UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Inverse modelling in wildfire spread forecasting: towards a data-driven system

Ph.D. Thesis

Oriol Rios i Rubiras

Ph.D. advisors:

Dra. Elsa Pastor i Ferrer

Dra. Eulàlia Planas i Cuchi

**Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
Programa de Doctorat en Enginyeria de Processos Químics**

Barcelona, September 2018

SÒCRATES

He encès un foc
que no sé on crema.
He pagat una casa
que no sé on és.
He corregut una cursa
que no sé quan acaba.

Viure sense saber, saber viure

Núria Domènech i Amador,
AFONIA, Ed. Viena, 2016

Dedico aquesta tesi
a la meva mare, Montse
i al meu pare, Paco.

Agraïments

Escriure una tesi és una mica com una expedició de muntanya. L'objectiu de partida és clar; fer cim. Però a la pràctica, i n'ets conscient des del principi, el cim és lluny, i només l'intueixes quan ja hi ets molt a prop. Saps que hi haurà dies de fred que costarà sortir de la tenda, moments de soledat, que a voltes creuràs que no pots més, que viuràs petites victòries..., però el que és important és anar pujant, pas a pas, amb totes les forces.

Pel camí trobaràs algunes traves addicionals, sorpreses inesperades i sobretot, gent amb qui compartiràs un xic de te calent o una mica de via. I t'ajudaran, i fins i tot, pot ser que obrin traça una estona. Així doncs, vull fer esment a tota aquesta gent, que des de la plana fins a l'aresta, des de passar-te la méteo fins a tirar de la corda, han contribuït, indispensablement, a assolir aquest cim.

Primer de tot, vull agrair a l'Eulàlia i a l'Elsa, haver-me ofert l'oportunitat de fer un doctorat sobre foc, i fer-lo a casa. Certament, les condicions el 2011 no eren gens propícies i l'esporgada pressupostària en ciència causà estralls. Us dono les gràcies per haver confiat en mi des del principi i haver-me obert les portes de la docència i la recerca en el camp del foc a Catalunya.

I would like to thank Prof. Bart Merci, Prof. Guillermo Rein and the entire IMFSE program (lecturers, students and colleagues). This was my entrance gate to the Fire Science field and it is clear that if I had not been thrilled by this scientific domain, I would not have pursued this PhD.

Gracias también a Wolfram, por acogerme en una estadía en Santiago, en las antípodas de donde nos conocimos, y ayudarme a sacar muy buen provecho de unos meses de programación intensa en una oficina con inmejorables vistas a la cordillera. Gracias también a Léo, Carlos, Maria, Toño y Munita por hacerme descubrir en 3 meses las profundidades Santiagueñas y las alturas Andinas.

Vull agrair també a tots els CERTECUS: en Miguel, l'Ari, la Diana, en Xavi, en Christian, en Borja, en MMiguel i en Joaquim. Perquè el bon ambient del dia a dia és el que et permet realment avançar.

A tots els FísicsPelMón, per fer equip i convertir les nostres trobades en quelcom més que interaccions gravitatòries. Al Xavier, al Dani, a l'Eloi a l'Ula i al Pol, per ajudar-me a mantenir un peu a la física amb cafès a Palau Reial. A la Nú, per les converses dolces i les amargues, sempre massa breus però amb llarg regust.

A la família Clapés, a tots set, per il·luminar les nits Barcelonines amb una taula parada. Sentir-vos a prop és fer sempre costat.

Als Cintus: la Magalí, l'Ares i en Yago, perquè certament hem estat una segona família durant més de mig doctorat. Conviure és una acció del dia a dia, i les envolades discussions dels ressopons intra-setmanals, responsables de més lleganyes matutines del compte, quedaran sempre enganxades a la pell, com les engrunes a les estovalles.

Al Parera i al Martí, per ser uns companys de cordada, vosaltres literals, que ompliu de sentit el sortir a gaudir de les muntanyes. Sense oxigenar les neurones amb dosis de pedra, gel i neu, aquí no hauria sortit res.

En rapprochant le bout, je veux remercier mes amies du Pays de Gex; Nico, Elli, Rafa, Laura, Art, Gaud, Jon et Lucie. Vous m'avez accompagné et motivé à mettre fin à ce que j'avais presque fini lorsque nous nous sommes rencontrés dans ce beau coin. Merci pour l'accueil dans cette découverte active et l'accompagnement dans les derniers pentes pour gravir ce sommet.

A la Mari, per passar al davant quan flaquegen les forces, per haver patit i gaudit junts aquesta, i tantes altres, ascensions.

I finalment, gràcies a la meva família pel suport més que incondicional. Al Roger, per fer de puntal, atent i sempre disponible; al Bernat per capgirar els rols i ensenyar-me tant, malgrat ser el petit. A la meva mare Montse, per l'amor i l'ajuda que només poden donar les mares i al meu pare Paco, per ser-hi sempre present.

Perquè l'alegria de fer cim és immensa, però en el fons, només ho és per tot el que has compartit fins a arribar-hi. Ítaca és sempre només l'excusa.

Thoiry, Setembre del 2018

Clarification

To write this thesis, the author has deliberatively given preference to the active voice to contribute to the text comprehension and readability. This choose is further motivated by recent journal's writing guidelines that favor the use of the active voice for reasons of clarity and conciseness.

Additionally, the author generally preferred to use the 1st person of the plural (i.e. *we*) to highlight the cooperative work and necessary assistance throughout the developing process of this thesis. In any case, the author does not share the authorship of this manuscript or any flaw contained in it.

Resum

Els incendis forestals són, al cap i a la fi, un fenomen ecològic inherent a la dinàmica de la terra i àmpliament escampats per tot el món. A més de l'impacte ambiental, quan els incendis forestals excedeixen, en magnitud i intensitat, la capacitat d'extinció, representen una amenaça per a propietats i vides humanes. En molts països, l'abandó rural de les últimes dècades, el creixement de la continuïtat forestal i l'augment de la interfície urbana-forestal (Wildand-Urban Interface) està comportant l'augment de comunitats exposades al foc forestal alhora que les fan més vulnerables a un gran incendi. A més a més, l'escalfament global sembla que està afavorint i facilitant la recurrència de les condicions climàtiques propícies pels incendis forestals. El problema de l'incendi forestal s'està convertint en un perill natural recurrent i repetitiu que clama avanços urgents en recerca, planificació i gestió per tal de minimitzar-ne el seu impacte.

En aquest context, la present tesi se centra en el desenvolupament, la implementació i la validació inicial d'un model de predicció de la propagació del perímetre d'incendis forestals que podria ajudar als responsables de l'emergència a prendre decisions més oportunes per a emprar els recursos de forma eficient tot protegint els actius valuosos. Aquest model predictiu és una implementació particular d'un sistema basat en dades. És a dir, les dades disponibles s'utilitzen per a millorar i calibrar els resultats del model de propagació del front amb l'objectiu de proporcionar un pronòstic més precís i oportú de l'evolució del perímetre del foc en les pròximes hores.

Aquesta tesi construeix el sistema esmentat pas a pas, tot incrementant-ne la seva complexitat i abordant les millores i adaptacions necessàries en cada etapa, per exemple, en la caracterització del combustible i la projecció i interacció del vent amb la topografia. Inicialment, s'utilitzen imatges infraroges de l'evolució del perímetre (isòcrones) de dos cremes experimentals dutes a terme a Austràlia per a realitzar una prova de concepte del sistema. Malgrat el resultat favorable d'aquesta primera investigació, s'identifiquen alguns avenços per millorar-ne l'efectivitat i permetre'n l'aplicació en incendis reals. Així, els capítols següents se centren en els submodels de combustible i vent juntament amb l'actualització topogràfica del model de propagació i els diferents algorismes i estratègies matemàtiques necessàries per dur a terme el procés d'assimilació basat en dades.

Pel que fa als combustibles, la tesi presenta una anàlisi en profunditat de la caracterització dels combustibles que han d'utilitzar els models de propagació de foc. Això es fa mitjançant una anàlisi minuciosa de la sensibilitat dels paràmetres de caracterització del sistema més utilitzats. A la llum d'aquests resultats, es proposa un model simplificat que integri totes les propietats de combustible diferents per ser utilitzat pel model predictiu desenvolupat en la present tesi.

Per resoldre adequadament la interacció del vent amb el terreny i acoblar-la al model de propagació bastat en dades, s'utilitza el programa de diagnòstic WindNinja. Els temps necessaris de computació, però, no permeten la seva directa integració en una estratègia d'assimilació de dades. Així doncs, en aquesta tesi, es desenvolupa i valida un marc interpolador que permeti actualitzacions ràpides i computacionalment assequibles del camp de vents a nivell topogràfic. Aquest element clau es converteix en una peça clau per aconseguir el model de propagació basat en dades que es cerca en aquesta tesi. Per al procés d'optimització (present en qualsevol model conduit per dades) es comparen i s'avaluen sis algorismes matemàtics diferents. Tres d'ells són estratègies de cerca basades en programació lineal i les altres tres són estratègies de recerca global. L'exploració conclou que la selecció d'algorismes té un gran impacte en els resultats finals en termes de la precisió de pronòstic i temps de computació.

Finalment, tot el sistema es verifica i valida utilitzant les dades de dues fonts disponibles: (1) incendis experimentals de mitjana escala realitzats a Portugal en un pendent homogeni ben caracteritzat i (2) amb fronts generats sintèticament que reprodueixen un incendi real a gran escala. Aquestes validacions estan orientades a estudiar el rendiment general, comprovar la funcionalitat del sistema així com ressaltar possibles defectes i millores necessàries per tal de poder utilitzar l'eina en una situació d'emergència real. Malgrat els resultats mostren els potencials del sistema tot proporcionant un pronòstic acceptable, utilitzable com a eina de suport per a la gestió d'una emergència, queda també palès que es requereixen més validacions per comprovar la robustesa i fiabilitat de l'eina abans d'utilitzar-la en situacions operatives.

Abstract

Wildfires are an ecological phenomenon inherent to earth dynamics and widely spread over the globe. In addition to the environmental impact, when wildfires grow beyond controllable magnitudes, they pose a principal threat to human lives and properties. On many countries, the rural abandonment of last decades, the forest continuity growth and the Wildland Urban Interface increase are exposing entire communities and rendering them vulnerable to a major fire event. Coupled together with a global warming that seems to be enlarging and worsening wildfire-prone weather conditions, the wildfire problem is becoming a recurrent and repetitive natural hazard that is in urgent needs of research development, planning and organizational changes to minimize its impact.

In this context, the thesis at hand focuses on the development, implementation and initial validation of a wildfire perimeter spread prediction model that might help emergency responders on taking sound decisions to efficiently employ resources and protect valuable assets. This forecasting model is a particular implementation of a data-driven system. That is, available data are used to improve and calibrate the spread model results with the aim of delivering a more accurate and timely forecast of the fire spread for the upcoming hours.

This thesis builds up the mentioned system by increasing its complexity and tackling required improvements and adaptations on fuel characterization and wind projection on topography. Initially, a simplified proof of concept that uses front perimeter (isochrones) evolution extracted from infrared imagery of the fire is challenged with data from real-scale burning experiments conducted in Australia. Despite the positive outcome of this initial investigation, some advancements are identified to further upgrade the system. Thus, the following chapters focus on the fuel and wind sub-models together with the spread model topographic upgrade and the different mathematical algorithms and strategies necessary to conduct the data-driven process.

Regarding fuels, the thesis presents an in-depth analysis of fuel characterization to be used by fire spread models. This is done by a thorough sensitivity analysis of the most commonly used fuel characterization systems. In the light of these results, a simplified model that integrates all different fuel properties is proposed to be used by the data-driven framework at hand.

To properly resolve the wind interaction with the terrain and to couple it into the data-driven system, the WindNinja diagnose software is employed. However, long computational times do not allow for its integration into any data-assimilation strategy. Thus, a full interpolating framework is developed and validated to allow fast and computationally inexpensive wind field updates. This key element becomes then a cornerstone of

the full data-driven approach. For the optimization process (embedded into any data-driven systems) six different mathematical algorithms were compared and evaluated. Three of them being line-search strategies and the other three being global search. It was found that the algorithm selection has an impact on the final results in terms of forecast accuracy and computing time.

Finally, the overall system is verified and validated using two source of available data: (1) well characterized, homogeneous slope, medium-scale experimental fires conducted in Portugal and (2) with synthetically generated fronts reproducing a real large-scale fire. These validations were aimed at studying the overall performance, checking the system functionality and highlighting possible flaws and necessary improvements if the tool is to be deployed in a real emergency situation. Whereas the results show the potentials of the approach by delivering an acceptable forecast usable for emergency responders, further validations are required to check the robustness and reliability of the tool before using it in operational situations.

Contents

List of Figures

List of Tables

Nomenclature

ANOVA	Analysis of Variance
AD	Automatic Differentiation
AE	Absolute Error
APE	Absolute Percentage Error
AW	assimilation windows
BBOF	Black-Box Optimization Framework
Cat	Categories standard fuel models
DA	Data Assimilation
DEM	Digital Elevation Model
GCP	Ground Control Points
GIS	Geographic Information System
GLUE	Generalized Likelihood Uncertainty Estimation
GPU	Graphics Processing Unit
GSA	Global Sensitivity Analysis
HPC	High-Performance Computing
Idty	Identity
IGCC	Institut Geogràfic i Cartogràfic de Catalunya
IR	Infra-Red
KKT	Karush-Kuhn-Tucker
lat	Latitude
LH	Live herbaceous.
lon	Longitude
LW	Live Woody
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MPI	Message Passing Interface
NWP	Numerical Weather Prediction
OpenMP	Open Multi Processing
PE	Percentage Error
Res	Resolution
RoS	Rate of Spread
SAV	Surface-area-to-volume ratio
SFM	Standard Fuel Model
TLM	Tangent Linear Method
USGS	Unites States Geological Survey
WUI	Wildland-Urban Interface

Variables

A	burnt area [m2]
a	lateral firelet propagation velocity [m s^{-1}]
α	adjacent angle between two nodes [rad]
b	firelet partial front velocity, starting at the center of the firelet [m s^{-1}]
β	fuel packing ratio [-]
c	firelet partial front velocity, starting at the ignition point [m s^{-1}]
d	Euclidean distance [m]
δ	fuel depth [m]
$\boldsymbol{D}$	set of wind directions
D	principal wind direction
D_b	wind direction base map
D_I	interpolated principal wind direction map
D_s	spatial threshold value [m]
D_a	angular threshold value [deg]
ϵ	effective heating factor [-]
ϵ	relative error [%]
ϵ_{abs}^{area}	absolute areal error [m2]
ϵ_r^{area}	relative areal error [%]
χ	propagating flux ratio [-]
f	fuel type
F	proportionality factor
Phi_s	Slope factor [-]
Phi_w	Wind factor [-]
h	fuel height
$\mathbb{I}$	identity matrix
I_R	reaction intensity [J m^{-2} s^{-1}]
I_U	wind speed invariant [m s^{-1}]
I_θ	wind direction invariant [rad]
I_{mfw}	fuel-wind invariant [1 s^{-1}]
$\mathbb{J}$	cost function
$\mathbf{M}$	forward model
M	reduce model characteristic moisture content
M_f	moisture content [%]
M_x	moisture of extinction [%]
m_x	moisture content for a given particle class x [%]
m_i	moisture content for a given particle class i [%]
N	number of isochrones
n	number of front nodes
p	fire front node

Q_{ig} heat of pre-ignition [J kg^{-1}]

q   parameters vector

RoS rate of spread [m s^{-1}]

ρ_b bulk density [kg m^{-3}]

SAV surface-area to volume ratio [m^{-1}]

S reduced model characteristic SAV

s front discretization parameter [-]

s_x surface-area-to-volume fraction for a given particle class x [m^{-1}]

t simulation time [s]

T digital terrain model

U   mid-flame wind speed [m s^{-1}]

$\boldsymbol{U}$ set of wind speeds

U principal wind speed

U_b wind speed base map

U_I interpolated principal wind speed map

W   weighting matrix

W, Wo reduced model characteristic fuel load [kg m^{-2}]

$\boldsymbol{W}$ wind maps matrices (direction and magnitude)

W_O oven-dry fuel loading [kg m^{-3}]

x set of front coordinates

θ slope-wind direction [rad]

Part I

Thesis

Chapter 1

Introduction

Wildfires are a global phenomena inherent to earth dynamics. Their development and impact, however, are closely related to human activity and the capacity to create resilient structures to mitigate their thread. A principal way to built up those structures is to foresee the wildfire occurrence and predict its behaviour. In this chapter the complex nature of wildfire is briefly introduced as a way to understand the current existing approaches and limitations to simulate its behaviour. Data assimilation and inverse modelling are presented as an alternative methodology to potentially fill the knowledge gap of current existing models and develop a reliable operational forecasting tool.

1.1 Wildfire phenomena

Wildfire exists since the appearance of terrestrial plants (Bowman et al., 2009) as those provided the fuel and the oxygen necessary to produce combustion. First wildfire evidences, in the form of charcoal fossils, are dated from 420 to 440 million years ago (Pausas and Keeley, 2009; Glasspool et al., 2004) when oxygen arose beyond 13% concentration in the atmosphere enabling flammable combustion. Since then, wildfire has manifested in multiple forms and ecosystems ranging from smouldering fires to crown fires.

Over the last decades, the direct yearly wildfire suppression cost exceeded 2.300 M \$ only in US federal lands and Canada (NIFC, 2013; Georg et al., 2013). When the economic, personal and environmental losses are added, the figure is dramatically increased. A single event such the recent Fort McMurray wildfire in Canada in 2016 produced an expenditure of 3.500M\$ (Statistics Canada, 2016), making it the costliest insured event in the history of Canada (Statistics Canada, 2016).

Despite being an ancient phenomenon wildfire consequences still threaten societies and endanger ecological systems while the human capacity to control them is still limited (Georg et al., 2013). Moreover, wildfires have recently shown the potential to reach colossal dimensions that were never seen before. Fires in the Victoria region in Australia, 2009 Cruz et al. (2012), or Colorado, USA, 2012, as examples, have led to a new scientific term; *mega-fires* Williams (2013), used to describe the large, highly-destructive wildfire that grows beyond any currently existing extinguishing capacity.

1.1.1   Wildfire underlying physics

From the physics point of view, a wildfire is a multi-scale phenomenon that spans over several order of magnitudes in both spatial and temporal dimensions. Starting from smallest scales, the pyrolysis process (of the order of mm and ms) is the responsible of transforming solid fuel into gas, which, after mixing with oxygen, will be combusted. This pyrolysis is driven by the energy balance. As it is an endothermic reaction, the supply of energy, by means of any of the three heat transfer modes; conduction, convection or radiation, will control the gaseous combustible production rate. The flames produced by combustion can range from mm to tens of meters. Beyond the flames, the smoke plume that is responsible of transfer of heat and firebrands (both closely related to fire spread), can extend to the order of kilometres. The range of these scales, in order of magnitudes, for the principal involved processes are gathered in Table 1.1. Besides those phenomena, the external conditions as terrain shapes and weather characteristics (relative humidity, air temperature, wind speed and direction, etc.) also play a principal role on the wildfire behaviour. Their relevant scales go from local wind burst and terrain heterogeneities to mesoscale weather situation or general slope characteristics. Moreover, it is not only that those scales are all important to explain the wildfire phenomena, but they are also highly coupled. A paradigmatic example is the potential influence of a wildfire to the local weather conditions being able, in some cases, to drastically modify them.

Despite the description of the principal physical mechanisms is available in the literature and can be directly taken from other research disciplines, there is a lack of a unifying theory that explain all the coupling effects and interactions within the wildfire field. In fact, Finney et al. (2013) state (in the articles entitled *On the need for a theory of wildland fire spread*) that there is an urgent need for a sound theory to unify and explain all those phenomena. After highlighting some examples of wrong assumptions made when trying to model wildfires, they conclude that a fundamental theory of wildland fire spread is missing. Such a theory should set the basis of fire spread based on physical principle, but resolving all processes, from ignition mechanisms to live fuel

TABLE 1.1: Order of magnitude range of the spatial and temporal scales involved in a wildfire. Adapted from Sullivan (2009a)

Type	Time scale [s]	Vertical scale [m]	Horizontal scale[m]
Combustion reactions	$10^{-20} - 10^{2}$	$10^{-4} - 10^{-2}$	$10^{-4} - 10^{-2}$
Fuel particles	–	$10^{-3} - 10^{-2}$	$10^{-3} - 10^{-2}$
Fuel complex	–	$10^{0} - 10^{1}$	$10^{0} - 10^{2}$
Flames	$10^{-2} - 10^{1}$	$10^{-1} - 10^{1}$	$10^{-1} - 10^{1}$
Radiation	$10^{-9} - 10^{1}$	$10^{-2} - 10^{1}$	$10^{-2} - 10^{1}$
Conduction	$10^{-2} - 10$	$10^{-2} - 10^{1}$	$10^{-1} - 10^{1}$
Convection	$10^{0} - 10^{2}$	$10^{-1} - 10^{2}$	$10^{-1} - 10^{1}$
Turbulence	$10^{-1} - 10^{3}$	$10^{0} - 10^{3}$	$10^{0} - 10^{3}$
Spotting	$10^{0} - 10^{3}$	$10^{0} - 10^{3}$	$10^{0} - 10^{5}$
Plume	$10^{0} - 10^{5}$	$10^{0} - 10^{5}$	$10^{0} - 10^{2}$

combustion. Without advancing on this fundamental theory, model engineering will be limited to simulate known problems in narrow ranges of conditions.

To support this statement, Finney's group focused on studying fundamental questions and revising all the assumptions in the field. Results of their work showed that convective heat transfer played a more important role than radiation heat transfer on fire spread as it is a more efficient way heating up a thin fuel particle as there is no convecting cooling (Finney et al., 2015). Additionally, they identified and reported buoyancy-inertia–induced flame bursts structures (see Figure 1.1) that where closely coupled with this convective ignition and, thus, with fire spread.

1.1.2   Propagation modes

Wildfires can exhibit different dynamics depending on the fire spread driving factor, such as fuel or weather conditions for example. One way of classifying them is regarding the principle combustion mechanisms, that is, flame and flameless combustion. Even though one might be the driving force of a given wildfire, usually a combination of those are found.

Flameless fires (normally known as smouldering fires) are driven by solid-phase combustion and, despite having less visual impact that flaming fires, they are considered to be the longest burning fires on earth (Rein, 2013, 2015). Indeed, the *Burning Mountain* in Australia is estimated to have been burning continuously during more than 5500 years (Ellyett and Fleming, 1974). Due to they large extension, duration and heat released, peat fires (one of the most common flameless fires) can burn 100 times more mass than flaming fires (Davies et al., 2013) and they are responsible of the 15% of greenhouse gas emission worldwide (Rein, 2015). Although they show slower spread velocities than flaming fires (Rate of Spread of the order of 1 mm min^{-1}), they are difficult and arduous
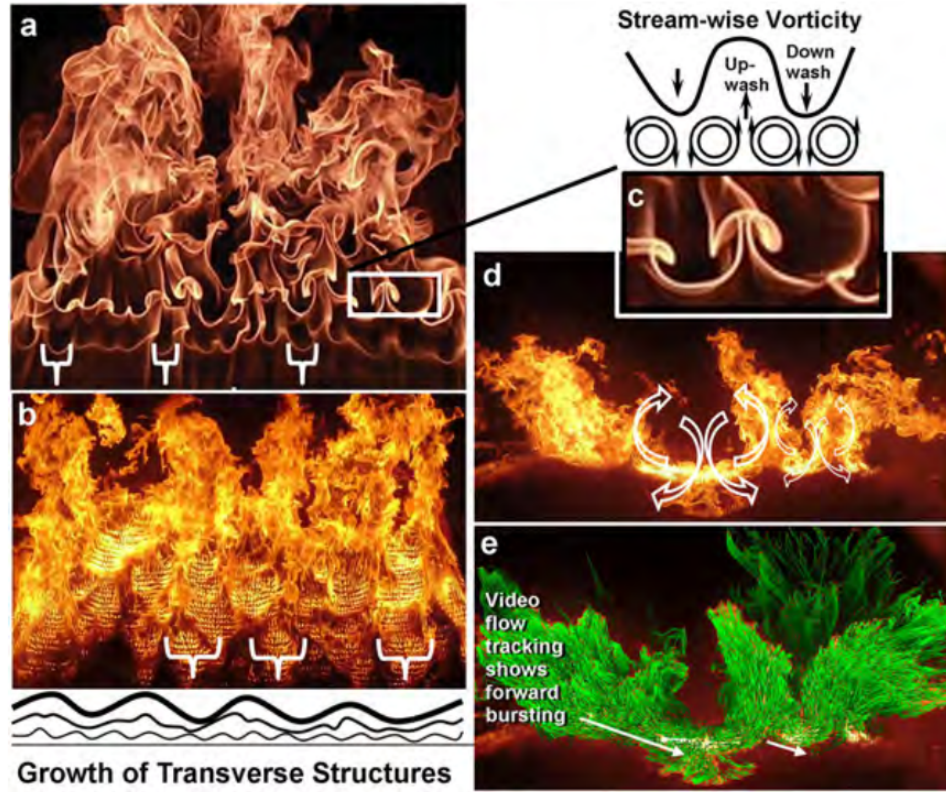
FIGURE 1.1: Example of buoyancy-inertia–induced flame bursts. (a) is a stationary ethylene gas burner and (b) a wind-tunnel experimental fire spreading in cardboard fuel. Longitudinal vortices (c) induce flame peaks and troughs at up-wash and down-wash convergence zones. The animated images of the leading edge of spreading fire in plot (d) and its flow-tracking analysis (e) show flame vortex circulations and forward flame bursts through flame troughs. Source: Finney et al. (2015)

to extinguish. The reason is the impediment to clearly identify the front location (most of the times, the only visible sign is smoke and haze) and, even when the front can be identified, it is difficult to access it (several meters underground) and to create physical barriers to stop it. Usually, only an extensive and generalized precipitation period can definitely put it out. In the wildfire operational field this propagation mode is usually called ground fires. This type of fires can have large ecological impacts (as in *Tablas de Daimiel* national park Moreno et al. (2011), see Figure 1.2, for example) and impact on life safety due to the large amount of toxic haze produced.

FIGURE 1.2: Peat fire smoking at *Tablas de Daimiel* National Park. Source: Mendez (2009)

However, the most extensively known form of wildfire propagation mode is the flaming combustion. Within this type, several propagating forms are commonly identified. The most standardized ones are surface fires and crown fires. The first type are characterized to burn litter, duff, grasses and shrubs combustible types, that is, the lower part of the canopy cover. Due to their limited intensity, the fire does not spread to the upper part of the trees and thus, they cause the least damage to the forest. On the contrary, crown fires (see Figure 1.3) involve all the canopy. This phenomenon produces high intense fires whose rate of spread can exceed 10 km h^{-1}(Cruz et al., 2013). This types of fires pose great challenges to fire fighters as the front heat intensity overwhelms the extinguishing capacity. Thus, they can produce major impact on environment and particularity threaten life and properties when impacting a Wildland-Urban Interface (see section 1.1.4).

Finally there is a propagation mode, that does not use direct heat transfer mechanisms but the lofted incandescent particles (called embers or firebrands) that are transferred far from the fire front (see Figure 1.4). This mechanism is called spotting and is normally linked to crown fires as high intensity is required. Due to its dynamics (solid lofting, solid-state combustion, transient ignition, etc.), and the difficulty to reproduce it in a controlled environment, this phenomena is complicated to study and remains one of the less characterized wildfire events. The usual spotting distance range from several meters to few kilometres for fully developed fires. The maximum spotting distance reported was of about 30 km (Australia (Koo et al., 2010)).

In the operational field, beyond differentiating the fires by the type of combustion (flaming vs flameless) they are normally identified regarding the principal propagation

(a)   (b)

FIGURE 1.3: Two images of a fully developed crown fire in Òdena, 2015. Source: Bombers de la Generalitat de Catalunya

mechanism. The two principal ones are, wind-driven and topographic driven fire. This classification is useful to foresee change on the fire dynamics. For example, if wind changes dramatically in a wind-driven fire, it will have an important impact to the spread. Similarly, the front arrival on a ravine enclave or at a ridge will have different effects on the propagation for a topographic fire than for a wind-driven one. Within the topographic fire type, there is a particularly dangerous sub-case called eruptive fire (Viegas and Simeoni, 2010). More than a particular propagation driver, it could be seen as an extreme behaviour of a topography driven fire. Although researches agree that there exists some conditions that favours and eruptive fire (such as canyon alike terrain), further investigation and devoted studies are still needed to fully understand and clarify this phenomena that has caused major fatal accidents to emergency responders during lasts decades (Viegas and Simeoni, 2010).

FIGURE 1.4: Spotting propagation example. The principal fire (on the right) has launched firebrands that caused two new fires hundreds of meters ahead the fire front. Source: Mendez (2009)

1.1.3 Wildfire and global warming

As global warming has become an irrefutable fact, its effect to wildfires are already being studied and reported. Jolly et al. (2015) have studied the global wildfire danger variation from 1979 to 2013. They used three global climate data sets and three danger indicators to study the fire seasons trends during this period. Results show that fire seasons duration is increasing worldwide. Moreover, the burnable area exposed to large fire weather seasons has doubled from the first half (1979-1996) with respect to the second half (1996-2013) of the studied period. In general trends, this change in the fire season distribution affects principally areas of eastern Africa, Asia and Mediterranean shore, west US and eastern south America (see Figure 1.5). With those results, they conclude that if those wildfire danger indicators correlate with ignition sources, wildfire events will have a major impact on ecosystems, societies, economies and even climate.
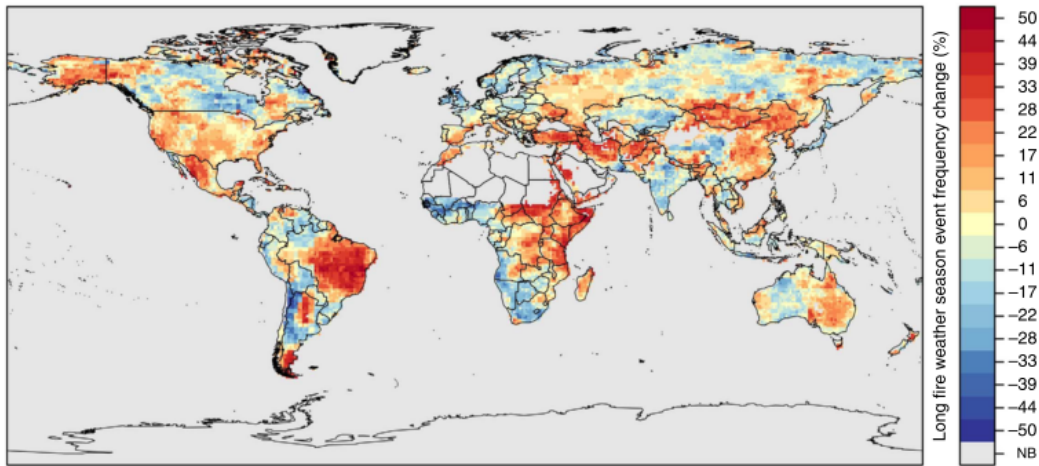
FIGURE 1.5: Regions that have experienced changes in the frequency of long fire weather seasons ($> \sigma$ above historical mean) from 1996-2013 compared to 1979-1996 period. Reds indicate areas where long fire weather seasons have become more frequent. Blues indicate areas where long fire weather seasons have become less frequent. Source: Jolly et al. (2015)

Georg et al. (2013) write a White Paper called *Vegetation Fires and Global Change* directed to the United Nations and International Organizations. More than 50 experts participated in the elaboration of this document by stating the current fire regimes around the globe and the probable change due to global warming. Some of the final conclusions could be synthesized as:

- Climate change research estimates and increases in the incidence and severity of forest fires, resulting in larger areas burned, shorter fire-return intervals, and a net loss of carbon to the atmosphere. These raises doubts over the economic sustainability and success of current fire management programs.

- In some regions (Canada, US, Mediterranean shore) the wildland-urban interface (WUI) is expanding what results in escalating property losses and evacuations, and bringing wildfire closer to homes.

- Forest health and productivity has suffered as effective suppression programs have led to a shift to older, more disease-prone forests with significant accumulations of dead woody fuel.

Beyond this final statements, the White Paper also describes the potential impact on ecosystems, population health, economics and society of an increasing wildfire regime worldwide to justify the final conclusion; there is the need to reduce the threats of destructive wildfires collectively and cooperatively.

1.1.4   Wildland-Urban Interface

The Wildland-Urban Interface (WUI) is defined as the intersection between forest and urban areas. It is the region where a potential wildfire will encounter urban structures (homes, industrial facilities, social infrastructures, etc.) and might cause an enormous impact on property protection and live safety.

Based on factors as rural areas abandonment, WUI area increase and higher recurrence of fire events that escape extinguishing capacities, it is generally believed in the wildfire community that fire impacting WUI will increase in both frequency and consequences. In this direction, Modugno et al. (2016) recently created the first European map of the extent of WUI areas (depicted in Figure 1.6) in order to analyse the risk of large forest fire impacting the WUI. Their analysis indicated that the presence of WUI areas is a common phenomenon in European countries, particularly in Central Europe and some Mediterranean areas. Additionally, they compared those WUI identified areas with wildfire historical data from 2002-2013 period to work out an estimation of the probability of having a wildfire at a given distance from the WUI region. The conclusion was that Mediterranean regions (including Cyprus, Albania and Bulgaria), are largely exposed to the WUI problem. Although this work represents the first systematic characterization of the WUI region, the authors also request a better definition of WUI zones in order to better exploit all current Geographic Information System (GIS) capabilities.
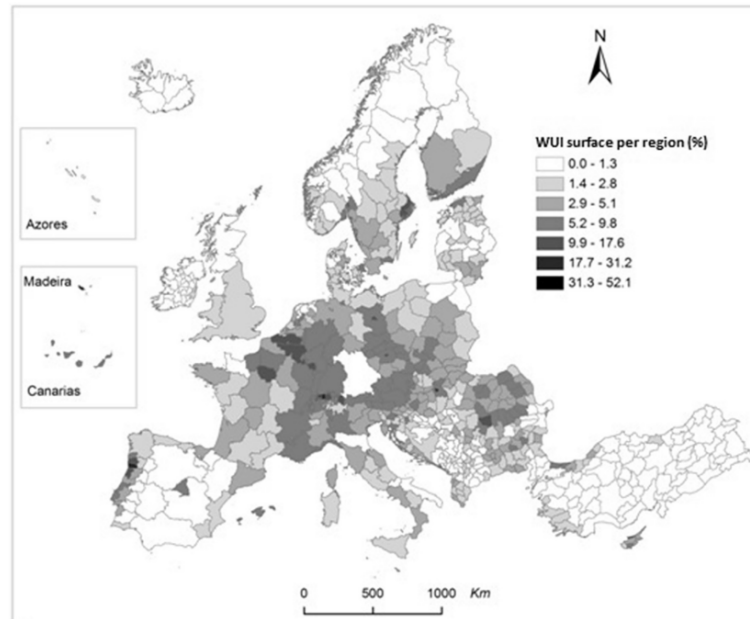
FIGURE 1.6: Europe WUI surface map divided in regional units. Source: Modugno et al. (2016)

Recent wildfire events impacting WUI with dramatic outcomes and numerous fatalities such as Victoria fires in 2009 in Australia (Cruz et al., 2012), Fort McMurray in 2016 in Canada (Westhaver, 2017), Madeira Island and Pedrógão Grande (Tonbul and Kavzoglu, 2017; De Sousa Dantas, 2017) in 2016 and 2017 in Portugal, (as example and among many others) fostered the interest and dedicated resources for the fire community to focus on WUI problem.

As WUI is a highly interdisciplinary problem, multiple approaches are taken to study it. NIST (National Institute for Standards and Technology) developed a Firebrand Generator (Manzello and Suzuki, 2013) to study how incandescent embers interact with urban structures and how could those be engineered in a more resilient way. In the same direction Zhou et al. (2014) study fundamentals of Firebrand transport by generating glowing firebrands showers and statistically quantifying the lofted firebrands size and mass under different wind conditions. Moving forward in this direction Caton et al. (2016) and Hakes et al. (2017) conducted and in-depth review analysis of the propagation ways within the WUI. The aim of their study was to detail the known pathways by which fires can spread into and within a WUI region. The idea behind was to highlight key aspects to reduce the pathways to ignition via hardening structures, communities, and surrounding wildland.

In another direction Ronchi et al. (2017), together with NFPA (National Fire Protection Association) thoroughly reviewed critical events where the fire impacting the WUI caused property and lives loses. They proposed an *Open Multi-Physics Framework for Modelling Wildfire Urban Evacuation* with the aim of setting the needs and the specifications of simulation tools capable of finding and securing safe escape routes to evacuate the WUI zones. After the study of potential models to be used, they proposed a research roadmap on the areas which require further analysis in the future.

1.2   Wildfire modelling

Wildfire discipline utilises multiple approaches and strategies to try to model and simulate the fire phenomena. They range from modelling the long term ecological impact of a wildfire to predicting the amount and type of pollutants emitted. In this thesis, we focus on the modelling of the fire spread (normally quantified by the Rate of Spread, RoS) and the fire perimeter evolution. The later (normally called fire perimeter or fire-line) is the macroscopic interface between unburnt and burnt combustible. It is, indeed, a macroscopic scale quantity, as if we scale it down to fuel particle elements (i.e. leaves or branches) it automatically loses all its practical interest.

There are several ways to classify the existing models; by the nature of the underlying equations, the variables studied, the propagation modes included (surface fires, crown

fires, spotting, etc.), the mathematical methodology used, the principal assumptions, weather they are based on empirical correlations or its applicability. Pastor et al. (2003) classified them regarding the nature of the underlying equations and the different propagation modes. Focusing on fire spread, Sullivan published the most extended review up to date (Sullivan, 2009a,b,c), and classified them upon: physical and quasi-physical, empirical and quasi-empirical and mathematical analogues. A comparison between applications, features, output capacities, advantages and disadvantages of the four first categories is laid down in Table 1.2.

The fifth category, mathematical analogues, comprises all models that are generally known as simulators. The reason is that almost all wildfire simulators are primarily aimed, (for obvious reasons) at computing the fire perimeter evolution over time. And thus, they need a mathematical strategy to propagate the front. Additionally, they can incorporate as sub-models elements of the previous categories. Papadopoulos and Pavlidou (2011) recently studied and compared 23 wildfire simulators (all existing ones they could find) to highlight their features, commonalities and differences. After discussing on their features they conclude (in line with Sullivan's conclusion) that FARSITE Finney (1998) and Prometheus Tymstra et al. (2010) where the most extendedly validated and over-performing ones.

In this section, we will provide a brief overview of four selected models, that could be candidates for the data assimilation system to be developed. The reader is directed to previously mentioned sources (Pastor et al., 2003; Sullivan, 2009a,b,c; Papadopoulos and Pavlidou, 2011) as well as Rochoux (2014) for an up-to-date and more in-depth review.

Due to the methodological approach that will be presented in chapter 2 and 5 we were interested in vector-based simulators. Those strategy, also known as a Lagrangian-based, keeps track of markers that define the front contour. This approach has been successfully applied on most broadly used fire simulators, FARSITE, Phoenix, or Prometheus (see below). The final front shape equivalence with the level-set alternative approach was recently demonstrated by Bova et al. (2015) over complex topographical domains.

- **PHOENIX** (Tolhurst et al., 2008) is a risk management model currently in use in Australia, that includes different tools to compute fire danger index, estimate fire impacts and simulate the fire spread. The fire spread is simulated with SiroFire (Coleman and Sullivan, 1996) a dynamic, spatially continuous, empirical fire characterization computational model now maintained and developed by Bushfire CRC. The fire spread is computed using Huygens' wavelet principle (Anderson, 1982) and implemented as proposed by Knight and Coleman (1993). PHOENIX also contains a deterministic firebrand transport and ignition module.

The model is aimed, calibrated and validated for large scale fire spread (Pugnet et al., 2013).

- **FIRESTATION** (Lopes et al., 2002) Is a fire spread simulator developed in Portugal that also uses the Huygens' principle for front expansion but applies it to a raster-based spread. That is, the propagated front is being projected into a gridded domain to account for active (burning) and inactive (non-burning) cells. It also accounted for the wind field by means of two selectable sub-routines, called Nuatmos or Canyon. It was partially validated with a real fire scenario providing good agreement between simulation and observations. To the best of this thesis author's knowledge, it is not in use.

- **PROMETHEUS** (Tymstra et al., 2010) is the current Canadian Wildland Fire Growth Simulation Model, produced and maintained by the Canadian Interagency Forest Fire Centre. It is a deterministic fire growth simulation tool that uses spatial fire behaviour input data on topography (slope, aspect and elevation) and a particular Canadian-based Fire Behaviour Prediction (FBP) fuel types, along with weather information layers. It uses a simple ellipse as the underlying template to shape fire growth. The model also simulates fire growth based on Huygens' principle of wave propagation, by implementing the differential equations derived by Richards (Richards, 1993; Richards and Bryce, 1995). The simulation code is written in Visual C++ and supports 2D and 3D graphical interface.

- **FARSITE** (Finney, 1998) is the most extensively used and validated wildfire spread simulator. It's surface fire spread is based on the Rothermel Rothermel (1972a) spread model coupled with Huygens wavelet expansions. Besides the surface fire spread it incorporates four additional sub-models: crown fire spread (Rothermel, 1991), fire acceleration, fuel moisture, and spotting from torching trees. The elliptical wave shape is determined locally at each vertex by the additive combination of wind speed and slope. Raster spatial data on fuels and topography is required for these calculations. Weather conditions can be obtained from weather stations databases, or can be simulated over the topography by means of an additional sub-routine called WindNinja Forthofer (2007).

TABLE 1.2: Overview and comparison of four different approaches to model wildfires. Adapted from Papadopoulos and Pavlidou (2011)

	Physical	Quasi-physical	Empirical	Quasi-empirical
Application	- Used for simulation purposes or to explore a particular physical process or aspect. - Applicable to medium-scale landscape domains, or to small-scale laboratory domains.	- Used in a laboratory-scale domain.	- Estimates the likely spread in the direction of the wind for suppression planning purposes. - Conducted in the form of simple calculations for plotting on a wall map.	- Particularly suited to faster than real time computation with low requirements for data.
Main feature	-Based on theory (both physical and chemical processes of the phenomenon). -Differential equations with numerical solutions used to describe the spread process. -Complex nature.	- Based on physics theory. - Lack of combustion chemistry and reliance on the heat release transfer. - Parameters determined by laboratory experiments.	- Based on observation and experiment and not on physics/chemistry theory. - Pragmatic nature. - Determination of the key characteristics used to describe the behaviour of the fire.	- Data gathered from experimental observation, statistically analysed using a physical framework for the functional relationships between dependent and independent variables.
Output	- Principally RoS. - Energy released from the fuel. - Radiation transfer to surrounding unburnt fuel and to the atmosphere. - 2-dimensional or 3-dimensional models of fire spread.	- RoS. - fire-line intensity. - flame geometry. - Radiation transfer to surrounding unburnt fuel and to the atmosphere. - l-dimensional models.	- RoS. - Rate of increase of the perimeter. - fire damage area. - Height, angle and depth of the flames. - l-dimensional models.	- RoS. - Convective heat output to rate of buoyancy production. - Ratio of combustion time to time characteristic of flame dynamics. - Fuel moisture content. - Geometry of the fire perimeter.
Advantages	-Analytical in nature. -Detailed fuel descriptions to better represent wildland fuels. -Ability for computer-based tests of the model.	-Rely upon a higher level model compared to physical models.	- Relatively straightforward implementation. - Direct relation to the behaviour of real fires.	- Alleviating the influence of the size of the dataset and personal choice.
Disadvantages	- Limited domain size due to computational requirements - Required detail in input data. Not suitable for landscape simulation or operational context.	- Not suitable for landscape simulation or operational context.	- Highly dependent on the conditions obtained by the source data. - Requires major approximations.	- Highly dependent on the conditions obtained by the source data. - Requires major approximations.

Note on theoretical models and CFD

The increase of computational power has fostered the implementation of physical models by means of computational fluid dynamics (CFD) strategies. Some examples of this strategy are FIRETEC (Linn and Harlow, 1998; Linn, 1997), FOREFIRE (Filippi et al., 2009, 2014b), WFDS-FDS (Mell et al., 2007; McGrattan and Miles, 2016). This promising approach has shown large potential, specially for the study of particular problems such as the Wildland-Urban Interface (Overholt et al., 2012; Fanlo et al., 2016). Intense efforts are being made to turn them into operational tools. As an example, a hybrid mixed of CFD and semi-empirical model, that couples 3D weather modelling with Rothermel's fire spread (Rothermel, 1972a) named WRF-SFIRE (Mandel et al., 2011, 2014) has recently been reported to be applied operationally, as a 6 hours forecast could be launched with a 30 minutes run in a multi core processor.

However, for the moment, those simulators are generally restricted to research use and small scale applications due to the high computational costs and required initializing data (Viegas, 2011).

1.2.1   Data-driven systems in wildfire modelling

A way to reduce wildfire impact is to improve and optimize wildfire fighting capacities by providing the emergency responders with sound data on the upcoming wildfire dynamics. However, detailed models (such us CFD based, for example) are not suitable for operational purposes due to the amount of data required to initialize them and the computational time requirements.

Thus, at present, only simplified semi-empirical models are available to be used for such endeavour. However, those simplified models have a common generalised drawback: they need to be calibrated and adapted to each different scenario. As an illustrative example, Cruz and Alexander (2013) reviewed 49 fire spread models datasets and found that only 3% of the simulations acceptably predicted the observed rate of spread (RoS) and revealed the need to adjust those models if they were to be used operationally.

Moreover, the currently most advanced and commonly used wildfire models such as FARSITE, PHOENIX or PROMETHEUS (see section 1.2), have a lack of accurate operational forecasting capacities due to the scarcity of precise data available to initialize them (Finney et al., 2013) and the empirically developed sub-models that they contain Sullivan (2009c). Those drawbacks make them ardours and problematic to be exported to all sorts of different fire scenarios.

To deal with this problem, data assimilation (Tarantola, 2005) is a promising mathematical field that has shown great potential to efficiently improve models by observation (weather prediction, as a paradigmatic example). The investigation of its application to

wildfire field is considerably recent and one of the most promising directions to develop a reliable and operational wildfire spread forecasting system.

1.2.2   Data Assimilation and Inverse Modelling problem

In general terms data assimilation (DA) is a method for combining observations with outputs from a given model to improve it. The comparison between those is performed by means of some optimization process. Resolving a DA problem requires to study and evaluate the model starting from the observations, and this is, indeed, a particular case of an inverse problem. This sets one of the cornerstones of employing DA strategies to improve models; the *ill-posedness* problem.

The conditions that a well-posed problem must comply, as opposed to an ill-posed one, are defined by Ash et al. (2016) as:

- Existence of a solution

- Uniqueness of the solution

- Continuous dependence of the solution on the data

Usually, an inverse problem will not fully comply with some of those conditions (Ash et al., 2016). This notorious particularity will ultimately decide if a given problem is resolvable or not by means of DA.

Typically, there are two main strategies to set the DA algorithms; the classical (or variational) and the statistical (see Figure 1.7). The first one, comprises the minimisation of a certain cost function that expresses the error between observations and predictions. This minimization normally relies on the computation of the gradient of the cost function which can be extremely expensive for high-dimensional problems and thus, it normally takes advantage of adjoint methods. The second strategy involves the minimization of the variability or uncertainty of the model error and is based on statistical estimation theory. It is generally more complex and time-consuming than the variational approach and can give characteristics of the variability of the solution as a probability distribution.

Both approaches could be (and actually are) applied to the wildfire simulating problem. Current works using both techniques are commented in the following section (see 1.2.2.1) although the thesis at hand will only explore the variational DA approach. The author believes that the lack of a consistent wildfire spread theory (see section 1.1) justifies the selection of this simpler approach. Within the classical approach, different adjoint methods will be explored together with investigation of the *ill-posedness* problem and its potential limitations.
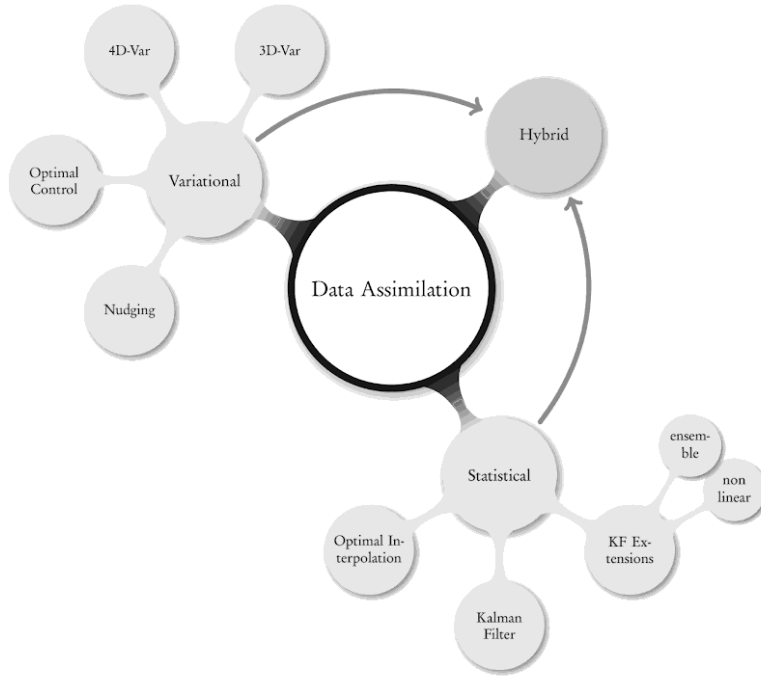
FIGURE 1.7: Schematic representation of the two main approaches within the data assimilation field and its most common algorithms. Source: Ash et al. (2016)

1.2.2.1   Existing data-driven systems on wildfire

Mandel et al. (2009) pioneered the use of data assimilation for wildfire modelling, predicting flame temperature and location using an ensemble Kalman filter framework and an atmosphere-coupled wildfire model. Their work showed promising results while raising some concerns about spurious fire corrections and the computing time required. Following their idea, Rochoux et al. (2014a,b,c, 2015) explored a data-driven wildfire simulator based on parameter and state estimation that assimilates fire front positions and corrects the wildfire forecast by means of a level set method based on Rothermel's model. They explored a parameter and state estimation strategy with stochastically based estimation of the error covariance matrices. The simulator was run using parallel computing and showed great capacity both in synthetically generated data and a small-scale controlled experiment. Lautenberger (2013) also used Rothermel's model and a data assimilating approach to calibrate fire propagating parameters to meet observed fire shape. All these authors used statistical DA techniques and an Eulerian level set method to track fire front propagation. Although this approach offers advantages when dealing with crossovers and large perimeters, it also has some drawbacks such as the level set solution dependency on the mesh (Mallet et al., 2009) and the requirement of a flux limiter to filter fire front shapes (Bova et al., 2015).

Some authors also worked to couple Lagrangian-based fire models with data assimilation. Denham et al. (2012) successfully applied genetic algorithms (GA) to find the wind configurations that best resembled observations to launch an improved forecast. A combination of weather and fuel calibration using fire perimeters has also been implemented using FARSITE and high-performance computing, showing great improvements and potential for long-term predictions (Wendt et al., 2011; Artés et al., 2014). Combining these ideas, Rios et al. (2014a) developed a data-driven algorithm based simplified version of Rothermel's RoS model (Rothermel, 1972a) and Huygens' elliptical propagation and optimized it by means of tangent linear method and automatic differentiation. The authors proved that this flat-terrain-restricted 2D system provides a short-term accurate forecast of wind-driven wildfires when tested with synthetically generated data.

Bianchini et al. (2010) proposed an alternative approach to classical data-driven prediction methods based on factorial experimentation (i.e. run multiple model simulation varying all parameters at each time) which they name Statistical System for Forest Fire Management (S^2F^2M) (Bianchini et al., 2005a). They applied this data driven strategy to a cell automata implementation of BEHAVE called FireLib (Bevins, 1996) simulator and compared the performance to two other methodology called GLUE (Generalized Likelihood Uncertainty Estimation), based on uncertainty prediction, and BBOF (Black-Box Optimization Framework), based on evolutionary algorithms. Whereas they find that S^2F^2M provided a considerable improvement on the results, the system requires to much computational resources to be used in real applications, as more than 5 minutes of computation were required to optimize 1 minute forecast on a High Performance Computer.

Finally, WIFIRE (Gollner and Trouvé, 2015) is the more recent, and still ongoing, project to explore data-driven techniques applied to wildfires. The project gathers experts from mathematics, computing and fire engineering disciplines with the aim of build an end-to-end cyberinfrastructure for real-time and data-driven simulation, prediction and visualization of wildfire behaviour (Altintas et al., 2015). Focusing on the data assimilation part and the currently available work (Srivas et al., 2016), they used Ensemble Kalman Filters (a statistical approach algorithm) together with FARSITE simulator. This enhancement of FARSITE shows great capacity to simulate 18 hours of observed fronts (with one hour time resolution). The paper also studies the impact of changing the variance (i.e. observations error) to show that convergence is still guaranteed although more iterations are needed. Nevertheless, no mention is made on the computational time and resources required to deliver the forecast, and weather it reaches positive lead times or not.

Chapter 2

Inverse modelling approach to wildfires: a proof of concept and its validation

In this chapter I introduce the algorithm rationale to create an inverse model capable of assimilating fire front observations and deliver an improved front spreading forecast. The model is prepared to cope with real scale dimensions and complexities and deliver positive lead time predictions by means of automatic differentiation. It is validated with infrared images acquired from real scale experimental shrub fires conducted in Ngarkat national park, Australia. Fragments of this chapter have been presented in the *VII International Conference on Forest Fire Research* (Rios et al., 2014a) as well as published in the paper entitled *Short-term fire front spread prediction using inverse modelling and airborne infrared images* at International Journal of Wildland Fires (Rios et al., 2016).

2.1 A data driven wildfire spread simulator

The system developed in this chapter combines two main sequential tasks that are envisaged to take place cyclically. The first one consists in solving an inverse fire modelling problem, by which the parametric space of certain observed fire front behaviour is found. The second task involves launching a forecast with the forward fire behaviour model, relying on the converged parameters. Thus, at its core, this approach is based on the assumption that there are several parameters (named invariants) that remain constant during a certain period of time. Those invariants can be either physical magnitudes or unknown model parameters that determine the fire front dynamics. With the aid of a

proper assimilation framework those invariants can be resolved using the available data
(e.g. fuel characteristics, topography, weather conditions, etc.) to reproduce certain
fire behaviour observations (i.e. fire perimeters sensed by airborne IR) and then input
back to the model to deliver a more accurate forecast. The algorithm begins with the
input of the observed fire fronts isochrones, the complementary available data and an
initialization of the invariants. Then, an optimization loop identifies the proper invari-
ants that best suit the fire dynamics observed, and those are finally used to launch the
forecast. The algorithm flow is depicted in Figure 2.1 and shows the key aspects that
will be explained in detail in the following sections.
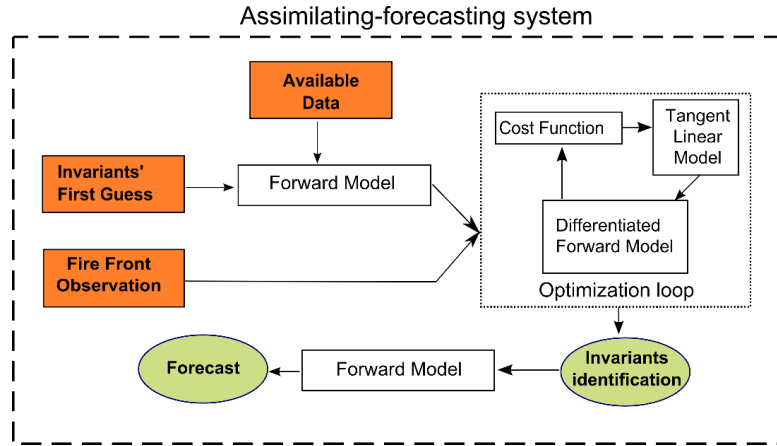
FIGURE 2.1: Flow diagram showing the algorithm rationale. Orange rectangular boxes
are inputs and green elliptical boxes are outputs.

2.1.1 Forward model

The first key aspect of the overall system is the forward model. The forward model is
the mathematical model that allows to explain the phenomenon and exhibits capacity
to reproduce the assimilated wildfire dynamics. To be suitable for an operational inverse
modelling scheme the forward model has to be computationally efficient to ensure the
forecast is delivered before the event takes place (i.e. positive lead time). The model
is composed by the rate of spread (RoS) model and the spreading model (i.e. way to
propagate this RoS in a 2D surface)

2.1.1.1 Spreading model

In the present work we chose a spread model based on Rothermel's (Rothermel 1972)
to determine RoS and a Huygens' expansion based approach as spreading model to
propagate the front in 2D (Richards 1990; Richards 1993). The Rothermel model (see
eq. (1)) is based on a simple energy balance that expresses the fire heat source by means

of the reaction intensity (γ_r) and the propagating flux ratio (ξ), and the fuel heat sink by means of the heat of preignition (Q_{ig}), the effective heating factor (ε) and the bulk density (ρ_b). The RoS is corrected by two experimental factors to account for the slope (Φ_s) and wind effects (Φ_w).

$$RoS = \frac{\gamma_r \xi}{\rho_b \varepsilon Q_{ig}}(1 + \Phi_s + \Phi_w) \tag{2.1}$$

Using Rothermel's (1972) correlations, the RoS can be ultimately expressed in terms of the fuel depth (δ), ovendry fuel loading (W_O), the surface-area to volume ratio (SAV), the packing ratio (β), the moisture of extinction and fuel moisture content (M_x, M_f) and the wind and slope factors (Φ_w, Φ_S):

$$RoS = \mathcal{F}(\delta, W_O, SAV, \beta, M_x, M_f, \Phi_w, \Phi_s) \tag{2.2}$$

At this stage, the slope factor is not considered since the algorithm is implemented for flat terrain as the available data for validation belongs to experimental fires in flat plots. In order to identify the invariants we performed a sensitivity study of the Rothermel's model to explore the variables dependency and influence on the RoS. Thousand randomly generated fuels models sets (W_O, β, SAV, M_x, M_f, δ), ranged among physically possible values taken from (Scott and Burgan, 2005a), were used to explore variables correlation with RoS. For each given set, the variables were swept from their minimum to their maximum while calculating the Pearson's product-moment correlation coefficient to search for linear dependencies. The fuel depth (δ) was found to be the variable that exhibited the greatest linear behavior as demonstrates the Person's coefficient histogram for the random generated fuel sets showed in Figure 2.2.
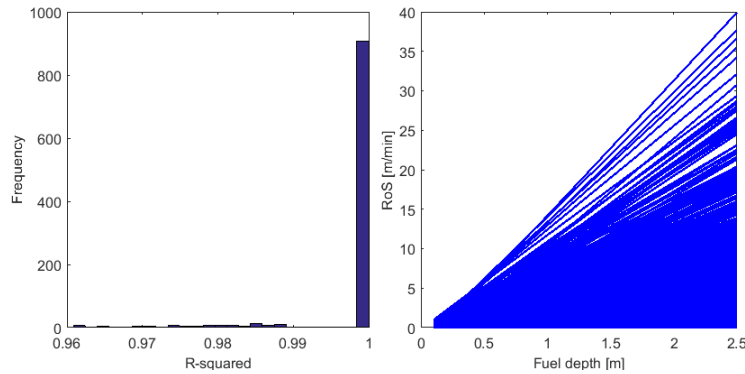
FIGURE 2.2: Sensitivity study of Rothermel's model dependency with fuel depth δ. The Pearson's product-moment correlation distribution (left) and the corresponding RoS vs. fuel depth for 1.000 model sets (right)

Thus, to determine useful invariants the Rothermel RoS was factorized as:

$$RoS = I_{mfw} \cdot \delta \tag{2.3}$$

Where I_{mfw} is an invariant that accounts for the overall effect of fuel moisture, moisture of extinction, bulk density, surface-area to volume ratio and wind factor.

The Huygens expansion approach considers every point in the fire front as a virtual ignition point that will expand following an elliptical geometry called firelet. The curve that envelops all firelets is then the expanded perimeter. The lateral (a) and the front velocity (b+c) of each firelet (Figure 2.3) can be derived from geometrical relationships given RoS and the length-to-breadth ratio LB of the firelet as:

$$a(s,t) = RoS(s,t) \cdot \frac{1 + \left(LB - \sqrt{LB^2 - 1}\right)^2}{2 \cdot LB} \tag{2.4}$$

$$b(s,t) = RoS(s,t) \cdot \frac{1 + \left(LB - \sqrt{LB^2 - 1}\right)^2}{2} \tag{2.5}$$

$$c(s,t) = b(s,t) - \frac{RoS(s,t)}{\left(LB + \sqrt{LB^2 - 1}\right)^2} \tag{2.6}$$

Where s is a mathematical parameter to discretize the front and t is the spreading time.
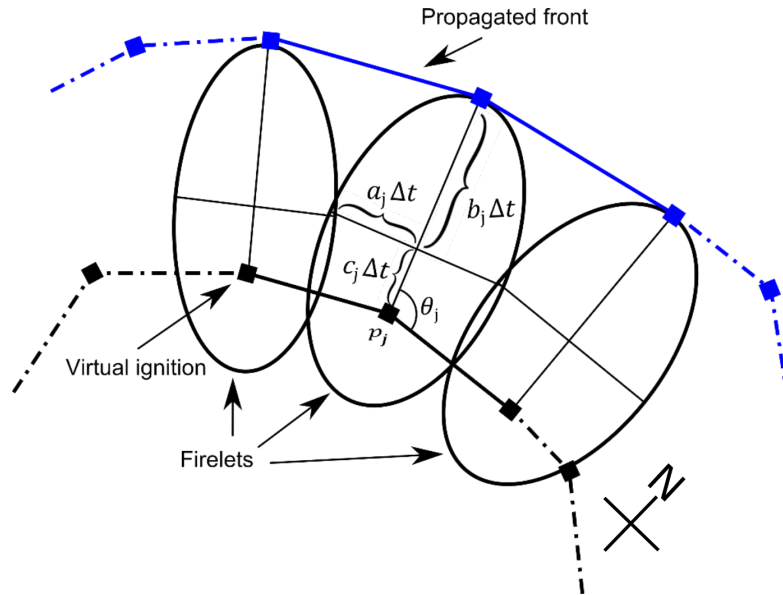
FIGURE 2.3: Geometrical relationship between a_j, b_j and c_j velocities for the elliptical firelet corresponding to the node pj of a given perimeter. θ is the wind-slope direction (i.e. towards where the fire is being pushed, counterclock wise, from the eastern axis) and the subscript j identifies a node of a given fire front.

To determine the length-to-breadth ratio LB we used Anderson's experimental correlation (Anderson, 1982):

$$LB = (0.936)^{0.2566U} + 0.461^{-0.1548U} - 0.397 \tag{2.7}$$

Where U is the mid-flame wind speed (in m s^{-1}). Note that equations 2.5 and 2.6 satisfy that:

$$b(s,t) + c(s,t) = RoS \tag{2.8}$$

As RoS is the distance, over a time unit, between the virtual ignition point on the perimeter and its perpendicular expansion as illustrated in Figure 2.3. The firelets enveloping curve is calculated using the partial differential equations derived by Richards (Richards, 1990, 1993) (eq. 2.9 and 2.10) that are resolved using a predictor-corrector method.

$$\frac{\partial x(s,t)}{\partial t} = \frac{a^2 cos(\theta)\left(\frac{\partial x}{\partial s}sin(\theta) + \frac{\partial y}{\partial s}cos(\theta)\right) - b^2 sin(\theta)\left(\frac{\partial x}{\partial s}cos(\theta) + \frac{\partial y}{\partial s}sin(\theta)\right)}{\sqrt{b^2\left(\frac{\partial x}{\partial s}cos(\theta) - \frac{\partial y}{\partial s}sin(\theta)\right)^2 + a^2\left(\frac{\partial x}{\partial s}sin(\theta) + \frac{\partial y}{\partial s}cos(\theta)\right)^2}} + c \cdot sin(\theta) \tag{2.9}$$

$$\frac{\partial y(s,t)}{\partial t} = \frac{-a^2 sin(\theta)\left(\frac{\partial x}{\partial s}sin(\theta) + \frac{\partial y}{\partial s}cos(\theta)\right) - b^2 cos(\theta)\left(\frac{\partial x}{\partial s}cos(\theta) + \frac{\partial y}{\partial s}sin(\theta)\right)}{\sqrt{b^2\left(\frac{\partial x}{\partial s}cos(\theta) - \frac{\partial y}{\partial s}sin(\theta)\right)^2 + a^2\left(\frac{\partial x}{\partial s}sin(\theta) + \frac{\partial y}{\partial s}cos(\theta)\right)^2}} + c \cdot cos(\theta) \tag{2.10}$$

Note that temporal dependency of a, b and c as stated in equations 2.4, 2.5 and 2.6. θ is the slope-wind direction, that corresponds to the wind blowing direction (i.e. towards where it is blowing) if the terrain is flat, measured counterclockwise from the east. Variables x and y are the coordinates of the nodes that constitutes a front $\{\mathbf{x_i}\}_t$ in a given time t. In the present implementation the mid-flame wind speed U in equation 2.7 and direction θ used in equations 2.9 and 2.10 are also considered invariants to be identified and named I_U and I_θ respectively. With equations [2.4-2.10] the final forward model $\mathcal{M}$ that provides the output of the set of front perimeters $\{\mathbf{x_i}\}_t$ can be defined as a function of the initial observed perimeter $\mathbf{x_o}$, the fuel depth δ and the invariants vector $\mathbf{q}$ as:

$$\{\mathbf{x_i}\}_{i=0}^{i=t} = \mathcal{M}(\mathbf{x_o}, \delta, \mathbf{q}) \tag{2.11}$$

$$\mathbf{q} = [I_{mfw}, I_u, I_\theta] \qquad\qquad (2.12)$$

2.1.1.2   Loop clipping schema

Owing to the Lagrangian formulation of the forward model at hand, when the front exhibits convex regions, the Huygens expansion approach may create some loops or overlaps along the perimeter (see Figures 2.4 and 2.5). These topological problems, already identified by previous developers using Huygens' principle (Barber et al., 2008; Finney, 1998), are physically meaningless and a loop-clipping filter is required to remove these loops and reclose the remaining front. The proper optimization of this loop clipping algorithm is a cornerstone to the development of a system to be used operationally as this filter has a dramatic influence on the computational time required to perform a single forward model run. The reason is that the filter is needed at every integration step to solve equations 2.9 and 2.10. This means that to create 5 isochrones every 10 minutes (i.e. 50 min simulation) the filter is required to be run 120 times. Thus, an optimal implementation is fundamental. Different methods have been proposed in the literature to tackle it. The Turning number, also known as winding number Richards and Bryce (1995), Bryce and Dueck (2009), can be used to identify the front sections that are internal to the curve and filter them out. By definition, the Turning number of a vector is the number of times a particle traversing a directed path around the curve will rotate counterclockwise around the vertex, (with clockwise rotations cancelling counterclockwise rotations), and will be zero if and only if the vertex is external to the curve. Figure 2.4 illustrates the Turning number mapping of a tangled perimeter together with two different algorithms to select the nodes to be removed (black markers) and nodes to be kept (red markers).

Whereas this method gives correct results in most of the cases, the Figure 2.4 exemplifies two approaches that fail to properly correct the front. Moreover, determining the Turning number for a large domain with multiple loops can be extremely time consuming (Bryce and Dueck, 2009). An alternative solution that considerately speeds up the calculation is to use the 2-colour theorem (Bryce and Dueck, 2009) and a set of filtering rules. This is the current implementation of Prometheus (the operational Canadian Forest Service's spread simulator, Tymstra et al. (2010)) but it also fails in some configurations (when triple loops are encountered, for example) and thus, we developed and alternative algorithm that copes with all issues highlighted so far.

Our routine is based on a two step process: (1) the determination and creation of all intersections (those on the perimeter and on the inner and outers loops) and (2) the construction and simplification of the adjacent matrix (containing all intersections) to clip all closed loops. The main hipothesis of the algorithm is that the initial node in the set is out of any existent loop (i.e. is a true perimeter marker). Although this rarely
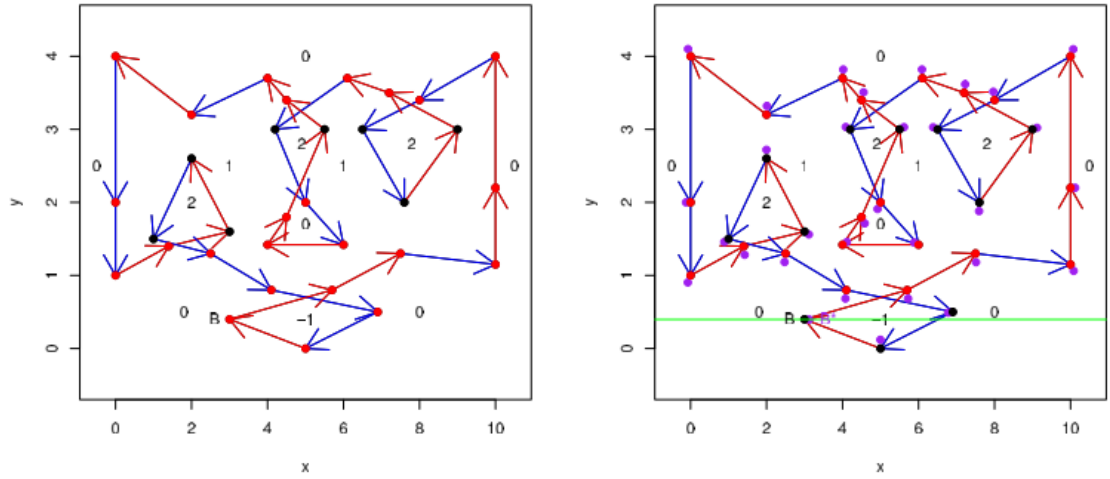
FIGURE 2.4: Turning number based algorithms. Despite domain mapping is the same in both cases the active nodes (red markers) differ depending on the algorithm used. Left plot uses the Richards and Bryce (1995) algorithm whereas right plot uses the scan line approach (green line as example) proposed by Barber et al. Source: Barber et al. (2008)

happens, the while loop is included to directly delete this first node if it happens to be part of a loop. All the process is fully vectorized (except for the while loop) to optimize the computational resources. The full code can be found in the Appendices (annex C). The downside of vectorizing is that the number of intersections to be analysed is the combinations-without-repetition of all segments taken 2 by 2. This introduce a factorial scaling that obliges to keep the total number of nodes below a certain threshold. The algorithm work-flow is as follows:

```
%% LOOP CLIPPING ALGORITHM
function loop_clipping()
while numel(XY_no_loop) < numel(XY_loop)/4

        combination_without_repetition()
        edges_matrix()
        intersect_edges()

        if intersection>0 % eliminate full inner and outer loops

                add_all_intersect_points() %all intersect points

                XY_no_loop=eliminate_loops() %eliminate all inner points

        end

        delete_last_marker()

end
end
```

LISTING 2.1: Loop clipping algorithm programming schema

The developers of Prometheus (Tymstra et al., 2010), found a loop case particularly complicated to be filtered out by common loop clipping algorithms and called it *Prometheus Nightmare* (Barber et al., 2008). This topological nightmare is a front markers distribution that can easily arose when applying Huygens principle in complex terrain or fuel and is displayed in green solid lines in Figure 2.5. The double loop that is originated and its intersection to the main perimeter are what make it difficult to be correctly filtered out. To validate our algorithm and its implementation we challenge it with the *Prometheus Neightmare*. The results are highly positive and our algorithm is able to cope with this tangled configuration and resolve it smartly. Results are presented in Figure 2.5.
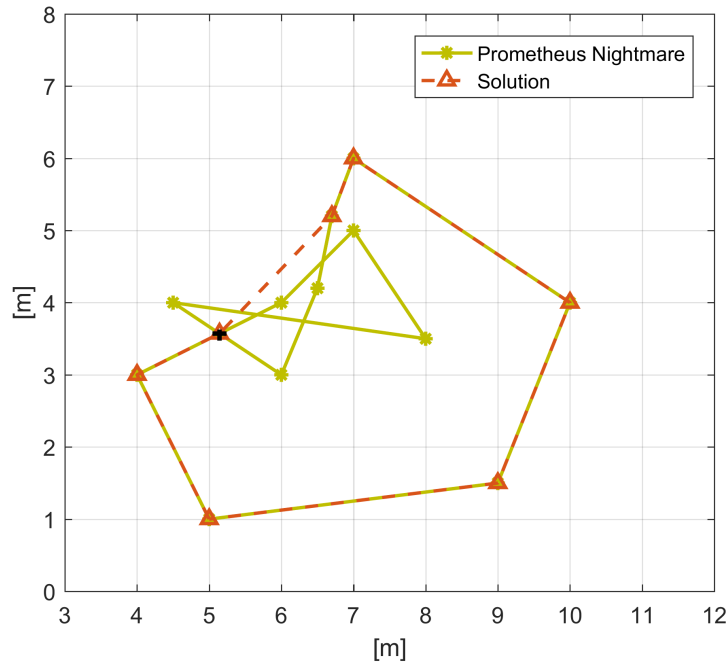
FIGURE 2.5: Prometheus nightmare front markers position (solid green line) and implemented solution (dashed orange line). Note that the black cross marker is automatically added to the front perimeter solution.

To further validate the algorithm we apply it to a double an triple loop configuration (see Figure 2.6). This triple node configuration is reported to be unresolved by Prometheus (Bryce and Dueck, 2009). In our case, the algorithm properly untangles the loop and closes the correct perimeter.

2.1.1.3   Regridding and degridding

The Lagrangian formulation also yields another problem. As the fire front grows and spreads out, the nodes distance increases and more nodes are needed to avoid formation
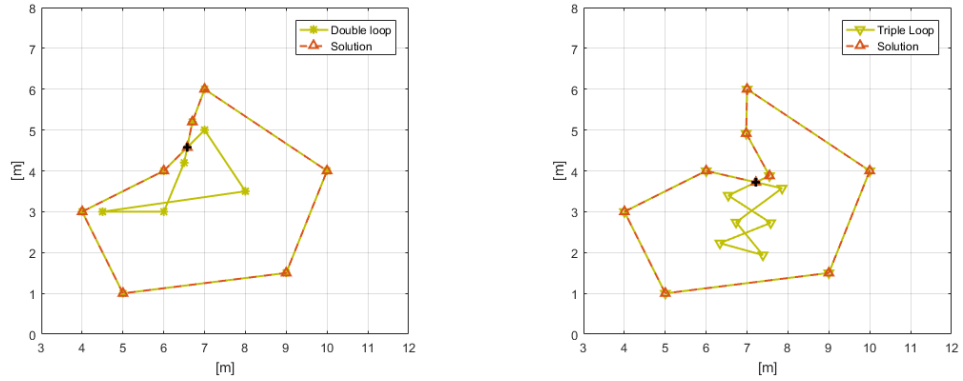
FIGURE 2.6: Double (left) and triple loops (right) problems correctly handled by our implementation

of sharp unreliable regions (see upper-left plot in Figure 2.8). This acute angles formation also occurs when the fronts overpass fuel patches with different characteristics. Therefore, a regridding process is needed to avoid both large distance between nodes and sharp vertex formation.

The acute angles correction process is conducted by comparing the two consecutive front edges adjacent angles α_j and α_{j+1} (see Figure 2.7). If the largest angle exceeds a given threshold D_a and the edge length is larger than a given threshold D_a, an extra node is added in the midpoint of the corresponding edge. By this methodology the new markers are added on the correct side of the acute vertex. To avoid the increasing separation between nodes, all edges distances are checked and those greater than a given threshold D_s are halved regardless of the adjacent angle. To effectively apply the filter both corrections must be implemented in a proper boolean logical way as illustrated in Listing 2.2 (the complete code can be found in the appendices C). As a result of the application of the whole filter, more nodes are automatically generated in abrupt and sharp areas smoothing the forthcoming simulated isochrones. The overall effect of the filters is displayed in Figure 2.8.

```
%% REGRIDDING ALGORITHM
function regridding()

if adj_angle >= T && edge_distance > D_a
        Add_node()

elseif edge_distance > D_s
        Add_node()
end
```

LISTING 2.2: Regridding algorithm rationale. Note that the loop is vectorized so it does not have to sweep one by one over all front nodes

However, the loop clipping process (described in the previous section) and the merging process (when two different front section merge together after overpassing an unburned fuel path for example) can cause the emergence of multiple nodes close together in particular regions (see left panel in Figure 2.6). The existence of such a nearby nodes can potentially generate more tangling on the subsequent spreading runs effecting both the accuracy and computing time. To filter this out, a degridding algorithm must be applied after the loop clipping and regridding steps. This algorithm remove all nodes that generate an edge shorter than a given threshold D_s by selecting one node per edge (see Figure 2.7. Since the routines are vectorized and it can sometimes arise that three or more nodes in a row are closer than the threshold the process is embedded in a recursive while loop to avoid removing unnecessary markers as illustrated in Listing 2.3. All routines contained in the while loop are vectorized. The effect of the degridding algorithm is depicted in Figure 2.8 panels (e) and (f). Notice that the presence of close markers in the sharp sections of the third isochrone do not propagate to following ones thanks to the filtering process. After some exploration the spatial and angular threshold are set to:

$$
\begin{aligned}
D_s &= 5 \quad [m] \\
D_a &= \frac{\pi}{10} \quad [rad]
\end{aligned}
\tag{2.13}
$$

```
%% DEGRIDDING ALGORITHM

function deggriding()

while nodes2remove >0
        nodes2remove=find_nodes2remove()
        remove_front_nodes()
end
```

LISTING 2.3: Degridding algorithm programming schema
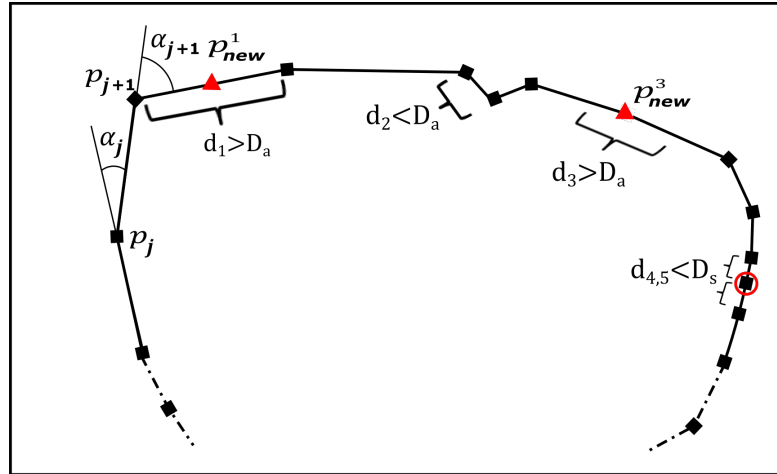
FIGURE 2.7: Schematic representation of regridding and degridding strategy. Squared black markers are the original nodes, red triangles are added nodes and the circle sign indicates eliminated nodes after filtering.

(a) no filter

(b) loop clipping

(c) loop clipping and regridding

(d) loop clipping, regridding and degridding

(e) zoom at (c)

(f) zoom at (d)
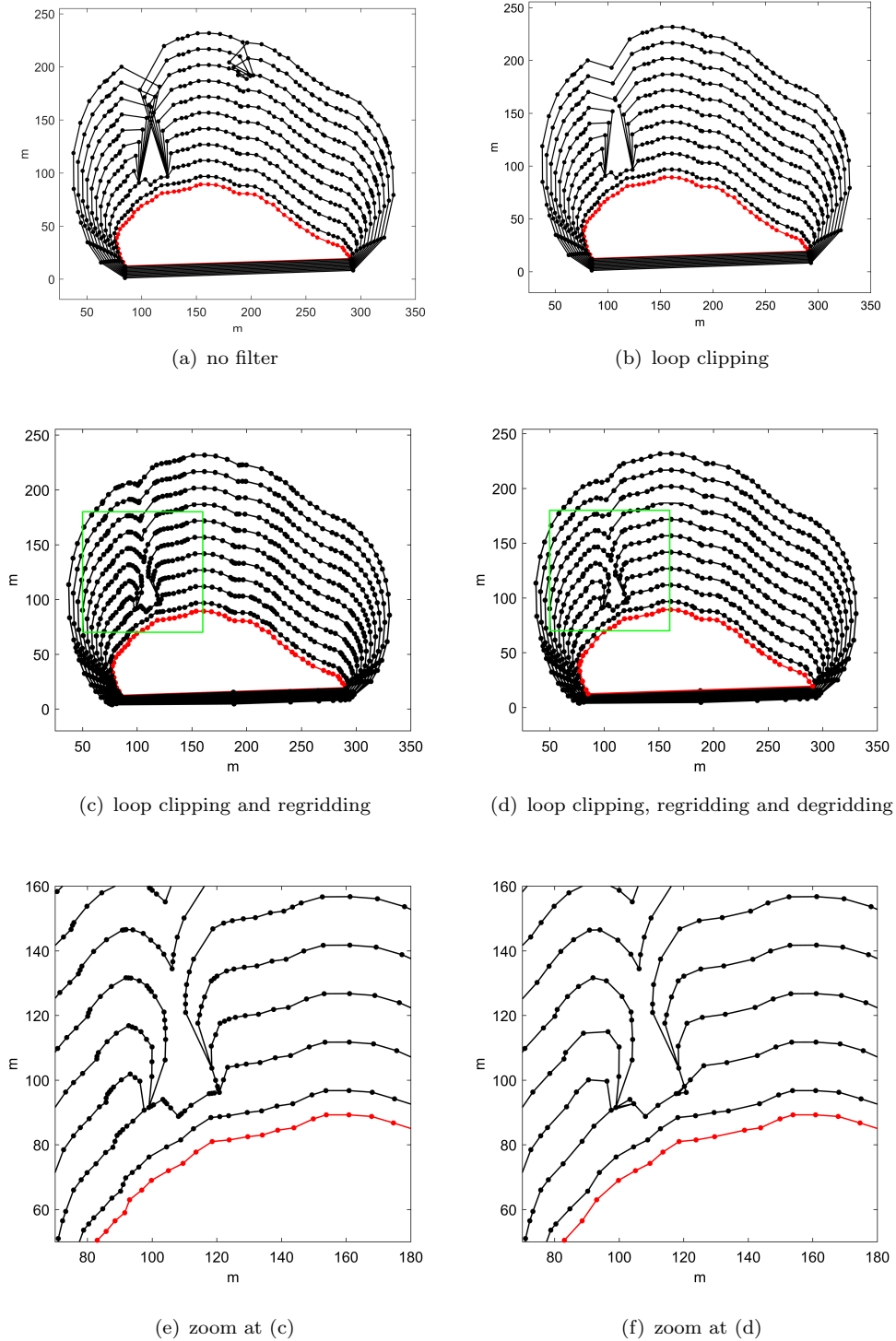
FIGURE 2.8: Effect of consecutively applying all three filters (loop clipping, degridding and regridding). Note that the filters are applied at every front before the propagation (actually at every integration step) and not once the propagation is generated. Panels (e) and (f) compare the degridding effect when the threshold $D_s = 5m$. The angular regridding threshold, D_a is set to $\pi/10$ rad. The red-markers line is the initial fire perimeter.

