ABSTRACT

Title of dissertation: DATA-DRIVEN SIMULATIONS OF WILDFIRE SPREAD AT REGIONAL SCALES

Cong Zhang
Doctor of Philosophy, 2018

Dissertation directed by: Professor Arnaud Trouvé
Department of Mechanical Engineering

Current wildfire spread simulators lack the ability to provide accurate prediction of the active flame burning areas at regional scales due to two main challenges: a modeling challenge associated with providing accurate mathematical representations of the multi-physics multi-scale processes that induce the fire dynamics, and a data challenge associated with providing accurate estimates of the initial fire position and the physical parameters that are required by the fire spread models. A promising approach to overcome these limitations is data assimilation: data assimilation aims at integrating available observations into the fire spread simulator, while accounting for their respective uncertainties, in order to infer a more accurate estimate of the fire front position and to produce a more reliable forecast of the wildfire behavior.

The main objective of the present study is to design and evaluate suitable algorithms for regional-scale wildfire spread simulations, which are able to properly handle the variations in wildfire spread due to the significant spatial heterogeneity

in the model inputs and to the temporal changes in the wildfire behavior. First we developed a grid-based spatialized parameter estimation approach where the estimation targets are the spatially-varying input model parameters. Then we proposed an efficient and robust method to compute the discrepancy between the observed and simulated fire fronts, which is based on a front shape similarity measure inspired from image processing theory. The new method is demonstrated in the context of Luenberger observer-based state estimation strategy. Finally we developed a dual state-parameter estimation method where we estimate both model state and model parameters simultaneously in order to retrieve more accurate physical values of model parameters and achieve a better forecast performance in terms of fire front positions. All these efforts aim at designing algorithmic solutions to overcome the difficulties associated with spatially-varying environmental conditions and potentially complex fireline shapes and topologies. It paves the way towards real-time monitoring and forecasting of wildfire dynamics at regional scales.

DATA-DRIVEN SIMULATIONS OF WILDFIRE SPREAD AT REGIONAL SCALES

by

Cong Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Dr. Arnaud Trouvé, Chair/Advisor
Dr. Mélanie Rochoux
Dr. Michael Gollner
Dr. André Marshall
Dr. Shapour Azarm
Dr. Kayo Ide, Dean's Representative

Acknowledgments

This research project is funded by the US National Science Foundation (Award number 1331615) through the WIFIRE project, as well as from the University of Maryland's Council on the Environment seed grants. Financial support provided by the Department of Fire Protection Engineering, the French National Program LEFE/INSU, and LabEx AMIES is also gratefully acknowledged. The participation to the 2016 CEMRACS Summer program (FLAM project in collaboration with CERFACS, INRIA and LIMSI) was beneficial for this work.

First and foremost, my deep gratitude goes to my advisor Dr. Arnaud Trouvé, for his continuous support throughout my PhD study. His patient guidance and broad knowledge helped me in all the time of exploring this interdisciplinary project. I would also like to thank my co-advisor Dr. Mélanie Rochoux, who contributed many insightful discussions to this project. She was always there to help and instruct me on both research and writing, and I learned a lot from her. It was my pleasure to work with both of them in the past four years.

I would also like to thank my committee members, Dr. Michael Gollner, Dr. Kayo Ide, Dr. Shapour Azarm and Dr. André Marshall, for serving on my thesis committee, and offering invaluable suggestions to make this work better. Special thanks to Dr. Michael Gollner and Dr. Kayo Ide for the helpful discussions on wildland fire physics and data assimilation.

My sincere thanks also go to our research collaborators, Dr. Annabelle Collin, Dr. Philippe Moireau and Dr. Didier Lucor, for their help and support during the

2016 CEMRACS summer school. This project would not be possible without their support. Thanks are also due to my research partners, Wei Tang, Mickaël Durand and Maria Theodori, who have been great colleagues and friends to work with. Besides, thanks to Cheng Da, Salman Verma and Alexis Marchand for their ideas and suggestions on this project, and thanks to all my colleagues from the fire modeling group and weather chaos group for their help and encouragement during my study.

Finally I thank all my friends and family members for supporting and accompanying me along this journey.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Wildfire, a Burning Global Issue

Wildland fire has been prevalent on a global scale in recent years, due to the increasing drought and extreme weather conditions [1]. Figure 1.1 shows the active fire detections at global scale in June, 2009, using MODIS data from the Terra satellite. Whether it is ignited by the natural forces (mostly by lightning), or human beings (both intentionally or unintentionally), wildfires continue to threaten our home and communities around the world. In June, 2017, a series of deadly wildfires across central Portugal have caused at least 66 deaths and 204 injured people. The Tomas fire, largest wildfire on record in California happened in December 2017, destroyed more than 1000 buildings, including many homes. Thousands of people were forced to flee and over 8500 firefighters were mobilized to fight it. In addition to these tremendous losses, future climate changes are likely to enhance increases in mean temperature (about 2-4° globally) with significant drying in some regions, with the consequences of favoring the occurrence of wildfires and lengthening the fire season (by about twenty days per year) [2–4].

Narrow down to North America area, Fig. 1.2 presents the number of wildfires that occurred in the United States, together with the total burnt area from year 1985

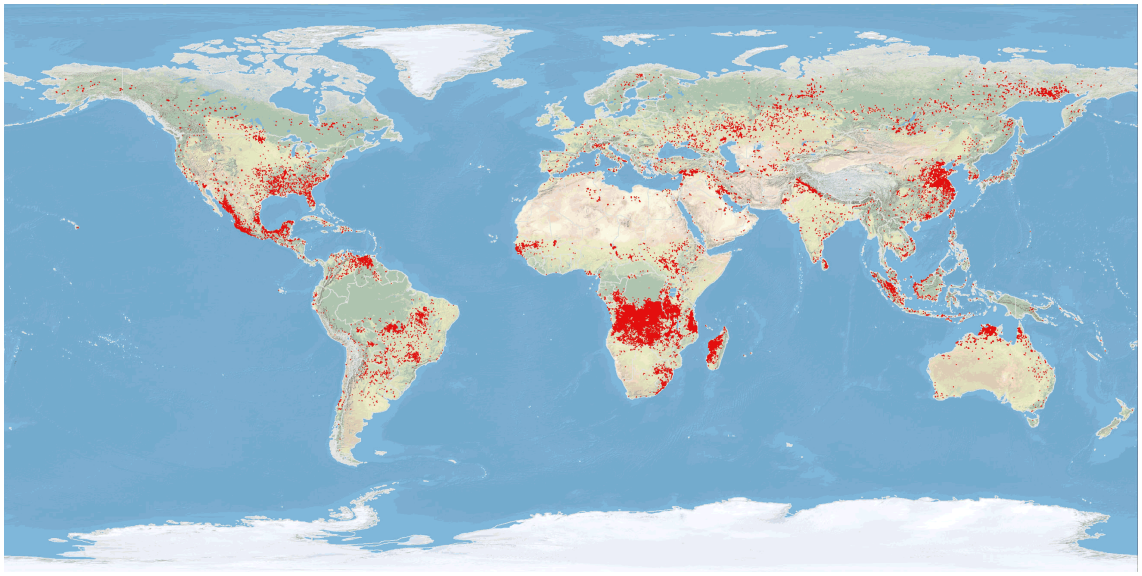

Figure 1.1: MODIS rapid response active fire detections in June, 2009. Source: NASA, http://rapidfire.sci.gsfc.nasa.gov/

to year 2017. The total number of wildfires has gradually decreased in recent years, possibly due to human factors such as fire suppression and fire exclusion policies, increased firefighting efficiency, improved fire prevention. However, the burnt areas have been multiplied by a factor of three, implying that there are more wildfires affecting a larger area in recent years; these large wildfires (sometimes referred to as megafires) feature higher fire intensity and fire severity with important consequences for public safety and ecosystem [1].

Since the early 1950s, formal research initiatives by federal and state government forestry agencies started concerted efforts to build fire danger rating systems, which embodied a fire behavior prediction component in order to better prepare for wildfire hazards. These efforts include:

- The Canadian Forest Fire Danger Rating System (CFFDRS) is a globally

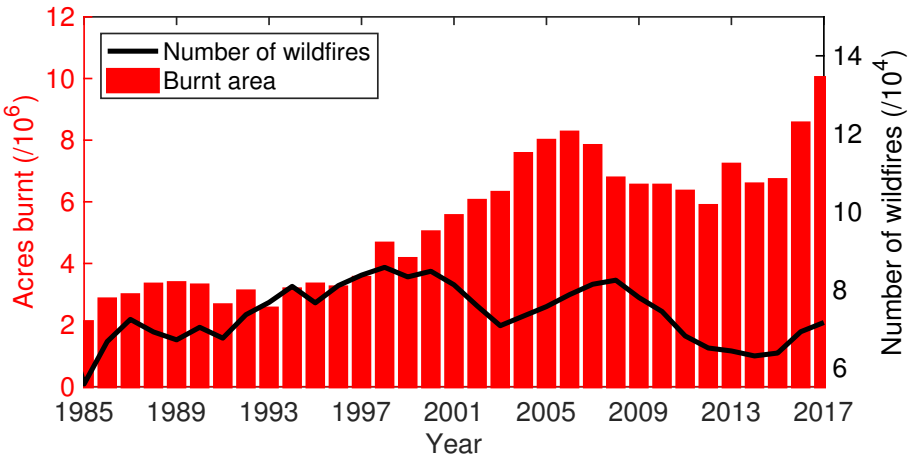

Figure 1.2: Five-year moving averages of wildfire data from the National Interagency Fire Center: solid black line indicates the number of wildfires; red bar indicates the total burnt area (acres).

known wildland fire risk assessment system, which has been developed through 60 years of research and field study efforts by the Canadian Forest Service [5]. It has two major components: the fire weather index (FWI) system and the fire behavior prediction (FBP) system. It is widely used today to predict the potential risk for daily fire ignition across the landscape as well as to assess the behavior of a specific wildfire in a particular forest type. The CFFDRS system has also been introduced and adapted into the Fire Danger Rating Systems (FDRS) in southeast Asia countries such as Indonesia and Malaysia [6].

- The US National Fire Danger Rating System (NFDRS) has been developed to provide local indices of fire occurrence or behavior based on daily measurements of vegetation, terrain topography, weather and risk of ignition (i.e. human-caused and lightning) [7]. More recently, the Wildland Fire Deci-

sion Support web-based System (WFDSS) emerged to provide state-of-the-art risk analysis for individual fires, including national weather forecast data, fire behavior prediction, economic assessment, smoke management assessment and landscape databases. In particular, this analysis produces a fire spread probability map and an inventory of the assets (for instance critical infrastructure) most likely to be at risk [8].

- In Australia, the McArthur grassland and forest Fire Danger Rating Systems (FDRS) [9] along with the CSIRO Grassland Fire Spread Meter (GSFM) [10] serve as the two main tools for fire danger forecasting. The Fire Danger Rating Systems are used by rural fire authorities to forecast fire danger in Australian forest and grassland areas. The Fire Spread Meter is used to predict how quickly the grassland wildfire will spread.

- Across european countries, the European Forest Fire Information System (EFFIS) has been established by the Joint Research Centre and the Directorate General for Environment of the European Commission to provide up-to-date, reliable information on forest fire risk assessments during both pre-fire and post-fire phases [11]. The EFFIS serves as a complementary system in addtion to national and regional systems in european countries, and provides harmonized information for the sake of international collaboration on large scale wildfire fighting.

These systems serve as the main components for wildfire danger prediction and prevention across the world, but they feature some limitations in operational

context. Fire danger rating systems such as the FWI mostly rely on meteorological information and only partially account for various vegetation properties, for instance they usually do not differentiate between different vegetation fuel types. When it comes to wildfire spread modeling, fire prediction systems are subject to significant uncertainties due to modeling simplifications and knowledge gaps in the environmental and meteorological conditions as detailed in Sec. 1.2.

1.2 Modeling of Wildfire Spread

Alongside numerous efforts to predict the probability of wildfire occurence, there is a great need for accurate predictions of wildfire spread after ignition, which can benefit both fire risk management and fire emergency response, for instance to design efficient firefighting strategies and thereby avoid the development of large-scale highly-destructive fires. Computer-aided modeling approach has been a popular choice to analyze the behavior of wildfire spread, especially the impact of a number of environmental factors on the wildfire behavior. In recent years, advances in computational power and spatial data analysis (GIS, remote sensing, etc) has led to an increase in wildfire spread modeling efforts.

1.2.1 Multi-scale Multi-physics Problem

Wildfire spread can be regarded as a succession of ignitions towards the unburnt vegetation fuel region. Once a certain area of vegetation fuel is ignited (human-induced ignition, thunderstorm lightning, etc), heat will be released from

chemical reactions (broadly categorized as an oxidation reaction) in the process of combustion; then transferred to surrounding unburnt vegetation through convection, radiation and conduction heat transfer mechanisms, leading to the thermal degradation of the vegetation, to the release of flammable gases and to their subsequent ignition. The former is the domain of chemistry (e.g. chemical kinetics) and occurs on the scale of mean free paths, and the latter is the domain of physics (e.g. heat transfer and fluid mechanics) and occurs on scales ranging from millimetres up to kilometres (Table 1.1) [12].

Table 1.1: Major biological, physical and chemical components and processes occuring in a wildfire and relevant temporal and spatial scales [12].

Type	Time scale (s)	Vertical scale (m)	Horizontal scale (m)
Combustion reactions	0.0001 - 0.01	0.0001 - 0.01	0.0001 - 0.01
Fuel particles	-	0.001 - 0.01	0.001 - 0.01
Fuel complex	-	1 - 20	1 - 100
Flame	0.1 - 30	0.1 - 10	0.1 - 2
Radiation	0.1 - 30	0.1 - 10	0.1 - 50
Conduction	0.01 - 10	0.01 - 0.1	0.01 - 0.1
Convection	1 - 100	0.1 - 100	0.1 - 10
Turbulence	0.1 - 1,000	1 - 1,000	1 - 1,000
Spotting	1 - 100	1 - 3,000	1 - 10,000
Plume	1 - 100,000	1 - 10,000	1 - 100

The interactions of these processes occur over a wide range of temporal and spatial scales (Fig. 1.3): vegetation scales that characterize biomass fuels at centimeter scale or less; flame scales that characterize combustion and heat transfer

processes at meter scale; topographical scales and landcover scales that characterize terrain and vegetation boundary layer at hundred-meter scale; and meteorological micro-/meso-scales that characterize atmospheric conditions at hundred-kilometer scale [13]. Due to these complicated interactions at multiple temporal and spatial scales, understanding the key mechanisms driving a wildfire and their interactions is still an active research area [14]. Modeling wildland fire behavior is therefore a challenging problem.

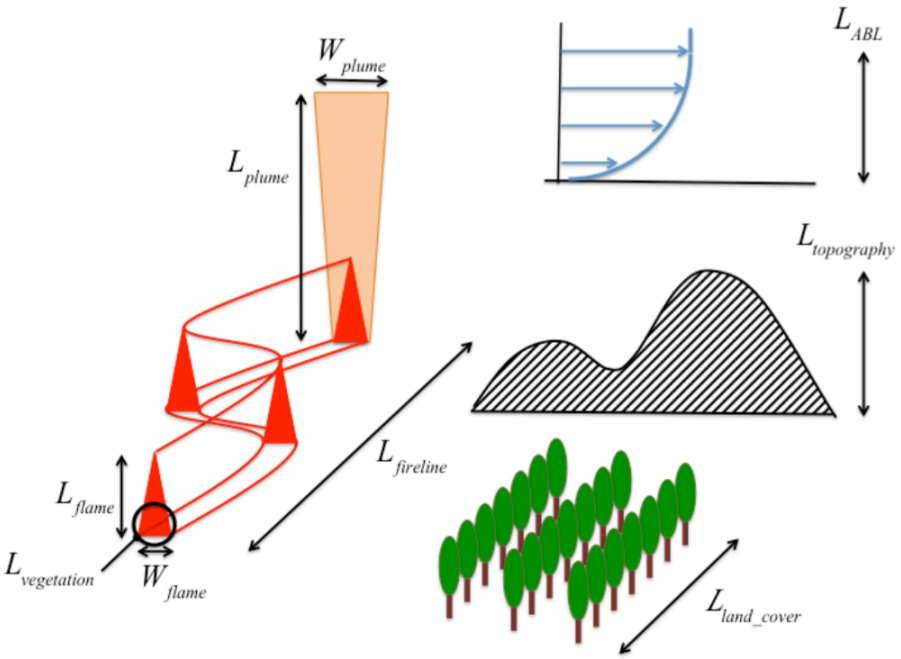

Figure 1.3: The different spatial scales invoved with wildfire behavior: vegetation scale, flame scale, topographical scale and meteorological scale [15].

In general we differentiate between three main types of wildfires depending on the fuel being consumed: ground fire, surface fire and crown fire. Ground fire is a slow mode of combustion occurring beneath the surface layers of the forest ground,

at low temperatures and usually without any flame; a typical example is a peat fire [16]. Surface fire consumes fine particles in the surface litter and undergrowth (tree leaves, conifer needs, bark, branches, shrubs, etc); usually surface fuels are less compact than ground fuels and provide conditions more favorable for propagation. Crown fire only occurs when surface fire spreads vertically up to the canopy and tree crowns under extreme dry and windy conditions. It barely happens due to the higher moisture content of live vegetation crown fuels, but when it does, it corresponds to a drastic increase in the heat release rate and the fire size. Crown fire often enhances the production of embers, and these embers could be transported a long distance by the strong wind. They can potentially initiate new fire sources called spotting fires and therefore, drastically enhance the fire spread. In the present work, we focus on surface fires since this is the main mode of wildfire spread and also the most studied one.

1.2.2 Wide Range of Modeling Approaches

There are different length scales involved in wildfire spread as seen from Table 1.1. Depending on the scale(s) of interest and the purpose of the study (operational or research), several types of models are reported in the literature [12, 17, 18].

On the one hand, research-level models aim at a better understanding of the fundamental physical and chemical processes involved in the combustion of vegetation fuel and driving the fire behavior. For this purpose, high fidelity numerical simulations are performed at flame scale (1 m) to resolve interactions between the

vegetation and the flame as well as between the flame and the atmosphere. These physical models explicitly solve for mass, momentum and energy balance equations using computational fluid dynamics (CFD) tools, see for instance WFDS [19] and FIRETEC [20]. Figure 1.4 shows an example of a coupled modeling system, where an atmospheric hydrodynamics model, HIGRAD, is coupled to a wildfire behavior model, FIRETEC, to produce a coupled atmosphere/wildfire behavior model based on conservation of mass, momentum, species, and energy. It is used to study the physical processes of the interactions between heterogeneous vegetation, topography, and atmospheric conditions during a wildfire event [21].

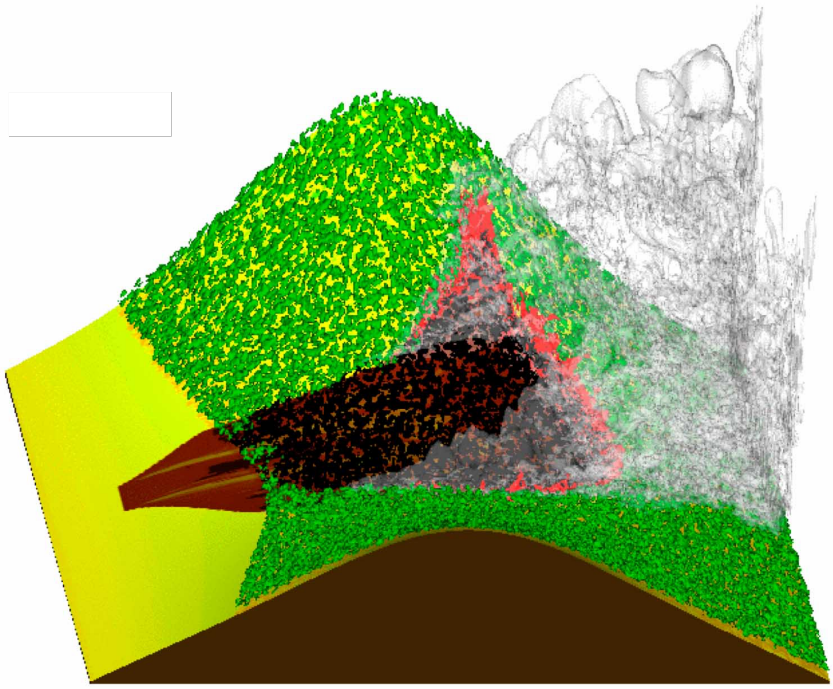

Figure 1.4: HIGRAD/FIRETEC coupled simulation of wildfire burning upslope into a saddle. This simulation is used to illustrate the effects of transient wind conditions, the effects of nonhomogeneous terrain and the effects of nonuniform fuels. Source: Los Alamos National Laboratory, http://www.lanl.gov/.

While the physics-based CFD approach has the potential to accurately describe flame-scale processes and provides us a better understanding of the physical and chemical processes controlling fire propagation, this approach is currently restricted to academic research projects due to its high computational cost and high-resolution input data requirement [13]. It is not suitable for real-time forecast of wildfire spread which supports an operational decision-making process. On the other hand, operational-level models adopt empirical or semi-empirical models to predict the growth of a wildfire at regional scales (i.e. at scales ranging from a few hundreds of meters up to several kilometers). They generally treat a wildfire as a propagating interface from the burnt to the unburnt vegetation that self propagates in the normal direction. This propagating interface is referred to as the fire front or fireline, whose local propagation speed is called the rate of spread (ROS). In this context, work on wildfire spread modeling aims at predicting the ROS of the fire front for a given set of environmental conditions. The ROS is usually modeled using an empirical or quasi-empirical function with respect to a reduced number of factors characterizing environmental conditions such as fuel moisture content, local wind conditions and terrain topography. Empirical modeling is based on phenomenological descriptions or statistical correlations of observed fire behaviors such as wind-tunnel experiments and field-scale controlled burning experiments [10]. Semi-empirical modeling relies on a physical framework in combination with statistical modeling. In particular, it formulates the ROS using the energy balance equation and the resulting model parameters are calibrated using experimental data. Both empirical and semi-empirical approaches are simple and computationally efficient compared with physics-based

modeling approach, making them consistent with an operational framework. The semi-empirical approach has the additional advantage of including some relevant physical aspects of wildfire spread through the energy balance equation. The most widely-used semi-empirical model is Rothermel's model [22, 23].

Current operational wildfire growth simulators such as FARSITE [24], PHOENIX RapidFire [25] or PROMETHEUS [26] adopt a regional-scale perspective and use semi-empirical ROS models to propagate the fire front, providing a quick model response for real-time forecasting purposes. However, their accuracy is limited due to the simplified representation of wildfire behavior. First, these models have a domain of validity that is limited to the domain of experimental conditions used during their original development and calibration. Their extension to study regional-scale wildfire spread problem is thus questionable. For instance, they do not include extreme fire behavior conditions due to high wind conditions and/or steep slopes. Besides, these models do not explicitly account for fire-atmosphere interactions, which are important to account for the fire feedbacks on the near-surface winds [27–30]. And third, they rely on uncertain input parameters that may not be known or may only be known with limited spatial and temporal resolution [14, 31]. Therefore, current operational models suffer from these simplifications and uncertainties; when operated in real time forecasting mode, their performance is often found to be dissatisfactory compared with observations. A new approach to this problem is to couple existing operational models and real-time observations, with the objective of reducing the uncertainties in model fidelity and input data in order to achieve better forecast performance. This approach belongs to the large category called

data-driven modeling and is detailed in Sec. 1.3.

1.3 Data-driven Wildfire Spread Modeling

Data-driven modeling is meant to dynamically incorporate additional data into an executing application for model optimization purpose. And data assimilation is one popular statistical method to integrate available observations into the modeling application, which accounts for the modeling and observation uncertainties. Dating back to the application of steering the Apollo moon spaceships in the 1960s, data assimilation methodologies have proven successful over the past decades for a wide range of applications in geosciences and engineering sciences, e.g. numerical weather forecasting, oceanography, atmospheric chemistry, biomechanics, reservoir engineering [32,33]. This is a promising approach to overcome the limitations in operational wildfire spread modeling. The objective is to deploy an inverse modeling procedure taking advantage of available sensor observation data to provide the best possible prediction of the future fire front position simulated by operational-level models.

1.3.1 Data Assimilation

Data assimilation provides an optimized description of the state of a given system by including all sources of information, i.e. from numerical models and observations, and their estimated uncertainties. A data assimilation framework typically features the following main components (see Fig. 1.5 corresponding to sequential Kalman filtering): a forward model that simulates the state of a physical system

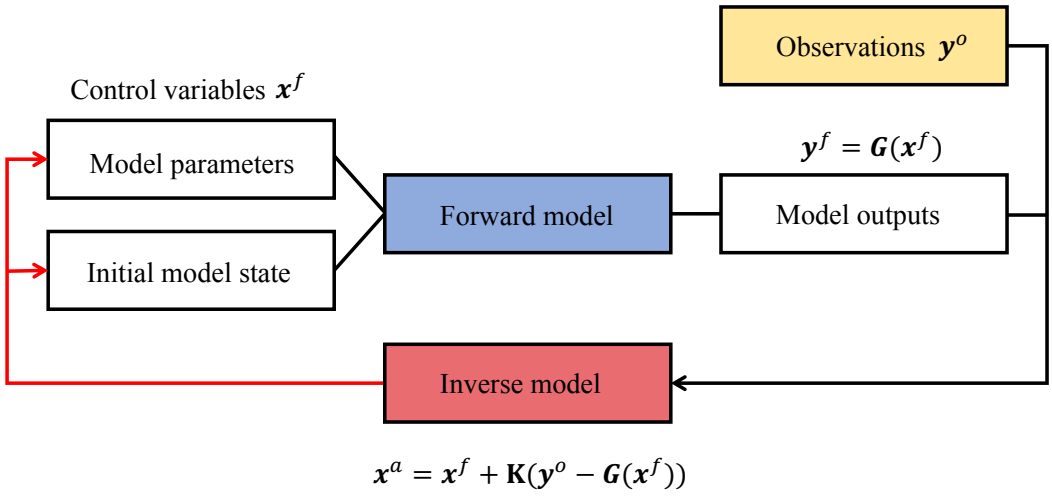

Figure 1.5: Data assimilation flowchart for a typical data-driven model based on sequential Kalman filtering; the ensemble Kalman filter (EnKF) that is part of this family of methods and that is an ensemble-based data assimilation method will be detailed in Chapter 2.

(with some modeling uncertainty) given a set of physical and numerical parameters as well as initial and boundary conditions; a series of observations $\mathbf{y}^o$ (with some measurement and processing uncertainty); and an inverse model that

1. defines the estimation targets (or control variables) included in the control vector $\mathbf{x}$. For wildfire spread modeling, the control variables could be the model state (fire front positions) in a state estimation operation mode, and the model parameters (physical parameters controlling fire propagation) in a parameter estimation operation mode;

2. computes the distance between the observation $\mathbf{y}^o$ and the simulated prediction $\mathbf{y}^f = \mathcal{G}(\mathbf{x}^f)$ corresponding to a priori (or forecast) determined by the control variables $\mathbf{x}^f$ (this is the forecast step): $\mathcal{G}$ is referred to as the observation

operator mapping the control vector onto the observation space;

3. computes the posterior (or analysis) values $\mathbf{x}^a$ according to some algorithm that minimizes the distance $(\mathbf{y}^o - \mathbf{y}^f)$ given some weight $\mathbf{K}$ (or confidence) to the available information (this is the update step): the updated state $\mathbf{y}^a$ can be obtained by applying the observation operator to $\mathbf{x}^a$ such that $\mathbf{y}^a = \mathcal{G}(\mathbf{x}^a)$ is a more accurate estimate of the (unknown) true state of the system than the observations $\mathbf{y}^o$ or the forecast $\mathbf{y}^f$ taken separately.

This two-step procedure (forecast and update) is performed sequentially as observations become available; the next forecast is initialized by considering the previous analysis and so on. The resulting updated state of the system is referred to as the *data-driven run*. The performance of the data assimilation algorithm over a given time period is usually evaluated by its ability to reduce the distance between the simulated state of the system and the observations, i.e. by comparing the differences between the *free run* (i.e. the simulation provided by the forward model without data assimilation over the whole time period) and the *data-driven run*.

Formulating the weight $\mathbf{K}$ in adequation with the uncertainties in the observations and in the forward model is a key feature of any data assimilation algorithm to make it successful [32]. The uncertainties inherent in geosciences and engineering sciences generally go beyond the limitations of deterministic model capabilities. The error is unquantifiable when considering a single deterministic run of the forward model. Thus, the uncertainties suggest the use of ensembles to consider multiple scenarios of the system behavior and to thereby better represent its potential vari-

ability over a time period given sources of uncertainties in the input parameters and in the model. In this framework, the estimation targets are considered as random variables and are therefore characterized by a probability density function (PDF). Thus, each scenario corresponds to a different realization of the estimation targets and to a different realization of the model state by integrating the forward model. Ensemble-based data assimilation algorithms such as the ensemble Kalman filter (EnKF) [33] therefore produce an ensemble of deterministic model state predictions to have a better representation of the errors. EnKF is essentially a Monte-Carlo implementation of the Bayesian update problem, and the weight $\mathbf{K}$ is then formulated in a statistical sense using the ensemble of estimation targets and the ensemble of model states.

As stated before, the control variables can include the model state; this is useful to retrieve a more accurate initial condition from which the forward model can restart to produce forecasts (this is referred to as "state estimation"). They can also include physical parameters required as inputs to the problem; this is in general useful to reduce the bias in the model predictions and to extend the forecast quality over time (this is referred to as "parameter estimation"). The state and the parameters can be estimated separately or simultaneously depending on the problem [34–36].

1.3.2 Wildfire Applications

While still at an early stage of development, the idea of data assimilation has been considered over the last decade for possible applications to wildfire behavior problems [37]. While there are some variations in the literature, the forward model is typically a wildfire growth simulator that uses a Rothermel-type ROS description; observations are generally fireline positions; the inverse model is either some ensemble-based algorithms accounting for both modeling and observation errors, or deterministic optimization techniques to minimize a certain cost function; and control variables are generally the parameters of the ROS model or the fireline positions. This discussion on the wildfire applications, as well as the following discussion on the available observed fire data, provide a brief summary of all relevant researches, and they are adapted from the NSF-WIFIRE technical report [15].

Table 1.2 provides an overview of the recent data assimilation methods designed for wildfire applications. In particular, Mandel et al. [37] pioneered the idea of combining a data assimilation method with a forward model for wildland fire to sequentially update model simulations using ensemble-based Kalman filter algorithms. Towards the ultimate objective of building a real-time coupled atmospheric-wildland fire modeling system, their work has shown promising results while raising some concerns about performing data assimilation for wildfire applications. In Ref. [38] they used a morphing EnKF technique to assimilate measured temperatures into running fire growth models. The wildfire problem is found to be challenging for usual data assimilation methods since the PDF of the temperature is assumed to be Gaussian

while it is actually bimodal (burnt region or unburnt vegetation). This is typically the case for the EnKF algorithms. The idea of morphing was derived from image processing to map all fronts in the same reference frame and to thereby avoid having bimodal PDF. However, this led to technical difficulties for implementation in practice. Denham et al. [39] adopted a dynamic data-driven genetic algorithm to automatically adjust input parameters of the ROS model, thus enhancing forecast quality. However, genetic algorithms usually require many realizations (also referred to as "members" or "particles") in the ensemble and so many integrations of the forward model, which can become prohibitive in practice. Rochoux et al. [40–42] designed an EnKF with a Lagrangian representation of the fire front; the transformation of the observed quantities from the burning area (or temperature) field like in [37] onto a set of marker positions avoids the problem of bimodal PDF (the PDF on each front marker along the fireline features a Gaussian PDF) and formulates the discrepancies between the observed and the simulated fronts as an Euclidean distance, hereby allowing the application of a standard EnKF. This strategy was found promising to apply both parameter estimation [41] and state estimation [42] on controlled burn trials and to increase forecast performance. Rios et al. [43] also demonstrated the reliability of inverse modelling-based algorithm to improve short-term fire spread forecast using a Lagrangian front-tracking solver.

In the present study, we continue developing a dynamic data-driven application system for wildfire spread prediction based on previous work by Rochoux et al [40–42]. The main challenge is now to extend this work to regional-scale wildfires and in particular to be able to address the heterogeneities in the environmental

Table 1.2: Literature review of data assimilation methods for wildfire applications over the last decades - adapted from NSF-WIFIRE report [15].

References	Control variables	Forward model	Inverse model
Bianchini et al. [44, 45]	fire/no fire field	cell automata	particle filter
Wendt et al. [46] Denham et al. [39]	ROS parameters	cell automata	genetic algorithm
Gu and Hu [47, 48]	fire/no fire field	cell automata	particle filter
Mandel et al. [37]	temperature field	PDE-based model	EnKF
Mandel et al. [38]	temperature field	coupled WRF /level-set	EnKF
Vejmelka et al. [49, 50]	fuel moisture	fuel moisture model*	EKF/UKF**
Mandel et al. [51]	fire arrival time	coupled WRF-SFIRE	least squares
da Silva et al. [52]	ROS parameters	Eulerian level-set	particle filter
Lautenberger [53]	ROS parameters	Eulerian level-set	genetic algorithm
Rochoux et al. [41]	ROS parameters	Eulerian level-set	EnKF
Rochoux et al. [42]	fire front position	Eulerian level-set	EnKF
Rios et al. [43]	ROS parameters	Lagrangian model	gradient-based optimization

* Fuel moisture model is used to advance moisture content in time, not fire propagation.
** EKF/UKF stands for Extended Kalman filter and Unscented Kalman filter, respectively. Both are variants of Kalman filter to address nonlinear problems.

conditions and the nonlinear wildfire behavior. These heterogeneities induce significant uncertainties in the simulated fire front by operational-level models, make the

topology of the fire front quite complex and require the data assimilation algorithm to properly address position and topological errors.

1.3.3 Available Observed Fire Data

The challenge in real-time modeling is to provide access to useful real-time data. Whether collected from prescribed burns or uncontrolled wildfires, the types as well as the spatial and temporal resolution of the available data are paramount to the development and use of data-driven wildfire spread models. On the one hand, wildfire data have primarily been collected via prescribed burn experiments. These experiments are typically less intense than an accidental wildfire but they are sufficient to provide a basis for model validation. They still require significant time and resources. Early data collection efforts began with single goals in mind (e.g. FIRE-FLUX I to evaluate coupled fire/atmosphere models [54]). However, more recent data collection efforts have tried to collect data on prescribed burns that serve multiple user groups (e.g. RxCADRE [55]); Table 1.3 lists in-situ measurements available during RxCADRE experiments. The RxCADRE prescribed burn experiments yield a comprehensive dataset of fire behavior, fire effects, smoke chemistry and dynamics, with measurements taken systematically at multiple scales ranging from centimeter to kilometer. The objective is to help fire modelers and scientists to validate and refine physics-based fire behavior models, as well as provide insights to improve empirical and semi-empirical models given a well-founded understanding of fire physics from the experimental dataset.

Table 1.3: In-situ measurements available during the RxCADRE experiments [55].

Discipline	Measurement data
Fuel characteristics	Mass, cover,depth, moisture
Fire effects	Thermal radiometry, HD visual imagery, stem temperatures
Local event-scale meterology	Plume properties, fine-scale wind and thermodynamic fields
Fire behavior	Fire intensity, ROS, convective/radiative power and energy, soil heating, IR imagery, wind/flame velocity
Event-scale fire mapping	Fire radiative power and energy, flame front development, Satellite imagery of fire and effects
Emissions and event-scale plume behavior	CO, CO_2, H_2O, $PM_{2.5}$ black carbon, plume height

For developing and validating a prototype data-driven wildfire spread simulator, we will primarily focus on the measurement data of fire behavior and spread rather than fire emissions, even though the latter database is also important for many other user groups [56, 57]. In the perspective of tracking wildfire behavior at regional scale, there is a great need to take advantage of remote sensing technologies [58, 59], aboard airplanes [60], satellites [61] or unmanned aerial vehicles (UAV) [62]. They provide overhead measurements of infrared (IR) images or fire radiative power (FRP) estimates, which are useful to detect active fire location and estimate fire intensity for each pixel. They can be processed to track the fireline for use with real-time wildfire spread modeling.

Firelines with spatial resolution of approximately 10 m and temporal resolution

of approximately 10 minutes are ultimately desired to achieve a reliable forecasting tool with accurate-enough predictions for fire behavior. These requirements can theoretically be met with current satellite technology; however, these requirements may also be cost-prohibitive at the moment. While polar orbiting satellites such as Terra, Aqua, and S-NPP (with MODIS and VIIRS sensors, respectively), provide autonomous, synoptic observations of fire activity, both day and night, nominally twice a day from each sensor, their temporal resolution, and the corresponding spatial resolution, may not be adequate for real-time fire modeling. Figure 1.6 (top) shows one example of active fire detections by VIIRS sensor on the Suomi National Polar-orbiting Partnership (S-NPP) satellite. While the majority of the burning area is detected, this dataset is not sufficient to be fed into current data-driven fire spread modeling system due to the limited spatial resolution. NOAAs Geostationary Operational Environmental Satellite system (GOES) offers greater temporal resolution, but still suffers from limited spatial resolution. One way to alleviate the resolution problem is to fuse data with various sources of remotely sensed data, to fill gaps and then improve remotely sensed data resolution. GeoMAC (Geospatial Multi-Agency Coordination) provides fire perimeter data based upon multiple data sources, including incident intelligence sources (onsite crew, apparatus measurements), GPS data, infrared imagery from fixed wing and satellite platforms. Figure 1.6 (bottom) shows that the spatial resolution of GeoMAC dataset is good enough for current data-driven modeling system. However, it is worth noting that the data uncertainties associated with data fusion should be properly addressed and quantified, which remains a challenging task so far. NIROPS (USDA Forest

Service National Infrared Operations) have also shown that it is possible to produce firelines at good spatial resolution using an airborne infrared sensor. However, the low frequency of the fireline mapping (maps are made only once per night) is a limitation, and part of the problem is that the process is not automated. In addition, the use of drones, for instance the use of an MQ-1 Predator Remotely Piloted Aircraft (RPA) on the Rim fire in California, was successful in observing particular fires, but no permanent program has been established, most likely because of the high cost and because of UAV safety concerns.

While numerous challenges are present at the route to develop an operational data-driven application system for wildfire forecasting, new technologies are under development. Emerging advances in remote sensing (such as UAVs and commercial satellites) are foreseen to meet the high spatial and temporal resolution requirements of data assimilation. The increasing capability of computational power and intelligent algorithms will make real-time wildfire spread prediction a reality in the near future [63].

1.4 Thesis Overview

This thesis work is carried out as part of the NSF-WIFIRE project [15]. The goal is to design and evaluate a data-driven wildland fire spread modeling strategy that is applicable at regional scale and that is able to address the heterogeneities in the environmental conditions and the nonlinear wildfire behavior, which induce significant uncertainties in the simulated fire front by operational-level models, make

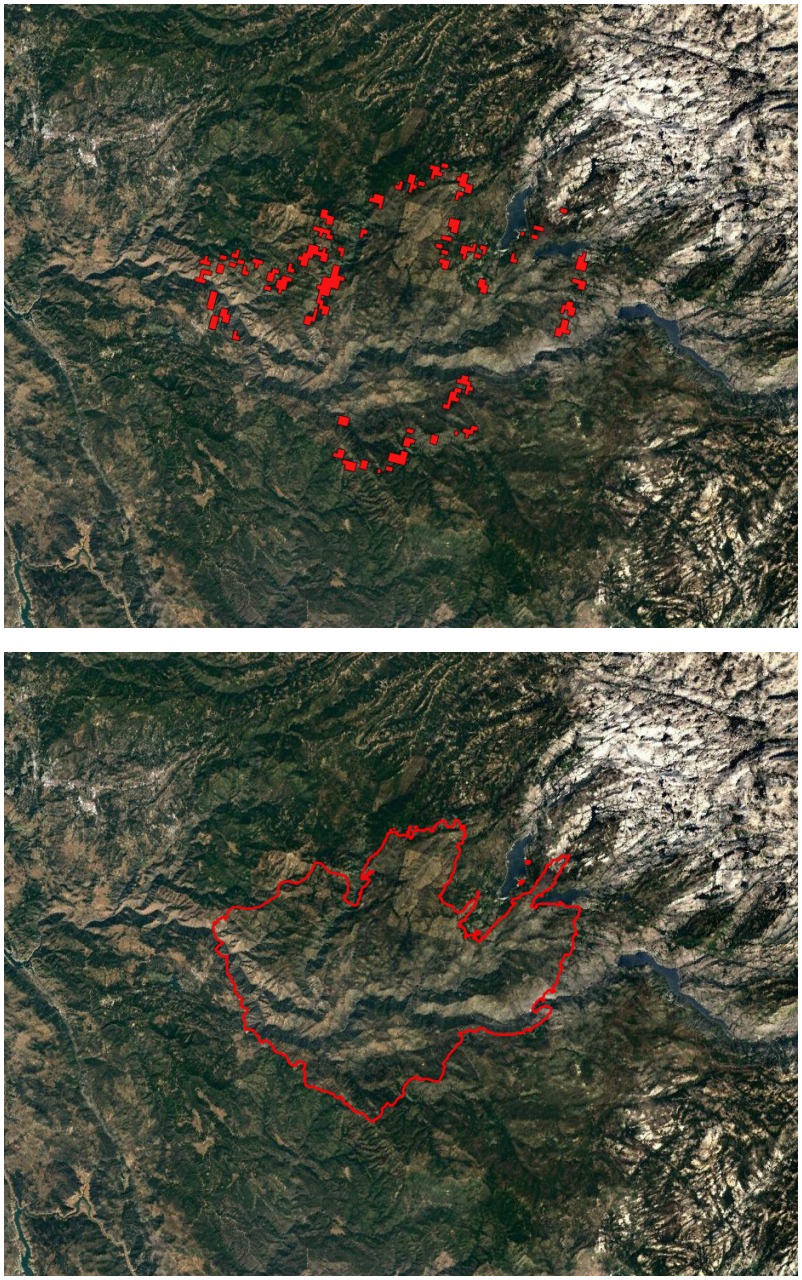

Figure 1.6: Illustration of the available observation data for the Rim fire, a massive wildfire occurred in California in August 2013. Top figure shows the active fires on 23 August 2013 (20:35 UTC) detected by S-NPP VIIRS. Bottom figure shows the fire perimeter data on 23 August 2013 (21:00 UTC) derived from GeoMAC dataset.

the topology of the fire front quite complex and require the data assimilation algorithm to properly address position and topological errors.

There exists a variety of data assimilation schemes featuring different inverse modeling procedures in the literature. The present work builds on past studies (based on an EnKF algorithm, see Sec. 1.3.2) by Rochoux et al. [41, 42] and on the data-driven wildfire spread prototype simulator jointly developed by UMD and CERFACS: This data-driven wildfire spread modeling sytem will be presented in more details in Chapter 2. The main challenges that are addressed in the present work are summarized as follows:

- The first challenge being addressed is to cope with heterogeneous environmental conditions. A new grid-based spatialized parameter estimation approach is developed where the estimation targets are the spatially-varying input parameters of the ROS model. This approach changes the sensitivity of the simulated fire front location with respect to the input parameters along the fire front, especially between the head fire and the flank fires. This gives model simulations more degrees of freedom to find a physically consistent solution. A dynamic distance-based localization scheme is developed to restrict the correction of the control parameters in the vicinity of the propagating fire front and to avoid spurious corrections far away from the fire.

- The second challenge being addressed is to provide is a more efficient and robust method to compute the distance between the observed and simulated fire fronts. The evaluation of such a distance is required by any data assimilation

algorithm in order to give weights to the forecast ensemble simulations and nudge them toward the observations. In past work, we treated the observed fire front as a discretized contour with a finite set of markers. The discrepancy between these simulated and observed fronts was then computed by pairing observed front markers with the same number of simulated markers. However, such pairing may become difficult to operate for regional-scale wildfires that feature strong heterogeneities in the land surface conditions. This issue is addressed by introducing a new front shape comparison method where the burning area is treated as a moving object that can feature both position and topological errors. This method borrowed from object detection in image processing theory formulates a shape similarity measure based on the Chan-Vese contour fitting functional.

