

SMOOTHING WITH APPLICATION TO STOCHASTIC FIRE GROWTH MODELLING

by

Xi Wang

B.Sc., Simon Fraser University, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

August 2017

© Xi Wang, 2017

The undersigned certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis entitled: SMOOTHING WITH APPLICATION TO STOCHASTIC FIRE GROWTH MODELLING submitted by XI WANG in partial fulfilment of the requirements of the degree of Master of Science

Dr. W.J. Braun, Faculty of Arts & Sciences
Supervisor, Professor

Dr. J. Loeppky, Faculty of Arts & Sciences
Supervisory Committee Member, Professor

Dr. S.X. Wang, Faculty of Arts & Sciences
Supervisory Committee Member, Professor

Dr. J. Brinkerhoff, Faculty of Applied Sciences
University Examiner, Professor

August 3, 2017
(Date Submitted to Grad Studies)

Abstract

Modelling wildland fire spread stochastically is an important way to incorporate the uncertainty associated with this phenomenon. Fitting such a model to data from remote-sensed images could be used to provide accurate fire spread risk maps. We study a particular model from this perspective. One objective of this thesis is to verify the model on data collected under experimentally controlled conditions. We present the analysis of data from small-scale experimental fires that were digitally recorded. Data extraction and processing methods and issues are discussed, along with an estimation methodology. A critical part of the estimation methodology revolves around the smoothing of observed counts of burning and burnt out pixels as functions of elapsed time. We employ nonparametric regression for this purpose and consider two bias reduction strategies as possible ways to obtain more accurate estimates of the parameters underlying the stochastic fire spread model. An argument for partial validation of the model is also provided.

Table of Contents

Abstract	iii
Table of Contents	iv
List of Tables	vi
List of Figures	viii
Acknowledgements	xiv
Dedication	xv
Chapter 1: Introduction	1
Chapter 2: Literature Review	4
2.1 Wildland Fire Behaviour and Models	4
2.2 The Boychuk Fire Spread Models	5
2.3 Wildfire and Experimental Data	6
2.4 Nonparametric Regression	7
2.4.1 Nadaraya-Watson Estimator	9
2.4.2 Local Linear Estimator	9
2.4.3 Local Cubic Estimator	10
2.5 Bias Reduction Strategies	10
2.5.1 Double-smoothing Local Linear Estimator	10
2.5.2 Data Sharpening	11
2.5.3 Choice of Bandwidth	13
2.5.4 Boundary Correction	13
Chapter 3: Smouldering Experiments	14
3.1 Apparatus	14
3.2 Data Extraction and Segmentation	16

TABLE OF CONTENTS

Chapter 4: Fitting a Stochastic Fire Growth Model to Data .	19
4.1 Model Fitting	19
4.2 Moment and Continuous Least-Squares Estimation	20
4.3 Results	21
4.3.1 Parameter Estimation Bias and Variability	22
4.3.2 Model Assessment	23
Chapter 5: Applying Bias Reduction to the Fitted Stochastic Model	27
5.1 Double-smoothing	27
5.2 Data Sharpening	28
5.3 Results	28
5.3.1 Assessing the Estimators by Simulation	29
5.3.2 Median and Spread Comparisons	29
5.3.3 Bias and MSE Comparisons	32
Chapter 6: Conclusions and Future Work	36
Bibliography	38
Appendix	42
Appendix A: Figures	43

List of Tables

Table 3.1	Record of six statistics for the third micro-fire at a sequence of times (in the column headed by “elapsed time”), measured in seconds. Statistics include burning site counts, burnt out site counts and the nearest neighbour counts: N refers to the number of unburned fuel sites that are nearest neighbours of currently burning sites in the northerly direction; S refers to the corresponding number of unburned fuel sites to the south; W refers to the corresponding counts of western nearest neighbours; and E refers to the corresponding counts of eastern nearest neighbours.	18
Table 4.1	Parameter estimates from the smoothing/DE-based estimation scheme on 6 experimental data sets. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.	21
Table 4.2	Parameter estimates from data simulated from the stochastic spread model (Model B) using μ and λ as given in columns 2 and 4. Columns 3 and 5 give the medians of the estimates of μ and λ for 30 simulated data sets.	23
Table 5.1	Estimates of stochastic spread model parameters, μ and λ , based on data from each of 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the corresponding scale factors listed in Table 4.1.	29

LIST OF TABLES

Table 5.2	Performance of the smoothing/DE-based estimation scheme on 6 simulated data sets, using Double-smoothing. True values of μ and λ are listed together with the medians of the corresponding estimates from 30 simulated data sets generated according to the stochastic spread model rules. The true values of μ and λ were chosen to match the observed estimates for the 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.	30
Table 5.3	Performance of the smoothing/DE-based estimation scheme on 6 simulated data sets, using Data Sharpening. True values of μ and λ are listed together with the medians of the corresponding estimates from 30 simulated data sets generated according to the stochastic spread model rules. The true values of μ and λ were chosen to match the observed estimates for the 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.	30

List of Figures

Figure 3.1	A sequence of burn patterns on a sheet of wax paper observed at times: 1s, 6s, 11s, 16s, 21s, 26s, 31s, 36s, 41s, 46s, for the third micro-fire. Time increases from left to right, and then down.	16
Figure 3.2	Thresholded patterns for the third micro-fire. Unburned areas are coded as the light grey (green, if in colour), burning regions are at an intermediate grey shade (or red), and burnt regions are black.	16
Figure 4.1	Comparisons of parameter estimates between simulations and observed data for the third microfire. The side-by-side boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the estimated value of μ and the blue dashed horizontal line represents the estimate of λ , based on the observed data.	22
Figure 4.2	A sequence of images based on a simulation of the stochastic spread model (Model B) using $\mu = .510$ and $\lambda = .403$. Colour coding is as in Figure 3.2. . . .	24
Figure 4.3	A second simulation run using the same parameters as in Figure 4.2.	25

LIST OF FIGURES

Figure 4.4	Comparisons between data from 30 simulation runs and observed data from 6 micro-fires. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.	26
Figure 5.1	Boxplots for comparisons of parameter estimates between simulations and truth for the third simulation experiment. The boxplots represent samples of estimates based on 30 simulations. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ . Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	31
Figure 5.2	Bias for μ based on each technique used in each experiment.	32
Figure 5.3	Bias of λ based on each technique used in each experiment.	33
Figure 5.4	RMSE of μ based on each technique used in each experiment.	34
Figure 5.5	RMSE of λ based on each technique used in each experiment.	35
Figure A.1	Experiment 1: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.	43

LIST OF FIGURES

Figure A.2	Experiment 2: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.	44
Figure A.3	Experiment 4: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.	45
Figure A.4	Experiment 5: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.	46
Figure A.5	Experiment 6: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.	47

LIST OF FIGURES

Figure A.6	Boxplots for comparisons of parameter estimates between simulations and observed data for the first microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	48
Figure A.7	Boxplots for comparisons of parameter estimates between simulations and observed data for the second microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	49
Figure A.8	Boxplots for comparisons of parameter estimates between simulations and observed data for the fourth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	50
Figure A.9	Boxplots for comparisons of parameter estimates between simulations and observed data for the fifth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	51

LIST OF FIGURES

Figure A.10	Boxplots for comparisons of parameter estimates between simulations and observed data for the sixth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.	52
Figure A.11	Conventional local linear: comparisons between data from 30 simulation runs and observed data from the first micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.	53
Figure A.12	Conventional local linear: comparisons between data from 30 simulation runs and observed data from the second micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.	54

LIST OF FIGURES

- Figure A.13 Conventional local linear: comparisons between data from 30 simulation runs and observed data from the fourth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based. . . . 55
- Figure A.14 Conventional local linear: comparisons between data from 30 simulation runs and observed data from the fifth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based. . . . 56
- Figure A.15 Conventional local linear: comparisons between data from 30 simulation runs and observed data from the sixth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based. . . . 57

Acknowledgements

I would first like to express my gratitude to my supervisor Professor John Braun for the continuous support through out my M.Sc. program and related research. I really appreciate him taking me as his student. Thanks for giving me the opportunities to attend those inspiring workshops, to meet all professional and interesting people, and to widen my general knowledge. Thanks for all of his feedback to my courses, my research and my work. He definitely played a crucial role in making me the researcher I am today. Also, thanks for being supportive during the six-month NSERC program, and not giving up on me.

Besides my supervisor, I would like to thank my thesis committee members: Professor Shawn Wang, and Professor Jason Loepky for their insightful comments, which incented me to refine my thesis from various perspectives. Also, many thanks to Dr. Andrew Jirasek for letting me conduct the smouldering experiments in his lab. Many thanks to John Thompson for those informative discussions about the experiments.

I would also like to say a big thank you to Professor Tim Swartz and Professor Carl Schwarz. It was because of Tim's entry level Statistics lecture, I changed my major to Statistics. Thanks for the life-changing lecture, and thanks for introducing me to John. If it were not Tim, I would not find my interest in Statistics, and not even mention the opportunity of studying at UBCO. Many thanks to Carl, for teaching me SAS and R, for giving me the collaboration opportunity for the research work, and for opening my mind to applications of statistical knowledge.

I am also thankful to my friends at UBC Okanagan who I shared laughs with. I enjoyed the inspiring conversations with Alireza, Jimit and Mayang in the office, as well as our hearty potluck and those random dinners together. I will miss our time together, our laughs, and those mean jokes. I am thankful to Abby, Dongying, Jessica, Jiayu, Xukun and Zhen for those brainstorming board games, and the amazing Nakusp trip.

The last but not least, I am eternally grateful to my parents, my brother and my fiance for their unconditional love. Life has not been easy, but thank you for always being the solid anchor in my life.

Dedication

To my parents, my brother and my fiance

Chapter 1

Introduction

In this era of uncertainty due to climate change and other human impacts on the environment, it is as important as ever to quantify the risks associated with wildland fires. The risk of large catastrophic wildland fires appears to be increasing in many countries as evidenced by recent devastating wildfires in Europe, Australia and the United States, not to mention the 2016 Fort McMurray fire in Canada.

Risk due to wildfire has been studied in the past, for example, by Braun et al. [BJL⁺10], where the fire risk was studied in a localized area using the Burn-P3 fire growth simulator of Parisien et al [PKH⁺05]. This simulator is an attempt to incorporate uncertainty into the deterministic fire spread model, Prometheus, which was developed under the direction of Tymstra [Tym05]. Deterministic models such as Prometheus and its American counterpart, Farsite (Finney, [Fin04]), are the end result of large amounts of research into wildfire behaviour. However, as fires such as the one at Fort McMurray demonstrate, more work needs to be done. Modelling uncertainty more effectively is one aspect of this. The stochastic lattice spread model of Boychuk et al. [BBK⁺07] was proposed as a possible way to approach uncertainty in fire spread directly. Although the model has certain intuitive appeal, its use operationally is still a long way off, largely because it has not been validated on real fire data. The kind of data that would be ideal for this is not yet widely available, since direct observation of growing wildfires is not straightforward, even with remote sensing. On the other hand, we can take the viewpoint that biological and medical researchers often take when they cannot study human subjects directly; they consider animal models, realizing that not all observations and inferences made on these models are transferable to the human context, but rather, particular features and trends might be applicable.

Some typically measured factors in fire data are slope aspect, wind speed, wind direction, hourly precipitation, temperature, relative humidity, type of fuel, and so forth. However, there also exists some

sources of uncertainty. It might be the cases that the measured factors are not known at a very precise resolution or at the appropriate time or location; and there may be considerable heterogeneity in fuel structure, for instance, different sized trees, differences in understory, and so on.

In this thesis, then, we study a fire smouldering experiment that produces data that can be used to calibrate a stochastic fire growth model, and to quantify model performance. This small-scale experiment is our version of a “mouse model”. Not all characteristics of this model carry over to the wildfire context, but certain insights might be drawn from these kinds of experiments.

Videos of smouldering wax paper, treated with potassium nitrate, are studied and processed into data that can be analyzed using stochastic forest fire model. Each experiment will have controlled and uncontrolled conditions. The controlled conditions are elevation, wind speed, wind direction, and ignition location. An uncontrolled condition can be non-uniformly coating potassium nitrate on wax paper. The effects of the uncontrolled conditions are to be captured in the uncertainty of the stochastic model. After the movies of smouldering wax paper have been collected, an image processing algorithm will be used to convert the videos to snapshots, similar to remote sensing data obtained by satellites. The snapshots are converted into data images with three regions; unburnt fuel, burning fuel, and burnt out fuel.

The objectives of this thesis are twofold. The first goal is to study the Boychuk model from the point of view of data analysis, and show that the results from the experiments validate the stochastic models performance. Given data from a fire, is it possible to fit the model to the data, and is it possible to do model assessment? The second goal is to apply modern smoothing techniques to assist in the fitting of the stochastic model and to determine whether some previously proposed bias reduction strategies in nonparametric regression lead to improved parameter estimates in this setting. Chapter 2 will provide a literature review on fire behaviour and models, as well as the background of local polynomial regression that we employ when fitting the stochastic model under study. The experimental design and apparatus used are described in Chapter 3, along with the data extraction and image segmentation procedure. Chapter 4 will argue that it is possible to show that the parameters for the simplest special case of the stochastic model can be estimated from a sequence of pictures of a single fire.

Chapter 5 outlines the application of two bias reduction techniques, double-smoothing and data sharpening, to the fitting of the stochastic fire spread model. Results and conclusions will be provided in Chapter 6 with a discussion of some possible further work.

Chapter 2

Literature Review

2.1 Wildland Fire Behaviour and Models

Weber and Stocks [WS98] postulated that increasing temperatures could lead to increased numbers of wildland fire ignitions, a longer fire season and numbers of days with severe fire-weather. In some regions of Canada, fire seasons are getting longer (Albert-Green et al. [AGDMW12]) and fire risk has been shown to be increasing (Woolford et al., [WCDM10], [WDM⁺14]). Annual area burned has increased and has been connected to human-induced climate change (Gillett et al., [GWZF04]). Studies that analysed data output from climate model scenarios have suggested increased severity ratings [FvW], area burned [FLA⁺05], ignitions [WNF10] and a longer fire season [WF93]. Consequently, the development of accurate, spatially explicit fire spread models is of crucial importance for understanding aspects of fire behaviour and forecasting fire spread risk. Such models can be used at the incident level for individual fire management, or be coupled to fire occurrence and fire duration models in a simulation-based approach for longer-term strategic planning by wildland fire management agencies.