2.1.2   Cost function

To perform a proper assimilation process a cost function ($\mathcal{J}$) is defined to minimise differences between the observed and simulated front. The most direct way to define this cost function is using the averaged Euclidean distance between fronts summed all over the simulated isochrones. These distances are computed by selecting the intermediate points ($\tilde{p}_{j,i}^m$) between simulated nodes and their perpendicular intersection to observed front ($\tilde{p}_{j,i}^{ob}$) as depicted in Figure 2.9.

Then the cost function ($\mathcal{J}$) can be written in terms of the invariants vector $\mathbf{q}$ as:

$$
\mathcal{J}(\mathbf{q}) = \sum_{t=t_i}^{t_f} \frac{\sqrt{\left[\mathbf{x}_i^{ob} - \mathbf{x}_i^m\right]^T \mathbf{W}_i \left[\mathbf{x}_i^{ob} - \mathbf{x}_i^m\right]}}{n_i} = \sum_{t=t_i}^{t_f} \frac{\sqrt{\left[\mathbf{x}_i^{ob} - \mathcal{M}_i(\mathbf{q})\right]^T \mathbf{W}_i \left[\mathbf{x}_i^{ob} - \mathcal{M}_i(\mathbf{q})\right]}}{n_i}
$$

$$(2.14)$$

Where $\mathbf{x}_i^m$ and $\mathbf{x}_i^{ob}$ are vectors containing the concatenated Cartesian coordinates of all $\tilde{p}_{j,i}^m$ and $\tilde{p}_{j,i}^{ob}$ respectively for a given modelled and observed front i. n_i is the number of nodes in each isochrone, subscript i is the identifier of the simulated (and assimilated) front at any given time. $\mathbf{W}_i$ is the weight matrix that can be set to give more weight to the perimeters assimilated later in time, and give less importance to those assimilated a longer time ago. Several definitions of $\mathbf{W}_i$ will be explored and discussed in section 5.4.3.

The vector-based cost function definition $\mathcal{J}(\mathbf{q})$ is critical because it will drive the invariant's convergence to the final optimised value. Whether on not or not $J(\mathbf{q})$ is smooth and differentiable will effect the convergence capacity and even the availability to find an absolute minimum (the reader is referred to section 5.2.1 where this is extensively discussed). The actual vector-based method used to quantify the dissimilarity between two perimeters may entail some converging discrepancy when compared with the actual area between curves. Figure 2.9 illustrates this effect. When the simulated perimeter has convex regions and the observed does not, the intersection identification diverge and gives rise to large euclidean distances – i.e. large values of $J(\mathbf{q})$– even though the area enclosed is kept small (dashed arrows in Figure 2.9). Whereas this effect can be locally filtered, if both perimeters are extremely different it can affect the majority of the nodes. This effect is, thus, more important when handling large time steps and irregular front shapes (normally caused by highly heterogeneous terrain or fuel distribution). Therefore to asses whether the vector-based cost function works as expected we also compute the area enclosed between curves at each optimisation loop. This enclosed area contain both the over-predicted and the under-predicted areas (as displayed in fig 2.9). The reason why we do not directly use the enclosed area as a cost function is that to explore the automatic differentiation approach (see section 2.1.2.2 an analytical differentiable

function is needed. Calculating the area enclose between curve is a non-differentiable non invertible topological problem.
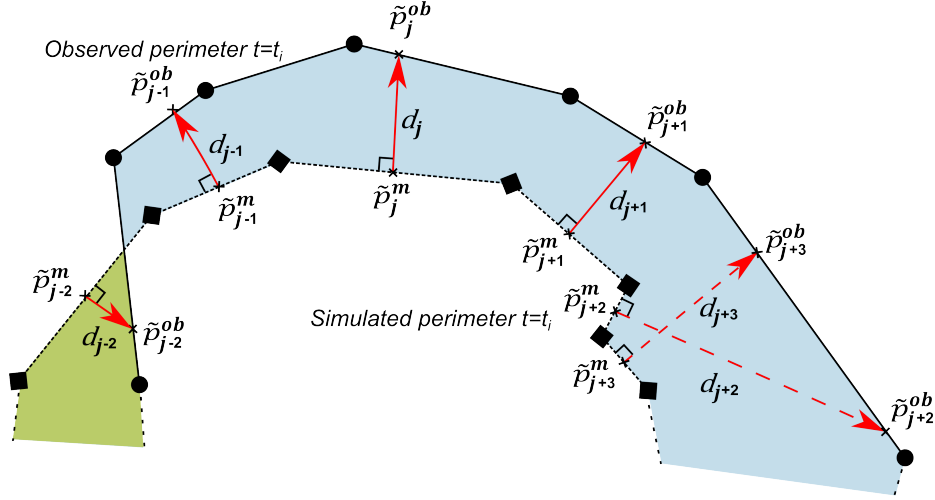
FIGURE 2.9: Cost function geometrical definition between simulated front (dashed line) and observed front (solid line). Red arrows are the Euclidean distances between simulated and observed perimeters. Blue shaded zone represents the under-predicted area whereas light green zone is the over-predicted area. The sum of both is used as a convergence controller

Once the cost function is defined, the inverse problem to find $\mathbf{q}$ can be formulated as the following optimisation problem:

$$min(\mathcal{J}(\mathbf{q})) \quad s.t. \quad \mathcal{M}(\mathbf{x}_o, \mathbf{q}, \delta, t) \approx \{\mathbf{x}_i^{ob}\}_{i=0}^{i=t} \tag{2.15}$$

For a given initial front observed coordinates $\mathbf{x}_o$, fuel depth δ and assimilating period t. For the present implementation, the invariants vector ($\mathbf{q}$) is:

$$\mathbf{q} = [I_{mfw}, I_U I_\theta] \tag{2.16}$$

And the search parametric domain is bounded with physically meaningful values as:

$$\begin{aligned} I_{mfw} &\in [0 - I_{mfw}^{max}](s^{-1}) \\ I_U &\in [0 - U^{max}](ms^{-1}) \\ I_\theta &\in [0 - 2\pi](rad) \end{aligned} \tag{2.17}$$

Where the invariants I_{mfw}, I_U and I_θ are ranged considering physical meaning with I_{mfw}^{max}, I_U^{max} as the maximum invariants values for a given scenario. This turns the optimization problem into a constrained problem, and allows for finite domain optimization techniques. In the present implementation, if any invariant exceeds these ranges during an assimilation loop its value is reset to the previous estimated value.

2.1.2.1 Optimization

The optimization strategy is one of the key factors when developing a data assimilation algorithm. Different approached have been explored in the literature. Whereas no in-depth comparison of optimizing schemas exists when it comes to wildfire forecasting problem, two main approaches can be used to solve optimization problems such as the one at hand (Onwubolu and Babu, 2013); the gradient free and the gradient based algorithms. The first methods smartly search on the optimization domain to find a global minima, but they are computationally expensive if multiple invariants are used, as the cost function needs to be evaluated multiple times for each invariant change and so the forward model. On the other hand, the gradient based strategies do not guarantee an absolute minima, however, if the forward model behave smoothly and the algorithm is initialized in the vicinity of the solution it will provide the correct answer with great computational efficiency (Nocedal and Wright, 1999). The more suitable strategy depends on the particularity of the system and its requirements. The problem at hand is a constrained non-linear optimization that might have to be solved multiple times during an operational event (i.e. every time new data is available to feed the system). In order to to deliver a solution in tight time constraints, thus, it has to be computationally efficient. Moreover, we want the algorithm to be ready to handle larger number of invariants what might increase the computational cost. For all these reasons, the gradient based approach is chosen.

To speed up the optimization process, we take advantage of a Tangent Linear Model (TLM) converging algorithm. This algorithm is chosen as it was reported to over-perform other strategies (such as steepest descent) according to Jahn (2010). TLM method iteratively linearise the forward model ($\mathcal{M}i(\mathbf{q})$) on the vicinity of the most recent invariants' estimation ($\mathbf{q}_b$). The TLM is applied by means of first order Taylor expansion. When the iterative loop begins the expansion is performed around a first educated guess $\mathbf{q}_0$ based on available data and user experience.

The gradient of the linearized cost function is then written as equation 2.18. Note that to efficiently apply the optimisation condition ($\nabla \mathcal{J}(\mathbf{q}) = 0$) of equation 2.14, we take the square of the Euclidian distance (removing the squared root) as it does not affect the optimisation.

$$\nabla \mathcal{J}(\mathbf{q}) = \sum_{i=t_i}^{t_f} \frac{\left[\nabla \mathcal{M}_i(\mathbf{q}_b)(\mathbf{q} - \mathbf{q}_b)\right]^T \mathbf{W} \left[\mathbf{x}_i^{ob} - (\mathcal{M}_i(\mathbf{q}) + \nabla \mathcal{M}_i(\mathbf{q}_b)(\mathbf{q} - \mathbf{q}_b))\right]}{n_i} \qquad (2.18)$$

Which can be rewritten as:

$$\sum_{i=t_i}^{t_f} \frac{\nabla\mathcal{M}_i(\mathbf{q}_b)^T \mathbf{W}\mathcal{M}_i(\mathbf{q}_b)}{n_i}(\mathbf{q}-\mathbf{q}_b) = \sum_{i=t_i}^{t_f} \frac{\nabla\mathcal{M}_i(\mathbf{q}_b)^T \mathbf{W}(x_i^{ob}-\mathcal{M}_i(\mathbf{q}_b))}{n_i} \qquad (2.19)$$

Equation 2.19 is a linear system that can be solved for $(\mathbf{q}-\mathbf{q}_b)$ using a QR factorization algorithm (Nocedal and Wright, 1999). Despite the fast convergence of TLM method for smooth problems, the gradient $\nabla\mathcal{M}_i(\mathbf{q}_b)$ needs to be solved to compute the converging step at each iteration $(\mathbf{q}-\mathbf{q}_b)$. As a novelty, when it comes to wildfire applications, and following the promising results found for enclosed fires (Jahn et al., 2012) we make use of automatic differentiation codes to directly compute the gradient avoiding multiple forward model evaluation (i.e. difference-base) methods. This approach was already successfully implemented in the synthetic data test performed by Rios et al. (2014a).

2.1.2.2   Automatic Differentiation

One of the principal novelties of this algorithm is the use of automatic differentiation (AD) to resolve the optimization problem. AD algorithms have been used in multiple and divers fields (Bücker and Corliss, 2006) but, to the knowledge of the author, they have never been applied to wildfire forecasting problem. AD algorithms are based on the fact that the function itself is scripted in a command-line program. Thus, the scripted program can be obtained by applying the chain's rule for derivatives (Bischof et al., 2006). There are two possibles approaches to reach this: the forward and the adjoint (reverse) accumulation. Those strategies are illustrated with the following short programs:

Forward and adjoint automatic differentiation example:

Let us consider the function:

$$y(x_1, x_2) = e^{\frac{x_1}{x_2}} \qquad (2.20)$$

The code that evaluates this function is for $x_1 = 1$ and $x_2 = 2$

Program
x1=1
x2=2
x3=x1/x2
y= exp(x3)

The analytical partial derivative with respect to x_2 is:

$$\frac{\partial y}{\partial x_2} = \frac{\partial e^{\frac{x_1}{x_2}}}{\partial x_2} = e^{\frac{x_1}{x_2}} \cdot (-\frac{x_1}{x_2^2}) \tag{2.21}$$

That when evaluated gives:

$$\left.\frac{\partial y(x_1, x_2)}{\partial x_2}\right|_{x_1=1, x_2=2} = e^{\frac{1}{2}} \cdot (-\frac{1}{2^2}) = -0.41 \tag{2.22}$$

The forward differentiation way to calculate this derivative is by scripting the forward differentiation of the `Program`. The `dProgram` computes the partial derivative with respect to x_2 at every command line without directly knowing the analytical derivative of the original function $y(x_1, x_2)$:

dProgram
x1=1
dx1=0
x2=2
dx2=1
x3=x1/x2
dx3=-x1/(x2*x2)*dx2
y= exp(x3)
dy= exp(x3)*dx3

Which gives `dy`=-0.41.

The adjoint differentiation (sometimes called reverse or backwards differentiation) works as follows:

aProgram
x1=1
x2=2
x3=x1/x2
y= exp(x3)
dy=1
ax3=exp(x3)*dy
ax2=ax3*(-1/(x2*x2))
ax1= ax2* x1

Which also gives `ax1`=-0.41.

Note that the number of statements in both `dProgram` and `aProgram` are twice the initial number and thus, the time to run it could be expected to double. However, just by running the `dProgram` once, the Jacobian (i.e. gradient) matrix can be computed. This elemental example shows how this algorithm can be applied to differentiate any kind of code. Because it does not require to run the forward model multiple times (as more common methods do) (Griewank, 2000) it is a potential strategy to speed up optimization time.

The forward approach is indicated when the number of observables (i.e. model outputs) is much larger than the quantity of independent variables (i.e. the number of invariants in our case). By contrast, calculating the adjoint differentiation is more convenient and efficient when there is a large number of independent variables.

Therefore, in the present work, we use forward differentiation approach. The differentiation of the code was generated with the help of `ADiMat` (Bischof et al., 2002), a set of functions and libraries that work with `Matlab` software.

2.2   Real Data Validation

To investigate whereas this proof of concept has the potential to be deployed in an operational way, it is necessary to validate it with real-scale fire front data as one of the main requirements is to be able to handle large number of nodes and complexities inherent to real wildfires and deliver a useful forecast.

The algorithm is tested with data from two large-scale experimental fires performed in South Australia in March 2008. Despite these being within the framework of a scientific experimental burning program, the fires exhibited real wildfire behaviour patterns as they were performed in large plots under extremely severe weather conditions. These experiments were conducted in Ngarkat Conservation Park (35°45'S, 140°51'E), which consists of a characteristic dune and swale system comprising large flat areas (130 m above sea level) of mallee-heath shrublands. Fire behaviour in mallee-heath fuel types is characterised as being discontinuous and highly variable owing to the heterogeneous characteristics of the various fuel layers that comprise mallee-heath fuel complexes (Cruz et al., 2013).

2.2.1   Ngarkat experiments

Data were gathered from two experimental burns. These burns were performed in two different sites named hereafter *Shrub site* and *Woodland site* (plots A and AS2E respectively according to Cruz et al. (2013)). The burn in the Shrub site was performed in a 9-ha, 9 year-old heath plot on 4 March 2008. This fuel complex was characterised

by scattered small-leafed shrubs, organised in clumps, and a discontinuous litter layer partially buried by sand. The main fire-carrying fuel layer was the discontinuous shrub canopy. Ambient conditions were a temperature of 32°C and relative humidity of 25%. Wind speed (10-m open) averaged 15 km h^{-1}, with gusts up to 30 km h^{-1}. Wind direction was south–south-westerly. The characteristic dead fuel moisture of the fuel complex was 7.7%. The fire was ignited with a 150-m long line and 2 min after ignition, flame heights were $\approx$ 2–2.5 m, with flashes up to 4 m. The fire spread vigorously throughout the plot with sustained flames heights of 4–5 m. Eight minutes after ignition, the main flame front hit the northern border of the plot, concluding a 350-m head fire run (Planas et al., 2011a).

The burn in the Woodland site was performed on 5 March 2008, in a 25-ha plot of a 22 year-old woodland dominated by mallee eucalypts. The fuel complex in this plot was characterised by a mallee overstorey 2 to 3 m tall and a shrubby understorey that constituted the main fuel layer supporting fire propagation. Air temperature was 36°C and relative humidity 13%. The 10-m open wind speed averaged 17 km h^{-1}, gusting up to 36 km h^{-1}. Sampled dead fuel moisture of dead suspended fuels was 7%. The fire was ignited using a 250 m line and initially spread in heath vegetation with flame heights averaging 4 m. As the fire burnt into mallee vegetation, crowning ensued, with flame heights averaging 4.5 m, and peaking at 10 m. A short lull in wind speed 5 min after ignition slowed fire propagation and revealed multiple spot fires along the fire perimeter. As wind speed increased again, the fire made an intense crown fire run in the mallee vegetation, with flame heights between 8 and 10 m and spot fires developing 40–60 m ahead of the fire.

The Woodland site was used to study and characterise aerial suppression efficiency (Plucinski and Pastor, 2013). Thus, for the evaluation of the algorithm developed in the present work, only those perimeters before the first aerial drop were used.

Both plots were filmed from a hovering helicopter equipped with an IR camera. This camera operated within the 7.5–13 µm range and stored sequences of IR images (240 × 320 pixel) at an approximate rate of 5 frames s^{-1}. The helicopter was positioned so that the majority of the plot was in view for the duration of each fire.

The images were georeferenced using beacons deployed along the plot as referencing markers and georectified assuming a flat terrain (see Figure 2.10). Fire front positions were extracted every 10 s to create isochrones and RoS maps for both plots by applying a methodology for IR analysis described in (Pérez et al., 2011). This methodology has 3 principal processes:

- **Georeference**
  This is the process to assign real world coordinates to the image. Theoretically, it

can be performed by only using platform inertial (i.e. camera orientation) and camera specifications (i.e. resolution, field of view, etc.). However, the inherent sensor error and noise, specially when using airborne platforms, are extremely magnified and renders this approach unfeasible. The robustest way to perform georeference is by using ground control points (GCP). Those GCPsS are any identifiable point in the image whose GPS coordinates are available. For visible-spectrum images those points cant be road intersections, plot corners, stationed vehicles, etc. However, those features are not detected by infrared images when the temperature ranged is set to [300-800ºC] (range needed to prevent image saturation when filming fire). Thus, hot (i.e. detectable by IR) beacons consisting of georeferenced pre-ignited bonfires were used.

- **Orthonormalization**

  This process consists in projecting a non ortho-image to an orthonormalized plane. In the raw captured images all pixels correspond to different real-world area as the ground-to-camera distance is not constant for the whole picture (panel (b) in Figure 2.10). To correct for this, an homography transformation generated with GCP must be performed (Pastor et al., 2006; Rudz et al., 2009). As the camera view and orientation changes at every frame, the process must be individually repeated for every image. For the data sets at hand, a simplification of flat terrain is made to ease this homography transformation. Although the plots had some protuberances, the digital elevation model (DEM) did not capture them and the simplification is considered to unalter the results.

- **Flame filtering and edge detection**

  Finally, once the image is geo-orthonormalized (see panel (c) in Figure 2.10) the active fire front must be identified and depicted to create fire front isochrones (plotted in (d) in Figure 2.10). Whereas edge detection algorithms are normally used for common image segmentation, the flame tilting effect (blurry lines in panel (c) in Figure 2.10) requires a particular filtering step before its application. The flame tilting effect is due to the fact that flames emerge out of the picture plane and are then wrongly homographied. This process was manually performed as expert judgement was the best available tool at the moment. Recently, advances in automatising this filtering methodology that showed high agreement with the manually performed step (Valero et al., 2015, 2016).

Visible channel

(a)

Raw IR channel

(b)

Geo-Orthonormalized

(c)

Isochrones extraction

(d)
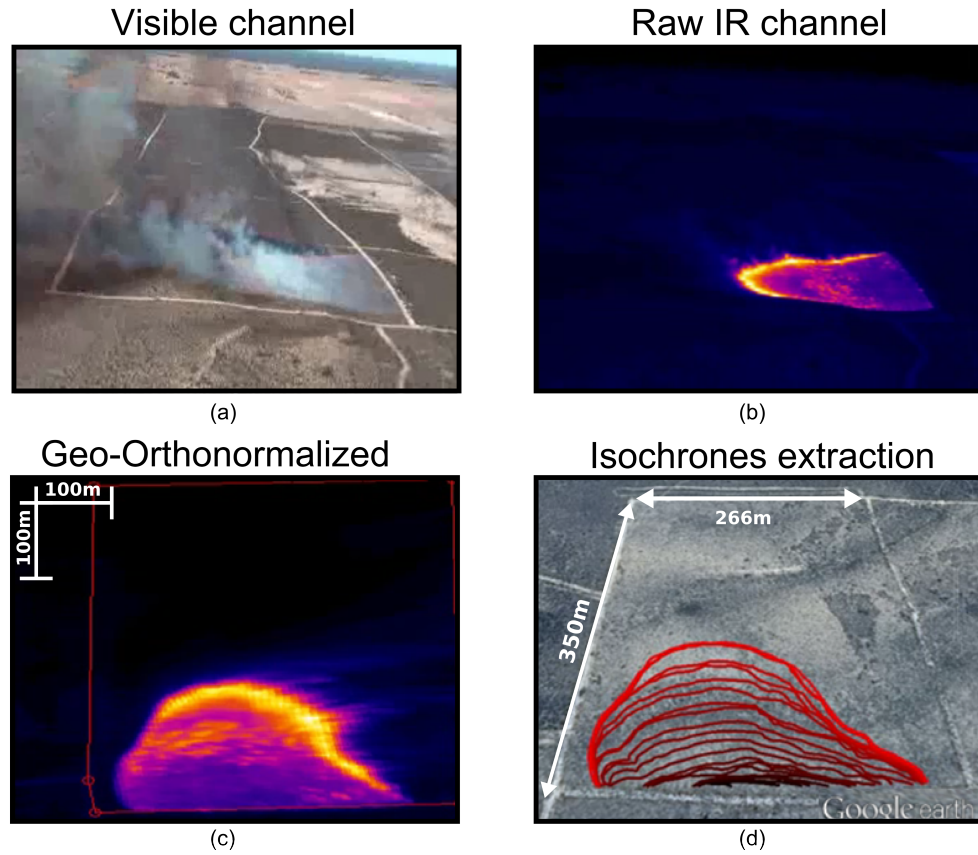
FIGURE 2.10: Shrub site composite of (a) the visible channel, (b) the raw infrared (IR) image, (c) the orthonormalisation of the image using reference points and (d) the isochrone extractions and representation in a geographic information system (GIS) interface. All the images represent t = 210 s after ignition.

The fire in the Shrub site lasted 370 s, giving 38 perimeters at 10-s frequency, although the initial 130 s are considered to be within the artificial linear ignition and were discarded, leaving 25 perimeters available for the current algorithm validation ((a) plot in Figure 2.11). The fire in the Woodland site lasted 450 s but the first drop discharge took place 240 s after ignition. Thus, taking the perimeters from 60 s on to avoid the effect of the artificial ignition, 19 perimeters at 10-s frequency were left for the purpose of the current study. Both set of isochrones are depicted in (a) and (b) plots in Figure 2.11.
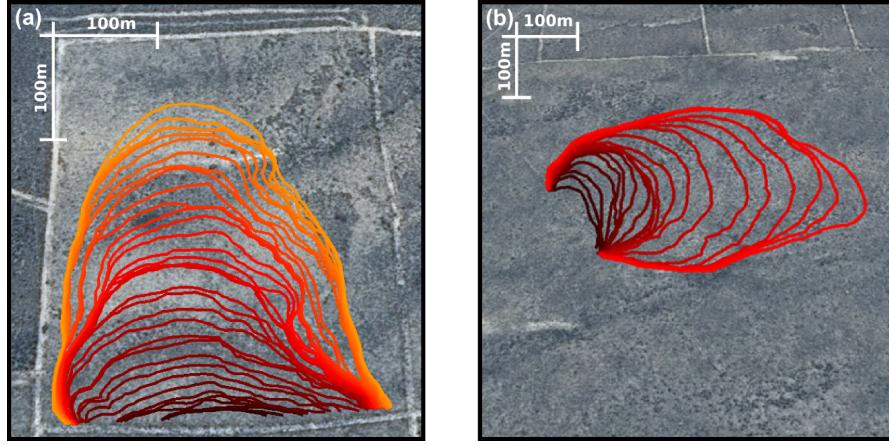
FIGURE 2.11: Shrub site (a) and Woodland site (b) burnt during the Ngarkat experiments campaign. Isochrones plotted every 10 s are represented on GoogleEarth after georeferencing and orthonormalising. Map data: Google, CNES/Astrium

2.2.2   Results

The two presented plots are chosen to illustrate the algorithm performance when tested with real data. The wind speed and direction invariants are initialised using approximate values registered during the experiment (simulating the way the system could be used in a real operation). The I_{mfw} invariant initial value (Table 2.1) is calculated with equation 2.1 and 2.3. Required parameters are taken from the standard fire behaviour fuel model SH5 (Scott and Burgan, 2005b) corresponding to *Very high load, dry climate shrubs* and complemented with values of moisture measured during the Ngarkat experiments (Table 2.1). This moisture content was measured using the oven-dry method. The upper limits for the invariants I_{mfw}^{max} and U^{max} are an upper bound needed to set the optimisation problem. Their values were set considering conservative maximum reasonable values for the scenario. In the present examples, those values were never reached.

The fuel depth δ is considered homogeneous. The system is tested using this approximation as in a real-case scenario where an accurate fuel map might not be available. In the following sections, different outputs of the system are separately evaluated.

2.2.2.1   Assimilation step

The first factor to validate the system at hand is the assimilation stage. Only if the optimisation algorithm manages to find the proper invariants that model the observed fronts will it be possible to launch a reliable forecast. The assimilating process is iterative and stops when the percentage difference of the invariants between two consecutives

TABLE 2.1: All values needed to initialise the algorithm and their source. SH5 corresponds to the *Heavy Shrub Load* fuel model from Scott and Burgan (2005b)

Variable	Initial Value		Source
	Shrub Site	Woodland Site	
SAV	$4107\ m^{-1}$	$4107\ m^{-1}$	SH5
M_x	15%	15%	SH5
β	$2.06 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$	SH5
W_o	$1.45\ kg \cdot m^{-2}$	$1.45\ kg \cdot m^{-2}$	SH5
M_f	7%	7.7%	Measured
δ	1.5 m	1.4 m	Measured
U^0	$4.16ms^{-1}$	$4.7ms^{-1}$	Measured
I^0_{mfw}	$0.388s^{-1}$	$0.664s^{-1}$	Estimated
I^0_{θ}	0 rad	0 rad	Estimated
I^{mas}_{mfw}	$10\ s^{-1}$	$10\ s^{-1}$	Estimated
U^{max}	20 m s^{-1}	$20\ ms^{-1}$	Estimated

loops (named relative error, ξ_k) reaches a threshold of $T_\xi = 0.05\%$ as:

$$\xi_k = \frac{\mathbf{q}_k^i - \mathbf{q}_k^{i-1}}{\mathbf{q}_k^{i-1}} \cdot 100 \leqslant T_\xi = 0.05 \tag{2.23}$$

where the index k represents each invariant, the superscript i each loop iteration.

As mentioned before, the area between the curves is tracked to check the correctness of the cost function. The absolute areal error is calculated by summing all the enclosing areas between the simulated front, M, and the corresponding observed front, O:

$$\xi_A = \sum_{i=0}^{i=N_f} \cdot (M_i \oplus O_i) \tag{2.24}$$

where $\oplus$ is the XOR logical operator, ξ_A is the absolute areal error and the subscript i corresponds to a given time (i.e. pair of fronts).

Apart from using the area enclosed between observed and modelled fire perimeters there are other error metrics found in the literature that allow us to asses quantitative comparison between shapes. We take three of those indexes to further evaluate the capacity of our system to recreate observed fronts and forecast those to come. Those metrics are defined as:

- **SDI'** is the inverted shape deviation index (SDI) (Cui and Perera, 2010):

$$\text{SDI'}_i = 1 - \frac{M_i \oplus O_i}{O_i} \tag{2.25}$$

- **Sørensen** (Sørensen, 1948):

$$\text{SØRENSEN}_i = \left(1 - \frac{2 \cdot (M_i \cup O_i)}{M_i + O_i} \right) \tag{2.26}$$

- **Jaccard** (Jaccard, 1901):

$$JACCARD_i = \left(1 - \frac{(M_i \cup O_i)}{M_i + O_i - (M_i \cup O_i)} \right) \tag{2.27}$$

Figure 2.12 illustrates the modelled M_i and observed O_i areas taken into account for error metrics calculations as well as XOR ($\oplus$) and $\cup$ topological operations.
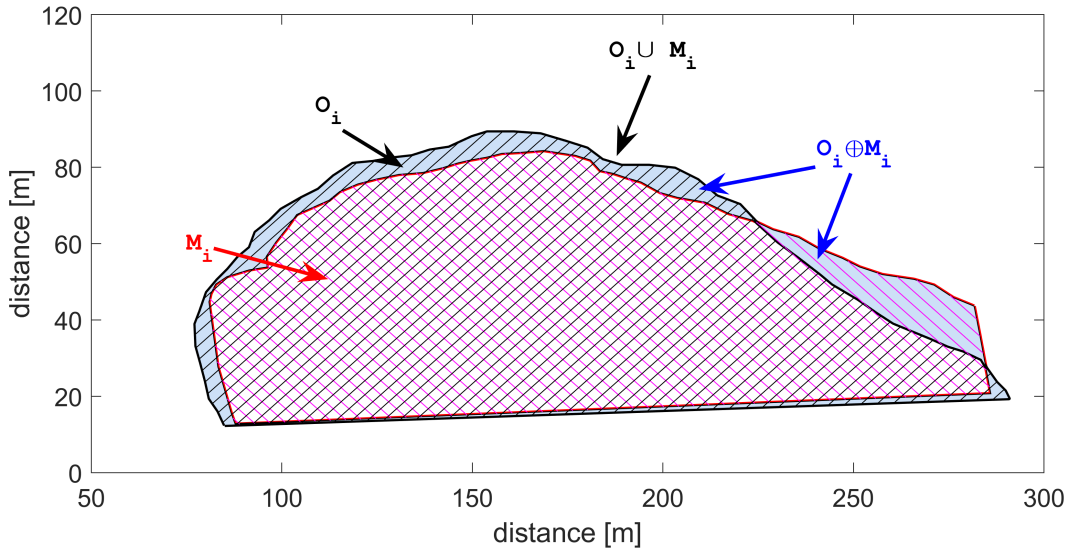
FIGURE 2.12: Modelled area (M_i, red patch) and observed area (O_i black patch) used to create the error metrics index. $O_i \oplus M_i$ is the blue shaded area whereas $O_i \cup M_i$ is composed by all patched regions.

The three similarity indices have a score range of [0–1], 1 being the best score, indicating identical shapes. Their main difference is the denominator identification or, in other words, the normalising factor. Because it is not clear which of the mentioned indexes performs best to quantify simulation correctness (Filippi et al., 2014a), we analyse the three of them.

Figure 2.13 shows six assimilated isochrones for the Woodland site. The isochrone captured at 40 s after the ignition is used as the initial perimeter x_0. The subsequent assimilated isochrones span 10 s. The optimisation loop converges after three iterations and the similarity indices are kept over 0.9, showing great resemblance.
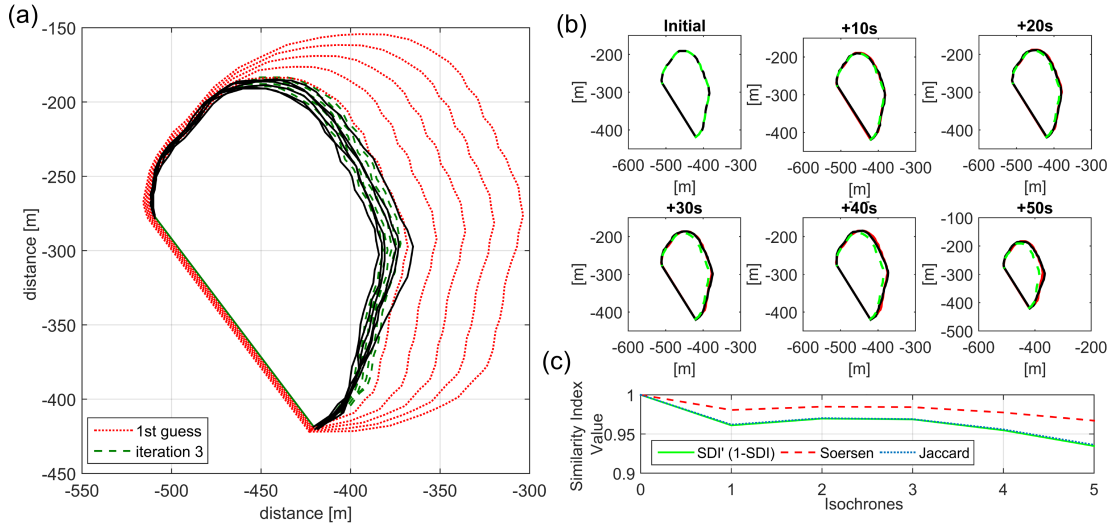
FIGURE 2.13: Assimilating process for the Woodland site: (a) The assimilation is resolved after three optimisation iterations. (b) Detailed areal difference after optimisation convergence for every simulated isochrone. Solid (black) lines are the observed front, dotted (red) lines the simulated front and dashed (green) lines the initial front considered. (c) Similarity indexes. Note that the y-axis is expanded between 0.90 and 1.00.

The same procedure is applied to the Shrub site; in this case, six isochrones spanning 10 s are assimilated. The first isochrone corresponds to 140 s after fire ignition. The optimisation loop converges after eight iterations (Figure 2.14). The similarity indices are over 0.9 for the three initial fronts and decrease to 0.8 owing to imperfect simulation of the right flank. This flank showed an extremely low RoS that could not be fully replicated by the forward model as only one common lateral RoS (a) is resolved.
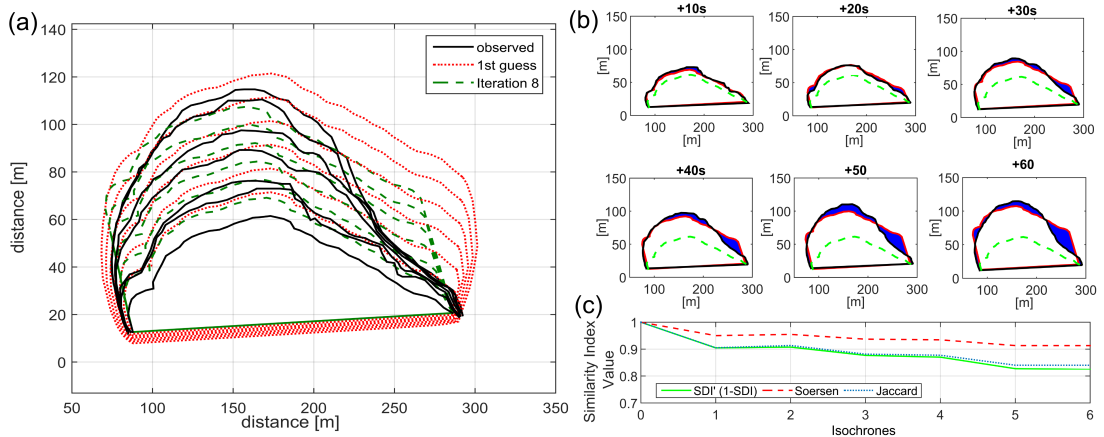
FIGURE 2.14: Assimilating process for the Shrub site: (a) The assimilation is resolved after eight optimisation iterations. (b) Detailed error after optimisation convergence for every simulated isochrone. The shaded region (blue) represents the areal error for a given simulated and observed isochrone. (c) Similarity indexes. Note that y axis is expanded between 0.50 and 1.00. For lines definition, see Figure 2.13 caption

In both cases, the assimilation stage shows great ability to model the front data regardless of the number of assimilated isochrones.

2.2.2.2   Invariants convergence

The effectiveness of the algorithm relies on the individual convergence of each invariant. Only if they are all individually resolved will their value be meaningful and useful to run the forecast. For each assimilation, the convergence of the cost function and the invariants' individual convergence is tracked and displayed in Figure 2.15 and Figure 2.16.
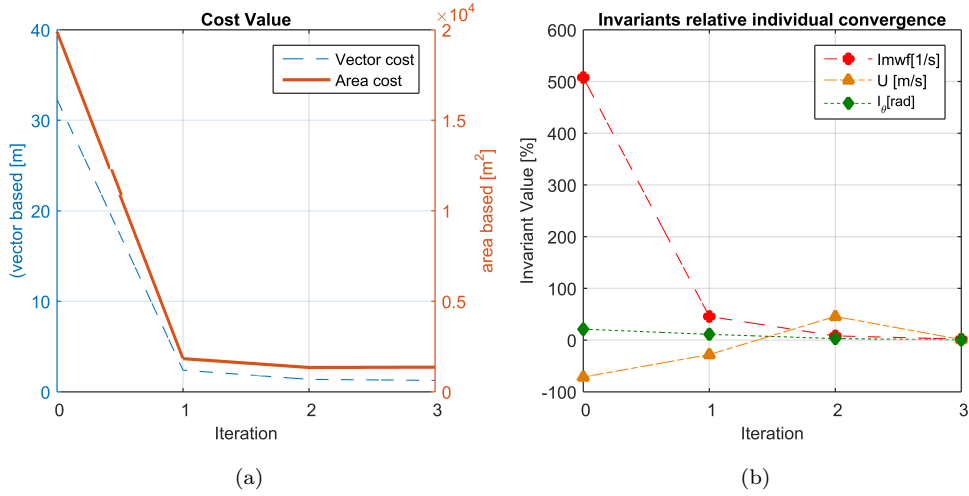
FIGURE 2.15: Woodland site convergence: (a) The vector cost function value (dotted blue line) and the areal cost value (solid orange line) during the optimisation loop. (b) The relative convergence for the three invariants

The invariant relative error (ξ_k, equation 2.23) shows how far from the final value the invariant was initialised. For the Woodland site, the invariants converge to values of 0.542 s^{-1}, 2.651 m s^{-1} and 0.569 rad (see Table 2.1), which represents a change from the initial values of 500%, $-$ 80% and 25% (Figure 2.15). Applying equation 2.1 and assuming an average fuel depth of 1.4 m, the I_{mfw} invariant can be expressed as a RoS with the value of 45.5 m min^{-1}. This value represents a RoS average for the whole perimeter during the assimilation period and is in agreement with the values reported in Planas et al. (2011b) where RoS was found to range between 20 and 60 m min^{-1} along the whole perimeter. The wind speed and direction also had reasonable values although more detailed data are not available for comparison.

The vector cost function calculated as 2.14 converges to 1.2 m (see Figure 2.15), which indicates the average normal distance between simulated and observed fronts. At
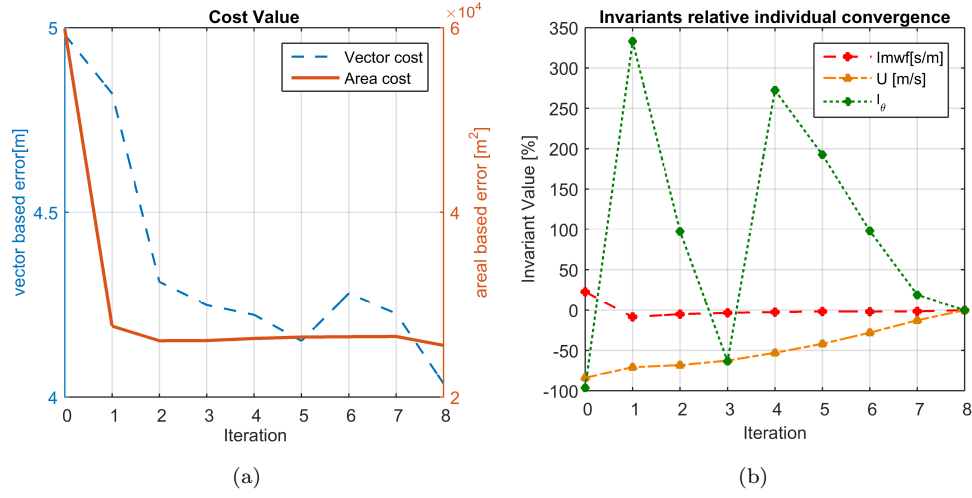
FIGURE 2.16: Shrub site convergence: (a) The vector-based (blue dotted line) and the areal absolute error (orange solid line). (b) The invariants' relative error. The I_θ invariant fluctuation is due to the nature of angular direction restricted to $[0-2\pi]$.

the same time, the sum of the absolute areal error for all perimeters converges to a value of 1363 m^2.

Similar results are obtained for the Shrub site (see Figure 2.16). In this case, invariants converge to the values 0.447 s^{-1}, 5 ms^{-1} and 5.95 rad, which represent an average RoS of 10.4 min^{-1}. The values observed in Planas et al. (2011b) ranged from 5 to 40 min^{-1}. In this case, the average distance between simulated and observed fronts is 4 m whereas the areal distance is 2135 m^2. As mentioned before, both convergence indicators do not follow exactly the same behaviour and the vector-based one finds a minimum before stabilising. This highlights the need for an areal-based cost function instead of a vector-based one.

2.2.2.3 Forecasting step

Once the invariants are resolved, we can launch a forecast using the last assimilated isochrone as the initial perimeter ($\mathbf{x}_o$) and propagate it with the forward model while assessing its efficiency. For each case explored, the observed available fronts are split into two groups. The first is used as data assimilation input whereas the second group is reserved for evaluating the correctness of the forecast. In addition to the forecasting error, to evaluate the appropriateness of the system, the lead time must be computed. The lead time is defined as the period of validity of the model once the computational time is subtracted. In order for the model to be operative, the lead time needs to be positive. For the current exploration, we considered the forecast to be valid as long as the SDI' index was kept over 0.85.
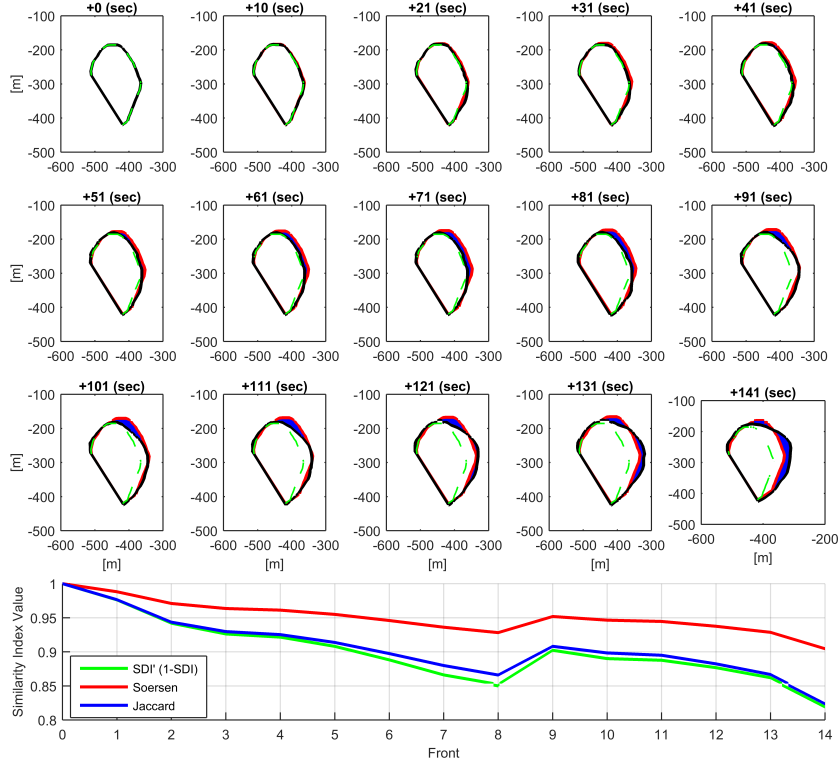
FIGURE 2.17: Forecasting error for Woodland site when six fronts are assimilated: (a) Similarity index for every 10-s spanned front. The lead time is 130 s (until 13th forecast front). For lines definition, see Figure 2.13 caption.

Following the already defined scenarios, when five isochrones are assimilated every 10 s in the Woodland site, the lead time reaches 130 s as the similarity index gradually decreases, reaching the 0.85 threshold (see Figure 2.17). The forecast run for the Shrub site, when six fronts are assimilated, yields a larger lead time. Although the similarity indices decrease to 0.8 at 50 s, they stabilise at 0.9 for the rest of the available fronts, as shown in Figure 2.18.

From this point on, the fire behaviour in the right flank changes, accelerating the rate of spread, and the forecast front is therefore underpredicted. The underlying cause might be due to the fact that Rothermel's model was initially derived for surface fire spread. As fire propagates through crowns, the model is no longer valid, although a proper assimilation structure extends its validity as fire in the Woodland site was erratically torching and crowning.

It is important to mention that the forecasting step can be improved by adding extra layers of information, particularly those that can change in time and space (i.e. fuels
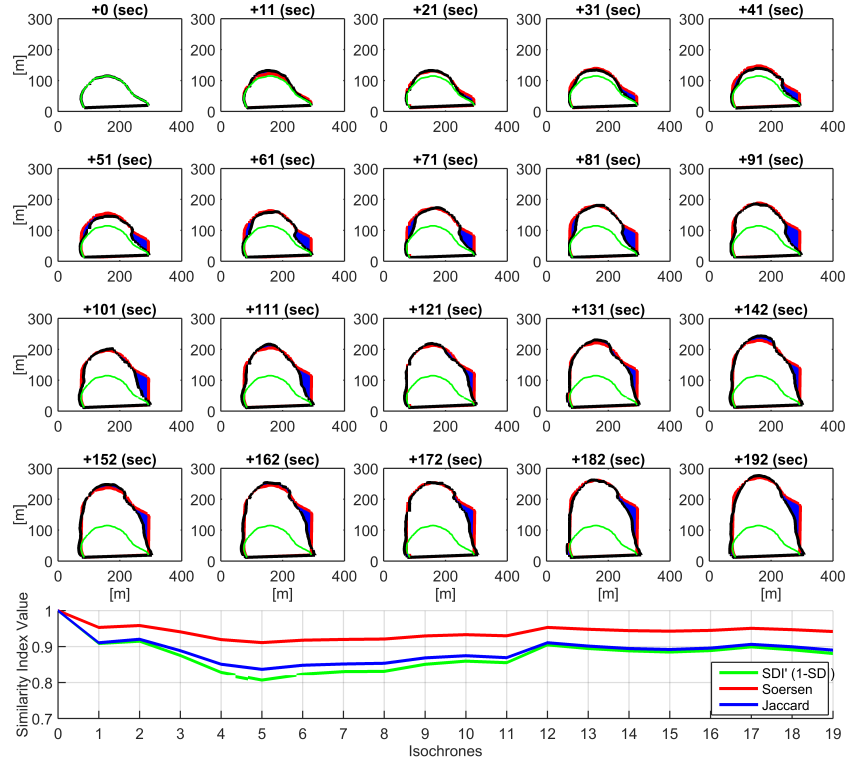
FIGURE 2.18: Forecast fronts and error for shrub Shrub site (six fronts assimilated with 10-s spacing each). Lead time reaches 190 s as all available fronts are predicted within the accepted error, predicting all the available fronts. SDI is shape deviation index. For lines definition, see caption of Figure 2.14.

depth map, wind speed and direction, etc.). If those layers are available, the lead time can be improved as the present algorithm cannot resolve time-dependent variables.

2.2.2.4   Assimilation window exploration

If the algorithm is to be used operationally, the consequences of the amount and frequency of assimilated data need to be investigated. The assimilation window (AW) is defined as the number of fronts assimilated before launching a forecast. To explore this, we use the Shrub site to run our assimilation and forecasting system 12 times, changing the number of assimilated fronts. The assimilation window varies from 8 assimilated fronts (i.e. 70 s) up to 19 assimilated fronts (i.e. 180 s) in the last run. The last assimilated front is kept constant for all runs and corresponds to 190 s after ignition. Then, the forecast is launched to generate the 18 fronts corresponding to the 180 s left of observed data. The areal relative error is computed to assess the validity of each forecast.

The results are shown in Figure 2.19 As more fronts are assimilated, the SDI' index tends to stabilise at 0.87. It is interesting to note that, for the current experiment, there is an optimum AW at ≈ 11–12 assimilated fronts where the lead time is maximised and reaches 180 s. As more than 12 fronts are assimilated, the fronts corresponding to times close to ignition are taken into account. In particular, when 17 fronts are assimilated, the first assimilated isochrone corresponds to 10 s after the ignition. At this stage, the fire is still driven by initial acceleration effects, and thus the forecast slightly overestimates the RoS and the forecasting areal error tends to grow, dramatically decreasing the lead time.
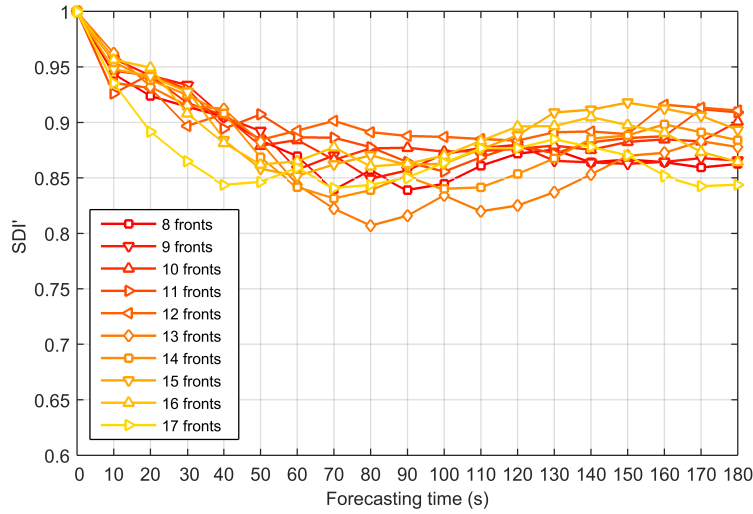
FIGURE 2.19: Forecasting areal relative error for shrub site and 10 different assimilation windows (from 70 to 160 s, assimilating one front every 10 s). SDI is shape deviation index

2.2.3   Conclusions

We presented a data assimilation framework that enhances a Rothermel-based model to make it capable of forecasting short-term wildfire dynamics even when using it beyond its original applicability scenarios. The assimilation framework previously applied to synthetically generated data has been improved to deal with real scenarios and is tested in two large-scale shrubland fire experiments conducted in South Australia, yielding similarity index scores over 0.8 and obtaining positive lead times of 130 and 160 s, depending on the scenario. The system performance when changing the assimilating windows has been also explored to conclude that the more fronts are assimilated, the better the forecast validity, up to a point where the initial fire dynamics perturb the forecast.

Even though in the present work the available fire fronts data lasted a few minutes and extended up to several hundred metres, the algorithm structure is envisaged to be applicable at larger time and space scales. At the operational level, the time required to perform image processing tasks such as georeferencing and fire front edge detection must be considered as they will directly affect the lead time. Author's contributions to this field, out of the scope of the thesis at hand, could be found in Valero et al. (2015) and Valero et al. (2016).

The present proof of concept demonstrated the promising potential of the assimilating and forecasting framework at hand. However, if the system is to be deployed operationally, some necessary features must be added and their performance evaluated:

- Firstly, the invariant I_{mfw} must be revisited to input a more detailed representation of fuel characteristics independently. To perform this step a revision of standard fuels model should be conducted to define the proper invariants (see chapter 3).

- Secondly, non-homogeneous high resolution spatial-dependant topographic winds must be coupled with the model as they have a crucial impact on the fire front development. A methodology to rapidly obtain such maps must also be explored (see chapter 4).

- Thirdly, the algorithm must be prepared to handle non-flat terrain scenarios by interacting with digital terrain models and solving the slope-wind interaction. This will be done in chapter 5.

Chapter 3

On fuels modelling

After setting the basis for a data-driven wildfire spread simulator the need for a better treatment of fuel characteristics was identified as one of the features to be improved in order to handle operational situations. In this chapter the standard fuel model, based on the original 13 Albini (1976a) and Anderson (1982) fuels, is revisited to identify the most important components to be transformed as invariants in a data-driven framework. Advanced analysis of variance (ANOVA) strategies are employed to deeply explore standard fuel model formulation dependencies. Finally, a reduced model that contains 5 fuel variables is proposed together with the translating framework to convert standard fuel models sets to the reduced formulation and vice-versa.

3.1   Introduction

Fuel elements are the ultimate combustible that burns during a forest fires. They are distributed in a range of different shapes, dimensions and arrangements. Thus, it is needed to group fuel elements in a more identifiable and useful way: the *Fuel Type*. Fuel type is defined as "an identifiable association of fuel elements of distinctive species, form, size arrangement, and continuity that will exhibit characteristic fire behaviour under defined burning conditions" Merrill and Alexander (2003). To generate fuel types one needs to classify similar vegetation structures and quantify their main properties in order to group regions with similar fire behaviour. Fuel types, however, are not designated specifically to be used in a given fire propagation model. They must be defined in a way that their physical properties can be input to the simulator to properly capture the important features relevant to the model. That is, for example, if the model resolve the surface-to-crown fire transition, the fuel type description must incorporate the crown's height among others. In other words, a fuel model is the numerical description of the

physical parameters that characterize each fuel type for a given fire propagating system (Arroyo et al., 2008). Fuel types may be similar in different ecosystems, but may have different fuel models associated with them (Chuvieco et al., 2003).

There exist three principal sets of fuel models used for fire modelling. They were generated in Canada, Australia and US and adapted (or directly transposed) to the rest of regions and ecosystems where fire has been modelled and studied (Arnaldos et al., 2004). Their main difference relies on the ecosystems described (as those are different in every region) as well as the propagation model they are generated for. As an example, fire models used for the Canadian Forest Service to generate their Forest Fire Behaviour Prediction system (FBP, Taylor et al. (1996)) have been updated to directly incorporate the geometrical properties of a Huygens based spreading algorithm. Table 3.1 shows the Canadian fuel models (first column) and its numerical simulating equivalence. Note that columns 2, 3 and 4 correspond to the variables of equation 2.4, 2.5 and 2.6 stated in chapter 2.

Fuel type	a_{ROS}[a]	b_{ROS}[a]	c_{ROS}[a]	BUI_0[b]	q[c]	Max. BE[d]	CBH[e] (m)	CFL[f] (kg/m^2)
C-1	90	0.0649	4.50	72	0.90	1.076	2.0	0.75
C-2	110	0.0282	1.50	64	0.70	1.321	3.0	0.80
C-3	110	0.0444	3.00	62	0.75	1.261	8.0	1.15
C-4	110	0.0293	1.50	66	0.80	1.184	4.0	1.20
C-5	30	0.0697	4.00	56	0.80	1.220	18.0	1.20
C-6	30	0.0800	3.00	62	0.80	1.197	7.0	1.80
C-7	45	0.0305	2.00	106	0.85	1.134	10.0	0.50
D-1	30	0.0232	1.60	32	0.90	1.179	NA[k]	NA
M-1/M-2[g]	C-2 110	C-2 0.0282	C-2 1.50	50	0.80	1.250	6.0	0.80
	D-1 30	D-1 0.0232	D-1 1.60					
M-3[h]	M-3 120	M-3 0.0572	M-3 1.40	50	0.80	1.250	6.0	0.80
	D-1 30	D-1 0.0232	D-1 1.60					
M-4[h]	M-4 100	M-4 0.0404	M-4 1.48	50	0.80	1.250	6.0	0.80
	D-1 30	D-1 0.0232	D-1 1.60					
O-1a[ij]	190	0.0310	1.40	1	1.00	1.000	NA	NA
O-1b[ij]	250	0.0350	1.70	1	1.00	1.000	NA	NA
S-1	75	0.0297	1.30	38	0.75	1.460	NA	NA
S-2	40	0.0438	1.70	63	0.75	1.256	NA	NA
S-3	55	0.0829	3.20	31	0.75	1.590	NA	NA

TABLE 3.1: Fuel Models of the Canadian Forest Fire Behaviour Prediction System. The reader is directed to the source for the meaning of the rest of the abbreviations. Source: Tymstra et al. (2010)

The most widely used set of fuel models is the one published by the US Forest Service in 1972 by Rothermel (1972a) and lately updated by Scott and Burgan (2005b). As it was developed as the core fuel model set for Rothermel propagation system, which, in turn, provides the underlying spread equations for our forward model, this set is the one that will be explored in depth.

The ultimate aim of such exploration is that in chapter 2, the I_{mfw} invariant was chosen after the linear relation was found when exploring Rothermel model dependency

on fuel model inputs (see Figure 2.2). Such a simplification prevented to track independently every physical magnitude that effected the rate of spread as they were all cast in one value. On the contrary, now we aim to simplify the model yet keeping as many physical magnitudes as possible.

3.2 Original Rothermel Fuel Modelling Approach

Rothermel (1972a) proposed to break down fuels types into unit cells wherein the fuel mean values were characterized. The classification ought to be performed upon fuel elements and particle size (surface-area-to-volume ratio SAV, particularly). This approach was simplified by reducing fuel types into 5 classes that gather similar SAV and fuel elements and categorize the different fuel moisture time-lag: 1h, 10h, 100h, Live Woody (LW), Live Herbaceous (LH). The first three classes were directly associated to fuel diameter sizes as presented in Table 3.2. Fuel particles properties were held constant the fuel particle properties (total and effective mineral content, heat content, and particle density) and established a constant extinction moisture content (m_x) for all classes. The former 11 models where rapidly complimented with a model characterized by dead understory cover, and a model for highly flammable grass documented by Albini (1976b) to become what is now called the original 13 fuel models. Albini (1976a) also proposed to calculate the extinction moisture content individually as a weighted sum of 1h, 10h and 100h classes and gave weighting values depending on the fuel model. The 13 fuel model set was finally incorporated in BEHAVE Andrews (1986) and FARSITE Finney (1999) with little modification. Despite their extensive use and great potential showed for both, estimating fire propagation and calculating fire danger indexes, Scott and Burgan (2005b) revisited the approach creating a new set of 62 models to better describe more precisely canopy covers and include the heterogeneity of fuel types (see Table 3.5). The aim was to better represent fuel configurations, improve the accuracy of fire behaviour predictions and increase the ability to model fuel treatment and changes in fuelbeds. This new set is now called: the Standard Fire Behaviour Fuel Models.

The quantifiable inputs of the model where identified by 9 parameters, which are:

- the ovendry fuel loading: W_o [kg m^{-2}]

- fuel depth: δ [m]

- surface-area-to-volume ratio: SAV [m^{-1}]

- fuel heat content: ΔH [kJ kg^{-1}]

- fuel ovendry particle density: ρ_p [kg m^{-3}]

- fuel moisture content: M_f [%]

- fuel moisture of extinction: M_x [%]

- fuel total mineral content: S_t [%]

- fuel effective mineral content: S_e [%]

Additionally, for each of the five particles classes, the characterized fuel load ($w0$), surface-area-to-volume ratio SAV and moisture content (m_f) must be provided. However, SAV values for all classes (except for the thinner one, i.e. $0.0 - 0.6$ cm) were found to have little variation and were considered constant for the whole fuel models set (see values in Table 3.2). Hence, the total number of description parameters for the Rothermel fuel model reaches 24 although 8 were given a constant values throughout the whole fuel model set. All the variable and constant parameters are reported in Table 3.2. Note that the moisture content values for each particle class (m) are added for completeness but do no directly belong to a given fuel model. Their value can vary independently for every model although *Moisture Scenarios* might be created to define dead and live particle's moisture content at once (Scott and Burgan, 2005b). Note that the original Rothermel fire spread model allowed for class-dependent additional variables as: heat content, effective and total mineral content. However, those are considered constant for every fuel particle class in both the original and the standard fuel model (SFM) set. The reason might be the (still current) difficulties on characterizing those magnitudes and its variability.

In order to adapt the Standard Fire Behaviour Fuel Model set to a data assimilation schema, the number of parameters describing every fuel model must be simplified. The reason is that a real-time system that optimizes 17 invariants (15 from the standard model plus wind speed and direction) is computationally costly and unstable as the parameters domain is of high dimension.

Hence, the proposed strategy is to focus on particle classes (1h, 10h, 100h, Live Woody and Live herbaceous) to explore their influence on the final RoS value calculated with the Rothermel equation (see equation 2.1). To perform this exploration a sensitivity analysis is thus required.

TABLE 3.2: Constant and variable Rothermel inputs for Standard Fuel Model Scott and Burgan (2005b). Heat content is considered homogeneous for all particle classes. W_0 is the fuel load, SAV the surface-area-to-volume ratio and m_f the fuel moisture. The abbreviation *var* stands for variable.

Particle properties

Particle classes	W_o [kg m^{-2}]		SAV [m^{-1}]		M_f[%]	
Dead 1h (0.0–0.6cm)	$w0_{d1}$	var	sv_{d1}	var	m_{d1}	var
Dead 10h (0.6–2.5cm)	$w0_{d2}$	var	sv_{d2}	357	m_{d2}	var
Dead 100h (2.5–7.6cm)	$w0_{d3}$	var	sv_{d3}	98	m_{d3}	var
Live Herbs	$w0_{lh}$	var	sv_{lh}	var	m_{lh}	var
Live Woody	$w0_{lw}$	var	sv_{lw}	var	m_{lw}	var

Fuel bed properties

Depth	d	var	[m]
Moisture of Extinction (dead fuel)	M_x	var	[%]
Ovendry Particle density	ρ	513	[kg m^{-3}]
Total mineral content	s_t	5.5	[%]
Eff. mineral content	s_e	1	[%]
Heat content (x5 classes)	ΔH	18609	[kJ kg^{-1}]

3.3 Sensitivity analysis of Rothermel model

There are multiple works in the literature that performed an in-depth sensitivity analysis on Rothermel model parameters. Bachmann and Allgöwer (2002) reduced the independent variables to 6 and used Taylor series to find the inputs correlation matrix for a particular fuel model (see Table 3.3). Jimenez et al. (2008) enlarged the study to 13 input variables using a sensitivity derivative-enhanced sampling method to accelerate the convergence of the classical Monte Carlo method. They applied it to *Chaparral* and *Short Grass* standard fuel models (from Scott and Burgan (2005b)) to evaluate the three Rothermel outputs (RoS, effective wind speed and spread direction). However, as the aim was to speed up Monte Carlo process rather to study all dependences, only two dependencies where investigated.

TABLE 3.3: Correlation matrix for Rothermel reduced inputs. Source: Bachmann and Allgöwer (2002)

Coniferous	$w0_{d1}$	$w0_{d2}$	$w0_{lw}$	sv_{d1}	$depth$	mx
$w0_{d1}$	1.000					
$w0_{d2}$	0.678	1.000				
$w0_{lw}$	0.161	0.172	1.000			
sv_{d1}	0.403	-0.085	-0.063	1.000		
$depth$	0.224	0.128	0.702	-0.242	1.000	
mx	-0.199	0.001	-0.177	0.079	-0.615	1.000
Burned areas	$w0_{d1}$	$w0_{d2}$	$w0_{lw}$	sv_{d1}	$depth$	mx
$w0_{d1}$	1.000					
$w0_{d2}$	-0.170	1.000				
$w0_{lw}$	0.550	-0.336	1.000			
sv_{d1}	0.280	-0.175	-0.106	1.000		
$depth$	0.439	-0.076	0.733	-0.082	1.000	
mx	0.590	0.010	0.393	0.119	-0.135	1.000
Chestnut	$w0_{d1}$	$w0_{d2}$	$w0_{lw}$	sv_{d1}	$depth$	mx
$w0_{d1}$	1.000					
$w0_{d2}$	-0.334	1.000				
$w0_{lw}$	-0.072	0.022	1.000			
sv_{d1}	0.261	-0.388	-0.356	1.000		
$depth$	0.025	0.029	0.832	-0.321	1.000	
mx	0.189	0.061	0.017	-0.124	-0.282	1.000

Liu et al. (Liu et al., 2013, 2015a) extended Jimenez et al. work using global sensitivity analysis to identify and rank the important input parameters and reduce the dimensionality of the model. They performed a deep Monte Carlo exploration with an analysis of variance and applied it to 6 standard fuel models. The parameters that are class-dependant (i.e. surface-area-to-volume ratio (SAV), fuel load (W_o) and moisture content (M_f) were grouped to look for its sensitivity as some particle classes might be absent in a given model. They found that ovendry particle density (ρ), mineral contents and slope do not have much effect on RoS. Inversely, the wind speed, fuel depth and heat content were the most susceptible ones (Figure 3.1). The low sensitivity of slope (slp in Figure 3.1) might be due to the base value taken to perform the sensitivity (14.04 deg) although it is not discussed in depth. The Sobol's global sensitivity analysis (Sobol', 2001) showed that the selected reduced parameters (see Figure 3.1) were independent as the overall index sum converge to 1 Liu et al. (2015a).

Liu et al. also performed a Sobol's analysis on the complete set of 24 Rothermel parameters (Liu et al., 2015b) although only for the *chaparral* Rothermel original fuel model (A4 in Table 3.6). They identified 7 parameters that were tagged as significant since they contributed together up to 97% of the total variance. Those parameters were: Wind speed (wsp), fuel depth (d), surface-area-to-volume ration and fuel load of 1h
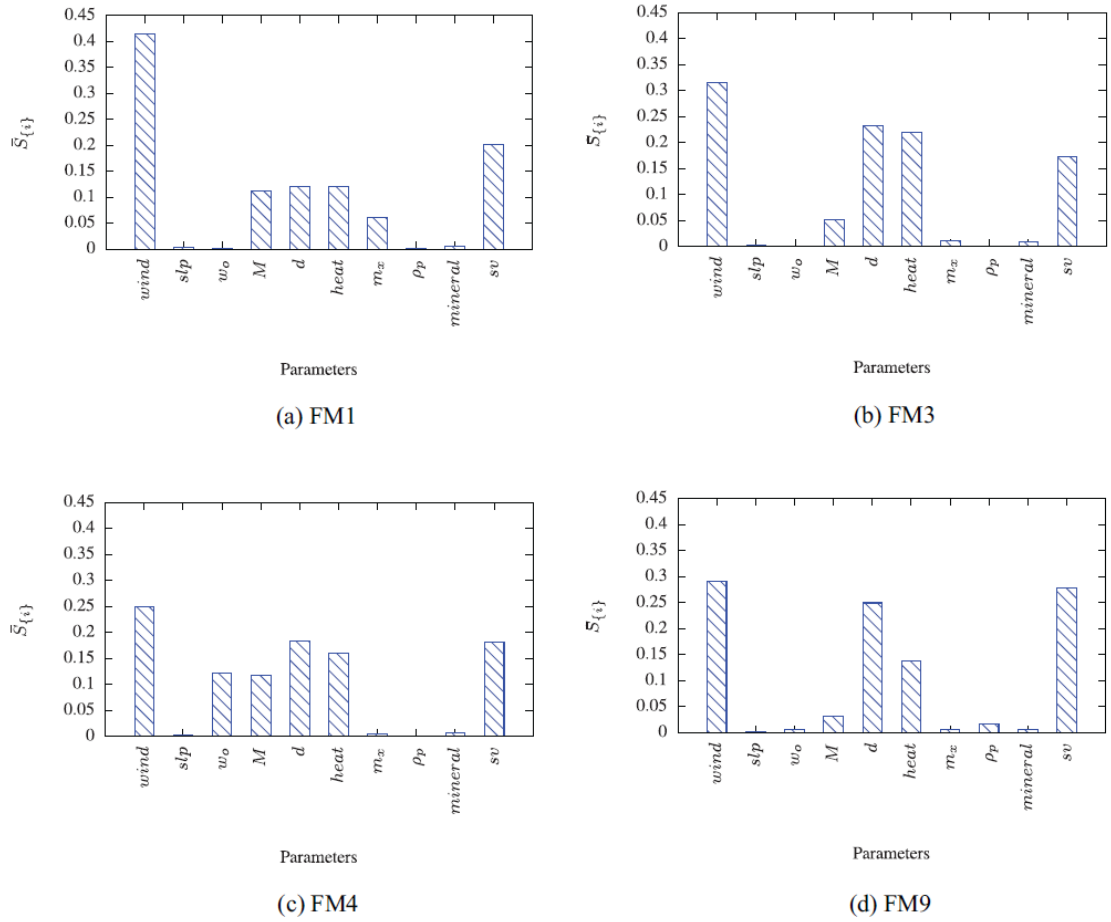
FIGURE 3.1: RoS Sobol' global sensitivity index ($\bar{S}$, y axis) for 4 different fuel models. FM1: *short grass*; FM3: *tall grass*; FM4: *chaparral*; FM9: *hardwood litter*; FM10, *timber*; FM11, *light logging slash*). *Mineral* corresponds to the effective mineral content variable as used in this thesis. Source: Liu et al. (2015a)

particle class (sv_{d1} and $w0_{d1}$), heat content (ΔH) and moisture content and fuel load of live woody class (m_{lw} and $w0_{lw}$). Their contribution to RoS is depicted in Figure 3.2a.

It has to be noted that the wind direction (θ) is the angular alignment to the up-slope direction as this is the way wind-slope interaction is resolved by Rothermel. Its influence is highly noticeable in the direction of maximum spread, a secondary Rothermel output (see (b) plot in Figure 3.2) that have a dramatic effect when used in a shape propagating method (i.e. as Huygens expansion, introduced in section 2.1.1). Thus, it must also be considered when the fuel model is to be used for a fire spread simulator.
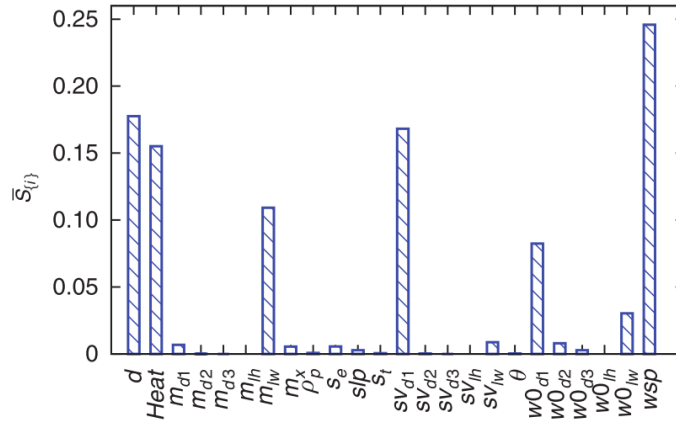
(a) RoS
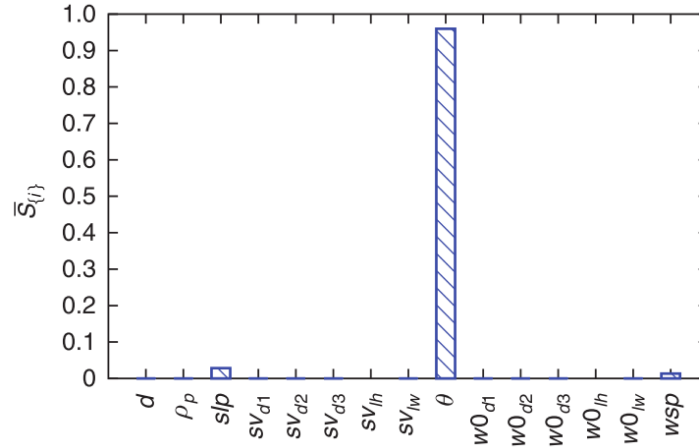
(b) direction of maximum spread

FIGURE 3.2: Sobol' global sensitivity index for RoS (a) and maximum spread direction (b). All parameters from fuel model FM4 (*chaparral*) are considered. The seven *significant* parameters stand out from the rest of model variables. Source: Liu et al. (2015b)

Whereas those studies give interesting insights on variables dependency, they do not answer the question of how much important the sub model classes are (1h, 10h, 100h, live wood and live herbaceous) on the final RoS characterization as those subclass parameters were grouped and discarded or only evaluated on a given fuel model for a given static moisture and weather scenario and slope. To answer this, and improve the integration of the standard fuels models to a data assimilation schema, we will perform a fuel particle class sensitivity study with Rothermel fire spread model and Standard fuels package (Ascoli et al., 2015) in **R** software (Team, 2008).

3.4 Sub fuel class influence in standard fuels model

To carry out this study, we compute the RoS for the standard fuel models generated by Scott and Burgan (2005b). We take the 53 burnable ones, as the non-combustible models (5) are not of our interest. All models are briefly described in Table 3.5.

The study is performed as follows: First, only 1h time-lag fuel class is taken into account and the following class is added subsequently until all 5 standard classes are considered (1h, 10h, 100h, Live Woody, Live Herbaceous). Those virtual classes analysed are summarized in Table 3.4.

TABLE 3.4: Virtual fuel classes generated to study the influence of fuel sub classes. Recall that *LH* stands for and *Live Herbaceous* and *LW* means *Live Woody*.

Virtual Composite Fuel Class	Containing Fuel Classes
VC1	1h
VC2	1h, 10h
VC3	1h, 10h, 100h
VC4	1h, 10h, 100h, *LH*
VC5	1h, 10h, 100h, *LH*, *LW*

Whereas it seems that conducting the sensitivity analysis in such a way will alter the total fuel load of each fuel model the justification relies on the way Rothermel's model incorporates all sub fuel classes. For all class-dependent variables (i.e. fuel load ($w0$), surface-to-area-volume ratio (sv) and moisture content (m_f)) a weighted averaged or a mean value is calculated. The same is also performed to secondary variables (i.e. no direct inputs) such us bulk density (ρ_b) or the packing ratio (β). For the fuel load, as example, the total fuel load of a given fuel model is an averaged sum weighted by surface area ratio between its class and the average surface of the particles contained in the model. Thus, for most of the fuel models, the weighted final fuel load is mainly influenced by thinner elements as they have more area for the same fuel load. Applying this strategy, if a sub-class other than the 1 h time-lag is removed from a given fuel model, the total fuel load is mostly maintained. On the other hand, some other parameters (as the overall moisture content or mean surface-area-to-volume ratio has changed. Those changes are what we want to explore.

It is important to highlight that standard fuel models that contain a live herbaceous component are defined as *dynamic* and *static* (Scott and Burgan, 2005b). For the *dynamic* fuel models the live herbaceous fuel load its dynamically transferred to a new dead fuel class as function of the live herbaceous moisture content as described by Burgan (1976). The moisture content of this new dead herbaceous class is taken equal to that of 1h time-lag, the SAV value is conserved from the live herbaceous original class

and the transferred load is calculated as showed in Figure 3.3.
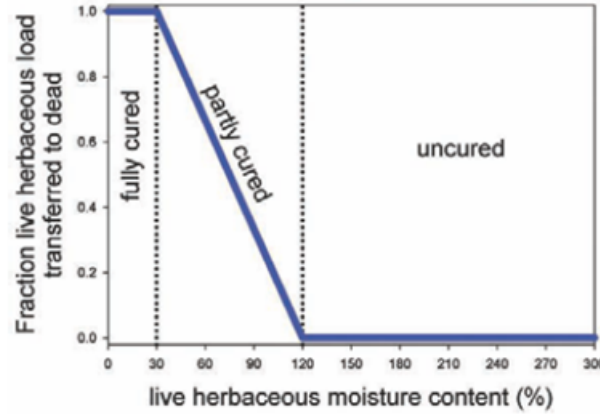
FIGURE 3.3: Graphical representation of dynamic live-to-dead herbaceous fuel trans-
ferring process. Thresholds are set to 120% and 30%. (Source: Scott and Burgan
(2005b))

To perform the analysis the wind and slope are systematically changed within a range
of 0 to 30 km h^{-1}and 0 to 20 deg respectively. However, recalling fundamental Rother-
mel's equations, the wind and slope are set as factors multiplying the fuel contribution
to RoS (see equations 2.2). Thus modifying this two parameters did not altered the
sensitivity dependencies. Thus the arbitrary value of 10 km h^{-1}and the slope to 4%.

The moisture content of each sub-class is initialized according to D1L1 (D1 for dead
fuels and L1 for live fuels) standard moisture scenario established by Scott and Burgan
(2005b) (see Table 3.7 and Table 3.6).

The results of the sensitivity analysis are depicted as follows: first, a grouped bar
plot Figure (3.4) with the absolute RoS value per each sub class modified standard fuel.
Secondly, a normalized stacked bar plot, Figure 3.5 where all RoS corresponding to the
same SFM class are stacked, where the initial standard model RoS with only 1h fuel is
used as a normalizing reference. Lastly, two RoS normalized box plot (Figure 3.6) are
computed to capture the statistical parameters. Note that if a fuel model is defined as
Static (i.e. no live herbaceous load), then the RoS evaluation incorporating this class
(yellow bars in plots 3.4, 3.5 and 3.7) will be the same as the previous one (orange bars)
as no contribution will be made when *Live Herbaceous* subclass is added.