- The third challenge being addressed is to develop a new dual state-parameter estimation method to retrieve more physical values of control parameters and therefore improve short-term forecast performance. The state estimation part is achieved with a Luenberger observer (LO) where the simulation counterpart is gradually nudged to the observation as controlled by a weighting factor. The parameter estimation part is done in a similar fashion to grid-based spatialized parameter estimation using EnKF framework. The dual state-parameter estimation method is found to produce more physical posterior parameters compared with standalone parameter estimation method, making it a promising approach in real world wildfire scenarios where both model state and model

parameters suffer inevitable bias and uncertainties.

- Futhermore, the enhanced data-driven wildfire spread simulator is evaluated with these new algorithmic features against large-scale controlled burn experiments and an accidental wildfire event. The validation test against the Fire-Flux I field-scale burn experiment (30 ha) shows that a spatially-distributed parameter estimation can successfully reconstruct a more realistic set of the control parameters and thereby better capture the anisotropy in the observed fire front. The validation test against the RxCADRE S5 burning experiment (4 ha) and the Rim Fire hazard (1041 km^2) show that the global shape comparison method is a promising method that can accurately track observed fire front location and shape. Dual state-parameter estimation method is validated against the RxCADRE S5 experimental dataset and is shown to be a better algorithm candidate to improve short-term forecast performance.

The thesis is organized as follows. First the data-driven wildfire spread modeling system is presented in Chapter 2. Then we proceed to a presentation of the new parameter estimation method intended to address spatial variations of the parameters in large domains (Chapter 3). Next, the new global shape comparison method is introduced in the context of a LO-based state estimation approach, which is able to handle any complex fire front topology (Chapter 4). Next, the dual state-parameter estimation method is detailed in Chapter 5. Finally, in Chapter 6, a summary of the current work and the author's contributions are detailed, with some discussion of future work.

Chapter 2: Data-driven Wildfire Spread Modeling System

Introduction

The present study is a continuation of previous work on data-driven wildfire spread modeling that has led to the development of a prototype wildland fire spread simulator called FIREFLY[1] [40–42, 64]. The past version of FIREFLY features the following main components: a forward model using an Eulerian-based front-tracking solver and a description of the ROS based on Rothermel's formulation; a series of observations of the fire perimeter location; and an inverse model based on an ensemble Kalman filter (EnKF) and a Lagrangian representation of the fire front to compute discrepancies between observations and simulated counterparts. The EnKF algorithm is a statistical data assimilation algorithm in the sense that a statistical sample (or "ensemble") of the inputs is obtained using a Monte Carlo random sampling and results into multiple predictions of the fire front positions through multiple forward model integrations. The differences between the observed and the simulated fire fronts are then translated into a correction of the inputs to the Rothermel-based ROS model or directly of the fire front location. Thus, the inverse model featured a choice between a parameter estimation approach and a state estimation approach.

[1] http://firefly.cerfacs.fr/

FIREFLY has been previously evaluated in a wide series of verification tests using synthetic observations (these tests are called Observing System Simulation Experiments or OSSE). The simulator has also been evaluated in a preliminary validation study corresponding to a small-scale (4 m $\times$ 4 m) controlled grassland fire experiment in which environmental conditions were treated as uniformly-distributed and the predicted fireline positions were compared with observation data [40–42, 64].

Present work aims at improving the components and expanding the functionalities of FIREFLY to build a data-driven wildfire spread modeling system which is capable of real-time wildfire spread predictions at regional scales. The data-driven modeling system is built upon the previous simulator and it has the main components presented in Fig. 2.1. Given the inputs of the fire initial location and environmental conditions (biomass fuel, local meterology, topographical conditions, etc) at time t_0, the forward model, typically a wildfire spread simulator, yields a series of fire front locations at different leading times. These locations could be represented either using Lagrangian front-tracking markers or a two-dimensional progress variable field showing burnt/unburnt area. At a certain time t_1, the observation data become available and is fed into the modeling system, the inverse model computes the difference between the simulation and observed counterpart, and infer the corrections of either model input parameters, or model state. Using the optimized model parameters and model state, the forward model is integrated until a later time t_2 with a better forecast of wildfire spread and behavior. Once new observation data becomes available, the inverse model can perform either parameter estimation or state estimation again by assimilating the observation data and

update the forecast. It is done in a typical sequential data assimilation framework.

In terms of the forward model, a Lagrangian front-tracking model is incorporated into the data-driven modeling system in addition to the original Eulerian level-set front-tracking model. This Lagrangian solver is based on Huygens' principle, which is similar to the fire propagation model adopted in FARSITE [24]. Both Eulerian and Lagrangian forward models use Rothermel's formulation to calculate the ROS in the head fire direction (i.e., the mainstream wind direction), but they differ when calculating the ROS in the flank fire and rear fire directions. The perspective is to have an ensemble that is as rich as possible in terms of simulated fire front shapes. Better performance is expected with multiple fire propagation solvers, especially when using EnKF algorithms, since ensemble members can benefit from the different model components in the spectrum of fire spread models.

For the parameter estimation problem, we adopted a new grid-based spatially-distributed parameter estimation approach to treat heterogeneous environmental conditions. The control variables are the spatially-varying input parameters of the ROS model; they correspond to parameter fields defined at a resolution that is usually coarser than that used by the front-tracking simulator. This approach changes the sensitivity of the simulated fire front location with respect to the input parameters along the fire front, especially between the heading part of the fire and its flanks. This gives model simulations more degrees of freedom to match observations. A dynamic distance-based localization is required to restrict the correction of the control parameters to the vicinity of the propagating fire front and to avoid spurious corrections far away from the fire.

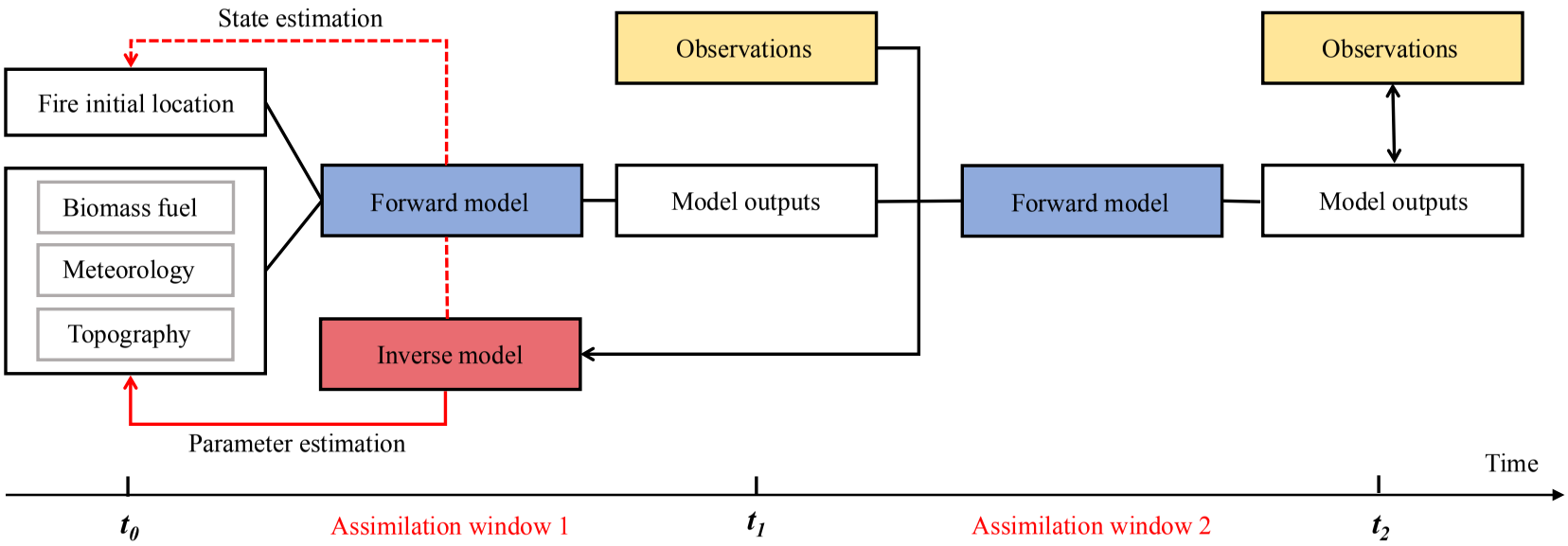

Figure 2.1: A dynamic data-driven modeling system for wildfire spread simulations at regional scales. Each assimilation cycle can be decomposed into two stages: the forecast stage from time t_0 to t_1 when the model is integrated until the observation time; the analysis stage when model optimization is performed by assimilating the observation data, either by state estimation or parameter estimation.

For the state estimation problem, a new deterministic data assimilation algorithm based on Luenberger Observer is developed and tested in the context of regional-scale wildfire spread simulations. The central idea of the LO algorithm is to nudge between simulation and observation based on the confidence level of observation data. We introduced a new shape comparison method in the LO algorithm where we directly consider the burning area as a moving object that can deform under heterogeneous conditions and thus represents the discrepancies between simulated and observed fire state. This method is expected to be more robust when comparing simulated and observed fire fronts, and thus to be better able to handle arbitrary fire front topology.

A new dual state-parameter estimation method is also developed and validated where we combine the EnKF-based parameter estimation and the LO-based state estimation approaches. The objective is to distribute the correction inferred from observations to model parameters and model state properly in order to retrieve more physical model parameters and gain a better forecast at large leading times. This dual state-parameter estimation method alleviates the problem of overcorrecting the model parameters when both the model initial state and model parameters are known with biased information.

2.1 Forward Model

The forward model in our data-driven modeling system is a wildland fire spread solver which takes the fire initial location and a number of environmental factors as

inputs, and produces a series of fire front locations at different leading times. For real-time forecasting purposes, wildland fire spread solvers adopt a regional-scale perspective since they should run faster than real time. These solvers simulate a wildland fire as a two-dimensional interface that self-propagates normal to itself into unburnt vegetation. This interface is usually referred to as a fireline. The depth of the fire front is often neglected; meaning that the fire is treated as an infinitesimally thin front. In this representation, the main quantities of interest are the rates of spread of the fireline (the propagation speed of the fireline, m/s). These quantities are described using empirical or semi-empirical functions related to meteorological and environmental conditions, i.e. information characterizing the fuel (moisture content, surface loading, etc.), weather (wind speed, direction), and terrain (slope, aspect ratio). In present work, the Rothermel's model [22] is used to describe the ROS at the head of the fire with respect to local meteorological and environmental conditions, which is the most widely used model in the United States.

ROS Model

The Rothermel's model is a one-dimensional model that computes the fire propagation rate ROS during wind-aided or up-slope fire propagation scenarios [22]. It assumes that wind-aided and slope-aided ROS is additive and proportional to the default no-wind no-slope ROS. For the default no-wind no-slope ROS computation, it follows the work done by Frandsen [65] and improves the formulation as follows:

$$\mathrm{ROS}_0 = \frac{\xi I_r}{\rho_b \varepsilon Q_{ig}} \tag{2.1}$$

where ξ is the dimensionless propagating flux ratio, which describes the proportion of the flame heat release transferred to the vegetation in the non-flaming zone. I_r [kJ/(min m^2)] is the energy release rate of the combustion. ρ_b [kg/m^3] is the ovendry bulk density, ε is the dimensionless effective heating number. $(\rho_b\varepsilon)$ describes the effective fuel density, i.e., the amount of vegetation per unit volume of the fuel bed raised to ignition ahead of the advancing fire. Q_{ig} [kJ/kg] is the heat of pre-ignition, i.e., the heat required to bring a unit weight of fuel to ignition.

For wind-aided and slope-aided fire propagation (Fig. 2.2), Rothermel assumes two additional terms: Φ_w and Φ_{sl} representing the additive wind and slope effects on the default fire propagation rate, respectively.

$$\mathrm{ROS}_{1D} = \mathrm{ROS}_0(1 + \Phi_w + \Phi_{sl}) \tag{2.2}$$

where ROS_0 represents the no-wind no-slope ROS value represented in Eq. 2.1.

This one-dimensional formulation of the ROS model from Rothermel requires 11 input parameters. The physical quantities involved in Eqs. 2.1 and 2.2 such as the combustion-induced energy release rate I_r, the wind and slope correction coefficients Φ_w and Φ_{sl} were parameterized with respect to these input parameters using the experimental dataset from s series of wind-tunnel experiments. Thus, the one-dimensional ROS model can be expressed as follows:

$$\mathrm{ROS}_{1D} \equiv \mathrm{ROS}_{1D}\left([M_\mathrm{v}, M_{\mathrm{v,ext}}, \delta_\mathrm{v}, m''_\mathrm{v}, \rho_\mathrm{v}, \Sigma_\mathrm{v}, \Delta h_\mathrm{v}, s_t, s_e], \alpha_\mathrm{sl}, \mathbf{u}_\mathrm{w}\right), \tag{2.3}$$

with M_v [%] the fuel moisture (mass of water divided by mass of dry vegetation), $M_{\mathrm{v,ext}}$ [%] the fuel moisture content at extinction, δ_v [m] the fuel depth (vertical thickness of the vegetation layer), m''_v [kg/m^2] the fuel surface loading, ρ_v [kg/m^3]

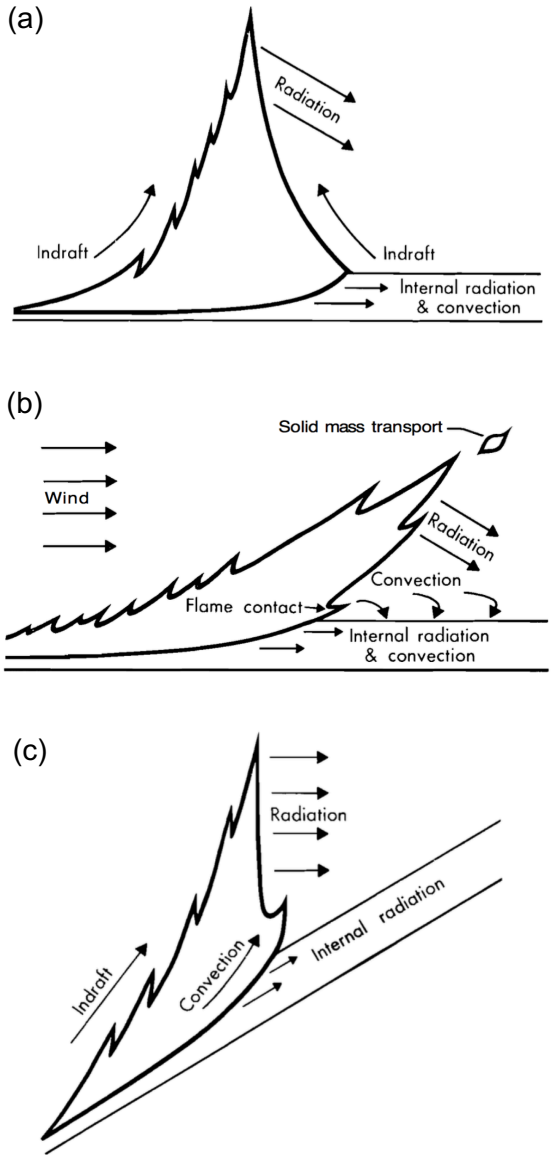

Figure 2.2: Illustration of the wildfire spread for different environmental conditions. (a) no-wind no-slope fire propagation. (b) wind-aided fire propagation. (c) slope-aided fire propagation. [22]

the fuel mass density, Σ_{v} [m^{-1}] the fuel particle surface-to-volume ratio, Δh_{v} [J/kg] the fuel heat of combustion, s_t [%] the fuel particle total mineral content, s_e [%] the fuel particle effective mineral content, α_{sl} [rad] the terrain slope angle and $\mathbf{u}_{\mathrm{w}}$ [m/s] the wind velocity (at mid-flame height). In this list of input parameters, the moisture content at extinction $M_{\mathrm{v,ext}}$, the fuel particle total mineral content s_t and the fuel particle effective mineral content s_e are usually assumed to be independent of the biomass fuel type. Other input parameters can be customized by the user to compute the fuel-dependent ROS values.

Rothermel's model is used for one-dimensional surface fire propagation calculation, it cannot be applied to represent fire spread through a tree canopy (crown fire). For such purposes, models for transition between the surface to the canopy should be used, such as Van Wagner's model [66], followed by adjustments to surface models to account for the drastically different fire spread regimes in crown fuels. In the present work, we mainly focus on the surface fire propagation problem. In the following sections, an extension of the one-dimensional ROS model to a two-dimensional ROS model will be detailed for two different forward modeling approaches (Eulerian and Lagrangian forward models).

2.1.1 Eulerian Front-tracking Solver

The Rothermel's model is used to compute the fire propagation rate in the head fire direction, i.e., the up-wind up-slope direction. There is a need to extend the originial one-dimensional (1D) ROS formulation to two-dimensional (2D) surface

propagation scenario, considering the fire propagation rate in the flank fire and rear fire directions. The 2D extension for Eulerian model is implemented in the same manner as in previous work [67], so here we only provide a brief summary.

To illustrate the extension of the 1D ROS formulation to two dimensions in an Eulerian modeling framework, a geometrical reference frame is defined in Fig. 2.3. First of all, any terrain topography can be locally characterized by the pair of aspect angle and slope angle noted (θ_a, θ_{sl}) in a point-wise manner: θ_a represents the downhill direction, defined in a clockwise representation, where 0° indicates the North direction; θ_{sl} takes values between 0° (flat terrain) and 90° (vertical wall). It is worth noting that even though the terrain topography can be complex in a three-dimensional configuration, the front-tracking problem remains two-dimensional by projecting the ROS onto the two-dimensional horizontal plane. The wind angle θ_w represents the direction from which the wind blows, and is defined on the horizontal plane, starting from the North direction (0°) and in the clockwise direction.

In the context of fire front propagation, the front angle θ_{fr} indicates the outward-pointing normal direction to the fire front. Thus the following equation is used to formulate a two-dimensional ROS [67]

$$\mathrm{ROS}_{2D} = \frac{\mathrm{ROS}_0 \max(1, 1 + \cos[\theta_{fr} - (\theta_w + \pi)]\Phi_w + \cos[\theta_{fr} - (\theta_a + \pi)]\Phi_{sl})}{\sqrt{1 + \tan^2 \theta_{sl} \cos^2 (\theta_a - \theta_{fr})}} \quad (2.4)$$

Assuming the terrain is flat and we only consider the wind effect, this 2D ROS formulation implies that:

- when the fire propagates in the upwind direction (meaning $\theta_{fr} = \theta_w + \pi$), the

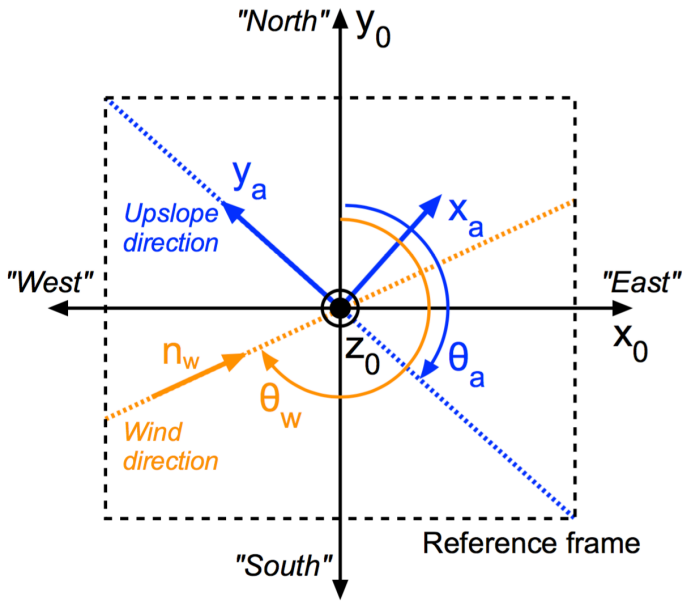

Figure 2.3: Representation of the aspect angle θ_a and the wind angle θ_w on the two-dimensional horizontal plane (x, y). *Credit: Mélanie Rochoux.*

wind contribution to the ROS is maximal;

- when the fire propagates in the normal direction to the wind direction (meaning $\theta_{fr} = \theta_w + \pi/2$), the wind does not affect the fire propagation: the fire propagates at the no-wind ROS;

- when the fire propagates in the opposite direction to the wind (meaning $\theta_{fr} = \theta_w$), the wind correction coefficient is negative and the fire propagates at the no-wind ROS.

Similarly to the wind effect, assuming now that there is no wind and we only consider the terrain slope effect, this ROS formulation implies that:

- when the fire propagates in the uphill direction (meaning $\theta_{fr} = \theta_a + \pi$), the

terrain slope contribution to the ROS is maximal;

- when the fire propagation occurs in the normal direction to the uphill direction (meaning $\theta_{fr} = \theta_a + \pi/2$), the slope does not affect the propagation: the fire propagates at the no-slope ROS;

- when the fire propagates in the downhill direction (meaning $\theta_{fr} = \theta_a$), the slope correction coefficient is negative and the fire propagates at the no-slope ROS.

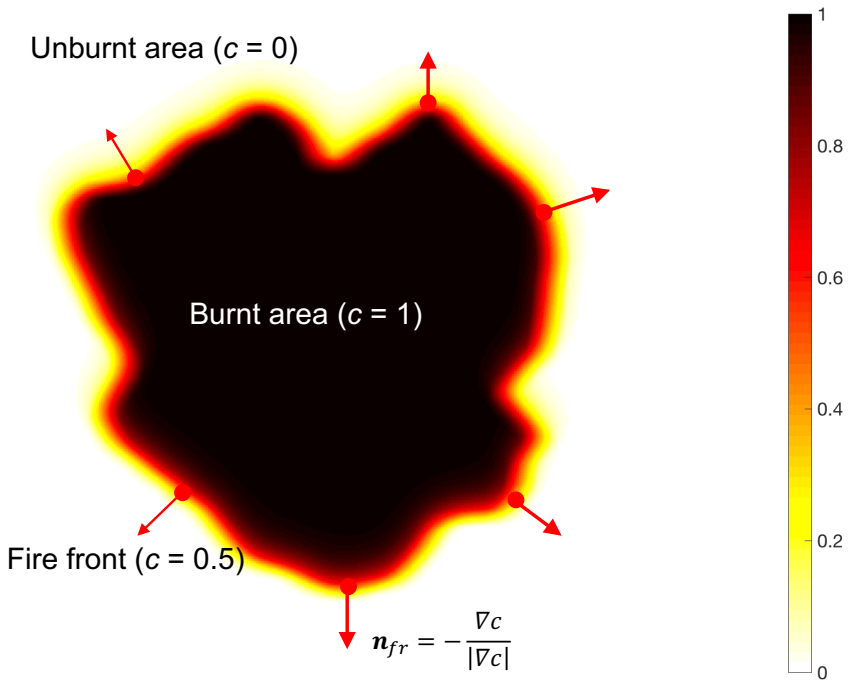

Figure 2.4: Eulerian front representation: the front is represented as the 2-D contour line $\Gamma_c(t)$ corresponding to $\phi_c(x, y, t) = c(x, y, t) - c_{fr} = 0$.

The Eulerian fire spread model in our data-driven modeling system adopts a classical approach taken from the premixed combustion literature, in which a

progress variable $c \equiv c(x, y, t)$ is used as the prognostic variable over the computational domain Ω and is introduced as the front marker: $c = 0$ in the unburnt vegetation, $c = 1$ in the burnt vegetation, and the fire front is identified at a given time as the contour line $c_{fr} = 0.5$ [40–42, 64]. Using this formalism, the active burning areas $\mathcal{B}_c$ are defined as $\mathcal{B}_c = \{(x, y) \in \Omega \,|\, c(x, y, t) > c_{fr}\}$ (Fig. 2.4). The progress variable $c \equiv c(x, y, t)$ is used as the prognostic variable and is calculated as a solution of the propagation equation:

$$\frac{\partial c}{\partial t} = \mathrm{ROS}_{2D} \, |\nabla c| \,, \quad c(x, y, t_0) = c_0(x, y), \quad (x, y) \in \Omega, \; t \geq t_0, \tag{2.5}$$

with $c_0(x, y)$ the initial condition at time t_0 and with ROS_{2D} the local value of the Rothermel-based ROS (Eq. 2.4) defined along the normal direction to the fire front satisfying $\mathbf{n}_{fr} = -\nabla c / |\nabla c|$. Using this formalism, the fire front propagation can be represented by a level-set function $\phi_c(x, y, t) = c(x, y, t) - c_{fr}$, which also satisfies Eq. 2.5. The fire front is represented by the surface $\phi_c(x, y, t) = 0$ denoted by $\Gamma_c(t) = \{(x, y) \in \Omega \,|\, \phi_c(x, y, t) = 0\}$. To solve Eq. 2.5, we follow the choices made by [68] using a second-order Runge-Kutta scheme for time-integration and a second-order total variation diminishing scheme with a Superbee slope limiter for spatial discretization (more details on the numerical solver can be found in [67]).

2.1.2 Lagrangian Front-tracking Solver

Another Lagrangian fire spread solver has also been implemented in the data-driven modeling system, which represents the fire front at a given time, as a finite set of markers (or vertices) located through their coordinates $\{x_{fr}(t), y_{fr}(t)\}$ on the

two-dimensional horizontal plane. We simulate the evolution of the interface as the trajectory of the markers without an underlying grid to represent the state of the system. In the present work, we still use Rothermel's model to determine the ROS, and compute two-dimensional fire propagation using Huygens' principle. This approach is also used in FARSITE [24]: all these Lagrangian fire growth models are inspired by the work done by Gwynfor Richards [69, 70] where we consider the two-dimensional fire shape as ellipsoidal under uniform environmental conditions.

In this approach, a two-dimensional fire is assumed to feature an elliptical shape under uniform conditions (conditions that are rarely found in reality); the ellipse is distorted by the near-surface wind conditions and terrain topography. The fire perimeter is propagated from each marker assuming Huygens' principle and thus orienting an elliptical shape at each time step (any marker on the fire front is considered as an independent source of a new fire of elliptical shape). The shape and direction of the ellipse are determined by a wind-slope vector, while its size is determined by the ROS and the time step length. Thus, the Rothermel-based ROS model is used to predict the spread of the heading portion of the fire, while the spread in all other directions are inferred from the mathematical properties of the ellipse. The fire front is expanded over each time step and can be represented as an envelope of all such individual ellipses around the previous fire front.

As we see in Fig. 2.5, the marker locations at time t represent ignition points that expand ignition over the time step Δt as a small ellipse. The new marker locations at time $(t + \Delta t)$ are obtained from a nonlinear system of first-order differential equations with parameters directly computed from meteorological and environmen-

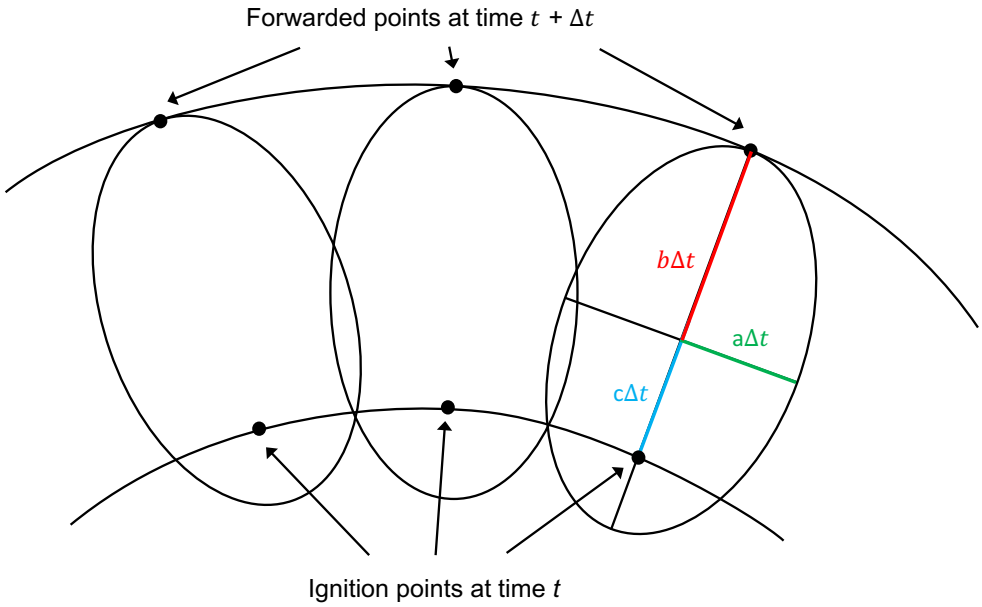

Figure 2.5: The envelope of ellipses forming the fire front at time $(t+\Delta t)$ from ignition points at time t. The parameters a, b and c are model parameters calculated from meteorological and environmental conditions.

tal conditions. Assuming that the model inputs of the environmental conditions are available, first we can calculate the Rothermel-based ROS_0 using Eq. 2.1 and also the dimensionless coefficients representing both local wind effect and slope effect, Φ_w and Φ_{sl}. Following the same techniques implemented in FARSITE, three vectorized quantities can be achieved: the effective mid-flame wind speed U [m/s], the resultant wind-slope vector θ [rad] and the vectorized fire spread rate ROS [m/min] (details can be found in Ref. [24]). The effective mid-flame wind speed U represents the virtual windspeed that by itself would produce the combined effect of slope and wind on the fire spread rate. θ represents the angle of the resultant wind-slope vector for the direction of maximum fire spread on the local slope at a given vertex. Based on the Huygens' expansion method, the length to breadth ratio (LB), the

head to back ratio (HB) of a single ellipse can be computed as:

$$LB = 0.936e^{0.2566U} + 0.461e^{-0.1548U} - 0.397 \tag{2.6}$$

$$HB = \frac{LB + (LB^2 - 1)^{0.5}}{LB - (LB^2 - 1)^{0.5}} \tag{2.7}$$

Using these expressions, the geometrical parameters a, b, c can be computed as:

$$a = \frac{ROS + \frac{ROS}{HB}}{2LB} \tag{2.8}$$

$$b = \frac{ROS + \frac{ROS}{HB}}{2} \tag{2.9}$$

$$c = b - \frac{ROS}{HB} \tag{2.10}$$

The elliptical curve is updated using the angle differentials x_s and y_s which determines the direction normal to the fire front for each vertex $\{x_{fr}(t), y_{fr}(t)\}$:

$$\frac{\partial x}{\partial t} = \frac{a^2 \cos\theta(x_s \sin\theta + y_s \cos\theta) - b^2 \sin\theta(x_s \cos\theta - y_s \sin\theta)}{(b^2(x_s \cos\theta - y_s \sin\theta)^2 + a^2(x_s \sin\theta + y_s \cos\theta)^2)^{0.5}} + c \sin\theta \tag{2.11}$$

$$\frac{\partial y}{\partial t} = \frac{-a^2 \sin\theta(x_s \sin\theta + y_s \cos\theta) - b^2 \cos\theta(x_s \cos\theta - y_s \sin\theta)}{(b^2(x_s \cos\theta - y_s \sin\theta)^2 + a^2(x_s \sin\theta + y_s \cos\theta)^2)^{0.5}} + c \cos\theta \tag{2.12}$$

Equation 2.11 and 2.12 are initially derived by Gwynfor Richards and solved in a predictor-corrector manner [69, 70]. Thus the Lagrangian forward model can output a set of fire front markers $\{x_{fr}(t), y_{fr}(t)\}$ at different leading time. It is worth noting that as the fire front grows and spreads, the distance between two front markers increases and there is a need to add more front markers in between in order to obtain a realistic front shape. In the current model, we define a critical distance based on the initial distance between two successive front markers at the ignition time, then we gradually add more front markers during the simulation once the distance between two successive front markers is larger than the critical distance.

In addition, it is possible that the ellipitical curve crosses over itself and creates some small loops. A loop-clipping filter algorithm has also been implemented to remove these loops.

Despite the sophisticated techniques required to deal with markers interpolation and loop clipping, the Lagrangian model is considered to be computationally more efficient since it does not rely on a 2D grid system, compared with the Eulerian level-set model. However, the level-set model is a very good candidate to model merging fronts process due to the underlying mathematics and thus does not require special treatment as in the Lagrangian model [71]. This feature is important for wildfire spread modeling since often, there are massive spotting fires generated seperately from the main fire.These spotting fires will grow and merge with others. With an Eulerian level-set model, it is natural to model this process.

We incorporated both Eulerian and Lagrangian front-tracking solvers in our forward model. It is of great interest to compare these two modeling approaches when it comes to fire propagation on a two-dimensional terrain. Figure 2.6 shows a comparison between Eulerian and Lagrangian front-tracking simulators in a fire simulation case with moderate southern wind (0.5 m/s). While both simulators provide the same fire propagation in the head fire region, the Lagrangian simulator yields a faster ROS on the flanks, making the shape of the burning area significantly different from the Eulerian simulator. This is due to the modeling assumptions that while both simulators rely on Rothermel's model to compute the ROS in the head fire direction, they use different formulations to propagate the fire in the flank and rear directions. The Lagrangian modeling approach uses a Huygens' wavelet

model which assumes an ellipsoidal fireline shape and correlations for the width of the fireline are used to spread the fire on the flanks [24, 69, 70]. In contrast, in the Eulerian level-set simulator, we modified the formulation of the Rothermel's model for 2-D propagation following choices made by Chris Lautenberger [53]. This formulation involves the wind direction angle and is applied locally according to the local normal vector to the front. Since both models are empirically formulated, we cannot assess which simulator should be prefered in terms of forecast performance. One objective of implementing multiple forward models is to have an ensemble that is as rich as possible in terms of simulated fire front topologies, especially when using ensemble-based modeling procedure. since ensemble members can be built using different simulators and thereby represent a wider spectrum of fire front shapes and positions that are useful to have a representative ensemble.

2.2 Inverse Model

Inverse model aims at finding the model parameters (the "cause") given a set of the observation data, it is the "inverse" to the forward problem. In our current data-driven modeling system, the forward model yields successive fire front locations given an initial fire location and Rothermel-based ROS parameters. The inverse modeling procedure estimates either a selection of these ROS parameters or the initial fire location given the observed fire front locations at the observation time. The discrepancies between the observed and the simulated fire fronts are translated into a correction of the inputs to the Rothermel-based ROS parameters

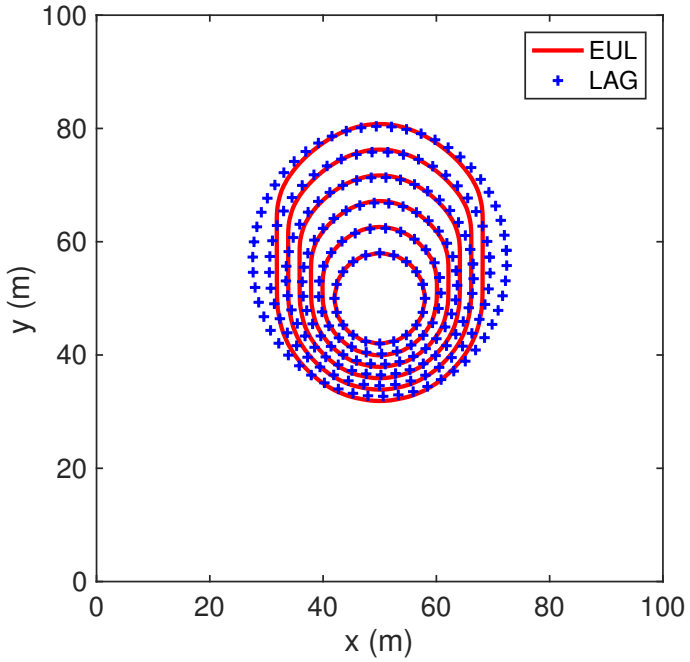

Figure 2.6: Comparison of fire front propagation between Eulerian (EUL) and Lagrangian (LAG) front-tracking simulators. Red solid line represents the fire front line location where $c = 0.5$ in the Eulerian model; blue symbols represents the fire front markers' location in the Lagrangian model.

(parameter estimation) or directly of the fire front location (state estimation), or both (dual state-parameter estimation). A list of model inputs to be corrected forms the control vector, the control vector will be optimized in light of the discrepancy between simulation and observation. It is worth mentioning that the objective of the current data-driven wildfire spread modeling system is to improve wildfire spread forecast using the optimized model inputs, so we rely on the forecast performance to evaluate the effectiveness and goodness of the inverse modeling approach.

Data assimilation is considered as a probabilistic formulation of an inverse model problem, where the uncertainties associated with both the model and the

observation are taken into account. Neither the numerical model or observations can provide a complete and accurate discription of the physical process when used alone, data assimilation can combine these two information, assign weights to them using their individual uncertainties, and infer a more accurate estimate of the model state or parameters. Data assimilation methods were largely developed for operational numerical weather prediction, but they have been applied to an increasing range of earth science disciplines in recent years [32]. There are two main data assimilation methods: variational and sequential methods. The variational methods (3D-Var, 4D-Var) are a mature and experienced technique in atmospheric applications, where such approaches try to find the optimal value of the control vector by minimizing a cost function measuring its distance to the prior simulations and to the observations. The sequential data assimilation methods are mainly based on a Kalman filter algorithm which allows for dynamic error covariances, where new observations are sequentially assimilated into the model when they become available. For nonlinear problems, the extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) algorithms are developed to deal with such conditions which are very common in most data-driven application systems. In recent decades there is a special focus on the ensemble-based data assimilation methods (EnKF and similar variants). These methods have become very popular, both due to their simple implementation and interpretation and their properties with nonlinear models [33]. In this thesis work, we use the ensemble approaches as the main data assimilation technique applied to wildfire spread modeling.

2.2.1 Front Distance Measure

The evaluation of the distance between observed fire fronts and simulated fire fronts is required by any data assimilation algorithm in order to give weights to the forecast simulations and nudge them toward the observations according to the modeling and observation uncertainties. Our current data-driven modeling system adopts two ways to compute the simulation-observation discrepancy.

On the one hand, we use a discretization of both the simulated and observed fire fronts. The discrepancy between these simulated and observed fronts is then computed by pairing observed front markers with the same number of simulated markers. Notice that usually the numerical model resolution is much higher than observation data, thus the number of simulated front markers is much more than the number of observed front markers. A selection operator is required to select the same number of simulated markers in order to pair them with observed markers one by one. In the current model, we use a simple treatment (taking one out of several points as seen in Fig. 2.7); and it provided reasonable results in previous tests [41, 42, 72].

After applying the selection operator, we have a set of N_{fr} simulated markers which are characterized by the following two-dimensional coordinates:

$$\mathbf{y^s} = [(x_1^s, y_1^s), (x_2^s, y_2^s), ..., (x_{N_{fr}}^s, y_{N_{fr}}^s)]$$

Similarly, the observation data is discretized as a set of N_{fr} observed markers; the observation vector $\mathbf{x^o}$ is defined as:

$$\mathbf{y^o} = [(x_1^o, y_1^o), (x_2^o, y_2^o), ..., (x_{N_{fr}}^o, y_{N_{fr}}^o)]$$

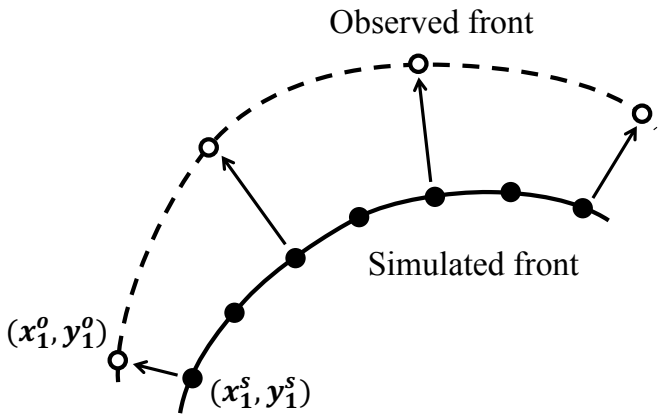

Figure 2.7: Front marker pairing method to quantify the differences between simulated and observed fire fronts.

The distance between simulated and observed fire fronts gives the discrepancy vector $\mathcal{D}$ of dimension $2N_{fr}$ and is simply defined as the vector formed by the direct distance between the paired markers:

$$\mathcal{D} = \mathbf{y^o} - \mathbf{y^s} = \begin{pmatrix} x_1^o - x_1^s \\ x_2^o - x_2^s \\ . \\ . \\ . \\ x_{N_{fr}}^o - x_{N_{fr}}^s \\ y_1^o - y_1^s \\ y_2^o - y_2^s \\ . \\ . \\ . \\ y_{N_{fr}}^o - y_{N_{fr}}^s \end{pmatrix} \tag{2.13}$$

This distance vector based on the Euclidean distance and the marker-pairing method is then used in inverse modeling to infer the corrections of the control parameters.

The marker pairing method has shown to be effective in relatively simple front

topology scenarios [41, 42, 72]. However, such method may become difficult to use for regional-scale wildfires that feature strong heterogeneities in the land surface conditions. Figure 2.8 shows a case where this marker-pairing method becomes questionable when the observed fire front features a complex front topology compared with the simulated fire front. In this case, it is difficult to correctly represent the Euclidean-type distance error by a marker-pairing method, thus the distance measure may become unsuitable when moving to realistic real-world wildfire events.

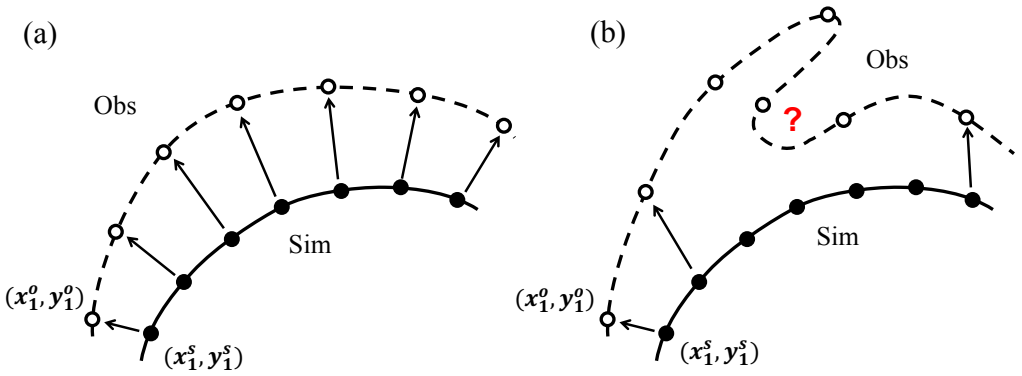

Figure 2.8: Sketch of the marker-pairing method limitations for a case where observed fire front features a simple front geometry (a) and ineffectiveness where observation features complex front geometry (b).

On the other hand, in collaboration with researchers from INRIA and CERFACS though the 2016 CERMACS summer school, we have developed an alternative called global shape similarity measure to represent the observation-simulation discrepancy vector in the data assimilation framework [73, 74]. This new shape comparison method directly considers the burning area $\mathcal{B}_c$ as a moving object that can deform under heterogeneous conditions and thus represents the match (or mismatch)

between simulated and observed fire burning states. It is derived from image segmentation theory and was already adapted in the context of electrophysiology data assimilation [75–77]. The idea of assimilating the observation by comparing two images essentially belongs to the research category of "image assimilation", which has also been a popular topic for atmospheric prediction applications [78, 79].