Deterministic fire spread models, such as PROMETHEUS [Tym05] and FARSITE [Fin04], play an important role in this arena. Although these simulators are well established and used frequently in Canada, the United States and in several other countries, their chief weakness is that they are not stochastic. Fire managers would benefit from probability maps to indicate where a currently burning fire may spread. Burn-P3 [PKH⁺05] is an ensemble-type simulation procedure which randomizes weather sequences in order to induce randomness, but this kind of procedure may be more appropriate for studying fire risk on large temporal and spatial scales; modelling the uncertainty associated with a single fire requires a different approach.

The paper by Garcia et al. [GBBT08] attempts to introduce stochasticity to the PROMETHEUS model via a block bootstrap procedure,

and Han and Braun [HB14] incorporate uncertainty through introducing an error component into the underlying model for rate of spread (ROS), as a parametric bootstrap. Much work remains to be done in order to make these procedures operational.

2.2 The Boychuk Fire Spread Models

A stochastic lattice-spread model of Boychuk et al. [BBK⁺07] was studied by Braun and Woolford [BW13]; in the latter paper, an interesting variant was introduced as well. As in [BW13], we refer to the original version of the model as Model A and the variant as Model B. According to the simplest special case of Model A, the landscape is assumed to be flat, the weather conditions are constant, and it assumed that there is no wind. The fuel type and density is also assumed to be homogeneous. On this landscape, a regular square $n \times m$ lattice is imposed. Each of the grid cells can be in one of three possible states:

- unburned fuel (**F**)
- burning fuel (**B**)
- burnt out (**O**).

Transitions between these states occur as follows: initially (i.e. at time $t = 0$), the grid cell at some location (i, j) is in state **B**, while all other cells are in state **F**; the fire burning in cell (i, j) will spread to each of its four nearest neighbours (i.e. north, south, east and west) in random amounts of time $T_{0,1}, T_{0,-1}, T_{1,0}$, and $T_{-1,0}$, provided it does not burn out first; specifically, at time $T_{0,1}$, the cell at $(i, j + 1)$ makes the transition from state **F** to **B**, if the cell is not already in state **B**. Similar transitions are made by cell $(i, j - 1)$ at $T_{0,-1}$, cell $(i + 1, 0)$ at $T_{1,0}$ and cell $(i - 1, 0)$ at $T_{-1,0}$. These times are assumed to be independent and exponentially distributed with mean $1/\lambda$. Once a cell has made a transition to state **B**, fire spreads from that cell to the sites of its nearest neighbourhood at a new set of independent exponential random times. Each burning cell waits an exponentially distributed amount of time until burning out, entering state **O**. The burn-out rate parameter is μ . Once in state **O**, a grid cell will make no further transitions.

Note that because of the minimum property of independent and identically distributed exponential random variables, when there are k

burning cells, the time until the first of these sites burns out is exponentially distributed with rate $k\mu$. This, together with the memoryless property of the exponential distribution, provides an equivalent way of specifying the burn-out rule for Model A: a cell is randomly selected from the set of k sites that are burning at the time of the last event and that cell makes the transition to O after the expiry of a simulated exponential $k\mu$ time period. Under Model B, a site is chosen to burn out at this same time, but not at a randomly selected site; instead, the site which has been burning longest is chosen to burn out. This may be a more realistic rule. This innovation had not been considered prior to [BW13]. The authors demonstrated that Model A has some deficiencies. In particular, its burn out rule can lead to simulations with large numbers of unburnt “islands”, a phenomenon that does not occur in the actual micro fire experiments.

2.3 Wildfire and Experimental Data

The data requirements for modelling wildland fire risk are massive, and the current available databases are limited in both the time and spatial dimensions for a thorough examination of the accuracy of stochastic fire spread models. The relevant weather data, including wind speed, wind direction, relative humidity, temperature and precipitation, are often only available at a sparse network of weather stations. Data available from remote sensing is limited in its accuracy and frequency. Topographical information is of somewhat limited accuracy as well. Satellite observations of evolving wildfires are occasionally available, but again, temporal frequency is low, and spatial resolution is limited.

In recent years, laboratory experiments have been conducted in order to study wildfire behaviour on a small-scale. A typical example of such a “burning-table” experiment is described in the paper of Martinez-de-Dios et al. [MdDAG⁺06] where the fire behaviour of different fuel types are studied under various scenarios. Such experiments are important because they provide physical models of wildfire where the same types of fuels are burned; of course, such models do not capture all aspects of a wildfire, such as spotting behaviour, and scaling the results is difficult. Larger scale experiments have been conducted in fire-proof wind tunnels and these, again, provide important insights

into fire behaviour, but there are, again, limitations. The paper by Zhang et al. [ZZAL92] provides another type of physical model, using ordinary bond paper, coated with potassium nitrate, to reduce flaming; the authors argue that the fibers of the paper relative to the size of the fire are roughly proportional to the stems of trees in an actual forest, providing more satisfactory scaling properties. [BW13] conducted some experiments using wax paper because it burned more cleanly than bond paper. The experimental data that [BW13] analyzed came from 31 small fires under identical conditions except the measured factor - slope. The results from previous experiments showed that smouldering wax paper, coated in potassium nitrate, exhibited similar patterns to what was observed in forest fires [BW13].

We adopt a similar physical modelling approach to that of [ZZAL92] and [BW13] in this thesis. In order to examine a numerical fire spread model on real data, a fire smouldering experiment, described in Chapter 3, will be configured to control variables important in estimation of fire growth rates and the implementation of the Boychuk lattice spread model.

2.4 Nonparametric Regression

Nonparametric or kernel regression represents a set of techniques for smoothing data. An important subset is local polynomial regression which is frequently used to estimate unknown regression functions. The underlying idea is premised on the use of a nonnegative symmetric density function $K(x)$ which is used to assign weights to observations upon applying least-squares polynomial regression. By assigning low weights to observations distant from the region of interest, one is able to obtain reasonable approximations to arbitrary functions using relatively low order polynomials. The bandwidth, often denoted by h , controls the amount of smoothing, through its effect on the kernel weights.

We are given n independent observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ from the model:

$$Y = g(X) + \sigma(X)\varepsilon, \tag{2.1}$$

where $g(\cdot)$ is an unknown function that has at least k derivatives, $\sigma(X)$ is a positive valued function of X , X and ε are independent, and ε has

mean 0 and variance 1.

The basic definition of local k th degree polynomial regression estimator of $g(x)$, which follows from a local least-squares argument, is:

$$\hat{g}_{LP}(x) = e_1^T \overbrace{\{X(x)^T W(x) X(x)\}^{-1} X(x)^T W(x) Y}^{\hat{\beta}(x)}, \quad (2.2)$$

where $e_1 = (1, 0, \dots, 0)^T$, $W = \text{diag}(W_i(x))$, $Y = (Y_1, \dots, Y_n)^T$,

$$W_i(x) = K\left(\frac{X_i - x}{h}\right) \quad (2.3)$$

and

$$X(x) = \begin{bmatrix} 1 & (X_1 - x) & \cdots & (X_1 - x)^k \\ \vdots & \vdots & \cdots & \vdots \\ 1 & (X_n - x) & \cdots & (X_n - x)^k \end{bmatrix}.$$

Since e_1 is the $(k+1)$ -vector with 1 in the first position and 0 elsewhere, it ensures that only β_0 is obtained from equation (2.2). Details of the argument giving rise to the least-squares estimator, based on Taylor Series are given below for the special case of $k = 1$. (See Section 2.4.2.)

Nadaraya-Watson (NW) regression ($k = 0$) is one of the popular kernel regressions as it is relatively easy to construct. Local linear regression ($k = 1$) is often used, as well, since it has better asymptotic properties, but higher order local polynomial regressions suffer from sparse data issues.

Kernel regression estimates of an unknown non-linear function are subject to bias. Therefore, methods have been developed to reduce bias while attempting to maintain the order of the asymptotic variance. This thesis will focus on “double-smoothing” and “data sharpening” as proposed in [HH09] and [CHR00], respectively.

Four estimators are considered in [HH09]. They include the local linear estimator, local cubic estimator, and the proposed double-smoothing local linear estimator as well as a previously proposed estimator of Choi and Hall [CH98]. We summarize this literature in the next few subsections. The methods for bias reduction will be introduced in detail in Section 2.5.

2.4.1 Nadaraya-Watson Estimator

Suppose independent vectors X_i and Y_i are from a multivariate distribution for $1 \leq i \leq n$. X_i 's are p -vectors and Y_i 's are scalars. The goal is to estimate $g(x) = E[Y|X = x]$, which is the expected value of Y given X is x . The Nadaraya-Watson estimator offers the simplest approach to nonparametric regression. It estimates $g(x)$ as a locally weighted average, using a kernel as a weighting function. Let $\hat{g}_{NW}(x)$ denote the conventional Nadaraya-Watson estimator:

$$\hat{g}_{NW}(x) = \frac{\sum_{i=1}^n Y_i W_i(x)}{\sum_{i=1}^n W_i(x)}, \quad (2.4)$$

where the weights $W_i(x)$ are defined at (2.3).

2.4.2 Local Linear Estimator

In contrast to the Nadaraya-Watson estimator, the local linear estimator preserves linear data. It fits locally a straight line instead of constants, given n independent observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ from the model:

$$Y = g(X) + \sigma(X)\varepsilon, \quad (2.5)$$

where $g(\cdot)$ is an unknown function that has at least one derivative, X and ε are independent, and ε has mean 0 and variance 1. The goal is to find $g(X)$ that minimizes:

$$\sum_{i=1}^n \{Y_i - E[Y_i|X_i]\}^2 = \sum_{i=1}^n \{Y_i - E[g(X_i) + \sigma(X_i)\varepsilon_i]\}^2 \quad (2.6)$$

$$= \sum_{i=1}^n \{Y_i - g(X_i)\}^2. \quad (2.7)$$

Without further assumptions, it's difficult to determine $g(X)$. However, Taylor's Theorem suggests that $g(X_i) \approx g(x) + g'(x)(X_i - x)$. By substituting the Taylor Expansions for $g(X_i)$ in (2.6) and including a symmetric kernel function $K(\cdot)$ ¹, (2.6) becomes a local least squares

¹In [HH09], the Epanechnikov kernel: $K(u) = \frac{3}{4}(1 - u^2)$, where $|u| \leq 1$ is used for all the estimators.

objective:

$$\sum_{i=1}^n \left[Y_i - (g(x) + g'(x)(X_i - x)) \right]^2 K\left(\frac{X_i - x}{h}\right). \quad (2.8)$$

The minimizer to (2.8) is the local linear estimator, which is obtained by minimizing

$$\sum_{i=1}^n \left[Y_i - (\beta_0 + \beta_1(X_i - x)) \right]^2 W_i(x), \quad (2.9)$$

where β_0 is the regression mean $g(x)$, as given in the definitions at (2.2) and (2.3), and β_1 is its first derivative $g'(x)$.

2.4.3 Local Cubic Estimator

In a very similar fashion, the local cubic estimator can be obtained by using the Taylor Expansions of $g(X_i)$ and minimizing the local least squares as below:

$$\sum_{i=1}^n \left[Y_i - \left(\beta_0 + \beta_1(X_i - x) + \frac{\beta_2(X_i - x)^2}{2!} + \frac{\beta_3(X_i - x)^3}{3!} \right) \right]^2 W_i(x), \quad (2.10)$$

where β_2 is the second derivative of $g(x)$ and β_3 is its third derivative.

Another related method that we mention in passing, due to Choi and Hall [CH98], is a weighted average of three local linear estimators. Since this method is not a focus of the thesis, it is not described further here.

2.5 Bias Reduction Strategies

2.5.1 Double-smoothing Local Linear Estimator

By generalizing [CH98], [HH09] came up with the idea of double-smoothing. Instead of three local linear estimators as in [CH98], double-smoothing weighted-averages all the local linear estimators. It is named double-smoothing because it involves two steps of smoothing. The first step of smoothing is to calculate the local linear estimator and its first derivative (i.e., β_0 and β_1). The second step is to obtain the weighted

average of the Taylor Expansions of the estimator. Its mathematical expression is as follows:

$$\hat{g}(t) = \int \left[\hat{\beta}_0(x) + \hat{\beta}_1(x)(t - x) \right] K\left(\frac{t - x}{h}\right) dx. \quad (2.11)$$

Double-smoothing is developed to estimate the regression mean with a bias reduction without changing the order of the asymptotic variance. The motivation behind this new estimator was to make use of information from both the intercept (β_0) and the slope (β_1), while LL considers only the intercept (β_0). In addition to this intuitive reason, LL has relatively large bias and LC is susceptible to sparse data issues when the sample size is small [HH09]. DS is, therefore, a compromise method between LL and LC as it reduces asymptotic bias and it is less likely to be affected by sparse data. [HH09] suggest that LC and DS are comparable as they all reduce the asymptotic bias by two orders of magnitude to h^4 from h^2 , but it's hard to determine which one is a better estimator than the others since the orders of variance are the same [HH09]. That is why He and Huang compared the relative variability for the estimators in their paper using simulation.

2.5.2 Data Sharpening

Another important procedure is data sharpening. In [CHR00], Choi, Hall and Rousson introduced some techniques to move the data in such a way that the fume hood methods have reduced bias. Although a wide selection of conventional estimators are available, Choi, Hall and Rousson considered two estimators in detail in their paper: NW and LL. Let $\hat{g}_{LL}(x)$ denote the conventional local linear estimator.

