-organized structure. Mature biofilms typically consist of differentiated mushroom-like structures of cells embedded within extracellular polymer matrix, which contains voids open to the bulk fluid, to allow the transport of nutrients from the interface to the inner parts of the biofilm, and removal of metabolic wastes. These three-dimensional structures characterized by macrocolony morphology rely on self-produced extracellular matrix components. EPS, amyloid-forming proteins, adhesins, and exopolysaccharides (all inside of the biofilm matrix) are necessary to generate such structures, where spatial distribution (gradients) of water,nutrients, signaling compounds (and waste prod- ucts) can be noted along the spatial extent of the biofilm, changing thus the methabolism of the cells.

The fifth (and last) stage of bio-film life cycle is when bacterial cells detach from the biofilm or seed reentering the planktonic state, and may start a new biofilm formation cycle in other settings. While detachment is a passive process of cell loss resulting from sloughing of cells and/or erosion from the biofilm, active or seeding dispersal is coordinated via regulatory systems in response to a number of cues (e.g., alteration in the availability of nutrients, oxygen depletion, levels of iron) and signals (e.g., acyl-homoserine lactones, diffusible fatty acids, cell-cell autoinducing peptides) [103]. Thus, seeding dispersal can occur in the complete absence of flowing conditions, and does not depend upon shear forces that removes cells from the biofilm. Another interesting feature of seeding dispersal is that cells appear to have a distinct phenotypes different from those of biofilm and planktonic cells, increasing cell ability to colonize a greater range of habitats important for niche expansion [54, 63]. Thus, dispersal represents an important adaptive strategy with profound impacts on the survival and fitness of microorganisms. It allows biofilm populations to spread and

colonize new surfaces, avoiding overcrowding, depletion of resources and competition among cells in the local environment, and promoting the rejuvenation of biofilms [21]. Furthermore, dispersal is linked to the generation or maintenance of genetic variation, with significant outcomes for the success of those bacteria in the environment [227, 54, 63].

Although dispersal is advantageous from the microbial standpoint, it may negatively affect some industrial and medical processes. For instance, through dispersed cells, biofilm can spark new infections within the host and result in the transmission of bacteria between different hosts [132]. Furthermore, dispersal may promote, for example, the spread of parasitism phenomena in animals and plants [297], biodeterioration of historical and artistic objects [48, 298] and fouling in food-processing equipment [47].

The existence of a programmed generation of dispersed cells appears increasingly clear, but the challenge now is to provide the mechanistic understanding of biofilm dispersal.

In 6.2, we propose a modeling approach to study the growth of mono-layer microbial biofilm on inert surfaces by focusing on the biofilm spread induced by dispersal, predicting the formation and growth of satellite colonies generated by dispersing biofilms.

6.1.1 Mathematical modeling of biofilms

Understanding the mechanisms of biofilm formation, growth, and detachment/dispersal plays an important role for promoting biofilms which help human activities while reducing the damage provoked by detrimental ones. Without any doubt, mathematical modeling is one of the essential tools for shedding some light on such complex biological phenomena.

Since most biofilms are complex systems [303], a biofilm model that attempts to capture all its complexity should necessarily include mass balance equations for all processes occurring for all biofilm components, continuity and momentum equations for the fluid inside and outside the biofilm , and boundary conditions for every variable in the interface with bulk liquid. The implementation of such an exact model is impractical, at least. As a consequence, even the most complex biofilm models make use of several assumptions and simplifications. Most biofilm models capture therefore only a fraction of the total complexity of a biofilm system, but they still prove to be rather useful. Simplifications, as a matter of fact, are a natural and necessary part of any kind of mathematical model.

Many biofilm models have been proposed in the literature over the last decades [172, 129]. Some of them have been derived in the framework of continuum mechanics and formulated as differential equations based on (mass, volume, momentum, energy) conservation
principles [304, 75, 7, 57, 318, 228]. Others have been introduced as bottom-up models and assume biofilms to be inherently stochastic living systems [220, 134, 273, 120, 140]. Still, biofilm modeling remains a challenge, in particular since the biological processes involved in biofilm formation and

growth are highly nonlinear and since there is no agreed-upon methodology to guide the user in the selection of the most appropriate model(s) and in the choice of the input parameters. For instance, no reference values have been defined for these inputs [35], while they may affect the nonlinear system in unpredictable ways.

An interesting view that links this biological topic with the front propagation in random/complex models is the fact that biofilm is made of bacteria that instead of living dispersed in an aqueous media, prefer to live in a neatly separated structure. It is easy to detect thus a front that advances due to the push of biological activity on the biofilms, and may be subject to external velocity field of hydrodynamic nature [272]. In the following Sections, this view of biofilm in the context of statistical front propagation is pursued in two different context. In Section 6.2, the statistical front propagation model of Chapter3 is used to model planar biofilm that perform (random) seeding and dispersal of bacteria, creating new colonies. This work corresponds to the recent report [288] In Section 6.2, conversely, a one-dimensional biofilm model of engineering relevance that includes for planktonic invasion in a biofilm composed by different species is studied. This work corresponds to the recently published paper [285]. For the sake of such study, UQ and SA with some of the mathematical instruments presented in Chapter 2 is performed. In this case, the biofilm is seen as growing in only one direction, and the front is just a point separating the biofilm phase and the bulk liquid. Nevertheless, since its evolution depends on a set of parameters that are supposed of random nature (because no precise information can be inferred), we have indeed a front that propagates as a random process.

6.2 Model Application: Planar Bioilm Modeling

In this work, summarized in the recent report [288], we focus on a planar geometry, neglecting the vertical growth of the biofilm, to better catch a bidimensional pattern of biofilm colonies all over the surrounding environment. Such planar setting is proper of biofilm growth in oligotrophic environments (e.g., reverse osmosis membranes, stone monuments, surgical gauze, contact lenses, water supply pipes), where nutrient constraints limit microbial growth to thin mono-layered biofilms. As already stated in the previous Section, the growth of this biofilm is characterized by two main phenomena: the biomass expansion due to the growth of primary existing colonies, and the formation of new colonies due to the attachment of dispersed cells released by the primary ones, i.e., seeding dispersal. The biofilm colony growth is modeled by using the Level Set Method introduced in Chapter 1, while seeding dispersal is simulated, using the same modeling approach of Chapter 3, through the Probability Density Function (PDF) corresponding to the diffusive process that governs the bacteria dispersal behavior. The concepts of Chapter 3 are in the following adapted to this biological application.

The surface of mature biofilm colony Ω is generally composed by an ensemble of biofilms spots Ω_i with $i = 1, \ldots, n(t)$ where the total number n depends on time t because of possible merging and birth of colonies. Let $\varphi : \mathcal{S} \times [0, +\infty[\to \mathbb{R}$ be a function defined on the domain of interest $\mathcal{S} \subseteq \mathbb{R}^2$ such that the iso-line $\varphi(\mathbf{x}, t) = c$ describes the evolution the boundaries of Ω_i, i.e., the evolution of the colonies fronts. Then the motion of the fronts of biofilm colonies is determined by the Level Set Equation:

$$\frac{\partial \varphi(\mathbf{x}, t)}{\partial t} = u(\mathbf{x}, t) \left\| \nabla \varphi(\mathbf{x}, t) \right\| . \tag{6.1}$$

For the sake of simplicity, the outward normal velocity $u(\mathbf{x}, t)$ is assumed constant, i.e., $u(\mathbf{x}, t) = u$. Let the mature colonies be able to release a sufficient large number of cells whose dispersion is characterized by a random motion. Let $\mathbf{X}^{\omega}(t, \overline{\mathbf{x}})$ be the ω-realization of the trajectory of a dispersed cell with an average position $\overline{\mathbf{x}} = \overline{\mathbf{x}}(t)$ and initially located in $\overline{\mathbf{x}}(0) = \overline{\mathbf{x}}_0$, such that $\mathbf{X}^{\omega}(0, \overline{\mathbf{x}}) = \overline{\mathbf{x}}_0$. Cell trajectories are described by the one-particle density function $p^{\omega}(\mathbf{x}; t) = \delta\left(\mathbf{x} - \mathbf{X}^{\omega}(t, \overline{\mathbf{x}})\right)$, where $\delta(\mathbf{x})$ is the Dirac δ-function. Moreover, let the regions Ω occupied by the colonies be conveniently marked by an indicator function $\mathcal{I}_{\Omega}(\mathbf{x}, t)$. Then, an *effective indicator* φ_e, $\varphi_e(\mathbf{x}, t) : \mathcal{S} \times [0, +\infty[\to [0, 1]$, of the region surrounded by a random front is obtained by using the sifting property of the δ-function and by averaging the indicator function:

$$\begin{aligned}
\varphi_e(\mathbf{x}, t) &= \langle \int_{\mathcal{S}} \mathcal{I}_{\Omega}(\overline{\mathbf{x}}, t) \delta(\mathbf{x} - \mathbf{X}^{\omega}(t, \overline{\mathbf{x}}))\, d\overline{\mathbf{x}} \rangle \\
&= \int_{\mathcal{S}} \mathcal{I}_{\Omega}(\overline{\mathbf{x}}, t) \langle \delta(\mathbf{x} - \mathbf{X}^{\omega}(t, \overline{\mathbf{x}})) \rangle\, d\overline{\mathbf{x}} \\
&= \int_{\mathcal{S}} \mathcal{I}_{\Omega}(\overline{\mathbf{x}}, t) p(\mathbf{x}; t \,|\, \overline{\mathbf{x}})\, d\overline{\mathbf{x}} \\
&= \int_{\Omega(t)} p(\mathbf{x}; t \,|\, \overline{\mathbf{x}})\, d\overline{\mathbf{x}} ,
\end{aligned} \tag{6.2}$$

where $p(\mathbf{x}; t \,|\, \overline{\mathbf{x}}) = \langle \delta\left(\mathbf{x} - \mathbf{X}^{\omega}(t, \overline{\mathbf{x}})\right) \rangle$ is the PDF of the seeding bacteria. In this work, $p(\mathbf{x}; t \,|\, \overline{\mathbf{x}})$ is assumed to be Gaussian.

Function $\varphi_e(\mathbf{x}, t)$ in this application fields provides an estimation of the probability that dispersed bacteria cells arrive in a point $\mathbf{x}$ from different sources $\Omega_i(t)$. However, to relate this probability of arrival to a successful formation of a new biofilm colony spot, a criterion associated with a reversible/irreversible attachment due to environmental conditions and biological time scales is needed. A parallelism can be made between this concept and the time delay in the secondary fires appearances due to firebrand showers treated in Chapter 5. In both cases we thus recur to a storing function that included a time-delay parameter, ψ.

The integral field ψ is defined here as

$$\psi(\mathbf{x}, t) = \int_0^t \frac{1}{\tau(\mathbf{x}, \epsilon)} \varphi_e(\mathbf{x}, \epsilon)\, d\epsilon . \tag{6.3}$$

Such field stores at any point of domain $\mathbf{x}$ the signals received from the active biofilm (that is, Ω) during the temporal interval $[0, t]$.

Unlike forest fire applications of Chapter 5, time delay is here actively determined by the dynamics of biofilm and the environmental conditions of the surroundings.

We denote in fact by $\tau(\mathbf{x}, t)$ the timescale of signal storing ([288]), defined as

$$\tau(\mathbf{x}, t) = \tau_e(\mathbf{x}, t) + V(\mathbf{x}, t)\,. \tag{6.4}$$

In the last formula, $\tau_e(\mathbf{x}, t)$ represents the environmental distribution of resources *neglecting the action of biofilms*, while the field $V(\mathbf{x}, t)$ accounts for the resource depletion carried out by the biofilm.

The feedback mechanism, in the same way as seen in Chapter 5 between ψ and φ is given by the procedure

$$\psi(\mathbf{x}, t) \geq 1 \rightarrow \mathcal{I}_\Omega(\mathbf{x}, t) = 1\,. \tag{6.5}$$

In this biofilm modeling framework, formula 6.5 has the following biological interpretation: when into a considered spot a certain amount of dispersed cells have established and endured a certain amount of time (that accounts for the environmental availability of resources) then a new colony is generated. We recall that the indicator function $\mathcal{I}_\Omega(\mathbf{x}, t)$ results to be

$$\mathcal{I}_\Omega(\mathbf{x}, t) = \begin{cases} 1\,, & \text{if } \varphi\,(\mathbf{x}, t) \leq c \text{ or } \psi(\mathbf{x}, \mathrm{t}) \geq 1\,, \quad \mathbf{x} \in \Omega\,, \\ \\ 0\,, & \text{elsewhere}\,, \quad \mathbf{x} \notin \Omega\,. \end{cases} \tag{6.6}$$

Equation (6.4) portrays an trade-off between the availability of resources offered from the surrounding environment and the resource depletion carried out by the growth of the biofilm colonies. This simple formulation of the timescale for the waiting times of free cells seeding might generate different patterns of biological interest. In the following, the term τ_e is assumed constant in time, because it represents the availability of resources *before* the action of biofilm, and this is assumed to change slower than the biofilm evolution.

The term $V(\mathbf{x}, t)$ instead, is modeled by the following Poisson problem

$$\alpha \Delta V(\mathbf{x}, t) = \rho_b\,, \tag{6.7a}$$

$$V(\mathbf{x}, t)|_{\mathbf{x} \in \partial \mathcal{S}} = 0\,, \tag{6.7b}$$

where ρ_b is the bacterial density inside the colonies and α an absorption kinetic coefficient. In our case, the bacterial density inside the colony is constant, and the latter equation becomes

$$\alpha_* \Delta V(\mathbf{x}, t) = I_\Omega(\mathbf{x}, t)\,, \tag{6.8}$$

where α_* corresponds to α in the rescaled setting and differs for the physical dimensions. The dynamic governed by (6.8) depends only on α_* and, in spite of its simplicity, it manages to represent availability of biofilm resources, determining the temporal dynamics of seeding dispersal.

Computational resources In the following two Subsections, numerical routines are used to integrate the equations [6.1–6.8]. Such computations were done by using the facilities of BCAM by running an OpenMp-parallel finite difference `C/Fortran` code. Its routines rely on a general-purpose library written in `Fortran2008/OpenMP`, `LSMLib` (`http://ktchu.serendipityresearch.org/software/lsmlib/`). The latter provides robust and efficient tools to track the evolution of co-dimensional fronts moving in one-, two- and three-dimensional domains.So called Essentially non oscillatory (ENO) algorithms are employed to compute accurate space derivatives, while for the advancement in time a second order Runge–Kutta scheme was implemented.

6.2.1 Synthetic test case

In the numerical experiments performed, the simulation starts with 4 active biofilms spots located at a random places in the square $\mathcal{S} = [0, L] \times [0, L]$. Such active biofilms are represented for the sake of simplicity as 4 circular spots of the same radius $L/20$. The distribution of resources not affected by biofilm activity, τ_e, is here modeled as expressed in the model description, via 14 spots of resources of the same radius of the active biofilms, $L/20$. The level of resources is therefore low outside of a favorable spot , wich implies a bigger attachment time scale $\tau = 10.0$. On the other hand, inside of a favorable spot the time scale of attachment is low, and the attachment time has been set to $\tau = 0.01$. The normal velocity of the Level Set Equation 6.1 has been deliberately been set low, in order to highlight the effect of the seeding and attachment phenomenon rather than the biomass expansion.

Figure 6.3 shows the colonization of the surrounding environment with resources spot marked as red dashed circles, and active colonies as magenta solid contours. In the z-axis the quantity φ_e that is related to the probability of seeding and dispersal is plotted. Three different time steps are plotted, with increasing simulated time from left to right. At the last time step, due to the action of ϕ_e (stored in ψ), a new active colony is born in a resource spot in the lower right corner of the simulated domain. This new colony is born uniquely due to the effect described by the presented model, and not by the standard front advance carried out by LSM equation. Field of ψ of the last time-step of the latter simulation is portrayed in Figure 6.4.

Figure 6.5 shows the dynamics of a second synthetic experiment, with 6 spots of resources. The active colonies contours are colored from brown to yellow, according to the different time steps. The black dashed lines are isocontours of V field at the time of the first plotted active colony, while the purple spots are location of resources. Note that due to the action of the V field, not always the new biofilm spot is generated in the nearest available location. Formation of new colonized spots due uniquely to seeding can be observed.

Figure 6.6 shows the effect of the variation of α on the biofilm dynamics for this numerical setting (more specifically, on the seeding-storing field ψ). From left to right, the parameters assumed increasing values $\alpha \in (0.1, 0.15, 2.5)$. Higher values for α meant higher consumption of resources put in action by the colonies, that prevented thus the establishment of new ones in their surroundings.

6.2.2 Experimental test case

In order to prove the potentiality of the proposed approach, an experimental test case has been designed and realized. *Pseudomonas aeruginosa* strain PAO1 (MH873) was used in this study as a model system of bacterial biofilms. In fact, the metabolically versatile *P. aeruginosa* PAO1 is an opportunistic pathogen of plants, animals, and humans and is ubiquitously distributed in soil and aquatic habitats. Furthermore, the bacterium is genetically characterized and amenable to mutagenesis and "omics" based approaches [263, 301]. The microorganism was maintained at -80°C in suspensions containing 20% glycerol and 2% peptone, and was grown aerobically in Tryptic Soy Broth (TSB medium) for 15h at 30°C. Dispersion experiments were conducted by using the colony-biofilm culturing system. Briefly, 2 sterile black polycarbonate filter membranes (0.22 μm pore size and 25mm diameter) were placed in each Petri dish containing Tryptic Soy Agar (TSA medium), at a distance of 2 mm from each other. Bacterial cells are trapped completely by the membrane filters having a pore size smaller than the bacterial size, while nutrients and metabolites diffuse across membranes easily. Fifty μl of cell suspension containing 1×10^8 cells were used to inoculate the central filter membrane. The plates were incubated at 30°C for 72h. Every 24h the Petri dishes were observed, and the dispersal phenomenon was documented by capturing imagines with both a camera and a stereomicroscope (magnification 12X). Numerical solution of model (6.1)–(6.8) has been computed by setting the physical non-dimensional parameters as follows: $\alpha_* = 0.05\,\mathrm{LT}^{-2}$, $u = 10.0\,\mathrm{LT}^{-1}$ inside the membranes and zero outside, and the diffusion coefficient of the Gaussian PDF equal to $10^3\,\mathrm{LT}^{-2}$. The numerical set-up is based on a 2D mesh $[0, 220]$x$[0, 370]$ with grid step $\delta x = \delta y = 1.0L$. The numerical test concerns the two membranes: the inoculated one and seeding target (the two external dashed lines in Fig. 6.7). These circular membranes have radius $R = 60$ in grid step units and center in $(110, 118)$ (the inoculated membrane) and $(110, 252)$ (the target membrane). At the initial instant, a mature biofilm colony is assumed to be present in the inoculated membrane with circular profile centered in the center of the membrane and radius $r = 35$. Furthermore, the availability of the environmental food needs to be set and it is represented by τ_e in (6.4). In particular, τ_e is assumed to be constant in time and ranging through a linear interpolation procedure from 0.01T, when $\mathbf{x}$ is inside the inner disk with radius $< 0.70R$, to 600.00T, when $\mathbf{x}$ is outside the membranes (see the

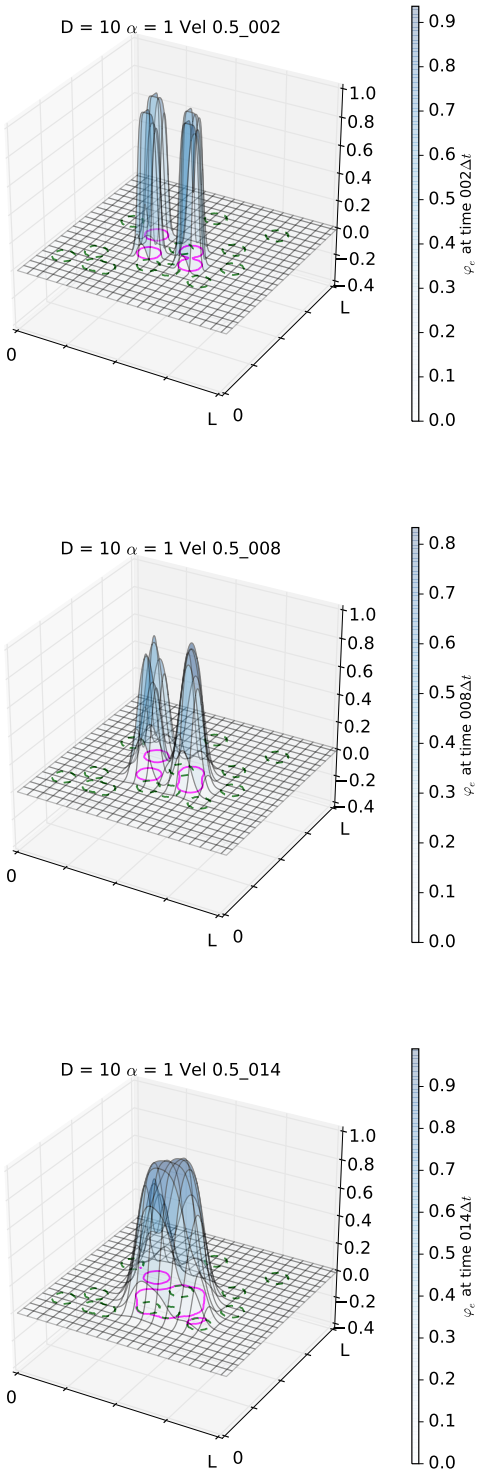


Figure 6.3: Colonization of the surrounding environment with resources spot marked as red dashed circles, and active colonies as magenta solid contours. In the z-axis the quantity φ_e that is related to the probability of seeding and dispersal is plotted. Three different time steps are plotted, with increasing simulated time from top to bottom. At the last time step, due to the action of ϕ_e (stored in ψ), a new active colony is born in a resource spot in the lower right corner of the simulated domain.