We now define a shape similarity measure between an observed front and a simulated front using the Chan-Vese contour fitting functional [80, 81]. In a level-set formalism, this measure can be written as

$$\mathcal{J}(\mathrm{y}^o, \phi_\mathrm{c}) = \int_\Omega H_v(\phi_\mathrm{c})\, [\mathrm{y}^o - C_{\max}(\mathrm{y}^o, \phi_\mathrm{c})]^2 + (1 - H_v(\phi_\mathrm{c}))\, [\mathrm{y}^o - C_{\min}(\mathrm{y}^o, \phi_\mathrm{c})]^2 \, dx\, dy \tag{2.14}$$

where y^o is the observation data (for current wildfire application it is a binary field seperating burnt and unburnt area), ϕ_c the level-set function ($\phi_\mathrm{c} = c - c_\mathrm{fr}$), H_v the Heaviside function ($H_v(\phi_\mathrm{c}) = 0$ if $\phi_\mathrm{c} < 0$; $H_v(\phi_\mathrm{c}) = 1$ if $\phi_\mathrm{c} > 0$); and where $C_{\min}$ and $C_{\max}$ are scalar coefficients defined by $C_{\min} = \min(C_0, C_1)$ and $C_{\max} = \max(C_0, C_1)$ with

$$C_0(\mathrm{y}^o, \phi_\mathrm{c}) = \frac{\int_\Omega (1 - H_v(\phi_\mathrm{c}))\, \mathrm{y}^o \, dx\, dy}{\int_\Omega (1 - H_v(\phi_\mathrm{c}))\, dx\, dy}, \; 0 \le C_0 \le 1 \tag{2.15}$$

$$C_1(\mathrm{y}^o, \phi_\mathrm{c}) = \frac{\int_\Omega H_v(\phi_\mathrm{c})\, \mathrm{y}^o \, dx\, dy}{\int_\Omega H_v(\phi_\mathrm{c})\, dx\, dy}, \; 0 \le C_1 \le 1 \tag{2.16}$$

C_1 corresponds to the mean of y^o across the simulated burnt region ($H_\mathrm{v} = 1$) and measures the match between the observed and simulated burnt areas. C_0 corresponds to the mean of y^o across the simulated unburnt region ($H_\mathrm{v} = 0$) and measures the mismatch between the observed and simulated unburnt areas. If the observed

and simulated fronts coincide, $C_1 = 1$ and $C_0 = 0$.

Having defined the discrepancy functional $\mathcal{J}$, our objective is now to propose a distance vector of least square type associated with our front shape similarity measure. The Chan-Vese functional in Eq. 2.14 involves a Heaviside distribution that cannot be multiplied univocally, so it is not a least square criterion [73]. Thus we adopt a pseudo least square strategy by decomposing the Chan-Vese functional into two parts and approximating them by the least square functionals following the choice made in [82]:

$$\mathcal{J}_{\text{data}}(\mathrm{y}^o, \phi_{\mathrm{c}}) = \mathcal{J}^+_{\text{data}}(\mathrm{y}^o, \phi_{\mathrm{c}}) + \mathcal{J}^-_{\text{data}}(\mathrm{y}^o, \phi_{\mathrm{c}}) \tag{2.17}$$

$$\mathcal{J}^+_{\text{data}} = \frac{1}{2}\int_\Omega \Big\{\Big[1 + \frac{2}{\pi}\arctan\Big(\frac{\phi_{\mathrm{c}}}{\epsilon}\Big)\Big]\Big[\mathrm{y}^o - C_{\max}(\mathrm{y}^o, \phi_{\mathrm{c}})\Big]\Big\}^2 \, dx\, dy \tag{2.18}$$

$$\mathcal{J}^-_{\text{data}} = \frac{1}{2}\int_\Omega \Big\{\Big[1 - \frac{2}{\pi}\arctan\Big(\frac{\phi_{\mathrm{c}}}{\epsilon}\Big)\Big]\Big[\mathrm{y}^o - C_{\min}(\mathrm{y}^o, \phi_{\mathrm{c}})\Big]\Big\}^2 \, dx\, dy \tag{2.19}$$

where ϵ is a parameter defined with respect to the contour sharpness [83]. Hence a distance vector can be defined as $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-)^T$ with $\mathcal{D}^+$ and $\mathcal{D}^-$ as follows:

$$\mathcal{D}^+ = \Big[1 + \frac{2}{\pi}\arctan\Big(\frac{\phi_c}{\epsilon}\Big)\Big]\Big[\mathrm{y}^o - C_{\max}(\mathrm{y}^o, \phi_{\mathbf{c}})\Big] \tag{2.20}$$

$$\mathcal{D}^- = \Big[1 - \frac{2}{\pi}\arctan\Big(\frac{\phi_c}{\epsilon}\Big)\Big]\Big[\mathrm{y}^o - C_{\min}(\mathrm{y}^o, \phi_{\mathbf{c}})\Big] \tag{2.21}$$

Here we present the formulation of the discrepancy vector derived from the shape similarity measure that we will integrate in the EnKF algorithm. The advantage of this discrepancy term is that it does not rely on the discretized front markers and is thus suitable for any type of heterogeneous environment and fire behavior. In addition, this distance vector using the new front shape similarity measure can differentiate between "hits", "false alarms" and "misses": the term $\mathcal{D}^+$ highlights the "hits" area where the simulated burnt area matches the observed burnt area,

and penalizes the "false alarms" area where the simulated burnt area is outside the observed burnt area; the term $\mathcal{D}^-$ highlights the "misses" area where the observed burnt area is missed by the simulated burnt area [73]. In Chapter 4, the robustness of this shape similarity measure is demonstrated with a deterministic state estimator based on a Luenberger Observer. In Chapter 5, more verification and validation tests of the shape similarity measure are shown in the context of the dual state-parameter estimation.

2.2.2 EnKF Algorithm

In current data-driven modeling systems, we rely on the ensemble methods to do the estimation work when using parameter estimation (whether in standalone parameter estimation mode or in dual state-parameter estimation mode). Ensemble methods belong to a general class of particle methods where a Monte Carlo ensemble is used to represent the probability density function (PDF) of the control variables. Then the time integration of every ensemble member is processed with the forward model until the observation time. There are different schemes to update the ensemble and describe their PDFs given the observation data. In the present study we rely on ensemble Kalman filter method. EnKF assumes that all PDFs involved are Gaussian, while in practice the Gaussian assumption may not be satisfied well, especially when they are used for highly nonlinear problems. There exist multiple solutions towards non-Gaussianity problems including Gaussian mixture model, multiple Gaussian kernels and other variants of particle filter method [84, 85].

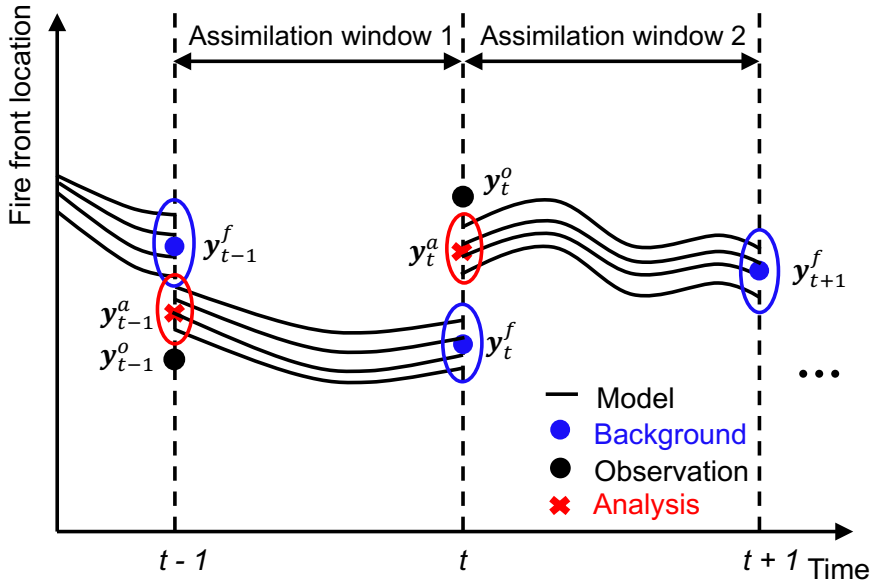

Figure 2.9: Flowchart of the EnKF algorithm applied to two successive assimilation windows $[t-1, t]$ and $[t, t+1]$. For clarity purpose, only 4 model trajectories are represented (instead of a large ensemble).

The control vector includes the uncertain parameters to be estimated over one assimilation cycle. A prior estimates of these control parameters referred to as the forecast $\mathbf{x}^{\mathrm{f}}$ is generated through a Monte Carlo random sampling based on an assumed Gaussian PDF. The EnKF algorithm is sequentially applied over each assimilation window, each sequence corresponding to a two-step procedure (the forecast and analysis steps) triggered by new observation data. Figure 2.9 shows the EnKF flowchart for two successive assimilation windows $[t-1, t]$ and $[t, t+1]$. Assume that we have N_e ensemble members at time $(t-1)$, each member carries one realization of the initial guess of the control parameters $\mathbf{x} = \mathbf{x}^{\mathrm{f}}$ (the forecast ensemble). During $[t-1, t]$, the forward model produces an ensemble of predictions

of the fire front position at time t, designated as $\mathbf{y}_{\mathrm{t}}^{\mathrm{f}} = \mathcal{G}\left(\mathbf{x}^{\mathrm{f}}\right)$, where $\mathcal{G}$ is the observation operator $\mathbf{G}_t$ and describes the mapping from the control parameters to the fire front position at the observed time t. These predictions are then compared to the observed fire front position at time t, designated as $\mathbf{y}_{\mathrm{t}}^{\mathrm{o}}$. Thus, the discrepancy $\mathcal{D}_{\mathrm{t}} = \mathcal{D}\left[\mathbf{y}_{\mathrm{t}}^{\mathrm{o}}, \mathcal{G}\left(\mathbf{x}^{\mathrm{f}}\right)\right]$ provides an estimate of the distance between observations and model predictions. This distance can then be used to update the control variables:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{K}_{\mathrm{t}}\mathcal{D}\left[\mathbf{y}_{\mathrm{t}}^{\mathrm{o}}, \mathcal{G}(\mathbf{x}^{\mathrm{f}})\right] \tag{2.22}$$

$$\mathbf{K}_{\mathrm{t}} = \mathbf{P}_{\mathrm{t}}^{\mathrm{f}}\,\mathbf{G}_{\mathrm{t}}^{\mathrm{T}}\left(\mathbf{G}_{\mathrm{t}}\mathbf{P}_{\mathrm{t}}^{\mathrm{f}}\,\mathbf{G}_{\mathrm{t}}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{2.23}$$

$$\mathbf{P}_{\mathrm{t}}^{\mathrm{a}} = \left(\mathbf{I}_n - \mathbf{K}_{\mathrm{t}}\,\mathbf{G}_{\mathrm{t}}\right)\mathbf{P}_{\mathrm{t}}^{\mathrm{f}} \tag{2.24}$$

where $\mathbf{P}_{\mathrm{t}}^{\mathrm{f}}$ is the forecast error covariance matrix representing errors in the control variables, and $\mathbf{R}$ is the observation error covariance matrix representing observation errors. Thus the new estimates of the control parameters $\mathbf{x} = \mathbf{x}^{\mathrm{a}}$ (the analysis) are defined as a correction to the forecast vector $\mathbf{x}^{\mathrm{f}}$ and the magnitude of the correction is controlled by the Kalman gain matrix $\mathbf{K}_{\mathrm{t}}$. The updated control vector $\mathbf{x}^{\mathrm{a}}$ and ensemble predictions $\mathbf{y}_{\mathrm{t}}^{\mathrm{a}} = \mathcal{G}\left(\mathbf{x}^{\mathrm{a}}\right)$ can then be used to initialize the forecast over the next time period $[t, t+1]$.

The EnKF algorithm does not require the explicit use of an error covariance matrix $\mathbf{P}_{\mathrm{t}}$ and an observation operator $\mathbf{G}_{\mathrm{t}}$. Instead, these matrices can be estimated using the finite number of members in the ensemble:

$$\mathbf{P}_{\mathrm{t}}^{\mathrm{f}}\mathbf{G}_{\mathrm{t}}^{\mathrm{T}} = \sum_{k=1}^{N_{\mathrm{e}}} \frac{\left(\mathbf{x}_{\mathrm{t}}^{\mathrm{f},(k)} - \overline{\mathbf{x}}_{\mathrm{t}}^{\mathrm{f}}\right)\left(\mathbf{y}_{\mathrm{t}}^{\mathrm{f},(k)} - \overline{\mathbf{y}}_{\mathrm{t}}^{\mathrm{f}}\right)^{\mathrm{T}}}{N_{\mathrm{e}} - 1} \tag{2.25}$$

$$\mathbf{G}_{\mathrm{t}}\mathbf{P}_{\mathrm{t}}^{\mathrm{f}}\mathbf{G}_{\mathrm{t}}^{\mathrm{T}} = \sum_{k=1}^{N_{\mathrm{e}}} \frac{\left(\mathbf{y}_{\mathrm{t}}^{\mathrm{f},(k)} - \overline{\mathbf{y}}_{\mathrm{t}}^{\mathrm{f}}\right)\left(\mathbf{y}_{\mathrm{t}}^{\mathrm{f},(k)} - \overline{\mathbf{y}}_{\mathrm{t}}^{\mathrm{f}}\right)^{\mathrm{T}}}{N_{\mathrm{e}} - 1} \tag{2.26}$$

where $\mathbf{x}_{\mathrm{t}}^{\mathrm{f}}$ is the forecast ensemble of the control variables at time t, and $\mathbf{x}_{\mathrm{t}}^{\mathrm{f,(k)}}$ represents one of the N_e ensembles ($k = 1, ..., N_e$). The overline denotes the mean value over the ensemble. After updating all ensemble members, the posterior (analysis) error covariance matrix $\mathbf{P}_{\mathrm{t}}^{\mathrm{a}}$ can also be estimated from the analysis ensemble $\mathbf{x}_{\mathrm{t}}^{\mathrm{a}}$:

$$\mathbf{P}_{\mathrm{t}}^{\mathrm{a}} = \sum_{k=1}^{N_{\mathrm{e}}} \frac{\left(\mathbf{x}_{\mathrm{t}}^{\mathrm{a},(k)} - \overline{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a}}\right)\left(\mathbf{x}_{\mathrm{t}}^{\mathrm{a},(k)} - \overline{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a}}\right)^{\mathrm{T}}}{N_{\mathrm{e}} - 1} \tag{2.27}$$

In Eq. 2.22 the observation data $\mathbf{y}_t^{\mathrm{o}}$ are used to compute the discrepancy between the simulations and the observations. It is known that if the same observations are used to update every ensemble member, the ensemble will systematically underestimate the analysis error covariance [86, 87]. In the standard EnKF formulation, random perturbations were added to the observation data:

$$\mathbf{y}_{\mathrm{t}}^{\mathrm{o},(k)} = \mathbf{y}_{\mathrm{t}}^{\mathrm{o}} + \mathbf{e}^{\mathrm{o},(k)} \tag{2.28}$$

where $\mathbf{e}^{\mathrm{o},(k)}$ is drawn from $\mathcal{N}(0, \mathbf{R})$.

This method can be classified as perturbed observations (or stochastic) EnKF, and has been widely used in the early stages of the EnKF methodology [86, 88–90]. An alternative which avoids the perturbations of the observation is a class of square root (or deterministic) filters [87, 91, 92]. The central idea is to update the ensemble mean and spread instead of updating directly each individual ensemble member. In this formulation, the ensemble mean (single analysis) denoted by $\overline{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a}}$ reflects an estimate of the true control parameters, while the ensemble spread denoted by $\hat{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a}}$ reflects the uncertainty of this estimate, and calculated as $\hat{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a,(k)}} = \mathbf{x}_{\mathrm{t}}^{\mathrm{a,(k)}} - \overline{\mathbf{x}}_{\mathrm{t}}^{\mathrm{a}}$ ($k = 1, ..., N_e$). Once the analysis ensemble mean and spread are calculated, the analysis ensemble can be easily constructed from these two quantities. In our current data-driven

modeling system, an Ensemble Transform Kalman filter (ETKF) algorithm has been implemented, in order to avoid observation perturbations and accompanying sampling error introduced in the perturbation process. In the following, we provide a summary of ETKF algorithm adapted from Ref. [93]. We have left out the subscript t since they are all computed at the same analysis time step.

In the ETKF algorithm, we seek ensemble spreads which satisfy

$$\mathbf{P}^{\mathrm{f}} = \frac{1}{N_{\mathrm{e}}-1}\hat{\mathbf{x}}^{\mathrm{f}}(\hat{\mathbf{x}}^{\mathrm{f}})^{\mathrm{T}} \tag{2.29}$$

$$\mathbf{P}^{\mathrm{a}} = \frac{1}{N_{\mathrm{e}}-1}\hat{\mathbf{x}}^{\mathrm{a}}(\hat{\mathbf{x}}^{\mathrm{a}})^{\mathrm{T}} = \frac{1}{N_{\mathrm{e}}-1}\hat{\mathbf{x}}^{\mathrm{f}}\mathbf{W}(\hat{\mathbf{x}}^{\mathrm{f}}\mathbf{W})^{\mathrm{T}} \tag{2.30}$$

Let $\mathbf{Z}^{\mathrm{f}} = \frac{1}{\sqrt{N_{\mathrm{e}}-1}}\hat{\mathbf{x}}^{\mathrm{f}}$, and then we try to find $\mathbf{Z}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{f}}\mathbf{W}$ so that

$$\mathbf{P}^{\mathrm{f}} = \mathbf{Z}^{\mathrm{f}}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}} \tag{2.31}$$

$$\mathbf{P}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{a}}(\mathbf{Z}^{\mathrm{a}})^{\mathrm{T}} = \mathbf{Z}^{\mathrm{f}}\mathbf{W}\mathbf{W}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}} \tag{2.32}$$

where $\mathbf{W}$ is the transform matrix. From the classical Kalman filter algorithm, we know the analysis error covariance $\mathbf{P}^{\mathrm{a}}$ should satisfy

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{P}^{\mathrm{f}} \tag{2.33}$$

Meanwhile, $\mathbf{K}$ can be written as

$$\mathbf{K} = \mathbf{P}^{\mathrm{a}}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1} \tag{2.34}$$

where $\mathbf{R}$ is the observation error covariance. Combining Eqs. 2.31, 2.32, 2.33 and 2.34 we have

$$\mathbf{Z}^{\mathrm{a}}(\mathbf{Z}^{\mathrm{a}})^{\mathrm{T}} = \mathbf{Z}^{\mathrm{f}}\left[\mathbf{I} - \mathbf{W}\mathbf{W}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\mathbf{Z}^{\mathrm{f}}\right](\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}} \tag{2.35}$$

Compare Eqs. 2.32 and 2.35 we know that one solution is

$$\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{I} - \mathbf{W}\mathbf{W}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\mathbf{Z}^{\mathrm{f}} \tag{2.36}$$

Since $\mathbf{G}\mathbf{Z}^{\mathrm{f}} = \dfrac{1}{\sqrt{N_{\mathrm{e}}-1}}\hat{\mathbf{y}}^{\mathrm{f}}$, we can re-arrange Eq. 2.36 and get

$$\mathbf{W}\mathbf{W}^{\mathrm{T}} = \left[\mathbf{I} + \frac{1}{N_{\mathrm{e}}-1}(\hat{\mathbf{y}}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{y}}^{\mathrm{f}}\right]^{-1} = (N_{\mathrm{e}}-1)\left[(N_{\mathrm{e}}-1)\mathbf{I} + (\hat{\mathbf{y}}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{y}}^{\mathrm{f}}\right]^{-1} \tag{2.37}$$

Therefore we can take

$$\mathbf{W} = \sqrt{N_{\mathrm{e}}-1}\left[(N_{\mathrm{e}}-1)\mathbf{I} + (\hat{\mathbf{y}}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{y}}^{\mathrm{f}}\right]^{-\frac{1}{2}} \tag{2.38}$$

Thus we are able to compute the analysis ensemble spread $\hat{\mathbf{x}}^{\mathrm{a}} = \hat{\mathbf{x}}^{\mathrm{f}}\mathbf{W}$. The analysis ensemble mean $\overline{\mathbf{x}}^{\mathrm{a}}$ is computed via the traditional Kalman gain update using Eq. 2.32, 2.34 and 2.37

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{f}} + \left[(N_{\mathrm{e}}-1)\mathbf{I} + (\hat{\mathbf{y}}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{y}}^{\mathrm{f}}\right]^{-1}(\hat{\mathbf{y}}^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\mathcal{D}(\mathbf{y}^{\mathrm{o}}, \overline{\mathbf{y}}^{\mathrm{f}}) \tag{2.39}$$

where $\mathcal{D}(\mathbf{y}^{\mathrm{o}}, \overline{\mathbf{y}}^{\mathrm{f}})$ denotes the innovation term evaluating the discrepancy between simulated front shape and observed front shape.

EnKF methods rely on an ensemble of model simulations to characterize the mean and covariance of its probability distribution. The accuracy of the sampled mean and covariances are ensured when using an optimal ensemble size. In the present study, we have examined the optimal ensemble size for spatialized parameter estimation method in Chapter 3 and dual state-parameter estimation method in Chapter 5. It is worth noting that the optimal ensemble size is case-specific, whether the selected ensemble size is directly applicable to other cases is unknown.

Besides, in order to compensate for the covariance underestimation issue due to the limited ensemble size, a covariance inflation scheme should be used in practice. This problem occurs in all EnKF algorithms and can lead to filter divergence if not taken care of. There are various inflation schemes in literature, here we adopt the relaxation to background method to inflate the analysis ensemble spread, which also inflate analysis error covariance matrix. To be specific, we use the relaxation-to-prior perturbations (RTPP) scheme which can be found in [94]:

$$\hat{\mathbf{x}}_t^a = (1-\alpha)\hat{\mathbf{x}}_t^a + \alpha\hat{\mathbf{x}}_t^f \tag{2.40}$$

where α ranges from 0 to 1 featuring a small to large relaxation-to-prior perturbations.

Another important aspect of any EnKF algorithm is the treatment of the spurious correlations that appear between the variables that are known to be uncorrelated, or from the sampling error introduced by the finite ensemble size [33]. Covariance localization is primarily used to remove such spurious correlations in practice. In parameter estimation mode, spurious correlations imply that control parameters that are supposed to be uncorrelated with an observation, will experience a small unphysical update. In Chapter 3, a dynamic distance-based localization scheme is implemented to restrict the correction of the control parameters in the vicinity of the front and to avoid spurious corrections far away from the front. Note that this is our first step to implement a localization scheme in fire spread simulation scenario, while future work is needed towards better addressing the spurious correlations.

An important aspect of the EnKF method is that the "analysis" solution relies on the ratio of model error over observation error. Assume the observation error is very large, then the posterior solution after EnKF update will still remain close to the prior (free forecast). In contrast, if the observation error is very small, then the posterior solution will move quite close to the observation data. In the present study, we always assume the observation error is small, thus our objective is to check whether the data assimilation algorithms are able to nudge the simulation to match the observation data after EnKF update. As a premilinary test, we also assume the observation error covariance matrix is diagonal. While such assumption of using a diagonal error covariance matrix may not be entirely representative of the true error structure, future work should consider the spatial correlations of the error characteristics [95, 96].

2.3 Technical Implementations

Our data-driven wildfire spread modeling system has two main components: the forward model and the inverse model. The forward model consists of an Eulerian front-tracking simulator and a Lagrangian front-tracking simulator. The inverse modeling procedure consists of three components: a standalone parameter estimation algorithm, a standalone state estimation algorithm and a dual state-parameter estimation algorithm. The present thesis work is a continuation of previous work that has led to the development of a prototype wildland fire spread simulator called FIREFLY through a collaboration work between UMD and CERFACS. Initially

FIREFLY was implemented in Fortran using the OpenPALM dynamic code coupler[2], so the following spatially-distributed parameter estimation algorithm, which is an extension of the previous FIREFLY system, is also developed in Fortran with the OpenPALM code coupler. Later through the collaboration with researchers from INRIA and CERFACS, we started building a toy model of the current data-driven modeling system in Matlab using the VerdandinMatlab library. The state estimation and dual state-parameter estimation work presented in Chapter 4 and 5 is thus done in a Matlab coding environment. Code implementation in Matlab with the VerdandinMatlab library offers great efficiency in coordinating with other researchers on this project, however we will unify the spatially-distributed parameter estimation based on the OpenPLAM environment and state estimation (as well as dual estimation) based on the VerdandinMatlab library in future.

OpenPALM Coding Environment

The OpenPALM dynamic code coupler has been implemented on the University of Maryland Deepthought2 high performance computing cluster[3]. OpenPALM, developed at CERFACS since 1998, allows for data parallelism as well as for the intercommunication of several programs through the message passing interface (MPI) technology. A coupled component can be launched and release resources upon termination at any moment during the simulation. Computing resources (such as the required memory and the number of concurrent processors) are handled by the

[2] http://www.cerfacs.fr/globc/PALM_WEB/

[3] https://www.glue.umd.edu/hpcc/

OpenPALM driver.

An OpenPALM application can be described as a set of computational units arranged in a coupling algorithm. The different units are controlled by conditional and iterative constructs and belong to algorithmic sequences called computational branches. A branch is structured like a program in a high-level programming language: it allows the definition of sequential algorithms. Inside a branch, the coupled independent programs, the units, are invoked as if they were subroutines of the branch program. It is then possible to define communication points between units from different branches to exchange, send and receive data. OpenPALM applications are implemented via a graphical user interface called PrePALM (Fig. 2.10). In this interface, the programmer initially defines the coupling algorithm: number of components (units), sequential (units within the same branch) and parallel (separated branches) sections, resources management. Then, the actual communications are materialized by points at the top of the unit box (corresponding to received data) or at the bottom (corresponding to sent data); these variables are provided by/to other program units through MPI communications (dashed lines).

In the FIREFLY system, OpenPALM is helpful to exchange data between the fire spread model, the routines processing geophysical data and the mathematical units required for data assimilation. Figure 2.10 presents the PrePALM interface related to a deterministic run of the fire spread model in the data-driven modeling system. There are five branches in total: four branches correspond to the preprocessing of the input parameters, the generation of the terrain topography in yellow, the near-surface wind velocity in blue, the biomass fuel distribution and

properties in green, as well as the fire initial condition in orange. The final branch in red corresponds to the spatial and temporal integration of the fire spread model itself, receiving data from the four other branches. Besides, an important component of the current modeling system is PARASOL, which is written in Tcl language, used to run multiple instances of the same code (task parallelism). In the EnKF framework, this is used to manage the ensemble running of the fire spread solver.

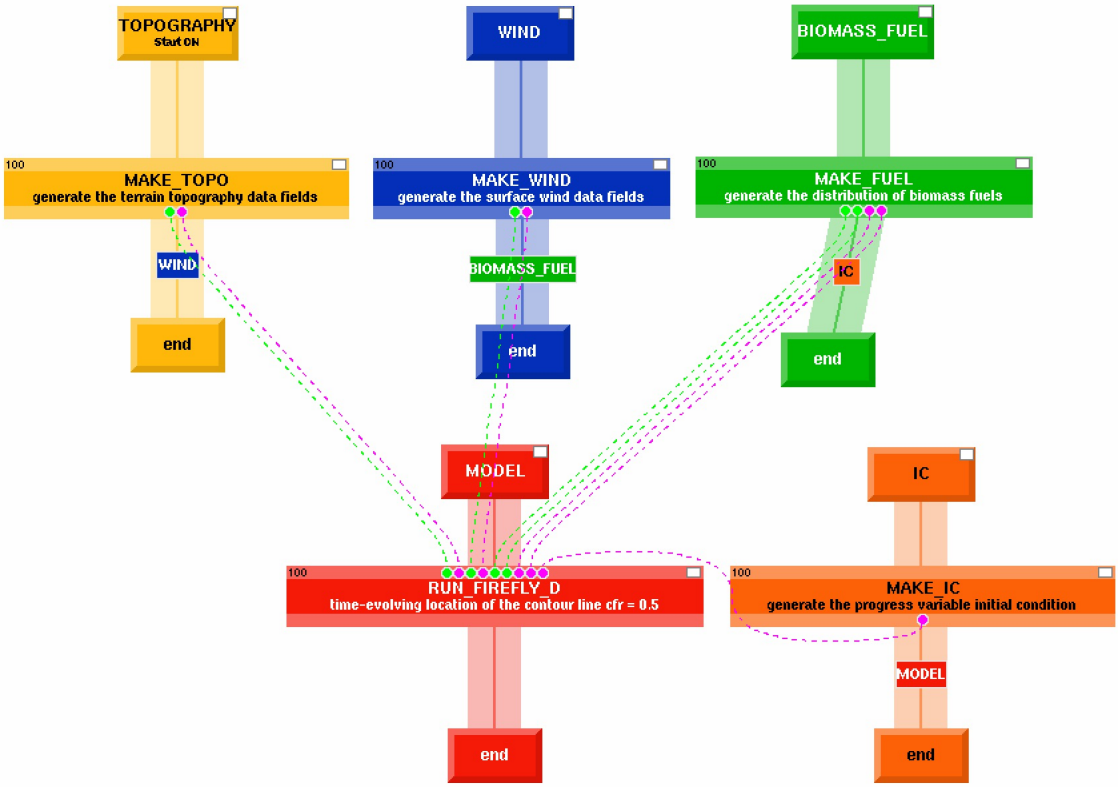

Figure 2.10: PrePALM interface corresponding to the deterministic version of the data-driven modeling system (only forward model, without data assimilation).

Chapter 3: Spatially-distributed Parameter Estimation

Introduction

In a data-driven modeling system, the parameter estimation approach is a way to correct and optimize physical parameters required as inputs to the forward model, in order to account for a significant part of the model bias and uncertainties with a positive effect on the short-to-medium range prediction performance [36]. In wildfire applications, the ROS formulation (Eq. 2.3) requires terrain topography (slope and aspect), vegetation properties (moisture content, fuel type, fuel layer depth, etc) and near-surface wind conditions. These inputs are introduced as a set of parameterizations that correspond to simplifications of the fire dynamics. The optimal value of some of these input parameters is thus intrinsically uncertain. While the terrain topography and the biomass fuel property data could be relatively easy to measure and access, the effect of the local wind on fire propagation remains challenging: a wildfire will create its own meteorological condition near the active flame area, these local wind profiles continue changing the fireline propagation [97]. Current fire spread models do not explicitly account for these interactions. Therefore, it is of significant interest to rely on parameter estimation to retrieve the realistic local wind conditions in order to overcome (at least partially) the lack of fire-atmosphere

interaction representation.

This chapter is dedicated to designing and evaluating a stochastic EnKF strategy to estimate the near-surface wind direction and speed; the evaluation is carried out for the 30-hectare controlled grassland burn trial, FireFlux I. This work has been published in Fire Safety Journal [72].

3.1 Towards Large Scale Problems

Initially FIREFLY adopted a parameter estimation approach using the EnKF algorithm, in which the control parameters were assumed to be uniformly distributed [41, 42]. This assumption is reasonable for small-scale fires. The data assimilation methodology was first evaluated against a 16 m^2, flat, controlled grassland burn experiment (personal communication of the data by King's College London – Ronan Paugam, Martin Wooster). The observations of the fire front positions were derived at a high temporal frequency from thermal infrared imaging; the observation error was supposed to be very low [58]. The estimation targets (i.e. the wind speed and direction, the vegetation moisture content, the vegetation fuel surface-to-volume ratio) were considered uniform; their sequential estimation at 14-s time intervals was found to increase the model forecast performance over several assimilation cycles with a time persistence that was longer than for a state estimation approach (in which the estimation targets are directly the front marker positions). Figure 3.1 shows the performance of the data-driven run. The accuracy of the forecast, while still significantly better than that obtained in the free run simulation

(without EnKF update over the fire duration), rapidly decreases over time. The correction remains valid longer for the parameter estimation approach than for the state estimation approach. However, the variations we see in the heading portion of the observed fire cannot be accounted for when estimating uniform parameters, while the state estimation approach provides enough degrees of freedom to modify the shape of the simulated fire front and to match that of the observed fire front at the update time. These results demonstrate that parameter estimation is required to increase forecast performance; and the parameters need to be spatially distributed to represent the anisotropy in wildfire behavior.

While the initial focus on a small-scale experiment offered the benefits of controlled quasi-uniform environmental conditions, there is an obvious need to extend the FIREFLY validation effort to the case of field-scale experiments and scenarios that are more representative of accidental wildfires, for which we have to deal with highly heterogeneous biomass fuel and winds and with fire behavior that may not be isotropic (the wildfire behavior may significantly change between the head of the fire and its flanks).

To overcome these limitations, we have developed a grid-based spatially distributed parameter estimation approach, whose objective is to control the values of the near-surface wind speed and direction at different spatial locations. Note that this feasibility study focuses on the wind parameters but could easily be extended to biomass fuel parameters such as the fuel moisture content. Note also that the extension of FIREFLY to regional-scale simulation is far from trivial because this extension requires an increase in the number of unknown parameters to treat het-

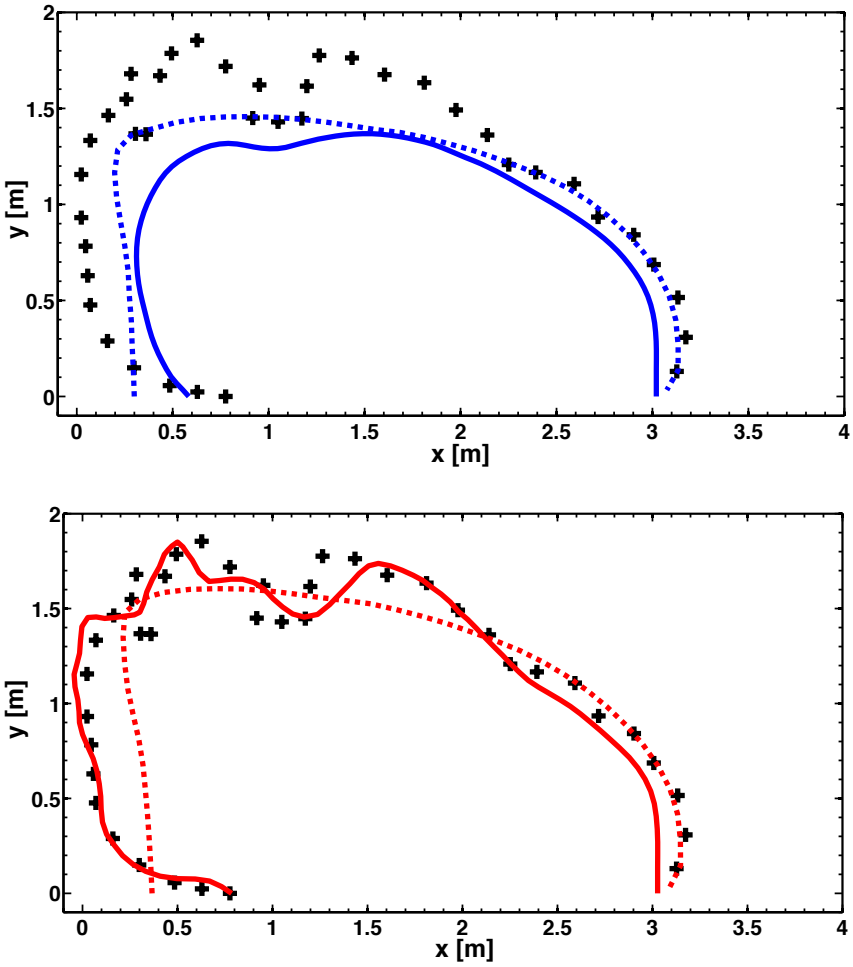

Figure 3.1: Uniformly-distributed parameter estimation approach applied to a small-scale experiment: comparison between simulated (lines) and observed (symbols) front positions at 106 s. The simulated front position is the mean position calculated as the average of the EnKF ensemble; dashed lines (solid lines) correspond to parameter estimation (state estimation). (top) Forecast (with an EnKF update at 92 s and model integration until 106 s) (bottom) Analysis (with an EnKF update at 106 s). *Credit: Mélanie Rochoux* [42].

erogeneous conditions (i.e. spatial variations in biomass fuel and wind), which in turn results in new challenges in the EnKF algorithm as well as an increase in the computational cost. The present chapter shows the results of this new approach on the 30-hectare field-scale controlled burn experiment named FireFlux I and referred

to as FireFlux in the following [54]. Synthetic test cases representative of FireFlux (OSSE), where the ROS is approximately 1 m/s in the wind direction and where the wind parameter estimation is sequentially performed at 100-s time intervals, were carried out in a preliminary study to verify the newly implemented features.

3.2 Control Vector Definition

The control vector denoted by $\mathbf{x}$ gathers all the variables (here the wind speed and direction parameters at different spatial locations) that are corrected by the EnKF. A priori estimate of these control parameters referred to as the forecast $\mathbf{x}^f$ is generated through a Monte Carlo random sampling based on an assumed Gaussian PDF; specific details on the sampling are provided in Sec. 3.3.1. An ensemble of wind direction and speed values is then obtained. Each member of the ensemble is corrected through the EnKF update equation (Eq. 2.22) to formulate the posterior estimate of the control parameters referred to as the analysis $\mathbf{x}^a$ as we discussed previously in Chapter 2.

The control parameters are typically estimated at a much coarser resolution than the computational grid used in the front-tracking solver. First, because all inputs such as meteorological data are usually provided at much coarser resolution. For instance, the finest resolution used by the fire-atmosphere coupled system Fore-Fire/MesoNH is 50 m [28]. Second, if the input parameters were independently estimated at each grid point of the computational domain, the correction would only have an impact where the prior fire fronts propagate during the assimilation

time window (i.e. in-between two successive observation times). Such an approach would limit the persistence of the correction over time and the impact on the forecast quality. Third, if the number of control parameters is much larger than the number of possible fire front topologies, the problem would probably become highly under-determined and computationally heavy.

In the present application to FireFlux, vegetation and terrain topography parameters are still treated as uniform, but wind parameters (magnitude and direction) are now treated as functions of space. These wind spatial variations are described on a uniform rectangular Cartesian grid. In the following, we will present results obtained with a grid featuring 1-m resolution and a wind grid featuring 76-m resolution in the x-direction and 79-m resolution in the y-direction; both wind speed and direction are linearly interpolated between the wind grid nodes. The maximum size of the control vector $\mathbf{x}$ is thus equal to the total number of wind grid nodes (66 nodes) multiplied by two (66 values for the wind speed magnitude and direction); further details on the size of $\mathbf{x}$ are provided in Sec. 3.3.2. The control vector is assumed constant in time over an assimilation window and is only modified when a new observation is available and the algorithm proceeds to the next assimilation cycle.

Note that the wind information manipulated in FIREFLY corresponds to near-surface wind defined at mid-flame height (here at 2-m elevation). Note also that numerical tests with wind grids featuring from 1 to 231 nodes have shown that while the FIREFLY solution is sensitive to the wind grid at resolutions coarser than 100 m, it becomes grid-converged (i.e. independent of the wind grid) at resolutions

finer than 100 m. Our baseline configuration with 66 nodes is thus grid-converged.

3.3 Adaptation of the EnKF Algorithm

3.3.1 Constraints on the Monte Carlo Random Sampling

The generation of an ensemble is an essential part of the EnKF algorithm. It shall include a variety of fire front shapes, positions and topologies to represent all possible scenarios over the fire duration. The ensemble is generated based on perturbations of the control parameters; no perturbation is added to the fire ignition state since we only perform parameter estimation in this case.

There are two main requirements in this perturbation procedure for making it effective [41]: these control parameters need to be uncertain; and the fire front characteristics need to be sensitive to the perturbations in the control parameters. Stated differently, the EnKF algorithm becomes ill-posed when perturbations in the control parameters do not result in observable changes of the system state. A solution to this problem consists in introducing bounds in the search space for the wind parameters. The EnKF algorithm was therefore modified to use truncated Gaussian PDF when generating these ensemble members. In the FireFlux experiment, the surface wind blew mainly from North to South (where we define wind from North as 0°). The bounds introduced in the wind direction are due to a lack of sensitivity of the ROS model to this parameter when the wind blows against the direction of fire propagation (see Fig. 3.2a where wind-opposed spread corresponds to wind angles smaller than -90° or greater than +90°). For the FireFlux case, the search space for

the wind direction was restricted to the South-West/South and South/South-East quadrants. Similarly, the bounds introduced in the magnitude of the wind speed are due to a lack of sensitivity of the ROS model to the wind speed and direction when the ROS takes low values. For the FireFlux case, the ROS of the head fire was close to 1 m/s, which corresponds to a wind speed close to 2 m/s (see Fig. 3.2b); the search space for the wind speed was therefore restricted to values above 1 m/s.

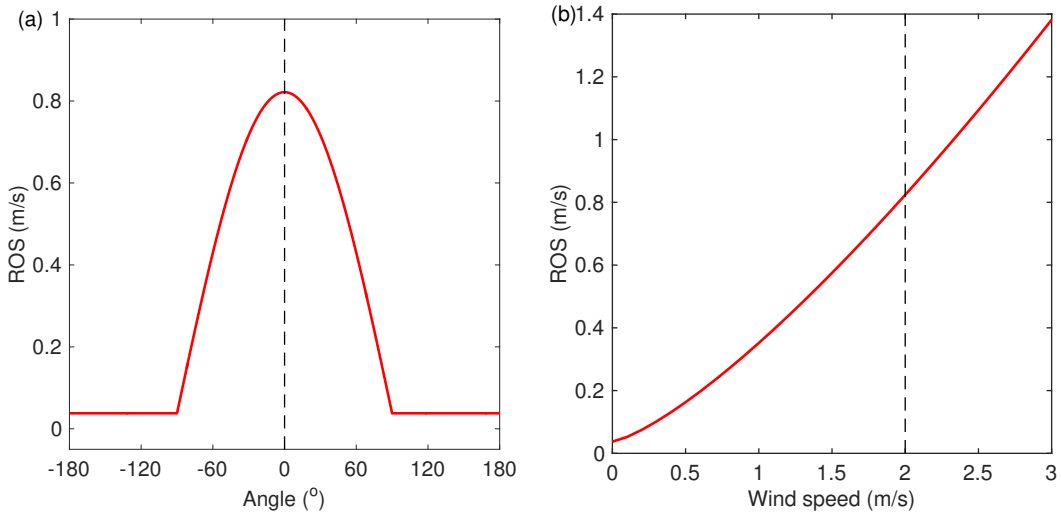

Figure 3.2: Sensitivity of the Rothermel-based ROS model implemented in FIREFLY to the wind angle (angle between the wind direction and the fire spread direction) (a) and wind speed (b). (b) The wind direction is 0°. (a) The wind speed is 2 m/s. In wind-assisted directions, FIREFLY uses a correction to Rothermel's ROS expression based on the cosine of the wind angle; in wind-opposed directions, FIREFLY uses Rothermel's ROS value without wind.

3.3.2 Localization Adapted to Front-tracking Problem

An ill-posed problem is also found when the EnKF algorithm tries to optimize the values of wind speed and direction at wind grid nodes that are remote from

the fire region. In a front-tracking problem, the information is localized in the vicinity of the fire front. Thus, to avoid this problem, we adopt a simple localization methodology in which only the wind grid nodes that are sufficiently close to the fireline (i.e. within a user-specified threshold distance) are included in the control vector $\mathbf{x}$. Thus, the size of the control vector is not constant in time and changes dynamically over successive assimilation cycles. For the FireFlux case, we use a wind grid resolution of 76 m $\times$ 79 m, and a threshold distance approximately equal to twice the wind grid spacing (140 m). Figure 3.3 shows the activated wind points with a threshold distance to the observed fire front equal to 140 m for several observations. Thus, in addition to filtering out noise in the Kalman gain estimation, localization allows a significant reduction in the problem size (starting from a wind grid with 66 nodes, the number of nodes included in the control vector $\mathbf{x}$ remains below 32 during the full FireFlux case).