Figure 3.4 already highlights the significant influence of the subclass fuel models.
Whereas RoS can reach values of 150 [m min^{-1}] with only 1h particle class (model A4
for example) it is lowered to less than 40 [m min^{-1}] when all subclasses added. In all
cases, the RoS decreases as more subclass are added. This has a physical reasoning as

TABLE 3.5: Standard fire behaviour fuel models name and brief description. Source: Scott and Burgan (2005b)

Fudl Model Name	Description
Original Rothermel Fuel Models	
A1	Short Grass
A2	Timber Grass and Understory
A3	Tall Grass
A4	Chaparral
A5	Brush
A6	Dormant Brush
A7	Southern Rough
A8	Compact Timber Litter
A9	Hardwood Litter
A10	Timber Understory
A11	Light Slash
A12	Medium Slash
A13	Heavy Slash
Nearly pure grass and/or forb type (Grass)	
GR1	Grass is short, patchy, and possibly heavily grazed. Spread rate moderate; flame length low.
GR2	Moderately coarse continuous grass, average depth about 1 foot. Spread rate high; flame length moderate.
GR4	Moderately coarse continuous grass, average depth about 2 feet. Spread rate very high; flame length high.
GR7	Moderately coarse continuous grass, average depth about 3 feet. Spread rate very high; flame length very high.
GR1	Grass is short, patchy, and possibly heavily grazed. Spread rate moderate; flame length low.
GR3	Very coarse grass, average depth about 2 feet. Spread rate high; flame length moderate.
GR5	Dense, coarse grass, average depth about 1 to 2 feet. Spread rate very high; flame length high.
GR6	Dryland grass about 1 to 2 feet tall. Spread rate very high; flame length very high.
GR8	Heavy, coarse, continuous grass 3 to 5 feet tall. Spread rate very high; flame length very high.
GR9	Very heavy, coarse, continuous grass 5 to 8 feet tall. Spread rate extreme; flame length extreme.
Mixture of grass and shrub, up to about 50 percent shrub coverage (Grass-Shrub)	
GS1	Shrubs are about 1 foot high, low grass load. Spread rate moderate;flame length low.
GS2	Shrubs are 1 to 3 feet high, moderate grass load. Spread rate high; flame length moderate.
GS3	Moderate grass/shrub load, average grass/shrub depth less than 2 feet. Spread rate high; flame length moderate.
GS4	Heavy grass/shrub load, depth greater than 2 feet. Spread rate high; flame length very high.
Shrubs cover at least 50 percent of the site; grass sparse to nonexistent (Shrub)	
SH1	Low shrub fuel load, fuelbed depth about 1 foot; some grass may be present. Spread rate very low; flame length very low.
SH2	Moderate fuel load (higher than SH1), depth about 1 foot, no grass fuel present. Spread rate low; flame length low.
SH5	Heavy shrub load, depth 4 to 6 feet. Spread rate very high; flame length very high.
SH7	Very heavy shrub load, depth 4 to 6 feet. Spread rate lower than SH5,but flame length similar. Spread rate high; flame length very high.
SH3	Moderate shrub load, possibly with pine overstory or herbaceous fuel, fuel bed depth 2 to 3 feet. Spread rate low; flame length low.
SH4	Low to moderate shrub and litter load, possibly with pine overstory, fuel bed depth about 3 feet. Spread rate high; flame length moderate.
SH6	Dense shrubs, little or no herb fuel, depth about 2 feet. Spread rate high; flame length high.
SH8	Dense shrubs, little or no herb fuel, depth about 3 feet. Spread rates high; flame length high.
SH9	Dense, finely branched shrubs with significant fine dead fuel, about 4 to 6 feet tall; Spread rate high, flame length very high.
Grass or shrubs mixed with litter from forest canopy (Timber-Understory)	
TU1	Fuelbed is low load of grass and/or shrub with litter. Spread rate low; flame length low.
TU2	Fuelbed is moderate litter load with shrub component. Spread rate moderate; flame length low.
TU3	Fuelbed is moderate litter load with grass and shrub components. Spread rate high; flame length moderate.
TU4	Fuelbed is short conifer trees with grass or moss understory. Spread rate moderate; flame length moderate.
TU5	Fuelbed is high load conifer litter with shrub understory. Spread rate moderate; flame length moderate.
Dead and down woody fuel (litter) beneath a forest canopy (Timber Litter)	
TL1	Light to moderate load, fuels 1 to 2 inches deep. Spread rate very low; flame length very low.
TL2	Low load, compact. Spread rate very low; flame length very low.
TL6	Moderate load, less compact. Spread rate moderate; flame length low.
TL9	Very high load, fluffy. Spread rate moderate; flame length moderate. Fuelbed composed of long-needle pine litter.
TL8	Moderate load and compactness may include small amount of herbaceous load. Spread rate moderate; flame length low.
TL4	Moderate load, includes small diameter downed logs. Spread rate low; flame length low.
TL7	Heavy load, includes larger diameter downed logs. Spread rate low; flame length low.
TL3	Moderate load conifer litter. Spread rate very low; flame length low.
TL5	High load conifer litter; light slash or mortality fuel. Spread rate low; flame length low.
TL9	Very high load broadleaf litter; heavy needle-drape in otherwise sparse shrub layer. Spread rate moderate; flame length moderate.
Activity fuel (slash) or debris from wind damage (blowdown) (Slash-Blowdown)	
SF1	Fine fuel load is 10 to 20 tons/acre, weighted toward fuels 1 to 3 inches diameter class, depth is less than 1 foot. Spread rate moderate; flame length low.
SF2	Fine fuel load is 7 to 12 tons/acre, evenly distributed across 0 to 0.25, 0.25 to 1, and 1 to 3 inch diameter classes. Spread rate moderate; flame length moderate.
SF3	Fine fuel load is 7 to 12 tons/acre, weighted toward 0 to 0.25 inch diameter class, depth is more than 1 foot. Spread rate high; flame length high.
SB2	Blowdown is scattered, with many trees still standing. Spread rate moderate; flame length moderate.
SB3	Blowdown is moderate, trees compacted to near the ground. Spread rate high; flame length high.
SB4	Blowdown is total, fuelbed not compacted, foliage still attached. Spread rate very high; flame length very high.
Insufficient wildland fuel to carry wildland fire under any condition (Nonburnable)	
NB1	Urban or suburban development; insufficient wildland fuel to carry wildland fire.
NB2	Snow/ice.
NB3	Agricultural field, maintained in nonburnable condition.
NB8	Open water.
NB9	Bare ground.

the added particle classes are always less combustible (their SAV is smaller) and so it is the composite fuel model.
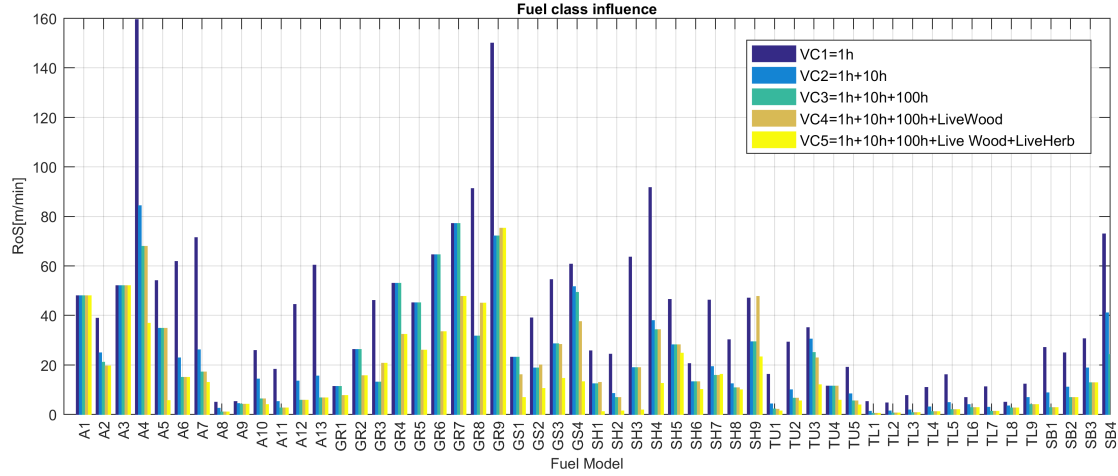
FIGURE 3.4: Absolute RoS [m min^{-1}] value for every fuel model subclass combination. For models name description see Table 3.5

Figure 3.5 allows to distinguish the particular influence of each particle fuel class over each standard fuel models as the RoS is normalized by 1h particle sub class fuel model. That is, the effect of adding subclasses is compared to the 1h class only (i.e. VC1). If the stacked sum reaches 5 (see model A1 for example) it means that no influence at all is observed. The largest reduction on RoS is identified from VC1 to VC2. That is, 10h time lag class is the one (apart from 1h class) that creates the highest difference. The box-plots (a) and (b) in Figure 3.6 help to validate this. The RoS normalized median value for all fuel models is statistically studied by plotting its values as 1st and 3rd quartile (blue box). The fuel models that do not originally contain a certain fuel particle class are removed for the statistical metrics calculations as they add a flawed contribution. RoS normalized value lowers roughly to a half when the 10h fuel class is added (red line in Figure 3.6a). The changes when the next 100h particle class is added (i.e. VC3) is of 15% (see (b) panel of Figure 3.6). The addition of live herbaceous class reduces the RoS value a median of 20% (right plot in Figure 3.6) but introduces a high scattering of results (i.e. for 50% of the fuels, the reduction ranges from 10% to almost 80 %). This is the class that introduces the largest statistical dispersion. The remaining live woody class has a median reduction of slightly more than 20% being the second higher (after the 10h class).
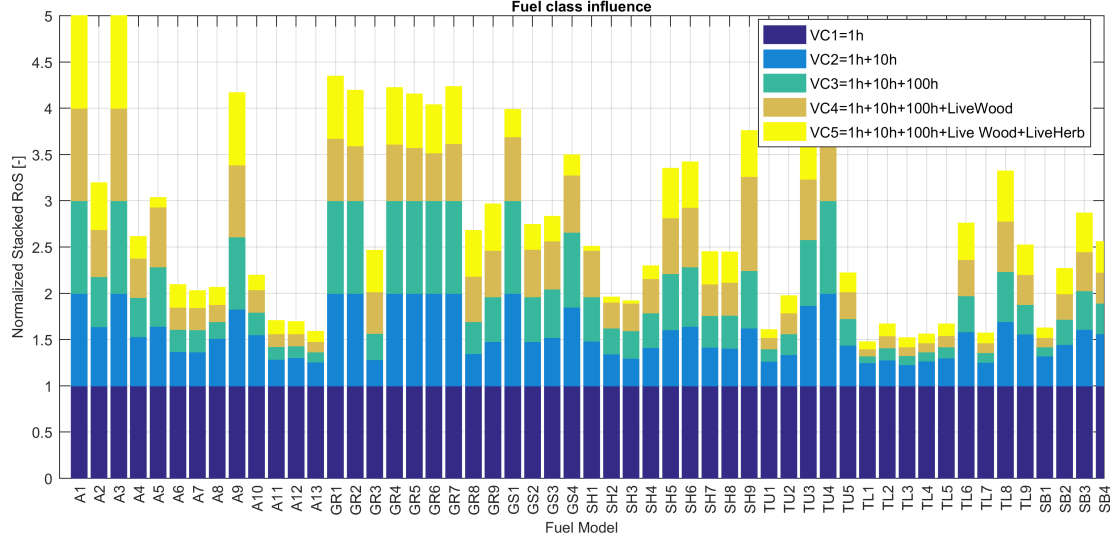
FIGURE 3.5: Stacked representation of 1h-class-normalized RoS [-] for each model and its Virtual Composite fuel classes as stated in Table 3.4. Note that y axis value is the normalized stacked summed of RoS. Equal stacked bar length reaching 5 means that the given model shows no dependency to addition of sub-classes.

(a) Normalized RoS

(b) Absolute difference on normalized RoS
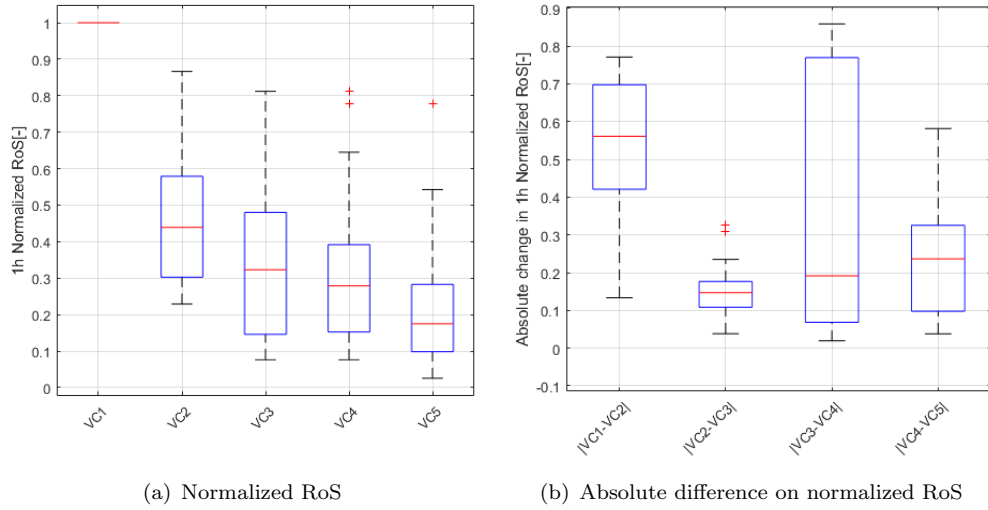
FIGURE 3.6: Box plot for normalized RoS (a) and virtual class absolute difference (b) of all standard fuel models versus each subclass fuel model. The blue box shows the limits of the 1st and 3rd quartile. Whiskers represent the 99.3% ($\pm2.7\sigma$) and the red line is the median value. The decreasing trend is clearly appreciated. The higher difference belongs to 10h class addition. Live Herbaceous inclusion (VC5) does not show much effect.

TABLE 3.6: SFM properties. Each fuel particle class variable is depicted in a colour scale to illustrate each parameter variability throughout the different models (green lower value, red higher value). D stands for Dynamic and S for Static models. Note that the heat content (h) is common for all models. Values are taken and converted to IS units from Scott and Burgan (2005b)

Id	Model		w_{d1}	w_{d2}	w_{d3} $\mathrm{tn \cdot ha^{-1}}$	w_{lh}	w_{lw}	s_{d1}	s_{d2}	s_{d3} $\mathrm{m^{-1}}$	s_{lh}	s_{lw}	δ m	m_x %	h $\mathrm{kJ \cdot kg^{-1}}$
1	A1	S	1,66	0	0	0	0	11483	0	0	0	0	0,3	12	18622
2	A2	S	4,49	2,25	1,12	1,12	0	9843	358	358	4921	0	0,3	15	18622
3	A3	S	6,74	0	0	0	0	4921	0	0	0	0	0,76	25	18622
4	A4	S	11,23	8,98	4,49	0	11,23	6562	358	358	0	4921	1,83	20	18622
5	A5	S	2,25	1,12	0	0	5,79	6562	358	0	0	4921	0,61	20	18622
6	A6	S	3,7	6,18	4,94	0	0	5741	358	358	0	0	0,76	25	18622
7	A7	S	2,72	4,69	3,71	0	0,91	5741	358	358	0	4921	0,76	40	18622
8	A8	S	3,37	2,25	5,61	0	0	6562	358	358	0	0	0,06	30	18622
9	A9	S	6,54	0,93	0,34	0	0	8202	358	358	0	0	0,06	25	18622
10	A10	S	6,74	4,49	11,23	0	4,49	6562	358	358	0	4921	0,3	25	18622
11	A11	S	3,37	10,1	12,35	0	0	4921	358	358	0	0	0,3	15	18622
12	A12	S	8,98	21,44	37,06	0	0	4921	358	358	0	0	0,7	20	18622
13	A13	S	15,72	51,66	62,89	0	0	4921	358	358	0	0	0,91	25	18622
14	GR1	D	0,22	0	0	0,67	0	7218	0	0	6562	0	0,12	15	18622
15	GR2	D	0,22	0	0	2,25	0	6562	0	0	5906	0	0,3	15	18622
16	GR3	D	0,22	0,9	0	3,37	0	4921	358	0	4265	0	0,61	30	18622
17	GR4	D	0,56	0	0	4,27	0	6562	0	0	5906	0	0,61	15	18622
18	GR5	D	0,9	0	0	5,62	0	5906	0	0	5249	0	0,46	40	18622
19	GR6	D	0,22	0	0	7,64	0	7218	0	0	6562	0	0,46	40	20934
20	GR7	D	2,25	0	0	12,13	0	6562	0	0	5906	0	0,91	15	18622
21	GR8	D	1,12	2,25	0	16,4	0	4921	358	0	4265	0	1,22	30	18622
22	GR9	D	2,25	2,25	0	20,22	0	5906	358	0	5249	0	1,52	40	18622
23	GS1	D	0,45	0	0	1,12	1,46	6562	0	0	5906	5906	0,27	15	18622
24	GS2	D	1,12	1,12	0	1,35	2,25	6562	358	0	5906	5906	0,46	15	18622
25	GS3	D	0,67	0,56	0	3,26	2,81	5906	358	0	5249	5249	0,55	40	18622
26	GS4	D	4,27	0,67	0,22	7,64	15,96	5906	358	358	5249	5249	0,64	40	18622
27	SH1	D	0,56	0,56	0	0,34	2,92	6562	358	0	5906	5249	0,3	15	18622
28	SH2	S	3,03	5,39	1,69	0	8,65	6562	358	358	0	5249	0,3	15	18622
29	SH3	S	1,01	6,74	0	0	13,93	5249	358	0	0	4593	0,73	40	18622
30	SH4	S	1,91	2,58	0,45	0	5,73	6562	358	358	0	5249	0,91	30	18622
31	SH5	S	8,09	4,72	0	0	6,52	2461	358	0	0	5249	1,83	15	18622
32	SH6	S	6,52	3,26	0	0	3,15	2461	358	0	0	5249	0,61	30	18622
33	SH7	S	7,87	11,91	4,94	0	7,64	2461	358	358	0	5249	1,83	15	18622
34	SH8	S	4,61	7,64	1,91	0	9,78	2461	358	358	0	5249	0,91	40	18622
35	SH9	D	10,11	5,51	0	3,48	15,73	2461	358	0	5906	4921	1,34	40	18622
36	TU1	D	0,45	2,02	3,37	0,45	2,02	6562	358	358	5906	5249	0,18	20	18622
37	TU2	S	2,13	4,04	2,81	0	0,45	6562	358	358	0	5249	0,3	30	18622
38	TU3	D	2,47	0,34	0,56	1,46	2,47	5906	358	358	5249	4593	0,4	30	18622
39	TU4	S	10,11	0	0	0	4,49	7546	0	0	0	6562	0,15	12	18622
40	TU5	S	8,99	8,99	6,74	0	6,74	4921	358	358	0	2461	0,3	25	18622
41	TL1	S	2,25	4,94	8,09	0	0	6562	358	358	0	0	0,06	30	18622
42	TL2	S	3,15	5,17	4,94	0	0	6562	358	358	0	0	0,06	25	18622
43	TL3	S	1,12	4,94	6,29	0	0	6562	358	358	0	0	0,09	20	18622
44	TL4	S	1,12	3,37	9,44	0	0	6562	358	358	0	0	0,12	25	18622
45	TL5	S	2,58	5,62	9,89	0	0	6562	358	358	0	0	0,18	25	18622
46	TL6	S	5,39	2,7	2,7	0	0	6562	358	358	0	0	0,09	25	18622
47	TL7	S	0,67	3,15	18,2	0	0	6562	358	358	0	0	0,12	25	18622
48	TL8	S	13,03	3,15	2,47	0	0	5906	358	358	0	0	0,09	35	18622
49	TL9	S	14,94	7,42	9,33	0	0	5906	358	358	0	0	0,18	35	18622
50	SB1	S	3,37	6,74	24,72	0	0	6562	358	358	0	0	0,3	25	18622
51	SB2	S	10,11	9,55	8,99	0	0	6562	358	358	0	0	0,3	25	18622
52	SB3	S	12,36	6,18	6,74	0	0	6562	358	358	0	0	0,37	25	18622
53	SB4	S	11,8	7,87	11,8	0	0	6562	358	358	0	0	0,82	25	18622

Finally, to asses if the moisture fuel scenario taken to initialize the models (D1L1) have a huge influence in the results, the stacked bar plot together with normalized RoS box plot and its variation along the virtual classes (same as shown in Figures 3.5 and 3.6) are generated for all possible moisture scenarios an depicted in Figures 3.7 and 3.9.

Those moisture scenarios are all possible combinations of live and dead standard moisture values proposed by Scott and Burgan (2005b) and gathered in Table 3.7.

TABLE 3.7: Composition of moisture scenarios proposed by Scott and Burgan (2005b). All values are percentage (water content over ovendry fuel mass).

Scenario	1h	10h	100h	LH	LW	Description
D1L1	3	4	5	30	60	Very dry dead FM, fully cured herb
D2L2	6	7	8	60	90	Dry dead FM, 2/3 cured herb
D3L3	9	10	11	90	120	Moderate dead FM 1/3 cured herb
D4L4	12	13	14	120	150	High dead FM, uncured herb
D1L2	3	4	5	60	90	Very dry dead FM, 2/3 cured herb
D1L3	3	4	5	90	120	Very dry dead FM, 1/3 cured herb
D1L4	3	4	5	120	150	Very dry dead FM, uncured herb
D2L1	6	7	8	30	60	Dry dead FM, fully cured herb
D2L3	6	7	8	90	120	Dry dead FM, 1/3 cured herb
D2L4	6	7	8	120	150	Dry dead FM, uncured herb
D3L1	9	10	11	30	60	Moderate dead FM, fully cured herb
D3L2	9	10	11	60	90	Moderate dead FM, 2/3 cured herb
D3L4	9	10	11	120	150	Moderate dead FM, uncured herb
D4L1	12	13	15	30	60	High dead FM, fully cured herb
D4L2	12	13	15	60	90	High dead FM, 2/3 cured herb
D4L3	12	13	15	90	120	High dead FM, 1/3 cured herb

The results of the analysis show that although the RoS value for each standard fuel changes when the moisture of its containing classes is altered (see Figure 3.7) the comparison by virtual classes remains extremely similar. The statistical outputs (median, 1st and 3rd quartile) of the normalized RoS over the different virtual classes are depicted as a box plot in Figure 3.8 per each scenario. They show almost identical results as for moisture scenario D1L1. The most remarkable difference is that all combinations containing the highest value for L4 live moisture scenario (i.e. 150% extremely uncured) show a larger divergence of results (larger box in Figure 3.9) when this live class is added (i.e. $|VC3 - VC4|$). The explanation for this effect might be the amount of highly incombustible matter introduced and the importance, in this specific case, of the live particle class load in the fuel model.
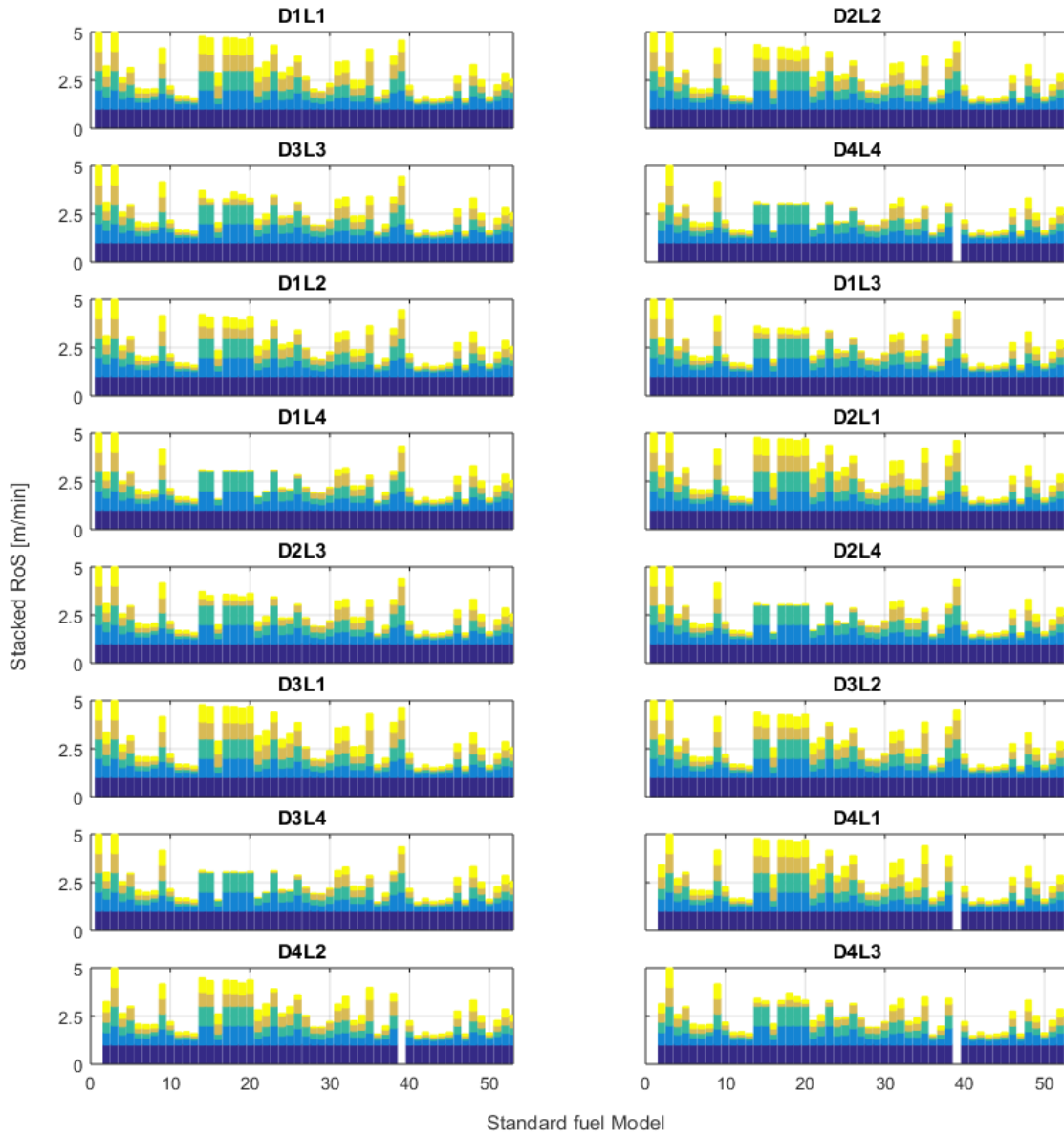
FIGURE 3.7: Composed stacked bar plot for all moisture scenarios. The influence of subclasses is recognized independently of the moisture scenario used as all the fuel model's RoS changes when additional subclasses are added. Horizontal axis displays the standard fuel model name ID as indicated in 3.6
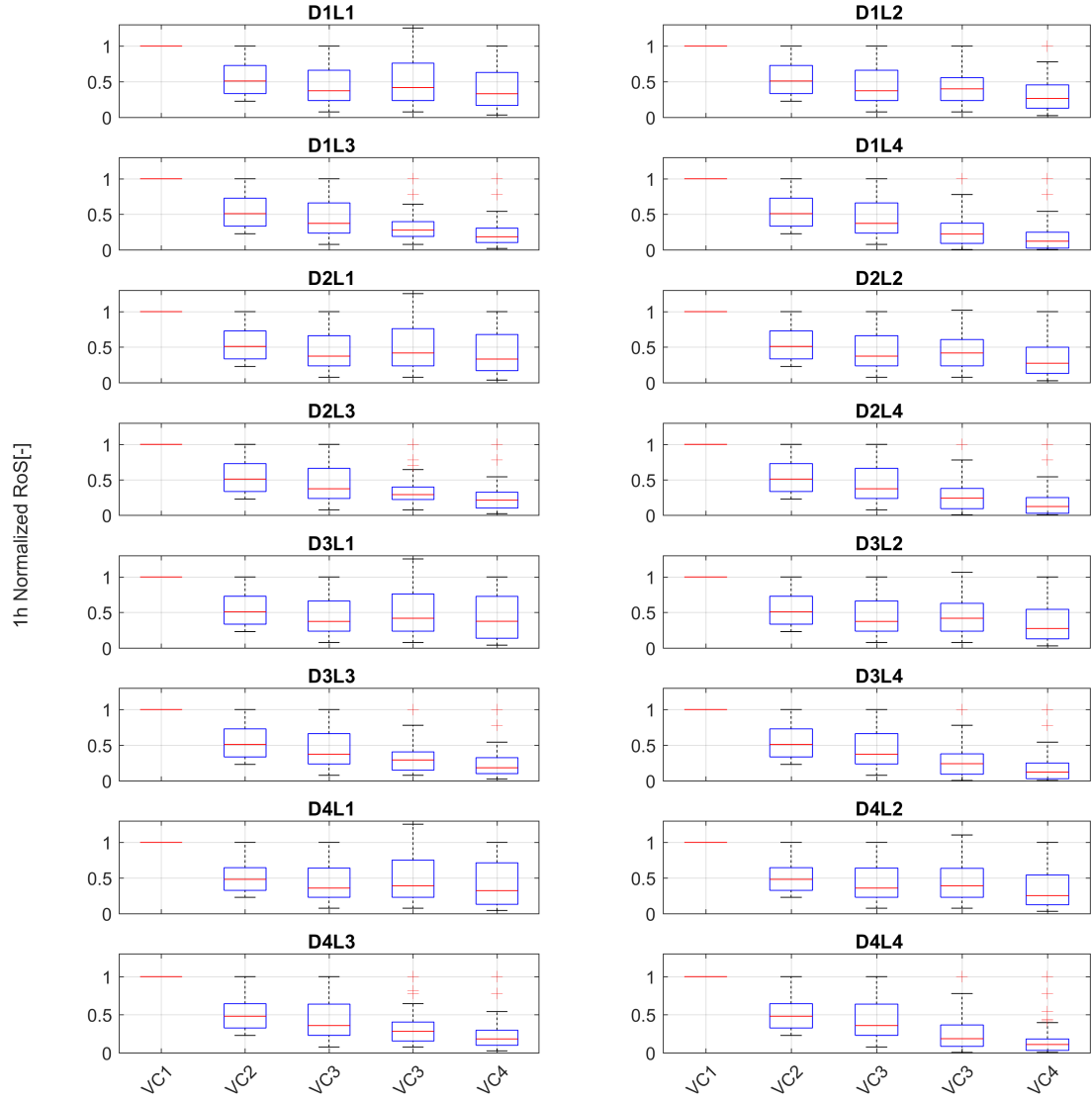
FIGURE 3.8: Normalized RoS, as shown in Figure 3.6 but now extended to all possible moisture scenarios. The blue box shows the limits of the 1st and 3rd quartile. Whiskers represent the 99.3% ($\pm 2.7\sigma$) and the red line is the median value.
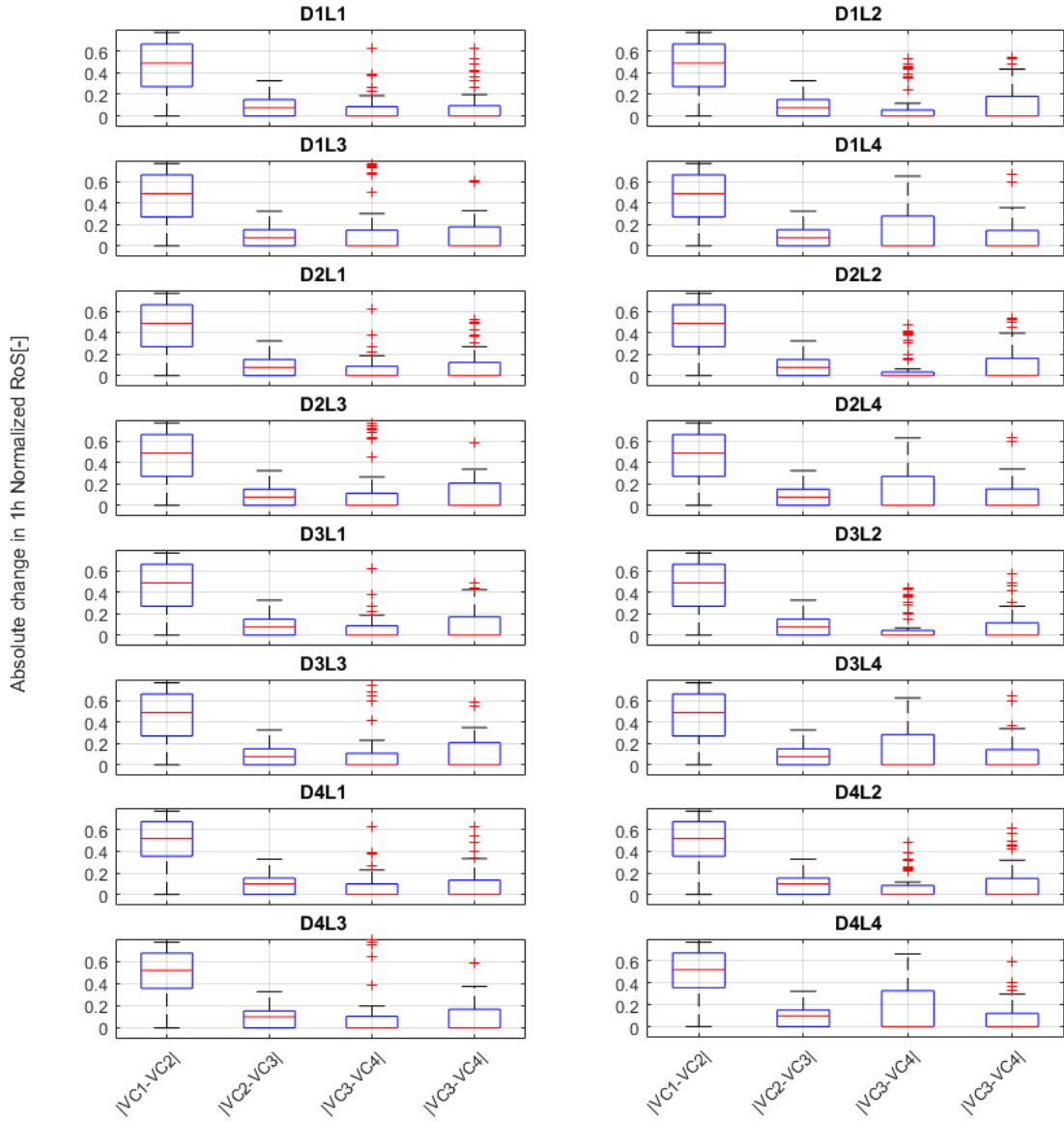
FIGURE 3.9: Absolute variation in normalized RoS, as shown in Figure 3.6b, but now extended to all possible moisture scenarios. The blue box shows the limits of the 1st and 3rd quartile. Whiskers represent the 99.3% ($\pm 2.7\sigma$) and the red line is the median value

3.5   Sobol' index exploration

The previous section helped to asses the variability and the influence of fuel particle classes in standard fuel models. To ultimately explore the dependency of each fuel parameter in a systematically way we expand the Liu et al. (2015b) work and conduct a Monte-Carlo Sobol' analysis for all 53 burnable standard fuel models. In addition, we enlarge the Rothermel variables exploration by incorporating the heat content for each fuel particle class, following original Rothermel's formulation and, thus, exploring a total of 27 different input variables.

3.5.1   Sobol' sensitivity index analysis

Sobol' analysis (Sobol', 2001; Sobol' et al., 2007) is a global sensitivity analysis method (GSA) based on analysis of variance (ANOVA) strategy that allows to quantify the significance of interactions between the input variables. It is based on the outputs decomposition of variance in a unique additive series of independent functions of increasing dimensionality (Salvador et al., 2001; Liu et al., 2013). In other words, ANOVA completely explores the parameter space of a given function (i.e. model) considering interactions among its parameters to examine how uncertainties on the input parameters affect the output and rank their influence on the overall result.

The mathematical formulation works is described in the following paragraphs:

Given any model ($Y = f(\{x_1, x_2, ..., x_n\})$) that depends on n independently and uniformly distributed variables within the unit hypercube (i.e. $x_i \in [0, 1]$), which is a normalization applicable to any variable set, the ANOVA-high dimensional model representation is given by:

$$f(\mathbf{x}) = f_o + \sum_{i_1} f_{\{i_1\}}(x_{i_1}) + ... \sum_{i_1 < i_2} f_{\{i_1, i_2\}}(x_{i_1}, x_{i_2}) + \\ + f_{\{i_1, i_2, ..., i_n\}}(x_{i_1}, x_{i_2}, ..., x_{i_n}) \tag{3.1}$$

where constant term f_o is the mean (expectation) of the model, $f_{\{i_1\}}(x_{i_1})$ represents the first order interactions, $f_{\{i_1, i_2\}}(x_{i_1}, x_{i_2})$ are those of second order and so on. This is an orthogonal decomposition in the sense that the total variance of the model is decomposed as sum of the variances of the component functions. The variance is then defined as:

$$\sigma^2 = \sum_{i \subseteq \{1, ..., n\}} \sigma_i^2 \tag{3.2}$$

where σ_i^2 is the variance of each $f_{\{i\}}$ function.

Sobol' (1990) introduced two global sensitivity indices for a given set of input parameters $u = \{1, ..., n\}$.

$$S_u = \frac{1}{\sigma^2} \sum_{v \subseteq u} \sigma_v^2 \qquad (3.3)$$

$$\bar{S}_u = \frac{1}{\sigma^2} \sum_{v \cap u = \varnothing} \sigma_v^2 \qquad (3.4)$$

The first order sensitivity index (S_u) (also known as main effect index) only considers the impact of each individual parameter to the function output, whereas the total effect index ($\bar{S}_u$) incorporates all the interactions between parameters. Those two indicators can be used as a lower (S_u) and upper ($\bar{S}_u$) bounds for the sensitivity index value (Liu et al., 2015b).

It must be remarked that the total number of partial variance to compute is of $2^n - 1$. To compute the expectation of those random parameters sets (e. g. $f_{\{i_1, i_2\}}$), Sobol' et al. (2007) designated a Monte-Carlo based algorithm to efficiently determine those indices. The algorithm uses two sets of randomly generated input parameters. To apply it to every standard fuel model, we considered uniform distributions with a standard deviation of 5% similar to what was performed by Liu et al. (2015b). Then, we can compute the first order sensitivity index S_u (hereafter Sobol' Index) for the 27 Rothermel's model input parameters using 10^6 runs. The computation of the indices for the 53 standard fuels models took 49 hours in a 44-core cluster.

3.5.2 Sobol' index results

The Sobol' index plots for the 53 burnable standard fuel models are attached in Appendix A. From all cases, a four representatives fuels models (*short grass*-A1, *chaparral*-A4, *hardwood litter*-A9 and *light slash*-A11) are depicted in Figure 3.10 and 3.11 and discussed in the following. The base moisture content scenario is the D2L2 gathered in Table 3.7. The wind speed (U), the fuel depth (δ), the SAV (sv_{d1}) and the low heat content (h_{d1}) of 1h time-lag class are the four more influencing parameters in all cases. For *light slash*, A11 fuel model, the sv_{d1} shows a lower value due to the lower net load of this class in this particular fuel model. After those most influencing four parameters the moisture content of 1h fuel class (m_{d1}), live woody heat content (h_{lw}) and some fuel particle loads are the rest of significant parameters depending on the model. For example, the 100h time-lag fuel load turns out to be the fifth more influencing parameter in A11, whereas in all other fuel models it has no influence. This is associated to the fuel load of this class in this particular model.

Those selected four fuel models illustrate the overall behaviour of the 53 fuels models. Whereas the four most influencing parameters are kept for the majority of the models (even though the rank order might be altered) the order of the remaining parameters depend largely on the fuel model particularities, e.g. the relation between fuel class loads or the presence or not of live matter.

(a) A1, short grass

(b) A4, chaparral

FIGURE 3.10: Sobol' index S_u (first order sensitivity index) for four original Rothermel fuel models A1 and A4. The SumChk in each plot title displays the sum of all indices. When its value gets closer to unity, it means the Monte-Carlo process has converged.
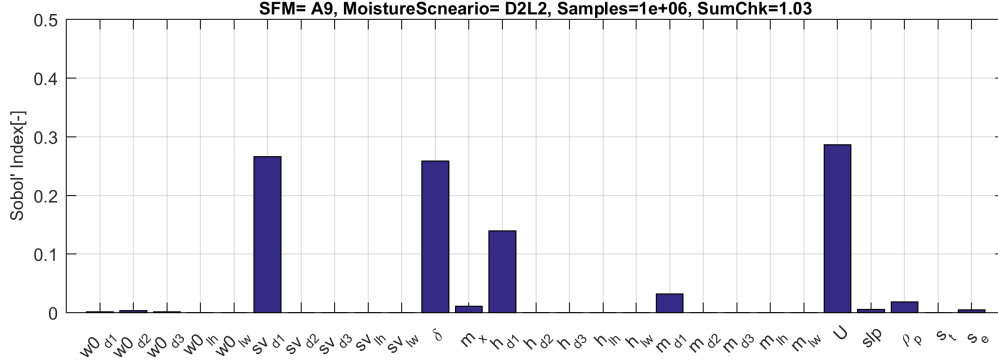
(a) A9, hardwood litter
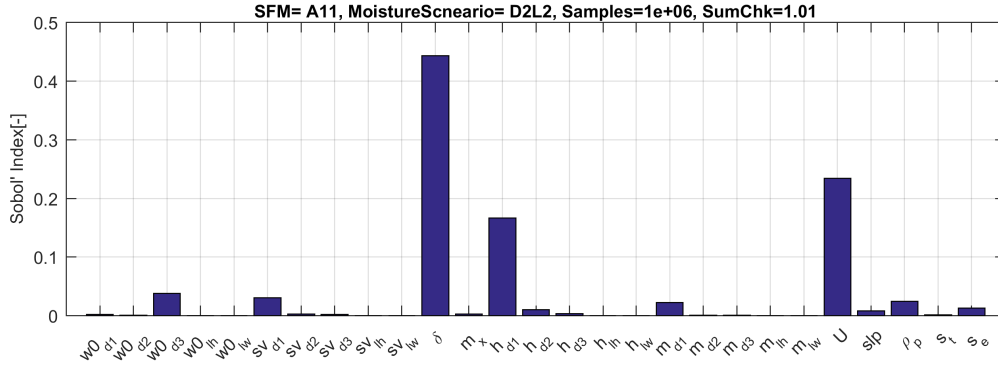
(b) A11, light slash

FIGURE 3.11: Sobol' index S_u (first order sensitivity index) for four original Rothermel fuel models A9 and A11. The SumChk in each plot title displays the sum of all indices. The more closer to unity, the better the Monte-Carlo process has converged.

To quantify a single Sobol' index value for all standard fuel models we perform a box plot statistical exploration of all Sobol' values gathered in appendix A. Results are displayed in Figure 3.12. All parameters are sorted by decreasing influence (determined by their median value, red line in Figure 3.12) and the 1st and 3rd quartile box to illustrate dispersion of Sobol' index values along fuel models. To calculate those statistical metrics the parameters that are not present in a given model are discarded. That is, for example, values of Sobol' index for live moisture content of model A1 (that does not incorporate live fuels) are not considered in the box plot, neither in the computation of the statistical metrics. All 27 input parameters are ranked by their median value at Table 3.8. Note that all parameters following the live herbaceous load ($w0_{lh}$) have a zero median value but different influence over scenarios and are further sorted given the standard deviation (i.e. related to the box size). Both moisture content for live herbaceous and woody, when present in a fuel model, are important parameters to consider. The first four parameters identified previously (wind speed (U), fuel depth (δ), SAV (sv_{d1}) and the low heat content (h_{d1}) for 1h time-lag class) are also the higher ranking

parameters in overall fuel models.

Regarding the less variance influencing parameters, the slope (slp), effective mineral content (s_e), total mineral content (s_t) and moisture of 100h time-lag fuel class (m_{d3}) are the ones that show less variability among fuel models (i.e. lower standard deviation in Table 3.8 and smaller box in Figure 3.12) and less outliers. Those results, justify the common characterization of effective and total mineral content as constant parameters. The slope, however, despite low influence on RoS variance, cannot be set constant as it becomes extremely important when coupled to a spreading model. Moisture content, surface-area-to volume ratio and heat content for 10h and 100h show the lower overall mean Sobol' value. The explanation may relay on the low contribution of these particular fuel classes to the final RoS. The two other parameters that are usually kept constant are the heat content (of all species) and the ovendry particle density (ρ_p). Focusing on the heat content, disjointing it into heat content for each particle class ($h_{d1}, h_{d2}, h_{d3}, h_{lh}, h_{lw}$) helped identify that 1h time-lag and live herbaceous are the most influencing ones whereas the rest (particularly 10h and 100h classes) do not show significant influence on overall variance and their value might be approximated.
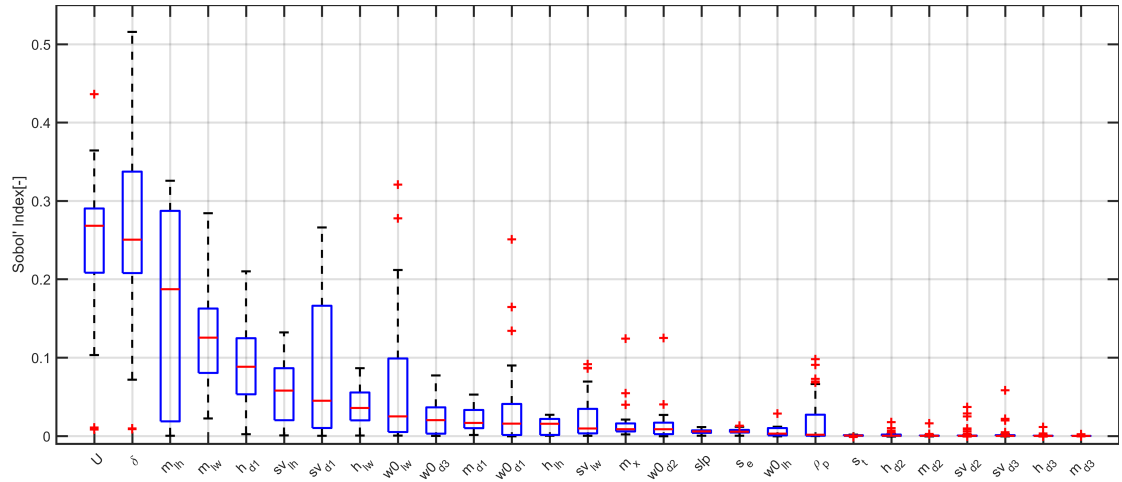
FIGURE 3.12: Sobol' index box plot for all standard fuel models. Parameters are sorted in decreasing influence (median value, red line). From $w0_{lh}$ all median values are zero.

TABLE 3.8: Ranked parameters over all standard fuel models according to their Sobol' index median value on all standard fuel models

Rank	Parameter	Median	Std	Rank	Parameter	Median	Std
1	U (u)	0,269	0,084	15	Mx dead (m_x	0,007	0,005
2	Fuel Bed Depth (δ)	0,248	0,110	16	Load 10h ($w0_{d2}$)	0,008	0,003
3	Moist Live Herb (m_{lh})	0,187	0,048	17	slope (slp)	0,005	0,070
4	Moist Live Woody (m_{lw})	0,125	0,083	18	Eff. mineral content (s_e)	0,005	0,008
5	Heat 1h (h_{dl})	0,085	0,013	19	Load Live Herb ($w0_{lh}$)	0,002	0,009
6	SAV Live Herb (sv_{lh})	0,057	0,049	20	Density (ρ_p)	0,0	0,036
7	SAV 1h (sv_{d1})	0,031	0,003	21	Total mineral content (s_t)	0,0	0,024
8	Heat Live Woody (h_{lw})	0,035	0,019	22	Heat 10h (h_{d2})	0,0	0,002
9	Load Live Woody ($w0_{lw}$)	0,024	0,003	23	Moist 10h (m_{d2})	0,0	0,009
10	Load 100h ($w0_{d3}$)	0,017	0,019	24	SA.V 10h (sv_{d2})	0,0	0,025
11	Moist 1h (m_{d1})	0,015	0,001	25	SA.V 100h (sv_{d3})	0,0	0,000
12	Load 1h ($w0_{d1}$)	0,013	0,020	26	Heat 100h (h_{d3})	0,0	0,113
13	Heat Live Herb (h_{lh})	0,015	0,026	27	Moist 100h (m_{d3})	0,0	0,077
14	SAV Live Woody (sv_{lw})	0,009	0,002				

3.6 Sensitivity analysis conclusions

After performing the fuel class influence exploration (section 3.4) and Sobol' index computation for all standard fuel models (section 3.5) we can draw the following conclusions:

- Among the 5 fuel classes, the 1h time-lag particle fuel class is the one that introduces the highest dependency on the RoS output. According to Sobol' analysis the SAV, fuel load and moisture content of this particle class are the more influencing parameters.

- The sensitivity of the fuel model parameters is not significantly altered by different moisture scenarios.

- The fuel particle class that induces the highest change on RoS, when added to the fuel model, is the 10h time-lag.

- Adding 100h to a 1h+10h fuel model is the less impacting change.

- Both, woody and herbaceous live class show a diversity of effect as their addition scatters the RoS values among the fuel models.

- For all fuels (except those that implicitly do not have a specific sub-class contribution) the RoS decreases as subsequent sub-classes are added. This could be explained as less combustible material is added to the fuel acting as a heat sink.

- Results of the global sensitivity analysis clearly showed the negligible effect of the variability of three of the model variables; the low heat content, the particle density and the mineral content.

- Sobol' index exploration over the complete set of standard fuel models reveals that no specific particular class can be discarded (i.e. kept constant) for the overall standard fuels models without affecting its RoS output. Extrapolating this result to a data assimilation framework, this implies that it is not possible to directly drop off some of the standard fuel models parameters without dramatically changing the subsequence rate o spread. Hence, and alternative way is required to simplify the standard fuel model parameters set.

3.7 A reduced model build up

The conclusions drawn in the previous section led to discard the idea of directly setting some of the fuel parameters constant (other than heat content, effective and total mineral content and ovendry particle density) to reduce the model dimensionality. The alternative way to lower the number of fuel model describing variables is then to apply a clustering strategy. The idea consists in gathering the fuel particle parameters in a characteristic representative value. The clustering strategy (i.e. functional relation) between fuel particle properties (load, SAV and moisture content) is an arbitrary step that attempts to capture the fuel particle effect into a single variable. The clustering equations should also have the property to be used to convert fuel models from the full standard description to the reduced one. Mathematically this means they must be invertible and uniquely determined.

Following the idea of Scott and Burgan (2005b) we firstly explore the viability of using their so-called characteristic values.

3.7.1 Characteristic variables approach

Scott and Burgan (2005b) suggested that characteristic values for fuel load (W), surface-area-to-volume ratio (SAV) and moisture content and moisture of extinction (m and m_x) could be generated to describe each fuel model.

For the fuel load, since the fine fuel is the main driver in a wildfire, the *Fine Fuel Load* (W) characteristic variable could be created as 1h time-lag load plus the live herbaceous and woody:

$$W = w0_{d1} + w0_{lh} + w0_{lw} \qquad (3.5)$$

The characteristic surface-area-to-volume ratio (SAV) will be the area-weighted sum of all class SAV:

$$SAV = \sum_{i}^{5} \frac{A_i}{A_t} \cdot sv_i \qquad (3.6)$$

where A_i and A_t are the class and the total area respectively defined as:

$$A_i = \frac{sv_i \cdot w0_i}{\rho_{p_i}} \tag{3.7}$$

$$A_t = \sum_i^5 A_i \tag{3.8}$$

If we take the ovendry particle density (ρ_{p_i}) constant for all classes, equation 3.6 can be written as:

$$SAV = \frac{1}{\rho_p A_t}\big(sv_{d1}w0_{d1} + sv_{d2}w0_{d2} + $$
$$sv_{d3}w0_{d3} + sv_{lh}w0_{lh} + sv_{lw}w0_{lw}\big) \tag{3.9}$$

The last characteristic variables are the moisture content (m_f) and moisture content of extinction (M_x). The first is taken as the 1h time-lag class (that is why lower case letter is chosen). This choice is motivated by the difficulty to measure moisture contents for all classes in a operational situation. Moreover, moisture content is a dynamic variable that changes daily (due to insulation). The complexity in resolving this dynamic effect for al 5 classes makes it unfeasible to use any weighted sum as in previous parameters. On the other hand, the moisture of extinction (M_x) is taken to be a representative value for all classes considered.

The complete set of variables, their range and the constant values taken to set up our reduced fuel modelling approach are presented in Table 3.9. The ranges and constant values are derived form the minimum and maximum values appearing in the standard fuel models (Scott and Burgan, 2005b).

Note that the total number of inputs is reduced to 7 but an inequality bound must be applied between moisture content and moisture of extinction to prevent non-physical parameters relationships. For definition moisture content must comply with:

$$Mx > Mf \rightarrow Mx - Mf > 0 \tag{3.10}$$

Otherwise, the fuel is incombustible and thus RoS cannot be computed.

To check if this characteristic variable approach accomplish the requirements of a reduced model, we run it for all 53 standard fuels (see Table 3.5) and the 16 moisture scenarios (see Table 3.7) and compare the obtained RoS values to those generated with the original 27 input parameters model. The process is repeated for different input wind speed and slopes. The results are presented for a 14 deg slope and 10 km h^{-1} wind speed as a box plot of the difference between models (Figure 3.13a) and the absolute percentage error (APE) (see Figure 3.13b). Each box is thus composed by the 53 fuel models that are compared over the 16 different moisture scenarios.

TABLE 3.9: Summary of the proposed fuel model variables, ranges and constant values.

Parameter		Range (value)	Units
Characteristic variables			
Fine Fuel Load	W	0.067-2.925	[kg m^{-2}]
Surface-area-to-volume ratio	SAV	3753-7270	[m^{-1}]
Moisture content	M_f	5-40	[%]
Moisture of extinction	M_x	12-40	[%]
Original Variables			
Fuel Bed Depth	D	0.06-1.83	[m]
Mid-flame Wind Speed	Uf	0.06-1.83	[m s^{-1}]
Mid-flame Wind Dir	θ	π-3π	[rad]
Constant parameters			
Low heat content (all fuel particles)	h	18609	[kJ kg^{-1}]
Ovendry particle density	ρ_p	512.59	[kg m^{-3}]
Effective mineral content	s_e	5.55	[%]
Total mineral content	s_t	1.0	[%]
Terrain input parameters			
Slope	slp	-	[rad]
Aspect	α	-	[rad]

The first plot allows to identify that the reduced model tends to over-predict RoS (i.e. predicts a higher value). The median of this over-prediction is of about 20 m min^{-1} for all scenarios. The reason for this behaviour is the omission of 10h and 100h time-lag classes that slower RoS down. In terms of absolute percentage error, this over-prediction reaches unacceptable limits as the median value is around of 200% of committed error. Moisture configurations containing L1 live moisture scenario (i.e. a live moisture content for herbaceous and woody class of 120% and 150% respectively) show a particular dramatic behaviour. Their median APE reach 500% whereas third quartile grows up to 3000%. This is a clear sign that the live moisture dynamics (see sec. 3.4) cannot be reduced to one variable. In the light of this results, an alternative approach must be taken to built the reduced model if it has to resemblance the complete standard fuel model.

(a) Error

(b) APE
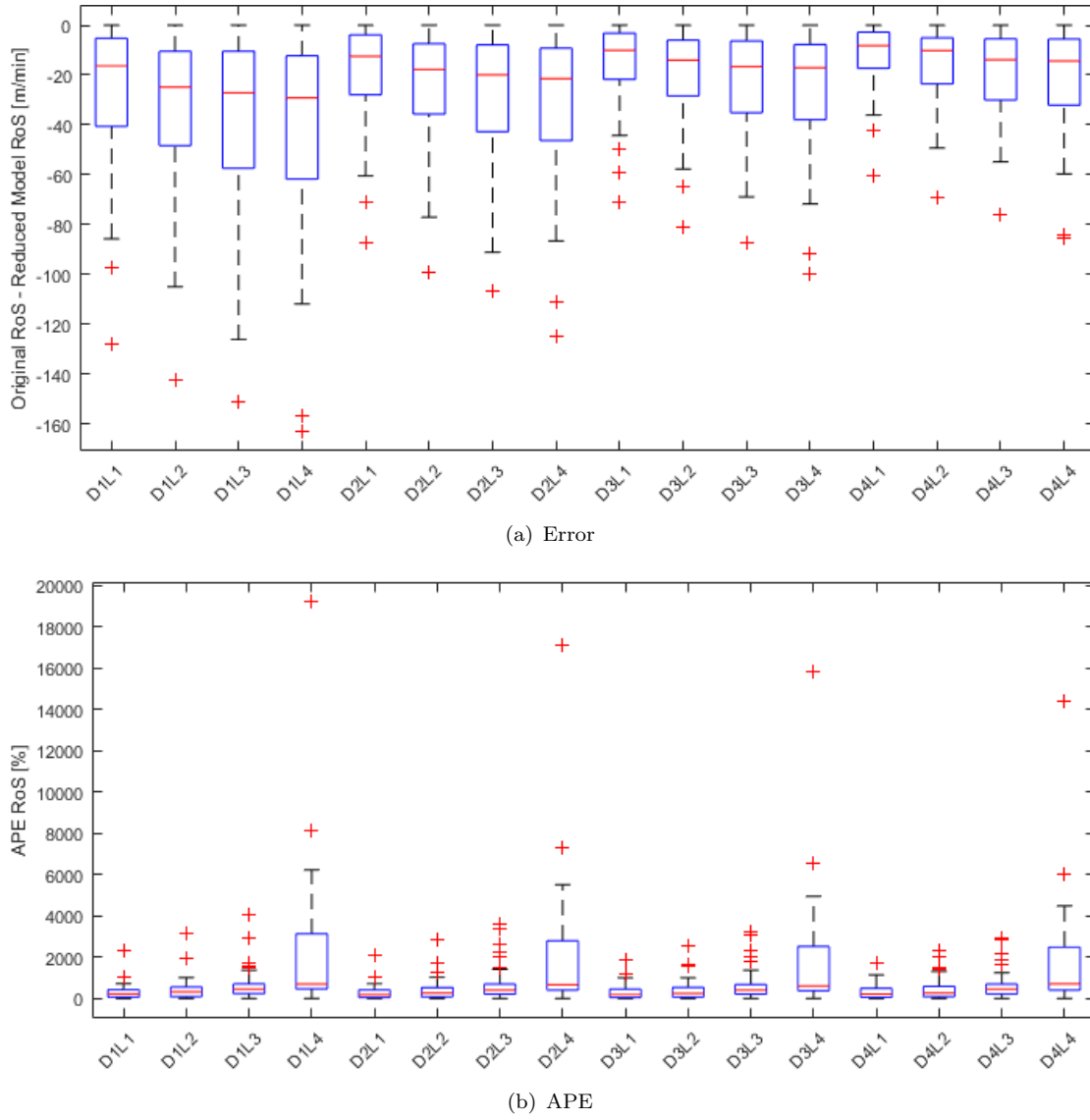
FIGURE 3.13: Box plots of error (i.e. difference) (plot (a)) and absolute percentage error (APE) (plot (b)) for all 53 burnable standard fuel models and 16 different moisture scenarios. Slope is set to 14 deg and wind speed to 10 km h^{-1}

3.7.2   Numerical correlation approach

The numerical correlation approach consist in finding the coefficients that allow to establish a representative value from the 15 fuel particle describing variables. That is, deriving the complete set of parameter per category given a characteristic fuel parameters (w, σ, m) and vice-versa.

$$W, S, M \longleftrightarrow \begin{cases} w_{d1}, w_{d2}, w_{d3}, w_{lh}, w_{lw} \\ s_{d1}, s_{d2}, s_{d3}, s_{lh}, s_{lw} \\ m_{d1}, m_{d2}, m_{d3}, m_{lh}, m_{lw} \end{cases} \tag{3.11}$$

These mathematical relations are important as the physical variables are the ones that can be directly measured and afterwards used to generate the reduced variables set. Moreover it allows for the model validation with other spread simulator that uses Rothermel's input complete set (which are the majority of common used simulators). To find the equivalence equations it is necessary to identify the fuel particles classes present in the given fuel (as not all fuels contain all classes). By exploring the SFM set, and discarding dynamic models for the aforementioned reasons, one can identify 6 categories (Cat.) depending on the combination of fuel particle classes. Those categories are defined in Table 3.10. Note that out of the 8 possible classes, two have zero members (i.e. no SFM falls into the given class combination) and thus, are discarded in Table 3.10 (second and sixth rows). It must also be commented that by definition, dynamic models are all the new fuel model introduced by Scott and Burgan (2005b) that include live herbaceous fuel category (i.e. $w_{lh} \neq 0$). This, however, does not apply to the 13 original Rothermel's fuel models. The 6th category (Cat. 6) contains the only fuel model that has live herbaceous load and is at the same time considered a *static* model.

TABLE 3.10: Categories of fuel models within the standard fuel model (SFM) system. X sign marks the corresponding particle fuel class is present in the category.

	w_1h	w_10h	w_100h	w_lh	w_lw	# of SFM
Cat. 1	x	-	-	-	-	2
-	x	x	-	-	-	0
Cat. 2	x	x	x	-	-	19
Cat. 3	x	x	x	-	x	9
Cat. 4	x	x	-	-	x	4
-	x	-	x	-	x	0
Cat. 5	x	-	-	-	x	1
Cat. 6	x	x	x	x	-	1

The correlation factors that links the characteristic reduced variables to the standard fuel model load, moisture content and SAV values for each particle class are numerically generated for each fuel category. Correlations are of the form:

$$w_i = \frac{w}{a_i} \tag{3.12}$$

$$s_i = \frac{S}{b_i} \tag{3.13}$$

$$m_i = \frac{m}{c_i} \tag{3.14}$$

Where i runs for all fuel particle class present in the given fuel category (see Table 3.10). The complete form for the fuel load variable, for example, will be:

$$\begin{cases} w_{d1} = \frac{w}{a_1} \\ w_{d2} = \frac{w}{a_2} \\ w_{d3} = \frac{w}{a_3} \\ w_{lh} = \frac{w}{a_4} \\ w_{lw} = \frac{w}{a_5} \end{cases} \tag{3.15}$$

By setting those equations we can directly convert from a reduced model set of variables to the Rothermel full set and vice-versa.

The strategy used to find the coefficients a_i, b_i and c_i is based on global search minimization of the absolute percentage difference (APE) between RoS generated with the reduced variables (W, S, M) and the RoS generated with the full variables. The optimizing algorithm makes use of the sequential quadratic programming (SQP) explained by Nocedal and Wright (1999) which is particularly effective for non-linear constrained optimization.

The optimization process must be repeated for each particle class and each category. The optimized coefficients values are gathered in Table 3.11. Note that they only exist for the fuel particles included in each category. Hence (and inevitably), the choice of the fuel category must be done beforehand. This assumption is deemed to be acceptable since recognizing the fuel particle classes present in a given scenario is a task that might be performed at first glance or using canopy cover maps.

TABLE 3.11: Coefficients to convert values from the reduced model (characteristic w, σ and m) and the five fuel particle dependent model. a_i, b_i and c_i coefficients are for w, σ and m respectively and depend on the selected category (Cat.).

Cat.	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5	c_1	c_2	c_3	c_4	c_5
1	1,00	-	-	-	-	1,00	-	-	-	-	1,00	-	-	-	-
2	0,14	4,15	3,41	-	-	3,12	3,24	3,55	-	-	0,89	6,89	15,00	-	-
3	0,88	0,15	0,12	-	4,49	4,03	3,24	2,40	-	0,21	0,68	15,00	15,00	-	0,49
4	15,00	2,47	-	-	0,56	0,15	2,87	-	-	-	0,16	2,92	-	-	0,23
5	1,06	-	-	-	10,32	2,12	-	-	-	0,05	0,81	-	-	-	1,06
6	0,00	9,97	7,54	1,47	-	1,04	0,04	0,53	0,30	-	1,07	6,82	15,00	2,11	-

Note that in the standard fuel model, the surface-area-to-volume ratio for 10h and 100h time-lag particle class is a constant value for all fuel models. The reason relies on the traditional use of the original Rothermel fuels model (Scott and Burgan, 2005b). In

order to represent a more general scenario and to better mimic reality, we consider them variable and thus, they are also calculated with the conversion equations.

The accuracy of the correlations equations (eq. 3.12-3.14) is explored by applying them to the 36 static standard fuel models with changing wind speed, slope and fuel particle moisture values. Given that the conversion system of equations is overdetermined when translating from the full model to the reduced one (see equation 3.15) the mean value of all present classes is taken to compute the characteristic variables. The validation results are presented in terms of mean absolute percentage error (MAPE) (Figure 3.14a) and a box plot of the absolute percentage error (APE) (Figure 3.14b). Categories 1, 5 and 6 give perfect results as both MAPE and the median value of APE (red lines in Figure 3.14b) are zero. The category 3 (that contains 9 members) reaches values of 22% of error in terms of MAPE and 11% in terms of median APE. For category 2 (the largest one, containing 19 members) the errors metrics increase to 28 and 23 respectively.

Lastly, the four fuel models contained in category 4 give an unacceptable error as MAPE grows to 58% and the box plot shows that some values reaches almost 100% of APE. These large errors are attributed to the fact that live wood load changes largely among the 4 members of this category. The conversion correlations are hence not recommended for category 4 fuels (1h, 10h and live herbaceous particle class only). For the rest of categories the error might be high but acceptable. Note, that given the global optimization scheme used to generate the correlation coefficients, those errors are the minimum possible given the current correlation formulation. An alternative formulation is difficult to be implemented as the main problem relies on the fact that equations must be linearly independent since 9 variables must be associated to 3 and vice-versa. The presented equations are then the best possible conversion between reduced and complete fuels model.

(a) MAPE   (b) APE
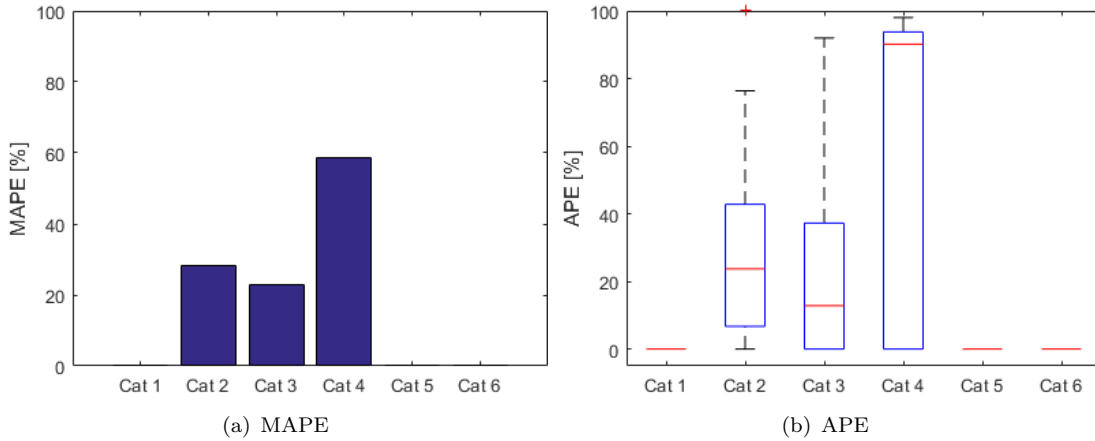
FIGURE 3.14: (a) Mean Absolute Percentage Error (MAPE) and (b) box plot of Absolute Percentage error (APE) for each fuel category when comparing RoS computed with the reduced and the complete fuel model variables set. Wind speed is set to 10 km h^{-1}, slope is set to 14 and the moisture scenario corresponds to D1L1.

The new variables range of validity is presented in Table 3.12. It is important to remark that, other than category 1 (when the reduced model directly mimics the extended one), the reduced values do not try to be fuel physical describing values. As highlighted by Ascoli et al. (2015) the purpose of calibrating a fuel model is to provide the best combination of model parameters that maximises the fit between observed and predicted fire behaviour, and not to exactly reproduce fuel characteristics observed in nature (Burgan, 1976; Cruz and Fernandes, 2008). With the present reduced model, we also assume that fuel model parameters do not represent mean or median values measured in the field but the values that physically and observably describe the fire spread. In fact, fire is not driven by average fuel conditions because it follows the path of least resistance; that is, the combustion wave spreads on the finest and driest of the fine fuels present in the scenario (Cruz and Fernandes, 2008).

TABLE 3.12: Ranges of validity for the numerically reduced variables

Parameter		Range (value)	Units
Reduced variables			
Fine Fuel Load	W	0.16-4.72	[kg m^{-2}]
Surface-area-to-volume ratio	S	596-67399	[m^{-1}]
Moisture content	M	5-40	[%]

Those reduced variables, together with the 4 additional variables gathered in Table 3.9 (i.e, low heat content, ovendry particle density, effective mineral content and total mineral content) will be used to update the data assimilation and forecasting algorithm

enlarging the parameters space from 4 to 7, gaining on the physical description and thus better capturing the fire dynamics.

3.8   Conclusions

This chapter was aimed to conduct and in-depth exploration of the fuel models used to run fire spreading simulations. After a brief revision of the most common and used strategies to characterize the fuels and describe them in a useful manners, the chapter focuses on the Standard Fuel Models (Scott and Burgan, 2005b) based on the original Albini and Rothermel characterization.

Initially, we performed a sensitivity analysis using the 53 fuels to determine the influence of the fuel particle time-lag subclasses (i.e. 1h, 10h, 100h, lh, lw) with 16 different moisture scenarios. To achieve this, a set of virtual classes are created. The results of the analysis shows that fuel particle classes have a major influence in the computation of the rate of spread. Moreover, the moisture of each subclass is a fundamental value that drives the overall RoS. This is extremely important for all fuels containing live herbaceous (LH) and live woody (LW) classes. To systematically explore the influence of each fuel model variable we carried out an analysis of variance (ANOVA) for al 53 SFM. While an ANOVA had been performed before by Liu et al. (2015b), only two fuel model where explored, and no the complete set. As an influence indicator, the Sobol' index is used. Running this analysis using Monte Carlo strategy, we obtained the complete list of variables sorted by influence for all 53 SFM. Although this helps to extract valuable insights on the variables interaction for the 53 SFM fuels and under 16 different moisture scenarios, the possible reduction of the dimension of the problem (i.e. reducing the number of variables necessary to describe a fuel model), is too limited for the need of the inverse modelling approach.

Finally, in order to generate a reduced model framework that limits the required variables necessary to describe a fuel model while keeping the RoS values, a correlation investigation is performed. By grouping SFM models by categories, composed of same fuel particle classes, we derived a correlation that allows to compute a characteristic SAV (S), fuel load (w) and moisture content (M) for each category. The validation exploration with different moisture scenarios shows good acceptable accuracy for all categories except Cat. 4, most probably due to the scatter generated by the presence of live woody particle class. For the rest of the categories, the strategy successfully sets a framework to translate from a reduced fuel model description to the SFM complete one. This last findings will then be used in the following chapters to enhance the inverse modelling approach developed in this thesis.

Chapter 4

Speeding up topographic wind simulations

Wind fields at flame height are essential to properly simulate wildfire spread in realistic scenarios as homogeneous wind conditions do not allow for accurate representation of fire spread in other than flat terrains. Moreover, topographic wind modelling can capture orography interaction and help to simulate complex fire dynamics. In the data-driven approach (when simulations are corrected by real observations) resolving the wind diagnostic model at every iteration (i.e. when wind values are updated) is computationally unaffordable as it compromises the lead time. Assuming we are to run multiple WindNinja computations in a constant scenario changing only some parameters such as the principal wind values, we made the hypothesis that general high-resolution wind fields can be approximated from a finite set of reference fields previously obtained. This hypothesis was studied and interpolation relations were found. The proposed methodology was compared to the standard WindNinja solver in 8 different topographic scenarios for multiple resolutions and base wind direction sets. Results show this approach leads to a minimal loss of accuracy while dramatically decreasing computation times. Therefore, this methodology may have an important effect in data assimilation frameworks and probabilistic risk assessment where high-resolution wind fields must be computed for multiple weather scenarios. In this chapter we develop an interpolation framework to rapidly correct wind maps generated with the WindNinja wind topographic model while keeping the error in low bounds. The work presented in this chapter was published in the *International Journal of Wildland Fires* (Rios et al., 2018).

4.1   Introduction

Wind speed and wind direction are two of the most influential parameters for wildfire spread simulations. If wind is present, it is the principal driver for the fire front shape and direction of propagation (Rothermel, 1972b; Albini, 1982). However, the wind speed and direction are not spatially homogeneous along the fire-front and they are highly influenced by topography (Forthofer et al., 2014). This fact increases the difficulty of predicting the propagation directions and might cause some extreme behaviour in complex terrain (Sharples et al., 2012; Viegas and Simeoni, 2010). Thus, it seems necessary to produce accurate wind fields and employ them in fire propagation models in order to deliver detailed fire front forecasts to emergency responders.

Wildfires are known to alter the local weather by inducing strong wind currents that in turn affect the flame structure and ultimately accelerate the fire spread (Clark et al., 2004; Filippi et al., 2011). Attempting to predict wildfire spread from first principles would thus require a fire-spread model to be coupled to a mesoscale numerical weather prediction model. An example of this approach is WRF-SFIRE (Weather Research and Forecasting model-Spread Fire), which is being used operationally in Israel. Although this tool has shown promising results, its complexity, required resources (several hundred cores) and initialisation issues mean that it is beyond the current computing capacity of most emergency response services.

An uncoupled strategy to generate high-resolution wind diagnostic models consists of running common numerical weather prediction (NWP) models on smaller grids than their normal application (3 km or larger horizontal grid resolution). Although this prognostic approach has given positive results with resolutions of less than 1 km (Ching et al., 2014; Seaman et al., 2012), when used in complex terrain, they encounter problems with the computational grids necessary for finite difference discretization (Wyngaard, 2004) and the turbulence closure models (Wyngaard, 2004). This, together with the costly running time, render them unappropriated for most wildfire scenarios. An alternative approach is to exploit mesoscale weather predictions or localised wind field measurements and apply a diagnostic model to downscale the wind maps to high spatial resolutions (up to 15 m) at a low height (below 2 m). The advantages of this strategy are the reduction in computational time (owing to simplified models and the avoidance of time stepping) and the accurate integration of digital elevation maps (DEM) with higher resolution than the prognostic approach.

High-resolution diagnostic wind models can be classified into three categories according to the level of physics considered and the strategy followed to resolve it. The diagnostic models of the first type are called mass-conserving models and use fundamental mass-balance equations often combined with empirical correlations for non-neutral vertical stability or diurnal heating dynamics (Forthofer et al., 2009; Butler et al., 2015).

The models within the second category are based on the numerical approximate solution of Navier–Stokes equations together with a turbulence sub-model for closure. They can also incorporate energy balance equations. These computational fluid dynamics (CFD) models can deliver a high degree of accuracy but usually require large computational resources. Examples include WindStation (Lopes, 2003) –specifically aimed at assisting wildfire simulations – and the more general OpenFOAM (Weller and Tabor, 1998). The third category falls in between the two previous categories, as momentum is solved by linearising the conservation equations. Their performance in terms of computing time is similar to mass-conserving models, but despite including more physics they tend to produce less accurate results than mass-conserving models, as demonstrated in scenarios with dispersion of hazardous materials (Homicz, 2002).

WindNinja (Forthofer, 2007; Wagenbrenner et al., 2016) is an open-source software developed and maintained by the USDA Forest Service. It takes into account the terrain (DEM) and three types of canopy cover (grass, brush and trees), and it can use either a mass-conserving or a mass-and-momentum model. The mass-conserving model estimates the result of the mass balance equation, whereas the mass-and-momentum model uses the OpenFOAM toolkit (Weller and Tabor, 1998). Although the use of the second approach gives more accurate results for strong winds on the lee side of mountains and ridges (where eddies can occur), the first option is 60 times faster to deliver the wind map (Forthofer et al., 2014). Additionally, the mass-conserving approach allows the use of point measurements and is able to handle non-neutral atmospheric stability effects. Moreover, its performance might be improved by using mesoscale forecast data for initialisation. This combination can thus account for both mesoscale data and local terrain effects (Forthofer et al., 2014).

WindNinja is one of the most extensively used diagnostic wind models used to generate high-resolution near-surface wind maps for the use in fire propagation models, which are then used by fire propagation models. WindNinja is the core near-surface wind simulator for recognised fire spread simulators such as FARSITE (Finney, 1998), FireStation (Lopes et al., 2002), FlamMap (Finney, 2006) and WildfireAnalyst (Monedero et al., 2011). Figure 4.1 shows typical output of WindNinja mapped onto a terrain rendered in Google Earth. Despite the software's reasonably high speed, computation times required even by the fast mass-balance solver are of the order of 5 min for an 800 × 800-cell map using multi-core workstations (Sanjuan et al., 2016a). Computing times of such magnitude are unaffordable if the wind field is to be computed recursively. This is a critical feature of any data assimilation algorithm incorporated into data-driven simulators. Those simulators (see examples in Altintas et al. (2015); Rios et al. (2016); Zhang et al. (2017)) try to match observed fire locations with simulation results. To achieve this, they need to run the fire spread algorithm (which includes a surface wind

module) multiple times. Thus, resolving the wind field for different values of mesoscale wind is a major bottleneck when aiming to achieve positive lead times.

In recent years several attempts have been made to develop data-driven wildfire propagation systems. They have been proven to be a promising strategy to overcome the lack of information (such as fire front location, accurate vegetation maps, weather conditions, etc.) inherent in emergency operations, and to reduce the high uncertainty of initial model parameters (Mandel et al., 2009; Rios et al., 2014b; Rochoux et al., 2014c; Altintas et al., 2015; Rios et al., 2016). Although different strategies are being investigated, they all require multiple runs of the core spread model, which consists of a wildfire spread algorithm and normally includes the computation of near-surface wind diagnostic models. Therefore, computation time has become a principal constraint if the whole system is to deliver useful information to end-users in due time. Recent efforts to parallelize WindNinja exploit domain decomposition methods Sanjuan et al. (2016c), computational parallelization based on GPU (graphics processing unit) (Sanjuan et al., 2016a) and hybrid integration using message passing interface (MPI) and open multi-processing (OpenMP) (Sanjuan et al., 2016b). Despite the remarkable speed-up achieved with those strategies, the computing time required for a single run of a 1500×1500-cell map exceeds 90 s on a 64-node processor using the most efficient approach (hybrid MPI-OpenMP integration). WindNinja is thus still not sufficiently fast to be used in an optimisation framework where hundreds of runs are required.

In order to minimise the computational cost, we propose to downscale only a few combinations of wind speed and direction (gridded wind output from WindNinja) and then interpolate the wind field at 1 m above ground level (the lower limit of the downscaling height suitable for this methodology). The interpolation is performed using a set of base fields generated with WindNinja for different wind speeds and directions. This scheme allows to produce wind maps that take into account terrain characteristics while avoiding to run WindNinja iteratively. This approach has an intrinsic limitation when resolving thermally driven winds that might be dominant in complex terrain during quiescent synoptic conditions. However, the approach is valid for any non-quiescent situations where mechanical effects of the terrain dominate over thermal effects. Those are the cases of principal interest for fire applications, as they also yield the highest fire front rate of spread.

In this article, we perform an in-depth analysis of this strategy using the WindNinja mass-conserving solver without thermal parametrization. The effects of grid resolution, canopy cover and the number of available base maps are studied in order to find the best interpolation framework. The proposed methodology is assessed in different scenarios (illustrating diverse topography configurations) under different conditions by comparing the resulting wind fields with the results obtained using WindNinja directly. Finally, a

comparative study is presented to evaluate the effect of the proposed wind field inter-
polation methodology in a fire spread simulation.
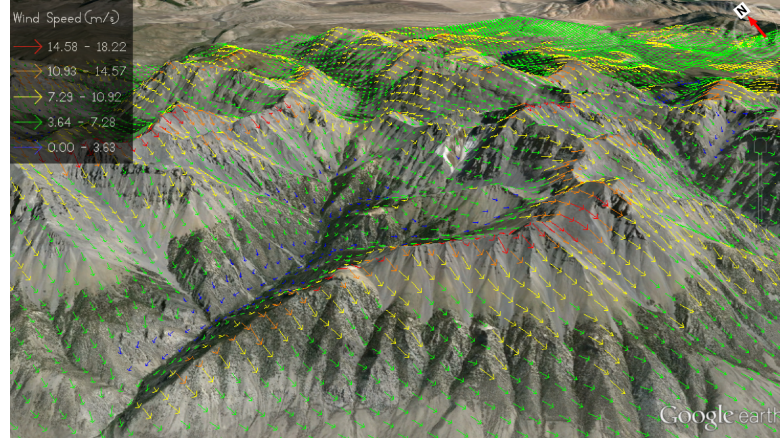
FIGURE 4.1: WindNinja output at 1 m height and grass canopy for a given hilly
scenario. Wind vector map representation (direction and magnitude) in Google Earth
for 10 m s^{-1} and 0 deg (north) principal wind speed and direction.

4.2 Interpolation Framework

A possible approach to speed up multiple WindNinja runs is to exploit the fact that,
within a data-driven application strategy, diagnostic wind models need to be run mul-
tiple times for the same scenario (e.g. same DEM, canopy cove, height) changing only
the boundary conditions such as the initialisation wind speed and direction (hereafter
referred as the principal wind speed and direction), canopy cover, etc. An interpolation
framework can be set as follows:

A WindNinja generated wind map ($\boldsymbol{W}$) consists of two matrices that contain wind
speeds ($\boldsymbol{U}$) and wind directions ($\boldsymbol{D}$) (in angles) at a requested height (h) . A wind map
W is produced by WindNinja from a given representative wind speed, called principal
wind speed (U) and principal wind direction (D) at a certain height (h), for a given
digital elevation model (T), and a fuel type (f) that can be either grass, bush or trees.

A set of base map ($\{Wb\}$), are those wind maps that are used as basis for the
interpolation framework. Thus, they are composed by base wind speed (U_b) and base
wind directions (D_b). As before, those maps are generated with principal wind speeds
(U_b) and directions (U_b). The subscript 'b' is to recall that they are principal wind
components that define the correspondent base map.

For a singular principal wind speeds and an n-directions WindNinja produces a set
of wind maps $\{W1, W2, \dots, Wn\}$ that can be employed as base wind maps (namely
base maps, $U_{b,i}$, $D_{b,i}$). Then, instead of running WindNinja again for a new desired

principal wind speed and direction, the wind map (U_I, D_I) is obtained by interpolating the base maps to the desired principal wind speed (U) and direction (D) according to the following equations:

$$U_I = \mathbf{U_{b,i}}(U_{b,i}, D_{b,i}, h, T) \cdot F_{U,i}(U_{b,i}, D_{b,i}, h, T, D, U), \qquad (4.1)$$

$$D_I = \mathbf{D_{b,i}}(U_{b,i}, D_{b,i}, h, T) \cdot F_{D,i}(U_{b,i}, D_{b,i}, h, T, D, U), \qquad (4.2)$$

The wind speed proportionality factor ($F_{U,i}$) could in principle depend on the fuel type (f), the original principal wind speed (U_b). Similar dependencies might be found for the wind direction proportionality factor ($F_{D,i}$). Although the proportionality factor can also depend on the DEM, the hypothesis is that this dependency is negligible. To assess whether such an interpolation scheme exists and whether it has a coherent form for the different dependencies, 11 scenarios were defined and explored. The scenario descriptions are given in 4.1. For each scenario all wind speeds, directions and canopy covers were combined to explore the sensitivity to the parameters.

In total, 1320 WindNinja runs were launched and their output compared. The work flow is represented in Figure 4.2 together with the nomenclature used in following sections. To estimate the factors $F_{D,i}$ and $F_{U,i}$ (eq. 4.1 and 4.2), maps generated at a height of 1 m were compared with maps where both U, D and U_b, D_b were changed. Seeing as the values of F_D and F_U might differ at each pixel point in the map, mean values and standard deviations were computed.

TABLE 4.1: Simulated scenarios used for model testing. The letters g, b and t stands for the *grass*, *brush* and *trees* canopy cover types. IGCC is the *Institut Geogràfic i Cartogràfic de Catalunya*. USGS stands for *Unites States Geological Survey*. U_b and D_b values are expressed as lower bound (lb), step (s), upper bound (up) and the resulting number of elements (#) as: lb:s:up (#).

