Anisotropic kernel smoothing for change-point data with an analysis of fire spread rate variability

John Ronald James Thompson
The University of Western Ontario

Supervisor
Braun, W. John
University of British Columbia Okanagan Co-Supervisor
Woolford, Douglas G.
The University of Western Ontario Joint Supervisor
Dean, Charmaine B.
University of Waterloo

Graduate Program in Statistics and Actuarial Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© John Ronald James Thompson 2018

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Statistical Methodology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/5889

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.
Abstract

Wildland fires are natural disturbances that enable the renewal of forests. However, these fires also place public safety and property at risk. Understanding forest fire spread in any region of Canada is critical to promoting forest health, and protecting human life and infrastructure. In 2014, Ontario updated its Wildland Fire Management Strategy, moving away from “zone-based” decision making to “appropriate response” decision making. This new strategy calls for an assessment of the risks and benefits of every wildland fire reported in the province. My research places the emphasis on the knowledge and understanding of fire spread rates and their variabilities. To satisfy these needs for a forest fire risk-benefit assessment tool that incorporates the effects of ignition, extinction, and spreading rate, the research herein explores new methods for spread rate estimation with an emphasis on understanding spread rate variability, for use in stochastic forest fire models.

In this research, we develop a novel anisotropic smoothing method for change-point data that uses estimates of the underlying data generating process to inform smoothing. We show that our anisotropic local constant and local linear kernel regression estimators are consistent with convergence rate $O\left(n^{-1/(q+2)}\right)$. We demonstrate their effectiveness on simulated one- and two-dimensional change-point data that are motivated by fire spread data. We detail the design and experimentation procedure of a micro-fire spread apparatus. We consider these micro-fire experiments a mouse model for wildland fire spread. We apply the anisotropic smoothers as image processors for measured data, and as an estimator for ignition and extinction event times at the pixel resolution. Those event times are then used to estimate instantaneous and average fire spread rates, and residency times for burning cells.

**Keywords:** Forest fires, nonparametric statistics, kernel smoothing, change-point estimation, image processing, computer vision, micro-fire experiments, fire growth, fire spread
Co-Authorship Statement

This work was completed under the supervision of Dr. John Braun, Dr. Charmaine Dean, and Dr. Douglas Woolford. All papers resulting from this thesis will be co-authored with Dr. Braun, and some with Drs. Dean and Woolford. Any papers that include the fire smoldering apparatus will be co-authored with Ms. X. Joey Wang, where my contribution is in the paper below.

Paper title: A Mouse-model For Studying Fire Spread Rates Using Experimental Micro-fires
Publication: Submitted to the Journal of Environmental Statistics
List of authors: John R.J. Thompson, X. Joey Wang, W. John Braun, and Douglas G. Woolford

The fire smoldering apparatus was inspired by the work of Zhang et al. (1992) and conceived by Dr. Braun. I designed the apparatus based on previous test experiments by Boychuk et al. (2007). I conducted experimentation to develop the apparatus, and Ms. Wang conducted the majority of experimentation for data collection under my supervision, and contributed to improving the design of the apparatus. Dr. Braun suggested using anisotropic diffusion to process the data and using data sharpening framework to estimate the colour channel signal. I implemented the anisotropic diffusion methods and designed the data sharpening methodology specifically to estimate ignition and extinction event times. I proposed kernel smoothing to estimate intra-pixel ignition and extinction times. Dr. Braun and I contributed to the preparation of this paper.
This one is for Dad.

Thank you for teaching me everything.
Acknowledgements

I first and foremost thank my supervisors, W. John Braun, Douglas Woolford, and Charmaine Dean, for their tireless support throughout my time at the University of Western Ontario. Their guidance in my academic work and personal life have helped me grow as a professional and as an individual. I am grateful to Dr. David Bellhouse and Dr. Vera Tai for their encouragement to participate in statistics-related consulting projects outside of my research; these projects have enriched my understanding of the application of statistics to the modern data analysis questions around the university.

I would like to thank Karen Nygard at the Integrated Microscopy lab at the University of Western Ontario for providing instruction on and usage of their scanning electron microscope to obtain microscopic images of wax paper and potassium nitrate coating.

Thank you to Andrew Jirasek at UBCO for providing the lab space needed for fire smoldering experiments and insights into experimental methodology. A thank you to Joey Wang for her help with the design of the apparatus and experimentation. I would like to thank the team at SHARCNET, and the members of the Computer Science Club and Computational Mathematics Club at the University of Waterloo for providing computational resources. I would also like to thank Mikko Karttunen, Jason Leopky, Neal McLoughlin, Matthew Thompson, and Cordy Tymstra for conversations and materials on forest fire research that guide the motivation of this dissertation.

I am thankful for the many insightful conversations with my labmates, particularly my academic brother-in-arms Devan Becker. Lastly, I would like to thank my friends and family, especially Claire Doran, whose love and support make life wonderful.
### Contents

Abstract .......................... i  
Co-Authorship Statement .......... ii  
Dedication ........................ iii  
Acknowledgements ................ iv  
List of Figures ..................... viii  
List of Tables ...................... xvi  

1 Introduction ..................... 1  
1.1 Motivating Study ............... 3  
1.2 Intended audiences and dissertation outline .............. 5  

2 Background theory ............. 7  
2.1 Nonparametric kernel regression .......... 7  
2.2 Gaussian and PDE-based anisotropic diffusion .......... 14  
2.3 Data sharpening ................ 16  
2.4 Fire spread modelling .......... 18  
2.4.1 The Canadian Forest Fire Behaviour Prediction System .... 18  
2.4.2 The BEHAVE system .......... 21  
2.4.3 Prometheus fire growth model .......... 23  
2.4.4 Boychuk fire growth model .......... 25  
2.5 Motivation for studying fire growth .......... 26  

3 Anisotropic smoothing ........... 28  
3.1 Regression with change-point data .......... 29  
3.2 Nonparametric anisotropic kernel regression .......... 33  
3.2.1 Anisotropic local constant estimator .......... 35  
3.2.2 Pilot estimator for anisotropic local constant estimator .......... 42  
3.2.3 Iterative estimation .......... 43  
3.2.4 Anisotropic local linear estimator .......... 44  
3.2.5 Anisotropic local polynomial estimator .......... 51  
3.2.6 Bandwidth selection methods .......... 52  
3.3 Simulation studies .......... 53
List of Figures

2.1  A Prometheus fire spread ellipse. ................................. 24

3.1  An example of change-point data, where there is an abrupt change in the regression function at $X = 1.5$. ................................. 30

3.2  The three one-dimensional data generating processes used for simulations. The left most figure is the piecewise constant regression function. The middle figure is the continuous regression function. The right most figure is the same continuous regression function with a jump at $X = 1.5$. ............ 54

3.3  Fitting the isotropic and anisotropic local constant estimators to the piecewise constant regression function data and comparing to the underlying data generating process (truth). This plot uses simulation parameters $n = 400$ and $\sigma = 0.5$, the uniform kernel function for smoothing, and least-squares cross-validation bandwidth selection method. ......................... 55

3.4  The MESEs for the three kernel estimators on the piecewise constant function. Note that $Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. ................................. 57

3.5  The MESEs for the three kernel estimators on data from the continuous regression function with a jump. Note that $Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. ................................. 59

3.6  Fitting the isotropic and anisotropic local constant estimators to the continuous regression function with a jump data and comparing to the underlying data generating process (truth). This plot uses simulation parameters $n = 400$ and $\sigma = 1$, the uniform kernel function for smoothing, and least-squares cross-validation bandwidth selection method. ......................... 61

3.7  Iteratively fitting the ALC estimator with LC as a pilot estimator on the piecewise constant regression function data. The bandwidth is selected once using least-squares cross-validation for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 62

3.8  Iteratively fitting the ALC estimator with LC as a pilot estimator on the piecewise constant regression function data. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 63
The MESEs for the ALC estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS). The simulated data is the piecewise constant regression function, and the parameters for the simulation are $\sigma = 1$ and $n = 800$. 64

The MESEs for the ALC estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS). The simulated data is the continuous regression function with a jump, and the parameters for the simulation are $\sigma = 1$ and $n = 800$. 65

Examples of a simulated fire data generating process with and without noise. The simulation’s video length is 70 equally spaced (in time) frames. 71

The isotropic local constant estimator applied to image frame $t_{35}$ of the simulated fire spread data, and the estimated errors. 72

The anisotropic local constant estimator applied to image frame $t_{35}$ of the simulated fire spread data, and the estimated errors. The first pair of plots use an ad hoc selected bandwidth. The second pair have the range kernel bandwidth increased by a factor of five. 73

The anisotropic diffusion estimator applied to image frame $t_{35}$ of the simulated forest fire spread data, and the estimated errors. 74

The mean of the MESE of fires at each simulated video frame. LC and ALC are the isotropic and anisotropic local constant estimator with least-squares cross-validated bandwidths. ALC.Z takes the cross-validated bandwidth for the range kernel, multiplies it by Z, and refits the anisotropic local constant estimator. The number of Monte Carlo replicates is 15 fires. 75

A fire smoldering experiment in action, using a $32^\circ$ slope configuration. The camera position in this figure is not the current standard. Here the camera is pointed orthogonally to the tray, but all experiments considered in this dissertation have the camera pointed orthogonally to the ground. 82

SEM images of uncoated wax paper. Note that for all SEM images, the distance on the scale (e.g. 300 $\mu$m) is for the entire length of the scale. 86

SEM images of the edge of wax paper coated in potassium nitrate. 87

SEM images of thickness of wax paper, with and without the potassium nitrate coating. 87

SEM images of burned uncoated wax paper. 88

SEM images of burned uncoated wax paper with flaking. 89

SEM images of burned coated wax paper with clumping. 90

SEM images of burned coated wax paper with clumps and hole formations. 91

SEM images of burned coated paper with signs of melting wax. 92

Fire spread images from the $11^\circ$ slope experiment 2. 94

A fire spread image from ten seconds into the $11^\circ$ experiment 2. The individual RGB channel values are shown separately next to the original image. The $z$-axis is the RGB channel value, and $(x, y)$-directions are the pixel locations of the image. 95
4.12 The differences of the RGB channels for the fire spread image from ten seconds into the $11^\circ$ slope experiment 2. The $z$-axis is the difference in the RGB channel values, and $(x,y)$-directions are the pixel locations of the image.

4.13 The red, green, and blue channel values of a single pixel over the course of the $11^\circ$ slope experiment 2.

4.14 The difference of the red and green channel values for a single pixel over the course of the $11^\circ$ slope experiment 2.

4.15 The result of applying an anisotropic diffusion filter on a smoldering image. The anisotropic diffusion procedure is 50 tensor field calculation iterations with 10 smooths per iteration.

4.16 The result for the RGB spectrum after applying image weighting and anisotropic diffusion on a smoldering image. The anisotropic diffusion procedure is 50 tensor field calculation iterations with 10 smooths per iteration. The $z$-axis is the RGB channel value, and $(x,y)$-directions are the pixel locations of the image.

4.17 The result for the red channel value after applying isotropic and anisotropic local constant estimators on a smoldering image. The bandwidths for the isotropic local constant smoother is selected using least-squares AIC, and the anisotropic range kernel is selected using the ad hoc procedure. The $z$-axis is the RGB channel value, and $(x,y)$-directions are the pixel locations of the image.

4.18 The red, green, and blue channel values for a single pixel over the course of a video that has been smoothed using anistropic local constant estimator.

4.19 The difference of the red and green channel values for a single pixel over the course of an entire video that has been smoothed using anistropic local constant estimator.

5.1 An example of a data sharpened result for the evolution over time of a single pixel’s red channel value.

5.2 An example of a data sharpened result for the evolution over time of a single pixel’s red and green channel values simultaneously.

5.3 The ignition event times estimated using data sharpening for $11^\circ$ slope fire experiment 2. The $z$-axis is time, and $(x,y)$-axes are the pixel coordinates.

5.4 A single pixel’s red channel value over time, smoothed using isotropic and anisotropic smoothers. The vertical red dotted line represents the ignition event time estimated for this pixel.

5.5 The ignition event times estimated using anisotropic smoothing for $11^\circ$ slope fire experiment 2. The $z$-axis is time, and $(x,y)$-axes are the pixel coordinates.

5.6 Using the np package’s (Hayfield and Racine, 2008; R Core Team, 2018) local constant estimator with least-squares cross-validation (Hurvich et al., 1998) for bandwidth selection on the ignition time data in Figure 5.5. The $z$-axis is time, and $(x,y)$-axes are the pixel coordinates.
5.7 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 2. The colour gradient is time in the first plot, and magnitude of the gradient in the second. ......................................................... 123
5.8 Uphill and downhill instantaneous spread rates for 0° slope experiments. 124
5.9 Uphill and downhill instantaneous spread rates for 11° slope experiments. 125
5.10 Uphill and downhill instantaneous spread rates for 32° slope experiments. 126
5.11 The estimated residency times for all videos. The residency times are calculated from the data sharpening estimated ignition and extinction event times. ........................................ 127
5.12 The estimated residency times for a 21 × 21 grid centered around the ignition point for all videos. The residency times are calculated from the data sharpening estimated ignition and extinction event times. .... 128

A.1 The MESEs for the three kernel estimators on the piecewise constant regression function with \( \sigma = 0.1 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 133
A.2 The MESEs for the three kernel estimators on the piecewise constant regression function with \( \sigma = 0.5 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 134
A.3 The MESEs for the three kernel estimators on the piecewise constant regression function with \( \sigma = 1 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 135
A.4 The MESEs for the three kernel estimators on the piecewise constant regression function with \( \sigma = 2 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 136
A.5 The MESEs for the three kernel estimators on data from the continuous regression function with a jump and \( \sigma = 0.1 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 137
A.6 The MESEs for the three kernel estimators on data from the continuous regression function with a jump and \( \sigma = 0.5 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 138
A.7 The MESEs for the three kernel estimators on data from the continuous regression function with a jump and \( \sigma = 1 \). Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \). ......................................................... 139
A.8 The MESEs for the three kernel estimators on data from the continuous regression function with a jump and $\sigma = 2$. Note that .Z means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$.

A.9 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.10 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.11 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using the least-squares cross-validation method once for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.12 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.13 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.14 Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.15 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.16 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.17 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using the least-squares cross-validation method once for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 

A.18 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.19 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

A.20 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.
A.18 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 150

A.19 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 151

A.20 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 152

A.21 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 153

A.22 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 154

A.23 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the least-squares cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 155

A.24 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 156

A.25 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 157

A.26 Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. ................................. 158

B.1 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of $11^\circ$ slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ................................. 160
B.2 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 161

B.3 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 162

B.4 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 163

B.5 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 164

B.6 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 165

B.7 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 166

B.8 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 167

B.9 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 168

B.10 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 169

B.11 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 170

B.12 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. ............................. 171
B.13 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 5. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 172

B.14 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 6. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 173

B.15 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 7. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 174

B.16 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 8. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 175

B.17 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 9. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 176

B.18 The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 10. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second. .................................................. 177
List of Tables

3.1 The mean of the MESEs for each nonparametric estimator fitted to the simulated piecewise constant regression function data. .......................... 56
3.2 The sample standard deviation of the MESEs for each nonparametric estimator fitted to the simulated piecewise constant regression function data. ............................... 56
3.3 The mean of the MESEs for each nonparametric estimator fitted to the simulated continuous regression function data. ................................. 58
3.4 The sample standard deviation of the MESEs for each nonparametric estimator fitted to the simulated continuous regression function data. ................................. 58
3.5 The mean of the MESEs for each nonparametric estimator fitted to the simulated continuous data with a jump in the regression function. ................................. 60
3.6 The sample standard deviation of the MESEs for each nonparametric estimator fitted to the simulated continuous data with a jump in the regression function. ................................. 60
3.7 The mean of the MESEs for the anisotropic local constant estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS). .................................................. 62
3.8 The global mean and sample standard deviation of the MESEs over all simulated fires for the isotropic and anisotropic local constant estimators and anisotropic diffusion estimator. .................................................. 67

4.1 Number of completed experiments and usable fire spread measurements. .......................... 92

5.1 Parameter estimates for the uphill (West) direction. .............................. 117
5.2 Parameter estimates for the downhill (right) direction. ............................ 117
5.3 The maximum likelihood parameter estimates of a gamma distribution for the three sloped configuration residency times. .............................. 119
5.4 The maximum likelihood parameter estimates of a gamma distribution for the three sloped configuration residency times of a $21 \times 21$ grid centered around the ignition point. .............................. 120

A.1 R-packages used in simulations, data visualizations, type-setting and dissertation creation. .......................... 133
Chapter 1

Introduction

Regression analysis is the study of the relationship between a response variable and a set of covariates. Our interest is in the expected relationship (that is, the regression function) between a continuous response variable and one or more continuous explanatory covariates. Our focus is on change-point data where there is a sudden change or “jump” in the regression function (Qiu, 2005). We have developed consistent anisotropic local constant and local linear kernel estimators for a univariate continuous response and $q$-dimensional ($q \in \mathbb{Z}^+$) multivariate explanatory change-point data, where the data has a finite unknown number of change-points with unknown locations and change magnitudes. Bandwidth selection methods for these estimators have been chosen to capture regions with change-points while minimizing error by balancing the bias and variance of the estimator. These estimators automatically smooth change-points in $q$-dimensional explanatory data, which was previously impossible due to the nature of specifying criteria for change-point location and magnitude in high dimensions.

For an example of a change-point data set, consider setting up a light intensity detector in an empty, windowless basement office. The response data is the measured light intensity and the covariate is time. At 6 am, when a post-doc flips on the light switch to find that a Ph.D. student has fallen asleep on their desk, the light intensity almost instantly jumps. This jump from one level to another, with some noise in the measurements, generates a change-point data set.

It is well known that determining the underlying data generating process for the
expected relationship between response and covariate variables of a data set is impossible. When choosing a stochastic model, it is impossible to know all physical and statistical processes of a relationship, even with the best of intuitions. We believe this causes some of those processes to be overly simplified or incorrectly specified, and that specifying the correct physical model and correct distribution of the noise (not accounted for in the physical understanding) is untenable. For example, we could apply an increasing number of physical constraints to a stochastic model, but any errors we make in our guess bias our results away from the true behaviour of the system we are trying to model. Any broad assumptions about the noisy data generating process further bias our model’s results away from that process.

The main contribution of this dissertation is the development of novel nonparametric anisotropic smoothing methods. Nonparametric statistics methodology concentrates on the ability to estimate and predict an underlying data generating process with minimal assumptions and restrictions on the process (Li and Racine, 2007). The statistical perspective of this dissertation strongly favours using nonparametric statistical methodology, aided by the physical understanding of a real system, to estimate regression functions and conduct analysis and inference. We use nonparametric methods to allow the data to guide our understanding, estimations, and predictions. We compare nonparametric and parametric methods to further show that data-driven methods with fewer assumptions provide superior results, particularly in the context of our application to estimating fire spread rates in micro-fire experiments.

The nonparametric methods that appear in this dissertation are data sharpening (Braun and Hall, 2001) and kernel smoothing (Li and Racine, 2007). Data sharpening is a technique used to improve statistical estimation methods by perturbing the data. Kernel smoothing is the estimation of the underlying regression function where we only specify the smoothness, not shape, of that regression function.

We have developed a kernel regression estimator that automatically detects and smooths change-point data. The “anisotropic” local constant and local linear kernel estimators are similar to their isotropic versions, where we add an additional kernel weighting term that allows for the underlying data generating process to be incorpo-
rated into the smoothing procedure. Anisotropic local smoothing allows us to smooth a one-dimensional response variable for any number of explanatory variables, without specifying either the shape of the regression function or the locations of its change-points. We prove these kernel estimators for \( q \) explanatory variables have weak consistency with a rate of convergence of \( O(n^{-1/(q+2)}) \) as sample size increases. We discuss bandwidth selection methods for our anisotropic estimators in the presence of change-point data. We study the ability of these estimators to preserve regions of a fire image using simulated data and real data from micro-fire spread experiments.

1.1 Motivating Study

The prediction of forest fire behaviour is key to protecting people, human infrastructure, and forest ecology in Canada and across the globe. Fire managers and academic researchers work together developing statistical models to improve forest fire event and attribute predictions, and to aid operational decision-making on the detection, monitoring, and suppression of fires (Rothermel, 1983; Alexander and Thomas, 2004). Academic researchers contribute in the creation, development, refinement, and understanding of statistical models. Managers drive the design of statistical tools based on their needs as operations managers, and contribute their expert understanding and intuition on forests and forest fire phenomena.

Deterministic and stochastic fire spread models require parameters as inputs to simulate a fire spread event. The ability to accurately estimate fire spread parameters directly affects the fire spread model’s ability to predict fire spread behaviour (Alexander and Cruz, 2013). One contribution of this dissertation is the development of statistical methodology that estimates fire spread rates. We look to estimate those rates using a nonparametric perspective and encourage using nonparametrically-determined fire spread rate distributions in the stochastic modelling of fire growth.

We study experimental fire spreads with the goal of improving fire spread rate estimation methods that directly influence every fire spread simulation model (Sullivan, 2009c). We are interested in quantifying a fire spread rate distribution’s variability. That vari-
ability affects the uncertainty in the prediction of forest fire behaviour, which affects the tools that fire managers use to understand the possible outcomes of fire spread scenarios (Alexander and Thomas, 2004), which affects their decisions for monitoring and suppression, which impacts the safety of firefighters in the field (Butler, 2014), urban and industrial infrastructure, human life, and the ecology of forests (Agee, 1996).

Forest fires are very complex physical phenomena influenced by many environmental factors. Fires are influenced by the topography of the landscape, the local and macro fuel types, the age of the forest, the humidity and moisture in the fuel, the weather leading up to and during a fire, the intensity of the fire, the type of fire (e.g. ground and crown fires), and so on. Fire managers use their understanding of the environmental variables of their region to monitor, predict, and react to forest fire behaviours. We do not develop new methods to estimate the environmental variables for forest fires, but we do use environmental conditions in modelling and calculations.

In science, the use of mice to study human conditions and diseases—prevalent in the medical sciences—is known as a “mouse model”. We introduce a mouse model for studying fire spread through fire smoldering experiments at the laboratory scale. The fire smoldering apparatus we have developed for these experiments involves the following: a piece of wax paper is coated with potassium nitrate, attached to a metal tray, and placed in a fume hood with a digital camera centered above. The paper is ignited and measurements of fire spreading across the paper are collected by the camera in the visual spectrum into red, green, and blue (RGB) colour channels. Using this experimental procedure, a data set of fire spread measurements with different environmental configurations has been created. These configurations involve different slopes, wind velocity, moisture levels, firebrand spotting, surrounding lighting, and types of paper. We believe the data collected from the smoldering experiments will provide insight into the mechanisms of fire spread and interaction with environmental conditions, and the statistical analysis methods for fire spread imagery data.

Statistical tools developed for fire managers analyze collected data for decision-making on a fire spread scenario. The imagery data collected from smoldering experiments can be interpreted as two-dimensional change-point data, where there are three regions: fuel,
burning, and burnt-out material. We believe there is knowledge to be gained about fire spread from the statistical analysis of fire spread images. The methods developed in this dissertation are intended to improve the analysis and inference of fire spread rates in deterministic and statistical fire spread models. Our research provides new methods of estimating fire spread rates from the satellite perspective of fire spread images.

We use our anisotropic local constant kernel estimator to estimate the underlying data generating process of fire spread images. After smoothing each image of a smoldering experiment video, we apply data sharpening and anisotropic smoothing methodologies to estimate each pixel’s ignition and extinction event times. These event time estimates are smoothed using local constant regression estimators to estimate the behaviour of fire spread between sequential fire spread images, and to calculate instantaneous fire spread rates at each pixel location. Then, we study residency time from ignition to extinction, and the effect of slope on the distribution of fire spread rates and their variability.

1.2 Intended audiences and dissertation outline

This dissertation is intended for statisticians and applied practitioners with high-dimensional change-point data sets who want to analyze the underlying data generating process, and fire managers interested in quantifying the distribution of fire spread over a region. We focus on developing new methods of analysis for fire spread rates that are intuitive to researchers and fire managers. We hope that the discoveries in fire spread rate estimation will be directly useful to fire management operations, and that fire managers may incorporate the understanding of fire spread rates gained from this research.

The chapter structure of the dissertation is: (2) a literature review of relevant non-parametric statistical methods, image processing methods, and forest fire spread models, (3) the theory and simulation study of anisotropic local constant and local linear kernel smoothing to analyze $q$ dimensional change-point data, (4) the design and experimentation procedure of a fire smoldering apparatus, and the processing of measurements from experimentation, (5) the estimation of ignition and extinction event times and fire spread rates, and the analysis of fire spread rates and burning residency times, and (6)
a discussion on future work to close the dissertation.
Chapter 2

Background theory

We begin by introducing the nonparametric statistical methodology and philosophy that is omnipresent in this dissertation. The methodologies used are kernel smoothing in the regression framework and data sharpening. We introduce image processing for denoising images while preserving boundaries using partial differential equation methods. We introduce fire spread modelling, where we look at the Canadian FBP System and the BEHAVE system for estimating fire rates of spread, and the Prometheus and Boychuk fire growth models that use these estimated fire rates of spread as inputs. Any further background information on specific topics that fit into only one chapter is found in the exposition. The notation in Section 2.1 and Chapter 3 adheres as closely as possible to the notation of Li and Racine (2007). We choose this notation as it is commonly used in the nonparametric literature. For clarity, we specify the forms of quantities (scalar, vector, matrix, etc.) when those details may not be clear from the context.