3.3.3 Cold Start Issue

In FireFlux simulations, at ignition time, the initial guess of the wind parameters (the first forecast) corresponds to a uniform wind field. This initial guess leads to predictions of the fire front position at the first analysis time (100 s) that are far from the observations and that cannot be totally corrected by the analysis step, even when increasing the size of the ensemble N_e. To obtain more optimal fire front positions and overcome this cold start, the EnKF (forecast and analysis) is applied iteratively during the first assimilation cycle. When applying the

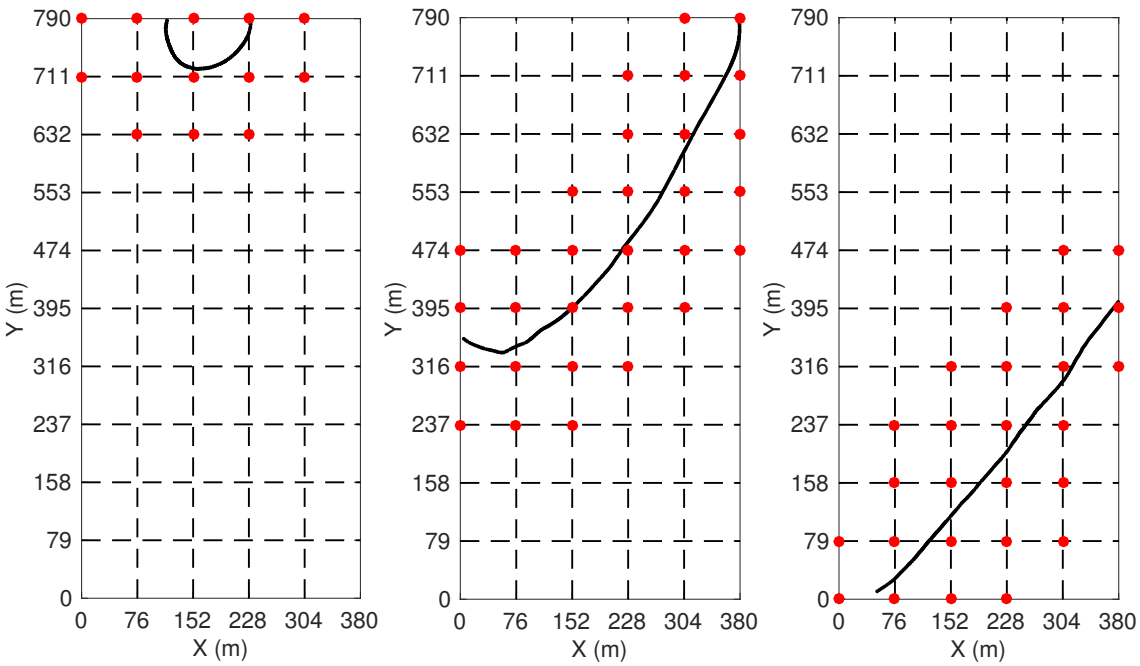

Figure 3.3: Activated nodes (that are part of the wind grid – dashed lines) with a threshold distance to the observed fire front equal to 140 m; several observed fire fronts are represented as black line.

next EnKF iteration, we use the analysis of the previous iteration as initial guess for the next iteration. To be specific, we use the posterior mean value of control wind parameters as the prior mean value for a second EnKF iteration; note that the standard deviation (STD) of control wind parameters is kept unchanged in the second iteration in order to have a robust ensemble spread. We found that only two iterations are sufficient to retrieve a consistent analysis in this case. For the present test cases, this iterative procedure is only applied to the first assimilation cycle and is implemented following the simplified quasi-outer-loop (QOL) algorithm, which is a simplified version of "running in place" (RIP) algorithm proposed by Yang et al [98]. This is a way to address model nonlinearities.

3.4 Verification Test: Observing System Simulation Experiment

We first evaluate the EnKF strategy presented in Sec. 3.3 on a synthetic test (OSSE) that is representative of the FireFlux conditions and that controls the spatially-varying wind speed and direction.

3.4.1 Numerical Settings

In the OSSE framework, synthetic observations are generated using the Eulerian front-tracking solver in FIREFLY and user-prescribed values of the control parameters. This corresponds to a data assimilation case in which the "true" value of the control vector $\mathbf{x}$ exists and is known. The observations are therefore obtained by integrating the forward model using the "truth" $\mathbf{x}^t$ and by perturbing the true fire front positions. The prior estimate of the control vector $\mathbf{x}^f$ is given by directly perturbing the truth $\mathbf{x}^t$. These perturbations are introduced as errors that follow a Gaussian PDF (fully characterized by their mean and STD). In this case, the performance of FIREFLY is simply evaluated by examining its ability to re-construct the true fire behavior and the corresponding prescribed wind field. We focus here on the accuracy of the EnKF analysis with respect to the truth.

In this test, the computational domain size is 380 m $\times$ 790 m, with a spatial resolution of 1 m. Observations are available at time 100 s to be consistent with the FireFlux case presented in Sec. 3.5. Since we focus on the analysis performance, the total simulation time is 100 s, with a time step of 0.02 s. Each simulated fire front is discretized using 100 markers. The EnKF ensemble has 200 members. Synthetic

observations are generated using a spatially-varying wind field created on a (76 m $\times$ 79 m) wind grid, the same wind grid resolution as the FireFlux case. Figure 3.4a shows the true wind field which we use to generate the observations. The threshold value for activating wind grid nodes is 140 m. The prior values of the control parameters are selected using (truncated) Gaussian PDF with an assumed STD of 1 $\mathrm{m\,s^{-1}}$ for wind speed and 30° for wind direction; recall that we consider a uniform wind to initialize the EnKF (see Fig. 3.4b).

Since the observation error STD is small (1 m), a successful EnKF would reduce the spread (i.e. the level of uncertainty) in the analysis ensemble and provide an analysis estimate of the control parameters that corresponds to a mean fire front position that is "closer" to reality (in terms of shape and position). We consider the following performance metrics: the PDF of the wind speed and direction values at the activated wind grid nodes; the PDF of the fire front marker positions at different simulation times; and the distance to the observed fire front. These metrics are calculated for the free run simulation (without data assimilation) as well as for the analysis for the data-driven simulation over the 100-s time window.

3.4.2 Performance of the Two-iteration EnKF Algorithm

Figure 3.5 presents a comparison, at time $t = 100$ s, between the true fire front position (the observation since the observation error STD is small) and different predictions based on FIREFLY: the mean (i.e. average over the EnKF ensemble at a given front marker position) fire front position produced without any analysis,

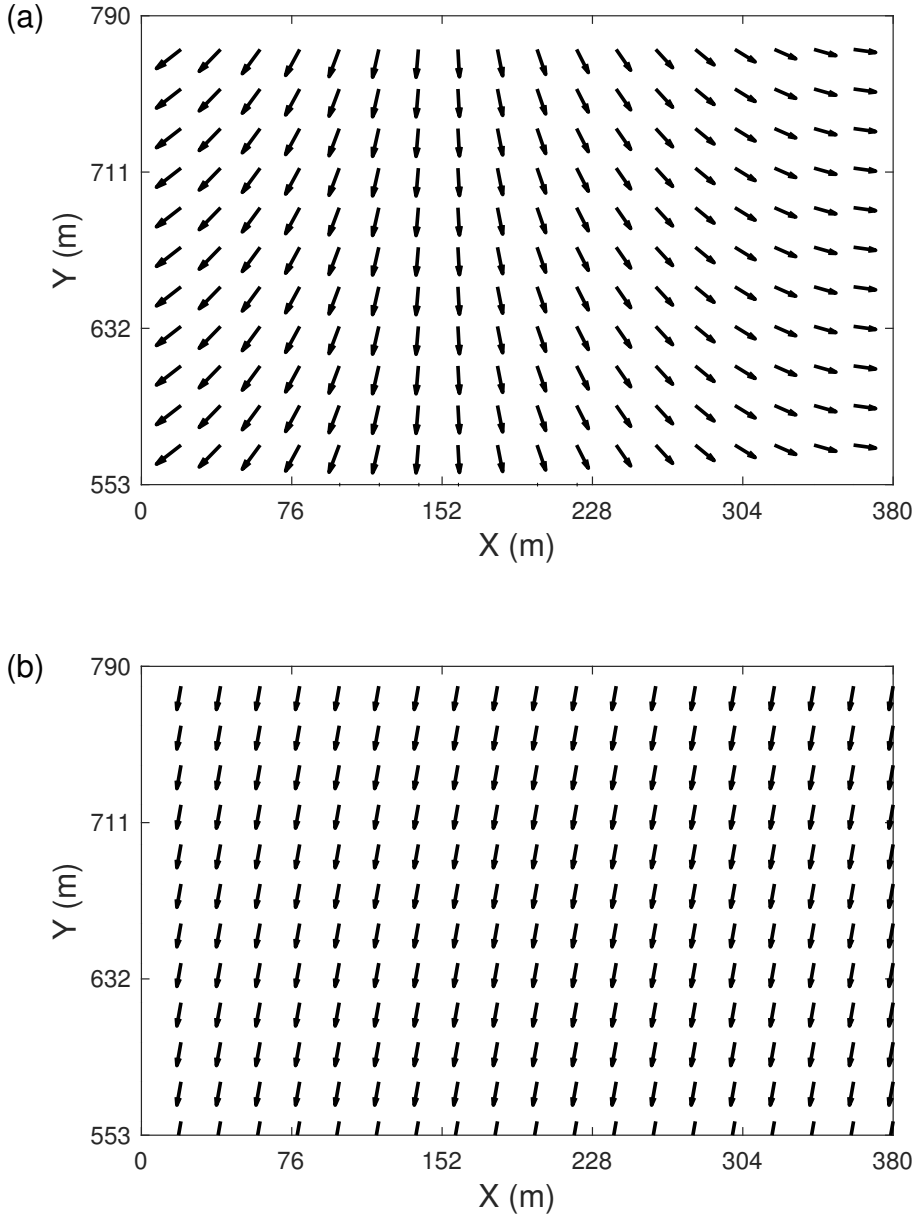

Figure 3.4: Wind field comparison between (a) truth which we use to generate observations and (b) uniform wind field we use to initialize the EnKF.

called the free run or free forecast; the mean fire front position based on an analysis but produced without the two-iteration procedure, called analysis-1; and the mean fire front position based on an analysis with the two-iteration procedure, called

analysis-2. The free forecast features a wrong direction of the head fire. In contrast, the analysis-1 and analysis-2 predictions correctly capture the South-East spread direction. Still, the analysis-1 prediction underestimates the ROS at the head of the fire and overestimates the spread on the eastern flank of the fire. This is further corrected in the second iteration. The analysis-2 prediction is indeed located very close to the true fireline position, thereby demonstrating the benefits of the proposed two-iteration procedure. The first iteration primarily addresses the spread direction error when this is significant, while the second iteration reduces the error in the ROS amplitude.

The differences in fire perimeters observed in Fig. 3.5 are due to differences in the predicted wind field. We examine these differences in Fig. 3.6 in terms of estimates of the wind direction at one particular point located near the head fire, (x, y) = (228 m, 711 m) (marked as point w in Fig. 3.5). This point w is selected since the wind direction at this particular location has the most significant impact on the heading part of the fire front. Figure 3.6 shows at the analysis time 100 s, the PDF of the wind angle at point w, before (top panel) and after (bottom panel) the EnKF analysis. In both plots, the mean estimate of the wind direction is indicated and compared to the true value (-30° – black cross) used to generate the synthetic observations. The prior PDF features an incorrect ensemble mean (8.5° – black circle) and a large scatter (STD of 32°) in which the true value is included. In contrast, the posterior PDF features a reduced scatter by a factor of two (STD of 14.5°) and an ensemble mean (-31°) much closer to the true value. Similar conclusions could be drawn for the wind speed and for other locations in

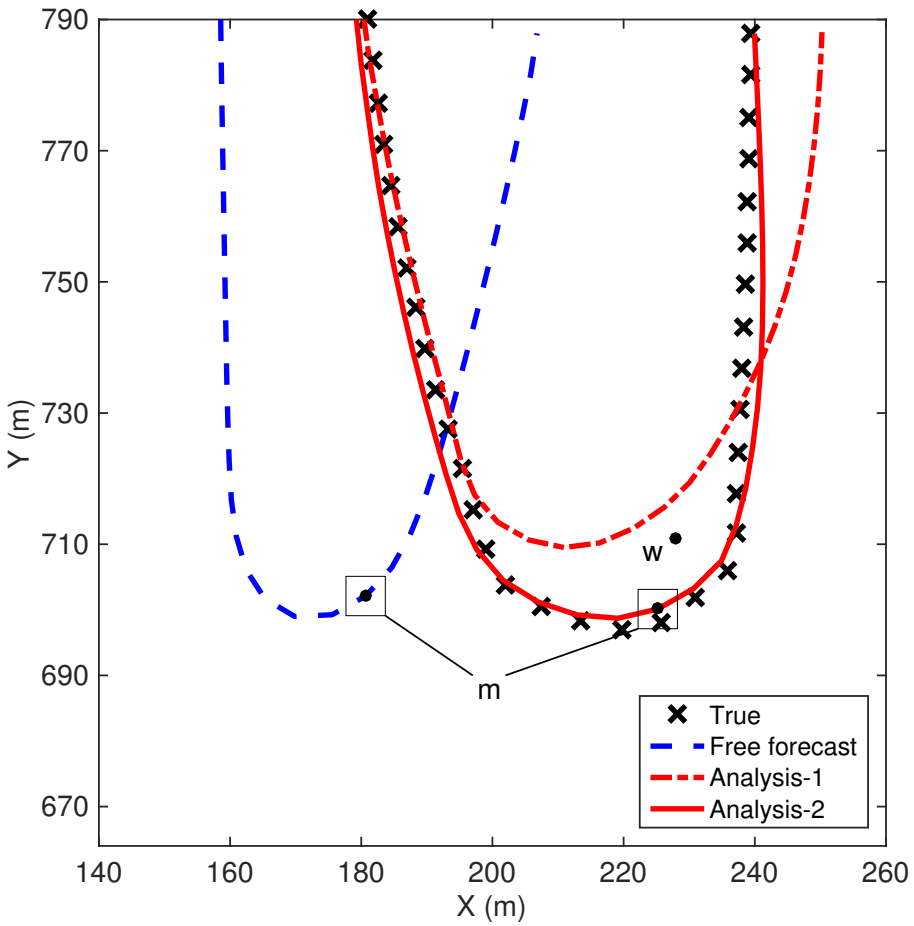

Figure 3.5: Verification OSSE test with synthetic observations and a prescribed spatially-varying wind field. Comparison of fire front positions at time $t = 100$ s: true fireline (cross symbols); mean free forecast (dashed blue line); mean analysis with one iteration, analysis-1 (dashed red line); mean analysis with two iterations, analysis-2 (solid red line). The location of the wind point indexed by w and of the front marker indexed by m are indicated.

the vicinity of the fire front (not shown here).

We also examine how the PDF of the fire front marker positions is changed according to the variations in the wind speed and direction. We focus here on the analysis of the front marker m that is located near the head fire (see Fig. 3.5). Figure 3.7 shows the PDF of the marker m position in terms of x-coordinate (left

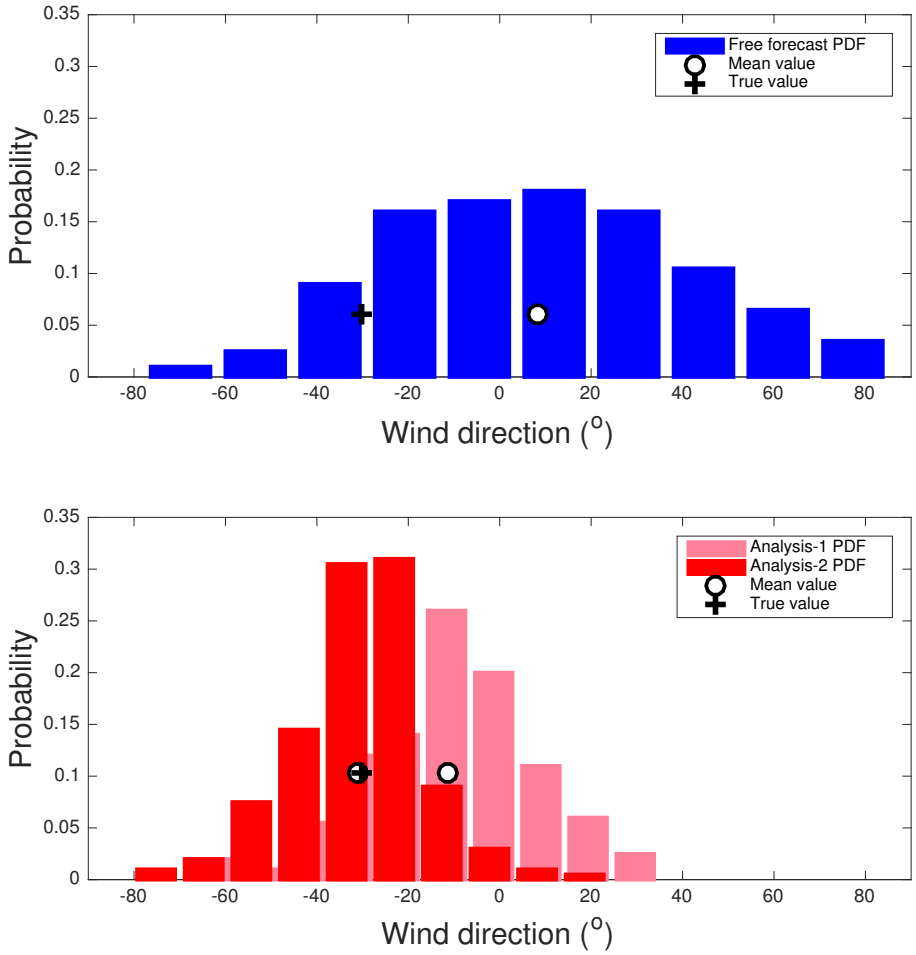

Figure 3.6: PDF of the wind direction at the grid point w over the time window [0; 100 s], associated with the free forecast (without data assimilation – top panel) and with the analysis (after the EnKF update at time 100 s – bottom panel). The wind grid node w is located at (228 m, 711 m), see Fig. 3.5. Histograms are reconstructed from the ensemble; circles correspond to the mean estimate of the ensemble; and crosses correspond to the reference (true value of the wind direction).

panels) and y-coordinate (right panels) at time 100 s. Consistently with Fig. 3.6, Fig. 3.7 shows that the correction of the front marker m is able to correctly track the true front marker with a significantly reduced scatter and reduced bias on the marker position x-coordinate (the mean value of the marker position y-coordinate

is already accurate in the prior ensemble since there are no significant uncertainties in the North-South direction).

In summary, these good results obtained in a synthetic case representative of FireFlux confirm that the new features of the EnKF-based parameter estimation approach (truncated PDF, localization, iterative EnKF) are correctly implemented and valuable. They demonstrate the ability of FIREFLY to match the observed fire front position at the analysis time in a case where observations are assumed to have a very small error.

3.5 Validation Test: Application to the FireFlux Experiment

We now turn to the evaluation of our EnKF strategy presented in Sec. 3.3 on the FireFlux experiment (30-ha) that is more representative of large-scale wildfires and is useful to validate our new spatially-distributed parameter estimation approach. In the present study, the observations are produced by a reference ForeFire/Meso-NH simulation that is assumed to be physically consistent with reality [30].

3.5.1 Overview of the FireFlux Experiment

The FireFlux experiment corresponds to a 30-hectare fire burn [54]. The main biomass fuel was tall grass, which in the modeling, is assumed to be homogeneous. During the experiment, the surface wind blew mainly from North to South (approximately 10° into the South-West direction). The ROS is approximately 1 m/s

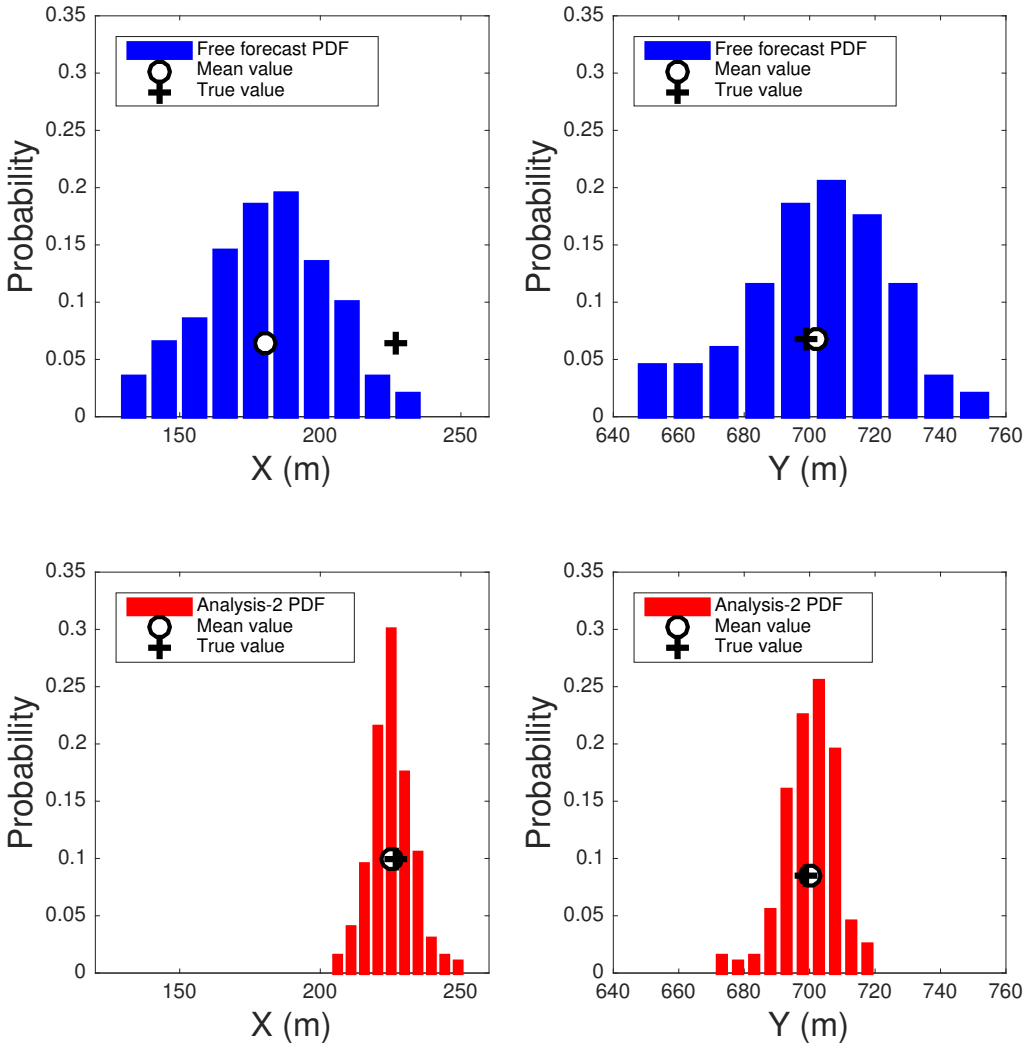

Figure 3.7: PDF of the front marker m (see Fig. 3.5) at time 100 s, associated with the free forecast (without data assimilation – top panels) and with the analysis-2 (after the EnKF update at time 100 s – bottom panels). Left (right) panels correspond to the x-(y-)coordinate of the front marker m. Histograms are reconstructed from the ensemble; circles correspond to the mean estimate of the ensemble; and crosses correspond to the reference (true value of the front marker location).

in the upwind direction; and the fire duration is approximately 15 min. The fire was ignited on the North side of the lot and propagated into the southern direction (see Fig. 3.8). Details on the environmental conditions required to set up the ROS

model in FIREFLY can be found in Table 3.1. FireFlux represents a typical case of wind-driven fire propagation over a flat terrain.

Table 3.1: Input parameters of Rothermel's ROS model corresponding to the FireFlux conditions. These parameters are treated as invariant (in space and time) except for U_w and α_w that are the control variables.

Parameter name (unit)	Symbol	Value
Fuel depth (m)	δ_v	1.5
Fuel loading (kg/m^2)	m''_v	1.08
Fuel moisture content (dead fuel) (%)	M_v	9
Fuel heat of combustion (J/kg)	Δh_c	1.543e+7
Fuel particle surface-to-volume ratio (m^{-1})	Σ_v	5000
Fuel particle mass density (kg/m^3)	ρ_p	400
Fuel moisture content at extinction (%)	$M_{v,ext}$	30
Wind speed (at mid-flame height) (m/s)	U_w	2.0
Wind direction (at mid-flame height) (°)	α_w	10

3.5.2 Observation Data

The FireFlux experiment was originally proposed as a validation experiment for coupled fire-atmosphere models and it has been used by a number of research groups as a benchmark test for Computational Fluid Dynamics (CFD) model validation. One limitation is that the instrumentation was mainly dedicated to the analysis of the smoke plume rather than the fire spread: for instance, the time evolution of the fireline was not tracked. Still, the arrival time of the fire front was recorded at two instrumented towers; this information was used by Filippi et al. [30] to establish the accuracy of ForeFire/Meso-NH simulations when configured in a

two-way coupling mode (ForeFire is a 2-D front-tracking solver used to represent the fire propagation at the land surface and Meso-NH is a 3-D meso-scale atmospheric solver used to account for the vertical and horizontal wind patterns of the turbulent atmospheric flow). The two-way coupling mode accounts for the effects of the fire-induced micrometeorology and was shown in [30] to be an important factor in providing good agreement between numerical results and experimental data. As mentioned in the introduction, we use the ForeFire/Meso-NH fire perimeters produced by Filippi et al. [30] as a surrogate for experimental observations. In the ForeFire/Meso-NH simulations, the horizontal resolution is 10 m and the vertical resolution at ground level is 3 m. We consider observations at 100-s time intervals over a period of 800 s. Figure 3.8 shows the full series of available observations; we can notice an acceleration of the fire spread between 200 and 400 s due to the interactions between the fire and the near-surface wind.

3.5.3 Numerical Settings

The Eulerian front-tracking solver in FIREFLY uses a computational domain that covers the size of the grass area burnt during the experiment, and is 380-m long in the x-direction (the West-East direction) and 790-m long in the y-direction (the South-North direction); the total size of the computational domain is therefore 380×790 m^2 or 30 hectares (ha). The spatial resolution used by the solver is 1 m. Initial conditions for the fireline are treated in an approximate way using the observed fireline shortly after ignition time (14-m wide in the x-direction and 4-m

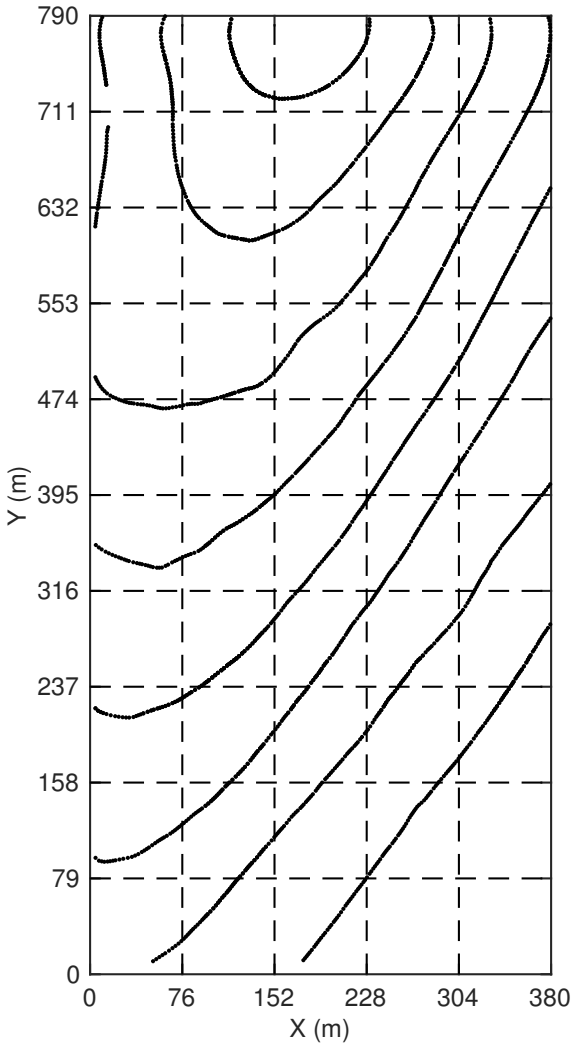

Figure 3.8: Available observation data (solid lines) produced by a reference ForeFire/Meso-NH simulation at 100-s time intervals between 100 and 800 s. The fire spreads from top to bottom (North to South). The dashed lines correspond to the wind grid (76 m $\times$ 79 m).

deep in the y-direction) and without accounting for the fact that in the experiment, fire ignition was set by two walking fire crews [30]. The total simulation time is 800 s. The temporal resolution of the fire propagation solver is 0.02 s.

As in the previous OSSE test case, the control vector is made of the wind direction and speed at selected grid nodes on the 76 m $\times$ 79 m wind grid (the threshold

distance for activation is 140 m). We start with a wind field of uniform values at all wind grid nodes. The initial wind speed has a mean value of 2 $\mathrm{m\,s^{-1}}$ with a STD value of 1 $\mathrm{m\,s^{-1}}$; the initial wind direction has a mean value of about 10° with a STD value of 30°. These values are selected based on rough estimates of uncertainties as well as with the intent to avoid the classical problem of ensemble collapse [99]. The EnKF ensemble has 200 members. The wind values for all ensemble members are then generated using (truncated) Gaussian PDFs corresponding to these mean and STD properties. These settings are identical to the OSSE test case (see Sec. 3.4.1).

For the observation data, we discretize the observed fireline using 100 equally spaced markers. We adopt the previously discussed Lagrangian treatment to evaluate the discrepancy between simulated and observed fire fronts by calculating the Euclidean distance between the 100 observed markers and the 100 simulated markers that are paired one by one (see Sec. 2.2.1). The observation error is assumed small (1 m) and independent for each fire front marker.

The 800-s simulation time is divided into eight 100-s-long data assimilation cycles: the cycles are characterized by seven analysis events, called A1–A7 with Ai designating an analysis performed at time (i × 100 s) based on an observation made at the same time, and by seven forecast events, called F1–F7 with Fi corresponding to a forecast performed at time ((i+1) × 100 s) based on the analysis Ai. The quality of the analysis is evaluated through comparisons between predicted and observed fireline positions at time (i × 100 s); the quality of the forecast is evaluated through similar comparisons at time ((i+1) × 100 s). For each assimilation cycle, we evaluate FIREFLY's performance first by comparing the ensemble average of the fire front

positions (mean values of each front marker over the 200 members) to the observed fire front and second by calculating a measure of the "distance error".

3.5.4 Performance of Spatially-distributed Parameter Estimation

Figure 3.9 presents a comparison between model predictions and observations at analysis and forecast times and in terms of the mean fireline position. Figure 3.9a presents the mean free forecast estimate, i.e. without data assimilation; Fig. 3.9b and Fig. 3.9c present the mean analysis and forecast estimates, respectively, based on data assimilation. The results in Fig. 3.9a illustrate the poor accuracy of the forward model in FIREFLY, primarily due to an incorrect description of the fire spread on the flanks. Also, while the observations indicate that the fire accelerates between 200 and 300 s, the model incorrectly predicts a constant ROS in the wind direction.

The results are much improved with data assimilation. The results in Fig. 3.9b evaluate the quality of the EnKF: they demonstrate the ability of FIREFLY to match observed fireline positions at analysis times and thereby confirm the general ability of data assimilation to steer an inaccurate model towards observations. The results in Fig. 3.9c evaluate the quality of the forecast (performed 100 s after the EnKF analysis). Due to both model inaccuracies and input parameter uncertainties, the predicted fire fronts are found to deviate from the observations and to under-estimate the observed ROS at the head of the fire. The magnitude of these deviations, however, remains moderate and is much smaller than that observed without data

assimilation (Fig. 3.9a). We will show below that the magnitude of the deviations increases in time, and remains acceptable in Fig. 3.9 because of the availability of observations at relatively short time intervals (100 s). In other words, the quality of the forecast depends on the frequency of the observations and of the corresponding updates of the model.

As mentioned above, the data-driven predictions (forecast and analysis) are based on an EnKF statistical ensemble of 200 members and while we choose in Fig. 3.9 to demonstrate the quality of these predictions by plotting a single fire perimeter (the mean perimeter), it is also worth considering the entire PDF of 200 perimeters and the associated deviations from the mean. Figure 3.10 presents a representative sample of 10 fire fronts plotted with the observed fire front at analysis time A1 ($t = 100$ s), corresponding to the free forecast in Fig. 3.10a and to the analysis in Fig. 3.10b. In Fig. 3.10a, the predicted fire perimeters deviate significantly from the observations and feature a large scatter. In contrast, in Fig. 3.10b, the analyzed fire perimeters are close to the observations and feature a much reduced scatter. Because of this reduced scatter, we adopt in the following the viewpoint of Fig. 3.9 and use the mean fire perimeter to quantify the performance of FIREFLY.

Figure 3.11a shows the time variations of the FIREFLY error defined as the average distance between the predicted and observed fireline positions. This average distance is evaluated through an approximate treatment in which at any given time, the predicted and observed fronts are first discretized by the same number of markers, called P- and O-markers, the P- and O-markers are then paired together, and the FIREFLY error is finally calculated as the average distance between paired

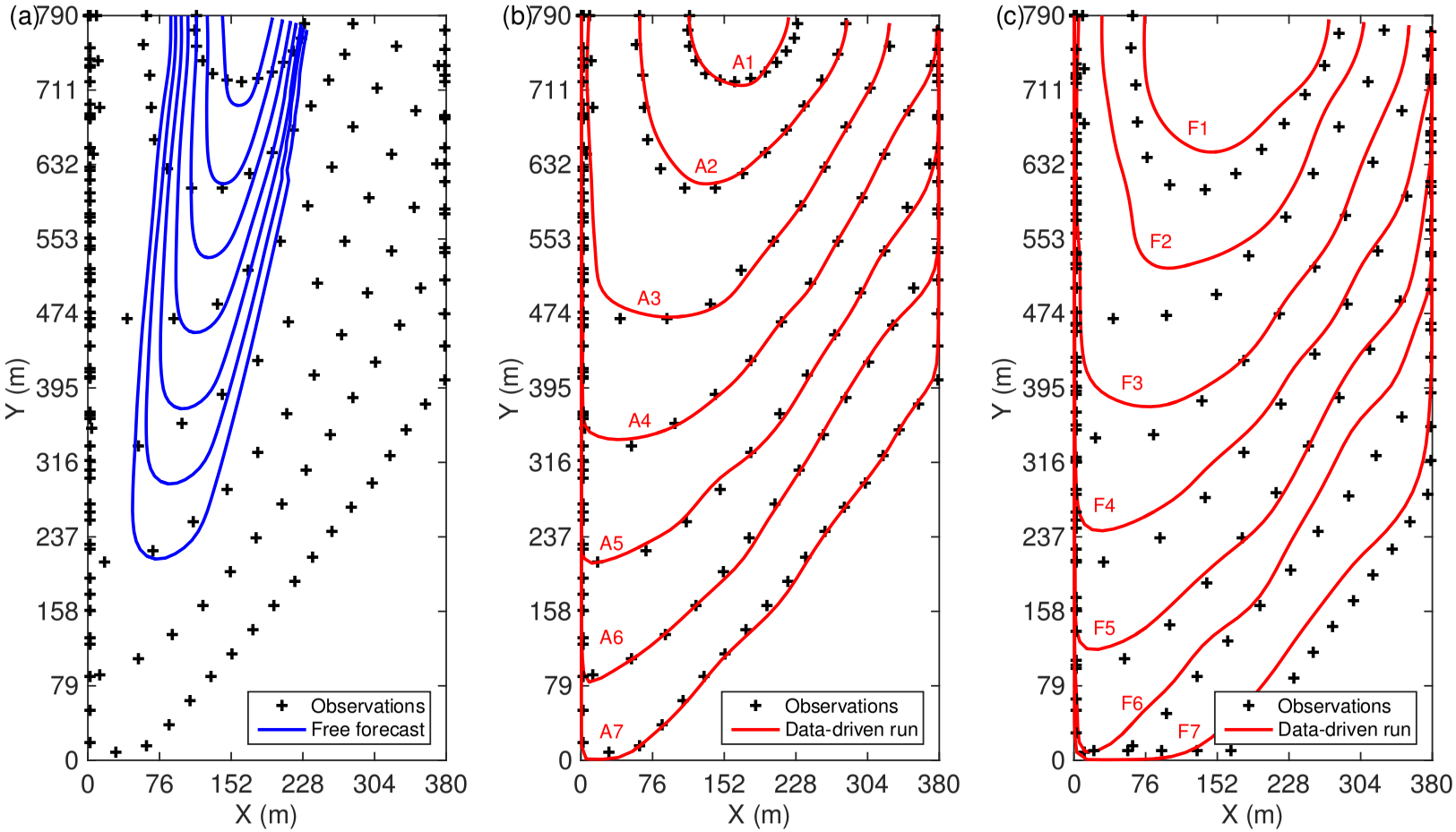

Figure 3.9: Comparison of predicted (solid lines) and observed (cross symbols) fire front positions for FireFlux. (a) The predictions correspond to the mean free forecast ($100 \leq t \leq 700$ s). (b) The predictions correspond to the mean analysis performed at events A_1-A_7 ($100 \leq t \leq 700$ s). (c) The predictions correspond to the mean forecast performed at events F1-F7 ($200 \leq t \leq 800$ s). The fire spreads from top to bottom (North to South). The time interval between plotted perimeters is 100 s.

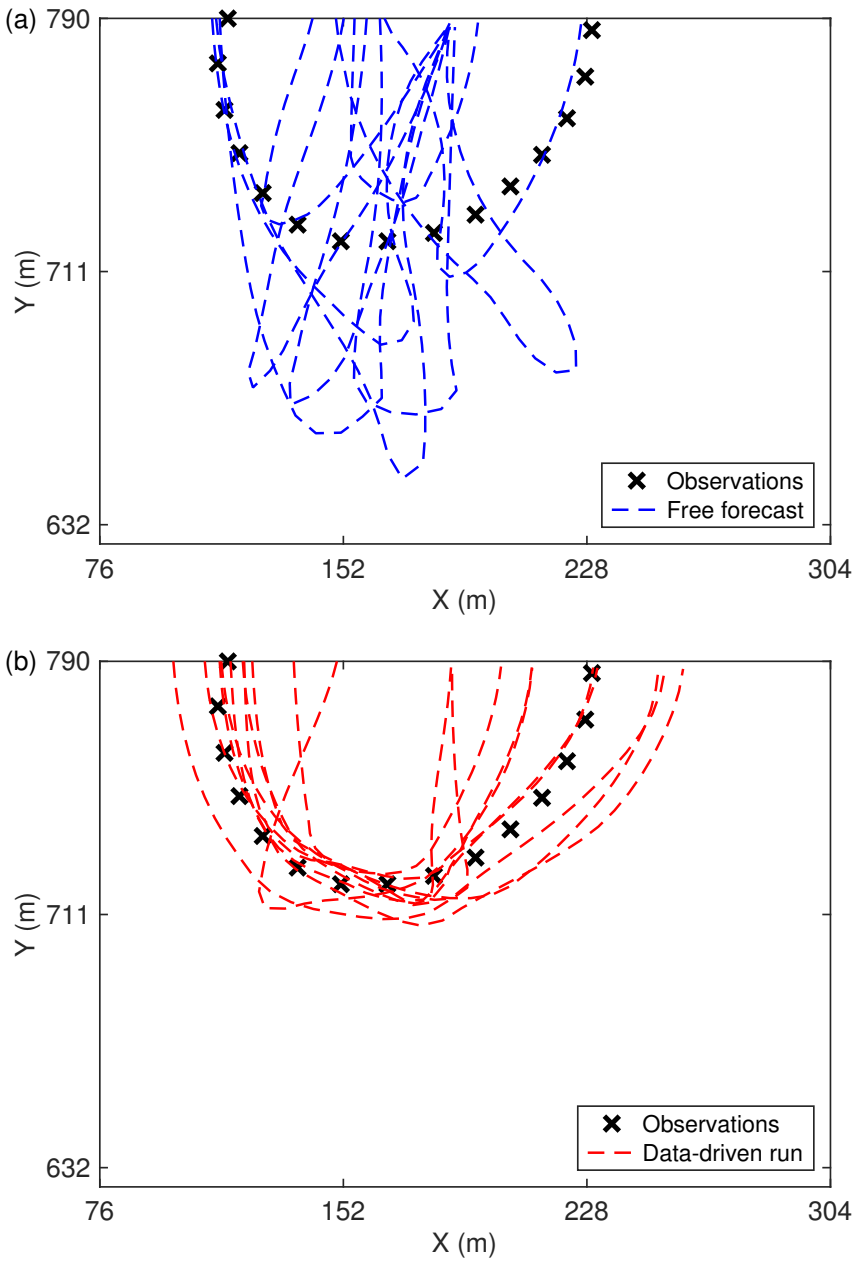

Figure 3.10: Comparison of predicted (dashed lines) and observed (symbols) fire front positions at $t = 100$ s. The predictions correspond to a representative sample of 10 EnKF ensemble members (corresponding to different possible values of the wind parameters) for (a) a free forecast and (b) an analysis.

P- and O-markers. Figure 3.11a compares the evolution of the distance error in the free forecast and data-driven modes. The evolution of the error in the free forecast mode is a smooth, continuously increasing function of time. Consistent with the observed fire propagation acceleration between 200 and 300 s (Fig. 3.9), the error grows at a faster rate (twice faster) during that period. This error takes very large values (more than 150 m at time $t = 600$ s) primarily due to the incorrect description of the fire spread on the flanks. In contrast, the evolution of the error in the data-driven mode is a discontinuous function: deviations of model predictions from observations are periodically reduced (to less than 10 m) during the analysis events A1–A7. After each analysis event, the error increases but remains bounded and takes small-to-moderate values: the error features a peak value of 40 m at time $t = 300$ s (due to changes in the fire dynamics) and decreases to less than 10 m (after 600 s, the fire can be considered as a flank fire and is therefore easier to track).

Figure 3.11b presents a slightly different perspective that illustrates the effect of the assimilation frequency by comparing the evolution of the distance error between the following variations: the free forecast; a first (second; third) data-driven curve, called DD1 (DD2; DD3), with analysis at time $t = 100$ s ($t = 100$ and 200 s; $t = 100$, 200 and 300 s) and a free forecast thereafter; and a fourth data-driven curve, called DDB, corresponding to our baseline case with model updates at all analysis events A1–A4. The growth of the forecast error is slower for DD3 and DDB than for DD1 and DD2, confirming that the fire intensity decreases after its peak around 300 s and that the forecast is becoming more reliable at late times. At time $t = 500$ s, the distance error of the free forecast is approximately 140 m; the error

decreases to 115 m, 75 m, 30 m and 10 m, in cases DD1, DD2, DD3 and DDB, respectively.

The information in Fig. 3.11b is particularly valuable for the design of a data-driven model. Let us assume for instance that the design objective is to predict the fire location with a 50-m accuracy: the results in Fig. 3.11b suggest that the model requires data assimilation and that the assimilation frequency should correspond to observations made at least every 200 s. Lower assimilation frequencies will result in less accurate predictions of the fireline location. We conclude here that a 50-m accuracy on the fireline location requires relatively frequent observations made (approximately) every 3 min. Note that this conclusion is specific to the FireFlux experiment: it will be further examined in the future for other burns. Note also that in addition to being decreased by increasing the assimilation frequency, the forecast error can also be decreased by providing a more accurate fire spread model.

