Analysis of Spatio-Temporal Data for Forest Fire Control

by

Yi Xiong

B.Sc. (With Distinction, Computing Science), Simon Fraser University, 2012

Project Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2015

APPROVAL

Name: Yi Xiong

Degree: Master of Science

Title: Analysis of Spatio-Temporal Data for Forest Fire Control

Examining Committee: **Chair**: Dr. Tim B. Swartz, Professor,
Statistics and Actuarial Science

Dr. X. Joan Hu, Professor
Statistics and Actuarial Science
Senior Supervisor

Dr. Derek Bingham, Professor
Statistics and Actuarial Science
Supervisor

Dr. W. John Braun, Professor
Statistics, The University of British Columbia - Okanagan
Internal Examiner

Date Approved: January 23rd, 2015

Partial Copyright Licence

Abstract

This project aims to establish the relationship of forest fire behavior with ecological /environmental factors, such as forest structure and weather. We analyze records of forest fires during the fire season (May to September) in 1992 from the Forest Fire Management Branch of Ontario Ministry of Natural Resource (OMNR). We start with a preliminary analysis of the data, which includes a descriptive summary and an ordinary linear regression analysis with fire duration as the response. The preliminary analysis indicates that the fire weather index (FWI) used by Natural Resource of Canada is the most relevant together with fire location and starting time. We apply semi-variogram and Moran's *I*, the conventional methods for exploring spatial patterns, and extend them to investigate spatio-temporal patterns with the fire data. Evaluations of the extended Morans *I* statistic with the residuals of the ordinary linear regression analysis reveal a large departure from the independence and constant variance assumption on the random errors. It motivates two sets of partially linear regression models to accommodate possible nonlinear spatial/temporal patterns of the forest fires. We integrate univariate and bivariate Kernel smoothing procedures with the least squares procedure for estimating the model parameters. Residual analysis indicates satisfactory fittings in both sets of regression analysis. The partially linear regression analyses find that the association of fire duration with FWI varies across different fire management zones, and depends on the fire starting time.

Keywords: Kernel Smoothing; Local Constant/Linear Regression; Morans *I*; Partially Linear Models

To my parents.

“Love all, trust a few, do wrong to none”

— William Shakespeare,
ALL’S WELL THAT ENDS WELL, *1623*

Acknowledgments

I would like to express my deep gratitude to my supervisor, Professor Joan Hu. She has been very supportive since the days I joined the master program in statistics at Simon Fraser University. I feel lucky to have a supervisor who cared so much about my work throughout the two years of my masters study. Her excellent guidance and scholarly inputs lead me to complete my thesis project step by step. What I learn from her is not just how to meet the program requirements, but also how to work as a good statistician.

I am grateful to Dr. Derek Bingham, Dr. John Braun and Dr. Tim Swartz for serving on my project committee. I am also thankful to Dr. Steve Cumming, Dr. Meg Krawchuk and Dr. Mike Wotton for their precious advice and comments on the fire data that my project is based on.

I would like to thank Professors Ian Bercovitz, David Campbell, Jinko Graham, Robin Insley, Richard Lockhart, Tim Swartz, Carl Schwarz, Boxing Tang in particular and all the professors in the department of Statistics and Actuarial Science for teaching me statistics and helping me to become a qualified statistician. I also thank Charlene Bradbury, Kelly Jay and Sadika Jungic for their kind help in the past two years.

Additionally, I want to thank my fellow graduate students Sherry Chen, Michael Grosskopf, Bobby Han, Kunasekaran Nirmalkanna, Nate Payne, Werjindra Premarathna, Pulindu Ratnasekera, Gerald Smith, Biljana Stojkova, Elena Szefer, Fei Wang, Huijing Wang, Vicky Weng, Annie Yu, Sabrina Zhang and friends Annie He, Yanfang Le, Cindy Sun for their help, encouragement, and friendship.

Finally, I record my special thanks to my parents, Junshun Xiong and Lixia Jiang, whose love and support have been always with me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Forest fire is a major cause of damage to both the forest ecosystem and human society. In Canada, an average of 8,000 forest fires occur and the disruption costs are ranged from 500 million to 1 billion annually (Wotton, M. 2012). While less than half of the fires are caused by lightning, the damage of those fires is greater than human-caused fires.

Forest fire management planning requires an understanding of the mechanism of fire occurrence process. The process of a lightning-caused fire occurrence can be broken down into 3 phases: ignition from a lightning strike, smouldering (either in the forest floor or in surface), and detection (Kourtz and Todd 1992; Anderson et al. 2002). The first stage of fire occurrence, i.e. the ignition, is crucial for the fire occurrence and the degree of ignition will influence the way that fire behaves afterwards. Forest structure and climate play the most important roles in the ignition of fire and these ecological factors often have strong spatial and temporal characteristics. Hence, establishing the relationships between the ecological factors and fire activities with embedded spatial and temporal characteristics is in demand for fire management planning.

To understand the spatio-temporal patterns of fire activities, many researchers investigated the spatio-temporal patterns of fire occurrences and fire behavior (e.g. Gralewicz et al. 2012; Krawchuk et al. 2006; Cumming 2001) from the historical data. Regression analyses are carried out to study the relationship between a fire and associated ecological

factors together with spatial and temporal characteristics of the fire. For example, Podur (2001) presents spatial and temporal patterns of annual area burned by forest fire in Ontario, Canada from 1917 to 2001, and Martell and Sun (2008) use spatial autoregressive models to evaluate the association between environmental factors and fire's burned area.

1.2 Project Objective and Outline

The previous studies suggest the importance of fires' spatio-temporal characteristics in evaluating the relationship between ecological risk factors and fire activities. This leads us to consider a regression model for fire activity accounting for the spatio-temporal correlation. Let $Y_i = Y(\mathbf{s}_i, t_i)$ be a measurement (e.g. duration of the fire) of fire i, occurring in location $\mathbf{s}_i$ and starting at time t_i. A general regression model is :

$$Y_i = \mu(\mathbf{s}_i, t_i; \mathbf{z}_i) + \epsilon_i;\ i = 1, 2, \cdots, n, \tag{1.1}$$

where $\mathbf{z}_i$ are the ecological factors of interest, $\mu(\mathbf{s}_i, t_i; \mathbf{z}_i) = E[Y_i|\mathbf{z}_i]$ and $\epsilon_i = \epsilon(\mathbf{s}_i, t_i)$ is the random error with $E[\epsilon_i] = 0$. The spatial and temporal patterns of the response can be modeled by specifications of the mean function, together with the distribution of random errors.

We aim to study the relationship between fire activities and the ecological factors (e.g. moisture content in the forest, the weather condition) in order to provide insights to forest fire control. The forest fire data provided by Forest Fire Management Branch of Ontario Ministry of Natural Resource (OMNR) and described in Chapter 2 are used to motivate and illustrate our approaches. Regression analyses with various specifications of $\mu(\mathbf{s}_i, t_i; \mathbf{z}_i)$ in (1.1) are conducted with the fire data. We begin with the ordinary linear regression analysis and use it to motivate two sets of partially linear models to capture better spatial and temporal correlation in fire activities.

Another statistical objective is to check for the spatio-temporal correlation to validate the regression model assumptions. We review the conventional approaches first and adapt them to analyze the forest fire dataset in Chapter 3.

We organize the rest of the project as follows. In Chapter 2, we present descriptive data analysis and an ordinary linear regression analysis with the forest fire data. Chapter 3 reviews some methodologies for detecting spatio-temporal correlation and adapts the methods to the forest fire data. We present procedures for estimating the parameters in the partially linear models and conduct residual analyses in Chapter 4. A summary of the study and a list of future work are provided in Chapter 5.

Chapter 2

Preliminary Analysis of Forest Fire Data

This chapter first presents a descriptive summary of the forest fire data set. Then we analyze the forest fire data with the ordinary linear regression model as a preliminary analysis.

2.1 Descriptive Data Analysis

The Forest Fire Management Branch of Ontario Ministry of Natural Resource (OMNR) have records of 259 lightning-caused fires during the fire season of 1992 (May to September). The dataset includes information of fire's size, durations, location, related fuel information and weather condition. The variables of the dataset are listed in Table 2.1. The following presents numerical and graphical summaries of the data.

2.1.1 Response Variables: Measurements of Fire

Fire size and fire duration are commonly-used measures of fire impact. Table 2.2 summarizes the means, medians and standard deviations of the fire durations and sizes. The standard deviations for both fire durations and sizes are quite high. The big differences between means and medians indicate severe skewness of the original data, and the large standard deviations show great variations in the data. The histograms of fire sizes and fire durations are plotted in Figure 2.1a and 2.1b. We can see that the distributions of

Table 2.1: Description of the Fire Dataset

Variable Name	Abbreviation	Description	Units of Measurements	Value Range
Fire Measurement Variables				
Size	N/A	The area of each fire has burned	hectare	0.1 to 44200
Duration	N/A	Difference between fire's start date and end date	days	1 to 58
Location and Time Variables				
Start Date	N/A	Estimated start date of the fire	N/A	1992-05-16 to 1992-09-15
Out Date	N/A	Date the fire was declared out	N/A	1992-05-16 to 1992-09-17
Month	N/A	Month of fire's start date	N/A	May to September in 1992
Latitude	lat	Latitude of fire's start location	degrees	N44.47 to N53.97
Longitude	long	Longitude of fire's start location	degrees	W95.10 to W76.26
Fmz	Fmz	Fire Management Zones	Extensive(E), Measured(M), Intensive(I)	N/A
Region	N/A	Fire regions	NWR, NER, SCR	N/A
District	Cur_dist	Administrative Districts	APK, BAN, CHA, COC, DRY, FOR, HEA, KEM, KEN, KLK, NIP, NOR, PAR, PEM, PET, RED, SAU, SLK, SUD,THU, TIM, WAW	N/A
Weather and Fuel Variables				
FFMC	FFMC	Fine Fuel Moisture Code	numeric ratings	14 to 93
DMC	DMC	Duff Moisture Code	numeric ratings	0 to 74
DC	DC	Drought Code	numeric ratings	3 to 390
ISI	ISI	Initial Spread Index	numeric ratings	0 to 22
BUI	BUI	Buildup Index	numeric ratings	1 to 82
FWI	FWI	Fire Weather Index	numeric ratings	0 to 33

both duration and size are highly skewed. On the other hand, the distribution of the log-transformed fire duration is closer to a normal distribution; See Figure 2.1d.

From Figure 2.2b, we can see that the log-transformed fire sizes and durations are closely related. It suggests to use either size or duration as the response variable in the regression analysis. We choose fire duration as the response variable in the following.

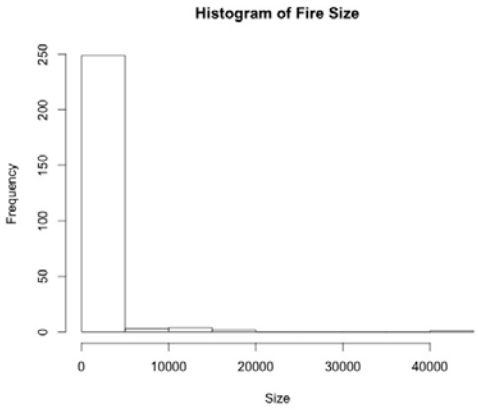
	Duration	Size
Mean	8.4	658.6
Median	4.0	0.3
SD	11.1	3591.1

Table 2.2: Summary of Fire Duration and Size

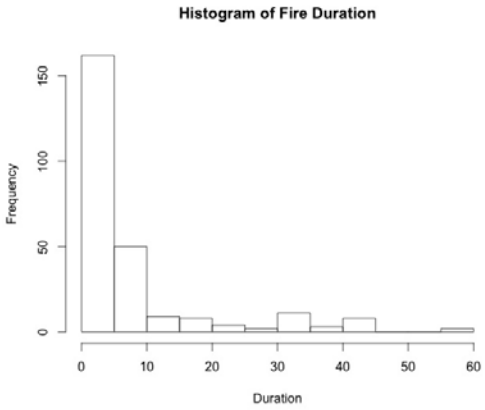
2.1.2 Explanatory Variables: Location and Time

Fire Location

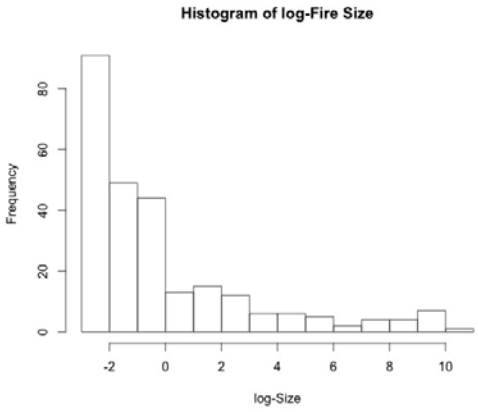
The location information of each fire included in the dataset is: *Longitude, Latitude, Region, District* and *Fire management zones (Fmz).* There are 3 regions: Northeast Region(NER), Northwest Region(NWR), South and Central Region(SCR). Each region contains a set of



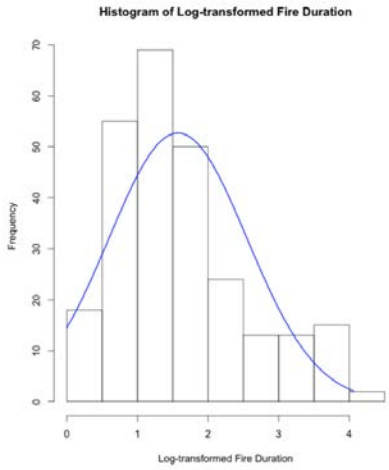
(a) Histogram of Fire Size



(b) Histogram of Fire Duration

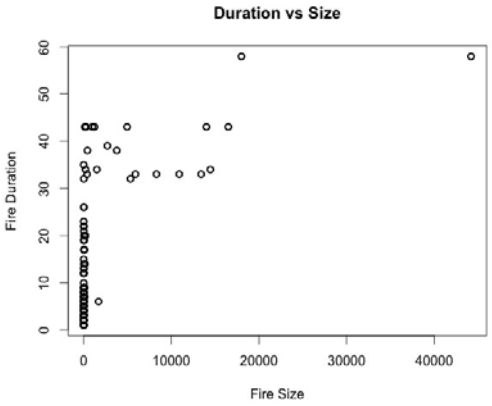


(c) Histogram of the Log-transformed of Fire Size

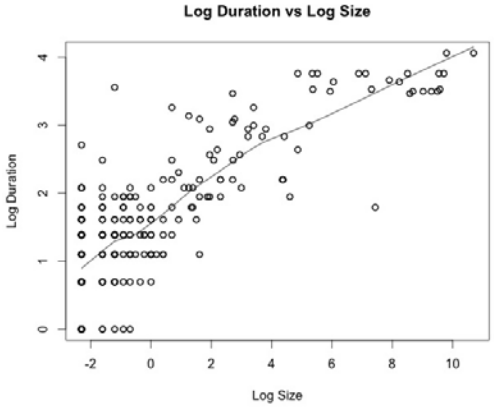


(d) Histogram of the Log-transformed of Fire Duration

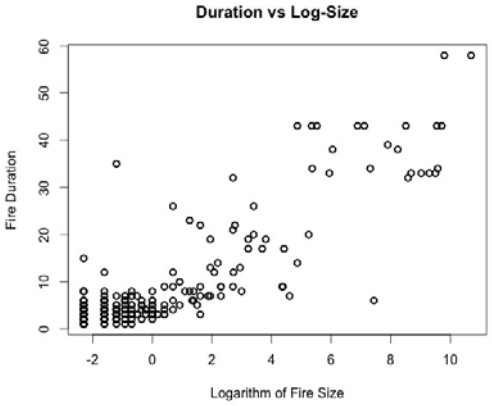
Figure 2.1: Histograms of Fire Duration and Fire Size



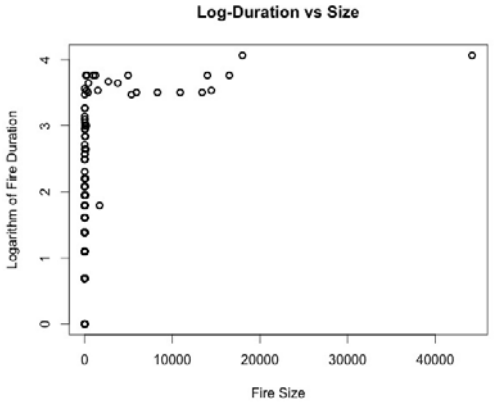
(a) Fire Duration vs Fire Size



(b) Log(Fire Duration) vs Log(Fire Size)



(c) Fire Duration vs Log(Fire Size)



(d) Log(Fire Duration) vs Fire Size

Figure 2.2: Scatterplots of Fire Duration and Fire Size

districts. Table 2.3 lists the districts. It appears that regions and districts are determined according to the geographical coordinates, i.e. longitude and latitude.

Table 2.3: Regions and Districts in Ontario

Zones	NER	NWR	SCR
Districts	NOR, CHA, COC, HEA, KLK, SAU, SUD, TIM, WAW	DRY, FOR, KEN, NIP, RED, SLK, THU	APK, BAN, KEM, PAR, PEM, PET

Ontario is divided into 3 fire management zones based on fire management strategy: extensive(E), measured(M) and intensive(I). Zone I is the area close to cities with large populations. All fires in this zone are fought aggressively. Zone M is the area where a fire has less potential to damage public safety and fires in zone M are only provided with full resources to fight at the initial attack (Martell and Sun, 2008). Zone E is close to remote areas and fires in this zone have less chance to threaten the public safety. So the fires in zone E are not fought aggressively. Figure 2.3a displays the variation of fire duration in each fire management zone and the summary statistics are listed in Table 2.4. From the plot and the summary statistics, we see that durations of fires in zone E are longer than other two zones, which is in agreement with the fire management strategy.

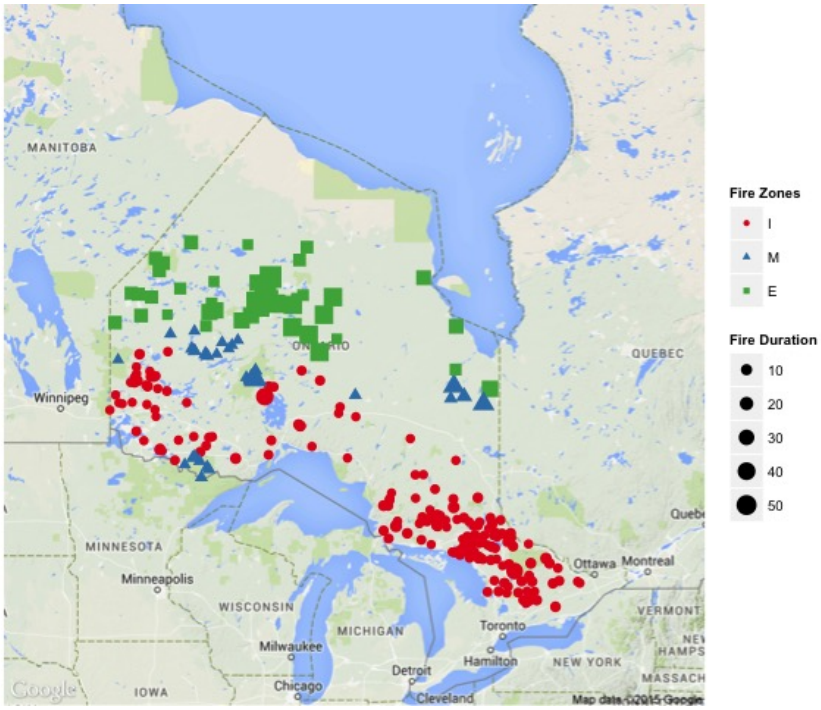
The variations of fire duration in different regions and zones suggest that, in addition to considering the geographical coordinates (longitude and latitude), *Region* and *Fmz* might be possible covariates to examine the association between fire location and fire duration.

Zone	Number of Fires	Mean of Fire Duration	SD of Fire Duration
Extensive(E)	45	24.89	14.57
Measured(M)	31	8.90	11.19
Intensive(I)	183	4.27	4.19

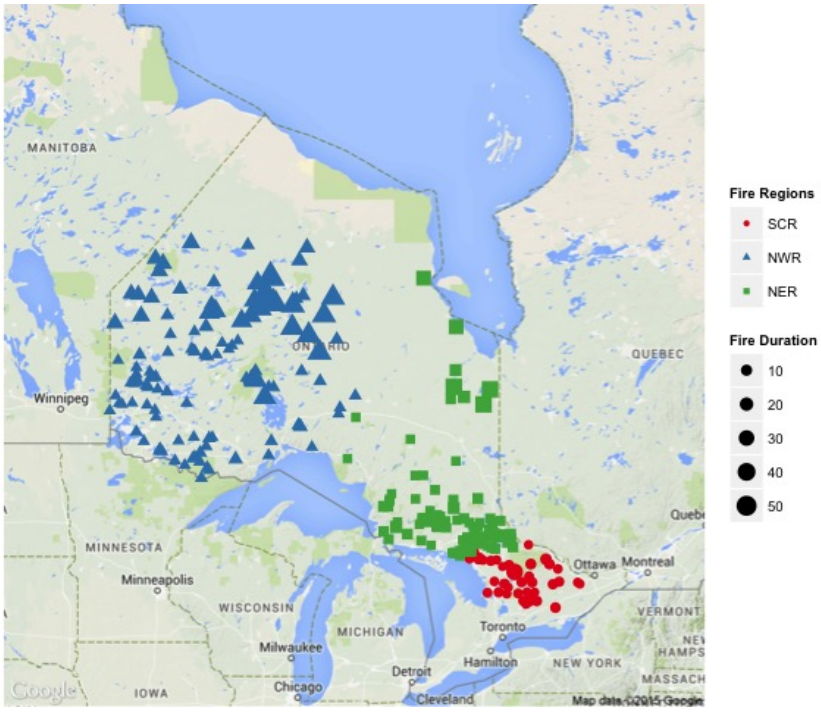
Table 2.4: Summary Statistics of Fire Duration in Each Fire Zone

Region	Number of Fires	Mean of Fire Duration	SD of Fire Duration
NWR	122	11.76	1.41
NER	97	5.40	0.78
SCR	40	5.48	0.65

Table 2.5: Summary Statistics of Fire Duration in Each Region



(a) Fire Durations across Zones



(b) Fire Durations across Regions

Figure 2.3: Fire Durations of Different Zones and Regions

Fire's Starting Time

All fires in the dataset started between May to September in 1992. Figure 2.4 plots fire duration against fire's start date, and Table 2.6 shows summary statistics of the fire duration. It appears that more than half of the fires started in June and fire duration in June was higher than those in May and July. Plus the fire duration in June had the highest variation.

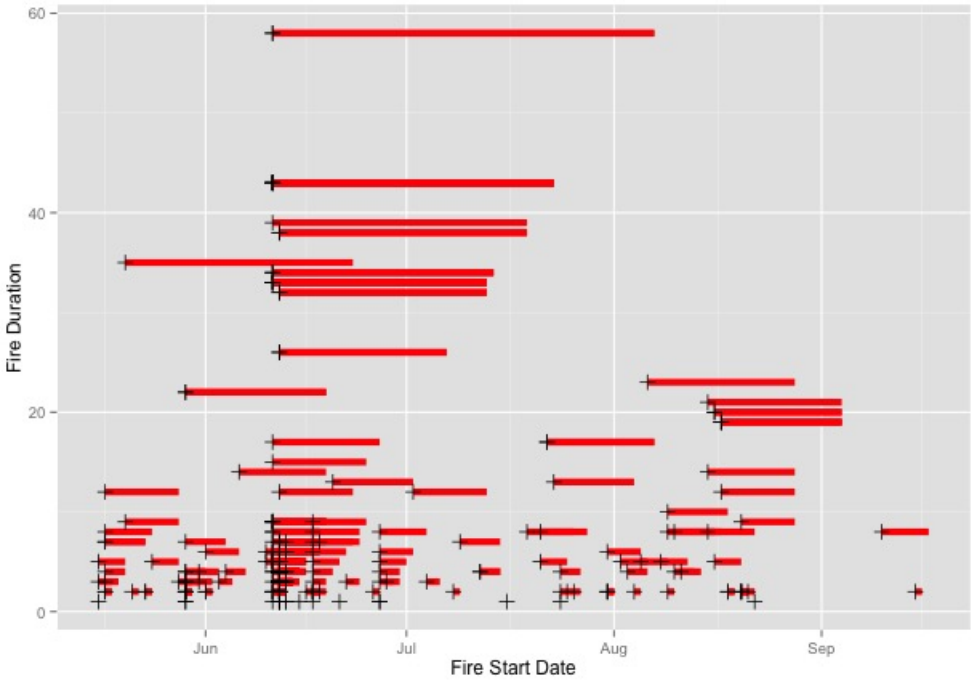


Figure 2.4: Fire Duration vs. Fire's Start Date

Month	No.of Fires	Mean of Fire Duration	SD of Fire Duration
May	35	5.68	7.11
June	169	9.37	12.74
July	22	5.64	4.96
August	31	8.42	7.19
September	2	5.00	4.24

Table 2.6: Summary Statistics of Fire Duration in Each Month

Plots in Figure 2.5 show the association of fire duration with fire's start date in different fire zones and regions. Most of fires with long durations were from region NWR and extensive zone.