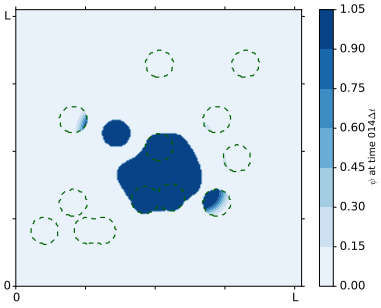


Figure 6.4: ψ field of the last time-step of the simulation described by Figure 6.3. The favorable spots (here dotted lines), far from the action of the active colonies, are colonized with the help of the storage function ψ.

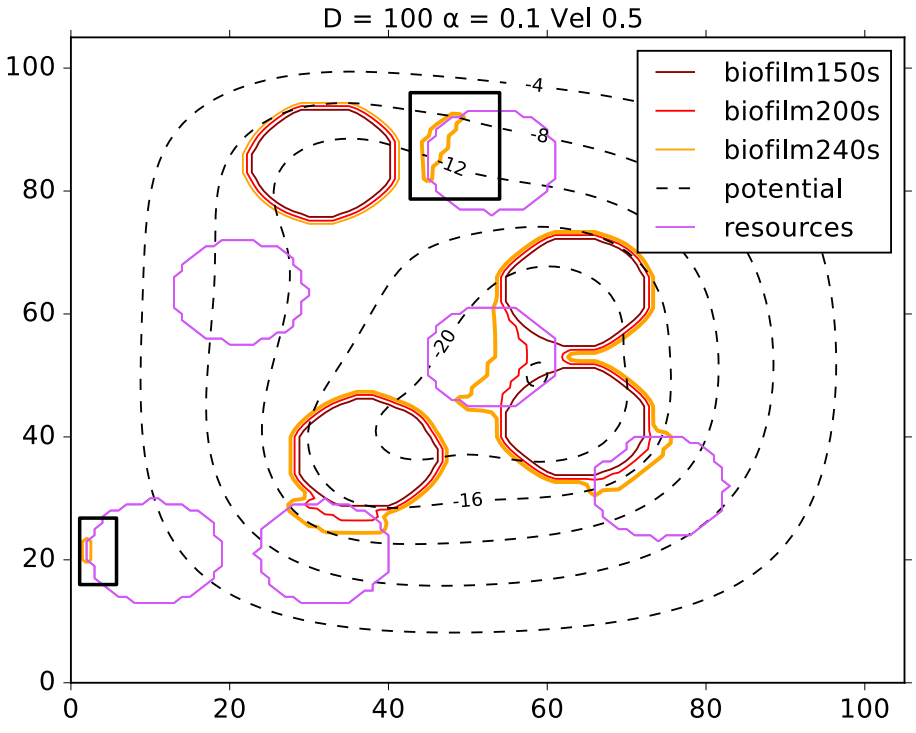


Figure 6.5: From brown to yellow, time evolution of biofilm spots. The black dashed lines are isocontours of V field, while the purple spots are location of resources. Note that due to the action of the V field, not always the new biofilm spot is generated in the nearest available location. The part of the domain that have experienced a biofilm colonization that must be due uniquely to seeding and dispersal effect is highlighted by a black square.

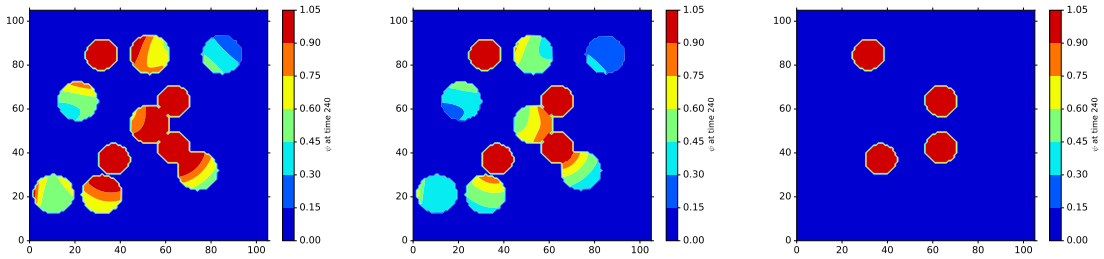


Figure 6.6: Plot of the ψ field, for increasing α (from left to right, $\alpha \in (0.1, 0.15, 2.5)$, at a fixed time $t = 240\Delta t$. This comparison shows the crucial role of biofilm resource consumption in the dynamic of biofilm colonization of a complex environment.

dashed circles in the right side of Fig. 6.7). This assumption corresponds to a very large timescale for generating a new colony outside the membranes, which corresponds to unfavorable conditions. Figure 6.7 shows the growth of a primary colony in the inoculated membrane and its colonization of the target membrane by a seeding dispersal mechanism both for the experimental data (the pair of membranes on the left) and the proposed modeling approach (the pair on the right). The Level Set Method describes the growth of the colony: first the primary one that is living in the inoculated membrane (left side membrane) and later the secondary one in the target membrane. The seeding and the attachment mechanism, which are responsible for the colonization of the target membrane, are well reproduced by the model. In spite of the fact that the present comparison is qualitative, it shows that the present approach is able through its modular structure to model the growth of the biomass colony and to take into account the different processes that simultaneously occur. In particular, the present approach provides a method to link a sharp interface model for the growth of biofilm colonies and a statistical treatment for biofilm seeding. The modular structure allows for a detailed front propagation through a more detailed expression for the normal velocity of the colony front $u(\mathbf{x}, t)$ and a more detailed bacterial migration through a new statical characterization. The comparison between the experimental pictures and some frame of evolution of the model is promising and, thanks to the modular structure, the present approach emerges as a novel and useful method for understanding the complex dynamics displayed by microbial biofilm.

Conclusions

The work described so far shed some light on the possibility of linking a sharp interface model for biofilm spots growth and a statistical treatment of biofilm seeding. That should help model and understand dynamics that are of paramount importance in a various spectra of applications, given that any new stochastic characterization of bacterial migration could be easily implemented in this framework. As a final remark, one of the main motivations for researching on biofilm dispersal is to provide a mechanistic model to predict how cells attach and proliferate, seeding then new biofilms. A better understanding of the evolution of dispersed cells may offer a broad conceptual

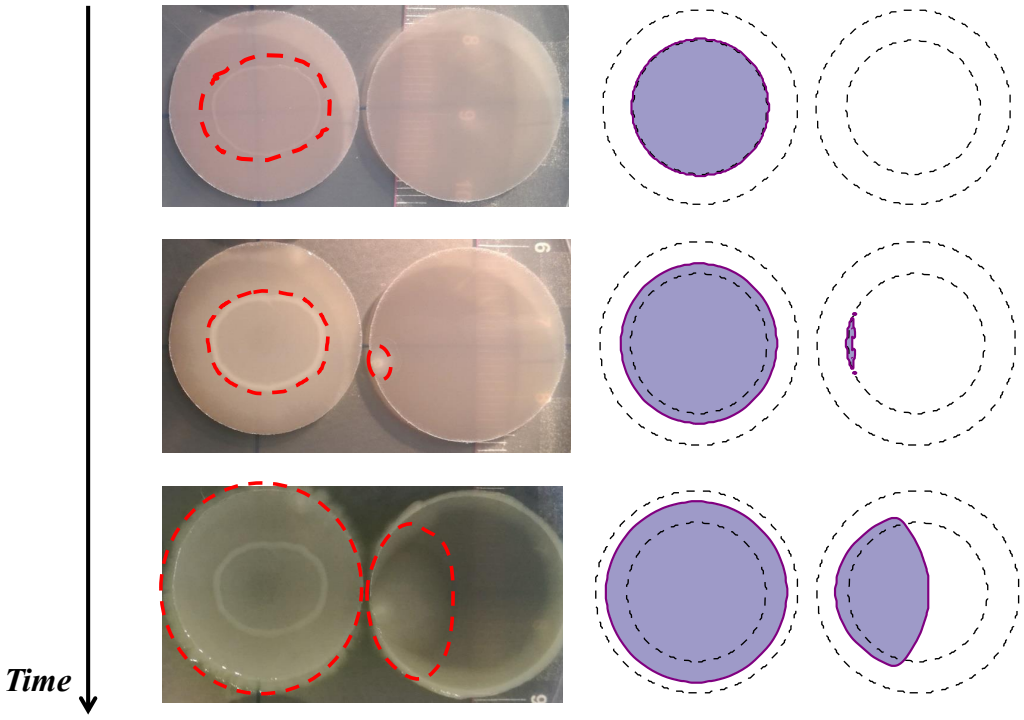


Figure 6.7: Left column: pictures from the inoculated membrane and the host membrane, taken at $t = 24$h, 48h and 72h. The biofilm is contoured by a red dashed line. Right column: three stages of the numerical simulation of the experiment, where the biofilm is marked by the purple bold surface.

framework for constructing new approaches and techniques to manipulate biofilm formation (either discouraging or promoting biofilm development) in environmental, industrial and medical applications. Hence, the ability to model and predict the mechanisms of dispersal would have a great socio-economical impact, with many implications for global health, as well as for the management of environmental microorganisms in biogeochemical cycling processes and biotechnological biofilm-related applications.

6.3 Uncertainty Quantification and Sensitivity Analysis of a multi species biofilm modeling

In this section, an existing model of 1D biofilm that had recently been improved with new equations accounting for species invasion, is studied with UQ and SA procedures. This Section will follow the results of the recent paper [284].

6.3.1 Importance of sensitivity analysis in existing biofilm models

In the context of biofilm modeling, studying the sensitivity of the biofilm model predictions to the variability in the input parameters is a way to better understand the response of the model to an certain choice of parameters and ultimately to highlight new features into the biological processes. In this Section the same notation of Chapter 2 is adopted. As a consequence, for each set of input parameters $\boldsymbol{\theta} = \{\theta_1 \dots \theta_d\}$, the output of the model is codified into a set of quantities of interest $\mathbf{y} = \{y_1 \dots y_n\}$. This leads to the definition of the functional relation $\mathcal{F}$

$$\boldsymbol{\theta} \in \mathbb{R}^d \quad \rightarrow \quad \mathbf{y} = \mathcal{F}(\boldsymbol{\theta}) \in \mathbb{R}^n. \tag{6.9}$$

In the framework of uncertainty quantification [143, 256], the set of input parameters $\boldsymbol{\theta}$ is considered uncertain and the objective is to propagate the input uncertainties through the numerical model and to estimate the subsequent uncertainties in the quantities of interest $\mathbf{y}$. In complement, global sensitivity analysis methods [118, 65] provide valuable ways to characterize the input-output model dependency $\mathcal{F}$: they are helpful to derive a relevant screening of the input parameters, spot unimportant parameters and focus the attention on the most relevant ones. Although the parameters involved in biofilm models may vary considerably and interact with each other to influence in depth the model output, only few attempts have been made in the past years to apply uncertainty quantification [261, 51] and sensitivity analysis to biofilm models at both local and global levels [109, 40, 36, 137, 294, 309, 55]. Most of these studies refer to an application of the original Wanner-Gujer model [304], which nowadays the most widely used biofilm model in engineering applications. This model has been integrated in AQUASIM [230], a numerical simulator designed for simulating aquatic systems and also for performing parameter estimation and sensitivity analysis, see Refs. [40, 137, 294] related to global sensitivity analysis: Ref. [40] presents a comparison between the qualitative Morris screening method and the quantitative variance-based Fourier amplitude sensitivity test for a two-step nitrification biofilm model; Ref. [137] presents variance-based sensitivity analysis applied to a one-dimensional biofilm model for ammonium and nitrite oxidation for varying biofilm reactor geometry; and Ref. [294] calculates sensitivity by performing model output linear regression for a complete autotrophic nitrogen removal biofilm.

Unfortunately, Wanner-Gujer-type biofilm modeling is not detailed enough to study bacterial invasion mechanisms, which occur frequently and play an important role in most of engineering applications. To overcome this modeling limitation, a new class of continuum models for multi-species biofilm formation and growth, which explicitly accounts for invasion mechanisms, has been recently introduced [71, 72]. The novelty in such biofilm modeling class relates to the introduction of a new state variable, which describes the concentration of planktonic species inside of the biofilm. The diffusion of the free cells from the bulk liquid into the biofilm and the other

way around is modeled by a diffusion-reaction equation; the growth processes are governed by a system of nonlinear hyperbolic partial differential equations; and finally, substrate dynamics are described by a system of semi-linear parabolic partial differential equations. All equations are inter-connected so that the resulting system of differential equations corresponds to a *free boundary value problem*, where the free boundary is represented by the biofilm thickness. This model formulation aims at reproducing the colonization of new species diffusing from bulk liquid to biofilm and the development of latent microbial species within the biofilm, without explicitly prescribing boundary conditions for the invading species at the free boundary. These boundary conditions are consistently specified by the model, instead of being set arbitrarily [305].

This new class of continuum models can handle any number of microbial species, both in sessile and planktonic states, as well as dissolved substrates. One difficulty is that this type of model involves parameters related to species invasion that are rather new in the literature and whose reference values are not obvious to specify. To overcome this issue, we present in this study, a variance-based sensitivity analysis approach that makes use of the well known Sobol' indices [257, 243] described thorougly in Chapter 2, to spot the most important parameters related to bacterial invasion mechanisms. These Sobol' indices derived from variance decomposition quantify the contribution of each uncertain parameter to the variance of the quantities of interest. One non-intrusive way to compute them could have been the execution of a Monte Carlo random sample of inputs and simulated outputs [80]. While this approach may be regarded as generic and robust, it is computationally expensive due to a slow rate of convergence with respect to the sample size. Due to the complexity of the biofilm model, this would require at least the order of 10^4–10^5 biofilm model simulations: this solution would of course exceed the available computational budget. An alternative is to derive (or "train") an emulator of the biofilm model using a limited sample of inputs and simulated outputs (or "training set") and taking advantage of the regularity of the model response $\mathcal{F}$. Stated differently, the objective is to fit the emulator (or "surrogate") over a dataset of biofilm model simulations and then to mimic in an accurate and efficient way, the model response $\mathcal{F}$ for any set of parameters $\boldsymbol{\theta}$ without solving the original system of differential equations. Statistical information on the quantities of interest and Sobol' indices can then be computed using the emulator. Emulating can be regarded as a supervised learning procedure and belongs to the field of machine learning [111].

In this study, the objective is to build a surrogate that accurately represents bacterial invasion as described by a recent multi-species biofilm model and use it to perform uncertainty quantification and global sensitivity analysis. In order to provide results that are not algorithm-dependent, we compare two families of popular surrogate models, namely generalized Polynomial Chaos (gPC) [265, 224, 312, 68, 30, 70] and Gaussian Processes (GP) [192, 167, 149, 141, 166]. Comparison of gPC-expansion and GP-model have been reported in the

literature [246, 142, 197, 238]; Ref. [197] highlights that one approach does not systematically outclass the other in terms of surrogate accuracy and computational efficiency, the best surrogate being application-dependent. It is therefore of interest to compare gPC and GP approaches for biofilm applications. The training step of the surrogate requires a sampling of the uncertain input space, with the creation of a Design of Experiment (DoE). The GP approach is known to be more accurate for less structured design than tensor grid when performing sensitivity analysis [293]. Consistently, the sampling is performed here using a low-discrepancy Halton's sequence with a given budget $N = 216$. Due to the nonlinearities of the biological processes involved, we investigate the effect of several choices of the gPC polynomial basis (full or sparse) on the surrogate performance for a fixed sample size N. We recall that adopting a sparse polynomial basis may reduce the size of the stochastic problem by only selecting the most significant basis components, and help to better capture a complex model response to variations in the input parameters [286]. We consider here the least-angle regression (LAR) approach to build a sparse gPC basis [33, 32], which was found to provide the best performance among several sparse methods in Ref. [286]. Again, further details on LAR regression can be found in Chapter 2.

It is worth noting that the biofilm model we analyze belongs to the category of hyperbolic partial differential equations, meaning that the quantities of interest can exhibit sharp variations (and even discontinuities) for a subset of the input stochastic space. Given that, building an accurate surrogate that covers the whole input space when dealing with model nonlinearities results to be quite challenging [224, 68, 30, 217]. One strategy found in literature is to partition the input space, bulding local surrogates and finally combine them into a mixture-of-experts model [144]. It is thus of paramount interest to investigate if building a global surrogate for biofilm applications could give acceptable results, before embarking on more complex settings such as mixture of experts.

In this Section, the target problem is a typical microbial interaction occurring in waste-water treatment plants. Initially, the biofilm is only made of heterotrophic bacteria and latent autotrophic bacteria are present in the bulk liquid; then autotrophic bacteria infiltrate the biofilm, switch their state from planktonic to sessile mode and start to proliferate, where they meet the best environmental conditions for their growth. The gPC and GP surrogates are exploited to quantity the uncertainties in the microbial species volume fractions and analyze their dependency with respect to three parameters related to the autotrophic bacterial invasion (the problem dimension is $d = 3$ in Eq. 2.1). Note that in the literature, global sensitivity analysis and uncertainty quantification mostly deal with scalar outputs, while the biofilm model output here is functional with spatial and temporal discretizations, $n > 1$ in Eq. (2.1). Our approach consists here in building a surrogate at each time step of interest, over the spatial grid associated to the model output [45, 99, 65].

6.3.2 Biofilm model

In the following the recent continuum model [71] is described. We recall that the aim of this model is to forecast in a quantitative and deterministic way, the bacterial invasion in multi-species biofilms [172]. Such model essentially consists of a modified Wanner-Gujer formulation accounting for the dynamics of the invading planktonic species as well as substrate diffusion, attachment, detachment, microbial growth and biomass spreading. This model has been derived in one dimension and subsequently generalized to three dimensions [129]. However, in the present study, we focus only on the one-dimensional model.

6.3.3 Free boundary value problem

The invasion model is formulated as a *free boundary value problem* for the three state variables: (1) the concentration of microbial species in sessile form $X_i(z,t)$, $i = 1,\dots,N_s$, $\mathbf{X} = X_1,\dots,X_{N_s}$; (2) the concentration of planktonic species $\psi_i(z,t)$, $i = 1,\dots,N_s$, $\boldsymbol{\psi} = \psi_1,\dots,\psi_{N_s}$; and (3) the concentration of the dissolved substrates $S_j(z,t)$, $j = 1,\dots,N_m$, $\mathbf{S} = S_1,\dots,S_{N_m}$, including the substrates provided by the bulk liquid and the metabolic waste products related to microbial metabolism. Note that the state variables are functions of time t and space z, with z denoting the one-dimensional spatial coordinate assumed perpendicular to the substratum surface located at $z = 0$. Note also that for generality, both the microbial species in sessile and planktonic states are in number of N_s, although in most of applications N_s denotes the number of all particulate components, such as extracellular polymeric substance, inert material and all the phenotype variants of the microbial species.