While the previous results demonstrate that the FIREFLY manipulation of the wind field allows for a successful reconstruction of the observed fireline positions, there is an open question about the exact meaning of the inferred wind parameters and whether these parameters can be used to provide information on the near-flame wind dynamics. To answer this question, we present in Fig. 3.12 a comparison between the "true" values of the wind field produced by the reference ForeFire/Meso-NH simulation and the predicted values produced by the baseline data-driven FIREFLY simulation. The comparison is made through instantaneous snapshots of the spatial variations of horizontal wind velocity, at 2-m elevation and at time $t = 300$ s. In Fig. 3.12, the near-surface ForeFire/Meso-NH data, initially

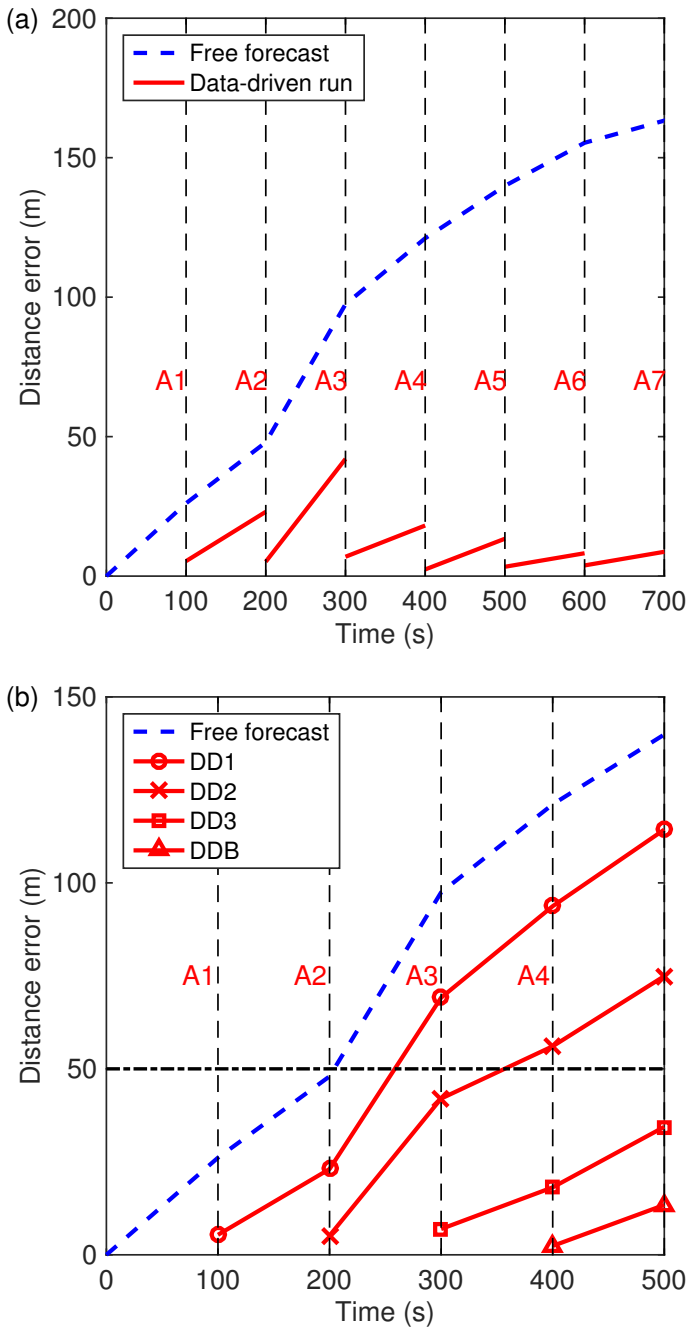

Figure 3.11: Time variations of the mean distance between the predicted and observed fire front positions for FireFlux. (a) Comparison between the free forecast (dashed line) and baseline data-driven run (solid line), $0 \leq t \leq 700$ s. (b) Comparison between the free forecast (dashed line) and different variations of the data-driven run, $0 \leq t \leq 500$ s: DD1 (circles); DD2 (crosses); DD3 (squares); DDB (triangles). The dash-dotted horizontal line at 50 m indicates a possible target value for the distance error.

estimated at 3-m elevation, are linearly interpolated to 2-m elevation; in addition, the FIREFLY data, initially estimated at 1-m horizontal resolution, are plotted on the 10-m resolution ForeFire/Meso-NH grid. The "true" and predicted wind fields are found to be quite different, especially near the fire front: the wind field in the ForeFire/Meso-NH simulation features turbulent-like small-scale variations (Fig. 3.12a) that are absent in the FIREFLY simulation (Fig. 3.12b). The presence of small-scale variations in Fig. 3.12a may be explained by the fire-atmosphere interactions that are captured in ForeFire/Meso-NH when configured in two-way coupling mode. FIREFLY cannot represent those small-scale variations, partly due to the coarse resolution of the wind grid and partly due to the limited modeling components (there is no description of the atmospheric boundary layer flow and of the fire plume in FIREFLY).

The discrepancies between ForeFire/Meso-NH and FIREFLY results are further characterized in Fig. 3.13 through scatter plots of the wind magnitude and direction: a perfect correlation between ForeFire/Meso-NH and FIREFLY results would correspond in these plots to a cluster of points near the 45-degree line; by the same token, large deviations from this line indicate a lack of correlation between the "true" and predicted values of the wind parameters. Note that the points near the 2 $\mathrm{m\,s^{-1}}$ vertical line in Fig. 3.13a and near the 10° vertical line in Fig. 3.13b correspond to data at inactive wind nodes (i.e. points located at a distance greater than the threshold distance from the fireline). The results in Fig. 3.13 suggest that the inferred wind parameters in FIREFLY are not accurate and should be interpreted as effective values that incorporate multiple sources of uncertainties that are

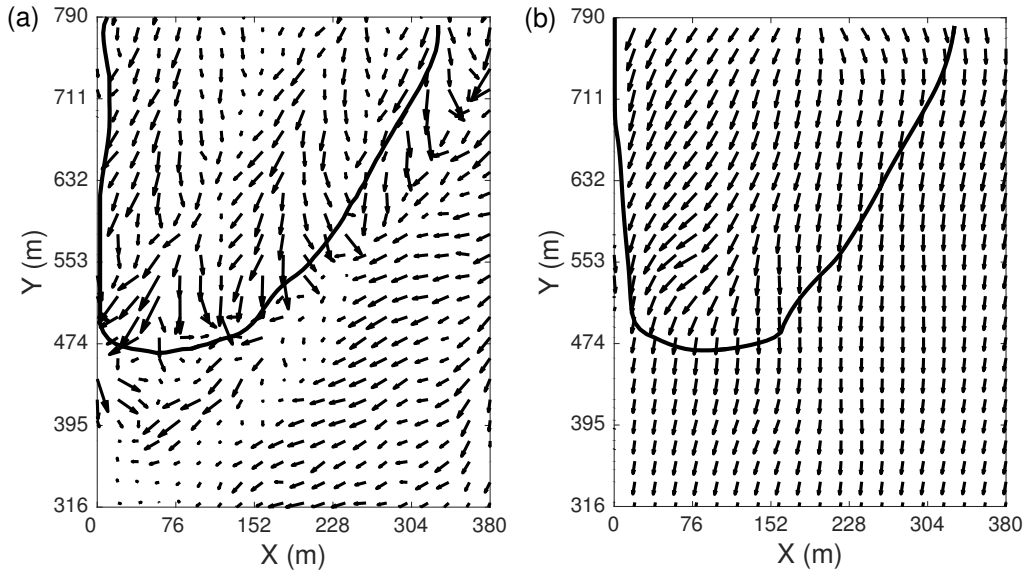

Figure 3.12: Instantaneous spatial variations of the horizontal wind velocity vector at 2-m elevation and at time $t = 300$ s. Comparison between (a)"observations (reference ForeFire/Meso-NH simulation) and (b) predictions produced by the baseline data-driven FIREFLY simulation (ensemble-averaged field). The solid line indicates the fire front location.

not identified here. More realistic values of the wind parameters may be obtained if the control vector $\mathbf{x}$ is extended to include additional parameters (e.g. vegetation parameters) and/or if the accuracy of the ROS model is increased (in particular to better describe the fire spread on the flanks).

Scalability Study

In this section, firstly we present scalability test results regarding to the ensemble size and the control vector size. The purpose of the scalability test is to estimate the computational cost when goes to large-scale wildland fire spread simulations,

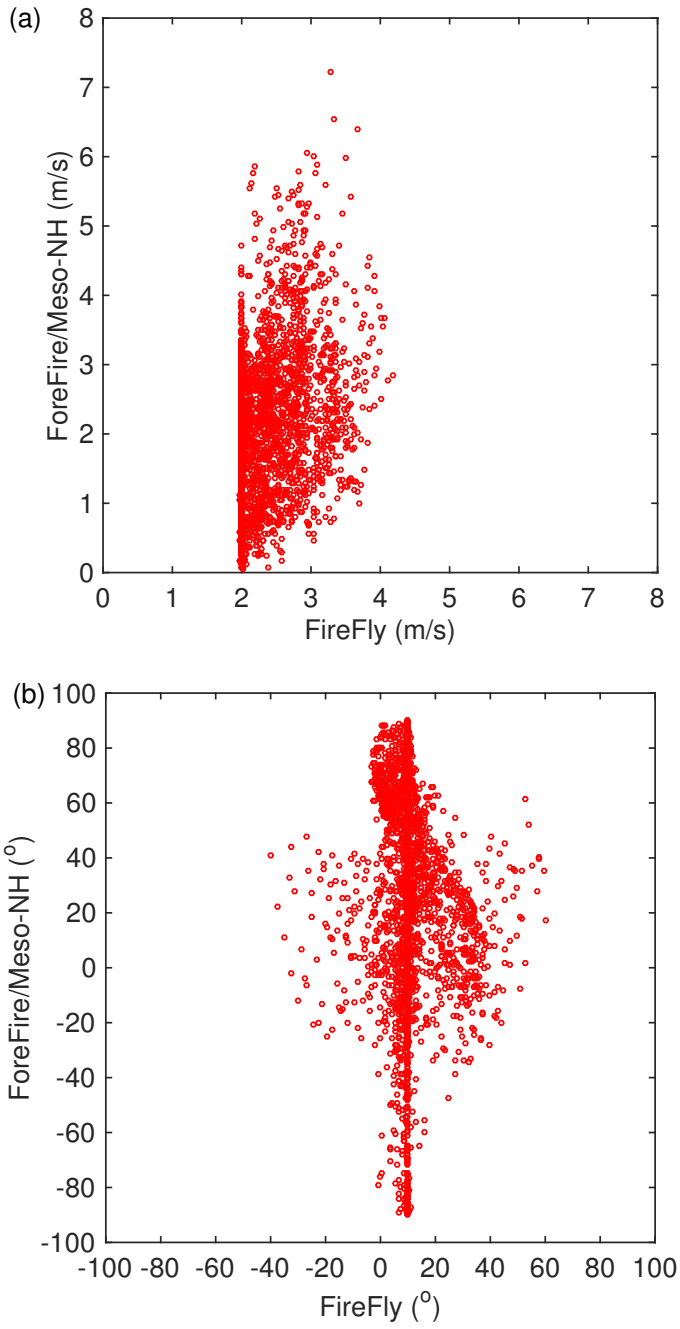

Figure 3.13: Scatter plots comparing FIREFLY (x-axis) and ForeFire/Meso-NH (y-axis) results for: (a) wind magnitude; (b) wind direction. The points correspond to the wind data presented in Fig. 3.12.

which is typically on the order of kilometers. For this test, first we run the simulation of first assimilation cycle with different ensemble size. Figure 3.14 presents the computational cost (evaluated as CPU time) and estimation accuracy (evaluated as mean distance error bewteen simulated and observed fronts) with different ensemble size ranging from 50 to 250. The CPU time is calculated as $T_{cpu} = T \times n$, where T is the simulation running time (hr), and n is the number of processors per node (20 in this case). It shows that the CPU hours will increase gradually as ensemble size increases. For this specific simulation case, the CPU hours generally increase 33 hours as 50 more ensemble members added each time. The distance error between simulation and observation decreases as ensemble size increases initially, but it will reach to a stable value when ensemble size is large enough. For current test, we could state that 100-150 members is the minimum requirement of the ensemble size (previous simulation results are obtained with 200 ensemble members).

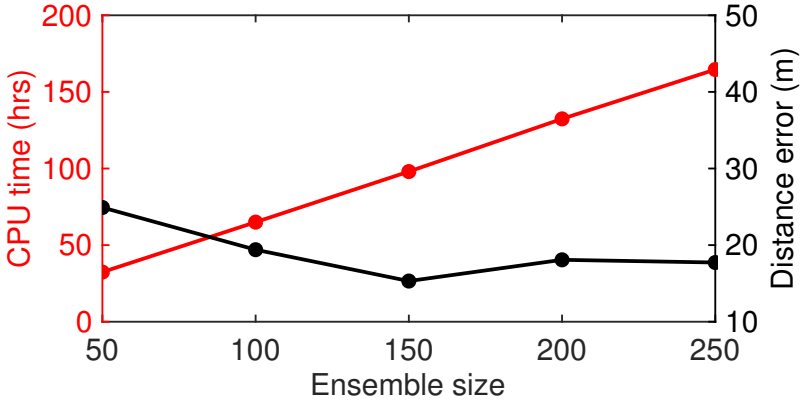

Figure 3.14: Variations of computational cost (CPU time) and estimation accuracy (mean distance error bewteen simulated and observed fronts) with different ensemble size.

As to the size of control vector, it is equal to the number of controlled wind grid nodes multiplied by two, since both wind speed and wind direction are included. Figure 3.15 presents the computational cost (CPU time) against different number of controlled wind grid nodes. In the present study, the CPU time does not increase as control vector size increases. Indeed, for parameter estimation purpose: the size of control parameter is relatively small; even if the control parameters go to several thousand for large-scale wildfire spread simulations, there is no need for special treatments on the computational cost. In contrast, when going to state estimation such as Numerical Weather Prediction (NWP) applications, the control vector (state variable) size is at least on the order of 10^7 [32]. Such large control vector makes both the storage and the computation of the matrices unfeasible, thus special techniques are needed for NWP applications.

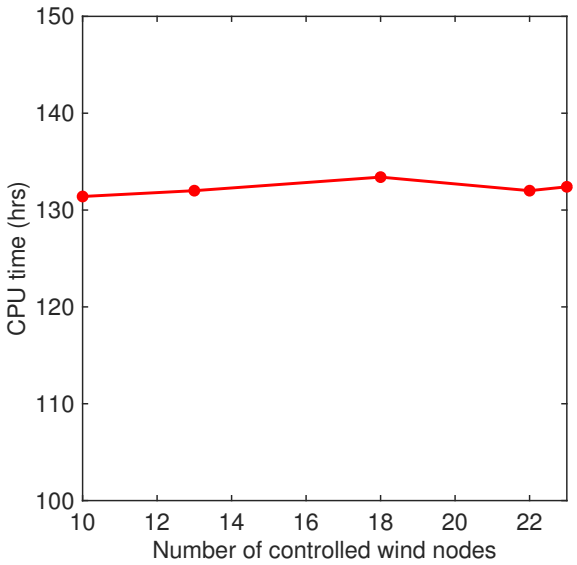

Figure 3.15: Variations of CPU time with different number of controlled wind grid nodes.

There is also of interest to check the computational cost regarding to the spatial and temporal resolution of the simulations. The spatial resolution is usually chosen in order to have a complete representation of fire front topologies, and we use 1 m spatial resolution for the FireFlux simulation case. After that the temporal resolution is set according to the Courant-Friedrichs-Lewy (CFL) condition to ensure the accuracy of the numerical scheme. In the present study, we test four different time step values which satisfying the CFL condition. Figure 3.16 presents the computational cost (evaluated as CPU time) and estimation accuracy (evaluated as mean distance error bewteen simulated and observed fronts) with different time step values ranging from 0.02 to 0.2. It is clear that the CPU time decreases proportionally as the time step value increases. As the distance error remains low throughout the test, we choose 0.02 s to simply ensure a convergent solution is met for results presented in this chapter.

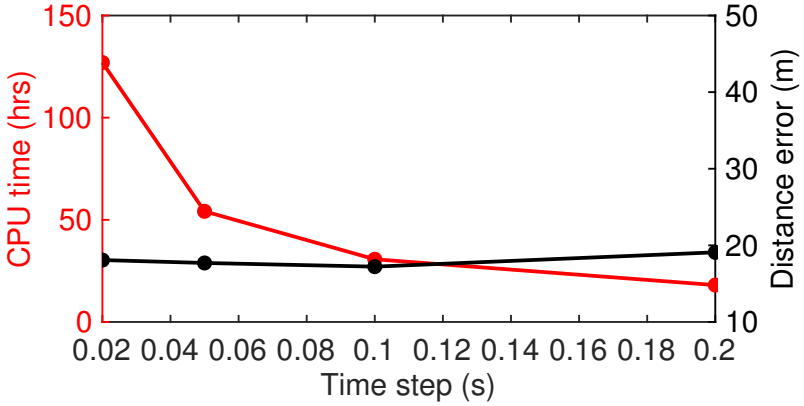

Figure 3.16: Variations of computational cost (CPU time) and estimation accuracy (mean distance error bewteen simulated and observed fronts) with different temporal resolution.

Conclusion

The results demonstrate the ability of FIREFLY to steer an inaccurate fire spread model towards observed fireline positions and to provide an improved forecast of the fire behavior compared to the standalone fire spread model. The quality of the forecast depends on the frequency of the observations. In the present study, the performance of the FIREFLY forecast is quantified for the first time: it is found that a 50-m accuracy on the fireline location requires a frequency of observation of 3 min or better. While this figure of merit is not general and is specific to the FireFlux experiment, it provides a valuable benchmark for future studies. The results also show that the inferred wind parameters should be viewed as effective values that incorporate multiple sources of uncertainties that are not identified here. While the objective of the parameter estimation method is not to provide a detailed representation of fire-wind interactions, these interactions play a role in the FIREFLY forecast capability and therefore should be captured with some level of accuracy.

Chapter 4: Shape Similarity Measure-based State Estimation

The inverse problem in a typical data-driven model relies on the measure of the discrepancies between the observations and simulated fire front locations to infer the magnitude of the corrections to the control vector. For wildfire spread application, it is thus of paramount importance to have an adequate representation of the observation-simulation discrepancies. In Chapter 3, observed fire fronts are treated as discretized contours with a finite set of markers when applied to the FireFlux I experimental dataset. The simulated fire front line is treated as discretized front markers in the same way. The Euclidean distance between these simulated and observed fronts is then computed by pairing each observed marker with its closest neighbor along the simulated front. The innovation involved in the EnKF algorithm corresponds to a difference in terms of x and y coordinates for the set of observed markers, which is consistent with Gaussian-type position errors (see Eq. 2.13). However, such pairing may become difficult to handle for regional scale wildfires induced by strong heterogeneity in the surface conditions (Fig. 2.8). In this case, it is difficult to correctly represent the Euclidean-type distance error by a marker-pairing method, thus this type of distance measure may become unsuitable when moving to realistic real-world wildfire events.

This chapter is dedicated to designing a front shape similarity measure to address position errors rather than amplitude errors in the fire problem, since the observed quantity is the fireline position. This Chapter is also dedicated to illustrating how this measure applies through a state estimation approach based on a Luenberger observer. We apply it to the three-hectare RxCADRE S5 field-scale experiment and results show that the front shape similarity measure is able to accurately track fire fronts and improve short-term fire spread forecast performance. We also demonstrate that this front shape similarity measure is suitable for both Eulerian and Lagrangian-type front-tracking solvers and thereby can provide a unified framework to track moving structures such as flame front position and topology in combustion problems. This work has been accepted for presentation at the International Symposium of Combustion and is currently under review for publication in Proceedings of the Combustion Institute [74].

4.1 Front Shape Similarity Measure

In Chapter 2 we have introduced how to compute the distance vector in the context of Kalman-type filter algorithm using the shape similarity measure. The similarity measure $\mathcal{J}$ between a target front and a simulated front reads as Eq. 2.14, it can be treated as a data-fitting term with two components:

$$\mathcal{J}_0(\mathrm{y}^o, \phi_\mathrm{c}) = \int_\Omega \left(1 - H_v(\phi_\mathrm{c})\right) \left[\mathrm{y}^o - C_{\min}(\mathrm{y}^o, \phi_\mathrm{c})\right]^2 \, dx\, dy \tag{4.1}$$

$$\mathcal{J}_1(\mathrm{y}^o, \phi_\mathrm{c}) = \int_\Omega H_v(\phi_\mathrm{c}) \, \left[\mathrm{y}^o - C_{\max}(\mathrm{y}^o, \phi_\mathrm{c})\right]^2 \, dx\, dy \tag{4.2}$$

where y^o represents the observation data (binarized field where $y^o = 1$ in the burnt area and $y^o = 0$ in the unburnt area), ϕ_c is the level-set associated with the progress variable c field.

This data-fitting term will reach minimum value if the simulated contour ($\phi_c = 0$) matches the observed contour. It is illustrated through the various examples in Fig. 4.1 [73]: different scenarios of the comparisons between simulation and observation, and it clearly shows that $\mathcal{J}$ is minimum when the observed and simulated contours coincide. In the following sections, we show that a state estimator is derived using the gradient of the shape similarity measure.

4.1.1 Luenberger State Observer

We now present Eulerian- and Lagrangian-type state estimators, which are derived from the data fitting functional presented in Eq. 2.14 and introduce new relaxation terms in the propagation equation used in the forward modeling component. This type of state estimator is known as a Luenberger observer. The gradient of the shape similarity measure in Eq. 2.14 reads

$$\nabla \mathcal{J} = \delta(\phi_c)\, \mathcal{D}(y^o, c) \tag{4.3}$$

where δ is the Dirac delta-function, and where $\mathcal{D}(y^o, c)$ is the discrepancy term defined as

$$\mathcal{D}(y^o, c) = [y^o - C_{\max}(y^o, \phi_c)]^2 - [y^o - C_{\min}(y^o, \phi_c)]^2 \tag{4.4}$$

The function δ plays the role of a localization operator (the correction associated with state estimators is only active along the simulated fire front) [76].

$$\mathcal{J}_{\text{data}} = \underbrace{\int_{\Omega} (1 - H_0(c)) \left[\mathrm{z}^{\text{obs}} - C_{\min}(\mathrm{z}^{\text{obs}}, c)\right]^2 \mathrm{dx}}_{\mathcal{J}_0}$$
$$+ \underbrace{\int_{\Omega} H_0(c) \left[\mathrm{z}^{\text{obs}} - C_{\max}(\mathrm{z}^{\text{obs}}, c)\right]^2 \mathrm{dx}}_{\mathcal{J}_1}$$

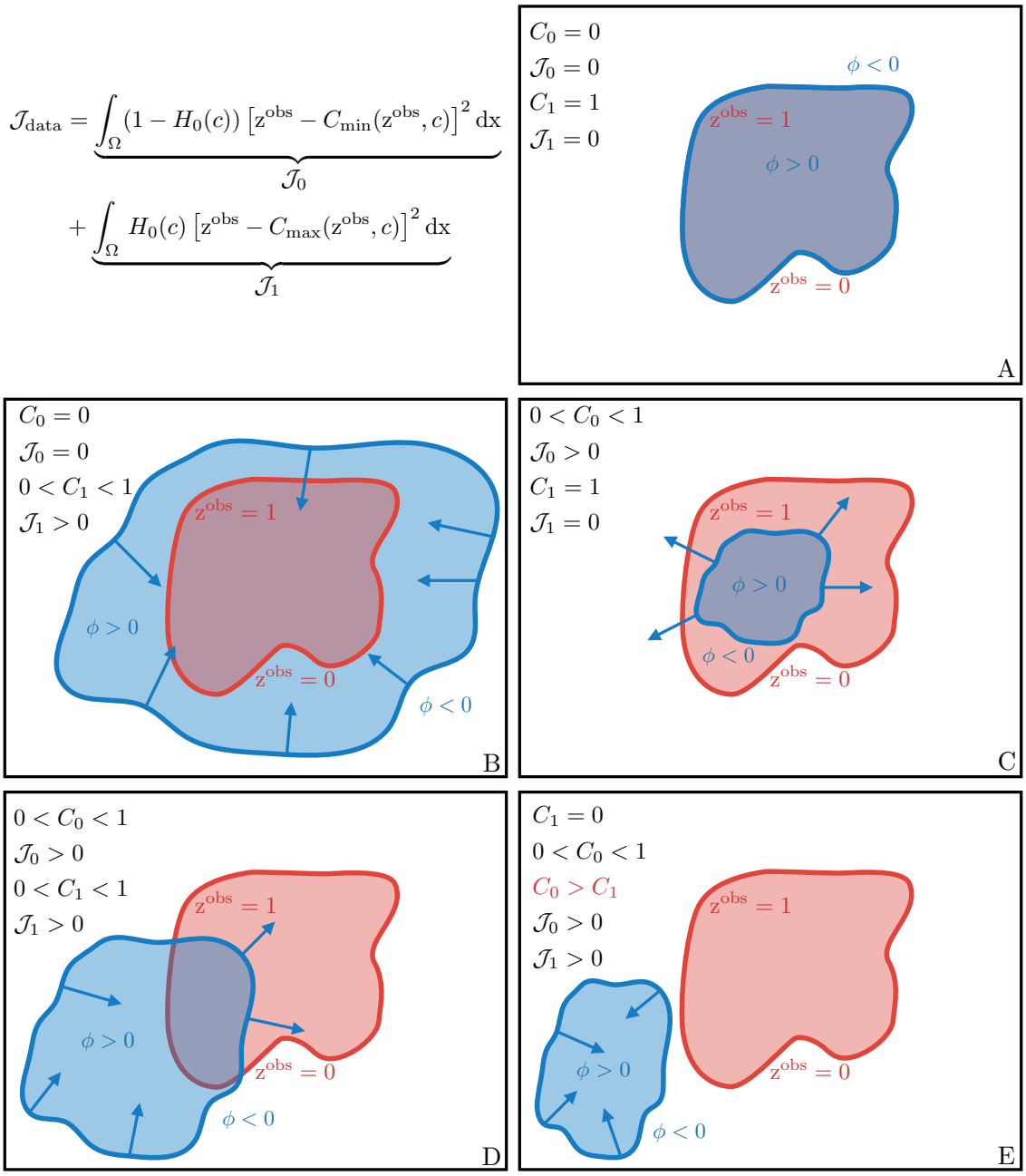

Figure 4.1: Illustration of Chan-Vese functional for possible configurations of simulated and observed fields. The arrows indicate the direction towards the minimization of the functional $\mathcal{J}$ [73].

For the Eulerian fire spread model, Eq. 2.5 is modified as follows:

$$\frac{\partial c}{\partial t} = \mathrm{ROS}_{2D}\ |\nabla c| - \lambda\delta(\phi_c)\left\{[\mathrm{y^o} - C_{\max}(\mathrm{y^o}, \phi_c)]^2 - [\mathrm{y^o} - C_{\min}(\mathrm{y^o}, \phi_c)]^2\right\} \tag{4.5}$$

where the feedback term is expressed as $-\lambda\nabla\mathcal{J}$, and where $\lambda \equiv \lambda(x, y)$ is the gain that describes confidence in the observations (λ can be intepreted as the inverse of the observation error variance). Starting from given (possibly incorrect) initial conditions c_0 at time t_0, λ controls the rate at which the simulated progress variable c converges towards the observations $\mathrm{y^o}$, it provides a balance between simulated model state and observations: a larger value of λ means more weight on the observation data. The LO-based state estimator allows a fast modification of the model state without requiring much additional computational efforts. This method is perfectly suitable for level-set forward modeling system since the fire front is represented by the surface contour ($\phi_\mathrm{c} = 0$). For the Lagrangian-type state estimator, since we only represent the fire frontline using a finite number of markers, it requires additional techniques to implement this state estimator which will be discussed later.

Note that there exists various numerical schemes to describe the Dirac delta-function [100], in the present study, the Dirac delta-function is numerically approximated as $\delta(\phi_c) = |\nabla c|$ since the initial fire front is thin. Equation 4.5 is referred to as the "Eulerian Luenberger observer" (LO–EUL).

The "Lagrangian Luenberger observer" (LO–LAG) can be derived by analogy to the LO–EUL. In Chapter 2 we presented the Lagrangian front-tracking solver implemented in the current data-driven modeling system, where we assume that the fire front features a local elliptical shape. The front could be parameterized by the

closed curve $(x(s,t), y(s,t))$ with $0 < s < 2\pi$. The Lagrangian equations are for $(x, y) \in \Omega$ and $t \geq t_0$

$$\begin{cases} \partial_t x = F\left(s, t, \partial_s x, \partial_s y\right) \\ \partial_t y = G\left(s, t, \partial_s x, \partial_s y\right) \end{cases}, \begin{cases} x(s, t_0) = x_0(s) \\ y(s, t_0) = y_0(s) \end{cases} \tag{4.6}$$

where $(x_0(s), y_0(s))$ is the initial condition at time t_0, and where F and G are functions of the local topography, biomass and meteorological conditions (the shape and orientation of the local ellipse are determined by the wind and slope conditions, while its size is determined by the ROS).

If $X(s,t) = (x(s,t), y(s,t))^T$ denotes the front parameterization and $V = (F, G)^T$ the associated velocity vector, then Eq. 4.6 is modified as follows:

$$\frac{\partial X}{\partial t} = V(s, t, X(s,t)) - \lambda \left\{ [\mathrm{y^o} - C_{\max}(\mathrm{y^o}, c)]^2 - [\mathrm{y^o} - C_{\min}(\mathrm{y^o}, c)]^2 \right\} \mathrm{n_{fr}} \tag{4.7}$$

where $\mathrm{n_{fr}}$ is the normal vector to the fire front and $\lambda \equiv \lambda(s)$. Equation 4.7 is referred to as the "Lagrangian Luenberger observer" (LO–LAG). Notice that the Dirac delta-function is no longer necessary since the Lagrangian formulation only acts on the fire front by definition. Also notice that we use a ray casting algorithm [101] to construct the binary burnt/unburnt fields from the front marker positions given by Eq. 4.7; these fields are required in the evaluation of the discrepancy term $\mathcal{D}(\mathrm{y^o}, c)$.

The feedback term features a similar formulation in the Lagrangian and Eulerian models. In principle, $\mathrm{ROS} = (V \cdot \mathrm{n_{fr}})$ provides the equivalence between the Eulerian and Lagrangian formulations of a front propagation problem. However, we adopt here the usual formulations of the wildland fire research field so that the way to handle fire propagation on the flanks and at the rear of the fire is different in the

present Eulerian and Lagrangian models [24,68] (see discussion in Chapter 2 related to Fig. 2.6).

The "Eulerian Luenberger observer" (LO–EUL) features additional benefits to be able to detect the spot fire. Inspired by previous work in [77], a topological gradient, can be computed as:

$$\nabla_{\text{top}}\,\mathcal{J}_{\text{data}} = \left\{[\mathrm{y^o} - C_{\max}(\mathrm{y^o},\phi)]^2 - [\mathrm{y^o} - C_{\min}(\mathrm{y^o},\phi)]^2\right\} \tag{4.8}$$

This term typically computes the sensitivity of the discrepancy measure to new appearing fronts [102]. From this sensitivity measure, we add an additional correction term to be able to detect new appearing fronts [77] to the Eq 4.5:

$$\begin{aligned}\frac{\partial c}{\partial t} &= \mathrm{ROS}_{2D}\;|\nabla c| - \lambda\delta(\phi_c)\left\{[\mathrm{y^o} - C_{\max}(\mathrm{y^o},\phi_c)]^2 - [\mathrm{y^o} - C_{\min}(\mathrm{y^o},\phi_c)]^2\right\}\\ -\lambda_{\text{top}}\,H_v(\nabla_{\text{top}}\mathcal{J}_{\text{data}} \times (c - c_{\text{fr}}))&\left\{[\mathrm{y^o} - C_{\max}(\mathrm{y^o},\phi_c)]^2 - [\mathrm{y^o} - C_{\min}(\mathrm{y^o},\phi_c)]^2\right\}\end{aligned} \tag{4.9}$$

where $\lambda \equiv \lambda(x, y)$ is the gain associated with the shape gradient and where $\lambda_{\text{top}} \equiv \lambda_{\text{top}}(x, y)$ is the gain associated with the topological gradient. The impact of this additional topological feedback can be illustrated through case E in Figure 4.1. Without this topological gradient, the simulated front (also the simulated burnt area) will decrease until it disappears. However, the topological gradient-based estimator can create a new front which corresponds to the observation data in this case. Unlike the shape gradient, the topological gradient correction term is only active once at the time when the new observation data becomes available. In contrast, the shape gradient correction term is applied at successive time steps during the whole assimilation window. In the following discussions, unless mentioned specifically, we focus on the shape gradient correction with λ referring to the gain associated with

the shape gradient.

4.1.2 Illustration of the Effect of Luenberger Observer Corrections

We illustrate the behavior of the front shape similarity measure by considering the Eulerian framework (LO–EUL) and a case for which the simulated fireline is enclosed by, and initially quite different from the observed fireline (see Fig. 4.2).

We consider a computational domain of 200 m $\times$ 200 m; the grid resolution is $\Delta x = \Delta y = 1$ m and the temporal resolution is $\Delta t = 0.1$ s. The gain is $\lambda = 4$. For illustration purposes, the propagation equation in Eq 4.5 is solved without the ROS term, implying that the λ parameter only features the speed of the correction process; a smaller λ value would provide the same correction but at a slower rate. As we discussed before, the coefficient C_1 corresponds to a perfect match for the simulated burnt area, therefore $C_1 = 1$; the coefficient C_0 corresponds to some level of mismatch for the simulated unburnt area, therefore $0 \leq C_0 \leq 1$; thus, $C_{\min} = C_0$, which gives $0 \leq C_{\min} \leq 1$, and $C_{\max} = C_1 = 1$. In the vicinity of the simulated fireline (at $c \approx c_{\text{fr}}$), the discrepancy term is negative, $\mathcal{D}(\mathrm{y}^{\mathrm{o}}, c_{\text{fr}}) = \left\{ [\mathrm{y}^{\mathrm{o}} - C_{\max}]^2 - [\mathrm{y}^{\mathrm{o}} - C_{\min}]^2 \right\} = -[1 - C_{\min}]^2 < 0$, which corresponds to a positive term on the right-hand-side of Eq 4.5 and as seen in Fig. 4.2, to outward propagation of the simulated fireline. Figure 4.2 (top) shows how the simulated fireline is progressively modified to match the observed front shape; Fig. 4.2 (bottom) presents the discrepancy term $\mathcal{D}(\mathrm{y}^{\mathrm{o}}, c)$ at time $t = 5$ s; negative (positive) values of $\mathcal{D}(\mathrm{y}^{\mathrm{o}}, c)$ correspond to outward (inward) propagation; this term is multiplied by

$\delta(\phi_c)$ so that the correction in Eq 4.5 is only active at $c \approx c_{\text{fr}}$.

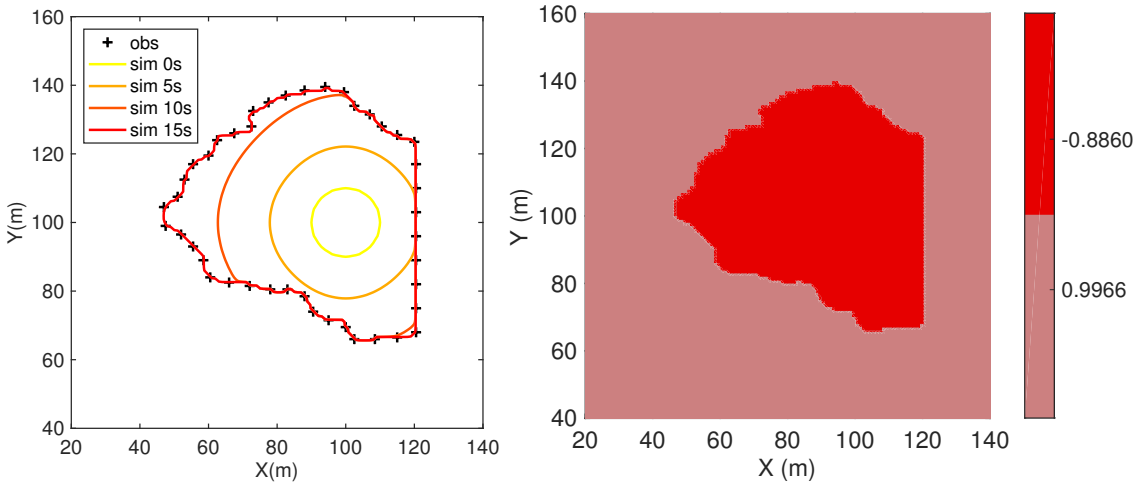

Figure 4.2: Simulation of a representative test case. (a) Comparison of simulated (lines) and observed (symbols) firelines at $t = 0, 5, 10$ and 15 s. (b) Discrepancy term $\mathcal{D}(\mathrm{y}^{\mathrm{o}}, c)$ (Eq 4.4) at $t = 5$ s.

4.2 Diagnostic of Front Distance

In Chapter 3, we evaluate the performance of data-driven model by calculating the average distance error between the same number of simulated front markers and observed markers. The formulation of the global shape comparison has the drawback of not providing a direct diagnostic in terms of Euclidean distance. To overcome this problem of interpretation, we need to introduce a diagnostic that has the dimension of a metric distance. In this chapter, we introduce a performance matrix based on the Hausdorff distance to determine the similarity of two objects (i.e., simulated fire front and observed fire front). One advantage of the Hausdorff distance is that it has practical meaning in physical space: its unit is meter (or kilometer), which

makes it consistent with the Euclidean distance concept.

The Hausdorff distance is widely used in computer vision to find a given template in an arbitrary target image [103–105]. In wildfire spread applications, this idea can be applied similarly to evaluate the match (or mismatch) between the observed and simulated fire fronts. It can be defined as follows: it is the greatest of all the distances from a point in one set representing the observed fire front to the closest point in the other set representing the simulated fire front. This distance has the meaning of maximum distance error between observation and simulation, which is a more conservative estimate than the mean distance error. To calculate the Hausdorff distance, first we discretize both the observed fireline Γ_o and the simulated fireline Γ_ϕ with a large enough number of points (O and P points respectively) to represent the fire front topology, then we apply the following algorithms to find the Hausdorff distance HD

$$\mathrm{hd}(\Gamma_o, \Gamma_\phi) = \{\max_{O\in\Gamma_o}\{\min_{P\in\Gamma_\phi} \mathrm{d}(O,P)\}\} \tag{4.10}$$

$$\mathrm{hd}(\Gamma_\phi, \Gamma_o) = \{\max_{P\in\Gamma_\phi}\{\min_{O\in\Gamma_o} \mathrm{d}(P,O)\}\} \tag{4.11}$$

$$\mathrm{HD}(\Gamma_o, \Gamma_\phi) = \max\{\mathrm{hd}(\Gamma_o, \Gamma_\phi), \mathrm{hd}(\Gamma_\phi, \Gamma_o)\} \tag{4.12}$$

For verification purposes, we first apply the Hausdorff distance calculation to the previous parameter estimation work against FireFlux experimental data, and compare it with the marker-pairing distance. Figure 4.3 plots the time variations of the Hausdorff distance for each assimilation cycle (left panel), showing the same trend of error growth as previously seen in Fig. 3.11. The quasi-linear correla-

tion between the Hausdorff distance and the marker-pairing distance (right panel) demonstrates the suitability of the Hausdorff distance. In addition, all the points on the right plot of Fig. 4.3 are above the 45-degree line meaning that the Hausdorff distance always yields a larger value than the marker-pairing distance, we consider it acceptable since a conservative estimate of the distance error (considering a worst case scenario and thus overestimating the distance error) is always better than underestimating it for fire risk management purposes.

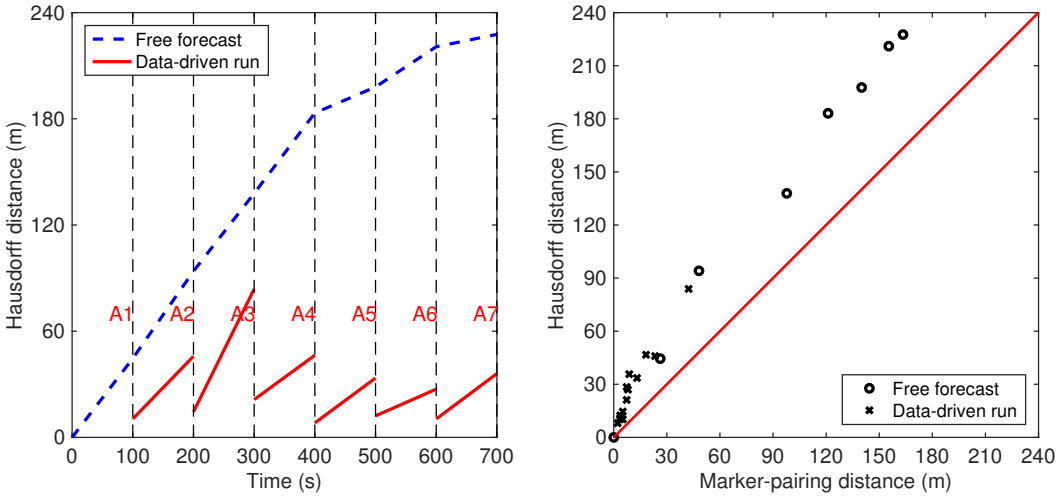

Figure 4.3: Hausdorff distance applied to the FireFlux experiment case. Left figure shows the time variations of the Hausdorff distance between the predicted and observed fire fronts. Right figure shows the comparison between the Hausdorff distance and the distance error adopted in Fig. 3.11.

4.3 Verification Test: Observing System Simulation Experiment

We first present in this section results from an OSSE case study in which synthetic observations are generated using the in-house Eulerian and Lagrangian

front-tracking solver. The purpose of these tests is to verify that the LO-based state estimation can nudge the simulation towards observation when imposing an appropriate gain value λ. Figure 4.4 and 4.5 show a test case in a computational domain of size 100 m $\times$ 100 m, and with a spatial resolution of 1 m. The total simulation time is 60 s and the time step is 0.02 s for the Eulerian model, 0.05 s for the Lagrangian model. In this case, we assume homogeneous fuel (tall grass in Rothermel's fuel database [22]) with a moisture content of 5% and uniformly-distributed wind (1 m/s, 225°) across a flat terrain. Observation data are generated with a known fire initial location (the fire is represented as a circle of radius 10-m, and whose center corresponds to the center of the computational domain). These observation data are provided every 20 s. We start our simulation with a shifted and deformed fire initial condition (the fire is represented as a circle of radius 6-m, and whose center is located at (60 m, 40 m)) and impose a medium nudging level ($\lambda = 1$) using Luenberger state observer through Eq. 4.5. During each assimilation cycle $[t, t+1]$ (whose duration is 20 s in the present case), the simulated model state is constantly modified at each time step Δt based on a linear combination of two observations at times t and $(t+1)$. As we can see in Figs. 4.4 and 4.5, at simulation time 20 s, the data-driven run (red dashed line) already matches part of the observation, and the small residual discrepancy is totally removed at time 40 s for the LO-LAG data-driven model, 60 s for the LO-EUL data-driven model. The difference between LO-EUL and LO-LAG is because the forward model yields different shapes of fire front, thus the discrepancy term $\mathcal{D}(\mathrm{y}^{\mathrm{o}}, c)$ is different, resulting in a different magnitude of correction at each time step. This test case demonstrates

the effectiveness of both LO-EUL and LO-LAG state estimation methods, they correctly drive the simulations towards observations given the user-specified value of the gain λ without adding much additional computational time.

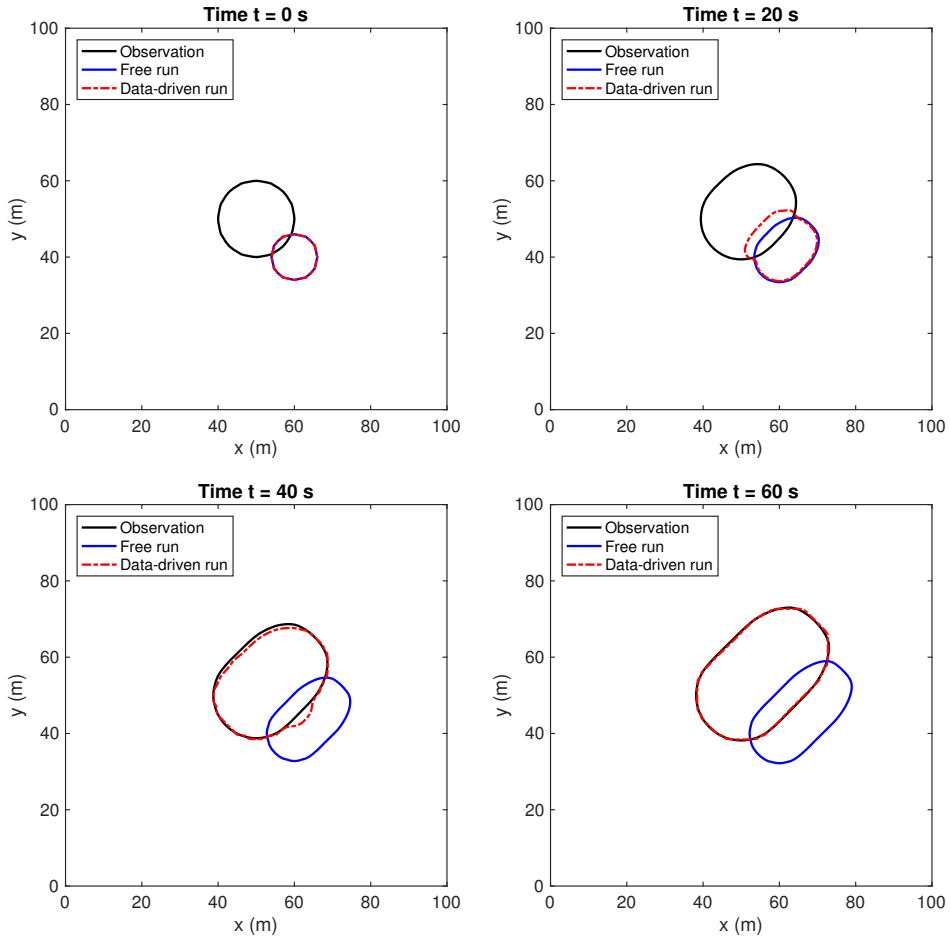

Figure 4.4: LO-EUL state estimation based on the shape gradient correction: black line represent the observed fire front, blue line represents the free run without data assimilation, red line represents the data-driven run with state estimation.