ID	DEM Res. [m]	Source	Size [km]		Size [px]		Canopy	U_b [ms^{-1}](#)	D_b [deg](#)
Montseny05m	5	IGCC	4.7	3.65	940	730	g, b, t	1:1:2 (2)	0:15:285 (8)
Montseny15m	15	IGCC	6.81	5.46	454	364	g, b, t	1:2:15 (8)	0:45:315 (8)
Alaska30m	30	USGS	28.2	21.39	940	713	g, b, t	1:2:15 (8)	0:45:315 (8)
Boulder30m	30	USGS	7.2	5.13	240	171	g	5,11 (2)	0:5:355 (32)
Canada30m	30	USGS	15.6	10.74	520	358	g, b, t	1:2:15 (8)	0:45:315 (8)
Colorado30m	30	USGS	7.2	5.13	240	171	g	1:2:15 (8)	0:15:345 (25)
Idaho30m	30	USGS	28.2	21.39	940	713	g, b, t	1:1:16 (16)	0 (1)
Missoula30m	30	USGS	6.42	3.78	214	126	g, b, t	1:2:15 (8)	0:45:315 (8)
Maipo90m	90	USGS	19.17	11.88	213	132	g	1:2:15 (8)	0:45:315 (8)
Santiago90m	90	USGS	18.63	12.06	207	134	g	1:2:15 (8)	0:45:315 (8)
Valpo90m	90	USGS	18.81	12.87	209	143	g	1:2:15 (8)	0:45:315 (8)
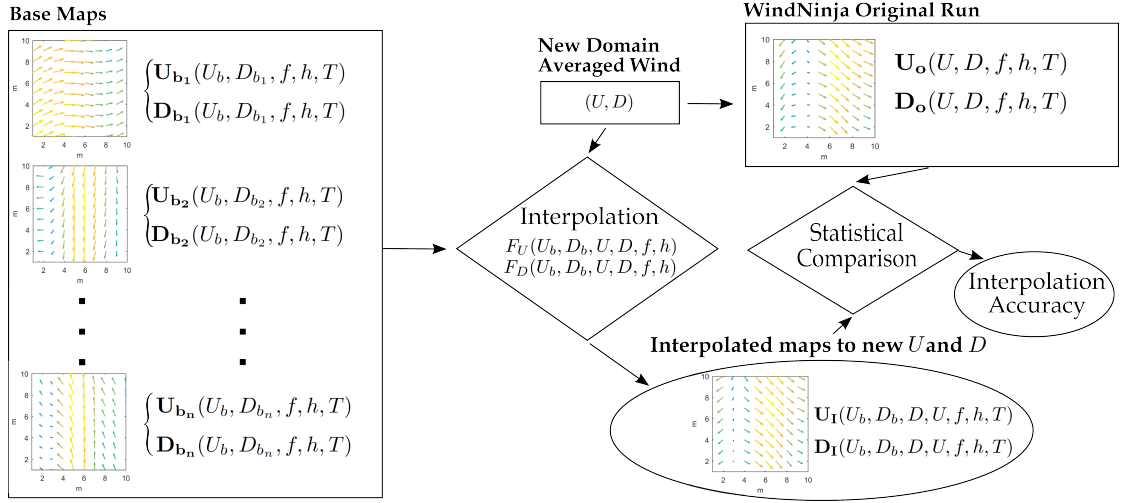
FIGURE 4.2: Work flow diagram. Squared boxes are input, diamonds are computational process and oval are outputs. Illustrative wind field maps of a 10x10 cells scenario are displayed.

4.2.1   Principal wind speed dependency

The dependence of the factors (F_D and F_U) on the speed (U) and the input base speed (U_b) is analysed by comparing speed and direction maps generated with a constant principal wind direction D_b. F_U shows a strong linear dependence on U in all simulated scenarios. FD, however, shows no clear dependence on U, as the mean value is very close to unity (no dependence at all). The dispersion of values for averaged speed lower than 3 m s–1 does increase, however. Figure 4.3 shows the mean values of F_U (solid lines) and their standard deviation (shadowed areas) for an illustrating scenario, ($Alaska30m$) as a function of 15 different wind speeds (U) and 8 base wind speeds (U_b).

The linear dependency between F_U and U (the slopes of the different curves in Figure 4.3a) changes as function of the principal wind (U_b) used to create the base map (U_b). To quantify this slope, a linear regression of the form $F_U = A_U + B$ was adjusted for every principal wind speed and every testing scenario in Table 4.1. Note that the intercept is added for consistency although a value of zero is expected. The results for grass canopy are depicted in Figure 4.4. All scenarios show exactly the same slope behaviour for different base wind speeds. This behaviour actually corresponds to the inverse of the base wind speed itself. The intercept does show some dependence on the scenario, but as it is of the order of 10^{-3}, it can be neglected (as anticipated). The linear dependence on U when the direction is fixed could be expected because mass-consistent models simply solve the Poisson's equation and Figure 4.3 can thus be regarded as a verification result.

The wind speed interpolating factor F_U can thus be generated using the base map with the closest direction available according to the following relation,

(a)   (b)
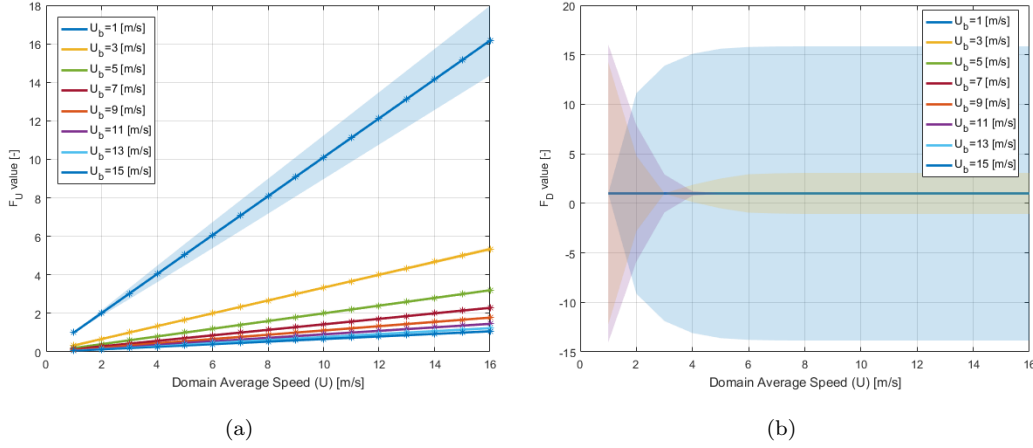
FIGURE 4.3: F_U (a) and F_D (b) interpolating wind factors for different principal speed ($\mathbf{U_b}$) on Alaska30m. Median values along the domain are depicted as solid lines and standard deviations as shaded areas. The principal wind direction (D_b) is set constant to 0 deg. For the sake of clarity only three standard deviation (i.e. shaded areas) are depicted in (b). Those are $U_b = 1$ (light blue), 3 (yellow) and 11 ms^{-1} (purple). Note F_D mean values for all U_b speeds overlap at unity.
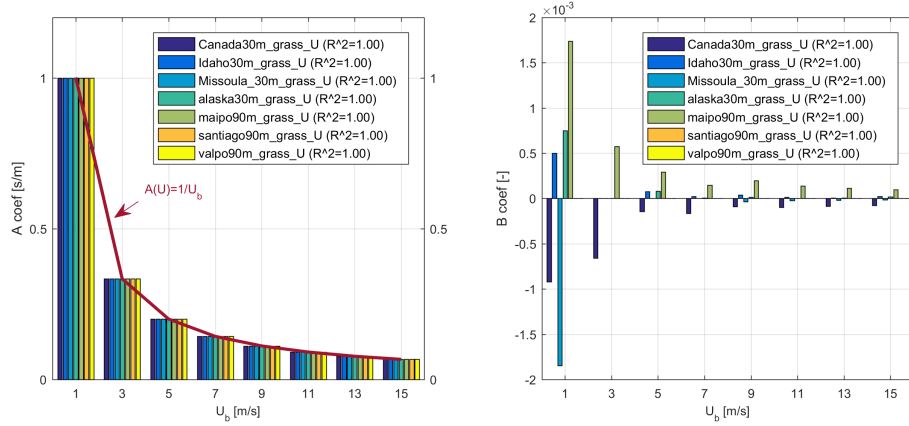
FIGURE 4.4: F_U linear coefficients dependency on wind speed and grass canopy cover. The slope (left plot) evolution is equivalent for every scenario and corresponds to the slope $1/Ub$. On the other hand, the origin coefficient (right plot) has a negligible dependency and can be assumed to be zero. The r-squared value is displayed in the legend for each scenario.

$$F_U(U_b, D_b, U, D) = \frac{1}{U_b} \cdot U \Big|_{D_b} \tag{4.3}$$

4.2.2   Principal wind direction dependency

An analogous analysis is conducted for the dependence of F_U and F_D on the wind direction. Results for the Alaska30m case with a base wind speed of 15 m s–1 are presented

in Figure 4.5 as an illustration (other scenarios, again, show similar behaviour). The speed downscaling factor F_U shows an averaged value of unity (i.e. no significant influence) when the wind direction changes. The standard deviation, however, shows large values as the angle between the input and the base wind direction increases, reaching a maximum when they are perpendicular to each other (shaded area in Figure 4.5). That is, if the base map plot is generated with a base wind direction $D_b = 0°$, the discrepancy will be at its maximum with an input wind direction (D) of 90°. The standard deviation can reach values five times larger than the absolute mean downscaling value (see Figure 4.5a). The standard deviation rapidly decreases as the interpolating direction becomes closer to reference direction maps and for symmetrical base wind directions (i.e. 0–180, 90–270). This symmetry implies that the wind speed directional dependence is cancelled out for opposite directions. This indicates that wind direction must not differ significantly from the base map in order to get an acceptable downscaling result. This effect increases as the base wind speed gets higher. The base wind direction factor F_D shows a linear dependence on wind direction for a reasonably large range of averaged wind directions (D), but the standard deviation (shaded areas) is larger when changing the averaged wind direction (D) than when changing the wind speed (previous case, Figure 4.4). The directional dependence of F_D on the updated direction (D) has a large standard deviation when comparing maps that are only some degrees off the base direction (see Figure 4.5b). The comparison shows different behaviours for directions to both sides of the base map direction. Four different base reference directions are displayed in Figure 4.5b
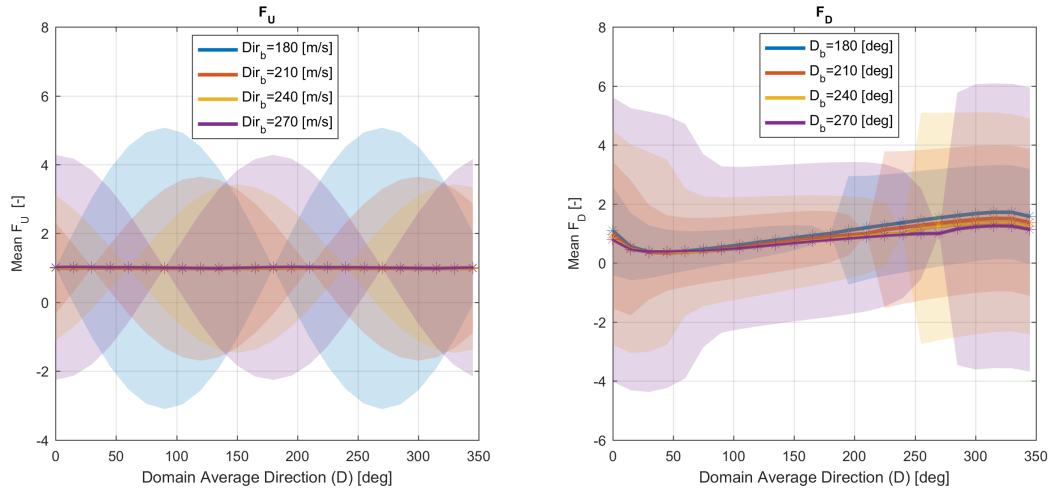
FIGURE 4.5: Alaska30m downscaling wind factors dependency on wind direction at 1 m height and with grass canopy cover. Four different principal directions (D_b) are displayed. Wind principal speed is set to (U_b) 15 ms^{-1}. Shaded are represent the associated STD.

Figure 4.5 led to the conclusion that more than one base wind direction map is required in order to keep the interpolating error in acceptable bounds because F_D does not show a clear pattern that would allow for a one-point correction. The proposed framework thus requires a set of pre-run wind directions so that the wind direction factor F_D can be defined as a direct linear interpolation between the closest two maps. Different direction sets are investigated in the following sections.

The interpolation is formulated as:

$$F_D(U_b, D_b, D) = \mathbf{D}\left(U_b, D_{b_k}\right) + \frac{\mathbf{D}\left(U_b, D_{b_k}\right) - \mathbf{D}\left(U_b, D_{b_{k+1}}\right)}{D_{b_k} - D_{b_{k+1}}} \cdot (D - D_b) \qquad (4.4)$$

Where k subscript represents the closest direction map for a given downscaling direction (D).

It is worth noting that as wind direction is a cyclic dimensionless magnitude it must be properly scaled and bounded to apply equation 4.4. This is, all angles are devalued to $[0, 2\pi]$ and all subtractions operation bounded between $[-\pi, \pi]$ to preserve linearity.

4.2.3   Fuel Canopy Dependency

As stated in eq. 4.1 and eq. 4.2, WindNinja supports three canopy cover: *grass*, *trees* and *brush*. To explore their effect on wind simulations we compare wind speed and direction maps at 1 m above ground for the three available canopies by means of histogram plots (see fig. 4.6).

The results are equivalent for all explored scenarios although all the effects are maximized when resolution cell coarsens. Grass and trees canopies show similar influence in downscaled wind speed and direction. When comparing their downscaling, the difference distribution is centred at 0 (see middle row of sub-plots in Figure 4.6). For this particular scenario (*Montseny15m*, with input wind set at 15 ms^{-1}and direction at 90 deg), the standard deviation (representing 68% of the results) is less than 10% of the input wind speed and less of 2 deg regarding the initial wind blowing direction. Brush canopy, however, shows a differentiated behaviour compared to *tree* and *grass* canopy. This cover considerably reduces the overall speed values. The averaged reduction decreases as the input speed is lowered. This caused the effect to be negligible for input values lower than 5 ms^{-1}but can be of importance for higher speed (see fig. 4.6), reaching a 50% percent of the input velocity. This difference is not homogeneous for the whole plot although the standard deviation of the distribution is kept in the same order of magnitude than when compared *grass* vs *trees* canopies. Those results imply that the *brush* canopy induces a drag effect (at 1 m heigh) non-existent for grass and trees covers. Regarding wind direction, the *brush* canopy scatters the directions rising

the standard deviation value up to 30 deg in 30 m resolution maps, whereas the mean of the difference is kept close to 0 indicating no particular drift in direction is detected.
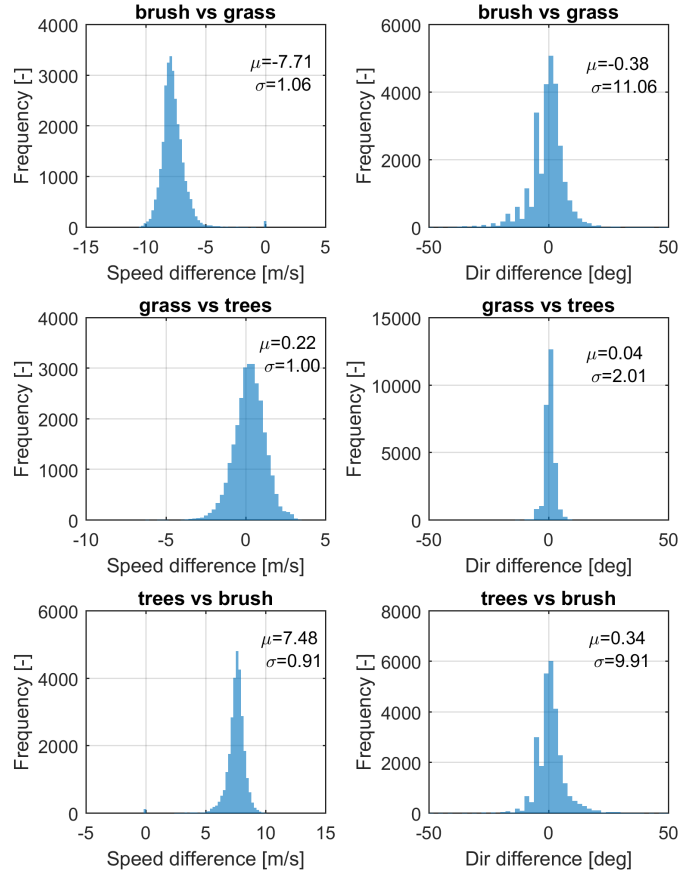
FIGURE 4.6: Histogram representation of the difference in direction and speed when changing canopy layer for a principal wind speed and direction of 15 m/s and 90 deg. *Montseny15m* scenario. Mean values (μ) and standard deviation (σ) are displayed in each sub-plot.

Although the comparison shows the canopy effect is not negligible (at least for brush case) we still need to evaluate its effect on the speed interpolating factor F_U. That is: If one base map is used (either grass, brush or trees), does the same interpolation scheme holds for all of them? To answer this, the correlation analysis illustrated in Figure 4.4 is repeated for all three canopy covers. The speed factor linear coefficients ($F_U = A \cdot U + B$) follow same correlation on the slope (A) and negligible effect on the intercept (B) as illustrated in Figure 4.7. This implies that the interpolating factor F_U (eq. 4.3) is valid for the three canopy layers given that the base maps are generated with the appropriate cover.
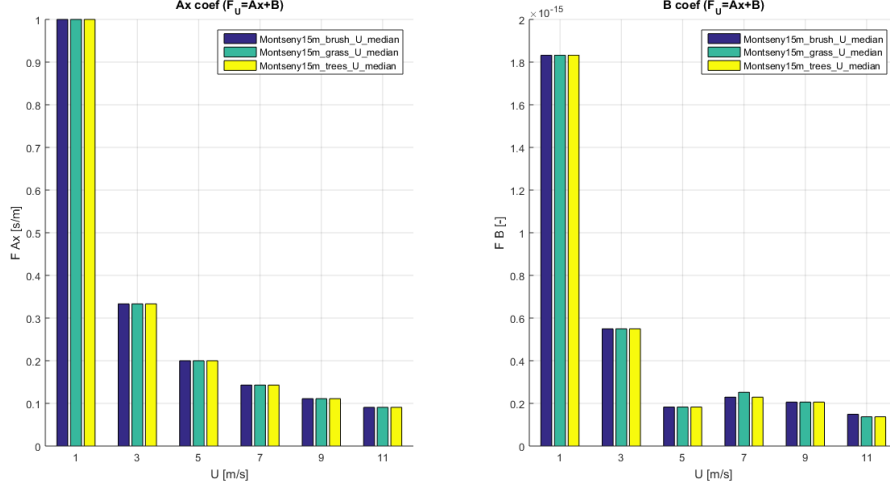
FIGURE 4.7: Fuel canopy comparison of F_U for a high resolution DEM. Not influence at all is identified.

To sum up, after studying the downscaling factor response the updated equations that simplifies the dependency stated in eq. 4.1 and 4.6) are:

$$F_U(U_b, D_b, U, D, f) \mapsto F_U(U_b, D_b, U) \tag{4.5}$$

$$F_D(U_b, D_b, U, D, f) \mapsto F_D(U_b, D_b, D) \tag{4.6}$$

4.2.4   Validation methodology

To validate the overall downscaling model we explore the absolute error committed by comparing interpolated wind maps with correctly initialised WindNinja simulations. The validation is performed step by step. First, only the wind speed is evaluated, second only the directional correction, and lastly, both direction and speed are corrected analysing the effect of mesh resolution. For assessing the discrepancy, the error (E) and the absolute error (AE) are defined together with their percentages (PE, APE) as:

$$E_i^U = \mathbf{U_I}(T, f, U, D)_i - \mathbf{U_O}(T, f, U, D)_i, \tag{4.7}$$

$$AE_i^U = |\mathbf{U_I}(T, f, U, D)_i - \mathbf{U_O}(T, f, U, D)_i|, \tag{4.8}$$

$$PE_i^U = \frac{E_i^U}{\mathbf{U_O}(T, f, U, D)_i} \cdot 100, \tag{4.9}$$

$$APE_i^U = \frac{AE_i^U}{\mathbf{U_O}(T, f, U, D)_i} \cdot 100, \tag{4.10}$$

where $||$ is the absolute value, subscript i represents a particular pixel in the wind speed matrix and $\mathbf{U_O}$ and $\mathbf{U_I}$ are the original WindNinja simulated and interpolated speeds, respectively. The same definition applies to $\mathbf{D_O}$ (direction maps) although percentage errors in this case are meaningless. When dealing with averaged value of AE_i and APE_i metrics, MAE and MAPE common abbreviations will be used. The Figure 4.8 displays the validation process for an illustrative subset of 11×11 cells. The wind result of the vector subtraction of the original and interpolated field is displayed as a histogram for both speed and directions and these data is further analysed with the error metrics mentioned above.
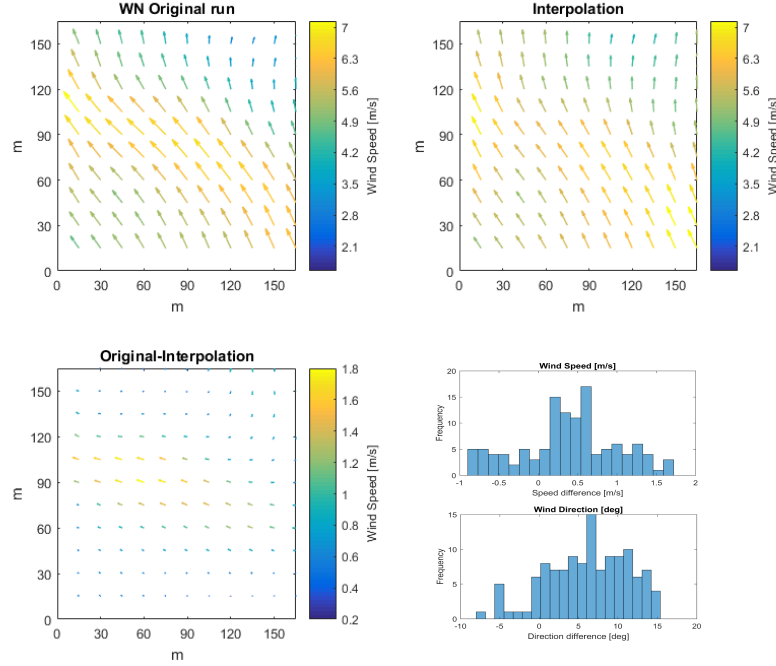
FIGURE 4.8: Composite illustrating the validation process for a Montseny15m scenario. Only a subset of 11×11 cells is shown for clarity. The WindNinja original field (U=9 ms^{-1} and D=105 deg) is compared to an interpolated field (from U_b=5 ms^{-1} and $Db = \{90, 180\}$). The subtraction result is then statistically studied, in this case in terms of histogram plots for each magnitude.

The eight different scenarios used for validating the interpolation framework are gathered in table 4.2. Each scenario is represented with different DEM resolutions to assess the influence of this parameter on the interpolation. The combination of wind speed and direction (columns 8 and 9) produces 7776 scenarios that were run with WindNinja.

TABLE 4.2: Simulated scenarios used for model testing. Abbreviations: IGCC, Institut
Geogràfic i Cartogràfic de Catalunya; USGS, Unites States Geological Survey; DEM,
digital elevation maps; Res, resolution. All plots are squared.

ID	DEM Res. [m]	Source	Type	Altitude min-max [m]	Size (side) [km]	[px]	U [ms^{-1}](#)	D [deg](#)
Atlanta		USGS	Hill	1616-1821				
Atlanta_300x300_90m	90				27	300	1:2:23(12)	0:15:359(24)
Atlanta_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Atlanta_108x108_250	250				27	108	1:2:23(12)	0:15:359(24)
Boulder		USGS	Hill-	1535-2615				
Boulder_108x108_250m	250		Flat		27	108	1:2:23(12)	0:15:359(24)
Boulder_300x300_90m	90				27	300	1:2:23(12)	0:15:359(24)
Boulder_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Idaho		USGS	Hill-	123-219				
Idaho_300x300_90m	90		Flat		27	300	1:2:23(12)	0:15:359(24)
Idaho_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Kansas		USGS	Flat	207-286				
Kansas_108x108_250	250				27	108	1:2:23(12)	0:15:359(24)
Kansas_300x300_90m	90				27	300	1:2:23(12)	0:15:359(24)
Kansas_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Montseny		IGCC	Moun-	527-1972				
Montseny_50x50_90m	90		tain		4,5	50	1:2:23(12)	0:15:359(24)
Montseny_75x75_60m	60				4,5	75	1:2:23(12)	0:15:359(24)
Montseny_150x150_30m	30				4,5	150	1:2:23(12)	0:15:359(24)
Montseny_300x300_15m	15				4,5	300	1:2:23(12)	0:15:359(24)
Montseny_900x900_5m	5				4,5	900	1:2:23(12)	0:15:359(24)
Needham		USGS	Moun	1919-3480				
Needham_108x108_250	250		tain		27	108	1:2:23(12)	0:15:359(24)
Needham_300x300_90m	90				27	300	1:2:23(12)	0:15:359(24)
Needham_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Olost		IGCC	Hill	421-641				
Olost_50x50_90m	90				4,5	50	1:2:23(12)	0:15:359(24)
Olost_75x75_60m	60				4,5	75	1:2:23(12)	0:15:359(24)
Olost_150x150_30m	30				4,5	150	1:2:23(12)	0:15:359(24)
Olost_300x300_15m	15				4,5	300	1:2:23(12)	0:15:359(24)
Olost_900x900_5m	5				4,5	900	1:2:23(12)	0:15:359(24)
Riverside		USGS	Hill	293-752				
Riverside_300x300_90m	90				27	300	1:2:23(12)	0:15:359(24)
Riverside_900x900_30m	30				27	900	1:2:23(12)	0:15:359(24)
Riverside_108x108_250	250				27	108	1:2:23(12)	0:15:359(24)

4.3 Results and discussion

4.3.1 Framework validation

4.3.1.1 Wind speed interpolation

First, the downscaling correction algorithm is applied to evaluate the wind downscaling interpolation when the direction of the base map considered matches the correcting direction (i.e only equation 4.3 is evaluated). Results show that speed interpolation leads to an error lower than 0.05 ms^{-1} when using base maps generated with velocities higher than 5 ms^{-1}. On contrary, when using low speeds maps, the error can reach similar values to the actual speed (100% relative error). The reason is that low speed maps input a high uncertainty on the downscaling factor (as highlighted in Figure 4.3). Despite this fact, the overall error committed is considered to be extremely low when wind direction is similar as the base wind map. Figure 4.9 shows the absolute error AE^U box-plot using different base speed maps. The blue box represent the Q1 and Q3 quartiles (50% of the values) whereas the whiskers length enclose the 99.3% of map pixels. All maps generated with speed higher than 10 ms^{-1} have similar error and thus any base speed beyond this threshold is considered to be an optimum reference speed.
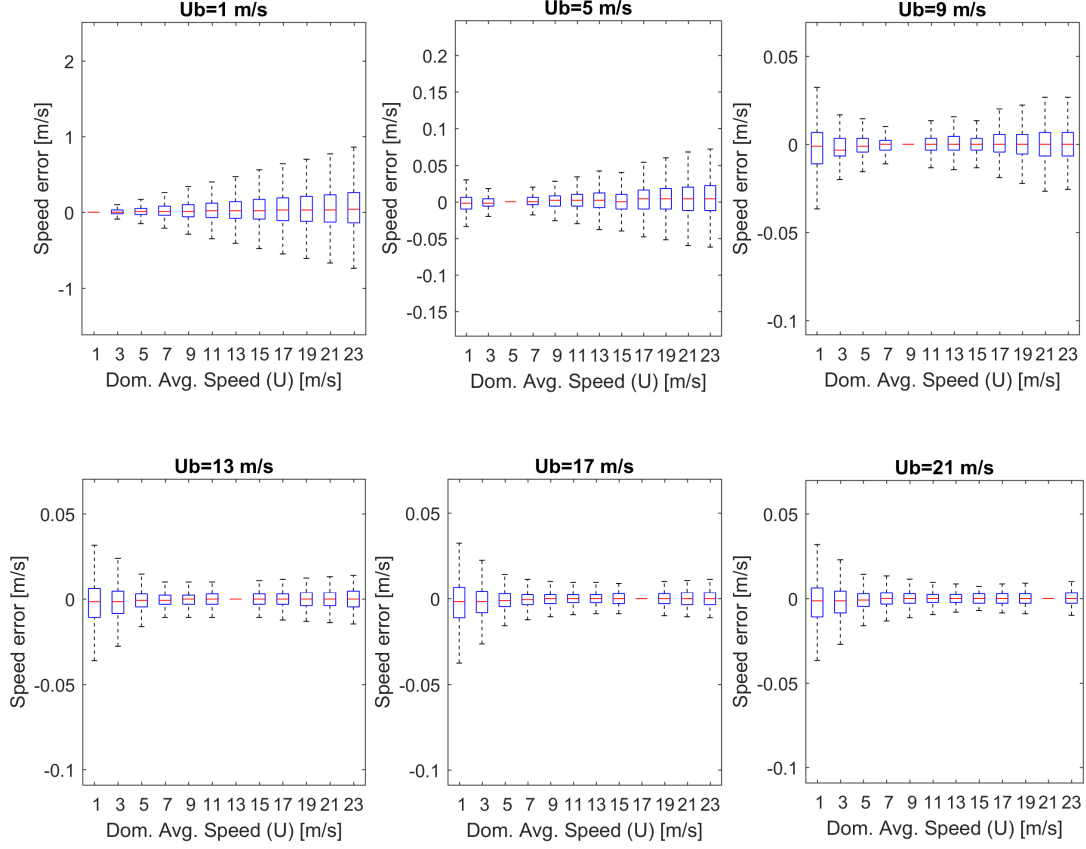
FIGURE 4.9: Error in wind speed for Atlanta_300x300_90m scenario when using different base speed maps and keeping the direction constant and equal to the available base map (180 deg at present case).

4.3.1.2   Wind direction interpolation

To assess the directional downscaling, WindNinja simulations are compared every 15 deg to interpolations generated with three different sets of base maps ($\mathbf{D_b}$). Although all scenarios are processed only the results for *Boulder_900x900_30m* are shown in figure 4.10. The sets are composed of: 12 maps spanning 30 deg (figure 4.10a), 8 maps spanning 45 deg (Figure 4.10b) and 4 maps at cardinal directions (Figure 4.10c). The wind speed is set to 11 ms^{-1}, same as the base maps. The mean of the absolute committed error is 0 for all cases (error compensation) and the whiskers length (99% of the values) grows from 2 degrees when 12 base maps are used, to over 15 degrees when only 4 directions are taken as reference. The error between base direction maps shows a similar behaviour regardless of the direction. However, a close look exhibits small differences due to topography interactions.

(a) 30 deg step

(b) 45 deg step

(c) 90 deg step
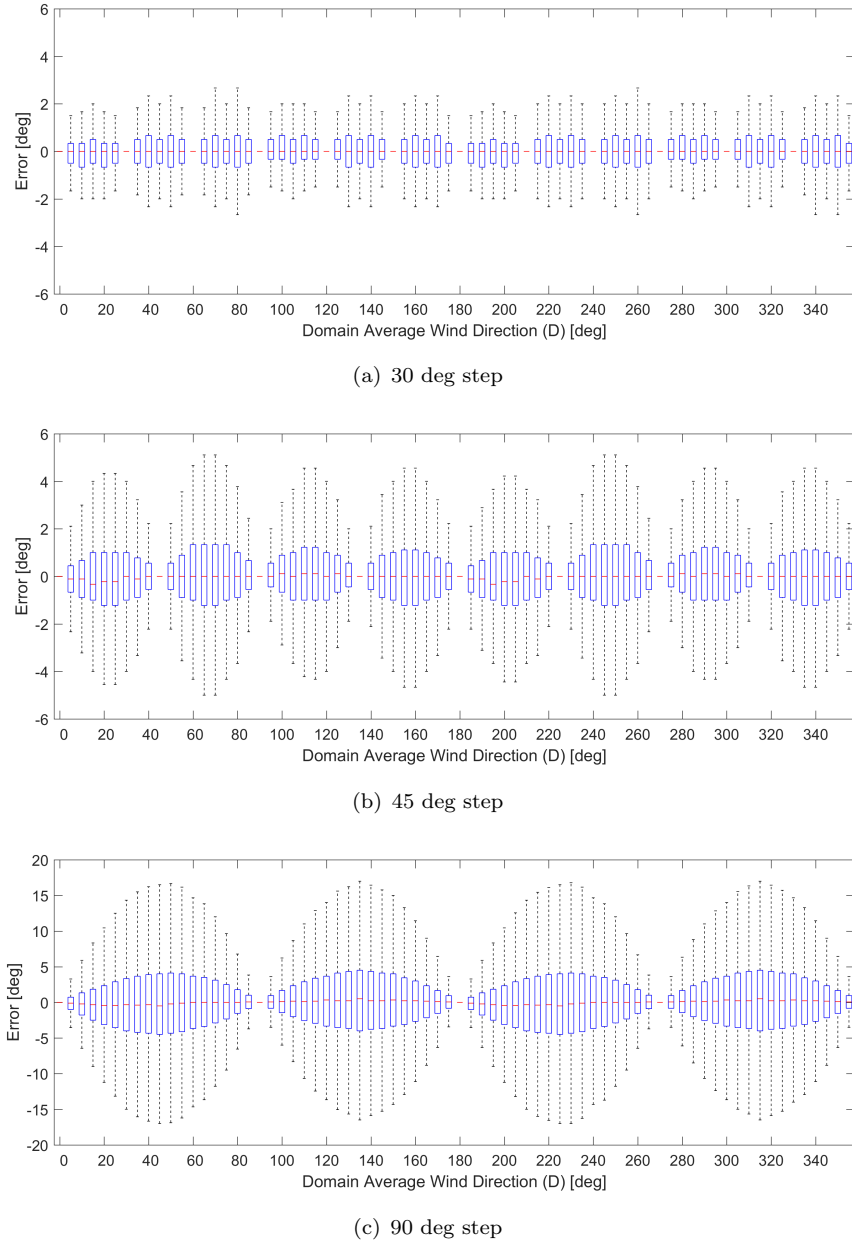
FIGURE 4.10: Box plot comparison of directional downscaling error when changing the step to generate the set of reference maps.

4.3.1.3 Wind and direction correction

To validate the overall accuracy of the interpolating scheme the domain average direction and speed are freely changed for all 27 cases gathered in Table 4.2. The reference maps are generated with $U_b = 11$ ms^{-1} principal speed and two sets of directional maps ($\{D_{b_i}\}$) are used: 4 directions (spanned 90 deg) and 8 direction (spanned 45 deg). The interpolation correctness is quantified by the mean absolute error (MAE) and standard

deviation (STD) of every interpolated map ($\mathbf{U_I}, \mathbf{D_I}$) compared to the WindNinja original maps (($\mathbf{U_o}, \mathbf{D_o}$)). The results are depicted as filled contour plots for the wind speed (U) and direction (D). That is, every pixel represent a MAE and STD value for a given combination of U and D. To condense the overall response, the scenarios are gathered by DEM resolution and their MAE is averaged and depicted in one figure. Figure 4.11 shows the mean absolute error committed when the two scenarios having 15 m resolution (res.) DEM ((b) and (d) panels) and the eight scenarios having 30 m resolution ((a) and (c) panels) are considered. In all cases the compared maps span from 0 to 359 deg every 15 deg (i.e. 24 directional maps) and from 1 to 23 ms^{-1} every 2 ms^{-1} (i.e. 12 maps). The maximum error is found in-between directional maps (either at 45 deg or at 22.5 deg from the reference map) and increases with the wind speed up to 3 ms^{-1} at 23 ms^{-1} reference speed in the four directional maps case and at 1.1 ms^{-1} at 23 ms^{-1} on the 8 maps scheme and it is kept constant along directional variation. This indicates that the angular interpolation uncertainty dominates over the wind speed interpolating error. A percentage exploration shows that the error is kept constant along same direction reaching a maximum of 15%, confirming the directional uncertainty source.

(a) U error, 15 m res., 90 deg step   (b) U error, 30 m res., 90 deg step

(c) U error, 15 m res., 45 deg step   (d) U error, 30 m res., 45 deg step
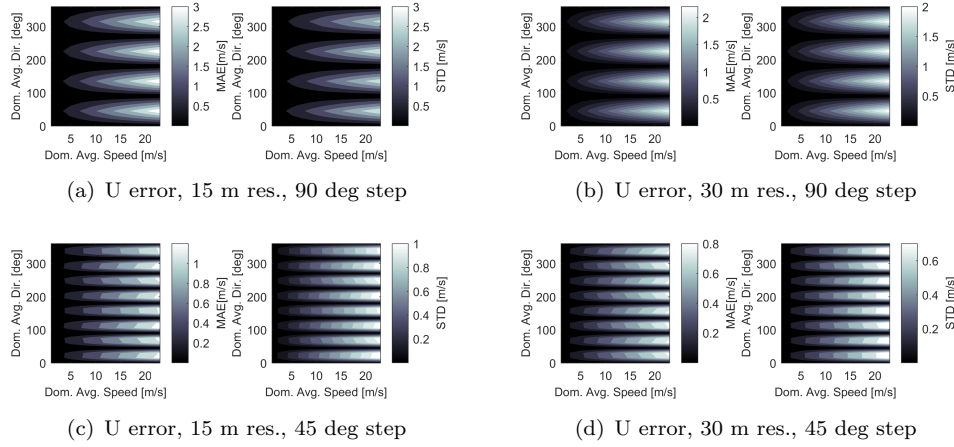
FIGURE 4.11: Wind speed validation for all 15 m and 30 m resolution scenarios. The MAE and STD are represented.

The wind direction is less correctly resolved as mean absolute errors easily grow to 10 deg with a standard deviation, indicating that in the worst case interpolating scenarios at least 68% of the corrected directions have an offset lower than 20 deg. The directional error is completely independent of both the input base principal direction and the interpolating direction. Doubling the available base maps halves the incurred error.

(a) Dir. error, 15 m res., 90 deg step   (b) Dir. error, 30 m res., 90 deg step

(c) Dir. error, 15 m res., 45 deg step   (d) Dir. error, 30 m res., 45 deg step
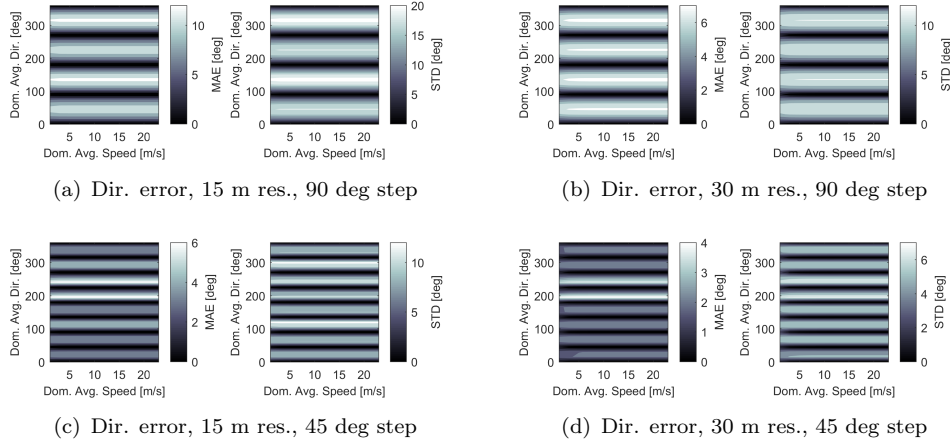
FIGURE 4.12: Direction speed validation for all 15 m and 30 m resolution scenarios. The MAE and STD are represented.

4.3.2 Fire spread comparisons

To assess the quality of the interpolating scheme we use the fire perimeter spread simulator described in Rios et al. (2016) with added topography (see section 5.1) to perform a 30 min synthetic run over the Montseny scenario (see Figure 4.13) and compare the perimeters using interpolated maps, original WindNinja maps and constant wind. This comparison is not intended to validate the propagation model, but to assess the application of the interpolation framework proposed, and to highlight the improvement when using topographic wind fields.

To represent the worst case the base maps are generated at $U_b = 5$ ms^{-1} (18 kmh^{-1}) and with two different sets ($\{D_{b_i}\}$): 4 and 8 directions (i.e. every 90 and 45 deg.). The updated principal wind is set to be at $U = 10$ ms^{-1} and blowing from 240 deg (towards 60 deg north). This conditions illustrates the highest error committed by the interpolation, as discussed in previous sections. The DEM is taken to be of 15 m resolution and 600×600 px (i.e. 9×9 km). The resolution is conserved for wind maps. Results are shown in figure 4.13. Isochrones spacing 120 s are depicted over the DEM hillshade map. Both, interpolated and original maps, allow the propagation to capture topographic effects better than the homogeneous wind case, as the wind is also topography dependent, whereas for the homogeneous wind case the overall terrain influence is insignificant. This is clearly noted on the two sides of the ridge (at the centre of the Figure 4.13). As the fire has difficulties to propagate downwards on the lee side, the fire front is broken into two sub-fronts separated by a lower rate-of-spread zone. This is particularly affected by a region with almost 0 wind (see figure 4.14).

Regarding the differences between WindNinja original maps and interpolated maps, when 4 reference maps are used (panel (a) 4.13) the simulated forefront at 30 min has

a maximum discrepancy of 250 m, as the interpolated maps do not correctly solve the lee side of the ridge and the front shows a stagnation point. This lee side effect is not captured by the simulation with a homogeneous wind field, mostly unresolved by 90 deg stepping interpolation and considerably captured by the 45 deg step case. On the flanks, the discrepancy is reduced to less than 20 m. When 8 reference wind fields maps are used, the stagnation point is properly resolved and the forefront difference are below 15 m. The southern flank matches perfectly with the original runs whereas on the northern flank the error grows to less than 10 m for the last isochrone (i.e. after 30 min of propagation). To help to discern between fire spread runs, isochrones unfolding plots (i.e. a pair of isochrone at a time) are attached in the annex (figures B.11, B.12, and B.13).
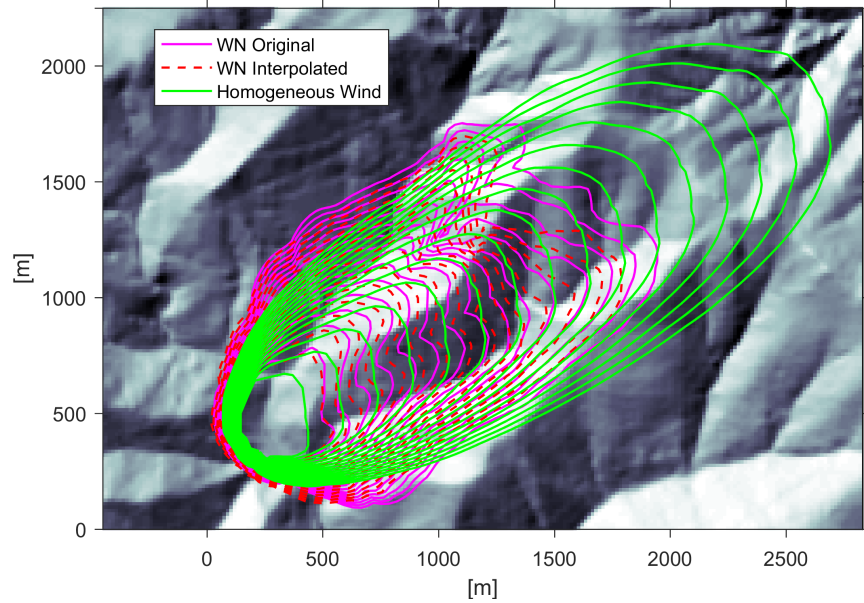
(a) 90 deg step
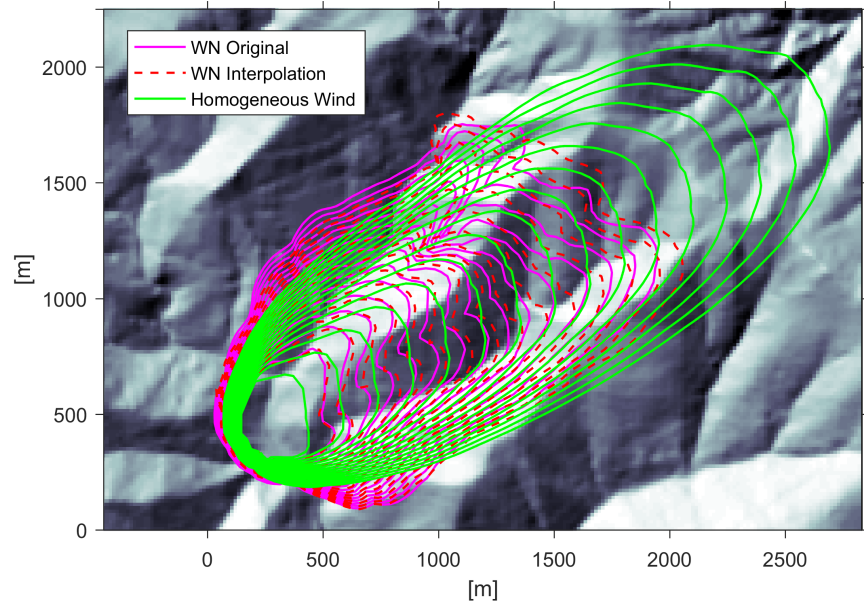
(b) 45 deg step

FIGURE 4.13: Comparison of a 30 min synthetic run with input wind speed of 5 ms^{-1} and wind blowing to 60 deg from north. Spread using wind ninja maps (magenta solid lines) is depicted every 2 min. Red dashed lines are the fronts generated applying the interpolating framework with different base maps settings. Green solid lines are the propagation with homogeneous wind field. The hillshade map is displayed in the background.
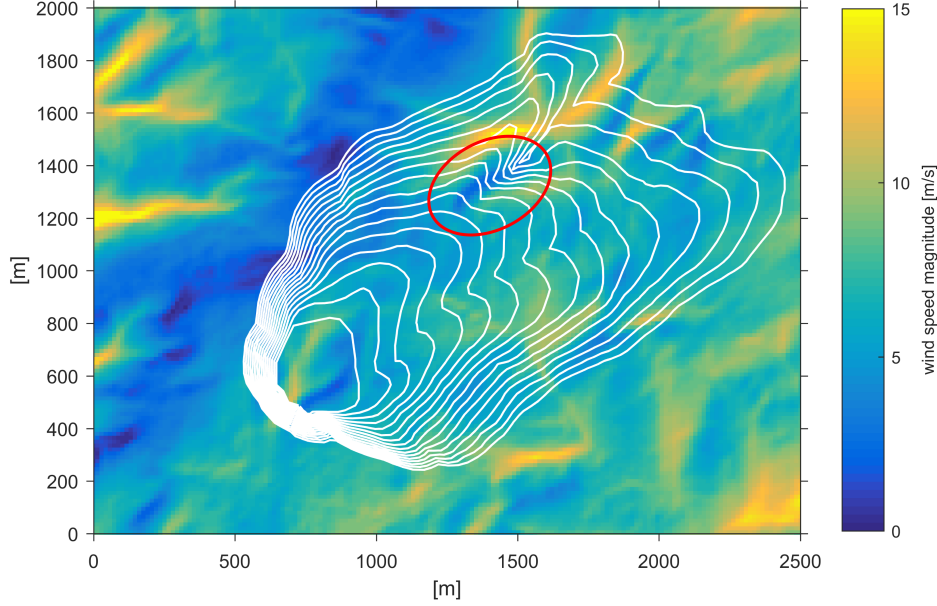
FIGURE 4.14: Fire spreading perimeters run with a WindNinja original wind map displayed over the Wind magnitude ($\|\vec{U}\|$) map. The lee side stagnation region is highlighted with a red circle.

To perform a front-to-front comparison, the shape deviation index (SDI) is calculated for every isochrone and displayed in Figure 4.15. The SDI_i is an error metric commonly used to compare front agreement (Cui and Perera (2010); Rios et al. (2016)) at a given time i, and is defined as:

$$\text{SDI}_i = 1 - \frac{A_i^o \cap A_i^c}{A_i^o} \tag{4.11}$$

where A_o^i is the area of the isochrone run using WindNinja original wind field at a time (i) and A_c^i are the isochrones run with the interpolating scheme. Note that as an absolute match is achieved the SDI value tends towards 0. Figure 4.15 clearly reflects the improvement of using the interpolation framework as the error is kept below 20% for 90 deg step and 10% for a 45 deg step.

Differences are only noticeable after 20 min, and are in the order of 20 m on the southest part of the front. Both spread patterns overlap in the majority of the fronts. Each wind field map simulated with WindNinja takes 49 sec parallelized on an eight processors (Intel Xeon 2.3Ghz E5-2697 v4) dedicated workstation. Applying the interpolating framework presented in this chapter gives almost instant solution (run time $< 10^{-3}$ s) given the pre-run scenarios.
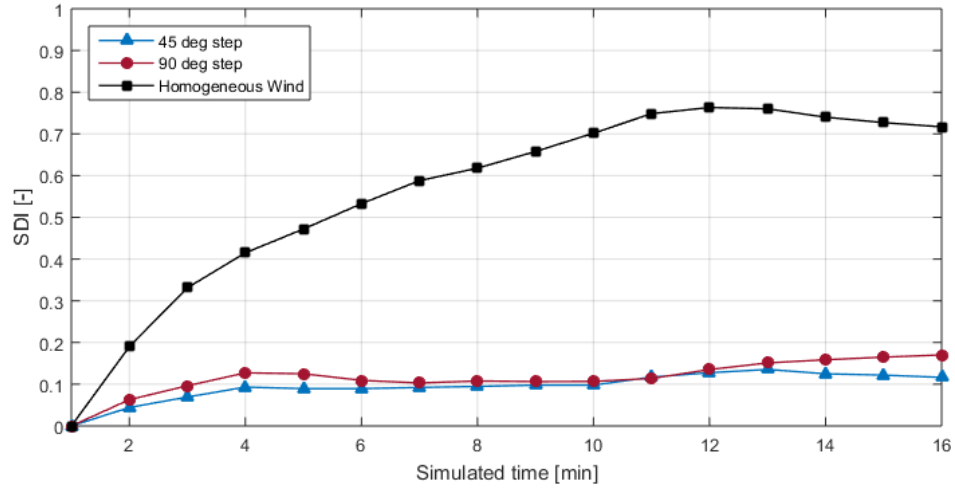
FIGURE 4.15: Similarity index comparison for fire-fronts simulated with a homogeneous wind field (black squared markers), interpolation using 90 deg step (round red markers) and 45 deg step (blue triangles).

4.4   Conclusions

A framework to downscale and interpolate high resolution topographic wind fields is presented and validated. The dependency of the downscaling factors is explored in terms of wind speed, wind direction, canopy cover and DEM resolution. The overall system is validated with WindNinja simulations performed on 8 different scenarios representing divers topographic configurations. The wind speed chosen to generate reference wind maps does not have a considerable influence as long as it is high enough to capture main topographic wind flows (i.e. $> 3ms^{-1}$). Wind direction, however, has a large influence on downscaled wind maps and therefore must be interpolated. Interpolations are compared using two reference wind direction sets (with 4 and 8 base direction maps). The wind speed mean absolute percentage error (MAPE) grows up to 7% for 20 ms^{-1} when using only 4 reference maps (see figures B.1, B.2 and B.3 in appendix B) and the directional error has a maximum mean absolute error (MAE) of 5 deg. When using reference maps every 45 deg, the wind speed MAPE decreases to 3% and the directional MAE decreases to 3 deg. A 30 minutes long fire spread simulation is launched to compare and validate the two interpolating framework configurations. The present implementation offers a significant reduction in computing time if multiple wind simulations are to be run in a particular scenario (as it would be the case in data assimilation procedures). This speed-up can be in the order of hours for optimisation routines requiring a high number of model evaluations (e.g. genetic algorithms and gradient-based methods). If eight base directions are used, the discrepancy with respect to WindNinja simulations is kept low enough so that the fire spread simulations initialised with those interpolated data are equally acceptable for operational purposes (less than 80-m discrepancy in the

fire front, and less than 10 m on the flanks after 30-min propagation). The results prove this methodology is reliable and capable of speeding up recursive near-surface wind evaluations, which would enable data assimilation and probabilistic risk assessment applications currently unfeasible.

Chapter 5

SmartQFIRE built up. An advanced data-driven algorithm

The successful proof of concept conducted in chapter 2 motivated the further development of the inverse methodology approach to create an operational tool that efficiently helps to forecast wildfire spread behaviour. To set up this tool, the fuel modelling strategy needed to be revisited and adapted to satisfy the inverse modelling constraints. This was carried out in chapter 3. Another key aspect to be solved was the wind computation which has been deeply investigated in chapter 4 to finally create a framework that allows for high precision topographic wind maps with minimal computational cost. Once all those parts are available, we should assemble them together in a seamless tool. Thus, chapter 5 is dedicated to set the basis of the overall system. Moreover, the model presented in chapter 2 is further developed with features necessary to handle real world scenarios; terrain interaction and weight factor optimization. Additionally, the cost function schema introduced in chapter 2 is revised and upgraded to overcome some identified drawbacks that were preventing the strategy to effectively work in a real-world fire scenarios.

5.1   Forward Model

In the proof of concept presented in chapter 1 the slope interaction with the spread rate and direction was announced but was not implemented. Its implementation has to be carefully analysed since, due to the front-markers tracking strategy in place, it can have major impact on the loops formation and overall running efficiency (see section 2.1.1.2). After analysing different strategies, we based our implementation on the Prometheus

Tymstra et al. (2010) software report and tailor it to meet our requirements. Verification tests are conducted in section (5.1.4) to ensure the validity of the implemented algorithm.

5.1.1   Terrain transformation

Richards equations introduced in chapter 2 (eq. 2.9 and eq. 2.10) were developed for flat terrains and, thus, must be adapted to work in non flat conditions.

Richards (2000) presented an update formulation of the Huygens' enveloping equations that account for the 3D complexity and the wind-slope vector. This generalization relies on a plane transformation also proposed by Finney (1998). The horizontal-to-surface-plane transformation (see Figure 5.1) must be applied to all angular differentials ($x_s = \frac{\partial x}{\partial s}$, $y_s = \frac{\partial y}{\partial s}$) in equations (eq. 2.9 and eq. 2.9). The transformation depends on local slope (ϕ_i) and aspect (ω_i) and has a different value for each i node.

$$\widetilde{x}_s = x_s \pm D_i \sin(w_i) \tag{5.1}$$

$$\widetilde{y}_s = y_s \pm D_i \cos(w_i) \tag{5.2}$$

where D_i is the difference between nodes in the surface plane and the horizontal plane defined in terms of slope ϕ as:

$$D_i = (x_s^2 + y_s^2)^{\frac{1}{2}} \cdot \cos(\xi_i)(1 - \cos(\phi_i)) \tag{5.3}$$

and ξ_i is the difference between the aspect direction (ω_i) and the tangential angle (α_i) of a given node as referenced on the coordinate system of the surface plane:

$$\xi_i = \tan^{-1}\left(\frac{\tan(\omega_i - \alpha_i)}{\cos(\phi_i)}\right) \tag{5.4}$$

where the tangential angle is geometrically defined as:

$$\alpha_i = \tan^{-1}\left(\frac{y_s}{x_s}\right) \tag{5.5}$$

The outputs of equations 2.9 and 2.9 are the spread rate time differentials ($x_t = \frac{\partial x}{\partial t}$, $y_t = \frac{\partial y}{\partial t}$) expressed in the surface plane. Since the final front coordinates are stored in world horizontal plane projection ([lat, lon], for example) they must be converted back from surface plane to the horizontal plane using:

$$x_t = \widetilde{x}_t \pm \widetilde{D}_i \sin(w_i) \tag{5.6}$$

$$y_t = \widetilde{y}_t \pm \widetilde{D}_i \cos(w_i) \tag{5.7}$$

where:

$$\widetilde{D}_i = \left[(x_t^2 + y_t^2)\right]^{\frac{1}{2}} \cdot \cos\left(\omega_i - \tan^{-1}(\frac{y_t}{x_t})\right)(1 - \cos(\phi_i)) \tag{5.8}$$

Note that the $\pm$ symbol in equations 5.1, 5.2, 5.6 and 5.7 is due to D_i (and $\widetilde{D}_i$) ambivalence given by the argument of $\tan^{-1}$ in 5.4. The Cartesian quadrant must be examined to choose the proper sign. Also note that the aspect (ω_i) is defined from North and following clockwise sense.

The transformation process is illustrated in Figure 5.1. Due to 3D projection representation shortcomings not all transformation variables could be clearly represented and some are therefore omitted in the figure.
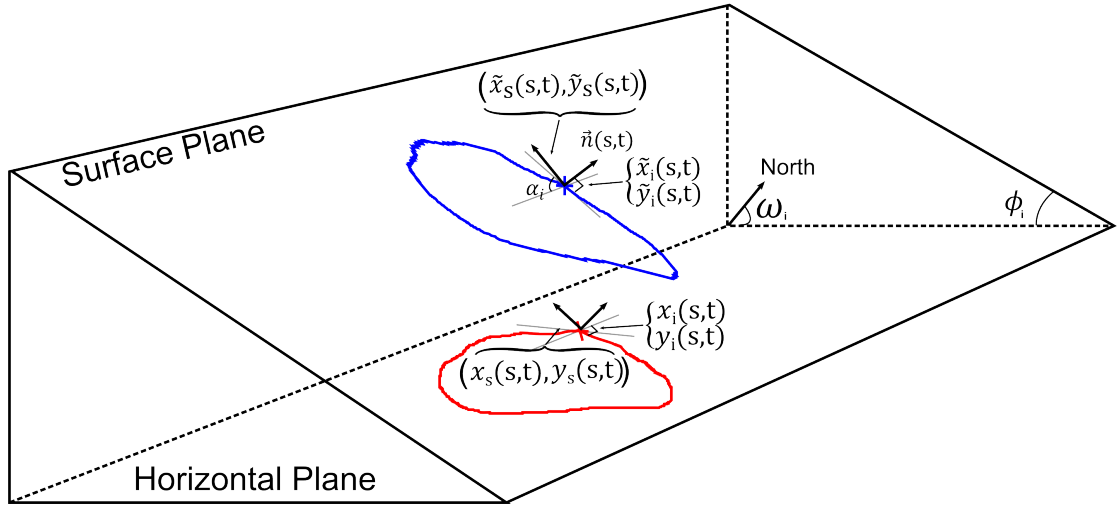
FIGURE 5.1: Surface to horizontal plane coordinates and angular transformation. Blue perimeter spreads on surface plane whereas red perimeter is the horizontal transformation.

Notice that Figure 5.1 illustrates a homogeneous slope case. That is, for a given slope (ϕ_i) and aspect (ω_i). The generalization for heterogeneous topography requires creating the surface local plane for each of the nodes. Furthermore, the implicit transformation assumes that the local surface plane of each node is kept constant throughout the integration step of differential equations 2.9 and 2.10. While this assumption might be valid for slow spreading fronts, it must be born in mind when setting the time step in rapid spreading fires cases (see section 5.5).

5.1.2 Slope interaction

In addition to the geometrical terrain transformation of fire spread equations resolved in the above section, the slope has an inherent interaction to RoS, speeding it up if the propagation is upwards or slowing it down if otherwise. In the literature, the slope is

generally treated as a correction factor (Φ_s) that is either multiplied or added to the RoS generated without slope. McArthur (1966) was the first to generate a correcting *rule of thumb* that was afterwards transformed to a mathematical correlation by Noble et al. (1980) that also included a correction down to -10 degrees.

$$\Phi_s(\phi) = 2^{\phi/10} \tag{5.9}$$

Rothermel (1972a) proposed a slope factor based on results of laboratory experiments performed with excelsior fuel beds and four different fuel bed packing ratios (β) on three slopes:

$$\phi_s(\phi, \beta) = 5.275(\beta)^{-0.3} \tan(\phi)^2 \tag{5.10}$$

Note that differently to the Noble factor, Rothermel's factor has a null value for flat terrains (0 slope). The reason is that the coefficient multiplies the non-slope (and no wind) RoS and the result is added back to the flat RoS value (see equation 2.1 for clarification).

In Canada, Wagner (1988) explored and compared five slope factors found in the literature that included the above mentioned correlations and since he found no foundations to make a chose, he took the averaged of all those functions, leading to:

$$\phi_s(\phi) = \exp\left(3.533 \cdot \tan(\phi)^{1.2}\right) \tag{5.11}$$

The comparison of the RoS calculated with all slope factors is depicted in Figure 5.2. The validity range for each case is consider to determine the ending of each curve.

Although the validity of some of the slope factors can be extended to downslope propagations (i.e. backfire) they were not developed (neither validated) to serve this purpose. Highlighting this fact, Sullivan et al. (2014) geometrically derived a new correlation called *kataburn* that, coupled with the mentioned factors, improves the performance of downwards propagation. Despite the outperformance when compared to experimental data, its application is reduced to simple constant slope hills (i.e. triangles). Moreover, when hills are not symmetrical (i.e. same upslope than downslope) the results are considerable worse. Due to those aforementioned reasons this approach was not appropriate to be implemented in a real complex topographic scenario and thus it was discarded for the present implementation. The original Rothermel factor (eq. 5.10) was implemented instead in the current version.

5.1.3   Wind-slope interaction

The direction of maximum spread (θ) appearing in equations 2.9 and 2.10 was previously set equal to the wind direction as no slope was present ($\theta = \theta_s$). However, the wind-slope
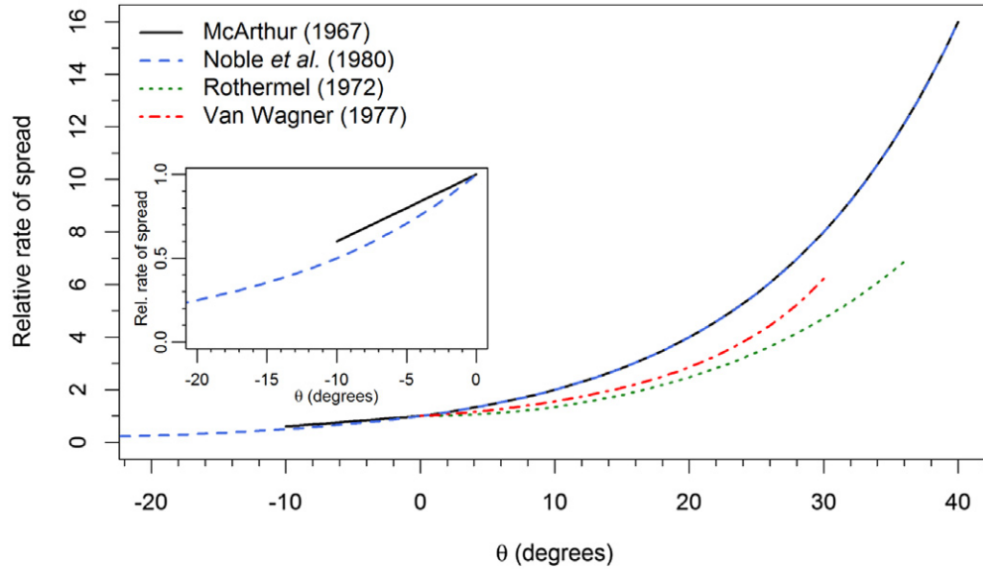
FIGURE 5.2: Functional forms of operational slope factor functions in use around the world over their effective domains. Noble et al. (1980) is a literal interpretation of McArthur's rule of thumb and differs from the function given by McArthur (1966) on lee slopes (see inset). Source: Beck (2000) and Sullivan et al. (2014)

interaction must now be resolved to determine this maximum spread direction. In the literature, there exists mainly two different approaches; the scalar and the vector based. Whereas the first does not account for the directional misalignment of wind and slope, the second considers it for the interaction calculation. For the present, a vector method is required as we need to determine the resulting wind-slope direction.

The vectors methods decompose the wind blowing direction into components aligned and perpendicular to the maximum slope direction (i.e. mathematically speaking, the aspect direction). Then the slope-wind interaction is applied to the component aligned to the slope whereas the other component is kept for the final wind-speed direction determination. Thus, those methods assume that the slope acting direction is aligned to the maximum slope direction (i.e. higher terrain gradient) as this fits the intuition that the heat will be steered upwards (Sharples, 2008).

Sharples (2008) reviewed 7 slope-wind interaction models and proposed to further classify them as additive and multiplicative methods. The additive methods consider the RoS of a no-wind no-slope fire and add correcting factors (Φ_s) (that might, at a time, depend on the original RoS). On the other hand, multiplicative models directly multiply the wind-induced RoS by a slope correcting factor. The classification can be mathematically illustrated by:

Additive methods:

$$\text{RoS} = \text{RoS}_o \cdot (1 + \Phi_s + \Phi_w) = \text{RoS}_o + \Phi_s \text{RoS}_o + \Phi_w \text{RoS}_o \tag{5.12}$$

Multiplicative methods:

$$\text{RoS} = \text{RoS}_o \cdot \Phi_s \cdot \Phi_w \tag{5.13}$$

The main problem with the multiplicative forms is that they do not admit the possibility of a fire burning upslope against the wind as the wind factor (Φ_w) is always 0 for counter-wind propagation. Thus, the only suitable models for the simulator at hand are vector-additive models.

Sharples found that Nelson (2002) and McRae (2004) vector-additive models were the more appealing and promising models due to their mathematical derivation and physical basis (Sharples, 2008). Nelson (2002) method however, was not validated directionally (i.e. only the RoS corrected magnitude without considering its direction was validated) and this is crucial in a perimeter spread simulator. Nelson (2002) and McRae (2004) methods are analogous in the sense that both take into account the vertical motion of air. This is included in Nelson (2002) method as a flame tilting angle (θ_f) and in McRae (2004) as a vertical buoyancy velocity (U_b). The inclusion of this variables could be added as an additional degree of freedom to be calibrated in a data driven framework. However, although this might be interesting to be explored in a future work, a previous attempt conducted by the author (Rios, 2013) showed that converting wind factors into degrees of freedom might lead to an ill-definition of the inverse problem as the wind and its related factors are multivalued and cannot be correctly individually resolved. Therefore their approach was discarded.

An alternative approach is to use Rothermel and Albini slope and wind additive correction factors (Rothermel, 1972a) and combine them in a vectorial way following Finney (1998) approach. The slope factor correlation proposed by Rothermel was a power law form:

$$\Phi_w = C \cdot \left\| \vec{U_{mf}} \right\|^B \tag{5.14}$$

where $\|\vec{U_{mf}}\|$ is the magnitude of the mid-flame wind (i.e. the wind at mid-flame height) and the factors C and B ($0 < B < 1$) were experimentally determined by Rothermel and are function of the fuel hight (δ), load (w_o) and surface-area-to-volume ratio (σ). Beer (1991) pointed out that the upper limit of equation 5.14 could exceed the magnitude of the wind itself meaning that the fire front will propagate faster than the wind. Andrews et al. (2013) recently revisited this effect and stated that it should be bounded such us the calculated RoS do not exceed the mid-flame wind speed.

Rothermel 1D model can be extended following Finney (1998) wind-slope vectorial addition. The wind-slope vector ($\vec{U}_{ws}$) can be defined as:

$$\vec{U}_{ws} = \Phi_s \cdot \big(\sin(\phi - \omega),\, \cos(\phi - \omega)\big) + \Phi_w \cdot \big(\sin(\theta_s),\, \cos(\theta_s)\big) \tag{5.15}$$

Remember ϕ is the local slope, ω the local aspect and θ_s the wind blowing direction (towards where it blows). All angular distances are defined from north cardinal direction.

From vectorial equation 5.15, the module of the wind-slope is:

$$\left\| \vec{U}_{ws} \right\| = \sqrt{\big(\Phi_s \cdot \sin(\phi - \omega) + \Phi_w \cdot \sin(\theta_s)\big)^2 + \big(\Phi_s \cdot \cos(\phi - \omega) + \Phi_w \cdot \cos(\theta_s)\big)^2} \tag{5.16}$$

And its direction:

$$\theta = \tan^{-1}\left(\frac{\Phi_s \cdot \cos(\phi - \omega) + \Phi_w \cdot \cos(\theta_s)}{\Phi_s \cdot \sin(\phi - \omega) + \Phi_w \cdot \sin(\theta_s)} \right) \tag{5.17}$$

Equations 5.16 and 5.17 allows us to compute the resulting wind-slope interaction magnitude and direction. However, this resulting angle is expressed in the horizontal plan. In order to be able to use in into Richard's equations (2.9 and 2.10) it must be expressed in the surface plane, and thus, we must apply the re-projecting equations mentioned in the section above (5.1.1).

5.1.4   Wind-slope verification

A verification process is needed to asses the proper implementation of slope-corrected equations and their correct behaviour. In the following, four simple propagation cases are presented in order to check the proper behaviour of the model as well as to illustrate the practical implications of equations discussed in the previous sections. In all cases we represent the firelets originating from Huygens' elliptical propagation concept (see 2.1.1.1) as they illustrate the difference between the surface and the horizontal plan. Note that the axis represent arbitrary lat-lon coordinates, and thus, they represent the horizontal plane (recall Figure 5.1). Moreover in all cases, north direction is aligned to the positive y-axis direction.

The first case (Figure 5.3) is a simple configuration without slope (flat terrain), and only wind blowing to 225 deg (azimuth) from north. Green and red solid lines are the original and the 60 s-propagated-front respectively. As mentioned before, firelets (grey ellipses) and the propagated front overlap as no slope transformation is required as all expansion takes place at the horizontal plane.
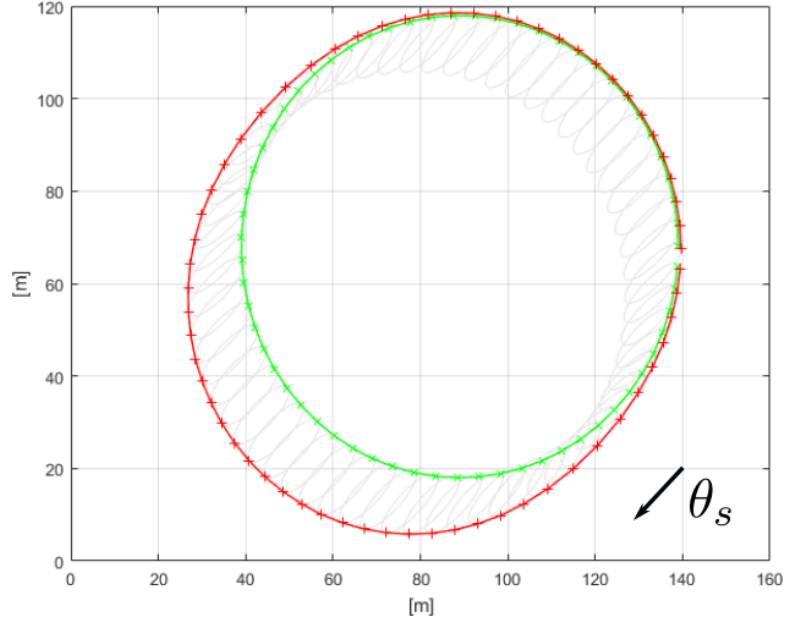
FIGURE 5.3: No slope. Wind at 5m s^{-1} blowing to 225 deg. Green line is original front, red the propagated front after 60 s and grey lines are the elliptical firelets.

For the second verification, we only use a homogeneous slope and no wind. The aspect is set to 0 azimuth (meaning it represents a north face) and the slope is set to 30 deg. Note that now, the firelets are circular since the a and b coefficients in equations (2.4 and 2.5) are the same due to the absence of wind. They, now, do not match the enveloping red curve due to the transformation of plans commented in section 5.1.1. Moreover, firelets live in the surface plane, which is overlapped to the horizontal for illustrative purposes. The pink solid line shows the wind-slope direction propagation in the horizontal plane, that is, the projection of the wind-slope direction (equation 5.17) from the surface to the horizontal plane.
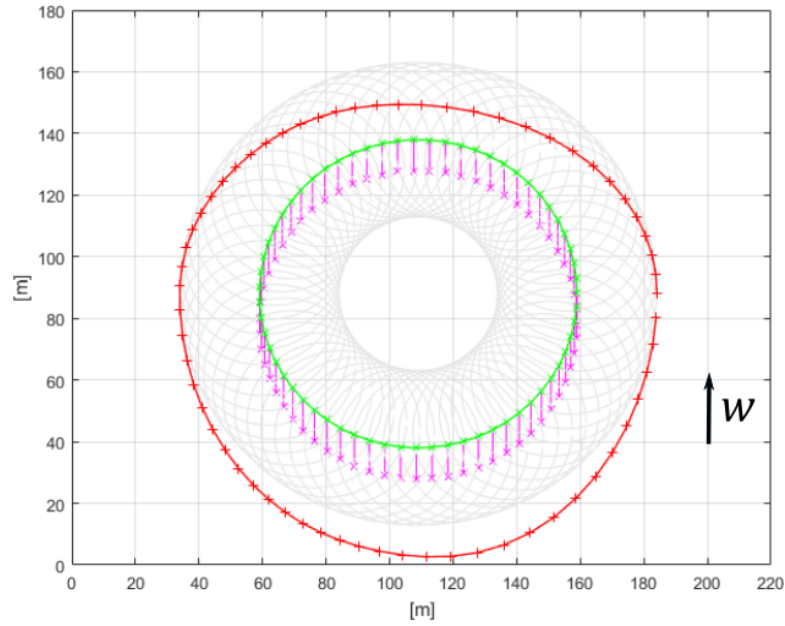
FIGURE 5.4: Aspect=0 deg azimuth (i.e. north face). Slope = 30 deg. Green line, original perimeter. Red line, after 60 s expansion. Pink arrows, firelet propagation direction. The firelets are circular since the a, b coefficients are identical due to no wind condition. They, however, point to the wind-slope direction that is -135 deg (from north).

For the third case, we add together wind and slope. First, wind and aspect have opposite directions. That is, wind flows uphill and, thus, the rate of spread is reinforced. The Figure 5.5(a) shows the results for a 60 s propagation. The final perimeter then (red line) is pushed forward following this direction on the local plan to horizontal plan transformation. Moreover, to see the effect of both terrain transformation and slope impact to the rate of spread, the solid blue line represents the propagated front only considering the wind. Note that the slopes multiplies the RoS by more than three (distances travelled is three times bigger). This is due to the high angle (45 deg) considered for the example at hand. Secondly, in plot 5.5(a) the wind and the aspect directions are not aligned as now the slope faces east (aspect of 90 deg) whereas wind is kept blowing to 270 deg. The firelets enveloping curve are shifted north, showing the slope-wind interaction result.

(a) Aspect 45

(b) Aspect 90
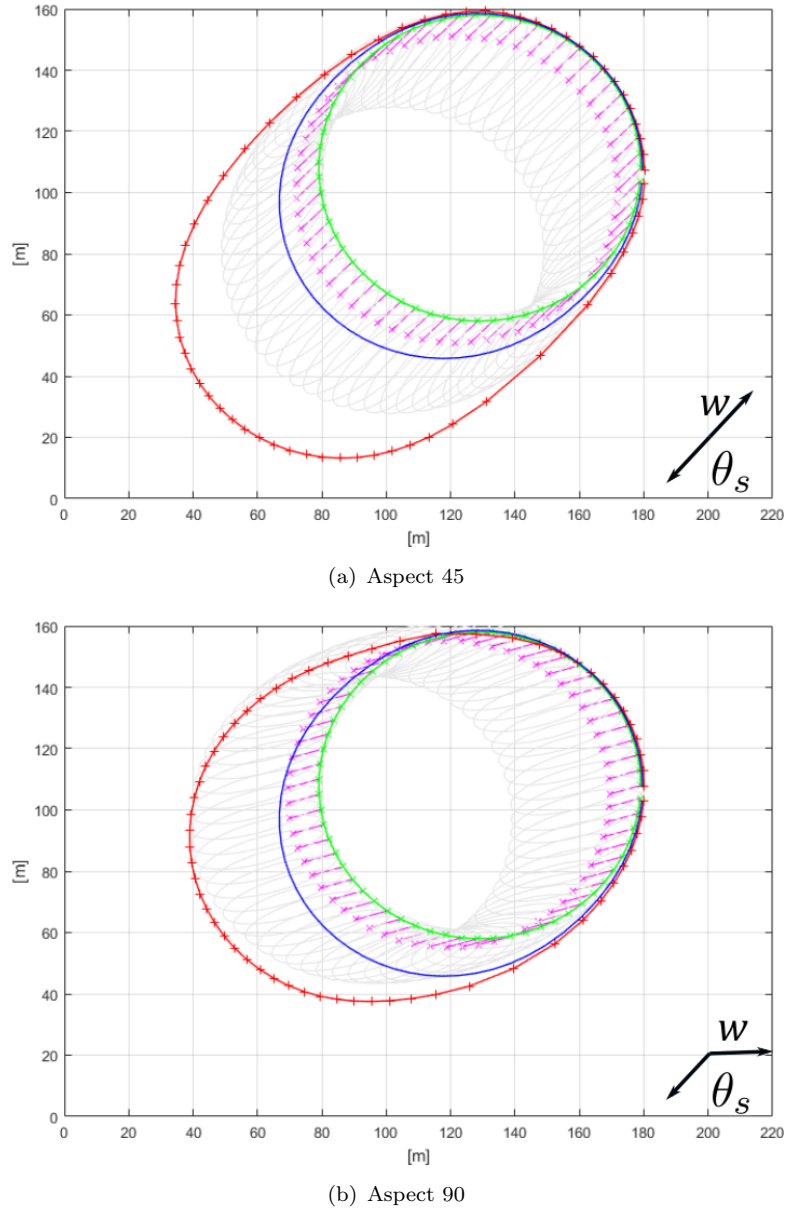
FIGURE 5.5: Effect of adding up wind and slope. Green lines are original fronts, blue lines are propagated fronts with wind but not slope and red lines are fronts with both, wind and slope.

Finally, in order to illustrate the overall behaviour when the wind-slope interaction changes heterogeneously, we run the model over an arbitrary terrain map. The result is plotted in Figure 5.6. Green solid lines are fire fronts depicted every 2 min during 32 min of propagation. The wind is set to 5 ms^{-1}(18 km h^{-1}) blowing to 270 deg. In the background, a hillshade map helps us to identify the slopes and aspect directions. If we analyse the behaviour of the simulations we realize that at the beginning the fire propagates fast uphill the valley versus point A (red letter in Figure 5.6). This

propagation is speeded up by the common contribution of wind and slope as both are aligned. After reaching the pass, the fire starts descending slower towards point B, as now, wind and slope oppose. At same time, the right flank opens to the north as it can find positive slopes in his right side slope. It also tends to spread to the west as this is still the wind blowing direction. This flank opening effect during the descend of a valley is an observation frequently reported by operational fire analysts and highlights the reliable output of the implemented wind-slope interaction.

Between points D and C, the fire is initially spreading fast until going over a ridge. Then the isochrones start getting closer as the front is propagating downhill to the bottom of the valley at point C. The flank opening effect is no possible here as the wind actively prevents it. Last, the backfire spread (toward east) around point D is kept limited due to wind intensity and only some lateral spread is identified.

This case example, should not be seen as a validation (proper validation are provided in chapter 6) but as an illustrative case that verifies the correct implementation of wind-slope interactions and equations projections in the system at hand.
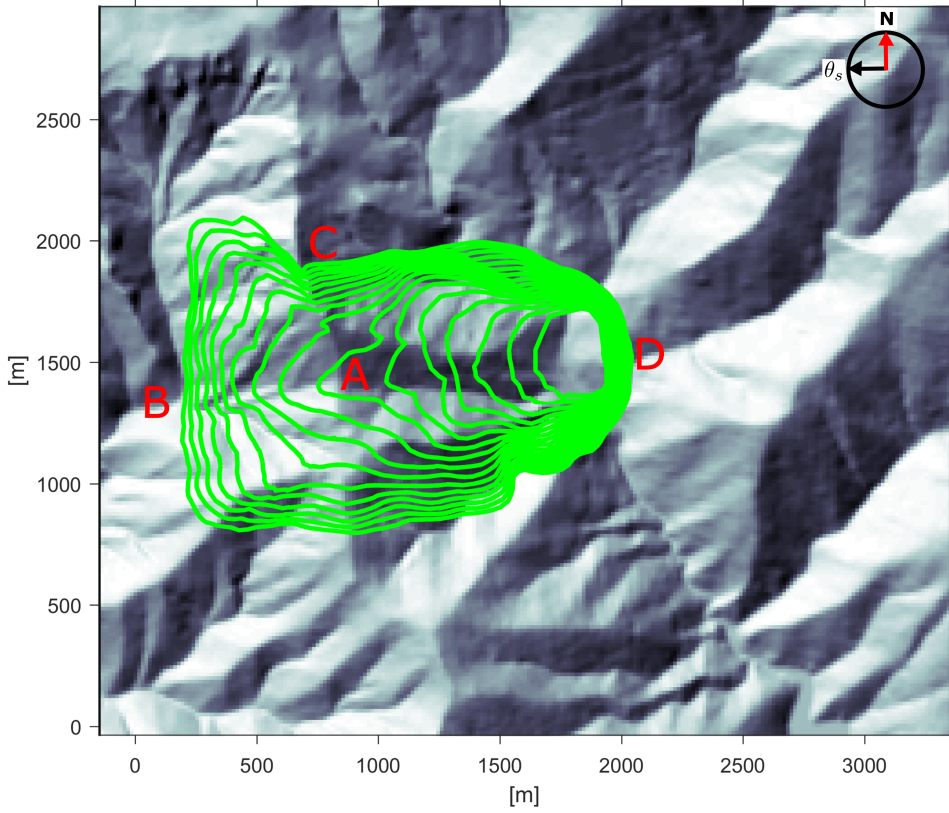
FIGURE 5.6: Illustrative example of wind-slope interaction. WindNinja maps are generated at 15×15 m resolution. 16 isochrones are plotted every 2 minutes simulating a 30 min fire spread under high wind conditions (18 km h^{-1} wind-flame speed) blowing to west (see upper right corner). Letters are highlighted points to ease the discussion in the text.