In the following, the concentration of the ith microbial species in sessile form $X_i(z,t)$ is defined as:

$$\begin{cases} \frac{\partial X_i}{\partial t}(z,t) + \frac{\partial}{\partial z}\left(u(z,t)X_i(z,t)\right) = \rho_i\, r_{M,i}(z,t,\mathbf{X},\mathbf{S}) + \rho_i\, r_i(z,t,\mathbf{S},\boldsymbol{\psi}), \\ X_i(z,0) = \varphi_i(z),\ t = 0,\ 0 \le z \le L(0). \end{cases} \tag{6.10}$$

Equation (6.10) describes the growth of the ith microbial species constituting the biofilm and derives from mass conservation. Biofilm expansion is driven by biomass accumulation: biomass spreading is modeled as an advective mass flux of each species. The reaction terms $r_{M,i}$ model the growth of sessile cells (which is controlled by the local availability of nutrients, usually described with standard Monod kinetics) and the natural death of cells. The terms r_i stand for the growth rates of the ith microbial species due to colonization, which induces the switch of planktonic cells to a sessile growth mode. This phenotypic change is catalyzed by the formation of specific environmental niches within the biofilm matrix. Please note that Eq. (6.10) can be reformulated in terms of volume fractions

$$f_i = X_i/\rho_i,\ \sum_{i=1}^{N_s} f_i = 1, \tag{6.11}$$

where f_i is the volume fraction at a particular location that is occupied by the ith species, and where ρ_i represents the biomass density for the ith species, usually assumed the same for all microbial species. Note that $\varphi_i(z)$ in Eq. (6.10) denotes the initial distribution of biofilm particulate components at initial time; for invading microbial species, $\varphi_i(z) = 0$. Note also that the advective biomass velocity $u(z,t)$ corresponding to the velocity at which the microbial mass is displaced with respect to the film-support interface is formulated as

$$\begin{cases} \frac{\partial u}{\partial z}(z,t) = \sum_{i=1}^{N_s} \left(r_{M,i}(z,t,\mathbf{X},\mathbf{S}) + r_i(z,t,\mathbf{S},\boldsymbol{\psi}) \right), \\ u(0,t) = 0,\ z = 0,\ t \geq 0. \end{cases} \tag{6.12}$$

$u(z,t)$ is determined by the mean observed specific growth rate of the biomass; it is assumed identical for all considered species. $u(z,t)$ also depends on the specific growth rates related to invasion process. The boundary condition at $z = 0$ is derived from a no-flux condition imposed at the substratum surface.

Moreover, the biofilm extent (or "thickness") varies with time, i.e. $L \equiv L(t)$. Equation (6.13) governs the evolution of the free boundary, which depends on the displacement velocity of microbial biomass as well as on the attachment and detachment fluxes:

$$\begin{cases} \frac{dL}{dt}(t) = u(L(t),t) + \sigma_a(t) - \sigma_d(L(t)),\ t > 0, \\ L(0) = L_0,\ t = 0, \end{cases} \tag{6.13}$$

where L_0 corresponds to the initial biofilm thickness. Equation (6.13) is derived from conservation principles at global scale.

The concentration of the ith planktonic species $\psi_i(z,t)$ follows a diffusion-reaction equation:

$$\begin{cases} \frac{\partial \psi_i}{\partial t}(z,t) - \frac{\partial}{\partial z}\left(D_{M,i} \frac{\partial \psi_i}{\partial z}(z,t) \right) = r_{\psi,i}(z,t,\mathbf{S},\boldsymbol{\psi}), \\ \psi_i(z,0) = \psi_{i,0}(z),\ t = 0,\ 0 \leq z \leq L(0), \\ \frac{\partial \psi_i}{\partial z}(0,t) = 0, z = 0,\ t > 0, \\ \psi_i(L(t),t) = \psi_i^*(t),\ z = L(t),\ t > 0. \end{cases} \tag{6.14}$$

Equation (6.14) governs the movement of planktonic cells within the biofilm matrix. The reaction terms $r_{\Psi,i}$ represent a loss term for invading species when biofilm colonization occurs. $D_{M,i}$ denotes the diffusion coefficient of the ith planktonic species within the biofilm. For all considered microbial species, the initial concentration of planktonic cells within the biofilm is usually set to 0 (implying that invasion occurs at initial time) or using a spatially-distributed specific function $\psi_{i,0}(z)$. Homogeneous Neumann conditions are adopted on the substratum surface at $z = 0$ due to a no-flux condition. Dirichlet boundary conditions are prescribed at the free boundary $z = L(t)$. The functions $\psi_i^*(t)$ represent the concentrations of planktonic cells within the bulk liquid; they can be prescribed or derived from mass conservation within the bulk liquid.

The concentration of the jth dissolved substrate $S_j(z,t)$ is also governed by a reaction-diffusion equation

$$\begin{cases} \frac{\partial S_j}{\partial t}(z,t) - \frac{\partial}{\partial z}\left(D_j \frac{\partial S_j}{\partial z}(z,t)\right) = r_{S,j}(z,t,\mathbf{X},\mathbf{S}), \\ S_j(z,0) = S_{j,0}(z),\ t=0,\ 0 \le z \le L(0), \\ \frac{\partial S_j}{\partial z}(0,t) = 0,\ z=0,\ t>0, \\ S_j(L(t),t) = S_j^*(t),\ t>0, \end{cases} \tag{6.15}$$

where the term $r_{S,j}$ represents the jth substrate production or consumption rate due to microbial metabolism, and where D_j denotes the diffusion coefficient of the jth substrate within the biofilm. The initial concentration of the jth dissolved substrate is prescribed using the function $S_{j,0}(z)$. As for the concentrations of planktonic species $\psi_i(z,t)$, homogeneous Neumann conditions are adopted for $S_j(z,t)$ on the substratum surface at $z=0$ due to a no-flux condition, and Dirichlet boundary conditions $S_j^*(t)$ are prescribed at the free boundary $z=L(t)$.

6.3.4 Autotrophic colonization

In the present study, we consider the following target problem: the biofilm is constituted by three particulate components, heterotrophic bacteria X_1, autotrophic bacteria X_2, and inert material X_3 (X_3 directly results from the decay of the two active microbial species X_1 and X_2).
At initial time, we assume that the biofilm is only composed of heterotrophic bacteria and we enhance autotrophic colonization. We consider heterotrophic-autotrophic competition with oxygen as common substrate as in Ref. [304]. Three dissolved substrates are taken into account: organic carbon S_1, ammonia S_2, and oxygen S_3. Oxygen is used for both organic carbon oxidation and nitrification. Note that the waste products of the metabolic reactions are not explicitly modeled. The establishment and proliferation of X_2 strictly depend on the formation of an environmental niche, where the growth of heterotrophic bacteria X_1 is limited by the low concentration in organic carbon. Planktonic cells ψ_2 are considered for X_2 as the biofilm model is aimed at simulating the invasion of a constituted biofilm by autotrophic bacteria after the establishment of a favorable environmental niche.
The stoichiometry and the process rates needed to close the model equations (Eqs. 6.10–6.15), including the expressions for $r_{M,i}$, $r_{S,j}$, r_i and $r_{\psi,i}$, are taken from Refs. [305, 71].
The biomass growth rates $r_{M,i}$ in Eq. (6.10) are

$$r_{M,1} = \left(\mu_{\max,1}\,\frac{S_1}{K_{1,1}+S_1}\,\frac{S_3}{K_{1,3}+S_3} - k_{d,1}\right) X_1, \tag{6.16}$$

$$r_{M,2} = \left(\mu_{\max,2}\,\frac{S_2}{K_{2,2}+S_2}\,\frac{S_3}{K_{2,3}+S_3} - k_{d,2}\right) X_2, \tag{6.17}$$

$$r_{M,3} = k_{d,1}\,X_1 + k_{d,2}\,X_2, \tag{6.18}$$

where $\mu_{\max,i}$ denotes the maximum net growth rate for the ith biomass, $K_{i,j}$ is the affinity constant of the jth substrate for the ith biomass, and $k_{d,i}$ stands for the decay constant for the ith

biomass. The specific growth rates induced by the switch of the planktonic cells to the sessile mode of growth, also required as inputs to Eq. (6.10), are defined as

$$r_1 = r_3 = 0, \tag{6.19}$$

$$r_2 = k_{col,2}\,\frac{\psi_2}{k_{\psi,2}+\psi_2}\,\frac{S_2}{K_{2,2}+S_2}\,\frac{S_3}{K_{2,3}+S_3}, \tag{6.20}$$

where $k_{col,2}$ corresponds to the maximum colonization rate of autotrophic bacteria, and where $k_{\psi,2}$ corresponds to the affinity-type constant for ψ_2.

The conversion rates for the three substrates required as inputs to Eq. (6.15) can be formulated as

$$r_{S,1} = -\frac{1}{Y_1}\,\mu_{\max,1}\,\frac{S_1}{K_{1,1}+S_1}\,\frac{S_3}{K_{1,3}+S_3}\,X_1, \tag{6.21}$$

$$r_{S,2} = -\frac{1}{Y_2}\,\mu_{\max,2}\,\frac{S_2}{K_{2,2}+S_2}\,\frac{S_3}{K_{2,3}+S_3}\,X_2, \tag{6.22}$$

$$\begin{aligned} r_{S,3} = &-\frac{1-Y_1}{Y_1}\,\mu_{\max,1}\,\frac{S_1}{K_{1,1}+S_1}\,\frac{S_3}{K_{1,3}+S_3}\,X_1 \\ &-\frac{4.57-Y_2}{Y_2}\,\mu_{\max,2}\,\frac{S_2}{K_{2,2}+S_2}\,\frac{S_3}{K_{2,3}+S_3}\,X_2, \end{aligned} \tag{6.23}$$

with Y_i denoting the yield of biomass i.

The conversion rate of the planktonic cells associated with the ith species, required as input to Eq. (6.14), is given by

$$r_{\psi,i} = -\frac{1}{Y_{\psi,i}}\,r_i, \tag{6.24}$$

where $Y_{\psi,i}$ is the yield of sessile species on planktonic ones. The terms $r_{\psi,i}$ describe the consumption rates of planktonic cells due to invasion process. $r_{\psi,i}$ are assumed proportional to r_i, meaning that they are modeled using the same Monod kinetics.

6.3.5 Simulation settings

To numerically solve the free boundary problem presented in Section 6.3.3 and Section 6.3.4, we use a straightforward upgrade of the numerical method proposed in Ref. [73]. The method of characteristics is used to track the biofilm expansion. Finite difference method is then adopted to solve the diffusion-reaction equations. We extend this method to account for the new independent variables $\{\psi_i\}$, which account for invasion processes and which follow Eq. (6.14); $\{\psi_i\}$ are treated similarly as the variables $\{S_j\}$ characterizing dissolved substrates in Eq. (6.15). The solver is implemented in MATLAB.

In the present application, simulations are run for the target simulation time $T = 15$ days. The initial and boundary conditions related to the free boundary problem are listed in Table 6.1.

Table 6.1: Initial-boundary conditions for biofilm growth, Eqs. (6.10)–(6.15).

Variable	Symbol	Value	Unit
Initial volume fraction of f_1	$\varphi_1(z)$	0.0	–
Initial volume Fraction of f_2	$\varphi_2(z)$	1.0	–
Initial volume Fraction of f_3	$\varphi_3(z)$	0.0	–
Bulk liquid concentration of S_1	S_1^*	3	$g_{COD}m^{-3}$
Bulk liquid concentration of S_2	S_2^*	13	$g_N m^{-3}$
Bulk liquid concentration of S_3	S_3^*	5	$g_{O_2}m^{-3}$
Bulk liquid concentration of ψ_1	ψ_1^*	0.0	$g_{COD}m^{-3}$
Bulk liquid concentration of ψ_2	ψ_2^*	1.0	$g_{COD}m^{-3}$
Initial biofilm thickness	L_0	300	μm
Initial concentration of S_1	$S_1(z,0)$	0.0	$g_{COD}m^{-3}$
Initial concentration of S_2	$S_2(z,0)$	0.0	$g_N m^{-3}$
Initial concentration of S_3	$S_3(z,0)$	0.0	$g_{O_2}m^{-3}$
Initial concentration of ψ_1	$\psi_1(z,0)$	0.0	$g_{COD}m^{-3}$
Initial concentration of ψ_2	$\psi_2(z,0)$	0.0	$g_{COD}m^{-3}$

6.3.6 Sources of uncertainty, quantities of interest and experimental designs

6.3.7 Functional output

The state of the biofilm evolves in time $t \in [0,T]$ and space $z \in [0, L(t)]$. The biofilm is characterized by biomass volume fractions, $f_i, i \in \{1,\dots,N_s\}$, and substrates $S_j, j \in \{1,\dots,N_m\}$, with $N_s = 3$ and $N_m = 3$ (see Section 4.1). Since the objective here is to analyze invasion mechanisms, we focus our attention on the species volume fractions f_i defined in Eq. (6.11).
The quantities of interest could be in principle formulated as

$$y_i(t) = \frac{\int_0^{L(t)} f_i \, dz}{L(t)}, \quad i \in \{1,\dots,N_s\}. \tag{6.25}$$

However, this choice would not express the spatial variability of the biofilm properties and would lead to an analysis of the different species as if the biofilm were concentrated in a single point. To circumvent this problem, the following discretization of the biofilm is proposed:

$$y_{ijk} = f_i(z_j, t_k), \quad i \in \{1,\dots,N_s\}, \tag{6.26}$$

where the spatial discretization is given by $z_j = j\,\Delta z$, $\Delta z = L(t)/N_z$ and $j \in \{0,\dots,N_z-1\}$, and where the time discretization is given by $t_k = k\,\Delta t$, $\Delta t = T/N_t$ and $k \in \{0,\dots,N_t-1\}$.
In particular, we consider $N_t = 4$ times at which the biofilm extension is discretized into $N_z = 5$ locations. Note that the inert volume fraction f_3 is retrieved by mass conservation (Eq. 6.11). Hence, the model output $\mathbf{y}$ is of functional type and includes the elements y_{ijk} with $i = \{1,2\}$; $j = 1,\dots,5$; and $k = 1,\dots 4$ ($\mathbf{y} \in \mathbb{R}^n$ with $n = 40$) in the present study. This functional output is referred to as the "quantities of interest".

Note that the quantities of interest are considered as Lagrangian markers assigned to a relative position of the biofilm, whose spatial extent $L \equiv L(t)$ depends on time and on the biofilm model parameters (see Section 6.3.8).

6.3.8 Sources of uncertainty

In biological applications, a major source of uncertainty resides in the parameters associated with species or substrates. In the present modeling approach, parameters such as $\mu_{\max,i}$, $k_{d,i}$, $K_{i,j}$ and Y_i ($i = 1, \ldots, N_s$, $j = 1 \ldots N_m$) are well characterized in Ref. [304] and are therefore assigned to reference values. We thus shift our attention to the parameters related to autotrophic bacteria biofilm invasion: $k_{col,2}$ and $k_{\psi,2}$ involved in r_2 in Eq. (6.20) to model the growth rate of autotrophic bacteria in sessile mode on the one hand, and $Y_{\psi,2}$ involved in Eq. (6.24) to model the consumption rate of planktonic cells denoted by $r_{\psi,2}$ on the other hand. Hereafter, $k_{col,2}$, $k_{\psi,2}$ and $Y_{\psi,2}$ are respectively denoted by k_{col}, k_ψ and Y_ψ for clarity purposes. The uncertain input vector $\boldsymbol{\theta}$ is thus defined as

$$\boldsymbol{\theta} = (k_{\text{col}}, k_\psi, Y_\psi) \in \mathbb{R}^3. \tag{6.27}$$

such that the problem dimension is $d = 3$, see Table 6.2.

We recall that these parameters are not well characterized in literature and their determination still requires an accurate experimental activity based on ad-hoc techniques. In this work, we consider stochastic methods to represent input uncertainty. Thus, according to the theory of Chapter 2, the uncertain input parameters are modeled by a random vector $\boldsymbol{\Theta}$, meaning that their values are supposed to depend on a random parameter ω such that $\boldsymbol{\Theta} \equiv \boldsymbol{\Theta}(\omega)$. ω is to be taken from the set of all outcomes Ω, which is equipped with a $\sigma-$algebra $\mathcal{S}$ and a probability measure $\mathcal{P}$. The triplet $(\Omega, \mathcal{S}, \mathbb{P})$ forms a probabilistic space [51].

The functional output $\mathbf{y}$ is considered as an element of $L^2(\Omega, \mathcal{S}, \mathbb{P})$ and is therefore represented as a vector of stochastic process, i.e.

$$\mathbf{Y}(\omega) = \mathcal{F}\left(\boldsymbol{\Theta}(\omega)\right), \tag{6.28}$$

with $\mathcal{F}$ the mapping of the input parameters onto the space of the functional output given by the biofilm model (see Eq. 2.1).

Table 6.2: Uniform marginal PDF associated with k_{col}, k_ψ and Y_ψ. Note that $\mathcal{U}(a, b)$ stands for the uniform distribution with a the minimum value of the parameter and b the maximum one.

Parameter	Uniform distribution
k_{col}	$\mathcal{U}(10^{-4}, 10^{-2})$
k_ψ	$\mathcal{U}(10^{-5}, 10^{-2})$
Y_ψ	$\mathcal{U}(10^{-5}, 10^{-3})$

Stochastic methods require to characterize the probability density function (PDF) associated with the input random vector $\boldsymbol{\Theta}$ denoted by $\boldsymbol{\rho_\Theta}$. We need to introduce some assumptions on the nature of such uncertainty sources. First, we assume the components of $\boldsymbol{\Theta}$ are independent. Second, we consider uniform marginal PDF for each random variable Θ_i $(i = 1, \ldots, d)$ in $\boldsymbol{\Theta}$, denoted by ρ_{Θ_i}. The following restrictions apply: $k_{\text{col}} > 0$, $k_\psi > 0$ and $Y_\psi \in [0; 1]$; see Table 6.2. The objective here is to analyze under uncertainty, the relation between inputs $\boldsymbol{\Theta}$ and outputs $\mathbf{Y}$ and to build an emulator of the relation $\mathcal{F}$ in Eq. (6.28).

6.3.9 Experimental designs, databases, adopted surrogate mdoels

It is recalled that a design of experiments refers to the way of discretizing the uncertainty space (or "hypercube") $Z_\Theta \in \mathbb{R}^d$ $(d = 3)$, in which the three parameters k_{col}, k_ψ and Y_ψ evolve. It is a way to define the N realizations of parameters $\boldsymbol{\theta}$, for which the biofilm model is integrated as a "black-box" to obtain the ensemble of N functional outputs $\mathbf{y}$ from which statistics can be derived. This ensemble forms a database $\mathcal{D}_N$:

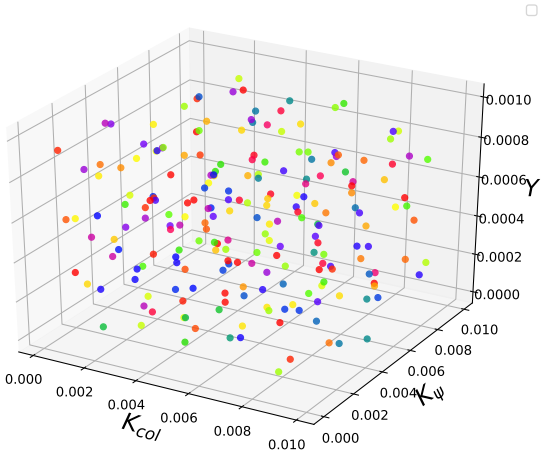
$$\mathcal{D}_N = \left\{ \left(\boldsymbol{\theta}^{(l)}, \mathbf{y}^{(l)} \right)_{1 \leq l \leq N} \right\}, \tag{6.29}$$

where $\mathbf{y}^{(l)} = \mathcal{F}\left(\boldsymbol{\theta}^{(l)}\right)$ stands for the integration of the biofilm model $\mathcal{F}$ associated with the lth set of input parameters $\boldsymbol{\theta}^{(l)}$.