Another advantage of the LO-EUL state estimation method is that it is able to detect the appearance of multiple fires using the additional topological gradient correction. This will be particularly convenient to address spot fires. As we dis-

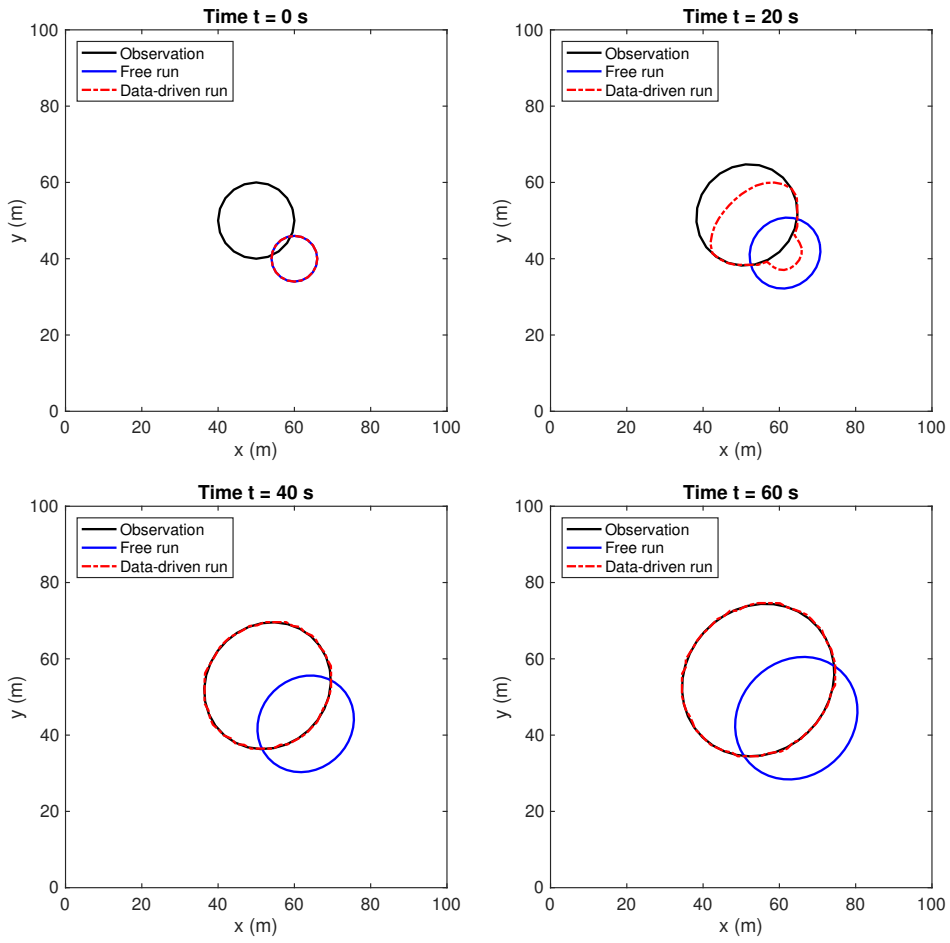

Figure 4.5: LO-LAG state estimation based on the shape gradient correction: black line represent the observed fire front, blue line represents the free run without data assimilation, red line represents the data-driven run with state estimation.

cussed before, a spot fire is a fire ignited outside the perimeter of the main fire by flying sparks or embers. It is very common in wildland fire propagation scenarios, and it remains one of the major challenges when using Langrangian front-tracking method. Figure 4.6 demonstrates that the topological gradient-based state estimation approach can detect a second fire contour present in the observed image but not included initially in the simulation. Here we use a numerical configuration similar

to that used in Fig. 4.4 except that a spot fire (modified configuration with a radius of roughly 5 m on the right side of the main fire location) is observed at time 20 s. As we can see, the data-driven simulation can consider the main fire at time 20 s, but is unable to detect the spot fire at that time. At time 40 s, the data-driven run is able to retrieve both the main fire and the spot fire with a topological gradient gain value $\lambda_{\text{top}} = 1$.

It is worth noting that so far we only consider a scalar λ. In the future we plan to implement a spatially-distributed λ, to better address incomplete observation data, the plan is to impose a larger value of λ in the areas where we have high fidelity observations and smaller values of λ in the areas where we have low fidelity observation or no observation, thus allowing a flexible spatially distributed state estimation approach [73].

4.4 Validation Test: Application to the RxCADRE Experiment

For validating the new front shape comparison method, we first apply the LO-based state estimation approach to a 4-hectare field-scale controlled burn from the RxCADRE S5 dataset. We will then show its potential on the Rim accidental wildfire.

4.4.1 Overview of the RxCADRE Experiment

We use the RxCADRE field-scale dataset from the Prescribed Fire Combustion and Atmospheric Dynamics Research Experiment (RxCADRE). RxCADRE yielded

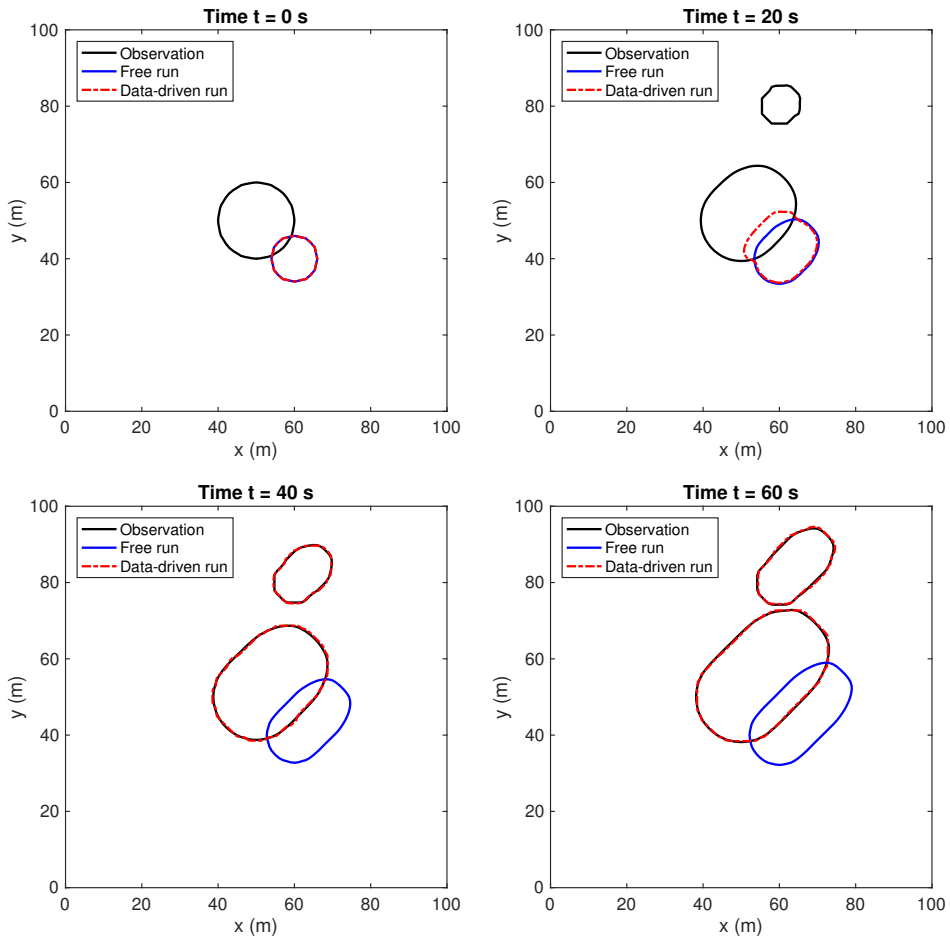

Figure 4.6: LO-EUL state estimation using the same Luenberger state observer but with additional topological gradient correction: black line represents the observed fire front, blue line represents the free run without data assimilation, red line represents the data-driven run with state estimation.

a comprehensive dataset of fire behavior, fire effects, and smoke chemistry and dynamics (see Table 1.3). These data were collected in 2008, 2011 and 2012, on small replicate and large operational prescribed fire burn blocks.

In the present work, we focus our attention on the 2012 S5 prescribed fire (4 ha), which took place at Eglin Air Force Base, Florida [106]. The S5 experiment

was a 15-minute-long prescribed fire conducted on a flat terrain characterized by a surface area of approximately 180 m $\times$ 180 m and a mixed grass and shrub vegetation. During the fire, the surface wind blew mainly from the North direction. In the simulations, we assume uniformly-distributed vegetation fuel as well as uniform and constant wind. The input parameters to the Rothermel model are based on experimental measurements: the fuel depth is 0.2 m; the fuel surface loading is 0.28 kg/m^2; the fuel particle surface-to-volume ratio is 9000 m^{-1}; the fuel moisture content is 10%; the wind velocity at mid-flame height is 2 m/s and the wind direction is 345° (corresponding to a northwest wind).

4.4.2 Observation Data

During the S5 experiment, the fire was ignited on the North side of the lot and propagated into the southern direction. Fire propagation was recorded through a series of temperature maps starting at time $t = 34$ s after ignition and recorded at 1-Hz frequency using a long-wave thermal infrared imaging system [106]. Since the initial fire only covers a very small area, our initial condition for fire spread simulations is used to mimic the observed fire spread at time $t = 60$ s. So in the following, time $t = 0$ s corresponds to time $t = 60$ s in the RxCADRE dataset. Figure 4.7 shows the moving fire front indicated by the temperature record by the IR imaging system. Based on these temperature measurements, we are able to generate a map showing the time at which the flame front arrives at a given pixel of the S5 burn lot (see Fig. 4.8), the left figure shows the arrival time map from 0 s to

480 s, which corresponds to the time interval [60 s; 540 s] in the RxCADRE dataset. In this chapter, we use the first 480 s fire data and assume that observations are available at 60-s time intervals. For example, Fig. 4.8 (right) shows the observed unburnt/burnt binary field that is assimilated at time $t = 480$ s. This binary field is obtained using the corresponding map of flame arrival times and after filtering to remove small-scale holes and outliers.

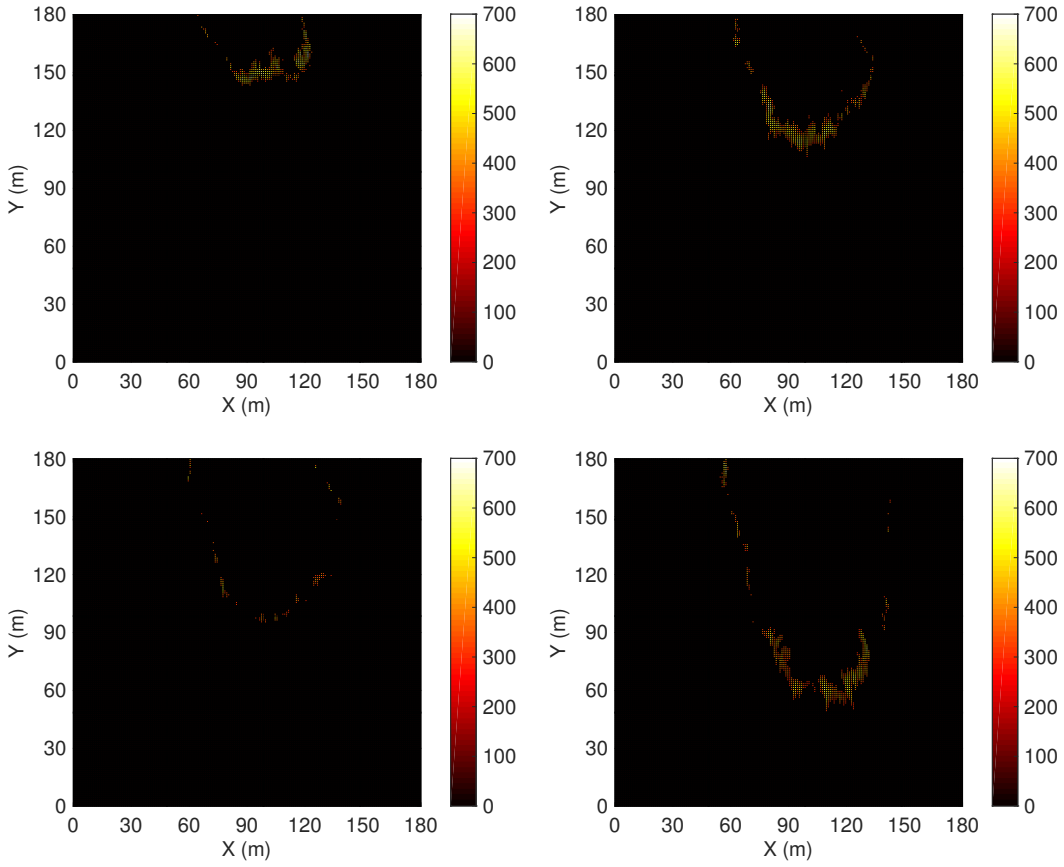

Figure 4.7: RxCADRE S5 fire temperature (unit: °C) distribution measurements at time 2 min (top left), 4 min (top right), 6 min (bottom left) and 8 min (bottom right).

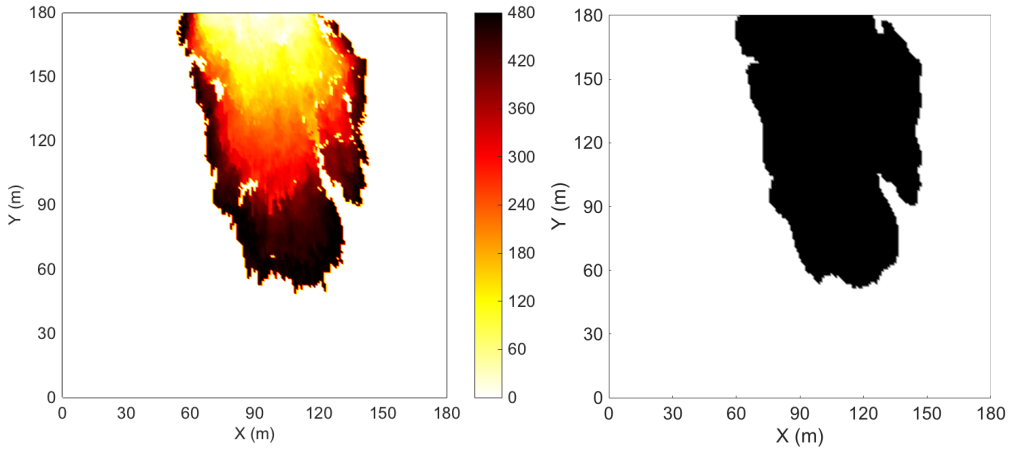

Figure 4.8: Dataset of the RxCADRE S5 fire. (a) Map of flame arrival times (0-480 s). (b) Binary image showing unburnt (white) and burnt (black) vegetation at time $t = 480$ s.

4.4.3 Numerical Settings

Model input parameters are selected using available experimental data. The computational domain is 180 m $\times$ 180 m; the total simulation time is 480 s with a fixed temporal resolution Δt. For LO–EUL, $\Delta x = \Delta y = 1$ m and $\Delta t = 0.05$ s. For LO–LAG, $\Delta t = 0.5$ s and the initial number of front markers is 40 (the number of front markers increases during the simulation due to the increasing length of the fireline). The initial condition c_0 (indicating the location of the fireline at ignition time) is supposed to be unknown. The objective is to demonstrate that our state estimation algorithm is able to overcome an imperfect knowledge of the fire situation at initial time. To initialize the fire spread model, we thus consider a simple, approximate fireline location, i.e. a semi-circular front with a 15-m radius and a center located at $(x, y) = (90$ m, 180 m). Observations are assimilated at 60-s time

intervals. Between two observations made at time $t = t_n$ and $t = t_{n+1}$ (i.e. during an analysis), the model state is continuously steered towards the observations made at time t_{n+1}. The intensity of the steering process is controlled by the gain λ (the higher the value of λ, the higher the level of confidence in the observations and the lower the value of the observation error standard deviation). Beyond the last observation (i.e. during a forecast), the relaxation terms are de-activated and the model solves the original propagation equation.

4.4.4 Performance of State Estimation

We first consider simulations with a large prescribed value of λ for both LO–EUL and LO–LAG models, in order to push the data-driven strategy to its limits in a situation where observations are considered accurate, the standalone model prediction (or "free run") is far away from the observations and thus where the correction is significant. In practice, the λ-value should be set according to the available information on the observations. Here we set $\lambda = 1$, which corresponds to a very large weight on the observation data in this case. Figure 4.9 compares the observed and simulated firelines during the first 8-minutes of the S5 experiment. The simulated firelines include free runs (that are not informed by the observations) and data-driven simulations using LO–EUL in Fig. 4.9 (left) and LO–LAG in Fig. 4.9 (right). It is seen that the free runs underestimate the ROS in both the head fire and flank regions indicating the presence of significant model errors for both Eulerian and Lagrangian models. In contrast, the data-driven simulations successfully reproduce

the location and shape of the fire perimeters. This shows the capability of the data-driven model to accurately retrieve the shape of the observed firelines.

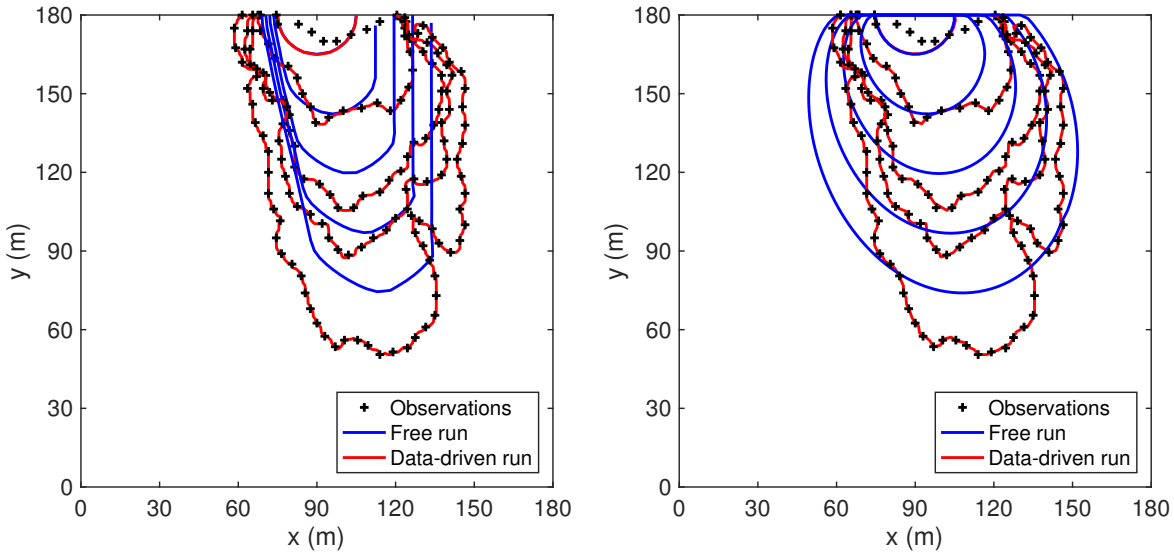

Figure 4.9: Comparison of simulated (lines) and observed (symbols) firelines at 120-s time intervals during $0 \leq t \leq 480$ s; $\lambda = 1$. Blue (red) lines correspond to free runs (data-driven runs) using the Eulerian model and the Lagrangian model, respectively.

The previous results are obtained with a prescribed constant value of the gain λ. As λ measures the level of confidence between the observed and predicted fronts, it should typically be proportional to the ratio of the observation error covariance over the prediction error covariance in terms of mean distance between fronts. A very illustrative justification of the impact of λ can be produced by treating λ as a stochastic parameter in an ensemble of test simulations. As we see from Fig. 4.9, $\lambda = 1$ is a very large value meaning high fidelity of the observations over simulation. In the following test, we generate an ensemble of simulations with different small-to-moderate λ values to consider a more realistic situation, in which uncertainties in

the observations are not negligible. Figure 4.10 illustrates the sensitivity of the data-driven solution to the value of λ for the LO–EUL model. In these test simulations, λ features a Gaussian statistical distribution characterized by a mean value $\mu = 0.2$ and a standard deviation $\sigma = 0.08$. We use a statistical ensemble with 40 members and Monte-Carlo-based random sampling. Figure 4.10 (left) presents the corresponding discrete probability density function. Figure 4.10 (right) presents the ensemble of simulated fireline positions at time $t = 60$ s. When the λ-value increases, the simulated fireline is closer to the observation, implying a higher confidence in the observation and thus a faster rate of convergence towards the observed fireline. This figure shows that we can use an ensemble of LO-state estimations to represent uncertainties. Similar results can be obtained with LO–LAG.

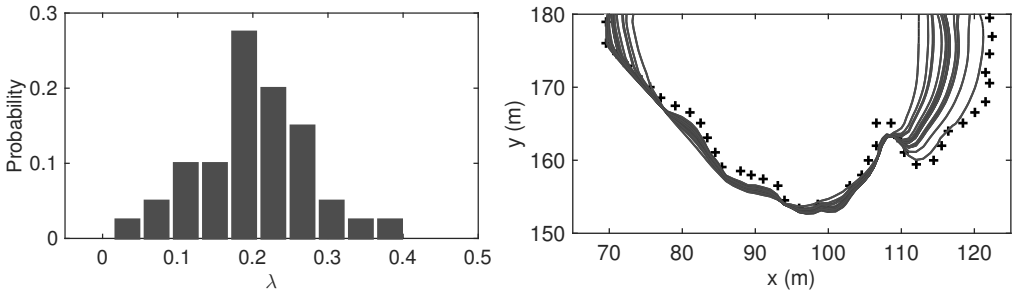

Figure 4.10: LO–EUL simulations using a statistical distribution for the gain λ. (a) Prescribed PDF of λ. (b) Comparison of simulated (lines) and observed (symbols) firelines at $t = 60$ s (the figure displays a subset of 15 simulations for clarity purpose; each line corresponds to a given value of λ).

We now examine the forecast performance at 60-s lead time of the LO–EUL and LO–LAG data-driven runs with $\lambda = 1$. Figure 4.11 shows the time variations of the distance error between the predicted and observed fireline positions. The

distance error is evaluated using the Hausdorff distance (in meter). Note that we choose the Hausdorff distance as a diagnostic tool since it is easier to interpret for an end-user than the front shape similarity measure (the latter is still used in the data assimilation algorithm to calculate the simulation-observation discrepancies, from which the new estimation is derived) and since it is a more conservative estimate than the mean distance error. Figure 4.11 compares the evolution of the Hausdorff distance in the free runs (blue lines) and data-driven runs (red lines). The evolution of the error in free forecast mode is a continuously increasing function of time. This error takes moderate values (approximately 25 m at time $t = 480$ s) primarily due to the limited scale and duration of the S5 experiment but the error is unbounded and keeps increasing in time. In contrast, the evolution of the error in data-driven mode is a discontinuous function: deviations of model predictions from observations are periodically reduced (to less than 5 m) during the analysis events (when integrating new observations). After each analysis event, the error increases but remains bounded and takes small-to-moderate values (on the order of 10 m). LO–EUL and LO–LAG models provide similar results. During a forecast, the LO feedback terms are de-activated and the model solves the original propagation equation; the rapid increase of the error seen in Fig. 4.11 after each analysis indicates that the ROS model has limited accuracy and that the benefits of assimilating new observations have limited persistence. Assimilation at 60-s frequency results in a distance error of 10 m; assimilation at lower (higher) frequencies would result in larger (lower) errors. Thus the forecast error can be decreased by providing higher assimilation frequencies or a more accurate fire spread model. In any case, Fig. 4.11 clearly

shows the ability of data assimilation to steer inaccurate fire spread models towards observed firelines and to provide an improved forecast of the fire behavior compared to a standalone fire spread model.

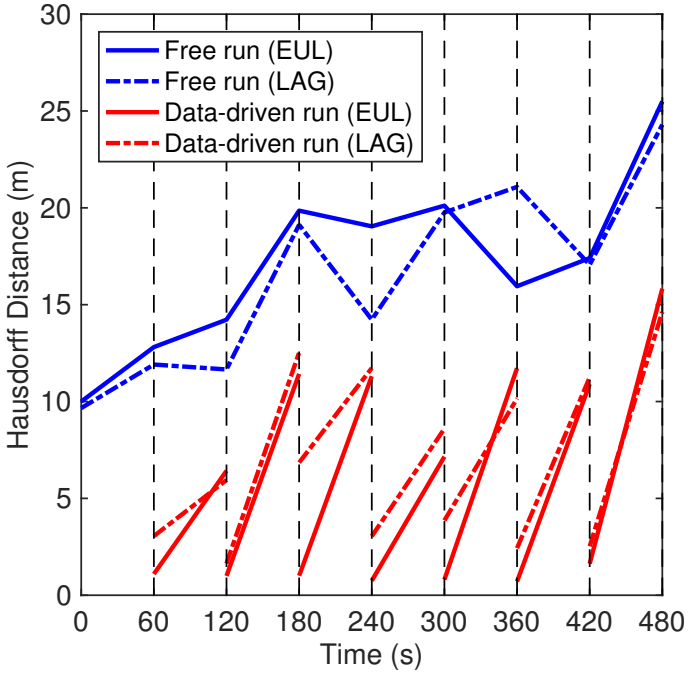

Figure 4.11: Time variations of the Hausdorff distance between the simulated and observed firelines. Blue (red) lines correspond to free runs (data-driven runs) using LO–EUL (solid line type) and LO–LAG (dashed line type).

4.5 Validation Test: Application to the Rim Fire

4.5.1 Overview of the Rim Fire

We continue our validation work and now consider a real accidental fire, the Rim Fire. The Rim Fire occurred during the 2013 California wildfire season on the edge of Yosemite National Park. It started on August 17 from an illegal campfire in

a remote canyon in central Sierra Nevada and extended up to 1040 km^2 (the third largest wildfire in California's history). It was fully contained on October 24 after a 9-week firefighting battle. The Rim Fire occurred under extreme drought and weather conditions, with notably unstable weather occurring soon after ignition, leading to two days of extreme fire growth in its early stage. By August 23, 400 km^2 were already burnt. So as a preliminary test, we focus here on the time period August 20-23 when the fire was moving at the fastest rates during its early development. It is worth noting that the Rim Fire occurred in forest vegetation types (chaparral, ponderosa and Jeffrey pines, white and red fir, etc), involving crown fire under severe environmental conditions which favor faster fire propagation and stronger fire intensities [107]. Thus a crown fire model is needed for more accurate calculations of fire propagation. This is beyond the scope of the present work. The current propagation model in our data-driven modeling system is used for surface fire spread, which include grasses, shrubs and other low-lying vegetation. This simplification of the ROS model is part of a modeling error addressed when estimating the model state. In other words, we should give more weights to observations over simulations due to the absence of a crown fire model.

4.5.2 Observation Data

For the Rim Fire, the Geospatial Multi-Agency Coordination (GeoMAC) data provide complete fire perimeter on a daily basis. GeoMAC provides fire perimeter data based upon multiple data source: incident intelligence sources (onsite crew,

apparatus measurements), GPS data, infrared (IR) imagery from fixed wing and satellite platforms. It is manually updated by interpreters located at a duty station everyday. We also have the National Infrared Operations (NIROPs) nighttime airborne data available for the Rim Fire. Like the GeoMAC dataset, the NIROPS dataset is postprocessed by an interpreter after raw data received. These raw data come from the Phoenix imaging system with a dual IR line scanner installed on an airplane.

In our preliminary test, we decided to use the GeoMAC dataset for Rim Fire over August 20-23 as our observation data. While their uncertainty are difficult to evaluate since the GeoMAC dataset combines multiple inputs which may have high or low fidelity individually, it offers complete fire perimeter information, which facilitates the construction of a binary field of burnt area, suitable for data-driven simulations using a LO-based state estimation. Figure 4.12 shows the 6 observation snapshots available over August 20-23 from the GeoMAC data source. The fire was ignited in the southwestern area, and propagated mainly towards the northeast and northwest directions over this time period.

4.5.3 Numerical Settings

The vegetation fuel map was downloaded from Google Maps and we adopted three spatially-distributed vegetation fuel types (forest fuel, bare soil and lake water) based on the fuel map. The terrain topography properties were set up using slope and aspect values at all cells extracted from Google image by QGIS tool, with a

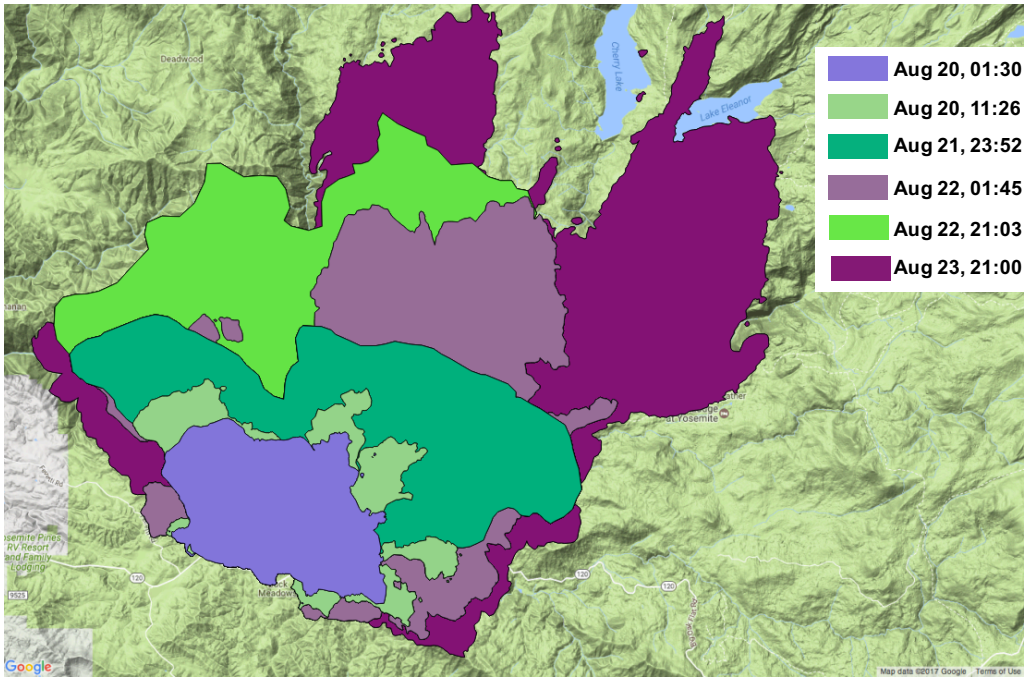

Figure 4.12: 6 observations available from GeoMAC data source for Rim Fire over August 20-23. *Credit: Dr. Evan Ellicott, UMD.*

cell size 70 m × 70 m. Although several weather stations nearby recorded the meteorological data during the Rim Fire event, we assume that the wind is uniformly distributed (average middle flame wind speed is 0.8 m/s, wind direction is 225°) in our preliminary test. This simplification will be removed in future work by coupling with WindNinja tool to have a specially-distributed wind field input.

In the present study, we tested the Rim fire data with the LO–EUL framework, LO–LAG work can be carried out in a similar fashion. The computational domain size is 69 km × 49 km (33800 ha), the spatial resolution is 70 m. The total simulation time is 100 hours, and the temporal resolution is 1 min. We consider a simplified representation of the fireline and we initialize the fire as a circular front (radius

3600 m) at location (27 km, 15 km). The purpose of this ignition configuration is to have a good approximation of the first observation. We will test whether our state estimation method is able to track the observed fire propagation if start with biased information. This is a good representation of a real world case where usually the initial fire location cannot be accessed.

4.5.4 Performance of State Estimation

Figure 4.13 shows the comparison of firelines between a free forecast (without state estimation) and a data-driven run (with LO–EUL state estimation). Six available observations are displayed in black cross symbols, the blue solid lines represent the free forecast at corresponding observation times, and red solid lines represent the data-driven run. Given the significant differences between the free forecast and observation, we use a relatively large value of λ ($\lambda = 0.5$) that allows the posterior fire front to accurately track the observations in this case. This test case demonstrates the ability of the new front shape comparison method to retrieve the correct front shape, even when this shape is complex due to highly heterogeneous environmental conditions. One advantage of this approach is that it allows a fast modification of the model state without requiring much additional computational resources. The CPU running time is 30 min on a personal laptop for the simulation of the Rim Fire case with a 100 hours simulation time. This computational cost is consistent with an operational framework.

We tested the forecast performance using the same LO–EUL state estimation

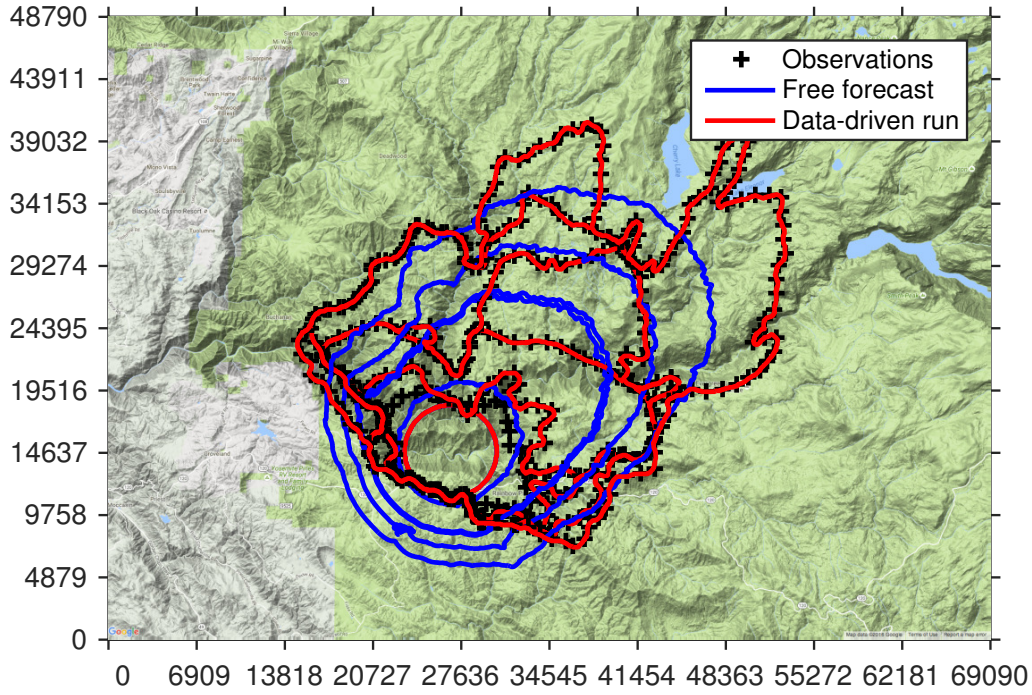

Figure 4.13: Comparison of fire fronts using LO-EUL state estimation for the Rim Fire: blue line represents prior fire front without data assimilation, red line represents posterior fire front with data assimilation, black crosses correspond to the 6 observed fire fronts during August 20-23.

approach by examining the Hausdorff distance between simulations and observations at different times. Figure 4.14 shows the time variations of the Hausdorff distance. The blue line indicates how the prior Hausdorff distance increases during the free run without data assimilation. The discrepancy between simulation and observation keeps increasing during the simulation time. In contrast, when applying data assimilation, the posterior Hausdorff distance was significantly decreased to the lowest value at all observation times. This is quite understandable since we impose a relatively large gain value ($\lambda = 0.5$) to nudge the forward model simulation towards observation, so that the simulated fire front is similar to the observed fire front over

the whole fire duration. Based on these results, we examine the Hausdorff distance at the forecast time; corresponding here to the next observation time. We found that the Hausdorff distance computed for the data-driven run at the next observation time is of the same order of magnitude or even larger than that corresponding to the free run. This is due to the fact that the state estimation approach usually has a limited time persistence and is only valuable at short forecast leading times. Here, the observation time interval is very large (> 10 hours). In addition, it is hard to keep track of accurate fire propagation by doing standalone state estimation, especially when there is a sudden change in the fire front topology due to changes in the environmental conditions and/or firefighting actions. The input parameters need to be modified to have a more consistent forecast, for instance we could act on the vegetation moisture content to account for firefighting actions. In the future, a better forecast capability is foreseen by applying also a parameter estimation approach, as was shown in Chapter 3 for the FireFlux case. In addition, a better representation of the spatially-distributed fuel and wind parameters will certainly help to construct a more realistic model simulation on a large scale wildfire event like the Rim Fire.

Conclusion

The results demonstrate the performance and robustness of the shape similarity measure applied in the context of LO-based state estimation approach. In the Lagrangian distance formulation implemented prior to this work, the observed fire front is treated as a discretized contour with a finite set of markers, and calculates

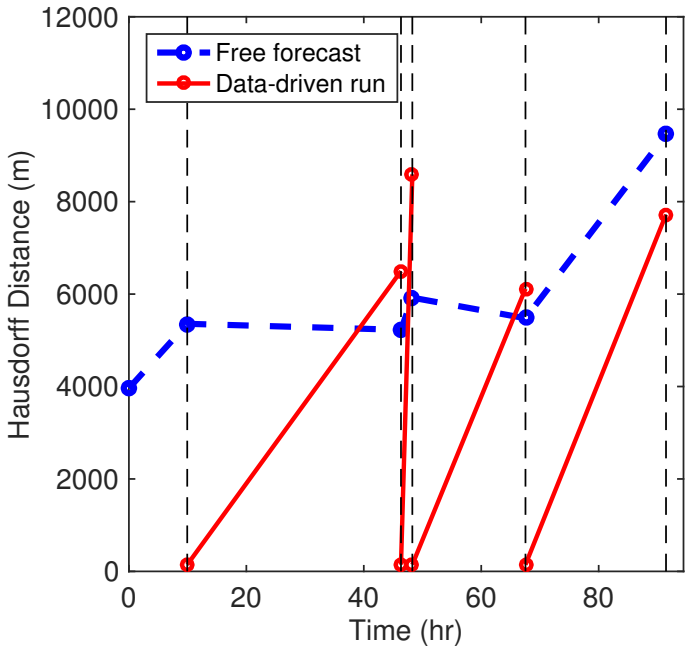

Figure 4.14: Time variations of the Hausdorff distance between simulated and observed fire front postions for the Rim Fire. The blue dashed line represents the free forecast and the red solid line represents the data-driven run.

the distance error by pairing the observed markers with the simulated markers. This method is not accurate for large scale wildfires that feature strong heterogeneity in the surface conditions. With the new shape similarity measure derived from image segmentation, we directly consider the burning area as a moving object that can deform under heterogeneous conditions and thus represents the similarity (or dissimilarity) in terms of topologies between the simulated and observed burning areas.

The performance of the new front shape similarity measure combined with a state estimation approach is evaluated in the three-hectare RxCADRE S5 controlled burn, in both analysis (i.e. assimilation) and forecast modes, and using an Eulerian

or Lagrangian fire spread model. The results demonstrate the ability of the proposed models to reduce uncertainties and to provide an improved forecast. Our preliminary test with the Rim fire data also shows the viability and robustness of this front shape similarity based on a state estimation approach to deal with km-scale wildfire hazards.

It is worth noting that including the forward model in the estimation is particularly useful when we only have partial observations of the fireline [42]. Although it is not the case here, observed firelines are often incomplete in wildfire hazards due to cloud cover or scale-inefficient detection techniques. Simple extrapolation of the observation data will not work well in such scenarios, especially in the presence of surface heterogeneity (e.g. lakes, roads) that induces changes in fire dynamics. The state estimation approach uses the forward model to provide physical constraints and define correlations of the errors along the fireline in order to reconstruct a complete fireline during the analysis step. Note also that observation data may be subject to uncertainties and that data assimilation also provides an efficient framework to account for observation uncertainties, which would be a useful capability to monitor wildfire hazards where observation uncertainties are large.

In this chapter we focus on state estimation and the proposed models are shown to provide much improved short-term forecast performance. However, longer-term forecast performance requires a correction of the fire spread model through parameter estimation. Therefore in Chapter 5 we will show the development of a dual state-parameter estimation approach where we distribute the corrections to both model parameters and model state appropriately, in order to retrieve more

physical values for the control parameters and gain a long-term forecast leading time.

Chapter 5: Dual State-parameter Estimation

In Chapter 3, our data-driven wildfire spread modeling system was successfully evaluated against reduced-scale and field-scale (FireFlux I) experiment in spatially-distributed parameter estimation mode. In that work, we use an EnKF algorithm to find better estimates of the Rothermel-based model parameters (i.e., spatially distributed wind parameters), thereby improve forecast performance. However, we did not find a good agreement between posterior wind values and analytical wind values from which the observation data were generated. Thus we treated the posterior wind values as effective values which incorporate multiple sources of uncertainties. In addition, the innovation term is calculated with discretized markers representation, which suffers limitation when moving to regional-scale wildfire spread simulations.

In Chapter 4, we introduced a LO-based state estimation method to directly update the fireline position and shape based on the front shape similarity measure derived from image segmentation theory. This method allows a fast modification of the model simulation towards the observed fire front location. But such state estimation approach yields a good forecast only for a short leading time, since the ROS parameters may be baised and uncertain.

Thus the objective in this chapter is to show the benefits of a dual state-

parameter estimation approach in terms of analysis and forecast performance. The key idea of this approach is to simultaneously control the model state and a subset of significant parameters, in order to obtain more accurate physical values of the estimated parameters by attributing uncertainties not only to the model parameters, but also to the model state. The correction in both the model parameters and the model state is expected to improve forecast performance due to the improved knowledge in the rate-of-spread model. Parameter estimation is carried out using an ETKF algorithm [92], which is a deterministic ensemble-based data assimilation algorithm. State estimation is carried out using a Luenberger observer suitable for front-tracking problems [73]. Both the state estimation and parameter estimation approaches rely on a front shape similarity measure to represent the differences in position between the observed and simulated firelines. The performance of the dual state-parameter estimation approach is demonstrated for verification tests against synthetic observations as well as for a validation test against the 2012 RxCADRE S5 field-scale experiment [106]. This work is currently being written and will be soon submitted for publication in an archival journal.

5.1 Dual Algorithm

In wildland fire applications, the discrepancy between simulated and observed fire fronts could result from either the biased initial fire location, or unknown model parameters. In Chapter 2 and 3, we use $\mathbf{x}$ to denote the control vector, which will be corrected based on the discrepancy between simulation and observation. The

control vector could include either model state or model parameters depending on the application system, whether it is executed in state estimation mode or parameter estimation mode. For the dual estimation method, for clarity purpose, we will use θ to specifically denote the control variables (ROS parameters) for the parameter estimation part, c to denote the model state (two-dimensional progress variable field) for the state estimation part, x to denote the "extended state" (both model parameters θ and model state c included). In the present study, the uncertain input variables are the initial location of the fire (i.e. the initial condition c_0) and some ROS parameters θ entering the ROS model (i.e. the parameters underlying the ROS model, in particular those that are associated with the near-surface wind U_w [41]). Note that the model formulation itself (in particular the ROS model) is subject to uncertainties; this can also be addressed by data assimilation algorithm [108], however this is beyond the scope of the present study.