5.2   Assimilation Strategy

The assimilation process, as explained in section 2.1.2, consists in observing different timestamps of the fire front and optimizing the desired parameters (also called invariants) in order to minimize a given cost function, in our case, linked to the euclidean distance between markers in the simulated and observed front. The strategy implemented in chapter 3 took advantage of fast calculation of the differentials of the model (via an automatic differentiation of the forward model's code). This was coupled to a Tangent Linear Model (TLM) as a methodology to iterative converge the cost function towards the closest relative minimum.

However, after the work developed in chapter 3, the number of parameters to be optimized was upgraded to 7. Those are the ones recalled in Table 5.1. This enlargement of the optimization domain constitutes a challenge for the TLM implemented (see section 2.1.2.1) since the requirement for its successful application is that the objective function

needs to be differentiable and smooth on all 7 directions. Additionally the automatic differentiation step also increases in complexity as every new dimension requires a new program differentiation.

TABLE 5.1: Summary of the reduced Rothermel model parameters (also called invariants).

Parameter		Range (value)	Units
Fine Fuel Load	W	0.067-2.925	$[\text{kg m}^{-2}]$
Surface-area-to-volume ratio	SAV	3753-7270	$[\text{m}^{-1}]$
Moisture content	M_f	5-40	$[\%]$
Moisture of extinction	M_x	12-40	$[\%]$
Fuel Bed Depth	D	0.06-1.83	$[\text{m}]$
Mid-flame Wind Speed	Uf	0.06-1.83	$[\text{m s}^{-1}]$
Mid-flame Wind Dir	θ	0-2π	$[\text{rad}]$

In the following sections we will demonstrate that the cost function used so far display non-smoothness behaviour when the propagation is conducted on complex scenarios. This behaviour might be attributed to complex terrain, heterogeneity of fuel canopy cover, and to the use wind maps (instead of one homogeneous direction). This lack of smoothness will require the development and implementation of a different optimization approach in order to work it around

5.2.1 Objective function domain exploration. An ill-defined problem

To explore the optimization domain topology, the shape of the cost function within the parameters space (as defined in 2.1.2) is plotted for a given set of invariants. The objective is to explore how the cost function changes with respect to the different parameters and thus, assess its smoothness. For simplification, this is conducted for the case with only 4 parameters ($U, I_{Wo}, I_{mf}, I_\delta$, same model as used in 2) but it is extensible to the 7-parameters case. Since the optimization function lives in $\mathcal{R}^4$ the 6 corresponding pairwise 3D projections were represented. Each sub-plot in Figure 5.7 represents then the the projection $\mathcal{R}^4 \to \mathcal{R}^3$ of the cost function when two parameters are changed while the other two are kept constant. The green arrow represents the direction (and step) towards where the optimization algorithm evolves the parameters' set. The roughness (peaks) of the cost function are identified for each combinations, what explains the flawed performance of the optimization algorithm.

In the light of this result, it is clear that the cost function definition must be changed as its inherent roughness causes the optimization strategy to crash to converge to a relative minimum, far away from the absolute one.
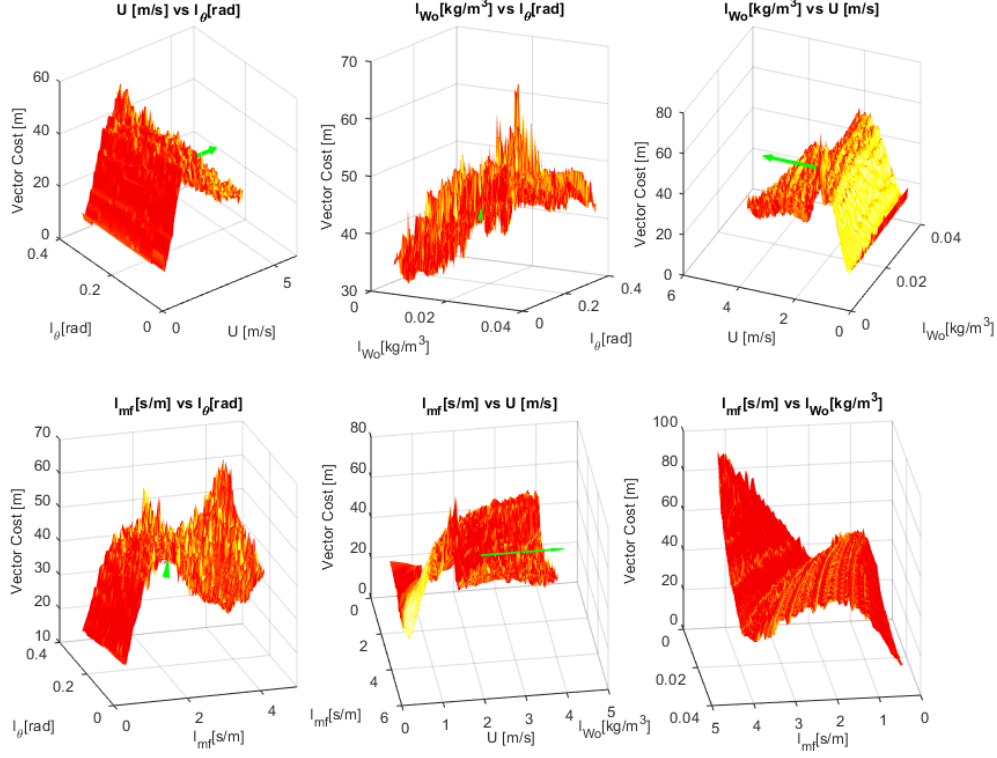
FIGURE 5.7: Domain exploration of the cost function. Roughness is present in all
variables.

The first attempt is to revise the mathematical strategy to account for the distance
between two consecutive fire fronts. As explained in chapter 3, the vector-based strat-
egy was based on quantifying the perpendicular distance between the markers of the
simulated fronts and the observed fronts. This approach however, gives many non-
intersected nodes when the perpendicular direction does not reach the observed front.
This can be seen in Figure 5.8. The currently implemented strategy only provides the
average distance of blue segments, as they where the only ones intersecting the observed
front.

In order to overcome this drawback, we implement a solution based on proxy defi-
nition: when a node of the simulated front does not find a perpendicular intersection
with the observed one, the minimum distance to the model is taken instead. Those new
segments are depicted as orange lines in Figure 5.8.

The change in the cost calculating algorithm does have an effect on the optimization
process as both the cost function values and the optimized fronts are altered. However,
the optimization continues to fail in some situations driving some of the invariants out
of their physical reasonable bounds and the optimized fronts away of observations. This
affects and renders the strategy not robust enough for an operational deployment and
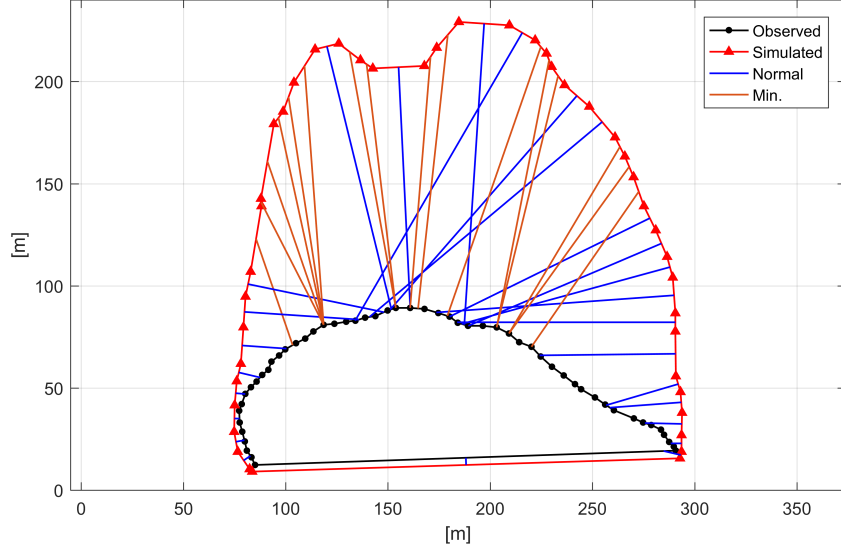
FIGURE 5.8

required to explore other solutions as directly changing the objective function to be optimized.

5.3 Alternative optimization strategy

In order to improve the algorithm, we explore an alternative optimization technique that does not make use of neither the automatic differentiation nor the TLM approach. The strategy is based on an *interior point* method for non-linear programming. This algorithm has the flexibility of switching between a line search (based on Newton step method) and a conjugate gradient method. Steps computed by direct factorization are always tried first, but if they are deemed ineffective, a trust region iteration that guarantees progress toward stationary is invoked (Waltz et al., 2006).

The strategy is evaluated with the reduced 4-invariants model for a known case (see section 2.2.2) The constrained variables region is defined based on feasible values as:

$$
\begin{aligned}
I_{mfw} &= [0.001 - 10](m^3/kgs) \\
I_w &= [0.01 - 1.2](kg/m^3) \\
I_U &= [0 - 20](m/s) \\
I_\theta &= [-2\pi - 2\pi](rad)
\end{aligned}
\tag{5.18}
$$

The first implementation of the method directly uses the sum of the difference area as the objective function. The convergence criteria are set according to two indicators; the step tolerance ($StepTol$) and the function tolerance ($FuncTol$). The first is the relative variation of the optimized parameters, whereas the second is the relative variation of the cost function. Both must be fulfilled for the iterative process to stop. In the present exploration those values are set as:

$$StepTol = 10^{-6}$$
$$FuncTol = 10^{-6}$$

(5.19)

Fronts assimilation results are shown in Figure 5.9. As can be seen in Figure 5.10, the convergence rate is high at initial iterations and flatters until the restrictive tolerance conditions are met.
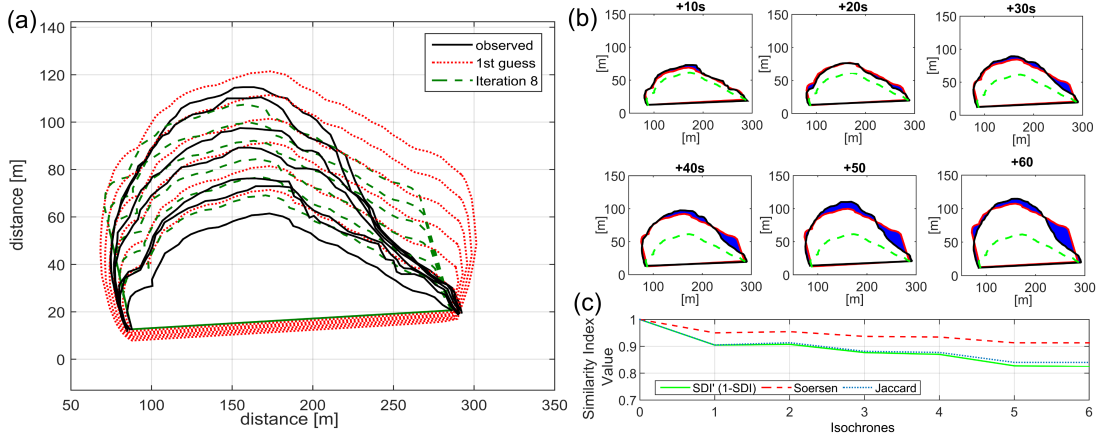
FIGURE 5.9: Plot A assimilation. Green dashed lines are first guess, black solid lines, observations and dashed red lines are best optimization.
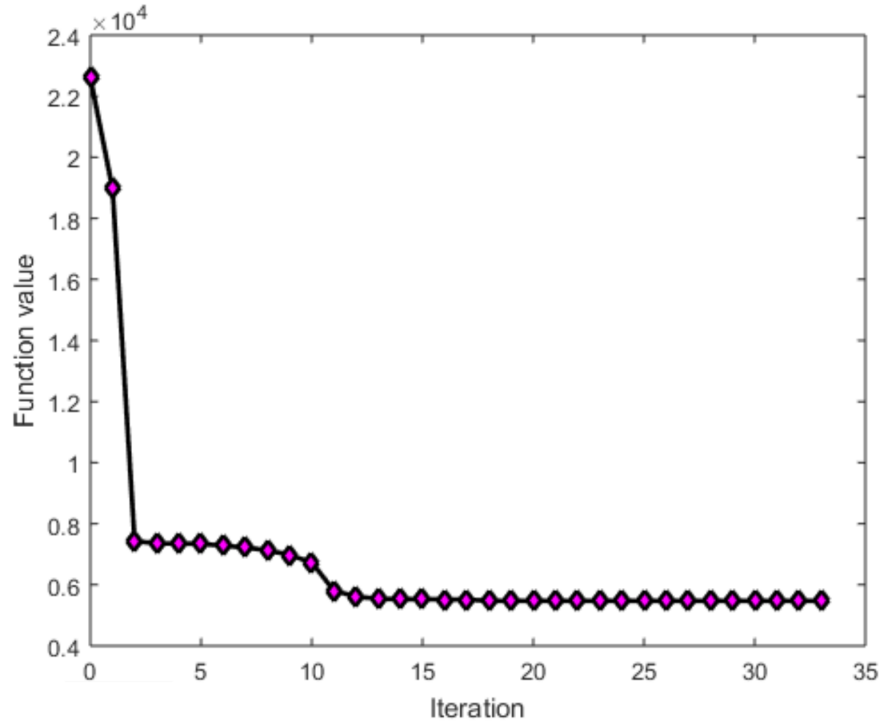
FIGURE 5.10: Convergence of the cost function.

To proof that convergence and objective function behave properly, in Figure 5.11 we plot the cost function centred on initial guess values. Since the cost function lives in $\mathcal{R}^5$ the representation is made by slicing it in pair of parameters. The remaining 3 parameters are taken constant around the initial guessed value. Additionally each sub-plot shows the convergence path (red solid line) to the minimum (pink arrow). Note that, although some plots show the correct convergence direction, some others seem to lead to the opposite direction. Since it is difficult to judge if the minimum is an actual extremum point, Figure 5.12 shows same optimization example but seen from the final converged values. Thus, all final points (pink arrows) should be directed to the local minima. This is clearly not the case, and even some of them point to the opposite direction showing that the cost function is ill-defined and not correctly optimized. The conclusion is, thus, that a new cost function is required in order to properly resolve the parameters of interest.
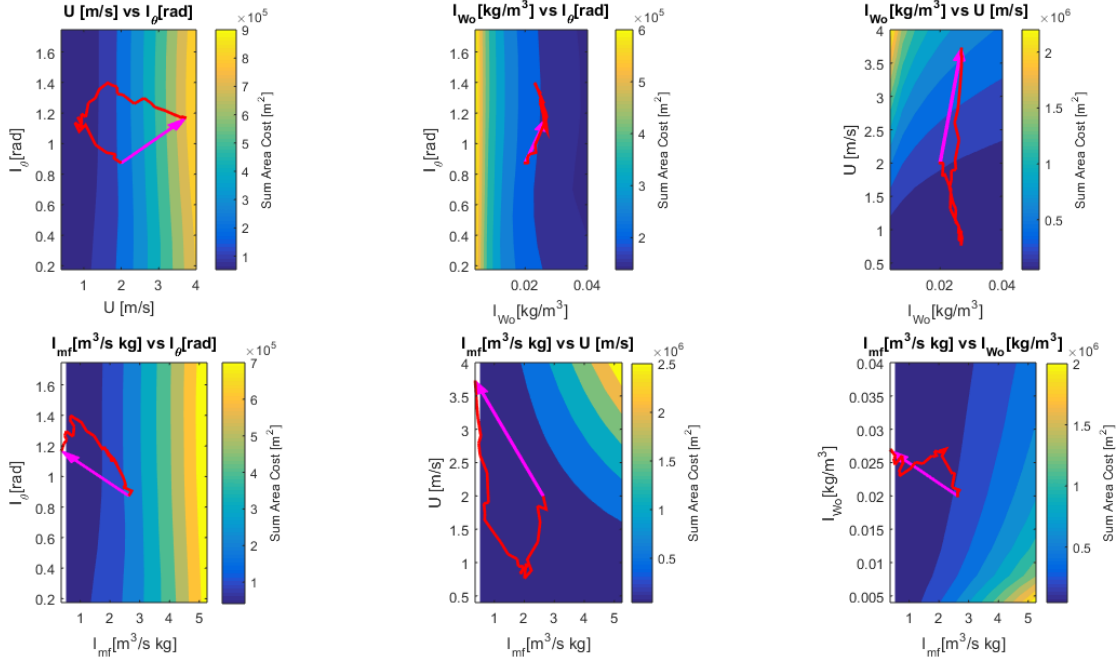
FIGURE 5.11: Convergence path represented in 2D objective function diagram generated from the initial guess. Note that the reason why it does not seem to reach to a minima is the fact that the missing parameters in each sub-plot are taken constant to the initial value.
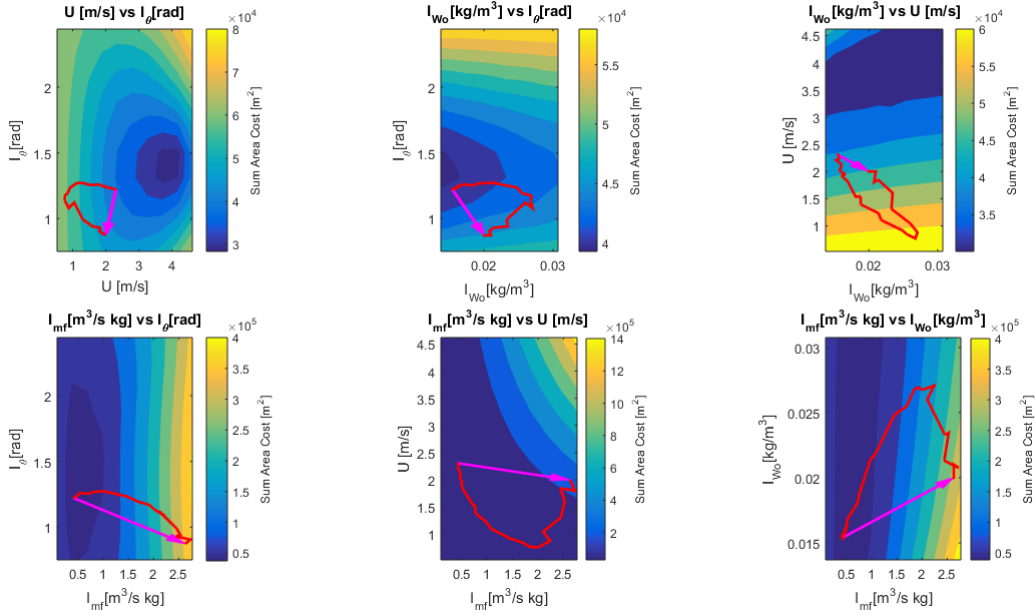
FIGURE 5.12: Convergence path represented in 2D objective function diagram generated from the final convergence point. Note that magenta arrows are pointing to the opposite direction in most cases.

5.4 New cost function exploration

The strategy to find a more suitable cost function to be optimized is based on finding a perimeter match indicator based on areal difference and not in marker-to-marker (vector-based) difference. However, before exploring the possible alternatives it is important to look at the mathematical constraints to avoid an ill-defined optimization problem.

5.4.1 Note on Scaling

The performance of an optimization is crucially determined by how the problem is formulated. One of the most important issues, is the scaling of the input variables (Nocedal and Wright, 1999). If the response of the model varies differentially when similar change is applied to input variables the problem is said to be poorly scaled and optimization might loose performance and can even lead to divergence. To check whereas our RoS model is poorly scaled or not a tornado plot is generated. Tornado plots illustrate the absolute variation in the output when individual input variables are changed from a base line value. To generate a tornado plot, a base case set of values are required for every input variable. Those values are extracted from the range of validity of each invariant previously reported in Table 3.9 and are presented in Table 5.2.

TABLE 5.2: Base case values to generate the tornado plot. Note that slope variable is kept constant during the sensitivity analysis as it is not an invariant (i.e. is input with the DEM).

	w [kg m^{-3}]	SAV [m^{-1}]	δ [cm]	m_x [%]	m [%]	u [km h^{-1}]	slope [%]
Base Value	0.3	15000	30	12	6	10	10

The fact that not all internal default parameters units are in the SI system responds to a scaling need already. Units change is the first strategy to balance the response of a model (Nocedal and Wright, 1999) as it alters relationship with the rest of variables. The generated tornado plot is depicted in Figure 5.13. A 20% change is applied to each input variable. Base case is depicted with a vertical line at 76 m min^{-1}. Blue bars represent direct proportionality (that is, if the input value grows, so does the output) whereas red bars represent the contrary (i.e. inverse response). Note that wind direction is the only invariant missing as its effect is not present in the RoS value but in the front spread shape and direction. All 6 invariants except for fuel load (w) respond to the model in the same order of magnitude meaning that the problem is correctly scaled for them. Fuel load, however, has a lower response. It is less than two order of magnitude lower sensitive than wind speed response. Although it should be considered when analysing optimization results and particularly when selecting the initial guess value, the difference is tractable by most common optimization strategies.
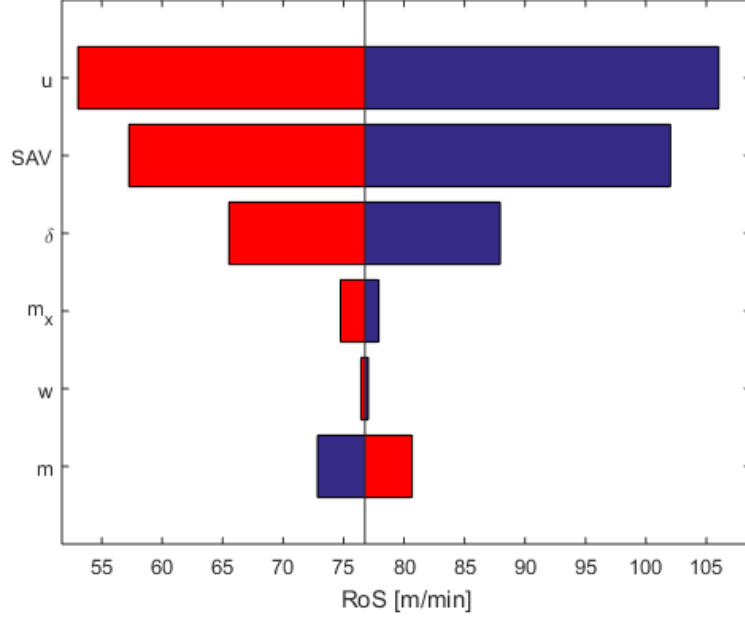
FIGURE 5.13: Tornado plot for the reduced RoS model. Blue bars are direct proportionality whereas red represent inverse relationship between input and output values.

5.4.2 Alternative cost function

As previously introduced in the section 2.2.2.1, the difference between two perimeters can be expressed using several alternative index. Cui and Perera (2010) and Filippi et al. (2014c) conducted several investigations on possible approaches and compared them. The most adequate ones for our problem are described in the following:

Original Index

$$
\begin{aligned}
\text{AreaCost} &= \sum_{i=0}^{i=N_f} W_i \cdot (M_i \oplus O_i) \\
\text{SDI} &= \sum_{i=0}^{i=N_f} W_i \cdot \frac{M_i \oplus O_i}{O_i} \\
\text{S\o RENSEN} &= \sum_{i=0}^{i=N_f} W_i \left(1 - \frac{2 \cdot (M_i \cup O_i)}{M_i + O_i} \right) \\
\text{JACCARD} &= \sum_{i=0}^{i=N_f} W_i \left(1 - \frac{(M_i \cup O_i)}{M_i + O_i - (M_i \cup O_i)} \right)
\end{aligned}
\tag{5.20}
$$

However, given that our systems needs a first front to be initialized, the area covered by this initial front could bias the final index. Te reason is that if the initial area is

much larger than the predicted one, final value is masked by the initial match. Thus, all index presented above can be modified to account for this by removing the initial area. The Venn diagrams in Figure 5.14 helps to illustrate the modification.

Redefined Index

$$\text{SDI}_a = \sum_{i=0}^{i=N_f} W_i \cdot \frac{M_i \oplus O_i}{O_i - I_a}$$

$$\text{SøRENSEN}_a = \sum_{i=0}^{i=N_f} W_i \left(1 - \frac{2 \cdot (M_i \cup O_i - I_a)}{M_i + O_i - I} \right) \qquad (5.21)$$

$$\text{JACCARD}_a = \sum_{i=0}^{i=N_f} W_i \left(1 - \frac{(M_i \cup O_i - I_a)}{M_i + O_i - (M_i \cup O_i - I_a)} \right)$$

Where M_i and O_i are de modelled and observed fronts' areas in a given time i. The I_a is the initial front area used to perform the perimeter expansion (see Figure 5.14). W_i is the weighting matrix that will be deeply explored in 5.4.3. Note that subtracting the area in the numerator of the SDI_a case do not modify the result.
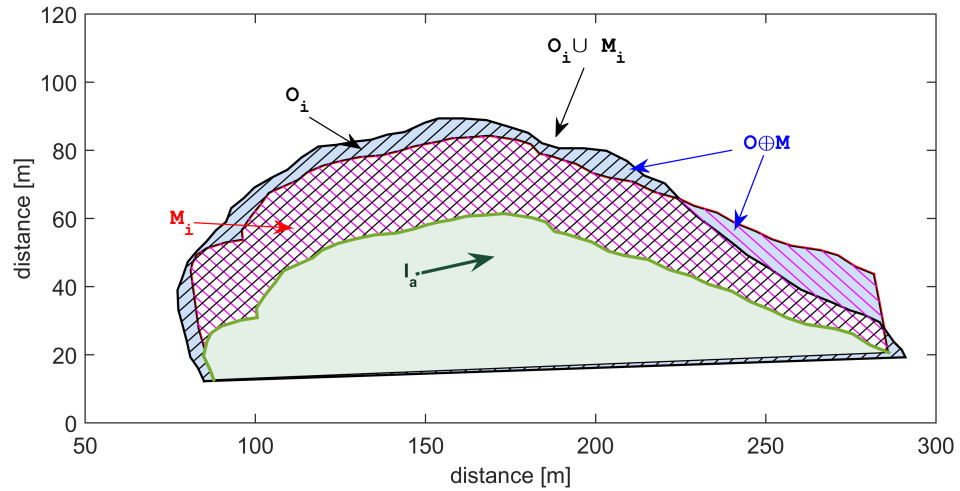
FIGURE 5.14: Venn diagram to identify similarity index calculations

The transposition of union and intersection operators into an analytical formulation is not straight forward. Despite the area under a curve (i.e. integral) could be easily formulated, a general closed perimeter does not comply with a function definition of single element image ($f : X \rightarrow Y$) and thus, cannot directly be integrated. In practice, this obliges to implement those operators as piecewise functions, rendering them non-differentiable. This limitation, does not affect our strategy any more as the new

algorithms explored are differentiation-free. In consequence, in the following, we will explore the index presented in 5.20 and 5.21 and compare their performance.

5.4.3   Weight factor exploration

As introduced in the cost function definition (see section 2.1.2) a weighted factor can also be included in the index-based cost function formulation as:

$$\mathcal{J}(\mathbf{q}) = \sum_{t=t_i}^{t_f} \frac{\sqrt{S_i^T \cdot W_i \cdot S_i}}{n_i} \tag{5.22}$$

Where S_i is the similarity index value for a pair of observed-simulated fronts in a given time. The purpose of the weighting factor (W_i) is to give different weight to isochrones assimilated in different times or with divers accuracy. That is, for example, more value could be attributed to the latest acquired isochrones than to the ones assimilated at the beginning of the fire. Four different weighting strategies are proposed as a study cases. Those strategies could be better tuned considering extra factors that effect fire front identification and extraction such as platform used, fire intensity or segmentation performance (see Valero et al. (2016) for more details on fire front extraction).

The four weight function explored in this chapter are called and defined as:

$$\text{idty} \longrightarrow W_i = \mathbb{I} \tag{5.23}$$

$$\text{lin} \longrightarrow W_i = i \tag{5.24}$$

$$\text{linN} \longrightarrow W_i = \frac{i}{N_f} \tag{5.25}$$

$$\text{expN} \longrightarrow W_i = \frac{e^{i/N_f}}{e^{N_f}} \tag{5.26}$$

Where the i index is the cardinal of every isochrones processed, from the oldest to the newest.

The identity case ($\mathbb{I}$) considers all isochrones equivalently and, thus, all have the same influence on the cost function. If isochrones are homogeneously distributed on time, the linear case (lin, eq. 5.24) distributes the weight inversely to the isochrones elapsed time. That is, if the time step is two times greater than another one, it has half the weight on the cost function. The linear normalized case (linN, eq. 5.25) implies similar weight

distribution but it is independent of the number of the assimilated fronts. That is, 10 fronts will be identically weighted as 40 fronts. This was not the case in the lin distribution, as the last front in the 40 fronts assimilated case will have four times more influence than the last one of the 10 assimilated fronts case. Finally, the exponential distribution (expN, eq. 5.26) gives additional importance to the last assimilated fronts whereas the first ones are nearly neglected. This represents an extreme case where initial data is to be discarded due to uncertainty or a change in fire dynamics conditions.

5.4.4 Verification

In order to identify which of the assimilation index (see section 5.4.2) and weighting factors (see section 5.4.3) performs better within our optimization strategy we run optimizations scenarios using all possible combinations of index and weight factors. The plots A and A2SE previously used to evaluate the system performance (see section 2.2.1) are not of interest for current evaluation as they are flat terrains, thus, the main challenge and source of front irregularity is missing. The evaluation is then performed with a synthetic case generated with a simplified version of the spread algorithm that does not contain the fuel model introduced in chapter 3. By this, we can use a synthetic case generated on terrain that do not perfectly match our algorithm (i.e. there is no perfect solution as the generating spread model and the one used in the optimization differ), which actually mimics the case in reality, where all the physical phenomena are never captured entirely.

Figure 5.15 shows the 4 isochrones representing the 15 first minutes of the fire in a foot of a hill. Propagation is then being driven by a mixed effect of slope and wind.

Starting from the same initial guess, the assimilation process is launched using the seven different objective function implementation functions introduced in section 5.4.2. The front spread results are depicted in Figure 5.16. The first element to highlight is the dispersion of the results depending on the cost function employed. Despite this dispersion, all the runs try to resemble the observations and there is no flagrant misconvergence.
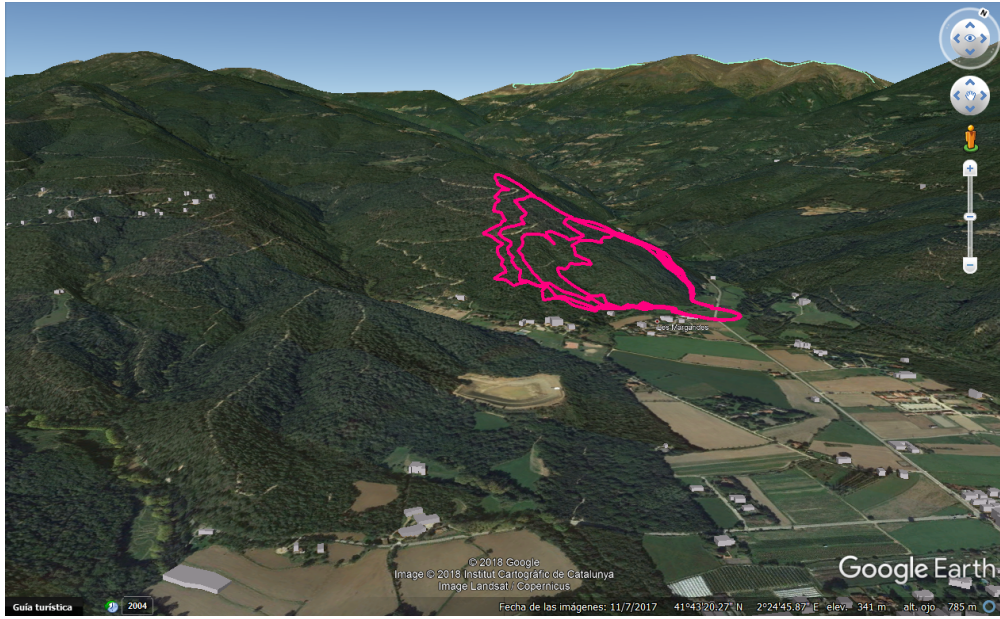
FIGURE 5.15: Synthetically generated fronts displayed in Google Earth viewer.
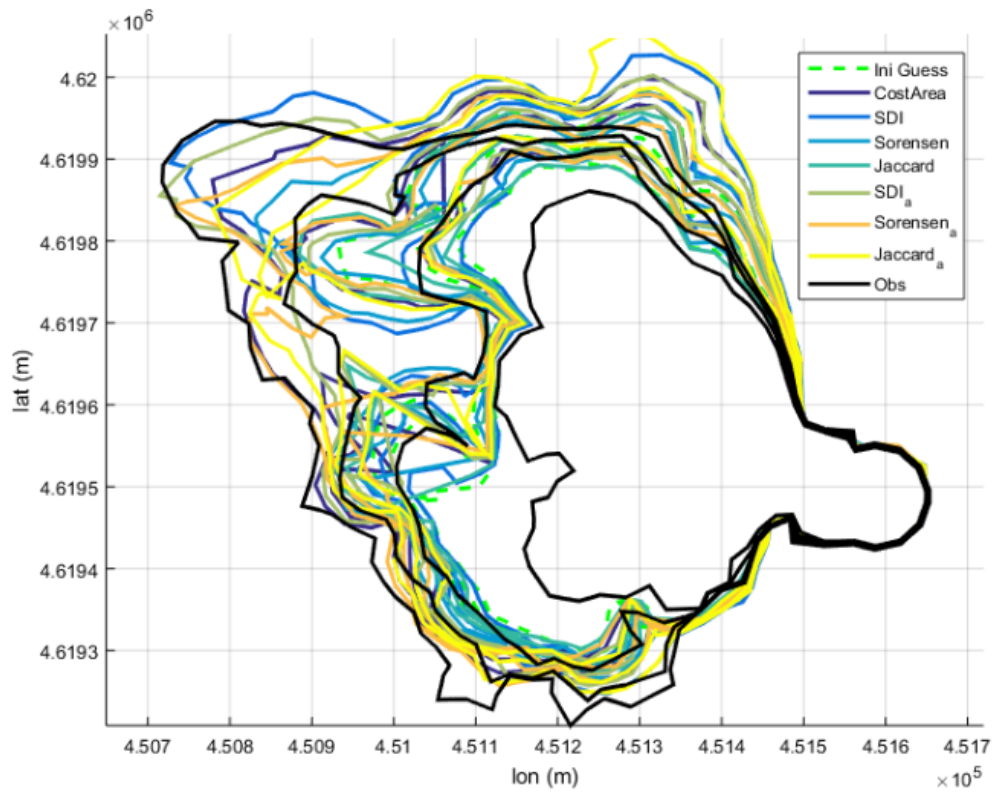
FIGURE 5.16: Comparative fronts plots for the 7 invariants in plot Montseny (synthetic data, 4 assimilated fronts)

To better compare the performance of each cost function, the final cost value for

each of the three fronts (when the simulation converges) is summed and displayed in the third panel of Figure 5.17. Recalling the meaning of this index, the closer to unity the more identical are the simulated and observed fronts. Thus, a result of three (for the addition of the three fronts studied) would give perfect similarity. The SDI_a and the $Jaccard_a$ strategies are the ones giving better results followed by the original SDI. Additionally, regarding the computational time required and the number of iterations to reach convergence (first and second panels in Figure 5.17), $Jaccard_a$ and $Jaccard$ are the best performing ones. Additionally, $Sorensen$ is the one that provides the better perimeters matching results
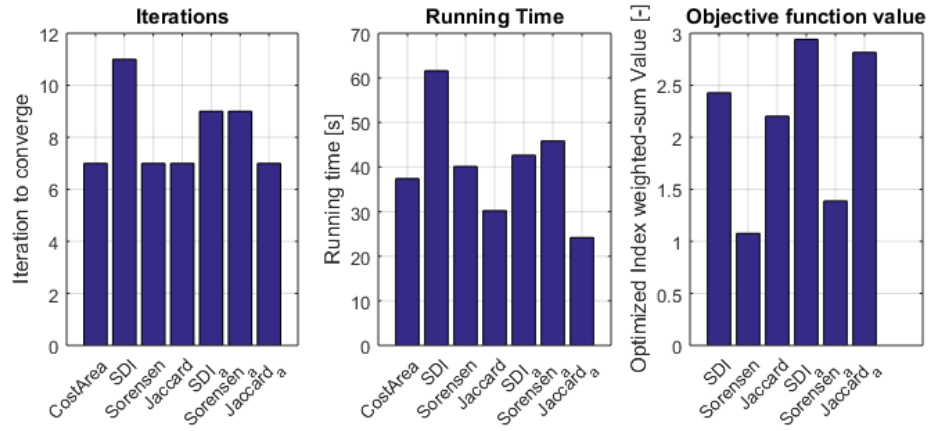
FIGURE 5.17: Convergence performance for the 7 invariants in plot Montseny for different cost functions (synthetic data, 4 assimilated fronts)

Regarding the individual convergence of each of the seven invariants, not all converge to same value. The main discrepancy is found in the fuel moisture for the $Sorensen_a$ objective function and more remarkably, on the fuel load. For this case, both the $CostArea$ and the $Jaccard_a$ values differ largely from the rest of the cases showing that different relative minima are still present in the problem.
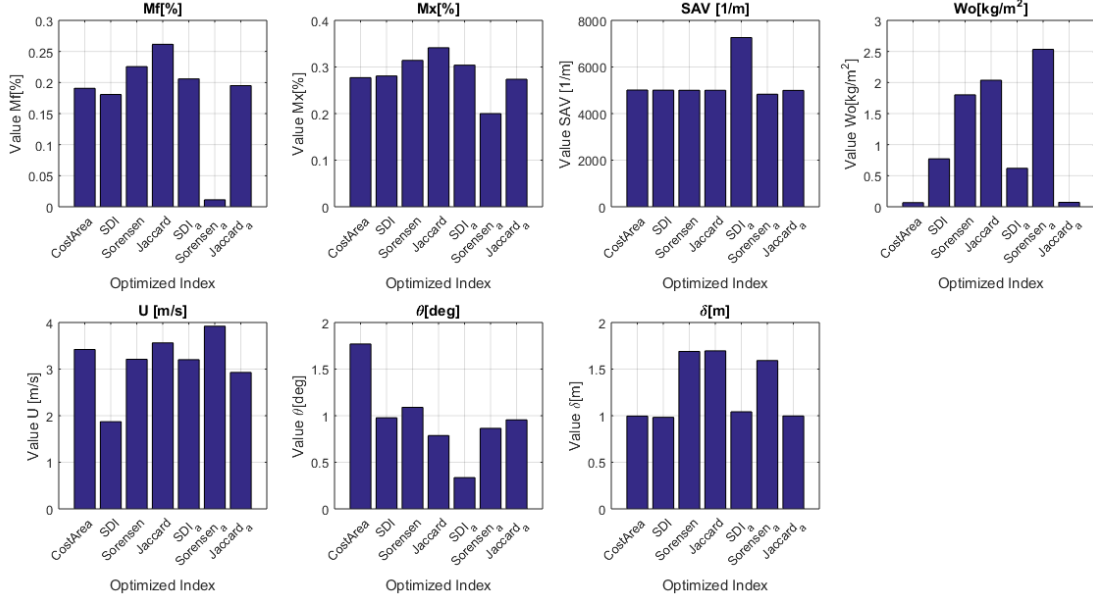
FIGURE 5.18: Individual convergence for the 7 invariants in plot Montseny (synthetic data, 4 assimilated fronts)

Finally, an exploration of the impact of the weighting factor in the assimilation process (see section 5.4.3) is conducted by comparing the system performance and convergence for a given cost function (SDI for the example at hand) and the 4 different weighting strategies. The results on the converged fronts and on the invariants convergence are gathered in Figure 5.19 and 5.20 respectively.
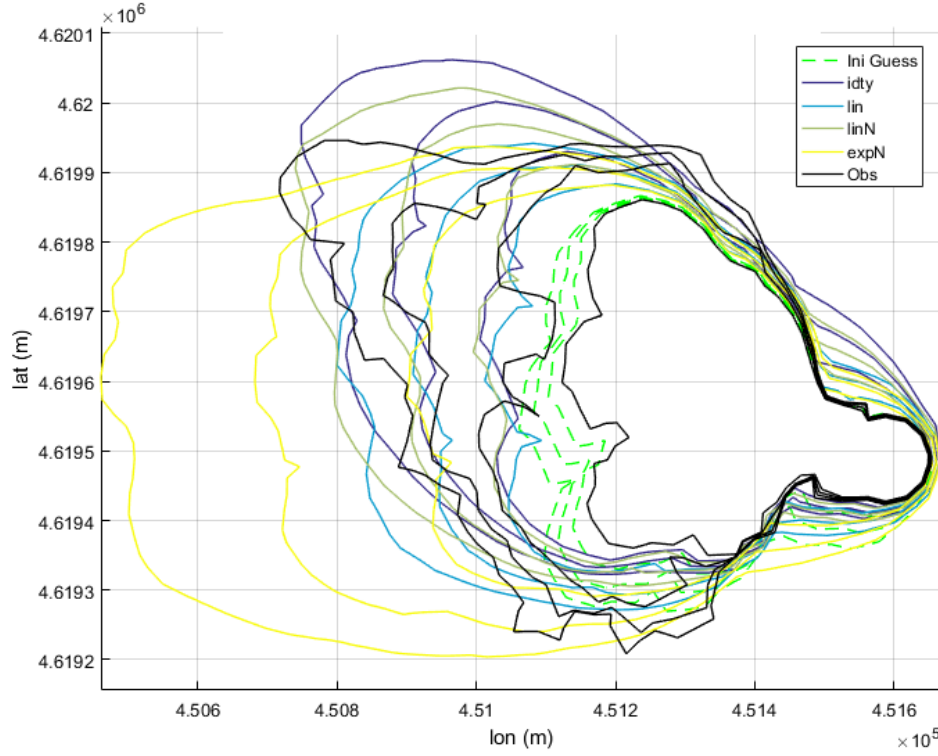
FIGURE 5.19: Individual convergence for the 7 invariants in plot Montseny (synthetic data, 4 assimilated fronts) for the four different weighting strategies
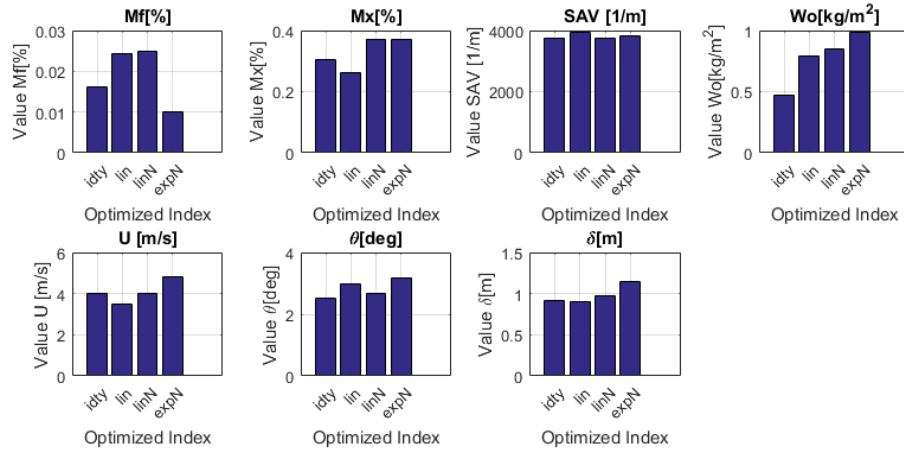
FIGURE 5.20: Individual convergence for the 7 invariants in plot Montseny (synthetic data, 4 assimilated fronts) for the four different weighting strategies

Figure 5.19 and 5.20 show that the choose of a weighting factor does have an influence on the final solution. Both, the shape and the final invariants value are altered. The *Idty* (identity) and *linN* (linear normalized) cases are the ones that better matches the observation. Since the true physical impact of this factor could not be clarified, more

research is needed to use it safely for operational proposes For this reason, the identity case is the one selected to be employed for the rest of this thesis.

5.5   Integration time step

In order to computationally solve the front spread differential equations (see 2.9 and 2.10) it is necessary to stablish an integration time step. When Richards and Bryce (1995) originally derived those equations they used 10 s as integration time which provided good convergence. However, they made use of homogeneous fuel and terrain scenarios and, thus, it is necessary to explore the effect of the integration time in more generic and complex scenarios. To achieve this, we selected two real digital elevation models, namely, Cardona and Montseny. Their hillshade representation is displayed in Figure 5.21

(a) Cardona, 9.0km x 9.0km   (b) Montseny, 4.5km x 4.5km
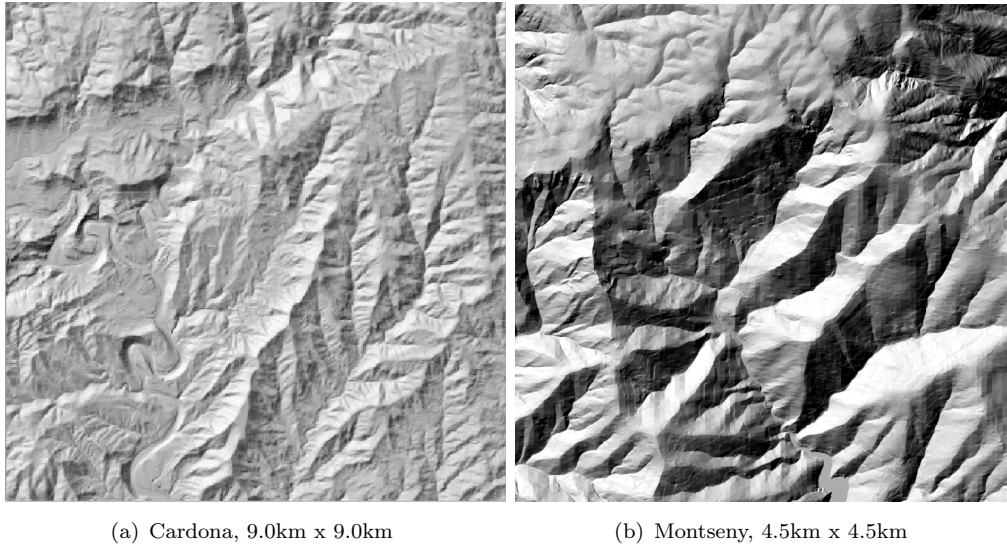
FIGURE 5.21: Hillshade representations of both terrain scenarios used to generate the synthetic data.

The effect of the integration time is analysed by comparing 10 different time steps, uniformly distributed from 05 s to 95 s. The fronts are only displayed every 1000 s (16.6 min) for a total simulation time of 4000 s (66 min). Results for both scenarios are depicted in Figure 5.22.

In both cases, the convergence of the results for the integration time steps up to 25 s is similar (blue lines in Figure 5.22). When exceeding the 30 s, the effect is more relevant. This phenomena is even more remarkable in Cardona case as the results for the different time steps (in particular the ones exceeding 60 s) are more scattered in space. For this scenario, the 95 s step run differs from the 5 s run by more than 500
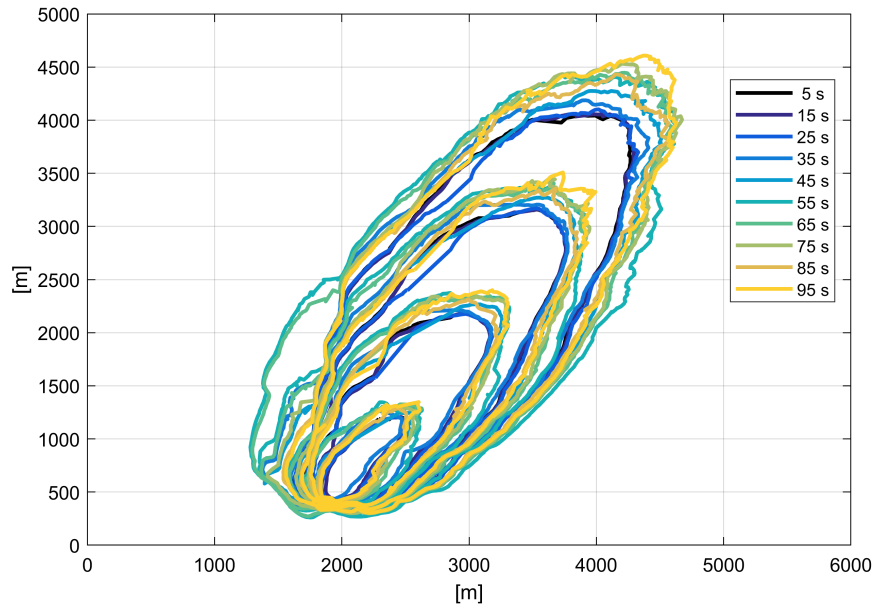
(a) Cardona case
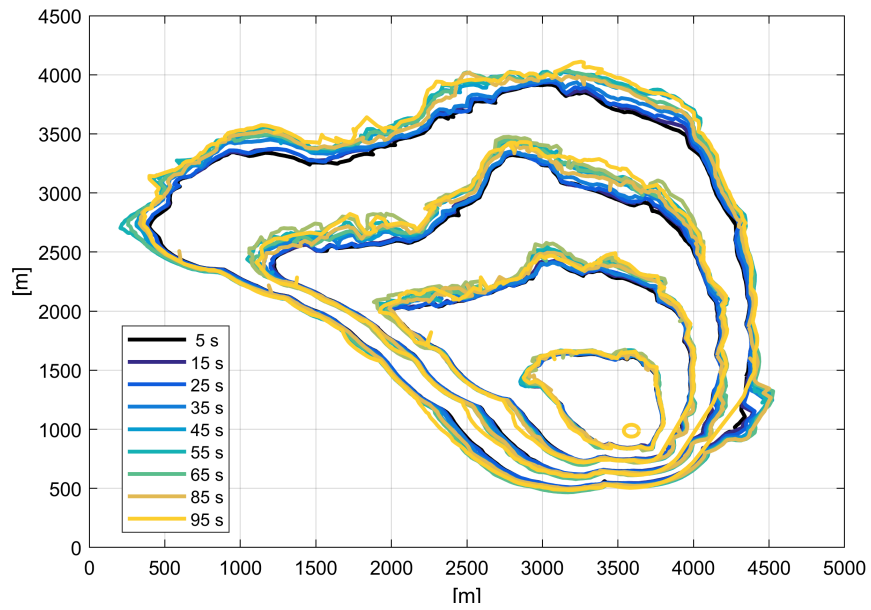
(b) Montseny case

FIGURE 5.22:   Fronts differences when using different integration time steps in Montseny case.

m in the head of the fire (yellow front and dark-blue front). This difference exceeds the acceptable uncertainty for operational purposes.

Besides exploring the uncertainty linked to the integration time, it is necessary to explore its impact on the computational time. Elapsed computing time for both cases are depicted in Figure 5.23. Results show that the required computing time decreases drastically when extending the integration time step from 5 to 35 seconds. From this point on, it tends to stabilize. The speed up is of the order of 8 times (5 s over 40 s) for Montseny case and of about 5 times (3 sec over 17 s) for the Cardona case. Those speed up are extremely relevant for an operational application as they might allow delivering the results in a positive lead time. In Montseny case, the peak at 75 s of time step corresponds to a mathematical instability of the model (a particularly difficult loop clipping scenario). Although it should be looked at as a singular event, it is demonstrated that the effect of increasing too much the integration time may lead to a clipping problem that ends up requiring more computational time than the one saved due to its increase.
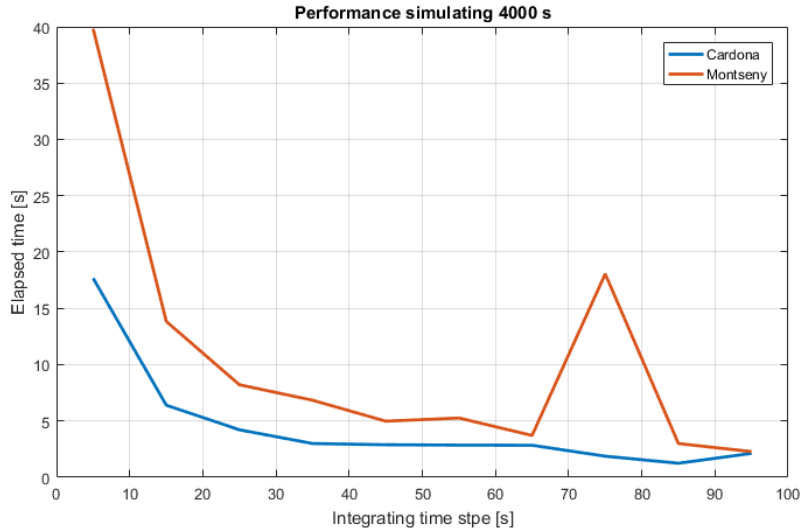
FIGURE 5.23: Elapsed computing time dependency on integration time step for 4000 s (using wind maps version)

Considering all this observations, the default time step is set to 25 s which is more than 4 times faster than 5-10 s.

It is worth to mention that the total computing time is also affected by the number of front markers of a given front. This quantity depends on both, the desired spatial resolution and the physical length of the fire front itself. The number of calculations of the `loop_clipping` (see 2.1.1.2) scales up as a power of two. Thus, changing from 10 to 100 markers to describe the fronts isochrones involves 1000 times mores operations. For

the current case, the threshold values mentioned in sections equation 2.13 are considered as good trade-off between accuracy and computational efficiency.

5.6 Conclusions

This chapter presented the improved implementation of the assimilation algorithm initially explored in Chapter 2. The new algorithm (named *SmartQFIRE*) can now handle terrain and fuel complex heterogeneities. The wind-slope interaction was revisited in order to implement a proper solution consistent with last research findings. During this chapter it was also shown that the previous automatic-differentiation based optimization framework was not capable of resolving the new complex scenarios due to lack of smoothness of the vector-based cost function. Thus, new solution had to be worked out and implemented. For this, seven different cost functions were developed and analysed to identify the best performing one. The results do not allow to select a single one, but indicated that $Sorensen$, $Jaccard_a$, $Jaccard$ and SDI are the most effective and over-performing ones. Finally, the impact of the integration time step to solve the front propagation differential equations is also explored. After taking into consideration the spatial dispersion of results and the increase on the computational time, the best trade-off solution between accuracy and computational efficiency was stated at 25 s.

Chapter 6

SmartQFIRE: Optimization Exploration

The inverse modelling methodology presented until chapter 5 has proofed to be a capable system that helps to gain information of a real fire evolution. However, in order to be effective, the forecasting horizon needs to be enlarged into several hours to become a useful tool for operational managers. Whereas the forecast model can easily be ran on those time scales, (once the proper invariants' set is identified) the inverse modelling pre-process is much limited. The computational cost required to process several hours of minutely provided fire fronts is unaffordable from an operational point of view. Thus, in this chapter, we perform an in-depth analysis of different optimization strategies to upgrade the SmartQFIRE implemented version in order to foster its applicability using the most suitable optimization algorithm for the problem at hand. The work included in this chapter was accepted and presented on an oral talk in the *International Conference on Computational Science (ICCS), 11-14 June, 2017* (Rios et al., 2017b).

6.1   Optimization strategies testing

The successful implementation and performance of the overall system is crucially dependant on the capability of the optimization algorithm to rapidly converge the invariants to the closest possible vicinity of its true value. In this section, six different optimization strategies are compared in terms of computing time and assimilation accuracy (capability to mimic observations). Additionally, the forecast potential and adequacy after each strategy is also compared. Out of the six strategies used, three are line search algorithm (with single starting point) and three are global optimization strategies. The need for

the exploration of both types of optimizations is due to the fact that whereas line search strategies are usually faster than global search, they do no guarantee the convergence to a global minima.

The correct way to compare the different strategies is to benchmark them with a long-enough scenario that allows to check their efficiency and capability to produce a set of invariants that correctly matches the observations. However, the lack of ready-to-use methodologies and sensors systems to monitor real fire-event during long period makes it impossible for us to use real data. Instead, synthetic data generated with a widely spread wildfire simulator are considered to challenge all different algorithms.

6.1.1 Synthetic data generation

The synthetic data is generated using FARSITE (Finney, 1998) simulator, as it is one of the most extendedly used and validated wildfire spread model. FARSITE was developed in 1998 by coupling the Rothermel fire spread model with Huygens light propagation mechanism. The software solves Richards and Bryce (1995) equations for fire spread in a vector approach using, as input data, raster layers to describe the simulation domain characteristics (canopy cover, terrain, moisture content, etc.). Additionally, it also accepts WindNinja wind maps as an input wind raster data layer to account for the local wind conditions. All this features render FARSITE a widely used tool in the US (broadly spread through Federal and State Land Management agencies) for both operational and prevention applications (Sullivan, 2009c). The fuel modelling is based on BEHAVE (Andrews, 1984) although the user can self-define its own models and calibration parameters to apply to the standardized fuel descriptions.

In order to compare the assimilation and forecasting performance of different optimization algorithms, two base cases scenarios (Case A and Case B) are generated in a homogeneous fuel cover whose characteristics are gathered in Table 6.1. Those scenarios are based on *Vall-llobrega* real fire that will be used for a more detailed algorithm validation later on this thesis (see section 7.3.1). The only difference between both cases is the wind speed. Whereas Case A has no wind (only topography and fuel governs the fire spread), Case B has a 5 m s^{-1} mid-flame wind speed blowing homogeneously from 240 deg (see Figure 6.1b). The run is launched for three hours depicting isochrones every 5 min WindNinja maps are input to FARSITE to account for topographic wind with a 15 m grid resolution. All initialization parameters (except for the wind) are gathered in Table 6.1. Since isochrones are output every 5 minutes, 37 isochrones are available for the entire 180 minutes run. Fire fronts locations for both scenarios overlapping the hillshade terrain representation are plotted in Figure 6.1. Although at first sight it might seem that difference between both cases are small, a closer look at Figure 6.1 shows that the final burnt area already differs some hectares. Another remarkable difference

is that Case B (with wind blowing from north-eastern corner in 6.1b plot, grows faster and further than Case A in that direction. Additionally backwards fire spread is also slightly faster although propagating against the wind. Whereas it is out of the scope of the current section to validate FARSITE simulations, a possible explanation for this effect might be the fact that in the presence of wind, the RoS aligned to wind direction increases dramatically as stated by equation 2.1. This, when coupled to a Huygens' propagation model by equations 2.4, 2.5 and 2.6 can lead to a greater backwards RoS than a no-wind case. Finally, is worth noticing that not only the fire fronts spread faster but their shape also changes in both scenarios. This is indeed due to the global wind difference but is further amplified by the fact that WindNinja maps introduce local terrain-based variability in the wind magnitude and direction.

TABLE 6.1: Input parameters to generate FARSITE case simulations. Bold parameters are those that are also considered constant in SmartQFIRE simulator.

Particle properties

Particle classes		Fuel Load [kg m^{-2}]		SAV [m^{-1}]		M_f[%]
Dead 1h (0.0–0.6cm)	$w0_{d1}$	0.449	sv_{d1}	6562	m_{d1}	3
Dead 10h (0.6–2.5cm)	$w0_{d2}$	0.898	sv_{d2}	357	m_{d2}	4
Dead 10h (2.5–7.6cm)	$w0_{d3}$	0.449	sv_{d3}	98	m_{d3}	5
Live Herbaceous	$w0_{lh}$	0	sv_{lh}	-	m_{lh}	-
Live Woody	$w0_{lw}$	1.123	sv_{lw}	4921	m_{lw}	90

Fuel bed properties

Depth	d	1.83	[m]
Moisture of Extinction (dead fuel)	m_x	20	[%]
Average bulk density	ρ	**513**	[kg m^{-3}]
Total mineral content	s_t	**5.5**	[%]
Effective mineral content	s_e	**1**	[%]
Heat content	$heat$	**18609**	[kJ kg^{-1}]

Regarding the parameters initialization (initial guess), it was deliberately established be far from the true values (i.e. values used in FARSITE) in order to challenge the system with an ambitious case scenario. The true values gathered in Table 6.2 are the reduced model equivalent variables generated from the complete set presented in Table 6.1. This conversion to the SmartQFIRE input values can be performed using the coefficients presented in Table 3.11. The fronts produced with the considered initial guess (dashed lines in (a) ad (b) panels in Figure 6.2) are different for each case and differ largely from the observed fronts. The explanation is that the initial synthetic front taken to initialize the run is also different in both cases. This minor difference is amplified by the terrain-wind interaction and outputs a slightly different first initial guess even though the set of parameters are the same. This already highlights the sensitivity of the model
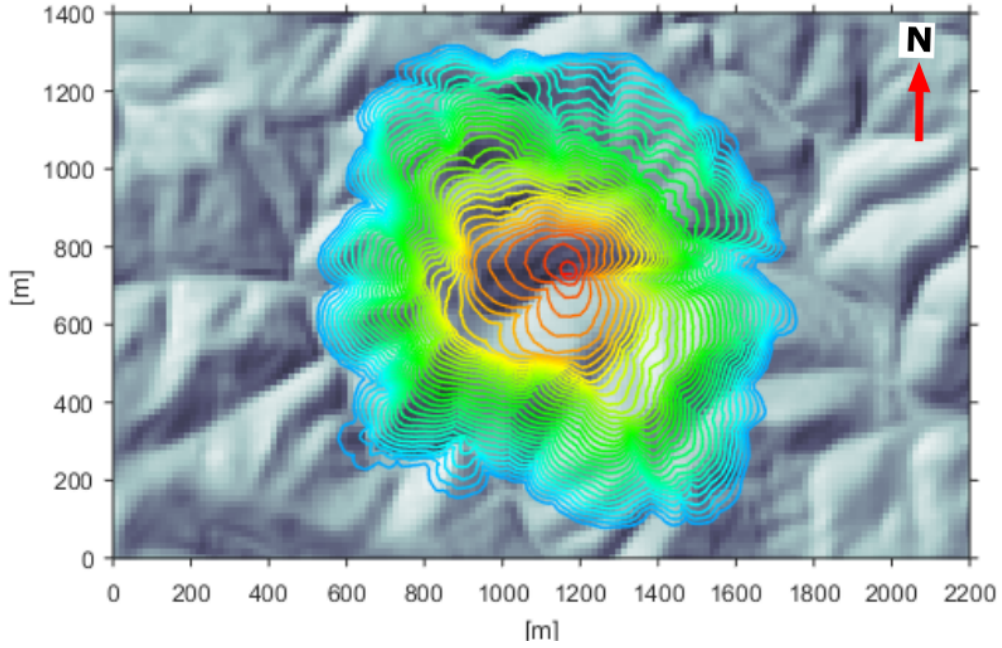
(a) Case A: No wind
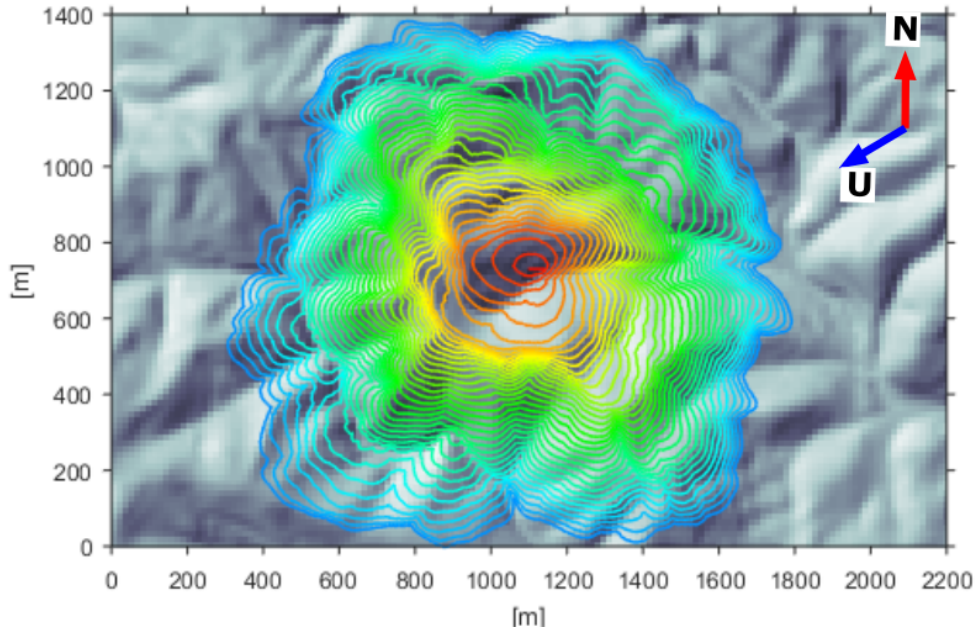
(b) Case B: $U = 5$ m s^{-1}, 30 deg

FIGURE 6.1: Two synthetic isochrones cases used for comparing the optimization algorithm performance. Isochrones span 5 minutes and a total simulation time of 3 h is represented. U is the mid-flame wind speed. Background image is a hillshade representation of terrain model. North and wind direction indicated in the upper-right corner.

to the initializing front. From an optimization point of view, it is expectable that the closer the initial guess is to the true solution the easier for the algorithm to find the

global minima. However, this will be evaluated in the following sections.

TABLE 6.2: Equivalent reduced values (true value) extracted form Table 6.1 and the initialization values used to launch the assimilation for both cases. Note that initial guess is the same for both cases. W is the representative fuel load, SAV the representative surface-area-to-volume ratio, δ the fuel height, M_f, M_x the moisture content of the fuel and the extinction one respectively and U and Dir, the wind speed and direction

		W [kg m^{-2}]	SAV [m^{-1}]	δ [cm]	m$_x$ [%]	M$_f$ [%]	U [m s^{-1}]	Dir [deg]
Ini. Guess	*both*	2	3000	100	40	30	5	270
True value	Case A	0.255	5589	183	20	8	0	-
	Case B	0.255	5589	183	20	8	5	240
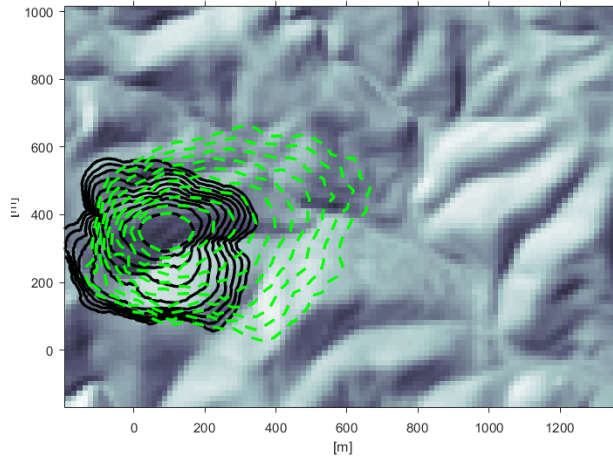
(a) Case A
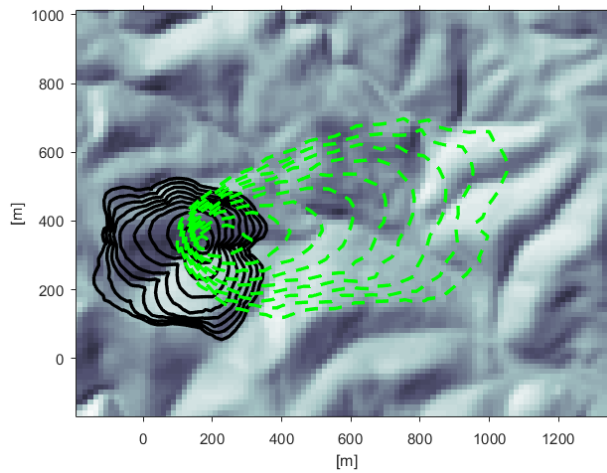
(b) Case B

FIGURE 6.2: Synthetic fronts (black solid lines) and initial guessed (green dashed lines) for Case A (left) and Case B (right).

As mentioned before, the selected optimization algorithms to be compared can be split up into two different categories: line search and global search. Table 6.3 lists the names taken to identify the algorithm explored in the present thesis. A detailed description for each algorithm is provided in the sections 6.2 and 6.3.

To accomplish the final objective of investigating their performances, first, line search algorithms are explored during the assimilation and forecasting step. Secondly, same analysis is conducted for the global search category. Finally, all algorithms are compared by means of performance indicators.

TABLE 6.3: Classified algorithms used for the optimization comparison. Note the name is for labelling purposed only.

Line search	Global search
interior-point	particleswarm
active-set	patternsearch
SQP	globalsearch

6.2   Line search algorithms

We tested three strategies based on line search algorithm. All those strategies start the search from form the stated initial point (i.e. initial guess) and make use of different mathematical techniques to calculate the line search step (length and direction) that leads towards the local minima. Whereas some strategies can compare two local minima to converge to the deeper one, the finding of a global minima is not guaranteed in this kind of algorithms. Their principal characteristics are briefly explained as follows:

- **Interior-point algorithm**

  The interior point algorithm used as first optimizing strategy is based on sequential approximation of the original problem. Mathematically, if the original problem was formulated with the objective function $f(f)$ and the bounds and inequalities $h(x)$ and $g(x)$ respectively. μ is called the barrier parameter (Waltz et al., 2006):

  $$\min_x f(x), \quad \text{subject to } h(x) = 0 \quad \text{and} \quad g(x) \leqslant 0 \tag{6.1}$$

  The sequentially resolved problem can be formulated as:

  $$\min_{x,s} f_\mu(x) = \min_{x,s} f_\mu(x) - \mu \sum \ln(s_i), \quad \text{subject to } h(x) = 0 \quad \text{and} \quad g(x) + s = 0 \tag{6.2}$$

  Where s_i are as many slack variables as inequalities in $g(x)$ and μ should decrease to zero as f_μ approaches the minimum of f.

The added logarithmic term $(\ln(s_i))$ is called a barrier function. The particularity of the strategy is that, in order to solve the approximate problem, the algorithm uses a hybrid strategy. Firstly, it attempts a direct step approach by applying a linearised form of Karush-Kuhn-Tucker (KKT) (Kuhn and Tucker, 1951) conditions. KKT are analogous to the condition of zero gradient value applied in non-constrained optimization but take into account the bounds and inequalities of constrained problems using Lagrangian multipliers. This first step can be also regarded as Newton step. The alternative strategy when the application of first step fails (when the approximate problem is not locally convex, for example) is to apply a conjugate gradient step that numerically computes gradients with centred differences in perpendicular directions using a trust region hypothesis (explained in the following algorithm). This method is fully described in Byrd et al. (2000)

- **Active-set**

  Active-set methods can be applied to both convex and non-convex problems and have been the most widely used methods since 1970s (Nocedal and Wright, 1999). It is based on a trust-region strategy. This technique establishes a sub-domain near the current objective function evaluation point (trust region) where the cost function is approximated using a quadratic model. Then, the constrained problem is solved using a sequence of parametrized unconstrained optimizations, which in the limit converge to the constrained problem. The method can be mathematically formulated with the Kuhn-Tucker equations:

$$\nabla f(x*) = \sum_{i=1}^{m} \lambda_i \cdot \nabla G_i(x*) = 0$$
$$\lambda_i \cdot \nabla G_i(x*) = 0, \quad i = 1, ..., m_e \tag{6.3}$$
$$\lambda_i \geq 0, \quad i = m_e + 1, ..., m$$

The first equation describes a cancelling of the gradients between the objective function and the active constraints at the solution point. For the gradients to be cancelled, Lagrange multipliers $(\lambda_i, i = 1, ..., m)$ are necessary to balance the deviations in magnitude of the objective function and constraint gradients. Because only active constraints are included in this cancelling operation, constraints that are not active must not be included in this operation and so are given Lagrange multipliers equal to 0. This is stated implicitly in the last two Kuhn-Tucker equations (MathWorks, 2016).

Spatial care on the implementation of this algorithm must be taken regarding *Maratos effect*. This effect states that bounds and equalities might be violated to find the fastest way to the minimum. In our present problem, this is important in

two of the bounds. Lower wind speed (i.e. 0 ms^{-1}) and the inequality $m_f \leq m_x$. If any of those constraints is violated, negative and complex ($\mathbb{C}$) RoS are output producing an error in the spread algorithm. To solve this fact (or at least prevent the optimization to halt) the algorithm must be tweaked to handle those scenarios and produce a `NaN` output that can be further interpreted by the optimization method.

- **SQP**

  The algorithm name stands for *sequential quadratic programming* and is described in Nocedal and Wright (1999), chapter 18. This algorithm is of the same family as *active-set* but its implementation provides some advantages: First, the *SQP* algorithm takes every iterative step in within the region constrained by bounds, that is, considering them as hard constraints that are strictly satisfied at each iteration. Such strict feasibility with respect to bounds is also applied when computing the finite differences. This represents an advantage as some problems, as the one at hand, throw an error if bounds are violated, causing the optimization process to terminate.

  Due to its particular implementation, during the iterative process, the *SQP* algorithm can attempt to take a step that fails (i.e. returns a non-real number). This is also particularly suitable to our problem as in some specific conditions the loop clipping algorithm (see section 2.1.1.2) cannot resolve the front entanglement and fails. In this case, the algorithm attempts to take a smaller step and continue. The *SQP* algorithm sets the penalty parameter described above according to the suggestion in Byrd et al. (2000).

6.2.1 Assimilation

The first element to compare those three different line search strategies performance is their capability to generate the fronts that best resemble the assimilated fronts. For each of the two cases previously presented (Case A and Case B), two different assimilation situations are also compared. In the first one, only 5 isochrones are assimilated (i.e. observed) whereas in the second, the number of assimilated fronts is increased to 10. Given that isochrones are homogeneously spaced 5 min each, the assimilation process is conducted during 20 min and 45 min respectively. In both cases, the first assimilated isochrone is taken to be 45 min after the ignition time. This delayed time is considered to be representative of the time needed for emergency responders to deploy the data acquisition system in a hypothetical real operation. Additionally, this time permits the initial line ignition source to develop as a more realistic fire front influenced by local wind and topography. Composite plots including the three optimization strategies applied to

Case A and Case B observation (i.e. six panels) are generated for 5 front assimilation situation (figure 6.3) and 10 fronts one (figure 6.4). In all cases, the observed fronts (black solid lines), the initial guess fronts (green dashed lines) and the fronts after assimilation (red dashed lines) are depicted together with the hillshade of the scenario as a background to illustrate the topographic effects.

The *SQP* strategy fails to correctly reproduce the observed fronts regardless of the case and the assimilation situation. This strategy is not only failing to find a global minima but also seems to worsen the initial guess solution as fronts expand even more towards the initial wind direction instead of decreasing towards observations. The two other strategies (*interior-point* and *active-set*) seem to better mimic the observations. The *interior-point*, however, fails to properly resolve the assimilation of 5 fronts in Case A and the 10 fronts assimilation for Case B. In general, the results suggest that Case B is better assimilated than Case A. This consideration will be further explored in the following section 6.4, when similarity index will be used to extract quantitative comparisons.

(a) Case A: interior-point

(b) Case B: interior-point

(c) Case A: active-set

(d) Case B: active-set

(e) Case A: SQP

(f) Case B: SQP

FIGURE 6.3: Line search optimization strategies performance when 5 fronts are assimilated for Case A (left) and Case B (right). Black solid lines are observations (i.e. FARSITE simulations), green dashed line are initial guess launch and red dashed line represents best run after assimilation.

(a) Case A: interior-point

(b) Case B: interior-point

(c) Case A: active-set

(d) Case B: active-set
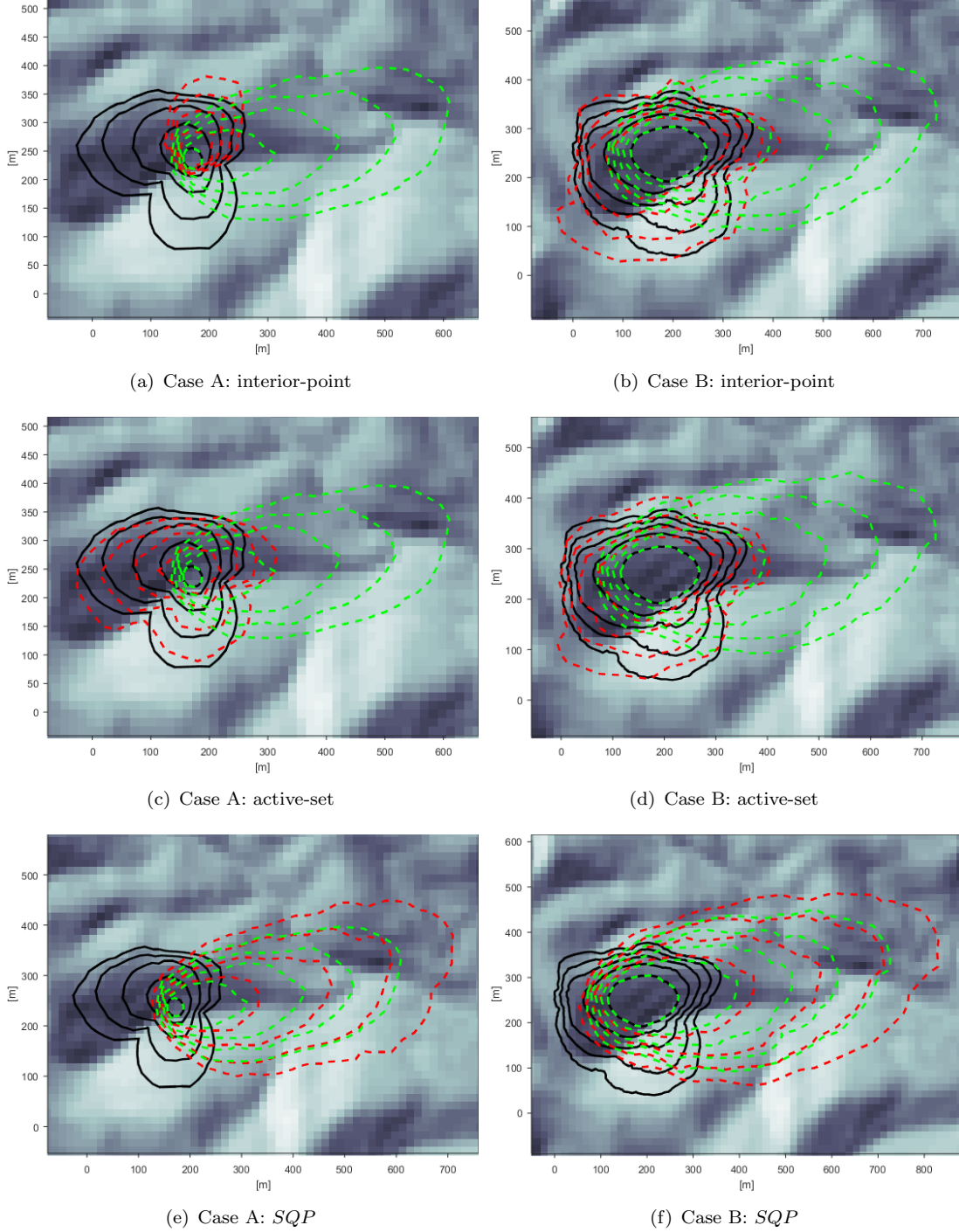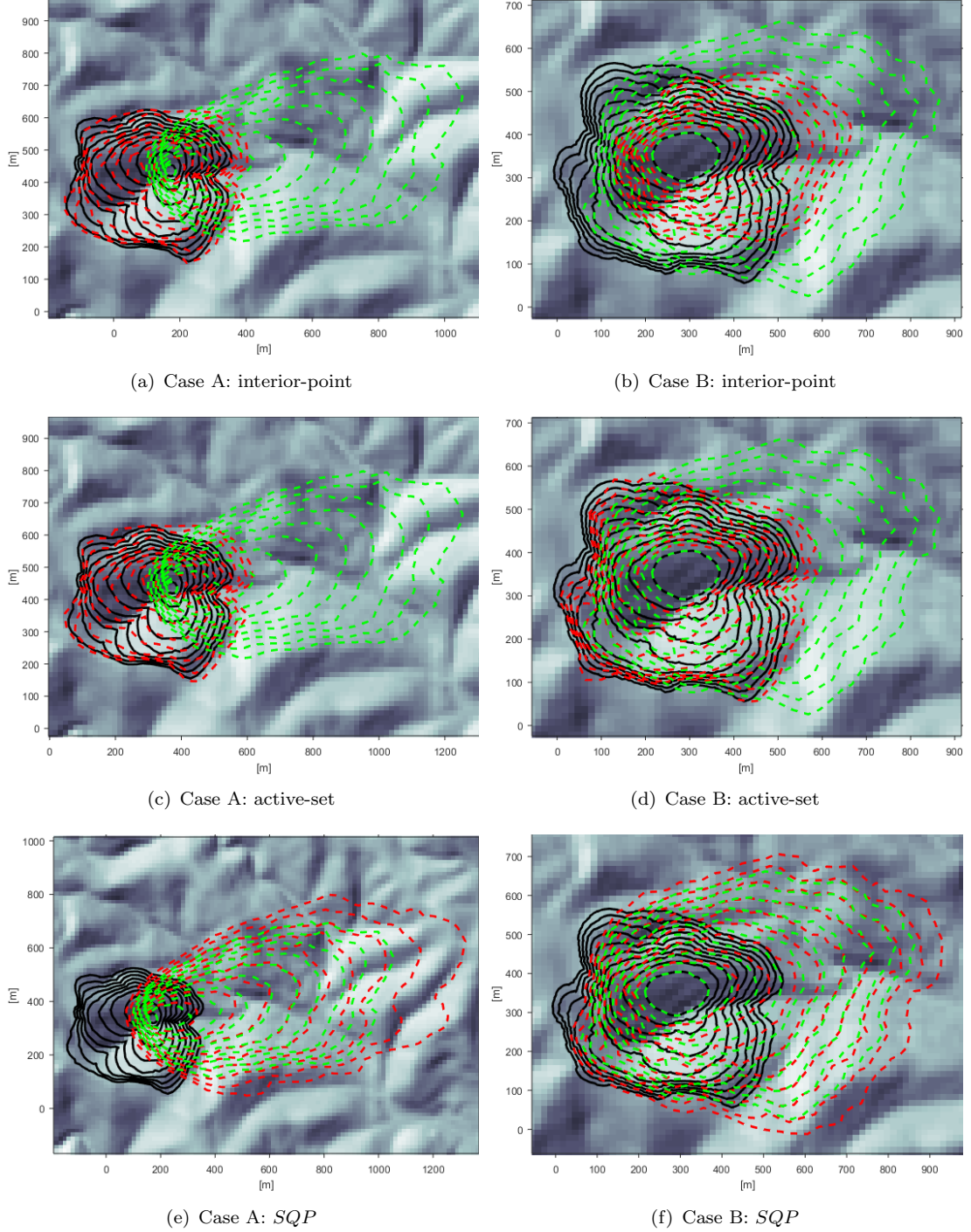
(e) Case A: SQP

(f) Case B: SQP

FIGURE 6.4: Line search optimization strategies performance when 10 fronts are assimilated for Case A (left) and Case B (right). Black solid lines are observations (i.e. FARSITE simulations), green dashed line are initial guess launch and red dashed line represents best run after assimilation.

