

UCLA

UCLA Electronic Theses and Dissertations

Title

Evaluating the Use of Various Distance Metrics for Assessing a Model's Wilfire Prediction Performance

Permalink

<https://escholarship.org/uc/item/0nz4z9mf>

Author

Chao, Jeffrey

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Evaluating the Use of Various Distance Metrics
For Assessing a Model's Wildfire Prediction Performance

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

by

Jeffrey Chao

2018

© Copyright by
Jeffrey Chao
2018

ABSTRACT OF THE THESIS

Evaluating the Use of Various Distance Metrics
For Assessing a Model's Wildfire Prediction Performance

by

Jeffrey Chao

Master of Science in Statistics

University of California, Los Angeles, 2018

Professor Frederic R. Paik Schoenberg, Chair

While being able to predict wildfires is crucial in being able to safely control them, also important are valid methods and metrics to evaluate the effectiveness and performance of these predictions. Thus, in this paper we adapt fourteen existing distance metrics for use in assessing the accuracy of wildfire predictions. These fourteen distance metrics include: Simple Matching, Simpson, Kulczynski, Jaccard, Yule, and Phi. The theoretical implications of each distance metric are examined, and these metrics are applied to comparisons of actual wildfire perimeter data from ten wildfires and their predicted perimeters generated by the FARSITE fire prediction software. By doing so, we are able to illuminate the characteristics of each of the distance metrics (takes rotations into account, punish overburn/underburn, etc.), as well as illustrate how these distance metrics can be easily applied and chosen from based on one's requirements of a prediction model.

The thesis of Jeffrey Chao is approved.

Hongquan Xu

Nicolas Christou

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

1	Introduction	1
2	Prediction Model Methodology	3
2.1	Introduction of FARSITE	3
2.2	FARSITE Data Inputs	4
2.3	Features of Data	7
3	Prediction Model Results	10
3.1	Shapes of the Actual fire with Couple of Layers	10
3.2	Shapes of the Actual fire with Single Layer	12
3.3	Shapes of the Predictive Fire	14
3.4	Shapes of the Intersection between Predictive Fire and Actual Fire	18
4	Introduction to Distance Metrics	20
4.1	Simple Matching, Rogers, Hamann	21
4.2	Simpson, Braun, Kulczynski	22
4.3	Dice, Jaccard, Ochiai	24
4.4	Sneath, Rao, Yule	25
4.5	Phi, Kappa	26
5	Application of Distance Metrics to Wildfire Data	28
5.1	Results Under Original Observation Region	29
5.1.1	Simple Matching, Rogers, Hamann	29
5.1.2	Simpson, Braun, Kulczynski	29
5.1.3	Dice, Jaccard, Ochiai	30

5.1.4	Sneath, Rao, Yule	31
5.1.5	Phi, Kappa	31
5.2	Results Under Smallest Possible Observation Region	32
5.2.1	Simple Matching, Rogers, Hamann	32
5.2.2	Rao, Yule	33
5.2.3	Phi, Kappa	33
6	Discussion of Results of Application of Distance Metrics	35
6.1	Simple Matching	35
6.2	Rogers	36
6.3	Hamann	36
6.4	Simpson	37
6.5	Braun	37
6.6	Kulczynski	37
6.7	Dice, Jaccard, Sneath	38
6.8	Ochiai	38
6.9	Rao	39
6.10	Yule, Phi	39
6.11	Kappa	40
7	Conclusion	42
A	Appendix	44
	References	61

LIST OF FIGURES

2.1	Huygens principle	4
2.2	An example of a gridded weather file	6
3.1	Actual maps with elevation (1)	10
3.2	Actual maps with elevation (2)	11
3.3	Actual maps with elevation (3)	11
3.4	Actual maps with elevation (4)	11
3.5	Actual maps with elevation (5)	12
3.6	Actual maps with a single layer (1)	12
3.7	Actual maps with a single layer (2)	13
3.8	Actual maps with a single layer (3)	13
3.9	Actual maps with a single layer (4)	13
3.10	Actual maps with a single layer (5)	14
3.11	Predictive maps for Aspen	14
3.12	Predictive maps for Carstens	15
3.13	Predictive maps for Chariot	15
3.14	Predictive maps for Gobblers	15
3.15	Predictive maps for Hathaway	16
3.16	Predictive maps for Mountain	16
3.17	Predictive maps for Pfeiffer	16
3.18	Predictive maps for Rim	17
3.19	Predictive maps for Sharp	17
3.20	Predictive maps for Bridge	17
3.21	Intersection between Predictive and Actual Maps (1)	18

3.22	Intersection between Predictive and Actual Maps (2)	18
3.23	Intersection between Predictive and Actual Maps (3)	19
3.24	Intersection between Predictive and Actual Maps (4)	19
3.25	Intersection between Predictive and Actual Maps (5)	19
A.1	Occurrence Matrix	44
A.2	Simple Matching and Rogers Theoretical Demonstration	44
A.3	Hamann Theoretical Demonstration (1)	45
A.4	Hamann Theoretical Demonstration (2)	45
A.5	Hamann Theoretical Demonstration (3)	45
A.6	Hamann Theoretical Demonstration (4)	46
A.7	Simpson Theoretical Demonstration	46
A.8	Braun, Kulczynski, Dice, Jaccard, Ochiai, and Sneath Theoretical Demonstration	47
A.9	Rao Theoretical Demonstration (1)	47
A.10	Rao Theoretical Demonstration (2)	48
A.11	Yule, Phi, and Kappa Theoretical Demonstration (1)	48
A.12	Yule, Phi, and Kappa Theoretical Demonstration (2)	49
A.13	Aspen fire, original observation region	49
A.14	Aspen fire, smallest observation region possible	50
A.15	Bridge fire, original observation region	50
A.16	Bridge fire, smallest observation region possible	51
A.17	Carstens fire, original observation region	51
A.18	Carstens fire, smallest observation region possible	52
A.19	Chariot fire, original observation region	52
A.20	Chariot fire, smallest observation region possible	53

A.21 Gobblers fire, original observation region	53
A.22 Gobblers fire, smallest observation region possible	54
A.23 Hathaway fire, original observation region	54
A.24 Hathaway fire, smallest observation region possible	55
A.25 Mountain fire, original observation region	55
A.26 Mountain fire, smallest observation region possible	56
A.27 Pfeiffer fire, original observation region	56
A.28 Pfeiffer fire, smallest observation region possible	57
A.29 Rim fire, original observation region	57
A.30 Rim fire, smallest observation region possible	58
A.31 Sharp fire, original observation region	58
A.32 Sharp fire, smallest observation region possible	59

LIST OF TABLES

2.1	Summary of wildfires in this project(Pimlott,Laird,&Brown)	5
2.2	Duration of each wildfires	9
A.1	Distance Metrics Results under original observation area	59
A.2	Distance Metrics Results under smallest possible observation area	60
A.3	Summary of Distance Metric Characteristics	60

ACKNOWLEDGMENTS

Chapters 2 and 3 are largely based off the following dissertation, with the permission of the author, whom I would like to thank: Hao, Y. (2018). *California wildfire spread prediction using FARSITE and the comparison with the actual wildfire maps using statistical methods* (Order No. 10790287). Available from Dissertations & Theses @ University of California; ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (2050095894). Retrieved from <https://search.proquest.com/docview/2050095894?accountid=14512>

I would also like to thank Rick Schoenberg for his patience with me as the head of my committee, as well as the immensely helpful guidance in structuring my research and ultimately this thesis.

Finally, I would like to thank Nicolas Christou and Hongquan Xu for providing extremely valuable advice in the writing of my thesis.

CHAPTER 1

Introduction

Wildfires, being immensely destructive phenomena that occur essentially every year, are something that authorities want to obviously manage. To give an idea of their damage, over the past ten years damages from wildfires have been estimated to be a total of \$5.1 billion in the United States alone. (Samanta, n.d.).

In order to actually control such fires so as to minimize loss of life and property, authorities need to be able to predict their trajectories and general behavior as accurately as possible (Alexander & Cruz, 2013). However, with many predictions models that already exist and more still being invented, which ones are the best to use?

Thus, there is a need for metrics that can evaluate the prediction performance of various models. Ideally there would also be different metrics for different usage scenarios - since there are divergent criteria for what makes a prediction “good” under various situations. This paper will attempt to examine fourteen distance metrics and illustrate how they can be applied to evaluating wildfire predictions, as well as providing some guidelines on what types of scenarios are suited for which metric.

The structure of the paper will be as follows. Chapter 2 introduces the wildfire data used in this investigation as well as the FARSITE software used to generate the predicted wildfire perimeters based on that input data. Chapter 3 presents the predictions generated by FARSITE as well as how they visually compare to the actual wildfire perimeters. Chapter 4 will provide an introduction to the various distance metrics that could potentially be used to determine how well a particular model does in predicting the actual fire spread and perimeter some specified time into the future. Chapter 4 will also discuss theoretical implications of each measure, describing from a theoretical standpoint each measure's pros and cons. In

Chapter 5, we apply these various distance metrics to actual data and examine numerically what the results are. In Chapter 6, we provide possible explanations on the results seen in Chapter 5, which also leads to further insights on the nature of each distance metric. Finally Chapter 7 concludes with why the study of these metrics is practically important, as well as other related avenues that warrant further investigation.

CHAPTER 2

Prediction Model Methodology

The FARSITE fire prediction software is used in this paper to generate wildfire perimeter predictions that will later be evaluated under various distance metrics. This section introduces FARSITE, its required inputs, and the methodology behind which perimeters are compared.

2.1 Introduction of FARSITE

FARSITE is a fire growth simulation modeling system that utilizes spatial information on topography and fuels, as well as weather and wind data. Its prediction model 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), and thus its outputs are designed to be viewed on such devices running GIS software (Finney, 1998). Data types such as GRASS and ARC/INFO GIS raster data are accepted as inputs by FARSITE.

Huygens principle of wave propagation was originally used to describe light wave behavior but has since been applied to explain expanding fire fronts (Finney, 1998). In its original application, Huygens' 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).

Huygens' principle only requires information from points on a fire's edge, allowing computers to save memory and to more quickly do the calculations in models that incorporate this principle. For example, in order to minimize the distortion in a prediction resulting from the grid in cellular models, the fire spread to unburned cells within a wide radius of each active cell must be calculated.

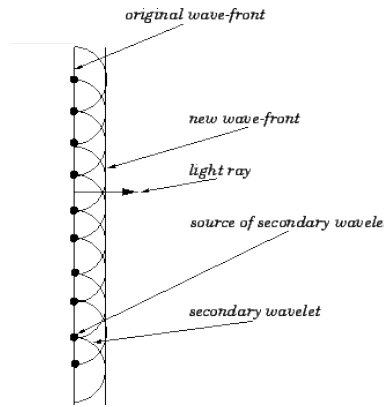


Figure 2.1: Huygens principle

For this paper, FARSITE version 3 is utilized instead of FARSITE version 4. This is because the previous version is able to process gridded temperature and relative humidity fields, which the data used has. Version 4 removed gridded temperature and relative humidity processing in exchange for better fuel moisture processing.

2.2 FARSITE Data Inputs

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 used the simulations in FARSITE to generate prediction

wildfires maps, and compared 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.

Table 2.1: 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

The wildfire spatial data to be used in generating predictions with FARSITE are sourced from LANDFIRE (LF), a program supported by the U.S. Department of Agriculture Forest Service. The LF program provides data such as geographic data, canopy data, fuel data, weather data, and actual wildfire perimeters which were utilized in this paper. The Forest Service is an agency of the USDA responsible for monitoring 154 national forests and 20 national grasslands (LANDFIRE Partners, 2018).

One of the components that FARSITE requires in order to generate predictions on wildfire spread is a landscape file, which contains information on vegetation, fuel, and topography. These landscape files can be generated using data files compatible with the Geographic Information System (GIS). Specifically, FARSITE requires five raster data themes, which are data on elevation, slope, aspect, fuels, and canopy. 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), and the same extent (the corners of the rectangular spatial region must be the same)(Data Requirements, 2018). Further optional data themes can also be added to the landscape profile to create an even more precise picture of the landscape. These data themes include information on: crown bulk density, crown base height, stand height, duff loading, and coarse woody. The resulting landscape file 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).