2.1 Nonparametric kernel regression

Statistical regression is an extremely broad topic that encompasses many models, methods, and approaches. If given a vector of explanatory variables \( X = \{X_j\}_{j=1}^q \) and a univariate response \( Y \), an objective of statistical models is to gain understanding on quantities such as the expectation of the response variable given some explanatory data, how the response changes with different explanatory data, if this relationship is due to
random chance or realistic, and what conclusions we make about the relationship between $X$ and $Y$. Most parametric regression models are generalizable to

$$ Y = g(X, \beta) + \epsilon, $$

where $\epsilon$ is a noise term, $\beta$ is a vector of shape parameters, and $g$ is a function that relates the response variable to the explanatory variable. The number and type of assumptions on $g$ vary in parametric and nonparametric models. Parametric modellers focus on specifying the functional form of $g$ and the distributions of $Y$, $X$, and $\epsilon$. A popular choice for $g$ is the first order polynomial linear regression model $g(X, \beta) = \beta_0 + \sum_{i=1}^{q} \beta_i X_i$, where the estimators for $\beta$ are well known. Nonparametric modellers forgo specifying the functional form and specify less restrictive assumptions on $g$, where the assumptions are typically on differentiability conditions on the function $g$ and on moment conditions on the noise term $\epsilon$. Assuming an additive noise structure, a nonparametric regression model is written as

$$ Y = g(X) + \epsilon, $$

where the functional form of $g$ is not specified and no longer depends on shape parameters. These assumptions and lack of parameters are used in the design and derivations of nonparametric estimators. The first nonparametric regression estimator in this review begins with the assumption (Watson, 1964) that there exists a joint distribution $f(x,y)$ between $X$ and $Y$ such that

$$ g(x) := E[Y|X = x] = \frac{\int y f(x,y) dy}{\int f(x,y) dy}, \quad (2.1) $$

where $x = \{x_j\}_{j=1}^{q}$ is the set of realizations of $X$ and we forego the subscripts of the joint distribution for simplicity, i.e. $f(x,y) = f_{X,Y}(x,y)$. It was shown that $E[Y|X = x]$ is the optimal predictor of $Y$ given $X$ (Li and Racine, 2007, p.59, Theorem 2.1) such that

$$ E[(Y - r(X))^2] \geq E[(Y - E[Y|X])^2] \quad \forall \ r(\cdot) \in \mathcal{G}, $$
where $G$ is the set of all Borel measurable $r(X)$ with a finite second moment. For this property, all continuous functions with countably many discontinuity points are Borel measurable, and this fact is important for derivations of change-point regression estimators in Chapter 3.

The next step is to consider estimating the numerator and denominator of Equation (2.1). The denominator is the marginal distribution $f_X(X)$, which can be estimated using the nonparametric multivariate Rosenblatt-Parzen density estimator (Rosenblatt, 1956; Parzen, 1962). In the numerator, the integral is estimated similarly using a multivariate Rosenblatt-Parzen estimator for the joint distribution $f(x,y)$. This approach is one way to obtain the most popular nonparametric regression estimator, the Nadaraya-Watson or local constant kernel regression estimator (Watson, 1964; Nadaraya, 1965), written as

$$
\tilde{g}(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right)},
$$

(2.2)

where $K(\cdot)$ is a multivariate kernel function, the $i$th vector of random variables $X_i = \{X_{ij}\}_{j=1}^{q}$. The kernel function is a weighting function with useful statistical properties for kernel regression estimation, such as proving the local constant estimator converges in probability to the data generating process $g$. In this dissertation, we choose $K(\cdot)$ to be a product kernel function with support $\mathbb{R}^q$ given by

$$
K \left( \frac{X_i - x}{h} \right) = k \left( \frac{X_{i1} - x_1}{h_1} \right) \times k \left( \frac{X_{i2} - x_2}{h_2} \right) \times \ldots \times k \left( \frac{X_{iq} - x_q}{h_q} \right),
$$

(2.3)

where $k(\cdot)$ is a univariate kernel function and $h = (h_1, h_2, \ldots, h_q)$ is a vector of bandwidths. We choose the product kernel as it assumes orthogonal weights between directions. The $\nu$-order univariate kernel function $k(\cdot)$, where $\nu \in \mathbb{N}$, is any nonnegative
bounded kernel function that satisfy the properties:

\begin{align}
(i) \quad & k(z) \geq 0 \quad \forall \ z \in \mathbb{R}, \\
(ii) \quad & \int_{\mathbb{R}} k(z) \, dz = 1, \\
(iii) \quad & \int_{\mathbb{R}} z^\ell k(z) \, dz = 0, \ \ell \in \{1, 2, \ldots, \nu - 1\}, \\
(iv) \quad & \int_{\mathbb{R}} z^\nu k(z) \, dz \neq 0.
\end{align}

(2.4)

The order of the kernel function directly affects the rate of weak convergence for the local constant and other kernel estimators. Some examples of kernel functions (Silverman, 1986) are:

- Uniform: \( k(z) = \frac{1}{2} \) for \(|z| < 1\), otherwise 0,
- Gaussian: \( k(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}, z \in \mathbb{R}, \)
- Epanechnikov: \( k(z) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}z^2) \) for \(|z| < \sqrt{5}\), otherwise 0,
- Biweight: \( k(z) = \frac{15}{16}(1 - z^2)^2 \) for \(|z| < 1\), otherwise 0.

The Gaussian and biweight kernels are more commonly used in practice, as they handle derivatives well and have efficient rates of convergence to the data generating process relative to the most efficient Epanechnikov kernel (Silverman, 1986; Wand and Jones, 1994). The uniform kernel suffers from the greatest efficiency loss and the least smoothness, but has advantages in simplicity from a theoretical perspective. The Gaussian kernel has an infinite support which may provide greater smoothness Wand and Jones (1994). A truncated version of the Gaussian kernel is also used in the literature (Cheng, 1995), where the support is a bounded interval. Therefore, the Gaussian and uniform kernels are used in this dissertation.

The vector \( h \) contains the smoothing parameters that act as bandwidths (or window widths) in the function \( K \left( \frac{X - x}{h} \right) \). The choice of bandwidths determines the weight of each \( Y_i \) local in \( x \) that is used to estimate \( g(x) \). This is seen by changing the bandwidths
in Equation (2.3); as \( h_j \) decreases, \( \frac{X_j - x_j}{h_j} \) increases, and \( k \left( \frac{X_j - x_j}{h_j} \right) \) decreases (except when \( X_j = x_j \) such that \( K(0) \) is constant for any \( h_j \)). The bandwidth does not affect the shape of the regression relationship, but controls the smoothness of the estimated function.

There are many bandwidth selection methods in the nonparametric literature. The first option is to guess, which almost never yields an optimal result. The next option is a rule-of-thumb method where the bandwidth is given by \( h = c X_{sd} n^{-1/(q+4)} \) and \( X_{sd} \) is the \( q \)-vector of standard deviations of each \( X_j \) where \( j \in \{1, \ldots, q\} \), and \( c \) is a constant close or equal to 1. A third option is plug-in methods that minimize the mean integrated squared error (Sheather and Jones, 1991), but these methods are subject to some weaknesses not shared by cross-validation (Li and Racine, 2007). A fourth option is least-squares cross-validation that is related to a predicted sum of squares (PRESS) statistic and obtains an optimal bandwidth asymptotically (Li and Racine, 2007). We choose \( h_1, \ldots, h_q \) to minimize the objective function

\[
CV_{lc}(h_1, \ldots, h_q) = \sum_{i=1}^{n} (Y_i - \hat{g}_{-i}(X_i))^2 M(X_i),
\]

where \( \hat{g}_{-i}(X_i) \), the “leave-one-out” estimator \( \hat{g}_{-i}(X_i) \) is estimated on all data except the \( i \)th observation and then evaluated at the \( i \)th observation. \( M(\cdot) \) is a weight function that avoids dividing by zero and any slow convergence of bandwidth selection caused by boundary effects. A fifth option is another cross-validation procedure that is an improved Akaike information criterion (AIC) approach (Hurvich et al., 1998) that can be used for any estimator that is a linear combination of \( Y_i \). For this dissertation, all the nonparametric estimators are sums of weighted \( Y_i \), e.g. \( \hat{g}(x) = \sum_{i=1}^{n} w_i Y_i \) where \( w_i = K \left( \frac{X_i - x}{h} \right) / \sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) \) for the local constant kernel estimator. The improved AIC method is to select a bandwidth that minimizes the criterion

\[
AIC_c = \ln(\hat{\sigma}^2) + \frac{1 + \text{tr}(H)/n}{1 - \{\text{tr}(H) + 2\}/n},
\]
where
\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \hat{g}(X_i)\}^2 = Y'(I - H)(I - H)Y/n, \]

and the matrix \( H \) has elements
\[ H_{ij} = \frac{\prod_{s=1}^{q} \frac{1}{h_s} k \left( \frac{X_{is} - X_{js}}{h_s} \right)}{\sum_{i=1}^{n} \prod_{s=1}^{q} \frac{1}{h_s} k \left( \frac{X_{is} - X_{js}}{h_s} \right)}. \]

The advantage of this improved AIC selector is that it performs well on finite samples (Hurvich et al., 1998) compared to other plug-in and generalized cross-validation methods (Craven and Wahba, 1979). Although no theoretical results for the improved AIC asymptotic optimality exist, simulations show that it outperforms least-squares cross-validation for relatively small samples (Li and Racine, 2004). We will use all of these methods in this dissertation, where the bandwidth selection method will be made clear for each estimator.

The question for all nonparametric estimators is: how do they compare to their parametric counterparts? The usual method is to first determine if the estimator is consistent \( \left( \hat{g}(x) \xrightarrow{P} g(x) \right) \) and if so, calculate how fast this estimator converges to the true data generating process as \( n \to \infty \). The local constant kernel regression estimator is a consistent estimator of \( g(x) \) with a rate of convergence \( O \left( n^{-2/(q+4)} \right) \) when using a second-order kernel function (\( \nu = 2 \)) (Li and Racine, 2007), while the optimal rate of convergence is \( O \left( n^{-1/2} \right) \) for correctly specified parametric estimators.

A useful property of local constant regression is its ability to smooth out independent or “irrelevant” explanatory variables in regression through the bandwidth selection method. This is attainable for the local constant kernel estimator as the bandwidth is selected using an appropriate data-driven method such as cross-validation. To demonstrate what smoothing out variables means in the regression setting, suppose that there are \( q = q_1 + q_2 \) regressors in a model, such that there are \( q_1 \) relevant and \( q_2 \) irrelevant variables. That is, let \( X_{q_1} \) be the \( q_1 \) columns of relevant regressors and \( X_{q_2} \) be the \( q_2 \) columns of irrelevant regressors for the design matrix \( X = [X_{q_1}, X_{q_2}] \). The assumption
of irrelevance for \( q_2 \) variables gives that \( X_{q_2} \) is independent of \( (Y, X_{q_1}) \). This means
\[
\hat{g}(X) = E[Y|X] = E[Y|X_{q_1}, X_{q_2}] = E[Y|X_{q_1}].
\]

Least-squares cross-validation has been shown to asymptotically smooth out these irrelevant regressions without having to remove them from the model (Li and Racine, 2007). Irrelevant regressor bandwidths, selected using cross-validation, will be very large and essentially be \( h_{\{q_2\}} \to \infty \). This is illustrated by
\[
\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i \prod_{s=1}^{q_1} k\left(\frac{x_s - X_{is}}{h_s}\right) \prod_{s=q_1+1}^{q} k\left(\frac{x_s - X_{is}}{h_s}\right)}{\sum_{i=1}^{n} \prod_{s=1}^{q_1} k\left(\frac{x_s - X_{is}}{h_s}\right) \prod_{s=q_1+1}^{q} k\left(\frac{x_s - X_{is}}{h_s}\right)},
\]
and \( h_s \to \infty \) for \( s > q_1 \) regressors. This gives
\[
\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i \prod_{s=1}^{q_1} k\left(\frac{x_s - X_{is}}{h_s}\right) \prod_{s=q_1+1}^{q} k(0)}{\sum_{i=1}^{n} \prod_{s=1}^{q_1} k\left(\frac{x_s - X_{is}}{h_s}\right) \prod_{s=q_1+1}^{q} k(0)},
\]
where the estimator no longer depends on the irrelevant variables.

The local constant estimator is derived from the following minimization of an objective function
\[
\min_a \sum_{j=1}^{n} (Y_j - a)^2 K\left(\frac{X_i - x}{h}\right),
\]
where \( a \equiv a(x) \) is a constant with respect to \( x \) and hence, the quantity that minimizes this equation is called the “local constant” kernel estimator. We extend this objective function to the local linear framework (Stone, 1977; Cleveland, 1979) by
\[
\min_{\{a,b\}} \sum_{j=1}^{n} (Y_j - a - (X - x)'b)^2 K\left(\frac{X_i - x}{h}\right).
\]
This minimization problem is rewritten as

$$\min_{\{a, b\}} (Y - \mathcal{X}\delta)'\mathcal{K}(x)(Y - \mathcal{X}\delta),$$

where \( \delta = \delta(x) = (a(x), b(x))' \), \( Y = (Y_1, Y_2, \ldots, Y_n)'_{n \times 1}, \mathcal{X} = (1, X_i - x)'_{n \times (1+q)}, \) and

$$\mathcal{K}(x) = \begin{bmatrix}
K \left( \frac{X_1 - x}{h} \right) & 0 & 0 & 0 \\
0 & K \left( \frac{X_2 - x}{h} \right) & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & K \left( \frac{X_n - x}{h} \right)
\end{bmatrix}.$$ 

The solution to this weighted least-squares minimization problem is

$$\hat{\delta}_h(x) = (\hat{a}, \hat{b})' = [\mathcal{X}'\mathcal{K}(x)\mathcal{X}]^{-1}\mathcal{X}'\mathcal{K}(x)Y,$$

$$= \left[ \sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right) \left( \begin{array}{c} 1 \\ X_j - x \end{array} \right) (1, (X_j - x)') \right]^{-1} \sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right) \left( \begin{array}{c} 1 \\ X_j - x \end{array} \right) Y_j.$$ 

The solutions for isotropic local linear regression yield \( \hat{a}(x) = \hat{g}(x) \), and \( \hat{b}(x) = \hat{g}^{(1)}(x) = \left( \frac{\partial g(x)}{\partial x_1}, \ldots, \frac{\partial g(x)}{\partial x_q} \right) \). Note that \( \hat{b} \) is not the derivative of the estimate of \( \hat{g}(x) \), but rather an estimate of the derivative of \( g(x) \). Also, notice that as \( h \to \infty \), the local constant estimator becomes constant for all \( x \) and the local linear estimator becomes the least-squares line—hence, “local linear” estimator. Therefore, these two estimators are viewed as the local versions of a global average and of linear regression. For this dissertation, all smoothers use the above nonparametric framework.

### 2.2 Gaussian and PDE-based anisotropic diffusion

In image processing, an image filter is any technique that changes one or more characteristics of an image to create a new “filtered” image. In this dissertation, we are interested in filters that remove undesirable noise in images. A popular method is Gaussian filtering, which is analogous to kernel smoothing. A Gaussian filter is in essence a local constant
regression, where the kernel is a mean-zero Gaussian distribution and the variance of the distribution acts as the smoothing parameter (Jain et al., 1995). The algorithm is a convolution between the Gaussian filter and the image (Nixon and Aguado, 2012) as

$$I(x, y, \sigma) = I(x, y, 0) \ast \frac{1}{2\pi\sigma^2}e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)},$$

where \(I(x, y, 0)\) is RGB channel values at (x,y) of the original image and \(I(x, y, \sigma)\) is the convoluted image. The smoothing method of local constant kernel regression is equivalent to Gaussian filtering, where \(h\) and \(\sigma\) are analogous in controlling the smoothing. This method is also equivalent to the simplified heatflow diffusion (Nixon and Aguado, 2012) of

$$\frac{\partial I(x, y, t)}{\partial t} = \nabla^2 I(x, y, t),$$  \hspace{1cm} (2.5)

where \(I(x, y, t)\) is the RGB value of a pixel at coordinates \((x, y)\) and time \(t\), \(\nabla\) is the spatial gradient operator, and the initial condition \(I(x, y, 0)\).

The area of image processing we are interested in is concerned with distinguishing boundaries or “edges” in an image, and smoothing between the boundaries. Anisotropic diffusion filtering, also called Perona-Malik diffusion (Perona and Malik, 1990; Black and Marimont, 1998) or anisotropic diffusion, is a partial differential equation (PDE) method for edge-preserve smoothing between boundaries. Anisotropic diffusion is an edge-preserving extension to (isotropic) Gaussian filtering. A physical interpretation of anisotropic diffusion is that it takes a convolution of the image and an isotropic kernel function, using an edge-stopping function to preserve boundaries between regions without supervision. Essentially, they are taking locally weighted averages around each pixel. This is modeled analogously by a more general heatflow diffusion PDE given as

$$\frac{\partial I(x, y, t)}{\partial t} = \nabla \cdot [f(\|\nabla I\|)\nabla I],$$  \hspace{1cm} (2.6)

where \(\nabla \cdot\) is the divergence operator, the edge-stopping function is \(f(\|\nabla I\|) \to 0\) as \(\|\nabla I\| \to \infty\), and initial condition \(I(x, y, 0)\) of the PDE is the original image. If \(f(\|\nabla I\|) :=\)
Chapter 2. Background theory

1, this becomes Gaussian filtering (Koenderink, 1984; Hummel, 1987). The imaginary time $t$ is for the diffusion of pixel values of a single image. Many possible choices exist for $f(\cdot)$, such as $f(x) = 1 / \left(1 + \frac{x^2}{k^2}\right)$ or $f(x) = \exp\left\{\frac{-x^2}{k^2}\right\}$, where these edge-stopping functions are the same at all imaginary times $t$. The quantity $k$ is analogous to a heat conduction coefficient or smoothing parameter $h$. The PDE in Equation (2.6) has no analytical solution for most initial and boundary conditions, but the behaviour of the system is approximated by a numerical solution. The anisotropic diffusion numerical software used in this dissertation (Pilny and Janacek, 2006) uses an algorithm based on a framework for image regularization (Tschumperle and Deriche, 2005). This anisotropic diffusion software is implemented into our own software for the purpose of processing images of a fire spread movie (Lanz-O’Brien and Braun, 2018). Anisotropic diffusion is compared to the edge-preserving kernel smoothing methods developed in this dissertation.

2.3 Data sharpening

Data sharpening is a useful method to allow data to satisfy a restriction such as unimodality in density estimation by perturbing or “sharpening” the data. We use data sharpening in this dissertation to estimate ignition and extinction times of pixels in a fire spread video in Section 5.1. Data sharpening was introduced to perturb data values to improve density estimation by reducing bias (Choi and Hall, 1999). Given a sample $X = \{X_j\}_{j=1}^n$, their methodology is to perturb the data using

$$\psi(X_j) = X_j + c \frac{\sum_{i=1}^n (X_i - X_j) k \left(\frac{x_i - X_j}{h}\right)}{\sum_i k \left(\frac{x_i - X_j}{h}\right)},$$

where $c > 0$. The rearranged $\psi(X_j) - X_j$ is the local constant kernel estimator in Equation (2.2). This shows that the data sharpening method transforms the original data $X_j$ into the perturbed data $\psi(X_j)$ using locally averaged information subject to scaling $c$. The subtraction of the kernel density estimator decreases the bias of the density
estimator—effectively increasing the rate of convergence—while preserving the kernel density estimator’s consistency. Data sharpening is also used to improve the performance of statistical methods by perturbing the data to minimize a distance metric with the original data subject to constraints (Braun and Hall, 2001). Any constraints on the data can be made, as long as there exists an analytic or numerically optimal solution subject to those constraints. For the practitioner, a constraint could be unimodality on a probability density estimate, where data sharpening perturbs the data to produce a unimodal kernel density estimate and the amount of nudging is minimized in the data sharpening algorithm.

The data sharpening methodology that we use to perturb the data minimizes a quadratic objective function subject to linear constraints, and utilizes quadratic programming to find an optimal solution. Let \( D(X_j, X_j^*) \) be a distance measure between the sample and the perturbed data \( X^* = \{ X^*_j \}_{j=1}^n \). The objective function for the minimization problem is
\[
D(X, X^*) = \sum_{j=1}^n D(X_j, X_j^*),
\]
where the distance metric used in this dissertation is the \( L_2 \)-norm, \( D(X_j, X_j^*) = \| X_j, X_j^* \|^2 \). If the \( L_2 \)-norm is used and the constraints imposed on the data are linear, then a numerical optimization algorithm for quadratic programming is used to find solutions to the minimization problem. The quadratic programming algorithm is to optimize a quadratic objective function subject to linear constraints (Braun and Murdoch, 2001, see ch. 7.5.7), i.e. to minimize the objective function
\[
\min_{\beta} \left\{ \frac{1}{2} \beta^T D \beta - d^T \beta \right\},
\]
subject to constraints \( A^T \beta \geq b \), where \( \beta \) is \( p \times 1 \) vector of unknowns, \( D \) is \( p \times p \) positive definite matrix, \( d \) is a \( p \times 1 \) vector, \( A \) is a \( p \times k \) matrix and \( b \) is a \( k \times 1 \) vector. The minima of the objective functions do not have analytic solutions, therefore numerical solutions are found using the \texttt{quadprog} (Turlach and Weingessel, 2013; R Core Team, 2018) package in \texttt{R}.
2.4 Fire spread modelling

There are many fire spread models in literature with a variety of physical and stochastic methods dedicated to estimating, analyzing, and simulating fire growth. A current state-of-the-art review of wildland fire spread models is found in the three following reviews and details can be found in references therein: physical and quasi-physical models (Sullivan, 2009a), empirical and quasi-empirical models (Sullivan, 2009b), and simulation and mathematical analogue models (Sullivan, 2009c). We explore a subset of these models that are used by provincial governments across Canada. We review two systems for modelling fuel types and fire behaviour under a variety of environmental conditions: the Canadian Forest Fire Behaviour Prediction System and the (American) BEHAVE system. We specifically limit the review to parts of each model relevant to fire growth. Then, we review the usage of fire spread rates in the Prometheus and the Boychuk fire spread simulation models.

2.4.1 The Canadian Forest Fire Behaviour Prediction System

The Canadian Forest Fire Behaviour Prediction (FBP) System was originally developed in an unpublished paper (by C.E. Van Wagner et al.) in the early 1980s for predicting fire behaviour characteristics in a range of fuel types (Fire Danger Group, Forestry Canada, 1992), and was issued for use in the field in 1985 (Lawson et al., 1998). The original paper included rates of spread for fourteen Canadian regional fuel types. Since the inception of FBP, the model has been expanded with additional fuel types, estimates of fuel consumption and fire intensity, and more complex models for fire growth and crown fires. FBP has a main component for forward fire rate of spread and secondary components for flank and back rate of spread, and fire spread distances. Understanding these spread rates and their variabilities from a statistical perspective is one of motivations of this dissertation. Next, we focus on the components of the FBP system that are relevant to fire spread rate calculations.
Chapter 2. Background theory

**FBP System slope adjusted rate of spread**

Note that we are closely following the notation of equations used by FBP (and later BEHAVE) with some minor changes to allow for consistency in the dissertation. Italicized acronyms in this section are considered to be non-random variables.

FBP’s fire spread rate component starts by determining the initial spread index ($ISI$). The $ISI$ is calculated from daily, hourly, or forecasted weather (temperature, wind speed and direction, relative humidity, precipitation), and fine fuel moisture codes ($FFMC$) from the Fire Weather Index (FWI) system that describes the potential for ignition and flammability of fine fuels. The $ISI$ is calculated using the equation

$$ISI = 0.208 \times p(W) \times q(F),$$  \hspace{1cm} (2.7)

where

$$p(W) = \exp\{0.05039 \times |\overrightarrow{WSV}|\},$$

$$q(F) = q(FFMC) = 91.9 \times \exp\{-0.1386 \times m\} \times \left[1 + \frac{m^{0.31}}{4.93 \times 10^7}\right],$$

$$m = \frac{147.2 \times (101 - FFMC)}{59.5 + FFMC},$$

and $\overrightarrow{WSV}$ is the wind speed vector. The $ISI$ and rate of spread calculations are a model for when the spread rate has reached equilibrium. The initial rate of spread ($RSI$) for the majority of coniferous fuels is calculated by

$$RSI = \alpha \left[1 - e^{-\beta \times ISI}\right]^\gamma,$$ \hspace{1cm} (2.8)

where $\alpha, \beta, \gamma > 0$ are parameters associated with fuel type and $RSI$ has units of metres per minute. The functional form of $RSI$ was chosen for its “S” shape and the parameters were estimated from experimental fire and wildfire data for each fuel type. Other functional forms for the remaining coniferous, deciduous, grass, and slash fuel types are modified forms of Equation (2.8).

Slope is included into the $RSI$ by converting it to an effective wind speed and adding
it to the observed wind speed in the ISI calculation. To perform this, first the percentage ground slope \((GS)\) is calculated by

\[
GS = 100 \times \frac{\text{Elevation rise}}{\text{Horizontal ground distance}} = 100 \times \tan(\theta),
\]

where the first equality is for map calculations (noting that measurements of the slope are always in the direction of the gradient) and the second is for on-site calculations where \(\theta\) is the number of degrees for the angle above the horizontal. The percentage ground slope is converted into a spread factor \((SF)\) by

\[
SF = \exp\left\{ 3.533 \times \left( \frac{GS}{100} \right)^{1.2} \right\}, \quad (2.9)
\]

where \(GS\) should not exceed 60% as \(SF\) increases rapidly and has not been shown to be empirically valid. The \(SF\) value is adjusted to a zero-wind rate of spread with slope adjustment \((RSF)\) by

\[
RSF = RSZ \times SF,
\]

where \(RSZ\) is the zero-wind rate of spread without the slope adjustment. The \(RSF\) is used to update the ISI to the initial spread factor \((ISF)\) given by

\[
ISF = \frac{1}{-\beta} \ln \left[ 1 - \left( \frac{RSF}{\alpha} \right)^{1/\gamma} \right].
\]

The \(ISF\) is transformed into the wind speed equivalent \((WSE)\), where the magnitude is given by

\[
|\vec{WSE}| = \frac{1}{0.05039} \ln \left[ \frac{ISF}{0.208 \times q(F)} \right],
\]

and the direction of \(WSE\) is given by the angle of the gradient of the slope direction relative to the \(y\)-direction \((SAZ)\). The resultant wind speed vector \((\vec{WSV})\) includes both the observed wind speed \(\vec{WS} = (WS_x, WS_y)\) and the effective wind speed \(\vec{WSE} = \)
(\(WSE_x, WSE_y\)). The \(\overrightarrow{WSV}\) is calculated by

\[
\overrightarrow{WSV} = (W_S \sin(WAZ) + WSE_x \sin(SAZ), W_S \cos(WAZ) + WSE_y \cos(SAZ)),
\]

where \(WAZ\) is the observed wind speed angle relative to the \(y\)-direction. The \(ISI\) is re-calculated with the new \(\overrightarrow{WSV}\) using Equation (2.7) and the head rate of spread that includes slope and wind (\(ROS\)) is calculated using the \(RSI\) Equation (2.8). The back rate of spread (\(BROS\)) is calculated very similarly, except where \(p(W)\) is replaced with \(Bp(W)\) given by

\[
Bp(W) = p(W)^{-1} = \exp\{-0.05039 \times WSV\},
\]

\(\Rightarrow\) \(BISI = 0.208 \times p(W) \times Bq(F)\),

\(\Rightarrow\) \(BROS = \alpha \left[1 - e^{-\beta \times BISI}\right]^\gamma\).