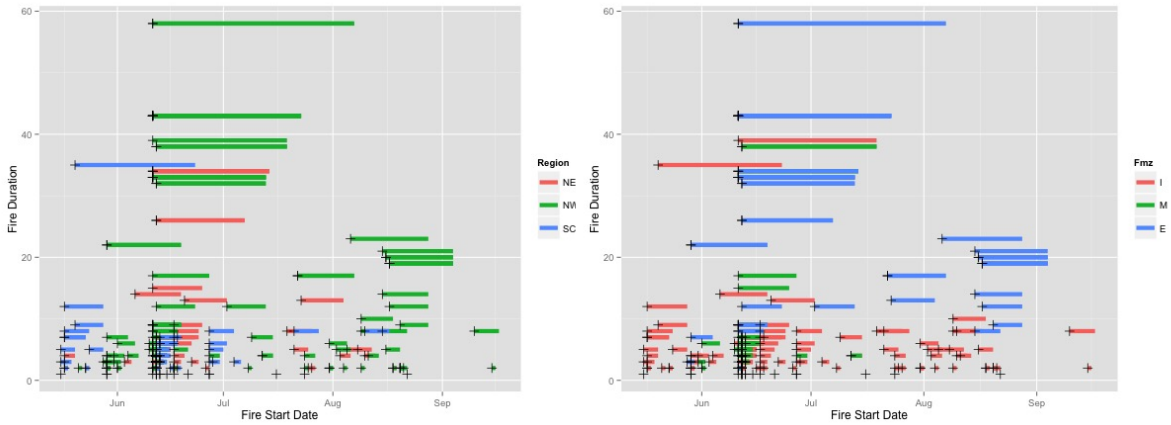


Figure 2.5: Fire Duration vs. Fire's Start Date in terms of Fire Management Zones and Regions

2.1.3 Explanatory Variables: Weather and Fuel Information

Description

Weather and fuel moisture of the forest play important roles in fire duration. For example, fires in a dry area might burn longer than a moist area. Canadian forest fire management agencies use the Canadian Forest Fire Weather Index System to measure the weather and fuel moisture data to predict the behavior of fire. Figure 2.6 from Natural Resource of Canada shows how the 6 standard components are calculated based on daily temperature, relatively humidity and wind speed.

There are 3 fuel moisture codes: FFMC, DMC and DC. Each of them represents the moisture content in different forest floors. The Fine Fuel Moisture Code (FFMC) mainly measures the moisture content of the fine fuels in the surface, which indicates the ease of fire ignition. The Duff Moisture Code (DMC) is a numeric rating of the average moisture content of loosely compacted organic layers with moderate depth. The Drought Code(DC) is a numeric rating of average moisture content in the deep compact organic layer, which indicates the ease of fire smoldering. These terms are explained in Merrill et al. (1987). Figure 2.7 shows the structures of forest floors and its corresponding fuel moisture codes. Figure 2.8 shows the relationship between moisture content in the forest floors and fuel moisture codes. The decreasing trends in plots suggest negative associations between

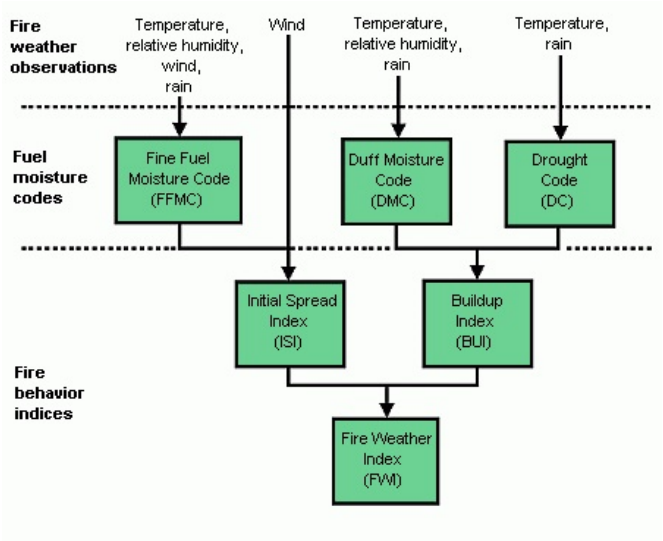


Figure 2.6: The FWI System from Natural Resource of Canada (2008)

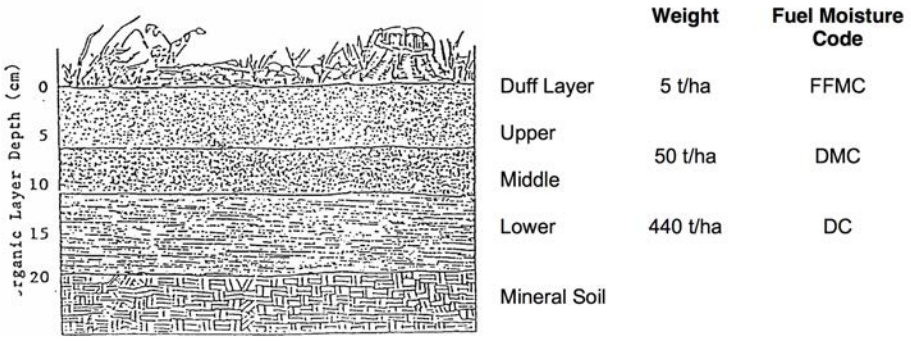


Figure 2.7: Interpreting the Canadian Forest Fire Weather Index(FWI) System from De Groot et al. (1998)

moisture content and fuel moisture codes, i.e. if the forest floor is dryer, the fuel moisture codes will be higher.

The remaining fire behavior indices *ISI, BUI, FWI* summarize daily weather and fuel information to represent the ratings of fire danger. The Initial Spread Index(ISI) is generated based on wind speed and FFMC and indicates the rate of fire spread. Buildup Index(BUI) is generated based on DMC and DC, which represents the amount of fuel available for fire burning. Fire Weather Index(FWI) is derived from ISI and BUI and works as a numeric rating of fire intensity. The formulas listed below are given in Van Wagner et al. (1987).

$$\textbf{ISI} = 0.208 \times (e^{0.0504W})(91.9e^{-0.138\text{FFMC}})(1 + \frac{\text{FFMC}^{5.31}}{4.93} \times 10^7)$$

$$\textbf{BUI} = \frac{0.8(\text{DMC})(\text{DC})}{\text{DMC} + 0.4\text{DC}}$$

$$\textbf{FWI} = 0.1(\text{ISI})[0.626(\text{BUI}^{0.809}) + 2]$$

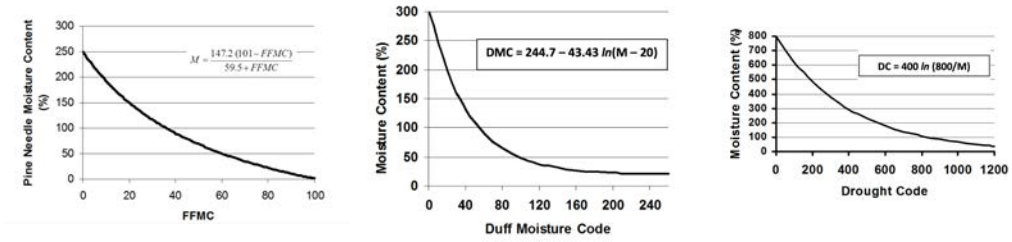


Figure 2.8: Moisture Content and Fuel Moisture Code from Van Wagner et al. (1987)

Summary Statistics

In this dataset, 6 codes on each fire's start date are provided. Table 2.7 lists the summary statistics of 6 codes. To make these codes more comparable, we standardized these codes to make them range over $[0, 1]$. Table 2.8 displays the summary statistics of standardized codes. We will use the standardized codes as covariates in future regression analysis.

FWI has been used commonly to summarize the weather and fuel information. Our initial regression analysis confirmed that *FWI* is a good environmental factor. Thus in the following, we only consider it as the environmental factor.

	FFMC	DMC	DC	ISI	BUI	FWI
Mean	84.45	26.93	148.29	7.07	35.34	13.27
Median	89.00	26.00	151.00	6.60	34.00	13.00
SD	12.33	13.10	67.10	4.88	15.49	7.80

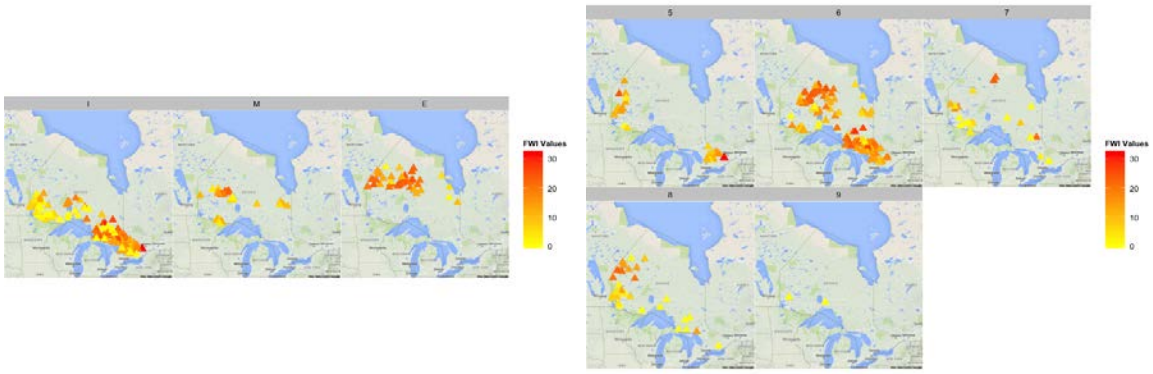
Table 2.7: Statistical Summary of Weather and Fuel Information

	FFMC	DMC	DC	ISI	BUI	FWI
Mean	0.89	0.36	0.38	0.32	0.42	0.40
Median	0.95	0.35	0.38	0.30	0.41	0.39
SD	0.16	0.18	0.17	0.22	0.19	0.24

Table 2.8: Statistical Summary of Standardized Weather and Fuel Information

The Association of *FWI* with Time and Location

To include *FWI* as a covariate in a regression model together with fire's starting time and location, we need to examine their correlation. Figure 2.9 shows the distributions of *FWI* values in different months and different fire management zones. The yellow spot indicates a low value of *FWI* and a red spot indicates a high *FWI*. There seems no obvious pattern in *FWI*'s distributions, but *FWI* values in zone I and September are lower comparing with other zones and other months.



(a) *FWI* in Each Fire Zone (b) *FWI* in Each Month

Figure 2.9: Association of *FWI* with Fire Starting Time and Fire Location

2.2 Ordinary Linear Regression Analysis

The descriptive data analysis provides an information summary of the dataset and suggests variables potentially related to fire duration. We conduct an ordinary linear regression analysis to explore how fire duration is associated with spatial and temporal variables as well as the environmental factor *FWI*. For the *fire duration* mentioned in the rest of the project, it stands for the logarithm of fire duration in the dataset.

2.2.1 Notation and Model Specification

We assume that $\mu(\mathbf{s}_i, t_i; \mathbf{z}_i)$ in Model (1.1) is a linear combination of functions of all the spatial, temporal and environmental variables. The model is specified as follows:

$$Y_i = \alpha_1 * g(\mathbf{s}_i) + \alpha_2 * h(t_i) + \boldsymbol{\beta}^{'}\mathbf{z}_i + \epsilon_i,\ i = 1, 2, \cdots, 259, \tag{2.1}$$

where $g(\cdot)$ and $h(\cdot)$ are pre-determined functions. In this model, we assume that residuals $\epsilon_i = \epsilon(\mathbf{s}_i, t_i)$ are identically and independently from a normal distribution with mean 0 and a constant variance σ^2.

We use $g(\mathbf{s})$ to represent a spatial variable considered as a predictor. It can be (a) the geographical coordinates: **Longitude, Latitude**, that is $g(\mathbf{s}) = \mathbf{s}$; (b) **the fire management zone code**; or (c) **the region code**.

To assess the relationship between fire duration and fire's starting time t, we can use $h(t)$ as the fire's starting date. We define the start date of each fire as the days between fire's start date and 1992-05-16. Variable z is the environmental factor. We focus on only *FWI* $= z$ in the analysis, which uses the standardized *FWI* ranging between 0 to 1. Table 2.9 lists the models with different $g(\mathbf{s})$ specifications.

Model	Explanatory Variables
Model 2.1(a)	Longitude, Latitude, Start date, FWI
Model 2.1(b)	Fire management zone, Start date, FWI
Model 2.1(c)	Region, Start date, FWI

Table 2.9: Ordinary Linear Regression Models

2.2.2 Analysis Results

Table 2.10 summarizes parameter estimates from Model 2.1 (a), which includes *Longitude* and *Latitude* in the model. Table 2.11 summarizes the estimates from Model (b) and (c), which incorporate the spatial effect as category variables. The standard errors and p_values from the Wald-test on whether the parameter is zero are given below the estimates. R^2 and mean squared error (MSE) are also given in the tables. The significance level is 0.05. The significant predictors are marked as bold.

The spatial variables $g(\mathbf{s})$ are found to be statistically significant among all three models. Model 2.1(a) suggests that *Longitude* and *Latitude* are statistically significant predictors for fire duration. Positive coefficients of *Longitude* and *Latitude* indicate that fire occurring in northeastern part of Ontario have a longer duration. Model 2.1 (b) takes zone I, as the reference level for variable *Fire management zone (Fmz)*. From the results in Table 2.11, duration of fires occurring in zone M and E, which are farther north parts, is significantly longer than the ones in zone I. The analysis results of Model 2.1(c) indicate that duration of fires occurring NWR regions are significantly longer than the ones in other regions.

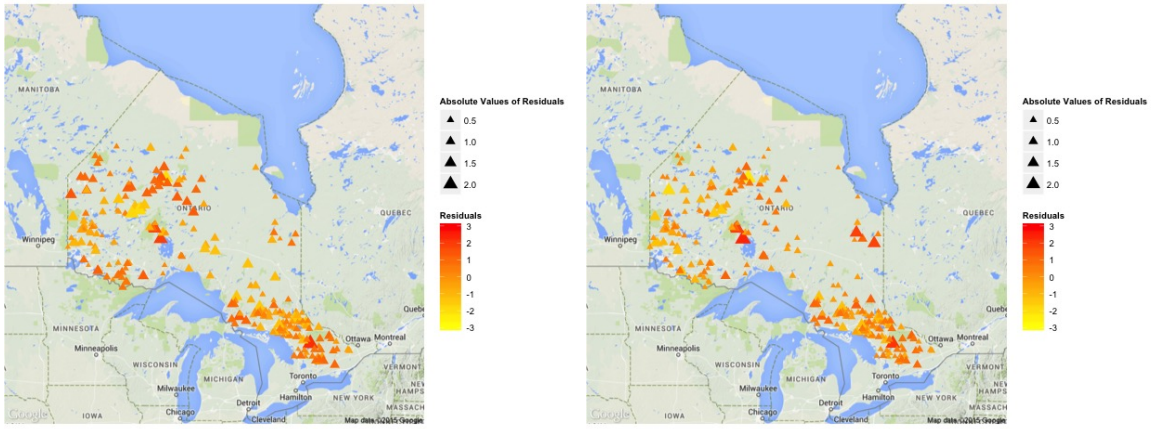
The time variable, *Start date*, is significant with Model 2.1(b). The fuel and weather index, *FWI*, is significant with Model 2.1(c). Comparing the R^2 and MSE of three models, Model 2.1(b) has the highest R^2 and smallest MSE. The small R^2 value with Model 2.1(c) indicates that there is a lot of variation left unexplained, making it the least appropriate model.

Table 2.10: Regression Coefficients Estimates in Model 2.1(a).

	Model a:				
	intercept	α_{long}	α_{lat}	α_{date}	β_{FWI}
Estimates	-6.23	**0.124**	**0.381**	-0.001	-0.270
Std.Error	0.927	0.015	0.032	0.002	0.229
P$-$value	$<$0.001	$<$0.001	$<$0.001	0.560	0.239
R^2			0.367		
MSE			0.784		
AIC			615.9		

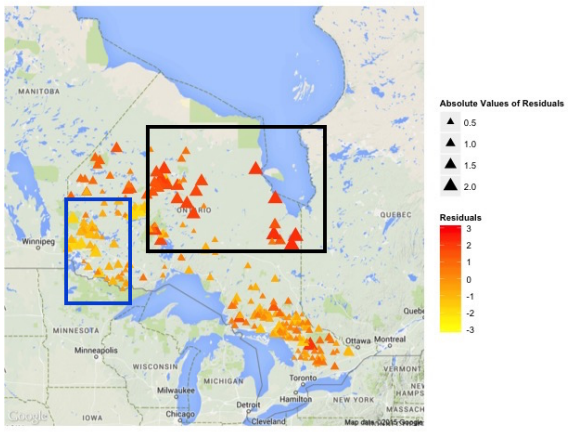
Table 2.11: Regression Coefficients Estimates in Model 2.1(b) and (c).

	Model b:					Model c:				
	intercept	**Fmz:E**	**Fmz:M**	α_{date}	β_{FWI}	intercept	**Region:NWR**	Region:SCR	α_{date}	β_{FWI}
Estimates	1.51	**1.86**	**0.360**	**-0.004**	-0.379	1.05	**0.562**	0.192	-0.0001	**0.584**
Std.Error	0.137	0.125	0.141	0.002	0.210	0.221	0.132	0.182	0.003	0.268
P−value	<0.001	<0.001	<0.001	0.037	0.072	<0.001	<0.001	0.292	0.966	0.030
R^2			0.473					0.076		
MSE			0.716					0.949		
AIC			569.5					714.7		



(a) Residual Map for Model 2.1(a)

(b) Residual Map for Model 2.1(b)



(c) Residual Map for Model 2.1(c)

Figure 2.10: Residual Maps with Ordinary Linear Regression

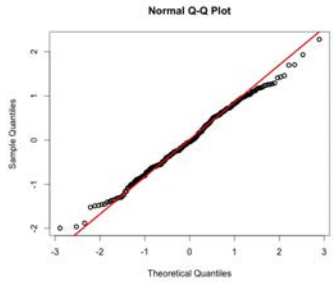
2.2.3 Residual Analysis

We produce a map of residuals in Figure 2.10 to further study the goodness of fit for the three models. We use the size of each point to represent the absolute values of residuals and the gradient of a point's color from yellow to red to illustrate the values of residuals. If the color of point is close to yellow, then the residual is negative; If it is close to red, then the residual is positive. The residual map of Model 2.1(c) shows a pattern of clustering: positive residuals in the black rectangle area and negative residuals in the blue rectangle area. The patterns in the residual map suggest that Model 2.1(c) might not be an appropriate model.

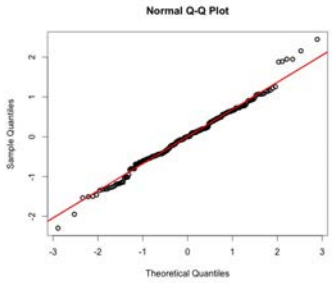
The interpretation of ordinary linear regression model is based on the assumption that the error terms are independently and identically from the standard Normal distribution $N(0, 1)$. Because of the obvious pattern in the residual map and low R^2 value, we focus on the residual analysis for Model 2.1(a) and (b) in the following.

Checking Model Assumptions

The normal distribution assumption for the models is checked by plotting Normal Q-Q plots in Figure 2.11a and Figure 2.11b. The plots indicate that residuals from Model 2.1(a) and (b) roughly follow normal distributions. The scatterplots of residuals versus predicted values, residuals versus FWI and start dates for both models are plotted in Figures 2.12 and 2.13. The red lines are the locally weighted scatterplot smoothing curves drawn by **lowess** function in R. These plots show that there may be more appropriate models, but the fits of the two models could be acceptable.



(a) Normal QQ Plot for Model 2.1(a)



(b) Normal QQ Plot for Model 2.1(b)

Figure 2.11: Normal QQ-Plots for Residuals

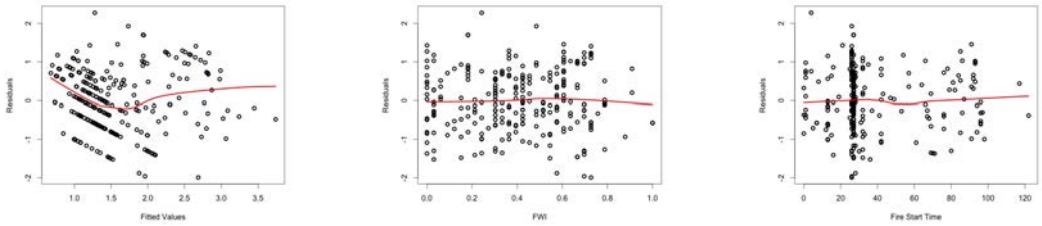


Figure 2.12: Scatterplots of Residuals in Model 2.1(a)

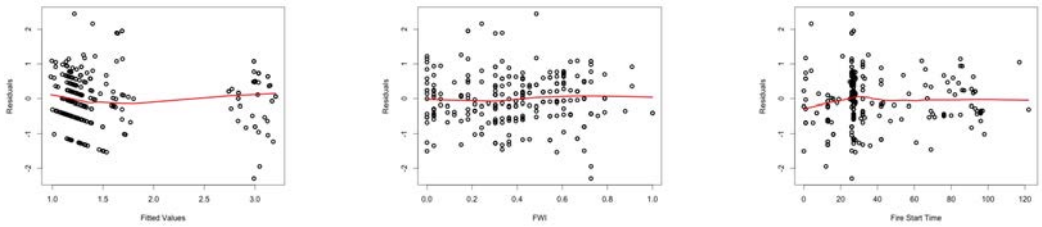


Figure 2.13: Scatterplots of Residuals in Model 2.1(b)

Another key assumption of the linear model is that all the observations are independent. In other words, all the durations of fires, regardless of the starting time and location, are independent. In fact, fires in near locations are likely related to each other. It is often desirable to find spatio-temporal correlations in fire activities.

Chapter 3

Detecting Spatio-Temporal Correlation

In this chapter, we first review commonly-used methods for measuring and detecting spatial autocorrelation. Then we extend the methods to examine the spatio-temporal correlation in the forest fire data. The methods are applied for detecting spatio-temporal correlation with the residuals of linear regression analysis in Chapter 2.

3.1 Detecting Spatial Autocorrelation

Spatial data analysis is challenging due to the presence of spatial autocorrelation. Failing to address the spatial autocorrelation in the data could lead to biased or inefficient inference. In this section, we briefly review two well-established spatial autocorrelation measurements: Semi-variogram and Moran's *I* Statistics.

3.1.1 Semi-variogram

In geostatistical data analysis, *variogram* is used to summarize the covariance structure of a spatial stochastic process. The *variogram* of a stochastic process $\{X(\mathbf{s}_i) : i = 1, 2, \cdots, n\}$ is defined as a function $V(\mathbf{s}_i, \mathbf{s}_j) = Var\{X(\mathbf{s}_i) - X(\mathbf{s}_j)\}$, which is the variance of the difference in $X(\cdot)$ at two locations ($\mathbf{s}_i$ and $\mathbf{s}_j$). The *semivariance* $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ is defined as half of variogram: $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}V(\mathbf{s}_i, \mathbf{s}_j)$. The *semi-variogram* of a set of spatial

data can be plotted as the scatterplot of the points $(d(\mathbf{s}_i, \mathbf{s}_j), \gamma(\mathbf{s}_i, \mathbf{s}_j))$, where $d(\mathbf{s}_i, \mathbf{s}_j)$ is the distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$.

For example, with the regression model (1.1) in Chapter 1 for the forest fire dataset, we may view the random errors as a spatial process $\boldsymbol{\epsilon}(\mathbf{s}) = \{\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2), \cdots, \epsilon(\mathbf{s}_{259})\}$ with $E[\epsilon(\mathbf{s})] \equiv 0$ and $V(\epsilon(\mathbf{s})) = \sigma^2$. The semivariance of the residual spatial process is then

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2} Var[\epsilon(\mathbf{s_i}) - \epsilon(\mathbf{s}_j)] = \frac{1}{2} E[\{\epsilon(\mathbf{s}_i) - \epsilon(\mathbf{s}_j)\}^2];\ i, j = 1, 2, \cdots, 259.$$

We call $\boldsymbol{\epsilon}(\mathbf{s})$ an isotropic process if its covariance depends only on the distance d between locations $\mathbf{s}_i$ and $\mathbf{s}_i$. For such a process, the semivariance can be written as a function of the distance d:

$$\begin{aligned} \gamma(d) &= \frac{1}{2} Var(\epsilon(\mathbf{s}_i)) + \frac{1}{2} Var(\epsilon(\mathbf{s}_j)) - Cov\{\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)\} \\ &= \sigma^2 - C(d), \end{aligned} \tag{3.1}$$

where $C(d)$ is the covariance function of any pair of observations with distance d.

The empirical semi-variogram is obtained by plotting the estimates of the semivariance versus the distance between any pair of locations. An estimator for semivariance is given by Cressie (1993):

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i,j:|\mathbf{s_i} - \mathbf{s_j}| = d} (X(\mathbf{s_i}) - X(\mathbf{s_j}))^2 \tag{3.2}$$

where $N(d)$ is the set of distinct pairs separated by distance d.

The empirical semi-variogram with a dataset can be used to detect possible spatial autocorrelation in the data. If observations close to each other have high correlation, the covariance $C(d)$ increases as the distance between observations becomes smaller, and the semivariance $\gamma(d)$ decreases. On the other hand, the semivariance $\gamma(d)$ increases as the distance increases. When the distance goes up to a certain value (leading to little spatial autocorrelation between observations), the covariance $C(d)$ reduces to 0 and the semivariance reaches to the variance of the underlying spatial process. Therefore, the

increase trend in semi-variograms as distance increases indicates a spatial autocorrelation in the data. If there is no spatial autocorrelation in the data, the semivariance should be distributed randomly around the true variance of the data.