Besides the landscape file, other input data files are required for FARSITE to run. These are files on weather, wind, adjustment, and initial fuel moisture. The weather file is an ASCII text file containing daily observations of temperature, humidity, and precipitation. This data can be used to generate a temporal weather stream, which greatly simplifies actual variations in the weather. As for weather data, for this paper gridded weather inputs were used with the files being of the following format:

```
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

Above, TESTxx.TMP, TESTxx.HMD, TESTxx.PPT, etc. all refer to text files that contain grids (in ArcASCII format) of weather variables.

The wind file is also an ASCII text file, contained in a Wind (.WND) or as a gridded weather (.ATM) file as a stream of data. The wind files in this paper were of the latter kind. Winds are measured across both space and time. However, in FARSITE winds are assumed

to be constant in space for a given wind stream but variable in time. This means there are no topographic effects on winds.

The adjustment factor file is also a ASCII text file, and represents the rate of spread adjustment factors within the forecast. Users can use their judgment from experience or local data to fine-tune the simulation to better the observed 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. For this paper, we used an adjustment factor of 1.0.

The initial fuel moistures files are also ASCII text file, and must be set for each fuel type at the beginning of the simulation. This is because FARSITE needs initial fuel moistures in order to calculate site specific fuel moistures for the rest of the time steps in the simulation (Data Requirements, 2018). The fuel model utilized in this investigation is the Anderson fire behavior, which serves as an input to Rothermels mathematical surface fire behavior and spread model. The thirteen fuel model layers that are needed for the Anderson model represent 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. For this paper, the only optional input to be employed was information on gridded weather and winds .

2.3 Features of Data

Data from ten different wildfires were used for this investigation. These wildfires are: 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. Data on geography, canopy, fuel, weather, and actual perimeters as required by FARSITE simulation models

were all available for these wildfires.

The data for the wildfires originally came in the WGS84 coordinate reference system (CRS). However, the data was projected (transformed) from the WGS84 to the Albers CRS in R with the package Rgdal to make them consistent with the predictive wildfire map projection, which are of the latter CRS.

The duration of simulation in FARSITE varies depending on the case of each wildfire. However, the data collected for each wildfire is not perfectly complete. For example, while the Aspen fire started on July 22nd 2013, the data for that fire only exists from July 24th 2013. Another less than ideal aspect of the data is that the frequency of data collection is different between distinct days. For example, on the 24th staff recorded data on the Aspen fire once around 10 p.m., but he or she recorded data twice on the 25th. We chose to simulate and predict only over several hours or several days, because there is no need for the simulate the entirety of the actual fire for our purposes. There is also the fact that wildfires become more unpredictable over time, and so wildfire predictions based on simulations over a longer duration are less likely to be even somewhat representative of the actual fire perimeter by that time. Also, no corrections have been made to the data to account for the effect of firefighting activity on the wildfire, since there is no data that explicitly points out when this happens.

The simulation duration is chosen so as to produce predictions of a wildfire's perimeter of a time that directly matches up with an actual wildfire perimeter, so as to directly compare the two. For instance, one of the times the Aspen fire's perimeter was recorded is on July 24th, 22:11. We thus choose to run the simulation from July 23rd 1:00 (the earliest we have data from the Aspen fire) to July 24th, 22:00. We then compare the predicted Aspen fire perimeter on July 24th 22:00 to the actual on July 24th 22:11.

The following table depicts the simulation durations for each fire, as well as the actual fire perimeters used for comparison with the predicted perimeters.

Table 2.2: Duration of each wildfires

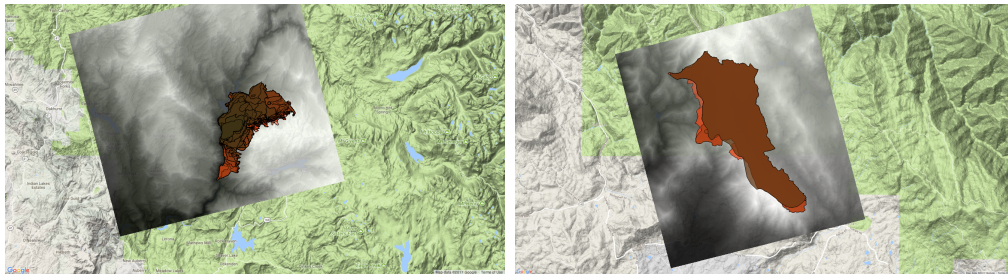
Year	Name	Start Date	Stop Date	Simulation Duration	Time Compared with Actual Fire
2013	Aspen	7.22 22:15	9.8 18:00	7.23 1:00-7.24 22:00	7.24 22:11
2013	Carstens	6.16 14:12	6.20 00:00	6.16 14:00-6.18 22:00	6.18 22:01
2013	Chariot	7.6 12:55	7.22 6:00	7.7 14:00-17:00	7.7 17:31
2013	Gobblers	8.20 13:32	8.30 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

CHAPTER 3

Prediction Model Results

3.1 Shapes of the Actual fire with Couple of Layers

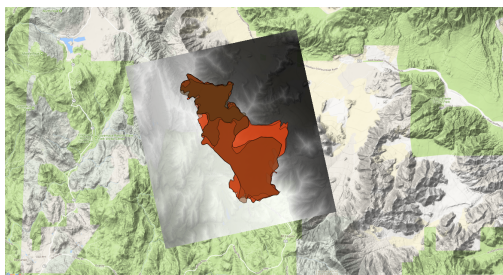
The perimeters of the wildfires in the data record the actual shape of the fire at different times. Therefore, if one were to regard each actual perimeter at a certain time as a layer, these layers can be drawn on an actual map of the surrounding area via QGIS to demonstrate the spread of the fire over time. These maps are as follows. Notably, the elevation data for the wildfires is fairly consistent with those of the actual map.



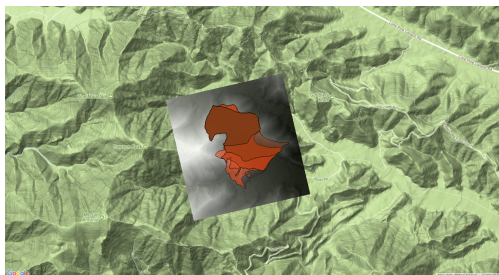
(a) Aspen

(b) Carstens

Figure 3.1: Actual maps with elevation (1)

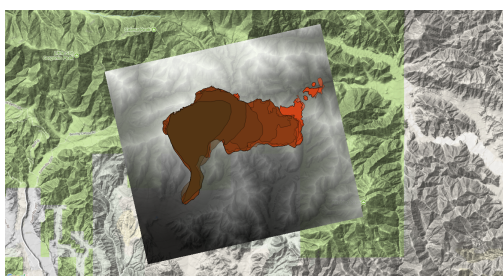


(a) Chariot

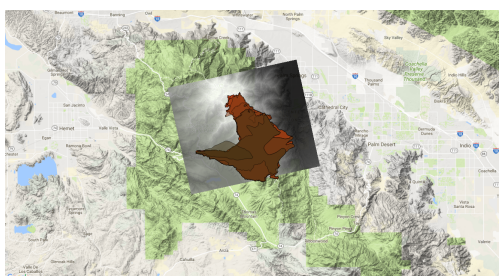


(b) Gobblers

Figure 3.2: Actual maps with elevation (2)

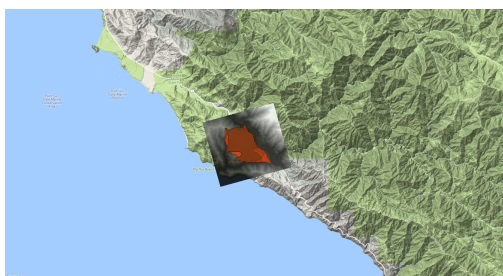


(a) Hathaway

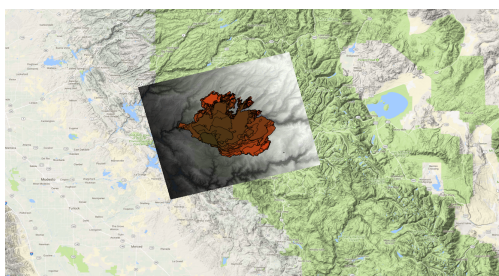


(b) Mountain

Figure 3.3: Actual maps with elevation (3)

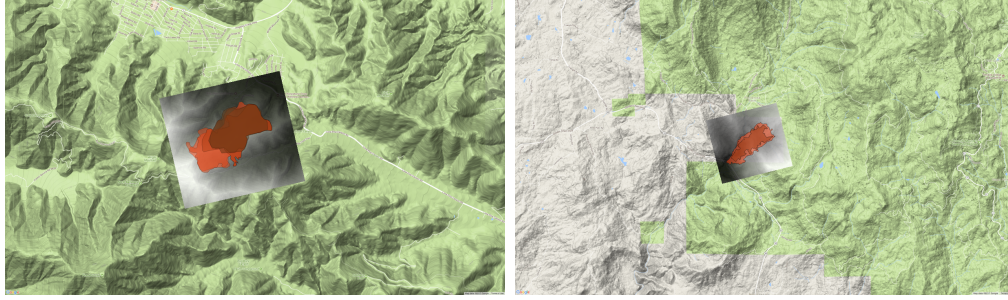


(a) Pfeiffer



(b) Rim

Figure 3.4: Actual maps with elevation (4)



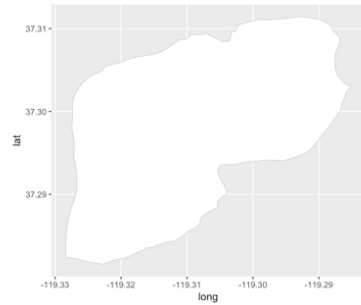
(a) Sharp

(b) Bridge

Figure 3.5: Actual maps with elevation (5)

3.2 Shapes of the Actual fire with Single Layer

The shapes of the actual fire perimeters used, represented as a single layer via R, are as follows:

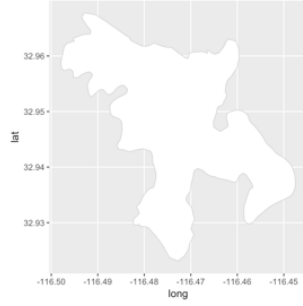


(a) Aspen

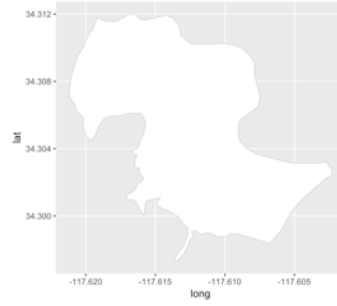


(b) Carstens

Figure 3.6: Actual maps with a single layer (1)

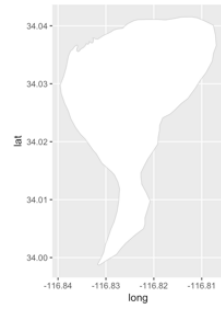


(a) Chariot

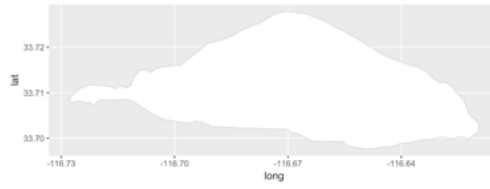


(b) Gobblers

Figure 3.7: Actual maps with a single layer (2)

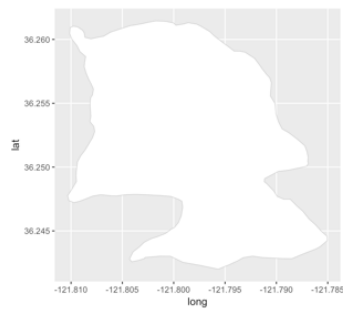


(a) Hathaway

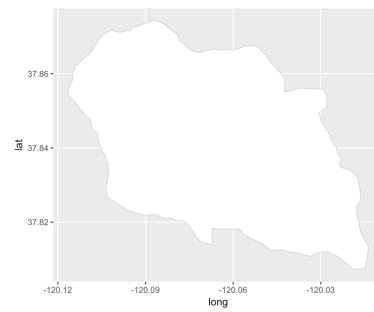


(b) Mountain

Figure 3.8: Actual maps with a single layer (3)

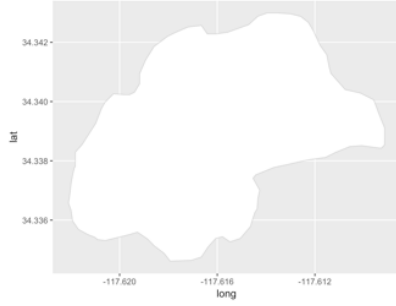


(a) Pfeiffer

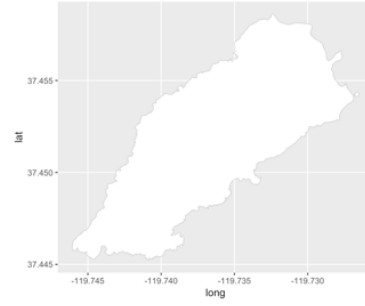


(b) Rim

Figure 3.9: Actual maps with a single layer (4)



(a) Sharp

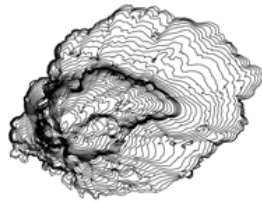


(b) Bridge

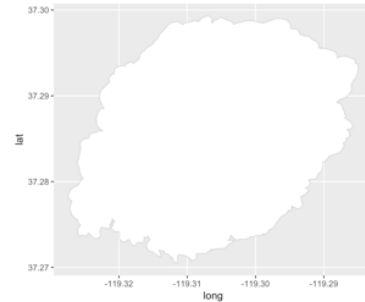
Figure 3.10: Actual maps with a single layer (5)

3.3 Shapes of the Predictive Fire

The predicted wildfire perimeters generated by FARSITE come in the form of Shapefiles, which contains information on their shape. The spatial data frame of these Shapefiles contain multiple polygons coming from the simulation over time and thus several simulated perimeters over that time. After reading the Shapefiles into R, the intersecting polygons as a raw result of the predictions need to be combined, and then the most outside shape is kept. The shapes of predictive fires are as the following.

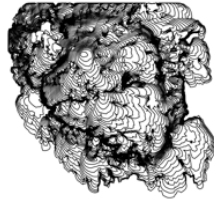


(a) With visible simulations

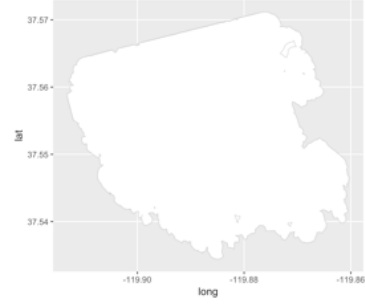


(b) Without visible simulations

Figure 3.11: Predictive maps for Aspen



(a) With visible simulations

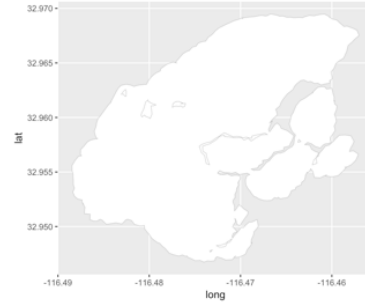


(b) Without visible simulations

Figure 3.12: Predictive maps for Carstens

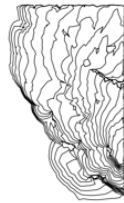


(a) With visible simulations



(b) Without visible simulations

Figure 3.13: Predictive maps for Chariot

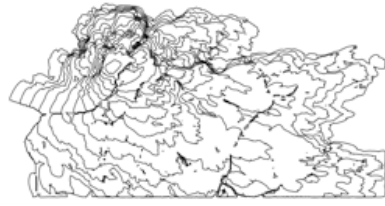


(a) With visible simulations

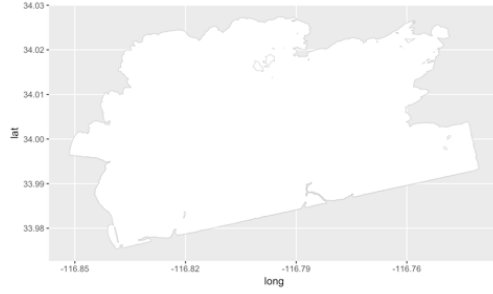


(b) Without visible simulations

Figure 3.14: Predictive maps for Gobblers

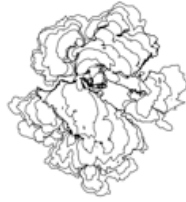


(a) With visible simulations

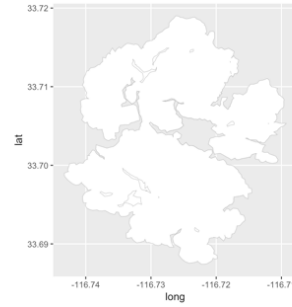


(b) Without visible simulations

Figure 3.15: Predictive maps for Hathaway

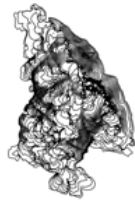


(a) With visible simulations



(b) Without visible simulations

Figure 3.16: Predictive maps for Mountain

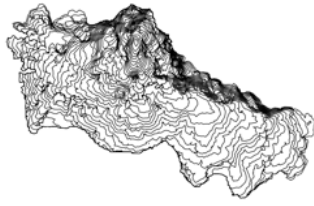


(a) With visible simulations

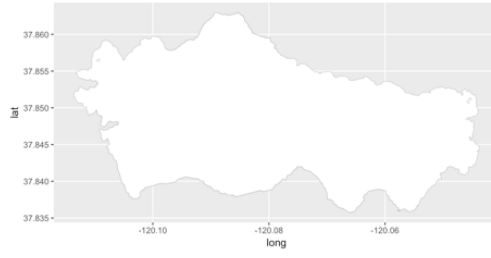


(b) Without visible simulations

Figure 3.17: Predictive maps for Pfeiffer



(a) With visible simulations

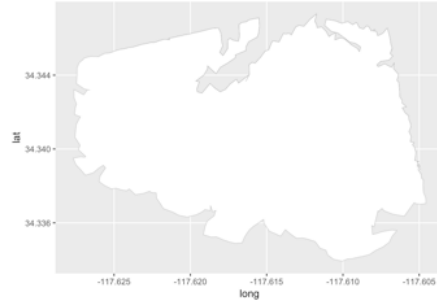


(b) Without visible simulations

Figure 3.18: Predictive maps for Rim

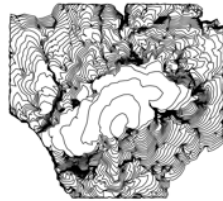


(a) With visible simulations

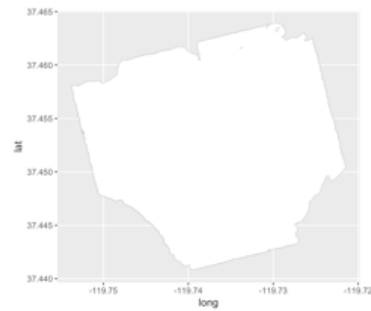


(b) Without visible simulations

Figure 3.19: Predictive maps for Sharp



(a) With visible simulations

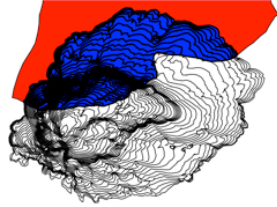