Flank rate of spread (\(FROS\)) is calculated from these two quantities by

\[
FROS = \frac{1}{2} \frac{ROS + BROS}{LB}.
\]

### 2.4.2 The BEHAVE system

The BEHAVE system is a series of two subsystems, named FUEL and BURN, that allow fire managers to construct site-specific fuel models and predict fire behaviour for operations, respectively. We are interested in BURN, as it describes fire spread. The BURN system is based on the spread rate formulas originally derived by Frandsen (1971).

Suppose there is a fire front along a homogeneous fuel bed and there is a unit volume of fuel with the \(x - y\) plane on the surface of the earth and \(z\) distance from the earth surface. Frandsen (1971) applied the conservation of energy to the heat flux into a unit volume of homogeneous fuel bed from the advancing fire front. The theoretical head rate of spread for the BEHAVE system (\(ROS_{BEHAVE}\)) in the \(x\)-direction is given by

\[
ROS_{BEHAVE} = \frac{I_{xig} + \int_{-\infty}^{0} \left(\frac{\partial I}{\partial z}\right)_{\xi C} dx}{\rho_{be} Q_{ig}}.
\]
The $x$-direction is along the horizontal and $z$ is the vertical direction. The fuel-fire interface is fixed at $x = 0$ and the unit volume of fuel moves into the interface. The parameters in the equation are:

\[ I_{xig} = \text{horizontal heat flux into a unit volume of fuel, i.e. horizontal heat intensity}, \]
\[ \left( \frac{\partial I}{\partial z} \right)_{zc} = \text{the change in vertical heat intensity at height } z_c, \]
\[ \rho_{be} = \text{fuel density in the unit volume of fuel at fuel-fire interface}, \]
\[ Q_{ig} = \text{heat energy required to ignite one unit weight of fuel}. \]

The rate equation in (2.10) shows that the vertical heat flux is an additive radiation component to the horizontal heat flux. Rothermel (1972) points out that the sum in the numerator is a special case of the propagating heat flux $I_p = I_{xig} + \int_{-\infty}^{0} \left( \frac{\partial I}{\partial z} \right)_{zc} dx$. To generalize, Rothermel proposes that there is a no-wind, no-slope heat flux $(I_p)_o$ given by

\[ (I_p)_o = R_o \rho_{be} Q_{ig}, \]

where $R_o$ is the no-wind, no-slope rate of spread–equivalent to $RSI$. Rothermel posits that the base intensity is some unknown function of the fire intensity of the fire front $(I_R)$ given by $(I_p)_o = f(I_R)$ such that

\[ I_R = \frac{dw}{dt} h, \]

where $\frac{dw}{dt}$ is the mass loss rate of the fire front and $h$ is the potential energy per weight of the fuel. Finally, he posits that the effects of wind and slope are additive such that Equation (2.10) is approximated by

\[ ROS_{BEHAVE} = \frac{(I_p)_o (1 + \phi_{wind} + \phi_{slope})}{\rho_{be} Q_{ig}} = R_o (1 + \phi_{wind} + \phi_{slope}), \quad (2.11) \]

where $\phi_{wind}$ and $\phi_{slope}$ are the additional effects of wind and slope on the flux. In this BURN model, slope and wind are seen as additive and scaled to the base rate of spread.
Rothermel explicitly states that, for expediency, rate of spread due to wind and slope are assumed to have no interaction, such that in zero wind or zero slope you have $\phi_{\text{wind}} = 0$ or $\phi_{\text{slopes}} = 0$, respectively. Equation (2.11) with zero wind was rearranged for the slope coefficient (Rothermel, 1972) to be

$$\phi_{\text{slope}} = \frac{ROS_{\text{BEHAVE}} - R_o}{R_o} = 5.2757\beta^{-0.3} \tan^2(\theta),$$

where $\theta$ is the angle of the slope (0 to 90 deg) and $\beta$ is the fuel packing ratio, that is the fraction of the packing made up of fuel. Also, notice that the slope coefficient is the contribution to the rate of spread by slope that is normalized by the base rate of spread. Similarly, Equation (2.11) with zero slope was rearranged for the wind coefficient (Rothermel, 1972) to be

$$\phi_{\text{wind}} = \frac{ROS_{\text{BEHAVE}} - R_o}{R_o} = C v^B \left( \frac{\beta}{\beta_{op}} \right)^{-E},$$

where

\begin{align*}
v & = \text{wind velocity,} \\
C & = 7.47 \exp \left\{ -0.133\sigma^{0.55} \right\}, \\
B & = 0.02526\sigma^{0.54}, \\
E & = 0.715 \exp \left\{ -3.59 \times 10^{-4}\sigma \right\}, \\
\sigma & = \text{fuel particle surface-area-to-volume ratio}, \\
\beta_{op} & = 3.348\sigma^{-0.8189}, \text{ the optimal } \beta \text{ for } \sigma.
\end{align*}

### 2.4.3 Prometheus fire growth model

The Prometheus fire growth model is an elliptical fire growth model, derived from Huygen’s principle of wave propagation (Wagner, 1969; Tymstra, C. (Project Leader), 2005). A fire starts as a circle, and then propagation is simulated as an envelope of ellipses on the boundary of the circle. Figure 2.1 shows one of the ellipses of the Prometheus where, from an ignition point, fire moves in the direction of WSE with a forward rate of spread.
(ROS), back rate of spread (BROS) and flank rate of spread (FROS). The dimensional parameters of this ellipse \((a, b, c)\) are determined through

\[
\begin{align*}
    a &= \frac{[(\text{ROS} + \text{BROS}) \times t]}{2}, \\
    b &= \text{FROS} \times t, \\
    c &= \frac{(\text{ROS} - \text{BROS}) \times t}{2}.
\end{align*}
\]

Spread rates in Equation (2.12) are calculated using the FBP and FWI systems. At each time step of a Prometheus simulation, the simulation algorithm (i) selects new vertices around the perimeter, (ii) calculates the ellipse at each vertex according to Equation (2.12) and the length of the time step, and (iii) draws the new fire perimeter as a tangential envelope of the ellipses. Prometheus fire simulations are calculated using deterministic quantities of the FBP system, which implies the system assumes that using the same initial conditions always results in the same fire spread simulation. To address this issue, Monte Carlo methods have been used to introduce uncertainty in the ROS inputs, which may not be representing true prediction uncertainty (Garcia et al., 2008). Another model is Dionysus, a stochastic extension of the Prometheus fire simulation model (Han and
Braun, 2013) named after the son of Zeus associated with chaos and disorder. Dionysus adds randomness to the Prometheus model by introducing uncertainty into the RSI and the model’s primary function is to calculate burn probability regions. Extending this approach to sloped terrain, and properly accounting for back and flank rate of spread is a priority, and this dissertation can provide some of the tools to move this project forward.

### 2.4.4 Boychuk fire growth model

Another common structure for fire spread models is to split the fuel’s spatial region into cells. A grid-based cellular fire spread model was designed by Boychuk et al. (2007) to incorporate the local randomness of fire spread. The Boychuk model is an interacting particle system where a region is drawn as a $n \times m$ grid of square cells with three states: fuel, burning, and burnt-out (FBO). Under this model, a cell $(i,j)$ interacts with its four nearest neighbours $(i, j + 1), (i + 1, j), (i, j - 1), (i - 1, j)$. A fuel cell transitions into a burning cell if one of its nearest neighbours is a burning cell, and a burning cell transitions into a burnt-out cell, both at exponential rates.

Suppose you have a cell $(i, j)$ in a burning cells with four nearest neighbours are fuel cells. Under the simplest conditions, the time until each of these neighbouring cells transition from fuel to burning (ignition) are exponential random variables $T_{(i, j + 1)}, T_{(i + 1, j)}, T_{(i, j - 1)}, T_{(i - 1, j)}$ with ignition rate $\lambda$. The ignition rates depend on neighbouring cells states, weather, and topography. The burning cell at $(i, j)$ also has a time from ignition to extinction events known as the residency time $R_{(i, j)}$. The residency time is also considered to be exponentially distributed with a constant rate $\mu$ that depends on weather and topography, but does not depend on neighbouring cells.

The Boychuk model is a type of a broader class of cellular automaton models, where these models need cells with defined states, interactions, and local rules. Another cellular automaton model, used to study fire in the Mkuze game reserve (Berjak and Hearne, 2014), bases their cell interaction on fuel heterogeneity, wind speed and direction, and topography, and a base rate of spread from the BEHAVE system that accounts for fuel bulk density, propagating flux ratio, reaction intensity, pre-ignition heat, and a
dimensionless effective heating number (Rothermel, 1972; Burgan and Rothermel, 1984).

2.5 Motivation for studying fire growth

The estimation of fire spread rates is a topic of discussion when using a fire spread simulator. Regardless of which simulator is employed, a spread rate is needed for simulation. A study of the accuracy and precision of spread rate estimates (Cruz and Alexander, 2013) shows that estimates for an experimental grasslands fire in West Australia were highly variable over short periods of time (1-3 min), but consistent over long periods (15-30 min). They compare 49 studies with recorded rate of spread observations and predictions, where some studies use more than one model to predict fire spread. About half of the studies (25 out of 49) have 51% to 75% mean absolute percentage error of estimating the rate of spread, with no bias towards under- or over-predicting the rate of spread. They also show that the range of experimental burns versus wildfire data sets are 26% to 35% and 46% to 52% mean absolute percent error, respectively, which demonstrates the uncertainty of the environmental conditions effect on the ability to estimate fire spread rates. These results demonstrate the prevalent underestimation of spread rates for fire models when compared to reality.

The set of equations for calculating rate of spread in the FBP and BEHAVE systems contain a few subtle details about the mechanisms of fire spread. Wind and slope are environmental factors that can be expressed as similar “forces” that drive fire spread. This is analogous to how gravitational and electric forces both obey the inverse square law. In the last century, the interaction of these two fundamental forces was discovered in gravitational lensing (Bartelmann and Schneider, 2001), which was a direct result of Einstein’s theory of general relativity (Einstein, 1916). FBP and BEHAVE do not include the interaction effects of wind and slope in their models, where these fire growth systems assume the functional form of the rate of spread. Also, FBP and BEHAVE do not include uncertainty in spread rate predictions. The Dionysus model introduces uncertainty into the FWI inputs, but they also assume the shape of the distribution for fire spread rates and that shape is constant over FWI inputs. One of the main goals of this dissertation is
to quantify fire spread rates and their variability from fire smoldering experiments, and to investigate the effects of environmental conditions on fire spread rate distribution.
Chapter 3

Anisotropic smoothing

In this chapter, we introduce an anisotropic smoothing framework for kernel estimators that automatically detects and smooths between change-points of the underlying data generating process of a regression model. We begin by reviewing previous work in detecting and smoothing change-point data in the contexts of regression and image processing. The local constant and local linear estimators from Section 2.1 smooth isotropically in the domain. We propose our anisotropic estimators as extensions to the local constant and local linear regression estimators such that the smoothing is no longer isotropic in the domain. We show weak consistency for the anisotropic local constant and local linear estimators, and we calculate the rates of convergence of the anisotropic local constant and local linear estimators to be $O\left(n^{-\frac{1}{q+2}}\right)$, and compare to isotropic kernel methods. We discuss iterative estimation methods and bandwidth selection methods for anisotropic smoothing, and show that the bandwidth selection methods may be inappropriate for change-point regression function estimation. We investigate the isotropic and anisotropic local constant estimator through one- and two-dimensional simulation studies, and compare their performance. The two-dimensional data is designed to be similar to the fire spread data in Chapter 4. We compare our anisotropic local constant and local linear estimators to PDE-based anisotropic diffusion, where we show our method better captures the change-points for the two-dimensional data.

The two biggest challenges when analyzing change-point data are: (1) correctly detecting change-points in the regression function, and (2) smoothing change-point regres-
sion functions. Anisotropic smoothing implicitly detects change-points while smoothing the regions between them simultaneously. Our focus is on smoothing between change-points, so further discussion on detection in this section is extraneous. We note that our methods could be used for change-point detection by studying the derivative of the estimate, such as in Section 5.1.2. Henceforth, we assume that change-points are detectable, and we estimate the regression function in the presence of change-points.

3.1 Regression with change-point data

Nonparametric kernel regression is a method of estimating a regression function without specifying the functional form of the relationship between regressors and outcome variables. Data with an abrupt change in the regression function is known as change-point data, and an example is shown in Figure 3.1. Kernel regression methods use local (about the regressors) information (of the outcome) to estimate the regression function. Kernel estimators are known to oversmooth in regions of data where there is an abrupt change in the regression function, such as when \( g'(x)/g(x) \) is very large (Fan and Gijbels, 1996) and information local in the \( X \) direction is not local in the \( Y \)-direction. Popular kernel estimation methods, such as the Nadaraya-Watson kernel estimator, are not well-equipped to differentiate between regions of the data that are separated by a change-point. The Nadaraya-Watson kernel estimator inappropriately uses data from both sides of a change-point when estimating the regression function and tends to oversmooth the regression function around change-points.

To review change-point smoothing methods, assume that we have a set of explanatory and response continuous data \( \{(X_i, Y_i)\}_{i=1}^{n} \). Our regression model is \( Y_i = g(X_i) + \epsilon_i \) where \( g \) is a regression function with unknown form. We note that many of the models presented here are appropriate for one or two dimensional explanatory data, as they are predicated on the definition of a change-point in those dimensions for the context of their research.

There exist many methods to detect and preserve jumps or “edges” in regression and image processing. One method is the split linear fit that combines left, right, and center linear fits of to make smooth functions (McDonald and Owen, 1986; Hall and
Figure 3.1: An example of change-point data, where there is an abrupt change in the regression function at $X = 1.5$.

Titterington, 1992). This work leads to further study of asymmetric kernel functions in change-point regression that extends to change-points in the derivative of the regression function (Muller, 1992). Jump locations are estimated from the derivative of the local constant estimator (Gijbels et al., 1999) with appropriate bandwidth selection (Gijbels and Goderniaux, 2004) and smoothing between jump locations. Another method is to estimate the regression relationship with a jump regression function (Qiu et al., 1991) of the form

$$g(x) = f(x) + \sum_{i=1}^{p} d_i 1_{(s_i, s_{i+1})}(x),$$  \hspace{1cm} (3.1)

where $f(x)$ is continuous, $\{s_i\}_{i=1,2,...,p}$ are the known jump locations, $d_i$ are the scalar
jump magnitudes, and \(1_{(s_i, s_{i+1})}(x)\) are indicator functions. This function is further generalized to estimate the derivatives of \(g(x)\) (Qiu and Yandell, 1998), and non-constant forms of jumps \(d_i = d_i(x)\). The jump regression method is extended to local kernel smoothing methods (Fan and Gijbels, 1996; Qiu, 2003) by extending the local linear minimization problem to a piece-wise minimization problem of

\[
\min_{a_{l,0}, a_{l,1}; a_{r,0}, a_{r,1}} \sum_{i=1}^{n} \{y_i - [a_{l,0} + a_{l,1}(x_i - x)] \cdots \cdots - [(a_{r,0} - a_{l,0})I(x_i - x) + (a_{r,1} - a_{l,1})(x_i - x)I(x_i - x)]}^2 K \left(\frac{x_i - x}{h}\right), \tag{3.2}
\]

where the solution \(\{a_{l,0}, a_{l,1}; a_{r,0}, a_{r,1}\}\) can be found by realizing that Equation (3.2) can be treated as two local linear minimization problems (Qiu, 2003). The piece-wise linear minimization method is used for surface estimation with known or unknown number of jumps with jump location detection (Qiu, 2007).

Jump regression methods fall into two categories. Generally, the first collection of algorithms is to detect jumps, define subintervals, and then smooth between those intervals. The second collection of algorithms removes the effects of the jumps by subtracting \(\sum_{i=1}^{p} d_i 1_{(s_i, s_{i+1})}(x)\) in Equation (3.1) and then smoothing to obtain a continuous regression function estimate. For extensive detail on jump regression analysis methods, see Qiu (2005) and references therein.

Jump regression depends on the ability to detect change-points before smoothing. A more robust method to estimate a regression relationship with change-points is a family of estimators that can be viewed as subtle versions of each other. This unified framework of estimators has four types (Mrázek et al., 2006): \(M\)-estimators, local \(M\)-smoothers, Bayesian / regularization / diffusion filtering, and bilateral filtering. Recall our regression framework \(Y_i = g(X_i) + \epsilon_i\) for paired data \(\{(X_i, Y_i)\}_{i=1}^{n}\). An \(M\)-estimator is a solution to the minimization problem (Huber et al., 1964; Hampel et al., 2011) written as

\[
\min_{u} \sum_{i=1}^{n} \Psi(u - Y_i), \tag{3.3}
\]
where $u$ is a constant and $\Psi(\cdot)$ is usually a norm. For example, if $\Psi(x) = x^2$ then minimum $\hat{u} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ is the least-squares solution. If $\Psi$ is the negative log-likelihood under a hypothesized probability model, then $\hat{u}$ is the maximum likelihood estimate—hence the “$M$” in $M$-estimator. If $u$ is also a function of the index $i$, then the $u_i$ that minimizes Equation (3.3) is $u_i = Y_i$, which is not desirable. Instead, the $M$-estimator is extended in two ways for change-point data. One extension is the local $M$-smoother given by

$$\min_{\{u_1, \ldots, u_n\}} \sum_{i=1}^{n} \sum_{j \in B_i} \Psi(u_i - Y_j) w(X_i - X_j), \quad (3.4)$$

where $w$ is some weighting function (e.g. a kernel function) and $B_i$ is a set of neighbouring points that are local in $X_i$. Kernel smoothing can be viewed as a subcategory of the $M$-smoothing family. In fact, the anisotropic diffusion method is a specific case for a Bayesian/regularization framework approach to image smoothing, which is closely related to local $M$-smoothers (Mrázek et al., 2006). The other extension of $M$-estimators is the bilateral filter (Tomasi and Manduchi, 1998) given by

$$\min_{\{u_1, \ldots, u_n\}} \sum_{i=1}^{n} \sum_{j \in \mathcal{B}} \Psi(u_i - u_j) w(X_i - X_j).$$

$M$-smoothers and bilateral filters that have no analytic solutions can be approximated by a $W$-estimator, which is a variant of a $M$-estimator (Hampel et al., 2011). A $W$-estimator is an iterative procedure starting at $u_i^0$ as the median value of the $Y_i$’s and

$$u_i^{k+1} = \frac{\sum_{j \in \mathcal{B}(i)} g(u_i^k - Y_j) w(X_i - X_j) Y_i}{\sum_{j \in \mathcal{B}(i)} g(u_i^k - Y_j) w(X_i - X_j)} \quad (3.5)$$

This iterative procedure is for local $M$-smoothers. For bilateral filters, replace $Y_j$ with $u_j^k$ in Equation (3.5). The $W$-estimator in Equation (3.5) converges to a local minimum of the local $M$-smoother in Equation (3.4) as $k \to \infty$ (Hampel et al., 2011). Diffusion filtering is similar to these methods, and is discussed in Section 2.2 as an image processing method. In fact, the local constant kernel estimator—called “adaptive smoothing” in the
image processing literature—is similar to anisotropic diffusion and bilateral filters, and has been used to show a connection between those two filters (Barash, 2002). The advantage of bilateral filtering is that optimized versions exist, which do not require a smoothing parameter selection when using a median filter (Weiss, 2006).

Change-point data analysis from the kernel regression perspective currently has limitations. $M$-estimators as edge preserving estimators have been studied for local constant (Chu et al., 1998), local linear (Rue et al., 2002; Lin et al., 2009), and local polynomial (Hwang, 2004) regression. An image processor detects edges and preserves them (Lopez-Molina et al., 2013), but does not always smooth a function with an estimator that converges to the data generating process in probability. Our goal is to investigate estimators, inspired by the above family of estimators, that we have named “anisotropic” smoothers. Our estimators do not isotropically smooth in the domain by including a range kernel—a kernel function that incorporates the range of the regression function $g$. Anisotropic smoothing has potential for local polynomial estimation of a multivariate regression function with multivariate regressors, using a data-driven method that implicitly detects change-points and naturally smooths change-point regression functions. The anisotropic smoothing methods described herein provide new knowledge and insight into the consistency of estimating a regression function with change-points. We investigate nonlinear regression functions with unknown number and magnitude of change-points. The path from imagination to realization of a consistent anisotropic kernel estimator for change-point regression functions starts with this dissertation, where we make the first steps by studying anisotropic local constant and local linear kernel estimators for multivariate regressors and a univariate outcome.

### 3.2 Nonparametric anisotropic kernel regression

Before we develop the anisotropic estimator for change-point data, we explicitly write out the nonparametric regression framework. Consider the $q$-D multivariate regressors $X_i = (X_{1i}, X_{2i}, \ldots, X_{qi})$, $q \in \mathbb{N}$, univariate outcome $Y_i$, $i \in \{1, \ldots, n\}$, and the regression
model

\[ Y_i = g(X_i) + \epsilon_i, \]  

(3.6)

where \( g(\cdot) \) is a possibly nonlinear smooth function. The noise \( \epsilon_i \) is assumed to have mean 0, variance \( \sigma^2 \), and \( \text{Cov}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j \). The minimization problem for the isotropic local constant kernel estimator is

\[
\min_a \sum_{j=1}^{n} (Y_j - a)^2 K\left( \frac{X_i - x}{h} \right),
\]  

(3.7)

which could be characterized as a local M-smoother minimization problem in Equation (3.4). Without loss of generality, we assume that the kernel functions are order \( \nu = 2 \) from Equation (2.4), i.e. the kernel function satisfies the properties:

\[
(i) \quad \sup_{-\infty < x < \infty} |k(x)| < \infty, \int_{-\infty}^{\infty} |k(x)| \, dx < \infty, \lim_{x \to \infty} xk(x) \, dx = 0,
\]

\[
(ii) \quad \int_{-\infty}^{\infty} k(x) \, dx = 1,
\]

\[
(iii) \quad k(x) = k(-x) \Rightarrow \int_{-\infty}^{\infty} xk(x) \, dx = 0.
\]

The solution to this minimization problem in (3.7) is the local constant estimator

\[
\tilde{g}(x) \equiv \tilde{a}(x) = \frac{\sum_{i=1}^{n} Y_i K\left( \frac{X_i - x}{h} \right)}{\sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right)}.
\]  

(3.8)

The local constant estimator is a locally weighted average of \( Y_i \)'s using \( \tilde{g}(x) = \sum_{i=1}^{n} Y_i w_i \), with weights \( K\left( \frac{X_i - x}{h} \right) / \sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right) \), to estimate the regression function \( g(x) \). The order of the kernel function affects bias calculations for the local constant estimator, and thus rate of convergence calculations. However, nonnegative kernels do not exist for orders \( \nu > 2 \) (Li and Racine, 2007). Using a second order kernel, the local constant estimator was shown to have a consistent rate of convergence to \( g(x) \) of \( O\left( n^{-2/(q+4)} \right) \) (Li and Racine, 2007). For more on isotropic kernel regression methods and asymptotics, see Section 2.1.
3.2.1 Anisotropic local constant estimator

We develop the anisotropic nonparametric regression estimator, and demonstrate its potential for change-point regression estimation. The development of the anisotropic estimator starts with a modification to the minimization problem in Equation (3.7). Consider the minimization problem

\[
\min_b \sum_{j=1}^n (Y_j - b)^2 K \left( \frac{X_j - x}{h} \right) k \left( \frac{Y_j - y}{h_{q+1}} \right),
\]

and solution

\[
\hat{b}(x, y) = \frac{\sum_{i=1}^n Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{Y_i - y}{h_{q+1}} \right)}{\sum_{i=1}^n K \left( \frac{X_i - x}{h} \right) k \left( \frac{Y_i - g(x)}{h_{q+1}} \right)}.
\]

The kernel function \( k \left( \frac{Y_i - y}{h_{q+1}} \right) \) is known in the image processing literature as a range (or tonal) kernel (Mrázek et al., 2006). Range kernels allow for the range of outcome \( Y_i \) to be included in the kernel weighting procedure. Points more local in both the domain and range of the regression function contribute more to the smoothing of the regression function.

Equation (3.9) is almost a local \( M \)-smoother minimization problem, but not quite. If we set \( y := b \) then this would be an \( M \)-estimator. This equation is similar to the bilateral filter from image processing, but our goal is to study it from a regression perspective. We want to understand the entire function over all \( x \), and not just estimate at the observed data \( X = x \). We want the regression function estimate to not depend on \( y \), since we are interested in predicting \( Y | X = x \) and to develop a consistent change-point regression estimator. We propose replacing \( y \) with \( g(x) \), the true underlying regression function. Consider rewriting the estimator in (3.10) as

\[
\hat{b}(x) = \frac{\sum_{i=1}^n Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{Y_i - g(x)}{h_{q+1}} \right)}{\sum_{i=1}^n K \left( \frac{X_i - x}{h} \right) k \left( \frac{Y_i - g(x)}{h_{q+1}} \right)}.
\]

The range kernel no longer depends on our outcome variable \( y \), retains the anisotropic
filter through the range kernel function, and we write our estimator as a function without $y$. However, the estimator contains $Y_i$ embedded inside the kernel function and the error $\epsilon_i$ for the regression model in Equation (3.6) becomes recursively embedded in the estimator. To avoid this issue, we propose replacing $Y_i$ with $E[Y_i|X_i] = g(X_i)$. Therefore, the anisotropic local constant estimator is

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i-x}{h}\right) k\left(\frac{g(X_i)-g(x)}{h_{q+1}}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i-x}{h}\right) k\left(\frac{g(X_i)-g(x)}{h_{q+1}}\right)},$$

where it can be viewed as an anisotropic extension of the local constant kernel estimator. The intuition behind the anisotropic local constant estimator is that we are locally averaging the regression function over $x$ and the regression function itself $g(x)$. Data that are farther away in $x$ or $g(x)$ from $X_i$ are given lower weights than more local points. If there is a large jump in the data, points local in $x$ but distant in $y$ will contribute less and we are able to smooth across jump regions. This estimator is notably similar to the nonparametric estimator for the conditional cumulative distribution function (Li and Racine, 2008), where the conditional cumulative distribution function estimator is considered to be more robust for estimating regression functions, particularly in the presence of censoring (Li and Racine, 2007, ch. 6).