We simulated 259 observations independently from a standard normal distribution $N(0, 1)$ and assigned each of them to the fire location. The semi-variogram with this simulated data is displayed in Figure 3.1. The red smoothing curve of the semi-variance is added to the plot. The red curve remains constant and doesn't increase as the distance increases. Comparing with the blue dashed line, which is the true variance of these observations, the red curve is very close to the true variance of the data. It indicates that there is no spatial autocorrelation in the data, which is in agreement with the settings of simulation.

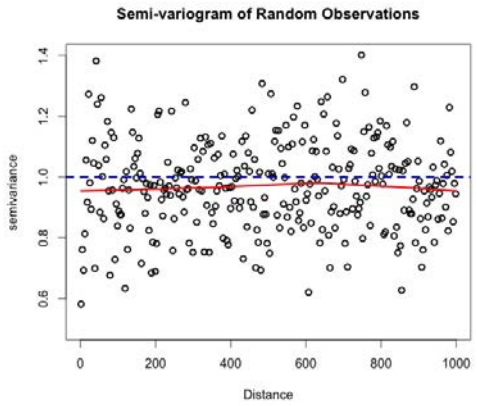


Figure 3.1: Semi-variograms of Simulated Independent Data

The semi-variogram is commonly employed in descriptive geostatistics (F Dormann, Carsten, et al. 2007) to describe the spatial autocorrelation of a spatial stochastic process. However, it requires the stationary assumption of the underlying process and it is hard to interpret with non-stationary process.

3.1.2 Moran's *I* Statistics

The semi-variograms provide a visualization of the correlation structure in a spatial process. However, we need some statistics to quantify the spatial autocorrelation. Moran's *I* introduced by Moran (1950) is a commonly-used statistic to measure spatial autocorrelation. We review below how to use Moran's *I* to check for spatial autocorrelation.

Global Moran's *I* Statistics

The global Moran's *I* is defined by Moran(1950) with a dataset $\{x_i, i = 1, 2, \cdots, n\}$ as:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W_0 \hat{\sigma}^2}, \tag{3.3}$$

where x_i is observation at location $\mathbf{s}_i$, and $\bar{x} = \frac{\sum_i^n x_i}{n}$, $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$, $W_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ with w_{ij} the weight chosen according to the locations of x_i and x_j. If the two observations are neighbors, $w_{ij} = 1$; otherwise, $w_{ij} = 0$. We define neighbors based on distance between the locations of two observations. Usually, for a pre-determined distance d, we call x_i and x_j are neighbors if distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$ is less than d.

Moran's *I* is a weighted average of the cross-product of observations, centered to the average value of the observations and standardized to adjust for the variance of the observations (Walberg, 1985).

Theoretic properties, including the expectation and variance of global Moran's *I*, are derived by Cliff and Ord (1981). They show the expectation and variance of Moran's *I* are as follows.

$$E[I] = \frac{-1}{n-1}\ , Var[I] = \frac{n^2 W_1 - n W_2 + 3 W_0^2}{(n-1)(n+1) W_0^2} - \frac{1}{(n-1)^2}. \tag{3.4}$$

where $W_0 = \sum_i \sum_j w_{ij}$, $W_1 = \frac{1}{2} \sum \sum_{i \neq j} (w_{ij} + w_{ji})^2$, $W_2 = \sum_{i=1}^n (\sum_{j=1}^n w_{ij} + \sum_{i=1}^n w_{ji})^2$. Cliff and Ord (1981) prove that the Moran's *I* is asymptotically normally distributed. Based on these results, global Moran's *I* can be used to test on spatial independence assumption.

We conducted Moran's *I*-based test with the forest fire dataset. For 259 records of fire in the dataset, we use d =400 km as a pre-determined distance to define neighbors based on fires' locations. Figure 3.2 plots a map of the fires of the dataset. The colors represent the durations of fires and the lines between points represent the neighbors of the fire. The similar colors of neighbor points suggests a similar value of fire durations, which indicates a possible spatial autocorrelation.

We obtained the Moran's *I* statistic with the data as *I*=0.149, and the p_value for testing on a spatial autocorrelation is lower than 0.01. It indicates a significant spatial autocorrelation.

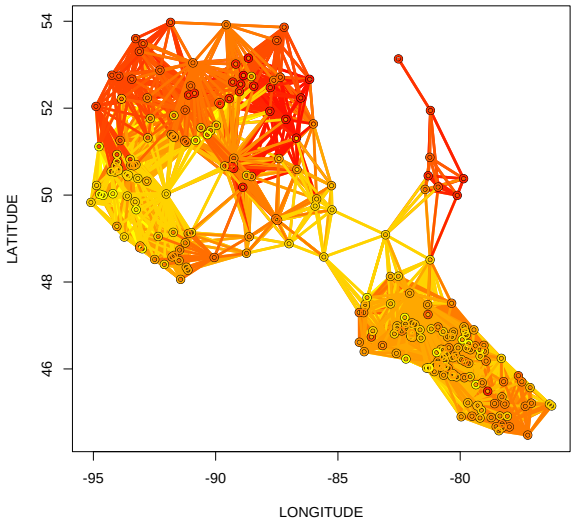


Figure 3.2: Neighbors Within 400 km of Each Fire

The permutation tests can be used to verify the above test outcomes. With the fire data, we permuted the fire duration and assigned them randomly to fire location at each replication. Moran's *I* can be computed at each replication. We repeated the procedure 999 times and obtained the sample distribution of Moran's *I*. The observed Moran's *I* above can be compared with the simulated Moran's *I*. Figure 3.3 shows the sample distribution of Moran's *I* and the observed Moran's *I* with the fire data. Table 3.1 summarizes the statistics of the simulated Moran's *I*. The sample mean of simulated Moran's *I* is -0.00393, which is close to the mean $\frac{-1}{259-1} = -0.00388$ computed by theoretical results in formula (3.4). The blue dotted line is the observed Moran's *I*. It is significantly deviated from the mean of simulated Moran's *I*, indicating a significant spatial autocorrelation.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.0171	-0.0081	-0.00531	-0.00393	-0.00143	0.031410

Table 3.1: Summary Statistics of Simulated Moran's *I*

The Moran's *I* depends on the definition of neighbors. Denote the Moran's *I* calculated using neighbors with distance within d by $I(d)$. We defined locations within 400 km as neighbors above. To see how Moran's *I* varies among the distance between neighbors, we plot a *spatial correlogram*, a scatterplot of Moran's *I* statistics $I(d)$ against the distances d.

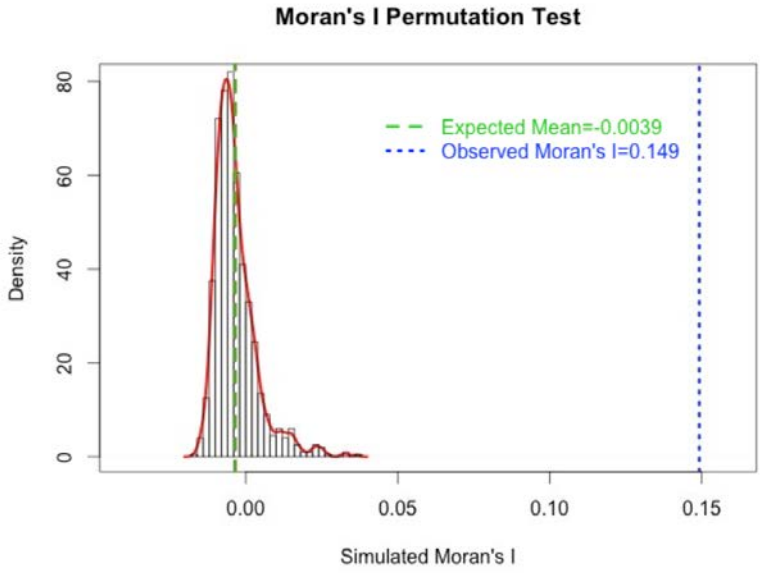


Figure 3.3: Moran's I Permutations

Figure 3.4 shows the spatial correlogram of Moran's *I* statistics with fire duration. The 95% acceptance region is also plotted for each Moran's *I*. The red dashed line shows the expectation value. The plot shows the Moran's *I* decreases as the distance between neighbors increases. According to the 95% acceptance regions and the expectation values, the spatial autocorrelation in fire duration is significant up to the distance by 1000 km.

Local Moran's *I* Statistics

The global Moran's *I* summarizes the spatial correlation with a single value. Local Moran's *I*, introduced by Anselin (1995), helps to investigate the local spatial clustering and identify spatially correlated hotspots.

Local Moran's *I* is calculated for each observation. Specifically, the local Moran's *I* associated with the i^{th} observation at location $\mathbf{s}_i$ is:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{j=1}^{n} w_{ij} \hat{\sigma}^2} \tag{3.5}$$

Comparing with formula (3.3), the global Moran's *I* is a weighted average of local Moran's

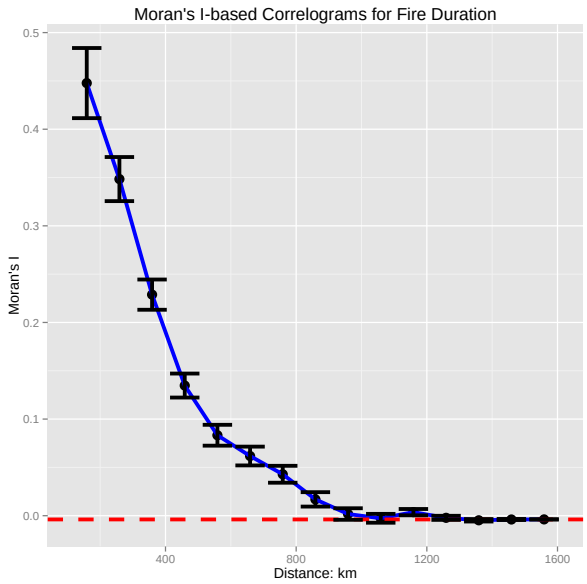


Figure 3.4: Moran's I-based Spatial Correlogram for Fire Duration

I's:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} I_i}{W_0}$$

The expectation and variance of I_i can be derived based on the results of global Moran's *I*. We can conduct test of local spatial autocorrelation for each point based on these results following the test procedure of global Moran's *I*.

3.2 Extended Methods for Spatio-Temporal Correlation

A fire occurs at a location and a time point. It is neither purely a spatial process nor a time series. The usual statistics to describe spatial autocorrelation need to be extended to incorporate the time dimension (Dub, J. & Legros, D., 2013). We will first review the semi-variogram computed for spatio-temporal process given by Cressie (1999) and then extend Moran's *I* methods to the spatio-temporal dimension.

3.2.1 Spatio-Temporal Semi-variogram

One method to check the spatio-temporal correlation in the data is to treat the data as a spatio-temporal process and then summarize its covariance structure. For a spatio-temporal process $\{X(\mathbf{s}_i, t_i), i = 1, 2, \cdots, n\}$, the estimator of spatio-temporal semivariance is given by Cressie (1999):

$$\hat{\gamma}(d;\tau) = \frac{1}{2N(d;\tau)} \sum_{\substack{i,j:|\mathbf{s_i}-\mathbf{s_j}|=d,\\ |t_i-t_j|=\tau}} (X(\mathbf{s_i}, t_i) - X(\mathbf{s_j}, t_j))^2 \tag{3.6}$$

where $N(d;\tau)$ is the number of pairs of observations separated by distance d and time differing at τ.

From this definition, semivariance is calculated for each combination (d, τ). However, for forest fire data, it is not guaranteed that there are fires with respect to each combination (d, τ). So this method might not be applicable for the fire data.

3.2.2 Extended Moran's *I*

We consider an extension of the global Moran's *I* by specifying spatio-temporal neighbors to account for time dimension.

Define the extended global Moran's *I* as follows:

$$I^* = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^*(x_i - \bar{x})(x_j - \bar{x})}{W_0^* \hat{\sigma}^2}. \tag{3.7}$$

$W_0^* = \sum_{i=1}^n \sum_{j=1}^n w_{ij}^*$, where $w_{ij}^* = 1$ if fire i and j's locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are within a pre-specified distance d and their starting times t_i and t_j differ at most by a pre-specified values τ. The extended global Moran's *I* have the same properties as the global Moran's I given in equation (3.3), except W_0, W_1, W_2 in (3.3) are determined by w_{ij}^*. Thus we can test on significance of spatio-temporal correlation based on I^*.

Similarly, the local Moran's *I* can be extended by defining the neighbors to account for the time dimension. Maps of the extended local Moran's *I* test results can be plotted to evaluate local spatio-temporal correlation.

Moran's *I*-based correlograms can also be plotted to describe how the extended global Moran's *I* varies among the distance and time lag between neighbors. For forest fire data, we define spatio-temporal neighbors based on fires' locations and starting dates. The extended global Moran's *I* can be calculated to detect spatio-temporal correlation for fire duration.

Figure 3.5 produces a perspective plot of the extended global Moran's *I* against time lags and distance. We choose distance between neighbors as 200 km, 400 km, 600 km, 800 km, 1000 km, 1200 km and 1400 km. The sequence of time lags is defined as 0 day, 5 days, 10 days,$\cdots$, 120 days. To give a more detailed presentation, we produce the correlograms which plot the extended Moran's *I* with different distances and time lags. The plots are shown in Figure 3.6. The 95% acceptance regions and expectation values of extended Moran's *I* are also presented in the plot. The red dashed lines represent the expectation values and the orange dot-dashed lines represent original Moran's *I* calculated from (3.3) at each distance.

The correlogram shows that the extended Moran's *I* decreases till it reaches the value of original Moran's *I*. Moreover, the significance of extended Moran's *I* varies according to both distance and time lag. Correlograms at distance 200 km, 400 km, 600 km and 800 km show significant correlation regardless of the starting time. At distance 1000 km and 1200 km, we only find that the significant correlation exists when fires occurring on the same day or within 5 days are considered as neighbors.

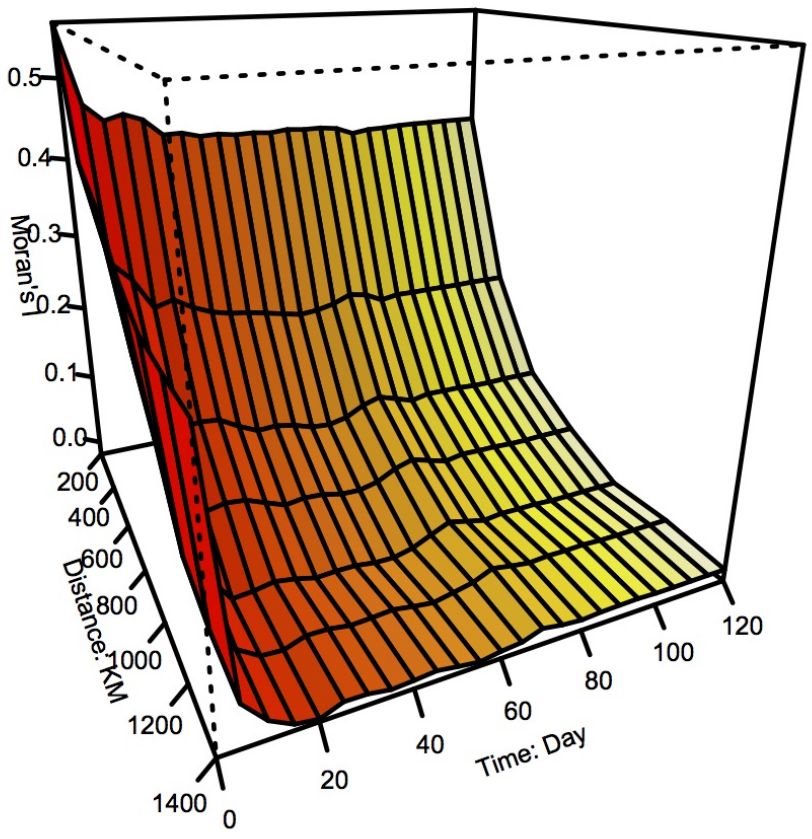


Figure 3.5: Perspective Plot of Extended Moran's I

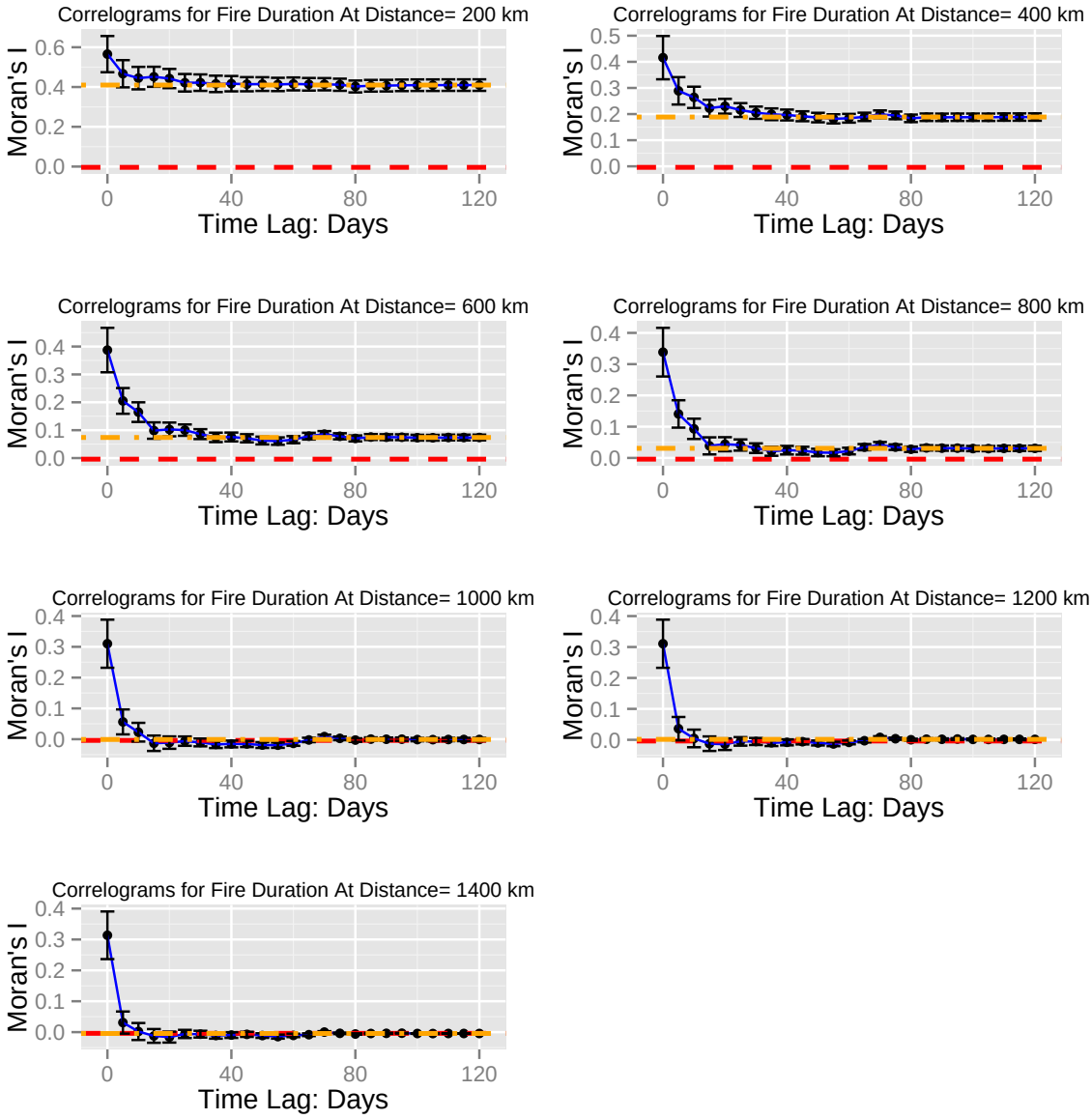


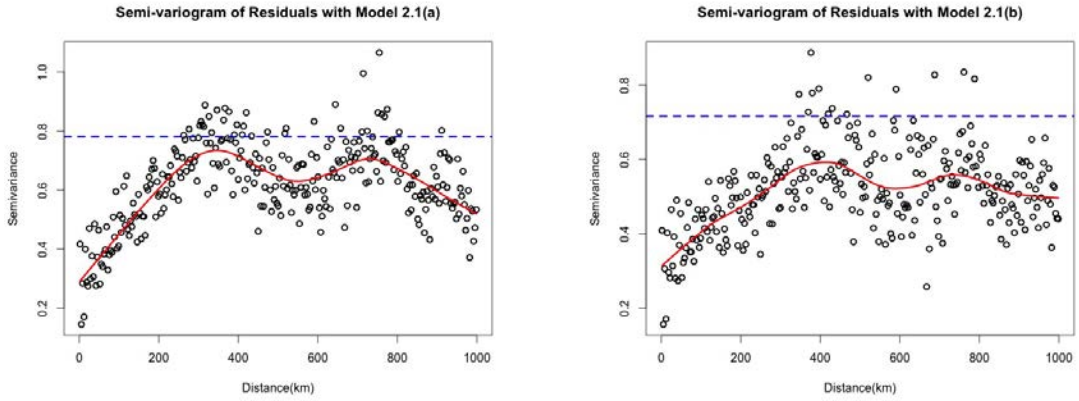
Figure 3.6: Extended Moran's I-based Spatio-Temporal Correlogram for Fire Duration

3.3 Application of Moran's *I* in Residuals Analysis

To validate the independence assumption of residuals from Model 2.1(a) and (b) in Chapter 2, we apply first the existing methods to evaluate the spatial autocorrelation and then apply the extended Moran's *I* to examine the spatio-temporal correlation with the residuals in the ordinary linear regression analysis.

3.3.1 Check Spatial Correlation in Residuals

First, we show the empirical semivariograms of residuals to check the spatial correlation graphically. Figure 3.7 presents the empirical semivariograms of residuals from Model 2.1(a) and (b). The red curves are the smoothing curves drawn by **lowess** function in R and the blue lines stand for the estimated variance of the residuals from the models. The semivariance increases as the distance of pair of fire locations becomes larger. The increase trends of semivariance with both Model 2.1(a) and (b) suggest a spatial autocorrelation within distance 400 km. Because of the fact that the variance of the residuals may



(a) Semi-variogram of Residuals in Model 2.1(a) (b) Semi-variogram of Residuals in Model 2.1(b)

Figure 3.7: Semi-variograms

not be constant, the semivariance for both models keep fluctuating below the estimated variance of residuals. As reviewed in section 3.1.1, a semivariogram may not indicate the spatial autocorrelation if the variance of the underlying process is not constant. We proceed to conduct the Moran's *I* test with the residuals in the following.

The spatial correlograms of obtained global Moran's *I* with different neighbors' defining distances are shown in Figure 3.8 with the 95% acceptance regions. The plots show that the significant spatial autocorrelation exists between fires till the distance between their locations up to 400 km. The test results obtained by defining fires within 400 km as neighbors are summarized in Table 3.2. The p_values with both models indicate the presence of strong spatial correlation within 400 km.

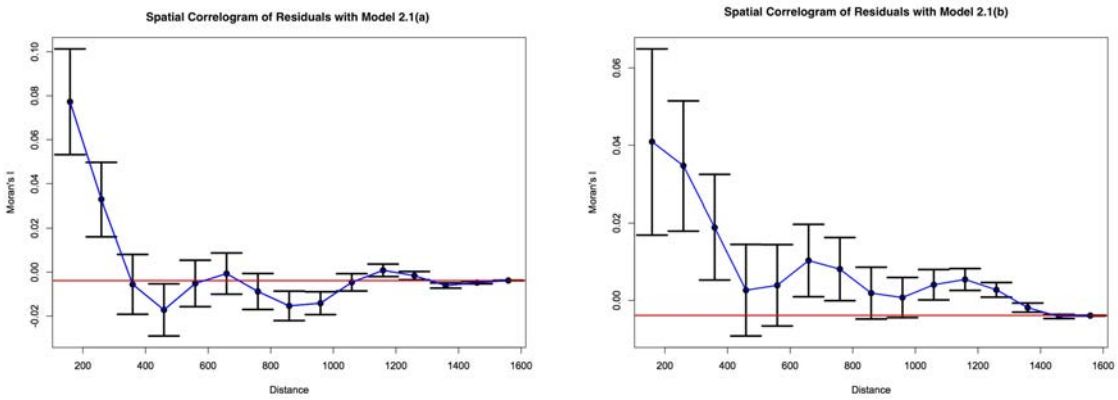
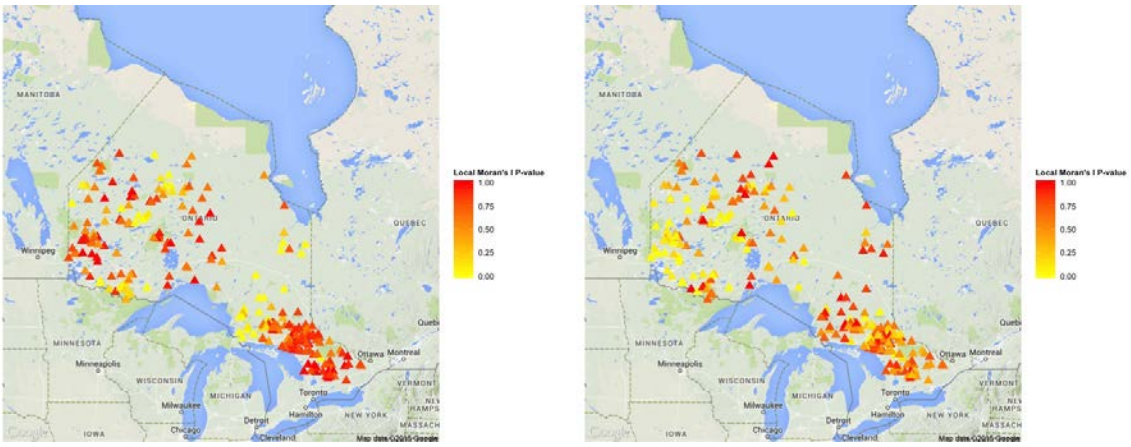


Figure 3.8: Spatial Correlograms for Model 2.1(a) and (b)

	Observed Moran's *I*	Expectation	Variance	p_value
Model 2.1(a)	-0.016	-0.0038	0.000043	0.06
Model 2.1(b)	0.011	-0.0038	0.000043	0.02

Table 3.2: Moran's *I* test on Model 2.1(a) and (b)

To investigate the local spatial clustering, we produce local Moran's *I* maps in Figure 3.9. The p_value of each point is adjusted for multiple testing and is plotted by different colors in the map. The yellow spots are those with $p\_value < 0.05$, indicating a significant spatial autocorrelation around the points. From the local Moran's *I* map, the yellow spots for Model 2.1(b) are more than the one of Model 2.1(a), indicating a stronger local spatial autocorrelation of residuals in Model 2.1(b).