(b) Without visible simulations

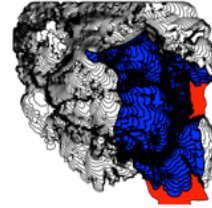
Figure 3.20: Predictive maps for Bridge

3.4 Shapes of the Intersection between Predictive Fire and Actual Fire

Below, the red area in each map represents the wildfire's perimeter at the specific time given in table 2.2. The blue area is the intersection of the predicted and actual wildfire perimeters.

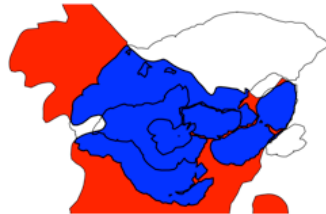


(a) Aspen



(b) Carstens

Figure 3.21: Intersection between Predictive and Actual Maps (1)

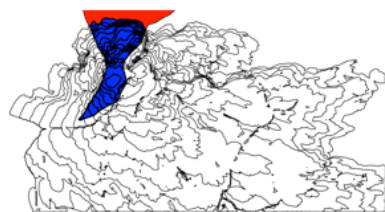


(a) Chariot

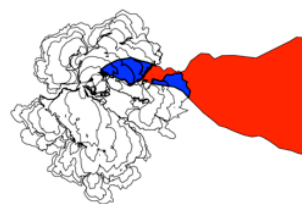


(b) Gobblers

Figure 3.22: Intersection between Predictive and Actual Maps (2)

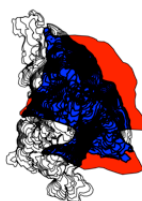


(a) Hathaway

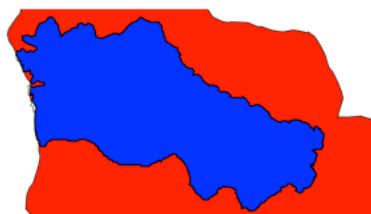


(b) Mountain

Figure 3.23: Intersection between Predictive and Actual Maps (3)

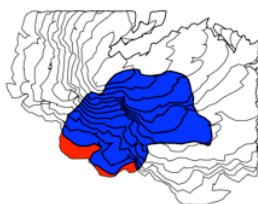


(a) Pfeiffer

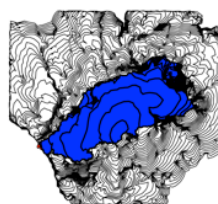


(b) Rim

Figure 3.24: Intersection between Predictive and Actual Maps (4)



(a) Sharp



(b) Bridge

Figure 3.25: Intersection between Predictive and Actual Maps (5)

CHAPTER 4

Introduction to Distance Metrics

The fourteen distance metrics in this paper used to characterize the performance of the prediction model in a particular wildfire are all based on looking at where the predicted and actual fire perimeters overlap and do not overlap in a given observation region. As shown in Figure A.1, a refers to the amount of area that both the predicted and actual fire perimeters say will burn, while b specifies the amount of area in which the predicted fire perimeter says will burn but in the actual fire perimeter does not burn. The letter c states the amount of area in which the predicted fire perimeter says will not burn but burned in the actual fire perimeter, while d is the measure of how much area both the predicted and actual fire perimeters say will not burn. Another way to sum up these variables is to say that a represent mutual presence, b and c represent mismatches, and d represents mutual absences.

Other notes to mention are that: most measures do not take into account rotational differences between the predicted and actual fire perimeter. For example, if the predicted fire perimeter is exactly a ninety-degree rotation of the actual fire perimeter, most of the distance metrics will simply not take that into consideration and the prediction will just fare badly in terms of that metric. Also, these distance metrics do not inherently place more importance on particular areas. For instance, if one wants to correctly predict whether a fire will spread to a town and especially punish models that do not predict this movement (so the authorities can evacuate people from that town as needed), further modifications to the distance metrics would be needed. A final note is that these metrics range from a value of 0 to a value of 1, unless otherwise noted.

4.1 Simple Matching, Rogers, Hamann

The *Simple Matching* coefficient (Sokal and Michener 1958, Rand 1971) is of the form:

$$\frac{a + d}{a + b + c + d} \quad (4.1)$$

is the sum of all overlapping areas divided by the total area in the observation region. The simple matching metric is intuitive and takes into account both matches of burning and non-burning areas. However, the metric depends heavily on the size of the bounded prediction region of interest, and in particular, even if the predicted and actual fire perimeters are wildly different, the simple matching distance will be low provided the predicted and actual fire perimeters are small relative to the observation region. An illustration is in Figure A.2, where the prediction is totally incorrect but the simple matching metric is low because d is much larger than a , b , and c .

The *Rogers* distance metric (Rogers and Tanimoto 1960) has the form:

$$\frac{a + d}{a + 2b + 2c + d} \quad (4.2)$$

which is quite similar to that of simple matching. Thus, this metric logically has all of the pros of the simple matching metric, but doubling the weight of mismatches somewhat mitigates the effect of large d . This also makes it harder for any model's prediction to do well in general. However, like simple matching, if d is much bigger than the sum of $a + b + c$, then the effect of d dominates. An example of this would be Figure A.2.

The *Hamann* coefficient (Hamann 1961) has the following form:

$$\frac{a + d - (b + c)}{a + b + c + d} \quad (4.3)$$

The Hamann metric includes a penalty for mismatches that somewhat lessens the effect of having large d , and the coefficient ranges from -1 to 1 allowing one to easily tell if there are

more matches or mismatches. However, d is included in the metric, so it is still dependent on the size of the observation region. Also, the mismatches penalty may lead to some strange results. For instance, the Hamann coefficient would indicate that the prediction in figure A.4 does much better than the prediction in figure A.3, since in figure A.4 d is much higher than b or c while in figure A.3 $b + c$ is much greater than $a + d$. But in A.4 the prediction fails to predict that the fire will spread to the town, while in figure A.3 the prediction at least does foresee this, even if it is not a good prediction overall.

In figures A.5 and A.6, assume that the total area of the square is 100. In figure A.5, assume that $a = 25$, $b = 0$, $c = 25$, and $d = 50$. For figure A.6, assume that $a = 50$, $b = 25$, $c = 0$, $d = 25$. If so, the predictions in figures A.5 and A.6 do equally as well under Hamann. However, one may argue that the A.6 prediction should do better than the A.5 prediction, since in figure A.6 the prediction actually does get all of the actual area burned, unlike in A.5.

4.2 Simpson, Braun, Kulczynski

The *Simpson* coefficient (Simpson, 1943) is characterized as:

$$\frac{a}{\min(a + b, a + c)} \quad (4.4)$$

which is the proportion of matches as a fraction of the lesser of $a + b$ (the total burned area in the prediction), or $a + c$ (the total burned area in the actual map). It is not influenced by the amount of area of mutual absences and thus the size of the observation region. However, if the discrepancy between $a + b$ and $a + c$ is large, then this measure may actually make a prediction look better than it actually is. For instance, in figure A.7 the Simpson coefficient here would be calculated as the red area over the blue area and not the red area over the green area - making the prediction performance look much better than it really is. Essentially, the Simpson coefficient fails to punish overburn, which is when the actual burned area is far more than what the predicted burn area is.

The *Braun* distance metric (Braun-Blanquet, 1932) has a form of:

$$\frac{a}{\max(a+b, a+c)} \quad (4.5)$$

Being like a more strict (harder to do well under) version of the Simpson coefficient), the pros of the Braun coefficient are quite similar. However, like with Simpson, if in Braun $a+b$ and $a+c$ are very different, the distance metric may give a misleading picture of the prediction's performance. In figure A.8, the prediction exhibits much underburn - the phenomenon of the actual fire perimeter only being a smaller area inside of the predicted fire perimeter. Under the Braun coefficient this is not a good prediction, reflecting the coefficient's tendency to punish underburn. But a prediction with underburn may be better than a prediction that does not fully cover the actual fire perimeter. For instance, it may be better to evacuate towns that do not actually end up being destroyed by the fire than not evacuating towns that actually do get consumed by the fire.

The Kulczynski measure (Kulczynski, 1927) is of the form:

$$\frac{1}{2} \left[\frac{a}{a+b} + \frac{a}{a+c} \right] \quad (4.6)$$

and is the arithmetic mean of the Simpson and Braun coefficients (Warrens 2008). The Kulczynski measure shares the same pros as those of Simpson and Braun, but also with extra advantages. For instance, a prediction with overburn is limited in how well it does in this measure (overburn benefits only from either $\frac{a}{a+b}$ or $\frac{a}{a+c}$, whichever is larger). In generally a model has to limit mismatches as much as possible to get a decent Kulczynski score. A potential con also stems from this weighted average structure in that it also punishes underburn (which benefits only from either $\frac{a}{a+b}$ or $\frac{a}{a+c}$), which may be undesirable as discussed before. An example is figure A.8 ($\frac{a}{a+c}$ is high, but $\frac{a}{a+b}$ is low), in which the Kulczynski score is unlikely to be much greater than $\frac{1}{2}$ here.

4.3 Dice, Jaccard, Ochiai

The *Dice* coefficient (Dice 1945, Sorenson 1948, Gleason 1920) is written as:

$$\frac{2a}{2a + b + c} \quad (4.7)$$

This is the harmonic mean of the Simpson and Braun coefficients (Warrens, 2008). Dice is not influenced by d , and may be useful in scenarios in which there are low number of matches relative to mismatches. Specifically, if b and c are high relative to a for several different models' predictions, without the double weighting on a we would find that the coefficient would be close to zero for most predictions - making it difficult to distinguish between them. However, if the amount of matches is high relative to mismatches, giving greater weight to a will make it harder to distinguish between the predictions then. Also, Dice punishes underburn (refer again to figure A.8), which may not be wanted.