**Weak consistency of the anisotropic local constant estimator**

We show the weak consistency of the anisotropic local constant estimator in Equation (3.11). A list of assumptions for this proof are:

T1 Without loss of generality, we select a second-order kernel function ($\nu = 2$) from Equation (2.4). This allows us to compare rates of convergence with the isotropic kernel estimator that using same order.

T2 The regressors $X_i$ are fixed and not random. For the data in Chapter 4, it is natural to assume pixel locations are fixed in an image. This is also a typical assumption for kernel estimator consistency derivations in the literature (Ruppert and Wand, 1994).
T3 Bandwidths are of similar orders of magnitude. This ensures that \( h_1, h_2, \ldots, h_q \to 0 \) at the same rate. We also assume \( h_{q+1} \to 0 \) at a faster rate than any of \( h_1, h_2, \ldots, h_q \).

**Theorem 1** Under the assumptions T1-T3 above, and the assumptions that \( x \) is an interior point and \( g(x) \) is at least three times differentiable (except on a set \( D \) that forms a set of points (univariate \( x \)) or curves (multivariate \( x \)) such that \( \mathbb{P}(D) = 0 \)), then as \( n \to \infty, h_j \to 0, n_j h_j \to \infty, \forall j \in \{1, 2, \ldots, q, q+1\} \), we have, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} \mathbb{P}(|\hat{g}(x) - g(x)| < \epsilon) = 1,
\]

where \( n = n_1 \times n_2 \times \ldots \times n_q = n_{q+1} \).

**Proof.** To show weak consistency, we equivalently show that

\[
\lim_{n \to \infty} \text{Var}(\hat{g}(x)) = 0, \quad \lim_{n \to \infty} \mathbb{E}[(\hat{g}(x) - g(x))] = 0.
\]

First, consider rewriting the estimator in (3.11) as

\[
\hat{g}(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i)}{h_{q+1}} \right),
\]

where \( Y_i \) is univariate and \( X_i \) is \((q \times 1)\) vector. Substituting that \( Y_i = g(X_i) + \epsilon_i \) yields

\[
\hat{g}(x) = \frac{1}{n} \sum_{i=1}^{n} (g(X_i) + \epsilon_i) K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i)}{h_{q+1}} \right).
\]

Taking the expectation of the above equation, noting that \( X_i \)'s are not random and the assumption that \( \mathbb{E}[\epsilon_i] = 0 \), gives

\[
\mathbb{E}[\hat{g}(x)] = \frac{1}{n} \sum_{i=1}^{n} g(X_i) K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i) - g(x)}{h_{q+1}} \right).
\]

Therefore, if we further assume that the data is equally spaced by \( \frac{1}{n_j} \), then the nu-
Chapter 3. Anisotropic smoothing

The numerator and denominator are Riemann sums where \( \Delta x = \Delta x_1 \Delta x_2 \ldots \Delta x_q = (X_{1i} - X_{1(i-1)}) (X_{2(i-1)} - X_{2(i-1)}) \ldots (X_{ qi} - X_{ q(i-1)}) = \frac{1}{n} \). As each \( n_i \) becomes large, we approximate these Riemann sums as integrals in

\[
E[\hat{g}(x)] \approx \frac{\int g(w)K\left(\frac{w-x}{h}\right)k\left(\frac{g(w)-g(x)}{h_{q+1}}\right)dw}{\int K\left(\frac{w-x}{h}\right)k\left(\frac{g(w)-g(x)}{h_{q+1}}\right)dw},
\]

where \( w = (w_1, w_2, \ldots, w_q) \), \( dw = dw_1 dw_2 \ldots dw_q \). Using the substitution \( z_i = \frac{w_i-x_i}{h_i} \), \( dz_i = \frac{1}{h_i}dw_i \), \( z = (z_1, z_2, \ldots, z_q) \), \( dz = dz_1 dz_2 \ldots dz_q \), \( hdz = h_1 dz_1 h_2 dz_2 \ldots h_q dz_q \) and \( zh = (z_1 h_1, z_2 h_2, \ldots, z_q h_q) \) gives

\[
E[\hat{g}(x)] \approx \frac{\int g(zh+x)K(z)k\left(\frac{g(zh+x)-g(x)}{h_{q+1}}\right)hdz}{\int K(z)k\left(\frac{g(zh+x)-g(x)}{h_{q+1}}\right)hdz}. \tag{3.12}
\]

A Taylor series expansion of \( g(zh+x) \) about \( z = 0 \) is

\[
g(zh+x) = g(x) + zh \cdot \nabla g(x) + \frac{1}{2} (zh)' Hzh + \ldots, \tag{3.13}
\]

where

\[
g_{z_i}(x) = \frac{\partial g(z)}{\partial z_i}\bigg|_{z=x}, \quad g_{z_i z_j}(x) = \frac{\partial^2 g(z)}{\partial z_i \partial z_j}\bigg|_{z=x}, \quad \nabla g(x) = (g_{z_1}(x), g_{z_2}(x), \ldots, g_{z_q}(x)),
\]

\[
zh \cdot \nabla g(x) = \sum_{j=1}^{q} z_j h_j g_{z_j}(x) \propto O(h_*),
\]

\[
H = (H_{ij})_{q \times q}, H_{ij} = g_{z_i z_j}(x),
\]

\[
(zh)' Hzh = \sum_{j=1}^{q} \sum_{k=1, k \neq j}^{q} 2g_{z_j z_k}(x) z_j z_k h_j h_k + \sum_{j=1}^{q} g_{z_j z_j}(x) z_j^2 h_j^2 \propto O(h_*^2),
\]

and we consider \( h_* \) to be any bandwidth \( h_j, j \in \{1, \ldots, q\} \). Inserting the Taylor series
expansion in Equation (3.13) into Equation (3.12) yields

\[
E[\hat{g}(x)] \approx \int \left( g(x) + zh \cdot \nabla g(x) + \frac{1}{2} (zh)'Hzh + \ldots \right) K(z)k\left( \frac{g(zh+x) - g(x)}{h_{q+1}} \right) dz,
\]

\[
= g(x) + \sum_{i=1}^{q} h_i g_{z_i}(x) \int K(z)k\left( \frac{g(zh+x) - g(x)}{h_{q+1}} \right) zdz + O(h_i^2).
\]

(3.14)

Recall the assumptions for the kernel function \( k \) that as \( w \rightarrow \pm \infty, k(w) \rightarrow 0 \), and \( k(w) < \infty \ \forall w \in \mathbb{R} \). Consider the Taylor series expansion inside the kernel function

\[
k\left( \frac{g(zh+x) - g(x)}{h_{q+1}} \right) = k\left( \frac{zh \cdot \nabla g(x) + \frac{1}{2} (zh)'Hzh + O(h^3)}{h_{q+1}} \right).
\]

We denote the direction where there is a jump in \( g(x) \) as \( * \). Specifically at a jump in \( g(x) \), we have \( g_{x_+}(x) \) is very large such that \( g_{x_+}(x) \rightarrow \pm \infty \). We also have that as \( n \rightarrow \infty, h_+ \rightarrow 0 \). Either \( g_{x_+}(x)h_+ \rightarrow \pm \infty \) or \( g_{x_+}(x)h_+ \rightarrow 0 \) but slower than \( h_{q+1} \), thus the first term \( \frac{g_{x_+}(x)h_+}{h_{q+1}} \rightarrow \pm \infty \). Similarly, \( g_{x_+}(x) \rightarrow \pm \infty \) gives that the second term either goes to 0 or \( \pm \infty \). Thus, we have that \( k\left( \frac{g(zh+x) - g(x)}{h_{q+1}} \right) \rightarrow 0 \) for \( h_+ \rightarrow 0 \) and \( g_{x_+}(x) \rightarrow 0 \). The other terms are order \( O\left( \frac{h^2}{h_{q+1}} \right) = O(h_+) \rightarrow 0 \). We assume that as \( n \rightarrow \infty \), then \( h_+ \rightarrow 0 \). Therefore as \( n \rightarrow \infty \), we get that \( E[\hat{g}(x)] = g(x) \) and that the anisotropic local constant estimator is asymptotically unbiased.

The variance of the estimator is

\[
\text{Var}(\hat{g}(x)) = \text{Var}\left( \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i-x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i-x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right)} \right),
\]

\[
= \text{Var}\left( \frac{\sum_{i=1}^{n} (g(X_i) + \epsilon_i) K \left( \frac{X_i-x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i-x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right)} \right),
\]
Since $X_i$ is fixed and $\text{Var}(\epsilon_i) = \sigma^2$, this simplifies to

$$\text{Var}(\hat{g}(x)) = \sigma^2 \sum_{i=1}^{n} \left[ K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i) - g(x)}{h_{q+1}} \right) \right]^2,$$

$$= \sigma^2 \frac{1}{n} \sum_{i=1}^{n} \left[ K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i) - g(x)}{h_{q+1}} \right) \right]^2.$$

Taking $n$ to be very large allows us to approximate these sums as integrals in

$$\text{Var}(\hat{g}(x)) \approx \sigma^2 \frac{1}{n} \int \left[ K \left( \frac{w - x}{h} \right) k \left( \frac{g(w) - g(x)}{h_{q+1}} \right) \right]^2 dw.$$

Substituting $z = \frac{w - x}{h}$ as before yields

$$\text{Var}(\hat{g}(x)) \approx \sigma^2 \frac{1}{n} \int K(z) k \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) dz.$$

We have assumed that as $n \to 0$ and $h \to 0$, then $nh \to \infty$. This assumption implies that $\text{Var}(\hat{g}(x)) \to 0$ as $n \to \infty$. Therefore, $\hat{g}(x)$ is a consistent estimator of $g(x)$. 

We have established that the anisotropic local constant estimator in Equation (3.11) converges in probability to the data generating process $g(x)$ under certain conditions. The next step is to establish the rate that the anisotropic local constant estimator converges to the underlying data generating process compared to competing methods.

**Rate of convergence of the anisotropic local constant estimator**

The rate of convergence is important when comparing estimation methods, as it guides a practitioner on which estimator has the minimum expected squared error for a given
sample, assuming the regression model is correctly specified. Typically, the fewer assumptions on the structure of the data generating process the slower an estimator converges. A pointwise estimate of the mean squared error (MSE) is used to find the optimal rate of convergence. Using the expressions in Equations (3.14) and (3.15) for the bias and variance, the MSE for this estimator is proportional to the sample size and lowest order of bandwidth by

\[
\text{MSE}(\hat{g}(x)) = \text{bias}(\hat{g}(x))^2 + \text{Var}(\hat{g}(x)) \propto h^2 + \frac{1}{nh^q}. \tag{3.16}
\]

The assumption that each bandwidth of the kernel functions has the same magnitude and converges to zero at the same rate allows us to treat each \(h_1, \ldots, h_q\) as the “same” \(h_\star\) in this context. Since the bandwidth is a smoothing parameter we specify, we calculate an optimal smoothing parameter that minimizes the MSE and determine the relationship of sample size to the optimal smoothing parameter. Taking the first derivative of the MSE with respect to bandwidth \(h_\star\) yields

\[
\frac{\partial \text{MSE}}{\partial h_\star} \propto 2h_\star + \frac{-1}{nh^q},
\]

\[
\Rightarrow 0 \propto 2h_\star + \frac{-1}{nh^q},
\]

\[
\Rightarrow h_\star \propto n^{-1/q}. \tag{3.17}
\]

This shows that as \(n \to \infty\), we have that bandwidth \(h_i \to 0\), which we assumed for the proof of Theorem 1. Inserting the optimal bandwidth in Equation (3.17) into (3.16) shows the MSE converges to zero at a rate \(O\left(\frac{n^{-2}}{n^{q+2}}\right)\). This implies the rate of convergence for our estimator is \(O\left(\frac{1}{n^{q+2}}\right)\), which is slower than the isotropic estimator’s rate \(O\left(\frac{n^{-2}}{n^{q+2}}\right)\).

Recall that the parametric rate of convergence is \(O\left(n^{-\frac{1}{2}}\right)\), so the parametric model should converge to the data generating process faster than the isotropic estimator. However, this is for a correctly specified parametric model. If you misspecify the parametric model, then the parametric estimator never converges to the data generating process, and the isotropic estimator converges to the data generating process faster than the parametric one. In a similar way, the isotropic estimator should outperform the anisotropic, as
long as it is the correctly specified estimator. If there are change-points in the regression function, we argue that the isotropic estimator is a misspecification and the anisotropic estimator is the correct specification and will outperform the isotropic estimator.

There is a not-so-subtle practical issue with the anisotropic estimator that you may have noticed; we are using the data generating process \( g(x) \) in a kernel function to estimate itself. This implies we have access to the full knowledge of \( g(x) \) that we are trying to estimate. In fact, we do have access to information about \( g(x) \) through \( Y_i \) and we obtain that information through a pilot estimator \( \hat{g}(x) \).

### 3.2.2 Pilot estimator for anisotropic local constant estimator

We have shown the weak consistency of the \( q \)-dimensional anisotropic local constant estimator when using the true regression function \( g(x) \). This estimator allows local information to be filtered if there is a large change-point in the regression function \( g(x) \). However, in practice, we do not have access to the full information of \( g(x) \). In order to use this estimator, we propose using a pilot estimator \( \tilde{g}(x) \). The anisotropic local constant estimator becomes

\[
\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}} \right)}.
\] (3.18)

Equation (3.18) is no longer the same estimator as Equation (3.11), so we need to re-establish weak consistency. We show the anisotropic local constant estimator that uses \( \tilde{g}(x) \) as a pilot estimator also converges in probability to \( g(x) \) using two probability theorems (Grimmett and Stirzaker, 2016) in Appendix A.1 and the assumption that the pilot estimator has the property \( \tilde{g}(x) \xrightarrow{P} g(x) \).

**Theorem 2** Assume that \( \tilde{g}(x) \xrightarrow{P} g(x) \). For the estimator in (3.18), then \( \hat{g}(x) \xrightarrow{P} g(x) \).

**Proof.** Consider the estimator in Equation (3.18) with a Taylor series expansion from
Equation (3.13) for $\tilde{g}(X_i)$ about $x$

$$\tilde{g}(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{(X_i - x) \cdot \nabla \tilde{g}(x) + \frac{1}{2} (X_i - x)' \tilde{H}(X_i - x) + \tilde{R}(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) k \left( \frac{(X_i - x) \cdot \nabla \tilde{g}(x) + \frac{1}{2} (X_i - x)' \tilde{H}(X_i - x) + \tilde{R}(x)}{h_{q+1}} \right)}.$$

(3.19)

Theorem A.1 gives that

$$\frac{(X_i - x) \cdot \nabla \tilde{g}(x) + \frac{1}{2} (X_i - x)' \tilde{H}(X_i - x) + \tilde{R}(x)}{h_{q+1}} \to \frac{(X_i - x) \cdot \nabla g(x) + \frac{1}{2} (X_i - x)' H(X_i - x) + R(x)}{h_{q+1}},$$

and the right side is constant if each term converges, where $\nabla g(x), H,$ and $R(x)$ are bounded. However for this estimator, we assume that as $n \to \infty$ that $h \to 0$. So, we need that $\tilde{g}(x) \to g(x)$ faster than $h \to 0$. Thus, the numerator in Equation (3.19) converges to the random variable numerator in Equation (3.11) and the denominator in Equation (3.19) converges to the constant denominator in Equation (3.11). Theorem A.2 gives that Equation (3.18) converges in probability to Equation (3.11) and we are done.

A candidate for a pilot estimator that converges in probability is the isotropic local constant kernel estimator in Equation (2.2). Theorem 2 applies for this pilot estimator, but only with specific assumptions on the bandwidths of the pilot estimator in relation to the bandwidths of the anisotropic estimator. We need the bandwidths for the pilot estimator to converge to zero faster than the bandwidths for the anisotropic estimator, such that $\tilde{g}(x) \to g(x)$ faster than the bandwidths for the anisotropic estimator go to zero. This implies that the anisotropic estimator can be viewed as a corrective method that updates the isotropic estimator if we think the data has change-points. Section 3.3.1 contains a simulation study that compares data with and without change-points, and the result of iteratively smoothing the regression function.

### 3.2.3 Iterative estimation

We have shown that the anisotropic local constant converges in probability to the data generating process $g(x)$ if the pilot estimator $\tilde{g}(x)$ converges in probability to $g(x)$. Therefore, we need to a any pilot estimator that converges to the data generating process and
we choose the anisotropic local constant estimator. The algorithm for iterative estimation is:

1. Fit the isotropic model using

\[ \tilde{g}(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right)}. \]

2. Fit the anisotropic model using

\[ \tilde{g}_1(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}} \right)}. \]

3. Iteratively fit the anisotropic model \( d \) times using

\[ \tilde{g}_j(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}_{j-1}(X_i) - \tilde{g}_{j-1}(x)}{h_{q+1}} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) k \left( \frac{\tilde{g}_{j-1}(X_i) - \tilde{g}_{j-1}(x)}{h_{q+1}} \right)}, \quad j \in \{2, 3, \ldots, d\}. \]

### 3.2.4 Anisotropic local linear estimator

Consider the nonparametric regression model \( Y = g(X) + \epsilon \) where \( \epsilon \) is random noise, \( X \) and \( Y \) are explanatory and dependent variables, and \( g(\cdot) \) is a nonlinear smooth regression function that is the mean of \( Y \) given \( X \). The local linear minimization problem is

\[ \min_{\{a,b\}} (Y - \mathcal{X}\delta)' K(x)(Y - \mathcal{X}\delta), \]

where \( \delta = \delta(x) = (a(x), b(x)' \prime)' \), \( \mathcal{Y} = (Y_1, Y_2, \ldots, Y_n)'_{n \times 1} \), \( \mathcal{X} = (1, X_i - x)_{n \times (1+q)}' \), and

\[ K(x) = \begin{bmatrix}
  K \left( \frac{X_1 - x}{h} \right) & 0 & 0 & 0 \\
  0 & K \left( \frac{X_2 - x}{h} \right) & 0 & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & K \left( \frac{X_n - x}{h} \right)
\end{bmatrix}. \]
The solution to this least-squares minimization problem is

$$
\hat{\delta}_{ll}(x) = (\hat{a}, \hat{b}') = \left[ X'K(x)X \right]^{-1}X'K(x)Y,
$$

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

The solutions for isotropic local linear regression yield \( \hat{a}(x) = \hat{g}(x) \) and \( \hat{b}(x) = \hat{g}'(x) \). The anisotropic local linear estimator is an extension of the local linear estimator. The range kernel is multiplied at each row of the product kernel matrix \( K(x) \) which gives solution

$$
\hat{\delta}_{all}(x) = \left[ \sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right) \left( \frac{1}{X_i - x} \right) (1, (X_i - x)') \right]^{-1}
$$

$$
\times \sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right) \left( \frac{1}{X_i - x} \right) Y_i. \tag{3.20}
$$

Notice that this estimator does not include the data generating process in the design matrix such that \( X = (1, X_i - x, g(X_i) - g(x))'_{n \times (1+q)} \). By including the range kernel in the numerator and denominator, the estimator remains a weighted average of \( Y_i \) where the weights sum to one. Similar to the anisotropic local constant estimator, we show that the anisotropic local linear estimator is a consistent estimator of the underlying data generating process.

**Weak consistency of the anisotropic local linear estimator**

In this section, we show the weak consistency of anisotropic local linear estimator in Equation (3.20). The same assumptions as Theorem 1 need to be made to establish the consistency of the anisotropic local linear estimator, and are repeated here for convenience of reading:

T1 Without loss of generality, we select a second-order kernel function \( (\nu = 2) \) from Equation (2.4). This allows us to compare rates of convergence with the isotropic kernel estimator using the same order.
The regressors $X_i$ are fixed and not random. For the data in Chapter 4, it is natural to assume pixel locations are fixed in an image. This is also a typical assumption for kernel estimator consistency derivations in the literature (Ruppert and Wand, 1994).

Bandwidths are of similar orders of magnitude. This ensures that $h_1, h_2, \ldots, h_q \to 0$ at the same rate. We also assume $h_{q+1} \to 0$ at a faster rate than any of $h_1, h_2, \ldots, h_q$.

**Theorem 3** Under the assumptions T1-T3 above, and the assumptions that $x$ is an interior point and $g(x)$ is at least three times differentiable (except on a set $D$ that forms a set of points (univariate $x$) or curves (multivariate $x$) such that $\mathbb{P}(D) = 0$), then as $n \to \infty$, $h_j \to 0$, $n_jh_j \to \infty$, $\forall j \in \{1, 2, \ldots, q, q+1\}$, we have, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|\hat{a}(x) - g(x)| < \epsilon) = 1,$$

where $n = n_1 \times n_2 \times \ldots \times n_q = n_{q+1}$.

**Proof.** Similar to the anisotropic local constant estimator consistency proof, we study the bias and variance of the estimator. Let $e_1 = (1, 0, \ldots, 0)$ be a $(q+1) \times 1$ column vector. We write the local local linear regression function estimator from Equation (3.20) as

$$\hat{a}(x) = e_1'[\mathcal{X}'K(x)\mathcal{X}]^{-1}\mathcal{X}'K(x)Y.$$

Since $\mathcal{X}$ is fixed, the expectation of this estimator yields

$$E[\hat{a}(x)|\mathcal{X}] = e_1'[\mathcal{X}'K(x)\mathcal{X}]^{-1}\mathcal{X}'K(x)g,$$

where $g = (g(X_1), g(X_2), \ldots, g(X_n))$. Consider the Taylor expansion

$$g = \mathcal{X}'' \begin{bmatrix} g(x) \\ \nabla g(x) \end{bmatrix} + \frac{1}{2} Q(x) + \ldots,$$

where $Q(x) = [(X_1 - x)'H(x)(X_1 - x), (X_2 - x)'H(x)(X_2 - x), \ldots, (X_n - x)'H(x)(X_n - x)]$. 
Inserting these expansions into Equation (3.21) yields

\[
E[\hat{\alpha}(x)|\mathcal{X}] = e' \left[ \frac{g(x)}{\nabla g(x)} \right] + e'_1 [\mathcal{X}'\mathcal{K}(x)\mathcal{X}]^{-1} \mathcal{X}'\mathcal{K}(x) \left( \frac{1}{2} Q(x) + \ldots \right),
\]

\[
= g(x) + e'_1 [\mathcal{X}'\mathcal{K}(x)\mathcal{X}]^{-1} \mathcal{X}'\mathcal{K}(x) \left( \frac{1}{2} Q(x) + \ldots \right). \tag{3.22}
\]

To determine if this estimator is asymptotically unbiased, we must show that the second term goes to zero as \( n \to \infty \). First, we calculate the behaviour of the denominator \( \mathcal{X}'\mathcal{K}(x)\mathcal{X} \) as \( n \to \infty \). Consider

\[
\mathcal{X}'\mathcal{K}(x)\mathcal{X} = \sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \left( \frac{1}{X_j - x} \right) (1, (X_j - x)'),
\]

\[
= \left[ \sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \right] \left( \sum_{j=1}^{n} \frac{1}{X_j - x} \right) (1, (X_j - x)').
\]

Similar to arguments leading to the denominator in Equation (3.12), this denominator is written as

\[
\frac{1}{n} \mathcal{X}'\mathcal{K}(x)\mathcal{X} \approx \left[ \int K(z) k \left( \frac{g(z+h_x) - g(x)}{h_{q+1}} \right) h dz \right] \left[ \int K(z) k \left( \frac{g(z+h_x) - g(x)}{h_{q+1}} \right) z'h^2dz \right].
\]

We write this matrix in terms of orders of \( h \) for clarity, as any other constants are not important in the sequel. Thus,

\[
\frac{1}{n} \mathcal{X}'\mathcal{K}(x)\mathcal{X} \approx \begin{bmatrix} O(h) & O(h^2) \\ O(h^2) & O(h^3) \end{bmatrix}.
\]

The inverse of this matrix is given by

\[
\left[ \frac{1}{n} \mathcal{X}'\mathcal{K}(x)\mathcal{X} \right]^{-1} = \frac{1}{O(h^3)} \begin{bmatrix} O(h^2) & O(h) \\ O(h) & O(1) \end{bmatrix}.
\]
Similarly, consider the term

\[
\frac{1}{n} \mathcal{X}' \mathcal{K}(x) Q(x) = \left[ \frac{1}{n} \sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right) k \left( \frac{g(X_i) - g(x)}{h_{q+1}} \right) (X_i - x)' H(x) (X_i - x) \right] (X_i - x),
\]

which is approximated as integrals, by making the same arguments as above, in

\[
\frac{1}{n} \mathcal{X}' \mathcal{K}(x) Q(x) \approx \left[ \int K(z) k \left( \frac{g(z h + x) - g(x)}{h_{q+1}} \right) z'H(x) z \right] h^3 \begin{bmatrix} O(1) \\ O(h) \end{bmatrix},
\]

Inserting back into Equation (3.22) yields

\[
\left[ \frac{1}{n} \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \frac{1}{n} \mathcal{X}' \mathcal{K}(x) Q(x) \approx \frac{1}{O(h^3)} \begin{bmatrix} O(h^2) & O(h) \\ O(h) & O(1) \end{bmatrix} h^3 \begin{bmatrix} O(1) \\ O(h) \end{bmatrix},
\]

\[
= \frac{1}{O(1)} \begin{bmatrix} O(h^2) \\ O(h) \end{bmatrix},
\]

\[
\Rightarrow \epsilon_1 \left[ \frac{1}{n} \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \frac{1}{n} \mathcal{X}' \mathcal{K}(x) Q(x) \approx O(h^2).
\]