(a) Local Moran's *I* for Model 2.1(a) (b) Local Moran's *I* for Model 2.1(b)

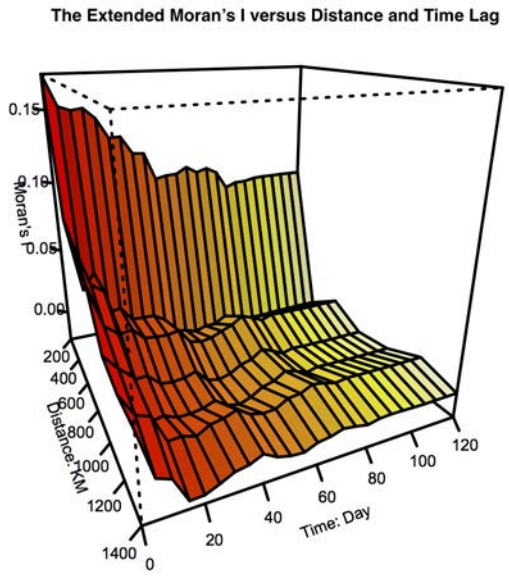
Figure 3.9: Local Moran's *I* Map for Model 2.1(a) and (b)

3.3.2 Check Spatio-Temporal Correlation in Residuals by Extended Moran's *I*

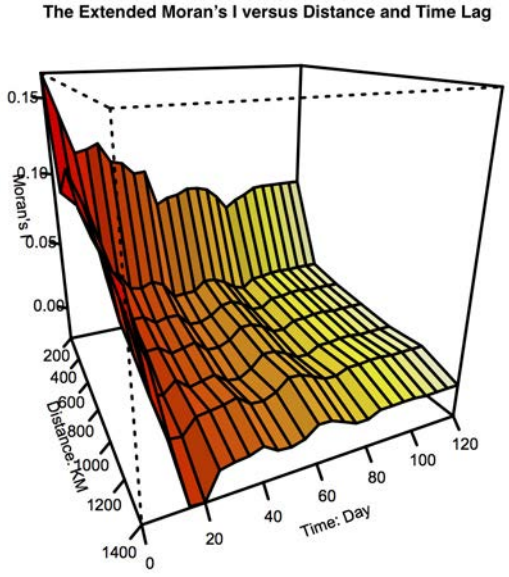
Since the fires occurred over time, it is often of interest to examine their spatio-temporal correlation. We apply the extended Moran's *I* to check the spatio-temporal correlation in the residuals. Following the settings in section 3.2, the correlograms based on extended Moran's *I* are produced.

Figure 3.10a and 3.10b show perspective correlograms plots which display extended Moran's *I* versus distance and time lag of fires. Figure 3.11 and 3.12 present the correlograms which plot extended Moran's *I* versus time lag of two fires according to different distance. Comparing with the correlograms of fire duration in Figures 3.4 and 3.5, the extended Moran's *I* is smaller for residuals, on average. The patterns of correlograms are similar: the extended Moran's *I* decreases as the distance and time lag between two fires increase.

From these plots, spatio-temporal correlation is found in residuals. Especially at distance 200 km, we see the residuals are highly correlated regardless of the starting time. With Model 2.1(a), the spatio-temporal correlation is eliminated when distance increases to 800 km. With Model 2.1(b), the spatio-temporal correlation exists for residuals within 10 days when distance increases to the maximum.



(a) Perspective Plot of Extended Moran's I for Residuals in Model 2.1(a)



(b) Perspective Plot of Extended Moran's I for Residuals in Model 2.1(b)

Figure 3.10: Perspective Plots of Extended Moran's I for Residuals with Ordinary Linear Regression Models

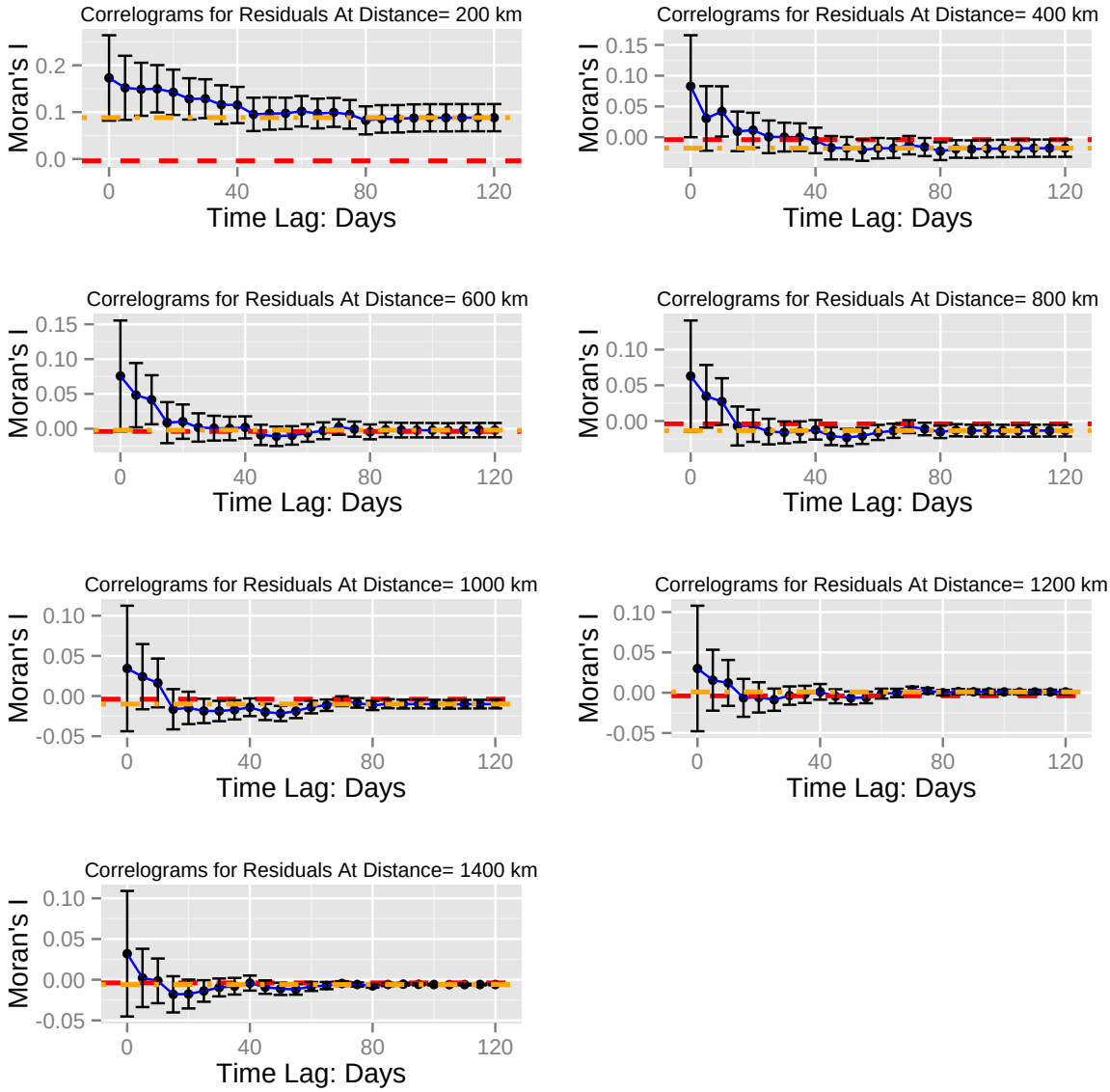


Figure 3.11: Extended Moran's I-based Spatio-Temporal Correlogram for Residuals in Model 2.1(a)

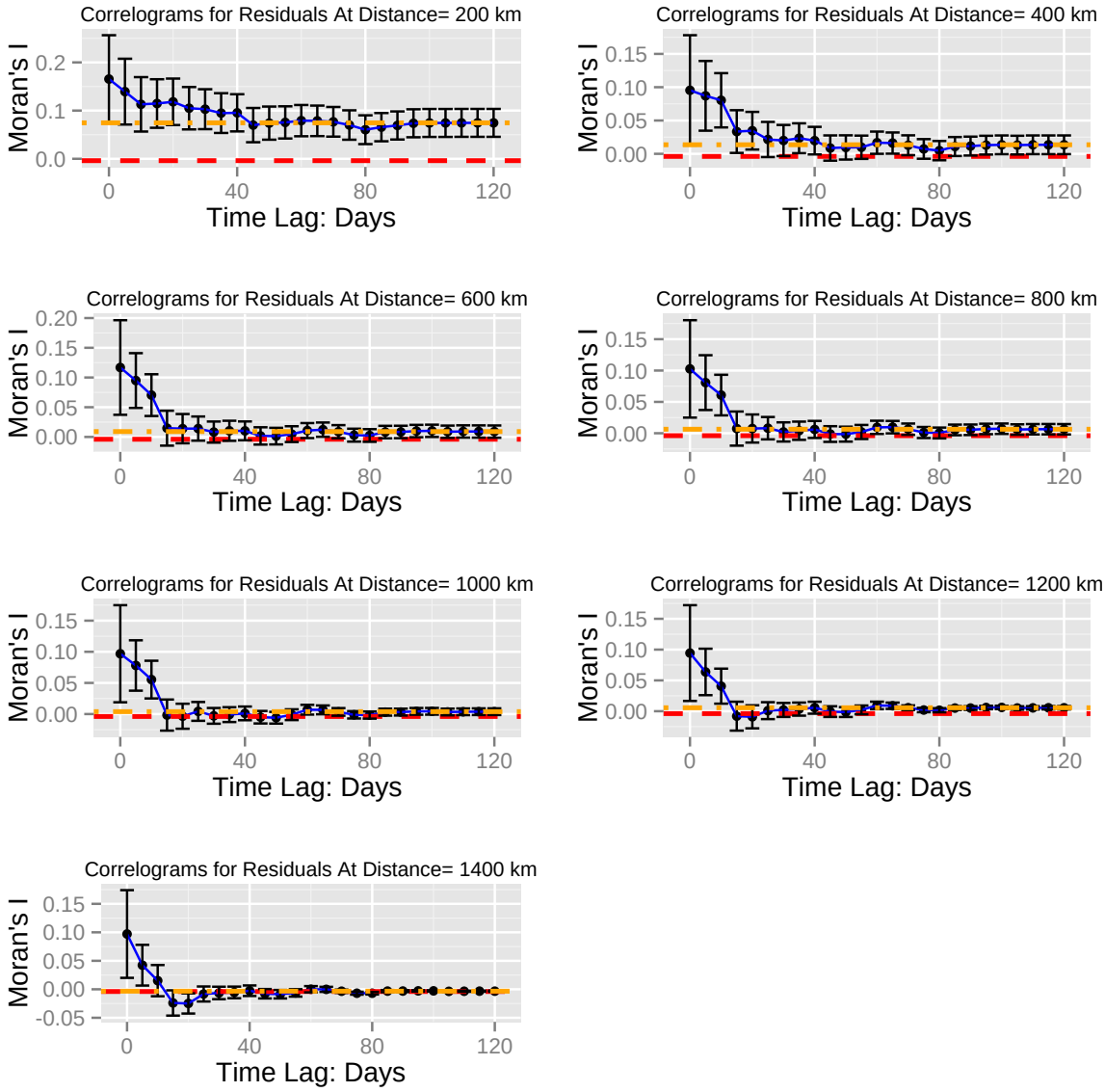


Figure 3.12: Extended Moran's I-based Spatio-Temporal Correlogram for Residuals in Model 2.1(b)

3.3.3 Discussion

The diagnostic results based on Moran's *I* statistics indicates the presence of spatio-temporal correlation in residuals from ordinary linear regression with Model 2.1(a) and (b). However, the expectation and variance of Moran's *I* derived by Cliff and Ord (1981) are based on the assumption that there is no spatial autocorrelation in the data and imply the stationarity of the data. So the large values of observed Moran's *I* might indicate either the presence of correlation or the nonstationarity. In order to address these issues, we can consider a more appropriate regression model to reduce the nonstationarity and then better explore the spatio-temporal correlation in the residuals. Alternatively, we can also focus on the specification of spatio-temporal correlation and obtain estimates by accounting for the correlation. In this project, we adopt the first approach and consider a partially linear regression model.

Chapter 4

Partially Linear Regression Analysis of Forest Fire Data

The linear regression analysis in Chapter 2 provides preliminary results, indicating fire duration is associated with fire's location, fire's starting time and *FWI*. However, the residual analysis suggests that the linear relationship might not be appropriate. The correlation tests suggest strong spatio-temporal correlation and nonstationarity in fire duration. This motivated us to consider a partially linear regression model to explore the association of fire duration with *FWI*, the fuel and weather index, adjusting for the potential spatial and temporal effects. We adopt nonparametric regression with Kernel smoothers in the analysis, and integrate it with the least squares estimator to assess the association. The partially linear regression may partially address the concern about the potential spatio-temporal correlation with conventional methods.

We propose two partially linear regression models with nonparametric time and location terms. The following starts by introducing the notation and modeling, and reviewing the estimation procedure with Kernel smoother by Speckman (1988). The regression analysis with the models are then presented. After that, we check the spatio-temporal correlation in residuals by using the extended methodology in Chapter 3.

4.1 Notation and Modelling

Recall that the mean function $\mu(\mathbf{s}, t; \mathbf{z})$ is specified as a linear function of a spatial variable, a starting time variable and the environmental factor in the ordinary linear regression analysis in Chapter 2. Here, we model $\mu(\mathbf{s}, t; \mathbf{z})$ into a nonparametric component of time and location and a linear component of covariates $\mathbf{z}$:

$$Y_i = \mu_0(\mathbf{s}_i, t_i) + \beta' \mathbf{z}_i + \epsilon_i \tag{4.1}$$

Two special cases of $\mu_0(\mathbf{s}_i, t_i)$ are considered. Specifically, we consider the following models,

$$Y_i = h_{g(\mathbf{s}_i)}(t_i) + \beta_{g(\mathbf{s}_i)} z_i + \epsilon_i, \tag{4.2a}$$

$$Y_i = h_{g(\mathbf{s}_i)}(t_i) + \beta z_i + \epsilon_i,\ i = 1, 2, \cdots, 259, \tag{4.2b}$$

where $g(\mathbf{s}_i)$ is the code of the management zone of fire i, and $h_{g(\mathbf{s}_i)}(\cdot)$ is to be estimated. We also consider:

$$y_i = g_{h(t_i)}(\mathbf{s}_i) + \beta_{h(t_i)} z_i + \epsilon_i, \tag{4.3a}$$

$$y_i = g_{h(t_i)}(\mathbf{s}_i) + \beta z_i + \epsilon_i,\ i = 1, 2, \cdots, 259, \tag{4.3b}$$

where $h(t_i)$ is fire i's starting time in month of May, June, July, August or September, and $g_{h(t_i)}(\cdot)$ is unspecified. We group fires starting in August and September in the same category to balance the number of fires in each category in the analysis.

The random errors $\epsilon_i = \epsilon(\mathbf{s}_i, t_i)$ are assumed to be identically and independently from a normal distribution with mean 0 and a constant variance σ^2.

4.2 Partially Linear Regression Analysis with Models 4.2(a) and 4.2(b)

4.2.1 Estimation Procedures

We adopt local constant and local linear estimation methods in estimating the nonparametric function $h_{g(\mathbf{s})}(t)$, and apply the least squares estimation (LSE) for estimating parameter β. The variance σ^2 is estimated using the adjusted residuals.

Local Constant Estimation

With Model (4.2a) for any fixed t, we approximate $h_{g(\mathbf{s}_i)}(t_i)$ by $h_{g(\mathbf{s}_i)}(t)$. Then we minimize the weighted sum squares in each fire management zone r:

$$\sum_{i:g(\mathbf{s}_i)=r} (y_i - \beta_{g(\mathbf{s}_i)} z_i - h_{g(\mathbf{s}_i)}(t))^2 K(\frac{t-t_i}{d})$$

where $r = 1, \cdots, R$, (in this dataset $R = 3$ for three fire management zones: I, M and E), and $K(\cdot)$ is a Kernel function of the bandwidth d. Here, we use the Tukeys Tricube weight function:

$$K(u) = \begin{cases} (1-|u|^3)^3 & \text{if } |u| < 1; \\ 0 & \text{if } |u| > 1. \end{cases}$$

By taking derivatives and after some algebra, we obtained

$$\hat{h}_r(t) = \frac{\sum_{g(\mathbf{s}_i)=r}(y_i - \beta_r z_i) K(\frac{t-t_i}{d})}{\sum_{g(\mathbf{s}_i)=r} K(\frac{t-t_i}{d})}. \tag{4.4}$$

In a matrix form, we write $\mathbf{Y}_r$ as a sub-vector of $\mathbf{Y} = [y_1, y_2, \cdots, y_{259}]^{'}$ and $\mathbf{Z}_r$ as a sub-vector of $\mathbf{Z} = [z_1, z_2, \cdots, z_{259}]^{'}$with y_i and z_i from stratum $g(\mathbf{s}_i) = r$. Then the matrix form of equation (4.4) is:

$$\hat{h}_r(t) = (\mathbf{1}^{'}\mathbf{K}_r\mathbf{1})^{-1}\mathbf{1}^{'}\mathbf{K}_r(\mathbf{Y}_r - \beta_r\mathbf{Z}_r) \tag{4.5}$$

where $\mathbf{K}_r$ is a diagonal matrix for fire management zone r with diagonal elements being

$K_r(i,i) = K(\frac{t_i - t}{d})I(g(\mathbf{s}_i) = r)$. We use $\mathbf{S}_r = (\mathbf{1}'\mathbf{K}_r\mathbf{1})^{-1}\mathbf{1}'\mathbf{K}_r$ as a smoothing matrix of fire management zone r.

By plugging $\hat{h}_r(t)$ of (4.5) in equation (4.4), we can estimate β_r by minimizing $||(\mathbf{I} - \mathbf{S}_r)(\mathbf{Y}_r - \beta_r\mathbf{Z}r)||^2$. It yields $\hat{\beta}_r = (\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Y}}_r$, where $\tilde{\mathbf{Y}}_r = (\mathbf{I} - \mathbf{S}_r)\mathbf{Y}_r, \tilde{\mathbf{Z}}_r = (\mathbf{I} - \mathbf{S}_r)\mathbf{Z}_r$. The fitted values $\hat{\mathbf{Y}}_r$ are then

$$\begin{aligned}\hat{\mathbf{Y}}_r &= \mathbf{S}_r(\mathbf{Y}_r - \hat{\beta}_r\mathbf{Z}_r) + \hat{\beta}_r\mathbf{Z}_r \\ &= \mathbf{S}_r\mathbf{Y}_r + \tilde{\mathbf{Z}}_r(\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}\tilde{\mathbf{Z}}_r(\mathbf{I}_r - \mathbf{S}_r)\mathbf{Y}_r \\ &= \{\mathbf{S}_r + \tilde{\mathbf{Z}}_r(\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}\tilde{\mathbf{Z}}_r(\mathbf{I}_r - \mathbf{S}_r)\}\mathbf{Y}_r \end{aligned} \tag{4.6}$$

For Model (4.2b)(in the following), where β is fixed across zones, we minimize the objective function below instead,

$$\sum_{r=1}^{R}\sum_{i:g(\mathbf{s}_i)=r}(y_i - \beta z_i - h_{g(\mathbf{s}_i)}(t))^2K(\frac{t - t_i}{d}). \tag{4.7}$$

From the estimation procedure with Model (4.2b), the estimators of $h_r(t)$ and β are

$$\begin{aligned}\hat{h}_r(t) &= (\mathbf{1}'\mathbf{K}_r\mathbf{1})^{-1}\mathbf{1}'\mathbf{K}_r(\mathbf{Y}_r - \beta\mathbf{Z}_r), \\ \hat{\beta} &= (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-\mathbf{1}}\tilde{\mathbf{Z}}'\tilde{\mathbf{Y}}\end{aligned}$$

where $\tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S})\mathbf{Y}, \tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S})\mathbf{Z}$ and $\mathbf{S} = Diag[\mathbf{S}_1, \cdots, \mathbf{S}_r, \cdots, \mathbf{S}_R]$. Then the fitted values $\mathbf{Y}$ can be obtained as: $\mathbf{Y} = \{\mathbf{S} + \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-\mathbf{1}}\tilde{\mathbf{Z}}'(\mathbf{I} - \mathbf{S})\}\mathbf{Y}$.

Local Linear Estimation

Local constant estimator approximates the function $h_{g(\mathbf{s}_i)}(t_i)$ at a fixed t by taking an average of the values for observations such that t_i are in a neighborhood of t. Similar with the local constant estimator, the idea of a local linear estimator is to approximate $h_{g(\mathbf{s}_i)}(t_i)$ for any fixed t by a linear function $h_{g(\mathbf{s}_i)}(t) + \dot{h}_{g(\mathbf{s}_i)}(t)(t_i - t)$, where $\dot{h}_{g(\mathbf{s}_i)}(t)$ is the first derivative of $h_{g(\mathbf{s}_i)}(t)$.

With Model (4.2a), we minimize weighted sum squares

$$\sum_{i:g(\mathbf{s}_i)=r} (y_i - \beta_r z_i - h_{g(\mathbf{s}_i)}(t) - \dot{h}_{g(\mathbf{s}_i)}(t)(t_i - t))^2 K(\frac{t - t_i}{d}). \tag{4.8}$$

Then the estimators of $h_r(t)$ and its first derivative $\dot{h}_r(t)$ are obtained:

$$\left(\hat{h}_r(t), \hat{\dot{h}}_r(t)\right)' = (\boldsymbol{X}_r(t)'\mathbf{K}_r\boldsymbol{X}_r(t))^{-1}\boldsymbol{X}_r(t)'\mathbf{K}_r(\mathbf{Y}_r - \beta_r\mathbf{Z}_r)$$

where $\boldsymbol{X_r(t)} = (\mathbf{1}, \triangle)$ where $\triangle$ is a vector of components $(t_i - t)$ with $g(\mathbf{s}_i) = r$. The estimator of $h_r(t)$ is obtained as the first row component of $\left(\hat{h}_r(t), \hat{\dot{h}}_r(t)\right)'$.

We use $\mathbf{S^*}_r(t) = (\boldsymbol{X}_r(t)'\mathbf{K}_r\boldsymbol{X}_r(t))^{-1}\boldsymbol{X}_r(t)'\mathbf{K}_r$ as a smoothing matrix for fire management zone r. We can then obtain the estimator of β_r by minimizing (4.8) with respect to β_r:

$$\hat{\beta}_r = (\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Y}}_r$$

where $\tilde{\mathbf{Y}}_r = (\mathbf{I} - \mathbf{S^*}_r)\mathbf{Y}_r$, $\tilde{\mathbf{Z}}_r = (\mathbf{I} - \mathbf{S^*}_r)\mathbf{Z}_r$. The fitted values are then $\hat{\mathbf{Y}}_r = \{\mathbf{S^*}_r + \tilde{\mathbf{Z}}_r(\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}\tilde{\mathbf{Z}}_r'(\mathbf{I} - \mathbf{S^*}_r)\}\mathbf{Y}_r$.

Similarly we can estimate $h_r(t)$ and β with Model (4.2b):

$$\left(\hat{h}_r(t), \hat{\dot{h}}_r(t)\right)' = (\boldsymbol{X}_r(t)'\mathbf{K}_r\boldsymbol{X}_r(t))^{-1}\boldsymbol{X}_r(t)'\mathbf{K}_r(\mathbf{Y}_r - \beta\mathbf{Z}_r)$$
$$\hat{\beta} = (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{Y}}$$

where $\tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S^*})\mathbf{Y}, \tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S^*})\mathbf{Z}$ and $\mathbf{S^*} = Diag[\mathbf{S^*}_1, \cdots, \mathbf{S^*}_r, \cdots, \mathbf{S^*}_R]$. Then the fitted values $\hat{\mathbf{Y}}$ can be also be obtained as $\hat{\mathbf{Y}} = \{\mathbf{S^*} + \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-\mathbf{1}}\tilde{\mathbf{Z}}'(\mathbf{I} - \mathbf{S^*})\}\mathbf{Y}$.

4.2.2 Bandwidth Selection

Choosing the bandwidth in Kernel smoothing is important. Various bandwidth selection methods are discussed in the literature. We use the generalized Cross-Validation criterion proposed by Craven and Wahba (1978) to select the bandwidth. The GCV function is defined as

$$GCV(d) = \frac{RSS(d)}{[1 - n^{-1}trace(\mathbf{A})]^2},$$

where RSS is the average residual sum of squares and $\mathbf{A}$ is the hat matrix which satisfies $\hat{\mathbf{Y}} = \mathbf{AY}$. The optimal bandwidth is selected as the lowest GCV value.