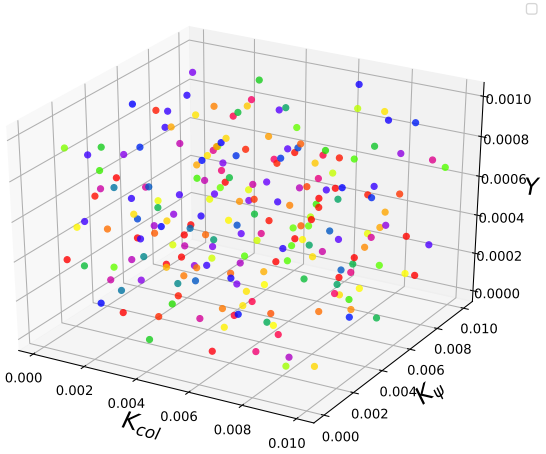
In the present study, two databases of size $N = 216$ are compiled using quasi-Monte Carlo sampling methods. They rely on low-discrepancy sequences to explore the hyperspace given by the support of the three PDFs without any bias and to capture most of the variance [64]. The first database built using Halton's sampling serves as a training set and corresponds to the ensemble of simulations over which the surrogates are trained (Fig. 6.8a). The second one is built using Faure's sampling. It serves as a validation set and corresponds to the ensemble of simulations that is not part of the experimental design and that is used to evaluate the surrogate accuracy (Fig. 6.8b). Note that the biofilm model is characterized by high nonlinearities. Figure 6.9 presents 10 representative biofilm model snapshots at different times, (a) 5 days, (b) 10 days and (c) 15 days. The spatial distribution of the heterotrophic bacterial volume fraction f_1 is represented for each time, each line corresponds to a different realization of input parameters $\boldsymbol{\theta} = (k_{\text{col}}, k_\psi, Y_\psi)$ that is a point of the Halton's low-discrepancy sequence presented in Fig. 6.8a and each line is colored with respect to the autotrophic bacterial volume fraction f_2. The biofilm length $L(t)$ effectively varies with time from 0.0010 to 0.0016 m.

Adopted surrogate models

In this work we built an emulator of the biofilm model, using generalized Polynomial Chaos (gPC) expansion or Gaussian Process model. The common idea of both approaches is to design for each

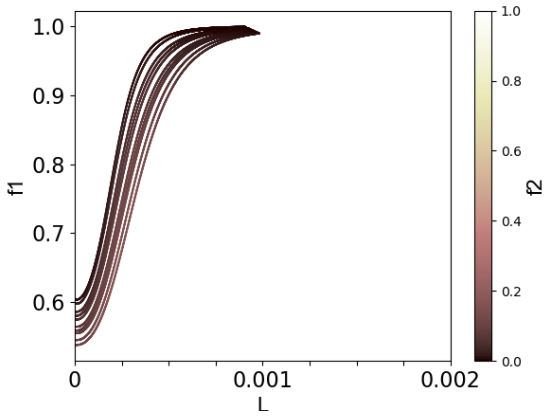


(a) Halton's sampling

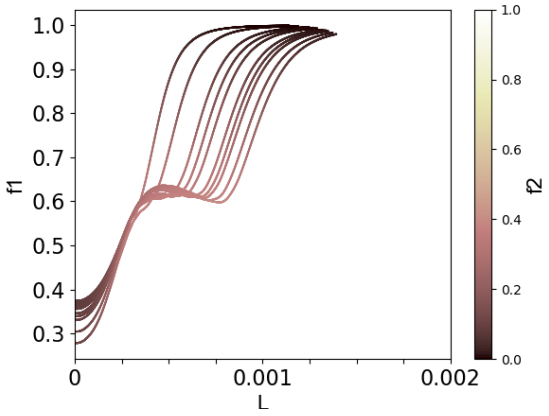


(b) Faure's sampling

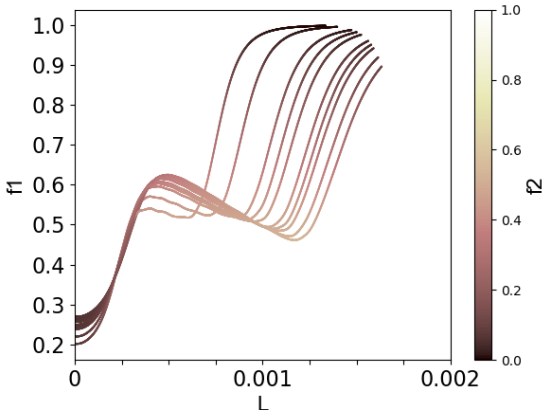
Figure 6.8: Cloud representation of the two databases $\mathcal{D}_N$ with $N = 216$, corresponding to different sets of the three parameters k_{col} (x-axis), k_ψ (y-axis) and Y_ψ (z-axis). The two databases correspond to low-discrepancy sequences, (a) Halton's sampling (training set) and (b) Faure's sampling (validation set).



(a) Time $t = 5$ days



(b) Time $t = 10$ days



(c) Time $t = 15$ days

Figure 6.9: Time-evolving species volume fractions f_1 and f_2 for varying uncertain input vector $\boldsymbol{\theta} = (k_{\text{col}}, k_\psi, Y_\psi)$ (Eq. 6.27). The x-axis corresponds to the biofilm thickness $L(t)$; the y-axis corresponds to f_1; and the colormap corresponds to f_2. The simulated physical time is (a) 5 days, (b) 10 days and (c) 15 days.

quantity of interest Y in the vector $\mathbf{Y}$ ($Y \equiv Y_{ijk}$) a surrogate by means of a weighted (finite) sum of basis functions:

$$Y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \gamma_{\boldsymbol{\alpha}} \, \Psi_{\boldsymbol{\alpha}}, \tag{6.30}$$

The coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ and the basis functions $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ are calibrated with the information contained in the Halton's training set $\mathcal{D}_N$ with $N = 216$ (see Section 6.3.9).
In practice, the three adopted surrogate that are analyzed in the following sections are:

- generalized Polynomial Chaos (gPC) with linear projection strategy (SLS)
- generalized Polynomial Chaos (gPC) with LAR sparse projection strategy (LAR);
- Gaussian Process with RBF kernel.

In practice, the computation the gPC-expansion and GP-model used *OpenTURNS* [25] Python package (see www.openturns.org); *batman* [239] was used to build Halton's and Faure's datasets.

6.3.10 Results

A posteriori error estimation of the surrogate models

The surrogates are ranked by their accuracy by the means of two error estimator introduced in Section 2.4.5 of Chapter 2: Empirical Error ϵ_{emp} and Q_2 coefficient.
Since in the following, for any tested configuration, we have $\epsilon_{\text{emp}} < 10^{-4}$, we shall focus on the Q_2 coefficient alone. Moreover, it is remarked that GP surrogate has automatically $\epsilon_{\text{emp}} = 0$ because it is an interpolator of the DoE points. We validated thus the surrogates using the Q_2 predictive coefficient that corresponds to a cross-validation error metric using the independent dataset based on Faure's low discrepancy sequence:

$$Q_2 = 1 - \frac{\sum_{l=1}^{N_{\text{faure}}} \left(y^{(l)} - \widehat{y}^{(l)} \right)^2}{\sum_{l=1}^{N_{\text{faure}}} \left(y^{(l)} - \overline{y} \right)^2}, \tag{6.31}$$

with $\overline{y}$ the empirical mean over the Faure's validation set ($N_{\text{faure}} = 216$). We remind that the target value for Q_2 is 1.
Figures 6.10–6.11 present the Q_2 predictive coefficient along the biofilm after 5 days, 10 days and 15 days for the three different surrogates: SLS-based gPC-expansion (black-star line); LAR-based gPC-expansion (red-dotted line); and RBF-based GP-model (blue-squarred line). Figure 6.10 is obtained when considering the species volume fraction f_1 – heterotrophic bacteria – as observable; Fig. 6.11 is the counterpart of Fig. 6.10 for f_2 – autotrophic bacteria. Results show that the LAR gPC-expansion gives the best performance with a Q_2 close to 1 over the whole time period and all along the biofilm thickness. The SLS gPC-expansion exhibits to significant error after 10 days and

15 days, when the biological processes at play become more complex. The minimum value for Q_2 moves along the biofilm over time, with Q_2 going down to 0.6 at $z \approx L/4$ after 10 days and 0.82 at $z = 2/4L$ after 15 days. The GP-model achieves intermediate performances between LAR-based gPC-expansion and SLS-based gPC-expansion; the corresponding Q_2 being at minimum equal to 0.9 when it reaches 0.6 for SLS-based gPC-expansion after 10 days. After 15 days both LAR-based gPC-expansion and GP-model share similar performance.

Figure 6.12 presents the polynomial terms that are retained in the LAR gPC-expansion built to emulate the species volume fraction f_1 at a particular location of the biofilm ($z = L(t)/4$); time evolution of these polynomial terms is presented (after 5, 10, 15 days). Note that we consider the case $z = L(t)/4$ since the LAR gPC-surrogate tends to outperform the SLS gPC-surrogate and the GP model at this location (see Fig. 6.10). Each active polynomial $\Psi_{\boldsymbol{\alpha}}$ is associated with a colored symbol, where the color represents the magnitude of the coefficient $\gamma_{\boldsymbol{\alpha}}$. The x-/y-/z-axis of the plots represent the degree of the polynomial. It is evident that in this case LAR offers some flexibility (due to the sparse structure of the polynomial basis) to integrate high-order polynomial terms in the gPC-expansion, in particular along the direction associated with the parameter k_{col} (x-axis), where polynomial degrees go up to 16 after 10 days. The full basis considered in the SLS gPC-surrogate cannot include these terms due to the limited size of the training set ($N = 216$, implying that $P \leq 5$). The increase in complexity of the biofilm structure with respect to time is highlighted by the increasing number of terms retained in the gPC-expansion over time.

To sum up, the sparse truncation strategy underlying the LAR-based gPC-expansion seems to provide a clear advantage to build an emulator of the biofilm model. The magnitude and number of LAR gPC-coefficients give insight into the complexity of the biological processes occurring in multi-species biofilm; this complexity growing over time. The latter can only be captured by a flexible adaptative surrogate approach that identifies inline the required polynomial degree to accurately capture the system dynamics. The following analysis is therefore carried out using the standalone LAR approach.

6.3.11 Uncertainty quantification of the biofilm model predictions

Using the LAR gPC-expansion, the statistics of each quantity of interest y can be derived analytically from the coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathcal{A}}$. The mean value μ_y and STD σ_y of y can be estimated as

$$\mu_y = \gamma_0, \tag{6.32}$$

$$\sigma_y = \sqrt{\sum_{\substack{\boldsymbol{\alpha}\in\mathcal{A}\subset\mathbb{N}^d \\ \boldsymbol{\alpha}\neq 0}} \gamma_{\boldsymbol{\alpha}}^2}. \tag{6.33}$$

The PDF of each quantity of interest is retrieved through kernel smoothing techniques by sampling the uncertain input space Z_Θ using 10,000 members based on Monte Carlo random sampling and by evaluating the LAR gPC-expansion for all these points.

Figure 6.13 presents the PDF of the species volume fractions f_1 and f_2 with respect to the biofilm thickness $L(t)$, along with the mean (solid line) and STD (dashed lines); each panel from left to right corresponds to a different time step over the 15-day time period under consideration. Results show that the uncertainty on the model output is driven rightwards as the simulation runs forward in time: after 5 days the largest variance is observed near $z = L(t)/4$ and moves to $z = 3/4\, L(t)$ after 15 days. The same trend is observed for both species volume fractions f_1 and f_2.

The fact that the central part of the biofilm is subject to the highest level of uncertainty can be interpreted as the increase in complexity of the biofilm structure, which is correlated to the establishment of the invading species, is essentially due to the niche formation occurring far from the biofilm boundaries (substratum surface on the left and bulk liquid on the right). Recall that the adopted boundary conditions refer to a fixed bulk liquid concentration at $z = L(t)$ as well as a no-flux condition at $z = 0$ (see Table 6.1). Figure 7 shows the trends for the three substrates $S_j (j = 1, \ldots, 3)$ over time; the organic carbon S_1 and the oxygen S_3 feature a signifcantly reduced spread at the bottom of the biofilm, independently of the choice of the input vector $\boldsymbol{\theta}$. This is due to a combined effect of substrate diffusion and microbial metabolism which leads to the decrease of substrate concentration with respect to the constant value prescribed at the bulk liquid inteface. More specifically, S_1 is mainly consumed in the outermost part of the biofilm and tends to become zero in the central part of the biofilm where the invading species finds favourable environmental conditions for its growth. Moreover, S_3 is completely depleted in the outer part of the biofilm and thus the microbial complexity due to the invasion process is significantly reduced at the bottom of the biofilm. Note that all the results have been obtained for the specific case study, reproducing a typical microbial interaction occurring in wastewater treatment plants, which is of relevant interest for engineering applications. Diverse boundary conditions may lead to a different result in terms of invasion phenomenon and thus in uncertainty quantification.

It is worth mentioning that some PDFs associated with f_1 and f_2 have more than one mode, see for instance Fig. 6.15 corresponding to the PDF of the autotrophic species volume fraction f_2 at $z = L/4$ after 10 days. This bimodal PDF has a physical explanation: for the given range of the input parameters under consideration, the autotrophic invasion at some location features two distinct behaviors, either a successful or unsuccessful niche formation.

6.3.12 Analysis of the biofilm structure

Using the Halton's training set, we can compute the covariance matrix $\mathbf{C}_{\mathrm{yy}} \in \mathbb{R}^{N_z \times N_z}$, also known as dispersion matrix, to characterize the covariance between the model state $\mathbf{y}$ at different

locations $z \in [0, L(t)]$ at a given time. $\mathbf{C}_{\text{yy}}$ can be empirically estimated as

$$\mathbf{C}_{\text{yy}} = \sum_{l=1}^{N} \frac{\left(\mathbf{y}_{ik}^{(l)} - \overline{\mathbf{y}}_{ik}\right)\left(\mathbf{y}_{ik}^{(l)} - \overline{\mathbf{y}}_{ik}\right)^{T}}{N-1}, \tag{6.34}$$

where $\mathbf{y}_{ik}^{(l)} = \{y_{ijk}^{(l)}\}_{j=1,\cdots,N_z}$ is the vector containing the ith quantity of interest y_{ijk} at a given time index k for the ensemble member l. In this matrix, the diagonal terms correspond to the variance of the model state variable at a given location j. The off-diagonal terms represent the covariances in the model state variable between two locations along the z-axis. The covariance matrix is symmetric by definition. By normalizing the covariance matrix by the variance, we can derive the correlation matrix shown in Fig. 6.16 (by definition diagonal terms are equal to 1). One column of the correlation matrix therefore provides the correlation function of a particular point with the rest of the z-axis.

Figure 6.16 presents the evolution of the correlation matrix over the 15-day time period for both f_1 and f_2 state variables. Results show that at early times (after 5 days), the biofilm can be considered as a single entity with respect to its internal structure since the correlation factor is very high (above 0.99 for both f_1 and f_2). At later times, the internal structure becomes more complex and decorrelates. This evolution is due to the growth in spatial complexity of the biofilm, with the mechanism of autotrophic invasion that alters the species composition of the biofilm in a non-linear way via species niche formation. This is inline with the complex structure of the LAR polynomial basis presented in Fig. 6.12, which includes for instance high-order polynomial terms in the three directions k_{col}, k_ψ and Y_ψ.

In summary, the spatial structure of the biofilm after 10 days seems to be organized as two main clusters: one related to the scarcity in substrates at $z = 0$ (the blue cluster at the bottom-left corner of the correlation matrix in Fig. 6.16), a second one related to the fixed bulk concentration of substrates at $z = L(t)$ (the blue cluster at the top-right of the correlation matrix in Fig. 6.16).

6.3.13 Input-output sensitivity analysis

Sobol' indices [257, 243] are widely used for global sensitivity analysis based on variance decomposition. They are described in Section 2.3.2 of Chapter 2.

It is here briefly recalled that the first-order Sobol' index S_i associated with the ith parameter of $\mathbf{\Theta}$ is given by

$$S_i = \frac{\mathbb{V}_i(Y)}{\mathbb{V}(Y)}, \tag{6.35}$$

and corresponds to the ratio of the output variance $\mathbb{V}(Y)$ that is due uniquely to the ith input parameter; S_i ranges between 0 and 1. The corresponding total Sobol' index S_{T_i} measures the whole contribution of the ith input parameter (including interaction with other parameters of $\mathbf{\Theta}$)

on the output variance. By definition, $S_{T_i} \geq S_i$. If both first-order and total indices are not equal, this indicates that the input parameter Θ_i has some interactions with other parameters of $\boldsymbol{\Theta}$ to explain the variance of the output QoI.

Figure 6.17 shows the first-order and total Sobol' indices obtained with the LAR gPC-expansion related to the autotrophic bacteria volume fraction f_2. These indices are presented at different times $t \in \{5, 10, 15 \text{ days}\}$ (from left to right panels), and at different locations along the biofilm thickness $z \in \{0, L/2, L\}$ (from top to bottom panels).

Results clearly express the prevalence of the input parameter k_{col} with Sobol' indices close to 1 for all times and locations. From a physical viewpoint, k_{col} is therefore a key parameter to represent colonization by autotrophic species X_2 at the expense of heterotrophic species X_1.

k_{col} reproduces in fact the attitude of the microrganisms to switch their state from planktonic to sessile. That is, k_{col} represents the key parameter for the invasion phenomenon to occurr, so changes in Y_ψ and k_ψ have a negligible effect on the overall invasion process. This is due to the fact that the concomitant presence of planktonic species and of specific environmental niches allows the invasion phenomenon to occur only when the planktonic species are characterized by significant values of the colonization rate for the investigated simulation times. This gives us precious information about which measurements may be improved, to use the invasion modeling to improve our understanding of the colonization process overall.

This is inline with the high-order terms contained in the LAR polynomial basis in the direction of k_{col} (see Fig. 6.12). The total polynomial order of the sparse gPC-expansion is due to k_{col}: k_{col} exhibits polynomial terms of degrees up to $P = 16$ after 10 days and $P = 14$ after 15 days.

Note that similar sensitivity is observed along the biofilm thickness after 5 days (first column of panels in Fig. 6.17), which is consistent with the uniform correlation matrices obtained at the same time in Fig. 6.16 and the subsequent interpretation: the biofilm can be considered as a single entity at early times.

In complement, the sensitivity of the model output to the parameters k_ψ and Y_ψ is slightly higher after 15 days than after 5 days ($10^{-2}/10^{-3}$), in particular in the first portion of the biofilm ($z \geq L/2$). These results are also consistent with the two clusters observed in the correlation matrices after 15 days in Figure 6.16. The biofilm is gaining in spatial complexity as time advances: more parameters with respect to the standalone k_{col} could act on the spatial distribution of the invading species. Results show that the input parameter k_ψ is usually more influential than Y_ψ, especially at $z = L$ (third row of panels in Fig. 6.17), even though the relevance of these parameters is of several orders of magnitude below that of k_{col} (about 10^{-4}).

First-order and total Sobol' indices are not identical, implying that some interactions occur between the two parameters k_ψ and Y_ψ. Recall that k_ψ is related to the colonization rate of autotrophic bacteria and that Y_ψ is related to the consumption rate of autotrophic planktonic

cells. These two processes are thus interrelated, the consumption of planktonic cells being affected by the switch of planktonic cells to a sessile mode of growth (see Eq. 6.14).
It is worth noting that at location $z = L$, we obtain nearly constant Sobol' indices over time. This may be caused by the constant boundary conditions imposed at the bulk liquid interface. In contrast, in the central part of the biofilm (second row of panels in Fig. 6.17 corresponding to $z = L/2$), where the niche formation occurs, the sensitivity of the model output to Y_ψ becomes higher than that of k_ψ for long times of biofilm evolution.

6.3.14 Conclusions

In this work, uncertainty quantification and global sensitivity analysis non-intrusive methods were applied to a novel and promising multi-species microbial biofilm model, which explicitly accounts for bacterial invasion processes. Invasion can rapidly alter biofilm populations and could even result in the loss of the resident species. It is therefore a key biological process that requires deeper understanding to improve engineering design. For instance, the continuum biofilm model could be helpful to predict the optimal operational conditions (dilution rates, oxygen concentration, carbon addition, etc.), which favor the establishment of a specific microbial syntrophy between resident and invading species.
The simulation of these biological processes is directly affected by the choice of the biofilm boundary conditions as well as by the range of variation of the input parameters, in particular those related to the planktonic species. The present study focused on the sensitivity of the autotrophic and heterotrophic bacteria volume fractions to the parameters characterizing the colonization rate of autotrophic bacteria and the consumption rate of planktonic cells, i.e. $\boldsymbol{\theta} = (k_{col,2}, k_{\psi,2}, Y_{\psi,2}) \in \mathbb{R}^3$. This sensitivity has been measured here through the computation of spatial and temporal Sobol' indices using a cost-effective surrogate.
It is worth mentioning that Sobol' indices measure the relative contribution of a given parameter on the output variance among the perturbed parameters and of its possible interactions with other parameters. The sensitivity analysis results therefore depend on the choice of $\boldsymbol{\theta}$. The biofilm model may depend on a rather large set of parameters, even on those that were fixed to nominal values in this work. For this reason, the output variance obtained here is necessarily a fraction of the potential variance that could be measured for a fully randomized model.
We presented a detailed analysis of the surrogate performance for a given simulation budget N. Two families of surrogates, gPC-expansion and GP-model, were compared in terms of Q_2 predictive coefficient. One difficulty in building surrogates is the choice of the basis. In particular, for gPC-expansion, the choice of the total polynomial order P and of the basis components (full basis with all elements of degree less or equal to P, or sparse basis) is an essential step to insure the surrogate accurately represents the model response over the whole input parameter space. In

the present test case, the LAR gPC-expansion was found to be the best emulator of the biofilm model over the different time snapshots and biofilm locations, the sparse basis providing more flexibility on the total polynomial order for each input parameter than the full basis. The sparse basis is then an asset to fit the nonlinear biological processes with a limited training set. A single global surrogate was enough to achieve the target Q_2 criterion for the LAR gPC-expansion.
This investigation carried out via the LAR gPC-expansion provided new insights into the biofilm invasion mechanisms.
First, the spatial correlation functions along the biofilm thickness highlighted the temporal changes in the biofilm structure: the young biofilm (after a few days) featured some homogeneity in its spatial structure but the mature biofilm (after ten-to-fifteen days of growth) lost spatial correlation due to the increase in complexity of the biological processes involving niche formation and ongoing resident/invading species competition.
In complement, Sobol' sensitivity indices highlighted the key role of $k_{col,2}$, which represents the maximum colonization rate of autotrophic bacteria and which outclasses by several orders of magnitude the contribution of $k_{\psi,2}$ (affinity-type constant for planktonic species associated with autotrophic bacteria) and $Y_{\psi,2}$ (yield of sessile species on planktonic ones for autotrophic bacteria). This prevalence of $k_{col,2}$ is not only related to its key role in regulating the switch from planktonic to sessile modes of growth, but also to the specific setting of the case study. A relative increase in the relevance of $(k_{\psi,2}, Y_{\psi,2})$ was noticed as biofilm increased in complexity over time.
Finally, the PDF and statistics of the biofilm state provided an interesting viewpoint on the biofilm structure and its temporal evolution. While the mean values retrieved autotrophic invasion trends already documented in Ref. [71], the present study found that the invading and resident species concentrated both their variance in the central part of the biofilm, far from the free boundary, where restrictive conditions on substrates have been imposed, and far from the inert surface, where lack of substrates limited the variability. The variance trends showed for both heterotrophic and autotrophic species, a shift in the location of the maximum spread towards the free boundary $L \equiv L(t)$ for increasing time t.
Uncertainty and global sensitivity analysis is found to be a promising way to identify the most influential parameters in any given regime or application scenario and to quantify their effects on the biofilm structure and evolution. More generally, this provides guidelines to orient further biofilm model developments and design in the long-term prediction capability that could answer some of the medical, environmental and industrial issues related to bacterial invasion. Further work might be related to the extension of the present analysis to more complex biological situations, which are related to the dispersal phenomenon and involve the modeling of planktonic species dynamics in multi-species biofilm.

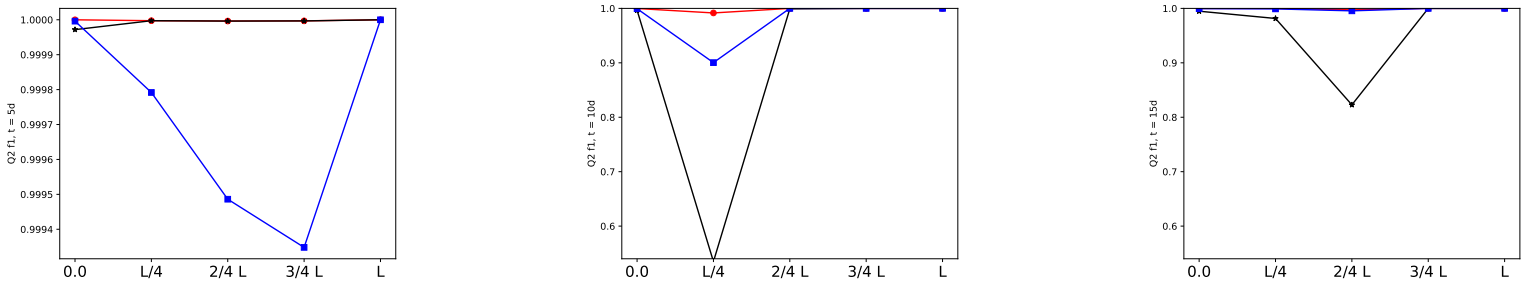


Figure 6.10: Q_2 predictive coefficient along the biofilm thickness $L \equiv L(t)$ at three different time steps: 5 days, 10 days and 15 days (from left to right panels); Halton's experimental design is used as the training set with $N = 216$. Comparison of SLS-based gPC-expansion (black star line), LAR-based gPC-expansion (red dotted line), and RBF-based GP-model (blue squarred line) for the species volume fraction f_1 associated with heterotrophic bacteria.

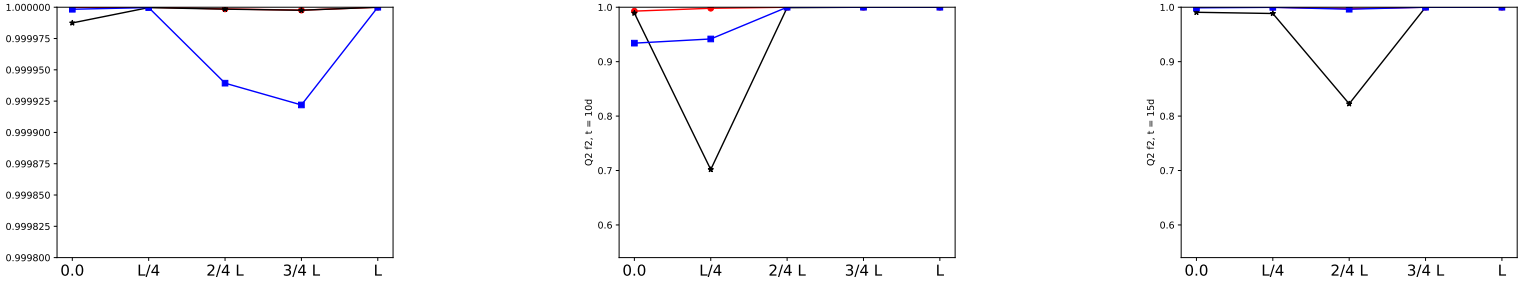


Figure 6.11: Similar caption as Fig. 6.10 but for the species volume fraction f_2 associated with autotrophic bacteria.

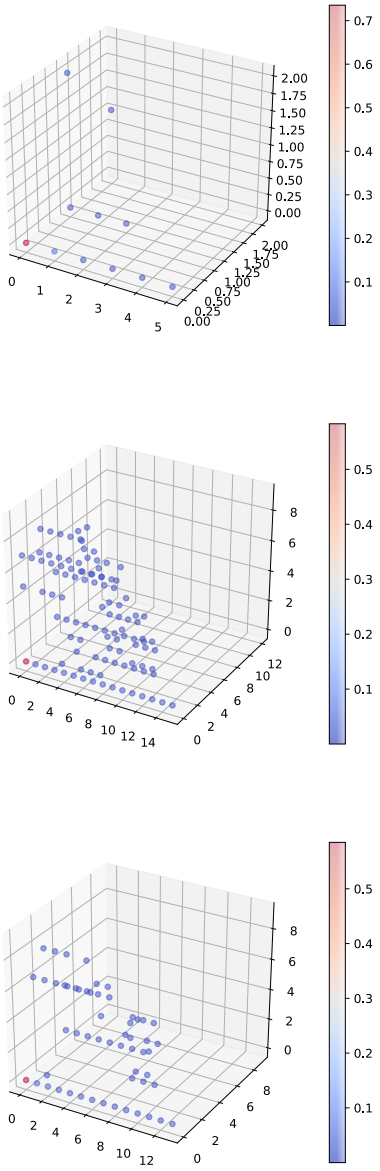
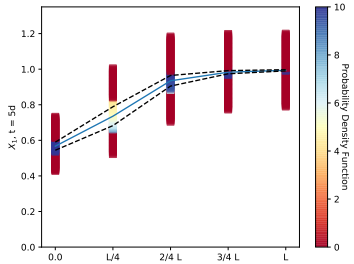
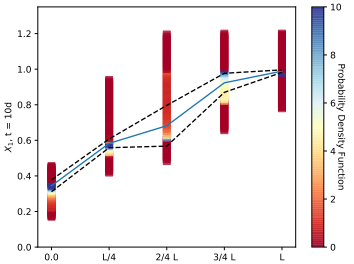


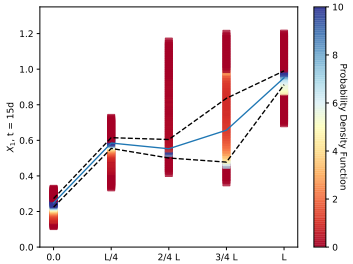
Figure 6.12: Sparsity plots representing the magnitude of the LAR gPC-coefficients $\{\gamma_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in mathcal A}$ with respect to the three-dimensional input space, $\boldsymbol{\theta} = (k_{\text{col}}, k_{\psi}, Y_{\psi})$ $(d = 3)$ and time evolution from 5 to 15 days (from left to right panels). x-, y- and z- axis correspond to the polynomial degrees of the gPC-expansion terms associated with k_{col}, k_{ψ} and Y_{ψ}, respectively. The gPC-expansion under consideration represents the model response for the species volume fraction f_1 (heterotrophic bacteria) at $z = L(t)/4$. The color of the symbols indicates the magnitude of the gPC-coefficients.



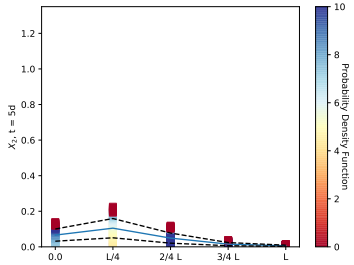
(a) $f_1, t = 5$ days



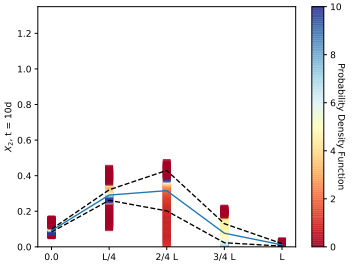
(b) $f_1, t = 10$ days



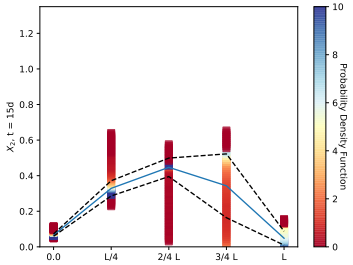
(c) $f_1, t = 15$ days



(d) $f_2, t = 5$ days



(e) $f_2, t = 10$ days



(f) $f_2, t = 15$ days

Figure 6.13: Statistical moments and PDF of each model output $y_{ijk} = f_i(x_j, t_k)$ where i corresponds to the species index, j corresponds to the space index and k corresponds to the time index. The colormap represents the model output PDF at each location and time step. The solid line represents the mean value computed using Eq. (6.32). The dashed lines represent the STD computed using Eq. (6.33), $\mu_y \pm \sigma_y$.

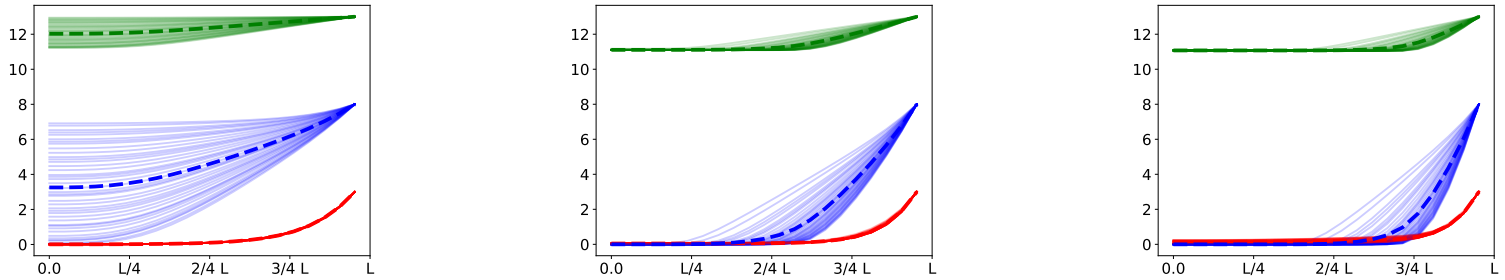


Figure 6.14: Spatial and temporal evolution of the three substrates S_1 (red), S_2 (green) and S_3 (blue) from $z = 0\ \mu m$ to $z = L(t)$ after $t = 5, 10, 15$ days (from left to right panels). The thin solid lines correspond to 40 representative simulations of the biofilm model from Halton's training database. The dashed thick lines correspond to the sample means.

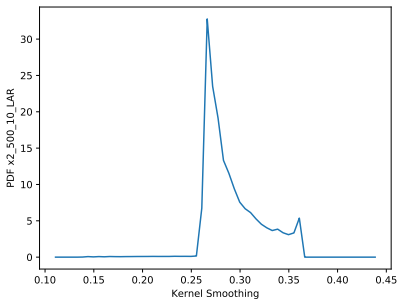
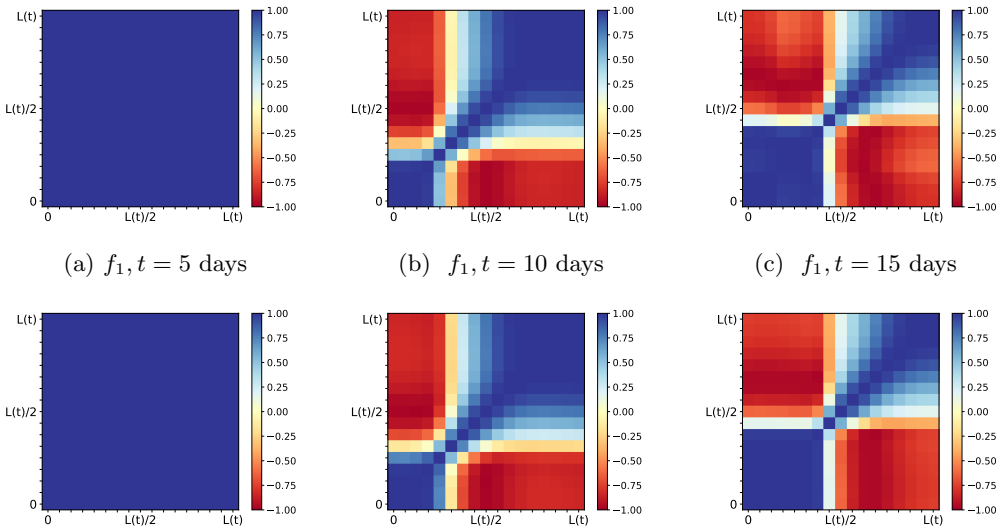


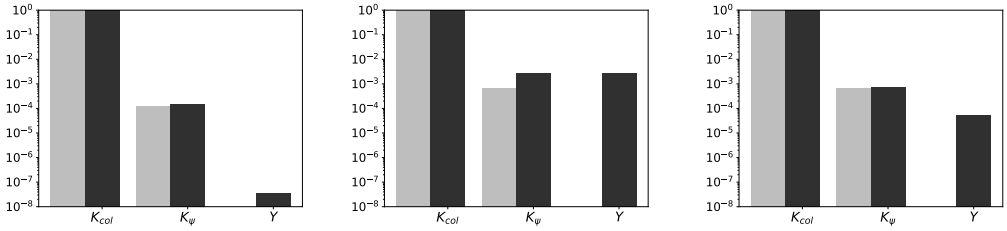
Figure 6.15: Bimodal PDF of the autotrophic species mass fraction f_2 at location $z = L/4$ after 10 days obtained through kernel smoothing.



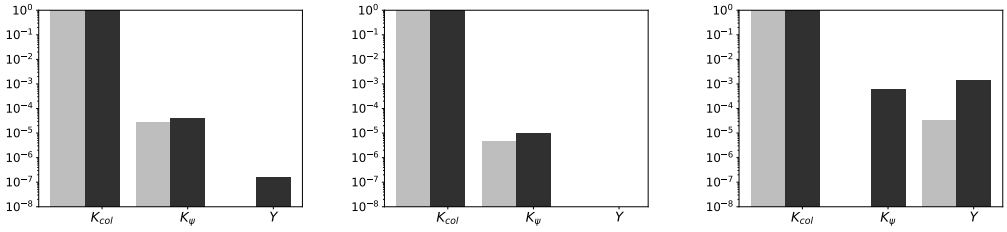
(a) $f_1, t = 5$ days (b) $f_1, t = 10$ days (c) $f_1, t = 15$ days

(d) $f_2, t = 5$ days (e) $f_2, t = 10$ days (f) $f_2, t = 15$ days

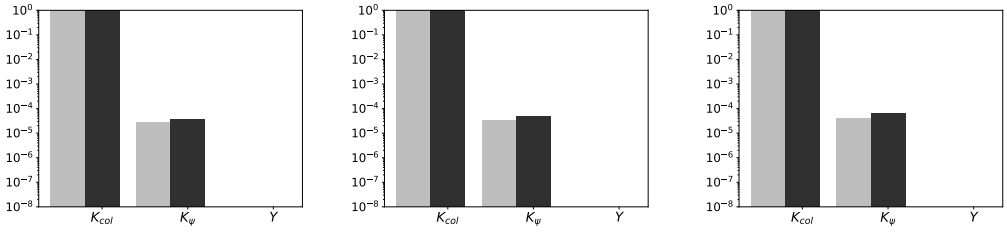
Figure 6.16: Spatial correlation matrices for species volume fractions f_1 (top panels) and f_2 (bottom panels) evolving over time (5 days to 15 days from left to right panels) and computed using Halton's training set with $N = 216$.



(a) $t = 5$ days, $z = 0$ (b) $t = 10$ days, $z = 0$ (c) $t = 15$ days, $z = 0$



(d) $t = 5$ days, $z = L/2$ (e) $t = 10$ days, $z = L/2$ (f) $t = 15$ days, $z = L/2$



(g) $t = 5$ days, $z = L$ (h) $t = 10$ days, $z = L$ (i) $t = 15$ days, $z = L$

Figure 6.17: First-order and total Sobol' indices (in logarithmic scale) associated with input parameters $\boldsymbol{\theta} = (k_{col}, k_{\psi}, Y_{\psi})$ and species volume fraction f_2 (autotrophic bacteria). Time evolution from 5 to 15 days of biofilm growth is portrayed from left to right panels; spatial distribution along the biofilm thickness ($0 \leq z \leq L(t)$) is shown from top to bottom panels. For each panel, light gray colors correspond to first-order Sobol' indices, while dark gray colors correspond to total Sobol' indices. The indices are presented in the following order from left to right bars: k_{col}, k_{ψ}, Y_{ψ}.

Chapter 7

Conclusion

In this final Chapter, conclusions of this PhD work and potential future paths to extend the contributions are reported.

7.1 Conclusions

During this PhD work, the model presented in Chapter 3 ultimately reached a good level of maturity. This has been possible through both a careful study of its theoretical insights and by strengthening the bonds with applications. As a consequence of this study, the use of UQ and SA techniques has been a mandatory step, setting the possibility for a detailed study of several models (even models not pertaining to the class of models that use the formulation of Chapter 3) via *ad hoc* UQ and SA algorithms.
The major contributions are listed as follows:

- The random front propagation model has been tested via a versatile probability density functions borrowed from Fractional Calculus. In particular, at the end of Chapter 3, the influence of the Gaussian shape of the pdf and the variance of the underlying particle displacement is studied through surrogate-based UQ and SA techniques.

- The application to Turbulent Premixed Combustion saw a possible application in the study of hydrodynamic instabilities in combustion. Since the particle displacement, for what concerns the studied case, is connected to a quasi-probability, the limitation of the proposed formalism for tackling a spectra of non-linear phenomena where the random front propagation cannot be ascribed totally to fluctuations around a mean flow are detected [287].