The *Jaccard* distance metric (Jaccard, 1912) has the following form:

$$\frac{a}{a + b + c} \quad (4.8)$$

Jaccard shares much of the same pros and cons as Dice. However, because it weighs a less than in Dice, it is less suitable for situations in which $b + c$ is much higher than a , and more suitable for situations in which a is much greater than $b + c$. Also, predictions which exhibit underburn will do worse under Jaccard than under Dice. For example, in figure A.8 let $a = 5$, $b = 10$, and $c = 0$. Here, the Dice coefficient would be $\frac{10}{25} = \frac{2}{5}$, while the Jaccard coefficient would be $\frac{5}{15} = \frac{1}{3}$.

The Ochiai coefficient (Ochiai 1957, Driver and Kroeber 1932) is calculated as follows:

$$\frac{a}{\sqrt{a + b}\sqrt{a + c}} \quad (4.9)$$

This is actually the binary version of cosine similarity, which measures similarity in terms of orientation between the predicted and actual fire perimeters (Warrens, 2008). Also, like with Kulczynski it theoretically will do well in separating the good and bad predictions models due to the coefficient's "toughness." However, this metric punishes underburn which may be undesirable, even more so than with Kulczynski due to the denominator being a product rather than a sum. But Ochiai may be higher than Dice and Jaccard in an identical situation. An example is again figure A.8 letting $a = 5$, $b = 15$, $c = 0$, then Kulczynski would be $\frac{5}{8}$ while Ochiai would be $\frac{1}{2}$. In the same scenario, Dice would be $\frac{2}{5}$ and Jaccard would be $\frac{1}{3}$.

4.4 Sneath, Rao, Yule

The *Sneath* distance metric(Sokal and Sneath, 1963) takes on the following form:

$$\frac{a}{a + 2b + 2c} \tag{4.10}$$

Compared to Jaccard, b and c are doubly weighted - making it tougher for a prediction to do well in this coefficient than in Dice and Jaccard. Thus, Sneath has the potential to better illustrate to distinguish the truly good and bad predictions. Sneath is also useful in cases in which $b + c$ is much greater than a , as the extra weight put on mismatches ensures the performance of different predictions do not all just end up being 1. Sneath is not affected by d . However, under Sneath model predictions with underburn are hit particularly hard, and using this measure may not be useful $b + c$ is much greater than a . For instance, in figure A.8 again assuming $a = 5$, $b = 15$, $c = 0$, Sneath is just $\frac{1}{7}$.

The *Rao* coefficient (Rusel and Rao, 1940) is written as:

$$\frac{a}{a + b + c + d} \tag{4.11}$$

This is similar to Simple Matching, except the numerator is only a (occurrence matches)

instead of $a + d$ (matches of any kind). While Rao is easy to understand and simple to implement, any prediction will do badly if d is much greater than a . For instance, if in figure A.9 $a = 4$, $b = c = 0$ and $d = 96$, then Rao is $\frac{1}{25}$. In figure A.10, assuming $a = 75$, $b = c = 0$ and $d = 25$, Rao is $\frac{3}{4}$. The Rao coefficients are totally different despite both predictions totally matching the actual perimeters.

The *Yule* distance metric (Yule, 1900) is of the following form:

$$\frac{ad - bc}{ad + bc} \quad (4.12)$$

Also known as Yule's Q, this coefficient measures the strength of association between the predicted and the actual fire perimeters, ranging from -1 to 1 (Yule, 1912). Two fire perimeters are positively associated if ad is greater than bc , resulting in a positive value for this measure, and vice versa. This means that Yule takes into account similarity in direction between the predicted and actual fire perimeters and if they are (near) rotations of each other. However, Yule more directly measures the association between two fire perimeters rather than the performance of the prediction. For instance, in figure A.11 assume $a = 90$, $b = 5$, $c = 5$, and $d = 0$. Here, Yule is -1 , demonstrating how the prediction is the “opposite” of what really happens in the actual perimeter but does not give much indication of the overall performance of the model. Thus, Yule may be better when used with other measures that give a more direct assessment of a prediction's performance. Also, under Yule: it is dependent on the size of observation region, and it is easy to get a zero denominator and thus render the coefficient unusable (an example is figure A.12, assuming $a = 95$, $b = d = 0$, $c = 5$).

4.5 Phi, Kappa

The *Phi* (Yule, 1912) is written as follows:

$$\frac{ad - bc}{\sqrt{(a + b)(a + c)(d + b)(d + c)}} \quad (4.13)$$

Phi is actually the Pearson correlation coefficient in a binary scenario (Zysno, 1997). Thus, this metric measures the association (positive or negative) between the predicted and actual fire perimeters, like with Yule. However, compared to Yule, Phi gives less strongly associated results. For instance, in figure A.11 assuming $a = 90$, $b = 5$, $c = 5$, and $d = 0$, Phi is approximately -0.05 while Yule coefficient is -1 . Drawbacks to Phi include: relying on a high d value to get a highly positive score, being influenced by the size of the observation region, and being relatively easy to get zero in the denominator (see figure A.12, again assuming $a = 95$, $b = d = 0$, $c = 5$).

The *Kappa* distance metric (Cohen, 1960) can be written as:

$$\left(1 + \frac{(b + c)(a + b + c + d)}{2ad - 2bc}\right)^{-1} \quad (4.14)$$

Kappa measures how well two perimeters overlap after taking into account “agreement that would be expected purely by chance.” (Sim and Wright, 2005, p. 258) The coefficient ranges from -1 (if all mismatches) to 1 (if all areas match). If there are more mismatches than matches, then the denominator $2ad - 2bc$ is negative and drags down Kappa, potentially below zero depending on the numerator (Sim and Wright, 2005). Like the Yule and Phi coefficients, Kappa describes the association between two perimeters. However, with Kappa it is easy to get a zero in the fraction (if $ad = bc$; or $ad = 0$ and/or $bc = 0$, such as in figure A.12 assuming $a = 95$, $b = d = 0$, $c = 5$) which leads to an undefined coefficient, and that it is not an easily interpretable measure.

CHAPTER 5

Application of Distance Metrics to Wildfire Data

We now turn our discussion to the applying these various distance matrices to the wildfire data previously introduced. To briefly recap, ten different wildfires with the necessary data for predictions are examined. The FARSITE program is used to predict what the perimeter of the wildfire will look like at a given time, which is compared to the actual wildfire perimeter at the same specified time. Each of the fourteen distance metrics discussed earlier are then calculated for each fire, based on both the original observation region and the smallest possible square/rectangular boundaries that encompass both the predicted and actual perimeters, as this does noticeably affect the results of several distance metrics.

All actual and predicted wildfire maps were analyzed (including calculating the distance metrics) with the R statistical software, specifically using the packages: *rgdal*, *shapefiles*, *rgeos*, *raster*, *sp*, and *maptools*. Visual maps that provide an alternate view of the predicted and actual perimeters of each fire as well as their overlapping areas can be found in figures A.13-A.32 of the appendix.

In this chapter, we will first examine the performance numerically of the ten wildfire predictions using the various distance metrics under the original observation region, and then examine how the results under the various distance metrics are affected if looking at the smallest possible boundary that encompasses both the predicted and actual fire perimeters. Tables A.1 and A.2 contains these exact numerical values.

5.1 Results Under Original Observation Region

5.1.1 Simple Matching, Rogers, Hamann

Looking at the Simple Matching numerical results of the predictions under the original observation region, several observations can be made. Any wildfire predictions with high d (such as in the Aspen, Carstens, Chariot, Gobblers, Hathaway, Mountain, and Rim fires), actually do quite well here with values very close to 1, implying a near-perfect match. Also, notably the Rim prediction does quite well despite it suffering from overburn, and cases of underburn like with Bridge and Sharp are heavily punished.

Since the Rogers coefficient is quite similar to Simple Matching, the trends in results exhibited under the former are similar to the latter. Numerically, most of the actual results under Rogers drop compared to the Simple Matching counterparts, but the ranking in terms of prediction performance of each fire remains the same.

Under Hamann, we find that numerically the Aspen, Rim, Chariot, and Mountain predictions do quite well even though they visually do not look the best. Also, despite Pfeiffer looking like the best prediction it does worse than the aforementioned four fires. The Hathaway and Gobbler predictions do much worse here than under Rogers and Simple Matching. The Bridge and Rim predictions (particularly the former) also suffer heavily compared to the Simple Matching and Rogers cases, with Bridge now having a negative result.