4.2.3 Estimation of Variance

To estimate the variance of residuals in the partial linear model with kernel smoothing, we use the estimation method by Speckman (1988), which can be written as:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - trace(\mathbf{A})}, \text{where } \mathbf{A} \text{ is the hat matrix.}$$

Then we can estimate the standard error of β in the parametric part. Specifically, with Model (4.2a), we have $\hat{Var}(\hat{\beta}_r) = \hat{\sigma}^2\tilde{\mathbf{Z}}_r(\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}(\mathbf{I} - \mathbf{S}_r)(\mathbf{I} - \mathbf{S}_r)\tilde{\mathbf{Z}}_r(\tilde{\mathbf{Z}}_r'\tilde{\mathbf{Z}}_r)^{-1}$, and with Model (4.2b), $\hat{Var}(\hat{\beta}) = \hat{\sigma}^2\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}(\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}$.

4.2.4 Analysis Results

In this section, we present the analysis results, including bandwidth selection and estimation results with Models (4.2a) and (4.2b).

Bandwidth Selection of $h(t)$ in Each Zone

We first choose the bandwidth for $h(t)$ in each fire zone by the lowest GCV values. Table 4.1 displays the bandwidth in different fire management zones for local constant and local linear estimators with Model (4.2a) and (4.2b). Figures 4.1, 4.2, 4.3, 4.4 show the scatterplots of GCV versus different bandwidths. The chosen bandwidths for local linear estimator with both the models are the same. Due to the fact that there are fewer fires in zone M and zone E, the bandwidths in zone M and zone E are larger than the one in zone I.

	Model (4.2a)			Model (4.2b)		
	$h_{\text{Fmz=I}}(t)$	$h_{\text{Fmz=M}}(t)$	$h_{\text{Fmz=E}}(t)$	$h_{\text{Fmz=I}}(t)$	$h_{\text{Fmz=M}}(t)$	$h_{\text{Fmz=E}}(t)$
Local Constant	11	11	15	11	29	5
Local Linear	9	29	29	9	29	29

Table 4.1: Bandwidth Selection of Model (4.2a) and (4.2b)

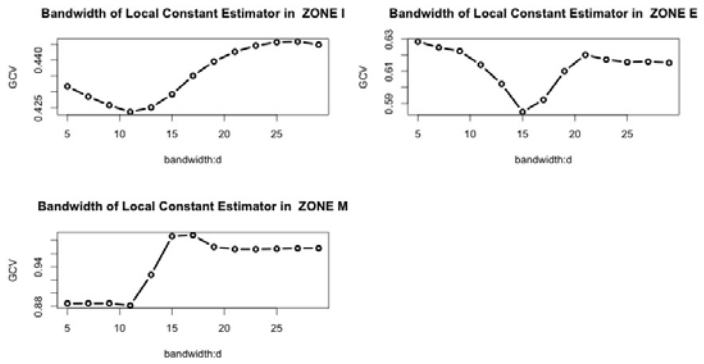


Figure 4.1: Bandwidth Selection for Local Constant Estimator of $h(t)$ with Model (4.2a)

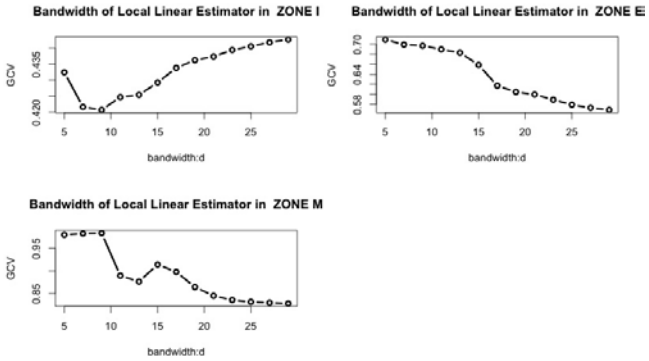


Figure 4.2: Bandwidth Selection for Local Linear Estimator of $h(t)$ with Model (4.2a)

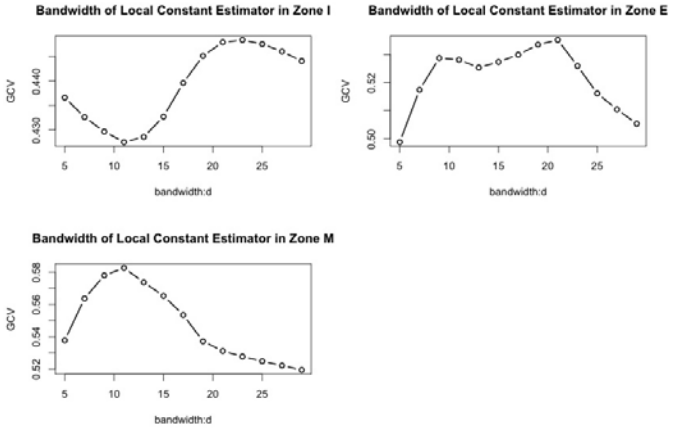


Figure 4.3: Bandwidth Selection for Local Constant Estimator of $h(t)$ with Model (4.2b)

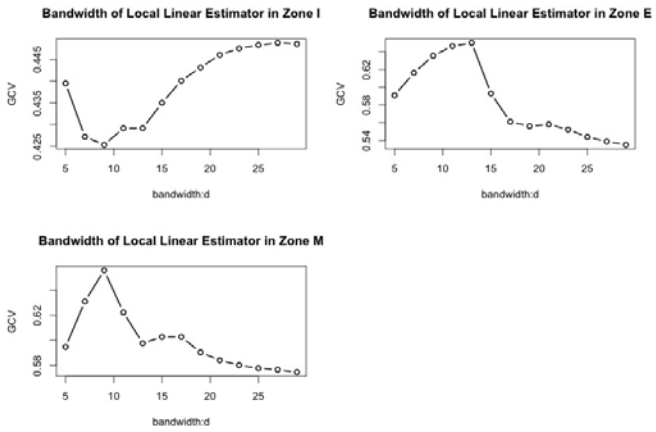


Figure 4.4: Bandwidth Selection for Local Linear Estimator of $h(t)$ with Model (4.2b)

Estimates of Regression Parameter and Nonparametric Function

Table 4.2 presents the estimates of regression coefficient β with Model (4.2a) and (4.2b) using the optimal bandwidth chosen by GCV. The estimated standard errors and the p_values of the Wald-test on the significance are also displayed in the table. The statistically significant predictors' effects are marked as bold. Model (4.2a) considers the situation that the effect of *FWI* on fire duration is stratified to each fire management zone. We see that the estimates of *FWI* coefficients vary across different fire zones with Model (4.2a). The estimated coefficients of *FWI* with Model (4.2b) are similar and statistically significant by both local constant and local linear estimators.

When we compare the results of Model (4.2a) and (4.2b), the estimated coefficient of *FWI* β in Model (4.2b) can be approximated by a weighted average of β_I, β_M, β_E in Model (4.2a). The similarity of estimates by local constant and local linear estimators with both the models suggest the robustness of estimators and important effect of *FWI*.

The estimated coefficients of *FWI* with both Model (4.2a) and (4.2b) are negative, which is in agreement with the linear regression analysis in Chapter 2. Since a high value of *FWI* indicates that the fire might be dangerous, the fire agency will provide more resources to put out that fire. Thus the fire duration might decrease as the value of *FWI* increases.

The smoothing curves of fire's starting time are obtained by local constant and local linear estimators in three different zones. Figure 4.5 and Figure 4.6 show the smoothing

	Model (4.2a)						Model (4.2b)	
	Local Constant			Local Linear			Local Constant	Local Linear
	$\beta_{\mathbf{I}}$	$\beta_{\mathbf{M}}$	$\beta_{\mathbf{E}}$	$\beta_{\mathbf{I}}$	$\beta_{\mathbf{M}}$	$\beta_{\mathbf{E}}$	$\beta_{\mathbf{local\ constant}}$	$\beta_{\mathbf{local\ linear}}$
Estimates	-0.23	**-2.99**	-0.24	-0.21	**-2.86**	-0.16	**-0.47**	**-0.49**
Std.Error	0.36	0.91	0.71	0.36	0.91	0.71	0.23	0.23
P−value	0.26	0.005	0.36	0.28	<0.001	0.41	0.02	0.01
MSE	0.450			0.447			0.485	0.467

Table 4.2: Estimates of Regression Coefficients for Model (4.2a) and (4.2b)

curves of fire's starting time obtained by local constant and local linear estimators with Model (4.2a). Figure 4.7 and Figure 4.8 show the corresponding smoothing curves with Model (4.2b). Both sets of curves fit the data well and behave similarly with the two underlying models.

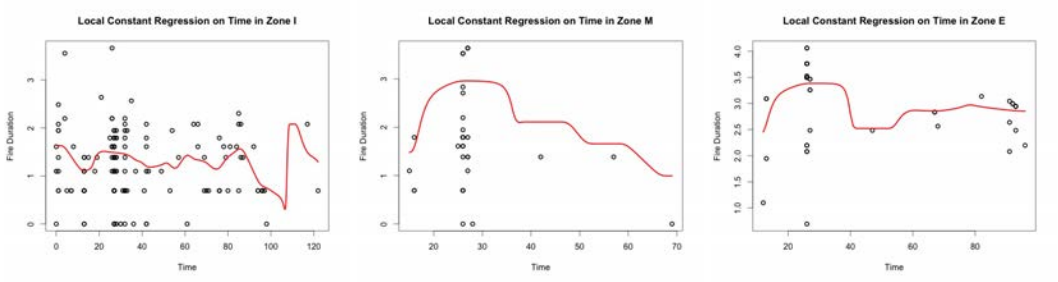


Figure 4.5: Local Constant Smoothing Curves in Zone I, M and E for Model 4.2(a)

.

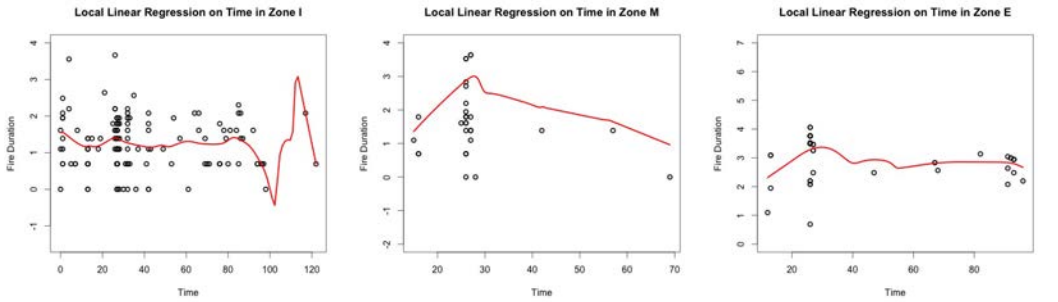


Figure 4.6: Local Linear Smoothing Curves in Zone I, M and E for Model 4.2(a)

.

4.2.5 Residual Analysis

For both Model (4.2a) and (4.2b), we have assumed that ϵ_i are distributed independently and identically from a normal distribution with mean 0 and constant variance σ^2. Now we

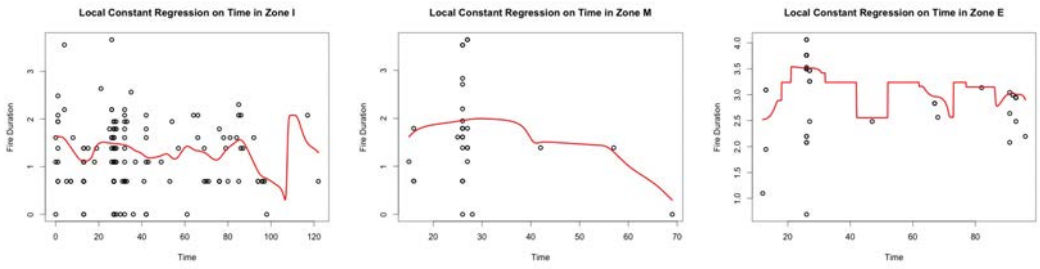


Figure 4.7: Local Constant Smoothing Curves in Zone I, M and E for Model 4.2(b)

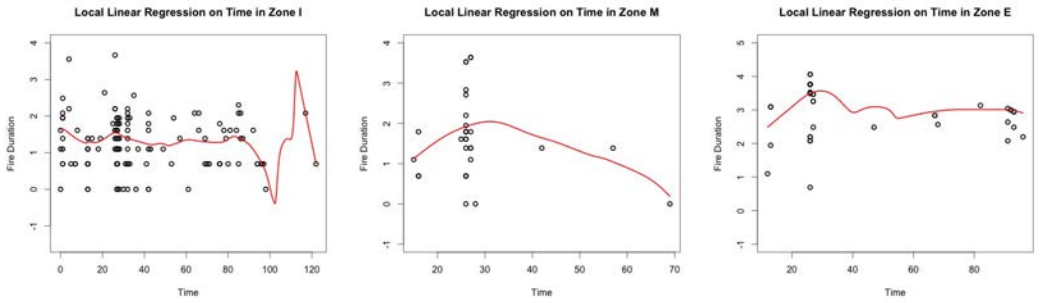


Figure 4.8: Local Linear Smoothing Curves in Zone I, M and E for Model 4.2(b)

explore whether the corresponding model assumptions are valid. Since the results from local constant and local linear estimators are similar, we only present the residual analysis of Model (4.2a) and (4.2b) with local linear estimators.

Normal Q-Q plots are used to examine the normal assumption of residuals. Figure 4.9 shows the Q-Q plots for residuals from Model (4.2a) and (4.2b). The points are distributed very closely to the red Q-Q lines. So the normal assumption is reasonable for the residuals. The scatterplot of residuals versus predicted value, residuals versus FWI and residuals versus start date are plotted in Figure 4.10 and 4.11 to check the mean and variance of residuals. The red lines are the locally weighted regression smoothing curves computed by **lowess** function in R. The plots show that residuals are distributed around zero so the mean of residuals is approximately 0 for both models. However, there seems a residual variation decreasing over time: the variance of residuals becomes smaller when fire starts in late part of fire season, i.e. in August and September.

The residual maps of Model (4.2a) and (4.2b) are shown in Figure 4.12a and 4.12b. The residual distributions are quite similar with both models. The maps don't show obvious

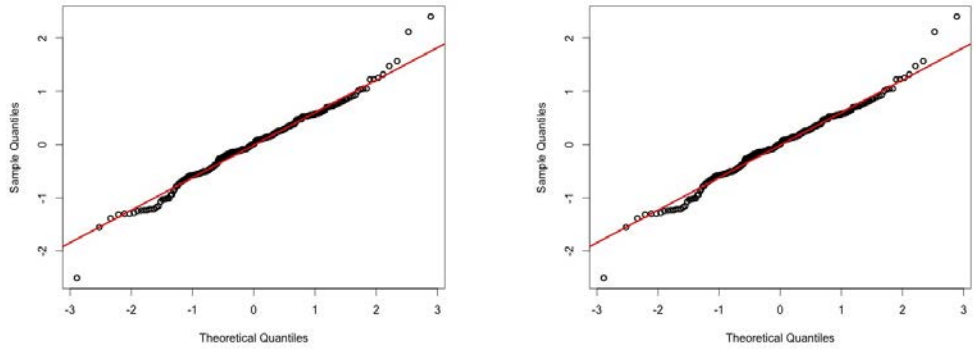


Figure 4.9: Normal QQ Plot of Residuals in Model (4.2a) and Model (4.2b) by Local Linear Estimators

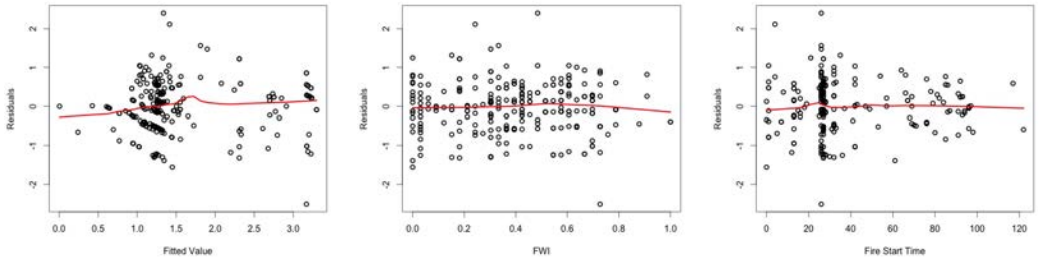


Figure 4.10: Residual Plots of Model (4.2a) by Local Linear Estimator

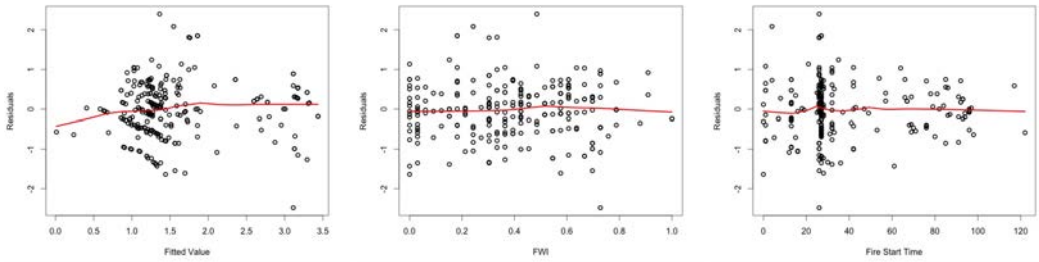
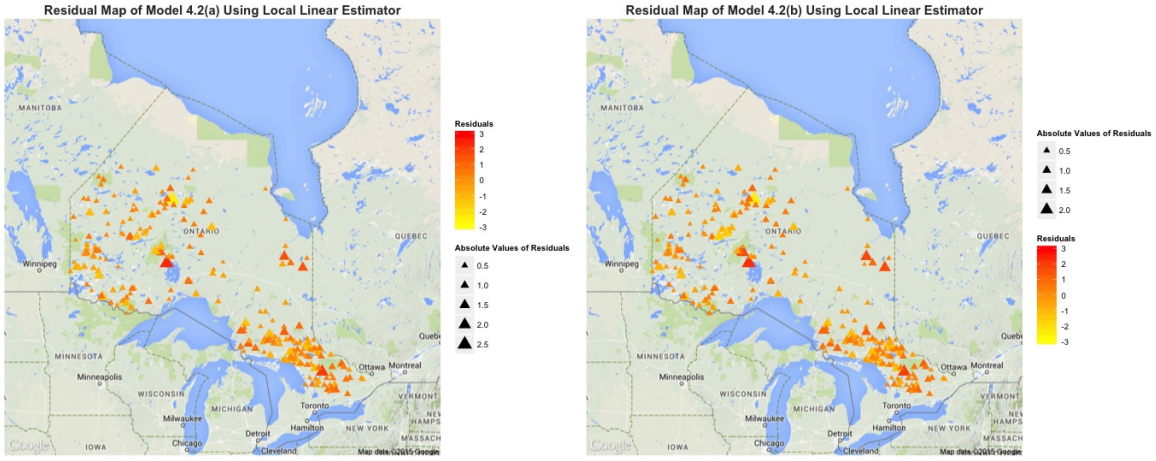


Figure 4.11: Residual Plots of Model (4.2b) by Local Linear Estimator

patterns of residuals, which suggests that residuals are distributed with mean 0. Comparing with the residual maps in Figure 2.10a and Figure 2.10b in ordinary linear regression, there are fewer large and red spots in the residual maps of Model (4.2a) and (4.2b). As discussed previously, this suggests better fits with the current models.



(a) Residual Map for Model (4.2a)

(b) Residual Map for Model (4.2b)

Figure 4.12: Residual Maps

We proceed to examine the independence assumption by checking if residuals are correlated in space and time. Figure 4.13 shows the semivariogram of residuals from Model (4.2a) and (4.2b). From the semivariograms, it is clear that there is an increase trend of semivariance in a small area within 400 km. This indicates that there is a spatial autocorrelation of fires whose starting locations are within 400 km. However, the semivariance goes down instead of reaching to the estimated variance of residuals, making it likely that the residuals are non-stationary and leading us to use Moran's *I* to assess the spatial correlation.

First, we conduct the global Moran's *I* test on residuals. Based on the information from the semivariograms, we define the fires whose starting locations are 400 km apart from each other as neighbors. Table 4.3 summarizes the test results. The global Moran's *I* test results reveal that there is no statistically significant spatial autocorrelation among residuals with Model (4.2a) and (4.2b).

To provide more evidence of the improvement with the partially linear models, we produce the spatial correlograms to show how Moran's *I* varies according to the distance of

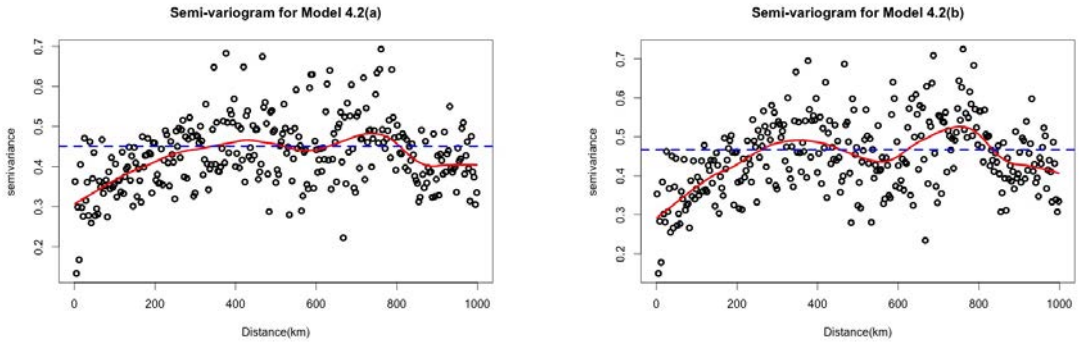


Figure 4.13: Semivariogram of Residuals of Model (4.2a) and (4.2b) by Local Local Linear Estimator

	Observed Moran's *I*	Expectation	Variance	p_value
Model (4.2a)	-0.010	-0.0038	0.000043	0.32
Model (4.2b)	-0.011	-0.0038	0.000043	0.31

Table 4.3: Moran's *I* test on Model (4.2a) and (4.2b)

neighboring fires. Figure 4.14 shows the spatial correlograms of Model (4.2a) and (4.2b). Comparing with the spatial correlograms in Figure 3.8 from ordinary linear regression, Model (4.2a) and (4.2b) reduce the spatial autocorrelation because the 95% acceptance regions include the expectation value of Moran's *I* except at distance 1100 and 1200 km.

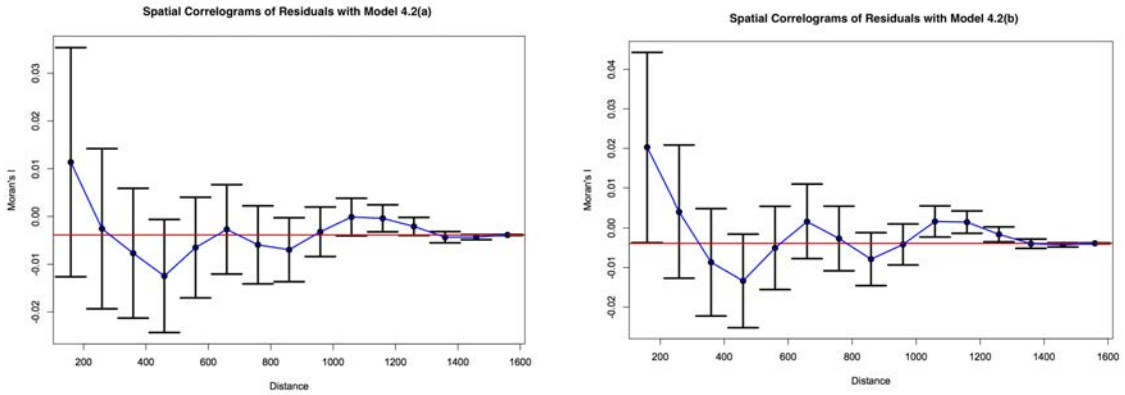
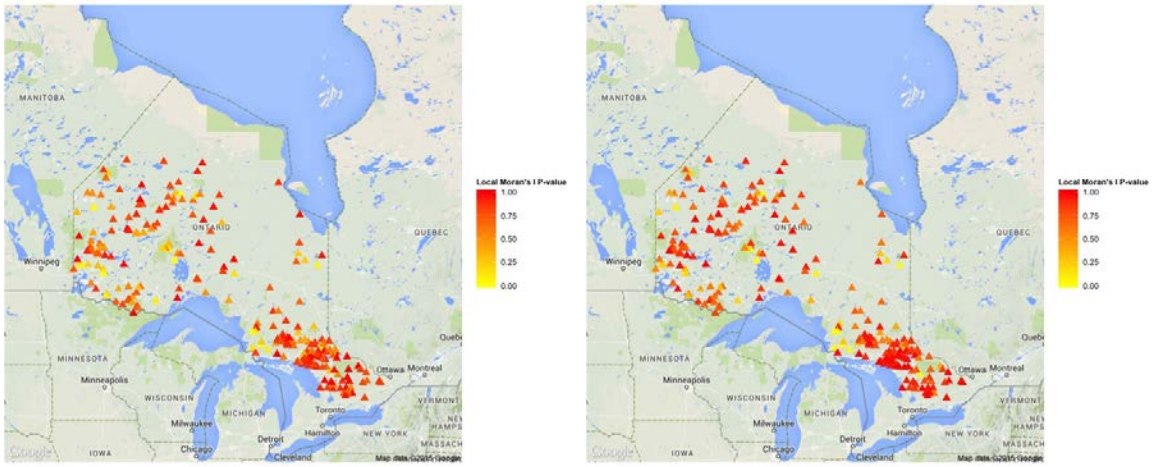


Figure 4.14: Spatial Correlograms for Model (4.2a) and (4.2b)

Furthermore, the local Moran's *I* maps for residuals also show that spatial autocorrelation is mostly eliminated. Figure 4.15 displays maps of local Moran's *I* test results for residuals of Model (4.2a) and (4.2b). Comparing with maps by the same settings in Figure 3.9 in Chapter 3, we can see the numbers of yellow spots for Model (4.2a) and (4.2b) are

much less than the ones from linear regression analysis.



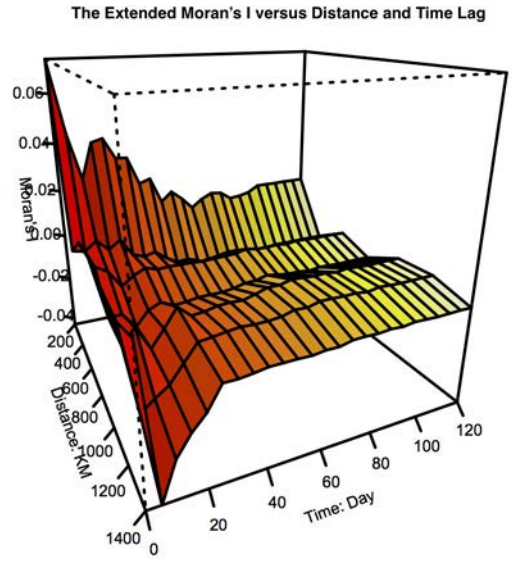
(a) Local Moran's I for Model (4.2a)

(b) Local Moran's I for Model (4.2b)

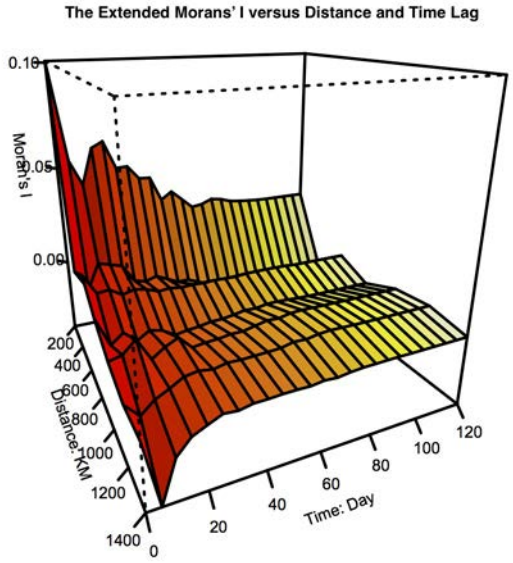
Figure 4.15: Local Moran's I Map for Model (4.2a) and (4.2b)

After detecting the presence of spatial autocorrelation in residuals, we proceed to check the spatio-temporal correlation by using the extended methods of Moran's *I*. Figure 4.16a and 4.16b show perspective correlograms of extended Moran's *I*. Figures 4.17 and 4.18 present the correlograms which plot extended Moran's *I* with the difference of fires' starting dates at different distance. For Model (4.2a), the correlograms at each distance show that 95% acceptance regions for extended Moran's *I* test contain the expectation values of extended Moran's *I* under the null hypothesis that there is no spatio-temporal correlation. For Model (4.2b), the correlogram at distance 200 km shows that 95% acceptance region doesn't contain the expectation value at time lag 0, which indicates a significant spatio-temporal correlation for fires whose locations are within 200 km and starting on the same day.

The maps of p_values based on the extended local Moran's *I* and adjusted for multiple tests are also presented in Figures 4.19 and 4.20 to evaluate local clusters. We calculate the extended local Moran's *I* by choosing the pre-determined distance as 200 km and time lags as $0, 5, \cdots, 30$ days. The patterns are similar for both models while the number of yellow points (with $p\_value < 0.05$, indicating a significant local spatio-temporal correlation) is lessened with Model (4.2b).



(a) Perspective Plot of Extended Moran's I for Residuals in Model (4.2a)



(b) Perspective Plot of Extended Moran's I for Residuals in Model (4.2b)

Figure 4.16: Perspective Plot of Extended Moran's I of Residuals with Model (4.2a) and (4.2b)

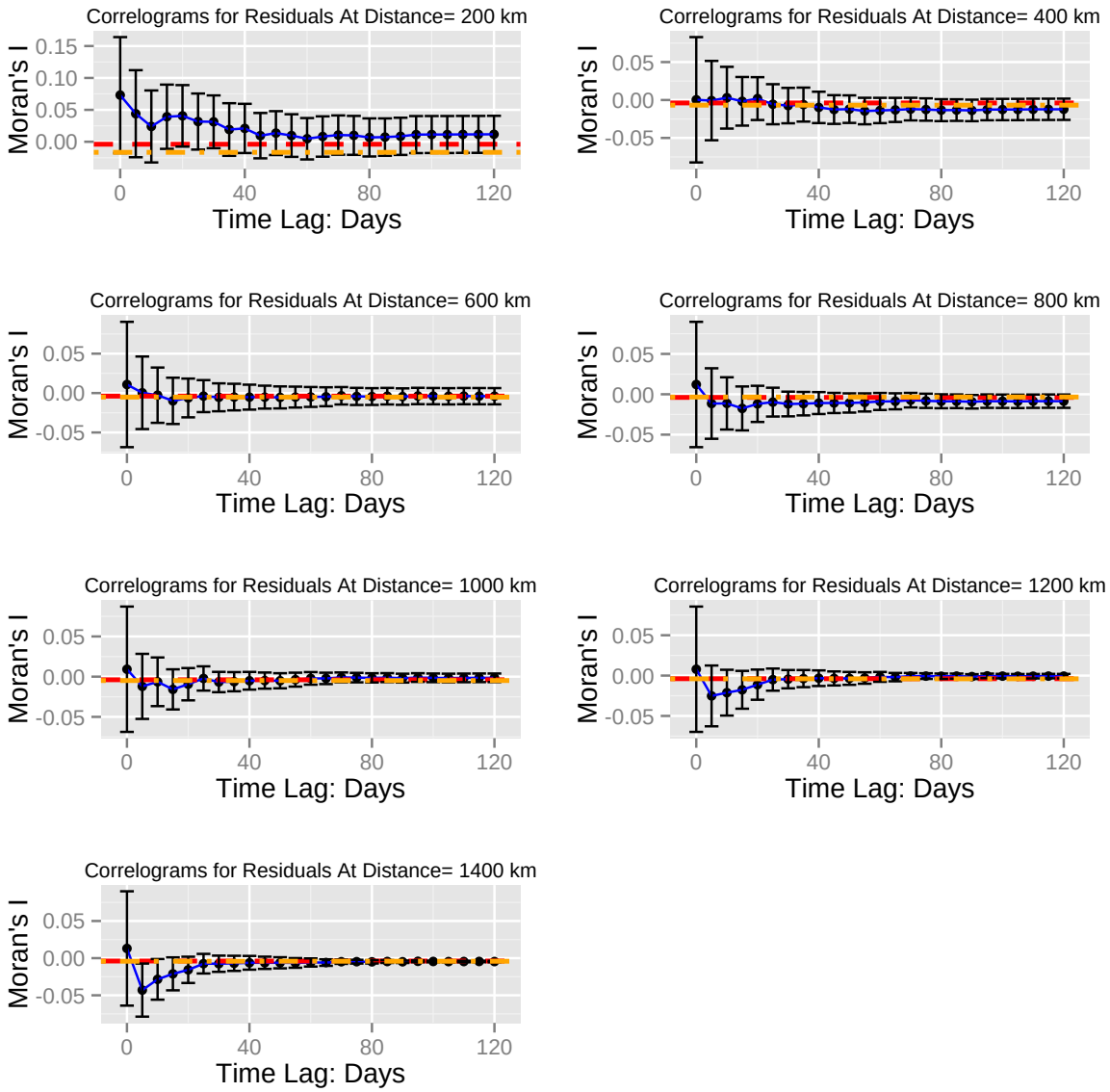


Figure 4.17: Extended Moran's I-based Spatial-Temporal Correlogram for Residuals in Model (4.2a)

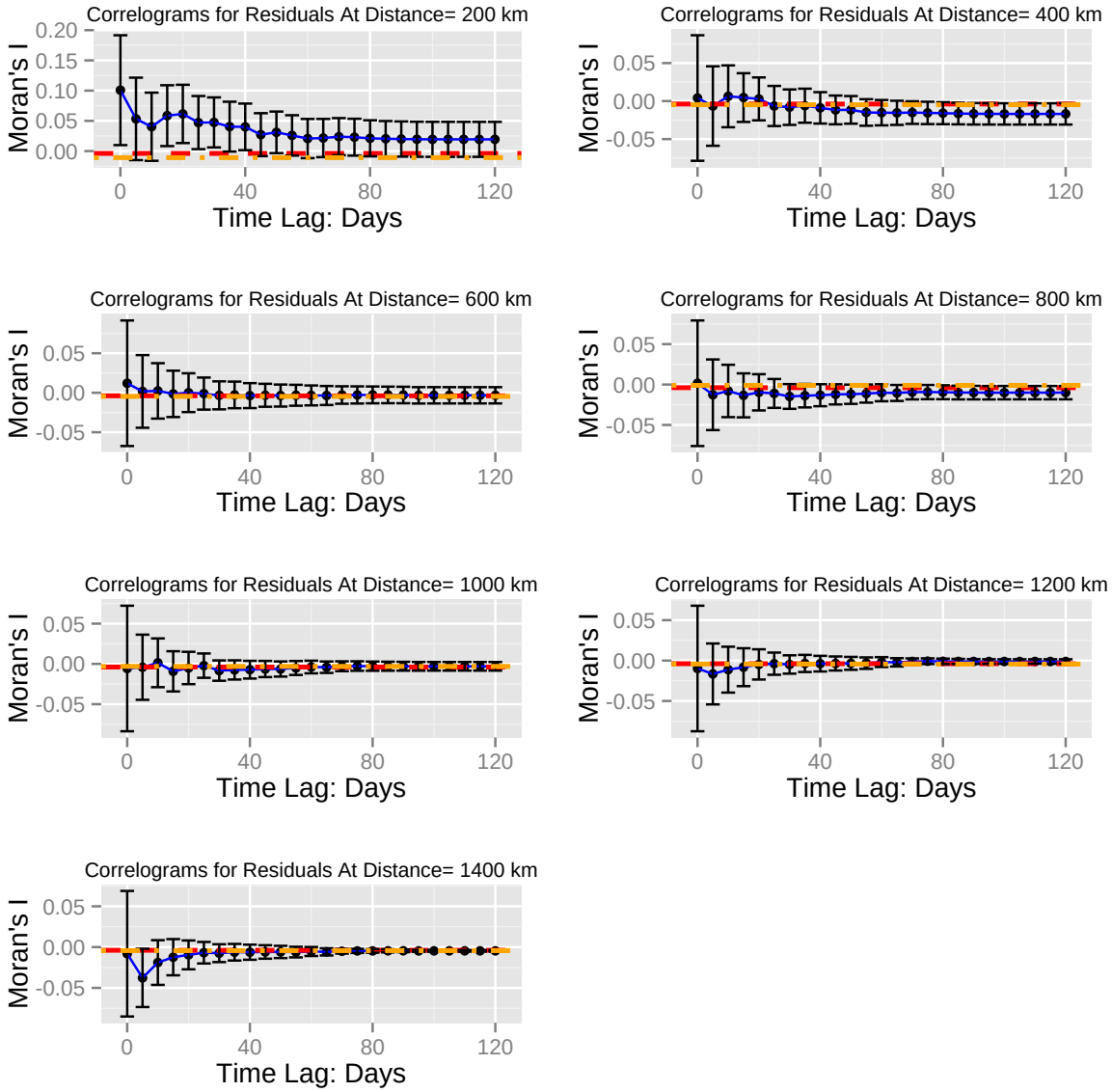


Figure 4.18: Extended Moran's I-based Spatial-Temporal Correlogram for Residuals in Model (4.2b)

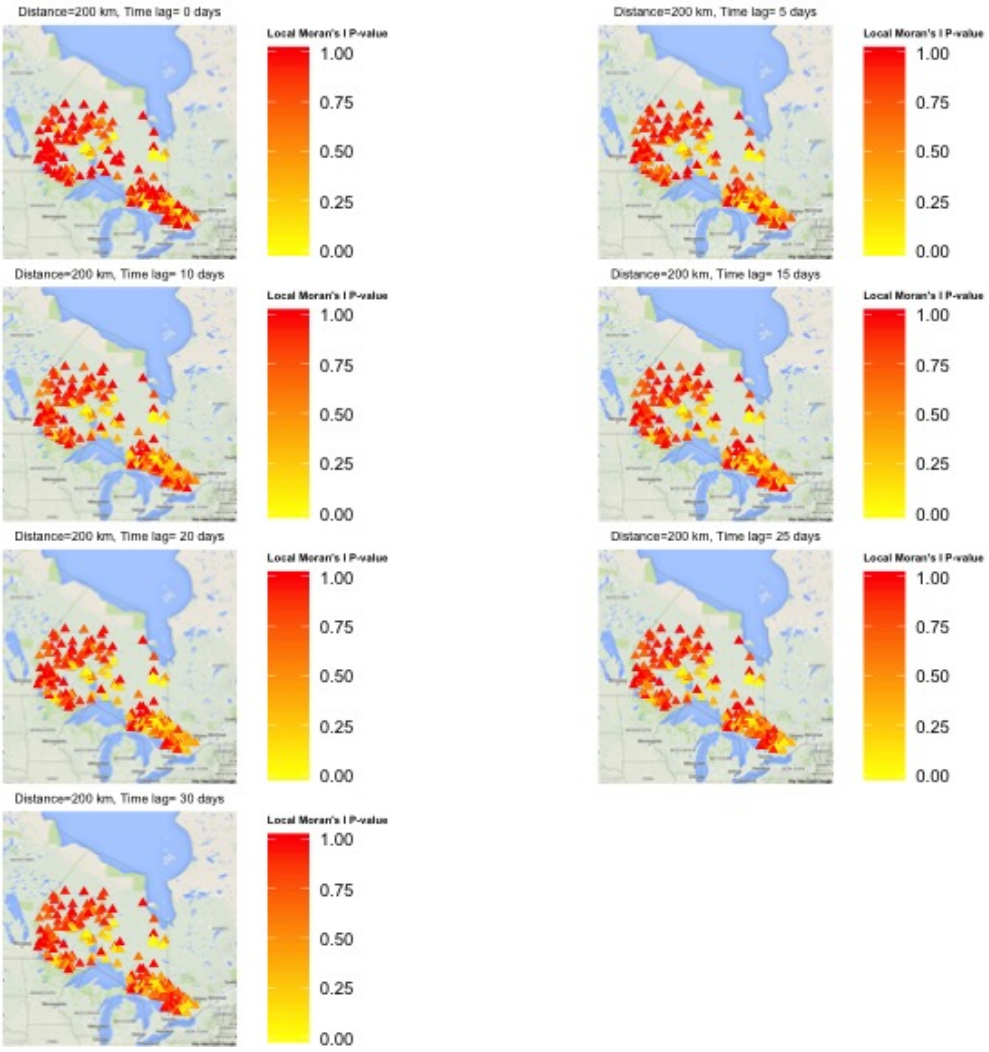


Figure 4.19: Local Moran's I-based Map for Residuals in Model (4.2a)

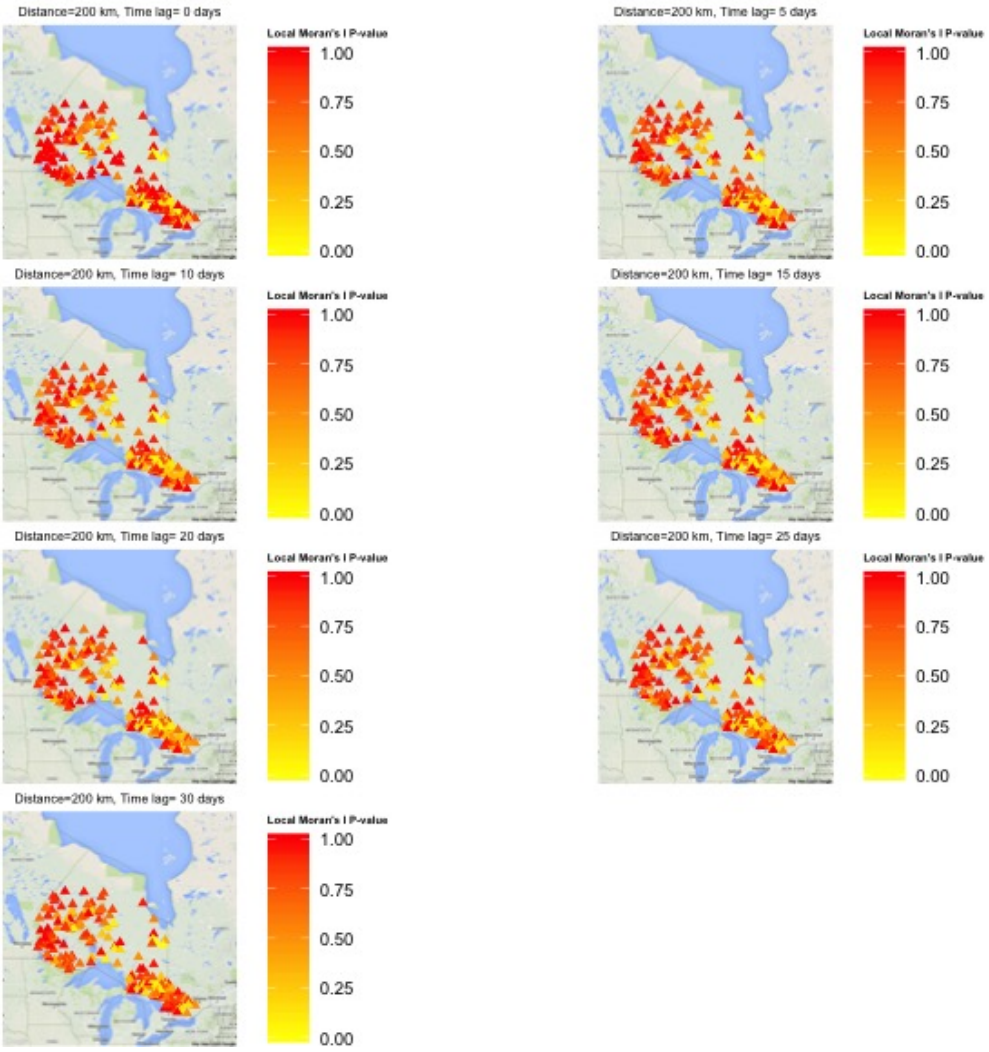


Figure 4.20: Local Moran's I-based Map for Residuals in Model (4.2b)

The residual analyses show that the normal and independence assumptions for partially linear models (4.2a) and (4.2b) are reasonable. Recalling the residual analysis results of ordinary linear regression models in Chapter 3, the current models have reduced the spatio-temporal correlation in data and provide improvements. However, the spatial autocorrelation are not removed completely with current models and the constant variance assumption is still not valid. We will try another set of partially linear models (4.3a) and (4.3b) to provide a better fit.

4.3 Analysis of Partially Linear Regression with Model 4.3(a) and 4.3(b)

In order to provide a better fit accounting for the underlying correlation of the data, we consider another partially linear regression. The estimation procedures and results as well as residual analyses with the models are presented.

4.3.1 Estimation Procedures

Based on the estimation procedures of partially linear Model (4.2a) and (4.2b), we extend local constant and local linear methods to estimate the 2-variate nonparametric function $g_{h(t)}(\mathbf{s}_i)$ and apply LSE to estimate the parameter β.

Similar to the estimation procedures in section 4.2, with Model (4.3a), we minimize the weighted sum squares in each month m:

$$\sum_{i:h(t_i)=m} (y_i - \beta_{h(t_i)} z_i - g_{h(t_i)}(\mathbf{s}))^2 K(\frac{\mathbf{s}-\mathbf{s}_i}{d}) \tag{4.9}$$

where $m = 1, 2, 3, 4$ representing month May, June, July, August and September of fire starting time in the dataset.

With Model (4.3b), we minimize the objective function below instead,

$$\sum_{m=1}^{4} \sum_{i:h(t_i)=m} (y_i - \beta_{h(t_i)} z_i - g_{h(t_i)}(\mathbf{s}))^2 K(\frac{\mathbf{s}-\mathbf{s}_i}{d}) \tag{4.10}$$

We consider $K(\cdot)$ as a 2-variate Kernel function of bandwidth d. Here, we use the Tricube function

$$K(\mathbf{u}) = \begin{cases} (1-|\mathbf{u}|^3)^3 & \text{if } |\mathbf{u}| < 1; \\ 0 & \text{if } |\mathbf{u}| > 1. \end{cases}$$

where $\mathbf{u} = (u_1, u_2)'$ and $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2}$. The bandwidth d controls the size of the local neighbors in both *Longitude* and *Latitude* direction.

The local constant and local linear estimation procedures for Model (4.3a) and (4.3b) are similar with the presented procedures in section 4.2, except that we use bivariate

Kernel smoothers. The estimators we obtained below are based on the procedures in section 4.2.

Local Constant Estimation

Accordingly, with Model (4.3a) for any fixed location $\mathbf{s} = (s_1, s_2) \in R^2$, we approximate $g_{h(t_i)}(\mathbf{s}_i)$ by $g_{h(t_i)}(\mathbf{s})$. By minimizing the weighted sum squares in (4.9), the estimator of function $g_m(\mathbf{s}_i)$ and β_m are:

$$\begin{aligned} \hat{g}_m(\mathbf{s}_i) &= (\mathbf{1}'\mathbf{K}_m\mathbf{1})^{-1}\mathbf{1}'\mathbf{K}_m(\mathbf{Y}_m - \beta_m\mathbf{Z}_m) \\ \hat{\beta}_m &= (\tilde{\mathbf{Z}}_m'\tilde{\mathbf{Z}}_m)^{-1}\tilde{\mathbf{Z}}_m'\tilde{\mathbf{Y}}_m \end{aligned} \tag{4.11}$$

where $\mathbf{Y}_m$ is a sub-vector of $\mathbf{Y} = [y_1, y_2, \cdots, y_{259}]'$, $\mathbf{Z}_m$ is a sub-vector of $\mathbf{Z} = [z_1, z_2, \cdots, z_{259}]'$ with y_i and z_i from stratum $h(t_i) = m$, and $\tilde{\mathbf{Y}}_m = (\mathbf{I} - \mathbf{S}_m)\mathbf{Y}_m$, $\tilde{\mathbf{Z}}_m = (\mathbf{I} - \mathbf{S}_m)\mathbf{Z}_m$. The smoothing matrix in each month is $\mathbf{S}_m = (\mathbf{1}'\mathbf{K}_m\mathbf{1})^{-1}\mathbf{1}'\mathbf{K}_m$, where $\mathbf{K}_m$ is a diagonal matrix $Diag[K(\frac{\mathbf{s}-\mathbf{s}_i}{d})$: i with $h(t_i) = m]$.

With Model (4.3b), the estimators of function $g_m(\mathbf{s}_i)$ and parameter β are obtained as:

$$\begin{aligned} \hat{g}_m(\mathbf{s}_i) &= (\mathbf{1}'\mathbf{K}_m\mathbf{1})^{-1}\mathbf{1}'\mathbf{K}_m(\mathbf{Y}_m - \beta\mathbf{Z}_m) \\ \hat{\beta} &= (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{Y}} \end{aligned} \tag{4.12}$$

where $\tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S})\mathbf{Y}$, $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S})\mathbf{Z}$, and the smoothing matrix $\mathbf{S}$ is a diagonal matrix $Diag[\mathbf{S}_m : m = 1, 2, 3, 4]$.

Then the fitted values $\hat{\mathbf{Y}}$ with Model (4.3a) are:

$$\hat{\mathbf{Y}}_m = \{\mathbf{S}_m + \tilde{\mathbf{Z}}_m(\tilde{\mathbf{Z}}_m'\tilde{\mathbf{Z}}_m)^{-1}\tilde{\mathbf{Z}}_m'(\mathbf{I} - \mathbf{S}_m)\}\mathbf{Y}_m,\ m = 1, 2, 3, 4, \tag{4.13}$$

and with Model (4.3b),

$$\hat{\mathbf{Y}} = \{\mathbf{S} + \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'(\mathbf{I} - \mathbf{S})\}\mathbf{Y}. \tag{4.14}$$

Local Linear Estimation

For local linear estimator with any fixed $\mathbf{s} = (s_1, s_2) \in R^2$, we approximate $g_{h(t_i)}(\mathbf{s}_i)$ by $g_{h(t_i)}(\mathbf{s}) + \dot{\mathbf{g}}_{h(t_i)}(\mathbf{s})'(\mathbf{s}_i - \mathbf{s})$ where $\dot{\mathbf{g}}_{h(t_i)}(\mathbf{s}) = \begin{bmatrix} \frac{\partial g_{h(t_i)}(\mathbf{s})}{\partial s_1} \\ \frac{\partial g_{h(t_i)}(\mathbf{s})}{\partial s_2} \end{bmatrix}$, and $(\mathbf{s}_i - \mathbf{s}) = (s_{i1} - s_1, s_{i2} - s_2)'$. Then following the same settings with local constant estimation procedure, with Model (4.3a), the local linear estimator is the first row component of

$$\left(\hat{g}_m(\mathbf{s}), \hat{\dot{g}}_m(\mathbf{s})\right)' = (\boldsymbol{X}_m(\mathbf{s})' \mathbf{K}_m \boldsymbol{X}_m(\mathbf{s}))^{-1} \boldsymbol{X}_m(\mathbf{s})' \mathbf{K}_m (\mathbf{Y}_m - \beta_m \mathbf{Z}_m),$$

where $\boldsymbol{X_m(s)} = (\mathbf{1}, \triangle^*)$, and $\triangle^*$ is a matrix with ith row equal to $(s_{i1} - s_1, s_{i2} - s_2)$. The estimator of β_m is $\hat{\beta}_m = (\tilde{\mathbf{Z}}_m' \tilde{\mathbf{Z}}_m)^{-1} \tilde{\mathbf{Z}}_m' \tilde{\mathbf{Y}}_m$, and $\tilde{\mathbf{Y}}_m$, $\tilde{\mathbf{Z}}_m$ are defined in (4.11).

Similarly, with Model (4.3b), the estimator of $g_m(\mathbf{s})$ is the first row component of

$$\left(\hat{g}_m(\mathbf{s}), \hat{\dot{g}}_m(\mathbf{s})\right)' = (\boldsymbol{X}_m(\mathbf{s})' \mathbf{K}_m \boldsymbol{X}_m(\mathbf{s}))^{-1} \boldsymbol{X}_m(\mathbf{s})' \mathbf{K}_m (\mathbf{Y}_m - \beta \mathbf{Z}_m).$$

The estimator of β is $\hat{\beta} = (\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}' \tilde{\mathbf{Y}}$, where $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{Z}}$ are defined in (4.12).

The fitted values $\hat{\mathbf{Y}}$ using local linear estimator with Model (4.3a) and (4.3b) are in the same form with the ones using local constant estimator in equation (4.13) and (4.14).

Bandwidth Selection and Estimation of Variance

The GCV criterion presented in section 4.2.2 and the estimation method for variance in section 4.2.3 can be applied to Model 4.3(a) and (b).

4.3.2 Analysis Results

We apply GCV criterion to select the optimal bandwidth for Model (4.3a) and (4.3b) by using local constant estimator and local linear estimator.

Table 4.4 lists the optimal bandwidths chosen by GCV criterion for Model (4.3a) and (4.3b). The corresponding plots of GCV values versus bandwidths are produced in Figure 4.21, 4.22, 4.23 and 4.24. The bandwidths selected for local constant estimators remain the same through the whole fire season with both the Model (4.3a) and (4.3b). Using local

linear estimators, the selected bandwidths are consistent with both models for June, July, August and September but are different in May. The bandwidths chosen in May are much larger than other months.

	Model (4.3a)				Model (4.3b)			
	$g_{\textbf{May}}(\mathbf{s})$	$g_{\textbf{June}}(\mathbf{s})$	$g_{\textbf{July}}(\mathbf{s})$	$g_{\textbf{AugSep}}(\mathbf{s})$	$g_{\textbf{May}}(\mathbf{s})$	$g_{\textbf{June}}(\mathbf{s})$	$g_{\textbf{July}}(\mathbf{s})$	$g_{\textbf{AugSep}}(\mathbf{s})$
Local Constant	5	5	5	5	5	5	5	5
Local Linear	30	5	7	5	17	5	7	5

Table 4.4: Bandwidth Selection of Model (4.3a) and (4.3b)

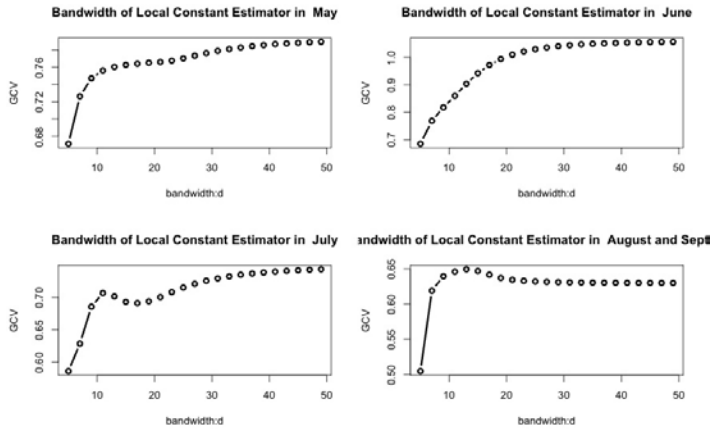


Figure 4.21: Bandwidth Selection for Local Constant Estimator of $g(\mathbf{s})$ with Model (4.3a)

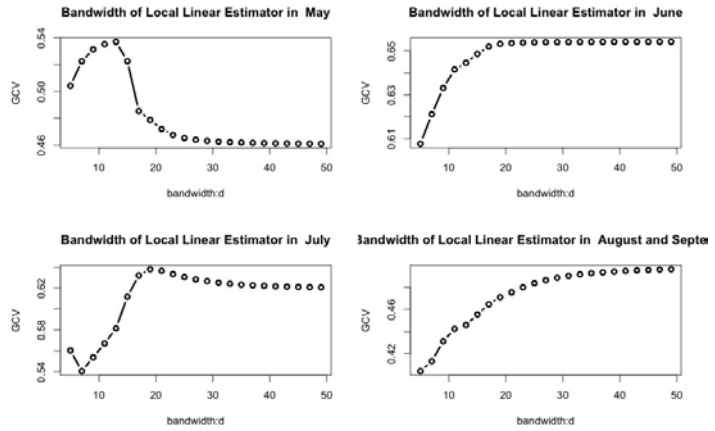


Figure 4.22: Bandwidth Selection for Local Linear Estimator of $g(\mathbf{s})$ with Model (4.3a)

Using the optimal bandwidths, we then obtain the analysis results including estimates of β and the smoothed function values of $g_m(\mathbf{s})$ in each month.

Table 4.5 presents the estimates of β_m and β in Model (4.3a) and (4.3b). The estimated

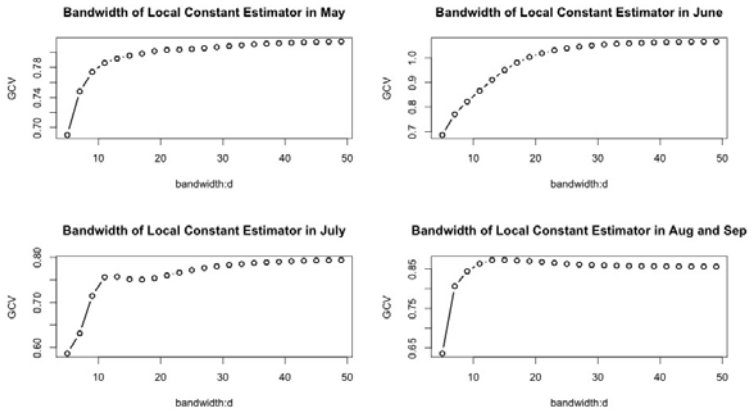


Figure 4.23: Bandwidth Selection for Local Constant Estimator of $g(\mathbf{s})$ with Model (4.3b)

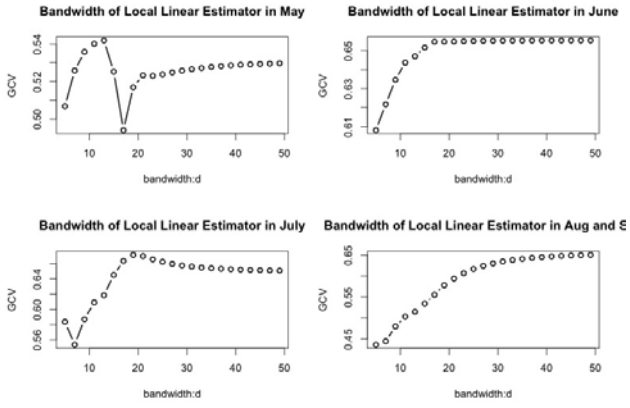


Figure 4.24: Bandwidth Selection for Local Linear Estimator of $g(\mathbf{s})$ with Model (4.3b)

standard errors are given below the estimates with the p-values of Wald-test on the significance. The statistically significant predictors' effects are marked as bold. The results show that estimates of β_m are different through the fire season with Model (4.3a). The β_{AugSep} is statistically significant by using local constant estimator to smooth the spatial effect while β_{May} is significant by using local linear estimator. This difference can be explained by looking at Figure 4.25 and Figure 4.26, which produce the maps of values of smooth function $g_m(\mathbf{s})$ in each month using local constant and local linear estimators. The values of smoothed function $g_m(\mathbf{s})$ are represented by colors. From the map of the local constant estimator in Figure 4.25, we can see the color of points are almost the same in August and September, which means the variation of fire duration are mostly explained by the parametric part in the model, i.e. by variable *FWI* in August and September. In contrast, Figure 4.26 displays the smoothed values by local linear estimators. Comparing with Figure 4.25, it exhibits a more clear pattern of fire duration varied from locations for each month. In addition, the standard error and mean squared error(MSE) are smaller by local linear estimators than local constant estimators.

The estimates of *FWI*'s effect are different by local constant and local linear smoothers with Model (4.3b). The difference is in agreement with the results of Model 4.3(a) because the obtained estimates β can be viewed as a weighted average of β_{May}, β_{June}, β_{July} and β_{AugSep} in Model (4.3a). The values of $g_m(\mathbf{s})$ by local constant and local linear estimators are presented in Figure 4.27 and 4.28. The exhibited pattern by local linear estimator is similar to the one with Model (4.3a).

Comparing the results of Model 4.3(a), (b) with Model 4.2(a) and (b), we find *FWI*'s effects on fire duration become less significant when we smooth the spatial effects for each month, indicating that the influence of fire's locations can not be omitted.

	Model (4.3a)								Model (4.3b)	
	Local Constant				Local Linear				Local Constant	Local Linear
	$\beta_{\textbf{May}}$	$\beta_{\textbf{June}}$	$\beta_{\textbf{July}}$	$\beta_{\textbf{AugSep}}$	$\beta_{\textbf{May}}$	$\beta_{\textbf{June}}$	$\beta_{\textbf{July}}$	$\beta_{\textbf{AugSep}}$	$\beta_{\textbf{local constant}}$	$\beta_{\textbf{local linear}}$
Estimates	-0.653	-0.160	0.112	**1.69**	**-1.49**	-0.375	0.209	0.668	0.00146	-0.274
Std.Error	0.689	0.318	0.834	0.599	0.578	0.318	0.741	0.527	0.252	0.255
P−value	0.343	0.614	0.447	0.004	0.011	0.226	0.616	0.205	0.984	0.283
MSE	0.622				0.502				0.621	0.501

Table 4.5: Estimates of Regression Coefficients for Model (4.3a) and (4.3b)

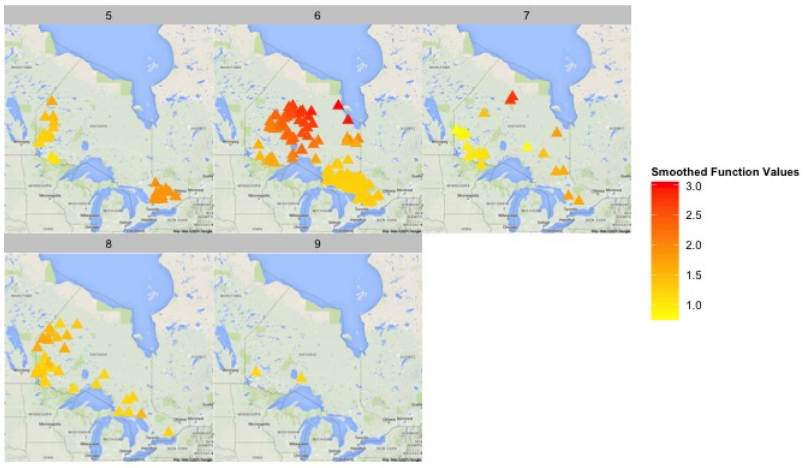


Figure 4.25: Smoothed Values for Model (4.3a) by Local Constant Estimator

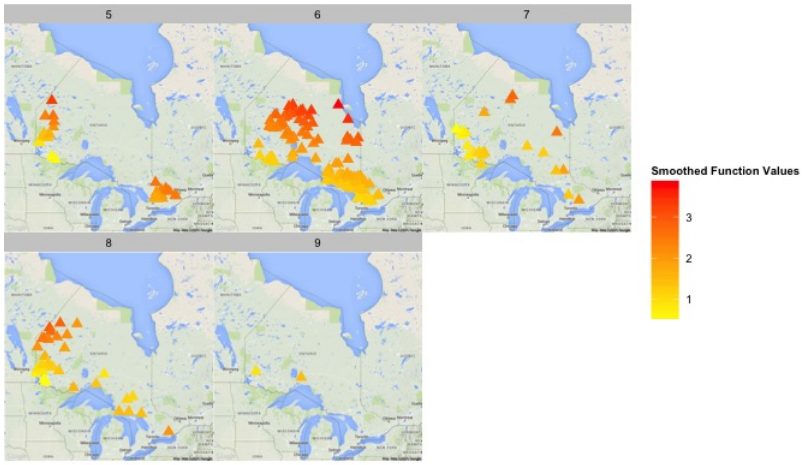


Figure 4.26: Smoothed Values for Model (4.3a) by Local Linear Estimator

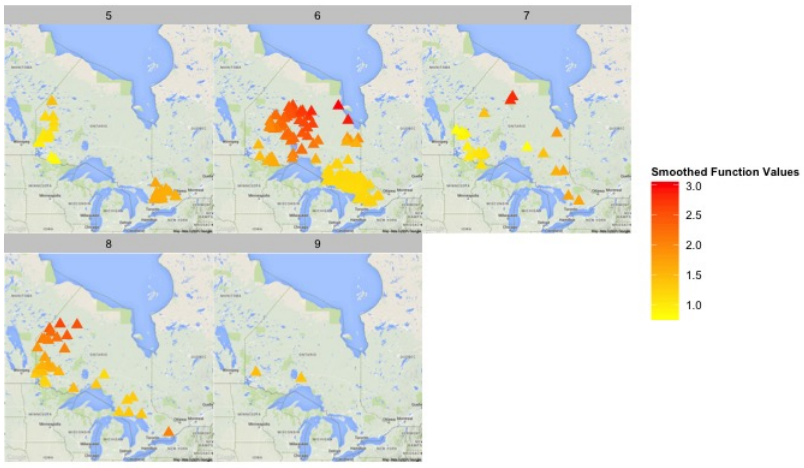


Figure 4.27: Smoothed Values for Model (4.3b) by Local Constant Estimator

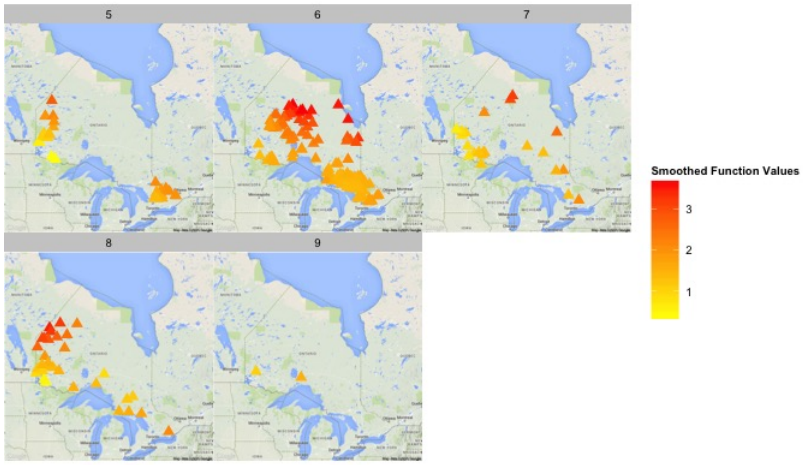


Figure 4.28: Smoothed Values for Model (4.3b) by Local Linear Estimator

4.3.3 Residual Analysis

We use the same procedure of residual analyses in section 4.2 for Model (4.2a) and (4.2b) to check the normality, constant variance and independence assumption of the model. The residual plots of Model (4.3a) and (4.3b) by local linear estimators are presented in the following.

Normal Q-Q plots are displayed in Figure 4.29 and points are distributed closely to the Normal Q-Q lines, therefore the normal assumption is reasonable. The scatterplot of residuals versus predicted value, *FWI* and fire's start date are plotted in Figures 4.30 and 4.31 to check the mean and variance of residuals. The red lines obtained from **lowess** in R are close to 0 so the mean of residuals can be approximated to 0. Figure 4.32 displays the residual maps and it is hardly to see the similarity in values for neighboring spots. So the residual map doesn't reveal any obvious pattern of residual distribution.

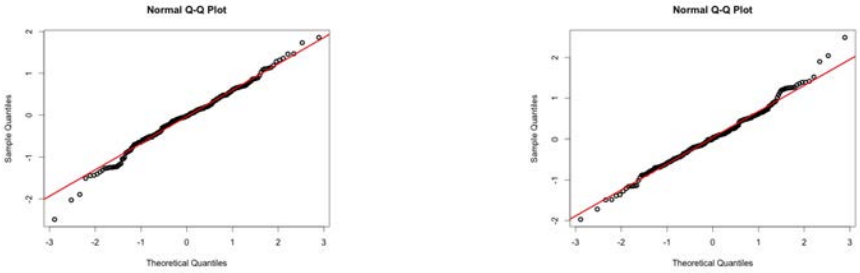


Figure 4.29: Normal QQ Plot of Residuals in Model (4.3a) and Model (4.3b) by Local Linear Estimators

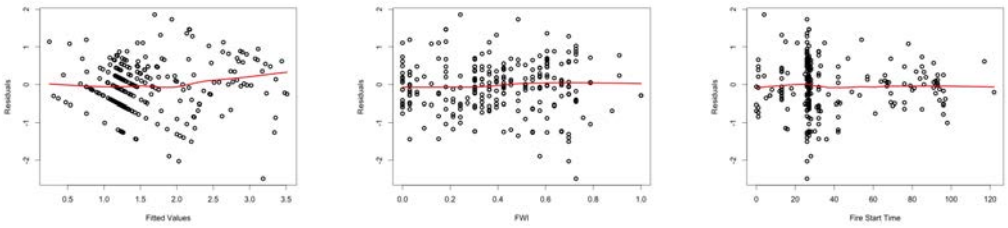


Figure 4.30: Residual Plots of Model (4.3a) by Local Linear Estimator

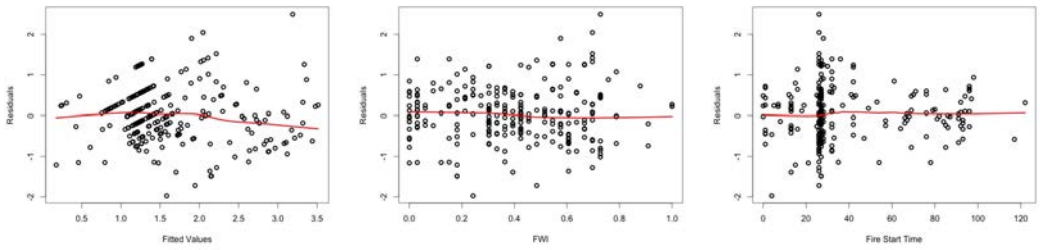
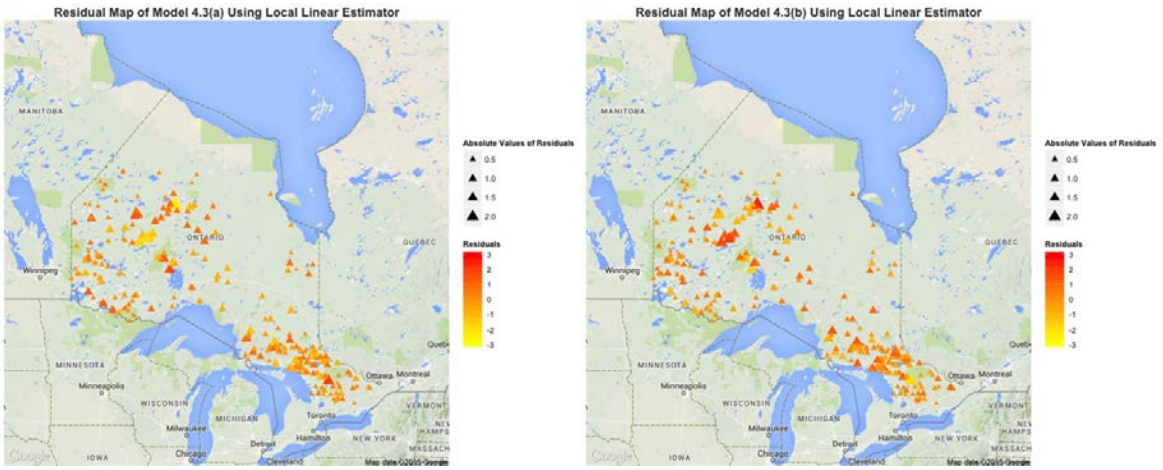


Figure 4.31: Residual Plots of Model (4.3b) by Local Linear Estimator



(a) Residual Map for Model (4.3a) (b) Residual Map for Model (4.3b)

Figure 4.32: Residual Maps

We check the independence assumption of residuals by investigating whether correlation is present among residuals in time and space. Empirical semivariogram are produced to visualize the possible correlation structure. The semivariograms in Figure 4.33 show that the semivariances keep increasing up to 300 km and fluctuates around MSE with both models. The rise and fall of semivariance after 300 km might be due to the violation of constant variance assumption of residuals. So we proceed to use Moran's *I* to further assess the spatial autocorrelation of residuals.

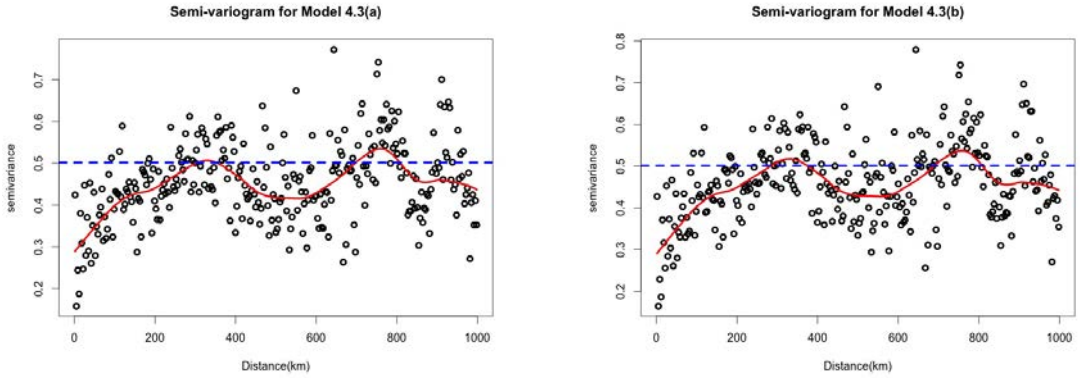


Figure 4.33: Semivariogram of Residuals of Model (4.3a) and (4.3b) by Local Linear Estimator

By defining that fires within 300 km as neighbors, the results of global Moran's *I* test are listed in Table 4.6. The test results indicate that there is no statistically significant spatial autocorrelation among the residuals with both the models. The spatial correlograms are displayed in Figure 4.34. Both spatial correlograms show that the 95% acceptance region includes the expectation value of Moran's *I* regardless of specification for distance. This indicates that there is no significant spatial autocorrelation in residuals. Additionally, local Moran's *I* maps are produced to identify the points with high spatial autocorrelation. The yellow points are those with $p\_value < 0.05$, indicating there is a significant spatial autocorrelation around that point. Comparing with local Moran's I map of Model (4.2a) and (4.2b) in Figure 4.15, the number of points with significant local spatial autocorrelation is lessened with Model (4.3a) and (4.3b).

Following the settings in section 4.2, we then produce the spatio-temporal correlograms to check spatio-temporal correlation in residuals. From the perspective plots in Figures

	Observed Moran's I	Expectation	Variance	p_value
Model (4.3a) Local Linear	-0.013	-0.0038	0.000043	0.14
Model (4.3b) Local Linear	-0.014	-0.0038	0.000043	0.12

Table 4.6: Moran's I test on Model 4.3(a),(b)

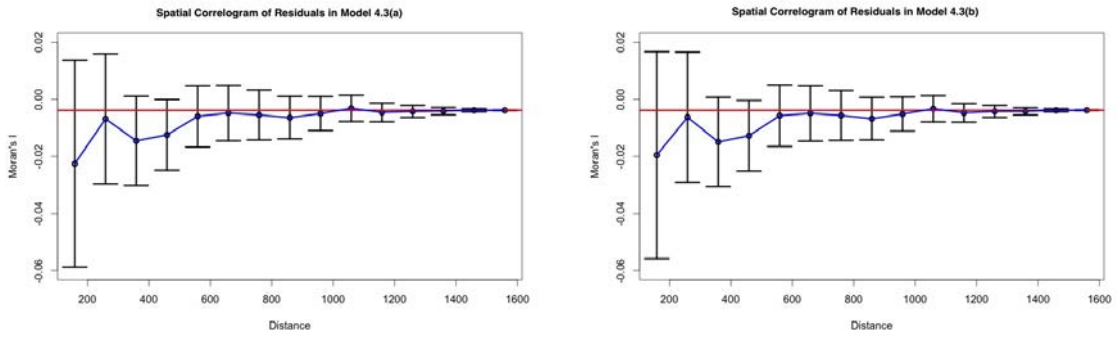
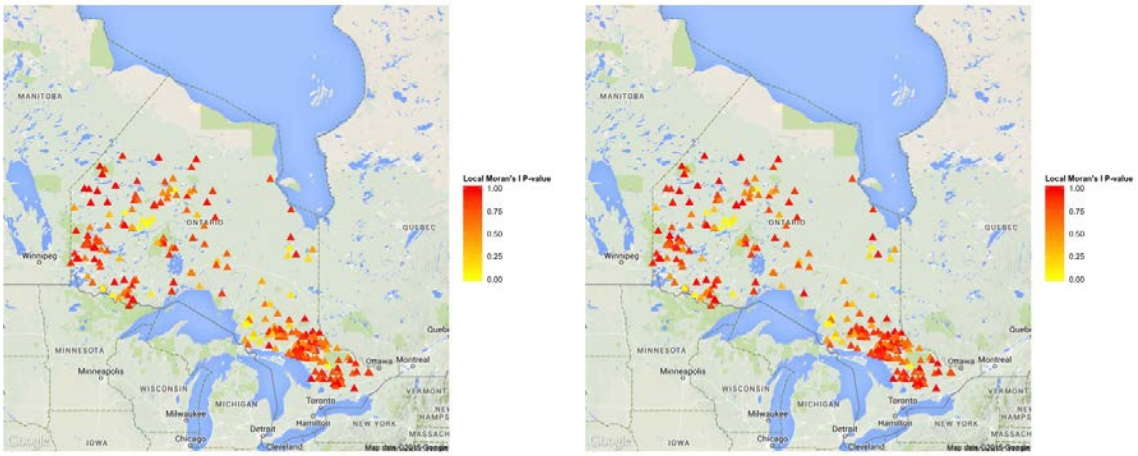


Figure 4.34: Spatial Correlograms for Model (4.3a) and (4.3b) by Local Linear Estimator



(a) Local Moran's I for Model (4.3a)

(b) Local Moran's I for Model (4.3b)

Figure 4.35: Local Moran's I Map of Residuals with Model (4.3a) and (4.3b)

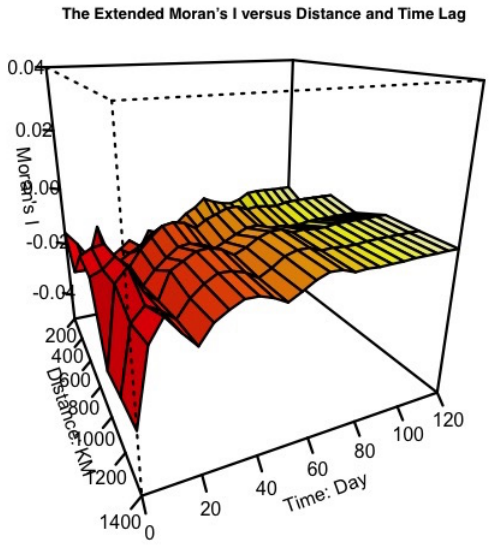
4.36a and 4.36b, the values of extended Moran's *I* change greatly from time lag 0 to the next and remain unchanged spatially and temporally afterwards. The correlograms in Figures 4.37 and 4.38 show that the values of extended Moran's *I* are close to the expectation values regardless of time at all distance, indicating that there is no significant spatio-temporal correlation in residuals.

The p_values testing on local spatio-temporal correlation by using extended local Moran's *I* are plotted in Figures 4.39 and 4.40. We find that there are fewer yellow points compared with the maps for Model 4.2 (a) and (b). This suggests that Model 4.3 (a) and (b) behave better in removing the spatio-temporal correlation.

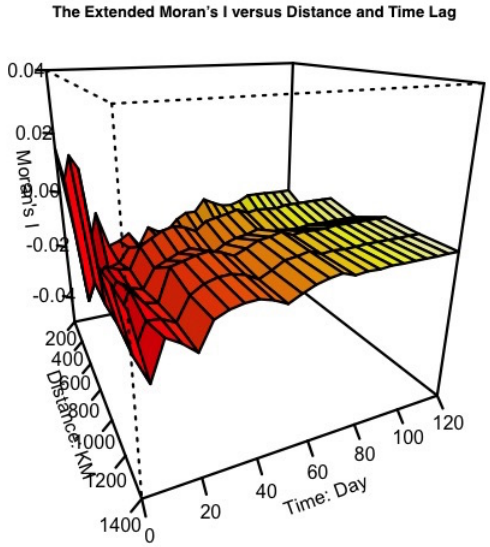
4.4 Summary

To tackle the problems arising from preliminary analysis and filter out the correlation of underlying data, this chapter proposed 2 sets of partially linear regression models to analyze the forest fire data. The analysis with the first model used the univariate Kernel smoothing methods to handle fire's starting time effects on fire duration. This model had great improvements on removing the correlation of underlying data and obtained robust and significant estimates in the parametric components. The analysis with the second model extended univariate Kernel smoothing methods to 2-variate to estimate the spatial effects of fire duration. The second model exhibited a more clear pattern of fire duration variation spatially and the residuals are independent in time and space with the second model.

We considered fixed and stratified parametric components in each set of models to examine if *FWI*'s effects vary from different fire management zone or different months. The different estimates of the coefficients of *FWI* in the regression analysis suggests that *FWI*'s effects on fire duration change from different fire management zone and different months.



(a) Perspective Plot of Extended Moran's I for Residuals in Model (4.3a)



(b) Perspective Plot of Extended Moran's I for Residuals in Model (4.3b)

Figure 4.36: Perspective Plot of Extended Moran's I for Residuals with Model (4.3a) and (4.3b)

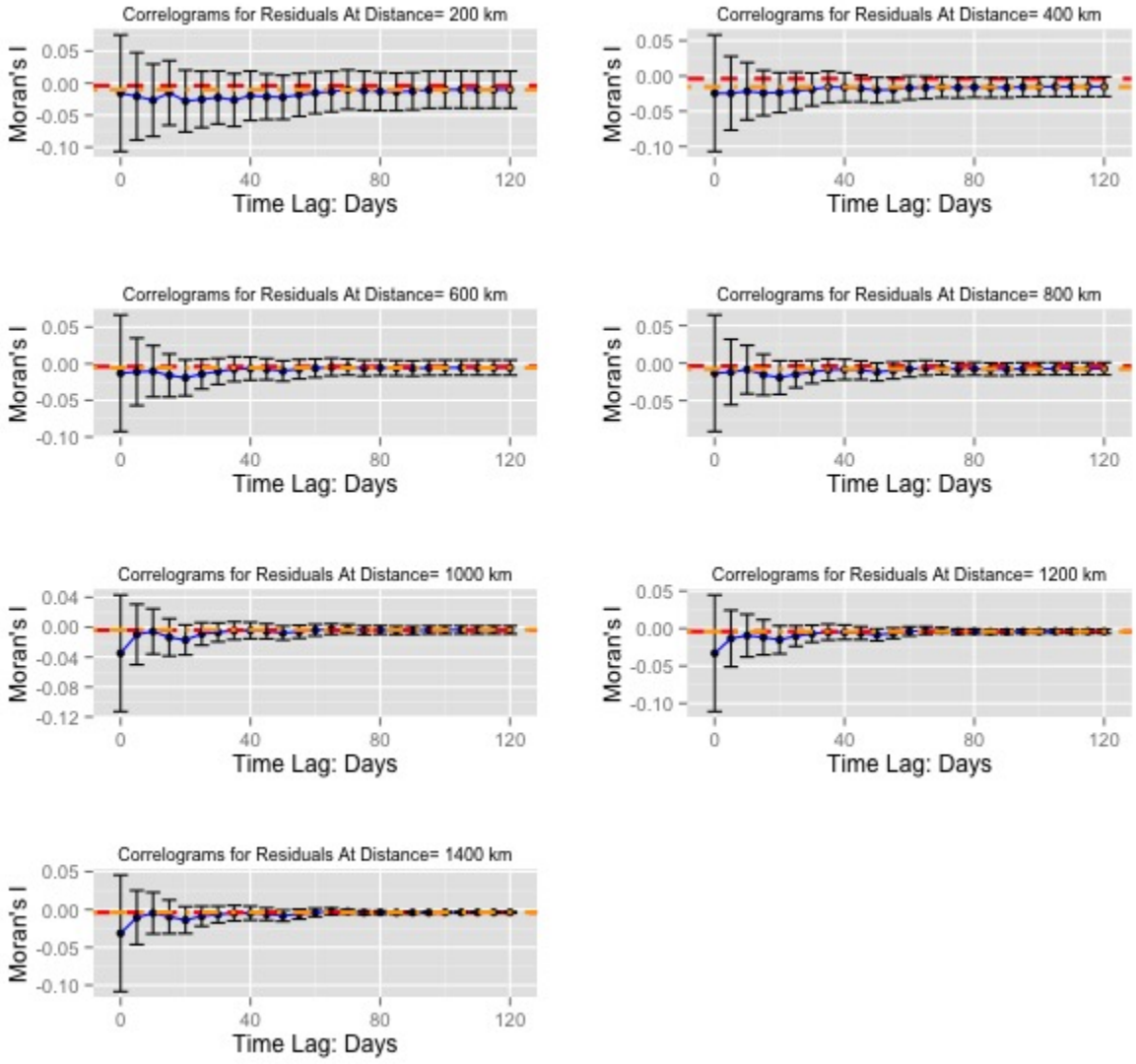


Figure 4.37: Extended Moran's I-based Spatio-Temporal Correlogram for Residuals in Model (4.3a)

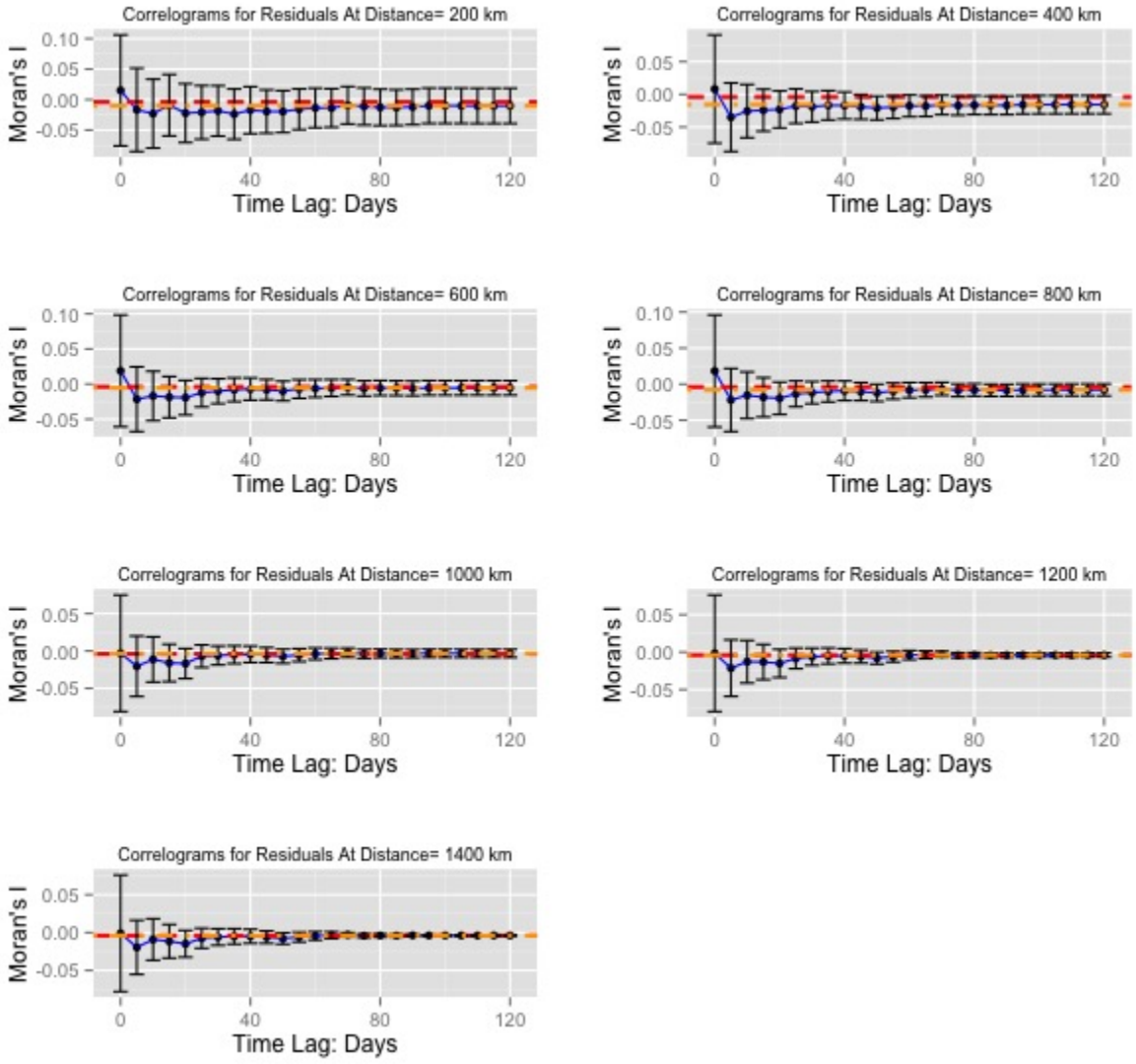


Figure 4.38: Extended Moran's I-based Spatio-Temporal Correlogram for Residuals in Model (4.3b)

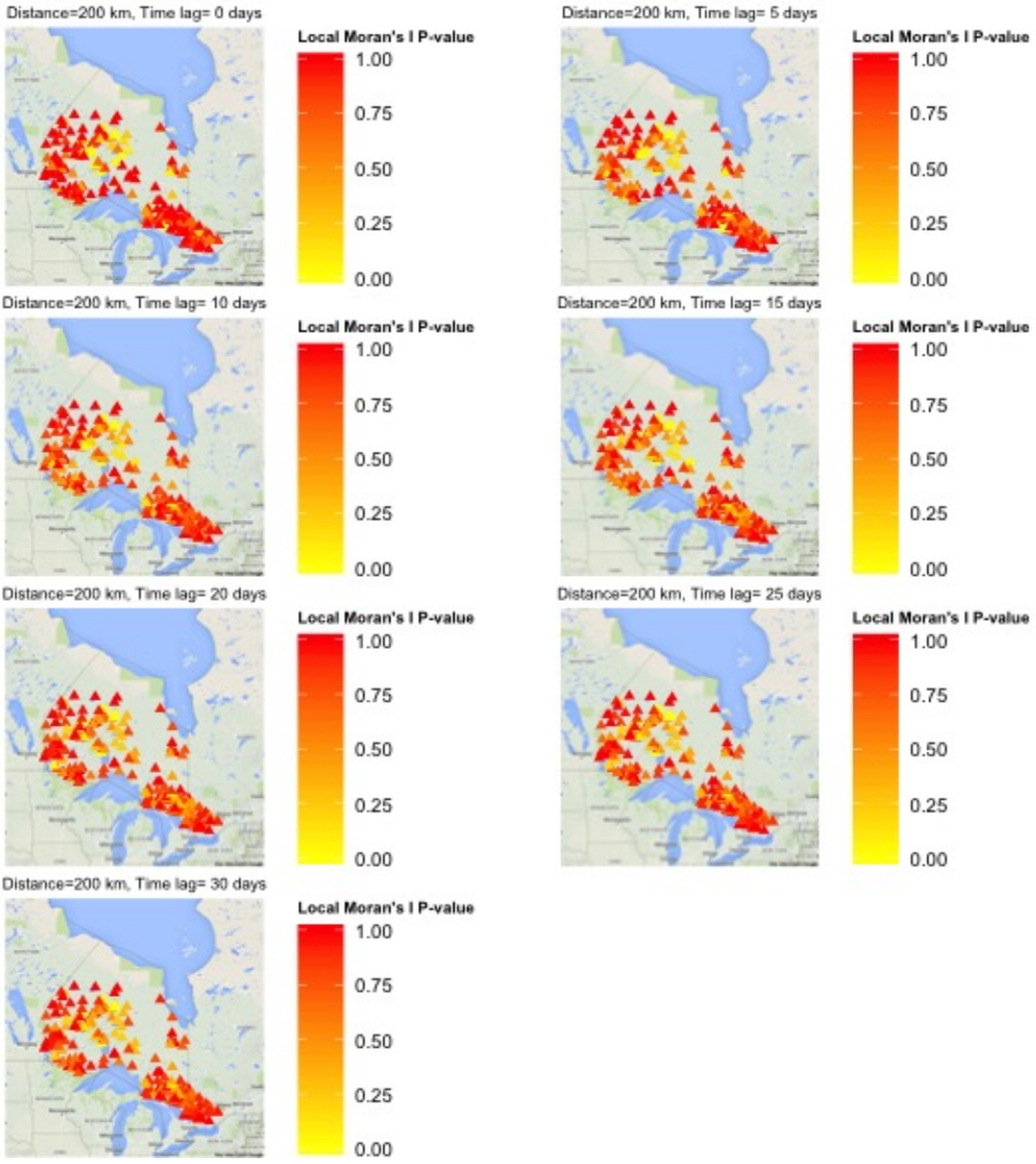


Figure 4.39: Local Moran's I-based Map for Residuals in Model 4.3(a)

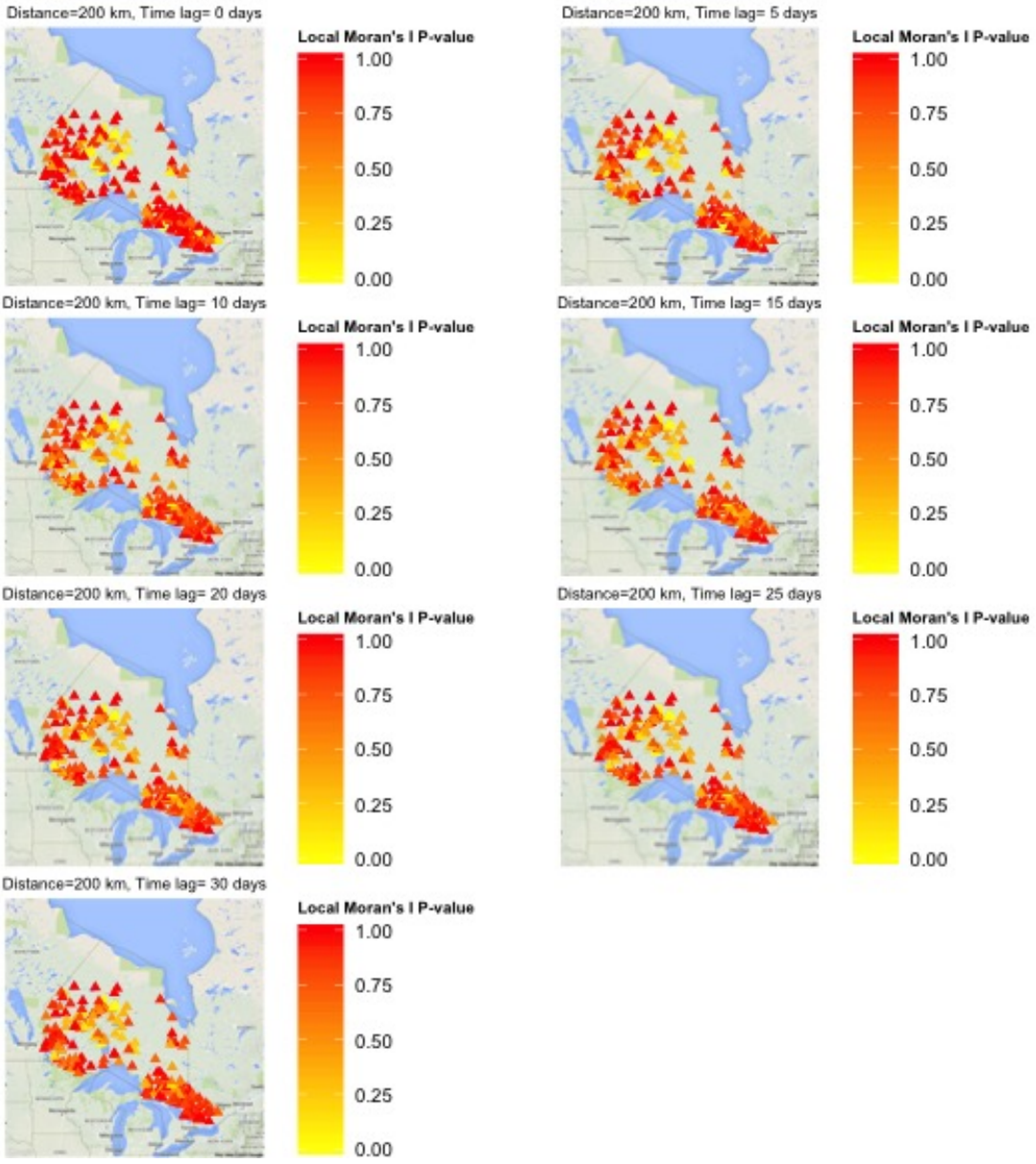


Figure 4.40: Local Moran's I-based Map for Residuals in Model 4.3(b)

Chapter 5

Final Remarks

5.1 Summary

This project analyzes the forest fire dataset from Ontario and studies the association between fire duration and environmental variables, adjusting for spatio-temporal correlation in fire duration. By extending the Moran's *I* to account for the temporal correlation, we examine spatio-temporal correlation in the dataset and the residuals of linear regression models. The ordinary linear regression is the preliminary analysis motivating partially linear regression analysis to filter out the embedded spatio-temporal correlation in the data.

With the partially linear regression models, we apply Kernel smoothing methods to handle the nonparametric components. Residual analyses are performed to check goodness of fit and the spatio-temporal correlation. Based on the results in Chapter 4, partially linear regression models can capture the spatio-temporal effects and the model assumptions are appropriate. Both the ordinary linear regression and partially linear regression indicate that fire duration is significantly associated with fire's location, fire's starting time, and the fuel and weather index FWI. The association varies in time and space.

5.2 Future Work

There are a few issues which need to be addressed. We listed them in the following:

- As discussed in Chapter 3, the significant results from Moran's *I* test indicate either

the presence of correlation or the nonstationarity of the data. There are 2 approaches to address this issue. The partially linear regression models used in this project reduce the nonstationarity and the correlation of the residuals by further modelling the systematic component. The alternative approach could incorporate the correlation structures in the models. The generalized estimating equation (GEE) approach developed by Liang and Zeger (1986) can be an alternative approach.

- We detected the correlation in the data by extending Moran's *I* to consider neighbors in space and time. An alternative approach to extend Moran's *I* by integrating a spatio-temporal weight matrix proposed by Dub, J.& Legros, D. (2013) is worth being studied.

- The ecological variable we used in this study is *FWI* related to i^{th} fire on fire's start date. It will be of interest to consider its effect on fire duration is time-variant if the records of *FWI* on each day could be provided.

- The conclusions in this project are based on the Ontario dataset in a specific time period. Thus, these conclusions need to be confirmed by investigating forest fire data of other years to make it a more comprehensive analysis.

Bibliography

Kerry Anderson. A model to predict lightning-caused fire occurrences. *International journal of wildland fire*, 11(4):163–172, 2002.

Luc Anselin. Local indicators of spatial association—lisa. *Geographical analysis*, 27(2): 93–115, 1995.

Andrew David Cliff and J Keith Ord. *Spatial processes: models & applications*, volume 44. Pion London, 1981.

Peter Craven and Grace Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1978. 43

N Cressie. Statistics for spatial data: Wiley series in probability and statistics. 1993.

Noel Cressie and Hsin-Cheng Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339, 1999.

William J De Groot et al. Interpreting the canadian forest fire weather index (fwi) system. In *Proc. of the Fourth Central Region Fire Weather Committee Scientific and Technical Seminar*, 1998. xi, 12

Peter Diggle and Paulo Justiniano Ribeiro. *Model-based geostatistics*. Springer, 2007.

Jean Dubé and Dièg o Legros. A spatio-temporal measure of spatial dependence: An example using real estate data*. *Papers in Regional Science*, 92(1):19–30, 2013.

Carsten F Dormann, Jana M McPherson, Miguel B Araújo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard G Davies, Alexandre Hirzel, Walter Jetz, W Daniel Kissling, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5):609–628, 2007.

Mario Francisco-Fernandez and Jean D Opsomer. Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics*, 33(2):279–295, 2005.

Nicholas J Gralewicz, Trisalyn A Nelson, and Michael A Wulder. Spatial and temporal patterns of wildfire ignitions in canada from 1980 to 2006. *International Journal of Wildland Fire*, 21(3):230–242, 2012.

Bradley E Huitema and Joseph W McKean. Autocorrelation estimation and inference with small samples. *Psychological Bulletin*, 110(2):291, 1991.

P Kourtz, B Todd, et al. Predicting the daily occurrence of lightning-caused forest fires. In *Central Region Fire Weather Committee Scientific and Technical Seminar*, page 37, 1992.

MA Krawchuk, SG Cumming, MD Flannigan, and RW Wein. Biotic and abiotic regulation of lightning fire initiation in the mixedwood boreal forest. *Ecology*, 87(2):458–468, 2006.

David L Martell and Hua Sun. The impact of fire suppression, vegetation, and weather on the area burned by lightning-caused forest fires in ontario. *Canadian Journal of Forest Research*, 38(6):1547–1563, 2008.

DF Merrill, Martin E Alexander, et al. Glossary of fire management terms. 1987. 11

Patrick AP Moran. Notes on continuous stochastic phenomena. *Biometrika*, pages 17–23, 1950.

Natural Resource of Canada. Canadian forest fire weather index (fwi) system. URL http://cwfis.cfs.nrcan.gc.ca/background/summary/fwi. xi, 12

Justin J Podur. *Spatial and temporal patterns of forest fire activity in Canada*. PhD thesis, University of Toronto, 2001.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

Paul Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 413–436, 1988. 39

CE Van Wagner et al. *Development and structure of the Canadian forest fire weather index system*, volume 35. 1987. xi, 13

BM Wotton et al. A lightning fire prediction system. frontline express 60. 2012.

Scott L Zeger and Kung-Yee Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pages 121–130, 1986.