[CHR00] introduced three versions of data sharpening for Nadaraya-Watson estimator. The first version involves moving the design variables (X_i 's) so that they become more concentrated in places where the design density is high and more sparse in places where it is relatively low [CHR00]. Let $\hat{g}_{NW,1}(x)$ denote the first data sharpening version of \hat{g}_{NW} . By replacing the design points X_i 's with $\hat{X}_j = \frac{\sum_i X_i W_i(X_j)}{\sum_i W_i(X_j)}$, we can obtain

$$\hat{g}_{NW,1}(x) = \frac{\sum_i Y_i K\left(\frac{\hat{X}_i - x}{h}\right)}{\sum_i K\left(\frac{\hat{X}_i - x}{h}\right)}.$$

The second version adjusts the response variables (Y_i 's) and keeps the explanatory variables unchanged. Set $\hat{Y}_j = \hat{g}_{NW}(X_j)$, and $\tilde{Y}_j = 2Y_j - \hat{Y}_j = Y_j + e_j$, where e_j are the residuals. We replace Y_i with their sharpened form \tilde{Y}_j . That is, the second data sharpening version of \hat{g}_{NW} is

$$\hat{g}_{NW,2}(x) = \frac{\sum_i \tilde{Y}_i W_i(x)}{\sum_i W_i(x)}.$$

The third version alters both design and response variables. We take the sharpened \hat{X}_j from version one and put $\check{Y}_j = 2Y_j - \hat{g}_{LL}(X_j)$ as the new sharpened response points. Therefore,

$$\hat{g}_{NW,3}(x) = \frac{\sum_i \check{Y}_i K\left(\frac{\hat{X}_i - x}{h}\right)}{\sum_i K\left(\frac{\hat{X}_i - x}{h}\right)}.$$

On the other hand, there are two versions of data sharpening for the local linear estimator. Version one is to move both variables. Set $\tilde{X}_j = 2X_j - \hat{X}_j$. Each (X, Y) is replaced by (\tilde{X}, \tilde{Y}) , as defined previously, and

$$\hat{g}_{LL,1}(x) = e_1^T \{\tilde{X}(x)^T \tilde{W}(x) \tilde{X}(x)\}^{-1} \tilde{X}(x)^T \tilde{W}(x) \tilde{Y}.$$

However, in contrast to the Nadaraya-Watson estimator, this time we slightly shift the design points so that they are a little more closer together in places where the design density is relatively low and a little further apart where the concentration is high [CHR00]. This is an attractive feature when dealing with sparse data.

The second data sharpened local linear estimator is similar to $\hat{g}_{NW,2}(x)$. By replacing Y by \check{Y} , which is previously defined, we obtain

$$\hat{g}_{LL,2}(x) = e_1^T \{X(x)^T W(x) X(x)\}^{-1} X(x)^T W(x) \check{Y}.$$

Correspondingly, $\hat{g}_{LL,1}(x)$ and $\hat{g}_{LL,2}(x)$ have the definition of \hat{g}_{LL} .

To conclude this chapter, both double-smoothing and data sharpening are ways to reduce bias while maintaining a similar level of variability. There is always a trade-off between bias and variance.

2.5.3 Choice of Bandwidth

The performance of local polynomial regression estimates depends critically on the bandwidth h , and much has been written on the subject of bandwidth selection. A standard reference is [WJ95]. Among the methods that have been proposed are leave-one-out cross-validation and so-called direct plug-in methods.

The method that we will use is the plug-in method, available in coded form in the R package *KernSmooth* as `dpill`. This function attempts to balance the asymptotic bias with the asymptotic variance in an optimal way for local linear regression. It is not optimized for the data sharpening or the double-smoothing applications, but it provides a reasonable balance between speed of computation and accuracy.

2.5.4 Boundary Correction

Another problem associated with nonparametric regression, particularly with local constant regression, is boundary bias. This is due to the use of symmetric kernels in the determination of the regression weights.

Without some form of adjustment, estimates of the regression function at the boundaries can be quite distorted, due to a lack of data beyond those boundaries. Double-smoothing, in particular, is susceptible to such boundary effects.

In this thesis, “mirroring” is employed as a boundary adjustment. It is a simple technique that reflects the data points through the boundary, on both sides. The mirroring approach mitigates the problems seen in double-smoothing.

A number of other approaches have been suggested to handle such boundary issues, including asymmetric boundary kernels but these lie outside the scope of this thesis.

Chapter 3

Smouldering Experiments

In this chapter, we describe the experimental set-up and design used to collect video data for small micro-fires. In this case, the fires were prevented from open flaming, for safety reasons. A number of different factors were considered in the actual experiments, but our subsequent analysis will be confined to the simplest case – where the fire burns on a flat horizontal surface. The other cases were recorded for a larger study to be carried out by other members of the lab.

3.1 Apparatus

The apparatus used for the smouldering experiment is one sided wax paper, a laboratory scale to weigh the potassium nitrate, a walled tray to soak wax paper, a flat tray to smolder the wax paper, a hot plate to dry the wax paper, a metal breadboard (as the base) and a few metal pegs to hold and fix the flat tray, a fume hood for a steady and safe experimental environment, a camera to film the experiment, and an ignition device. The one-sided wax paper is chosen so that it does not absorb potassium nitrate, and it is still easy to ignite. The potassium nitrate-to-water ratio in the solution is 1:10. The dimensions of the flat tray are approximately $38.74 \times 26.04 \times 1.91$ cm. It is painted black (by non-flammable spray) to minimize reflected light. There are holes drilled on the pan to allow for the attachment to the breadboard, as well as for ignition from below. The ignition device is a barbecue lighter with a bendable head. Metal pegs are screwed into the base, and the camera is attached at a fixed position, 48cm high from the base so that it can record the experiments from the top. The camera model is the Olympus Stylus 600, manufactured in 2010.

The experiments are conducted to investigate the impact of environmental effects, such as elevation and wind velocity, to fire spread rate. There are four groups of experiments: “flat”, “sloped”, “wind-flat”, and “wind-sloped”. There are three steps involved in the experimen-

3.1. Apparatus

tal design. In the first step, wax paper is soaked in potassium nitrate solution for an hour, with 30 minutes for each side, in the walled tray. Then it is dried on a hot plate with the target temperature at 60°C . The soaking and drying steps are identical for all the experiments. The last step is to burn the dried wax paper. During the smouldering, wax paper is fixed on the flat tray by tape with some space in between to allow enough oxygen for burning. Four pegs of the same length are used to hold the corners of the pan for “flat” experiments. The wax paper is ignited at the center from below by the bendable lighter. For the “sloped” experiments, different lengths of pegs are used to create the angles. Using pegs, two slope angles are accomplished for different slopes, 11° and 32° . The ignition spot is also at the center. Different slopes allow us to investigate the effect of elevation on fire rate of spread. “Wind-flat” and “wind-sloped” experiments are same to “flat” and “sloped” experiments, respectively, except including the effect of wind velocity by using a fan. The camera attached from the top starts filming the experiments before the ignition and stops when the smouldering is out of scope or when the boundary of the paper has been reached.

It is worth mentioning that all the “flat” experiments are conducted in the dark. This gives easier analyses than the crude estimate of fire rate of spread taken by [BW13]. By conducting the experiments in the dark, problems with segmenting the images are avoided since the visible light camera detects only the light from the fire. Thus, only black and non-black are registered for each movie frame. Tracking the colour of each pixel through time results in a sequence of blacks at or near zero, followed by a clearly nonzero value at the time the pixel burns. A return to zero occurs when the pixel is burnt out. Segmenting follows easily because each pixel is classified as “fuel” (green), followed by “burning” (red), and then “burnt out” (black). No further processing is needed. Once the data collection step is done, it is necessary to smooth the data in terms of both estimating the regression function and its derivative for bias reduction. The nonparametric methods discussed in Chapter 2 play a role here.

3.2 Data Extraction and Segmentation

The open source program *ffmpeg* [dev13] was used to freeze-frame each movie at approximately half-second intervals to obtain clear image-captures with time-stamps. These captured images were then saved as JPEG files, readable into the R system [Tea16]. For illustrative purposes, the images for the third experimental run are shown in Figure 3.1. The original images were a combination of several colours: yellow, red, grey, black, etc. In order to convert the image matrices into a usable form, conversion to three colours red (burning), green (fuel), black (burnt out) was necessary.

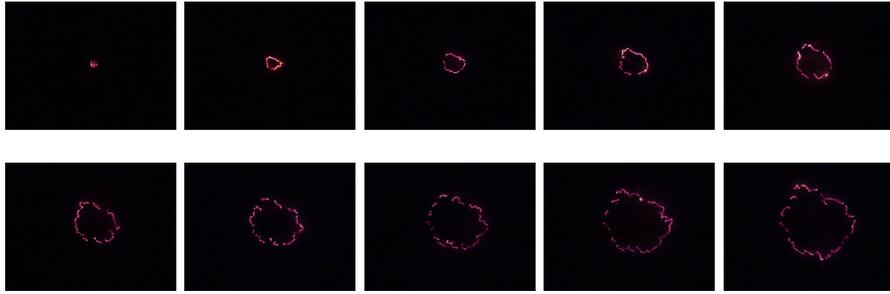


Figure 3.1: A sequence of burn patterns on a sheet of wax paper observed at times: 1s, 6s, 11s, 16s, 21s, 26s, 31s, 36s, 41s, 46s, for the third micro-fire. Time increases from left to right, and then down.

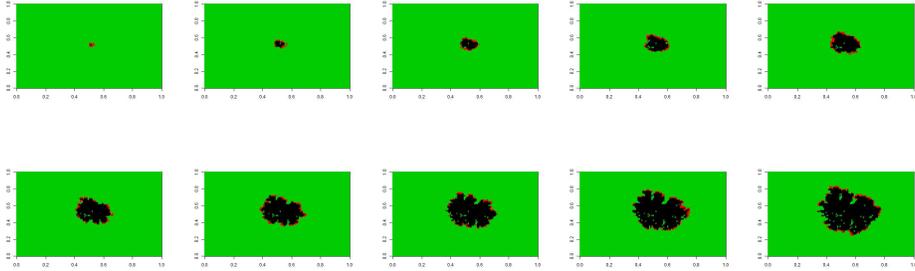


Figure 3.2: Thresholded patterns for the third micro-fire. Unburned areas are coded as the light grey (green, if in colour), burning regions are at an intermediate grey shade (or red), and burnt regions are black.

3.2. Data Extraction and Segmentation

This image segmentation problem was greatly simplified because of the use of darkness in the experimental setup. Each grid cell is unburned until it burns (and is lit up in the video footage) and is burnt out for all remaining time. Therefore, the time(s) at which each grid cell burns can be identified from the time series of the red, green and blue (rgb) measurements. The series corresponding to red is most useful for this purpose; prior to burning, the values are at or near 0; thus all images corresponding to these times can be set to “green”; after burning, the values are again at or near 0; these can be set to “black”. The values at the time(s) of burning can be set to “red”. The resulting patterns, corresponding to the images in Figure 3.1 are displayed in Figure 3.2.

Finally, the colour-coded images were converted to a numeric matrix corresponding to the green, red, and black pixels of the image array. We assigned the colour green to the value 0, red to the value 1, and black to the value 2. The numbers of 0, 1, and 2’s were counted which gave the numbers of unburned fuel sites, the number of burning and the number of burnt out sites. In addition, the number of fuel sites neighbouring burning sites at each time point were also counted, at half-second intervals for the entire duration of each fire.

For illustration purpose, the counts for each of these statistics are listed in Table 3.1 for the third micro-fire, corresponding to the first 30 seconds of the fire. The data for the other fires were collected in a similar manner, but not recorded here in tabular form, in order to save space, but graphical summaries of the 6 statistics for each of the 6 fires are displayed as heavy black curves in the figures contained in the Appendix.

3.2. Data Extraction and Segmentation

Table 3.1: Record of six statistics for the third micro-fire at a sequence of times (in the column headed by “elapsed time”), measured in seconds. Statistics include burning site counts, burnt out site counts and the nearest neighbour counts: N refers to the number of unburned fuel sites that are nearest neighbours of currently burning sites in the northerly direction; S refers to the corresponding number of unburned fuel sites to the south; W refers to the corresponding counts of western nearest neighbours; and E refers to the corresponding counts of eastern nearest neighbours.

elapsed time	number burning	number burnt out	N	S	W	E
0	1	0	1	1	1	1
1	11	1	5	4	4	3
2	7	9	5	4	2	4
3	13	13	6	7	6	6
4	30	17	8	9	7	7
5	15	31	5	10	6	6
6	21	39	6	10	9	9
7	33	46	10	13	10	11
8	30	67	12	14	9	11
9	34	77	11	14	7	12
10	33	97	13	15	9	13
11	29	115	12	14	9	11
12	34	131	17	17	9	13
13	39	145	17	17	13	15
14	45	157	16	19	13	17
15	55	176	18	22	16	21
16	56	201	16	21	16	22
17	56	225	15	21	17	23
18	65	249	23	23	19	22
19	59	284	17	24	18	23
20	57	313	18	22	18	23
21	65	349	19	23	26	23
22	62	385	20	25	24	23
23	63	417	23	26	23	22
24	70	444	21	25	28	26
25	79	470	24	27	28	27
26	72	518	28	29	30	25
27	92	553	30	33	31	30
28	81	598	31	34	32	33
29	89	632	32	39	31	35
30	83	682	27	33	30	36

Chapter 4

Fitting a Stochastic Fire Growth Model to Data

This chapter proceeds as follows. In Section 4.1, we describe a method to estimate the two parameters of the basic interacting particle model using data on numbers of burning grid sites and numbers of neighbouring unburned sites. The use of the nonparametric regression estimates in a continuous least-squares framework is described in Section 4.2. Section 4.3 summarizes the results from the experiments, and then provides specific tools to assess the fit of the model to the data. We conclude the chapter with our observations and our ideas about future work on related problems.