5.1.2 Simpson, Braun, Kulczynski

Under the Simpson coefficient, the Pfeiffer prediction does the best, matching intuition unlike the other distance metrics thus far. Also of note are the particularly good performance of the Bridge and Sharp predictions, which have a large amount of underburn, with values quite close to 1. Similarly, despite much overburn, the Chariot and Rim predictions also have Simpson coefficients somewhat close to 1.

The Braun coefficient produces quite different results compared to Simpson. Most of the fire predictions have a much lower Braun coefficient than Simpson coefficient. Especially

large drop-offs are present for Bridge and Sharp (due to underburn) as well as for Chariot and Rim (due to overburn). As a result, Pfeiffer by far does the best here, as it should, unlike under Simpson. Also, no longer are predictions such as Bridge, Sharp, Rim, and Chariot categorized as being much better than that of something like Carstens.

The Kulczynski measure at first glance seems to produce numerical results that match up with intuition. For instance, the Pfeiffer prediction does the best here. The Gobbler and Hathaway predictions do badly because their predictions are just not that good. However, worth mentioning is that with Kulczynski, the Rim prediction does extremely well relative to the other predictions (second best here). This is undesirable as there is much overburn in the prediction.

5.1.3 Dice, Jaccard, Ochiai

Under the Dice measure, the Gobbler, Hathaway, and Mountain fire predictions do poorly due to them not lining up with the actual. Pfeiffer is the best-performing prediction here by far, matching intuition. Cases like the Bridge and Sharp are hurt under due to large underburn. Also notable is that the Chariot and Rim fires, despite much overburn, do well above average.

Compared to Dice, the empirical Jaccard results are smaller but the performance of the predictions rank in the same exact order. Thus, at least for these set of results, the Dice and Jaccard metrics behave very similarly.

Under Ochiai, for some predictions like Aspen, Chariot, Gobbler, Hathaway, Mountain and Pfeiffer, the actual score falls little compared to Kulczynski. However, for predictions with massive underburn (Bridge and Sharp) or overburn (Rim), their Ochiai scores decrease significantly compared to Kulczynski which illustrates just how much better the Pfeiffer prediction is. Despite these changes, how the predictions rank relative to each other are the same as under Kulczynski, except for Carstens now ranking higher than Bridge.

5.1.4 Sneath, Rao, Yule

Empirically, the scores of the predictions' performance under the Sneath coefficient are much lower than in Dice and Jaccard due to the former's form. Otherwise, how the predictions rank relative to each other and how far apart they are in terms of the actual coefficient are similar to Jaccard and Dice.

The numerical results for the Rao are significantly different from that of Simple Matching, despite these measures' similar forms. Predictions with high d such as Aspen, Chariot, Gobbler, Hathaway, and Mountain do not do well in Rao unlike with Simple Matching. However, predictions that visually look good such as Pfeiffer regardless of observational area do fairly well under this measure. Cases like Gobbler and Hathaway, which intuitively do not look like good predictions, predictably do among the worst.

Looking at the scores using Yule (which measures the correlation between the predicted and actual perimeters), the highly negative scores of the Gobbler and Hathaway predictions correctly reflect how the predicted and actual fire perimeters tend to span different locations and directions. Predictions with much underburn also suffer under this measure due to the prediction spanning areas that the actual fire perimeter does not cover. It also means that something like Pfeiffer gets a highly positive score. However, predictions with much underburn (such as Bridge and Sharp) receive highly positive Yule scores. This is not technically incorrect because the prediction does tend to be in the same area and direction as the actual fire perimeter, even if the prediction only spans a small portion within.

5.1.5 Phi, Kappa

The Phi coefficient seems to behave similarly to Yule, at least in the empirical results calculated under the original observation region. The predictions all rank relative to each other in the same order, and gaps of the predictions' scores relative to one other seem to be similar percentage-wise. However, as stated in the theoretical analysis, the Phi coefficient tends to produce numbers further away from -1 or 1, giving less strongly correlated results.

The Kappa distance metric gives numbers similar to those under Phi. However, notably

the Rim and Sharp Kappa scores fall moderately compared to their Phi scores, and the Bridge Kappa score is much less than its Phi score.

5.2 Results Under Smallest Possible Observation Region

5.2.1 Simple Matching, Rogers, Hamann

We now switch our discussion to how well the predictions do under each measure if the smallest rectangle possible that covers both the predicted and actual fire perimeters is used as the boundaries, rather than the entire original observation region. Note that since the Simpson, Braun, Kulczynski, Dice, Jaccard, Ochiai, and Sneath coefficients do not depend on the size of the boundary box, the results discussed previously for those measures also apply here.

For Simple Matching, the smaller bounding area does mostly curb coefficient inflation for predictions with high d , such as Aspen and Mountain. Predictions for which the size of the observation region did not change much after using the smallest bounding rectangle, such as Bridge, did not see significant changes to their Simple Matching score.

Under the smallest bounding rectangle, the Rogers scores are much like the Simple Matching scores, only decreased. Thus, a smaller observation area affects Rogers scores much like with Simple Matching scores.

Hamann scores drop massively under the smallest possible boundary rectangle compared to under the original observation region, resulting in a more accurate depiction of the predictions' effectiveness. For instance, the Pfeiffer prediction is now by far the best performing and Aspen is now no longer among the “best” predictions and instead is among the lower-performing ones. Predictions with much underburn (Bridge and Sharp) are also heavily penalized, but interestingly the predictions with much overburn (Chariot and Rim) are among the better performing ones despite the massive score drop compared to under the original observation region. Also of note is that the Mountain prediction, which intuitively is the worst one, is actually middle-of-the-pack in terms of Hamann.

5.2.2 Rao, Yule

Rao now much more accurately illustrates the predictions' performance. For instance, compared to when using the original observation area, the Pfeiffer prediction now becomes the best performing one by far, while predictions like Aspen, Gobbler, Hathaway, and Mountain now are rated significantly worse than the other predictions. Interestingly, here predictions with underburn (Bridge and Sharp) are less penalized than predictions with overburn (Chariot and Rim), which may be desirable and is similar to the case under the original observation region.

As stated before, Yule measures how the predicted and actual perimeters are correlated rather than the prediction' performance. For instance, predictions that match up well with the actual (such as Sharp) are ranked higher, while predictions that obviously do not look good (such as Hathaway, Mountain, and Gobbler) get more negative scores. However, there are oddities in this scenario such as the score of Pfeiffer dropping noticeably compared to under the original observation area and now having a lower score than Sharp. Also, like under the original observation area, predictions with overburn (Chariot and Rim) continue to receive highly positive Yule scores.

5.2.3 Phi, Kappa

The Phi scores all either drop or stay around the same here compared to under the original observation region, which results in different rankings of the predictions relative to each other. Notably, the Pfeiffer prediction performs best here unlike with Yule. Also, while predictions with much underburn (Bridge and Rim) are among the better performing ones, they are still quite far away from the best predictions like Pfeiffer which is desirable. However, even under a reduced observation area, predictions with overburn (Chariot and Rim) receive highly positive scores again, while not necessarily incorrect may be misleading and suggests that the Phi coefficient is best used with a measure that more directly measures the predictions' accuracy.

Scores for Kappa, under the reduced observation area, compared to under the original

observation area are either around the same or somewhat decreased. Interestingly, scores for Kappa compared to Phi are quite similar for many predictions (Aspen, Carstens, Chariot, Chariot, Gobbler, Mountain, and Pfeiffer). However, both predictions with heavy overburn (Rim) or heavy underburn (Bridge and Sharp) seem to be more heavily penalized than with the Yule and Phi coefficients.