6.2.2   Invariants' convergence

Besides properly representing the observed front, a main question to pose to the system is its capacity to resolve the individual convergence of each invariant. That is, if the last optimized value converges to the theoretical value used to run the FARSITE simulations. Such *true* invariants value are calculated by reducing the complete set of the variables input to FARSITE to the seven reduced ones used by SmartQFIRE (as previously explained in section 6.1). Despite this theoretical value being the most-accurately reduced model found (see section 3.7), the conversion is susceptible to errors as discussed in section 3.7.2. Bearing this in mind, the best assimilated values for each optimization strategy are plotted together with the initial guess and the true value in Figure 6.5 (for Case A) and 6.6 (for Case B). Along with comparing the performance of each optimization strategy, the 5 fronts and 10 fronts assimilation situations are depicted side-by-side to highlight difference and similarities.

Figure 6.5, Case A, shows the complexity of correctly resolving all parameters at same time. All algorithms fail to resolve all the parameters at these same time. The interior-point is the one that best performs the assimilation stage. This strategy, is the only one capable of correctly identifying the representative fuel load (W) and the fuel depth (δ). However, this happens only when 10 fronts are assimilated. For the 5 fronts case, those two parameters are not properly resolved as the final convergence is still far from the true value. The surface-area-to-volume ratio (SAV) is the less correctly resolved parameter. All three strategies converge it to the similar incorrect point. The 5 and 10 fronts scenarios do not cause any variation on this regard. The possible explanation for this result is the presence of large relative minima that none of the strategies manage to avoid. This conclusion is further sustained by the results of the global search algorithms (see section 6.3.1). Unexpectedly, the moisture parameters (M and M_x) are better resolved for the 5 fronts cases, where the SQP and the interior-point method makes them converge close to the true value. Those two strategies also resolve the wind speed (set to 0 m s^{-1} but initialized at 5 m s^{-1}) except for the case with 10 fronts, where SQP converge at 15 m s^{-1}, far from the true value. This result could be justified by the fact that SQP is also wrongly resolving the representative fuel load, and those are two parameters that can effect the RoS value at same time. The active-set strategy does not manage to properly resolve the assimilation as it converges all the parameters close to the their initialization value, except for the SAV and the wind direction. The latter, however, is not relevant as the wind speed is zero, and thus its direction is meaningless and has been kept into the Figure for consistency with the Case B.

FIGURE 6.5: Invariants' convergence value for Case A (no wind) and each strategy (coloured bars) grouped by 5 and 10 fronts (5F and 10F, respectively) assimilation situations. True theoretical values (black line) and initial guess (green line) are also plotted. Note that true value for wind direction does not exist as there is no wind speed.

For the Case B, (synthetic data generated with $5\,\mathrm{m\,s^{-1}}$), the active-set strategy shows the same low performance as the previous case, except for the wind speed, that resolves it perfectly, outperforming the other strategies. The SQP also shows similar performance as in the previous case, except for the 5 fire fronts scenario, where the moisture related parameters, the representative fuel load and the fuel depth final convergence move away from the true value. Similar effect is found for the interior-point strategy and the 5 fronts scenario. However, this strategy resolves better the wind speed, particularly, for the 5 fronts scenario. Regarding the wind direction, all strategies fail to get closer to the 90 deg (wind originating direction) and keeps close the initial guess at 270 deg. No reason is identified to explain the difficulty to resolve this parameter in comparison with the other ones.
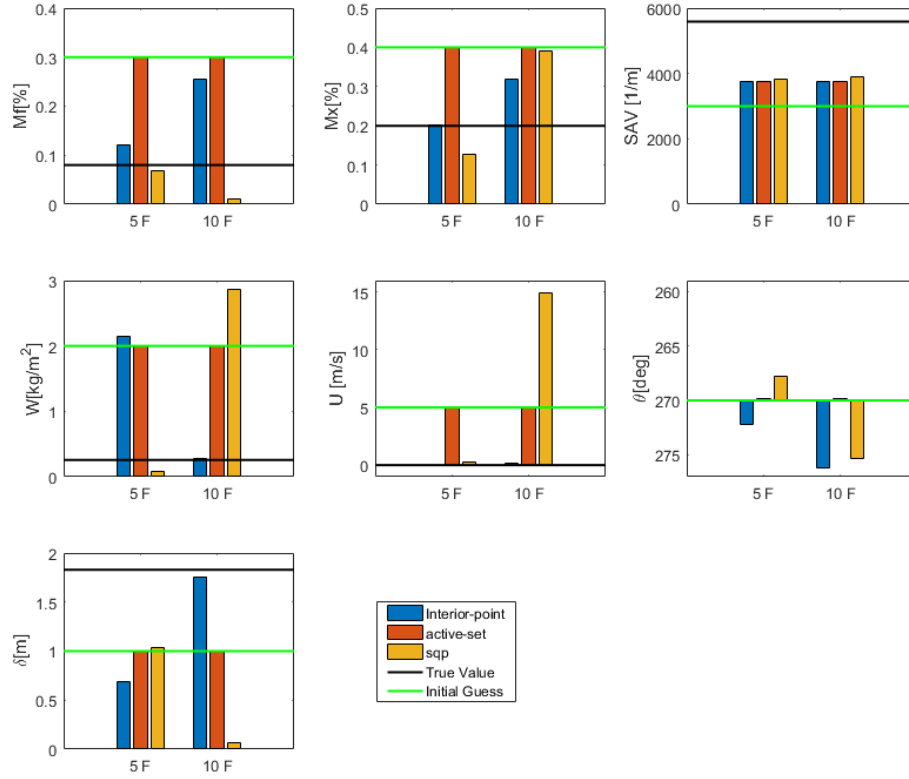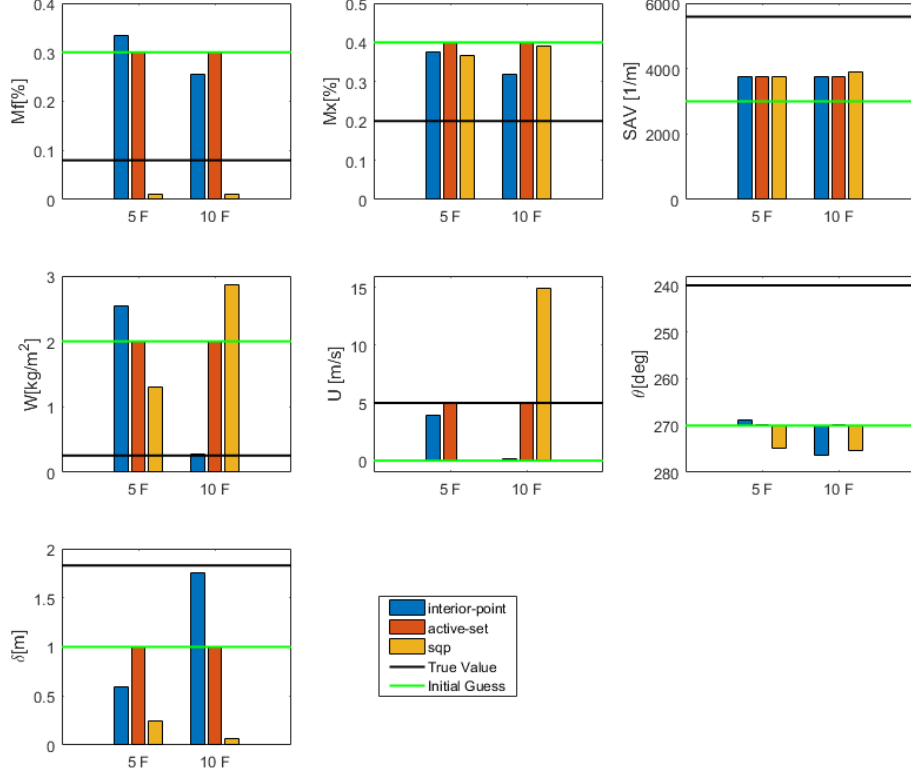
FIGURE 6.6: Invariants' convergence value for Case B (with wind) and each strategy (coloured bars) grouped by 5 and 10 fronts (5F and 10F, respectively) assimilation situations. True theoretical values (black line) and initial guess (green line) are also plotted.

6.2.3 Forecast

After the assimilation process is performed, the updated invariants value can be used to launch a forecast starting from the last assimilated front. For the easy identification and further comparison of each scenario, we chose to forecast the same number of fronts as assimilated. Obviously, this does not come from any limitation of the system, and the number of assimilated fronts and the available fronts could differ. The only restriction is that the sum up of assimilated fronts and forecast fronts has to equal the total number of available isochrones.

Figure 6.7 displays the Case A and B forecast for the 5 fronts assimilation scenario. The interior-point strategy when applied to Case B is the one that shows better forecasting performance. It is important to remark that both active-set cases also show an acceptable forecast capacity despite that the parameters where not properly identified (as mentioned in the section above). This supports the hypothesis that certain wrong

set up of parameters can lead to similar fronts solutions. This multiplicity renders the optimization challenging for line search methods as they converge to the local minima created by those sets.

(a) Case A: interior-point

(b) Case B: interior-point

(c) Case A: active-set

(d) Case B: active-set

(e) Case A: *SQP*

(f) Case B: *SQP*

FIGURE 6.7: 50 min forecast of 10 fronts spaced 5 min for Case A and B using line search methods. Black solid lines are the synthetic fronts used to check the forecast and the blue dashed lines are the forecast fronts after the assimilation process.

The forecast results when 10 fronts are assimilated is presented in Figure 6.8. Similar conclusions to the 5 fronts scenario are withdrawn. The active-set strategy leads to the best forecast fronts. The main explanation is the correct identification of the wind speed, a key parameter that has a significant influence if 50 min are to be forecast Additionally, due to same reason SQP strategy fails to get closer to the observed fronts as the wind speed was overestimated (to 15 m s^{-1}). Finally, the interior point works correctly for the Case A (all parameters where closely resolved) while for the Case B the fronts are over-expanded towards the east.

(a) Case A: interior-point

(b) Case B: interior-point

(c) Case A: active-set

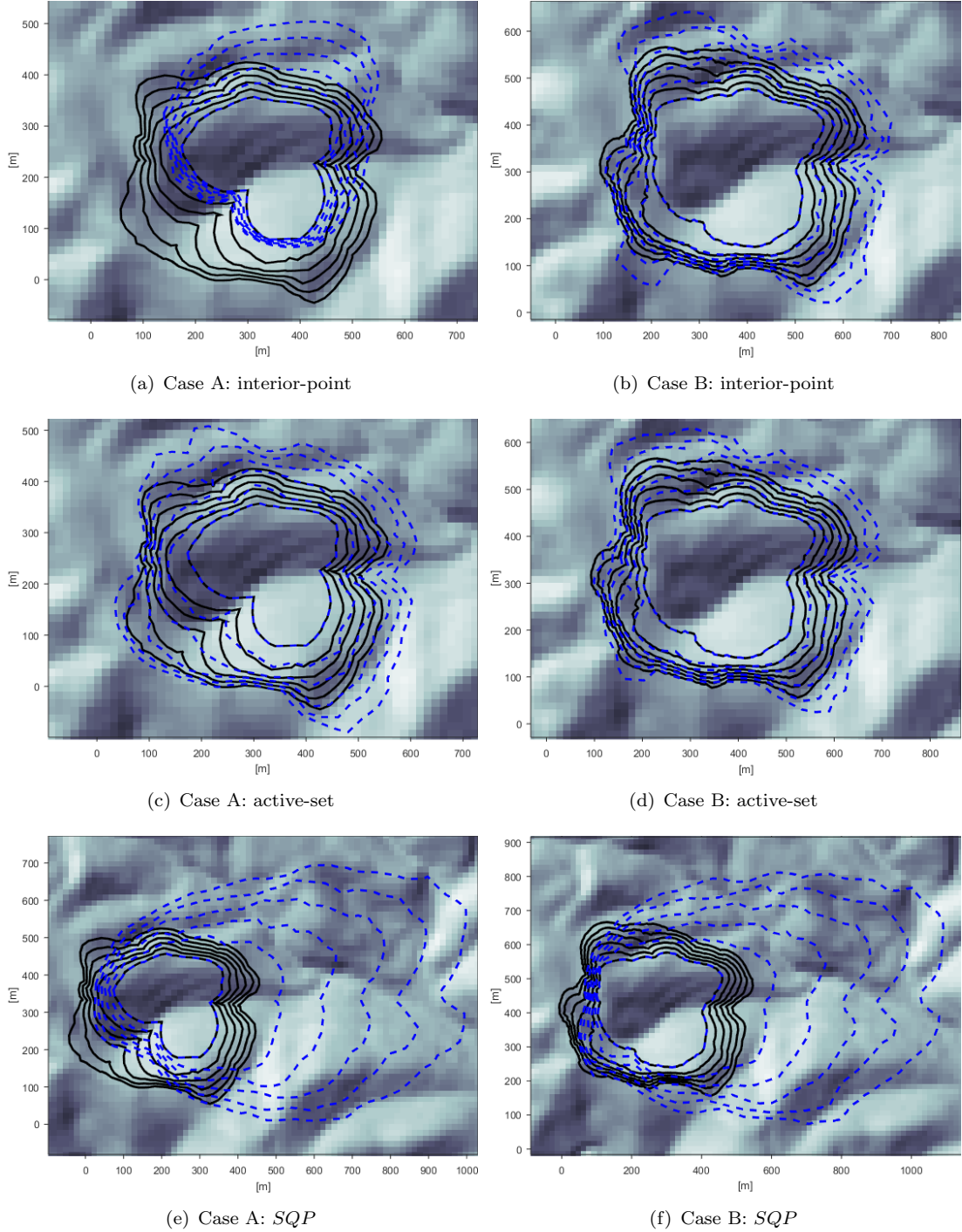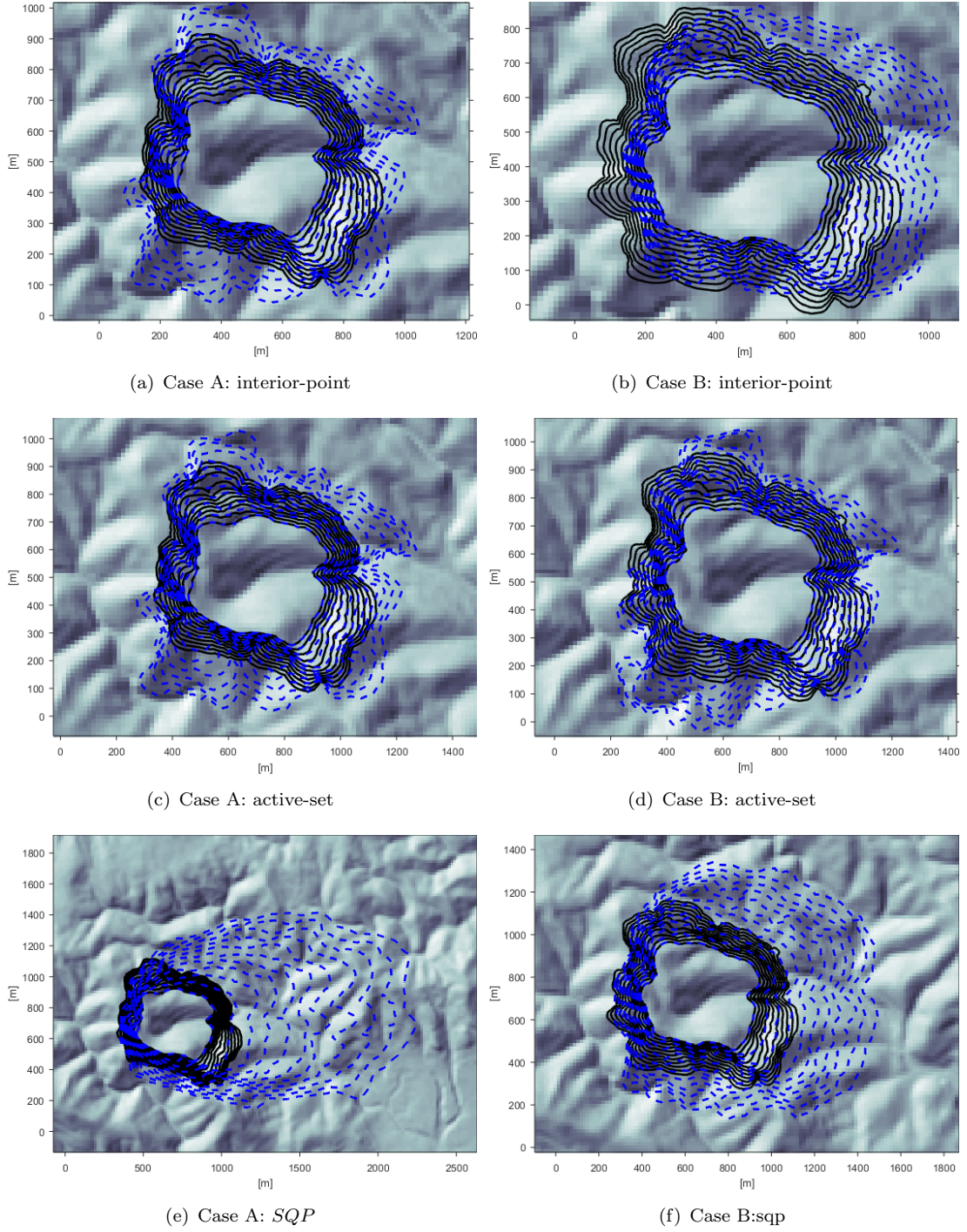(d) Case B: active-set

(e) Case A: SQP

(f) Case B:sqp

FIGURE 6.8: 50 min forecast of 10 fronts spaced 5 min for Case A and B using global search methods. Black solid lines are the synthetic fronts used to check the forecast and the blue dashed lines are the forecast fronts after the assimilation process.

6.3   Global search strategies

Global search strategies are generally based on multiple initiation points. Instead of starting off from an initial point (initial guess in the line search algorithms) multiple individuals are generated to look for the global minima. The interaction between the different individuals and the governing rules that guide them toward the minima are the main differences between the algorithm within this family. In the following, the three algorithms used for the comparison at hand are briefly introduced.

- **Particleswarm**

  Particle swarm is a population-based algorithm, similar to the genetic evolutionary algorithms (Wendt et al., 2011). At the initialization stage, a swarm of particle is generated and scattered over the optimization domain. Each of those initial particles is given a different speed to perform one iteration or step. The evaluation of the cost function at their new location provides an update to their individual speed for the following iteration. As a result, a collection of individuals move in steps throughout the optimization domain.

  The algorithm is inspired with a flock of birds or a swarm of bees. The innovative feature is that each particle is attracted to some degree to the best location it has found so far, and also to the best location any member of the swarm has found. After some steps, the population can coalesce around one location, or can coalesce around a few locations, or can continue to move (Rios and Sahinidis, 2013).

- **Patternsearch**

  Pattern search is a direct search algorithm that does not require gradients to be computed (in contrast to most more traditional optimization methods) and works with non-differentiable cost functions. The strategy is to conduct a set of cost functions evaluations around the current point, looking for the lower value and directing the next step towards there. Although it is not properly a global search algorithm, the generalized pattern search implementation, that creates a pool of point cardinally distributed to be evaluated at each iteration, is considered to lead to a global solution (MathWorks, 2016).

  Direct search strategies are particularly suitable for problems where the objective function is not differentiable, or is not even continuous, being those two properties a characteristic of the problem at hand.

  A complete description of the strategy together with convergence evaluation can be found in Kolda et al. (2003).

- **Globalsearch**

  This strategy uses the interior-point algorithm presented in section 6.2 but instead of applying line search strategy at the initial guess point, a multi-start approach is used to generate a scatted initial points following the Ugray et al. (2007) algorithm designed to find global minima.

6.3.1 Assimilation

As for the line search strategies, the first evaluation is conducted by observing the assimilated fronts. Figure 6.9 shows the 5 assimilating fronts scenario. The first remarkable difference with the line search strategies is that global search get closer to the observables. All three strategies manage to correct the initial guess and find the parameters combination that resembles the assimilated fronts. The Case A is similarly resolved and no strategy over-performs the others. On the contrary, for Case B, particleswarm and patternsearch underestimates the north-west flank whereas globalsearch overestimates it.

The results when 10 fronts are assimilated are gathered in Figure 6.10. As for the 5 front scenario, all cases are extremely well resolved. The particleswarm strategy might be the one best resolving the both cases at same time, although differences are limited. Already from this first assimilation evaluation it is clear that global search strategies over-perform line search ones in all scenarios and for both studied cases.

(a) Case A: particleswarm

(b) Case B: particleswarm

(c) Case A: patternsearch

(d) Case B: patternsearch

(e) Case A: globalsearch

(f) Case B: globalsearch

FIGURE 6.9: Global search optimization strategies performance when 5 fronts are assimilated for Case A (left) and Case B (right). Black solid lines are observations (i.e. FARSITE simulations), green dashed line are initial guess launch and red dashed line represents best run after assimilation

(a) Case A: particleswarm

(b) Case B: particleswarm

(c) Case A: patternsearch

(d) Case B: patternsearch

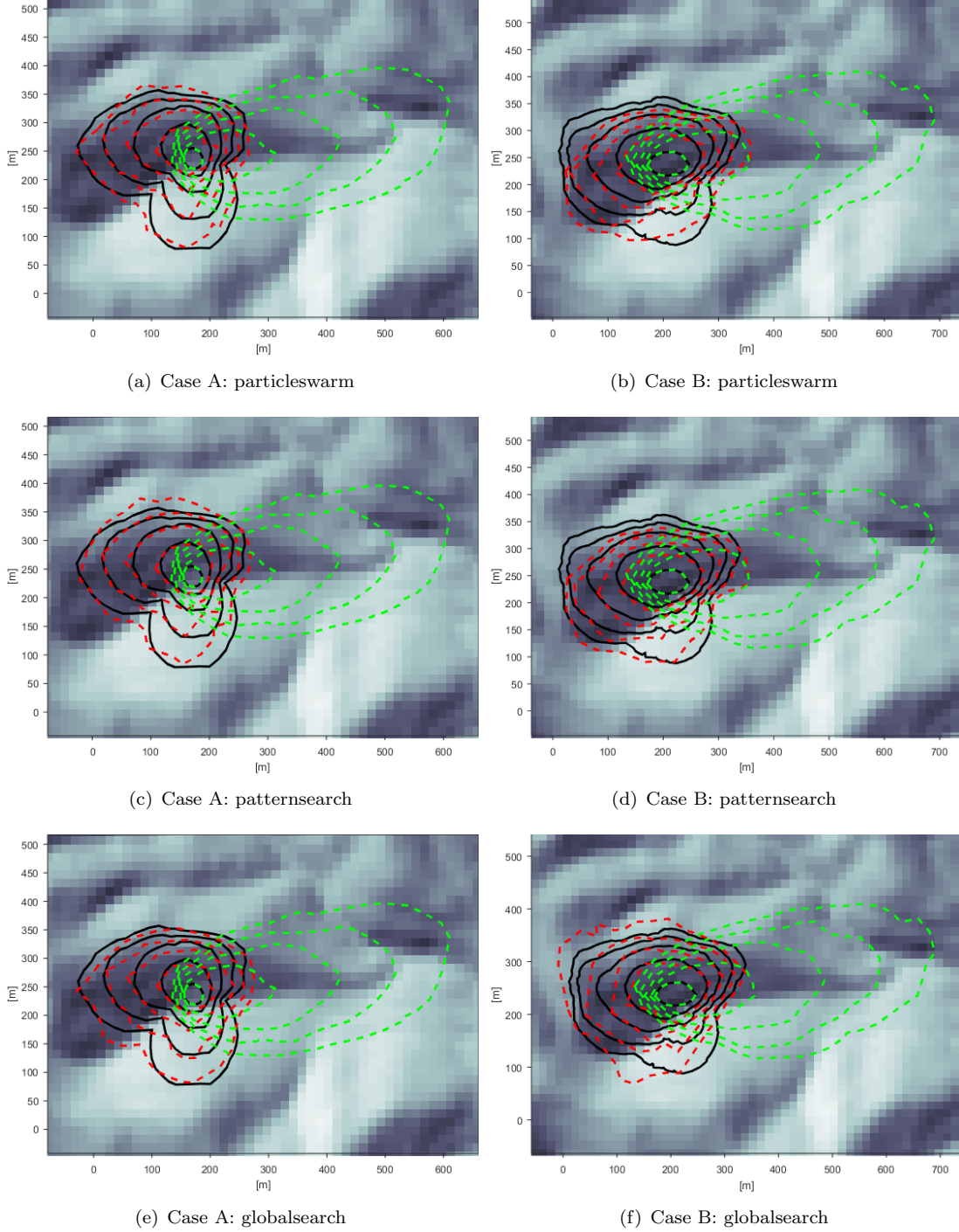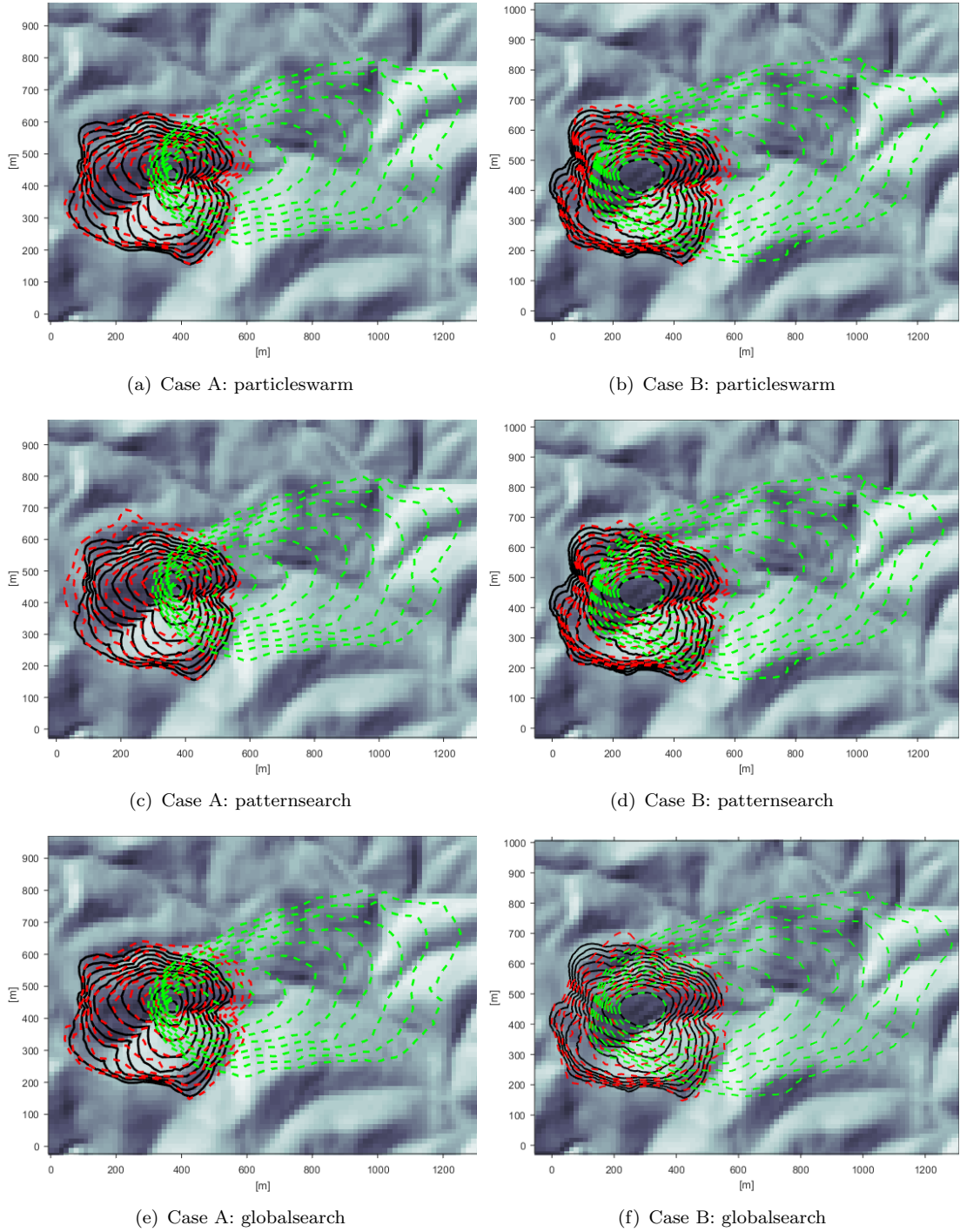(e) Case A: globalsearch

(f) Case B: globalsearch

FIGURE 6.10: Line search optimization strategies performance when 10 fronts are assimilated for Case A (left) and Case B (right). Black solid lines are observations (i.e. FARSITE simulations), green dashed line are initial guess launch and red dashed line represents best run after assimilation

6.3.2   Invariants' convergence

Observing the individual parameter convergence we see that, although the range of the final converged values is lower than the line search methods case, the three optimization strategies do not converge to the same end state (i.e. identical parameter set). This is a surprising result because global search algorithm should identify the absolute minima of the problem. Although this is not guaranteed, the fact that for the same problem they reach different end states indicates the optimization complexity of the problem itself.

Looking at Figure 6.11 (convergence results for Case A) we can identify one important difference from line search results; the SAV parameter varies and gets closer to the true value (except for the patternsearch). This might be one of the reasons of the overall performance improvement of these strategies with respect to the front shape. On the contrary, the representative fuel load, the fuel moisture and the fuel depth converge far from the true value. The moister of extinction is better identified, except for the patternsearch strategy that again, converges far from the correct value. No considerable difference is identified between the assimilation scenarios.
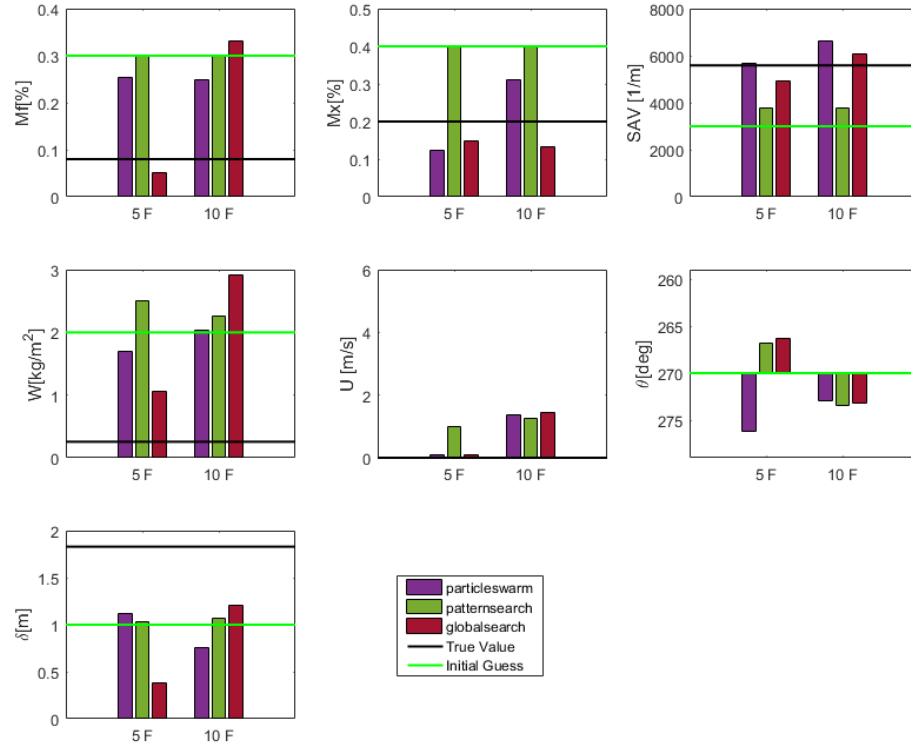
FIGURE 6.11: Invariants' convergence value for Case A (no wind) and each global optimization strategy (coloured bars) grouped by 5 (5F) and 10 (10F) fronts assimilation situations. True theoretical values (black line) and initial guess (green line) are also plotted

For the Case B (6.12), the wind direction is now well resolved as all strategies converge close to the 90 deg (the true value). On the contrary, the wind speed is only partially resolved by the patternsearch strategy (and the 5 fronts scenario) that converges it towards 3.5 ms^{-1}(being 5 m s^{-1}the true value.). The other strategies give it a value around zero. Regarding the fuel moisture, the fuel load and the fuel depth parameters, they also converge far from their true value. Although there is some improvement with respect to Case A, particularly the particleswarm strategy and the globalsearch (for the 10 fronts scenario on both cases), the added complexity of wind do not alter the performance of the optimization, neither improves it nor diminishes it.
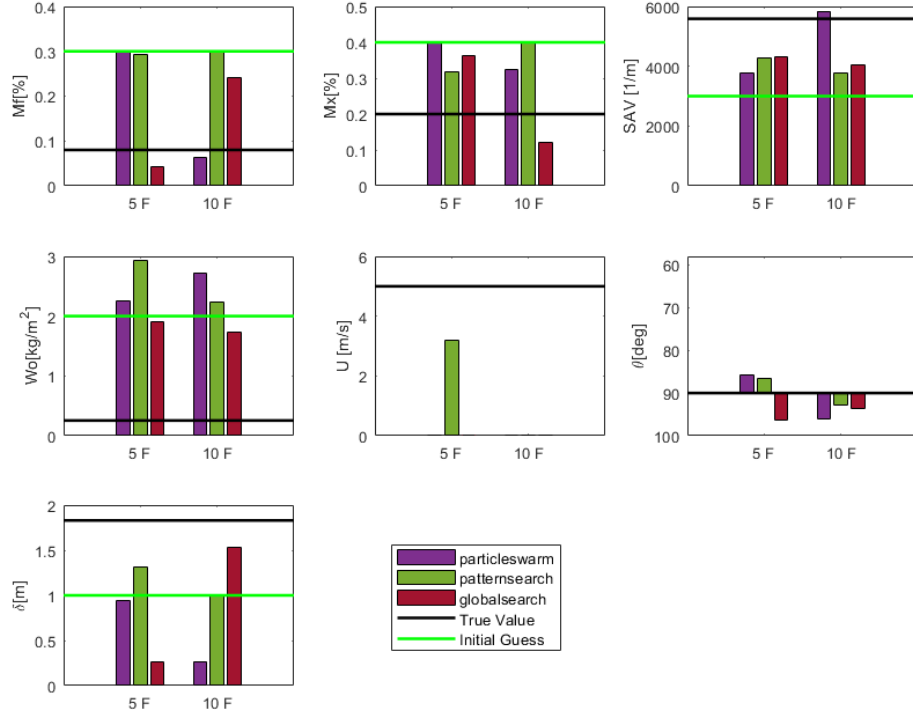
FIGURE 6.12: Invariants' convergence value for Case B (with wind) and each global optimization strategy (coloured bars) grouped by 5 (5F) and 10 (10F) fronts assimilation situations. True theoretical values (black line) and initial guess (green line) are also plotted

6.3.3   Forecast

Similarly as presented in section 6.2.3, the forecasting results for the 5 and 10 assimilation scenarios are presented in Figures 6.13 and 6.14 respectively.

The forecast for the 5 assimilated fronts scenario shows better results for the Case B (with wind) than Case A, where all strategies tend to over-predict the north-west flank. On the contrary, all three strategies perform correct forecast for the Case B a none of them can be identified for over-performing the others. In particular, it is remarkable that the thalweg appearing on the east flank, due to the fire spread-terrain interaction, is correctly forecast in all cases.

In the 10 fronts scenario (Figure 6.14), the Case A over-predictions identified before are no longer observed for particleswarm and globalsearch. Both strategies correctly mimics the north-west flank fronts. The over-prediction is now observed on the south flank. Regardless of the set of converged parameters, all strategies fail to properly

forecast the development of this flank and it is always over-predicted. For Case B, the rest of the flanks are correctly forecast.

(a) Case A: particleswarm

(b) Case B: particleswarm

(c) Case A: patternsearch

(d) Case B: patternsearch

(e) Case A: globalsearch

(f) Case B: globalsearch
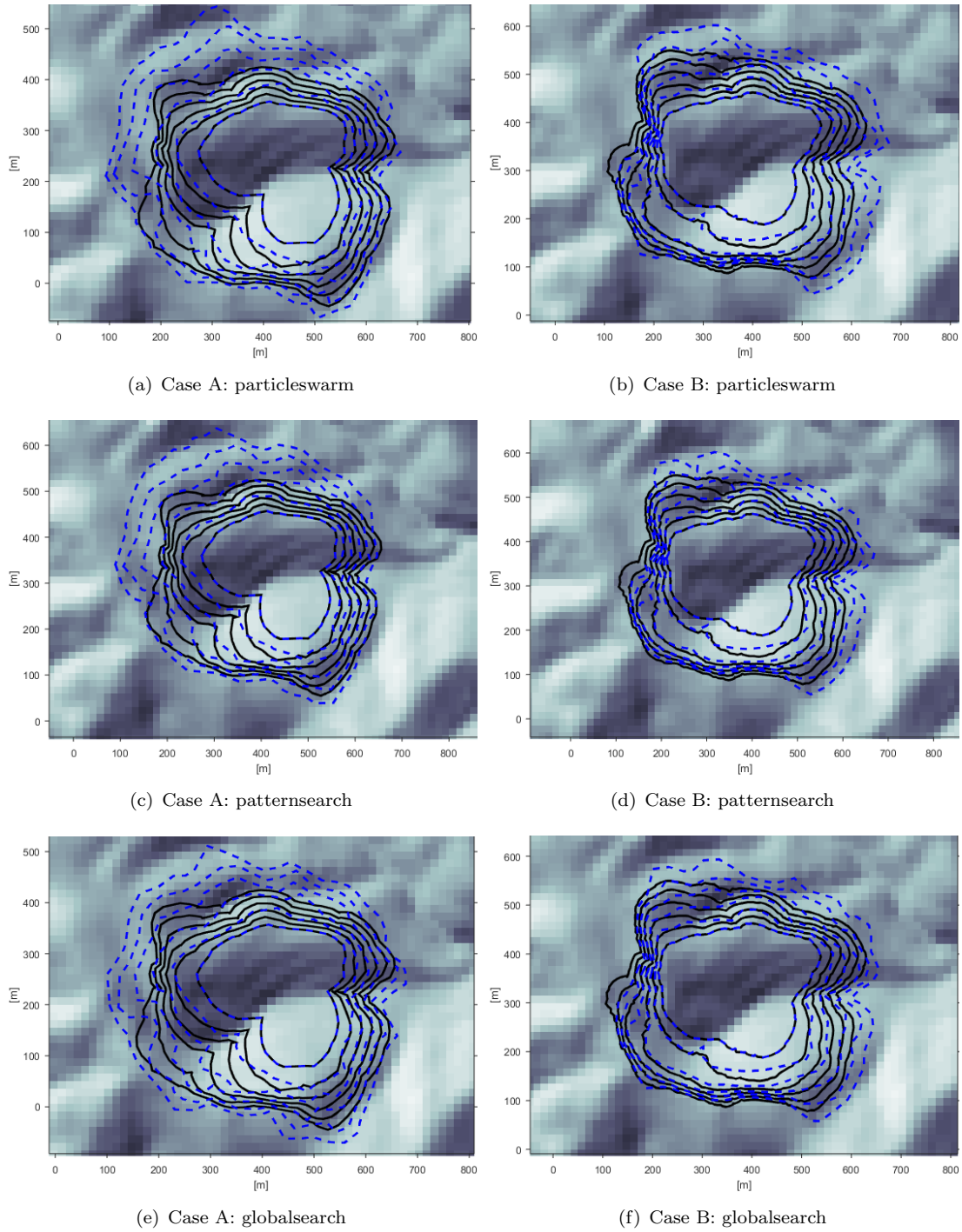
FIGURE 6.13: Forecast of 05 fronts (25 min) for Case A (left) and Case B (right). For legend information see caption of Figure 6.3.

(a) Case A: particleswarm

(b) Case B: particleswarm

(c) Case A: patternsearch

(d) Case B: patternsearch

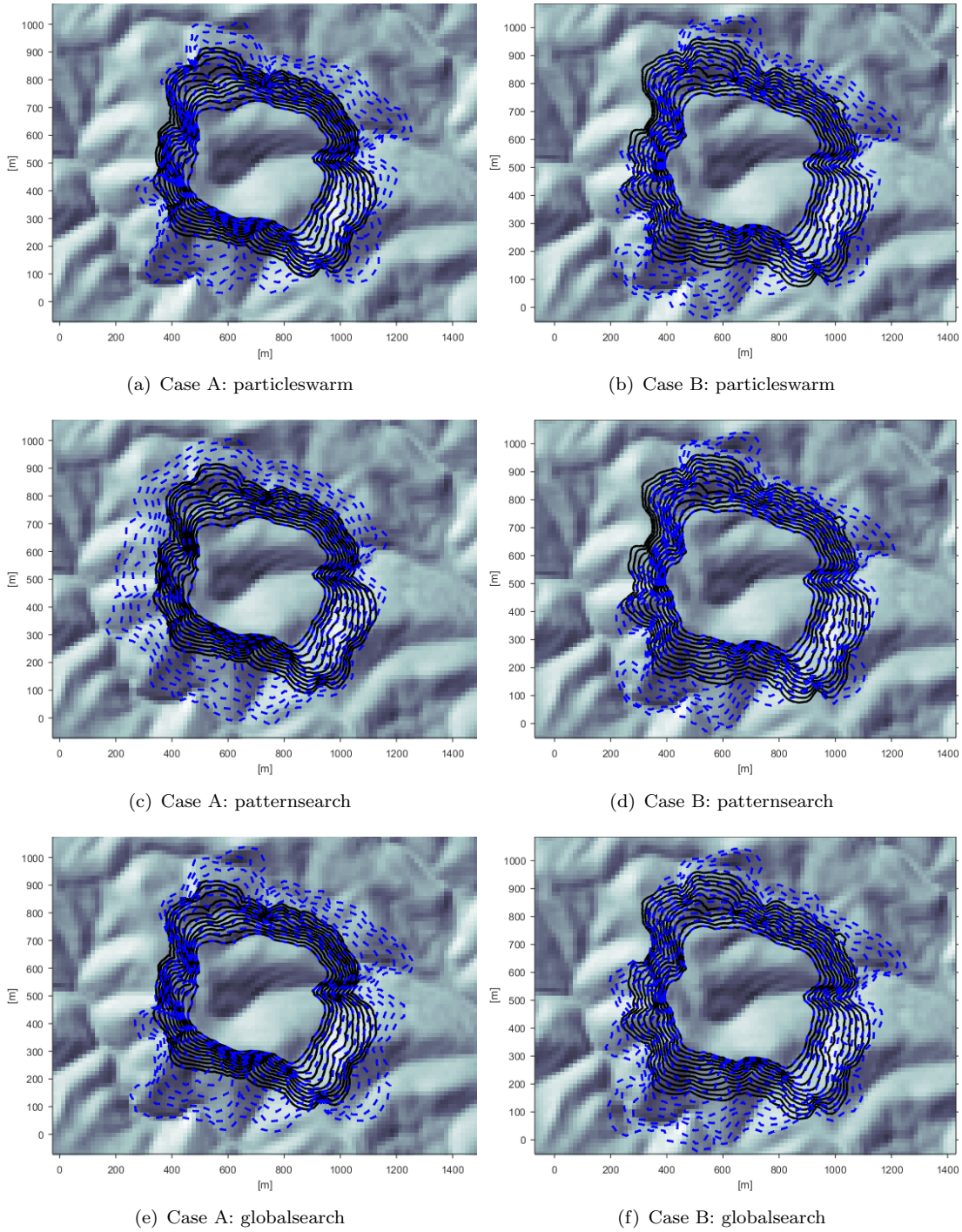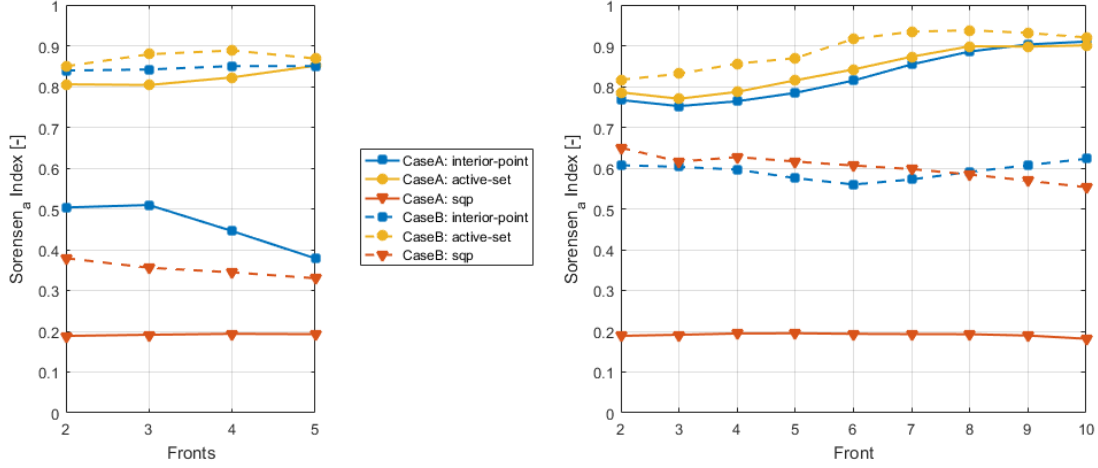(e) Case A: globalsearch

(f) Case B: globalsearch

FIGURE 6.14: Global search 10 fronts (50 min) forecast. Black lines are synthetic observations whereas blue dashed lines are improved forecast.

6.4   Global search versus line search

A quantitative way to account for the discrepancy between observed and simulated fronts is to compute a similarity index for each pair of observed and simulated front after the assimilation step. The modified version of Sorensen index (that discards the initial area to give more emphasis on actual simulated front, as described in sec. 5.4) is taken for this purpose. This index is chosen as it is the only one (together with Jaccard) that is bounded between [0-1]. Being unity total agreement, and zero complete mismatching. Displaying this index can help to compare the assimilating performance of line search versus global search strategies.

Figure 6.15(a) displays all linear search cases (dashed and solid lines) and scenarios (left and right panel). Figure 6.15(b) shows same results and distribution for the global search strategies. Looking at both panels, it is evident to conclude that in general terms global optimization performs better assimilations that the line-search approach. Only the active-set strategy and some interior-point cases are comparable to the global optimization performance. Another conclusion to be withdrawn is that, as time passes by, the simulated fronts similarity to observation is kept constant or even improve. This can be seen as all curves in Figure 6.15 are horizontal or have positive slope. This might be seen as an unexpected result but following the logics of the assimilation system at hand, as more fronts are assimilated the selected parameter become more representative of the overall behaviour.

As a partial conclusion, it is remarkable the fact that, in general terms, the parameters convergence diverge from the true value whereas the fronts do reassembles observations. There are two potential elements that contribute to this fact. First, the considered true value at hand is the one for the FARSITE simulations, thus the correspondence to the spread model is not perfect. The spreading system is not identically implemented (despite based on same principles) and the fuel is not equally described as the reduction fuel model is used for transforming it into input data for the assimilation process. Secondly, it has been already pointed out in previous chapters the multiplicity of existing set of parameters that produce similar (if not exact) solutions.

(a) Line search strategies
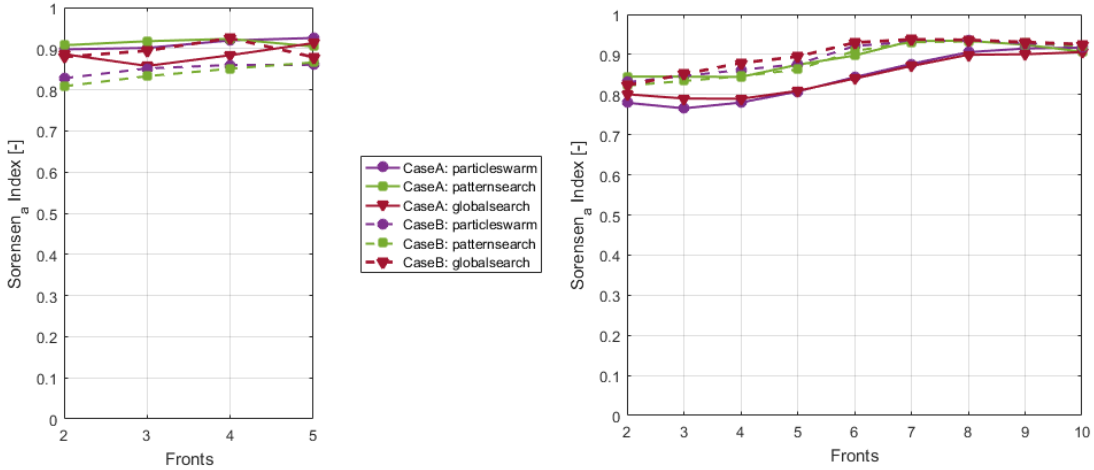
(b) Global search strategies

FIGURE 6.15: Similarity index comparison within the line search strategies during 5 fronts optimization (left panel) and 10 fronts optimization (right panel). Isochrone steps are 5 min thereby assimilation times are 20 min and 45 min respectively (first assimilated front is used for model initialization. Notice that horizontal axis starts at 2 front as the last assimilated one is used to initialize the forecast)

6.5   Running time performance

The last indispensable comparison is to explore the computing time required to run each strategy. This is a crucial aspect if SmartQFIRE is to be used operationally. To examine computing elapsed time it is necessary to look into the parallelization performance of the optimization. Multi core machines are easy accessible as a result of ever-lowering cost of hardware together with the expansion and availability of remote computing, such as

cloud-computing solutions. Thus, although emergency services might not be currently equipped with such supercomputers, they could be easily accessed if required.

To explore such parallel implementation, the assimilation cases presented in the previous sections are repetitively launched within a size-changing pool of workers. At the end of each assimilation stage the wall-clock elapsed time is recorded and plotted. The code parallelization is attained with the `Matlab Parallel Toolbox` (The MathWorks, 2016). A dedicated 44 cores (Intel (r) Xeon E5-2699 v4) is employed to perform all simulations.

The results for the line search optimization strategies are gathered in Figure 6.16. The effect of adding workers to speed up the assimilation step is unfruitful regardless of the assimilation scenario (5 or 10 fronts, sub-plots 6.16(a) and 6.16(b)) and the case (sub-plots 6.16(c) and 6.16(d)). This could be explained as the step-wise approach followed by the line-search strategies is eminently a serial process, that can be only limitedly parallelized. Thus, no speed up can be achieved by means of parallel computing with these methods.

(a) Case A: 5 fronts (20 min)

(b) Case A: 10 fronts (45 min)

(c) Case B: 5 fronts (20 min)
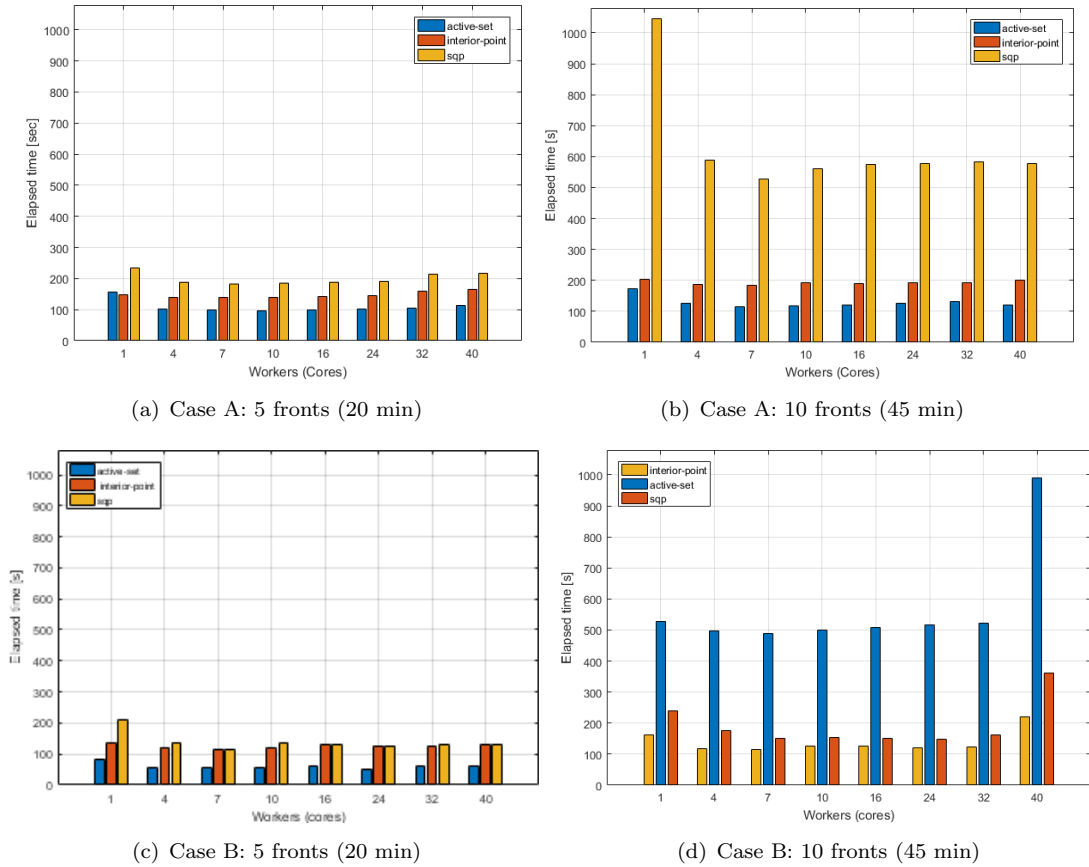
(d) Case B: 10 fronts (45 min)

FIGURE 6.16: Line search optimization strategies. Assimilation process elapsed time for Case A and Case B and both assimilation scenarios (5 and 10 fronts) when parallelized with different number of workers.

On the opposite, global optimization strategies might be more parallelizable as they are based on multiple search steps that can be executed in parallel. For the benchmarking exercise at hand, only Case B with the 10 fronts scenario is analysed, as we know from the previous sections that it is the worst case in terms of optimization performance (larger elapsed time and higher number of fronts to optimize). Figure 6.17 reports the wall-clock elapsed time for different number of workers (from 1 to 40). Only the particleswarm strategy substantially benefits from the parallelization. It shows an optimal point around 7 workers where the computing time reduces to one fifth of the original computing time (20 min with respect to the initial 100 min single-worker time). After this, the addition of workers do not further speed up the process. This could be justified for the computing overhead and the input-output processes (reading- writing). Globalsearch strategy shows a poor parallelization capacity as the computing time when multiple workers are available doubles the original single-worker time. Finally, the patternsearch strategy does no see any significant advantage on increasing the number of workers.
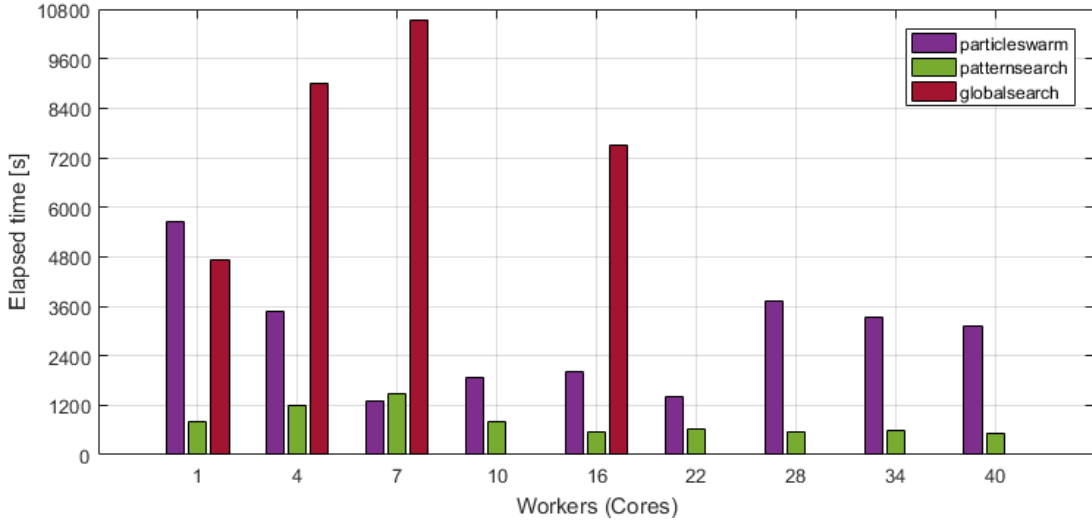

FIGURE 6.17: Global optimization strategies. Assimilation process elapsed time for Case B and 10 fronts scenario when parallelized with different number of workers. Note that he vertical axis scale is larger than previous case.

It is worth mentioning that the parallelization performance of the strategies studied in this section is highly dependant on the native code implementation (i.e. `Matlab Parallel Toolbox`) and it is out of the scope of this thesis to evaluate the efficiency of such implementation.

6.6　Conclusion

We have deeply explored a set of six different optimization strategies, three of which could be classified as a line search and three as global search strategies. To perform the

comparison we used SmartQFIRE algorithm (see chapter 5) including the interpolating system developed in this thesis (see chapter 4) to create fast terrain-based wind maps (by means of WindNinja interpolations). This is the first time a global optimization is launched together with dedicated topographic wind simulations and successfully reaches operation time constraints.

To compare the performance of the different strategies we have analysed the assimilation capacity to produce fronts that match observations, the values of the converged invariants and the forecast capability. The analysis is conducted using synthetic data generated with FARSITE as an observable data. Two different cases (with and without wind) and two different assimilation scenarios (5 (15 min) and 10 (45 min) fronts). Additionally, the computing time required for each strategy is compared using different configurations cores workers form a 44 core cluster to explore the parallelization performance of each strategy.

As a general conclusion, the active-set strategy is considered to be the most convenient and over-performing one, both, in terms of efficiency and accuracy. However, it has been also identified that line search strategies can be flawed as relative minima might lead to wrong parameters convergence. Global search methods perform a better in the assimilation phase and deliver better parameters convergences, nonetheless, they are much more computationally costly (some times jeopardizing the positive lead time) and they can also encounter local minima solutions and deliver a faulty result.

Finally, by means of parallelization, the patternsearch strategy could be speeded up 5 times when using 7 dedicated cores. The rest of the strategies do not show any significant improvement with the parallelization. Although this was, up to some extend, expectable for line search strategies, the fact that two of the global search do not show any improvement is most highly probable to inappropriate source code development (out of the scope of this thesis).

Chapter 7

SmartQFIRE validation

After implementing and exploring a data-driven system to forecast wildfires, this chapter is devoted to its validation. One of the principals challenges when dealing with validation is to properly gather data from a real event. The logic prioritization of emergency response together with the lack of an easy-to-deploy system to monitor the fire spread, make it difficult to collect the necessary information to set the ground truth to perform the validation. Thus, due to the lack of correctly characterized large scale real fires (i.e. minutely fire fronts, canopy cover maps, weather information, etc.), this chapter uses two large-scale experimental burns together with a tuned and validated simulation recreation of a wildfire that took place in northern Catalunya. Part of the results displayed in this chapter were presented in a talk at *WCNR2017 - World Conference on Natural Resource Modeling, 2017* (Rios et al., 2017a) and are accepted for peer-review in a special issue of the *Frontiers in Mechanical Engineering*[1].

7.1   Validation data sources

One of the main difficulties when it comes to system validation in real-scale and non-flat topography is to find convenient sound data. Most of properly reported experiments are small burns performed in indoor facilities (Viegas (2004); Pastor et al. (2006); Rossa et al. (2015)). Although those experiments are important to develop deep understanding of fire dynamics, they are not suitable to validate the system at hand which is foreseen to be used in large scale (two or three order of magnitudes higher than indoor experiments). Gathering proper data in real scale scenarios has two principal downsides. First, in the development of an emergency the priorities are clear. Controlling and putting out the blaze will always be prior to record data. Although this is logically conceived,

---

[1]https://www.frontiersin.org/research-topics/7938/wildland-fire

a paradigm shift on data-driven simulations and decision support tools that provide responders with key forecasts to efficiency manage the emergency, encompasses the need to properly monitor real fire events. The second difficulty relies on the technology needed to provide such data. Although infrared cameras have experienced a dramatic decrease in price and robustness due to microbolometer sensor technology, an integral system able to be rapidly deployed in a wildfire event is not yet ready. Research is being conducted to develop low-cost equipment to be boarded in aerial platforms (both manned and unmanned) and to generate the required algorithms for real time processing and convert raw data into useful information to feet decision support tools (Valero et al. (2015, 2017b)). Meanwhile, medium-scale experiments together with recreation of fire development made by the operational responders, are the only available source of data to accurately validate the model.

7.2 Small scale assimilation validation

In this section we use data from an experiment campaign conducted in Gestosa (Portugal) to validate the assimilating capabilities of SmartQFIRE. The forecast evaluation cannot be validated, in this case, due to the few isochrones (and spread time) available and the unstable nature of the initial phase of a fire in a slope terrain.

The two runs were selected to perform such validation. Both experiments were set on a non-flat terrain of about 1 ha delimited shrub fuel plot. Figure 7.1 shows and example of one of this plots.

FIGURE 7.1: Example of Gestosa campaign plot. Source: Bianchini et al. (2005b).

Whereas this scale cannot be considered large-scale, it is indeed a real fire that can be used to specific features validation such as slope and short-term forecast capability. The

mentioned experimental campaign was conducted in *Sierra de Lusã*, Gestosa, Portugal in 2004 (Viegas (2008)). The two specific experiments selected to validate the SmartQFIRE are two constant slope plots of similar sizes delimited by a fire-break (non-combustible strips). The exact plot dimensions together with their slope are reported in Table 7.1:

TABLE 7.1: Gestosa experiments in sloped plots. The available isochrones are spaced every 2 min.

Plot Name	Width [m]	Length [m]	Slope [º]	Duration [min]	Available Isochrones
Experiment 1	90	80	14	14	6
Experiment 2	90	60	6	12	5

Pyrotechnic ignition was used to create the initial fire-line. The fuel complex was composed by Mediterranean shrubs (*Erica umbellata, Chamaespartium tridentatum* and *Halimium sp.*). In both cases the fire breaks are input in the simulation canopy cover to define the real combustible domain. As a result, the flanks of the fire are fully determined by those fire breaks and the validation is focused on the front dynamics.

Although it has been impossible to track back the exact value of the parameters for this given specific plots, we gathered data of the adjacent plots in Gestosa campaign to set up the ground truth range to evaluate the parameters convergence. Those ranges are summarized in Table 7.2.

TABLE 7.2: Gestosa experiments parameters. Source: Viegas et al. (2006) and Pastor (2004). Data on wind speed and direction could not be found.

Fuel Load [kg m^{-2}]	SAV [m^{-1}]	Moisture content [%]	Moisture of extinction [%]	Fuel Depth [m]
$0.35 - 7$	$5000 - 7000$	$5 - 15$	$20 - 40$	$0.5 - 1.5$

Despite these data do not allow for a conclusive validation in terms of final parameter convergence, it will show if the best-matching solutions are within the expectable range of values found in the experimental campaign.

For both cases, the first isochrone at 2 min after the fire ignition has been taken as a first observable front. The reason why we avoided starting from the linear front is imperfect ignition (i.e. not all front is light up with same intensity) and the fact that taking the irregular first isochrone is actually a more realistic case, as the front is then allowed to spread following the front-shape dynamics.

7.2.1   Experiment 1

The 6 available isochrones of this experiment were extracted from static images used to report the time evolution of the fire presented in Figure 7.2. It can be easily seen that

there is an evident acceleration of the fire front. This is due to the close-to-ignition time and it will become a steady spread speed as time advances. As a recall, the system at hand, does not handle this acceleration if it is not caused by change of given parameters (slope, weather, fuel properties, etc.). Thus, the challenging part of this validation is to see if the final perimeter will match with the given data.
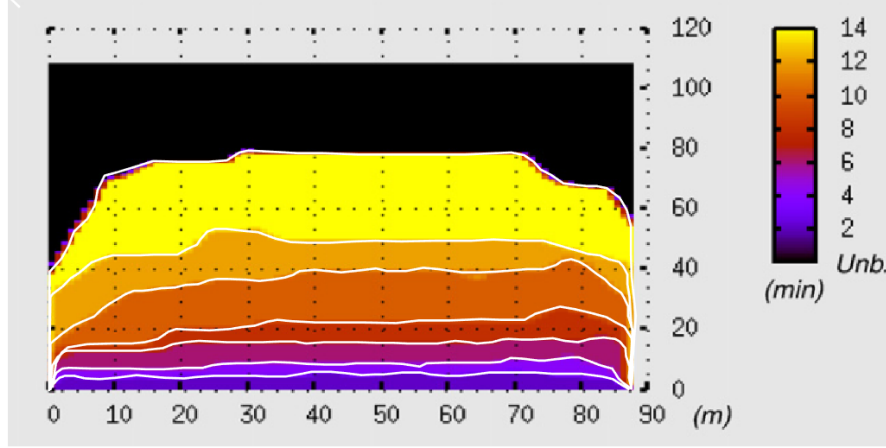

FIGURE 7.2: Experiment 1 isochrones. White solid lines are the extracted isochrones. Note that axis are not equally scaled. Source: Bianchini et al. (2010), modified.

The simulations where initiated with the guess values reported in first row of Table 7.3. In order to assess the effect of the fire acceleration into the assimilation process two different weight factor approaches (see section 5.4.3) were explored. Figure 7.3a shows the best possible assimilation without using a weighting factor (i.e. identity weight). The solution converges to a representative rate of spread that overestimates the initial fronts and underestimates the later ones. This is perfectly logic and shows correct behaviour of the assimilation process. Alternatively, when the normalized exponential factor is used (see eq. 5.26) the last fronts are better assimilated while firsts fronts are overestimated. In this case, as the acceleration is clear, using the exponential factor might help in properly resolving the parameters for a correct forecast.

The correctness of the assimilation can be quantified by using the similarity index defined in section 5.4.2. The results for both cases are gathered in Figure 7.4. In both cases, the modified version of the SDI index that subtracts the initial front area scores very high (i.e. wrong simulation-observation matching) for the first isochrones as it is very sensitive to close-to-origin front where the spread area is small. As mentioned before, the last fronts for the expN case are four times better matched than for the identity weight case. Finally, from both panels in Figure 7.4 we can see that the isochrones 2 and 3 (i.e. 6 and 8 min after ignition) show quasi steady state behaviour as the error is kept almost constant for both cases. Before and after those times, the fire does not spread in a steady state manner.
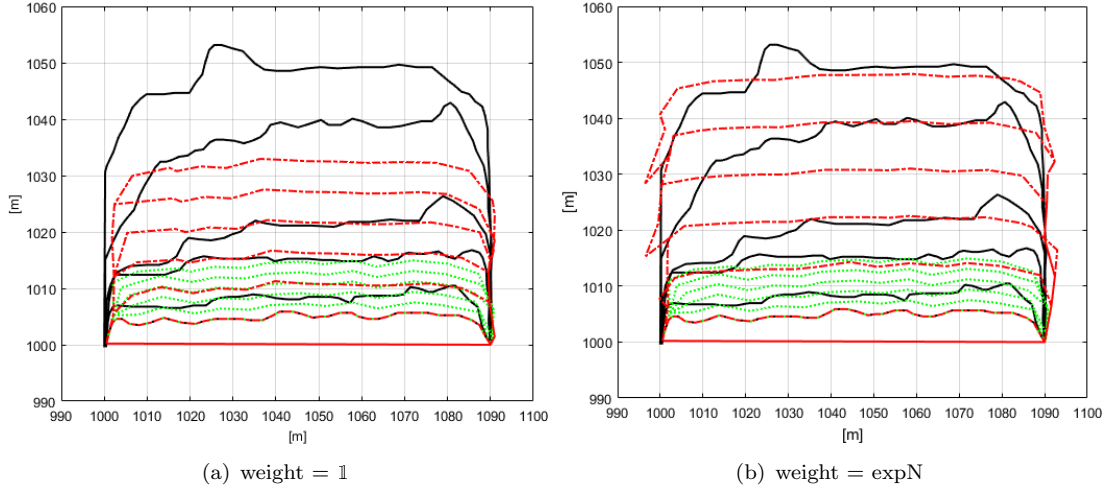
(a) weight = $\mathbb{1}$ (b) weight = expN

FIGURE 7.3: Gestosa assimilation. Black lines are observed fronts, green dotted lines are initial run and red dot-dashed lines are optimized fronts.

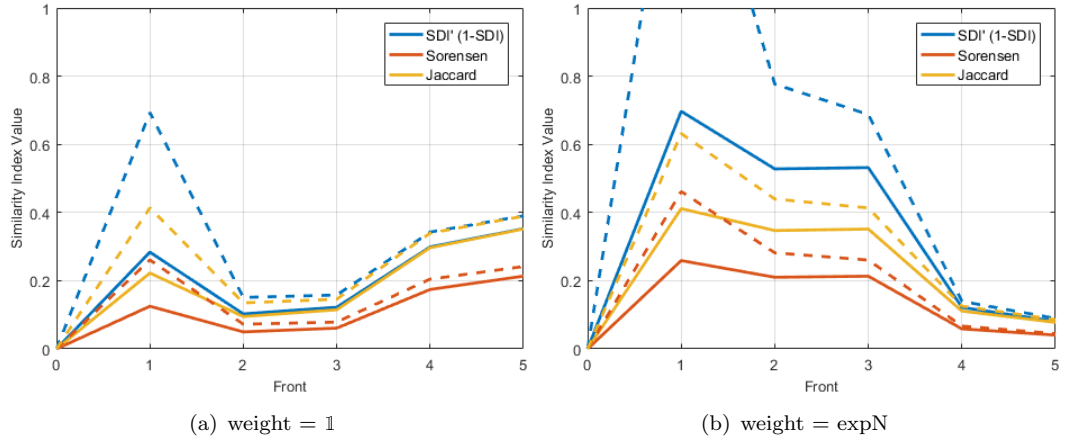(a) weight = $\mathbb{1}$ (b) weight = expN

FIGURE 7.4: Similarity index quantification for experiment 1. Dashed lines are index modified by subtracting the initial burn area (see section 5.4.2)

Finally we can compare the converged values of the simulation parameters for each case and against the estimated validity range. Results are gathered in Table 7.3. Fuel load is the parameter that changes the most when using the different weighting factor (almost 50%) whereas SAV is practically unaffected. All the converged values are within the ground truth range established in Table 7.2.

TABLE 7.3: Initial and optimitzed values for the invariants in each case. Last raw represents the percentual difference between the two assimilation strategies.

	Mf	Mx	SAV	W	U	θ	δ
Initial Guess	0,20	0,30	5754,00	1,90	5,00	0,00	1,00
No Weight	0,15	0,31	5757,80	0,98	3,73	0,00	0,96
Weight = exp	0,10	0,40	5754,00	1,90	3,80	0,00	1,20
% Change	25,82	-29,67	0,06	-48,28	-1,49	0,00	-24,44

7.2.2   Experiment 2

This second experiment was conducted under similar conditions as the previous one with the only change of the slope that went down to $6^{\underline{o}}$, less than half of previous case. In this case, there are 5 available isochrones, representing 8 min of fire development. Differently to the previous one, in this case we are interested on the effect of the initial wind direction and speed, as this was unknown for us. The temporal development used to extract the isochrones is plotted in Figure 7.5.
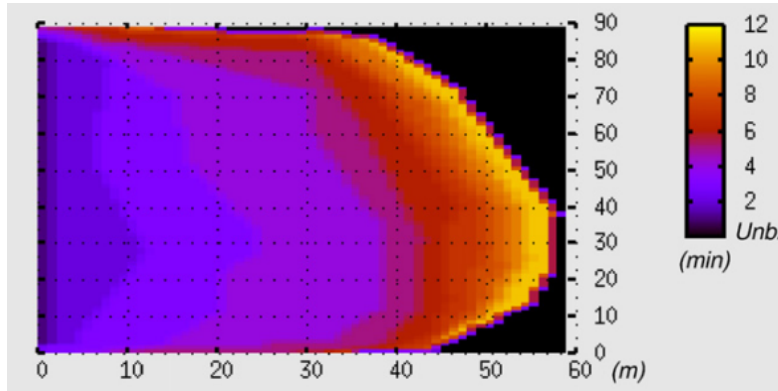

FIGURE 7.5: Experiment 2 from Gestosa campaign. Note that axis are not equally scaled. Source: Bianchini et al. (2010)

In this case, the accelerating effect of the fire is less noticeable than the previous case and thus, the identity weighting function case has been used. Figure 7.6 displays the results for both assimilation runs initialized with the data gathered in the first row of Table 7.4. The final assimilated fronts show that, in terms of similarity, both cases manage to properly mimic the observed fronts. Looking at the similarity indexes displayed in Figure 7.7, it is clear that both cases resolve the problem in a similar way, in particular the last two isochrones as all similarity index give almost identical values. However, in terms of absolute parameter convergence (Table 7.4) we can identify non-negligible differences in the moisture content (M_f) and the fuel load (W). Although both cases still fall into the acceptable range this outcome highlights the need to narrow initial estimation as the final convergence can still provide multiple solutions.
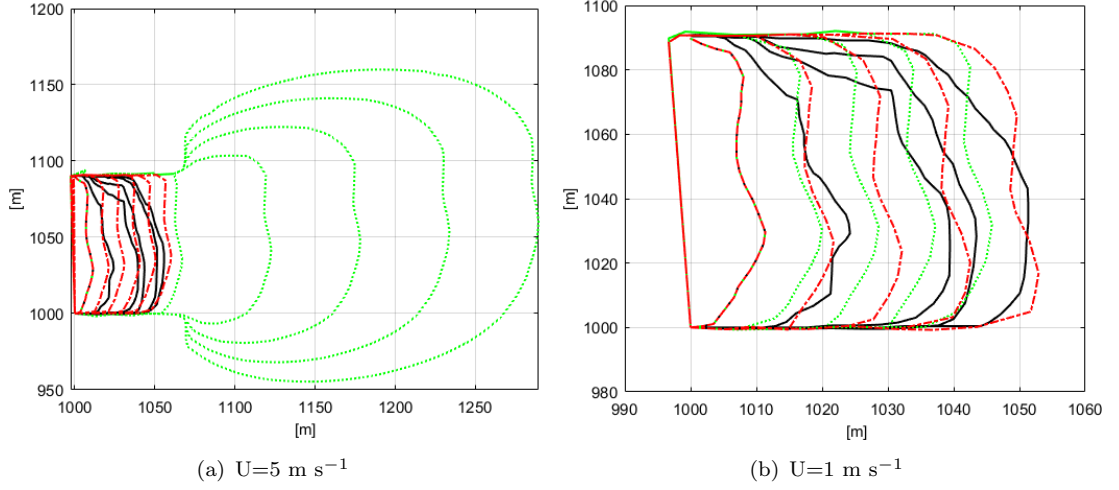
(a) U=5 m s^{-1} (b) U=1 m s^{-1}

FIGURE 7.6: Gestosa experiment 2 assimilation. Black lines are observed fronts, green dotted lines are initial run and red dot-dashed lines are optimized fronts.

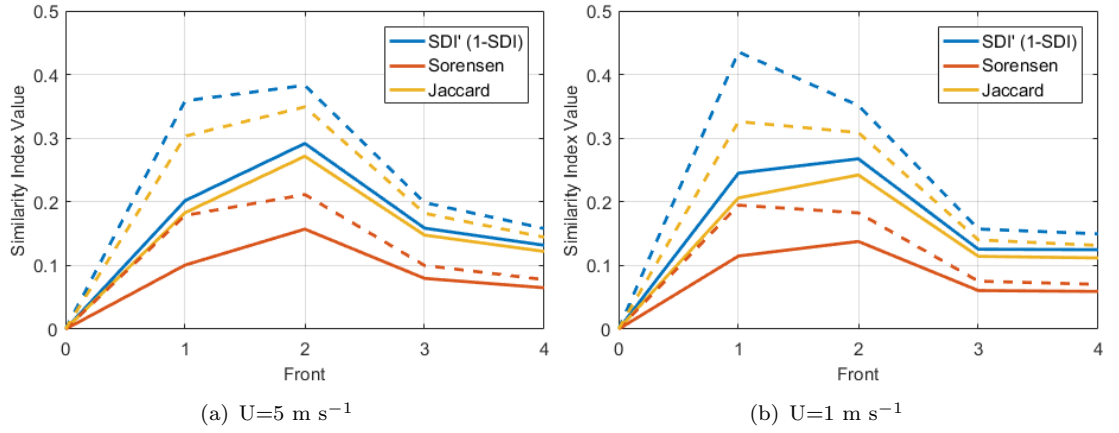

(a) U=5 m s^{-1} (b) U=1 m s^{-1}

FIGURE 7.7: Assimilation performance. Dashed lines are the modified index as introduced in section 5.4.2).

TABLE 7.4: Assimilated values for the seven invariants. Wind speed (U) and direction θ have two initial guess values due to the different initialization.

	Mf	Mx	SAV	W	U	θ	δ
Initial Guess	0,2	0,3	5754	1,9	5/1	355/ -91	1,70
Uo=5 m s^{-1}	0,40	0,40	5818,60	1,30	3,50	-0,03	0,80
Uo=1 m s^{-1}	0,20	0,30	5754,20	1,80	0,60	0,00	1,00

7.3 Validation with real wildfires: Vall-llobrega study case

Finally, we want to test the data-driven algorithm with a case as close as possible to real wildfire emergency scenario. Despite the efforts, we could not find a wildfire event that

gathered all ideal conditions to be used as a study case. The limiting factor was always the characterization of the fire front. Only approximate fire front positions in a time and spatial resolution of the order of hours and tens of meters, respectively, were available in the majority of the cases inspected. This limitation mostly arises due to the lack of an easy-to-deploy airborne system that could automatically track the front with high temporal and spacial resolution. Whereas this is not into the scope of this thesis, we have also been working on this direction (Valero et al., 2017a,c, 2018) to contribute on the prompt development of such a tool. Thus, it will be a future work task to fully validate this data-driven system with the necessary data gathered with those news technologies.

The approach taken here, is to use a real fire event that was recreated and validated with operational observations by the Catalan Fire Service. The aim is to analyse how SmartQFIRE algorithm performs in a pseudo-real emergency situation, when all the upgrading features explained in detail in the precedent sections taken into account. Thus, we will consider different assimilation scenarios with the common lack of accurate data on fuels and weather conditions to imitate different runs of the algorithm along the first 2 h 50 min of the intervention.

7.3.1   Vall-llobrega wildfire

The study case corresponds to a wildfire of significant repercussions occurred in Catalonia in March 2014. Detailed information about fuel and meteorology was collected by the fire brigades and allowed reconstructing the fire development with Farsite, a widely employed simulator. This reconstruction was validated using observed data about the burned perimeter evolution by L. Castell (personal communication, 2016).

The fire started near the village of Vall-llobrega, on the northern shore of Catalunya (see origin sign in Figure 7.9) at 13:04h of 16th march 2014 and it ended burning 359 ha. The extinguishing operation involved more that 50 fire engines, 7 aerial means, and more than 150 fire fighters. The fire took place off-season but showed similar behaviour to summer fires (Figure 7.8) in terms of propagation speeds and fire intensity. Reasons were twofold. The severe drought regime present at the time (see draught code values (Turner (1972)) displayed in Figure 7.11) and the presence of large dead fuel on the ground due to an uncommon snowfall in a precedent year (2010), gave additional available fuel.