4.1 Model Fitting

The methodology required was developed [BW13] to fit the grid-based fire spread model to the data extracted from a sequence of images of a growing fire. Referring to data such as in Table 3.1, let $X(t)$ denote the number of burning sites at time t , and $Y(t)$ denote the number of sites that are burnt out by time t . In addition, let $B_N(t)$, $B_S(t)$, $B_W(t)$ and $B_E(t)$ denote the number of burning sites with unburnt fuel in the site immediately above (north), below (south), on their left (west) or right (east), respectively. Also, let

$$\begin{aligned}x(t) &= E[X(t)], \\y(t) &= E[Y(t)], \\b_N(t) &= E[B_N(t)], \\b_S(t) &= E[B_S(t)], \\b_W(t) &= E[B_W(t)],\end{aligned}$$

and

$$b_E(t) = E[B_E(t)].$$

According to the model rules and arguing as in Braun and Kulperger [BK93], we have the relations

$$y'(t) = \mu x(t), \tag{4.1}$$

$$x'(t) = -\mu x(t) + \lambda (b_N(t) + b_S(t) + b_W(t) + b_E(t)), \tag{4.2}$$

and initial conditions of $y(0) = 0$ and $x(0) = 1$.

These equations hold for both the original Boychuk model, Model A, as well as the variant, Model B. This result seems surprising, since Model B is non-Markovian, while Model A is. It is important to note that the differential equations are for population level quantities. They are also only a partial description of the process dynamics. However, they contain enough process information to allow for construction of moment estimators for the process parameters, which we now demonstrate.

The notation in the ensuing discussion can be simplified by making the substitution

$$g(t) = b_N(t) + b_S(t) + b_W(t) + b_E(t).$$

Then

$$x'(t) = -\mu x(t) + \lambda g(t). \tag{4.3}$$

By adding (4.1) and (4.3), we obtain:

$$x'(t) + y'(t) = \lambda g(t). \tag{4.4}$$

4.2 Moment and Continuous Least-Squares Estimation

We can estimate the functions $x(t)$ and $y(t)$ using $X(t)$ and $Y(t)$, respectively. In fact, improved estimates of these functions can be obtained by applying a local linear kernel smoother to the $(t, X(t))$ data, and $(t, Y(t))$ data respectively. Similarly, estimates of $g(t)$ can be improved by a local constant smoother. We used the `locpoly` function in the *KernSmooth* package [Wan15] with the automatically selected smoothing parameter. The software also allows for estimation of the

4.3. Results

first derivatives $x'(t)$ and $y'(t)$ from the same data sets (see [WJ95], for example).

Continuous least-squares can then be applied to the estimate of equation (4.1) to estimate μ . That is,

$$\hat{\mu} = \arg \min_{\mu} \int_{t_1}^{t_n} (\hat{y}'(t) - \mu \hat{x}(t))^2 dt \quad (4.5)$$

where $\hat{y}'(t)$ and $\hat{x}(t)$ are the smoothed estimates described in the preceding paragraph. An estimator for λ can be obtained from the estimated version of (4.4) similarly:

$$\hat{\lambda} = \arg \min_{\lambda} \int_{t_1}^{t_n} (\hat{x}'(t) + \hat{y}'(t) - \lambda \hat{g}(t))^2 dt. \quad (4.6)$$

4.3 Results

As noted earlier, the sequence of fire images are displayed in Figure 3.1 for one of the micro-fires. The corresponding thresholded images are displayed in Figure 3.2. Based on these images, counts of the various statistics (discussed in Section 4.1) were taken, at a number of different grid resolutions (scale factors). Table 3.1 displays the counts at one of the resolutions, and the third row of Table 4.1, that is Replicate 3, contains the resulting parameter estimates and the amount of scaling done to increase the pixel size.

Table 4.1: Parameter estimates from the smoothing/DE-based estimation scheme on 6 experimental data sets. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.

Replicate	scale factor	$\hat{\mu}$	$\hat{\lambda}$
1	2.9	0.339	0.301
2	2.6	0.481	0.391
3	2.8	0.490	0.383
4	2.4	0.347	0.287
5	2.1	0.409	0.365
6	2.2	0.445	0.376

The remaining rows of Table 4.1 contain the parameter estimates and scalings required for the other 5 micro-fires. We see that the fires

4.3. Results

required different amounts of scaling, and the estimates of μ range from 0.34 per second to 0.49 per second. The estimates of λ range from 0.29 to 0.39. In all cases, the burn out parameter exceeds the spread rate parameter.

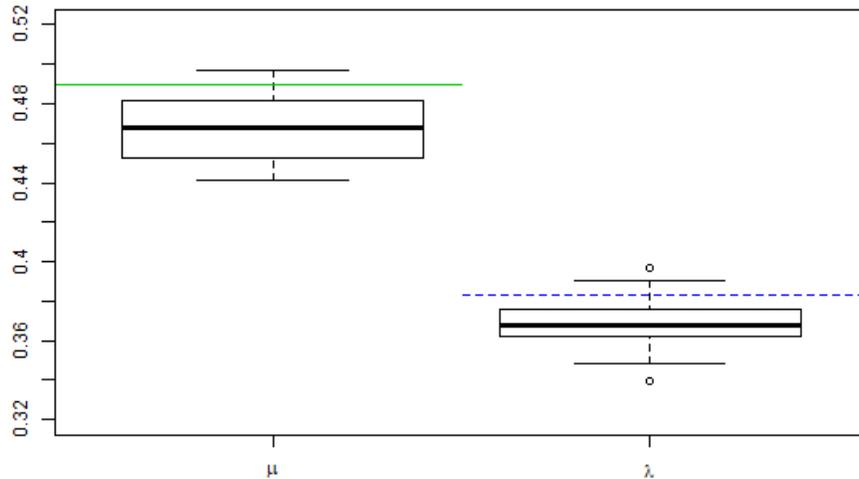


Figure 4.1: Comparisons of parameter estimates between simulations and observed data for the third microfire. The side-by-side boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the estimated value of μ and the blue dashed horizontal line represents the estimate of λ , based on the observed data.

4.3.1 Parameter Estimation Bias and Variability

By simulating from the fitted model and re-estimating the parameters, it is possible to assess bias and variability of the parameter estimates. All simulations are done using Model B. We have chosen to display the results of this assessment for the third micro-fire in Figure 4.1 using side-by-side boxplots to graphically summarize the distribution of the estimates of μ and λ for the simulated data sets. Horizon-

4.3. Results

tal lines have been drawn to indicate the locations of μ (solid) and λ (dashed). What is immediately evident from the graphs is that both μ and λ appear to be underestimated in the simulations: the medians of the estimates from the simulated data are both about .02 less than the values used to produce the simulated data.

The side-by-side boxplots also show that the distributions of the parameter estimates are approximately symmetric, and the amount of variability is in the same order as the bias. That is, if these simulations were viewed as a parametric bootstrap, the standard errors (SE) for the two estimators are approximately .01. Thus, the bias dominates the MSE ($\text{MSE} = \text{bias}^2 + \text{SE}^2$), and the root-MSE (RMSE) is approximately .022 for both estimators.

Boxplots for the other five fires (shown in the Appendix) are quite similar to what appears in Figure 4.1. All six fires exhibit the same degree and direction of bias. Table 4.1 contains information on the medians for the estimates of λ and μ from simulations of each of the six fitted models. In all cases, the medians of the parameter estimates from the simulated data are slightly below the observed parameter values used in the simulations.

Table 4.2: Parameter estimates from data simulated from the stochastic spread model (Model B) using μ and λ as given in columns 2 and 4. Columns 3 and 5 give the medians of the estimates of μ and λ for 30 simulated data sets.

Experiment	μ	median $\hat{\mu}_{\text{sim}}$	λ	median $\hat{\lambda}_{\text{sim}}$
1	0.359	0.333	0.321	0.301
2	0.501	0.474	0.411	0.394
3	0.510	0.492	0.403	0.395
4	0.367	0.343	0.307	0.288
5	0.429	0.398	0.385	0.360
6	0.465	0.439	0.396	0.376

4.3.2 Model Assessment

To assess the adequacy of the grid-based fire spread model for this particular data set, we again simulated realizations of fire spread at the estimated parameter values, using the chosen grid resolution. For all simulations, we used the variant Model B.

4.3. Results

Since a bias of .02 has been observed in both parameter estimates, we add this quantity to the observed estimates of both μ and λ and use the resulting values as the “truth” in our simulation validation study.

Table 4.2 lists the true parameter values together with the medians of estimates coming from 6 simulation experiments, each using a different pair of values for μ and λ . Each experiment involves 30 simulation runs. What is clearly evident from the table is the consistent, but slight, underestimation of each parameter. Further examination is required before we can make firm conclusions about the model and the fitting technique.

The results from one simulation run are displayed in the sequence of pictures in Figure 4.2. The images have been plotted at approximately 5 second intervals. We see from these pictures that the fire size and the thickness of the fire perimeter are similar to the analogous quantities for the thresholded images at corresponding times. The boundary is somewhat less smooth in the simulation pictures than in the actual fire, but the overall shapes are fairly similar. The simulated fire appears to have grown somewhat larger than the observed fire. However, Figure 4.3 shows the results of another simulation run where similar qualitative behaviour is evident but where the fire sizes tend to be somewhat smaller than the observed.

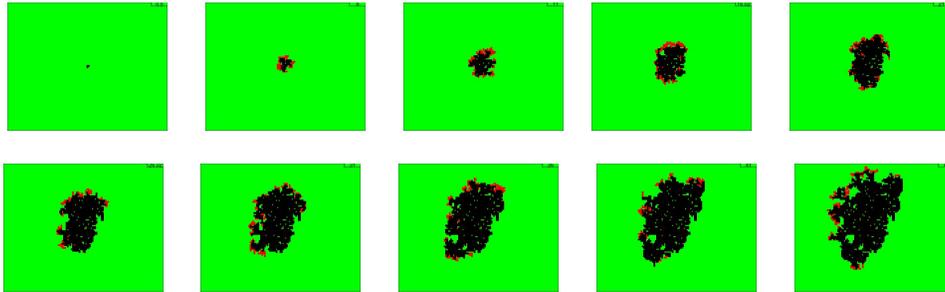


Figure 4.2: A sequence of images based on a simulation of the stochastic spread model (Model B) using $\mu = .510$ and $\lambda = .403$. Colour coding is as in Figure 3.2.

Such plots are limited in their usefulness. In this case, we can see that there are differences between the simulated pictures and the actual data, but it is difficult to tell if these differences are due to the variability we are trying to model, or if these are failures of the model

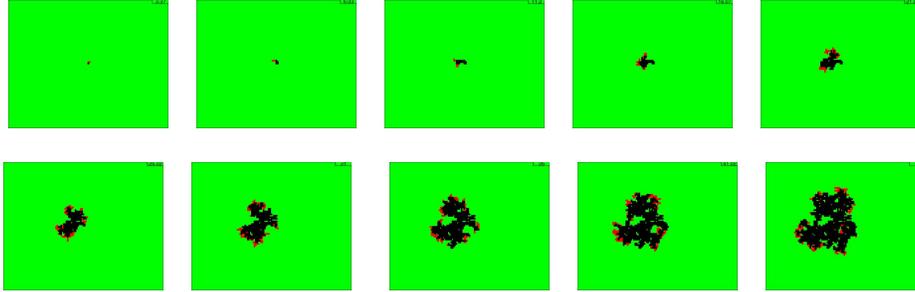


Figure 4.3: A second simulation run using the same parameters as in Figure 4.2.

itself.

As another check on the appropriateness of the model, we can compare the burning cell counts, burnt out cell counts and nearest neighbour statistics for the original data with simulated data from the fitted model (using data from the third micro-fire only). Figure 4.4 shows the results of 30 simulated realizations (plotted in grey) with the observed counts for all other fires (plotted in blue) against time; the observed data for the third micro-fire is plotted in black.

What is evident from this set of plots is that the range and distribution of simulated counts of burning sites (based on the observed data from one fire) matches the observed range and distribution of burning sites for other fires very well. Except for a location-shift in the distribution of simulated burnt out sites and neighbourhood statistics, we also see similarities in the range and distribution in these cases. The location shift is likely due to the estimation bias discussed earlier. Note that by increasing both μ and λ slightly, we will not see much change in the total number of burning sites over time, but we will see many more burnt out sites, for example. It is also noteworthy that for one of the actual fires, there were instances of flaming which distorted the camera images at two time points, reflecting in the anomalous spikes in Figure 4.4.

These observations provide strong evidence that the model is doing well at capturing distributional behaviour in these dimensions. The corresponding count graphs for the other five micro-fires are attached in the Appendix.

4.3. Results

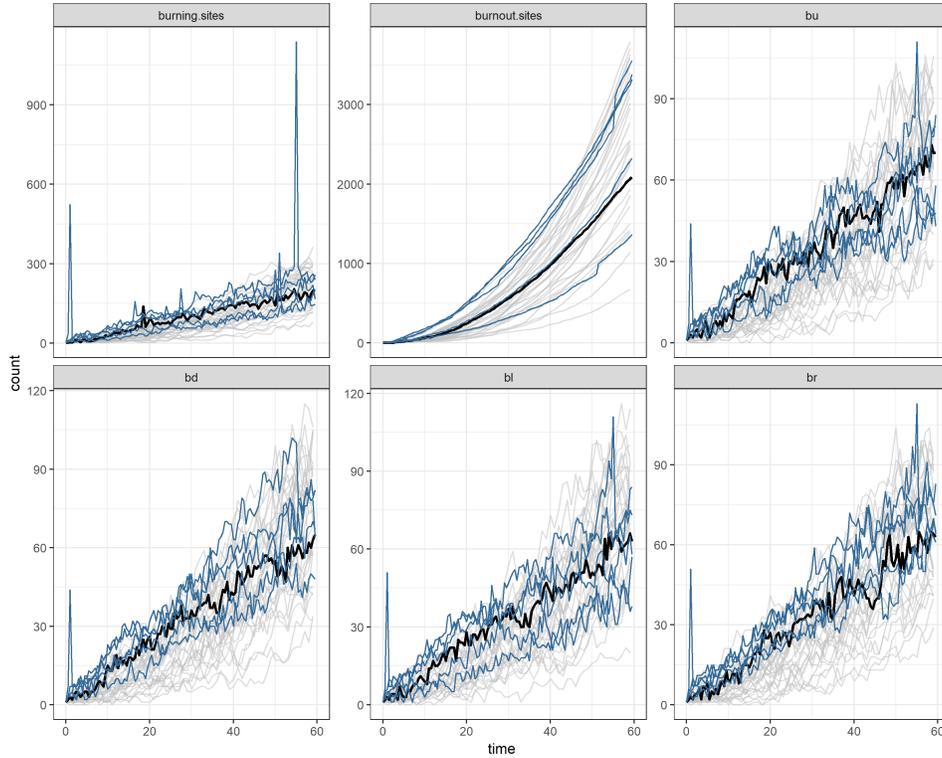


Figure 4.4: Comparisons between data from 30 simulation runs and observed data from 6 micro-fires. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.