In the following, the front position uncertainties due to uncertainties in θ are addressed through parameter estimation, meaning that the parameters included in the vector θ are explicitly updated when observations become available. We assume that the errors on the ROS model parameters vary slowly in time. Moreover, the front position uncertainties due to the initial fire c_0 are addressed through state estimation, meaning that the state $c(x, y, t)$ is updated when observations become available to locally correct the shape of the fireline and restart the forward model from a better-informed fire situation. This is consistent with the objective of forecasting the wildland fire behavior at future leading-times. This is in line with what is usually done for numerical weather predictions. The objective of sequential data

assimilation in this framework is to find the time-evolving dynamics of the "extended state" x (also referred to as the "control vector"). The state and the parameters are estimated using different data assimilation algorithms in this work for cost issues. For the state estimation part, a cost-effective LO-based state estimator is implemented during model integration stage (i.e., forecast stage), it continuously nudge every simulation member towards the observation at a certain extent (see Chapter 4). For the parameter estimation part, we still rely on an ensemble framework since the dimension of model parameter θ is usually much lower than the dimension of model state c. Each ensemble member of the control parameters is updated at the analysis time when the observation data is compared with the "nudged" simulation model state (see Chapter 3). The idea of this dual state-parameter estimation is that: the discrepancy between the observation and simulation is partially distributed to state correction during the model integration stage, and the remaining discrepancy is distributed to the ROS parameter correction during the parameter update stage at analysis time.

The dual state-parameter estimation framework is similar to the framework of the parameter estimation using EnKF (see Chapter 2 and 3). During the assimilation window $[t-1, t]$, we start the simulation with a finite number of ensemble member N_e. Each member starts with the same initial model state but with different control parameter values generated from a Gaussian PDF at time $(t-1)$. The main difference is that during the model integration $[t-1, t]$, we use the propagation equation 4.5 instead of the original equation 2.5 for each ensemble member. With an appropriate choice of the gain value λ (typically based on the inverse of the observa-

tion error covariance), the discrepancy between simulation and observation has been decreased due to the "nudging" process introduced in the forecast stage; thus at the analysis time t, the remaining discrepancy is used to update every model parameters in the ensemble. Note that we use an ETKF to update control parameters, but the corresponding model simulations at time t are already "contaminated" by the observation due to the nudging process, while the Kalman-type filter assumes that model errors and observation errors should be uncorrelated. Also the distribution of the corrections to either model state or model parameters is controlled by the gain factor λ, while it is a practical and direct method, the optimality of the solution from a mathematical proof perspective is outside the scope of the present study.

5.1.1 First Stage State Estimation

In the present study, we correct the model state using a deterministic Luenberger observer, i.e. by directly modifying the propagation Eq 2.5 through the introduction of relaxation terms towards the observations y^o. In the present front-tracking problem, the front shape similarity measure-based LO implicitly propagates the correction from one point of the fireline to the rest of the fireline and thus preserves a coherent structure for the simulated front. Note that this is different from the state estimation algorithm used in [42], where an EnKF was implemented and tested against reduced-scale controlled burnings. Using a LO avoids the formulation of high-dimensional matrices required to compute the Kalman gain matrix when the model state is the two-dimensional progress variable c.

Assume N_e ensemble members are generated at initial time $(t-1)$, the k-th member carries an "extent state" $\mathbf{x}^{\mathrm{k}}$ including model state c^{k} and control parameters θ^{k} $(k = 1,...,N_e)$. During the forecast stage $([t-1, t])$, the model state of the k-th ensemble member is forwarded in time:

$$\frac{\partial c^{\mathrm{k}}}{\partial t} = \mathrm{ROS}_{2D} \left|\nabla c^{\mathrm{k}}\right| - \lambda\delta(\phi_{c^{\mathrm{k}}}) \left\{[\mathrm{y}^{\mathrm{o}} - C_{\max}(\mathrm{y}^{\mathrm{o}}, \phi_{c^{\mathrm{k}}})]^2 - [\mathrm{y}^{\mathrm{o}} - C_{\min}(\mathrm{y}^{\mathrm{o}}, \phi_{c^{\mathrm{k}}})]^2\right\} \quad (5.1)$$

Since the model parameters θ^{k} varies for each ensemble member, the simulated fire front shape will vary too. Thus in Eq 5.1 the nudging effect by the additional correction term is different for every ensemble member. At analysis time t, the model state c^{k} varies among N_e ensemble members, although all of them are nudged towards the observation somehow. In principle, the gain value λ should be chosen wisely to eliminate the discrepancy between simulation and observation that is not caused by the biased control parameters. It cannot be a very large value in dual estimation approach, otherwise at the analysis time t, the model state are completely nudged to match the observation, and the distance vector $\mathcal{D}$ will vanish, and will not provide an update on the control parameter space.

5.1.2 Second Stage Parameter Estimation

At analysis time t, we update the N_e ensemble of the control parameters (i.e., the ROS submodel parameters) using ETKF algorithm. The general update equation is:

$$\theta_{\mathrm{t}}^{\mathrm{a}} = \theta_{\mathrm{t}}^{\mathrm{f}} + \mathbf{K}_{\mathrm{t}}\mathcal{D}\left[\mathbf{y}_{\mathrm{t}}^{\mathrm{o}}, \mathcal{G}\left(\theta_{\mathrm{t}}^{\mathrm{f}}\right)\right] \quad (5.2)$$

where the distance vector $\mathcal{D}$ is computed based on the shape similarity measure (see Eq 2.21). $\mathbf{K}_t$ stands for the generalized Kalman gain. In the ETKF algorithm, the ensemble mean is updated using the traditional Kalman gain, and the ensemble spread is updated using the a matrix square root form of that gain. The details of these computation can be found in Chapter 2 .

After we update the control parameters θ_t^a we re-run the forward model from time $(t-1)$ to time t using the updated parameter values for all ensemble members, so that we can also get the updated "extended state" vector $\mathbf{x}_t^a$. The new extended state vector $\mathbf{x}_t^a$ serves as the prior-known information for next assimilation window. Note that for the validation test against the RxCADRE S5 experimental data, we use the RTPP scheme presented in Chapter 2 to do covariance inflation work when moving from one cycle to another. We use a α value of 0.5 featuring a medium inflation process.

5.2 Verification Test: Observing System Simulation Experiment

In this framework, the true value of the uncertain variables (the input parameters θ and/or the initial condition c_0) exists and is known. In the present study, we choose the mid-flame wind parameters in the ROS submodel as our control parameters and we assume they are uniformly-distributed, so the control vector θ includes both the wind speed u_w and the wind direction d_w that are uniform over the computational domain Ω. Note also that the initial condition c_0 is formulated here in a parametric form with respect to the "center of mass" of the initial burning

area, whose position is denoted by $(x_{\text{ign}}, y_{\text{ign}})$. So the parameters that can be perturbed in the following to generate observations or prior information are $(u_{\text{w}}, d_{\text{w}})$ and $(x_{\text{ign}}, y_{\text{ign}})$.

The observations y^{o} are synthetically generated by integrating the fire spread model (Eq. 2.5) using the true values of the uncertain variables and by adding noise to the observed fireline. In the present OSSE test case, the error in the observations ϵ^{o} is assumed to be very small; the observation is thus considered as the "target".

The performance of the dual state-parameter estimation is evaluated by examining its ability to retrieve the true value of the control parameters θ and the true structure of the fire state c in a situation where observations are considered accurate. This configuration aims at pushing the data-driven model to its limit in a situation where the error in the prior is large and where the estimation involves an important correction to match the observations.

Numerical Settings

The 2-D computational domain Ω is 180 m $\times$ 180 m (with a step size $\Delta x =$ 1 m); the time window is 100 s (with a constant time step $\Delta t = 0.1$ s) with the analysis carried out at time 50 s. In this case, we assume flat terrain and heterogeneous biomass moisture content, $M_{\text{v}} = 5$ % (yellow areas in Fig. 5.1) and $M_{\text{v}} = 20$ % (blue areas in Fig. 5.1). The rest of the input parameters to the ROS model is assumed uniform. The biomass fuel is characterized by $\delta_{\text{v}} = 0.2$ m (layer thickness), $\rho_{\text{v}} =$ 1 kg/m^3 (layer mass density), $\Sigma_{\text{v}} = 9000$ m^{-1} (particle surface-to-volume ratio),

$\Delta h_{\mathrm{v}} = 18.6 \times 10^6$ J/kg (heat of combustion).

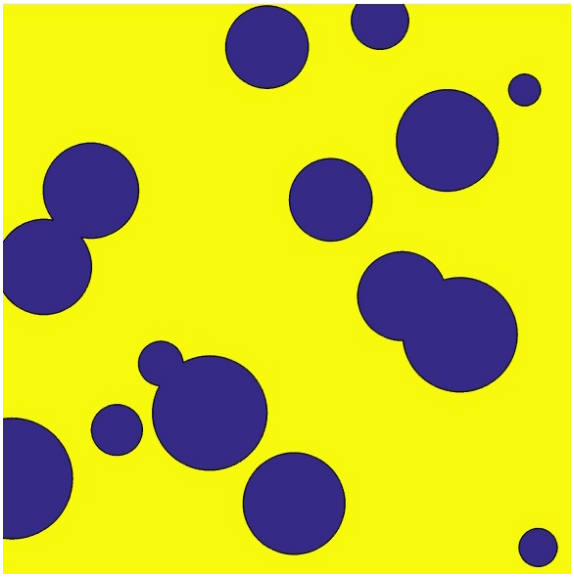

Figure 5.1: Heterogeneous biomass moisture content for OSSE test case. Blue area represents 20 % moisture content, yellow area represents 5 % moisture content.

Performance of Standalone Parameter Estimation

First, we verify the standalone parameter estimation algorithm based on the ETKF works well. For this purpose, we assume uncertainties only come from the near-surface wind; the wind speed u_{w} and the wind direction d_{w} are considered constant over time so that $\mathrm{x} = (u_{\mathrm{w}}, d_{\mathrm{w}})$.

The fire is ignited at time 0 s at $(x_{\mathrm{ign}}, y_{\mathrm{ign}}) = (90\ \mathrm{m}, 90\ \mathrm{m})$ as a circular front with radius $r_{\mathrm{ign}} = 10$ m. The true wind speed is $u_{\mathrm{w}}^{\mathrm{t}} = 3$ m/s, the true wind direction is $d_{\mathrm{w}}^{\mathrm{t}} = 180°$. Figure 5.2 shows the true trajectory of the fire in black solid lines. The prior estimate of the fire propagation is ignited at the same position

$(x_{\text{ign}}, y_{\text{ign}}) = (90\text{ m}, 90\text{ m})$ but with prior wind values that are subject to significant error with $u_{\text{w}}^{\text{b}} = 4$ m/s (25 % error) and $d_{\text{w}}^{\text{b}} = 225°$ (20 % error). The corresponding background trajectory of the fire is presented in Fig. 5.2 in blue lines (this is the mean background trajectory). Heterogeneities in the biomass fuel moisture content induce different front shapes between the true trajectory and the background trajectory.

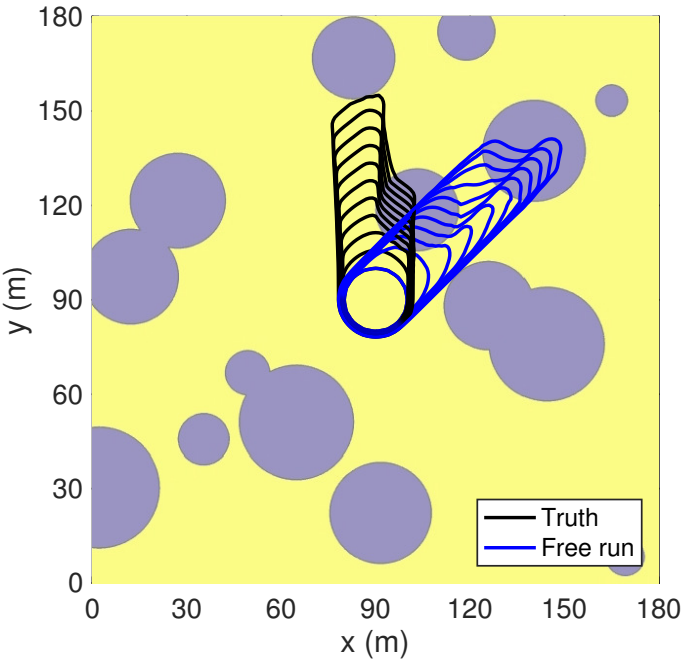

Figure 5.2: Time-evolving locations of the firelines at 10-s time intervals from initial time $t_0 = 0$ s until 100 s; OSSE test case – Comparison between true firelines (black solid lines) and free run or background (blue solid lines). The colormap corresponds to the biomass fuel moisture content field in Fig. 5.1.

We model the uncertainties of the wind parameters using the following standard deviation σ^{b}: 0.5 m/s for the wind speed u_{w}^{b}, and 25° for the wind direction d_{w}^{b}. In this configuration, we analyze the sensitivity of the results to the ensemble size N_e varying from 10 to 200. Figure 5.3 presents the mean and STD of the

analysis wind speed and wind direction, which are empirically derived from the N_e-ensemble. Figure 5.3 shows that the analysis mean values tend to converge towards the true values for $N_e > 30$; the STD being relatively constant as the ensemble size N_e increases. An ensemble of $N_e = 40$ members is thus enough for achieving a good performance of the ETKF algorithm. Note that in Chapter 3 we used an ensemble of 200 members for spatially-distributed parameter estimation; a higher number of members was required due to the larger dimension of the control vector that included multiple wind grid points in the vicinity of the fireline.

Figure 5.4 shows at time 50 s the ensemble of fireline positions before (left panel) and after (right panel) ETKF-based parameter estimation when $N_e = 40$. Before parameter estimation there is a wide scatter of the fireline positions due to the uncertain wind speed and wind direction. Once parameter estimation is applied, all members of the ensemble move towards the observed fireline at time 50 s. The analysis ensemble features a much reduced spread compared to the background ensemble. Note that we applied a medium RTPP inflation scheme ($\alpha = 0.5$) to keep a robust analysis ensemble spread, which is essential to avoid filter divergence issue. So the analysis ensemble still features some variability at the headfire.

Performance of Dual Estimation

Now we move to the dual state-parameter estimation. In this configuration, we assume uncertainties are due to both wind parameters and ignition location so that $\mathrm{x} = (u_\mathrm{w}, d_\mathrm{w}, c)$. We keep the same configuration with $u_\mathrm{w}^\mathrm{t} = 3$ m/s, $d_\mathrm{w}^\mathrm{t} = 180°$

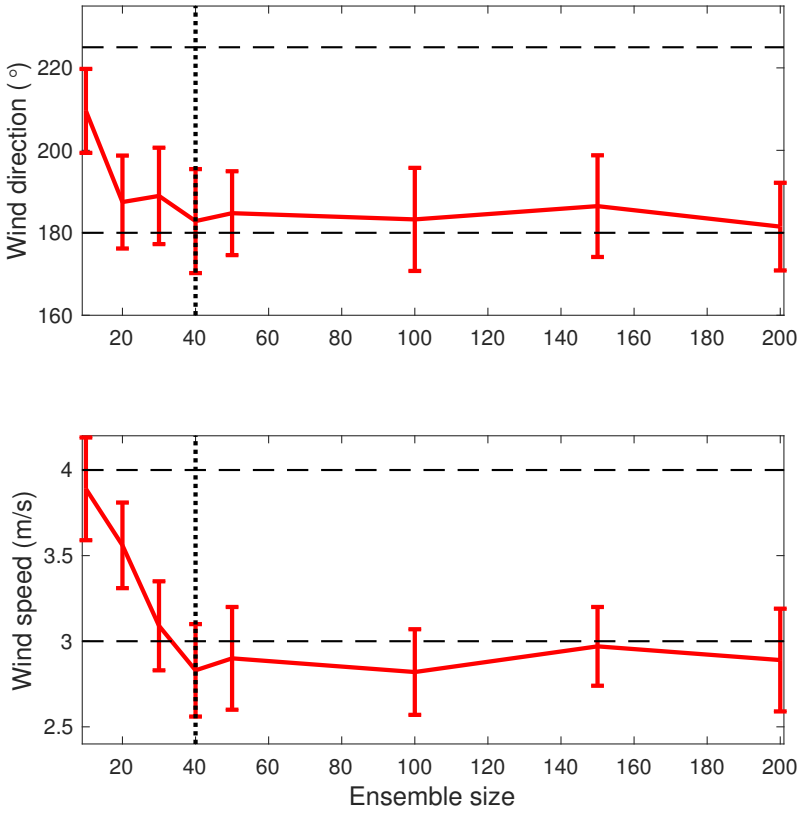

Figure 5.3: Convergence test of the ETKF-based parameter estimation algorithm for varying ensemble size N_e; OSSE test case with $\mathrm{x} = (u_\mathrm{w}, d_\mathrm{w})^T$ – Red solid lines represent the mean analysis estimate of the wind parameters (wind speed u_w in top panel; wind direction d_w in bottom panel); error bars represent their associated STD. Top dashed lines represent the mean background value of the wind parameters; bottom dashed lines represent their true value. Vertical dotted lines represent the ensemble size ($N_e = 40$) used in the present work.

for the true value of the control parameters and $u_\mathrm{w}^\mathrm{b} = 4$ m/s, $d_\mathrm{w}^\mathrm{b} = 225°$ for the background value of the control parameters. However, the background fire is also subject to an uncertain initial condition: the background fire is ignited at time 0 s at $(x_\mathrm{ign}, y_\mathrm{ign})^\mathrm{b} = (85\ \mathrm{m}, 85\ \mathrm{m})$ as a circular front with radius $r_{mathrmign} = 10$ m, while the true fire is ignited at $(x_\mathrm{ign}, y_\mathrm{ign})^\mathrm{t} = (90\ \mathrm{m}, 90\ \mathrm{m})$. For the state estimation part, the parameter λ value is set to $\lambda = 0.2$ featuring a medium nudging effect in

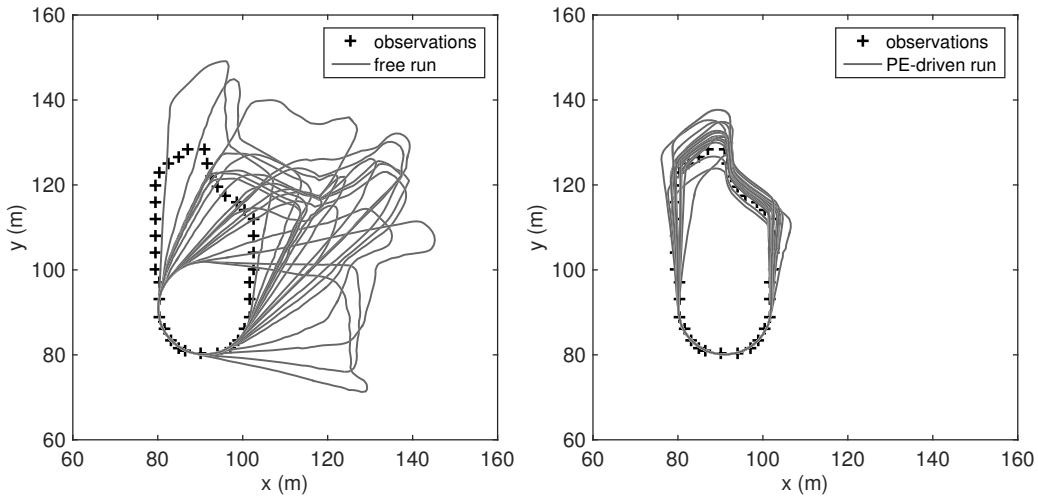

Figure 5.4: Comparison of fireline locations at time 50-s; OSSE test case with $N_e = 40$. Black cross symbols correspond to the true trajectory; gray lines correspond to the simulated firelines (left panel presents the background ensemble; right panel presents the posterior ensemble after parameter estimation–PE). For clarity purpose, only 20 members of the ensemble are plotted.

this case.

We compare in Fig. 5.5 the fireline positions obtained at the analysis time 50 s with different algorithms. The top left figure presents the free run obtained with an incorrect initial location and incorrect wind parameters (there is the additional uncertainty in the initial condition compared to Fig. 5.4 – left panel). Top right figure shows the posterior fireline positions obtained if we only estimate the wind parameters $(u_{\mathrm{w}}, d_{\mathrm{w}})$ using the ETKF algorithm. Due to the bias induced by the uncertain initial condition, the ETKF algorithm is unable to retrieve the true fireline position (i.e. to find values of the wind parameters that correspond to a simulated fireline position matching the true position). There is still a significant difference in the position of the head fire between the true fireline position and the analysis

ensemble. Bottom figures show the posterior fireline positions when applying the dual state-parameter estimation algorithm. The bottom left figure shows the results after the first step of the algorithm, i.e. the state estimation procedure. The bottom right figure shows the result after the second step of the algorithm, i.e. when both state estimation and parameter estimation are applied. Compared with the free run in the top left figure, we see that all ensemble members are already slightly nudged towards the observed fireline. The bias in the fireline position induced by the wrong ignition is decreased by state estimation (the magnitude of this correction is controlled by the gain λ value). When combined with parameter estimation, the algorithm is able to retrieve the observed fireline position with a small scatter among the ensemble members. The head fire is correctly located. There are still some small uncertainties at the rear fire due to the uncertainty in the initial condition. Figure 5.5 thus indicates that the dual state-parameter estimation algorithm achieves the best performance among the algorithms tested and can overcome uncertainties in both wind parameters and fire ignition position.

As a complement, we analyze the distribution of the posterior wind values when using the dual state-parameter estimation method with respect to the true wind values. Figure 5.6 shows the probability density function for both wind direction (top panel) and wind speed (right panel) control variables; this probability density function is derived empirically from the 40-member ensemble. This indicates that the dual state-parameter estimation approach provides posterior values that are closer to the truth than the standalone parameter estimation approach. It also significantly reduces the spread of the ensemble around the mean posterior

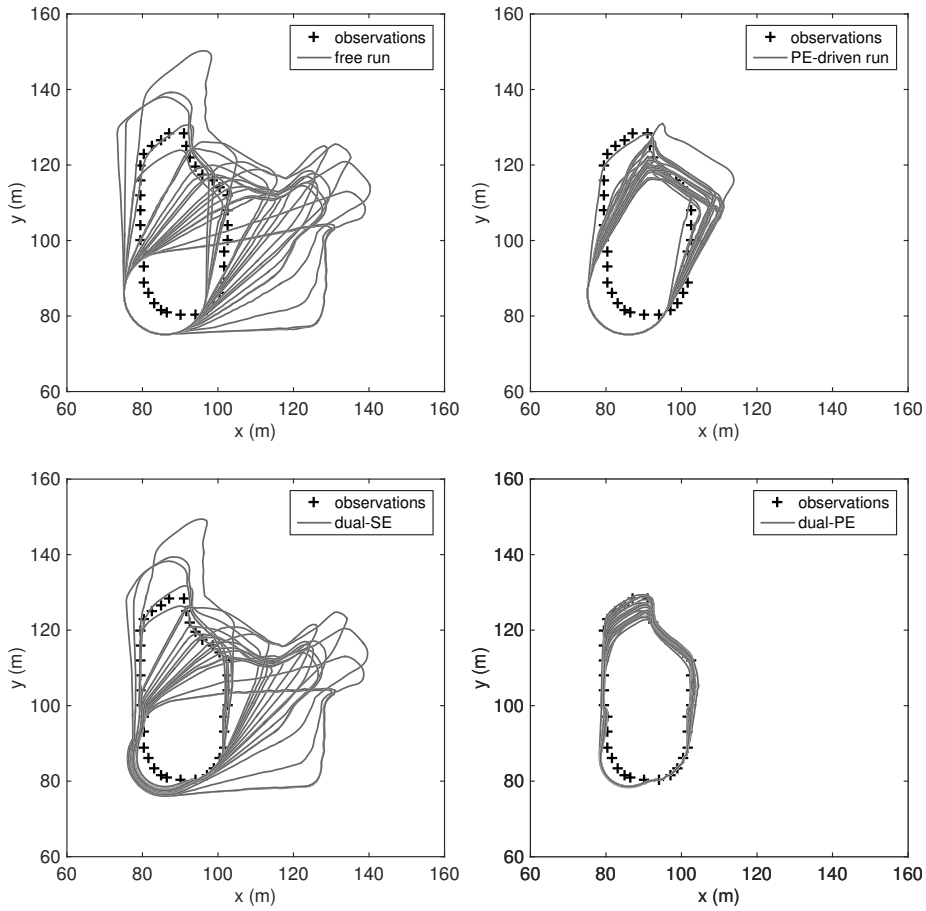

Figure 5.5: Comparison of fireline locations at time 50-s; OSSE test case with $N_e = 40$. Black cross symbols correspond to the true trajectory; gray lines correspond to the simulated firelines. Top left panel presents the background ensemble. Top right panel presents the posterior ensemble after standalone parameter estimation (PE). Bottom left panel presents the posterior ensemble after the first step of the dual state-estimation algorithm (state estimation). Bottom right panel presents the posterior ensemble after the second step of the dual state-estimation algorithm (both state estimation and parameter estimation). For clarity purpose, only 20 members of the ensemble are plotted.

value. The standalone parameter estimation is unable to find the truth due to the extra bias introduced by the wrong ignition location. This confirms the good performance of the dual state-parameter estimation approach that is able to (1) retrieve a

realistic fire behavior and (2) find more accurate estimates of the near-surface wind (in terms of mean and standard deviation).

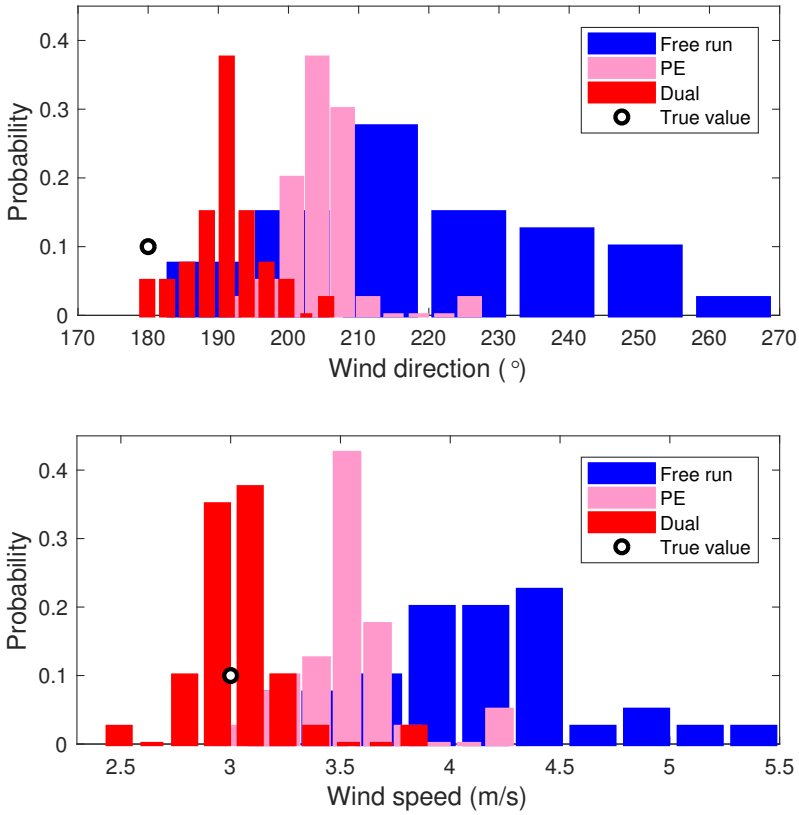

Figure 5.6: Probability density function of the wind parameters; OSSE test case with $N_e = 40$. Circle symbols represent the true value of the wind parameters. Blue histogram indicates the distribution of the wind parameters without data assimilation. Pink histogram indicates the posterior distribution of the wind parameters under standalone parameter estimation, Red histogram indicates the posterior distribution of the wind parameters under dual state-parameter estimation.

5.3 Validation Test: Application to the RxCADRE Experiment

5.3.1 Observation Data

For this validation work, we use the whole dataset from the RxCADRE experiment (duration: 780 s). Figure 5.7 (left) presents the map of flame arrival times, from $t = 0$ to 780 s (corresponding to the time interval [60 s; 840 s] in the RxCADRE dataset), showing the time at which the fireline arrives at a given pixel of the S5 burn lot. In the assimilation procedure, we assume that observations are available at 60-s time intervals. We generate the binary fields required at these observation times, which are obtained using the corresponding map of flame arrival times after filtering to remove small-scale holes and outliers. Figure 5.7 (right) shows the contour of these binary fields with a 60-s time interval.

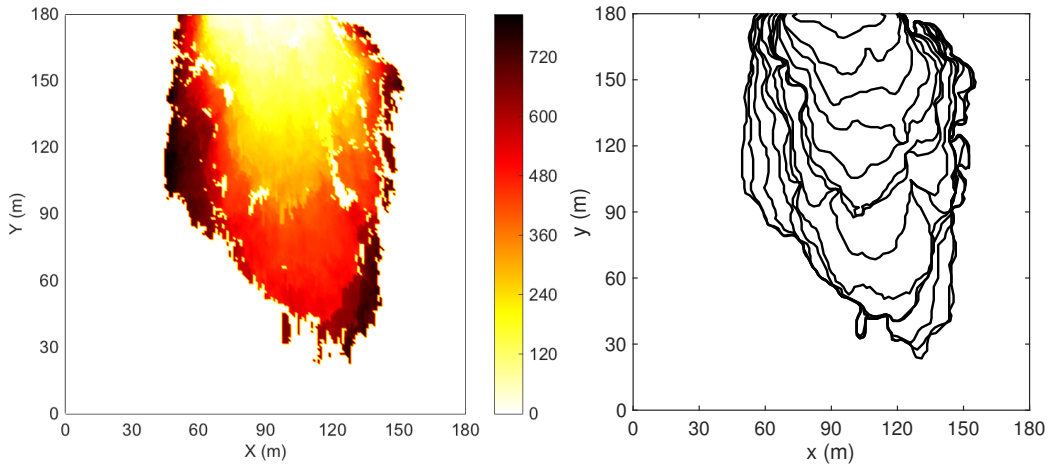

Figure 5.7: Dataset of the RxCADRE S5 fire. (left) Map of flame arrival times (0-780 s). (right) Series of contours (or "fireline") separating unburnt and burnt vegetation over the time period (0-780 s) with a 60-s time interval.

5.3.2 Numerical Settings

We now apply our dual state-parameter estimation method to the RxCADRE S5 experimental dataset with observations available at 60-s time intervals. The 780-s simulation time is divided into thirteen 60-s-long data assimilation cycles: the cycles are characterized by twelve analysis events, called $A_1 - A_{12}$ with A_i designating an analysis performed at time ($i\times$ 60 s) based on an observation made at the same time, and by twelve forecast events, called $F_1 - F_{12}$ with F_i corresponding to a forecast performed at time ((i+1) $\times$ 60 s) based on the analysis A_i. The quality of the analysis is evaluated through comparisons between predicted and observed fire propagation positions at time ($i\times$ 60 s); the quality of the forecast is evaluated through similar comparisons at time ((i+1) $\times$ 60 s).

The 2-D computational domain Ω is 180 m$\times$180 m (with a step size $\Delta x = 1$ m); the time window is 780 s (with a constant time step $\Delta t = 0.05$ s). In the simulations, we assume uniformly-distributed vegetation fuel as well as uniform and constant wind. The input fuel parameters to the Rothermel model are based on experimental measurements: the fuel depth is 0.2 m; the fuel surface loading is 0.28 kg/m^2; the fuel particle surface-to-volume ratio is 9,000 m^{-1}; the fuel moisture content is 10 %. For the input wind parameters, we start with an initial guess wind speed value 2 m/s, wind direction value 360°. Note that in Chapter 4 we use a wind direction value of 345° during a state estimation mode with the RxCADRE S5 data. Here we use a biased wind value from the measurement to check the performance of the dual estimation method on retrieving this 345° wind direction value. The corresponding

background STD is 0.5 m/s for wind speed and 25° for wind direction. The size of the ETKF-ensemble is $N_e = 40$. The LO-gain value is $\lambda = 0.2$. Also during the experiment, the fire is ignited near the northern boundary and propagated from north to south direction. In the simulation, we set the ignition at time 0 s at (90 m, 180 m) as a semi-circular front with radius = 15 m.

5.3.3 Performance of Dual Estimation

Comparison of Estimation Algorithms

We consider observations are available at 60-s time intervals. Figure 5.8 presents a comparison between simulated firelines and observation over the first assimilation cycle at analysis time 60 s. Top left figure corresponds to the free run. Top right figure corresponds to the analysis fireline estimate obtained by standalone parameter estimation. The analysis mean values are $u_{\mathrm{w}}^{\mathrm{a}} = 2.36$ m/s for wind speed and $d_{\mathrm{w}}^{\mathrm{a}} = 335°$ for wind direction; associated STD values are 0.30 m/s and 11.61°, respectively. Bottom left figure corresponds to standalone state estimation. Bottom right figure corresponds to the analysis fireline estimate obtained by dual state-parameter estimation. The analysis mean values are $u_{\mathrm{w}}^{\mathrm{a}} = 2.09$ m/s for wind speed and $d_{\mathrm{w}}^{\mathrm{a}} = 325°$ for wind direction; associated STD values are 0.27 m/s and 13.31°, respectively. These results show that there is a significant change in the wind parameters to track the observed fireline compared to the free run; the scatter of the wind parameters is reduced by a factor of two with respect to the background but is comparable for the analysis obtained with standalone parameter estimation

and dual state-parameter estimation. Dual state-parameter estimation outperforms standalone parameter estimation in terms of analysis performance by retrieving a more accurate fireline shape thanks to state estimation.

Note that free run and standalone state estimation have a similar computational cost, meaning that computing the LO-feedback term does not increase the computational cost required for the resolution of the propagation equation. When running ensemble simulations, either in standalone parameter estimation mode or in dual state-parameter estimation mode, the computational cost is approximately multiplied by a factor of two per model simulation (due to the simulation of reanalysis event). For a fixed number of processors, the total computational cost linearly scales with the ensemble size N_e. However, an advantage of EnKF methods is that model simulations during the forecast stage are independent and can be run simultaneously according to available computing facilities.

We now evaluate the forecast performance at future leading terms. Figure 5.9 compares the fireline positions at times 60 s, 120 s and 180 s. The fire spread model is run freely (without data assimilation) starting from the analysis fireline estimates shown in Fig. 5.8 at time 60 s until times 120 s and 180 s. Observation at time 60 s is assimilated; observations at times 120 s and 180 s are only used to compute diagnostics. Top left figure presents the simulation-observation comparison in free run mode. Top right figure presents the simulation-observation comparison in standalone parameter estimation mode. Bottom left figure presents the comparison in state estimation mode. Bottom right figure presents the comparison in dual state-parameter estimation mode. Results show that standalone state estimation cannot

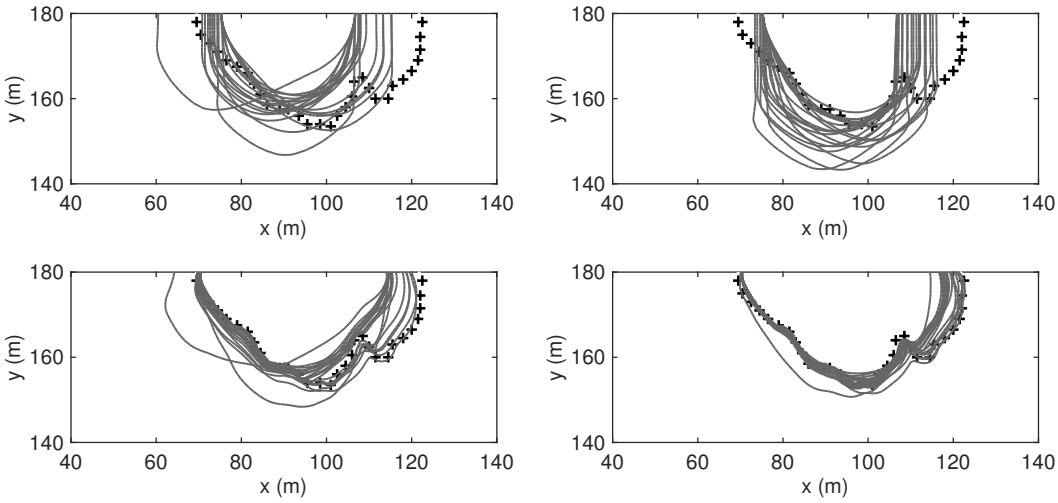

Figure 5.8: Comparison of fireline locations at analysis time 60 s with $N_e = 40$. Black cross symbols represent the observations, grey lines reprensent simulated fronts. Top left panel presents the background ensemble. Top right panel presents the posterior ensemble after standalone parameter estimation (PE). Bottom left panel presents the posterior ensemble after the first step of the dual state-estimation algorithm (state estimation). Bottom right panel presents the posterior ensemble after the second step of the dual state-estimation algorithm (both state estimation and parameter estimation). For clarity purpose, only 20 members of the ensemble are plotted.

capture the direction of fire spread unlike other approaches. Standalone parameter estimation and dual state-parameter estimation are able to forecast fireline positions that are consistent with the observations at times 120 s and 180 s. This implies that the background wind direction is not realistic and needs to be updated. In addition, combining state estimation with parameter estimation provides a better forecast of the fireline geometry on the flanks and of its irregularities than standalone parameter estimation. These results indicate the advantage of dual estimation method to correctly track the fire front propagation and to update the input parameters of the ROS model to retrieve the actual main wind conditions.

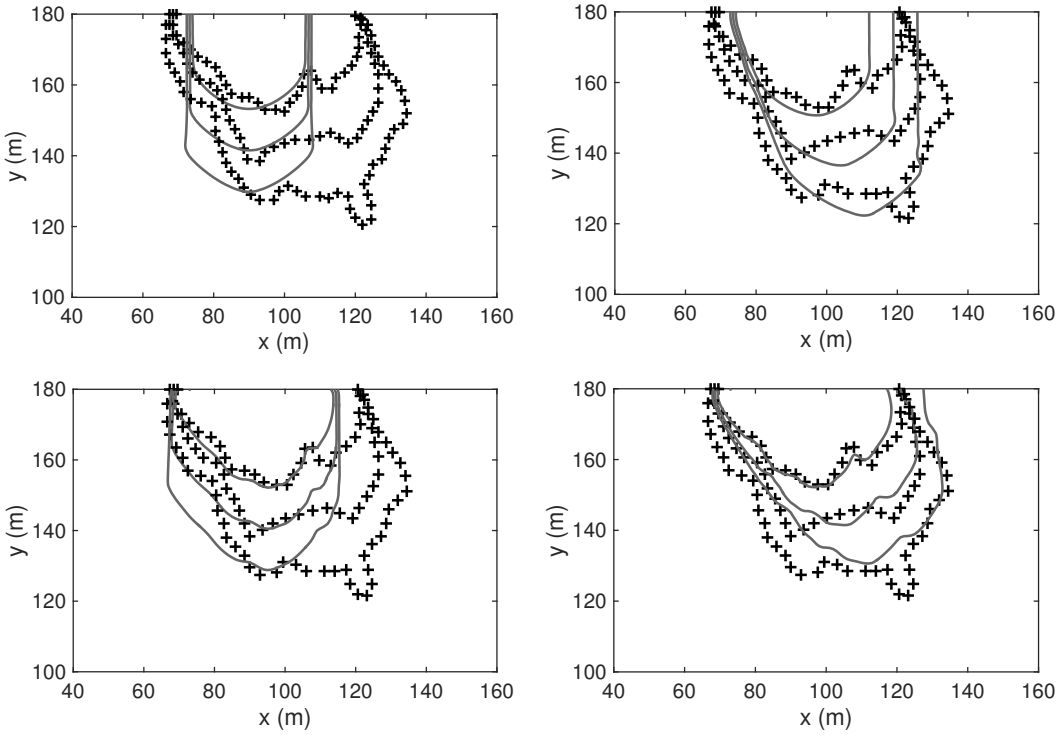

Figure 5.9: Front comparison at analysis time 60 s and forecast time 120 s and 180 s. Black cross symbols represent observation data, gray lines represent simulated fronts. Top left panel presents the free run. Top right panel presents the posterior front driven by standalone parameter estimation. Bottom left panel presents the posterior front driven by standalone state estimation, with the same λ value used in dual estimation (λ=0.2). Bottom right panel presents the posterior front driven by dual state-parameter estimation. For clarity purpose, only the mean fireline location is plotted.

To quantitatively evaluate the forecast performance, we introduce the Hausdorff distance as in Chapter 4 to compute the difference between simulated front and observed front in a unit of physical distance (meter or kilometer). While Hausdorff distance is our perferred choice when we need a diagnostic metric with the Euclidean distance concept, there are other various metrics to evaluate the discrepancy between two shapes. One instinctive choice is the Chan-Vese data-fitting functional [80] (see

Eq 2.14). Basically the smaller this similarity measure, the better match between two objects being compared with each other. Chan-Vese data fitting functional is not a normalized indice, so we choose another two similarity indices: Sorensen index and Jaccard index, which are normalized similarity indices [43]. Their score ranges between 0 and 1, while 1 indicates identical shapes. In practice, there is no rule to determine which indice should be used to best quantify the discrepancy between simulation and observation [109], so here we adopt these four indices as our diagnostic metrics.

Figure 5.10 presents the temporal variations of the diagnostics. For the free run, the observation-simulation discrepancy is high at the analysis time 60 s and increases with time over the forecast time period $[120; 180$ s$]$. State estimation drastically reduces the discrepancy at the analysis time but the positive effects of the analysis are rapidly lost during the forecast time period and the discrepancy rapidly follows the same trend as the free run. This indicates that there is no long persistence of the correction in the model state due to model bias (i.e. wrong near-surface wind conditions). Parameter estimation scores better than state estimation during the forecast time period, indicating the ability of the algorithm to retrieve physical values for the wind parameters. For parameter estimation, the observation-simulation discrepancy increases at a slower rate than the other approaches. So dual state-parameter estimation provides a good compromise with a good-quality analysis and a good-quality forecast over the time period $[120; 180$ s$]$, with an observation-simulation discrepancy that is similar to that of standalone parameter estimation at time 180 s.

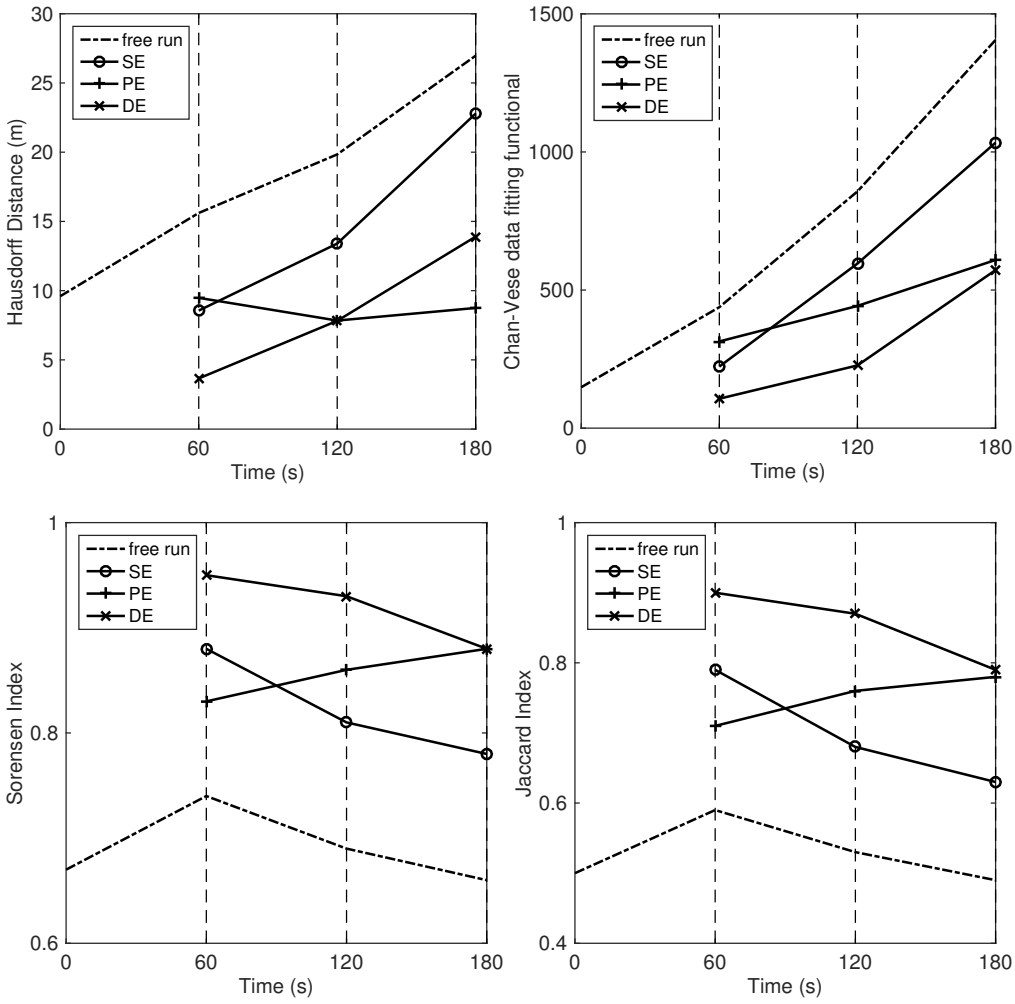

Figure 5.10: Time variations of observation-simulation discrepancy using diagnostics for the first assimilation cycle (analysis time 60 s, forecast time 120 s and 180 s). Dashed lines correspond to free runs; solid lines correspond to data-driven runs. Circle symbols correspond to standalone state estimation (SE); plus sign symbols correspond to standalone parameter estimation (PE); cross symbols correspond to dual state-parameter estimation (DE).