We have shown that the slowest term converges to zero at \( O(h^2) \). Therefore, the expectation in Equation (3.22) becomes

\[
E[\hat{a}(x)|\mathcal{X}] \approx g(x) + O(h^2) \overset{n \to \infty}{\rightarrow} g(x),
\]

and the estimator is asymptotically unbiased. The next step is to consider the variance of the estimator:

\[
\text{Var}(\hat{a}(x)|\mathcal{X}) = \epsilon_1 \left[ \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \left( X' \mathcal{K}(x) \Sigma \mathcal{K}(x) X \right) \left[ \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \epsilon_1,
\]

\[
= \frac{1}{n} \epsilon_1 \left[ \frac{1}{n} \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \left( \frac{1}{n} X' \mathcal{K}(x) \Sigma \mathcal{K}(x) X \right) \left[ \frac{1}{n} \mathcal{X}' \mathcal{K}(x) \mathcal{X} \right]^{-1} \epsilon_1,
\]

where \( \Sigma = \text{diag} \left( K \left( \frac{X_i - x}{h} \right)^2 k \left( \frac{g(X_i) - g(x)}{h_{q+1}} \right)^2 \right) \sigma^2 \). The upper left entry of \( \frac{1}{n} X' \mathcal{K}(x) \Sigma \mathcal{K}(x) X \)
is approximated, similar to the numerator in Equation (3.15), by

\[
\frac{1}{n} \sum_{j=1}^{n} \left[ K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \right]^2 \sigma^2,
\]

\[
\approx \sigma^2 \int \left[ K \left( \frac{w - x}{h} \right) k \left( \frac{g(w) - g(x)}{h_{q+1}} \right) \right]^2 dw,
\]

\[
= \int \left[ K(z) k \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) \right]^2 h dz = O(h).
\]

The upper right block is

\[
\frac{1}{n} \sum_{j=1}^{n} \left[ K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \right]^2 (X_i - x)' \sigma^2,
\]

\[
\approx \sigma^2 \int \left[ K \left( \frac{w - x}{h} \right) k \left( \frac{g(w) - g(x)}{h_{q+1}} \right) \right]^2 (w - x)' dw,
\]

\[
= \int \left[ K(z) k \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) \right]^2 h^2 z' dz = O(h^2).
\]

The lower left block is

\[
\frac{1}{n} \sum_{j=1}^{n} \left[ K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \right]^2 (X_i - x) \sigma^2,
\]

\[
\approx \sigma^2 \int \left[ K \left( \frac{w - x}{h} \right) k \left( \frac{g(w) - g(x)}{h_{q+1}} \right) \right]^2 (w - x) dz,
\]

\[
= \int \left[ K(z) k \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) \right]^2 h^2 dz = O(h^2).
\]

The lower right block is

\[
\frac{1}{n} \sum_{j=1}^{n} \left[ K \left( \frac{X_j - x}{h} \right) k \left( \frac{g(X_j) - g(x)}{h_{q+1}} \right) \right]^2 (X_i - x)(X_i - x)' \sigma^2,
\]

\[
\approx \sigma^2 \int \left[ K \left( \frac{w - x}{h} \right) k \left( \frac{g(w) - g(x)}{h_{q+1}} \right) \right]^2 (w - x)(w - x)' dw,
\]

\[
= \int \left[ K(z) k \left( \frac{g(zh + x) - g(x)}{h_{q+1}} \right) \right]^2 h^3 z' dz = O(h^3).
\]
Chapter 3. Anisotropic smoothing

This gives the matrix

\[
\frac{1}{n} X'K(x) \Sigma K(x) X \approx \begin{bmatrix}
O(h) & O(h^2) \\
O(h^2) & O(h^3)
\end{bmatrix},
\]

and the variance becomes

\[
\text{Var}(\hat{a}(x)|X') \approx \frac{1}{O(nh^6)} e_1 \begin{bmatrix}
O(h^2) & O(h) \\
O(h) & O(1)
\end{bmatrix} \begin{bmatrix}
O(h) & O(h^2) \\
O(h^2) & O(h^3)
\end{bmatrix} \begin{bmatrix}
O(h^2) & O(h) \\
O(h) & O(1)
\end{bmatrix} e_1',
\]

\[
= \frac{1}{O(nh^6)} e_1 \begin{bmatrix}
O(h^4) & O(h) \\
O(h) & O(1)
\end{bmatrix} \begin{bmatrix}
O(h^3) & O(h^2) \\
O(h^2) & O(h^3)
\end{bmatrix} e_1',
\]

\[
= \frac{1}{O(nh^6)} e_1 \begin{bmatrix}
O(h^5) & O(h) \\
O(h^4) & O(h^3)
\end{bmatrix} e_1',
\]

\[
= O \left(n^{-1}h^{-1}\right) \xrightarrow{n \to \infty} 0.
\]

since we have assumed that as \( n \to \infty \), then \( h \to 0 \) and \( nh \to \infty \). Thus, the local linear estimate \( \hat{g}(x) \) converges in probability to \( g(x) \).

\[\blacksquare\]

Rate of convergence of the anisotropic local linear estimator

We have determined that the anisotropic local linear estimator is a consistent estimator. Similar to the anisotropic local constant estimator, we need to determine the rate of convergence. Interestingly, the anisotropic local constant and local linear estimators have the same rate of convergence—not on the boundary, since local constant has a slower rate of convergence on non-interior points—since the MSE in Equation (3.17) are similar and \( \hat{g}(x) - g(x) \) convergence rates are the same for both. See Section 3.2.1 for details on convergence rates.
3.2.5 Anisotropic local polynomial estimator

Consider the nonparametric regression model $Y = g(X) + \epsilon$ where $\epsilon$ is random noise, $X$ and $Y$ are explanatory and dependent variables, and $g(\cdot)$ is a nonlinear smooth regression function that is the mean of $Y$ given $X$. The local polynomial minimization problem requires extensive notation (Masry, 1996a,b)

$$r = (r_1, \ldots, r_q), \quad r! = r_1! \times \ldots \times r_q!, \quad \bar{r} = \sum_{j=1}^{q} r_j,$$

$$x^r = x_1^{r_1} \times \ldots \times x_q^{r_q}, \quad \sum_{0 \leq r \leq p}^{p} \sum_{j=1}^{j} \sum_{r_q=0}^{j} \text{ where } \bar{r} = j,$$

$$(D^r)g(x) = \frac{\partial^r g(x)}{\partial x_1^{r_1} \ldots \partial x_q^{r_q}}.$$

Under the assumption that $g(x)$ has order $p + 1$ derivatives at point $x$, then $g(z)$ is approximated locally by using a multivariate polynomial of total order $p$ given by

$$g(x) \approx \sum_{0 \leq r \leq p}^{p} \frac{1}{r!} (D^r)g(x)|_{v=x}(x-x)^r.$$

Then, the minimization problem becomes

$$\sum_{j=1}^{n} \left\{ Y_i - \sum_{0 \leq r \leq p}^{p} b_r(x)(X_i - x)^r \right\} K\left( \frac{X_i - x}{h} \right).$$

The solutions for the isotropic local polynomial minimization problem yield $\hat{g}(x) \equiv \hat{b}_b(x)$ and $\hat{g}^{(l)}(x) \equiv ll\hat{b}_l(x).$ This estimator has uniform almost sure rate of convergence and is pointwise asymptotically normal—for details see (Masry, 1996a,b). The anisotropic local polynomial (ALP) estimator is an extension of the local linear estimator, by adding the anisotropic term to the kernel function

$$\sum_{j=1}^{n} \left\{ Y_i - \sum_{0 \leq r \leq p}^{p} b_r(x)(X_i - x)^r \right\} K\left( \frac{X_i - x}{h} \right) k\left( \frac{g(X_i) - g(x)}{h_{q+1}} \right).$$
The local polynomial estimator’s bias and variance are omitted, as we do not investigate the finite sample performance of this estimator, and the derivation would be similar to the above anisotropic local constant and local linear estimators.

3.2.6 Bandwidth selection methods

The three options for bandwidth selection for isotropic kernel selection used in this dissertation are: (1) ad hoc bandwidth selection, (2) least-squares cross-validation, and (3) least-squares AIC (Li and Racine, 2007) described in Section 2.1. By far, the most cavalier method for choosing a smoothing parameter is to select the one that appears to work well. Arbitrarily selecting the smoothing parameter for any smoothing is all but guaranteed to give a sub-optimal bandwidth selection that directly affects the rate of convergence of the estimator to the data generating process.

We need to select bandwidths \( h_j, j \in \{1, \ldots, q + 1 \} \) and the ad hoc method needs adjustments for the anisotropic estimators. Recall that the ad hoc approach for an isotropic estimator is to use \( h_j = c_j X_{sd(j)} n^{-1/(q+4)} \), where \( c_j \) is a constant (usually 1 or close to 1) and \( X_{sd(j)} \) is the sample standard deviation of the \( X_{ij} \)’s. This approach is specifying the bandwidth based on the spacing/density of the explanatory data. We have assumed specifically that \( X_i \)’s are not random and are evenly spaced. This allows for the approach to be calculated once for any number of curves with equally spaced data, and is therefore computationally efficient, but may not yield an optimal bandwidth for our smoother.

Extending this ad hoc procedure to include a range kernel bandwidth, the bandwidth with the optimal rate of convergence would suggest \( h_j = c_j X_{sd(j)} n^{-1/(q+2)} \) for \( m \in \{1, \ldots, q \} \) and \( h_{q+1} = c_t Y_{sd(q+1)} n^{-1/(q+2)} \), where \( Y_{sd} \) is the sample standard deviation of the output data minus the pilot estimator that we are using for the anisotropic procedure, i.e. \( Y_i - \tilde{g}(X_i) \).
3.3 Simulation studies

3.3.1 One-dimensional data

We have chosen a few simple simulated data examples with change-point regression functions that allow us to explore the performance of our estimator on change-point data. The simulated data use a piecewise constant function with two change-points and a continuous function with and without change-points. The three data generating processes used in simulation are shown in Figure 3.2 using the equations

\[
\text{Piecewise constant: } g(x) = \begin{cases} 
1, & 0 \leq x \leq 1, \\
7, & 1 < x \leq 2, \\
3, & 2 < x \leq 3, 
\end{cases}
\]

\[
\text{Continuous: } g(x) = 50 \left( \left( \frac{x}{3} \right)^2 - \left( \frac{x}{3} \right)^3 \right), \quad 0 \leq x \leq 3,
\]

\[
\text{Continuous with jump: } g(x) = \begin{cases} 
50 \left( \left( \frac{x}{3} \right)^2 - \left( \frac{x}{3} \right)^3 \right), & 0 \leq x \leq 1.5, \\
50 \left( \left( \frac{x}{3} \right)^2 - \left( \frac{x}{3} \right)^3 \right), & 1.5 < x \leq 3,
\end{cases}
\]

and i.i.d. noise \( \epsilon \sim N(0, \sigma^2) \). In our Monte Carlo simulations, we consider 125 Monte Carlo replicates with \( \sigma \in \{0.1, 0.5, 1, 2\} \). The Monte Carlo simulation algorithm is:

1. Simulate \( n \) \( N(0, \sigma^2) \) and calculate \( Y_i = g(X_i) + \) such that \( X_i \) start at 0 and are evenly spaced in the interval \( 0 \leq X_i \leq 3 \).

2. Fit the isotropic model using

\[
\tilde{g}(x) = \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right)}{\sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right)}.
\]

3. Fit the anisotropic model using

\[
\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i \frac{X_i - x}{h} \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}}}{\sum_{i=1}^{n} \frac{X_i - x}{h} \frac{\tilde{g}(X_i) - \tilde{g}(x)}{h_{q+1}}}
\]

4. Repeat steps 1-3 for a total of \( M \) Monte Carlo replicates.
Chapter 3. Anisotropic smoothing

Figure 3.2: The three one-dimensional data generating processes used for simulations. The left most figure is the piecewise constant regression function. The middle figure is the continuous regression function. The right most figure is the same continuous regression function with a jump at $X = 1.5$.

The piecewise constant data type simulates a simple change-point regression function that is often analyzed in the image processing literature. The continuous regression function without the change-point allows us to investigate the scenario when a change-point estimator is inappropriate, and the continuous regression function with a change-point allows for a more complex regression function on either side of the change-point.

An example of the isotropic and anisotropic local constant estimators is shown in Figure 3.3. To determine the theoretical performance of regression estimators, we have looked at the pointwise convergence rate—that is, the rate at which the mean square error between our estimator and underlying data generating process. In practice, to compare performance of nonparametric estimators on simulated data (since we need to know the function $g$ at all $X_i$) we calculate the mean of the estimated squared errors (MESE) (Thompson, 2014) given by

$$\text{MESE}(\hat{g}) = \frac{1}{n} \sum_{i=1}^{n} (g(X_i) - \hat{g}(X_i))^2.$$  \hspace{1cm} (3.23)

Three estimators are fit to simulated data: the isotropic local-constant estimator (LC), our anisotropic local constant estimator with LC as the pilot estimator (ALC), and our
anisotropic local constant estimator with the data generating process (the “truth”) as the pilot estimator (ALCT). All three simulated data types are used with four error standard deviations ($\sigma \in \{0.1, 0.5, 1, 2\}$) with 125 Monte Carlo replicates, the uniform kernel, and sample sizes of 400 to 1600. The bandwidth selection method is the cross-validation AIC method. The relative performance of these estimators is shown in Tables 3.1, 3.3, and 3.5 using the mean of the MESEs, where the MESE of each Monte Carlo replicate is calculated using Equation (3.23) and then averaged over $M$ Monte Carlo simulations. Tables 3.2, 3.4, 3.6 show the sample standard deviations of the MESEs.

Figure 3.3: Fitting the isotropic and anisotropic local constant estimators to the piecewise constant regression function data and comparing to the underlying data generating process (truth). This plot uses simulation parameters $n = 400$ and $\sigma = 0.5$, the uniform kernel function for smoothing, and least-squares cross-validation bandwidth selection method.

For the piecewise constant regression function data in Table 3.1, the ALC estimator
consistently has a smaller MESE than the LC estimator, showing that the ALC estimator improves the fit for the LC estimator by better accounting for the change-point. The ALCT estimator consistently has a smaller MESE than the ALC estimator, showing that anisotropic local constant estimator is further improved by using a better pilot estimator. Given a good pilot estimator, we get relatively better estimates of the data generating process, occluded only by the noise of the data. Table 3.2 shows that as noise increases, the ALC estimator tends to have a large sample standard deviation for the MESE. However, Figure 3.4 shows that the ALC is still an overall improvement over the LC estimator. For box plots of MESE for each simulation, see Appendix A.3.

Table 3.1: The mean of the MESEs for each nonparametric estimator fitted to the simulated piecewise constant regression function data.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$n = 400$</th>
<th>$n = 800$</th>
<th>$n = 1600$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>0.1</td>
<td>0.00363</td>
<td>0.00179</td>
<td>0.00012</td>
</tr>
<tr>
<td>0.5</td>
<td>0.079</td>
<td>0.02119</td>
<td>0.003</td>
</tr>
<tr>
<td>1</td>
<td>0.1855</td>
<td>0.06835</td>
<td>0.01181</td>
</tr>
<tr>
<td>2</td>
<td>0.43075</td>
<td>0.24942</td>
<td>0.04743</td>
</tr>
</tbody>
</table>

Table 3.2: The sample standard deviation of the MESES for each nonparametric estimator fitted to the simulated piecewise constant regression function data.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$n = 400$</th>
<th>$n = 800$</th>
<th>$n = 1600$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>0.1</td>
<td>0.00276</td>
<td>0.00105</td>
<td>0.00013</td>
</tr>
<tr>
<td>0.5</td>
<td>0.01993</td>
<td>0.01877</td>
<td>0.00316</td>
</tr>
<tr>
<td>1</td>
<td>0.03004</td>
<td>0.04527</td>
<td>0.0126</td>
</tr>
<tr>
<td>2</td>
<td>0.09884</td>
<td>0.12813</td>
<td>0.05201</td>
</tr>
</tbody>
</table>

Table 3.3 shows that the ALC estimator has a larger MESE (generally worse fit) than the LC estimator on continuous data that does not contain a jump in the regression function. This comparison demonstrates the effect on smoothing when the ALC is used inappropriately. Interestingly, the ALCT estimator shows an improvement over the LC estimator. This is consistent with similar range kernel methods used to improve regression estimates in the literature (Li and Racine, 2007). However, a practitioner may choose not to use the ALC estimator in this context, given the LC estimator is performing adequately
Figure 3.4: The MESEs for the three kernel estimators on the piecewise constant function. Note that \(n = Z00\), e.g. LC.4 is the isotropic local constant estimator with \(n = 400\) and ALC.16 is the anisotropic estimator with \(n = 1600\).
at estimating the continuous regression function, and change-points are not present in the
data. Table 3.4 shows the sample standard deviations of the MESE for each estimator
on the continuous regression function data.

Table 3.3: The mean of the MESEs for each nonparametric estimator fitted to the sim-
ulated continuous regression function data.

<table>
<thead>
<tr>
<th></th>
<th>n = 400</th>
<th>n = 800</th>
<th>n = 1600</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>0.1</td>
<td>0.00126</td>
<td>0.00189</td>
<td>0.00101</td>
</tr>
<tr>
<td>0.5</td>
<td>0.01473</td>
<td>0.01938</td>
<td>0.01239</td>
</tr>
<tr>
<td>1</td>
<td>0.0416</td>
<td>0.05359</td>
<td>0.03765</td>
</tr>
<tr>
<td>2</td>
<td>0.12214</td>
<td>0.15806</td>
<td>0.11457</td>
</tr>
</tbody>
</table>

Table 3.4: The sample standard deviation of the MESES for each nonparametric estima-
tor fitted to the simulated continuous regression function data.

<table>
<thead>
<tr>
<th></th>
<th>n = 400</th>
<th>n = 800</th>
<th>n = 1600</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>0.1</td>
<td>0.00031</td>
<td>0.00053</td>
<td>0.00025</td>
</tr>
<tr>
<td>0.5</td>
<td>0.00412</td>
<td>0.00525</td>
<td>0.00409</td>
</tr>
<tr>
<td>1</td>
<td>0.01616</td>
<td>0.01882</td>
<td>0.01604</td>
</tr>
<tr>
<td>2</td>
<td>0.06245</td>
<td>0.07175</td>
<td>0.06931</td>
</tr>
</tbody>
</table>

Table 3.5 shows the continuous data, except there is a jump in the regression function.
We see a similar result as when smoothing the piecewise constant regression function
data, where the ALC estimator has a smaller MESE than the LC estimator. However,
the smaller sample standard deviations of the MESE for the LC estimator over the ALC
estimator in Table 3.6 suggest that these smooths may not be significantly different from
each other. Figure 3.5 shows a box plot of MESEs for fitting the LC, ALC, and ALCT
to data of a continuous regression function with a jump. The figure shows that ALC
is an improvement over the LC estimator. Figure 3.6 shows a single fit of the kernel
estimators to data of a continuous regression function with a jump. We see that for
the ALC estimator, the improvement at the change-point is balanced by the loss of
information in areas where \( g'(x) \) is large, and the range kernel loses information local in
the \( X \) direction when there is no jump causing undersmoothing.
Figure 3.5: The MESEs for the three kernel estimators on data from the continuous regression function with a jump. Note that \( Z \) means \( n = Z00 \), e.g. LC.4 is the isotropic local constant estimator with \( n = 400 \) and ALC.16 is the anisotropic estimator with \( n = 1600 \).
Table 3.5: The mean of the MESEs for each nonparametric estimator fitted to the simulated continuous data with a jump in the regression function.

<table>
<thead>
<tr>
<th></th>
<th>n = 400</th>
<th>n = 800</th>
<th>n = 1600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>σ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.01732</td>
<td>0.00319</td>
<td>0.00152</td>
</tr>
<tr>
<td>0.5</td>
<td>0.07338</td>
<td>0.04232</td>
<td>0.01516</td>
</tr>
<tr>
<td>1</td>
<td>0.15469</td>
<td>0.12098</td>
<td>0.04529</td>
</tr>
<tr>
<td>2</td>
<td>0.33026</td>
<td>0.28736</td>
<td>0.14387</td>
</tr>
</tbody>
</table>

Table 3.6: The sample standard deviation of the MESES for each nonparametric estimator fitted to the simulated continuous data with a jump in the regression function.

<table>
<thead>
<tr>
<th></th>
<th>n = 400</th>
<th>n = 800</th>
<th>n = 1600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC</td>
<td>ALC</td>
<td>ALCT</td>
</tr>
<tr>
<td>σ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.00109</td>
<td>0.00043</td>
<td>0.00030</td>
</tr>
<tr>
<td>0.5</td>
<td>0.00735</td>
<td>0.01848</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>0.03002</td>
<td>0.03732</td>
<td>0.01904</td>
</tr>
<tr>
<td>2</td>
<td>0.08068</td>
<td>0.08431</td>
<td>0.07131</td>
</tr>
</tbody>
</table>

3.3.2 Simulation results for iterative anisotropic smoothing

Figure 3.7 shows the box plot of the MESEs of the LC estimator, the ALC estimator using the LC as the pilot estimator (ALC.1), and each ALC estimator that uses the ALC estimator as the previous step (ALC.2, e.t.c.). The bandwidth is calculated once and then reused for each fit. We see that the initial anisotropic smooth performs worse and worse for each re-smoothing. We believe this is caused by sub-optimal bandwidth selection at each stage and finite sample properties. Figure 3.8 shows the same estimation procedure with the bandwidth recalculated at each re-smooth. While this estimator performs better, we see an increase in variance of the MESEs. For each of these figures, regardless if the bandwidth for the ALC smoothers is calculated at each step or only at the first step, the iterative ALC estimator generally increases the MESE relative to the first iteration of the ALC estimator. Therefore, we have shown that using our algorithm, the anisotropic local constant estimator that uses the isotropic local constant estimator as a pilot will be the estimator for change-point data in Chapters 4 and 5. See Appendix A.4 for simulations with similar results using any combination of bandwidth selection method and data type for reiterative smoothing. Next, we compare bandwidth selection methods used for the
Figure 3.6: Fitting the isotropic and anisotropic local constant estimators to the continuous regression function with a jump data and comparing to the underlying data generating process (truth). This plot uses simulation parameters $n = 400$ and $\sigma = 1$, the uniform kernel function for smoothing, and least-squares cross-validation bandwidth selection method.

3.3.3 Bandwidth selection comparison

Table 3.7 contains the MESEs for each of the bandwidth selection methods on all data types, and shows that all methods perform similarly. Figures 3.9 and 3.10 support the claim that there is no difference in the bandwidth selection methods for the anisotropic smoother. Since the cross-validation methods are computationally heavy, we find that the ad hoc method works best for our efforts as it performs similarly to cross-validation.
Figure 3.7: Iteratively fitting the ALC estimator with LC as a pilot estimator on the piecewise constant regression function data. The bandwidth is selected once using least-squares cross-validation for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

Table 3.7: The mean of the MESEs for the anisotropic local constant estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS).

<table>
<thead>
<tr>
<th>DGP</th>
<th>$n = 200$</th>
<th>$n = 400$</th>
<th>$n = 800$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH</td>
<td>AIC</td>
<td>LS</td>
</tr>
<tr>
<td>UJ</td>
<td>0.12023</td>
<td>0.14784</td>
<td>0.12824</td>
</tr>
<tr>
<td>C</td>
<td>0.08141</td>
<td>0.07322</td>
<td>0.07341</td>
</tr>
<tr>
<td>CJ</td>
<td>0.14366</td>
<td>0.152</td>
<td>0.16583</td>
</tr>
</tbody>
</table>

3.3.4 Two-dimensional simulated fire spread data

We have shown that for one-dimensional data, our anisotropic smoothing methods show significant improvement of fit in the presence of change-points. In Chapter 4, we process
Figure 3.8: Iteratively fitting the ALC estimator with LC as a pilot estimator on the piecewise constant regression function data. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$.

Fire spread data collected from fire spread experimentation. In this section, we demonstrate the effectiveness of using anisotropic diffusion filtering and smoothing on simulated data similar to measurements collected from fire smoldering experiments.

Fire spread videos start with a sheet of fuel where all RGB channel values are high. When a fire begins, areas that are burning have a high red channel value, and low green and blue channel values. Burnt-out areas all have low RGB channel values. The simulated fire data reflects the red value of RGB channels, is calculated using the data generating process

$$g_j(x_{1i}, x_{2i}) = \begin{cases} 
80, & (x_{1i} - x_{1o})^2 + (x_{2i} - x_{2o})^2 < r(t_j)^2, \\
130, & \text{otherwise},
\end{cases}$$

where the origin of the fire is at $(x_{1o}, x_{2o})$, $t_j$ is $j^{th}$ image frame and $r(t_j)$ is the radius
Figure 3.9: The MESEs for the ALC estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS). The simulated data is the piecewise constant regression function, and the parameters for the simulation are $\sigma = 1$ and $n = 800$.

of the fire front at frame $t_j$. The model is $Y_i(t_j) = Y_{ij} = g_j(x_{1i}, x_{2i}) + \epsilon_{ij}$, where the error is identically and independently normally distributed with mean 0 and variance $\sigma^2 = 20$. The simulation parameters is 15 Monte Carlo replicates with one simulated $80 \times 80$ pixel-grid fire per replicate. The choice for the variance is based the observed variability in experimental fire measurements.