CHAPTER 6

Discussion of Results of Application of Distance Metrics

While the previous chapter presented the numerical distance metric results, this chapter will attempt to explain why those were the results to bring to light particular characteristics of the metrics not immediately obvious just by thinking about the theoretical implications. A summary of the characteristics of each distance metric, as discussed in this chapter and the previous chapter, can be found in table A.3 in the appendix.

6.1 Simple Matching

For the Simple Matching coefficient under the original observational area, of particular note is the prediction for the Mountain wildfire, which is obviously not a good prediction but its performance under Simple Matching is above average compared to the others, due to its high d . Actually, the Pfeiffer prediction does not do the best here, most likely due to its low amount of d . As for scenarios of underburn such as the Bridge and Sharp predictions, those get low scores under Simple Matching simply because one can hardly declare a prediction having done particularly well if it simply covers every possible outcome. However, the Bridge and Sharp predictions do look unnecessarily incompetent because the coefficient inflation induced by the size of the original operating area.

Under the smallest possible observation region, Simple Matching coefficient inflation due to high d (such as for Aspen) has mostly been curbed. However, depending on how the predicted and actual fire perimeters are oriented relative to each other, there still may exist high d . Thus, it is possible to end up with situations like the Mountain prediction, which

still has a somewhat inflated score (and does better than some of the other predictions even though it should not).

6.2 Rogers

The Rogers score of Aspen and Rim mostly remain the same compared to Simple Matching because of high d compared to a , b , and c . Thus, the seemingly great performance of the Aspen and Rim predictions (visually obviously not the case) relative to the other predictions is even more exaggerated under this metric than with Simple Matching. Of course, under the reduced observation region, this behavior is greatly reduced to more correctly reflect the performance of Aspen and Rim due to the drastic decrease of d in many of the predictions.

6.3 Hamann

With the Hamann measure under the original observation area, the Aspen, Rim, Chariot, and Mountain predictions do well presumably due to the large d . Also due to that reason, Pfeiffer gets a lesser score than the aforementioned four fires, which should not be the case. Hathaway and Gobbler correctly end up looking much worse here compared to the other predictions than under Rogers and Simple Matching, mainly because the prediction and actual perimeter directions do not line up. The Bridge and Rim predictions (Bridge especially so) do much worse relative to the other predictions because of massive underburn, which Hamann seems to heavily punish.

Under the reduced observation region, predictions with much underburn are heavily penalized, presumably due to reduced d but with no change to b and c . These same changes also seem to allow predictions with much overburn in this data set to perform relatively better. Also, like with the original observation area, the Mountain prediction actually does average among the predictions despite looking like the worst prediction. This is most likely due to the large d that exists even after shrinking the observation area. Overall, the Hamann coefficient seems to be much more useful in accurately portraying the actual performance of

predictions when considering a reduced observation area, but is still not without its quirks.

6.4 Simpson

Under Simpson, both underburn (Bridge and Sharp predictions) and overburn (Chariot and Rim) are not heavily punished. This occurs because of the massive discrepancy between the prediction's total area $a + b$ and the actual perimeter total area $a + c$. As long as the predicted and actual fire perimeters intersect “enough” (a is high enough), the denominator in this measure is based on the smaller of $a + b$ and $a + c$ thus giving a very good result for the prediction. Because of this tendency, for example Chariot does about as well as Carstens, despite the latter looking like a better prediction. Here, the coefficient is calculated by taking the blue area over the relatively small green area, which does not give the true picture of the massive overburn represented by the red area.

6.5 Braun

As noted earlier, for Braun predictions with severe underburn (Sharp and Bridge) or overburn (Chariot and Rim) are punished for their lack of precision. Predictions for which the predicted area and actual fire perimeter are not really in the same place or direction are also punished under this measure, such as with Hathaway and Mountain. Overall, it seems that only great predictions will do well under the Braun coefficient, which overall is favorable for correctly evaluating model predictions but it does at the same time also heavily punish predictions with much underburn, which may not be desirable.

6.6 Kulczynski

Under Kulczynski, the Rim prediction despite much overburn does better than all of the others except Pfeiffer. It seems that empirically, only giving half weight to $\frac{a}{a+b}$ is enough to drive up the value of this coefficient when there is just “enough” overlap between the

predicted and actual fire perimeters. However, predictions with underburn (Carstens and Sharp) do okay, if not excellent relative to the other predictions, unlike in other measures such as Hamann and Braun. Overall, the Kulczynski measure does seem to do a fairly good job at properly recognizing what predictions are good, but it surprisingly seems to not appropriately punish certain cases of overburn like with Simpson.

6.7 Dice, Jaccard, Sneath

Like mentioned earlier, under Dice there are cases which do match up with intuition, such as Pfeiffer being the best prediction and Mountain being by far the worst. However, for cases like Bridge and Sharp with much underburn, the actual score will be low, simply because $b+c$ is so high relative to a , even when doubly weighted. The Chariot and Rim fires do fairly well under this measure, despite much overburn, most likely because the doubly weighted a is seemingly more than enough to offset $b+c$ - this may be of concern. Since the Jaccard and Sneath metrics are quite similar in form to the Dice, much of the insights here on Dice also apply to the other two as the different weights do not seem to affect the relative performance ranking of the predictions.

6.8 Ochiai

Under the Ochiai coefficient, the relative performance of the predictions seemed to match up fairly well with intuition. For instance, the Pfeiffer prediction does much better than any of the predictions with significant underburn or overburn. Predictions like Gobbler, Hathaway, and Mountain are appropriately deemed to be much worse than a fairly good prediction like Carstens. Thus, the Ochiai coefficient seems to keep the strengths of Kulczynski while punishing overburn, underburn, and general inaccuracies more heavily. The Ochiai coefficient, at least going by the empirical results here, seems to be a good candidate for a measure to distinguish between the exceptional and poor predictions.

6.9 Rao

For Rao with the original observation region, predictions with high d tend to do poorly due to the lack of d in the numerator. This seems to indicate to get a high Rao score, a prediction has to match the actual fire perimeter well. This is borne out with the Pfeiffer, Gobbler, and Hathaway predictions (the latter two miss the mark in terms of both similar size and direction between the predicted and actual perimeters). However, strange results can manifest under this measure. One notable example is Bridge getting the highest Rao result, presumably due to very low c and d . Rao does not seem to take into account if the shapes of the predicted and actual perimeters are similar to each other. Also, empirically under Rao underburn is less punished than overburn, as the predictions with much underburn (Bridge, Sharp) performed better than the predictions with much overburn (Chariot, Rim), unlike the other metrics.

Under the reduced observational area with the Rao, overburn is still more heavily punished than underburn. For this data set, this is most likely because the underburn predictions already covers most of the original observation region, and so in the new observation region d and thus the entire measure does not change much. Even the massive boost to scores to predictions with overburn due to decreased d is not able to propel them past the ones with underburn. However, there is still no indication that Rao even in this scenario takes into account if the predicted and actual fire perimeters have similar shapes (see how Bridge and Sharp do relatively well).

6.10 Yule, Phi

As mentioned earlier, the Yule coefficient may be misleading at first glance, because it measures the correlation between the predicted and actual fire perimeters rather than the prediction performance directly. For instance, the Bridge and Sharp predictions under the original observation region have much underburn but still receive relatively high scores close to 1 under Yule. Similarly, for Chariot and Rim under the original observation region with