- In the application to wild-land fire simulators, the model needed to be integrated in a framework useful for practitioners and engineers. By the end of the thesis, this model has been implemented into `WRF-Sfire` [283] and `ForeFire`, reducing thus the gap between an otherwise theoretical model of statistical front propagation and the real challenges that

society must tackle in order to reduce wild-land fire related risk. Another aspect that need to be remarked is that wild-land fire application field saw, chronologically, the first use of UQ and SA techniques on the parameters related to the parametrization `RandomFront` of the statistical front propagation model of Chapter 3. This work, summarized by the paper[284], gave us important insights on the role of the different parts of `RandomFront` parameterization on the overall behavior of a wild-land fire. On the other hand, it constitutes one of the most complete and extensive sensitivity analysis for wildfire simulation currently available in literature, that contributed to the ongoing open problem of the comparison of surrogate models (gPC vs. GP) with the use of cross-validation. The of the problem dimension using sparse approach confirmed once more the usefulness of sparse regression techniques in gPC surrogate modeling for UQ and SA of computational models. The last works on wild-land fire are resumed in [77] and [78] (the last one currently under review), which focus on concurrent scales (meso-scale and macro-scale) that share influence on the wild-land fire problem, and the factors in `RandomFront` that express such effects. More specifically, in [77] the effects of the atmospheric stability conditions on the behavior of fire-spotting generated fires have been studied with the proposed formulation for statistical fronts. It emerged that during stable conditions, there are many simultaneous active fires because turbulent heat transport does not merge them rapidly. During unstable conditions, turbulent heat transport emerged to be strong enough for merging the independent fires causing an acceleration in the front propagation and the formation of unburned islands surrounded by the fire. The key idea is that different atmospheric stability conditions cause different kind of risks, which in turn require different management strategies.

- The last application of the model in chronological order is the one regarding bacterial biofilm spread. On the one hand, the model of Chapter3 is adapted to study, with the help of additional analytical tools, the spread of planar bacterial colonies, with the formation of new detached colonies under the assumption of random seeding and dispersal [288]. On the other hand, the knowledge on biofilm modeling thanks to the aforementioned application, as well as the use of UQ and SA tools adopted in [284] laid the basis for a probabilistic analysis of a detailed one-dimensional biofilm model that explicitly models planktonic bacterial invasion in a multi-species biofilm [285]. In particular, the final aim was to quantify and understand how the uncertainty in the input factors of the invasion submodel impacts the overall biofilm model predictions, and to spot which parameters are the most important factors enhancing the response of the biofilm model.

Possible further works starting from the basis laid in this PhD thesis are:

- Completion and publication of all the works cited in Chapter "List of Publications" under Section *"Papers in preparation"*.
- After the publication of [78], a high-dimensional UQ and SA study of the effects of parameters concerning both macroscale (atmospheric stability) and meso-scale (flame geometry - flame length) on the number of separate spotting fires and the ovarall burnt area, possibly using the tools described in [284].
- Concerning biofilms, completion and publication of the work [288] with increased experimental and numerical studies. Subsequent UQ and SA study of the obtained model to spot the most influential factors.

Bibliography

[1] Nwcg, s-290 intermediate wildland fire behavior course. unit 12: Gauging fire behavior and guiding fireline decisions. Url checked on 21-03-2019.

[2] A. Mentrelli ad G. Pagnini. Front propagation in anomalous diffusive media governed by time-fractional diffusion. *J. Comput. Phys.*, 293:427–441, 2015.

[3] F. A. Albini. Spot fire distance from burning trees: a predictive model. Technical Report INT-56, U.S. Department of Agriculture Forest Service Intermountain Forest and Range Experiment Station, 1979.

[4] F. A. Albini. Potential spotting distance from wind-driven surface fires. Research Paper INT-309, U.S. Department of Agriculture Forest Service Intermountain Forest and Range Experiment Station, 1983.

[5] F. A. ALBINI. A model for fire spread in wildland fuels by-radiation. *Combustion Science and Technology*, 42(5-6):229–258, 1985.

[6] M. E. Alexander. Calculating and interpreting forest fire intensities. *Can. J. Bot.*, 60:349–357, 1982.

[7] Erik Alpkvista and Isaac Klapper. A multidimensional multispecies continuum model for heterogeneous biofilm development. *Bulletin of Mathematical Biology*, 69(2):765–789, Feb 2007.

[8] D. H. Anderson, E. A. Catchpole, N. J. De Mestre, and T. Parkes. Modelling the spread of grass fires. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 23(4):451466, 1982.

[9] H. E. Anderson. Aids to determining fuel models for estimating fire behavior. general technical report int-122. Technical report, Intermountain Forest and Range Experiment Station, Ogden, UT, 1982.

[10] Patricia L. Andrews. Behave: fire behavior prediction and fuel modeling system-burn subsystem, part 1. general technical report. General Technical Report INT-194, Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station., July 1986. URL checked on 01-03-19.

[11] Patricia L. Andrews. Current status and future needs of the behaveplus fire modeling system. *Int. J. Wildland Fire*, 23(23):21–33, 2014.

[12] Patricia L. Andrews. The rothermel surface fire spread model and associated developments: A comprehensive explanation. General Technical Report (GTR) RMRS-GTR-371, Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, July 2018. URL checked on 27-02-19.

[13] P.L. Andrews and C.H. Chase. Behave: Fire behavior prediction and fuel modeling system: Burn subsystem, part 2. Research Paper INT-260, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah 84401, 1989.

[14] M. I. Asensio and L. Ferragut. On a wildland fire model with radiation. *Int. J. Numer. Meth. Engng.*, 54:137–157, 2002.

[15] Michael Asher, Barry Croke, Anthony Jakeman, and L.J.M. Peeters. A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51:5957–5973, 2015.

[16] Benjamin Auder, Agns De Crecy, Bertrand Iooss, and Michel Marqus. Screening and metamodeling of computer experiments with functional outputs. application to thermal-hydraulic computations. *Reliability Engineering and System Safety*, 107:122–131, 2012.

[17] Art B. Owen. Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2:439–452, 01 1992.

[18] P. Babak, A. Bourlioux, and T. Hillen. The effect of wind on the propagation of an idealized forest fire. *SIAM J. Appl. Math.*, 70:1364–1388, 2009.

[19] J. H. Balbi, F. Morandini, X. Silvani, J. B. Filippi, and F. Rinieri. A physical model for wildland fires. *Combust. Flame*, 156:2217–2230, 2009.

[20] E. Barkai, Y. Garini, and R. Metzler. Strange kinetics of single molecules in living cells. *Phys. Today*, 65(8):29–35, 2012.

[21] N. Barraud, S. Kjelleberg, and S. A. Rice. Dispersal from microbial biofilms. *Microbiol. Spectr.*, 3, 2015.

[22] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge Mathematical Library. Cambridge University Press, 2000.

[23] R. A. Bates, R. J. Buck, E. Riccomagno, and H. P. Wynn. Experimental design and observation for large systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):77–94, 1996.

[24] Michael Baudin, Khalid Boumhaout, Thibault Delage, Bertrand Iooss, and Jean-Marc Martinez. Numerical stability of Sobol' indices estimation formula. In *8th International Conference on Sensitivity Analysis of Model Output,*, Réunion Island, 2016.

[25] Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin. *OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation*, pages 2001–2038. Springer International Publishing, 2017.

[26] H.R. Baum and B.J. Mccaffrey. Fire induced flow field - theory and experiment. *Fire Safety Science*, 2:129–148, 1989.

[27] Keith Beven. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1):41 – 51, 1993. Research Perspectives in Hydrology.

[28] C. D. Bevins. Firelib: User manual and technical reference. Technical report, US Forest Service, Missoula Fire Sciences Laboratory, Fire Behavior Research Work Unit Systems for Environmental Management, 1996. Accessed 16th August 2018.

[29] S. Bhutia, M. A. Jenkins, and R. Sun. Comparison of firebrand propagation prediction by a plume model and a coupledfire/atmosphere largeeddy simulator. *J. Adv. Model. Earth Syst.*, 2:Art. # 4, 2010.

[30] A. Birolleau, G. Poëtte, and D. Lucor. Adaptive Bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws. *Communications in Computational Physics*, 16:1–34, 2014.

[31] G. Blatman. *Adaptative sparse Polynomial Chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Université Blaise Pascal, Clermont-Ferrand, 2009.

[32] G. Blatman and B. Sudret. Adaptative sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, 2011.

[33] Graud Blatman and Bruno Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 95(11):1216–1229, 2010.

[34] A. Blumen, A. A. Gurtovenko, and S. Jespersen. Anomalous diffusion and relaxation in macromolecular systems. *J. Non-Cryst. Solids*, 305:71–80, 2002.

[35] Joshua P Boltz, Barth F Smets, Bruce E Rittmann, Mark van Loosdrecht, Eberhard Morgenroth, and Glen T Daigger. From biofilm ecology to reactors: a focused review. *Water Science and Technology*, 75(8):1753–1760, 2017.

[36] J.P. Boltz, E. Morgenroth, D. Brockmann, C. Bott, W.J. Gellner, and P.A. Vanrolleghem. Systematic evaluation of biofilm models for engineering practice: components and critical assumptions. *Water Science and Technology*, 64(4):930–944, 2011.

[37] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.

[38] Emanuele Borgonovo. Sensitivity analysis of model output with input constraints: A generalized rationale for local methods. *Risk Analysis*, 28(3):667–680, 2008.

[39] David Bowman, Jennifer Balch, Paulo Artaxo, William Bond, Jean M Carlson, Mark A Cochrane, Carla D'Antonio, Ruth Defries, John Doyle, Sandy Harrison, Fay Johnston, Jon Keeley, Meg Krawchuk, Christian A Kull, J Brad Marston, Max Moritz, Iain Prentice, Christopher Roos, Andrew Scott, and Stephen J Pyne. Fire in the earth system. *Science (New York, N.Y.)*, 324:481–4, 05 2009.

[40] D. Brockmann and E. Morgenroth. Comparing global sensitivity analysis for a biofilm model for two-step nitrification using the qualitative screening method of morris or the quantitative variance-based fourier amplitude sensitivity test (fast). *Water Science and Technology*, 56(8):85, 2007.

[41] B. W. Butler, M.A. Finney, P. L. Andrews, and F. A. Albini. A radiation-driven model for crown fire spread. *Canadian journal of forest research*, 34(8):1588–1599, 2004.

[42] Neil A. Butler. Optimal and orthogonal latin hypercube designs for computer experiments. *Biometrika*, 88(3):847–857, 2001.

[43] G. M. Byram. Combustion of forest fuels. In K. P. Davis, editor, *Forest Fire: Control and Use*, pages 61–89. McGraw Hill, New York, 1959.

[44] R.E. Caflisch and C. Ratsch. Level set methods for simulation of thin film growth. *in the Handbook of Materials Modeling, eds. R. Catlow, H. Shercliff, and S. Yip*, 2004.

[45] K. Campbell, M.D. McKay, and B.J. Williams. Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, 91(10–11):1468–1472, 2006.

[46] Francesca Campolongo and Andrea Saltelli. Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliab. Eng. Syst. Saf.*, 57(1):49–69, 1997.

[47] F. Cappitelli, A. Polo, and F. Villa. Biofilm formation in food processing environments is still poorly understood and controlled. *Food Eng. Rev.*, 6:29–42, 2014.

[48] F. Cappitelli, O. Salvadori, D. Albanese, F. Villa, and C. Sorlini. Cyanobacteria cause black staining of the national museum of the american indian building (washington, d.c., usa). *Biofouling*, 28:257–266, 2011.

[49] W. G. Characklis and P. A. Wilderer, editors. *Structure and Function of Biofilms*, volume 4. John Wiley and Sons, Chichester, 1989.

[50] Karen. Chatto, Kevin G. Tolhurst, and Victoria. *A review of the relationship between fireline intensity and the ecological and economic effects of fire, and methods currently used to collect fire data / Karen Chatto and Kevin G. Tolhurst.* Dept. of Sustainability and Environment East Melbourne, Vic, 2004.

[51] Benito M. Chen-Charpentier and Dan Stanescu. Biofilm growth on medical implants with randomness. *Mathematical and Computer Modelling*, 54(7):1682–1686, 2011.

[52] Ronald Christensen. *Linear Models for Multivariate, Time Series, and Spatial Data.* Springer-Verlag New York, 1 edition, 1991.

[53] K. T. Chu and M. Prodanović. Level set method library (LSMLIB). 2009. `http://ktchu.serendipityresearch.org/software/lsmlib/`.

[54] S. L. Chua, Y. Liu, J. K. Yam, Y. Chen, R. M. Vejborg, B. G. Tan, S. Kjelleberg, T. Tolker-Nielsen, M. Givskov, and L. Yang. Dispersed cells represent a distinct stage in the transition from bacterial biofilm to planktonic lifestyles. *Nat. Commun.*, 5:4462, 2014.

[55] F. Clarelli, C. Di Russo, R. Natalini, and M. Ribot. A fluid dynamics multidimensional model of biofilm growth: stability, influence of environment and sensitivity. *Mathematical Medicine and Biology*, 33(4):371–395, 2016.

[56] Janice L. Coen, Marques Cameron, John Michalakes, Edward G. Patton, Philip J. Riggan, and Kara M. Yedinak. Wrf-fire: coupled weather-wildland fire modeling with the weather research and forecasting model. *J. Appl. Meteor. Climat.*, 52:16–38, 2012.

[57] N. G. Cogan. Two-fluid model of biofilm disinfection. *Bulletin of Mathematical Biology*, 70(3):800–819, Apr 2008.

[58] R. Confalonieri, G. Bellocchi, S. Bregaglio, M. Donatelli, and M. Acutis. Comparison of sensitivity analysis techniques: A case study with the rice model warm. *Ecological Modelling*, 221(16):1897 – 1906, 2010.

[59] J William Costerton, GG Geesey, KJ Cheng, et al. How bacteria stick. *Sci Am*, 238(1):86–95, 1978.

[60] J.W. Costerton, Zbigniew Lewandowski, Douglas Caldwell, Darren Korber, and Hilary Lappin-Scott. Microbial biofilms. *Annual review of microbiology*, 49:711–45, 02 1995.

[61] Francesco Creta, Rachele Lamioni, Pasquale Eduardo Lapenna, and Guido Troiani. Interplay of darrieus-landau instability and weak turbulence in premixed flame propagation. *Phys. Rev. E*, 94:053102, Nov 2016.

[62] Jose Luis da Silva and Mohamed Erraoui. Singularity of generalized grey brownian motions with different parameters. *Stochastic Analysis and Applications*, 11 2016.

[63] B. D'Acunto, L. Frunzo, I. Klapper, and M. R. Mattei. Modeling multispecies biofilms including new bacterial species invasion. *Math. Biosci.*, 259:20–26, 2015.

[64] G. Damblin, M. Couplet, and Iooss. B. Numerical studies of space filling designs : optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*, 2013.

[65] M. De Lozzo and A. Marrel. Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, 31(6):1437–1453, 2017.

[66] F. S. de Sousa, N. Mangiavacchi, L. G. Nonato, A. Castelo, M. F. Tomé, V. G. Ferreira, J. A. Cuminato, and S. McKee. A front-tracking/front-capturing method for the simulation of 3d multi-fluid flows with free surfaces. *J. Comput. Phys.*, 198(2):469–499, August 2004.

[67] D. del Castillo-Negrete. Non-diffusive, non-local transport in fluids and plasmas. *Nonlin. Processes Geophys.*, 17:795–807, 2010.

[68] Bruno Després, Gael Poëtte, and Didier Lucor. *Robust Uncertainty Propagation in Systems of Conservation Laws with the Entropy Closure Method*, pages 105–149. Springer International Publishing, 2013.

[69] Cristian Dinescu, Sergey Smirnov, Charles Hirsch, and Christian Lacor. Assessment of intrusive and non-intrusive non-deterministic cfd methodologies based on polynomial chaos expansions. *International Journal of Engineering Systems Modelling and Simulation*, 2:87–98, 2010.

[70] S. Dubreuil, M. Berveiller, F. Petitjean, and M. Salan. Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion. *Reliability Engineering & System Safety*, 121:263–275, 2014.

[71] B. DAcunto, L. Frunzo, I. Klapper, and M.R. Mattei. Modeling multispecies biofilms including new bacterial species invasion. *Mathematical Biosciences*, 259:20–26, 2015.

[72] B. DAcunto, L. Frunzo, I. Klapper, M.R. Mattei, and P. Stoodley. Mathematical modeling of dispersal phenomenon in biofilms. *Mathematical Biosciences*, 2018.

[73] Berardino DAcunto and Luigi Frunzo. Free boundary problem for an initial cell layer in multispecies biofilm formation. *Applied Mathematics Letters*, 25(1):20–26, 2012.

[74] John W. Eaton, David Bateman, Soren Hauberg, and Rik Wehbring. *GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations*. 2015.

[75] H J. Eberl, D. F. Parker, and M. C. M. Van Loosdrecht. A new deterministic spatio-temporal continuum model for biofilm development. *Journal of Theoretical Medicine*, 3(3):161–175, 2001.

[76] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004.

[77] Vera N. Egorova, Andrea Trucchia, and Gianni Pagnini. Fire-spotting generated fires. part i: The role of atmospheric stability. *Applied Mathematical Modelling*, 2019.

[78] Vera N. Egorova, Andrea Trucchia, and Gianni Pagnini. Fire-spotting generated fires. part ii: The role of flame geometry. *Applied Mathematical Modelling*, 2019. Submitted.

[79] M. El Houssami, E. Mueller, J. C. Thomas, A. Simeoni, A. Filkov, N. Skowronski, M. R. Gallagher, K. Clark, and R. Kremens. Experimental procedures characterising firebrand generation in wildfires. *Fire Technol.*, 52:731–751, 2016.

[80] Charlotte M. Emery, Sylvain Biancamaria, Aaron Boone, Pierre-André Garambois, Sophie Ricci, Mélanie C. Rochoux, and Bertrand Decharme. Temporal variance-based sensitivity analysis of the river-routing component of the large-scale hydrological model isba-trip: Application on the amazon basin. *Journal of Hydrometeorology*, 17(12):3007–3027, 2016.

[81] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.

[82] Henri. Faure. Discrpances de suites associes un systme de numration (en dimension un). *Bulletin de la Socit Mathmatique de France*, 109:143–182, 1981.

[83] A. C. Fernandez-Pello. Wildland fire spot ignition by sparks and firebrands. *Fire Safety Jornal*, 91:2–10, 2017.

[84] J. B. Filippi, F. Bosseur, and D. Grandi. ForeFire: open-source code for wildland fire spread models. In D. X. Viegas, editor, *Advances in forest fire research*, pages 275–282. Imprensa da Universidade de Coimbra, 2014.

[85] J. B. Filippi, F. Bosseur, X. Pialat, P. A. Santoni, S. Strada, and C. Mari. Simulation of coupled Fire/Atmosphere interaction with the MesoNH-ForeFire models. *J. Combust.*, 2011:540390, 2011.

[86] J. B. Filippi, V. Mallet, and B. Nader. Evaluation of forest fire models on a large observation database. *Nat. Hazards Earth Syst. Sci.*, 14:3077–3091, 2014.

[87] J. B. Filippi, F. Morandini, J. H. Balbi, and D. Hill. Discrete event front tracking simulator of a physical fire spread model. *Simulation*, 86:629–646, 2009.

[88] J. B. Filippi, X. Pialat, and C. B. Clements. Assessment of ForeFire/Meso-NH for wildland fire/atmosphere coupled simulation of the FireFlux experiment. *Proc. Combust. Inst.*, 34:2633–2640, 2013.

[89] M. Finney. Farsite: fire area simulator - model development and evaluation. Research Paper RMRS-RP-4, USDA Forest Service, Rocky Mountain Research Station, Ogden, Utah, 1998.

[90] M. Finney. Fire growth using minimum travel time methods. *Can. J. For. Res.*, 32:420–1424, 2002.

[91] M. Finney. Calculation of fire spread rates across random landscapes. *Int. J. Wildland Fire*, 12:167–174, 2003.

[92] Mark A. Finney. Farsite: Fire area simulator-model development and evaluation . General Technical Report RMRS-RP-4, Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1998. URL checked on 01-03-19.

[93] H. C. Flemming, J. Wingender, U. Szewzyk, P. Steinberg, S. A. Rice, and S. Kjelleberg. Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.*, 14:563–575, 2016.

[94] Hans-Curt Flemming, Jost Wingender, Ulrich Szewzyk, Peter Steinberg, Scott A Rice, and Staffan Kjelleberg. Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology*, 14(9):563, 2016.

[95] Navin Fogla. *Interplay between background turbulence and Darrieus- Landau instability in premixed flames via a model equation.* PhD thesis, 2010.

[96] Navin Fogla. *Propagation of large scale premixed turbulent flames*. PhD thesis, 2014.

[97] W. H. Frandsen. Fire spread through porous fuels from the conservation of energy. *Combust. Flame*, 16:9–16, 1971.

[98] Jim Freer, Keith Beven, and Bruno Ambroise. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach. *Water Resources Research*, 32(7):2161–2173, 1996.

[99] Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agns Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603, 2014.