The left column of Figure 3.11 shows the data generating process, and the right column shows the data generating process with noise. The application of an isotropic local constant kernel estimator yields Figure 3.12a, where there is relatively large error around the change-point, shown by a plot of the estimated errors in Figure 3.12b. The
Figure 3.10: The MESEs for the ALC estimator where the bandwidth is selected using an ad hoc approach (AH), AIC cross-validation, or least-squares cross-validation (LS). The simulated data is the continuous regression function with a jump, and the parameters for the simulation are $\sigma = 1$ and $n = 800$.

two-dimensional anisotropic local constant kernel estimator is shown in Figure 3.13a, where Figure 3.13b shows the estimated error, with less error around the jump, but more error in the flat regions. Figure 3.13c shows the same ALC smoother with the range bandwidth increased by a factor of five. Figure 3.13d shows the estimated error of the ALC with an increased range kernel bandwidth. By increasing the bandwidth for the kernel in the $z$-direction, the regions between boundaries have a decrease in error while minimally affecting the error in estimating the change-point.
3.3.5 Anisotropic diffusion

Figure 3.14 shows anisotropic diffusion applied to the simulated data and the spatial estimated error. Since anisotropic diffusion methods naturally account for any sharp changes in the function, they should perform well around the change-point of the simulated fire spread data. However, the noise of the change-point negatively affects the anisotropic diffusion’s ability to separate the distinct regions. Figure 3.14b shows large estimated errors around the change-point, which demonstrates poor performance at estimating the data generating process at the change-point. There is also an overall bias created by the anisotropic diffusion method, and we believe this causes the poor performance. It may be that the number of smoothing steps is not optimal, but this is the best we could obtain using the available anisotropic diffusion software.

To evaluate each fire, the MESE in Equation (3.23) is calculated for each image frame at \( t_j \) of each simulated fire video. Figure 3.15 shows the mean of the MESEs for each frame \( t_j \) across all simulated fire videos for the kernel estimators and anisotropic diffusion. This figure demonstrates the performance of each smoother at each time point, where the lowest mean MESE shows the least relative error. The estimators behave similarly when the radius of the “burning edge” is small. As the radius increases, the ALC and LC estimators perform similarly with the ALC estimator generally performing slightly worse. The larger range kernel bandwidth smoother ALC.10 performs worse than the smaller ALC.3, but they eventually switch around time 22. They switch back around time 70, but problems here are expected since the burning edge is so close to the boundary of the data. This figure demonstrates the inability to choose a good bandwidth. Table 3.8 shows the mean and standard deviation of the MESE across all frames and simulated fire videos using each kernel smoother and anisotropic diffusion. This table shows that the ALC is not performing as well as the LC estimator using least-squares cross-validated bandwidths. By simply increasing those bandwidths by factor of three (ALC.3) and ten (ALC.10) times, we decrease the global error significantly. This decrease in global mean of the MESE is caused by improved estimation away from change-point, but there is a decline in performance around the change-point. The anisotropic diffusion (AD)
estimator performs better than the LC and ALC estimators, but the ALCs with increased range kernel bandwidths outperform the AD estimator.

Table 3.8: The global mean and sample standard deviation of the MESEs over all simulated fires for the isotropic and anisotropic local constant estimators and anisotropic diffusion estimator.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td>27.72091</td>
<td>11.35974</td>
</tr>
<tr>
<td>ALC</td>
<td>29.01993</td>
<td>11.66351</td>
</tr>
<tr>
<td>ALC.3</td>
<td>12.18208</td>
<td>4.98715</td>
</tr>
<tr>
<td>ALC.10</td>
<td>14.48930</td>
<td>7.00527</td>
</tr>
<tr>
<td>AD</td>
<td>20.13268</td>
<td>8.02140</td>
</tr>
</tbody>
</table>

3.4 Discussion

There are a few areas where improvement is needed for the anisotropic smoothing framework, but the most important one is bandwidth selection. We show that theoretically we need relatively larger bandwidths for the anisotropic local constant estimator than the bandwidths used for the pilot isotropic or anisotropic local constant estimators. We show that using a slightly larger bandwidth than those chosen by our automatic bandwidth selectors for our two-dimensional simulated data estimator has the beneficial effect of decreasing the global mean of the MESE and thus improving the smooth. The development of a better bandwidth selector for single and reiterative anisotropic smoothing could lead to improved estimates close to and remote from change points, and in regions where $g'(x)/g(x)$ is large—an area where nonparametric estimators have been known to perform sub-optimally (Fan and Gijbels, 1996). Our current version of the estimator and bandwidth selection does not perform well in these regions, and undersmooths local information unlike the isotropic estimator.

We acknowledge that we have not presented the anisotropic local linear simulation results. That estimator performs sub-optimally on the one- and two-dimensional data. The issue we believe for that estimator is it suffers a greater detriment from sub-optimal bandwidth selection. That procedure needs to be improved before passing judgement on the performance of the anisotropic local linear estimator.
The task of quantifying the efficiency of any number of iterative re-smooths requires investigation. Our current automatic bandwidth selections may be causing the iterative re-smoothing procedure to behave sub-optimally. We want to use iterative smoothing to improve \( g(x) \) as best we can, but we don’t want to perform any unnecessary iterative smooths that do not improve the fit, or cause under- or over-smoothing. We choose a conventional criterion to compare nonparametric smoothing methods: the MESE. Currently, the optimal number of iterative re-smooths is chosen by the minimum MESE of the iterations, which is generally after one anisotropic smoothing step. We see that anisotropic diffusion and anisotropic smoothing have very different smooths visually, but similar mean of MESEs over time in Figure 3.15. Therefore, we make the claim that minimum MSE or MESE may not be the most appropriate measure of an optimal smooth in the context of change-point regression function estimation.

Figures 3.7 and Appendix A.4 show that the ad hoc re-selection of the bandwidth for each iterative re-smoothing step performs worse (median of MESE increase) when considering only the constant and continuous regression functions with jump(s). This difference in performance demonstrates that the ad hoc method is likely a poor choice of bandwidth selection for the iterative approach. For the cross-validated bandwidth re-selections for each smoothing iteration, we can see that the median of the MESE hovers around a similar value, but the spread of the MESEs increases. We found that some of the smooths within each graph benefits from iterative smoothing (decrease in MESE), whereas others do not improve (small to no change in MESE) or even worse (increase in MESE), where these changes in MESE values increase the variance of the MESE. We believe the poor performance of iterative smoothing is caused by bandwidth selection that is not optimized for anisotropic estimators, where sometimes the cross-validation chooses a bandwidth that improves or worsens the overall smoothing performance. We believe there is a need to identify an optimal bandwidth selection method for anisotropic smoothing that balances accuracy around the change-point with accuracy farther away from the change-point, and this is future work.

Theorem 2 is a useful result, but there are some implicit assumptions that need some care to complete the proof. The assumption that the gradient, Hessian, and remainder
Chapter 3. Anisotropic smoothing

terms of the pilot function all converge in probability to the gradient, Hessian and remainder terms of the underlying data generation process needs to be demonstrated in general to solidify the proof. This detail is necessary to the validity of the proof, and based on simulation studies shown in this thesis, we believe that it is a reasonable assumption.

Simulation sample sizes are deliberately small \( n = 400 \) for 1D, \( n = 80 \times 80 \) for 2D) to demonstrate the effectiveness of capturing change-points with less information. Larger simulation sizes were conducted (up to \( n = 128000 \) for 1D), which showed similar results to those presented in this chapter. We believe that the sample sizes shown here demonstrate the improvement in performance as sample size increases, particularly when using the underlying data generating process as the pilot estimator. We are interested in the applications in Chapters 4 and 5 that have low data resolution. Another issue is the number of Monte Carlo replicates; \( M = 125 \) for one-dimensional data and \( M = 15 \) for two-dimensional data. These step counts are quite small, however some simulations were run on larger number replicates (such as ad hoc bandwidth selection) where the results are similar.

Another interesting and subtle result arises from the rate of convergence of the anisotropic estimator. As shown in this chapter for an order \( \nu = 2 \) kernel, the convergence rates for the isotropic and anisotropic local constant estimators are \( O\left(n^{-\frac{2}{2q+4}}\right) \) and \( O\left(n^{-\frac{2}{2q+2}}\right) \) respectively, where \( q \) is the number of dimensions of the explanatory variables. The rate of convergence for a \( q = 1 \) anisotropic local constant estimator is \( O\left(n^{-\frac{2}{3}}\right) \), which is equivalent to the rate of convergence of a \( q = 2 \) isotropic local constant estimator. This implies that adding the range kernel to the isotropic estimator is equivalent to adding another dimension to the problem. Similarly, the rate of convergence for a \( q = 2 \) anisotropic local constant estimator is \( O\left(n^{-\frac{2}{3}}\right) \), which is equivalent to the rate of convergence of a \( q = 4 \) isotropic local constant estimator. Therefore, the underlying data generating process \( g(X_i) \) that has dimension \( q \) inputs into the range kernel is equivalent to doubling the number of dimensions in the rate of convergence of the isotropic estimator—which is \( O\left(n^{-\frac{2}{2q+4}}\right) \) becomes \( O\left(n^{-\frac{2}{2q+4}}\right) = O\left(n^{-\frac{1}{2q+2}}\right) \) when a range kernel is added that depends on \( g(X_i) \).

In this chapter, we have shown that the anisotropic local constant estimator with the
local constant estimator as a pilot performs the best on one- and two-dimensional data. We believe that this estimator implicitly interprets jumps in the regression function as “new boundaries” between regions. Next, we look at the experimental apparatus and measurements, where we will apply the anisotropic local constant estimator as an image processor.
Figure 3.11: Examples of a simulated fire data generating process with and without noise. The simulation’s video length is 70 equally spaced (in time) frames.
Figure 3.12: The isotropic local constant estimator applied to image frame $t_{35}$ of the simulated fire spread data, and the estimated errors.
Chapter 3. Anisotropic smoothing

(a) Anisotropic local constant estimator
(b) Anisotropic local constant estimator minus data generating process

(c) Oversmoothed anisotropic local constant estimator
(d) Oversmoothed anisotropic estimator minus data generating process

Figure 3.13: The anisotropic local constant estimator applied to image frame $t_{35}$ of the simulated fire spread data, and the estimated errors. The first pair of plots use an ad hoc selected bandwidth. The second pair have the range kernel bandwidth increased by a factor of five.
Figure 3.14: The anisotropic diffusion estimator applied to image frame $t_{35}$ of the simulated forest fire spread data, and the estimated errors.

(a) Anisotropic diffusion estimator

(b) Anisotropic diffusion estimator minus data generating process
Figure 3.15: The mean of the MESE of fires at each simulated video frame. LC and ALC are the isotropic and anisotropic local constant estimator with least-squares cross-validated bandwidths. ALC.Z takes the cross-validated bandwidth for the range kernel, multiplies it by Z, and refits the anisotropic local constant estimator. The number of Monte Carlo replicates is 15 fires.
Chapter 4

Fire smoldering experiments

In this chapter, we describe a micro-fire smoldering apparatus, which we use to develop a prototype methodology to study the effects of environmental conditions on fire spread. We look at how the design and experimentation of the apparatus are used to develop statistical analysis tools. These tools could be used to process fire spread satellite imagery data in future applications. We explain how the micro-fire experiments are a mouse model for forest fire spread. We discuss the fire spread measurements recorded from our experiments, investigate the physical structures of the data generating processes of the apparatus, and explain how we control the contributing environmental factors. We process the fire spread measurements using anisotropic smoothing methods and a PDE-based edge-preserving anisotropic diffusion filter, and processed measurements are analyzed in Chapter 5 to estimate fire spread rates. It should be noted that while our primary interest is in using and developing statistical designs and analytic tools, our scientific investigation of experimental fire spread is important, as it provides data that simulates and gives insights into fire spread.

Motivations for a fire spread mouse model

The physical manifestation of any complex, large-scale physical phenomenon in science is a result of the contributions from the macroscopic physical structure, atomic chemical structure, and interactions with environmental conditions. In many scientific fields, including physics and statistics, there is a constant desire to distill measurements of a
process down to a set of factors, equations, or rules that explain the physical manifestations of the process. In some fields, the contributing factors behind a complex physical process are themselves complicated physical processes that are difficult to model; this may obscure the extent of their contribution. For example, the impact of weather on forest fire spread is a complex process hierarchy. In this chapter, we simulate fire spread scenarios while controlling environmental variables using an experimental micro-fire apparatus. The imagery measurements collected allow us to develop statistical techniques for analyzing controlled fire spread data, which may provide insight into the behaviour of wildfire spread. Ultimately, the statistical techniques developed in this dissertation are intended for applications of forest fire spread rate estimation, and introduce new knowledge and understanding of the physical process of fire spread.

Fire spread has an underlying structure that we aim to study, and we want to consider the environmental factors that contribute to fire growth. As a simple example of one-dimensional fire spread, consider a piece of woven string. The string is a fuel bed for flame, and is ignited when there is enough energy to trigger the required action potential for ignition in a local area. Suppose the string is ignited in the center and the fire spreads along the string—“forward” and “backward”—at certain rates. We consider three simplified macroscopic regions of the string: fuel, burning, and burnt-out.

Let’s consider some environmental factors that affect how quickly this string burns. If the string is damp with water, it takes more energy to ignite and burn each portion. If the string is thick or densely woven or “packed”, it may take longer to burn. The orientation of the string relative to Earth’s gravitational field affects the burning rate. One of the goals of this dissertation is to investigate how environmental factors affect fire spread rate mean and variability. We have constructed an apparatus that can allow for investigation of the environmental factors that affect fire spread in a 2-D medium, and the statistical models to process measurements into data and extract information from data to study the fire spread rates. Before introducing the technical details of our fire smoldering apparatus, we first investigate the previous experimental work in fire spread that inspired the design of our micro-fire spread apparatus.
4.1 Prior work in laboratory fire research

In advance of our fire smoldering apparatus design, we discuss relevant and inspirational previous research on experiments related to understanding fire structure and spread for forest fires, and statistical fire modelling. Note that while this is a dissertation that primarily involves invoking statistical methodology on data sets, the work also includes the design of fire spread laboratory experiments that produce the data set. Our non-exhaustive literature review focuses on papers of more recent experiments, where Sullivan (2009c) gives an extensive review of experimentation and further references to past techniques.

One interesting sloped fire spread experiment apparatus consisted of fuel beds of vertical white birch sticks (13.97 x 0.455 x 0.110 cm) arranged in a rectangular grid (Hwang and Xie, 1984; Wolff et al., 1991). The experiments using this apparatus studied the effect of wind speeds (heading, backing, none) and slopes (positive, negative and none) on fire spread, conducted inside various spaces such as an open-topped tilting wind tunnel (Weise and Biging, 1994). A heading fire is when the wind speed and fire spread are in the same direction, and a backing fire is when they are opposed. To ensure fire spread for experiments involving no wind or slope, coarse excelsior made from quaking aspen was uniformly distributed over every experimental setup (Weise and Biging, 1994). Temperature was measured with two rows of nine evenly spaced thermocouples to track fire spread rate. The flames were measured using a video camera, and geometric properties of the photographs were used to calculate flame length. These experiments found that backing fires extinguished themselves most often among all configurations, where the wind cooled the wood ahead of the fire boundary and the ignition temperature could not be reached. They also observed that a heading downslope fire—known as a “sundowner” in southern California—spread faster than a no wind, no slope fire.

Another experiment simulating forest fire growth consisted of a metal tray with an 11 by 9 equispaced (20 mm) grid of compressed paper sticks (tree trunks) with balls of cotton on top (tree crowns) and ravelled cotton on the floor of the tray (underwood) (Buscarino et al., 2015). Since the cotton had high flammability and could absorb liquid,
the rate of spread was controlled using moisture levels of the bottom and top layers. These researchers were simulating the region of reforestation of maritime (Pinus pinaster) or Aleppo (Pinus halepensis) pines and the natural 4:1 average ratio of tree height to spacing. The apparatus was capable of simulating three types of fire: ground, crown, and surface. They modelled fire spread using a reaction-diffusion equation that was a combination of the diffusion PDE in Equation (2.5) with an extra non-linear reaction term on the right-hand side that was responsible for adding (or absorbing) heat from a reaction. They observed the behaviour of the three fire types to choose parameters in the reaction-diffusion equation and did not do any fitting of the model to experimental data.

More recent horizontal-only fire spread experiments used uniform laser-cut cardboard for evenly spaced large-scale fuel beds (Finney et al., 2013, 2017). These fuel beds were burned to observe the structure of fires and measure temperature from an array of 64 thermocouples, two sets of 32 with 1.5 cm and 3 cm spacings. The beds were burned under various wind speeds and temperatures were measured as times series at frequencies of 500 Hz. The authors analyzed linear relationships between flame structure (length, depth) and wind speed. Further experiments with laser-cut cardboard investigated wind and slope effects on fire spread, and temperature and pressure measurements showed that fires exhibited a time-averaged flame burst effect (Finney et al., 2017). Other similar fuel bed experiments used matchstick arrays (Fons, 1946; Vogel and Williams, 1970; Prahl and Tien, 1973) and toothpicks (Wolff et al., 1991). Similar to the experiments with no slope or wind mentioned earlier, these experiments used locally non-uniform fuel types such as excelsior (Catchpole et al., 1998) and pine needles (Rothermel and Anderson, 1966), where thermocouples were used to measure temperature to calculate rate of spreads. Other horizontal fire spread experiments also used gas burners and ponderosa pine lathes to study flame structure (Finney et al., 2015).

A recent study investigated methods to identify coherent streaks of fire in wildfire boundaries (Miller et al., 2017a). Their experiments involved a heated plate in a wind tunnel, where they used an IR camera to detect and track hot streaks on the surface of the plate. They processed their measurements using smoothing techniques to determine
the spacing, length and width of the hot streaks. This same group conducted further research in the area of understanding fire streak structure and its relationship to forest fire smoke plumes (Miller et al., 2017b).

The experiments that inspired our apparatus design were paper burning experiments (Zhang et al., 1992; Maunuksela et al., 1997, 1999). These experiments consisted of burning optical lens tissue paper soaked with potassium nitrate (KNO$_3$, or more commonly known as saltpeter) for flameless, uniform, slow smoldering. They positioned the paper vertically, igniting the paper from below by a straight heating wire to ignite the full length of the bottom edge. They measured the fire spread using four cameras (512x720 pixel videos) to obtain multi-level grayscale images that were later combined using software into a composite image (512x3000) for each frame. The experiments were designed to provide data to fit the Kardar–Parisi–Zhang nonlinear stochastic PDE model (Kardar et al., 1986) to study the kinetic roughening of growing fire fronts. Another example of a vertical experiment involved an array of protruding standard wood matchsticks with the heads removed (Gollner et al., 2012). The spacings between the matchsticks were varied to discover the relationships between spacing and fire front progression (spread rate), time to burnout, and mass-loss rate. The rate of upward flame spread increased as spacing increased, with a dramatic jump between the range of smaller spacings, attributed to convective heat transfer. The final example of vertical burning was an experiment of alternating fuel and insulation layers that was burned to observe how flame height affects burning between insulated vertical fuel beds (Miller and Gollner, 2015).

The predecessor to the work in this dissertation is a wax paper burning apparatus used to study fire spread (Boychuk et al., 2007), which was designed to provide data to be used for an application of the Boychuk model. These experiments had a reduced amount of smoke generated during burning, but experienced flaming that interrupted the signal of burning paper to the camera and caused significant errors during fire spread rate estimation, and they did not have a formal investigation of their apparatus. We develop the design of a micro-fire spread apparatus, scientifically investigate parts of the apparatus, and define the methodology of flameless paper smoldering experiments.

A shared theme of the analysis for all of these experiments was studying how fire
spreads locally. The tendency was to look at the temperature of the fire to get a better understanding of how fire’s structure was related to the local spread rate. By contrast, our fire image measurements from level and sloped burns give insight into fire spread rates and rates’ variabilities over a region.

### 4.2 Fire smoldering apparatus and experimentation

A micro-fire spread apparatus that measures paper smoldering experiments under controlled environmental conditions allows for a mouse model investigation of wildfire spread. The measurements can be processed, using statistical techniques, into different forms of data for the statistical analysis of fire spread mean behaviour and variability. Fire spread rates can be estimated from each smoldering experiment, and those estimates can be used to calibrate a fire growth model, and quantify spread rate estimation and model performance. Each experiment has controlled environmental conditions (e.g. elevation, fuel moisture, surface wind speed, and ignition location) and uncontrolled conditions (e.g. the local inconsistency of the wax paper manufacturing and local smoldering susceptibility from the non-uniform coating of potassium nitrate). The effects of the uncontrolled conditions, such as local inhomogeneity of the wax paper fibers and potassium nitrate coating, simulate the natural local randomness in the fuel bed of a forest fire.

From our experience, an apparatus for studying fire spread must meet the following conditions: First, the apparatus must incorporate both random and deterministic controls that are known within each experiment, and reproducible between experiments. This is accomplished through both apparatus design, incorporating different configurations of the experiment, and experiment procedure. Second, the apparatus must measures the fire spread of a fully smoldering experiment. The measurements are recorded by a camera positioned over the experiment, and the potassium nitrate coating ensures smoldering by regulating oxygen flow to the combustion reaction. The smoldering combustion is desired as large flames block the view of the fire front. Smoke occlusion of the fire front can be mitigated by image processing methods that we have developed. Third, the apparatus must simulate fire spread scenarios with consistent environmental conditions.
A fume-hood provides a fire-safe space with a consistent environment, and a consistent experimental procedure allows for control over fuel moisture, fuel type, and application of a potassium nitrate coating. Lastly, the apparatus must provide novel raw data that can be analyzed to provide insight into fire spread rate estimation, particularly the current challenges of modelling fire spread variability and firebrand spotting.

4.2.1 Apparatus description

Figure 4.1 shows the smoldering apparatus mounted in a laboratory fume hood (H.H. Hawkins Ltd. 111-72PR). The apparatus base is a nonflammable solid anodized aluminum optical breadboard (Newport) with dimensions $12 \times 18 \times 0.5$ inch, with 1/4-20 mounting holes on a 1 inch grid. An optical breadboard allows for consistent and controlled configurations of the experiment, and seamlessly integrates with mounting and camera equipment. Optical breadboard handles (Newport) allow for a fully set-up experiment to be easily moved inside the fume hood for experimentation. A $15.25 \times 10.25 \times 0.75$ inch metal tray is used to contain the smoldering paper experiment. To reduce glare off the bottom of the tray as observed in early experimental trials, the tray is coated in black heat-resistant paint to improve data extraction. The metal tray has holes machined

Figure 4.1: A fire smoldering experiment in action, using a $32^\circ$ slope configuration. The camera position in this figure is not the current standard. Here the camera is pointed orthogonally to the tray, but all experiments considered in this dissertation have the camera pointed orthogonally to the ground.
into the base to allow for mounting on supports, airflow to the smoldering experiments from below, and initiate smoldering by flame from underneath. The tray is supported by nonflammable, passivated stainless steel optical mounting posts (Newport). The camera (Olympus Stylus 600) is mounted on similar posts which hold it 19 inches above the center of the smoldering experiment, and pointed down and orthogonal to the ground. The camera captures $320 \times 240$ pixel images at 15 frames per second, and the spatial resolution of the camera for this experiment is approximately 1.2 mm.

The apparatus can be configured for level and sloped experiments by using different sets of mounting holes on the metal tray with different combinations of mounting post lengths. The apparatus currently has three slope configurations: $0^\circ$, $11^\circ$, and $32^\circ$. The $11^\circ$ angle configuration was chosen to be just slightly different than a horizontal regime, and the $32^\circ$ angle was chosen to be significantly different from the other two configurations. Figure 4.1 shows a smoldering experiment being run with the apparatus in a $32^\circ$ slope configuration.

The ability to set the wind speed was attempted, but this endeavour proved difficult. A fume hood does not provide an ideal environment for blowing wind over the apparatus. There is already a constant slight wind due to a negative pressure pulling smoke towards the back wall. This negative pressure is part of the design of a fume hood, as the pressure prevents harmful fumes from escaping the fume hood into the lab environment. The wind also protects the overheard camera from smoke damage. The fans available are either too large for the fume hood or too powerful at the lowest setting and caused the smolder to extinguish prematurely. We must also consider the turbulence structure of the wind being blown over the experiment. The variety of fans available in the lab provide unknown flow structure, turbulence, and interactions with the fume hood wind. An ideal environment is a ballistics wind tunnel that provides uniform wind conditions with a known turbulence structure on a continuum of speeds and permits smoke from experiments. Therefore, the incorporation of wind as a controllable variable in experiments is future work. We note here that all experiments had identical wind conditions.
4.2.2 Wax paper and potassium nitrate coating

A major factor behind fire spread experimentation that is often discussed in the forest fire experiment research literature is the fuel bed and mitigating environmental effects on the fuel bed (Burrows, 1999a,b; Sullivan, 2009b). When considering fire smoldering regimes, we tested tissue, printer, and wax papers as fuel beds. We have a fuel bed that contains local inhomogeneity of fuel with no discernible macroscopic patterns, is susceptible to local inhomogenous burning, does not require large energy to ignite when coated in potassium nitrate, contrasts white fuel with red burning and black burnt-out areas in measurements, is safe to work with inside a fume hood, and is cost-effective. Ultimately, we chose wax paper since it satisfied all of these criteria and had the added benefits of minimal flame-ups with a potassium nitrate coating, did not absorb an excessive amount of potassium nitrate or water, held its overall structural integrity during preparation and experimentation, and introduced appropriate levels of measurement noise. Additionally, a novel and unexpected outcome of using wax paper—observed by Lindsay Reinholz at UBC Okanagan—was that the wax paper experienced fire spotting behaviour. White paraffin wax paper, commonly available from local stores, is used for all experiments in this dissertation. The wax paper consists of paper fibers coated in paraffin wax. To get a better idea of the local inhomogeneous structure and confirm claims about the microscopic inhomogenous structure, microscopic images of burnt and unburnt coated and uncoated wax paper were taken using scanning electron microscopy.

4.2.3 Scanning electron microscopy of prepared paper samples

The resolving power of a human eye is approximately 0.1 mm, an optical electron microscope is 1 µm, and a scanning electron microscope (SEM) can be less than 1 nm. An SEM scans the surface of an object by irradiating it with a converged electron beam in a vacuum, and utilizes electrons released from near the sample surface, called secondary electrons, to form an image. Wax paper is imaged using a Hitachi S-3400N SEM at the University of Western Ontario’s Integrated Microscopy lab. The wax paper samples are prepared for the SEM by first sputter coating them in an ultra-thin
electrically-conducting gold/palladium layer. The gold/palladium coating is commonly used for imaging non-conductive materials with an SEM, as the metal layer prevents the build-up of charge and increases the production rate of secondary electrons, which in turn increases the signal-to-noise ratio since and amount of information available to form images, and protects the material’s surface.