much overburn receive fairly high scores, because the prediction, being a subset within, is highly correlated with the actual perimeter. Also, since this measure is influenced by d , something like the Mountain prediction with the original observational area which should have a highly negative score, may even end up getting a positive score.

One of the oddities with the Yule measure under the reduced observation region is that the Pfeiffer prediction score is less positive than that of Sharp, unlike with the original observation area. This is perhaps because of d dropping more heavily in Pfeiffer than in Sharp with the switch to a smaller boundary box. Also, like under the original observation region, predictions with overburn (such as Chariot and Rim) score highly positive. This is not necessarily wrong but may be misleading if the reader is unfamiliar with Yule.

These observations also apply to the Phi coefficient operating under the original observation region, since the trends in how the predictions compare to each other are similar to those observed in Yule.

6.11 Kappa

The Kappa results are largely similar to those of Phi under the original observation area, except for Bridge, Rim, and Sharp. Bridge demonstrates how underburn is punished somewhat heavily, most likely due to the large b value relative to a , c , and d - the fraction in the Kappa coefficient is high and thus, the entire sum when inverted is low. Rim demonstrates how overburn is punished heavily, probably due to the high c and d which lead to a large numerator in the fraction and thus when inverted, leads to a low score. Sharp is a less extreme version of Rim. Overall, Kappa behaves much like Yule and Phi but punishes overburn more heavily making it potentially a good measure to use. However, this measure still suffers from heavy influence of d as seen with Mountain, which should receive a negative Kappa score but does not due to the large d . While the influence of d is heavily curbed when using a reduced observation region, it still tends to give overburn predictions a relatively positive score, which can be misleading to those unfamiliar with Kappa.

While scores for Kappa are similar to those of Phi under the reduced observation region,

predictions with severe overburn or underburn seem to be punished more heavily. This is likely due to the structure of the coefficient which reacts more heavily when less matching area (here, from reduced d) is present. This means that the good predictions (such as Pfeiffer) are shown to be much better than the merely okay (Rim, Bridge, or Carstens) and the bad ones (Mountain), which would seem to indicate the Kappa coefficient may be especially useful for evaluating the predictions of different models.

CHAPTER 7

Conclusion

In conclusion, this paper presented and adapted fourteen different existing distance metrics for use in wildfire study, each with their own strengths and weaknesses, to give researchers and authorities more systematic tools (rather than methods such as visually comparing the predicted and actual wildfire perimeters) to better evaluate the prediction performance of various wildfire models. Using these distance metrics to summarize the performance of predictions in just a single number has its merits - a well-designed metric can effectively communicate to audiences who may not be familiar with wildfires and their prediction just how well a prediction model works.

These fourteen distance metrics are also particularly useful because not only are they relatively easy to implement, but they are guaranteed to work as long as one has data on the actual wildfire perimeter and can generate a predicted perimeter on the same observation region, so as to calculate the amount of overlapping and non-overlapping areas. While this paper presents how predictions under the FARSITE software model fare under these distance metrics, it would not be difficult to apply the predictions of other prediction models to the distance metrics, as long as they output predicted wildfire perimeters.

Also, different groups of people (whether they be researchers, authorities, or even hobbyists) have different criteria when it comes to evaluating how well a prediction model does in actually predicting a wildfire. The fourteen distance metrics should easily cover the vast range of requirements of these different groups, so anybody can just choose among the distance metrics the one(s) that best suit their needs.

That being said, there are numerous avenues of further research regarding evaluation metrics that can be explored in the future. In this paper, we have adapted fourteen existing

distance metrics for use in wildfire prediction evaluation. However, there are many more distance metrics that can be found in the literatures of various fields that can potentially also be utilized for assessment of wildfire predictions further research is needed to see if any of these other potential measures would be an even more “perfect” way to summarize the performance of a particular model, and if there are any models that suit any needs that may not be currently covered under the fourteen distance metrics.

Another aspect worth further looking into is to see how these distance metrics would perform over a greater variety of wildfire data. While our sample of ten different wildfires do cover a wide range of possible outcomes in terms of predicted and actual fire perimeters, they of course do not cover all possibilities. It is entirely possible that certain weaknesses and strengths of these distance metrics were not revealed with our limited sample, would be brought to light with further empirical wildfire data.

Appendix A

Appendix

		Actual		
		1	0	
Predicted	1	a	b	$(a + b)$
	0	c	d	$(c + d)$
		$(a + c)$	$(b + d)$	$(a + b + c + d)$

Figure A.1: Occurrence Matrix

All of the following figures are generated by the R statistical software unless otherwise noted.

For Figures A.2 - A.12, all plots on the left are the prediction of the fire perimeter, while the plots on the right depict the actual fire perimeter, unless otherwise noted. T represents a hypothetical town in the area.

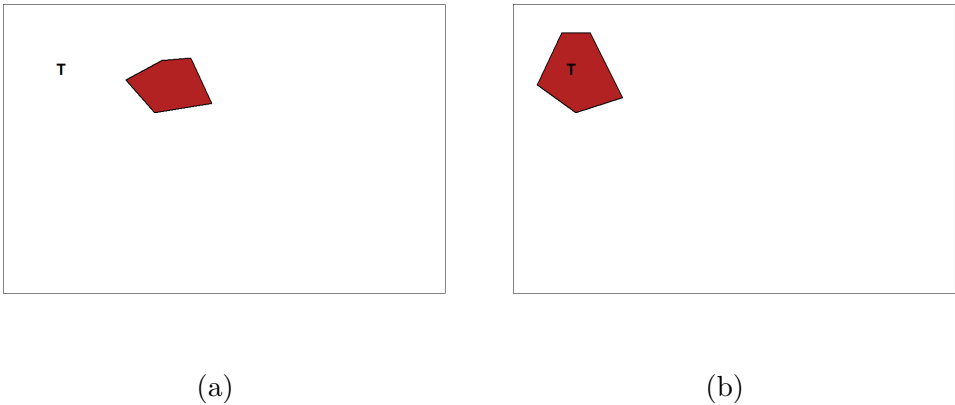
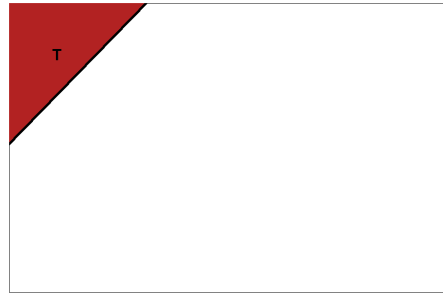


Figure A.2: Simple Matching and Rogers Theoretical Demonstration



(a)



(b)

Figure A.3: Hamann Theoretical Demonstration (1)

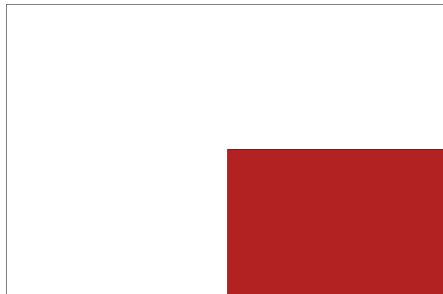


(a)

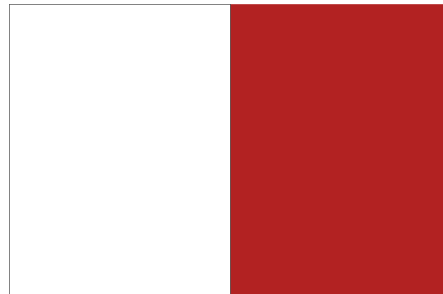


(b)

Figure A.4: Hamann Theoretical Demonstration (2)

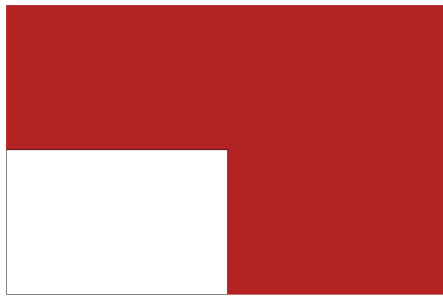


(a)

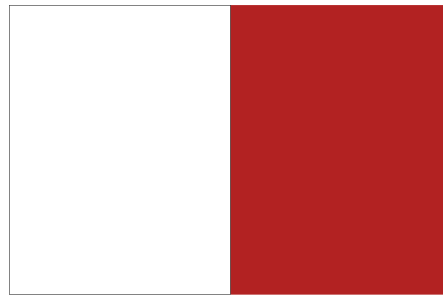


(b)

Figure A.5: Hamann Theoretical Demonstration (3)

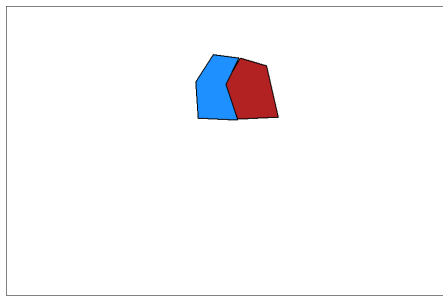


(a)

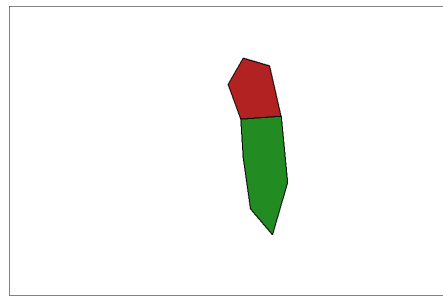


(b)

Figure A.6: Hamann Theoretical Demonstration (4)

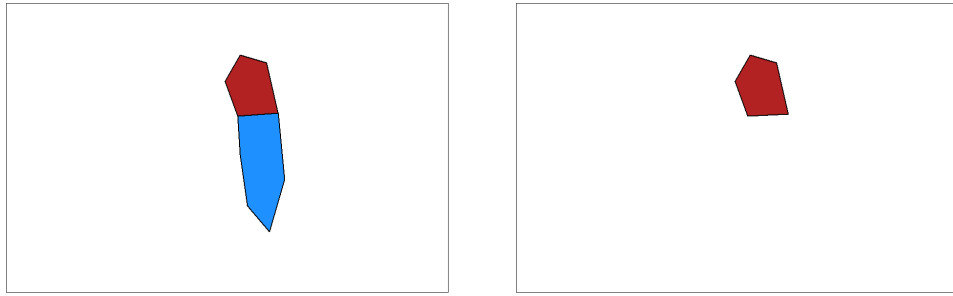


(a)



(b)

Figure A.7: Simpson Theoretical Demonstration



(a)

(b)

Figure A.8: Braun, Kulczynski, Dice, Jaccard, Ochiai, and Sneath Theoretical Demonstration

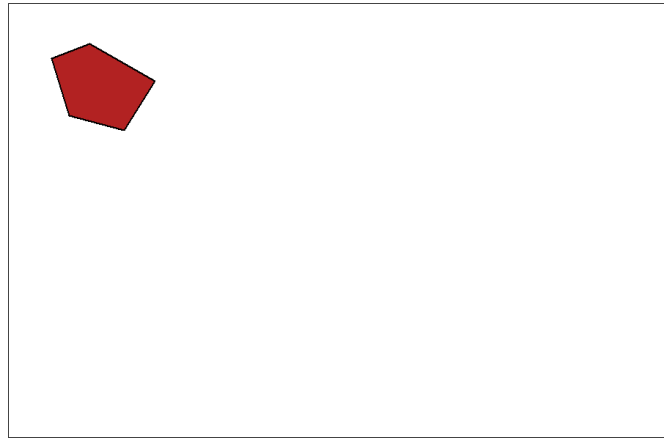


Figure A.9: Rao Theoretical Demonstration (1)

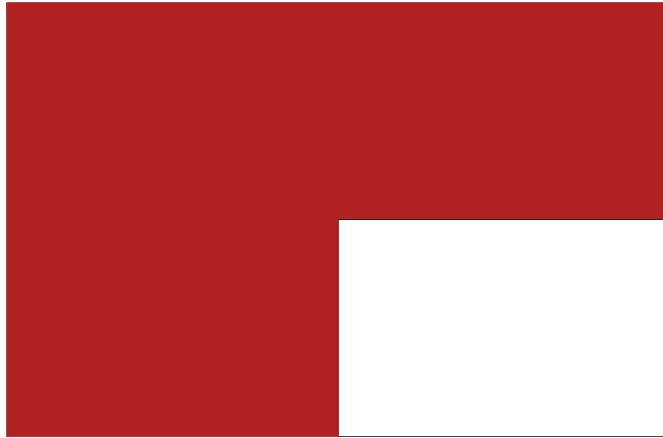
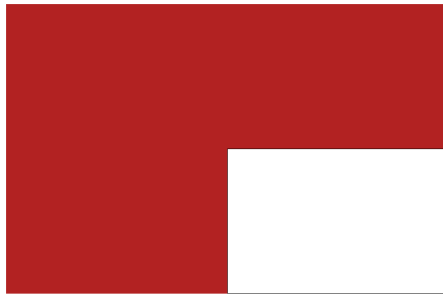
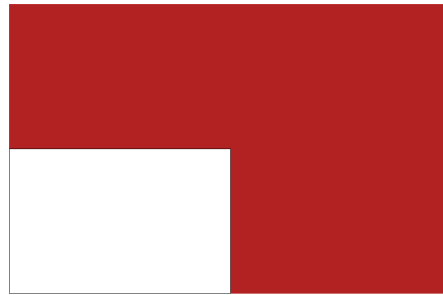


Figure A.10: Rao Theoretical Demonstration (2)



(a)



(b)

Figure A.11: Yule, Phi, and Kappa Theoretical Demonstration (1)

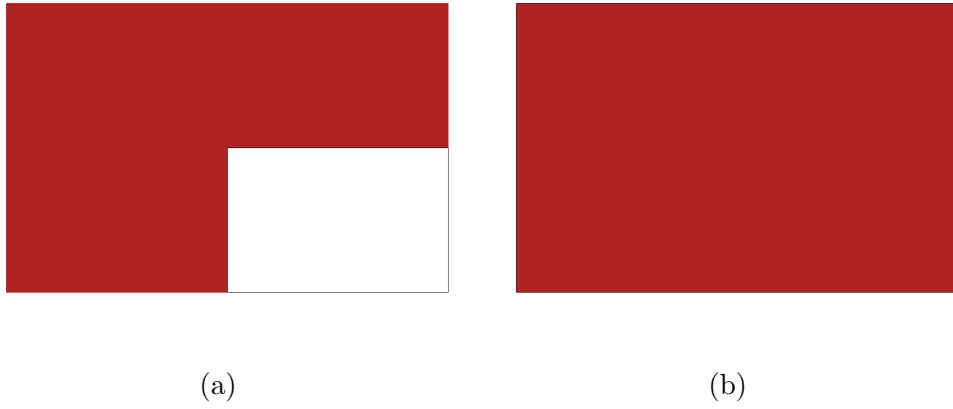


Figure A.12: Yule, Phi, and Kappa Theoretical Demonstration (2)

In the following figures A.13 - A.32, the green area is the predicted fire perimeter. The red represents the actual fire perimeter, and the blue represents the overlap between the predicted and actual fire perimeters.

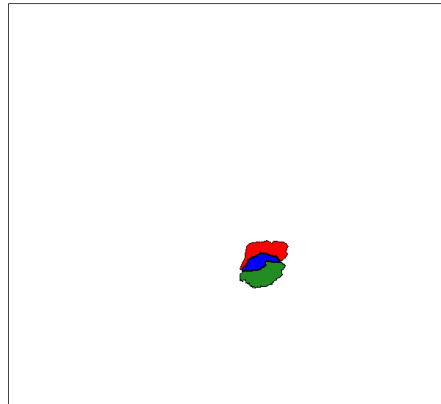


Figure A.13: Aspen fire, original observation region

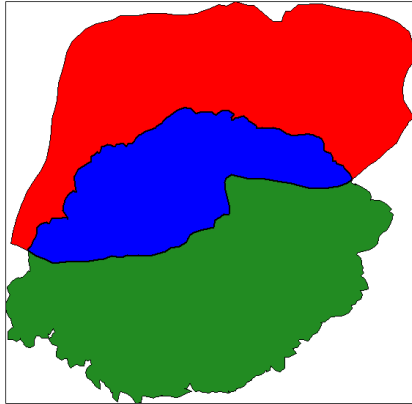


Figure A.14: Aspen fire, smallest observation region possible

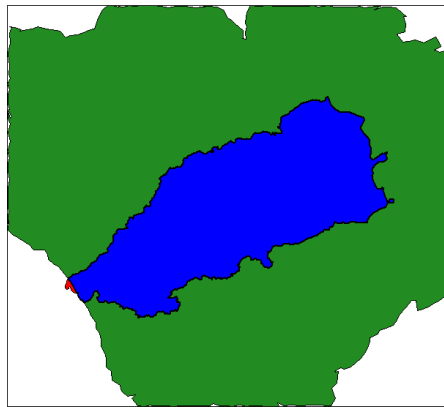


Figure A.15: Bridge fire, original observation region

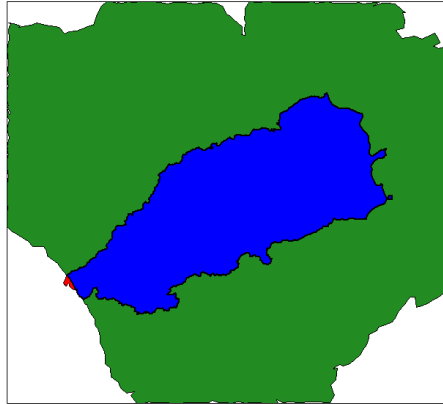


Figure A.16: Bridge fire, smallest observation region possible

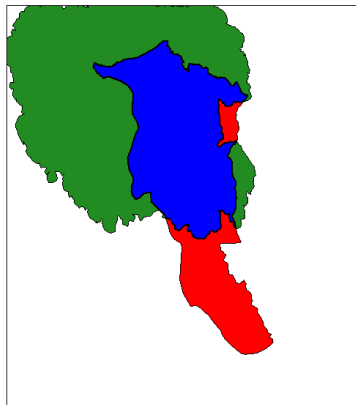


Figure A.17: Carstens fire, original observation region

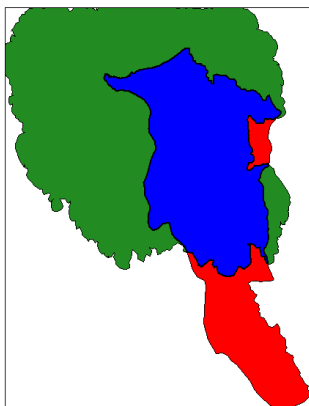


Figure A.18: Carstens fire, smallest observation region possible

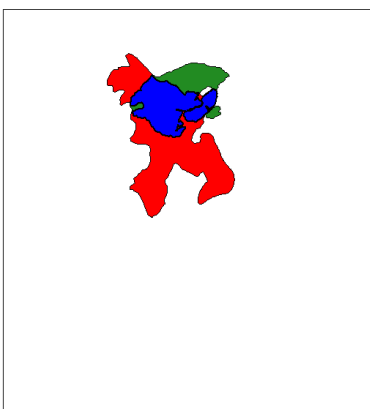


Figure A.19: Chariot fire, original observation region

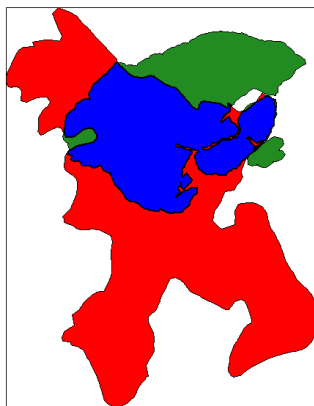


Figure A.20: Chariot fire, smallest observation region possible

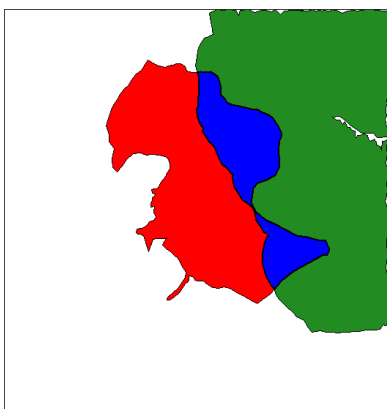


Figure A.21: Gobblers fire, original observation region

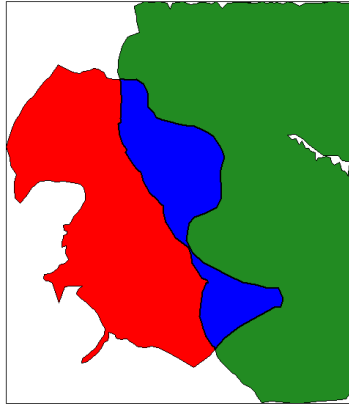


Figure A.22: Gobblers fire, smallest observation region possible

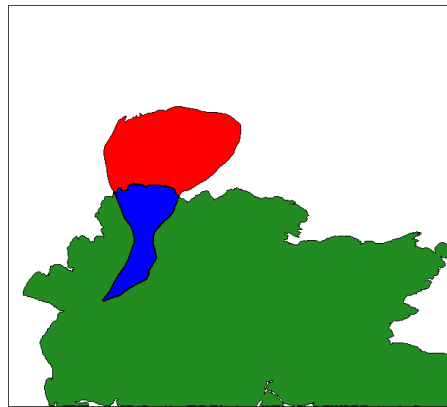


Figure A.23: Hathaway fire, original observation region

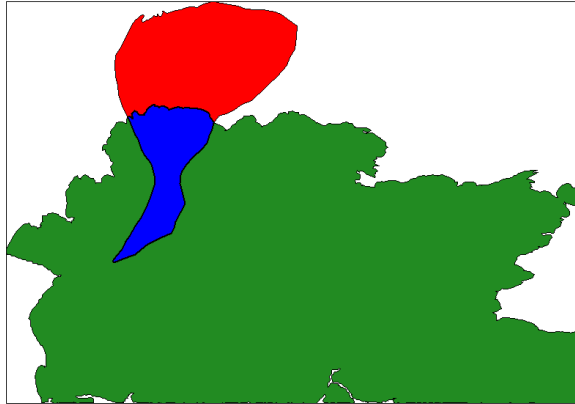


Figure A.24: Hathaway fire, smallest observation region possible

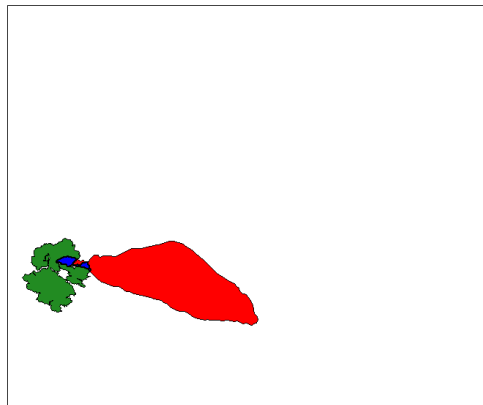


Figure A.25: Mountain fire, original observation region

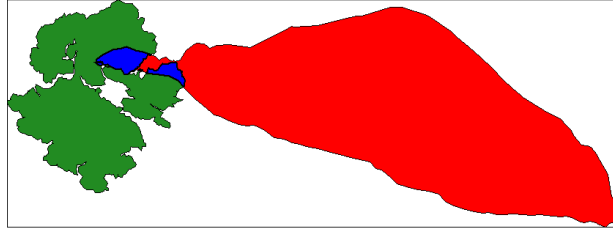