[100] Samoradnitsky Gennady and Taqqu M.S. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall CRC, 1994.

[101] Y. Giga. Surface evolution equations: A level set approach. *Springer Monographs in Mathematics*, 2006.

[102] Joo Paulo Gois, Anderson Nakano, Luis Gustavo Nonato, and Gustavo C. Buscaglia. Front tracking with moving-least-squares surfaces. *Journal of Computational Physics*, 227(22):9643 – 9669, 2008.

[103] C. Guilhen, C. Forestier, and D. Balestrino. Biofilm dispersal: multiple elaborate strategies for dissemination of bacteria with unique properties. *Mol. Microbiol.*, 105:188–210, 2017.

[104] W. Guo and H. H. Sawin. Review of profile and roughening simulation in microelectronics plasma etching. *J. Phys. D: Appl. Phys.*, 42:194014, 2009.

[105] M.E. Gurtin. *An Introduction to Continuum Mechanics*. Mathematics in Science and Engineering. Elsevier Science, 1982.

[106] K. D. Hage. On the dispersion of large particles from a 15-m source in the atmosphere. *J. Appl. Meteor.*, 18:534–539, 1961.

[107] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2(1):84–90, December 1960.

[108] A. Hanyga. Multidimensional solutions of time-fractional diffusion-wave equations. *Proc. R. Soc. Lond. A*, 458:933–957, 2002.

[109] Xiaodi Hao, Joseph J. Heijnen, and Mark C. M. van Loosdrecht. Sensitivity analysis of a biofilm model describing a one-stage completely autotrophic nitrogen removal (canon) process. *Biotechnology and Bioengineering*, 77(3):266–277, 2002.

[110] Elizabeth B. Harper, John C. Stella, and Alexander K. Fremier. Global sensitivity analysis for complex ecological models: a case study of riparian cottonwood population dynamics. *Ecological Applications*, 21(4):1225–1240, 2011.

[111] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction.* Springer, 2 edition, 2009.

[112] Jon C Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2):327 – 367, 1993.

[113] K. Himoto and T. Tanaka. Transport of disk-shaped firebrands in a turbulent boundary layer. In D.T. Gottuk and B.Y. Lattimer, editors, *The Eighth International Symposium on Fire Safety Science, 1823 September 2005, Beijing, China*, pages 433–444. nternational Association for Fire Safety Science: Baltimore, MD, 2005.

[114] F. Höfling and T. Franosch. Anomalous transport in the crowded world of biological cells. *Rep. Prog. Phys.*, 76:046602, 2013.

[115] Ronald A. Howard. Decision analysis: Practice and promise. *Management Science*, 34(6):679–695, 1988.

[116] J. D. Hunter. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, 9(3):90–95, 2007.

[117] R I. Cukier, C M. Fortuin, K E. Shuler, A G. Petschek, and J H. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i. theory. *The Journal of Chemical Physics*, 59:3873–3878, 10 1973.

[118] B. Iooss and A. Saltelli. Introduction to Sensitivity Analysis. In *Handbook of Uncertainty Quantification*, pages 1–20. Springer International Publishing, 2016.

[119] Bertrand Iooss and Paul Lematre. *A review on global sensitivity analysis methods.* Springer US, 1 edition, 2015.

[120] Pahala Gedara Jayathilake, Prashant Gupta, Bowen Li, Curtis Madsen, Oluwole Oyebamiji, Rebeca Gonzlez-Cabaleiro, Steve Rushton, Ben Bridgens, David Swailes, Ben Allen, A. Stephen McGough, Paolo Zuliani, Irina Dana Ofiteru, Darren Wilkinson, Jinju Chen, and Tom Curtis. A mechanistic individual-based model of microbial communities. *PLOS ONE*, 12(8):1–26, 08 2017.

[121] E. Jettestuen, J. O. Helland, and M. Prodanović. A level set method for simulating capillary-controlled displacements at the pore scale with nonzero contact angles. *Water Resour. Res.*, 49:4645–4661, 2013.

[122] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[123] O. Juan, R. Keriven, and G. Postelnicu. Stochastic motion and the Level Set Method in Computer Vision: Stochastic active contours. *Int. J. Comput. Vision*, 69:7–25, 2006.

[124] H. Karimabadi, J. Driscoll, Y. A. Omelchenko, and N. Omidi. A new asynchronous methodology for modeling of physical systems: breaking the curse of Courant condition. *J. Comput. Phys.*, 205:755–775, 2005.

[125] I. Kaur and G. Pagnini. Fire spotting modelling and parametrisation for wild-land fires. In Sauvage S., Snchez-Prez J.M., and Rizzoli A., editors, *International Environmental Modelling and Software Society (iEMSs) 8th International Congress on Environmental Modelling and Software*, volume 55, 2016.

[126] Inderpreet Kaur, Andrea Mentrelli, Frédéric Bosseur, Jean-Baptiste Filippi, and Gianni Pagnini. Turbulence and fire-spotting effects into wild-land fire simulators. *Commun. Nonlinear Sci. Numer. Simul.*, 39:300 – 320, 2016.

[127] Marc C. Kennedy and Anthony OHagan. Bayesian calibration of computer models. *J. R. Statist. Soc. B*, 43:425–464, 2001.

[128] Virginia Kiryakova. Generalized fractional calculus and applications. *Pitman Research Notes Mathematics Series*, 301, 01 1993.

[129] Isaac Klapper and Jack Dockery. Mathematical description of microbial biofilms. *SIAM review*, 52(2):221–265, 2010.

[130] E. Koo, P. J. Pagni, D. R. Weise, and J. P. Woycheese. Firebrands and spotting ignition in large-scale fires. *Int. J. Wildland Fire*, 19:818–843, 2010.

[131] E. Koo, P.J. Pagni, and R.R. Linn. Using firetec to describe firebrand behavior in wildfires. In *The Tenth International Symposium of Fire and Materials, 2931 January 2007, San Francisco, CA*. Interscience Communications: London, UK, 2007.

[132] H. Koo, R. N. Allan, R. P. Howlin, P. Stoodley, and L. Hall-Stoodley. Targeting microbial biofilms: current and prospective therapeutic strategies. *Nat. Rev. Microbiol.*, 15:740–755, 2017.

[133] S. Kortas, P. Mindykowski, J. L. Consalvi, H. Mhiri, and B. Porterie. Experimental validation of a numerical model for the transport of firebrands. *Fire Safety J.*, 44:1095–1102, 2009.

[134] Jan-Ulrich Kreft, Cristian Picioreanu, Julian W. T. Wimpenny, and Mark C. M. van Loosdrecht. Individual-based modelling of biofilms. *Microbiology*, 147(11):2897–2912, 2001.

[135] B Krzykacz-Hausmann. Epistemic sensitivity analysis based on the concept of entropy. In *Proceedings of SAMO2001*, pages 31–35, 01 2001.

[136] A. Kurenkov and M. Oberlack. Rans-simulation of premixed turbulent combustion using the level set approach. *PAMM*, 5(1):469–470, 2005.

[137] Susanne Lackner and Barth F. Smets. Effect of the kinetics of ammonium and nitrite oxidation on nitritation success or failure for different biofilm reactor geometries. *Biochemical Engineering Journal*, 69:123–129, 2012.

[138] Pierre Lallemand, Li-Shi Luo, and Yan Peng. A lattice boltzmann front-tracking method for interface dynamics with surface tension in two dimensions. *Journal of Computational Physics*, 226:1367–1384, 10 2007.

[139] L. D. Landau. On the theory of slow combustion. *Acta Physiochimica*, 19:77–85, 1944.

[140] Laurent A. Lardon, Brian V. Merkey, Snia Martins, Andreas Dtsch, Cristian Picioreanu, Jan-Ulrich Kreft, and Barth F. Smets. idynomics: next-generation individual-based modelling of biofilms. *Environmental Microbiology*, 13(9):2416–2434, 2011.

[141] L. Le Gratiet, C. Cannamela, and B. Iooss. A bayesian approach for global sensitivity analysis of (multifidelity) computer codes. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):336–363, 2014.

[142] L. Le Gratiet, S. Marelli, and B. Sudret. Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes. In *Handbook of Uncertainty Quantification*, pages 1–37. Springer International Publishing, 2017.

[143] O. Le Maitre and O. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.

[144] Rhea P. Liem, Charles A. Mader, and Joaquim R.R.A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis. *Aerospace Science and Technology*, 43:126–151, 2015.

[145] R. R. Linn and P. Cunningham. Numerical simulations of grass fires using a coupled atmosphere-fire model: Basic fire behavior and dependence on wind speed. *J. Geophys. Res.*, 110:D13107, 2005.

[146] Rodman Linn, Judith Winterkamp, J Canfield, Jeremy Sauer, J Colman, Jon Reisner, C Edminster, Franois Pimont, J Dupey, and Philip Cunningham. Versitility of firetec, a physics-based wildfire model. *Forest Ecology and Management - FOREST ECOL MANAGE*, 234, 11 2006.

[147] A. Lipatnikov. *Fundamentals of Premixed Turbulent Combustion.* CRC Press, 2012.

[148] A. N. Lipatnikov and V. A. Sabel'nikov V.A. Some basic issues of the averaged G-equation approach to premixed turbulent combustion modeling. *Open Thermodyn. J.*, 2:53–58, 2008.

[149] B.A. Lockwood and M. Anitescu. Gradient-enhanced universal kriging for uncertainty propagation. *Nucl. Sci. Eng.*, pages 1–32, 2012.

[150] M. Machacek and G. Danuser. Morphodynamic profiling of protrusion phenotypes. *Biophys. J.*, 90:1439–1452, 2006.

[151] F. Mainardi. Fractional relaxation-oscillation and fractional diffusion-wave phenomena. *Chaos Solitons Fract.*, 7:1461–1477, 1996.

[152] F. Mainardi. The fundamental solutions for the fractional diffusion-wave equation. *Appl. Math. Lett.*, 9(6):23–28, 1996.

[153] F. Mainardi. *Fractional Calculus and Waves in Linear Viscoelasticity.* Imperial College Press, London, 2010.

[154] F. Mainardi, Yu. Luchko, and G. Pagnini. The fundamental solution of the space-time fractional diffusion equation. *Fract. Calc. Appl. Anal.*, 4(2):153–192, 2001.

[155] F. Mainardi, A. Mura, and G. Pagnini. The M-Wright function in time-fractional diffusion processes: A tutorial survey. *Int. J. Diff. Equations*, 2010:104505, 2010.

[156] F. Mainardi, G. Pagnini, and R. Gorenflo. Mellin transform abd subordination laws in fractional diffusion processes. *Fract. Calc. Appl. Anal.*, 6:441–459, 2003.

[157] V. Mallet, D. E. Keyes, and F. E. Fendell. Modeling wildland fire propagation with level set methods. *Comput. Math. Appl.*, 57:1089–1101, 2009.

[158] J. Mandel, S. Amram, J. D. Beezley, G. Kelman, A. K. Kochanski, V. Y. Kondratenko, B. H. Lynn, B. Regev, and M. Vejmelka. Recent advances and applications of wrf-sfire. *nhess*, 14(10):2829–2845, 2014.

[159] 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.

[160] J. Mandel, L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. Douglas, M. Kim, and A. Vodacek. A wildland fire model with data assimilation. *Math. Comput. Simulat.*, 79:584–606, 2008.

[161] S. L. Manzello, T. G. Cleary, J. R. Shields, and J. C. Yang. On the ignition of fuel beds by firebrands. *Fire Mater.*, 30:77–87, 2006.

[162] S. L. Manzello, J. R. Shields, T. G. Cleary, A. Maranghides, W. E. Mell, J. C. Yang, Y. Hayashi, D. Nii, and T. Kurita. On the development and characterization of a firebrand generator. *Fire Saf. J.*, 43:258–268, 2008.

[163] Samuel L. Manzello, Alexander Maranghides, and William E. Mell. Firebrand generation from burning vegetation. *Int. J. Wildland Fire*, 16:458–462, 2007.

[164] Samuel L. Manzello, Alexander Maranghides, John R. Shields, William E. Mell, Yoshihiko Hayashi, and Daisaku Nii. Mass and size distribution of firebrands generated from burning korean pine (pinus koraiensis) trees. *Fire Mater.*, 33(1):21–31, 2009.

[165] G.H. Markstein, North Atlantic Treaty Organization. Advisory Group for Aeronautical Research, and Development. *Nonsteady flame propagation*. Agardograph Series. Published for and on behalf of Advisory Group for Aeronautical Research and Development, North Atlantic Treaty Organization by Pergamon Press, 1964.

[166] A. Marrel, G. Perot, and C. Mottet. Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, 29(3):959–974, 2015.

[167] Amandine Marrel, Bertrand Iooss, Beatrice Laurent, and Olivier Roustant. Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751, 2009.

[168] Amandine Marrel, Nathalie Saint-Geours, and Matthias De Lozzo. *Sensitivity Analysis of Spatial and/or Temporal Phenomena*, pages 1–31. Springer International Publishing, Cham, 2016.

[169] K.C. Marshall, R. Stout, and R. Mitchell. Mechanism of the initial events in the sorption of marine bacteria to surfaces. *J Gen Microbiol*, 68:337–348, 1971.

[170] J. Martin and T. Hillen. The spotting distribution of wildfires. *Appl. Sci.*, 6(6):177–210, 2016.

[171] J-M. Martinez. Analyse de sensibilite globale par decomposition de la variance. In *Presentation in the meeting of GdR Ondes and GdR MASCOT-NUM*, Paris, France, Jaunary 2011.

[172] MR Mattei, L Frunzo, B D'Acunto, Y Pechaud, F Pirozzi, and G Esposito. Continuum and discrete approach in modeling biofilm development and structure: a review. *Journal of Mathematical Biology*, 76(4):945–1003, 2018.

[173] D. McDougald, S. A. Rice, N. Barraud, P. D. Steinberg, and S. Kjelleberg. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nat. Rev. Microbiol.*, 10:39–50, 2011.

[174] E. L. McGrath-Spangler and A. Molod. Comparison of geos-5 agcm planetary boundary layer depths computed with various definitions. *Atmospheric Chemistry and Physics*, 14(13):6717–6727, 2014.

[175] William Mell, Mary Jenkins, James Gould, and Phil Cheney. A physics-based approach to modeling grassland fires. *International Journal of Wildland Fire - INT J WILDLAND FIRE*, 16, 01 2007.

[176] A. Mentrelli and G. Pagnini. Random front propagation in fractional diffusive systems. *submitted*, 2014.

[177] A. Mentrelli and G. Pagnini. Modelling and simulation of wildland fire in the framework of the level set method. *Ricerche Mat.*, 65:523–533, 2016.

[178] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.*, 16:24128–24164, 2014.

[179] R. Metzler and J. Klafter. The restaurant at the end of the random walk: recent developments in fractional dynamics descriptions of anomalous dynamical processes. *J. Phys. A: Math. Theor.*, 37(31):R161–R208, 2004.

[180] D.M. Michelson and G. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames - ii. numerical experiments. *Acta Astronautica*, 4:1207–1221, 1977.

[181] Ian M. Mitchell. The flexible, extensible and efficient toolbox oflevelset methods. *Journal of Scientific Computing*, 35(2):300–329, Jun 2008.

[182] Daniel Molina-García, Tuan Minh Pham, Paolo Paradisi, Carlo Manzo, and Gianni Pagnini. Fractional kinetics emerging from ergodicity breaking in random media. *Phys. Rev. E*, 94:052147, Nov 2016.

[183] J. J. Monaghan. Smoothed particle hydrodynamics. *Rep. Prog. Phys.*, 68:1703–1759, 2005.

[184] R. Montenegro, A. Plaza, L. Ferragut, and M. I. Asensio. Application of a nonlinear evolution model to fire propagation. *Nonlinear Anal. Theory Methods Appl.*, 30:2873–2882, 1997.

[185] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[186] A. Mura, M.S. Taqqu, and F. Mainardi. Non-markovian diffusion equations and processes: Analysis and simulations. *Physica A: Statistical Mechanics and its Applications*, 387(21):5033 – 5064, 2008.

[187] Antonio Mura and Gianni Pagnini. Characterizations and simulations of a class of stochastic processes to model anomalous diffusion. *Journal of Physics A: Mathematical and Theoretical*, 41:285003, 06 2008.

[188] A. Muraszew. Firebrand phenomena. Aerospace Report ATR-74(8165-01)-1, The Aerospace Corp., El Segundo, CA, 1974.

[189] Vicen Mndez, S Fedotov, and W Horsthemke. *Reaction-Transport Systems: Mesoscopic Foundations, Fronts, and Spatial Instabilities*. Springer-Verlag Berlin Heidelberg, 01 2010.

[190] C. D. Nadell, K. Drescher, and K. R. Foster. Spatial structure, cooperation and competition in biofilms. *Nat. Rev. Microbiol.*, 14:589–600, 2016.

[191] J. J. Niemela, Skrbek L., K. R. Sreenivasan, and Donnelly R. J. Turbulent convection at very high rayleigh numbers. *Nature*, 404, 2000.

[192] J.E. Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769, 2004.

[193] M. Oberlack, H. Wenzel, and N. Peters. On symmetries and averaging of the g-equation for premixed combustion. *Combust. Theory Modelling*, 5:363–383, 2001.

[194] T. Ohta, D. Jasnow, and K. Kawasaki. Universal scaling in the motion of random interfaces. *Phys. Rev. Lett.*, 49:1223–1226, 1982.

[195] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer New York, 2006.

[196] S. J. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations. *J. Comput. Phys.*, 79:12–49, 1988.

[197] N.E. Owen, P. Challenor, P. P. Menon, and S. Bennani. Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):403–435, 2017.

[198] G. Pagnini. The M-Wright function as a generalization of the Gaussian density for fractional diffusion processes. *Fract. Calc. Appl. Anal.*, 16(2):436–453, 2013.

[199] G. Pagnini. A model of wildland fire propagation including random effects by turbulence and fire spotting. In *Proceedings of XXIII Congreso de Ecuaciones Diferenciales y Aplicaciones XIII Congreso de Matemática Aplicada. Castelló, Spain, 9–13 September 2013*, pages 395–403, 2013.

[200] G. Pagnini. Fire spotting effects in wildland fire propagation. In F. Casas and V. Martínez, editors, *Advances in Differential Equations and Applications*, volume 4 of *SEMA SIMAI Springer Series*, pages 203–216. Springer International Publishing Switzerland, 2014. ISBN: 978-3-319-06952-4. (eBook: 978-3-319-06953-1).

[201] G. Pagnini. Short note on the emergence of fractional kinetics. *Physica A*, 409:29–34, 2014.

[202] G. Pagnini and E. Bonomi. Lagrangian formulation of turbulent premixed combustion. *Phys. Rev. Lett.*, 107:044503, 2011.

[203] G. Pagnini and L. Massidda. Modelling turbulence effects in wildland fire propagation by the randomized level-set method. Tech. Rep 2012/PM12a, CRS4, July 2012. Revised Version August 2014. arXiv:1408.6129.

[204] G. Pagnini and L. Massidda. The randomized level-set method to model turbulence effects in wildland fire propagation. In D. Spano, V. Bacciu, M. Salis, and C. Sirca, editors, *Modelling Fire Behaviour and Risk. Proceedings of the International Conference on Fire Behaviour and Risk. ICFBR 2011, Alghero, Italy, October 4–6 2011*, pages 126–131, May 2012. ISBN 978-88-904409-7-7.

[205] G. Pagnini and A. Mentrelli. Modelling wildland fire propagation by tracking random fronts. *Nat. Hazards Earth Syst. Sci. Discuss.*, 1:6521–6557, 2013.

[206] G. Pagnini and A. Mentrelli. Modelling wildland fire propagation by tracking random fronts. *Nat. Hazards Earth Syst. Sci.*, 14:2249–2263, 2014.

[207] G. Pagnini and A. Mentrelli. *Progress in Industrial Mathematics at ECMI 2014*, volume 74, chapter The Randomized Level Set Method and an Associated Reaction-Diffusion Equation to Model Wildland Fire Propagation. Springer International Publishing, 2014.

[208] Gianni Pagnini. Erdélyi-kober fractional diffusion. *Fractional Calculus and Applied Analysis*, 15(1):117–127, Mar 2012.

[209] Gianni Pagnini, Antonio Mura, and Francesco Mainardi. Two-particle anomalous diffusion: probability density functions and self-similar stochastic processes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1990):20120154, 2013.

[210] Gianni Pagnini and Paolo Paradisi. A stochastic solution with gaussian stationary increments of the symmetric space-time fractional diffusion equation. *Fractional Calculus and Applied Analysis*, 19, Nov 2016.