The local inhomogeneity of wax paper and the non-uniform coating of potassium nitrate are the main sources of variation in the fire smoldering experiments, and simulate the local, fine-scale heterogeneity crucial to a realistic fire spread. The next set of figures show the inhomogeneity of wax paper, with and without the potassium nitrate coating. Figure 4.2 shows the individual paper fibers and wax coating at different microscopic scales. Figure 4.2a shows the local non-uniform thickness and direction of the fibers, and how the wax coats the paper’s surface, at the microscopic scale. Figure 4.2b shows the torn edges of the wax paper, where we can further see the fibers’ non-uniform sizes. These images demonstrate the local inhomogeneity of the wax paper, where paper fibers have no microscopic weaving pattern and do not have a uniform wax coating. Note that the fibers do not appear to have chemically reacted with the wax.

Figure 4.3 shows an edge sample of wax paper with the potassium nitrate coating. Figures 4.3a and 4.3b show that the potassium nitrate coats the fibers and does not appear to react with the wax or fibers. This imagery shows that the potassium nitrate does not react with paper during the coating process. Figure 4.4a shows the thickness of the wax paper through a torn edge. Figure 4.4b shows the thickness of the potassium nitrate coating, and how the potassium nitrate coats the surface of the wax paper and does not penetrate the fibers or wax. This is further evidence that the potassium nitrate does not react with the wax paper because it adheres to the surface of the wax paper.

Next, we look at some burned samples of wax paper, with and without the potassium nitrate coating. Figure 4.5 shows a piece of wax paper which was burned before applying the gold/palladium layer. Figure 4.5a (when compared to Figure 4.2) shows that the structure of the fibers has remained intact through combustion, but the integrity of the paper is compromised and it crumbles. The wax around the paper has melted and burned, as evidenced by the newly formed holes on the surface. The view of the edge in
Figure 4.2: SEM images of uncoated wax paper. Note that for all SEM images, the distance on the scale (e.g. 300 µm) is for the entire length of the scale.

Figure 4.5b shows a similar story where the residual wax has been melted on the wax paper. Figure 4.6 shows a flake of burned wax paper being held by paper fibers to the main body. This shows that even though the area is burnt, the wax binds the burnt paper fibers together. This binding nature of the residual wax after burning is evidence of a firebrand phenomena that was observed in experimentation. The black colour of the
burnt paper suggests leftover carbon from combustion, which we believe makes up the residual fibers after burning.

Figure 4.7 shows the overall structure of a burnt coated wax paper. The main feature of these images is the clumping of the wax and fibers that has formed around a burnt region (shaped like a U) in Figure 4.7a. Figure 4.7b and 4.7c show that the clumps are actually more hollow than they appear macroscopically. The phenomenon here is that the melting of the wax and potassium nitrate causes the burning regions to clump. The oxygen from the melted potassium nitrate and airflow through holes feeds the combustion reaction inside the clump, but there is a lower abundance of oxygen that slows combustion and causes smoke from incomplete combustion. These slow-burning clumps break off and
are carried aloft by local air turbulence, becoming firebrands that could land on unburnt areas away from the fire front. This series of mechanisms creates an environment more likely to experience firebrand spotting. Figure 4.8 shows edge images of burnt coated wax paper. These look very different from burned uncoated wax paper in Figure 4.5b: the flaky, thin edge that we see on uncoated wax paper becomes smooth and wide from clumping wax when the paper is coated before burning. The holes in the paper around
burning areas still allow some oxygen to the combustion reaction, but they are not large enough to allow oxygen for complete combustion, resulting in smoke from smoldering. Figure 4.9 shows how the potassium nitrate crystals that coat the wax paper melt on the surface of the combusting paper.

4.2.4 Challenges in experimentation

Measuring the moisture content of the wax paper is difficult for several reasons. We made attempts to weigh the wax paper to see the effect of drying times, and measure the amount of moisture and potassium nitrate on each sheet. The available weighing scales were not large enough to weigh the paper without folding it, and the folding created creases that affected its uniform inhomogeneity. In fact, this discovery has lead to some experiments using fully crumpled paper, which tended to experience more firebrand spotting than uncrumpled paper. Future experiments should use a moisture meter to determine the
Figure 4.7: SEM images of burned coated wax paper with clumping.

amount of water in the coated wax paper and better understand moisture’s role in fire spread experiments.
Chapter 4. Fire smoldering experiments

Figure 4.8: SEM images of burned coated wax paper with clumps and hole formations.

4.2.5 Smoldering experiments

To date, we have conducted over forty smoldering experiments with various apparatus configurations and environmental conditions; Table 4.1 shows the current number of experiments with each configuration.
Figure 4.9: SEM images of burned coated paper with signs of melting wax.

Table 4.1: Number of completed experiments and usable fire spread measurements.

<table>
<thead>
<tr>
<th>Angle</th>
<th>No. of Experiments</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>Moderate</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Steep</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

Experimental procedure

The experimental set-up begins with cutting a sheet of wax paper to the same size as the tray and soaking the sheet in a 10 mL/g H₂O/KNO₃ solution inside a deep metal tray for 30-60 minutes. Inside the fume hood, the soaked paper is dried using a warming element set to 50-70°F for 20 to 120 minutes. The dry coated wax sheet is attached edge-to-edge onto the tray using paper tape, and the apparatus is assembled and placed inside the fume hood. With the camera recording and pointed down and orthogonal to the ground, the sheet is ignited using a directable butane ignitor from below through the holes in the bed of the tray and the smoldering is closely monitored. The experiment lasts for 40 to 60 seconds and ceases when the experimenter extinguishes the flame with water from a spray bottle.

The videos from each experiment are cropped to remove any boundary areas that are not part of the smoldering area (e.g. edges of the tray), which prevents any undesirable false burning recognition or boundary effects during image processing. The videos are cropped in time to only include start to finish of each smoldering experiment. Addi-
tionally, frames are removed when the flame from the ignitor appears in the video, as those frames cause error in the “baseline” for processing. Finally, videos are segmented such that we only keep the first frame of each second. This introduces uncertainty of fire spread between images, similar to the time between satellite images of a forest fire. Figure 4.10 shows one fire smoldering experiment at three points in time after ignition. Notice that smoke occlusion increases as the amount of smoldering wax paper increases. The uneven coating of potassium nitrate can be seen as white discolorations on the paper. These white spots show inhomogeneous oxidation control of burning areas over local smoldering regions of the fire front. With a plethora of raw measurement data in hand, we process the measurements into useful data formats.

### 4.3 Processing measurements into fire spread data

In this section, we describe a procedure to process the raw fire spread measurements into fire spread data. The images are denoised while preserving fire spread boundaries using anisotropic filtering and smoothing methods. This section employs the PDE-based anisotropic diffusion filtering, and isotropic and anisotropic smoothing methods from Chapters 2 and 3 to process fire spread images.

#### 4.3.1 Raw measurements

We explore the nature of the collected data, and apply methods to smooth it without imposing any bias or making any assumptions about its distribution. Figure 4.11 shows a raw unprocessed image and a visualization of each channel that demonstrates the variability over space. The RGB colour channels are similar throughout, except around the burning regions where the red channel differs substantially from the green and blue channels around the burning region. The channel differences are a result of the physical process of a fire. When the white paper changes from fuel to burning, the camera reads it as a decrease in blue and green and an increase in red. When the burning paper extinguishes, the red channel decreases.

The difference between each of the channels in Figure 4.12 shows that noise is cor-
related at the individual pixel level across RGB channels. The noise from pixel to pixel is likely due to the camera’s light detector and how it gathers information. The noise from measurement is insubstantial compared to the “noise” of the colours on the wax paper—that is the colour of the wax paper is uniform in the underlying data generating
Figure 4.11: A fire spread image from ten seconds into the 11° experiment 2. The individual RGB channel values are shown separately next to the original image. The z-axis is the RGB channel value, and (x,y)-directions are the pixel locations of the image.

process, but noisy due to the local inconsistencies in the construction of the paper and the potassium nitrate coating.

Figure 4.13 shows the channel value of each RGB channel for one pixel over the course of the entire video. The pixel’s RGB channels are not stationary: there is a clear change in the mean value over time at ignition and extinction events, and there is a change in variance when the mean value changes. We also see an aberration, where the value jumps
Chapter 4. Fire smoldering experiments

(a) Red minus green

(b) Red minus blue

(c) Green minus red

Figure 4.12: The differences of the RGB channels for the fire spread image from ten seconds into the $11^\circ$ slope experiment 2. The $z$-axis is the difference in the RGB channel values, and $(x, y)$-directions are the pixel locations of the image.

up to 200, after the ignition and extinction time change-points. Figure 4.14 shows the red channel minus the green channel from Figure 4.13. The events of ignition and extinction are contained within the spike, however those two event times are harder to determine in the differenced image than when each colour channel is considered separately. To estimate the underlying data generating process of the fire spread images, we smooth the surfaces in Figure 4.11 using anisotropic diffusion and local constant anisotropic smoothing such that we preserve the boundaries between the fuel, burning, and burnt-out regions of the fire image.
Figure 4.13: The red, green, and blue channel values of a single pixel over the course of the 11° slope experiment 2.

4.3.2 Anisotropic diffusion for image processing

An example of calculating an iterative solution for the anisotropic diffusion filter on a fire image (in Figure 4.15a) is shown in Figure 4.15b, and the histories of each of the RGB channels are shown in Figure 4.16. The images appear to be blurred, and we visually identify three distinct fairly uniform colour regions of fuel (grey), burning (red), and burnt-out (black). The boundaries between regions are preserved, but are interrupted by smoke occlusion and oversmoothed by anisotropic estimators. The smoke occlusion increases as time increases since there is a larger burning area to produce more smoke. Figure 4.16 shows the result of anisotropic diffusion on each of the RGB channel values in the right column, next to the original fire images in the left column. We see how the
4.3.3 *Anisotropic local constant smoothing for image processing*

In this section, anisotropic smoothing is applied to fire spread measurements to preserve edges and smooth between the boundaries. For full details on the methods of anisotropic local constant smoothing, see Chapter 3, and for processing simulated fire spread, see Section 3.3.4. Figure 4.17 shows the isotropic and anisotropic local constant estimators applied to the red channel of Figure 4.15a. This figure demonstrates the difference in two smoothers, where one accounts for change-points in the regression function. The
isotropic estimator oversmooths the jump, while the anisotropic estimator captures the
two boundaries between the fuel, burning, and burnt-out regions.

Figure 4.18 shows a pixel’s RGB channel values after spatially smoothing each fire
image of a segmented video; we are smoothing the fire images over space at each time
point. This figure compared to Figure 4.13 shows the difference in the mean values before
and after the change-points in the pixel values post spatial smoothing. The variance
around the estimated values is increased before and after the change points, and the
aberrant point at time 38 has been smoothed out. The difference between red and
green channels is shown in Figure 4.19 and the peak is more gradual than in Figure
4.14. This figure demonstrates that the fire image smoothing is not smoothing out the
relationship between channels, except near the change-point boundaries where we are
trying to capture the features in the data.
Figure 4.16: The result for the RGB spectrum after applying image weighting and anisotropic diffusion on a smoldering image. The anisotropic diffusion procedure is 50 tensor field calculation iterations with 10 smooths per iteration. The $z$-axis is the RGB channel value, and $(x, y)$-directions are the pixel locations of the image.

### 4.4 Discussion

The fire smoldering experiments described in this research are similar to the experiments performed by Zhang et al. (1992), with a few key differences. First, wax paper is used in place of optical lens cleaning tissue. Wax’s hydrophobic nature prevents large amounts of moisture from being retained in the material during soaking, which decreases preparation time and reduces the amount of potassium nitrate bound within the paper fibers. Our
Chapter 4. Fire smoldering experiments

(a) LC smoothing  
(b) LC residuals

c) ALC smoothing  
(d) ALC residuals

Figure 4.17: The result for the red channel value after applying isotropic and anisotropic local constant estimators on a smoldering image. The bandwidths for the isotropic local constant smoother is selected using least-squares AIC, and the anisotropic range kernel is selected using the ad hoc procedure. The $z$-axis is the RGB channel value, and $(x, y)$-directions are the pixel locations of the image.

Preliminary experiments showed that excessive potassium nitrate deposits cause premature extinction of smoldering. Second, previous videos (Zhang et al., 1992) were recorded using a higher resolution camera, with an effective pixel grid of 512 × 3000 that is much larger than our pixel grid of 320 × 240. Lastly, the purpose of their videos was to investigate an interface growth model of a linear fire front, whereas we are interested in determining fire spread rate distributions over a fire that burns radially outward from a point of ignition. Our work is similar to the processing and analysis of infrared fire streak data studied by Miller et al. (2017b), where they were interested in identifying specific regions of the data. Our methods could be applied to smooth those streak regions and find the change-point boundaries of flame and non-flame.

An issue throughout this section is smoke occlusion. To remove occlusion caused by smoke so that we can better identify fuel, burning, and burnt-out regions, images
Figure 4.18: The red, green, and blue channel values for a single pixel over the course of a video that has been smoothed using anistropic local constant estimator.

could be weighted by the previous time stamp image. An area occluded by smoke in the current image may not be occluded in the previous image. Therefore, burnt-out aberrant points in a pixel’s channel values would have a value close to black when re-weighted. This was performed on some videos with mixed results, however the optimal re-weighting of images was not thoroughly pursued. The image re-weighting causes burning regions to appear pre-smoothed, this caused more errors when identifying ignition times than removing smoke occlusion remedied.

Firebrand spotting is caused by the clumping of wax paper into small firebrands that were carried aloft by local turbulence was a regular occurrence in the videos. They initially caused some difficulty during smoothing in Chapter 5, until we relaxed our assumption that fire spreads out in a strict cone shape. The observation of firebrand
Figure 4.19: The difference of the red and green channel values for a single pixel over the course of an entire video that has been smoothed using anistropic local constant estimator.

spotting was unexpected, and the observed phenomena provides another area of future study for this fire spread apparatus.

Fire spread rates can be estimated from each smoldering experiment’s processed measurements, and those estimates can be used to calibrate a fire growth model, and quantify model and spread estimation performance. Each experiment is conducted with controlled conditions, such as elevation, surface wind speed, and ignition location, and uncontrolled conditions, such as local inconsistency of the wax paper manufacturing and local smoldering susceptibility from the non-uniform coating of the potassium nitrate. The effects of the controlled conditions are observed in the directional rates of burning; simply, fire spreads faster when it moves uphill and slower downhill. The uncontrolled conditions
attempt to capture the effects of local inhomogeneity of fuels that introduce uncertainties into forest fire spread. We next use the fire spread data processed in this chapter to estimate spread rates for each experiment.
Chapter 5

Fire spread rate estimation and analysis

The study of wildfire behaviour includes fire occurrence modelling, atmospheric science modelling, fire growth modelling, and much more. In fire growth modelling, researchers consider environmental and human factors that contribute to a fire’s ignition, growth, and extinction. Every fire growth model is governed by parameters, directly or indirectly related to fire spread rates, such as the Prometheus and Boychuk models that use the FBP and Behave systems (discussed in Section 2.4). Our goal is to develop methods to estimate fire spread rates from sequences of images extracted from a video, and analyze their mean function and variability.

The processing of fire spread measurements from Chapter 4 provides data to estimate the fire spread rates of the smoldering experiments. To estimate fire spread rates, we start by estimating the ignition and extinction event times of a location in an image with minimal inputs and supervision. The event time estimates indicate when the fire front reaches that location, and how long that location burns. We estimate ignition and extinction times using a two-step procedure. In the first step, we apply data sharpening and anisotropic smoothing techniques to estimate the ignition and extinction event times for each pixel of a fire video. In the second step, we smooth the ignition and extinction event times using the local constant kernel estimator. With the estimated event times at each location, we estimate the instantaneous and average fire spread rates for each video.
Chapter 5. Fire spread rate estimation and analysis

We integrate the scientific understanding of fire spread into our spread rate calculations and investigate a regression relationship that models the effects of slope on average spread rates. We investigate the effect of slope on the the mean and variability of instantaneous fire spread rates. Lastly, we look at the distributions of residency times for burning cells.

5.1 Estimators for ignition and extinction times

The evolution of an individual pixel over time contains events that indicate when a pixel changes from fuel to burning, and from burning to burnt-out. The first method to estimate ignition and extinction event times is a data sharpening algorithm that uses quadratic programming with a set of constraints uniquely designed for smoldering fire spread data. The second method is to use anisotropic smoothing that retains the sudden change in the regression function at ignition and extinction event times. The estimation of pixel ignition and extinction event times allows us to calculate estimations of spread rates and their variability over the course of a fire smoldering video. The processed fire spread measurements have been denoised such that the fuel, burning, and burnt-out regions have been smoothed and the boundaries between the regions are preserved. Next, we estimate when a pixel is ignited or extinguished using data sharpening.

5.1.1 Data sharpening for event time estimation

In this section, data sharpening from Section 2.3 is used to estimate ignition and extinction event times. For this data sharpening algorithm, the sharpening distance to be minimized is the sum of the squares of the Euclidean distance \( \sum_{j=1}^{k} (r_j - r_j^*)^2 \) between the red channel value of a pixel \( r_j = r_j(x,y) \in [0, 255] \) and the sharpened data \( r_j^* = r_j^*(x,y) \in [0, 255] \). The sum is over the number of frames \( k \) in each fire video. The distance metric is minimized with respect to ignition and extinction times that separate fuel, burning, and burnt-out states of a pixel. Figure 5.1 shows an example of the evolution of an individual pixel’s red channel over the course of a fire. The pixel’s state transitions from fuel (green circle) to burning (red circle) at the ignition time \( t_I \). The pixel’s state transitions from burning to burnt-out (black circle) at the extinction time \( t_E \).
These times are unknown and we can use data sharpening to estimate them by writing the pixel states of Figure 5.1 as linear constraints on the individual pixel at \((x, y)\) using

\[
\text{Red} \begin{cases} 
\text{Fuel:} & |r_j^* - c_1| < \epsilon, \quad 1 \leq t_j \leq t_I, \\
\text{Burning:} & r_j^* \geq c_1 - \epsilon, \quad t_I \leq t_j \leq t_E, \\
\text{Burnt-out:} & r_j^* \leq \epsilon, \quad t_j \geq t_E,
\end{cases}
\tag{5.1}
\]

where \(t_j\) is the recorded time at frame \(j\). For our segmentation resolution, \(t_j = j\) seconds but this does not need to be true in general. The constraints are all linear inequalities and the minimization problem is solved using quadratic programming. The algorithm for this set of data sharpening constraints is

**Algorithm 5.1:** Data sharpening red channel’s constraints

**Data:** Processed fire measurements, \(\epsilon\), and \(c_1\)

**Result:** Data sharpened ignition and extinction event time estimates

**Initialization:**

\[
\text{for each pixel with location } (x, y) \text{ and red channel values } r_j = r_j(x, y) \text{ such that } j \in \{1, \ldots, k\}
\]

\[
\text{for each possible } t_I \in \{1, 2, \ldots, k - 1\}
\]

\[
\text{for each possible } t_E \in \{t_I + 1, t_I + 2, \ldots, k\}
\]

\[
\text{construct the constraints according to Equation (5.1);}
\]

\[
\text{find new } r_j^*(x) \text{ using quadratic programming (Turlach and Weingessel, 2013)};
\]

\[
\text{if } \sum_{j=1}^{k} (r_j(x, y) - r_j^*(x, y))^2 \text{ is smaller than the previous data sharpened solution}
\]

\[
\text{record the new } r_j^*(x, y) \text{ as the data sharpened solution;}
\]

\[
\text{record ignition and extinction time estimates for } r_j(x, y) \text{ from } r_j^*(x, y)
\]

\[
\text{according to the constraints in Equation (5.1)};
\]

Figure 5.2 shows another example of the evolution of an individual pixel’s red and green channels over the course of a fire. The green channel changes at the time of ignition and red changes at the time of extinction. We write the regions of Figure 5.2 as linear
constraints on individual pixel at \((x, y)\) using

\[
\begin{align*}
\text{Red} \quad \left\{ \begin{array}{l}
\text{Fuel & burning:} \quad |r_j^* - c_1| < \epsilon, \quad 1 \leq t_j \leq t_E, \\
\text{Burnt-out:} \quad r_j^* \leq \epsilon, \quad t_j \geq t_E,
\end{array} \right. \\
\text{Green} \quad \left\{ \begin{array}{l}
\text{Fuel:} \quad |g_j^* - c_1| < \epsilon, \quad 1 \leq t_j \leq t_I, \\
\text{Burning & burnt-out:} \quad g_j^* \leq \epsilon, \quad t_j \geq t_I,
\end{array} \right.
\end{align*}
\]

(5.2)

(5.3)

where \(g\) represents the green channel value. For this algorithm, the parameter \(c_1 = c_1(x, y) = r_1(x, y) + 70\), where \(r_1(x, y)\) is the first red value of a pixel at \((x, y)\), and \(\epsilon = 10\).

![Figure 5.1: An example of a data sharpened result for the evolution over time of a single pixel’s red channel value.](image)

To find the optimal ignition and extinction times, the quadratic objective function is
Figure 5.2: An example of a data sharpened result for the evolution over time of a single pixel’s red and green channel values simultaneously.

minimized separately for each colour channel’s constraint set. The algorithm for this set
of data sharpening constraints is

**Algorithm 5.2: Data sharpening for red and green channels’ constraints**

**Data:** Processed fire measurements, $\epsilon$, and $c_1$

**Result:** Data sharpened ignition and extinction event time estimates

initialization;

for each pixel with location $(x, y)$ and red channel values $r_j = r_j(x, y)$ such that $j \in \{1, \ldots, k\}$

for each possible $t_I \in \{1, 2, \ldots, k - 1\}$

for each possible $t_E \in \{t_I + 1, t_I + 2, \ldots, k\}$

construct the constraints according to Equation (5.2);

find new $r^*_j(x)$ using quadratic programming (Turlach and Weingessel, 2013);

if $\sum_{j=1}^{k}(r_j(x, y) - r^*_j(x, y))^2$ is smaller than the previous data sharpened solution

record the new $r^*_j(x, y)$ as the data sharpened solution;

record extinction time estimate for $r_j(x, y)$ from $r^*_j(x, y)$ according to the constraints in Equation (5.2);

for each pixel at $(x, y)$ with green channel values $g_j = g_j(x, y)$ such that $j \in \{1, \ldots, k\}$

for each possible $t_I \in \{1, 2, \ldots, k - 1\}$

for each possible $t_E \in \{t_I + 1, t_I + 2, \ldots, k\}$

construct the constraints according to Equation (5.3);

find new $g^*_j(x)$ using quadratic programming (Turlach and Weingessel, 2013);

if $\sum_{j=1}^{k}(g_j(x, y) - g^*_j(x, y))^2$ is smaller than the previous data sharpened solution

record the new $g^*_j(x, y)$ as the data sharpened solution;

record ignition time estimate for $g_j(x, y)$ from $g^*_j(x, y)$ according to the constraints in Equation (5.3);

These two algorithms yield sharpened data $r^*_j$ that is minimum of the objective function for their constraint sets. However, we are not directly interested in the optimal
values of the sharpened data $r_j^*$, but rather the values of $t_I$ and $t_E$ that yield the optimal values of the large changes in $r_j^*$.

Figure 5.3 shows a perspective plot of the estimated extinction times from the fire video processed in Section 4.3. The area of fire ignition starts around a point near the base of the event time surface and spreads as a warped cone as time increases. There are areas where we are getting unusable estimates of ignition and extinction times. Evidence of this poor performance are the vertical regions where estimated ignition and extinction times of neighbouring pixels are relatively large, with some greater than 10 seconds. We know this is an error, where we can watch the fire spread in the raw videos, but these vertical regions in the estimates suggest no spread. Inspection of the channel values over time at these neighbouring pixels reveal the inaccuracy in estimation is caused by smoke occlusion—the amount of smoke increases as the amount of smoldering material increases—that interrupts the signal from the true image to the camera. We also found that the constraints associated with the two colour channels in Equations (5.2) and (5.3) perform similarly for estimating ignition times as the single colour channel constraints in Equation (5.1), but perform worse at estimating extinction times.

We note that the ignition times estimated by data sharpening have substantial systemic errors inside of the burning regions in some videos. This problem is mitigated through the intelligent choice of $c_1$. Using static ($c_1$ is constant for each pixel of a video) or adaptive ($c_1$ changes for each pixel of a video) parameter values both yield results that contained regions of ignition and extinction time with substantial errors, rendering the results unusable for analysis. We find that choosing the first pixel value in the time sequence plus a constant between 70 to 100, i.e. $c_1(x, y) = r_1(x, y) + 70$, yield an improved estimate. The sources of error of this estimation procedure arise from systemic problems that include smoke occlusion, flare-ups, ambient laboratory light, and automatic internal adjustments of the camera such as re-focusing. Again, we believe the error occurs during data measurement and not the data sharpening algorithm. The data sharpening method shows consistent results throughout the image except in small regions. Even with these errors present, we have demonstrated that our methods are fairly robust at estimating ignition and extinction times. To compare with this method, we look at finding event
times using anisotropic smoothing.

5.1.2 Anisotropic smoothing for event time estimation

A second method to estimate ignition and extinction event times is to smooth each pixel’s channel values over time and to search for peaks, valleys, and large changes. Each pixel’s channel values are interpreted as a smooth function over time, although the channel values and recorded times are discrete. We estimate the pixel value function using isotropic and anisotropic methods and then search for features that indicate ignition and extinction events. Figure 5.4 shows the red channel value for one pixel’s signal, the isotropic $\hat{g}_{isc}(t)$ and anisotropic $\hat{g}_{alc}(t)$ smoothers, and a vertical red line that indicates the ignition time $t_I$ is calculated from the the maximizer of $\hat{g}_{alc}(t)$. There is a sudden change in the pixel’s
value at the indicated ignition time, which is captured by the anisotropic smoother. When compared with other estimation approaches to infer ignition time, we found that maximizing the anisotropic smooth of a pixel’s red channel value contained the least number of errors and captured the ignition of each pixel through a visual evaluation.

Figure 5.4: A single pixel’s red channel value over time, smoothed using isotropic and anisotropic smoothers. The vertical red dotted line represents the ignition event time estimated for this pixel.