(a)   (b)

FIGURE 7.8: Aerial images of the Vall-llobrega wildfire. Source: La Vanguardia
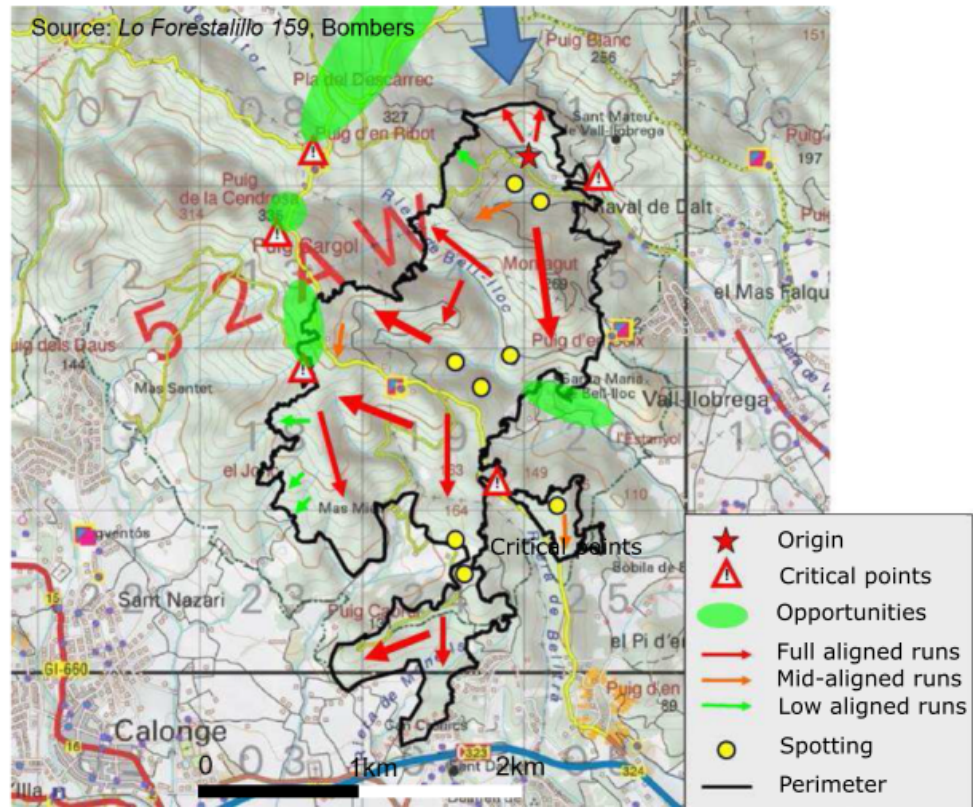

FIGURE 7.9: Vall-llobrega fire development and final perimeter. Source: GRAF (2014), adapted

The weather conditions were dry and windy. When the fire was initiated, the relative humidity was of 20% and the wind was blowing from north at speed over 36 km h^{-1} with reported wind gusts of 80 km h^{-1}. Data logged by the automatic weather station located at 10 km from the fire are reported in Figure 7.10.
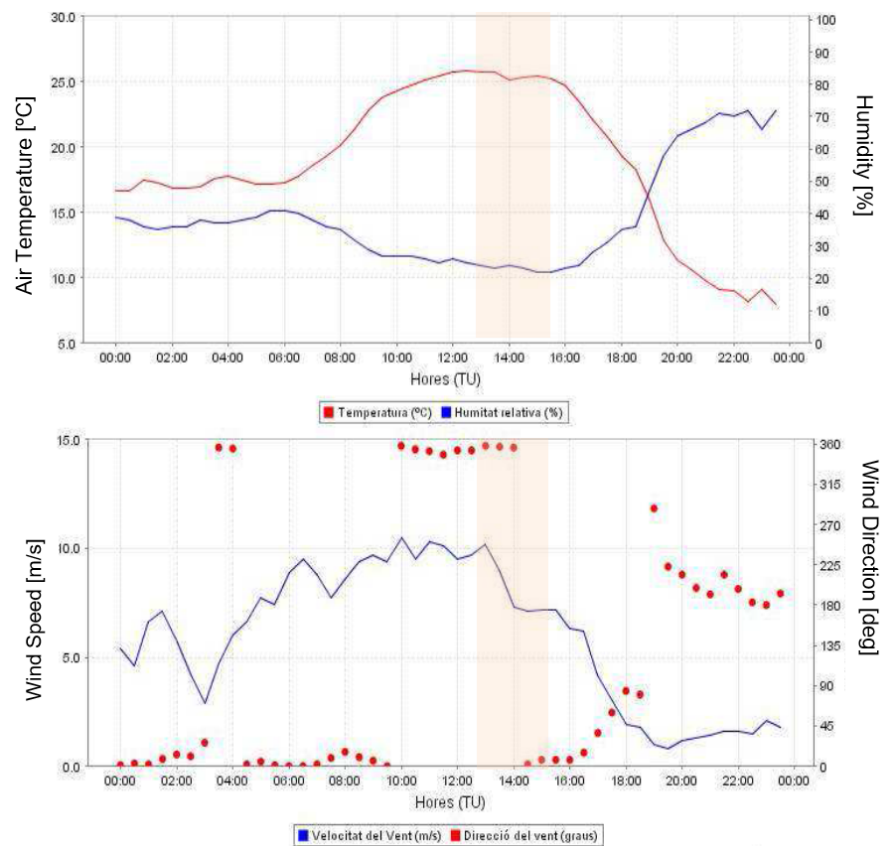
FIGURE 7.10: Wind, temperature and relative humidity data logged at the automatic station of La Bisbal d'Empordà (10 km from the fire) on the 16/03/2014. Source: Servei Meteorològic de Catalunya
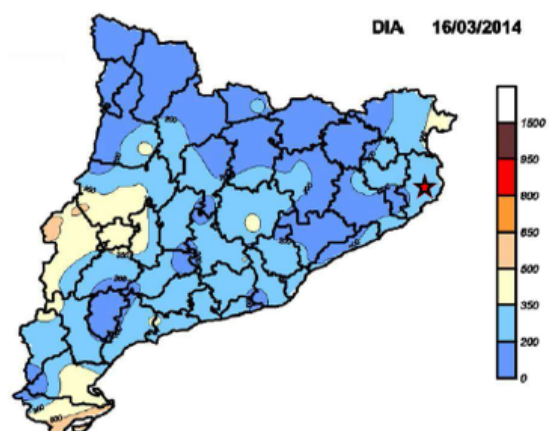
FIGURE 7.11: Drought index values on 16/03/14 over Catalunya. The star symbol marks Vall-llobrega location. Source: GRAF (2014)

7.3.2   Vall-llobrega ground truth data

As previously mentioned, this fire was recreated using Farsite (Finney (1998)) by members of Catalan Fire Department closely involved with the emergency response operations to tackle the fire. The Farsite input files, together with the adjustment factors, were manually tuned to closely reassemble the 2 h 50 min initial development of the fire. Those are depicted in Figure 7.12.
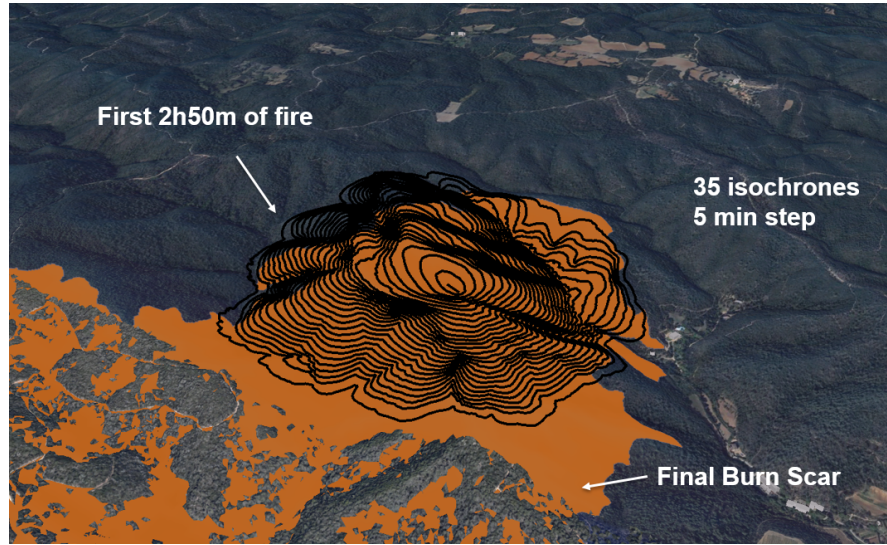
FIGURE 7.12: Farsite validated recreation of Vall-llobrega fires for the initial 170 min. Black lines are 35 isochrones at 5 min frequency. Orange solid contour is the satellite-based final scar (source: IGCC (2018)).

Taking those 35 isochrones as ground-truth, we have run the following three cases.

- **Case A**: 5 assimilated fronts (20 min), 29 forecast (145 min)

- **Case B**: 10 assimilated fronts (45 min), 24 forecast (120 min)

- **Case C**: 22 assimilated fronts (105 min), 12 forecast (60 min)

7.3.3   Assimilation runs

To initialize SmartQFIRE software the 6 parameters to be optimized were estimated with the values gathered in Table 7.5. This could be a blind estimation done in real situations as it only requires a rough idea on the scenario conditions. As example, for the case at hand, the fuel characteristics had been paired to the fuel model TU5 *Very High Load, Dry Climate Timber-Shrub* (Scott and Burgan, 2005c) as fire responders already had the information that packed and dead understory was present.

The results of the assimilated fronts for the three cases explored are plotted in Figure 7.13. It is clear that the estimated spread velocity (mainly due to the high wind speed

TABLE 7.5: Values of the parameters estimated to initialize the assimilation run

	w [kg m^{-1}]	SAV [m^{-1}]	δ [cm]	m_x [%]	m_f [%]	u [km h^{-1}]	Dir [deg]
Ini. Guess	1.73	4015	110	25	20	50	360

value) was overestimated. The dashed-green lines in 7.13 are the estimated propagation with values of Table 7.5. For the cases where 5 and 10 fronts are assimilated (panels a and b), the assimilation process converges into a situation that closely reassembles the observations. In particular, the topographic effect in both flanks is well resolved. The 10 front case, however, already show some divergence on the tail as converged fronts spread slightly further than observations. This effect is more remarkable in the third case, (7.13c) where this overspread is larger. In this case, however, flanks and fronts are still correctly resolved. The difficulty to properly simulate the tail part of the fire might lie on the fact that backwards spread (fire spreading against wind and topography) is not well characterized yet. A dedicated sub-model should be implemented, once available, to improve this weakness.

(a) Case A: 5 fronts   (b) Case B: 10 fronts   (c) Case C: 22 fronts
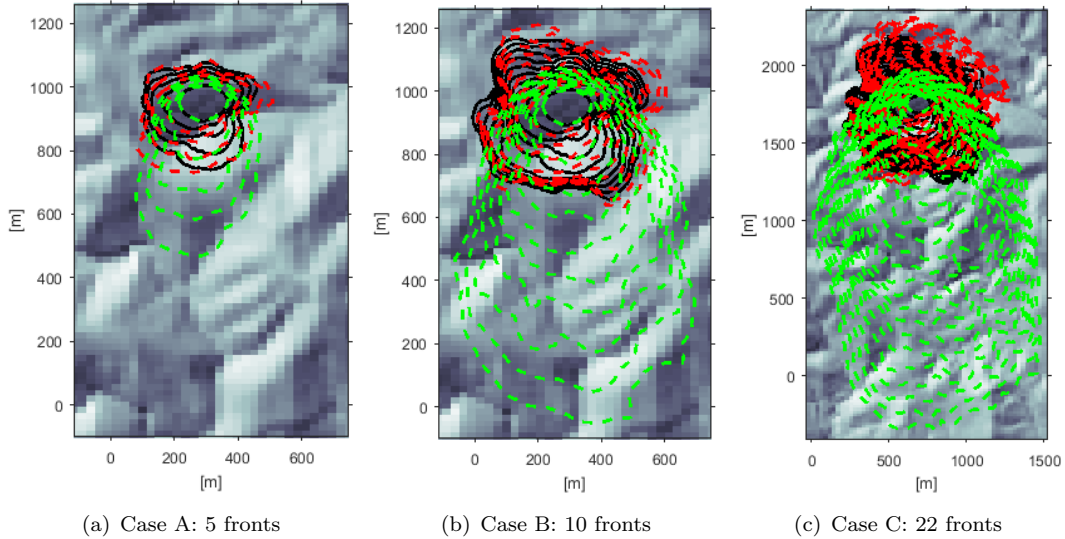
FIGURE 7.13: Three different assimilations runs. North direction is towards the top of the image. Black solid lines are observations, dashed green lines are initial run and dashed red lines are the optimized fronts at the end of the assimilation period. Background image is the hillshade representation of the domain.

7.3.4   Invariants convergence

To assess the convergence of the assimilation process we analyse, on one side, the convergence of the simulated fronts and, on the other, the convergence of the seven assimilated

parameters. Figure 7.14 shows those two convergences for the first case presented here (20 minutes of assimilation). As we use the active-set strategy (see Section 6.2), each of the iterations indicated in the Figure make use of multiple evaluations of the forward model in order to define the following step direction. The mean area cost value depicted in the left panel of Figure 7.14 is the absolute areal difference between each of the observed and simulated fire perimeters averaged by the number of simulated fronts. It is remarkable that in the 9 th iteration this value diverges rising to 240 m^2. The explanation for this behaviour might lie on several aspects. First, the cost function to be optimized is not this area difference but the SDI index (see Section 5.4.2). This peak could, thus, be related to some mismatch between the SDI index and the real matching area. Secondly, as explained before, the optimization is constraint, and thus, it has been observed that when converging process goes to the boundaries of the optimization domain, the parameter are forced to fall back into the valid range. Those processes have been sometimes identified to produce this local divergence of the optimization process.

The relative parameters value convergence (Figure 7.14b) is expressed as a percentage of the final converged value. Thus, all parameters converge to 100%. We can identify here that the convergence is not smooth and steady as they are large jumps (SAV and u at iteration 17 per example). This supports the finding mentioned on the previous chapters that the optimization problem at hand was not a smooth problem and thus, simple linear search algorithm did not provide an acceptable solution.
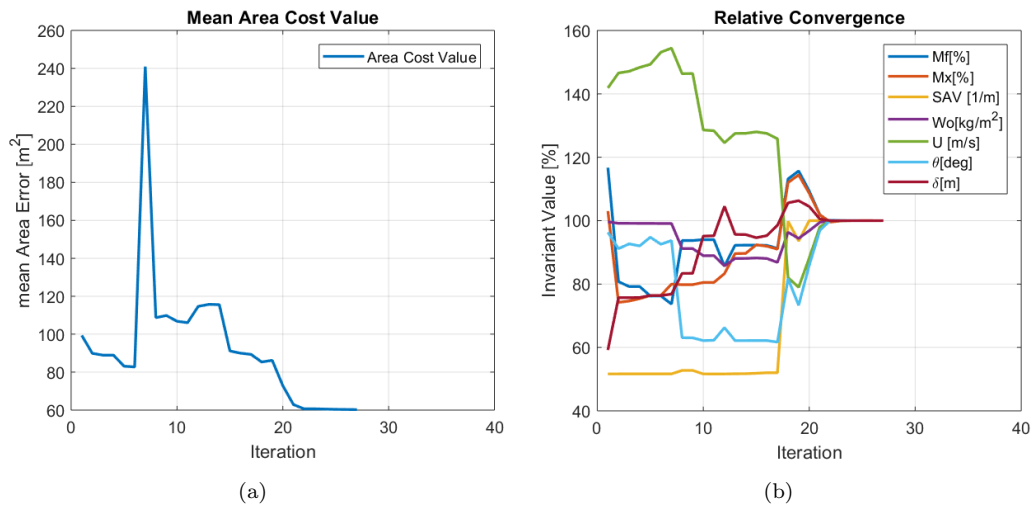

FIGURE 7.14: Mean area error between observed and assimilated fronts (left) and the relative convergence for each of the 7 optimized parameters.

The converged values for each parameter and case are gathered in Table 7.6. Analysing them one-by-one we find that all three cases generated here mostly converge towards the same set of parameters, except for the wind speed of case C, that is considerably

lower than the other cases. The fine fuel load (w) is reduced for all cases with respect to the initial value as well as the moisture of extinction and moisture content. This could provide a larger fire spread, which would be inexplicable regarding the fronts produced with the initial guess (see green dashed lines in Figure 7.13), but the dramatic wind reduction prevents this. Indeed, for all cases the effective representative wind speed is lowered by about 10 km h^{-1}. The wind direction pivots around the 360$^{\text{o}}$ value. That is, wind flowing from north. It is worth to recall here that the direction (Dir) is internally expressed in radians and values are always projected into [π - 3π] bounds for numerical reasons (see Table 3.9). Another remarkable output is the converged value of the surface-area-to-volume ratio (SAV). In all cases, it increases largely, and even in case A, it gets close to the upper bound (stated at 7270 m^{-1}). These values indicate that from our simulator standpoint, mostly the thinner fuel is the principal contributor to fire spread. The correctness of this results would need of post-fire data to assess the principal fuels involved. Despite the lack of this information, the observations of high intense fire behaviour and its high propagation speed might lead to think that this outcome is reasonable.

TABLE 7.6: Final absolute values for all seven parameters after the assimilation process. The initial estimated values are also depicted in the first row.

	w [kg m^{-1}]	SAV [m^{-1}]	δ [cm]	m$_x$ [%]	m$_f$ [%]	u [km h^{-1}]	Dir [deg]
Ini. Guess	1.73	4015	110	25	20	50	360
Case A	1.72	7234	183	24.0	16.9	34.9	371.1
Case B	1.51	6915	167	25.3	17.4	32.3	351.6
Case C	1.32	6859	172	23.6	16.3	27.7	358.2

7.3.5   Forecast runs

Once the seven parameters are calibrated by means of the assimilation process, the forecast is launched with the fronts left available in each case. The results are depicted in Figure 7.15. For all cases, the last assimilated front is the one used to initiate the forecast model run.

As could be expected, case A (with the less assimilations) performs poorer than the two other cases. This case correctly forecasts the back of the fire and the right flank after 145 min (2 h 25 min) (7.15a), however it over-predicts the front part of the perimeter by a couple hundreds of meters. The case B also manages to forecast the tail of the fire and it matches correctly the front and the flanks. The last three isochrones show a burst on the rear part of the right flank. This might be due to bad terrain handling, as it seems that this part of the front gets accelerated all at a sudden. Contrary to the

previous cases, the case C, the one that has assimilated more fronts, under-predicts the tail of the fire and slightly over-predicts the front. Nevertheless, the right flank is better resolved than previous cases.
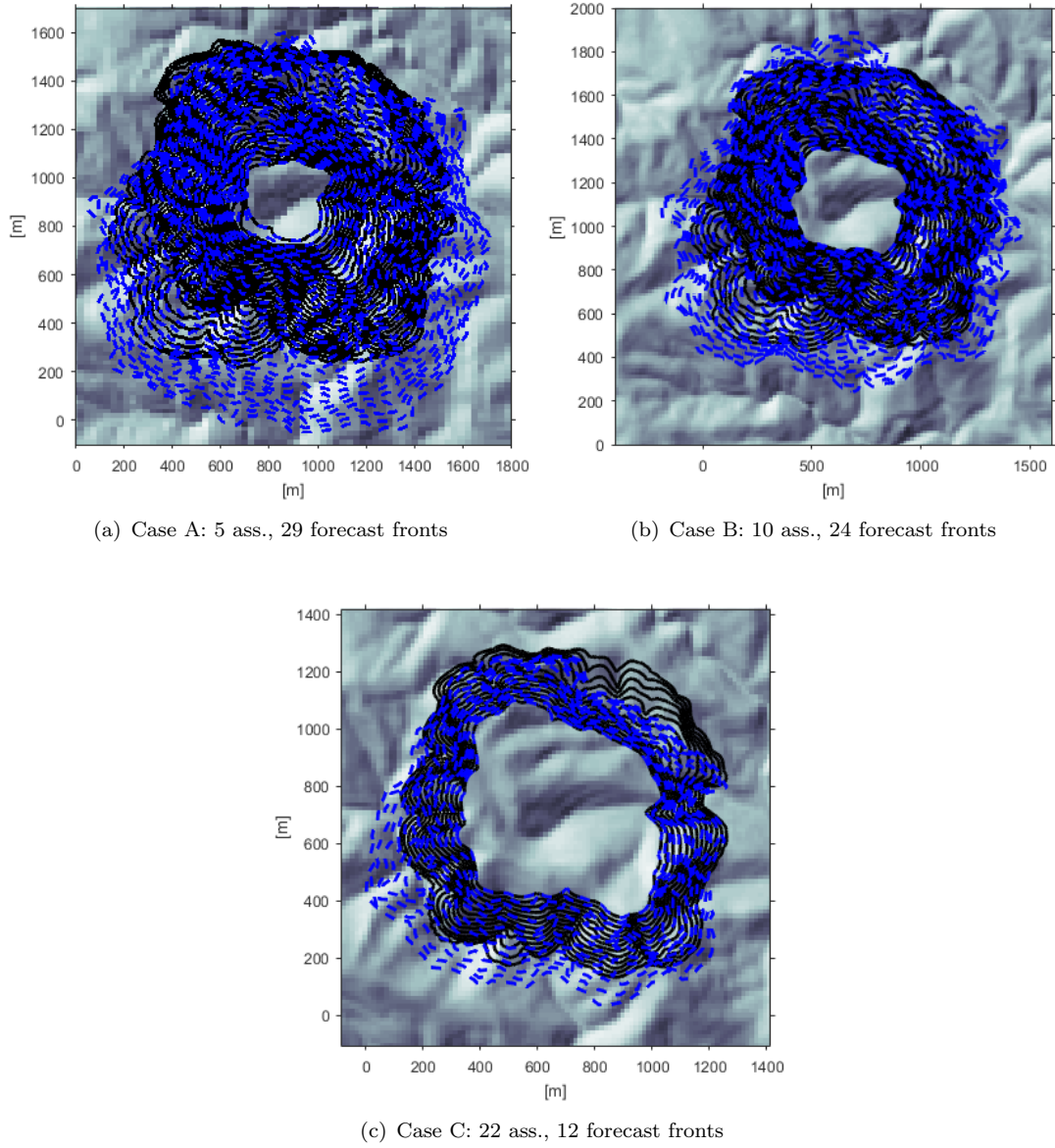

(a) Case A: 5 ass., 29 forecast fronts

(b) Case B: 10 ass., 24 forecast fronts

(c) Case C: 22 ass., 12 forecast fronts

FIGURE 7.15: Forecast run (blue dashed lines) versus observations (black solid lines) for the three cases.

To help in the analysis of the forecast performance of each case, the SDI index for each forecast front are plotted in Figure 7.16. Each marker in the Figure corresponds to a single SDI value for a given time. That is, at time 110 min the forecast of the three cases for the same front is compared. For this exact isochrone, case C is the over-performing one as its SDI index is larger than the rest of the cases. However, this is probably due

to the fact that case C was initialized at minute 90 so it is actually the forecast at 20 min horizon. Indeed, as time passes by case C performs worse than the B case. Case A, delivers an acceptable 30 minutes forecast. Beyond this time, the performance decreases with time, ending up with rather unacceptable forecast for the last 50 min. It is also remarkable the fact that at 125 min the Case B over-performs case C. This means that at this point, less assimilated fronts produce a better forecast. Indeed, as the whole systems is build up on finding representative parameters that are kept constant, there is a trade-off between the assimilating time and this averaging effect. Ideally, once enough validations are performed, the optimal assimilation windows for a given case could be defined and it should be shifted over time (without increasing the number of assimilated fronts) as the emergency evolves.
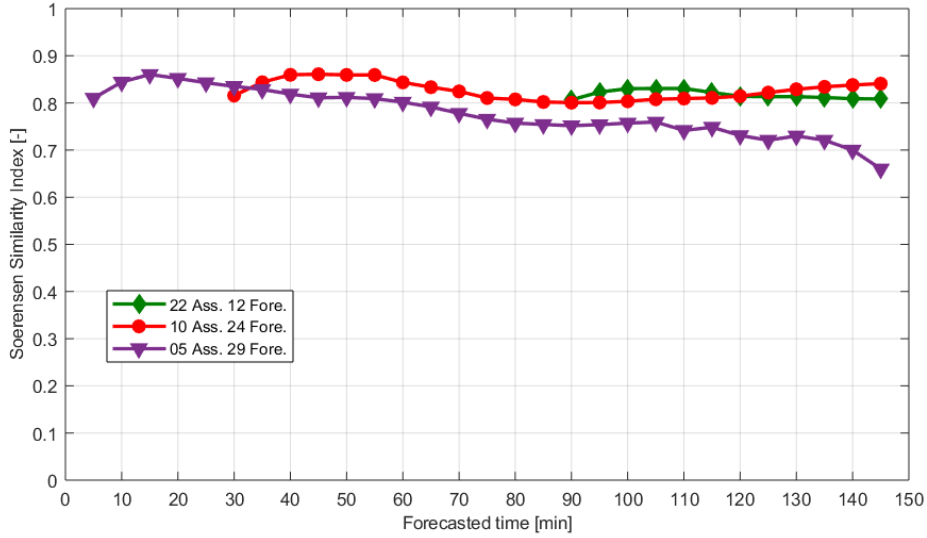
FIGURE 7.16: Similarity index (SDI) performance for all forecast fronts and the three different strategies explored.

Finally, the running time to perform all the assimilations and forecasting runs for each case are reported in Table 7.7.

TABLE 7.7: Running times and lead times for all cases. The simulations are conducted with an Intel(r) Core(MT) i7-6700CPU 3.4Ghz (8 physical cores).

Case	Elapsed wall-clock Time [s]	Forecast horizon [s]([min])	Lead time [s]([min])
Case A	115	8700 (145)	8585 (143)
Case B	251	7200 (120)	6949 (115.8)
Case C	380	3600 (60)	3220 (53.6)

All the cases, the lead time (time between the results delivery and the forecast time) exceeds the 50 minutes. This aspect, together with an accuracy estimation to set the

reliable forecasting horizon, are key aspects if SmartQFIRE is to be employed in operational situations.

7.3.6 Conclusions

In this final chapter, the SmartQFIRE data-driven algorithm was tested with real case scenarios that entangled wind and topography interaction.

Initially, data from controlled experiments are used to explore the algorithms capacity to assimilate few fronts in a constant slope terrain. The test also served to check the effect of weight factors in the initial stage of the fire, when an acceleration is present. The results, showed that even though the first estimated value produced fronts away from observations, the assimilation process is capable of correcting the parameters to reassemble the assimilated fronts. It was concluded that normalized exponential factor provides better results for accelerating front behaviours.

The impact of the initial wind estimation was also evaluated using those homogeneous slope experiments. The conclusions were that despite initializing the wind with a value 5 times larger, the observed fronts were properly matched. However, the convergence of the seven parameters was slightly effected. This is due to the optimization strategy, as line search methods are ultimately dependant of the initial searching point.

Secondly, SmartQFIRE was tested with the recreation of a real emergency scenario. The evaluation was performed with the initial 2 h and 50 min of isochrones with a frequency of 5 min. Three different assimilation cases were studied. They could be representative of different times where the assimilating system could be deployed. The algorithm showed great potential to handle complex wind and terrain entangled scenarios. The coupling with WindNinja enhanced successfully delivered realistic fronts that could closely match observation without jeopardizing the operational application due to computing time. Indeed, for all cases studied, the lead time exceeded the 50 minutes. Finally it was found that, in general terms, more assimilated fronts provides better results, the analysis also showed that it exists a trade-off on the averaging calibration parameters and correct forecast. Thus, more research is needed to determine the ways to stablish the correct window that optimizes the forecasting results.

Chapter 8

Conclusions and future work

The aim of this thesis was to develop and explore a data-driven system to forecast wildfire spread in operational conditions by assimilating wildfire perimeters (isochrones) to resolve and calibrate the spread model unknown parameters.

In Chapter 2, we studied a simplified data-driven system implementation only valid for flat terrains. The spread model was based on a linearised version of Rothermel's spread theory coupled with Huygens expansion principle. The data-driven module, used direct automatic differentiation and a tangent linearisation of the forward model to solve the optimisation problem. A regridding and a loop-clipping algorithms needed to be implemented to handle the irregularity of the assimilated fronts. Those two algorithms were identified as one of the efficiency bottle-neck but are unavoidable in a marker-tracking approach (i.e. expanding the front through markers in the perimeter).

The overall system was explored with spread results from a real-scale wildfire campaign conducted in Ngarkat Conservation Park (Cruz et al., 2013). The fire front isochrones where extracted from infrared imagery captured by means of a hovering helicopter. The data-reduction process was composed by an image orthonormalization step plus a flame filtering and edge extraction algorithms.

The system showed great efficiency finding the invariants within less than a minute and resolved the optimization process in less than 10 iterations. However, due to the forward model simplicity, a potential problem of multiplicity in the determination of the RoS was highlighted. In some scenarios, and depending on the initial guess and the assimilation window, the invariants I_{mfw} and U could not be correctly identified and they converged to a wrong value. The multiplicity can reduced when more fronts are assimilated although a minimum number of assimilations could not be determined. The reason is that the incorrect convergence is also linked to the scenario and the first guess configurations. The fact that two parameters could be misidentified (i.e. not converged to the real true value) reduces the forecasting capabilities. Possible ways to overcome

this multiplicity problem are resolving extending the complexity of the forward model so it predicts extra parameters (such as flame height or heat release rate) that are also assimilated, or including extra information about the system to break the multiplicity (such us irregular terrain or vegetation maps). The later was chosen for the further development of the system. Finally, this chapter states the required system upgrades to make it capable of handling real wildfire fronts. The results of this chapter where published in the *International Journal of Wildland Fire* (Rios et al., 2016).

One of the necessary steps to improve the initial system was to rethink the required parameters to describe the fuel to be incorporated into the spread modelling. This was conducted in Chapter 3. The revision of currently available fuels classification systems that might be used in a spread model brought up the need of a throughout exploration of all sub parameters before proposing a reduced model. Firstly, we explored in detail the effect of the different fuel particle classes (1h, 10h, 100h, live woody and live herbaceous) to the final RoS value. This methodological sensitivity analysis was also conducted in 16 different moisture scenarios. Secondly, we performed an Analysis of Variance (ANOVA) of all Scott and Burgan (2005b) burnable models. After an initial statistical exploration, the ANOVA analysis for RoS was performed by computing the Sobol' index to quantify the interactions significance of 27 input parameters. To the author's knowledge, this is the first time that Sobol' analysis is conducted for all Scott and Burgan (2005b) fuels. The collection of all cases allows for a general parameter influence prioritization (see Figure 3.12) . The conclusions of this analysis, however, did not allow for direct fuel model simplification. The reason was the mixture of main influencing parameters among different fuel models and their classification variability. Despite this fact, the analysis allowed to state that variations of low heat content, the particle density and the mineral content have a negligible effect on the final RoS.

Chapter 3 finalises with the development of a reduced fuel model. The rationale is to classify the standard fuel models (Scott and Burgan, 2005b) into 6 different categories (Cat.) depending on the containing particle fuel class. Then, for its category, a representative parameters for the fuel load (W), surface-area-to-volume ration (SAV) and moisture content (M) are calculated. Fuel load is taken to be the sum of the 1h, live herbaceous and live woody values. The other two are computed as a weighted sum of the different sub-classes values. The weighting factors depend on each category and are established by global search minimization of the absolute percentage difference (APE) between the RoS values generated with the reduced variables (W, SAV, M) and the RoS generated with the full fuel model parameters. Those coefficients are gathered in Table 3.11. This system allows for automatic conversion of a Standard Fuel Model into a reduced model and vice-versa. The performance of this reduction strategy is explored by applying it to 36 static standard fuel models with changing wind speed, slope and

moisture scenarios. The results show that this reduction strategy gives acceptable errors for all categories except for one, highly influenced by the live woody particle class. This reduction model is then selected to be incorporated to the data-driven system.

Due to the strong influence on the fire spread, accurate wind maps were a necessary feature of the final data-driven system. However, generating terrain dependent wind speed and direction is extremely computationally expensive. In a data-driven logic, these maps need to be updated at every model evaluation (which means of the order of 100 or 1000 times) and thus, the 5 min current computing time is a show-stopper for such implementation. Chapter 4 develops an interpolation system to conduct this update at almost no time expenses. The strategy consists on pre-running a set of wind speed and directions (4 and 8 direction sets are studied in the chapter) and then interpolate the new wind map out of those base values. WindNinja (Forthofer, 2007) was the selected diagnose model to project the domain-averaged wind speed and direction onto 1 m above the terrain as it is one of the most extensively used in the community and is also integrated in the FARSITE simulator. The strategy is validated in eight different terrain scenarios with changing Digital Elevation Model (DEM) resolutions. The in-depth statistical exploration of the results showed that the committed error was less than 7% in the worse case (base-case-off-aligned direction and 21 m s^{-1} wind speed). Fire spread simulations using the original WindNinja simulator and two different interpolations showed the successful application of this strategy as the error after a 20 min spread was of less than 20 m. The developed strategy provides an speed-up of the order of hours, and makes it fully integrable into our data-driven system. In addition to this application, this developed and validated framework might contribute on other fields as probabilistic risk assessment where high-resolution wind fields must be computed for multiple weather scenarios within Monte-Carlo exploration approach. The development and findings of this chapter have been published in the *International Journal of Wildland Fire* (Rios et al., 2018) [1].

Once the fuel characterization and the wind calculations are solved, Chapter 5 develops the upgraded data-driven system called SmartQFIRE. It has important changes regarding the one used in Chapter 2: first, the terrain is now included in the spread model, and thus, also the wind-slope interaction to give RoS values. Secondly, following the results of Chapter 3, the parameters to be calibrated after the assimilation process are increased to 7. More importantly, the cost function to be optimized is changed from a vector-based approach to a similarity index quantification based on areal difference. This change obliged to modify the optimization strategy from an automatic differentiating approach to a gradient-based minima search methods. Several similarity index are explored in this chapter together with three weighting factors to give distinct

---

[1]Figure 4.14 of this chapter was additionally selected to illustrate the cover page of the printed version of the IJWF (27th volume, 4th issue)

weight to fronts assimilated at different times. Whereas the weighting factor exploration concluded that the identity weight case might be the most reasonable one to use, no single similarity index could be identified to give better convergence results. Finally, we use two different scenarios to explore the impact on the fire spread results and the computational requirements of modifying the internal time step needed to integrate the spreading partial differential equations. In the light of the results a 25 s default time step is taken as a default value as it strikes the balance between performance and accuracy.

One of the key aspects to be able to deploy the data-driven system at hand is to provide results in a positive lead time, that is, delivering forecasts before the actual forecast time. The key element of this process is the optimization strategy. Thus, in Chapter 6 we deeply explored six different strategies. Three of them are line search algorithms whereas the other three are considered global optimization methods. By using two synthetically cases generated with FARSITE with different wind conditions we compared the ability of each strategy to resolve the true set of parameters. The comparison looked at the front spread results during the assimilation and the forecast phases together with the computing time and final parameters values. The conclusions are that the active-set strategy is considered to be the most convenient and over-performing one, both, in terms of efficiency and accuracy. However, it was also noticed that line search strategies can be flawed as relative minima might lead to wrong parameters convergence. On the contrary, global search methods perform a better assimilation stage and deliver better parameters convergences, although the strategies used here are not absent of flaws as they can encounter local minima solutions instead of global ones. The end of this chapter is devotes to try to speed up those global approaches by means of parallelization. The results do not show a significant improvement for neither the linear search nor the global search methods. Whereas this outcome was expected on the first case, for the second group it is considered to be a pitfall in the parallel implementation, either on the author's side or in the software one. The findings of this chapter were presented on the *International Conference on Computational Science (ICCS) in Zurich at 11-14 June, 2017* (Rios et al., 2017b).

The final Chapter 7 presents a two-fold validation of the entire data-driven system described in all previous sections. First, we use data from a real-scale experimental campaign conducted in Gestosa (Portugal), Viegas (2008). Although the two selected experiments show real-scale fire propagation they were set in a homogeneous slope terrain and lasted only several minutes. Due to the few available isochrones, only the assimilation stage is validated in this case. For both experiments, different initial guesses were used to explore the robustness of the system to the initialization stage. Although for both cases the calibrated fronts were in an operationally acceptable range (i.e. order of meters of mismatch), the results lead to two main conclusions. First, the fact that the spread model does not resolve fire accelerations for homogeneous conditions is an

important limitation, specially on the initial phases of the fire (as the experiments analysed). Secondly the selection of an initial set of parameters that are far from the true state can lead to a convergence that is not the global correct one. This second conclusion highlights, once more, the multiplicity problem highly present in low complexity scenarios.

The second part of Chapter 7 is devoted to validate the full capabilities of the data-driven system in a real emergency wildfire event. Due to the impossibility of finding a real-scale well-characterized candidate, recreated validation of the Vall-llobrega fire made by the Catalan Fire Service was taken as a ground truth to benchmark the data-driven system. The evaluation was performed with the initial 2h and 50 min of isochrones with a frequency of 5 min. Three different cases were studied as a realistic representative assimilation deployable situations. The data-driven system managed to handle complex wind and terrain entangled scenarios by delivering realistic fronts that closely matched observation. For all three cases, the available forecast time exceeded the 50 minutes. Finally it was found that it exists a trade-off on the assimilating time and correct forecast. Since the calibrated parameters are time-averaged there is a threshold where more assimilated fronts do not provide better solutions. More research will be needed to determine this correct temporal window that optimizes the forecasting results.

Future work

Despite the advancements made during this thesis to develop a data-driven system to be used operationally by emergency responders some important tasks are still needed to deploy such a tool.

- **Time changing parameters**
  One of the principal limitations of the current implementation is the fact that the resolved parameters are considered to keep constant as long as the forecast is run. Whereas for some of them it can be a correct hypothesis (fuel load or moisture fraction might not change much in a few hours), for other parameters and for long forecasting horizons it might be a major limitation. There are two possibilities that are worth exploring to overcome this limitations. First, it is possible to rerun the assimilation process continuously, as new data arrive, and thus, update the parameters values as time passes by. However, this strategy will never foresee acceleration events due to external conditions changes (wind sudden variation or diurnal relative humidity evolution, per example). The other strategy could be to define specific forecast sub-models for those varying parameters. That is, for example, updating the calibrated wind speed value with a dedicated weather

forecast model. Although this strategy could really enhance the data-driven system at hand it should be properly validated with real wildfire scenarios.

- **Spatially changing parameters**
  The spatial limitation of the system at hand stands for the fact that parameters are considered, at the present implementation, spatially homogeneous. The solution to this limitation was already explored by Zhang et al. (2017). It consists in regrinding the simulation domain with sub-cells where parameters are static. Then, different parameters values are allowed for each sub cells. Although this strategy showed great capacity to resolve an experimental real-scale fire, the question still arises on how to project those cells in the unknown forecasting domain. Research needs to be conducted in this field to better work around this limitation.

- **Forward model**
  To keep developing this methodology and make it suitable for more real scenarios the spread model (currently based on Rothermel's) should be coupled with additional models for spotting fires and surface-to-crowning transitions. Although some empirical model exists for this spreading mechanisms the lack of a sound theory makes it difficult to quickly solve this limitation. The exploration of more physically-based sophisticated forward models (i.e. pyrolysis, CFD) might require a full adaptation of the framework presented here as the optimization strategy selected will not deliver positive lead times in those cases.

- **Extinguishing**
  In case of operational deployment, the extinguishing effort will definitely reshape and impact on the fire spread. Although this is not a principle limitation for the model's interest (due to the fact that in large fires the fire-front section is usually out of extinguishing capabilities), it is true that flanks and the back of the fire are highly influenced by extinguishing operations. Adding a sub-model that accounts for this will also give the methodology an additional application as it could also be used to benchmark different extinguishing strategies.

- **Further validation**
  Finally, the principal underlying limitation, that is a common need for all research developments highlighted above, is the improvement of data sources for validation. As previously commented, the logic prioritization during an emergency response together with the lack of an easy-to-deploy system to monitor the fire spread, make it difficult to collect the necessary information to set the ground truth to perform the validation. Current efforts performed by CERTEC (Valero et al., 2017a, 2018) and other research groups (Manzano-Agugliaro et al., 2014), are trying to develop the necessary tools to gather these valuable data.

Part II

Appendices

Appendix A

Sobol' Index of all Standard Fuel Models

To create the ranking of the influence of the Standard Fuel Model parameters, the Sobol' analysis for al 53 models was performed. Since, to the best knowledge of the author, the exhaustive exploration has not been published, we attach below the results for all 53 models characterized in Scott and Burgan (2005b).
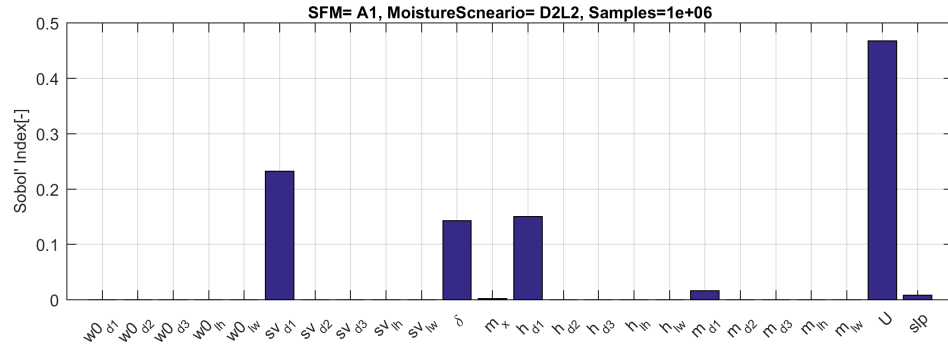
(a)
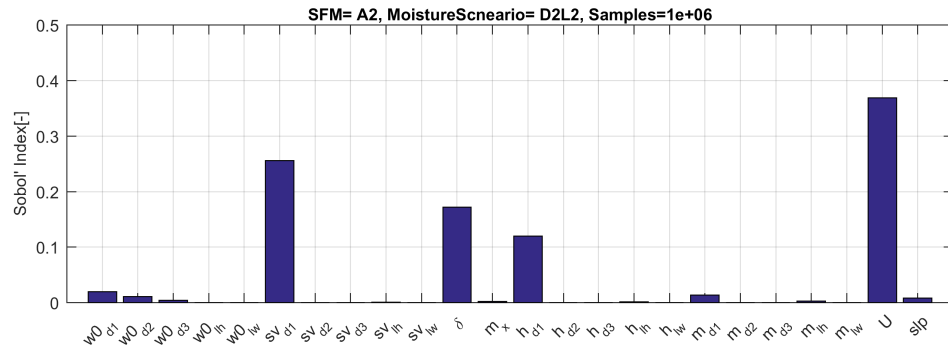
(b)
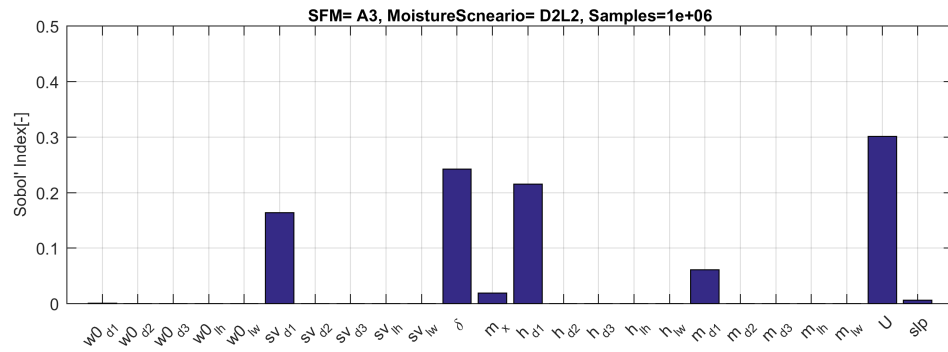
(c)
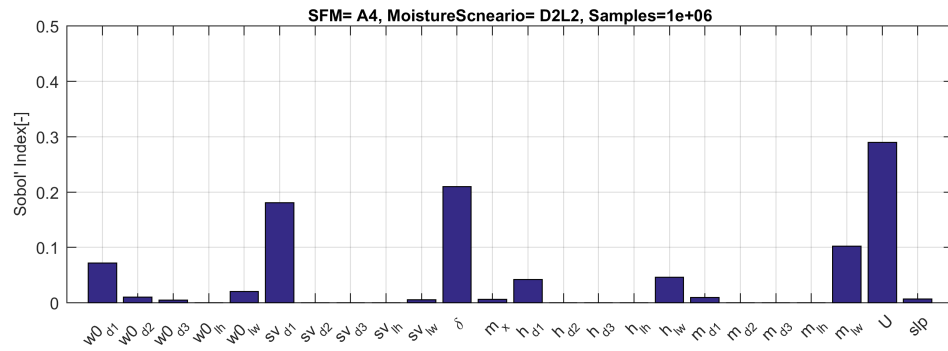
(d)

FIGURE A.1: asda

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

SFM= GS3, MoistureScneario= D2L2, Samples=1e+06

SFM= GS4, MoistureScneario= D2L2, Samples=1e+06

SFM= SB1, MoistureScneario= D2L2, Samples=1e+06

SFM= SB2, MoistureScneario= D2L2, Samples=1e+06

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(d)

(a)
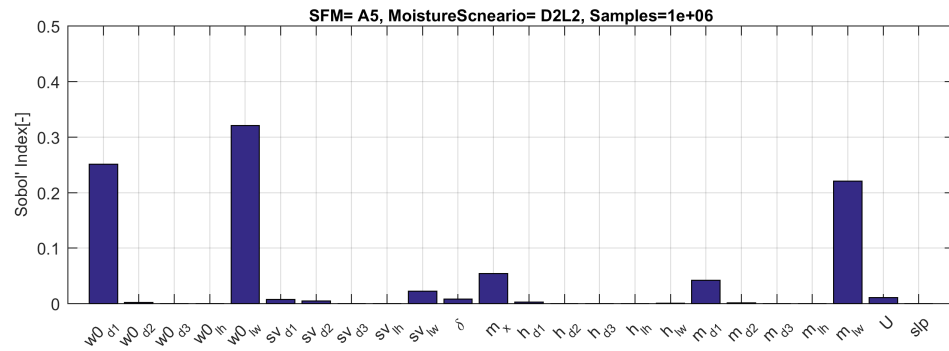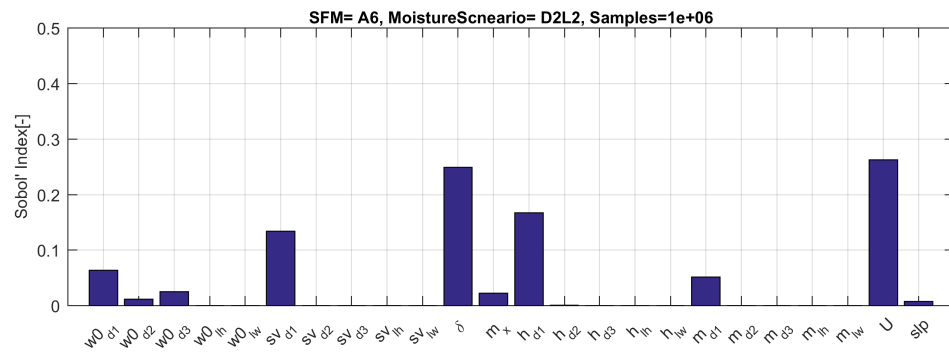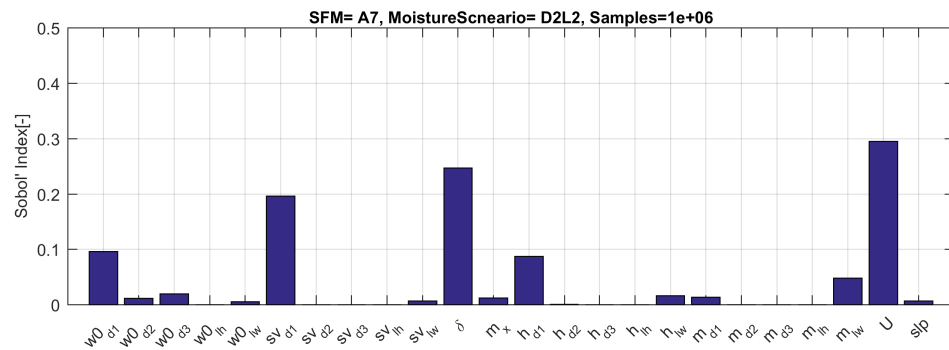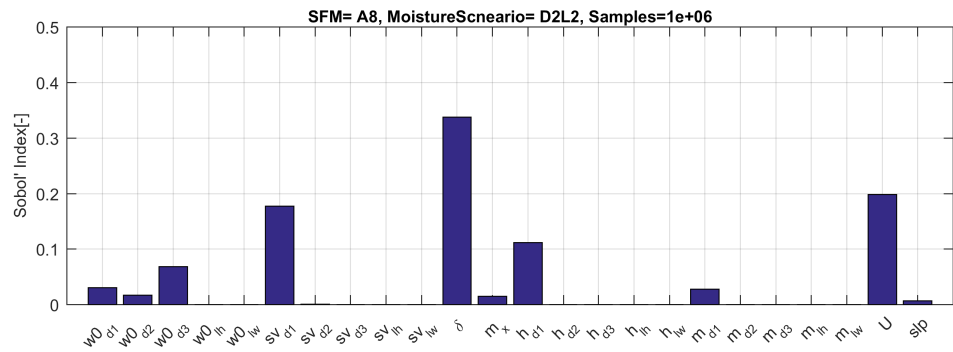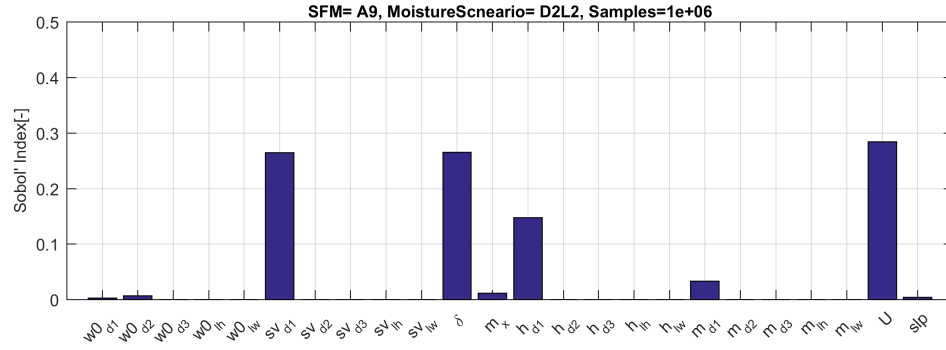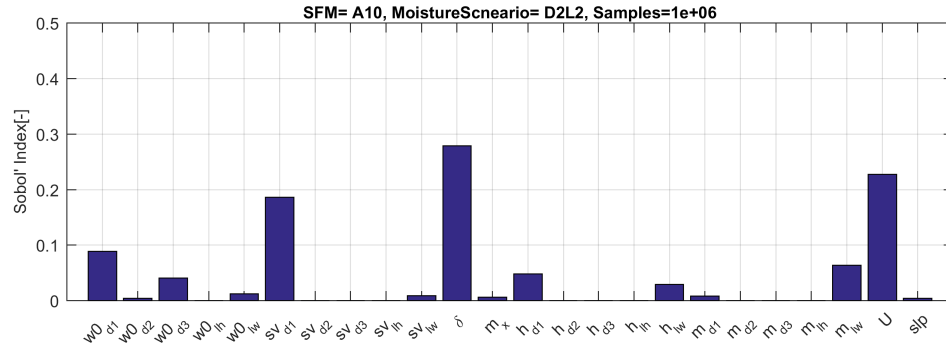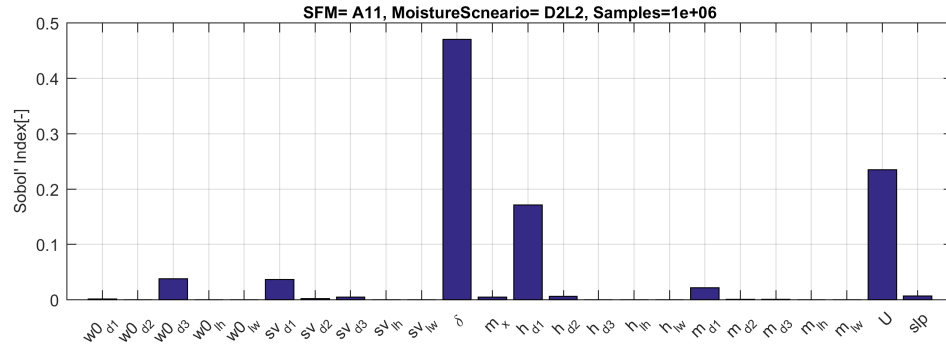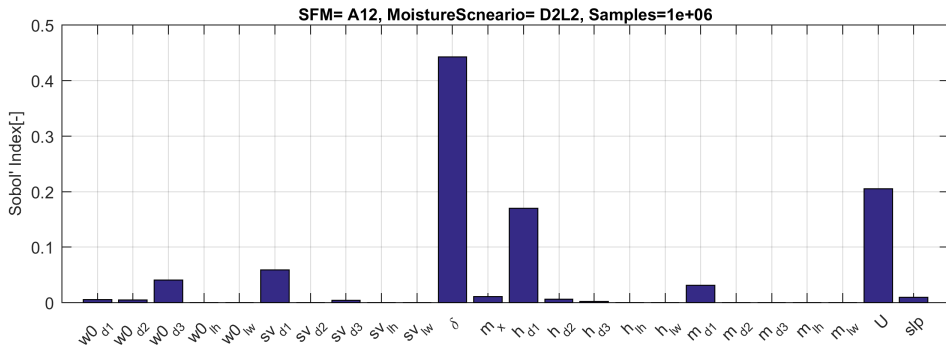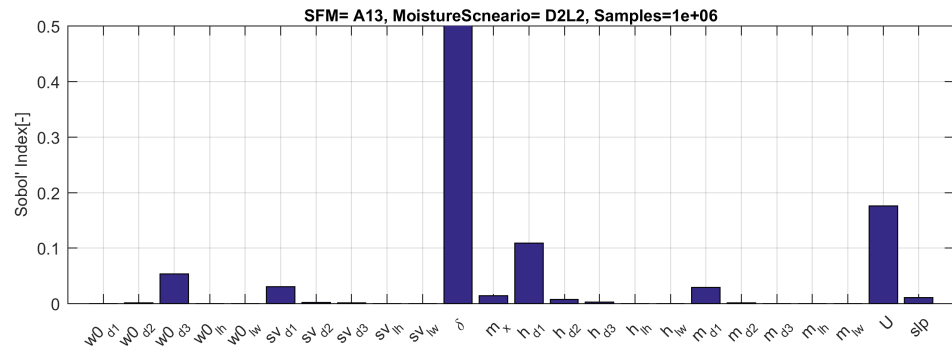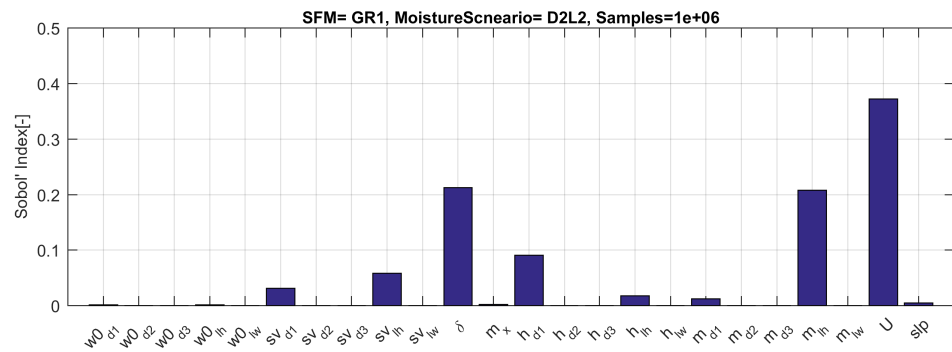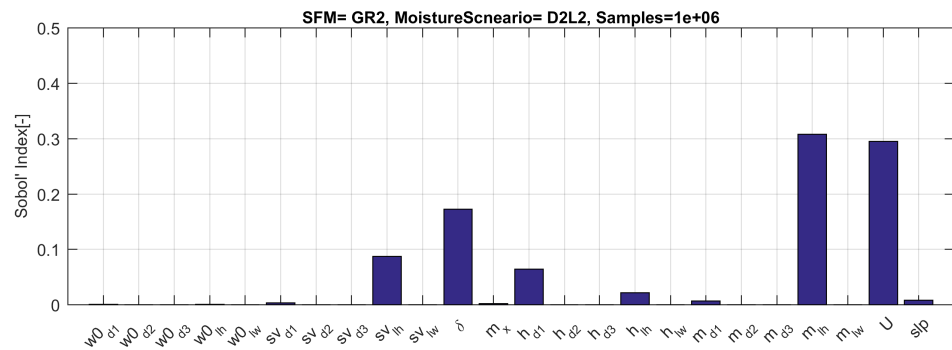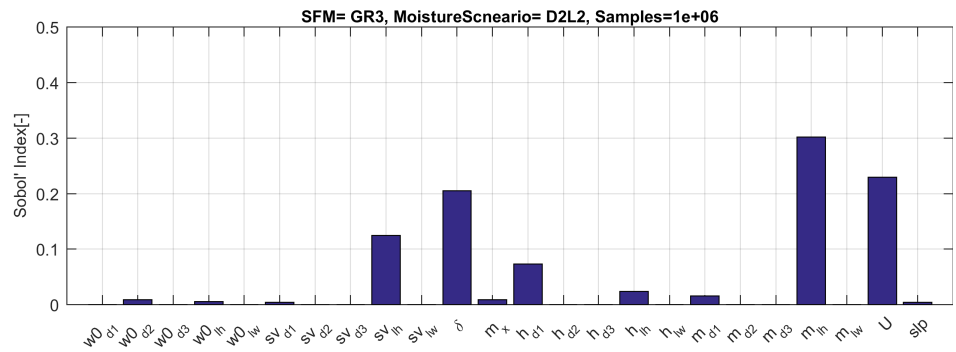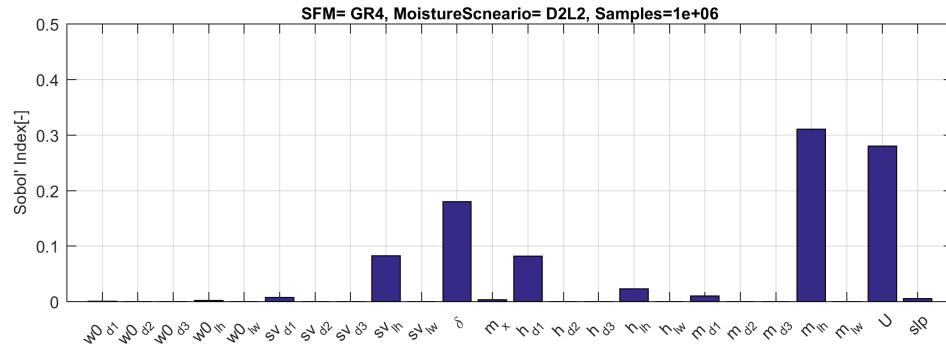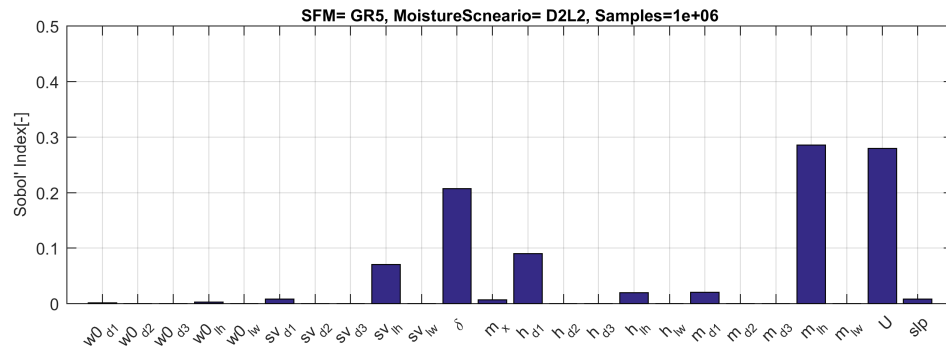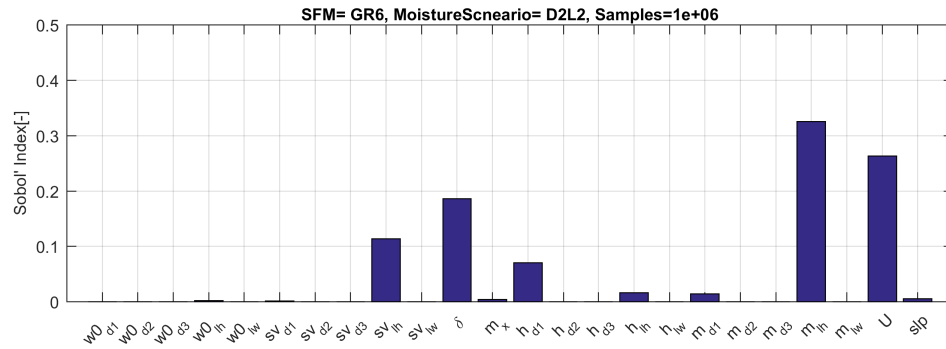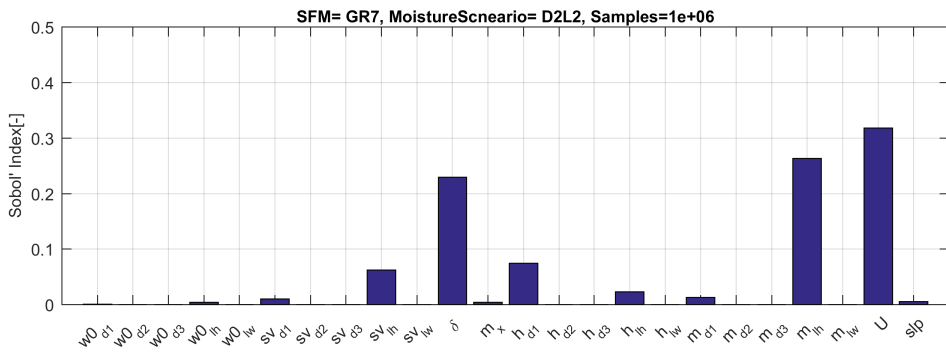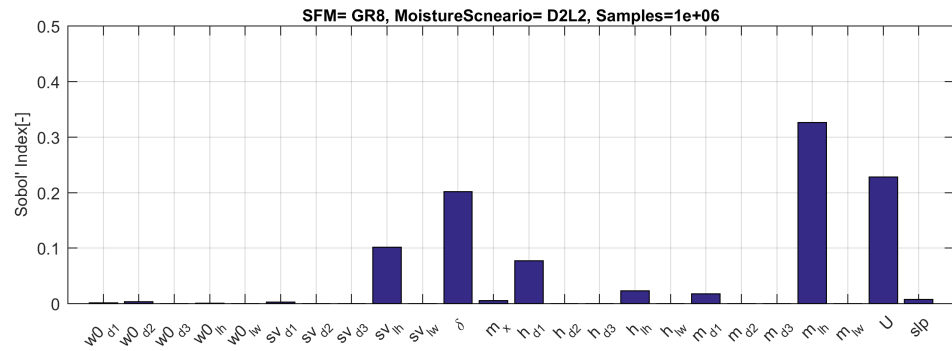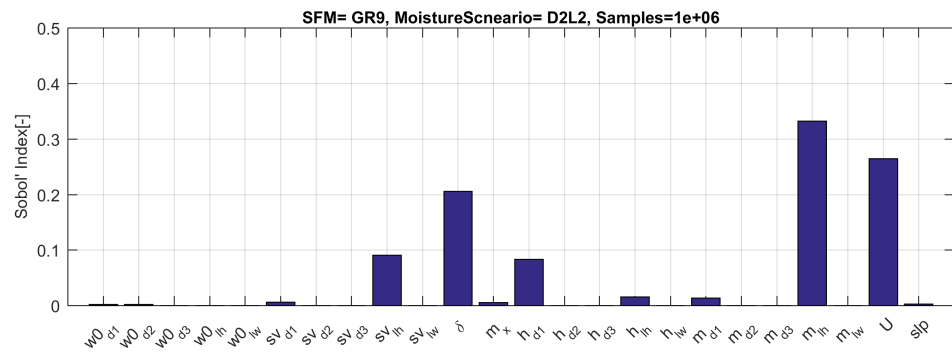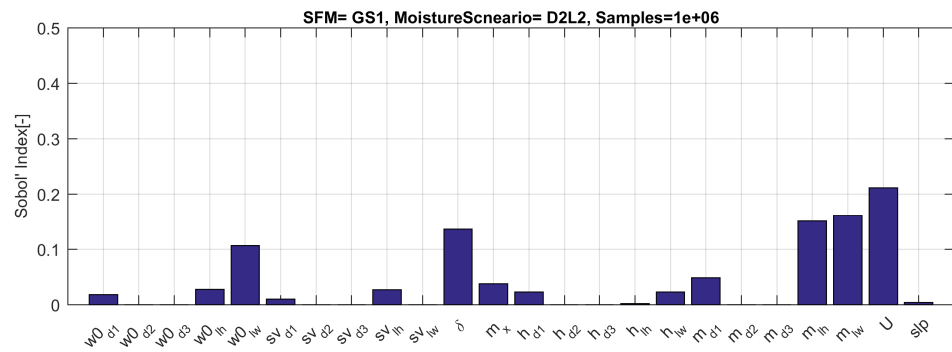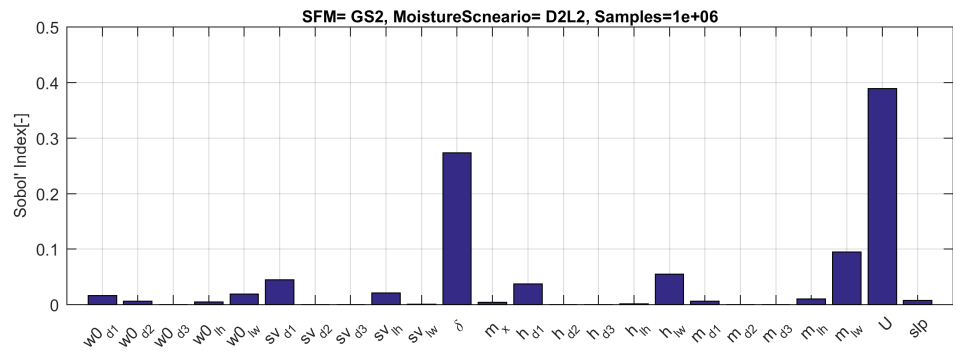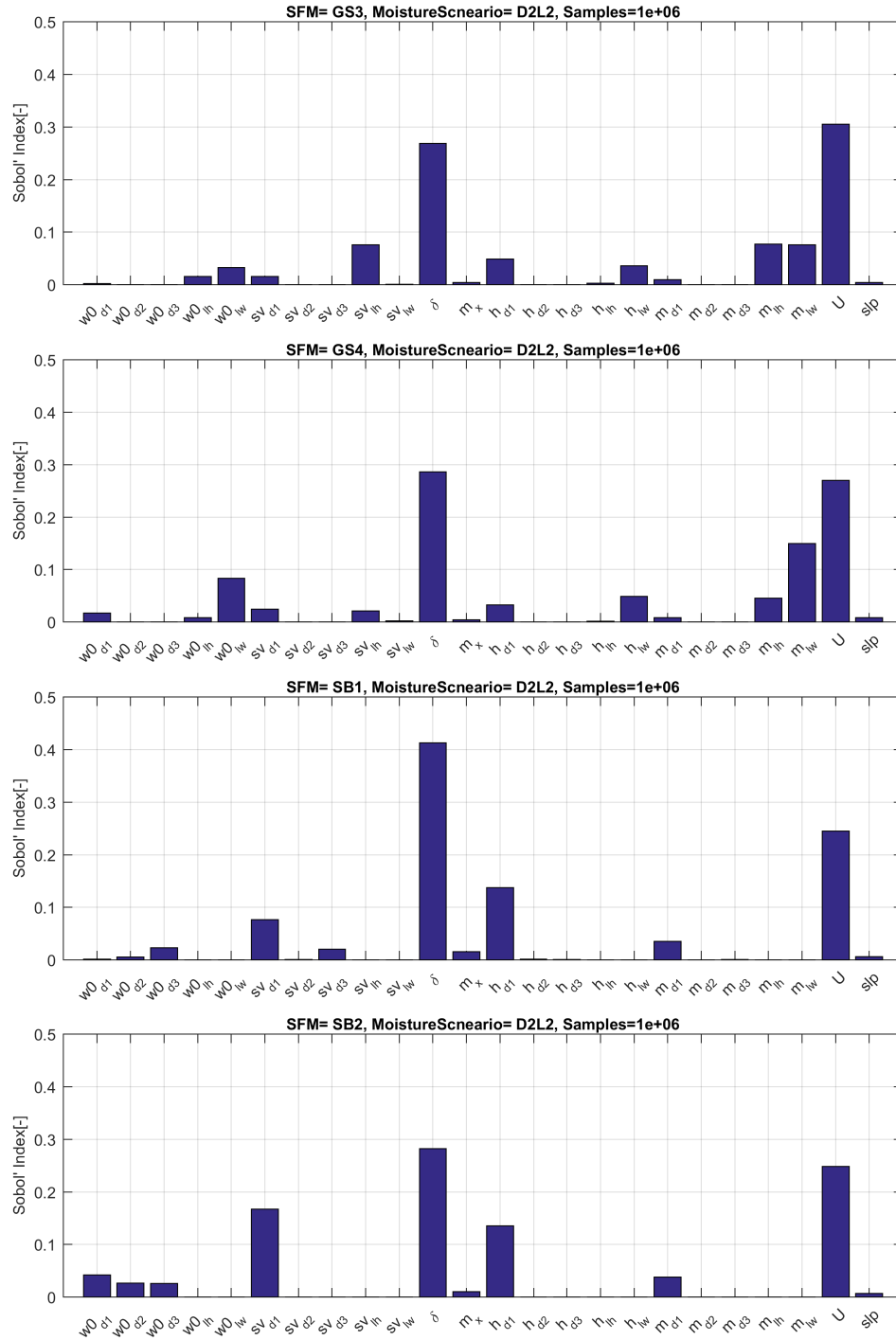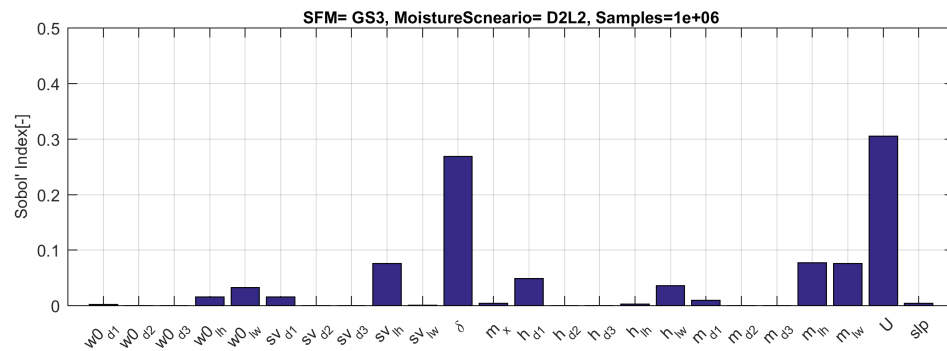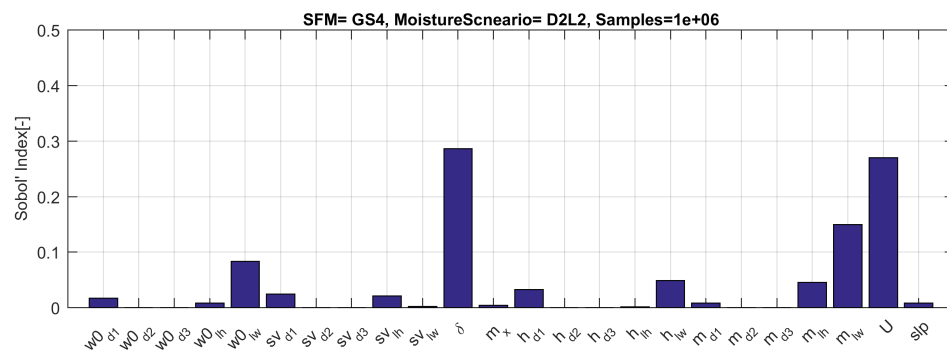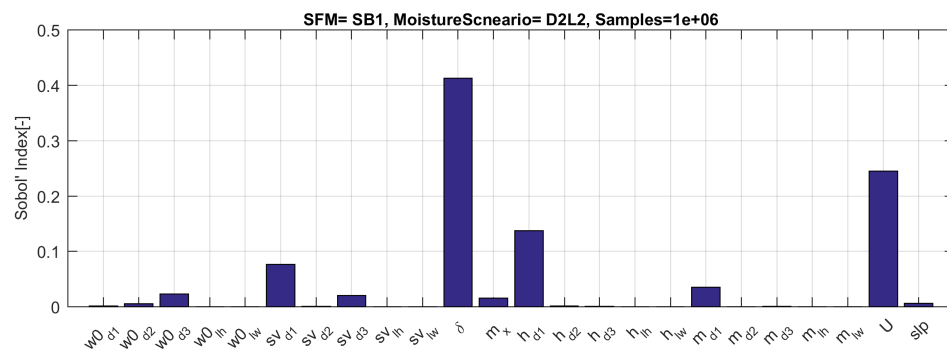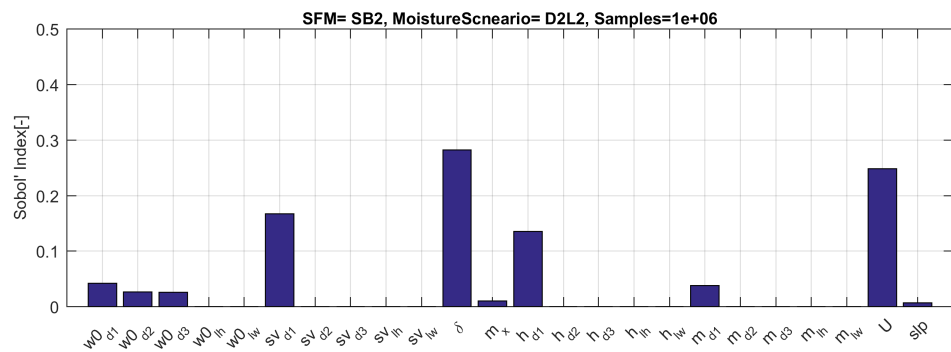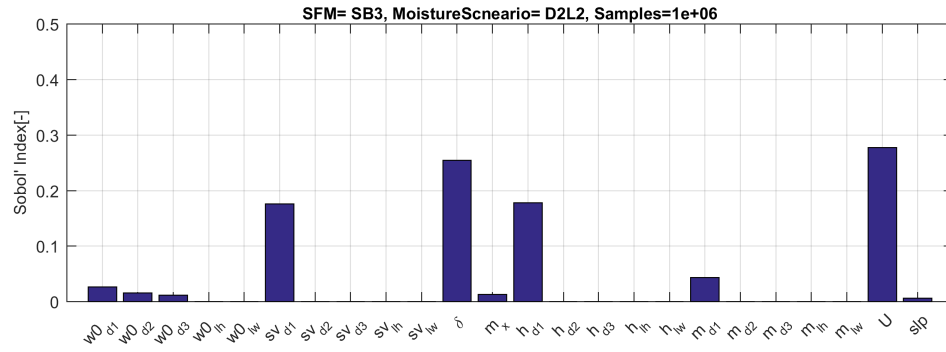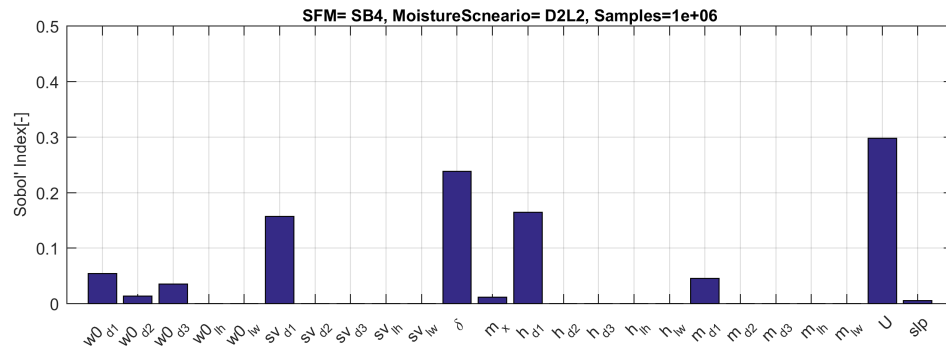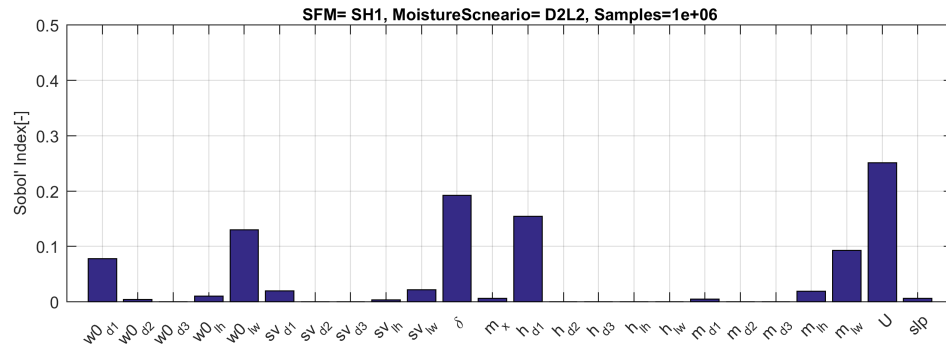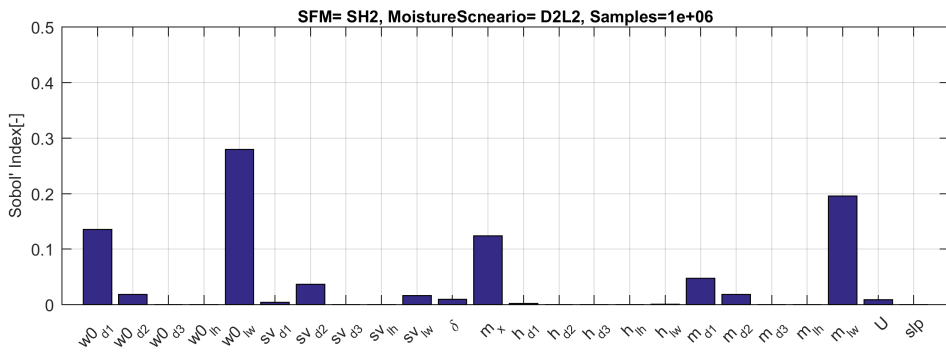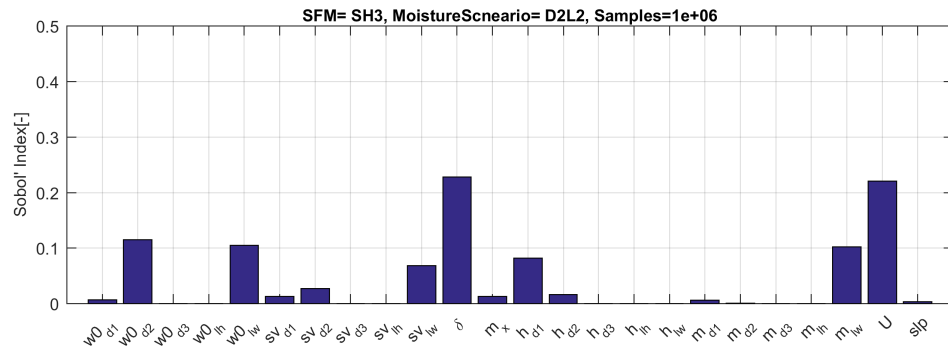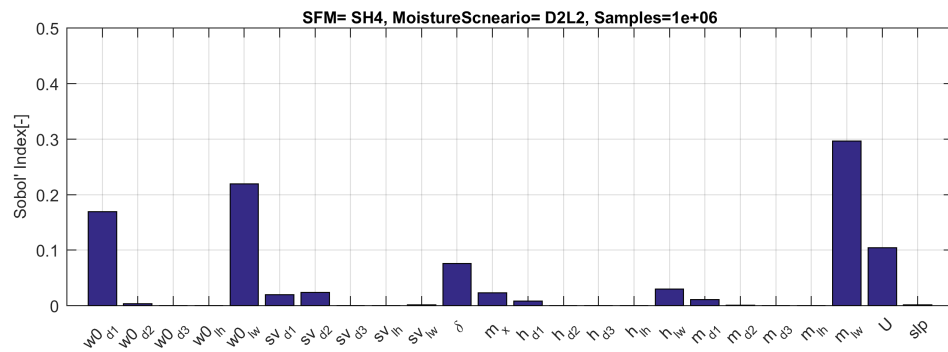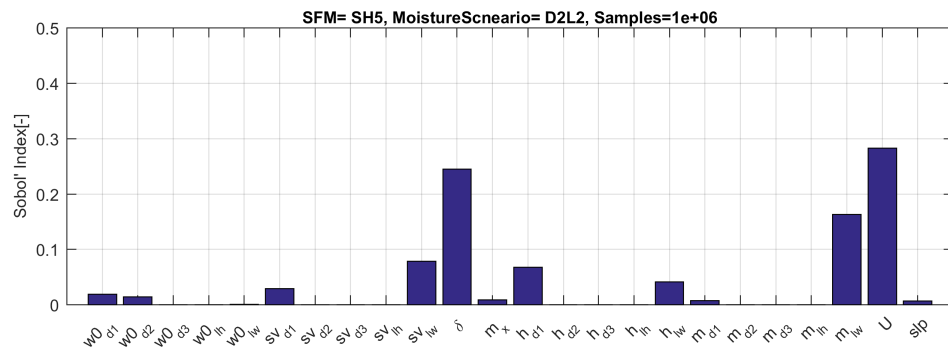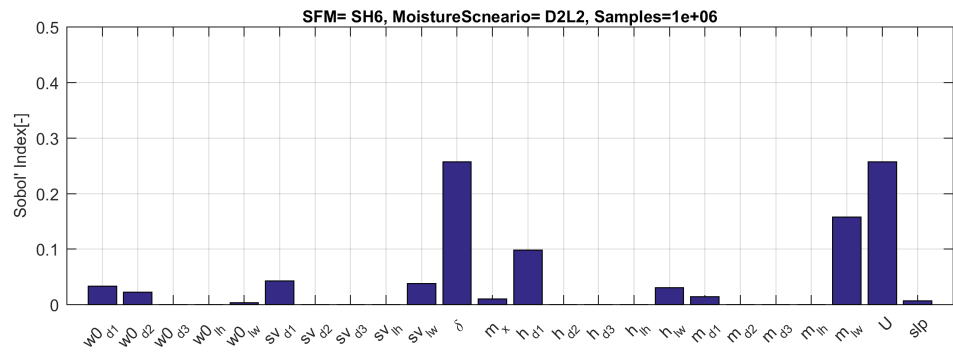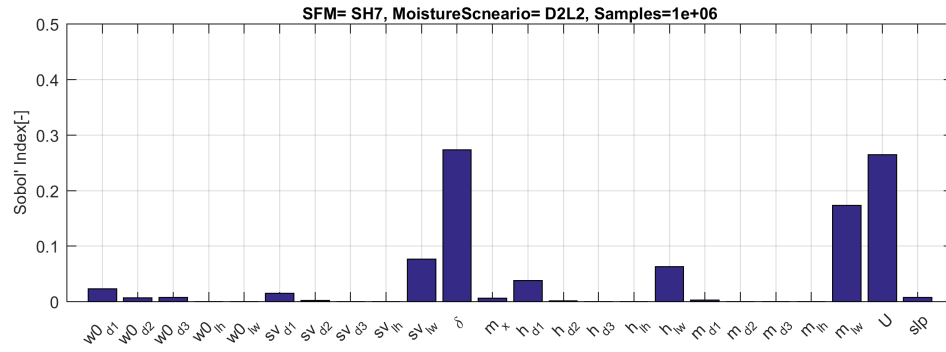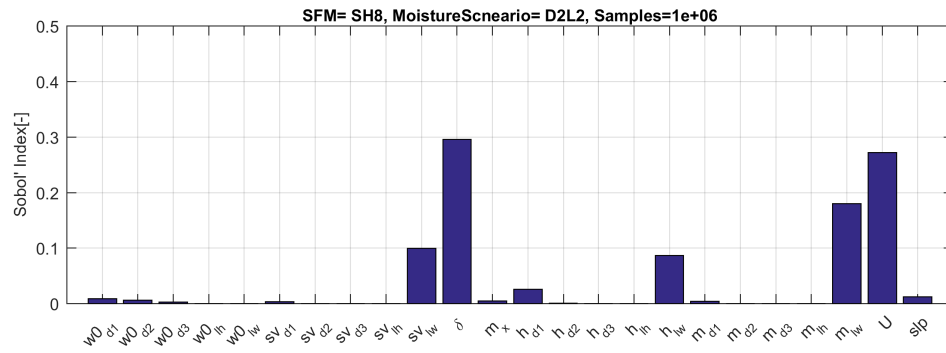
(b)
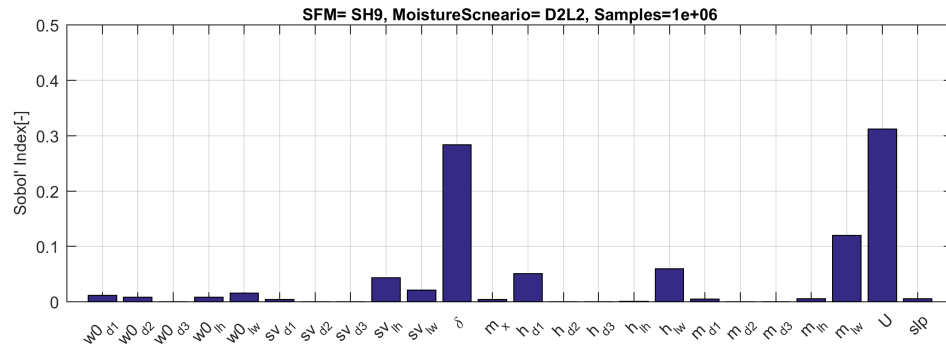
(c)
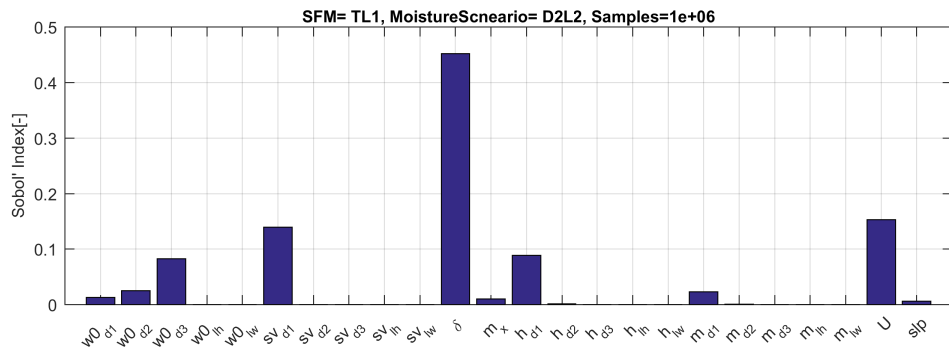
(d)
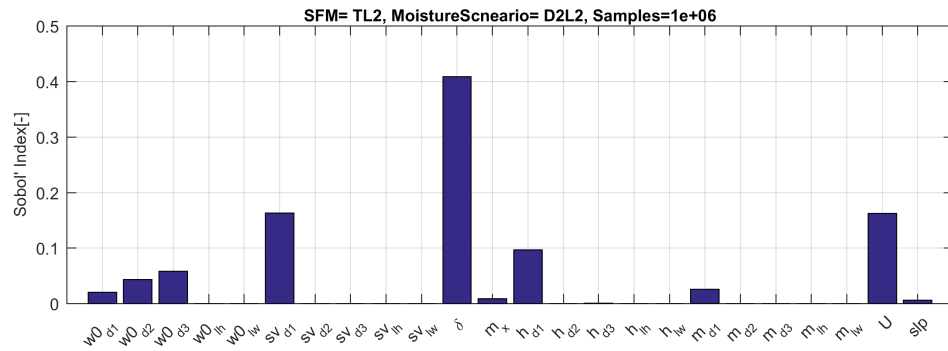
(a)
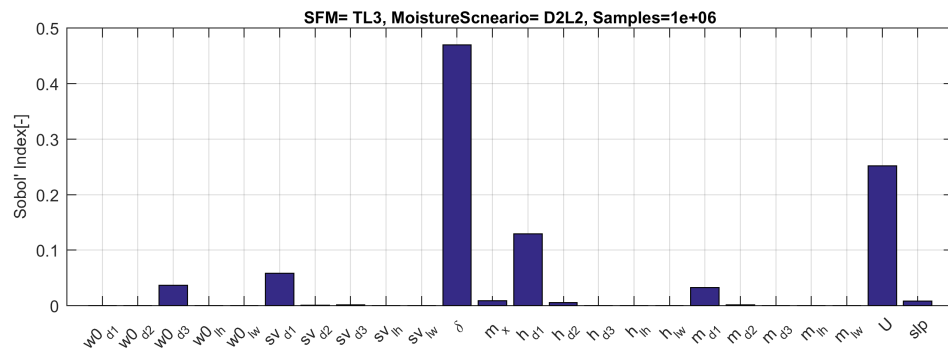
(b)
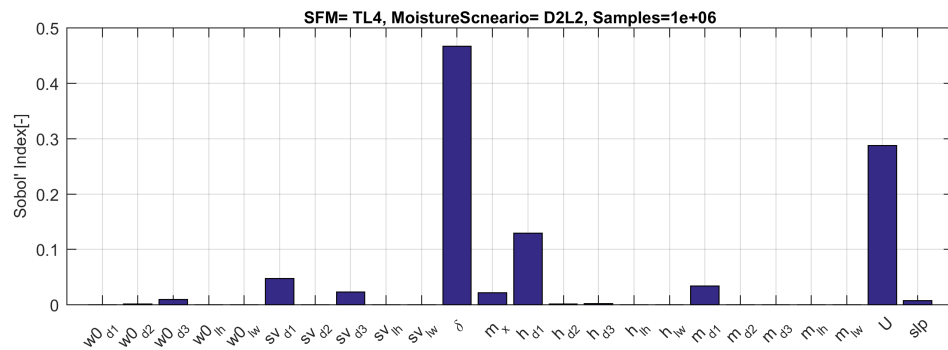
(c)
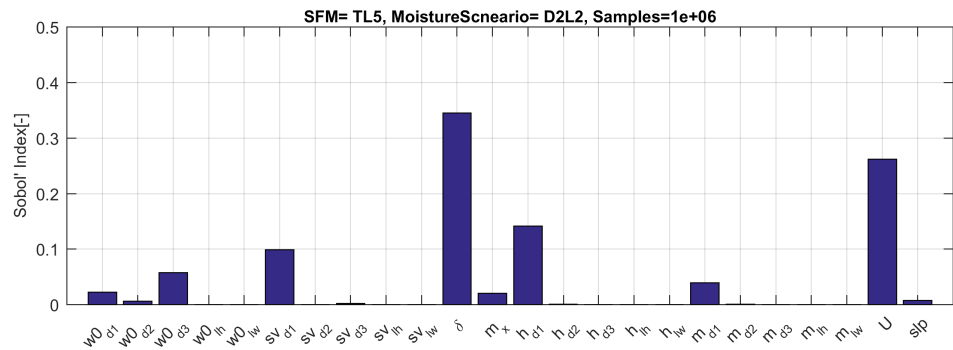
(d)
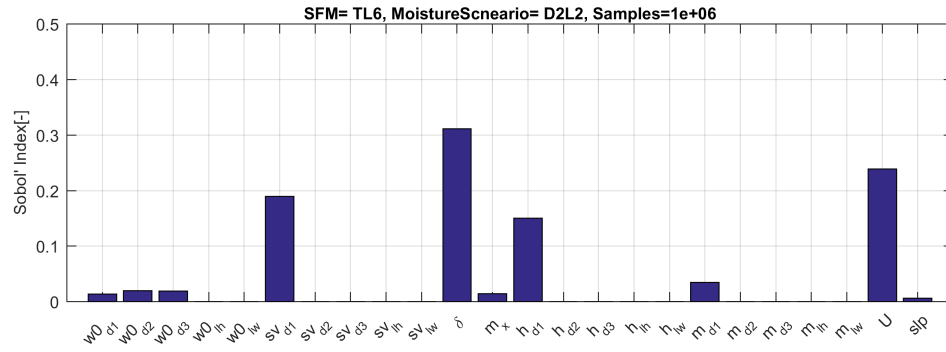
(a)
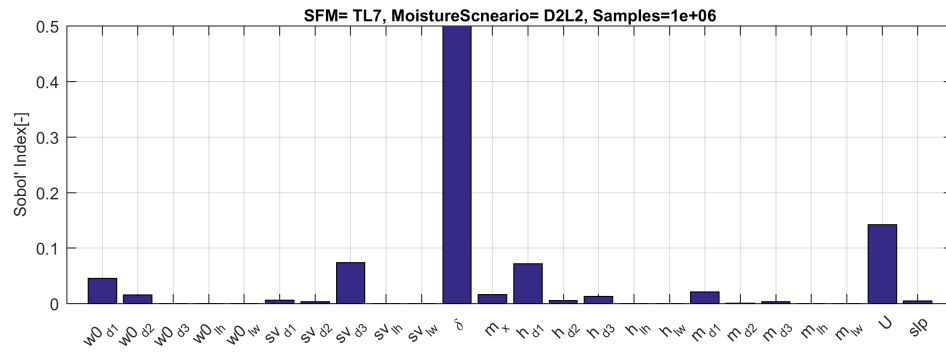
(b)
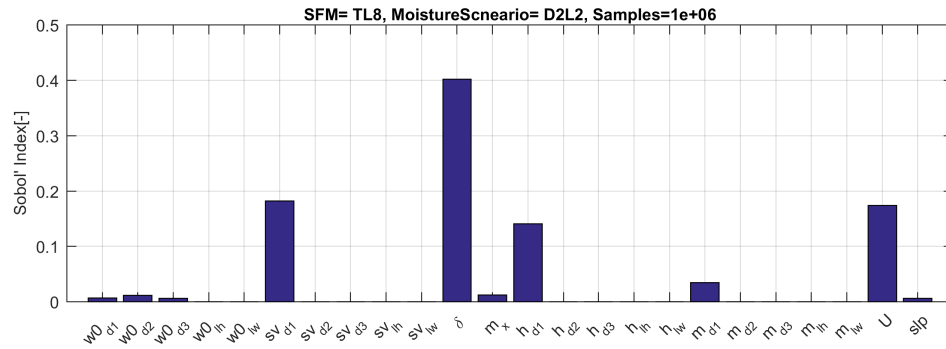
(c)
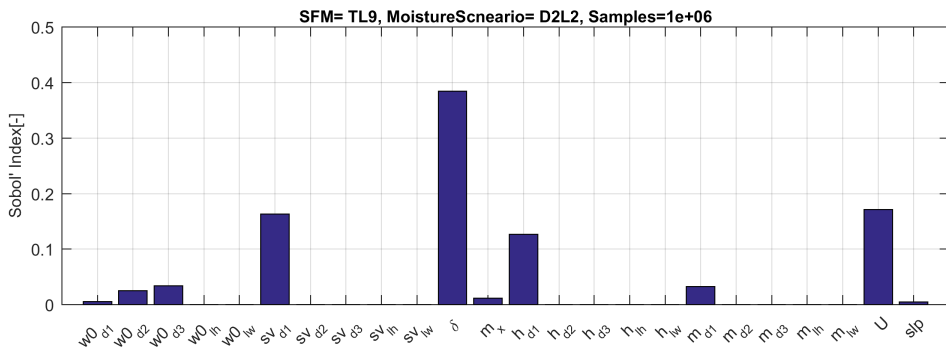
(d)
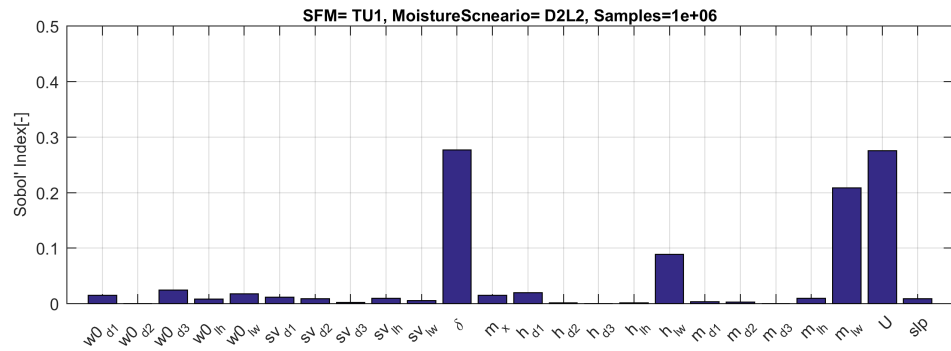
(a)
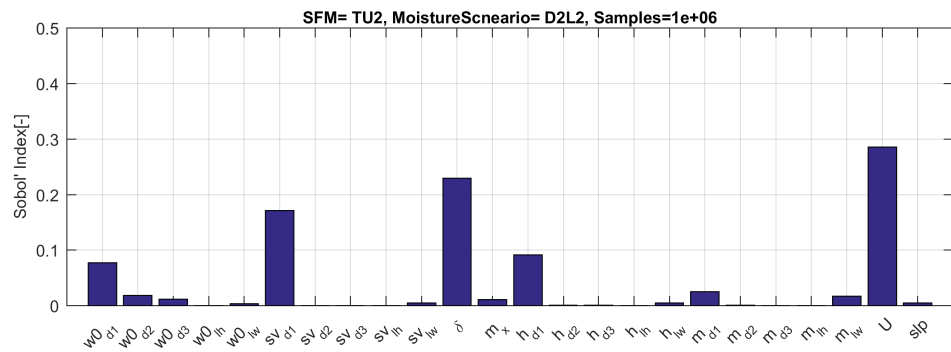
(b)
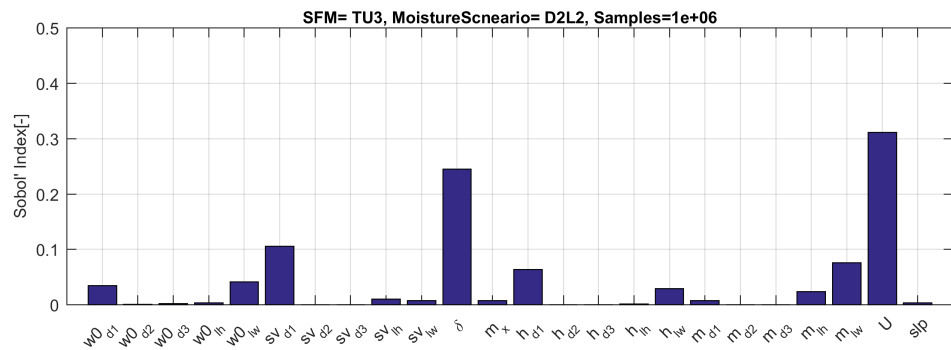
(c)
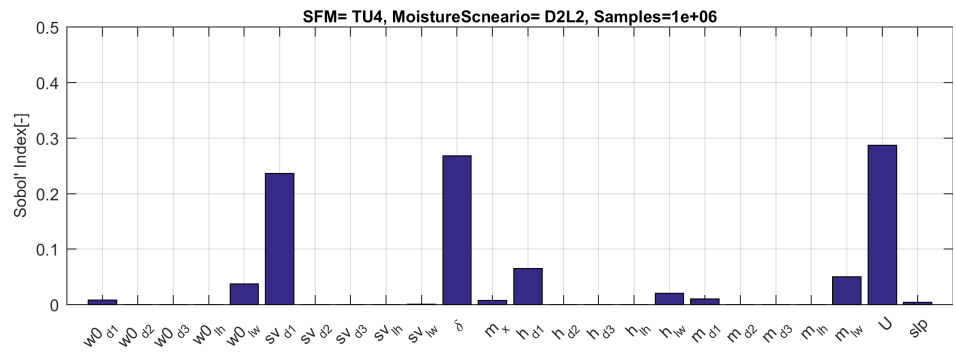
(d)
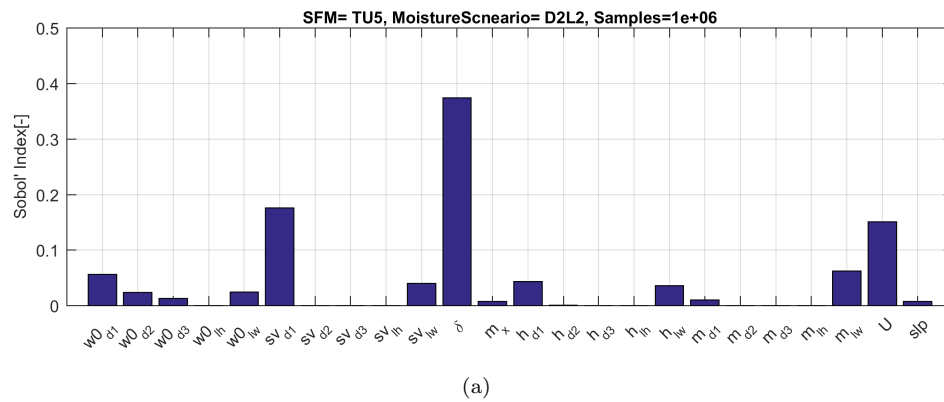
FIGURE A.2: All standard fuel models sobol index