[211] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12(2825-2830), 2012.

[212] Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Comput. Sci. Eng.*, 9(3):21–29, May 2007.

[213] H. A. Perryman, C. J. Dugaw, J. Morgan Varner, and D. L. Johnson. A cellular automata model to link surface fires to firebrand lift-off and dispersal. *Int. J. Wildland Fire*, 22:428–439, 2013.

[214] N. Peters. A spectral closure for premixed turbulent combustion in the flamelet regime. *J. Fluid Mech.*, 242:611–629, 1992.

[215] N. Peters. The turbulent burning velocity for large-scale and small-scale turbulence. *J. Fluid Mech.*, 384:107–132, 1999.

[216] N. Peters. *Turbulent Combustion*. Cambridge University Press, Cambridge, 2004.

[217] Per Pettersson, Alireza Doostan, and Jan Nordström. Level Set Methods for Stochastic Discontinuity Detection in Nonlinear Problems. *Under review in Journal of Computational Physics*, page arXiv:1810.08607, 2018.

[218] Francesca Pianosi, Keith Beven, Jim Freer, Jim W. Hall, Jonathan Rougier, David B. Stephenson, and Thorsten Wagener. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79:214 – 232, 2016.

[219] Francesca Pianosi and Thorsten Wagener. A simple and efficient method for global sensitivity analysis based oncumulative distribution functions. *Environmental Modelling & Software*, 67:1 – 11, 2015.

[220] Cristian Picioreanu, Jan-Ulrich Kreft, and Mark C M Van Loosdrecht. Particle-based multidimensional multispecies biofilm model. *Applied and environmental microbiology*, 70(5):30243040, May 2004.

[221] Len Pismen. *Patterns and Interfaces in Dissipative Dynamics*. Springer-Verlag Berlin Heidelberg, 01 2017.

[222] I. Podlubny. *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. Mathematics in Science and Engineering. Elsevier Science, 1998.

[223] I. Podlubny. *Fractional Differential Equations*. Academic Press, San Diego, 1999.

[224] Gaël Poëtte, Bruno Després, and Didier Lucor. Uncertainty quantification for systems of conservation laws. *Journal of Computational Physics*, 228(7):2443–2467, April 2009.

[225] Bernard Porterie, Nouredine Zekri, Jean-Pierre Clerc, and Jean-Claude Loraud. Modeling forest fire spread and spotting process with small world networks. *Combust. Flame*, 149(1-2):63–78, 2007.

[226] L. Pugnet, D.M. Chong, T.J. Duff, and K.G. Tolhurst. Wildlandurban interface (WUI) fire modelling using PHOENIX Rapidfire: A case study in Cavaillon, France. In J. Piantadosi, R.S. Anderssen, and J. Boland, editors, *MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2013*, pages 228–234, 2013.

[227] B. Purevdorj-Gage, W. J. Costerton, and P. Stoodley. Phenotypic differentiation and seeding dispersal in non-mucoid and mucoid Pseudomonas aeruginosa biofilms. *Microbiology*, 151:1569–1576, 2005.

[228] Kazi A. Rahman, Rangarajan Sudarsan, and Hermann J. Eberl. A mixed-culture biofilm model with cross-diffusion. *Bulletin of Mathematical Biology*, 77(11):2086–2124, Nov 2015.

[229] C.E. Rasmussen and C Williams. *Gaussian processes for machine learning*. MIT Press, 2006.

[230] Peter Reichert. Aquasim - a tool for simulation and data analysis of aquatic systems. *Water Science and Technology*, 30(2):21, 1994.

[231] G. D. Richards. The properties of elliptical wildfire growth for time dependent fuel and meteorological conditions. *Combust. Sci. Technol.*, 95:357–383, 1993.

[232] G. D. Richards. A general mathematical framework for modelling two-dimensional wildland fire spread. *Int. J. Wildland Fire*, 5:63–72, 1995.

[233] O. Rios, W. Jahn, and G. Rein. Forecasting wind-driven wildfires using an inverse modelling approach. *Nat. Hazards Earth Syst. Sci. Discuss.*, 14:1491–1503, 2014.

[234] B.E. Rittmann and P.L. McCarty. *Environmental Biotechnology: Principles and Applications*. McGraw Hill, New York, USA., 2001.

[235] Bruce Rittmann, Alex Schwarz, and P.B. Saez. Biofilms applied to hazardous waste treatment. *Biofilms II*, pages 207–234, 01 2000.

[236] Jean-Louis Rossi, Khaled Chetehouna, Anthony Collin, Basiliu Moretti, and Jacques-Henri Balbi. Simplified flame models and prediction of the thermal radiation emitted by a flame front in an outdoor fire. *Combustion Science and Technology*, 182:1457–1477, 10 2010.

[237] R. C. Rothermel. A mathematical model for predicting fire spread in wildland fires. Tech. Rep. Research Paper INT-115, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah 84401, 1972. available at: http://www.treesearch.fs.fed.us/pubs/32533.

[238] Pamphile T. Roy, Nabil El Moçayd, Sophie Ricci, Jean-Christophe Jouhaud, Nicole Goutal, Matthias De Lozzo, and Mélanie C. Rochoux. Comparison of polynomial chaos and gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. *Stochastic Environmental Research and Risk Assessment*, 32(6):1723–1741, Jun 2018.

[239] Pamphile T. Roy, Sophie Ricci, Romain Dupuis, Robin Campet, Jean-Christophe Jouhaud, and Cyril Fournier. Batman: Statistical analysis for expensive computer codes made easy. *Journal of Open Source Software*, 3(21):493, 2018.

[240] V. A. Sabel'nikov and A. N. Lipatnikov. Rigorous derivation of an unclosed mean G-equation for statistically 1D premixed turbulent flames. *Int. J. Spray Combust. Dyn.*, 2:301–324, 2010.

[241] V. A. Sabelnikov and A. N. Lipatnikov. Towards an extension of TFC model of premixed turbulent combustion. *Flow Turbul. Combust.*, 90:387–400, 2013.

[242] A. Saltelli, S. Tarantola, and K.P.S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56, 1999.

[243] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, Chichester, UK, dec 2007.

[244] N. Sardoy, J. L. Consalvi, A. Kaiss, A. C. Fernandez-Pello, and B. Porterie. Numerical study of ground-level distribution of firebrands generated by line fires. *Combust. Flame*, 154:478–488, 2008.

[245] N. Sardoy, J. L. Consalvi, B. Porterie, and A. C. Fernandez-Pello. Modeling transport and combustion of firebrands from burning trees. *Combust. Flame*, 150:151–169, 2007.

[246] R. Schoebi, B. Sudret, and J. Wiart. Polynomial-Chaos-based Kriging. *Int. J. Uncertain. Quan.*, 5(2):171–193, 2015.

[247] S. Serafin and D. Zardi. Critical evaluation and proposed refinement of the Troen and Marht (1986) boundary layer model. *ICAM/MAP conference*, pages 464–467, 2005.

[248] F. J. Serón, D. Gutiérrez, J. Magallón, L. Ferragut, and M. I. Asensio. The evolution of a wildland forest fire front. *Visual Comput.*, 21:152–169, 2005.

[249] J. A. Sethian. Curvature and the evolution of fronts. *Commun. Math. Phys.*, 101:487–499, 1985.

[250] J. A. Sethian and P. Smereka. Level set methods for fluid interfaces. *Ann. Rev. Fluid Mech.*, 35:341–372, 2003.

[251] James A. Sethian. *AN ANALYSIS OF FLAME PROPAGATION.* PhD thesis, 1982. LBNL Report #: LBL-14125. Retrieved from https://escholarship.org/uc/item/87d4t5q3.

[252] James Albert Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999.

[253] G. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames - i. derivation of basic equations. *Acta Astronautica*, 4:1177–1206, 1977.

[254] G. Sivashinsky. Instabilities, pattern formation, and turbulence in flames. *Annual Review of Fluid Mechanics*, 15:179–199, 1983.

[255] W. C. Skamarock, J.B. Klemp, J. Dudhia, O. D. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers. A description of the advanced research wrf version 3. Technical report, 2008.

[256] R.C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Computational Science and Engineering. Society for Industrial and Applied Mathematics, 2013.

[257] I.M Sobol. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1(4):407–414, 1993.

[258] M. Sofiev, T. Ermakova, and R. Vankevich. Evaluation of the smoke-injection height from wild-land fires using remote-sensing data. *Atmos. Chem. Phys.*, 12(4):1995–2006, 2012.

[259] H. M. Soner and N. Touzi. A stochastic representation for the level set equations. *Commun. Part. Diff. Eq.*, 27:2031–2053, 2002.

[260] Sara Soto. Importance of biofilms in urinary tract infections: New therapeutic approaches. *Advances in Biology*, 2014:1–13, 07 2014.

[261] Dan Stanescu and Benito M. Chen-Charpentier. Random coefficient differential equation models for bacterial growth. *Mathematical and Computer Modelling*, 50(5):885–895, 2009.

[262] Paul Stoodley, Karin Sauer, David Gwilym Davies, and J William Costerton. Biofilms as complex differentiated communities. *Annual Reviews in Microbiology*, 56(1):187–209, 2002.

[263] C. K Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature*, 406:959–964, 2000.

[264] R. B. Stull. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 1988.

[265] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.

[266] G. Sugiyama and J.S. Nasstrom. Methods for determining the height of the atmospheric boundary layer. Technical report, UCRL-ID-133200, Lawrence Livermore National Laboratory Report, 1999.

[267] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *Int. J. Wildland Fire*, 18:349–368, 2009.

[268] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *Int. J. Wildland Fire*, 18:369–386, 2009.

[269] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *Int. J. Wildland Fire*, 18:387–403, 2009.

[270] A. L. Sullivan. Inside the inferno: Fundamental processes of wildland fire behaviour. Part 1: Combustion chemistry and heat release. *Curr. Forestry Rep.*, 3:132–149, 2017.

[271] A. L. Sullivan. Inside the inferno: Fundamental processes of wildland fire behaviour. Part 2: Heat transfer and interaction. *Curr. Forestry Rep.*, 3:150–171, 2017.

[272] D. Taherzadeh, C. Picioreanu, and H. Horn. Mass transfer enhancement in moving biofilm structures. *Biophysical Journal*, 102(7):1483–1492, 04 2012.

[273] Youneng Tang and Albert J Valocchi. An improved cellular automaton method to model multispecies biofilms. *Water research*, 47(15):5729–5742, October 2013.

[274] Stefano Tarantola and Andrea Saltelli. Samo 2001: methodological advances and innovative applications of sensitivity analysis. *Reliability Engineering & System Safety*, 79(2):121 – 122, 2003. SAMO 2001: Methodological advances and innovative applications of sensitivity analysis.

[275] C.S. Tarifa, P.P. del Notario, and F.G. Moreno. On flight paths and lifetimes of burning particles of wood. In *Tenth Symposium on Combustion, 1721 August 1964, Cambridge, UK*, pages 1021–1037. The Combustion Institute: Pittsburgh, PA, 1965.

[276] C.S. Tarifa, P.P. del Notario, F.G. Moreno, and A.R. Villa. Transport and combustion of firebrands. Technical Report Grants FG-SP-114, FG-SP-146, Instituto Nacional de Tecnica Aeroespacial, 1967.

[277] Douglas Thomas, David Butry, Stanley Gilbert, David Webb, and Juan Fung. The costs and losses of wildfires. a literature survey. Spec. Publ. 1215, Natl. Inst. Stand. Technol., U.S. Department of Commerce, October 2017. 72 pages, CODEN: NSPUE2, https://doi.org/10.6028/NIST.SP.1215.

[278] W. Thurston, J. D. Kepert, K. J. Tory, and R. J. B. Fawcett. The contribution of turbulent plume dynamics to long-range spotting. *Int. J. Wildland Fire*, 26:317–330, 2017.

[279] A. Tohidi, N. Kaye, and W. Bridges. Statistical description of firebrand size and shape distribution from coniferous trees for use in Monte Carlo simulations of firebrand flight distance. *Fire Safety Jornal*, 77:21–35, 2015.

[280] A. Tohidi and N. B. Kaye. Stochastic modeling of firebrand shower scenarios. *Fire Safety Jornal*, 91:91–102, 2017.

[281] Kevin Tolhurst, Brett Shields, and Derek Chong. Phoenix: Development and application of a bushfire risk management tool. *Aust. J. Emerg. Manage.*, 23(4):47–54, 2008.

[282] I. B. Troen and L. Mahrt. A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Boundary-Layer Meteorology*, 37:129–148, 1986.

[283] A. Trucchia, V. Egorova, A. Butenko, I. Kaur, and G. Pagnini. Randomfront 2.3: a physical parameterisation of fire spotting for operational fire spread models – implementation in wrf-sfire and response analysis with `LSFire+`. *Geoscientific Model Development*, 12(1):69–87, 2019.

[284] A. Trucchia, V. Egorova, G. Pagnini, and M.C. Rochoux. On the merits of sparse surrogates for global sensitivity analysis of multi-scale nonlinear problems: Application to turbulence and fire-spotting model in wildland fire simulators. *Communications in Nonlinear Science and Numerical Simulation*, 73:120 – 145, 2019.

[285] A. Trucchia, M.R. Mattei, V. Luongo, L. Frunzo, and M.C. Rochoux. Surrogate-based uncertainty and sensitivity analysis for bacterial invasion in multi-species biofilm modeling. *Communications in Nonlinear Science and Numerical Simulation*, 73:403 – 424, 2019.

[286] Andrea Trucchia, Vera Egorova, Gianni Pagnini, and Mlanie C. Rochoux. On the merits of sparse surrogates for global sensitivity analysis of multi-scale nonlinear problems: application to turbulence and fire- spotting model in wildland fire simulators. *Communications in Nonlinear Science and Numerical Simulation (submitted)*, 2018.

[287] Andrea Trucchia and Gianni Pagnini. Restoring property of the michelson-sivashinsky equation: A challenge for modelling. *Combustion Science and Technology – Submitted*, 2019.

[288] Andrea Trucchia, Federica Villa, Luigi Frunzo, and Gianni Pagnini. Seeding dispersal modeling for systems of planar microbial biofilms, 2017.

[289] G. Tryggvason, B. Bunner, A. Esmaeeli, D. Juric, N. Al-Rawahi, W. Tauber, J. Han, S. Nas, and Y.-J. Jan. A front-tracking method for the computations of multiphase flow. *Journal of Computational Physics*, 169(2):708 – 759, 2001.

[290] Tams Turnyi. Applications of sensitivity analysis to combustion chemistry. *Reliability Engineering & System Safety*, 57(1):41 – 48, 1997. The Role of Sensitivity Analysis in the Corroboration of Models and its Links to Model Structural and Parametric Uncertainty.

[291] C. Tymstra, R.W. Bryce, B.M. Wotton, S.W. Taylor, and O.B. Armitage. Development and structure of Prometheus: the Canadian wild land fire growth simulation model. Information Report NOR-X-417, Canadian Forest Service, Northern Forestry Centre, 2010.

[292] Salih Ozen Unverdi and Grtar Tryggvason. A front-tracking method for viscous, incompressible, multi-fluid flows. *Journal of Computational Physics*, 100(1):25 – 37, 1992.

[293] Nathan M. Urban and Thomas E. Fricker. A comparison of latin hypercube and grid ensemble designs for the multivariate emulation of an earth system model. *Computers & Geosciences*, 36(6):746–755, 2010.

[294] Anna Katrine Vangsgaard, Miguel Mauricio-Iglesias, Krist V. Gernaey, Barth F. Smets, and Grkan Sin. Sensitivity analysis of autotrophic n removal by a granule based bioreactor: Influence of mass transfer versus microbial kinetics. *Bioresource Technology*, 123:230–241, 2012.

[295] Dimitri Vaynblat and Moshe Matalon. Stability of pole solutions for planar propagating flames: Ii. properties of eigenvalues/eigenfunctions and implications to stability. *SIAM Journal on Applied Mathematics*, 60:703–728, 2000.

[296] D.X. Viegas, editor. *Forest Fire Research & Wildland Fire Safety*. Millpress, 2002. Proceedings of IV International Conference on Forest Fire Research / 2002 Wildland Fire Safety Summit, Coimbra, Portugal, November 18-23, 2002.

[297] F. Villa, F. Cappitelli, P. Cortesi, and A. Kunova. Fungal biofilms: Targets for the development of novel strategies in plant disease management. *Front. Microbiol.*, 8:654, 2017.

[298] Federica Villa, Philip S. Stewart, Isaac Klapper, Judith M. Jacob, and Francesca Cappitelli. Subaerial biofilms on outdoor stone monuments: Changing the perspective toward an ecological framework. *BioScience*, 66:biw006, 02 2016.

[299] R. Wadhwani, D. Sutherland, A. Ooi, K. Moinuddin, and G. Thorpe. Verification of a Lagrangian particle model for short-range firebrand transport. *Fire Safety Jornal*, 91:776–783, 2017.

[300] David J. Wales and Jonathan P. K. Doye. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[301] T. S. Walker, H. P. Bais, E. Deziel, H. P. Schweizer, L. G. Rahme, R. Fall, and J. M. Vivanco. Pseudomonas aeruginosa-plant root interactions. pathogenicity, biofilm formation, and root exudation. *Plant Physiol.*, 134:320–331, 2004.

[302] H. H. Wang. Analysis on downwind distribution of firebrands sourced from a wildland fire. *Fire Technol.*, 47:321–340, 2011.

[303] O Wanner, Hermann Eberl, Eberhard Morgenroth, Daniel Noguera, Cristian Picioreanu, Bruce Rittmann, and Mark van Loosdrecht. *Mathematical Modeling of Biofilms*, volume 18. 04 2006.

[304] O. Wanner and W. Gujer. A multispecies biofilm model. *Biotechnology and Bioengineering*, 28(3):314–328, 1986.

[305] Oskar Wanner and Peter Reichert. Mathematical modeling of mixed-culture biofilms. *Biotechnology and Bioengineering*, 49(2):172–184, 1996.

[306] MS Waterman and DE Whiteman. Estimation of probability densities by empirical density functions. *International Journal of Mathematical Education in Science and Technology*, 9(2):127–137, 1978.

[307] F.A. Williams. *The Mathematics of Combustion*, chapter Turbulent Combustion, pages 97–131. SIAM, Philadelphia, 1985.

[308] A.A.G. Wilson and Ian Ferguson. Predicting the probability of house survival during bushfires, victoria, australia. *Journal of Environmental Management*, 23:259–270, 01 1986.

[309] M-K.H Winkler, K.F. Ettwig, T.P.W. Vannecke, K. Stultiens, A. Bogdan, B. Kartal, and E.I.P. Volcke. Modelling simultaneous anaerobic methane and ammonium removal in a granular sludge reactor. *Water Research*, 73:323–331, 2015.

[310] J. P. Woycheese, P.J. Pagni, and D. Liepmann. Brand propagation from large-scale fires. *J. Fire Prot. Eng.*, 10:32–44, 1999.

[311] Jack Xin. Front propagation in heterogeneous media. *Society for Industrial and Applied Mathematics*, 42:161–230, 06 2000.

[312] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, 2010.

[313] D. Xiu and G.E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[314] R. Yondo, E. Andrés, and E. Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 96:23–61, January 2018.

[315] B. P. Zeigler. Hierarchical, modular discrete-event modelling in an object-oriented environment. *Simulation*, 49:219–230, 1987.

[316] Bernard P. Zeigler. Hierarchical, modular discrete-event modelling in an object-oriented environment. *Simulation*, 49(5):219–230, November 1987.

[317] Dawei Zhang, Weilin Li, Xiaohua Wu, and Tie Liu. An efficient regional sensitivity analysis method based on failure probability with hybrid uncertainty. *Energies*, 11:1684, 06 2018.

[318] Tianyu Zhang, N. G. Cogan, and Qi Wang. Phase field models for biofilms. i. theory and one-dimensional simulations. *SIAM Journal of Applied Mathematics*, 69(3):641–669, 2008.

[319] V. L. Zimont. Gas premixed combustion at high turbulence. turbulent flame closure combustion model. *Exp. Therm. Fluid Sci.*, 21:179–186, 2000.

[320] V. L. Zimont and A. N. Lipatnikov. A numerical model of premixed turbulent combustion of gases. *Chem. Phys. Reports*, 14:993–1025, 1995.

[321] C.E. Zobell and D.Q. Anderson. Observations on the multiplication of bacteria in different volumes of stored sea water and the influence of oxygen tension and solid surfaces. *Biological Bulletin*, 71:324–342, 1936.