Figure A.26: Mountain fire, smallest observation region possible

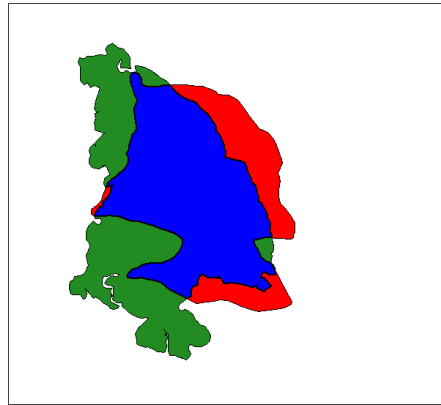


Figure A.27: Pfeiffer fire, original observation region

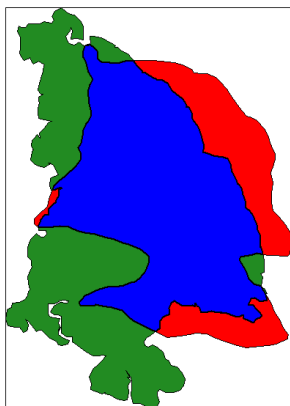


Figure A.28: Pfeiffer fire, smallest observation region possible

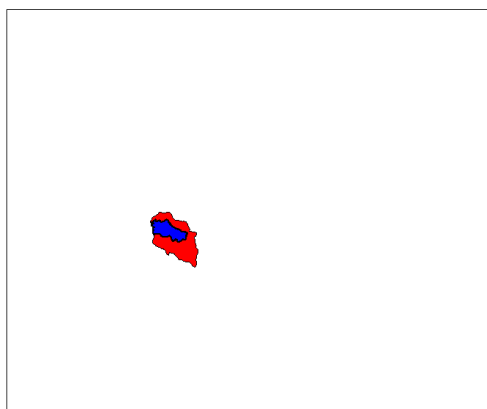


Figure A.29: Rim fire, original observation region

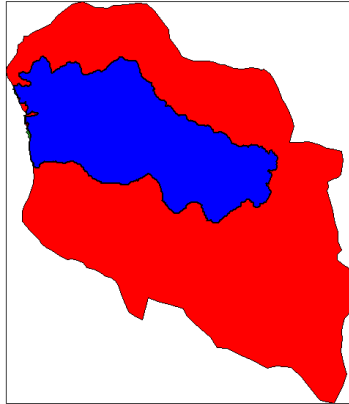


Figure A.30: Rim fire, smallest observation region possible

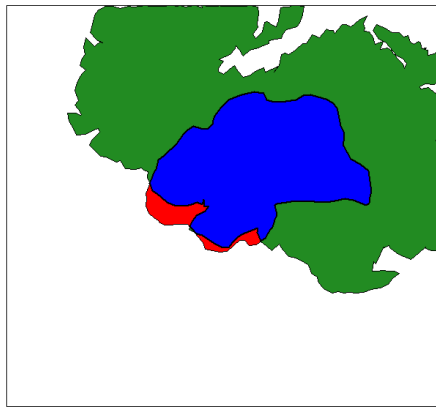


Figure A.31: Sharp fire, original observation region

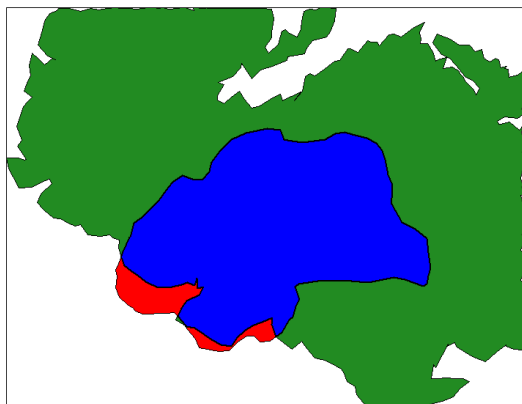


Figure A.32: Sharp fire, smallest observation region possible

Table A.1: Distance Metrics Results under original observation area

Metric	Aspen	Bridge	Carstens	Chariot	Gobblers	Hathaway	Mountain	Pfeiffer	Rim	Sharp
Simple Matching	0.9924	0.3421	0.7435	0.9464	0.6108	0.5640	0.9478	0.8684	0.9941	0.6828
Rogers	0.9848	0.2064	0.5917	0.8983	0.4397	0.3927	0.9008	0.7674	0.9882	0.5183
Hamman	0.9847	-0.3157	0.4870	0.8928	0.2216	0.1280	0.8956	0.7368	0.9882	0.3655
Simpson	0.3534	0.9978	0.6909	0.6459	0.2989	0.2946	0.0646	0.7290	0.9977	0.9259
Braun	0.3334	0.2198	0.3529	0.3069	0.1718	0.0512	0.0227	0.5829	0.2696	0.2798
Kulczynski	0.3434	0.6088	0.5219	0.4764	0.2353	0.1729	0.0437	0.6559	0.6336	0.6029
Dice	0.3431	0.3602	0.4762	0.4161	0.2182	0.0873	0.0336	0.6478	0.4245	0.4298
Jaccard	0.2071	0.2197	0.3048	0.2627	0.1225	0.0456	0.0171	0.4791	0.2694	0.2737
Ochiai	0.3433	0.4683	0.4938	0.4452	0.2266	0.1228	0.0383	0.6518	0.5186	0.5090
Sneath	0.1155	0.1234	0.1798	0.1512	0.0652	0.0233	0.0086	0.3150	0.1557	0.1585
Rao	0.0020	0.1852	0.1125	0.0191	0.0543	0.0208	0.0009	0.1210	0.0022	0.1195
Yule	0.9854	0.9817	0.7450	0.9503	-0.0493	0.2597	0.2521	0.9174	0.9999	0.9162
Phi	0.3394	0.2034	0.3523	0.4217	-0.0175	-0.0631	0.0150	0.5735	0.5171	0.3882
Kappa	0.3393	0.0805	0.3209	0.3917	-0.0164	-0.0378	0.0131	0.5681	0.4225	0.2887

Table A.2: Distance Metrics Results under smallest possible observation area

Metric	Aspen	Bridge	Carstens	Chariot	Gobblers	Hathaway	Mountain	Pfeiffer	Rim	Sharp
Simple Matching	0.4114	0.3420	0.6037	0.6173	0.3428	0.3975	0.5072	0.6764	0.5604	0.4862
Rogers	0.2590	0.2063	0.4324	0.4464	0.2069	0.2481	0.3397	0.5110	0.3892	0.3212
Hamman	-0.1772	-0.3160	0.2075	0.2346	-0.3144	-0.2049	0.0144	0.3527	0.1207	-0.0276
Simpson	0.3534	0.9978	0.6909	0.6459	0.2989	0.2946	0.0646	0.7290	0.9977	0.9259
Braun	0.3334	0.2198	0.3529	0.3069	0.1718	0.0512	0.0227	0.5829	0.2696	0.2798
Kulczynski	0.3434	0.6088	0.5219	0.4764	0.2353	0.1729	0.0437	0.6559	0.6336	0.6029
Dice	0.3431	0.3602	0.4762	0.4161	0.2182	0.0873	0.0336	0.6478	0.4245	0.4298
Jaccard	0.2071	0.2197	0.3048	0.2627	0.1225	0.0456	0.0171	0.4791	0.2694	0.2737
Ochiai	0.3433	0.4683	0.4938	0.4452	0.2266	0.1228	0.0383	0.6518	0.5186	0.5090
Sneath	0.1155	0.1234	0.1798	0.1512	0.0652	0.0233	0.0086	0.3150	0.1557	0.1585
Rao	0.1537	0.1852	0.1737	0.1363	0.0917	0.0288	0.0086	0.2976	0.1621	0.1936
Yule	-0.3715	0.9816	0.5022	0.4804	-0.6101	-0.5520	-0.8292	0.6541	0.9949	0.7600
Phi	-0.1894	0.2033	0.2303	0.2099	-0.3133	-0.1776	-0.2522	0.3628	0.3565	0.2606
Kappa	-0.1892	0.0804	0.2014	0.1819	-0.2810	-0.0951	-0.2022	0.3552	0.2266	0.1600

Table A.3: Summary of Distance Metric Characteristics

	Simple Matching	Rogers	Hamann	Simpson	Braun	Kulczynski	Dice	Jaccard	Ochiai	Sneath	Rao	Yule	Phi	Kappa
Dependence on boundary box	Y	Y	Y	N	N	N	N	N	N	N	Y	Y	Y	Y
Takes rotations into account	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N
Punishes overburn	N	N	N	N	Y	N	Y	Y	Y	Y	Y	N	N	Y
Punishes underburn	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	Y/N (depending on obs. area)	Y
Considers if overall sizes are similar	N	Y	Y	N	Y	N	Y	Y	Y	Y	Y/N (depending on obs. area)	N	N	Y
Considers if overall shapes are similar	N	N	N	N	Y	Y	Y	Y	Y	Y	N	N	N	Y
Considers if main directions are similar	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

REFERENCES

- [1] 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
- [2] Rothermel, R.C. 1972. A mathematical model for predicting fire spread in wildland fuels. *USDA For. Serv. Res. Pap.* INT-115.
- [3] Van Wagner, C.E. (1977). Conditions for the start and spread of crownfire. *Canadian Journal of Forest Research*7: 23-24.
- [4] Rothermel, R.C. 1991. Predicting behavior and size of crown fires in the northern Rocky Mountains. *USDA For. Serv. Res. Pap.* INT-438.
- [5] Albini, Frank A.; Reinhardt, Elizabeth D. 1995. Modeling ignition and burning rate of large woody natural fuels. *International Journal of Wildland Fire*. 5(2): 81-91.
- [6] Albini, Frank A.; Brown, James K.; Reinhardt, Elizabeth D.; Ottmar, Roger D. 1995. Calibration of a large fuel burnout model. *International Journal of Wildland Fire*. 5(3): 173-192.
- [7] Albini, Frank A.; Reinhardt, E.D. 1997. Improved calibration of a large fuel burnout model. *International Journal of Wildland Fire*. 7(1): 21-28.
- [8] 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
- [9] Fitzpatrick, R. (2007, July 14). Huygens' principle. Retrieved April 05, 2018, from <http://farside.ph.utexas.edu/teaching/3021/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] Data Requirements. (n.d.). Retrieved April 05, 2018, from http://www.fire.org/downloads/farsite/WebHelp/usersguide/ug3_data_requirements.htm
- [12] 13 Anderson Fire Behavior Fuel Models. (n.d.). Retrieved April 05, 2018, from <https://www.landfire.gov/fbfm13.php>
- [13] LANDFIRE Partners. (n.d.). Retrieved April 06, 2018, from https://www.landfire.gov/lf_partners.php
- [14] LANDFIRE: Home. (n.d.). Retrieved April 06, 2018, from <https://www.landfire.gov/index.php>

- [15] Warrens, M. J. (2008). *Similarity coefficients for binary data: properties of coefficients, coefficient matrices, multi-way metrics and multivariate coefficients*. Psychometrics and Research Methodology Group, Leiden University Institute for Psychological Research, Faculty of Social Sciences, Leiden University.
- [16] Yule, G. U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75, 579-652.
- [17] Zysno, P. V. (1997). The modification of the phi-coefficient reducing its dependence on the marginal distributions. *Methods of Psychological Research Online*, 2, 41-52.
- [18] Rogers, D. J., & Tanimoto, T. T. (1960). A computer program for classifying plants. *Science*, 132(3434), 1115-1118.
- [19] Hamann, U. (1961). Merkmalsbestand und Verwandtschaftsbeziehungen der Farinose. Ein Betrag zum System der Monokotyledonen. *Willdenowia*, 2, 639-768.
- [20] Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, 241(1), 1-31.
- [21] Braun-Blanquet, J. (1932). Plant sociology. *The study of plant communities*. Plant sociology. *The study of plant communities*. First ed.
- [22] Kulczynski, S. (1927). Die Pflanzenassociationen der Pienenen. *Bulletin International de LAcademie Polonaise des Sciences et des Letters, Classe des Sciences Mathematiques et Naturelles, Serie B, Suppl'ement II*, 2, 57-203.
- [23] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
- [24] Sorenson, T. (1948). A method of stabilizing groups of equivalent amplitude in plant sociology based on the similarity of species content and its application to analyses of the vegetation on Danish commons. *K Dan Vidensk Selsk Biol Skr*, 5, 1-34.
- [25] Gleason, H. A. (1920). Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club*, 47(1), 21-33.
- [26] Bray, J. R. (1956). A study of mutual occurrence of plant species. *Ecology*, 37, 21-28.
- [27] Jaccard, P. (1912). The distribution of the flora in the Alpine zone. *The New Phytologist*, 11(2), 37-50.
- [28] Driver, H. E., & Kroeber, A. L. (1932). Quantitative expression of cultural relationship. *The University of California Publications in American Archaeology and Ethnology*, 31, 211-256.
- [29] Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Bulletin of the Japanese Society for Fish Science*, 22, 526-530.

- [30] Sokal, R. R., & Sneath, P. H. (1963). *Principles of Numerical Taxonomy*. San Francisco: W. H. Freeman and Company.
- [31] Russel, P. F., & Rao, T. R. (1940). On habitat and association of species of Anopheline larvae in South-Eastern Madras. *Journal of Malaria Institute India*, 3(1), 153-178.
- [32] Yule, G. U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society, A*, 194, 257-319.
- [33] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- [34] Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, 85(3), 257-268.
- [35] Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38, 1409-1438.
- [36] Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846-850.
- [37] Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75(6), 579-652.
- [38] Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, 85(3), 257-268.
- [39] Hao, Y. (2018). *California wildfire spread prediction using FARSITE and the comparison with the actual wildfire maps using statistical methods* (Order No. 10790287). Available from Dissertations & Theses @ University of California; ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (2050095894). Retrieved from <https://search.proquest.com/docview/2050095894?accountid=14512>
- [40] Alexander, M. E., & Cruz, M. G. (2013). Limitations on the accuracy of model predictions of wildland fire behaviour: a state-of-the-knowledge overview. *The Forestry Chronicle*, 89(3), 372-383.
- [41] 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/content=842017>