Sequential Application of Dual State-parameter Estimation

In previous section, we show the results obtained for the first assimilation cycle A_1/F_1. Here we also study assimilation cycles A_2–A_{12}/F_2–F_{12} by performing dual state-parameter estimation every 60 s, implying that the wind speed and wind direction are sequentially updated to track temporal changes in-between the assimilation cycles.

Figure 5.11 presents the temporal variations of the analysis wind speed and wind direction in terms of mean value and STD. Figures 5.12–5.13 present a comparison between the observed fireline and the mean model predictions for all assimilation cycles at analysis time and at forecast time (60 s after analysis time), respectively. Results show that the mean wind direction significantly changes over the first 180 s but then stabilizes near 350° until 720 s. This indicates that the wind is modified from northwestern wind to northern wind during the experiment. This is consistent with the location of the headfire along the observed firelines. As for the mean wind speed, it suddenly decreases over the time period [300; 360 s] due to possible wind stagnancy; this is also seen in the observed firelines that propagate rather slowly during the corresponding time period. Hence the analysis obtained at time 300 s and used to forecast the fireline over the time period [300; 360 s] propagates too fast in the headfire region compared to the observation (see panel at time 300 s in Fig. 5.12 and corresponding panel at time 360 s in Fig. 5.13). Also the analysis obtained at time 420 s and used to forecast the fireline over the time period [420; 480 s] propagates too slowly compared to the observation (see panel at time 420 s in Fig. 5.12

and corresponding panel at time 480 s in Fig. 5.13). The correction in the wind parameters that adequately tracks the observed fireline is thus obtained with a 60-s temporal shift. This means that if the wind conditions significantly change over the forecast time window, the forecast performance can degrade. Still, the error in the predicted fireline position can be controlled by updating sequentially near-surface wind parameters.

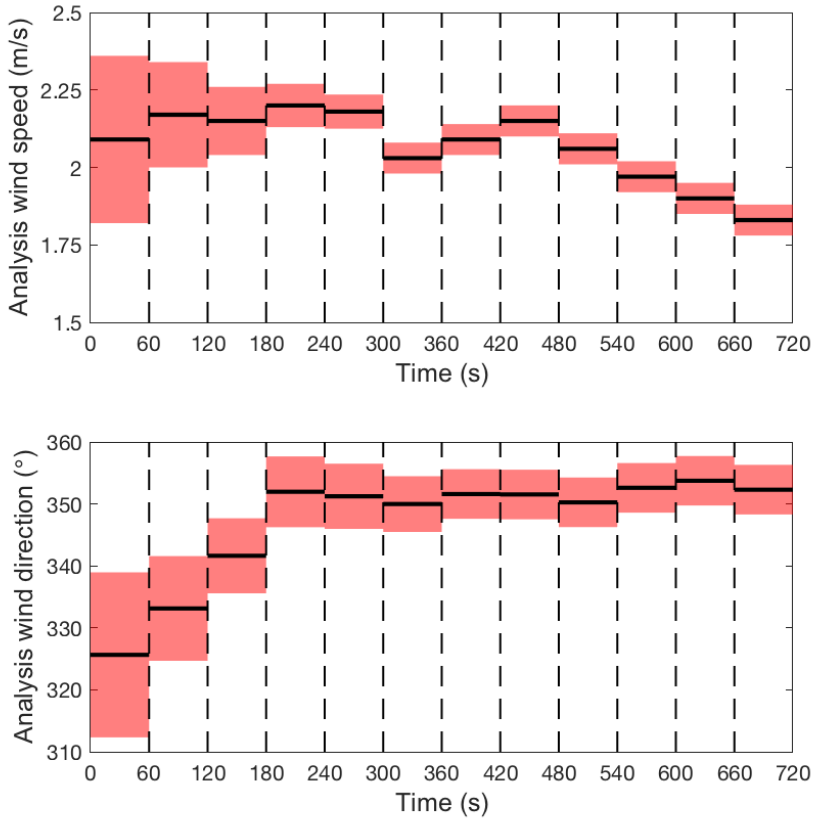

Figure 5.11: Time variations of wind speed (top) and wind direction (bottom) for all assimilation cycle $A_1 - A_{12}$. Wind values are updated every 60 s. Solid line corresponds to the mean value, light red area corresponds to the STD value, the STD values determine the lengths of each error bar above and below the mean value.

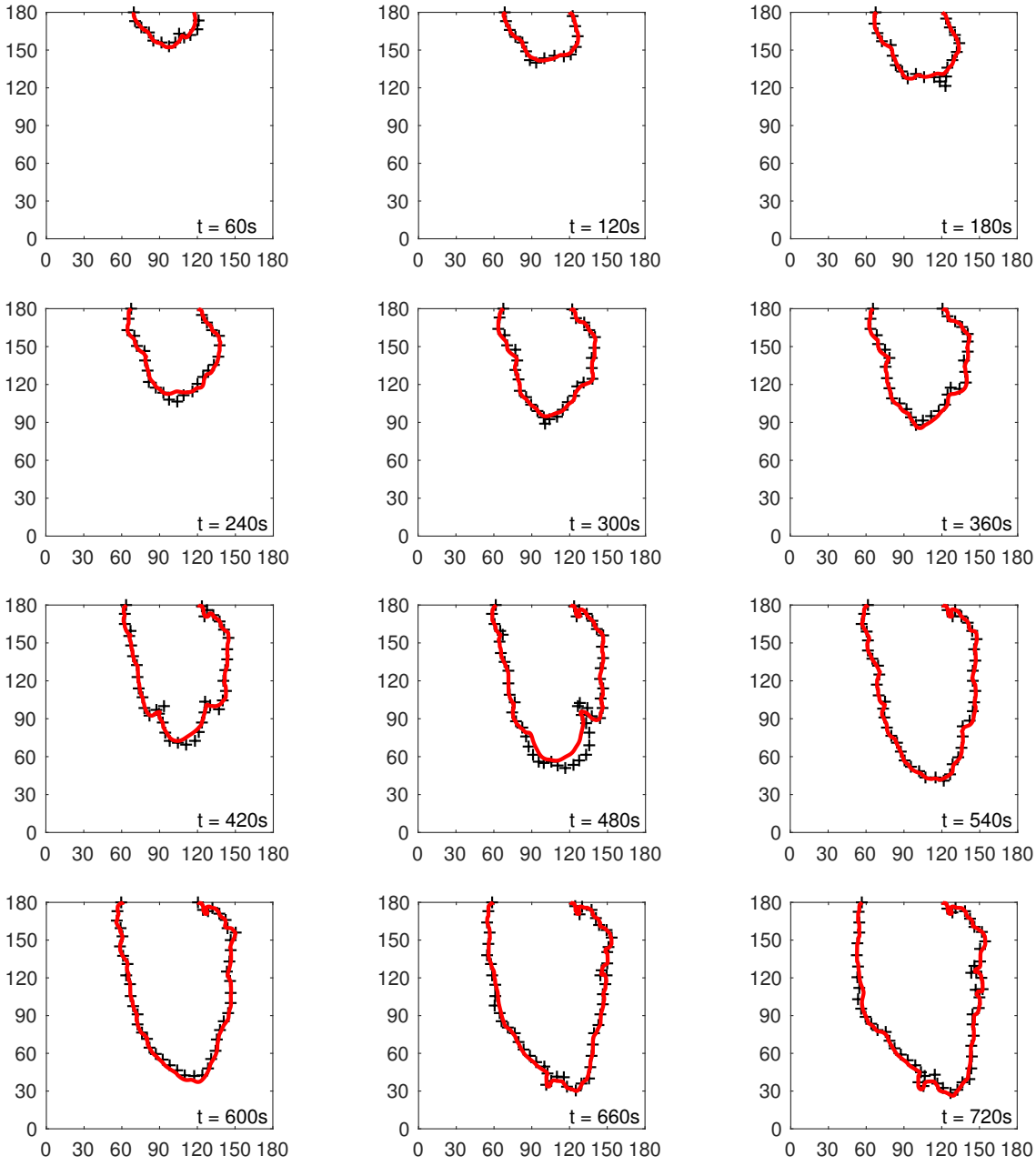

Figure 5.12: Comparison of simulated (solid red lines) and observed (cross symbols) fire front positions for RxCADRE S5 fire. The simulated firelines correspond to the mean analysis performed at events $A_1 - A_{12}$ ($60 \leq t \leq 720$ s). The fire spreads from top to bottom (North to South). The time interval between plotted perimeters is 60 s.

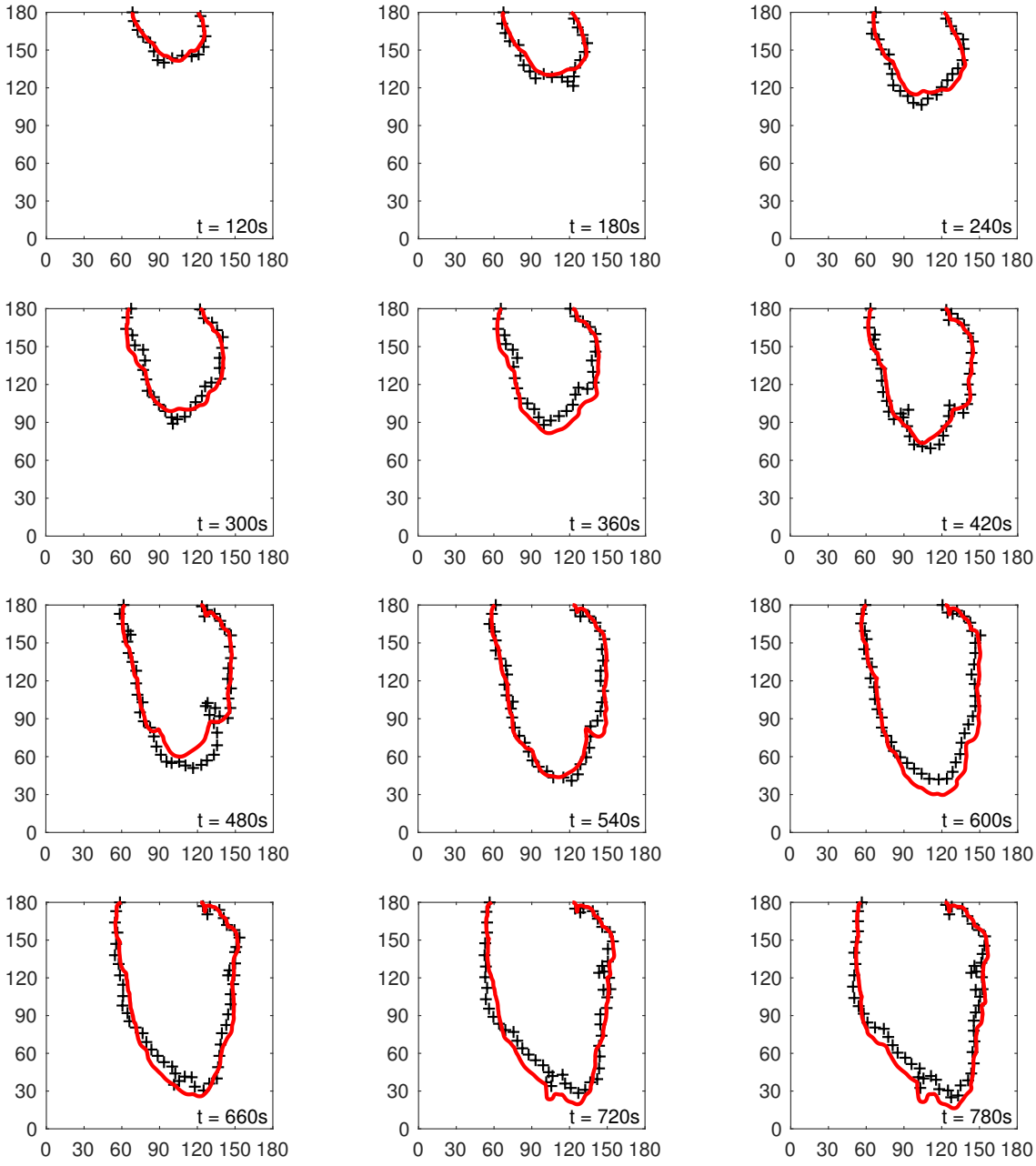

Figure 5.13: Comparison of predicted (solid red lines) and observed (cross symbols) fire front positions for RxCADRE S5 fire. The predictions correspond to the mean forecast performed at events $F_1 - F_{12}$ ($120 \leq t \leq 780$ s). The fire spreads from top to bottom (North to South). The time interval between plotted perimeters is 60 s.

Figure 5.14 presents the temporal variations of the Chan-Vese data-fitting functional and Hausdorff distance computed in free run mode and in data-driven mode. The evolution of the error in the free run mode is a smooth, continuously increasing function of time. In contrast, the evolution of the error in the data-driven mode is a discontinuous function: deviations of model predictions from observations are periodically reduced during the analysis events $A_1 - A_{12}$. After each analysis event, the error increases but remains bounded. Notice that the Hausdorff distance evaluates the maximum distance between simulated front and observed front in a physical space, it does not reflect the overall similarity between simulation and observation. When the observed front presents strong heterogeneities in a small local area which differentiates from the data-driven simulation (see for instance at time 180 s and 480 s), Hausdorff distance will capture this local discrepancy and resulting in larger values at the corresponding time.

Conclusion

In the present work, our main objective was to demonstrate the added value of dual state-parameter estimation to simultaneously reduce the main sources of uncertainty in wildland fire spread modeling. Parameter estimation was achieved with an ETKF algorithm that updates near-surface wind conditions as observations become available. State estimation was achieved using a cost-effective Luenberger observer adapted to the propagation equation to retrieve a more accurate shape and topology of the fireline for model restart. The merits of the dual state-parameter estimation

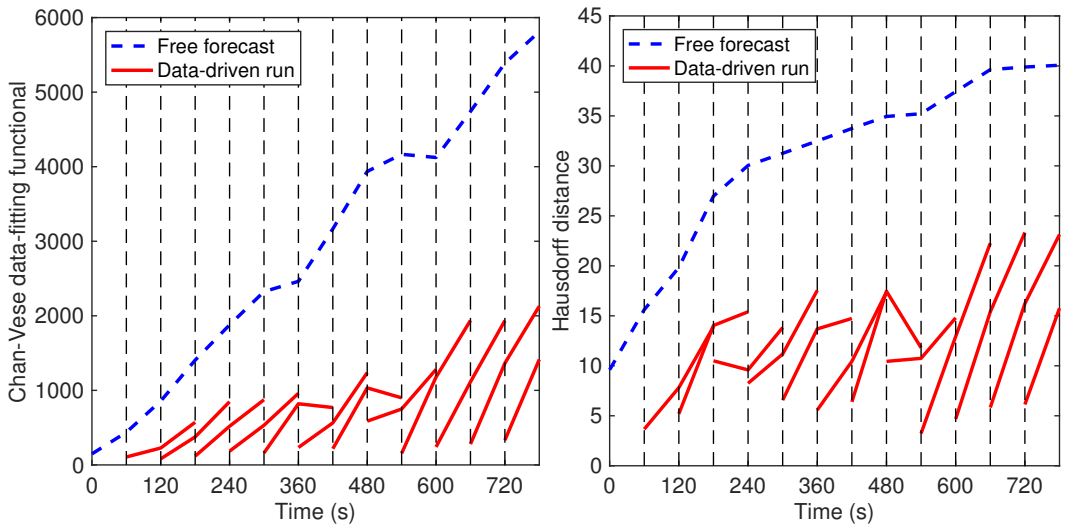

Figure 5.14: Time variations of the mean discrepancy (evaluated as Chan-Vese data-fitting functional (left) and Hausdorff distance (right)) between the predicted and observed fire front positions. Dashed blue line represents the comparison between the free forecast and observation, solid red line represents the comparison between data-driven run and observation. Time period: $0 \leq t \leq 780$ s.

algorithm were shown on OSSE verification test cases and on a validation test case against the 2012 RxCADRE S5 experimental data. Results show that parameter estimation is an essential component of data-driven wildfire spread modeling. This is useful to retrieve the environmental conditions that enhance wildland fire spread and to thereby produce more accurate forecasts of wildland fire spread. Results also show that combining state estimation and parameter estimation is powerful to address multiple sources of uncertainty, even those that are not part of the estimation targets, to account for model bias and thus retrieve physically-consistent model parameters. We believe these new methodological features are important to move towards data-driven wildfire spread simulation at regional scales and to address specificities of combustion applications more generally. This is useful to preserve a

coherent structure of the flaming front over time.

Chapter 6: Conclusions and Future Work

The objective of this research is to design and evaluate algorithmic solutions towards a dynamic data-driven modeling system for wildland fire spread monitoring and forecasting at regional scales. In the past, a data-driven wildfire spread prototype simulator was developed and validated against a small-scale grassland burn trial (4 m $\times$ 4 m) where the environmental conditions were considered uniform and where the fire front shape presented a simple geometry over time. The original prototype simulator suffers limited functionality when operated at regional scale, the extension of the modeling strategy requires careful design of the algorithms to deal with spatially-varying model inputs as well as highly heterogeneous fire front topologies. The author's contributions are therefore resolving some of the significant challenges for regional scale wildfire spread simulations which are detailed below.

6.1 Author's Contributions

The first contribution by the author is to extend the estimation system to cope with heterogeneous environmental conditions. A new grid-based spatialized parameter estimation approach is developed where the estimation targets are the spatially-varying input parameters of the ROS model. This approach changes the

sensitivity of the simulated fire front location with respect to the input parameters along the fire front, especially between the head fire and the flank fires. This approach gives model simulations more degrees of freedom to find physically consistent solutions. A dynamic distance-based localization scheme is also implemented to restrict the correction of the control parameters in the vicinity of the propagating fire front and to avoid spurious corrections far away from the fire. The effectiveness of this spatialized parameter estimation strategy has been demonstrated in the simulation case of FireFlux I controlled burning experimental data.

The second contribution by the author is to evaluate a more efficient and robust method to compute the distance between the observed and the simulated fire fronts. In past work, we treated the observed fire front as a discretized contour with a finite set of markers. The discrepancy between these simulated and observed fronts was then computed by pairing observed front markers with the same number of simulated markers. Such pairing may become difficult to operate for regional-scale wildfires that feature strong heterogeneities in the land surface conditions. This issue is addressed by introducing a new front shape comparison method where the burning area is treated as a moving object that can feature both position and topological errors. This method borrowed from object detection in image processing theory formulates a front shape similarity measure based on the Chan-Vese contour fitting functional. The effectiveness and robustness of this method have been demonstrated in the context of LO-based state estimation strategy against the RxCADRE S5 experiment, and also the Rim fire dataset.

The third contribution by the author is that a new dual state-parameter esti-

mation method is developed to estimate model state and model parameters simultaneously, in order to retrieve more physical values of control parameters and therefore improve short-time forecast performance. With test cases against the RxCADRE S5 experimental data, the proposed dual state-parameter estimation method is found to retrieve physical values for model parameters and yield better forecast performance compared with standalone state/parameter estimation method, making it a promising approach for real world wildfire scenarios where both model state and model parameters are known to suffer inevitable bias and uncertainties.

Futhermore, the enhanced data-driven modeling system has been evaluated against large scale controlled burn experiments and an accidental wildfire event, which is missing in previous work. In the present study, we have multiple datasets from the FireFlux I field-scale burn experiment (30 ha), the RxCADRE S5 burning experiment (4 ha) and the Rim Fire hazard (1041 km^2) to validate our model and algorithms. This is of paramount importance to quantitatively evaluate the new features before applied to operational wildland fire spread simulations.

We also provide a summary of the features of our EnKF algorithm used to either estimate model parameters or model state. Compared with the standard EnKF algorithm [33], we made following adaptations to make it work for wildfire spread simulations. First of all, we generally use a large ensemble size to aviod the covariance underestimation issue; and then we limit the control parameters within a certain range when generating the ensemble members. For instance, we set the minimum wind speed value to be larger than 0 m/s to aviod the unphysical values. Thirdly, we implement a dynamic distance-based localization scheme, and

a covariance inflation scheme as required for all kinds of EnKF methods. In the present study, we also assume that the observation error is very small, thus our objective is to check whether the data assimilation algorithms are able to nudge the simulation to match the observation data after EnKF update process. Note in general data assimilation provides an efficient framework to account for observation uncertainties, which will be a useful capability to monitor wildfire hazards where observation uncertainties are large.

6.2 Future Work

Additional work is required towards an operational data-driven application system for real-time wildfire spread monitoring and forecasting. At operation level, model inputs such as the wind field and biomass fuel should be fed and corrected in a more realistic way. It is known that the wind is one of the most influential environmental factors affecting wildland fire behavior. And in Chapter 3 we have discussed a grid-based spatially-distributed parameter estimation strategy to correct a spatially varying wind field. Because the fire front propagation is not sensitive to the wind values that are far away from the fire front, we adopted a dynamic distance-based localization to restrict the correction to the vicinity of the propagating fire front and to avoid spurious corrections far away from the fire. While this grid-based strategy is successful in providing a wind field that mimics observations, whether the posterior wind field has physical meaning is questionable. One way to improve the parameter estimation on the wind field could be to rely on a surface wind

model, such as WindNinja [110], WindNinja could be used to generate the required spatially varying wind field for the fire forward model, and to control the inputs of the surface wind model in parameter estimation mode. A more realistic posterior wind field could be achieved in this way since the prior ensemble would be the results of a model and not only of a Monte Carlo random sampling.

Besides, using now a long-term perspective, we could couple a cost-effective wildfire propagation model with a CFD-based atmospheric model to better account for time-varying weather conditions at regional scales [29, 30]. Thus a data-driven model with a better description of the wind forcing and of the fire feedback could contribute to reducing the model errors in the simulations. For instance, Filippi et al. [30] demonstrated the effectiveness of the Meso-NH/ForeFire coupling model for producing accurate atmospheric behavior simulations (such as plume size, transport dispersion and smoke concentration). The meso-scale atmospheric solver Meso-NH is a non-hydrostatic large eddy simulation solver [111] and is able to describe kilometer-scale to meter-scale atmospheric dynamics. The front-tracking solver ForeFire is a Lagrangian front-tracking solver applied to wildfire spread, evolving the location and width of the flame front according to a ROS model. The Meso-NH model forces wildfire behavior through the surface wind field, while ForeFire imposes heat and species fluxes as surface boundary conditions to Meso-NH. It is expected that a data-driven modeling strategy applied to a fire front-tracking model will have a beneficial effect on calculations of heat and species fluxes injected into an atmopsheric model, leading to a better representation of fire-atmospheric interactions.

Bibliography

[1] Mike D. Flannigan, Meg A. Krawchuk, William J. de Groot, B Mike Wotton, and Lynn M. Gowman. Implications of changing climate for global wildland fire. *International journal of wildland fire*, 18(5):483–507, 2009.

[2] Susan Solomon, Dahe Qin, Martin Manning, Kristen Averyt, and Melinda Marquis. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge university press, 2007.

[3] Richard Seager, Mingfang Ting, Isaac Held, Yochanan Kushnir, Jian Lu, Gabriel Vecchi, Huei-Ping Huang, Nili Harnik, Ants Leetmaa, Ngar-Cheung Lau, et al. Model projections of an imminent transition to a more arid climate in southwestern north america. *Science*, 316(5828):1181–1184, 2007.

[4] Mike Flannigan, Alan S. Cantin, William J. de Groot, Mike Wotton, Alison Newbery, and Lynn M. Gowman. Global wildland fire season severity in the 21st century. *Forest Ecology and Management*, 294:54–61, 2013.

[5] Brian J. Stocks, T.J. Lynham, B.D. Lawson, M.E. Alexander, C.E. Van Wagner, R.S. McAlpine, and D.E. Dube. Canadian forest fire danger rating system: an overview. *The Forestry Chronicle*, 65(4):258–265, 1989.

[6] William J. De Groot, Robert D. Field, Michael A. Brady, Orbita Roswintiarti, and Maznorizan Mohamad. Development of the indonesian and malaysian fire danger rating systems. *Mitigation and Adaptation Strategies for Global Change*, 12(1):165, 2007.

[7] Larry S. Bradshaw, John E. Deeming, Robert E. Burgan, and Jack D. Cohen. The 1978 national fire-danger rating system: technical documentation. *General Technical Report INT-169. Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 44 p.*, 169, 1984.

[8] Erin K. Noonan-Wright, Tonja S. Opperman, Mark A. Finney, G. Thomas Zimmerman, Robert C. Seli, Lisa M. Elenz, David E. Calkin, and John R. Fiedler. Developing the us wildland fire decision support system. *Journal of Combustion*, 2011, 2011.

[9] Alan G. McArthur. *Weather and grassland fire behaviour*. Forestry and Timber Bureau, Department of national Development, Commonwealth of Australia, 1966.

[10] Phil Cheney and Andrew Sullivan. *Grassfires: fuel, weather and fire behaviour*. CSIRO PUBLISHING, 2008.

[11] Jesús San-Miguel-Ayanz, Ernst Schulte, Guido Schmuck, Andrea Camia, Peter Strobl, Giorgio Liberta, Cristiano Giovando, Roberto Boca, Fernando Sedano, Pieter Kempeneers, et al. Comprehensive monitoring of wildfires in europe: the european forest fire information system (effis). In *Approaches to managing disaster-assessing hazards, emergencies and disaster impacts*. InTech, 2012.

[12] Andrew L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18(4):349–368, 2009.

[13] Domingos Viegas. Overview of forest fire propagation research. *Fire Safety Science*, 10:95–108, 2011.

[14] Mark A. Finney, Jack D. Cohen, Sara S. McAllister, and W Matt Jolly. On the need for a theory of wildland fire spread. *International journal of wildland fire*, 22(1):25–36, 2013.

[15] Michael Gollner, Arnaud Trouve, Ilkay Altintas, Jessica Block, Raymond de Callafon, Craig Clements, Anna Cortes, Evan Ellicott, Jean Baptiste Filippi, Mark Finney, et al. Towards data-driven operational wildfire spread modeling - report of the nsf-funded wifire workshop. Technical report, University of Maryland, 2015.

[16] Guillermo Rein, Natalie Cleaver, Clare Ashton, Paolo Pironi, and José L. Torero. The severity of smouldering peat fires and damage to the forest soil. *Catena*, 74(3):304–309, 2008.

[17] Andrew L. Sullivan. Wildland surface fire spread modeling, 1990-2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire*, 18:369–386, 2009.

[18] Andrew L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire*, 18(4):387–403, 2009.

[19] William Mell, Mary Ann Jenkins, Jim Gould, and Phil Cheney. A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire*, 16(1):1–22, 2007.

[20] Rodman Linn, Jon Reisner, Jonah J Colman, and Judith Winterkamp. Studying wildfire behavior using firetec. *International journal of wildland fire*, 11(4):233–246, 2002.

[21] Rod Linn. Coupled fire/atmosphere modeling, firetec. http://www.lanl.gov/orgs/ees/ees16/FIRETEC.shtml.

[22] Richard C. Rothermel. A mathematical model for predicting fire spread in wildland fuels. Technical report, U.S. Department of Agriculture, Forest Service, Intermountain forest and range experiment station, Ogden, Utah, USA, 1972.

[23] Patricia L. Andrews, Miguel G. Cruz, and Richard C. Rothermel. Examination of the wind speed limit function in the rothermel surface fire spread model. *International Journal of Wildland Fire*, 22:959–969, 2013.

[24] Mark A. Finney. FARSITE: Fire Area Simulator - Model Development and Evaluation. Technical Report February, US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1998.

[25] D. Chong, K.G. Tolhurst, T.J. Duff, and B. Cirulis. Sensitivity analysis of phoenix rapidfire. *Bushfire CRC, University of Melbourne*, 2013.

[26] C. Tymstra, R.W. Bryce, B.M. Wotton, S.W. Taylor, O.B. Armitage, et al. Development and structure of prometheus: the canadian wildland fire growth simulation model. *Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Information Report NOR-X-417.(Edmonton, AB)*, 2010.

[27] Jean-Baptiste Filippi, Frédéric Bosseur, Céline Mari, Christine Lac, Patrick Le Moigne, Bénédicte Cuenot, Denis Veynante, Daniel Cariolle, and Jacques-Henri Balbi. Coupled atmosphere-wildland fire modelling. *Journal of Advances in Modeling Earth Systems*, 1(4), 2009.

[28] Jean-Baptiste Filippi, Frédéric Bosseur, Xavier Pialat, Paul-Antoine Santoni, Susanna Strada, and Céline Mari. Simulation of coupled fire/atmosphere interaction with the mesonh-forefire models. *Journal of Combustion*, 2011, 2011.

[29] Jan Mandel, Jonathan D. Beezley, and Adam K. Kochanski. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development*, 4:591–610, 2011.

[30] Jean-Baptiste Filippi, Xavier Pialat, and Craig B Clements. Assessment of forefire/meso-nh for wildland fire/atmosphere coupled simulation of the fireflux experiment. *Proceedings of the Combustion Institute*, 34(2):2633–2640, 2013.

[31] Edwin Jimenez, M. Yousuff Hussaini, and Scott Goodrick. Quantifying parametric uncertainty in the Rothermel model. *International Journal of Wildland Fire*, 17:638–649, 2008.

[32] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 2003.

[33] Geir Evensen. *Data assimilation: the ensemble Kalman filter*. Springer, Berlin, 2007.

[34] Hamid Moradkhani, Soroosh Sorooshian, Hoshin V. Gupta, and Paul R. Houser. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28:135–147, 2005.

[35] Philippe Moireau, Dominique Chapelle, and Patrick Le Tallec. Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 197(6–8):659 – 677, 2008.

[36] Juan Jose Ruiz, Manuel Pulido, and Takemasa Miyoshi. Estimating model parameters with ensemble-based data assimilation: A review. *Journal of the Meteorological Society of Japan. Ser. II*, 91(2):79–99, 2013.

[37] Jan Mandel, Lynn S. Bennethum, Jonathan D. Beezley, Janice L. Coen, Craig C. Douglas, Minjeong Kim, and Anthony Vodacek. A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, 79(3):584–606, 2008.

[38] Jan Mandel, Jonathan D. Beezley, Janice L. Coen, and Minjeong Kim. Data assimilation for wildland fires. *IEEE control systems*, 29(3), 2009.

[39] Mónica Denham, Kerstin Wendt, Germán Bianchini, Ana Cortés, and Tomàs Margalef. Dynamic data-driven genetic algorithm for forest fire spread prediction. *Journal of Computational Science*, 3:398–404, 2012.

[40] Mélanie C. Rochoux, B. Delmotte, B. Cuenot, S. Ricci, and A. Trouvé. Regional-scale simulations of wildland fire spread informed by real-time flame front observations. *Proceedings of the Combustion Institute*, 34:2641–2647, 2013.

[41] Mélanie C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé. 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, 2014.

[42] Mélanie C. Rochoux, C. Emery, S. Ricci, B. Cuenot, and A. Trouvé. 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(8):1721–1739, 2015.

[43] O. Rios, E. Pastor, M.M. Valero, and E. Planas. Short-term fire front spread prediction using inverse modelling and airborne infrared images. *International Journal of Wildland Fire*, 25:1033–1047, 2016.

[44] Germán Bianchini, Ana Cortés, Tomàs Margalef, and Emilio Luque. S 2 f 2 m–statistical system for forest fire management. In *International Conference on Computational Science*, pages 427–434. Springer, 2005.

[45] Germán Bianchini, Ana Cortés, Tomàs Margalef, and Emilio Luque. Improved prediction methods for wildfires using high performance computing: A comparison. *Computational Science–ICCS 2006*, pages 539–546, 2006.

[46] Kerstin Wendt, Mónica Denham, Ana Cortés, and Tomas Margalef. Evolutionary optimisation techniques to estimate input parameters in environmental emergency modelling. In *Computational Optimization and Applications in Engineering and Industry*, pages 125–143. Springer, 2011.

[47] Feng Gu and Xiaolin Hu. Towards applications of particle filters in wildfire spread simulation. In *Proceedings of the 40th Conference on Winter Simulation*, pages 2852–2860. Winter Simulation Conference, 2008.

[48] Feng Gu and Xiaolin Hu. Analysis and quantification of data assimilation based on sequential monte carlo methods for wildfire spread simulation. *International Journal of Modeling, Simulation, and Scientific Computing*, 1(04):445–468, 2010.

[49] Martin Vejmelka, Adam K Kochanski, and Jan Mandel. Data assimilation of fuel moisture in wrf-sfire. *arXiv preprint arXiv:1309.0159*, 2013.

[50] Martin Vejmelka, Adam K Kochanski, and Jan Mandel. Data assimilation of dead fuel moisture observations from remote automated weather stations. *International Journal of Wildland Fire*, 25(5):558–568, 2016.

[51] Jan Mandel, Adam K Kochanski, Martin Vejmelka, and Jonathan D. Beezley. Data assimilation of satellite fire detection in coupled atmosphere-fire simulation by wrf-sfire. *arXiv preprint arXiv:1410.6948*, 2014.

[52] W.B. da Silva, M.C. Rochoux, HRB Orlande, M.J. Colaço, O. Fudym, M. El Hafi, B. Cuenot, and S. Ricci. Application of particle filters to regional-scale wildfire spread. *High Temperatures High Pressures*, 43:415–440, 2014.

[53] Chris Lautenberger. Wildland fire modeling with an eulerian level set method and automated calibration. *Fire Safety Journal*, 62, Part C:289 – 298, 2013.

[54] Craig B. Clements, Ryan Perna, Meongdo Jang, Daegyun Lee, Monica Patel, Susan Street, Shiyuan Zhong, Scott Goodrick, Ju Li, Brian E. Potter, et al. Observing the dynamics of wildland grass fires: Firefluxa field validation experiment. *Bulletin of the American Meteorological Society*, 88(9):1369–1382, 2007.

[55] Gail Wells. Capturing fire: Rxcadre takes fire measurements to whole new level. 2013.

[56] Darold E. Ward and Colin C. Hardy. Smoke emissions from wildland fires. *Environment International*, 17(2-3):117–134, 1991.

[57] Roger D. Ottmar. Wildland fire emissions, carbon, and climate: modeling fuel consumption. *Forest Ecology and Management*, 317:41–50, 2014.

[58] Ronan Paugam, Martin J. Wooster, and Gareth Roberts. Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread. *Geoscience and Remote Sensing*, 51:3385–3399, 2013.

[59] Martin J. Wooster, Gareth Roberts, Alistair MS Smith, Joshua Johnston, Patrick Freeborn, Stefania Amici, and Andrew T. Hudak. Thermal remote sensing of active vegetation fires and biomass burning events. In *Thermal Infrared Remote Sensing*, pages 347–390. Springer, 2013.

[60] Philip J. Riggan and Robert G. Tissell. Airborne remote sensing of wildland fires. In A. Bytnerowicz, M. Arbaugh, C. Andersen, and A. Riebau, editors, *Wildland Fires and Air Pollution. Developments in Environmental Science 8*, pages 139–168. Elsevier, Amsterdam, The Netherlands, 2009.

[61] S. Clandillon and H. Yesou. Gmes emergency response - fire mapping experience within eo-based rapid mapping. In *EGU General Assembly 2011*, volume 13, 2011.

[62] Vincent G. Ambrosia and Steven S. Wegener. Unmanned airborne platforms for disaster remote sensing support. In Pei-Gee and Peter Ho, editors, *Geoscience and Remote Sensing*, chapter 5, pages 91–114. InTech, Vukovar, Croatia, 2009.

[63] Marcos Rodrigues and Juan de la Riva. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental modelling & software*, 57:192–201, 2014.

[64] Mélanie C. Rochoux, Charlotte Emery, Sophie Ricci, B Cuenot, and A Trouvé. Towards predictive simulation of wildfire spread at regional scale using ensemble-based data assimilation to correct the fire front position. *Fire Safety Science*, 11:1443–1456, 2014.

[65] William H. Frandsen. Fire spread through porous fuels from the conservation of energy. *Combustion and Flame*, 16(1):9–16, 1971.

[66] CE Van Wagner. Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*, 7(1):23–34, 1977.

[67] Mélanie C. Rochoux. *Towards a more comprehensive monitoring of wildfire spread: contributions of model evaluation and data assimilation strategies*. PhD thesis, Ecole Centrale Paris, France, 2014.

[68] Ronald G. Rehm and Randall J. McDermott. Fire front propagation using the level set method. Technical Report March, National Institute of Standards and Technology, 2009.

[69] Gwynfor D. Richards. An elliptical growth model of forest fire fronts and its numerical solution. *International Journal for Numerical Methods in Engineering*, 30(6):1163–1179, 1990.

[70] Gwynfor D. Richards. A general mathematical framework for modeling two-dimensional wildland fire spread. *International Journal of Wildland Fire*, 5(2):63–72, 1995.

[71] Anthony S. Bova, William E. Mell, and Chad M. Hoffman. A comparison of level set and marker methods for the simulation of wildland fire front propagation. *International Journal of Wildland Fire*, 25:229–241, 2016.

[72] Cong Zhang, Mélanie Rochoux, Wei Tang, Michael Gollner, Jean-Baptiste Filippi, and Arnaud Trouvé. 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*, 91:758–767, 2017.

[73] Mélanie C. Rochoux, Annabelle Collin, Cong Zhang, Arnaud Trouvé, Didier Lucor, and Philippe Moireau. Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for eikonal equation. *ESAIM: Proceedings and Surveys*, pages 1–22, 2017.

[74] Cong Zhang, Annabelle Collin, Philippe Moireau, Arnaud Trouvé, and Mélanie Rochoux. Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the rxcadre field-scale experiment. Under review for publication in Proceedings of the Combustion Institute.

[75] Annabelle Collin. *Asymptotic analysis in cardiac electrophysiology: applications in modeling and in data assimilation*. PhD thesis, Université Pierre et Marie Curie-Paris VI, France, 2014.

[76] Annabelle Collin, Dominique Chapelle, and Philippe Moireau. A Luenberger observer for reaction–diffusion models with front position data. *J Comput Phys*, 300(C):288–307, November 2015.

[77] Annabelle Collin, Dominique Chapelle, and Philippe Moireau. Sequential State Estimation for Electrophysiology Models with Front Level-Set Data Using Topological Gradient Derivations. In *Proceedings of the 8th International Conference FIMH*, volume LNCS 9126, pages 402–411, 2015.

[78] François-Xavier Le Dimet, Innocent Souopgui, Olivier Titaud, Victor Shutyaev, and M Yussuf Hussaini. Toward the assimilation of images. *Nonlinear Processes in Geophysics*, 22:15–32, 2014.

[79] Olivier Titaud, Arthur Vidard, and Innocent Souopgui. Assimilation of image sequences in numerical models. *Tellus A: Dynamic Meteorology and Oceanography*, 62(1):30–47, 2010.

[80] Tony F. Chan and Luminita A. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. 2001.

[81] Stanley Osher and Ronald Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153. Applied Mathematical Sciences - Springer, 2003.

[82] Jung-ha An and Yunmei Chen. Region based image segmentation using a modified mumford-shah algorithm. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 733–742. Springer, 2007.

[83] Björn Engquist, Anna-Karin Tornberg, and Richard Tsai. Discretization of dirac delta functions in level set methods. *Journal of Computational Physics*, 207(1):28–51, 2005.

[84] Thomas Bengtsson, Chris Snyder, and Doug Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.

[85] Jeffrey L. Anderson and Stephen L. Anderson. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999.

[86] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[87] Jeffrey S Whitaker and Thomas M Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002.

[88] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162, 1994.

[89] Geir Evensen and Peter Jan Van Leeuwen. Assimilation of geosat altimeter data for the agulhas current using the ensemble kalman filter with a quasi-geostrophic model. *Monthly Weather Review*, 124(1):85–96, 1996.

[90] Peter L. Houtekamer and Herschel L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811, 1998.

[91] Jeffrey L. Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.

[92] Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar. Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.

[93] Cheng Da and Chu-chun Chang. Fortran-version maooam-das user guide. Technical report, Department of Atmospheric and Oceanic Science, University of Maryland, 2018.

[94] Fuqing Zhang, Chris Snyder, and Juanzhen Sun. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. *Monthly Weather Review*, 132(5):1238–1253, 2004.

[95] Gérald Desroziers, Loic Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613):3385–3396, 2005.

[96] L.M. Stewart, Sarah L. Dance, Nancy K. Nichols, J.R. Eyre, and J. Cameron. Estimating interchannel observation-error correlations for iasi radiance data in the met office system. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1236–1244, 2014.

[97] Rodman R. Linn and Philip Cunningham. Numerical simulations of grass fires using a coupled atmosphere–fire model: basic fire behavior and dependence on wind speed. *Journal of Geophysical Research: Atmospheres*, 110(D13), 2005.

[98] Shu-Chih Yang, Eugenia Kalnay, and Brian Hunt. Handling nonlinearity in an ensemble kalman filter: experiments with the three-variable lorenz model. *Monthly Weather Review*, 140(8):2628–2646, 2012.

[99] Zhiyong Meng and Fuqing Zhang. Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part iii: Comparison with 3dvar in a real-data case study. *Monthly Weather Review*, 136(2):522–540, 2008.

[100] Björn Engquist, Anna-Karin Tornberg, and Richard Tsai. Discretization of dirac delta functions in level set methods. *Journal of Computational Physics*, 207(1):28–51, 2005.

[101] Scott D. Roth. Ray casting for modeling solids. *Computer Graphics and Image Processing*, 18(2):109–144, 1982.

[102] Lin He, Chiu-Yen Kao, and Stanley Osher. Incorporating topological derivatives into shape derivatives based level set methods. *J Comput Phys*, 225(1):891–909, 2007.

[103] Felix Hausdorff. *Mengenlehre*. Walter de Gruyter Berlin, 1927.

[104] Daniel P. Huttenlocher, Gregory A. Klanderman, and William J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.

[105] M.P. Dubuisson and Anil K. Jain. A modified hausdorff distance for object matching. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 566–568. IEEE, 1994.

[106] Joseph J. O'Brien, E. Louise Loudermilk, Benjamin Hornsby, Andrew T. Hudak, Benjamin C. Bright, et al. High-resolution infrared thermography for capturing wildland fire behaviour: RxCADRE 2012. *International Journal of Wildland Fire*, 25(1):62–75, 2016.

[107] Jamie M. Lydersen, Malcolm P. North, and Brandon M. Collins. Severity of an uncharacteristically large wildfire, the rim fire, in forests with relatively restored frequent fire regimes. *Forest Ecology and Management*, 328:326–334, 2014.

[108] Yannick Trémolet. Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 133:1267–1280, 2007.

[109] Jean-Baptiste Filippi, Vivien Mallet, and Bahaa Nader. Evaluation of forest fire models on a large observation database. *Natural Hazards and Earth System Sciences*, 14(11):3077–3091, 2014.

[110] Jason Forthofer, Kyle Shannon, and Bret Butler. 4.4 simulating diurnally driven slope winds with windninja. *USDA Forest Service, Rocky Mountain Research Station, Missoula, MT*, 2009.

[111] Jean Philippe Lafore, Joël Stein, Nicole Asencio, Philippe Bougeault, Véronique Ducrocq, Jacqueline Duron, Claude Fischer, Philippe Héreil, Patrick Mascart, Valéry Masson, et al. The meso-nh atmospheric simulation system. part i: Adiabatic formulation and control simulations. In *Annales Geophysicae*, volume 16, pages 90–109. Springer, 1997.