Chapter 5

Applying Bias Reduction to the Fitted Stochastic Model

In the previous chapter, we applied kernel regression techniques to the problem of estimating the mean counts of burning and burnt out sites, as well as the derivatives of these functions. The resulting functions were used to obtain estimates of the parameters of a stochastic fire spread model. It is of interest to see what, if any, effect bias reduction techniques would have on the estimates of these parameters. To this end, we consider the double-smoothing local linear estimator and data sharpening, applying them to the estimation of the mean counts and subsequently applying least-squares, as in (4.5) and (4.6) to obtain estimates of μ and λ . Application to derivative estimation is outside the scope of this thesis.

5.1 Double-smoothing

We apply the double-smoothing local linear (DSLL) regression estimator with mirroring to the $(t, X(t))$ data to obtain $\hat{x}_{DSLL}(t)$. Similarly, $\hat{g}_{DSLL}(t)$ is obtained from the time series of nearest-neighbour statistics.

Continuous least-squares is then applied to the estimating equation (4.1) to estimate μ . That is,

$$\hat{\mu}_{DSLL} = \arg \min_{\mu} \int_{t_1}^{t_n} (\hat{y}'(t) - \mu \hat{x}_{DSLL}(t))^2 dt \quad (5.1)$$

where $\hat{y}'(t)$ remains as described in Chapter 4. An estimator for λ can be obtained from the estimated version of (4.4) similarly:

$$\hat{\lambda}_{DSLL} = \arg \min_{\lambda} \int_{t_1}^{t_n} (\hat{x}'(t) + \hat{y}'(t) - \lambda \hat{g}_{DSLL}(t))^2 dt. \quad (5.2)$$

5.2 Data Sharpening

In a similar fashion, we want to apply data sharpening (DS) to solve 4.5 and 4.6 instead of conventional local linear estimator.

Again, only $g(t)$ and $x(t)$ are sharpened and smoothed. The first version of data sharpening for local linear estimator is used (see the definition of $\hat{g}_{LL,1}(x)$ in Section 2.5.2). The number of burning sites ($x(t)$) and the sum of the burning sites in neighbourhood ($g(t)$) are first smoothed using the Nadaraya-Watson estimator, individually. These estimated counts are then “sharpened” by adding the residuals to the observed data. The next step is to apply local linear regression to obtain the estimated expected counts based on an equally spaced gridsize. Although the boundary effect in this case does not have a large impact, “mirroring” is also applied for data sharpening to ensure that the results can be compared across all methods. The resulting estimates are referred to as $\hat{x}_{DS}(t)$ and $\hat{g}_{DS}(t)$

Continuous least-squares is then applied to the estimating equation (4.1) to estimate μ . That is,

$$\hat{\mu}_{DS} = \arg \min_{\mu} \int_{t_1}^{t_n} (\hat{y}'(t) - \mu \hat{x}_{DS}(t))^2 dt \quad (5.3)$$

where $\hat{y}'(t)$ remains as described in Chapter 4. An estimator for λ can be obtained from the estimated version of (4.4) similarly:

$$\hat{\lambda}_{DS} = \arg \min_{\lambda} \int_{t_1}^{t_n} (\hat{x}'(t) + \hat{y}'(t) - \lambda \hat{g}_{DS}(t))^2 dt. \quad (5.4)$$

5.3 Results

Table 5.1 provides the estimates of the parameters for the six micro fires, based on the three approaches: conventional local linear, double-smoothing local linear and data sharpening.

According to the table, the estimates tend to be quite similar, but without a notion of standard error, it is not possible to make firm conclusions.

5.3. Results

Table 5.1: Estimates of stochastic spread model parameters, μ and λ , based on data from each of 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the corresponding scale factors listed in Table 4.1.

Replicate	$\hat{\mu}_{LL}$	$\hat{\mu}_{DSLL}$	$\hat{\mu}_{DS}$	$\hat{\lambda}_{LL}$	$\hat{\lambda}_{DSLL}$	$\hat{\lambda}_{DS}$
1	.3387	.3419	.3420	.3010	.3016	.2974
2	.4808	.4807	.4806	.3911	.3908	.3902
3	.4897	.5008	.5012	.3833	.3855	.3836
4	.3467	.3481	.3479	.2872	.2874	.2849
5	.4090	.4092	.4086	.3650	.3648	.3635
6	.4452	.4482	.4480	.3757	.3769,	.3715

5.3.1 Assessing the Estimators by Simulation

In Section 4.3, we learned that the observed μ and λ are both underestimated by .02 when compared with estimates based on data simulated from the fitted model. In order to obtain the bias and MSE of μ_{DSLL} , μ_{ds} , λ_{DSLL} , and λ_{ds} we use the true values listed in Table 4.2 to conduct an additional simulation check. Again, 30 simulation runs are conducted at each parameter setting.

For each set of simulated data, we obtain new estimates $\hat{\mu}_{DSLL}$, $\hat{\mu}_{DS}$, $\hat{\lambda}_{DSLL}$ and $\hat{\lambda}_{DS}$ using the double-smoothing local linear and data sharpening based estimators.

5.3.2 Median and Spread Comparisons

Tables 5.2 and 5.3 contain simple summaries of the results from the 6 simulation experiments, for the double-smoothing local linear estimates and the data sharpening estimates, respectively. For each experiment, the median of the estimates of μ is recorded as is the median of the estimates of λ .

What we see in the tables is that when estimating μ , there is not a lot of difference among the estimators in terms of median performance, though the data sharpening estimator might be very slightly better than the others.

In terms of λ , the median of data sharpening estimates is clearly closer to the true value than either of the other estimates, for all 6 experiments.

5.3. Results

Table 5.2: Performance of the smoothing/DE-based estimation scheme on 6 simulated data sets, using Double-smoothing. True values of μ and λ are listed together with the medians of the corresponding estimates from 30 simulated data sets generated according to the stochastic spread model rules. The true values of μ and λ were chosen to match the observed estimates for the 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.

Experiment	scale factor	μ	median $\widehat{\mu}_{DSLL.sim}$	λ	median $\widehat{\lambda}_{DSLL.sim}$
1	2.9	0.359	0.336	0.321	0.303
2	2.6	0.501	0.476	0.411	0.394
3	2.8	0.510	0.487	0.403	0.392
4	2.4	0.367	0.346	0.307	0.290
5	2.1	0.429	0.400	0.385	0.362
6	2.2	0.465	0.439	0.396	0.377

Table 5.3: Performance of the smoothing/DE-based estimation scheme on 6 simulated data sets, using Data Sharpening. True values of μ and λ are listed together with the medians of the corresponding estimates from 30 simulated data sets generated according to the stochastic spread model rules. The true values of μ and λ were chosen to match the observed estimates for the 6 micro-fires. The grid cell sizes used in the spread model were taken as the camera pixel size multiplied by the given scale factor.

Experiment	scale factor	μ	median $\widehat{\mu}_{DS.sim}$	λ	median $\widehat{\lambda}_{DS.sim}$
1	2.9	0.359	0.336	0.321	0.309
2	2.6	0.501	0.476	0.411	0.398
3	2.8	0.510	0.487	0.403	0.392
4	2.4	0.367	0.346	0.307	0.294
5	2.1	0.429	0.398	0.385	0.360
6	2.2	0.465	0.439	0.396	0.381

5.3. Results

To help visualize the distributions of the parameter estimates, three side by side boxplots are provided in Figure 5.1. The boxplots represent samples of estimates based on 30 simulations for the third simulation experiment. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ . Techniques used from left to right are conventional local linear, double-smoothing, and data sharpening. We can immediately see from the graph that the spread of the interquartile range and the range is decreasing from left to right, which indicates a decreasing variance. Boxplots of simulated μ and λ for the other five experiments (shown in the Appendix) are quite similar. Most fires exhibit the same degree and direction of median bias and spread.

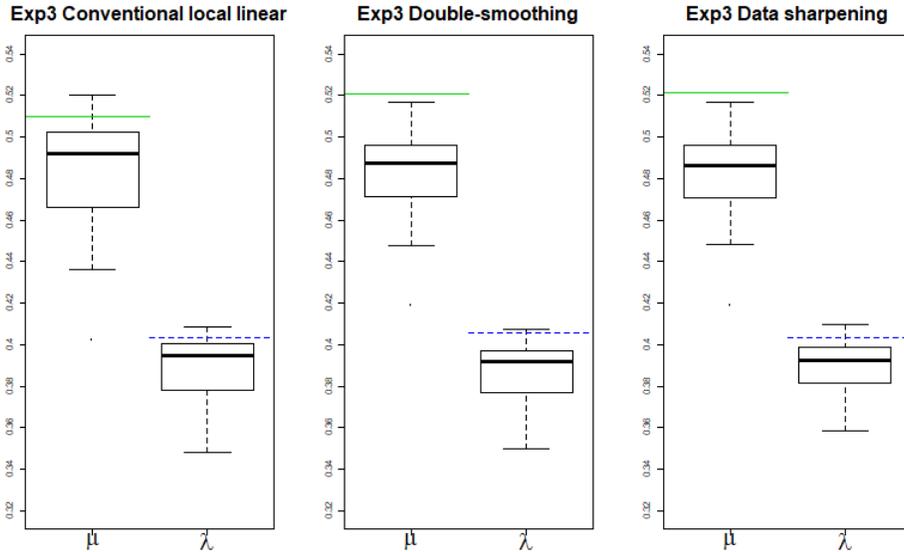


Figure 5.1: Boxplots for comparisons of parameter estimates between simulations and truth for the third simulation experiment. The boxplots represent samples of estimates based on 30 simulations. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ . Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

5.3.3 Bias and MSE Comparisons

The absolute value of bias is calculated by taking the absolute value of the difference between the true parameter values and the mean of the estimates from the simulations. In Figure 5.2, the absolute value of the bias in the estimates of μ is plotted for each of the 6 simulation experiments, for each of the three estimation techniques: local linear, double-smoothing local linear and data sharpening.

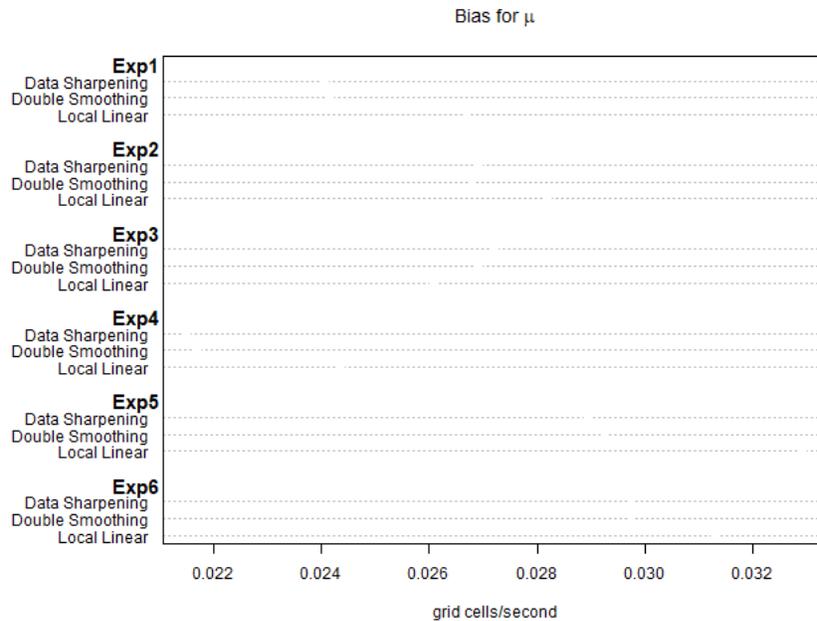


Figure 5.2: Bias for μ based on each technique used in each experiment.

Within each experimental group of estimates, the three techniques are represented by a different colour - conventional local linear smoother, double-smoothing, and data sharpening are presented in dark green, red, and blue, respectively.

The graph shows that in most cases, double-smoothing and data sharpening result in reduced bias, and these two techniques are competitive with each other since the blue dot and the red dot in each group are always close to each other. It is noticeable from the range of x-axis that the bias for μ is between .02 and .03 rather than exact .02.

5.3. Results

Hence, bias depend on actual values of parameter estimates.

Figure 5.3 the absolute value of the bias in the estimates of λ is plotted for each of the 6 simulation experiments, for each of the three estimation techniques: local linear, double-smoothing local linear and data sharpening. The colour coding is the same as in Figure 5.2.

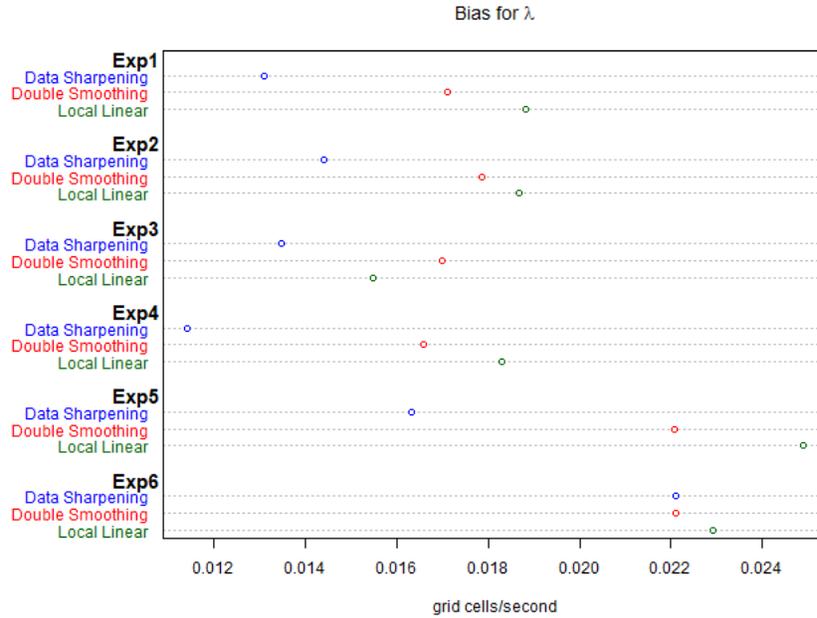


Figure 5.3: Bias of λ based on each technique used in each experiment.