Figure 5.5 shows one example where the technique of maximizing the anisotropic smooth was employed. The estimation contains some systemic error from smoke occlusion, where there are areas of ignition at time 1 (dark blue) that do not occur in the fire spread video. This error is not caused by the deployment of anisotropic smoothing, but rather because the pixels having no discernible change-point signal in their data over the course of the fire. The isotropic smoothing result performed similarly or worse in
these areas. Other videos induce similar errors estimating ignition and extinction event times near or at the ignition point of the fire, but those areas could be corrected since fire videos showed those pixels were ignited at the beginning of the video.

Figure 5.5: The ignition event times estimated using anisotropic smoothing for 11° slope fire experiment 2. The z-axis is time, and (x, y)-axes are the pixel coordinates.

5.1.3 Smoothing ignition time surfaces and estimating instantaneous spread rates

Figures 5.3 and 5.5 show the estimated ignition times of each pixel location of a fire video. The discrete format of the ignition time estimates cannot be used to calculate the instantaneous spread rates at each pixel location. The functional form of the ignition time surface across space is not known, but the warped cone shape of ignition times is found across all videos. We know that the ignition of a pixel must occur during the time interval between being a fuel cell and burning. To estimate the data generating process
behind ignition times, we use the local constant estimator in Equation (2.2) from Section 2.1. This allows us to estimate the ignition and extinction time surface between the pixel locations. The derivative of this estimate yields the instantaneous fire spread rate at each pixel location. Figure 5.6 shows the local constant estimate of the data from Figure 5.5, which demonstrates how the “roughness” from the discrete data is smoothed.

Figure 5.6: Using the \texttt{np} package’s (Hayfield and Racine, 2008; R Core Team, 2018) local constant estimator with least-squares cross-validation (Hurvich et al., 1998) for bandwidth selection on the ignition time data in Figure 5.5. The \( z \)-axis is time, and \((x, y)\)-axes are the pixel coordinates.

Instantaneous spread rates are calculated by taking the derivative of Equation (2.2), given by

\[
\frac{\partial}{\partial x} \hat{g}(x) = \frac{1}{h \sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right)} \left[ - \sum_{i=1}^{n} Y_i K' \left( \frac{X_i - x}{h} \right) + \frac{\sum_{i=1}^{n} Y_i K \left( \frac{X_i - x}{h} \right) \sum_{j=1}^{n} K' \left( \frac{X_j - x}{h} \right)}{\sum_{j=1}^{n} K \left( \frac{X_j - x}{h} \right)} \right].
\]

Figure 5.7 shows a filled contour plot of Figure 5.6, and the vector field of the estimator’s
gradient $\frac{\partial}{\partial x} \hat{g}(x)$ in the level plot. These are not spread rates, but rather reciprocals $1/\frac{\partial}{\partial x} \hat{g}(x)$ are the spread rates. For the vector field, this implies that the shorter the arrow, the higher the spread rate. The areas that are never burned around the fire spread area are reporting very large spread rates. These areas are not the areas of interest. See Appendix B.1 for the ignition time surface and spread rate plots for all experiments. In those plots, you can see a visual change in shape of the fire spread which indicates a difference in fire spread rate. We next develop the statistical methodology to estimate the differences in spread rates within and between fires.

## 5.2 Fire spread rates and variability

### 5.2.1 Average fire spread rate estimation

In this section, we explore methods to analyze the smoothed ignition time estimates from Section 5.1. We then develop a regression model to calculate fire spread rate estimates and variability from the estimated ignition. We calculated spread rate estimates using our models in regions where the fire spread rate has stabilized, similar to the conditions for the FBP and BEHAVE formulas.

Suppose $d(t_i|\theta_i)$ is a random variable representing the distance of the fire front from the estimated ignition location, $t_i$ is the ignition time random variable, and $\theta_i$ is the angle relative to the horizontal. The ignition location is estimated by using a centre of mass approach for each pixel burning at time 0. For the analysis in this section, we consider only the sets of observations $i$ such that they are directly uphill ($\{i \in 1, \ldots, n|\theta_i = \pi\}$) or downhill ($\{i \in 1, \ldots, n|\theta_i = 0\}$). We estimate the average rate of spread from a regression model given by

$$d(t_i|\theta_i) = ROS(\theta_i) \times t_i \times \epsilon_i,$$

$$\log(d(t_i|\theta_i)) = \log(ROS(\theta_i)) + \log(t_i) + \epsilon_i.$$ 

We transform this equation into a linear regression of the explanatory variable and we are able to interpret the intercept as the logarithm of the rate of spread. If we consider that
the slope has three discrete values, then we treat them as categories to determine if slope has a significant effect on the ROS. We consider only the directions “East” and “West”, which are downhill and uphill respectively on the sloped experiments. The regression model is

$$\log(d_i) = a + bI(\text{severe slope}) + cI(\text{slight slope}) + d\log(t_i) + \epsilon_i.$$ (5.4)

Table 5.1 shows the estimated parameters for the regression model in Equation (5.4) fitted to the uphill data. The uphill spread rates for the 11° slope experiments and the 0° slope experiments are not significantly different, while the 32° slope rate of spread uphill is significantly different from the 0° slope. Table 5.2 shows the estimated parameters for the regression model in Equation (5.4), where only the 11° slope experiments’ spread rates are significantly different from the 0° and 32° slopes’ spread rates.

Table 5.1: Parameter estimates for the uphill (West) direction.

| Parameter | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------|----------|------------|---------|----------|
| a         | 1.8143   | 0.0508     | 35.74   | 0.0000   |
| b         | 0.2861   | 0.0224     | 12.77   | 0.0000   |
| c         | -0.0267  | 0.0211     | -1.26   | 0.2074   |
| d         | 0.5457   | 0.0148     | 36.83   | 0.0000   |

Table 5.2: Parameter estimates for the downhill (right) direction.

| Parameter | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------|----------|------------|---------|----------|
| a         | 2.5470   | 0.0419     | 60.84   | 0.0000   |
| b         | 0.0081   | 0.0208     | 0.39    | 0.6972   |
| c         | -0.1549  | 0.0181     | -8.55   | 0.0000   |
| d         | 0.4107   | 0.0120     | 34.34   | 0.0000   |

5.2.2 Spread rate variability

Figure 5.8 shows the estimates of the instantaneous spread rates West (uphill) and East (downhill) of the origin of ignition for the 0° slope experiments. This figure (and all similar figures) shows the stable spread rate regions, where the regions are spatially different for all slopes in all directions. The instantaneous spreads for West and East
are similar, and estimates show a global sample average of 1.620 and 1.778 for West and East, respectively. The global sample variance is a little higher for East, where the global sample variances for West and East are 1.901 and 2.907.

Figure 5.9 shows the estimates of the instantaneous spread rates West (uphill) and East (downhill) of the origin of ignition for each fire of the 11° slope experiments. The figure shows that the uphill spread is somewhat similar to the down hill, but the spread rates are more volatile. The downhill slope is much less volatile than the uphill for zero or 11° slope experiment rates of spread. The instantaneous spreads for West are greater than East, and estimates show a global sample average of 1.760 and 1.135 for West and East, respectively. The global sample variances are very different, where the global sample variances 1.767 and 0.1010 for West and East, respectively.

Figure 5.10 shows the estimates of the instantaneous spread rates West (uphill) and East (downhill) of the origin of ignition for each fire of the 32° slope experiments. The figure shows that the uphill spread for a 32° slope has a larger spread rate and volatility than the downhill. The instantaneous spreads for West are greater than for East, and estimates show a global sample average of 2.614 and 1.016 for West and East, respectively. The global sample variances are very different: 4.078 and 0.2460 for West and East, respectively.

5.3 Residency time from ignition to extinction

Cellular automaton models, such as the Boychuk model, require inputs of the rate of ignition with respect to time, but they also require the rate of extinction for the residency time of a burning cell to change to burnt-out. The residency times are assumed to be exponentially distributed with a constant rate over time and space. We estimate the residency times for each pixel from the smoothed ignition and extinction times, estimated using data sharpening.

The videos considered in this analysis are 11° slope experiments 1 to 4, 32° slope experiments 1 to 3, and 0° slope experiments 1 to 3 and 6 to 10 (where excluded videos contained many extinction estimation errors near the origin). To estimate the residency
time, we calculate the difference between the smoothed ignition and extinction times for each slope configuration. All estimated residency times are shown in Figure 5.11. The plot shows the residency times with an overlaid gamma distribution marked by a blue line, using maximum likelihood for parameter estimates. The parameter estimates are in Table 5.3, where we see that the shape parameters are all similar, and the rate parameters are decreasing. This implies that the residency time maybe be decreasing as slope increases. However, using the full residency time surfaces may not be appropriate.

Table 5.3: The maximum likelihood parameter estimates of a gamma distribution for the three sloped configuration residency times.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$0^\circ$ slope</th>
<th>$11^\circ$ slope</th>
<th>$32^\circ$ slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>1.541</td>
<td>1.598</td>
<td>1.543</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>2.4e-05</td>
<td>5.6e-05</td>
<td>6.1e-05</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.5997</td>
<td>0.4804</td>
<td>0.4783</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>5e-06</td>
<td>7e-06</td>
<td>8e-06</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>164143</td>
<td>76544</td>
<td>65003</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>2.57</td>
<td>3.326</td>
<td>3.226</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>4.285</td>
<td>6.924</td>
<td>6.745</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>4.285</td>
<td>6.924</td>
<td>6.745</td>
</tr>
</tbody>
</table>

In Section 5.1.1, we comment on how the estimation procedure was poor for areas of the video where there is increased smoke occlusion. To avoid these areas, we consider the residency time on a $21 \times 21$ pixel grid centered around the ignition point of a fire spread in Figure 5.12. The centre of the ignition is calculated using a spatially weighted average of the estimated minimum ignition times. This figure is similar to Figure 5.11, except that there are very few values below 1 second for the $11^\circ$ and $32^\circ$ slope residency times. In Table 5.4, we see a difference in parameters estimated from residency times of the whole surface in Table 5.3. These parameters suggest that increasing the slope causes the estimated spread rate to increase, while causing the estimated expected residency time and variance to decrease.
Table 5.4: The maximum likelihood parameter estimates of a gamma distribution for the three sloped configuration residency times of a $21 \times 21$ grid centered around the ignition point.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$0^\circ$ slope</th>
<th>$11^\circ$ slope</th>
<th>$32^\circ$ slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>2.105</td>
<td>3.766</td>
<td>5.324</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.8007</td>
<td>1.9369</td>
<td>2.5714</td>
</tr>
<tr>
<td>$\hat{Var}(\alpha)$</td>
<td>0.002281</td>
<td>0.015781</td>
<td>0.040341</td>
</tr>
<tr>
<td>$\hat{Var}(\beta)$</td>
<td>0.00042</td>
<td>0.004777</td>
<td>0.010349</td>
</tr>
<tr>
<td>$n$</td>
<td>3372</td>
<td>1653</td>
<td>1323</td>
</tr>
<tr>
<td>$E[R]$</td>
<td>2.629</td>
<td>1.944</td>
<td>2.07</td>
</tr>
<tr>
<td>$\hat{Var}(R)$</td>
<td>3.283</td>
<td>1.004</td>
<td>0.805</td>
</tr>
</tbody>
</table>

5.4 Discussion

To summarize, the procedure presented in this chapter is: (1) apply data sharpening or the anisotropic local constant estimator each RGB channel value of each pixel from the processed fire spread measurements from Chapter 4, (2) estimate the ignition and extinction event times from the smoothed pixel, (3) smooth the event times spatially to estimate intra-pixel times, and (4) analyze the smoothed event times to estimate instantaneous and average fire spread rates, and residency times.

The data sharpening and anisotropic methods used to estimate ignition and extinction event times do not come without some issues. The anisotropic smoothing method of estimating ignition is more robust than data sharpening, as it captures the sudden change in the pixel channel values over time. However, difficulty arises when trying to estimate extinction times. The derivative of the anisotropic smoother could be used to estimate extinction times, but we do not have a satisfactory estimate of a extinction event using the derivative of the estimator. The gradient of the isotropic smoother did not yield useful results. The data sharpening approach appears to work better for extinction, except for large regions of poor estimation caused by smoke occlusion.

One of the distinguishing features of our work in the area of smoothing fire spread is that it specifically smooths real fire spread data, whereas previous work smooths simulated data from the Prometheus fire spread model (Garcia et al., 2008). There are not many data sets that are available to estimate spread rate variability over a fire from
satellite imagery of fire spread using a smoothing approach. We hope the work in this dissertation will motivate the experimental and satellite measurements of fire spread to create fire spread data sets for analysis.

Originally, we considered each fire spread to be a warping cone-shaped pattern, and we designed our data sharpening algorithm to include this restriction in our estimates for ignition and extinction times. Upon inspection of our data, we observed considerable numbers of firebrand spotting events that violated this restriction, and surmised that the restriction was not a realistic representation of fire spread. We relaxed this assumption in favour of allowing our estimation procedure to realize the shape of the fire spread.

The data sharpening algorithm can be viewed as change-point location estimation. We omit a lengthy and in-depth literature review on change-point location estimation, as we are not evaluating procedure against other methods. We are interested in the change-point estimation task at hand, and while many methods exist for estimating change-point location, none are specifically designed for this task. Additionally, there is no literature that uses data sharpening as a method to estimate change-point locations and therefore the application of these methods are novel. Therefore, we submit that data sharpening can be used as an effective change-point location estimator in this setting.

The Boychuk model for fire spread assumes that the residency times for burning are exponentially distributed. If the shape parameter of a gamma distribution is equal to 1, then the gamma distribution simplifies to an exponential distribution with the same rate parameter. The shape parameters of the estimated gamma distributions for all slope configurations within and between Tables 5.3 and 5.4 are clearly different than 1, and the shapes of the densities in Figures 5.11 and 5.12 are not exponential but better approximated by the gamma distribution. This is not an issue with the Boychuk model’s assumption, but a subtle issue with the grid size of the images. If a pixel’s area is too large (the grid is not fine enough), the grid points behave like the sum of exponential random variables, which is a gamma distribution (Braun and Woolford, 2013). If the grid were the correct size for exponentials, then our maximum likelihood estimate of the shape parameter should be approximately 1. We believe the grid size also has an effect on the rate parameter, where the rate would increase if we could make the grid more fine.
Another issue with the fire spread rate and residency time estimates is sample size. Many fire spread experiments, once analyzed, contains too many errors resulting from smoke occlusion. The sample size for analysis is nine $0^\circ$, four $11^\circ$, and three $32^\circ$ slope experiments. More experiments need to be completed to provide enough data to accurately estimate spread rates and residency times. We believe that the techniques proposed here can provide good accuracy, given more data.

When extracting the data from the videos, the choice of 1 frame per second was arbitrarily made, and it could be viewed as suboptimal from an image-processing or time series analytic viewpoint, since sampling at the highest possible frequency is necessary to obtain information about high frequency signals. However, we are attempting to find a methodology that will be ultimately be applicable to satellite imagery where there will inevitably be low temporal resolution. We also did not consider other statistical frameworks such as a time series (Hamilton, 1994) or kriging models (Gelfand et al., 2010) to analyze single pixel channel values over time, as we already have a natural and effective change-point estimator. In this chapter, we have demonstrated two effective methods for estimating ignition and extinction event times. Data sharpening is effective for estimating ignition and extinction event times for videos with low smoke occlusion. Anisotropic diffusion is better for estimating ignition event times, but lacks a reconcilable method for extinction times. We have used the estimated ignition and extinction event times to estimate instantaneous fire spread rates from an isotropic local constant estimator, as well as stable average fire spread rates using a stochastic regression model. We have estimated the residency times of burning cells and shown that the residency times can be approximated using a gamma distribution.
Figure 5.7: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 2. The colour gradient is time in the first plot, and magnitude of the gradient in the second.
Figure 5.8: Uphill and downhill instantaneous spread rates for $0^\circ$ slope experiments.
Figure 5.9: Uphill and downhill instantaneous spread rates for 11° slope experiments.
Figure 5.10: Uphill and downhill instantaneous spread rates for 32° slope experiments.
Figure 5.11: The estimated residency times for all videos. The residency times are calculated from the data sharpening estimated ignition and extinction event times.
Figure 5.12: The estimated residency times for a 21 × 21 grid centered around the ignition point for all videos. The residency times are calculated from the data sharpening estimated ignition and extinction event times.
Chapter 6

Conclusions and future work

In this dissertation, we’ve developed a novel anisotropic smoothing method for change-point data, based on existing methods in nonparametric statistics and image processing. We’ve shown that our methods for anisotropic local constant and local linear kernel regression estimators are consistent. We’ve demonstrated their effectiveness on simulated one- and two-dimensional data similar to fire spread data, and validated their performance against other nonparametric estimators. We have designed and developed an apparatus that can perform fire spread experiments at the laboratory scale. We applied anisotropic estimators to measured data as an image processor, and to processed data to estimate ignition and extinction event times at the pixel level. Those estimated times were then used to estimate instantaneous and average fire spread rates, and residency time for burning cells.

The future work in anisotropic smoothing would be to prove consistency for the anisotropic local polynomial estimator. Another avenue for work is to assume the covariates to be random. Some progress has been made in this area, but difficulty has arisen in proof of consistency for the local linear estimate. Additionally, we plan to conduct simulation studies with more covariates to demonstrate the effectiveness of the anisotropic smoothing algorithm on higher dimensional regression relationships. We plan to investigate better bandwidth selection methods for estimating domain and range kernel bandwidths simultaneously. The optimal bandwidth selection procedures that we want to investigate further are adaptive cross-validation bandwidth selection methods that use
the derivative of the pilot estimator as criteria for an optimal bandwidth. Lastly, metrics that quantify the performance of a change-point estimator beyond the minimum MSE and the MESE is future work.

A new fire smoldering experimental apparatus is under development and currently in the design stage. The experiments are larger scale fire spreads contained in a ballistics hall wind tunnel where the flow structure and speed of the wind is easily controlled. We plan to implement higher resolution cameras that capture a broader spectrum of light wavelengths including infrared light, which experiences less smoke occlusion than the visual spectrum. This new design will allow us to study the interaction between slope and wind on the micro-fire experiments. Other work on current experiments includes studying the effects of slope on fire shape and fire spotting rates. We have a second data set of fire spreads where the wax paper was crumpled before being coated with potassium nitrate, as mentioned in Section 4.2.4. These crumpled papers cause more firebrand spotting, and we are interested in the effects of slope and wind on those rates. If we assume that the physical phenomenon that drives fire spread across these regions is a firebrand spotting process, we are interested in future models where we assume fire fronts may travel using a self-similar fractal process.

Image analysis in this research followed a two-stage procedure: the first stage was a spatial anisotropic smoothing across each image of the fire spread, and the second stage was a temporal anisotropic smoothing of each individual pixel. This two-stage smoothing structure suggests that time and space are independent and orthogonal. Attempts to fit this have been met with inferior performance relative to our two-stage procedure, however we believe there is a possibility we could improve our model by using a one-stage spatio-temporal smoothing model.

The fire spread rate regression model needs to be expanded to include directions other than directly up or downhill. This could be accomplished using a spatial statistics model (Cressie, 1992; Gelfand et al., 2010), where the angle $\theta$ from the ignition point relative to the East direction is included as a variable in the model. The ignition and extinction time estimates could also be used in a nonparametric regression model such as a functional data analysis to estimate average fire spread rates and study the effects of slope on fire
spread shapes. We are also interested in using more in-depth error checking methods to evaluate ignition and extinction times estimates. Furthermore, we are interested in applying bootstrapping techniques to better understand the variability of fire spread, similar to Garcia et al. (2008).

In Chapter 5, we estimated ignition and extinction times for individual pixels and calculated instantaneous spread rates over entire fires. We did not look at estimating ignition and extinction times for individual pixels in the simulated fire spread videos. The complexity of the a spatial covariance of pixel colour channels is not well understood, where we found this difficult to accurately specify in simulation. A simulation study for effectiveness of estimating fire spread rates from images with a proper spatial dependency for variability of pixel colours is future work.

We are currently developing software for analyzing forest fire spread code-named Diana. The main function of Diana will be to accept experimental and satellite imagery data, and perform a statistical fire spread analysis and simulation using the methodologies described in this dissertation. Finally, we envision three papers based on the work from each of the Chapters 3 through 5, with plans to include some of the future work in those papers.
Appendix A

Supplementary material for Chapter 3

A.1 Statistical theorems

This section contains probability theorems used in proofs, where the theorems are from the textbook by Grimmett and Stirzaker (2016).

Theorem A.1 (Continuous mapping theorem) Suppose $m$ is a continuous function that maps from one metric space $S$ to $S'$, that is $m : S \to S'$. This function $m$ has a set of discontinuity points $D$ such that $\mathbb{P}(X \in D) = 0$. Thus, if $X_n \xrightarrow{p} X$, then $m(X_n) \xrightarrow{p} m(X)$.

Theorem A.2 Let $a$ be constant. If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{a} a$, then $\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{a}$.

A.2 R-packages

Table A.1 is dedicated to all the R-packages (R Core Team, 2018) used in the makings of this dissertation, in order of amount of usage and importance.

A.3 Box plots of MESEs for anisotropic smoothing.
<table>
<thead>
<tr>
<th>Package</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>knitr</td>
<td>(Xie, 2017)</td>
<td>Dynamic report generation with embed R-code used to generate dissertation document</td>
</tr>
<tr>
<td>np</td>
<td>(Hayfield and Racine, 2008)</td>
<td>Kernel regression estimation and bandwidth selection methods; for theoretical support (Li and Racine, 2007, see).</td>
</tr>
<tr>
<td>parallel</td>
<td>(R Core Team, 2018)</td>
<td>Parallel computation using R on computers and SHARC-NET clusters with shared memory between cores.</td>
</tr>
<tr>
<td>quadprog</td>
<td>(Turlach and Weingessel, 2013)</td>
<td>Solving quadratic programming problems for data sharpening</td>
</tr>
<tr>
<td>ggplot2,</td>
<td>(Wickham, 2009),</td>
<td>Data visualizations of fire spreads and vector fields</td>
</tr>
<tr>
<td>plotly,</td>
<td>(Sievert et al., 2017),</td>
<td></td>
</tr>
<tr>
<td>raster,</td>
<td>(Hijmans, 2017),</td>
<td></td>
</tr>
<tr>
<td>rasterVis</td>
<td>(Perpiñán and Hijmans, 2018)</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: R-packages used in simulations, data visualizations, type-setting and dissertation creation.

Figure A.1: The MESEs for the three kernel estimators on the piecewise constant regression function with $\sigma = 0.1$. Note that $\cdot Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.2: The MESEs for the three kernel estimators on the piecewise constant regression function with $\sigma = 0.5$. Note that .Z means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.3: The MESEs for the three kernel estimators on the piecewise constant regression function with $\sigma = 1$. Note that $Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.4: The MESEs for the three kernel estimators on the piecewise constant regression function with $\sigma = 2$. Note that .Z means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.5: The MESEs for the three kernel estimators on data from the continuous regression function with a jump and $\sigma = 0.1$. Note that $.Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.6: The MESEs for the three kernel estimators on data from the continuous regression function with a jump and $\sigma = 0.5$. Note that $Z$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.7: The MESEs for the three kernel estimators on data from the continuous regression function with a jump and $\sigma = 1$. Note that $\mathcal{Z}$ means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
Figure A.8: The MESEs for the three kernel estimators on data from the continuous regression function with a jump and $\sigma = 2$. Note that .Z means $n = Z00$, e.g. LC.4 is the isotropic local constant estimator with $n = 400$ and ALC.16 is the anisotropic estimator with $n = 1600$. 
A.4 Iterative smoothing results

Supplemental tables for some of the simulations when doing iterative smoothing.

Figure A.9: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.10: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.11: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using the least-squares cross-validation method once for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.12: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.13: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.14: Iteratively fitting the ALC estimator with LC as a pilot on the piecewise constant regression function data. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.15: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.16: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.17: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using the least-squares cross-validation method once for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.18: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 

Figure A.19: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.20: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.21: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the ad hoc method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.22: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the AIC cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.23: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected once using the least-squares cross-validation method for ALC.1 and reused on all subsequent smooths. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.24: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using the ad hoc method for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.25: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using AIC cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Figure A.26: Iteratively fitting the ALC estimator with LC as a pilot on the data for the continuous regression function with a jump. The bandwidth is selected using least-squares cross-validation for each ALC smooth. The parameters for this simulation are $\sigma = 1$ and $n = 800$. 
Appendix B

Supplementary material for Chapter 5

B.1 Time ignition surfaces
Figure B.1: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.2: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.3: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.4: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 11° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.5: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of $32^\circ$ slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.6: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.7: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.8: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 32° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.9: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 1. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.10: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 2. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.11: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 3. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.12: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 4. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.13: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of $0^\circ$ slope experiment 5. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.14: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of $0^\circ$ slope experiment 6. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.15: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of $0^\circ$ slope experiment 7. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.16: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 8. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.17: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 9. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Figure B.18: The local constant smoothed ignition times in a level plot and vector field of the gradient for the smoothed ignition time surface of 0° slope experiment 10. The colour scale is time (s) in the first plot, and magnitude of the gradient (mm/s) in the second.
Bibliography


Curriculum Vitae

Name: John Ronald James Thompson

Education: Ph.D. (Statistics), University of Western Ontario, 2014 - 2018
M.Sc. (Statistics), McMaster University, 2012-2014
B.Sc. (Physics, Applied Physics Specialization, Co-operative Program), University of Waterloo, 2006 - 2011

Awards: Ontario Graduate Scholarship, 2017 - 2018
Research Scholarship (Dr. Jeffrey S. Racine), Summer 2013
Department of Mathematics & Statistics Masters Scholarship, 2012 - 2013

Work Experience:
Sessional Lecturer – Statistical Science 1024A (Introduction to Statistics), University of Western Ontario, Fall 2018
Instructor – Introduction to R Workshops, University of Western Ontario, 2017 - 2018
Reviewer, Journal of Statistical Education, 2017 - 2018
Statistical Consultant, Western Data Science Solutions, University of Western Ontario, 2016 - 2018
Teaching Assistant, Department of Statistical and Actuarial Sciences, University of Western Ontario, 2014 - 2018
Teaching Assistant, Department of Mathematics & Statistics, McMaster University 2012 - 2013

Peer-Reviewed Publications: