UCLA
UCLA Electronic Theses and Dissertations

Title

California Wildfire Spread Prediction using FARSITE and the Comparison with the Actual Wildfire Maps using Statistical Methods

Permalink

https://escholarship.org/uc/item/8nz6p4hc

Author

Hao, Yujia

Publication Date

2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

California Wildfire Spread Prediction using FARSITE and the Comparison with the Actual Wildfire Maps using Statistical Methods

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Applied Statistics

by

Yujia Hao

2018

© Copyright by
Yujia Hao
2018

ABSTRACT OF THE THESIS

California Wildfire Spread Prediction using FARSITE
and the Comparison with
the Actual Wildfire Maps using Statistical Methods

by

Yujia Hao
Master of Applied Statistics
University of California, Los Angeles, 2018
Professor Frederic R. Paik Schoenberg, Chair

The unpredictability of wildfires has always been a major problem that brings a vast amount of devastation to the environment and human lives every year. This project uses the program R, QGIS 3 and FARSITE version 3. I implemented the simulations in FARSITE on 10 separate wildfire datasets in California. The datasets include canopy, fuels, weather, perimeters and geographic setting. Map projection was transformed from WGS84 to Albers. Predictions of wildfire maps are generated from FARSITE models in terms of vector data, raster data, and shapefiles. Statistical methods were applied to measure the similarity between the predictive wildfires area and the actual wildfire areas. The methods include Sorensens Q statistic, Jaccard similarity coefficient, and Hamming distance. Area of intersection and union were calculated. The results of these statistics show that the performance of FARSITE simulation model is acceptable, which can be an option for predicting future wildfires.

The thesis of Yujia Hao is approved.

Nicolas Christou

Yingnian Wu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2018

To my beloved parents . . .

thank you for always standing by my side

offering unconditional love, support, and encouragement

To my fiancé . . .

thank you for your love and patience

making me see this world in a different way

TABLE OF CONTENTS

1 Introduction **1**

1.1 Objective 2

1.2 Wildfires in this project 2

1.3 Problem Statement 3

1.4 Related Work 4

2 FARSITE **5**

2.1 Introduction 5

2.2 FARSITE Data Inputs 6

2.3 FARSITE Data Outputs 8

3 Understanding Geographic Information System (GIS) **10**

3.1 Introduction of GIS 10

3.2 Vector Data 11

3.3 Raster Data 11

3.4 Advantages and Disadvantages for Vector and Raster Data 12

3.5 Coordinate Reference Systems (CRS) 13

4 Data Preparation **14**

4.1 Data Sources 14

4.2 Features of Dataset 14

4.3 Shapes of the Actual fire with Couple of Layers 15

4.4 Shapes of the Actual fire with Single Layer 17

4.5 Shapes of the Predictive Fire 19

4.6 Shapes of the Intersection between Predictive Fire and Actual Fire 23

5 Methodologies to Compare Actual Fire and Predictive Fire 26

5.1 Sorensen's Q statistic 26

5.2 Jaccard similarity coefficient 27

5.3 Hamming Distance 27

6 Results 29

7 Conclusion 31

8 Future Work 32

References 33

LIST OF FIGURES

2.1 Huygens principle 6

2.2 An example of a gridded weather file 7

3.1 Trees, mountains, water, lands are represented in this image can be regarded as vector data. The image was captured at Frank G Bonelli Regional Park, San Dimas, CA 12

4.1 Actual maps with elevation (1) 16

4.2 Actual maps with elevation (2) 16

4.3 Actual maps with elevation (3) 16

4.4 Actual maps with elevation (4) 17

4.5 Actual maps with elevation (5) 17

4.6 Actual maps with a single layer (1) 18

4.7 Actual maps with a single layer (2) 18

4.8 Actual maps with a single layer (3) 18

4.9 Actual maps with a single layer (4) 19

4.10 Actual maps with a single layer (5) 19

4.11 Predictive maps for Aspen 20

4.12 Predictive maps for Carstens 20

4.13 Predictive maps for Chariot 20

4.14 Predictive maps for Gobblers 21

4.15 Predictive maps for Hathaway 21

4.16 Predictive maps for Mountain 21

4.17 Predictive maps for Pfeiffer 22

4.18 Predictive maps for Rim 22

4.19 Predictive maps for Sharp 22
4.20 Predictive maps for Bridge 23
4.21 Intersection between Predictive and Actual Maps (1) 23
4.22 Intersection between Predictive and Actual Maps (2) 24
4.23 Intersection between Predictive and Actual Maps (3) 24
4.24 Intersection between Predictive and Actual Maps (4) 24
4.25 Intersection between Predictive and Actual Maps (5) 25

LIST OF TABLES

1.1 Summary of wildfires in this project(Pimlott,Laird,&Brown) 3

4.1 Duration of each wildfires 15

6.1 The area over intersection, predictive, and actual fire 29

6.2 Statistics to measure the similarity 30

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Professor Schoenberg of the University of California, Los Angeles, who gave me great moral support to learn wildfires in California from a statistical perspective. The door to Prof. Schoenberg office has always open whenever I bumped into a road block in my study. He consistently allowed this paper to be my own work, but steered me in the right direction and introduced me to the U.S. Forest Service. I sincerely thank Prof. Schoenberg for all the help and support he provided throughout the project.

I would also like to thank Haiganoush K. Preisler and John Benoit, statistical specialist and computer specialist at the U.S Forest Service. I wouldn't have finished without their help. Mr.Benoit always explained patiently and gave me detailed instructions about our data and work. He provided me with suggestions to make my study more efficient. Thank you so much for your help and support.

Finally, I must express my extremely profound gratitude to my parents and to my fiancé for providing me with unconditional support and continuous encouragement throughout my study and the process of writing this thesis. Thank you all.

CHAPTER 1

Introduction

Wildfire, also called wildland fires or forest fires typically burn between 1.6 million to 2.0 million hectares of land in the United Statees each year. Wildfires can move at speeds of up to 23km/hour, burning down almost everything in thier paths including trees, animals, homes, even humans (Thiessen, 2017).

The massive wildfires that burned in California, Oregon, Montana, Idaho, and other parts of North America in 2017 in many cases exhibited a disturbing trend: a marked increase in the amount of area burned. The Thomas Fire, one of the most catastrophic wildfires in California of 2017, did not just destroy neighborhoods and livelihoods, but also broke records for wildfire size in California, consuming 281,893 acres in Californias Santa Barbara and Ventura counties in December (Gabbert, 2018).

According to Verisks 2017 Wildfire Risk Analysis, 4.5 million U.S. homes were identified at high or extreme risk of wildfire, with more than 2 million in California alone. Losses from wildfires added up to $5.1 billion over the past 10 years. Beginning October 6 and continuing until October 25, the devastation of wildfires hit eight counties in Northern California. Resulting at least 23 fatalities, destroyed 245,000 acres and over 8700 structures. In Southern California, five major fires occurred during 2017. Just the Thomas Fire alone acused damages estimated by Catastrophe modeling firm RMS, up to $2.5 billion in an insured loss. The California Department of Insurance reported that insurance claims from the October December fires add up to almost $12 billion, which makes the 2017 fire season the costliest on record (Samanta, 2018).

Experts and researchers dedicate long shifts and endure strenuous workloads to looking for an accurate early-warning system like those that exist for other natural disasters, for

example, hurricanes and tornadoes, which are the direct result of certain atmospheric conditions. However, California wildfires are sometimes deemed inherently more unpredictable due to the nature thst more than 90% of them are caused by human activity (Abramson, 2017).

1.1 Objective

FARSITE is a fire growth simulation modeling system that is widely used by the U. S. Forest Service, National Park Service, and other federal and state land management agencies (FARSITE, 2018). In this project,we use the simulations in FARSITE to generate prediction wildfires maps, and compare them with the actual maps using different statistical methods to measure the accuracy of this model. The wildfires used in this project are Aspen, Carstens, Chariot, Gobblers, Hathaway, Mountain, Pfeiffer, Rim, Sharp, and Bridge.

1.2 Wildfires in this project

The information is collected from the 2013 Wildfire Redbook and the California Deparment of Forestry and Fire Protection. The areas burned ranges from 200 to 260,000 acres, and the most causes are due to human activity.

Table 1.1: Summary of wildfires in this project(Pimlott,Laird,&Brown)

Fire Name	County	Acres Burned	Cause
Aspen	Fresno	22,992	Lighting
Carstens	Mariposa	1,708	Human
Chariot	San Diego	7,055	Vehicle
Gobblers	San Bernardino	413	Lighting
Hathaway	Riverside	3,870	Under Investigation
Mountain	Riverside	27,531	Human
Pfeiffer	Monterey	917	Under Investigation
Rim	Tuolumne	257,314	Under Investigation
Sharp	San Bernardino	243	Under Investigation
Bridge	Mariposa	300	Vehicle

1.3 Problem Statement

In regression analysis or spatial-temporal statistical analysis with gridded data, one typically evaluates forecasts by simply comparing predicted to actual observed values. However, for wildfire maps generated by FARSITE, visualization of the comparison between the predicted fires will occur on the map to the actual area in which the wildfires occurred was implemented, where the model predicts the fires will burn, and the actual wildfires, where the burn actually happened. In this project, we used three statistical methods to measure the accuracy, Sorensen's Q statistic, Jaccard similarity coefficient, and Hamming distance.

The project is still new, therefore there is still a long way to go to improve the model and accuracy.

1.4 Related Work

The FARSITE fire growth model for fire predictions is the simulation model used in this project. Some works had been done using FARSITE modeling. Thomas M. Williams, Brian J. Williams, and Bo Song were able to obtain sufficient historical data to create a FARSITE input dataset for the 1894 Hinckley Fire. Model results can also explain aspects of the historical account that were not used in the calibration (Williams, Williams, &Song, 2014). Mark A. Finney and Kevin C. Ryan applied FARSITE on naturally ignited fires of Horizon fire and the Howling fire. As both fires were simulated successfully, they discovered that the simulations were more accurate for the Horizon fire than the Howling fire. The simulations clearly showed the need for accurate fuel maps and descriptions (Finney & Ryan, 2016). John Brakeall also applied FARSITE throughout Mexico's forest fires, where FARSITE has been used extensively throughout the United States (Brakeall, 2013). Also, FARSITE can be viewed as an option for fire modeling in the Southern Appalachian Mountains. With developing fuel models it can lead a way to better represent existing conditions of fuels (Phillips, Waldrop, &Simon, 2006).

The previous FARSITE works showed that the simulation model can work quite well on different spatial locations and provide insight about the procedure of the system, by comparing the actual and predicted fires by features. This project is however build up the visualization between maps, and utilizes statistical methods to measure the accuracy.

CHAPTER 2

FARSITE

2.1 Introduction

FARSITE is a fire growth simulation modeling system in which uses spatial information on topography and fuels along with weather and wind files. It incorporates existing fire behavior models for surface fire spread (Rothermel, 1972), crown fire initiation (Van Wagner, 1977), crown fire spread (Rothermel, 1991), post-frontal combustion(Albini and others 1995; Albini and Reinhardt 1995), and dead fuel moisture (Nelson, 2000) into a 2-dimensional fire growth model.

FARSITE has been developed for use only on personal computers (PCs) (Finney, 1998). Therefore, it produces outputs that are compatible with PC and Workstation graphics and GIS software for later analysis and display. GRASS and ARC/INFO GIS raster data themes are accepted by FARSITE.

Huygens principle of wave propagation has been applied to expand fire fronts(Finney, 1998). When the Huygens principle applied to propagation of light waves, the principle states that(Fitzpatrick, 2007):

> Every point on a wave front may be considered a source of secondary spherical wavelets which spread out in the forward direction at the speed of light. The new wave-front is the tangential surface to all of these secondary wavelets.

Huygens principle requires information only from points on the fire edge, which saves time and memory of computer compared to other models. For example, in cellular models, the distorted fire shape resulting from the grid must be minimized by calculating fire spread to

unburned cells within a wide radius of each active cell (French,1992).

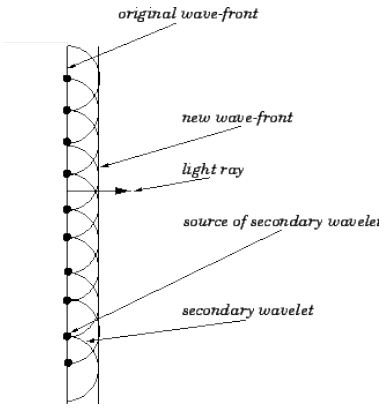


Figure 2.1: Huygens principle

In this project, I used FARSITE version 3 instead of FARSITE version 4, because the previous version is able to process gridded temperature and relative humidity fields. Version 4 removed gridded temperature and relative humidity processing, by replacing it with better fuel moisture processing.

2.2 FARSITE Data Inputs

To start it all off, a landscape file is required. FARSITE is compatible with Geographic Information System (GIS) to generate, manage, and provide spatial data themes containing vegetation, fules and topography. Also, FARSITE requires 5 raster data themes, which include elevation, slope, aspect, fuels, and canopy. The 5 raster data themes are the GIS file data units, and they are required to build up a landscape file. For elevation, slope, aspect, fuels, and canopy, they must be co-registered (e.g. have the same reference point, projection, and units), identical resolution (e.g. cell size must be the same for all themes), same extent (the corners of the rectangular spatial region must be the same)(Data Requirements, 2018). Also, optional themes can be chosen to precisely create the landscape. There are crown bulk density, crown base height, stand height, duff loading, and coarse woody. As a result, with the at least 5 raster data themes, we can get a landscape file, which is a binary file comprised of a header and a body of short integers for each of the themes it contains. The

header contains information on the bounds of the area, the resolution of the cells, and the units of the themes (Data Requirements, 2018). The file ext is .LCP, and the file type is GIS.

Then, FARSITE input data files are required to run the system. There are 5 basic files, landscape, weather, wind, adjustment, initial fuel moisture. For the weather file, it is a ASCII text files, which contains daily observations on temperature, humidity, and precipitation that depicts a temporal weather stream. The weather stream greatly simplified actual variations in the weather. For this project, I used gridded weather inputs, it is an . ATM file, which has a format like the following:

```
WEATHER_AND_WINDS
7 15 0200 TEST01.TMP TEST01.HMD TEST01.PPT TEST01.SPD TEST01.DIR TEST01.CLD
7 15 0600 TEST02.TMP TEST02.HMD TEST02.PPT TEST02.SPD TEST02.DIR TEST02.CLD
7 15 1200 TEST03.TMP TEST03.HMD TEST03.PPT TEST03.SPD TEST03.DIR TEST03.CLD
7 15 1600 TEST04.TMP TEST04.HMD TEST04.PPT TEST04.SPD TEST04.DIR TEST04.CLD
7 15 2000 TEST05.TMP TEST05.HMD TEST05.PPT TEST05.SPD TEST05.DIR TEST05.CLD
…
…
```

Figure 2.2: An example of a gridded weather file

Where TESTxx.TMP, TESTxx.HMD, TESTxx.PPT,etc. are all other text files containing grids (in ArcASCII format) of weather variables.

For the wind file, it is also a ASCII text file, and contained in a Wind(.WND) file or as a gridded weather (.ATM) file as a stream of data. As usual, winds are variables in space and time. However, for FARSITE, winds are assumed to be constant in space for a given wind stream but variable in time. This means there is no topographic effects on winds. The input format is similar to that for a Weather (.WTR) File. As the same as Weather File, the Wind file used in this project is also gridded.

For the adjustment factor file, it is also a ASCII text file. It represents the rate of spread adjustment factors within the forecast. Users can use their experienced judgement or local

data to fine-tune the simulation to the observed or actual fire spread patterns. For example, if the adjustment factor is 0.5, the spread rate for a given fuel type would be reduced by half, resulting with 0.25. In this project, we used 1.0 as adjustment factors in which maintains the original spread rate.

For initial fuel moistures files, it is also a ASCII text file, which must be set for each fuel type at the beginning of the simulation. Because these fuel moistures are required to begin the process of calculating site specific fuel moistures at each time step throughout the simulation (Data Requirements, 2018). The fuel models in this project is Anderson fire behavior, which serve as an input to Rothermels mathematical surface fire behavior and spread model. The 13 Anderson fire behavior fuel model layer represents distinct distributions of fuel loading found among surface fuel components (live and dead), size classes, and fuel types. The fuel models are described by the most common fire carrying fuel type (grass, brush, timber litter, or slash), loading and surface area-to-volume ratio by size class and component, fuelbed depth, and moisture of extinction (13 Anderson Fire Behavior Fuel Models, 2018).

There are 8 optional data inputs: fuel model conversion, custom fuel models, fire acceleration, air attack resources, coarse woody profiles, burn period, ground attack resources, and gridded weather and winds. In this project, we only used gridded weather and winds as optional FARSITE input data files.

2.3 FARSITE Data Outputs

The outputs for FARSITE is not as complicated as the inputs. There are three types of outputs, vector files, shapefiles, and raster files. Vector data and raster data will be introduced in next chapter about understanding Geographic Information System (GIS). Shapefile format is a popular geospatial vector data format in which can spatially describe vector features. For example, points, lines, and polygons, representing waters from map and so on. Attributes can be assigned on each item.

In this project, for determining the ignition, a .VCT file was created for each of our

historic fire cases to define a small polygon around the ignition location. It can provide a relatively accurate ignition.

CHAPTER 3

Understanding Geographic Information System (GIS)

3.1 Introduction of GIS

A geometric information system (GIS) is a framework aimed to gather, manage, and analyze data. It can present spatial and geographic data and organize layers of information into visualizations using maps and 3D scenes. GIS system can help users to have better understand of the patterns, relationships, and situations.

A GIS system consists of three components (Introducing GIS, 2018):

1. Digital data, which is the geographical information that we are going to view and analyze using computer hardware and software. For example, such as natural or constructed features, spatial attributes, and so on.

2. Computer hardware, which is usually a computer used for storing data, displaying graphics and processing data.

3. Computer software, which is called a GIS application, can run on the computer hardware and deal with our digital data. For example, ArcGIS, QGIS, ArcMap. In this project, I utilized QGIS.

GIS has a common feature that the non-geographic data can be associated with geographic data. Non- geographic data usually includes numbers, characters, and logical types. Geographic data used in GIS are vector and raster data. For example, the GIS Application can draw the layer based on non-geographic data, such as gender, disease type, and so on (Introducing GIS, 2018).

3.2 Vector Data

Vector data is used to present the real-world entities or its original resolution and form using points, lines and polygons within the GIS environment. For example, in regards to figure 3.1, we can see entities such as trees, mountains, water and so on. The attributes of features consist of text or numerical information. Geometry is used to represent the shape of a vector feature. As we know, the geometry is made up by one or more interconnected vertices. X, Y, and optionally Z axis define the position of a vertex in the space. There are three types of vector data, points, lines, and polygons.

For points, the geometry of the feature consists of only a single vertex. For example, a tree in our image can be a point, and maybe the year can be an attribute of this point. When the geometry of the feature consists of two or more vertices with the unequal first and last vertex, we can define the feature as a line, resulting in a sequence of joined vertices. For example, the road in our image can be a line, the length of the road can be an attribute of the feature. When the geometry of the feature consists of at least four vertices with the equal first and last vertex, we define the feature as a polygon. The polygon is represented as an enclosed shape. For example, the tent in the image can be defined as a polygon. In this project, the fire ignition is defined as a closed polygon (Vector Data,2018).

3.3 Raster Data

A raster data consists of a matrix of pixels (or cells), where each pixel carries a value that can represent the conditions for the region or the information of the area covered by the pixel. For instance, in a raster data, green pixels represent agriculture, blue pixels represent water, yellow pixels represent road, and red pixels represent residential. To sum it all up, a geographical region can be represented by each pixel and the value assigned to the pixel can represent some characteristic of the region and area.



Figure 3.1: Trees, mountains, water, lands are represented in this image can be regarded as vector data. The image was captured at Frank G Bonelli Regional Park, San Dimas, CA

3.4 Advantages and Disadvantages for Vector and Raster Data

As we put vector data and raster data together, we can see that vector data is more likely to separate the space into different parts, and each part represents respective object. However, raster data divides the space based on the region, and then divides again into pixels as a square tessellation of two or three dimensions. Therefore, raster data has the advantages that the data structure is relatively simple to interpret. The implementation of overlays is simple as well, and the image processing is comparably efficient, while the vector data have an efficient representation of topology, can adapt well to scale changes, allows representing networks and easy association with attribute data. The disadvantage for raster data is that it has less compact data structure, difficulties in representing topology, and the cell boundaries are independent of feature boundaries. While for vector data, the disadvantages are the data structure is complex, the implementation of overlays are more difficult, and image processing

is inefficient (Burrough & McDonnell,1998)

In this project, I used Shapefiles that stored the vector data with the advantages that is easy to process in R.

3.5 Coordinate Reference Systems (CRS)

A map projection is used to transform latitudes and longitudes of locations from a sphere or an ellipsoids surface into location on a plane (Snyder, 1989). A coordinate reference system (CRS), also called a spatial reference system (SRS), defines how the two-dimensional projected map locates the real geographical entities on the earth based on coordinates. A Spatial References System Identifier (SRID) is a unique value to identify different coordinate system. The most common authority of SRID is called European Petroleum Survey Group (EPSG), different codes represent different coordinate systems. There are two common coordinate systems, called UTM and WGS84. In this project, we used Albers projection due to the simple latitude and longitude which is not as sensitive as Albers projection to detect the change of fire spread. The Albers projection algorithm (Snyder, 1987) is as following (Weisstein, 2018),:

$$\mathrm{x} = \rho \sin\theta$$

$$\mathrm{y} = \rho_0 \text{ - } \rho \cos\theta$$

where

$$\mathrm{n} = \tfrac{1}{2}\left(\sin\varphi_1 + \sin\varphi_2\right)$$

$$\theta = n\left(\lambda - \lambda_0\right)$$

$$C = \cos^2\varphi_1 + 2n\sin\varphi_1$$

$$\rho = \tfrac{1}{n}\left(C - 2n\sin\varphi\right)^{0.5}$$

$$\rho_0 = \tfrac{1}{n}\left(C - 2n\sin\varphi_0\right)^{0.5}$$

λ is the longitude,λ_0 is the reference longitude, φ is the latitude, φ_0 is the reference latitude, φ_1,φ_2 are the standard parallels.

CHAPTER 4

Data Preparation

4.1 Data Sources

The wildfire spatial data includes geographic data, canopy data, fuels data, weather data, and perimeters, are all collected from LANDFIRE (LF). The information is supported by the U.S. Department of Agriculture Forest Service. The Forest Service is an agency of the USDA that monitor 154 national forests and 20 national grasslands (LANDFIRE Partners,2018). LANDFIRE (LF) is a shared program with USDA that provides over 20 national geo-spatial layers(LANDFIRE Home,2018). The start dates were from Projected Significant Resource Demobilization Start Date, and the stop dates were taken from ICS 209 Anticipated Incident Containment or Completion Date.

4.2 Features of Dataset

I've obtained 10 significant wildfire data that can run successfully in R. The data is from Aspen in 2013, Carstens in 2013, Chariot in 2013, Gobblers in 2013, Hathaway in 2013, Mountain in 2013, Pfeiffer in 2013, Rim in 2013, Sharp in 2013, and Bridge in 2014. All the datasets include geographic data, canopy data, fuels data, weather data, and perimeters are required by FARSITE simulation models.

The original coordinate reference system (CRS) for each wildfire is in WGS84, which means it describes the burned area in earths original size and shape in latitude and longitude. I transformed map projection from WGS84 to Albers in R with the package Rgdal to make it consistent with the predictive wildfire map projection.

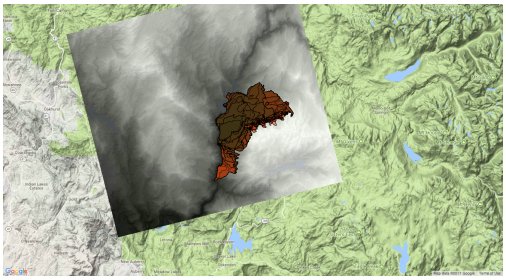
The duration of simulation in FARSITE varies depending on the case of each wildfire. The data collected for each wildfire is not perfectly complete. For example, the start date for Aspen is July 22nd 2013, but the data collected starts from July 24th 2013. The recorded time varies on different days. On the 24th, staff recorded the fire at around 10 p.m. for once, but he or she recorded the fire for couple of times on 25th. I chose a few hours or a few days as spread duration to run simulations, because there is no need for the simulation to go as long as the actual fire. Since we do not have definite method to measure when the fires behavior is strongly influenced by the firefighting activity, therefore it will definitely affect the actual fire. The following table is the duration I chose for each fire to run the simulations.

Table 4.1: Duration of each wildfires

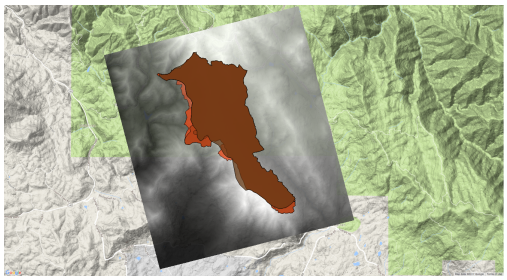
Year	Name	Start Date	Stop Date	Simulation Duration	Time Compared with Actual Fire
2013	Aspen	7.22 22:15	9.8 18:00	7.23 1:00-7.24 22:00	7.24 22:11
2013	Carstens	6.16 14:12	6.20 00:00	6.16 14:00-6.18 22:00	6.18 22:01
2013	Chariot	7.6 12:55	7.22 6:00	7.7 14:00-17:00	7.7 17:31
2013	Gobblers	8.20 13:32	8.30 18:00	8.22 1:00-22:00	8.22 20:52
2013	Hathaway	6.9 12:30	10.15 00:00	6.9 13:00 -6.10 4:00	6.10 3:20
2013	Mountain	7.15 13:43	7.31 18:00	7.15 13:00-7.16 7:00	7.16 6:42
2013	Pfeiffer	12.16 00:20	12.30 8:00	12.16 1:00- 12.17 17:00	12:17 17:17
2013	Rim	8.17 15:25	10.24 18:00	8.17 16:00 - 8.19 11:00	8.19 10:43
2013	Sharp	8.8 13:15	8.18 18:00	8.8 13:00 - 8.9 1:00	8.9 1:17
2014	Bridge	9.5 12:39	9.12 00:00	9.5 13:00-9.6 23:00	9.6 23:51

4.3 Shapes of the Actual fire with Couple of Layers

In our dataset, perimeters data records the shape of actual fire in different time. Therefore, regarding each shape at different time as a layer, we got couple of layers to draw on the real map with elevation via QGIS. The shapes are as followed, and we can see the elevation data is considerably consistent with the actual map.

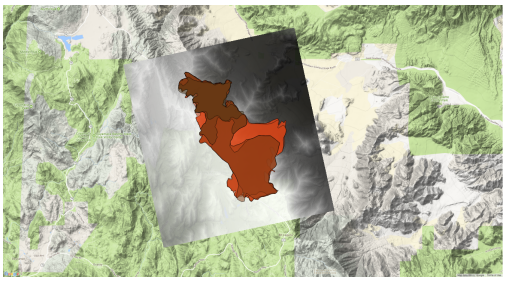


(a) Aspen

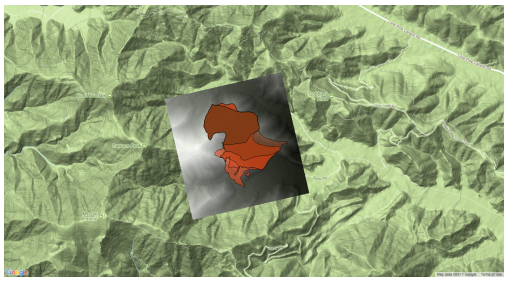


(b) Carstens

Figure 4.1: Actual maps with elevation (1)

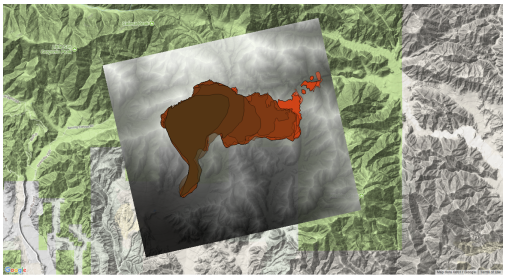


(a) Chariot

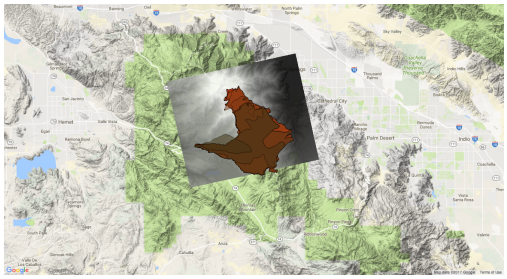


(b) Gobblers

Figure 4.2: Actual maps with elevation (2)

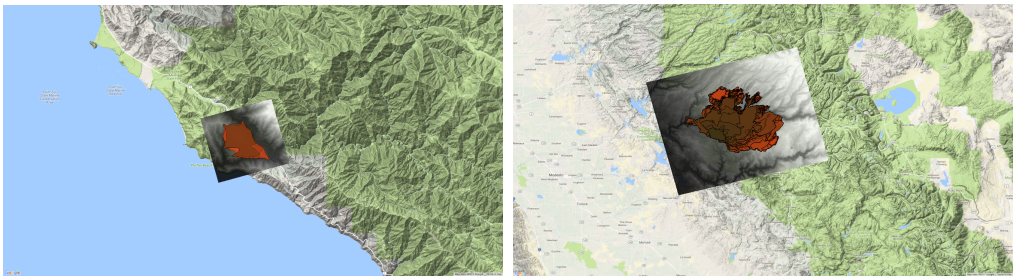


(a) Hathaway



(b) Mountain

Figure 4.3: Actual maps with elevation (3)



(a) Pfeiffer (b) Rim

Figure 4.4: Actual maps with elevation (4)



(a) Sharp (b) Bridge

Figure 4.5: Actual maps with elevation (5)

4.4 Shapes of the Actual fire with Single Layer

The longer the simulation duration is, the more unpredictable the wildfire is. Therefore, the methodology applied in chapter 5 depends on the shape at a specific time frame, I only chose a few hours or a few days as spread duration to run simulations. The time is the last column in table 4.1 The shapes of the actual fire with single layer via R are as the following.

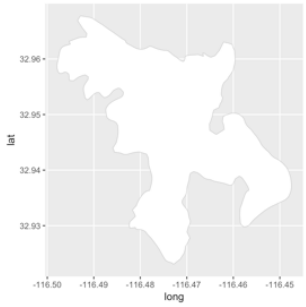


(a) Aspen



(b) Carstens

Figure 4.6: Actual maps with a single layer (1)

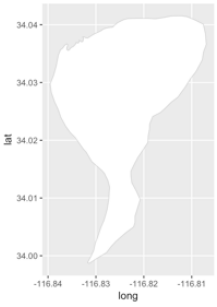


(a) Chariot

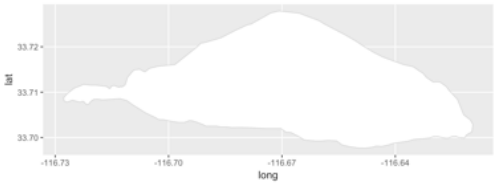


(b) Gobblers

Figure 4.7: Actual maps with a single layer (2)

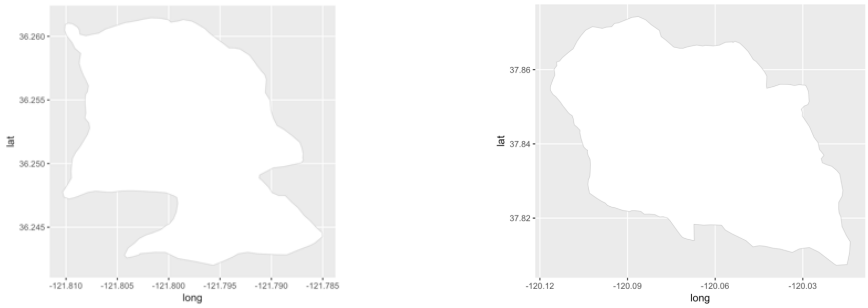


(a) Hathaway



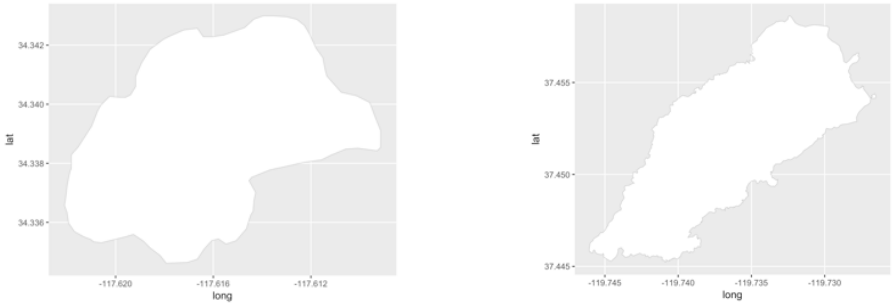
(b) Mountain

Figure 4.8: Actual maps with a single layer (3)



(a) Pfeiffer (b) Rim

Figure 4.9: Actual maps with a single layer (4)

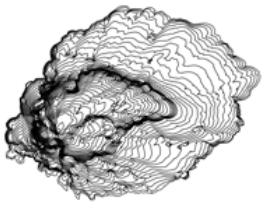


(a) Sharp (b) Bridge

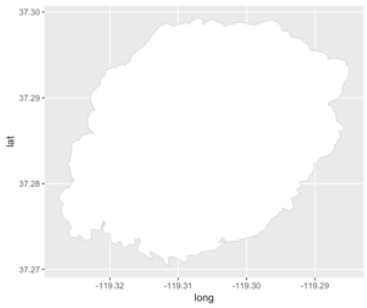
Figure 4.10: Actual maps with a single layer (5)

4.5 Shapes of the Predictive Fire

The predictive fires were simulated from FARSITE is based on simulation duration in table 4.1. I processed the shape in R using the vector data in terms of Shapefiles, which contains the attributes of each fire. The spatial data frame of the predictive fire contains multiple polygons based on the simulation process. After reading the Shapefiles in R, we need to combine the intersecting geometries and keep the most outside one. The shapes of predictive fires are as the following.

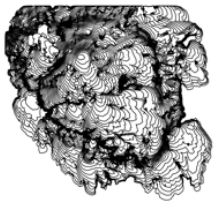


(a) With visible simulations

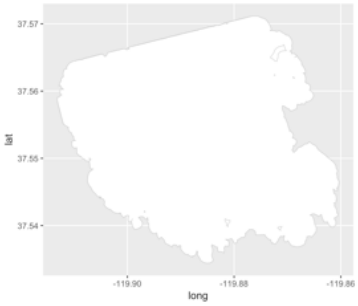


(b) Without visible simulations

Figure 4.11: Predictive maps for Aspen



(a) With visible simulations

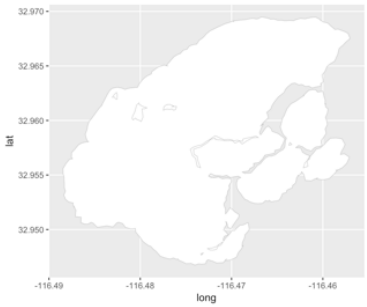


(b) Without visible simulations

Figure 4.12: Predictive maps for Carstens

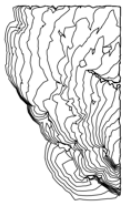


(a) With visible simulations



(b) Without visible simulations

Figure 4.13: Predictive maps for Chariot

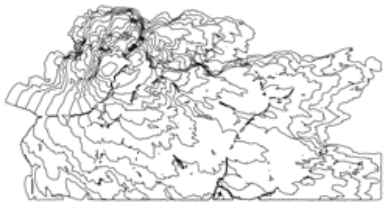


(a) With visible simulations

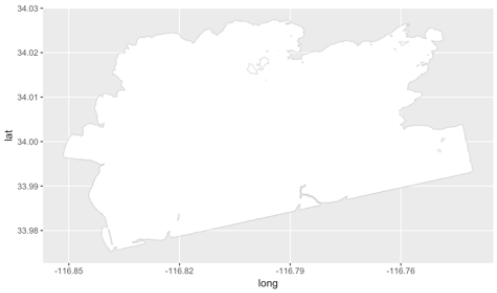


(b) Without visible simulations

Figure 4.14: Predictive maps for Gobblers

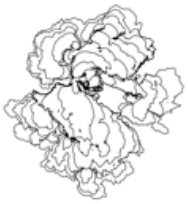


(a) With visible simulations

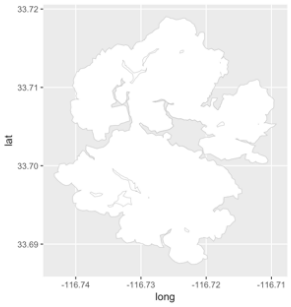


(b) Without visible simulations

Figure 4.15: Predictive maps for Hathaway

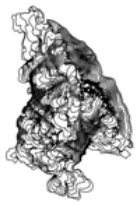


(a) With visible simulations

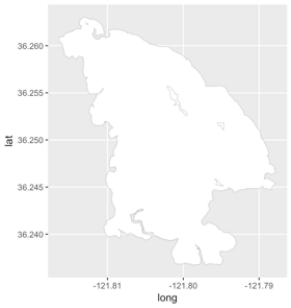


(b) Without visible simulations

Figure 4.16: Predictive maps for Mountain

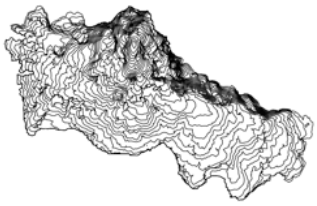


(a) With visible simulations

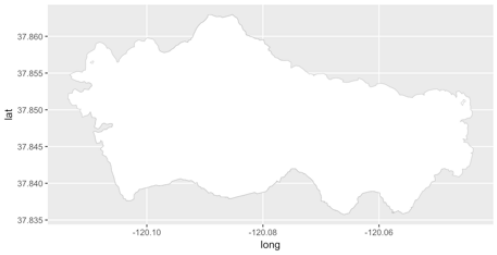


(b) Without visible simulations

Figure 4.17: Predictive maps for Pfeiffer

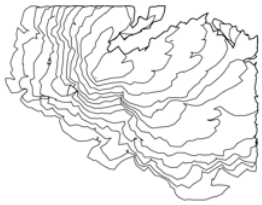


(a) With visible simulations



(b) Without visible simulations

Figure 4.18: Predictive maps for Rim

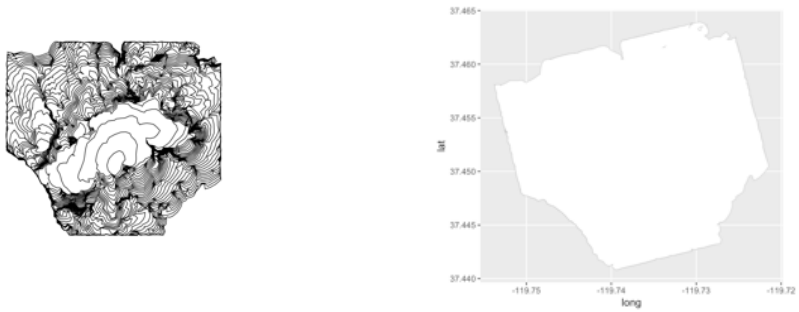


(a) With visible simulations



(b) Without visible simulations

Figure 4.19: Predictive maps for Sharp



(a) With visible simulations (b) Without visible simulations

Figure 4.20: Predictive maps for Bridge

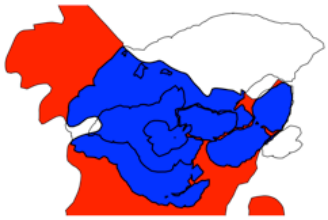
4.6 Shapes of the Intersection between Predictive Fire and Actual Fire

The red parts below is the actual fire of a single layer, therefore means that the actual fire at the given specific time. The blue parts is the intersection of the predicative fire and the actual fire. This is the part used to measure the accuracy of the prediction in chapter 5.

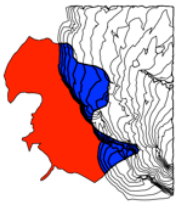


(a) Aspen (b) Carstens

Figure 4.21: Intersection between Predictive and Actual Maps (1)

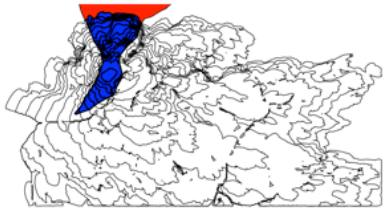


(a) Chariot

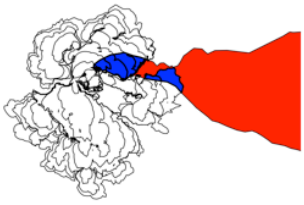


(b) Gobblers

Figure 4.22: Intersection between Predictive and Actual Maps (2)

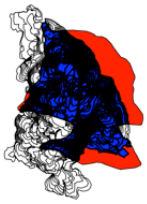


(a) Hathaway

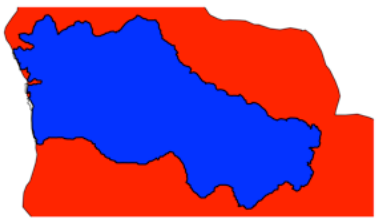


(b) Mountain

Figure 4.23: Intersection between Predictive and Actual Maps (3)



(a) Pfeiffer



(b) Rim

Figure 4.24: Intersection between Predictive and Actual Maps (4)



(a) Sharp (b) Bridge

Figure 4.25: Intersection between Predictive and Actual Maps (5)

CHAPTER 5

Methodologies to Compare Actual Fire and Predictive Fire

In chapter 4, we visualized the intersection of actual fire and the predictive fire using FARSITE simulation model. In chapter 5, I wanted to use methodologies to statistically compare the two maps, and discover how close the prediction fires to the actual fires are. Based on the discrete variables of spatial data, I used Sorensens Q statistic, Jaccard similarity coefficient, and Hamming distance are the three statistical methods.

5.1 Sorensen's Q statistic

The Sorensens Q statistic, also known as the Sorensen-Dice coefficient, is a statistic to measure the similarity of the given two samples. The value of Sorensens Q statistic is between 0 and 1, the closer to 1, the more similar the two sets are. When there is no similarity, Sorensens Q statistic is 0. In this project, the two samples are from the predictive fire maps and the actual fire maps. Sorensens formula was originally applied to binary data, and shown as below (Sorensen Dice coefficient,2018):

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|}$$

Where $|X|$ and $|Y|$ are the variables from two sets. In this project, $|X \cap Y|$can be regarded as the intersection of the two maps. It is the blue section in chapter 4, while $|X| + |Y|$ is the sum of the two maps. When the two maps are completely the same, the value of Sorensens Q statistic will be 1, and when they have no intersection, the value will be 0 instead.

5.2 Jaccard similarity coefficient

The Jaccard similarity coefficient is also a statistic used for comparing the similarity of two sets, which represents the intersection over the union of the two sets. It is a statistical method similar to Sorensens Q statistic as the maximum value is 1, and the minimum value is 0, but the algorithm is slightly different as shown below (Jaccard index,2018).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

where $|A \cap B|$ is the intersection of two maps, and $|A \cup B|$ is the union of the two maps.

5.3 Hamming Distance

Hamming distance is a statistical method to measure the difference between two strings in equal length. Not only does the value matters, but also the position needs to be the same. For example, the difference between 101 and 110 is 2, this is because only their first letter is the same. Therefore, applying Hamming distance in measuring the similarity of two maps is a different method from spatial perspective rather than geometry perspective compared with Sorensens Q statistic, and Jaccard similarity coefficient.

First of all, for this method I sampled the point locations in actual fire maps randomly, and generated the same numbers of point locations in predictive fire maps randomly. We need a large sample size, for all the maps the number of points is 100,000.When I plotted the samples of actual fires, the shape of samples is almost the same with the actual fire maps, which is also in the same situation for predictive fire maps. Then, I used intToBits function from R to convert integer vector to binary vector with 32 times the length of the integer vector. Thus, we got a considerably long string that encodes the actual point locations into 0 and 1. Then, I did the same conversion to the predictive fire maps. I compared longitude and latitude respectively, resulting in the hamming distance of longitude and latitude is almost the same. In this project, the Hamming distance can tell us the difference between the two strings, one is from the actual fire, and the other is from the predictive fire. The

result is not enough to give a rough sense about the degree of the differences. Therefore, I used a ratio to represent the dissimilarity in which is the Hamming distance over the length of strings.

CHAPTER 6

Results

Table 6.1: The area over intersection, predictive, and actual fire

Year	Name	Area of Intersection(m^2)	Area of Predictive Fire(m^2)	Area of Actual Fire(m^2)
2013	Aspen	2,777,905	8,331,804	7,860,680
2013	Carstens	4,534,180	12,847,241	6,562,459
2013	Chariot	2,933,366	4,541,595	9,558,277
2013	Gobblers	413,513	2,406,910	1,383,596
2013	Hathaway	1,951,838	38,104,121	6,624,449
2013	Mountain	399,963	6,189,205	17,630,820
2013	Pfeiffer	2,269,201	3,893,184	3,112,872
2013	Rim	11,094,026	11,119,548	41,151,209
2013	Sharp	604,216	2,159,255	652,585
2014	Bridge	1,090,843	4,963,576	1,093,251

Table 6.2: Statistics to measure the similarity

Year	Name	DSC	J(A,B)	HD(lon)	HD(lat)	HD Ratio(lon)	HD Ratio(lat)
2013	Aspen	0.34	0.21	621,472	626,674	0.19	0.20
2013	Carstens	0.44	0.30	639,883	624,942	0.20	0.20
2013	Chariot	0.42	0.26	639,240	710,777	0.20	0.22
2013	Gobblers	0.22	0.12	568,524	597,395	0.18	0.19
2013	Hathaway	0.09	0.05	700,681	831,177	0.22	0.26
2013	Mountain	0.03	0.02	715,920	645,427	0.22	0.20
2013	Pfeiffer	0.65	0.48	716,590	616,462	0.22	0.19
2013	Rim	0.42	0.27	657,078	687,880	0.21	0.21
2013	Sharp	0.43	0.27	563,758	594,801	0.18	0.19
2014	Bridge	0.36	0.22	599,572	671,006	0.19	0.21

DSC is the Sorensen's Q statistic, J(A,B) is the Jaccard Similarity Coefficient, HD is the Hamming Distance.

CHAPTER 7

Conclusion

According to the results in chapter 6, we can see that the most Sorensens Q statistic is higher than 0.40. The highest Sorensens Q statistic is around 0.65, which indicates the predictive wildfires map generated from FARSITE is quite similar to the actual fire map. For Jaccard similarity coefficient, most of them are around 0.20, and the highest being 0.48. This means if we have 100 samples in each fire dataset, 48 will be matched in both maps. For the trend of Hamming distance ratio, it is plausible that when Sorensens Q statistic and Jaccard similarity coefficients are high, the ratio of Hamming distance will be relatively low.

Also, I used bootstrap method to calculate the mean of Sorensens Q statistic, Jaccard similarity coefficient, and Hamming distance ratio. By bootstrapping for 100,000 times, the 95th percentile of the bootstrap distribution of the mean for Sorensens Q statistic is 0.4284884, and the standard deviation is 0.05763724. For Jaccard similarity of coefficient, the 95th percentile of the bootstrap mean is 0.2874254, and the standard deviation is 0.005155877. For Hamming distance ratio, the mean is 0.2094719, and the standard deviation is 0.005640608. These results show that FARSITE simulation model is quite an option to predict the wildfire.

CHAPTER 8

Future Work

Although FARSITE simulation model seems perform acceptably in this project, it still over-predicts according to the maps and area of predictive fires. The reason can be attributed to the adjustment factor files of fire spread rate. I used 1.0 to maintain the original spread rate, however, the fire spread rate varies and significantly depends on the topographical data, fuels data, and weather data. Also, the inaccuracy of data on weather model, fuel model, like the fuel moistures, and fuel descriptions may lead to over prediction. Information collected often has many errors and guesses, since it is usually recorded by active fire crews at the scene. In the future, more work should be done in calibrating the rate of fire spread and the variation of fire behaviors (Finney, 1998).

REFERENCES

[1] 13 Anderson Fire Behavior Fuel Models. (n.d.). Retrieved April 05, 2018, from https://www.landfire.gov/fbfm13.php

[2] Abramson, A. (n.d.). California Fires: Why They're So Hard to Predict. Retrieved April 04, 2018, from http://time.com/4977123/wildfire-california-nor-calprediction/

[3] Brakeall, J. (n.d.). Wildfire Assessment Using FARSITE Fire Modeling: A Case Study in the Chihuahua Desert of Mexico. Retrieved April 04, 2018, from http://digitalcommons.fiu.edu/etd/923/

[4] Burrough, P.A. and McDonnell, R.A. (1998) *Principles of Geographical Information Systems*. Oxford University Press, New York, 151 p.

[5] Data Requirements. (n.d.). Retrieved April 05, 2018, from http://www.fire.org/downloads/farsite/WebHelp/usersguide/ug3_data_requirements.htm

[6] FARSITE. (n.d.). Retrieved April 09, 2018, from https://www.firelab.org/project/farsite

[7] Finney, M. A. (1998). FARSITE: Fire Area Simulator-model development and evaluation. *FARSITE: Fire Area Simulator-model Development and Evaluation.* doi:10.2737/rmrs-rp-4

[8] Finney, M. A., & Ryan, K. C. (2016, February 18). Use of the FARSITE Fire Growth Model for Fire Prediction in U.S. National Parks. Retrieved April 4, 2018, from https://pdfs.semanticscholar.org/165c/dcb34394413ca456a155dbefe123e01ef72d.pdf

[9] Fitzpatrick, R. (2007, July 14). Huygens' principle. Retrieved April 05, 2018, from http://farside.ph.utexas.edu/teaching/302l/lectures/node150.html

[10] French, I.A. 1992. Visualisation techniques for the computer simulation of bushfires in two dimensions. University of New South Wales, Australian Defence Force Academy. M.S. Thesis.

[11] Gabbert, B. (n.d.). Wildfire problem to increase in coming decades. Retrieved April 04, 2018, from http://wildfiretoday.com/tag/research/

[12] Introducing GIS. (n.d.). Retrieved April 05, 2018, from https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/introducing_gis.html

[13] Jaccard index. (2018, April 06). Retrieved April 07, 2018, from https://en.wikipedia.org/wiki/Jaccard_index

[14] Jr, R. M. (2000). Prediction of diurnal change in 10-h fuel stick moisture content. *Canadian Journal of Forest Research*, 30(7), 1071-1087. doi:10.1139/x00-032

[15] LANDFIRE Partners. (n.d.). Retrieved April 06, 2018, from https://www.landfire.gov/lf_partners.php

[16] LANDFIRE: Home. (n.d.). Retrieved April 06, 2018, from https://www.landfire.gov/index.php

[17] Phillips, Ross J.; Waldrop, Thomas A.; Simon, Dean M. (2006). Assessment of the FARSITE model for predicting fire behavior in the Southern Appalachian Mountains. Proceedings of the 13th biennial Southern Silvicultural Research Conference. Gen. Tech. Rep. SRS-92. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 521-525

[18] Pimlott, K., Laird, J., & Brown, E. G. (2014). 2013 Wildfire Activity Statistics. *2013 Wildfire Activity Statistics*. Retrieved April 5, 2018, from http://www.fire.ca.gov/downloads/redbooks/2013Redbook/2013_Redbook_Final.pdf

[19] Richards, G.D. (1990). An elliptical growth model of forest fire fronts and its numerical solution. *Int. J. Numer. Meth. Eng.*30: 1163-1179.

[20] Rothermel, R.C. 1972. A mathematical model for predicting fire spread in wildland fuels. *USDA For. Serv. Res. Pap.* INT-115.

[21] Rothermel, R.C. 1991. Predicting behavior and size of crown fires in the northern Rocky Mountains. *USDA For. Serv. Res. Pap.* INT-438.

[22] Samanta, A. (n.d.). Key findings from the 2017 Verisk wildfire risk analysis. Retrieved April 04, 2018, from https://www.verisk.com/insurance/visualize/key-findings-from-the-2017-verisk-wildfire-risk-analysis/?utm_source=Social\&utm_medium=Twitter\&utm_campaign=VeriskSM\&utm_content=842017

[23] Snyder, J.P. (1989). Album of Map Projections, *United States Geological Survey Professional Paper*. United States Government Printing Office. 1453.

[24] Snyder, John P. (1987). Map Projections A Working Manual. *U.S. Geological Survey Professional Paper 1395*. Washington, D.C.: United States Government Printing Office.

[25] SorensenDice coefficient. (2018, April 05). Retrieved April 07, 2018, from https://en.wikipedia.org/wiki/SrensenDice_coefficient

[26] Thiessen, M. (2018, January 18). Wildfires Information and Facts. Retrieved April 04, 2018, from https://www.nationalgeographic.com/environment/natural-disasters/wildfires/

[27] Van Wagner, C.E. (1977). Conditions for the start and spread of crownfire. *Canadian Journal of Forest Research*7: 23-24.

[28] Vector Data. (n.d.). Retrieved April 06, 2018, from https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/vector_data.html

[29] Weisstein, Eric. "Albers Equal-area Conic Projection". Wolfram MathWorld. Wolfram Research. Retrieved 2018-04-05.

[30] What Is GIS. (n.d.). Retrieved April 05, 2018, from https://www.esri.com/en-us/what-is-gis/overview

[31] Williams, T. M., Williams, B. J., & Song, B. (2014). Modeling a Historic Forest Fire Using GIS and FARSITE. *Modeling a Historical Forest Fire Using GIS and FARSITE*, 6(2), 80-88. Retrieved April 4, 2018, from http://mcfns.net/index.php/Journal/article/view/6_80/183pdf