The graph confirms that data sharpening does an excellent job at bias reduction as the blue dots are much closer to zero than the other two. Double-smoothing also provides reduced bias in general, compared to the conventional local linear estimator. The range of x-axis exhibits that the bias for λ is between .01 and .02, which is slight smaller than the bias for μ .

In order to take the variability of the different estimates into consideration, the root-mean-square error (RMSE) is also computed. The definition of RMSE for μ that we use here is (for 30 simulation runs)

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{30} (\mu - \hat{\mu}_j)^2}{30}}.$$

5.3. Results

RMSE for λ is in a similar fashion. Figures 5.4 and 5.5 illustrate the RMSE for μ and λ , respectively, for each experimental group. Colour-coding in these figures also follows the scheme used in Figure 5.2.

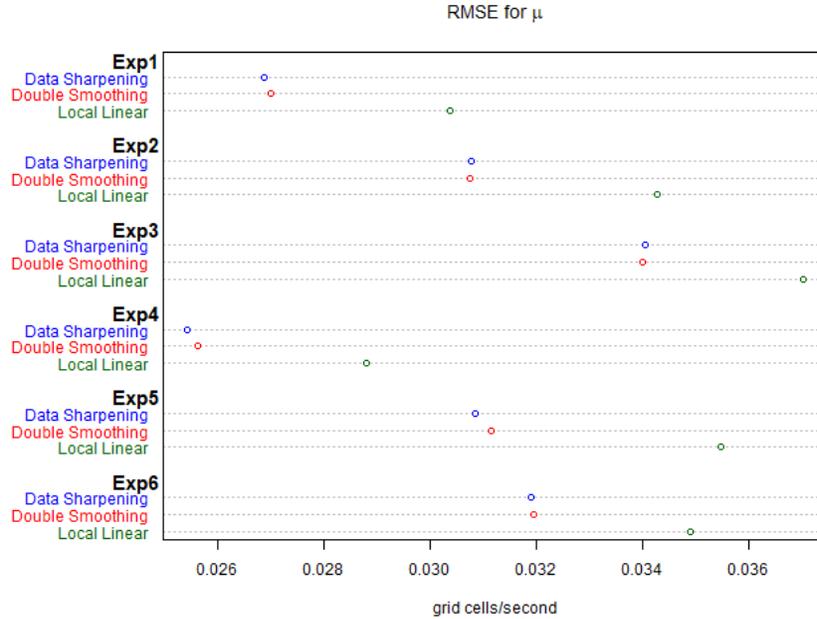


Figure 5.4: RMSE of μ based on each technique used in each experiment.

The two bias reduction techniques are very competitive in terms of RMSE for μ in Figures 5.4; data sharpening provides the best performance in terms of RMSE reduction in most cases, followed by double-smoothing. In Figure 5.5, we see that data sharpening outperforms both of the other techniques very clearly. Also, double smoothing provides an improvement over the conventional approach, though not as spectacularly.

5.3. Results

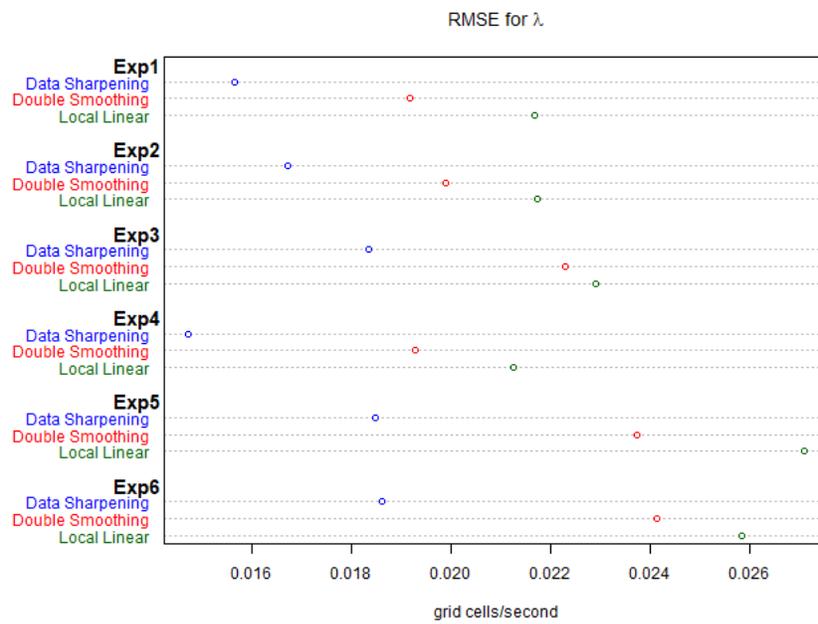


Figure 5.5: RMSE of λ based on each technique used in each experiment.

Chapter 6

Conclusions and Future Work

This work is part of an ongoing investigation into the suitability of a simple grid-based interacting particle system for stochastically modelling forest fire spread. Ultimately, we wish to fit such a model to sequences of satellite-based photographs of wildfires. Then, simulations of the model could be used to produce the maps of fire spread risk that are in demand by forest fire managers.

What can be firmly concluded is that the parameters for the simplest case of the model can be estimated from a sequence of photographs from a fire, using differential equations for the moments of certain statistics derivable from a video clip of a fire.

We have developed some goodness of fit methods. A simple visual assessment based on comparing burn patterns simulated from the fitted model with the observed pattern is a useful, if limited, first step. This method of assessment gives some assurance that the model appears to reasonably fit the data. However, such a comparison is highly subjective, and will not necessarily generalize to cases where, for example, the assumption of isotropy is invalid. The very nature of a stochastic model leads to different possible patterns under the same conditions. Hence, the following question arises: how different can the patterns be from the observed pattern before one might conclude that the model has failed?

What is needed, in general, is a metric for scoring burn pattern maps in terms of their shape and boundary characteristics. This thesis has proposed the four nearest-neighbourhood statistics as belonging to such a set of measures. On the basis of an informal bootstrap procedure applied to these statistics, we have a fair degree of confidence that the model is capturing much of the stochastic behaviour of the actual fire.

In terms of smoothing, it seems, from Chapter 5, that the bias reduction techniques work impressively well in variance reduction as

well. We can now draw the conclusion with confidence that double-smoothing and data sharpening can indeed provide improved performance in terms of bias reduction for standard nonparametric estimators of a regression mean. In the few cases where we do not see reduced bias, variability is dramatically smaller than when we do not apply bias reduction techniques. There is always a trade off between having small bias and variance. Double-smoothing and data sharpening are simple to implement, and data sharpening has a wide application to many estimators.

In the Introduction, reference was made to the use of “mouse” models in medical research; now, the question could be posed: how could this smouldering “mouse” model be used in fire research? We have now demonstrated a possible methodology that could be used to fit this model to large-scale fires given appropriate remote-sensing data, at least for homogeneous fuel types on relatively flat landscape under constant weather conditions. We also have provided a methodology for checking the accuracy of such a fitted model. With the estimated μ and λ parameters for a given fuel type and calm weather, we can use simulations from the model, after checking its validity, to compare probabilities of burning at given locations and times, assuming a fire has ignited in such an area. For fires in sloped and more windy environments, a slightly more complicated model is needed, involving spread rate parameters such as λ_N , λ_S , λ_W and λ_E associated with the four different nearest neighbour statistics. Extensions of our model fitting and assessment methodologies are needed, but once in hand, this model could be used in practice.

Another goal is to apply double-smoothing and data sharpening to the estimation of the derivative of the regression functions. The combination of these two techniques can also be investigated for bias or variance reduction; however, more variability might be introduced. Bandwidth selection is crucial and should be optimized in future work. Various combinations of estimators and l -fold data sharpening (i.e., apply data sharpening l times) can be investigated to determine whether better performing estimators can be found.

Bibliography

- [AGDMW12] A. Albert-Green, C.B. Dean, D.L. Martell, and D.G. Woolford. A methodology for investigating trends in changes in the timing of the fire season with applications to lightning-caused forest fires in Alberta and Ontario, Canada. *Canadian Journal of Forest Research*, 43(1):39–45, 2012. → pages 4
- [BBK⁺07] D. Boychuk, W.J. Braun, R.J. Kulperger, Z.L. Krougly, and D.A. Stanford. A stochastic model for forest fire growth. *Information Systems and Operational Research*, 45(1):9–16, 2007. → pages 1, 5
- [BJL⁺10] W.J. Braun, B. Jones, J. Lee, D.G. Woolford, and B.M. Wotton. Forest fire risk assessment in Muskoka, Ontario. *Journal of Probability and Statistics*, 2010:4–29, 2010. → pages 1
- [BK93] W.J. Braun and R.J. Kulperger. Differential equations for moments of an interacting particle process on a lattice. *Journal of Mathematical Biology*, 31:199–214, 1993. → pages 20
- [BW13] W.J. Braun and D.G. Woolford. Assessing a stochastic fire spread simulator. *Journal of Environmental Informatics*, 22(1):1–12, 2013. → pages 5, 6, 7, 15, 19
- [CH98] E. Choi and P. Hall. On bias reduction in local linear smoothing. *Biometrika*, 85(2):333–345, 1998. → pages 8, 10
- [CHR00] E. Choi, P. Hall, and V. Rousson. Data sharpening methods for bias reduction in non-parametric regression. *The Annals of Statistics*, 28(5):1339–1355, 2000. → pages 8, 11, 12

- [dev13] Libav developers. *ffmpeg. version 0.8.10-6:0.8.10-Oubuntu0.13.10.1*, 2000-2013. → pages 16
- [Fin04] M.A. Finney. Farsite: Fire area simulator-model development and evaluation. *The Annals of Statistics*, 2004. → pages 1, 4
- [FLA⁺05] M.D. Flannigan, K.A. Logan, B.D. Amiro, W.R. Skinner, and B.J. Stocks. Future area burned in Canada. *Climatic Change*, 72:1–16, 2005. → pages 4
- [FvW] → pages 4
- [GBBT08] T. Garcia, W.J. Braun, R. Bryce, and C. Tymstra. Smoothing and bootstrapping the prometheus fire spread model. *Environmetrics*, 19:836–848, 2008. → pages 4
- [GWZF04] N.P. Gillett, A.J. Weaver, F.W. Zwiers, and M.D. Flannigan. Detecting the effect of climate change on canadian forest fires. *Geophysical Research Letters*, 31(18):1–4, 2004. → pages 4
- [HB14] L.S. Han and W.J. Braun. Dionysus: A stochastic fire growth scenario generator. *Environmetrics*, 25:431–442, 2014. → pages 5
- [HH09] H. He and L Huang. Double-smoothing for bias reduction in local linear regression. *Journal of Statistical Planning and Inference*, 139(3):1056–1072, 2009. → pages 8, 9, 10, 11
- [MdDAG⁺06] J. R. Martinez-de Dios, J.C. Andre, J.C. Goncalves, B.C. Arrue, A. Ollero, and D.X. Viegas. Laboratory fire spread analysis using visual and infrared images. *International Journal of Wildland Fire*, 189(3-4):79–186, 2006. → pages 6
- [PKH⁺05] M. A. Parisien, V. G. Kafka, K. G. Hirsch, J. B. Todd, S. G. Lavoie, and P. D. Maczek. *Mapping wildfire susceptibility with the BURN-P3 simulation model*. 2005. → pages 1, 4

- [Tea16] R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0*, 2016. → pages 16
- [Tym05] C. Tymstra. Prometheus: The Canadian wildland fire growth model. *Forest Protection Division, Alberta Sustainable Resource Development, Edmonton, AB*, 2005. → pages 1, 4
- [Wan15] M.P. Wand. *KernSmooth: Functions for Kernel Smoothing*. R package version 2.23-15, 2015. → pages 20
- [WCDM10] D.G. Woolford, J. Cao, C.B. Dean, and D.L. Martell. Characterizing temporal changes in forest fire ignitions: looking for climate change signals in a region of the canadian boreal forest. *Environmetrics*, 21:789–800, 2010. → pages 4
- [WDM⁺14] D.G. Woolford, C.B. Dean, D.L. Martell, J. Cao, and B.M. Wotton. Lightningcaused forest fire risk in northwestern Ontario, Canada, is increasing and associated with anomalies in fire weather. *Environmetrics*, 25:406–416, 2014. → pages 4
- [WF93] B.M. Wotton and M.D. Flannigan. Length of the fire season in a changing climate. *The Forestry Chronicle*, 19:187–192, 1993. → pages 4
- [WJ95] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, 1995. → pages 13, 21
- [WNF10] B.M. Wotton, C.A. Nock, and M.D. Flannigan. Forest fire occurrence and climate change in Canada. *International Journal of Wildland Fire*, 19:253–271, 2010. → pages 4
- [WS98] M.G. Weber and B.J. Stocks. Forest fires and sustainability in the boreal forests of Canada. *Ambio*, 27:545–550, 1998. → pages 4

Bibliography

- [ZZAL92] J. Zhang, Y.-C. Zhang, P. Alstrom, and M.T. Levinsen. Modeling forest fire by a paper-burning experiment, a realization of the interface growth mechanism. *Physica A*, 15(1):383–389, 1992. → pages 7

Appendix

Appendix A

Figures

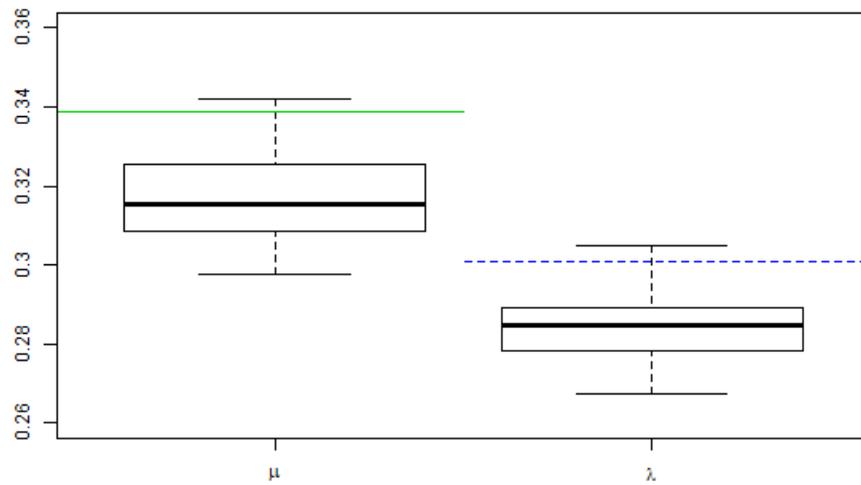


Figure A.1: Experiment 1: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.

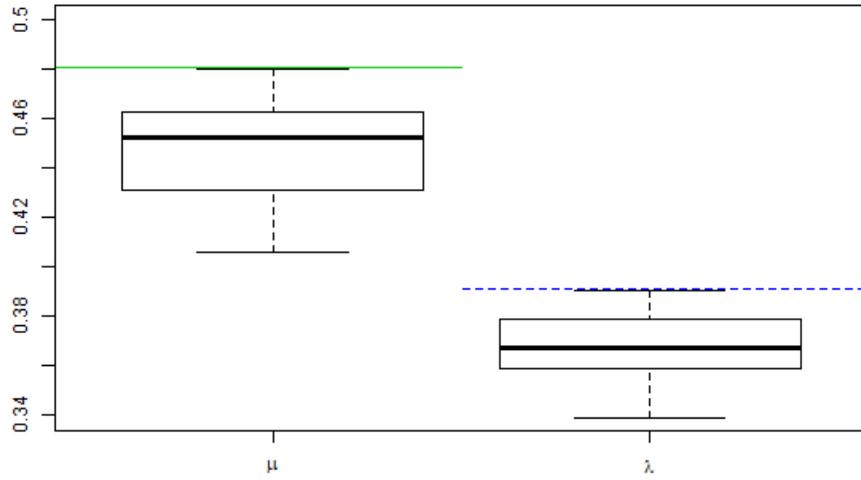


Figure A.2: Experiment 2: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.

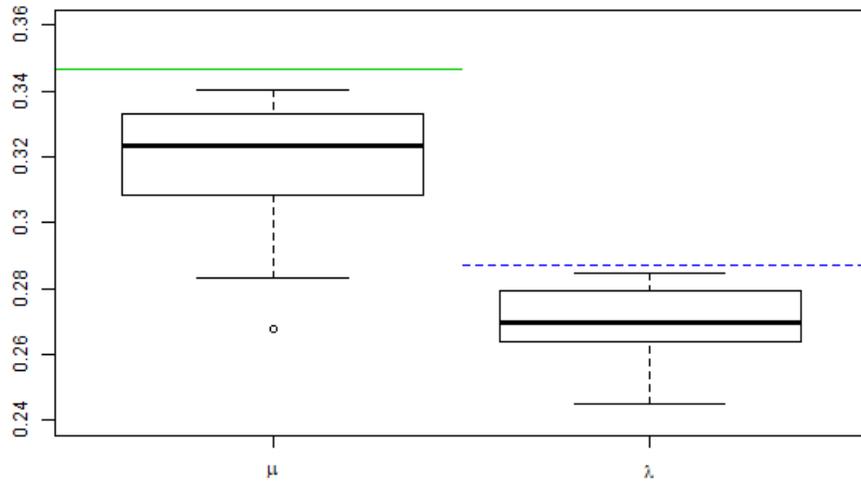


Figure A.3: Experiment 4: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.

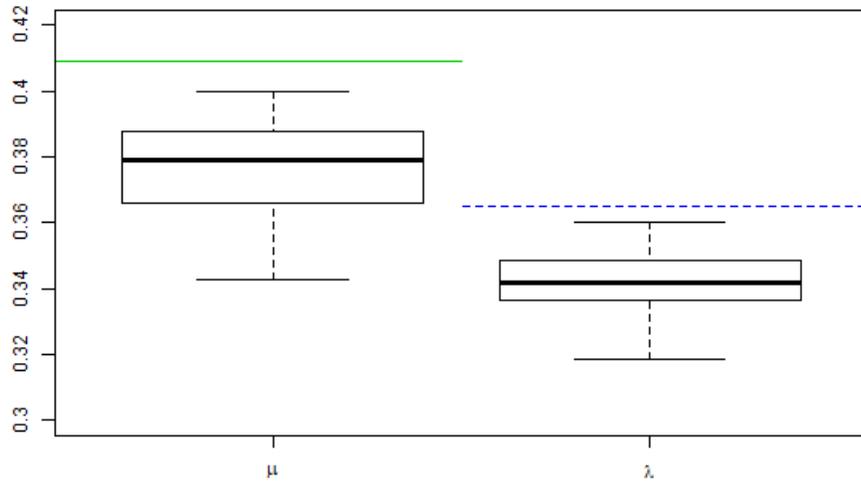


Figure A.4: Experiment 5: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.

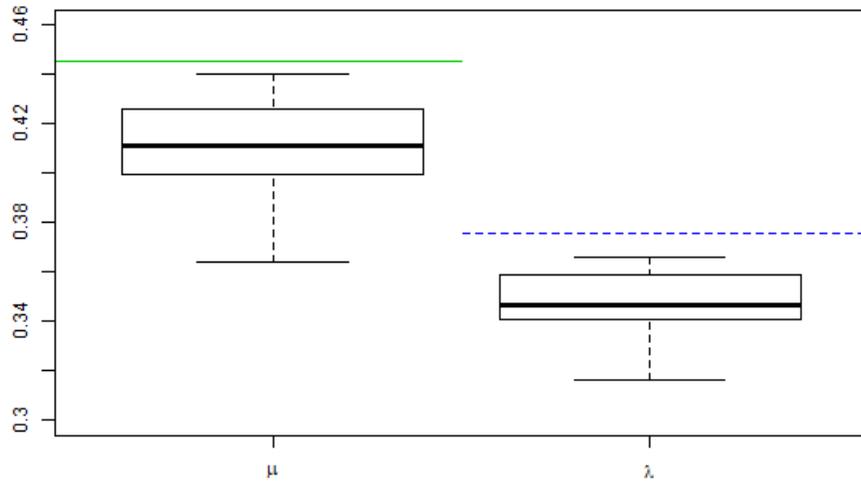


Figure A.5: Experiment 6: side-by-side boxplots for comparisons of parameter estimates between simulations and observed data for the third microfire. The boxplots represent samples of estimates for burn out rate μ (left) and burning rate λ (right) based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data.

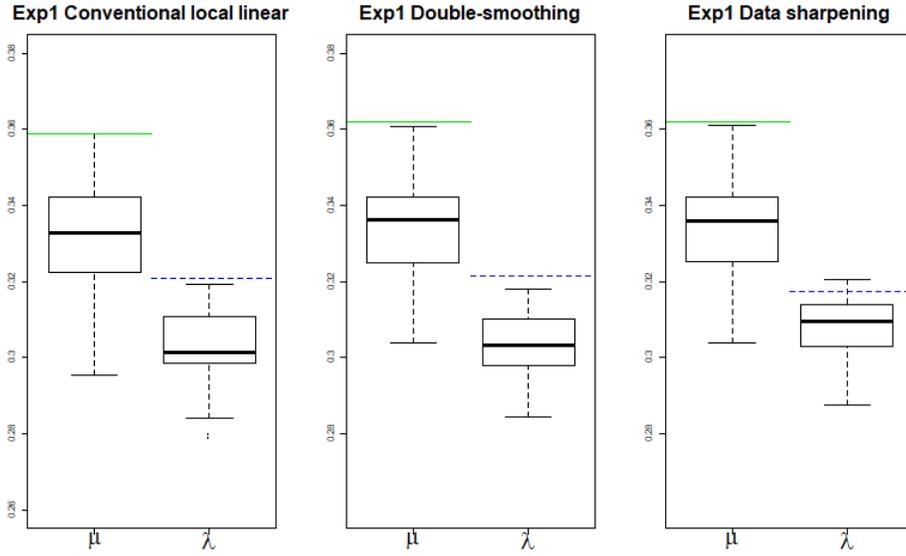


Figure A.6: Boxplots for comparisons of parameter estimates between simulations and observed data for the first microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

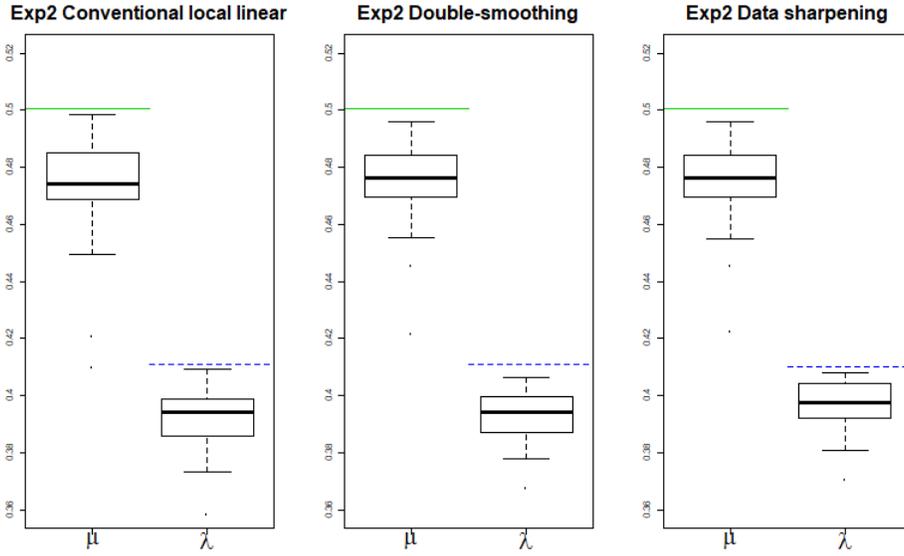


Figure A.7: Boxplots for comparisons of parameter estimates between simulations and observed data for the second microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

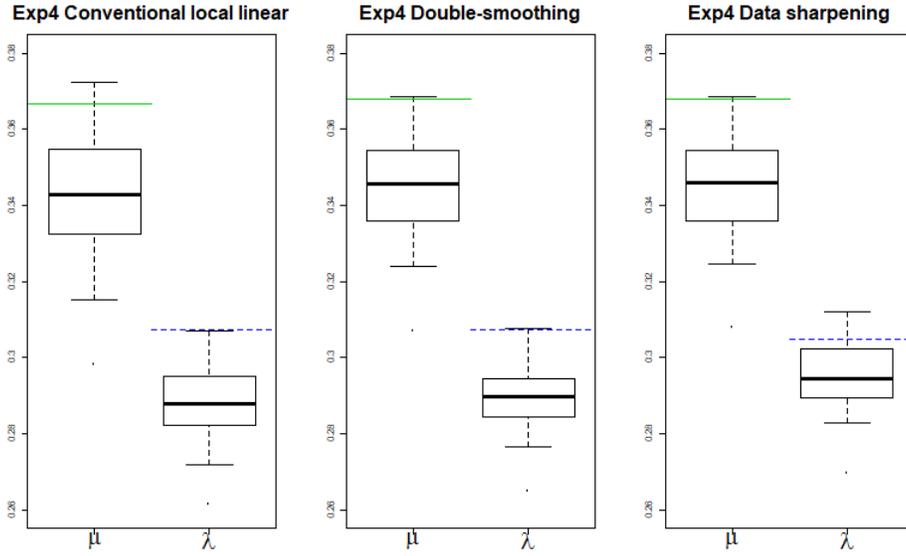


Figure A.8: Boxplots for comparisons of parameter estimates between simulations and observed data for the fourth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

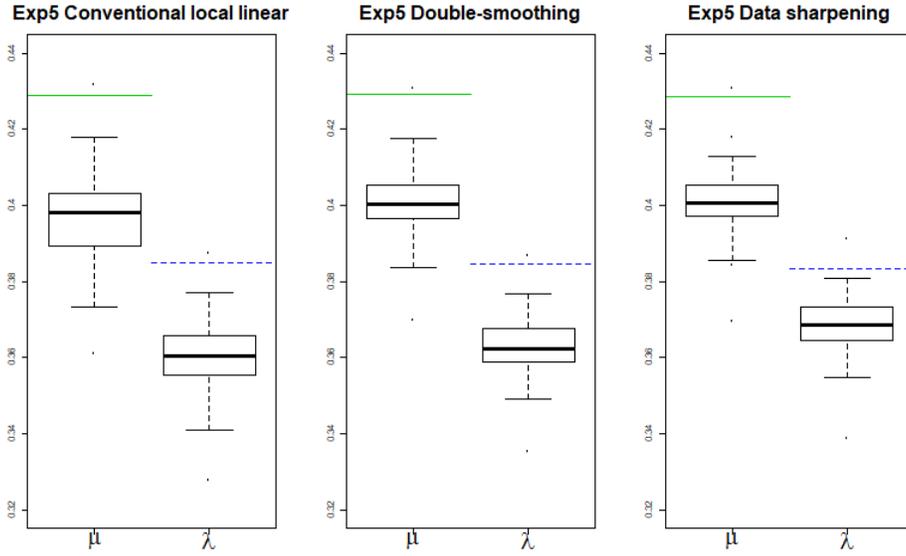


Figure A.9: Boxplots for comparisons of parameter estimates between simulations and observed data for the fifth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

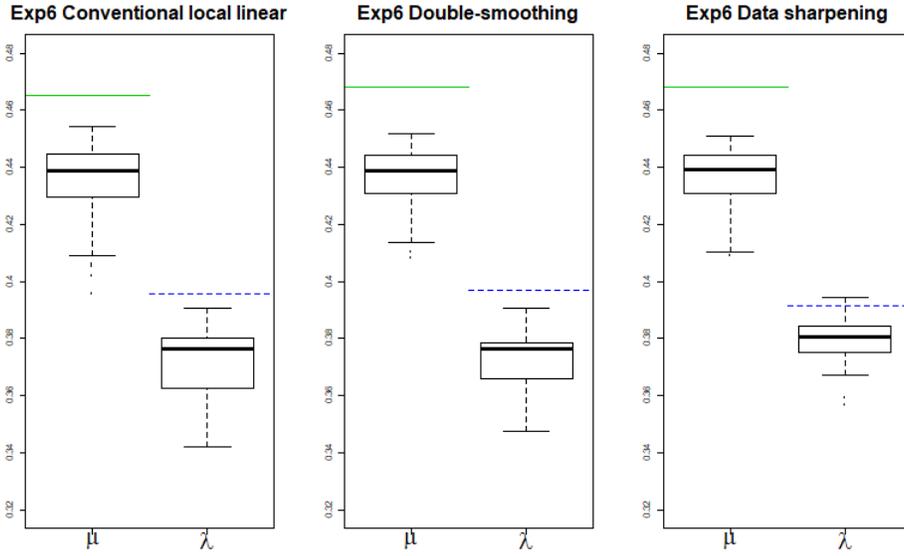


Figure A.10: Boxplots for comparisons of parameter estimates between simulations and observed data for the sixth microfire. The boxplots represent samples of estimates based on 30 simulations of the fitted stochastic spread model. The green solid horizontal line represents the true value of μ and the blue dashed horizontal line represents true λ , based on the observed data. Techniques used from left to right: conventional local linear, double-smoothing, and data sharpening.

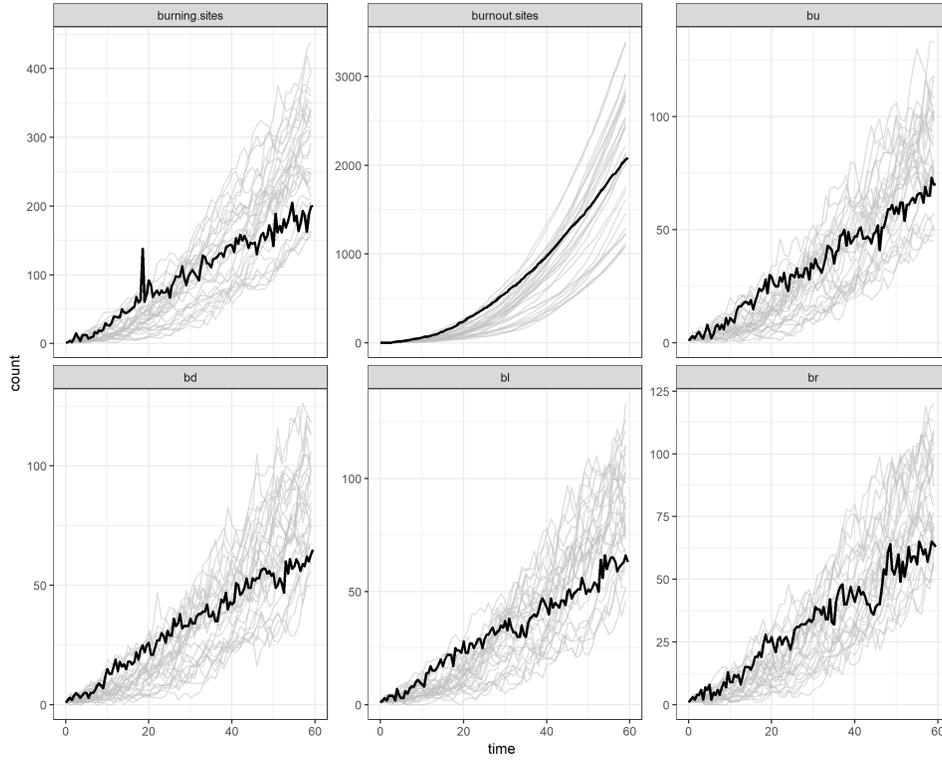


Figure A.11: Conventional local linear: comparisons between data from 30 simulation runs and observed data from the first micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.

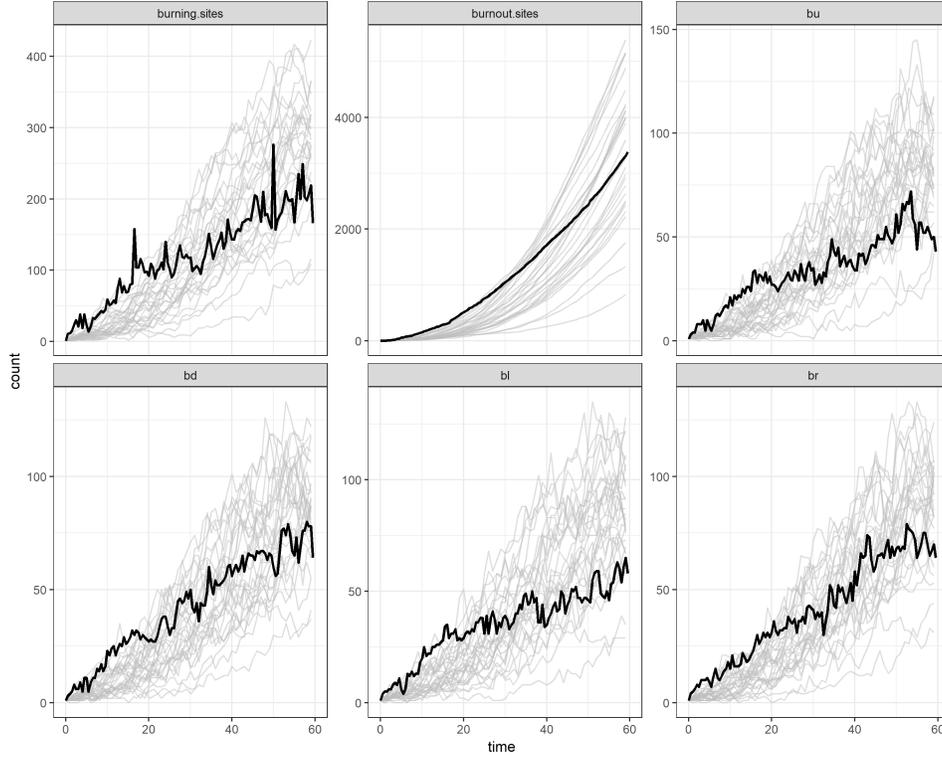


Figure A.12: Conventional local linear: comparisons between data from 30 simulation runs and observed data from the second micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.

Appendix A. Figures

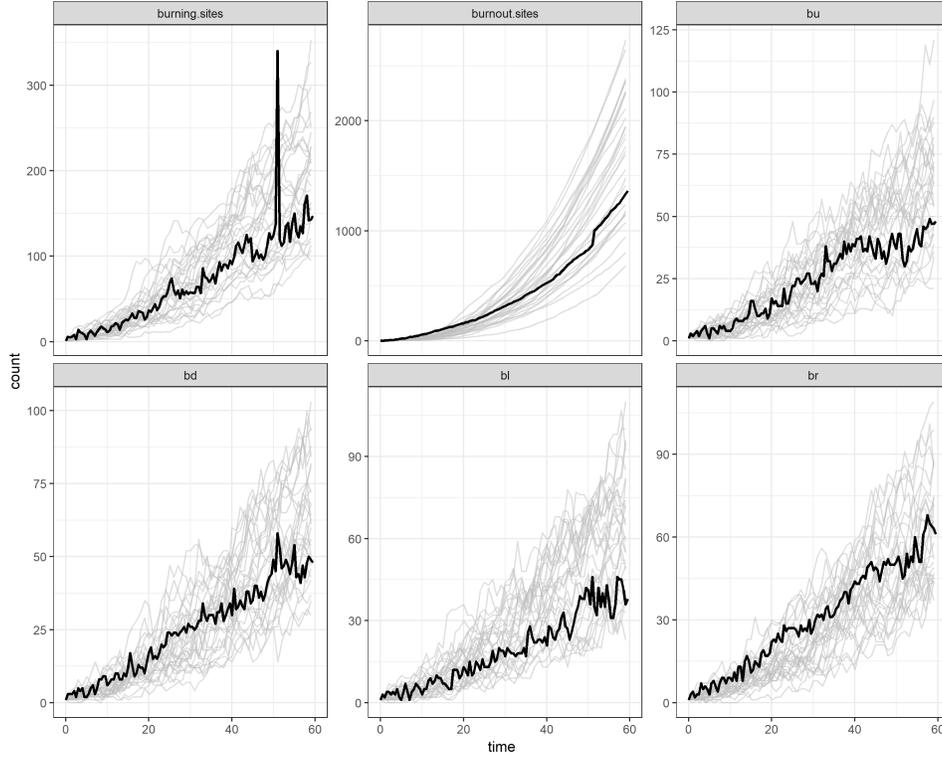


Figure A.13: Conventional local linear: comparisons between data from 30 simulation runs and observed data from the fourth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.

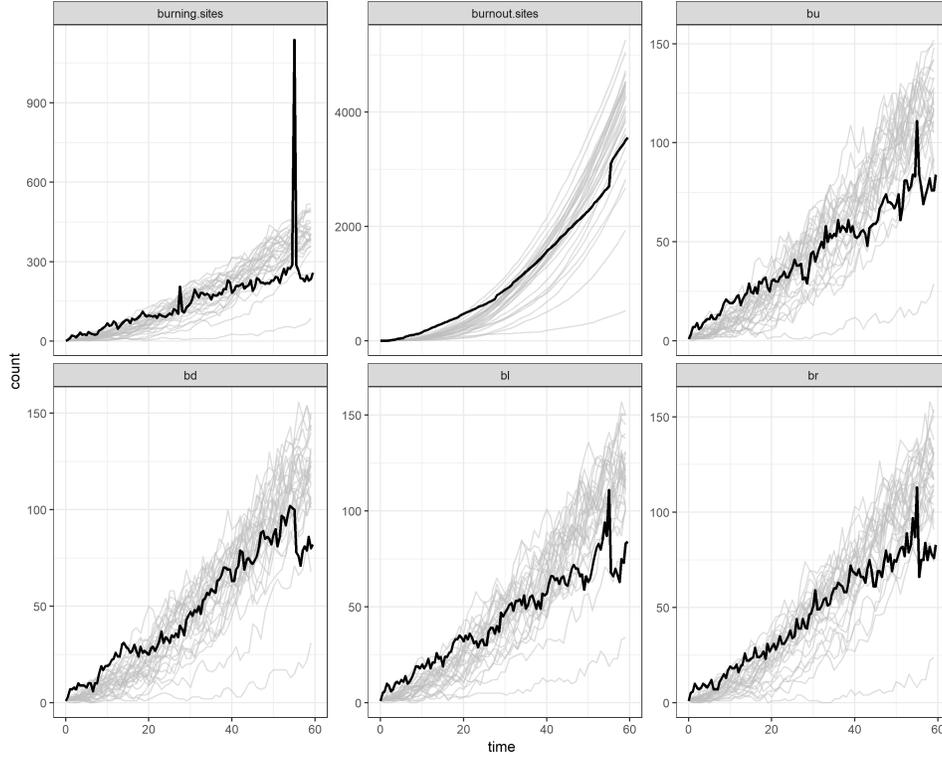


Figure A.14: Conventional local linear: comparisons between data from 30 simulation runs and observed data from the fifth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.

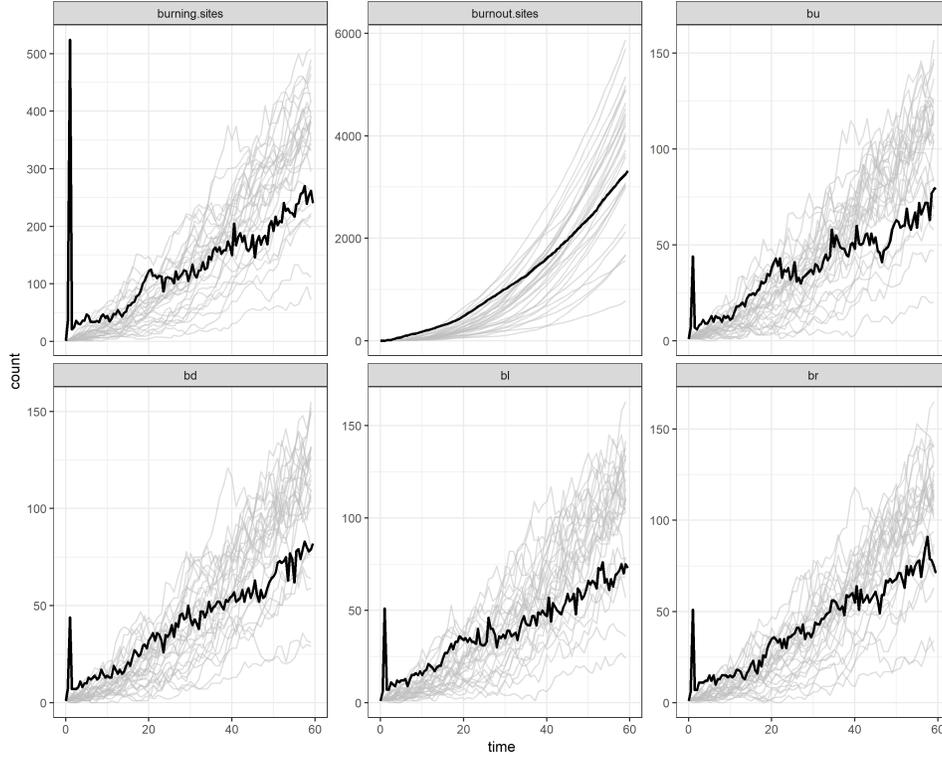


Figure A.15: Conventional local linear: comparisons between data from 30 simulation runs and observed data from the sixth micro fire. Top left: number of simulated (grey) and observed (blue) burning sites versus time; middle top: Number of simulated (grey) and observed (blue) burnt out sites versus time; remaining panels: numbers of simulated (grey) and observed (blue) neighbourhood counts versus time. The black curve on each panel corresponds to the observed data from the third micro-fire, upon which the estimates of μ and λ underlying the simulated data are based.