Appendix B

Wind Ninja individual scenarios exploration

B.1 Annex: Additional Plots

To provide additional material, we present the mean averaged percentage error (MAPE) for same resolution scenarios as well as all the individual exploration.

B.1.1 Averaged Results

Mean absolute percentage error classified by DEM resolution for a domain averaged speed of 1 ms^{-1} and different directional sets (see captions).

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
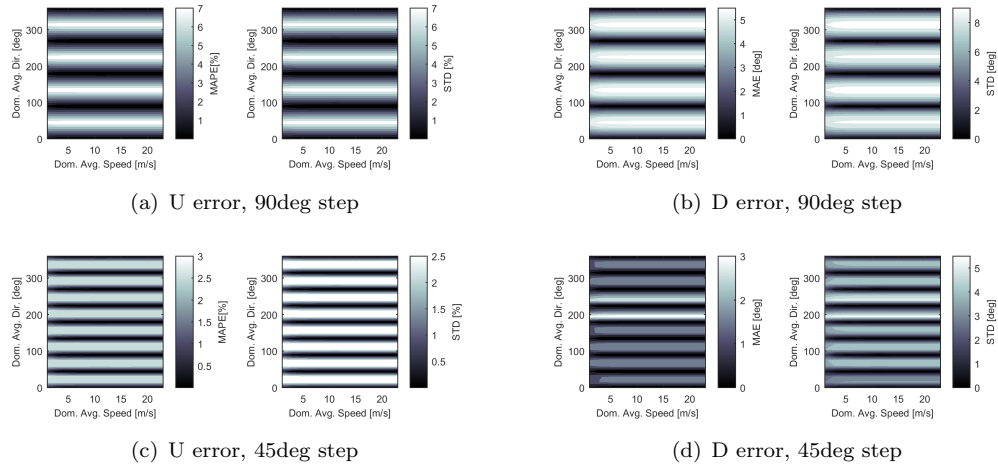
FIGURE B.1: MAPE for 90m resolution scenarios.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
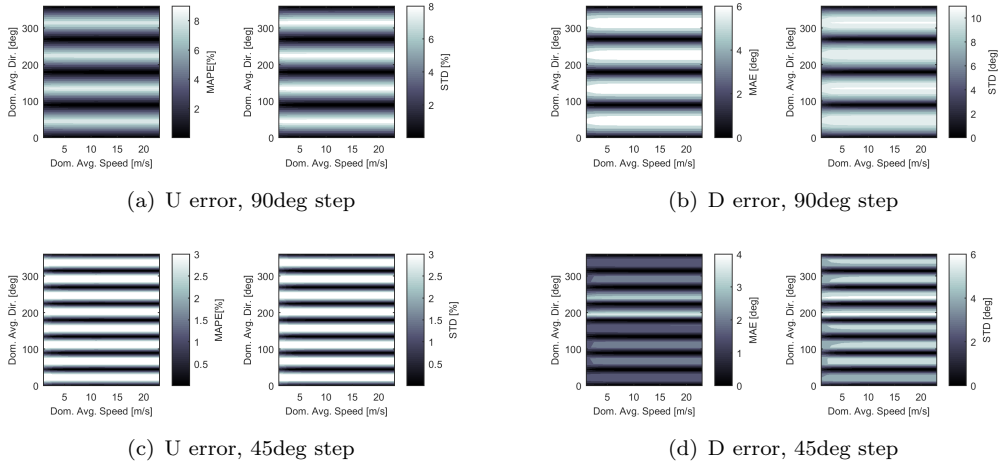
FIGURE B.2: MAPE for 30m resolution scenarios.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
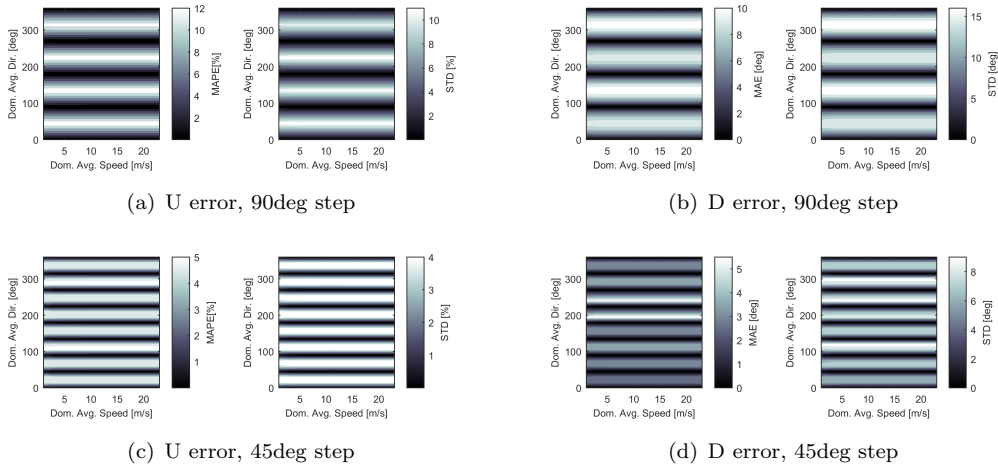
FIGURE B.3: MAPE for 15m resolution scenarios.

B.1.2 Individual scenarios

The results for different reference wind and direction configuration (see caption of each figure) are presented as contour filled plots for MAE and STD. For all cases, 24 directions (spaced 15 deg) and 12 wind speed (from 1 ms^{-1} to 23 ms^{-1} spaced by 2 ms^{-1}) are used to produce the contour plots.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
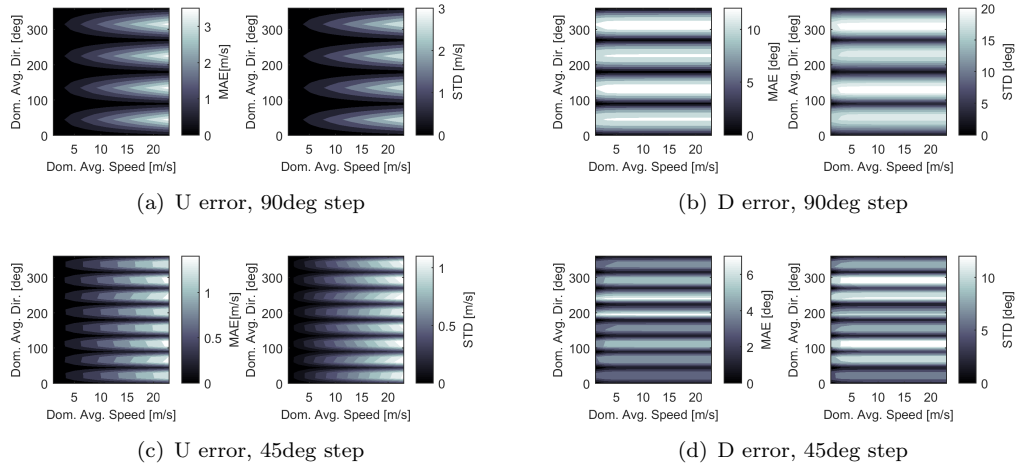
FIGURE B.4: Atlanta_90 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
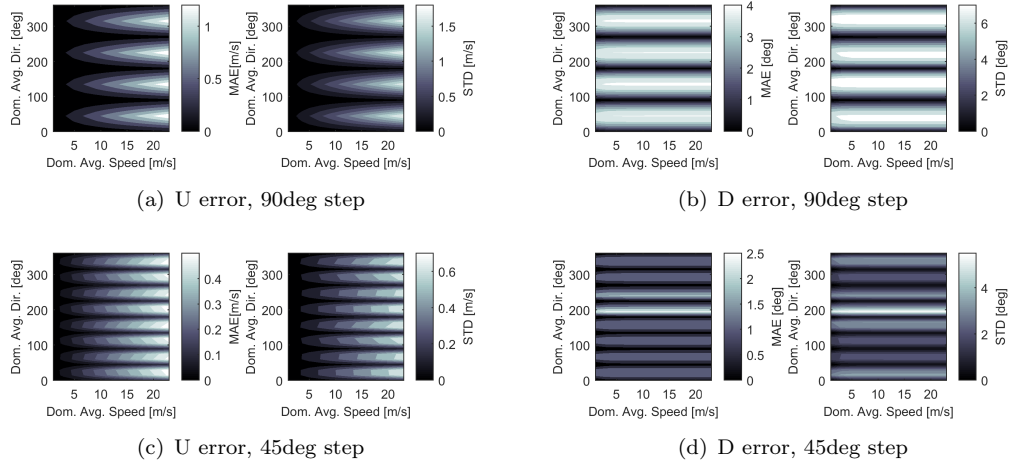
FIGURE B.5: Boulder_90 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
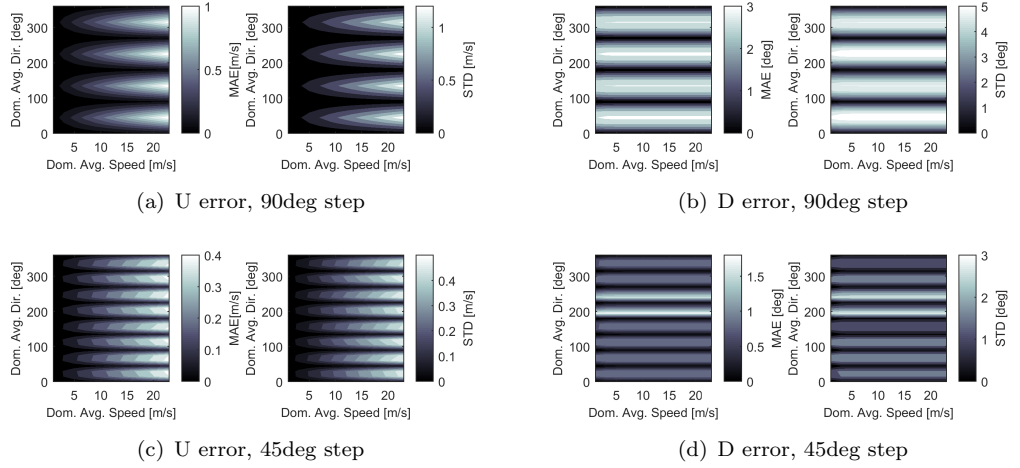
FIGURE B.6: Idaho_90 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
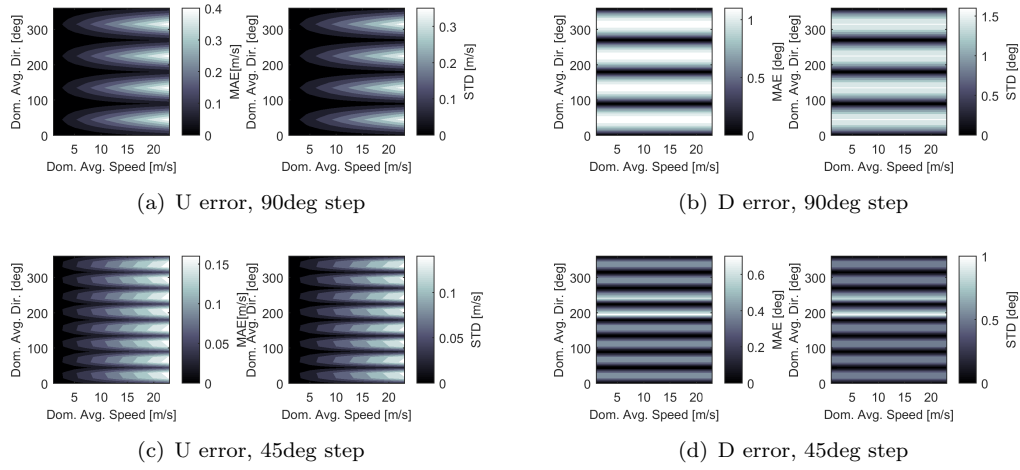
FIGURE B.7: Kansas_30 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
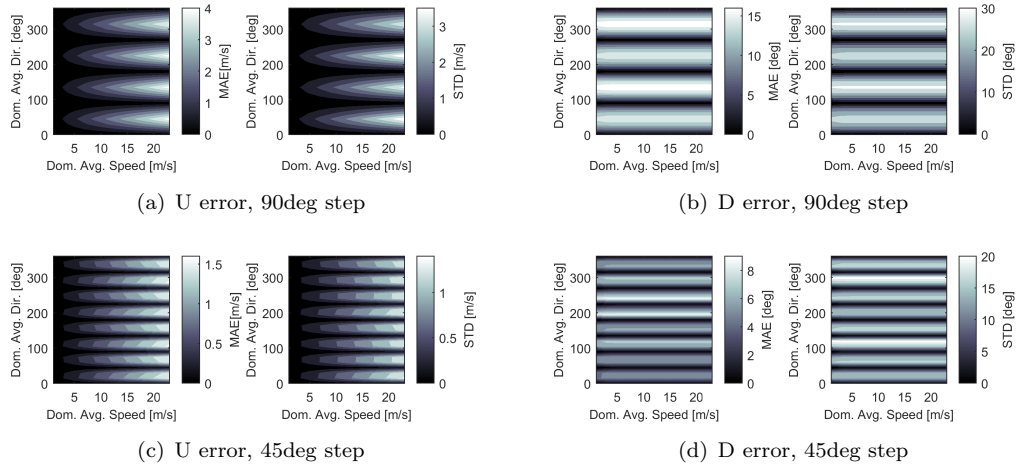
FIGURE B.8: Montseny_15 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
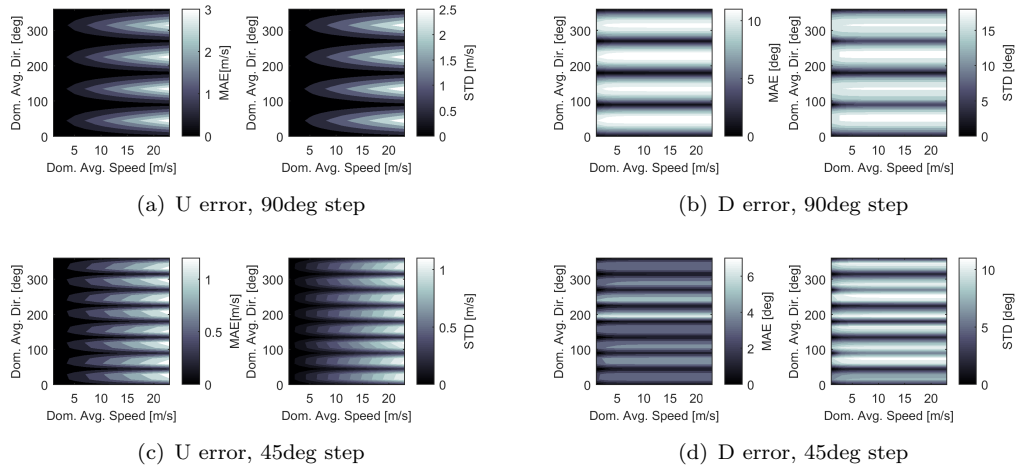
FIGURE B.9: Needham_90 scenario.

(a) U error, 90deg step

(b) D error, 90deg step

(c) U error, 45deg step

(d) D error, 45deg step
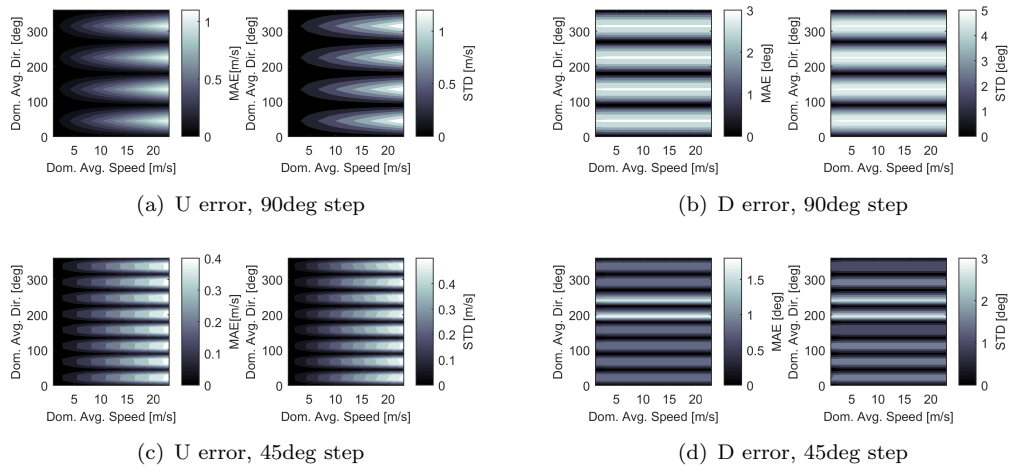
FIGURE B.10: Riverside_30 scenario.

B.2   Forecast performance fronts unfolding

. The Figure 4.13 in Chapter 4 compares 30 min of synthetic run data with two assimilation framework cases: (1) using WindNinja original wind maps, and (2) using the interpolating solution proposed. To better visualize this comparison, all isochrones are unfolded and plot against the synthetic data (observations) in Figures B.11 and B.12.
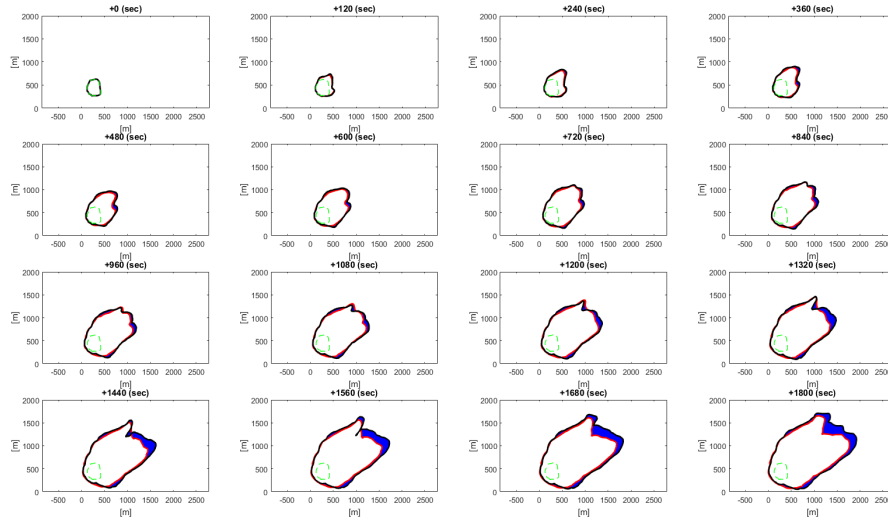
FIGURE B.11: WN interpolation vs WN original front-by-front areal comparison. Simulation performed with 90deg step set
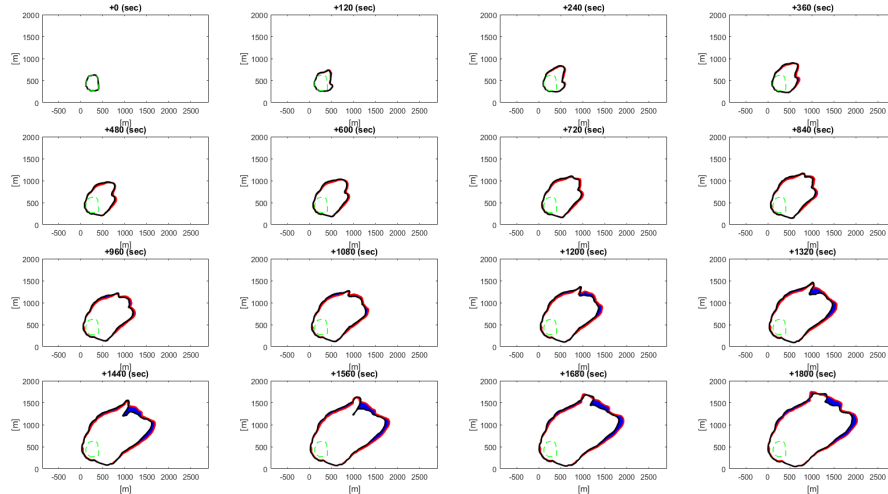
FIGURE B.12: WN interpolation vs WN original front-by-front areal comparison. Simulation performed with 45deg step set
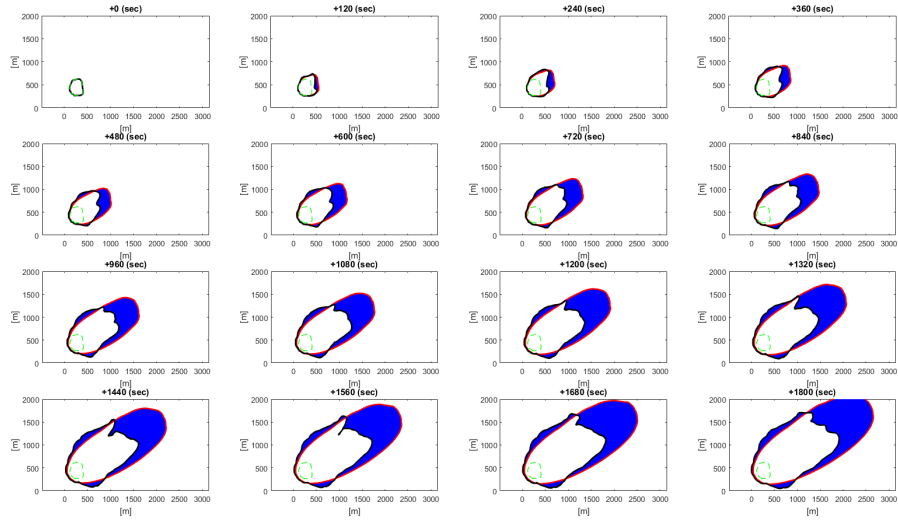
FIGURE B.13: Homogeneous wind vs WN original front-by-front areal comparison.

Appendix C

SmartQFIRE Scripts

All scripts used to develop the manuscript at hand are available at the following public repository:

   https://github.com/oriolrios/PhDThesis

List Of Publications

Published articles

- **Rios, O.**, Jahn, W., Valero, M.M., Pastor, E. and Planas, E., Framework to accurately interpolate downscaled wind simulations using WindNinja *International Journal of Wildland Fires* , 27(4), pp. 257-270, 2018,

- **Rios, O.** Pastor, E. Valero, M. Planas, E. Short-term fire front spread prediction using inverse modelling and airborne infrared images. *International Journal of Wildland Fires* 20(10), pp.1015–1032, 2016.

Other published articles

- Valero M., **Rios, O.**, Pastor E and Planas E, Automated mapping of active wildland fires through aerial infrared imaging and unsupervised edge detectors, *International Journal of Wildland Fires* , 27(4), pp. 241-256, 2017.

- Valero, M., **Rios, O.**, Mata C. Pastor, E. Planas, E. An integral system for automated tactical monitoring and data-driven spread forecasting of wildfires. *Fire Safety Science*, pp.1–10.

- **Rios, O.** Jahn, W. Rein, G. Forecasting wind-driven wildfires using an inverse modelling approach. *Natural Hazards and Earth System Science* 14(6) 1491–1503 2014.

Publications in preparation

- A data-driven fire spread simulator: validation in Vall-llobrega's fire, *Frontiers in Mechanical Engineering*, to be submitted on November 2018

Oral presentations

- **Rios, O.**, Valero, M., Pastor, E. Planas, Optimization strategy exploration in a wildfire propagation data driven system. *International Conference on Computational Science*, ICCS, 12-14 June 2017, Zurich, Switzerland

- **Rios, O.**, Valero, M., Pastor, E. Planas, A data-driven fire spread simulator and its validation in Vall-llobrega's fire. *World Conference on Natural Resource Modeling*, WCNRM, 06-09 June, 2017. Barcelona, Spain

- **Rios, O.** , E. Pastor, E. Planas, Short term forecasting of large scale wind-driven wildfires using thermal imaging and inverse modelling techniques, *VII International Conference on Forest Fire Research*. Coimbra, Portugal, December, 2014.

Contributions to oral presentations

- Valero, M., **Rios, O.**, Mata, C., Pastor, E., Planas, E. An integrated approach for tactical monitoring and data-driven spread forecasting of wildfires. *IAFSS 12th Symposium* 2017. Lund, Sweden 2017.

- Valero, M., Verstockt, S., **Rios, O.**, Pastor, E., Vandecasteele, F., Planas, E. Flame filtering and perimeter localization of wildfires using aerial thermal imagery. *Thermosense 2017: 39th Thermal Infrared Application*s: Anaheim, USA, April 10-13, 2017

- Rengel, F., Pastor, E., **Rios, O.**, Mata, C., Bartolomé, A., Planas, E. Assessment of the predictive capabilities of different modelling tools to forecast fire effects in residential compartments. *Research and advanced technology in fire safety*. Universidad de Cantabria. Santander 2017.

- Blond, J., **Rios, O.** Simulation of premixed flames with FDS. Application to the hot smoke testing system Izar. *Fire and Evacuation Modeling Technical Conference*, 16-18 Noviembre 2016. Malaga, Spain

- J. Fanlo, **Rios, O.**, M.M. Valero, E. Planas, X. Navalón, S. Escolano, E. Pastor. Designing CFD simulation scenarios for the assessment of dwellings vulnerability in the wildland urban interface. *International Conference on Forest Fires and WUI Fires*. Aix-en-Provence, France, Mayo 2016

- Pastor, E. Corberó, B. **Rios, O.** Giraldo, P. Haurie, L. Lacasta, A. Cuerva.E., Planas, E Compartment and Façade Large Scale Test: Behaviour comparison of different insulation materials in case of fire. *Applications of Structural Fire Engineering*, October 2015. Dubrovnik, Croatia

- Miralles, A., Pastor, E., Planas, E., **Rios, O.**, Muñoz, M., Tarrago, D., Martins, R. Toward the validation of CFD simulations for dispersion analysis: The USP-UPC experimental campaign. EMCHIE 2015: *7th European Meeting on Chemical Industry and Environment*, Tarragona, Spain: June 10-12, 2016

- Valero, M. **Rios, O.** Pastor, E. Planas, E. Automatic detection of wildfire active fronts from aerial thermal infrared image. *13th International Workshop on Advanced Infrared Technology and Applications*. Pisa, Italy. 2 Octubre 2015.

References

Albini, F. A. (1976a). Computer-based models of wildland fire behavior: A users' manual. Technical report, Intermountain forest and range experiment station, Ogden, Utah.

Albini, F. A. (1976b). Estimating wildfire behavior and effects. Technical report, USDA forest service, Odgen, USA.

Albini, F. a. (1982). Response of Free-Burning Fires to Nonsteady Wind. *Combustion Science and Technology*, 29(3-6):225–241.

Altintas, I., Block, J., De Callafon, R., Crawl, D., Cowart, C., Gupta, A., Nguyen, M., Braun, H. W., Schulze, J., Gollner, M., Trouve, A., and Smarr, L. (2015). Towards an integrated cyberinfrastructure for scalable data-driven monitoring, dynamic prediction and resilience of wildfires. *Procedia Computer Science*, 51(1):1633–1642.

Anderson, H. E. (1982). *Aids to determining fuel models for estimating fire behavior.* United States Department of Agriculture Forest Service, Ogden.

Andrews, P. L. (1984). Behave: Fire Behavior Prediction and Fuel Modeling System - FUEL Subsystem. Technical Report May 1984.

Andrews, P. L., Cruz, M., and Rothermel, R. C. (2013). Examination of the wind speed limit function in the Rothermel surface fire spread model. *International Journal of Wildland Fire*, 22(7):959.

Andrews, P. P. (1986). BEHAVE : Fire Behavior Prediction and Fuel Agriculture BEHAVE : Fire Behavior Prediction and Fuel Modeling. Technical report.

Arnaldos, J., Navalon, X., Pastor, E., Planas, E., and Zarate, L. (2004). *Manual de ingeniería básica para la prevención y extinción de incendios forestales.* Ediciones Mundi-Prensa.

Arroyo, L. A., Pascual, C., and Manzanera, J. A. (2008). Fire models and methods to map fuel types: The role of remote sensing. *Forest Ecology and Management*, 256(6):1239–1252.

Artés, T., Cencerrado, A., Cortés, A., Margalef, T., Rodríguez-Aseretto, D., Petroliagkis, T., and San-Miguel-Ayanz, J. (2014). Towards a dynamic data driven wildfire behavior prediction system at European level. *Procedia Computer Science*, 29:1216–1226.

Ascoli, D., Vacchiano, G., Motta, R., and Bovio, G. (2015). Building Rothermel fire behaviour fuel models by genetic algorithm optimisation. *International Journal of Wildland Fire*, 24(3):317.

Ash, M., Bocquet, M., and Nodet, M. (2016). *Data Assimilation: Methods, Algorithms, and Applications*. SIAM: Society of Industrial and Applied Mathematics.

Bachmann, A. and Allgöwer, B. (2002). Uncertainty propagation in wildland fire behaviour modelling. *International Journal of Geographical Information Science*, 16(2):115–127.

Barber, J., Bose, C., and Bourlioux, A. (2008). Burning issues with PROMETHEUS–the Canadian wildland fire growth simulation model. *Canadian Applied Mathematics Quarterly*, 16(4):337–378.

Beck, J. (2000). *Towards an Operational Geographical Information and Modelling System for Fire Management in Western Australia*. PhD thesis, Curtin University of Technology.

Beer, T. (1991). Bushfire rate-of-spread forecasting: Deterministic and statistical approaches to fire modelling. *Journal of Forecasting*, 10(3):301–317.

Bevins, C. D. (1996). FireLib User Manual & Technical Reference. Technical Report October, Systems for Environmental Management.

Bianchini, G., Cortés, A., Margalef, T., and Luque, E. (2005a). S2 F2 M - Statistical System for Forest Fire Management. *Computational Science – ICCS 2005*, pages 427–434.

Bianchini, G., Cortés, A., Margalef, T., Luque, E., Chuvieco, E., and Camia, A. (2005b). Wildland fire propagation danger maps based on factorial experimentation. *ITEE 2005 - 2nd International ICSC Symposium on Information Technologies in Environmental Engineering, Proceedings*, (April 2015):173–185.

Bianchini, G., Denham, M., Cortés, A., Margalef, T., and Luque, E. (2010). Wildland fire growth prediction method based on Multiple Overlapping Solution. *Journal of Computational Science*, 1(4):229–237.

Bischof, C., Bücker, H. M., Lang, B., Rasch, A., Vehreschild, A., Bucker, H., Lang, B., Rasch, A., and Vehreschild, A. (2002). Combining Source Transformation and Operator Overloading Techniques to Compute Derivatives for MATLAB Programs. In *Proceedings. Second IEEE International Workshop on Source Code Analysis and Manipulation*, pages 65–72, Los Alamitos, CA, USA. IEEE Comput. Soc.

Bischof, C., Büucker, H., and Vehreschild, A. (2006). A macro language for derivative definition in ADiMat. *Lecture Notes in Computational Science and Engineering*, 50:181–188.

Bova, A. S., Mell, W. E., and Chad, M. H. (2015). A comparison of level set and marker methods for the simulation of wildland fire front propagation. *International Journal of Wildland Fire*, 25(25):229–241.

Bowman, D., Balch, J., Artaxo, P., Bond, W., Carlson, J., Cochrane, M., D'Antonio, C., DeFries, R., Doyle, J., Harrison, S., Johnston, F., Keeley, J., Krawchuk, M., Kull, C., Marston, J. B., Moritz, M., Prentice, I. C., Roos, C., Scott, A., Swetnam, T. W., van der Werf, G., and Pyne, S. (2009). Fire in the Earth System. *Science*, 324(5926):481–484.

Bryce, R. and Dueck, G. (2009). Untangling the Prometheus Nightmare. In *18th World IMACS / MODSIM Congress*, number July, pages 240–246, Cairns, Australia.

Bücker, H. M. and Corliss, G. (2006). Automatic Differentiation: Applications, Theory, and Implementations. In *Lecture Notes in Computational Science and Engineering*.

Burgan, R. E. (1976). Estimating live fuel moisture for the 1978 National Fire-Danger Rating System. INT-226. Technical report, Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT: U.S.

Butler, B. W., Wagenbrenner, N. S., Forthofer, J. M., Lamb, B. K., Shannon, K. S., Finn, D., Eckman, R. M., Clawson, K., Bradshaw, L., Sopko, P., Beard, S., Jimenez, D., Wold, C., and Vosburgh, M. (2015). High-resolution observations of the near-surface wind field over an isolated mountain and in a steep river canyon. *Atmospheric Chemistry and Physics*, 15(7):3785–3801.

Byrd, R. H., Gilbert, J. C., Nocedal, J., Byrd, R. H., Gilbert, J. C., Nocedal, J., Region, A. T., and Based, M. (2000). A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming. 89:149–185.

Caton, S. E., Hakes, R. S. P., Gorham, D. J., Zhou, A., and Gollner, M. J. (2016). Review of Pathways for Building Fire Spread in the Wildland Urban Interface Part I: Exposure Conditions. *Fire Technology*.

Ching, J., Rotunno, R., LeMone, M., Martilli, A., Kosovic, B., Jimenez, P. A., and Dudhia, J. (2014). Convectively Induced Secondary Circulations in Fine-Grid Mesoscale Numerical Weather Prediction Models. *Monthly Weather Review*, 142(9):3284–3302.

Chuvieco, E., Riaño, D., Wagtendok, J. V., and Morsdof, F. (2003). Fuel Loads and Fuel Type Mapping. In Chuvieco, E., editor, *Wildland Fire Danger Estimation and Mapping*, chapter Fuel Loads, pages 119–142. World Scientific.

Clark, T. L., Coen, J., and Latham, D. (2004). Description of a coupled atmosphere-fire model. *International Journal of Wildland Fire*, 13(1):49–63.

Coleman, J. and Sullivan, A. L. (1996). A real-time computer application for the prediction of fire spread across the Australian landscape. *Simulation*.

Cruz, M. and Alexander, M. E. (2013). Uncertainty associated with model predictions of surface and crown fire rates of spread. *Environmental Modelling and Software*, 47:16–28.

Cruz, M., McCaw, W., Anderson, W., and Gould, J. (2013). Fire behaviour modelling in semi-arid mallee-heath shrublands of southern Australia. *Environmental Modelling & Software*, 40:21–34.

Cruz, M., Sullivan, A. L., and Gould, J. (2012). Anatomy of a catastrophic wildfire: the Black Saturday Kilmore East fire in Victoria, Australia. *Forest Ecology and Management*.

Cruz, M. G. and Fernandes, P. M. (2008). Development of fuel models for fire behaviour prediction in maritime pine (Pinus pinaster Ait.) stands. *International Journal of Wildland Fire*, 17(2):194–204.

Cui, W. and Perera, A. H. (2010). Quantifying Spatio-Temporal Errors in Forest Fire Spread Modelling Explicitly. *Journal of Environmental Informatics*, 16(1):19–26.

Davies, G. M., Gray, A., Rein, G., and Legg, C. J. (2013). Peat consumption and carbon loss due to smouldering wildfire in a temperate peatland. *Forest Ecology and Management*, 308:169–177.

De Sousa Dantas, A. (2017). Incêndio em Pedrógão Grande e Góis queimou quase 45 mil hectares.

Denham, M., Wendt, K., Bianchini, G., Cortés, A., and Margalef, T. (2012). Dynamic Data-Driven Genetic Algorithm for forest fire spread prediction. *Journal of Computational Science*, 3(5):398–404.

Ellyett, C. D. and Fleming, A. W. (1974). Thermal infrared imagery of The Burning Mountain coal fire. *Remote Sensing of Environment*, 3(1):79–86.

Fanlo, J., Valero, M. M., Rios, O., Planas, E., Navalon, X., Escolano, S., and Pastor, E. (2016). Designing CFD simulation scenarios for the assessment of dwellings vulnerability in the wildland urban interface. In *International Conference on Forest Fires and WUI Fires*, Aix-en-Provence, France.

Filippi, J. B., Bosseur, F., and Grandi, D. (2014a). ForeFire: open-source code for wildland fire spread models. In Viegas, D. X., editor, *Advances in forest fire research*. Imprensa da Universidade de Coimbra.

Filippi, J. B., Bosseur, F., Mari, C., Lac, C., Le Moigne, P., Cuenot, B., Veynante, D., Cariolle, D., and Balbi, J.-H. (2009). Coupled atmosphere–wildland fire modelling. *Journal of Advances in Modeling Earth Systems*, 1(11):9.

Filippi, J. B., Bosseur, F., Pialat, X., Santoni, P. A., Strada, S., and Mari, C. (2011). Simulation of coupled fire/atmosphere interaction with the MesoNH-ForeFire models. *Journal of Combustion*, 2011.

Filippi, J. B., Mallet, V., and Nader, B. (2014b). Evaluation of forest fire models on a large observation database. *Natural Hazards and Earth System Science*, 14(11):3077–3091.

Filippi, J. B., Mallet, V., and Nader, B. (2014c). Representation and evaluation of wildfire propagation simulations. *International journal of wildland fire*, (Fujioka 2002):46–57.

Finney, M. (1998). FARSITE: Fire area simulator: model development and evaluation. Technical report, USDA Forest Service, Ogden,UT: U.S.

Finney, M. (2006). An Overview of FlamMap Fire Modeling Capabilities. *Fuel Management-How to Measure Success: Conference Proceedings. 28-30 March*, pages 213–220.

Finney, M., Cohen, J. D., Forthofer, J. M., McAllister, S. S., Gollner, M. J., Gorham, D. J., Saito, K., Akafuah, N. K., Adam, B. a., and English, J. D. (2015). Role of buoyant flame dynamics in wildfire spread. *Proceedings of the National Academy of Sciences*, page 201504498.

Finney, M., Cohen, J. D., McAllister, S. S., and Jolly, W. M. (2013). On the need for a theory of wildland fire spread. *International Journal of Wildland Fire*, 22(1):25.

Finney, M. A. (1999). Spatial Modeling of Post-Frontal Fire Behavior. *Final Report RMRS-99557-RJVA, Missoula*, (Anderson 1983):8.

Forthofer, J., Shannon, K., and Butler, B. (2009). Simulating diurnally driven slope winds with WindNinja. *Proceedings of 8th Symposium on Fire and Forest Meteorological Society*, page 13.

Forthofer, J. M. (2007). *Modeling wind in complex terrain for use in fire spread prediction*. PhD thesis, Colorado State University.

Forthofer, J. M., Butler, B. W., and Wagenbrenner, N. S. (2014). A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. Part I. Model formulation and comparison against measurements. *International Journal of Wildland Fire*, 23(7):969–981.

Georg, J., Spessa, A., Werf, G. V. D., and Thonicke, K. (2013). *Vegetation Fires and Global Change – Challenges for Concerted International Action*. House, Kessel Publishing.

Glasspool, I. J., Edwards, D., and Axe, L. (2004). Charcoal in the Silurian as evidence for the earliest wildfire. *Geology*, 32(5):381–383.

Gollner, M. and Trouvé, A. (2015). Towards Data-Driven Operational Wildfire Spread Modeling. In *A report of the nsf-funded WIFIRE workshop*, number January, UCSD.

GRAF (2014). Incendis, estratègies i ecologia del foc forestal. Nº 159. 2014: Incendis tot l'any, llamps i pluja a l'estiu i cremes a la tardor. Technical Report 159, Bombers de Catalunya. Unitat tècnica GRAF.

Griewank, A. (2000). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Number 19 in Frontiers in Appl. Math. SIAM, Philadelphia, PA.

Hakes, R. S., Caton, S. E., Gorham, D. J., and Gollner, M. J. (2017). A Review of Pathways for Building Fire Spread in the Wildland Urban Interface Part II: Response of Components and Systems and Mitigation Strategies in the United States. *Fire Technology*, 53(2):475–515.

Homicz, G. (2002). Three-dimensional wind field modeling: a review. *Sandia National Laboratories, Albuquerque, SANDIA*, (August).

IGCC (2018). Perímetres del cercador d'incendis forestals. Technical report.

Jaccard, P. (1901). *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz.

Jahn, W. (2010). *Inverse modelling to forecast enclosure fire dynamics*. PhD thesis, University of Edinburgh.

Jahn, W., Rein, G., and Torero, J. L. (2012). Forecasting fire dynamics using inverse computational fluid dynamics and tangent linearisation. *Advances in Engineering Software*, 47(1):114–126.

Jimenez, E., Hussaini, M. Y., and Goodrick, S. (2008). Quantifying parametric uncertainty in the Rothermel model. *International Journal of Wildland Fire*, 17(5):638–649.

Jolly, W. M., Cochrane, M. a., Freeborn, P. H., Holden, Z. a., Brown, T. J., Williamson, G. J., and Bowman, D. M. J. S. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications*, 6(May):7537.

Knight, I. and Coleman, J. (1993). A Fire Perimeter Expansion Algorithm-Based on Huygens Wavelet Propagation. *International Journal of Wildland Fire*, 3(2):73.

Kolda, T., Lewis, R., and Torczon, V. (2003). Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods. *SIAM Review*, 45(3):385–482.

Koo, E., Pagni, P. J., Weise, D. R., and Woycheese, J. P. (2010). Firebrands and spotting ignition in large-scale fires. *International Journal of Wildland Fire*, 19(7):818–843.

Kuhn, H. W. and Tucker, A. (1951). Nonlinear Programming. *Proceedings of the Second Symposium on Mathematical Statistics and Probability*, (x):481–492.

Lautenberger, C. (2013). Wildland fire modeling with an Eulerian level set method and automated calibration. *Fire Safety Journal*, 62:289–298.

Linn, R. (1997). *A transport model for prediction of wildfire behavior*. PhD thesis, Los Alamos National Laboratory.

Linn, R. and Harlow, F. (1998). FIRETEC: a transport description of wildfire behavior. In *Conference: 78. American Meteorological Society annual meeting*, Phoenix, AZ.

Liu, Y., Hussaini, M. Y., and Ökten, G. (2013). Global sensitivity analysis for the Rothermel model based on high-dimensional model representation. In *Proceedings of 4th Fire Behavior and Fuels Conference*, Raleigh, North Carolina, USA.

Liu, Y., Hussaini, M. Y., and Ökten, G. (2015a). Global sensitivity analysis for the Rothermel model based on high-dimensional model representation. *Canadian Journal of Forest Research*, 45(11):1474–1479.

Liu, Y., Jimenez, E., Hussaini, M. Y., Ökten, G., and Goodrick, S. (2015b). Parametric uncertainty quantification in the Rothermel model with randomised quasi-Monte Carlo methods. *International Journal of Wildland Fire*, 24(3):307.

Lopes, A. (2003). WindStation—a software for the simulation of atmospheric flows over complex topography. *Environmental Modelling & Software*, 18:81–96.

Lopes, A. M. G., Cruz, M. G., and Viegas, D. X. (2002). Firestation - An integrated software system for the numerical simulation of fire spread on complex topography. *Environmental Modelling and Software*, 17(3):269–285.

Mallet, V., Keyes, D. E., and Fendell, F. E. (2009). Modeling wildland fire propagation with level set methods. *Computers & Mathematics with Applications*, 57(7):1089–1101.

Mandel, J., Amram, S., Beezley, J. D., Kelman, G., Kochanski, a. K., Kondratenko, V. Y., Lynn, B. H., Regev, B., and Vejmelka, M. (2014). Recent advances and applications of WRF-SFIRE. *Natural Hazards and Earth System Science*, 14(10):2829–2845.

Mandel, J., Beezley, J. D., Coen, J. L., and Kim, M. (2009). Data assimilation for wildland fires. *Control Systems, IEEE*, 29(3):47–65.

Mandel, J., Beezley, J. D., and Kochanski, a. K. (2011). Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development*, 4(3):591–610.

Manzano-Agugliaro, F., Pérez-Aranda, J., and De La Cruz, J. L. (2014). Methodology to obtain isochrones from large wildfires. *International Journal of Wildland Fire*, 23(3):338.

Manzello, S. L. and Suzuki, S. (2013). Experimentally simulating wind driven firebrand showers in wildland-urban interface (WUI) fires: Overview of the NIST firebrand generator (NIST dragon) technology. *Procedia Engineering*, 62:91–102.

MathWorks (2016). Optimization Algorithms in Matlab.

McArthur, A. (1966). Weather and grassland fire behaviour.

McGrattan, K. and Miles, S. (2016). Modeling Fires Using Computational Fluid Dynamics (CFD). In *SFPE Handbook of Fire Protection Engineering*, pages 1034–1065. Springer New York, New York, NY.

McRae, R. (2004). Virtually volvelles. In *Proceedings, Bushfire 2004*, Adelaide. Department of Environment and Heritage, South Australia.

Mell, W., Jenkins, M., Gould, J., and Cheney, P. (2007). A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire*, 16(1):1.

Mendez, R. (2009). Un insólito incendio subterráneo azota las Tablas de Daimiel.

Merrill, D. F. and Alexander, M. E. (2003). Glossary of forest fire management terms. Technical report, Canadian Interagency Forest Fire Centre.

Modugno, S., Balzter, H., Cole, B., and Borrelli, P. (2016). Mapping regional patterns of large forest fires in Wildland-Urban Interface areas in Europe. *Journal of Environmental Management*, 172:112–126.

Monedero, S., Buckley, D., and Ramírez, J. (2011). New approaches in fire simulations analysis with Wildfire Analyst. In *The 5th International Wildland Fire Conference Sun City, South Africa*, number May, pages 1–17.

Moreno, L., Jiménez, M. E., Aguilera, H., Jiménez, P., and de la Losa, A. (2011). The 2009 Smouldering Peat Fire in Las Tablas de Daimiel National Park (Spain). *Fire Technology*, 47(2):519–538.

Nelson, R. M. (2002). An effective wind speed for models of fire spread. *International Journal of Wildland Fire*, 11(2):153–161.

NIFC (2013). Federal firefighting costs (suppression only).

Noble, I., Bary, G., and Gill, A. (1980). McArthur's fire-danger meters expressed as equations. *Australian Journal of Ecology*, 5(2):201–203.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York.

Onwubolu, G. and Babu, B. (2013). *New optimization techniques in engineering.* Springer.

Overholt, K. J., Cabrera, J., Kurzawski, A., Koopersmith, M., and Ezekoye, O. A. (2012). Characterization of Fuel Properties and Fire Spread Rates for Little Bluestem Grass. *Fire Technology*, pages 1–30.

Papadopoulos, G. D. and Pavlidou, F. N. (2011). A comparative review on wildfire simulators. *Systems Journal, IEEE*, 5(2):233–243.

Pastor, E. (2004). *Contribució a l'estudi dels efectes dels retardants en l'extinció d'incendis forestals.* PhD thesis.

Pastor, E., Àgueda, A., Andrade-Cetto, J., Muñoz, M., Pérez, Y., and Planas, E. (2006). Computing the rate of spread of linear flame fronts by thermal image processing. *Fire Safety Journal*, 41(8):569–579.

Pastor, E., Zárate, L., Planas, E., and Arnaldos, J. (2003). Mathematical models and calculation systems for the study of wildland fire behaviour. *Progress in Energy and Combustion Science*, 29(2):139–153.

Pausas, J. G. and Keeley, J. E. (2009). A Burning Story: The Role of Fire in the History of Life. *BioScience*, 59(7):593–601.

Pérez, Y., Pastor, E., Planas, E., Plucinski, M., and Gould, J. (2011). Computing forest fires aerial suppression effectiveness by IR monitoring. *Fire Safety Journal*, 46(1-2):2–8.

Planas, E., Pastor, E., Cubells, M., Cruz, M., and Greenfell, I. (2011a). Fire behaviour variability in mallee-heath shrubland fires. *5th International Wildland Fire . . .*, (May).

Planas, E., Pastor, E., Cubells, M., Cruz, M., and Greenfell, I. (2011b). Fire behaviour variability in mallee-heath shrubland fires. In *5th International Wildland Fire Conference*, 9-13 May. Sun City, South Africa.

Plucinski, M. and Pastor, E. (2013). Criteria and methodology for evaluating aerial wildfire suppression. *International Journal of Wildland Fire*, 22(8):1144–1154.

Pugnet, L., Chong, D. M., Duff, T. J., and Tolhurst, K. (2013). Wildland–urban interface (WUI) fire modelling using PHOENIX Rapidfire: A case study in Cavaillon, France. *20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013*, (December):1–6.

Rein, G. (2013). Smouldering Fires and Natural Fuels. In Belcher, editor, *Fire Phenomena and the Earth System: An Interdisciplinary Guide to Fire Science*, number 1984, chapter Fire Pheno, pages 15–33. Wiley and Sons.

Rein, G. (2015). Smoldering-Peat Megafires: The Largest Fires on Earth. In Strache, G. B., Prakash, A., and Rein, G., editors, *Coal and Peat Fires: A Global Perspective*, volume 4, chapter 1, pages 1–11. Elsevier.

Richards, G. (1990). An elliptical growth model of forest fire fronts and its numerical solution. *International Journal for Numerical Methods in Engineering*, 30(6):1163–1179.

Richards, G. (2000). The mathematical modelling and computer simulation of wildland fire perimeter growth over a 3-dimensional surface. *International Journal of Wildland Fire*, 9(3):213–221.

Richards, G. and Bryce, R. (1995). A Computer Algorithm for Simulating the Spread of Wildland Fire Perimeters for Heterogeneous Fuel and Meteorological Conditions. *International Journal of Wildland Fire*, 5(2):73.

Richards, G. D. (1993). The properties of elliptical wildfire growth for time dependent fuel and meteorological conditions. *Combustion science and technology*, 95(1-6):357–383.

Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293.

Rios, O. (2013). *Forecasting Wind-Driven Wildfires Using An Inverse Modelling Approach*. PhD thesis, University of Edinburgh.

Rios, O., Jahn, W., Pastor, E., Valero, M. M., and Planas, E. (2018). Interpolation framework to speed up near-surface wind simulations for data-driven wildfire applications. *International Journal of Wildland Fire*, 27(4):257–270.

Rios, O., Pastor, E., Tarragó, D., Rein, G., Planas, E., and Viegas, D. X. (2014a). Short term forecasting of large scale wind-driven wildfires using thermal imaging and inverse modelling techniques. In *Advances in forest fire research*, pages 949–960, Coimbra. Imprensa da Universidade de Coimbra.

Rios, O., Pastor, E., Valero, M., and Planas, E. (2016). Short-term fire front spread prediction using inverse modelling and airborne infrared images. *International Journal of Wildland Fire*, 20(10):1015–1032.

Rios, O., Valero, M., Pastor, E., and Planas, E. (2017a). A data-driven fire spread simulator: validation in Vall-llobrega's fire. In *WCNR2017 - World Conference on Natural Resource Modeling*, 6-9 June, Barcelona.

Rios, O., Valero, M., Pastor, E., and Planas, E. (2017b). Optimization strategy exploration in a wildfire propagation data driven system. In *International Conference on Computational Science, ICCS 2017*, 12-14 June, Zurich.

Rios, O., W., J., and Rein, G. (2014b). Forecasting wind-driven wildfires using an inverse modelling approach. *Natural Hazards and Earth System Science*, 14(6):1491–1503.

Rochoux, M. (2014). *Vers une meilleure prévision de la propagation d ' incendies de forêt : évaluation de modèles et assimilation de données*. PhD thesis.

Rochoux, M., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A. (2014a). Towards predictive simulation of wildfire spread at regional scale using ensemble-based data assimilation to correct the fire front position. In *Fire Safety Science – Proc. Eleventh International Symposium*, page 217. International Association for Fire Safety Science.

Rochoux, M., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A. (2015). Towards predictive data-driven simulations of wildfire spread – Part II: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread. *Natural Hazards and Earth System Science*, 15:1721–1739.

Rochoux, M., Emerya, C., and Riccia, S. (2014b). A Comparative Study of Parameter Estimation and State Estimation Approaches in Data–Driven Wildfire Spread Modeling. In Viegas, D., editor, *VII International Conference on Forest Fire Research*, number Rochoux, pages 1–12.

Rochoux, M., Ricci, S., Lucor, D., Cuenot, B., and Trouvé, A. (2014c). Towards predictive data-driven simulations of wildfire spread – Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation. *Natural Hazards and Earth System Science*, 14(11):2951–2973.

Ronchi, E., Rein, G., Gwynne, S., Wadhwani, R., Intini, P., and Bergstedt, A. (2017). Open MultiPhysics Framework for Modelling Wildfire Urban Evacuation - Fire Protection Research Foundation. Technical Report December, NFPA, Report: FPRF-2017-22.

Rossa, C. G., Davim, D. A., and Viegas, D. X. (2015). Behaviour of slope and wind backing fires. *International Journal of Wildland Fire*, 24:1085–1097.

Rothermel, R. (1972a). A mathematical model for predicting fire spread in wildland fuels. Research Paper INT-115. USDA Forest Service. Intermountain Forest and Range Experiment Station. Odgen, USA. Technical report, USDA Forest Service. Intermountain Forest and Range Experiment Station, Odgen, USA.

Rothermel, R. C. (1972b). A mathematical model for predicting fire spread in wildland fuels. Technical report, USDA Forest Service, Ogden, Utah.

Rothermel, R. C. (1991). Predicting behavior and size of crown fires in the northern Rocky Mountains. Technical Report January.

Rudz, S., Chetehouna, K., Séro-Guillaume, O., Pastor, E., and Planas, E. (2009). Comparison of two methods for estimating fire positions and the rate of spread of linear flame fronts. *Measurement Science and Technology*, 20(11):115501.

Salvador, R., Piol, J., Tarantola, S., and Pla, E. (2001). Global sensitivity analysis and scale effects of a fire propagation model used over Mediterranean shrublands. *Ecological Modelling*, 136(2-3):175–189.

Sanjuan, G., Margalef, T., and Cortés, A. (2016a). Applying domain decomposition to wind field calculation. *Parallel Computing*, 57:484–490.

Sanjuan, G., Margalef, T., and Cortés, A. (2016b). Hybrid application to accelerate wind field calculation. *Journal of Computational Science*.

Sanjuan, G., Tena, C., Margalef, T., and Cortés, A. (2016c). Applying vectorization of diagonal sparse matrix to accelerate wind field calculation. *Journal of Supercomputing*, pages 1–19.

Scott, J. and Burgan, R. (2005a). Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. Gen. Tech. Rep. RMRS-GTR-153. USDA Forest Service. Rocky Mountain Research Station. Fort Collins, USA. Technical report, USDA Forest Service. Rocky Mountain Research Station, Fort Collins, USA.

Scott, J. H. and Burgan, R. E. (2005b). Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. *The Bark Beetles, Fuels, and Fire Bibliography*, page 66.

Scott, J. H. and Burgan, R. E. (2005c). Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. Technical report.

Seaman, N. L., Gaudet, B. J., Stauffer, D. R., Mahrt, L., Richardson, S. J., Zielonka, J. R., and Wyngaard, J. C. (2012). Numerical Prediction of Submesoscale Flow in the Nocturnal Stable Boundary Layer over Complex Terrain. *Monthly Weather Review*, 140(3):956–977.

Sharples, J. J. (2008). Review of formal methodologies for windslope correction of wildfire rate of spread. *International Journal of Wildland Fire*, 17(2):179–193.

Sharples, J. J., McRae, R. H. D., and Wilkes, S. R. (2012). Wind - terrain effects on the propagation of wildfires in rugged terrain: fire channelling. *International Journal of Wildland Fire*, 21(3):282.

Sobol', I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates I.M. *Mathematics and Computers in Simulation*, 5517:271–280.

Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118.

Sobol', I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S., and Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering and System Safety*, 92(7):957–960.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, 5:1–34.

Srivas, T., Artes, T., de Callafon, R., and Altintas, I. (2016). Wildfire Spread Prediction and Assimilation for FARSITE Using Ensemble Kalman Filtering. *Procedia Computer Science*, 80:897–908.

Statistics Canada (2016). Fort McMurray 2016 Wildfire economic impact. Technical report, statcan.gc.ca.

Sullivan, A. L. (2009a). Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18(4):349–368.

Sullivan, A. L. (2009b). Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire*, 18(4):369–386.

Sullivan, A. L. (2009c). Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire*, 18(4):387–403.

Sullivan, A. L., Sharples, J. J., Matthews, S., and Plucinski, M. P. (2014). A downslope fire spread correction factor based on landscape-scale fire behaviour. *Environmental Modelling and Software*, 62:153–163.

Tarantola, A. (2005). *Inverse Problem Theory*. SIAM.

Taylor, S., Pike, R., and Alexander, M. (1996). Field Guide to the Canadian Forest Fire Behavious Prediction (FBP) System.

Team, R. C. (2008). R: A Language and Environment for Statistical Computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria.

The MathWorks, I. (2016). MATLAB and Parallel Toolbox Release 2016b,.

Tolhurst, K., Shields, B., and Chong, D. M. (2008). Phoenix: Development and Application of a Bushfire Risk Management Tool. *Australian Journal of Emergency Management, The*, 23(4):47.

Tonbul, H. and Kavzoglu, T. (2017). Object- based Forest Fire Analysis for Pedrógão Grande Fire Using Landsat 8 OLI and Sentinel-2A Imagery. In *American Geophysical Union*, number July, page 41400, New Orleans, Louisiana.

Turner, J. (1972). The drought code component of the Canadian forest fire behavior system. Technical Report 1316, Canadian Forestry Service, Ottawa.

Tymstra, C., Bryce, R., Wotton, B., Taylor, S., and Armitage, O. (2010). Development and structure of Prometheus: the Canadian wildland fire growth simulation model. Technical report, Natural Resources Canada, Information Report NOR-X-417.

Ugray, Z., Lasdon, L., Plummer, J., Glover, F., Kelly, J., and Marti, R. (2007). Scatter search and local NLP solvers: A multistart framework for global optimization. *INFORMS Journal on Computing*, 19(3):328–340.

Valero, M., Rios, O., Pastor, E., and Planas, E. (2015). Automatic detection of wildfire active fronts from aerial thermal infrared images. In *13th International Workshop on Advanced Infrared Technology & Applications: Proceedings*, pages 63–67, Pisa.

Valero, M., Rios, O., Pastor, E., and Planas, E. (2016). Automated mapping of active wildland fires through aerial infrared imaging and unsupervised edge detectors. *(in review)*, pages 1–35.

Valero, M., Rios, O., Pastor, E., and Planas, E. (2017a). An integral system for automated tactical monitoring and data-driven spread forecasting of wildfires. *Fire Safety Journal*, 91(March):835 – 844.

Valero, M., Rios, O., Pastor, E., and Planas, E. (2017b). Automated location of wildfire active fronts through aerial infrared imaging and edge detection algorithms. *(in review)*.

Valero, M. M., Rios, O., Planas, E., and Pastor, E. (2018). Automated location of active fire perimeters in aerial infrared imaging using unsupervised edge detectors. *International Journal of Wildland Fire*, 27(4):241–256.

Valero, M. M., Verstockt, S., Rios, O., Pastor, E., Vandecasteele, F., and Barcelonatech, D. C. (2017c). Flame filtering and perimeter localization of wildfires using aerial thermal imagery. In *SPIE 10214, Thermosense: Thermal Infrared Applications XXXIX*, Anaheim, California, United States.

Viegas, D., Palheiro, P., Pita, L., Ribeiro, L., Cruz, M., Ollero, A., Arrue, B., and Ramiro, M. D. (2006). Analysis of fire behaviour in Mediterranean shrubs: The Gestosa fire experiments (Portugal). *Forest Ecology and Management*, 234:S101.

Viegas, D. C. (2008). Project Spread – Forest Fire Spread Prevention and Mitigation, 2004.

Viegas, D. X. (2004). On the existence of a steady state regime for slope and wind driven fires. *International Journal of Wildland Fire*, 13(1):101–117.

Viegas, D. X. (2011). Overview of Forest Fire Propagation Research. *Fire Safety Science*, 10:95–108.

Viegas, D. X. and Simeoni, A. (2010). Eruptive Behaviour of Forest Fires. *Fire Technology*, 47(2):303–320.

Wagenbrenner, N. S., Forthofer, J. M., Lamb, B. K., Shannon, K. S., and Butler, B. W. (2016). Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja. *Atmospheric Chemistry and Physics*, (January):1–32.

Wagner, C. V. (1988). Effect of slope on fires spreading downhill. *Can. J. For. Res.*

Waltz, R. A., Morales, J. L., Nocedal, J., and Orban, D. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3):391–408.

Weller, H. G. and Tabor, G. (1998). A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6):620–631.

Wendt, K., Denham, M., Cortés, A., and Margalef, T. (2011). Evolutionary optimisation techniques to estimate input parameters in environmental emergency modelling. *Studies in Computational Intelligence*, 359:125–143.

Westhaver, A. (2017). Why some homes survived: Learning from the Fort McMurray wildland/urban interface fire disaster. *ICLR research paper series*, (56):81.

Williams, J. (2013). Exploring the onset of high-impact mega-fires through a forest land management prism. *Forest Ecology and Management*, 294(-):4–10.

Wyngaard, J. C. (2004). Toward Numerical Modeling in the "Terra Incognita". *Journal of the Atmospheric Sciences*, 61(14):1816–1826.

Zhang, C., Rochoux, M., Tang, W., Gollner, M., Filippi, J. B., and Trouvé, A. (2017). Evaluation of a data-driven wildland fire spread forecast model with spatially-distributed parameter estimation in simulations of the FireFlux I field-scale experiment. *Fire Safety Journal*, 91(April):758–767.

Zhou, K., Suzuki, S., and Manzello, S. L. (2014). Experimental Study of Firebrand Transport. *Fire Technology*, pages 785–799.